

岭南经济论坛

(第二辑)

舒元 王珺 陈平 主 编

中山大学出版社

·广州·

序 言

2002年我们编辑出版了本院教师的第一本论文集《岭南经济论坛》，当时没有考虑到这么快又要再次出版系列的论文集，因而也没有称之为第一辑；但是，由于上次的论文集在学术界引起了较大的反响，并且，在短短的两年时间内我院教师又发表了大量高质量的论文，这激励我们将原来的论文编辑工作继续下去，并把上次的论文集视为第一辑，而以后的依次称为第二辑、第三辑……从而形成系列。

实际上，自2002年11月以来，我们教师又发表了近300篇学术论文。其中，在《经济研究》、《管理世界》等权威杂志中就发表了近30篇，在《世界经济》、《金融研究》、《经济学动态》、《国际金融研究》、《经济学（季刊）》、《数量经济技术经济研究》、《统计研究》、《财政研究》等全国性重要期刊上发表了近80篇文章；在国外杂志上也有10多篇。在这些文章中，我们根据刊物的影响因子对论文进行了筛选，同时补添了前一辑中遗漏的一些发表于权威刊物上的文章，这样，本辑我们一共选编了68篇论文。最后，根据我院的学科特点以及所收集论文的内容分布，我们将其粗分成六大部分，以便读者能够各择所需。这里我们将所选论文的内容简要说明如下：

第一部分是有关政治经济学与制度经济学的内容，该部分收集了15篇文章。储小平的“家族企业研究：一个具有现代意义的话题”梳理了家族企业研究的重要文献，并揭示了其所具有的现代理论价值和实践意义，提出了自己的研究思路。王珺的“企业簇群的创新过程研究”分别从政府、公司化工作和市场扩散三方面探讨了企业创新的动力和途径，并提供了案例分析。张建琦和黄文锋的“职业经理人进入民营企业影响因素的实证研究——对经理人进入意愿的分析检验”根据抽样调查对经理人进入民营企业的影响因素进行了实证分析。孙洛平的“竞争力与企业规模无关的形成机制”使用边际分析方法提出了竞争力的提高与企业的规模无关的看法。陈宏辉等的“企业利益相关者三维分类的实证分析”借鉴西方学者常用的“多维细分法”和“米切尔评分法”，研究我国企业利益相关者的分类问题。储小平的“职业经理与家族企业的成长”通过调研发现：相当多的私营家族企业的成长瓶颈主要不是金融资本，而是管理资源这种最重要的人力资本。储小平和李怀祖的“信任与家族企业的成长”通过对家族企业成长过程的分析，论证家族企业成长瓶颈的实质是信任资源约束，进而对华人社会的信任资源及其结构对华人家族企业成长的影响进行了剖析。张建琦和汪凡的“民营企业职业经理人流失原因的实证研究——对广东民营企业职业经理人离职倾向的检验分析”通过实证分析指出了影响民营企业经理人流失的5个最重要因素依次为：企业的发展前景，工资福利水平，收入分配与晋升的公平性，职业发展的机会，雇主对经理人的信任与信用。杨永福等的“经济交易规则发生机制的信息流

程解构”沿着信息演化的路径，即从数据上升为公共知识的路径，对交易规则发生机制进行基于信息流程的解构性分析。陈宏辉等的“企业社会责任观的演进与发展：基于综合性社会契约的理解”从历史的视角审视了企业社会责任观的演进，发现企业社会责任与利益相关者的利益要求是紧密结合在一起的。孙洛平等的“产业集群的分工优势”说明，我国产业集群的竞争优势来源有其特殊性，这就是它能够在市场规制和企业制度不完善的情况下促进企业间的分工深化。杨永福等的“制度性规则建构方法研究综述”介绍和评述其中最具代表性的、影响广泛的几种从制度与规则建构的过程来寻找不同国家在人力资本的发展水平和物质资本的积累方面所出现的速度和效果差异的原因的方法。王珺等的“社会资本结构与民营企业成长”以个人关系资本与集群网络资本构成的社会资本结构为理论工具，通过两者的禀赋差异与组合，将我国经济转轨时期民营经济成长的环境分为四种类型。王则柯的“交易虽然互利，但是未必公平”认为，交易的确有一个公平与否的问题，即交易利益是否能够公平分割；在发达国家与欠发达国家的双边贸易中，发达国家往往处于垄断的位置，取得大部分交易利益。李胜兰的“国家所有权客体分类与主体界定——一种经济分析的思路”结合我国物权立法和国有产权改革的实践，运用制度经济学的理论和方法，从国家所有权制度创新的角度，在肯定我国国家所有权物权立法模式的前提下，就国家所有权如何进行客体分类和主体界定，提出了具体的思路和途径。

第二部分是有关宏观经济增长与改革方面的内容，该部分收集了6篇文章。王珺的“增长取向与适应性调整：对地方政府行为演变的一种理论解释”以增长为取向的适应性调整是20多年来我国地方政府行为演变的基本轨迹。王曦的“当期收入还是永久收入：转型时期中国居民消费模式的检验”根据标准的实证检验程序检验了我国居民的消费模式，发现有证据显示，总消费对当期收入是过度敏感的。徐现祥和李郁的“中国城市经济增长的趋同分析”采用趋同分析的标准方法，就我国216个地级及其以上城市展开讨论，结果发现我国城市层面上同时存在新古典增长理论和新增长理论所强调的趋同机制。徐勇和孙洛平的“市场规制和企业制度不完善条件下的经济增长机制探讨”指出“人际关系网”看似不利于市场经济的发展，但在一定条件下可以形成更高的竞争力。孙群燕和李杰等的“寡头竞争情形下的国企改革——论国有股份比重的最优选择”通过构建一个混合寡头垄断竞争的博弈模型，分别就封闭经济和开放经济的情形，探讨国企在背负一定的社会性负担的条件下，国企内部国有股份比重的变化如何影响整个社会的总体福利以及政府的支付。李广众和王美今的“金融中介发展与经济增长：中国案例研究与国际比较”引入VaR系统中协整向量的确定与相应的因果关系分析方法研究了我国金融中介发展的不同方面与经济增长之间的关系。

第三部分是有关国际贸易与投资方面的内容，该部分收集了8篇文章。黄静波和付建的“FDI与广东技术进步关系的实证分析”实证分析了FDI对广东技术进步的作用，认为从总体上看FDI对广东省在技术进步方面的促进作用不明显。喻世友等的“跨国公司R&D投资的国别选择”在实证分析的基础上，分析了跨国公司在中国R&D投资的区位现状，提出了中国在吸引跨国公司R&D投资的政策建议。许罗丹和谭卫红的“外商直接投资聚集效应在我国的实证分析”借助邓宁的区位优势理论，构建局部调整模型，分析FDI的聚集效应，以及其他影响外商在我国进行直接投资的因素。许罗丹等的另一篇文章“四组外

商投资企业技术溢出效应的比较研究”通过对广东省华资、欧盟、日资、美资四组外商投资企业的调查数据,总结外商直接投资溢出效应的途径,试图从华资、欧盟、日资、美资四个投资方,对比分析这些外商直接投资在产品水平、R&D投入、上游产品来源、员工素质、培训方面对我国企业的技术溢出。黄静波的“新自由主义贸易政策论及其演进”梳理了具有重要影响的新自由主义代表人物的观点,并分析了这些新自由主义者有关的贸易政策理论的主要观点。许罗丹等的“对外直接投资理论的发展”从直接投资与国际贸易的关系、直接投资的动机与成因分析、直接投资对东道国的影响三个方面梳理了对外直接投资理论的发展。黄静波和赖伟娟的“广东省出口产业比较优势分析”通过使用显示性比较优势指数和竞争力指数研究广东省对外贸易行业比较优势的变化情况。李杰和孙群燕的“从啤酒市场整合程度看WTO对消除地方保护的影响”运用协整分析的方法,以啤酒行拥作为切入点,具体考察我国啤酒市场从1997年1月至2003年3月的整合情况。

第四部分是有关税收与财政方面的内容,该部分收集了8篇文章。陈平和黄健梅的“我国出口退税效应分析:理论与实证”通过引入实际有效汇率的概念,从理论上系统阐述出口退税对出口盈利性,继而对出口规模的作用;同时运用基于ECM模型的协整分析、Panel Data分析等实证分析工具,研究了我国出口退税政策对出口的促进作用。杨卫华和尹淑平的“发展中小企业 亟需改革税制”分析了现行与中小企业发展相关的税收制度存在的问题,并提出了完善税制促进中小企业快速发展的构想。郭小东和王倩的“优化纳税服务与交易费用的降低”论证了纳税服务在降低交易费用方面所发挥作用的几个方面,并提出了优化纳税服务的相关政策建议。林江和王雅莉的“国有资产管理的改革方向”尝试通过综合分析国有企业经营绩效和股权结构、公司规模、行业和经营年限等因素之间的关系,探讨国有资产管理的改革方向。郭小东的“论公共部门运行中交易费用合理的‘度’”分析了在我国当前进行的公共财政建设中如何合理看待、把握交易费用问题以及如何控制好交易费用的“度”的问题。郭小东和王倩的“网络经济中的屏蔽因子与政府规制”从“屏蔽因子”的角度对电子商务中的信息不对称现象进行分析,并对政府应该采取的措施提出建议。张勇和王美今的“中国企业年金税收优惠政策的研究——我国企业年金税收支出的精算统计分析”运用精算技术构建了我国企业年金税收支出的测算模型,提供了减少这种偏差的解决办法。夏南新的“税收诱致性现金持有量模型因果性检验及对我国地下经济规模的估测”指出地下经济存在和发展的主要原因在于居高不下的税率,由于偷逃漏税是地下经济活动参与者获取非法收入的主要动机,所以,一个拟合优良的通货需求模型应当将纳税负担作为一个关键的解释变量引入。

第五部分是有关金融与资本市场方面的内容,该部分收集了16篇文章。陈浪南和屈文洲的“资本资产定价模型的实证研究”运用上海股票市场的数据,对资本资产定价模型进行实证检验,尤其注重在CAPM模型中起着举足轻重的 β 值的分析和测量上。宾国强和舒元的“股权分割、公司业绩与投资者保护”构建了一个更为广泛的上市公司1995~2001年的面板数据样本,对上市公司股权分割与公司业绩的关系以及公司上市后的业绩变迁进行了实证研究。王美今和孙建军的“中国股市收益、收益波动与投资者情绪”从我国股市的现实情况出发,构造理论模型证明:投资者接受价格信号时表现出来的情绪是影响均衡价格的系统性因子。李涛的“商业银行监管的国际比较:模式及影响——兼论中国的商业

银行监管模式选择”系统地研究了世界上 118 个国家的商业银行监管模式及其对金融发展水平的影响。陈浪南和黄洵的“联合外汇干预的实证研究”采用干预分析模型来考察日、美两国央行对日元/美元汇率进行联合干预的效果，以得到外汇干预是否对汇率产生影响以及该影响能否持久的结论。邹功达和陈浪南的“中国 A 股与 B 股的市场分割性检验”以布莱克版 CAPM 作为理论模型，修正国外先进的市场分割检验模式，运用较复杂的计量经济估计方法对我国 A、B 股市场的一体化（或分割性）进行实证研究。王宣喻和储小平的“信息披露机制对私营企业融资决策的影响”通过实证研究检验了有关私营企业融资决策的一个假说，即在不存在制度歧视和政策倾斜的前提下，如果不考虑资金的利息成本和执行成本，那么，影响私营企业融资决策的主要因素是各层级资本市场的信息披露机制。黄后川和陈浪南的“中国股票市场波动率的高频估计与特性分析”发现股票指数与个股的高频交易数据中的微观摩擦影响正好相反，使用极高频的数据会大大增加个股的波动率估计值，相反却会大大降低指数的波动率估计值。牛鸿和詹俊义的“中国证券投资基金市场择时能力的非参数检验”运用传统的参数检验方法（TM 模型和 HM 模型）和 Jiang 提出的非参数方法对我国 33 只封闭式证券投资基金的择时能力进行了实证研究，并对研究结果进行了对比。王曦和舒元的“中国货币市场运行：内生性、调整时滞与动态”考虑了我国货币和商品市场的调整时滞，建立了我国货币市场的动态系统理论模型，描述了我国货币市场的运行态势，并讨论了这种运行的含义。陈平和龙华的“中国股市涨跌停绩效的经验分析及政策建议”结合中国股市特征和实际结果，提出了放宽涨跌停幅度和实施盘中暂停交易的政策思考。李涛和李红的“养老金投资规模、投资者保护与股票市场发展——论经合组织国家的经验与中国的选择”以经合组织国家为例，考察了养老金投资规模对投资者保护程度和股票市场发展水平的影响。李仲翔、李仲飞和陆军的“投资基金业的跨界活动与障碍”研究了发达国家投资基金跨界活动的类型、程度和基金跨界活动的法律和监管问题，以及影响跨界活动的因素，最后总结出了解决跨界活动障碍的通行做法。林江和夏育松的“影响港元汇率的货币政策因素分析”通过实证分析指出，在香港的特殊经济和政策环境下，通货膨胀的压力和货币供应量的独立增长均不是导致港元汇率不稳定的因素，同业拆借利率与港元汇率之间却表现出有力的反馈关系，而同业拆借利率正是受香港货币政策工具控制的。劳平的“金融结构历史演变的初步考察”在现有的纵向研究基础上，以交易成本为突破口考察一国金融结构的历史变迁，并得出一国金融结构的演进大致遵循从银行主导型到市场主导型再到相互融合阶段的结论。何兴强的“中国股票市场收益非线性相关结构的经验分析”利用 McLeod-Li, BDS, Hsieh 检验和随机非线性模型，探讨中国股票市场收益非线性相关结构的存在性和基本类型。

第六部分是有关数量经济学方面的内容，该部分收集了 15 篇文章。陆军和钟丹的“泰勒规则在中国的协整检验”在全面分析泰勒规则的理论含义和实际应用的基础上，运用协整分析方法估计我国泰勒规则的具体形式。陈浪南和王艺明的“技术交易规则与超常收益研究”旨在通过 White 真实性检验试算法的应用，实现对数据窥察偏倚的调整，从而能确定技术交易规则所取得的超常收益在多大程度上基于其自身实际预测能力。韦华全和王燕鸣的“Skiba 的一个未解决问题”肯定回答了 Skiba 最近在《The Kourouka Notebook》中提出的一个未解决问题，并获得了比原问题更一般且深刻的结果。杨智元和陈浪南的“基

于跳跃过程的指数期权模型”抓住组成股票价格指数的个股运动变化相对独立的特点，引入跳跃过程描述股票价格的运动变化，并在一定的限制条件下得出指数期权的定价方程及定价模式。李仲飞和姚京的“安全第一准则下的动态资产组合选择”在 Black-Scholes 金融市场设置下，利用 Roy 提出的安全第一（Safety-First）准则，导出了最优常数再调整投资组合投资策略的显式表达式。王艺明和陈浪南的“我国寿险业粗死亡率的实证研究”应用具有转折点的 Logistic 模型对我国寿险业经验生命表作实证分析。孙建军和王美今的“股市政策对个体证券投资者交易行为的影响——行为金融理论对我国股市低迷现状的解释”从行为金融的角度，考察股市政策对风险厌恶的个体证券投资者交易行为的影响。李仲飞和汪寿阳的“EaR 风险度量与动态投资决策”在 Black-Scholes 设置下，给出了均值 - 在险收益意义下的最优常数再调整证券组合投资策略以及有效边界的显式表达式。杨智元、王艺明和陈浪南的“保险精算中单损失和双损失环境的一些重要理论再辨析”就《保险精算中单损失和双损失环境的一些重要理论辨析》探讨的问题提出一些不同看法。姚京和李仲飞的“基于 VaR 的金融资产配置模型”根据均值 - 方差模型的框架，建立了用 VaR 代替方差或标准差作为风险的测量指标时的均值 - VaR 模型，同时使用等 VaR 线分析了两种模型的内在联系。最后是收录了周作领等的 5 篇探讨数理工具的文章，这些研究对混沌经济学的发展具有基础性作用。

《岭南经济论坛》（第二辑）的选编一如既往地得到了岭南学院师生的大力支持与配合，众多师生积极寄来自己的优秀文章，但由于篇幅有限，这里无法完全吸纳，令人不免遗憾，也只有等今后有机会再作弥补了；在出版过程中，中山大学出版社的叶侨健社长、周建华主任也付出了很大的热心和努力，并使《岭南经济论坛》（第二辑）得以在最短的时间内出版。因此，我们在这里对他们表示衷心感谢；同时，也希望广大读者认可和支持我们的工作。

岭南学院
2005 年 1 月 16 日

目 录

第一部分 政治经济学与制度经济学

家族企业研究：一个具有现代意义的话题	储小平/3
企业集群的创新过程研究	王 珺/15
职业经理人进入民营企业影响因素的实证研究	
——对经理人进入意愿的分析检验	张建琦 黄文锋/26
竞争力与企业规模无关的形成机制	孙洛平/35
企业利益相关者三维分类的实证分析	陈宏辉 贾生华/44
职业经理与家族企业的成长	储小平/57
信任与家族企业的成长	储小平 李怀祖/70
民营企业职业经理人流失原因的实证研究	
——对广东民营企业职业经理人离职倾向的检验分析	张建琦 汪 凡/79
经济交易规则发生机制的信息流程解构	杨永福 段红涛 张智革/88
企业社会责任观的演进与发展：基于综合性社会契约的理解	陈宏辉 贾生华/100
产业集群的分工优势	孙洛平 孙海琳/110
制度性规则建构方法研究综述	杨永福 段红涛 孟 磊/116
社会资本结构与民营企业成长	王 珺 姚海琳 赵 祥/124
交易虽然互利，但是未必公平	王则柯/134
国家所有权客体分类与主体界定——一种经济分析的思路	李胜兰/143

第二部分 宏观经济增长与改革

增长取向的适应性调整：对地方政府行为演变的一种理论解释	王 珺/155
当期收入还是永久收入：转型时期中国居民消费模式的检验	王 曦/165
中国城市经济增长的趋同分析	徐现祥 李 郇/177
市场规制和企业制度不完善条件下的经济增长机制探讨	徐 勇 孙洛平/188
寡头竞争情形下的国企改革	
——论国有股份比重的最优选择	孙群燕 李 杰 张安民/193
金融中介发展与经济增长：中国案例研究与国际比较	李广众 王美今/216

第三部分 国际贸易与投资

FDI 与广东技术进步关系的实证分析	黄静波	付建	229
论跨国公司 R&D 投资的国别选择	喻世友	万欣荣	史卫
外商直接投资聚集效应在我国的实证分析	许罗丹	谭卫红	252
四组外商投资企业技术溢出效应的比较研究	许罗丹	谭卫红	刘民权
新自由主义贸易政策论及其演进		黄静波	283
对外直接投资理论的发展	许罗丹	谭卫红	289
广东省出口产业比较优势分析	黄静波	赖伟娟	295
从啤酒市场整合程度看 WTO 对消除地方保护的影响	李杰	孙群燕	305

第四部分 税收与财政

我国出口退税效应分析：理论与实证	陈平	黄健梅	323
发展中小企业 亟需改革税制	杨卫华	尹淑平	335
优化纳税服务与交易费用的降低	郭小东	王倩	341
国有资产管理的改革方向	林江	王雅莉	348
论公共部门运行中交易费用合理的“度”		郭小东	356
网络经济中的屏蔽因子与政府规制	郭小东	王倩	363
中国企业年金税收优惠政策的成本研究			
——我国企业年金税收支出的精算统计分析	张勇	王美今	366
税收诱致性现金持有量模型因果性检验及对我国地下经济规模的估测		夏南新	376

第五部分 金融与资本市场

资本资产定价模型的实证研究	陈浪南	屈文洲	385
股权分割、公司业绩与投资者保护	宾国强	舒元	399
中国股市收益、收益波动与投资者情绪	王美今	孙建军	411
商业银行监管的国际比较：模式及影响			
——兼论中国的商业银行监管模式选择	李涛	423	
联合外汇干预的实证研究	陈浪南	黄洵	435
中国 A 股与 B 股的市场分割性检验	邹功达	陈浪南	445
信息披露机制对私营企业融资决策的影响	王宣喻	储小平	457
中国股票市场波动率的高频估计与特性分析	黄后川	陈浪南	471
中国证券投资基金市场择时能力的非参数检验	牛鸿	詹俊义	481
中国货币市场运行：内生性、调整时滞与动态	王曦	舒元	493
中国股市涨跌停绩效的经验分析及政策建议	陈平	龙华	507

养老金投资规模、投资者保护与股票市场发展 ——论经合组织国家的经验与中国的选择	李 涛	李 红	520
投资基金业的跨界活动与障碍	李仲翔	李仲飞	530
影响港元汇率的货币政策因素分析	林 江	夏育松	535
金融结构历史演变的初步考察	劳 平		544
中国股票市场收益非线性相关结构的经验分析		何兴强	550

第六部分 数量经济学

泰勒规则在中国的协整检验	陆 军	钟 丹	565
技术交易规则与超常收益研究	陈浪南	王艺明	578
Skiba 的一个未解决问题	韦华全	王燕鸣	589
基于跳跃过程的指数期权模型	杨智元	陈浪南	595
安全第一准则下的动态资产组合选择	李仲飞	姚 京	602
我国寿险业粗死亡率的实证研究	王艺明	陈浪南	609
股市政策对个体证券投资者交易行为的影响 ——行为金融理论对我国股市低迷现状的解释	孙建军	王美今	615
EaR 风险度量与动态投资决策	李仲飞	汪寿阳	621
保险精算中单损失和双损失环境的一些重要理论再辨析 ...	杨智元	王艺明	632
基于 VaR 的金融资产配置模型	姚 京	李仲飞	636
黎曼流形上不变集的分形维数估计	瞿成勤	周作领	647
一类广义 Sierpinski 海绵的填充测度	朱智伟	周作领	656
三分 Cantor 集自乘积的 Hausdorff 测度的估计	贾宝国	周作领	666
自相似集的 Hausdorff 测度与连续性	罗 俊	周作领	673
Cantor 集的自乘积集的 Hausdorff 测度的下界	贾宝国	周作领	680

第一部分

政治经济学与制度经济学

家族企业研究：一个具有现代意义的话题^{*}

储小平

据众多课题组的调查报告和专家们的研究：今日中国的私营企业，普遍采用家庭家族拥有的形式，在企业内部的管理上广泛存在着家族制管理^①。可以预料，家族企业将越来越多，其活动会日益成为我国的一个重要经济现象。不仅私营企业普遍存在家族制管理，而且其他类型的企业，如乡镇企业、集体企业、合伙企业、股份合作制企业、民营承包企业等也大量存在家族制或泛家族制管理。改革开放以来，我国一直倡导学习国外先进的管理理论和方法，并提出要建立现代企业制度。谁也没提倡发展家族企业和采用家族制方法来管理企业，谁也没想到家族企业会大量出现。然而，家族企业雨后春笋般地发展起来。长期忽视对家族企业的研究，是大陆经济学和管理学界的一大缺陷。只是在近几年，学界才开始重视家族企业问题，并引发一些争论。但是，如何认识家族企业？家族企业是不是一种有效率的组织形式？家族企业应如何向现代企业组织形式演变？这些问题尚须进行深入的探讨。

一、家族企业研究：一个不容忽视的重要课题

在现代人的印象中，家族企业是一种落后的企业形式，用家族的规则来管理企业也是一种落伍的管理方法。但是，无论是发达国家还是发展中国家，家族企业都大量顽强地生存和发展着。据克林·盖尔西克等人的研究，最保守的估计也认为由家庭所有或经营的企业在全世界企业中占 65% ~ 80% 之间。世界 500 强企业中有 40% 由家庭所有或经营。家族企业创造了美国生产总值的一半，雇佣的劳动力也占一半。在欧洲，家庭公司支配着中小规模的公司，并在一些国家里占较大公司的大多数。在亚洲各国，家庭公司大都居主导地位。在拉美，由家族建立和控制的大型企业在绝大多数产业部门中都占主导地位^②。著名管理学家德鲁克也指出：大部分企业——包括美国的和所有其他发达国家的——都是由家族控制和管理的。进而，德鲁克还指出了—一个奇怪的现象：有关管理的书籍和课程却几乎完全是针对公共的和专业管理的企业——它们难得提到家族经营的企业^③。由此可见，很长时间，家族企业的研究往往不被人们看成是一个具有现代意义的课题，几乎被人们忽略了。这种状况在我国显得更为明显。

^{*} 原载：《中国社会科学》，2000 年第 5 期；转载：《新华文摘》，2000 年第 10 期。

^① 《中国私有企业主阶层研究》课题组：《我国私有企业主的经营状况与私有企业主的群体特征》，《中国社会科学》，1994 年第 4 期；张厚义，明立志主编：《中国私营企业发展报告》（1978 ~ 1998），社会科学文献出版社 1999 年版。

^② 克林·盖尔西克等：《家族企业的繁衍——家族企业的生命周期》，经济日报出版社 1998 年版，第 2 ~ 3 页。

^③ 彼得·德鲁克：《大变革时代的管理》，上海译文出版社 1999 年版，第 29 页。

近代以来,我国家族企业已有一定的发展。随后,由于外敌的入侵和政治的因素,家庭、家族、家文化、家族企业受到强烈冲击。改革开放前,大陆的家族企业已绝迹。促使人们逐渐重视家族企业研究的原因主要有两点:①90年代以来,港澳台及分布在东南亚和世界其他地区的华人企业及其经济成就特别令世人刮目。约翰·奈斯比特不无耸言地说,世界华人经济对日本的经济地位大有取而代之势,华人企业家团体“把亚洲经济推向巅峰”,“并逐渐侵蚀着西方经济领域”①。虽然海外华人所处的环境并不相同,“他们却都一样地发展并保持了一种特殊的企业组织形式——家族企业”②,有时从表面上看,海外华人的新的跨国集团和其他企业完全一样,比如,它们按公司方式组建,有董事会和公司职员。但它们的运行方式和世界经济中任何东西都完全不同,描述它们的最好方式也许是把它们当作一起营运操作的家族③。②20世纪70年代末至80年初,中国大陆平地涌起家庭经营的浪潮。周其仁比较充分地认识到“家族经营再发现”的学术与实践意义,“家庭这个古老的组织形式,在今天和明天的中国社会主义农业的发展中,正在并将继续焕发出它所容纳的全部活力”④。但是,周其仁并没有把它的研究延伸到“家族制与企业组织”这一课题上。由德国大众汽车基金会资助的,并由德国著名学者何梦笔(Carsten Herrman - Pillath)和中国社会科学院农村发展研究所陈吉元教授共同主持的大型研究项目“中国乡镇企业的领导与决策:文化和经济转变模式”,对家族规则和现代的商业规则的包容问题作了很有价值的探讨⑤。但相关的研究依然很少。李新春敏锐地看到,中国经济学者对家族制度的研究的漠视是令人惊异的,“在缺乏对家族制度深入了解的情况下,对中国经济特别是企业组织的发展将是难以准确把握的”⑥。

对我国而言,研究家族企业有着特殊的理论与实践意义。中国是一个家文化传统最为悠久和深厚的国度,台湾著名学者李亦园认为中国文化是“家的文化”⑦;杨国枢进一步认为:“家族不但成为中国人之社会生活、经济生活及文化生活的核心,甚至也成为政治生活的主导因素。”⑧汪丁丁指出:“从那个最深厚的文化层次中流传下来,至今仍是中国人行为核心的,是‘家’的概念。”⑨费孝通主张要重视家庭的重要作用,“这个细胞有很强的生命力”,农村中“真正有活力的就是家庭工业”⑩。虽然解放以来,家文化、家族组织受到强烈冲击,但是,中国的现实表明:“家族主义”或“泛家族主义”倾向在中国的各种组织或单位中都惊人地相似,并普遍地存在着⑪。即使在计划经济体制最鼎盛的时期

① 约翰·奈斯比特:《亚洲大趋势》,外文出版社、经济日报出版社、上海远东出版社1996年版,第6~12页。

② 雷丁(G. Redding):《海外华人企业家的管理思想——文化背景与风格》,上海三联书店1993年版,第4页(原英文书名“ The Spirit of Chinese Capitalism”,1990)。

③ 彼得·德鲁克:《大变革时代的管理》,上海译文出版社1999年版,第158~157页。

④ 周其仁:《农村变革与中国发展(上卷)》,香港牛津大学出版社1994年版,第100~101页。

⑤ 何梦笔和陈吉元:《当代中国的村庄经济与村落文化丛书》,山西经济出版社1996年版。

⑥ 李新春:《中国的家族制度与企业组织》,《中国社会科学季刊》(香港),1998年秋季卷。

⑦ 李亦园:《中国人的家庭与家的文化》,文崇一,萧新煌主编:《中国人:观念与行为》,台湾远流图书公司1988年版,第113页。

⑧ 杨国枢:《家族历程、泛家族主义及组织管理》,《海峡两岸之组织与管理》,台湾远流出版公司1998年版。

⑨ 汪丁丁:《经济发展与制度创新》,上海人民出版社1995年版,第21页。

⑩ 参看《费孝通、李亦园对话录》,《北京大学学报》(哲社版),1998年第6期。

⑪ 樊江春:《中国微观组织中的“家族主义”》,《新华文摘》,1992年第5期。

也是如此，只不过是有一种变相的或潜伏的形态存在。在体制转轨时期，家文化与家族企业的勃然复发，应是不足为怪的。正如爱森斯塔（Eisenstadt）所言，虽然现代化会削弱旧传统的某些层面，然而在社会变迁的过程中，旧传统的某些层面有时会被再度提出和强调，以解决文化断层的危机和建立新的集体认同^①。在中国社会转型状态中，计划经济体制规则渐渐失效，健全的市场经济规则还在逐步的建立之中。因此，家庭与家族规则自然就成为在夹缝中发展起来的私营企业创建和发展的支撑构件。杰纳曾深刻地指出，20世纪的中国历史固然伤痕累累，唯一比其他机制更强韧、更蓬勃的就是父系制度的中国家庭，因为家庭一向是中国人对抗外在险恶政治环境的避风港^②。

从学理背景看，家文化体现了中国但统文化的突出特征，几千年家文化传统的社会心理积淀对企业的组织与经营行为、对家族企业的生命周期都产生着重大影响。因此，重视对当代中国家族企业，包括海外华人家族企业的研究，可以使人们加深对华人企业组织制度形式与组织行为以及与社会经济制度环境之间的互动变迁关系的认识，可以引发我们去深入探讨中国传统文化规则与现代市场经济规则以及现代企业制度规则的衔接点、结合点，进而可以从学理上探讨富有效率的中国特色的企业管理模式。从这个意义上言，家族企业是一个值得深入研究的课题。

二、家族企业组织的合理性及其效率

人们对家族企业的歧见，主要原因在于人们对这种企业组织的合理性及其管理效率有不同的看法。归纳起来大致有以下三种：

(1) 对家族企业持消极的评价。现代工业和传统家庭是相互抵触的，传统的中国家庭制度下蕴育出的裙带关系对高效率的工业制度是一最大的障碍^③。家族式经营不能摆脱传统规范，他们要想增强自己的竞争能力，就必须改变“企业家族化”的状况^④。家族主义会妨碍企业现代化，不能考虑事业的长期发展，因为过分重视私利，不肯也不愿多花研究开发费用，在技术更新上不能有创新与突破^⑤；用人为私，排斥家族圈外的人才，没有制度，造成较高的流动率；建立在家族关系和传统行为模式上的企业结构不符合现代资本主义的经营管理原则^⑥。

(2) 肯定家族企业的积极意义。注重家族血亲和同乡关系，而不注重企业所有权和经

① 爱森斯塔（Eisenstadt），Tradition, Change and Modernity, New York, 1973, pp209~210。

② 参看福山：《信任——社会道德与繁荣的创造》，远方出版社1998年版，第113页。

③ Levy, Marion, The Family Revolution in Mordern China. New York: Octagon Books, 1963, p354; 赵耀东：《我的经营观》，经济管理出版社1986年版。

④ 刘可复，饶美蛟：《香港华资银行的经营困境与出路》，香港《信报》月刊，1987年第11卷第2期。

⑤ 鲍遵信：《儒家伦理与亚洲四小龙——儒家复兴说驳议》，《改革》，1988年第6期。

⑥ 陆民仁：《人是经济升级的关键》，台湾《中国论坛》，1984年第199期；杨君宏，杜念中：《儒家伦理与经济发展》，台湾，联经出版社1988年版；黄光国：《中国式家族企业的现代化》，见黄光国编：《中国人的权力游戏》，台湾，远流图书公司1988年版。

营权的分离,并没有成为经济发展的障碍^①。家族式组织和企业网络是亚洲经济组织的一种特征,这种特征本身是效率中性的,即家族式组织不一定是低效率的,在特定的情况下甚至比市场或科层更有效率和竞争力^②。家族企业不是落伍的代名词,家族成员的参与常常是创业最需要的低成本组织资源;家族成员更易建立共同利益和目标,从而更易进行合作;家族企业的性质更能保证企业领导的权威,有更强的凝聚力^③。作为一种企业制度的家族制,更适合市场经济的初创阶段^④。

(3)对家族企业持一分为二的评价。大多数学者既肯定家族企业在特定条件下存在的积极意义,又注意分析其消极的一面。对海外华人家族企业深有研究的雷丁从纵向合作、横向合作、控制、适应性四个方面探讨了华人家族企业组织形式的优点和缺点,认为华人家族企业既是一种高效的工具,也可能成为失败的根源^⑤。李新春指出,1978年以来,大陆的家庭制度不是限制,而是大大地加速了改革和为改革作出了贡献,但是家族内外有别的伦理关系会造成企业组织内部的帮派体系和组织内耗^⑥,因此也可能出现效率低下的状况。有的人简要概括为一句话:“成也家族,败也家族。”^⑦

上述每一种见解在历史和现实中都有不胜枚举的案例予以佐证。而对家族企业组织的合理性与管理效率持不同看法的学者大都没有深究家族企业的定义,大概是认为家族企业就是企业资产所有权与经营控制权都归家族所有。但这种认识过于笼统,另一些学者则对家族企业做出了界定。

美国著名企业史学家钱德勒的定义是:“企业创始者及其最亲密的合伙人(和家族)一直掌有大部分股权。他们与经理人员维持紧密的私人关系,且保留高阶层管理的主要决策权,特别是在有关财务政策、资源分配和高阶人员的选拔方面。”^⑧从钱德勒的定义看,这种家族企业并不是指由家族成员掌握全部所有权和经营控制权,而是一种大部分和基本掌握上述两种权利的企业组织形式。这个定义未将家族企业的全部外延包括进来。潘必胜认为当一个家族或数个具有紧密联盟关系的家族拥有全部或部分所有权,并直接或间接掌握企业的经营权时,这个企业就是家族企业。他还根据家族关系渗入企业的程度及其关系类型,把家族企业分为三种类型:①所有权与经营权全为一个家族所掌握;②掌握着不完全的所有权,却仍能掌握主要经营权;③掌握部分所有权而基本不掌握经营权。^⑨这个定义的外延较宽,不把家族企业看成一个固定的模式。但第三种类型是否仍是家族企业就含糊不清了。台湾学者叶银华根据以前学者的研究,提出以临界控制持股比率将个别公司的

① 杜恂诚:《中国传统伦理与近代资本主义——兼评韦伯 中国的宗教》,上海社会科学院出版社1993年版,第60~61页。

② 陈凌:《信息特征、交易成本和家族式组织》,《经济研究》,1998年第7期。

③ 王军武:《家族企业不是落后的代名词》,《中国企业家》,1999年第8期。

④ 孙早:《家族制与中国私营企业的成长》,《山西财经大学学报》,1999年第10期。

⑤ 雷丁(G. Redding):《海外华人企业家的管理思想——文化背景与风格》,上海三联书店1993年版,第277页(原英文书名为“The Spirit of Chinese Capitalism”,1990)。

⑥ 李新春:《中国的家族制度与企业组织》,《中国社会科学季刊》(香港),1998年秋季卷。

⑦ 杨继绳:《成也家族,败也家族》,《经济参考报》,1999年12月23日。

⑧ 小艾尔弗雷德·D. 钱德勒:《看得见的手——美国企业的管理革命》,商务印书馆1987年版,第9页。

⑨ 潘必胜:《乡镇企业中的家族经营问题》,《中国农村观察》,1998年第1期。

股权结构的差异性与家族的控制程度纳入家族控股集团的认定，于是，具备以下三个条件就可认定为家族企业：①家族所控制的持股比例大于临界控制持股比例；②家族成员或具二等亲以内之亲属担任董事长或总经理；③家族成员或具三等亲以内之亲属担任公司董事席位超过公司全部董事席位的一半以上。^① 这个定义比较精确，而且从股权和经营控制权的角度把家族企业看成是一个连续分布的状况，从家族全部拥有两权到临界控制权，都是家族企业。一旦突破了临界控制权，家族企业就蜕变为公众公司。

从上述家族企业的定义分析看，实在难以确定哪一种状态中的家族企业组织形式是不合理的和低效的。即使是最极端的类型——家族拥有全部所有权和经营控制权——也难以对其作出上述判断。在给定的条件下，如特定的规模要求、产品特征、技术特性、资金需求等，只要家族企业资源能满足这些特性和要求，那么它的存在就是合理的，管理也是有效的。不能简单地从王安将总经理座位传给自己的儿子，导致企业破产等案例而断定家族企业组织是不合理的，管理是低效的。因为中外大量的案例也表明企业控制经营权在家族成员内传递也是相当成功的。同样，根据部分家族企业的衰亡而断定“成也家族，败也家族”也不能深化人们的认识，因为即使是建立了所谓现代企业制度的股份有限公司，也是每天都有衰亡，每天都有成功，由此，不能简单地推论“成也现代企业制度，败也现代企业制度”。判定家族企业组织形式的合理性和管理效率的标准可以借用哈耶克的一个思想，即人类合作秩序的不断扩展^②。大量的案例表明：家族规则有利于创业，不利于发展。当市场竞争的各种条件要求家族企业突破自身的界限，需要以家族资本去有效融合社会的财务资本，需要与非家族成员共享企业的资产所有权、剩余索取权和经营控制权时，或者甚至需要完全放弃家族控制时，家族企业主不能与世推移，依然在家族财务资本和人力资本的封闭圈子内运作，依然用家族的规则来管理企业，那么这时的家族企业组织就是不合理的，其管理也必然是低效的。用制度经济学的语言来描述就是：在特定条件下，家族成员及其之间的忠诚信任关系作为一种节约交易成本的资源进入，家族伦理约束简化了企业的监督和激励机制，这时家族企业就能成为有效率的经济组织。但是，当家族企业在市场竞争中，其内部有限资源和家庭或家族成员管理能力不高而导致的内部交易成本大于那些非家族制企业的竞争对手，造成竞争力低下时，那么家族企业就是不合理的和低效的。我们也可以常常发现，在家族企业中，由于亲情关系的纠缠，家族规则往往不能或难以抑制家族成员的违规行为和内讧，因而造成企业的衰亡^③。

三、中国家族企业的发展

发达国家的经济发展表明：家族企业本身在工业化初期对经济发展起了重要的推动作用。但是，工业化或现代化的历史研究也表明：在落后的经济中，报酬多根据至亲关系、

① 叶银华：《家族控股集团、核心企业与报酬互动之研究——台湾与香港证券市场之比较》，《管理评论》（台湾），1999年第18卷第2期。

② 参看汪丁丁：《经济发展与制度创新》，上海人民出版社1995年版，第108页。

③ 杜恂诚：《中国传统伦理与近代资本主义——兼评韦伯《中国的宗教》》，上海社会科学院出版社1993年版，第98页。

地位和性别来分配；安排经济职务多根据特殊主义（亲情规则）的标准。而在发达经济中，专业技能越来越决定着劳动的报酬，用人的标准更多的是能力而不是特殊的社会关系。因此，随着现代化过程的进展，随着社会经济活动中的分工合作体系在一国和世界范围内的日益扩展和繁复，家族企业能否突破家族封闭的圈子，能否超越亲情熟识这种人格化社会网络交易而进入非人格化的制度性交易，能否从非规范的管理向现代企业的科层制管理转化，对一国现代经济的发展显得至关重要。正是在这个意义上，道格拉斯·诺斯指出：从人格化交换到非人格化交换的转变是经济发展中的关键性制约因素^①。按照小艾尔弗雷德·D. 钱德勒对美国企业的管理革命的研究，上述转变的主要特征是：家族企业超越自身，由一群支薪的中、高层经理来管理企业并协调在其控制下的各单位的工作，使管理层级制本身变成企业持久性、权力和持续成长的源泉，也就是使企业超越了工作于其间的个人或集团的限制，而成为经理式企业、现代化的大企业，使之成为支配主要经济部门，甚至改变整个经济结构的重要因素^②。从钱德勒的分析中可以看出，这个过程是通过两条途径实现的。一是起初的家族小企业在各种压力下，自身的财务资本不足，不得不通过兼并、合并的方式扩展企业规模；二是引入非家族的专业经理人员进入企业，逐步使他们掌握经营控制权，使企业成为支配美国主要经济活动的现代意义上的经理式企业。因此，这种转变的实质内容主要就是家族企业与社会财务资本和社会人力资本等因素融合的过程。在美国，这个过程大约经历了近 100 年的时间。中国家族式企业发展的核心问题也是如何有效融合社会资本，这是一个重大的理论研究领域。

在这个研究领域中，其中一个重要问题就是信用与家族企业的发展，也就是说，家族企业如何将信用扩展到家族成员以外来创业和使企业发展。汪丁丁曾深刻地揭示到，外国和中国的经济发展都表明，说服资源所有者把他们所支配的资源汇集到一起，由企业家支配，并实现企业家的创新方案，是最困难的事。在企业家的合作精神里，最核心的东西是一个“信”字。各种企业，经常造成企业生存危机的，是合伙人相互之间的信任关系发生了问题^③；“资本”，特别是大规模资本的筹集，在根本上依赖于市场经济的现代模式的最核心的一项制度——信用制度。人与人之间的大规模的分工合作，没有“信用制度”的支持是不可能扩展到家族或血缘关系的支撑范围以外的^④。钱德勒虽然详细研究了美国家族企业向现代经理式企业的发展历程，但对此问题的探讨却语焉不详。对现代中国的家族企业而言，这个问题是实实在在的最难逾越的障碍。

“信用制度”当然包括法律形式和文化观念两种形态，而后者则是根本的。当我们深入探讨中国传统文化中的信任资源与企业发展的关系时，又发现一些矛盾的令人困惑的现象。中华民族是不是一个讲“信用”的民族？从大量的古训和格言中，我们看到两种相反的状态：君子一言，驷马难追；士为知己者用（甚至死），为朋友两肋插刀，赴汤蹈火，在所不辞；人心隔肚皮，知人知面不知心，见人未可全抛一片心等。两种说法都是有大量

① 道格拉斯·诺斯：《制度变迁理论纲要》，《改革》，1995 年第 3 期。

② 小艾尔弗雷德·D. 钱德勒：《看得见的手——美国企业的管理革命》，商务印书馆 1987 年版，第 9 页。

③ 汪丁丁：《我思考的经济学》，三联书店 1997 年版，第 199~200 页。

④ 汪丁丁：《回顾“金融革命”》，《经济研究》，1997 年第 12 期。

的历史故事支撑。马克思·韦伯曾明确断言：“儒家君子只顾表面的‘自制’，对别人普遍不信任，这种不信任阻碍了一切信贷和商业活动的发展。”“在中国，一切信任，一切商业关系的基石明显地建立在亲戚关系或亲戚式的纯粹个人关系上面，这有十分重要的经济意义。伦理宗教，特别是新教的伦理与禁欲教派的伟大业绩，就是挣断了宗族纽带，建立了信仰和伦理的生活方式共同体对于血缘共同体的优势，这在很大的程度上是对于家族的优势。”^① 韦伯认为阻碍资本主义在中国发展的首要因素是儒教过于重视家族和家族关系。韦伯的论断产生了广泛而持久的影响，直到 70 年代以来，才逐渐有学者挑战韦伯的论断。近来，引起国际学术界很大反响的福山的新著《信任——社会道德与繁荣的创造》对世界一些有代表性的国家的家族企业特征进行了比较研究。他发现华人社会的工业结构和天主教拉丁语系国家雷同得令人惊讶，企业都倾向于由家族拥有和管理，所以在规模上也显得比较小。在这些国家里，企业主多半不愿为公司引进专业经理人，因为这需要跨出家族联系的圈子，而他们对外人的信任感太低了。这就是福山这本论著中研究的核心问题，即文化差异、社会交往中的信任度是国家之间繁荣和竞争能力差异的原因。他认为，存在着三种资本：经济资本、人力资本和社会资本。他的社会资本的定义就是：人们在一个组织中为了共同的目的去合作的能力，是从社会或社区中流行的信任中产生的能力。他将不同的文化区分为低信任度的文化和高信任度的文化，前者的信任只存在于血亲关系上的社会，后者是指信任超越血亲关系的社会。他认为，中国、意大利南部地区、法国等属于低信任文化，而日本、德国、美国等属于高信任文化。注重家族主义的社会比较难以创建大规模的经济组织，而规模的限制终将阻碍中小企业参与全球经济的幅度。福山还对同属于东方文化的中日家族企业作了比较，认为，日本到了 30 年代，财阀企业都不再依赖家族成员担任高阶管理工作了。“二战”后，日本又经历了财阀的解体与企业及其网络的重构。日本公司已经制度化，变成专业化管理的科层组织。因此，它们从来没有面临过像某些华人公司创办人过世之后所留下来的家族企业崩解或企业精神溃散的问题。日本企业之所以能够迅速成长，早早就超越家族事业的范围，原因是日本的家庭与华人家庭有天壤之别^②。福山虽然指出了日本财阀企业与华人企业的差异，但未能看出日本企业在名义上具有现代层级制特点，而家文化的灵魂仍渗透其中。否则，70 年代末至 80 年代，日本企业的管理艺术就不会令那么多的西方学者惊讶和探讨了。韦伯和福山的见解虽有偏颇之处，但所提出的仍然是值得深入研讨的重大课题。

相比而言，雷丁的见解更加慎重和全面。他指出：中国人十分重视“信任”这一观念，中国人的信任建立在个人的承诺以及保持名誉和面子的基础上。“在其他国家需要通过合同、律师、担保、调研、广泛征求意见以及各种时间耽搁才能完成的各种交易，海外华人只要依靠电话、握手或清茶一杯即可解决问题”，当然，雷丁也谈到，这并不是说合法的手续完全免掉了。同时，雷丁也指出了另一面：华人之间的信任是有限的，是个人之间的信任，“主要特点似乎是对家庭的信任是绝对的，对朋友或熟人的信任只能达到建立相互依赖关系，双方都不失面子的程度”。对个人的忠诚，否定了客观性和公正性，从而

① 马克思·韦伯：《儒教与道教》，商务印书馆 1995 年版，第 296，289 页。

② 弗兰西斯·福山：《信任——社会道德与繁荣的创造》，远方出版社 1998 年版。

阻止了真正合乎理性的和专业的严密管理制度发展。“要推动华人家族企业通过实行西方所谓的‘管理革命’，即把权力移交给职业经理人员并把控制权与所有权分离，明显地存在着很大的困难”^①。这种权力的不可分享是限制海外华人企业组织发展的重要因素。当然，雷丁似乎也看到海外华人企业也处在发展变化之中，部分家族企业不是一味地注重亲和关系，而是同时注重能力、表现。对这一点，井上隆一郎看得更清晰一些，在亚洲，“在经营方面，打破同一家族经营的旧格局，起用有能力的专业经营者的倾向也广泛出现”^②。

完全断定中国人固守家族观念，中国的家族不能突破自身的局限性，是缺乏历史和现实根据的。一些学者的研究表明，清末以来，家族企业的领导层中实际上就有了亲属以外的人员，并在经商活动中始终起着重要的作用，华人家族企业的人员网络是富有弹性的(Hamilton, 1991; Zelin, 1991; 彭允松, 1994; 刘可为, 1997)^③。金耀基对香港家族企业的研究发现：企业主对亲属的雇佣，多半出于理性上的考虑，而不是单单由于儒家家族主义文化传统的结果^④。现实中，浙江温州民营家族企业的发展具有一定的典型意义，一些上规模、上水平的大公司不仅在股权结构上，而且在治理结构上打破了家族的垄断，从而造成执行机构、董事会和股东会对家族制的全面突破^⑤。浙江金义集团更是突破家族企业模式的突出案例。笔者的调研发现，有些私营企业主严格限制家庭和家族成员在企业中担任管理职务，甚至不允许他们成为企业的员工。

上述分析表明：韦伯、福山和雷丁的见解虽然发人深思，但他们基本上把中国传统文化看成一个常量，特别是把“家”看成是范围固定与封闭的，并由此考察其对家族企业发展的影响。但实际上，中国传统文化有很大的包容性和弹性，比如“家”这个概念就有很大的伸缩性，正如费孝通(1948)所言，“在中国乡土社会中，家并没有严格的团体界限，这社群里的分子可以依需要，沿亲属差序向外扩大。”“家里的”可以包罗任何要拉入自己的圈子、表示亲热的人物。自家人的范围是因时因地可伸缩的，大到数不清，真是天下可成一家^⑥。由此可以得出这样一个推论性的判断：中国家族企业的发展具有突破狭隘的亲情“信用”限制的文化基因，可以实现从人格化的交换向非人格化交换的制度创新。由何梦笔和陈吉元共同主持的项目研究发现：传统的血缘、亲缘和地缘关系，与适应市场需要而建立的各种带有很强的商业性质的利益关系，能够耦合成一个有机的整体。这里提出了

① 雷丁(G. Redding)：《海外华人企业家的管理思想——文化背景与风格》，上海三联书店1993年版，第287，85，183，240页(原英文书名为“The Spirit of Chinese Capitalism”，1990)。

② 井上隆一郎：《亚洲的财阀和企业》，三联书店1997年版，第49页。

③ Hamilton G., Business Networks and Economic Development in East and Southeast Asia, Hong Kong: Centre for Asian Studies. Zelin M., The Rise and the Fall of the Fu-Rong Salt-Yard Elite: Merchant Dominance in Late Qing China, In: Esherick, Rankin, 1991, pp82~109; 彭久松：《中国契约股份制》，成都科技大学出版社1994年版；刘可为：《从山西票号的衰亡探析企业的经营与制度创新》，《管理世界》，1997年第4期

④ 金耀基：《东亚经济发展的文化诠释》，刊《中国社会和文化》，香港牛津大学出版社1992年版，第52~69页。

⑤ 马津龙：《关于温州民营企业制度创新和管理创新若干问题的思考》，《经济研究》，1999年第12期

⑥ 费孝通(1948)：《乡土中国》，转引自《东方之子·大家丛书·费孝通卷》，华文出版社1999年版，第142，131页。

一个很值得深入探讨的学术方向。

四、泛家族主义与华人企业组织行为

上述分析都涉及家文化与中国家族企业行为这个大问题。中国家文化之所以重要，是因为它不只是给家庭或家族提供一套规则，而是把它泛化到社会经济生活的方方面面。任何家族以外的社群、机构，包括企业或国家，都可视为“家”的扩大。因此，泛家族主义是中国文化的一大突出特征。这一点，上述费孝通的见解已有论及。中西方的“家”的概念，家族规则与家族企业虽有一定的差异，但其共性特征十分突出，如亲情规则、内外有别、家长权威等。但把家族观念、家族制度、家族行为规则泛化至社会的各个层面，却是中西方文化截然不同的一个基本点，而华人企业的组织行为的奥秘深蕴其中。台湾著名学者杨国枢曾对泛家族主义作了长期的极有意义的探讨。他对泛家族主义作了简明而准确的描述：在传统社会内，在家族中的生活经验与习惯常是中国人惟一的一套团体或组织生活的经验与习惯，因而在参与家族以外的团体或组织活动时，他们自然而然地将家族中的结构形态、关系模式及处事方式推广、概化或带入这些非家族性的团体或组织。也就是说，在家族以外的团体或组织中，中国人会比照家族主义的取向而进行。更具体言之，中国人的泛家族化历程主要表现在三个层次：①将家族的结构形态与运作原则，概化到家族以外的团体或组织；亦即，比照家族的结构形式来组织非家族团体，并依据家族的社会逻辑（如长幼有序）来运作。②将家族中的伦理关系或角色关系，概化到家族以外的团体或组织；亦即，将非家族性团体内的成员予以家人化，成员间的关系比照家族内的情形而加以人伦化。③将家族生活中所学得的处事为人的概念、态度及行为，概化到家族以外的团体或组织；亦即，在非家族性团体或组织内，将家族生活的经验与行为，不加修改或稍加修改概予采用①。此后，杨国枢又进一步深化了他的研究，认为中国人是经由刺激类化的途径将家族的组织特征、人际特征及行为特征推广到家族以外的团体②。台湾其他学者的研究成果进一步证实了杨国枢的分析。刘兆明等的研究发现：在台湾非家族企业中，甚至在位于工业化都会区的大企业中，其企业文化也包含了以下现象：①组织领导有意无意都会形成长式的权威，且将此种权威建立在道德伦理基础之上；②组织内强调家庭气氛，特别重视和谐，鼓励团队精神，形成组织是个大家庭或大家都是一家人的一体感；③组织内形成类似家庭伦理中之长幼与辈分，并建立私人感情以维持此种特殊伦理关系；④依关系亲疏形成组织内的差序格局，进而导致以组织领导为中心的内团体，使组织内的层级化更为明显③。本文前面也提到，大陆学者樊江春也揭示企业单位中充满类似家族的文化特色及组织功能。

① 杨国枢：《中国人的社会取向：社会互动的观点》，见杨国枢，余安邦主编：《中国人的心理与行为——观念及方法篇（1992）》，台湾：桂冠图书公司1993年版，第95页。

② 杨国枢：《家族化历程、泛家族主义及组织管理》，见：《海峡两岸之组织与管理》，台湾远流出版公司1998年版，第19~60页。

③ 刘兆明，黄子玲，陈干玉：《台湾企业文化的解读与分析》，1995；转引杨国枢《家族化历程、泛家族主义及组织管理》一文，《海峡两岸之组织与管理》，台湾远流出版公司1998年版。

上述分析展开了一个极有现代学术价值与实践意义的研究领域，即中国传统家文化与华人企业组织行为。在这方面，台湾学者的研究特别引人注目。他们认为，在探讨家族主义与华人企业组织行为的关系上，至少有三种不同的思考面向：①讨论儒家伦理与家族主义的关系；②探讨家族主义与组织行为的关系；③理清家族主义与经营绩效的关系。陈其南在比较中、日、美的经营体制之后，指称华人的企业组织是“差序关系导向的纯营利经营方式”企业，与日本的“身份关系与共同体理念的经营方式”、美国的“契约关系与市场规范的经营方式”企业是大不相同的^①。郑伯壘认为，企业主持人对员工的差序认知结构，实为影响华人组织行为的重要关键。根据关系、忠诚、才能等三项归类标准，企业主持人会依差序程度，将员工归类为八种类别原型。由于企业主持人与各类别原型之员工的互动有异、信任有别、所持自己人意识也有不同，而导致了华人企业内部复杂、多样而极具特色的组织行为^②。

笔者认为，当代华人社会与华人企业并非完全受中国传统家文化和泛家族主义的支配，儒家伦理在近代以来，在海外，在核心价值观上都有重大变化，并对华人经济的发展产生重大影响^③。华人社会正处在传统家文化与现代工商社会文化的互动、融合、重构的变迁过程之中。探讨这种变迁过程对学者展示了极为诱人的学术前景。

泛家族主义虽然是中国文化的一大特征，对华人企业组织行为产生至关重要的影响。但是，这种泛家族主义并不是普天之下同等兼爱，而是在社会关系与社会交易方面有着独特的形态，这就是费孝通在半个多世纪前所提出的著名的“差序格局”论。一个差序格局的社会，是由无数私人关系搭成的网络。这个网络像个蜘蛛网，有一个中心，就是自己。以“己”为中心，像石子一般投入水中，和别人联系成的社会关系像水的波纹一般，一圈圈推出去，愈推愈远，也愈推愈薄。也就是说，愈往外推，关系的紧密程度和信任程度是递减的。费孝通的假设是：中国乡土社会采取了差序格局，利用亲属的伦常去组合社群，经营各种事业，使这基本的家变成民族性了；但家的结构不能限于亲子的小组合，必须加以扩大^④。西方社会是团体格局，华人社会是差序格局，对此，西方学者也逐渐谙识其中的内涵。《纽约时报》前驻外记者 Fox Butterfield 有如下观察：中国人对其心理世界进行组织时采取了一种完全不同于西方人的方式。西方人倾向于把人视为个体，虽然也对所认识的人与不认识的人作出某些区分，但从根本上说，西方人有一套对所有的人都适用的行为准则。……中国人则本能地将人分成两类：一类是他们与之已有固定关系的，另一类则是没有固定关系的。这些关系就像一根根看不见的线，把中国人紧紧串连在一起^⑤。

差序导向不仅是华人企业自身的组织特征，而且也是华人企业利用这种社会关系网络来配置资源和经营扩展业务的主要途径。这方面，众多学者作了大量有价值的研究。这里仅评述部分研究成果。Hamilton 从宏观社会学角度比较了西方和中国历史之后，得出结

① 陈其南：《家族伦理与经济理性——试论韦伯与中国社会研究》，《当代》（台湾），1986年第10期。

② 郑伯壘：《差序格局与华人组织行为》，《本土心理学研究》（台湾），1995年第2期。

③ 储小平：《儒家伦理与海外华人经济的发展》，《汕头大学学报》，1998年第6期。

④ 费孝通：《乡土中国》，转引自《东方之子·大家丛书·费孝通卷》，华文出版社1999年版。

⑤ Fox Butterfield (1983)，转引自金耀基：《关系和网络的建构——一个社会学的诠释》，见金耀基：《中国社会与文化》，香港牛津大学出版社1993年版，第65页。

论：经济的基本单位实际上是家庭企业的网络；这些网络根据不同的形势和需要调整自己的范围和基础^①。所以，华人家庭是一个灵活开放的单位，其范围可以扩大也可以缩小。雷丁认为：关系网是起决定作用的，关系网是理解海外华人社会与经济生活所必须注意的基本情况，中国人的家族企业通过复杂的外界网络扩大交易与势力^②。在何梦笔和陈吉元主持的研究项目中，胡必亮通过对湖北、山西、陕西、广东、浙江五省乡镇企业发展的典型调查研究，发现“关系”规则对乡镇企业经济资源的组织表现出了十分明确的积极作用，在市场规则、行政规则并存的环境中，“关系”规则的作用也是相当大的。“关系”已经成为我国社会生活中一种普遍存在的“游戏规则”^③。何梦笔^④和胡必亮认为家族制度只是中国特殊的“关系”制度的一种表现形式，但他们未能进一步认识到其他各种关系往往是家族关系的泛化。张其仔将社会资本定义为社会网络，并考察了私有（家族）企业通过亲缘关系网络来配置资源的效应。他们认为，不但一系列的重要经济资源通过社会网络的管道流向了个人和经济单位之手，而且社会网络被“资本化”，被不少人纳入一定的经济体系，以节约费用，提高效率^⑤。石秀印直接提出了“社会网络基础论”。他重点考察了企业家在社会网络中的位置与作用，认为社会关系媒介可分为血缘的、姻缘的、地缘的和事缘的。他的主要假设是：那些与资源拥有者具有更亲密的私人社会关系的人，更可能成为私有企业家^⑥。郑伯壖也指出，位于网络节点的各企业主持人实为经济发展的枢纽^⑦。这些学者的一个共同特点是：忽视了网络关系对健全的市场经济规则形成的负面影响。

五、简短的结语

本文旨在对家族企业研究的重要文献作出理论评述，揭示其所具有的现代理论价值和实践意义，并进一步阐述自己的研究思路。

从历史角度看，家族企业始于18世纪英国工业革命时代，200多年来，家族企业呈现多样化的发展模式，只有少数家族企业在竞争中因时演变而得以延续或扩展。从世界范围看，各国的经济活动都为家族企业留下了比较广阔的空间。全世界每天都有成千上万的家族企业在生生灭灭。但是，家族企业在全球经济活动中始终扮演着重要的角色。

中国家文化积累之深厚、对人心理与行为的影响之大是其他民族的家文化所难以比拟的。尽管家文化也处在动态变化之中，也在与现代市场经济的规则处在融合之中，但其深层的核心结构仍有顽强的生命力。深入研究家族企业，特别是华人家族企业，可望能总结出华人社会特有的有效率的企业组织制度、企业内部有效的治理结构、行为管理模式、社

① Hamilton G.: Business Networks and Economic Development in East and Southeast Asia. Hong Kong: Centre for Asian Studies, 1991.

② 雷丁(G. Redding):《海外华人企业家的管理思想——文化背景与风格》，上海三联书店1993年版，第47，4~5页(原英文书名“*The Spirit of Chinese Capitalism*”，1990)。

③ 胡必亮：《“关系”规则与资源配置》，《中国社会科学季刊》(香港)，1996年秋季卷。

④ 何梦笔：《网络文化与华人社会经济行为方式》，山西经济出版社1996年版。

⑤ 张其仔：《社会资本论——社会资本与经济增长》，社会科学文献出版社1997年版。

⑥ 石秀印：《中国企业家成功的社会网络基础》，《管理世界》，1998年第6期。

⑦ 郑伯壖：《差序格局与华人组织行为》，《本土心理学研究》(台湾)，1995年第2期。

会关系网络与资源配置机制等极有价值的结论。同时也需进一步深入研究以下问题：

家族亲情及泛家族式社会网络作为资源配置的手段，是否会导致原子式的家族企业之间陷入恶性关系网的竞争？家族企业组织与政府组织之间是否会陷入相互报酬递减的恶性循环之中？是否会妨碍制度化的市场经济规则的形成及对资源的有效配置？此外，如果说家族规则与市场经济制度规则能够结合，那么其结合的机制、条件是什么？

笔者并不认为世界范围内的家族企业都要转变成非家族式的企业。大量的家族企业的存在是在任何经济结构中不可缺少或必然应有的内容。但是，在现代经济成长中，家族企业本身的生存与发展必须要有效地融合社会资本，而家族企业向现代企业的转化对一国经济的发展起着至关重要的作用。科尔曼^①和福山等从社会关系、社会组织结构、信任及合作能力、权威和规范等方面界定了社会资本的内容、特点与作用，富有极大的启发意义。笔者认为，从个体与社会之间的关系角度，可进一步把社会资本界定为：个体（包括企业个体）以外的以社会状态存在的各种资源要素，大致包括四个层面：社会财务资本、社会人力资本、社会关系网络资本和社会文化资本。家族企业与社会财务资本的结合往往伴随着与人力资本的结合，在与社会财务资本和人力资本结合的过程中，华人社会特有的社会关系网络起着特别重要的作用，而这一切都与中国的文化资本积累有关。因此，有必要深入研究家族企业与这四个层面的社会资本有效融合的途径、方式、条件和障碍，以及这四个层面的资本之间的内在联系。

钱德勒的研究告诉我们：在美国这样家文化观念较淡薄的国家，其家族企业演变为现代经理或企业花了 100 年的时间。而中国当代的家族企业才只有 20 年的发展历程。相信具有弹性和容纳能力的华人家文化也会适应当代经济既高度分工又高度一体化的时代要求。钱德勒的研究中有一点要特别注意，家族企业演变成现代经理式企业，“使管理层级制本身变成企业持久性、权力和持续成长的源泉”，而华人家族企业中的家长权威制度、遗产继承制度、亲情规则纠缠形成的内耗等恰恰使企业难以形成持久性、权力和持续成长的源泉。中国式家族企业在发展过程中如果不能有效地与社会资本融合，就难以逃脱“一代创业、二代守成、三代衰亡”的规律。家族和家族式企业发展状况如何将直接影响中国经济与社会的发展前景。

^① 詹姆斯·科尔曼：《社会理论的基础》，社会科学文献出版社 1990 年版。

企业集群的创新过程研究^{*}

王 珺

一、引言

近年来，企业集群作为突破中小企业技术创新自身限制的有效组织方式，已引起了学术界与决策部门的普遍关注（Piore、Sabel，1984；UNCTAD，1998；Giovana Ceglie and Marco Dini，1999；M. P. van Dijk，1999；M. Porter，2000）。然而，这种关注仍然集中在集群所形成的知识外溢与信息共享对创新能力的增强作用上，而较少地从集群创新能力的产生、增长与扩散过程来全面地加以把握与认识。事实上，如果集群本身没有产生创新能力，或没有形成这种自我增强的成长机制，那就谈不上创新能力的增强与扩散。M. 波特（M. Porter，2000）给出的相当部分企业集群的萎缩和衰落反映了集群内企业并不天然地具备这种创新能力的事实。因此，这需要研究企业集群的创新能力来源问题。

在研究企业集群创新能力的文献中，可以分为两类：一方面，创新能力是内生的，即通过技术创新的外部性、协同性和范围经济等方面从内部引致创新。萨克森宁（Saxenien，1996）对美国以硅谷为代表的IT企业聚集性创新网络研究提供了这个方面的实例。在以传统产业为主的企业集群中，当大多数中小企业缺乏创新能力与动机时，依托集群中较大型企业的技术力量与研发活动被认为是一条突破创新困境的可行之路。如中小企业按照大企业要求的技术与质量标准提供承接下来的各种产品，大企业为保证产品质量也会为中小企业提供一定的技术服务等。日本大企业与中小企业之间的订单承接与分包联系提供了这种网络创新的实例（今井贤一、小宫隆太郎，1995）。另一方面，创新能力是外生的，即技术范式的变更与引入构成了外生的创新能力来源。从外生性创新能力看，一些学者论述了集群与跨国公司等技术领先者在生产链条之间建立相对稳定的转包与承包关系，从而给作为技术后来者的集群创造模仿与学习的机会等（Birkinshaw，J. M.，and Neil. Hood，1998；Hobday，1994）。

然而，无论是内生型或是外生型创新能力，其生长机制都是有条件的。内生型创新能力的生长限于集群本身的创新性质或集群内大企业的存在。如果集群是以传统产业为主的企业网络，而不是类似硅谷的创新网络；如果集群都是中小企业，而不存在中小企业技术创新可依赖的大型企业，那么，内生型创新能力就无法在这些集群中应用与推广。外生

^{*} 本研究是作者主持的2002年度国家社会科学基金项目“企业集群的创新过程研究”中的一部分，该项目批准号为02BJL024。原载：《管理世界》，2002年第10期。

型创新能力的依赖条件是跨国公司或跨地区大企业的进入。如果没有这些大公司的进入,企业集群也就无法获得创新能力的来源。事实上,现实中大量存在的企业集群都是既不具有内生型创新能力条件,也缺少外生型创新能力条件的以传统产业为主的中小企业网络。如何使技术创新能力在这些中小企业网络中生长出来,并不断地得到强化和发展,这是关系到大多数中小企业集群的兴衰而现有的理论尚未更多关注的问题。

本文以不具备内生型与外生型创新能力条件的中小企业集群为研究对象,着重探讨中小企业集群的技术创新过程。由于技术创新过程包括了创新能力的来源、创新资源的组织与创新活动的扩展等三个方面,为此,本文针对以传统产业为主的中小企业集群创新能力从哪里来、如何最有效地组织与使用创新资源以及扩展创新成果等问题,提出以地方政府为主体的外部引入、创新组织的公司运作和创新活动的市场化扩散为特征的集群创新能力动态性增强与扩展的理论观点,并以广东省南海市西樵镇和中山市小榄镇企业集群的创新实践印证这个理论看法,目的在于为大多数以传统产业为主的企业集群创新能力生成、强化与扩展提供可行的理论指导。

二、以政府引入创新源,突破中小企业创新困境

企业集群(Clusters of enterprises)是指在某一特定产业形成相互联系的公司与机构在地理位置上的聚集(Becattini, 1991)^①。现实经济中的企业集群形成大约可以概括为组织设计和自发生成两类。如创新科技园、IT软件园、出口加工区等都属于前一种类型,这种类型体现了作为设计主体的地方政府的发展目标与战略意图;依据地理位置、文化传统、产业联系和市场聚集资源等因素自然形成的企业聚集是后一种类型。无论是发达国家还是发展中国家,大多数以传统产业为特色的中小企业集群都具有自发生成的特点。如主要集中在意大利东北部的70多个小企业专业区^②,在东京都大田地区的金属制品小企业集群,玻利维亚集中在Santa Cruz De la Sierra的服装企业集群,加纳Kumasi地区的汽车配件制造专业区,洪都拉斯在San Pedro Sula地区的家具、金属制品专业区,印尼的Aentral Java地区的棕榈糖加工区,巴西在Sinos Valley地区的鞋类专业区(Hubert Schmitz, 1995),印度在Tiruppur地区的纺织产品专业区,以及国内浙江省、广东省等相继出现的服装、陶瓷、灯饰、钟表、制鞋、五金、玩具和家具等几百个以传统产业为主的产业区和专业镇等(仇保兴, 1999;王珺, 2002)。本文专注于我国,特别是广东省自发生成的中小企业集群创新能力形成与发展过程。

自发形成的中小企业集群往往因缺乏创新能力而自发地衰落下去。这不仅是由于构成

^① 目前,国内外学术界对于具有某一产业内部分工联系的企业聚集现象给出了许多不同的概括性定义,如产业区、新产业区、企业集群、小企业集群、地方企业网络、特色产业区、专业产品区、专业镇与块状经济等。虽然每个概念强调的重点都有所不同,但是在现有的理论文献中,使用频率较高的专业术语是企业集群,因此本文在这里也使用这个概念。

^② 这个以小企业专业区聚集为特征的意大利东北部被Becattini等学者称为“第三意大利”,这是相对于南部相对贫穷的“第一意大利”和西北部历史上相对比较富裕的“第二意大利”而言的,即指东北部以小企业专业化的聚集促使意大利东北部的迅速崛起,使其与南部和西北部形成了明显的差别。

集群的大多数中小企业资本规模小，靠中低档产品的标准装配线和传统工艺组织生产，在产品开发的后期阶段漫游，属于生产技术上的后进者（Hobday, 1995），更重要的是，传统产业内在的中小企业模仿动机远远超过了其创新动机。因为创新投资的风险较大和创新产品的外部性效应较强^①，这使中小企业愿意成为“免费搭车者”，而不愿意自己投资。每个企业都企盼着别人开发出新产品，而自己仿效，以减少研发投资和降低创新风险。然而，如果每个企业都选择这种策略，那么，企业集群的创新就会陷入纳什定义的“囚徒困境”，这就是一些企业集群萎缩甚至衰落的根源所在。广东省于20世纪80年代曾十分兴盛的沿着马路边形成的以农副产品、日常工业用品为主的专业市场及专业生产网络到了90年代中期纷纷萎缩的事实印证了这种推论。

要避免以传统产业为主的中小企业集群衰落，由地方政府引入外部创新能力与资源就变得非常重要。一方面，处于发展初期的地方政府充当整个集群的代表会比任何一家企业更为合适，因为由任何一家企业承担这种角色都可能会使该企业利用这种权力引进对自己的技术开发有利而不一定对整个集群的企业都适用的技术，而地方政府则以集群内大多数企业需要的通用型技术为出发点，从而可能避免由任何一家企业引进所带来的倾向。另一方面，地方政府在引进外部资源中，从谈判能力到组织信誉，从合作网络的扩展能量到各项条款的承诺，都比一般的中小企业更有优势。

由地方政府扮演引入者角色，就需要对地方政府的引入动机做些分析，因为这直接涉及引入效果。在经济转轨中，除了财税制度改革导致“地方政府公司化”（Jean, Oi, 1992）的经济诱因之外，还有两个条件对地方政府的引入动机产生着重要的影响。一是集群环境。地方政府受到财力和体制限制，不可能为每个企业购置先进的技术设备。如果地方政府引入的技术能力能够为大多数企业所共享，那么引入行为就产生了规模效果。企业集群恰恰提供了这样一种环境，因为企业集群是以专业化和规模化市场为特征的^②，即生产与经营同一类产品的企业聚集在一起，他们对技术设备与设计的需求基本上是相似的，这种对技术与设计的类似需求，一方面成为技术供应者进入的主要吸引力，以集群环境引入科研院所、高等院校等合作伙伴比非集群环境更容易成功的事实说明了这一点（Chandler, Alfred D. Orjan Solvell, and Peter Hagstrom, 1998；）。另一方面是地方政府降低引入成本，提高引入规模收益的基础。因为引入一套技术设备与资源对集群企业技术升级的辐射范围比非集群企业的技术的辐射范围大得多、成本低得多。二是集群边界。一个地方政府管辖范围内的企业集群与一个跨地区的企业集群相比，前者更容易使地方政府发生引入行为。因为前者的行政边界是清楚的，在本地区内产生的知识外溢与信息共享效应是明显的，作为地方政府业绩也是有显示性的。相反，在一个跨地区集群中，引入行为所带来的外部效应必然会波及地区外的企业，而这些得到技术外溢收益的企业增长被看成是没有发生引入行为的地方政府的业绩。任何一方政府都不愿意做“自己栽树，别人乘凉”的事

① 这里所说的创新产品外部性效应是指某一个中小企业在开发出新产品后，由于缺乏知识产权的有效保护，必然导致模仿行为的大量发生，从而使创新者无法实现创新的收益。

② 这种专业化指的是成百上千个企业进入同一产业或少数几个相关产业，并沿着这种产业链条形成了分工联系。所谓规模化，是指进入同一产业或少数相关产业的成百上千个企业共同构成了整个产业供求的市场规模。

情，而愿意免费分享知识的外溢效应。其结果，跨地区企业集群创新能力的引入就会受到制约。近年来广东省大部分专业镇^①相继设立技术创新组织，而跨乡镇的产业区却很少建立这种技术创新中心的事实就说明了这一点。

受到地方财力、技术信息与本地企业生产能力的限制，地方政府主要以引入技术趋于成熟而本地集群发展所急需的应用型技术设备与设计能力为主。这种引入方式主要有以下几种：①以地方政府垫付资金、购置关键性技术设备为重点，通过与本地区技术能力的结合，提升整个集群企业的技术开发能力。广东省中山市以生产灯饰为主的古镇就是通过这种方式使整个集群的灯饰品种开发数量由每年新增十几种变为上百种，产品合格率由40%左右上升到80%以上，整个集群灯饰产品占据国内市场的60%以上。②以技术设备引入为核心，成建制地引入外部技术资源，包括了对技术设备的调试、安装、操作、维修和使用新设备、开发新产品等工艺所需要的各类技术与管理人才等。广东省南海市金沙镇五金模具设计中心提供了这样的例子。该设计中心是镇政府在成建制地引入华中科技大学机械设计与制造专业的技术设备与人员基础上设立的。自该设计中心建立以来，金沙专业镇不仅五金产品种类有了大幅度增长，产值也增长了4倍以上。③以不同渠道引入不同技术资源，通过地方政府整合，形成有效的技术创新能力。广东省顺德市以木工机械产品为主的伦教镇就是这样一个例子。该镇政府先垫付资金从意大利和德国等国家引进一批国际上比较先进的生产家具的木工机械设备，再从国内科研院所聘任了几个“星期六工程师”，从市场上招聘了一批技术与管理人才，组建了木工机械设计中心。该中心设计与开发出来的各种新款式木工机械设备极大地替代了作为其邻居且占全国家具市场近三分之一的乐从镇近2000多家家具企业对进口木工机械设备的依赖。在乐从镇，使用伦教镇制造的木工机械的企业从1999年的不到20%上升到2002年底的接近60%。

三、以公司化运作，形成创新能力增强与扩散机制

随着创新资源的引入，选择一个创新资源有效使用的组织机制是必不可少的。从一些国家创新组织的发展经验看，一个有效使用创新资源的组织机制应具有三个特性：一是主体性。集群内创新资源的使用主体是企业，而不是政府。因此，地方政府需要设计一种机制将自己引入的创新资源转移到企业手中。二是持续性。一次性引入的创新资源应转化为具有自我增强的持续性创新能力，而不应使集群内企业的创新能力越来越严重地依赖于引入外部的创新资源。三是扩散性。引入的创新资源应成为整个集群创新能力的种子，通过技术外溢与信息共享机制，使得创新能力与技术信息迅速地扩散到集群内几乎所有的企业，从而促使这些企业利用低成本获得的创新资源提升自身技术开发能力。以这三个特性为参照，评价与选择现实经济中出现的三种组织机制。

第一种是以扶持大企业为重点的组织机制。这种机制是将地方政府引入的创新资源直接投放到集群中最具开发能力的较大型企业中去，通过增强该企业的研发能力，使其尽快

^① 现阶段广东省大部分的企业集群是以乡镇等行政区划为边界的，“一镇一业”、“一乡一品”成为这类企业集群的基本特征，因此，我们将这种以乡镇行政边界为特征的企业集群称为专业镇。

成为企业集群新技术和经济的生长点。这种机制设计的好处在于：一方面，直接将地方政府掌握的创新资源转移到企业手里，强化企业作为使用创新资源的主体地位；另一方面，较大企业的研发投入与地方政府引入的创新资源相结合，有利于增强持续性的创新能力。不足之处在于：在地方政府把创新资源投放给较大型企业后，该企业往往会根据本企业的发展战略与竞争策略，开发出适合于本企业需要的专有技术产品，而不一定适合于大多数企业的通用型技术产品，其结果，技术扩散会受到限制。特别是这类企业为维持在新技术开发等方面的垄断，往往会人为地对技术扩散设置一些障碍，如技术转让定价过高、技术信息提供不充分等。

第二种是以公共服务为宗旨的组织机制。地方政府把引入的创新资源看成是公共产品，通过建立一个公共性的非盈利机构，为集群内企业无偿地提供新产品设计与开发试验等方面的技术服务。这种机制的好处在于：这种技术服务的无偿性有利于降低中小企业采用新技术、新设计和新品种的费用，加速创新能力的扩散。不足之处是：这不是一个公司化运作的组织，不仅缺乏盈利动机与活力，也无法形成创新能力的持续性增强机制，只能靠地方政府不断地增加投入来维持这类公共组织的运作与公共设施的维护、保养、管理与更新，致使地方政府的财政负担越来越重。特别是创新资源的无偿使用会使中小企业对公共性创新资源产生越来越强的依赖性。其结果，整个集群内企业都依赖这个公共性技术组织，而这个组织因缺乏盈利动机和持续发展能力而陷入萎缩的困境。我国计划经济时期大多数科研机构面向企业服务的动机不足、市场化的开发能力有限就是这种组织机制的结果。

第三种是以公司化运作为原则的组织机制。地方政府根据引入者数量与投入规模，组建一个股份制或政府全部出资的技术开发公司，来专门经营与开发引入的创新资源，目的在于通过市场竞争，不断开发适销对路的新设计、新工艺和新产品，在满足中小企业需要的同时，增强公司持续性创新能力。这种公司化组织的优点在于：企业是创新资源的运作主体，盈利动机与市场压力使该公司必须增强持续性创新能力。但是，创新组织的公司化定位也会带来一个问题，即在假定新开发的技术设计与产品符合中小企业技术更新需要条件下，其价格水平往往成为困扰着技术扩散的一个难点。因为创新组织是以商业规则来确定新开发的技术产品定价的，对于中小企业支付能力来说，这种定价往往比较高。如果大多数中小企业无力以这种价格发生交易，那么以这种交易方式形成的技术转移与扩散就会受到限制，反而以低于购买价来获取技术秘诀的违规行为，如盗版设计、窃取信息和低劣仿效等会大行其道，其结果，创新公司也无法得到应有的收益。如果降低技术转让价格，那么创新组织又会遭受损失。因此，突破交易双方难以成交的困境是这种组织机制在起步阶段面临的挑战之一。

以上三种创新组织机制都是不完整的，因此，使用创新资源的组织只能是一种次优选择，而不是最优选择。在这三种组织机制中，第一种组织是以集群中存在着较大型企业为前提的，对于大多数企业集群来说，这个前提往往是集群发展的结果，因此不适合处在成长初期的中小企业集群。第二种创新组织的设计是远离市场经济制度的，因而也无法使创新主体在盈利动机与竞争压力下推动其技术成果的持续性开发与扩散。第三种创新组织的设计是定位在已经进入市场经济正常运作轨道的公司组织，然而，这种组织设计的推行难

点在于如何从成长期进入到成熟期。从上述比较中可以看出，次优选择主要集中在第一和第三种组织设计上，第一种组织可能产生的问题是大型企业专用性技术开发需要与中小企业对通用性技术产品与服务需求之间的冲突，第三种组织的难点是创新组织设立初期双方交易能否顺利实现。两者相比，作为组织设计主体的地方政府来说，往往会选择第三种组织机制，因为企业自主发展机制使地方政府对由第一种组织引发企业间利益冲突的协调力度是有限的，而通过地方政府的介入促成交易的实现却是可行的。

在发展初期，政府以补贴方式的扶持是促成技术交易实现的关键。地方政府根据自己的财力，可以采用直接补贴到技术转让的价格里，也可以补给出售者或购买者以及几种方式共用等，目的在于保证创新组织获取平均资本回报率基础上，使中小企业能够买得起技术、设计与设备等新开发产品。不可否认，补贴在一定程度上造成了市场信号的扭曲，但是，与成长初期难以形成技术交易的发展困境相比，这种扭曲的代价是较低的。从这个意义上说，由补贴支撑的第三种组织机制可以看成是“帕累托改进”的次优模式。然而，补贴不能无限地持续下去，因为技术交易量的不断扩大不仅使补贴量增大，而且对经济运行的扭曲影响也会不断加深，如中小企业过分依赖补贴，开发产品价格的扭曲日益严重以及地方政府为支付不断增加的补贴而致使财政负担日益加重等，因此，选择合适的时机退出补贴应成为逐步发展起来的集群地区政府扶持政策调整的一个重点。

四、以市场扩展，推进创新组织与企业集群的创新互动

地方政府对补贴政策的调整是以扩大创新组织的技术扩散来推动技术服务的市场拓展为条件的。通过市场拓展形成的规模收益，使得补贴在减少甚至取消的情况下，中小企业可以获得以低价支付的创新成果，而创新组织仍能有盈利空间。

加速技术成果与能力的扩散是创新组织实现价值、中小企业更新技术的关键。创新组织的技术扩散主要是通过四个渠道加以推进的：一是增加新技术设计与产品的有效供给。这类技术成果的有效供给是创新组织的研发能力、技术产品的有效性和技术转让的可行性等三个因素共同作用的结果。研发能力被视为创新组织所拥有的研发资源与资产（尼古莱·J. 福斯，1998）。它反映了创新组织聚集研发人员、资本数量、技术设备的先进程度与数量以及资源整合的制度安排等。技术产品的有效性是指开发出来的设计与产品是否符合集群中小企业的需要。本集群企业不需要的技术产品，开发成本再低，质量再好，也难以卖得出去。技术转让的可行性涉及技术产品价值的实现机制。比如，在大多数中小企业既对新开发的技术产品缺乏购买力，也对其应用的不确定程度缺少预测力的条件下，补贴有助于技术产品价值的实现。因此，增大新技术成果的有效供给就是挖掘这三个因素的潜力，形成有力的技术辐射源。二是加强对中小企业技术人员的培训。随着技术成果的转让，培训构成了技术扩散所不可缺少的一个环节。通过各种技能培训，促使中小企业增加对采用创新成果的使用兴趣，帮助中小企业尽快地掌握其使用性能，提高对新产品和新设计的吸收、消化能力等，以增强对新技术、设备的新产品设计的深层次开发与应用。三是扩大对创新成果的售后服务以及再开发项目，这包括了对中小企业使用的新技术、设备提供维修、检验等专项的技术服务项目；利用技术人员相对密集的优势，对中小企业发展所

共同面临的工艺设计和开发实验等技术难题组织集体攻关，以解决每个企业普遍需要但又不愿意自己投资的通用型技术。四是加大促进技术信息交流、资源共享的中介力度，这涉及在集群内定期举办本行业内各种新技术产品、新工艺设计和新材料的展销会与演示会，为集群内中小企业提供技术信息传播与交流的平台，帮助中小企业与集群外创新组织与行业协会建立广泛的技术与市场联系，解决每个中小企业收集技术与市场信息成本过高的问题。

技术成果的迅速扩散产生了集群创新活动的聚集效应，由此拓展了技术服务的市场规模。这集中表现在两个方面：一种是集群外企业为获得知识外溢与信息共享的外部收益而迁入这个集群，致使集群规模扩大。随着创新成果的应用，中小企业集群显示了市场竞争力，于是，集群外越来越多的竞争对手产生了购置这类创新成果的需求，加上地方政府“不求所有，只求所在”的政策引力以及更宜于技术信息的交流与传播的集群环境，如在集群内有各种各样的广告，有各种商品展览、促销活动，有各种各样的学术或商业会议等，这些都以极为低廉的费用，甚至免费为企业提供专业性的商业信息（慕继丰，冯宗宪 2001），使得一些企业在非集群环境之间重新配置会亏本，而非集群地区向集群环境转移配置就不一定亏损，这就会引致大量的外部企业进入集群。广东省中山市古镇以 1300 多家灯饰企业对上游及配套产品的巨大需求，把深圳、惠州以及浙江省的一些灯饰设计、塑料、钢管、玻璃和包装以及配件生产的近 100 家企业和 200 多家生产灯饰的外来企业吸引到古镇的事实反映了集群内企业聚集的过程，从而拓展了古镇灯饰产品创新与设计中心的市场服务范围。

另一种是随着研发能力的增强，创新组织在为集群内企业提供技术产品与服务的同时，越来越多地承接集群外企业专用性技术产品开发与设计的订单。在地方政府对集群内的技术服务存在着价格补贴的条件下，承接集群外订单的价格往往高于对集群内企业进行技术服务的价格，这使得创新组织在完成地方政府作为控股者所规定的为本地企业提供必要的技术产品和服务前提下，愿意更多地承接外部订单。随着集群内企业创新能力的增长以及创新组织由集群内向集群外企业的市场扩展，对集群内企业技术服务的损失可以通过对集群外获得的收益加以弥补。在这种条件下，地方政府就可以采取以政策调整逐步替代补贴的方法，即在适当地降低地方政府对创新组织服务于本地企业数量规定的同时，也相应地减少价格补贴数量。当然，承接外部订单意味着创新组织技术服务的市场规模扩展了，这种扩展在一定程度上又会降低技术开发的单位成本，使技术服务价格趋于下降，这也创造了在补贴减少后创新组织收益不会下降，而中小企业对创新成果也有一定购买力的条件。本文第五部分给出的两个案例印证了技术服务的市场扩展对降低补贴的作用。

随着创新组织盈利能力的增强，集群内越来越多的企业开始设立研发机构，进而使得创新组织的功能发生结构性调整。企业集群与创新组织之间的互动，除了集群给创新组织带来了研发活动的盈利机会，使得企业孤立地开展研发活动远远没有他们之间聚集在一起通过交流与信息共享所进行的研发活动更加有效之外（Nelson，1993），创新组织通过技术扩散也使得创新活动变得异常活跃。这种创新活动的推动力量主要来自几个方面：一是越来越多的企业家们在体会到应用创新成果有利于提高生产率的甜头后以及在观察到创新组织盈利能力不断增长的事实后，他们就对本企业开展技术研发活动发生了越来越浓厚的使

用兴趣，并通过对创新组织的效仿，降低自己研发的风险；二是日益激烈的市场竞争使得有实力的企业越来越注重开发自己特有的专用性技术，而以开发通用型技术为主的创新组织则无法满足这种需要，这使得一些有实力的企业试图通过研发活动，增加个性化设计，提高质量，促进企业品牌的成长；三是随着企业技术与资本积累能力的增强，他们对研发活动具有了越来越强的投资能力；四是集群形成的技术外溢与信息共享的创新环境也使得集群内企业比集群外的企业更愿意自己增设研发机构，因为集群环境有利于降低企业搜寻信息、把握技术和获取资源的成本。随着集群内越来越多的企业对研发机构的设立，他们对创新组织研发出来的通用型创新成果的需求就相对下降了，这一方面可能促使创新组织进一步增大对外部市场的技术服务范围；另一方面会使创新组织承担的各种功能形成结构性调整，例如，在适当地降低创新设计品种的同时，增大对技术人员的技能培训，加强对新技术设备的使用、检验和维修等方面的技术服务，以及努力扩展信息交流与技术资源共享的中介职能等，如建立行业协会、市场信息机构、技术设备维修机构、培训机构、出口服务机构和信贷机构等中介组织，促进科研院所的技术转移与集群内企业建立产学研合作等。

总之，随着技术成果与信息资源的扩散，集群内中小企业在接受了创新能力的同时，也产生了对创新产品的更大需求，这样，一个创新组织带动了一批企业研发机构的兴起，形成了整个集群创新能力动态性强化的过程。当越来越多的企业开始研发本企业特有的专用性技术时，创新组织就会由设计与开发一般通用型技术产品为主转变为以信息传播、技术服务和深层次产品功能开发为主，这种结构性调整是集群内企业专业化分工与动态演进的一个客观结果。

五、案例分析

下面以广东省两个专业镇的创新过程为例，验证集群内地方政府如何有效地引入创新源、选择创新组织，以及促进创新成果有序扩散、使集群内企业创新能力得以成长的过程。

广东省南海市西樵镇是全国四大纺织生产和销售基地之一。早在明朝嘉靖年间，西樵镇已被称为“纺织之乡”，享有“广纱甲天下”的美誉。目前，集群内的纺织企业总数为1670家，其中织造厂1636家，染整厂10家，印花厂14家，服装厂25家，化纤厂2家。纺织机总数为28000台，其中剑杆机13600台，有梭机14400台，年产各类布料5.8亿米。西樵镇在纺织产业就业人数达到6万多人，占全镇本地劳动力与外来劳动力总量的43.8%。

20世纪80年代以来，西樵镇进入了有史以来纺织业最兴旺的时期。“千家店、万台机、亿米布”以及连接全国各地的布匹批发市场和销售网络，使西樵镇成为国内外知名的“家纺名城”。90年代中期以后，随着国内买方市场的出现，西樵镇纺织品的花色品种单调、产品质量档次低等，供给能力已无法适应收入水平不断提高的市场需求，市场占有率不断下降，产量不断收缩，企业关闭和转产的数量不断增加。到1994年，生产纺织品的企业只剩下了300多家，产量比最兴旺的80年代后期减少了60%以上。在西樵镇大部分

个体私营企业难以靠自身实力解决投资不足、设备老化和新产品开发率低等约束的情况下，镇政府于1997年从财政和管理区上缴的费用中拿出了200多万元，从韩国引进了一套世界上比较先进的电脑制版系统，用于纺织中的大提花工艺设计，并高薪聘请了一些技术人员，组建了一家专门从事新产品开发的研究机构——制版公司。该公司自1998年5月开业以来，开发了上万个面料新品种，并以低于开发新产品成本的300元价格转让给本地纺织企业。镇政府规定，制版公司设计开发出来的新产品，只能一次性出售给一家企业，不能向多家企业进行多次转让。此外，镇政府通过贴息帮助企业购置无梭剑杆机，替代有梭剑杆机，以改进产品质量，提高生产效率。1995~2000年间，无梭剑杆机的使用量从不足500台增加到13000台，无梭化率达到了48%，远远超过了全1国10%的平均水平，其结果，企业劳动生产率不仅提高了20%~30%，而且，新开发出来的品种面料的每米价格平均提高了20%以上。

当许多中小企业在尝到了购买新品种有利于提升盈利能力甜头的同时，也感到了制版公司开发的通用性设计品种越来越不适应每个企业追求专用性设计的要求，在这种条件下，具有一定实力的企业开始设立了自己的研发部门，研制与开发有企业自己特色并支撑自有品牌成长的新品种。到2000年底，整个西樵镇有100多家规模相对较大的企业自己设置了研发部门。随着研发部门的设立，制版公司把越来越大量的技术资源用于整个镇1900多家企业的设备维护、人员培训与信息交流上，新产品、新技术交易则主要集中在其与外部企业之间。到2000年底，制版公司已承接了6000多家外部企业对产品开发与设计样品的订单，这对制版公司创新能力的持续性发展产生了重要影响。

总之，西樵镇通过设立以技术开发为主的制版公司成为西樵纺织业从衰退走向复兴的关键。1998~2001年间，西樵镇纺织品销售收入从59.5亿元增加到了153.4亿元，年增长率为26.7%，成为仅次于浙江绍兴纺织城的全国第二大纺织生产基地。

另一个实例是以生产球形锁制品而闻名全国的中山市小榄镇。目前，全镇有3050个企业，其中以生产锁制品为主的五金企业以及与此相关的模具生产、零配件加工、电镀、塑料和包装等企业占企业总数的55%。2000年，该行业销售总收入为35.6亿元，占小榄镇工业销售总额的32%。在全镇近16万劳动力中，在以制锁为主的五金企业就业的人数占总劳动力的35%，整个小榄镇锁制品等五金产品的国内市场占有率达到了30%左右。本来小榄镇是没有金属制品资源的，那么，她是如何发展成一个以锁制品生产为主、纵向分工网络明显的专业镇的？

20世纪80年代初期以来，小榄镇一些具有企业家眼光的创业者抓住了球型锁制品在当时国内市场短缺，并随着居民收入水平的提高，其市场潜力还会迅速扩展的时机，利用在海外考察、访问的机会和亲戚朋友的渠道，从境外采购了各种各样的球形锁制品，然后将其分解，翻版其设计和制造工艺，开始模仿性生产，并赚取了第一桶金。当其他人发现这种产品的技术与资本门槛较低，且市场盈利空间较大后，也纷纷效仿跟进。这就形成了小榄镇生产球型锁制品的第一个浪潮。90年代初期，小榄镇从事球形锁制品的企业达到了300余家，其产品市场也得到了迅速扩展。然而，简单地模仿是缺乏持续竞争力的。当内地一些有技术积累的制锁企业在模仿的基础上加以改良，从而生产出款式更新、质量更好的球型锁时，小榄镇以简单仿效为支撑的产品就遇到了竞争对手的严峻挑战。90年代

中后期，小榄镇球型锁制品的市场占有率下降，特别是一些企业生产出来的产品合格率不高，退换率上升，致使整个小榄镇的锁制品质量声誉受到了影响。在这种压力下，镇政府通过让利、让股权等方式，把一些研制金属模具制品的科研院所引入小榄镇，同时专门从财政与管理费中拨出 500 万资金，引进国际上比较先进的 SLA 的设计模具快速成型的设备，共同组建了专门以研制开发各种五金制品模具为主的技术创新中心。该技术中心通过展示自己研制与开发的几百个新品种的模具样品与设计，承接了大量企业的外形设计、结构与功能设计等订单，并以成本价格帮助企业设计各种新型的生产模具以及提供技术服务。在全镇内制造球形锁的几百家企业中，有 35% 左右的企业从技术开发与设计中心获得了生产模具的新样品和新设计。得到了新模具设计的企业不仅在球形锁的品种档次上有了明显的改进和提高，而且，优质高价的新品种不断投入市场，使企业销售年收入增长了 30% 以上。其结果，一方面，与生产球形锁相配套的五金生产、电镀、零部件、塑料、包装以及运输等上百家企业相继进入小榄镇，形成了以球形锁制品为主的产业链条上的企业分工网络。另一方面，一些有实力的企业在感受到了研发投入给企业竞争力带来实际利益的条件下，为追求个性化、高质量产品与品牌，开始自己设置技术开发部门。到 2000 年底，年产值 2000 万元以上的 70 多家企业设立了技术开发部门，且大多数企业的研发开支占整个企业销售总额的 3% 左右。10 多家较大型企业在全国的一些大学、研究所共建了技术工程中心。随着技术开发部门在企业内的广泛设立，由镇设立的技术创新中心越来越注重技能培训、设备维修以及技术与市场的信息服务等，这反过来又进一步扩大了创新知识的外溢效应，促进了技术资源与信息的共享，降低了集群内企业研发的成本。

六、结论

通过上述的理论与案例分析，可以得出以下几点结论：首先，技术创新是以传统产业为主的中小企业集群兴衰的关键。在这些集群缺乏与作为技术先进者的跨国公司承包联系、内生的创新动机与能力不足的情况下，地方政府在引入外部创新源中应扮演重要的角色。其次，设立以公司性运作机制为标志的技术创新中心是有效地使用外部创新源，通过为本地中小企业集群提供有效的技术服务，使其创新能力得到持续增强的组织保证。在技术创新中心有偿提供创新成果的初期，补贴成为推动这些创新成果与资源向本地中小企业有效扩散的扶持性手段。最后，在技术创新中心通过市场扩展逐步产生了技术服务的规模收益，中小企业也通过新技术设计与产品的引入而增强了竞争能力，其结果，一方面，这为地方政府逐步取消补贴创造了条件；另一方面，一些具有技术与资本实力的企业为追求个性化和品牌化发展而开始自己设立研发机构。技术创新中心与企业内研发机构的创新互动必然会推动企业内研发机构更加注重个性化技术与产品的开发，而技术创新中心则更加注重开发技能培训、设备维护与信息交流等服务项目与功能。

参考文献

- [1] 萨克森宁 (A. Saxenien):《地区优势:硅谷和 128 公路的文化与竞争》，曹逢、杨宇光等译，上海远东出版社 1996 年版。

- [2] 今井,小宫隆太郎主编:《现代日本企业制度》,陈晋等译,经济科学出版社1995年版。
- [3] 慕继丰,冯宗宪:《企业簇与企业竞争力》,《工业企业管理》,2001年第7期。
- [4] 尼古莱·J. 福斯 (Foss, N. J.):《企业能力理论的兴起》,载尼古莱·J. 福斯,克努森主编:《企业万能:面向企业能力理论》,李东红译,东北财经大学出版社1998年版,第1~21页。
- [5] 仇保兴:《小企业集群研究》,复旦大学出版社1999年版。
- [6] 王珺:《产业组织的网络化发展》,《中山大学学报》,2002年第1期。
- [7] 王珺:《论集群经济的阶段性演进》,《学术研究》,2002年第7期。
- [8] Giacomo Becattini, Italian Industrial Districts: Problems and Perspectives, International Studies of Mgt & Org, 1991, Vol.21, No.1, pp 83~90.
- [9] M. P. van Dijk, Small Enterprise Clusters in Transition, A Proposed Typology and Possible Policies per Type Cluster, Working Paper, 1999.
- [10] Giovana Ceglie and Marco Dini, Small Enterprises Cluster and Network Development in Developing Countries: The Experience of UNIDO, Issues Paper of UNIDO, 1999.
- [11] UNCTAD, Promoting and Sustaining SMEs Clusters and Networks for Development, Commission on Enterprise Business Facilitation and Development, Geneva, 1998, 2~4, Sept.
- [12] 波特 (Michael E. Porter), Location, Competition and Economic Development: Local Clusters in a Global Economy, Economic Development Quarterly, 2000, Vol.14.
- [13] Piore, M. and C. Sabel, The Second Industrial Divide: Possibilities for Prosperity: Basic Books, New York, 1984.
- [14] Jean, Oi, Fiscal Reform and the Economic Foundations of Local State Corporatism in China, World Politics, 1992, Vol.45, No.1.
- [15] Birkinshaw, J. M., and Neil. Hood. Birkinshaw, Julian M., and Neil. Hood, Roles of Foreign Subsidiaries in Industry Clusters, Working paper, Institute of International Business, Stockholm School of Economics, 1998.
- [16] Hobday, M., Innovation in East Asia: Diversity and Development, In M. Dodgson and R. Rothwell (eds) The Handbook of Industrial Innovation, Hants, UK: Edward Elgar, 1994, 94~105.
- [17] Hobday, M., Innovation in East Asia: The Challenge to Japan, Hants, Edward Elgar, 1995.
- [18] Chandler, Alfred D. Orjan Solvell, and Peter Hagstrom, The Dynamic Firm: The Role of Technology, Strategy, Organization, and Regions, Oxford: Oxford University Press, 1998.
- [19] Hubert Schmitz, Collective Efficiency: Growth Path for Small-Scale Industry, The Journal of Development Studies, 1995, Vol.31, No.4, pp529~566.
- [20] Nelson, R. R., National Innovation Systems: A comparative Analysis, New York: Oxford University Press, 1993.

职业经理人进入民营企业影响因素 的实证研究*

——对经理人进入意愿的分析检验

张建琦 黄文锋

企业成长的重要因素是现有管理资源所能达到的企业扩张极限，吸纳新管理能力的数量和速度限制着企业的发展（Penrose，1959）。缺乏管理资源是企业扩张的主要约束力量（Richardson，1964）。我国民营企业管理人才普遍匮乏，但引入职业经理人却困难重重。其中尤以中小民营企业为甚。管理资源这种最重要的人力资本已经成为我国相当多民营企业成长的瓶颈（储小平，2002）。本文试图对经理人进入民营企业的内部制度性影响因素进行实证研究，对以往的理论进行验证，找出重大影响因素，比较和评价民营企业对经理人的吸引力，进而提出改进的途径。

一、对以往实证研究的回顾与评价

国外的学者对于员工进入企业的影响因素进行了大量的实证分析，证明雇员进入企业受到多种因素的影响（Richard & David，2000）。其中与我们研究相关的主要变量包括：①与工作相关的变量：奖惩的公平性（Mueller，1994），薪酬水平与工作的满意度（Mobley，Hand & Meglino，1979）；②个人的变量：受教育水平（Robert，John & Timothy，1994），个人价值取向（Hom and Griffeth，1995），对家族的责任感（Iverson，1992）；③环境变量：再就业机会（Mueller，1994），工作搜索成本（Kormendi，1979）。文献表明，上述变量在不同的企业和条件下作用的方向与大小不同。国外的研究通常隐含地假定：第一企业内部的人力资源管理制度和治理结构完善；第二，雇主和员工行为是诚信和规范的；第三，雇员的能力和价值观对于进入具有决定性作用，而雇主个人素质则无关紧要。显然，这些假定对于转轨经济中的中国民营企业难以成立，因而其结论的有效性也是有限的。

一些国外学者对中国择业者的选择偏好也进行过初步的研究。Fung（1996）的研究表明，个人兴趣、发挥所学知识的机会和初始工资是影响职业选择最重要的因素。Ding（1999）等人调查显示，高额薪酬和优越的工作条件对于中国的年轻人最具吸引力。Bu & McKeen（2001）则发现职业选择者对于物质的和精神的奖励都很敏感。Irene 和 Hang-yue（2001）证明工作保障、公司的规模和企业领导者的国籍等因素对学生的择业影响存在差

* 本研究由广东省高等学校人文社科规划项目（03SJA630002）和中山大学岭南学院资助。原载：《经济研究》，2003年第10期。

别。这些研究在一定程度上显示了择业者的价值取向。但大多数研究的对象宽泛，缺乏代表性，且未对所获数据进行严格的计量检验。

由于研究背景的差异，国内学者提出了不尽相同的影响因素。周立群和邓宏图(2001)的企业家流动模型说明，企业家追求的是在未来时间内的福利最大化，而福利(收入、偏好的实现、声誉)也就成了企业家流动的重要考虑因素。企业家流动主要受到流动主体的经济目标和价值取向、体制环境和市场机制等因素的影响和制约(邓宏图，2002)。储小平(2002)论证了家族企业成长瓶颈的实质是信任资源的约束，并指出中国的企业主和经理人都缺乏良好职业道德。由于雇主的价值取向和制度的缺陷，家族企业在某种程度上的“隐私”经营给企业从外部引入经理带来风险(李新春，2003)。而雇主诚信和价值取向也对经理人的进入具有重要影响(张建琦，2002)。国内的研究显示，经理人进入民营企业的障碍主要是制度性的。但这类研究目前仍处在定性分析阶段，充分的实证研究尚未开展，因而有必要对国内外研究提出的影响因素的可靠性及其在民营企业的作用进行进一步的实证检验分析。

二、调查对象、指标选择和分析方法

我们把职业经理人定义为具有一定的工作经历、受过系统的专业训练并以企业管理为职业的专门人才。为此我们选择调查的对象是具有大学本科学历以上，并且具有3~6年工作经验的职业经理人。我们发出500份调查问卷，回收337份，其中有效问卷312份。在有效答卷中，受访者中具有硕士学位的103名，占33%，具有学士学位的209人，占67%。有效样本的地区分布见表1。

表1 样本的地区分布

地区	广东	湖北	湖南	河南	浙江	山东	江苏	江西
样本量	191	15	14	12	11	8	7	6
比例	61.22%	4.81%	4.49%	3.85%	3.53%	2.65%	2.24%	1.92%
地区	福建 安徽	四川、 辽宁	重庆、河 北、北京	广西、云南、山西、 黑龙江、贵州、陕西	上海、宁夏、新疆、 甘肃、吉林、天津		未知	
样本量	各5	各4	各3	各2	各1		3	
比例	各1.60%	各1.28%	各0.96%	各0.64%	各0.32%		0.96%	

综合国内外的研究，我们选择了较为重要的10个企业内部制度因素进行调查(见表2)。首先由经理人对这些因素进行重要度评价，以便选择需要进行计量检验的因素。其次由经理人对民营、国有和外资企业的10项因素的状况进行评价(见表4)。最后根据重要度评价数值，对其中5个最重要的因素进行了计量分析和统计检验。

表 2 经理人对进入民营企业影响因素重要度排序

重要度排序	影响因素	重要度得分	标准差	权重 ^①
1	职业的发展机会与空间	3.141	2.456	0.166
2	现金收入与福利水平	3.939	2.340	0.150
3	雇主对经理人的诚信	4.219	2.562	0.144
4	企业规模与形象	5.270	3.340	0.121
5	收入分配与职务晋升公平性	6.170	2.811	0.102
6	人际关系与组织氛围	6.865	2.810	0.088
7	雇主的素养与受教育水平	7.569	3.562	0.073
8	领导民主性	7.727	3.161	0.069
9	社会保障制度的完善程度	8.752	3.133	0.048
10	工作的物质环境和条件	9.136	3.067	0.039

注：表中的得分接近 1 的因素为最重要，接近 10 的为最不重要。

三、对调查数据的回归分析与检验

1. 研究假设与分析方法

根据以往的研究，我们对调查中的 5 项最重要的影响因素提出以下假设：

假设 1：企业提供的职业发展的机会与进入意愿正相关。企业不仅是经理人人力资本使用发挥和实现价值的场所，更是经理人人力资本不断成长和提升的载体。经理人进入民营企业不仅仅是为了追求短期的现金收入，同时也是为了实践自己的知识，获得更多的管理能力和经验。因此，企业的职业发展的机会越多，空间越大，对经理人的吸引力越大。

假设 2：当前收入和福利水平与进入意愿之间具有正相关关系。现金收入和福利水平既是对职业经理人教育和人力资本投资的补偿，也是经理人自身经济和社会价值最直接的体现，因此高额收入和优厚的福利条件会增大企业对经理人的吸引力。

假设 3：雇主对经理人的诚信与经理人的进入意愿正相关。雇主的诚信是指雇主对经理人和企业在诚实和信用方面的态度和行为。经理人与雇主之间是一种长期交易关系，在过程中会面临很大的不确定性。如民营企业大量存在的“隐私”经营会给经理人带来很大的风险；企业做假账会使经理人的股权收益化为泡影，对顾客和社会的欺诈会使经理人怀疑雇主对自己的诚信。雇主诚信的改善有利于降低上述风险，从而提高经理人的进入意愿。

假设 4：收入分配和晋升的公平性与进入意愿呈正相关关系。在进入过程中，经理人与雇主之间不仅是后验性的关系交易，也存在一种互动的博弈行为，其中具有很大的不确定性。公平能够稳定博弈收益的预期和变化，构成了良好的交易维系机制。因此雇主的公平对于雇员的进入影响很大（Ronald and Donald, 1994）。收入分配和职务晋升的公平性反映企业管理的科学化和规范化程度，以及经理人的工作与发展是否具有一个公平竞争的环境。这一因素越完善，经理人的进入意愿越高。

① 各因素权重 = (11 - 各因素的重要度得分) / ∑ (11 - 各因素的重要度得分)。

假设 5：企业的规模和形象与经理人的进入意愿正相关。不同规模的组织在工作条件、报酬和保险、工作环境的支持性和工作的匹配上具有很大的不同 (Jodie, Katherine & John, 2001)。大企业的管理通常比较先进和规范, 经理人可以从中学到更多的管理知识和经验, 因而对人力资本的增长具有重要影响。同时在规模大、声誉高的企业的工作经历会提高经理人再就业的谈判地位, 因而规模大、形象好的企业更容易引入经理人。

我们采用 Logit 模型来建立经理人的决策行为模式。假设决策者为效用最大化追求者, 效用函数为线性可加成形式。在我们的模型中, 经理人只有愿意进入和不愿意进入两种选择, 故我们采用二元选择的 Logit 模型。计量模型为:

$$\text{Logit}P_j = \alpha_j + \sum_1^i \beta_{ij}x_{ij} + e_j$$

其中: P 为进入的概率; α 为常数项; x 为解释变量; β 为待估参数; e 为随机干扰项; j 与 i 分别表示第 j 个样本和第 i 个解释变量。被解释变量是经理人进入民营企业的意愿; 愿意进入的变量设为 1, 不愿意的设为 0。在 312 份的有效问卷中, 203 人选择了进入, 109 人选择了不进入。表 2 中的前 5 项最重要的因素为解释变量。

2. 实证检验的结果及其解释

表 3 的检验结果表明, 职业发展机会与空间变量的系数为 0.753, 在 1% 水平下显著。表明该因素与经理人进入意愿呈强正相关关系, 从而强有力地支持了我们假设 1。该变量对经理人的进入意愿的影响最大。这与 Fung (1996) 对于中国年轻的职业选择者调查的结果接近。但在 Fung 研究中职业发展的机会在所有因素中位居第二, 而在我们的直接统计分析 (表 2) 和回归分析 (表 3) 中这一因素都居首要地位。这表明职业经理人的观念已经开始变化, 自身的成长和发展上升到了第一位, 反映出我国职业经理人阶层的成长性要求在不断提高。

表 3 5 项最重要因素的回归分析结果

Variable	Coefficient	Std. Error	z-Statistic	Prob.
截距	- 6.228	0.955	- 6.523	0.0000**
职业发展的机会与空间	0.753	0.182	4.139	0.0000**
工资与福利水平	0.407	0.206	1.973	0.0485*
雇主对经理人的诚信	0.392	0.191	2.056	0.0397*
收入与晋升的公平性	0.360	0.178	2.031	0.0422*
企业规模与形象	0.182	0.229	0.798	0.4249
LR statistic (5 df)	81.191**			

注: ** 表示在 1% 的水平显著; * 表示在 5% 的水平显著。

工资与福利水平变量的系数为 0.407, 且在 5% 水平下显著。这一结果支持了我们的假设 2。这表明现阶段现金收入在经理人的就业选择中依然占据十分重要的地位。这一结

果与 Ding (1999) 等人的结论相近。不过在 Ding 的调查中, 高额薪酬对中国的年轻人的吸引力是第一位的。在 Fung (1996) 的研究中, 初始工资的重要性是第三位的。我们的检验结果说明, 在转轨经济条件下, 经理人普遍偏好现实收益。就业市场的机会多, 创业的进入障碍低, 必然引发职业经理人的短期性行为冲动。目前中国经理人还很少是职业化或专业化的, 目标在不少情况下是游移不定 (李新春, 2000, 2003)。同时, 现阶段民营企业普遍缺乏长期战略, 经营不稳定, 且寿命周期短。经理人的短期利益追求也是对人力资本与非人力资本合作关系难以长期化和规范化的一种理性反应。

雇主对经理人的诚信变量的系数为 0.392, 并在 5% 的水平下显著。这一结果支持了我们的假设 3。表明雇主的诚信不仅影响经理人的职业发展和经济价值的实现, 而且影响雇主与经理人之间信任关系的建立。

收入分配与职务晋升公平性的系数为 0.360, 且在 5% 水平下显著, 支持了假设 4。说明收入和晋升的公平性是人力资本价值实现的保证和前提, 分配和晋升制度的完善对于经理人的进入有积极的影响。而国内外的实证研究尚未对这两个因素进行过检验。

企业规模与形象变量的系数为 0.182, 在统计检验上不显著。没有支持我们的假设 5。这表明对上述 5 个因素共同考虑的时候, 经理人对企业规模和形象的要求不突出, 或者在其他因素条件较好的情况下, 这一因素可以忽略。这说明中小民营企业即使名声和规模不大, 但只要相关的制度完善, 对高素质的管理人才仍然具有吸引力。我们的结论和 Irene & Hang-yue (2001) 的实证研究结果相一致。他们的研究显示, 在对晋升的可能、学习的机会、薪酬、海外旅游的机会、工作保险、公司的规模等工作因素的重要性排序中, 规模因素被就业者排到了最后一位。

四、民营企业对职业经理人吸引力的比较评价及其改进方式

表 4 显示了经理人对民营企业 10 项内部制度因素水平和满意程度的评价。令人欣慰的是, 最受经理人重视的职业发展机会与空间在所有因素中获得了最高的评价 (3.875), 且高于外资企业 (3.666) 和国有企业 (2.749)。这是在 5 项最重要的因素中惟一超过其他类型企业的因素。这表明尽管民营企业各种制度尚不完善, 但经理人仍有可能发挥自己的才能, 也反映出民营企业近年来的迅速成长为经理人提供了更大的发展空间。但该因素评价的标准差 (0.912) 较高, 高于国有企业 (0.906) 和外资企业 (0.765), 说明经理人对这一因素的评价的差异较大, 反映出民营企业在提供经理人学习和发展机会方面还存在明显的缺陷, 企业之间具有较大的不平衡性。工资福利水平在所有因素中居第二位, 对民营企业的评价 (3.489) 高于国有企业 (2.720), 低于外资企业 (4.434), 吸引力居中, 说明民营企业的效益较好。从标准差来看, 国有企业间 (1.932) 目前的收入差距最大。收入分配和晋升的公平性评价居第四位, 也高于国有企业 (2.720), 低于外资企业 (4.434), 吸引力亦居中间水平, 说明民营企业的收入分配和晋升制度正在向着规范化的方向不断完善。

表4 经理人对民营、国有和外资企业相应的各项因素水平的评价

项 目	民营企业			国有企业 ^①			外资企业		
	平均得分	加权得分	标准差	平均得分	加权得分	标准差	平均得分	加权得分	标准差
职业的发展机会与空间	3.875	0.643	0.912	2.749	0.456	0.906	3.666	0.609	0.765
工资与福利水平	3.489	0.523	0.799	2.720	0.408	1.932	4.434	0.665	0.597
人际关系与组织氛围	3.484	0.307	0.935	4.100	0.361	0.884	2.778	0.244	0.940
收入分配与职务晋升公平性	3.186	0.325	0.886	2.331	0.238	0.789	4.141	0.422	0.616
工作的物质环境和条件	2.968	0.116	0.774	3.070	0.120	0.788	4.367	0.170	0.623
雇主的素养与受教育水平	2.675	0.195	0.728	3.174	0.232	0.715	4.650	0.339	0.517
企业规模与形象	2.672	0.323	0.697	3.640	0.440	0.876	4.534	0.549	0.615
雇主对经理人的诚信	2.592	0.373	0.852	3.341	0.481	0.947	4.116	0.593	0.745
社会保障制度的完善程度	2.196	0.105	0.670	4.013	0.193	0.745	4.257	0.204	0.651
企业领导的民主性	2.006	0.138	0.861	2.418	0.167	0.890	3.707	0.256	0.775
加权平均得分	0.305			0.310			0.405		

注：表4的因素评价采用 Likert-type scale 五点量表计分法，按调查因素的水平分别赋予数值1, 2, 3, 4, 5。

令人遗憾的是，在统计分析中影响力居第三位的雇主对经理人的诚信因素在表4中却被经理人排到了第八位，其得分（2.592）远低于国有企业（3.341）和外资企业（4.116）。这表明民营企业雇主对经理人的诚信度相对低下，这对职业经理人有明显的排斥力。企业规模与形象的影响力在统计分析中位居第五位，但在经理人的评价中位居第七位，评价分数（2.672）远低于国有企业（3.640）和外资企业（4.534）。它反映出相当多的民营企业规模偏小，轻视信用和品牌，短期行为和急功近利的倾向严重。这种状况对经理人的进入有明显的消极影响。在5项影响力最大的因素中，民营企业有2项低于国有企业，有4项低于外资企业。

总体来看，在10个因素中，经理人对民营企业的评价有7项低于国有企业，有9项低于外资企业。我们用各因素的评价分数加权平均数表示各类企业对经理人的综合吸引力。民营企业对经理人的综合吸引力为0.305，国有企业的综合吸引力为0.310，外资企业的综合吸引力为0.405。可以看出，民营企业的综合吸引力已经接近国有企业。但在三类企业中，民营企业的综合吸引力仍然最低，尤其是与外资企业的差距较大。

从表4可以看到，与国有和外资企业比较，单项因素吸引力（加权得分）差距最大的仍然是企业的规模与形象、雇主对经理人的诚信和社会保障制度的完善程度。这三个因素是造成民营企业综合吸引力低下的主要因素。但从博弈论的角度来看，这并不意味着人才竞争中民营企业需要首先从最弱的方面进行改进。下面我们将采用计量检验方法说明各因

① 国有企业含国有全资和国有控股企业。

素改进的优先序。^①

我们建立三个 Logit 计量模型。模型 1 说明与国有企业比较，民营企业各因素的差距（或优势）对经理人进入意愿的影响。模型 2 解释与外资企业比较，民营企业各因素的差距（或优势）对经理人进入意愿的影响。模型 3 说明与国有或外资企业各因素中评价最优的数值比较，民营企业各影响因素的差距（或优势）对经理人进入意愿的影响。

$$\text{模型 1: } \text{Logit} P_j = a_j + \sum_1^i \beta_{ij} (X_{ij} - G_{ij}) + e_j$$

$$\text{模型 2: } \text{Logit} P_j = a_j + \sum_1^i \beta_{ij} (X_{ij} - W_{ij}) + e_j$$

$$\text{模型 3: } \text{Logit} P_j = a_j + \sum_1^i \beta_{ij} [X_{ij} - \max(G_{ij}, W_{ij})] + e_j$$

其中：P 为进入意愿的概率；a 为常数项；X 为民营企业解释变量；G 为国有企业的解释变量；W 为外资企业的解释变量； β_{ij} 为待估参数；e 为随机干扰项；j 和 i 分别表示第 j 个样本和第 i 个解释变量。

在 Logit 模型中，解释变量发生一单位变化时，所带来的进入行为期望概率的变化称为因素对于进入行为的边际效果。我们利用 Logit 函数的条件期望 $E(y/x) = \frac{1}{1 + \exp(-x\beta)}$ ，推导出解释变量对于期望概率的微分： $\frac{\partial E(y/x_i)}{\partial x_{ij}} = \frac{\exp(x_i\beta)}{[1 + \exp(x_i\beta)]^2} \beta_j$ 。由于 $\frac{\exp(x_i\beta)}{[1 + \exp(x_i\beta)]^2}$ 恒正，各解释变量的边际效果等于正的固定参数乘以计量估计系数，因此我们用模型的计量估计系数直接认定各解释变量边际效果的大小排序。

表 5 三类企业比较的计量结果

变 量	与国有企业比较的 系数 (模型 1)	与外资企业比较 的系数 (模型 2)	与国企外企综合比较 的系数 (模型 3)
截距	0.319 (0.232)	1.698 (0.35)**	3.667 (0.561)**
职业发展的机会与空间	0.344 (0.115)**	0.608 (0.112)**	0.922 (0.234)**
工资与福利水平	0.301 (0.132)*		0.641 (0.247)**
雇主对经理人的诚信	0.319 (0.115)**		0.284 (0.138)*
企业规模与形象		0.324 (0.159)*	0.442 (0.216)*
收入分配与职务晋升公平性		0.476 (0.143)**	0.756 (0.184)**
LR statistic	48.933**	61.421**	66.246**

注：**表示 1% 的水平显著；*表示 5% 的水平显著。

① 前面的分析显示民营企业大部分的因素落后于其他企业。我们把民营企业各因素水平对于其他企业的超越占优的情况也简单地视为一种改进。

表 5^① 的计量结果显示, 边际收益上, 当与国有企业进行人才竞争时, 民营企业在职业发展机会和空间 (0.344)、对经理人的诚信 (0.319) 和工资福利水平 (0.301) 三方面的改进能最有效地提高对经理人的吸引力。当与外资企业竞争时, 首先对职业发展机会和空间 (0.608)、收入分配与职务晋升公平性 (0.476) 以及企业规模与形象 (0.324) 等因素的改善则最为有效。当与上述两类企业同时竞争时, 民营企业在同等程度地改进各因素的情况下, 对综合吸引力提高的效果由大到小的顺序是: 职业发展的机会与空间、收入分配与职务晋升公平性、工资福利水平、企业规模与形象、雇主对经理人的诚信。

五、结论与启示

实证结果表明, 影响经理人进入民营企业最重要的因素依次为职业发展的机会与空间、工资与福利水平、雇主对经理人的诚信、收入与职务晋升公平性、企业规模与形象。它们与民营企业的吸引力均正相关, 其中前四个因素在统计检验上更为显著。综合我们考察的 10 个因素, 与国有和外资企业相比, 民营企业的综合吸引力仍然最低, 尤其与外资企业的差距较大; 但其综合吸引力已经接近国有企业。值得一提的是, 民营企业在经理人的职业发展机会与空间这一首要影响因素上的吸引力已经超过了国有企业和外资企业, 它显示出民营企业的家族制经营模式可能最终会被打破, 经营者的升级换代有望在企业内部实现。经理人对现金收入的高度重视, 表明民营企业需要具有长期稳定发展的战略观念, 同时在完善监督机制的前提下, 对经理人持股采取谨慎与渐进的态度。民营企业不能只是强调改善对经理人的信任, 更重要的是兑现对社会和经理人的承诺。尽管中小企业的声誉和规模有限, 但只要相关的制度完善, 对经理人仍然具有吸引力。

在边际效益上, 各因素对综合吸引力提高的贡献大小依次是: 职业发展的机会与空间、收入分配与职务晋升公平性、工资与福利水平、企业规模与形象、雇主对经理人的诚信。在整体的人才竞争中, 民营企业应当按照上述优先序改善其综合吸引力。当与国有企业进行人才竞争时, 应优先改进民营企业在职业发展机会和空间、对经理人的诚信和工资福利水平三方面的因素; 当与外资企业竞争时, 应重点改善职业发展机会和空间、收入分配与职务晋升公平性以及企业规模与形象, 从而长期有效地提高吸引经理人的竞争力。

参考文献

- [1] 储小平:《职业经理与家族企业的成长》,《管理世界》,2002年第4期。
- [2] 邓宏图:《企业家流动的博弈模型:经济含义和企业家能力配置》,《经济科学》,2002年第3期。
- [3] 邓宏图,周立群:《经理人市场:供求与交易关系研究》,《江苏社会科学》,2001年第4期。
- [4] 李新春:《企业家过程与国有企业的准企业家模型》,《经济研究》,2000年第6期。
- [5] 李新春:《经理人市场的失灵与家族企业治理》,《管理世界》,2003年第4期。
- [6] 张建琦:《经理人“背叛”的机理与雇主的对策取向》,《管理世界》,2002年第5期。
- [7] Bu, N., Mckeen, C. A., Work goal among male and female business student in Canada and C China: the

① 经检验,表中系数空白的因素在该模型下可忽略。

- effect of culture and gender , *International Journal of Human Resource Management* , Vol.12 , pp166 ~ 183.
- [8] Devine T. J. & Kiefer N. M. , *Empirical labor economics* , New York : Oxford University Press , 1991.
- [9] Ding , X. , *Twenty years through the eyes of Chinese youth : A survey* , *China Today* , May , 1999.
- [10] Fung , C. K. , *Job choice and job mobility for Chinese young professionals* , *Hong Kong Manager* , Nov ~ Dec , pp20 ~ 25.
- [11] Irene , Hang-yue , *Gender differences in job attribute preferences and job choice of University student in China* , *The Journal of Applied Business Research* , 1996 , Vol.18 , No.2.
- [12] Jodie , L. Hertzog , John F. Zip , Shell M. MacDermid , & Katherine B. Kensinger , *The role of organizational size and industry in job quality and work-family relationships* , *Journal of Family and Economics Issues* , 1994 , Vol.22 , No.2 , Summer , pp191 ~ 216.
- [13] Kopelman R. , *Rational and construct validity evidence for the job search behavior index : Because intentions often come naught* , *Journal of Vocational Behavior* , 1992 , Vol.40 , pp269 ~ 287.
- [14] M. Ronald Buckley , Donald B. Fedor and David S. Marvin , *Ethical considerations in the recruiting process : a preliminary investigation and identification of research opportunities* , *Human Resource Management Review* , Vol.4 , No.1 , pp35 ~ 50.
- [15] Penrose , Edith T. , *the theory of the Growth of the Firm* , Oxford : Basil Blackwell Publisher , 1959.
- [16] Richardson , G. , *The Limit to a Firm 's Rate of Growth* , *Oxford Economic Papers* , 1964.
- [17] Richard R. Cline and David A. Mott , *Job Matching in Pharmacy Labor Markets : A Study in Four States* , *Pharmaceutical Research* , 2000 , Vol.17 , No.12.
- [18] Robert J. Vandenberg and Jai Hyun Seo , *Placing Recruiting Effectiveness in Perspective : A Cognitive Explication of The Job-choice and Organizational-entry Period* , *Human Resource Management Review* , 1992 , Vol.2 , No.4 , pp239 ~ 273.

竞争力与企业规模无关的形成机制^{*}

孙洛平

一、问题的提出

当产出增加引起单位产出成本下降，就称为规模经济；若投入加倍而产出大于加倍，就称为规模报酬递增，反之，称为规模报酬递减^①。从个别企业来看，具有递增规模报酬似乎是合乎逻辑的。比方说，两个一样的企业合并，生产方法的可选择范围只会扩大，最起码也能够保持原有的方法不变，所以随着生产规模的扩大，单位产品的成本不应该增加。

在竞争力与企业规模的关系上，理论推论与实际观察一开始就存在冲突。单个企业规模报酬递增预示企业的规模越大其生产的效率就越高，将这个推论用于整个经济，可以得出以下结论：在产品结构一致的情况下，一个经济系统整体的生产效率较高是与企业的平均规模较大联系在一起。马歇尔（1890）显然注意到了产业集聚现象，即若干中小企业集聚在一个产业区内形成高度分工的体系，产业区里的企业规模并没有简单地增大，可是整个产业区显示出很强的整体竞争力。在马歇尔看来，单个企业可以没有规模经济，产业区作为一个整体则具有规模经济，他称之为外部规模经济，意指若干企业通过相互分工而使得生产效率提高和最终产品的成本下降。产业集聚现象提出了另一种机制，即企业的平均规模较小与经济整体生产效率较高是有一定关联的。

杨格（1928）在《递增报酬与经济进步》一文中，从分工的角度来解释企业的平均规模与经济整体竞争力的问题。他认为，决定一国产业效率的不是企业的规模，而是分工专业化的水平。在他看来，分工可以分为生产迂回程度的增加和不同行业的分工两种。生产迂回程度的增加才是提高生产效率的主要原因。斯蒂格勒（1951）更明确将“市场容量限制劳动分工”的斯密定理解释为：“依斯密定理，在不断壮大的产业中，典型的情况应该是垂直非一体化，而垂直一体化倒是衰落产业的特征。”^② 上述观点可以归纳为：当纵向

^{*} 原载：《经济研究》，2004年第3期。

^① 若用 q 表示产量， AC 表示平均成本， MC 表示边际成本，用 S 来度量规模经济： $S(q) = AC(q) / MC(q)$ ，那么 $S(q) > 1$ 表示在产出水平 q 有规模经济， S 恒大于 1 表示在全局有规模经济。规模经济与规模报酬之间存在一定的关系。若所有的要素扩大 t 倍，产出扩大的倍数大于 t ，则称之为正规报酬或规模报酬递增，反之则为负规模报酬或规模报酬递减。规模经济是基于成本来考虑的，而规模报酬是基于技术来考虑的。本文主要关注产品的成本与定价的比较，因此不严格区分这两个概念。

^② 斯蒂格勒：《市场容量限制劳动分工》，见：《产业组织和政府管制》，三联书店 1989 年版，第 29 页。

分工在企业间进行时,经济整体生产效率提高将伴随着企业规模的缩小,这与马歇尔对产业集聚现象的描述是一致的。不过,问题并没有得到解决,究竟是何种原因决定分工应该在企业内进行还是在企业间进行,这一点并不明确。

科斯(1937)的《企业的性质》一文明确提出了企业与市场关系的问题,他引入交易费用的概念来解释企业的规模。按照科斯的见解,利用市场进行交易是有成本的,企业之所以存在,就是因为它可以节约交易费用。当市场的交易效率提高时,企业的平均规模将变小,反之则变大。虽然科斯并不关注利用市场还是利用企业来组织分工的问题,但是由科斯的逻辑可以得出,当一个产业在一个地理区域集中并导致交易效率提高时,整体规模经济的出现和竞争力的提高将伴随企业平均规模的减小。

不过,上述关于生产效率、分工和企业规模的洞见却难以形式化。杨小凯和黄有光(1993)从企业的内部分工结构入手,提出间接定价企业理论,即管理者的贡献难以在市场上直接定价和交易,因而借助于企业间接测定其贡献。他们用超边际分析方法将分工专业化、交易费用和企业规模(或制度)三个因素包容于一个模型中,解决了将上述思想形式化的问题。超边际分析方法的做法是比较若干个可能的不同分工状态(角点解)的效率,以得出最优的或最可能出现的解。分工是通过企业内部还是市场来实现,取决于模型中关于生产效率和交易费用的参数值的变化,一个直接的结论就是生产效率与企业平均规模无关。然而超边际分析方法也有不尽人意的地方,几乎所有的模型都是一种处理范式^①,这种范式的单调性和分析方法的特殊性使之难以为主流的新古典经济学所容纳。

与古典经济学关注分工不同,新古典经济学关注的是竞争及其均衡的存在。单个企业的规模报酬递增性显然排斥竞争,它预示一个行业只应该有一个企业存在,这显然与实际不符。因此,协调单个企业的规模报酬递增性与不同企业竞争均衡的存在性,在新古典经济学家看来始终是产业组织领域里的一个难题。在新古典经济学的分析框架中处理这个难题的,首推迪克西特和斯蒂格利茨(1977)关于垄断竞争的论文,它开创性地提出了一个新的分析思路。克鲁格曼(1979)将这一方法应用到国际贸易的分析得到非常漂亮的结果,以及后来若干学者在地理经济学、城市经济学、内生经济增长理论等诸多方面的应用,由此形成新古典经济学的一个标准的分析框架。这个分析方法的基本要点就是递增规模报酬效应,尤其是基于单个企业的递增规模报酬效应,随之而来的是经济整体竞争力的提高伴随着企业平均规模扩大的预言。

迪克西特和斯蒂格利茨关于垄断竞争的模型虽然是高度简化的,但显然包含有合理的经验内容,正因为如此,它才能在国际贸易、新经济地理学等领域得到广泛的应用。不过,经济实践表明,伴随经济整体竞争力的提高,企业的平均规模可以增大,也可以缩小,或者保持不变^②。因此,新古典经济学的垄断竞争分析框架只有相对的合理性,它并不能涵盖经济整体竞争力与企业平均规模关系的全部内涵。

① 用超边际分析方法建立和求解模型的一般步骤:建立描述分工专业化的非线性规划模型;在各种可能的角点解中排除绝大多数无效角点解;对余下的角点解利用边际分析方法求解局部最优解;通过比较各个局部最优解获得全局最优解;利用交易效率 and 生产条件这两个参数和产品的相对价格,比较各个角点解的个人效用水平;利用决策前完全相同的人在决策后应获得相同的真实收入条件求解产品的市场价格。

② 参见张永生:《厂商规模无关论——理论与经验证据》,中国人民大学出版社2003年版,第38~109页。

然而，我们面临的难题并不在于哪一种理论的解释力更强，而在于面对杨格-斯蒂格勒与迪克西特-斯蒂格利茨这两种似乎是不可调和的分析框架，要作出非此即彼选择的困难。本文试图说明，只要将分工种类数目作为一个可选择的变量引入企业模型，就可以使用新古典经济学的边际分析方法来处理利用市场还是利用企业来组织分工的古典经济学问题。

本文的分析表明，在一般情况下，我们找不到一个科斯式的市场交易费用的平衡点，在这个平衡点上，利用企业组织分工与利用市场组织分工具具有相同的效率。即使这个平衡点存在，它也不是一个稳定的平衡。分工一旦稍微偏向于在企业内组织，企业组织分工的成本会相对于市场越来越小；反之，分工一旦偏向于利用市场来组织，市场组织分工的成本会相对于企业越来越小。文中的模型能够解释为什么在经济实际中，要么发挥企业组织的优势，从而企业的平均规模较大；要么发挥市场机制的优势，从而企业的平均规模较小。我国若干中小城镇的产业集聚现象，就是利用市场组织分工从而发挥其竞争优势的典型例子。

二、企业和市场组织分工的费用

分工，尤其是纵向分工是提高生产效率的最重要途径，问题的核心集中在应该利用市场还是利用企业来组织分工。要解决这个问题，需要模型化企业组织的成本和市场交易费用，然后比较两者的定价水平和平均成本，以确定各自的优势。

先从生产的成本分析入手。影响企业生产成本的因素可以分为两个方面：其一，由生产技术条件决定的成本，比方说生产中投入的原材料、水电、折旧、直接作业的人力等成本；其二，企业的组织管理成本。为简化表述，假定企业只生产一种产品，而且单位产品要求投入的原材料、折旧、一线劳动工资等成本保持不变，这部分成本可以归之为变动成本。管理成本与产量的关系要复杂一些，通常表现出随产量阶段性变化的特征，即在一定的产量变化范围内保持不变，因此不能简单的将其归之为固定成本或变动成本，我们对此作进一步分析。

企业管理成本与管理工作的复杂程度和产量有关。假定管理工作的复杂程度与分工的种类数相关，为简化讨论，设企业内部的分工有 K 种，产量为 Q ，管理的费用可以写为：

$$M = M(K, Q) \quad (1)$$

满足： $M(K, 0) > 0$ ，表示只要有分工，就有管理费用，即使产量为零也是如此； $\partial M / \partial K > 0$ ，表示管理费用随分工种类数的增加而增加； $\partial^2 M / \partial K^2 > 0$ ，表示管理的效率随分工种类数的增加而递减，可以理解为随着管理对象的增加，管理的效率降低； $\partial M / \partial Q \geq 0$ ，表示管理费用至少不因产量增加而减少。

上面关于企业管理费用与企业包容的分工种类数相关的假定与通常的假设不同，这里需要对此加以说明。设想一种产品的生产由 K 个不同的纵向分工构成，分别由 K 个工人进行生产，由于每个工人的工作各不相同，协调、评价和考核工人的工作构成企业管理费用的最重要来源。在管理人员具有有限管理幅度的假定下， K 的增大势必使管理的科层增加，必将导致信息扭曲和官僚成本产生。如果在原有 K 个工人构成的纵向分工生产链之

外再增加一条相同的生产链，即产量增加一倍，那么管理费用显然不会也加倍，主要原因在于管理的复杂程度没有增加。产量翻一翻对于管理费用有正反两方面的影响：一方面，一线生产人数增加了，需要评价和考核的工作量增加；另一方面，同一分工的工作有了相对比较的可能，甚至有可能通过建立竞赛机制而减少评价和考核的工作量。因此，产量变化对于管理费用的总的影 响是不容易确定的。不过，有一点可以肯定，产量变化对管理费用的影响远不能与分工的种类数变化的影响相比。

为简化分析，假定管理费用与产量无关，即 $\partial M/\partial Q = 0$ ，企业生产的总成本函数可以写为：

$$C(K, Q) = cQ + M(K) \quad (2)$$

式中： c 为变动成本。以上成本函数的一个基本特征是，企业生产有规模经济，即平均成本随产量的增加而不断减小，除非管理费用为零。

再来考察市场交易费用的描述。利用市场组织分工的交易费用可以分为两类：交易性费用和契约性费用^①。如果是利用市场价格机制，交易费用主要表现为交易性费用，应该是搜寻交易对象和讨价还价的成本。如果交易物品在规格、品质等方面具有不确定性，人们只能借助于契约的方式来减少事后的机会主义行为，交易费用主要表现为契约性费用，应该是契约的签订、执行、监督以及违约处理等费用。本文只关心给定产品和生产技术条件下的分工组织问题，因此，市场交易费用可以简化为单纯的搜寻交易对象和讨价还价的费用。

搜寻交易对象的费用随着独立的可交易对象数目的增加而减小，不妨假定搜寻费用与交易对象的数目成反比。讨价还价的费用一般与议价双方成交价格的不确定性成比例，成交价格越是不确定，人们越是愿意花费时间和精力去讨价还价以增加自己的利益。成交价格的不确定性取决于交易双方市场议价力量的对比，在卖方垄断或买方垄断的市场上，买卖双方的力量对比是确定的，价格由占有优势的一方说了算，议价的费用并不大。在有多个买方与卖方的市场上，议价的费用也不大，因为交易双方都有退路，谁也要挟不到谁，成交价格的不确定性也小。双边垄断的市场成交价格具有最大的不确定性，议价费用也最高。综合以上分析，假定在给定产品和技术条件下，市场交易费用可以写为：

$$\frac{a}{n_i \times n_j} \quad (3)$$

式中： a 为交易费用参数； n_i 和 n_j 分别表示买者和卖者的数目。如果只有一个买者和一个卖者， a 就是双边垄断情况下的搜寻和议价费用。

三、市场分工模式和企业分工模式的产品定价

利用市场来组织分工，对于最终产品的价格有两方面的影响：其一，每一个市场交易

^① 科斯（1937）认为的交易费用主要有：其一，“通过价格机制‘组织’生产的最明显成本就是所有发现相对价格的工作”；其二，“市场上发生的每一笔交易的谈判和签约的费用也必须考虑在内”。参见《企业的性质》，见：《企业、市场与法律》，第5~6页。类似的叫法为外生交易费用和内生交易费用。

环节都会在投入的中间产品上加价，层层加价因素使产品的价格升高；其二，每一个市场交易环节因为有多多个同类的交易者竞争，竞争因素使产品的价格降低。考察两个相邻分工部门之间的定价关系。假定上游分工 A 共有 m 个企业，每个企业包含 K_A 个分工环节，下游分工 B 共有 n 个企业，每个企业包含 K_B 个分工环节。两个市场上的产品分别是同质的，企业之间没有合谋。每生产 1 单位最终产品要求 1 单位 A 和 1 单位 B 作为中间产品投入，生产过程如图 1 所示。

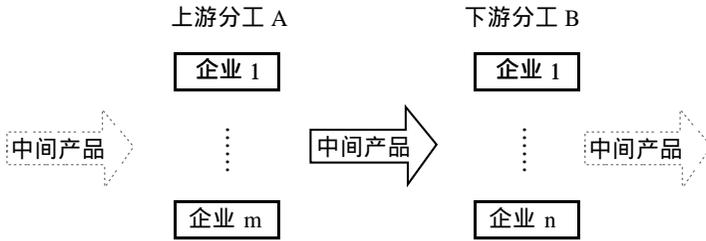


图 1

先分析下游分工部门的价格决策。部门 B 的逆需求函数为 $p_B = f(Q)$ ，其中 Q 为总产出。假定同一分工部门中所有的企业都相同，那么有 $Q = nq_B$ ，其中 q_B 表示部门 B 的企业产出。部门 B 的企业把投入的中间产品 A 的价格 p_A 作为参数，B 的一个企业的利润函数为：

$$\pi_B = (p_B - p_A - c_B)q_B - M_B(K_B) \quad (4)$$

式中： c_B 是部门 B 投入的单位产品的变动成本，假定它保持不变； $M_B(K_B)$ 是管理费用，按照前文的简化处理，它不受产量变化影响。由上式，利润最大化条件为：

$$\frac{d\pi_B}{dq_B} = p_B + q_B \frac{dp_B}{dq_B} - p_A - c_B = 0 \quad (5)$$

假定同一部门中的企业的决策是相互独立的，那么满足推测变化为零的条件，即成立 $dq_B = dQ$ ^①。利用上式和 $Q = nq_B$ ，可以得到：

$$p_B + \frac{Q dp_B}{n dQ} = p_A + c_B \quad (6)$$

因 dp_B/dQ 的值为负，表示定价总是在边际成本 $p_A + c_B$ 之上增加一个利润边际 $-\frac{Q dp_B}{n dQ}$ ，这是加价因素。同一部门中企业的数目越多，利润边际就越小，这是竞争因素。

因为产出的度量 Q 同时适用于 A 和 B，所以由 (6) 式还可以得出对上游部门 A 的产品引致需求（逆需求函数）：

① 如果同一分工部门中企业的决策是相互依赖的，那么 B 部门中的企业 i 选择 q_i 的利润最大化条件为：

$$d\pi_i dq_i = p_B(Q) + q_i \frac{dp_B}{dQ} \frac{dQ}{dq_i} - p_A - c_B = 0$$

其中： $\frac{dQ}{dq_i} = 1 + \sum_{j \neq i} \frac{dq_j}{dq_i}$ ，右边第二部分为推测变化，体现企业 i 对于所有其他企业 j 对 i 产量增加会作何反应的估计。推测变化等于零意味着 $dq/dq_i = 1$ ，即 $dq_i = dQ$ 成立，或对应文中的 $dq_B = dQ$ 。

$$p_A(Q) = p_B + \frac{Q}{n} \frac{dp_B}{dQ} - c_B \quad (7)$$

上式右边的前两项就是部门 B 中单个企业的边际收益 $MR_B(n, Q)$ ，利用边际收益符号，部门 A 的引致需求可以简写为：

$$p_A(Q) = MR_B(n, Q) - c_B \quad (8)$$

可以看出，上游部门企业的需求等于下游部门企业的边际收益减去边际成本。

用类似的分析，可以得出部门 A 中一个企业的利润函数（忽略上游中间产品的投入）：

$$\pi_A = (p_A - c_A)q_A - M_A(K_A) \quad (9)$$

式中： c_A 和 $M_A(K_A)$ 分别是部门 A 的企业的边际成本和管理费用。将 (8) 式代入 (9) 式，利用 $mq_A = Q$ 和零推测条件 $dq_A = dQ$ ，得到利润最大化的一阶条件：

$$MR_B(n, Q) + \frac{Q}{m} \frac{d}{dQ} MR_B(n, Q) = c_A + c_B \quad (10)$$

上式的右边是边际收益，准确地说，是下游部门 B 的企业的“边际收益的边际收益”^①；左边是部门 A 和 B 的边际成本之和。这个结论可以推广到整个纵向分工链，最终产品的定价高低，如同一个虚拟的企业在以下条件下定价：边际成本等于各个分工环节的边际成本之和，需求则由最终产品的需求开始，每越过一个市场交易环节，下游企业的边际收益曲线构成上游企业的需求曲线，直至最上游的分工部门为止。

上述结论可以用一个特例来说明。假定最终产品的逆需求函数为 $p = 1 - Q$ ，生产过程由 K 个纵向分工环节构成，并分别由 K 个部门生产。设第 i 个分工环节有 n_i 个相同的在决策上独立的生产者，序号从上游部门开始计算，最终产品部门就是部门 K ，因此满足 $p = p_K$ 。部门 K 中每个生产者的边际收益为：

$$MR_K(n_K, Q) = p_K + q_K \frac{dp_K}{dq_K} = p_K + \frac{Q}{n_K} \frac{dp_K}{dQ} = p_K - \frac{Q}{n_K} = 1 - \frac{n_K + 1}{n_K} Q \quad (11)$$

上式利用了零推测变化条件 $dq_K = dQ$ 。上式构成 $K-1$ 部门的逆需求函数，并可得到 $K-1$ 部门企业的边际收益。将此关系依次用于 K 个分工环节，可以得到由最终产品需求派生到部门 1 的边际收益：

$$1 - \frac{\prod_{i=1}^K (n_i + 1)}{\prod_{i=1}^K n_i} Q \quad (12)$$

用 c 表示 K 个分工部门的边际成本之和，即单位最终产品的总边际成本。由边际收益等于边际成本条件和 $p_K = 1 - Q$ ，可以得到最终产品的定价：

$$p_K = 1 - \frac{\prod_{i=1}^K n_i}{\prod_{i=1}^K (n_i + 1)} (1 - c) \quad (13)$$

① 参见多纳德·海和德里克·莫瑞斯：《产业经济学与组织》（上册），经济科学出版社 2001 年版，第 274 ~ 280 页。

不难得到利用企业来组织同一分工的定价。当 K 个分工在一个总的企业内时，企业的利润为：

$$\pi = (p - c)Q - M(K) \quad (14)$$

最终产品的最优定价为：

$$p = 1 - \frac{1}{2}(1 - c) \quad (15)$$

由此，可以得到利用市场组织分工的定价低于利用企业组织分工的定价的条件为：

$$\frac{\prod_{i=1}^K n_i}{\prod_{i=1}^K (n_i + 1)} > \frac{1}{2} \quad (16)$$

四、市场分工模式和企业分工模式的平均成本

以上的定价遵循边际原则，实际中的竞争效率还要考虑平均成本因素。我们来分析市场分工模式的平均成本。假定每个分工环节都构成一个独立生产部门，并假定市场交易费用参数 a 都相同，利用 (3) 式，将全部分工部门的企业的交易费用加总，得到完全利用市场组织分工的模式的总交易费用为：

$$C_M(a; n_1, \dots, n_K) = a \times \left(\frac{1}{n_1} + \sum_{i=2}^{K-1} \frac{2}{n_i} + \frac{1}{n_K} \right) \quad (17)$$

式中右边括号中的第 1、3 两项的系数为 1，表示第 1 个和第 K 个分工部门只有买或卖一种交易，而其余的分工部门则有买和卖两种交易。显然，交易费用 C_M 随着分工种类数 K 的增大而增大，随着系统中的企业数的增多而减小。

假定只有一个分工环节的企业可以由一个生产者构成，这个个体生产有着最强的激励，无需监督和业绩考核费用，那么，可以认为管理费用 $M(1, q) = 0$ 。由 (17) 式得到，完全利用市场来组织分工生产的单位最终产品的平均成本为 $c + C_M/Q$ 。由 (2) 式得到，完全利用企业来组织分工生产的最终产品的平均成本为 $c + M(K)/Q$ 。利用 $p = 1 - Q$ 替换产出 Q ，并利用 (13) 和 (15) 式消去价格变量，可以将平均成本表示为市场结构因素的函数。

五、市场分工模式和企业分工模式的演化稳定性

市场需求的大小是决定市场分工模式还是企业分工模式具有优势的一个重要因素。考虑两种极端的组织分工的情形：一种是由一个企业包容所有的 K 个分工环节，不妨称之为总企业模式；另一种是每个分工环节都分别构成一个独立的部门，每一部门都由若干个小企业构成，不妨称之为小企业模式。如果最终产品的需求不大，每个分工环节允许的独立生产者的数量不多，那么小企业模式的边际定价和平均成本都会太高，总企业模式是唯一可行的。如果最终产品的需求很大，允许的分工参与者足够多，那么根据 (13) 式和 (17) 式，小企业模式的边际定价和平均成本会足够接近边际成本，小企业模式具有竞争

的绝对优势。

不过，即使小企业模式的边际定价和平均成本都低于或高于总企业模式，何种组织分工的模式能成为实际中的模式，也是不确定的，原因在于两种生产模式都具有状态锁定性，或称之为演化稳定均衡特征，即一旦一种模式占了上风，它的优势就会得到强化，从而遏制了另一种模式的发展。我们来说明这个形成机制。

比方说，一种产品的生产已经形成了总企业模式，要将企业内的某个分工环节交由市场来进行，将面临两个困难：其一，总企业与分离出来的中间产品生产企业之间的交易是双边垄断的，交易费用很高，而且总企业一方更容易受到要挟。其二，容易证明，中间产品分离之后的生产是一个垄断链，定价水平高于单一垄断企业，因而缺乏整体的价格竞争优势。一个可能的解决办法是总企业同时与若干个独立的生产同一中间产品的小企业交易，利用市场竞争降低交易费用。这个办法在理论上可以成立，然而在实际上因为独立的生产者总是逐个产生的，几乎不可能同时出现一个中间产品生产群以供交易。所以，一旦形成总企业生产模式，这种模式就具有稳定性。这时，产量的扩大会导致平均成本降低，经济整体竞争力的提高是与企业的规模扩大相关联的。

另一方面，如果小企业模式能够存在，它的竞争优势同样会越来越强，生产方式也将锁定于小企业模式。其原因在于，小企业模式是一个开放的系统，它对于新的进入者无法构成障碍，一方面因为对进入的资金规模要求不高，没有资金障碍；另一方面因为生产环节单一，技术要求也不高，没有技术障碍。每进入一个新的生产者，利用(13)式容易证明最终产品的定价就会低一点，利用(17)式同样容易证明交易费用也同时减小，平均成本也会降低一点，整体的竞争力也随之提高一点。当生产方式锁定于小企业模式后，是否会因为同一中间产品的生产者联合起来，构成垄断呢？理论上是不会的，因为联合是为了提价，而一旦提价，将诱使更多的生产该中间产品的新生产者进入，新进入者只要按原来的价格定价就足以使联合者失去价格优势^①。因此，一旦形成小企业生产模式，这种模式也具有稳定性。这时，经济整体竞争力的提高是与较小的企业平均规模联系在一起的。

那么何种分工模式会成为现实中的模式呢？如同演化博弈理论预示的那样，初始状态具有偶然性和不确定性。不过，不同分工模式的初始出现概率是不同的，如果交易费用参数 a 较大，小企业模式出现的可能性较小，甚至不可能出现。如果存在一种方法可以普遍降低交易费用参数 a ，那么小企业模式出现的可能性就会大幅提高，成为实际中的生产方式。我国许多中小城镇的产业集聚生产方式，就是通过相关企业在一地理区域的集中来降低搜寻交易对象和议价的成本，从而满足企业外组织分工条件的典型例子。

由以上的分析可以得出以下三点主要结论：第一，我们一般不能断定一个区域的生产或一种产品的生产会锁定于总企业模式还是小企业模式，因此，不能简单地在经济整体竞争力与企业平均规模之间建立确定的对应关系。第二，我们甚至不能通过“超边际分析方

^① 管理费用与产量无关的假定导致企业生产具有全局规模经济，随之而来的是自然垄断性，预示一种产品只能由一个企业生产。本文通过个体生产者没有管理费用的假定回避了这一结果的出现。实际中只要有任何形式的固定成本存在，并假定新增成本随产量的增加而递增，就将失去全局规模经济性，有最小平均成本存在，成本函数将与实际情形更为符合。可以证明，在一定的需求范围内，即使存在最小平均成本，分工状态的演化稳定特征也仍然存在。就本文的目的而言，文中的简化假定已经够了。

法”比较何种模式具有更高的竞争力,从而认为它就是实际中的生产模式。经济增长和生产模式的形成是一个演化过程,具有路径依赖性,一般会表现出演化稳定均衡的特征。第三,科斯的交易费用大小影响企业平均规模的观点只具有相对合理性。如果交易费用的大小不受分工是在企业内组织还是在企业外组织的影响,那么科斯的结论是正确的;而如果交易费用本身就是分工组织方式的函数,即如本文所描述的那样,那么科斯的结论就有局限性。

参考文献

- [1] Dixit, A. and Stiglitz, J., Monopolistic Competition and Optimum Product Diversity, American Economic Review, 1977, Vol.67, pp297 ~ 308.
- [2] Krugman, Paul, Increasing Returns, Monopolistic Competition, and International Trade, Journal of International Economics, 1979, Vol.9, pp469 ~ 479
- [3] Young, Allyn, Increasing Returns and Economic Progress, The Economic Journal, 1928, Vol.38, pp527 ~ 542.
- [4] 多纳德·海, 德理克·莫瑞斯:《产业经济学与组织》, 经济科学出版社 2001 年版。
- [5] 科斯:《企业的性质》, 见:《企业、市场与法律》, 上海三联书店 1990 年版。
- [6] 马歇尔:《经济学原理》(上卷), 商务印书馆 1994 年版。
- [7] 斯蒂格勒:《市场容量限制劳动分工》, 见:《产业组织和政府管制》, 三联书店 1989 年版。
- [8] 杨小凯, 黄有光:《专业化与经济组织——一种新兴古典微观经济学框架》, 经济科学出版社 1999 年版。
- [9] 盛昭瀚, 蒋德鹏:《演化经济学》, 上海三联书店 2002 年版。
- [10] 张永生:《厂商规模无关论——理论与经验证据》, 中国人民大学出版社 2003 年版。

企业利益相关者三维分类的实证分析*

陈宏辉 贾生华

现代经济学和管理学界已经达成的一个共识是：企业是一种人格化的组织，是其利益相关者之间综合性社会契约的汇集点（多纳德逊、邓非，2001）。事实上，现代企业理论的发展已经把企业理解为不同个人和群体之间一组复杂的显性契约和隐性契约交汇所构成的法律实体（Jensen & Meckling, 1976）。在这种法律实体中，交汇的契约既有经营者与所有者之间的契约、经营者与雇员之间的契约，还有企业作为债权人与债务人之间的契约、企业作为供应商（或消费者）与消费者（供应商）之间的契约、企业作为法人与政府之间的契约等等。这些被称作“利益相关者”的个人和群体不仅会影响企业目标的达成，而且还会受到企业目标达成过程中所采取的各类行动的影响（Freeman, 1984）。如何有效地识别和应对这些利益相关者的利益要求已经成为现代公司治理理论和实践的前沿课题。主流企业理论诟病利益相关者理论的一个主要理由就是“企业的利益相关者难以界定和分类”，因为倘若“连谁是企业的利益相关者都弄不清的话，那么利益相关者理论也就无从谈起了”（Donaldson & Preston, 1995；Shankman, 1999）。事实上，在过去的40年里，西方经济学界对于企业利益相关者的界定和分类进行了广泛的探索，并完成了从概念思辨到实证研究的转变。只是我国诸多学者仍将目光停留在主流企业理论的框架之中，忽视了利益相关者理论在西方学术界的快速发展，现有的一些中文文献更多地是关注利益相关者理论本身的立意，而对于像“我国企业的利益相关者分类”、“我国企业利益相关者利益要求的内容、实现方式、实现程度”等基础性问题的实证研究还尚未涉猎。本文借鉴西方学者对企业利益相关者的分类方法，着眼于如何对我国企业的利益相关者进行科学分类，以此作为探讨我国企业合理安排公司治理架构的基础。

一、文献回顾

米切尔和伍德（Mitchell & Wood, 1997）曾经详细研究了利益相关者理论产生和发展的历史，总结了自1963年开始到20世纪90年代中期为止前后共30多年时间里西方学者所给出的27种有代表性的利益相关者定义。总的来看，西方学者关于企业利益相关者的界定可谓形式多样，“没有一个定义得到普遍的赞同”（多纳德逊、邓非，2001）。但是如果仔细剖析近30年来利益相关者理论的发展史，还是可以发现西方学者对利益相关者认识的主流趋势，即倾向于将利益相关者界定为那些与企业有一定的关系，并在企业中进行了一定的专用性投资的人（Clarkson, 1994；Carroll, 1994；Starik, 1994）。一个被越来越

* 原载：《经济研究》，2004年第4期。

多的学者所认同的企业利益相关者的定义是：利益相关者是指那些在企业中进行了一定的专用性投资，并承担了一定的风险的个体和群体，其活动能够影响该企业目标的实现，或者受到该企业实现其目标过程的影响（陈宏辉，2003）。

然而，界定出了企业的利益相关者，并不意味着就完全把握了利益相关者的特性，因为“列举出一个大公司的每一个可能有资格作为利益相关者的人，造成的结果往往是把具有极不相同的要求和目标的相互交接的群体混在一起”（多纳德逊、邓非，2001）。换句话说，虽然企业的生存和发展离不开所有的利益相关者的支持，但利益相关者作为一个群体还可以从多个角度进行分类，细分出的不同类型的利益相关者对于企业管理决策的影响以及受到企业活动影响的程度是不一样的。对企业而言，“所有的利益相关者都可能是十分重要的，但没有人说过他们是同等重要的”（沃克、马尔，2003）。梳理利益相关者理论文献，我们可以发现到20世纪80年代末期以后，西方学者就已经普遍认识到仅仅界定出企业的利益相关者是不够的，由此兴起了对企业的众多利益相关者进行分类的高潮。其中“多维细分法”的思想和“米切尔评分法”是最为引人瞩目的成果。

所谓“多维细分法”，就是从多个维度对“企业利益相关者”这一群体进行分类，以期找出不同利益相关者在某些特征上的差异。在20世纪80年代初期至90年代中期，“多维细分法”逐渐成为利益相关者分类中最常用的分析工具。其中有代表性的分析结果有以下一些：弗里曼（Freeman，1984）从所有权、经济依赖性^①和社会利益三个不同的维度对利益相关者进行了开创性的分类。弗雷德里克（Frederick，1988）则认为利益相关者是“对企业的政策和方针能够施加影响的所有集团”。他将利益相关者分成了直接利益相关者和间接利益相关者。查克汉姆（Charkham，1992）按照相关群体与企业是否存在交易性的合同关系，将利益相关者分为契约型利益相关者（contractual stakeholders）和公众型利益相关者（community stakeholders）。克拉克森（Clarkson，1994；1995）提出了两种有代表性的分类方法：①根据相关群体在企业经营活动中承担的风险的种类，可以将利益相关者分为自愿的利益相关者（voluntary stakeholders）和非自愿的利益相关者（involuntary stakeholders）。②根据相关群体与企业联系的紧密性，可以将利益相关者分为首要的利益相关者和次要的利益相关者。英国学者威勒（Wheeler，1998）则将社会性维度引入到利益相关者的分类中，并产生了深远的影响。结合克拉克森提出的紧密性维度，威勒将所有的利益相关者分为以下四种：首要的社会性利益相关者、次要的社会性利益相关者、首要的非社会性利益相关者、次要的非社会性利益相关者。

上述从多个维度来细分利益相关者的思路大大深化了人们对企业的利益相关者的认识。然而这些分类方法普遍都停留在学院式的研究中，缺乏可操作性，从而制约了利益相关者理论的实际运用（贾生华、陈宏辉，2002）。20世纪90年代后期，美国学者米切尔和伍德（Mitchell & Wood，1997）提出了一种评分法（score-based approach）来对利益相关者进行分类，思路清晰，简单易行。米切尔等人从三个属性上对可能的利益相关者进行评分，然后根据分值的高低来确定某一个体或者群体是不是企业的利益相关者、是哪一类型的利益相关者。这三个属性是：①合法性（legitimacy），即某一群体是否被赋有法律上的、道义上的或者特定的对于企业的索取权；②权力性（power），即某一群体是否拥有影响企业决策的地位、能力和相应的手段；③紧急性（urgency），即某一群体的要求能否立即引

起企业管理层的关注。米切尔评分法的提出大大改善了利益相关者分类的可操作性，极大地推动了利益相关者理论的应用，并逐步成为利益相关者分类的最常用的方法。

国内学者对利益相关者理论的关注始于 20 世纪 90 年代中期以后，并且迅速认识到对利益相关者进行界定和分类是“一个很关键的问题”（杨瑞龙、周业安，2001）。万建华、李心合分别从企业利益相关者合作性和威胁性两个维度对利益相关者进行了分类，细分出四种类型的利益相关者：支持型的利益相关者、边缘型的利益相关者、不支持型的利益相关者以及混合型的利益相关者（万建华，1998；李心合，2001）。应该说，这种分类方法是直接脱胎于西方学者 20 世纪 90 年代的研究成果，具有典型的代表意义^①。

但是笔者对于诸多学者仅仅停留在从多个维度上来区分众多利益相关者的研究思路是不赞同的，尽管其分类结果对于我们详察不同的利益相关者的特性大有裨益。换句话说，倘若仅仅提出两个或三个维度来对企业利益相关者进行想象式的分类，而不是依靠取自于实际中的数据进行验证的话，我们永远也无法深入开展利益相关者理论的研究。例如，我们完全可以接受西方学者的直接利益相关者、间接利益相关者、首要利益相关者和次要利益相关者等分类概念，但在我国企业中究竟哪些利益相关者是直接利益相关者？哪些又是间接利益相关者？哪些是首要利益相关者，哪些又是次要利益相关者呢？在不同性质的企业中，在不同规模的企业中，在不同发展阶段的企业中，同一种利益相关者（如员工或供应商）在某一维度上的特性会一样吗？在这一点上，我们也可以非常清楚地看出中外学者研究方法上的差异：西方学者在详细讨论了如何确定利益相关者的分类维度后，迅速转入将这些分类维度应用于企业实践的实证研究之中；我国学者则习惯于构造出多种多样的分类框架，但止步于深入的实证分析，也就难以得出令人信服的、有实践意义的结论了。笔者认为，对于利益相关者的分类研究，如果是针对某一具体企业或产业而言的话，我们完全可以借鉴米切尔的分类方法对众多的利益相关者进行详细地评分，进而将他们分为不同的类型，采取不同的管理策略；如果是从一般意义上对企业的利益相关者进行分类的话，则必须依赖于实证调查的资料，最好是让众多企业界的人士来进行选择分类，进而得出具有统计意义的分类结果。本文即尝试利用实际调研资料，从一般意义上对我国企业的利益相关者进行分类，并探讨不同的个体因素和企业因素对分类结果的影响。

二、研究假设

借鉴西方学者的“多维细分法”和“米切尔评分法”的分析思路，笔者拟从利益相关者的主动性、利益相关者的重要性和利益相关者利益要求的紧急性三个维度上对所界定出

^① 萨维奇等人曾经根据利益相关者对企业的威胁或合作的潜在可能性，把利益相关者区分为支持的（合作的可能性高）、边缘的（两种可能性都低）、不支持的（威胁的可能性高），或者威胁与合作兼有的（两种可能性都高）四种类型的利益相关者（Savage et al., 1991）。

的 10 种利益相关者进行分类^①。首先，我们认为在企业的利益相关者中，有的利益相关者会对企业主动施加影响，从而也往往主动地承担着企业经营的风险；而另外一些利益相关者则是被动地受到企业经营行为的影响，从而被动地承担企业经营风险。也就是说，不同的利益相关者影响企业经营行为的主动性存在差异（Frederick, 1988；Clarkson, 1994）。其次，我们认为不同的利益相关者对于企业生存和发展的重要性是有差异的。在我们所界定的十类利益相关者中，对于某一特定的企业而言，有的利益相关者是绝对不可或缺的，有的则可能影响不大（Freeman, 1984；Clarkson, 1995）。再次，在某一特定的时点上，不同的利益相关者的利益要求的紧急性会存在差异。有的利益相关者在一般情况下可能既不主动，也不重要，但是在某一状态下其利益要求必须很快得到满足，否则就会影响到企业的正常运营（Mitchell & Wood, 1997）。根据上述分析，我们可以提出假设 1：

假设 1：企业中的众多利益相关者在多个维度上具有特征差异。

如果企业中的众多利益相关者的确在多个维度上存在特征差异的话，是否就意味着对这些利益相关者难以进行分类呢？笔者认为这一判断是不成立的，因为虽然存在差异，但应该可以寻找出一些统计规律。首先，在任何一个企业中，必然离不开股东、管理人员和员工这三类人员^②，他们作为企业经营运作的直接参与者，其利害关系必然与企业紧密相关（Freeman, 1984）。无论如何，他们都应该被视作企业的核心利益相关者。其次，如果利用某些综合性指标来进行排序的话，有些利益相关者的顺序将会明显偏后，从而游离于企业利益相关者的边缘地带。例如，虽然特殊利益团体和社区这两类群体在西方学者的眼中几乎都肯定是企业的利益相关者（Charkham, 1992），我国企业也越来越重视清洁生产、环境管理，并有意识地加强与社区的沟通和交流，但我们推测在我国现实情况下，他们恐怕还只能是边缘型利益相关者。据此，我们提出假设 2 和假设 3：

假设 2：作为企业经营运作的直接参与者，股东、管理人员和员工是企业的核心利益相关者。

假设 3：在我国当前企业中，特殊利益团体和社区还很难被认为是利益相关者。

更进一步的问题是，倘若对利益相关者进行了分类，譬如说分成了四类利益相关者（甲类、乙类、丙类、丁类），那么不同的人对于将某一利益相关者归入甲类或乙类会有不同的看法吗？例如，企业高层管理人员可能认为政府是企业中的非常重要的利益相关者，而普通员工却可能并不这样认为，他们也许认为消费者才是不可或缺的。类似的问题是，这种分类结果在不同类型的企业中是否会有差别呢？我们可以想象，一个初创的、小型的

① 本研究的前期工作是根据利益相关者的定义，利用专家评分法对我国企业的利益相关者进行了界定。在界定过程中约请 24 位来自企业界的经理人员，向他们每个人提供了一份共包含 20 种利益相关者在内的名单，分两轮请他们选出认为符合这一定义的利益相关者，数量不限。每一种利益相关者入选率的统计结果如下：股东 100%，管理人员 100%，员工 100%，消费者 83.33%，债权人 83.33%，政府 79.17%，供应商 66.67%，分销商 62.50%，特殊利益团体 58.33%，社区 54.17%，行业协会 25.00%，媒体 12.50%，公众 8.33%，政治团体 8.33%，专业投资机构、教育机构、宗教团体、竞争对手都是 4.16%，人类下一代和非人物种入选率为 0.00%。以入选率 50% 作为标准，笔者将股东、管理人员、员工、消费者、债权人、政府、供应商、分销商、特殊利益团体和社区这十个群体作为我国企业的利益相关者开展系列研究。

② 当然，在某些企业中，股东、管理人员和员工的身份可能重合，例如在古典企业中股东可能同时又是管理人员，甚至也还是普通员工。但这并不影响本文的分析结论。

民营企业与一个处于成熟期的、大型的国有企业相比，他们对于供应商、消费者、社区等各种利益相关者的认识可能会有很大的差异。因此，我们可以提出假设 4：

假设 4：不同的人对企业利益相关者的认识是有差异的；在不同类型的企业中，人们对利益相关者的认识也是有差异的。

三、数据来源与处理方法

本研究主要通过实地访谈和问卷调查两种形式取得数据。从 2002 年 6 月开始，笔者与杭州市、宁波市、慈溪市、舟山市、深圳市、广州市等 6 个城市的 22 家企业的 67 名人员进行了深入访谈。访谈的目的主要是为了了解该企业主要利益相关者的构成、不同利益相关者的特性及其变迁情况。在进行了 45 份的试调查以后，笔者从 2002 年 6 月底开始进行正式的问卷调查，截至 2003 年 1 月初完成，共历时 6 个多月。问卷内容主要包括五个方面：填写者的基本情况、所在企业的基本情况，分别从主动性、重要性和紧急性三个维度对给出的 10 种利益相关者进行排序。问卷调查范围包括浙江省、广东省、福建省、江苏省、湖北省、陕西省、海南省、北京市、上海市等 9 个省市，主要以东部和南部为主，但也涵盖了全国东南西北中的地理位置。共计发放问卷 795 份，实际回收 462 份，回收率 58.1%，回收问卷中有效问卷数 423 份，回收问卷有效率 91.6%。

样本的基本分布情况如下：从性别来看，男性 319 份，占 75.4%，女性 104 份，占 24.6%；从年龄来看，20~30 岁 143 份，31~40 岁 182 份，41~50 岁 56 份，51 岁以上 42 份，分别占样本总数的 33.8%，43.0%，13.2% 和 9.9%；从职务来看，一般员工 82 份，基层管理人员 132 份，中层管理人员 139 份，高层管理人员 70 份，分别占样本总数的 19.4%，31.2%，32.9% 和 16.5%；从所调查的企业性质来看，国有企业 202 份，集体企业 46 份，民营企业 119 份，“三资”企业 56 份，分别占样本总数的 47.8%，10.9%，28.1% 和 13.2%；从企业规模来看，大型企业 160 份，中型企业 183 份，小型企业 80 份，分别占样本总数的 37.8%，43.3% 和 18.9%；从企业上市情况来看，103 家企业为上市公司，320 家企业没有上市，分别占样本总数的 24.3%，75.7%；从企业发展阶段来看，创业阶段企业 50 份，发展阶段企业 159 份，成熟阶段企业 134 份，衰退阶段企业 80 份，分别占样本总数的 11.8%，37.6%，31.7% 和 18.9%。从样本的分布来看，最少的分类变量数据也有 40 个，数据量已经满足了进行统计分析的基本要求。

四、数据分析

数据分析主要分为以下三个步骤：

1. 从三个维度上对利益相关者进行评分

在调查问卷的第二部分中，首先要求问卷填写者结合本企业的实际情况，对于给出的 10 种利益相关者的主动性程度按照从高到低进行排序。将排序转化为数值型数据以后，利用 SPSS10.0 进行描述性统计，结果如表 1 所示。

表 1 十类利益相关者在主动性维度上评分的描述性统计

	有效样本 (N)	最小值 (Min)	最大值 (Max)	均值 (Mean)	标准差 (Std D.)
股东主动性	410	1	10	2.95	2.67
管理人员主动性	410	1	10	3.47	2.04
供应商主动性	410	1	10	4.71	2.46
员工主动性	410	1	10	5.03	2.45
债权人主动性	410	1	10	5.13	2.49
分销商主动性	410	1	10	5.37	2.15
政府主动性	410	1	10	5.67	2.73
消费者主动性	410	1	10	6.57	2.58
特殊团体主动性	410	2	10	7.86	2.04
社区主动性	410	1	10	8.25	2.06

注：填写人若对某一利益相关者的主动性评分越小，则表明他认为该利益相关者越主动。以下类似。

显然，不能简单地根据表 1 中均值大小的排列来判断某一利益相关者就一定比另一利益相关者更主动，因为这没有统计意义（马庆国，2002）。我们利用“配对样本 T 检验”（paired-samples T test），进一步判断上述每两个变量均值之差与 0 是否具有显著性差异。统计结果如表 2 所示。表 2 中数据的含义是：未加括号的数据表示某一利益相关者在主动性维度上评分的均值与另一利益相关者在主动性维度上评分的均值之差，括号内的数据是 T 检验值。如果均值之差通过了 95% 或 99% 置信度的检验，则标上 * 号或 ** 号；如果没有通过检验，则在均值之差的数据下方划有横线^①。综合分析表 2 的统计结果，我们可以发现：从主动性维度上来看，供应商与员工、员工与债权人、员工与分销商、债权人与分销商、分销商与政府虽然评分的均值不同，但是这种均值的差异与 0 没有统计意义上的显著性差别，其他的排序都具有显著的或非常显著的统计意义上的差别。

表 2 主动性维度评分均值差异的配对样本 T 检验结果

	1	2	3	4	5	6	7	8	9
1 股东主动性									
2 管理人员主动性	0.52** (3.259)								
3 供应商主动性	1.76** (8.670)	1.24** (7.197)							
4 员工主动性	2.08** (11.807)	1.56** (13.915)	<u>0.32</u> (1.719)						
5 债权人主动性	2.17** (12.438)	1.66** (9.315)	0.42* (2.355)	0.09 (0.524)					

① 在本文以下所做的配对样本 T 检验的结果中，数据含义都与此类似，不再赘述。

续上表

	1	2	3	4	5	6	7	8	9
6 分销商主动性	2.41** (13.585)	1.90** (11.790)	0.66** (4.778)	0.33 (1.809)	0.24 (1.372)				
7 政府主动性	2.72** (12.527)	2.20** (12.011)	0.96** (4.842)	0.64** (3.014)	0.54** (2.750)	0.30 (1.809)			
8 消费者主动性	3.61** (19.498)	3.10** (18.240)	1.86** (9.452)	1.53** (8.727)	1.44** (7.448)	1.20** (7.113)	0.90** (4.526)		
9 特殊团体主动性	4.91** (25.353)	4.39** (27.488)	3.15** (18.926)	2.83** (15.281)	2.73** (16.054)	2.50** (16.343)	2.19** (16.623)	1.29** (7.537)	
10 社区主动性	5.30** (28.654)	4.78** (28.932)	3.54** (22.896)	3.22** (18.266)	3.12** (18.232)	2.89** (19.009)	2.58** (16.763)	1.68** (9.176)	0.39* (3.137)

注：* $p < 0.05$ ，** $p < 0.01$ 。下同。

按照同样的方法，可以根据 10 种利益相关者在重要性和紧急性维度上的评分进行排序，并进行配对样本 T 检验。限于篇幅，数据从略。根据统计结果，我们实际上已经明确了这 10 种利益相关者在这三个维度上的排序情况。在我们的统计中，排序的最大分值为 10，我们将 10 分划分为 1~4 分、4~6 分、6~10 分三段，根据这 10 种利益相关者在各个维度上得分的均值将其填入相应的单元格中，形成表 3。在表 3 中有三种利益相关者（用上标 a、b、c 标出）的归类值得特别说明。从紧急性维度来评价股东，其得分的均值为 4.08，已经大于 4 分，似乎应该归入紧急性维度的“4~6 分”单元格中。但是配对样本 T 检验结果表明，股东在紧急性维度上的得分均值 4.08 与管理人员在紧急性维度上的得分均值 3.74 并没有统计意义上的显著差异，而应该将他们归入一类之中。同样的情况还发生在从重要性维度来评价债权人的得分和从紧急性维度来评价政府的得分上。

表 3 十种利益相关者的三维分类结果

评分 维度	[1, 4)	[4, 6)	[6, 10]
主动性	股东、管理人员	供应商、员工、债权人、分销商、政府	消费者、特殊团体、社区
重要性	管理人员、员工	消费者、股东、政府、分销商、供应商、债权人 ^b	特殊团体、社区
紧急性	员工、管理人员、股东 ^a	消费者、分销商、债权人、供应商、政府 ^c	特殊团体、社区

根据表 3 中各个利益相关者的位置，我们可以对这 10 种利益相关者进行分类，结果如下：

核心利益相关者：至少在两个维度上得分在 4 分以下，他们是企业不可或缺的群体，与企业具有紧密的利害关系，甚至可以直接左右企业的生存和发展。在我们的统计结果中

包括管理人员、员工、股东三类利益相关者。

蛰伏利益相关者：至少在两个维度上得分在 4 分以上、6 分以下，他们往往已经与企业形成了较为密切的关系，所付出的专用性投资实际上使得他们承担着企业一定的经营风险。在企业正常经营状态下，他们也许只是表现为一种企业的显性契约人而已；然而一旦其利益要求没有得到很好的满足或是受到损害时，他们可能就会从蛰伏状态跃升为活跃状态，其反应可能会非常强烈，从而直接影响企业的生存和发展。在我们的统计结果中，这些利益相关者包括供应商、消费者、债权人、分销商和政府等五类。

边缘利益相关者：至少在两个维度上得分在 6 分以上，他们往往被动地受到企业的影响，在企业看来他们的重要性程度很低，其实现利益要求的紧迫性也不强。在我们的统计结果中包括特殊团体、社区两类利益相关者。

至此，我们也就部分地验证了假设 1，即这十类利益相关者在主动性、重要性和紧急性三个维度上表现出不同程度的特征差异，而且大部分的差异都十分显著，但也有一些差异并不显著。

2. 计算综合得分，进行进一步的验证

为了进一步地验证以上分类结果，并从总体上进行比较，我们可以计算出每一个利益相关者在主动性、重要性和紧急性三个维度上的综合得分，称之为综合相关度。记第 i ($i = 1, 2, \dots, 10$) 个利益相关者的综合相关度为 PX_i ，该利益相关者在第 j ($j = 1, 2, 3$) 个维度上的得分为 V_{ij} ， PX_i 的计算公式为：

$$PX_i = \frac{1}{3} \sum_{j=1}^3 V_{ij}$$

由此生成 10 个新的变量，其描述性统计结果如表 4 所示。

表 4 利益相关者综合相关度的描述性统计

	有效样本 (N)	最小值 (Min)	最大值 (Max)	均值 (Mean)	标准差 (Std D.)
管理人员综合相关度	402	1.00	9.00	3.5050	1.5425
股东综合相关度	403	1.00	10.00	3.6716	2.0941
员工综合相关度	403	1.00	10.00	4.0639	1.6956
消费者综合相关度	400	1.00	10.00	4.9717	2.0193
分销商综合相关度	399	1.00	10.00	5.3826	1.6105
政府综合相关度	402	1.00	10.00	5.4154	1.9717
供应商综合相关度	398	1.00	10.00	5.4765	1.6731
债权人综合相关度	398	1.33	10.00	5.6633	1.8668
特殊团体综合相关度	397	2.00	10.00	7.9748	1.6480
社区综合相关度	397	1.33	10.00	8.4971	1.5528

同样，我们需要对这 10 个新的变量进行配对样本 T 检验，以判断上述排序是否具有显著性统计意义。结果表明，管理人员与股东、分销商与政府、分销商与供应商、政府与供应商、政府与债权人、供应商与债权人虽然综合相关度的均值不同，但是这种均值的差

异与 0 没有统计意义上的显著性区别。该表的统计信息同样表明, 管理人员、股东、员工可以归为一类 (综合相关度均值在统计意义上的 4 分以下), 消费者、分销商、政府、供应商和债权人可以归为一类 (综合相关度均值在统计意义上的 4 分以上、6 分以下), 而特殊利益团体和社区可以归为一类 (综合相关度均值在统计意义上的 6 分以上)。这一结果完全印证了我们将这十类利益相关者分为核心利益相关者、蛰伏利益相关者和边缘利益相关者的结论, 从而也就验证了假设 2 和假设 3, 即股东、管理人员和员工的确是企业不可或缺的核心利益相关者, 而在我国企业中将特殊利益团体和社区作为需要考虑的利益相关者还显得很勉强。

3. 差异比较分析

为了进一步分析不同的分类因素对企业利益相关者认识的差异, 我们需要进行单因素方差分析 (ANOVA)。也就是说, 我们试图来寻找不同性别、年龄、工龄、职务的人以及不同性质、上市与否、不同规模、不同发展阶段的企业对于十类利益相关者综合相关度的排序是否存在差异, 分析的结果如表 5 所示。

表 5 利益相关者综合相关度的差异比较分析

	性别	年龄	工龄	职务	企业性质	是否上市	企业规模	发展阶段
	F 值	F 值	F 值	F 值	F 值	F 值	F 值	F 值
股东综合相关度	0.501	9.867**	5.971**	4.343**	2.000	0.404	12.593**	3.967**
管理人员综合相关度	3.514	8.310**	1.668	5.435**	2.150	6.871**	1.780	1.015
员工综合相关度	0.352	12.829	2.810*	1.509	1.869	2.891	3.787*	1.692
债权人综合相关度	0.021	2.157**	0.759	0.426	1.390	4.499*	3.483*	4.954**
供应商综合相关度	1.881	1.360	1.340	0.174	2.347	0.530	1.807	0.261
分销商综合相关度	13.487**	3.358*	1.735	1.705	4.811**	2.118	1.767	1.399
消费者综合相关度	0.009	19.104**	18.692**	5.520*	4.640**	0.001	0.522	9.367**
政府综合相关度	2.827	6.777**	2.097	3.564*	3.539**	0.008	3.487*	7.154**
特殊团体综合相关度	2.359	2.666	4.205**	1.330	1.513	1.297	6.283**	3.720**
社区综合相关度	10.466**	8.617**	8.233**	5.243**	2.372	14.752**	10.698**	1.925

可以看出, 8 个分类变量对于不同利益相关者的认识都产生了或多或少的影 响。在表 5 中, 标有星号的数据表明某一利益相关者综合相关度的均值在某一分类变量上存在显著的或非常显著的差异。例如, 不同职务的人对于股东、管理人员、消费者、政府和社区的综合相关度都存在显著的或非常显著的差异 (相应的数据右上方标有 * 或 ** 号)。然而, 根据表 5 我们还无法判断这种差异究竟是发生在一般员工与中层管理人员之间, 还是发生在基层管理人员与高层管理人员之间。为了找出具体的差别来源, 需要做详细的多重比较, 结果如表 6 所示。

表6 利益相关者综合相关度的分组均值

		股东	管理人员	员工	债权人	供应商	分销商	消费者	政府	特殊团体	社区
性别	男	3.63	3.42	4.39	5.66	5.41	5.22	4.97	5.50	8.04	8.63
	女	3.81	3.77	4.27	5.69	5.68	5.91 [△]	4.99	5.11	7.74	8.05 [△]
年龄	20~30岁	3.29 [△]	3.78 [△]	4.48	5.67	5.70	5.24	4.65 [△]	5.52 [△]	8.09	8.50 [△]
	31~40岁	3.52 [△]	3.41 [△]	4.14	5.86 [△]	5.32	5.52	4.55 [△]	5.64 [△]	8.12	8.75 [△]
	41~50岁	5.02	3.91 [△]	4.86	5.15	5.40	5.75 [△]	5.73 [△]	4.32	7.55	7.57
	51岁以上	3.78	2.35	4.24	5.48	5.48	4.83	6.76 [△]	5.60 [△]	7.58	8.72 [△]
工龄	1年以内	3.40 [△]	3.35	3.54	5.95	6.00	5.17	3.99 [△]	5.94	8.65 [△]	8.94 [△]
	1~5年	3.32 [△]	3.68	4.38 [△]	5.75	5.35	5.45	4.84 [△]	5.60	8.04 [△]	8.55
	6~10年	3.86 [△]	3.45	4.54 [△]	5.63	5.48	5.33	4.66 [△]	5.24	8.08 [△]	8.75 [△]
	11~20年	3.58 [△]	3.53	4.59 [△]	5.53	5.34	5.75	4.93 [△]	5.12	7.60	8.42
	20年以上	5.02	3.02	4.25	5.30	5.66	4.93	7.27	4.98	7.29	7.28
职务	一般员工	4.09 [△]	4.10	4.25	5.46	5.45	5.49	4.46	5.48 [△]	7.73	8.43 [△]
	基层管理人员	3.48	3.41 [△]	4.32	5.69	5.46	5.58	5.35 [△]	5.45 [△]	7.97	8.28 [△]
	中层管理人员	3.94 [△]	3.44 [△]	4.60	5.68	5.44	5.29	5.18 [△]	5.05 [△]	7.96	8.40 [△]
	高层管理人员	3.04	3.15 [△]	4.11	5.80	5.61	5.07	4.40	3.98	8.28	9.16
企业性质	国有独资企业	4.29	3.49	4.34	5.50	5.23	5.61 [△]	5.18	5.02 [△]	8.00	8.37
	国有控股企业	3.46	3.90	4.22	5.85	5.29	5.73 [△]	4.70 [△]	5.45	8.02	8.35
	国有参股企业	3.78	3.14	3.88	5.67	5.74	5.86 [△]	5.35	4.94 [△]	8.12	8.26
	集体企业	3.73	3.45	4.81	6.11	4.99	5.22	6.06	4.97	7.40	8.08
	民营企业	3.47	3.37	4.42	5.73	5.80	4.84	4.61 [△]	5.93	8.15	8.74
	“三资”企业	3.30	3.82	4.61	5.27	5.43	5.50 [△]	4.53 [△]	5.68	7.79	8.90
上市	已经上市	3.79	3.86	4.61	6.01	5.37	5.18	4.96	5.43	7.81	7.99
	没有上市	3.63	3.39 [△]	4.28	5.55 [△]	5.51	5.45	4.97	5.41	8.03	8.67 [△]
规模	大型企业	4.23 [△]	3.69	4.57	5.98	5.32	5.24	5.07	5.08	7.72 [△]	8.04
	中型企业	3.57 [△]	3.41	4.37	5.46 [△]	5.49	5.38	4.97	5.63 [△]	7.95 [△]	8.76 [△]
	小型企业	2.82 [△]	3.35	3.93 [△]	5.51 [△]	5.76	5.67	4.77	5.57	8.54 [△]	8.77 [△]
发展阶段	创业阶段	3.04	3.60	3.91	5.87	5.45	4.94	4.28	6.26 [△]	8.66	8.92
	发展阶段	3.53	3.58	3.52	5.90 [△]	5.50	5.42	5.08 [△]	5.17	7.79 [△]	8.35
	成熟阶段	3.69	3.30	4.29	5.74	5.55	5.41	4.55	5.74 [△]	8.03 [△]	8.58
	衰退阶段	4.28 [△]	3.61	4.43	4.96	5.34	5.51	5.85 [△]	4.86	7.82 [△]	8.38

表 6 中的粗体数字表示该组均值与其他某些组的均值有显著性差异，我们在与这一均值存在显著性差异的组别均值的右上角标有“△”符号，右上角没有标“△”符号的均值则与之没有显著性差异。例如，41~50 岁的人对于股东的综合相关度的得分均值为 5.02，这一均值与 20~30 岁、31~40 岁的人的判断有显著性差异，而与 51 岁以上的人对于股东综合相关度的均值则没有显著性差异。

结合表 5 和表 6，我们就可以清楚地看出 8 个分类因素对于企业利益相关者综合相关度的差异了。具体来说，可以总结如下：

(1) 从性别来看，男性较之女性将分销商排序明显靠前，而女性对于社区的认识程度明显强于男性。

(2) 从年龄来看，差异主要发生在 41 岁以上的老职工与 20~30 岁、31~40 岁的中青年职工之间。这些老职工对于管理人员、债权人、分销商、政府、社区的认识程度明显优于中青年职工，但他们对于股东和消费者的排序明显弱于中青年职工。

(3) 从工龄来看，工作 1 年以内的新职工明显地将员工这一利益相关者排序靠前，而工作 20 以上的老职工对于股东、消费者的排序明显偏后，但对于特殊团体、社区作为企业的利益相关者的认同感明显更强。

(4) 从职务来看，一般员工对于股东、管理人员的认识明显弱于其他层次的员工，而高层管理人员对于消费者和政府作为企业利益相关者的认识明显要强得多，但他们对于社区的认识要明显弱于其他层次的员工。

(5) 从企业性质来看，民营企业对于分销商的认识要明显强于国有企业，但民营企业对于政府的排序却明显偏后，集体企业对消费者的排序也明显偏后。

(6) 从是否为上市公司来看，没有上市的企业对于管理人员、债权人的排序要明显强于已经上市的企业，但它们对于社区的排序明显偏后。

(7) 从企业规模来看，大型企业明显看重政府和特殊团体，而它们对于股东、债权人和社区的排序明显偏后。

(8) 从企业发展阶段来看，处于创业阶段的企业明显注重股东和消费者，但对特殊团体的排序明显偏后，而处于衰退阶段的企业明显看重的是债权人和政府。

至此，我们已经验证了假设 4，即在不同的人看来，在不同的企业中其利益相关者的轻重缓急是有差异的。应该说，上述绝大部分的差异比较分析结果与我们现实中所观察到的现象都是吻合的，或者说这些差异都有其现实意义。限于篇幅，本文不一一赘述这些差异的现实体现，进一步的研究可以通过国际上流行的案例分析法来详细阐明。

六、研究结果与讨论

我们可以将本文研究的结论总结为以下几点：

(1) “多维细分法”和“米切尔评分法”是对企业利益相关者进行分类的基本方法，但其基础应该是实际调研的数据资料，而不能仅仅停留在学院式的思辨研究过程。现实企业中利益相关者的类型是复杂多变的。

(2) 我国企业的 10 种利益相关者在其主动性、重要性和紧急性三个维度上是存在一

定的差异的，从统计结果来看，股东、管理人员和员工是我国企业的核心利益相关者，供应商、消费者、债权人、分销商和政府是企业中的蛰伏利益相关者，而特殊利益团体和社区则是企业的边缘利益相关者。

(3) 不同的个体因素和企业因素会对企业利益相关者的认识产生差异。不同性别、不同年龄、不同工龄、不同职务的人可能会对某一群体作为企业利益相关者的认识具有很大的差异；同样，这种差异也会发生在不同性质、不同规模、不同发展阶段以及是否上市的企业之中。

总体而言，关于我国企业利益相关者的分类研究所蕴涵的意义在于：企业决不是只为股东而生存的，诸多利益相关者受企业决策活动的影响，也会影响企业的决策；同时，我们需要对利益相关者这一群体进行分类，企业在经营运作过程中对不同类别的利益相关者并不需要也不可能“等量齐观”，在科学分析的基础上对不同的利益相关者进行“分类治理”是企业保持持续发展的必然选择。

长期以来，当代主流企业理论和公司治理理论对利益相关者理论的研究存在着两大批评意见：一是利益相关者难以确定，从而使得利益相关者理论难以应用于企业实践；二是与主流企业理论层出不穷、细致入微的实证研究结果相比，利益相关者理论的实证研究结果太少。应该说，这些批评意见正中利益相关者理论的“软肋”，致力于利益相关者理论研究的西方学者也一直在不断地加以改进。首先应该承认，早期泛化界定利益相关者的方法的确曾使利益相关者理论的发展陷入困境，但是 20 世纪 80 年代以后，尤其是 90 年代中期以来，西方学者对企业利益相关者分类的研究取得了重大进展。总体来看，西方学者关于利益相关者的分类实际上是经历了一个“窄定义 - 宽认识 - 多维细分 - 属性评分”的过程：20 世纪 60~70 年代是利益相关者理论的初创期，对于利益相关者的定义是狭隘的；80~90 年代初期，对利益相关者从广义上认识后，迅速兴起了对众多利益相关者进行细分的热潮；90 年代中期以后，量化的评分法大大促进了利益相关者的分类工作，也使利益相关者理论具有了很强的可操作性。针对第二种批评意见，西方学者在 90 年代以后开始进行了大量的实证研究，在国际知名刊物上涌现出 100 多篇有分量的论文，并概括总结出利益相关者理论研究的方法体系 (Donaldson & Preston, 1995)。在国内学术界于 90 年代中后期还在将“股东至上理论”奉作经典的时候，利益相关者理论实际上已经成为西方经济学界炙手可热的话题，只不过由于多方面的原因使得我国许多学者对同期西方企业理论的最新进展持一种漠视和怀疑的态度，以至于“许多人认为利益相关者理论是西方近年来开始流行的一种企业治理结构观，这是一种误解，或者说是缺乏文献的结果” (杨瑞龙、周业安, 2001)。要想完善至今仍处于“非主流企业理论”地位的利益相关者理论，我国学界必须在开展规范分析的同时，更多地从企业实践入手开展实证研究，发掘利益相关者理论在我国经济转型背景下的深刻含义。

参考文献

- [1] [美] 多纳德逊, 邓非著:《有约束力的关系——对企业伦理学的一种社会契约论的研究》, 赵月瑟译, 上海社会科学院出版社 2001 年版。
- [2] 陈宏辉:《企业的利益相关者理论与实证研究》, 浙江大学博士论文, 2003。

- [3] 贾生华, 陈宏辉:《利益相关者的界定方法述评》,《外国经济与管理》,2002年第5期。
- [4] 李心合:《面向可持续发展的利益相关者管理》,《当代财经》,2001年第1期。
- [5] 万建华:《利益相关者管理》,海天出版社1998年版。
- [6] [美]沃克,马尔:《利益相关者权力》,经济管理出版社2001年版。
- [7] 杨瑞龙,周业安:《企业共同治理的经济学分析》,经济科学出版社2001年版。
- [8] Carroll, A.B., A speech. In Jones T. M., *The Toronto Conference: Reflections on stakeholder theory, Business and Society*, 1994, Vol.33, No.1, p128.
- [9] Charkham, J., *Corporate governance: Lessons from abroad. European Business Journal*, 1992, Vol. 4, No. 2, pp8 ~ 16.
- [10] Clarkson, M., A risk-based model of stakeholder theory, *Proceedings of the Toronto Conference on Stakeholder Theory, Center for Corporate Social Performance and Ethics, University of Toronto, Toronto, Canada*, 1994.
- [11] Clarkson, M., A stakeholder framework for analyzing and evaluating corporate social performance, *Academy of Management Review*, 1995, Vol.20. No.1, pp92 ~ 117.
- [12] Donaldson, T. & Preston, L. E., The stakeholder theory of the corporation: Concepts, evidence, and implications, *Academy of Management Review*, 1995, Vol.20. No.1, pp65 ~ 91.
- [13] Frederick, W. C., *Business and society, corporate strategy, public policy, ethics* (6th ed.), McGraw-Hill Book Co., 1988.
- [14] Freeman, R.E., *Strategic management: A stakeholder approach*, Boston, MA: Pitman, 1984.
- [15] Freeman, R. E., & Reed, D. L., Stockholders and stakeholders: A new perspective on corporate governance, *California Management Review*, 1983, Vol.25. No.3, pp88 ~ 106.
- [16] Jensen, M. C. & Meckling, W. H., Theory of the firm: Managerial behavior, agency costs, and ownership structure, *Journal of Financial Economics*, 1976, Vol.3, pp305 ~ 360.
- [17] Knut, H. M. & Svein, J., From user-groups to stakeholders? The public interest in fisheries management, *Marine Policy*, 2001, Vol.25. No.1, pp281 ~ 292.
- [18] Michael, J. P. & Jacquelyn, A. O., Exploratory examination of whether marketers include stakeholders in the green new product development process, *Journal of Cleaner Production*, 1998, Vol.6, No.3 ~ 4, pp269 ~ 275.
- [19] Mitchell, A. & Wood, D., Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts, *Academy of Management Review*, 1997, Vol.22. No.4, pp853 ~ 886.
- [20] Savage, G. T., Nix, T. W., Whitehead, C. J. & Blair, J. D., Strategies for assessing and managing organizational stakeholders, *Academy of Management Executive*, 1991, Vol.5. No.1, pp61 ~ 75.
- [21] Shankman, N. A., Reframing the debate between agency and stakeholder theories of the firm, *Journal of Business Ethics*, 1999, Vol.19. No.4, pp319 ~ 334.
- [22] Starik, M., Should trees have managerial standing? Toward stakeholder status for non-human nature, *Journal of Business Ethics*, 1995, Vol.14. No.1, pp207 ~ 217.
- [23] Thayer, C.E. & Fine A. H., Evaluation and outcome measurement in the non-profit sector: Stakeholder participation, *Evaluation and Program Planning*, 2001, Vol.24. No.1, pp103 ~ 108.
- [24] Wheeler D. & Maria S., Including the stakeholders: the business case, *Long Range Planning*, 1998, Vol.31. No.2, pp201 ~ 210.

职业经理与家族企业的成长^{*}

储小平

家族企业在发展过程中要不断融资，而不管通过哪种途径融合金融资本，特别是股权稀释化融资，必然伴随着与社会人力资本的融合。因为，融资表明企业规模的扩展，业务的增多，技术含量的提高以及经营活动的半径扩大等等。这时，家族内的人力资本供应，特别是管理资源的供应必然不足。因此，吸纳和集成新的管理资源就成为私营家族企业成长的重要课题。

为了简化要研究的对象，本文把家族企业要融合的人力资本主要界定为管理资源，特别是职业经理。做这样的限定，是因为管理资源，特别是职业经理是企业中最主要的人力资本^①。企业与管理资源特别是与职业经理融合的有效性是企业成长中至关重要的环节。

任何企业在成长的过程中，都必然要不断地吸纳和集成新的管理资源。家族企业在吸纳和集成新的管理资源的过程中，其特点是有家文化规则和泛家族规则在其中起着复杂的作用。吸纳的对象、吸纳的过程、吸纳后对管理资源集成的方式、成本及其成效都与非家族企业有很大的不同。

笔者在调研中发现：相当多的私营家族企业的成长瓶颈主要不是金融资本，而是管理资源这种最重要的人力资本。因为这些私营家族企业的内源融资能力和亲属网融资能力都很强，企业的技术、机器设备也都很先进，但仍然陷入重重的发展困境之中。其主要原因就是不能有效吸纳和集成新的管理资源。

1998年，浙江金义集团的业主陈金义对企业家族制进行革命，将金义集团内担任管理职务的30多位家族成员，包括他的妻子、兄弟等几乎全部从集团中退出，引起媒体广泛关注，中央电视台“实话实说”栏目还做了专题节目，引起社会很大的轰动。实际上，突破企业家族制并不是始于陈金义，不少私营企业已经意识到家族制的局限性，并逐步淡化家族制。比如，温州的正泰集团、天正公司、新华公司等都在20世纪90年代中期以前就开始突破企业家族制^②。此后，突破家族制、聘用职业经理的各种成功与失败的案例屡屡见诸报端。1999年，兰州黄河集团差点被经理人颠覆，董事长杨纪强又重新把他的几个儿

^{*} 本文为国家自然科学基金课题“家族企业与社会资本的融合”（项目批准号：79970049）的研究成果之一，并作为征文提交中山大学管理学院和四川大学工商管理学院共同主办的“成长中的中国企业”学术研讨会。原载：《管理世界》，2002年第4期。

^① 当然，企业家资源对企业的创业与成长也是至关重要的。但是，究竟什么是企业家？什么是职业经理？学界并没有一个明晰的划分。毛蕴诗提出了一个“企业家与职业经理特征识别模型”。毛教授提出的这个识别模型很有意义，但仍让人感觉难以在“企业家、资本家、职业经理”三者之间划出清晰的界限。职业经理在一定程度上也具有企业家的功能。本文并不主张企业家与职业经理是绝然不同的两类人。

^② 马津龙：《关于温州民营企业制度创新和管理创新若干问题的思考》，《经济研究参考资料》，1999年第12期。

子安排在高层管理岗位上；2000年，广东中山华帝集团的7位创始人退出“老总”岗位，专职于董事会，聘集团原雇员姚吉庆担任总经理；广东顺德万和集团聘请职业经理人李洪峰（原大学教授）担任主管营销的副总经理；2001年，广东创维集团营销总经理陆强华与老板黄宏生翻脸，率营销队伍近百人哗变；2001年，广西喷施宝公司职业经理王惟尊因揭露企业主做假账，而他自己又被老板以职务侵占和商业受贿罪加以控告，并被拘留8个月之久才被取保候审。此事引起广泛关注，王惟尊和喷施宝老板王祥林同时成为凤凰卫视2001年度华人世界十大新闻人物；2001年，众多IT界和网络公司的职业经理纷纷离职……于是，2001年，关于家族企业利弊、私营企业老板与职业经理的矛盾引发了大范围的讨论，众多著名经济学家、政府要员、企业董事长、首席执行官、总裁等都参与讨论。遗憾的是，对这一重大课题广泛关注的多，深入研究的少。

张维迎（2001）对这一问题给予了高度的重视。他认为，中国企业能不能长大，中国的民营企业能不能发展，很大程度上取决于职业经理人的道德水平。“我们不需要担心缺少企业家，中国人充满了企业家冒险、创新意识，中国要担心的是没有具有足够的良好职业道德和职业行为的职业经理”。他还指出，职业经理人队伍的建设将是中国企业继产权改革之后又一个重大难题。

总之，能否有效地融合社会人力资本，特别是经理管理资源是当前中国私营家族企业成长的关键问题。当深入思考这一问题时，就会发现，困扰私营家族企业融合社会金融资本的两大因素，即披露企业信息的风险和企业主的“集权情结”也成为影响企业融合社会人力资本的主要因素。

直接地看，一方面，中国当代确实缺乏具有足够的良好职业道德和职业行为的职业经理；另一方面，还应该指出，中国当代也缺乏具有足够的良好企业家道德和企业家行为的企业主。深入一层看，即使在市场经济制度比较完善的国家，并非天生地就具备足够的合格的企业主和经理人。我们从钱德勒研究中就可以发现：在美国，足够的合格经理阶层的形成经历了差不多100年的历史。这里有一个复杂的社会经济制度发展，特别是社会信用制度逐步建立健全的过程^①。无论在何种社会经济制度环境中，只要有企业存在，那么企业主与经理人之间的信息不对称是普遍存在的。不过在不同的社会经济制度环境中，企业主与经理人之间信息不对称的程度、表现方式，以及降低这种信息不对称的各种制度安排及其成效有很大的差异。

人力资本与金融资本的区别是：人力资本天然属于活生生的个人，而且，人力资本的“所有权限于体现它的人”（Rosen, 1985），是“独一无二的所有权”，而其他经济资源包括各种金融资本、土地所有权等非人力资本，既可以属于个人，也可以属于家庭、社区、其他共同体或国家，还可以不属于任何人或人的群体（周其仁，1996）。而人力资本市场的信息不对称与金融资本市场的信息不对称也是有差异的。

张维迎（1999）^②分析了金融资本市场信息不对称的主要内容：①关于融资项目质量

① 储小平：《信任与中美家族企业演变的比较及其启示》，《学术研究》，2001年第5期。

② 张维迎：《从非对称信息看公司融资与资本市场》，见：《企业理论与中国企业改革》，北京大学出版社1999年版，第255页。

的非对称信息；②关于融资者选择行为的非对称信息；③关于融资者经营能力的非对称信息。正是由于以上三个方面的信息不对称，才限制了私营企业在资本市场上的融资能力。在这里，第①点的信息不对称可以通过一些技术的手段来降低，比如对项目的可行性研究报告进行评估，只不过有时成本会很高。而第②、③两点的信息不对称可以通过抵押、担保等来降低，尽管成效有限。而人力资本市场的信息不对称包括：①契约前交易双方的人的素质或能力的非对称信息。实物资本与货币资本等可以量化、评估，其功能可以较准确地预测，有些实物资本，比如机器设备的工效在技术上是事先设定的。人力资本的素质或能力是很难直接观察的，尽管学历文凭、经历经验等在一定程度上是人力资本素质的显示，但都不能绝对地作为依凭。文凭不完全等同于能力，经验也可能是此一时彼一时。实物资本、货币资本本身不会有夸大、缩小、掩饰、隐瞒信息等主体主观行为，但人力资本所有者既可以诚实地披露自己的信息，也可以扭曲有关自己的人力资本信息。②契约履行中的非对称信息，特别是相互之间守信程度的非对称信息。由于存在信息不完全和不确定性，人力资本的市场交易不可能签订完备的契约。在契约履行过程中，时时都会出现新的不确定性，经理人的行为选择的机会主义空间非常大。因而，正式契约的约束是有限的。这时守信、声誉机制就是非常重要的。张维迎（2001）认为，无论在产品市场还是资本市场，声誉都是维持交易关系的一种不可缺少的机制。实际上，产品市场、资本市场的交易关系的实质是人与人之间的交易关系。声誉机制约束的是人，而不是没有生命的产品或实物资本。③人力资本的产出有非常大的弹性，有机缘兴会效应，有激励乘数效应，激励者、激励诱因、被激励者的千差万别及其组合所产生的激励效果也是各不相同的。因而，关于人力资本产出，特别是经理人力资本的产出信息是很难事先设定或准确评估的。由于人力资本市场信息不对称的特性，使其与金融资本市场信息不对称有很大的区别：①人力资本市场交易双方在全部的交易过程中，时时都受各自主观因素的强烈影响，其中有很多人格化因素，如性格、兴趣、态度、情感联结度、伦理价值观、习惯性心理与行为等都影响着人力资本的交易，特别是与经理人的交易；相对而言，在金融资本市场，人格化因素的影响主要发生在通过亲属网融资这一渠道中，并且着重关注是否守信这一特质上。充分发达的资本市场体系中，是人格化因素起作用最小的一种经济制度。②人力资本交易达成之后，职业经理人要进入企业内部这个“黑箱”，不同程度地进入企业主的信息密级圈，并分享企业内部的部分控制权。闹不好，“就像一把钢刀插入敌心脏”。而在金融资本市场中，贷款人、债权人基本上是站在企业外部来掌握信息的，即使是上市公司的股东，也只是通过《公司法》和公司章程有限度地对企业内部信息有一定的知情权。③在人力资本市场，交易双方的信息披露并没有法律、法规的明确规定，披露多少、如何披露、披露的程度取决于双方的互信程度和在利益基础上的博弈。而在正式的金融资本市场，信息披露都有法律、法规 and 政策的明文规定。④相对于金融资本市场交易而言，人力资本的交易合约，特别是企业主与经理人之间的交易合约会更不完备，比如，绞尽脑汁尽心尽力地履行职务、主动的企业公民行为、创新精神、协作与团队精神等都不可能用契约来解决。人力资本的产权相当特别：只能属于个人，非“激励”难以调度（周其仁，1996）。激励不当，经理人不仅会限制、压抑、关闭自己的人力资本供给，而且会在企业内部造成人力资本内耗，更有甚者，会策反人力资本，并带走企业的商业机密或资产。⑤传统的华人社会的信

息低规范度和高集中度 (Max Boisot, 1987;) 加重了企业主与经理人之间的信息不对称。学者们的研究表明, 西方社会是“高文本文化”, 信息是清晰和非人格化的, 人们通过各种契约来规范各自的行为。在“低文本文化”中 (典型的是日本、中国), 人们更喜欢做含糊和间接的交流, 而且信息交流较多依靠的是以前人们在共同的文化背景下形成的共识, 经常是三言两语就能使对方明白 (陈凌, 1998)。华人社会的信息交流有更多的察言观色、心领神会的特色。华人企业主 (当然, 不仅仅是企业主, 几乎任何组织中的领导者) 都自然而然地使自己成为信息的集中者和垄断者。在笔者的个案调研中, 发现不少私营家族企业, 其客户状况、营销网络、竞争状况、原料采购、价格等信息很少有系统的文字资料积累下来, 差不多都集中在企业主的大脑中。在很多情况下, 企业主不会轻易地让非家族成员的经理知晓、掌握这些信息。于是, 笔者发现两种情况: 一是经理人, 特别是非家族成员的经理由于没能获得岗位工作所必需的信息, 难以有效地履行职责, 致使企业主认为他无能。英雄无用武之地, 这些经理就会另谋高就了。二是一部分非家族成员的经理获得一些机密信息, 特别是一些客户资料, 并与客户建立了个人间的关系, 于是另立门户, 同自己原来的老板竞争。这种情形已是比较普遍了^①。正是由于上述人力资本市场交易的特性, 使得私营家族企业融合人力资本, 特别是融合经理人力资本的难度很大。

一般而言, 随着市场经济体系的逐步完善, 特别是信用制度的完善, 加上企业规模的扩大、层次的提高和企业主管理理念境界的提升, 企业融合社会人力资本的能力会提高。本文在此将着重考察中国国家家族企业融合社会人力资本的发展过程, 并通过有关调研和案例资料检验以下假设:

(1) 私营家族企业主突破家族人力资本的封闭性的动机和行为正在日益增强。

(2) 信息分享的风险和企业主的“集权情结”成为家族企业融合经理人力资本的主要障碍。

(3) 家族企业内部的特殊主义的用人规则会发生蜕变, 并与“能力至上”的普遍主义规则相融合。

与家族企业融合社会金融资本相比较而言, 与人力资本的融合更多地表现为企业主与经理人分享经营控制权, 企业信息机密程度、企业主与经理人之间的信任程度、企业主的“集权情结”的强度等影响着企业内经营控制权的分配。而家族企业与社会金融资本的融合, 更多地会涉及企业所有权和剩余控制权的分配。

首先, 对第一个假设做一初步的检验。

根据《全国首次私有企业抽样调查数据及分析》所反映的到 1992 年底的情况, 在私

^① 笔者个案调研中发现一个典型的案例: 一家生产出口工艺品的家族企业, 产品质量、交易信用都很好, 业务进展比较好。但这个行业的进入壁垒比较低, 只要掌握一定的客户, 有 20 万 ~ 30 万元的启动资金就可以创办企业、生产经营。所以, 这家企业的中层经理工作几年后, 就辞职另立门户, 同老板分割市场。企业主对此一直很苦恼。于是, 他决定不用那些社会经历较丰富的人, 从大学毕业生中找了一位国际贸易专业的学生, 精心的栽培她, 重用她, 厚待她。几年后, 这位经理人也提出辞职, 老板十分不解, 多次劝留无效后, 只得同意她辞职。随后, 在广州交易会上, 这位企业主发现自己摊位对面、经营同样工艺品的老板竟是他过去精心栽培的那位经理。由于那位经理与原公司的客户有更多的直接接触, 所以不少客户纷纷与她签约。此事对企业主刺激太大。他只有小学文化程度, 不会用英语直接与外商谈判。从那以后, 他发愤学习英语。经过一段时间的努力后, 他终于可以直接用英文同外商谈判签约了, 但他却比以前忙得更加不可开交, 工作常常顾此失彼。如此这般, 企业如何能成长壮大?

有企业内部，所有权与管理权紧密结合。全体所有者都参与经营、管理企业的占被调查户的 73.2%，另有 4.1% 是多数投资者参与管理，15.0% 是投资者中少数人参与管理（一般发生在有限责任公司里），至于投资者不参与管理、经营的企业，只有 2.0%。即使是在有限责任公司里，仍由主要投资者或由投资者中的一部分人管理，极少出现聘请专职经理全权管理的现象^①。

从 20 世纪 90 年代初到 90 年代中，私营企业的发展速度很快。其管理人员的来源及其构成有很大的变化，如表 1 所示。

表 1 私营企业员工进入企业渠道（%）

项 目	与企业主有亲属关系	与企业主是邻居或朋友	亲朋好友介绍	官员、关系介绍	社会招聘
管理人员	29.5	9.5	8.0	0.6	45.4
技术人员	17.6	8.5	10.3	1.2	62.4
工 人	11.4	10.1	11.8	2.4	64.3

资料来源：《中国私营经济年鉴（1996）》，中华工商联合出版社 1996 年版。

表 2 企业主对不同人员的选用标准（%）

选用标准	老实可靠	关系密切可信任	关系户	有技术	有关系网	其他
管理人员	32.2	43.1	0.8	21.8	1.0	1.0
技术人员	12.8	11.3	0.8	74.4	0.4	0.3
工 人	65.7	10.8	4.6	15.5	1.9	1.5

资料来源：同表 1。

表 3 私营企业管理人员的来源（%）

	社会招聘	从本企业基层提拔	董事长或总经理亲属	董事长或总经理亲戚朋友	政府部门委派	其他来源
管理人员总数	43.8	29.24	11.67	8.59	0.58	6.11
总经理	16.00	25.00	42.00	3.00	1.00	12.00
副总经理	29.33	23.55	24.23	11.05	1.50	10.34
三总师	50.10	28.19	9.86	5.63	0.69	5.52
财务部门主管	40.58	24.57	16.37	12.90	1.01	4.57
采购主管	36.15	36.45	15.85	6.79	0.75	4.01
销售主管	41.87	35.60	13.59	3.89	0.46	4.60
仓库主管	34.50	42.10	13.76	4.50	1.50	3.63
一般管理人员	37.50	41.44	9.01	3.11	0.17	8.76

资料来源：《中国私营企业发展报告（2001）》，社会科学文献出版社 2002 年版，第 181 页。

从表 1 至表 3 看：①私营企业经营管理来源渠道已是多元化的，主要有社会招聘、企业主的亲属、基层提拔三种形式。就总体而言，社会招聘逐步成为私营家族企业管理人员进入的主要渠道，但是与技术人员和工人相比，经营管理从社会招聘来的比例要小一些。就高层经营管理而言，社会招聘、基层提拔和企业主亲属三种方式的从

① 参看张厚义等：《中国的个体和私营经济》，社会科学文献出版社 1995 年版，第 379 页。

80年代以亲属为主到90年代中后期, 社会招聘占到40%以上。②与技术人员(如三总师)和工人相比较, 管理人员来源具有更强的内生性。在管理人员中, 与企业主有亲属关系的占20%以上, 而技术人员和工人与企业主有亲属关系的要少得多。这说明管理人员与企业主的关系比技术人员和工人与企业主的关系更为密切, 而且选用管理人员第一位的考虑因素是关系密切、可信任。总之, 可以大致判断, 在私营企业选择经营管理人员方面, 社会招聘机制和亲属网产生机制同时发挥作用, 但社会招聘机制渐渐增强。

《2000年中国第四次私营企业抽样调查数据及分析》指出, 部分私营企业主逐渐抛弃家族制管理办法, 开始建立权力分散、授权经营的科层制管理方式, 有的甚至提出“以资产换管理”, 聘用社会上的专业人士来担任主要管理者。

对此, 结合笔者的问卷调查结果做进一步分析, 见表4、表5。

表4 非家族成员经理进入公司的途径(%)

亲戚介绍	朋友推荐	老乡介绍	同学朋友关系	人才市场招聘	大中专院校招聘	从员工中提拔	其他
7.02	43.86	3.51	17.54	63.16	21.05	66.67	12.28

表5 你是否同意“家族制用人已经阻碍了企业的发展”(%)

同意, 因为家族内的人才毕竟有限	同意, 因为即使家族内人采够用, 也不利于引进人才	同意, 因为本公司已经感觉到了	不同意, 只要家族人才培养得好就行	不同意, 因为人才市场不健全	不同意, 因为我的公司就没受影响	其他
68.63	27.45	21.57	0	0	3.92	1.96

笔者的调研表明: 通过社会招聘的方式产生经理比各种亲属网推荐的比重要高一些。而且, 多数企业主都认为企业用人家族制阻碍了企业的进一步发展。由此, 可以大致判断: 私营家族企业主突破家族人力资本的封闭性的动机和行为正日益增强。

再看“中国私营企业治理结构研究”课题组的调研数据(见表6)。

表6 管理层与经理层交互任职情况(%)

	兼任总经理	副总经理	三总师	一般管理人员	一般职工	不在本企业任职	其他
股东会成员	36.89	33.44	2.76	14.08	5.61	3.85	3.37
董事会成员	28.50	30.35	5.94	22.57	1.42	7.55	3.68
监事会成员	26.67	9.64	5.42	30.77	15.42	5.71	6.37
职代会成员	4.08	2.24	6.73	33.69	46.54	5.38	1.34
工会委员	8.43	16.41	6.58	37.76	30.82	0.00	0.00
党组织成员	22.83	16.01	8.54	29.76	16.30	4.94	1.61

表 6 反映出，企业所有者参与经营管理的比例仍很高。但也可看出，与表 5 反映的情况一样，对三总师这一类专业技术知识要求比较高的管理人员，社会招聘的比例很高。另外，在董事会中还有 7.55% 的人不在本企业中任职。这些董事也许是专业人士，也许是政府官员，也许是完全不介入经营管理的股东。若是前两者，则分别起着专业作用或疏通关系网的作用。但总之可以视为突破家族企业用人的封闭性。

其次，我们尝试检验第二个假设：信息共享的风险和企业主的“集权情结”成为家族企业融合经理人力资本的主要障碍。

要检验第二个假设是非常不容易的。因为各行各业各个地区的各种私营家族企业是千差万别的，企业内的各种信息的规范度、集中度的差异很大。企业内的信息在企业主与经理人之间如何分享？分享程度如何？这都是很难加以测定的。在这里，笔者主要以“管理岗位机密程度的分配序”和“非家族成员的经理在企业内进入的经营管理的层次”这两个变量来判断企业内信息的分享和企业主“集权情结”的程度。之所以选取这两个变量，是因为，一方面，从横向角度看，任何企业内的各个部门、各个岗位的信息机密程度及其对企业风险的影响程度是不同的。有的是要害机密部门，有的是次级机密岗位，有的是一般管理岗位。私营家族企业主对这个问题更加敏感，在管理岗位的分配中，受“信息共享风险”这一因素的支配。另一方面，从企业内纵向科层体制来看，管理层级越高，信息的机密程度越高。如果非家族成员的经理进入管理层级越高，那么企业内的信息分享度越高，企业主的“集权情结”的强度越低。

在上面检验第一个假设时，可以看出，私营家族企业越来越多地从社会招聘管理人才，体现出企业主突破企业用人家族制的倾向越来越明显。这只是从企业吸纳管理资源的对象和渠道的角度进行分析，还没有分析管理资源进入企业内以后，企业对管理岗位是如何分配的。

2000 年“中国私营企业研究”课题组的调研发现，虽然部分私营企业主突破家族制管理的倾向日趋明显，开始建立权力分散、授权经营的科层制管理方式，“用股权换管理”、聘用社会上的专业人士来担任主要管理者的做法也成为热门话题，但其中相当多的是把生产过程剥离出来让“外人”去管理，而财务、销售等环节牢牢地抓在自己手中。

笔者的私营企业问卷专门调查了这一问题，见表 7。

表 7 贵公司里，有外人担任经理的部门所占比例（%）

技术和质检部门	生产部门	财务部门	人事部门	营销部门	总经理助理	副总经理	分公司
76.79	75.12	69.64	66.07	64.29	46.43	48.21	41.07

从这一问题的回答中可以看出：技术和生产部门由外人担任经理的最多，这说明这些部门的工作内容比较清晰具体，企业主与技术生产部门的管理层的信息不对称的程度较低，业绩考核与监控相对容易。如果企业主本人是懂技术的，而且技术与生产过程模仿比较多，那么这两个岗位的信息机密程度更低些，由外人担任经理，企业主比较放心。财务部门由外人担任经理的占 69.64%，这个比重比实际观察要高，企业主对这个问题的回答可能有掩饰行为。其他几个部门由外人担任经理的比重也挺高，我们的猜测是：“外人”

在一定程度上被泛家族化了，这些“外人”中有相当一部分已成为企业主的亲信。这一问题是从静态的角度看“外人”在各种不同的经理岗位中的分布，下一问题是从时序上看企业主在安排外人做经理时，对各种不同的经理岗位做何种顺序安排，以期从中得出有价值的发现。

在问卷中，我们专门设计一道题，看企业主对聘请外人做经理在岗位分配上是如何做顺序安排的，从回答中可以发现有以下时序安排：①生产部经理；②质管部经理；③设计开发部经理；④办公室主任；⑤副总经理；⑥总经理助理；⑦营销经理；⑧财务经理；⑨人事经理；⑩采购经理。很显然，管理岗位排序与上一问题的回答结果大致相同。表明企业主首先是将机密程度较低的生产技术部门和例行公事管理（办公室）的经理岗位对外人开放；对副总经理和总经理助理的时序安排上，实际上出现两个极端，小部分排在前面，多数排在最后面，这说明有值得信赖的就先安排这样的岗位给他们，要不就到最后才考虑把这样的岗位安排给值得信赖的外人；营销、财务、人事、采购的岗位机密程度高，一般是到最后才安排给值得信赖的外人。

关于从纵向的角度看非家族成员的经理进入管理的层级，从而了解信息分享和控制权分配，更是一个比较困难的问题。已有的调研在这个变量上未给予重视，笔者的调研也存在这个遗憾。但是，可以就现有的调研资料的一些统计，从间接的角度做一分析。

《中国私营经济年鉴（2000）》中的数据表明：私营企业在重大经营决策方面，由主要投资人决定的比重呈递减趋势，分别从1993年的63.6%、1995年的54.4%下降到2000年的43.7%，而由主要投资人和主要管理人员共同决定的比重略有上升，由1993年的20.7%、1995年的25.6%上升到2000年的29.1%，这种变化可从一个侧面显示专职经理人员已有一定的比例进入决策高层。从数据中可以看出，在总经理、副总经理和三总师这几个高层管理岗位上，董事长或总经理的亲属、亲戚朋友参与管理的比例分别为45%、35.28%和15.59%。这大致表明有相当比例的非家族成员的经理进入管理高层。这个统计中有一些含糊的地方，比如朋友就是非家族成员，但可视作已成为泛家族成员。非家族成员经理进入管理高层的状况与企业规模有紧密关联，对此可看《1995年全国私营企业抽样调查数据及分析》所提供的—个表（见表8）。

表8 不同规模私营企业决策方式（%）

企业规模（实有资产总额：万元）	企业主本人单独决策	董事会共同决策	企业主和主要管理人员共同决策
30以下	68.9	5.3	25.6
30~100	62.0	13.4	24.5
100~200	51.1	20.4	28.5
200~500	50.1	24.9	24.9
500~1000	45.4	26.5	28.1
1000以上	28.8	47.7	23.4

表 8 表明，随着企业规模的扩大，由企业主本人单独决策的比重呈明显下降趋势，董事会共同决策的比重呈显著上升趋势，这说明权力随企业规模的扩大而分散。但由企业主与主要管理人员共同决策的比重没有提高，说明非家族成员的经理进入管理高层的比例很低。

以上分析表明，企业内信息分享的风险和企业主的“集权情结”依然是当前私营家族企业融合社会人力资本的主要障碍。但应该看到，这个障碍在逐渐降低，没有充分的根据说明华人家族企业主绝对不能与经理人分享信息，企业主的“集权情结”的强度也是可以逐步降低的。下面看作者的问卷调查，见表 9。

表 9 对部门经理是否充分授权 (%)

充分授权	有限授权	有监督地授权	依部门授权	依人授权	以事授权	依制度授权	其他
8.9	41.1	14.3	19.6	16.1	12.5	42.9	1.8

从对这一问题的回答中可以看出，企业主在一定条件下是愿意授权的，特别注重以制度授权。在笔者个案调查中，发现不同的企业差异很大，这与企业规模、企业主对管理规范化的意识及其追求有关。再看下一问卷题，见表 10。

表 10 对非常信任的非家族成员经理所可能采取的激励方式 (%)

激励方式	所占比例
给他更高的物质收入	38.3
给更好的福利待遇	42.5
晋升到更高、更重要的岗位	38.3
争取多的社会荣誉和地位	10.6
给更多的学习深造机会	23.4
当自家人参与讨论重大决策	36.2
给部分股权	25.5
给更大的管理权限	40.4
其他	0

对这一问题的回答中，有 36.2% 的企业主注意将非家族成员的经理予以泛家族化，在个案调查观察中，在一些家族企业内部，下属对企业主及其担任管理岗位的家族成员，一般不以职务相称，而以“哥”、“姐”等家族成员角色称呼，在企业中弥漫着一种泛家族氛围。

总体看，企业内信息分享的风险和“集权情结”对私营家族企业融合人力资本的障碍还没有充分降低。现实报道中发现，一些出身高级知识分子的私营企业主也极力为家族企业辩护，如西安海星科技实业集团公司董事长兼总裁荣海就说：“在 1993 年的时候，我在公司里进行了一次大‘清洗’，包括亲戚、同学、朋友，但是后来发现，家族式管理在中国有它的道理。”重庆力帆集团老板尹明善声称家族企业 50 年不变。之所以会出现这种情形，是因为当代中国私营家族企业的发展历史还不够长，另外，企业外部的社会制度安排

还不能充分保证降低企业内信息分享的风险和“集权情结”松动后的控制权失控的风险。另外，从表 10 中可以看出，企业主对经理人员让渡部分经理管理控制权的比重远大于让渡部分股权。重股权控制，逐步授让经营管理权是私营家族企业治理结构优化的一个基本途径。对此，将在另文中予以探讨。

下面将检验以下第三个假设：家族企业内部的特殊主义的用人规则会发生蜕变，并与“能力至上”的普遍主义规则相融合。对这个假设的检验，主要是想进一步探讨私营家族企业在企业内部是如何整合管理资源的。

社会学家 (Parsons and Shils, 1951; B. Hoselitz, 1960; Blau, 1987) 将特殊主义和普遍主义作为传统社会与现代社会的重要区别之一。所谓特殊主义，是指根据行为者与对象的特殊关系而认定对象及其行为的价值高低，维护与行为者的属性有着密切关系的那些价值的至高无上性。而普遍主义则是独立于行为者自己的地位与对象的地位的特殊关系，坚持效率、利润和机会均等的原则。比如，在挑选所长时根据专业能力是普遍主义，因为某人是亲友而重用他则是特殊主义的。在落后的经济中，报酬多根据宗亲关系、地位和性别来分配；安排经济职务多根据特殊主义的标准。而在发达经济中，通过教育而获得的专业技能越来越决定着劳动报酬；对经济职务的竞争是公开的，择人用人是根据客观可变的能力标准而非特殊的社会关系。^① 现在，我们用“特殊主义 - 普遍主义”这一对变量来考察家族企业内部整合管理资源的特点及其变化。在前面的分析中，实际上已涉及这一问题。在家族企业创办及发展的一定阶段，其管理资源主要是内生供应的，用人的准则是根据特殊主义的规则。随着企业规模的扩大与企业主素质的提高，相当多的私营家族企业在吸纳和整合管理资源时开始从特殊主义向普遍主义转化，在转化的过程中，往往是将特殊主义的用人规则和普遍主义的用人规则融合起来或相互之间形成制衡关系。概括起来大约有以下几种情形：

(1) 用泛家族规则整合管理资源。1997 年的私营企业调研报告指出，在家族制企业中以业主为核心，家庭近亲占据财务、采购、销售等关键职位，家族成员分布在生产、技术、管理的各个环节。对于非家族成员，则用“家族化”的方法，把他们变为“准家庭”、“家族式”的成员。对于邻居、乡党、亲近的同学、同事、朋友，逐步植入家庭内部长幼、兄弟之间的“孝、悌”观念，把他们视为家人，以便他们也能信守互助、互惠和信任的家庭价值观，通过把公共关系变为私人关系，利用传统文化来促进企业的利益。在这里，课题组提出了一个中国私营家族企业整合企业内部人力资源，特别是管理资源的一种特殊的重要手段，即把非家族成员予以“泛家族化”。用“泛家族化”的规则吸纳和整合企业的人力资本，特别是管理资源，将特殊主义的用人规则与普遍主义的用人规则结合起来，将亲情、信任和能力融合联结起来，是华人家族企业组织行为的一大重要特色。

(2) 以亲制疏。从前面内容的分析中可以看到，相当多的私营家族企业在吸纳、整合管理资源方面是采用以亲制疏的规则。笔者的调研也发现大量这种情形。把“家里人或自己人”安排在重要关键岗位，“外人”安排在一般岗位；或在同一部门既安排“自家人”，也安排“外人”，相互之间形成制约关系。在有的企业，专门安排“自家人”做督察工作。

^① 表 3 的统计数据为中国社会科学院“中国私营企业治理结构”课题组对 179 家私营企业所做的抽样调查。

这种以亲制疏的规则整合管理资源的做法是比较普遍的。表 11 是笔者对这一问题的问卷调查。从对上述问题的回答中可以看出，企业主以亲制疏的动机是非常明显的，同时也注重让“自家人”向有能力的外人学习，以提高才干。

表 11 由家族成员与外人共同负责的好处 (%)

通过家族成员来监督	让家族成员更好地学习	让他们相互竞争、相互制约	中层经理辞职时，有人接替	其他
20.69	55.17	48.28	13.79	6.9

(3) 业主与多类别员工的互动。企业内的员工类别不能简单地以“亲疏”来划分，以亲制疏的规则不足以概括华人家族企业整合管理资源的全部特性。

华人社会人际关系的突出特征是“差序格局”（费孝通，1948）。与西方的团体格局相比较，中国传统的社会结构是一种同心圆波纹性质的差序格局。波纹的中心是自己，与别人发生的社会关系，就像水的波纹一样，一圈圈推出去，随着波纹与中心的远近，而形成种种亲疏不同、贵贱不一的差序关系。著名管理学家、台湾大学的郑伯（1995）由此对华人企业组织行为做了非常精湛的研究。他从企业主持人的认知结构来探讨这一问题，指出，华人企业主通常兼具家长、企业所有人及主要经营这三种角色，它不但是企业的讯息中心、资源分配者，而且须负企业成败之责任，因此，对企业经营拥有举足轻重的影响力。企业主根据关系亲疏、忠诚高低及才能大小等三项归类标准，将员工归类为八种类别原型，分别称之为：经营核心（亲/忠/才）、事业辅佐（亲/忠/庸）、恃才傲物（亲/逆/才）、不肖子弟（亲/逆/庸）、事业伙伴（疏/忠/才）、耳目眼线（疏/忠/庸）、防范对象（疏/逆/才）及边际人员（疏/逆/庸）。企业主持人与这八种类型员工之间的互动法则，即构成华人企业组织行为运作的基础。这八种类型的员工在组织里面分别扮演不同的角色，与企业主的互动中形成较为稳定的关系，但也会在一定条件下出现类别异位的情形。探讨企业主持人与类别原型间的互动，及与自己人意识或信任程度的关系，将可掌握华人企业主持人不能一视同仁对待员工的关键，进而能够进一步了解华人企业内部复杂、多样而具有特色的组织行为。郑伯提出了一个非常有价值的深入研究华人企业组织行为的方向。如果企业主的领导艺术高明，用这种方法整合人力资本，特别是管理资源，会提高管理绩效，反之就不一定了。企业主与多类别员工互动式组合人力资源，特别是管理资源的组织行为模式，有一定的经验支持，但还需要进一步做实证研究。

(4) 一视同仁。有关企业的报道和笔者的调研也发现，少数私营企业在所有权上保持家族制，但在用人上却极力淡化家族制，大力倡导能力至上，比如太太药业集团、浙江金义集团、浙江正泰集团等。随着竞争压力的增大和人力资本市场规范度的提高，用“一视同仁”的规则整合管理资源的家族企业胜出的概率会提高。表 12 是笔者的问卷调查。

表 12 任命经理的主要依据 (%)

亲戚、朋友	同学、同事、战友	能力和品德	业绩和责任心	对您的忠诚	其他
0	0	84.3	60.8	21.6	2.0

对这一问题的回答不完整，与现实有蛮大的差异，但重视能力、品德和业绩这三个因素是反映绝大多数企业主的心态的。

笔者在问卷调查中有这样一道题：“吸收非家族的外人当经理，您看重的是什么？对选项回答的顺序依次是：①个人的经验和才干；②个人的品德；③对企业的忠诚程度和责任心；④有良好的社会背景，能拉到客户；⑤受过良好的教育；⑥其他方面的素质和能力。

再看对下一问题的回答（见表 13）。对上述问题的回答也表明企业主非常重视能力因素，同时做法上很有中国那种人情味，同时没有忘记运用“以亲制疏”的规则。

表 13 对能力不适应企业发展的家族成员或公司创业元老做何安排（%）

安排方法	所占比例
直接从领导岗位上撤下来	21.4
离职，让他们做督察工作	62.5
把年轻的送去学习	19.6
让年纪大的退休，但保留待遇	32.9
给笔钱，让他们自己创业	16.1
其他	5.4

从发展趋势上看，会有越来越多的家族企业主意识到，企业毕竟不是家族；充满商业竞争和一定程度市场化的现代社会，也毕竟不是传统的家族社会。因而，在用人上必然要向着注重人才的素质、技术和受教育水平等各种普遍主义的标准发展，向社会招聘人才，将亲情、忠诚和才能三者结合起来，将逐渐成为私营家族企业最重要的融合社会人力资本的手段。真正完全以普遍主义的规则吸纳、整合人力资本，在中国可能是一个比较长的过程，也可能不会成为华人企业组织行为的普遍模式。从现实情况看，有相当多的家族企业在突破用人家族制的变革中受到重创，兰州黄河集团就是一个典型。由于缺乏有效的制度资源的支撑，那些在突破用人家族制方面碰了壁的企业又返回来强化了家族制，形成了企业家族化的路径依赖。突破企业家族制的风险和损失越大，则返回来，企业主又重新动员、组合、启用家族和泛家族成员担任管理职务，由此，虽然可以降低企业信息分享的风险和控制权失控所可能造成的损失，但不可避免地抑制了企业吸纳、集成新管理资源的能力，从而限制了企业的进一步成长。私营家族企业要突破这种负效应的路径依赖，还任重道远。请看笔者的调研中企业主对下面两道问题的回答（见表 14、表 15）。再看“中国私营企业治理结构研究”课题组的调研统计（见表 16、表 17）。

表 14 企业主对能否实现职业经理式管理运作的态度（%）

非常乐观	很乐观	看好	不好说	不看好	很悲观	非常悲观
3.9	13.7	47.1	25.5	11.7	0	0

表 15 对职业经理式管理运作不看好的理由 (%)

理 由	所占比例
职业经理市场没有完全建立	52.2
即使建立了经理市场，运作也不规范	23.9
没办法查询经理人以往的个人信用	37.0
人才交流中心不值得信赖	8.7
很难完全了解一个人	17.4
法制不健全	39.1
其他	2.2

表 16 经理人才市场发育状况

项 目	企业数	百分比 (%)
已经成熟	11	6.4
比较成熟	45	26.3
已经起步，但需要改进	76	44.4
很不成熟	12	7.0
不清楚	27	15.8
合计	171	100

资料来源：《中国私营企业发展报告（2001）》，社会科学文献出版社 2002 年，第 190 页。

表 17 是否可以找到合格的经理人才

项 目	企业数	百分比 (%)
可以	55	32.4
也许可以	75	44.1
招不到	12	7.1
不清楚	28	16.5
合计	170	100

资料来源：同表 16。

对上述问题的回答表明，一方面，企业主非常盼望能实现职业经理式管理（当然，他们心目中对什么是职业经理式管理、如何运作职业经理式管理可能还是非常含糊的），职业经理人市场也初步发育起来；另一方面，职业经理人市场的建设、职业经理式管理的外部制度建设等在我国还是一项长期的艰巨的任务。

信任与家族企业的成长^{*}

储小平 李怀祖

任何企业成长的实质都可以归结为其是否能有效地与社会金融资本、社会人力资本、社会网络资本和社会文化资本融合的问题。家族企业，尤其是华人家族企业在融合不同层面的社会资本的过程中，会受到家文化规则的更大的影响。初步研究表明：在不同国家，信任资源和结构是有差异的，家族企业在成长的不同阶段对信任资源的需求也是不同的。因此，本文将从分析家族企业成长的瓶颈入手，着重分析中国社会的信任结构及其对家族企业融资及其成长的影响。

一、家族企业的定义与成长模式

学者对家族企业有多种不同的定义。大多是从所有权与经营权的角度来界说的，并认为家族企业并不是仅指所有权与经营权完全为家族成员所拥有的封闭式企业组织形态。本文认为，家族企业是家族资产占控股地位、家族规则与企业规则的结合体。它的所有权和控制权表现为一种连续的状况，包括从所有权与控制权不可分离的紧密持有形式到企业上市后，家庭成员对企业资产和经营管理保持临界控制权的企业（叶银华，1999）。

一般而言，企业成长的标志主要体现在三个方面：资本规模的扩大、组织结构的扩展和市场盈利能力的提高。对家族企业而言，资本规模的扩大，表明家族企业要突破家族资本的封闭性，不断吸纳外部的社会金融资本；组织结构的扩展，表明企业的委托-代理链条要不断拉长，为此要逐步突破家族管理资源的封闭性，要吸纳外部职业经理进入企业，并要有效地对人力资本进行有效的整合；市场盈利能力的提高，表明家族企业要在有效融合金融资本和人力资本的基础上，建立广泛的商业网络或企业联盟，并能塑造具有特色企业文化的核心竞争力。总之，家族企业要在上述互有内在联系的三个方面获得成长，就必须不断地吸纳和有效地整合各种社会资本。

由此可见，家族企业成长的核心是如何与各种社会资本的融合。企业的成长实质上就是不断地有效融合社会资本的过程，企业成长的障碍就是不能有效地融合社会资本，企业的衰落或夭折就是融合社会资本过程的中断。由此，笔者建构了家族企业成长的一个路径模式，见图1。此模式表明：从家庭/家族企业到泛家族企业、泛家族企业向企业网络联盟的演变或向家族掌握临界控制权的演变是华人家族企业成长的两个主要路径。这就是说，并不是所有的家族企业成长的最终形态都是向所谓的股份上市公司方向发展，有的保存家

^{*} 本文为国家自然科学基金项目“家族企业与社会资本的融合”成果，批准号：79970049。原载：《管理世界》，2003年第6期。

族企业的基本形态，但与其他类型的企业结成网络联盟，这在台湾中小企业发展中表现得比较明显；有的家族企业的成长可能停留在某一阶段，甚至一直保持纯家族形态；有的企业在成长过程中可能越过一些阶段，有的可能在成长中越不过某一结点而被淘汰，有的可能在不同路径之间跳跃发展，这表明并非所有的家族企业成长都是遵循一种发展路径或都必须越过每一个结点。



图1 家族企业成长的路径模式

二、家族企业成长瓶颈的实质

家族企业如果不能与各种社会资本实现有效的融合，则其成长都会遇到瓶颈。笔者所勾画的家族企业成长路径模式表明，很多家族企业在成长过程中迈不过一些结点、突不破瓶颈而长不大或衰败。国际上研究企业成长理论颇有盛名的彭罗斯（Edith Penrose，1959）认为，企业成长受到企业吸纳新管理能力的数量和速度的限制。这着重是从企业与社会人力资本，特别是与管理资本融合的角度来探讨企业成长问题的。看起来，家族企业的成长也必然会遇到彭罗斯效应中吸纳与集成新管理能力的问题。家族企业进入成长阶段，必然增加对管理资源的需求，有效的管理服务增加的速度越快，则企业成长的速度也越快；反之，则企业成长受到约束。从笔者对较大范围的私营家族企业实地调研和问卷调查中发现，大多数私营家族企业的管理规范化程度低，管理资源严重不足。因而，企业主事必躬亲，忙累不堪，企业家精神或能力困厄于日常经营管理活动中，而不能用于企业的扩张和创新发展。笔者所做的问卷调查表明，有90%以上的企业主认为家族制用人已经阻碍了企业的发展；有57%的企业主不赞成一定要用自己的子女来掌管企业的经营控制权。从全社会的角度看，经过20多年的改革开放，企业外部的管理资源不足状态大为缓和，可企业内部的管理资源为何依然严重不足？

如果稀缺的是管理资源，那么，总可以通过人力资本市场价格的诱导来提高人力资源的供给。虽然企业主个人甚至家（庭）族的管理资源的不足制约了企业的成长，但可以融合他人的管理资源来弥补。因此，真正稀缺的可能不是管理资源，而是别的因素。李新春、胡骥（2000）敏锐地看到，企业家型的企业的成长约束主要不是吸纳和集成管理资源

的能力，而是在吸纳新管理资源时企业主必然要受让部分控制权；在集成管理资源时企业主要付出很多整合成本，一旦吸纳和集成新管理资源超出了企业家个人能力及其控制幅度，则企业内控制权的分配就成为企业成长的阻碍。因此，“企业家控制的企业的成长主要地受到企业家个人经营能力和控制能力的约束”。这种见解富有启发之处在于：管理资源的约束与企业家能力的约束是不同的，前者可以通过经理人市场去吸纳和集成，而企业家能力的约束则很难通过这一渠道来解除。因为企业家能力很难由人力资本市场的供应者提供现成品，企业家能力可能是由个人的天赋、后天的机缘与学习曲线等多种因素构成的，其能力是一个过程（李新春，2000）。缓解企业家能力约束的一个重要条件是新吸纳和集成的管理资源，使企业家能从日常经营管理中解脱出来而致力于企业的扩张性活动。从日常经营管理中解脱出来，其实质也就是使企业家受让部分控制权，而这有可能导致分权后的失控（如发生财务信息泄露或经理层“政变”）。因此，企业家可能通过在各部门安插“自己人”来充当“耳目”，以此来降低信息不对称的程度和处理失控问题（李新春、胡骥，2000）。

上述分析实质上揭示的是：真正制约家族企业成长的是信任资源。由于信任不足，因而家族企业主难以从经理人市场吸纳管理资源，因为有效的经理人市场的缺失也就是社会缺少制度来保证企业能吸纳到守信的经理人；由于信息不对称和信任不足，企业家受让控制权时心有疑虑，对授权后可能导致的风险、成本有很高的敏感度和预期。因此，大量的家族企业发展陷入以下困境之中：爱也经理人，恨也经理人，因此，授权—失控—收权也成为私营家族企业中一个较为普遍的现象。笔者在个案调研中发现，在一些家族企业内部弥漫着“做贼与防贼”的氛围。有位家族企业主苦恼地说：下面有人在不断地“埋雷”，而我要不停地“挖雷”。企业内部的这种低信任状态怎么能使企业有效地成长？张维迎（2001）指出：“职业经理人的职业道德对企业的扩张十分重要。如果没有职业道德，对所有者缺乏忠诚，就不会有资本家信任他们，那企业只能是一个家族企业。”中国当前不缺有能力的经理人，但却缺乏既有能力又讲信用的经理人。张维迎强调职业经理人的职业道德是十分重要的，但他似乎忽略了另一方面，即企业主的信用道德水平。据笔者的调研发现，部分私营企业主为了掌控非规范经营活动的信息，而不愿或不能与外人经理分享企业机密信息，不肯信任和授权给经理人，使经理人的的人力资本供给和产出受到极大压抑，从而抑制了企业的成长。

理论分析和现实中家族企业成长的困境使我们必须对中国社会的信任资源及其结构做进一步的探讨。

三、华人社会信任资源的考察

西方很多学者都认为华人之间的信任度很低或是有限的。100多年前的传教士亚瑟·亨·史密斯（2001）在其《中国人的性格》一书中就认为，不诚实和相互不信任是中国人的两大特点。马克斯·韦伯认为，“儒家君子只顾表面的‘自制’，对别人普遍不信任，这种不信任阻碍了一切信贷和商业活动的发展。”“在中国，一切信任、一切商业关系的基石明显地建立在亲戚关系或亲戚式的纯粹个人关系上面，这有十分重要的经济意义。伦理宗

教，特别是新教的伦理与禁欲教派的伟大业绩，就是挣断了宗族纽带，建立了信仰和伦理的生活方式共同体对于血缘共同体的优势，这在很大的程度上是对于家族的优势。”^① 福山认为，“因为华人文化对外人的极端不信任，通常阻碍了公司的制度化，华人家族企业的业主不让专业经理人担任管理重任，宁愿勉强让公司分裂成几个新公司，甚或完全瓦解。”^② 雷丁虽然并不认为华人对外人绝对不信任，但他也指出华人之间的信任是有限的，是个人之间的信任，“主要特点似乎是对家庭的信任是绝对的，对朋友或熟人的信任只能达到建立相互依赖关系，双方都不失面子的程度。”^③ 这些学者的基本评判是华人之间的信任度低，所以华人家族企业难以成长为大规模的现代企业。这是华人家族企业成长实践中，也是研究华人企业成长的学者不能回避的一个大问题。

有些学者对这一问题作了一些回应性探讨。King (1991) 认为，华人社会的特殊主义的信任含有工具理性，因而可推论中国文化可以容纳某种程度的普遍信任；Kao Cheng-shu (1991) 在研究了台湾大型企业中的信任问题时发现：“‘私人信任’是一个关键机制”，在商业组织之中和组织之间，“私人信任”是基本的组织原则，是商业交易的基础。同时他也指出，这种“私人信任”也具有工具理性。Wong Siu-lun (1991) 认为，近代传统中国的商业信用，在本质上仍是个人化、特定化的，经过与西方的接触，系统信任与普遍信任已在香港和海外华人社团中建立起来。相对而言，彭泗清 (1999) 对这一问题的回应更有深度。他认为“对外人不信任”有两种情形：一是起点上的不信任；二是永远的不信任。由于对外人的信息了解不多，与外人的关系较为疏远，所以对外人信任度较低是合情合理、举世皆然的，西方社会也是如此。起点上的不信任可以通过交往互动中的关系运作达到了解，从而建立信任关系。西方学者认为华人社会的信任突不破血缘亲情的屏障，表明他们对华人社会信任资源特性的认识极为肤浅。

显然，对华人社会的信任资源要做新的考察。初看起来，相对于西方社会的普遍信任而言，华人社会是以家和家族为核心的内外有别的特殊主义信任。但是，中国古代的历代先哲们认识到“一己之私”信任的弊端，因而制定了超越狭隘的家族信任屏障的一整套规则，即泛家族的信任资源。泛家族信任的演进机制是：华人在合作互动的历史实践中不断地将与他们有着地缘、业缘、学缘、神缘等联系的外人予以“家人化”，变成“一家人不说两家话”，信任边界在不断扩展着。久之，先哲们就概括提炼出一套家文化的价值观念和治理天下的理论体系，并被皇朝接受作为统治的制度化了的意识形态，即纲常伦纪、礼法制度。数千年制度强化的结果，使整个国家的上中下各层对这一套规则达到很高的共识度。所以，在民间社会，即使人们之间没有血缘、地缘、业缘、学缘和神缘等联系，但可以通过对忠义规则的共识结拜成兄弟，如《三国演义》中的刘、关、张桃园三结义，《水浒传》中 108 条汉子的兄弟结盟等无数的历史文学形象和民间传说；士绅阶层之间也可“以道为朋”，突破血缘信任的局限。整个天下的治理也靠泛家族化的道义礼法制度的信任来维系，这一点宋代的司马光在《资治通鉴》中说得很明白：“夫以四海之广，兆民之众，

① 马克思·韦伯 (1920):《儒教与道教》，商务印书馆 1995 年版，第 296, 289 页。

② 弗兰西斯·福山：《信任——社会道德与繁荣的创造》，远方出版社 1998 年版，第 96 页。

③ 雷丁：《海外华人企业家的管理思想——文化背景与风格》，上海三联书店 1993 年版，第 85 ~ 86 页。

受制于一人，虽有绝伦之力，高世之智，莫不奔走而服役者，岂非以礼之纪纲哉！是故天子统三公，三公率诸侯，诸侯制卿大夫，卿大夫治士庶人。贵以临贱，贱以承贵。上之使下犹心腹之运手足，根本之制枝叶，下之事上犹手足之卫心腹，枝叶之庇本根。”^①

中国社会传统中发展出这一套理论体系和信任资源，突破了血缘亲情信任的局限，但同时也为自身的信任扩展规定了限度。因为这套理论体系和信任规则建立在以人治为基础的身份等级制的基础上，一方面，一旦高层统治者的行为不符合纲常伦纪的基本规范，则中下层社会对这套规则的共识度就会降低，然后经过一段时间长度不等的混乱以后，这套规则又渐渐赢得高共识度，新的皇朝又重构起来；另一方面，这套信任规则毕竟是差序格局，是长幼尊卑的身份等级社会，而且长期的人治规则造成“中国人更多地忠于人，而较少地忠于原则或思想”^②的“私忠、私人信任”的国民性格。显然，这种信任体系有扩展的空间，但却是难以无限扩展的。相比而言，西方社会并不是通过发展出一套泛家族规则来扩展信任，而是如韦伯所言，新教伦理挣断了宗族纽带，建立了信仰和伦理的生活方式共同体对于血缘共同体的优势，这在很大的程度上是对于家族的优势。笔者认为，这种优势突出体现在近代以来的社会发展过程中，它将信任扩展建立在理性化的法律制度基础上，法律制度有更大的普适性和强制性，通过法律契约和对非个人化的市场经济规范的更高的共识度，在信任扩展中大大降低了中国式的泛家族信任扩展中那种对交往对象的亲疏关系识别、信任度的鉴别、交往类型的差序弹性互动的高成本，同时由公共司法机构对失信行为实施惩罚的成本在总体上也远远低于泛家族社会中对失信行为的惩罚成本，惩罚的效率和公正性的广度也大大提高，因而，社会经济交易中，特别是在超出“熟识社会”以外的交易中可以产生更高、更广泛的行为预期。这一点在人类历史从乡土“熟识社会”发展到现代的“匿名社会”的阶段显得更具有关键意义。

四、华人社会的信任结构

关于信任问题有着众多的讨论，不同学科的学者对信任所定义的侧重点有所不同。相对而言，一般比较倾向于赞同甘比塔（Gambetta，1988）所下的定义：“信任……是一个特定的主观概率水平，一个行为人以此概率判断另一个行为人或行为人群体将采取某个特定行动……当我们说我们信任某人或某人值得信任时，我们隐含意思就是，他采取一种对我们有利或至少对我们无害的行動的概率很高，足以使我们考虑与他进行某种形式的合作”。^③信任既然是人的主观概率预期，那么就与他所生存的社会历史文化传统紧密关联。不同的社会可能会演变出不同的信任产生机制，但也有一些共性。祖克尔（Zucker，L. G.，1986）对信任产生机制所作的概括具有经典意义。她划分了三种机制：①基于过程的信任模式。这种信任模式依赖于过去的交易的经验，如信誉和礼物的交换。这表明可根据一个人过去的行为和声誉的了解而决定是否给予信任。②基于特征的信任模式。此模式依

① 司马光：《资治通鉴》，中华书局1956年版，第2~3页。

② 雷丁：《海外华人企业家的管理思想——文化背景与风格》，上海三联书店1993年版，第177页。

③ 转引自 Oliver E. Williamson：《治理机制》，中国社会科学出版社2001年版，第319页。

赖于个人的家庭背景和种族。这是一种由社会相似性而产生的信任，如根据他人与自己在家庭背景、种族、价值观念等方面的相似性的多少来决定是否给予信任。一般而言，相似性越多，信任度越高。因为相似的社会背景往往意味着有相近的行为规范，容易相互理解，在交往或经济交换中容易达成共识等。^③基于制度的信任模式。此模式基于非个人性的社会规章制度，促使这种制度信任产生的原因大致有三个：一是交易双方的社会距离，即交易跨越团体进行，而这些职业和工业团体有着不同的交易预期。二是交易双方的地理距离，制度信任更容易跨越地理距离而传播开来。三是社会体系中不可分离的交易数量，尤其是大型的相互依赖的网络交易的产生。

祖克尔所提出的三种信任产生机制对我们考察华人社会信任机制的产生提供了很好的借鉴。传统的华人社会中，信任产生的机制大约也是三种类型：①家庭、家族生活中所习得的核心层信任。这依然是华人社会信任层级结构中的基石。②传统中由纲常伦纪的礼法制度所产生的信任。这种信任在传统中国社会也能跨越地理距离和团体，使人们对相互的交往产生预期。但这种信任在现代化进程中已经不复存在了，代之而起的是现代理性制度信任的逐步建构。从西方社会发展的历史看，制度对信任的广泛形成产生了重要的支持，使得更进一步的风险承受能力与信任行为成为可能。这种支持在组织中可以产生更好的团队和工作绩效；在社会层面里，可以利用法律制度等来更广泛、更有效地保护个人财产和相互间交易的权益。③泛家族关系运作中由习俗、道德规范和礼尚往来所产生的信任。这是华人社会中最复杂的信任，即泛家族信任。这种信任的产生与祖克尔所讲的基于过程的信任产生机制有相同的一面，即与人们之间重复性交往的次数、频率以及相互对预期回报兑现的感受相关。但与祖克尔所讲的也有不同的一面，华人社会泛家族信任的产生不仅与当下的人们之间交往的次数、频率和回报预期感受相关，而且与华人社会长期形成的一些社会交往的文化规则有关。这些文化规则对信任的产生有一种先验性的奠基作用。泛家族信任既包含有亲情的信任，又包含有算计性的工具信任。

威廉森 (Williamson, 2001) 等对算计性信任 (calculative-trust) 作了较为深入的研究。他们认为，以算计为基础的信任是建立在理性选择的基础上的，是在经济交易中对得失权衡比较中所产生的信任。当委托人认识到受托人采取了一种对其是有利的行动的时候，信任就会出现。在这种信任中，信任的产生不仅是可能存在某种程度上的威慑，即如果采取不守信的行为可能会受到威胁或惩罚，而且可以通过一些信号显示 (文凭、资格证书、有声望者的推荐等) 掌握受托人的意愿和能力的可靠信息。在西方，算计性信任一般出现在经济的而非个人的特殊交易条件下，失败的迹象在短期内能够得到控制，委托人的利益一般不会受到很大的损失。在华人社会的泛家族信任中，算计性信任往往掩盖在面子、交情等表象之中，骨子里可能想锱铢必较，但表面上却“碍于情面、不好意思”，因而华人之间的交往常出现“先君子后小人”的尴尬局面。

亲情化的私人信任、算计性信任和制度信任三者之间并不是绝然分离的，在算计性信任中那些持续的长期性交往可能会建立起关系密切的信任。郑伯埙提出了“信任穿透模式”：在企业内部，企业主持人与部属之间会出现亲疏异位、忠逆异位；在企业之间，长期交易关系中的信任的形成与断裂有以下几个穿透环节：初步人际信任—经济信任—深度人际信任—义利共生。人们在交易中，互换的不只是限于经济资源，而且还包括人情面子

等社会资源。郑伯坝的这一观点是很有见地的。就制度信任而言，威廉森也认为，制度信任指的是合约被嵌入其中的社会与组织环境。这表明制度信任也会受到社会中个人化的、传统文化因素等的影响。Wong Siu-lun (1991) 在研究香港和海外华人企业时也认为，制度化信任或系统信任不可能在真空中存在，必须嵌入关系网络之中，因此制度化信任与私人信任之间是相互加强，而非相互排斥。彭泗清 (1999) 认为，在华人社会中，关系运作是建立和增强信任的重要机制。关系运作不仅包括利用关系网络或请客送礼等工具性色彩较强的方法，而且还有互相尊重、交流思想感情等情感性色彩较强的方法。在经济合作关系中，为了增强信任，人们除了进行关系运作之外，还会采用法制手段。二者可以共存。这些都表明，华人社会特别注重关系运作，因而，信任的产生、发展和变化是非常复杂的。

五、华人社会的三层信任结构与家族企业成长模型

根据上述分析，可以把华人社会的信任结构概括为三层：家族信任、泛家族信任与制度化信任。由此可建构一个信任与家族企业成长的简要模型，如图 2 所示。

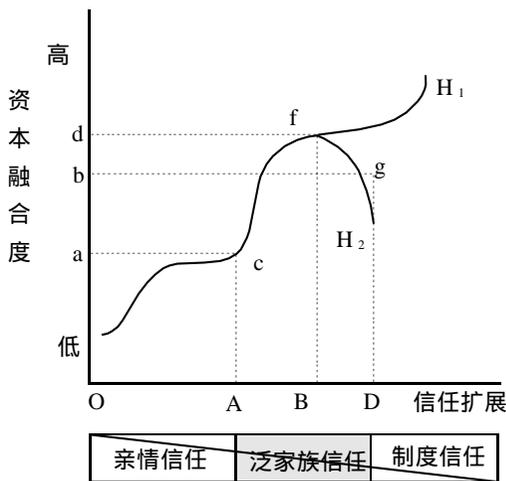


图 2 信任结构与家族企业的成长

对这一模型可作以下分析：

(1) 横轴表示信任的扩展，纵轴表示家族企业与各种社会资本的融合程度。从信任资源的分布看，从右到左，亲情信任资源渐重；从左到右，制度信任资源渐重；在 A—D 之间弥漫着泛家族信任资源。三种信任资源之间没有绝对清晰的界限，而是相互之间有程度不同的渗透。

(2) 在 O—A 区间，亲情信任非常迅速地将家庭和家族资本集成起来，强有力地支撑着家族企业的创办与初期成长；在 A—D 区间，家族企业的成长得到泛家族信任的支撑，对资本的融合突破了纯血缘亲情信任的局限。在 f 点，泛家族信任促进家族企业融合各种

社会资本达到最高点。随后，由于泛家族信任的内在矛盾，部分家族企业的信任扩展遇到障碍，融资能力下降，如曲线 H_2 所示。另一部分家族企业在 B—D 方向继续扩展信任广度，企业得到制度信任的强有力的支撑，因而对各种社会资本的融合度大大提高，企业形态也将演变成家族临界控股的大型（上市）公司或公众上市公司，职业经理将掌握经营控制权，如曲线 H_1 所示。

(3) f 点是家族企业成长的关键结点。在这个结点上，家族企业要将亲情信任和制度信任协调起来是不容易的。尽管有学者认为私人信任与制度信任能够相互容纳共存，但内在的矛盾可能会使家族企业陷入成长的困境之中。李新春（2002）将这种困境称为“家族主义困境”。他指出：在一定的制度文化形成的价值体系下，只有家族化的信任才能形成最优的合作（忠诚），但这种合作却不是最有效率的，因为最具代理能力的人才不能或难以家族化。按他的见解，家族企业能比较有效地解决降低代理成本问题，但却难以有效地激发代理能力。笔者认为，那些注重制度信任建设的家族企业主会突破家族主义的困境，从而能更多地融合社会资本。中国大陆和海外一些大型的华人家族企业的实践也证实了这一点。例如，浙江温州一些上规模、上水平的民营家族企业不仅在股权结构上，而且在治理结构上打破了家族的垄断，从而造成执行机构、董事会和股东会对家族制的全面突破（马津龙，1999）。井上隆一郎也指出，在亚洲的大型家族企业中，“在经营方面，打破同一家族经营的旧格局，起用有能力的专业经营者的倾向也广泛出现。”^①

(4) 当企业突破家族主义的困境时，得到了制度信任的支撑，能更大程度地融合社会资本。特别要指明的是，制度信任包含有法律契约的正式制度信任和非正式制度信任，如声誉机制、心理契约等。这两者是相辅相成、互相促进的，而且这里所讲的声誉机制、心理契约是由市场经济的公平公正、平等自愿等价值观规定的，而不是由传统社会中的身份面子、私忠、私人信任所规定的。还应看到，进入制度信任阶段，并不表明企业的成长就会一帆风顺，因为还有很多其他的因素会影响企业的成长，如社会政治制度、政府的法律政策、总体经济环境，特别是社会的信用制度状况等。所以上述模型是在舍弃了有关约束条件的前提下建构的。

稍微考察华人企业史就会发现：晋商和近代以来的民族企业之所以能发展到相当大的规模，主要是得到了积累深厚的传统社会信任资源的有力支撑。这种社会信任资源的特征是：在血缘基础上形成的以家族及泛家族信任为核心内容的伦理信用规则，以及以这种规则所形成地缘或业缘性的帮、会、社等社会关系网络。相对于西方社会而言，传统中国政府“不会增强契约的有效性，也不会支持信誉机构。……而这恰恰是西方市场经济的基石”（Hamilton，1991）。由此可见，中国传统的伦理信任资源的功能是非常强大的。但是，非常遗憾的是，这种伦理信任并未能支撑中国的传统企业成长为现代化大企业。除了社会政治动荡、外国列强入侵等原因外，这种突出体现了传统文化中的尊卑长幼有序的人治特征的伦理信任使企业融资的广度和效度低于与市场经济中的理性契约化的制度信任，因此，传统的华人家族企业难以向现代企业演变发展。

当代中国的家族企业处在社会转型的背景之中，传统的伦理信用规则的功能虽有一定

① 井上隆一郎：《亚洲的财阀和企业》，三联书店 1994 年版，第 49 页。

程度的重新恢复,支撑了家族企业的兴起,但在一个多世纪的动乱和反复批判中,家文化受到严重破坏,特别是连结人际交往的泛家族信任规则受到极大破坏,它在家族企业融合各个层面的社会资本中的纽带功能大大弱化,超出家庭、家族以外的通过泛家族规则来寻求企业发展的信任资源支撑的习惯做法效用递减,20世纪90年代初期以来,甚至出现了严重的“杀熟”现象。而社会转型状态,法制型的制度信用远未充分建立健全,这使得私营家族企业主一方面难以像晋商、徽商、近代民族企业主那样获得泛家族规则这种传统的伦理信任资源的有力支撑,企业成长受传统伦理信任资源弱化的制约;另一方面又受到法律制度信用严重残缺的制约。可以说,由于当代中国社会信任资源的双重残缺,使绝大多数私营企业不得不呈现为家族制形态,并难以从家族制管理向现代专业化管理转变。因此,要促进家族企业有效成长,就必须尽快解决当代中国双重信任残缺的状况。

参考文献

- [1] Penrose, Edith T., *The Theory of the Growth of the Firm*, Oxford: Basil Blackwell Publisher, 1959.
- [2] 李新春,胡骥:《企业成长的控制权约束》,《南开管理评论》,2000年第3期。
- [3] 李新春:《企业家过程与国有企业的准企业家模型》,《经济研究》,2000年第6期;《信任、忠诚与家族主义困境》,《管理世界》,2002年第6期。
- [4] 张维迎:《产权、政府与信誉》,三联书店2001年版。
- [5] 张正明:《晋商兴衰史》,山西古籍出版社1995年版。
- [6] 唐力行:《商人与中国近世社会》,中华书局1995年版。
- [7] 石骏:《汇通天下的晋商》,浙江人民出版社1997年版。
- [8] 费孝通(1948):《乡土中国》,见:《东方之子·大家丛书·费孝通卷》,华文出版社1999年版。
- [9] G. Hamilton, *Business Networks and Economic Development in East and Southeast Asia*, Centre of Asian Studies, Hong Kong University, 1991.
- [10] 雷丁(G. Redding):《海外华人企业家的管理思想——文化背景与风格》,上海三联书店1993年版。
- [11] 钱德勒:《看得见的手——美国企业的管理革命》,商务印书馆1987年版。
- [12] 储小平:《家族企业研究:一个具有现代意义的话题》,《中国社会科学》,2000年第5期。
- [13] 汪丁丁:《回顾“金融革命”》,《经济研究》,1997年第12期。
- [14] 弗兰西斯·福山:《信任——社会道德与繁荣的创造》,远方出版社1998年版。
- [15] 祖克尔(Lynne G. Zucker), *Production of Trust: Institutional sources of economic structure, 1840~1920*. *Research in Organizational Behavior*, Vol.8, pp53~111, Copyright 1986, JAI Press Inc.
- [16] 亚瑟·亨·史密斯:《中国人的性格》,学苑出版社2001年版。
- [17] 彭泗清:《信任的建立机制:关系运作与法制手段》,《社会学研究》,1999年第2期。
- [18] 奥利弗·E. 威廉森:《治理机制》,中国社会科学出版社2001年版。
- [19] Gary G. Hamilton, *Asian Business Networks*, Walter de Gruyter, Berlin, New York, 1991.
- [20] 奥利弗·威廉森(Oliver E. Williamson):《治理机制》,中国社会科学出版社2001年版,第319页。
- [21] 郑伯坝:《差序格局与华人组织行为》,《本土心理学》(台湾),1995年2月,第141~219页。
- [22] 郑伯坝,刘怡君:《义利之辨与企业间的交易历程》,《本土心理学》(台湾),1995年8月,第2~41页。

民营企业职业经理人流失原因的实证研究^{*}

——对广东民营企业职业经理人离职倾向的检验分析

张建琦 汪 凡

一、对以往研究的回顾与评价

企业的“个人所有企业——创业家式或家族式企业——经理式企业”的成长历程（钱德勒，1960），暗示着经理人的成长与企业成长的同步性。管理资源是构成企业扩张的主要约束力量（Richardson，1964）。是否拥有并留住足够的高素质的经理人是企业成长与发展的关键。目前我国民营企业管理人才严重匮乏，管理资源这种最重要的人力资本当前已经成为我国相当多民营企业成长的瓶颈（储小平，2002）。职业经理人不但难以进入民营企业，而且流失严重（张建琦，2002）。其中尤以中小民营企业为甚。本文试图对广东民营企业经理人离职的影响因素进行实证研究，对以往的理论进行验证，探索经理人离职的主要影响因素，并提出政策建议。广东是一个民营经济的大省，经理人流失问题具有一定的代表性，因此本文的研究可能会对全国民营企业和相关的研究具有一定启发意义。

离职被分为自愿离职和非自愿离职（Price，1977），而自愿离职又被分为对企业有利的良性（functional）离职和不利的非良性（dysfunctional）离职。本文的研究属于经理人的自愿非良性离职。西方对这种离职的四类影响因素进行了大量实证研究：①雇员的个人特性：年龄和任职期限、性别、家庭责任等（Farrell & Rusbult，1981）。②与工作相关的因素：工作收益和工作成本（Farrell and Rusbults，1981）。工作收益包括自主性、公正、伙伴与上司的支持等（Iverson，1996；Mueller et al.，1994），以及薪酬（Cotton and Tuttle，1986）与绩效（James R. Van Scotter，2000）。工作成本则包括工作重复性（Routilization）、角色混淆、角色冲突、工作风险等（Kahn et al.，1964）。当工作收益降低或者工作成本上升时，员工倾向于退出（Farrell and Rusbults，1981）。③员工个人倾向：员工对工作和组织的感性反应，包括工作满意度（Price & Mueller，1986）、组织融入度（Porter，et al.，1974）、退出倾向（Iverson & Roy，1994）和个人与组织价值观的匹配度（Iverson，1999）。④外部环境：工作机会（Bluedorn，1982；Mueller，1994）、非正式组织的存在（Friedman & Holtom，2002）、失业率（Blau and Kahn；1981）等。研究表明，上述因素在不同的企业和条件下作用的方向与大小不同，具体的观点我们将在本文的分析中展开。国外对雇员离职的因素的研究已经十分广泛和深入，但其通常隐含地假定：企业内部的人力资源制度和治理结构完善；雇

^{*} 本研究得到中山大学岭南学院资助。原载：《管理世界》，2003年第9期。

主和员工行为是诚信和规范的，双方都具有较高的自我约束能力；经理人的能力和价值取向对于离职是决定性的，雇主的个人素质和价值观没有明显作用。显然，这些假定对于转轨经济中的中国民营企业难以成立，因而其结论的有效性也是有限的。

根据我国特殊的经济市场法律环境，国内学者提出了民营企业经理人离职的影响因素。①雇主个人变量：价值观、信用以及对雇员以往信用的重视程度。民营企业赢得经理人忠诚，其关键在于雇主真正地尊重和关心经理人的发展（张玉波，2001）。雇主的人格缺陷与企业发展难以协调矛盾的突出也会影响经理人的退出决定，如过度造名、盲目扩张、决策独断和奢侈腐化等（吴光炳，2001）。②环境变量：法制建设滞后、企业内部治理结构的不完善和经理人职业道德缺失是雇主与职业经理人之间产生信用危机的主要原因（刘湘国，2003）；与工作相关的变量：其流动主要受到流动主体的经济目标和价值取向、体制环境和市场机制等因素的影响和制约（邓宏图，2002）。③经理人恶意退出行为发生的根源在于企业内部机制的缺陷和雇主诚信的缺乏，从而导致激励机制失灵（张建琦，2002）。④经理人个人变量：企业家追求的是在未来时间内福利最大化，而福利（收入、偏好的实现、声誉）也就成了企业家流动的重要考虑因素（周立群、邓宏图，2001）。组织因素中个人成就需求、职业发展需求和工作应激三个因素对员工离职影响最大（孙海法，2003）。

国内的研究表明，民营企业经理人离职的主要原因是制度性的。但这类研究目前仍处在定性分析阶段，充分的实证研究尚未开展，因而有必要对国内提出的影响因素的可靠性，以及国外研究的相关因素在民营企业的作用作进行进一步的实证检验。限于篇幅，本文主要研究企业内部制度因素的影响。

二、调查对象、指标选择和分析方法

我们选择调查的对象是广东省民营企业在职的经理人。我们向广东省的民营企业发出800份调查问卷，回收453份，占56.6%，其中有效问卷383份，有效率达84.5%。有效问卷中，经理人的平均工龄为4.8年，其中5年以上占总数的40.21%。有效问卷的行业分布如表1所示。

综合国内外的研究成果，我们选择了较为重要的12个企业内部因素（见表2）。首先由经理人对这些因素进行重要度评价（见表2）。其次由经理人对本企业的12项因素的满意程度进行评价。最后根据经理人的评价数值，我们选择了其中重要度最高的5个因素进行了计量分析和统计检验。

我们采用Logit模式来建立经理人的决策行为模式。假设决策者为效用最大化追求者，效用函数为线性可加成形式。在我们的研究当中，经理人只有未来离职和不离职两种选择，故我们采用二元选择的Logit模型。计量模型为：

$$\text{Logit}P_j = a_j + \sum_i \beta_{ij}x_{ij} + e_j$$

其中：P为离职的概率；a为常数项；x为解释变量； β 为待估参数；e为随机干扰项，j与i分别表示第j个样本和第i个解释变量。被解释变量是经理人退出民营企业的意愿；

退出的变量设为 1，不退出的设为 0。在 383 份的有效问卷中，有 169 人选择了未来将离职，有 214 人选择了未来不离职。由于 12 个变量间存在着多重共线性，因此我们最终只选择了表 2 中前 5 项最重要因素进行了计量检验。

表 1 有效问卷的行业分布

行 业	有效样本量	比率 (%)
农、林、牧、渔	7	1.82
制造业	148	38.64
电力、燃气及水的生产和供应业	10	2.61
建筑业	8	2.08
交通运输、仓储和邮政业	10	2.61
信息传输、计算机服务和软件业	30	7.83
批发和零售业	3	0.78
住宿和餐饮业	10	2.61
金融业	11	2.87
房地产业	18	4.70
租赁和设备服务业	40	10.44
科学研究、技术服务和地质勘查业	6	1.56
水利、环境和公共设施管理业	2	0.52
居民服务和其他服务业	18	4.70
教育	2	0.52
卫生、社会保障和社会福利业	9	2.35
文化、体育和娱乐业	24	6.26
公共管理与社会组织	10	2.61
不明	14	3.65
共计	383	100

表 2 经理人对离职影响因素的重要度评价排序

(对各因素的期望) 重要度排序	影响因素	重要度 平均得分	标准差	权重 ^①
1	企业的发展前景	4.576	3.485	0.120
2	企业的薪酬和福利水平	4.624	2.865	0.119
3	晋升与分配的公正性	4.987	2.893	0.113
4	培训和职业发展空间	6.173	3.118	0.094
5	雇主对经理人的信任与信用	6.557	3.848	0.088
6	雇主的个人品格和魅力	6.898	3.854	0.082
7	企业管理的民主性与规范性	6.939	3.161	0.081
8	雇主对社会的诚信程度	7.035	3.891	0.080
9	人际关系的复杂性	7.281	3.279	0.076
10	对公司文化和氛围的适应性	8.092	3.149	0.063
11	工作自主性	8.513	3.335	0.056
12	雇主的受教育水平	10.204	3.384	0.029

注：表中的重要度平均得分接近 1 的因素为最重要，接近 12 的为最不重要。

① 各因素权重 = (12 - 各因素的重要度平均得分) / \sum (12 - 各因素的重要度平均得分)。表示某因素在经理人对全部因素的期望满足状况中的影响程度。权重为 1% 时，当对该因素的期望获得满足时，则表明 1% 的期望获得了满足，反之亦然。

三、对调查数据的回归分析与检验

根据以往的研究，我们对调查中的 5 项最重要的影响因素提出以下假设：

假设 1：良好的企业的发展前景，预示着经理人今后的经济收入和社会地位会不断提高，施展才能的平台将得到进一步的扩大。因此企业发展前景与经理人的离职倾向呈负相关关系。

假设 2：薪酬和福利水平既是对经理人人力资本投资的补偿，也是其自身经济和社会价值的体现，同时也是支撑家庭和谋求学习发展的基础。因此当前工资福利水平与离职倾向负相关。

假设 3：经理人与雇主之间是一种长期的重复博弈关系。公平性能够稳定博弈收益的预期和变化，构成良好的交易维系机制，从而使经理人长期留在企业。同时，公平性反映企业管理的科学化和规范化程度，以及经理人的工作与发展是否具有一个公平竞争的环境。因此收入分配和晋升的公平性与离职倾向负相关。

假设 4：企业不仅是经理人的人力资本使用、发挥和实现价值的场所，更是其人力资本不断成长和提升的载体。经理人进入企业不仅仅是为了追求短期的现金收入，同时也是为了实践自己的知识，获得更多的管理能力和经验。因此学习与职业发展的机会与退出倾向是负相关的。

假设 5：研究表明，在家族经营的民营企业中，雇主对家族以外的经理人信任和信用程度不足。而雇主的信任和信用对于经理人的能力发挥具有很大的影响。因此雇主的信任程度与经理人的离职倾向之间应当具有负相关关系。

回归分析的结果如表 3 所示。

表 3 6 项最重要因素的回归分析结果

Variable	Coefficient	Std. Error	z-Statistic	Prob.
截距	1.231244	0.490318	2.511113	0.0120
薪酬和福利水平	0.288292	0.107080	2.692307	0.0071
分配和晋升的公平性	- 0.175040	0.095078	- 1.841010	0.0656
培训学习和职业发展机会	- 0.168586	0.091446	- 1.843564	0.0652
雇主对经理人的信任与信用	- 0.179758	0.090829	- 1.979075	0.0478
企业发展的前景	- 0.181007	0.087192	- 2.075955	0.0379
LR statistic (5 df)	22.72958			

大多数学者的实证结果都证明薪水与退出倾向负相关 (Cotton and Tuttle, 1986)。但令人意外的是，我们的实证结果却与假设相反，证明收入和福利水平与退出倾向正相关 (0.277)，且在 1% 的水平下显著。即经理人的收入水平越高，离职倾向越高。但是，该结

果与 Krau (1981) 年的研究结论相一致。Krau 发现 35 岁是雇员退出的“高峰期”，而这个年龄的雇员通常经验丰富，收入较高。我们的解释与 Krau 接近，即通常工作年限长、经验丰富、能力较强的经理人收入较高，这部分员工容易诱发自我创业动机，或者离职谋求更高收入的工作，因此离职倾向会较为显著。但这并不表明经理人的收入在整体上已经偏高了。

收入分配与晋升的公平性的系数为 - 0.175，且在略高于 5% 的水平下显著，支持了假设 3。说明企业内部的公平性对于经理人的离职有重要的影响。因为收入和晋升的公平是人力资本的报酬实现和职业发展提升的保证。这与国外的研究结论一致 (Mueller et al., 1994)。

学习和职业发展机会与空间变量的系数为 - 0.169，在略高于 5% 的水平下显著。表明该因素与经理人退出倾向呈负相关关系，从而支持了假设 4。表明经理人在一定经济收入和制度的公平性下会追求人力资本的升值和长远的职业发展，而且该因素是职业转换的一个重要选择标准。我们的结果支持了孙海法 (2003) 的结论。

雇主对经理人的信任与信用变量的系数为 - 0.180，并在 5% 的水平下显著，支持了我们的假设 5，即雇主信任和信用与经理人的离职倾向呈负相关关系。因为信任和信用是经理人的个人价值实现和职业发展的重要社会基础。Price & Mueller (1986) 的实证研究也证明雇主的信任和支持可显著降低雇员的退出倾向。

企业发展前景变量的系数为 - 0.181，在 5% 的水平下显著，支持了我们的假设 1，即企业的发展前景与退出倾向负相关。说明良好的发展前景是吸引经理人的一个重要因素。即使是薪酬和福利水平较差的中小民营企业，也可凭借良好的发展前景给经理人提供充分的职业发展空间，从而吸引和长期挽留经理人。

四、对职业经理人评价的分析

表 4 显示了经理人对企业内部因素的评价。表 4 中各因素的重要度评价反映出了经理人对该因素的期望，对本企业各因素满意度的评价反映了经理人对该因素的满意程度。经理人的期望和满意度的匹配状况可以分为以下三类：

(一) 满意度超过经理人期望的因素

雇主对顾客、同事、朋友、亲属和社会的诚信程度会影响到经理人对雇主和企业的信任和诚信。如果雇主对社会缺乏诚信，不仅会降低经理人对雇主的信任，甚至会诱发经理人对雇主的欺诈和恶意离职 (张建琦, 2002)。雇主对社会的诚信的重要度排在了第 8 位，但经理人对其评价却达到了第 2 位。反映出社会的信用观念和体系正在不断地完善，民营企业雇主对社会的诚信已经有了较大的改善。这将间接地提高经理人对雇主的信任，从而有助于长期留住更多的经理人。然而，在民营企业偷漏税款和弄虚作假行为还大量存在的现实面前，这一因素的重要度排位居后，同时经理人又对雇主的社会诚信给予如此高度的评价，反映出经理人对雇主的社会信用的要求不高，甚至对雇主的不诚信行为有着某种程度的“理解”和支持。因此，企业的不诚信行为在很大程度上是经理人和雇主互相纵容的

结果。这是经理人认识上的一种误区，纵容和支持雇主的不诚信行为，最终会导致雇主对经理人的不诚信（张建琦，2002）。

表 4 经理人对企业各项因素状况的评价

满意度 评价排序	影响因素	满意度 得分	标准差	重要度 排序
1	当前所在企业的发展前景	3.649	1.361	1
2	雇主对社会的诚信	3.489	1.229	8
3	晋升和分配的公正性	3.478	1.251	3
4	雇主个人的个人品格和魅力	3.377	1.231	6
5	对公司文化和氛围的适应程度	3.295	1.095	10
6	培训学习和职业发展的机会与空间	3.243	1.238	4
7	工作的自主性	3.198	1.293	11
8	雇主对家族外经理人的信任与信用	3.183	1.281	5
9	对薪酬和福利水平的满意程度	3.149	1.095	2
10	人际关系的复杂程度	3.146	1.300	9
11	企业管理的民主性	3.052	1.100	7
12	雇主受教育水平	2.242	0.997	12

注：①本表采用 Likert-type scale 五点量表计分法，按调查因素的程度分别赋予数值 1, 2, 3, 4, 5。②接近 1 分者表示水平最低，接近 5 分者表示水平最高。

作为企业领导者的雇主如果具有高尚的品格与修养，则有可能使经理人产生较高的向心力，即使在企业困难的情况下也能凝聚经理人共同奋斗。雇主的个人品格和魅力在重要度中位居第 6，但对现实状况的评价得分位居第 4。这一方面说明多数民营企业的领导人是经过商界风雨考验的“能人”，在经理人中具有较高的威望，其素养也有所改善，同时表明经理人对雇主的行为和素养具有较高的理解和接受程度。另一方面也反映出经理人目前对雇主的个人品格和魅力的要求不高。因为现实中毕竟有不少民营企业的雇主作风粗俗，格调低下甚至生活糜烂（吴光炳，2001）。尽管在现阶段经理人的地位和能力尚不足以阻止这种倾向，但经理人对雇主个人的不良品格的默许和宽容，有损于企业文化的改善。

企业文化和组织氛围是雇主和企业领导人价值观的反映。如果经理人对这些明文或默许的要求不能认同，就无法在企业生存和发展。对公司文化和氛围的适应程度的重要度评价位居第 10，而对现实状况的评价位居第 5。表明目前经理人在很大程度上能够接受民营企业的组织文化和氛围。

工作的自主性是经理人能力发挥的前提，自主性越高，表示雇主对经理人的信任程度越高，经理人发展的空间越大，从而企业对经理人的吸引力就越大。工作的自主性的重要度评价居 11 位，但经理人的满意度居第 7 位。表明即使在家族式经营和“一言堂”的领导 and 决策体制中，由于雇主的管理水平不高以及管理制度的不完善，雇主仍需给予经理人一定的授权，因此经理人的能力仍有一定的发挥余地和创新空间。但同时也说明目前经理

人在很大程度上能够接受民营企业这种缺乏员工参与的家长制的管理模式。

对以上两个因素的分析说明,由于经理人的接收和容忍的程度较高,家族经营和家长制的领导模式在民营企业中还有相当的生存空间,不会在短期内发生根本的变化。

(二) 满意度与经理人期望相匹配的因素

令人兴奋的是,留住经理人最重要的因素(重要度 4.500)——所在企业的发展前景在所有因素中获得了最高的评价(3.878),它反映出民营企业目前总体发展的状况和前景良好,给了经理人很大的希望,这是留住经理人最大的吸引力。同时也显示出经理人已经把企业的长远发展放在了第一位,这将对民营企业的发展有更大的促进作用。但该因素评价的标准差在所有因素中也是最高的(1.375),表明民营企业的发展存在着不平衡性。

晋升和分配的公平性的重要度排在第3位,而经理人对其评价也排在了第3位。说明由于存在效率的压力,民营企业的收入分配和晋升是严格按照工作能力和绩效来决定的,而且这方面的制度正在向着规范化的方向不断完善,基本达到经理人的满意程度。

一般而言,雇主的受教育水平越高,其修养和领导水平就越高,企业的凝聚力也越强。雇主的受教育水平的重要度和满意度评价都位居12。说明民营企业雇主的受教育水平普遍不高。但经理人对雇主的受教育水平并不苛求,对这种状况可以接受。因此,雇主不必急于搞假文凭,更重要的是加强自身的学习,并且把企业变成学习型组织。

(三) 满意度低于经理人期望的因素

学习和职业发展的机会的重要度排在第4位,而满意度位居第6。反映出民营企业更重视经理人的使用,而较少提供培训学习的机会,缺乏系统的职业发展计划,使得经理人只有人力资本的输出,而缺乏人力资本的输入,难以得到成长和发展。经理人进入民营企业不仅仅是为了经济利益,同时也是为了获得更多的管理能力和经验,否则即使收入较高,也难以留住经理人。

对雇主对家族外经理人的信任和信用程度的期望评价位居第5,而满意度排位第8。显示出多数民营企业是家族经营模式。家族以外的经理人得到的信任依然十分有限,其能力的发挥和经济利益的实现受到较大制约。

差距最大的是收入与福利水平。对薪酬和福利水平的期望排位第2,满意度位居第9。显示经理人对近期的现金收入高度重视,同时反映出民营企业的经理人的总体收入不高,与经理人的期望反差较大。除工龄长和职务高的经理人外,中下层的经理人的收入偏低。加之现阶段民营企业普遍缺乏长期战略,经营不稳定,寿命周期短,因此追求近期收入也是对这种状况的一种理性反应。在经济收入上,民营企业对经理人的凝聚力不强。

人际关系的复杂程度的重要度位居第9,满意度评价位居第10。表明民营企业的家族经营模式带来了较多的裙带关系,从而使经理人无法按制度和规范实施自己的管理职能,才能难以发挥,受信任感下降。但期望与满意程度差距不大,说明经理人对民营企业的人际关系仍有一定的容忍空间,不会对经理人造成严重的排斥。

企业管理的民主性的重要度排位第7,但满意度评价位居第11。离经理人的要求差距较大。民营企业的家族制经营模式必然带来的家长制的领导模式和决策的“一言堂”,导

致企业经营管理存在着很大的主观随意性和盲目性,无法在科学规范的制度框架下运行,经理人难以参与管理与决策,作用无法发挥。评价的差距较大显示经理人对这一点反应强烈。

在重要度评价中,前4位最重要的因素都是与经理人个人发展和收入直接相关的因素,后8位都是雇主和组织的因素。而在满意的7项因素中,5项是雇主个人和组织因素;在不满意的5项因素中,3项是与经理人个人收入和发展直接相关的因素。从上述分析和重要度评价可以看出,经理人在现阶段并不十分重视民营企业雇主的个人素质和企业组织的状态改善,而更关注自身的发展。

在所有因素中,有4项的满意度超出经理人的期望,有3项达到了经理人的期望,有5项低于经理人的期望。5项最重要的因素中,有3项没有达到经理人的期望。达到和超过经理人期望的因素权重之和为0.542,而低于经理人期望的因素的权重之和为0.458。表明民营企业内部因素对经理人的凝聚力只是略大于排斥力。因此,民营企业在防止经理人流失方面还面临着较大的挑战。

五、结论与政策建议

研究结果表明,影响经理人离职的5个最重要因素依次为:企业的发展前景,工资福利水平,收入与晋升的公平性,职业发展的机会,雇主对经理人的信任与信用,它们与民营企业的经理人的流失都显著相关。在所考察的因素中,有58%达到了经理人的期望,42%低于经理人的期望。民营企业内部因素对经理人的凝聚力只是略大于排斥力。民营企业面临的人才形势依然严峻。

对雇主的个人因素,包括雇主对社会诚信、个人品格和雇主的受教育水平,经理人的要求不高。对与工作相关的因素,包括家族经营模式带来的裙带关系、家族式的企业文化和组织氛围,经理人显示出了相当的容忍度,反映出家族经营和家长制领导模式还有一定的生存空间。但对雇主提高对经理人的信任与信用程度和参与管理,经理人的要求强烈。在自身因素方面,经理人高度重视企业的前景、现金收入、自身的职业发展机会,显示出经理人在现阶段并不十分重视民营企业雇主的个人素质和企业组织的状态改善,而更关心自身发展。

民营企业要提高对经理人的凝聚力,首先需要注重三个因素:①雇主要具有长期稳定发展的战略观念,使经理人具有安全感和对未来充满新信心;②改善经理人的收入和福利,对不同层次的经理人采取不同的分配方式,支持中低层经理人的发展,实行分配方式多元化,留住高层和有经验的经理人;③完善职业发展制度,为经理人提供更多的学习和发展机会,在完善监督制度的前提下,对经理人充分授权,提高经理人的被信任感。由于这些因素在不满意因素中的权重最高,因而改进边际效益将最为显著。

如前所述,上述问题并不完全是由雇主单方面造成的,因此经理人不能对企业人才制度的改善坐视等闲,而需要通过管理参与,积极地影响雇主的价值观和企业文化及组织氛围,促进企业制度的改善,为自己和企业的长远发展创造一个良好的环境。

参考文献

- [1] Aaron Cohen, Turnover among professionals: A longitudinal study of American layers, *Human Resource Management*, 1999, Vol.38, pp 61 ~ 75.
- [2] Allen C. Bluedorn, A Taxonomy of Turnover, *Academy of Management Review*, 1978, pp647 ~ 650.
- [3] Blau, F. D., & Kahn, L. M., Race and sex differences in quits by young workers, *Industrial and Labor Relations Review*, 1981, Vol.34, pp 563 ~ 577.
- [4] Cotton, J. L. & Tuttle, J. M., Employee turnover: A meta-analysis and review with implications for research, *Academy of Management Review*, 1986, Vol.11, pp55 ~ 70.
- [5] Farrell, D., & Rusbult, C. E., Exchange variables as predictors of job satisfaction, job commitment, and turnover: The impact of rewards, costs, alternatives, and investments, *Organizational behavior and human performance*, 1981, Vol.28, pp 78 ~ 95.
- [6] Hom P. W., Griffith R. W., A structural equation model of the process of employee turnover, *Journal of Applied Psychology*, 1991, Vol.76, pp 350 ~ 366.
- [7] Krau, E., Turnover analysis and prediction from a career developmental point of view, *Personnel Psychology*, 1981, Vol.34, pp 771 ~ 790.
- [8] Penrose, Edith T., the theory of the Growth of the Firm, Basil Blackwell Publisher, Oxford, 1959.
- [9] Porter, L. W., Steers, R. M., Mowday, R. T., & Boulian, P. V., Organizational commitment, job satisfaction and turnover among psychiatric technicians, *Journal of Applied Psychology*, 1974, Vol.59, pp603 ~ 609.
- [10] Price, J. L., The study of turnover, Ames, IA: Iowa State University Press, 1977.
- [11] Price, J. L. & Mueller, C. W., Professional turnover: The case of nurses, New York: SP Medical and Scientific, 1981.
- [12] Price, J. L. & Mueller, C. W., Absenteeism and turnover of hospital employees, Greenwich, CT: JAI Press, 1986.
- [13] Raymond A.Friedman and Brooks Holtom, The effects of network groups on minority employee turnover intentions, *Human Resource Management*, 2002, Vol.41, pp405 ~ 421.
- [14] Richardson, G., The Limit to a Firm's Rate of Growth, *Oxford Economic Papers*, 1964.
- [15] Roderick D. Iverson, An event history analysis of employee turnover: The case of hospital employees in Australia, *Human Resource Management Review*, 1999, pp397 ~ 418.
- [16] 储小平:《职业经理与家族企业的成长》,《管理世界》,2002年第4期。
- [17] 储小平,李怀祖:《信任与家族企业的成长》,《管理世界》,2003年第6期。
- [18] 邓宏图:《企业家流动的博弈模型:经济含义和企业家能力配置》,《经济科学》,2002年第3期。
- [19] 邓宏图,周立群:《经理人市场:供求与交易关系研究》,《江苏社会科学》,2001年第4期。
- [20] 李新春:《经理人市场的失灵与家族企业治理》,《管理世界》,2003年第4期。
- [21] 刘湘国:《民企老板与职业经理人信用危机探析》,《嘉兴学院学报》,2003年第15卷第1期。
- [22] 钱德勒:《看得见的手——美国企业的管理革命》,商务印书馆1987年版。
- [23] 孙海法:《领导策略与团队管理》,中山大学出版社2003年版。
- [24] 吴光炳:《民营企业家人格障碍及克服》,《福建论坛》,2001年第220期。
- [25] 张建琦:《经理人“背叛”的机理与雇主的对策取向》,《管理世界》,2002年第5期。
- [26] 张维迎:《家族企业的成长与职业经理人》,《中国工商》,2001年第10期。
- [27] 张玉波:《民营企业如何赢得忠诚经理人》,《企业人力资源管理》,2001年第10期。

经济交易规则发生机制的信息流程解构^{*}

杨永福 段红涛 张智革

一、引言

“创业”自古中外有之。我们要全面建设小康社会，要提高全社会的文明水准，“创业”将成为我们社会的中心话题。长期以来，人们一直将创业的视野聚焦于物产、货币等一类有形资源，视之为事业成功的要件。但在信息成为战略资源的知识经济时代，《信息空间》一书的作者布瓦索（Boisot, Max H.）认为：经济过程既不是使用价值的流通过程，也不是价值流通过程，而是信息流通过程。信息流转，要求人类系统，也就是它的文化、制度环境与之配合。信息流转的本质，是在人类系统与环境系统的对话交流中，有序程度不断发生变化的过程^[1 2]。布瓦索的观点是对当前我国强调的“以人为本、统筹发展”科学发展观的佐证。

本文将沿着信息演化的路径，即从数据上升为公共知识的路径，对交易规则发生机制进行基于信息流程的解构性分析。从规范的角度说，数据和信息的基本特征就是：未扩散 - 扩散、具体 - 抽象、未编码 - 编码及其组合。在认识论范畴，人的活动就是由“未编码”走向“编码”等，即由波普（Popper, Karl R.）的“世界 1”上升到“世界 2 和 3”，如此往返等等。

交易规则机制是一组由处于不同状态的数据、信息和知识构成的空间。交易规则的差异便简约为它们在状态空间的位置差异。

二、交易规则机制的空间组态及其整合

根据发生认识论原理，依据历史与逻辑统一的方法，以及数据生成信息、信息构成知识的过程，我们构造以下几类“知识”空间。

1. “知识”空间 A

这是由编码和抽象构成的二维空间，如图 1^① 所示。

^{*} 本文得到中山大学岭南学院科研基金资助。原载：《管理世界》，2004 年第 4 期。

^① 图 1、图 2、图 3、图 5 根据布瓦索（Boisot, Max H.）《信息空间》一书中的有关资料绘制。

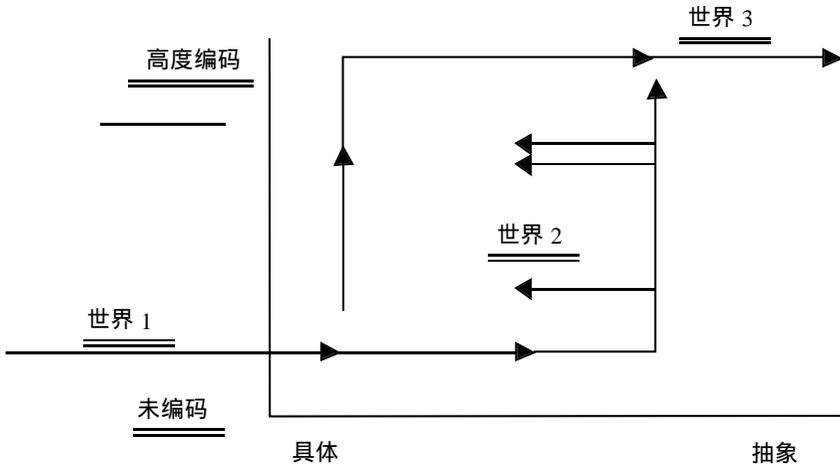


图1 “知识”空间 A

当数据由世界 1 进入世界 2 以后，它勾画出一条缓慢的近似对角线的路径，走向空间 A 的右上角，并以一种抽象的和具有“生命”的对象物而进入世界 3。世界 3 中的对象物对新的来自世界 1 的输入进行编码，赋予它们以形式和意义。与未被世界 3 调和之前相比较，现在的输入能够在较高的水平上进入世界 2。如技术熟练的员工和“生手”的区别，就在于前者已经掌握大量的技艺，而后者则刚刚由世界 1 进入世界 2，接受有关的训练。

个体的空间 A 混合了他的经历和作为文化传统的输入物。世界 1, 2 和 3 是互相渗透、相互作用的。世界 2 是意识世界，它的存在依赖能量状态的变化，世界 3 则依赖世界 2 的创造。但世界 2 不能归入世界 1，世界 3 亦不能简约为世界 1。

如果世界 3 中的某物被认为具有足够的价值，或者是经过同世界 3 中其他对象的竞争后得以幸存，那么它可能以各种方式被外部化而存在，成为人工制品。人工制品是人类社会历史和思想及文化的储存器。它们包含着丰富的知识和信息。这些知识和信息通过各种解读、诠释成为负载特定意义的数据，进入其他人的空间 A 而使自身得到扩展，产生延伸。

一般来说，生物体在其生存阶段的经历并不能在生物体内传给下一代。但在人类社会，人的经历，比如他们的知识、经验、技能以及从祖辈那里接受下来的传说、风俗、禁忌等，可以通过文字、身传言教、仪式、社会活动和演示等各种方式传给下一代。最近，受到理论界关注的拟子学 (memetics) 在这方面有深入的研究。它表明：文化发展与社会进化是拉马克式 (Lamarckian) 的而非达尔文式 (Darwinian) 的。

2. “知识”空间 B

知识空间 B 是将信息的传播和它的抽象程度相联系的空间，形如图 2 中带箭头的曲线所示。信息理论证明：抽象与扩散构成正相关关系，随着扩散程度的增加，抽象程度亦逐步提高，其一般形式为： $f_a(\dots) = kF_d(\dots)$, ($k > 0$)， f_a 表示抽象程度， F_d 表示扩散能力。这是一个惟像模型。宗教、神话、某些科学理论能够流传数百年、数千年而不衰竭，很大程度上就在于它们高度的抽象性。

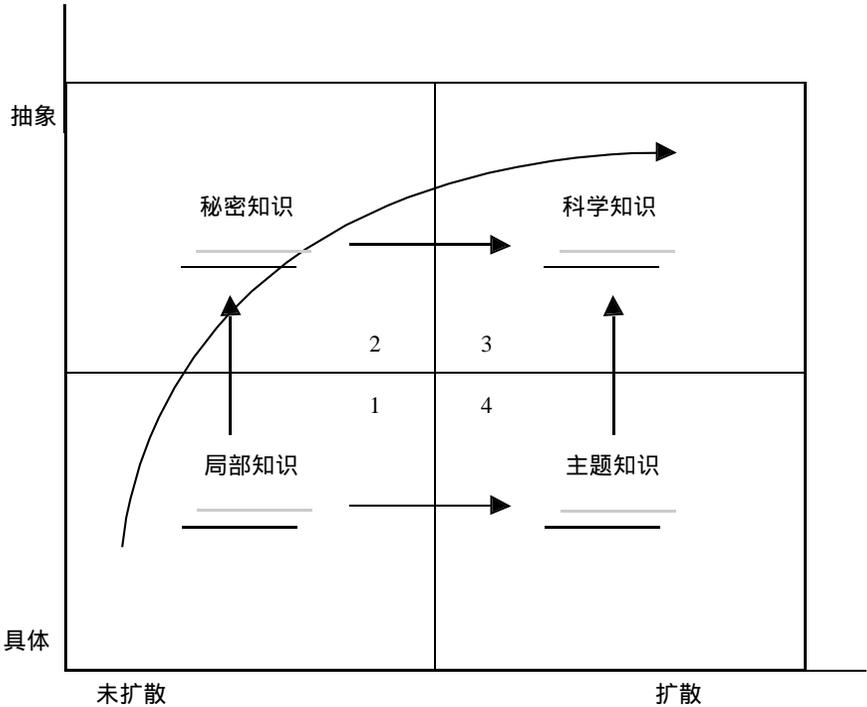


图2 “知识”空间 B

在图2中，位于象限1的是未扩散的局部知识，即哈耶克所说的特定时间和地点的知识。显然，它们是一些具体知识。

在象限2的是秘密知识，是个体或集团的专用知识，在性质上颇像法律。这一类知识未必抽象，但一般都是未扩散的。比如，史书记载，毕达哥拉斯（Pythagoras）曾秘密结社，社团里有男有女，地位一律平等，一切财产都归公有，社团的组织纪律很严密，加入组织要经历一系列神秘的仪式，遵守很多的规范和戒律等等。

象限3是一类已经扩散而且抽象的科学知识，比如水的沸点、汽油易燃、火药会爆炸、金子是贵金属、价格波动会引起供需变化等。

位于象限4的是已经扩散的具体知识，即主题知识，比如新闻、传闻、趣闻、闲话、逸事等。主题知识产生传说、神话和历史故事等等。

3. “知识”空间 C

知识空间 C 是将信息的编码和它的扩散程度相联系的空间，形如图3中带箭头的曲线所示。

知识空间 C 也涉及4类知识，但它们与空间 B 中的知识不同。

公共知识：这是一类在社会里最普遍传播的非个体化的知识，是经过世代的人们构建的经过生活检验的知识。从图3看出，在时间序列中，公共知识可由其他类型的知识转化而来。比如，群体中的个体不仅知道某个事实，而且每个个体知道该群体的其他人知道这

个事实。但作为公共舆论，这个人所共知的“事实”仍然处于默契的、心照不宣的状态，是一种前公共知识（pre-public knowledge）。一旦有某个契机，它即变成可公开“谈论”的而不是“私下里”的公共知识，变成诺斯（North, Douglass C.）意义上的正式规则。

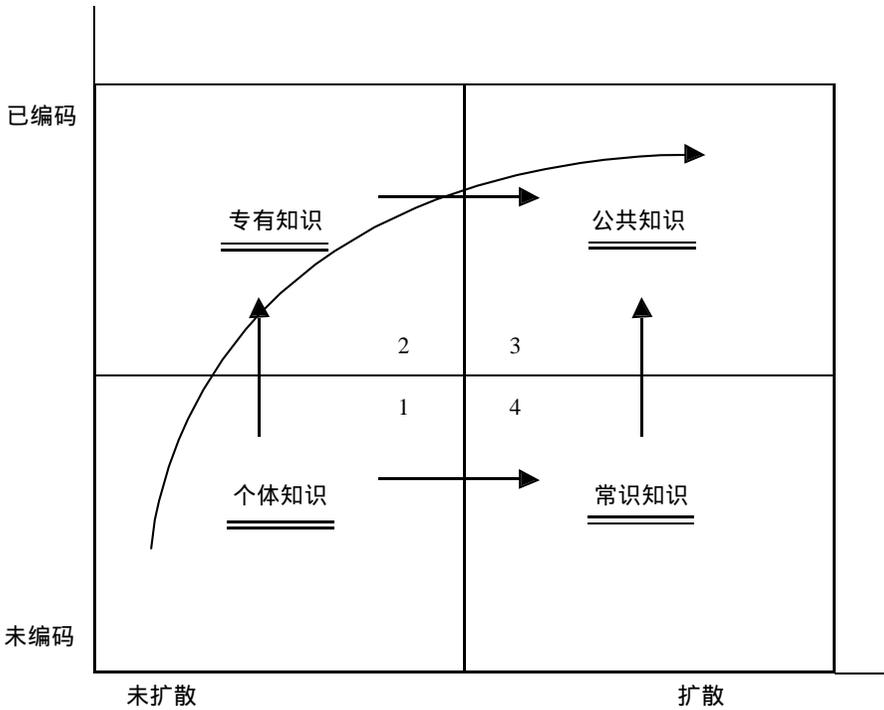


图3 “知识”空间 C

常识：它是一种强烈依赖于行为主体的社会和生活背景的知识。对于常识，爱因斯坦（Einstein, Albert）的见解很有趣，他认为，常识（common sense）是18岁前所积累的偏见。至于常识的作用，布迪厄（Bourdieu, Pierre）认为：它们（ordinary-scholarly common senses）往往嵌入在制度之中，从而既体现在社会组织的客观性上，又反映在社会组织参与者的思想里，形成所谓社会的前建构性之物（social pre-constructions）。

个体知识：对于个体知识（personal knowledge），波兰尼（Polanyi, M.）认为，它既非主观，也非客观。就个体服从于他所承认的独立于他的外在要求而言，它不是主观的；但就其行动受个体激情的支配而言，它也不是客观的。它超越了主观和客观的选言命题（disjunction）。

个体知识中包含着一类重要的知识，就是默会知识（tacit dimension），它是一种“只可意会不可言传”的知识，或者如波兰尼（Polanyi, M.）所说的“我们所知道的要比我们所能言传的多”中的“所知道”减去“能言传”的差额。而一经“言语”说出的可能往往又不是所欲说的，处于词不达意的“尴尬”之境。所以，叔本华（Arthur Schopenhauer）不无悲怆地说：思想在语言中表达的时候死去。

专有知识：这是一种可明确表达且归属明确的知识，具有一定的稀缺性。但若进入运

用环节，则应取得一定的使用权限，所以称为专有知识（proprietary knowledge）。

4. “知识”空间 A、B、C 的整合状态

三类“知识”空间表现了数据、信息和知识在由具体、未扩散和未编码状态演变成抽象、扩散和编码状态的不同侧面。它们的整合可以反映个体和群体的学习过程以及知识的发展，进而展示创业环境中交易规则机制形成的复杂和多样性。从图 4 可见，个体和群体在社会性的活动中学习知识的过程大致是这样的：由 D_i 向 B_e 方向运动，即个体或者群体从社会的知识发布和扩散活动中，如各种出版物、教育与教学机构等，获取知识，进而变成竞争力或者一种优势。这是知识内化或者收敛的过程，与社会性的知识扩散过程方向相反。在由 D_i 向 B_e 运动的同时，叠加着向 A_b 的运动，如图 4 中的 DA 曲线所示。这就是所谓边干边学的过程。在这个过程中，既产生了个体和群体的许多默会知识，比如不同的心智活动和对知识对象的不同理解；也有大量的默会知识转化为明晰知识，比如创新过程及其成果等。

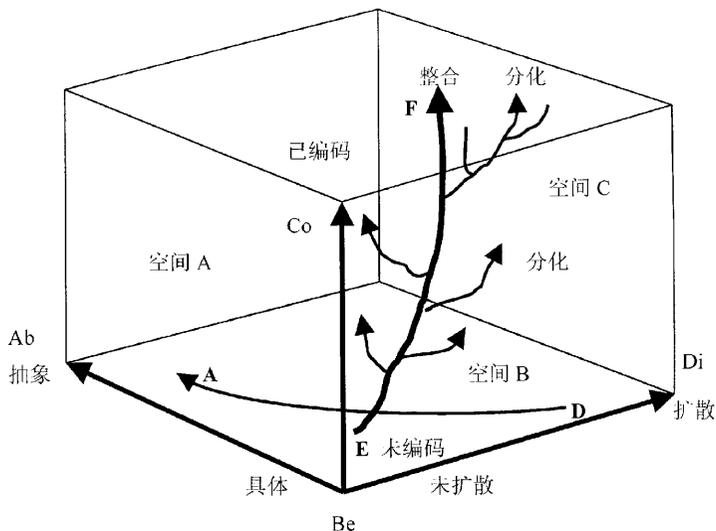


图 4 三类“知识”空间整合

EF 曲线表明，在社会 - 历史层面上，知识的发展表现为两重性：分化与整合并存。整合就是随着知识的大规模的世俗化扩散，知识的编码程度和抽象程度逐步提高；而分化则指知识在量与质两个维度上的多样性，并出现相互渗透，形成知识网络。

三、交易规则发生机制解构

根据上面的分析，下面我们将讨论 4 种经济活动中常见的交易规则的发生机制。它们在整合状态空间中的分布如图 5 所示。

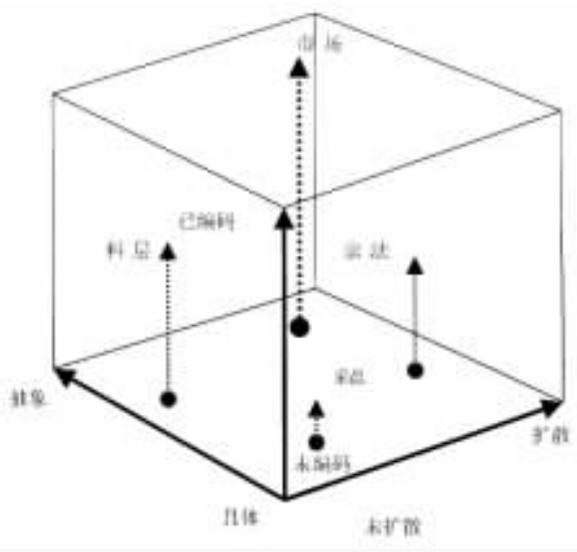


图5 交易规则类型示意

（一）市场交易规则机制

在完全竞争的市场环境中或者在非个性化的社会交换中，人们普遍关心的是货物的价格、内容和数量，本质上并不关心交易对方的身份；极端的情况下，甚至不需开口说话。你可能注意到，在销售晚报、电视报或者另类卡通漫画的摊位前，买卖双方很少开口说话。人们认为这是信息高度“编码”化的结果，“编码”化的信息确定了买卖双方预设的（presupposing）规则和秩序。经济学对这种交易类型的研究广泛而深入。

众所周知，市场的基本特点就是进入和退出交易不受阻碍，而且市场价格应当反映供求的真实状况。但在市场的不同阶段，情况并非总是如此。因为任何一种交易，在其起始阶段，参与交易者的规模是一个很重要的参数。一般来说，参与交易的人数越少，就越有可能使这些参与者的身份以及他们的相对规模变得重要。他们的行为在事实上表明他们知道而别人不知道的有关交易的信息。相对于市场中其他信息不灵通的参与者，他们因而就可以采取“先声夺人”的战略行动，像确定交易程序、交易规则、率先进入利润丰厚“地带”等。这是“自发规则”和“自发秩序”的源头。随着交易的不断发展和规模扩大到“圈子”之外，这时的市场才受到“看不见的手”的引导而保护交易者进出市场不受限制。

布瓦索（Boisot, Max H.）认为，如果小数量条件本身持续下去，并且交易者的规模逐步缩小，交易的类型有可能要么走向信息空间的科层制度，要么走向宗法制度^[3]。这是市场失灵后的两种退出机制。在与市场相关的信息没有完全被价格抽象、编码和捕捉的地方，交易更可能走向宗法制度而不是科层制度。有些更为地方性的与交易有关的未编码信息只能传递给很接近的人，像同事、朋友、亲戚或者“圈子内的人”。在交易上，这些人比“圈子外的人”具有信息优势：他们得到的是一手的甚至是机密的“信息”，而“圈子

外的人”得到的肯定是“二手”或者是来自传媒的信息。

对此，布瓦索接着指出，市场在大多数文化中和在大多数历史时期以这种或者那种形式存在着，但总是在其他制度的辅助下发挥作用^[2.4]。尽管在自由理论的推动下，市场被提升到社会秩序的支配性地位，但它却不能独立于国家权力的支撑和保障。沃恩(Vaughn, Karen I.)在为《新帕尔格雷夫经济学大辞典》撰写的“看不见的手”的词条中认为，在一般均衡的表述中，自利心被解释为对于一切商品（但不是慈善心或荣誉感）的偏好秩序，而社会的政治体制和社会体制是可以通过政府的纠正行动去改变的^[5]。根据这种看法，看不见的手仍然会使制度运转，但不一定能保证结果是最佳的。因而，如果人们遵循这种论点的逻辑，那末看不见的手从最好的方面说也是瘫痪的。因为它实际上只能在现实世界不可能得到满足的条件下，才能起产生善果的作用。看不见的手的另外一种表述是把一个社会的经济体制看成是自利的经济行为的副产品，把这些体制看成是对于自行建立秩序的过程所不可缺少的。这一点，连同这类体制服务于比计划制定人所知道的更多的见解，使得人们在解决市场安排时更加谨慎。

（二）科层制度交易规则机制

布瓦索认为，科层制度的运转依赖编码良好且大多为抽象的信息的流动，比如，财政预算、月报、经济评估、统计调查、各种“红头”文件等^[2.6]。这些信息与它们在市场中所起的作用完全不同，编码并不能完全抓住与科层交易有关的一切信息，而且这些信息也不能为所有的交易者得到。它们被数量有限的个体通过授权而获得，信息的扩散受到有组织的控制。

信息理论认为，经过编码和抽象的信息具有自然的扩散趋势。所以，处于整合状态空间科层制度区域的信息就具有内在的不稳定性。而为了阻止信息扩散就必须设置保障措施。这就是世界上所有的科层政府为了防止他们的“机密”泄露而制定的繁文缛节的保密规则以及为执行这些规则而设置的叠床架屋般的机构的原因之所在。正是这些保障的存在和作用，使得科层交易具有了特殊的格调。在整合空间中，这一区域内交易的制度化，其目的就在于确保个体拥有有价值的信息以及对信息的使用在社会意义上被赋予合法权益。因此，权力与利益融为一体。

在科层制度下，个体不能像在市场环境下那样自由地追求一己的个体目标。目标来自组织的规定。只有服膺此一目标，个体才享有相应的权力，所以“权力”和“利益”都应归组织所有。但对于可交易信息的占有和使用的不可分割性，经常使个体很容易抓住这些权利和可乘之机为制度上并不合法的目的服务。因此，这些人的行动必须受到约束。其办法：要么是建立旨在使他们的行为曝光的外部控制系统，要么是建立起内部化的具有调节功能的个体价值系统。在处理得到良好编码、建立在抽象原则基础之上，且可记录在案和公布于众的交易时，前一种办法较可取。而在交易中所用的信息是具体的、不能很好编码的地方，只有那些熟悉所要求的价值系统、表现出对这种价值系统有可信的承诺的人才能获得拥有这种信息的委任。这种情况不再是纯粹的科层交易，它属于整合空间的采邑领域。这是一种基于权威而不是威权的正式的专门通讯网络。所以，科层制度下，权力受制于结构性规则。

在此有必要提及韦伯的科层制度观点。韦伯 (Weber, Max) 认为任何一种合乎需要的统治都有其合理性 (rationality) 或者合法性 (legitimacy) 基础。科层制是特定权力的施用和服从关系的体现, 自愿服从又是以形成个体价值氛围的信仰体系为基础的^[7]。个体需认同信仰体系, 才能取得行动的一致性、连续性而不致导致内心的紧张, 并最终获得自愿的服从。韦伯把个体自愿服从的体系视为合理性或合法性体系, 即合理性并不表现在事实的好坏之分, 而在于它是否被人们在信仰上认可。他在逻辑上划分出两种合理性: 形式合理性和实质合理性。形式合理性是指在统治关系中, 行动方式倾向于在其手段和程序等方面尽可能地加以量化, 从而使得行动本身以及对行动后果的预测成为可以被计算的任务。这是一种纯粹客观的合理性。实质合理性则完全基于价值判断的基础。它对行动的目的和后果进行价值评价, 如是否合乎宗教信仰或宗教教义、是否符合习惯、是否表现出某种社会美德或善行等。它仅对行动实现价值判断, 极力强调行动的社会关注, 忽视行动效率, 是一种主观合理性。而形式合理性把行动过程和目的本身都看成是可计算的, 或者说使它具有程序化的可计算性, 目的在于社会秩序的理性化。现代社会日趋繁复的生产与生活, 要求把行动的效率提到重要的位置。由于受市场经济法则的支配, 公司不得不连续而精确地并尽可能以更好的成本 - 效益比和更快的速度处理它的业务; 现代国家治理需采用科层化的管理方式, 使它的军事、司法和行政管理人员普遍地雇员化; 在公共生活领域中, 报刊等社会舆论不仅在内部形成了分工明确、运作有序的机制, 而且在外部和科层化的国家相互信赖, 并受到经过专业训练的各类活动家或党派官员的指导等等。

在正当性观念的支持之下, 任何来自权威的命令都会得到个体的遵从, 而不论这些命令是否来自统治者个体, 或通过契约、协议产生的抽象法律条文、规章等命令形式。

韦伯认为, 现代科层化制度由于其明确的技术化、理性化和非人格化而表现出它的合理性: ①现代科层化制度表现为一整套持续一致的程序化的命令 - 服从关系, 是法律化的等级制度。②从属关系一般是由严格的职务或任务等级序列先在地安排的。权力矩阵并不反映在权力的个性特点方面, 而是基于职务本身的组织构造。在科层体制内部, 每一个个体单元被分割成各自独立的部分, 并且要求完全排除个体的情感纠葛。对权力义务体系的规定细致而明晰, 使得每个个体都能够照章办事而不致越出权力义务体系范围, 不允许科层个体随意扩大其行动的范围并表现出“能动性”。就是说, 个体在科层体制中已经被物化与原子化了。③现代科层的非人格倾向。由于权力是源于建立在实践理性基础上的形式法学理论和形式法律规定的制度, 科层运作的主要指标是可操作性与效率, 实证主义甚至功利主义就大占上风, 个体的性格和意志在这里难以有所作为, 人身依附关系因为科层内部流动的物化标准与程序化而弱化乃至消失了。个体和国家财产的实际分离以及个体权力资源与管理手段的分离导致个性化权威的虚拟化。官员的体制内流动取决于制度所规定的行动的程序化、客观化以及他的年资、工作经验、责任心和敬业精神等。个体的服从对象不再是拥有特定职务的个体, 而是个体拥有的特定职务, 他为客观的、个体的组织和组织目标服务。④现代科层的技术化倾向。在技术化导向的支配下, 现代科层倚重各类专家, 在管理的方法和途径上也越来越科学化、合理化, 组织行为的科学化业已成为各类科层的共识。

市场化规则旨在处理大量编码为价格和数量的抽象数据。但科层化制度要常常处理不

那么容易标准化的数据。这些非标准化的数据中不可避免地有一部分要超出“常规”，而成为泰勒（Taylor, Frederick Winslow）所说的“例外”。这些“例外”将按组织规则上交，流向科层等级的上层，直到它们找到一种处理程序——这种程序可能成为常规，也可能成为“例外”的例外。

“例外”的向上传递表明，科层化制度的规则，使模糊的、不可分析和非常规的决定不可避免地要由最高一级首长或者“一把手”来做最后抉择。因此，科层化制度蕴涵着向采邑制度演变因素。

（三）宗法制度交易规则机制

当交易各方之间的信息不对称造成小数量的状况以及交换过程的自我调节功能不再有效时，就会出现市场失灵。处于此种状态的市场会出现各种走势，如停滞、走向管制、重新组合或者分叉（bifurcation），以取代或者补充市场过程。未编码和具体的信息所带来的传播限制会形成小规模参与者的市场格局。布瓦索将这种维持在面对面交往关系的小团体称为带宗法色彩的群体。他们代表有限的信息扩散的状况^[8]。在这种情况下，一群“内部人”集体地享有信息优势，而“外人”是享受不到的。这就是宗法制度型交易市场。在这里，所谓宗法制度（clan system），是指无等级的、有限规模的群体，相互间的交易是在共同的知识价值基础上进行的，如具有大致相同的生活背景、共同的信仰以及对于同一种“图腾”（Totem）的崇拜等。宗法制度型交易把局部的、特殊的、影响日常人们生活的数据或如哈耶克（Hayek, Friedrich A.）所说的“局部知识”（local knowledge）看作是关键性的。这种“局部知识”构成两个极端情境之间的一系列情景交易（context bargaining）的基础：一个是高情景交易（high context bargaining），另一个是低情景交易（low context bargaining），如图6所示。高情境交易涉及丰富、具体和未编码的面对面交流的数据，其特点是复杂、多样且微妙；而低情境交易则在清晰、简单和客观的情况下，把编码的抽象数据的选择使用作为交易的窗口。

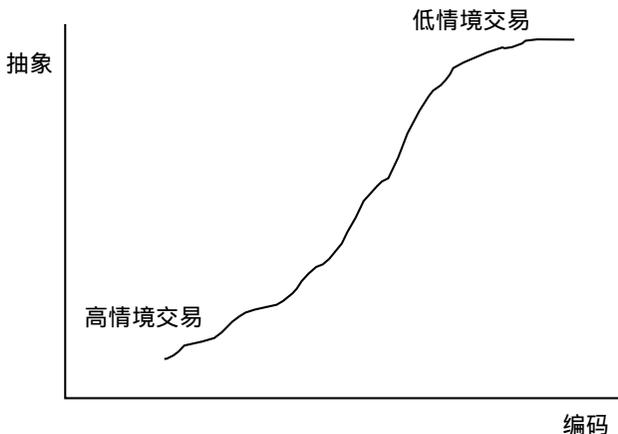


图6 高-低情境交易示意

宗法制度型交易是一种基于人际信任，以及共同道德、共同价值的交易，交易者处于一种集体学习的环境中。因此，人们相互之间建立信任的速度要比那些缺少这种基础的人们来得快，传递信息和交易的过程也会更加有效。老同学关系、某种口音或用词、穿着的特殊方式等都可能是已经“嵌入”（implanted）的社会资本。正是由于这种人际信任，需要将宗法制度型交易规模限定在较小的范围，特别是限定在价值可以共享、期望可以趋同的那些人之间。但期望和价值的趋同并不意味着交易各方实际上必须互相喜欢或者他们必须具有共同的目标^[7]。但他们必须能预测对方，在互相调整的过程中能够利用“同样的频率”，同时由高级的价值或目标联合起来。宗法制度型交易深深地嵌入并形成了长期的社会关系。

在社会层面，如果宗法制度型交易成为一种稳固的结构，它有可能产生内卷化（involution）现象。而在企业层面，则可能构成内部人控制的局面。

1. 内卷化

内卷化概念是吉尔茨（Geertz, Clifford）在1963年提出来，用以描述印尼爪哇地区一种生态稳定、内向、人口快速增长、高密度的农耕社会。“Involution”源于拉丁语 involutum，原义是“转或卷起来”，表达了“一种盘绕起来的、复杂的、纷繁混乱的事物——一种特化的产物”。内卷化描述社会的一种演化形态，即如果在结构上把变迁视为内向的传统力量与外向的变革力量之间相互作用的结果，那么内卷化的一般理论形式就是传统的力量在实际上占据了优势，并以非正式规则的形式决定变迁的路径，新的正式规则和程序经常被转化成某种因袭传统的不确定的制度。它好比水流中的漩涡，向前的推动力都被转化成做原地打转的运动^[9]。

内卷化强调社会演变中过去和现在的紧密联系。但围绕什么样的序参量内卷、以什么方式来选择和决定哈肯（Haken, Hermann）所说的“慢变量”及“快变量”，则可能更加关键。吉尔茨当时所观察到的爪哇社会是一种高度弹性和充满不确定性的社会。这种社会允许变化，也强调发展、吸收和采纳新知识，但不支持真正的变迁。在传统、惯例和共同价值的作用下，其内部不断精致化，每一个体和每一个环节都被安排得十分精细和复杂。历史显示，爪哇人并没有通过现代化进入经济的持久变革，而是内卷于传统的农业生产方式，有增长而无发展，尽管社会产值在增长，但资本积累的能力却在降低，单位产值的实际劳动成本增加，实际人均劳动投入边际收益递减，劳动生产率降低，经营效益恶化等。

2. 内部人控制

内部人控制（insiders control）概念由青木昌彦在1994年提出。青木教授认为内部人控制是转轨过程中所固有的一种潜在现象，是从计划经济制度的遗产中演化而来。那些已经从国家机关获得很大控制权的管理者们，可能利用计划经济解体后留下的真空，进一步加强自己的权力。当然，内部人控制的潜在可能转化为现实的过程在不同经济中是不同的。在私有化的场合，多数或相当大量的股权由内部人持有；在企业仍为国有的场合，在企业的重大战略决策中，内部人的利益得到有力的强调。有经济学家认为，内部人控制是企业经营活动的发展趋势。但国内多从其负面影响来讨论内部人控制现象。国内有学者主张，首先要社会上开展对代理人问题及内部人控制问题的讨论，让更多人意识到问题的严重性，要警惕和防范内部人控制现象等。但也有学者对此持不同观点。他们认为，目前国内

所出现的内部人控制问题，根源在于控制企业的人不是人力资本。如果在责、权、利对称的条件下，企业由人力资本进行控制，使得出资人资本增值，这是出资人所希望的，应该是件好事。在企业里，出资人的权利集中体现为产权的利益回报，人力资本在保证货币资本保值、增值的前提下，是可以独立地经营企业的。

CEO的产生在本质上就是人力资本在企业中的权利和地位得到保障的结果，表明人力资本作为一支重要力量已登上历史舞台。但CEO既不是总经理，也不是总裁。CEO的权力非常大，除了拥有总经理的全部权力外，还要加上董事长50%的权力。在CEO环境下，董事会的主要作用是选择、考评以CEO为中心的管理层以及订立薪酬制度。

可见，如果缺少必要而合适的外部激励与制约机制，不改写既定的规则和刷新当前的秩序，上述两种事态都可能走向在外部看来是自锁（lock-in self）的状态。

（四）采邑制度交易规则制度

从交易规则角度看，采邑是与市场相对的一种交易方式。后者是由于对抽象的编码信息的迅速而非个性化的扩散产生自调节（self-adjustment）的结果；前者指一种处理未编码的具体数据的交易结构。这种数据难以传播。即使面对面传授，也表现出高度的个性化。通常是在有限的范围内，交易各方积累的信息不对称。

布瓦索认为，采邑制度可能出现在未编码、未扩散的具体数据的人际化过程中，也可能从科层制度或者宗法制度的失败中凸现（emergence）。前一种失败是因为科层制度的规则不能有效地促使复杂的交易对象得到编码；后一种失败是由于存在信息分享障碍以及交易各方之间持久的信息不对称^[10]。其共同结果就是导致难以编码、难以扩散且与交易密切相关的信息置于一人或少数人之手，产生“克里斯马”（Charisma）现象。

作为一种交易结构，采邑制度以默会的和个性化的知识为基础。这些知识具有潜在用途。只要条件适宜，尽管它们有着内在的不扩散性，也可以发挥极大的作用。那些成功的企业家或者家族式企业的首脑就可能在采邑制度中运作。他们丝毫不亚于日本德州幕府（Tokugawa Shogunate）时期的谱代大名（fudai daimyo）或者外样大名（tozama daimyo）的所作所为。

研究表明，引导企业家“决胜”市场风云的就是他的“知识”，一种默会的、可能无法编码、既抽象又具体的知识。这种“知识”在知识类型学上是个矛盾体，至少目前还找不到确切的定位。依靠这种“知识”，企业家能把那些深层的真相传达给其他市场主体。格雷（Gray, John）认为，企业家的洞见或者感觉跟书本学习无关，而完全是意外发现珍宝的运气和天赋，是某种不受规律、法则制约的创造性活动，处于我们有意识控制的范围以外，企业家的感觉往往只是灵光一闪，突如其来，事先并没有征兆；而它一旦现身，就会重新安排宇宙万物。虽然他的用词不失夸张之嫌，但却道出了其中那种难以明言的“不可说”之意境。

四、结束语

劳动经济学的研究表明，职工担心他们的工会代表和资方达成不符合职工利益的协

议；金融经济学中，有许多资料表明董事会和管理人员之间存在合谋行为；公共经济学指出，税收和补贴并不必然总是为了使福利函数最大化而设置的，可能是为了满足特殊社会阶层福利的需要，有时甚至是出于合谋而危害税收的征收；规制（regulation）经济理论证明：如果政府机构或者官员和产业或者其他利益集团之间有合谋行为，将极大地影响规制制度和政策及其实施效果；发展经济理论向公众揭示出腐败的种种现象及其危害；多重代理人——如纳什和贝叶斯执行、拍卖和公共商品机制、团队道德风险、相对业绩评价以及等级结构等——之间的契约设计理论证明，代理人之间可能存在合谋动机和行为。

凡此种种表明，现实世界发生的经济行为就是上述交易规则型态的混合体。这个混合体在不同的维度（dimensions）和不同的层面（lays）上反映出信息作用的复杂、多样、异质和不确定性。与此对应的就是诸如“逆向选择”、“道德风险”、履“约”（合同）诚信、搭便车、“合同不完全”、“合谋”（collusion）现象的反复发生，以及层出不穷的逆向防范对策。

参考文献

- [1] 姜奇平：《信息空间——建立辨识 21 世纪财富地图的感觉》，《互联网周刊》，2003 年 11 月 3 日。
- [2] K. Sobczyk, Information Dynamics: Premises, Challenges And Results, Mechanical Systems and Signal Processing, 2001, Vol. 15, No. 3, pp475 ~ 498.
- [3] 布瓦索 (Boisot, Max H.): 《信息空间》，中译本，上海译文出版社 2000 年版，第 342 页。
- [4] 布瓦索 (Boisot, Max H.): 《信息空间》，中译本，上海译文出版社 2000 年版，第 346 页。
- [5] 沃恩 (Vaughn, Karen I.): 《看不见的手》，见：[英] 约翰·伊特韦尔编：《新帕尔格雷夫经济学大辞典》，中译本，经济科学出版社 1996 年版。
- [6] 布瓦索 (Boisot, Max H.): 《信息空间》，中译本，上海译文出版社 2000 年版，第 348. 页。
- [7] 焦文峰：《韦伯科层制理论分析》，《齐齐哈尔师范学院学报》，1998 年第 2 期。
- [8] 布瓦索 (Boisot, Max H.): 《信息空间》，中译本，上海译文出版社 2000 年版，第 360 页。
- [9] 申明民：《组织内卷化与中国的腐败》，《二十一世纪》，2002 年 6 月号。
- [10] 布瓦索 (Boisot, Max H.): 《信息空间》，中译本，上海译文出版社 2000 年版，第 372 ~ 373 页。

企业社会责任观的演进与发展： 基于综合性社会契约的理解*

陈宏辉 贾生华

一、企业社会责任之争

20世纪50年代中期以后，关于企业社会责任（Corporate Social Responsibility, CSR）问题的讨论一直都是西方经济学界和管理学界的热门话题，争论的核心是两大问题：一是企业为什么需要承担社会责任？二是企业应该在多大程度上承担社会责任？近半个世纪以来，在此问题上不同学派的观点纷呈而出，引导着学术界和企业界人士对于企业的社会功能进行全方位的思考。在此期间，利益相关者理论对企业社会责任问题的深入探讨最为引人注目。在主流企业理论不断的质疑和批评声中，利益相关者理论在企业社会责任问题上形成了一套相对完整的理论体系，并得到了企业实践的检验。1963年，斯坦福研究院（Stanford Institute）首次提出了“利益相关者”（Stakeholder）的概念，受到学术界的关注。在弗里曼、布莱尔、多纳德逊、米切尔等学者的努力下，利益相关者理论的分析框架、核心理念和研究方法逐渐明晰，并明确指出“企业对界定清晰的利益相关者负有社会责任”。与传统的股东至上主义的主要区别在于，该理论认为任何一个公司的发展都离不开各种利益相关者的投入或参与，企业追求的是利益相关者的整体利益，而不仅仅是某个主体的利益。这些利益相关者包括企业的股东、债权人、雇员、消费者、供应商等交易伙伴，也包括政府部门、本地居民、当地社区、媒体、环境保护主义者等压力集团，甚至还包括自然环境、人类后代、非人物种等受到企业经营活动直接或间接影响的客体（Freeman, 1984; Mitchell, 1997）。这些利益相关者都对企业的生存和发展注入了一定的专用性投资，他们或是分担了一定的企业经营风险，或是为企业的经营活动付出了代价，企业的经营决策必须要考虑他们的利益，并给予相应的报酬和补偿（Blair, 1995）。从这个意义上讲，企业是一种治理和管理专业化投资的制度安排，其生存和发展取决于它能否有效地处理与各种利益相关者的关系，而股东只是其中之一罢了（布莱尔，1999）。在利益相关者理论看来，企业的发展前景有赖于管理层对公众不断变化的期望的满足程度，也就是说依赖于企业管理层对利益相关者利益要求的回应质量，企业对于社会负有包括经济责任、法律责任、道德责任和慈善责任在内的多项社会责任（Donaldson, 1995; 万建华，1998）。

然而，利益相关者理论及其社会责任观受到了主流企业理论的猛烈抨击，后者认为利

* 原载：《中国工业经济》，2003年第12期。

益相关者管理“仅仅只是自由主义代言人精心策划的一个攻关活动而已”（约瑟芬·马特比、路易·威尔金森，1999/1998）。1976年诺贝尔经济学奖获得者、美国经济学家米尔顿·弗里德曼可能是反对利益相关者理论的学者中名气最大的一位。在20世纪60年代出版的一部著作中，弗里德曼曾经明确指出：“企业有一个并且只有一个社会责任——使用它的资源，按照游戏的规则，从事增加利润的活动，只要它存在一天它就如此。……如果企业管理者接受这种（广泛的）社会责任观念，而不是尽可能地为企业创造价值的话，那就几乎没有什么倾向能如此彻底地破坏我们这个自由社会的基础了。这种（利益相关者的）观点基本上是一种败坏社会的信条。”（Friedman, 1962）在这里，弗里德曼实际上是拒绝承认企业需要承担社会责任，企业即使有社会责任的话，也只有经济责任。企业只需要为其股东赚取足够多的利润就是对社会最大的贡献，无需关注什么利益相关者问题。弗里德曼认为，在自由企业制度中，企业管理者必须要对股东负责；而股东想尽可能多地获取利润，因此管理者的唯一使命就是力求达到这个目的。如果管理者把“股东的钱”花到公众利益上，他实际上就是在未经股东许可的情况下花钱，因而必须遭到反对。同样，如果企业从事某些社会行动而花费的成本通过提高产品价格而转嫁到了消费者身上，这个管理者就是在花消费者的钱，这种做法也应该遭到拒绝（Friedman, 1970）。作为一名复兴自由放任主义经济学的旗手，弗里德曼认为利益相关者管理模式会使得企业因承担社会责任而使其成本增加，但市场的严酷性最终将会威胁到它的竞争地位。

简而言之，利益相关者理论与主流企业理论在企业社会责任问题上存在着根本的分歧：前者认为企业应该对其利益相关者负起包括经济责任、法律责任、道德责任和慈善责任在内的多项社会责任，而后者则强调企业经营惟一的任務就是在法律许可的范围内追求利润最大化，过多地关注社会责任会使企业经营迷失方向。这两种理论在企业社会责任问题上的争论成了20世纪80年代以后西方管理学界的一道独特风景线。

二、企业社会责任观的演进：历史的视角

笔者认为，单从理论和逻辑的层面来争论某一种社会责任观的对错是不够的，也难以得出服众的结论，而从历史的视角来剖析企业社会责任观的演进也许会有益于我们对企业社会责任内涵的理解。传统经济理论认为，企业如果尽可能高效率地使用资源以提供社会需要的产品和服务，并以消费者愿意支付的价格销售它们，企业就尽到了自己的社会责任（乔治·斯蒂纳、约翰·斯蒂纳，2002/1997）。虽然这一来源于古典主义经济学的企业行为标准的假设在亚当·斯密之后200多年的时间里依然稳立不倒，但从来就没有在企业的实践中被无条件地实行过。就连亚当·斯密自己也承认，由于社会的原因，这一标准肯定会有无数的例外（亚当·斯密，2001/1776）。事实上，这些“例外”在现实之中往往就衍生成了企业不得不考虑其应有的社会责任。

1. 早期企业社会责任观

在18世纪末期以后，企业的社会责任观就已经开始发生了一些微妙的变化。那时西方企业的规模普遍都还很小，企业家们行为节俭。但与此同时，也有一些小企业的业主们经常捐助学校、教堂和穷人们。当然他们这样做完全是个人行为，被认为是一种乐善好施

的举动。随着企业财富的积累，企业的社会活动开始持续增加。进入 19 世纪以后，已经有一些企业初具规模，进而可以在一定程度上影响当地的经济、政治和文化生活了。

然而，在整个 19 世纪中，这些参与社区建设、向穷人捐款、兴办教育的慈善活动都是些个人行为，而不是企业行为。换句话说，企业主完全可以支配他个人的财富来“行善事”，但企业是不能承担什么社会责任项目的。当时的法律在企业管理者如何使用公司的资金上也有明确的规定，认为企业没有权力去做其业务范围之外的事，否则就是“过度活跃”（*cultra vires*）了。这一法律短语的意思是企业的活动超出了它的权力。一个处于“过度活跃”状态的企业往往容易遭受股东的诉讼。虽然法官们一般都会允许企业在其所在的城镇里建立学校和教堂，但对于企业过多地考虑利益相关者的利益要求是有限制的。因此，在 19 世纪，真正有责任心的可能是企业家，而不是企业。

其实，在整个 19 世纪，人们对企业的社会责任观是持消极态度的。之所以如此，据说是“社会达尔文主义”思潮起了很大的作用。许多人认为社会生活也是符合“达尔文主义”的，弱肉强食、适者生存也是社会生活中的普遍规律。因此，企业捐款资助弱者是与自然进化过程相违背的，由此造成的社会保护只会降低人类的适应能力。那些给予捐款的好心人实际是支持了那些在人类竞赛中的失败者，违背了自然法则。赫伯特·斯宾塞是带头倡导这种严酷主张的人，他在 1850 年出版的一本书中写道：“这看起来很残酷，一个劳动者生了病，就丧失了与其他强壮竞争对手进行竞争的能力，就必须遭受穷困，那些寡妇和孤儿必须自己为生或死而挣扎。虽然如此，如果不是孤立地看，而是与整个人类的利益联系起来看，这些无情的命运就充满了最高的仁慈，正是这种仁慈使得那些父母患病的孩子早早入土为安，也正是这种仁慈使得意志消沉者、酗酒者以及那些受流行疾病折磨的虚弱者早早告别人间”。（Spence, 1970/1850）在这种“弱者死有余辜”思潮的影响下，许多企业不仅不主动地承担社会责任，而且对那些与企业有密切关系的供应商、分销商、员工极尽盘剥，以求尽快变成社会竞争的强者。因此，从我们现在的视角来看，在早期消极的企业社会责任观的影响下，企业认为与其利益相关者之间是一种赤裸裸的市场竞争关系。至于众多利益相关者的非经济性的利益要求，企业是不需要过多考虑的；即使加以考虑，那也是出自于企业主的怜悯之心。

2. 近现代企业社会责任观

进入 20 世纪以后，工业的大力发展产生了许多负面的社会影响。批评家们开始指责“社会达尔文主义”的残酷和冷漠，并意识到企业必须对那些与其有关联的群体负起责任。到 20 世纪 20 年代，共出现了三种支持扩大企业社会责任的观点（乔治·斯蒂纳、约翰·斯蒂纳，2002/1997）。第一种观点是“受托人观”，即认为管理者是受托人（trustee），公司赋予他们相应的权力和地位，他们的行为不仅要满足股东的权益，而且要满足顾客、雇员和社会的需要。第二种观点是“利益平衡观”，即管理者有义务来平衡（balance）那些与企业有关联的集团之间的利益。也就是说，企业管理者就是各种各样的互相冲突的利益团体之间的利益协调人。第三种观点是“服务观”，他们认为企业有义务承担社会项目去造福或服务于公众，而管理者个人也可以通过成功地运营企业来减少社会不公、贫穷、疾病，从而为社会作出贡献（Lunden, 1988）。这些观点的迅速传播促进了一些企业领导人在实践中开始实施完全不同于以往的社会责任活动。例如在 1924 ~ 1954 年间，领导 Sears Roebuck

公司的罗伯特·伍德就是一位全面关注利益相关者利益要求的先行者。他相信，一个大公司除了是一个经济机构，还是一个社会机构。在 Sears Roebuck 公司 1936 年的年度报告中，他写道：“在最近一段社会经济和政治价值观发生改变的时期，提交一份管理层职位与职责的报告，不仅仅从财务的角度来报告，而且考虑更为一般性的广泛的社会责任，这些社会责任虽不能用数字予以说明，但非常重要。这样的做法是值得提倡的。”伍德为 Sears Roebuck 公司制定出了一套方法，详细说明公司应该如何向其主要的支持者——顾客、公众、雇员、供应商以及股东履行责任。在类似于伍德这样的先行者的带领下，20 世纪早期的西方企业普遍已经不再对其社会责任抱着冷漠的态度，它们开始主动捐款，资助社区活动和红十字会事业，帮助当地政府完善义务教育和公共健康制度。当然，企业经营活动中最大的转变是开始真正地重视与其有密切关系的各种利益相关者的利益要求，许多企业都设立了正式的工会组织，建立起专门的组织机构来处理供应商、分销商、贷款人、特殊团体、社区的意见和建议。同时，许多大公司还设立了慈善基金，并取得了相应的免税特权，企业积极从事社会责任活动已经不再是个别现象，反而成了有企业拥有远大抱负的象征。最终，1953 年，新泽西州的最高法院认为，“过度活跃”条款是不合理的限制，并拒绝执行它，从而使该条款寿终正寝（乔治·斯蒂纳、约翰·斯蒂纳，2002/1997）。

20 世纪中期以后，虽然一些崇尚自由经济主义的学者仍然拒绝承认企业的社会责任^①，但是随着利益相关者理论的迅速发展，广义企业社会责任的观念继续得到了发展和扩大，并且得到了学术界的广泛支持。同时，现实中的企业开始积极地承担起各种社会责任，主动地甚至是前瞻性地考虑其利益相关者的利益要求。从根本上而言，企业社会责任观在 20 世纪中期以后逐渐扩大的原因在于：加速的工业活动不断地改变着社会，企业活动对社会产生了前所未有的巨大影响，履行多种社会责任是企业的应有之义。例如，在 20 世纪初，工业废水所产生的致癌物还不被世人所熟知，雇佣中的种族歧视所产生的危害也没有被详细考察。然而，几十年过去以后，企业生产所带来的“副产品”已经非常明显了。正如管理学大师彼得·德鲁克曾指出的那样，“一个健康的企业和一个病态的社会是很难共存的”（Drucker, 1973），况且社会之所以进入“病态”，企业本身负有不可推卸的责任。

20 世纪 80 年代以后，企业承担的社会项目范围不断扩大，有的企业甚至开始实施大范围的社会行动，项目范围涉及教育、公共健康、就业福利、住房、城区改造、环境保护、资源保护、双职工家庭的婴儿护理中心等，在每一个领域，不同的企业所实施的项目多达几千个，同时涌现出了一批积极承担社会责任、主动关注利益相关者利益要求的令人尊敬的企业。例如，位于美国新泽西州的制药企业默克公司（Merck Co.）在 1981 年花费了 2.3 亿美元来研发一种名叫 ivermectin 的药品，该药品用于治疗一种能使人变瞎的“河盲症”。几个世纪以来，这种疾病都在折磨着临近河流的赤道地区的贫困居民。尽管非常清楚这些居民无力购买这种药品，但是默克公司最终还是决定生产 ivermectin，并免费配送给那些需要这种药品的居民，从而挽救了上百万人的健康和生命。另外，美国 Ben &

^① 例如我们已经阐述过的弗里德曼式的企业社会责任观——企业没有社会责任，如果一定要说有的话，那就是尽量为股东赚钱。

Jerry 冰淇淋公司的创办人 Ben Cohen 和 Jerry Greenfield 就以一种他们称之为“共同繁荣”的理念来经营公司，他们尽量避免企业的经营对社会造成伤害，深信企业的命运与社会的命运是不可分割地交织在一起的。该企业遵循的经营原则是“在促进社会繁荣的同时促进企业的繁荣”。每年，Ben & Jerry 公司都花费巨资来种植足够多的树，以补充那些被它砍下来用于制作巧克力和冰棍棒的树木。另外，当那些无家可归者和残疾人要求加盟 Ben & Jerry 公司连锁经营店时，公司免收了每人 25000 美元的特许经营费（乔治·斯蒂纳、约翰·斯蒂纳，2002/1997）。

到了 20 世纪 90 年代以后，像 Ben & Jerry 公司、默克公司这样品行优良的企业一时间成了社会学习的楷模，衡量一个企业经营活动优劣的指标也从早期单纯的经济指标发展为综合性的“企业社会绩效指标”。即判断一个企业的经营效果不仅要看它的经济绩效，也不仅仅看它是否接受社会责任这一观念，而且要看它在主动寻求社会需求、实施具体项目以帮助实现这些需求过程中的表现。应该说，与早期的企业社会责任观念相比，企业社会绩效的概念是全新的。在这种企业社会责任观的指引下，企业必须充分了解其利益相关者利益要求的内容，并尽量调用企业资源来满足这些利益相关者合理的利益要求。也就是说，现代企业中利益相关者利益要求之所以能够受到企业越来越多的重视，正是因为现代企业社会责任观的确立。

三、综合性社会契约：联系企业社会责任与利益相关者利益要求的纽带

至此，我们回顾了企业社会责任观的形成和发展过程，试图从历史和逻辑相一致的角度来剖析企业的社会责任问题。我们的基本结论是：企业是一种人格化的组织，它能够而且必须对其经营活动所处的社会系统的要求做出回应，承担相应的社会责任是其应有之义。事实上，现代企业理论也已经把企业理解为不同个人之间一组复杂的显性契约和隐性契约的交汇所构成的一种法律实体（Jensen & Meckling, 1976）。在这种法律实体中，交汇的契约既有经营者与所有者之间的契约、经营者与雇员之间的契约，还有企业作为债权人与债务人之间的契约、企业作为供应商（或消费者）与消费者（供应商）之间的契约、企业作为法人与政府之间的契约等等。这样，企业的行为实际上就成了一组复杂契约系统的均衡行为，这种复杂的契约系统的主体就是一系列目标不同且可能相互冲突的利益相关者。

在制度经济学看来，制度是一种人为设计的界定人们相互关系的约束机制。它既包括正式的约束机制，如规则、法律、法规，也包括非正式的约束机制，如行为准则、习惯、自我行为规范等（诺斯，1994）。既然所有的制度安排都是显性契约和隐性契约的混合体，企业作为制度安排的一种自然也不能例外，在构成企业的各种契约交汇中，显性契约和隐性契约也是并存的。例如，在雇佣契约中，雇主和雇员尽管可以就工作时间长度、从事的工作岗位、不同岗位的工资水平等等做出明确的规定，但在大多数情况下，雇员对企业价值的贡献是难以客观地加以衡量的，只能求助于各级管理人员的主观评价，而由企业管理人员或经营者对雇员的绩效进行评估（进而作为确定雇员报酬的根据）往往存在着“败德行为”的可能，即经营者可能会有意低估雇员绩效的价值，以减少企业的工资支付额。为

此，雇佣双方事前必须达成一种隐性契约，而这种契约的实施是建立在双方的信任基础之上的。正是由于隐性契约的存在和有效地实施，才补充和改进了显性契约的不足（李向阳，2000）。从这个意义上讲，如果企业只考虑某些利益相关者的显性契约而忽视了其隐性契约，或者是根本上就没有考虑另外一些利益相关者的利益要求，那么就会导致错误的企业社会责任观。美国管理学家多纳德逊和邓非（Donaldson & Dunfee, 1994/1995）将企业与其利益相关者之间所遵循的所有契约形式总称为综合性社会契约（integrative social contracts），进而将企业社会责任和企业利益相关者的利益要求统一起来。他们认为企业对利益相关者的利益要求必须做出反应，这是因为“企业是社会系统中不可分割的一部分，是利益相关者显性契约和隐性契约的载体”。倘若企业忽视其社会责任，对其利益相关者的合理利益要求不作慎重考虑且尽量满足的话，那么这种企业的长久生存和持续发展就很可能成问题了（多纳德逊、邓非，2001/1999）。

在论证综合性社会契约之所以能成为联系企业社会责任与利益相关者利益要求的纽带时，多纳德逊和邓非详细考察了两种不同的支持性观点^①。第一种观点被称为“工具性观点”，其核心思想是企业之所以要承担社会责任、关注利益相关者的利益要求，是因为这样做将使企业变得更有利可图（Jones, 1995）。这一观点的否定说法就是，忽视企业的社会责任、忽视利益相关者利益要求的企业实际上是在冒风险，而与重要的利益相关者对立则可能危及企业自身的生存。以一家公司有关迁址的决策为例，有些利益相关者可能会采取支持性行动，如政府机构放松有关划区的法令或提供退税，工会同意更灵活的工作规定，以及供货商答应给予更优惠的条件等；而另外一些利益相关者则可能表示强烈反对，如分销商由于需要重新规划分销渠道而发动联合抵制进货，社区由于就业机会的减少而表示抗议等，可能的行动包括工会发起罢工，社区各群体采取联合抵制，现有的分销商会因为合同的中断而起诉等等。公司就应当考虑某些利益相关者对抗这个决策的行动所可能造成的损失。因此，企业管理者需要对利益相关者行动所可能造成的得失做出预算，再根据公司的财务状况做出权衡。总而言之，工具主义的观点认为企业需要承担社会责任、考虑利益相关者的利益要求，是因为可以将其作为一种实现企业经营目的（End）的手段和工具（Instrument）。古德帕斯特曾把这种观点描述为“策略性的利益相关者分析”（Goodpaster, 1991）。第二种观点则是“规范性观点”（Donaldson & Preston, 1995），其核心思想是不论企业的经营状况如何，它都有一种伦理性的社会责任，应当对利益相关者的要求做出恰当的回。规范性观点强调做“正确的事”，做“应该做的事”，它不再将“关注利益相关者利益要求”作为“实现企业经济利益”的一种手段和工具，而是超出对企业的净成本的简单分析。一个经典的例子是，当企业决定它提供给顾客的货物有多安全时，它所考虑的必须是作为利益相关者的消费者和用户的身体安全方面的合法利益，而不再是“权衡公司的财务状况”。因此，公司应该立即从市场召回一种不太安全的产品，而不是花费

^① 作为利益相关者理论研究方法的开拓者，多纳德逊在此也是从不同的研究方法的角度来分析综合性社会契约论的。不过，Donaldson & Preston（1995）曾经归纳了三种而不是两种利益相关者理论，即规范的、描述的/实证的、工具性的利益相关者理论。多纳德逊和邓非在此只提到两种（工具性的和规范的），是因为对于公司策略选择来说，只有它们是备选者；而描述的/实证的利益相关者理论“只是描述或解释公司的活动，而没有它们应当如何做的含义”。（多纳德逊、邓非，2001/1999）

时间去计算这样做的成本有多大。尽管在很多情况下公司的律师会从纯法理的角度劝告说,这么做会增加公司被追究责任的可能性^①。持这种观点的研究者跳出了如何看待利益相关者利益要求的“目的-手段”之争,而是从更根本的价值判断视角来论证企业必须关注利益相关者的利益要求,认为这样做既不是企业的“目的”,也不是企业为了达到某种目的而采用的“手段”,其性质完全类似于一个人在社会系统中所需要扮演的角色任务(Clarkson, 1995; Freeman, 1984; Mitchell & Wood, 1997)。古德帕斯特把这种规范性的观点称为“多方面信托”的利益相关者分析(Goodpaster, 1991)。

在多纳德逊和邓非所建立的综合性社会契约论中,兼收并蓄了上述两种观点。他们认为工具性观点是企业必须承担其社会责任、考虑其利益相关者利益要求的“最流行的辩护理由”,也容易为企业界所接受;但是规范性观点从更根本的基础上奠定了企业与利益相关者之间的契约关系,明确了企业承担社会责任的本质(多纳德逊、邓非, 2001/1999)。综合社会契约论之所以强调规范性观点,是因为企业与其利益相关者的契约性质使然。虽然许多利益相关者的利益要求都是通过他们与企业所签订的显性契约来实现的,但是还有许多利益要求是无法显化的,或是显化的成本极高以至于双方都愿意放弃这种显化的努力。但这并不意味着当某些事前没有在契约中明示的或然事件发生时,企业可以以“契约中没有这一规定”为由而推卸责任,因为这既不符合规范性的道德伦理,也会对企业生存发展产生不利的影响。当然,有些利益相关者的利益要求是合理的,有些利益相关者的利益要求则是不合理的,这需要企业建立起一套严格的审计程序和规范的决策过程来加以甄别,但这是一个技术层面的问题,并不能从根本上否定企业应该承担的社会责任(Quinn & Jones, 1995)。事实上,在当今社会,企业在其经营活动中必须考虑承担更多的社会责任的观点开始深入人心,而那种单纯追求股东利益最大化的经营思想开始受到质疑。当然,绝大多数人仍然认为企业的经营活动必须考虑利润,但人们对利润的态度已经发生了变化。“今天,如果经理试图增加公司股东的利润,最好方式是兼顾企业中相关的主要社会团体的利益。”(乔治·斯蒂纳、约翰·斯蒂纳, 2002/1997)正如美国管理学大师德鲁克所阐述的那样:“在企业经营活动中重要的是管理者应该意识到他们必须考虑公司政策和公司行为对于社会的影响。他们必须考虑一定的行为是否有可能促进公众的利益,有利于社会基本信仰的进步,有利于社会的稳定、强盛与和谐。”(Drucker, 1995)

有学者研究表明,从20世纪80年代以来,许多公司已经开始越来越多地从事一些与股东利益最大化目标无直接联系的活动,如资助大学或研究机构的研究项目,向社会公益项目捐款等等。这表明公司已经是一个“社会存在”的概念,而不仅仅是为股东利益而生存,公司是为所有相关利益者服务的(黄少安、宫明波, 2002)。同时,一些西方学者的研究也发现,即使是在英国和美国,在20世纪大部分时间里,实行利益相关者纳入、考虑社会利益的企业在经营绩效上要比奉行“股东至上主义”的企业更胜一筹(威勒、西兰琶, 2002)。相反,那种认为企业只为“股东利润最大化”而生存的思想很可能导致非道德经营和不道德经营。因为企业一旦把利润作为惟一的追求时,往往就会忽视综合性社会契约的要求,在经营行为上就难免把职工作为获利的工具,把顾客视为争夺市场份额的

^① 强生公司的律师就根据这个理由反对收回 Tylenol 胶囊的(多纳德逊、邓非, 2001/1999)。

对象，把竞争对手看成是对头冤家，把媒体视作祸水，把政府法令当作儿戏，把自然资源作为肆意攫取的目标。从这个意义上来讲，正是由于遗忘了企业与诸多利益相关者之间所缔结的综合性社会契约，才会爆出南京冠生园的变质月饼风波和金华火腿的“敌敌畏”杀虫丑闻。

由于现代企业社会责任观的发展已经要求企业积极主动地承担社会责任，在遵守其综合性社会契约的前提下开展经营活动，因此企业管理者必须清醒地认识到，企业决不是游离于现实生活之外的一个产品加工器，企业的社会责任感和伦理道德观须臾不可缺失。我们认为，企业要想真正遵循综合性社会契约，形成良好的社会责任感，必须认真开展以下针对利益相关者的工作。

(1) 将企业的合法收入及时在股东、债权人、供应商之间合理分配，不拖不欠，形成良好的企业信誉。企业像人一样，也有自己的性格、气质、个性、风貌和品德。企业要在市场立足，首先要有信誉，这是企业社会责任的第一体现。

(2) 尊重员工，创造人性化的工作环境。现代企业社会责任要求企业对待员工决不仅仅是支薪了事，还需要给员工提供符合安全和卫生要求的工作场所，制定公正、公开的奖惩制度，公平地对待每一位劳动者，为每一位雇员提供广阔的生存和发展空间，不搞任何形式的种族、性别和工种歧视。

(3) 真正将消费者满意视作企业销售产品的最高目标，决不欺瞒消费者。“顾客是上帝”已经成为了许多企业管理者的口头禅，但能否言行一致，则有赖于企业是否真正具有社会责任感。据有关市场研究表明：蒙骗而开罪了一个消费者，其现身说法所引起的连锁反应会影响 25~30 名消费者的购买决心，最终受损的还是企业自己。

(4) 依法经营、照章纳税，并尊重生产地、销售地社区居民的生活习惯和生活规律。正如一个讲道德的人往往是自觉守法并且不会粗鲁地打搅别人的生活一样，一个讲道德的企业自然会将法律作为最低的道德标准，也不会破坏当地社区、居民的既有生活秩序。

(5) 保持与相关媒体的良好合作关系。如果万一出现了管理纰漏，对社会其他成员造成伤害时，则决不推诿，要真心实意地致以歉意，并尽力为相关者弥补损失。正如一个有道德修养的人在犯错误后会主动向受到伤害的人赔礼道歉、给予适当地赔偿一样，企业经营也是如此。

(6) 在力所能及的情况下积极从事慈善事业，回报社会，树立健康的公众形象。在我国，为富不仁始终为中华民族文化所不齿。只有将企业的经营行为融入到社会发展的潮流之中，并主动为社会解难分忧，这样才可能真正实现道德化经营，并获得长远的发展。

(7) 维护生态环境，注重可持续发展。当前，全球环境问题日益成为人们关注的焦点，企业作为破坏环境的“罪魁祸首”，必然要对其负责。以德治企要求企业在解决环境问题上发挥更加主动积极的作用。

美国 Sears Roebuck 公司总裁罗伯特·伍德在谈到公司的社会责任和持续发展问题时，反复强调要把股东的利益放在最后一位来加以考虑，其原因“不是因为股东不重要”，“而是因为，在很多情形下，除非他先满足顾客、雇员和其他利益相关者的需要，否则，他是得不到满意的回报的”。

四、评述性结语

分析表明，不同的理论学派对于企业社会责任问题的回答大相径庭。通过历史的视角来审视企业社会责任观的演进可以发现，随着人们在不同时期对企业利益相关者利益要求认识的不断深化，企业社会责任的实践越来越多地倾向于利益相关者理论的论述。通过综合性社会契约这一概念，我们进一步将企业社会责任与企业利益相关者的利益要求紧密地结合在一起，而且还可以在公司治理的构架中寻求相应的措施和制度来保证企业承担相应的社会责任。

值得特别指出的是：我们强调企业必须履行其综合性社会契约、考虑利益相关者的利益要求、承担相应的社会责任，其含义决不是倡导“企业办社会”。对于我国许多企业，尤其是国有企业来说，由于因循了“企业办社会”的发展思路而背上了沉重的历史包袱，那些“大而全，小而全”的国有企业正在为当初不按经济规律办事的决策还债。这种过多地承担了社会职能——注意，不是社会责任——的企业实际上是按照政府的意愿，做了许多政府应该做的事情，做了许多别的企业该做的事情，而不是出于自身发展需要所做出的理性决策。这与我们所强调的“企业应该承担相应的社会责任、兼顾企业利益和社会利益”的原则具有完全不同的性质：“企业办社会”是随意扩大企业边界的行为，它实际上是不把企业当作“主动创造财富的主体”看待，而是把企业当作完成社会目标的工具看待；“主动承担社会责任”是在承认企业应该主动创造财富的前提下，强调企业不仅是为股东创造财富，而是为众多的利益相关者和社会提供服务。用管理学家古德帕斯特的话来讲，企业必须既要实现自身的利益，同时还要实现必须的社会利益，二者不可偏废，“既不能造成没有伦理学的企业，也不能造成没有企业的伦理学”（Goodpaster, 1991；多纳德逊、邓非，2001/1999）。

参考文献

- [1] 陈宏辉：《企业利益相关者理论与实证研究》，浙江大学博士论文，2003年。
- [2] [美]多纳德逊，邓非著：《有约束力的关系——对企业伦理学的一种社会契约论的研究》，赵月瑟译，上海社会科学院出版社2001年版。
- [3] [美]乔治·斯蒂纳，约翰·斯蒂纳著：《企业、政府与社会》，张志强，王春香译，华夏出版社2002年版。
- [4] [英]威勒，西兰琶著：《利益相关者公司——利益相关者价值最大化之蓝图》，张丽华译，经济管理出版社2002年版。
- [5] Blair M. M., *Ownership and Control: Rethinking Corporate Governance for the Twenty-first Century*, The Brookings Institution, Washington DC, 1995.
- [6] Blair M. M., *Corporate "Ownership"*, *Brookings Review* (Winter), 1995, pp16~19.
- [7] Clarkson, M., *A Stakeholder Framework for Analyzing and Evaluating Corporate Social Performance*, *Academy of Management Review*, 1995, Vol.20, No.1, pp92~117.
- [8] Donaldson, T. & Dunfee, T. W., *Toward, A Unified Conception of Business Ethics: Integrative Social Contracts Theory*, *Academy of Management Review*, 1994, Vol.19, No.1, pp252~284.

-
- [9] Donaldson, T. & Dunfee, T. W. , *Integrative Social Contracts Theory : A Communitarian Conception of Economic Ethics , Economics and Philosophy* , 1995 , Vol.11 , No.1 , pp85 ~ 112
- [10] Friedman, M. , *Capitalism and Freedom* , Chicago : University of Chicago Press , 1962.
- [11] Friedman, M. , *The Social Responsibility of Business is to Increase its Profits* , *New York Times Magazine* , 1970 , Vol.9 , p13.
- [12] Goodpaster, K. E. , *Business Ethics and Stakeholder Analysis* , *Business Ethics Quarterly* , 1991 , Vol.1 , No.1 , pp53 ~ 74.
- [13] Spence, H. , *Social Statics* , New York : Robert Schalkenbach Foundation , 1970/1850.
- [14] Jones, T.M. , *Instrumental Stakeholder Theory : A Synthesis of Ethics and Economics* , *Academy of Management Review* , 1995 , Vol.20 , No.2 , pp404 ~ 437.

产业集群的分工优势*

孙洛平 孙海琳

产业集群 (industrial cluster) 是指特定产业以及支撑和关联产业在一定地域范围内的地理集中, 形成产业集群的过程也被称为产业集聚 (industrial agglomeration)。产业集聚不是什么新机制, 而是一种在世界范围内较为普遍的经济现象。早在工业化发展时期, 它就成为产业发展的一种有效模式。随着发达国家市场制度的日益完善, 许多重要产业的发展开始突破地理区域的限制, 以大公司和跨国公司的方式发展。不过, 许多国家始终有一些传统产业保持着地理区域相对集中的特征, 如美国加州的葡萄酒业、意大利的皮具时装业等。在当代, 现代信息产业在硅谷的迅猛发展、引领和左右世界信息产业发展方向的现象, 重新引起人们对产业集聚现象的关注。

为什么特定产业在地理上的集聚会形成较高的竞争力, 人们从不同的角度出发往往给出不同的解释。概括而言, 有两种主要的思路: 一是从地理学的角度, 强调不同区域的吸引力及其相互作用; 二是从经济学的角度, 关注企业的经营环境和竞争优势的形成机制。除了学科方法的差异外, 产业集群的类型也影响人们观察问题的视角。技术创新型的产业集群与加工制造型的产业集群之间的差异是明显的。对于技术创新而言, 技术和知识外溢效应、人们的非正式交往、创新活动的共同技术平台的形成等是决定技术创新行为集聚的关键因素; 对于加工制造业来说, 企业的经营环境和条件往往是更为重要的考虑因素。

我国的产业集群一般都属于加工制造型, 而且多数产业集群并不在中心城市, 而是在中小城镇甚至乡镇发展起来的。本文试图从经济学的角度, 说明我国产业集群的优势形成机制与发达国家的 (加工制造型) 产业集群有所不同, 主要在于它能够降低集群内部企业之间的交易费用, 从而可以提高企业间的分工水平, 这是我国产业集群竞争优势的最主要来源。

一、产业集群竞争优势的来源

早在一个世纪前, 英国经济学家马歇尔就已注意到一些产业在特定地区集聚的现象, 他将这一特定地区叫做“产业区” (industrial district)。产业区内集中了大量种类相似的小型企业, 相互之间形成高度的分工专业化联系。马歇尔认为, 工业在一个地方聚集的优势在于能够产生知识和生产技能外溢效应, 导致高效率专业机器的采用, 形成熟练劳动力集中的市场。这三种优势来源的特点是, 企业层面的规模报酬不变和整个产业区的外部经济性 (external economies), 即竞争的优势不是来自于单个企业, 而是企业之间的分工协作。

* 原载:《经济学动态》, 2004年第6期。

波特 (Porter, 1990) 从企业如何获得竞争优势的角度对产业集聚现象进行了研究, 并提出了产业集群 (industrial clusters) 的概念。他认为, 产业集聚在一起增强了企业之间的相互作用强度, 它从两方面影响产业整体的竞争力: 一是能够形成专门的供应商和需求, 促进产业的纵向和横向分工发展, 有利于基础设施的建设和利用, 以及专门的研究和开发机构的产生等; 二是增强了企业之间的竞争强度, 促成企业的发明创新活动和新技术的利用等。波特立足于个别企业的竞争力来解释产业集群的优势, 这是与马歇尔不同的地方; 用企业之间相互作用导致的动态外部经济性来解释产业集群的优势, 这是与马歇尔类似的地方。

20 世纪 90 年代, 以 Krugman (1991)、Fujita (1988) 和 Venables (1996) 等人为代表的新经济地理学, 为产业集聚的机制提供了新的分析视角。新经济地理学以企业层面的规模报酬递增和不完全竞争的市场结构为前提, 在 Dixit and Stiglitz 垄断竞争模型的基础上, 从经济活动的内部机制来解释产业集聚现象。按照新经济地理学的观点, 消费者偏好多多样化的消费, 希望产品的品种越多越好, 而企业的生产具有规模报酬递增的特征, 产品的品种越多, 单个企业的生产规模优势就越难以发挥。正是在这个两难选择中, 一个地方的经济规模越大, 可容纳的产品品种就越多, 以致每一个产品的产量得以扩大, 其规模报酬递增效应得以发挥, 最终形成生产同一类产品的产业在一个地方集聚的现象。

另外, 新经济地理学还将企业间的分工引入对于产业集聚现象的说明 (Venables, 1996), 将企业内部的规模经济与企业外部的规模经济相结合。指出产业集群的规模扩大, 中间投入品的种类增多, 分工更细, 产品的价格也更低, 从而形成更强的地理上的集聚力。新经济地理学虽然引入了分工, 但不是以促进分工, 而是以给定分工来说明产业集群的优势。

不过, 在上述优势之外, 我们可以发现产业集群的一个重要的特点是, 它不以完善的市场规制和先进的企业制度为先决条件, 这也是为什么在不具备当代市场规制条件的工业革命时期, 就存在产业集群的原因。我国的经济发展与国外相比缺乏的正是市场规制和企业制度方面的优势, 按道理这一劣势将阻碍分工的深化。可是在我国东南沿海地区, 尤其是长江三角洲地区的许多中小城镇, 产业集群生产模式不仅大量涌现, 而且竞争力一点也不亚于中心城市。我们来说明产业集群推进分工深化的机制, 即便没有企业内部的规模报酬递增性和企业之间的外部经济性, 产业集群也同样能够形成相对竞争优势。

二、利用市场组织分工的费用

分工能够提高生产的效率是一目了然的, 不过, 分工不会自然而然地发展和深化, 原因是人们在利用分工提高生产效率的同时, 也为分工参与的每一方利用自己的相对垄断地位要挟他人提供了可能, 由此导致的利益争夺行为为足以耗竭部分乃至全部的分工潜在利益。因此, 在较大的范围内组织分工并非易事。市场经济主要是通过市场和企业这两种方式组织分工的, 这两种组织分工的方式都是有费用的。我们先来分析市场组织分工的费用。

为简化分析, 不计物质资本因素, 只考虑分工引起的人力资本投入的变化。假定人力

资本投入是以掌握生产技能的形式体现的,而且人力资本的数量可以用学习时间的长短来衡量。假定生产技能的学习时间可以分解为两部分:一是只能用于分工合作双方的专用技能学习时间;二是可用于合作体之外的通用技能学习时间。专用技能与通用技能是相对而言的。比方说,每个人都要学会种棉和织布的技能后才能制衣,如果只有两个人合作,一人种棉,一人织布,那么他们各自的技能将不是通用性的。原因是两人的合作要是破裂,没有人愿意雇用他们,因为其他的人已经掌握了全部制衣技能,没有必要为雇用他们而发生新的协调费用。如果种棉和织布已经成为普遍性的分工,那么这两种生产技能都是通用性的。我们不妨将因率先分工而引起的分工双方生产技能的相互依赖性,看作是一种分工专用性的生产技能。当然,分工总是会伴随物质资本和人力资本的相应改变,这时会产生普通意义上的专用性人力和物质资本。

分工的一个起码的优势来自于每个参与人只需要学习一部分生产技能,从而使合作体的总的学习时间缩短,使可用于生产的时间增加。我们用分工后节约的生产技能学习时间来衡量分工的好处,也可以理解为因可用于生产的时间增加而增加的产品或收益。以两个人的分工合作为例(也可以是两个企业或一个个体生产者与一个企业)。假定分工双方进行的工作要求的全部生产技能学习时间为 T ,如果单干,每一方都要付出时间 T 用来学习。利用分工,双方的学习时间加起来等于 T ,正好节约了一个学习时间 T 。分工要能够实行,先要将总的生产技能学习时间在两人之间分解,假定它满足以下关系:

$$T = t^S + t^G = t_1^S + t_2^S + t_1^G + t_2^G \quad (1)$$

式中: t^S 为专用技能学习时间; t^G 为通用技能学习时间;下标表示不同的参与人。

两人分工后的利益 T 如何在分工双方之间分割?我们假定可以用纳什议价博弈来说明。分工协议达成后,双方各自投入生产技能学习时间并合作生产。基于不完全合同的理念,事前合同的不完全性导致事后的再议价行为。由于双方已掌握各自的生产技能,因此都有保留收益,它影响各自对合作利益分割的讨价还价能力。简化分析,假定合作关系一旦破裂,双方各自投入的专用技能学习时间将失去价值,保留收益为各自花费的通用技能学习时间。设双方分别按 x_1 和 x_2 的比例分割分工利益 T ,纳什议价博弈的结果由下式确定:

$$\begin{aligned} & \max(x_1 T - t_1^G) \times (x_2 T - t_2^G) \\ \text{s.t. } & x_1 + x_2 = 1, x_i T - t_i^G \geq 0, i = 1, 2 \end{aligned} \quad (2)$$

可以得到双方的份额为:

$$x_i = \frac{1}{2} + \frac{t_i^G - t_i^S}{2T}, i, j = 1, 2, i \neq j \quad (3)$$

利用(1)、(2)和(3)式,参与人 i 得到的利益为:

$$T_i = x_i T = t_i^G + 0.5t^S \quad (4)$$

保留收益 t_i^G 可以理解为已经定分的利益,即合作破裂后仍然在手的收益。专用技能学习时间 t^S 属于尚未定分的利益,因为双方的贡献缺一不可,无法判定谁的贡献大。纳什议价博弈的结果表明,分工参与人的期望收益等于在各自保留收益之上加上未定分利益的一半。

由(4)可以得到 $\partial T_i / \partial t_i^S = 0.5$,表明参与人 i 每投入 1 单位时间用于专用技能学习,自

己只能得到一半的好处。 $\partial T_i / \partial t_i^G = 1$ 表明,通用技能学习时间的投入没有扭曲。专用性投入激励不足的根源在于套牢效应,利用市场机制组织专用技能成分较高的分工,会因为套牢效应引起的交易费用过大而使分工难以实现。我们来分析市场组织分工的交易费用。

虽然分工合作可以节约生产技能学习时间,但由于双方的机会主义行为,它也带来了相互协调的费用。假定参与人 i 为争夺利益而耗费时间 t_i^N , 它是以讨价还价和契约签订、执行、监督、仲裁等的时间耗费表现出来,称为交易费用。 t_i^N 是一种浪费,因为它既不能用于掌握生产技能,也不能用于生产。交易费用源自于对未定分利益的争夺,未定分的利益越大,争夺的激烈性就越强,因此,可以假定交易费用与未定分利益 t^S 成比例。另外,议价、签约等一般来说是一个双方都在场的行为,不妨假定分工双方付出的交易费用(时间)相等。交易费用简化为以下关系:

$$t_1^N = t_2^N = at^S \quad (5)$$

式中: $a > 0$, 为交易费用系数,它与市场机制的效率有关,随着一国市场机制完善程度的提高而减小。

分工要能够通过市场来实现,要求参与双方在分工合作之前的预期收益均为正。分工的预期收益等于分工利益扣除交易费用后的余额,利用(4)和(5),分工实现的条件为:

$$E_i = T_i - t_i^N = t_i^G + (0.5 - a)t^S \geq 0, i = 1, 2 \quad (6)$$

式中: E_i 为参与方 i 的预期收益。不难看出,当 $a > 0.5$ 时,上式有可能不成立;两人的工作要求的专用性投资的相对比例越高,分工实现的条件越不容易得到满足。同时,分工实现条件满足与否取决于通用技能学习时间投入较少的一方。如果双方分担的总的生产技能学习时间大约相当,那么专用技能学习时间投入较多的一方容易受到要挟,处于劣势地位。

三、利用企业组织分工的费用

依据前面的分析,市场是一个掌握通用技能的一方容易要挟掌握专用技能一方的机制,这一特征使得若干分工难以利用市场机制来实现。企业是一个制度创新,它将合作体的分工组织和利益分配权给予参与的一方,使一些在市场条件下不能实现的分工得以实现,为整个经济系统的分工向更高的水平推进提供了可能。企业的特点可以概括如下:①利用权威的方式决定分工组织和利益分配,自由议价的费用几乎降至0;②在企业的内部存在官僚、腐败、影响活动等费用,它的大小取决于企业制度的效率;③企业内部进行的分工在市场上一般不存在,否则将失去企业专用性的意义。

我们来考察利用企业组织分工的条件。不妨假定参与人 i 掌握企业的剩余控制权和剩余索取权,并决定如何进行分工和收益分配,称为企业组织者。在企业中,参与人 j 的行为方式发生变化,在其努力和投入难以测定的情况下,企业组织者 i 要花费时间建立一套激励和监督机制。参与人 j 甚至花费时间通过影响企业组织者 i 的决策来提高自己的收益。总之,双方仍然要花费时间来处理分工参与各方的利益协调。从生产的角度来说,这些时间的花费也是一种浪费,它构成利用企业来组织分工的费用,不妨用 t^M 表示。

假定参与人 j 用于企业专用的生产技能学习时间需要得到补偿,以此作为参与条件。用企业管理费用 t^M 替换市场交易费用 t^N , 由 (1) 不难得到利用企业组织分工的条件:

$$\begin{cases} E_i = T - t^M - E_j = t^S + t^G - t^M - E_j \geq 0 \\ E_j \geq t_j^S \end{cases} \quad (7)$$

企业组织者 i 优化其收益,有 $E_j = t_j^S$ 。代入 (7) 式后,得到利用企业组织分工的条件为:

$$E_i = t_i^S + t^G - t^M \geq 0 \quad (8)$$

可以看出,专用生产技能比例较大反而成为促成分工实现的有利条件,使得一些原本在市场条件下不能实现的分工可以用企业来实现。如果一国的企业制度比较完善,表现为企业组织分工的费用 t^M 较小,那么会有较多的分工利用企业来组织。

四、我国产业集群的分工优势

我国东南沿海地区,尤其是长江三角洲的一些经济相对发达的地方,中小城镇(包括乡镇)既没有大城市在市场规制上的优势,也没有现代企业制度的传统,它们只能集中力量,以产业集群的方式生产一种或少数几种产品。在产业集群内部,以家庭或小企业为独立生产单位,每一个生产单位在整个产品的产业链中只占据一个或少数几个分工环节,而产业集群作为一个整体则由众多的生产单位构成一个高度分工专业化的网络体系。这种模式在经济机制上产生两种效应,不妨称之为“集群锁定效应”和“专用投资通用化效应”,它们恰好能够克服市场不规范和企业制度不完善的弱点。

在我国中小城镇产业集群内部,每个生产者由于高度专业化加上直面市场竞争,具有很强的竞争力。可是一旦离开产业集聚区,由于掌握的生产技能过于单一,生产者会因缺乏相应的通用生产技能而难以生存。这种集群锁定效应能够有效地克服人们的机会主义行为。产业集群对于其中的生产者产生的另一个作用是专用投资通用化效应,它能够有效地克服因没有良好的企业制度而难以保护专用性投资利益的缺陷。在产业集群外部看来是专用性的生产技能,在产业集群内部则不然。一般情况下,(纵向)分工每深化一步,都会形成单个买方对单个卖方的双边垄断格局,机会主义带来的交易费用在所难免。可是当分工以多个买方对多个卖方的方式进行时^①,专用技能将以通用技能的形式表现出来,因为针对某个特定交易对象进行的投资可以有较大的部分转用于同其他交易对象进行合作。当然,由于技术创新和产品多样化趋势增强,总有部分生产技能是双边专用的,不合作即作废。

假定在双边垄断条件下的生产专用技能,在多对多的交易条件下有比例 b 部分可以转化为通用技能,满足 $0 \leq b \leq 1$ 。显然,产业集群内的同一分工层次的独立生产者越多, b 就越大。当双方不合作时,各自实际损失的专用技能学习时间为 $(1 - b)t_i^S$ 。从双方讨价还价的角度考虑,可以转作他用的生产技能学习时间属于保留收益,其数量分别为 $t_i^G + bt_i^S$ 。由有保留收益的纳什议价博弈可以得到双方不考虑交易费用时的利益为:

^① 本文没有考虑分工初始出现的情形。如果生产效率的提高可以替代他人的市场份额,那么初始分工的利益要大于学习时间的缩短。在产业集群里,因为同一类的竞争者较多,分工初始出现的条件一般容易满足。

$$x_i T = t_i^G + bt_i^S + 0.5(1 - b)t^S, i = 1, 2 \quad (9)$$

按照交易费用来自对未定分利益的争夺，而未定分利益用不能通用化的专用技能学习时间来量度，类似 (5) 式，交易费用为：

$$t_1^N = t_2^N = a(1 - b)t^S \quad (10)$$

考虑交易费用因素后，由 (9) 和 (10) 式，分工参与双方的预期收益为：

$$E_i = x_i T - t_i^N = t_i^G + bt_i^S + (0.5 - a)(1 - b)t^S, i = 1, 2 \quad (11)$$

不难看出，双方的专用性投资激励都增强了，通用性投资激励没有扭曲。当双方参与分工的预期收益都不小于 0 时，利用市场实现分工的条件得以满足。比较 (6) 式可知，在产业集群内部实现分工的条件更容易得到满足。

一个国家的市场机制比较完善，分工得以深化，企业制度比较完善，分工也能够深化，如果两者都比较完善，分工就可以达到更高的水平。我们面临的问题是，如果既没有市场机制优势，也没有企业制度优势，靠什么来深化分工水平？中国的经济实践回答了这个问题，这就是利用产业集群来深化分工。从一定的意义上说，完善的市场规制和先进的企业制度并不是一国经济增长的必要条件，这也可以看作是揭示“中国经济增长之谜”的一个思路。

参考文献

- [1] Fujita, M., A monopolistic competition model of spatial agglomeration: A differentiated product approach, *Regional Science and Urban Economics*, 1988, Vol. 18, pp 87 ~ 124.
- [2] Krugman, Paul., Increasing returns and economic geography, *Journal of Political Economy*, 1991, Vol. 99, pp 484 ~ 499.
- [3] Porter, M., *The Competitive Advantage of Nations*, Macmillan: Landon, 1990.
- [4] Venables, Anthony J., Equilibrium locations of vertically linked industries, *International Economic Review*, 1996, Vol. 37, pp 341 ~ 359.
- [5] 威廉姆森：《交易费用经济学：契约关系的规制》，见：《企业制度与市场组织——交易费用经济学文选》，上海三联书店、上海人民出版社 1996 年版。
- [6] 黄曼慧，黄燕：《产业集聚理论研究述评》，《汕头大学学报》（人文社会科学版），2003 年第 1 期。
- [7] 刘乃全：《产业聚集理论及其发展》，《上海财经大学学报》，2002 年第 2 期。
- [8] 李小建，李二玲：《产业集聚发生机制的比较研究》，《中州学刊》，2002 年第 4 期。
- [9] 苗长虹，崔立华：《产业集聚：地理学与经济学主流观点的对比》，《人文地理》，2003 年第 3 期。

制度性规则建构方法研究综述*

杨永福 段红涛 孟磊

20世纪80年代以来,全球社会的理论家们致力于寻找不同国家在人力资本的发展水平和物质资本的积累方面所出现的速度和效果差异的原因。其中,有些学者把关注的目光投向了制度与规则建构的过程,以及孕育它们的宽广、复杂和千变万化的社会环境。本文将介绍和评述其中最具有代表性的、影响广泛的几种方法。

一、波兰尼与格兰诺维特的“嵌入性”方法

经济史学家波兰尼(Polanyi, K.)主张,经济学家在做研究时,必须考虑人类的行为与心理因素,同时要在广泛的社会、文化、政治环境中进行制度分析。他主张:交易、货币、市场都是制度。在对经济、社会与文化交互影响的实证研究中,波兰尼建立了他的市场交换和嵌入(embedding)理论。

在《大转变》(The Great Transformation)一书中^[1],波兰尼从经济史的角度对交易(trade)、货币(money)、市场(market)等范畴作了深入的剖析。他认为,一种经济组织的存在是以承担特定的社会功能为前提的;人类社会的各种因素对于经济的影响都不可忽视,有必要研究经济过程在不同时空、不同条件下得以制度化的方式。与18~19世纪流行的假说相反,商业和货币起源于人类的史前时期,而市场则是近代才发展起来的。所以,把交易、货币、市场视为本质上不可分割的三组元素是对产生于市场经济观念的误解。在吸收马林诺斯基(Malinowski, B.)有关经济人类学研究成果的基础上,波兰尼提出了三种类型的交换模式,即“互酬”(mutual reward)模式、“再分配”(redistribution)模式和“市场交换”(market exchange)模式。我国有学者认为,“互酬”指对称集团间相应物品和服务的互动;“再分配”指物品的权利转让集中于社会的某一中心,通过中心组织进行再分配;“交换”则指在市场体系下,物品以货币为媒介在不同所有者之间的流动^[2,3]。这三种模式交织在一起,可能同时存在于一个社会中,也可能其中某一种方式在某个社会的某个层面或者某个时候占主导地位。一般来说,古代社会以“互酬”与“再分配”为主要模式,现代社会显然以“市场交换”为主。

不同的交换模式要求不同的制度性支撑。“互酬”要求存在对称性质的集团;“再分配”须有“社会中心”的存在;“市场交换”对应的是现代意义的市场体系。波兰尼认为,这三种交换模式“嵌入”在相应的社会制度中。并且不是个体行为的简单加总。制度的差

* 原载:《经济学动态》,2003年第6期;全文转载于中国人民大学书报资料中心复印报刊《理论经济学》,2003年第9期。

异决定了交换模式的不同：“互酬”模式 \longleftrightarrow 对称组织；“再分配”模式 \longleftrightarrow 社会中心；“市场交换”模式 \longleftrightarrow 市场体系。在前资本主义社会^[2]，“互酬”和“再分配”是资源配置的主要形式，经济演化的逻辑嵌入社会联系网络，“由于社会组织自身的稳定性，无需个体经济发挥作用。经济制度事实上只是作为社会组织的功能之一”。只是近代以来，“市场交换”模式才开始发挥主导作用。“互酬”模式、“再分配”模式和“市场交换”模式都能发挥资源配置的作用。因而，贸易和交换自身并不能解释近代自我调节市场制度的出现。只有经济体制的制度演化理论，才能对近代市场出现的制度基础及其根源作出有说服力的解释。现代市场经济中，许多经济制度从表面上看似乎不可分割地连接于一体。事实上，他们的起源各不相同。这种连接不过是在掩盖着各自不同的本质起源基础上的一种功能性连接而已。

波兰尼市场交换理论的基础是他率先提出的“嵌入性”(embeddedness)方法。在《作为制度过程的经济》^[4]一文中，波兰尼指出：人类的经济活动嵌入并缠结于经济与非经济的制度之中。把宗教和政府作为要素放到经济研究过程中去是非常必要的，因为“宗教和政府可能像货币制度或减轻劳动强度的工具与机器的效力一样重要”。因此所谓“嵌入”^[5]，就是行动者并非如原子个体一般地在社会网络之外行动或者决策，他不会一成不变地恪守其社会角色的职责。他力图进行的有目的的行动是嵌入在具体的、当前的社会关系网络中的。

但在古典和新古典经济学家看来，经济活动中的社会关系是市场运行的障碍物。斯密(Smith, Adam)及其后继者认为，人本能地具有用一种东西去换取、交易或者交换另一种东西的倾向；再之，由于劳动是原始社会中惟一的生产要素，物品是按其劳力成本的比例来交换的，所以个体的经济行为是独立于其社会关系的。社会制度及人的社会行为是由理性的原子化个体的自利追求而生成的东西。与此相反，帕森斯(Parsons, T.)及其后继者在回答“社会秩序如何可能”时则主张，由于社会个体非常在乎他人对自己的看法，所以他会遵守各种为大家一致公认的社会规则。这些规则通过社会化机制内化于每个社会个体的心中，成为个体行为的有机组成部分，且并不觉得它们是对自己的束缚。这种观点强调行动个体在社会环境中的嵌入深度，但它夸大了嵌入的程度，以致连冲突也被淡化了。

在社会经济学家的视野里，斯密的“原子化个体”理论是社会化不足的个体理论，帕森斯的“个体”理论是社会化过度的个体理论。这两种个体理论都不足以解释个体的行动与其所在的社会现实的关系。波兰尼的“嵌入”理论主张，个体行动者既不外在于社会环境，也不固执地坚守其既有的、普遍的社会规则与信条，而是“嵌入”于具体的、当前的(present)社会关系网络中；正是在这种“本土化”的格局中，社会个体作出符合自己目的、能实现自己愿望的选择。

经济的社会嵌入性概念是新经济社会学的重要表征。“嵌入”及其理论取向是多元的：有人认为经济是更大制度结构的有机组成部分，有人认为经济行动是在构成社会结构的社会关系网络中发生的，有人认为经济行动不仅在社会结构而且是在文化中发生的。因此，应该区别出认知的、结构的、文化的和政治的不同嵌入方式。它们综合在一起，有利于我们全面认识经济现象。因为经济行动依靠这些类型的因素。嵌入性理念对新古典经济学中忽视社会情感联系、不受个人情感影响的理性的行动者的观点提出了挑战，同时引入了社

会资本的概念。

所谓社会资本，是指存在于社会结构之中，通过促使行动者进行交易与协作等特定活动而产生效益的资源，这些资源表现为社会网络及某些特征，诸如信任、关系与规范。科尔曼 (Coleman, J.) 在《社会理论的基础》一书中一再指出^[6]，蕴含着行动者利益的事件，部分或全部处于他人的控制之下，行动者为了实现自己的利益而进行相互交换，进而形成了持续存在的社会关系。这些社会关系既是社会结构的组成部分，也是一种个人资源。必须强调的是，社会资本指一种结构资源，蕴含于结构本身，而不是像社会资源理论所指的通过关系网络可以汲取的实际资源。依据这种结构，行为主体可以通过两种社会联系获取 (access) 他所需要的资源，如权力、地位、财富、资金、学识、机会、信息等等。第一种社会联系是个体作为社会团体或组织的成员与这些团体和组织所建立起来的稳定的联系，个体可以通过这种稳定的联系从社会团体和组织获取稀缺资源。比如，通过单位可以得到住房，通过校友会可以获得工作机会，通过学会可以了解国际最新学术动态等等。对这种“社会成员关系”所形成的社会资本，政治学者尤为重视，其代表人物是哈佛大学的普特南 (Robert Putnam)。第二种社会联系是人际社会网络。与社会成员关系不同，进入人际社会网络没有成员资格问题，无须任何正式的团体或组织仪式，它是经过人们之间的接触、交流、交往、交换等互动过程而发生和发展的。社会学者重视社会网络以及个体由社会网络获取社会资源的过程。格兰诺维特 (Mark Granovetter) 曾阐述了劳动力市场的信息如何通过社会网络而传播，人与人之间的信任感是如何经过社会网络而建立、巩固和发展，并在《经济行动与社会结构：嵌入性问题》^[7]一文中提出了“嵌入性”问题。在这篇文章里，格兰诺维特指出应从“行动”的视角出发，用“嵌入”概念来对经济现象进行社会学意义上的解释，阐释了社会关系网络在经济生活中的重要意义。关系的嵌入被理解为一种约束条件。一些学者认识到了个体与市场的社会嵌入问题，认为在一个嵌入的网络关系中，这样的交换有助于维持网络的稳定性。从此，“嵌入性”方法得到广泛的重视。

格拉诺维特理解的“嵌入性”不像波兰尼认为的“经济活动嵌入于制度”，而是侧重于把经济活动放到人际关系——即社会网络——的背景下予以考虑。比如市场。在经济学的供求分析中，假定供求双方可以很顺利地见面，而经济理论认为供求状况主要依靠价格信号来作出调整。但格拉诺维特认为，这个命题忽略了匹配问题，即供给和需求究竟是怎么相遇的问题。若仅仅有斯蒂格勒 (Stigler, George J.) 的“搜寻是要付出成本的”观点也忽略了供求相遇的机制。这一点是不够的，因为这一命题没有把机制说出来。因此，格拉诺维特主张：市场的供给与需求在很大程度上是通过人际关系网络来解决的，并对人际关系网络作了进一步的分析。他指出，人际关系网络有两类：一类是强关系，另一类是弱关系。常常在一起的是强关系；偶尔联系的、互动频率低的是弱关系。他发现，在劳动力市场中，主要是弱关系在发挥作用。他解释道：通过人际网络能解决信息问题。但信息本身是有质量差异的，即它涉及信任之类的问题。因而，在人际网络中的信息质量通常比在陌生人之间的信息质量要高。在人际网络中传递信息的时候，如果是强关系，一般来说，信息的同构型较高。这样就不太容易建立供求之间的匹配关系。如果是弱关系，信息的异质性就会比较高。所以，供求就是要解决信息差异问题。处于社会网络中的厂商是在经过互相接触、交谈和观察后方达成交易的，并逐步建立起信任关系和市场制度。

波兰尼的“嵌入”理论在伯恩斯 (Burns, Tom R.) 的“行动者 - 系统动力学理论” (Actor-System Dynamics, ASD) 中有更进一步的发展。ASD 强调^[8], 行动者通过有目的地收集、组织、筛选信息, 并以社会规范——包括对社会控制的组织与规范——为基准调适 (adjustably-integrate) 个体行为, 进行角色扮演, 寻求期望的收益。行为者既是社会规范的执行者, 又是实施这些规范所必须的实践经验的载体, 同时可能对社会规范和行动环境作出新的有时甚至是出乎意料的个体的解释。更为普遍的是, 个体行为者能够改变环境, 并在重建社会规范、社会制度、创建新的技术及工艺过程中, 展示这种能力。行为者在行动和互动几率 (包括资源控制) 分配上的差异, 决定了他们在某一场景中的相互权利的大小及影响未来发展的能力。社会行动者的行为及其互动能够规范并改变自然、制度和文化环境, 但同时亦受到限制。

二、格雷夫的历史制度分析方法

历史制度分析 (Historical Institutional Analysis, HIA) 方法主要运用现代经济学的分析工具, 在历史比较博弈分析的理论建构中, 从制度的起源处对“文化信念” (cultural belief) 与社会制度的发生与构型关联机制做出解释。其代表人物是斯坦福大学的格雷夫 (Greif, Avner) 教授。他认为^[9], 在研究制度之前不能先入为主地给制度下定义。从博弈论的角度看, 制度起源于积极的文化信仰, 是自发演化的产物。他主张在处理制度之前, 应该首先处理组织。因为组织的导入反应了知识存量的增长, 并将引致有意识的追求或是无意识的产出。他指出, 一个社会的组织——它的经济、法律、政治和社会及道德强制制度——是伴随着它的社会建构、信息传递和协调机制, 并且对长期的制度演化具有非常重要的影响。

在对中世纪商业行会 (merchant guild) 的研究中, 格雷夫发现, 行会之所以作为一个制度普遍存在于许多城市, 是因为存在国家困境。国家困境是指作为强制契约和产权并提供公共物品制度的国家也具有权力抑制保护或是强制征收私人财富, 并削弱市场经济的基础。但行会存在的基础是商人群体在交易过程中自发形成的共同知识、诚信以及行业文化。有了这个基础, 行会才能够广泛地在不同的商人集团之间建立信誉传递机制, 并对行会成员建立诚信约束和道德强制。在多边声誉机制形成的过程中, 形成了分布式的信息交流网络, 长距离的贸易才成为可能, 产权最终才得以建立。也就是说, 产权是网络建立的结果。虽然不能说, 没有网络就没有产权, 但至少在最底的限度上却可以说, 没有网络, 产权就缺少稳固的并得以有效发挥作用的社会基础。

格雷夫创立的 HIA 方法具有如下的一些特点:

首先, 它是一种方法论。在《分析性叙述》 (Analytic Narratives)^[10]一书中, 格雷夫认为, HIA 是从历史叙述开始但不断返回历史叙述的过程。在这一过程中, 我们不仅可以看到博弈过程的演化, 而且能够看到每一次均衡所需要的条件和约束, 由此可以更加深刻地理解制度演化的路径依赖过程 (institutional path dependence process)。在格雷夫看来, 制度是自我强制的 (self-enforcing)、非技术性的行为约束。这种约束通过预期和组织而得以实现。所以预期和组织是 HIA 的两个主要的、交互性 (reciprocity) 的构成因素。显然, 一个

博弈者对其他博弈参与者行为的预期是非技术性的约束，而且这些约束是每一个博弈参与者都不能避免的。组织通过引入新的参与者影响博弈结构中的信息交流、博弈者的行为和各自的收益权衡。

其次，HIA 基本上是均衡分析，但它超越博弈论的经验均衡研究，强调制度选择过程和制度演化的路径依赖，认为制度选择过程和演化是一个历史的、耗时而复杂的非各态历经过程（nonergodic process）。

再次，HIA 特别强调国家的制度基础。格雷夫对热那亚（Genoa）和马格里布（Maghrib）的一系列研究表明，初始的文化信仰对于组织的演化和国家的出现是非常关键的内生变量。历史上，热那亚（Genoa）是拉丁世界的一部分，是后罗马帝国时期的一个城邦。马格里布（Maghrib）是穆斯林世界的一部分，地域覆盖现在的摩洛哥、阿尔及利亚和突尼斯大部分地区。经济史的研究发现，在 11~12 世纪，热那亚和马格里布商人面临同样的贸易环境，运用同样的航海和陆地交通技术，并差不多进行同类商品的贸易。但是，初始的文化信仰却使热那亚演变成为高度分权化的、以个人主义为基础的社会，并且发展出契约、长距离贸易、行会、产权；在穆斯林世界，由于信奉集体主义，只认可组织内部的通信、道德和诚信，因而成为集权主义社会，形成了集权主义国家。

最后，由于制度具有文化信仰的基础，制度变迁的过程就是知识增长的过程。它体现出来的特征，首先是斯密式的经济增长，但却要通过运用国家强制力影响经济运行的效率。HIA 正在修正诺斯（North, Douglass C.）对欧洲经济增长过程的结论。更富戏剧性的是，HIA 已经促使诺斯对他的理论基础进行反思。

三、奥斯特罗姆夫妇的多中心方法

多中心理论为美国印第安纳大学政治理论与政策分析研究所的奥斯特罗姆（Ostrom, Elinor & Ostrom, Vincent）教授夫妇共同创立。多中心理论提出了对社会秩序的全新洞见^[11]，发现了公共领域中的“看不见的手”，在市场秩序与国家主权秩序之外发现了社会运转的多中心秩序。传统理论认为，权力分散与交叠管理是“不正常”现象，它既是政府失灵的表现，也是政府失灵的根源。多中心理论否定了这个简单化认识，揭示了这一领域的复杂性及其规律。公共领域背后的这只“手”有其自身的机理。多中心理论正是对这一机理的阐述，它在市场与政府之外发现了存在于社会之中的秩序与繁荣涌现（emergence）的力量，并进一步发掘了使这种力量得以持续的制度性基础。

“多中心”一词首先由波兰尼所使用。他区别了组织社会任务的两种方法或者两种秩序。一是设计的或者指挥的秩序，它为终极的权威所协调，该权威通过一体化的命令系统实施控制。在设计或者指挥的秩序中，上级 A 指挥下级 B_1, B_2, B_3, \dots 执行特定的任务、完成特定的使命。这样的秩序可以概括为一元的或者单中心的秩序。另一个是“自发的”或者多中心的秩序，它指许多因素的行为相互独立，但以一般规则体系——如宪法——为依据而相互调适、激励和约束。

多中心体制设计的关键因素是自发性。波兰尼认为“自发的”与“多中心的”同义。自发意味着多中心体制内的组织模式在个人有动机创造或者建立适当的有序关系模式的意

义上自我产生或者自我组织。多中心体制在发展有序关系方面是“自发的”，自我组织的倾向在若干不同的行为层面上就必然发生。奥斯特罗姆的多中心理论是一个三层面、多条件、多组织的结构体系。

三个层面间的关系是：在操作层面，具体行动由直接受影响的个体或者政府官员实施。这些行动以某种明显的方式直接影响社会，并产生明显的政策结果；界定与制约操作领域中个体公民和官员行动的原则由集体选择层面的程序决定；而修改这些规则的更基本的规则由立宪选择层面确立。这样，第一个层面，即操作层面应用于进入和退出特定的多中心秩序中的条件。就简单市场来说，单个的个体可能不是富有活力的市场参与者。如果是这样，市场作为多中心秩序的活力就得依靠个体是否有激励形成有效参与市场活动的企业。第二个层面，即集体选择层面的组织应用于一般行为规则的实施，它为多中心秩序提供了法律框架。如果运作于多中心秩序中的个体或者单位有动机采取行动，实施一般的行为规则，多中心作为一种组织形式会日益富有活力。第三个层面，即立宪选择层面的组织涉及基本行为规则的制定和修改，它为任何特定的多中心秩序提供框架。如果个体知道特定规则及其在具体条件下可能引起的社会结果之间的关系，那么具体的多中心秩序就可以被看作是一种设计的事务。如果条件发生了变化，而特定组别的规则没有作出适当的反应，这种规则就有可能被改变。这时就可能发生制度创新。

具体到社会生活层面，面临的是大量的公共池塘资源的治理问题。所谓公共池塘资源(Common-Pool Resources, CPR)，是指这样的资源，它们犹如无主而开放的池塘之水，谁都可以取。比如，公共服务、公益性活动、公共财产资源或公有资源就是这一类性质的资源，它们在本质上与 CPR 并没有泾渭分明的差别。这类资源一经为“私人”取得，就可能成为“私人”拥有、为“私人”享用的物品。所以，公共事物的治理向来是一件令人头疼的事情。一谈起它，人们总是免不了要想到哈丁(Hardin, Garrett)的“公地悲剧”、“囚犯悖论”和奥尔森(Olson, Mancur)的“集体行动的逻辑”困境等，浮现出一群自私的“理性人”在搭便车和机会主义行为的泥潭里不能自拔、孤苦无助的场面。

但奥斯特罗姆认为，问题并不在于人们的行动本身，而在于常用的集体行动理论既不充分，亦难以理解。所谓不充分，是指这些理论并没有把解释自主组织所需的内部变量和外部关键变量纳入考虑之中；而难以理解则指已为人们熟知的集体行动理论并没有提供明确的公共政策建议。它们不适用于小规模公共池塘资源中的制度变迁。其不足在于：没有反映制度变迁渐进性和制度自主转化的本质；在分析内部变量如何影响规则的集体供给时没有注意外部政治制度特性的重要性；没有包括信息成本和交易成本。这三方面的因素对解释成功的自主组织和自主治理的制度安排具有重要意义：① 分权制企业表明规模较大的公共池塘资源中的人们是通过这样的逻辑来成功地采取集体行动的，即先建立较小的组织作为基础，进行不断的建构和发展。经验表明，这样做的成本要大大低于在没有任何基础的前提下建构组织的成本。而由于最基本组织的成功，则可使人们得以在已经创立的社会资本基础上通过更大、更复杂的制度安排来解决较大的问题。现行的制度变迁理论没有强调制度资本的自然增长过程，不重视在制度供给过程中经常涉及的渐进的自主转变。组织学习的重要性就在于它是一种组织性的自主转变过程。② 案例研究表明，外部政治制度对于生产者进行自主性的活动，可能是积极因素，也可能是消极因素。所以，有必要考

虑外部力量的策略对公共池塘资源占用者的成本和收益的影响。③ 信息成本和交易成本的模型是以完全信息、独立行动、信息完全对称、没有失误,并且缺乏可接受的行为规范、零监督成本和实施成本,以及环境本身不具备改变结构的能力等作为假设。但实际情况并非如此。公共池塘资源的环境很少能像竞争市场那样推动参与者追求效率,也不存在任何像竞争市场价格那样的单一变量,能作为公共池塘资源环境中进行理性选择的基础。所以,科学的态度是不把规则变更的决策视为机械的计算过程,而是把制度选择视为对不确定的收益和成本进行有根据的评估过程。

奥斯特罗姆夫妇指出,在研究制度变迁问题时,不能把制度起源和制度变迁看作是两个不同的东西,应把它们纳入统一的框架,用同样的理论分析它们,如把宪法选择和集体选择统一起来进行思考。他们主张,集体行动分析不能停留在操作层面。只有规则的使用、供给、监督与实施才与操作层面相关。影响集体行动的制度不止这一层。宪法决策与宪法层面相关,决策的制定与管理涉及集体选择层面。三个层面的行动规则中,层面越高,变动成本越大,越难完成;一个层面的行动规则变迁是在其上一层规则的框架中发生的。还有,非正式规则也能产生有效的影响。这样,非正式规则就进入了制度分析的视野。

社会中蕴藏着丰富的自组织和自主管理的资源。政府的全面管制和到处插手可能酝酿着阻碍这种资源发挥效能的因素。当我们试图反向操作,即把已经被阻碍的自主能力恢复过来,无疑会遇到巨大的挑战。这种挑战来自奥斯特罗姆夫妇所描述的情景:几乎所有人都处在面对搭便车、规避责任或其他机会主义行为的诱惑的情形下。正是这种情景,使得奥斯特罗姆夫妇的研究显得特别珍贵。在重建合作制度时,我们建立较小的次级组织的资源是很丰富的,比如浓厚的血缘关系、亲情和邻里关系、社区的互助等。艾克斯罗德(Axelrod, Robert)在他的博弈论研究中^[12-14]表达了类似的思想。他指出,一个合作的小群体有极强的鲁棒性(robustness),即使在不利的环境中也能够发展壮大,而且,如果条件合适,它甚至能够最终改变整个群体,并确定有关规则。

四、结语

从上述的介绍和评述中可以看出:人类的经济活动嵌入并缠绕于经济与非经济的制度之中,即个体行动者既不外在于社会环境,也不固执地坚守其既有的、普遍的社会规则与信条,而是“嵌入”于具体的、当前的(present)社会关系网络中。这个网络有其内部结构,即规则的使用、供给、监督与实施与操作层面相关;宪法决策与宪法层面相关,决策的制定与管理涉及集体选择层面;而且三个层面的行动规则中,层面越高,变动成本越大,越难完成;一个层面的行动规则变迁是在其上一层规则框架之下发生的。其间,有着复杂的、本土化的信息传递模式和协调机制。正是在这种由历史和文化积淀的格局中,社会个体和群体作出符合自己目的、能实现自己愿望的选择。

参考文献

[1] Polanyi, K., *The Great Transformation*, Beacon Hill Press, Boston, MA, 1957.

- [2] 林义：《制度分析及其方法论意义》，《经济学家》，2001年第4期。
- [3] 朱国宏：《经济社会学》，复旦大学出版社1999年版，第228页。
- [4] Polanyi, Karl, *The Economy as Instituted Process*, In *Primitive, Archaic and Modern Economics: Essays of Karl Polanyi*, Beacon Press, 1968, pp139 ~ 174.
- [5] 朱国宏：《经济社会学》，复旦大学出版社1999年版，第356~359页。
- [6] 詹姆斯·科尔曼：《社会理论的基础》，中译本，社会科学文献出版社1999年版。
- [7] Granovetter, Mark, *Economic Action and Social Structure: The Problem of Embeddedness*, *American Journal of Sociology*, 1985, Vol.91, pp481 ~ 510.
- [8] T.R. 伯恩斯：《结构主义的视野》，中译本，社会科学文献出版社2000年版，第4~5页。
- [9] 秦海：《阿弗纳·格雷夫的历史制度分析》，《经济社会体制比较》，2001年第2期。
- [10] Greif, Avner, et al., *Analytic Narratives*, Princeton: Princeton University Press, 1999.
- [11] 文森特·奥斯特罗姆：《多中心》；文森特·奥斯特罗姆，埃莉诺·奥斯特罗姆：《公益物品与公共选择》；埃莉诺·奥斯特罗姆等：《规则、博弈与公共池塘资源》，上海三联书店2001年版。
- [12] Axelrod, Robert, *The Evolution of Cooperation*, New York: Basic Books, 1984.
- [13] Axelrod, Robert, *The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration*, Princeton, NJ: Princeton University Press, 1997.
- [14] Axelrod, Robert and Hammond, Ross, *Evolution of Ethnocentric Behavior*, www.personal.umich.edu, 2003.

社会资本结构与民营企业成长^{*}

王 □ 姚海林 赵 祥

一、引言

在我国民营企业数量得到较快增长的同时，我们也关注到了这样一个现实，即我国民营企业生命周期普遍较短。2001年国家统计局城市调查队公布的数据显示，我国私营企业平均存活的时间为5.7年。其中在1988年前已经成为“雇工大户”的占9.2%，1989~1997年间成立的占48.3%，1997年以后成立的占42.5%^①。虽然每年新注册的企业数量大于倒闭数量构成了我国市场化进程中民营经济规模不断增大的宏观特征，但是，从微观角度看，任何一个企业都不愿意在较短时间内走完它的周期。是什么因素对我国民营企业持续性的发展产生着重要的影响？观察我国成功与失败企业的实践，我们认为，成功企业的原因是基本相同的，不成功的企业，其原因可能是各不相同的。与其解释各不相同的原因，不如从经营比较成功的企业中找出相同的因素，这对大多数民营企业延长存活时间、提高可持续发展的能力会更有意义。

目前对这个问题存在着三种解释：一种观点是用钱德勒给定的企业家式企业向公司制企业转变的制度范式来解释一部分民营企业走向成功的原因，这可以称为企业制度理论的解释。但是，在家族企业向公司制企业转轨的制度逐步完善过程中，有的成功，有的不成功，珠江三角洲地区一部分家族企业雇佣外来人作为经理，因受骗而蒙受损失的例子比比皆是。反过来，在仍然沿用家族方式进行管理的企业中，成功与不成功的实例也都存在。即使在美国这样市场经济比较发达的国家里，家族企业也占企业数量的95%以上。因此，用一个趋势性因素来区别某一个时段内的企业绩效差别是缺乏解释力的。另一种观点是政策性解释，即用政策因素解释民营企业的绩效差异，甚至一些学者把民营企业看成是“政策性经济”。虽然从限制、允许到现阶段扶持的全国政策与民营经济成长数量与规模是相关的，但是在某一个时期内，同样一种政策却有着不同的企业绩效。此外，在过去一段时间内，尽管不同地区制定的政策差别也对民营企业发展产生了不同影响，但是，各地区之间的不同政策是最容易通过模仿走向趋同的，在各地“争着做出让步”以至于形成政策趋

^{*} 本研究是2002年国家社会科学基金项目“企业集群创新过程研究”的一部分，该项目的批准号为02BJ1024。同时，本项目也得到了广东省社会科学规划基金、广东省软科学基金的资助，在此一并感谢。原载：《中国工业经济》，2003年第9期。

^① 国家统计局城调队：《中国民营企业的平均寿命》，《商界》，2002年4月。

同的环境下，民营企业经营与发展状况也仍然存在着较大的差别。因此，政策因素也不是关键。还有一种观点是社会资本理论的解释，即从社会资本角度来探究民营企业的发展期限与盈利差异（李路路，1998），也就是说，民营企业主在“下海”前的职业背景、亲戚朋友的关系网络等因素对其获取机会与资源都产生着重要的影响，进而导致企业之间的绩效差异。显然，在区分“成功与不成功”的民营企业中，社会资本理论作为个性化分析比制度与政策等共性解释具有较强的解释力。但是，这仅限于从社会资本的民营企业主个人关系的狭义视角来研究其对经营绩效的影响程度，这种视角难以阐释我国民营企业的成功数量不断增加，因为虽然每个人都有自己的社会关系网络，但并不是每个人的关系网络都具备获取机会与资源的能力。既然社会资本是民营企业走向成功的一个重要因素，那么，缺少这种社会资本的创业者在创业和经营过程中就会通过投资来编织这种能够带来机会与提高资源动员能力的关系网络。如果编织这个关系网络所支付的交易成本过高，那么，缺少这种社会资本的企业主会降低对此投资的热情，进而影响民营企业走向成功。从宏观角度看，如果缺乏这种社会资本的企业难以走向成功，那么，社会中具有创业能力的人就不愿意进入民营企业行列，其结果，民营企业发展数量就变得十分有限。显然，这种逻辑推理与我国民营企业的发展现实是不一致的。因此，本文从广义的社会资本入手，把社会资本分为两部分，即个人关系资本与集群网络资本，解释其各自对民营经济发展的支撑作用。再通过这两种社会资本的动态转换分析，揭示社会资本的结构变动与我国民营企业持续成长的趋势性联系。

二、社会资本及其分类

社会资本是20世纪80年代以来在西方社会学、经济学和社区发展领域被广泛使用的概念，它用来解释“处于网络或更广泛的社会结构中的个人动员稀有资源的能力”^①。根据这个定义，林南（2002）把它概括为三种成分：镶嵌于一种社会结构中的资源；个人涉取这些社会资源的能力；通过有目的地运用和动员这些社会资源以获得更大的回报。^②按照林南的说法，社会结构中的资源数量与特性决定着个人动员社会资源的数量与特性，而回报程度则直接是行动者动员社会资源的结果。因此，研究社会资本的关键在于解释社会结构中的资源是一种什么样的资源，这种资源是由什么样的社会结构产生出来的，这是确定个人动员社会资源的前提。首先，社会结构中的资源指的是蕴含在人际关系网络中的资源，如信任、规范和网络等，它能够通过推动协调的行动来提高社会的效率（R. Putnam, 1993）。信任关系决定着整个社会结构中的每个行为者从事经营活动的环境与风险水平。规范则是为一般社会群体成员所接受了的行为习惯与规则。网络是行为者在参与互动中使资源得到聚集、维持和强化的一种有效途径。无论什么样的社会结构，都存在着

① Alejandro Portes, *The Economic Sociology of Immigration: A Conceptual Overview*, In Portes (ed), *The Economic Sociology for Immigration: Essays on Network, Ethnicity and Entrepreneurship*, New York: Russell Sage Foundation, 1995, p12.

② 这种回报包括了直接获得收益的经济回报、改善政治经济地位的政治回报以及增强社会影响力的社会回报等。鉴于本文的研究范围，我们对社会资本回报的考察集中在经济回报上。

这些诸如信任、规范与网络聚集的社会资源，这种资源对提高行动结果的价值产生着重要的影响，许多学者论述了这方面的作用，如促进信息交流、强化共识行动、降低交易费用和协调利益关系等以提高回报率（布尔迪厄，1992；科尔曼，1990；彼得·埃文斯，1999；简·弗泰恩和罗伯特·阿特金森，1999；福山，1998）。一些学者还把它直接看成是全要素生产率的一个要素（金立群，2002）。其次，蕴含着社会资源的社会结构可以从狭义与广义的两个视角加以考察。狭义的社会结构指的就是作为个人的联系。伯特（Burt Bonald）指出，社会资本就是一个人与朋友、同事的关系或者更一般的联系，通过这些关系，人们可以使已经获得的金融资本和人力资本的效用最大化。奥斯特洛姆（Ostrom Elinor）把这个狭义的社会资本看成是微观层次。广义的社会结构是指建立在这种个人联系网络基础上的一种组织特征。普特南（1993）对意大利南方与北方长达20年的实证研究发现，意大利北方在经济与地方政府绩效水平总体上大大高于其南方，其根本原因在于两个地区之间的公民参与以及人们之间相互信任水平的差异。这项研究被看成是对社会资本进行宏观分析的一个经典案例。总之，狭义与广义的社会资本指的就是个人关系与集体组织中所蕴含的资源。

我们认为，社会资本对我国民营企业发展差异的影响也需要从狭义与广义的两个层面加以考察，任何单一层次的分析都是不完整的。为此，根据民营企业发展对社会资本的需求，本文将社会资本分为个人关系资本与集群网络资本。个人关系资本指的是民营企业主从其亲戚、朋友关系中获得机会与资源的能力。国内学者关于社会资本与民营企业成长的研究基本上是从这个视角展开的（李路路，1997，1998；张厚义等，1995）。这些研究得出了一个共性的结论，即我国经济转轨时期表现为正式制度规则处于变革过程，非正式的制度规则发挥着更加重要的作用。在这种体制背景下，民营企业主的亲戚、朋友关系的社会地位背景越显赫，动员稀缺资源的能力就越强。但是，这种以个人关系网络为基础的社会资本属于个人，因而具有一定的封闭性和排他性。换句话说，在这个关系网络内的人不会把这种社会资本无偿地提供给网络以外的其他人享用，不具有这种人际关系网络的人要获得这种社会资本，需要投入比在这个关系网络内的成员高得多的交易费用。集群网络资本是建立在企业参与专业化分工网络基础上的相互认同、互惠与信任。早在19世纪后期，马歇尔使用的“工业区”概念被看成是这种具有高效率、高灵活性的专业分工与合作的企业集群网络。波特（M. Porter，1990）认为，在现代经济中，产业集群“作为在某一领域内的相关工业及联合机构中，那些既相互竞争又彼此合作的相互依赖的公司、专业供应商、服务商、生产商等，集中在某一地区的一种工业现象”，在组织经济活动中具有重要的作用。这主要体现在企业之间基于专业化分工联系的稳定性以及由此产生的合作关系，集群内非正式的规则认同以及行会、互助会等组织的协调所产生的创新保护、有效竞争与资源共享等。这些作用构成了企业集群网络资本。与个人关系资本相比，集群网络资本具有三个特征：一是基于一个地区的地理条件与商业传统而形成了以某一产业内的市场分工深化，使产业内零售商、供应商与采购商以及服务机构之间的交易在这个地区内变得更加集中，因而具有地域专业化。二是地区内不断聚集的资源、信息和企业配套能力使得企业相互沟通、要素供给与技术扩散更加方便，由此生成的网络资源具有共享性。三是企业网络通过与外部市场、信息、资源和企业之间的扩大交换而保持着竞争活力，因而具有开放

性与群体性。换句话说，凡是希望获取这种集群网络资本的相关企业和配套厂家都可以在此投资设厂。在假设集群内开办新企业的费用与其他地区基本相同的条件下，由于集群内专业化需求信息集中、厂商分工细化、配套服务完整等优势，所以，这往往导致流入集群的资源大于流出，从而推动整个集群经济的快速增长。以发展 IT 行业为主的美国硅谷和以传统产品为主的意大利北部的专业产品区成为了这方面的成功范例。

从个人关系资本与集群网络资本构成的社会资本结构角度考察其对我国民营企业成长的影响有助于我们扩展对我国民营企业成功因素与环境类型的认识。对于民营企业获取资源与机会的能力来说，个人关系资本可分为强供给与弱供给两种。而集群网络资本可以通过民营企业在集群内外显示出来，换句话说，集群内的企业可享有这种公共性质的集群网络资本，而集群外的企业则无法享受。这样，影响民营企业发展的社会资本结构可扩展为四种环境：一是集群外与缺少个人关系资本组合的环境；二是集群外与拥有较强个人关系资本组合的环境；三是集群内与缺少个人关系资本组合的环境；四是集群内与拥有较强个人关系资本组合的环境。在这四种环境中，如果某种环境下的企业更有效率，我们就把它看成是比其他环境提供了更多的生存机会。对于民营企业来说，究竟哪一种环境具有更多的生存机会？

三、两种社会资本组合的实证比较

在上述四种环境中，区分可比较的环境是必需的。这种区分包括两种情况：一是在假设集群网络资本不变的条件下，观察个人关系资本的强弱与企业绩效差异之间的联系，由此区分第一与第二种环境的影响力。二是在假设个人关系资本不变的条件下，比较集群内外的企业绩效差异，从而判断第二和第三种环境，即有没有集群网络资本对民营企业影响程度。从前一个区分来看，大量的文献与数据已表明了我国转轨时期的个人关系资本强弱与民营企业经营绩效差异之间存在着相关联系的看法。如李路路（1997）通过对 1995 年的第二次全国私营企业问卷调查数据的对数线性回归分析，得出了民营企业主办企业前的家庭背景、个人地位以及政治面貌对企业成功并不具有重要性，而与其来往关系最密切的亲戚、朋友及其配偶的社会地位对企业成功具有明显的重要性。石秀印（1998）也以 1995 年全国工商联对私营企业家的调查数据为分析依据，通过概率统计，验证了存活时间长且经营成功的私营企业家往往是与拥有资源与权力支配者具有更密切关系的人的看法。边燕杰、丘海雄（2000）通过对广州 1998 年的 188 家企业问卷调查数据分析，证明了社会资本对企业经营能力与绩效具有重要作用的理论假设。这些研究对这种相关联系的解释是，在我国经济转轨过程中，由于正式制度规则的变动具有更大的不确定性，因而依赖非正式的社会关系网络便有了更大的作用空间，这种“关系”资源的多少决定着民营企业获取市场机会与资源的能力大小，进而导致了企业绩效差别。

后一个区别也为许多学者的实证研究所证明。意大利经济学家 S. Fabiani, G. Pellegrini, E. Romagnano, L. F. Signorini 等人使用了 10939 家意大利企业在 1995 年的财务报表数据，并将企业分为非集群内企业与集群内企业两类。通过统计分析，发现对于任何规模和行业而言，集群内的企业都比非集群内的企业具有更高的经营能力和更好的经营

绩效。比如在 13 个行业中，非集群内的企业平均盈利率为 11.55%，而集群内的企业平均盈利率为 13.54%。在企业集群内，从 10~19 人，20~99 人到 100~250 人等三类企业规模的平均盈利率都分别比非企业集群的企业平均盈利率高出 2 个百分点左右。同时，大部分行业内的集群企业比非集群企业的融资也更为平衡，集群企业的债务成本平均为 7.84%，而非集群企业为 8.03%。集群内企业在总支付利息/总生产利润的指标为 29.52%，非集群内企业的这个指标为 31.78%。这种融资方面的良好表现使其成为集群内企业获得高利润率的因素之一^①。意大利经济学家 Michele Bagella, Leonardo Becchetti, Simona Sacchi 等人的研究发现，意大利在非集群内的企业出口为 21.08%，而集群内企业出口比例为 25.61%^②。Evert-Jan Visser 对秘鲁生产服装的中小企业进行了分类研究，结果显示，集群内的工人每月平均销售额比非集群内企业的平均销售额高出了 30% 左右，工人工资也高出 30% 以上^③。国内学者（金祥荣，2002；王珺，2002）对浙江专业产品区及广东专业镇的研究也印证了在缺乏集群网络与具有集群网络的两个时期中民营经济发展的明显差异。由于这些研究揭示了集群网络所具有的信息交流方便、获取专业技术人员更加容易、技术快速传播与扩散、由交易成本降低而增进的合作等网络优势，因而本文将其概括为集群网络资本。

把上述两方面的比较综合起来，可以看出，与缺乏个人关系资本或集群网络资本的民营企业绩效相比，存在着较强的个人关系资本或集群网络资本环境下的民营企业都有更好的经营业绩。由此可以得出这样一个结论：在既没有集群网络资本又缺乏较强个人关系资本的第一种环境下的企业绩效一般是最差的。

四、两种社会资本优势叠加的不兼容分析

然而，这是否可以从逻辑上推出另外一个判断，那就是：在既有较强的个人关系资本又处于集群网络内的企业绩效是否是最优的？观察我国民营经济发展的现实，我们难以发现大量的民营企业在走向成功中几乎同时具备这两种社会资本，换句话说，个人关系资本与集群网络资本之间是缺乏“兼容性”的。之所以如此，我们认为，这与我国经济转轨时期社会资本结构中两种资本的配置取向与资源动员的顺序机制有关。

从配置取向上看，拥有较强个人关系资本的企业往往会利用这种优势进入资金门坎较高、市场准入限制也较多而盈利空间较大的行业。这是因为，一方面，进入的资金门坎较低、限制较少的竞争性行业不需要个人关系资本的大量投入；另一方面，个人关系资本需要通过回报加以维护和增值，如果拥有较强的个人关系资本进入到盈利机会十分有限的竞

^① S. Fabiani, G. Pellegrini, E. Romagnano, L. F. Signorini, Efficiency and Localisation: The Case of Italian Districts, The Competitive Advantage of Industrial Districts, Germany: Physica-Verlag Herdelberg, 2000, p52.

^② Michele Bagella, Leonardo Becchetti, Simona Sacchi, The Positive Link Between Geographical Agglomeration and Export Intensity: The Engine of Italian Endogenous Growth, The Competitive Advantage of Industrial Districts, Germany: Physica-Verlag Herdelberg, 2000, p107.

^③ Evert-Jan Visser, A Comparison of Clustered and Dispersed Firms in the Small-Scale Clothing Industry of Lima, World Development, 1999, Vol. 27, No. 9, pp1553 ~ 1579.

争性行业，那么使用个人关系资本的主体就无法获得相应的回报，进而使得个人关系资本无法积累与增长。我国经济转轨时期的房地产业就属于这种行业。1993年我国私营经济研究会当时中国私营经济的问卷调查结果表明，房地产建筑业通过“利用各种手段打通关节”方式来购买原料和进货的比例占其购货总量的25.7%，这在所有行业中是最高的^①。2002年《福布斯》公布了我国民营企业的百富名单。在这百富中，有15位曾是政府官员，26位曾是国有企业厂长、经理，12位曾具有海外经商的家族背景与社会关系。这三部分人就是我们通常所说的在经济转轨时期具有较强个人关系资本的人，这三部分人就占了民营企业百富中的53%。从他们从事的产业来看，47%是房地产开发商，其中前10位中就有8人来自房地产。来自建材的占6%，基础设施的占9%，来自制造业的占了25%，传媒、文教占了7%，农业及其他占了6%等。2003年在《新财富》杂志4月号公布的内地400富豪中，进入房地产行业的有76人，占19%。而《福布斯》全球首富的476人中，只有29人属于房地产业，占6.1%。之所以有较强个人关系资本的人多数进入了房地产业，这是因为自20世纪80年代，特别是90年代以来土地使用权的供给是由政府以“协议出让”的不完全市场化方式配置的，这种配置大约占了全部经营性土地使用权出让的95%以上。在这种体制下，拥有较强的个人关系资本的人以“协议出让”的较低价格获取了土地使用权，并利用一级土地市场与开发后房地产市场之间的差价可以获得了巨大收益。可见，没有较强的个人关系资本的中小企业是难以进入的。直到2002年7月1日国家开始实施《招标拍卖挂牌出让国有土地使用权规定》后，市场机制引入了土地一级市场，才停止了这种“协议出让”方式。

不具有较强个人关系资本的民营企业大部分只能进入以日常工业用品和小五金产品为主的竞争性行业。这些行业的资金、技术门槛较低，不存在市场准入的限制，因而也决定了进入这些行业的民营企业规模较小。这些企业不对个人关系资本产生过多依赖，但是，集群网络资本的存在却支撑了他们的快速发展。浙江与广东等省中小型民营企业的较快发展提供了这方面的例证。在浙江66个县（市、区）中，以中小企业聚集的专业化产业区超过1亿元的就有306个，平均每个产业区的规模达到了8.7亿元。在这些专业化产业区中，绝大部分是传统的竞争性行业，如纺织、服装、制鞋、打火机、皮革、眼镜、钮扣拉链、家具和小五金等。这些企业雇佣的人数大多数为30~50人，这些企业家的80%以上来自推销员、中小学教师、工人和手工业者。在温州以低压电器和服装制鞋为主的产业区中，生产规模居于前16位和前10位的企业家几乎没有一位具有政府部门的干部与国有企业的厂长、经理等背景。即使入选2002年中国百富的正泰集团董事长南存辉和德力西集团公司董事会主席胡成中也分别是从小作坊和小裁缝做起的。1998年，广东省超过10亿元社会总产值的建制镇为274个，占全省1551个建制镇的17.67%。这些总产值在10亿元以上乡镇的一个共同特点是，产值的90%以上是靠民营企业创造的，而民营企业基本上是以专业镇为组织依托的。如中山古镇的灯饰之都、小榄的五金制品之乡、黄圃的腊

^① 私营企业以“利用各种手段打通关节”方式购买原料和进货的比例，在工业部门为17.3%，交通运输业为24%，商业为19.3%，饮食业为8%，服务业18.8%，修理业14.6%，科技咨询18%，其他17.2%。参见《中国（民）私营经济年鉴》，香港经济导报1994年版。

味城、沙溪的服装基地、澄海市澄城的玩具城、南海西樵的纺织之都、大沥的铝型材制品和摩托车之城、盐步的全国内衣会展中心、平洲和里水的制鞋之都、南庄和佛山石湾的陶瓷之乡、张槎的针织城、顺德伦教的木工机械制造中心、乐从的中国家具城、东莞虎门的服装城、石龙、石碣和清溪等镇电子产品、厚街的鞋业、茂名市高州和化州的水果、梅州普宁的凉果等等。到2002年，经过了广东省科技厅评认定的58个专业镇的GDP总和已经达到了1000多亿元，占广东省国内生产总值的近十分之一。

再从资源动员的顺序机制上看，大、中、小型民营企业发展对市场、银行与政府的需要顺序是不同的。在2000年北京举办的一次民营企业座谈会上，小企业普遍认为，企业经营发展，首先要琢磨市场，有一定基础后会找银行，到做大了之后再找政府支持。中型企业代表的看法是，在做项目或企业进一步发展时，首先会想到找银行，以寻求资金支持，其次才会开拓市场。对于一些大型企业来说，往往是一立项或刚上马，首先会寻求得政府的支持，在政府的支持下到银行贷款，最后找市场^①。这些说法尽管未必完全正确，但却反映了不同的企业规模安排政府支持、银行贷款和市场拓展的差异性。

上述分析表明，在我国经济转轨时期，两种社会资本相结合作为民营企业发展最优的第四种环境的理论假设在现实中并没有大量地出现，这主要是两种社会资本在动员顺序与配置产业上的差异导致的“不兼容”造成的。

五、两种社会资本的结构转换

接下来是区分第二与第三种环境对民营企业发展的影响程度。这两种环境的影响程度是随着我国体制转轨的深化而呈现出不同的变化趋势。个人关系资本对民营企业成功的影响程度是一个由强变弱的过程。在任何社会中，个人结成的关系网络都存在，但是这种关系对资源配置的影响程度却取决于一个社会的正式与非正式制度安排的结构。如果一个社会缺乏正式的制度安排来配置资源，那么非正式制度安排就会填补这个空档。在由计划体制向市场体制的转型初期，适合计划体制的正式制度规则已经不适应市场体制的要求，但是适应市场经济的制度规范还没有完全建立起来，在这个新旧制度规则交替过程中，正式制度规则配置资源的支配、控制和约束的能力与范围会收缩，而以关系网络配置资源的非正式制度规则会相应地扩展。如注册审批、获取市场准入的许可、购买原料、搭建销售渠道甚至缴纳税收多少的谈判等等，企业经营活动的每个环节几乎都体现着个人关系资本的作用。在这种条件下，拥有较强个人关系资本的民营企业比缺乏这种资本的人具有更高的成功率。比如，当大多数民营企业不允许进入一些盈利较大的，诸如金融、保险、批发、电讯等具有一定行政垄断的行业时，少数企业可以凭借个人关系通过各种打“擦边球”方式突破这种体制约束，从而获得先进入的收益。随着经济体制转轨的深化，一方面，与市场经济制度相适应的正式制度规则相继出台与实施，使得依赖个人关系资本获取资源与市场机会的盈利空间会逐步减少，进而利用这种关系资本进行交易的数量与方式也降至最低限度。另一方面，限制民营企业进入的体制性与政策性“篱笆”逐步被取消，使得民营企

^① 杨惠妹：《民企如何走近政府》，《北京经济报》，2000年12月11日。

业与其他所有制类型企业享有的市场准入、融资条件等各种待遇逐步趋同，这样，民营企业通过个人关系来解决经济活动中的资源动员、纠纷处理等问题的空间就会逐步减少，而通过公开的制度程序解决这些问题的机会则相应地增加。因为在市场经济制度不断完善条件下，动用与不动用个人关系资本所获得的回报率大致趋同，因此，与其使用个人关系资本后，需要把获得的一部分收益分配给支付了这种个人关系资本的其他人，倒不如不使用这种个人关系资本。其结果，个人关系资本的作用会相应地减弱。这与一些学者（边燕杰、丘海雄，2000）提出的“企业社会资本的作用会随着市场制度的不断完善而相应减弱”的观点是相一致的。只是说，这种不断减弱的社会资本实际上指的就是这种个人关系资本，而不是整个社会资本。

集群网络资本对民营企业发展的影响则随着我国市场经济发展而呈现出一种不断增强的趋势，推动力量主要来自以下几个方面：第一，各地区积极发展民营经济的动机将会借鉴广东和浙江等省依托专业产品区或专业镇推进民营企业发展的成功经验，这将成为重要的扩展力量。目前，我国大多数省市都把民营企业发展作为新一轮经济的增长点。而大量的中小民营企业的竞争劣势不在于规模，而在于孤单（Schmitz, 1997）。以专业产品区、专业镇为主的产业集群网络恰恰是形成民营企业的群体力量，共同抵御市场波动与风险，从而克服这种分散、孤单的有效组织方式。浙江、广东等省份在这方面的成功经验又降低了其他省份进行探索的成本，因此，搭建具有集群网络资本的专业化组织平台，促进民营企业的自发性发展与地方政府有组织的引导相结合，已成为许多省份推进民营企业发展的一项主要措施。江苏、福建、四川、贵州和海南等省份都显示了依托集群网络促进民营企业发展的势头。第二，随着市场专业化分工的深化，以企业簇群为组织方式的集群网络资本还会不断“升值”。经过了十几年的迅速发展，广东省大多数专业产品区、专业镇等产业网络已由过去的集群与外部有市场分工联系，而集群内企业之间缺少专业化分工联系的阶段，开始进入到集群内企业之间围绕着产业链条形成纵向分工联系的阶段，这种集群网络内的市场分工深化不仅吸引外部相关产业及资源流入，也会强化企业对集群网络的依赖性，从而降低企业外迁的可能性。广东南海市西樵镇在纺织行业内的分工演化以及中山市古镇 1500 多家企业围绕着灯饰制品进行产品链条上的分工情况提供了这方面的例证。第三，随着电信、汽车、重型机械装备等行业的行政垄断性门槛大幅度削减，具有一定技术与投资实力的民营企业开始相继进入这些产业领域。由于这类产业大部分是产业链条较长的行业，且存在着比较明显的资源聚集性配置特征，因此，专业产品区等组织形式开始由传统的竞争性行业向这些产业链条比较长的重型机械装备工业扩展，这无疑给公共网络资本“增值”提供了新的市场空间。第四，在专业产品区中，随着有组织的行会、商会等中介机构的兴起，建立在规范价值基础上的广泛信任网络将会逐步地创造出来，进而提升公共网络资本的价值。社会资本理论认为，在资源缺乏流动的封闭社会中，没有行业协会、商会也容易形成有效规范和可强制推行的信任。在资源流动性增强的开放环境下，行业协会、商会就成为建立有效规范，强制推行各种约束因素，如行规行约、行业自律等，从而形成可强行的信任的必要组织形式。在广东省基本成型的 58 个专业镇中，基本上都相继组织了行业协会和商会，这给外部资源的融入提供了信任的制度基础。

总之，在我国经济转轨的社会资本结构对民营企业发展的影响中，个人关系资本是一

个由强变弱的过程，而集群网络资本则是一个由弱变强的过程。在经济转轨初期，拥有较强个人关系资本对民营企业发展的影响力超过了集群网络资本，列在首位。随着体制转轨的深化，集群网络资本的重要性会逐步增加，以至将会超过个人关系资本而排列在首位。

六、结论

根据社会资本构成社会环境的理论，本文首先把社会资本分为两个部分，即个人关系资本与集群网络资本，这两者之间的组合形成了民营企业发展的四种环境。然后，观察和分析这四种环境下的民营企业发展绩效，我们得出这样几点结论：第一，在既缺少个人关系资本又不存在集群网络的环境下，民营企业绩效是最差的。第二，在既存在集群网络联系又有较强个人关系资本的环境下，民营企业绩效应该最优。但是，在我国经济转轨时期，由于两种资本的形成机制和产业进入的领域不同，因而两种资本叠加为最优的环境假设是不现实的。第三，拥有较强的个人关系资本而不存在集群网络资本的环境与缺少个人关系资本而共享集群网络资本的环境之间存在着一个替代与转换过程。在经济转轨初期，由于非正式制度规则支配资源的领域扩展和企业专业化集群发展不十分普遍，因而拥有较强个人关系资本的民营企业的成功概率比较高。而随着适应市场经济的制度规则逐步完善，企业专业化分工基础上形成的集群网络逐步增多，集群网络资本对民营企业走向成功变得越来越重要。第四，个人关系资本对民营企业成功的影响是个体性的，而集群网络资本对民营企业成功的影响是群体性的，因而，随着个人关系资本的作用弱化与集群网络资本的作用强化，在其他制度与政策不变的情况下，民营企业的存活率与持续发展的可能性都会相应地提高。

参考文献

- [1] 李惠斌，杨雪冬主编：《社会资本与社会发展》，社会科学文献出版社 2000 年版。
- [2] 张厚义，陈光金主编：《走向成熟的中国民营企业家》，经济管理出版社 2002 年版。
- [3] 胡润主编：《福布斯 2002 中国百富》，中国商业出版社 2002 年版。
- [4] 李路路：《私营企业主的个人背景与企业“成功”》，《中国社会科学》，1997 年第 2 期。
- [5] 李路路：《向市场过渡中的私营企业》，《社会学研究》，1998 年第 6 期。
- [7] 石秀印：《中国企业家成功的社会网络基础》，《管理世界》，1998 年第 6 期。
- [8] 边燕杰，丘海雄：《企业的社会资本及其功效》，《中国社会科学》，2000 年第 2 期。
- [9] 迈克尔·E. 波特：《簇群与新竞争经济学》，《新华文摘》，2000 年第 7 期。
- [10] 金祥荣，朱希伟：《专业化产业区的起源与演化》，《经济研究》，2002 年第 8 期。
- [11] 王珺：“企业簇群的创新过程研究”，《管理世界》，2002 年第 10 期。
- [12] 林南：《建构社会资本的网络理论》，《国外社会科学》，2002 年第 4 期。
- [13] Michele Bagella and Leonardo Becchetti, Geographical Agglomeration in R&D Games, The Competitive Advantage of Industrial Districts, Germany: Physica-Verlag Heidelberg, p35.
- [14] S. Fabiani, G. Pellegrini, E. Romagnano, L. F. Signorini, Efficiency and Localisation: The Case of Italian Districts, The Competitive Advantage of Industrial Districts, Germany: Physica-Verlag Heidelberg, 2000, p52.

-
- [15] Evert-Jan Visser , A Comparison of Clustered and Dispersed Firms in the Small-Scale Clothing Industry of Lima , World Development 1999 , Vol.27 , No.9 , pp1553 ~ 1579 .
- [16] Alejandro Portes , The Economic Sociology of Immigration : A Conceptual Overview , In Portes (ed) , The Economic Sociology for Immigration : Essays on Network , Ethnicity and Entrepreneurship , New York : Rusell Sage Foundation , 1995 , p12 .
- [17] Hubert Schmitz and Khalid Nadvi , Clustering and Industrialization : Introduction , World Development , 1998 , pp1503 ~ 1513 .

交易虽然互利，但是未必公平^{*}

王则柯

大家知道，交易是互利的行为。对你有好处，对他也有好处，交易就做成了。本来，交易利益是交易的前提，可是对于这一点，人们有时候并不那么清醒。

何以见得？这只要看老太太买东西常常唠叨又吃了亏。

买东西是自愿的行为。如果你觉得吃亏，不买就是了。一方面唠叨吃亏，另一方面却还是要买，那不是口是心非是什么？没有人强迫你买呀。真的觉得吃亏，不掏钱就是了，为什么自愿掏钱买亏吃？

不过，我们也不要过多责备老太太，更不要讥笑老太太。如果说老太太口是心非，那是不自觉的口是心非，因为她不明白。虽然她其实并不吃亏，但是她真的觉得自己吃亏。她的感觉是失真的，但是她的话语却是她的感觉的忠实反映。值得注意的是，饱受经济学教育的学子和学者，也会犯老太太这样的糊涂。例如，2001年10月在北京大学中国经济研究中心举办的“首届中国经济学年会”上，学者们就在一个分会场争辩交易何以让人吃亏。在我看来，虽然我们的经济学教育也孤立地讲了交易利益，但是在融会贯通方面做得很差，以至于常常自觉或者不自觉否定了这样一个重要的事实：交易利益是交易的前提。

因为交易是自愿的，我们可以建立下面这样的命题：

命题 在交易的当时，交易各方都不会真的吃亏。

证明 若不然，只要有一方真的吃亏，他就没有道理参与和实施这笔交易。

这是很简单的反证法的证明。

区分事实需求和心理期望

既然交易一定是互利的，为什么却有那么多人感觉因为交易而吃亏了呢？这可能是两个因素在作怪：一是人们混淆了事实需求和心理期望；二是人们混淆了当时需求和事后检讨。

关于事实需求和心理期望，要注意人们是否吃亏的感觉，有两个层次：一个是事实上是否吃亏；一个是心理上是否吃亏。

举例来说，你到一个生疏的地方旅游或者探险，渴得不得了。这时候你遇到一个卖瓶装水的小贩，一瓶350毫升的纯净水卖5元钱。这种瓶装水你很熟悉，往常在你知道的一些超市，撑死了也就卖1元，想不到现在居然卖5元。“这小贩真黑，怎么没有反暴利警察之类管管”，你心里这么想，虽然你素来对“反暴利”之类不以为然。可是眼下你嗓子冒烟，渴得不得了。救命要紧，5元就5元，咬咬牙，你还是买了。

^{*} 原载：《江苏社会科学》，2004年第2期（论文题目作了适当修改）。

整个过程，是一番紧张的权衡取舍：5元钱留在兜里好，还是拿出去换来一瓶纯净水好。反复权衡取舍的结果，最后你掏钱了。不管你心里怎么感觉，掏钱的行为说明你认为5元钱买一瓶水值，因为这是你权衡取舍的结果。可见，你事实上并不吃亏。

买了纯净水，你咕噜咕噜就灌了下去。久旱逢甘雨，你痛快极了，仿佛从来没有喝到过这么甘甜的纯净水似的。

那么，为什么你内心却还是忿忿不平念叨被宰了呢？这就是心理因素在作怪了。前面说了你感觉好像从来没有喝到过这么甘甜的纯净水。本来，这应该是你那5元“大价钱”的心理补偿，可惜你算不过这个账来。买东西谁不想便宜？最理想的情况，是有人知道您老要到这里来巡视，老远把瓶装水替你运到这个地方来，等候你口渴的时候光顾，而且价钱还像城里的超市一样，1元钱1瓶。可是你想想，谁会做这样辛苦赚吆喝的生意？——除非他对你别有所求。不过，那将是他宁愿现在吃小亏要在将来占你大便宜的营销策略了。那是钓鱼放出的鱼饵，不是独自完整的交易。

所以，经济学家更加注视人们的行为而不是人们的言辞。

把故事接着讲下去：掏出5元钱买了一瓶纯净水，咕噜咕噜你就灌了下去。你不是在旅游吗，你不是还要走下去吗？想不到往下没走几步，刚拐了个角，你就看到一个小卖部，同样的纯净水在那里只卖3元钱一瓶。这下你真的火了：“被那个家伙坑了，我真是亏得紧！”

可不是？再走两步就可以买到3元钱一瓶的纯净水，你却买了5元钱一瓶的，真的很亏。可是，这仍然没有违反我们的上述命题，因为命题强调“在交易的当时”，交易各方都不会真的吃亏。

讨论人们的经济行为，不可不注意其中的信息要素。在你嗓子冒烟，渴得不得了，并且不知道旁边不远就有价钱合理得多的店铺的时候，权衡取舍的结果，你自愿掏出5元钱买下小贩的一瓶纯净水。这里，你不知道旁边不远就有价钱合理得多的店铺这一信息因素，至关重要。当信息因素变化的时候，人们的以货币表达的需求也要变化（股票市场是突出的例子）。事实上，如果你知道旁边不远有便宜许多的店铺，哪怕你嗓子已经冒烟，渴得不得了，再熬半分钟，你是熬得起的，从而你不会买5元钱一瓶的纯净水。可惜你先前决策买下那瓶纯净水的时候，你并不掌握旁边不远有便宜许多的店铺的信息。

信息结构变了，一切都可能变化。谈生意，为什么要讲究签订合同？就是为了防止因为信息结构变化和其他变化而变卦。

保留价格与交易利益

实际经济生活中，岂止在交易的当时交易各方都不会真的吃亏，而且在交易的当时通常各方都得到交易利益（trade benefit）。这也是一个可以证明的命题，同样可以用反证法证明：如果没有交易利益，人们没有道理参与交易。

回到我们的故事来。当你和那个小贩交易的时候，你已经渴得不得了，如果不是遇上这个出价5元钱的小贩，而是遇上暂时没有竞争对手的一个开价6元钱的小贩，甚至遇上开价8元钱的小贩，可能也会买，因为你不是守财奴，不至于因为不愿意感觉被宰，就宁可渴死。好了，据此我们可以知道这时候你对于一瓶纯净水的保留价格（reservation price）

至少是8元钱。为简单起见，就算是8元钱吧。

反过来，那个利用信息优势欺负你的小贩，他把那么些瓶装水老远运来，也不容易。购水成本加上运费和辛苦，合起来每瓶水的经济成本姑且算是2元钱吧。这样我们就知道，小贩出售瓶装水的保留价格，是2元钱。只要售价在这个保留价格上面，小贩做的就是赢利的生意。

这里注意，经济成本（economic profit）不是会计成本（accounting profit），经济成本通常比会计成本高，要把自己劳动的辛苦费等等加上去。还要注意，买方的保留价格，是他为了买这个商品顶多愿意支付多少钱；而卖方的保留价格，是你至少给我这么多钱我才肯把商品卖给你。简单说来，交易价格在卖方的保留价格上面，卖方就赚钱；交易价格在买方的保留价格下面，买方就买得值。

现在，买方的保留价格是8元钱，卖方的保留价格是2元钱。那么，如果他们交易成功一瓶纯净水，他们就共同实现了 $8 - 2 = 6$ 元钱的交易利益。

至于这总数6元钱的交易利益在买卖双方之间如果分割，那就要看具体的交易价格了。现在是以5元钱的价格成交，那么小贩得到 $5 - 2 = 3$ 元钱的交易利益，你得到 $8 - 5 = 3$ 的交易利益。如果小贩再狠一点要价6元钱，你不知道旁边不远还有地方可以买到纯净水，只好接受这个价格成交，那么总数6元钱的交易利益将按照小贩得 $6 - 2 = 4$ 元钱、你得 $8 - 6 = 2$ 元钱来分割。

总起来说，交易利益等于买方的保留价格减去卖方的保留价格。至于交易利益如何在买卖双方之间分割，则取决于具体的成交价格。具体的成交价格由市场决定。这里注意，只由一个卖方和一个买方组成的市场，应该说是最小的市场。不过人们常常不把这种只有一个买方和一个卖方的情况看作是一个市场，而是说“没有市场”。如果我们接受这样的说法，那么在没有市场的情况下，具体的成交价格，由买卖双方的讨价还价决定。整个过程，是双方讨价还价能力的角力。信息因素，自然是讨价还价能力的组成部分。设想你知道旁边不远就有价钱公道得多的铺子，你的讨价还价力量将大大提升。但是还有别的一些重要因素会影响讨价还价能力。

经济学讲究买卖双方对交易标的物的评价（valuation），即交易标的物对于买卖双方各值多少。

但是，一样东西或者一件商品对于交易当事人值多少钱，是当事人的私有信息。我们说人们对于同一标的物的评价是私有信息（private information），包含两层意思：首先，对于同一标的物的评价，因人而异，非常个性化。你渴得嗓子冒烟，愿意出最高8元钱的价格买一瓶纯净水；另外一个游客有备而来，饮用水原来就带了比较多，不那么渴，如果有2元钱一瓶的纯净水，他乐意补充一些，超过这个价位，他就不接受。在这个例子里面，对于同样一瓶纯净水，你的评价是8元钱，他的评价是2元钱，很不相同。

我们说人们对于同一标的物的不同评价是私有信息的另一层意思，则是他们不肯轻易把自己的评价说出来。经常有这样的情况：小贩其实赚得很开心，嘴里却说“亏了亏了，亏本卖给你了”，就是这个道理。反过来，设想你在跳蚤市场看中了一件小玩意儿，如果你让卖方知道你对这件小玩意儿喜欢得不得了，即你对这件东西评价很高，那么你是讨不到好价钱的。事实上，出于追求更大交易利益的考虑，买卖双方都有隐蔽自己对于交易标

的物的真实评价的动机。

人们对于商品的评价，就像人们消费一件商品所获得的“效用”即满意程度一样，具有主观心理特征，难以给予客观的度量。但是我们在前面说过，在经济学家眼里，人们是以他们的行动而不是以他们的言辞说话，所以经济学把买卖双方对于作为交易标的物的商品的保留价格，看作是他们对于商品的评价。撇开保留价格作为私有信息的隐蔽性，保留价格至少在理论上具有可度量性。

这样，设 x 为交易标的物，记卖方对交易标的物的评价为 $v_s(x)$ ，买方对交易标的物的评价为 $v_b(x)$ ，这里下标 s 和 b 分别是英文卖者 (seller) 和买者 (buyer) 的头一个字母，那么，双方就标的物 x 达成交易的前提条件是：

$$v_b(x) > v_s(x)$$

或者：

$$v_b(x) = tv_s(x), t > 1$$

而买卖双方交易这件标的物所实现的交易利益是 $v_b(x) - v_s(x)$ 。

交易利益如何产生

现在，我们讨论交易利益如何在买卖双方之间分割的问题。

对于一种商品，“潜在的买者愿意按照他的保留价格 P^* 购买数量为 Q^* 的商品”这一事实，可以在“数量 - 价格”QP 平面上表示为一条具有一个下行阶梯的折线：折线沿 P 轴从上面走下来，在 $P = P^*$ 的位置向右水平行走 Q^* 的距离，然后垂直下行到达 Q 轴结束。这条下行一个阶梯的“曲线”，就是这位潜在的买者个人的需求曲线。同样，“潜在的卖者愿意按照他的保留价格 P^* 出售数量为 Q^* 的商品”这一事实，可以在“数量 - 价格”QP 平面上表示为一条具有一个上行阶梯的折线：折线沿 P 轴从下面走上去，在 $P = P^*$ 的位置向右水平行走 Q^* 的距离，然后一直垂直上行。这条上行一个阶梯的“曲线”，就是这位潜在的卖者个人的供应曲线。这是大家在微观经济学课程已经熟悉的事情。

在对于一种商品“存在市场”的情况，即对于一种商品具有众多买者和众多卖者的情况，这种商品的市场需求曲线，由众多潜在买者的个人需求曲线水平相加而得，而这种商品的市场供应曲线，则由众多潜在卖者的个人供应曲线水平相加而得。商品的市场需求曲线和市场供应曲线相交或者重合的位置，决定该商品的交易价格和成交数量。这也是我们在微观经济学熟悉的事情。上行下行两条阶梯曲线未必相交于一个点，可能重合于一个线段。在这两条阶梯曲线重合于一个线段的时候，如果这个线段垂直，两条曲线的重合部位决定的不是一个交易价格，而是一个价格范围；如果这个线段水平，那么两条曲线的重合部位决定的不是一个成交数量，而是一个成交数量范围。（见图 1）

现在我们要说的是，在“存在市场”的具有众多买者和众多卖者的条件下，众多卖者所分享的交易利益的总和，就是微观经济学所说的生产者剩余 (producers' surplus)；众多买者所分享的交易利益的总和，就是微观经济学所说的消费者剩余 (consumers' surplus)。

在微观经济学里面我们已经熟悉，在竞争的情况下，对于一种商品的单一市场，供给曲线 S 和需求曲线 D 的交点决定商品的交易价格 P^* 和成交数量。有了这个市场价格，买

卖双方不必每次再讨价还价，实际上由于竞争，进一步的讨价还价应该已经没有意义。

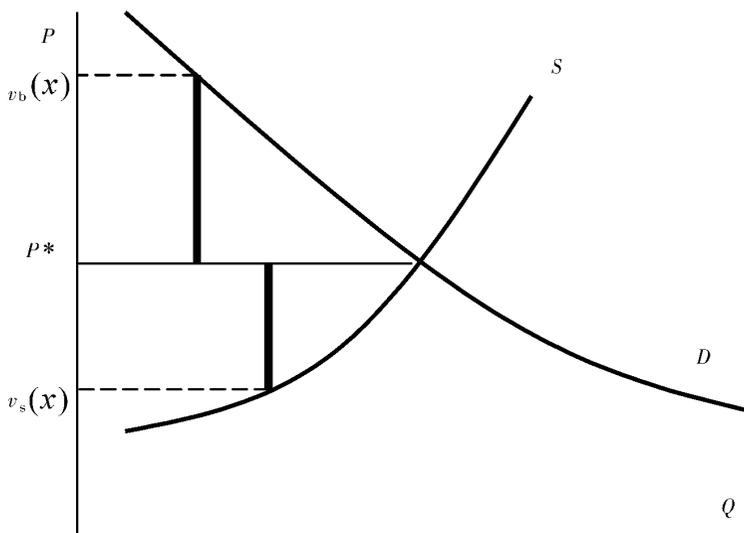


图1 交易利益与市场得益

虽然上述同质商品的竞争的市场价格 P^* 是划一的，但是参与这个市场的众多买者和众多卖者的保留价格并不一样。实际上，正是由于众多市场交易参与人的保留价格不一样，最后才合成一条下降的需求曲线和一条上升的供给曲线。这样，由于各人的保留价格不同，他们分享的交易利益也不相同。对于保留价格远远高于市场价格的买者，他分享的交易利益就很多，因为他原来愿意出很高的价钱买这种商品，但是现在只须付低得多的市场价格；相反，对于保留价格刚刚高于市场价格的买者，他分享的交易利益就很少，因为他原来就只愿意出差不多的价钱买这种商品，现在没有得到多少好处。同样，对于保留价格远远低于市场价格的卖者，他分享的交易利益就很多，因为他原来很低的价钱也愿意接受，可幸市场价格那么高；相反，对于保留价格略低于市场价格的卖者，他分享的交易利益就很少。

具体来说，对于保留价格是 $v_b(x)$ 的买者，因为最坏的情况他宁愿出 $v_b(x)$ 这个价钱，但是现在托市场的福他只需付 P^* 这个价钱，从而他享受的交易利益是 $v_b(x) - P^*$ ；对于保留价格是 $v_s(x)$ 的卖者，因为最坏的情况他宁愿接受 $v_s(x)$ 这个价钱，但是现在托市场的福他实现了 P^* 这个价钱，从而他享受的交易利益是 $P^* - v_s(x)$ 。上面的图演示，许许多多这样的卖者享受到的交易利益和买者享受到的交易利益，合成纵轴和供求曲线围成的表征总的交易利益的曲边三角形。这个三角形被表征市场价格的水平线分割成上下两个曲边三角形，上面的三角形给出总的消费者剩余，下面的三角形给出总的生产者剩余。

基于以上讨论，我们觉得“消费者剩余”和“生产者剩余”并不是好的翻译。我们建议把 consumers' surplus 和 producers' surplus 意译为消费者市场得益和生产者市场得益。有了市场，买方和卖方不但不必花费那么大力气搜寻对方，而且节省许多讨价还价的功夫。

买方和卖方都得益于市场。这样翻译，还因为英文 surplus 本身在经济学的另一个场合表示过剩，与表示短缺的 shortage 相对。在一词多义的情况下，采用不同的翻译，有时候能够起到比较准确地或者比较精细地把握概念的作用。同一个英文单词 auction，在一些场合翻译为拍卖，在另一些场合翻译为招标，就是成功的例子。

前面，在演示交易利益与市场得益的供求曲线分析图上，条形 $v_b(x) - P^*$ 表示一个买者因为在市场上购买一件商品所实现的市场得益，条形 $P^* - v_s(x)$ 表示一个卖者因为在市场上出售一件商品所实现的市场得益。我们特意把这两个条形画得错开。请读者仔细琢磨错开的含义。

如果“没有市场”，买卖双方能够相互发现配起对来，并不容易。开放集市贸易就是提供相互发现机会的重要制度设置。并不容易的相互发现，只是甲知道了乙想卖，乙知道了甲想买，接下去的讨价还价，通常更加艰苦。虽然买卖双方都希望交易成功，这样才能实现交易利益，但是双方都想得个好价钱，希望自己分享的交易利益大一些。事实上，固然双方都希望达成交易以便实现交易利益，但是对于交易价格之“好”、“坏”，双方却处于利益对立的位置：买方希望成交价越低越好，卖方希望成交价越高越好。

在一对一交易的情况，交易利益的分割，取决于双方的讨价还价能力。这一点我们以后会谈。讨价还价能力的构成十分复杂，例如，有时候忍耐力就是讨价还价能力的重要组成。

想象两家农户组成一个偏僻的小村落。现在两家的经济都不同程度地发展了，急需改善与外界的交通。这需要两家合伙投资。可是投资如何分摊，就是很大的问题。如果两家的关系比较疏远并且都想自己少出点钱，那么忍耐力比较强的一家通常占便宜。再想象大热天一个卖鱼虾的小贩和一个顾客讨价还价，眼看再也没有其他顾客，小贩的忍耐力就非常薄弱。他只能接受比较低的价格，不然的话，鱼虾热死，他的损失将更加惨重。这里要十分注意，即使小贩的鱼虾卖得比进货价格还低，在交易的当时，他仍然多少是获得一些交易利益的，因为这时候鱼虾已经不那么活泼了，鱼虾已经贬值。如果他嫌成交价格低不肯出售，到头来可能血本无归。

交易利益如何分割

至此我们应该已经清楚，人们不会在交易的当时真的吃亏，因为交易利益的存在是交易实现得前提。但是为什么人们还是常常对于不“如意”的交易耿耿于怀，认为自己吃亏呢？这主要是因为利益预期太高，但是实际实现得没有这么高，包括自己的东西已经贬值，却还想卖个好价钱的情况。这一节我们专门讨论这个问题。

我们还是集注于一对一交易的情形，因为一对一的交易比市场交易更容易使人感到吃亏。但是，我们不再限于交易一种商品，而是两个经济人一对一同时交易多种商品。从微观经济学的学习我们已经知道，多商品情形的讨论，常常可以简化为二商品情形的讨论。二商品情形讨论的结果，具有很好的一般性。

考虑经济人 A 和经济人 B 交换 x 和 y 两种商品，如下面的艾奇沃斯盒 (Edgeworth box) 所示，在初始时刻，A 对商品 x 和 y 的持有量是 X_A 和 Y_A ，B 对商品 x 和 y 的持有量是 X_B 和 Y_B 。这样，商品 x 和 y 的总量是 $X = X_A + X_B$ 和 $Y = Y_A + Y_B$ 。

艾奇沃斯盒分析，是读者在微观经济学已经熟悉的分析方法。在艾奇沃斯盒中，通过

表征两位主体人 A 和 B 对商品 x 和 y 的存量的初始持有点 E ，主体人 A 的无差异曲线和主体人 B 的无差异曲线围成一个纺锤形的交易互利区域（region of mutual advantages）。对于两位主体人来说，互利区域里面的任何一点，都表示效用（即各自满意程度）比初始持有点 E 高的一种资源配置，并且这种配置可以通过交换而实现。在这个艾奇沃斯盒中，我们知道主体人 A 的效用沿右上方向上升，主体人 B 的效用沿左下方向上升。由此可知，如果通过交易从 E 走到互利区域里面的任何一点，双方的效用都将得到提高。这也按艾奇沃斯盒分析的方式再次验证了我们前面讲的交易互利的命题。

问题在于同样是互利的交易，双方效用提高的情况却并不相同。按照两位主体人效用上升方向的不同，我们可以看到，无论是从 E 到 C 的交易，还是从 E 到 D 的交易，双方的效用都在提高，但是提高的情况不一样。在从 E 到 C 的交易和从 E 到 D 的交易之间，主体人 A 偏向于从 E 到 C 的交易，对于他来说， C 比 D 好得多；相反，主体人 B 偏向于从 E 到 D 的交易，对于他来说， D 比 C 好得多（见图 2）。这就是在交易互利的大格局下双方仍然存在利益冲突的道理。为什么建立世界贸易组织？为什么建立了世界贸易组织以后在组织里面还吵个不停？道理全在这里。

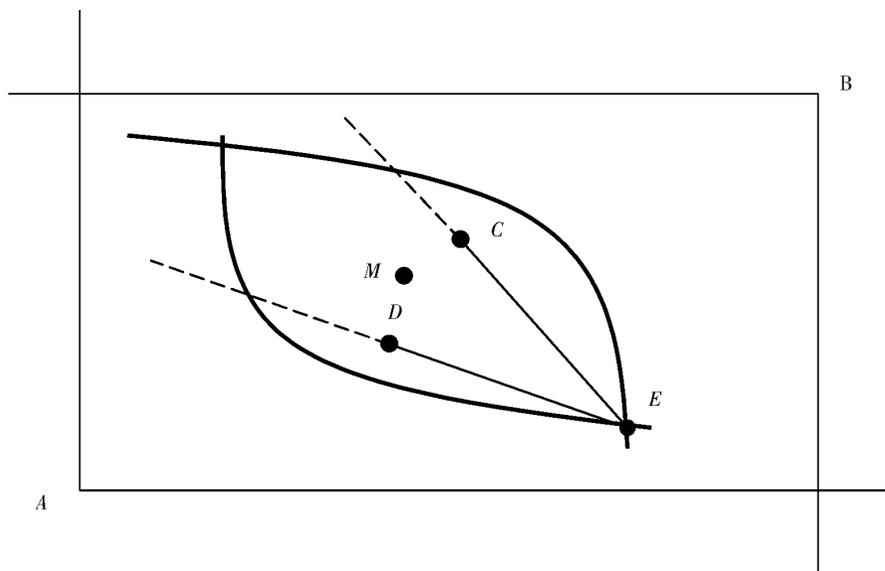


图 2 互利大局下的利益冲突

上述艾奇沃斯分析，深刻说明了人们的交易吃亏心理的根源：不是真的因为交易吃亏，而是比起期望来说得益不足，嫌好处还不够多。主体人 A 期望通过交易到达 C 这样的位置，实际上只到达 M 这样的位置；主体人 B 期望通过交易实现 D 这样的资源配置，实际上实现的却是 M 这样的资源配置。所以，虽然他们都在交易中得到好处，但是他们可能还是非常不满意。

可见，交易吃亏心理来自不能够分享更多交易利益的感觉。因为觉得分享交易利益少了，就觉得吃亏，就觉得不公平。

经济学非常讲究公平，但公平本身却是经济学上最难把握的概念。幸好，在上述“艾

奇沃斯交换”的情形，经济学家有比较一致的看法，认为公平指的是这样一种“竞争均衡”：想象一位“拍卖师”或者“公证人”尝试性地向两位主体人提出两种商品 x 和 y 的一个比价，按照这个比价，两位主体人各自盘算拿多少商品 x 换对方的商品 y 或者拿多少商品 y 换对方的商品 x ，一直尝试到 A 愿意交换出去的商品 x 的数量和 B 愿意交换进来的商品 x 的数量相等，或者 A 愿意交换进来的商品 x 的数量和 B 愿意交换出去的商品 x 的数量相等，A、B 双方就按照这个“竞争均衡”价格交换，实现这个“纯交换经济”的竞争均衡。

这个竞争均衡是可以在艾奇沃斯盒内实现的：大家知道，通过初始持有点 E 的每一条直线，表示商品 x 和 y 的一个比价。准确地说来，通过初始持有点 E 的每一条直线的斜率，表示商品 x 和 y 的一个比价。按照任何比价进行的交换，是“资源配置”沿着相应的直线运动的过程，双方都力图沿着表征比价的直线运动到自己效用尽可能高的位置。如果“拍卖人”或者“公证人”给出的是上述“竞争均衡”的比价，那么双方按照各自的盘算沿着这条直线运动，最后到达的对于各自来说效用最高的位置，将正好重合。

明白了这一点，就知道上述竞争均衡最终可以由两位主体人的提供曲线（offer curve）的交点确定，交点就是竞争均衡点。主体人的提供曲线，是通过初始持有点 E 的每一条直线和该主体人的无差异曲线的切点的轨迹（作为微观经济学的一个练习，请描出图“互利大局下的利益冲突”，并且尝试按照这里提供曲线的定义，把两位主体人的提供曲线画出来）。在下面的艾奇沃斯盒中，主体人 A 的提供曲线是箭头指向右上方的曲线，主体人 B 的提供曲线是箭头指向左下方的曲线。两位主体人的提供曲线的交点 M ，就是该纯交换经济的竞争均衡。见图 3。

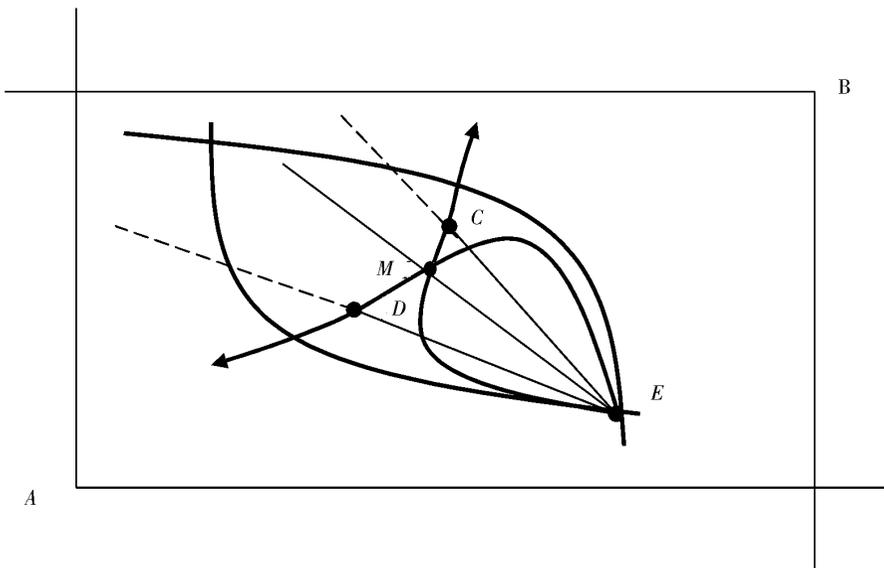


图3 艾奇沃斯交换的竞争均衡

前面说了，竞争均衡的情形，是想象为两位主体人听从“公证人”的价格尝试各自做效用最大化最后达到均衡的情形。如果不是某位“公证人”在主持“试价”，而是强势的一方

决定价格，结果就不一样。例如，假设主体人 A 处于垄断的强势地位，理论上他可以在 B 的提供曲线上寻找自己效用最高的点，如图中 B 的提供曲线和 A 的无差异曲线的切点 C（见图 4），按照从初始有点 E 出发通过 C 的直线提出交换价格，让 B 处于要么接受要么拒绝的位置。这时候，因为只有要么接受要么拒绝这样两种选择，主体人 B 出于自身的利益，将接受主体人 A 提出的垄断价格。结果，资源配置从原来的 E 走到 A 垄断的均衡 C。

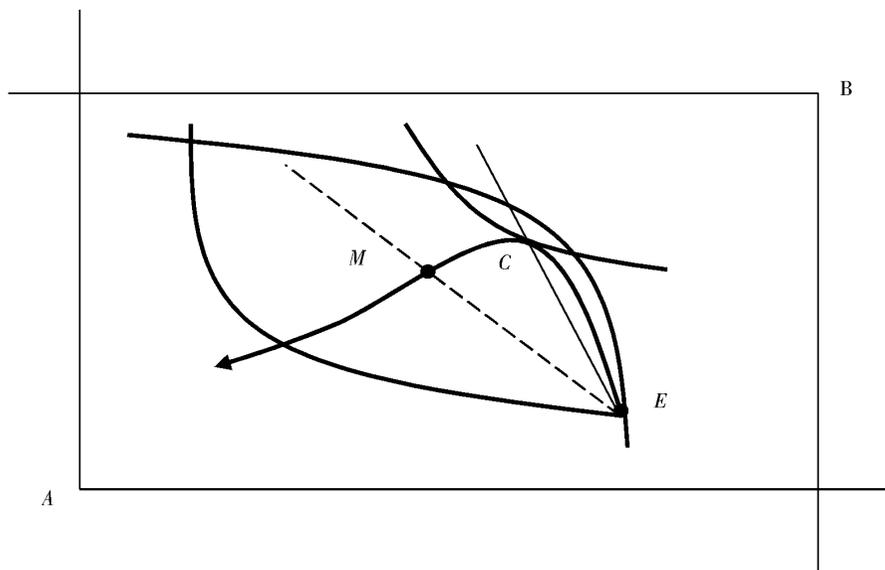


图 4 艾奇沃斯交换的垄断均衡

艾奇沃斯盒的垄断均衡，清楚地演示了交易仍然互利但是很不公平的情形。交易互利，说的是通过交易主体人 A 和主体人 B 的效用都得到提高，这从两位主体人的无差异曲线的情况可以看得很清楚。不公平，说的是因为主体人 A 处于交易的强势位置，所以他分享了绝大部分交易利益。事实上，在图 4 所示的情况，主体人 A 处于垄断的交易位置。

当前世界的国际贸易中，发达国家常常处于强势位置。设想一个只有初级产品可以出口的国家 B 与一个拥有某种垄断技术的国家 A 进行贸易谈判，谈判内容主要就是国家 B 购买国家 A 的这种垄断技术产品。进一步可以假设国家 B 的初级产品并不享有垄断的位置，即其他国家也向国际市场供应这种初级产品。这时候，虽然还是一对一的谈判，但是双方的地位不对等：国家 B 有求于国家 A，但是反过来国家 A 并不一定要从国家 B 进口那种初级产品。据此，我们可以明白欠发达国家在国际贸易中的苦衷：虽然他们在与发达国家的贸易中常常承受不公平的待遇，但是对于他们来说，贸易还是比不贸易好，因为如果闭关自守的话，他们连图 3 中 C 点表示的那一点可怜的交易利益也实现不了。

总的来说，利益是交易的前提，在理性人的条件之下，自愿的交易总是互利的。只是需要注意，虽然交易总是互利的，但是互利的交易未必公平。

国家所有权客体分类与主体界定^{*}

——一种经济分析的思路

李胜兰

国家所有权主体所代表的公共意志、客体所承载的公共利益，使得国家所有权成为我国所有权结构中不可或缺的部分。然而，现实中国家所有权因客体分类不清、主体界定不明晰而导致的国家所有权主体的“缺位”以及国有资产经营效率低下，都制约着国家所有权的行使和功能的发挥。因此，本文运用经济分析的方法，对我国国家所有权客体分类和主体界定提供一种新的思路和方法，也是立足于对既存的国有产权制度改革与创新实践问题的探讨。

一、国家所有权客体分类

依照传统的物权法学理论，国家所有权的客体从范围上观察，和一般所有权的客体是基本一致的。只要能够作为一般所有权的客体，也就能够成为国家所有权的客体。但就我国立法来看，某些财产可以作为国家所有权的客体，却不能作为个人所有权的客体，如土地资源等。应该说，这并不代表国家所有权和一般所有权的客体在本质上有所差异，只是国家出于一定的公共利益需要，对个人所有权的客体范围作出的一种立法限制。

虽然对国家所有权的客体的传统物权法学分类标准有多种，如有体物与无体物、动产与不动产等，但是我们在此从另一个角度，即结合国家财产的经济属性以及对其实际管理的情况，将国家所有权的客体分为四类：资源性财产、经营性财产、行政性财产和公益性财产。其中，自然资源资产并不反映在国有资产的账面上，而后三类分别以经营性资产与非经营性资产的名义列入国有资产的总量。根据财政部会计决算统计，1999年底我国境内外各类企业和行政事业单位占用的国有资产总额为90964.2亿元，比1998年增加8753.1亿元，增长10.6%。我国国有资产总量继续实现稳步增长。在全部国有资产总量中，经营性国有资产总额66748.4亿元，占73.4%，比1998年增长7%；非经营性国有资产总额24215.8亿元，占26.6%，比1998年增长22.3%。

新制度经济学理论认为，同一种财产由不同类型的权利人所有，会产生不同的利用效率；而不同的财产客体在同一种主体手里，也会导致不同的效用实现。例如，在经营性资产方面，国家所有权这种产权类型的存在不利于高效率的实现；在公益性财产方面，国家所有权却又发挥着不可替代的作用。相比经济学的效率研究，法学理论则在注重公平之时，只强调了维护各种所有权在权利内容上的平等性，而忽视了所有权的不同主体与客体

^{*} 原载：《中山大学学报》（社会科学版），2003年第1期。《新华文摘》2003年第5期推荐在热点论文。

之间的差异性。正是这种差异性，导致了对客体的不同利用效率，从而也可能影响到权利的实质平等。因此我们拟从效率实现的角度对国家所有权的客体作一具体探讨。

1. 资源性财产：国家所有权的分级收益

资源性财产是指自然界存在的天然物质财富。作为国家所有权的客体，是指国家所有的可供开发的自然资源。《宪法》第9条规定：“矿藏、水流、森林、山岭、草原、荒地、滩涂等自然资源，都属于国家所有”；第10条规定：“城市的土地属于国家所有，农村和城市郊区的土地，除由法律规定属于国家所有的以外，属于集体所有”。因此，国家所有的资源性财产既包括土地、矿藏、水等非生物资源，也包括森林、草原以及陆生动物、水生动物等生物资源。^①

因为大多数自然资源本身所拥有的稀缺性和不可再生性的特性，这类国有资产体现出的国家所有程度之高与立法保护强度之高是较明显的，这从林业、矿产等各种职能部门的设置和各项专门法规的制定就可看出。在这一领域，除去《宪法》与《民法通则》，我国已有的专项立法有：《土地管理法》、《矿产资源法》、《煤炭法》、《水法》、《森林法》、《草原法》、《渔业法》、《野生动物保护法》等，以及大量的实施细则与相关条例。但统计仍显示出：我国耕地资源由于城市建设等原因，每年减少700万亩左右；草场资源更以每年1000万亩的速度被沙化。对资源的盗采、盗伐表明国家所有权并未得到有效的保护，而管理者的破坏性采伐与高成本开发也表明国家财产未得到有效率的使用。这就是国有资源所面临的现实处境。究其原因，自然资源具有较高的经济价值，又有较强的稀缺性。面对经济发展的资源需求，此类财产能带来巨大的收益。因此，自然资源的国家所有权行使，很大程度上是保护与使用、支配与收益的权衡问题。

自然资源的“产权应包括两个部分：一个是对资源的收益权，另一个是对资源的控制权。如果两种权利相互分离，那么无论何时、何地、何人，由于自利行为的出现都会导致效率的损失”。目前资源性财产的国家所有权就存在这种支配与收益的分离，结果，一方面，资源的实际控制权被各地方与部门条块分割，国家的整体权益得不到维护；另一方面，缺乏收益激励的地方部门，或怠于保护，或变管理为占有，导致资源的流失与滥用。权能配置不当与利益冲突，严重影响着所有权的行使效率，造成对自然资源的破坏。

2. 经营性财产：国家所有权的适度保留

国家所有的经营性财产即国家作为出资者在企业中依法拥有的资本及其权益，是国家通过对各种形式的企业进行投资形成的。在经营性国有资产中，21.7万户国有工商企业占用国有资产总额为53306亿元，占有总资产的79.9%；国有金融企业占用8017.4亿元，占有总资产的12%；国有境外企业占用1015亿元，占有总资产的1.5%；建设基金及建设项目占用4410亿元，占有总资产的6.6%。在这一客体领域的国家所有权问题，一直是市场经济建设与国有企业改革的关注焦点。那么，面对市场经济体制，国有企业的经济合理性究竟如何？需要作全面分析。

首先，从总体上讲，经营性资产具有增值性，国有资产在经营性领域的投入是为了盈

^① 在《宪法》没有修改的条件下，物权法建议稿第62条仅将矿产资源定为国有，如果不是缺乏宪法依据的越权立法，也是一种导致法律体系混乱的制度安排。

利。若仅从企业的经营绩效而言，国家所有是低效率的。这已被现实所印证，也是世界性的共识。国家所有权在经营性领域中的低效率，一是源于国家所有权主体的特殊性，即国家的经济人与政治人双重身份。国家的双重行为目标——经济目标与社会政治目标，决定了国家不会单纯考虑经营活动的盈利，这一点与经营性资产的增值性相悖。二是源于国家所有权行使的复杂性。与非国有经济相比，国有经营性财产的所有权行使关系复杂，“国家——政府——企业”的行使环节，使意思表示到行为实施到监督落实的过程漫长，造成了过高的行使成本，自然会引引起经济效率的降低。因此，在竞争性行业领域，国家所有权的退出有助于实现资产的利用效率。而现今的市场经济改革中，国家所有权的适度退出正是在市场竞争中资产有效率地重新配置的体现。

其次，市场也是有缺陷的，在以下的部分经营领域，由于①资本规模要求高，②投资风险大，③盈利预期不充分，市场也无法有效地组织资本的配置与投入。而国家由于其在资本筹集、投资周期等方面的规模优势，以及对社会效益的考虑，能够有效地发挥对市场运作的替代。所以从宏观经济的角度，国家所有权在此中的一些领域，如重要经济命脉部门、大型基本建设、公用事业企业经营领域，仍有存在的必要。

3. 行政性财产：国家所有的必要性

行政性财产即由行政机构占有、使用，在法律上确认为国家所有的各种有价资产的总和，是国家行使其行政职能所必需的财产。国家的存在与运行需要一定的机构，这就意味着一定的场所、设备等固定资产，以及相应的物资和经费，这些都来自国家的财政投入，是机构运转必不可少的。无此财产，国家的实际运转就无法维持，甚至将不复存在，就只剩空洞的抽象人格。在非经营性国有资产中，行政事业单位占用 18557 亿元，占非经营性国有资产总额的 76.6%。

应该看到，虽然物权法建议稿将行政性财产谓之公有物，并定性为非经营性，不以盈利为目的不可融通物，旨在保护公共利益的完整，但由于行政性财产的不可交易，也就存在着无效率使用乃至浪费的情况。其一，行政冗员的存在，以及公务工作的低效率，造成以薪金方式支出的国家财产的浪费；其二，行政性财产的私用，如公车私用、公家电话私用等问题，造成国家财产被私人侵占；其三，超标的公款支出现象，诸如新闻舆论披露的近 40 人编制的机构建起了八层办公楼，东北一个仅需半天的会议千里迢迢行至黄山召开四五天的例子，造成行政费用的居高不下。公务人员对自身利益的谋取，使国家的行政性财产被广泛地“搭便车”占有与使用。然而，即使广泛存在上述情况，行政性财产的国家所有权也无退出的可行性。如何有效地行使国家所有权，管好这部分财产，不仅仅是一个物权法以至民法的问题。经济立法也不仅仅是管理财产的问题，而是管理财产的管理者的问题。例如，缩减编制、裁除冗员，是有关人事制度的问题；控制行政公务设施规模，如限制电话费、公车的配备等，是有关财政制度的问题；有偿使用行政性资产和扩大行政性资产的收益范围^①，则是物权和合同制度的问题。此外，行政性资产的运营管理还涉及纪律检查、监察、审计、检察等各部门的广泛参与。总的来讲，行政性财产的高效管理要加

^① 我们认为，物权法建议稿将收益性行政性资产排斥在行政性资产之外，即采取行政性资产“狭义说”的立法例，不利于行政性资产的有效利用。

强法治意识，综合运用多种法律手段进行综合治理，并主要依靠严格的行政与民事立法和执法加以实现。

4. 公益性财产：国家所有权的必需

公益性财产即国家所有的供公众免费使用的公共设施，如道路、桥梁、路灯、消防设备、绿地、花园等，由国家负责拨款兴建并向社会开放，公众使用时毋需再付费。对此类公益性财产，物权法建议稿谓之公用物，其性质与前述行政性财产的性质相同。

公益设施之所以要免费，第一是由于国家的职能体现。提供必要的、充足的公共设施，是国家承担的社会职能，是国家担负的义务。若是收费，则公众须付费购买使用权，国家的义务就成了权利。第二是由于公益设施的经济属性。这一类公益设施具有消费的非排他性，简单说，街头的交通信号灯，每个人使用时都不能排除，也不需排除他人的使用，若要付费使用的话，每个人都可能在他人付费使用时“搭便车”，从而期待由他人付费。大多数公共设施都是如此。这种存在收费困难的消费品，被经济学称为“公共物品”。

由于私人对公共物品的生产，难以确保能够独占其收益，因而缺乏生产的激励，所以公益性财产只能由国家免费提供，并从财政收入（如税收）中获取兴建与维护费用。更为根源的原因是国家就是为承担诸如提供“公共物品”的社会职能而产生的。公益性财产既然由国家提供，自然应归国家所有。^① 较为特殊的是，近来在现实生活中出现了公用设施付费使用的情况。国家利用借贷等方式筹集的资金修建道路、桥梁等设施，随后通过收费或转让收费权来还贷，这实质上接近于私人在公共物品上的直接投资。这表现了国家在履行公益职能时，运用市场方式对社会资源的引入。据此，物权法建议稿将公用物定性为不可融通物，是片面的。^②

当然，在肯定公用物可以交易的前提下，在以下两点仍坚持其公益性：一是收费仅限于收回投资，而不具盈利性；二是多限于公共设施需求大，而国家财力不足的情况。同时，也要注意这种收费趋势一度有任意扩张的趋向。一些政府机构作为国家所有权的行使者，开始在财政投入兴建的公益财产上收费，例如在修建人行道、护栏等城市公用设施时，向沿路住户摊派受益费等各种名目的费用，或是在一段公路完工后，便设立关卡收取过路费。当收费成为部分政府机构甚至个人的收入来源时，国家的公益性财产变成了某些管理者的经营性财产。也就是说，政府代理人过于强烈的经济人理性，违背了国家提供社会公益服务的政治人身份。因此，要维护国家对公益财产充分的所有权，前提就是要依法防止管理者利用国有财产的私自牟利行为，促使其履行国家法定的有关社会职能。

二、国家所有权主体界定

（一）国家所有权主体界定的意义

我们根据国家财产的经济属性，从其实际管理的情况和效率实现的角度，将国家所有

^① 经济学将使用时排他的消费品称为“私人物品”，使用时不排他的消费品称为“公共物品”。并认为私人物品适于私人所有，公共物品适于国家所有。

^② 为了提高公有物和公用物的使用效率，应该允许公有物和公用物在法律和法规允许的范围内交易。

权客体分为四类。产权经济学家发现，同一种财产客体由不同类型的权利主体所有，会产生不同的利用效率；而不同的财产客体由同一权利主体所有，也会导致不同的效率实现。因此，不同类型的国家所有权客体的主体归属是否界定清楚、合理，是决定国家所有权效率高低的因素。

国家作为所有权主体的意义在于：一部分社会资源的财产权利不为获取收益而存在，却为人们所必需，如公用设施、国防设施等公共物品。因其不能获利，对普通主体缺乏足够的激励，普通主体不会为该所有权支付保护成本、救济成本等排他性费用，因此需要国家以所有人身份出现，对此类财产进行投入和管理，并凭借公法的强制渠道（如税收），获取必要的费用。

现代社会的政治国家具有较复杂和特殊的身份。国家既是一个政治权利实体，又能以民事主体的身份参与到民事关系之中。因此，国家兼有国家主权者和所有权主体双重身份。国家究竟是作为一个民事主体享有所有权，还是作为公共权利的执掌者而享有所有权，是难以绝对分开的。西方国家的国家所有权接近与政府所有权，其所有权内容虽带有行政性特点，但仍受司法规范支配。我国根据国家所有权改革的现实及特色，国家所有权实际上体现的是国家分级所有。分级所有借鉴了财政税收的“分税制”，将国家所有权在国家所有的大前提下，细化为国家所有与地方所有。国家一级的所有权仍由中央政府行使，地方一级的国家所有权则由地方政府行使。将国家财产的所有权行使在国家与地方之间进行合理分配，在各司其责的基础上也可以各得其利。这样的国家所有权主体结构有以下优点：

其一，节约了权利行使成本。国家将一部分财产的支配、管理与收益下放到地方，由地方政府自主管理，大大减少了管理层次，减少了权利行使的环节，降低了成本耗费。

其二，形成了对地方政府的有效激励。分级所有中，地方所有财产的管理直接与地方利益息息相关，改变了以往国家利益与地方利益不一致的冲突，使地方政府产生了有效管理财产、发挥财产最大效用的激励与约束机制。

其三，能够更有效地保护国家财产。不能不说，国家财产遭受损失的原因之一，就是地方政府的怠于保护。而这又有其经济动因：地方政府支付了保护成本，却缺乏相应的收益补偿。分级所有能够改变这种收益与成本的脱钩局面，提高地方政府保护国有财产的主动性与积极性。

（二）国家所有权主体的清晰界定

根据以上思路，对国家所有权主体进行界定要根据不同类型的国家所有权客体的不同特征、不同功能进行多层次界定，使国家所有权主体真正明晰。下面我们也拟从效率角度对国家所有权主体界定问题进行分析。

1. 资源性财产主体界定

资源性财产即国家所有的可供开发和利用的自然资源。如前所述，我国将绝大多数自然资源界定为归属国家所有，国家对这部分资产的立法、司法和执法的保护强度是非常高的，这从林业、矿产等各种职能部门的设置和各项专门法规的制定就可看出。但现实中资源的盗采、盗伐表明国家所有权并未得到有效的保护和最有效率的使用。究其原因，国

土幅员之大、资源分布之广，使国家所有权的管理和保护面临巨大的成本负担。以野生藏羚羊资源为例，藏羚羊以其突出的经济价值遭受偷猎的灭顶之灾，但由于财力不足，常设保护机构不但缺乏必要的交通设备，甚至连人员工资都由个人垫付。大规模的跨省联合打击只能偶尔为之。这样的保护力度使资源的国家所有权归属成为一纸空文，更无效率可言。

再以森林资源为例，国家为保护数量有限而生态作用巨大的森林资源，于1998年发布了禁伐令。而1999年3月4日云南省玉龙雪山的一起森林火灾，表明由于禁伐，减少了林区管理者的收入，使得灭火投入严重不足，林区群众保护态度消极，见火不报。禁伐使森林在避免了刀斧之祸时又受火灾肆虐。禁伐与开禁之争使国有资源的保护举步维艰，归根到底是资源界定缺乏激励机制的问题。

作者认为，自然资源具有较高的经济价值，又有较强的稀缺性。对自然资源的归属界定，既要考虑到较小的保护成本，又要实现有效率地利用。在这样的思路下，自然资源的所有权应作如下界定：

对于小规模的资源，不再界定为国家所有，诸如小面积的林区、水域，小储量的矿藏等。从成本分析，这部分资源的规模较小，国家之外的主体能够承担其管理和保护成本。从利用考虑，非国家的主体由对国家资源的盗用变为对自己资源的利用，将一改不计后果的盲目掠夺，而转为有效率、可持续的利用。而且，私人所有权（即自物权）的设置比使用权（即他物权）更能提供稳定的预期，促进所有者对资源的投入。

对于大规模、大储量的自然资源，仍界定国家所有。这是因为：从成本上，只有国家能够提供大规模的管理和保护，具有规模经济的优势，私人则无此财力承担。从利用上，国家对规模性资源的掌握，有利于维护资源开发与经济发展的宏观均衡，维护资源总体上的可持续利用。

这样的界定，使得国家得以从为数众多的、零散的小规模自然资源的保护中解脱出来，集中力量从事大规模自然资源的保护与利用。而从总体上讲，对自然资源的保护，不仅仅是国家的事，也有了社会力量的参与。非国家的主体不再是单纯的资源利用者，也承担了相应的保护成本。

2. 经营性财产主体界定

经营性财产即国家所有的企业资产。也如前所述，仅从企业的经营绩效考察，国家所有权缺乏完善的效率机制，这已被现实所印证，也是全球性的共识。因此，经营性财产的所有权归属中，国家的退出有利于实现资产的利用效率。国家的退出也正是在市场竞争中，资产有效率地重新界定的体现。但从宏观经济的角度，一些领域（如国家的经济命脉部门、大型基本建设、公用事业企业领域）的经营性财产仍须界定为国家所有。这里强调的是，以上三类企业资产，依据企业规模是否在国民经济中具有全局性或跨地域性意义，可分为国家与地方二级所有。如跨地域的交通动脉、水利枢纽等，应由国家所有；而如城市公用事业之类的企业，宜界定为地方所有。

从微观的角度，企业经营性国有资产的界定，还应注意以下问题：

第一，在国有资产界定的法律依据上，应尽可能适用物权法规则，而非国有资产的行政管理法规。国家所有与非国家所有的界定，是所有权的主体界定，体现为所有权之间的

外部关系，也就是国家主体与非国家主体之间的物权关系。对此适用调整国有资产管理局制定的《国有资产产权界定和产权纠纷处理暂行办法》，有明显的越界之嫌。

第二，在企业新增资产的归属界定上，应遵循谁投资谁收益的原则。确认新增资产源于何种资本增值，则由相应的资本投入者取得所有权，而不应无限扩大国有资产的范围。那种将国有资产参与投入的增值，或仅是作为担保而产生的资本增值，都尽数界定为国家所有的做法，明显有悖公平。

第三，在国家所有权与企业经营权的权能界定上，遵循股权与法人经营权“两权分离”的思路。国家作为所有权主体掌握股权，对经营性资产行使价值形态的终极支配权能；企业拥有相对独立的经营财产权，取得了从国家所有权中分离出的占有、使用、处分权能，并对股东与企业自身的收益负责。

3. 行政性财产主体界定

行政性财产即国家行使其行政职能必需的财产。行政性财产只能界定为国家所有；并根据其所属的政府级别，划分为国家与地方所有。国家的存在与运行需要一定的机构，这就意味着一定的设施、场所等。这是机构运转必不可少的，无此财产，国家就只剩空洞的抽象人格。但应该看到，这部分财产的确存在无效率使用乃至浪费的情况，诸如近40人编制的机构建起了高达八层的办公大楼，科级干部配备高档进口轿车等。然而，行政性财产的归属没有界定给其他主体的可能性，只有强化国家所有权的管理。将国家所有限定在公务的必要范围之内，对于不属于公务行为所需的财产，尽量排除在国家所有之外，以防止将国家财产用于私人的非公务需要。

4. 公益性财产主体界定

公益性财产即国家所有的供公众免费使用的公共设施，如路灯、消防设备、街心花园等，由国家从公众缴纳的税收收入中拨款兴建，公众使用时无需再付费。公益性财产根据是由中央还是地方政府投资建设，分别归属中央或地方所有，总体上以地方所有居多。

如前所述，公益设施之所以要免费使用。第一是由于公益设施的经济属性使然。公益设施具有消费的非排他性，简单说，街头的交通信号灯，一人使用时不能排除他人的使用，每个人都可能在他人付费使用时“搭便车”，从而期待由他人付费。街心花园等公益设施更是如此。这种收费困难的消费品，被经济学称为“公共物品”，只能由国家免费提供。第二是由于国家的职能体现。即使公共物品存在收费的可能，免费提供社会存在和发展所必要的、充足的公共设施，仍然是国家承担的社会职能，是国家对公众所担负的义务。若是公共设施需要付费使用，则公共物品就变成了国家专有的私人物品；而国家提供设施的义务就成了索取报酬的权利。国家无端成了公众的剥削者。前者是更为根源的原因。一定意义上，国家就是为承担诸如提供“公共物品”的社会职能而产生的。公益性财产既然由国家提供、国家管理，自然也应界定为国家所有。

需要指出的是，一些政府机构作为国家所有权的行使者，在实际工作中表现出在公益财产上收费的倾向。例如，在修建人行道、护栏等城市公用设施时，向沿路住户摊派受益费等各种名目的费用，或是在一段公路完工后便设立关卡，收取来往车辆的过路费。当所收费用不是用来集资或还贷，而成为政府的收入时，一些人便真的将公益设施变成了用于牟利的、付费使用的私人物品。使公益性财产变成经营性资产，这是违背国家的社会职能

的。因此，确定公益性财产归属国家所有，前提条件使国家履行法定的有关公共服务职能。

（三）国家所有权主体界定与国有资产的效率

我国所有权主体界定制度主要包括《宪法》中对资源和财产归属的划分、《民法通则》中对原始取得和继受取得的规定、各项自然资源使用权的转让立法、《国有资产评估管理办法》和《国有资产产权界定和产权纠纷处理暂行办法》中针对国有资产界定作出的专门规定，以及从产权再界定意义上的《合同法》、《证券交易法》等保障产权交易秩序的法律，还有《关于企业兼并的暂行办法》、《关于出售国有小型企业产权的暂行办法》等直接规范产权交易的规定。从是否使资源在不同的效用需求间得到合理配置的角度分析，我国现行国家所有权主体界定制度必须进一步完善，表现在：

1. 注重激励机制

虽然我国立法对国家所有权的界定规定较为详细，并且制定了具体的国有资产划分的办法，但却未尽合理。例如，《国有资产产权界定和产权纠纷处理暂行办法》（以下简称《暂行办法》）中不合理地将国家的减税让利和补贴也应划为国家所有，而国家引进外资、发展高新技术企业也实行了许多减税和信贷的优惠，却又不将之列为国有资产，这就造成了国有企业和其他企业的不平等。

众所周知，国有企业为我国的建设事业作出了巨大贡献，在税制改革前缴纳的税率比其他类型企业要高出很多，国有企业的负担使其与其他竞争者不在同一条起跑线上。税制改革的目的是要消除这种竞争上的不平等，可是《暂行办法》的规定却沿袭了这种不平等，必然影响国有企业经营管理者 and 职工的投入积极性。再如，《暂行办法》中只将用于职工工资、奖励、福利等用于个人消费的基金排除在了国有资金之外，其他一切利润和原始资金都划为国有资产，这样的做法无疑抹煞了人力资本的价值。

我们认为，人力资本是生产要素中非常关键的一个环节，正是因为人力资本的差异，从事同一行业、统一规模的企业，有的亏损上亿，有的却能盈利上亿。而在国有企业的所有权界定中，却将人力资本的重要性视为零，企业无论经营的好坏，其资产一律归原始出资者所有，国家占有了过多的剩余。而经营者和劳动者除了其固定的报酬外，工作绩效与收入无关，因此不能给资产的直接经营者提供实现资产最大效用的经济激励。

进一步分析，国家所有权界定办法中存在的种种问题都是因为立法者站在国家的立场之上，试图尽量多地将企业资产划归国家所有。这种做法完全是从静态的角度来考虑的，忽视了企业资产的最终目的是为了增值。国家作为资产的所有者，真正主要的收益来源应该是企业的利润，国家正是为了获取生产经营的收益，才会将资产交由经营者直接支配。确保国有资产的保值增值，应立足于促进国有资产经营管理者 and 劳动者的工作绩效。因此，在制定国家所有权界定法规时，应该纠正尽量扩大国有资产范围的做法，而应该以为资产的增值提供有效的激励机制为核心目标。

2. 强化成本收益比较

应该注意国家所有权界定的收益与成本比较。按照产权经济学的观点，如果交易费用为零，那么防止国有资产流失的最佳途径就是国家所有权界定得越清晰越好，因为国家所

有权保护的净收益肯定大于零。但在现实世界中，国家所有权界定是需支付成本的，有时这种成本甚至很高。这是因为：

首先，确立排他性的国家所有权，需要支付国家所有权界定的技术成本。要确立主体对财产的权限范围，要对财产的性质和数量有一个准确的认识。这一认识有赖于估量、检测和信息的收集等行为，这些都要耗费一定的成本。获得对财产及其权利的准确认识后，还要根据不同的界定方式付出界定成本。例如，通过买卖界定产权，就要付出协商、履行、监督履行等费用；法律界定就要付出查找法规、咨询、申请、实施等活动的相关费用。所有这些都构成界定产权的成本。

其次，初始所有权的界定^①对所有权界定成本的影响。如果初始的所有权安排不具有排他性，那么为避免搭便车行为和资源的耗竭所进行的内部权利分配的成本在逻辑上是很高的。例如，在政府拥有国有资产剩余索取权的条件下，为了使企业产权独立化，就必须进行复杂的企业产权结构安排，形成多级委托代理关系，这不仅不一定能抑制国有资产的流失，而且将支付昂贵的总和代理成本。

再次，所有权的再界定^②对所有权界定成本的影响。国家所有权的再界定是源于所有权关系不可能处于静止的状态，总要借助不同的界定方式，在流动中重新配置。而且，法律关于原始取得即初始界定的制度安排不一定是有效的，通过再界定可以使所有权最终得到有效率的重新配置。大部分的所有权转移正是通过市场手段来重新界定的。但是国家所有权的再界定，无论是寻求行政、法律手段，还是通过市场化的协商合意，都要花费一定的成本。相对于国家所有权的初始界定，再界定的成本一般要高一些。

考虑到国家所有权界定成本，对提高国家所有权界定的收益强调以下两点：

第一，一般来说，资源的稀缺性越强，人们就越有意愿对其所有权进行精确的界定；界定所有权所能带来的收益越大，人们就越有动力对所有权进行精确的界定。如 20 世纪 60 年代以前，各国对领海的要求仅限于 12 海里以内，这是基于海洋在当时的价值——沿海航运与领土岸防。但在公海地相继发现重要的石油、天然气、锰结核等工业资源后，各国召开了第三届联合国海洋法会议，将对海洋的专属利用的相关权益扩大到海岸线外 200 海里。而至今各国仍在为公海的资源利用权甚至所有权要求争论不休。这个经济规律可以解释在现实生活中，越是缺水的地区，就会出现越多的关于水资源使用和管理的法规。水资源的强稀缺和高价值使人们有了承担用水管理的立法、执法、守法成本的必要和动力。而水资源充沛的地区，若不是为了未雨绸缪，人们不会愿意投入水资源的界定成本。企业资产也是这样，在我国没有形成企业产权交易市场时，没有人想到要明晰企业资产的归

① 初始权利界定一般涉及的是对所有权归属的确定，是指所有权及其权利的初次产生和确定，多采用法律界定的方式进行。初始界定既影响到资源配置的公平，也影响到资源配置的效率。一方面，对公平的影响。所有权的初始获得者，不仅意味着获取一定的财产存量，而且意味着获取一定的财产增量和依附在财产上的各种权利。不同的初始界定，影响到社会成员之间获取财产存量的机会及结果的不同，进而影响到财产增量的不同，使经济主体具备不同的竞争起点，最终影响到资源分配的公平。另一方面，对效率的影响。初始界定往往并非最优配置，初始界定的完善与否，决定了再界定的基础，在存在交易费用的情况下，也影响着资源配置的效率。

② 所有权的再界定是对初始界定后的所有权进行重新配置，包括行政界定、法律界定、市场界定三种方式。比如通过法律对被侵占的产权的返还，以及法定继承就是法律界定。而所有权的转移，在多数情况下是一种市场界定。

属，要成立专门的评估机构去对企业资产进行评估。

第二，所有权界定成本越高，人们就越缺乏界定的意愿与动力；而产权界定得越精确，要耗费的界定成本就越高。当界定成本大于预期的财产收益时，人们会放弃界定，该财产尽管有用，也会落入无人所有的境地。比如空气，因为确定归属、进行排他性占有的成本太高，所以人们不会对其进行所有权的界定。可是当这种无主财产的稀缺度进一步提高，价值增加到一定程度，直至收益大于界定成本时，人们不得不花费高昂的成本来界定它。在污染越来越重的情况下，面对清新空气的珍贵，西方已提出了“环境权”的概念。总的来说，随着经济的发展，自然资源的稀缺度和价值会越来越高，人造产品的稀缺度和价值会越来越低；而随着技术的发展，所有权的界定成本会越来越低。因此可预见，自然资源，已成为所有权客体的，如土地，会得到越来越清晰的界定；尚未成为所有权客体的，如宇宙空间、卫星轨道，会逐步被界定归属。

以上对国家所有权界定成本和收益的讨论，说明所有权界定收益与界定成本之间密切相关。简单地说，当界定的成本小于界定的收益时，清晰界定国家所有权归属和范围的要求就产生了。所有权明晰是交易的前提。由于交易促使财产向更高效用的方向流动，能增加财产的价值，随着交易在市场经济中越来越频繁，人们要求清晰界定所有权的愿望也越来越强烈。与此同时，也在尽量降低所有权界定的成本。降低成本可以通过提高技术水平、改进交易信息传递的手段来实现；也可以通过加强制度供给，制定界定规范，并促进交易规模的扩大和交易期限的延长来实现。经济学认为，交易规模越大，平均交易成本越低；交易持续时间越长，双方的风险点越稳定，有利于降低协议达成的成本。所以，一批促进国家所有权市场建立、发育的法规、规章，如《关于企业兼并的暂行办法》、《关于出售国有小型企业产权的暂行办法》，促进交易安全与交易便捷的《合同法》，以及规范资产度量方法的《国有资产产权界定和产权纠纷处理暂行办法》，都起到了降低国家所有权界定成本的作用，从而有利于明晰所有权，提高国有资产配置效率。

参考文献

- [1] 魏杰：《构建新的国有资产管理体制》，江苏人民出版社 1998 年版。
- [2] 肖耿：《产权与中国经济改革》，中国社会科学出版社 1997 年版。
- [3] 张军：《现代产权经济学》，上海三联书店 1994 年版。
- [4] 刘伟，李风圣：《产权通论》，北京出版社 1998 年版。
- [5] [美] 迈克尔·迪屈奇：《交易成本经济学》，王铁生，葛立成译，经济科学出版社 1999 年版。
- [6] 陈郁：《所有权、控制权与激励》，上海三联书店，上海人民出版社 1998 年版。

第二部分

宏观经济增长与改革

增长取向的适应性调整：对地方政府行为演变的一种理论解释*

王 □

在市场化改革与发展中，地方政府由生产活动的直接组织者转向公共产品的提供者已作为市场化改革的目标之一为人们所接受。然而，人们对这种角色转换背后的行为过程的认识却不十分清楚，改革以来地方政府转变职能的实际效果与市场发展的客观需要不适应反映了这一点^①。虽然这与人们对政府职能转变的希望较高有关，但是也反映了这个方面的理论解释力的薄弱。现有的研究主要集中于界定地方政府应该做什么和不应该做什么的行为规范上，而对于地方政府在转轨时期的行为调整过程则缺乏有说服力的实证研究。本文从改革以来的制度变化入手，运用制度与行为的分析框架，把 20 年来地方政府行为演变概括为建立在评价与激励制度基础上、以增长为取向的适应性调整的过程。随着经济条件与社会需求结构的变化，整个社会需要重新构建对地方政府行为业绩的评价与激励机制，这将使未来的地方政府行为取向做出适应性调整。

一、制度变革与行为特征

诺斯 (North, 1993) 认为，制度是经济行为变化的决定因素。市场化改革以来，随着行政性分权的推进，激励制度的变革促使地方政府的行为取向发生了变化。首先，通过把农村经济改革比较成功的联产承包责任制方式引入到中央与地方之间的分配关系中，创造了中央对地方政府的高能激励。在传统的计划体制下，地方政府无论产出多少，都要全额上缴，因此，地方政府缺少推动发展的经济诱因。在财政包干体制下，地方政府作为代理人通过分享剩余收益补偿她的努力 (Moe, 1984; Barzel, 1989)，产出越多，地方政府自留的剩余收益也越多，这成为地方政府尽可能增大产出的基本动力。财政包干制度的推行意味着中央对地方的管理制度由规则管理转向了目标管理。在实施目标管理中，容易测量的、可观察的财政收入、经济增长率等几个重要指标成为对地方政府业绩评价的关键 (Naughton, 2000)。只是随着市场化改革的推进，中央对地方政府的目標管理由以下达的计划指标为依据考核实际执行结果的评价体系转变为市场业绩的相对评价体系。例如，上级部门不再对基层政府下达各项经济计划指标，而是先选择几个关键性的指标，诸如

* 原载：《管理世界》，2004 年第 8 期。

^① 吴佩伦等人对 20 世纪 80 ~ 90 年代中期我国政府职能转换的评价是，“政府职能转换的进展是初步的，也是十分缓慢的，不适应市场经济发展的需要”。参见中国改革与发展报告专家组编著：《中国的道路：1978 ~ 1994 年中国改革与发展报告》，中国财政经济出版社 1995 年版，第 200 页。

GDP、固定资产投资、税收、出口以及利用外资等，然后用这些指标的每年实际完成情况，按照高低顺序对各地区进行排列，并对排列结果加以公布。如果一个地区排位向前移动了，就意味着经济业绩得到了改善；相反，一个地区的排位后移了，就意味着业绩下降。

这种制度变化促使地方政府把本地区经济增长最大化作为基本的行为导向。一方面，因为本地区最大化的经济增长是衡量行为业绩的最具有显示度的指标。在以经济建设为中心的社会发展中，把本地区经济搞上去就是地方政府的最大业绩。而经济搞上去的直接标志就是每年本地区 GDP、财政收入、出口以及利用外资等经济指标的快速增长。另一方面，以增长为取向也是地方政府主要领导人对于制度激励作出的有效反应。这种分配制度与评价业绩体制的改变不仅对地方政府增加收入产生了巨大的激励作用^①，而且对地方主要领导者的晋升与提拔也产生着重要的影响（郑永年、吴国光，1994）。即使业绩良好的地区官员也并不一定马上得到提拔，较快的经济增长也有利于地方政府的多方面利益的实现（Maria Edin，2000）。由于这种强激励，地方政府不仅仅是行政服务与公共产品的提供者，也成为了类似于公司组织的经济行为主体（Jean Oi，1992）。这集中体现在，地方政府因应业绩导向的考核与评价，总是把经济增长放在优先选择的安排上，而其他属于地方政府职能范围的但缺乏相应的指标评价体系，或缺少有效激励的公共服务项目，只能处于次要位置。

在经济转轨的不确定性环境下，适应性调整成为地方政府促进增长的另一个行为特征。转轨时期的不确定环境意味着在由封闭走向开放、由计划体制走向市场以及由二元结构走向一体化进程中的市场机会、要素结构与产业组织的变化以及相应的制度规则调整与重建。地方政府作为公共行政组织，受到法理权力体系固有的影响，对制度规则的依赖程度要比企业组织紧密得多，这使得地方政府把制度规则作为行为反应的基本准则（Cooper，1975；Scott，1995）。然而，在缺乏一套可遵从的制度规则的转轨时期，地方政府担负了企业家式创新者的角色（Jean Oi，1995）。而这种创新的主要目标是使组织规则适应内外部不断变化的环境（C. Banard，1938；F. Hayek，1945；O. Williamson，1996），因此，适应性调整成为地方政府增长取向在不确定性环境下得以实现的一种有效策略。

具体来说，地方政府的适应性调整包含了三方面含义：

一是适时地调整经济增长目标。在各地地方政府制定的发展战略目标中，由于受到依据市场业绩进行相对评价的指标压力，增长率指标的制定往往既不能低于过去，也不能低于与自己的经济实力大致相同的竞争对手。在实施过程中，既考虑不确定环境所能实现的程度，也参照与自己经济实力差不多的其他地区的增长业绩，这使得地方政府总是对发展战略中所设定的增长目标进行调整。以广东省为例，在 21 个地级市实施“九五”经济发展战略计划中，有 12 个地级市对五年中每一年的计划指标都进行了相应的调整，有 4 个地级市对其中四年的每一年计划指标做了相应修订，还有 5 个地级市对其中三年的每一年计划目标进行了重新制定。显然，适应性的目标调整已经成为一种十分普遍的行为。

^① 钱颖一：《中国市场化改革的制度基础》，见：胡鞍钢主编：《中国走向》，浙江人民出版社 2000 年版，第 171 页。

二是适时地调整增长途径。在传统的计划体制下，政府扮演了社会生产活动的决策者角色。20世纪80年代中期以来实施的所有权与经营权分离，使地方政府把生产经营权转给了企业，而自己利用所有权，大量地投资经营性项目，并采取公司化运作方式，对所投资的项目进行经营与管理，从而保证本地区经济的快速增长。20世纪90年代后期，当政府投资的大量经营性项目受到体制与宏观经济环境制约而陷入了经营困境的条件下，地方政府开始收缩投资战线，调整投资重点，把主要的投资领域转向公共产品与服务，这既可以实现地方政府的职能转变，又能拉动本地区的经济增长（吴敬琏等，1998）。而在转向公共产品的投资中，受到追求增长业绩最大化的利益驱动，各地方政府偏重于对增长产生直接效应的交通、道路、通讯和能源等生产性基础设施项目，而对供水、绿地、垃圾处理等生活性公共设施投资不足，这种不平衡又成为下一轮调整的主要内容（王战，2003）。

三是适时地调整地方政府的职能边界。区域试验被认为是中国经济体制转轨的一个基本特征（林毅夫等，1994）。但是，在授权进行率先试验的地区中，没有一个地区的地方政府职能是一步转变到位的。从安徽的农村联产承包责任制到四川的厂长、经理承包制，再到广东顺德的产权制度改革等等都说明了这一点。这表明，在缺少可遵从的制度规则条件下，一些通过授权率先进行变革试验的地方政府“在采取行动之前，往往会作一番仔细评估。首先，他们要知道上级部门及同伴对一项政策及实施方法的态度；其次，他们要关心一些内外力量对这些政策及实施方法的接受程度，这些力量对他们的职业生涯、经济福利以及组织本身的福利都有重要影响。再次，他们要调查清楚有什么东西会阻挡他们的雄心，包括制定与实施政策的愿望等”^①。因此，受到不确定性制度环境的影响，一个地区的率先试验也只是比其他地区“先走一步”，而不会走得太远。

20多年来地方政府的行为调整经历了从经营企业到经营城市（社区）再到经营园区的三个阶段。本文以这三个阶段演化的内在逻辑为实例来论述制度调整与地方政府行为取向以及策略之间的互动过程。

二、经营企业的行为与调整

自财政包干为发端的行政性分权制度推行以来，增加财政收入成为地方政府谋求经济最大化增长的基本目标，而这个目标的实现取决于大批企业的成长。在市场化改革初期，全社会范围内95%以上的企业都是由各级政府投资与经营的国有与集体企业，非公有经济受到法律、行政制度以及政策等方面的诸多限制，在国民经济中处于无足轻重的地位，这使得地方政府把发展企业的重点仍然放在国有及集体所有的企业身上，直接介入企业的投资、生产经营与管理自然成为地方政府促进企业发展的主要行为。许多学者（Walder，1997；Jean Oi，1996）都大量地论述了这个时期地方政府直接介入企业活动的各种表现，如直接从银行为企业筹措资金，或提供贷款担保，安排劳动力，任免企业领导人，征用土地，争取短缺资源，申请各种进出口贸易许可证及外汇额度的批文等等。

^① 约瑟夫·拉巴隆巴拉（Joseph Lapalombara）：《组织中的权力与政治：公司组织比较》，见：迈诺尔夫·迪尔克斯（Meinolf Dierkes）等主编：《组织学习与知识创新》，上海人民出版社2001年版，第435页。

这种地方政府直接经营企业行为不仅源于以公有制为主的制度基础，而且，卖方市场环境决定了这种行为的有效性。在卖方市场环境中，只要能够找到投资，无论上什么项目、生产什么档次的产品，都会找到销路。在地方政府比企业更有能力获得各种短缺性资源的体制下，地方政府与其把自己手里的各项企业经营管理权限下放给企业，不如自己充当市场经济活动的主角，而且，地方政府扮演主角的行为绩效也是比较明显的。林青松、伯德（1989）等人对我国乡镇政府与乡镇企业快速成长的研究，Kang Chen, Jefferson and Singh（1992）等人对改革以来中国国有企业全要素生产率有所改善的研究以及王珺（1999）对省、地（市）、县等地方政府所属的国有企业绩效高于中央所属的国有企业绩效的观点都印证了这一点。

尽管地方政府充当经济活动主角的行为后果以及对建立市场制度的阻碍影响早已被一些学者所认识^①，但是，在卖方市场环境下，这种行为所给地方政府带来的快速增长以及财政收入增加等方面的收益大于其付出的代价，因此地方政府并没有产生较强的调整这种行为的需求。在整个 20 世纪 80 年代乃至 90 年代中期，几乎没有一个地方政府针对这些问题率先全面地推进以产权制度改革为核心的市场化改革的事实就是例证。

引起地方政府对直接介入企业生产经营活动的行为进行调整的根本原因是卖方市场向买方市场的转变。随着买方市场的来临，许多由各级政府投资的国有及集体企业开始出现了持续性的经营亏损。在出现亏损初期，地方政府还试图采取各种措施帮助其渡过难关。比如，增加政府的财政补贴和直接出面说服银行不断贷款给其所属企业以及提供各种扶植性政策等^②。然而，当这些措施不仅没有从根本上解决企业持续性亏损问题，反而给地方政府带来了巨大的财政负担，使地方政府“甩包袱”的动机应运而生。特别是在金融秩序的规范化和银行部门实行垂直性管理体制以来，靠地方政府出面、通过与当地的银行和金融机构协调，来解决企业负债和资金周转困难的方式已经变得不适用。在这种情况下，少数地区，如广东顺德和山东诸城等地区率先推进了以界定政府与企业之间产权关系为核心的改革试验，随后，全国许多地区在模仿与改良的基础上，也迅速地开展了这方面的改革。

三、经营城市（社区）的行为与调整

20 世纪 90 年代中期以来，虽然地方政府不能直接介入企业的生产经营活动，但是以增长为取向的制度激励并未改变，这促使地方政府寻找新的增长途径与资源。而大规模地基础设施建设投资成为地方政府推进增长的主要途径与增长来源。一方面，随着买方市场

^① 董辅初在 1979 年就提出了国家与企业的财产关系是经济体制改革的关键，参见《关于我国社会主义所有制形式问题》，见：董辅初著：《经济体制改革研究》，经济科学出版社 1995 年版；蒋一苇也在 1979 年提出，“改革的根本任务是确定企业的位置”，参见《企业本位论刍议》，见：蒋一苇等著：《股份制的理论与实践》，中国人民大学出版社 1988 年版。

^② 这里所说的扶植性政策主要指诸如降低使用土地费用，帮助筹集资金，与科研院所主动沟通，出面协调资金周转困难，组织各种项目论证会和产品展销会，提高企业知名度等。

的来临，经济增长来自消费的拉动力减弱^①，进而对投资拉动的需求开始上升。特别是日常工业品及耐用消费品等竞争性工业的快速成长，对城镇交通、能源及公路等基础设施项目产生了极大的市场需求，使基础设施改善成为地方政府投资的重点。另一方面，城镇基础设施与公共环境是公共产品，地方政府充当这种公共产品的计划者与投资者符合市场化改革对地方政府职能定位的要求。因此，这种改革与发展的客观要求推动了地方政府行为向公共设施建设的转变。

基础设施的全面建设需要大量的资本投入。在公共财政不足以支撑巨大的基础设施投资需求的条件下，经营城镇就成为地方政府动员社会资本推进公共设施建设的创新性制度安排。经营城镇指的是把公共设施开发所提升的城镇价值作为主要资源，通过转让升值的城镇资源，吸引非政府的资本参与公共设施的开发与建设，从而实现以城兴城、滚动发展的良性循环。在城镇资源中，土地是最主要的经营资源，而城镇土地资源是通过公共设施的建设与开发得到增值的。这种增值的城镇资源可以通过两种方式转让：一种是事后转让。地方政府以土地使用权作为抵押，获取了第一笔基建贷款后，对规划用地先行投资，改善公共设施环境。然后，将公共设施开发所带来的升值的土地经营权进行招标、拍卖，吸引社会资金参与公共设施环境的滚动式开发。率先实施经营城市的大连、上海等城市主要采用了这种方式。20世纪90年代后期，大连市通过土地使用权的抵押从银行获得了新增建设投资268亿元。在公共设施环境得到改善后，再实施专营权有偿转让，进而弥补了后续的建设资金不足。上海市1991~2001年的十年间，通过土地批租共筹集了超过1000亿元的资金。把这笔资金用于新城区开发与旧城区改造，又带动了数倍的社会资金投入^②。另一种是事先转让。在城镇规划指导下，地方政府把公共设施改善后土地经营权的升值预期作为专营权，吸引社会资金参股进行公共设施的建设与开发。这是一种以未来地价升值的商业机会换取现期的公共设施建设资金的做法。浙江与广东的一些城市公共设施建设更多地采用了这种做法。“九五”期间，浙江省用这种方式动员了390亿的民间与境外资金用于城镇建设，这笔资金占整个城镇建设资金总量528亿元的74%。广东东莞市用于城镇建设的资金达到了254亿元，其中，以这种方式动员的社会资金接近了140亿元，占城镇建设总量的55%，使政府投入与社会资金投入的比例为1:1.24。在广州上百亿元的环城路建设中，将近1/3的投资是以这种方式吸纳的社会资金。当然，无论是哪一种转让方式，都有利于把社会资本引入基础设施建设项目，这不仅弥补了公共财政资金之不足，也对降低公共产品的供应垄断、改进效率具有重要意义（世界银行，2003）。

这种以城镇建设为中心的经营机制在一些市场经济较发达省市被率先采用后，全国各地的城镇政府也纷纷效仿。随着经营城镇在全国各城镇的普遍推广，以城镇建设和房地产投资为中心的投资迅速地增长。统计数据显示，1998~2002年间，全国的全社会固定资产

① 2002年我国消费率为58.2%，比“六五”时期的66.1%低7.9个百分点，为改革开放以来的历史最低水平，而世界平均消费率为80%左右。发展经济学家钱纳里等人的一项实证研究表明：在人均国内生产总值为1000美元时，世界各国居民消费率一般为61%。2002年我国人均GDP已接近1000美元，但我国居民消费率却仅为48%，比国际平均水平低13个百分点。

② 王战主编：《WTO元年与上海发展思路创新：2002~2003年上海发展报告》，上海财经大学出版社2003年版，第92页。

投资每年增长 11.3%，其中，基建投资与房地产投资分别增长了 11.4% 和 21.1%，两个项目投资总量占全社会固定资产投资的比重也由 54.7% 上升到了 58.5%。使国内生产总值对固定资产投资变动的弹性系数由 1998 年的 0.38 上升至 2002 年的 0.46^①。而基建投资主要集中在城市规模扩展与基础设施环境的改善上，从而加速了城市化进程。2002 年城市（不包括辖县）实现国内生产总值比 2001 年增长了 13%，高出全国国内生产总值的增长速度 5 个百分点，占全国国内生产总值的 63%。城市实现财政收入增长了 22%，比 2001 年提高了 3.2 个百分点。随着城市增长能力的增强，我国城市化水平从 1990 年的 19.0% 提高到了 2002 年的 37.3%^②。

然而，经营城镇的关键在于土地经营权的转让，而这种转让又取决于预期的商业机会。如果预期的商业机会多，升值机会较高的地价可以吸引社会资金进入。一些处于交通枢纽的大城市，如上海、广州、大连以及杭州等因本身的商业机会较多且资源聚集效应明显而有效地实现了较高地价的有效转让。如果预期的商业机会有限，那么，即使地价较低，也难以吸引社会资金进入。一些地理位置偏僻的城市以及中小城镇在通过土地使用权的有偿转让获取社会资金方面就受到了预期商业机会有限的制约，20 世纪 90 年代中期进入大规模基础设施建设的珠海市就是这样的例子。虽然珠海政府以自己掌握的土地为抵押，获得了大笔公共设施建设资金，但是，由于缺少商业机会，许多企业买了地、设了厂，最后还是搬走了，其结果，已改善的投资环境利用率不足。对于许多缺少商业机会的中小城镇政府来说，不推进公共设施的建设，经济增长率会受到影响。要推进公共设施建设，只能依赖政府财政投资以及银行贷款，而银行贷款又转化为对地方政府的财政压力，这使得许多中小城镇政府不能像大城市一样全面推进基础设施建设，而是将投资重点从整个经济缩小到一个局部范围，园区开发就成为这些地区率先突破建设能力与商业机会有限的新途径。

四、经营园区的行为与调整

经营园区指的是在城镇规划基础上，通过对划定的地区提供土地租金、管理收费等方面的优惠政策和供应完备的基础设施、引入投资咨询、工商登记以及技术服务等各种中介服务机构以及采用公司化运作机制，吸引和培植企业的聚集化发展。与经营城镇相同的是，园区基础设施环境作为公共产品的一部分，不仅需要由地方政府提供，而且地方政府对园区建设的配套开发符合投资推动增长的战略需要。同时，在此基础上，将园区经营权委托给园区管理公司或大企业的公司化运作方式也适应了市场化的改革要求。

然而，与经营城镇相比，经营园区还有许多不同之处。首先，园区开发所需要的投资规模要小得多。整个城镇环境的改善不仅需要旧城区的大范围拆迁与全面改造，也需要拓展新的城区范围。而一般设立在地价较低的城郊结合部的园区不仅仅面积要小得多，而且单一的生产功能不需要配置与生活环境相关的基础设施项目。因而开发园区与中小城市

① 全国的全社会固定资产投资的数据来自于相关各年度的《中国统计年鉴》。

② 上述指标数据来自历年的《中国城市统计年鉴》。

的有限投资能力相适应。其次，园区适合以增量带动存量的空间结构调整需要。增加新项目、投入新资源是经济增长的一个重要途径，而新项目和新兴产业的配置需要有大量的发展空间。在城镇已有的空间被传统产业填满的条件下，与其将传统产业挪出去，腾出空间来发展新项目，不如开辟新的空间来发展新的项目。这既可以节省传统产业迁移所支付的大量调整成本，也可以避免传统产业迁出后新项目尚未成长起来所带来的增长震荡，还可以通过传统产业向已设立的园区转移，进而保证城镇的功能规划得以实现。最后，设立园区也成为地方政府直接参与地区竞争的一种有效手段。随着市场制度的完善和交易费用的降低，生产要素的跨地区流动日益增多。由于生产要素流动对本地财政收入与经济增长都产生着重要影响，于是，各地都把实施优惠政策作为招商与保商的重要手段。对于地方政府来说，与其在整个城镇范围内实施优惠政策，不如设立一个实施优惠政策的园区更为有利。20世纪90年代后期，珠江三角洲等地的一些企业向长江三角洲地区的园区流动，从而促使珠江三角洲等地各级政府纷纷设立以优惠政策为主的各种产业园区的事实说明了这一点。

在增长压力与有限财力约束的条件下，地方政府开始大举兴办各种类型的园区。如以IT、软件、生物、医药为主的高新技术园区；以服装、陶瓷、家具、农产品加工为主的传统产业园区；以台商、外资、民营、留学生等投资来源界定的园区等。据国土资源部的调查统计，目前我国各类园区与开发区已达5658个，其中经国务院批准的只有232家，省级批准的1019家，各类开发区规划面积3.6万平方公里，超过了城镇建设用地总量。

随着园区数量的迅速增加，园区之间争抢投资项目的过度竞争突显了出来。在全国的园区数量有限的条件下，每个园区都可以从市场上引入足够数量的符合园区定位的投资项目，致使园区内土地资源闲置较少，园区经营收益也较高。例如，天津、广州、苏州等经国务院于20世纪80~90年代初期审批的43个技术开发区和工业园区平均每平方公里的工业产值为50亿~80亿元。近年来，园区数量的迅速增加使每个园区难以吸收到适合园区定位的、充足的投资项目，这必然造成园区土地资源的闲置与园区经营效益的下降。一项研究表明，2001年，我国经营较好的园区每平方公里的工业产值在1亿~10亿元之间。如上海等地开发区每平方公里的平均产出额为15亿元，区县和乡镇政府兴办的园区产出额不到1亿元^①。目前，土地利用率达不到50%的园区占了我国园区总量的接近40%。园区土地资源的闲置迫使地方政府采取两种过度竞争的手段：一是降低园区进入的技术门槛。一些高新技术产业园区在无法有效地吸纳诸如新材料、电子信息、生物医药等新项目的条件下，就把一些低附加值的传统产业项目吸引进来，造成了园区规划对产业发展的定位与已引入的产业不一致的结果。二是加大提高园区的政策优惠力度。比如，一些城市园区开发每亩地的成本（包括拆迁费用）约为30万元，而工业用地报价却在1.5万~5万元。在税收方面，虽然国家税务总局明确规定，地方政府不得擅自决定税收的减免，但是一些地方政府与企业私下达成协议，以奖励方式变相地将地税返还给企业，实行10年减免等。个别地方政府甚至对引入的项目贷款进行贴息1~2个百分点，并按照项目投资规

^① 王战主编：《WTO元年与上海发展思路创新：2002~2003年上海发展报告》，上海财经大学出版社2003年版，第125页。

模给予 15% 的资金配套，这部分投资不算是资本金，不参与分红、不收利息，也不干预经营，如果项目失败，也不求补偿等。其结果，园区的兴办大大地加剧了地方政府的财政负担。许多地方政府截留或延缓支付对农民的征地补偿费用就是这种财政压力的结果。

虽然兴办经营园区有利于城镇规划的实现与产业聚集化发展，但是，在“县县建园、镇镇设区”的重复建设中，园区的优势无法显示出来。目前，从中央到地方的各级政府开始对园区进行清理与撤并，提高园区的设立标准，这使得依靠设立经营园区来推动地方增长的行为又遇到了挑战。

五、结论性评述：未来调整的趋势展望

概括 20 多年来地方政府经营企业、经营城镇与经营园区的三个阶段，增长取向的适应性调整构成了其行为演变的一条主线。其中，追求本地区 GDP 的快速增长是这种行为取向的直接表现，业绩评价与激励机制是这种行为取向得以持续的制度基础。在未来的发展中，以增长为取向的适应性调整行为是否还会持续下去？无疑，面对不断变化的环境，地方政府的适应性调整作为一种策略并不会改变，但是增长取向将会发生以协调发展为取向的调整。这种调整主要是由对地方政府行为绩效的评价机制的改变所引起的。未来对地方政府行为绩效的评价机制将会发生三个方面的变化：

第一，评价地方政府行为业绩的指标体系将由以 GDP、人均 GDP 等反映经济增长的指标为主向经济、社会、环境与生态等协调发展的综合指标体系扩展。在经历了 20 多年的高速增长和人均收入翻两番之后，人们开始把市场秩序、公共教育、环境保护、社会福利、医疗保健以及垃圾处理等公共产品与服务作为生活质量整体改善的一个重要方面，这要求地方政府将工作重心由致力于整个地区的经济增长变为协调发展上来。评价指标体系的重构与扩展不仅反映了社会需求结构的变化，也成为引导地方政府行为调整的“指挥棒”。目前，一些地区开始尝试引入联合国开发计划署提出的，由平均寿命、平均受教育年限以及人均 GDP 等三个指标构成的人文发展指数作为衡量地方政府行为业绩的主要指标。还有一些地区鉴于这种人文发展指标没有包含自然资源因素，而用世界银行（1992）提出的“绿色 GDP”作为评价协调发展的核心指标等。目前，虽然在全国范围内尚未有一个统一的反映协调发展的指标体系，但是，这种探索表明了评价地方政府业绩的政策取向已开始发生了变化。

第二，评价地方政府政绩的主体将由以上级部门为主向上级部门、中介机构与公众共同评价转变。这种转变不仅仅是增加了评价主体，而是改变了评价机制。在由上级部门作为地方政府行为绩效的评价主体时，受到时间、精力以及信息不对称的约束，上级部门不可能对每个基层组织在经过了事无巨细地调查后作出评价，只能以统计指标、上报材料以及走马看花的考察等有限的信息作出评价。下级部门为适应上级部门的这种评价机制而披露有利于自己的信息，而隐匿不利的信息，“数字出官”和“形象工程”等就是这种评价机制的结果。改变这种评价的有效方式是引入中介组织与公众的评价。中介机构的非盈利性和专业性使其在反映公众评价方面具有不可替代的独特优势。奥斯特罗姆（Ostrom, 1999）认为，居民对公共产品属性的评价是很准确的。中介机构通过了解公众对地方政府

提供公共产品的满意程度对地方政府行为产生了重要的影响，使地方政府不仅关注上级部门的偏好，也要考虑公众的需求。事实上，由权威性的学术机构对地方政府行为绩效的评价已成为世界上许多国家的普遍做法^①。近年来，我国一些学术机构和中介机构也开始对地方政府的公共政策及创新行为进行评价^②。随着这种评价方式在全国范围内的扩大以及以这种方式全面评价地方政府行为绩效的展开，考虑公众的发展需求在评价地方政府业绩中将发挥越来越重要的作用。

第三，评价范围由决策结果扩展到决策过程。公共选择理论提出了政府决策往往因受到利益集团的影响而可能偏离社会需求的看法。而治理这种偏离性的办法是增加透明度与广泛的公众参与决策过程。因此，对地方政府业绩的评价不仅通过人们对公共政策实施结果的反应加以体现，而且通过公众参与制定政策的过程反映出来。公众参与决策过程既是对“长官意志”的一种约束，也是对其制定一些有利于本身利益团体而偏离公众需要的政策的一种制衡，因而作为改善地方政府治理结构的一个有效途径，被列入对地方政府业绩的评价范围。当然，公众参与决策过程与评价决策后果都要求增加政务信息的透明度。目前，各地政府开始加大政务公开化程度的尝试，如上海市已将政务公开由基层单位到市、区等政府部门的延伸，由执法部门到经济管理部门的扩展等，以及湖南省长沙市的四级联动政务公开制度、贵阳市的人大旁听制度、江苏省沭阳县首创的干部任前公示等。在增加信息透明度的基础上，一些地方政府也开始探讨政府与居民之间对话与沟通的方式与渠道，如四川省平昌县公开评税制度、广州市对公共服务项目价格调整的听证制度以及广东省省长与专家和企业家定期对话制度等。虽然处于尝试阶段的做法还不十分完善，但是它反映了评价机制调整的方向与趋势，沿着这个导向来设计对地方政府行为业绩的评价体系，公开性与参与性必然成为显示决策过程的重要质量指标。

因此，面对正在发生变化的评价机制，作为直接反应的行为取向也将会作出相应的调整，那就是地方政府通过适应性调整逐步确立协调发展的取向以适应业绩评价机制的变化。至于什么时候能够实现这种行为取向的转换，这主要取决于这种评价机制开始有效运作的时间。

参考文献

- [1] Banard, C. I., *The Functions of the Executive* Cambridge, MA: Harvard University Press, 1938.
- [2] Hayek, F. A., *The Use of Knowledge in Society*, *American Economic Review*, 1945, Vol. 35, pp519 ~ 530.
- [3] Williamson, O. E., *Strategy Research: Governance and Competence Perspectives*, Berkeley: Business and Public Policy, 1999.
- [4] Kang Chen, Jefferson and Singh, *Lessons from China's Economic Reform*, *Journal of Comparative Economics*,

① 例如，哈佛大学肯尼迪政府学院每年组织若干政治学与行政学专家，对美国各级政府的活动进行独立的研究与评估。阿根廷、巴西、菲律宾等也是如此。香港是在一些大学设立电话调查中心，调查公众对突发事件以及领导人威信等方面作出的反应。

② 中共中央编译局比较政治与经济研究中心和中央党校比较政治研究中心于2000年开始组织发起对“地方政府创新奖励”进行评选。广州市社情民意调查中心每年对出台的一些重大政策以及发展业绩通过电话访谈等方式听取公众的反应等。

1992, Vol.16.

- [5] Jean Oi, Fiscal Reform and the Economic Foundations of Local State Corporation, World Policies, 1992, Vol.45, No.1, pp99 ~ 126.
- [6] Jean Oi, The Role of the Local State in China's Transitional Economy, In Andrew Walder (ed), China's Transitional Economy, New York: Oxford University Press, 1996, pp170 ~ 187.
- [7] Yoram Barzel, Economic Analysis of Property Rights, New York: Cambridge University Press, 1989.
- [8] Terry Moe, The New Economics of Organization, American Journal of Political Science, 1984, Vol.28, No.4, pp739 ~ 777.
- [9] Cooper, J., Strengthening the Congress: An Organizational Analysis, Harvard Journal of Legislation, 1975, Vol.12, pp307 ~ 68.
- [10] Scott, R. W., Institutions and Organizations, Thousand Oaks, CA: Sage, 1995.
- [11] 奥斯特罗姆 (Ostrom, 1999) 等著:《制度激励与可持续发展》,陈幽泓等译,上海三联书店 2000 年版。
- [12] 世界银行:《2003 年世界发展报告:变革世界中的可持续发展——改进制度、增长模式与生活质量》,中国财政经济出版社 2003 年版。
- [13] 吴敬琏等:《国有经济的战略性改组》,中国发展出版社 1998 年版。
- [14] 王战主编:《WTO 元年与上海发展思路创新:2002 ~ 2003 年上海发展报告》,上海财经大学出版社 2003 年版。
- [15] 华尔德 (Andrew Walder):《中国产权改革的非私有化道路》,见:甘阳,崔之元编:《中国改革的政治经济学》,香港牛津大学出版社 1997 年版,第 79 ~ 98 页。
- [16] 林毅夫,蔡昉,李周:《中国的奇迹:发展战略与经济改革》,上海三联书店,上海人民出版社 1994 年版。
- [17] 林青松,伯德主编:《中国农村工业:结构、发展与改革》,经济科学出版社 1989 年版。
- [18] 王珺:《我国政企关系演变的实证逻辑》,《经济研究》,1999 年第 11 期。
- [19] Barry Naughton:《中国:在促进人类发展中的政府作用》,见:胡鞍钢,王绍光编:《政府与市场》,中国计划出版社 2000 年版,第 229 ~ 252 页。
- [20] 郑永年,吴国光:《论中央与地方关系:中国制度转型中的一个轴心问题》,《当代中国研究》(普林斯顿),1994 年第 6 期。
- [21] Maria Edin:《基层干部的政治激励机制:中国地方政府发展经济的动力》,《香港社会科学学报》(香港),2000 年,第 45 ~ 66 页。
- [22] 道格拉斯·诺斯 (Douglass C. North) 著:《制度、制度变迁与经济绩效》,刘守英译,上海三联书店 1993 年版。

当期收入还是永久收入：转型时期 中国居民消费模式的检验^{*}

王 曦

依据当期收入和依据生命周期 - 永久收入消费是消费的两种典型模式，这两种理论对于消费者行为基础的认定有着本质的区别，它们暗示的应用性宏观计量模型也不一样。检验转型时期我国居民到底符合什么样的消费模式是很必要的，这是深入研究我国消费的必然开端。

一、两种消费模式

依据当期收入（Current Income）和依据生命周期永久收入（Life Cycle-Permanent Income）消费是消费的两种典型模式，它们都有各自的理论依据。

消费的当期收入模式是指消费者只依据当期可支配收入水平（ Y_t ）来确定消费（ C_t ）。即：

$$C_t = C_0 + cY_t \quad (C_0 \text{ 和 } c \text{ 为正常数}) \quad (1)$$

对此的解释最早可以追溯到 20 世纪 30 年代凯恩斯的“绝对收入理论”。现代消费理论对这种消费模式有了更多的解释，主要是把它与短视行为和流动性约束联系起来。

凯恩斯之后，库兹涅茨研究了美国内战以来消费和储蓄的行为，发现宏观变量的实际运动与凯恩斯的预言在很多地方相悖，在当时被称为“消费函数之谜”。

为了解释“消费函数之谜”，从 50 年代开始，涌现出了很多消费函数理论。其中最有影响的是莫迪里安尼的“生命周期假说”和弗里德曼的“永久收入假说”。这两种假说对消费者行为基础有着共同的设定：①没有流动约束。②消费者具有时间偏好。③效用函数有规避风险的特征。④消费者在给定的预算约束下追求一生期望效用最大化。因此它们有着类似的理论框架和结论，后人因而合称它们为生命周期 - 永久收入理论，本文以后简称这两种理论为“永久收入理论”，依据这两种理论的消费称为消费的永久收入模式。下面的模型是该理论的一个现代版本（Attanasio, 1999）：

$$\text{Max } E_t \sum_{j=0}^{\infty} \beta^j U(C_{t+j}) \quad (0 < \beta < 1) \quad (2)$$

满足约束：

^{*} 本文是笔者博士论文《中国转型经济总需求分析——微观基础与总量运行》的一部分，研究得到□川博士论文重大创新项目基金的资助；成文过程中，美国普林斯顿大学计量经济中心 Han Hong 教授和中山大学计量经济学博士生导师王美今教授认真审阅了原稿并提出了很多宝贵意见，特此感谢。原载：《世界经济》，2002 年第 12 期。

$$A_{t+j+1} = R_{t+j}(A_{t+j} + Y_{t+j} - C_{t+j}) \quad (j = 0, 1, \dots, \quad ; A_t \text{ 给定}) \quad (3)$$

这里, E_t 表示以时间 t 能获得的信息为条件的预期; β 为时间折现系数; A 为实际财富; Y 为实际劳动收入; C 为实际消费; $R_t = 1 + r_t$, r_t 为实际利率; 效用函数 $U(C)$ 被假定是时际可分的。

求解这个随机规划问题, 通过一阶条件可以推导出下面的欧拉方程:

$$\beta E_t [R_{t+1} U'(C_{t+1})] = U'(C_t) \quad (4)$$

其中, $U'(C)$ 为消费的边际效用。式 (4) 表明, 消费者力图在一生中根据 β 和 R 使 $U'(C)$ 平滑化。

可见, 当期收入理论和永久收入理论在内涵上有着本质的区别, 它们对于消费模式的预言也截然不同。当期收入理论中消费由当期收入惟一决定, 而永久收入理论表明:

- ① 人一生中消费模式具有“独立性”, 消费的变化只与永久收入有关, 与收入方式无关。
- ② 消费水平具有“稳定性”, 只有收入的“剧变”和“持久冲击”才会影响消费。

在我国最近的文献中, 利用永久收入理论及其变形来分析中国的消费问题成为热点。但是如果现实中我国居民很大部分按当期收入模式消费, 那么这些研究的出发点就错了。因此, 对我国的消费模式进行检验是十分重要的, 这是进一步分析我国消费的必然开端。在我国, 对此的研究还比较少。樊纲等 (1992) 的分析支持了永久收入假说; 臧旭恒 (1994) 利用了 Hall 检验, 认为永久收入假说在中国的适应性逐渐加大; 王信 (1996) 的分析拒绝了永久收入假说; 而在赵志军 (1998) 和贺菊煌 (2000) 的分析中, 结果是不明朗的。

由于本文下一部分将要陈述的原因, 这些检验在方法上存在着不足。另外, 其检验结果也是相互矛盾的, 一个可能的原因就是两种消费模式是同时存在的, 而现有的检验不包含这种可能性。

二、永久收入理论的检验方法

检验居民消费是否符合永久收入模式, 有两个方向: 对理论本身的检验和对欧拉方程的检验。

在对于理论本身的检验方面, 对于式 (2) 和 (3) 表示的基本问题, 如果效用函数是二次型, R 固定, $\beta R = 1$, 则可推导出永久收入理论的简化公式:

$$\begin{aligned} C_t &= Y_t^p \\ &= (1 - R^{-1}) \left[\sum_{j=0}^{\infty} E_t(Y_{t+j}) R^{-j} + A_t \right] \end{aligned} \quad (5)$$

利用这个公式和美国的时间序列数据, Flavin (1981) 和 Sargent (1980) 研究表明, 消费对于当期收入是过度敏感的, 因此, 式 (5) 的消费模式被拒绝。

而对欧拉方程的检验成为后来检验的基础。Hall (1978) 首先迈出重要的一步。如果 R 为常数, U 为二次型, 则根据式 (4) 的欧拉方程, 有:

$$\Delta C_t = \epsilon_t \quad (6)$$

其中： ε_t 为扰动项。式 (6) 表明，消费服从一阶马尔可夫过程，并且没有其他变量是消费戈兰杰因果意义上的原因。这常常被称之为消费的“随机游动”性。Hall 考虑了多种可能的其他解释变量，发现滞后的股票市场价格对消费的变动有预测功效，因此，Hall 正式拒绝该模型。

Hall 对效用函数和利率的假定过于严格，并且他使用消费的绝对变动量。但是时间序列研究发现，宏观变量的绝对变动量一般是不平稳 (Non-stationary) 的，对不平稳的变量进行随机游动检验是有偏的。因此，Hall 之后的研究放松了他的假定，把常相对风险厌恶 (CRRA) 效用函数，即 $U(C) = (C^{1-\sigma} - 1)/(1 - \sigma)$ ，代入欧拉方程，可以推导出该方程的另一种形式：

$$\Delta \ln C_t = \sigma^{-1} \ln R_t + \alpha + \varepsilon_t \quad (7)$$

其中： α 为常数； σ 为自然对数； $\Delta \ln C$ 表示实际消费增长率。这方面的一个代表性工作见 Hansen 和 Singleton (1983)。利用式 (7) 进行检验的好处是使用实际消费增长率指标，避免了变量不平稳性的问题。因此，类似式 (7) 的公式成为后来检验的基础。

利用泰勒级数展开容易证明，在 r 接近 0 的时候， $\ln R = \ln(1+r) \approx r$ ，则式 (7) 可以改写为：

$$\Delta \ln C_t = \sigma^{-1} r_t + \alpha + \varepsilon_t \quad (8)$$

后来的发展中，一组很有特色的检验是 Campbell 和 Mankiw (1989, 1991) 做出的。他们的检验基于这样的理论设定：总收入 Y_t 可以分成两个部分。其中第一组消费者收入占 Y_t 的比重为 λ ($1 \geq \lambda \geq 0$)，他们按照“拇指原则”消费，即消费者完全消费当期收入。以下标 1 和 2 代表分组，则有： $C_{1,t} = Y_{1,t} = \lambda Y_t$ 。第二组消费者获得 $(1 - \lambda)Y_t$ ，按照永久收入模式消费，其消费 $C_{2,t}$ 符合式 (8)。

接着，Campbell 等认为，总消费变动率等于两组消费变动率的加权平均值，另外为了方便起见，以后我们以小写字母 c 和 y 代表 $\ln C$ 和 $\ln Y$ ，则有：

$$\begin{aligned} \Delta C_t &= \lambda \Delta c_{1,t} + (1 - \lambda) \Delta c_{2,t} \\ &= \lambda \Delta y_t + (1 - \lambda) \sigma^{-1} r_t + (1 - \lambda) \alpha + (1 - \lambda) \varepsilon_t \\ &= \lambda \Delta y_t + \theta r_t + \mu + v \end{aligned} \quad (9)$$

其中： λ 、 θ 和 μ 为常数； $v_t = (1 - \lambda) \varepsilon_t$ ，仍为扰动项。如果所有消费者都按照永久收入模式消费， λ 应为 0。如果 λ 显著不为 0，则可以根据 λ 的估计值计算出不同消费者的比例。但笔者发现，由上面的收入分组模式不能得出式 (9) 的结论，这是 Campbell 等原始推导中的一个失误。这个错误容易纠正，如果 λ 被定义为符合当期收入模式的消费占总消费的比重，问题就解决了。^① 修正后的 λ 定义对于 Campbell 检验的具体步骤没有任何影响，只是在说明检验结果时我们要小心。总体上，Campbell 检验还是适用的。

① 首先由 $C_t = C_{1,t} + C_{2,t}$ ，有 $\Delta C_t = \Delta C_{1,t} + \Delta C_{2,t}$ ；又有 $\Delta C_t / C_t = (C_{1,t} / C_t) (\Delta C_{1,t} / C_{1,t}) + (C_{2,t} / C_t) (\Delta C_{2,t} / C_{2,t})$ ，则 $\Delta c_t = (C_{1,t} / C_t) \Delta c_{1,t} + (C_{2,t} / C_t) \Delta c_{2,t}$ 。若式 (9) $\Delta c_t = \lambda \Delta c_{1,t} + (1 - \lambda) \Delta c_{2,t}$ 恒成立，应有 $\lambda = C_{1,t} / C_t$ ，即 λ 应代表总消费中第一组消费者消费所占的比重。但是原来对 λ 的设定是 $\lambda = Y_{1,t} / Y_t = Y_{1,t} / (Y_{1,t} + Y_{2,t})$ 。对比 λ 的两个方程式，并由 $C_{1,t} = Y_{1,t}$ ，我们将得到 $C_{2,t} = Y_{2,t}$ ，即第二组的消费者也按照“拇指原则”消费。若 $C_{1,t} = \alpha Y_{1,t}$ ，我们也会有类似的结论。

与其他的检验相比, Campbell 检验有明显的优势: ①避免了 Hall 检验的不足。②包含了两种消费模式同时存在的可能。③通过检验结果, 可以显性地计算出符合不同模式的消费者比率。④它考虑了解释变量和扰动项相关的可能性, 运用工具变量法进行两阶段最小二乘法估计, 使检验更加科学。Attanasio (1999) 因此认为, 它可能是所有检验中最重要。

另外, 若消费者消费当期收入的一个固定比例, 即 $C_{1,t} = cY_{1,t}$, 也会得到与式 (9) 一样的检验方程, Campbell 检验也适用。因此, 这里的“拇指原则”可以看作简化的当期收入模式, 则 Campbell 检验实际上是检验了当期和永久收入两种消费模式的相对重要性。对于中国的消费, 本文采用 Campbell 检验, 并且对于 λ 采用修正过的定义。

三、转型时期我国居民消费模式的检验

1. 估计方法

对于式(9)的检验方程, 一个问题就是解释变量 Δy_t 和 r_t 与扰动项 v_t 几乎肯定是相关的。一种可能是正相关。 ε_t 在经济含义上是指符合永久收入模式的消费者对于永久收入预测(从而当期消费)的修正, 而当期收入有利的波动通常也是消费者对于未来收入有良好预期的时候, 因此 ε_t (从而 v_t) 与 Δy_t 正相关。 Δy_t 与 ε_t 负相关的可能也同样存在, 例如以事后高税收弥补现时军费开支的增加。这时, 当期收入的增加对经济主体意味着永久收入的减少。 v_t 与 Δy_t 也有不相关的可能性, 此时收入的波动完全是由暂时性的冲击(如经济周期)造成的, 消费者不会修改对于永久收入的预期。类似地, r_t 的变动会通过改变预期财富收益从而改变永久收入, 因此 r_t 与 v_t 也存在很大的相关可能。

相关性意味着用普通最小二乘法 (OLS) 估计式 (9) 是无效的, OLS 估计会导致 λ 估计值的偏倚 (Biased)。解决这个问题的办法是使用工具变量 (IV), 用两阶段最小二乘法 (TSLS) 估计 λ 。另外, IV 法 TSLS 回归同时包容了 Δy_t 和 r_t 与 v_t 不相关的可能, 若不相关, 该方法对 λ 的一致估计与 OLS 估计是相同的。

2. 工具变量的选取

使用 IV 最重要的问题就是寻找合适的 IV, 如果一个变量或向量与 v_t 不相关而与 Δy_t 和 r_t 高度相关, 并且是平稳的 (Stationary), 它就是合适的 IV。

对于这个问题, 永久收入理论本身提供了重要的线索。消费者依照永久收入模式消费, 这种行为本身就意味着他对于未来收入和实际利率具有前瞻性的预见, 他的预期必然是理性的。理性预期是指利用所有可能的信息对于未来做出的最好预测, 因此这种预测的偏差 ε_t (从而 v_t) 必然与仅仅包含过去信息的任何滞后变量正交 (Orthogonal) 或不相关。这说明, 任何滞后变量都有可能成为合适的工具变量。具体确定 IV 也不是盲目的, 我们可以根据经济含义进行有的放矢的选择。为了保障检验的普遍意义, 我们同时考虑了三个滞后变量集: ①滞后的 Δy 和 r 。对于 Δy_t , 这意味着经济周期的一种形式; 对于 r_t , 由于 r_t 等于名义利率减去通货膨胀率, 这意味着货币政策在时间上的联系。②滞后的 Δc 和 r 。考虑滞后的 Δc , 是因为根据永久收入理论, 以前的消费包含着对当期收入和利率的预期, 因此它们之间可能存在

相关关系。③滞后的 Δy 和 r 再加上 $(c_{t-1} - y_{t-1})$ 。其含义在于，如果把方程(9)表示为 IV 的函数，我们将发现这是 Δc_t 关于 $\Delta y_{t,j}$ 和 $r_{t,j}$ 的一个误差修正模型 (ECM)。

3. 数据

样本期限为 1978~1999 年，为年度数据。

文献或资料中，我国可支配收入指标的来源或估算共有三种：①狭义口径，定义为消费加上居民储蓄。其中居民储蓄是指居民银行储蓄存款增量（郭树清和韩文秀，1991）。②广义口径，认为居民储蓄应该包括实物形式（固定资产投资、商品存货）和金融资产形式（包括银行存款、现金和有价证券）。中国人民银行研究局课题组（1999）持这种看法。③基于家庭抽样调查的指标。表 1 给出了这三种方法的比较。

表 1 全国居民人均可支配收入的三种口径

单位：人民币元

年份	1980	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997
广义口径	266	537	634	723	884	1007	1093	1216	1474	1696	2334	3063	3788	3971
狭义口径	244	463	522	602	743	984	965	1071	1276	1610	2259	2890	3339	3430
抽样调查	247	479	540	599	709	804	904	970	1127	1388	1872	2363	2782	3009

数据来源：广义口径的原始总量数据来自中国人民银行研究局课题组（1999），这里进行了人均处理；抽样调查指标的原始数据来自相关各年《中国统计年鉴》，这里根据农村和城镇的人口比重进行了加权平均；狭义口径为本文计算。

总的来看，广义口径指标的数值最高，狭义口径次之，抽样调查最低，这并不奇怪。首先，在抽样调查中，被调查对象出于逃税、避税和藏富的目的，倾向于隐瞒收入；其次，广义口径比狭义口径要广。从定义上看，广义口径显得更加合理，但是，使用广义口径的问题就是除了消费、储蓄存款和私人固定资产投资以外，我们没有任何有客观依据的数据。年鉴中对于商品存货、现金和有价证券的统计都是全国总量，而我们需要的是私人居民的数据。如果利用广义口径估计可支配收入，势必引起主观误差。

本文采用狭义口径计算可支配收入，主要是因为：①由于虚报，抽样调查指标肯定是不合理的。②可以避免广义口径的主观估计误差。③本文采用式（9）作为主检验方程，其解释变量均为增长率指标。如果狭义口径指标的增长率与广义口径指标的增长率高度相关，对检验结果不会有大的影响。实际上，这也是事实，详见后文对于“其他影响检验的可能”的进一步分析。

设定基期为 1978 年。定义：人均实际可支配收入 = 居民人均实际消费 + 居民人均储蓄存款实际余额增量。其中，人均实际消费 = 基期人均名义消费 × 定基比居民消费指数 / 100；居民人均储蓄实际余额增量 = Δ （居民人均储蓄名义余额 / 定基比商品零售物价指数 × 100）。最后计算出居民人均实际可支配收入增长率 (Δy) 和人均消费实际增长率 (Δc)。

实际利率 (r) 有两种：事前实际利率 = (名义利率年初值 - π^e)；事后实际利率 =

(名义利率年末值 - π)。其中, π 和 π^e 分别为通货膨胀率及其预期。由于居民的消费计划是在年初制定的, 因此这里应该采用事前实际利率。假定 π^e 为静态预期, 则 $\pi^e = \pi_{-1}$ 。 π 以商品零售价格指数计算。

所有基础数据均来源于《中国统计年鉴》, 详细数据见表 2。

表 2 基础统计数据

年份	消费指数 定基比	实际居 民可支 配收入 (亿元)	人口 (万人)	年初名 义利率 (%)	环比商品 零售价格 指数	人均实 际消费 (元)	人均实 际可支 配收入 (元)	$\Delta c(\%)$	$\Delta y(\%)$	$r(\%)$
1978	100.00	1962	96259	3.24	100.7	184.00	203.83	4.5	4.49	2.30
1979	106.50	2206	97542	3.24	102.0	196.07	226.16	6.5	10.96	2.54
1980	115.77	2447	98705	3.96	106.0	213.09	247.91	8.7	9.62	1.96
1981	125.03	2716	100072	5.40	102.4	230.10	271.40	8.0	9.48	- 0.60
1982	133.15	3033	101654	5.40	101.9	245.06	298.37	6.5	9.93	3.00
1983	143.54	3488	103008	5.40	101.5	264.24	338.61	7.8	13.49	3.86
1984	160.19	3995	104357	5.76	102.8	294.98	382.82	11.6	13.05	4.26
1985	181.17	4432	105851	5.76	108.8	333.65	418.70	13.1	9.37	2.96
1986	189.15	5087	107507	5.76	106.0	348.40	473.18	4.4	13.01	- 1.60
1987	199.74	5613	109300	7.20	107.3	368.10	513.54	5.6	8.53	1.20
1988	214.52	6211	111026	7.20	118.5	395.43	559.42	7.4	8.93	- 0.10
1989	213.45	6602	112704	7.20	117.8	393.37	585.78	- 0.5	4.71	- 9.86
1990	220.70	6879	114333	8.64	102.1	406.78	601.66	3.4	2.71	- 6.46
1991	239.02	7261	115823	11.34	102.9	440.51	626.90	8.3	4.20	6.54
1992	269.86	8257	117171	8.64	105.4	497.44	704.70	12.9	12.41	4.66
1993	291.71	8387	118517	7.56	113.2	537.77	707.66	8.1	0.42	2.16
1994	304.26	9743	119850	7.56	121.7	560.87	812.93	4.3	14.88	- 2.22
1995	327.08	11417	121121	10.98	114.8	602.95	942.61	7.5	15.95	- 10.72
1996	356.84	13466	122389	10.98	106.1	657.82	1100.26	9.1	16.72	- 3.82
1997	371.83	14087	123626	10.98	100.8	685.26	1139.49	4.2	3.56	1.37
1998	392.28	14948	124810	7.47	97.4	723.31	1197.68	5.5	5.11	5.27
1999	421.31	16754	125909	6.07	97.0	775.21	1330.68	7.4	11.10	6.38

(四) 检验

我们首先利用 ADF 检验 (Augmented Dickey-Fuller Test) 检验了 Δc 、 Δy 和 r 的平稳性, 发现所有变量都是平稳的。由于这个结论十分平凡^①, 检验的细节不再赘述。平稳性保证了我们以后的回归分析的有效性, 不会引起伪回归 (Spurious Regression)。

依据前述的实证检验程序, 检验结果见表 3。

表 3 经济转型中的消费模式——实证结果

工具变量	工具变量适用性		使用 IV 的 TSLS 估计		
	Δy	r	λ	θ	联合显著性
1 无(OLS)	—	—	0.42 (2.93)**	0.38 (3.41)**	$R^2 = 0.48$ $F = 8.73^{**}$ $\chi^2 = 17.5^{**}$
2 Δy_{-3} Δy_{-4} Δy_{-5} r_{-1} r_{-2}	$R^2 = 0.52$ $F = 2.374^*$ $\chi^2 = 11.87^{**}$	$R^2 = 0.69$ $F = 4.947^{**}$ $\chi^2 = 24.74^{**}$	0.51 (1.94)*	0.45 (2.56)**	$F = 3.55^*$ $\chi^2 = 7.1^{**}$
3 Δc_{-3} Δc_{-5} r_{-1} r_{-2}	$R^2 = 0.29$ $F = 1.21(0.36)$ $\chi^2 = 4.84(0.3)$	$R^2 = 0.45$ $F = 2.44^*$ $\chi^2 = 9.77^{**}$	—	—	—
4 Δy_{-3} Δy_{-4} Δy_{-5} r_{-1} r_{-2} $c_{-1} - y_{-1}$	$R^2 = 0.53$ $F = 1.85(0.19)$ $\chi^2 = 11.1^*$	$R^2 = 0.7$ $F = 3.9^{**}$ $\chi^2 = 23.4^{**}$	0.56 (2.09)*	0.48 (2.72)**	$F = 4.07^{**}$ $\chi^2 = 8.14^{**}$

说明: ①第 1 列为实际使用的 IV。为了对比, 第 1 行未采用任何 IV, 直接为 OLS 回归结果。第 2~4 行分别对应于前面预设的不同 IV 集。对于每个预设 IV 集, 我们考虑了 IV 最大滞后 6 阶的各种可能组合, 最后选出了对 Δy 和 r 最有解释力的 IV, 即为表中的 IV。

②IV 适用性的检验包括第 2 和第 3 列的第 2、3、4 行, 分别对应于 Δy 和 r , 是把 Δy 和 r 分别对每行 IV 进行 OLS 回归的结果。这两列的每一行中, 上面是 R^2 值, 中间和下面是对 IV 联合显著性的 Wald 检验(即所有 IV 的系数均为 0, 例如对于第 2 行, 零假设是 Δy_{-3} 、 Δy_{-4} 、 Δy_{-5} 、 r_{-1} 和 r_{-2} 的系数均为 0)的 F 值和卡方值(χ^2)。

③ TSLS 估计结果包括第 4、5 和第 6 列的第 2 和第 4 行, 通过 Eview 3.1 软件中使用 IV 的 TSLS 命令直接实现。其中, 第 4 和第 5 列分别对应 λ 和 θ 。这两列的每一行中, 上面的数据是 λ 和 θ 的估计值, 下面括号中的是系数显著性检验的 t 值。由于第二阶段回归的 R^2 并不重要, 第 6 列给出 λ 和 θ 联合显著性的 Wald 检验结果, 检验的 F 值和 χ^2 值全部列出。

④“**”和“*”分别表示在 5% 和 10% 的水平上通过显著性检验, 未通过 10% 检验的统计量的显著水平 (p 值) 在其后面的括号中给出; 所有回归均包含常数项。

① 对于宏观实际经济变量, 总量的增长率指标以及收益率指标一般都是平稳的。

由表 3, 经验分析有以下结果:

(1) 在 IV 的适用性检验方面: ① 第 2 行的 IV($\Delta y_{-3}, \Delta y_{-4}, \Delta y_{-5}, r_{-1}, r_{-2}$) 对 Δy 和 r 都通过了系数联合显著性检验, 有解释力, 是合适的 IV。② 第 3 行的 IV 集($\Delta y_{-3}, \Delta y_{-5}, r_{-1}, r_{-2}$) 虽然对 r 通过了系数联合显著性检验, 但是对 Δy 不能通过联合显著性检验, 不是合适的 IV。③ 第 4 行的 IV($\Delta y_{-3}, \Delta y_{-4}, \Delta y_{-5}, r_{-1}, r_{-2}, c_{-1} - y_{-1}$) 对 r 通过了系数联合显著性检验, 对 Δy 的联合显著性通过了 χ^2 检验, 但未通过 F 检验($p = 0.19$)。总地说, 这个 IV 对 Δy 和 r 是有解释力的, 也是合适的 IV 集。

(2) 在 TSLs 估计方面, 对于合适的 IV(第 2 行和第 4 行), λ 和 θ 均通过了显著性的 t 检验, 估计是有效的。对应于第 2 行和第 4 行的 IV 集, λ 的估计分别为 0.51 和 0.56; θ 分别为 0.45 和 0.48。

(3) 第 1 行 OLS 回归对 λ 和 θ 的估计分别是 0.42 和 0.38, 对比 TSLs 回归, 我们发现, OLS 回归对于 λ 和 θ 的估计都是向下偏的。

5. 其他影响检验的可能

我们还考虑了其他影响检验结果的可能性, 限于篇幅, 简述如下:

(1) 可支配收入的口径。为了避免由于统计口径问题引起的对 λ 和 θ 的估计误差, 我们还考虑了居民可支配收入的广义口径。采用中国人民银行研究局课题组(1999)对居民实际收入的总量数据, 我们计算了广义口径的 Δy 。再结合本文的 Δc 和 r 数据, 我们重复了以上的 TSLs 检验, 数据期间为 1979 ~ 1997 年。结果显示: ① ($\Delta y_{-3}, \Delta y_{-4}, \Delta y_{-5}, r_{-1}, r_{-2}, \ln c_{-1} - \ln y_{-1}$) 为合适的工具变量。对于 Δy 和 r , 系数联合显著性检验的 F 值分别为 3.85 ($p = 0.05$) 和 1.85 ($p = 0.22$), χ^2 值分别为 23.1 ($p = 0.001$) 和 11.1 ($p = 0.086$)。② TSLs 回归对 θ 的估计是 0.55 ($t = 1.76, p = 0.10$), 对 λ 的估计是 0.46 ($t = 2.28, p = 0.04$)。③ OLS 对 λ 和 θ 的估计仍然是向下偏的。对比表 ③ 可知, 使用居民收入的不同口径, 对 λ 和 θ 的估计基本上是一致的。

(2) 效用函数的时际不可分性 (Intertemporal Inseparability)。文献中, 有两种情况要求效用函数具有时际不可分的特征: 耐用品消费理论和对于消费习惯的考虑。耐用品消费理论强调效用取决于所拥有的资产存量, 它等于过去购买与折旧之差。于是, 某个时期的购买量就会影响下一个时期的效用。利用这个框架, Mankiw (1982) 的结论是, 支出的变化是一个一阶移动平均过程 (MA1)。另一个方向是对于消费习惯的研究, 这时耐用品消费理论中的耐用品存量被解释为广义的“心理”存货, 其基本含义是: 低于过去消费的现期消费将会导致效用的下降, 因此习惯将引致消费滞后, 见 Constantinides (1990)。

由于我国没有完整的耐用品和非耐用品消费指标, 本文检验全部消费品。为了考虑耐用品消费和消费习惯的影响, 结合 Mankiw 和 Constantinides 的理论, 我们也检验了式 (9) 的一个变形:

$$\Delta c_t = \Delta \phi c_{t-1} + \lambda \Delta y_t + \theta r_t + \mu + v_t \quad (\mu \text{ 是常数}) \quad (10)$$

由于前定变量 (Predetermined Variable) Δc_{t-1} 本身就是自己的合适 IV, 因此在最终选定的 IV 中应该包含 Δc_{t-1} 。在表 3 的三组 IV 中分别加入 Δc_{t-1} , 我们对式 (10) 重复了 TSLs 分析。结果发现: ① ($\Delta y_{-3}, \Delta y_{-4}, \Delta y_{-5}, r_{-1}, r_{-2}, \Delta c_{t-1}$) 构成了合适的 IV。对于 Δy

和 r ，系数联合显著性检验的 F 值分别为 6.68 ($p=0.003$) 和 3.3 ($p=0.04$)， χ^2 分别为 40.1 ($p=0.001$) 和 19.8 ($p=0.003$)。②TOLS 对 ϕ 的估计不能通过显著性检验：对 ϕ 的估计值是 0.23 ($t=1.22, p=0.24$)。式 (10) 的模型还不如式 (9) 表现得好。

(3) λ 不是常数。考虑到经济转型的巨大冲击，一些消费者可能会改变其消费模式，因而 λ 可能不是常数。

首先考虑 λ 的变化趋势 TR_t 。设 $\lambda_t = a + bTR_t$ ，其中 a 和 b 为待估计系数。我们考虑了 TR_t 的多种替代可能，分别以 t (时间)、 $\ln t$ 、 C_t 、 $\ln C_t$ 、 Y_t 和 $\ln Y_t$ 替代 TR_t 。使用 $(\Delta y_{t-3}, \Delta y_{t-4}, \Delta y_{t-5}, r_{t-1}, r_{t-2}, \Delta r_{t-2})$ 作为 IV，我们应用非线性 TOLS 方法估计了 b 。对应于 TR_t 的这些替代， b 的估计分别是 -0.03, -0.36, -0.001, -0.51, -0.001 和 -0.42； t 值分别是 -0.87, -0.85, -0.83, -0.82, -0.87 和 -0.85； p 值分别是 0.4, 0.41, 0.42, 0.43, 0.4 和 0.41。虽然实证估计表明 b 总体上为负，显示 λ 可能有下降的趋势，但是所有的趋势替代都不能通过显著性检验，说明 λ 下降的趋势并不显著。考虑 λ 趋势的模型还不如原来的模型表现得好。另外一个可能是 λ 含有结构性变化 (Structural Change)，利用 Chow 检验进行分析，我们同样拒绝了这种可能。

(4) 相对收入模式。某些居民可能会依据相对收入模式消费，但我们认为这种消费模式并不重要。因为：①相对收入理论预言，消费具有向下的刚性，而这与各国的宏观事实不符。也正因为如此，后来它在国外并没有成为主流的消费理论。②赵人伟等 (1997) 对于基尼系数的分析表明，我国居民收入分配差距的逐渐增大。如果相对收入理论适用，居民储蓄将减少，但这与我国总体上上升的人均储蓄率 (定义为居民实际储蓄增量除以实际可支配收入，数据来源于本文计算) 不符，见图 1。

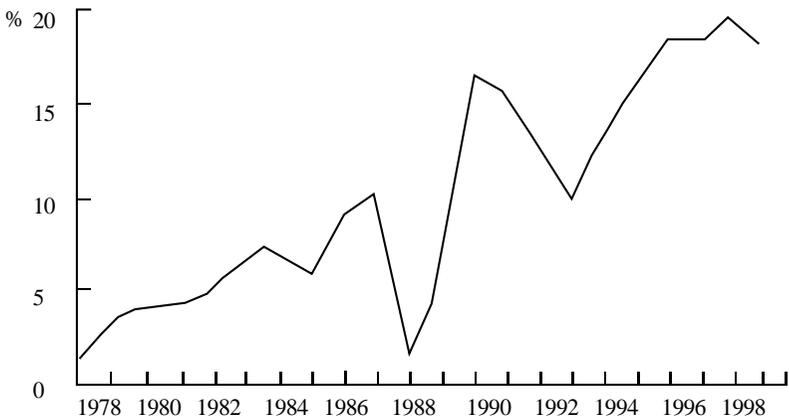


图 1 我国居民的实际储蓄率

但在臧旭恒 (1994) 的研究中似乎有证据支持相对收入模式。他使用了下面的模型：

$$C_t / Y_t = a + bY_a / Y_t \quad (11)$$

其中 a 和 b 为正常数； C_t 和 Y_t 代表个人消费和收入； Y_a 代表社会平均收入。臧旭恒认为上式的含义在于：个人消费倾向 (C_t / Y_t) 是他在收入分配中相对位置的函数。如果 Y_t 低，则个人

消费倾向就高,反之亦然。接着,他利用 1981 ~ 1991 年收入水平分组数据证明了上式中 b 的显著性。

我们仔细考察了他的研究方法,发现其中有一定的缺陷,因而他的研究不能看作是对相对收入模式的正确支持。首先,式(11)也可以用永久收入理论来解释:其中, Y_p 是消费者永久收入预期的某种代表; Y_t 是暂时收入的代表。在生命周期的不同阶段,消费者暂时收入会有波动。在 Y_t 低的时候,消费者为了平滑一生中的消费,消费倾向 C_t/Y_t 自然就高,反之亦然。其次,他忽略了利率的作用,使得模型过度简化,这可能会引起模型的设定错误 (Specification Error)。

四、结论与含义

实证分析表明, λ 在我国显著不为0,对 λ 的TOLS估计是0.51或0.56;并且其他影响检验结果的可能性也被排除。因此,我们有理由相信,我国总消费对当期收入是过度敏感的。在我国的总消费中,50%强的消费根据当期收入进行;其他的消费是按照永久收入模式进行的,两种消费模式都十分重要。对于深入研究我国消费行为,这个结论是一个很好的经验分析基础。

我国总消费的过度敏感性可能有两个方面的解释:①短视行为。长久以来,我国经济体制改革在探索中进行,很多时候,改革者都不知道未来会发生什么。对于消费者,这意味着他缺乏足够的信息来预测未来的收入。这样,部分消费者可能会选择把消费支出与当期或近期可预测的收入联系起来,即消费计划是短视的。②流动性约束。对此的深入讨论见 Deaton (1991),这里只考虑一个简单的情况:一个消费者生存两期,每期有收入 Y_1 和 Y_2 ,没有初始财富,没有效用的时间偏好,储蓄利率为0。由于消费的边际效用递减,消费者的最优消费是每期消费 $(Y_1 + Y_2)/2$;并且如果 $Y_1 < Y_2$,消费者必须在第一期借入 $(Y_2 - Y_1)/2$ 的资金。但如果存在流动性约束,消费者不能借钱,作为次优选择,他在第一期就只能消费 Y_1 ,第二期消费 Y_2 。这样,消费者就表现为依据“拇指原则”来消费,即 $C_t = Y_t$ 。可见,对于面临流动性约束的消费者,其支出和当期收入紧密相连。与发达国家相比,我国居民的可支配收入还很低;另外,我国的个人消费信贷市场还处于萌芽阶段,因此,我国居民面临很强的流动性约束。短视行为和流动性约束的共同作用表现在宏观上,就是总消费对当期收入的过度敏感性。在这方面,叶海云(2000)进行了初步的理论探讨,本文的结果可以视为对他工作的经验分析支持。

经验分析还发现,OLS方法对 λ 的估计是向下偏的。这意味着式(9)中 Δy_t 与 v_t (或 ε_t) 是负相关的,在经济含义上,这表示当期收入的增加对经济主体意味着永久收入预期的减少。其原因在于:转型时期特殊的产权制度安排产生了大量的短期行为,而短期行为中当期收入的增加是以未来收入的降低为代价的。①在农业领域,1978年“家庭承包责任制”的实行使得农业迅速发展,但1985年后农业的增长放慢。这是因为,没有土地的所有权,农民对将来土地使用权就存在不确定性,他们就会更加看重短期利益。其现实表现为农民和集体减少了对土地的许多长期改进:1994年农业基础设施真实投资水平只是1979年的

58%。②工业领域的现有制度安排，不能为国有企业经理追求短期利益提供足够的约束。首先，国企经理 60 岁就要退休；其次，经理的利益来自于行政提升、企业规模和效益的结合，而不是企业的长期发展；最后，监督者的收入与企业效益无关，没有足够的激励去监督企业。

本文分析的政策含义在于：①居民消费的增长率 (Δc) 与当期可支配收入的增长率密切正相关。这意味着要刺激消费，增加居民实际可支配收入的政策是有效率的。②理论 [见式 (8)] 和经验分析 (见表 3) 都表明，我国消费增长率与事前实际利率正相关。这是因为，事前实际利率上升，消费者就会预期其未来财富收益从而永久收入上升，因此，当期消费提高。这意味着，试图通过降低名义利率进而实际利率来刺激消费的政策不一定可行。如果政策当局降低利率的行为被预期到，消费者就会减少消费和增加储蓄，而不是相反。但这个作用只针对于事前利率，现时利率与消费的关系究竟如何，有待于进一步研究。③应该尽快建立和完善我国的社会主义市场经济体制，大力促进我国消费信贷市场的发展，以消除短视行为和缓解流动性约束，提高居民的福利水平。④经济中的短期行为有着重要的影响，要消除短期行为，促进经济的长期发展，必须深化产权制度改革。

参考文献

- [1] Attanasio, O. P., Consumption Demand, In Handbook of Macroeconomics, 1999, Vol.3, Amsterdam: North Holland.
- [2] Campbell, J. Y. and N. G. Mankiw, Consumption, Income and Interest Rates: Reinterpreting the Time Series Evidence, NBER Working Paper, 1989.
- [3] Campbell, J. Y. and N. G. Mankiw, The Response of Consumption to Income: A Cross-Country investigation, European Economic Review, 1991, Vol.35, pp715 ~ 721.
- [4] Constantinides, G. M., Habit Formation: A Resolution of the Equity Premium Puzzle, Journal of political Economy, 1990, Vol.104, pp219 ~ 240
- [5] Deaton, A. S., Savings and Liquidity Constraints, Econometrica, 1991, Vol.59, pp1221 ~ 1248.
- [6] Flavin, M., The Adjustment of Consumption to Changing Expectations about Future Income, Journal of Political Economy, 1981, Vol.9, pp974 ~ 1009.
- [7] Hall, R. E., Stochastic Implications of the Life Cycle-Permanent Income Hypothesis, Theory and Evidence, Journal of Political Economy, 1978, Vol.6, pp971 ~ 987.
- [8] Hansen, L. P. and K. J., Singleton Stochastic Consumption, Risk Aversion and the Temporal Behavior of Asset Returns, Journal of Political Economy, 1983, Vol.91, pp249 ~ 265.
- [9] Mankiw, N. G., Hall's Consumption Hypothesis and Durable Goods, Journal of Monetary Economics, 1982, Vol.10, pp417 ~ 425.
- [10] Sargent, T. J., Rational Expectations, Econometric Exogeneity and Consumption, Journal of Political Economy, 1980, Vol.86, No.4, pp673 ~ 700.
- [11] 樊纲, 余根钱:《体制改革时期的储蓄增长——对近年来中国城镇居民储蓄增长原因的分析》,《金融研究》,1992年第6期。
- [12] 贺菊煌:《消费函数分析》,社会科学文献出版社2000年版。
- [13] 郭树清, 韩文秀:《中国 GNP 的分配与使用》,中国人民大学出版社1991年版。
- [14] 叶海云:《试论流动性约束、短视行为与我国消费需求疲软的关系》,《经济研究》,2000年第11期。

- [15] 王信：《我国居民收入高速增长时期的储蓄存款分析》，《经济科学》，1996年第5期。
- [16] 臧旭恒：《中国消费函数分析》，上海三联书店，上海人民出版社1994年版。
- [17] 赵人伟，李实：《中国居民收入差距的扩大及其原因》，《经济研究》，1997年第9期。
- [18] 赵志军：《我国居民储蓄率的变动和因素分析》，《数量经济技术经济研究》，1998年第8期。
- [19] 中国人民银行研究局课题组：《中国国民储蓄和居民储蓄的影响因素》，《经济研究》，1999年第5期。

中国城市经济增长的趋同分析*

徐现祥 李 郇

一、引言

Lucas 在其 1988 年的经典文献中早已强调了城市在经济增长中的作用,认为城市是人力资本的集中地,城市的存在与发展是人力资本外部性的体现,可以作为人力资本存量的一种度量形式(Lucas, 1988)。2002 年, Lucas 把城市视为先进生产技术积聚的场所,进而把城市化视为劳动密集型技术向人力资本密集新技术转移的过程,构建城市化模型(Lucas, 2002)。Lucas 对城市在经济增长中作用的理解,暗示了可能存在着不同于以国家或省为单位的经济增长特性,但为现有趋同文献所忽视。

基于此,本文采用趋同分析的标准方法,就我国 216 个地级及以上城市经济增长的趋同情况展开讨论,结果发现,在 20 世纪 90 年代城市间的经济增长存在 δ 趋同和绝对 β 趋同,这与我国省区间的趋同模式明显不同。我们进而探索其背后的趋同机制,发现不仅存在新古典增长理论所强调的趋同机制(资本边际报酬递减),而且还存在新增长理论所强调的趋同机制(技术扩散、转移),这也有别于省区趋同机制。

本文第二部分是文献综述;第三部分是数据来源;第四部分是实证分析;第五部分是对实证结果的进一步讨论;第六部分是结论性评述。

二、趋同文献综述^①

趋同研究始于 Baumol (1986) 和 Abramovitz (1986)。Baumol (1986) 采用 Maddison 数据,发现工业化国家存在趋同。但该发现被认为是建立在样本选择有偏的基础上(Delong, 1988)。因为,如果把样本扩展到非工业化国家,则不存在趋同。同期以 Romer (1986) 和 Lucas (1988) 为代表的新增长理论,确切地说是 AK 类型的新增长理论,预言的恰是趋异。

基于此,新古典增长理论的拥护者把趋同进一步细化为 δ 趋同、绝对 β 趋同和条件 β

* 本文是在第三届经济学会年会会议论文《中国城市绝对 β 趋同》的基础上修改完成的,任若恩教授评论了会议论文;在修改过程中得到舒元教授、陈广汉教授和关智生副教授的帮助,本文的研究得到中山大学□川基金博士生重要创新研究项目(2003)、广东省重点研究基地 2003 年度重大项目(03JDXM79002)和广东省教育厅人文社会科学规划项目(02SJA790002)的资助。原载:《经济研究》,2004 年第 5 期。

① Temple (1999) 就趋同文献做了详细的综述。

趋同。 δ 趋同是指，一组经济体人均或劳均实际 GDP 的标准差具有下降的趋势，即 $\delta_{t+T} < \delta_t$ ，其中 δ_t 是 $\ln(y_{it})$ 的标准差， y_{it} 是 t 时期经济体的实际人均或劳均 GDP。绝对 β 趋同是指，穷经济体比富经济体增长得更快，即在回归方程 $g_{i,t,t+T} = \alpha + \beta \ln(y_{it}) + \epsilon_{i,t}$ 中 β 为负，其中 $g_{i,t,t+T}$ 是经济体 i 在 t 到 $(t+T)$ 期的平均增长速度。条件 β 趋同是指，经济体的增长速度与其自身初始状态到其稳定状态的距离大致成反比，简而言之，经济体向自身的稳定状态收敛，即在回归方程 $g_{i,t,t+T} = \alpha + \beta \ln(y_{it}) + \Psi X_{i,t} + \epsilon_{i,t}$ 中 β 为负，其中 $g_{i,t}$ 是刻画经济体 i 稳定状态的一组变量，需要强调的是，这个回归方程就是著名的 Barro 回归方程。在概念细化的基础上，大量的实证分析都揭示了，在世界范围内不存在绝对 β 趋同，但确实存在条件 β 趋同。具有代表性的工作是 Barro 和 Sala-i-Martin (1992) 以及 Mankiw, Romer 和 Weil (1992) (以后简称 MRW 分析框架) 等。

随后趋同研究大致沿着两个方向发展：趋同机制和趋同工具箱的拓展。趋同机制的拓展主要是从关注资本边际报酬递减到关注技术扩散、转移。趋同研究源于资本边际报酬递减还是不变的争论^①，随着新增长理论从 AK 类型发展到 R&D 类型，人们着重强调内生的技术在长期经济增长中的引擎作用。在这种情况下，从技术转移的角度研究趋同，就成为水到渠成的事。目前人们着重探讨技术转移在趋同中的作用，如 Bernard, Jones (1996) 和 Kumar, Russell (2002) 等。

趋同工具箱的拓展主要是从截面分析 (cross-section) 到面板分析 (panel data)。② 研究趋同一般始于 Barro 回归方程和 MRW 分析框架，Islam (1995) 明确指出，二者的估计结果是有偏的，建议采用面板分析。因为面板分析既能保持与新古典增长理论的结论一致（每个经济体的长期增长率相等，为外生的技术进步率），又能解决估计有偏问题，而且 Islam (1995) 发现 MRW 分析框架可以自然推导出面板分析的回归方程。

Dowrick 和 Rogers (2002) 同样采用面板分析，在 MWR 分析框架中引入资本变量，用以直接刻画资本边际报酬递减在趋同中的作用，而把初始劳均 GDP 解释为技术缺口，从而实现了在一个分析框架内同时考察新古典增长的趋同机制（资本边际报酬递减）和新增长理论的趋同机制（技术转移、扩散）。

目前，随着 Quah (1996) 批判不断为人们所接受，从增长分布的角度研究趋同已逐渐成为一个新的发展方向。沿着这条主线，现有研究主要在三个层次上展开：提炼全球增长分布演进的典型事实、探索全球增长分布演进的机制和展望全球未来增长分布的演进 (Quah, 1996; Jones, 1997; Kumar & Russell, 2002; Henderson, et al., 2002)。

纵观现有趋同文献中，人们研究的对象是国家、地区（省区），忽视研究对象内部的情况。Lucas 一直强调城市在经济增长中的作用，但人们还没有从实证的角度分析，城市是否存在不同于国家或省区层面的经济增长特性。因此，本文拟采用 Barro 回归方程以及 Dowrick 和 Rogers 框架分析我国城市经济增长的趋同情况。

① 新古典增长理论的基本假设是资本边际报酬递减，而新增长理论的初始面貌是 AK 类型的，强调资本边际报酬不变。因此，是否存在趋同就成为实证检验新古典增长理论和 AK 类型新增长理论的试金石。

② 已有部分学者采用时间序列方法来研究趋同问题。

三、数据

1. 研究对象

本文研究的对象是我国地级及其以上城市的市区，暂不包县级城市。

我国城市建制分为直辖市、地级市（其中省会城市为副省级）、县级市和镇。从现有的统计数据来看，《中国城市统计年鉴》对地级市分别列出“地区”和“市区”两项，县级城市只有“地区”一项。“地区”包括市区和下辖县、县级市，包含了农村地区的数据，不能真实地反映城市的经济活动；“市区”则仅包括城区和郊区，行政界线相对稳定，体现了城市中的经济活动。

为了避免样本选择有偏问题，我们初步选定了统计年鉴中有相关数据的 216 个样本^①，其中 8 个城市没有提供完整的固定资产投资数据，在必要时又剔除了这 8 个样本。

2. 数据来源

本文采用的数据全部来源于《中国城市统计年鉴》（1989~2000）。其中 GDP 分别按各自省区的 GDP 缩减指数进行平滑^②，GDP 缩减指数来源于《新中国 50 年统计资料汇编》和《中国统计年鉴》（2000）；人口是样本的总人口数；资本存量是采用永续盘存法模拟的，首先根据固定资产投资流量模拟出名义资本存量，然后按各省 GDP 缩减指数进行平滑，最后得到实际资本存量。

四、中国城市经济增长的趋同分析

1. δ 趋同分析

刻画 δ 趋同的常用指标有变异系数和 Theil 指数等。因此，本小节主要采用这两个指标考察中国城市人均 GDP 的趋同情况。考虑到 Theil 指数可以把总体差异分解为组内、组间差异，我们分组考察中国城市经济增长趋同时，则主要采用 Theil 指数。

首先采用变异系数和 Theil 指数对 1990~1999 年中国 216 个地级及以上城市的人均 GDP（对数形式）进行分析，计算结果见图 1。从图形上看，在 20 世纪 90 年代，中国 216 个城市人均 GDP 的 Theil 指数和变异系数均呈现出明显的下降态势，即存在 δ 趋同。



图 1 中国城市经济增长的 δ 趋同

① 具体城市名称见附录。

② 在统计年鉴中，1992 年和 1993 年的数据缺失，利用内插模拟补足。

为了更加细致地刻画我国城市的 δ 趋同，我们把 216 个城市分为东部城市和中西部城市两组，其中东部城市有 92 个、中西部城市有 114 个。采用 Theil 指数把城市间的差距分解为组内、组间差距，结果见图 2。从图形上看，东部城市内的差距在 20 世纪 90 年代初期迅速缩小，进入 90 年代中后期大致稳定下来；中西部城市内的差距，如果我们忽略了 1996 和 1998 这两年，则一直呈现出下降的态势；东部和中西部城市间的差距在 20 世纪 90 年代初期迅速扩大，进入 90 年代中后期大致稳定下来。但需要强调指出的是，与组内差距相比，组间差距下降了一个数量级，这表明东西部城市间的差距相对而言比较小。

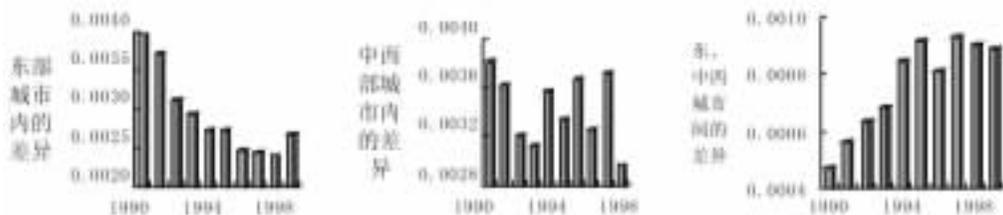


图 2 我国城市经济增长差异的分解：按地理位置

考虑到城市级别可能会影响到城市间的增长差距，进而把全国 216 个城市分为 30 个省会城市（包括直辖市在内）和 186 个非省会城市，同样采用 Theil 指数把城市间的差距分解组内、组间差距，结果见图 3。从图形上看，省会城市内的差距主要是集中在 0.0017 ~ 0.0020 之间，相对而言，增长差距不是很大；从变动态势上，整体而言呈现下降的态势。非省会城市内的增长差距明显呈现下降态势；而省会城市与非省会城市间的差距不大，整整下降了一个数量级，而且稳定在 0.0002 ~ 0.0003 之间。

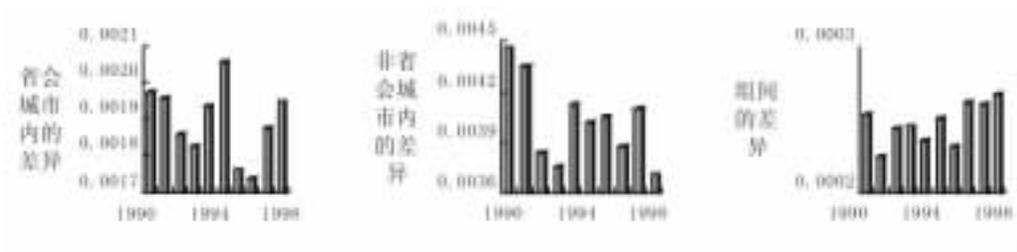


图 3 我国城市经济增长差异的分解：按行政级别

通过以上分析可知，在 20 世纪 90 年代，我国城市间存在 δ 趋同，而且大致不受城市地理位置和级别的影响。

2. 绝对 β 趋同分析

在本小节，我们主要是基于 Barro 回归方程考察我国城市经济增长的绝对 β 趋同情况，计量结果见表 1。

表 1 城市绝对趋同分析

	回归 1	回归 2	回归 3
固定效应	无	无	有
Ln(RY)	- 0.0144*** (- 3.3)	- 0.0314*** (- 7.0)	- 0.0407*** (- 5.2)
DEAST* Ln(RY)		0.0047*** (6.0)	0.0054*** (5.6)
DCAP* Ln(RY)		0.0028*** (2.8)	0.0031*** (3.6)
DOPEN* Ln(RY)		0.0034*** (2.7)	0.0039*** (3.1)
Adj. R ²	0.04	0.253	0.243
F	10.6	18.6	27.6
D- W	1.5	1.8	1.8
N	208	208	416
趋同速度	1.55		
东部城市		2.86	3.89
省会城市		3.09	4.16
沿海城市		3.01	4.07
其他城市		3.77	5.24

注 括号内为统计量,***表示通过显著水平为 1% 的检验,模型 1 2 的样本区间是 1989 ~ 1999,分析方法是截面分析,模型 3 的样本区间是 1989 ~ 1994,1994 ~ 1999,采用的是 Panel 分析;DEAST,DCAP 和 DOPEN 分别是刻画东部、省会和沿海城市的虚拟变量。

表 1 报告了全国 216 个城市绝对 β 趋同的计量分析结果。回归方程 1 是标准的 Barro 回归方程。从回归结果看,初始人均 GDP 的回归系数为负,而且能够通过显著水平为 1% 的统计检验。这表明,在 20 世纪 90 年代,我国城市间存在绝对 β 趋同,即初始人均 GDP 水平越低,增长得越快。我们根据回归系数计算了我国城市间的绝对 β 趋同速度,大致为 2%。^① 这意味着,在我国如果落后城市的人均 GDP 水平只有发达城市的一半,那么落后城市大致需要 35 年的时间来弥补该差距。回归 2 是在估计方程 1 的基础上引入虚拟变量,把全国 216 个城市分为沿海开放城市^②、东部城市、省会城市和其他城市等四种类型,考察每种类型城市的绝对 β 趋同情况。

① 趋同速度为 2%,是跨国或者地区趋同分析中的一个典型特征。

② 沿海开放城市包括 4 个特区和 15 个沿海开放城市(包括威海)。

从回归的估计结果看,引入反映城市类型的虚拟变量,回归方程的拟合度从4%提高到25.3%,而且回归变量都通过显著水平为1%的统计检验。这表明,在1989~1999年间,我国城市间存在绝对 β 趋同,但每类城市的趋同速度并不相同。从城市间绝对趋同速度上看,东部城市、沿海城市、省会城市和其他城市的趋同速度依次递增,分别为2.86%,3.01%,3.09%和3.77%。

由前面的文献综述可知,采用截面分析的结果可能是有偏差的。为了稳健起见,本文进而采用Panel分析,回归3是有固定效应的Panel分析,采用克服异方差的White一致协方差估计。从回归结果看,Panel分析仍然支持本文的结论,即在1998~1999年间,我国城市间存在绝对 β 趋同,每类城市的趋同速度并不相同。在Panel分析中,回归变量都通过显著水平为1%的统计检验,回归系数的符号与方程2的相同,但估计值的大小发生了变化,从而每类城市的收敛速度几乎都提高1个百分点。

需要强调指出的是,即使采用Panel分析,各类城市的速度大小次序保持不变,仍然是东部城市、沿海城市、省会城市和其他城市依次递增,分别为3.89%,4.07%,4.16%和5.24%。这与跨国趋同文献的发现一致,跨国趋同文献大都揭示采用Panel分析,趋同速度一般会提高,即截面分析的结果是有偏的。

以上分析表明,在20世纪90年代,我国城市经济增长存在绝对 β 趋同,而且该结论是比较稳健的。中国城市经济增长在20世纪90年代存在 δ 趋同和绝对 β 趋同,自然又引出一个有趣的问题,即为什么会趋同或者说中国城市经济增长的趋同机制是什么?下面,我们将就这个问题展开讨论。

3. 中国城市趋同机制分析

由前面的文献综述可知,人们所发现的趋同机制主要有两个,而Dowrick和Rogers的分析框架能够同时考察这两个机制在趋同中的作用。因此,我们采用该框架分析中国城市趋同的机制。

Dowrick和Rogers的分析框架是 $g_{i,t} = c + \beta \ln(y_{i,0}) + \alpha(k/k)_{i,t} + \epsilon_{i,t}$,其中 g 、 y 和 k/k 分别表示在平均增长速度、初始劳均GDP和劳均资本存量的增长速度。这个分析框架的经济含义非常直观,即在MWR(1992)的分析框架中引入资本变量,用以直接刻画资本积累在趋同中的作用,而把初始劳均GDP解释为技术缺口,从而实现了在一个分析框架内综合考察资本积累和技术转移在趋同中的作用。具体而言,实证分析中,如果 α 显著,且 $\alpha < 1$,则存在新古典增长理论的趋同机制;如果 β 显著,且 $\beta < 0$,则存在新增长理论的趋同机制。其中,二者所带来的趋同速度分别为 $\lambda_k = (1 - \alpha)(n + g + \delta)$ 和 $\lambda_l = -[\ln(1 + t\beta)]/t$ 。

表2报告两种机制在城市趋同中的作用。回归方程4是在回归方程2的基础上引入资本控制变量,从而初始人均GDP水平就可以用来反映技术扩散在城市趋同中的作用。从估计结果看,回归方程的拟合度明显提高,从25%提高到42%, α 和 β 显著,且 $\alpha < 1$, $\beta < 0$,这表明城市趋同中同时存在新古典增长的趋同机制和新增长理论的趋同机制。

表 2 技术和资本积累在城市趋同中的作用

	回归 4	回归 5	回归 6	回归 7
固定效应	无	无	有	有
Ln(RY)	- 0.0184*** (- 4.2)	- 0.0184*** (- 4.3)	- 0.0218*** (- 4.4)	- 0.0217*** (- 2.8)
DEAST* Ln(RY)	0.00264*** (3.5)	0.00271*** (3.6)	0.0023*** (2.8)	0.0024*** (2.8)
DCAP* Ln(RY)	0.00203** (2.3)		0.0022*** (2.2)	
DOPEN* Ln(RY)	0.00367*** (3.3)	0.00366*** (3.3)	0.00377*** (3.1)	0.0038*** (4.0)
GK0.2097*** (7.7)	0.2075*** (7.6)	0.2358*** (10.4)	0.234*** (8.0)	
DCAP* GK		0.0614** (2.5)		0.066*** (3.1)
Adj. R ²	0.421	0.424	0.329	0.329
F	31.1	31.5	34.9	34.9
D- W	2	2	1.9	1.9
N	208	208	416	416
技术趋同速度				
沿海城市	1.57	1.6	1.89	1.86
东部城市	1.69	1.71	2.05	2.03
省会城市	1.76	2.04	2.07	2.45
其他城市	2.01	2.04	2.31	2.45
资本趋同速度	4.74/2.77		4.59/2.67	
东部城市		4.75/2.77		4.59/2.68
省会城市		4.39/2.45		4.20/2.43
沿海城市		4.75/2.77		4.59/2.68
其他城市		4.75/2.77		4.59/2.68

注：括号内为统计量，***、**和*分别表示通过显著水平为1%，5%和10%，模型4和5是截面回归分析；模型6和7是Panel分析；在资本趋同速度栏中“/右边”是按照 $(n + d + g) = 0.035$ 计算；其他同表1，在固定效应分析中，共同截距为0.177，显著水平1%。

从相应的趋同速度上看，新古典增长理论的趋同机制所产生的趋同速度达到4.77%，

接近于 Dowrick 和 Rogers (2002) 的 5% 的资本趋同速度。就新增长理论的趋同机制所贡献的趋同速度而言,我们发现,沿海开放城市、东部城市、省会城市和其他城市的技术扩散趋同速度依次递增,分别为 1.57%, 1.69%, 1.76% 和 2.01%。这可能反映了我国的现实。作为发展中国家,我国是技术进步的积极接纳者^①,技术进步来自于对先进国家的学习与模仿。在经济全球化的背景下,城市作为人力资本的高度集中地 (Lucas, 1988, 2002),成为技术学习的主体。李郁 (2003) 认为,在我国实施梯度发展战略的过程中,城市存在不同的技术学习模式,沿海城市等通过不断“试错”寻求适合我国国情的技术、制度等,是“干中学”的模式;其他地区城市可以直接学习沿海城市已经探索成功的经验,是“看中学”的模式。因此,反映在技术扩散所贡献的趋同速度上,沿海城市应该低于其他城市。

方程 5 引入城市类型控制变量,进一步考察新古典增长理论的趋同机制在城市中趋同的作用。^②从计量结果看,新增长理论的趋同机制仍然存在,其所贡献的趋同速度的大小略微有点变化,沿海开放城市、东部城市、省会城市和其他城市的趋同速度分别为 1.60%, 1.71%, 2.04% 和 2.04%,仍然大致依次递增。

考虑到截面分析是有偏的,我们同样采用 Panel 数据分析两个机制在我国城市趋同中的作用,即回归方程 6 和 7,二者分别与回归方程 4 和 5 相对应。从估计结果看,即使采用 Panel 分析,新古典增长理论和新增长理论的趋同机制仍然存在,而且就后者所贡献的趋同速度而言,沿海开放城市、东部城市、省会城市和其他城市仍然大致依次递增。

需要强调指出的是,新古典增长理论的趋同机制不容忽视。对比表 1 和表 2,我们清楚地看到,只有在模型中引入资本变量后,沿海开放城市、东部城市、省会城市和其他城市的技术趋同速度才呈现出大致依次递增的态势,从而与我国梯度发展战略相吻合。另外,如果我们仔细观察表 2 中的技术趋同速度栏和资本趋同速度栏,就会发现,即使假定全国 208 个城市稳定状态增长速度非常低,如为文献中通常假定的一半 (3.5%),新古典增长理论的趋同机制所贡献的趋同速度仍然比技术转移所贡献的速度略高些。

五、进一步讨论

在本部分,我们将进一步讨论我国城市在 20 世纪 90 年代的趋同情况,如城市 δ 趋同与绝对 β 趋同的关系、城市趋同与省际趋同的异同,从而进一步凸现 Lucas (1988, 2002) 在理论上所强调的城市具有不同于国家或省区层面的经济增长特性。

1. δ 趋同与绝对 β 趋同

在前面的分析中,我们发现,我国城市经济增长中存在 δ 趋同和绝对 β 趋同,而且该结论是稳健的。一个有意思的问题是,二者之间的关系如何?

Barro 等 (2002) 在理论上证明了, β 趋同是 δ 趋同的必要但不充分条件,即如果存在

① 世界银行报告把我国定为积极的技术接纳国 (世界银行, 2001)。

② 在实证分析过程中,我们尝试采用其他虚拟变量刻画资本边际报酬递减在不同类型城市中的作用,但显著的只有表 2 中的回归方程 5。

δ 趋同, 则一定存在 β 趋同。本文发现, 在 20 世纪 90 年代我国城市经济增长中同时存在 δ 趋同与绝对 β 趋同的现象, 显然与 Barro 等人的理论分析相吻合。

2. 城市趋同与省区趋同

我国城市趋同明显不同于省区间的趋同, 主要表现在以下三点:

第一, 城市间存在 δ 趋同, 而省区间不存在。改革开放后, 我国省区间人均 GDP 的差距呈现出“V”字型变动(参见蔡昉, 2002)。^①

第二, 城市间存在绝对 β 趋同, 而省区间不存在。大量研究我国省区趋同的文献已发现, 改革开放后, 省区间几乎不存在绝对 β 趋同, 但存在条件 β 趋同(Jian, Sachs 和 Warner, 1996; 蔡昉等, 2002; 林毅夫等, 2003)。

第三, 趋同机制上, 二者也可能存在差异。在城市趋同中, 两大趋同机制同时存在; 在研究我国省区趋同的文献中, 如刘强(2001)发现, 我国不存在新古典经济增长理论所强调的趋同机制。

以上分析表明, 城市趋同与省区趋同的结果完全不同, 那么两者之间的关系如何? 能否加深我们对经济增长的理解? 本文以城市为研究对象实际上是对我国经济增长的地理空间进一步分解。经济地理学家认为, 城市是区域经济增长的核心, 区域是城市经济增长的基础, 由于城市的“核心”性质, 城市经济增长往往表现出领先于区域经济增长的现象。

一方面, 城市的绝对 β 趋同与省区的条件 β 趋同并不矛盾, 具有内在的一致性。从方法论的角度看, 省区的条件趋同, 是在对省区经济增长施加一系列控制变量(如制度条件、地理条件、产业结构等)后得出趋同的现象, 即每个省区所面临的为控制变量所反映的条件一致时, 省区间存在 β 趋同。就城市而言, 自 1985 年以来, 我国经济体制改革大部分集中在城市, 沿海城市、省会城市、大部分边境城市都是开放城市, 而且以开放经济为代表的各层次经济开发区遍布不同等级的城市中, 因此, 各类城市经济增长的制度条件十分相似。从地理条件看, 地级及其以上城市都是历史上或现代交通的枢纽, 集中于冲击平原或盆地。从生产要素配置的角度看, 各级城市是其所在地区的资源配置平台, 往往优先考虑城市自身的发展(李郁, 2003)。总之, 城市间的经济增长比省区间的经济增长表现出较强的条件一致性, 从而表现为绝对(无条件) β 趋同。但需要特别指出的是, 城市间增长条件的相对一致性, 不是我们在估计方程中人为控制的, 而是在增长过程中城市本身所具有的。

另一方面, 我国城市经济增长 δ 趋同, 进一步指出了我国省区增长差距的空间分布。进入 20 世纪 90 年代, 我国省区的增长差距确实不断拉大, 现有趋同文献主要把其归因于沿海与内地省区间的增长差距不断拉大。而本文本把我国经济增长的地理空间进一步分解为地级及以上城市(市区)和其他地区, 结果发现, 我国城市间的人均 GDP 差距越来越小, 而且沿海、内地城市间的人均 GDP 差距并不大。这表明, 我国省区间人均 GDP 差距的拉大, 可能主要归因于本文所考察的样本以外, 即归因于其他地区内部差距或者其他地

^① 更加细致的研究发现, 我国省区间存在双峰趋同(徐现祥、舒元, 2004), 甚至三俱乐部趋同(沈坤荣等, 2002)。本文没有明确考察我国城市俱乐部趋同的情况, 但发现, 按地理位置和行政级别对我国城市分组后, 组间差异整整下降了一个数量级。这从一个侧面反映了我国城市间可能不存在俱乐部趋同。

区与本文所考察样本间的差距不断拉大。

六、结论性评述

Lucas (1988, 2002) 在理论上一直强调城市具有不同于国家或省区层面的经济增长特性, 但为现有的趋同文献所忽视。本文采用趋同分析的标准方法, 就我国 216 个地级以及城市经济增长的趋同情况展开讨论, 结果发现:

与现有省区间的趋同模式不同, 我国城市经济增长存在 δ 趋同和绝对 β 趋同, 而且该发现是稳健的。

从趋同机制上看, 在我国城市层面上同时存在新古典增长理论所强调的趋同机制(资本边际报酬递减)和新增长理论所强调的趋同机制(技术转移、扩散)。

另外, 本文的发现有助于人们对我国省区差距在 20 世纪 90 年代重新拉大的进一步理解。目前, 相关文献把省区差距进入 20 世纪 90 年代重新拉大归因于沿海与内地省区间的差距不断拉大, 而本文发现, 沿海、内地城市间的差距不大, 也没有呈现出不断拉大的态势。这意味着把我国省区差距的重新扩大归因于沿海与内地省区间的差距不断拉大, 可能有失偏颇。

由于数据的原因, 本文并没有涉及县级城市和农村的经济增长, 但从本文所发现的地级及其以上城市所具有的不同于国家或省区层面的经济增长特性来看, 这显然值得我们进一步研究, 也许能够发现我国省区差距重新拉大的真正源泉。

附录: 本文所采用的 216 个样本分别是北京; 天津; 石家庄, 唐山, 秦皇岛, 邯郸, 邢台, 保定, 张家口, 承德, 沧州, 廊坊, 衡水; 太原, 大同, 阳泉, 长治, 晋城, 朔州; 呼和浩特, 包头, 乌海, 赤峰, 通辽; 沈阳, 大连, 鞍山, 抚顺, 本溪, 丹东, 锦州, 营口, 阜新, 盘锦, 铁岭, 朝阳; 长春, 吉林, 四平, 辽源, 通化, 白城; 哈尔滨, 齐齐哈尔, 鸡西, 鹤岗, 双鸭山, 大庆, 伊春, 佳木斯, 七台河, 牡丹江, 黑河; 上海; 南京, 无锡, 徐州, 常州, 苏州, 南通, 连云港, 淮阴, 盐城, 扬州, 镇江, 泰州, 宿迁; 杭州, 宁波, 温州, 嘉兴, 湖州, 绍兴, 金华, 衢州, 舟山; 合肥, 芜湖, 蚌埠, 淮南, 马鞍山, 淮北, 铜陵, 安庆, 黄山, 阜阳, 宿州, 滁州; 福州, 厦门, 莆田, 三明, 泉州, 漳州, 南平, 龙岩; 南昌, 景德镇, 萍乡, 九江, 新余, 鹰潭, 赣州; 济南, 青岛, 淄博, 枣庄, 东营, 烟台, 潍坊, 济宁, 泰安, 德州, 威海, 聊城, 临沂, 莱芜, 日照; 郑州, 开封, 洛阳, 平顶山, 安阳, 鹤壁, 新乡, 焦作, 濮阳, 许昌, 漯河, 三门峡, 商丘, 南阳, 信阳; 武汉, 黄石, 十堰, 宜昌, 襄樊, 鄂州, 孝感; 长沙, 株洲, 湘潭, 衡阳, 邵阳, 岳阳, 益阳, 常德, 郴州, 永州, 怀化; 广州, 韶关, 深圳, 珠海, 汕头, 佛山, 江门, 湛江, 茂名, 惠州, 肇庆, 潮州, 梅州, 中山, 东莞, 汕尾, 河源, 阳江, 清远; 南宁, 柳州, 桂林, 梧州, 北海, 钦州, 贵港; 海口, 三亚; 重庆; 成都, 自贡, 攀枝花, 泸州, 德阳, 绵阳, 广元, 遂宁, 内江, 乐山, 宜宾, 南充; 贵阳, 六盘水, 遵义; 昆明, 玉溪, 曲靖; 西安, 铜川, 宝鸡, 咸阳, 延安, 汉中, 渭南; 兰州, 嘉峪关, 金昌, 白银, 天水; 西宁; 银川, 石嘴山, 吴忠; 乌鲁木齐, 克拉玛依。

参考文献

[1] 蔡昉:《制度、趋同与人文发展》, 中国人民大学出版社 2002 年版。

- [2] 李郁：《中国城市中的经济增长》，中山大学博士论文，2003年。
- [3] 林毅夫，刘培林：《中国的经济发展战略与地区收入差距》，《经济研究》，2003年第3期。
- [4] 刘强：《中国经济增长的收敛性分析》，《经济研究》，2001年第6期。
- [5] 沈坤荣，马俊：《中国经济增长的“俱乐部收敛”特征及其成因研究》，《经济研究》，2002年第1期。
- [6] 世界银行：《2001人类发展报告》，中国财经出版社2001年版。
- [7] 徐现祥，舒元：《中国省区增长分布的演进：1978~1998》，《经济学季刊》，2004年。
- [8] Abramovitz, M., Catching Up, Forging Ahead, and Falling Behind, *Journal of Economic History*, 1986, Vol.46, pp 385~406.
- [9] Barro, R. and X. Sala-i-Martin, Convergence, *Journal of Political Economy*, 1992, Vol.100, pp 223~251.
- [10] Barro, R. and X. Sala-i-Martin, *Economic Growth* (Second edition), McGraw-Hill, Inc., 2002.
- [11] Baumol, W., Productivity Growth, Convergence, and Welfare: What the Long-Run Data Show, *American Economic Review*, 1986, Vol.76, pp 1072~1085.
- [12] Bernard, A. and C. Jones, Technology and Convergence, *The Economic Journal*, 1996, Vol.106, pp1037~1044.
- [13] Dowrick, S. and M. Rogers, Classical and Technological Convergence: Beyond the Solow-Swan Growth Model, *Oxford Economic Papers*, 2002, Vol.54, pp369~385.
- [14] Islam, N., Growth Empirics: A Panel Data Approach, *Quarterly Journal of Economics*, 1995, Vol.110, pp1127~1170.
- [15] Jian Luntian, J. Sachs and A. Warner, Trends in Regional Inequality in China, *China Economic Review*, 1996, Vol.7, No.1, pp1~21.
- [16] Jones, C., On the Evolution of the World Income Distribution, *Journal of Economic Perspectives*, 1997, Vol.11, No.3, pp19~36.
- [17] Kumar, S., and R. Russell, Technological change, Technological Catch-up and Capital Deepening: Relative Contributions to Growth and Convergence, *The American Economic Review*, 2002, Vol.92, pp527~548.
- [18] Lucas, R., On the Mechanics of Economics Development, *Journal of Monetary Economics*, 1988, Vol.22, pp3~42.
- [19] Lucas, R., Life Earnings and Rural-Urban Migration working paper, <http://home.uchicago.edu/~sogrodow/homepage/life-earnings.pdf>.
- [20] Mankiw, N. G., D. Romer and D. Weil, A Contribution to the Empirics of Economic Growth, *Quarterly Journal of Economics*, 1992, Vol.107, pp407~437.
- [21] Quah, D., Twin Peaks: Growth and Convergence in Models of Distribution Dynamics, *The Economic Journal*, 1996, Vol.106, pp1045~1055.
- [22] Romer J., Increasing Return and Long-Run Growth, *Journal of Political Economy*, 1986, Vol.94, pp1002~1037.
- [23] Temple, J., The New Growth Evidence, *Journal of Economic Literature*, 1999, Vol. XXXVII, pp112~156.

市场规制和企业制度不完善条件下的经济增长机制探讨*

徐 勇 孙洛平

一、中小城镇的经济高速增长现象

我国改革开放以来的经济持续高速增长早就引起世人的关注。美国经济学家弗里德曼曾经说过，中国有哪位经济学家能够把这场改革的整个过程描述下来，为什么能走向成功，怎么走向成功的……，中国学者有可能就此得一个诺贝尔奖。杨小凯、林毅夫等许多经济学者也多次提到这个被称为“中国经济奇迹”的问题。

世人之所以会关注所谓的“中国经济奇迹”问题，其原因在于在主流经济学的框架内，难以对我国改革开放以来经济的持续高速增长给予合理的解释。一国的经济增长需要具备的条件不外乎丰富的自然资源、各种层次的人力资源、完善的市场规制、高效率的政治体制和人与人之间的较高的社会信任度等五个方面的优势。从这五个方面的条件比较，大概只有劳动力成本低廉可以算作是我国的一个相对优势，还不要说我们的高级人才严重缺乏的劣势。不过，即便是低劳动成本优势也不是绝对的，因为有许多国家的劳动力成本比我们还要便宜。然而，正是在这样的不利条件下，我国却出现了长期持续的经济高速增长现象，这不能不说是一个谜。

不过，在不具备相应优势的条件下出现经济快速增长的现象在我国国内早就存在，只是人们对此熟视无睹而已。从我国地区经济发展的不平衡中，我们也可以看到所谓的“中国经济奇迹”的影子。以浙江经济为例。浙江地处江南丘陵地带，在自然资源、工业基础、外资优惠政策方面，相对于沿海其他地区没有什么优势。然而，浙江这些年成了全国的经济明星，在经济和外贸增长速度方面更是表现出色。令人特别感兴趣的是这些奇迹主要都是由若干中小企业群来实现的。按照经济学者们的见解，浙江的经济活力主要应该归结为所有制的优势，即经济民营化。

具体地说，浙江经济的发展主要应该归功于地方县市甚至乡镇的小企业群集生产模式。若干以中小企业甚至家庭为基础的生产单位聚集在一起，形成相互之间高度分工专业化的集群，使人力成本低廉的优势得以充分发挥。有的学者认为这样做是为了克服大资本缺乏的缘故，其实不然，因为这种观点暗示如果有大资本来组织同样的生产将会有更好的效果。考察一下长江三角洲和珠江三角洲区域经济的发展，许多中小城镇都是以产业集群的方式专业于生产一类产品的生产，由此形成高度的分工专业化优势，并以此来与中心城

* 原载：《管理世界》，2004年第3期。

市进行竞争的。在中国，中小城镇的产业群集具有较高的经济效率应该是一个较为普遍的现象。

二、中小城镇的竞争优势从何而来

组织一种产品的生产，依靠什么来形成竞争的优势，尤其是当其他的竞争者也可以生产同样的产品时，靠什么形成比其他竞争者更强的相对竞争优势？人们对这个问题的回答通常是从是否有良好的公共设施、宽松的融资条件、高素质的人才、完善的市场规制、高效率的企业制度等方面寻求解释。如果我们用这些条件来对比长三角和珠三角地区大量的小城镇经济发展的现状，那么你未必会坚持以上说法。同样的事实是，改革开放以来，国有企业的经营条件一般要比中小城镇的民营企业优越（除了在改革开放的前期国有企业历史包袱较重这个例外），可是国有企业的发展速度却远比一般的民营企业来得慢。是什么机制使得经济增长的效果与其条件之间有如此大的背离？我们对此提出一些想法。

任何竞争优势的形成都是利用一定的优势，这一点是没有疑问的。不过，如何看待经济增长的优势，却不能简单化。简化我们的分析，不考虑自然资源、政治体制、社会信任度方面的因素，因为它们对于不同所有制的企业来说都差不多。如果一个地区具有良好的市场竞争规制，比方说经济法规比较完善、经济纠纷的诉讼成本较低、政府积极打击制假售假行为等，那么这个地区的企业当然是充分利用市场机制来组织生产，以节约经营的成本，并形成相对于其他地区的竞争优势。同样，一个地区要是有良好的企业制度传统，比方说具有较多的职业经理人才，工人们也有在现代大企业中工作的经验，那么这个地区也可以利用这个优势，即利用企业形式来组织生产。即便市场法规方面有所欠缺，以企业替代市场，也可以形成相对的竞争优势。

可是考察一下长三角和珠三角地区大量小城镇的企业所面临的环境，它们既没有市场规制方面的优势，又没有企业制度的人才方面的优势，但是只要是这些乡镇企业能够生产的产品，其产品的竞争力一点也不亚于那些大中城市的企业。我们想说明的观点是，在一定的历史条件下，经济的活力正是来自于那些以传统的观点看起来一点优势也没有的地方。限于篇幅，本文不能对此给出正式的论证，下面仅用一种高度简化的方式说明，要是你既没有市场规制的优势，又没有企业制度的优势，那么反而可能通过一种新的方式形成甚至是更高的竞争力，至少我国的经济实践证实了这一点，这就是沿海地区的中小城镇的产业群集生产模式。

三、中国特色的产业群集竞争优势

产业群集是指特定产业以及支撑和关联产业在一定地域范围内的地理集中。产业群集不是什么新机制，而是一种世界范围内的较为普遍的经济现象。早在工业化发展时期，它就成为一个产业发展的有效模式。随着发达国家市场制度的日益完善，许多重要产业的发展开始突破地理区域的限制，以现代大公司和跨国公司的方式发展。不过，许多国家始终有一些传统产业保持着地理区域相对集中的发展模式，如美国加州的葡萄酒业、意大利的

皮具时装业、我国的瓷器产业等。在当代，现代信息产业在硅谷的迅猛发展、领导和左右世界信息产业发展方向的现象，重新引起人们对产业集群现象的关注。

为什么特定产业的群集会形成较高的竞争力，经济学者从不同的角度出发往往给出不同的解释。早在英国工业革命时期，产业集群就成为一种有竞争力的工业模式。经济学家马歇尔对此有专门的论述。他认为，工业在一个地方聚集的利益主要来自三个方面：知识和生产技能外溢效应、高效率专业机器的采用、熟练劳动力市场的集中。当代较为普遍的见解是，产业集群的竞争优势来自于以下四个方面：①接近生产的投入和相应的基础设施；②熟练劳动力和人力资源市场集中；③容易获得相关的信息和对生产业绩进行量度；④形成互补的产品链（Scorsone, 2002）。尽管具体的产业以及相应的产业集群模式各有不同，但上述优势一般都是存在的。不过，在上述优势之外，可以发现产业集群模式的一个重要特点是，它不以具备完善的市场法规和先进的企业制度为先决条件。这也是为什么在不具备当代市场经济条件的工业革命时期，就存在产业集群这种经济现象的原因。

我国的许多产业集群模式更有其特点。在长江三角洲和珠江三角洲的一些经济相对发达的地方，中小城镇（包括乡镇）既没有大城市在筹集资金上的优势，也没有企业制度和市场规制上的优势，它们只能集中力量，以产业集群的方式生产一种或少数几种产品。在产业集群区内部，以家庭或小企业为独立生产单位，每一个生产单位在整个产品的产业链中只能占据由高度分工形成的一个或少数几个环节，而集群区作为一个整体则由众多的生产单位构成高度专业化的分工网络体系。在这些中小城镇产业集群区内部，生产单位之间相互关系的一个特点是，作为买者，总是同时与多个卖者进行交易，有时不惜购买有差异的同类产品，并不断调整与各个卖者交易的数量。作为卖者，也总是同时与多个买者进行交易，甚至为不同的买者提供有差异的产品。这种交易关系既不像市场的一次性交易关系，也不像企业内部的固定交易关系，而是一种半固定的关系。尤其在人际关系方面，不仅在作为交易伙伴的买方与卖方之间保持着良好的私人关系，而且在处于同一产业链地位的作为竞争关系的生产者之间也保持着良好的信息交流渠道。用一句通俗的话来说，这种产业集群模式是一个建立在“熟人”关系基础上的复杂的“关系网结构”。

从经济学的角度来看，这种模式形成两个效应，不妨称之为“群集锁定效应”和“专用性投资通用化效应”，它们恰好能够克服市场不规范和企业制度不完善的弱点。在集群区内部，每个生产者由于高度专业化加上直接面对市场竞争，具有很强的竞争意识和市场适应能力。可是一旦离开所在的产业集群区，由于掌握的技能过于专门化而缺乏相应的通用性技能，所以无法生存。同时，建立在乡里乡亲关系网基础上的分工体系还是一个高效率的信息网，任何欺诈行为一旦被发觉，会很快在交易伙伴中散布开（主要通过同等产业链地位的竞争者之间的信息沟通机制），从而大幅提高欺诈成本。因此，即便没有一套完善的契约签订、执行、监督以及违约处理等市场法规体系，也能有效地克服人们的机会主义行为。另外，这样的产业集群也没有大企业带来的组织成本，这也是为什么以产业集群方式生产出来的产品总是特别便宜的原因。

产业集群对于其中的生产者产生的另一个作用是专用性投资通用化效应，它能够有效地克服由于没有良好的企业制度而产生的难以保护专用性投资者利益的缺陷。在产业集群区外部看来是专用性的生产技能，在产业集群区内部则不然。一般情况下，（纵向）分工

每深化一步，都会形成单个买方对单个卖方的双边垄断格局，机会主义带来的交易费用在所难免。可是当分工通过多个买方对多个卖方的方式实现时，专用性技能可以以通用性技能的形式表现出来，因为针对某个交易对象进行的专用性投资可以有较大的部分转用于同其他交易对象进行的合作。在产业集群区外看来只能通过企业内部化方式组织的专用性技术和人力的投资，在产业集群区内部则成为或大部分成为通用性的技术和投资。因此，建立在“人际关系网”基础上的群集生产模式，产品和工艺的创新激励并不弱于大中城市的建立在较完善的市场规制和企业制度基础上的创新激励。

分工是一切生产力提高的最根本的原因，如何提高分工的水平，这要以生产者所处的环境条件而定。如果市场规制比较完善，可以较好地利用契约来约定分工参与各方的利益，人们倾向于利用市场机制来组织分工。如果企业制度比较完善，在企业内部利用权威的方式组织分工比较有效。不过，这两种分工组织方式都要付出费用。利用市场机制组织分工，随之而来的是讨价还价、契约制定、执行监督以及违约仲裁等费用；利用企业来组织分工，随之而来的是委托代理、信息扭曲、企业官僚等费用。长三角和珠三角地区中小城镇的产业集群的整体竞争力来自于更高的分工专业化水平，同时却避免了高额的市场交易费用和大企业组织费用。当大中城市的市场规制和企业制度条件相对于成熟的市场经济来说有欠缺时，只会更加充分地体现出这些中小城镇的经济能力。产业集群生产模式如果有对手的话，那就只会是另一个产业集群。

四、低廉劳动成本优势的发挥

我国的劳动力成本低廉是一个公认的优势，其实这个优势的发挥是有条件的。在当代，空有力气而没有专业技能的劳动者是没有什么优势可言的。人们一般以为技术性和专业性人才是一个外生的变量，只能通过人才流动来获得，这个看法是片面的。如果是这样的话，中小城镇的经济发展将难以为继。人力资源在相当大的程度上是一个内生变量。在知识经济条件下，培养一个有创新能力的人才有时需要近 20 年的教育时间，考虑到技术创新经验的获得和积累，甚至要耗费更长的时间。在这种情况下，专业化人才的缺乏将会成为经济增长的重要制约因素。如何让人们尽快地成为专业人才，是一个地区形成整体竞争力的关键。

我国沿海地区中小城镇的产业集群生产模式，就是一个让人们在很短的时间里成为专业人才的机制。一个人的精力是有限的，如果能将有限的精力用于相对狭小的领域，其效能会大幅度地提高。比方说，传统的裁缝要学会从设计到剪裁，再到缝纫成衣的所有工序，因此，当一个好的裁缝是很不容易的。然而，正是因为每个传统裁缝都太有才能，才导致所有裁缝作为一个整体的低效率。如果能够进行分工，一个人专门设计，一个人专门剪裁，一个人专门缝纫，那么每个人要学习的东西都少了，可用于工作并积累经验的时间多了，合在一起，裁缝整体的生产力则可以提高。如果能将制衣的分工专业化水平进一步提高，童装、女装、男装等分别由不同的人来设计，再将剪裁和缝纫也分割成若干工序分别由不同的人来操作，那么不难想象生产力一定会有更大幅度的提高。在这样的生产环境中，连文盲都有可能成为高度专业化的人才。

在我国中小城镇的产业群集区里，企业间的分工不断深化，不仅可以提高技术创新的水平，还能够有效地提高本地区人力资源使用的效率。从一定意义上说，分工专业化可以创造出人才。每个人都有发展的巨大潜能，之所以平凡的人多，杰出的人少，就是因为要关注和学习的东西太多。比方说，当一个人可以集中精力于专门的服装设计，甚至是仅仅专注于某一类服装的设计，那么一个普通的人也有可能成为服装设计大师。当一个人集其毕生的精力钻研一种或少数几种产品的生产和创新，他就会是一个人才。

产业群集区内的高度专业化的分工以及相应人才的形成，又反过来将这些人才锁定于群集区的内部，因为多数人才的专业面高度集中，离开了所在的产业群集区将不再是一个人才。这个群集锁定效应又会进一步约束人们的机会主义行为，从而进入一个竞争力不断提高的良性循环过程。

市场价格机制和现代企业制度的一个共同特点是，其运行是建立在非人格化力量的基础上的，即人们通常所说的依靠法律和制度来约束人们的行为。一般认为，随着我国市场经济体制的不断完善，经济行为中的人情关系色彩会越来越淡薄，这一趋势确实已经在许多大中城市的经济发展中显现出来。不过，中国改革开放以来的多种经济成分和多种发展模式共存的经济实践也向世人表明，不要轻言一种模式一定会比另一种模式优越，也不要轻言东方人注重人格化力量的传统难以形成现代的生产力。中国经济的崛起迟早会向世人昭示，东方的传统也是有竞争力的。

参考文献

- [1] Scorsone, Eric A., *Industrial clusters: Enhancing rural economies through business linkages*, published by Southern Rural Development Center, <http://ext.msstate.edu/srdoc/>, 2002.
- [2] R. H. 科斯：《企业的性质》，见：《企业、市场与法律》，上海三联书店 1990 年版。
- [3] O. 威廉姆森：《交易费用经济学：契约关系的规制》，见：《企业制度与市场组织——交易费用经济学文选》，上海三联书店，上海人民出版社 1996 年版。
- [4] M. 波特：《群集区和新竞争经济学》，见：《战略——45 位战略家谈如何建立核心竞争力》，中国发展出版社 2002 年版。
- [5] 邱成利：《制度创新与产业集群的关系研究》，《中国软科学》，2001 年第 9 期。
- [6] 张元智：《基于企业核心竞争力的产业集群观点》，《人文杂志》，2001 年第 4 期。
- [7] 李小建，李二玲：《产业集群发生机制的比较研究》，《中州学刊》，2002 年第 4 期。
- [8] 骆静，聂鸣：《发展中国家集群比较分析及启示》，《世界经济研究》，2002 年第 3 期。
- [9] 魏守华，王缉慈，赵雅沁：《产业集群：新型区域经济发展理论》，《经济经纬》，2002 年第 2 期。

寡头竞争情形下的国企改革^{*}

——论国有股份比重的最优选择

孙群燕 李 杰 张安民

一、引言

多年来，国有企业改革一直是中国经济体制改革的中心环节。由于国企承担着调控经济、提供就业、保障社会稳定等多重任务，使其出现了经济效率低下、冗员过多、资产负债率过高等问题。如何进行改革，提高生产效率，将国企从沉重的社会性负担中解脱出来，一直受到政府与众多经济学家的普遍关注。从20世纪80年代至90年代中期，中央政府通过出台一系列措施，包括中央与地方的财政分权、促进企业内部改善经营与激励机制、引入市场竞争以及转换国有企业的经济角色等，力图增强国企的经济活力。20世纪90年代中期以后，中央政府又通过鼓励国有企业间的横向兼并、强化国有银行的贷款约束、裁减富余人员、将某些国企重组为有限责任公司或股份制企业等方式，对国企进行更深层次的改革（Zhang等，2001）。与此同时，对于国企私有化的程度与速度，经济学界也从多个角度进行了研究和探讨，但迄今仍未形成比较统一的看法。

Zhang（1998）从公司治理结构的角度出发，论述了为什么中国的国有企业改革只能解决管理层的短期激励而不能解决长期激励问题。关于整个国民经济中国有企业所占比重的问题，Bai等（2000）给出了新的认识角度：对于经济转轨国家而言，由于完备的社会保障体系尚未建立，所以政府在经济中保留一些国有企业是最优的。平新乔（2000）讨论了企业生产成本函数形状以及生产目标函数的变化对于国有经济或国有企业在产业乃至整个经济中的比重的影响，论证了在一定条件下国有经济在国民经济中的比重不会趋于零。王红领等（2001）从实证的角度出发，说明对国有企业的产权改革可以采用一个将政府目标函数考虑在内的次优民营化或破产清算方案。林毅夫（2001）从企业自生能力的角度入手，说明只有通过消除国企的社会性政策负担和战略性政策负担，增强其自生能力，才有可能取得国企改革的成功。刘小玄（2003）从实证的角度检验了中国转轨经济中产权结构和市场结构对于产业绩效的影响作用，发现国有产权结构变量对于产业绩效具有明显的负效应，产业集中度和规模变量则具有正效应。陆铭（2003）通过构建一个合作博弈模型，为

^{*} 本文得到国家自然科学基金的资助（批准号：10131030）和中山大学□川基金博士生创新项目提供的资助，特此致谢。中山大学岭南学院王则柯教授与王美今教授对本文的初稿提出了宝贵的意见，在此一并致谢。原载：《经济研究》，2004年第1期。

不同所有制企业间劳动生产率差距拉大的现象提供了一个理论解释，其基本观点是，放松政府管制对于成功的市场化改革和企业重构而言是至关重要的。

然而，上述文献都没有从国企内部国有股份所占比重的角度来考虑国企改革，而这恰恰是当前亟需解决的难题。近几年来，政府一直尝试通过减持和转让国有股的方式深化国企改革，而这实质上就是国有股份在国企中的比重问题。目前，关于减少国有股比重的方式和方法，社会各界仍存在很大的争议。另外，由于中国已经成为世界贸易组织的正式成员，国企所面临的竞争格局在今后几年内将出现重大变化，尤其是贸易自由化程度的加深必将对国企改革产生重大影响。在这一背景下讨论国有股份的比重问题，具有非常重要的现实意义。

一般而言，转轨经济国家一般具有两个重要特点：①原有计划经济体制下已经形成的大量社会性负担，如国企富余劳动力的安置问题。②国有企业与私有企业在混合寡头的市场结构下展开竞争。本文抓住这两个重要特点，借鉴 Fraja 和 Delbono 的方法 (Fraja and Delbono, 1989)，就一般情形构建一个两阶段的混合寡头垄断竞争博弈模型：在第一阶段，政府选择国有股份的适当比重，从而最大化社会总体福利或自身的支付；在第二阶段，国有企业在给定企业内部国有股份比重的条件下，与私有企业（和外国企业）展开古诺竞争，以最大化自身的经营目标。

具体而言，我们分别就封闭经济与开放经济的情形，从国企中国有股份比重的角度入手，通过在政府政策目标函数中加入对国有企业总收益的考虑，探讨国企在背负一定的社会性负担的条件下，国有股份比重的变化如何影响社会总体福利水平和政府的支付 (payoff)。我们的分析表明，无论是从社会总体福利的角度还是从政府政策目标的角度，最优的国有股份比重的选择都取决于各类企业的生产成本、它们产出之间的边际替代效应以及整个社会的就业压力等因素之间的相互关系。在满足某些标准条件的情形下，完全的国有化或完全的私有化都不能最大化社会总体福利或政府支付。另外，关于关税变化对各类企业均衡产出的影响，我们还得出了一个出人意料的结论：关税的降低并不一定会减少国企的均衡产出。

在模型结果的模拟部分，我们通过设置具体的需求与成本函数，就封闭经济的情形分别给出了最大化社会总体福利目标以及政府政策目标的最优的国有股份比重的具体值及其与整个社会的就业压力以及国有企业的相对生产成本三者之间的相互关系。在给定就业压力的条件下，无论从社会总体福利还是从政府政策目标的角度考虑，最优的国有股份比重与国企的生产效率呈同向变动的关系；在给定国企的生产效率的条件下，从政府政策目标的角度看，最优的国有股份比重与就业压力呈同向变动。值得注意的是，从社会总体福利的角度来看，存在国企的相对生产成本的一个临界指标，当国企的相对生产成本低于这一临界指标时，最优的国有股份比重与就业压力同向变动，当国企的相对生产成本高于这一临界指标时，最优的国有股份比重与就业压力呈反向变动。

本文其余部分的结构如下：在第二部分，我们在封闭经济的情形下构建一个混合寡头竞争的模型框架并给出基本结论；第三部分我们通过设置具体的线性需求函数以及二次成本函数对第三部分的结果进行模拟分析；第四部分通过引入外国企业的竞争因素，对模型进行扩展；第五部分为结语。

二、基本模型

假设市场上有两家企业：国有企业^①和私有企业，它们生产相同的产品，产品间具有完全的替代性。设国有企业的产出为 Q_1 ，私有企业的产出为 Q_2 ，它们的生产成本分别为各自产出的函数，表示为 $C_1(Q_1)$ 、 $C_2(Q_2)$ ，市场反需求函数为 $P(Q)$ ，其中 $Q = Q_1 + Q_2$ ，并且假定消费者具有稳定不变的偏好。我们分别用企业 1 和企业 2 表示国有企业和私有企业，用 π_1 、 π_2 分别表示它们的利润，并用 CS 表示消费者剩余，则有：

$$\pi_1 = PQ_1 - C_1, \pi_2 = PQ_2 - C_2, CS = \int_0^{Q_1+Q_2} P(q)dq - P(Q_1 + Q_2)$$

与通常的讨论相类似，我们用社会福利函数衡量整个社会的总体福利，并将其定义为生产者剩余与消费者剩余之和，即：

$$W = \pi_1 + \pi_2 + CS = \int_0^{Q_1+Q_2} P(q)dq - C_1(Q_1) - C_2(Q_2) \quad (1)$$

我们假定政府可以通过增加就业达到维持社会稳定的目的，在我们的模型中具体表现为在政府的目标函数中加大对国有企业总收益（total revenue） PQ_1 的考虑，而这种考虑的偏重程度又取决于整个社会的就业压力，我们用参数 β 表示， β 越大，表明社会的就业压力程度越高。另外，政府还必须考虑整个社会的福利水平，于是政府的政策目标函数可表示为：

$$G = W + \beta PQ_1 = \int_0^{Q_1+Q_2} P(q)dq - C_1(Q_1) - C_2(Q_2) + \beta PQ_1 \quad (2)$$

值得注意的是，我们所说的就业压力是一个广义的概念，它包括诸如适龄劳动人口的就业、失业保险以及退休养老等因素所造成的社会压力。

另外，正如上文所交代的，国有企业是一个混合产权的股份制企业，其股本由两部分构成：私人股份与国有股份。我们假设国有股份所占比重为 α ($0 \leq \alpha \leq 1$)， $\alpha = 0$ 表示完全的私有控股，而 $\alpha = 1$ 则表示完全的国有控股。国有企业按照国有股份比重的大小决定其目标函数中对政府支付的兼顾程度，同时作为一个市场经济的主体，它还必须考虑自身的盈利。于是国有企业的目标函数也由两部分构成：企业自身的利润以及政府的政策目标，两者在目标函数中的权重分别为 $1 - \alpha$ 和 α 。我们把国企所兼顾的政府支付定义为社会性负担，即 αG 。国有企业的目标就是最大化企业自身的利润以及政府支付的加权平均和^②：

$$\max_{Q_1} U_1 = \max_{Q_1} (1 - \alpha)(PQ_1 - C_1) + \alpha G$$

① 这里所说的国有企业比通常意义上的国有企业的定义范畴更大，它是指存在一定比例（不一定是全部）的国有股份的企业，更确切的说法可能是混合产权企业（mixed-property firm）。

② 在国企的目标函数中， αG 所体现的是一种社会性负担，因此，只要 $\alpha \neq 0$ ，国企的生产决策就会偏离利润最大化点，如果这种负担足够大的话，可能导致国有企业的亏损。这种亏损也就相当于林毅夫（2001）中所指出的政策性亏损。政府必须为这种政策性亏损负起责任，给予国企各种政策性优惠、补贴和保护。但由于我们是在一个局部均衡的分析框架内进行讨论，这种补贴的来源已经超出了我们模型框架所讨论的范围。

$$\begin{aligned}
 &= \max_{Q_1} (1 - \alpha)(PQ_1 - C_1) + \alpha \left[\int_0^{Q_1+Q_2} Pdq - C_1(Q_1) - C_2(Q_2) + \beta PQ_1 \right] \\
 &= \max_{Q_1} (1 - \alpha + \alpha\beta)PQ_1 - C_1 - \alpha C_2 + \alpha \int_0^{Q_1+Q_2} Pdq \quad (3)
 \end{aligned}$$

与通常的讨论相类似，我们把私有企业看作一个单纯追求利润最大化的经济主体，因而其目标函数可表示为：

$$\max_{Q_2} U_2 = PQ_2 - C_2 \quad (4)$$

需要指出的是，在本文的讨论中，我们不考虑国有股份比重变化所产生的效率提高效应^①，即假定国有股份比重的变化不会引起企业成本函数的变化，我们仅仅分析国有股份比例的变化与社会总体福利及政府支付之间的内在联系。

另外，为讨论的方便，我们还需要借助以下基本假设：

假设 1: $\forall Q \geq 0, P > 0, P(Q)$ 是二次可微的，且 $P'(Q) < 0$ ，其中 $Q = Q_1 + Q_2$ 。

假设 2: $\forall Q_i \geq 0, C_i(Q_i)$ 是二次可微的，且 $C_i'(Q_i) > 0, C_i''(Q_i) \geq 0$ 其中 $i = 1, 2$ 。

我们假设国有企业与私有企业所进行的是完全信息条件下的古诺 (Cournot) 竞争，各自的均衡产出分别满足以下一阶条件：

$$\begin{aligned}
 P'(1 - \alpha + \alpha\beta)Q_1 + (1 + \alpha\beta)P - C_1' &= 0 \\
 P'Q_2 + P - C_2' &= 0 \quad (5)
 \end{aligned}$$

(6)

由方程组(5)、(6)可以决定国有企业与私有企业的均衡产出，分别表示为 $E_1(\alpha, \beta)$ 和 $E_2(\alpha, \beta)$ ，并把均衡总产出表示为 $E = E_1 + E_2$ 。

假设 3: 与方程组(5)、(6)相关的二阶条件成立。

下面我们首先考虑国有股份比重的变化对均衡产出的影响，即 α 的变化对 E_1 和 E_2 的影响。由(5)、(6)分别对 α 求偏导，可得：

$$\begin{aligned}
 \frac{\partial E_1}{\partial \alpha} [P''(1 - \alpha + \alpha\beta)E_1 + P'(2 - \alpha + 2\alpha\beta) - C_2''] & \\
 + \frac{\partial E_2}{\partial \alpha} [P''(1 - \alpha + \alpha\beta)E_1 + (1 + \alpha\beta)P'] &= (1 - \beta)E_1 P' - \beta P \quad (7)
 \end{aligned}$$

$$\frac{\partial E_1}{\partial \alpha} [P'E_2 + P'] + \frac{\partial E_2}{\partial \alpha} [P'E_2 + 2P' - C_2''] = 0 \quad (8)$$

将方程(7)、(8)联立，我们可以解出：

^① 如果我们在模型的讨论中加入对效率提高效应的考虑，则国企业的成本函数变为 $C_1(Q_1, \alpha)$ 。为了在模型中体现国有股份比重下降所产生的效率提高效应，我们假设 $\frac{\partial C_1}{\partial \alpha} > 0$ 。然而，由于现有文献中没有足够的实证分析能够使我们明确对国有企业边际成本的影响方向，我们难以判定 $\frac{\partial^2 C_1}{\partial Q \partial \alpha}$ 的符号。如果我们假定 $\frac{\partial^2 C_1}{\partial Q \partial \alpha} \neq 0$ ，则 α 的变化对国有企业均衡产出的影响方向是不确定的，从而使得接下来的所有讨论都难以得到明确的结果。相反，如果我们假设 $\frac{\partial^2 C_1}{\partial Q \partial \alpha} = 0$ ，则 α 的变化对国有企业均衡产出的影响与没有在国企的生产成本函数中考虑 α 的情形相同。在这种情形下，我们就没有必要在国企的成本函数中引入 α 对其影响的考虑。

$$\frac{\partial E_1}{\partial a} = \frac{[(1 - \beta)P' E_1 - \beta P' [P'' E_2 + 2P' - C_2'']]}{\Delta} \quad (9)$$

$$\frac{\partial E_2}{\partial a} = - \frac{[(1 - \beta)P' E_1 - \beta P' [P'' E_2 + P']]}{\Delta} h = \quad (10)$$

其中：

$$\Delta \equiv \begin{vmatrix} K & H \\ P'' E_2 + P' & P'' E_2 + 2P' - C_2'' \end{vmatrix}$$

$$H = P''[(1 - \alpha + \alpha\beta)E_1] + P'(1 + \alpha\beta)$$

$$K \equiv \{P''[(1 - \alpha + \alpha\beta)E_1] + P'(2 - \alpha + 2\alpha\beta) - C_1''\}$$

我们把根据一阶条件得到的反应函数分别记为 $R_1(Q_2), R_2(Q_1)$ ，其中：

$$R_1(Q_2) \equiv \arg \max_{Q_1 \geq 0} U_1(Q_1, Q_2)$$

$$R_2(Q_1) \equiv \arg \max_{Q_2 \geq 0} U_2(Q_1, Q_2)$$

根据 Zhang 等 (1996)，为保证由古诺竞争所决定的经济系统的稳定性以及体现不同企业间产品的战略性替代关系，我们增加以下的假设：

$$\text{假设 4: } -1 < \frac{\partial R_i}{\partial Q_j} < 0, i, j = 1, 2, i \neq j.$$

由以上假设以及相关的一阶条件，我们可得到引理 2.1：

$$\text{引理 2.1: 如果满足假设 1} \sim 4, \text{且 } E_1 > 0, E_2 > 0, \text{则 } \frac{\partial E_1}{\partial \alpha} > 0, \frac{\partial E_2}{\partial \alpha} < 0, \frac{\partial E}{\partial \alpha} > 0.$$

证明见附录 A。

引理 2.1 说明，在均衡状态下， E_1 是 α 的增函数， E_2 是 α 的减函数，并且总的均衡产出 $E = E_1 + E_2$ 也是 α 的增函数。其经济含义是显然的： α 越大，意味着国企中国有股份的比重越大，企业生产决策受政府政策目标影响的程度也就越高，因而所承担的社会性负担也就越重。正如我们在模型的设计中所述，社会性负担在很大程度上影响国有企业的产出水平，两者成同向变动的关系。因而国有股份比重的增大必然带来国有企业均衡产出的增加，即 α 的增加导致了 E_1 的增加；另一方面，由于在本文的讨论中，我们已经假设国企与私企生产的产品具有完全的替代性且消费者的偏好稳定不变，国企产出增加的结果必然会导致私企产出的减少，也就是说， α 的增加导致了私企均衡产出 E_2 的减少。由于企业 2 对企业 1 产出变化的敏感程度小于 1，即 $\left| \frac{\partial R_2}{\partial Q_1} \right| < 1$ ，所以国有股份比重的加大所带来的企业 1 的均衡产出的增加要大于企业 2 均衡产出的减少，因而总产出水平增加。

下面我们再考虑国企中国有股份比重的变化对整个社会总体福利的影响，即 α 的变化对 w 的影响：

我们定义 $w^F(\alpha, \beta) \equiv W[E_1(\alpha, \beta), E_2(\alpha, \beta)]$ 为整个社会的均衡福利水平，并定义 $U_G^E(\alpha, \beta) \equiv U_G[E_1(\alpha, \beta), E_2(\alpha, \beta)]$ 为政府的均衡支付。

由引理 2.1, 我们容易得出以下命题:

命题 2.1: 若假设 1~4 成立, 则

(1) 当 $\alpha = 1$ 时, 如果 $E_1(1, \beta) > 0, E_2(1, \beta) > 0$, 且 $(P - C_1) < -R_2'(P - C_2)$, 则

$$\left. \frac{\partial W}{\partial \alpha} \right|_{\alpha=1} < 0, \text{ 即 } 1 \notin \arg \max_{\alpha} W^E.$$

(2) 当 $\alpha = 1$ 时, 如果 $E_1(0, \beta) > 0, E_2(0, \beta) > 0$, 且 $(P - C_1) > -R_2'(P - C_2) > 0$,

$$\text{则 } \left. \frac{\partial W}{\partial \alpha} \right|_{\alpha=0} > 0, \text{ 即 } 0 \notin \arg \max_{\alpha} W^E.$$

命题 2.1 的证明见附录 A。

我们将 $|R_2'(P - C_2)|$ 定义为国有企业的边际产出变化对私有企业的利润空间所产生的替代效应, 我们将它简称为企业 1 对企业 2 的边际替代效应。它由两部分构成: 企业 1 对企业 2 的边际产出替代效应 $(\frac{dR_2}{dQ_1})$ 以及企业 2 自身的盈利能力 $(P - C_2)$ 。

命题 2.1 具有深刻的政策含义, 它从社会总体福利的角度, 分别给出了完全由国家控股的国有企业是否需要进行私有化改革以及国有企业是否需要进行彻底的私有化改革的相关条件: 当国企的利润空间小于其对私有企业所产生的边际替代效应时, 采取完全的国有控股形式就不能最大化社会总体福利; 当国企的利润空间 (mark up) 大于其对私有企业所产生的边际替代效应时, 对国有企业进行完全的私有化改革同样不能最大化社会总体福利。事实上, 国企的利润空间大于其对私企所产生的边际替代效应, 意味着国企的生产效率已经达到一定的水平, 在这种条件下, 就没有必要对国企进行完全的私有化, 保留一定程度的国有控股还可以增加消费者剩余。相反, 如果原来完全由国家控股的国企的利润空间小于其对私企所产生的边际替代效应, 即国企的利润空间较小, 因而它的生产相对于社会平均水平而言处于低效率的生产状态。如果此时它完全以国家的政策目标作为企业自身的目标, 必然会带来过量的生产, 加重低效率生产的程度。所以说, 从整个社会总体福利的角度看, 如果完全由国家控股的国企的盈利能力较弱, 就必须进行私有化改革。

以下我们讨论在均衡状态下, 国有股份比重 α 对政府均衡支付 G^E 的影响。

国有股份比重的变化对政府均衡支付的影响可分解为两种效应的乘积: α 的变化对均衡产出 E_1 的影响 $\frac{\partial E_1}{\partial \alpha}$, 以及均衡产出的变化对政府支付的影响 $\frac{\partial G}{\partial E_1}$ 。

由 $G = W + \beta PQ_1 = \int_0^{Q_1+Q_2} Pdq - C_1(Q_1) - C_2(Q_2) + \beta PQ_1$ 对 E_1 求偏导, 可得:

$$\frac{\partial G^E}{\partial E_1} = \beta P' E_1 (1 + R_2') + (1 + \beta) P - C_1' + (P - C_2) R_2' \quad (11)$$

经过一些必要的数学推导 (见附录 A), 我们得出以下命题:

命题 2.2: 若假设 1~4 成立, 则

(1) 当 $\alpha = 0$ 时, 如果 $E_1(1, \beta) > 0, R_2'(E_1) > 0$, 且 $C_1' > C_2' + \beta P$, 则 $\left. \frac{\partial G^E}{\partial \alpha} \right|_{\alpha=1} < 0$, 即:

$$1 \notin \arg \max_{\alpha} G^E;$$

(2) 当 $\alpha = 0$ 时, 如果 $E_1(\alpha, \beta) > 0, R_2(E_1) > 0$, 且 $P - C_1 > -(P - C_2)R_2$, 则有:

$$\left. \frac{\partial G^E}{\partial \alpha_1} \right|_{\alpha=0} > 0, \text{ 即 } 0 \notin \arg \max_{\alpha} G^E.$$

命题 2.2 的(1)的经济含义是: 当国有企业的边际生产成本超出私有企业的边际生产成本达到一定程度(用 βP 衡量)时, 从最大化政府支付的角度出发, 就必须对完全由国家控股的企业进行私有化改革。换句话说, 当国有企业的生产效率太低时, 即使政府希望国有企业更多地承担社会性负担, 也不能通过对企业采取完全国有控股的方式来达到这一目的。

命题 2.2 的(2)的经济含义是: 当国有企业的利润空间高于其对私有企业的边际替代效应时, 完全的私有化也不是政府的最优选择。在这一条件下, 虽然仍存在国有企业的生产效率低于私有企业生产效率的可能, 即 $C_1 > C_2$, 但它们之间的差距已经不大。此时, 政府出于维持社会稳定的考虑, 对国有企业不进行完全的私有化改革而保留部分的国有控股, 让国企承担部分的社会性负担, 可起到增加政府支付的作用。因而, 这一命题就给出了国有企业不进行完全私有化改革的一个具有可操作性的条件。

近年来, 我国国有企业的产值份额在不断下降, 其工业产值占总产值的比重已不到 30%, 但其就业份额仍超过 50%。这表明, 政府将经济增长的希望寄托在非国有部门身上, 而国有部门对于社会目标的实现则显得越来越重要了。特别是由于国有企业的产权变革总是从一些绩效较差的国有企业开始的, 目前的存量国有企业的绩效是相对较好的(即与私有企业的生产效率差距不大), 这些企业似乎就更加没有理由要求放弃社会目标了(陆铭, 2003)。这与命题 2.2 的结论是非常吻合的。也就是说, 完全的私有化改革应集中于那些生产效率过于低下的国有企业, 对于那些具有一定生产效率的国企, 仍应当承担部分的社会性负担。

从命题 2.1 与命题 2.2 我们可以看出, 无论是出于社会总体福利的角度, 还是出于政府政策目标的角度考虑, 最优的国有股份比重的选择都必须考虑各类企业的生产成本、它们产出之间的边际替代效应以及社会就业压力等因素的影响。在满足某些标准条件的情形下, 完全的国有化或完全的私有化都不能最大化社会总体福利或政府支付。当然, 以上讨论仅仅是从国企承担社会性负担的程度进行分析, 而没有考虑国有企业在私有化进程中所产生的效率提高效应, 有兴趣的读者可将这一效应纳入到模型中作进一步的深入探讨。

三、结果模拟

为了更好地理解第二部分的两个命题, 我们下面通过构造具体的需求以及成本函数, 分析最优国有股份比重与就业压力之间以及最优国有股份比重与国企的生产效率之间的相互关系。

我们令:

$$P = \alpha - Q_1 - Q_2, C_1 = \frac{k}{32} Q_1^2, C_2 = \frac{1}{2} Q_2^2 \quad (\alpha > 0, k > 0) \quad (12)$$

显然, 这些具体的函数形式都满足第二部分所作的基本假设 1~4。值得注意的是, 我们具体设定私有企业的成本函数, 并通过在国有企业的成本函数中引入参数 k 来衡量国有

企业的相对生产效率。显然,当参数 $k < 16$ 时,意味着国有企业的生产效率高于私有企业;当 $k = 16$ 时,国有企业与私有企业具有相同的生产效率;当参数 $k > 16$ 时,国有企业的生产效率低于私有企业。

根据一阶条件 (5) 与 (6), 可以计算出:

$$E_1 = \frac{2 + 2\alpha\beta}{5 - 3\alpha + 5\alpha\beta + \frac{3}{16}k} \alpha, E_2 = \frac{1 - \alpha + \alpha\beta + \frac{1}{16}k}{5 - 3\alpha + 5\alpha\beta + \frac{3}{16}k} \alpha \quad (13)$$

我们首先从社会总体福利的角度出发,考虑以下两种变化趋势:在给定国有企业不同相对生产效率(用 k 衡量)的条件下,最优的国有股份比重 α_w^* 随 β 变化的趋势;在给定不同就业压力(用 β 衡量)的情况下, α_w^* 随 k 变化的趋势。

由 (A.7), 令 $\frac{\partial W^E}{\partial E_1} = 0$, 可得:

$$\alpha_w^* = \frac{k - 80}{\beta(80 - 6k) - 80} \quad (14)$$

由 (14) 式对 β 求偏导, 可知:

$$\frac{\partial \alpha_w^*}{\partial \beta} = - \frac{(k - 80)(80 - 6k)}{[\beta(80 - 6k) - 80]^2} \quad (15)$$

显然,当 $k < \frac{40}{3}$ 时, $\frac{\partial \alpha_w^*}{\partial \beta} > 0$, 即 α_w^* 随着 β 的增加而增加;当 $\frac{40}{3} < k < 80$ 时, $\frac{\partial \alpha_w^*}{\partial \beta} < 0$, 即 α_w^* 随着 β 的增加而减少。鉴于当 $k \geq 80$ 时,国有企业的相对生产效率太低,这与实际情况不符,因而我们不予以考虑。

从上面的结果可以看出, $k = \frac{40}{3}$ 是一个临界点。当 $k < \frac{40}{3}$ 时,表明国有企业具有较高的生产效率,此时,从社会总体福利的角度看,就业压力越大,高效率的国有企业就应当承担越多的社会性负担;相反,当 $k > \frac{40}{3}$ 时,表明国有企业的生产效率相对于私有企业而言不具有太大的优势,此时,就业压力越大,政府的政策目标中就越发偏重国有企业的产出,如果 α_w^* 保持不变,则国有企业的社会性负担势必加重,这将影响国有企业的生产效率。从社会总体福利的角度看,就必须相应地减少最优的国有股份的比重 α_w^* 。

我们再由 (14) 式对 k 求偏导, 可得:

$$\frac{\alpha_w^*}{\partial k} = \frac{-400\beta - 80}{[\beta(80 - 6k) - 80]^2} < 0 \quad (16)$$

也就是说,在给定就业压力 β 的情况下,随着 k 的增加,即随着国有企业效率的下降,最优的国有股份比重 α_w^* 会相应地下降。也就是说,从社会总体福利的角度考虑,如果国有企业的相对生产效率下降,则应当相应地减少国有企业的社会性负担,从而减少其产出以降低过量生产的程度,避免由生产效率下降所带来的社会总体福利损失。也就是说,在给定就业压力的条件下,可以通过减少国有股份比重的方式来弥补国有企业相对生产效率的下降所带来的社会总体福利损失。

综合以上讨论,我们可得出如下命题:

命题 3.1: 假设需求函数与成本函数具有 (12) 的形式, 则从最大化社会总体福利的角度考虑, 有:

(1) 在给定 k 的条件下, 当 $k < \frac{40}{3}$ 时, $\frac{\partial \alpha_w^*}{\partial \beta} > 0$; 当 $\frac{40}{3} < k < 80$ 时, $\frac{\partial \alpha_w^*}{\partial \beta} < 0$ 。

(2) 在给定 $\beta (\beta \geq 0)$ 的条件下, $\frac{\partial \alpha_w^*}{\partial k} < 0$ 。

为了得到具体的最优国有股份比重的值并更好地反映 α_w^* 随着 β 变化的具体趋势, 我们对 k 赋予一些特殊值, 给出相应于不同 k 值的最优的国有股份比重 α_w^* 的值 (见表 1)。类似地, 我们对 β 赋予一些特殊值, 并给出相应于不同 β 值的最优的国有股份比重 α_w^* 的值 (见表 2)。

从表 1 可以看出, 当 $k = 12$ 时, α_w^* 随 β 增加而增加; 当 $k = 16, 24$ 以及 32 时, α_w^* 随 β 增加而减少。同样, 从表 2 可以看出, 当 $\beta = 0.05, 0.10, 0.15$ 及 0.20 时, α_w^* 随 k 增加而减少。上述模拟结果与命题 3.1 的结论是吻合的。

现在我们从政府政策目标的角度出发, 考虑在给定国有企业的不同相对生产效率 (用 k 衡量) 的条件下, α_G^* 随 β 变化的趋势。

由 (11) 式, 令 $\frac{\partial G^E}{\partial E_1} = 0$, 可得:

$$\alpha_G^* = \frac{k - 6\beta k - 32\beta - 80}{32\beta^2 - 6k\beta - 16\beta - 80} \quad (17)$$

由 (17) 对 β 求偏导,

$$\frac{\partial \alpha_G^*}{\partial \beta} = \frac{2(96k + 512)\beta^2 + 2(2560 - 32k)\beta + 2(3k^2 + 8k + 640)}{(32\beta^2 - 6k\beta - 16\beta - 80)^2} \quad (18)$$

可以证明, 当 $0 < k < 80$ 时, $\frac{\partial \alpha_G^*}{\partial \beta} > 0$, 也就是说, 从政府的角度看, 当就业压力越大时, 要提高政府的收益, 应当相应地增加国有企业中国有股份所占比重。

我们再由 (17) 对 k 求偏导, 可得:

$$\frac{\partial \alpha_G^*}{\partial k} = - \frac{16(12\beta^3 + 4\beta^2 + \beta + 5)}{(32\beta^2 - 6k\beta - 16\beta - 80)^2} < 0 \quad (19)$$

(19) 式表明, 在给定就业压力的条件下, 当国有企业的生产效率越低时, 要增加政府的收益, 就必须相应地减少最优的国有股份的比重。

由此可得以下命题:

命题 3.2: 假设需求函数与成本函数具有 (12) 的形式, 则从最大化政府收益的角度考虑, 有:

(1) 在给定 $k (0 < k < 80)$ 的条件下, $\frac{\partial \alpha_G^*}{\partial \beta} > 0$ 。

(2) 在给定 $\beta (\beta \geq 0)$ 的条件下, $\frac{\partial \alpha_G^*}{\partial k} < 0$ 。

与前面的讨论相类似, 为了得到具体的最优国有股份比重的值并更好地反映 α_G^* 随着 β 变化的具体趋势, 我们对 k 赋予一些特殊值, 给出相应于不同 k 值的最优的国有股份比重

α_G^* 的值(见表 3)。类似地,我们对 β 赋予一些特殊值,并给出相应于不同 β 值的最优的国有股份比重 α_G^* 的值(见表 4)。

从表 3 可以看出,当 $k = 12, 16, 24$ 以及 32 时, α_G^* 随 β 增加而增加。同样,从表 4 可以看出,当 $\beta = 0.05, 0.10, 0.15$ 及 0.20 时, α_G^* 随 k 增加而减少。上述模拟结果与命题 3.2 的结论是吻合的。

综合命题 3.1 与命题 3.2,我们发现,在给定国有企业相对生产成本的条件下,如果国有企业具有较低的相对生产成本(即 $k < \frac{40}{3}$),则无论是从社会总体福利的角度考虑还是从政府政策目标的角度考虑,最优的国有股份比重与就业压力应当同向变动,即随着 β 的增加, α_G^* 与 α_W^* 也会相应增加,然而,值得注意的是,如果国有企业的相对生产成本高于某一临界指标(即 $k > \frac{40}{3}$)时,则社会总体福利目标所要求最优的国有股份比重与就业压力呈反向变动,而政府政策目标则要求两者呈同向变动。此时,要增进社会总体福利就必然损害政府收益,而兼顾政府收益就必须牺牲社会总体福利,所以政府政策目标与社会总体福利目标两者不可能同时兼顾。

另外,在给定社会就业压力 β 的条件下,无论国有企业的生产效率如何,社会总体福利目标与政府的政策目标都要求最优的国有股份比重与国有企业的生产效率呈同向变动,即 α^* 与 k 呈反向变动。换句话说,在既定的社会压力条件下,无论是从社会总体福利的角度还是从政府政策目标的角度看,国有企业的生产效率越低,则国企私有化改革的程度就应当越大。

需要进一步说明的是,当 $\beta > 0$ 且 $k > 0$ 时,方程(14)与(17)没有公共解。从直观上看,这一结果也是显然的,因为政府的政策目标与社会总体福利目标是不一致的,因而,最大化政府的政策目标与最大化社会总体福利目标所要求的最优国有股份比重是不同的,即 $\alpha_W^* \neq \alpha_G^*$ 。但值得注意的是,从表 1、表 3 可以看出,在给定相同的就业压力的条件下,当 $k < \frac{40}{3}$ 时最大化政府政策目标所要求的最优国有股份比重 α_G^* 与最大化社会总体福利目标所要求的最优国有股份比重 α_W^* 之间的差距,小于当 $k > \frac{40}{3}$ 时 α_G^* 与 α_W^* 之间的差距。也就是说,当国有企业的生产效率较高($k < \frac{40}{3}$)时,政府政策目标 G 与社会总体福利目标 W 之间的冲突程度,要小于当国有企业生产效率较低($k > \frac{40}{3}$)时 G 与 W 之间的冲突程度。

四、模型的扩展：外国竞争者的情形

鉴于中国已经正式成为世界贸易组织的成员,并将于今后几年中逐步降低对 WTO 成员国的关税,最终实现贸易自由化,因而在本节进一步引入外国企业的竞争因素,对基本模型进行扩展,讨论关税的变化对各类企业产出的影响以及在实行完全的贸易自由化条件下,国有股份比重的变化对政府支付和社会总体福利的影响。

假设现在市场上有三个企业：国有企业、私有企业以及外国企业，我们将其分别表示为企业 1, 2 和 3。与前面的假设相类似，这三个企业生产相同的产品且产品间具有完全的替代性。作为一个外国企业，在没有实现贸易自由化以前，企业 3 必须对单位产品支付一定的关税，我们用 $t \geq 0$ 来表示，其中 $t = 0$ 表示完全取消关税壁垒或实现完全的贸易自由化。因而，企业 3 的利润函数 π_3 可表示为：

$$\pi_3 = P(Q)Q_3 - C_3(Q_3) - tQ_3 \quad (20)$$

此时，政府通过对企业 3 征收关税可获得关税收入 tQ_3 。所以，社会总体福利由三部分构成：生产者剩余、消费者剩余以及政府的关税收入，即：

$$W = \pi_1 + \pi_2 + CS + tQ_3 = \int_0^{Q_1+Q_2+Q_3} Pdq - C_1(Q_1) - C_2(Q_2) - PQ_3 + tQ_3 \quad (21)$$

相应地，政府的目标函数变为：

$$G = W + \beta PQ_1 = \beta PQ_1 - C_1 - C_2 - (P - t)Q_3 + \int_0^{Q_1+Q_2+Q_3} Pdq \quad (22)$$

与前面的讨论类似，国有企业、私有企业与外国企业分别最大化他们各自的效用函数 U_1, U_2 和 U_3 ，其中：

$$U_1 = (1 - \alpha + \alpha\beta)PQ_1 - C_1 - \alpha C_2 + \alpha \int_0^{Q_1+Q_2+Q_3} Pdq - \alpha Q_3(P - t) \quad (23)$$

$$U_2 = PQ_2 - C_2 \quad (24)$$

$$U_3 = \pi_3 = PQ_3 - C_3 - tQ_3 \quad (20)$$

为讨论方便，我们将原来的假设 1 和 2 作如下调整：

假设 1' : $\forall Q \geq 0, P > 0, P(Q)$ 是二次可微的，且 $P'(Q) < 0$ ，其中 $Q = Q_1 + Q_2 + Q_3$ 。

假设 2' : $\forall Q_i \geq 0, C_i'(Q_i) > 0$ 且 $C_i''(Q_i) \geq 0, i = 1, 2, 3$ 。

由(23)、(24)、(20)式分别对 Q_1, Q_2 和 Q_3 求导，根据一阶条件可得：

$$P'(1 - \alpha + \alpha\beta)Q_1 + (1 + \alpha\beta)P - C_1' - \alpha P'Q_3 = 0 \quad (25)$$

$$P'Q_2 + P - C_2' = 0 \quad (26)$$

$$P'Q_3 + P - (C_3' + t) = 0 \quad (27)$$

联立方程(25)、(26)、(27)，可以得到三个企业各自的均衡产出 $E_1(\alpha, \beta, t), E_2(\alpha, \beta, t)$ 和 $E_3(\alpha, \beta, t)$ ，我们同样把总产出记为 $E = E_1 + E_2 + E_3$ 。

进一步，我们对基本模型的假设 3 作如下调整：

假设 3' : 与方程组(25)、(26)、(27)相关的二阶条件成立。

与前面的讨论相类似，我们将假设 4 作如下调整：

假设 4' : $-1 < \frac{\partial R_i}{\partial Q_j} < 0, i, j = 1, 2, 3$ ，且 $i \neq j$ 。

下面我们考虑国有股份比重 α 的变化对均衡产出 E_i 的影响。

由(25)、(26)、(27)分别对 α 求偏导，得到：

$$\frac{\partial E_1}{\partial \alpha} K_1 + \frac{\partial E_2}{\partial \alpha} H_1 + \frac{\partial E_3}{\partial \alpha} L_1 = (1 - \beta)P'E_1 - \beta P + P'E_3 \quad (28)$$

$$\frac{\partial E_1}{\partial \alpha}(P''E_2 + P') + \frac{\partial E_2}{\partial \alpha}(P''E_2 + 2P' - C_2'') + \frac{\partial E_3}{\partial \alpha}(P''E_2 + P') = 0 \quad (29)$$

$$\frac{\partial E_1}{\partial \alpha}(P''E_3 + P') + \frac{\partial E_2}{\partial \alpha}(P''E_3 + P') + \frac{\partial E_3}{\partial \alpha}(P''E_3 + 2P' - C_3'') = 0 \quad (30)$$

将方程(28)、(29)、(30)联立,记其系数行列式的值为 Δ_1 , 则:

$$\frac{\partial E_1}{\partial \alpha} = \frac{\Delta_{23}}{\Delta_1} [(1 - \beta)P'E_1 - \beta P' + P'E_3] \quad (31)$$

$$\frac{\partial E_2}{\partial \alpha} = -\frac{1}{\Delta_1} [(1 - \beta)P'E_1 - \beta P' + P'E_3] (P''E_3 + P') (P' - C_3'') \quad (32)$$

$$\frac{\partial E_3}{\partial \alpha} = -\frac{1}{\Delta_1} [(1 - \beta)P'E_1 - \beta P' + P'E_3] (P''E_3 + P') (P' - C_2'') \quad (33)$$

其中:

$$\Delta_1 \equiv \begin{vmatrix} K_1 & H_1 & L_1 \\ P''E_2 + P' & P''E_2 + 2P' - C_2'' & P''E_2 + P' \\ P''E_3 + P' & P''E_3 + P' & P''E_3 + 2P' - C_3'' \end{vmatrix}$$

$$K_1 \equiv P''[(1 - \alpha + \alpha\beta)E_1 - \alpha E_3] + P'(2 - \alpha + 2\alpha\beta) - C_1''$$

$$H_1 \equiv P''[(1 - \alpha + \alpha\beta)E_1 - \alpha E_3] + P'(1 + \alpha\beta)$$

$$L_1 \equiv P''[(1 - \alpha + \alpha\beta)E_1 - \alpha E_3] + P'(1 - \alpha + \alpha\beta)$$

$$\Delta_{23} \equiv \begin{vmatrix} P''E_2 + 2P' - C_2'' & P''E_2 + P' \\ P''E_3 + P' & P''E_3 + 2P' - C_3'' \end{vmatrix} \text{ 为 } \Delta_1 \text{ 的二阶顺序主子式}$$

我们容易得到以下引理:

引理 4.1 如果假设 1 ~ 4 满足,则有 $\frac{\partial E_1}{\partial \alpha} > 0$, $\frac{\partial E_2}{\partial \alpha} < 0$, $\frac{\partial E_3}{\partial \alpha} < 0$, 且 $\frac{\partial E}{\partial \alpha} > 0$ 。

证明见附录 B。

引理 4.1 的经济含义是:随着国有股份比重的降低,国有企业的均衡产出会相应减少,而私有企业与外国企业的均衡产出都会相应增加,并且总产出水平会下降。这说明国有股份比重的变化所引起的企业 2 与企业 3 的联合产出变化幅度小于企业 1 的产出变化幅度。导致这一结果的内在经济机制与引理 2.1 中所述是相同的,此处不再赘述。

接下来我们再考虑国有股份比重的变化对均衡的社会总体福利 W^E 的影响:

给定均衡产出 E_1, E_2 , 并假定 $Q_3 = R_3(Q_1, Q_2)$, 由 W 的定义,有:

$$W = \pi_1 + \pi_2 + CS + tQ_3 = \int_0^{Q_1+Q_2+Q_3} Pdq - C_1(Q_1) - C_2(Q_2) - PQ_3 + tQ_3$$

则均衡状态下的社会总体福利为:

$$W^E = \int_0^{E_1+E_2+R_3(E_1, E_2)} Pdq - C_1 - C_2 - PR_3(E_1, E_2) + tR_3(E_1, E_2)$$

由 W^E 对 α 求偏导,可得:

$$\frac{\partial W^E}{\partial \alpha} = \frac{\partial W^E}{\partial E_1} \frac{\partial E_1}{\partial \alpha} + \frac{\partial W^E}{\partial E_2} \frac{\partial E_2}{\partial \alpha}$$

其中：

$$\frac{\partial W^E}{\partial E_1} = P - C_1 + t \frac{\partial R_3}{\partial E_1} - P' R_3 \left(1 + \frac{\partial R_3}{\partial E_1}\right) \quad (34)$$

$$\frac{\partial W^E}{\partial E_2} = P - C_2 + t \frac{\partial R_3}{\partial E_2} - P' R_3 \left(1 + \frac{\partial R_3}{\partial E_2}\right) \quad (35)$$

所以有：

$$\frac{\partial W^E}{\partial \alpha} = \left[P - C_1 + t \frac{\partial R_3}{\partial E_1} - P' R_3 \left(1 + \frac{\partial R_3}{\partial E_1}\right) \right] \frac{\partial E_1}{\partial \alpha} + \left[P - C_2 + t \frac{\partial R_3}{\partial E_2} - P' R_3 \left(1 + \frac{\partial R_3}{\partial E_2}\right) \right] \frac{\partial E_2}{\partial \alpha}$$

当 $t = 0$ 时，有：

$$\begin{aligned} \frac{\partial W^E}{\partial \alpha} &= \left[P - C_1 - P' R_3 \left(1 + \frac{\partial R_3}{\partial E_1}\right) \right] \frac{\partial E_1}{\partial \alpha} + \left[P - C_2 - P' R_3 \left(1 + \frac{\partial R_3}{\partial E_2}\right) \right] \frac{\partial E_2}{\partial \alpha} \\ &= (P - C_1) \frac{\partial E_1}{\partial \alpha} + (P - C_2) \frac{\partial E_2}{\partial \alpha} + (P - C_3) \left(1 + \frac{\partial R_3}{\partial E_1}\right) \left(\frac{\partial E_1}{\partial \alpha} + \frac{\partial E_2}{\partial \alpha} \right) \end{aligned} \quad (36)$$

我们把(36)式的第一项定义为 α 的变化对企业 1 的利润空间所产生的效应，第二项为 α 的变化对企业 2 的利润空间所产生的效应，简称为效应 1 和效应 2。由于 $\frac{\partial R_3}{\partial E_1} = \frac{\partial R_3}{\partial E_2}$ (见附录

B) 我们将 $(P - C_3) \left| \frac{\partial R_3}{\partial E_1} \right|$ 定义为企业 i 对企业 3 的利润空间的边际替代效应，则 $(P - C_3) \left| \frac{\partial R_3}{\partial E_1} \right| \left(\frac{\partial E_1}{\partial \alpha} + \frac{\partial E_2}{\partial \alpha} \right)$ 可以看成是 α 的变化所引起的国内企业产出变化对企业 3 利润空间

的替代效应。另一方面， $(P - C_3) \left(\frac{\partial E_1}{\partial \alpha} + \frac{\partial E_2}{\partial \alpha} \right)$ 可以看成是 α 的变化所引起的国内企业边际产出变化对企业 3 利润空间所产生的直接效应，则(36)式的第三项可以看成是 α 的变化所引起的国内企业产出变化对企业 3 利润空间的净效应，简称为国内企业对企业 3 的净效应。

并将 $(P - C_3) \left(1 + \frac{\partial R_3}{\partial E_1}\right)$ 称为国内企业对企业 3 的边际净效应。由于 $\frac{\partial E_1}{\partial \alpha} > \left| \frac{\partial E_2}{\partial \alpha} \right|$ ，(36) 式的第三项的符号为正，即 α 的增加会使使得国内企业对企业 3 的净效应为正。也就是说，单纯从这一效应来看，增加 α 有助于增进社会总体福利。但从总体上来考虑，社会总体福利最终取决于以上三种效应所产生的综合影响。

由(36)式及引理 4.1，我们可以得到以下命题：

命题 4.1 若假设 1' ~ 4 成立，则

(1) 当 $\alpha = 1$ 时，如果 $P - C_1 < 0$ ，且 $|P - C_1| > (P - C_3) \left(1 + \frac{\partial R_3}{\partial E_1}\right)$ ，则 $\frac{\partial W^E}{\partial \alpha} \Big|_{\alpha=1} < 0$ ，即 $1 \notin \arg \max_{\alpha} W^E(\alpha, \beta, \theta)$ 。

(2) 当 $\alpha = 0$ 时，如果 $C_1 < C_2$ ，则 $\frac{\partial W^E}{\partial \alpha} \Big|_{\alpha=0} > 0$ ，即 $0 \notin \arg \max_{\alpha} W^E(\alpha, \beta, \theta)$ 。

证明见附录 B。

命题 4.1 从社会总体福利的角度给出了我国在实行贸易自由化政策后，采取完全国有控股形式的国企是否必须进行私有化改革以及国有企业私有化改革程度的相关条件。命题

4.1 中的(1)的经济含义是:在实行完全的贸易自由化条件下,如果完全由国家控股的国有企业生产效率 q_1 率低下,处于亏损状态,且亏损幅度大于国内企业对企业3的边际净效应,则从社会总体福利的角度考虑,采取完全的国有化形式不能最大化社会总体福利,因而必须对完全由国家控股的国企实行私有化改革。

当 $C_1 < C_2$ 时,由(36)式可知,效应1为正,效应2为负,且效应1大于效应2的绝对值,而国内企业对企业3的净效应为正,因而总体社会总体福利会增加。命题4.1中的(2)为一个单纯追求利润最大化的经济主体进行生产决策时,倘若其边际成本小于私有企业的边际成本,则不应该对国有企业进行彻底的私有化改革,否则就会降低总体的社会总体福利。也就是说,如果国有企业的生产效率不低于私有企业的生产效率,那么从社会总体福利的角度考虑,就应当对国有企业保留一定程度的国有控股。其中的原因在于总体社会福利必须同时兼顾消费者剩余与国有企业利润两个因素,而完全的私有化虽然能够提高企业的利润,但由引理4.1,社会总产出就会减少,由此带来消费者剩余的减少,进而降低总体社会总体福利。

以下再讨论国有股份比重的变化对均衡的政府支付 G^E 的影响。

由于 $\frac{\partial G^E}{\partial \alpha} = \frac{\partial G^E}{\partial E_1} \frac{\partial E_1}{\partial \alpha} + \frac{\partial G^E}{\partial E_2} \frac{\partial E_2}{\partial \alpha}$,我们需要求出 $\frac{\partial G^E}{\partial E_1}$ 与 $\frac{\partial G^E}{\partial E_2}$ 。同样,给定均衡产出 E_1 , E_2 ,并假定 $Q_3 = R_3(Q_1, Q_2)$ 。

由 G^E 分别对 E_1 和 E_2 求偏导,得到:

$$\begin{aligned} \frac{\partial G^E}{\partial E_1} &= P' \left(1 + \frac{\partial R_3}{\partial E_1}\right) (\beta E_1 - R_3) + (1 + \beta)P - C_1 + t \frac{\partial R_3}{\partial E_1} \\ \frac{\partial G^E}{\partial E_1} \Big|_{t=0} &= P' \left(1 + \frac{\partial R_3}{\partial E_1}\right) (\beta E_1 - R_3) + (1 + \beta)P - C_1 \end{aligned} \quad (37)$$

$$\begin{aligned} \frac{\partial G^E}{\partial E_2} &= P' \left(1 + \frac{\partial R_3}{\partial E_2}\right) (\beta E_1 - R_3) + P - C_2 + t \frac{\partial R_3}{\partial E_2} \\ \frac{\partial G^E}{\partial E_2} \Big|_{t=0} &= P' \left(1 + \frac{\partial R_3}{\partial E_2}\right) (\beta E_1 - R_3) + P - C_2 \end{aligned} \quad (38)$$

则有:

$$\begin{aligned} \frac{\partial G^E}{\partial \alpha} \Big|_{t=0} &= \left[P' \left(1 + \frac{\partial R_3}{\partial E_1}\right) (\beta E_1 - R_3) + (1 + \beta)P - C_1 \right] + \frac{\partial E_1}{\partial \alpha} \\ &\quad + \left[P' \left(1 + \frac{\partial R_3}{\partial E_2}\right) (\beta E_1 - R_3) + P - C_2 + t \frac{\partial R_3}{\partial E_2} \right] \frac{\partial E_2}{\partial \alpha} \end{aligned} \quad (39)$$

由(39)及引理4.1,我们有以下命题:

命题4.2 若假设1~4满足,则

(1) 当 $\alpha = 1, t = 0$ 时,如果 $(P - C_1) + \beta P < 0$,则 $1 \notin \arg \max_{\alpha} G^E$ 。

(2) 当 $\alpha = 0, t = 0$ 时,如果 $(P - C_1) + \beta P > P - C_2$,则 $0 \notin \arg \max_{\alpha} G^E$ 。

命题4.2的经济含义是:对于一个完全由国家控股的国企而言,如果其生产效率十分低下,其亏损额度低于临界指标 βP ,则在实行完全的贸易自由化条件下($t = 0$),政府从自身支付最大化的角度出发,应当对其进行私有化改革;相反,如果国有企业与私有企业的生产效率差异不大,小于临界指标 βP ,则政府应当对国有企业保留一定的股份控制权,即对国有

企业采取完全私有控股的形式不是政府的最优选择。因此,命题 4.2 就从政府支付的角度,具体明确了国有企业不能采取完全国有或完全私有形式的充分条件。

下面我们再考虑关税率的变化对均衡产出 $E_i (i = 1, 2, 3)$ 的影响。

由一阶条件对 t 求偏导,得到:

$$\frac{\partial E_1}{\partial t} K_1 + \frac{\partial E_2}{\partial t} H_1 + \frac{\partial E_3}{\partial t} L_1 = 0 \quad (40)$$

$$\frac{\partial E_1}{\partial t} [P''E_2 + P'] + \frac{\partial E_2}{\partial t} [P''E_2 + 2P' - C_2] + \frac{\partial E_3}{\partial t} [P''E_2 + P'] = 0 \quad (41)$$

$$\frac{\partial E_1}{\partial t} [P''E_3 + P'] + \frac{\partial E_2}{\partial t} [P''E_3 + P'] + \frac{\partial E_3}{\partial t} [P''E_3 + 2P' - C_3] = 1 \quad (42)$$

从方程组(40)、(41)、(42)我们可以得到:

$$\frac{\partial E_1}{\partial t} = \frac{1}{\Delta_1} \begin{vmatrix} H_1 & L_1 \\ P''E_2 + 2P' - C_2 & P''E_2 + P' \end{vmatrix} \quad (43)$$

$$\begin{aligned} \frac{\partial E_2}{\partial t} &= -\frac{1}{\Delta_1} \begin{vmatrix} K_1 & L_1 \\ P''E_2 + P' & P''E_2 + P' \end{vmatrix} = -\frac{1}{\Delta_1} (P''E_2 + P')(K_1 - L_1) \\ &= -\frac{1}{\Delta_1} (P''E_2 + P')[P'(1 + \alpha\beta) - C_2] \end{aligned} \quad (44)$$

$$\frac{\partial E_3}{\partial t} = \frac{\Delta_{12}}{\Delta_1} \quad (45)$$

其中 $\Delta_{12} \equiv \begin{vmatrix} K_1 & H_1 \\ P''E_2 + P' & P''E_2 + 2P' - C_2 \end{vmatrix}$ 是 Δ_1 的顺序主子式。

由假设 1 ~ 4 以及二阶条件可知:(44)式的符号为正,(45)式的符号为负。但(43)式的符号却是不确定的,我们可以举一个简单的例子来予以说明。假设需求曲线为线性的,并且

$P' = -1$ 私有企业的边际成本为常数,则 $\frac{\partial E_1}{\partial t} = \frac{1}{\Delta_1} [\alpha(1 - \beta) - (1 - \alpha)]$,显然它的符号取

决于 α, β 的具体取值。例如,当 $\alpha = 1/4, \beta = 1/2$ 时, $\frac{\partial E_1}{\partial t} > 0$; 但当 $\alpha = 3/4, \beta = 1/2$ 时,却有

$$\frac{\partial E_1}{\partial t} < 0。$$

由以上讨论我们有命题 4.3:

命题 4.3 如果满足假设 1 ~ 4, 则 $\frac{\partial E_2}{\partial t} > 0, \frac{\partial E_1}{\partial t} < 0, \text{sign} \frac{\partial E_1}{\partial t}$ 不确定。

命题 4.3 表明:随着关税率的降低,外国企业的成本会相应降低,因而外国企业的产出会增加,私有企业的产出会减少,但对国有企业产出的影响却是不确定的,它取决于国有股份的比重 α 以及政府对就业压力的重视程度 β 。

导致上述结果的原因在于,关税下降所产生的冲击效应主要体现在对企业利润的冲击,而国有企业在进行生产决策的时候不是以追求利润最大化为单一目标,它必须同时兼顾社会性负担。这样,与其他类型的企业相比,其利润目标在生产决策中的重要程度就相对降低了,国企的均衡产出水平在一定程度上还得取决于它所承受的社会性负担,在本文中体现为

α, β 的具体取值。所以, 关税下降对国企均衡产出影响的方向是不确定的。 α 的取值越小, 国企的生产决策行为就越接近于以单纯追求利润最大化为目标的私有企业, 关税下降对它的影响就会越大。在上述特殊例子中, 如果 $\alpha = 0$ 则有 $\frac{\partial E_1}{\partial t} > 0$, 即在完全私有化的条件下, 若外国企业的成本降低(关税税率 t 下降), 则会导致国有企业产出的缩减。这与直观是吻合的, 因为此时国有企业的生产决策就与私有企业完全一样, 以利润最大化为目标而无须兼顾社会性负担。

由于 $\text{sign} \frac{\partial E_1}{\partial t} < 0$ 的不确定, 我们难以在现有的模型框架下讨论 t 对社会总体福利 w 与政府支付 G 的影响, 有兴趣的读者可以对这一问题作进一步的探讨, 或对模型作进一步的扩展。

五、结语

国有企业改革的成功与否, 已经成为关系到国计民生和社会稳定的关键因素。长期以来, 国有企业一直在吸纳城镇劳动力、提供社会保障等方面发挥着社会稳定器的作用。改革开放 20 多年, 尽管私有经济得到了蓬勃发展, 提供了相当多的新增就业机会, 但国有企业仍然是吸收城镇劳动力的主要渠道。冗员现象的存在和庞大的退休职工队伍, 使国有企业在追求利润的同时, 必须承担一定的社会性任务, 因而在分析国企改革问题时, 就不能不考虑它所承担的社会性负担。

本文通过在国企的目标函数中引入其所承担的社会性负担, 将之与国有股份的比重联系起来, 并假定两者之间存在正比例关系, 分别就封闭经济与开放经济的情形, 讨论国有股份比重的变化对社会总体福利以及政府支付的影响。我们的分析表明, 无论是从社会总体福利的角度还是从政府政策目标的角度, 最优的国有股份比重的选择都取决于各类企业的生产成本、它们产出之间的边际替代效应以及整个社会的就业压力等因素之间的相互关系。在满足某些特定条件的情形下, 完全的国有化或完全的私有化都不能最大化社会总体福利或政府支付。另外, 关税的降低并不一定会减少国企的均衡产出。在模型结果的模拟部分, 我们还分别给出了最大化社会总体福利目标以及政府政策目标的最优的国有股份比重的具体值及其与整个社会的就业压力以及国企的相对生产效率三者之间的相互关系。

我们的模拟分析中有一个值得注意的结论: 政府的政策目标与社会总体福利目标是有冲突的, 而这种冲突程度的大小又取决于国企本身的相对生产成本。当国企的生产成本低于某一临界值时, 从社会总体福利目标与政府政策目标两个角度来看, 最优的国有股份比重都与就业压力的大小呈同向变化。此时, 政府的政策目标与社会总体福利目标具有相对一致性, 因而, 要尽可能地协调政府政策目标与社会总体福利目标, 就必须提高国有企业的相对生产效率。从这个意义上考虑, 我们的结果可以在一定程度上为现行的国有企业的所有制结构改革提供一定的理论借鉴。

附录 A:

引理 2.1 的证明:

由(5)、(6)分别对 Q_1 和 Q_2 求导 根据假设 3 二阶条件成立 可得:

$$P''(1 - \alpha + \alpha\beta)Q_1 + P'(2 - \alpha + 2\alpha\beta) - C_1' < 0$$

$$\text{即: } K < 0 \quad (\text{A.1})$$

$$P''Q_2 + 2P' - C_2' < 0 \quad (\text{A.2})$$

$$\Delta > 0$$

又由(5)、(6) 我们可推出以下等式:

$$\frac{\partial R_1}{\partial Q_2} = - \frac{P''(1 - \alpha + \alpha\beta)Q_1 + P'(1 + \alpha\beta)}{P''(1 - \alpha + \alpha\beta)Q_1 + P'(2 - \alpha + 2\alpha\beta) - C_1'} = - \frac{H}{K} \quad (\text{A.3})$$

$$\frac{\partial R_2}{\partial Q_1} = - \frac{P''Q_2 + P'}{P''Q_2 + 2P' - C_2'} \quad (\text{A.4})$$

又因为 $-1 < \frac{\partial R_1}{\partial Q_2} < 0$ 且 $-1 < \frac{\partial R_2}{\partial Q_1} < 0$ 有:

$$-1 < - \frac{P''Q_2 + P'}{P''Q_2 + 2P' - C_2'} < 0 \quad (\text{A.5})$$

由(A.2)式和(A.5)式可知 $P''Q_2 + P' < 0$ 并且有:

$$|P''E_2 + P'| < |P''E_2 + 2P' - C_2'| \quad (\text{A.6})$$

现在我们已经证明了 $P''Q_2 + 2P' - C_2' < 0$, $P''Q_2 + P' < 0$ 以及 $\Delta > 0$ 又因为 $(1 - \beta)P'Q_1 - \beta P < 0$. 故有结论: $\frac{\partial E_1}{\partial \alpha} > 0$, $\frac{\partial E_2}{\partial \alpha} < 0$.

又由(9)与(10)以及(A.6) 我们容易得出: $\frac{\partial E_1}{\partial \alpha} > \left| \frac{\partial E_2}{\partial \alpha} \right|$, 从而有结论 $\frac{\partial E}{\partial \alpha} > 0$.

命题 2.1 的证明:

我们记 $\frac{\partial W}{\partial Q_1} \Big|_{Q_1=E_1} = \frac{\partial W^E}{\partial E_1}$ 则有:

$$\frac{\partial W^E}{\partial E_1} = P(1 + R_2') - C_2' - C_2'R_2' = (P - C_1') + R_2'(P - C_2') \quad (\text{A.7})$$

从而 $\frac{\partial W^E}{\partial \alpha} = \frac{\partial W^E}{\partial E_1} \frac{\partial E_1}{\partial \alpha}$ 由于 $\frac{\partial E_1}{\partial \alpha} > 0$, $\frac{\partial W^E}{\partial \alpha}$ 与 $\frac{\partial W^E}{\partial E_1}$ 的符号相同。

由于 $R_2' < 0$ 又由方程(6) $P - C_2' = -P'Q_2 > 0$ 则 $R_2'(P - C_2') < 0$.

当 $\alpha = 1$ 时 如果 $(P - C_1') < -R_2'(P - C_2')$ 则 $\frac{\partial W^E}{\partial E_1} = (P - C_1') + R_2'(P - C_2') < 0$,

即有 $\frac{\partial W^E}{\partial \alpha} \Big|_{\alpha=1} < 0$.

当 $\alpha = 0$ 时 如果 $(P - C_1') > -R_2'(P - C_2') > 0$ 则 $\frac{\partial W^E}{\partial E_1} = (P - C_1') + R_2'(P - C_2') >$

0, 从而有 $\left. \frac{\partial W^E}{\partial \alpha} \right|_{\alpha=1} > 0$ 。

命题 2.2 的证明：

将 $\alpha = 1$ 代入一阶条件(5)式, 得到：

$$\beta P' E_1 + (1 + \beta)P - C_1' = 0$$

代入(11)式, 得到：

$$\left. \frac{\partial G^E}{\partial E_1} \right|_{\alpha=1} = (C_1' - C_2' - \beta P)R_2'$$

所以, 当 $\alpha = 1$ 时, 如果满足条件 $C_1' > C_2' + \beta P$, 我们有：

$$\left. \frac{\partial G^E}{\partial E_1} \right|_{\alpha=1} < 0$$

因而有 $\left. \frac{\partial G^E}{\partial \alpha} \right|_{\alpha=1} < 0$ 。

又将 $\alpha = 0$ 代入一阶条件(5)式, 得到：

$$P' Q_1 + P - C_1' = 0 \text{ 则 } P - C_1' = -P' Q_1 > 0$$

将这一条件代入(11)式, 则有：

$$\left. \frac{\partial G^E}{\partial E_1} \right|_{\alpha=0} = (P - C_1')(1 - \beta R_2') + \beta C_1' + (P - C_2')R_2'$$

又因为 $1 - \beta R_2' > 1$, 则当 $\alpha = 0$ 时, 如果满足条件 $P - C_1' > -(P - C_2')R_2'$, 则 $\left. \frac{\partial G^E}{\partial \alpha} \right|_{\alpha=0} > 0$ 。

附录 B：

引理 4.1 的证明：

由二阶条件以及系统的稳定性条件, 可得：

$$\Delta_1 < 0 \tag{B.1}$$

$$\Delta_{12} > 0 \tag{B.2}$$

$$\Delta_{23} > 0 \tag{B.3}$$

$$K_1 < 0 \tag{B.4}$$

$$P''E_2 + 2P' - C_2'' < 0 \tag{B.5}$$

$$P''E_3 + 2P' - C_3'' < 0 \tag{B.6}$$

由一阶条件, 可以求得：

$$\frac{\partial R_2}{\partial Q_1} = -\frac{P''R_2 + P'}{P''R_2 + 2P' - C_2''} \frac{\partial R_3}{\partial Q_1} = \frac{\partial R_3}{\partial Q_2} = -\frac{P''R_3 + P'}{P''R_3 + 2P' - C_3''} \tag{B.7}$$

又因为 $-1 < \frac{\partial R_2}{\partial Q_1} < 0$ 以及 $-1 < \frac{\partial R_3}{\partial R_1} < 0$, 可得：

$$P''E_2 + P' < 0 \tag{B.8}$$

$$P''E_3 + P' < 0 \tag{B.9}$$

由(B.1)、(B.3)两式可知, $\frac{\partial E_1}{\partial \alpha}$ 与 $(1 - \beta)P'E_1 - \beta P + P'E_3$ 的符号相反, $(1 - \beta)P'E_1 - \beta P + P'E_3$ 的符号显然为负, 因此有:

$$\frac{\partial E_1}{\partial \alpha} > 0$$

由(B.8)式, $P''E_2 + P' < 0$, 又因为 $P' - C_3'' < 0$, 得到:

$$\frac{\partial E_2}{\partial \alpha} > 0$$

同理可得:

$$\frac{\partial E_3}{\partial \alpha} > 0$$

并且有:

$$\begin{aligned} \frac{\partial E}{\partial \alpha} &= \frac{\partial E_1}{\partial \alpha} + \frac{\partial E_2}{\partial \alpha} + \frac{\partial E_3}{\partial \alpha} \\ &= \frac{1}{\Delta_1} [(1 - \beta)P'E_1 - \beta P + P'E_3] (P' - C_2'')(P' - C_3'') > 0 \end{aligned}$$

故总产出随 α 增加而增加。

命题 4.1 的证明:

由(36)式, 令 $t = 0$, 得到:

$$\left. \frac{\partial W^E}{\partial \alpha} \right|_{t=0} = (P - C_1' - P'R_3(1 + \frac{\partial R_3}{\partial E_1})) \frac{\partial E_1}{\partial \alpha} + [P - C_2' - P'R_3(1 + \frac{\partial R_3}{\partial E_2})] \frac{\partial E_2}{\partial \alpha} \quad (\text{B.10})$$

又因为 $-1 < \frac{\partial R_3}{\partial E_1} < 0$, $-1 < \frac{\partial R_3}{\partial E_2} < 0$, 又因为 $\frac{\partial E_1}{\partial \alpha} > 0$, $\frac{\partial E_2}{\partial \alpha} < 0$, 从而(B.10)式的后一项符号为负。

由一阶条件, $P - C_3' - t = -P'E_3$, 且 $0 < 1 + \frac{\partial R_3}{\partial E_1} < 1$, 所以:

(1) 当 $\alpha = 1$ 时, 如果 $P - C_1' < 0$, 且 $|P - C_1'| > (P - C_3')(1 + \frac{\partial R_3}{\partial E_1})$, 则(B.10)前一

项的符号为负, 这样, 我们有: $\left. \frac{\partial W^E}{\partial \alpha} \right|_{\alpha=1} < 0$ 。

(2) 因为 $\Delta_{23} = (P''E_2 + P')(P' - C_3'') + (P' - C_2'')(P' - C_3'') + (P''E_3 + P')(P' - C_2'')$
 $> (P''E_2 + P')(P' - C_3'')$ (B.11)

由(31)与(32)式可得:

$$\frac{\partial E_1}{\partial \alpha} > \left| \frac{\partial E_2}{\partial \alpha} \right| \quad (\text{B.12})$$

由(B.7), $\frac{\partial R_3}{\partial E_1} = \frac{\partial R_3}{\partial E_2}$, 代入(B.10)得到:

$$\left. \frac{\partial W^E}{\partial \alpha} \right|_{t=0} = [P - C_1 - P' R_3 (1 + \frac{\partial R_3}{\partial E_1})] \frac{\partial E_1}{\partial \alpha} + [P - C_2 - P' R_3 (1 + \frac{\partial R_3}{\partial E_1})] \frac{\partial E_2}{\partial \alpha}$$

则当 $\alpha = 0$ 时, 如果满足 $C_1 < C_2$, 则有: $\left. \frac{\partial W^E}{\partial \alpha} \right|_{\alpha=1} > 0$ 。

命题 4.2 的证明:

(1) 令 $\alpha = 1$, 由一阶条件(25)可得:

$$P' (\beta E_1 - R_3) = - [(1 + \beta)P - C_1] \quad (B.13)$$

代入(37)与(38)式, 有:

$$\left. \frac{\partial G^E}{\partial E_1} \right|_{\substack{t=0 \\ \alpha=1}} = \frac{\partial R_3}{\partial E_1} P' (\beta E_1 - R_3) = - \frac{\partial R_3}{\partial E_1} [\beta P + (P - C_1)]$$

$$\left. \frac{\partial G^E}{\partial E_2} \right|_{\substack{t=0 \\ \alpha=1}} = - (1 + \frac{\partial R_3}{\partial E_2}) [\beta P + (P - C_1)] + P - C_2$$

当 $\beta P + (P - C_1) < 0$ 时, $\left. \frac{\partial G^E}{\partial E_1} \right|_{\substack{t=0 \\ \alpha=1}} < 0$, $\left. \frac{\partial G^E}{\partial E_2} \right|_{\substack{t=0 \\ \alpha=1}} > 0$, 又由引理 4.1, $\frac{\partial E_1}{\partial \alpha} > 0$, $\frac{\partial E_2}{\partial \alpha} < 0$, 可

得:

$$\left. \frac{\partial G^E}{\partial \alpha} \right|_{\substack{t=0 \\ \alpha=1}} < 0$$

即 实行完全的贸易自由化后, $\alpha = 1$ 不是政府的最优选择。

(2) 令 $\alpha = 0$, 由一阶条件(25)、(26)、(27)式可得:

$$-(P - C_1) = P' E_1 < 0$$

$$-(P - C_2) = P' E_2 < 0$$

$$-(P - C_3 - t) = P' E_3 < 0$$

将上述条件代入(37)与(38)式, 得到:

$$\left. \frac{\partial G^E}{\partial E_1} \right|_{\substack{t=0 \\ \alpha=0}} = -\beta \frac{\partial R_3}{\partial E_1} (P - C_1) + (1 + \frac{\partial R_3}{\partial E_1}) (P - C_3) + (1 - \beta) (P - C_1) + \beta P > 0 \quad (B.14)$$

$$\left. \frac{\partial G^E}{\partial E_2} \right|_{\substack{t=0 \\ \alpha=0}} = -\beta \frac{\partial R_3}{\partial E_2} (P - C_1) + (1 + \frac{\partial R_3}{\partial E_2}) (P - C_3) - \beta (P - C_1) + (P - C_2) \quad (B.15)$$

(B.15) 式的符号是不确定的。

由(B.7)式 $\frac{\partial R_3}{\partial E_1} = \frac{\partial R_3}{\partial E_2}$, 如果有条件 $P - C_1 + \beta P > P - C_2$, 则有:

$$\left. \frac{\partial G^E}{\partial E_1} \right|_{\substack{t=0 \\ \alpha=0}} > \left| \left. \frac{\partial R_3}{\partial E_2} \right|_{\substack{t=0 \\ \alpha=0}} \right|$$

又由(B.12)式 $\frac{\partial E_1}{\partial \alpha} > \left| \frac{\partial E_2}{\partial \alpha} \right|$, 所以有:

$$\left. \frac{\partial G^E}{\partial \alpha} \right|_{\substack{t=0 \\ \alpha=0}} > 0$$

即 $\alpha = 0$ 不是政府的最优选择。

表 1 最优的国有股份比重 α_w^* 随 β 变化的情况

β	$\alpha^* (k = 12)$	$\alpha^* (k = 16)$	$\alpha^* (k = 24)$	$\alpha^* (k = 32)$
0.02	0.851703	0.796813	0.688976	0.583658
0.06	0.855131	0.790514	0.667939	0.553506
0.10	0.858586	0.784314	0.648148	0.526316
0.14	0.862069	0.778210	0.629496	0.501672
0.18	0.865580	0.772201	0.611888	0.479233
0.22	0.869121	0.766284	0.595238	0.458716
0.26	0.872690	0.760456	0.579470	0.439883
0.30	0.876289	0.754717	0.564516	0.422535
0.34	0.879917	0.749064	0.550314	0.406504
0.38	0.883576	0.743494	0.536810	0.391645
0.42	0.887265	0.738007	0.523952	0.377834
0.46	0.890985	0.732601	0.511696	0.364964
0.50	0.894737	0.727273	0.500000	0.352941

表 2 最优的国有股份比重 α_w^* 随 k 变化的情况

k	$\alpha^* (\beta = 0.05)$	$\alpha^* (\beta = 0.10)$	$\alpha^* (\beta = 0.15)$	$\alpha^* (\beta = 0.20)$
10	0.886076	0.897436	0.909091	0.921053
12	0.854271	0.858586	0.862944	0.867347
14	0.822943	0.820895	0.818859	0.816832
16	0.792079	0.784314	0.776699	0.769231
18	0.761671	0.748792	0.736342	0.724299
20	0.731707	0.714286	0.697674	0.681818
22	0.702179	0.680751	0.660592	0.641593
24	0.673077	0.648148	0.625000	0.603448
26	0.644391	0.616438	0.590810	0.567227
28	0.616114	0.585586	0.557940	0.532787
30	0.588235	0.555556	0.526316	0.500000
32	0.560748	0.526316	0.495868	0.468750

表 3 最优的国有股份比重 α_G^* 随 β 变化的情况

β	$\alpha^* (k = 12)$	$\alpha^* (k = 16)$	$\alpha^* (k = 24)$	$\alpha^* (k = 32)$
0.02	0.857277	0.809464	0.715495	0.623669
0.06	0.871722	0.827668	0.743813	0.665188
0.10	0.886076	0.845070	0.769231	0.700637
0.14	0.900398	0.861837	0.792353	0.731477
0.18	0.914737	0.878099	0.813636	0.758742
0.22	0.929137	0.893966	0.833427	0.783182
0.26	0.943636	0.909526	0.851999	0.805356
0.30	0.958269	0.924855	0.869565	0.825688
0.34	0.973067	0.940018	0.886299	0.844506
0.38	0.988061	0.955068	0.902340	0.862069
0.42	1.003277	0.970055	0.917804	0.878582
0.46	1.018743	0.985020	0.932787	0.894210
0.50	1.034483	1.000000	0.947368	0.909091

表 4 最优的国有股份比重 α_G^* 随 k 变化的情况

k	$\alpha^* (\beta = 0.05)$	$\alpha^* (\beta = 0.10)$	$\alpha^* (\beta = 0.15)$	$\alpha^* (\beta = 0.20)$
10	0.891065	0.907424	0.924129	0.941227
12	0.868121	0.886076	0.903979	0.921927
14	0.845502	0.865299	0.884599	0.903566
16	0.823199	0.845070	0.865945	0.886076
18	0.801208	0.825369	0.847977	0.869397
20	0.779520	0.806175	0.830658	0.853474
22	0.758131	0.787468	0.813953	0.838257
24	0.737034	0.769231	0.797831	0.823699
26	0.716222	0.751445	0.782261	0.809760
28	0.695691	0.734095	0.767216	0.796399
30	0.675435	0.717164	0.752668	0.783582
32	0.655447	0.700637	0.738595	0.771277

参考文献

- [1] Bai, Chongen, Li, David, Tao, Zhigang and Wang, Yijiang, A Multitask Theory of State Enterprise Reform, *Journal of Comparative Economics*, 2000, Vol.28, pp716 ~ 738.
- [2] Fraja, G. D. and Delbono, F., Game Theoretic Models of Mixed Oligopoly, *Journal of Economic Surveys*, 1989, Vol.4, No.1, pp1 ~ 17.
- [3] Zhang, Anming and Zhang, Yiming, Stability of a Cournot-Nash Equilibrium: The Multiproduct Case, *Journal of Mathematical Economics*, 1996, Vol.26, pp441 ~ 462.
- [4] Zhang, Weiyong, China's SOE Reform: A Corporate Governance Perspective, Working Paper Series, Mimeo, 1998.
- [5] 林毅夫等:《中国国有企业改革的回顾与展望》,北京大学中国经济研究中心讨论稿,2000年。
- [6] 王梦奎:《中国经济转轨二十年》,外文出版社1999年版。
- [7] 吴敬琏:《当代中国经济改革:战略与实施》,上海远东出版社1999年版。
- [8] 平新乔:《论国有经济比重的内生决定》,《经济研究》,2000年第7期。
- [9] 王红领,李稻葵,雷鼎鸣:《政府为什么会放弃国有企业的产权》,《经济研究》,2001年第8期。
- [10] 林毅夫,刘培林:《自生能力和国企改革》,《经济研究》,2001年第9期。
- [11] 刘小玄:《中国转轨经济中的产权结构和市场结构》,《经济研究》,2003年第1期。
- [12] 陆铭:《为何改革没有提高国有企业的相对劳动生产率》,《经济学季刊》,2003年第2卷第4期。

金融中介发展与经济增长： 中国案例研究与国际比较^{*}

李广众 王美今

一、理论综述与问题的提出

戈德史密斯（1969）对 35 个国家的研究表明，在其考察期间内，各国的金融部门与经济之间存在着平行发展的关系，但他同时也指出，“目前我们仍无法自信地说明这两者之间的因果关系方向”。此后，大量的金融发展文献在理论上清楚地说明，一个运行状况良好的金融部门能够动员储蓄并将储蓄有效地配置到高效率的生产性部门，对企业生产性技术选择以及证券市场外在性等方面的作用也十分明显，从而将对经济增长产生促进作用。

Patrick（1996）的研究为考察金融发展与经济增长之间的动态因果关系提供了一个有用的参考框架。他认为金融发展存在着“供给导向型的金融发展”与“需求导向型的金融发展”两种模式。在前一种模式中，由于金融部门的发展先于对金融服务的需求，因此金融部门在动员传统部门的资源并将之用于具有增长趋势的现代部门方面发挥着基础性作用，金融发展与经济增长之间的相互作用方向是：金融发展促进经济增长；而在后一种模式中，只有经济发展到达一定的阶段与水平之后，经济部门对金融部门产生了新的需求，这时候金融部门才开始发展，此时金融发展与经济增长之间的相互作用方向是“经济增长促进了金融发展”。

因此，金融发展与经济增长之间的因果关系及其作用机制问题一直是理论界研究的重要论题之一，对其研究则主要采用：①利用传统的格兰杰因果检验方法（Granger Causality Test）对金融发展指标与经济增长指标的时间序列数据进行分析；②在 Barro 式跨国增长方程中引入反映金融发展水平的变量来考察金融对经济增长的作用。

上述两种方法中，传统的格兰杰因果分析不可避免地会遇到数据的平稳性问题，而通过在 Barro 跨国增长回归方程中引入金融发展因素来考察二者之间的关系，不仅无法说明二者的因果作用方向，而且可能面临控制变量选择的合理性与完整性问题。作为此方面研究的代表人物之一，Levine（1997）在对自己以往工作的评价中曾指出：“运用该方法将受

^{*} 本研究得到中山大学□川优秀青年研究基金资助，在此表示感谢。同时，在本文的研究过程中，中山大学岭南学院陈广汉教授、陈平教授提供了许多有价值的参考意见，香港岭南大学冉齐鸣教授提供了许多宝贵的研究资料，在此一并表示感谢。原载：《统计研究》，2003 年第 1 期。

到度量、统计以及概念上的影响……同时跨国回归方程并没有说明变量之间的因果关系……因此，分析必须还在时间序列分析的基础上进一步展开。”

就我国的情形而言，迄今为止对中国经济增长问题的研究大多强调实际因素的作用，如转轨时期我国不断上升的进出口、持续流入的外资、产业结构升级因素等等，对金融部门在经济增长中的作用并没有给予足够的重视。谈儒勇（1999）利用中国近年来的季度数据对金融发展与经济增长之间的关系进行了实证分析，但该文同时也指出，研究并没有能够说明金融发展与经济增长之间的因果关系。另一方面，该研究也无法揭示金融中介发展的不同方面对经济增长的作用机制。

为此，本文拟针对金融发展理论中金融中介发展与增长之间的关系问题联系我国实际进行研究。本文具体安排如下：第二部分给出了本文的实证研究框架；第三部分给出本文的模型设计及相关估算与假设检验；第四部分是模型测算结果分析及其国际比较；最后是本文的结束语。

二、实证研究框架：多变量协整分析

本文分析主要在 AK 模型和现代金融发展理论基础上展开。具体实证估算则借助 VAR 模型，利用 Johansen 极大似然估计法对 VAR 系统中变量之间的协整与因果关系进行估计与检验。

为了分析的简便以及不失一般性，我们假定讨论的变量只涉及服从 I(1) 过程的非平稳变量。k 阶 VAR 模型可以表述为：

$$Y_t = \sum_{i=1}^n \Pi_i Y_{t-i} + U_t = \Pi_1 Y_{t-1} + \Pi_2 Y_{t-2} + \dots + \Pi_k Y_{t-k} + U_t, U_t \sim \text{IID}(0, \Omega) \quad (1)$$

其中： Π_1, \dots, Π_k 都是 $N \times N$ 阶参数矩阵， U_t 是 $N \times 1$ 阶随机误差列向量。 Ω 是 $N \times N$ 阶方差协方差矩阵。对最大滞后期数 n 的选择主要根据赤池（Akaike）信息准则（AIC）或者施瓦茨（Schwartz）准则（SC）进行。

若 VAR 模型中的非平稳变量存在协整关系，我们就可以在 VAR 模型的基础上经过协整变换建立向量误差修正模型，表示为：

$$\Delta Y_t = \Gamma_1 \Delta Y_{t-1} + \Gamma_2 \Delta Y_{t-2} + \dots + \Gamma_{k-1} \Delta Y_{t-k+1} + \Pi Y_{t-k} + U_t \quad (2)$$

其中，压缩矩阵 $\Pi = -I + \Pi_1 + \dots + \Pi_k$ 。

此时，若 ΠY_{t-k} 是非平稳的，则意味着 Y_t 中的分量之间不存在协整关系。相反，如果 ΠY_{t-k} 是平稳的，则 Y_t 中的分量之间一定存在协整关系。当协整向量个数 r 在 $(0, N)$ 之间时， Π 可以分解为：

$$\Pi = \alpha \beta \quad (3)$$

其中： α 和 β 都是 $N \times r$ 阶矩阵。 β 为协整参数矩阵，其每一列都是一个协整向量。 α 为调整系数矩阵，其中的每个元素表示相应每个误差修正项对差分的被解释变量的调整速度。

为了解决多变量之间协整关系的识别问题, Johansen (1991) 认为必须对协整向量施加 $r \times r$ 个恰好识别约束 (just-identifying restrictions)。Pesaran 与 Shin (1994) 则证明了: 由于与任何恰好识别约束组合相联系的极大对数似然值都一样, 这使人们仍然无法在不同的约束组合之间进行选择, 因此上述恰好确认识别约束是不充分的。为此, 他们二人建议通过对包含 k 个过度识别约束在内的 $r^2 + k$ ($k \geq 1$) 个约束条件进行检验的方法来解决识别问题。每一个协整向量中至少都必须包括 r 个约束条件, 这些约束条件中还必须包括一个正规化约束。这时候对约束条件的提出就不能依靠统计方法来实现, 而必须在经济理论基础之上予以恰当的假设, 再通过统计分析来检验。

除此之外, 威肯斯 (Wickens, 1996) 的研究还表明, 协整方程对应的调整系数的符号及其显著性也是一个重要的帮助识别的依据。在 VECM 模型中, 为反映对偏离长期均衡关系的情况进行调整的调整系数必须在符号上为负数, 同时统计检验显著。

在因果关系分析方面, Johansen et al. (1992), Hall 与 Milne (1994) 说明了在一个存在协整关系的 VAR 系统中对变量的弱外生性 (weakly-exogenous) 进行检验可以等同于对变量之间长期因果关系的检验。根据 Demetriades 与 Hussein (1996) 的研究, 我们认为, 对这种长期因果关系的检验可以通过对 VECM 中协整方程对应调整系数的参数显著性检验来实现。

三、实证模型设计与数据来源

1. 实证模型设计

由 AK 模型, 我们可以得到 $LYP = F(LKP)$, 而根据现代金融发展理论, 则可以得到 $FD = (LYP, R)$, $FD = (BANK, PRI)$ 。

$LYP, LKP, BANK, PRI, R$ 分别为实际人均 GDP 的自然对数、实际人均资本存量的自然对数、金融中介发展规模指标、金融中介发展效率指标以及实际利率。本文对以上变量之间协整关系与因果关系的讨论主要就是通过构建包括这五个变量的 VAR 系统来实现。

经过协整变换建立向量误差修正模型可以表示为:

$$\Delta Y_t = \Gamma_1 \Delta Y_{t-1} + \Gamma_2 \Delta Y_{t-2} + \dots + \Gamma_{k-1} \Delta Y_{t-k+1} + \prod Y_{t-k} + U_t \quad (4)$$

其中: $Y_t = (BANK, LYP, LKP, PRI, R)$ 。具体分析可以根据变量的数据生成过程以及我国现实, 考虑在基本模型中加入常数项、虚拟变量或将实际利率视为弱外生变量。

根据理论分析, 我们认定当变量之间存在三个协整关系时, 可以将其中一个正规化为反映金融中介规模发展方程; 一个正规化为增长方程, 另一个正规化为金融中介效率方程。具体可以表示为:

$$\begin{pmatrix} \Delta BANK_t \\ \Delta LYP_t \\ \Delta LKP_t \\ \Delta PRI_t \\ \Delta R_t \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} \\ \alpha_{51} & \alpha_{52} & \alpha_{53} \end{pmatrix}_{5 \times 3} \begin{pmatrix} 1 & \beta_{12} & \beta_{13} & \beta_{14} & \beta_{15} \\ \beta_{21} & 1 & \beta_{23} & \beta_{24} & \beta_{25} \\ \beta_{31} & \beta_{32} & \beta_{33} & 1 & \beta_{35} \end{pmatrix}_{5 \times 3} \begin{pmatrix} BANK_{t-1} \\ LYP_{t-1} \\ LKP_{t-1} \\ PRI_{t-1} \\ R_{t-1} \end{pmatrix}_{5 \times 1} \quad (5)$$

撇开正规化约束 (normalization restrictions) 使 $\beta_{11} = \beta_{22} = \beta_{34} = 1$ 这三个假设条件外, 根据经济理论及其我国现实, 我们可以认为资本存量、金融中介效率变动对金融中介规模发展没有直接的作用, 因此 $\beta_{13} = \beta_{14} = 0$ 。根据 AK 模型, 若资本存量具有常数报酬, 则假设 $\beta_{23} = 1$, 当数据拒绝了同一性 (homogeneity) 时, 计算结果有可能不等于 1。这里当 β_{24} , β_{25} 不等于 0, 即意味着金融中介效率、实际利率对资本存量的产出效率具有一定影响。最后, 假定我国金融中介效率主要受到金融中介规模的影响, 因此 $\beta_{32} = \beta_{33} = 0$ 。所有以上为达到识别目的而进行的假设都可以获得检验。

2. 数据来源

对资本存量的估计, 我们根据段宾与张瑞 (1993) 的研究, 采用张军扩 (1991) 的估算方法, 即以美国经济学家估计的我国 1953 年资本总额为 2145 亿元为基准, 然后估计出我国投资转化为资本的时滞分布为: 约 26% 的投资在当年转化为资本, 约 26% 的投资在第二年转化为资本, 约 20% 的投资在第三年转化为资本, 约 12% 的投资在第四年转化为资本, 约 9% 的投资在第五年转化为资本, 约 7% 的投资在第六年转化为资本。这样把每年的投资额按时滞系数均摊到各有关年份中, 估计出我国 1952 ~ 1999 逐年的资本总额。

对金融发展指标的定义, 根据李广众 (2002) 的研究, 金融中介规模指标 BANK 定义为金融机构贷款占 GDP 的比重, 金融中介发展效率指标 PRI 定义为非国有经济获得银行贷款占银行资金运用总额的比重。对指标的计算, 我们主要利用《中国金融年鉴》与《新中国统计资料五十年》中全国金融机构信贷资金平衡表中有关资金运用方面的统计予以计算。对 PRI 的估算, 我们结合北京大学中国经济研究中心 (2000) 的研究, 1978 ~ 1995 年的数值等于城镇集体企业及个体工商户贷款在国家银行资金运用的比重, 1995 年之后的数值等于乡镇企业贷款、私营企业及个体贷款、“三资”企业贷款获得的贷款占国家银行资金运用的比重。

表 1 列出了有关金融中介发展指标的统计描述。从表 1 中我们可以发现, 改革开放以后我国金融中介发展指标 BANK, PRI 均有较大幅度增长, 但其变动幅度也有所上升。笔者认为这种情况的出现是合理的, 它反映了我国金融中介的发展, 同时还从另一个侧面说明了市场因素已经在我国金融中介资金配置过程中起越来越大的作用。

表 1 金融中介发展指标的统计描述

	时期	最小值	最大值	均值	标准差
Bank	1952 ~ 1999	0.159	1.144	0.594	0.249
Pri	1952 ~ 1999	0.003	0.104	0.034	0.029
Bank	1952 ~ 1978	0.159	0.665	0.428	0.140
Pri	1952 ~ 1978	0.003	0.028	0.012	0.007
Bank	1979 ~ 1999	0.505	1.144	0.807	0.188
Pri	1979 ~ 1999	0.028	0.104	0.062	0.020

为了便利在国际比较中看待中国的金融中介发展水平，表 2 与表 3 给出了 King 与 Levine (1993a) 对近 80 个国家 1960 ~ 1989 年数据的分析结果以及 King 与 Levine (1993b) 对 1985 年跨国样本分析的结果。

表 2 1969 ~ 1989 年各国金融中介发展指标的统计分析

	最小值	最大值	均值	标准差
PRIY	0.037	1.119	0.256	0.185
Private	0.176	0.920	0.598	0.172

表 3 1985 年跨国样本金融中介发展指标的统计分析

	十分富裕国家	富裕国家	贫穷国家	十分贫穷国家
PRIY	0.53	0.31	0.20	0.13
Private	0.71	0.58	0.47	0.37

注：国家富裕程度的划分标准为：十分富裕国家：人均实际 GDP > 4998 美元；富裕国家：4998 美元 > 人均实际 GDP > 1161 美元；贫穷国家：1161 美元 > 人均实际 GDP > 391 美元；十分贫穷国家：391 美元 > 人均实际 GDP。

在采用该两项指标来衡量中国金融体系的发展程度的基础上，我们进行的国际比较说明，与瑞士国际管理发展学院 (IMD) 的研究结果相似^①，中国金融体系在金融机构与金融工具的数量和规模方面处于较高水平，而在金融体系的质量和效率方面则排名很低。

四、模型的测算结果及其国际比较

1. 数据的平稳性分析

在进行协整分析之前，我们首先必须对 VAR 系统所涉及变量的数据生成过程的平稳性进行检验。表 4 报告了根据 Augmented Dickey-Fuller 方法 (ADF 检验法) 与 Phillips-Perron 方法 (PP 检验法) 对变量平稳性检验的结果。结果说明，所涉及变量除实际利率外均为 I (1) 过程的非平稳变量，并没有出现 I (2) 过程的非平稳变量。

^① 根据 IMD 研究，我国金融体系在反映金融机构和金融工具的数量和规模的指标往往排名比较靠前，大致在其研究的 47 个国家和地区中排在中上位置，而在反映金融体系质量与效率方面则排名比较靠后，大致在倒数 5 名以内。资料来源：干杏娣，徐明祺：《浴火重生：入世后中国金融的结构变革》，上海社会科学院出版社，高等教育出版社 2001 年版。

表 4 变量的单位根检验

变量	ADF 检验法		PP 检验法	
	ADF 检验值	临界值	PP 检验值	临界值
LYP	- 2.138	- 3.5088	- 1.693	- 3.5066
Δ LYP	- 5.172 **	- 3.5112	- 5.944 **	- 3.5088
LKP	- 1.180	- 3.5088	- 1.472	- 3.5066
Δ LKP	- 5.438 **	- 3.5112	- 9.469 **	- 3.5088
R	- 2.703 **	- 1.948	- 3.710 **	- 1.9478
BANK	- 1.116	- 2.9256	0.608	- 2.9241
Δ BANK	- 3.885 **	- 2.9271	- 4.336 **	- 2.9256
PRI	1.700	- 2.9256	- 2.118	- 2.9241
Δ PRI	- 5.553 **	- 2.9271	- 7.344 **	2.9256

注：从经济意义出发，对变量的单位根检验方程中分别采用包括了截距和趋势项、包括截距以及不包括二者三种情况进行，因此其临界值也有所不同；以上提供的为显著性水平为 5% 的临界值；加 * * 表示通过显著性水平为 5% 的单位根检验。

2. 协整分析结果

利用 Hansen 与 Juselius (2001)，我们分别对将实际利率视为弱外生变量与将实际利率视为内生变量的模型进行估计。结果如下：

模型 1：将实际利率视为弱外生变量，加入虚拟变量 D（改革开放之前取 0，改革开放之后取 1）。见表 5。

表 5 模型 1 的 Johansen 检验结果

I (1) ANALYSIS				
Eigenv.	L-max	Trace	$H_0 : r$	p-r
0.6190	44.39 **	88.71 **	0	4
0.4559	28.00 **	44.32 **	1	3
0.2792	15.06 **	16.32 **	2	2
0.0269	1.26	1.26	3	1

表 5 的分析结果支持存在 3 个协整关系。根据金融发展理论与中国现实，我们对协整向量施加限制的约束矩阵为（结果见表 6）：

$$H_1 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}; H_2 = \begin{pmatrix} 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}; H_3 = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

表 6 LR 检验、协整向量与相应的调整系数 (模型 1)

The LR test , CHISQ (1) = 0.15 , p-value = 0.69						
协整向量 :						
BANK	LYP	LKP	PRI	R	CONSTANT	
1.000	- 0.439	0.000	0.000	- 1.358	1.865	
0.000	1.000	- 1.000	- 28.075	- 10.496	1.462	
- 0.100	0.000	0.000	1.000	0.000	0.036	
调整系数 :						
ALPHA				T-VALUES	FOR	ALPHA
DBANK	- 0.164	0.020	1.318	- 1.757	1.715	1.256
DLYP	- 0.096	- 0.114	- 2.696	- 0.682	- 6.374	- 1.709
DLKP	- 0.230	- 0.080	- 5.028	- 1.167	- 3.167	- 2.262
DPRI	- 0.061	0.001	- 0.567	- 6.337	0.588	- 5.224

模型 2 : 为了方便国际比较, 我们对将实际利率视为内生变量, 并加入虚拟变量 D 的模型再次进行估计。见表 7。

表 7 模型 2 的 Johansen 检验结果

I (1) ANALYSIS				
Eigenv.	L-max	Trace	H0 : r	p-r
0.6533	48.73**	110.26**	0	5
0.4594	28.29**	61.53**	1	4
0.4216	25.19**	33.23**	2	2
0.1555	7.78	8.05	3	2
0.0059	0.27	0.27	4	1

注 : 加 * * 表示可以在 5% 的显著性水平上拒绝原假设。

检验结果同样支持存在 3 个协整关系。对协整向量施加同样约束的结果见表 8。

表 8 LR 检验、协整向量与相应的调整系数 (模型 2)

The LR test , CHISQ (1) = 0.28 , p-value = 0.59						
协整向量 :						
BANK	LYP	LKP	PRI	R	CONSTANT	
1.000	- 0.460	0.000	0.000	- 1.376	1.899	
0.000	1.000	- 1.000	- 29.500	- 9.915	1.197	
- 0.092	0.000	0.000	1.000	0.000	0.044	
调整系数 :						
ALPHA				T-VALUES	FOR	ALPHA
DBANK	- 0.135	0.039	1.718	- 1.547	2.967	1.704
DLYP	- 0.054	- 0.107	- 2.192	- 0.410	- 5.513	- 1.451
DLKP	0.038	- 0.066	- 0.473	0.193	- 2.242	- 0.207
DPRI	- 0.056	- 0.001	- 0.517	- 6.089	- 0.487	- 4.889
DR	0.212	0.025	3.714	3.056	2.386	4.631

两个模型的分析结果证明了以上计量结果具有较强的鲁棒性 (robust)。在协整方程 1 中, LYP 的系数分别为: 0.460, 0.439, R 的系数分别为: 1.376, 1.385; 在协整方程 2 中, PRI 的系数分别为 29.5, 28.075, R 的系数分别为: 9.914, 10.496; 在协整方程 3 中, BANK 的系数分别为 0.092, 0.100。所有结果满足 Wickens (1996) 的识别条件。结合 Hall 与 Milne (1994) 的研究, 我们认为模型 1 的分析较为强硬地支持了以下结论:

(1) 金融中介规模不构成经济增长的原因, 相反, 经济增长可以成为具有较高金融中介规模的原因。这一分析结果证明了在金融中介规模方面的确存在如 Robinson (1952) 所强调的情况, 即金融部门是随着实际经济部门的发展而发展的。

实际利率提高同样有利于金融中介规模的扩大。从系数大小的比较来看, 我们强调促进金融中介规模的扩大的主要原因在于经济发展。金融中介规模扩大不构成经济增长原因的本身说明了, 长时间以来我国金融中介规模的扩大对整体经济的作用还仅仅停留在一种被动适应的地位上。

(2) 金融中介效率则是促进经济增长的原因, 而经济增长不成为具有较高金融中介效率的原因。这一分析结果则符合内生经济增长理论对金融中介发展作用的认识, 即强调金融中介效率提高将有利于经济增长。

这一结果同时说明, 我们必须通过外部的改革措施来推动金融中介效率的提高, 希望通过经济发展本身带来金融中介在资金配置方面效率的提高是不合适的。表 6 与表 8 的分析还说明, 与 Díaz-Alejandro 与 Carlos (1985) 对拉丁美洲金融自由化实践的研究相类似, 中国金融中介效率的提高主要通过提高资本产出效率来影响经济增长。从系数大小的比较来看, 分析结果说明金融中介效率提高对经济增长的促进作用大于实际利率的作用。

(3) 在以上多变量 VAR 系统中, 对 AK 模型的检验是可以接受的。同时, 金融中介规模扩大有利于金融中介效率的提高, 但作用系数偏小。从长期来看, 金融中介各项贷款占 GDP 的比重上升 1%, 非国有经济获得贷款占各项贷款的比重才上升 0.1%。

3. 国际比较

为了更好地了解我国金融中介发展与经济增长之间的关系, 我们将上述测算结果与 Luintel 与 Khan (1999) 在分别将实际利率水平视为内生变量或弱外生变量的不同情况下对印度、韩国、菲律宾、斯里兰卡和泰国的同类研究结果进行比较。结果说明:

(1) 中国金融中介规模发展对其长期均衡关系的偏离进行调整的速度相对较慢 (大致为 0.15), 而根据 Luintel 与 Khan (1999) 的研究, 合理的调整速度应在 15% ~ 36% 之间。与印度的情况相近似, 笔者认为造成这一结果的主要原因在于政府对金融部门的过度控制与干预。

(2) 在产出增长方程方面, 从资本的产出弹性看, 与大多数国家表现不同的是, 中国的资本积累对经济增长表现出规模报酬不变的假设是可以接受的。

五、结论性评述

本文首次引入 VAR 系统中协整向量的确定与相应的因果关系分析方法研究了我国金融中介发展的不同方面与经济增长之间的关系。结合李广众 (2002) 的研究, 笔者认为整

个分析结果说明：

第一，我国金融中介发展的不同方面与经济增长之间的因果关系存在一定的不对称性，即经济增长是促进金融中介规模扩大的原因，而金融中介规模扩大并不是促进经济增长的原因；相反，金融中介效率是促进经济增长的原因，而经济增长本身并不是促进金融中介效率的原因。因此，要想从金融部门角度为经济增长提供支持，关键在于提高金融中介效率，加大金融中介对非国有经济的支持力度，而非简单地扩大金融中介的资金配置规模（尽管金融中介规模的扩大有利于投资数量的上升）。实现这一效率的提高，依靠经济发展本身还是行不通的，必须借助于外力（如金融体制改革、加强金融中介之间的竞争等）的作用来实现。

第二，金融中介效率指标对经济增长单方向上的因果关系及其与非国有工业增长之间的双向因果关系是同时存在的；而金融中介规模指标对经济增长单方向上的无因果关系及其与非国有工业增长的双向无因果关系也是同时存在的。这一结果意味着金融中介发展能否促进经济增长关键在于能否与以市场为主导的“非国有经济”形成良好的相互促进机制。在这一意义上说，从促进长期经济增长的角度出发，鼓励银行同业之间竞争、加强金融部门与非国有经济部门的经济联系也就显得尤为重要了。

第三，对金融中介发展与经济增长之间关系的协整分析还证明了，在一个实行利率管制的金融政策与金融中介垄断程度较高的市场结构下，金融中介发展的不同方面与经济增长之间无法形成双向的因果关系。在金融中介发展规模方面，研究支持 Robinson（1952）年提出的实际经济主导观点，而在金融中介效率方面则支持内生经济增长理论中所强调的金融发展促进观点。

参考文献

- [1] Demetriades, P.O. and Hussain, K.A., Does Financial Development Cause Economic Growth? Time Series Evidence from 16 Countries, *Journal of Development Economics*, 1996, Vol.51, No.3, pp313 ~ 328.
- [2] Díaz-Alejandro, Carlos F., Good-bye financial repression, hello financial crash, *Journal of Development Economics*, 1985, Vol.19, pp1 ~ 24.
- [3] Goldsmith, Raymond W., *Financial structure and development*. New Haven, CT: Yale University Press, 1969.
- [4] Hall, S.G., Milne, A., The relevance of p-star analysis to UK monetary policy, *The Economic Journal*, 1994, Vol.104, pp597 ~ 604.
- [5] Hansen, H., Juselius, K., CATS in RATS, Estima, Evanston, 2001.
- [6] King, Robert G. and Levine, Ross, Finance and Growth: Schumpeter Might Be Right, *Quarterly Journal of Economics* August, 1993, Vol.108, No.3, pp717 ~ 38.
- [7] King, Robert G. and Levine, Ross, Finance, Entrepreneurship, and Growth: Theory and Evidence, *Journal of Monetary Economics*, December, 1993, Vol.32, No.3, pp513 ~ 542.
- [8] Johansen, S., Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregression models, *Econometrica*, 1991, Vol.59, pp551 ~ 580.
- [9] Levine, Ross, Financial Development and Economic Growth: Views and Agenda, *Journal of Economic Literature*, 1997, pp688 ~ 726.

-
- [10] Robinson, Joan, *The generalization of the general theory, in The rate of interest, and other essays*. London: Macmillan, 1952, pp67 ~ 142.
- [11] Wickens, R. W. , *Interpreting cointegration vectors and common stochastic trends*, *Journal of Econometric*, 1996, Vol.74, pp255 ~ 271.
- [12] 段宾, 张瑞:《中国资本、资本效率、工资水平和技术进步状况的历史分析(1952~1990)》,《中国经济史研究》,1993年第4期。
- [13] 李广众:《金融中介发展与经济增长:多变量VAR系统研究》,《管理世界》,2002年第3期。
- [14] 谈儒勇:《中国的金融发展与经济增长关系的实证研究》,《经济研究》,1999年第10期。
- [15] 张军扩:《“七五”期间经济效益的综合分析》,《经济研究》,1991年第4期。

第三部分

国际贸易与投资

FDI 与广东技术进步关系的实证分析*

黄静波 付 建

一、FDI 与技术转移：实证文献举要

技术进步是经济增长的重要源泉。技术进步的途径主要分为两大方面：一是本国技术能力和水平的提升，其基本形式是技术开发和创新；二是技术获取或技术转移，即通过从国外直接引进先进技术、干中学（learning by doing）、国外 R&D 溢出、引进外资获取技术外溢等方式来提高东道国的技术水平，其中，吸引外国直接投资（FDI）是发展中国家普遍重视的促进技术进步的途径。

关于 FDI 的技术转移效果，国内外已有大量相关文献，其中相关的实证研究十分引人注目，但同时也令人迷惑，因为众多的实证研究并没有得出一致的结论。坎波斯和基诺什塔（Campos and Kinoshita, 2002）列举了一些结论不相同的相关研究^[5]。例如，萨尔茨（Saltz, 1992）曾指出 FDI 对经济增长有负面影响，而布隆斯敦等人（Blomstron et al., 1994）认为 FDI 对经济增长有积极作用。林辛克和莫里西（Lensink and Morrissey, 2001）则发现 FDI 对经济增长有积极作用，但不是完全稳固的。

有不少实证研究发现，FDI 有助于国际技术转移。布伦斯廷等人（Borensztein et al., 1998）对 OECD 国家流向 69 个发展中国家的 FDI 的数据进行分析，其结果表明，如果发展中国家有足够的人力资本的话，FDI 是向发展中国家技术扩散的有效途径，并对经济增长起积极作用。^[4]坎波斯和基诺什塔以转型国家为例，指出 FDI 在技术转移方面起到一种引擎的作用。^[5]

不少实证结果表明，进入发展中国家的跨国公司及其投资对东道国的出口都有重要影响，无论是对跨国公司在东道国的分支机构还是对东道国本身的企业，这种影响都是重要的^[10,11]。这些研究结果实际上揭示了外国的知识流入对本地的企业存在巨大的积极影响，而由于外国知识流入导致生产增加，从而有更多的出口。不过，这种效果通常是短期的，并且作用呈递减趋势。

跨国公司的 FDI 是向东道国进行技术转移的最重要的形式，FDI 本身涉及资本、技术和知识从跨国公司本国向东道国的流动，投资期间跨国公司使用更先进的技术提供了提高生产率的可能性，从而促进经济增长。在技术转移方面，跨国公司 FDI 具有一种“全套型”的性质，它不仅仅涉及技术转移的本身，还提供了使那些与被转移技术相关的各种技

* 本文为广东省哲学社会科学规划项目“广东省对外贸易发展的实证研究”（02c33）的部分成果。原载：《管理世界》，2004 年第 9 期。

术诀窍、管理经验、企业家能力配套转移的可能性。^[2]另一方面,跨国公司的 FDI 可能使技术和技术诀窍的使用更为有价值或更加低成本,特别是当这些技术本身是由这些跨国公司所开发的时候尤为如此。^[7]

一般认为,通过 FDI 发生技术转移的途径包括四种基本类型:垂直型关联、水平型关联、劳动力流动以及 R&D 的国际化。^[7]而从 FDI 发生技术转移的效果上看,FDI 促进东道国技术进步的渠道分为直接效应和间接效应两类:一是外资企业技术进步的直接效应,指外资企业相对于国内企业的要素生产率优势的提高。二是间接效应,指外资企业对国内企业的技术外溢。^[13]

里奇腾伯格等人 (Lichtenberg and van Pottlsberhe, 1996) 通过实证研究认为,内向流动的 FDI 的溢出效应并不是很明显。^[9]而凯勒 (Keller, 1999) 的研究表明,在 OECD 国家中,导致生产率增长的主要源泉不是来自国内,而是来自国外。对于在 R&D 投入比发达国家更少的发展中国家来说,国外 R&D 的溢出也就显得很重要了。^[8]鲍德温等人 (Baldwin et al., 1999) 认为,从产业层面上考察时,发现内向流动的 FDI 是存在溢出效应的。^[2]

一国自身的技术水平和人力资本的积累会直接影响到吸收国外先进技术的能力。外国技术和一国自身吸收能力存在一种协调关系,如果一国本身的吸收能力有限的话,一国从国外技术溢出中的获益是有限的。科尔和莫格登 (Coe and Moghadam, 1996) 认为,国内的 R&D 积累是决定一国生产率水平的关键因素。对于一个国家来讲,从国外获得技术是一国迅速赶上发达国家的捷径,但对于一国长期发展来讲,自身技术的发展才是关键。^[6]对于人力资本,伯纳比伯和斯皮尔格 (Benhabib and Spiegel, 1994) 实证了尼尔森和菲尔普斯 (Nelson and Phelps, 1966) 强调的一点:一国的人力资本水平会影响到该国吸收先进技术的能力。^[3]

另一方面,技术溢出还要受到技术供给方提供意愿的限制。对于一个外国投资者而言,在东道国采用何种技术受到以下几个方面的影响:一是上面提到的东道国的吸收能力,这点对于投资者来说有着正反两种效应。东道国的技术力量强时,投资者的先进技术可以在东道国运用,这样投资者会使用较先进的技术;同时东道国的技术能力强时,外国投资者采用先进技术存在潜在的竞争威胁。二是东道国对知识产权的保护程度。三是东道国市场的竞争程度。四是其他商业利益。

需要注意的是,在国家层面分析 FDI 在技术溢出中的作用与在产业甚至是企业层面分析所用的数据是不一样的。另外,所采用的数据也可能存在缺陷,因此结论也可能会存在差异。

在 FDI 技术外溢以及东道国技术进步的实证方面,不少国内学者就中国的情况进行了不同角度的研究,较有代表性的有何洁 (2000)、赖明勇和包群 (2002)、姚洋和章奇 (2001)、潘文卿 (2003) 等。何洁认为,FDI 的效果受当地经济发展特征的影响。^[12]赖明勇和包群指出,FDI 的技术进步作用有较大的波动性。^[13]沈坤荣、耿强 (2001) 认为 FDI 可以通过技术外溢效应提高国民经济的综合要素生产率。^[14]姚洋和章奇区分了 FDI 对 GDP 的贡献的两个来源,即外资企业的技术外溢和外资企业相对于内资企业的生产率差异。^[15]潘文卿强调 FDI 的外溢效应存在着地区性差异,存在着一个当地经济发展水平的门槛,跨过此门槛后,FDI 外溢效果才会是正向的、积极的。^[16]

张建华和欧阳轶雯以广东的数据讨论了 FDI 技术外溢与经济增长的关系,认为 FDI 在行业中的外溢效应小于在地区内的效应,省内各城市技术水平和政策因素会强烈影响 FDI 的外溢效果。^[17]

二、广东省外商直接投资概观

广东省利用外资在全国占有重要地位。2002 年广东批准利用外资项目的合同外资额 189.01 亿美元,比 2001 年增长 45.1%;实际引资 165.89 亿美元,增长 26.2%。其中,全年批准的外商投资企业合作合同外资金额 161.71 亿美元,增长 51.7%;实际外商直接投资 131.11 亿美元,增长 26.6%,占当年全国实际外商直接投资 527.43 亿美元的 24.86%。FDI 对广东省的技术进步起什么作用,与在广东省的 FDI 的来源地结构和技术结构有十分密切的关系。

首先,从外商直接投资的来源地结构看,广东的 FDI 来源结构有以下几个特点:一是广东的 FDI 主要来源于港澳台地区,港澳台投资一直占 FDI 总额的 80% 左右。二是来自西方国家的 FDI 中,以美国和日本为首,占实际利用金额的 17%。三是英、法、德对广东的 FDI 呈波动增长,其投资项目规模较大,但流入总量规模较小。

其次,从 FDI 的技术结构来看,早期广东引进的 FDI 项目多数技术层次偏低,且集中在服务业和加工工业,这种行业性的结构制约了其技术的发展。20 世纪 90 年代后,FDI 的内部结构发生了变化:一是服务业的 FDI 由餐饮、酒店等项目向房地产及相关的物业管理投资等项目转移;二是在制造业的 FDI,由资源加工型和劳动密集型的一般工业制造项目逐步向化学、电子电气等资本和技术密集型制造业转移。这表明外商对资本和技术密集型行业的投资比重有所增大,但是,技术密集型投资还是较少,技术层次偏低的基本格局仍未改变。

上述特点是我们评价广东省利用外资时必须考虑的重要因素,但我们希望更进一步从量化的角度来考察 FDI 与广东省技术进步的关系。我们首先通过对全要素生产率 (TFP) 的分析,获取 FDI 与技术进步的关系的基本信息,然后进一步考虑上面提到的两方面因素,即通过对 FDI 的来源地结构和技术结构的详细分析来讨论 FDI 与技术进步的关系。

三、广东历年 TFP 变化和 FDI 的关系

经济学家通过研究生产函数发现,资本、劳动力和科技进步是影响经济产出的主要因素。科技进步的概念是通过全要素生产率 (TFP) 来体现的。即在其他条件不变情况下,由于时间变化而引起产出的增加部分,被认为是技术进步的作用。我们采用索洛对全要素生产率的度量方法求出广东省 1986~2001 年的 TFP,回归方程为:

$$\ln(Y_t/L_t) = \ln A_t + \alpha_T t + \alpha_K \ln(K_t/L_t) + \varepsilon_t \quad (1)$$

式中: Y_t , L_t , K_t 分别代表 GDP、劳动力投入、资本存量; A_t 代表 t 时期的全要素生产率。本文采用的数据全部来自历年《广东统计年鉴》和《广东工业统计年鉴》。

通过 (1) 式的回归,得到资本的产出弹性为 $\alpha_K = 0.42$, 劳动的产出弹性为 $\alpha_L = 1 -$

$\alpha_K = 0.58$ 。可以看出,劳动力对广东总产出的贡献明显大于资本的贡献。令 1986 年的 TFP 指数为 100,可以计算出 1986~2001 年的 TFP 指数及其增长率。

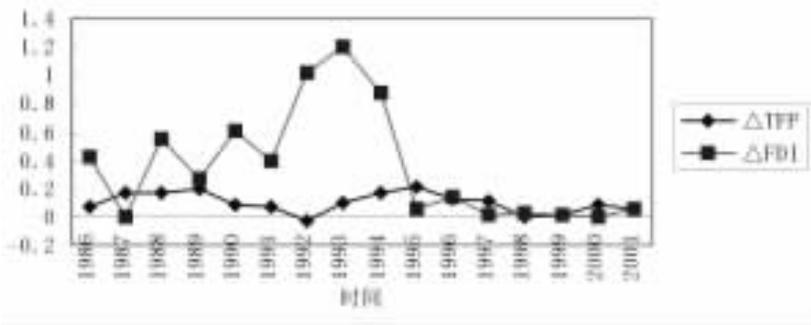


图 1 广东省 TFP 指数及其增长率 (1986~2001)

从图 1 中我们可以看出,在 20 世纪 90 年代中期以前,广东省 TFP 的增长同进入广东的 FDI 的增长有着比较明显的负相关性。90 年代中期之后,广东的 TFP 和进入广东的 FDI 之间的负相关性降低,出现了 TFP 和 FDI 比较平稳的增长期。

接下来考虑 FDI 促进技术进步的直接效应和间接效应。前者表现为外资企业相对于国内企业的要素生产率优势的提高。后者通过外商直接投资占国内总投资的比重(反映外资企业技术外溢效果)来反映。^[8]

$$\ln Y_t = \ln B_t + \eta \text{SHARE}_t + \ln \text{FDI}_t + \alpha \ln L_t + \beta \ln K_t \quad (2)$$

式中: $\ln Y$, $\ln K$, $\ln L$, $\ln \text{FDI}$ 分别是国内生产总值、投资、劳动力、外商直接投资取对数; SHARE 是外商直接投资额占广东省国内总投资的比重。利用广东省 1986~2001 年 16 年的数据,按照公式进行回归,结果如下:

$$\begin{aligned} \ln Y_t = & -77 + 2.3 \text{SHARE}_t - 0.25 \ln \text{FDI}_t + 0.42 \ln L_t + 5.69 \ln K_t \quad (3) \\ & (-6.9) \quad (-1.23) \quad (2.45) \quad (1.96) \quad (7.52) \\ R^2 = & 0.995 \quad \text{DW} = 1.5 \end{aligned}$$

从整个回归结果来看,在广东省的外资企业对于广东的国内企业有着比较强的溢出效应,而外商直接投资的进入在技术进步方面的促进作用十分不明显,甚至有负的效应。这个结果和人们通常认为的进入东道国的跨国公司相对于发展中东道国国内企业有着更高的生产率优势的假设有出入。原因可能与我们前述的 FDI 的来源地(即 FDI 本身技术含量不高)和 FDI 所投资的产业不尽理想有关。接下来我们进一步分析这两个因素对广东省技术进步的作用。

四、FDI 来源地对广东技术进步的影响

将进入广东的 FDI 分成两部分:一部分是来自于港澳台地区;另一部分来自于西方七国,以考察 FDI 来源地的不同所导致的 FDI 对广东技术变迁的影响的不同。

首先，将前面计算的 TFP 与来自港澳台的 FDI 和来自西方七国的 FDI 的变化趋势进行比较，这里 ΔTFP 是 TFP 的变化， $\Delta WEST$ 是西方七国进入广东的 FDI 的变化， ΔGAT 是港澳台进入广东的 FDI 的变化（见图 2）。

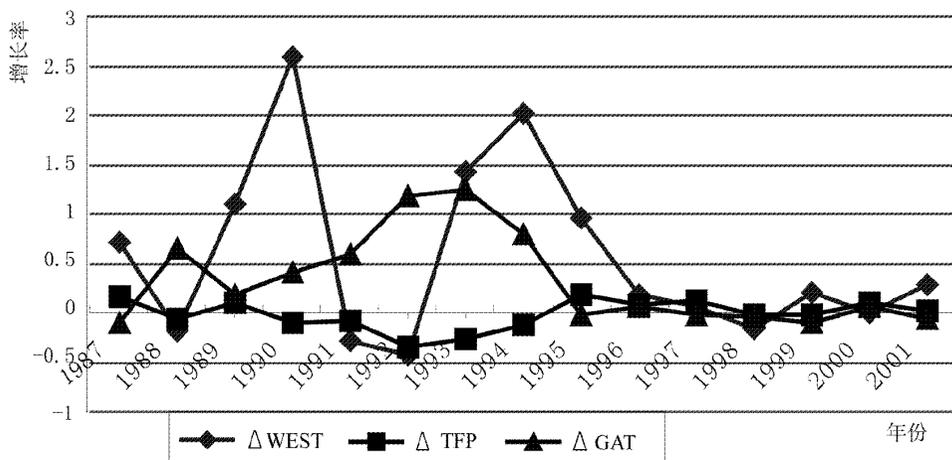


图 2 广东 TFP 与来自港澳台的 FDI 和来自西方七国的 FDI 的变化趋势

从图 2 中可以看出，广东 TFP 的变化与来自港澳台的 FDI 的变化有比较明显的负相关性；而广东 TFP 的变化与来自西方七国的 FDI 的变化有相对正的相关性。据此，我们将影响 TFP 的因素分为来自港澳台的外资 FDI_g 、来自西方七国的外资 FDI_w 和其他因素 B，并建立回归方程：

$$\Delta TFP = B + \alpha \Delta FDI_g + \beta \Delta FDI_w + \epsilon \quad (4)$$

这里， ΔFDI_g 和 ΔFDI_w 分别是指来自港澳台和西方七国的外商资本变化； ΔTFP 是全要素生产率的变化，可以认为是广东技术的变化。回归结果如下：

$$\begin{aligned} \Delta TFP = & 0.07 - 0.3\Delta FDI_g + 0.015\Delta FDI_w & (5) \\ & (2.74) \quad (-6.83) \quad (1.66) \\ & R^2 = 0.8 \quad DW = 1.8 \end{aligned}$$

从回归结果上看， R^2 等于 0.8，这表明模型的拟合效果比较好，同时 F 检验的相伴概率为 0.000000，说明变量之间呈高度线性，回归方程高度显著。对于 $n = 15, p = 3$ 查表得 DW 检验的 1% 临界值为 $du = 0.59, dl = 1.46$ 根据判定法残差序列不存在自相关。另外进行不含交叉项的 White 检验可以得到检验的相伴概率是 0.279，因此可以认为基本模型不存在异方差。

从回归结果可以看出，来自港澳台的 FDI 对广东全要素生产率的影响是负面的；而来自西方七国的 FDI 对广东的全要素生产率的影响是正的，来自西方七国的资本增长 1%，广东全要素生产率上升 1.5%。随着来自西方发达国家的 FDI 的不断增加，FDI 将会更好地促进广东技术进步。

五、FDI 的投资产业与技术进步

投入不同产业的 FDI 有可能对广东技术进步产生不同影响。为验证这种想法，我们将广东的 FDI 分成投向劳动密集型产业和资本技术密集型产业的 FDI。利用广东 1998 ~ 2001 年 27 个工业部门的数据，建立下面这个一般的线性 Panel Data 模型。

$$\ln Y_{it} = \ln B_{it} + \eta_{it} \text{SHARE}_{it} + \theta_{it} \ln \text{FDI}_{it} + \alpha_{it} \ln L_{it} + \beta_{it} \ln K_{it} + \mu_{it} \quad (6)$$

在应用 EVIEWS 软件的分析中，我们选用了使用可行的广义最小二乘法（GLS）估计，目的是为了减少由于截面数据造成的异方差影响。

从表 1 中的回归结果看，我们可以发现调整后的 R^2 等于 0.9998，这表明模型的拟合效果很高，DW 值等于 1.735878，根据判定法残差序列不存在自相关。从整体上讲，该模型的效果不错。

表 1 劳动密集型产业的 FDI 与技术进步

因变量：ln Y				
变量	系数	标准误差	t - 统计量	概率
Ln B	5.499865	0.678196	8.109547	0.0000
SHARE	- 0.217537	0.123153	- 1.766389	0.0822
FDI	0.044613	0.027642	1.613987	0.1115
K	0.604895	0.043702	13.84143	0.0000
L	0.248689	0.029801	8.345043	0.0000
加权的统计量 (Weighted Statistics)				
R^2	0.999865	因变量均值 (Mean dependent var.)		44.04527
\bar{R}^2	0.999857	因变量标准差 (S.D. dependent var.)		23.85186
标准误差	0.285557	残差和 (Sum squared resid.)		5.137191
DW 统计值	1.735878			
F - 统计量	116846.5	F 统计量相伴概率 (Prob. <F- statistic >)		0.000000
未加权的统计量 (Unweighted Statistics)				
R^2	0.812775	因变量均值 (Mean dependent var.)		23.84905
\bar{R}^2	0.800887	因变量标准差 (S.D. dependent var.)		0.671746
标准误差	0.299747	残差和 (Sum squared resid.)		5.660434
DW 统计值	1.960728			

从表 2 中的回归结果看，我们可以发现调整后的 R^2 等于 0.999995，接近 1，这表明模型的拟合效果很高，DW 值等于 1.884373，根据判定法残差序列不存在自相关。从整体

上讲，该模型的效果不错。

表 2 资本密集型产业的 FDI 与技术进步

因变量：lny				
变量	系数 Coefficient	标准误差 Std. Error	t - 统计量	概率 Prob.
lnB	- 6.020849	0.995498	- 6.048076	0.0000
SHARE	6.543067	1.415478	4.622513	0.0000
FDI	- 0.350262	0.072226	- 4.849507	0.0000
K	1.481096	0.103108	14.36445	0.0000
L	0.086224	0.057645	1.495775	0.1037
加权的统计量 (Weighted Statistics)				
R ²	0.999995	因变量均值 (Mean dependent var.)		75.15095
\bar{R}^2	0.999995	因变量标准差 (S.D. dependent var.)		134.9355
标准误差	0.315745	残差和 (Sum squared resid.)		3.489315
DW 统计值	1.884373			
F - 统计量	1780667.	F 统计量相伴概率 (Prob. <F-statistic >)		0.000000
未加权的统计量 (Unweighted Statistics)				
R ²	0.916412	因变量均值 (Mean dependent var.)		24.23136
\bar{R}^2	0.906859	因变量标准差 (S.D. dependent var.)		1.034654
标准误差	0.315766	残差和 (Sum squared resid.)		3.489785
DW 统计值	1.569070			

通过上面两个 Panel Data 分析中可以看出，进入广东的 FDI 对广东劳动密集型产业和资本密集型产业的技术进步的影响是不同的。分析的结果表明，投资在劳动密集型的 FDI 的直接效应是正的效应，每增加 1% 的 FDI，对广东劳动密集型产业的直接效应为 0.044%，这说明进入广东劳动密集型产业的 FDI 总体技术水平是高于广东本地企业的技术水平，因此，FDI 企业本身就是提高广东技术水平的因素。但是溢出效应是负的，每增加 1% 的 FDI，对广东的劳动密集型产业的溢出效应为 - 0.22%，负的溢出效应是与劳动密集型产业本身的技术水平、劳动者的素质和产业关联度有着直接的关系。

从回归结果中可以看到，投资在资本密集型的 FDI 的直接效应是负效应，每增加 1% 的 FDI，对广东资本密集型产业的直接效应为 - 0.35%，这就与人们普遍认为的 FDI 企业的技术水平比东道国企业的技术水平要高有反差。这与在前面分析过的 FDI 来源地不同对技术进步有不同的影响相对应，较多来自港澳台的企业（当然也包括一部分来自发达国家的企业）主要投资在资本密集型产业的加工装配环节，而这些环节本身技术含量并不高；而来自发达国家的 FDI 虽然技术水平比较高，但这些企业的进入从某种角度不利于原有的

本地企业利用优质的资源。FDI 对广东资本密集型产业溢出效应是正，而且是比较明显，每增加 1% 的 FDI，对广东的资本密集型产业的溢出效应为 6.5%。回归结果证明了前面的推断，即投资于资本密集型产业的 FDI 比投资于劳动密集型产业的 FDI 有着更强的溢出效应，而从直接效应看，投资于资本密集型产业的 FDI 又的确对广东技术进步带来了负面的影响。

六、结论和政策建议

从以上的分析可以看出，从技术溢出效应和直接效应来看，FDI 厂商的进入对广东产生的主要作用在于溢出效应。而较弱的直接效应的产生则可以分成两部分：一部分是由于进入广东的 FDI 的质量不高导致的；另一部分就是 FDI 进入之后对国内厂商产生的挤出效应。当然，FDI 影响技术进步的效应远不止这些，我们所分析的仅是其中某些方面。

从技术进步的角度看，较为理想的情形是 FDI 能够同时有积极的直接效应和技术溢出效应。但实际的情况往往不尽如人意，何洁（2000）曾经指出过 FDI 对东道国技术进步可能有的负面影响，认为 FDI 会将原本服务于国内企业的优秀人力资本吸引过去，造成国内企业人力资本的减少和积累速度的减慢；FDI 企业利用雄厚的资本和广泛的信息来源，抢占了原本由本国企业利用的稀缺资源，从而造成本国企业生产率的下降。^[12]广东利用外资方面存在类似的情况。不容否认的事实是，不少 FDI 厂商进入广东主要是为了利用劳动力廉价的优势，将广东作为其全球战略中的加工基地，缺乏技术转移的动机。总体上看，广东引进外资的规模已是较大，FDI 在广东的投资规模、出口创汇、增加就业以及经济增长方面起着重要的作用，但在通过引进外资获取技术的质量方面则还不甚理想。

我们认为，要充分发挥 FDI 对广东技术进步的推动作用，应当特别注意以下几个方面：

第一，需要将引进外资的观念从重视规模转向重视质量，应调整外资政策，把推动技术进步作为引进外资的重点，积极吸引西方发达国家跨国公司投资。正如本文前面提到的，进入广东西方七国的 FDI 对广东技术变迁有着积极的影响，应该着力吸引跨国公司到广东投资。着重完善有利于跨国公司投资的环境建设，推动各类大中型企业与跨国公司的合作，促进高科技产业的发展。

第二，要积极引进高产业关联度的项目，积极培育与外资相关产业的发展，形成有效的产业关联和聚集效应，充分发挥外资的产业带动效果，实现产业结构高度化的目标。尤其是一些由跨国公司主宰的主导产业群，如汽车、微电子、生物工程和航天工业等作为主导产业，一种企业层次上的复合一体化国际分工体系已经形成，产业关联系数极高，而且国际化程度也很高，它正在越来越大的程度上决定和引导着全球产业结构的变化。简言之，通过发展本地的配套产业来吸引国际重要主导产业的跨国公司进入广东。

第三，国内外大量研究表明，实现技术的国际扩散或转移，东道国的技术吸收能力是关键，一国的劳动力的素质在一定程度上决定了能够吸引的 FDI 的技术含量。因此，必须大量投资于人力资本。受过良好教育和培训的工人具有更高的生产率，而且更能应对先进技术，拥有大量受过良好教育、具有良好沟通能力的高技能的劳动力，就有可能吸引需要

使用这些技能的 FDI, 使得 FDI 的溢出效应更加显著。

广东的经济发展已具有了相当实力, 为把目光转向更多地注重引进外资的技术含量提供了必要的物资基础, 但需要更多的信息交流和政策引导。同时, 本地企业需要有一种在开放经济条件下竞争的危机感, 以及长期发展的眼光, 加快引进和学习先进技术。当本地企业有强烈的获取技术的需求和能力、政府政策有利于形成一个适宜国际技术转让的环境, 再加上外资自身的技术扩散需要, 提高广东引进外资的技术水平是完全可能的。

参考文献

- [1] Aitken, Brian, Gordon H. Hanson, and Ann E. Harrison, Spillovers, Foreign Investment, and Export Behavior, *Journal of International Economics*, 1997, Vol.43, pp103 ~ 132.
- [2] Baldwin, Richard; Henrik Braconier and Rikard Forslid, *Multinationals, endogenous growth and technological spillovers: theory and evidence*, London: CEPR Discussion Paper, 1999, No. 2155.
- [3] Benhabib, J. and M. Spiegel, The Role of Human Capital in Economic Development: Evidence from Aggregate Cross-Country Data, *Journal of Monetary Economics*, 1994, Vol. 34, pp143 ~ 174.
- [4] Borensztein, E., De Gregorio, J. and J-W. Lee, How does foreign direct Investment affect economic growth? *Journal of International Economics*, 1998, Vol.45, pp115 ~ 135.
- [5] Campos, Nauro F. and Yuko Kinoshita, Foreign Direct Investment as Technology Transferred: Some Panel Evidence from the Transition Economies, *William Davidson Working Paper*, 2002, Number 438.
- [6] Coe, D. T. and R. Moghadam, Capital and Trade as Engines of Growth in France, *IMF Staff Papers*, 1993, Vol.40, pp542 ~ 566.
- [7] Fortanier, Fabienne, FDI and technology transfer, in OECD (2001), *The Costs and Benefits of FDI*, Paris: OECD, 2001.
- [8] Keller, Wolfgang, International Technology Diffusion, *NBER Working Papers 8573*, National Bureau of Economic Research, 2001.
- [9] Lichtenberg, Frank and Bruno van Pottelsberghe de la Potterie, International R&D Spillovers: A Re-Examination, *NBER Working Papers 5668*, National Bureau of Economic Research, 1996.
- [10] Lipsey, Robert E., Trade and Production Networks of US MNCs and Exports by their Asian Affiliates, *NBER working paper 5255*, National Bureau of Economic Research, 1995.
- [11] Naujoks, Petra and Klaus-Kieter Schmidt, Foreign direct investment and trade in transition countries: Tracing the links, *Kiel Working Paper 704*, 1995.
- [12] 何洁:《外国直接投资对中国工业部门外溢效应的进一步精确量化》,《世界经济》,2000年第12期。
- [13] 赖明勇,包群:《中国外商直接投资与技术进步的实证研究》,《经济评论》,2002年第6期。
- [14] 沈坤荣,耿强:《外国直接投资、技术外溢与内生经济增长——中国数据的计量检验与实证分析》,《中国社会科学》,2001年第5期。
- [15] 姚洋,章奇:《中国工业企业技术效率分析》,《经济研究》,2001年第10期。
- [16] 潘文卿:《外商投资对中国工业部门的外溢效应:基于面板数据的分析》,《世界经济》,2003年第6期。
- [17] 张建华,欧阳轶雯:《外商直接投资、技术外溢与经济增长——对广东数据的实证分析》,《经济学季刊》,2003年第2期。

论跨国公司 R&D 投资的国别选择^{*}

喻世友 万欣荣 史 卫

一、引言

跨国公司 R&D 投资国际化 (Internationalization of R&D by Multinational Firms) 就是跨国公司将其 R&D 投资活动伸展到本国以外的其他国家, 目的是利用各国研发资源, 在全球范围内组织起 R&D 活动。从 20 世纪 70 年代起, 跨国公司 R&D 投资国际化现象开始引起学术界的广泛注意, 有迹象表明, 大多数跨国公司都显示出提高 R&D 投资国际化比例的趋势。不可否认, 跨国公司是先进技术创新和扩散的主体, 广泛开展与跨国公司的合作是东道国获取技术、提高技术水平的重要渠道。鉴于可观的 R&D 溢出效应, R&D 投资的国际化有利于东道国的技术创新能力的提高。而技术创新又推动了跨国公司 R&D 能力的发展, 从而促使跨国公司 R&D 投资的国际化。可以说, 跨国公司 R&D 的国际化现象对母国和东道国的技术发展呈现出一种双赢关系。特别进入到 20 世纪 80 年代中期, 随着经济全球化趋势的迅猛发展和国际竞争的日趋激烈, 跨国公司在海外设立 R&D 机构的趋势不断加强, 标志着跨国公司的发展进入到一个新的发展阶段。

本文以跨国公司 R&D 投资的国际化问题作为研究对象, 研究的目的是分析跨国公司 R&D 投资的区位分布规律, 从而为中国制定吸引跨国公司 R&D 投资的区位政策提供依据。本文采用了美国跨国公司海外 R&D 投资数据, 通过选取 21 个国家和地区的数据, 检验了东道国吸引跨国公司 R&D 投资的决定因素。运用多元线性回归分析的方法, 力求从另一个角度分析影响跨国公司 R&D 投资国家选择的因素。同时结合跨国公司 R&D 投资动机, 从中发现一些规律性的东西, 从而为中国制定吸引跨国公司 R&D 投资的区位政策提供依据。中国是一个发展中国家, 也是一个经济快速增长的国家, 虽然整体技术水平还相对落后, 但在某些技术领域处于领先地位, 通过吸引跨国公司来华进行 R&D 投资, 必将促进中国整体技术水平的提高。因此, 正确规划 R&D 投资的区位布局已刻不容缓, 塑造怎样的区位优势来吸引跨国公司 R&D 投资是一个必须尽快回答的问题。

(一) 国外学者关于跨国公司 R&D 投资国际化的理论研究

跨国公司 R&D 的专门研究比 FDI 理论研究要晚得多。长期以来, 经济学家一直认为 R&D 活动发生在国内, Vernon (1966) 和 Kindleberger (1969) 曾经用一个准殖民地模型论证了这一观点。进入到 20 世纪 80 年代后, 随着跨国公司 R&D 国际化这一经济现象的迅

^{*} 原载:《管理世界》, 2004 年第 1 期。

猛发展,经济学家们逐渐关注跨国公司 R&D 国际化问题的研究。

Mansfield (1979) 运用 55 家美国公司的资料,发现海外 R&D 与海外子公司的销售呈正相关关系。Lall (1979)、Hewitt (1980)、Hirschey 和 Caves (1981) 运用美国统计局的数据分析了美国的跨国公司在海外 R&D 投资的区位偏好,发现海外 R&D 活动与其所服务的国外市场规模呈现明显的正相关关系,产业的 R&D 强度与海外 R&D 活动也具有一定的相关性,R&D 投资于国外是为了节省成本。他们还发现在美国从事 R&D 人员的工资成本与海外 R&D 活动呈正相关性。Pearce (1989) 创建了 R&D 分散与集中理论,认为影响跨国公司建立海外 R&D 实验室是由于离心力和向心力相互作用而造成的。离心力即促使 R&D 机构分散化的因素,如关键的投入、相互作用交流和政治因素等。向心力即促使 R&D 机构集中化的因素,如经济规模、内部交流、安全和秘密等。Pearce (1989) 还从公司层面分析了 1982 年 122 家主要跨国公司海外 R&D 活动的区位决定因素,发现海外生产比例与海外 R&D 活动呈正相关。Zejan (1990) 通过对瑞士跨国公司海外 R&D 活动的分析,发现海外子公司的 R&D 与母公司的 R&D 强度有关,东道国的特征(市场规模、人均收入)均对跨国公司海外 R&D 活动有促进作用。Fors 和 Zejan (1996) 通过对瑞士跨国公司海外 R&D 活动的研究,发现海外 R&D 的区位选择很大程度上是为了适应当地生产的需要,而且跨国公司倾向于将海外 R&D 布局在其行业领域技术相对先进的地区。Granstrand 将影响跨国公司 R&D 对外直接投资的驱动因素归纳为:支持当地化生产,满足当地消费者的生产需求,获得国外先进科学技术,降低 R&D 人力成本,东道国政府政策影响(尤其是知识产权保护法)。Kuemmerle (1997) 将跨国公司海外 R&D 直接投资分为两类:充分利用跨国公司的现有技术知识和开拓国际市场。所以,影响跨国公司 R&D 对外直接投资的主要因素是东道国市场规模和成长潜力。东道国公共部门和私有机构的 R&D 投资量、国内人力资源质量、有关科技领域取得的卓越成就等,成为影响跨国公司选择海外区位、建立 R&D 分支机构的关键因素。

(二) 国内学者关于跨国公司 R&D 投资国际化的理论研究

林进成、柴忠东(1998)指出,跨国公司投资于国外的 R&D 经费占其总额比例不断上升;在国外企业中,从事 R&D 人数的比重大幅度提高;在高新技术领域内相互交叉投资的现象较为突出;倾向于在科研政策宽松、服务设施完善、市场潜力巨大的东道国从事技术开发。沈群红(1999)分析了国际 R&D 的发展状况,指出 R&D 国际化使企业获得网络知识,并进一步促使 R&D 活动的全球化,加快了世界范围内的知识流动与技术扩散。冼国明、葛顺奇(2000)认为市场一体化的发展、知识经济的兴起和信息技术的进步,促使跨国公司 R&D 实行国际化战略,主要表现在:跨国公司海外 R&D 分支机构数量不断增加;海外 R&D 支出在公司总支出中的比例不断上升;海外 R&D 分支机构的专利日益增多。江小涓(2000)对跨国公司在华 R&D 行为的特点作了研究,认为:从总体上看,外商在华研发机构,多数从事适应型 R&D 活动,真正从事创新型 R&D 活动的很少。但以上文献还仅限于对 R&D 全球化现象的描述和分析,尚没有涉及跨国公司 R&D 的区位选择问题,对于跨国公司 R&D 的区位选择的实证分析更是很少。不过最近已有学者开始注意到这一问题。郑京淑(2000)探讨了跨国公司 R&D 职能的海外区位特征,对我们研究各国

各地区在跨国公司全球战略中的位置以及在跨国公司技术转移轨迹中所处的地位具有重要意义。杜德斌(2001)采用1997年美国跨国公司R&D支出数据,从跨国公司在当地生产的规模和特性、东道国技术资源状况和东道国政策法规三个方面来考察影响跨国公司R&D投资区位选择的因素。但作者采用一元线性回归的方法,不能够从总体上说明东道国有关的社会经济变量对美国海外R&D支出的影响,忽视经济变量之间的相互影响。采用一元线性回归的方法不能对东道国有关的社会经济变量进行特征分析,容易遗漏重要的解释变量,产生有偏估计。

二、跨国公司 R&D 国家选择模型的建立及数据分析

中国学者关于跨国公司 R&D 国际化问题的研究还主要集中于现象描述和原因分析,跨国公司在选择目标国家进行 R&D 投资时,会受东道国区位条件的引力作用影响。本文着重对跨国公司 R&D 国家选择展开实证分析,采取多元线性回归分析的方法,力求从另一个角度分析东道国吸引跨国公司 R&D 投资的决定因素。

(一) 变量的归纳及数据分析

本文选取了美国跨国公司在 1997 年海外 R&D 投资数据作为检验样本,分别选取了 21 个国家和地区的数据。在建立回归模型的过程中,所涉及的变量可能会很多,我们把可能影响跨国公司 R&D 投资国家选择的因素归为三大类。

(1) 东道国生产规模的程度和特性,包括 FDI、GDP、人口。外商直接投资(FDI)可以被认为是吸引跨国公司 R&D 投资的决定因素,跨国公司 R&D 海外投资与其海外生产有关。市场规模包括东道国的 GDP 和人口,一些经济学家(如 Zejan, 1990)通过对瑞士跨国公司的实证研究,发现市场规模对 R&D 的区位选择有决定作用。

(2) 东道国的资源,包括每千人拥有计算机数、每百万人从事 R&D 的科学家和工程师人数、R&D 占 GNP 的比例、工资水平。一国每千人拥有计算机数越多,其通讯设备越好,会对跨国公司 R&D 区位选择产生影响;每百万人从事 R&D 的科学家和工程师人数反映了一国的人才资源,跨国公司进行海外 R&D 投资是看重东道国大批高素质且廉价的人才资源;R&D 占 GNP 的比例反映一国对 R&D 资金投入程度的高低,R&D 投入是衡量一国科技实力大小的重要指标;工资水平的高低会成为影响跨国公司海外 R&D 投资的因素,一般而言,东道国低工资水平可降低跨国公司海外 R&D 投资的进入成本。

(3) 东道国的政策制度,包括知识产权保护水平。世界银行(2001)对知识产权保护的结论是:知识产权保护水平取决于经济和社会环境,在经济学分析中,知识产权保护是对知识运用的暂时垄断状态,是对知识和信息市场失灵的次优解决方法,知识产权保护可以鼓励发明者进行 R&D 活动。R&D 活动的直接产品是知识产品,本文拟用知识产权保护来测度跨国公司 R&D 活动的国家选择,而每千人拥有的专利数即发明者进行专利注册是衡量知识产权保护水平的一个重要依据,因为知识产权保护水平与发明者的回报呈正相关关系。如果一个国家专利权利缺乏保护的话,它将成为专利可专用性的障碍,发明者因此失去了 R&D 投资的回报利润,而这部分回报利润被模仿者无偿获得。可以假设东道国完

善的知识产权保护水平将刺激发明者进行专利注册，每千人拥有的专利数可以反映一国的知识产权保护水平。

(二) 变量的选择

在上述诸多影响因素中，要从中选择几个合适的自变量是不容易的，跨国公司 R&D 投资受到诸多因素的影响，且各因素之间又交织在一起，相互作用，要完全分离出某一因素对跨国公司 R&D 投资所产生的影响是比较困难，因为某些变量对 R&D 投资的国家选择可能不是最重要的，或者是因为有较大的测量误差，或者是与其他自变量高度相关，须剔除这些自变量。^① 自变量太多的回归模型易于引起多重共线性，也不便于分析研究。为了不遗漏重要的解释变量，我们首先引入全部变量，提出多个假说，来考察各种因素对 R&D 投资区位分布的影响程度，从而分析在诸多影响因素中那些因素是最重要的，建立初步模型：

$$RD = C + a_1 \times FDI + a_2 \times GDP + a_3 \times POP + a_4 \times RDPOP + a_5 \times COMPUTER + a_6 \times WAGE + a_7 \times PPP + a_8 \times RDGDP + \epsilon$$

其中：RD：跨国公司海外 R&D 支出（单位：百万美元）；FDI：外商直接投资（单位：亿美元）；GDP：东道国的国民生产总值（单位：百万美元）；POP：人口（单位：百万人）；RDPOP：每百万人从事 R&D 的科学家和工程师人数；COMPUTER：每千人拥有计算机数；WAGE：工资水平（单位：美元每月）；PPP：知识产权保护水平（每千人拥有专利数）；RDGDP：R&D 占 GDP 的比例。

我们预期以上所有因素的符号都为正，以海外 R&D 为因变量对自变量进行回归，结果如表 1 所示。

表 1 截面数据的 OLS 估计

常数项 - 412.9797 (- 1.302817)	FDI 1.029058** (2.326409)	GDP 0.000363*** (1.843986)	POP 0.125184 (0.27434)	COMPUTER 0.544504 (0.369595)
RDPOP - 0.33668*** (- 1.777464)	RDGDP 223.1979 (0.939406)	WAGE 0.333528 (1.371529)	PPP 297.9316 (0.68801)	
R ² = 0.747325 F = 4.436482 (0.01)				

注：系数下面括号里的是 t- 统计值；** 指在 5% 水平上显著，*** 指在 10% 水平上显著。

如表 1 所示，经初步回归，通过 t 值检验和经济意义的比较，我们最终选取了 FDI

^① 自后淘汰变量法 (BACKWARD)。先将全部 K 个自变量引入回归方程，然后对每个自变量做显著性检验，剔除不显著变量中最“不重要”的（即所谓“偏回归平方和”最小的）；第二步是用剩下的自变量与 Y 重新拟合回归方程，然后再按前面的做法剔除最“不重要”的变量；这样继续进行，直到方程中的所有变量都显著为止。

(外商直接投资)、RDPOP (每百万人从事 R&D 的科学家和工程师人数)、GDP (东道国的国内生产总值)、PPP (每千人拥有专利数) 4 个指标作为模型分析的变量, 删除了 POP (人口)、COMPUTER (每千人拥有计算机数)、RDGDP (R&D 占 GDP 的比例)、WAGE (工资水平) 4 个指标。

(三) 跨国公司 R&D 国家选择模型的建立

通过上述因素的筛选, 重新建立模型:

$$RD = \beta_1 + \beta_2 \times FDI + \beta_3 \times PPP + \beta_4 \times GDP + \beta_5 \times RDPOP + \epsilon$$

为了验证海外 R&D 支出与 FDI、每百万人从事 R&D 的科学家和工程师人数、东道国的国内生产总值、知识产权保护间的相关关系, 笔者对模型进行回归, 回归结果如下:

$$RD = -110.3776 + 1.453019 \times FDI + 0.0003 \times GDP + 314.9644 \times PPP - 0.0833 \times RDPOP$$

(- 0.58063) (4.258836) (1.979906) (0.834043) (- 0.723373)

$$R^2 = 0.681563 \quad \text{调整后的 } R^2 = 0.601954 \quad F \text{ 检验} = 8.561353 \quad DW = 1.405017$$

经多元线性回归结果分析, 整个模型在 1% 水平上通过 F 检验 (F 值为 8.561353), 但 PPP、RDPOP 不能通过 t 检验, 进而检验解释变量之间的多重共线性问题。发现 $\text{Corr}(PPP, RDPOP) = 0.866267994584$, 这说明 PPP 和 RDPOP 之间具有高度相关性, 整个回归方程存在多重共线性问题。另外, RDPOP 的系数为负数, 与理论假设不符, 而知识产权保护比每百万人从事 R&D 的科学家和工程师人数更能反映东道国政策制度的指标, 为了避免信息的重复反映, 因此删除 RDPOP 因素。

删除 RDPOP 后重新建立模型:

$$RD = \beta_1 + \beta_2 \times FDI + \beta_3 \times PPP + \beta_4 \times GDP + \epsilon$$

对模型进行回归, 回归结果如下:

$$RD = -147.893 + 1.410202 \times FDI + 0.000324 \times GDP + 83.07466 \times PPP$$

(- 0.820233) (4.257098) (2.191782) (0.42214)

$$R^2 = 0.671149 \quad \text{调整后的 } R^2 = 0.613116 \quad F \text{ 检验} = 11.56503 \quad DW = 1.377834$$

经多元线性回归结果分析, 所有因素的符号都为正, 与理论假设相符, $R\text{-squared} = 0.671149$, 通过 F 检验, FDI、GDP 均通过了 t 检验, 只有 PPP 未通过 t 检验, 怀疑存在异常点。经残差分析检验, 剔除第 1 个和第 4 个样本数据后, 对模型再进行回归, 结果如下:

$$RD = -163.6913 + 1.4706 \times FDI + 0.00012 \times GDP + 112.7474 \times PPP$$

(- 2.761296) (13.55887) (2.304693) (1.743511)

$$R^2 = 0.942 \quad \text{调整后的 } R^2 = 0.93 \quad F \text{ 检验} = 80.561$$

残差呈正态分布, 回归结果完全令人满意。R-squared = 0.941562, FDI、GDP、PPP 均通过了 t 检验, 整个模型可以通过 F 检验 (F 值为 80.56148)。

三、实证结果及分析

由上述结果分析, 这一回归分析的拟合程度很好, 显示跨国公司进行海外 R&D 投资

与东道国的 FDI 规模和东道国的国民生产总值以及知识产权保护水平呈高度相关关系，这三个因素是东道国影响跨国公司进行 R&D 投资国家选择的决定因素。

(1) 跨国公司 FDI 的流向影响其海外 R&D 投资，回归结果表明 FDI 通过 t 值在 0.001 水平上的显著性检验，证明 FDI 是跨国公司 R&D 投资国家选择最为重要的一个因素，这也支持了前面的假设：东道国的 FDI 规模是吸引跨国公司 R&D 投资的决定因素。这也与实际情况相符，跨国公司到东道国 R&D 投资的一个重要目的是支持当地的生产。但是，从数据分析，跨国公司 R&D 投资与 FDI 相关性并不完全一致。例如，美国在 1997 年海外 R&D 投资最大的国家是德国，R&D 投资金额达到 2964 百万美元，海外 R&D 投资第四位的国家是法国，R&D 投资金额达到 1238 百万美元，但德国、法国吸引的 FDI 并不是很多，分别为 385 亿美元和 358 亿美元，德国吸引的外商直接投资金额不仅远远少于居第一位英国的 1531 亿美元，甚至少于加拿大的 960 亿美元和荷兰的 644 亿美元；而法国吸引的外商直接投资金额也少于意大利的 377 亿美元和荷兰的 644 亿美元。这说明跨国公司 R&D 投资除受东道国的 FDI 规模影响外，还受到其他因素的制约。

(2) 东道国的国民生产总值是另一影响跨国公司 R&D 投资国家选择的重要因素，东道国的经济规模和水平决定着一个国家的有效需求，人口的数量多少会影响有效需求，但就市场规模而言，人口数量多少并不决定跨国公司 R&D 投资，因此人口 (POP) 被删除了。实证结果表明：东道国的国民生产总值 (GDP) 与跨国公司 R&D 投资具有正相关性，与人口数量的多少相关性不大。东道国的经济规模越大，跨国公司就越会增加对该国的 R&D 投资，因此有效需求和潜在需求大的国家是跨国公司 R&D 投资的理想区域。

(3) 知识产权保护水平与每百万人从事 R&D 的科学家和工程师人数之间具有高度相关性，另外实证结果表明 RDPOP 的系数为负数，与理论假设不符。笔者认为知识产权保护比每百万人从事 R&D 的科学家和工程师人数更能反映东道国政策制度的指标，为了避免信息的重复反映，因此删除 RDPOP 因素。但由于知识产权保护水平的测量没有特定的标准，本文采用每千人拥有专利数作为知识产权保护水平的测度标准，带有一定的主观成分，因此，对待该结论应遵循谨慎原则。

(4) 从直观上看，工资水平的高低是影响跨国公司 R&D 投资的主要因素，就“三极”地区 (美国、欧盟、日本) 而言，这些国家的工资水平普遍很高，但实际情况是“三极”地区是跨国公司 R&D 投资较大的国家。这说明工资水平不是影响跨国公司 R&D 投资的主要因素。这可能是 R&D 是一种高智力、高知识的投资活动，与 FDI 相比，其投资规模相对较小，如果东道国在其他方面具有吸引力的话，跨国公司不会因为高工资水平而放弃 R&D 投资。同样，删除每千人拥有计算机数这个因素并不意味着一个国家拥有计算机数不重要，只是每千人拥有计算机数不是影响跨国公司 R&D 投资的主要因素。从数据分析我们知道，尽管每千人拥有计算机数，瑞士 (421.3 台)、澳大利亚 (411.8 台)、中国台湾 (455.2 台)、新加坡 (399.5 台) 都超过了德国 (255.9 台)，但是这 4 个国家的接受海外 R&D 投资金额与德国相比相距甚远。

R&D 本身是一个比较复杂的过程，而跨国公司 R&D 投资的国家选择更是一个系统化过程，其中所涉及的因素很多，这些因素又交织在一起，使得影响跨国公司 R&D 投资的国家选择的原因多种多样。通过实证分析，笔者提炼出一些关键的因素，以期对中国制定

吸引跨国公司 R&D 投资的政策提供理论支持。

四、跨国公司在华 R&D 投资现状及对策

近年来,伴随着世界经济和科技全球化进程的加快,越来越多的跨国公司到中国投资技术密集型产业,中国正在成为跨国公司竞相追求技术密集型产业的投资场所。自 1994 年加拿大北方电讯公司在北京投资成立北京邮电大学 - 北方电讯电信发展研究中心以来,跨国公司在中国设立 R&D 机构的发展趋势日益迅猛。已有 IBM、英特尔、太阳微、宝洁、杜邦、诺基亚、爱立信、松下、罗纳郎克、微软、GM、联合利华等世界著名跨国公司相继在中国设立 R&D 机构。另外,法国的罗纳普朗公司甚至在北京、上海建设了“科技园”,爱立信公司还投资建立了“爱立信中国学院”,跨国公司 R&D 机构的快速涌入使其成为我国 R&D 体系中一个不容忽视的组成部分。

(一) 跨国公司在华 R&D 投资的特征

据统计数据显示,到 2002 年底,跨国公司在华至少设立了 96 家规模较大的 R&D 机构。

1. 跨国公司在华设立的 R&D 机构多集中于北京、上海等发达城市

从海外 R&D 机构在中国的区位布局看,跨国公司 R&D 机构大部分集中在北京、上海、广东(深圳、广州)和江苏(南京、苏州)等沿海大城市,尤以北京和上海最为集中。到 2002 年底,在北京设立 R&D 机构的大型跨国公司有: SUN、诺基亚、松下、微软、诺克韦尔、富士通、贝尔、INTEL、德尔福、惠普等。据上海报导,在 2001 年头 10 个月已有 406 家外国投资公司进入上海,大量跨国公司的 R&D 机构在上海安家落户,包括 20 个国家级中心。在亚太经合组织上海会议结束后的一个月,7 个跨国公司,包括 IBM、NESTLE、MICROSOFT 在上海建立了 R&D 中心。跨国公司 R&D 机构在中国的分布高度集中于沿海省市(见表 2): 96 家跨国公司的 R&D 投资区域较为集中, R&D 机构主要设在北京、上海、广东等地。北京有 49 家、上海 27 家、广东(深圳、广州) 7 家、江苏(南京、苏州) 6 家、西安 3 家、青岛 2 家、成都 1 家、杭州 1 家、南通 1 家,这 11 个城市的国外 R&D 机构共 96 家。

这些地区特点是: 外商直接投资分布集中(FDI 累积额达 1345 亿美元,占 53.4%); 技术水平发达(R&D 经费 283 亿元,占 47.2%; R&D 机构人数 276256,占 34.2%; R&D 机构科技活动课题 37162 个,占 43.6%); 市场规模大(GDP 为 119066 亿元,占 26.9%); 知识产权保护水平高(国内专利申请受理为 36221 件,占 32.9%)。以 3 个城市为例,北京、上海、深圳是跨国公司 R&D 投资的重点城市。这 3 座城市的经济实力、科技实力、开放程度和知识产权保护政策环境是中国最好的。而且,这些城市拥有较先进的基础设施和技术配套设施,交通、信息流通方便,可以随时了解国际最新技术动态。

表 2 跨国公司 R&D 机构在中国大陆的主要分布

主要城市	R&D 机构数	FDI 累积额 (亿美元)	R&D 经费 (亿元) (1999)	R&D 机构人数 (1999)	GDP (亿元) (1999)	R&D 机构科技活动课题 (个) (1999)	国内专利申请受理 (件) (1999)
北京	49	101 (4%)	121 (20.2%)	148980 (18.5%)	11485 (2.6%)	22357 (26.2%)	7723 (7%)
上海	27	207 (8.2%)	51 (8.5%)	60097 (7.4%)	20999 (4.7%)	5622 (6.6%)	4605 (4.2%)
广东(深圳、广州)	7	687 (27.3%)	65 (10.8%)	25087 (3.1%)	45262 (10.2%)	3639 (4.3%)	16802 (15.3%)
江苏(南京、苏州)	6	350 (13.9%)	46 (7.7%)	42092 (5.2%)	41320 (9.3%)	5544 (6.5%)	7091 (6.4%)
合计	89	1345 (53.4%)	283 (47.2%)	276256 (34.2%)	119066 (26.9%)	37162 (43.6%)	36221 (32.9%)
全国	96	2517 (100%)	600 (100%)	806910 (100%)	443224 (100%)	85224 (100%)	109958 (100%)

说明：“FDI 累积额”指 1995~2000 年主要省市外商直接投资的累积额。

资料来源：根据外经贸部发布规模较大的跨国公司在华 R&D 机构名单；科技部在全国性统计调查取得的 R&D 机构名单和有关数据；科技部和北京市科委委托长城战略研究所跟踪取得了 2001 初至 2002 年新建 R&D 机构；《国际商报》2001 年 6 月 13 日；《光明日报》，2000 年 5 月 27 日；江小涓：《积极吸引外资，促进我国高新技术产业发展》，提供给外经贸的研究报告，2000；《中国科技统计年鉴》(2000)，《中国统计年鉴》(1995~2000)制作。

2. 在华建立 R&D 机构的跨国公司多数为欧美国家和日本

从海外 R&D 机构的来源看，来自 10 多个国家和地区的跨国公司在中国设立了 96 家 R&D 机构（见表 3）。其中，美国在华 R&D 机构数量最多，达到 36 家；其次是欧洲 26 家、日本 22 家，共占总数的 91.7%。其他为：中国台湾 5 家、韩国 4 家、加拿大 2 家、中国香港 1 家，分别占总数的 5.2%，4.2%，2.1% 和 1.1%。

表 3 在华跨国公司 R&D 机构来源

国别	美国	日本	德国	法国	中国台湾	英国	韩国	瑞典	合计
R&D 机构的数	36	22	6	6	5	4	4	3	96
国别	丹麦	芬兰	加拿大	荷兰	瑞士	香港			
R&D 机构的数	2	2	2	2	1	1			

资料来源：根据外经贸部发布规模较大的跨国公司在华 R&D 机构名单；科技部在全国性统计调查取得的 R&D 机构名单和有关数据；科技部和北京市科委委托长城战略研究所跟踪取得了 2001 初至 2002 年新建 R&D 机构；《中国科技统计年鉴》(2000)制作。

美国、欧洲、日本是在华设立 R&D 机构的主要国家，其中又以美国为最多。美国企业进入中国的市场时间较长，对中国的投资已经相当成熟，而且美国与中国距离较远，文

化差异较大，因此更需要在中国本土设立 R&D 机构以支持中国的市场。事实上，在我国设立 R&D 机构最多的跨国公司所属的国家，也正是在中国直接投资最大国家（见表 4）。以跨国公司在华 R&D 机构的数量为衡量标准，我们可以观察到这样一种趋势：跨国公司 R&D 机构在中国的发展与各国在中国的直接投资呈正相关。跨国公司来华设立 R&D 机构主要是为本国在中国的直接投资服务。

表 4 跨国公司在华 R&D 机构数量与“三极”地区对华直接投资相关性

国别	美国	日本	欧盟（英、德、法）
R&D 机构的数 （名次）	36 (1)	22 (2)	16 (3)
对华直接投资（亿美元） （名次）	248 (1)	239 (2)	158 (3)

说明：“对华直接投资额”指截至 1999 年对华直接投资总额。

资料来源：根据《中国统计年鉴》（1990~1999）年制作。

3. 跨国公司在华 R&D 机构的行业分布

跨国公司在华 R&D 机构涉及计算机、通讯、电讯、生物医药、精细化工、汽车、光机电一体化、石油勘探等研究领域，根据对 96 家跨国公司在华 R&D 机构的研究，从事 IT 行业（计算机、通讯、电讯）的 R&D 机构有 70 家，占总数的 72.9%；其次为精细化工 8 家（8.3%）、生物医药 8 家（8.3%）、汽车 7 家（7.3%），另外光机电一体化与石油勘探分别为 2 家（2.1%）和 1 家（1.0%）（见表 5）。

表 5 跨国公司 R&D 投资在华的行业分布

	R&D 机构数	通讯	电讯	计算机	生物医药	精细化工	汽车	光机电一体化	石油勘探
北京	49	17	7	14	6	1	1	2	1
上海	27	7	4	4	2	6	4		
深圳	5	2	2	1					
广州	2		1				1		
苏州	3		1	2					
南京	3		2				1		
杭州	1	1							
西安	3	1	1	1					
成都	1		1						
青岛	2			1		1			
总计	96	28	19	23	8	8	7	2	1

资料来源：根据外经贸部发布规模较大的跨国公司在华 R&D 机构名单；科技部在全国性统计调查取得的 R&D 机构名单和有关数据；科技部和北京市科委委托长城战略研究所跟踪取得了 2001 初至 2002 年新建 R&D 机构；《中国科技统计年鉴》（2000）（附件 4）制作。

跨国公司在华 R&D 机构的研究领域以 IT 业（计算机、通讯、电讯）为主，计算机、通讯、电讯等信息产业是 21 世纪的支柱产业，同时它也是研发力度最大的行业，这一行业的技术升级速度之快是其他行业所无法比拟的。

（二）吸引跨国公司来华 R&D 投资的对策

通过分析，我们知道跨国公司在新兴市场上的 R&D 投资正呈现出广泛地理位置上的分散，跨国公司 R&D 机构国际化不再是发达国家的专利。有证据表明，跨国公司 R&D 投资会发生溢出效应，东道国公司的技术效率会随着跨国公司竞争的加剧而提高，中国虽不可能在所有技术领域吸引跨国公司 R&D 投资，但在某些领域大规模吸引跨国公司 R&D 投资并不是不可能的，关键是中国应充分了解自身的区位特征，构建具有国际竞争力的区位，塑造出吸引跨国公司 R&D 投资的区位优势，通过制定出有效的政策，集中有限资源培养出有优势的区位因素。在区位布局上，坚持“分类指导、梯度推进”的原则，遵循非均衡发展的客观规律，从实际出发，因地制宜，继续发挥区域优势。吸引更多的跨国公司来华设立 R&D 机构，在本地实现技术化，把先进的运作模式和管理方式带入中国。

随着中国经济持续高速增长，跨国公司 R&D 机构在中国的发展将进入高速增长阶段，即 R&D 机构数量和投资额呈逐渐增长的趋势。并且新创立的 R&D 机构普遍会起点高，投资规模大。而这些 R&D 的选址则会依旧在北京、上海、深圳等经济发达城市。跨国公司 R&D 机构高度集中于这些城市，同外资企业在这些区位的直接投资分布高度吻合，同这些区位的技术水平和经济实力高度吻合。中西部各省市应大力吸引外商直接投资，从而为跨国公司到中西部设立 R&D 提供契机。如果外资项目不多，跨国公司为节约成本，就不会在中西部设立 R&D 机构。另外，跨国公司不会像在沿海省市那样广泛设立 R&D 机构，只会在重庆、成都及西安三大城市设立 R&D 机构。对于到中西部设立 R&D 机构的跨国公司，国家可提供更优惠的待遇，如税收减免等。

从以上实证分析看，生产规模的程度、东道国的资源特征以及知识产权保护政策对跨国公司 R&D 投资的具有重大影响，结合跨国公司 R&D 投资的内部经济诱因分析，提出以下政策建议：

第一，区位布局上，遵循非均衡发展的原则。从跨国公司 R&D 投资在中国的发展趋势分析，在区位布局上应遵循非均衡发展的原则，吸引跨国公司在沿海省市广泛设立 R&D 机构，适度吸引跨国公司到中西部设立 R&D，西部中心城市西安、重庆、成都应在比较优势领域吸引跨国公司设立 R&D 机构，例如，资源深加工、汽车和化工等行业。而由北京、上海、深圳，由东向西组成的“金三角”，正像硅谷、班加罗尔、新竹成为全球的“创新高地”一样，正在构成中国的创新区域。因此对这三座城市重点领域的研究是十分必要的。这一结论与跨国公司在中国 R&D 机构的区位分布高度吻合，“金三角”地区的经济水平、科技水平最高，因而吸引的跨国公司 R&D 机构最多，集聚形成的马太效应又会加速这一进程。可以说，在信息经济时代，北京、上海、深圳会成为中国最高层次的全球性技术创新的源头。而任何一座城市都有其自身的特点，在资源禀赋、重点吸引 R&D 投资行业、拓展空间等方面都有所不同。不同区域在行业分布、国别分布上应形成分工体系。

第二，吸引外商直接投资的重新定位。在一个国家大规模外商直接投资进入的情况下，一般而言，FDI 会成为东道国吸引跨国公司 R&D 投资的主要因素。在其他因素保持不变的情况下，FDI 规模扩张为吸引跨国公司 R&D 投资提供了更多机会，子公司为适应当地市场的生产需求，大多会设立 R&D 机构来支持它们的生产，适应当地市场的差异性成为吸引海外 R&D 投资的主要动机，尤其是在发展中国家。中国应从引进 FDI 以弥补资金缺口为重点转向通过引进 FDI 以吸引 R&D 投资，从而获取外资的技术外溢效应。中国今后应将引进 FDI 的功能定位在技术创新上，以吸引更多的跨国公司来华进行 R&D 投资。在吸引外商直接投资的同时，一个国家有较大的市场规模更会成为吸引跨国公司 R&D 投资的因素。市场规模与经济发展速度及经济发展水平有关。从目前情形看，世界经济持续低迷，国际经济环境仍存在很大的不确定性，而中国经济却“一枝独秀”，依然保持 7% 以上的增长速度，中国良好的宏观经济环境决定了市场的有效需求，使得市场容量呈现不断扩大的趋势。因此，在国外需求不足的情况下，进一步扩大开放和竞争领域，继续刺激外商来华进行直接投资，同时，通过制定一系列优惠政策，如在项目申请、土地使用权以及税收等方面享有优惠，这无疑会对跨国公司的 R&D 投资产生巨大的吸引力。

第三，完善科技资源和基础结构。通过检验变量 PPP 分别与 RDPOP、RDGDP 之间的相关性分析。发现 $\text{Corr}(\text{PPP}, \text{RDPOP}) = 0.866267994584$ ， $\text{Corr}(\text{PPP}, \text{RDGDP}) = 0.784566041743$ ，PPP 与 RDPOP、RDGDP 之间具有高度相关性，这说明具有良好的知识产权保护环境的国家，R&D 经费往往投入大，拥有从事 R&D 的科学家和工程师人数多。而 R&D 经费投入、R&D 机构人数多寡是衡量一国科技实力大小的重要指标，因此中国应加大这些科技基础资源的投入力度。只有科技资源和基础结构得到完善，才能吸引跨国公司来华进行 R&D 投资。据有关资料显示，自 1965 年以来，美国工业部门每年投入到 R&D 活动中的经费一直高达数百亿乃至上千亿美元，这就不难理解为什么美国的科技实力在世界范围内长期处于领先地位。

从表 6 分析，中国的 R&D 经费支出明显不足，在世界范围内 R&D 占 GDP 比重处于下游水平，1997 年的 R&D 占 GDP 比重仅为 0.64，与美国、日本、德国等发达国家（2% ~ 3%）相比有较大的差距，与以色列、韩国、新加坡、中国台湾等新兴市场相比也有不小差距，中国应加大 R&D 资金的投入。

表 6 1997 年 R&D 经费及占 GDP 比重的国际比较

国 家	中国	美国	日本	荷兰	法国	德国	丹麦	瑞典
R&D 占 GDP 比例	0.64	2.71	2.92	2.09	2.23	2.31	2.03	3.85
国 家	以色列	中国台湾	新加坡	韩国	南非	墨西哥	阿根廷	印度
R&D 占 GDP 比例	2.4	1.92	1.47	2.89	0.7	0.31	0.4	0.05

资料来源：《中国科技统计年鉴 2000》

第四，重视人力资源的配置与使用，加强对科学技术人员的培训。跨国公司在对 R&D 大量资金投入的同时，十分重视人力资源的配置与使用，人才高层化战略是跨国公

司在 R&D 活动中的主要特点。一些技术先进国家在不断增加其 R&D 资金投入的同时,还大力加强人力资源建设。事实上,在激烈的市场竞争中,跨国公司十分重视对高级人才的争夺。在这方面,日本的表现尤为突出(见表 7)。

如表 7 所示,在世界各国的科技队伍中,中国从事 R&D 活动的科技人员远低于美国、日本、英国等发达国家,也低于新加坡、韩国等新兴工业化国家,只略高于越南、墨西哥。科技人才是吸引跨国公司 R&D 投资的重要因素。只有从事 R&D 的科学家的数量和质量得到普遍提升的条件下,跨国公司在中国的 R&D 投资才会持续地发展。因此,中国应加大对从事 R&D 科学家的培养,建立起尊敬知识、崇尚科学的良好环境。

表 7 百万人口中从事 R&D 科技人员的国际比较

国 家	中国	日本	美国	加拿大	法国	意大利	澳大利亚	英国
R&D 人数	350	6309	3732	2656	2659	1318	3357	2448
国 家	中国台湾	南非	韩国	新加坡	以色列	墨西哥	泰国	越南
R&D 人数	3522	1031	2638	2728	2638	213	119	308

资料来源: World Bank:《知识与发展:21 世纪新追赶战略》,1998,第 208 页。

第五,加强知识产权保护,鼓励企业、大学、科研机构 and 跨国公司的合作。东道国的技术规则和政策是吸引跨国公司 R&D 投资的一个重要因素,其中良好的知识产权保护机制尤为重要。中国加入 WTO 以后,研发活动要遵循一些大家共同遵守的规则,中国应加大对知识产权保护的力度,促使企业、科研机构、大学普遍建立和健全知识产权保护制度,强化知识产权保护意识,以保护外国投资者的合法利益。鼓励企业、大学、科研机构和跨国公司的合作,基于 R&D 具有创造性和外溢性的特点,海外 R&D 投资对东道国的专利保护十分敏感,因此对知识产权保护有更强的需求。通过对美国跨国公司海外 R&D 投资的检验,分析出知识产权保护会对跨国公司 R&D 本地化产生影响。中国除了要制定完善的知识产权保护法律外,还应加大严格执行知识产权保护法律的力度。如果对知识产权保护执行不利,一方面,国内的 R&D 成果会被海外跨国公司轻易获取,从而有损中国的技术创新;另一方面,跨国公司 R&D 成果易被竞争对手模仿、窃取,损害跨国公司 R&D 机构的利益,不利于跨国公司来华设立 R&D 机构。因此,中国入世后,立法重点应尽快转向规范无形市场,具体说就是转向知识产权保护上。只有加强了知识产权保护,原创性 R&D 成果才会更多,这样才能吸引更多的跨国公司与中国企业、大学、科研机构进行 R&D 合作。

第六,鼓励科技人员组建 R&D 公司,条件成熟的公司到海外创立 R&D 机构。国有企业是 R&D 发展的主体,但有的国有企业亏损极为严重,根本没有能力进行 R&D 投资,所以应加大 R&D 机构的市场化力度,鼓励科技人员组建 R&D 公司,创办直接面向社会的市场化的 R&D 服务企业和盈利性科研机构,改变创新效率低下和竞争性企业 R&D 活动强度不足的现状,使得科研成果与生产实践有机地结合起来,同时国家应给予这些企业同样的优惠政策,激发起国内企业的 R&D 能力。扶植条件成熟的公司到海外创立 R&D 机构,也

正是由于我们自己的技术落后,就更有必要到技术发达的国家学习先进经验。已在海外投资生产的劳动密集型企业,应在东道国建立 R&D 机构,以根据当地的需求特征和消费习惯,设计出适合当地市场的产品,进一步扩大市场占有率。对于技术密集型企业,在海外设立 R&D 机构,以从竞争对手的技术外溢中获益,提升本企业的技术水平。

五、结束语

现在让我们用简短的注解总结这篇论文,从长期的动态趋势分析,中国呈现出与跨国公司 R&D 投资一般趋势趋同的特征,跨国 R&D 投资正在受到我国企业和跨国公司的重视,可以预期其份额将不断上升,跨国公司在华投资出现由制造环节为主转向制造与 R&D 并重的阶段。随着经济全球化和科学技术的一日千里,R&D 日益成为一国经济增长的关键因素,因此,吸引跨国公司 R&D 投资对中国发展有着重要的意义。跨国公司海外 R&D 投资高度依靠东道国提供必要的技术资源和当地技术资本的能力。跨国公司为保持发达国家和发展中国家之间的技术差距,大都把 R&D 投资聚集在发达国家。虽然中国的技术能力还很落后,但中国应在某些技术领域加大 R&D 投入。知识产权保护机制会显著地影响 R&D 投资,应加强对知识产权保护法律的完善和执行。最后,就市场规模而言,它是决定 R&D 区位的重要因素,作为东道国通过在国家领域扩展有效的市场规模和吸引外商直接投资来增加对 R&D 投资的吸引力。

参考文献

- [1] 长城企业战略研究所:《R&D 拥抱中国:跨国公司在华 R&D 的研究》,广西人民出版社 2002 年版。
- [2] 杜德斌:《跨国公司 R&D 全球化区位模式研究》,复旦大学出版社 2001 年版。
- [3] 邱立成:《跨国公司研究与开发的国际化》,经济科学出版社 2000 年版。
- [4] 肖静华:《IT 业跨国投资与国家竞争力》,中山大学出版社 2001 年版。
- [5] 江小涓,杜玲:《国外跨国投资理论研究的最新进展》,《世界经济》,2001 年第 6 期。
- [6] 林进成,柴忠东:《试析跨国公司技术研究与开放的国际化趋势》,《世界经济》,1998 年第 7 期。
- [7] 冼国明,葛顺奇:《跨国公司 R&D 的国际化战略》,《世界经济》,2000 年第 10 期。
- [8] 姚战琪:《跨国公司研发投资的经济分析》,《世界经济》,2001 年第 11 期。
- [9] 郑京淑:《跨国公司海外 R&D 机构及其区位选择》,《国际地理研究》,2000 年第 1 期。
- [10] Cantwell, J., *Historical Trends of International Patterns of Technological Innovation, New Perspectives on Late Victorian Economy*, Cambridge University Press, 1991.
- [11] Cantwell, J., *The Globalization of technology: What Remains of Product Cycle Model*, *Cambridge Journal of Economics*, 1995, Vol.19.
- [12] Cheng, J. and Bolon, D., *The Management of Multinational R&D: A Neglected Topic in International Business Research*, *Journal of International Business Studies*, 1993, Vol.24, pp1 ~ 18.
- [13] Caves, R. E., *Multinational Enterprise and Economic analysis*, *Cambridge Surveys of Economic Literature*, 1996.
- [14] D. H. Donald and Serapio. M. G., *Globalizing Industrial Research and Development*, U.S. Department of Commerce Technology Administration office of Technology Policy, 1999.
- [15] Hakanson, L and R. Nobel, *Foreign Research and Development in Swedish Multinationals*, *Res. Policy*,

1993 , Vol.22 , pp373 ~ 396.

- [16] Hewittl , Gary. , Research and Development Performed Abroad by U.S , Manufacturing Multinationals. *Kyklos* , 1980 , Vol.33 , pp308 ~ 327.
- [17] Kuemmerle , The Drivers of Foreign Direct Investment into R&D : An Empirical Investigation , *Journal of International Business Studies* , 1999 , Vol.130 , pp11 ~ 24.
- [18] Lall , S.The International Allocation of Research Activity by US Multinationals.*Oxford Bulletin of Economics and Statistics* , 1979 , Vol.41 , pp313 ~ 331.
- [19] Massimo , P. and Simone. G. , R&D Internationalisation in the Strategic Behaviour of the Firm , *STEEP DISCUSSION PAPER* , 1997 , No.39.
- [20] Manfield , E. D. Teece and A. Romeo , Overseas R&D by U.S. Based Firms.*Economica* , 1979 , Vol.46 , pp187 ~ 196.
- [21] Pearce , R. D. , The Internationalization of Research and Development by Multinational , Enterprises.London : McMillan , 1989.
- [22] Zejan , Mario C. , R&D Activities in Affiliates of Swedish Multinational Enterprises , *Scandinavian Journal of Economics* , 1990 , Vol.92 , No.3 , pp487 ~ 500.

外商直接投资聚集效应在我国的实证分析*

许罗丹 谭卫红

一、理论文献回顾

自改革开放，特别是 90 年代以来，外商在我国的直接投资有了飞速的增长。到 2000 年，我国累计吸引外商直接投资 3400 多万美元，年平均增长率高达 13 %。我国成为吸引外商直接投资最多的发展中国家。同时，我国吸引的外商直接投资主要是来自于香港、台湾、澳门及其他华人华侨的投资；虽然近年来华资所占的比重有所下降，但还是其中的重头之戏。这与国际资本市场主要是发达国家的投资有些差别。为此，研究我国的外商直接投资成为经济学的重点，他们从不同的角度说明了外商直接投资选择我国的动因。在国外，大量有关中国外商直接投资区位因素的文献不断涌现；在国内，也有不少学者对这一问题进行过比较深入的研究。然而，这些研究大多是综合性的，没有考虑到不同投资类型和投资来源国之间的差别，结合聚集效应进行实证分析的则更少。

聚集经济是指由于经济活动和相关生产设施的区域集中而形成的正外部性以及规模和范围经济 (Smith and Florida, 1994)。对聚集经济的研究，可以追溯到 100 多年以前。当时 Marshall 指出了聚集效应对工业区位选择的重要性。“当某一工业定位于某一区域时，就极有可能定位于此。同类厂商彼此相邻并从事类似的经济活动能够产生巨大的利益，厂商也倾向于选择在某种特定技能的劳动力集中的区域设厂，这种就业上的优势同工业在特定区位的集中组合形成制造业的聚集效应”。韦伯 (1909) 从企业最小生产成本出发，认为费用最小的区位是最好的区位，而聚集能使企业获得成本节约。Caves (1971) 考虑交易成本的高低，通过不确定性分析认为，与东道国企业相比，外国投资者在所处的投资环境中处于一种“外来身份”劣势，在原材料购买、识别和发现市场机会、熟练劳动力来源、管理以及许多不可预见到的不确定性等方面，比东道国企业有着更高的信息成本，理性的外资企业会做出风险回避的区位选择，外资企业往往会考虑其他外国投资企业的选择，在已吸引外资较多的国家进行投资。Kickerbocker (1973) 建立 A、B 两国相互竞争模型，在一国首先向第三国投资的时候，另一国也会在第三国进行投资，也就是所谓的“跟进策略” (follow-my-leader)，从而逐渐形成聚集效应。Dunning (1977) 提出“国际生产折衷理论”，认为区位优势取决于要素投入和市场的地理位置状况、各国的生产要素成本的质量、运输成本、基础设施等，而这些正是聚集效应的体现 (吴丰，2002)。Krugman (1991) 从区域经济与规模经济的角度，解释了工业活动的集中性，聚集效应的关键是规模经济，特别是

* 原载：《管理世界》，2003 年第 7 期。

外部规模经济；聚集能导致生产某一产品的平均成本下降，进而产生递增的规模效应。Dunning (1995) 认为：随着经济全球化的迅速发展，跨国公司的区位选择不仅要考虑传统的要素及其成本，又要考虑聚集经济和交易成本、动态外在经济性。Dermot Leahy, Stephen Pavelin (2003) 以企业竞争动机为考虑对象，建立 A、B 两国，1、2 两种商品模型，分析得出彼此国家之间相互投资的激烈竞争，相互的示范效应会导致同时在一个国家进行直接投资。

国内外学术界对聚集效应与 FDI 关系的实证研究也表明，聚集效应对吸引 FDI 和 FDI 的区位选择有显著影响。Luger 和 Shetty (1985) 通过对三位数产业（产业分类标准）的研究，证实了聚集经济对外国公司区位选择的重要影响。Head et al. (1995) 和 Kogut & Chang (1996) 研究表明，日本在美国的直接投资倾向于采用“跟进策略”，同时也重视早期的直接投资。Head 和 Ries (1996) 也发现以前投资对外商在我国直接投资区位选择的重要性。Barrell (1999) 发现，同特定产业相关的聚集效应对美国在欧洲的直接投资产生显著的影响。Head (1999) 根据日本厂商从 1980 到 1992 年对美国直接投资的研究中发现了同特定国家相关的聚集效应的强烈作用。Leonard K. Cheng 和 Yum K. Kwan (2000) 设定动态调整模型 $d \ln Y_{it} / dt = \alpha (\ln Y_{it}^* - \ln Y_{it})$ ($0 < \alpha < 1$)，推导出 $y_{it} = (1 - \alpha) y_{it-1} + \alpha y_{it}^*$ ；并以此模型为基础，分析了外商直接投资的区位决定因素。Linda F.Y. Ng 和 Chyau Tuan (2003) 的“扩展中心 - 外围理论”，研究聚集效应在外商直接投资对我国广东省投资的影响。

国内对区位选择和聚集效应的研究，则局限于简单的模型回归和聚集效应的介绍及分类。Kevin Honglin Zhang (1999) 在研究大量外资涌入中国的原因时，分析了影响来自亚洲资金与欧美资金的因素；结果发现亚洲资金比欧美资金更看重我国的工资成本，欧美资金更看重我国广阔的市场，对我国 GDP 的增长率反应更敏感。Leonard K. Cheng 和 Yum K. Kwan (1999) 在研究外商直接投资的区位选择时，运用中国的数据，分析得出外商直接投资具有很强的聚集效应，并且巨大的市场、良好的基础设施、经济特区政策对吸引外商直接投资有巨大的吸引力，而工资的影响不显著。孙俊 (2002) 在中国 FDI 地点选择的因素分析中，考虑优惠政策、开放水平、市场化程度对外商直接投资的影响，发现上述因素对各地区吸引外商直接投资都具有正的影响。吴丰 (2003) 研究外商直接投资的聚集效应和我国吸引外资的新趋向，从结构化的聚集效应分析框架，透视了影响 FDI 区位选择的因素，着重讨论了聚集效应与 FDI 的关系。

本文在借鉴上述文章的基础上，发展 Leonard K. Cheng & Yum K. Kwan (2000) 的动态模型，结合邓宁的区位优势理论，分析了外商直接投资聚集效应对我国吸引外资的影响；并就华资（香港、澳门、台湾）与美国直接投资进行对比分析。本文的第二部分简要介绍我国引进的外商直接投资状况；第三部分是理论模型的设定；第四部分则根据理论模型设定计量模型，分析效应的影响，各种变量对外商直接投资的作用；最后则就计量分析的结果做简单的分析。

二、我国引进的外商直接投资状况

从投资来源看,我国所吸引的外商直接投资可以分为港澳台资本和外国资本两部分。外国资本中以日、美、德、法、英等西方发达国家为主,发达国家的投资又以美国投资为代表。其中港澳台资本一直以来占据我国吸引外商直接投资的第一位;来自发达国家的投资则相对较少。1992年华资所占比例达到最高点79.4%;2000年比例降到最低,占36.8%,但其总量仍呈增长趋势。从表1中我们也可以看出,华资更易受到外界的影响。受1997年金融危机的影响,华资无论是投资总量还是所占投资比例,在接下来的几年都呈下降趋势。而美资仅在1997年小幅度下降以后,从1998年开始又呈增长趋势。在发达国家对我国的直接投资中,美国占重要地位,从整体存量来看,1977~2000年居第一位,并且其所占比例有所上升。如图1、图2所示。

表1 FDI来源地区分配表

单位:百万美元

地区/国家	1977~1991			1992~2000		
	名次	FDI流入量	百分比	名次	FDI流入量	百分比
华资	1	13407	58.03	1	186163	57.58
美国	2	2382	10.31	2	27343	8.46
日本	3	1740	7.53	3	24779	7.66
新加坡	6	628	2.72	4	23149	7.16
韩国	12	51	0.22	5	8005	4.97
英国	4	862	3.73	6	5830	3.42
德国	5	655	2.88	7	3332	2.49
加拿大	9	146	0.63	8	2046	1.42
泰国	10	118	0.51	9	1876	0.87
法国	7	305	1.32	10	1620	0.80

资料来源:《中国统计年鉴》。

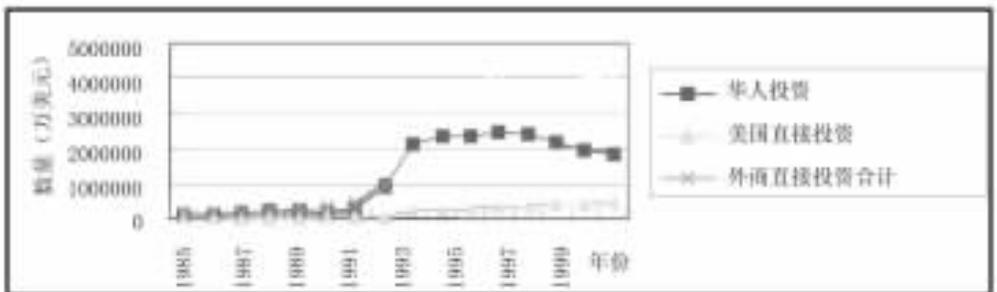


图1 外商直接投资流量

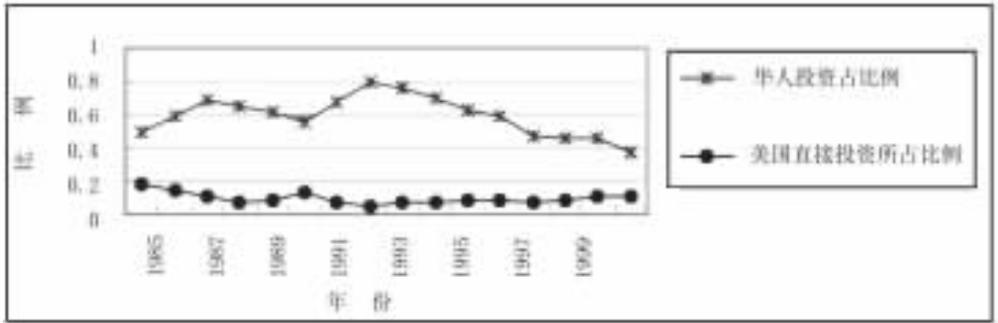


图2 华资、美资所占比例

从市场导向战略来看，由于各类资本所包含的技术和工业化水平存在阶段性差异，同时还与我国的发展战略密切相关，在外资引进过程中，逐步形成了港澳台资本的出口导向型和外国资本的国内市场导向型两种模式。王岳平（1999）利用出口率（出口额占销售收入的比重）指标，并结合相对劳动密集程度指数与相对技术水平指数，分析发现我国“三资”企业的发展确实存在出口导向和国内市场导向两种专门化类型。“三资”企业中，以港澳台资本为主的行业具有明显的出口导向特征，如港澳台资本参与度较高的两组其出口率高达 42.6% 和 65.03%；而以外国资本为主的行业具有明显的国内市场导向特征，外国资本参与度最高的一组其国内市场导向最为突出，其内销率高达 86.31%。魏后凯（2001）也分析发现，来自北美洲和欧洲的外商投资通常是针对中国市场而来的，而亚洲国家和地区一般把中国大陆作为生产基地和出口平台；从投资战略来看，欧美国家在华制造业投资更多地看重中国巨大的国内潜在市场，开发和占领当地市场是他们来华投资的重要目标之一。

基于华资与美国直接投资在我国引进外商直接投资中的重要性，而这两种投资代表了不同类型国家的投资和不同的投资导向，因此本文除分析我国引进外商直接投资的整体情况外，又分别分析了华资与来自美国的直接投资。本文为了分析问题的方便以及数据约束，在分析华资时，只是考虑来自香港、澳门、台湾的华人投资。

三、模型设定与变量选择

外商直接投资的聚集效应体现在增量 FDI 的区位选择受在特定区位的 FDI 存量的影响，即当某一区域的外商资本控制的厂商越多，新的外商将更倾向于投资该国家或区域。作为一种高风险和生产要素全面卷入的投资行为，FDI 的决策是相当谨慎的。由于需要对东道国的市场、法规、经济、文化等因素有较为全面的了解，FDI 通常面临较高的信息搜索成本，FDI 的实施也会面临较高的交易成本。针对这种不确定性和高成本，采取所谓“跟进策略”是一种好的选择。同时，大量研究也表明，聚集经济对外资制造业工厂的区位选择有着重要的正的影响（Head, Ries and Swenson, 1995）。这种聚集经济不仅体现在城

市规模经济效益方面,而且也体现在行业规模经济效益方面。大量的相关企业在地域上相互集中在一起,可以节约生产成本,扩大生产和消费需求,并有利于相互竞争和协作,提高管理和办事效能。

本文在现有文献的基础上,同时根据邓宁的区位优势理论,以及聚集效应分析,设定局部调整模型,根据中国近年来的数据,运用计量经济方法进行实证分析和检验,分析 FDI 的聚集效应和其他影响 FDI 水平的因素。本文现介绍一下局部调整模型:

局部调整模型的一个主要方程与 y_t 的理想水平 y_t^* 有关,这个理想水平为:

$$y_t^* = \alpha + \beta X_t$$

实际水平的调整方程为:

$$y_t - y_{t-1} = (1 - \lambda)(y_t^* - y_{t-1}) + \epsilon_t$$

也即当前实际水平的调整同理想水平与上一期的实际水平的差成比例。由上面的模型可得:

$$y_t = \alpha(1 - \lambda) + \lambda y_{t-1} + \beta(1 - \lambda)x_t + \epsilon_t$$

在局部调整模型思想的基础上,我们设定影响 FDI 存量的动态模型。假定 FDI_{it} 为我国截止 t 年从 i 国(地区)引进的 FDI 总量, FDI_{it}^* 是我国相应的 FDI 理论均衡需求量(本文因为数据的限制,FDI 的计算从 1985 年开始,数据统计年份为 1985~2001 年)。

$$\begin{aligned} FDI_{it}/FDI_{it-1} &= (FDI_{it}^*/FDI_{it-1})^\lambda \\ 0 < \lambda < 1 \end{aligned} \quad (1)$$

也就是说,一国引进的 FDI 总量的百分比变化与 FDI_{it}^* 与 FDI_{it-1} 的比成比例。又 $FDI_{it} = (FDI_{it}^*)^\lambda (FDI_{it-1})^{1-\lambda}$, 可得:

$$\ln FDI_{it} = \lambda \ln FDI_{it}^* + (1 - \lambda) \ln FDI_{it-1} \quad (2)$$

(2) 式等式右边体现了 FDI 的聚集效应,也就是已引进的 FDI 吸引更多的 FDI 的能力。聚集效应意味着 FDI_{it} 是 FDI_{it}^* 的正函数,并且如果无论 FDI_{it} 的水平如何,聚集效应都很强的话,在没有一般均衡约束的条件下,如资源约束、外部经济约束等, FDI_{it}^* 将会是 0 或无限大。但在我们的局部调整均衡模型中,我们假定 FDI_{it} 影响自己未来的量,而不是影响 FDI_{it}^* 。也就是说, FDI_{it}^* 受其他变量的影响,有自己的均衡水平。

根据邓宁的区位优势理论,影响一国引进 FDI 的因素包括:①投资东道国的经济发展情况;②投资东道国的国内市场容量;③当地的基础设施建设;④当地的工资水平;⑤当地政府对 FDI 的优惠政策。由此,FDI 的引进数量是一系列变量的函数,即:

$$\ln FDI_{it}^* = \alpha X_{it} + \epsilon_{it} \quad (3)$$

把 (3) 式代入 (2) 式,可得:

$$\ln FDI_{it} = (1 - \lambda) \ln FDI_{it-1} + \lambda \alpha X_{it} \quad (4)$$

其中: x_{it} 为市场容量、经济发展情况、基础设施、工资水平、优惠政策的组合。

邓宁的区位优势理论同时又指出,根据 FDI 的不同类型,相同的因素对不同的投资类型的影响是不同的。例如,出口导向型的 FDI 更看重当地低廉的生产及运输成本;而市场导向型的 FDI 则更注重当地市场的潜力和当地居民的需求。根据邓宁的这一理论,本文在分析我国整体情况的基础上,又对来自香港、澳门、台湾的华资与来自美国的直接投资分

别作了分析, 试图发现其不同的影响程度。

投资东道国的国内市场容量用外商直接投资平均的国民生产总值衡量; 经济发展情况则用我国 GDP 的水平来衡量; 对我国基础设施建设的衡量, 本文选取了我国的交通运输线路综合密集度, 即 (铁路营运里程 + 公路里程 + 内河航道里程) / 国土总面积。一项研究结果表明, 以全国 30 个省、市、自治区的基础设施和人均国民收入为样本, 分别计算了交通运输等七类基础设施与人均国民收入增长之间的相互关系, 其中邮电通讯的相关系数最高, 其次是交通运输, 其相关系数都在 0.9 以上。^① 所以本文选取交通运输线路综合密集度作为我国基础设施的代理变量。

我国对外商直接投资的优惠政策集中体现在税收优惠、信贷放宽、优先获得进出口权及外汇使用的优惠政策等方面。这些政策的实施, 主要通过建立经济特区、经济技术开发区、沿海开放城市、沿海开放地区等战略进行。所以就我国政府针对 FDI 的优惠政策, 选用我国设立的经济特区和浦东新区、经济与技术开放区、沿海开放城市、沿海开放地区作为代理变量。研究表明, 这种优惠政策有助于吸引外商直接投资 (Head and Ries, 1996; Wietal, 1999)。根据我国在不同地区实行不同程度的吸引 FDI 的优惠政策, 确定不同的经济类型的权数。五个经济特区和浦东新区的优惠政策最高, 权数为 4, 经济与技术开发区为 3, 14 个沿海开放城市为 2, 沿海开放地区为 1。经过累加计算, 得到历年的优惠政策指数。

四、数据来源与实证分析结果

有关 FDI、消费水平、GDP、基础设施、工资的数据来自于《中国统计年鉴》。有关政策优惠的相关资料来自于《地理位置与优惠政策对中国地区经济发展的相关贡献》和《中国改革报》(2001 年), 并通过计算获得。

根据第三部分的理论分析, 我们设定计量回归方程如下:

$$\ln(\text{FDI}_{it}) = C + (1 - \lambda) \ln(\text{FDI}_{it-1}) + \alpha\lambda \ln(\text{GDP}_{it}/\text{FDI}_{it}) + \alpha\lambda_2 \ln(\text{GDP}_{it}) + \alpha\lambda_3 \ln(\text{INFRA}_{it}) + \alpha\lambda_4 \ln(\text{WAGE}_{it}) + \alpha\lambda_5 \ln[\text{POL}(-1)] + \alpha\epsilon_{it}$$

其中: i 表示国别地区; 下标 t 表示时间。

$\ln(\text{FDI}_{it})$: 我国 FDI 存量水平的自然对数值, 代表 FDI 存量水平的增长速度。本文 $\ln(\text{FDI}_{it})$ 为被解释变量, 其余变量为解释变量。

$\ln(\text{FDI}_{it-1})$: 工作年份前一年 FDI 存量水平的对数, 其系数代表 FDI 的聚集效应, 与被解释变量成正相关关系, 以表示 FDI 有正的聚集效应。

$\ln(\text{GDP}_{it}/\text{FDI}_{it})$: 我国市场容量。从理论上讲, 市场导向型应该看重这一因素, 系数应为正; 出口导向型投资则不能确定。

$\ln(\text{INFRA}_{it})$: 我国的基础设施水平。基础设施包含多个方面, 对一国的经济发展有重要影响, 对吸引外商直接投资也应有正的影响。本文因为数据问题以及对一些基础设施

^① 《我国交通基础设施建设与区域协调发展》, 《铁路运输与经济》, 2002 年 10 月。

的量化问题，只选取了交通方面。

$\ln(WAGE_{it})$: 工资对投资者来说是一项成本，对吸引外商直接投资应有负的影响。

$\ln[POL(-1)]$: 针对 FDI 的优惠政策，对不同的投资类型有不同的影响。因为国家政策产生效用的滞后性，模型中选取政策的滞后一期作为因变量进行回归。

本文研究选用对数形式。采用对数形式模型有它自身的优点，如回归系数是外商直接投资规模相对于其他变量特性的弹性值，可以将可能的非线性关系转化成线性关系，减少变量的极端值等。回归结果如表 2 所示。

表 2 包含工资因素的回归结果

	FDI	HFDI	AFDI
C	- 8.960* (3.346)	- 7.399* (1.959)	- 22.665* (3.890)
$\ln[FDI(-1)_i]$	0.549* (0.105)	0.604* (0.052)	0.181** (0.167)
$\ln(GDP)$	0.391* (0.215)	0.264** (0.152)	0.619** (0.319)
$\ln(INFRA)$	0.746* (0.247)	0.687* (0.176)	1.672* (0.260)
$\ln(GDP/FDI)$	- 0.176** (0.062)	- 0.383* (0.030)	0.283* (0.106)
$\ln[POL(-1)]$	0.118** (0.130)	0.080** (0.075)	0.738* (0.197)
$\ln(WAGE)$	- 0.006*** (0.093)	- 0.0290*** (0.028)	- 0.007*** (0.043)
R-squared	0.999426	0.999784	0.999187
Adjusted R-squared	0.998393	0.999394	0.997723

注：①括号中数字为标准差；②“*，**”分别表示在 1% 和 5% 的水平上显著，“***”代表不显著。

由表 2 的回归结果可以看出，工资对我国吸引外商直接投资的影响不显著；工资变量的系数符号为负，这在一定程度上说明了劳动成本越高，吸引的外商直接投资量越少。这与 Leonard K. Cheng (2000) 的结论相一致。作者认为：外商在考虑对我国进行直接投资时，更看重基础设施、我国经济发展水平、外商直接投资的前期存量、我国居民的消费水平等，而把工资水平放在一个相对次要的位置。例如，广东省是我国引进外商直接投资最多的省份，并且从广东省的工资水平来看，在全国也是比较高的。另外一个原因可能是本文只是简单地考虑了我国自己的工资水平，而不是把投资母国或地区的工资水平与投资地区的工资水平相比较的相对工资水平。其他系数符号与我们的预想一致。不考虑工资因素的回归结果如表 3 所示。

表 3 不包括工资因素的回归结果

	FDI	HFDI	AFDI
C	- 8.853* (2.916)	- 6.973。 (2.831)	- 22.734。 (3.961)
ln [FDI (- 1) _i]	0.549* (0.078)	0.597* (0.078)	0.185** (0.171)
ln (GDP)	0.390* (0.139)	0.275** (0.218)	0.619* (0.299)
ln (INFRA)	0.734* (0.228)	0.641* (0.259)	1.678* (0.251)
ln (GDP / FDI)	- 0.183* (0.045)	- 0.371* (0.034)	0.276* (0.110)
ln [POL (- 1)]	0.118** (0.098)	0.072** (0.099)	0.730* (0.199)
R-squared	0.999753	0.999763	0.999172
Adjusted R-squared	0.999424	0.999448	0.998067

注：①括号中数字为标准差；②“*”，“**”分别表示在 1% 和 5% 的水平上显著。

计量回归的结果显示：以华资为代表的出口导向型直接投资，与我国整体外商直接投资具有相同的区位因素特征。这说明我国引进的外商直接投资，以华资为代表的出口导向型投资为主。截止到 2001 年底，我国实际利用的外商直接投资中，70.3% 属于出口导向型投资，26.39% 属于市场寻找型和自然资源寻找型（宋泓，2002）。以美资为代表的市场导向型直接投资，除都把我国的基础设施建设放在首要因素外，与二者有很大差别。本文重点比较华资与美资的不同，以说明出口导向型投资和市场导向型投资在这些方面的不同点；对我国的整体 FDI 水平只作简单分析说明。具体分析如下：

(1) 聚集效应分析：前期资本存量通过“试范效应”和“推动效应”引起在华直接投资的增加；并且投资也会引起上下游和相关行业的再投资，从而也会带来投资的增加。前期资本存量对当期吸引 FDI 的影响，体现了聚集效应的作用。由表 3 可以看出，聚集经济效应对华资有重要影响，其系数为 0.597，排在五项因素的第二位；这意味着前期华人直接投资每增加 1%，会带来 0.597% 当期直接投资的增加。对美资的影响相对弱一些，排在最后一位，系数值为 0.185。华资与美资聚集效应大小的不同，可以通过这两种投资在我国投资地区不同来说明。从投资地区来看，华资主要集中在广东、福建、上海、浙江，具有明显的地区聚集性；而美资的这一特征却并不明显。聚集效应是可以创造的，一经建立，它可能通过发挥其外部规模经济而不断加强体现出一种“路径依赖”（Path Dependency），进而影响后续的 FDI 的聚集（吴丰，2002）。我们可以利用聚集效应为我国吸引更多的外商直接投资。

(2) 基础设施建设：实证结果表明，以我国交通设施情况为代表的基础设施建设对我

国吸引外资有重要影响，对三种投资来说，都排在第一位。美资的基础设施系数高达 1.687，我国平均每平方公里营运历程增加 1%，就会相应带来 1.687% 的美国直接投资的增加。运输成本是企业生产的一项重要成本，落后的交通设施建设可能导致企业产品成本的增加，失去市场竞争力。Duggal, Saltzman 和 Klein (1999) 把基础设施作为技术的一个参数，并入生产函数，发现增加基础设施支出能提高资本和劳动的全要素生产率。

(3) 我国经济发展情况： $\ln GDP$ 对华资和美资的区位分布有显著的正的影响，在五项影响因素中，都排在第三位，系数分别为 0.275 和 0.619。这说明这些在华投资会考虑我国的经济基础，并且我国的经济基础对美资的影响比对华资的影响更大。

(4) 政策因子：我国政府的政策因子，体现在对外商直接投资逐步放开。例如，减少对外商直接投资的限制性条款，扩大外商直接投资的领域，减少对合资企业中跨国公司的股权限制，简化外商投资企业的审批登记手续的等，同时也给予一些税收、贷款方面的优惠。这些政策因子在吸引包括华资和美资在内的外商直接投资中起到很大作用。实证结果显示，政策因子对美资的影响比对华资的影响大，这与通常的观点有些不同。但是，在华资与美资的投资地域方面，我们可以看出美资多投资在经济特区的内部，而华资多投资在特区的边缘地带。

(5) 市场需求因素：市场需求因素对华资直接投资的影响比较复杂。实证结果表明，需求因素对华资的影响为负值，而对美资的影响为正值。从美资的回归结果我们知道，以美资为代表的市场导向型直接投资，把我国的市场容量作为重要的考虑因素。中国国内需求的扩张是由持续的经济增长引起的，中国巨大的市场是中国同其他发展中国家竞争国际资本的最有利因素。中国庞大的国内市场，为市场导向型直接投资提供了难得的销售场所。

不同类型的 FDI 投资，都对我国的经济有着不同的影响。从事加工贸易活动的出口导向型投资，对增强我国经济与世界经济的联系，解决我国大量农村转移的低素质劳动力的就业，对我国中小型企业规范经营和管理等有重要影响；但这种投资对我国产业结构的升级效益不高。市场导向型投资对我国技术进步、产业升级具有积极的影响。基础设施、GDP 水平、市场容量，FDI 显著的聚集效应对引进不同 FDI 类型存量水平的重要影响，都促使我们去重新考虑我们对 FDI 的政策：我国自身经济的发展，比给予外商投资优惠政策更有效。

参考文献

- [1] Keith Head, John Ries, Deborah Swenson, Agglomeration benefits and location choice: Evidence from Japanese manufacturing investments in the United States, *Journal of International Economics*, 1994, Vol.38, pp223 ~ 247.
- [2] Steven Goberman, Daniel M. Sapiro, *Journal of International Business Studies*, 1999, Vol.30, pp514 ~ 532.
- [3] David W. Loree, Stephen E. Guisinger, Policy and non-policy determinants of U.S. equity foreign direct investment, 1994.
- [4] Lionard K. Cheng, Yum K. Kwan, What are the determinants of the location of foreign direct investment? The Chinese experience, *Journal of International Economics*, 2000, Vol.51, pp379 ~ 400.
- [5] Yuan Lu, Shatin, Foreign Direct Investment policy in a transitional economy: The case of China, paper

presented at the Academy of International Business Annual Conference, 1998.

- [6] Chyau Tuan, Linda F. Y. Ng, FDI facilitated by agglomeration Economies: Evidence from manufacturing and services joint ventures in China, *Journal of Asian Economics*, 2003, Vol.13, pp749 ~ 765.
- [7] Dermot Leahy, Stephen Pavelin, Follow-my-leader FDI and tacit collusion, *International Journal of Industrial Organization*, 2003, Vol.21, pp439 ~ 453.
- [8] C. Keith Head, John C. Ries, Deborah L. Swenson, Attracting foreign manufacturing: Investment promotion and agglomeration, *Regional Science and Urban Economics*, 1999, Vol.29, pp197 ~ 218.
- [9] Linda F. Y. Ng, Chyau Tuan, Location decisions of manufacturing FDI in China: Implications of China's WTO accession, *Journal of Asian Economics*, 2003, Vol.14, pp51 ~ 72.
- [10] 孙俊:《中国 FDI 地点选择的因素分析》,《经济学季刊》,2002 年第 3 期。
- [11] Sylvie Démurger 等:《地理位置与优惠政策对中国地区经济发展的相关贡献》,《经济研究》,2002 年第 9 期。
- [12] 课题组:《看好潜在市场规模——欧盟在华投资中小企业业绩及经营战略调查》,《国家贸易》,2002 年第 4 期。
- [13] 王岳平:《我国外商直接投资的两种市场导向类型分析》,《国际贸易问题》,1999 年第 2 期。
- [14] 徐康宁,王剑:《美国对华直接投资决定性因素分析(1983~2000)》,《中国社会科学》,2000 年第 5 期。
- [15] 王洛林:《中国外商投资报告——外商投资的行业分布》,经济管理出版社 1997 年版。
- [16] 梁志成:《中国 FDI 的最优规模与内生增长研究》,《世界经济》,2000 年第 7 期。
- [17] 魏后凯,贺灿飞,王新:《外商在华直接投资动机与区位因素分析》,《经济研究》,2001 年第 2 期。
- [18] 魏后凯:《欧美日韩在华制造业投资的区位决定》,《中国工业经济》,2000 年第 11 期。
- [19] 贺灿飞,魏后凯:《信息成本、集聚经济与中国外商投资区位》,《中国工业经济》,2001 年第 9 期。

四组外商投资企业技术溢出效应的比较研究*

许罗丹 谭卫红 刘民权

一、引言

在经济全球化的背景下，任何国家都不能仅仅依靠本国的力量进行技术创新，还需要充分利用外部资源，促进本国的技术进步和经济发展。跨国公司的对外直接投资，是东道国吸收先进技术的重要途径；跨国公司在对外直接投资中，通过技术转让和技术溢出，对东道国企业的技术进步产生积极的影响。技术转让通过内部渠道、市场渠道向东道国企业进行技术转移，是一种交易行为；技术溢出是一种外部效应，是跨国公司对当地企业的一种非自愿行为。本文利用对广东省美资、日资、欧盟、华资四组企业的调查数据，重点分析外商直接投资技术溢出效应。

Kokko (1992) 在《外国直接投资、东道国特征和溢出》一书中，考察了 Multinational Corporation (以下简称 MNC) 在他国设立子公司，引致技术和生产力在当地溢出，对当地企业造成影响的情况。在该书中，Kokko 定义“溢出”为：由于 MNC 在东道国设立子公司，从而引起当地技术或生产力的进步，但 MNC 子公司又无法获取全部收益的现象。本文所指的“溢出效应”，一方面是指由于 MNC 子公司的进入和存在，引起和本地企业之间业务上的联系，从而不自觉地引起了前后产业关联效应；另一方面是由于 MNC 子公司的进入，引起本国公司面临的竞争加剧，迫使当地公司进行技术模仿、改进，以此实现生产效率的提高。

二、文献回顾

对溢出效应的研究可以追溯到 20 世纪 60 年代。MacDougall (1960) 在研究 FDI 福利效应时，首次把技术外溢效应作为 FDI 的一个重要现象进行分析。Caves (1974) 根据技术扩散对当地厂商的不同影响，第一次比较全面地把技术扩散可能存在的外在性分为三类：①本来具有强大行业壁垒的产业，由于 MNC 子公司的强行进入，垄断扭曲受到遏制，资源配置得到改善；②由于 MNC 子公司的进入，给当地企业带来的不断增加的竞争压力或示范效应，刺激当地厂商更加有效地利用资源，推动当地技术效率的提高；③由于竞争、反复模仿或其他原因，MNC 子公司的进入将加快技术扩散的速度。Findlay (1974) 构建简

* 原载：《管理世界》，2004 年第 6 期。

单的内生动态化模型,检验诸如技术差距、外资份额等静态特征对技术扩散的影响。模型假定:FDI可以通过传染效应(Contagion Effect)提高东道国的技术水平;FDI输出国与东道国的技术差距越大,技术扩散率越高;MNC子公司在当地的资本份额越高,扩散的速度越快。20世纪80年代,Romer和Lucas等人提出了新增长理论,突破了新古典增长理论关于技术外溢性的假定。Romer(1990)构建内生增长模型,着重强调技术扩散对于小国及发展中国家经济持续增长的决定性作用。FDI作为技术扩散的一个重要渠道,得到越来越多的重视。Kokko(1992)认为技术溢出效应的发生来自两个方面:一是来源于示范、模仿和传播;二是来源于竞争。前者是技术信息差异的增函数;后者主要决定于外国公司与当地厂商的市场特征及其相互影响。Markusea和Venables(1997)发现FDI与国内的投资具有互补性。

我国对外商直接投资技术溢出效应的研究比较晚。李向阳(1990)指出由外商直接投资引起的技术不仅会作用于本产业,还会向其他产业进行传播。王志乐(1996)全面介绍了各国著名跨国公司在中国的投资情况,包括详细的案例研究,还分析了外商在中国的直接投资对经济的正反两方面的影响。姚洋(1998)利用第三次全国工业普查资料,从中随机抽取了12个大类行业中的146704家企业作为样本,进行多因素回归分析后得出:与国有企业相比,国外“三资”企业的技术效率要高39%,港澳台“三资”企业要高33%;并且在行业内,如果国外“三资”企业数量的比重每增加一个百分点,东道国行业内每个企业的技术效率就会提高1.1个百分点。何洁、许罗丹(1999)借鉴Feder, Coe, Helpman(1995)的做法,利用有关外国直接投资工业企业的统计数据,通过计量分析外商直接投资企业对我国内资工业企业以及整个工业部门产生的外溢效应,得出外国直接投资带来的技术水平每提高1个百分点,我国内资工业企业的技术外溢作用(产量的增加)就提高2.3个百分点。沈坤荣、耿强(2001)构建内生增长模型,依据我国近几年的具体数据,运用计量方法,得出FDI通过技术外溢与经济增长呈正相关关系。

本文在借鉴前人研究的基础上,总结外商直接投资溢出效应的途径,利用对广东省外商投资企业的调查数据,试图从华资、欧盟、日资、美资四个投资方,对比分析这些外商直接投资在产品水平、R&D投入、上游产品来源、员工素质、培训方面对我国企业的技术溢出效应。

三、外商直接投资技术溢出的途径

根据以上对外商直接投资技术溢出效应研究的回顾,本文把外商直接投资技术溢出的途径归结为三种:示范模仿和市场竞争、厂商间的前项和后项联系、人力资本的流动。

示范模仿和市场竞争:对示范模仿效应的研究,由于理论上认为只有存在技术差距才可能出现示范和模仿,所以对此的研究主要集中在比较MNCs子公司和当地厂商间的技术水平上。而技术差距又具体体现在要素禀赋和劳动生产力两个经济指标上。厂商获取技术知识的另一种途径是“逆向工程”,即通过对产品的研究获取生产该产品的技术,是一种重要的模仿手段。

从竞争的角度考虑,发达国家MNCs凭借其先进的技术和管理经验等优势,进入发展

中国家后，打破了原来的市场均衡，加剧了当地的市场竞争。当地企业为了生存和发展，适应加剧的市场竞争，改进生产技术，提高技术水平。

厂商间前后项联系：联系也称关联（Linkages），是指外国跨国公司子公司与当地企业之间通过市场关系长期形成的一种供求方面的契约；广义的关联还应包括跨国公司子公司与当地政府、高校、金融服务机构、研发机构等的联系。关联有三种表现形式：前项（Forward Linkage）、后项（Backward Linkage）、水平（Horizontal Linkage）。前项联系指跨国公司子公司向当地企业销售产品或服务；后项联系则是跨国公司子公司向当地企业购买原材料和服务；水平联系强调的是跨国公司子公司与当地企业相互竞争，在竞争过程中形成的千丝万缕的联系。

与当地企业相比，MNCs 子公司通常具有先进的技术，当其当地的供应商和客户发生联系时，当地厂商就有可能从 MNCs 子公司学到先进的生产技术、管理经验以及先进的工序技术等。通常情况下，这种学习效应是一种“免费搭车”现象，即使 MNCs 向其收取一定的费用，这种费用往往也不能弥补其所费的成本，也不及当地厂商获得的全部利益。

对后项关联的溢出效应研究，主要包括：帮助供应商建立生产性设施；改善供应商产品的质量或促进其创新活动，为其提供技术帮助或信息服务；提供或帮助购买原材料和中间产品；提供组织管理上的培训和帮助；通过发掘新客户，帮助供应商从事多样化经营。当地价值含量是形成后项关联的重要因素。

相对于后项关联，有关前项关联的研究比较少。前项关联侧重于当地市场的培育，有助于尽快形成当地的生产体系，开发其制成品市场。

人力资本的流动：由于发展中东道国的人力资本，相对于 MNCs 母国来说，水平较低。发达国家的 MNCs 需要对东道国的当地雇员提供培训机会和培训设施。MNCs 对当地雇员培训形成东道国技术外溢的基础。这些培训从最简单的生产性操作，到监管人员，到高级技术人员，乃至高级经理人员，各个层次都有。这些雇员由 MNCs 子公司流向当地其他公司或自创企业时，其在 MNCs 工作时所学的专业技术和经营管理技术也随之外流。Caves（1982）发现，由日本向美国、欧洲的企业经理人员的流动，加速了管理技能的传播。

四、外商直接投资企业的生产技术及对我国的技术溢出

外商直接投资企业通过示范模仿和市场竞争、厂商间的前后项联系、人力资本的流动对我国的技术进步作出贡献，使得我国企业“免费搭车”，获得技术外溢效应。但华资、日资、欧盟、美资四组企业在途径上各有不同。

虽然资金来源不同的外资企业对生产技术的侧重有所不同，但它们都毫无例外地表现为产品的技术水平、企业的研发情况以及对雇员的培训。我们从这三个方面来剖析华资、日资、欧盟、美资四组外资企业的技术实力及其最终造成的我国市场竞争情况。

(一) 产品差异与产品水平

1. 产品差异

首先,我们来认识各组外资企业对产品差异的看法,看看它们认为自己企业提供的产品与国内同类产品之间的差异是由什么因素决定的。如表 1 所示。

表 1 产品差异

项 目	总计		华资		欧盟		日资		美资	
	企业数量	所占比例								
质量	160	38.55	101	35.94	14	31.11	21	58.33	24	45.28
质量、外观	75	18.07	48	17.08	10	22.22	3	8.33	14	26.42
质量、外观、其他	13	3.13	6	2.14	2	4.44	3	8.33	2	3.77
质量、其他	19	4.58	7	2.49	4	8.89	6	16.67	2	3.77
外观	33	7.95	28	9.96	2	4.44	2	5.56	1	1.89
外观、其他	1	0.24	1	0.36	0	0.00	0	0.00	0	0.00
其他	36	8.67	30	10.68	3	6.67	1	2.78	2	3.77
没有差异或不提供资料	78	18.80	60	21.35	10	22.22	0	0.00	8	15.09
	415	100	281	100	45	100	36	100	53	100

注:企业数量单位:家;所占比例单位:%。

这四组外资企业都认为产品质量在很大比重上决定了产品的差异,不过资金来源不同的外资企业对产品质量、外观及其他给予了不同的权重。根据表 1 可以看到:日资企业和美资企业远比华资企业和欧盟企业更强调质量在决定产品差异上的权威地位;58.33%的日资企业与 45.25%的美资企业认为,只有质量才是同类产品之间真正差异之所在,而持同一观点的华资企业与欧盟企业只有 30%多;有 9.96%的华资企业认为外观是决定产品差异的关键,这一比例分别是持相同观点的美资企业、日资企业的 5.19 倍、1.79 倍。表 1 的数据显示欧盟企业在关于产品差异问题上的观点与华资企业较为接近。另外,所有的日资企业均认为它们生产的产品与国内同类产品之间存在着差异,无论这种差异是由什么原因造成的,日资企业都肯定它们的产品至少在某一方面独特的。

至于造成差异的条件及其潜在原因,表 2 显示了各组外资企业所持的观点。

表 2 产品差异原因

项 目	总计		华资		欧盟		日资		美资	
	企业数量	所占比例								
新生产线	66	15.90	54	19.22	4	8.89	5	13.89	3	5.66
新生产线、新技术	53	12.77	28	9.96	6	13.33	8	22.22	11	20.75
新生产线、新技术、其他	8	1.93	4	1.42	2	4.44	1	2.78	1	1.89
新生产线、其他	4	0.96	3	1.07	1	2.22	0	0	0	0
新技术	109	26.27	70	24.91	13	28.89	10	27.78	16	30.19
新技术、其他	6	1.45	4	1.42	0	0	0	0	2	3.77
其他	67	16.14	46	16.37	6	13.33	5	13.89	10	18.87
没有差异或不提供资料	102	24.58	72	25.62	13	28.89	7	19.44	10	18.87
	415	100	281	100	45	100	36	100	53	100

注：企业数量单位：家；所占比例单位：%。

四组外资企业都认为新技术是造成产品差异的本质原因。在这一点上，四组外资企业给予新技术的权重差别不大，其重要性从最低的华资 24.91% 到最高的美资 30.19%。新生产线被看成造成产品差异的另一个重要因素，但对新生产线，四组企业的观点差别较大。仅有三家美资企业认为新生产线是产生产品差异的决定性因素，比例为 5.66%；而持同一观点的华资企业有 54 家，比例为 19.22%。还有相当部分的外资企业将新技术与新生产线同时列为产品差异形成的具体条件。

对新技术、新生产线的不同重视程度，对四组外资企业的产品造成不同的影响，产品的差异程度有所不同，产品所达到的水平、对我国的影响也不相同。

2. 产品水平

在我们了解了各组外资企业对产品差异及差异原因的观点后，我们来看看它们生产的产品达到的水平，我们通过各组企业生产产品符合国际、国内、其他标准的比例（见表 3）来说明这个问题。这个比例不仅能体现出企业在市场上所具有的竞争力，同时也是了解一个企业生产技术水平与产品技术含量的直观指标。

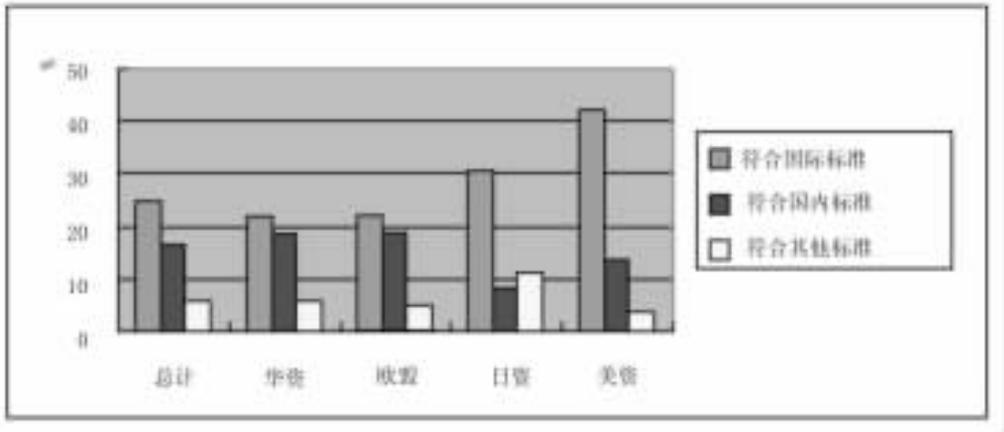


图1 产品水平

注：国家标准：在全国范围内统一的技术要求；国际标准：包括 ISO, IEC, UL, JIS, DIN, BSI, ASME, ASTM, API, IEEE, GOST 等各种标准；行业标准：指没有国家标准而又需要在全国某个行业范围内统一的技术要求

根据图1可以看出，42.31%的美资企业产品水平符合国际标准，是四组企业中最高；其次是日资企业，产品水平符合国际标准的日资企业比例为30.56%，而华资企业、欧盟企业的这一比例仅为20%左右。不过华资企业、欧盟企业产品符合国内标准的比例相对较高，都超过了15%。欧盟企业与华资企业相比，其产品符合各种标准的比例比华资企业更低。

(二) R&D 投入

企业对技术的重视与其技术实力除了在其产品水平上得以体现外，企业对自身研究与发展(R&D)活动的投入也是衡量其技术实力水平的一个重要标志。

表3列出了在不同资金来源的四组企业中，从事R&D活动和设立专门R&D部门的企业比重。

表3 R&D情况

项 目	总计		华资		欧盟		日资		美资	
	企业数量	所占比例								
进行R&D的公司数	182	43.86	118	41.99	27	60.00	11	30.56	24	46.15
有R&D部门的公司数	142	34.22	95	33.81	19	42.22	5	13.89	23	44.23

注：企业数量单位：家；所占比例单位：%。

有60%的欧盟企业进行R&D活动，是比例最高的一组；其他由高往低依次为美资企

业、华资企业、日资企业。至于“企业有专门的 R&D 部门”，美资企业具有最高比例，达到 44.23%。欧盟企业的比例也很高，为 42.22%，而日资企业只有 13.89%。

由表 3 也可以看出，即使企业不设立专门的 R&D 部门，也会有一些相关人员进行 R&D 活动。

考察四组外资企业对 R&D 的资金投入，我们来比较四组企业进行 R&D 活动的资金总量与平均量（见表 4）。

表 4 对 R&D 的投入

项 目	总计	华资	欧盟	日资	美资
每组企业的总值	573156118	338956118	168875000	23675000	41650000
每家企业的均值	1381099.08	1206249.53	3752777.78	657638.89	785849.06

注：投入单位：元/年。

可见，欧盟企业对 R&D 活动的投入远远高于另外三组外资企业，平均每家欧盟企业每年用于 R&D 的经费接近 380 万元。平均每家华资企业对 R&D 的投入也不少，达到 130 万元/年。日资企业对 R&D 的资金投入是最少的，几乎只有欧盟企业平均水平的 1/6。

更进一步分析四组外资企业 R&D 经费支出占销售额的比重，我们得到表 5。大部分日资企业用于 R&D 的经费仅占销售额的 10%，而一定数量的美资企业、欧盟企业的 R&D 经费占销售额比重的 50% 以上，甚至有部分美资企业用于 R&D 的经费超过了企业的销售额。华资企业用于 R&D 的经费占销售额的比重也相当大。与美资企业相比，甚至有更多比例的华资企业 R&D 经费超过了其销售额。

表 5 R&D 占销售额比重

项 目	总计		华资		欧盟		日资		美资	
	企业数量	所占比例	企业数量	所占比例	企业数量	所占比例	企业数量	所占比例	企业数量	所占比例
10%	29	6.99	21	7.47	3	6.67	4	11.11	1	1.92
20%	9	2.17	7	2.49	0	0.00	1	2.78	1	1.92
30%	9	2.17	7	2.49	2	4.44	0	0.00	0	0.00
40%	1	0.24	1	0.36	0	0.00	0	0.00	0	0.00
50%	9	2.17	7	2.49	1	2.22	0	0.00	1	1.92
60%	3	0.72	2	0.71	1	2.22	0	0.00	0	0.00
70%	1	0.24	1	0.36	0	0.00	0	0.00	0	0.00
80%	6	1.45	3	1.07	2	4.44	0	0.00	1	1.92
90%	3	0.72	3	1.07	0	0.00	0	0.00	0	0.00
100%	8	1.93	4	1.42	2	4.44	1	2.78	1	1.92
100% 以上	9	2.17	8	2.85	0	0.00	0	0.00	1	1.92

注：企业数量单位：家；所占比例单位：%。

另外，四组外资企业进行 R&D 活动基本上没有外部资助（包括来自各级政府或其他机构^①的科研经费资助）。调查发现，只有 7 家华资企业和一家美资企业的 R&D 活动获得了外部资金注入，但是所获得的外部资助在全部 R&D 经费中所占比例也不大。除了一家华资企业获得占其 R&D 总开支 50% 的外部资助外，其余企业得到的外部资助不到其科研经费的 10%。

（三）员工素质

人是接受技术的重要条件，员工素质是影响企业技术水平的重要因素。即使有先进的技术、生产线、机器设备，如果没有人会操作也是枉然。我国一些企业在引进大型机器设备方面，因为没有相应的技术人才，而造成机器闲置、资源浪费的现象，给我们以启示。

在过去人们对企业所做的分析中，员工素质（包括员工的受教育程度与企业对员工所提供的在职培训）往往被作为企业的管理因素而不是技术因素来进行分析。本文把员工素质作为影响企业产品与技术水平的重要因素来考虑。

1. 员工的受教育水平

（1）生产线工人。

图 2 展示了在四组企业中受过大学教育、高中教育、初中教育和小学教育的工人占总体工人的平均比重情况。接受大学及以上教育的工人，欧盟最多，其次依次为日资、美资、华资；可以看到，在四组企业中，接受高中教育的工人比重最高，其次是受过初中教育的工人。在四组企业中，华资企业生产线工人的受教育程度是最低的。华资企业具有高中教育水平工人的比重并不比初中文化程度工人的比重高多少，而且它们生产线工人只受过小学教育的比例也是四组企业中最高的。日资企业中，具有高中教育水平的工人比例比具有初中文化程度的工人比例高 41.72%，日资企业具有高中程度的工人比重（64.06%）

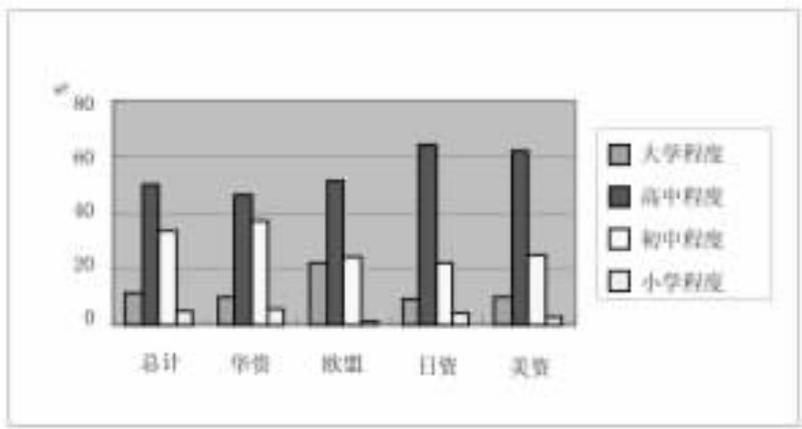


图 2 各组生产线工人平均受教育程度

^① 指上下游企业的资助。

在四组企业中是最高的。而欧盟企业拥有受过大学教育程度工人的比重为 22%，远比其他三组企业相应的 10% 的比重要高。在所受教育程度水平上，美资企业和日资企业工人的比重组成情况很相似。

(2) 管理人员。

生产线工人作为企业产品的直接制造者，他们的受教育程度与平均年龄固然在某种程度上与企业的科技水平相关联，管理人员作为生产线工人工作的组织者与管理者，他们的素质同样在一定意义上反映出企业的技术实力。表 6 提供了四组外资企业管理者在这方面的情况。

表 6 管理人员受大学及以上教育程度

项 目	总计		华资		欧盟		日资		美资	
	企业数量	所占比例								
10%	75	18.07	69	24.56	2	4.44	3	8.33	1	1.92
20%	40	9.64	32	11.39	2	4.44	5	13.89	1	1.92
30%	27	6.51	21	7.47	3	6.67	2	5.56	1	1.92
40%	24	39.04	20	7.12	2	4.44	2	5.56		0.00
50%	24	5.78	18	6.41	2	4.44	2	5.56	2	3.85
60%	13	70.36	9	3.20	3	6.67	1	2.78		0.00
70%	27	6.51	21	7.47	1	2.22	1	2.78	4	7.69
80%	33	7.95	19	6.76	3	6.67	5	13.89	6	11.54
90%	43	10.36	21	7.47	7	15.56	5	13.89	10	19.23
100%	73	17.59	30	10.68	14	31.11	7	19.44	22	42.31

注：企业数量单位：家；所占比例单位：%。

表 6 中第一列的百分数表示该企业中有若干比例的管理人员受过大学及以上教育。由此可以清楚看到，华资企业管理人员的受教育程度明显低于另外三组外资企业，这一点与其生产线工人的情况相同。华资企业中管理人员大学及以上教育程度在 50% 以下的占了多数，其中管理人员受大学及以上教育程度的比例为 10% 的华资企业就占 24.56%，而欧盟企业、日资企业和美资企业的管理人员大学及以上教育程度比例多集中在 70% 以上，并且管理人员 100% 接受大学及以上教育的企业比重均是所有比例中最高的，美资企业的这一比例更是高达 42.31%。由此可见，非华资企业比华资企业更重视管理人员的受教育程度，其管理人员的平均学历相应也较高。

2. 在职培训

为了让员工更好地适应其工作岗位，提高其工作能力，企业会对员工进行在职培训，员工接受的在职培训同样成为企业生产技术水平的一个考查指标。与国内企业相比，外资企业更重视提高员工的职业素质与技能。而外资企业对员工的培训，无形中提高了我国技术管理人员的素质，促进了我国技术与管理水平进步。因此在调查中，我们对外资企业向员工提供的在职培训进行了细致的比较与分析。结果发现，企业对各类人员的培训均以一个月以内的短期培训为主；日资企业对其员工所提供的培训是四组企业中最少的。

表7给出了各组企业中，为员工提供在职培训的企业的比重分布。高达95.56%的欧盟企业都为员工提供在职培训，其他依次为日资企业、美资企业、华资企业。而且超过半数的欧盟企业、美资企业内设有专门的培训部门。毫无疑问，与华资企业、日资企业相比，美资企业和欧盟企业对在职培训表现出更大程度的重视。

表7 提供在职培训

项 目	总计		华资		欧盟		日资		美资	
	企业数量	所占比例								
提供在职培训	343	82.65	219	77.94	43	95.56	32	88.89	46	88.46
有专门的培训部门	165	39.76	100	35.59	23	51.11	15	41.67	26	50.00
由外单位提供培训	99	23.86	62	22.06	15	33.33	4	11.11	18	34.62

注：企业数量单位：家；所占比例单位：%。

在当今大学生的择业选择中，因为对未来企业提供培训机会的重视，把美资、欧资等一些外资企业作为就业选择的首选对象，而把日资、华资企业放在相对次要的位置。这也在一定程度上显示了美资、欧资企业对人才、人员培训的重视，这些企业对提升我国的人员素质作出了一定的贡献。

下面考虑四组企业在职培训的时间长度。

前文已经提到，与产品制造关系最为密切的生产线工人与技术人员，还有作为生产的组织者与管理者的管理人员，他们的综合程度与企业的科技水平相关联；同样，其他部门的员工作为构成企业发展整合群体的部分，他们的素质在一定程度上同样反映了企业的技术实力。因此我们收集了四组企业中不同部门的人员一年内接受培训时期的具体数据，并加以整理，按企业员工每年接受的培训种类加以区分，得到表8。它概括了四组企业中技术人员、销售人员、行政人员、生产线工人和其他人员一年内接受培训时间长度的情况。

表 8 各类人员一年受训时间

项 目		总计		华资		欧盟		日资		美资	
		企业 数量	所占 比例								
技术人员	一周至一个月	198	47.71	136	48.40	21	46.67	15	41.67	26	50.00
	一个月至两个月	18	4.34	13	4.63	1	2.22	1	2.78	3	5.77
	两个月以上	12	2.89	9	3.20	2	4.44		0.00	1	1.92
销售人员	一周至一个月	170	40.96	110	39.15	21	46.67	13	36.11	26	50.00
	一个月至两个月	17	4.10	10	3.56	2	4.44	2	5.56	3	5.77
	两个月以上	3	0.72	2	0.71	1	2.22		0.00		0.00
行政人员	一周至一个月	198	47.71	132	46.98	23	51.11	14	38.89	29	38.46
	一个月至两个月	10	2.41	6	2.14	2	4.44	0	0.00	2	3.85
	两个月以上	3	0.72	3	1.07		0.00		0.00		0.00
生产线工人	一周至一个月	211	50.84	141	50.18	25	55.56	14	38.89	31	42.31
	一个月至两个月	19	4.58	13	4.63	2	4.44	0	0.00	4	7.69
	两个月以上	7	1.69	4	1.42	1	2.22	1	2.78	1	1.92
其他人员	一周至一个月	72	17.35	54	19.22	8	17.78	5	13.89	5	9.62
	一个月至两个月	4	0.96	3	1.07	0	0.00	0	0.00	1	1.92
	两个月以上	1	0.24	1	0.36	0	0.00	0	0.00	0	0.00

注：企业数量单位：家；所占比例单位：%。

很明显，在四组企业中，各类人员的培训均以一周至一个月的短期培训为主。就提供在职培训的量而言，欧盟企业与华资企业总体上是向各类人员提供岗位训练最多的两组企业。除管理人员外，日资企业与其他各组企业相比，为其员工提供的在职培训最少；不过对生产线工人提供两个月以上在职培训，比重最高的也是日资企业。除生产线工人和技术人员外，没有一家美资企业对其他部门的员工提供两个月以上的在职培训，同时，美资企业是四组企业中最注重对技术人员、销售人员提供一周至一个月、一个月至两个月培训的，但它们对行政人员提供的培训却比另外三组企业要少，这或许可以通过我们对表 8 的分析来获得答案。由表 8 可知，美资企业行政人员是四组企业中受教育程度、学历水平最高的。

概括各组企业员工的受教育程度与企业对其员工所提供的在职培训，不难发现：华资企业的员工无论是生产线工人还是管理人员，所受教育程度是四组企业中最底的，因此华资企业向员工提供的在职培训是最多的；欧盟企业的员工文化水平虽然不低，但它们同样提供了很多的岗位培训；日资企业的生产线工人具有最高和最平均的受教育程度，它们向各类人员提供的在职培训是最少的；美资企业的管理人员在四组企业中学历水平最高，比较注重对技术、销售人员的培训，而对管理人员的培训相对重视不够。

对员工进行训练的量，一方面依赖于企业对在职培训的重视程度，另一方面也受到企业所投入资金的限制。将企业员工的素质与企业对 R&D 活动的投入、企业产品的水平联系起来，我们假设企业对 R&D 的总投入包括了企业对员工提供各类在职培训的经费开支，那么前面所提到的企业资金投入与其产品水平之间存在的差距就可以获得解释。华资企业与欧盟企业对研究发展活动的重视与热忱还包括了对员工的培训，所以它们对 R&D 的投入虽然很大，但产品水平并不突出。

另外，我们必须说的是，关于企业的在职培训，外资企业的员工尤其是生产线工人大多是从农村流动到城市的打工者，外资企业对他们进行培训固然是出于自身生产目的的需要，实际上也完成了对这些打工者从农民到工人的提升，提高了他们的劳动技巧与能力，形成了人才方面对我国的技术外溢。

外资企业对员工提供在职培训，受培训的员工通过在各个企业间的流动，又形成了不同企业之间的技术溢出。下面我们从企业员工服务期的角度考虑技术外溢问题。

3. 企业员工的服务期

我们对各类人员的平均工作年数进行分析，以下两点值得注意：一是基本上各组企业的各类人员一般工作年数比重最高的是 3 年，其次是 5 年。二是美资企业的各类人员一般工作年数多为一年，普遍低于其他三组企业，欧盟企业各类员工平均工作年数明显偏长。

下面我们分别讨论技术人员、销售人员、行政人员、生产线工人的工作年数。

(1) 技术人员一般工作年数。

表 9 提供了技术人员在四组外资企业中一般工作年数的分布。对于各组外资企业来说，技术人员的一般工作年限多为 3 年，但是美资企业技术人员一般工作年数比重最大的是两年，只工作 1 年的比例也很大。另外，三组企业中有相当比重的技术人员工作 5 年，但美资企业的这一比例很低，几乎只是欧盟企业这一比例的 1/4。可见，美资企业技术人员的流动率要高于其他三组企业。

表 9 技术人员一般工作年数

项 目	总计		华资		欧盟		日资		美资	
	企业数量	所占比例								
1 年	25	6.02	17	6.05	1	2.22	2	5.56	5	9.62
2 年	46	11.08	28	9.96	4	8.89	5	13.89	9	17.31
3 年	63	15.18	44	15.66	6	13.33	7	19.44	6	11.54
4 年	20	4.82	17	6.05	1	2.22	1	2.78	1	1.92
5 年	35	8.43	25	8.90	6	13.33	2	5.56	2	3.85
6 年	4	0.96	1	0.36	1	2.22	1	2.78	1	1.92
7 年	4	0.96	2	0.71	1	2.22		0.00	1	1.92
8 年	1	0.24	1	0.36		0.00		0.00		0.00
9 年	1	0.24	0	0.00	1	2.22		0.00		0.00
10 年	4	0.96	4	1.42		0.00		0.00		0.00
10 年以上	2	0.48	0	0.00		0.00	1	2.78	1	1.92

注：企业数量单位：家；所占比例单位：%。

(2) 销售人员一般工作年数。

从表 10 中可以看出，美资企业销售人员的流动性相当高。美资企业销售人员只工作 1 年的比例要大大高于其他三组企业。就华资企业与日资企业自身而言，其销售人员一般工作 3 年的比重最大。欧盟企业销售人员工作 5 年的比例与工作 1 年的比例一样高，且比其他三组企业的相应比例都高。

表 10 销售人员一般工作年数

项 目	总计		华资		欧盟		日资		美资	
	企业数量	所占比例								
1 年	40	9.64	22	7.83	3	6.67	4	11.11	11	21.15
2 年	34	8.19	24	8.54	6	13.33	1	2.78	3	5.77
3 年	48	11.57	33	11.74	6	13.33	5	13.89	4	7.69
4 年	13	3.13	10	3.56	2	4.44	1	2.78		0.00
5 年	21	5.06	13	4.63	3	6.67	2	5.56	3	5.77
6 年	1	0.24	0	0.00		0.00	1	2.78		0.00
7 年	3	0.72	2	0.71		0.00		0.00	1	1.92
7 年以上	2	0.48	2	0.71		0.00		0.00		0.00

注：企业数量单位：家；所占比例单位：%。

(3) 行政人员一般工作年数。

对于华资企业、欧盟企业与日资企业自身来说，其行政人员工作 3 年与 5 年的比重比其他数不超过 3 年是主要趋势，尤以只工作两年为多，但也有一定比例的行政人员工作 7~8 年，并且这一比例超过了其他三组企业（见表 11）。

表 11 行政人员一般工作年数

项 目	总计		华资		欧盟		日资		美资	
	企业数量	所占比例								
1 年	25	6.02	14	4.98	2	4.44	2	5.56	7	13.46
2 年	35	8.43	21	7.47	2	4.44	4	11.11	8	15.38
3 年	61	14.70	41	14.59	8	17.78	5	13.89	7	13.46
4 年	10	2.41	8	2.85	1	2.22		0.00	1	1.92
5 年	34	8.19	24	8.54	6	13.33	3	8.33	1	1.92
6 年	7	1.69	3	1.07	2	4.44	2	5.56		0.00
7 年	4	0.96	3	1.07		0.00		0.00	1	1.92
8 年	3	0.72	2	0.71		0.00		0.00	1	1.92
8 年以上	7	1.69	5	1.78	1	2.22	1	2.78		0.00

注：企业数量单位：家；所占比例单位：%。

(4) 生产线工人一般工作年数。

就华资企业自身而言，生产线工人工作 3 年的比例在所有工作年数中比重最大，其次是工作 1 年。美资企业生产线人员工作 1 年的比例为 11.54%，远大于其他工作年数的比重。日资企业的受训生产线人员较为平均地工作 1~3 年。而欧盟企业的受训生产线人员大多工作 5 年，这一比例（11.11%）远远高于其他三组企业的相应比例（见表 12）。

表 12 生产线工人一般工作年数

项 目	总计		华资		欧盟		日资		美资	
	企业数量	所占比例								
1 年	45	10.84	32	11.39	2	4.44	5	13.89	6	11.54
2 年	30	7.23	20	7.12	4	8.89	5	13.89	1	1.92
3 年	54	13.01	44	15.66	3	6.67	4	11.11	3	5.77
4 年	8	1.93	6	2.14	1	2.22		0.00	1	1.92
5 年	22	5.30	13	4.63	5	11.11	3	8.33	1	1.92
5 年以上	4	0.96	1	0.36	2	4.44		0.00	1	1.92

注：企业数量单位：家；所占比例单位：%。

将各组企业员工的平均工作时期（即流动性）与企业对员工的管理措施相对比，我们得出一个令人疑惑的结论：欧盟企业与华资企业对员工的管理最为严苛，但是它们的员工流动性较低，平均工作年数较长；相对而言，美资企业对员工的管理是四组企业中最宽松的，但在美资企业工作的各类人员流动性最高，平均服务期最短。这一现象背后的原因是多方面的。可能因为欧盟企业的劳工合同规定较长的工作年限，也可能欧盟企业与华资企业多样化的奖励方式使员工愿意与企业共发展，还可能是欧盟企业与华资企业的工作环境和工作氛围最吸引人……。结合文中企业员工素质的调查，关于企业提供在职培训的分析结果表明，欧盟企业与华资企业每年向其各类员工提供最多的在职培训，使员工获得最多的可能来充实与发展自己，是否恰恰因此而使得这两组企业的员工能够接受或者说忍受企业内严格的管理以获取自身更大的提升？

无论如何，外资企业员工的流动性比起国有企业要高得多。当这些接受了外资企业在职培训的员工离开原来服务的公司后，他们不管进入哪一类型的企业，都有可能获得进一步自我发展的机会，如果是进入国有企业，他们在外企所受的培训与工作经历能够帮助他们更好地适应新岗位，同时推动国有企业的发展。而那些接受他们原来职位的人亦能够获得新一轮劳动技能的培训与提升。外资企业在无意间为社会培训与提升了一批又一批的劳动者，帮助社会完成农民到工人、工人到行政人员、行政人员到高级管理者的劳动力的升级与进化。这是外资企业在其自身发展过程中对中国经济与社会的又一贡献。

(四) 上游产品

了解了广东省外资企业生产技术与管理方面的情况，我们来关注外资企业其他的企业

行为。从外资企业上游产品的角度考虑对我国的技术外溢。

表 13 提供了四组企业上游产品的一些情况。其中“无主要上游产品”与“无上游产品”是有区别的。前者主要是指从事加工制造业的外资企业对各种原料进行加工，并不是针对某种或某几种特定的原材料进行生产。后者是指企业生产所需原料并非其他企业生产出来的产品。

表 13 上游产品

项 目	总计		华资		欧盟		日资		美资	
	企业数量	所占比例								
有主要上游产品	197	47.47	139	49.47	20	44.44	18	50.00	20	37.74
无主要上游产品	115	27.21	72	25.62	15	33.33	11	30.56	17	32.08
不提供资料	89	21.44	64	22.78	6	13.33	6	16.67	13	24.53
无上游产品	12	2.89	6	2.14	3	6.67	1	2.78	2	3.77

注：企业数量单位：家；所占比例单位：%。

广东省的外资企业大多从事加工制造业，无论是对特定原料进行加工还是不对特定原料进行加工，没有上游产品的企业只是少数，有主要上游产品的企业占了大多数，没有主要上游产品的企业比重也不低。

上游产品的来源

表 14 至表 17 分别说明四组外资企业上游产品的具体来源。

表 14 华资企业上游产品来源

项 目	来自本地		来自中国 其他地方		来自国外		来自 “三资”企业	
	企业数量	所占比例	企业数量	所占比例	企业数量	所占比例	企业数量	所占比例
10% 以下	94	33.45	102	36.30	64	22.78	41	14.59
20%	28	9.96	17	6.05	9	3.20	20	7.12
30%	10	3.56	10	3.56	11	3.91	8	2.85
40%	3	1.07	11	3.91	5	1.78	3	1.07
50%	7	2.49	7	2.49	13	4.63	13	4.63
60%	8	2.85	7	2.49	3	1.07	5	1.78
70%	4	1.42	2	0.71	13	4.63	4	1.42
80%	8	2.85	9	3.20	15	5.34	10	3.56
90%	9	3.20	9	3.20	9	3.20	3	1.07
100%	13	4.63	13	4.63	51	18.15	7	2.49

注：企业数量单位：家；所占比例单位：%。

由此可以看出，超过 1/3 的华资企业其上游产品来自本地和来自中国其他地方的比例不到 10%，所用原料 100% 来自中国的华资企业不到 5%，而接近 20% 的华资企业上游产品 100% 来自国外。

表 15 欧盟企业上游产品来源

项 目	来自本地		来自中国 其他地方		来自国外		来自 “三资”企业	
	企业 数量	所占 比例	企业 数量	所占 比例	企业 数量	所占 比例	企业 数量	所占 比例
10% 以下	14	31.11	12	26.67	3	6.67	9	20.00
20%	3	6.67	1	2.22	2	4.44	1	2.22
30%	2	4.44	1	2.22	3	6.67	2	4.44
40%	1	2.22	1	2.22	1	2.22	0	0
50%	1	2.22	3	6.67	0	0	1	2.22
60%	0	0	0	0	0	0	0	0
70%	1	2.22	2	4.44	2	4.44	1	2.22
80%	1	2.22	1	2.22	2	4.44	0	0
90%	0	0	0	0	1	2.22	0	0
100%	0	0	1	2.22	11	24.44	1	2.22

注：企业数量单位：家；所占比例单位：%。

来自本地、来自中国其他地方和来自“三资”企业的上游产品占其上游产品总和的比例都在 10% 以下，是多数欧盟企业的共同之处。与华资企业相比，更多的欧盟企业完全使用来自国外的上游产品，没有一家企业的上游产品 100% 来自本地。

表 16 日资企业上游产品来源

项 目	来自本地		来自中国 其他地方		来自国外		来自 “三资”企业	
	企业 数量	所占 比例	企业 数量	所占 比例	企业 数量	所占 比例	企业 数量	所占 比例
10% 以下	14	38.89	13	36.11	6	16.67	4	11.11
20%	3	8.33	5	13.89	0	0	0	0
30%	2	5.56	1	2.78	1	2.78	2	5.56
40%	1	2.78	0	0	1	2.78	0	0
50%	1	2.78	0	0	0	0	0	0

续上表

项 目	来自本地		来自中国 其他地方		来自国外		来自 “三资”企业	
	企业 数量	所占 比例	企业 数量	所占 比例	企业 数量	所占 比例	企业 数量	所占 比例
60%	0	0	2	5.56	2	5.56	0	0
70%	1	2.78	0	0	1	2.78	0	0
80%	1	2.78	0	0	3	8.33	0	0
90%	0	0	1	2.78	1	2.78	1	2.78
100%	0	0	0	0	9	25.00	3	8.33

注：企业数量单位：家；所占比例单位：%。

由表 16 可以看出，日资企业的上游产品很少来自中国，大部分上游产品来自国外，还有部分上游产品来自“三资”企业。将日资企业上游产品的来源与另外三组外资企业进行比较，发现上游产品 100% 来自国外和 100% 来自“三资”企业的日资企业比重是四组外资企业中比例最高的。

表 17 美资企业上游产品来源

项 目	来自本地		来自中国 其他地方		来自国外		来自 “三资”企业	
	企业 数量	所占 比例	企业 数量	所占 比例	企业 数量	所占 比例	企业 数量	所占 比例
10% 以下	17	32.08	12	22.64	10	18.87	7	13.21
20%	5	9.43	4	7.55	4	7.55	1	1.89
30%	1	1.89	0	0	2	3.77	1	1.89
40%	0	0	0	0	2	3.77	0	0
50%	1	1.89	4	7.55	2	3.77	1	1.89
60%	1	1.89	3	5.66	0	0	1	1.89
70%	1	1.89	1	1.89	1	1.89	0	0
80%	0	0	5	9.43	5	9.43	1	1.89
90%	0	0	2	3.77	1	1.89	1	1.89
100%	3	5.66	0	0	6	11.32	3	5.66

注：企业数量单位：家；所占比例单位：%。

相比之下，美资企业的上游产品较平均地来自中国其他地区、国外和“三资”企业。它们较少采用来自本地的产品。

上游产品的来源表现了外资企业与当地企业的后项联系；当地购买的比例取决于MNCs的市场取向、母公司的国籍、东道国的政策及其当地潜在供应商的技术能力。通常情况下，市场导向的MNCs比出口导向的MNCs从当地购买更多的商品和劳务；欧洲MNCs比美国和日本MNCs更多地依赖当地市场。

由上面的分析可以看出，四组外资企业从广东本地购买上游产品的比例较少。这就需要广东本地的企业练好内功，提高自己产品的竞争力，争取外资企业的供货市场。

外资企业通过其产品的示范模仿效应、对内部员工的培训、人力资本的流动、与当地厂商的联系，对我国的技术进步作出了贡献，它们的产品在我国市场面临的竞争也越来越激烈。

五、产品市场的竞争情况

一般情况下，如果企业的产品在市场上没有或较少有同类竞争产品，就表明企业拥有较强的技术优势，这可以作为企业技术水平高低的一个指标。表18至表21是对四组外资企业各主要产品在企业建立初期及现在的市场竞争情况的分析。其中“有（无）竞争品”表示企业的主要产品在市场上有（无）同类产品，“不知”表示企业对此情况并不了解，“无主产品”表示企业没有该项主要产品。

表18 主要产品一的竞争情况

		项 目	华资		欧盟		日资		美资		总计	
			企业数量	所占比例								
主要 产品 一	建 立 初 期	有竞争品	202	71.89	24	53.33	28	77.78	30	57.69	284	68.43
		无竞争品	29	10.32	10	22.22	4	11.11	13	25.00	56	13.49
		不知	5	1.78	2	4.44	0	0.00	1	1.92	8	1.93
		无主产品	45	16.01	12	26.67	4	11.11	8	15.38	69	16.63
	现 在	有竞争品	222	79.00	27	60.00	31	86.11	40	76.92	320	77.11
		无竞争品	6	2.14	4	8.89	1	2.78	5	9.62	16	3.86
		不知	4	1.42	1	2.22	0	0.00	0	0.00	5	1.20
		无主产品	49	17.44	13	28.89	4	11.11	7	13.46	73	17.59

注：企业数量单位：家；所占比例单位：%。

表 19 主要产品二的竞争情况

	项 目	华资		欧盟		日资		美资		总计		
		企业	所占	企业	所占	企业	所占	企业	所占	企业	所占	
		数量	比例	数量	比例	数量	比例	数量	比例	数量	比例	
主要产品二	建立初期	有竞争品	112	39.86	17	37.78	16	44.44	19	36.54	164	39.52
		无竞争品	24	8.54	4	8.89	6	16.67	9	17.31	43	10.36
		不知	1	0.36	0	0.00	0	0.00	1	1.92	2	0.48
		无主产品	144	51.25	24	53.33	14	38.89	23	44.23	205	49.40
	现在	有竞争品	134	47.69	19	42.22	20	55.56	25	48.08	198	47.71
		无竞争品	5	1.78	1	2.22	2	5.56	5	9.62	13	3.13
		不知	1	0.36	0	0.00	0	0.00	0	0.00	1	0.24
		无主产品	141	50.18	24	53.33	14	38.89	22	42.31	201	48.43

注：企业数量单位：家；所占比例单位：%。

表 20 主要产品三的竞争情况

	项 目	华资		欧盟		日资		美资		总计		
		企业	所占	企业	所占	企业	所占	企业	所占	企业	所占	
		数量	比例	数量	比例	数量	比例	数量	比例	数量	比例	
主要产品三	建立初期	有竞争品	60	21.35	11	24.44	12	33.33	16	30.77	99	23.86
		无竞争品	16	5.69	2	4.44	2	5.56	7	13.46	27	6.51
		不知	0	0.00	1	2.22	0	0.00	1	1.92	2	0.48
		无主产品	205	72.95	31	68.89	22	61.11	28	53.85	286	68.92
	现在	有竞争品	70	24.91	11	24.44	13	36.11	19	36.54	113	27.23
		无竞争品	6	2.14	2	4.44	1	2.78	4	7.69	13	3.13
		不知	1	0.36	0	0.00	0	0.00	0	0.00	1	0.24
		无主产品	204	72.60	32	71.11	22	61.11	29	55.77	287	69.16

注：企业数量单位：家；所占比例单位：%。

表 21 主要产品四的竞争情况

	项 目	华资		欧盟		日资		美资		总计		
		企业	所占	企业	所占	企业	所占	企业	所占	企业	所占	
		数量	比例	数量	比例	数量	比例	数量	比例	数量	比例	
主要产品四	建立初期	有竞争品	21	7.47	6	13.33	5	13.89	10	19.23	42	10.12
		无竞争品	8	2.85	2	4.44	0	0.00	5	9.62	15	3.61
		不知	0	0.00	1	2.22	0	0.00	0	0.00	1	0.24
		无主产品	252	89.68	36	80.00	31	86.11	37	71.15	356	85.78
	现在	有竞争品	27	9.61	8	17.78	5	13.89	12	23.08	52	12.53
		无竞争品	2	0.71	1	2.22	0	0.00	4	7.69	7	1.69
		不知	0	0.00	0	0.00	0	0.00	0	0.00	0	0.00
		无主产品	252	89.68	36	80.00	31	86.11	36	69.23	355	85.54

注：企业数量单位：家；所占比例单位：%。

我们对外资企业成立初期及现在国内市场上是否存在与其主要产品相同或类似的同类产品进行考察时，发现了颇为有趣的三点：

(1) 无论在企业成立初期或是现在，五成以上的华资企业只拥有一种主要产品，仅有一成华资企业生产四种（及以上）主要产品。欧盟企业只生产一种主要产品的企业数量也超过五成，但却有 20% 的欧盟企业拥有四种（及以上）主要产品，这一比例在四组企业中是相当高的。拥有最多主要产品的是美资公司，近六成美资公司拥有两种主要产品，生产四种（及以上）主要产品的美资公司比重超过 30%。

(2) 这四组外资企业，其四种主要产品现在在我国所面临的市场竞争，明显高于其成立初期在我国所面临的竞争。

(3) 在四组外资企业中，日资公司生产的各种主要产品在国内拥有最多的同类产品，换言之，日资公司生产的产品在国内面临的竞争最强。这一点可以通过表 22 反映出来。

表 22 日资企业主要产品在国内市场有同类产品的企业比例（%）

项 目		日资企业	平均水平
主要产品一	初期有同类产品	77.78	68.43
	现在有同类产品	86.11	77.11
主要产品二	初期有同类产品	44.44	39.52
	现在有同类产品	55.55	47.71
主要产品三	初期有同类产品	33.33	23.86
	现在有同类产品	36.11	27.23
主要产品四	初期有同类产品	13.89	10.12
	在有同类产品	13.89	12.53

越多日资公司生产的产品在国内有同类产品，就表示日资公司面临的竞争越多、越强。而日资企业在这方面的 8 个比例全部远高于平均水平。这一点似乎与前面提到的日资企业的产品水平高（见表 8、表 9）不符。我们将具体分析四组企业的投资行业，看看能不能从中获得答案。

四组企业由于外溢效应的作用，从不同的方面促进了我国企业产品水平的提高，也使它们在我国面临的竞争越来越激烈。但从前面的分析中也可以看出我国企业面临着一些问题，如外资企业从我国当地企业购买原料并不多。这就对我国企业提出新的要求：应该练好内功，提高自己产品的水平，抓住外资企业这块大市场。

参考文献

- [1] 何洁，许罗丹：《中国工业部门引进外国直接投资外溢效应的实证研究》，《世界经济文汇》，1999 年第 2 期。
- [2] 课题组：《看好潜在市场规模——欧盟在华投资中小企业业绩及经营战略调查》，《国际贸易》，2002 年第 4 期。

- [3] 李平：《技术扩散中的溢出效应分析》，《南开学报》，1999年第2期。
- [4] 沈坤荣，耿强：《外商直接投资的外溢效应分析》，《金融研究》，2000年第3期。
- [5] 王子君，张伟：《外商直接投资、技术许可与技术创新》，《经济研究》，2002年第3期。
- [6] 张雪：《跨国公司在中国的技术溢出效应分析——以汽车为例》，《世界经济研究》，2003年第9期。

新自由主义贸易政策论及其演进*

黄静波

新自由主义作为一种经济思潮形成于 20 世纪 30 年代，在 60 年代以后得到迅速发展，并开始渗入国际贸易领域，在这一领域有重要影响的新自由主义代表人物包括：巴拉萨 (B. Balassa)、巴格瓦蒂 (J. N. Bhagwati)、克鲁格 (Anne O. Krueger)、拉尔 (D. Lal)、里特尔 (I. Little)、斯瑞尼瓦桑 (T. N. Srinivasan) 等^[1]。新自由主义者从研究贸易战略和贸易体制入手，进一步在理论上进行阐释，从而探讨一系列相关的贸易政策问题，逐步形成了其贸易政策理论的主要观点：

一、保护成本论——对进口替代战略的全面评价

战后伊始，一大批发展中国家实行了进口替代战略。这些国家相信，由于发展中国家的市场功能和结构不完整，其经济发展不可能建立在完全的市场机制之上，国家干预和计划管理便成了加速资本形成、改进资源配置和实现迅速工业化仅可信赖的力量。^[2 3]进口替代被认为是低收入发展中国家实现工业化不可避免的发展阶段。

进口替代的特征是高关税、严格的进口限额和本币汇率高估等严重的保护措施，这些贸易措施通过影响资源配置从出口部门转向进口竞争部门来推动本国工业化，高关税和严格的进口限额保护本国工业免受外来竞争，而本币汇率高估则降低那些用于工业化过程的进口资本品的价格。

保护主义的觞殇引起了经济学家们的关注，20 世纪 60 年代起，关于有效保护率的研究迅速展开，衡量有效保护的各种方法相继提出，为评价进口替代战略的效果奠定了基础。1970 年，里特尔 (Little) 等人首次全面地评估进口替代战略的效果，指出，进口替代伴随着非常高的有效保护平均水平，这种高度保护是以本币高估抑制出口和巨大的国内资源成本作为代价的，它导致了资源的错置、对资本品进口的依赖和经济效率的低下。

此后，大量的相关研究围绕着进口替代战略的主要特征从各种角度进行了分析。重要的观点包括：①汇率高估的初始目的是有利于资本品和中间产品的进口，以便加速进口替代工业化。但是，本币定值过高将使得进口品的国内价格降低，而出口产品的国外价格却上升，出口创汇更为艰难，其必然结果是出口量减少，这是进口替代保护政策的反出口偏向 (anti-export bias)。②内向性的进口替代战略低估贸易在发展中的作用，有减少贸易的倾向。然而，由于那些新扩张的行业通常缺少充分的技术基础使生产合理化，进口替代本身又是一个相当长的过程，因此，随着进口替代产业的成长，相应的资本品进口也一同增

* 原载：《经济学动态》，2003 年第 11 期。

长,此为引致进口(induced import)。③出口价格下降,而进口资本品的需求不断增长,外汇短缺愈发严重。外汇缺口必定压迫出口,不得不在出口价格下降时增加出口量,但出口收益并不能得到增加,这种情况被称为“贫困化增长”(immiserizing growth)。

克鲁格在2000年发表的一篇文章中概括了进口替代战略的静态和动态的成本。^[4]关于进口替代的静态效应,克鲁格指出:第一,发展中国家在发展的初期严重依赖制成品进口,也比发达国家更为敏感于贸易政策,发达国家贸易保护的增减会导致产出构成的变化,而发展中国家贸易保护的结构则决定着资源配置的方向和生产的结构。第二,进口能力不足促使发展中国家以高成本的国内产品替代资本密集的进口品,将资源从具有比较优势的劳动密集领域转移到不具有优势的资本密集领域,由于生产所需的资本品价格高,这意味着实际投资水平降低,从而经济增长更慢。第三,发展中国家收入低、市场小,进口替代的内向性使生产达不到最低的效率规模,得不到规模经济效益。第四,进口替代吸收了新资源,出口被阻碍,而“进口替代本身是进口密集的”,外汇供给不足导致更严厉的进口限制和外汇管制。第五,进口替代使官方使用更多行政性措施来决定资源的分配。第六,进口替代体制越来越复杂,要耗费更多的资源来运作和管理。克鲁格也指出,除了造成资源错置和效率损失外,进口替代损失的静态源还包括:①贸易扭曲的生产成本;②进口许可证程序中的寻租和腐败造成的损失;③实施其他数量限制的成本和迟滞所带来的损失;④生产者在国内市场拥有垄断地位所产生的损失。“所有这些损失远远超过生产成本”。

克鲁格认为,进口替代的动态成本远高于其静态损失,因为该战略的成本一直增加而增长率则不断降低。第一,一国实行进口替代战略会始于较容易的部门,接近其比较优势,但随后由保护引致的活动就渐渐远离比较优势,这意味着经济增长率降低。第二,国内制成品市场小,而资本设备由于其具有不可分割性,其总体利用率低,加大了资本-产出的比率,降低增长率。第三,进口替代下的保护政策存在许多负面效果,使得政府不得不时而实施调整或稳定措施,进口替代具有一种时进时退(stop-go)的型式,也会降低增长率。第四,进口限制降低了人们积累和使用知识资本、外国先进技术的比率,从而降低经济增长。第五,进口替代体制下,政策的反馈机制很弱,如进口限额的关税当量不得而知,对本币汇率高估的程度不敏感等等。

作为对比,出口导向或促进出口战略显示出了其优越性。出口导向是以出口产业发展来带动整体经济,总体而言,出口导向战略立足于国内外两个市场,以比较优势理论为基础,能更有效配置资源。

二、贸易扭曲论——对贸易政策地位的界定

扭曲被定义为经济活动对帕累托最优状态的偏离。在贸易方面,新古典学派和新自由主义都秉承古典国际贸易理论的教条,认为无干预的自由贸易是一种最理想的状态。然而,符合帕累托最优状态的放任所要求的严格条件是不可能达到的,因为这些条件不仅要求所有的市场完美运转、没有外部性,而且要求所有的要素市场和商品市场(包括所有偶然事件)都是完美的,所有的市场参与者也都具有对这些市场的完备知识。

扭曲的根源，一方面是在于市场的不完善，或者说是市场失灵；另一方面则是由于政府政策所致。市场失灵或政策失当使得帕累托最优的条件 $DRS = DRT = FRT$ （即：国内边际消费替代率 = 国内边际生产转换率 = 国外边际转换率）中这个或那个条件不能得到满足而出现扭曲。巴格瓦蒂区分了扭曲的主要类型。就经济性扭曲而言， $DRS = FRT \neq DRT$ 时出现生产扭曲， $DRT = FRT \neq DRS$ 时出现消费扭曲，而 $DRS = DRT \neq FRT$ 导致国外市场扭曲即所谓贸易扭曲。消费扭曲和生产扭曲属国内扭曲，贸易扭曲则是国际扭曲。^[5]从产生扭曲的原因来看，内生性扭曲是致使自由放任的竞争性市场达不到帕累托最优的某种因素以及市场失灵所造成的扭曲。政策性扭曲是由各种经济政策产生的扭曲，又可分为自主性扭曲和非自主性扭曲。自主性扭曲是政策并非作为实现某种特定目标的工具时出现的扭曲；非自主性扭曲则是经济政策的工具性造成的扭曲。^[6]

扭曲理论的一个重要发展是对“直接非生产性寻利”（DUP）行为的分析。巴格瓦蒂与斯瑞尼瓦桑指出，外生型政策直接非生产性寻利会引致以另一种途径增加的额外扭曲，而内生型政策直接非生产性寻利则存在多种可能，减少内生型政策直接非生产性寻利在某些情形下可能改进福利，但在另一些情形下就未必。^[6]

因此就有经济政策的排序（ranking）问题，存在着扭曲的时候，经济就只能是追求一种“次优”（second-best）状态，即低于帕累托最优的相对最好状态。正是由于扭曲的存在，边际社会收益与边际社会成本间就存在差异，这为政府采取某种政策干预经济和贸易提供了空间。^[7]不过，按照“次优”理论，人们并不能先验地确定自由贸易或贸易保护就一定改善经济状况。

在一定情形下，贸易政策在纠正扭曲方面是可用的。例如，当贸易中存在垄断时，平均贸易条件不等于边际贸易条件，即一国平均的单位出口产品所能换取的进口品数量的相对比率与一国边际出口品所愿意换取的进口品数量的比率存在差异。以关税措施提高其出口品或降低其进口品的相对价格，使得征收关税后国内价格与边际贸易条件相同，即该国达到 $DMRS = DMRT = FMRT$ 的状态，其社会福利最大（虽然此时世界福利未必最大），这种关税即所谓最佳关税。在这个意义上，关税是纠正贸易扭曲的最佳政策。但如果采用国内税收或补贴来纠正贸易扭曲，结果可使得 $DMRT = FMRT$ ，但 $DMRS \neq DMRT$ ，即消除了贸易扭曲却又产生了国内扭曲。同样，以贸易政策（如关税）来纠正国内扭曲也会有纠正了国内扭曲但又产生贸易扭曲的情况。所以，纠正扭曲的最好政策是针对扭曲产生的根源采取措施。否则的话，政策运用不当又会产生政策扭曲。

应当说，建立和完善“最优 - 次优”说是经济学中的重要发展，但是，新自由主义者指出，不能以“次优”作为干预经济贸易的藉口。克鲁格猛烈抨击了学术界误用优秀经济理论（如比较成本说）从而包容干预和保护的现象。^[8]在新自由主义看来，包容干预的“中间道路”和“第三条道路”将导致制度上和功能上的冲突。只有市场才能解决一切问题，需要一种向自由市场经济制度的回归。如果市场有什么不能克服的缺陷，克服的办法是明晰产权，使市场机制赖以作用的基础得到完善，而不是政府干预；况且，政府行为也有可能产生所谓政府缺陷。

三、中性贸易体制论——对贸易干预的否定

任何一种贸易主张都必须有相应的体制来予以保证。新自由主义者在论证自由贸易体制的优越时无一例外地强调该体制的自由、无干预的所谓“中性”特征，要求实行中性贸易体制。

所谓贸易体制的中性，其核心在于对进口和出口的刺激具有对等性。这些对等的刺激不是来自人为的措施，而只是干预最小时市场机制自然作用的结果。必要的干预仅仅体现为以恢复对进口和出口刺激对等性为目的的偶然措施。而进口替代和极端出口促进均依赖于政府具有偏向性的政策，如本币汇率高估、严厉的进口限制、对出口的补贴等等，这些政策往往是形成扭曲的重要原因。

克鲁格提出贸易体制偏向指数的方法来测度贸易体制偏向，该指数（用 B 代表）比较进口品国内价格相对于进口品国际价格的扭曲和出口品国内价格相对于出口品国际价格的扭曲， $B = 1$ 代表贸易体制的中性， $B > 1$ 为内向型体制， $B < 1$ 为外向型体制。巴格瓦蒂采用有效汇率来衡量贸易体制偏向，即计算出口有效汇率（ EER_x ）和进口有效汇率（ EER_m ）并加以比较。有三种情况： $EER_x < EER_m$ 为进口替代战略，是一种具有“反出口偏向”的体制； $EER_x > EER_m$ 是极端的出口促进战略； $EER_x \approx EER_m$ 则是出口促进战略，是一种中性的贸易体制。

新自由主义者认为，中性贸易体制有以下几个方面的优点：第一，对进口和出口大体平衡的有效汇率（并且通常是单一的汇率）消除了“反出口偏向”，刺激国内资源有效配置，无需补贴或限制措施来进行调整，使国内资源成本接近国际的机会成本水平，产生有效率的结果。第二，由于市场直接配置资源，减少乃至消除扭曲，这就减少了直接非生产性寻利或寻租行为。第三，开放的环境有利于外国投资，外资流入不是为了绕过关税，而是出于在世界市场竞争的考虑，这与东道国的出口促进战略是一致的，有助于提高社会生产率。第四，可以导致更多的竞争和更少受保护的市场，从而促使更多的技术创新，并且开放的贸易也带来不断向国外学习的动态效果。出口促进有着提高生产率的增长率的积极作用。第五，出口促进是面向世界市场，市场的广阔性有助于实现更有效的专业化分工，改进生产要素的利用，获取规模经济效益。第六，尽管难以定量说明，但出口促进体制下， X -效率要优于进口替代体制，因为被保护或垄断的因素减少，对出口和进口替代的刺激基本对等，企业员工的积极性得到更充分发挥，使得企业非资源配置性效率（即 X -效率）得到提高。第七，切合实际的汇率机制可以缓解外汇短缺的瓶颈，更多地进入国际资本市场。^[9, 40]

四、贸易自由化论——向自由贸易的复归

以保护主义为特征的进口替代不利于发展，而中性贸易体制恰在这些方面具有优越性，因此，新自由主义的政策结论就是在一个存在扭曲、贸易保护过甚的世界里推行贸易自由化，进行贸易政策改革，从而建立起中性贸易体制，向自由贸易复归。

“贸易自由化是更多地应用价格机制使贸易制度更加透明以及使国内价格更接近于世界市场价格的过程。”^[11]这一定义强调贸易过程中市场价格机制的决定性作用，并且包含了贸易制度中立性的要求。通过减少反出口偏向，特别是取消数量限制，贸易自由化将导致一个更中立和透明的贸易体制。作为一个“过程”，贸易自由化也可能具有阶段性的特征。爱德华茨指出，这个定义接近于1978年由克鲁格和巴格瓦蒂各自主持的美国国家经济研究署(NBER)项目所作的自由化的定义。20多年之后，克鲁格仍然强调，“贸易自由化明确地意味着贸易体制偏向的减少，朝外向型贸易战略转变意味着变为很小甚至没有贸易体制偏差。”^[4]

贸易自由化要求进行贸易政策改革。“贸易政策改革包括贸易体制向中立性激励结构转变、趋向于更自由的贸易体制或二者兼而有之。中立性体制给予出口和进口、国内销售和出口、可贸易商品和非贸易商品的激励不偏不倚。更自由的贸易体制是指贸易控制的减少和以价格机制代替直接干预。”^[12]既然贸易自由化的方向是建立中立的或更自由的贸易体制，贸易政策改革的最重要的目标因此就是“使价格正常化”，要使市场机制成为制定贸易政策的准则。只有在市场机制充分发挥作用的基础上，各种贸易政策才能最有效地刺激贸易的发展，从而为改革的国家带来静态和动态的贸易利益。因此，恢复或建立起一个市场经济体制实际上是贸易政策改革和贸易自由化的根本所在。

一般认为，典型的贸易政策改革次序通常是：首先，通过本币贬值将汇率调整至切合实际，可以使进口数量限制变得多余，同时也有助于将这些限制迅速取消并且增加出口者的相对利益。其次，在进口改革之前进行出口政策改革（比如暂时出口许可或出口退税制等）。这些改革将为把关税统一起来提供支持。再次，将数量限制转化为等同的关税措施，因为关税一般而言更少产生紊乱而且能为政府带来税收收益。最后，为了进一步自由化，关税的总水平应降低。

参考文献

- [1] Colclough, Christopher, Structuralism versus Neo-liberalism: An introduction, In Colclough, Christopher and James Mannor (eds), States or Markets? — Neo-liberalism and the Development Policy Debate, Oxford: Clarendon Press, pp1 ~ 25.
- [2] Krueger, Anne O., Economic Policy Reform in Developing Countries (The Kuznets Memorial Lectures at the Economic Growth Centre, Yale University), Oxford: Blackwell, 1992.
- [3] Krueger, Anne O., Trade Policies and Developing Nations, Washington DC: The Brookings Institution, 1995.
- [4] Krueger, Anne O., Why trade liberalization is good for growth, In Dixon, Huw David (ed), Controversies in Macroeconomics: Growth, Trade and Policy, Oxford: Blackwell, 200, pp75 ~ 84.
- [5] Bhagwati, Jagdish N., The generalized theory of distortion and welfare, in J. Bhagwati et al. (eds), Trade, Balance of Payment and Growth: Essays in Honor of Charles Kindleberger, North-Holland, 1971, pp53 ~ 67.
- [6] Srinivasan, T. N., The generalized theory of distortion and welfare: Two decades later, In Feenstra, Robert C., Gene M. Grossman and Douglas A. Irwin (eds), The Political Economy of Trade Policy: Papers in Honor of Jagdish Bhagwati, Cambridge: The MIT Press, 1996, pp3 ~ 25.
- [7] Anderson, James E., The theory of protection, In David Greenaway and Alan Winters (eds), Surveys in International Trade, Blackwell, 1994, pp106 ~ 138.

- [8] Krueger , Anne O. , Trade policy and economic development : How we learn , The American Economic Review , 1997 , Vol.87 , No.1 , pp1 ~ 22.
- [9] Bhagwati , Jagdish N. , Export-promoting trade strategy : Issues and evidence , World Bank Research Observer , 1988 , Vol.3 , No.1 , pp27 ~ 57.
- [10] Salvatore , Dominick and Thomas Hatcher , Inward oriented and outward oriented trade strategies , Journal of Development Studies , 1991 , Vol.27 , No.3 , pp 7 ~ 25.
- [11] Edwards , Sebastian , Openness , Outward Orientation , Trade Liberalisation and Economic Performance in Developing Countries , Policy , Planning , and Research Working Papers , No.191 , The World Bank , 1989.
- [12] Thomas , Vinod and John Nash , Trade policy Reform : recent evidence from theory and practice , In Adhikari , R. , Kirkpatrick , C. and Weiss , J. (eds) , Industrial and Trade Policy Reform in Developing Countries [M] , Manchester University Press , 1992.
- [13] Jung , Woo S. and Peyton J. Marshall , Exports , growth and causality in developing countries , Journal of Development Economics , 1985 , Vol.18 , No.1 , pp1 ~ 12.
- [14] Grossman , Gene M. and Elhanan Helpman , The new growth theory : Trade , innovation , and growth , The American Economic Review , 1990 , Vol.80 , No.2 , pp86 ~ 91.
- [15] 丹尼·罗德瑞克：《让开放发挥作用：新的全球经济与发展中国家》，中国发展出版社 2000 年版。
- [16] Romer , Paul , Increasing Returns and Long-Run Growth , Journal of Political Economy , 1986 , Vol.94 , No.5 , pp1002 ~ 1037.
- [17] 马克·威廉姆斯：《国际经济组织与第三世界》，经济科学出版社 2001 年版。

对外直接投资理论的发展*

许罗丹 谭卫红

一、引言

第二次世界大战，特别是 20 世纪 60 年代以后，发达国家跨国公司及其对外直接投资急剧发展，对外直接投资在国际资本流动中逐渐占据主要地位，这些现象引起经济学界的关注，有关对外直接投资的理论研究越来越多。学者从不同的角度、不同的层次研究对外直接投资的起因及其决定因素，以及对东道国的影响，形成了有关对外直接投资的众多理论。由于对外直接投资是跨国公司最基本的特征，对外直接投资的主体是跨国公司，所以解释对外直接投资的各种理论又称为跨国公司理论。

严格说来，关于对外直接投资问题的研究，始于 20 世纪 60 年代^①。海默 (S.H.Hymer) 在其博士论文《国内企业的国际经营：关于对外直接投资的研究》中，首先提出了垄断优势理论；垄断优势理论的提出标志着对外直接投资理论的兴起。随后，学者们又相继提出了其他理论，这些理论根据研究的侧重点不同，大致分为直接投资与国际贸易的关系、直接投资的动机与成因分析、直接投资对东道国的影响三个方面。

二、对外直接投资与国际贸易关系的研究

对外直接投资与国际贸易有着密切的联系。不同的学者运用理论或实证的分析方法，对它们的关系得出不同的看法。有关对外直接投资与国际贸易关系的研究，主要集中在替代、互补两个方面。

投资与贸易的关系最初是由 Mundell (1957) 提出来的。Mundell 在赫克歇尔 - 俄林模型的基础上，建立两个国家、两种商品、两种生产要素的模型，考察贸易和投资相互替代的两种极端情况，即禁止性投资如何促进贸易和禁止性贸易如何刺激投资。他认为：在存在关税的情况下，投资与贸易具有替代性，贸易障碍会产生投资，投资障碍会导致贸易。Mundell 运用要素比例理论解释商品的国际流动，运用资本边际产量的差异解释资本的国际流动。贸易障碍会对两个国家之间的资本边际收益产生影响，因此贸易障碍在一定条件下会导致资本的国际流动或直接投资。Kojima (1975) 运用实证分析的方法，发现水平型外商直接投资与并购对出口有替代效应。Belderbos 和 Sleuwaegen (1996) 利用公司层面的

* 原载：《世界经济》，2004 年第 3 期。

① 小岛清：《对外贸易论》，南开大学出版社 1987 年版。

数据发现,在目标市场存在着潜在的或者实际的贸易保护的情况下,投资与贸易之间存在着替代关系。

Buckley 和 Casson (1981) 考虑运输成本、关税、投资经营的固定成本后认为:在较低的销售量的情况下,公司倾向于出口以避免在国外生产所需要的较高的固定成本;反之则倾向于在国外生产而不是出口。Markuson 和 Melvin (1983) 利用非要素比例模型阐述了要素流动与商品贸易之间的互补关系。Markuson 和 Svensson (1985) 利用要素比例模型,指出贸易和非贸易要素之间的“合作”与“非合作”关系,决定了商品贸易和投资之间是替代还是互补的关系;贸易和非贸易要素的合作关系,导致商品贸易和生产要素的流动相互促进,从而表现为互补关系,反之则表现为替代关系。Helpman (1984)、Helpman 和 Krugman (1985) 认为在要素禀赋不对称和规模报酬递增的情况下,由于跨国公司的专有资产很难通过外部市场达成交易,就会产生大量的公司内交易和对中间产品的需求,由此带动母国的出口贸易。Caves (1996) 认为:一般来说,对外直接投资对东道国出口贸易的促进作用有两个方面:一个是直接效应,即通过外商直接投资企业自身的出口来带动东道国的出口;另一个是间接效应,即对外直接投资通过对当地企业的影响促进其出口。Kojima (1975)、Graham (1996) 认为投资母国通过在东道国投资设厂,发展垂直型投资与贸易有互补关系。Swenson (1997) 发现:在美国投资的日资汽车生产企业,从日本本国进口大量的零部件。

我国学者对中国吸引外资与贸易的关系也进行了不少研究。冼国明等 (2003) 依据我国改革开放以来的数据,运用单位根检验、向量误差修正模型、Granger 因果检验等方法,对近 20 年来外商在华投资与中国出口之间的相关性进行研究,分析得出外商直接投资与我国的出口之间存在长期的均衡关系;出口的增加对我国吸引外国直接投资具有很强的促进作用。王洪亮、徐霞 (2003) 研究日本对华贸易与直接投资的关系,发现日本对华直接投资和中日贸易之间存在长期的互补关系。

三、外商直接投资的动因分析

对外商直接投资动机与原因的研究大致分为两个方面:早期的研究主要集中在跨国公司为什么对外进行投资;后期的研究则从投资东道国的区位优势角度,分析投资母国为什么选择这一国或地区而不是其他的国家和地区进行投资。

Hymer (1960) 在其博士论文《国内企业的国际经营:对外直接投资的一项研究》中首先提出了垄断优势理论。垄断优势理论认为:因为市场不完全即不完全竞争的存在,使企业可以拥有和保持垄断优势。市场不完全主要表现为:产品市场不完全、要素市场不完全、由规模经济引起的市场不完全及由政府管理行为造成的市场不完全;由此导致企业拥有产品市场优势、要素市场优势、规模经济优势、政府管理行为优势。投资母国企业具有的这些比东道国同类企业有利的垄断优势,是企业进行对外直接投资的动因。Aliber (1970) 在《对外直接投资理论》一文中,以垄断优势理论为基础,提出了资本化率理论。该理论认为,企业进行对外直接投资的动因在于不同国家的资本化率存在差异,并着重分析货币等宏观经济变量对跨国公司直接投资的影响。Vernon (1966) 提出产品周期理论。

该理论从产品比较优势和竞争条件变化的角度分析对外直接投资产生的原因,认为产品的生产地取决于该产品生命周期的不同阶段。在产品创新阶段,产品生产集中在国内;成熟阶段,企业开始进行防御性的对外直接投资;标准化阶段,企业完成出口向对外直接投资的转变,对外直接投资结构基本稳定。Buckley 和 Casson (1976) 在《跨国公司的未来》一书中提出了内部化理论,Rugman (1981) 在《跨国公司:内部化市场经济学》中又对内部化理论做了进一步发展。内部化^①理论引入交易成本,认为只要在某个地方国际资源配置内部化要比利用市场的成本少,那里就会出现跨国企业;跨国公司进行对外直接投资的动因是为了获得内部化利益。Dunning (1977) 提出了国际生产折中理论。该理论的核心是“OLL 模式”,即所有权优势、内部化优势、区位优势决定了跨国公司的行为和对外直接投资。只有当企业同时具有三种优势时,才会对外进行直接投资。当具有所有权优势和内部化优势时,企业选择出口;在只具有所有权优势时,企业采取转让无形资产的方式。Kojima (1978) 在《对外直接投资:跨国经营的日本模式》一书中提出了比较优势理论。该理论从国际分工合理化的角度出发解释跨国公司对外直接投资的决定因素;认为对外直接投资应该从本国(投资国)已经处于或即将陷于比较劣势的产业——可称为边际产业——依次进行;同时指出日本式对外直接投资是顺贸易导向型对外直接投资,美国式对外直接投资是反贸易导向型对外直接投资。

有关东道国投资优势的研究主要集中在区位优势理论研究上,国际上有关这方面的研究已有不少成果,这些研究多集中于传统的比较优势因素方面,如市场容量、劳动力成本、交通和通讯成本、相对技术水平等(Vernon, 1966)。外商直接投资区位因素分析的研究方法大致可以划分为两大类:问卷调查分析和区位模型分析。采用问卷调查的方法也是现代投资区位理论研究中最常用的研究方法之一(魏后凯等,2001)。区位模型分析一般运用各种数学模型,如多元线性和非线性回归模型、数量回归模型和“逻辑”选择模型等。Dunning 在国际生产折中理论中,认为区位优势理论区位优势理论从投资东道国的角度说明了投资东道国在吸引外商直接投资方面所具有的优势。在三优势理论中,Dunning 把直接投资分为 5 个类型:资源开发型、生产或加工专业化型、服务型、出口导向型、市场导向型。区位优势理论认为:影响外商直接投资的因素主要是东道国广阔的产品消费市场、政府的各种优惠政策、低廉的生产要素成本、当地原材料的可供性等;但市场导向型与出口导向型投资对这些政策的反应又不尽一致。市场导向型投资更看重当地的消费水平;出口导向型则更偏好于低廉的生产要素成本与运输成本。小岛清根据对外直接投资的动机,把外商直接投资分为三种类型、自然资源导向型、市场导向型和生产要素导向型,投资东道国在自然资源、市场、生产要素方面的优势决定了投资母国的区位选择和投资类型。

很多学者以区位优势理论为基础,利用实证或调查研究的方法,分析投资东道国的这些区位优势对他们吸引外商直接投资的影响弹性。Clegg (1992) 认为关税、非关税壁垒、税收优惠以及语言文化差异等也会影响外商直接投资地区分布。Gong 利用统计模型研究了

^① 内部化是指把市场建立在企业内部的过程,由内部市场取代外部市场,即建立由企业内部调拨价格起作用的内部市场,使之像固定的外部市场那样有效地发挥作用。

中国外商直接投资与一些关键区位因素间的定量关系，如基础设施、投资刺激、工业基础、政策优惠、教育水平、工资成本等。贺灿飞等（1997）对港澳地区在中国内地投资的区位选择及其空间扩散作了实证分析。Tatogluand Gaister（1998）认为市场规模和经济增长、材料和劳动力供应、政治和法律环境、东道国政策、东道国市场上产业竞争程度、地理接近程度以及交通运输成本、东道国的基础设施等都是决定外商直接投资区位选择的重要区位因素。Lipsev（2000）对发达国家的外资流向与产业结构作了研究，分析具体产业对外资的吸引力。孙俊（2002）认为除了政策、基础设施、教育、工资等变量外，一个地区的产业结构、开放水平以及市场化程度等对投资环境也有重大影响，从而对该地区吸收外商直接投资的水平起了举足轻重的作用。

Marshall 指出了聚集效应对工业区位选择的重要性以后，一些学者又从聚集的角度，研究聚集对外商直接投资的影响。Kickerbocker（1973）、Dunning（1977）、Krugman（1991）、Dunning（1995）、Leahy 和 Pavelin（2003）等从理论上研究了聚集效应的影响。Luger 和 Shetty（1985）通过对三位数产业（产业分类标准）的研究，证实了聚集经济对外国公司区位选择的重要影响。Head 等（1995）、Kogut 和 Chang（1996）研究表明，日本在美国的直接投资倾向于采用“跟进策略”，同时也重视早期的直接投资。Head 和 Ries（1996）也发现以前投资对外商在我国直接投资区位选择的重要性。Barrell（1999）发现，同特定产业相关的聚集效应对美国在欧洲的直接投资产生显著的影响。Head（1999）根据日本厂商从 1980 到 1992 年对美国直接投资的研究中发现了同特定国家相关的聚集效应的强烈作用。Cheng 和 Kwan（2000）设定动态调整模型，并以此模型为基础，分析了外商直接投资的区位决定因素。Tuan（2003）扩展中心 - 外围理论，研究聚集效应在外商直接投资对我国广东省投资的影响。许罗丹、谭卫红（2003）在 Dunning 区位优势理论的基础上，设定局部调整模型，分析聚集效应的作用，并比较了这些区位优势因素对华资、美资的不同影响。

四、外商直接投资对东道国的影响

关于外商直接投资对东道国影响的研究，主要集中在对经济增长、技术、出口与出口竞争力、劳工标准、政府政策等几个方面，对经济增长的影响体现在解决资金缺口、技术外溢与技术引进上。

Chenery 和 Strout（1966）在《国外援助和经济发展》一文中提出“两缺口模型”理论。该理论认为发展中国家为了保持一定的经济增长速度，必须积累足够的资本，而发展中国家在储蓄和外汇供给方面存在储蓄缺口和外汇缺口，而外商直接投资是弥补这些缺口的有效途径；通过弥补资金缺口促进发展中国家的经济发展。Koizumi 和 Kopecky（1980）构建国外商直接投资际资本长期流动模型，研究外商直接投资对异国经济增长的作用。Romer（1990）构建内生增长模型，着重强调技术扩散对于小国及广大发展中国家经济持续增长的作用。Basant 和 Fikkert（1996）利用印度厂商的数据，估计了 R&D 开支、技术购买、国内国际的 R&D 溢出对综合要素生产率的影响。王志乐（1996）分析了外商直接投资对我国经济正反两方面的影响。郑京平（1997）利用宏观经济模型检验了外商直接投资对中国经济的影响。姚洋（1998）利用工业普查数据，抽取 12 个行业分析“三资”企业

的影响。何洁、许罗丹(1999)考察了外商直接投资对我国内资工业企业的产量增加的作用。沈坤荣、耿强(2001)认为外商直接投资可以通过技术外溢效应,使东道国的技术水平、组织效率不断提高,从而提高国民经济的综合要素生产率。按照新古典经济学的理论,外商投资企业对整体经济的全要素生产率的贡献,主要表现为:向有关企业的技术转移、对同一行业或相关产业的企业技术溢出效应、促进产业结构向符合比较优势原则转化、促进经济体制向符合市场原则转化等(卢狄,2003)。

Markusen等人(1996)、Zhang和Markusen(1999)等从理论的角度阐述了外商直接投资与投资东道国出口的关系。Blomstrom, Graham, Brenton, Narula和Wakelin等利用发达国家的数据对外商直接投资与东道国对外贸易的关系进行了实证研究,这些实证结果认为外商直接投资与东道国的出口竞争力高度相关。Radosevic, Dyker, Guerreri, Camilla等分别研究了东欧转轨经济国家中外商直接投资与东道国对外贸易的关系,也得到了同样的结果。Lady(1992, 1994), Naughton(1996), Branstetter和Feenstra(1999), Chan等人(1999), Zhang(1998, 1999, 2000)对我国引进外资与我国的出口情况进行了分析。国内学者一般认为外商直接投资促进了中国产业结构升级,提升了中国产品的出口竞争力水平(詹晓宁,陈建国,葛顺奇,2002)。江小涓对外商直接投资企业与国内企业的高新技术产品出口份额进行比较,也得出了类似的结论。谢建国(2003)认为中国产品竞争力的变化有较强的路径依赖,FDI虽然不是中国产品出口竞争力的决定因素,但是打破这种路径依赖的重要原因。

对工资的研究,集中在外商投资企业支付的工资与当地企业支付工资水平的比较、“工资外溢”(wage spillovers)和对投资东道国平均工资水平的影响三个方面。

跨国公司在投资东道国树立良好的企业形象与建立良好的公共关系、减少员工的流动性以避免给当地的竞争对手带来技术外溢、信息不对称等,往往会支付较高的工资。Rippy(1931), Brash(1966), Harrison(1981), Hill(1990), Haddad和Harrison(1993), Lipsey(1994), Howenstine和Zeile(1994), Globerman, Ries和Vertinsky(1994), Aitken, Harrison和Lipsey(1996), Manning(1998), Feliciano和Lipsey(1999), Griffith和Simpson(2001), Girma等(2001), Oulton(2001), Driffield和Girma(2002)的研究都得出了类似的结论。跨国公司在支付较高工资的同时,是否会通过“工资外溢效应”对当地企业支付的工资水平造成影响,也引起了很多学者的兴趣。Findlay(1978), Aitken, Harrison和Lipsey(1996), Feliciano和Lipsey(1999), Figlio和Blonigen(2000), Lipsey等(2001), Girma等(2001), Lipsey(2002)对此进行了研究,结果发现有正的外溢效应。Aitken, Harrison和Lipsey(1996), Feenstra和Hanson(1997), Feliciano和Lipsey(1999), Figlio和Blonigen(2000), Blonigen和Slaughter(2001), Lipsey(2002)研究发现,跨国公司通过提高劳动需求,工资溢出对投资东道国的整体工资水平造成影响,投资东道国工资水平有上升的压力与趋势。跨国公司对当地员工的培训,提升了他们的技术水平,也变相地提高了这部分人的工资水平。许罗丹、刘民权(1998)调查发现,美资、欧资企业在我国的投资,为当地员工提供了大量的在职培训。李雪辉、许罗丹(2002)利用我国外资集中地区的宏观数据进行实证分析,验证了通过提高当地的熟练劳动的工资水平,外商直接投资可以提高我国外资集中地区的工资水平。

为了解决经济发展过程中的资金短缺问题，同时也逐渐认识到外资对经济发展的重要性，各国通过向外资提供各种优惠措施来吸引外资。1991~2000年，世界各国对外商直接投资政策有1185项改革，其中的1121项是向有利于吸引跨国公司直接投资的方向改进的(UNCD, 2001)。长久以来，政策优惠是我国吸引外资的重要手段，但是随着在政策优惠方面的竞争，政策对外资的吸引力越来越小。许罗丹、谭卫红(2003)发现，我国对外资的优惠政策对我国吸引外资影响不显著。通过对欧盟在华企业的调查也发现，优惠政策对欧盟投资的吸引力仅排在倒数第二位；欧盟企业更希望我国能提高政府工作效率，改进基础设施建设，实现彻底的国民待遇体制。世界投资报告(2001)提出了以关联为核心的第三代投资政策，促进东道国与外商直接投资关联的政策措施主要包括：提供信息、技术升级、人才培训和资金支持。至此，东道国对跨国公司外商直接投资的促进政策大致经历了三代：第一代以提供激励性优惠措施为主；第二代注重外商直接投资规制框架的自由化变革；第三代强调当地企业与跨国公司的关联(冼国明、葛顺奇，2002)。

广东省出口产业比较优势分析*

黄静波 赖伟娟

改革开放以来,广东的对外贸易得到迅速发展,已成为中国第一外贸大省并连续十几年保持这一地位。国际贸易理论认为,比较优势是对外贸易得以持续发展的基础,因此,有必要清楚地了解广东出口的比较优势是什么,以及广东出口贸易的比较优势所发生的结构性变化。特别是当前广东省正处于关键的产业调整时期,其出口比较优势的变化对产业调整的导向作用重大,因此,基于现阶段对外贸易比较优势的实证分析是广东省制定对外贸易发展战略、协调产业发展、推进改革的重要基础。

不少学者对中国的对外贸易比较优势进行了定量的研究。Yeats (1991)运用世界银行的“贸易分析与回溯系统”(Trade Analysis and Retrieval System),以联合国的贸易数据库为基础计算了中国1965~1987年的显现性比较优势(Revealed Comparative Advantage, RCA)指数,得出的结论是:中国在广泛的劳动密集型制成品方面有比较优势,并与亚洲其他经济形成竞争关系。^①王炳才(1999)通过计算中国各类商品的产业内贸易指数和产业内贸易指数从高比较优势、低比较优势、高比较劣势等几个层次分析了我国对外贸易的产业竞争力情况。^②岳昌君(2000)通过计算认为,总的RCA指数表明,中国杂项制品(SITC8)的出口具有最大的比较优势,其次是按材料分类的制成品(SITC6),这两类商品主要是劳动密集型产品。资本密集型的机械和运输制成品(SITC7)的RCA指数最低,说明我国的工业化程度相对较低,技术水平落后。^③刘重力、刘德江(2003)分别计算了中国农业、劳动密集型产业和资本密集型产业的RCA指数时间序列,并与主要竞争国家或地区作了相应的比较。他们的结论是:中国农产品的比较优势逐年下降;劳动密集型产品的RCA指数各年度均大于2.5,具有极强的比较优势;资本密集型产品还处于比较劣势,但有所改善。^④另外还有一些学者做过类似的计算和分析,本文不再一一列举。

但是就广东省出口比较优势的量化分析文献极少,陈雪梅(2001)计算了1999年广东21大类出口商品产品中出口额占前15位商品的竞争力指数,其中位居广东前5位的出口商品大类国际竞争力指数均大于零。在这5类出口商品中,具有高比较优势的产品有两

* 本文为广东省哲学社会科学规划项目“广东省对外贸易发展的实证研究”(项目批准号:02c33)的部分成果。原载:《国际贸易问题》,2004年第1期。

① Yeats, Alexander J., China's Foreign Trade and Comparative Advantage: Prospects, Problems, and Policy Implications, World Bank Discussion Papers, No. 141, November 1991, Washington DC: The World Bank.

② 王炳才:《中国90年代外贸优势及战略和政策选择》,《世界经济》,1999年第10期。

③ 岳昌君:《我国外贸出口结构变化与比较优势实证分析》,《国际经贸探索》,2000年第3期。

④ 刘重力,刘德江:《中国对外贸易比较优势变化分析》,《南开经济研究》,2003年第2期。

大类：鞋帽伞杖、加工羽毛、人造花、人发制品，杂项制品；具有低比较优势的产品有三大类：机械、电气设备、电视机及音响设备，纺织原料及纺织品，仪器、医疗器械、钟表及乐器。在第6至第15位出口商品中，大多为低比较优势或缺乏比较优势的产品。^①傅江景（2002）运用RCA指数、竞争力指数和出口优势变差指数等指标考察了广东省出口产品的比较优势变化，但分类较为粗略。傅江景认为，广东出口产品的比较优势集中在劳动密集型产品，资本和技术密集型产品具有潜在的比较优势，而资源密集型产品从来不具有出口比较和竞争优势，农产品渐渐失去竞争优势。^②

本文将利用实际数据，以更细一些的分类计算广东省出口产品的RCA指数，进而较深入地来分析广东出口产业的竞争力，在此基础上提出相关的政策建议。

一、出口产业比较优势的衡量

1. 衡量指标

为了考察广东省对外贸易的竞争力情况，本文引入了竞争力指数和显现性比较优势指数，对各产业的比较优势变化进行考察。竞争力指数可以表示为以下公式：

$$C_i = \frac{X_i - M_i}{X_i + M_i} \quad (1)$$

其中： C_i 表示*i*行业的竞争力指数或称净出口指数； X_i 表示*i*行业的出口值； M_i 表示*i*的进口值。一般认为 $C_i > 0$ ，出口大于进口，则表示该行业具有竞争力或比较优势；如果 $C_i < 0$ ，进口大于出口，则表示该行业缺乏竞争力或处于比较劣势。

根据竞争力指数和产业内贸易指数（ $IIT = 1 - [|X_i - M_i| / (X_i + M_i)]$ ，即劳埃德-格鲁贝尔指数）的联系，竞争力指数同时也可以反映产业内贸易状况，即 C_i 越接近于0，产业内贸易越发达。但竞争力指数只是一种静态分析，而现实中存在着各种出口鼓励或出口限制的措施，可能使这种静态分析不能完全准确现实中的贸易优势状况和在整个世界市场上的竞争力状况。因此，在本文中竞争力指数仅用作参照，而关注点在于按照国际通用的方法计算显现性比较优势来反映广东的出口产业竞争力。

本文主要采用Balassa首先提出的显示性比较优势指数，计算公式如下：^③

$$RCA_{ij} = (X_{ij} / X_i) / (W_j / W) \quad (2)$$

式中： RCA_{ij} 代表*i*国（地区）*j*产品的显现性比较优势指数； X_{ij} 代表*i*国（地区）对世界市场出口*j*产品的出口额； X_i 代表*i*国（地区）对世界市场的总出口额， W_j 代表世界市场*j*产品的出口额； W 代表世界市场产品的总出口额。RCA指数考虑了不同国家和不同产品在国际市场上的份额，侧重一国的出口绩效， X_{ij} / X_i 为*j*产品出口占该国出口的比例，*j*产品出口越多，该比例越大，比较优势越明显。按照日本贸易振兴会（JERTO）提出的

① 陈雪梅，蒋江敏：《入世后广货国际竞争力如何》，《新快报》，2001年11月7日。

② 傅江景：《广东出口贸易比较与竞争优势分析》，《学术研究》2002年第4期。

③ Balassa, B., Trade Liberation and Revealed Comparative Advantage, The Manchester School of Economic and Social studies, 1965, Vol. 33, pp92 ~ 123.

标准,当 RCA 数值大于 2.50 时,该产业具有极强比较优势;当 RCA 在 0.80~1.25 之间时,该产业具有中等比较优势;0.8 以下则处于比较劣势。^① 本文考察的是广东省的情况,因此 i 代表广东省。

2. 数据

国内学者研究这一专题时,产品的分类大多采用国际贸易商品标准分类法(SITC)。但我们从《广东统计年鉴》所获得的广东省出口商品分类数据,未采用这一分类法,而是现行的海关协调制度(HS)分类标准。考虑到 HS 分类较之 SITC 更为详细些,因此我们对地采用了 HS 分类标准的有关国际数据来计算广东省出口产业的 RCA 指数,这部分数据来源于联合国统计网站。在 HS 的类章总值表中,全部商品分为 21 类 97 章,这种分类可利用的另一个特点是:其分类较之 SITC 更符合产品的加工特性,各类以及同一类的各章基本上按加工程度从低到高排序。因而通过考察某些类当中各章的比较优势和产业内贸易情况,恰可反映广东省的加工贸易情况。

二、广东省出口比较优势的基本格局

广东省自改革开放以来对外贸易迅速增长,贸易结构也发生了根本性的变化。我们计算出主要年份 20 类产品的 RCA 指数如表 1 所示,从中我们可以看出具体行业在国际市场上的竞争力。

从表 1 可见,广东省在第 8、11、12、18 和 20 类的产品具有较强的比较优势(其中第 8、12、20 类的 RCA 指数达到 2.5 以上),这几类产品主要是服装、鞋帽、玩具、家居日用用品等劳动密集型的轻工业产品。广东省的农林业、矿产业、化工行业和运输设备等重工业处于明显的比较劣势。特别是农林业和矿业,第 1、2、4、5 类在 90 年代初期之前曾经有过较强或很强的显现性比较优势,例如 1988 年这四类产品的 RCA 指数分别为 1.61、6.58、8.44 和 1.50,但 1995 年后开始处于比较劣势,RCA 指数均低于 0.8,并且在 90 年代后半期,除第 4 类之外的其它三类产品的 RCA 指数更是大幅度下降,成为了强比较劣势产品。这些比较优势的变化在本文后面的分析中还将得到进一步佐证。

对照全国相关产业的 RCA 指数,可以看出,广东省的出口竞争力状况与全国的情况大体相同,但广东省的服装、轻工业产品的比较优势更为明显,而农产品、矿产品和主要重工业产品的比较劣势也更加突出。传统的国际贸易理论认为,一国的资源禀赋、传统产业的专业化生产所形成的劳动生产率差异决定了其贸易结构,因此,自然资源的差异、产业结构的不同会导致贸易结构的不同。广东省农业耕地面积小,农业人口相对较少,矿产资源较为贫乏,农产品和矿产品不能自足,更不能满足工业发展的需要,需要大量地进口谷物、原油等战略物质。由于缺乏重要的工业战略资源和某些历史原因,在全国工业布局上,广东省也少有大型重化工业的企业分布,所以在出口方面,广东省的化工行业、金属工业和运输工具等重工业的比较优势也比全国弱。

^① 刘重力,刘德江:《中国对外贸易比较优势变化分析》,《南开经济研究》,2003 年第 2 期。

表 1 广东省出口产品显示性比较优势 (RCA) 指数

类别	商品范围	广东省							中国		
		2001	2000	1999	1998	1995	1991	1988	2001	1999	1995
1	活动物, 动物产品	0.27	0.32	0.31	0.37	0.59	0.84	1.61	0.84	0.90	1.09
2	植物产品	0.21	0.20	0.21	0.23	0.44	1.05	6.58	0.83	0.90	0.88
3	动植物油脂及蜡	0.22	0.21	0.20	0.37	0.78	/	/	0.15	0.16	0.48
4	食品, 烟草及制品	0.45	0.49	0.42	0.44	0.45	2.72	8.44	0.75	0.72	0.78
5	矿产品	0.10	0.10	0.15	0.28	0.20	0.20	1.50	0.35	0.37	0.67
6	化工产品	0.21	0.24	0.23	0.24	0.26	0.42	0.43	0.57	0.63	0.66
7	塑料, 橡胶及制品	0.92	0.96	0.96	1.03	0.83	0.11	0.08	0.77	0.78	0.59
8	皮革, 毛皮及制品, 旅行用品, 手提包	3.55	3.91	4.30	4.25	4.58	/	/	3.60	3.81	3.63
9	木及木制品, 草柳编结品	0.60	0.64	0.56	0.55	0.68	/	/	0.93	0.83	0.97
10	木浆, 纸, 纸板及制品	0.48	0.47	0.45	0.48	0.30	/	/	0.31	0.26	0.21
11	纺织原料及纺织制品	2.01	2.39	2.57	2.71	2.55	4.15	3.35	3.21	3.47	3.37
12	鞋帽伞杖, 加工羽毛, 人造花, 人发制品	5.69	6.68	6.68	6.79	6.63	/	/	4.64	5.33	4.68
13	石材制品, 陶瓷产品, 玻璃及其制品	1.19	1.44	1.41	1.36	1.19	1.32	0.77	1.46	1.54	1.24
14	珠宝首饰, 硬币	0.97	0.86	1.01	0.76	0.82	/	/	0.54	0.64	0.59
15	贱金属及制品	0.75	0.74	0.70	0.62	0.50	0.49	0.82	0.97	1.00	1.00
16	机械, 电气设备, 电视机及音响设备	1.48	1.29	1.19	1.10	0.92	0.08	0.06	1.09	0.90	0.60
17	车辆, 航空器, 船舶及有关运输设备	0.19	0.24	0.21	0.22	0.77	0.18	0.27	0.31	0.29	0.94
18	仪器, 医疗器械, 钟表及乐器	1.38	1.47	1.52	1.61	1.56	0.84	0.21	0.90	1.03	0.84
20	杂项制品	4.12	4.55	4.40	4.39	4.12	/	/	2.79	2.91	2.72

注：“/”表示该项未能获得有效的数据而不能计算出相应的结果。

资料来源：广东和中国的数据来源于《广东统计年鉴》和《中国对外贸易统计年鉴》各年版；国际的数据来源于联合国统计网站 www.unstats.un.org。

另一方面，改革开放以来，随着引进外资的不断增加，以“三来一补”为特征的加工贸易行业率先利用广东省的沿海便利地理优势发展起来，充分利用了我国丰富的劳动力资源，逐步形成了世界上颇具规模的服装、杂项制品生产基地，该类产业的比较优势较全国更加明显。近年来，加工贸易出现技术水平提升的趋势，新兴的电子、家电、计算机通信等行业获得迅速发展，成为广东工业发展和对外出口新的重要力量，这些行业产品的出口比较优势逐渐增强。

总体上看，广东省资源状况和出口比较优势的状况决定了贸易结构的变化：工业制成品的出口比重不断上升，由 1987 年的 85% 上升到 2001 年的 96%，初级产品则从 15.2% 下降到 3.7%。相应地，初级产品的国际竞争力相对下降，至 2001 年，已处于比较劣势。按

照竞争力指数来衡量,广东省工业制成品的竞争力指数由 1987 年的 -0.08 上升到 2001 年的 0.13 (2000 年为 0.18)。具体如图 1 所示。

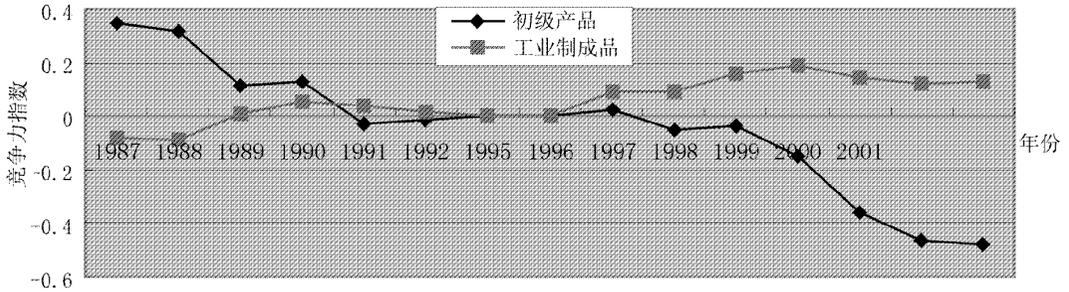


图 1 广东省初级产品和工业制成品的竞争力变化

资料来源:根据历年《广东统计年鉴》数据计算绘制而成

工业制成品出口的上升使得广东省的出口结构逐步优化,近年来,机电产品和高新技术产品出口保持较高的增长,到 2001 年,它们在广东省总出口中所占的比重已达 58.04% 和 23.36%。其中,机电产品出口 553.8 亿美元,增长 10.8%;高新技术产品(其中大部分属机电产品)222.9 亿美元,增长 30.9%,其增长主要是由计算机与通讯技术产品带动。2002 年机电产品出口 726.57 亿美元,增长 31.2%;高新技术产品出口 309.66 亿美元,增长 39.0%。正是由于这两类产品的快速增长所推动,使得广东省在传统商品出口出现一定降幅时,总出口仍能保持强劲的增长。同时,较合理的出口结构也可以使对外贸易在出现外部冲击或经济危机时免受过大损害,1997~1998 东南亚金融危机时广东省的出口能够继续保持增长就是一个很好的证明。

三、广东省出口行业的比较优势变化

但是,按照 HS 的分类标准,在同一类产品中,包含了各种加工程度不同的各章产品,在某类强(弱)比较优势产品中,可能包含了弱(强)比较优势的章,大类的 RCA 指数只能说明大体情况。为了更为细致地考察广东省对外贸易的特点,我们按照 HS 分类在两位数(即 98 章)层次上计算了 RCA 指数,并据 2001 年的指数值按很强比较优势、强比较优势、中等比较优势、比较劣势将主要产品分成四部分,分别列于表 2、表 3、表 4 和表 5。四个表中也计算了 2001 年度竞争力指数,以分析进口、出口和产业内贸易状况。

表 2 广东省具有很强比较优势产品的 RCA 指数

类 - 章别	商品范围	2001	2000	1999	1998	1995	2001 竞争力指数
20 - 94	家具、床上用品、照明装置、发光标志	2.79	2.99	2.85	2.49	2.10	0.97
20 - 95	玩具、游戏、运动用品及零附件	6.41	7.28	7.01	7.20	7.63	0.96
18 - 91	钟表及零件	4.80	5.71	6.15	6.25	5.98	0.36
17 - 86	铁道及电车机车、车辆及零件	4.31	5.73	4.60	5.12	3.06	1.00
13 - 69	陶瓷产品	2.50	3.40	3.40	3.16	2.25	0.90
12 - 64	鞋类及零件	5.45	6.44	6.36	6.39	6.14	0.97
12 - 66	伞、杖、鞭及零件	13.38	15.23	15.54	16.51	14.34	0.96
11 - 60	针织物及钩编织物	5.31	5.09	4.24	3.87	4.72	0.22
11 - 61	针织或钩编的服装及衣着附件	2.51	3.30	3.99	4.51	3.46	0.79
11 - 50	蚕丝	3.08	3.51	3.73	3.06	2.95	0.24
9 - 46	草柳编结品	5.72	8.03	6.69	7.62	14.43	0.92
8 - 42	皮革制品、旅行用品及手提包	5.80	6.42	6.97	7.33	8.65	0.97
6 - 36	炸药、烟火制品、易燃材料制品	3.31	3.50	5.46	3.97	4.82	0.99

资料来源：同表 1。

表 3 广东省具有强比较优势产品的 RCA 指数

类 - 章别	商 品	2001	2000	1999	1998	1995	2001 竞争力指数
18 - 92	乐器及零附件	2.08	2.08	1.92	1.74	1.25	0.81
16 - 84	核反应堆、锅炉、机械设备及零件	1.27	0.99	0.82	0.74	0.46	0.21
16 - 85	机电、电气设备、电视机及音响设备	1.70	1.57	1.58	1.50	1.43	0.05
15 - 73	钢铁制品	1.39	1.49	1.27	1.08	0.77	0.56
15 - 82	贱金属工具器具、利口器、餐具及零件	1.87	1.96	1.57	1.41	1.33	0.70
15 - 83	贱金属杂项制品	1.96	2.01	1.95	1.92	1.76	0.74
4 - 16	动物产品制品	1.41	1.64	1.16	1.36	1.42	0.98

资料来源：同表 1。

表 4 广东省具有中等比较优势产品的 RCA 指数

类 - 章别	商 品	2001	2000	1999	1998	1995	2001 竞争力指数
18 - 90	光学、照相电影、计量检验、医疗仪器设备	1.03	1.02	0.99	1.03	0.83	0.10
10 - 49	书籍、印刷品、设计图纸	1.08	1.08	0.95	0.83	0.40	0.65
7 - 39	塑料及其制品	1.12	1.17	1.18	1.30	1.04	- 0.41

资料来源：同表 1。

表 5 广东省具有比较劣势（极弱比较优势）产品的 RCA 指数

类 - 章别	商 品	2001	2000	1999	1998	1995	2001 竞争力指数
1 - 4	乳品、蛋品、天然蜂蜜、其他食用动物产品	0.08	0.10	0.09	0.12	0.06	- 0.19
2 - 10	谷物	0.00	0.01	0.01	0.01	0.01	- 0.96
4 - 18	可可及可可制品	0.02	0.02	0.01	0.02	0.01	- 0.70
5 - 26	矿砂、矿渣及矿灰	0.01	0.01	0.01	0.02	0.10	- 0.97
5 - 27	矿物燃料、矿物油及产品	0.10	0.10	0.15	0.29	0.19	- 0.70
6 - 29	有机化学品	0.09	0.10	0.10	0.11	0.14	- 0.76
6 - 30	药品	0.04	0.06	0.06	0.07	0.10	- 0.60
7 - 41	生皮及皮革	0.73	0.54	0.51	0.49	0.57	- 0.65
9 - 45	软木及软木制品	0.02	0.02	0.02	0.02	0.07	- 0.70
10 - 47	木浆及其他纤维素浆、废碎纸板	0.00	0.01	0.00	0.00	0.00	- 1.00
11 - 54	化学纤维长丝	0.31	0.34	0.36	0.46	0.35	- 0.81
11 - 55	化学纤维短丝	0.76	0.74	0.78	0.86	0.65	- 0.56
15 - 72	钢铁	0.07	0.09	0.08	0.10	0.10	- 0.94
17 - 87	车辆及零附件	0.14	0.18	0.15	0.14	0.12	0.21
17 - 88	航空器、航天器及零件	0.00	0.00	0.01	0.00	0.01	- 0.95

资料来源：同表 1。

从以上四个表中我们可以看到，广东省的出口结构呈现以下几个特点：

第一，广东省具有很强比较优势的产品种类基本上是技术含量不太高、劳动密集型的制成品，以日常用品为主，另有部分机械仪器零件（86 章、91 章），这与广东省的产业结构历来以轻工业为主是相对应的。再加上改革开放之后，广东省抓住了世界产业结构调整的机会，利用劳动力成本优势承接国际劳动密集型产品产业转移，迅速培育出一批以家电、日用消费品、旅游及运动用品为主的轻工出口行业。

第二，广东省的进出口贸易和加工贸易有密切联系。从以上 RCA 指数和竞争指数，

我们可以看到：在广东省的弱比较优势产品中，既包含了因为自然条件限制而需要进口的农产品、矿产品，也包含了大量加工贸易所需要的原料。这些原料或者用于制造同处一类的那些强比较优势制成品，或者是其他类制成品的重要原料。广东省大量进口的乙烯（29章）、皮革（41章）、纸浆（47章）、化学纤维（54、55章）、钢材（72章）分别是化工、皮革制品、造纸、纺织、钢铁等制造业的重要原材料，而这些行业加工制成的服装、鞋帽类、家具、塑料制品、贱金属产品、简单器具等更是广东省主要的出口创汇产品。以外商投资企业的进口为例，2001年，该类企业服装的出口是365470万美元，而用于制造服装而进口的羊毛、原棉、合成纤维、纱线、布匹等原料就达160757万美元，占44%。此外，加工贸易也是高新技术产品出口的主要方式，高新技术出口中加工贸易所占比重是94.8%。

第三，虽然服装、皮革制品、杂项等是广东省的强比较优势产品，出口总额占到了广东省出口额的30%以上，但是，这些劳动密集型产品的RCA指数已逐年下降。这里隐含着一些问题：为什么劳动密集型产业的比较优势会下降？RCA指数是否还会下降很多？广东省会丧失这部分产业的比较优势吗？

劳动密集型产业比较优势下降的原因之一是受国际贸易形势的影响。劳动密集型产品在国际市场上供大于求，价格趋于下降，随着东南亚经济的复苏，该产品竞争激烈。近年来，以美国、欧盟为代表的发达国家对发展中国家的服装、鞋类、陶瓷等杂项制品制定了配额、严格的技术标准等非关税壁垒，同时还对部分产品提出反倾销的诉讼，并采取相应的反倾销措施。这些贸易保护政策成了广东省劳动密集型产业出口的新障碍，严重削弱了广东省这些产品的比较优势。另一方面，广东省的产业政策调整也影响到出口产品比较优势格局的变化。为了改变加工贸易低附加值的状况，促进广东省的产业结构升级，广东省的产业政策实行了从粗放型向集约型的转变，劳动密集型产业不再作为发展的重点。但不少学者认为，劳动密集型产业是我国原始资本积累的重要基础，对于一个人口大国，劳动密集型产业在解决就业问题上发挥着重大的作用，而且我国的工业技术水平也不可能在短期内得到实质性的提高，在这种状况下，低成本的劳动力优势不可能动摇。因此，广东省劳动密集型产业的比较优势还将继续维持。

第四，广东省已在部分机械设备、钟表制品、电车机车、电子技术、计算机与通信技术、电器行业和钢铁加工业等行业具有一定的比较优势，虽然优势还不太强，但在逐步上升，表现出强劲的贸易潜力。集装箱、有线电话、彩电、电子手表、电路元件已成为主要出口商品，广东省的深圳、东莞市已成为世界电脑外部设备的重要生产基地，生产的电脑资讯产品在全球市场占有了相当的份额（电脑磁头、电脑机箱及半成品占40%，电脑驱动器占20%，高级交流电容器、行输出变压器占25%，电脑扫描仪、微型马达占20%）。这种外贸结构既是全球产业链分工、产品生命周期变化的结果，同时也归因于广东省的政策环境和外资引进为推进这种产业分工创造的有利条件。广东省有针对性地地为计算机通信、家用电器等资本密集度、知识含量较高的行业提供政策支持。外国直接投资大量进入广东也为这些行业的发展带来了资金、管理经验和先进技术。

但以上的实证结果也让我们看到了不足：广东省在机械设备、电气设备、电视机、钟表、仪器等产品所在的表3中的84章、85章和表4中的90章竞争力指数接近于0，即产

业内贸易非常大，这是因为在这些行业中广东省进口了大量含有关键技术的电子部件，如数据处理设备、半导体器件、显像管、集成电路（如电脑芯片）、电子件等，这些产品是我国目前技术水平难以大规模生产，甚至是无法制造的。而在出口的机电产品和高技术产品中，主要是电话机、金属餐具、金属工具、电子手表、小家电、电脑部件和电子元件等技术含量不高的产品。例如，广东电子工业的优势并不体现在技术上，而只是体现在最终产品总量上，它是港澳台电子工业加工装配环节转移的结果，产品的设计、材料和零部件的生产及产品销售均过分依赖于香港和海外。因此，广东省在机电产品和高技术产品上的出口优势并不能表明广东省的技术水平有了很大提高。

四、结论和建议

在对广东对外贸易比较优势进行实证分析的基础上，我们看到了广东对外贸易的比较优势格局：处于比较劣势的农林矿产等资源性行业、原材料产业、化工行业和一些重工业竞争力并没有进一步提高，其中一些还继续下降；广东对外贸易的持续发展主要是靠劳动密集型产业和比较优势日益上升的高新技术行业两股力量来推动的。

国际贸易利益的源泉，一是基于低成本比较优势，二是源于核心竞争力和品牌效应的竞争优势。相对而言，成本优势通常在成熟产品的竞争中显得更重要，但这时产品的附加价值已不高。而竞争优势在新产品、高新技术和高附加值的产品竞争中更为关键，拥有和不断创造竞争优势是对外贸易从而整体经济持续增长的重要方略。这两种优势的不同作用是贸易结构调整中必须予以充分注意的。

毫无疑问，我们应当充分发挥和利用当前的比较优势，获取最大限度的贸易利益。不过同时也要看到，广东引进外资开展加工贸易的规模可观，内资企业也多利用引进技术设备进行生产和出口，但存在企业重在引进而对如何消化、吸收、创新重视不够的缺憾，而以劳动密集型产业为主的出口结构又存在附加值较低、在国际市场上收入弹性小、面临激烈竞争的问题。所以广东的出口产业迫切需要实现从引进模仿向以自主创新为主的增长方式转变，从劳动密集、资金密集型向技术密集型为主转变。

首先，以“内地与香港更紧密经贸关系”的实施为契机，加快广东经济的对外开放，按照全面与国际市场接轨的要求完善广东的市场竞争环境。这意味着市场机制作用更大，市场竞争更加激烈，广东的许多企业可能有一个艰难时期，但竞争是技术进步的基本动力，是提升出口产业竞争力的有效途径。具体地说，香港先于内地的产业结构升级的优势和经验，有利于推动广东的贸易结构转换。

其次，重视与跨国公司的经济联系，吸引跨国公司到广东投资。应当注意到，通过技术贸易直接引进先进技术的策略往往会受到西方发达国际技术出口限制、本国技术配套能力和吸收能力等制约，而通过跨国公司投资带进技术可减少上述忧虑，跨国公司投资的技术外溢可以推动本地技术进步，带动相关行业的发展。

再次，针对现有的出口结构，一方面应在原有比较优势的劳动密集型产品上注入更多的知识含量，投入更多的人力资本，提高产品质量，以增强产品在世界市场的竞争力；另一方面努力发展具有潜在比较优势并有利于优化产业结构的产品，以获得更多的动态贸易

利益。尤其是抓住广东省当前高新技术行业 and 知识 - 资本密集型业发展势头良好的机遇，既要强化技术引进和消化能力，又要加大研究与开发力度，进行自主创新，让该行业产品真正成为高新技术产品。

最后，在信息服务、研究开发投入、体制保障、出口支持和保障等方面，政府政策可发挥重要作用来促进出口结构的转型。

参考文献

- [1] 广东省统计局：《广东统计年鉴》，广东统计年鉴出版社，各年。
- [2] 王炳生：《中国 90 年代外贸优势及战略和政策选择》，《世界经济》，1999 年第 10 期。
- [3] 刘重力，刘德江：《中国对外贸易比较优势变化分析》，《南开经济研究》，2003 年第 2 期。
- [4] 傅江景：《广东出口贸易比较与竞争优势分析》，《学术研究》，2002 年第 4 期。
- [5] 吴建伟：《国际贸易比较优势的定量分析》，上海三联书店，上海人民出版社 1997 年版。
- [6] 李荣林，张岩贵：《我国对外贸易与经济增长转型的理论与实证研究》，中国经济出版社 2001 年版。
- [7] Ballance, Robert H., Trade Performance as an Indicator of Comparative Advantage, In David Greenaway (ed), Economic Development and International Trade, Basingstoke: Macmillan, 1988, pp6 ~ 24.

从啤酒市场整合程度看 WTO 对消除地方保护的影响*

李 杰 孙群燕

一、引言

关于贸易保护，中外学者主要研究了不同国家间的贸易壁垒问题，但对于一个国家内不同地区间的贸易保护，学者们没有给予足够的关注。中国地区间的贸易保护是一个普遍的现象，而啤酒行业则是其中表现突出的一个行业。例如，一瓶在北京只卖 18 美分的燕京啤酒，在四川的售价竟高达 1 美元（Gilley, 2001）。其他诸如汽车、烟草、服装等产品的跨地区销售，也在不同程度上面临着地方保护的重重障碍。

目前，对中国地区间贸易保护问题的研究，比较有代表性的是 Young（2000）、银温泉和才婉茹（2001）以及陈抗等（2002），前两篇文章提供了大量关于中国地区间贸易保护的事实，后一篇文章则从中央与地方之间分配财政资源的不同方式入手，考察了财政集权程度的变化对地方政府保护行为的影响。以上文章都认为，行政性分权是造成地区间高贸易壁垒的重要原因。Li 和 Zhang（2003）在完全竞争的模型框架内，建立了一个中国地区间贸易保护的经济模型，分析中国在加入 WTO 后，各地区贸易保护壁垒对地区经济福利，进而对整个国家福利状况的影响。文章认为，加入 WTO 并对外实行自由贸易，有助于消除地区间的贸易保护壁垒。Li 等（2003）通过构建一个寡头垄断竞争模型，从理论上分析了地方政府设立地区贸易保护壁垒的激励与中国税制改革以及贸易自由化进程的关系。文章认为，低的中央税和高的对外贸易保护，是促成地区间形成贸易保护壁垒的重要原因。但到目前为止，从实证的角度来研究中国地区间的保护问题的文章尚不多见，而且涉及这一问题的文章，通常只是把地方保护作为其中的一个影响因素，如 Bai 等（2002）主要研究地区间分工专业化与地方保护之间的关系，其分析的重点是分工专业化而非地方保护行为。Ha 和 Fan（2003）用 ADF 方法研究了香港与大陆四个城市的价格序列的收敛趋势，其基本结论是，这些价格序列之间具有一致收敛的趋势。但该文的研究重点在于大陆四个城市价格收敛的一致性如何影响香港的价格变动趋势。据笔者所知，到目前为止还没有一篇文章从实证的角度考察中国加入 WTO 对地方政府保护行为的影响，本文试图沿着这个方向作一些尝试性的研究工作。

* 本文的研究得到国家自然科学基金之金融数学重大项目（项目号：10131030）以及中山大学四川优秀青年基金之博士生重要创新项目提供的资助，特此致谢。在本文的写作过程中，王美今教授提供了许多有价值的参考意见，在此表示感谢。原载：《世界经济》，2004 年第 6 期。

鉴于中国各行业都存在不同程度的地方保护行为，而且受到数据资料方面的限制，要对所有行业作一个全面的分析是不可行的，因此，本文仅仅以啤酒行业作为一个切入点，采用 1997~2003 年全国 40 个主要的大中城市 640 ml 当地主销啤酒价格的月度数据，着重分析以下两方面的问题：中国地区间是否存在贸易保护壁垒？加入 WTO 是否有助于消除地区间的贸易壁垒？

本文的基本结论是：从 1997 年 1 月到 2003 年 3 月，我国啤酒市场的整合程度不高，市场分割的现象比较普遍。从分时期的结果看，在中央政府出台取消地方保护壁垒的措施之前（无论有没有考虑政策出台的事前效应），我国啤酒市场的市场分割现象比较普遍；但在这一政策措施出台以后，啤酒市场却出现了明显的整合趋势。同时，以北京和青岛为中心，我国的啤酒市场出现了区域性的整合趋势。但是，北京与青岛这两个最大的啤酒生产地的啤酒价格序列之间，无论是从整段时期看，还是分时期看，都不存在协整关系。

文章的结构安排如下：在第二部分，我们首先回顾我国啤酒行业的发展状况及所存在的地区保护行为；在第三部分，我们提出研究的理论框架与研究方法；在第四部分，我们具体考察我国最大的两个啤酒产地与其他城市在啤酒价格方面的协整关系，据此分析中国地区间在啤酒行业方面的贸易保护程度及变化趋势；在第五部分，我们对基本结论提供一个合理的经济学解释。

二、地方保护与我国啤酒行业的发展状况

中国地区间的商品流动存在壁垒限制是一个众所周知的事实，这在很大程度上源于长期以来形成的地方保护主义，其表现形式主要是当地政府直接在当地投资设厂，并通过使用贸易壁垒，阻碍外地生产同类产品的企业进入本地市场。这种政府行为在啤酒行业表现得尤为突出。在新中国建立后，国家对啤酒行业采取了扶持政策，新建了一批啤酒厂，使啤酒业取得了一定的发展，但由于受当时购买能力和消费习惯的限制，啤酒的消费面仍比较小，全国多数地区啤酒生产均是一片空白。改革开放后，国家出台了一些政策，鼓励啤酒业快速发展。在这样的形势下，地方政府纷纷投资建立自己的啤酒厂，啤酒企业的数量与产量迅速扩张。到 1988 年，全国已有啤酒企业 813 家，产量 656.4 万吨，仅次于美国和德国。但企业规模非常小，大多为年产 1 万~2 万吨，并且在地域上高度分散。

从 1990 年代初期开始，啤酒产量以年均 20% 的速度增长，到 1997 年，全国啤酒总产量达到了 1888.5 万吨。在产业竞争格局上，各地政府纷纷扶持自己的啤酒品牌。在这一阶段，全国啤酒企业基本分化为三个层次：以青岛、燕京和珠江三大集团为第一层次，地方名牌啤酒企业为第二层次，如沈阳华润雪花啤酒、重庆啤酒、河南金星啤酒、哈尔滨啤酒、钱江啤酒、圣泉啤酒等，其他市县啤酒企业为第三层次。值得注意的是，啤酒行业的产业集中度仍然偏低，到 1997 年底，在全国 500 多家啤酒企业中，没有一家年产量超过 100 万吨，即使是产量最大的燕京和青岛啤酒，其产量总和也只占全国产量的 5% 左右。全国的啤酒市场仍处于群雄逐鹿的市场分割状态。导致这种局面出现的一个重要原因，在于各地政府保护本地企业利益的动机。这种动机的产生源于啤酒企业在地方财政收入中的重要地位。尤其在几次重大的税制改革之后，地方政府支配自身财政收入的权限大大增

加,而啤酒企业历来是地方的利税大户,为了维护自身的利益,地方政府都会千方百计地保护当地的啤酒厂。它们通过设置各种有形或无形的进入障碍,如设置质量检测标准和各种税费,提高外地啤酒企业的销售成本,甚至通过明令禁止外地啤酒在本地的销售等手段,把外地品牌阻挡在当地市场之外。我们在 Li 等(2003)中对地方政府保护当地企业的动机有更具体详细的论述。

出于加入 WTO 的需要,从 1995 年开始,中央政府多次表示要取消地区间的贸易保护壁垒。2001 年 4 月 21 日,国务院发布了《关于禁止在市场经济活动中实行地区封锁的规定》,明确规定“市场交易中的地区封锁”是违法行为。这一措施意在规范市场竞争,打击行政垄断,根除地方保护,为建立市场经济新秩序保驾护航。与此同时,由于我国啤酒消费进入年均 5% 的低增长状态,啤酒产能过剩,竞争激烈,行业盈利能力下降,出现大面积亏损。在这种形势下,为增强自身的竞争实力,青岛和燕京啤酒集团开始在全国进行大规模的资本圈地运动,拉开了战略竞争和行业大整合的进程。它们通过并购地方名牌企业,使产销区域不断扩张。到目前为止,青岛啤酒的产销区域已经向西延伸到陕西,向南延伸到广东,并打入香港市场,形成了很高的知名度;燕京啤酒也不甘示弱,先后进入湖南、湖北与江西,两者同时发展成为中国最大的啤酒生产厂家。近年来,两大啤酒集团的产能规模位次交错变化,燕京啤酒 1995~1998 年产量和市场占有率均居全国第一,1999 以后,青岛啤酒上升为第一,燕京退居第二。到 2002 年底,青岛啤酒在全国啤酒市场所占有的份额已经达到 11%,而燕京则为 10% 左右。总体上看,这两大啤酒集团在国内的啤酒市场上基本处于寡占的地位。

三、研究的理论框架、方法和数据说明

以下我们将从实证的角度研究地区间的贸易壁垒问题。地区间存在贸易壁垒的主要表现形式是市场分割。所谓市场分割,是指一个市场商品价格的变化,不能及时有效地传递到另一个市场,市场间的价格信息不存在相互或共同决定的趋势。市场分割是与市场整合相对的一个概念。市场分割往往会扭曲经济主体人的市场决策,导致商品的低效率流通,损害消费者的利益。

一般来说,从实证的角度研究市场分割问题的方法有很多,如相关分析法、Ravallion 模型法(The Ravallion Procedure)、协整分析法(The Co-integration Approach)及 Parity Bounds 模型法(The Parity Bounds Model)。鉴于本文所搜集的数据资料主要为城市的月度价格数据,同时缺乏城市间的运输成本资料,在这种情况下,我们选择采用协整分析方法是合适的^①。

运用协整分析方法研究市场分割问题,首先必须检验单个市场商品价格的时间序列是否平稳。检验单个时间序列的平稳性,通常采取 Augmented Dickey-Fuller (1979) 方法(ADF 法)。ADF 法是依据对以下模型中的统计值的显著性来判定:

^① 关于市场整合分析方法的比较,可参阅周章跃、万广华(1999)。

$$\Delta P_t = r_0 + r_1 P_{t-1} + \sum_{k=1}^N \delta_k \Delta P_{t-k} + \epsilon_t \quad (1)$$

其中： $\Delta P_t = P_t - P_{t-1}$ 表示当期价格与上一期价格的差； ϵ_t 为误差项； N 表示滞后期数。在方程(1)中，零假设 H_0 为序列 $\{P_t\}$ 具有一阶单整，即经过一阶差分后变平稳，用数学方法表示为 $P_t \sim I(1)$ 。如果 r_1 为负数并且在统计上显著，则拒绝零假设，这意味着序列为零阶单整，即 $P_t \sim I(0)$ 。如果 r_1 在统计上不显著，则不能拒绝零假设，这说明序列 $\{P_t\}$ 可能在更高阶上实现整合，我们应继续对以下方程(2)进行回归，作二阶平稳的单位根检验：

$$\Delta^2 P_t = \varnothing_0 + \varnothing_1 P_{t-1} + \sum_{k=1}^N \theta_k \Delta^2 P_{t-k} + \eta_t \quad (2)$$

与方程(1)相比，方程(2)用 ΔP_t 代替了方程(1)中的 P_t 。如果仍无法拒绝 $\varnothing_0 = 0$ 的零假设，我们还需进一步用 $\Delta^2 P_t$ 代替 ΔP_t ，对序列 $\{P_t\}$ 进行三阶平稳检验，并继续重复上述过程，直到能拒绝相应的零假设，得到一个平稳的序列为止。

在对单个价格序列进行了平稳性检验后，我们就可以进一步考察不同市场的商品价格之间是否具有协同变化的趋势。价格序列之间存在协同关系的一个基本前提，是价格序列必须同阶单整。如果不存在同阶单整性，则谈不上讨论序列间的协整关系问题。在两个市场的价格序列存在同阶单整的前提下，我们就可以分析它们之间的协整关系。

经济学家通常采用E-G两步法或Johansen方法来测定两个市场的价格序列之间的协整关系。本文采用Johansen方法来进行测定。Johansen方法的基本原理是：采用最大似然法估计包含有关变量一阶差分滞后项和水平量滞后项的向量自回归(VAR)模型。依据Johansen(1988)，Johansen和Juselius(1990)的方法，我们可以利用特征根最大值统计量 $-T \ln(1-\lambda)$ 即 λ_{\max} 和迹统计量 $-T \sum \ln(1-\lambda)$ (其中 T 为样本数， λ 为对应于不同秩数的特征根)来判断是否存在协整关系。

采用Johansen方法进行协整分析时，会涉及确定解释变量的最大滞后期数问题。我们的处理方法是：通过估计相应的VAR系统，并根据赤池(Akaike)信息准则(AIC)或者施瓦茨(Schwartz)准则(SC)，确定适当的最大滞后期数。由于我们采用的是啤酒价格的月度数据，如果不存在区域性的贸易壁垒，价格信息的传递是比较迅速的，因此我们所选取的最长滞后期不超过3期。

如果两个不同市场的价格序列之间存在协整关系，则这种关系可表述如下：

$$P_{it} = \alpha_0 + \sum_{k=1}^N \alpha_k P_{j,t-k+1} + \mu_t \quad (3)$$

其中： P_{it} ， P_{jt} 分别表示市场 i ， j 在时期 t 的价格； μ_t 表示误差项； N 表示滞后期数， α_0 表示除另一个市场的商品价格以外的所有影响价格变化的因素，如运输成本、加工成本、销售收入、政策因素等。

四、我国最大的两个啤酒产地与其他城市在啤酒价格方面的协整关系

下面我们主要研究两个问题：全国最大的两个啤酒生产地，即北京和青岛，与其他主

要城市的啤酒市场之间是否存在市场分割？如果存在的话，中国加入世贸组织是否有助于消除这种地区性的市场分割？我们首先对这些数据的总体变动趋势作一个粗略的描述，然后运用第三节所讨论的分析方法，构建经济计量模型，并对估计结果进行经济学解释。本文的数据分析结果主要通过使用计量分析软件 E-views 3.1 计算得到。

1. 数据来源及基本分析

本文采用 1997 年 1 月至 2003 年 3 月全国 40 个主要的大中城市 640 ml 当地主销啤酒价格的月度数据，数据来源于中国价格信息中心提供的工业消费品零售价格数据库。

由于我们使用的是月度数据，很有可能存在季节性波动，因此必须首先对其进行判别，在此基础上进一步考察数据波动的基本趋势。

图 1 显示了全国 40 个主要的大中城市 640 ml 当地主销啤酒价格随时间的变动趋势，图 2 显示了北京与青岛这两个城市 640 ml 当地主销啤酒价格随时间的变动趋势，表 1 则给出各个城市啤酒价格数据的描述性统计量。

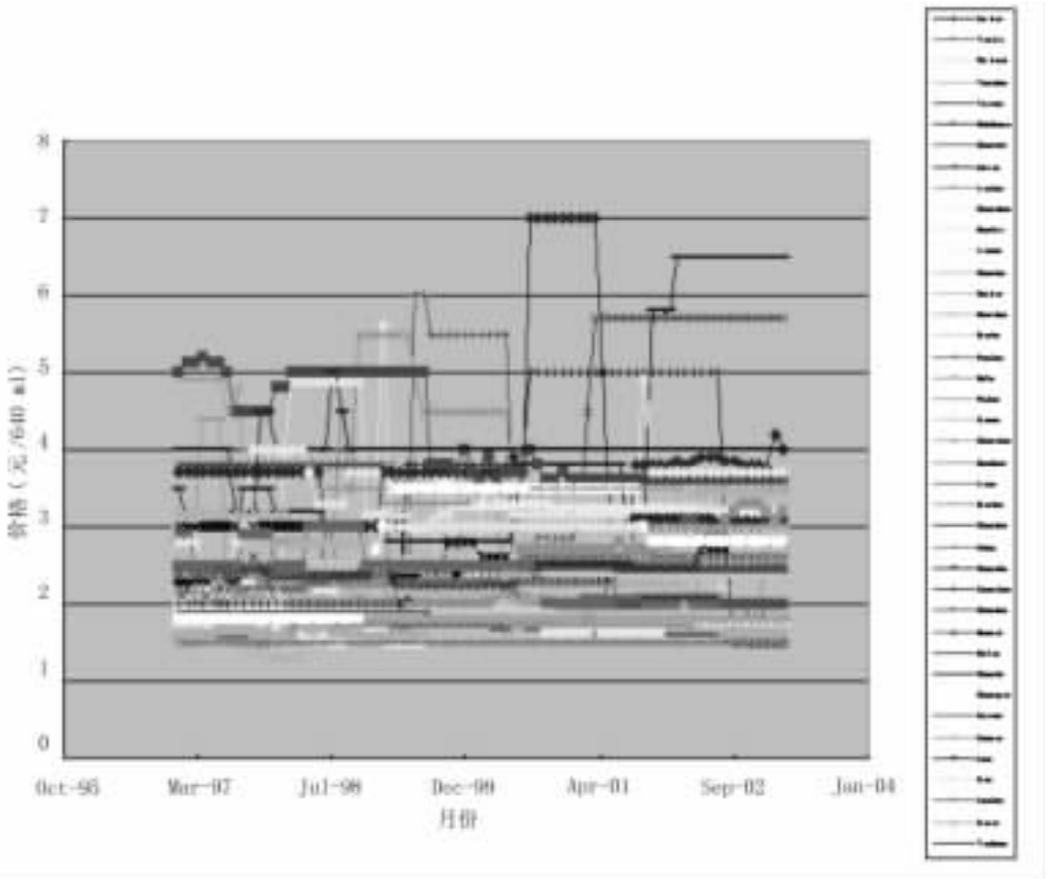


图 1 各城市 640 ml 当地主销啤酒价格变动趋势图 (1997.1~2003.3)

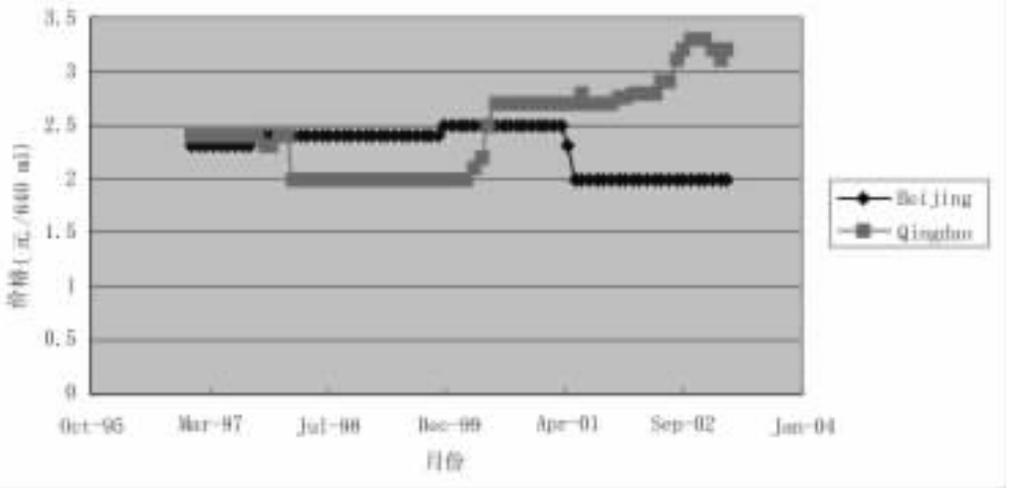


图2 北京 - 青岛啤酒价格变化趋势 (1997.1 ~ 2003.3)

表1 各城市啤酒价格序列的统计描述

	均值	中位数	最大值	最小值	标准差	标准差/均值	观察值个数
北京	2.29	2.40	2.50	2.00	0.20	0.09	75
天津	1.65	1.70	1.75	1.50	0.09	0.06	75
石家庄	3.21	3.20	4.90	2.00	0.78	0.24	75
唐山	1.48	1.50	1.60	1.30	0.06	0.04	75
太原	2.14	2.20	2.40	1.80	0.15	0.07	75
呼和浩特	2.84	3.00	3.10	2.50	0.27	0.09	75
沈阳	2.86	3.00	6.00	1.50	1.34	0.47	75
大连	1.84	1.80	1.90	1.70	0.05	0.03	75
锦州	2.26	1.70	3.60	1.50	0.85	0.38	75
长春	2.74	3.20	3.60	2.00	0.59	0.21	75
哈尔滨	1.79	1.60	3.60	1.50	0.63	0.35	75
佳木斯	1.97	2.00	2.10	1.80	0.07	0.04	75
上海	2.88	3.00	3.60	1.30	0.43	0.15	75
南京	2.77	2.70	3.60	2.40	0.18	0.07	75
杭州	1.56	1.55	1.70	1.50	0.06	0.04	75
宁波	2.30	2.00	3.50	1.80	0.49	0.21	75
温州	2.12	2.10	2.30	2.00	0.12	0.05	75

续上表

	均值	中位数	最大值	最小值	标准差	标准差/均值	观察值个数
合肥	2.84	3.00	4.40	2.10	0.42	0.15	75
福州	1.90	1.80	2.83	1.75	0.26	0.14	75
厦门	2.43	2.40	3.00	2.00	0.23	0.09	75
漳州	2.22	2.20	2.50	1.95	0.21	0.09	75
南昌	2.76	2.00	5.50	1.50	1.35	0.49	75
济南	2.44	2.40	3.00	2.20	0.18	0.07	75
青岛	2.46	2.40	3.30	2.00	0.41	0.17	75
郑州	3.30	3.80	5.00	1.50	1.39	0.42	75
武汉	1.50	1.50	1.60	1.50	0.02	0.36	75
长沙	3.91	3.20	6.50	2.80	1.39	0.28	75
广州	4.06	3.70	7.00	3.00	1.14	0.16	75
深圳	4.18	3.80	5.20	3.40	0.65	0.06	75
南宁	2.40	2.40	2.90	2.20	0.14	0.22	75
海口	4.33	3.80	5.70	3.50	0.96	0.10	75
成都	2.69	2.70	3.00	2.00	0.26	0.26	75
重庆	2.78	2.80	3.50	1.80	0.72	0.09	75
贵阳	2.37	2.50	2.55	2.00	0.21	0.18	75
昆明	3.46	3.50	4.90	2.50	0.63	0.06	75
拉萨	4.00	4.00	5.00	3.20	0.22	0.16	75
西安	3.34	3.00	5.60	2.50	0.53	0.14	75
兰州	2.37	2.50	3.00	1.98	0.34	0.19	75
西宁	2.71	2.40	3.50	2.00	0.50	0.06	75
银川	2.24	2.30	2.60	2.00	0.14	0.01	75

根据这些图表，我们可以初步得出以下几点结论：

(1) 从图 1 上看，各城市的啤酒价格基本不存在季节性波动的趋势，所以我们无须引入季节性虚拟变量对数据进行修正。

(2) 图 1 表明，从 1997 年 1 月至 2001 年 3 月，各城市的啤酒价格差异比较大，城市间的啤酒价格变化缺乏趋同性；2001 年 4 月以后，除少数城市以外，如长沙、海口、郑州，各城市的啤酒价格变化趋势趋于一致，价格差异比较稳定。

(3) 从图 2 上看，北京与青岛在啤酒价格方面的变化趋势差异较大，基本上不存在共同变化的趋势。

(4) 从基本的统计指标看, 均值最小的城市是唐山, 只有 1.48 元, 均值最大的城市是海口, 达到 4.33 元; 从标准差这一指标看, 价格波动比较大的城市是沈阳、南昌、郑州、长沙和广州; 变异系数 (标准差/均值) 的分布范围从 0.01 到 0.49。这些基本的统计指标同样表明, 各地价格波动的差异比较大, 价格分散化的趋势明显, 地区间的价格信息不易于沟通。

2. 单位根检验

根据上面提到的单位根检验方法, 我们需要对价格序列的平稳性进行测定, 这里采用了 ADF 检验方法。由于本研究所使用的价格序列数据并没有明显的随时间变动的趋势, 同时考虑到如果在平稳性分析中包含截距项这个外生变量, 会降低 ADF 检验的能力, 因此从增强 ADF 检验能力的角度考虑, 我们只选择进行不包含截距和时间趋势项的 ADF 检验。我们首先进行序列的水平值测试, 其零假设为各城市的啤酒价格水平序列存在单位根。从表 2 可以看出, 在 5% 的统计水平上, 只有拉萨的价格序列水平平稳, 其他城市的价格序列都不能拒绝存在单位根。因此, 我们必须作进一步的单位根检验, 继续对除拉萨以外的其余城市的啤酒价格序列作一阶单整测试。

表 2 各城市啤酒价格序列的单位根测试

城 市	水平值测试		一阶差分测试	
	ADF 测试 (1)	滞后值 (2)	ADF 测试 (3)	滞后值 (4)
北京	- 0.670	2	- 5.557*	1
天津	- 0.559	2	- 6.058*	1
石家庄	- 2.657	1	- 5.495*	1
唐山	- 3.313	2	- 4.123*	2
太原	- 1.186	1	- 7.572*	1
呼和浩特	- 1.380	1	- 8.485*	0
沈阳	- 2.155	1	- 6.390*	0
大连	- 1.181	0	- 8.544*	0
锦州	- 1.672	1	- 8.322*	0
长春	- 1.059	5	- 8.134*	0
哈尔滨	- 0.451	0	- 8.523*	0
佳木斯	- 1.013	0	- 8.485*	0
上海	- 1.998	2	- 7.941*	1
南京	- 3.145	2	- 7.495*	2
杭州	- 1.758	1	- 5.289*	2

续上表

城市	水平值测试		一阶差分测试	
	ADF 测试 (1)	滞后值 (2)	ADF 测试 (3)	滞后值 (4)
宁波	- 0.157	1	- 5.847 *	0
温州	- 0.113	4	- 4.495 *	3
合肥	- 0.586	1	- 8.333 *	0
福州	- 0.372	1	- 7.796 *	0
厦门	- 0.821	0	- 8.485 *	0
漳州	- 1.559	1	- 5.867 *	1
南昌	- 0.669	3	- 3.977 *	2
济南	- 2.285	2	- 8.672 *	0
青岛	- 0.190	2	- 4.156 *	2
郑州	- 0.630	3	- 4.099 *	2
长沙	- 0.473	1	- 8.458 *	0
广州	- 0.581	1	- 7.265 *	0
深圳	- 1.261	1	- 9.302 *	0
南宁	- 1.462	3	- 6.411 *	2
海口	- 0.648	4	- 3.938 *	3
成都	- 2.144	1	- 5.968 *	1
重庆	- 1.687	2	- 4.031 *	1
贵阳	- 2.487	2	- 5.799 *	2
昆明	- 1.667	1	- 5.824 *	1
拉萨	- 6.011 *	1	NA	NA
西安	- 3.016	1	- 8.322 *	1
兰州	- 0.663	1	- 7.965 *	0
西宁	- 1.150	1	- 6.414 *	1
银川	- 0.832	0	- 6.933 *	1

注：加 * 表示通过显著水平为 5% 的 ADF 单位根检验。

一阶单整测试的零假设为经一阶差分后各城市的价格序列存在单位根。从表 2 可以看出，在 5% 的水平上，所有城市的价格序列都在统计上显著，因此我们可以拒绝零假设，即除拉萨外的其余 39 个城市的价格序列均为一阶单整序列，服从 $I(1)$ 过程。因此，我们就可以对除拉萨外这 39 个城市的啤酒价格序列进行协整检验分析。

3. 协整分析结果

因为除拉萨外，其余 39 个城市的啤酒价格变量序列都是服从 $I(1)$ 过程的同阶序列，所以我们就可运用 Johanson 方法来测定除拉萨外各配对价格序列的协同变化程度。由于北京和青岛是中国最大的两个啤酒生产基地，它们的销售网络遍布全国，分别考察其余城市与这两个城市之间在啤酒价格序列方面的协同趋势，可在相当大程度上反映我国啤酒市场总体的整合或分割情况。我们首先对 1997 年 1 月至 2003 年 3 月整个时期的所有价格序列进行一个长周期的协整检验，考察这段时期内啤酒行业的总体市场整合情况。然后我们再区分两个阶段，即 1997 年 1 月到 2001 年 3 月以及 2001 年 4 月到 2003 年 3 月，分别进行阶段性的协整检验，考察中国加入 WTO 这个事件是否对我国啤酒行业的市场整合趋势产生影响。表 3、表 4 列出了北京（青岛）啤酒市场与其他城市的啤酒市场从 1997 年 1 月到 2003 年 3 月整段时期的整合情况，表 5 和表 6 则分别列出了这两个啤酒市场与其他城市的啤酒市场以 2001 年 4 月作为分界点的整合情况。

表 3 以北京市当地主销 640 ml 的啤酒市场价格作为参考 (1997.1 ~ 2003.3)

时期	与北京存在协整关系的城市个数	与北京存在协整关系的城市	与北京不存在协整关系的城市
1997.1 ~ 2003.3	10	呼和浩特、上海、南京、济南、武汉、长沙、南宁、海口、西安、银川	天津、石家庄、唐山、太原、沈阳、大连、锦州、长春、哈尔滨、佳木斯、杭州、宁波、温州、合肥、福州、厦门、漳州、南昌、青岛、郑州、广州、深圳、成都、重庆、贵阳、昆明、兰州、西宁

表 4 以青岛市当地主销 640 ml 的啤酒市场价格作为参考 (1997.1 ~ 2003.3)

时期	与青岛存在协整关系的城市个数	与青岛存在协整关系的城市	与青岛不存在协整关系的城市
1997.1 ~ 2003.3	11	天津、长春、南京、温州、合肥、福州、厦门、济南、武汉、南宁、西安	北京、石家庄、唐山、太原、呼和浩特、沈阳、大连、锦州、哈尔滨、佳木斯、上海、杭州、宁波、漳州、南昌、郑州、长沙、广州、深圳、海口、成都、重庆、贵阳、昆明、兰州、西宁、银川

表 5 以北京市当地主销 640 ml 的啤酒市场价格作为参考 (以 2001 年 4 月为分界点)

时期	与北京存在协整关系的城市个数	与北京存在协整关系的城市	与北京不存在协整关系的城市
1997.1 ~ 2001.3	8	长春、佳木斯、南京、厦门、济南、武汉、南宁、银川	天津、石家庄、唐山、太原、呼和浩特、大连、锦州、哈尔滨、上海、杭州、宁波、温州、福州、漳州、南昌、青岛、郑州、长沙、广州、深圳、海口、成都、重庆、贵阳、昆明、西安、兰州、西宁、沈阳、合肥
2001.4 ~ 2003.3	28	天津、石家庄、唐山、太原、呼和浩特、沈阳、大连、上海、宁波、温州、福州、厦门、漳州、南昌、济南、武汉、长沙、广州、深圳、南宁、海口、重庆、贵阳、昆明、西安、兰州、西宁、银川	锦州、长春、哈尔滨、佳木斯、南京、杭州、合肥、青岛、郑州、成都

表 6 以青岛市当地主销 640 ml 的啤酒市场价格作为参考 (以 2001 年 4 月为分界点)

时期	与青岛存在协整关系的城市个数	与青岛存在协整关系的城市	与青岛不存在协整关系的城市
1997.1 ~ 2001.3	9	长春、佳木斯、上海、南京、温州、合肥、厦门、武汉、南宁	北京、天津、石家庄、唐山、太原、呼和浩特、沈阳、大连、锦州、哈尔滨、杭州、宁波、漳州、南昌、济南、郑州、长沙、广州、深圳、海口、成都、重庆、贵阳、昆明、西安、兰州、西宁、银川、福州
2001.4 ~ 2003.3	25	天津、石家庄、唐山、太原、呼和浩特、沈阳、大连、锦州、长春、佳木斯、南京、宁波、合肥、福州、漳州、南昌、济南、郑州、武汉、广州、深圳、海口、成都、贵阳、兰州、	北京、哈尔滨、上海、杭州、温州、厦门、长沙、南宁、重庆、昆明、西安、西宁、银川

从整段时期看，我国啤酒市场整合的总体趋势还是比较低的，市场分割的趋势比较明显。具体而言，与北京存在协同价格变动趋势的城市只有 10 个，与青岛存在协同价格变动趋势的城市也只有 11 个。而与北京邻近、同在河北省境内的城市如天津、石家庄和唐山，基本上与北京不存在价格变动方面的协同趋势；与青岛邻近的城市中，除了济南外，基本上与青岛也不存在价格变动方面的协同趋势。另外，一个值得注意的趋势是，北京与青岛这两个最大的啤酒生产地的啤酒价格序列之间并不存在协整关系。

我们知道，中国于 2001 年 11 月正式加入世界贸易组织，此前，中央政府为了加速中国的贸易自由化进程，出台了一系列旨在取消地方保护壁垒的措施，其中最具有影响的一项措施，是国务院于 2001 年 4 月公开发文到各级政府，明令禁止区域间的贸易封锁行为。所以我们首先选择 2001 年 4 月作为分界点，分两阶段进行协整分析。我们发现，两段时期所呈现的啤酒市场整合趋势存在显著的差异。从 1997 年 1 月至 2001 年 3 月，与北京存在协整关系的城市个数只有 8 个，而从 2001 年 4 月至 2003 年 3 月，这一数字却上升到了 28 个。青岛与其他城市的市场整合情况也呈现出类似的趋势，在前一时期只有 9 个，而从 2001 年 4 月开始的后一时期却上升到了 25 个。所以上述协整变化的趋势说明，中国在加入 WTO 后，市场整合的趋势在加强。另外，与北京邻近但在前一时期不存在协整关系的城市，如天津、石家庄、唐山、太原、呼和浩特，在后一时期都与北京市场的价格变动呈现协同趋势。与青岛邻近的城市也有类似的结果。这说明，以北京和青岛为中心的区域性整合趋势在不断增强。但是，北京与青岛的啤酒价格序列，无论是从整段时期看，还是分时期看，都不存在协整关系。

另一方面，在一个政策出台以前，下级政府可能在事先就得到相关的信息，从而对其作出反应，调整自己的政策措施。考虑到政策出台的事前效应，我们分别选择了以 2001 年 4 月往前倒推三个月和六个月作为分界点，分两个时期进行协整测试。我们发现，往前倒推三个月作为分界点的测试结果，与以 2001 年 4 月作为分界点的测试结果基本无差异，即使是往前倒推六个月的测试结果，差异仍然不大。也就是说，我们以 2001 年 4 月作为分界点的处理是合理的。我们在表 7 和表 8 中列出了以 2000 年 10 月作为分界点的协整测试结果。当然，对以 2001 年 4 月和以 2000 年 10 月作为分界点的两个测试结果进行比较后我们看到，前者所体现的市场整合趋势比后者更明显。

表 7 以北京市当地主销 640 ml 的啤酒市场价格作为参考 (以 2000 年 10 月为分界点)

时期	与北京存在协整关系的城市个数	与北京存在协整关系的城市	与北京不存在协整关系的城市
1997.1 ~ 2000.9	8	长春、佳木斯、南京、厦门、济南、武汉、南宁、银川	天津、石家庄、唐山、太原、呼和浩特、沈阳、大连、锦州、哈尔滨、上海、杭州、宁波、合肥、温州、福州、漳州、南昌、青岛、郑州、长沙、广州、深圳、海口、成都、重庆、贵阳、昆明、西安、兰州、西宁
2000.10 ~ 2003.3	25	天津、石家庄、唐山、太原、呼和浩特、沈阳、大连、上海、宁波、温州、福州、厦门、漳州、南昌、济南、武汉、广州、深圳、南宁、海口、重庆、贵阳、西安、兰州、银川	锦州、长春、哈尔滨、佳木斯、南京、杭州、合肥、青岛、郑州、长沙、成都、昆明、西宁

表 8 以青岛市当地主销 640 ml 的啤酒市场价格作为参考 (以 2000 年 10 月为分界点)

时期	与青岛存在协整关系的城市个数	与青岛存在协整关系的城市	与青岛不存在协整关系的城市
1997.1 ~ 2000.9	8	长春、佳木斯、南京、温州、福州、厦门、武汉、南宁	北京、天津、石家庄、唐山、太原、呼和浩特、沈阳、大连、锦州、哈尔滨、上海、杭州、宁波、合肥、漳州、南昌、济南、郑州、长沙、广州、深圳、海口、成都、重庆、贵阳、昆明、西安、兰州、西宁、银川
2000.10 ~ 2003.3	20	石家庄、唐山、太原、大连、锦州、长春、佳木斯、南京、合肥、福州、厦门、南昌、济南、郑州、武汉、深圳、成都、贵阳、兰州、西宁	北京、天津、呼和浩特、沈阳、哈尔滨、上海、杭州、宁波、温州、漳州、长沙、广州、南宁、海口、重庆、昆明、西安、银川

五、结论及其经济解释

本文利用来自我国 40 个城市的月度市场价格资料，以啤酒市场为例，运用协整分析方法，对我国啤酒市场的市场整合趋势作了较深入的分析。结果表明：

(1) 从 1997 年 1 月到 2003 年 3 月整段时期看，我国啤酒市场的市场整合程度不高，市场分割的现象比较普遍。

(2) 从分时期讨论的结果看，在中央政府出台取消地方保护壁垒的措施之前（无论有没有考虑政策出台的事前效应），我国啤酒市场的市场分割现象比较普遍；但在这一政策措施出台以后，啤酒市场却出现了明显的整合趋势。并且，分别以北京和青岛为中心，我国的啤酒市场出现了区域性的整合趋势。

(3) 北京与青岛这两个最大的啤酒生产地的啤酒价格序列之间，无论是从整段时期看，还是分时期看，都不存在协整关系，也就是说，这两个啤酒市场基本上处于分割状态。

从以上结论看，中国加入 WTO 对于中国的市场整合产生了较大的影响，它促使中央政府及各级地方政府相应地调整自己的政策取向，消除地区间的贸易壁垒。我们可以结合中国的税制改革与贸易自由化进程，对政府的这种政策取向变化作出合理的经济学解释。1994 年的税制改革，明确了中央与地方政府之间的税收分配按照分税制的原则进行，地方的税收留成比例大大增加，这就给地方政府提供了进行地方保护的激励。微观经济学的基本理论告诉我们，一个政府在考虑实施一项政策决策时，它必须同时兼顾厂商与消费者两方面的利益。实施地方保护政策往往会增加厂商的利润，但同时也会损害消费者的利益。在中国，从政绩考核的角度考虑，地方政府会更偏向于保护当地企业的利益而甚于消费者的利益。并且，如果国家对外保持比较高的关税水平，外国企业对本国企业的竞争压力就会比较小，因而实行地区间的贸易保护政策就会比较奏效。这就能够在一定程度上解释，为什么中国的啤酒行业在中国加入 WTO 之前市场分割现象如此普遍，其主要的影响因素是分税制的作用。

中国于 2001 年 11 月正式成为 WTO 的成员，在中央政府与各成员国签订的协议中，其中一项重要的承诺，就是逐年降低各类产品的关税。也就是说，要逐步实现贸易自由化。在贸易自由化的过程中，关税水平会不断降低，这意味着外国企业进入本国市场的成本降低，竞争力加强，此时，设置地区间的贸易壁垒只是在一定程度上给外地企业进入本地市场造成了障碍，但由此让渡出来的市场份额很有可能被进入成本日益降低的外国企业所攫取。^① 这样，地方政府在设置地区间的贸易壁垒时，就必须考虑这一做法是否能够增加本地企业的利润，这种增加是否足以抵偿消费者的利益损失，如果这种增加不能抵偿消费者的利益损失，地方政府就有可能自动放弃地区间的贸易保护壁垒。这能够在一定程度上解释为什么中国啤酒市场的整合趋势日益明显。

对于北京与青岛的啤酒价格序列之间不存在协整关系这一结论，我们提供的一个解释

^① 关于这一解释，Li 等（2003）给出了严密的理论推导。

是,由于燕京和青岛啤酒边际生产成本都不高,运输成本在总成本中就成为比较重要的一个因素。燕京和青岛啤酒都是以本地市场为生产基地并逐渐向外地市场辐射而发展起来的,它们在本地市场的影响根深蒂固,其他品牌的啤酒往往难以进入当地市场。这样,它们在本地就具有足够的市场势力,并导致北京和青岛这两个啤酒市场的价格决定具有相对独立性。

本研究仍然存在许多有待改进之处。首先,由于数据的原因,本文只采用了协整分析的方法分析市场整合的变化趋势,而没有采用更多的方法对市场整合趋势作相应的检验。其次,由于缺乏财税改革方面的相关数据,我们没有验证税制改革对市场整合的具体影响。这些都是我们今后进一步研究的方向。

参考文献

- [1] Alwyn Young, *The Razor's Edge: Distortions and Incremental Reform in the People's Republic of China*, *Quarterly Journal of Economics*, 2000, Vol.115, No.4, pp1091~1135.
- [2] Chong-En Bai et al., *Protectionism and Regional Specialization: Evidence from China's Industries*, Working Paper, Mimeo, 2002.
- [3] Gilley B., *Breaking Barriers*, *Far Eastern Economic Review*, 2001, pp12~19.
- [4] Jie Li, Anming Zhang, *WTO Accession and China's Domestic Regional Liberalization: A Theoretical Analysis*, *Pacific Economic Review*, 2003.
- [5] Jie Li, Larry D. Qiu and Qunyan Sun, *Interregional Protection: Implications of Fiscal Decentralization and Trade Liberalization*, *China Economic Review*, 2003.
- [6] Jiming Ha, Kelvin Fan, *One country, Two Prices? A study of Deflationary Effect of Price Convergence on Hong Kong*, working paper, Mineo, 2003.
- [7] William H. Greene, *Econometric Analysis*, Fifth Edition, 2003.
- [8] 陈抗, Arye L. Hillman, 顾清扬:《财政集权与地方政府行为变化—从援助之手到攫取之手》,《经济季刊》,2002年第2卷第1期。
- [9] 银温泉,才婉如:《中国地区间市场分割成因和治理》,《经济研究》,2001年第6期。
- [10] 喻闻,黄季琨:《从大米市场整合程度看我国粮食市场改革》,《经济研究》,1998年第3期。
- [11] 周章跃,万广华:《论市场整合研究方法》,《经济研究》,1999年第3期。
- [12] 北京市经济信息中心:《中国啤酒:从“春秋”到“战国”》,2001年6月4日。

第四部分

税收与财政

我国出口退税效应分析：理论与实证^{*}

陈 平 黄健梅

2003年，我国政府对出口退税政策实施了一系列的调整。一方面，为化解“非典”对外贸出口的负面影响，2003年6月财政部宣布增加退税指标358亿，2003年上半年已经下达了1027亿退税指标，相当于2002年全年退税额的82%，同时国家税务总局发出加快出口退税进度的紧急通知。另一方面，2003年10月，政府又宣布出口退税率将平均下调3%左右。这些政策调整将对我国的出口产生怎样的影响，再度成为人们关注的焦点。

一、出口退税政策的演变与问题的提出

出口退税 (Export Drawback or Export Rebate) 是一国政府对出口商品退还其已征收的增值税、消费税，试图使商品以不含税价公平地参加国际市场竞争，促进本国出口。出口退税作为一种国际通行的做法，已被国际社会广泛地认同和采用，如法国、德国、意大利、韩国等都实施出口退税政策。

我国出口退税的做法可追溯到1950年，当时已有对出口货物退还已纳货物税的规定，而正式的出口退税制度则建立于1985年。1994年税制改革后，《增值税暂行条例》、《消费者暂行条例》和《出口货物退(免)税管理办法》等一系列法规对出口货物增值税、消费税的退还或免征范围、退税率、退税额计算等程序和管理制度作出了具体规定。出口退税的主要税种增值税^①，是目前我国税制中最主要的流转税之一，基本税率为17%，其征收范围很广。随着经济环境的变化，1994年以来出口退税率及退税范围不断调整(见表1)，其中1999年的大调整形成了目前17%，15%，13%，5%这四档增值税出口退税率^②，根据2003年10月13日出台的新改革方案，2004年起我国将执行17%，13%，11%，8%，5%五档退税率，平均退税率将下调3%左右。我国出口退税额随着出口的不断增长而不断攀升(见图1)，1985年退税额仅为17.95亿，2002年高达1258亿，增加了69倍。

^{*} 在本文的写作中，与国际商务系主任黄静波博士的讨论使笔者受益匪浅，特此致谢。原载：《管理世界》，2003年第12期。

① 增值税是对在我国境内销售货物或者提供加工、修理修配劳务，以及进口货物的单位和个人，就取得的货物或应税劳务的销售额以及进口货物的金额计算税款，并实行税款抵扣的一种流转税。

② 我国消费税只对特定的11类商品征收，消费税退税率为消费者征收率，基本不存在不完全退税问题，加上消费税退税占总退税额很小，因而本文主要探讨增值税退税政策变化的影响。

表 1 出口退税政策变化汇总表

时 间	内 容
1995 年 1 月	取消柴油出口退税
1995 年 7 月	农产品、煤炭税率为 3%，以农产品为原料加工生产的工业品和适用 13% 增值税税率的其他货物，退税率为 10%，适用 17% 增值税税率的其他货物，退税率为 14%
1995 年 10 月	新闻纸停止办理退税
1997 年 1 月	外商投资企业出口货物，均可向海关申请办理出口退税专用报关单
1997 年 10 月	新闻纸恢复办理出口退税，退税率为 9%
1998 年 1 月	纺织原料及制品的出口退税率提高到 11%（从一般纳税人购进的按 11% 的退税率办理，从小规模纳税人购进的按 6% 的退税率办理退税）；纺织机械的出口退税率提高到 17%
1998 年 6 月	煤炭出口退税率调为 9%，钢材出口退税率调为 11%，水泥出口退税率调为 11%，船舶出口退税率调为 14%
1998 年 7 月	七类机电产品、五类轻工产品的出口退税率从 9% 提高到 11%
1998 年 8 月	恢复对食糖的出口退税，按 9% 的退税率办理退税
1998 年 9 月	铝、锌、铅出口退税率调为 11%
1998 年 10 月	船舶出口退税率调为 16%
1999 年 1 月	机械及设备、电器及电子产品、运输工具、仪器仪表四大类机电产品的出口退税率提高到 17%，农机的出口退税率提高到 13%，有机化工原料、无机化工原料、涂料、染料、颜料、橡胶制品、玩具及运动用品、塑料制品、旅行用品及箱包的出口退税率提高到 11%，原来执行 6% 出口退税率的货物，包括以农产品为原料加生产的工业品及其他货物的出口退税率提高到 9%，农产品的出口退税率提高到 5%
1999 年 7 月	服装的出口退税率提高到 17%；服装以外的纺织原料及制品，机电产品中除机械及设备、电器及电子产品、运输工具、仪器仪表等目前已执行 17% 退税率的四大类产品以外的其他机电产品，以及法定税率为 17% 且现行退税率为 13%，11% 的货物，出口退税率统一提高到 15%；法定税率为 17% 且现行退税率为 9% 的其他货物的出口退税率提高到 13%；农产品以外的法定税率为 13% 且现行退税率未达到 13% 的货物，出口退税率提高到 13%
1999 年 12 月	出口计划内的一般贸易出口柴油恢复退税，退税率为 13%，消费税按照法定税额退（免）税
2001 年 7 月	纱布的出口退税率由 15% 提高到 17%
2002 年 1 月	列示的电解铜产品出口退税 17%

续上表

时 间	内 容
2002 年 9 月	出口加工区耗用水、电、气准予退税，退税率为 13%
2004 年 1 月	小麦粉、玉米粉、分割鸭、分割兔等出口退税率，由 5% 调高到 13%；取消原油、木材、纸浆、山羊绒、鳃鱼苗、稀土金属矿、磷矿石、天然石墨等所列明货物的出口退税政策；调低下列货物的出口退税率：汽油、未锻轧锌的退税率调低到 11%，未锻轧铝、黄磷及其他磷、未锻轧镍、铁合金、钨矿砂及其精矿等货物的退税率调低到 8%，焦炭半焦炭、炼焦煤、轻重烧镁、莹石、滑石、冻石等附件五所列明的货物的出口退税率调低到 5%；其余退税率为 17% 和 15% 的货物，其退税率一律调低到 13%；凡现行征税率和退税率均为 13% 的货物，其出口退税率一律调低到 11%

注：①资料来源于中国资讯行数据库 www.bjinfobank.com；②表中所列时间为政策开始实施时间。

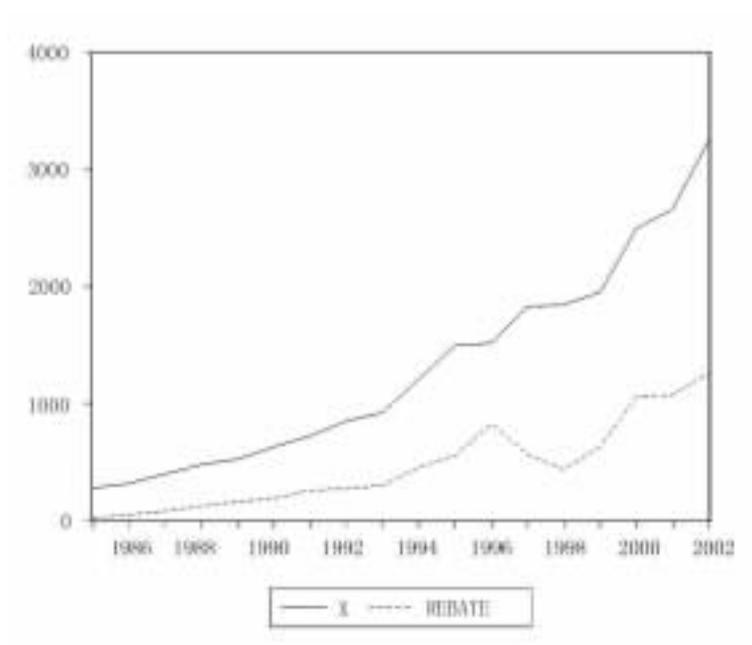


图 1 1985 ~ 2002 年我国出口与出口退税增长图

注：①数据来源于《中国税务年鉴》和《中国统计年鉴》。②出口额以亿美元计价；出口退税额以亿元计价

随着各国经济相互依赖程度的加强，进出口为各国政府所日益重视。尽管与储蓄一样，进口增长对于一国生产能力或潜在 GNP 的增长是十分重要的，但对于一国生产能力的充分利用则产生不利的影响；相应地，出口的增长通过提高总需求，对于一国充分利用现有的生产能力，驱使现实 GNP 等于潜在 GNP 至关重要，因而更为注重短期经济稳定（相对于长期经济增长）目标的各国政府所青睐。在政策现实中，不难发现各国政府无不在实施奖励出口、抑制进口的“重商主义”政策主张。

在我国，出口状况更是备受政府的关注。出口的增长除了对 GNP 的提高起促进作用

外,对于人口大国还具有特殊意义。人口大国的绝对和比较优势所在正在于劳动密集型产品,出口的增长对于解决我国长期存在的就业问题发挥着不可替代的作用。这预示在我国较长的时间内仍会也理应将出口增长作为政府一个无可辩驳的重要目标。

从宏观层面来看,一国出口状况除了受各国内外市场环境的影响,还受该国各种贸易政策和汇率政策的影响。政府根据市场环境变化,灵活地运用关税政策、出口信贷政策、出口退税政策、汇率政策等,以有效地促进外贸出口。随着我国政策环境的变化,出口退税政策对促进出口发挥越来越重要的作用。特别是1991年和1994年我国分别取消出口补贴和外汇留成后,出口退税政策事实上已成为我国最重要的一项出口奖励措施。特别是我国人民币汇率自1996年实行盯住美元的做法后,出口退税的税率和适用商品范围的调整更是成为我国政府应对市场因素变化、力保出口稳步增长的惟一有效的手段。在经受亚洲金融风暴的冲击、在邻国汇率贬值的压力下维持人民币不贬值的情况下,为保持外贸出口的增长势头,出口退税充分显示了其积极而灵活的政策杠杆作用。

2003年,我国政府颁布和实施(或即将实施)了一系列看似相互冲突的出口退税政策,将对我国的出口状况产生怎样综合的影响,是政府、企业界和学术界共同关注的问题。

在我国,关于出口退税的研究文献已有不少,但大部分的文献集中在从税制角度研究出口退税制度存在的缺陷及如何完善出口退税政策(罗丙志,2001;卓岩,2002;董俊英、姚坚,2002;等)。有关出口退税效应分析的研究,对于出口退税对出口、财政收支、GDP的影响也作了积极有益的探讨(隆国强,1998;李万甫、马衍伟,2000;牛锐,2002;等)。但是,在已有文献中,分析存在的一个缺陷在于:就出口退税谈出口退税。在理论上没有很好地综合考虑其他政策特别是汇率政策与出口退税政策的影响;在政策建议上无法考虑各种政策之间的协调,在人民币汇率升值压力存在的今天,该种分析就显得不足了。

针对已有研究的不足,本文在已有研究的基础上引入实际有效汇率的概念,分别从理论和实证两方面分析出口退税政策对出口盈利性以及出口的影响,为我国出口退税、汇率等政策的协调制定提供了一个有益的理论分析框架,并提出自己的政策建议。

二、理论阐述:出口退税与出口盈利性

一国的出口取决于出口供给与出口需求。影响出口供给与出口需求的因素中,有价格因素与非价格因素。但价格因素无疑是决定短期出口走势更重要的因素,因此绝大多数的贸易文献均着力于对价格因素以及相应的价格竞争力和贸易盈利性的分析。

分析出口盈利性及其变动一个最全面的指标是实际有效汇率(Real Effective Exchange Rate)。这一指标是有效汇率(Effective Exchange Rate)与实际汇率(Real Exchange Rate)的综合。其中,有效汇率的概念又涉及二个含义:一是为人们通常使用的,国际货币基金组织所倡导的,指的是贸易加权的多边汇率指数,用以衡量一国货币在一定时期内的总体走势。国际货币基金组织对各主要会员国货币的多边汇率指数进行测算并公布。但在经济分析中,有效汇率还被用来衡量一国一整套政策对进出口盈利性的效应。这一整套政策包括

汇率和各种价格性商业政策（贸易和外汇政策），如关税、出口补贴、出口信贷优惠、出口退税、复汇率等。用公式表示，则为：

$$Se = S \times (1 + w)$$

式中： Se ， S 分别为有效汇率和名义汇率； w 泛指各种价格性商业政策，如关税、退税、补贴率等。由于各种商品的关税、退税、补贴率可以是不同的，因此不同商品的有效汇率也往往是不同的。该有效汇率概念测度了贸易商出口获得 1 单位外汇后最终能够得到的本币单位数，或者是进口支付 1 单位外汇总共所需花费的本币单位数。该指标是与国内外比价配套使用来衡量贸易盈利性的高低变化的。以我国出口为例来看：

$$\begin{aligned} \text{出口利润额 (元)} &= \text{出口外汇净收入} \times \text{有效汇率} - \text{出口总成本} \\ &= \text{出口外汇净收入} \times (\text{有效汇率} - \text{出口换汇成本})^{\text{①}} \end{aligned}$$

$$\begin{aligned} \text{出口盈利率} &= (\text{出口人民币净收入} - \text{出口总成本}) / \text{出口总成本} \\ &= (\text{出口外汇净收入} \times \text{有效汇率} - \text{出口总成本}) / \text{出口总成本} \\ &= \text{利润总额} / \text{出口总成本} = \text{有效汇率} / \text{出口换汇成本} - 1 \end{aligned}$$

如果有效汇率相对于出口换汇成本上升，那么出口盈利率就得以提高，从而诱使出口的增加；反之亦然。

将有效汇率与国内外商品比价相结合，我们就有一个分析盈利性的综合指标——实际有效汇率（REER）。

$$REER = Se \times \frac{P_c}{P_n}$$

其中： P_c ， P_n 分别为国内价格（指数）、贸易伙伴国价格（指数）。

不难看出，在实际有效汇率的分析框架下，出口退税是通过 w 来影响出口盈利性的。简单以 2002 年为例，出口额为 3255.7 亿美元，出口退税 1258 亿元，平均出口 1 美元产品得到出口退税 0.386 元，人民币对美元的双边汇率为 8.277，这样，大体来看，出口的有效汇率即为 $8.277 + 0.386 = 8.663$ 。出口换汇成本只要低于 8.663，出口商就可以获利。从动态来看，出口退税的变化、名义汇率的变化和国内外价格的变化将共同决定实际有效汇率和出口盈利性的变化。出口退税率的提高，在名义汇率和出口换汇成本不变的情况下，名义和实际有效汇率均将上升，从而起着与人民币贬值同样的出口鼓励效应。

值得一提的是，对于有效汇率的两层含义，在本文中均有涉及。由于本文是从宏观经济（而非微观或企业）层面上关注出口盈利性，名义汇率的变动是用贸易加权的多边汇率指数，而非双边汇率来衡量的。因此，1996 年来，虽然名义汇率一直盯住美元，人民币对美元汇率始终维持在 8.27 左右，但用多边汇率指数表示的名义汇率是不断调整的，随着美元对其他货币的汇率变化而变化。2002 年美元改变强势政策前，人民币名义有效汇率下降，即人民币处于升值状态。

① 出口换汇成本是出口商品平均每取得 1 美元的外汇净收入所耗费的人民币成本。出口换汇成本 = 出口总成本（人民币元）/ 出口外汇净收入（美元）。其中出口总成本，包括进货（或生产）成本、国内费用（储运、管理费用）及税金。出口外汇净收入指的是扣除运费和保险费后的 FOB 外汇净收入。

三、实证分析

(一) 协整分析

为检验出口退税、实际有效汇率对我国出口的作用，在参照 Arize 等（2000）和 Chi-Chur Chao 等（2002）研究的基础上，设定我国出口的均衡模型为：

$$\ln(X_t) = c_1 + c_2 \times \ln(O_t) + c_3 \times \ln(Z_t) + c_4 \times \ln(V_t) + c_5 \times \ln(P_t) + c_6 \times \ln(Se_t) + \varepsilon_t \quad (1)$$

其中： \ln 表示取自然对数； X_t 为 t 期出口额； O_t 为 t 期的出口退税额； P_t ， Z_t ， V_t ， Se_t 分别表示 t 期的我国出口产品与世界价格的相对价格、外国收入、外汇波动率、有效汇率； ε_t 为 t 期的白噪声。各变量的数据来源见表2。

表2 各变量内容及数据来源表

	内 容	数据来源
X_t	我国各期出口额（亿元）	《中国统计年鉴》
O_t	我国各期出口退税额（亿元）	《中国财政年鉴》、《中国税务年鉴》
P_t	我国出口产品与世界价格的相对价格（1995年为100）	IMF出版的 International Financial Statistics
Z_t	以我国对美国、日本、中国香港、韩国、德国五国（地区）出口额为权重计算的外国GDP	IMF出版的 International Financial Statistics、《中国海关年鉴》
V_t	汇率波动率	《中国统计年鉴》
Se_t	有效汇率 Se_t ：在名义有效汇率基础上，根据出口信贷数额以及支付的出口信贷总费用、出口退税额，对平均汇率进行综合调整后得到	IMF出版的 International Financial Statistics、《中国统计年鉴》、《中国经济年鉴》
Se_0t	有效汇率 Se_0t ：在 Se_t 的基础上剔除出口退税的影响	同上栏

注：①样本期为1985~2002年。②汇率波动率的计算方法是：首先算 $V_t = d + K_{t-1} \times T + Q_t$ ，其中 $T = t - 15, \dots, t - 1$ ；接着算 $\sigma_t^2 = \sum_{i=t-14}^{t-1} (V_i - V_{i-1} - K_{i-1})^2 / 14$ ；最后计算进口加权的储备的标准差 σ_t 。③目前我国中有少数几种商品出口征收关税，这里的有效汇率不考虑关税的影响。

1. 实证结果

根据 Eagle 的研究，利用协整分析可探讨变量间的长期均衡关系以及短期变动关系，可分三个步骤完成。

第一步：平稳性 (Stationary) 检验。在检验 (1) 式是否存在协整关系前，必须对各变量进行平稳性检验。这里利用 ADF (Augmented Dickey-Fuller) 进行单位根 (Unit Root) 检验，当我们拒绝零假设时，便可判断序列平稳^①。ADF 检验结果见表 3。

表 3 ADF 检验结果

变量	差分阶数	滞后项	ADF 统计值	1% 临界值	DW 值	结 论
$\ln X_t$	2	0	- 6.1742	- 2.7411	2.46	I (2), 无自相关
$\ln O_t$	2	0	- 4.6016	- 2.7411	2.18	I (2), 无自相关
$\ln P_t$	2	0	- 6.1078	- 2.7411	2.13	I (2), 无自相关
$\ln Z_t$	2	0	- 4.1459	- 2.7570	2.13	I (2), 无自相关
$\ln V_t$	1	0	- 4.0798	- 2.7275	2.35	I (1), 无自相关
$\ln Se_t$	1	0	- 4.1297	- 2.7275	2.33	I (1), 无自相关

注：I (n) 表示该序列经过 n 次差分后达到平稳。

可见，各变量经过 2 阶或 1 阶差分后都显著地拒绝零假设，即各变量序列差分后达到平稳。

第二步：对 (1) 式进行 OLS 回归，生成残差序列 ee ，对序列 ee 进行平稳性检验。

回归结果如下：

$$\ln X_t = -4.541 + 0.498 \times \ln(O_t) + 0.693 \times \ln(Z_t) + 0.099 \times \ln(V_t) + 0.856 \times \ln(P_t) + 0.383 \times \ln(Se_t)$$

(2.715)(0.135) (0.200) (0.260) (1.546) (0.313)

$$DW = 2.20; F \text{ 统计量} = 177.72$$

其中，括号内数值为系数标准差。残差序列 ee 的 ADF 统计量为 - 4.3532，远小于 1% 的临界值 - 2.7275，且 DW 值为 2.00。不难看出，该序列平稳且不存在自相关。为避免 O_t 与 Se_t 产生估计有偏问题，利用剔除了出口退税影响的 Se_0 进行稳健回归，其结论不变。根据 Greene (1999) 的结论可知，当残差序列 ee 平稳时可判定各变量的线性组合存在协整关系， $c_1 \sim c_6$ 各系数为长期参数，表示各变量在长期均衡中的作用。

第三步：根据 (1) 式构建误差修正模型 (ECM) 进行回归分析，以确定短期参数 $b_1 \sim b_6$ ——表示各变量在短期波动中的作用。ECM 模型具体为：

$$\Delta_2 \ln(X_t) = b_1 + b_2 \times \Delta_2 \ln(O_t) + b_3 \times \Delta_2 \ln(Z_t) + b_4 \times \Delta_1 \ln(V_t) \\ + b_5 \times \Delta_2 \ln(P_t) + b_6 \times \Delta_1 \ln(Se_t) + \theta \times ECM_{t-1} + \varepsilon_t$$

回归结果如表 4 所示。

^① 具体分析请参看 William H. Greene, *Econometric Analysis* (fourth edition), pp781 ~ 784.

表 4 ECM模型估计结果

因变量： $\Delta_2 \ln(X_t)$			
变量	系数值	系数标准差	t 统计量
C	0.000	0.051	0.008
$\Delta_2 \ln(O_t)$	0.438 ^{**}	0.188	2.335
$\Delta_2 \ln(Z_t)$	0.559 [*]	0.296	1.889
$\Delta_1 \ln(V_t)$	0.047	0.353	0.133
$\Delta_2 \ln(P_t)$	- 0.392	0.888	- 0.442
$\Delta_1 \ln(Se_t)$	0.281	0.625	0.450
ECM _{t-1}	- 1.920 ^{**}	0.534	- 3.593
R-square	0.700	DW 值	2.210

注：① * * (*) 表示在 5% (10%) 水平上显著；② Δ_2 表示 2 阶差分，ECM 为均衡误差项， $ECM = \ln X - 0.438 \times \ln O - 0.559 \times \ln Z - 0.047 \times \ln V + 0.392 \times \ln P - 0.281 \times \ln Se$ 。

为验证步骤二的结论，在 ECM 的基础上进行 Johansen 协整检验，结果支持步骤二的结论（见表 5）。

表 5 Johansen 协整检验结果表

序列： $\ln X_t, \ln O_t, \ln Z_t, \ln V_t, \ln P_t, \ln Se_t$				
滞后间隔：0				
特征值	似然率	5% 临界值	1% 临界值	假定协整方程个数
0.979	138.411	94.150	103.180	0 ^{**}
0.892	76.458	68.520	76.070	At most 1 ^{**}
0.758	40.833	47.210	54.460	At most 2
0.550	18.118	29.680	35.650	At most 3
0.281	5.342	15.410	20.040	At most 4
0.005	0.074	3.760	6.650	At most 5

注：① * * (* * *) 表示在 5% (1%) 水平上拒绝零假设；② L.R. 检验表明在 5% 显著水平上存在两个协整方程。

2. 结论

从上面的结果容易看到：① 各变量之间存在线性协整关系，即存在长期均衡关系；出口退税的长期参数 c_2 为 0.498，表明出口退税对出口具有长期促进作用，出口退税增加 1%，出口将增加 0.498%。② 从 ECM 结果来看，短期参数 $b_1 \sim b_6$ 中只有国外收入的短期

参数在 5% 水平上显著，出口退税在 5% 的水平上显著。可见，在影响出口的因素中，短期内有效地促进出口的因素只有出口退税和国外收入。由于国外收入对我国政府而言属于外生变量，政府无法自主地改变国外收入。因而，政府手中刺激出口的有力政策工具为出口退税政策。实际上，我国频频利用调整退税率来刺激出口。从 ECM 的结果看到，出口退税短期波动 1%，出口将增加 0.438%，出口退税的短期效应显而易见。^③出口退税的短期系数在 5% 的水平上显著，而剔除出口退税影响的实际有效汇率的短期系数则不显著，表明目前我国一整套政策刺激出口的短期综合效果不理想。究其原因，可能在于我国的名义汇率过于僵化，除 1994~1995 年外人民币汇率的基本稳定与我国出口的大幅增长，形成了极其明显的反差，致使总体的政策配合效果差强人意。实证结果表明，从出口奖励的角度来考察，当前汇率政策应加以反思。

（二）Panel Data（面板数据）估计

前面利用 1985~2002 年的全国数据进行的协整分析可清楚看到出口退税对出口的有效促进作用。但由于上述实证分析的观测值有限（全国观测值个数 18 个），没有达到大样本要求，从而存在观测值不足而造成有偏估计的可能性。为确保实证分析的稳健性，接下来利用分地区的面板数据进行分析。我们利用历年出口退税额排名均在前 10 位以内的 9 个省 1992~2001 年的数据进行回归分析。

$$\ln(X_{it}) = d_{1i} + d_2 \times \ln(O_{it}) + d_3 \times \ln(Z_t) + d_4 \times \ln(V_t) + d_5 \times \ln(P_t) + d_6 \times \ln(Se_t) + \epsilon_{it} \quad (2)$$

此为总体估计模型，其中 $i=1, 2, \dots, 9$ 。1 = 北京、2 = 天津、3 = 辽宁、4 = 上海、5 = 江苏、6 = 浙江、7 = 福建、8 = 山东、9 = 广东。

由于数据原因——各省份的出口退税数据齐全程度不一，这里采用非平衡面板数据分析方法。回归结果如表 6 所示。

从表 6 看到，结合 Greene (1999) 关于 Models for Panel Data 的论述，① 总体估计弥补了单组数据不足的缺陷，其结果支持了协整分析的结论：出口退税对促进出口的作用显著，但总体估计结果表明有效汇率的作用明显；② 组间估计消除了时间影响，出口退税系数高达 1.043，反映了在长期中我国出口退税在促进出口中发挥了极其重要的作用；③ 组内估计则表明各变量在短期效应中仅出口退税显著，其余变量不显著；④ 地区效应分析结果表明：在九个省份中，出口退税系数除了辽宁、广东外，其他基本在 5% 水平下显著，即出口退税对出口起正向积极作用的结论成立，当前有效汇率对我国出口的促进作用显著。综合来看，Panel Data 估计的结论与前面协整分析的结论基本一致，反映有效汇率在促进我国出口方面发挥了一定作用。

（三）出口退税对出口盈利性的影响分析

从前面的分析可知，由于我国增值税制的特点导致外贸企业的出口换汇成本较高，能否退税将直接影响企业的盈亏。目前我国出口退税制度对出口盈利性的影响如何？接下来

笔者利用 1985 ~ 2001 年轻工业、纺织、化学工业、机械工业、电子工业五个行业^① 数据进行实证分析。

表 6 Panel Data 估计结果表

样本期：1992 ~ 2001

估计方法：GLS

总面板（非平衡）观测值：79

系数估计	d ₂	d ₃	d ₄	d ₅	d ₆	R ²			
总体估计	0.524*** (0.050)	0.029 (0.190)	0.045 (0.154)	4.278 (0.823)	2.208*** (0.375)	0.994			
组间估计	1.043*** (0.067)	3.11E-13 (0.279)	3.03E-13 (0.257)	2.32E-13 (1.326)	2.45E-12 (0.598)	0.743			
组内估计	0.775*** (0.038)	4.72E-13 (0.090)	-5.95E-13 (0.083)	2.24E-12 (0.426)	4.32E-12 (0.192)	0.998			
地区效应		0.450** (0.212)	3.736*** (0.705)	-0.216 (0.159)	2.343*** (0.309)	0.997			
ln(O _{it})的系数	d ₂₁	d ²²	d ₂₃	d ₂₄	d ₂₅	d ₂₆	d ₂₇	d ₂₈	d ₂₉
	0.987**	0.417***	0.003	0.341**	0.523***	0.411***	0.113*	0.424***	0.148
常数项	d ₁₁	d ₁₂	d ₁₃	d ₁₄	d ₁₅	d ₁₆	d ₁₇	d ₁₈	d ₁₉
	-12.97	-10.86	-9.22	-10.10	-10.89	-10.67	-9.72	-10.57	-7.84

注：①*，**，***分别表示在 10%，5%，1%水平上显著；括号内数值为系数标准差。②组间估计模型为： $\bar{Y}_i = \alpha + \beta \bar{K}_i + \bar{\varepsilon}_i$ ，其中 \bar{Y}_i 为 $\ln(X_{it})$ 对 t 的均值， \bar{K}_i 为各自变量对 t 的均值， α 为常数项；组内模型则为： $Y_{it} - \bar{Y}_i = \beta (K_{it} - \bar{K}_i) + \varepsilon_{it} - \bar{\varepsilon}_i$ 。详细分析请参看 *Econometric Analysis* (Greene, 1999), pp557 ~ 589。

为考核我国出口退税对出口盈利性的影响，构建模型如下：

$$\ln(\text{PROFIT}_{it}) = D_1 \times \ln(\text{Se}_{it} - C_{it}) + D_2 \times \ln(\text{EXPORT}_{it}) + U_{it} \quad (3)$$

$$\ln(\text{PROFIT}_{it}) = D_1 \times \ln(\text{Se0}_{it} - C_{it}) + D_2 \times \ln(\text{EXPORT}_{it}) + U_{it} \quad (4)$$

其中：PROFIT，EXPORT 的数据来源于《中国电子工业年鉴》； t 表示第 t 期；PROFIT，EXPORT 分别表示行业出口总利润、行业出口收入，且均将数据进行指数化处理； U_i 为白噪声； Se ， $Se0$ 分别为综合考虑一整套政策的有效汇率、剔除出口退税政策影响的有效汇率； C 为出口换汇成本，受资料来源所限， C 乃根据行业总利润、总销售收入、出口收入之间的联系估算出来； $(Se - C)$ 表示出口一美元产品最后得到的净利润，反映我国一整套政策对出口盈利性的作用。通过比较 (3) 式和 (4) 式的回归结果，可分析存在与不存在出口退税政策各自对出口盈利性的影响。如表 7 所示。

① 由于统计口径变化及数据来源所限，五个行业各自的数据齐全程度不一，故这里采用的是非平衡面板数据分析。

表 7 行业 Panel Data 估计结果

样本期：1985 ~ 2001

估计方法：GLS

总面板（非平衡）

观测值：79

(3) 式	系数估计					
	D_1		D_2		R^2	
总体估计	0.209*** (0.033)		0.895*** (0.034)		0.762	
组间估计	0.401*** (0.020)		0.726*** (0.017)		0.999	
组内估计	0.087* (0.054)		0.629*** (0.112)		0.979	
行业效应	轻工业	纺织工业	化学工业	机械工业	电子工业	R^2
$\ln(Se_{it} - C_{it})$	0.346***	0.677***	0.383***	0.083**	0.059**	0.967
$\ln(EXPORT_{it})$	0.785***	0.772***	0.771***	1.031***	0.944***	
(4)式 剔除出口退税政策影响的回归结果			D_1	D_2	R^2	
总体估计			0.080* (0.023)	0.931*** (0.022)	0.991	

注：①*，**，***分别表示在10%，5%，1%水平上显著；括号内数值为系数标准差。②行业效应利用克服自相关的Marquandt非线性回归进行Panel Data的FGLS估计。

我们不难看到：①总体估计结果表明 $(Se - C)$ 、 $(Se_0 - C)$ 均与行业出口利润显著正相关， $(Se - C)$ 增加1%时，出口利润将上升0.209%，足见我国目前一整套政策对出口盈亏的重要作用。②无论是组间估计还是组内估计，结果都反映了 $(Se - C)$ 对出口盈利性的重要影响。消除了时间效应的组间估计表明，我国一整套政策对出口盈利性起着决定性作用；在短期内，组内估计显示 $(Se - C)$ 的系数为0.087，仅相当于长期系数0.401的21.7%。③如果剔除出口退税政策的影响， $(Se_0 - C)$ 变动1%，出口利润只增加0.080%，出口利润增长降低了61.7%，凸现出口退税政策对企业出口盈利性的重要性。这从另一个角度说明出口退税滞后制约外贸出口企业发展的严重性。

四、结论与政策思考

无论是协整分析还是Panel Data的分析结果均表明，我国出口退税政策通过实际有效汇率对出口盈利性产生显著的影响，从而使得该政策对促进我国的出口增长无论在长期还是在短期里均起着极其重要的作用。可见，政府利用出退税政策来缓解“非典”对出口、经济的不利影响有着较科学合理的政策依据。实证结果支持了出口退税正效应的理论分析。我国政府出于减轻中央财政的负担和缓解人民币升值的压力，将出口退税率平均下调了3%。可以预计，在没有其他配套措施的情况下，这将会对我国的出口产生一定的不利影响，抑制我国的出口增长势头。

本文的实证分析还表明，剔除出口退税影响后的实际有效汇率对我国出口的影响并不

显著，主要原因应在于僵化的汇率体制。一方面，僵化的汇率使得出口奖励的政策压力落在了出口退税上；另一方面，实施出口退税所带来的出口增长及相应的贸易盈余、储备增长又反过来形成了对汇率升值调整的压力。从这一角度进行政策考察，似乎表明出口退税的政策力度偏大了，适度下调出口退税率也有合理性。

从长远的角度来看，目前较为僵硬的汇率政策需要调整，我国应实行较灵活的汇率政策与出口退税政策配合协调，更好地发挥实际有效汇率对出口的积极有效的调节作用，以适应出口增长的要求。换言之，可逐步放松人民币汇率浮动的幅度，随着货币市场的完善、利率市场化的推进，逐渐实现真正的有管理的浮动汇率制。

参考文献

- [1] Arize, Augustine C., Osang, Thomas and Slottje, Daniel J., Exchange-rate Volatility and Foreign Trade: Evidence from Thirteen LSC S, *Journal of Business Economic Statistics*, 2000, Vol.18, pp10 ~ 17.
- [2] Chi-Chur, Chou W. L. and Eden S. H. Yu, Export Duty Rebate and Export Performance: Theory and China's Experience, *Journal of Comparative Economics*, 2001, Vol.29, pp314 ~ 326.
- [3] William H. Greene, *Econometric Analysis* (fourth edition), Prentice Hall, 1999.
- [4] 董俊英, 姚坚: 《试水免抵退——改革出口退税体制与生产企业实行“免抵退”税的思考》, 《国际贸易》, 2002年第2期。
- [5] 李其谚: 《出口退税难题求解》, 《财经》, 2003年第13期。
- [6] 李万甫, 马衍伟: 《提高出口退税率的政策效应分析》, 《涉外税收》, 2000年第10期。
- [7] 隆国强: 《调整出口退税政策的效应分析》, 《国际贸易》, 1998年第7期。
- [8] 隆国强: 《出口欠退税风险聚积, 证券化破解历史难题》, 国研网 <http://www.drcnet.com.cn>.
- [9] 罗丙志: 《为出口退税而辩》, 《国际商报》, 2001年5月31日, 第5版。
- [10] 牛锐: 《对我国出口退税政策效应的实证分析》, 《西安金融》, 2002年第3期。
- [11] 卓岩: 《零税率是我国出口退税机制改革的必然选择》, 《企业经济》, 2002年第3期。

发展中小企业 亟需改革税制*

杨卫华 尹淑平

一、加快发展中小企业的必要性

在 20 世纪中期以前，由于企业竞相追逐规模效益，大企业盛极一时，不断扩大企业规模成为企业发展的方向。但 20 世纪 50 年代以后许多发达国家逐渐发现发展国民经济离不开中小企业，开始通过制定法律来扶持本国中小企业的发展。如美国在 1953 年通过了保护小企业的基本法《小企业法案》，1982 年又发布了《中小企业创新发展法》；英国先后出台 11 项保护中小企业的法案；日本专门设立管理和推动中小企业发展的中小企业厅；法国、德国也采取了包括税收在内的一系列政策措施来促进中小企业发展。可见，扶持中小企业发展渐成世界各国的共识。发展中小企业既有客观要求也有理论依据。

（一）从理论上讲，发展中小企业有其必要性

对中小企业的存在，国外有多种理论解释，其中最有说服力的是规模经济论。规模经济论的创始人奥斯汀·罗宾逊（Austin Robinson）认为，企业规模越大，管理层次越多，越容易产生官僚主义，导致企业内部的信息传递效率降低，决策效率下降；另外，由于技术非无限可分，分工超过一定限度反而会增加生产成本，降低生产效率。因此，他得出结论，当企业达到一定规模后，会出现规模收益递减，企业存在一个最优规模。大企业未必是企业的最佳选择，中小企业只要规模适度也能生存和发展。

新制度经济学派的代表人物、诺贝尔经济学奖得主罗纳德·科斯（Ronald Coase）进一步发展了规模经济论。他认为企业从市场购买原料等必需品时，除了购买成本外还要负担交易费用，企业进行内部生产则不存在交易费用。但是随着企业规模不断扩大，组织费用会不断增加，企业自然会在交易费用与组织费用的增加之间寻找一个平衡点。这表明，企业存在一个最佳规模，而非越大越好。这一理论不仅为中小企业的存在提供了理论依据，而且为中小企业的发展规模提供了有益判断标准。

美国著名经济学家乔治·施蒂格勒（George J. Stigler）从微观的角度对规模经济理论作了进一步解释和阐述。他假设平均成本曲线为 U 型曲线，认为 U 型曲线底部越平缓，即边际成本上升速度越慢，企业的灵活性越强；反之，企业的灵活性越弱。在一定产量区间内，大企业具有规模经济优势，单位成本较低；但产量未达到这一区间时，大企业会由于资产大量闲置而导致单位成本大幅度提高；而产量远大于这一区间时，大企业会由于超

* 原载：《税务研究》，2003 年第 12 期。

过最大生产能力，重新购置固定设备而导致平均成本大幅度提高。所以，从长期来看，厂商如果能在其短期成本曲线与长期成本曲线相切点或附近区域生产，大企业显然具有成本优势。但是由于市场需求的不稳定性，厂商不可能只在切点附近生产，一旦离开切点附近，大企业单位成本的变动速度大于小企业。因此，小企业更具灵活性，进而赢得竞争优势。

中小企业在世界范围内蓬勃发展，还有一个重要原因就是技术创新给人类社会带来的重大影响。第一次和第二次工业革命给世界带来了巨大的变化，使世界经济高速发展，但第三次技术革命给人类带来的影响却远远超过前两次，使人类在短短几十年间所创造的生产力超过过去几百年的总和。各个国家从中意识到高新技术给人类社会带来的深刻影响，鼓励创新成为各国的一项重要政策。英国学者罗斯韦尔（R. Rothwell）根据大企业和中小企业管理结构不同的特点，从理论上论述了中小企业在技术创新上的优势。中小企业在创新方面的优势逐渐引起各国政府的注意，在 1980 年经济合作与发展组织（OECD）部长级会议上，提出要给予中小企业的创新潜力以特别的关注，中小企业技术创新政策也成为政府创新政策的重要组成部分。

（二）从现实看，我国扶持中小企业发展具有重大意义

据统计，目前我国中小型企业有 791 万家左右，占全部注册企业总数近 99%，工业总产值占全国工业总产值的 60%；中小企业上缴的工商税收占全国工商税收总额的 50%，出口额占全国出口总额的 60% 左右，零售网点占全国网点的 90% 以上，就业人数占就业总人数的 70%，在新增就业机会中占 80%；从地区分布来看，中小企业 42% 分布在东部地区，东部中小企业的工业总产值是全国的 66%。^①由此可见，中小企业在我国国民经济中占有十分重要的地位，中小企业的发展对于推动经济增长、扩大就业、满足人民生活、增加财政收入具有重要作用。

二、我国推动中小企业发展的相关税收法律法规及存在的问题

（一）我国现行税收法律法规对中小企业的优惠规定

根据中小企业的特点，国家制定了包括税收政策在内的扶持政策，推动中小企业的发展。2000 年国家经贸委发布了《关于鼓励和促进中小企业发展的若干政策意见的通知》，要求加大财税政策的扶持力度，规定各类中小企业凡在我国境内投资符合国家产业政策的技术改造项目，可按规定享受投资抵免企业所得税的政策；国有企业下岗职工创办中小企业的，可按国家规定享受减免税收优惠政策。2002 年 6 月 29 日第九届全国人民代表大会常务委员会第二十八次会议通过了《中华人民共和国中小企业促进法》，从资金支持、创业扶持、市场开拓、技术创新、社会服务等几个方面规定了促进中小企业发展的措施。其中，第十二条规定，国家通过税收政策，鼓励对中小企业发展基金的捐赠；第十七条规

^① 李爱萍：《建立我国中小企业法的探讨》，《中国乡镇企业》，2002 年第 5 期。

定，国家通过税收政策鼓励各类依法设立的风险投资机构增加对中小企业的投资；第二十三条规定，国家在有关税收政策上支持和鼓励中小企业的创立与发展；第二十四条规定，国家对失业人员创立的中小企业和当年吸纳失业人员达到国家规定比例的中小企业，符合国家支持和鼓励发展政策的高新技术中小企业，在少数民族地区、贫困地区创办的中小企业，安置残疾人员达到国家规定比例的中小企业，在一定期限内减征、免征所得税，实行税收优惠。这些规定为我国通过税收政策促进中小企业发展提供了强大的法律依据。总之，我国现行税收法律法规对中小企业的优惠规定，主要体现在如下几个方面：

(1) 实行较低的税率。例如，从1998年7月1日起，对商业小规模纳税人的增值税征收率由6%降为4%；对年应纳税所得额在3万元及以下的企业，减按18%的税率征收企业所得税；对年应纳税所得额超过3万元至10万元的企业，减按27%的税率征收企业所得税。

(2) 对中小企业融资担保业务收入免税。例如，对纳入全国试点范围的非营利性中小企业信用担保、再担保机构，可由地方政府确定，对其从事担保业务取得的收入，3年内免征营业税。

(3) 鼓励就业的税收政策。例如，新办的城镇劳动就业服务企业，当年安置待业人员超过企业从业人员总数60%的，经主管税务机关审查批准，可免征所得税3年。对劳动就业服务企业中的加工型企业和街道社区具有加工性质的小型企业实体，凡安置下岗失业人员并签订一年以上劳动合同的，经劳动保障部门认定，税务机关审核，自2003年1月1日起，每吸纳1名下岗失业人员，每年可享受企业所得税2000元定额税收扣减优惠。当年不足扣减的，可结转至下一年继续扣减，但结转期不能超过两年。

(4) 鼓励第三产业和乡镇企业发展的税收政策。例如，新办的独立核算的从事咨询业、信息业、技术服务业的企业或经营单位，自开业之日起，第一、二年免征所得税。

(二) 相关税收政策存在的问题

(1) 税收法律级次不高，稳定性差。我国现行大多数税种都是依据行政法规开征的，而且几乎所有税种都是小条例（法律）、大细则，同时变动过快，稳定性不强，透明度不高。减免税规定大多采用实施细则、临时通知或补充规定的形式，没有形成完整而稳定的法律规定，因而人为因素多，随意性大，致使不少减免税规定被滥用，达不到应有的效果。

(2) 没有建立较为系统、规范的中小企业税收制度。现行税收制度虽然看起来不少与中小企业有关，但总的来说对中小企业针对性不强，税收政策目标不够明确，缺乏与实施《中华人民共和国中小企业促进法》相配套、针对性强且较为系统的中小企业税收制度。

(3) 优惠形式单一，优惠力度不足。我国税收优惠的主要形式是减免税和优惠税率，其他形式较少。实行的基本上是直接优惠，间接优惠较少。同时，我国中小企业税收优惠的力度很小，如企业所得税，对年应税所得额超过3万元至10万元之间的企业税率为27%，仍然偏高。在加拿大，对中小企业创办初期的所得税税率优惠为13.12%，比正常税率29.12%优惠了16个百分点；在英国，年产出低于30万英镑的小企业公司税率为20%，比大企业适用税率30%低了10个百分点。另外，我国对中小企业的创新鼓励不够，

如对新产品的广告费扣除没有规定合理的比例，影响了新产品的推广和研制。

(4) 增值税制不完善，不利于企业发展。一方面，增值税负担过重，影响了中小企业的发展。现行税制规定小规模工业企业适用征收率为 6%，商业企业为 4%。虽然商业企业税收负担比原来降低了 2 个百分点，降低的幅度不小，但实际税负仍高于一般纳税人。众所周知，工业企业 6% 的征收率是按收入额中有 35% 的增值额计算出来的，商业 4% 的征收率是按 23.5% 的增值额换算的，实际上一般工业企业的销售收入不可能有 35% 的增值额，商业企业一般增值额也高不过 20%。从我国实际征收的情况来看，1994 年我国独立核算的工业企业的产品销售收入为 423999 亿元，全国商业销售总额为 33178 亿元，同年的增值税收入为 2305 亿元，增值税的平均税负为 3.05%，^① 低于小规模纳税人 6% 和 4% 的水平。另一方面，生产型增值税不利于中小企业的技术更新和出口退税。我国实行生产型增值税，外购固定资产所含税金不予抵扣，造成固定资产价值重复征税，不利于鼓励中小企业投资，不利于企业技术创新和先进设备的采用，不利于提高劳动生产率。

三、进一步完善税收制度，促进中小企业发展

(一) 健全税收法律法规，提高税法级次

通过税收立法来保障中小企业的发展。必须健全税收制度，完善税收法规，提高税法级次，保证税法的稳定性、严肃性和权威性，减少执法的随意性，更好地发挥税收政策对中小企业发展的促进作用。

(二) 建立系统的税收法律体系，促进中小企业全面发展

根据《中华人民共和国中小企业促进法》在“税收政策上支持和鼓励中小企业的创立和发展”的规定，应在总结我国税收分配实践经验的基础上，借鉴外国扶持中小企业发展的税收政策，制定内容系统、针对性强的促进中小企业发展的税收法律或行政法规，发展税收调节作用，促进中小企业发展。

(三) 调整税率，为中小企业发展减负

降低企业的税收负担尤其是中小企业所得税负担是世界税制改革的趋势。根据我国的实际，应将增值税的税率由 17% 和 13% 降为 15% 和 11%，或原 17%、13% 的税率不变，但把生产型改为消费型，消除重复征税，降低税收负担。对小规模纳税人的征收率，工业企业应由现在的 6% 调到 3.5%（即按 23.5% 的增值额换算），商业企业由 4% 调到 2.5%（即按 17% 左右的增值额换算）。同时，在合并内、外资企业所得税的基础上，将企业所得税的税率由 33% 调到 25%，并对年应税所得额 10 万元以下的小型企暂时实行 18% 的税率。此外，对技术密集型的中小企业，其计税工资每月最高扣除额可适当提高至 1500 元，并视不同地区的消费水平，上下浮动 30%，以降低该类型企业的税收负担。税率下

^① 财政部财政科学研究所：《中国财经改革研究报告（1996～1997）》，经济科学出版社 1999 年版。

调带来的税收收入减少，可通过发展经济，扩大税基，减少税收优惠，杜绝随意减免、越权减免，加强征管等途径来弥补。

（四）制定科学的税收优惠制度，鼓励中小企业扩大生产

（1）鼓励中小企业增加投资的税收优惠。①为引导中小企业调整、优化产品结构，政府在税收制度上，应对生产“精、灵、特、优”产品的中小企业给予5年免征所得税的照顾，让中小企业朝“小而精，小而专，小而优”方向发展，增强中小企业的竞争力。②对企业用于科学研究和技术开发支出，允许在所得税前列支。同时对中小企业之间转让科研成果，给予一定时期（如三年）免征所得税照顾。此外，在中小企业及时、足额纳税的前提下，允许其在增长额中提取一定比例创建中小企业科技风险投资基金，专门用于中小企业的科学研究和技术开发，并对应用结果进行检查评估。另外，对中试产品给予免税。③对投资于高新技术等行业的中小企业，允许企业对固定资产实行加速折旧，以激励企业更新生产手段。④对中小企业转让带来的损失，购买企业可在5年内从应纳税所得额中扣减，以促进资本的快速自由流动，使中小企业能够迅速增加资本，扩大规模。⑤对新开办的符合国家产业政策的中小企业，从开办之日起，给予免征企业所得税3年。⑥借鉴国外经验，凡在欠发达地区（如我国西部地区）投资创办的中小企业，除从获利年度起“两免三减半”优惠外，在期满后5年内，再给予应纳税款30%~50%的减征。

（2）有利于中小企业融资的税收优惠。①为了解决我国中小企业的融资困难，国家除了对纳入全国试点范围的政策性中小企业信用担保、再担保机构，由地方政府确定对其从事担保业务的收入，3年内免征营业税外，还可对将所得税后利润用于再投资的中小企业，准予将再投资额已纳企业所得税税款的40%，抵免新增利润应缴纳的企业所得税税款。②为鼓励中小企业增加自有资金（不含本企业的利润作为注资），扩大企业生产规模，提高企业效益，对增加自有资金的中小企业，可在一定期间内给予一定比例的减、免税或降低适用税率的照顾，以引导民间资本向企业注资。

（3）引导投资方向的税收优惠。对直接为农业生产特别是粮食、油料等种植业服务从事农业产品深加工的中小企业，以及利用废气、废水、废渣等废弃物为主要原料的环保型中小企业，除了一般的减免税规定外，还应增加一定年度（如5年）的减免税优惠，以引导中小企业投资。

（4）促进就业的税收优惠。为了扩大就业，除保留现行对民政部门举办的福利企业、城镇劳动就业服务企业，安置失业人员达到一定数量和比例的给予一定时期的免征或减半征收企业所得税外，可考虑对所有中小企业，安置失业人员达到一定比例和一定数量的均给予一定时期的减免企业所得税优惠。

此外，在规范中小企业税收制度的前提下，应清除对中小企业的无理收费，规范税费征收，真正减轻中小企业的负担。同时，对收入比较少、所得比较少（如年所得额2万元以下）的小企业，应允许企业选择纳税方式，可以选择缴纳个人所得税，或缴纳企业（法人）所得税。

参考文献

- [1] 高金德：《中小企业与中小企业理论研究》，《华东经济管理》，2000年第8期。
- [2] Coase, Ronald H. , The Nature of the Firm, Economics, 1937, No.4.
- [3] Stigler, G. , Production and Distribution in the Short Run, Journal of Political Economy 47, June 1939.
- [4] 万兴亚：《中小企业技术创新与政府政策》，人民出版社2001年版。
- [5] 李爱萍：《建立我国中小企业法的探讨》，《中国乡镇企业》，2002年第5期。

优化纳税服务与交易费用的降低*

郭小东 王 倩

随着我国社会主义市场经济体制建设的不断推进和公共财政框架构建的不断完善，我们对税收征收管理的认识也在发生着积极的调整，纳税服务在整个税收征收管理制度安排中的地位正得到不断抬升。这种制度调整要求我们在理论认识上也应作出积极的跟进，形成更深入而全面的认识。本文将围绕纳税服务与税收征收过程中交易费用的关系展开分析。

一、纳税服务与交易费用

纳税服务是税收征管工作中的重要环节，是税务部门在税收征管过程中向纳税人提供的旨在方便纳税人履行纳税义务和享受纳税权利的各种服务的集合。

在过去相当一段时期里，税务部门主要侧重于保证国家财政收入，强调税收执法的刚性，税务人员也主要以执法者的身份出现在纳税人的面前，而对如何在依法治税的前提下更好地为纳税人服务，则显得重视不足。

随着政府施政中把“三个代表”重要思想和“立党为公、执政为民”治理理念不断落实到公共部门运行的各个方面，税务部门对纳税服务的重视程度在不断提高。2001年5月1日开始实施的新《税收征收管理法》就在强调纳税人义务和责任的同时，更加注重保护纳税人合法权益和纳税服务。2003年国家税务总局又下发了《关于加强纳税服务工作的通知》（国税发〔2003〕38号文），对纳税服务工作的有关问题进行了明确。纳税服务越来越成为税务工作的一个重要组成部分。国家税务总局局长也在2004年4月召开的“依法诚信纳税，共建小康社会”座谈会上要求，各级税务机关要进一步推进依法治税，强化科学管理，优化纳税服务，营造依法诚信纳税的社会环境，促进税收事业的发展，为全面建设小康社会作出新贡献。

目前，国税系统开展的“一窗式”管理模式就是税务工作向“管理服务型”迈出的的一大步，是加强纳税服务的具体体现。此外，各地出现的办税服务厅、首问责任制、低柜式服务、坐台咨询、表票供应、咨询辅导，以及多元化申报方式、电子缴款、12366呼叫中心等等，都是优化纳税服务的积极措施。

为有效提供纳税服务，税收部门必须做出相应的人力、物力投放，与之密切相连的，就是交易费用的变化。

* 原载：《税务研究》，2004年第8期。

交易费用 (transactions costs) 是指经济活动中因交易行为而产生的成本或费用^①。交换关系是经济活动中广泛存在的关系, 交易行为也是经济活动中广泛存在的行为。交易费用理论中所涉及的交易, 既包括人们通常所说的市场交易, 也包括现实中存在的企业内部的交易、政府 (公共) 部门内部的交易和公共部门与非公共部门之间的交易等其他类型。税收征纳过程中发生的征收成本、奉行成本等, 都是交易费用的具体表现。毋庸说明, 税收部门为提供纳税服务而投放的全部人力、物力、资金、时间等, 也均属交易费用范畴。

从总体上说, 纳税服务的加强, 不但会引起在服务项目上所作投放的变化, 而且还会导致整个税收征收体系中交易费用的变化。并且, 因纳税服务而造成的整个税收征收体系交易费用的变化不是一种单向的、简单的变化。其中, 有些费用会因此而增加, 但有些费用则会因此而减少。所以, 纳税服务引致的交易费用变化是一种多维的、较复杂的变化。

为便于分析说明, 我们把税收征纳过程中发生的全部交易费用, 按照与纳税服务的不同关系, 划分为三个大类:

第一类, 会因优化纳税服务而增加的交易费用 (用 F_1 表示), 这类费用主要由为提供纳税服务而投放的人力、物力所构成, 如税务机关为纳税人提供良好办税环境而支付的装修成本等。

第二类, 会因优化纳税服务而减少的交易费用 (F_2), 如税务机关为简便纳税而减少纳税环节, 从而为征纳双方节约的成本; 征纳双方加强沟通后, 纳税人对税务机关的满意程度提高, 通过减少纠纷而节约的诉讼成本等。

第三类, 与优化纳税服务无显著相关性的交易费用 (F_3), 如税务机关正常办公时所发生的经费等。

将这三类交易费用加总, 就可以得到税收征纳过程中总的交易费用, 即:

$$F = F_1 + F_2 + F_3$$

根据这一关系我们可以看出, 尽管纳税服务对税收总交易费用的影响是多维的, 但对这些影响, 我们还是可以做出综合性评价的, 这就是看纳税服务对总交易费用 (F) 的最终影响。这种最终影响, 实际上取决于 F_1 和 F_2 的对比。若纳税服务所造成的 F_1 的上升大于其所导致的 F_2 的下降, 则纳税服务在总体上就造成了 F 的上升; 反之, 如果 $F_2 > F_1$, 则纳税服务在总体上就造成了 F 的下降。

二、优化纳税服务对降低交易费用的促进作用

众所周知, 用尽可能小的成本按质、按量地完成法定税收征收任务, 是税收征管过程中追求的重要目标; 完成同一税收征收量所费成本的多少, 是衡量征管体系效率高低的根本指标。因此, 尽可能降低征收过程中的交易费用, 是提高征管水平的重要方面。

要降低征纳过程中发生的交易费用, 从纳税服务的角度讲, 一方面, 就是要在保证服

^① 斯蒂格利茨:《经济学》, 中译本, 中国人民大学出版社 1997 年版, 上册第 65, 448 页。交易费用迄今尚无真正统一的定义, 本文所用的是关于交易费用最一般性的定义。

务质量的前提下,控制好 F_1 的规模;另一方面,就是要发挥好纳税服务对 F_2 的拉高作用。对 F_1 的控制,尽管意义重大,在具体施行中也会面临许多挑战,但由于其内容相对较为直观,无需进行过多说明,因此,本文把通过纳税服务降低交易费用中的 F_2 作为分析说明的重点。

我们认为,优化纳税服务可从如下几个方面起到提高交易费用 F_2 的作用:

(1) 优化纳税服务,有利于降低税收征管过程中大量与信息沟通有关的交易费用。

交易费用是个涵盖范围非常广泛的概念。例如,按威廉姆森的划分^①,交易费用可分为事前、事后两个部分。事前交易费用包括起草、谈判和维护相关协议的成本。事后交易费用则包括:①当交易偏离了所要求的准则而引起的不适应成本;②交易双方为努力纠正事后偏离而导致的争执协商的成本;③如果交易纠纷需要管理机构(不一定是法庭)的介入来解决,伴随这种机构的建立和运作而来的成本;④为使承诺保证生效的约束成本。从威廉姆森的划分中就可看出,交易费用中的相当部分都与交易双方的信息沟通、意愿磨合、认识统一有关。通过加强纳税服务,就可有效降低因信息不对称、征纳双方的摩擦等原因而导致的交易费用,这与我们常说的“人和政通”、“家和万事兴”的道理是一致的。

(2) 纳税服务作为一种具有特定效用的公共产品,有着很强的正外部性,可在更广泛的范围内发挥降低交易费用的作用。

纳税服务作为公共部门提供的一项公共产品,具有很强的正外部性,能够有效强化征纳双方的合作关系。优化纳税服务,可以提高纳税人的地位,在全社会范围内形成对纳税人的尊重与关怀的氛围。这样,纳税人履行纳税义务就不再是简单的被动纳税行为,更重要的是体现其社会地位和价值,使纳税人产生一种社会责任感和使命感,在谋求自身发展的同时更注重国家和社会利益的实现与维护。因此,优化纳税服务,可以缓和征纳双方之间的矛盾,改变纳税人对缴纳税款的抵触情绪,提高其自觉纳税的意识,从而减少税收征纳过程中的逃税、漏税现象,进而降低在检查、强制执行等方面产生的交易费用。

(3) 优化纳税服务,对交易费用的长期降低有积极意义。

优化纳税服务,虽然会导 F_1 的上升,但 F_1 中相当部分的上升相对于 F_2 的提高来说,是次要的、短期的;而 F_2 的提高则是主要的、长期的。例如,通过培训提高税务人员的服务意识,虽然在短期内造成了 F_1 的上升,但是却可换来长期生效的优质服务,从而产生促使 F_2 长期提高的效果。因此,优化纳税服务可以成为降低税收征管过程中交易费用的一条长期有效的途径。通过简化税收征管的办事程序、对各项税收法律法规的及时宣传、加强与纳税人沟通以减少征纳双方矛盾等手段,就可以在降低税收征纳交易费用中发挥持续的作用。

(4) 优化纳税服务,对降低税收成本中的奉行成本有突出意义。

虽然税收成本是税收征纳中交易费用的组成部分,但从税收学角度说,它毕竟是一个颇为重要的独立概念,所以我们从降低税收成本,特别是从奉行成本的降低方面对纳税服务的积极意义加以说明。

① O. E. Williamson, *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*, Macmillan, 1985.

概括地讲，税收成本是指政府在征税过程中发生的征收成本、纳税人的奉行成本以及课税对经济运行产生的负效应之和。在整个征税成本中，优化纳税服务对于奉行成本降低的贡献特别值得注重。目前，我国研究税收成本基本上是从最直接的征收成本着眼，而对于奉行成本的研究相对较少。其实从国外统计资料看，奉行成本往往高于征税成本。如 20 世纪 80 年代后期，英国公司税的纳税成本为 2.22%，而征税成本仅为 0.52%^①。因此，通过优化纳税服务来降低纳税成本，对于整体税收成本的降低有着极其重要的作用。例如，优化纳税服务中的简化纳税环节、加强税法宣传、尊重纳税人权利等举措，都可以有效降低奉行成本，从而带来整个税收成本的明显下降。

(5) 优化纳税服务，有利于抑制税务机关的 X- 非效率。

X- 非效率是经济运行的低效率，其最直接表现就是铺张浪费、人浮于事。税收过程中存在的大量交易费用，都与 X- 非效率直接相关。

政府机关之所以会产生 X- 非效率，往往是由其所处的非竞争性环境决定的。在垄断的“公共品市场”，政府机关客观地得到了免于竞争压力的庇护场所，机关工作人员身上没有外部竞争压力，也就缺乏追求成本极小化的动力，在这种情况下，政府机关的浪费现象就较难避免。

政府机关 X- 非效率存在的程度，取决于约束、激励机制的设置状况，也就是说，取决于约束、激励机制能否有效排斥铺张和浪费。而优化纳税服务实际上就是促使税务部门排斥浪费的一个重要的约束机制。这是因为，优化纳税服务的各项措施，必然会体现在提高纳税人地位、提高税务部门运行的透明度等方面，这些改变，又必然会使纳税人享有的监督税务机关工作的权利得到更充分的发挥，从而形成对税务工作人员减少浪费、降低成本的积极压力，最终实现对 X- 非效率的有效抑制。

三、纳税服务与交易费用关系的数学模型

由于优化纳税服务可以通过提高纳税人对政府机关的满意程度等，在总体上降低税收征纳过程中的交易费用，特别是可以降低纳税人的奉行成本，这就可以在一定程度上对减少税收流失产生积极影响。这一结论可以在阿林汉姆（Allingham）和桑德莫（Sandmo）在 1972 年提出的 A-S 模型的基础上稍加改动后得出。以下进行具体论证。

为方便分析，我们做如下假设：①作为纳税人的个人是个风险厌恶者，符合理性经济人的假设，他以个人预期效用最大化为目标。②纳税人的效用仅由其可支配所得决定，不受其他因素影响。③税率（ t ）为固定比例税率。④纳税人在纳税过程中需支付交易费用，且该费用与纳税额呈正比。⑤逃税被发现后处罚的数额为纳税人偷逃税款的数倍。在这五项假设下，纳税人的期望效用函数是：

$$EU = (1 - P) \times U[\omega - tx - tf] + P \cdot U[\omega - tx - tf - \pi(\omega - x)] \quad (1.1 \text{ 式})$$

其中： P 为税务机关稽查偷逃税的概率； ω 是纳税人在某一时期的真实税前收入； x 为纳

^① 王磊，张瑞环，潘皓：《降低税收成本提高税收效率》，《辽宁经济管理干部学院学报》，2000 年第 2 期，第 7 ~ 10 页。

税人的申报纳税额； f 为纳税人在纳税过程中需支付的交易费用； π 代表逃税被发现后的罚款倍率； U 代表纳税人的效用函数； EU 为预期效用。

只有当纳税人预期能从偷逃税中得到的效用大于按照法律规定依法申报纳税的效用时，纳税人才会偷逃税。因此，纳税人作出偷逃税选择的前提条件是纳税人的目标效用函数关于 x 的一阶导数小于零。即：当纳税人每增加单位的申报额，预期效用就会因此而减少；反之，如果纳税人每减少单位的申报额，其预期效用会因此而增加，在这种情况下，纳税人会选择偷逃税。对 (1.1 式) 的 x 求导数，可得到：

$$\frac{dEU}{dx} = - (t + f) \cdot (1 - P) \cdot U[\omega - (t + f)x] - (t + f - \pi) \cdot P \cdot U[\omega - (t + f)x - \pi(\omega - x)] < 0$$

在 $x = \omega$ 处求解，即有：

$$(t + f)(1 - P) + (t + f - \pi) \cdot P > 0$$

也就是：

$$P < \frac{t + f}{\pi} \quad (1.2 \text{ 式})$$

根据这个模型结果 (1.2 式)，我们不难发现，通过优化纳税服务来提高纳税人对税务机关的满意程度，进而实现对纳税人交易费用 (f) 的降低，就可以在不影响纳税人对自愿缴纳税收 (x) 的选择的情况下，降低税务机关稽查偷逃税的概率 (P)，或者说，在保持相同稽查偷逃税概率 (P) 的前提下，提高纳税人自愿缴纳税款 (x) 的取向。

四、优化我国纳税服务的主要措施

优化纳税服务，是我国税收征管工作发展的必然趋势。在现有基础上，我国的征管工作应怎样进一步把优化纳税服务落到实处，我们认为应在以下几个方面采取积极措施：

1. 切实转变观念，进一步加强征管中的服务意识

如果说，过去的税收征管工作是以对“税收三性（强制性、确定性、无偿性）”的认识为基础的话，那么，随着纳税服务的加强，我们对税收的认识就应做出适当的调整，适度地把“税收自愿交易说”引进我们对税收的认知体系之中。

“税收自愿交易说”的典型理论是马斯格雷夫的税收价格论。在庇古、林达尔等人所作研究的基础上，马斯格雷夫于 1938 年发表的“The voluntary exchange theory of public economy”一文，较系统地阐述了公共经济运行中的自愿交换论观点。他认为，税收是私人经济部门为消费公共产品而向公共经济部门支付的价格。这种税收有偿的观点实际是认为公共产品的提供和费用补偿之间，存在着市场式的等价交换关系，即私人经济部门应向政府公共经济部门缴纳税收，而政府公共经济部门则应向私人经济部门提供公共产品。

在公共财政的自愿交换或价格理论的基础上，税收适用受益原则，即个人根据自身偏好对公共产品作出评价，并根据个人的边际效用来付款；政府则根据所提供公共产品的成本和个人的边际效用（效益）来征收赋税。

在税收价格论的理论框架下，政府和纳税人之间的关系实际上就是“交换关系”：一

方是公共产品的“消费者”，另一方是公共产品的“供给者”。而根据消费者主权理论，经济体系的主导者是消费者的需求，强调消费者应当得到适当而正确的保护。因此，政府及税务机关强调纳税服务，在一定程度上就是“注重顾客利益”的体现。

2. 合理借鉴新公共管理运动形成的相关经验

在改变思想观念的基础上，我们还可以从实践角度看看市场经济发达国家在税收征管体系中关于纳税服务的制度变化趋势。

我们可以从新公共管理运动形成的趋势上看。

20世纪70年代末80年代初，伴随着经济上新自由主义思潮的兴起和信息技术革命的浪潮，西方国家逐渐掀起了一场颇具规模的公共行政改革运动。这场主张运用市场机制和借鉴私营部门管理经验提升政府绩效的改革运动，被理论界称为新公共管理（New Public Management, NPM）运动^①。新公共管理运动首先发端于英国，此后波及澳大利亚、新西兰、美国、加拿大、荷兰、法国等众多国家。进入90年代以后，亚洲的日本及一些新兴工业化国家，如韩国、菲律宾等也开始加入这一公共行政改革的浪潮。

新公共管理运动的兴起对西方乃至世界各国公共管理理论与实践带来深刻的影响，促使传统的韦伯官僚制模式向以市场与服务为导向的政府管理模式转变。同时，它对西方传统的税收管理理念和运作方式也产生了深刻的影响，带来了四个方面的转变：

第一，税收管理从规制导向向服务导向转变。新公共管理以理性经济人的假设前提取代传统人性恶的假设前提，这一点给西方国家税收管理理念带来深刻影响，促使税务机关从过去视纳税人为偷税者处处设防，转变为重视对纳税行为的成本收益分析，通过提高服务质量、降低税收征收成本来提高纳税遵从率。正如前文所说，一份纳税服务方面的努力，可以相当于50份税务监督打击方面的收获。有鉴于此，美国联邦税务局于1998年实施了重组和改革法案，把“通过帮助纳税人了解和实现他们的纳税义务，并通过使税法公平适用于所有纳税人，为美国纳税人提供最高质量的服务”作为美国税务管理现代化的使命。

第二，纳税人从监管对象向商业客户转变。新公共管理主张将企业家精神引入政府管理之中，运用私营企业成功的商业化技术、手段和经验，为公众提供高品质的公共产品。不少国家把税务局的名称改成了service，如美国国税局（IRS）其实就是Internal Revenue Service（直译为国内收入服务机构），把纳税人（taxpayer）改成客户（client），把税收行政管理（tax administration）改成经营（business）。

第三，内部管理从行政管理向企业化经营转变。新公共管理主张通过建立激励机制和实行绩效评价，激励官员提供更优质的服务。如新加坡国会通过法律，同意税务局像承包企业一样经营，并允许在承包费中按中产阶级的收入标准确定税务人员的待遇；美国联邦税务局则建立了平衡业绩评价体系，依据纳税人满意度和经营业绩等对员工进行绩效评估，以此作为决定薪酬和升迁的依据。

第四，纳税服务从行政服务向社会化服务转变。新公共管理认为，要改善官僚制的运转效率，其关键在于消除公共垄断，在公共部门恢复竞争、引入市场，促进公共管理的社

^① 谢少华：《从新公共管理的兴起看西方税收管理理念的转变》，《涉外税务》，2002年第4期，第16~19页。

会化。像美国、英国等国家就充分利用中介组织和社会团体为纳税人提供大量的、优质的社会化纳税服务，从而大大降低了征税成本。

可见，在新公共管理运动的推进中，纳税服务在征管体系中的地位得到了明显的上升。

尽管我国国情与西方国家有较大差距，不能完全照搬西方的做法，不能全面套用外国新公共管理运动的做法，但随着改革开放的进一步深入，新公共管理运动所蕴涵的市场化、效率和顾客至上的理念也应该成为我国新一轮税收管理体制改革中加以积极借鉴的内容。

3. 优化纳税服务应与法制建设的完善同步进行

一方面，随着纳税人对于维护自身权益的愿望日益强烈，对纳税服务的要求也越来越高，应该用法的形式把对纳税服务的要求固定下来，取代当前各种临时性、应景式、不规范、形式主义的做法。因此，应按照国际惯例，在充分酝酿和广泛征求意见的基础上，适时制定出台《纳税人权利法》，建立健全纳税服务法律体系，提高纳税服务法律级次，使纳税服务真正从职业道德和思想政治工作的范畴变为税务机关应尽的法律义务和法律行为，实现纳税服务的制度化。

另一方面，其他税收方面的立法、执法也应加强。如尽快制定《税收基本法》，明确我国税收的立法原则、立法权限，明确纳税人、税务机关和社会各部门的权利和义务等基本法则，特别是要明确纳税人的隐私保密权、信息公告资料知悉权、税法适用公正权等基本权利等等；规范税收执法行为，在加大依法治税力度的基础上，健全税收执法权约束机制，增加税务机关行使权力的透明度，自觉接受全社会和纳税人的监督，使税收执法活动在“阳光”下运作，使纳税人在充分享有权利的同时有效地履行自己的纳税义务。

参考文献

- [1] (英) C. V. Brown, P. M. Jackson: 《公共部门经济学》(第四版), 中国人民大学出版社 2000 年版。
- [2] 郭小东: 《论公共部门运行中交易费用合理的“度”》, 《财政研究》, 2003 年第 5 期。
- [3] 谢少华: 《从新公共管理的兴起看西方税收管理理念的转变》, 《涉外税务》, 2002 年第 4 期。
- [3] 国家税务总局征收管理司纳税服务处: 《纳税服务: 中国税收征管新战略》, 《中国税务》, 2002 年第 12 期。

国有资产管理改革方向*

林 江 王雅莉

我国的国有资产管理改革，从目标和手段两个方面都和国有企业绩效紧密相联。就目标而言，国有资产管理委员会（以下简称“国资委”）的成立，进一步明确了我国国有资产管理体制改革的目标：既要保持对国有资产的所有权和控制权，又要使国有资产运营充满活力，使之保值增值，而国有企业经营绩效的好坏直接决定了国有资产能否保值增值；就手段而言，继续推进国有经济的布局调整和国有企业的公司化改制仍是国资委的基本职能和改革方向，那么在国有经济布局的调整过程中，国有资产该不该退出、退出的程度如何、采取什么方式退出都是应加以慎重考虑的问题。是沿用前期改革思路，还是转移改革重点，也应该基于对国有企业绩效的实证分析作出选择：有利于国有企业绩效提高的就保留，不利的就退出或者是进行改革。由此可见，国有资产管理改革的方向、方式和重点都体现于对国有企业的绩效建设。因此，本文意图通过国有企业绩效各影响因素的回归分析对国有资产管理改革的方向进行探讨。文章共分三个部分：第一部分综合分析，回顾相关文献；第二部分提出研究设计及得出研究结果；第三部分为研究结论及政策建议。

一、综合分析及相关文献回顾

如何判断国有企业绩效一直是学术界争论不休的问题。绩效是业绩和效率的统称，包括活动过程的效率和活动的结果两层含义。经营业绩是指经营者在经营管理企业的过程中对企业的生存与发展所取得的成果和所作出的贡献；管理效率是指在获得经营业绩的过程中所表现出来的盈利能力和核心竞争能力。我国现行的企业绩效评价体系是1999年6月由财政部、国家经贸委、人事部和国家计委联合颁布的，包括8项基本指标、16项修正指标和8项评议指标。其中，8项基本指标和16项修正指标是财务指标，8项评价指标是非财务指标。非财务指标包括企业领导班子基本素质、产品市场占有率、服务满意度、基础管理水平、在岗员工素质状况、技术装备更新水平、服务硬环境、行业或区域影响力、企业经营发展策略、长期发展能力预测，由于没有规定计量方法，只能通过定性分析得出结果，在对具体企业作分析比较的时候很难得出客观的结果，因此本文将视角主要投于财务指标。

国内学者在20世纪90年代中期提出了“中国国有企业非效率论”，认为改革以来国有企业财务状况日益恶化，亏损额和亏损面越来越大，与非国有企业相比处于严重不利的竞争地位。因此，我国对国有企业进行了战略性的改革，如何评价改革成效成为影响国

* 原载：《财政研究》，2004年第4期。

企业改革方向的重要因素。表 1 中给出了从 1997 年到 2000 年我国国有企业总体绩效指标比较。从表中内容来看,国有企业自身各项主要经济指标基本保持稳定中小幅度增长,说明我国国有企业经过一系列的改革之后情况稳定中有所好转。

表 1 1997~2000 年国有企业总体绩效指标

单位:亿元

	主要指标			
	1997 年	1998 年	1999 年	2000 年
盈利面 (%)	34.1	31.3	46.5	49.3
资产负债率 (%)	67.1	65.5	65.4	66
总资产报酬率 (%)	2.3	2.1	2.7	3.3
净资产利润率 (%)	1.7	0.4	2.1	4.9
销售利润率 (%)	1.2	0.3	1.7	3.8
流动比率 (%)	100.7	99.5	99.8	105.2
获利倍数	1.2	1.1	1.4	1.9
不良资产占权益比重 (%)	22.6	24.8	27.5	31.4
国有资产总额	44 340.2	48 051.6	53 306.0	57 554.4

资料来源:中华人民共和国财政部网站 <http://www.mof.gov.cn>。

要在众多举措中找出与国有企业绩效高度相关的决定性举措,则要求更加严格的定量回归分析。由于我国国有企业数量众多,信息披露制度又不甚完善,想采用所有的指标,对所有国有企业进行全面完整的绩效分析是不可能的。由于我国证券市场的上市公司多为国有企业且有较好的披露制度,数据来源相对具有完整性、全面性和代表性,因此我们可以从上市国有企业绩效分析管中窥豹,透视我国国有企业绩效状况全貌。从国内外文献研究成果来看,主要采用两大类型的指标进行回归:绝对指标和相对指标。绝对指标主要包括净资产收益率(ROE)、主营业务利润率(OPE)、每股收益;相对指标则主要将以上指标和股东权益进行加权处理以消除经理人控制的因素。由于我国股市经理人持股规模较小,我们采用相对简化的绝对指标,同时由于 ROE 是反映企业收益能力的国际通用指标和杜邦模型中的核心指标,综合能力很强,所以本文以 ROE 为衡量国有企业绩效的主要指标。

学术界对影响企业绩效因素的研究颇多。首先是股权结构与企业绩效之间的关系。Jensen 和 Meckling (1976) 认为,提高对企业有控制权的内部股东的股权比例,能有效产生管理激励,降低代理成本,提高企业价值。McConnell 和 Servaes (1990) 认为公司价值是其股权结构的函数,其经验结果表明 Tobin 的 Q 值与企业内部人持有股份之间具有曲线关系,当内部股东的持股比例从无到有并逐步增加时, Tobin 的 Q 值随其不断上升,并在内部股东持股比例达 40%~50% 时实现最大,然后开始下降。国外研究表明,股权结

构与企业绩效之间存在着区间效应，合理的股权结构能为企业带来更高的价值。我国学者在这一领域也做了不少研究，许小年（1997）对沪、深两市上市公司的经验研究结果表明，国有股比例越高的公司，其绩效越差；法人股比例越高的公司，其绩效越好。何浚（1998）分析了上市公司中的内部人控制问题，其经验结果显示，国有股在公司总股本中所占的比例越大，公司的内部人控制就越强。

其次是激励与企业绩效的关系。Coughlan and Schmidt（1985）的检验结果表明，经理报酬与企业业绩存在正相关性。然而魏刚（2000）的经验证据表明，公司高级管理人员的持股数量与企业绩效不存在“区间效应”，高级管理人员的报酬水平与企业规模存在显著的正相关关系，与其所持股份存在负相关关系，并受所处行业景气度的影响。可以说，在此领域，学术界尚未形成一致的观点。

再次，是对行业竞争因素的研究。Mark L. Defond 和 Chul W. Park（1999）发现高层管理人员的变动率与行业的竞争性水平正相关，从而对绩效评价产生影响。陈晓、江东（2000）认为，不同行业中影响企业绩效的因素不同，股权结构的多元化对公司业绩的正面影响取决于行业的竞争性。在提高行业竞争性的基础上，通过适当减持国有股比例，提高法人股和流通股的比例，将能改善公司的治理结构。

上述这些论文普遍没有考虑影响企业绩效的其他非股权因素，而且在进行回归分析时模型过于简单，方法也值得推敲。本文试图将以上几项国内研究未考虑的因素综合纳入模型内进行实证研究。

二、研究设计及研究结果

（一）研究假设

1. 规模因素

规模对企业绩效的有利之处可能体现在规模效应，雄厚的资金和良好的技术保障对于企业发展可能起到促进作用，其不利之处可能是导致机构庞大、管理缺乏效率。因此，我们提出第一个假设：企业规模与国有企业绩效具有正相关关系。

2. 股权结构

国有股比例的提高意味着获得政府保护，享受税收优惠的可能性上升，但在总体上内部人控制和由此引发的道德风险应对企业的绩效产生更大的负面影响，同时国家及其代理人具有特殊地位，很可能导致对企业过多的干预。因此，我们第二个研究假设是：国有股比例与国有公司绩效负相关。

目前的法人股虽然不能上市流通，但可以根据协议转让。且法人股都是倾向于长期投资，容易对经营者进行监督。Jensen 和 Meckling（1976）就持这种观点。因此，我们提出第三个研究假设：法人股比例与国有公司绩效负相关。

3. 激励因素

国有企业绩效是否也与高层管理人员的激励有关，我们提出第四个假设：高层管理人员薪酬与国有企业绩效正相关。

4. 时间因素

企业经营时间是否影响企业活力。我们提出第五个假设：企业经营时间与国有企业绩效正相关。

5. 行业因素

行业因素对于国有企业经营管理的绩效是否存在影响对于国有企业的产业分布政策具有重要的意义。因此，我们提出第六个研究假设：国有企业绩效受行业性因素影响。

(二) 模型设计

由于本文数据同时涉及时间序列与截面数据，因此采用了“panel data”的分析方法。基本计量模型为：

$$ROE_{it} = \alpha + \beta_1 \cdot X_{it} + \beta_2 \cdot TA_{it} + \beta_3 \cdot PAY_{it} + \beta_4 \cdot YEAR_{it}$$

其中：ROE 代表 i 公司在 t 年的业绩； X_{it} 为代表国有股（GYG）和法人股（FRG）股权结构的解释变量； TA_{it} 代表规模因素； PAY_{it} 代表激励因素； $YEAR_{it}$ 代表时间因素。变量定义见表 2。

表 2 变量定义

类型	符号	含 义	说 明
因变量	ROE	净资产收益率	直接取自年报
自变量	GYG	国有股比例	国有股股份/总股本
	FRG	法人股比例	法人股股份/总股本
	TA	企业总资产	直接取自年报（RMB）
	PAY	高管薪酬	直接取自年报（RMB）
	YEAR	上市年限	样本当年与上市年份之差

(三) 研究结果

1. 研究样本

本文以深、沪两市上市国有企业公开披露的年度报告作为数据来源。考虑到行业的竞争性，我们选择近似完全竞争的电子电器行业、垄断竞争性商业行业和接近完全垄断的公用事业行业为研究对象。研究窗口为 2000，2001 和 2002，去除 ST 公司数据。2000~2002 年期间，总样本量 270 个观测值。缺失的数据利用区间均值进行了估算。具体而言，电子电器企业 41 家，公用事业 9 家，商业企业 40 家。使用的计量软件是 E-views 3.1。

2. 回归结果

我们分别对混合、电子电器、商业和公用事业样本进行了回归估计。由于国有股和法人股具有共线性，因此分开两次估计。回归结果见表 3。

表 3 2000~2002 年多元线性回归结果

自变量	混合样本		电子电器		商业		公用事业	
	模型 1	模型 2	模型 1	模型 2	模型 1	模型 2	模型 1	模型 2
GYG		- 0.0389 * (- 4.589)		- 0.052 * (- 3.648)		- 0.028 (0.0503)		0.006 (0.191)
FRG	0.029 * (2.492)		0.034 * * (1.991)		0.034 (1.676)		0.042 (0.174)	
TA	1.09E - 13 (0.614)	1.26E - 14 (0.083)	1.59E - 13 (0.835)	- 2.03E - 14 (- 0.130)	4.70E - 10 (1.464)	4.29E - 10 (1.594)	- 2.38E - 11 (- 0.878)	- 1.51E - 11 (- 0.641)
PAY	0.0979 * (3.661)	0.01 * (4.011)	0.014 * (3.528)	0.014 * (3.797)	0.006 (1.583)	0.005 (1.546)	0.019 * (2.602)	0.02 * (2.794)
YEAR	- 0.308 * (- 4.278)	- 0.346 * (- 5.474)	- 0.257 (- 1.686)	- 0.397 * (- 2.557)	- 0.436 * (- 4.379)	- 0.391 * (- 3.997)	0.059 (0.311)	- 0.070 (- 0.279)
C	6.708 * (14.507)	8.849 * (14.892)	5.518 * (5.385)	9.348 * (8.235)	6.808 * (10.411)	8.366 * (9.252)	5.928 * (5.148)	7.118 * (5.353)
调整后 R ²	0.566	0.649	0.504	0.626	0.643	0.679	0.612	0.582
F 值	88.528 *	125.583 *	31.965 *	52.017 *	54.637 *	63.835 *	11.258 *	10.056 *
N	270	270	123	123	120	120	27	27

注：* 在 5% 水平下显著，** 在 0.1% 水平下显著。

方程显著性较好，拟合程度也很高。从表 3 中我们可以得出以下结论：

第一假设不成立。对于混合、电子电器、商业、公用事业行业的回归说明在竞争性不同的各行业中，规模因素和企业绩效都不相关，规模效应对于国有企业绩效的影响并不存在。企业绩效的变化来源于其他因素。

混合行业回归证实了第二、三假设，但是分行业回归说明这种结论只成立于完全竞争性行业，对于垄断竞争行业和垄断性行业假设并不成立。对行业因素的忽略在很大程度上误导了人们对股权结构和企业绩效关系的认识。在后两个行业中股权结构不具有解释性也验证了我们第六假设，即行业性影响确实存在。

对激励因素的检验只在商业行业没有被通过，这可能说明某些具体行业中高层管理人员的激励不仅仅依靠或来源于薪金，可能受到其他管理架构或因素的影响。

在竞争性行业中，经营年限和国有企业绩效存在负相关，企业活力可能随着经营时间的增长、市场竞争的加剧而逐渐丧失，这可能和现有机制中股权流通渠道不畅有关，特别是国有股流通的困难致使很多上市之前效益良好的企业绩效逐步下滑；而在垄断性行业中，由于垄断经营权的存在，企业绩效只和内部管理有关。

然而，我国在对国有企业的改革中经历了几个不同的阶段。1992~1997 年之间主要是

建立现代企业制度的试验和“抓大放小”、国有经济战略性重组阶段；2000年以后重点则主要落在调整国有资产管理体制和债转股方面。政策举措变化对回归结果是否有影响呢？考虑至此，我们又采用1995~1997年的数据进行回归分析，以期在因素分析横向比较的同时也能够进行结构性纵向比较，从而既具有因素上的现实性，又具有时间上的前瞻性。由于企业规模已经在检验中被证实为不显著，因此我们去掉这个解释变量，仍然选择近似完全竞争的电子电器行业、垄断竞争性商业行业和接近完全垄断的公用事业行业为研究对象。研究窗口为1995，1996和1997，去除ST公司数据。1995~1997年期间，总样本量171个观测值。由于此期间对于高层管理人员薪酬一般不予披露，因此去除PAY这一解释变量；同时由于公用事业行业上市国有公司数量很少，所以只针对电子电器企业和商业企业进行分析，两者分别代表竞争性行业和带垄断性色彩行业。具体而言，电子电器企业24家，商业企业33家。

表4 1995~1997年多元线性回归结果

自变量	混合样本		电子电器		商业	
	模型1	模型2	模型1	模型2	模型1	模型2
GYG		- 0.034 * (- 3.017)		- 0.078 * (- 2.018)		0.001 (0.087)
FRG	0.02 * (3.232)		0.023 * (3.878)		0.013 * (2.326)	
TA	- 2.91E - 10 (- 1.745)	- 2.06E - 10 (- 1.136)	- 3.02E - 10 (- 0.515)	- 1.49E - 10 (- 0.199)	- 2.88E - 10 (- 1.527)	- 2.15E - 10 (- 1.108)
YEAR	- 0.241 * (- 2.737)	- 0.337 * (- 3.439)	- 0.068 (- 0.535)	- 0.209 (- 0.639)	- 0.257 * (- 2.79)	- 0.247 * (- 2.493)
C	11.458 * (35.123)	13.436 * (21.757)	5.518 * (5.385)	13.78 * (7.837)	11.486 * (40.07)	11.625 * (19.258)
调整后 R ²	0.975	0.963	0.99	0.694	0.987	0.985
F 值	2183.795 *	1495.704 *	2269.662 *	54.557 *	2529.951 *	2158.510 *
N	171	171	72	72	99	99

注：* 在5%水平下显著。

将表4中的回归结果和表3中比较，可以发现结论基本相同。在各个回归中，企业规模和国有企业绩效都不相关。法人股比例在三个回归中都表现出和企业绩效的正相关，而国有股在竞争性行业显著，在带垄断色彩的行业中不显著。上市年限和企业业绩呈负相关。

三、研究结论及政策建议

通过以上分析，我们可以得出研究结论并对国有资产管理改革提出政策建议。

（一）调整国有经济布局要考虑行业因素

结合两个时期我国对国有企业改革侧重点的不同我们可以看出，在 1997 年之前进行的改革中，与法人股正相关对应的政策是调整股权结构，实现股权多元化；而和国有股在垄断性行业不显著相对应的是国有股逐步退出。这些政策已经取得了显著的成效。随着企业重组的逐步推进，股权调整在垄断性或是垄断竞争性行业已接近到位，因此改革重点应该加以适当调整，而在竞争性行业，仍然可以将股权多元化作为最重要的改革手段。因此，当前改革经营机制进行国有股减持和实现股权多元化必须和考虑行业性质，适度推进。盲目地进行国有股减持和股权多元化不利于国有企业治理的优化和绩效的提高。

对于一般竞争性领域的企业，由于它们无助于国家对经济的宏观调控，国有经济可以考虑从这些行业完全、平稳地退出，以提高其经营绩效；对于垄断竞争行业，例如商业行业，国有经济主要应该通过制度安排，致力于提高行业竞争性；对于具有自然垄断色彩的行业，股权结构并不是影响企业绩效的主要因素，因此保持国有股的控股地位并无不利影响，但从社会总福利来分析，也应该打破传统的垄断，致力于培育竞争机制。

目前国资委为 196 家中央企业圈定的未来发展的五大领域是：国家安全行业、自然垄断行业、提供重要公共产品和服务的行业、重要的资源行业以及支柱产业和高新技术产业中的骨干企业。这可以说是我们实证分析的最好注脚。

同时，股权多元化不一定都要通过国有股转让给私营或是外资企业，可以有多种方式，例如通过法人之间相互参股的方式来实现，只要一个公司里面有三个大股东，就有制约。

（二）增强国有资产的可流通性

另一个重要的因素在于资产的不可流通性。股权结构多元化既不是形成有效公司治理的目的，也不是公司治理有效的手段或必要前提，世界上很少有纯粹为完善治理结构进行股权多元化的公司，人为的股权分散化和多元化并不能有效解决公司治理问题。根源在于资产的不可流通性。回归中上市年限和国有企业绩效负相关的原因就在于此。

在我国，股权特别是国有股权的流动是相当困难的，收益组合中没有股票转让价差的诱惑，导致一些上市前资产质量较高、业绩较好的公司上市后同样出现业绩不佳、母公司掏空上市公司、败德关联交易、公司治理结构失效等问题。因此国资委成立之后，如何构建更加完善合理的国有资产管理体制，如何培育外部市场评价机制，对国有公司经营者形成硬的约束，才是解决公司绩效低下的关键问题。而对上市公司来讲，只有解决了国有股流通问题，才能够使最大的股东真正关心市场变化，才能够通过市场变化评价经营者业绩，同时，市场交易机制也才能对上市公司进行自发的优胜劣汰。

2003 年 6 月，国资委挂牌后批准了国有股转让的首家上市公司，同意大股东西藏自治区国有资产经营公司将所持有的 ST 明珠 4863.8 万股国家股分别转让给山东五洲投资集团和潍坊渤海实业公司。这可以说是一个可喜的突破，但是，市场呼唤更大力度的改革，国资委应该尽快通过国家股流通的相关法案，同时推出针对非上市公司的制度规定。

(三) 建立法人治理制度，完善国有企业内部管理

从实证结果我们发现，规模大小并不是直接影响国有企业经营绩效的因素。国有股份集中的大企业集团经营业绩较好，可能也是受到行业因素的影响。例如，在自然垄断行业，企业规模大部分比较可观，国家控股公司也较多，但分析证明这与经营绩效并不相关。抓大放小，似乎并不是提高国有企业经营绩效的直接手段。组建大企业集团的作用主要在于打破非完全竞争行业中的地域限制，提高行业的竞争性，继而通过股权多元化促进企业经营绩效的提高。因此对企业管理的放开更应该在细分行业性质的基础上进行。

同时，抓大放小并不是说在进行具体研究和管理的的时候只专注于大企业集团的资产管理，而应该是指通过对中小国有企业的改制建立起更有效的企业制度，包括法人治理制度和完善的内部管理制度。法人治理制度不仅包括体制，也包括机制，例如解决上市公司里设立独立董事体制问题，还要解决独立董事的任职调整、报酬、职业道德、社会责任等等一系列的问题。内部管理制度方面，应建立有效的激励机制和管理机制，将高层管理人员或是代理人的收入直接与国有企业经营绩效挂钩。同时，建立严格的财务审计制度和信息披露制度，强化资产经营公司、国家控股公司和持股公司中监事会的作用，加强对公司财务活动的定期审计，以防止管理机制由于长期固定而失去活力。

由此，我们明确了国有资产管理的改革方向，即建立完整、科学、有效的国有企业绩效评估指标体系，通过调整、治理、激励等手段实现评估指标体系的整体优化，不论是体制问题还是布局问题，都应该以提高国有企业绩效为出发点，以实现国有资产管理的整体优化为根本目标。

参考文献

- [1] 许小年：《以法人机构为主体建立公司治理机制和资本市场》，《改革》，1997年第5期。
- [2] 周业安：《金融抑制对中国企业融资能力影响的实证分析》，《经济研究》，1999年第2期。
- [3] 孙永祥，黄祖辉：《上市公司的股权结构与绩效》，《经济研究》，1999年第12期。
- [4] 陈晓，江东：《股权多元化、公司业绩与行业竞争性》，《经济研究》，2000年第8期。
- [5] 魏刚：《高级管理层激励与上市公司经营绩效》，《经济研究》，2000年第3期。
- [6] 刘小玄：《中国企业发展报告（1999～2000年）》，社会科学文献出版社2001年版。
- [7] Coughlan, Anne T., Schmidt, and Ronald M., Executive Compensation, Management Turnover, and Firm Performance, An Empirical Investigation, Journal of Accounting and Economics, 1985, Vol.7.
- [8] Jensen, Michael C., and William H. Meckling, Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure, Journal of Financial Economics, 1976, Vol.3, pp305~360.
- [9] Murphy, Kevin J. Corporation Performance and Managerial Remuneration: An Empirical Analysis. Journal of Accounting and Economics, 1985, Vol.7.
- [10] Myeong - Hyeon Cho, Ownership Structure, Investment, and the Corporate Value: An Empirical Analysis, Journal of Financial Economics, 1998, Vol.47, pp103~121.
- [11] McConnell, John J., and Henri Servaes, Additional Evidence on Equity Ownership and Corporate Value, Journal of Financial Economics, 1990, Vol.27, pp595~612.
- [12] Rosen, Sherwin, Authority, Control, and the Distribution of Earnings, Bell Journal of Economics, 1982, Vol.13.

论公共部门运行中交易费用合理的“度”^{*}

郭小东

交易费用是公共部门运行中一种相当突出也相当复杂的经济现象。公共部门的运行既有可使交易费用降低的一面，也有使交易费用“居高不下”的一面。这种“矛盾”现象，就使对公共部门交易费用的研究具有明显的理论和现实意义。在我国当前进行的公共财政建设中，如何合理看待、把握交易费用问题，如何控制好交易费用的“度”，就是个值得重视的问题。

一、公共部门与交易费用的降低

从经济学一般理论说，公共部门经济之所以能成为整体经济中的一个重要组成部分，或者说在混合经济中政府之所以要承担起资源配置、收入分配、宏观调控等经济职能，其基本原因是由于市场失灵现象的存在。而市场之所以失灵，则往往是因为交易费用过分高昂，使市场机制难以发挥作用。例如，要解决信息不对称所导致的市场失灵，须为之支付数额巨大的信息费用；要克服搭便车行为所导致的市场失灵，须为此花费极其高昂的谈判、监督费用，等等。由于这类交易费用数额极其庞大，甚至往往是无穷大，在现实中根本难以支付，从而导致市场失灵，进而要求政府的介入。按此理论，政府介入经济既然是为了弥补市场失灵，合乎逻辑的结论似乎就应该是，公共部门的存在和运行可以有效地降低交易费用。那么，情况是否如此？联系现实，情况并非那么简单。公共部门的存在，既有可以降低交易费用的一面，也有使得交易费用“居高不下”的一面。我们先看看公共部门与交易费用降低的情况。

从历史和现实的角度看，公共部门的存在和发展，在一定条件下确实具有降低交易费用的作用。这可用如下几个突出方面为例加以说明。

其一，通过政府的公信力以降低交易费用。合法产生的政府都具有一定的公信力。这样，由于公信力的存在，当人们将自己的经济收入通过纳税形式交给政府，由政府支配用以从事公共产品的生产或进行收入再分配时，就可省去大量为了了解政府资信而花费的信息费用以及为了说服自己把钱安心交给政府而花费的决策费用。作为对比，我们可以设想，如果没有政府的存在，人们若要把占自己收入相当比例的钱财交给某个私人机构，由其安排使用，大家必然要反复了解这一机构的状况，并肯定会花费大量时间、精力去判断到底要不要把钱交出去。而且，一个很可能出现的结果是，在花费了大量交易费用后，最终没有一个私人机构能够得到人们的普遍认可。特别是国防、治安等要害性公共品的提

^{*} 原载：《财政研究》，2003年第5期。

供，如果要社会将这类重要事项托付给某一私人机构，让其担当起组织军队、操纵警察的职能，肯定极难得到公众的认可。这时，要想说服公众把自己的安危托付于某个私人机构，交易费用无疑将趋向于无穷大。但是，当存在具有公信力的政府时，情况就简单得多。花费一定的交易费用，政府和公众之间就国防、治安等公共品的提供即可达成共识。交易费用因此从“无穷大”下降为了一个“有限”的量。

其二，通过政府的强制性作用以降低交易费用。这是公共部门运行可以降低交易费用的另一明显表现。突出的例证就是对搭便车行为的抑制。公共部门经济学的有关理论已充分证明，对较大群体规模内产生的搭便车行为，如果试图用市场机制或私人协商等方式加以解决，即使花费极其高昂的交易费用，也难以达到目的。而政府通过所拥有的公权实行强制性措施，则能用较少的成本对搭便车行为加以抑制。政府对税收的强制性征收、对污染的行政性控制等，都是属于这类例子。

其三，通过政府专职部门的设置以降低交易费用。经长期发展，全世界的政府都形成了各种职能部门各有分工、各司其职的运行模式。这种运行模式的好处之一，就是可以节省交易费用。这不但是由于各职能部门长期从事某一领域的工作，有利于对相关信息的掌握了解，可以节约信息费用；也不但是专职部门的决策往往可更快捷有效，可以节约决策费用；而且还因为这种机构设置方式能使民众更容易找到对口部门，更方便于民众和政府的沟通，从而可为公众、为全社会节约大量的交易费用。试想，如果公众面对的是一个没有分工、职权模糊的政府机构，有事根本不知找谁交涉办理，所花费的交易费用必然会高得多。

其四，通过合法简化程序以降低交易费用。这可以用公共决策过程中决策费用的降低为例。当公共决策涉及人们的普遍利益、需要人们共同参与投票抉择时，如果任何一项决议都按“一人一票”的方式逐个征求每个投票人的意见，为此花费的信息、决策费用将会是非常之高的。假如再加上一些公共选择过程中经常出现的现象，诸如投票制度较为苛刻，形成决议的条件较为严格；投票人的偏好存在较大差异，备选方案的数量较多等等，在这种情况下，要形成决议，交易费用更将是个庞大的数字。但是，通过政府部门合法运用决策机制，对相对次要的问题采用由民众所选代表进行表决的代议制度、由政府职能部门提出初始方案等方式作出决定，决策成本就可大大降低。此外，按照唐斯模型的有关理论，政党由于成员间具有共同的施政理念，让政党执政，也可在形成共识等方面起到节约交易费用的作用^①。

其五，通过公共部门所形成的规模经济以降低交易费用。当今世界，公共部门在各国经济中普遍占有一定比重，具有相当规模。而规模经济的好处之一就是能够降低交易费用。例如，对办公用品等的采购，由于各政府机构加总后能形成规模可观的需求量，通过政府职能部门统一进行大批量购买，比之单个企业、单个家庭个别式的零星采购，所花交易费用的单位成本必然较低。

其六，通过对某些产品的统一生产或联合提供以降低交易费用。有些产品通过改变提供方式可以起到降低交易费用的作用。诸如国防、治安、法规、制度、文字等等，都是明

^① (英) C. V. 布郎等：《公共部门经济学》，中国人民大学出版社 2000 年版，第 91 页。

显的例证。我们可以想象，如果没有政府部门的统一提供，而是五花八门，每村都有自己的法规制度，每县都有自己的文字，每省都有自己的军队，那么，姑且不说由此带来的混乱和冲突，即使假定这种状况可以维持，要让人们在文字不同、制度各异的地区之间经商办事，为此花费的交易费用也将是非常高昂的。

通过公共部门的作用以降低交易费用，从而提高相关资源的运用效率，实际上是公共部门得以存在和发展的重要经济原因之一。从这个角度，我们甚至可以这么说：如果公共部门在提供公共产品的过程中不能形成比非公共部门更低的交易费用，那么，它就失去了一个赖以生存和发展的重要经济依据，也就不具备从事公共品生产的充分资格。在一定条件下，降低交易费用应该是公共部门所具有的一项重要经济功能。然而，在当代公共部门经济学中论及政府弥补市场失灵问题时，往往侧重于强调政府的强制性作用，对公共部门在降低交易费用方面的作用则似乎重视不够，这是应该从理论认识上加以适当调整的。

二、公共部门与交易费用的“攀升”

公共部门降低交易费用的功能是个不争的事实，但这毕竟只是事情的一个方面。另一个方面的事实是，即使在正常情况下，在不少场合，公共部门的运行又会要求交易费用居于一定高度。公共部门中形成的交易费用在不少方面往往高于非公共部门。交易费用的“居高不下”同样构成了公共部门运行中一个值得注重的经济现象。

对导致交易费用的一般原因，已有不少学者进行了深入有效的分析探讨。例如，科斯(R. H. Coase)认为，交易费用的产生是因为“为了完成一项市场交易，必须弄清楚谁是参与交易者，必须通告人们，是谁愿意出售某物，以及愿意在何种条件下进行贸易谈判、签订合同并履行为保证合同条款得到遵守所必要的检查，如此等等。”^①赫伯特·西蒙(H. A. Simon)认为，在交易过程中，个人或组织在处理或利用可得到信息方面的能力将不可避免地受到限制，这种限制来自信息的复杂性和信息的不确定性。^②这是导致交易费用的重要原因。威廉姆森(O. E. Williamson)则认为，交易费用的存在取决于三个因素：受限制的理性思维、机会主义以及资产的专用性。^③

但是，这些分析实际上都主要是针对非公共部门的交易费用进行的。这类分析对公共部门的交易费用虽然也有一定解释力，但并不能覆盖全部原因。在公共部门产权条件下，交易费用之所以产生，特别是在不少场合之所以会居高不下，更主要是由公共部门本身的原因决定的。

就公共部门本身而言，使交易费用维持一定高度的原因基本与其产权特点有关，具体有如下几个主要方面。

其一，公共部门产权的内在属性会导致交易费用的上升。要言之，公共部门产权的内在属性起码会通过以下几方面对交易费用的上升产生影响：

① 科斯：《社会成本问题》，《法与经济学》，1960年第3期。

② 参见 H. A. Simon, *Models of Man*, London John Wiley & Sons Inc, 1957.

③ O. E. Williamson, *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*, Macmillan, 1985.

(1) 产权的清晰度特点。产权清晰度是决定交易费用高低的重要因素，清晰度越高，交易费用越低，反之亦是。而公共部门产权的清晰度从总体上远比非公共部门为低，所以会造成交易费用的上升。

(2) 产权结构特点。产权可分解为所有权、支配权、受益权等具体权利。公共部门产权的所有者、具体支配者和受益者之间往往存在各种程度的不一致。在此情况下，为保障公共部门运行的效率，必须通过各种机制发挥激励、约束作用，这些机制的力度越强，相关交易费用也必然越高。

(3) 产权分享度特点。公共部门产权是一种被广泛分享的产权，产权层次越高，其公共程度就越高，相关决策者、消费者群体的规模也越大，从而在权力的运用、相关产品的生产和消费等方面所面对的局面越复杂，交易费用就越高。

(4) 产权安排状况特点。公共部门产权在现实安排中往往具有层级结构、条块结构的特点，这一方面固然可节约一部分信息费用，但另一方面，职能部门之间的往来又会导致交易费用的增加。

正是因为公共部门产权具有与私人产权明显不同的特点，为适应这些产权特点的要求，导致了相关交易费用的“居高不下”。

其二，公共部门产权条件下交易过程的复杂性，决定了交易费用会具有一定的规模。从大的方面举例，如果按照税收交易说理论^①，公共部门提供的公共品，是用税收付价的。这形成了公共品和税收之间的交换关系。但是，这种交换和一般市场交易的“一手交钱，一手交货”明显不同。在市场交换中，交钱和交货的过程基本是统一的，一次性完成的；而公共品与税收的交换是分离的，是通过两个形式上完全分离且各自都相当繁复的过程完成的。例如，用税收购买“治安”，纳税是一过程，治安的提供是另一个过程。这是两个各自都需付出较高交易费用的复杂过程。政府对税收的征收，不但要通过专门的程序确定税基、税率，而且要设置专门的机构，配备专门的人员设备以进行征收监管，所有这些投入都属于征税过程花费的交易费用。同样，为完成“治安”这一公共品的提供，政府要了解与治安相关的各种情况（信息费用），决定治安的规模、质量（决策、谈判费用），要设置机构招聘警员购买器械（该过程中会发生不少交易费用），还须建立监督系统（监督费用），再加上治安部门内部的会议、公牍往来等，这些又都是在“治安”这一产品的提供过程中所须花费的交易费用。两个过程相加，就可看出交易费用所具有的总体规模是颇为可观的。

其三，公共部门产权条件下的决策费用较高。在私人部门，企业乃至家庭的决策，虽然也要在搜集信息、处理信息、形成决定等方面付出费用，但从理论上说，私人部门在市场机制的作用下，可以按照价格信号作出决定。而且，在形成决定的过程中，参与决策的人数较为有限，决策者之间商议、磨合的开支较少。而公共部门要形成决定时，所面临的则是一种远为复杂的状况。由于市场机制在公共品的生产方面是失灵的，因而不可能利用市场的价格信号作出决策。这本身已增加了决策的难度。而且，要明确公共部门应该生

^① 对税收的解释有多种理论，此处所用税收理论只是为了便于说明本文所要说明的问题，而不在于对税收理论进行探讨。

产什么、生产多少，还必须由公众共同作出决定。在公众的消费偏好、价值取向必然存在差异的情况下，就会使整个决策过程变得更为复杂。按照公共选择理论中有关决策成本的分析，决策成本与参与决策的人数之间存在函数关系：参与决策者越多，决策成本越高。而公共部门产权的公共性特点，又决定了公共部门的决策必须是由公众共同作出，这就使公共部门的决策成本高于私人部门成为必然。

其四，公共部门产权条件下的实施费用较高。即使通过较高昂的成本作出了决定后，要实施这一决定，公共部门所需付出的费用也往往较高。在私人部门中，交易决定一旦形成，或者说交易合同一经签订，所剩下实施交易的过程就变得较为简单，主要也就是有关资源的转移、交割问题。此时的交易费用主要体现为有关资源的运输费用和资金的支付费用。但公共部门的情况则并非如此。突出的例子是交税问题。即使是公众经过一定的程序对所需缴纳的税收作出了决定，但要完成税收的征收，所需花费的成本也仍然是相当高的。除政府要为此专门投入大量的人力、物力外，纳税人还要承担一定的奉行成本，在了解税收规则、缴纳税款上花费相应的费用。类似的道理，在实施一项已经决定的投资项目时，公共部门所要履行的手续也远比私人部门复杂，这也使其所要花费的实施成本高于私人部门。

其五，公共部门产权条件下的监督费用较高。在运用资源的过程中，虽然私人部门和公共部门一样，都要花费一定的成本用于监督，但公共部门的监督成本远远为高。为保障公共部门的正常运行，进一步说，为保证公共部门的效率，通过各种监督手段作为约束机制是非常必要的。受本身产权特点的制约，公共部门不可能像私人部门那样提供各种各样的物质鼓励以作为激励手段。同时，代理风险和道德风险又是公共部门中必须注重解决的问题。再加上还要防止滥用公权现象的出现。因而，投放于监督约束方面的交易费用必须达到一定水平。从全社会看，为监督或约束公共部门而投放的费用是非常高的。这不但体现在政府为此建立的专门的监察机构上，而且还体现在社会上报章传媒为实行舆论监督而花费的成本上，还体现在广大民众出于监督公共部门运行的目的而花费的人力、物力支出上。如果将这些费用加总起来，其数额是相当大的。

以上几个方面都说明，在公共部门产权条件下，交易费用往往高于非公共部门。这些费用都是为保证公共部门的运行效率所必须作出的开支。如果简单地追求这些费用的降低，反而可能造成更多的损失，带来更高的社会总成本。例如，在决策费用上，和私人部门相比，同样是对投资项目进行决策，并同样都用成本-效益分析方法作为决策手段，私人部门的投资项目仅计有形成成本和收益即可，而公共部门的项目则必须加计各种无形成本和收益，为此花费的开支显然较多。但如果为降低决策成本而像私人部门那样仅计有形成成本和收益，则公共部门的投资不可能得到合理的分析。如果再退一步，连成本-效益分析都不做，仅靠一两个人拍脑袋作决定，则决策失误的几率将很高，很可能给全社会带来重大损失。同样，监督费用的过分降低，则可能导致公共部门内部滥用公权、贪污舞弊、X-非效率行为的增加，导致代理风险和道德风险所造成的损失增加，从而使全社会为之承担更高的成本。因而，就公共部门而言，不能一味简单追求交易费用的降低。在必要的场合，适度的交易费用是必须予以投放的。

三、公共部门交易费用合理的“度”

如上所述，公共部门既有降低交易费用的一面，又有导致交易费用上升的一面，这就要求我们在现实经济中处理好着两方面的关系，在相关制度的安排上，在有关机制的设计上，合理把握好交易费用支出的“度”，把可以降低的交易费用尽可能降低下来，在需要投放交易费用的地方力求进行合理的、适度的投放，以使公共部门的运行效率得到有效保障。

要把握好交易费用支出的“度”，从总体上说应把交易费用分为两类，区别对待之：第一类是与公共部门产权的内在属性关系不大，可以尽量向“零交易费用”水平降低的费用。这类费用基本是可以利用市场机制加以控制的费用，如政府采购过程中发生的信息费用、洽商费用、签约费用、履约费用等大部分交易费用。这类费用的降低，既不会造成公共部门运行中的制度缺陷，操作中又有较现成的市场机制可以利用，把握起来相对较为容易。降低这类交易费用的途径，应该基本和企业降低交易费用的途径相同。第二类是按公共部门产权内在属性的要求不可以过分降低，需要保持一定支出水平的费用，如上面所说的决策费用、监督费用等，对这部分费用的“度”如何掌握，这才是真正的难点所在。就如前面所举例子，通过公共部门提供国防、治安等公共品，固然能降低交易费用，但这种降低只是相对于“无穷大”这个参照系而言的，到底应该降低到什么程度，需要保持到什么水平才为合理，把握起来难度就大得多。

本文认为，要掌握好第二类交易费用支出的“度”，其基本思路可以是：首先找出制约这类交易费用高低的关键因素，再根据这一关键因素与相关成本的关系，确定交易费用合理的支出水平。在此基础上，通过有效的制度安排，力求把交易费用降低到可以容许的程度。按此思路，试通过图1作较展开的说明。

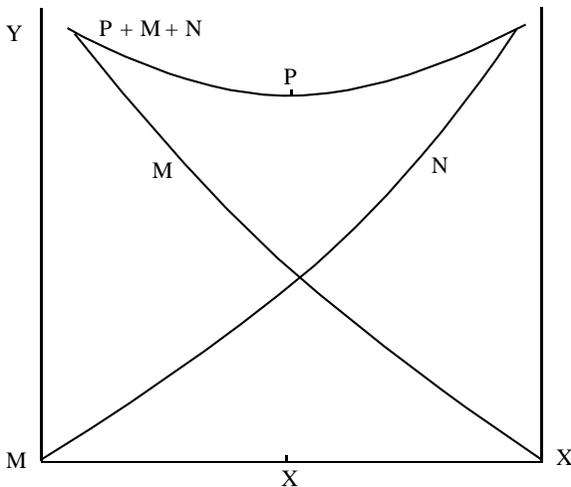


图 1

图 1 中各线的基本含义是：横轴（X）代表制约成本高低的关键性因素的变动；纵轴（Y）表示成本的变化；右边纵轴与左边纵轴的意义相同。坐标内南北走向的两条曲线（M 和 N）分别代表两种相关成本的变化状况。M 和 N 线上东西走向的曲线 P 代表 M 和 N 这两种成本相加的综合结果。将此图形作为基本分析工具，根据不同情况调整其中一些线条的具体含义，我们就可分别对几种较难把握的交易费用的“度”进行分析判断。

分析之一，决策费用合理“度”的决定。此时，横轴（X）代表参与决策的人数。曲线 M 表示因决策失误所导致的成本，曲线 N 则表示因决策行为而导致的成本。曲线 M 的变化趋势是，当决策人数少时，靠少数人“拍脑袋”，决策失误所导致的成本高；随着决策人数的增加，决策失误所导致的成本不断下降。曲线 N 的变化趋势是，当决策人数少时，决策行为导致的成本低；随着决策人数的增加，决策行为导致的成本不断上升。M 和 N 这两种成本相加后的综合结果，体现为曲线 P。曲线 P 的最低点 P'，就是综合成本最小的地方。横轴上与 P' 相对应的点 X，就是决策人数的最佳点。由于决策人数与决策费用之间存在函数关系，知道了决策人数的最佳点，也就知道了决策费用支出最佳的“度”。

分析之二，监督费用合理“度”的决定。此时，横轴（X）代表监督的力度。曲线 M 表示因监督不力所导致的成本。在监督力度弱时，贪污舞弊、滥用公权的现象较严重，因体制漏洞所造成的成本高；随监督力度的加强，因体制漏洞所造成的成本逐渐下降。曲线 N 表示因实施监督而导致的成本。当监督力度弱时，实施监督所导致的成本低；随着监督力度的增强，实施监督所导致的成本不断上升。综合这两种成本，我们同样能得到最低点 P'，横轴上与 P' 相对应的点 X'，就是监督力度的最佳点。监督力度与监督费用同样有函数关系，根据最佳监督力度，就可决定监督费用支出最佳的“度”。

分析之三，实施费用合理“度”的决定。此时，横轴（X）代表为维持实施效率而推行的约束激励机制的效力。曲线 M 表示因机制无效而导致的成本。在机制效力弱时，因工作消极、X-非效率等所造成的成本高，随机制效力的增强，因 X-非效率等所造成的成本趋于下降。曲线 N 表示为推行约束激励机制而付出的成本。当机制效力弱时，推行机制所导致的成本低；随着机制效力的增强，推行机制所导致的成本不断上升。根据综合成本最低点 P'，利用机制效力与实施费用间的函数关系，就可得出实施费用支出的最佳“度”。

当然，以上分析只是提供了理论上的标准。在现实经济中如何利用这些标准，仍须进行积极的探索。但是，只要参照这些理论上的最佳点进行制度安排，应该是可以达到既保证公共部门的有效运行，又把交易费用控制在尽可能低的水平这一目的的。

网络经济中的屏蔽因子与政府规制*

郭小东 王 倩

进入 21 世纪，电子商务以更加迅猛的势头发展。就信息的供应和获取而言，因特网堪称是一把双刃剑。一方面，它打破了信息传递的空间阻隔，以信息量大、检索方便、信息交流快捷廉宜等优点，大大降低了经济活动中的交易费用；另一方面，在不少场合，它也导致了经济中信息不对称现象的加强。

在网络经济条件下，尽管交易的具体方式在变化，获了信息的手段在变化，但经济的实质并未改变。在经济交往中，企业仍以利润最大化为目标，供应商为获利而控制、掩盖甚至扭曲信息的动机并未减弱。消费者在信息获取方面仍处于劣势，他们仍只能较被动地从企业提供的资料中获取相关信息。而且，在网络经济条件下，经济中的信息仍一如既往地分为“表层信息”和“深层信息”。供应商在网上披露的每每是有利于他们的表层信息，而关于商品质量、成本等的深层信息，消费者往往还是难以获取，这就决定了信息不对称现象在网络经济中必然继续存在。

同时，由于互联网这一新兴媒体被用作经济交往中的依托工具，又会使信息不对称有所扩大。我们把这种扩大信息不对称的因素称为“屏蔽因子”。“屏蔽因子”使网络经济的交易双方之间多了一道有利于供应商的“屏障”，使供应方在屏障的“保护”作用下更便于操纵信息，从而导致了信息不对称的扩大。在电子商务中，“屏蔽因子”之所以存在，是多方面因素综合作用的结果，这些因素与屏蔽因子的关系用函数式表示，即为 $E = E(X_1, X_2, X_3, Y_1, Y_2)$ 。按此函数式，以下分别对各因素作具体说明。

一、从网络经济中的供应商方面看

(1) 经营方式因素 (X_1)。网络经济与传统经济相比，它的最大特征表现为经营方式的虚拟性。特别是电子商务中大量网上经销商的涌现，更使虚拟企业的存在成为普遍现象。这些企业在虚拟办公室与客户打交道，通过 E-mail 处理往来业务。这种运作模式与传统商业相比，供需双方多了一层网络造成的隔阂，供应商因置身网络之后而变得“虚化”，变得难以直接沟通。对操纵信息的行为而言，网络无形中变成了一道保护屏障。网络经济为信息不对称的扩大提供了经营条件。

(2) 企业成本因素 (X_2)。网络经济条件下设立网站、开办虚拟企业的成本低下。虚拟企业投资规模小，内部组织结构简单，成员规模精简，对固定资产、流动资金的要求甚低。网上销售企业甚至不需要有固定的营业场所。这些，都使经营者逃避赔偿责任的方便

* 原载：《经济学家》，2003 年第 2 期。

程度提高,也使得网上销售企业扩大信息不对称为自己牟利的成本降低。

(3) 制度因素 (X_3)。当前,电子商务仍处于一个被鼓励、呵护的发展阶段。对于参与网络经济的供应商来说,由于经营成本低,再加上目前的监管制度不严格、监管体系不健全,没有设置有效的准入审核制度,就使准入门槛显得相当之低;同时,监管的宽松也导致事后惩罚难收阻吓之效,提供商的退出障碍也较低。这种“两头低”的状况,为供应商利用信息优势控制信息乃至扭曲信息在制度上提供了方便。

二、从网络经济中的消费者方面看

(1) 消费者的成本因素 (Y_1)。对网络经济条件下的消费者而言,随着信息量的扩大,随着可选择产品潜在范围的扩大,为甄别这些信息、弄清楚这些产品所需的时间和精力也随之增加。特别是在网络上,厂商们往往寻求以一种差别很小的方式去描述其产品,结果使得对产品进行直观比较几乎成为不可能,一些在传统交易中的非隐蔽信息也成了隐蔽信息,导致消费者辨别信息真伪、成功打假的成本上升。

(2) 交易地位因素 (Y_2)。尽管由于网络技术的发展,消费者获得信息的便利程度提高了,但他们接到的大量推销信息中不少带有泡沫成分,不少信息是按供应商的意愿提供的。网络经济中的消费者仍然处于被动地位。而且,网上交易使消费者失去了一些了解商品提供者诚信度的传统途径,使原本就处于劣势地位的消费者的交易地位进一步降低。

要解决信息不对称问题,一般说有市场和政府两条途径。其中,由市场本身来解决信息不对称问题主要是交易者使交易对方信任自己,相信自己提供的信息是真实的,以实现和扩大交易,其主要方法有:卖方向买方提供担保;投保商业保险;价格显示;交易者订立经济合同;卖方建立信誉等。但是,以上方法都有其局限性。①就卖方向买方提供担保来看。卖方向买方提供担保(质量担保等等)是市场经济中生产者最常用、最普遍的解决因信息不对称使消费者产生不信任问题的方法。虽然这种方法运用广泛,但并不意味着它能从根本上解决信息不对称所引起的不信任问题。其局限性主要表现在:真货、优品的生产者可以向买方提供担保,假货、次品的生产者同样也可以向买方提供担保。这样,产品市场上的难辨真伪现象就会延伸到担保领域,使得消费者对市场上众多生产者提供的担保本身难辨真伪。网络经济的虚拟性,进一步增加了消费者作出判断的难度。②就投保商业保险来看。在现代商业保险发展的条件下,卖方可以选择投保商业保险取得交易对方信任。相关的商业保险主要是信用保险,还包括责任保险等。但这种方式的投保金额、范围都会受到限制,存在大量不完全市场的现象。在网络经济条件下,完全靠市场力量形成一个广泛有效的商业保险体系,难度更大。而且,相对于日常生活中的零星小额买卖,这种方式施行起来也显得过于复杂。③就价格显示来看。最常见的价格显示法就是真货、优品的生产者将其产品价格长期保持在比较稳定的高水平上,使消费者通过价格来识别真货、优品。在前工业化经济中或在工业化初始时期,这种方法是比较有效的。而在现代经济中,这种方法的效力已日渐势微,特别是在网络经济中,新产品、新服务不断涌现,这种方法更是无法适用,而且它还可能助长假货生产商提价并获取更高利润。④就交易者订立经济合同来看。在信息不对称的条件下,卖方为实现或扩大交易可以采取与交易对方订立

经济合同的办法来明确双方的权利、责任与利益。这是从古至今的一种常用方法，在现代经济中也更普遍、更正式。但它在网络经济中也遇到了新问题——网上订立合同的合法性仍无保障，违约责任难以追究。^⑤就卖方建立信誉来看。在信息不对称的条件下，卖方要取得买方的信任，可以通过始终提供真货、优品在市场中树立讲信誉的形象。树立起信誉，可以赢得消费者，增加交易，扩大市场。这种做法的局限性也是明显的：如果网络经济中形成了不讲信誉的竞争，就会造成讲信誉的反而吃亏，甚至可能首先被挤出市场。

实际上，信息不对称本身就是导致市场失灵的重要原因。由于网络经济的实质未变，解决屏蔽因子造成的信息不对称问题，单靠市场机制是不够的，政府有必要行使管制职能。针对屏蔽因子导致的信息不对称，政府可以采取以下几种措施：

(1) 通过立法强制网络经济中处于信息优势的一方提供真实的、更多的信息。这一措施在一般市场交易中已有应用，在网络经济中，政府也应尽早立法，明确各项交易中商品或服务的供应者必须向消费者提供的信息，从而确保消费能以此为依据作出合理决策。

(2) 对网络经济中信息优势方利用信息优势进行不公平交易和不公平竞争的行为进行足够的处罚。要解决信息不对称问题，仅有立法的威慑作用是不够的，政府必须对违法行为进行足够的处罚。所谓“足够”，就是指行为人不仅不能从实行不公平交易和不公平竞争中得到任何好处，反会遭受损失。

(3) 努力降低消费者的信息成本。通过各种媒介向消费者介绍识别不公平交易的知识和方法，定期曝光利用网络进行欺诈的厂商。

(4) 政府直接搜集并提供必要的信息。信息在一定程度上具有公共品的性质，某些必要信息的提供可以由政府投资来进行。政府可以建立信息搜寻的专门体系，从而提高网络经济中交易的透明度，以克服信息不对称的弊端。

(5) 对网络经济中交易主体的交易资格进行审查。这是一种预防性措施，采取这项措施的目的在于通过对网络经济交易主体的资格审查，将不合格的交易者清除出市场，达到“净化”市场的目的，从而减少网络经济中出现欺诈等不公平交易和不公平竞争行为的可能性。

(6) 扶植一批有实力、有信誉的网络销售商。在我国现阶段网络经营秩序相对混乱的情况下，政府有必要在对网络经济中交易主体的交易资格进行审查的基础上，从中选择一批有一定经济实力、交易中严格遵守国家法律法规的网络销售商，由国家通过给予税收、融资、政府定点采购等政策优惠使其得到加速发展。

(7) 发挥消费者组织的作用。在我国，消费者协会是维护消费者合法权益的专门社会团体。由于该组织资源、人员素质等方面的局限，至今尚未有效介入对网上交易的消费者权益保护，政府应加强扶持，促进消协在网络经济中也发挥其应有的作用。

中国企业年金税收优惠政策的成本研究^{*}

——我国企业年金税收支出的精算统计分析

张 勇 王美今

一、文献回顾和问题的提出

由于企业年金在激励机制设计和养老保险制度等领域有着重要的作用，许多国家都通过税收优惠政策来促进企业年金的发展。如今企业年金与基本养老保险和个人储蓄性养老保险共同构成了我国的“三支柱”养老保险模式，并成为养老保险体系的“第二支柱”。2000年，国务院在《关于完善城镇社会保障体系的试点方案》中规定：企业缴费在工资总额4%以内的部分，可从成本列支。

但对于一项政策，不仅要注重它对经济和社会发展的促进作用，而且还要看到实施这项政策所付出的成本，后者直接影响到政策的运行效率和实际成效。税收优惠政策虽然能促进我国企业年金的发展，但不能忽视为此所付出的代价。

Surrey (1973)^①指出，与基准税收相比（即没有税收优惠政策时的税收），税收优惠政策会减少政府的税收收入，减少的数量就是通常所说的税收支出（tax expenditure），实际上是税收优惠政策的成本体现。现在许多国家每年都要对税收支出进行测算，并将其列入财政预算，为制定税收政策提供依据。那么如何测算我国企业年金税收优惠政策的成本——税收支出呢？影响税收支出的因素有哪些？它们的敏感性如何？本文将从精算统计的角度来进行研究。

由于企业年金与死亡率分布、工资和利率等因素密切相关，再加上养老基金积累方式的多样性，使得对企业年金税收支出的测算非常复杂。目前的研究可归结为两点：

第一是“跨期”问题。Knox (1990)指出，在测算税收支出时，必须计算从缴纳第一笔费用时起，到雇员死亡时领取的最后一笔养老金为止的整个时期的所有税收，但许多国家在测算税收支出时只考虑了雇员工作期间的税收。Clark (1997)用迭代模型分析了“跨期”对税收支出的影响，认为美国政府高估了企业年金税收支出。

第二是基准税收的选择。Knox (1990)，Dilnot 和 Johnson (2000)指出，雇员一般只会选择对自己最有利的养老基金积累方式，当没有税收优惠政策时，必须知道雇员的选择方

^{*} 本文得到中山大学岭南学院科研基金资助，以及哈佛大学陈琳博士和中山大学岭南学院徐现祥博士提出的极富建设性的指导意见，在此谨致谢意。原载：《统计研究》，2004年第8期。

^① Surrey 为哈佛大学教授，曾担任美国财政部长助理，他在1967年首次提出税收支出的概念。

式及其缴纳的税收，而此时的税收才是基准税收，它与企业年金的税收之差就是企业年金的税收支出，但许多人把直接银行储蓄的利息税当作基准税收，这将高估实际结果。Whitehouse (1999) 和刘佐 (2003) 认为，由于不同国家的基准税收不一样，它们之间的税收支出不具有可比性。

我国在这方面的研究还很少，刘云龙 (2002) 认为我国 2000 年的企业年金税收支出为 124 亿元，不过他没有考虑到上述两个问题，同时还扩大了企业年金的覆盖人数。

但是，在目前的研究中，都假设雇员的实际寿命等于预期寿命，实质上把企业年金当成了确定性年金。这是不正确的，因为它还受到死亡率分布的影响，本质上是不确定性年金。本文从不确定性出发，运用精算技术构建了我国企业年金的税收支出模型，并综合了企业年金的“跨期”特性和基准税收的选择。本文研究表明，把死亡率分布加入到模型后，目前的测算方法会高估实际结果。

下面首先阐释了我国企业年金税收制度，构建了我国企业年金税收支出的精算模型，然后是实证分析，最后为结语。

二、我国企业年金税收支出的精算模型

(一) 我国企业年金税收制度

明确我国企业年金的税收制度，是构建精算模型的基础。

首先是我国企业年金的征税模式。根据目前的税收制度^①，我国对企业年金的缴费和投资收益要征税，而雇员领取的养老金免税。产生税收支出的原因是因为在工资总额 4% 以内的缴费免交企业所得税。

其次是基准税收的选择。由于企业年金是由企业自主决定的，因而我国目前能代替企业年金的其他形式只有银行储蓄、商业性个人年金（以下简称“个人年金”）和个人储蓄性养老保险^②。但从目前的税收制度来看，个人储蓄性养老保险介于银行储蓄和个人年金之间，因此本文就从后两种形式来研究我国当前的企业年金税收支出。

最后是“跨期”特性。为了测算所有与税收支出相关的税收，本文把税收分为企业所得税、营业税、个人所得税、工作期间投资收益税和退休期间投资收益税等五种。

(二) 各种指标的精算现值公式

(1) 企业年金缴费的精算现值。根据缴费的税收优惠比率 (l)、工资水平 (w)、起始工作年龄 (x)、退休年龄 (r)，工资增长率 (g_w)、利率 (i) 和死亡率 (${}_kP_x$ ^③) 等参数，本文运用寿险精算技术得到了企业年金缴费在不确定性条件下的精算现值 (π) 以及

① 劳动和社会保障部与博时基金管理公司：《中国企业年金制度与管理规范》，第 5 页。

② 基本养老保险是强制性的，不是企业年金的替代形式。从安全性和稳健性要求来看，一般的证券投资也不是其替代形式。

③ ${}_kP_x$ 表示 x 岁的人在 k 年后活着的概率，显然 $0 \leq {}_kP_x \leq 1$ ，这正是企业年金不确定性的体现。

在确定性条件下的现值 (π')^①。

在确定性条件下, 必须假设每个雇员的实际寿命都等于预期寿命, 由于预期寿命大于退休年龄, 因而每个雇员在退休后都还活着, 这就意味着企业年金的缴费是从雇员参加工作到退休时为止的整个期间; 但在不确定条件下, 雇员在退休时活着的概率只有 $r_x P_x$, 如果雇员在退休前死亡, 缴费就停止了, 因而 $\pi < \pi'$ 。

(2) 企业所得税的精算现值。对于企业年金的缴费, 根据上述参数和企业所得税率 (t_c), 就可以得到没有税收优惠政策时企业所得税的精算现值:

$$T_1 = \pi \cdot t_c / (1 - t_c) \quad (1)$$

而存在税收优惠政策时企业所得税的精算现值为:

$$T_2 = T_1 \cdot t_c = [\pi \cdot t_c / (1 - t_c)] \cdot t_c \quad (2)$$

两者之差等于^②:

$$T_1 - T_2 = \pi \cdot t_c \quad (3)$$

(3) 营业税及其附加的精算现值。对于经营企业年金的机构来说, 缴费是营业收入, 因而要缴纳与营业收入有关的税收, 包括营业税、城市建设维护税及教育附加费。如果营业税率及附加为 t_s , 就可以得到营业税的精算现值为:

$$T_3 = \pi \cdot t_s \quad (4)$$

(4) 个人所得税的精算现值。对于雇员来说, 企业年金是其延期收入, 因而要缴纳个人所得税。本文用 t_x 表示个人所得税率, 雇员个人所得税的精算现值等于:

$$T_4 = \pi \cdot t_x \quad (5)$$

把式 (1) ~ (5) 中的 π 换成 π' , 就得到确定性条件下的税收现值, 但会高估实际结果。

(5) 投资收益税的精算现值。通过资金运用, 企业年金还可以产生投资收益。由于银行储蓄交利息税, 年金产品的投资收益交所得税, 两者的税率不一样。本文根据养老基金积累方式分别测算了工作和退休两个时期的投资收益及其税收。

根据利息税率 (t_i) 和养老金增长率 (g_p), 可以得出工作期间和退休期间的利息税的精算现值, 分别用 T_5 和 T_6 来表示。如果是年金产品, 只需把 T_5 和 T_6 中的 t_i 换成 t_x , 就得到了工作期间和退休期间投资收益所得税的精算现值, 分别用 T_7 和 T_8 来表示。

(三) 各种养老基金积累方式的税收

(1) 银行储蓄的税收。如果通过银行储蓄来积累养老基金, 那么雇员要交个人所得税和利息税, 企业要交企业所得税, 这些税收的精算现值之和等于:

$$T_b = T_1 + T_4 + T_5 + T_6 \quad (6)$$

政府得到的税收还有银行的营业税, 税收的精算现值总和为:

$$T_b = T_b + T_3 = T_1 + T_3 + T_4 + T_5 + T_6 \quad (7)$$

① 在精算术语中, 精算现值表示考虑死亡率时 (不确定性) 的数值, 而现值表示确定性条件下的数值。

② 这里假设 π 是税后缴费。若 π 为税前缴费, 则 $T_1 = \pi \cdot t_c$, $T_2 = 0$, 两者之差不变。测算税收支出只需要两者之差。

(2) 个人年金的税收。目前开办这种业务的机构主要是保险公司, 根据我国税法, 保险公司的养老金业务免交营业税, 但雇员要交个人所得税, 企业要交企业所得税, 个人年金的税收精算现值就等于:

$$T_a = T_1 + T_4 + T_7 + T_8 \quad (8)$$

(3) 企业年金的税收。目前经营企业年金的机构主要也是保险公司, 与个人年金的差别仅在于企业还可以免交企业所得税, 因此政府能得到的税收精算现值为:

$$T_p = T_2 + T_4 + T_7 + T_8 \quad (9)$$

(四) 我国企业年金税收支出的测算模型

(1) 如果 $T_a < T_b$, 当没有税收优惠政策时, 个人年金是积累养老基金的最佳选择, 应该以 T_a 作为基准税收。税收支出就是个人年金与企业年金的税收之差, 即:

$$LTE_a = T_a - T_p = T_1 - T_2 \quad (10)$$

(2) 如果 $T_a > T_b$, T_b 就是基准税收, 税收支出为银行储蓄和企业年金的税收差额:

$$LTE_b = T_b - T_p = T_1 + T_3 + T_5 + T_6 - T_2 - T_7 - T_8 \quad (11)$$

综上所述, 目前我国企业年金的税收支出为 $\min(LTE_a, LTE_b)$ 。

以上结果是一个雇员终生的税收支出精算现值, 对于每年的平均税收支出, 用 TE_a 和 TE_b 来表示, 他们分别等于 LTE_a 和 LTE_b 的年平均^①。

如果当年雇员总人数为 n , 那么该年度的企业年金税收支出总额就分别等于:

$$TTE_a = n \cdot TE_a, \quad TTE_b = n \cdot TE_b$$

三、我国企业年金税收支出的精算统计分析

本节首先根据统计资料对参数进行赋值, 并对各种指标作了基本分析; 其次是我国 2002 年企业年金税收支出的总量分析; 然后是敏感性分析; 最后研究了养老金水平、人口老龄化和退休年龄对税收支出的影响。

(一) 企业年金税收支出的基本分析

1. 参数值的设定

根据我国税收制度和统计年鉴的数据, 企业所得税率为 33%, 营业税率及附加为 5.5%, 利息税率为 20%, 企业年金税收优惠比例为 4%, 2002 年城镇单位就业人员平均工资为 12422 元。世界银行对我国的长期实际利率和长期实际工资增长率的预测分别为 2.5% 和 3.5%, 而对通货膨胀率的预测为 3%。由于要确保雇员的养老金购买力不会降低, 因此养老金增长率应与通货膨胀率保持一致。目前企业年金的覆盖人员基本上是城镇企业职工, 因此本文采用《全国城镇总人口生命表 (1989~1990 年)》。

① 这里的除数是由利率和死亡率构成的年金的精算现值。

本文把起始工作年龄设定为 17 岁^①，退休年龄为 60 岁^②。由于缴费只受雇员工资的影响，因此个人所得税就是雇员工资的所得税，而不包括其他收益的所得税，为此本文把长期个人所得税率设定为 5%^③。

2. 测算结果与基本分析

为了便于比较分析，本文测算了各个指标在不确定性和确定性两种情形下的结果，以及两者的差额（确定性数值 - 不确定性数值，下同），如表 1 所示。

表 1 测算结果

单位：元

指标	不确定数值	确定数值	差额
π	24834	26388	1554
T_1	12232	12997	765
T_2	4036	4289	253
T_3	1366	1451	85
T_4	1242	1319	77
T_5	2282	2475	193
T_6	1009	568	- 441
T_7	598	649	51
T_8	273	154	- 119
T_a	14344	15120	776
T_b	16764	17359	595
T_p	18130	18811	681
T_p	6149	6412	263
LTE_a	8195	8708	513
LTE_b	11981	12399	418
TE_a	278	288	10
TE_b	407	411	4

当把企业年金当成确定性年金时，只有退休期间投资收益的税收比不确定性情形下的税收低，其他所有指标的数值都高于不确定性情形下的数值。

① 本文根据学历对我国城镇就业人员进行了分类，求出各类人员的数量和起始工作年龄，并以他们的加权平均起始工作年龄作为本模型的起始工作年龄。

② 我国法定男退休年龄为 60 岁，女干部退休年龄为 55 岁，女职工退休年龄为 50 岁，但《中国劳动统计年鉴》(2003) 显示，60 岁以上的就业人员占总就业人员的 5.9%，因此本文把退休年龄设定为 60 岁。

③ 我国 2002 年个人所得税总额为 1211 亿元，工资总额为 13161 亿元，因此平均个人所得税率为 9.2%，其中来自于工资的个人所得税约占个人所得税总额的 40%，本文中的个人所得税率就为 3.68%，但从长期来看，个人所得税率还会上升。

如果雇员不要交任何税收，养老金精算现值就为 24834 元；若以银行储蓄为养老基金的积累方式，个人所得税的精算现值就为 4533 元 ($T_4 + T_5 + T_6$)，实际税率为 18%，是名义税率的 3.7 倍。由于雇员没有税收优惠，个人年金和企业年金对雇员是一样的，个人所得税的精算现值都为 2113 元 ($T_4 + T_7 + T_8$)，实际税率为 8%，也高于名义税率，可见目前的税制降低了企业年金的购买力。

在 5 种税收当中，企业所得税最高，达到 12232 元。一是由于企业所得税率高；二是企业所得税在雇员工作期间就要缴纳，而此时的死亡率比退休期间低，并且受利率的影响也更小。当存在税收优惠政策时，企业所得税为 4036 元，比优惠前减少了 8195 元。

由于 $T_a < T_b$ ，因此应该以 T_a 为基准税收，它与 T_b 之差就是一个雇员终生的税收支出的精算现值，为 8195 元。雇员的年均税收支出等于 278 元。

在确定性情形下，雇员的年均税收支出为 288 元，比不确定性条件下的数值高出 3.6%，从而高估了我国企业年金的税收支出。

(二) 企业年金税收支出的总量分析

到 2002 年底，企业年金覆盖人数为 657 万人，而基本养老保险人数为 1.47 亿。根据前文的分析， TTE_a 是我国企业年金的税收支出总量，为了便于比较，本文还计算了 TTE_b ，如表 2 所示。

表 2 2002 年的企业年金税收支出总量

单位：亿元

税收支出	不确定数值	确定数值	差额	占 GDP 百分比	占税收百分比
TTE_a	18.3	18.9	0.6	0.017%	0.104%
TTE_b	26.7	27.0	0.3	0.026%	0.152%

注：这里指不确定性条件下的企业年金税收支出占 GDP 和全国税收总额的百分比，我国 2002 年 GDP 为 104790.6 亿元，总税收为 17636.45 亿元。

我国 2002 年企业年金总税收支出 (TTE_a) 为 18.3 亿元，占 GDP 的 0.017%，总税收的 0.104%，而在确定性条件下， TTE_a 等于 18.9 亿元，比实际结果高出 3.6%。在目前的税收制度下，税收支出是比较低的，主要是因为参加企业年金的人数太少，只有基本养老保险的 4.5%，企业年金的发展状况与目标相差很远，这是否由于税收优惠政策力度不够才产生的呢？Whitehouse (1999) 指出，企业年金规模和税收优惠政策的关系很复杂，受到产业结构和养老保险制度等诸多因素的影响，税收优惠政策力度越大，并不意味着企业年金规模也越大。目前很多企业连基本养老保险费的缴纳都是迫于国家政策的强制才得以执行，而实行企业年金是企业的自愿行为，要其缴费的难度就可想而知了。

如果以银行储蓄的税收为基准税收，税收支出将达到 26.7 亿元，比实际结果高出 46%，可见基准税收的选择对测算结果有着根本性的影响。

前文指出，税收支出包括企业所得税、营业税、个人所得税、工作期间和退休期间投

资收益的税收等 5 种税收产生的税收支出，表 3 是他们在税收支出中所占的比重。

表 3 税收支出组成成分的比重

税收	企业所得税	营业税	个人所得税	工作期间投资收益税	退休期间投资收益税
TTE _a	100.0%	0	0	0	0
TTE _b	68.4%	11.4%	0	14.1%	6.1%

企业所得税产生的税收支出（不是企业所得税）的比重是最大的，当前占到了 100%。一些企业在雇员未达到退休年龄就将资金撤出保险公司，甚至采取“长险短做”的手法来躲避所得税，这些都增加了税收优惠政策成本，而且企业年金也没有得到正常发展。

保险公司免交营业税，因而营业税对税收支出没有贡献。

由于雇员不享受税收优惠，企业年金缴费的个人所得税在任何一种税制下都要缴纳，因此个人所得税不产生税收支出，但并不意味着个人所得税率对税收支出没有作用，这是因为 T_5 、 T_6 、 T_7 和 T_8 都受到个人所得税率的影响。雇员可以利用延迟缴纳的税收去投资，并把税后收益据为己有，其实这也是 T_5 、 T_6 、 T_7 和 T_8 的来源。即使对雇员缴费有税收优惠，但缴费的个人所得税并没有减少，雇员缴费是免税的说法是错误的。

（三）敏感性分析

根据测算模型，税收优惠比例、工资水平和雇员人数与税收支出是同步变化的，因此税收支出对这三个因素的敏感性为 100%。至于其他因素的敏感性，如表 4 所示。

表 4 税收支出的敏感性分析

参数	参数值	TTE _a	TTE _b
i	2.50%	- 2.50%	16.70%
g_w	3.50%	79.50%	76.10%
g_p	3.00%	0	0.90%
t_i	20.00%	0	25.40%
t_x	5.00%	0	- 7.50%
t_c	33.00%	100.00%	68.40%
t_s	5.50%	0	11.40%

税收支出对工资增长率和企业所得税率的敏感性是最强的，而且呈同方向变化。

从式 (10) 可以看出，在目前税制下， g_p 、 t_i 、 t_x 和 t_s 等因素并不影响税收支出，因

此敏感性等于零。企业所得税产生的税收支出是全部的税收支出，它对企业所得税率的敏感性就为 100%。

至于税收支出对利率的敏感性，情况要复杂一些，因为 T_5 和 T_7 先是随着利率上升而增加，而后随之减少。其他税收与利率是反方向变化的，因此税收支出对利率的敏感性取决于各组成部分对利率的敏感性，目前税收支出对利率的敏感性为 - 2.5%。

(四) 替代率对税收支出的影响

在目前税收优惠比例为 4% 的情况下，企业年金替代率水平只有 8.9%。表 5 反映了不同企业年金替代率水平下的税收优惠比例和税收支出增长率的关系。

表 5 替代率、税收优惠比例与税收支出增长率

养老金替代率		10%	20%	30%
税收优惠比例	个人年金	4.5%	9.0%	13.5%
	银行储蓄	5.0%	10.1%	15.1%
税收支出增长率	TIE_a	12.4%	124.7%	237.1%
	TIE_b	25.8%	151.5%	277.3%

企业年金替代率水平每提高 1%，税收优惠比例就要增加大约 0.45%，税收支出增长 11.2%。在当前政策下，要使替代率水平达到 20% 的改革目标，税收优惠比例应提高到 9%，同时税收支出增长 125%，达到 41 亿元。

(五) 人口老龄化对税收支出的影响

死亡率降低是人口老龄化的重要因素，2001 年我国居民的人均预期寿命为 71.8 岁，比 1981 年增加了 4 岁。本文分别测算了寿命增加 1~5 岁时的税收支出增长率，如表 6 所示。

表 6 人口老龄化对税收支出增长率的影响 (%)

增加寿命	1 年	2 年	3 年	4 年	5 年
TIE_a	0.019	0.036	0.051	0.064	0.077
TIE_b	0.028	0.054	0.078	0.100	0.121

寿命增加 1 岁，税收支出增加了大约 0.019%，为 34 万元，因此死亡率对税收支出的影响是很小的。

TIE_b 的增长幅度比 TIE_a 大，这是因为前者包括了更多的税收成分，比如营业税。

虽然人口老龄化对税收支出的影响很小，但对养老金的支付水平产生了很大压力，许

多国家对养老保险制度的改革都始于人口老龄化。为了缓解人口老龄化的影响，提高退休年龄或许是一种解决方案，但这也会增加企业年金的税收支出。

在目前税制下，退休年龄每提高 1 岁，税收支出就大约增加 2.56%，为 0.47 亿元。当把退休年龄提高到 65 岁，税收支出就增加了 12.6%，达到了 2.31 亿元（见表 7）。

表 7 不同退休年龄的税收支出增长率

退休年龄	61 岁	62 岁	63 岁	64 岁	65 岁
TIE _a	2.56%	5.10%	7.63%	10.13%	12.61%
TIE _b	2.55%	5.09%	7.61%	10.12%	12.58%

四、结论

税收支出作为税收优惠政策的一种成本，对税收政策的制定和运行效率有着重要的影响。目前对养老保险（不仅仅是企业年金）的研究，基本上都是在确定性的条件下进行的，但这可能导致较大的偏差，对于税收支出的测算也一样。本文运用精算技术构建了我国企业年金税收支出的测算模型，提供了减少这种偏差的解决办法。本文的主要结论有：

第一，我国 2002 年的企业年金税收支出为 18.3 亿元，如果把企业年金当成确定性年金，则会比实际结果高出 3.6%。若在目前要使企业年金替代率达到 20% 的改革目标，税收优惠比例应从目前的 4% 提高到 9%，税收支出将达到 41 亿元。

第二，目前我国企业年金税收支出比较低，原因是参保人数少，企业年金规模太小。但是税收政策越优惠，企业年金规模并非就越大，两者的关系受到很多因素的影响，税收优惠政策的力度不能超过一定限度，否则只会增加税收支出，导致税收优惠政策的低效率。刘云龙（2002）还指出，我国很多企业购买企业年金的动机并不是满足雇员养老的需要，而是负责人想获得保险公司的回扣或迫于人际关系等原因。

第三，目前的税收优惠政策实质上只是对企业的优惠，雇员的受益是非常间接的，这也是税收支出比较低的一个原因。而且参保雇员的实际税率高于名义税率，因此还应明确对雇员的税收优惠。直接提高对雇员的税收优惠力度，还能在一定程度上减少企业对企业年金的随意控制行为。

第四，虽然人口老龄化对税收支出的影响很小，但对养老金的支付水平产生很大压力，成为许多国家对养老保险制度进行改革的原因。提高退休年龄是解决人口老龄化的一种办法，但也会增加税收支出。

不过，本文有以下几点需要说明：首先，测算结果有赖于参数值的设定，本文的参数值都来源于国家统计局和世界银行等机构，力求有客观、合理的依据。但即使取不同的数值，结论都不会改变，即：从确定性角度测算的企业年金税收支出都会被高估。其次，目前我国只对企业缴费有税收优惠，因此本文只研究了它产生的税收支出，如果对雇员也有税收优惠，则应该从新构建模型，但思路是一样的，而且本文的思路在养老保险的其他领

域也同样适用。再次,本文研究的税收支出是税收优惠政策的直接成本,但还存在间接成本,如企业利用企业年金来避税,这种现象在我国已不少了,还有待于作进一步研究。

参考文献

- [1] Clark, Robert L., Wolper, Elisa, Pension Tax Expenditures: Magnitude, Distribution and Economic Effects, In Sylvester J. Schieber and John B. Shoven, eds., Public Policy Towards Pensions, Cambridge, Mass.: MIT Press, 1997.
- [2] Dilnot, Andrew and Johnson, Paul, Tax Expenditures: The Case of Occupational Pensions, Fiscal Studies, 2000, Vol.14, No.1, pp42~56.
- [3] Knox, D., The taxation support of occupational pensions: A long-term view, Fiscal Studies, 1990, Vol.11, No.4, pp 29~43.
- [4] Surrey, Stanley S., Pathways to Tax Reform: The Concept of Tax Expenditures, Harvard University Press, 1973.
- [5] Whitehouse, E. R., The Tax Treatment of Funded Pensions, World Bank, Social Protection Discussion Paper Series of World Bank, No.9910, 1999.
- [6] 劳动和社会保障部与博时基金管理公司:《中国企业年金制度与管理规范》,中国劳动出版社2002年版。
- [7] 刘云龙:《中国企业年金发展与税收优惠政策支持》,《管理世界》,2002年第4期,第45~54页。
- [8] 刘佐:《对税收支出问题的再探讨》,《税务研究》,2003年第3期,第28~33页。
- [9] 世界银行东亚及太平洋地区人类发展业务局:《中国社会保障改革的战略选择方案》,http://report.drc.gov.cn, 2000。

税收诱致性现金持有量模型因果性检验 及对我国地下经济规模的估测*

夏南新

一、现金是地下经济运作的“血液”

地下经济运转的主要特征之一是现金交易，因而，在国际上，地下经济通常又冠以“Cash economy”名称。现金是地下经济的工具和在现行体制的体外循环的血液。采用现金支付方式，其交易结算在银行账户上和买通各个关键环节上都不会留下痕迹，从而逃避政府和公安的检查和监督。为此，一旦地下经济活动猖獗，必然会引起社会对货币现金需求量的增加。由于地下经济往往是大笔交易，金额特别大，有时竟然出现用秤称或用尺量钞票的计钞方式，进而加剧了货币结构大面额化。有的国家和地区还把大面额钞票增长程度作为估算地下经济规模的基本方法。地下经济主要表现为走私、偷逃税、毒品、卖淫、寻租等黑色市场，它是在新旧体制转轨时期从监管薄弱和市场短缺的狭缝中孳生的。

所谓地下经济，是指未向政府申报，未纳入政府统计，逃避税收和监管的合法与非法的经济活动，它又称为隐形经济、影子经济、第二经济和平行经济等。在估测地下经济规模的方法中，货币分析法是具有一定技术性和代表性的方法。对货币分析法作出过贡献的学者主要有卡甘（Cagan）、法依格（Feige）和坦斯（Tanzi），他们先后提出了现金比率法、交易法和税收诱致性现金持有量法。以下主要针对税收诱致性现金持有量法进行分析和应用。

二、地下经济估测的现金持有量模型的引入

国外比较成熟的地下经济模型首推税率、利率等为独立外生变量的货币需求模型，这一模型是坦斯（Tanzi，1980，1983）经过对现金比率法修正后导出的，它的具体形式为：

$$\frac{C}{D} = \frac{C_o}{D_o} + \frac{C_u}{D_u} = K_o + \frac{C_u}{D} = f(y, r, \omega S) + f(\tau)$$

式中： C 、 C_u 和 C_o 分别表示通货实际持有量、地下和地上通货； D 、 D_u 和 D_o 分别表示实际、地下和地上活期存款，并且 $D_u + D_o = D$ ； K_o 表示地上的通货对活期存款的比率； y 为申报收入； r 为利息率； ωS 为工资和薪金在收入中所占的份额； τ 为平均实际边际税

* 广东省哲学社会科学规划项目“地下经济对国家经济安全影响的数量分析”，批准号为“02C90”。原载：《统计研究》，2004年第3期。

率。

坦斯以美国为背景，分别拟合了以净付税总额对调整过的总收入的比率（T）和利息收入的加权平均税率（TW）这样两种形式的通货需求模型。

(1) 使用平均税率 T：

$$\ln(C/M_2) = \beta_0 + \beta_1 \ln(1 + T) + \beta_2 \ln(WS/NI) + \beta_3 \ln R + \beta_4 \ln Y + e_1 U_{-1} + e$$

(2) 使用加权平均税率 TW：

$$\ln(C/M_2) = \beta_0 + \beta_1 \ln(1 + TW) + \beta_2 \ln(WS/NI) + \beta_3 \ln R + \beta_4 \ln Y + e_1 U_{-1} + e$$

式中：C 表示现金； M_2 表示广义货币供应量；WS/MI 表示工资与薪水占国民收入的比例；R 表示定期存款利率；Y 表示实际人均 GNP；U 表示校正项。坦斯使用了几种替换的纳税额变量，这些变量都与个人所得税相关。尽管美国的利息、红利和租金基本上是用支票支付的，但是部分工资仍然是用现金支付的，为此构造 WS/NI 形式的解释变量引入模型。随着经济的发展，Y 存在递增趋势，从而影响公众现金持有量。坦斯方法使用现金与货币供给 M_2 的比率来防止活期存款和定期存款的转移而影响估算结果。

三、对外生变量利率和税率的 Tintner 检验

根据中国国情，调整坦斯的现金持有量模型使之适应中国，建立如下回归方程：

$$\ln(C^d/M_2) = \alpha_0 + \alpha_1 \ln(TT/GDP) + \alpha_2 \ln(TW/GDP) + \alpha_3 \ln(AI) + \alpha_4 \ln(RPI) + \varepsilon$$

式中： C^d 表示流动中的货币； M_2 表示广义货币；TT 表示税收总收入；TW 表示工资总额；GDP 表示国内生产总值；AI 表示一年定期存款平均利率；RPI 表示环比零售价格指数。

1. 无约束回归

以下建模所采用的数据均来自《中国统计年鉴》（2002 年）和《中国金融统计年鉴》（2002 年），样本窗为 1978 ~ 2001 年，利用 SAS8.1 软件编程运算。

$$\ln(C^d/M_2) = 6.35631 + 0.18855 \ln(TT/GDP) + 1.68062 \ln(TW/GDP)$$

$$s.e. \quad 0.37091 \quad 0.10221 \quad 0.19623$$

$$P \text{ 值} \quad 0.0006 \quad 0.0807 \quad 0.0001$$

$$- 0.04852 \ln(AI) + 1.21479 \ln(RPI)$$

$$0.06307 \quad 0.43537$$

$$0.4512 \quad 0.0117$$

$R_{IR}^2 = 0.8795$ ， $F = 34.68$ （P 值 < 0.0001 ）。 $\ln(AI)$ 系数为负的，说明流动中的现金与利率是呈反向变化的，即利率越高通货越少。但是在以上拟合的模型中， $\ln(AI)$ 系数的标准差（s.e.）的 P 值（=0.4512）在 5% 的显著性水平上非常不显著，这表明年平均利率（AI）对流通中的货币占广义货币的比例（ C^d/M_2 ）没有显著影响，不过，税率也只能在 10% 的显著性水平上显著。再从 D - W 检验看， $DW = 1.701$ ，根据 $T = 24$ ， $\alpha = 0.05$ ，查表得 $d_L = 1.01$ ， $d_U = 1.78$ ，由此可得 $d_L < DW < d_U$ ，故不能确定扰动项 ε 是否存在一阶自相关。

2. Tintner 检验

$$(1) H_0 : \alpha_1 = 0 \quad H_1 : \alpha_1 \neq 0$$

$$(2) H_0 : \alpha_3 = 0 \quad H_1 : \alpha_3 \neq 0$$

Tintner 检验统计量：

$$F^* = \frac{(ESS_* - ESS)/c}{ESS/(T - p - 1)} \sim F(c, T - p - 1)$$

其中， ESS_* 为在 H_0 成立时的回归方程对应的残差平方和； T 为观察次数，即样本容量； $p+1$ 为包含的自变量个数（含截距）； $T - p - 1$ 为有效样本容量。若检验结果显著，认为在这 c 个约束条件不全为真；否则，接受 H_0 ，认为全部约束条件与观测的实际资料是一致的。

Tintner 检验结果：

$$(1) F_1^* = \frac{0.02634}{0.00774} = 3.40, P_1 \text{ 值} = 0.0807 ;$$

$$(2) F_3^* = \frac{0.00458}{0.00774} = 0.59, P_3 \text{ 值} = 0.4512 .$$

其中 $df_1 = c = 1$ ； $df_2 = T - c - 1 = 19$ ，从检验统计量 F_1^* 的 P_1 值（=0.0807）看，税率至少在 10% 的显著性水平上可以拒绝 H_0 ，即可以认为税率对流通中的货币占广义货币的比例（ C^d/M_2 ）有显著的影响；从检验统计量 F_3^* 的 P_3 值（=0.4512）看，在 10% 的显著性水平上，没有理由拒绝 H_0 ，进一步证实年平均利率（ AI ）对流通中的货币占广义货币的比例（ C^d/M_2 ）没有显著影响。

四、税率和利率对流通中的货币影响的显著性的诠释

尽管国外在对现金持有量模型应用时能同时将利率和税率作为关键解释变量而引入，而以上对现金持有量模型进行检验却发现，将该模型应用到中国时却只能保留税率，而不能引入利率，其原因在于：在国际上，已经被冠以“Cash economy”的地下经济通常表现在，无论何种制度的国家，其地下经济规模在一定程度上可以看作是法律和税制缺陷的积累，是地上公开经济欠缺的指示器。经济越发达，越显得经济立法监管滞后。税率过高，地下经济就会越猖獗，税收反而会减少，这是古今中外早已被证实了的事实。而利率对地下经济的影响很难直接挂上钩，所以，我国现在处于转轨时期，法规制度还不健全，此时也是最容易孳生地下经济活动的发展时期。

事实也是如此，我国利率对通货影响确实是不显著的。尽管自 1996 年起，央行连续 8 次降息，使利率降低了 78.4%。从 1999 年 6 月至今，居民储蓄存款利率维持在改革开放以来的最低水平上，像一年期的存款利率只有 2.25 个百分点，但是，居民储蓄存款额的增长势头却依然不减。据中国人民银行发布的统计数据表明，至 2003 年 5 月末，我国城乡居民储蓄存款余额高达 8.04 万亿元，首次突破 8 万亿元大关。

从 1978 年至今居民储蓄存款大致分为四个发展阶段：① 1978 ~ 1988 年为持续增长阶段，其间储蓄存款平均每年增长 30% 以上，1988 年储蓄存款余额达到 3822 亿元，为

1978年210亿元的18倍。②1989~1996年为高速增长阶段，在上期基数已经较大的情况下，储蓄存款余额仍然以31.6%的强劲势头增长，到1996年储蓄存款余额跃升为3.8万亿元。③1997~2000年为减速增长阶段，此间，央行连续降息，财政开征20%的利息所得税，实行存款实名制等，致使居民储蓄存款逐步分流，储蓄存款余额的增长速度开始放缓，2000年城乡居民储蓄存款余额为6.4332万亿元，当年只增加4977亿元，比上年同期少增1276亿元，增速为7.9%，增幅比上年下降3.7个百分点，1998~2000年的增长率依次为15.4%，11.6%和7.9%。④2001年以来为恢复性增长阶段，从总体上看，储蓄存款的增速比前几年有所提高，但是同1978~1996年的高速增长相比，其增速还略显得低，不过存款的稳定性有所增强。近年来储蓄存款出现回升的主要原因：一是股市低迷，新股发行少，从而使股市分流资金相对减少；二是政府公务员及事业单位人员加薪；三是新的证券交易保证金制度规定，居民由证券公司开户转到商业银行开户，从而增加了银行储蓄存款总额。

尽管此间的居民储蓄存款增长速度忽高忽低，但是其规模却总是在不断地扩张。由于居民的收入目前还不具备大规模投资的能力，从而在很大程度上只能将其转化为储蓄。再加上，中国人素有崇尚节俭的传统，因此储蓄率像亚洲一些国家一样居高不下也就不难理解了。

从储户结构看，占储蓄账户总数20%的储户，其储蓄额占总储蓄的60%以上，他们属于高收入群体；而占储蓄账户总数80%的储户却只占总储蓄的40%以下，这都是一些低收入者。绝大多数低收入者的储蓄动机是基于对未来收支预期的不确定而作的预防性储蓄。低收入储户对利率变化不太敏感，他们与其说是追求更多无风险收益，倒不如说是为了使自己的来自不易的收入有更大的安全性和流动性。而在高收入阶层中，像私营企业老板和个体工商户的储蓄大部分是被用作经营的流动资金，所以对利率变动也不敏感；而其他高收入者的储蓄可能是来源于消费的结余，尽管对利率变化较敏感，但在通货紧缩、投资风险大的情况下，只能持币观望，先将资金存进银行作间接投资。银行拥有的庞大储蓄存款，加上企业上市的严格条件，从我国目前融资的现状看，企业间接融资（向银行贷款）的比例远远大于直接融资（发行股票和债券）的比例。不过，社会上存在的一个巨大的现金流也是地下经济猖獗的表象之一，因为地下经济活动是以现金交割为特点的。

五、约束回归拟合我国税率诱致性现金持有量模型

根据以上样本窗资料拟合出税率诱致性现金持有量约束回归模型：

$$\ln(C^d/M_2) = 1.62957 + 0.21701\ln(TT/GDP) + 1.62599\ln(TW/GDP) + 0.96927\ln(RPI)$$

s. e.	0.34291	0.09430	0.18104	0.29308
P 值	0.0001	0.0323	0.0001	0.0035

不难看出，以上约束回归方程的参数只要在5%显著性水平上就能表现出显著。F = 47.00 (P 值 < 0.0001)， $R_r^2 = 0.8758$ ，均是显著的。DW = 1.676，因 T = 24， $\alpha = 0.05$ ，查表得 $d_L = 1.10$ ， $d_U = 1.66$ ，从而 $d_U < DW < 4 - d_U$ ，故扰动项不存在一阶自相关。

从约束回归模型中可以看出，流通中的货币占广义货币的比例 (C^d/M_2)，对税收总

收入占国内生产总值的比例 (TT/GDP) 的弹性为 0.2170; 对工资总额占国内生产总值的比例 (TW/GDP) 的弹性为 1.62599; 对环比零售价格指数 (RPI) 的弹性为 0.96927。即: 税收总收入占国内生产总值的比例增长 1%, 流通中的货币占广义货币的比例将增长 0.2170%; 工资总额占国内生产总值的比例增长 1%, 流通中的货币占广义货币的比例将增长 1.62599%; 环比零售价格指数增长 1%, 流通的货币占广义货币的比例将增长 0.96927%。

税率诱致现金能力尽管不如工资总额占国内生产总值的比例和环比零售价格指数, 但是它对现金的作用也不可低估。税率与通货占广义货币的比例呈正向变化, 主要是因为地下经济活动是以现金交割便于逃脱工商监管和法律的制裁的缘故, 这一现金暗流使得社会通货无形地膨胀。税率越高, 地下经济越猖獗 (即偷逃所得税、关税等税种的人越多), 税收反而越少, 这是约束回归模型对中外这一历史事实所给予的进一步验证。

六、利用税率诱致性现金持有量约束回归模型估测我国地下经济规模

经过试验筛选, 以上税率诱致性现金持有量约束回归模型拟合优良, 现利用该模型估测我国地下经济规模。其估测的基本思想, 也即坦斯方法的要点为:

(1) 假定地下经济中的绝大部分交易都是用通货进行的, 且 $D_u \rightarrow 0$ 。

(2) 地下经济存在和发展的主要原因在于居高不下的税率。

(3) 由于偷逃漏税是地下经济活动参与者获取非法收入的主要动机, 所以, 一个拟合优良的通货需求模型应当将纳税负担作为一个关键的解释变量引入。

(4) 假定合法和非法通货收入的流通速度相等, 用通货流通速度乘以地下经济活动所占有的现金量, 就可估测到地下经济规模。估测结果详见表 1 和表 2。

表 1

时 间	拟合优良的 C^d/M_2 方程		接近零税* 的 C^d/M_2 方程	
	拟合值 (C^d/M_2)	估计值 (C^d)	新生成 C_0^d/M_2 序列	估计值 C_0^d
1995	0.143926	8743.5829	0.087323	5304.9272
1996	0.125364	9539.6274	0.076062	5787.9053
1997	0.109711	9943.3924	0.066564	6032.8785
1998	0.123749	12931.6262	0.075081	7845.9068
1999	0.101243	12138.8581	0.061427	7364.9166
2000	0.103174	13888.3436	0.062598	8426.3686
2001	0.112671	17836.0690	0.068360	10821.5419

注: 由于双对数模型中的对数的首数不能为 0, 所以这里所拟合的模型中税率不能设定为零水平。假如将税率设定为接近零, 即若将年总税率均设为以 10% 的发展速度下降, 也就是每年税率乘以 10%, 则可以认为税率几乎降至零水平, 在这样一个十分低税率水平下, 可以假设纳税者基本上没有偷逃税动机。

表 2

时 间	非法通货 (亿元)	合法通货 (亿元)	合法货币收入 流通速度	未入账收入 (亿元)	未入账收入占 GDP 比例* (%)
	$C^d - C_0^d$	$M_1 - ①$	GDP/②	① × ③	④/GDP
	①	②	③	④	⑤
1995	3438.6557	20548.4443	2.845865	9785.949904	16.734384
1996	3751.7221	24763.0779	2.741364	10284.835903	15.150570
1997	3910.5139	30915.4861	2.418607	9457.996292	12.649047
1998	5085.7194	33867.9806	2.354672	11975.201071	15.016303
1999	4773.9415	41063.3585	1.998558	9540.998976	11.625795
2000	5461.9750	47685.2250	1.874870	10240.493068	11.454229
2001	7014.5271	52857.0729	1.814957	12731.065062	13.270747

注：这里所涉及的地下经济仅限于偷逃税方面，因此所估得的未入账收入占 GDP 的比例适中，在 11% ~ 17% 之间，若把其他地下经济考虑在内，此比例肯定还会加大。

七、结语

加拿大学者伊塞尔 (Ethier, 1980) 在应用坦斯方法时发现，货币持有量对纳税负担的反应十分敏感，其数量就是纳税申报人的平均边际税率，相比之下，坦斯方法的估测值只是局限在一个狭小的范围内，其值约在占在录国民生产总值的 5%，它大大低于现金比率法的估计值。以上估测表明，我国地下经济规模落在 11% ~ 17% 之间，远高于国外同类方法所估的 5% 的结果，但与国外利用现金比率法估得的结果以及作者曾使用现金比率法估得的结果持平。就表 2 的估测结果与近年来我国现实还是非常吻合的，我国地下经济活动有逐年蔓延之势，这与我国正处于经济转轨时期有关，我国经济制度和体制正处于逐步完善阶段。

当然，高税率加剧逃税只是一种简单表象，其实还应有对政府的态度、文化素养、惩处力度和逃税难易等因素决定逃税现象。

参考文献

- [1] [美] L. 法伊格：《地下经济学》，上海三联书店 1995 年版。
- [2] 夏南新：《地下经济估测》，中国财政经济出版社 2002 年版。
- [3] 伍超标：《经济计量学导论》，中国统计出版社 1998 年版。
- [4] Vito Tanzi, *The Underground Economy*, Lexington, Mass.: D. C. Heath, 1982.
- [5] Roderick Hill and Muhammed Kabir, *Currency Demand and the Growth of Underground Economy in Canada, 1991 ~ 1995*, *Applied Economics*, 2000, Vol. 32, pp 183 ~ 192.

第五部分

金融与资本市场

资本资产定价模型的实证研究*

陈浪南 屈文洲

本文试图运用上海股票市场的数据，对资本资产定价模型进行实证检验，尤其注重在CAPM模型中起着举足轻重的 β 值的分析和测量上（分为上升 β 值、下降 β 值以及S&W- β 值），并根据股市中的三种市场格局（上升、下跌和横盘）划分了若干的时间段来进行分析，进而检验 β 的解释力。

一、CAPM实证检验述评

（一）传统实证研究对CAPM的支持

Black, Jensen 和 Scholes (1972) 没有直接检验市场投资组合在高效集合上的预言，相反，他们把注意力集中在证券市场线上。我们知道，如果市场投资组合是高效的，那么它就意味着一个线性、正斜率的关系应该存在于 β 值与期望收益率之间。Black, Jensen 和 Scholes (1972) 的经验检验被设计用来证明证券市场线的这些性质。

另一个对CAPM的经典实证研究是由Fama和MacBeth (1973) 进行的。他们的研究与Black, Jensen 和 Scholes (1972) 最根本的不同之处是他们试图在前一时期估计的风险变量基础上预测投资组合未来的收益率。他们使用的样本是1906~1968年间在NYSE中交易的所有股票。他们的研究基本上支持CAPM的。

（二）对CAPM的挑战

Banz (1981) 对CAPM的检验中加入了公司规模这个因子，并对CAPM提出批评。在1936~1975年间，他发现在适当调整股票的风险后，CAPM中那些市值较小公司的股票平均收益率比那些大公司高出许多。这就是著名的公司股本规模效应。

Fama和French (1992) 考虑了是否存在其他因素能够解释横截面上的差异。除了加入公司股本这一因素外，他们还将公司收益的账面值与市值之比(BE/ME)作为解释因素。他们发现这个比率能解释平均收益率在横截面上的大部分变动情况，还发现这个比率(BE/ME)具有比公司规模更强的解释能力。

但是，Fama和French (1992) 采用与Fama和MacBeth (1973) 相同实证方法进行研究，却得到了完全相反的结论：Fama和MacBeth (1973) 的研究表明收益率与风险存在正

* 本文系国家自然科学基金课题《我国开放型市场经济条件下货币和财政政策的协调机制》、《国债管理系统研究》，国家教委社科基金课题《市场经济条件下财政政策和货币政策研究》，中加教育合作课题《公开信息与市场活动相关性研究》的成果之一。原载：《经济研究》，2000年第4期。

的相互关系，而 Fama 和 French (1992) 的研究却发现两者根本不存在任何关系。Fama 和 French (1992) 把产生不同结果的原因归咎于在他们各自的研究中采用了不同的样本周期。

上述几个著名的实证研究表明： β 值看来无法解释平均收益率在横截面上的变动。

(三) 近期实证研究对 CAPM 的支持

Kothari, Shanken 和 Sloan (1995) 对 Fama 和 French (1992) 研究结果的分析主要集中在 β 值的系数上。他们认为这个估计值具有较大的标准差。这说明，在较大的置信区间上 β 值的系数有较大的取值范围，统计上不能认为其只能是负值。Amihud, Christensen 和 Mendelson (1992) 进一步发现如果使用更有效的统计方法，那么对平均收益率和 β 值关系的估计值是正的而且是显著的。

Black (1993) 认为，Banz (1981) 提出的公司规模效应只是在某个样本时期内才能出现。这个结果也被 Jagannathan 和 McGrattan (1995) 的研究所证实。

至于权益的账面值与市值之比 (BE/ME)，Fama 和 French (1992) 发现其能有效地解释收益率在横截面回归中产生的差异。但是 Kothari, Shanken 和 Sloan (1995) 指出，在 Fama 和 French (1992) 从 Compustat 公司所获得的数据中存在着某些问题，即 Compustat 公司挑选上市公司以及处理这些数据的方法。他们采用了 Compustat 公司所挑选的样本，并使这些样本在整个样本期中保持不变，但没有采用 Compustat 公司后面又加进来的上市公司数据。其实证结果表明，在这些样本中权益账面值与市值之比的效应比 Fama 和 French (1992) 所得到的结果要小得多。

除了以上提到的 Fama 和 French (1992) 对 CAPM 中 β 值的挑战外，现在大部分经济学家对 Fama 和 French (1992) 实证结果的反映就是力图寻找另一个资本资产定价模型。但是 Jagannathan 和 Wang (1993) 并不这样认为。他们认为，缺乏对 CAPM 的实证支持应归咎于那些为了使实证检验方便而设立的不适当的假设条件。对 CAPM 的实证检验必须包括能反映全部资产的市场组合。因此 Jagannathan 和 Wang (1993) 把人力资本也包括到实证研究中来，提出了无约束条件的预期收益率三 β 方程。这一方程表明，预期收益率与 β 值因素成线性关系。

二、上海股票市场 CAPM 实证研究的设计

(一) 研究方法

采用双程回归技术，第一程是应用时间序列回归求 β 值，第二程回归在性质上是横断面的，通过观察值的最佳拟合线就是证券市场线的估计，最后构造相应的模型，并进行统计检验分析。

(二) 数据的选取

1. 研究对象 (股票市场)

本文选取上海股市为研究对象主要基于以下两个原因：①资料可得性。上海股市尽管

于 1996 年 7 月 1 日正式发布了上证 30 指数，但其综合指数依旧发布至今，因而为市场收益率计算提供了基本数据。②国内已有的关于 CAPM 实证研究的论文多是以上海股市为研究对象，故为本文的结果提供了对照比较的对象。

2. 股票样本数据的选取

数据的选取一方面考虑到要具有较长的历史时间，另一方面又要考虑到有较多样本的情况下。因此，本文选取了 100 家股票（代码从 600600 到 600699）作为股票样本。

3. 时间段的确定

本研究中，选取 1994 年 1 月 4 日到 1998 年 12 月 31 日作为研究的时间段；再根据股市中的三个市场格局（上升、下跌和横盘）划分为若干子时间段，以供进一步研究之用。时间段的划分如图 1 所示。

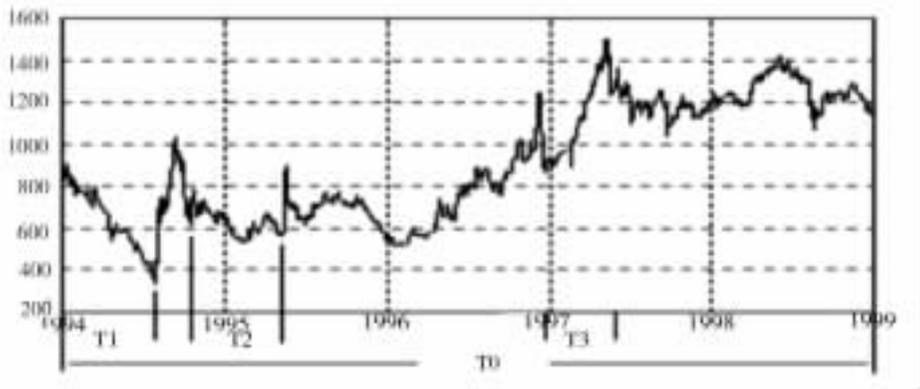


图 1

（全时间段）T0：1994 年 1 月 4 日至 1998 年 12 月 31 日

（下跌段）T1：1994 年 1 月 4 日至 1994 年 7 月 29 日

（横盘段）T2：1994 年 10 月 14 日至 1995 年 5 月 12 日

（上升段）T3：1996 年 12 月 27 日至 1997 年 5 月 9 日

4. 市场指数的选择

本研究选择上证综合指数作为市场指数，因为上证综合指数能够比较准确地反映股票整体行情的变化和股票市场的整体发展趋势，并包括了上海股市中各种的证券，比较符合 CAPM 中所描述的市场组合。而且上证综合指数是一种价值加权指数，符合 CAPM 市场组合构造的要求。

5. 无风险利率的确定

在国外研究中，常以一年期的短期国债利率或银行同业拆借利率来代替无风险利率。但是我国目前的利率还没有市场化，且国债以长期品种为多，因此无法用国债利率来代表无风险利率。在本研究中，决定采用三个月的定期居民储蓄存款利率作为无风险利率。

（三）股票收益率的确定

本研究在综合衡量后决定在 T0、T1、T2、T3 时间段选取日、周、月的收益率来分别

进行实证研究，并对得出的结果进行比较。

本文所采用的股票收益率的计算公式如下：

$$R_{it} = \frac{P_{it} - P_{i(t-1)} + D_{it}}{P_{i(t-1)}} \quad (i = 1, 2, \dots, 100) \quad (1)$$

其中： R_{it} 是第*i*种股票在*t*时刻的收益率； P_{it} 是第*i*种股票在*t*时刻的收盘价； $P_{i(t-1)}$ 是第*i*种股票在*t-1*时刻的收盘价； D_{it} 是第*i*种股票在*t*时刻每股所获得的红利、股息等收入。 $D_{it} = \text{每股现金股利} + P_{it} \times (\text{送股比例} + \text{配股比例}) - \text{每股配股价} \times \text{配股比例}$ 。

因为目前国家股、法人股不能上市流通，社会股东对它们的配股转让并不热心，因此国家股、法人股的转配股可忽略不计。

(四) 上证指数收益的计算

在上证综合指数计算原则中，对股票的分割、送配以及股票新上市等情况已分别作了相应的调整，因此，可以直接按下列公式计算指数收益率 R_{mt} ：

$$R_{mt} = \frac{t - \text{Index}_{t-1}}{\text{Index}_{t-1}}$$

式中： R_{mt} 表示市场组合*m*在*t-1*时刻到*t*时刻的收益率； Index_t 表示市场组合*m*在*t*时刻的收盘指数； Index_{t-1} 表示市场组合*m*在*t-1*时刻的收盘指数。

(五) 股票系数 (β 值) 的估计

估计单个股票的 β 系数采用单指数模型：

$$R_{it} = \alpha_i + \beta_i R_{mt} + e_{it} \quad (2)$$

其中： R_{it} 表示*i*股票在*t*时间的收益率； R_{mt} 表示上证指数在*t*时间的收益率； α_i ， β_i 为估计的系数； e_{it} 为回归的残差。

利用(2)式，进行时间系列的一元线性回归，得出 β 系数的估计值 β_i ，表示该种股票的系统性风险的测度。

(六) 组合的构造与收益滤计算

对CAPM的总体性检验是检验风险与收益的关系。由于单个股票的非系统性风险较大，用于收益和风险的关系检验易产生偏差。因此，通常构造股票组合来分散大部分的非系统性风险后进行检验。本文按个股 β 系数的由小进行分组构造组合，第一个股票组合包含 β 系数最小的 10 支股票。依次类推，最后一个组合包含 β 系数最大的 10 支股票。

按 β 值大小划分了 10 个股票组合后就可以计算组合的收益率。由于在现实股票市场中，投资者一般多采取等额的数量来进行投资并组成投资组合，因此在本文中对股票组合的收益率采用“简单算术平均法”，以此来更加接近现实股票市场的状况。

简单算术平均法：

$$\bar{R}_{pit} = \frac{\sum R_{it}}{N}$$

其中： \bar{R}_{pit} 是第 i 个投资组合在 t 时刻的收益率； R_{it} 是第 i 只股票在 t 时刻的收益率； N 是组合中所包含的股票家数。

(七) 组合系数 (β 值) 的估计

组合 β 值的估计，采用下面时间系列的 CAPM 模型：

$$R_{pt} - R_{ft} = \alpha_p + \beta_p (R_{mt} - R_{ft}) + e_{pt}$$

其中： R_{pt} 表示 t 时期组合的收益率； R_{ft} 表示 t 时期的无风险利率； R_{mt} 表示 t 时期的市场组合的收益率； e_{pt} 为回归的残差； α_p ， β_p 为估计参数。

(八) 对 β 值进行估计的其他方法

本文试图通过不同的 β 值估计方法来计算 β 值并对 CAPM 进行实证研究，最后比较各个不同 β 值的估计是否有利于提高 CAPM 的精确度。 β 值的其他估计描述如下：

1. 上升 β 值和下降 β 值的计算

许多股票在股票指数上升阶段相对指数的上升幅度与股票指数下跌阶段相对指数的下跌幅度是不一致的。如由于绩优股是投资者追捧的热点，与股票走势相比，其走势较强，股票指数处于上升阶段时，其涨幅大于指数的涨幅；股票指数处于下跌阶段时，其跌幅小于股票指数的跌幅。这种情况表明，应该用不同的 β 值来度量在股票指数上升时期和下跌时期的表现。由此引出了上升 β 系数和下降 β 系数。

某股票的上升 β 系数是指在一段时期内，选出所有指数上涨的日期，求出的该股票在这些对应日期内收益率对于股票指数的收益率作线性回归而得的 β 系数。某股票的下降 β 系数是指在一段时期内，选出所有指数下跌的日期，求出的该股票在这些对应日期内收益率对于股票指数的收益率作线性回归而得的 β 系数。

$$\text{上升 } \beta_i = \frac{\text{Cov}(R_{it}, R_{mt}^+)}{\text{Var}(R_{mt}^+)} \qquad \text{下降 } \beta_i = \frac{\text{Cov}(R_{it}, R_{mt}^-)}{\text{Var}(R_{mt}^-)}$$

其中： R_{mt}^+ 表示在 t 时期指数收益率大于零的收益率； R_{mt}^- 表示在 t 时期指数收益率小于零的收益率； R_{it} 表示在 t 时期第 i 只股票的收益率。

2. 运用休尔斯 - 威廉姆斯 (Scholes-Williams) 法来估计 β 值

为了避免 β 值的偏差估计，国外学者分别提出了不少解决的办法，如施瓦特 (Schwert) 等人提出的交易对交易 (trade-to-trade) 法，丁松 (Dimson) 对 β 值的一致统计估计量度称为丁松法，以及科恩 (K. J. Cohen)、浮勒 - 罗克 - 乔格 (Fowler-Rorke-Jog)、麦克英里希 (McInish-Wood) 等人都对 β 值的估计进行了一系列的探索，并得到了有益的结果。本文运用休尔斯 - 威廉姆斯 (Scholes-Williams) 对 β 值进行的一致统计估计得出的估计量 β_i (以下简称 S & W - β)，他们的估计量 β_i 表述如下：

$$\beta_i = \frac{b_i^0 + b_{i+1}^0 + b_{i-1}^0}{1 + 2b_{m+1}^0}$$

其中： $b_{m+1}^0 = \frac{\text{Cov}(R_{m,t+1}, R_{m,t})}{\text{Var}(R_{m,t})}$ ，时期内市场 β 观察值 (指数自相关)；

$$b_i^0 = \frac{\text{Cov}(R_{it}, R_{mt})}{\text{Var}(R_{mt})}, \text{ 证券 } \beta \text{ 理想值};$$

$$b_{i+1}^0 = \frac{\text{Cov}(R_{i,t+1}, R_{mt})}{\text{Var}(R_{mt})}, \text{ 证券超前一期 } \beta \text{ 观察值};$$

$$b_{i-1}^0 = \frac{\text{Cov}(R_{i,t-1}, R_{mt})}{\text{Var}(R_{mt})}, \text{ 证券滞后一期 } \beta \text{ 观察值}。$$

休尔斯 - 威廉姆斯针对非同步交易 (nonsynchronous trading) 现象提出了以上 β_i 的估计量, 他们用单个证券的收益率对同步和非同步市场指数收益率作回归。本文试图通过以上股市为研究对象来进行 S & W - β_i 估计, 以此来进一步检验 CAPM。

(九) 模型设计

本文主要是以上海股票市场为对象进行 CAPM 实践研究, 并对 β 估计运用不同的方法, 以此来区分 β 值在 CAPM 中的解释能力究竟有多大, 并根据上海股票市场不同的市场格局分别进行 CAPM 的实证研究。其回归模型主要有四种:

$$\bar{R}_p = \gamma_0 + \gamma_1 \beta_p + e_p \quad (\text{M1}) \qquad \bar{R}_p = \gamma_0 + \gamma_1 \beta_{\text{上升}} + e_p \quad (\text{M2})$$

$$\bar{R}_p = \gamma_0 + \gamma_1 \beta_{\text{下降}} + e_p \quad (\text{M3}) \qquad \bar{R}_p = \gamma_0 + \gamma_1 \beta_{\text{S\&W}} + e_p \quad (\text{M4})$$

M1 是经典模型

M2 是以上升 β 为解释因子的回归模型

M3 是以下降 β 为解释因子的回归模型

M4 是以 S & W - β 值为解释因子的回归模型

以上回归模型中的参数为: \bar{R}_p 为组合的算术平均收益率; β_p 为组合的普通 β 系数; $\beta_{\text{上升}}$ 为上升的 β 值; $\beta_{\text{下降}}$ 为下降的 β 值; $\beta_{\text{S\&W}}$ 为休尔斯 - 威廉姆斯 (Scholes-Williams) 法估计出来的 β 值。

本研究又将所选取的样本期 T_0 划分为三个子时间段 T_1 , T_2 和 T_3 。在四个时期段 (T_0 , T_1 , T_2 和 T_3) 进行四种模型的实证研究并分别进行回归, 计算出 β 值、解释变量的回归系数、回归系数的显著性检验 t 值、回归方程总体相关显著性检验 F 值及判定系数 R^2 , 再依这些统计数据作出分析和结论。

三、实证结果与分析

(一) β 值的分析

通过对 100 家样本股收益率进行处理, 可以得到在四个时间段 (T_0 , T_1 , T_2 和 T_3) 内 100 家样本股不同的 β 值, 由于数据量较为庞大, 在本文中仅作出了四个时间段内普遍 β 的频数分布图 (见表 1 至表 4)。

表 1

组限	频率
0.176663	1
0.322782	0
0.468902	1
0.615022	0
0.761141	5
0.907261	9
1.053381	21
1.199501	27
1.34562	23
1.49174	11
其他	2

表 2

组限	频率
- 0.11555	1
0.054966	0
0.225479	1
0.395993	0
0.566506	2
0.73702	0
0.907533	7
1.078047	26
1.24856	35
1.419074	22
其他	6

表 3

组限	频率
- 0.15507	1
0.135744	4
0.426554	3
0.717365	10
1.008175	17
1.298986	16
1.589796	20
1.880606	13
2.171417	7
2.462227	5
其他	4

表 4

组限	频率
0.310434	1
0.471106	2
0.631779	4
0.792451	14
0.953123	20
1.113795	35
1.274467	19
1.43514	2
1.595812	1
1.756484	1
其他	1

T0 时期 (1994.1.4 ~ 1998.12.31) 内 100 家样本股票以时间间隔为月计算出来的, 普通 β 值的频率分布情况如表 1 所示。T1, T2, T3 由于所选取的样本期间较短, 用时间间隔为月计算 β 值会造成回归数据较少的情况, 故在这三个时间内采用以周为时间间隔。T1 时期 (1994.1.4 ~ 1994.7.29) 内 100 家样本股票, 普通 β 值的频率分布情况如表 2 所示。T2 时期 (1994.10.14 ~ 1995.5.12) 内 100 家样本股票, 普通 β 值的频率分布情况如表 3 所示。T3 时期 (1996.12.27 ~ 1997.5.9) 内 100 家样本股票, 普通 β 值的频率分布情况如表 4 所示。通过对上述各时期内普通 β 值的频数分布表的观察, 我们可以对在不同的股票市场格局 (下降、横盘、上升) 中样本股 β 值的变动情况得到初步的认识。在本文所研究的 T0 时间中, 样本股 β 值主要集中在 1.05 ~ 1.35 这个区域内, 占样本股的 71%。在 T1 时间中, 样本股 β 值主要集中在 1.07 ~ 1.42 这个区域内, 占样本股的 83%。在 T2 时间中, 样本股 β 值主要集中在 1.30 ~ 1.88 这个区域内, 占样本股的 49%。在 T3 时间中, 样本股 β 值主要集中在 0.79 ~ 1.27 这个区域内, 占样本股的 88%。从上述分析中, 我们可以看出样本股在上海股票市场上是一组比较活跃的股票, 具有较高的系统风险。在股票市场处于下降的格局时, 样本股的系统风险会增大。在横盘的市场格局中, 样本股的系统风险得到较大地分散。而在上升市场格局中, 样本股表现出最大的系统风险, 有些股票的 β 值甚至达到 2.46。这说明在股票市场的三种格局中, 上升格局中股票所包含的系统风险最大, 下降格局中次之, 横盘格局中最小。因此, 样本股 β 值的变动在一定程度上反应出股票市场不同市场格局的风险状态。为了能更好地说明这个问题, 本研究还对股票市场的三种格局 (上升、下降、横盘) 的 β 值情况作了方差分析。其结果如下:

方差分析结果：单因素方差分析

T1、T2 和 T3 三个时期的方差分析

差异源	Df	F	P- value
组间	2	12.65	5.34E- 06
组内	297		
总计	299		

T1 和 T2 二个时期的方差分析

差异源	Df	F	P- value
组间	1	19.1	2.05E- 05
组内	198		
总计	199		

T2 和 T3 二个时期的方差分析

差异源	Df	F	P- value
组间	1	19.17	1.93E- 05
组内	198		
总计	199		

T1 和 T3 二个时期的方差分析

差异源	Df	F	P- value
组间	1	4.5	0.03495
组内	198		
总计	199		

查表得： $F_{2,297,0.05} \approx 1.70$ ， $F_{1,198,0.05} \approx 1.74$ 。

从方差分析的结果来看，各时期 β 值的方差检验 F 值均大于其检验统计量。因此我们可以有充分的理由来推翻原假设 H_0 ，说明在 T1、T2 和 T3 时期， β 值存在充分的差异，能解释不同市场格局的风险程度。

(二) 回归散点图分析

通过对 100 家样本股 β 值的排序，得到 10 个股票投资组合，并求得各组合的 β_p 值。然后对 10 个组合进行回归分析。由于数据量较大，回归模型数较多，在此只给出了四个回归统计表（见表 5 至表 8），全部回归模型将在下一部分列表表示。

表 5 T0 时期内模型 (M1) 的回归统计表 (时间间隔为月)

	Coefficients	标准误差	T Stat	P-value	回归统计	
Intercept	0.005089	0.006347	0.801845	0.445805	R Square	0.670384
X Variable 1	0.022658	0.005617	4.033689	0.003769	标准误差	0.39696

首先查表得： $T_{1,0.05/2}(8) = 1.8595$ ， $F_{1,0.05/2}(1,8) = 5.32$ 。

表 6 T1 时间内模型 (M3) 的回归统计表 (时间间隔为周)

	Coefficients	标准误差	T Stat	P-value	回归统计	
Intercept	- 0.04971	0.003448	- 14.4141	5.25E- 07	R Square	0.012265
X Variable 1	- 0.00086	0.002722	- 0.31518	0.760693	标准误差	0.003905

表 7 T2 时间内模型 (M4) 的回归统计表 (时间间隔为周)

	Coefficients	标准误差	T Stat	P-value	回归统计	
					R Square	
Intercept	- 0.01402	0.004564	- 3.07149	0.015313	R Square	0.283939
X Variable 1	0.008205	0.004607	1.781074	0.112766	标准误差	0.002831

表 8 T3 时间内模型 (M2) 的回归统计表 (时间间隔为周)

	Coefficients	标准误差	T Stat	P-value	回归统计	
					R Square	
Intercept	0.036671	0.003821	9.597251	1.15E- 05	R Square	0.066063
X Variable 1	- 0.00176	0.002344	- 0.75226	0.473448	标准误差	0.006732

从表 5 可观察到, 在上海股市中系统性风险 β 与收益存在一定的正相关关系。再从回归方程的估计系数中看, β 系数的估计值大于零而且检验值 $T = 4.033689 > T_{1-0.05/2} (8) = 1.8595$, 即该估计值显著地不为零, $F = 16.27065 > F_{1-0.05/2} (1, 8) = 5.32$, 即收益率 (Y) 与系统风险 (β 值) 存在显著线性相关, $R^2 = 0.670384$, 可见 β 值在这个时间对股票组合的收益率具有较强的解释能力。

从表 6 分析, β 系数的估计值为 $- 0.00086$, 其 T 值为 $- 0.31518 < T_{1-0.05/2} (8) = 1.8595$, 而且 $F = 0.099336 < F_{1-0.05/2} (1, 8) = 5.32$, $R^2 = 0.012265$ 。因此, 在 T1 时期内下降 β 值对股票组合的收益率几乎没有解释能力。而无风险收益 $R_f = - 0.04974 (T = |-14.4141| > T_{1-0.05/2} (8) = 1.8595)$, 这说明在下降格局中 β 值的作用较小, 在整个股票市场中各种股票不管其系统风险多大, 其下跌的幅度几乎相等。从上述情况可看出, 上海股票市场中投资者的行为投机性较大, 在大势下跌时, 不管手中的股票是绩优股还是垃圾股, 都非理性地大量抛售。

表 7 所描绘的收益率与系统风险的关系, 具有较高的相关性, 但回归结果 $R^2 = 0.283939$, 回归效果较差。这种结果很明显受到异常值 (第一个组合) 的影响。在这种情况下, 偏离程度很大的“异常值”一般为交易数据中的强影响观察值, 反映了股市运作中的重大信息, 将其强行删去虽可使数据和某些理论模型符合得更好, 但不利于客观地分析股票市场的实行运行状况, 所以有必要同时采用有和无“异常值”的两套数据进行对比分析, 以作参考。采用这种对比分析方法, 有助于分析在没有政府干涉及过度炒作情况下, 上海股市自然波动的特性。

从表 7 删去异常值后得到回归分析结果 (见表 9), 其中 $R^2 = 0.922648$, 有了较大幅度的提高。这说明在横盘时期, 股票交易较为清淡, S & W- β 值对这种非同步交易现象能进行适当的调整, 这样 S & W- β 值对 T2 时期的股票组合收益率有了较高的解释能力。

表 9 回归结果

	Coefficients	标准误差	T Stat	P-value	回归统计	
					R Square	标准误差
Intercept	- 2.66219592	0.225221	- 11.82033	7.03531E- 06	0.922648	
X Variable 1	2.003785201	0.219289	9.137617	3.86438E- 05	0.097263	

从表 8 分析, T3 时期内上升 β 值也同样像 T1 时期内下降 β 值一样, 对股票组合收益率没有较好的解释能力。其中 β 的系数的估计值为 - 0.00176, t- 检验为 - 0.75226。系统风险 β 值在此时无法成为显著的收益率的解释因素。

(三) 回归估计方程的分析

以下将列出各时期内用不同时间间隔计算出来的四个模型 (M1, M2, M3, M4) 共有 48 个回归估计方程, 列示如下:

T0 时间内不同时间间隔的回归模型

时间间隔为日的四个回归模型

模型类型	γ_0	T 值	γ_1	T 值	拟合系数	F 值
M1	- 9.42E- 06	- 0.011642	0.0011211	1.48049	0.215059	2.191851
M2	0.0110546	6.46859	0.0089081	5.643213	0.799226	31.84585
M3	- 0.004989	- 1.786923	- 0.012187	- 4.64655	0.729642	21.59045
M4	0.0010651	1.1685979	7.843E- 05	0.091808	0.001052	0.008429

时间间隔为周的四个回归模型

模型类型	γ_0	T 值	γ_1	T 值	拟合系数	F 值
M1	0.0030948	2.2051883	0.0031934	2.628122	0.46334	6.907024
M2	0.0310245	9.1203956	0.0168969	5.953599	0.815861	35.44534
M3	- 0.026929	- 7.357283	- 0.010328	- 3.09183	0.544404	9.559401
M4	0.0028814	1.5483514	0.0034058	2.109932	0.357523	4.451811

时间间隔为月的四个回归模型

模型类型	γ_0	T 值	γ_1	T 值	拟合系数	F 值
M1	0.00508913	0.8018449	0.0226579	4.033689	0.670384	16.27065
M2	0.0700989	4.9423353	0.0676232	5.523678	0.792267	30.51102
M3	- 0.087768	- 7.294353	0.0089103	0.798066	0.073743	0.63691
M4	0.0018287	0.2183035	0.0264419	3.558301	0.612807	12.6615

T1 时期内不同时间间隔的回归模型

时间间隔为日的四个回归模型

模型类型	γ_0	T 值	γ_1	T 值	拟合系数	F 值
M1	- 0.005031	- 3.238869	- 0.000705	- 0.51717	0.032351	0.267462
M2	0.0130879	8.62008	0.0064454	5.656649	0.799988	31.99768
M3	- 0.017272	- 9.637802	- 0.003641	- 2.27852	0.393556	5.191662
M4	- 0.004273	- 1.80487	- 0.001387	- 0.66502	0.052385	0.442251

时间间隔为周的四个回归模型

模型类型	γ_0	T 值	γ_1	T 值	拟合系数	F 值
M1	- 0.018667	- 3.271631	- 0.00944	- 1.8906	0.308817	3.574363
M2	0.0268162	13.847195	0.0076714	6.603342	0.844974	43.60412
M3	- 0.049707	- 14.41413	- 0.000858	- 0.31518	0.012265	0.099336
M4	- 0.027782	- 4.216039	- 0.001343	- 0.23145	0.006651	0.053567

时间间隔为月的四个回归模型

模型类型	γ_0	T 值	γ_1	T 值	拟合系数	F 值
M1	- 0.133032	- 13.41386	0.0156144	1.801424	0.288581	3.245128
M2	NaN ^①	NaN	NaN	NaN	NaN	NaN
M3	- 0.154117	- 15.41165	0.0180349	2.11462	0.358544	4.471619
M4	- 0.11917	- 14.79695	- 0.004564	- 0.61033	0.044491	0.372505

T2 时期内不同时间间隔的回归模型

时间间隔为日的四个回归模型

模型类型	γ_0	T 值	γ_1	T 值	拟合系数	F 值
M1	- 0.002119	- 2.426443	0.0021155	2.560881	0.450478	6.558109
M2	0.0055008	2.4318081	0.0159823	7.671931	0.880344	58.85853
M3	- 0.002839	- 1.67772	- 0.01613	- 9.91976	0.924813	98.40159
M4	- 0.002681	- 2.830596	0.0026482	2.949634	0.520968	8.70034

① 用此方程进行回归统计时在样本期内没有符合要求的数据存在。

时间间隔为周的四个回归模型

模型类型	γ_0	T 值	γ_1	T 值	拟合系数	F 值
M1	- 0.011651	- 3.400292	0.0057942	1.685942	0.262156	2.8424
M2	0.0273613	12.011179	0.0183015	8.256887	0.89498	68.17619
M3	- 0.029714	- 8.28561	- 0.00487	- 1.4438	0.206708	2.084565
M4	- 0.014018	- 3.071486	0.0082047	1.781074	0.283939	3.172225

时间间隔为月的四个回归模型

模型类型	γ_0	T 值	γ_1	T 值	拟合系数	F 值
M1	- 0.070136	- 6.544832	0.0336363	3.292768	0.575424	10.84232
M2	NaN	NaN	NaN	NaN	NaN	NaN
M3	- 0.082447	- 9.578854	- 0.015755	- 1.81441	0.29154	3.292091
M4	- 0.073122	- 2.661852	0.0359222	1.328155	0.180663	1.763995

T3 时期内不同时间间隔的回归模型

时间间隔为日的四个回归模型

模型类型	γ_0	T 值	γ_1	T 值	拟合系数	F 值
M1	0.0064108	1.9915112	- 0.000152	- 0.04933	0.000304	0.002433
M2	0.0151695	6.2407383	0.0014899	0.675445	0.053952	0.456226
M3	- 0.009757	- 3.352951	- 0.004771	- 1.63634	0.250769	2.677619
M4	- 0.003798	- 0.943578	0.009808	2.521256	0.44277	6.356734

时间间隔为周的四个回归模型

模型类型	γ_0	T 值	γ_1	T 值	拟合系数	F 值
M1	0.0272047	10.65402	0.0025263	1.385621	0.193544	1.919946
M2	0.0366713	9.5972513	- 0.001763	- 0.75226	0.066063	0.565888
M3	NaN	NaN	NaN	NaN	NaN	NaN
M4	0.0236557	5.0398885	0.0057226	1.670342	0.258576	2.790042

时间间隔为月的四个回归模型

模型类型	γ_0	T 值	γ_1	T 值	拟合系数	F 值
M1	- 0.005137	- 0.29211	0.0608456	4.581654	0.724058	20.99156
M2	0.1022947	9.2911708	0.0316724	4.355488	0.703377	18.97027
M3	NaN	NaN	NaN	NaN	NaN	NaN
M4	0.0138446	0.629904	0.045502	2.685766	0.474146	7.213338

从上述的回归方程中，我们可以发现在四个样本时期（不同的股票市场格局中）普通 β 值在对组合收益率横截面上的解释能力较为稳定，而上升 β 值和下降 β 值的解释能力在有的时期较高，而有些时段又很低，不是十分稳定。S & W- β 值最不稳定，但从 T3 时期可以发现，S & W- β 值比普通 β 值具有较高的解释能力，这说明在上升格局中，股票交易较为火爆，容易产生非同步交易现象，而 S & W- β 对这种现象进行了适当的调整，使 β 值的解释能力有了一定的提高。

在 T1 时期（下降市场格局中），下降 β 值比普通 β 值具有更强的解释能力，其判定系数及 F 值都较高。但在下降市场格局中，模型 M3 在以日和周为时间间隔的情况下，其 β 的系数的估计值小于零，这说明收益与风险成反比，收益高的组合其风险却是很小，这种情况比较符合现实股票在下降格局中的情况，即在下降市场格局中股票组合的收益率与其风险测量 β 成反比。

而在 T3 时期（上升市场格局中），上升 β 值与普通 β 值的解释能力差不多，这可能与所选取的样本期有关。因为在 1996.12.27 ~ 1997.5.9 这段时期内，上海股票市场连续攀升，这样就使用于计算普通 β 值与上升 β 值的数据基本上相同，因此从整体来看 M1 和 M2 在 T3 时期内没有显著的差别，其回归方程在横截面上的拟合程度基本上是相同的。

从上述对 T0, T1, T2 和 T3 四个时期的分析，可以看出在上海股市中普通 β 值（模型 M1）的作用还是比较大的。因为其较为稳定，而且也具有对收益一定的解释能力，总体上能反映出收益与风险成正比。其他的三个模型（M2, M3, M4），笔者主要是抱着尝试和创新的意图想找出上海股票市场上的新的一些规律，但从实证结果来看这三个模型在不同时期与不同时间间隔上的回归结果较不稳定。还有待以后选取较长的样本时期进行进一步的研究。

四、结论

通过上述的实证研究可以得到如下结论：

(1) β 值对市场风险的度量具有较为显著的作用。如果假设在各种市场格局（上升、下降、横盘）中， β 值呈正态分布，那么在这三种格局中其各自的 β 值的均值具有显著的差异。而且 β 值在上升格局中最大，在下降格局中次之，在横盘格局中最小。

(2) 普通 β 值、上升 β 值、下降 β 值以及 S & W- β 值对股票组合的平均收益率的解释能力都不很稳定。其中普通 β 值在这四种 β 值中较为稳定，对平均收益率具有 30% ~

50%的解释能力。在大部分样本期内下降 β 值与收益率成反比,即风险越大,收益率越低,这符合现实股票市场的情况。上升 β 值在对收益率的解释没有表现出比其他 β 值更强的能力。而 S & W- β 值随着回归时间间隔的扩大,其解释能力也逐渐下降,因此 S & W- β 值在排除非同步交易而产生的误差上有一定的作用。总得来讲, β 值与股票收益率的相关性较为不稳定性,这表明上海股票市场中股票的系统风险随时间而变化,市场上存在较大的投机性。

(3) 从各样本期的回归方程可以发现,无风险收益率在大部分时间内是负值(即 $r_f < 0$)。这表明在上海股票市场上的投资者具有较强的投机性,追求高风险所带来的高收益,而不关心资本的时间价值。这符合 Black (1972) 提出的 CAPM 的零 β 模式。在本研究中因为在样本期内大部分股票组合的 $\beta_p > 1$, 这样才导致 $r_f < 0$, 从而本研究也可以证明: CAPM 的零 β 模型比标准 CAPM 能更好地描述资产收益。

参考文献

- [1] Scholes, Myron, and Joseph Williams, Estimating Betas, from Non-Synchronous Data, *Journal of Financial Economics*, 1977, Vol.5, pp309 ~ 327.
- [2] Banz, Rolf W., The relationship between return and market value of common stocks, *Journal of Financial Economics*, 1981, Vol.9, pp3 ~ 18.
- [3] Jagannathan, Ravi, and Ellen, R. McGrattan, The CAPM Debate, *Federal Reserve Bank of Minneapolis Quarterly Review*, 1995, Vol.19, No.4, pp2 ~ 17.
- [4] Fama, Eugene F., and French, Kenneth R, The cross-section of expected stock returns, *Journal of Finance*, 1992, Vol.47, pp427 ~ 465.
- [5] Fama, Eugene F., and MacBeth, James D., Risk, return and equilibrium: Empirical tests, *Journal of Political Economy*, 1973, Vol.81, pp607 ~ 636.
- [6] Jagannathan, Ravi, and Wang, Zhenyu, The CAPM is alive an well. Research Department Staff Report 165, Federal Reserve Bank of Minneapolis, 1993.

股权分割、公司业绩与投资者保护*

宾国强 舒元

一、问题的提出

自“6·24”国有股减持叫停以后，我国股票市场的发展进入了一个“迷茫”时期，股权分割长期以来被许多人视为我国股票市场诸多问题的根源，一旦发现政府并无解决问题的良策与决心，投资者长期信心的崩溃自然如洪水决堤，一发不可收拾。每一次“利好”政策的出台都演变成为主力资金出逃的良机也就不难理解，缺少了对股市前景稳定的预期，中国股市也逐渐变成了一个扶不起来的“阿斗”。而具有中国特色“股权分割”就成为众矢之的，本文要提出的问题是——股权分割真的是中国股市问题的关键吗？通过对已有文献的回顾，作者发现，股权分割对上市公司业绩的负面影响事实上并没有得到坚实的实证支持，大部分的研究结果甚至得出了相反的结论。

二、文献回顾

关于我国上市公司股权结构与公司业绩的研究在 20 世纪 90 年代中期以后一直是研究的热点，早期最有影响的文章是 Xu & Wang (1997, 1999)，他们采用上市公司 1993 ~ 1995 年的样本进行的实证研究发现：上市公司的盈利能力与法人股比例正相关，而与国有股比例负相关，上市公司的股权集中度与业绩正相关。在他们的研究中，作者强调了法人股东在公司治理中的积极作用。Xu & Wang 的研究范式影响了后来很多研究成果，并且成为股权分割“负作用”最有力的证据。然而事实上，Xu & Wang 的结果并没有说明股权的全流通必然有利于上市公司的业绩，而他们所强调的法人机构和西方成熟市场上的机构投资者，如共同基金、保险公司、养老基金不同，事实上当时的大部分法人机构的最终控股权仍然落在国有企业手中，其营运方式与国有企业并无太大的差异。

与 Xu & Wang (1997) 相反的是，Chen (1998) 发现上市公司的业绩与国有股比例正相关，而不是法人股。Sun, Tong & Tong (2002) 把国有股和法人股视为同样的股权类别，作为政府持股的指标，采用 1994 ~ 1997 年的样本，他们发现公司业绩和政府股权比例之间是倒 U 型的关系，即适度的国有股比例对上市公司业绩是最优的。Sun, Tong & Tong 的研究无疑夸大了政府股权的规模与作用，因为事实上根据 Liu & Sun (2002) 的统计，仍然有 16% 的上市公司控制在非国有机构手中。Tian (2001) 利用 1994 ~ 1998 年的数据样本得

* 原载：《管理世界》，2003 年第 5 期。

出了与 Sun, Tong & Tong (2002) 完全相反的结论。他们发现政府股权和公司价值之间是 U 形的关系, 即适度的混合所有权形式是最差的模式, 国有股占少部分的时候, 增加国有股权会导致公司价值的下降, 而在国有股权足够高的时候, 公司的价值则随着国有股权的增加而增加。Liu & Sun (2002) 认为, 前面的文献研究存在一个共同的问题——没有对公司的最终控制权进行区分, 国有股和法人股的分割本质上是不清晰的, 因为很多的法人控股股东仍然是国有企业, 因而他们的研究结论可能是具有误导性的。Liu & Sun (2002) 发现中国 84% 的上市公司是由国家最终控股的, 其中 8.5% 由国家直接控制, 75.6% 的公司国家进行间接控制。他们在区分最终控制权的基础上比较了多种政府治理方式的效果, 从中找出最没有效率的治理模式, 但遗憾的是, 他们没有分析国有与非国有控股对公司业绩的影响。

国内其他一些学者对股权结构与公司业绩关系的研究重点都集中在股权集中度和公司业绩的关系上。如孙永祥、黄祖辉 (1999) 发现, 股权集中度和公司业绩是 U 形的关系, 股权高度集中和股权高度分散相比有利于经营激励、收购兼并监督机制发挥作用, 因而该种股权结构有利于公司业绩的提高。施东晖 (2000) 则发现国有股和流通股比例与上市公司业绩之间没有显著的关系, 法人股东在公司治理中的作用是不一样的, 社会法人控股而股权分散的公司业绩好于法人控股型公司, 法人控股型公司的业绩则好于国有控股型公司。陈小悦、徐晓东 (2001) 的实证研究结果表明: 在外部投资者利益缺乏保护的情况下, 流通股比例与企业业绩之间负相关, 在非保护性行业第一大股东持股比例与企业业绩正相关, 国有股比例、法人股比例与企业业绩之间的相关关系不显著。

综合已有文献研究的结论, 我们发现, 关于国有股和法人股对公司业绩的影响并没有形成统一的结论, 研究的样本和方法对结果的影响都比较敏感。而与通常的感觉恰恰相反, 国有股和法人股对于公司业绩的影响是复杂的甚至是积极的, 基本没有可靠的证据证明股权的分割尤其是国有股和法人股的非流通损害了公司的绩效。我国尚处于经济转轨时期, 由于采用了并无先例的渐进改革模式, 很多已取得的改革成果至今仍被主流经济学界视为一个“谜”, 我国独有的“股权分割”的资本市场有其发生和发展的历史根源, 也遗留了很多棘手问题, 但不能仅仅因为我们的模式偏离了已有的私有产权下的全流通的模式而不假思索地把我国股票市场问题的根源归咎于股权分割。

从理论上讲, 经济转轨时期政府所有权和治理并非一定劣于私有产权下的治理结构, Qian (1996, 2000) 通过中国乡镇企业实证研究证实了以上结论。他认为私有产权及其治理只有在市场制度环境是完善的情况下才会优于政府所有权及其治理, 在转轨时期政府治理尽管不是最优的, 但却是不完善制度环境中的次优选择。我国尚处于这样一个过渡时期, 缺乏法律对于私有产权的保护, 缺乏一个功能良好的资本市场, 缺乏有效的税收和财政制度, 在这样的制度环境中, 政府所有权仍然具有一定的优越性。

近年来发展起来的法律与金融发展的理论在解释国与国之间股权市场发展的深度和广度取得了丰硕的成果 (LLSV^①, 1998, 1999, 2000a, 2000b, 2002), 他们认为法律体系对于外部投资者 (股东和债权人) 的保护是决定金融市场发展水平的决定性因素。投资者保

① LLSV 是 La Porta, Rafael, Florencio Lopez-de-silanes, Andrei Shleifer, and Robert W. Vishny 四位作者的简称。

护之所以重要，是因为投资者一旦作出投资决策以后，他们就面临公司控股股东和经理侵吞他们的财富的可能，法律保护是他们免遭内部人剥削的机制。只有法律体系对于外部投资者利益实施有效保护才能维护金融市场的诚信与发展，降低委托代理的成本和资本的成本，有利于经济的长期增长。不同法律起源的国家对外部投资者的保护程度不一，其中以英美的普通法系保护最好，而法国的大陆法系对投资者保护最弱，德国法系和北欧的斯堪的纳维亚法系居中。不同法律体系及其实施的效率决定了投资者保护的效率并最终决定了金融市场的发展水平。

私有产权和全流通的市场远非股票市场成功的保证。基于内部人和外部股东的信息不对称及合约的不完美性，控股股东和经理对外部投资者的剥削在很多国家都是广泛存在的。很多实证研究的成果都发现在西欧、东南亚、中东、拉美和非洲等地区很多上市公司都存在大的控股股东，通常是家族，这些家族和上市公司的管理具有紧密的联系纽带，很容易形成合谋，通过金字塔型的股权结构控制大量的资产，然后采用低红利政策、投资于无效项目、关联交易等手段把公司资金转向对他们有利的用途，从而实现对外部投资者财富的掠夺（LLSV' 2000a；Cheassen et al.，2000；Faccio & Lang，2002）。这种掠夺行为在法律无法有约束的情况下投资者只能通过“用脚投票”来保护自己的财富，金融危机因此应运而生。

中国股市1年多以来的下跌正是这种“用脚投票”的后果的集中体现。国有股“高价减持”、高价增发配股、大股东占用上市公司资金、上市公司不分红、低成本造假、“庄家”操纵市场等等无一不体现政府、控股股东以及公司内部人对中小投资者财富的剥削。而最为严重的是，投资者无奈地发现，他们在面对种种的剥削行为时差不多是手无寸铁——没有任何有效途径可以保障他们的合法利益。我们认为中国股票市场在经历多年的“超常规发展”以后，已经到了夯实投资者保护这一市场基础的时候了。

为了能够更加客观地揭示股权结构与公司业绩之间的关系，本文的实证研究回避了以往研究的一些不足。我们建立了1995~2001年的基本涵盖所有深沪上市公司的面板数据样本（panel data），使我们的样本具有更加广泛而充分的代表性。基于面板数据的分析技术，我们比较了多个非平衡样本和平衡样本结果的稳定性，使我们的研究结论的可靠性更强。以往的文献基本延续了Xu & Wang（1997，1999）静态的业绩决定模型的研究框架，而没有考虑公司上市后业绩动态变化的模式，因而难以深刻反映我国股票市场的根本特征。虽然静态的公司业绩对评价公司的价值固然重要，但公司业绩预期的变动同样是不可忽视的因素。本文对上市公司业绩的动态变迁模式研究将是一个重点，同时也是对以往文献的有益补充。

三、数据样本与研究方法

Xu & Wang（1997，1999）等研究成果所采用的数据样本无论在时间长度和样本的代表性上都是具有一定局限性的，从而也影响了他们的研究结论的稳定性。我们的样本包含了1995~2001年几乎所有的上市公司（除了少数几个上市公司因为数据缺失而被剔除）数据，数据主要来源于证券交易所和巨灵证券信息系统。作为以往研究的一个延续，我们

建立了下述结构性模型来检验公司业绩与股权分割的关系。

$$y_{it} = \alpha_0 + X_{it} + \text{pshare}_{it} + \epsilon_{it}$$

其中： y 代表公司业绩的变量，本文主要选择了资产收益率（ROA）、销售利润率（ROS）、净资产收益率（ROE）三个指标。我们放弃 Tobin Q 和市值/账面价值，因为在中国股市的投机性比较强，由于容易受人为因素的操控，上述指标的意义不强。 X 代表控制变量； pshare 为国有股、法人股和非流通股（国有股 + 法人股）不同性质的股权所占的比例。由于国有股和法人股的划分不能准确反映最终所有权的差异，我们不想过分强调国有股与法人股和业绩之间的关系，主要目的在于分析非流通股对业绩的影响。由于可能存在的区间效应（entrenchment effect），下述回归模型引入了股权分割变量的二次项，通过其系数的符号和显著性，我们可以检验是否存在股权变量的区间效应。

$$y_{it} = \alpha_0 + X_{it} + \text{pshare}_{it} + \text{pshare}_{it}^2 + \epsilon_{it}$$

为了描述公司上市后不同年份业绩动态的变动模式，我们借鉴 Wang, Xu & Zhu (2001) 的研究方法，进行下面回归模型的分析：

$$y_{it} = \alpha_0 + X_{it} + \sum_{j=0}^6 \gamma_j D_{itj} + \epsilon_{it}$$

其中： D 为公司上市后 j 年的虚拟变量，我们分别考虑了公司上市后 0~5 年以及 6 年以上的情况；对于 t 年公司 i 的统计数据，如果公司已经上市 j 年（上市当年为 0），则取 1，其他情况下为 0。通过该虚拟变量的设置，我们可以分析公司的业绩在上市以后变动的模式。

四、实证研究结果

回归所用到的主要变量（见表 1）描述如下。

我们的变量主要分为业绩变量、股权变量和常用的控制变量。由于 ROE（净资产收益率）被认为与证监会对公司增发配股的要求有关而往往容易遭到操纵，因此我们同时选择了 ROA（资产收益率）和 ROS（主营业务利润率）作为参照。

我们静态的公司业绩模型结果与 Tian (2001) 相似——国有股、法人股^① 和非流通股比例与公司业绩之间呈显著的 U 形关系（见表 2、表 3 和附表）。也就是在一定的持股水平之下，国有股、法人股或非流通股的增加，公司的业绩下降；在拐点以上，则随着非流通股股份比例的增加而上升。与通常的感觉相反，国有股“独大”或“独小”对公司的业绩是有利的，而适度混合的股权结构反而对公司的业绩最不利。总体而言，股权分割对公司业绩的影响是复杂的，呈非线性的关系。

① 法人回归结果见附表。

表 1 加归变量描述

变 量	定 义
业绩变量	
ROA (资产收益率)	公司利润总额/公司总资产 × 100%
ROE (净资产收益率)	公司净利润/公司净资产 × 100%
ROS (销售利润率)	利润总额/主营业务收入 × 100%
股权变量	
ST (国有股)	国有股/总股本 × 100%
LPS (法人股)	法人股/总股本 × 100%
RIL (非流通股)	(国有股 + 法人股) / 总股本 × 100%
控制变量	
LEV (负债比率)	公司总负债/公司总资产 × 100%
LTA (规模变量)	公司总资产的自然对数

表 2 非平衡样本模型回归结果——国有股与公司业绩

因变量	ROA (1)	ROA (2)	ROE (1)	ROE (2)	ROS (1)	ROS (2)
Intercept	20.086*** (0.000)	20.378*** (0.000)	72.127*** (0.000)	61.250*** (0.000)	- 9.553*** (0.020)	- 17.941*** (0.000)
ST	0.036*** (0.000)	- 0.169*** (0.000)	0.104*** (0.000)	- 0.371*** (0.000)	0.102*** (0.000)	- 0.222*** (0.000)
ST 2		0.003*** (0.000)		0.006*** (0.000)		0.004*** (0.000)
LEV	- 0.253*** (0.000)	- 0.252*** (0.000)	0.076*** (0.000)	0.053*** (0.000)	- 0.494*** (0.000)	- 0.524*** (0.000)
LTA	- 0.152 (0.553)	- 0.095 (0.710)	- 3.324*** (0.000)	- 2.504*** (0.000)	2.319*** (0.000)	2.960*** (0.000)
R Square	0.28	0.32	0.15	0.20	0.07	0.07
DW	1.76	1.83	1.91	1.93	2.04	2.05
Prob (F Stat)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
样本数	6102	6102	6102	6102	6093	6093

注：回归在 Eviews 3.1 panel Date 的环境中完成，括号中数据为系数显著性检验的概率，其中 * * *，* * 和 * 分别代表通过 1%，5% 和 10% 的显著性检验。系数方差采用了 White 的一致估计。

由于绝大部分上市公司的控股权在政府手中，我们认为产生这种 U 形关系的原因主要取决于国有股东在不同股权结构下的行为方式。从理论上讲，国有产权的低效率源于政府的非价值最大化目标以及激励的不相容。在国有股权占少数的情况下，政府对于公司的行政干预弱化，公司的运行更接近于市场化的公司。随着政府投票权的增加则行政干预有所加强，更多的公司资源被用于非价值量化的目标，公司的整体经营绩效下降。但随着这种控制权的增强，尤其是政府处于绝对的控股地位以后，行政干预达到了最大。而政府对公司进行有效治理的激励增强，维护上市公司的市场形象及市场的融资能力变得更加重要。因而政府往往会运用手中的经济资源对上市公司进行“补贴”，通过转换公司的不良

表 3 非平衡样本模型回归结构——非流通股与公司业绩

因变量	ROA (1)	ROA (2)	ROE (1)	ROE (2)	ROS (1)	ROS (2)
Intercept	8.507*** (0.003)	10.866*** (0.000)	26.802*** (0.000)	34.225*** (0.000)	- 57.500*** (0.000)	- 50.999*** (0.000)
ST	0.118*** (0.000)	- 0.321*** (0.000)	0.289*** (0.000)	- 0.510*** (0.000)	0.291*** (0.000)	- 0.425*** (0.000)
ST ²		0.004*** (0.000)		0.007*** (0.000)		0.006*** (0.000)
LEV	- 0.247*** (0.000)	- 0.264*** (0.000)	0.040*** (0.000)	0.005*** (0.055)	- 0.564*** (0.000)	- 0.604*** (0.000)
LTA	- 0.126 (0.624)	- 0.562* (0.056)	- 1.730*** (0.000)	- 0.971*** (0.000)	4.113*** (0.000)	4.833*** (0.000)
R Square	0.32	0.37	0.19	0.23	0.08	0.08
DW	1.85	1.92	1.94	1.95	2.05	2.06
Prob (F Stat)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
样本数	6052	6052	6053	6053	6043	6043

注：回归在 Eviews 3.1 panel Date 的环境中完成，括号中数据为系数显著性检验的概率，其中 * * *，* * 和 * 代表通过 1%，5% 和 10% 的显著性检验。系数方差采用了 White 的一致估计。

资产等方式为上市公司“买单”，维持上市公司的业绩。相反，在股权结构适度混合的公司中，由于控股股东所占份额不大或存在多个份额相当的大股东，这种情况下不同利益团体的治理激励都不会很高，更容易产生“搭便车”问题，在上市公司资源被掏空以后，剩下一些没有管的“烂摊子”，对上市公司的业绩是最不利的。在我国经济转轨时期，由于政府仍然控制了相当多的经济资源，仍然有能力维持对上市公司的“补贴”，与政府“良好的关系”使上市公司在经营中能够得到更多的好处。因此，尽管在理论上国有产权的经营效率不是最优的，但在制度不完善的过渡时期仍然有可能是一种次优的选择。它对于公司业绩的影响是复杂的，还取决于上市公司的行业等诸多因素。对股权分割问题的分析不能脱离我国特定的经济环境，如果仅仅因为我们的股权结构是对成熟市场的私有产权下的全流通模式的偏离而断定股权分割问题是我国股票市场问题的核心，我们认为至少在实证上是缺乏依据的。

采用 Wang, Xu & Zhu (2001) 的研究方法，我们对我国上市公司上市后业绩的变动进行了研究。我们分别考察了公司上市后 0~5 年以及 6 年以后的业绩，结果显示我国上市公司的整体经营业绩呈现显著的逐年加速下滑的趋势（见表 4、表 5）。平衡样本和非平衡样本的结果完全一致，结果的稳定性相当好， $T_0 \sim T_{6+}$ 的系数全部都通过 1% 的显著性检验。我们进而实证检验了没有国有股和国有股“一股独大”两个子样本，发现公司业绩加速下滑的状况完全没有改变，股权结构对于业绩下滑的作用不显著。

表 4 上市公司业绩的动态变化

因变量	ROA (1)	ROA (2)	ROE (1)	ROE (2)	ROS (1)	ROS (2)
	(Unbalanced)	(Balanced)	(Unbalanced)	(Balanced)	(Unbalanced)	(Balanced)
Intercept	5.633* (0.052)	- 11.190*** (0.045)	13.854*** (0.000)	- 0.998*** (0.002)	- 62.369*** (0.000)	- 108.451*** (0.000)
T ₀	- 8.892*** (0.000)	- 5.700*** (0.000)	- 15.339*** (0.000)	- 13.086*** (0.000)	- 8.535*** (0.000)	- 6.780*** (0.000)
T ₁	- 8.998*** (0.000)	- 8.704*** (0.000)	- 15.668*** (0.000)	- 13.504*** (0.000)	- 14.967*** (0.000)	- 14.484*** (0.000)
T ₂	- 10.132*** (0.000)	- 9.627*** (0.000)	- 17.377*** (0.000)	- 15.191*** (0.000)	- 12.210*** (0.000)	- 12.440*** (0.000)
T ₃	- 11.351*** (0.000)	- 11.019*** (0.000)	- 21.834*** (0.000)	- 19.765*** (0.000)	- 18.177*** (0.000)	- 17.750*** (0.000)
T ₄	- 12.265*** (0.000)	- 11.892*** (0.000)	- 23.698*** (0.000)	- 21.917*** (0.000)	- 27.209*** (0.000)	- 26.169*** (0.000)
T ₅	- 12.482*** (0.000)	- 12.153*** (0.000)	- 19.988*** (0.000)	- 19.168*** (0.000)	- 18.003*** (0.000)	- 17.210*** (0.000)
T ₆₊	- 11.666*** (0.000)	- 11.648*** (0.000)	- 27.467*** (0.000)	- 22.342*** (0.000)	- 6.058*** (0.000)	- 19.047*** (0.000)
LEV	- 0.258*** (0.000)	- 0.272*** (0.000)	- 0.001*** (0.785)	- 0.056*** (0.000)	- 0.588*** (0.000)	- 0.615*** (0.000)
LTA	1.063*** (0.000)	1.904*** (0.000)	0.538*** (0.000)	1.281*** (0.000)	5.820*** (0.000)	8.128*** (0.000)
R Squsre	0.40	0.38	0.24	0.22	0.08	0.08
DW	1.95	1.99	1.98	1.99	2.05	2.09
Prob (F Stat)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
样本数	6310	3171	6306	3171	6300	3136

注：回归在 Eviews 3.1 panel Date 的环境中完成，括号中数据为系数显著性检验的概率，其中 * * *，* * 和 * 代表通过 1%，5% 和 10% 的显著性检验。系数方差采用了 White 的一致估计。

我们的实证结果验证了关于上市公司“第一年绩优、第二年绩平、第三年绩差”的说法。上市公司业绩的恶化在我国是一种系统性的现象，并不因为股权结构的差异而有所改变。上市公司业绩的恶化和中国宏观经济的高增长形成了强烈的反差。另外一些研究则证明，新股的上市对我国股票指数连年的上涨起到了很大的推动作用，通过投资中国股市不一定能够分享中国高经济增长的成果。

上市公司业绩整体的恶化导致我国股市的价值基础的丧失，它直接影响了投资者对股市的长期信心，对股市的打击无疑是最致命的。Ritter (1991) 认为，公司上市以后业绩下滑可能有两个原因：①公司上市前的业绩被夸大。这种现象在中国无疑是存在的，但它一般只会导致业绩的短暂下滑而不是连续。②委托代理成本增加。公司的 IPO 导致股东和经理人或者控股股东和外部投资者的利益冲突，公司的内部人把公司的资源转向对己有利的用途，实质上形成了对外部投资者的财富的掠夺。委托代理问题即使在私有化和全流通

表 5 回归结果稳定性检验

因变量	ROA (无国有股样本)	ROA (国有股 > 50% 的样本)
Intercept	- 9.911 (0.000)***	- 0.671 (0.906)
T ₀	- 7.024 (0.000)***	- 4.493 (0.000)***
T ₁	- 9.858 (0.000)***	- 6.902 (0.000)***
T ₂	- 10.543 (0.000)***	- 8.974 (0.000)***
T ₃	- 12.246 (0.000)***	- 10.554 (0.000)***
T ₄	- 11.721 (0.000)***	- 10.999 (0.000)***
T ₅	- 13.472 (0.000)***	- 12.437 (0.000)***
T ₆₊	- 12.859 (0.000)***	- 11.180 (0.000)***
LEV	- 0.255 (0.000)***	- 0.189 (0.000)***
LTA	1.851 (0.000)***	1.151 (0.001)***
R Square	0.350	0.390
DW	1.880	2.390
Prob (F Stat)	0.000	0.000
样本数	1008	658

注：回归在 Eviews 3.1 panel Date 的环境中完成，括号中数据为系数显著性检验的概率，其中 * * *，* * 和 * 代表通过 1%，5% 和 10% 的显著性检验。系数方差采用了 White 的一致估计。

的市场中同样是存在的，LLSV (1999)，Cheassen et al. (2000)，Faccio & Lang (2001) 证明了家族控制的上市公司对外部股东的剥削是广泛存在的。对于委托代理问题的约束取决于外部的法律和制度环境，在普通法系的国家对私有财产的保护要优于大陆法系国家。LLSV (2000a) 证明了作为委托代理成本的一个指标——红利支付比率是投资者保护机制的结果，在投资者保护比较完善的普通法系国家，红利支付比率要显著高于投资者保护较差的大陆法系国家。强有力的投资者保护机制迫使公司控制者把公司利润以红利的方式归还给股东，避免了他们把多余的现金投资于净现值为负的项目或增加经理人的在职消费，损害投资者利益。

我国股票市场上的委托代理问题是非常突出的，上市公司的不分红现象就是一个很好的明证。从理论上讲，只有现金红利或现金股份回购才是真正意义上的分红。而我国上市公司钟情的送股和转增股本的方式因为不涉及资金控制权的转移而不属此类。采用 LLSV (2000a) 的红利支付比率 (DTE) 指标，我们发现，2001 年以前我国上市公司 70% 以上采取了不分红政策，红利支付比率中位值为 0。2001 年，由于证监部门在上市公司再融资要求现金红利的要求后，才有所改善，但整体水平仍然低于 LLSV (2000a) 样本中红利率较低的大陆法系国家的平均水平（见表 6）。低红利政策和上市公司业绩弱化说明上市公司在没有良好的投资机会的情况下，仍然占用投资者的资金，进行低效率的投资，实质上构成了对外部投资者的财富的掠夺。而大股东和公司管理层通过占用上市公司资金、不公平的关联交易、弄虚作假等手段掏空上市公司、损害中小股东利益的事件更是屡见不鲜。

表 6 上市公司分红比率 (DTE) 的分布

	1995	1996	1997	1998	1999	2000	2001
dte = 0	65.92	77.45	74.97	71.68	71.21	82.23	39.64
0 < dte ≤ 30%	6.33	5.60	5.79	5.71	6.05	0.92	15.52
50% ≥ dte > 30%	10.43	4.51	5.78	7.13	8.15	0.92	20.38
dte > 50%	17.32	12.44	13.46	15.48	14.59	15.93	24.46
meadian	0	0	0	0	0	0	24.30
observations	537	643	847	911	1042	1092	1196

Stigliz (1994) 认为, 委托代理问题的本质不会因为所有权性质的改变而改变。它取决于外部的法律和制度环境对当事人违规行为的约束能力, 以及当事人违规行为可能承担的成本。外部的约束不能避免违规行为的发生, 但能够令违规者为自己的行为付出惨重的代价。即使在市场制度最完善的美国, 仍然会爆发“安然公司倒闭”事件, 但是我们也应该看到, 参与事件的安然公司的高层和安达信会计师事务所受到了相当严厉的惩罚。只有严刑峻法和高昂的违规成本才是维持市场诚信的有效手段。

我国股市严重的委托代理问题的根本症结不在股权分割, 即使在所有股权全流通的情况下, 外部监督机制的缺位仍然不能防范大股东剥削等行为的发生。“欧亚农业”事件以及其他内地民企—香港股票市场上的行为就是一个很好例子, 在没有良好的约束机制的情况下, 国有股东和民营股东的行为方式并没有本质上的差别。我国股票市场的问题恰恰在于监管机制的缺位以至违规成本太低, 无法有效保护投资者的利益。

这与我国法律体系的根源和司法效率是息息相关的。我国的法律体系源于日本的模式, 并融合了许多德国法律体系的特征, 本质上属于大陆法体系。根据 LLSV (1998) 的研究, 基于条文法的大陆法体系对私有产权和投资者的保护弱于基于案例法的普通法体系。尤其是在我国经历了几十年的以公有制为基础的计划经济阶段, 私有产权一度遭到政治上的否定, 即使在改革开放 20 多年以后仍然没有从法律上对私有财产的合法权益进行确认, 对投资者利益的保护的程序更无从谈起。因此, 在我国投资者利益受损害以后, 仍然处于投诉无门的状态。加上我国法律体系的独立性不高, 仍然容易受到行政干预, 在腐败盛行的转轨时期, 法律实施效率不高也是可以想像的。我国证监会的设立在规范上市公司行为和保护投资者利益方面起到了很大的作用, 但是行政干预毕竟近似于黑箱操作, 更容易演变成为不同利益团体的寻租场。因此, 我们认为在行政监督以外引入相对透明的法律诉讼机制是非常必要的。

大陆法体系虽然在投资者保护问题上天然地劣于普通法体系, 但这种特点并非是一成不变的, 通过两种体系优点的融合仍然可以起到扬长避短的效果。著名的证券监管专家朗咸平教授 (2002) 认为, 在大陆法体系中引进“辩方举证”和“集体诉讼”是弥补投资者保护不足的关键。美国 1929 年股市崩盘以后同样遭遇到了巨大的信心危机, 以肯尼迪 (Joseph Kennedy) 为首的第一任证监会主席及其继任者就是通过坚定不移地推行“辩方举证”和“集体诉讼”重建了投资者的信心, 并奠定了美国股市健康发展的百年基业。肯尼迪认为证券法规可以写得非常漂亮, 但关键仍在于它的实施, 而“辩方举证”和“集体诉讼”就是法律执行的利器, 它真正赋予了受损的投资者利用法律来保护自身利益的权力,

而不仅是“用脚投票”。在我国目前司法效率并不高的情况下无疑具有更加重要的意义，而且在技术上也不存在太大的难题。目前，我国在“医疗事故”诉讼方面已经引入了“辩方举证”机制。通过推行投资者保护方面的立法和司法才是促进我国股票市场健康发展的长远之计。

五、结论

本文的实证研究表明股权分割对上市公司业绩的影响是复杂的，并没有足够的证据证明股权分割损害了公司的业绩。我国上市公司在上市后呈现加速恶化的趋势，而这种趋势并不因为股权结构的差异而有所改变。我们认为，把中国股市问题的核心归咎于股权分割是具有误导性的，私有产权和全流通并不足以保证市场的成功。本文的结论并非要否定国有资产退出竞争性领域的大方向，事实上“十六大”以后国有资产的大规模退出也是我国渐进式经济转轨能否取得成功的关键步骤。但问题的关键不在于“是否退出”，而在于如何实现国有资产以“公平、公正、公开”的方式退出，在尽量保持社会稳定的前提下，实现经济制度的平稳过渡。国有股和非流通股问题的解决曾经有过良好的契机，但是往往由于所推出的方案不能得到市场的认同而受到市场力量极大的反弹。我们认为股权分割问题的解决必须以有效的投资者保护为基础，当务之急是我国大陆法系的架构下推动立法保护私有产权的合法权益，明确国有资产处理的根本原则，并通过引入“集体诉讼”和“辩方举证”机制加强投资者保护的效率，打击各种制度变迁过程中的剥削行为，重建投资者的信心。

参考文献

- [1] 陈小悦，徐晓东：《股权结构、企业绩效与投资者利益保护》，《经济研究》，2001年第11期。
- [2] 朗咸平：《保护投资者振兴股市新思维》，《新财富》，2002年第2期。
- [3] 施东晖：《股权结构、公司治理与绩效表现》，《世界经济》，2000年第2期。
- [4] 孙永祥，黄祖辉：《上市公司的股权结构与绩效》，《经济研究》，1999年第12期。
- [5] Chen, Jian, Ownership Structure as Corporate Governance Mechanism: Evidence from Chinese Listed Companies, *Economics of Planning*, 2001, Vol.34, No.1/2.
- [6] Claessens, Stijn, Simeon Djankov, and Larry H. P. Lang, Expropriation of Minority Shareholders in East Asia, *Journal of Financial Economics*, 2000, Vol.58, pp1 ~ 112.
- [7] Faccio, Mara, Larry H. P. Lang, and Leslie Young, Dividends and Expropriation, *American Economic Review*, 2001, Vol.91, Issue 1, pp54 ~ 78.
- [8] La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny, Legal determinants of external finance, *Journal of Finance*, 1997, Vol.52, pp1131 ~ 1150.
- [9] La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny, Law and Finance, *Journal of Political Economy*, 1998, Vol.106, pp1113 ~ 1155.
- [10] La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer, Corporate Ownership around the World, *Journal of Finance*, 1999, Vol.54, pp471 ~ 517.
- [11] La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert W. Vishny, Agency Problems

- and Dividend Policies around the World, *Journal of Finance*, February, 2000a, Vol.55, No1, pp1 ~ 3.
- [12] La Porta, R. F. Lopez-de-Silanes, A. Shleifer, and R. W. Vishny, Investor Protection and Corporate Governance, *Journal of Financial Economics*, 2000b, Vol.58, pp3 ~ 27.
- [13] La Porta, R. F. Lopez-de-Silanes, A. Shleifer, and R. W. Vishny, Investor Protection and Corporate Valuation, Forthcoming in *Journal of Finance*, 2002.
- [14] Liu, Guy & Sun Pei, The Class of Shareholdings and Its Impacts on Corporate Performance—A Case of State Shareholding Composition in Chinese Publicly Listed Companies, Working Paper of Department of Economics and Finance, Brunel University, 2000.
- [15] Qian, Yingyi, Enterprise Reform in China: Enterprise Problems and Political Control, *Economics of Transition*, 1996, Vol.4, No.2, pp427 ~ 447.
- [16] Qian, Yingyi, Government Control in Corporate Governance as a Transitional Institution: Lessons from China, Working Paper of University of Maryland, 2000.
- [17] Ritter, Jay, The Long-run Performance of Initial Public Offering, *Journal of Finance*, 1991, Vol.46, No.1, pp3 ~ 7.
- [18] Shleifer, A. and D. Wolfenson, Investor protection and equity markets, Forthcoming in *Journal of Financial Economics*, 2002.
- [19] Stiglitz, Joseph E., *Whither Socialism*. Cambridge, MA: The MIT Press, 1994.
- [20] Sun, Qian, Wilson Tong & Jing Tong, How Does Government Ownership Affect Firm Performance? Evidence from China's Privatization Experience, *Journal of Business Finance & Accounting*, 2002, Vol.29, No.1 & 2.
- [21] Tian, Lihui, Government Shareholding and the Value of China's Modern Firms, William Davidson Institute Working Paper 395, University of Michigan Business School Williamson, Oliver 1985, *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*, Free Press, New York, 2001.
- [22] Wang Xiaozu, Xu Lixin Colin & Zhu Tian, State-Owned Enterprises Going Public: The Case of China, Working Paper of Hong Kong University of Science and Technology, 2001.
- [23] Xu, Xiaonian & Yan Wang, Ownership Structure, Corporate Governance and Corporate Performance: The Case of Chinese Stock Companies, World Bank Policy Research Working Paper 1794, June, 1997.
- [24] Xu, Xiaonian & Yan Wang, Ownership Structure and Corporate Governance in Chinese Stock Companies, *China Economic Review*, 1999, Vol.10, No.1

附表 法人股与公司业绩

因变量	ROA (1)	ROA (2)	ROE (1)	ROE (2)	ROS (1)	ROS (2)
Intercept	17.171*** (0.000)	19.857*** (0.000)	75.194*** (0.000)	76.339*** (0.000)	- 7.580*** (0.054)	- 7.397*** (0.058)
LPS	0.024*** (0.000)	- 0.150*** (0.000)	0.031*** (0.000)	- 0.363*** (0.000)	0.035*** (0.000)	- 0.189*** (0.000)
LPS* 2		0.002*** (0.000)		0.005*** (0.000)		0.003*** (0.000)
LEV	- 0.249*** (0.000)	- 0.250*** (0.000)	0.100*** (0.000)	0.078*** (0.000)	- 0.465*** (0.000)	- 0.487*** (0.000)
LTA	0.003 (0.989)	- 0.070* (0.777)	- 3.375*** (0.000)	- 3.224*** (0.000)	2.300*** (0.000)	2.435*** (0.000)
R Squsre	0.27	0.30	0.13	0.16	0.07	0.07
DW	1.75	1.80	1.89	1.91	2.04	2.04
Prob (F Stat)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
样本数	6100	6100	6100	6100	6091	6091

中国股市收益、收益波动与投资者情绪*

王美今 孙建军

一、引言

金融风险资产价格对其价值的偏离究竟在多大的程度上能被套利者消除，这一问题近年来金融研究热点之一。在此类研究中，De Long et al. (1990) 提出的噪声交易者模型 (DSSW) 具有较大影响力；该模型表明：在有限套利的环境中，如果投资者情绪相互影响，套利者将无法消除非理性行为导致的错误定价，投资者情绪因而会成为影响金融资产均衡价格的系统性风险。DSSW 模型的意义还在于，它将封闭式基金折价现象作为例子运用自身理论进行剖析，从而引发了对“噪声交易者风险影响金融资产均衡价格”这一命题的广泛实证研究。内容涉及投资者情绪对收益的影响、投资者情绪对收益波动的影响等诸多方面。Lee et al. (1991)、Swaminatham (1996) 与 Neal wheatley (1998) 认为封闭式基金折价可以作为投资者情绪的代理变量。Brown (1999) 运用美国“个人投资者协会”(AAII) 提供的投资者情绪指数，证实投资者情绪与封闭式基金价格的波动紧密相关。Simon 与 Wiggins (1999) 考察了短期市场收益预测中封闭式基金折价代理投资者情绪的作用以及投资者情绪调查指数、技术指标的作用。Lee et al. (2002) 发现投资者情绪是影响价格的系统性因子，超额收益与情绪变化同期相关，情绪变化影响收益波动。Brown 与 Cliff (2004) 认为情绪的水平值、变化量与市场收益强相关。

从国内的研究来看，刘焯辉等 (2003)、施东晖与孙培源 (2002)、宋军等 (2003) 和李心丹等 (2002) 分别从不同的侧面，揭示我国股票市场上投资者心理与行为的影响作用，为研究投资者情绪与我国股票市场的关系提供了投资者情绪相互关联的客观依据。尽管我国封闭式基金也存在折价现象，但对它的解释至今仍没有较一致的观点 (金晓斌等，2002；张俊喜和张华，2002；刘焯辉和熊鹏，2003)。饶育蕾与刘达锋 (2003) 用央视机构看市与中证报机构看市水平值研究投资者情绪水平与未来收益的相关关系，发现统计上并不显著。

本文从我国股市的现实情况出发，推导出一个更具一般性的模型，证明投资者接受价格信号时表现出来的情绪，在一定条件下可能形成影响均衡价格的系统性因子；进而构造投资者情绪指数，建立 TGARCH - M (1, 1) 证实了理论模型的结论。实证结果表明沪深两市不仅具有相同的投资者行为特征和风险收益特征，而且均未达到弱式有效，我国机构

* 本文为国家自然科学基金课题 (批准项目号：70273060) 的阶段性研究成果。原载：《经济研究》，2004 年第 10 期。

投资者也是可能的噪声风险源。

本文的结构安排如下：第二部分构造理论模型，揭示投资者情绪对股票均衡价格的影响作用及其存在条件；第三部分说明数据来源、投资者情绪指数的构造方法及结果；第四部分设计实证模型，检验投资者情绪与收益及收益波动的关系并讨论其引申含义；第五部分总结全文。

二、理论模型

DSSW 是跨期代际结构模型，假设理性投资者考虑套利的长度是两期，也就是说套利者关心的是资产临时的转售价（interim resale prices）而非分红收益。此外，能否为某种既定的证券找到完全相同或近似的替代品是套利能否发挥作用的关键。Shleifer（2000）指出，当资产并非完全替代品时，代际模型的结构就不需要了。我国不存在卖空机制；证券没有明显合适的替代组合，套利者为了获得较高收益，只能简单地卖出或减持股票，套利行为缺乏有效性；上市公司很少分红，投资者买卖股票只是为了获得买卖差价。基于这些考虑，本文没有沿用套利的分析视角，而是将噪声交易者进一步细分为两大类：受情绪影响的噪声交易者和其他交易者^①，着眼于投资者在接受价格信号时可能产生的心态，设计一个单期模型，研究情绪型噪声交易者的交易活动对我国股市的影响。

假设在一个单期竞争性市场上有两种证券：一种是无风险的债券，其单位支付为 1；另一种是股票，其价格为 θ ， $\theta \sim N(\bar{\theta}, \sigma_\theta^2)$ 。同时假设市场上有三个投资者：理性投资者、噪声交易者与其他交易者^②。理性投资者与噪声交易者在期初都会收到关于股票价格的信号 s 。 s 由两部分组成：股票价格的真实信号与噪声信号。噪声信号不能影响理性投资者，但能引起噪声交易者产生情绪变化。用随机变量 ϵ 表示噪声交易者由于情绪变化而产生的对股票的错误估价， θ 与 ϵ 相互独立，则噪声交易者对信号 s 的理解是： $s = \theta + \epsilon$ ， $\epsilon \sim N(\bar{\epsilon}, \sigma_\epsilon^2)$ ， $\bar{\epsilon} \neq 0$ 表示噪声交易者的错误估价不能相互抵消。因此， $\epsilon - \bar{\epsilon}$ 越大，表明噪声交易者越乐观；相反，表明噪声交易者越悲观。 $\bar{\theta} + \bar{\epsilon}$ 表示噪声交易者对股票价格的评价； $\sigma_\theta^2 + \sigma_\epsilon^2$ 则表示由于噪声交易者的情绪变化导致股票价格偏离正常状态而加大了股票交易的风险， σ_ϵ^2 是噪声交易者风险的一种表现。理性投资者能从信号 s 中排除噪声信号的影响，对 s 的理解仍然是 $\theta \sim N(\bar{\theta}, \sigma_\theta^2)$ 。

为了达到期末财富期望效用最大化，两类投资者基于自己对收到的信号的理解选择债

① 本文将噪声交易者定义为受情绪（sentiment）影响而对风险资产未来收益分布形成错误理念的投资者（Shleifer, 2000）；其他交易者是指为了变现等流动性需求以及盲目跟风的交易者，他们本质上也是一种噪声交易者，但为了与前者相区别，称为其他交易者。Vandana（2001）将情绪定义为投资者关于投资的态度和看法；Brown 与 Cliff（2004）进一步将投资者情绪定义为市场参与人的预期与标准状态的比较。我们认为，投资者情绪是投资活动中一种“基于情感的判断”，由于心理或认知上的偏差而产生，导致对风险资产未来收益分布形成错误看法。Vandana, Brown 与 Cliff 的定义方法将“基于情感的判断”而对风险资产未来收益分布形成错误理念”具体化，以便于描述。因此，本文将“投资者情绪”具体化为噪声交易者对风险资产的估价与理性投资者的偏差。在实证分析时，由于涉及情感的价值难以测度，上述研究中用预期价格的涨跌作为对情绪的代理变量，这也给本文提供了借鉴。

② 投资者情绪相互关联，故同类型的交易者可以同质假定，为了分析上的方便，本文假定市场仅存在能代表这三类投资者的三个交易者。

券与股票组合。设期初财富为 w_{0i} ($i = r$ 表示理性投资者, $i = n$ 表示噪声交易者), 期末财富为 $w_{1i} = w_{0i} + X_i (\theta - \theta_0)$, 其中 θ_0 表示股票初始的购买价格, X_i 表示对股票的需求量 ($X_i > 0$ 表示买入, $X_i < 0$ 表示卖出)。假定理性投资者与噪声交易者都具有指数型的效用函数, 即 $U(w) = -e^{-aw}$, 其中 $a > 0$, 是绝对风险厌恶系数。 θ 的正态性决定了 w_{1i} 服从正态分布, 故财富的期望效用可以表示成财富的均方期望效用函数。

因理性投资者对信号的理解不受噪声的干扰, $E_r(\theta | s) = E_r(\theta) = \bar{\theta}$, 故:

$$\begin{aligned} E(U(w_{1r} | s)) &= -\exp\left\{-a\left[E_r(w_{1r} | s) - \frac{a}{2}\text{Var}_r(w_{1r} | s)\right]\right\} \\ &= -\exp\left\{-a\left[w_{0r} + X_r(E_r(\theta | s) - \theta_0)\right] - \frac{a^2}{2}X_r^2\sigma_\theta^2\right\} \\ &= -\exp\left\{-a\left[w_{0r} + X_r(\bar{\theta} - \theta_0)\right] - \frac{a^2}{2}X_r^2\sigma_\theta^2\right\} \end{aligned} \quad (1)$$

由极值条件知, 理性投资者对股票的最优需求量:

$$X_r = \frac{\bar{\theta} - \theta_0}{a\sigma_\theta^2} \quad (2)$$

在给定信号下, 噪声交易者关于财富期望效用:

$$\begin{aligned} E(U(w_{1n} | s)) &= -\exp\left\{-a\left[E_n(w_{1n} | s) - \frac{a}{2}\text{Var}_n(w_{1n} | s)\right]\right\} \\ &= -\exp\left\{-a\left[w_{0n} + X_n(E_n(\theta | s) - \theta_0) - \frac{a}{2}X_n^2\text{Var}_n(\theta | s)\right]\right\} \end{aligned} \quad (3)$$

由极值条件知, 噪声交易者对股票的最优需求量:

$$X_n = \frac{E_n(\theta | s) - \theta_0}{a\text{Var}_n(\theta | s)} \quad (4)$$

因为:

$$E_n(\theta | s) = \bar{\theta} + \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\epsilon^2}(\theta + \epsilon - \bar{\theta} - \bar{\epsilon}), \text{Var}_n(\theta | s) = \frac{\sigma_\theta^2\sigma_\epsilon^2}{\sigma_\theta^2 + \sigma_\epsilon^2} \quad \textcircled{1}$$

$$\text{故 } X_n = \frac{\bar{\theta} + \eta(\theta + \epsilon - \bar{\theta} - \bar{\epsilon}) - \theta_0}{av} \quad \text{其中: } \eta = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\epsilon^2}, v = \frac{\sigma_\theta^2\sigma_\epsilon^2}{\sigma_\theta^2 + \sigma_\epsilon^2} \quad (5)$$

令其他投资者对股票的需求量为 X ($X > 0$ 表示买入, $X < 0$ 表示卖出), 在期内不存在新股发行与配股, 则市场出清的条件为:

$$X_r + X_n + X = 0 \quad (6)$$

将 (2) 式与 (5) 式代入 (6) 式得:

$$\theta_0 = \frac{2\sigma_\epsilon^2\bar{\theta} + \sigma_\theta^2\theta + \sigma_\theta^2\epsilon - \sigma_\theta^2\bar{\epsilon} + aX\sigma_\theta^2\sigma_\epsilon^2}{\sigma_\theta^2 + 2\sigma_\epsilon^2} \quad (7)$$

① 依联合分布性质: 若 X 与 Y 是联合正态分布, 则有:

$$E(X | Y = Y) = EX + \frac{\text{Cov}(X, Y)}{\text{Var}(Y)}(Y - EY), \text{Var}(X | Y = Y) = \text{Var}(X) - \frac{(\text{Cor}(X, Y))^2}{\text{Var}(Y)}$$

采用 DSSW 模型将期内价格 θ 的分布等同于 θ_0 的方法, 化简 (7) 式:

$$\theta_0 = \bar{\theta} + \frac{\sigma_\theta^2(\epsilon - \bar{\epsilon})}{2\sigma_\epsilon^2} + \frac{aX\sigma_\theta^2}{2} \quad (8)$$

(8) 式表明, 股票的均衡价格, 首先取决于自身的分布, σ_θ^2 越大, 均衡价格偏离其基本价值越远。噪声交易者对股票均衡价格的影响由第二项“价格压力效应 (Price Pressure Effect)”^① 表现出来。由 $\partial\theta_0/\partial(\epsilon - \bar{\epsilon}) > 0$ 知, 在其他条件不变的情况下, 噪声交易者乐观程度与均衡价格同向变化。出现这一同向变化的原因在于 $\partial X_n/\partial(\epsilon - \bar{\epsilon}) > 0$, 即乐观情绪导致噪声交易者增加对股票的需求量, 这一需求量的变化推动均衡价格上升; 反之, 悲观情绪推动均衡价格下降。式中的 σ_θ^2 显示因情绪引致的噪声交易者风险反向修正均衡价格, 也就是说, 买入这种定价有偏差的资产所带来的收益被额外承担的价格风险抵消。尽管潜在的“价格压力效应”存在, 但其对均衡价格的影响 (即“价格压力效应”能否显现) 并不确定。原因在于均衡价格的形成还受到其他交易者需求量 X 的影响。 $\partial\theta_0/\partial X > 0$ 表明即使噪声交易者乐观, 但由于其他噪声交易者需求量的减少也可能导致均衡价格的不变或下降。

进一步分析理性投资者和噪声交易者的需求关系可以说明这一问题。将 (5) 式写成:

$$X_n = \frac{\bar{\theta} - \theta_0}{av} + \frac{\eta(\theta + \epsilon - \bar{\theta} - \epsilon)}{av}$$

将 η 与 v 代入, 化简后得:

$$\begin{aligned} X_n &= \frac{\bar{\theta} - \theta_0}{a\sigma_\theta^2} \left(1 + \frac{\sigma_\theta^2}{\sigma_\epsilon^2} \right) + \frac{(\theta - \bar{\theta}) + (\epsilon - \bar{\epsilon})}{a\sigma_\epsilon^2} \\ &= X_r \left(1 + \frac{\sigma_\theta^2}{\sigma_\epsilon^2} \right) + \frac{(\theta - \bar{\theta}) + (\epsilon - \bar{\epsilon})}{a\sigma_\epsilon^2} \end{aligned} \quad (9)$$

从 (9) 式看出, 当信号中股票市场价格不小于其基本价值, 即 $\theta \geq \bar{\theta}$, 噪声交易者充满乐观情绪时 ($\epsilon > \bar{\epsilon}$), 噪声交易者的需求量超过理性投资者; 反之, 市场低迷、噪声交易者一片悲观时, 噪声交易者的需求量可能小于理性投资者。而且, 作为理性投资者, 只有当它确信交易能获取资本利得时交易才可能发生; 换言之, 理性投资者识别了噪声交易者的错误估值后, 交易很可能无法进行, 但由于有足够多的其他噪声交易者的存在, 使得市场可能出清。我国股市的个体投资者中, 跟风从众、盲目信赖专家的现象很普遍^②, 因而该模型更适用于描述我国噪声交易者情绪的变化对股票市场的影响, 本文将在下面两个部分对 (8) 式进行实证检验。

综上所述, 我们认为该模型与模型 DSSW 模型相比, 有以下两个主要不同点:

第一, 放弃了套利的条件, 在更一般的背景下讨论问题。

第二, 将噪声交易者进一步细分成受情绪影响的噪声交易者和其他交易者, 后者的需

① 价格压力效应 (De Long et al., 1990) 指噪声交易者看淡后市, 对风险资产的需求量减少, 从而降低均衡价格; 噪声交易者看好后市, 对风险资产的需求量增加, 从而推动均衡价格上升。

② 李心丹等 (2002) 在两次大规模的个体证券投资者问卷调查中都发现个体证券投资者对专家跟风是普遍现象。宋军等 (2003) 认为目前中国证券市场中普遍存在着个体证券投资者对机构投资者的羊群行为, 并对此建立了头羊-从羊模型。

求是外生的,或由于流动性要求,或出于盲目跟风;他们与前者区别是不会衡量最终的财富水平。

三、数据来源和投资者情绪指数构造

(一) 数据来源

如果投资者情绪,即投资者投资决策中“基于情感的判断”在市场上广泛存在,则其效应不可避免地会由沪深两市表现出来,所以本文选取上综指与深成指日收益率数据进行研究。样本期自2001年4月23日至2003年10月31日,这与构造投资者情绪指数所能取得样本期一致。复权的收益率日线数据由南方基金提供。

(二) 投资者情绪指数构造

对(8)式进行实证检验的先决步骤是构造投资者情绪指数。从国外同类研究来看,常用的投资者情绪指数代理变量有开放式基金净赎回、情绪调查指数与封闭式基金加权折价率,运用得比较频繁和成功的是后两个。

情绪调查指数一般可以分为三种:一是“卖方指标”(the sell-side indicator),用华尔街主要分析师给出的资产配置建议数据编制,是一个由Merrill Lynch公司制作的月度反向指标,反映机构投资者情绪;二是“投资者智慧”(investors intelligence)指数,基于被调查的投资专家中看熊的比例而编制,是一个由Chartcraft投资服务公司制作的周指标,反映中等规模投资者情绪;三是“美国个人投资者协会”(American Association of Individual Investors)指数,基于被调查的协会会员中看熊的比例而编制,是一个反映个体证券投资者情绪的周指标。

在国外,封闭式基金加权折价率也往往作为投资者情绪指数的代理变量,其内在逻辑是,投资者情绪能较好地解释封闭式基金存在的四个谜团。如美国,封闭式基金主要由个体证券投资者持有并交易,故一般认为可用加权折价率代表个体证券投资者情绪。

我国至今对封闭式基金折价的解释仍没有较一致的观点,所以本文未采用封闭式基金加权折价率作为投资者情绪的代理变量。我国现有的另一种能代表投资者情绪的指数是中国证券分析师指数^①。本文也未采用该指数,原因有二:一是该指数从2001年7月开始编制至今,我国股市一直处于熊市,用它不能反映由牛到熊的过程;二是其构造方法上并不适合本文的研究(其值在-1000~+1000之间变化)。

基于上述原因,本文借鉴“投资者智慧”及“美国个人投资者协会”指数的编制方法,利用下式来计算投资者情绪指数,数据来源是“央视看盘”^②。

$$SI_t = \frac{BULLS_t}{BULLS_t + BEARS_t} \quad (10)$$

① 该指数由湘财研发中心和中央电视台“中国证券”节目合作,利用《央视看盘》统计的中国60家券商和咨询机构的分析师市场预测数据(含日预测和周预测数据),依公式“中国证券分析师指数=看涨比例×1000-看跌比例×1000”得出。

② www.cctv.com提供自2001年4月23日至今的日和周看涨、看跌与看平数据。

其中： SI_t 表示 t 时期投资者情绪指数； $BULLS_t$ 表示 t 时期看涨人数； $BEARS_t$ 表示 t 时期看跌人数。

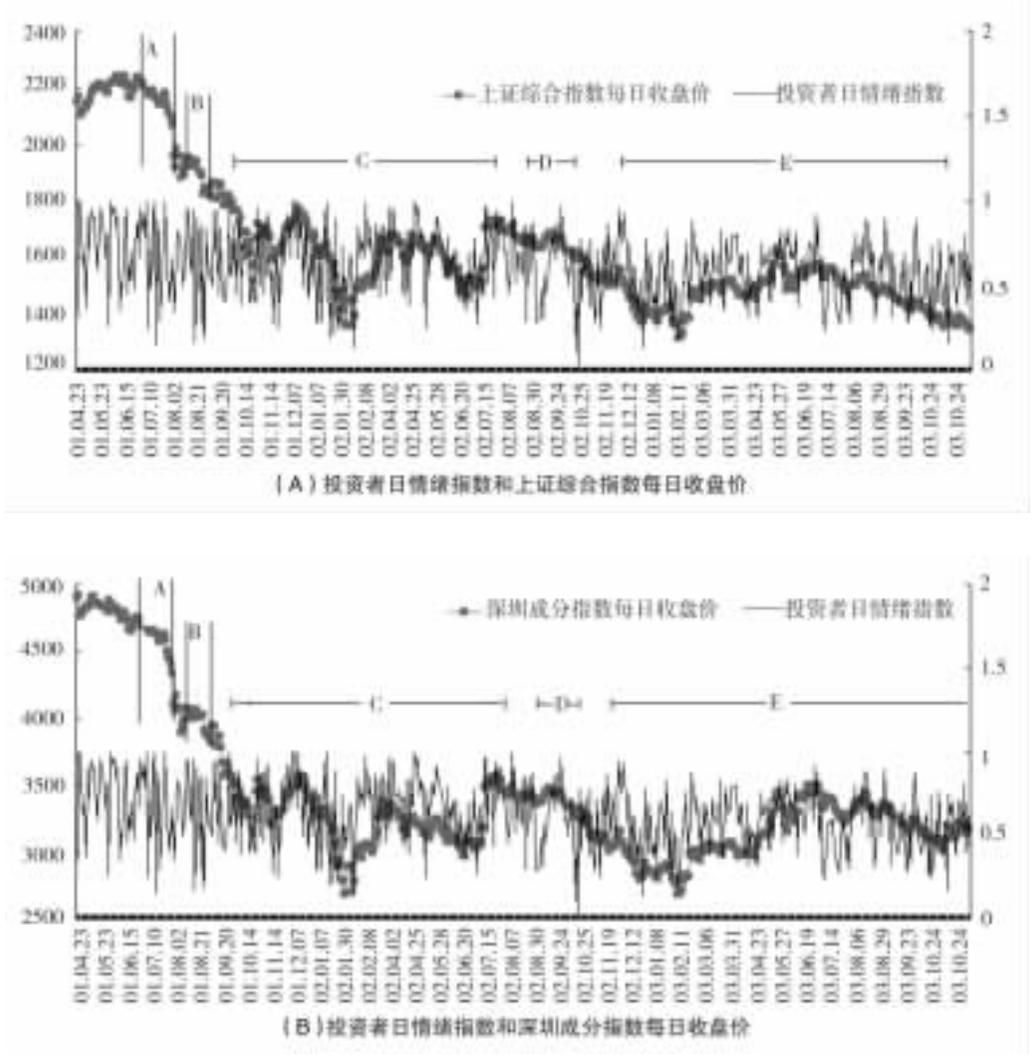


图 1

图 1 给出了投资者日情绪指数和上综指与深成指日收盘价。从图 1 可以看到，投资者情绪和上综指与深成指日收盘价在区间 A、B、C、D 和 E 的走势表现得较为一致，相比之下，深成指与投资者情绪走势拟合得更好。

令 $\Delta SI_t = SI_t - SI_{t-1}$ 表示 t 期投资者情绪的变化，则 $\Delta SI_t > 0$ 表示投资者在 t 期变得乐观或更乐观； $\Delta SI_t < 0$ 表示投资者在 t 期变得悲观或更悲观。

四、实证检验与结果

(一) 实证设计

由理论模型知，噪声交易者获得信号后，其情绪变化通过两个途径影响收益：一是引致对股票价格的错误估值；二是引致噪声交易者风险修正收益。这一作用过程的内在逻辑可以用图 2 表示。



图 2 情绪变化对收益与风险的影响

我们采用“从特殊到一般”的方法建立实证模型。高频金融时间序列分布通常具有波动率集群 (volatility clustering) 特性，GARCH 类模型一般能揭示这种波动特性。建模时首先考察上综指和深成指日收益率数据中是否平稳，是否存在序列依赖 (serial dependence) 和“日历效应” (calendar effect)。然后，在均值方程与方差方程中引入情绪指数变量，由诊断检验确定最终的模型。若数据支持 GARCH 类模型，且均值方程和方差方程中情绪变化的影响显著、方向正确 [结合 (8) 式和 (11) 式，应有： $\alpha_{1,1} \neq 0$ ； $\beta_{1,5} < 0$ 且 $\alpha_{1,2} > 0$]，则可以认为情绪变化是影响收益和收益波动的一个系统性因素，理论模型的结论从而得到实证支持。

(二) 实证结果^①

第一，样本期内沪深两市 ADF 检验的 τ 统计量分别为 -10.93747 和 -10.81436，均在 1% 的显著性水平上拒绝单位根，显示收益率平稳；第二，我们在均值方程中加入 AR (m)，并设置了反映“日历效应” (calendar effect) 的虚拟变量：根据样本容量沪深两市取 AR (25)，Box-Pierce Q 统计量分别为 25.536 和 20.320，表明两市收益率均为白噪声过程，不存在线性的序列依赖；虚拟变量显示两市均不存在“日历效应^②”；第三，上综指 ARCH 效应检验的 LM 统计量为 4.019714，深成指为 15.53782，在 5% 的水平上均显著，表明本文对类模型的选择是正确的。

本文最终确定了如下的 TGARCH - M (1, 1) 模型：

$$\begin{cases} R_{i,t} = \alpha_{1,0} + \alpha_{1,1} \sqrt{h_{i,t}} + \alpha_{1,2} \Delta SI_{i,t} + \varepsilon_{i,t} \\ h_{i,t} = \beta_{1,0} + \beta_{1,1} \varepsilon_{i,t-1}^2 + \beta_{1,3} \varepsilon_{i,t-1}^2 D_{1,t-1} + \beta_{1,4} h_{i,t-1} + \beta_{1,5} \Delta SI_{i,t} \\ \varepsilon_{i,t} | I_{i,t-1} \sim N(0, h_{i,t}) \end{cases} \quad (11)$$

① 本文的计量分析在 Eviews 4.1 上完成。

② “日历效应”在我国股市并无一致的结论 (薛继锐和顾岚, 2000; 奉立城, 2000; 汪伟和周宇, 2002)。

其中, $i=0$ 代表上综指; $i=1$ 代表深成指。 $\alpha_{i,1}$ 表示风险和收益之间的关系, 若 $\alpha_{i,1} > 0$ 并显著, 表示风险奖励; 若 $\alpha_{i,1} < 0$ 并显著, 表示风险惩罚。 $D_{i,t-1}$ 是表示“杠杆效应”的虚拟变量, 当 $\varepsilon_{i,t-1} < 0$ 时, $D_{i,t-1} = 0$; 当 $\varepsilon_{i,t-1} \geq 0$ 时, $D_{i,t-1} = 1$; 若 $\beta_{i,3} > 0$ 且显著, 则利空消息对收益波动的冲击要大于同等程度利好消息; 若 $\beta_{i,3} < 0$ 且显著, 则反之。 $\alpha_{i,2}$, $\beta_{i,5}$ 分别表示投资者情绪变化对收益和收益波动的影响。 $\beta_{i,5} > 0$ 并显著, 表示投资者情绪变化同向修正收益波动; $\beta_{i,5} < 0$, 则表示反向修正收益波动。表 1 列出了沪深两市模型 (11) 的最大似然估计结果。

表 1 沪深两市收益、收益波动与投资者情绪的实证检验结果

	上证	综合	指数		深圳	成分	指数	
	系数	标准差	z 统计量	p 值	系数	标准差	z 统计量	p 值
$\alpha_{i,0}$	-0.004619	0.001350	-3.421695	0.0006	-0.004078	0.001560	-2.613634	0.0090
$\sqrt{h_{i,t}}$	0.333479	0.125332	2.660766	0.0078	0.273076	0.128887	2.118734	0.0341
ΔSI_t	0.004960	0.001867	2.657123	0.0079	0.005501	0.002002	2.747396	0.0060
$\beta_{i,0}$	1.18E-05	2.80E-06	4.229733	0.0000	9.68E-06	3.21E-06	3.015207	0.0026
$\varepsilon_{i,t-1}^2$	0.124708	0.017060	7.309855	0.0000	0.141811	0.017524	8.092338	0.0000
$\varepsilon_{i,t-1}^2 D_{i,t-1}$	0.201394	0.052993	3.800416	0.0001	0.116239	0.046485	2.500559	0.0124
$h_{i,t-1}$	0.734102	0.028411	25.83865	0.0000	0.779265	0.026694	29.19250	0.0000
$\Delta SI_{t-2(1)}$	-6.70E-05	3.12E-05	-2.147451	0.0318	-6.63E-05	3.12E-05	-2.125740	0.0335
ML	1791.851				1741.069			
诊断检验统计量								
d-w	2.043898				2.015448			
BPQ ¹ (5)	3.8857			0.566	3.7638			0.439
BPQ ¹ (10)	6.3066			0.789	7.3504			0.692
BPQ ¹ (25)	23.4			0.554	21.677			0.654
BPQ ² (5)	2.7355			0.741	2.3407			0.800
BPQ ² (10)	7.9671			0.632	7.9351			0.635
BPQ ² (25)	24.573			0.487	25.015			0.462

注: ① BPQ¹(M), BPQ²(M) 分别代表估计的模型标准残差一次和二次项前 M 阶的 Box-pierce Q 统计量; ② 实证模型 (9) 标准化残差 Q-Q 图近似于一条直线, 显示近似正态。

从表 1 看出: 均值方程中, 投资者情绪显著地影响收益。当投资者乐观 (或更乐观) 时, 收益显著增加; 当投资者悲观 (或更悲观) 时, 收益显著减少, 这也印证了一个说法——市场乐观时都能赚钱, 市场悲观时都会赔钱。方差方程中, 沪深两市均表现出投资者情绪变化对收益波动有显著的反向修正作用, 但沪市中这一影响作用滞后两期。它们表明

投资者情绪的急剧变化能造成更剧烈的收益波动。

从估计结果还看出，沪深两市均存在显著的风险奖励。这表明两个市场的正向风险溢酬，高风险意味着高收益，高收益伴随着高风险；而且，投资者情绪的变化所造成的收益波动将通过这一机制进一步影响收益， $\beta_{i,5} < 0$ 和 $\alpha_{i,1} > 0$ 显示其作用是反向修正，印证了模型（8）的结论。模型（11）的估计结果还表明沪深两市的“杠杆效应”均很明显，这与成熟的资本市场反应方式相同（例如，Engle 和 Ng, 1993），也与洪永淼、成思危等（2003）对我国股市的研究结果相同，但与刘金全和崔畅（2002）的研究结果不一样。我们认为出现这一结果的原因在于样本期的差异，刘文样本期内^①，我国股票市场具有相当程度的政府主导或者政府影响特征，投资者对市场利空消息反应谨慎，对政府进行干预或者“政策救市”的预期十分稳固，导致利空消息产生较小的波动。

综上所述，沪深两市投资者情绪变化不仅能显著地影响收益，而且显著反向修正收益波动，并通过风险奖励影响收益。这一结论表明了沪深两市具有较为相同的投资者行为和风险收益特征，投资者情绪是一个影响收益的系统性因子。

（三）实证结果的引申含义

（1）机构投资者是不是噪声交易者风险源？本文构造的指数反映了机构投资者情绪，它对沪深两市均有显著影响作用，这一结果说明什么问题？心理学家的研究表明，很多心理偏误实际上谁都会犯，它们不仅仅出现在普通人身上，养老基金（Pension Fund）、共同基金（Mutual Fund）的经理人甚至诺贝尔奖获得者也会产生同样的偏误。Brown 和 Cliff（1999）曾发现机构投资者也会受到情绪左右，从而否定了情绪主要存在于个体证券投资者仅影响小公司股票收益的传统看法。Shleifer 与 Vishny（1997）指出：机构投资者也是普通人，影响个人投资者的偏见同样会影响他们。他们受委托代人理财，需要现实业绩，……。在某些时候，他们其实也是标准的噪声交易者。也就是说，由于种种原因，机构投资者也可能患各种认知偏差错误。我国培育和发展机构投资者的制度尚不完善，机构投资者的非理性行为还较为严重。何诚颖（2001）将中国的机构投资者分为“知情下注交易者、噪声交易者和其他类型投资者”三种类型，指出知情下注交易者也同时具有噪声交易者特征。施东晖（2001）发现，相对于个体投资者而言，理性程度较高、有较大影响力，以及在信息收集、处理与传递等方面具有优势的基金业中“羊群行为”普遍存在。王岗（2003）以基金为例说明了中国机构投资者存在“正反馈交易者”行为。本文的实证结果综合反映了机构投资者信念、情绪和反应的影响作用，表明他们是可能的噪声交易者风险源。另一方面，中国个体投资者整体素质并不高，有跟风以及做从羊的倾向，其情绪容易受机构投资者影响，两类投资者情绪的迭加导致噪声在市场上的广泛传播与放大。

（2）市场是有效的吗？本文的理论模型和实证结果均表明，噪声交易者的情绪能影响均衡资产价格，这一结论对有效市场假设有什么意义？Fama（1970）用鞅模型来刻画有效市场，即将有效市场定义为股票价格遵从鞅过程。市场弱式有效可正式表述为：

$$H_0 : E(r_t | I_{t-1}) = E(r_t) \quad a. s. \quad (12)$$

① 刘文的样本期自 1991 年 4 月 3 日至 2001 年 6 月 29 日。

其中： r_t 表示第期股票收益率，为 $t-1$ 期到 t 期的对数股价相对变化； I_{t-1} 表示在 $t-1$ 期已知的股票收益率集合。若原假设 H_0 成立，则市场达到弱式有效。

在本研究中，沪深两市收益率均为白噪声过程，但这并不能说明市场达到弱式有效，因为白噪声序列不一定是鞅差分序列，检验白噪声过程的统计方法只能检验收益率条件期望的线性依赖部分，而无法检验条件期望中的非线性依赖部分（陈灯塔和洪永森，2003）。表 1 的实证结果表明，样本期内沪深两市收益率均存在 TGARCH-M 效应，条件期望 $E(r_t | I_{t-1})$ 依赖于条件方差 h_t ，不符合式 (12) 的鞅过程定义。具体而言，不仅当期投资者情绪变化直接影响收益，而且情绪变化通过条件方差影响收益，股市期望收益随情绪变化表现出时变性、具有非线性依赖，这明显背离了市场弱式有效的性质。以上分析进一步说明了：尽管造成目前市场未能达到弱式有效的原因很多，但投资者情绪显然是一个非常重要的原因。

五、结语

在我国个体投资者素质不高，跟风、盲目信赖专家等现象普遍存在的情况下，投资者情绪变化不仅显著地影响沪深两市的均衡收益，而且显著地反向修正沪深两市的收益波动，通过风险奖励影响收益，投资者情绪是一个影响收益的系统性因子。这一结论表明了沪深两市不仅具有相同的投资者行为特征和风险收益特征，而且均未达到弱式有效，机构投资者是可能的噪声交易者风险源。上述结论为政府监管者、券商、中介机构、上市公司和投资者判断我国股票市场上的投资者行为和股市基本特征提供了一定的参照标准。把发展机构投资者作为我国证券市场重要发展战略的同时，如何提高机构投资者的素质、增强其投资行为的理性化程度也是一个同等重要的问题。这一问题解决得好，才能提高个体投资者素质，进而提高市场效率。

参考文献

- [1] 安德瑞·史莱佛 (Andrei Shleifer):《并非有效的市场——行为金融学导论》(“Inefficient Markets: An Introduction to Behavioral Finance”, 2000, Oxford), 中译本, 中国人民大学出版社 2003 年版。
- [2] 陈灯塔, 洪永森:《中国股市是弱式有效的吗——基于一种新方法的实证研究》,《经济学(季刊)》, 2003 年第 3 卷第 1 期。
- [3] 奉立城:《中国股票市场的“周内效应”》,《经济研究》, 2000 年第 11 期。
- [4] 金晓斌, 高道德, 石建民, 刘红忠:《中国封闭式基金折价问题实证研究》,《中国社会科学》, 2002 年第 5 期。
- [5] 何诚颖:《中国股市“板块现象”分析》,《经济研究》, 2001 年第 12 期。
- [6] 李心丹, 王冀宁, 傅浩:《中国个体证券投资者交易行为的实证研究》,《经济研究》, 2002 年第 11 期。
- [7] 李心丹, 傅浩:《中国证券投资者行为研究》, 上证联合研究计划第三期课题报告, 2002。
- [8] 刘焯辉, 熊鹏:《资产流动性、投资者情绪与中国封闭式基金之谜》, 第三届中国青年经济学者论坛论文集, 中山大学出版社 2003 年版, 第 439~453 页。
- [9] 刘焯辉, 贺菊煌, 沈可挺:《中国股市中信息反应模式的实证分析》,《管理世界》, 2003 年第 8

期。

- [10] 刘金全, 崔畅:《中国沪深股市收益率和波动性的实证分析》,《经济学(季刊)》,2002年第1卷第4期。
- [11] 饶育蕾, 刘达锋:《行为金融学》,上海财经大学出版社2003年版,第172~185页。
- [12] 施东晖:《证券投资基金的交易行为及其市场影响》,《世界经济》,2001年第10期。
- [13] 施东晖, 孙培源:《基于CAPM的中国股市羊群行为研究——兼与宋军、吴冲锋先生商榷》,《经济研究》,2002年第2期。
- [14] 宋军, 赵焯, 吴冲锋:《资本市场中的头羊——从羊模型》,《系统工程理论与实践》,2003年第1期。
- [15] 王岗:《机构投资者与股票市场稳定》,《证券市场导报》,2003年8月号。
- [16] 汪炜, 周宇:《中国股市“规模效应”和“时间效应”的实证分析》,《经济研究》,2002年第10期。
- [17] 薛继锐, 顾岚:《中国股票市场的日历效应分析》,《数理统计与管理》,2002年第2期。
- [18] 张俊喜, 张华:《解析中国封闭式基金折价之谜》,《金融研究》,2002年第12期。
- [19] 洪永淼, 成思危等: Extreme Risk Spillover Between Chinese Stock Markets and International Stock Markets, Working Paper, Cornell University & Tsinghua University, 2003.
- [20] Brauer, G. A., Investor Sentiment and The Closed-end Fund Puzzle: A 7 Percent solution, Journal of Financial Services Research, 1993, Vol.7, No.3, pp199~216.
- [21] Brown, G. W., Volatility, Sentiment, and Noise Traders, Financial Analysts Journal, 1999, Vol.55, No.2, pp82~90.
- [22] Brown, G. W., Cliff, M. T., Sentiment and The Stock Market, Working Paper, University of North Carolina, Chapel Hill, 1999.
- [23] Brown, G. W., Cliff, M. T., Investor Sentiment and the Near-term Stock Market, Journal of Empirical Finance, 2004, Vol.11, pp1~27.
- [24] De Long, J. B., Shleifer, A., Summers. L. G., Waldmann, R. J., Noise Trader Risk in Financial Markets, Journal of Political Economy, 1990, Vol.98, No.4, pp703~738.
- [25] Engle. R. F., Ng. V. K., Measuring and Testing The Impact of News on Volatility, Journal of Fiance, 1993, Vol.48, pp1749~1778.
- [26] Fama, E. F., The Behavior of Stock Market Prices, Journal of Business, 1965, Vol.38, No.1, pp34~105.
- [27] Fama, E. F., A Review of Theory and Empirical Work, Journal of Finance, 1970, Vol.25, pp383~417.
- [28] French, K. R., Roll, R., Stock Return Variances: The Arrival of Information and The Reaction of Taders, Journal of Financial Economics, 1986, Vol.17, No.1, pp5~26.
- [29] Friedman. M., The Case For Flexible Exchange Rates, Essays in Positive Economics, University of Chicago Press, Chicago, 1953.
- [30] Grossman. S. J., Stiglitz. J. E., On the Impossibility of Information ally Efficient Markets, The American Economic Review, 1980, Vol.70, No.3, pp393~408.
- [31] Kelly, M., Do Noise Traders Influence Stock Prices? Journal of Money, Credit, and Banking, 1997, Vol.29, No.3, pp351~363.
- [32] Lee, C. M. C., Shleifer, A., Thaler, R. D., Investor Sentiment and The Closed-end Fund Puzzle, Journal of Finance, 1991, Vol.48, No.1, pp75~106.
- [33] Lee. W.J., Jiang. C. X., Indro. D. C., Stock Market Volatility, Excess Returns, and The Role of

Investor Sentiment , *Journal of Banking & Finance* , 2002 , Vol.26 , pp2277 ~ 2299 .

- [34] Simon , D. P. , Wiggins , R. A. , *Stock Returns and Sentiment Indicators* , Working Paper , Bentley College , 1999 .
- [35] Swaminathan , B. , *Time-varying Expected Small Firm Returns and Closed-end Fund Discounts* , *Review of Financial Studies* , 1996 , Vol.9 , pp845 ~ 887 .
- [36] Shleifer , A. , Vishny , R. , *The Limit of Arbitrage* , *Journal of Finance* , 1997 , Vol.52 , pp35 ~ 55 .
- [38] Vandana Singhvi , *Investor Sentiment : Its Measurement and Dimensions* . Dissertation Submitted to the Degree of Doctor of Philosophy , New York University , 2001 .

商业银行监管的国际比较：模式及影响^{*}

——兼论中国的商业银行监管模式选择

李 涛

一、导言

现代经济中，与其他行业相比，银行业是政府监管最为严格的行业之一。文献中一般把这种严格的监管归结为以下原因：防止银行的流动性危机，维持金融业的稳定，保护分散的小储户和投资者，提高社会效率以及其他的社会目的（Herring 和 Santomero, 2000；Freixas 和 Rochet, 2001）。尽管商业银行受到如此严格的监管，但 20 世纪 80 年代以来，频繁发生的银行危机仍然成为经济不稳定的一个主要动因。因此，不恰当的银行监管成为人们批评的焦点（Caprio, 1998）。现有的文献对于恰当和不恰当的银行监管手段有着大量的研究。对此，Barth 等人（2002）在他们的文章里提供了一个很好的文献综述。同时，他们也分析了一些具体的银行监管措施对银行发展水平、银行经营效率和银行危机爆发的可能性的影响。

总结已有文献，尽管它们都详细地分析了某些具体的监管措施的效果，但是都没有系统地归纳研究世界上主要的银行监管模式，以及这几种模式对一国的金融发展水平（包括股市和银行）的影响。为弥补以上不足，我们在本文主要回答两个问题：第一，当今世界有几种主要的银行监管模式？第二，不同的监管模式对金融发展水平有什么样的影响？正如巴塞尔委员会发布的一系列关于银行监管的文件所显示，在当前银行危机频繁发生的情况下，回答以上问题对完善银行监管显得尤为迫切。

基于世界银行提供的 118 个国家商业银行监管措施的数据，本文首先提出了一个全面的银行监管措施分类，根据这个分类，本文分析了世界上主要的银行监管模式及其对金融发展水平的影响。

分析显示，世界各国监管商业银行的措施主要集中在以下四种维度：①政府拥有银行的程度；②政府直接监管银行的力度；③政府授权非政府机构进行监管的力度；④明确的存款保险制度涵盖范围。基于上述四种维度的不同组合，我们把世界各国的商业银行监管模式划分为三大类：①印度 - 中国型，其特点是极度依赖政府直接拥有银行；②南非 - 菲律宾 - 墨西哥型，其特点是强调政府对银行的直接监管，而忽视市场力量的作用；③德国

^{*} 本文来自于我的博士论文第一章。由衷感谢我的导师香港科技大学李稻葵教授在本文的写作过程中给予的指导与帮助，也感谢香港大学白重恩教授，香港科技大学陈松年教授、雷鼎鸣教授、朱天教授、北京大学和香港科技大学武常岐教授以及中国人民大学李红博士的宝贵意见。原载：《经济研究》，2003 年第 12 期。

- 美国 - 瑞士 - 法国型，其特点是主要依赖非政府力量来监管银行，而较少使用政府直接监管措施。回归分析结果表明，灵活的监管机构、非政府机构的监督，以及涵盖面广泛的存款保险制度都对金融业的发展有正面影响，而政府直接拥有银行，以及事先或事后的政府直接监管都不利于金融业的发展。比较不同的监管模式，在促进金融业发展方面，德国 - 美国 - 瑞士 - 法国型的监管模式最佳，南非 - 菲律宾 - 墨西哥型次之，而印度 - 中国型最差。

进一步比较中国与世界平均水平以及德国和美国的商业银行监管模式，我们发现：尽管中国的银行会计制度最为严格，这有助于中国金融业的发展；但中国的商业银行监管主要依赖政府拥有银行，表现为中国的国有银行比例世界最高，监管机构的独立性很低，同时又设置了世界上最为严格的事先监管要求，并且缺乏明确的存款保险制度，这些都不利于中国金融业的发展。

本文的结构如下：第二节介绍了我们使用的数据，以及对银行监管措施的一个四维划分；第三节对数据进行了一些简单的相关分析；第四节讨论了世界上三种主要的银行监管模式；第五节分析了不同维度进而不同模式的监管措施对金融发展水平的影响；第六节比较了中国与世界平均水平以及德国和美国的商业银行监管模式；在最后一节，我们总结了文章的发现并探讨了有关的政策含义。

二、数据和四维监管措施

（一）银行监管数据和四维监管措施

本文所使用的商业银行监管数据来自于世界银行资助的一个银行监管研究项目。该项研究采用问卷调查形式，搜集了世界上 118 个国家有关 12 个方面的银行监管数据。问卷搜集主要在 1999 年（65 份问卷在 1999 年搜集，其余的 53 份在 1998 或 2000 年搜集）。数据属于横截面格式。Barth 等人（2001）详细地介绍了数据的搜集过程和问卷的调查内容，并且进行了简单的统计分析。^①

参照巴塞尔委员会分别于 1997 和 2001 年发布的“有效银行监管的核心原则”（Core Principles for Effective Banking Supervision）和“新巴塞尔资本协定”（New Basel Capital Accord），我们把原始问卷涉及的 12 个方面的问题重新归纳，构造了四个宽维度的银行监管措施。第一维度是政府拥有银行的程度，“有效银行监管的核心原则”把它列为一种特殊的监管安排。La Porta 等人（2002）深入研究了国有银行的分布、成因，以及对金融和经济发展的影响。第二维度是政府直接监管银行的力度。根据“有效银行监管的核心原则”，我们可以进一步把它细分为四个子维度：①直接监管的先决条件，即监管机构的独立性和灵活性；②事先监管，即政府在银行开始营业前施加的监管措施；③日常监管，即政府对银行日常经营的监管措施；④事后监管，即政府在银行经营出现问题时的监管。第三维度是政府授权非政府机构进行监管的力度，即市场力量参与银行监管的程度。“新巴

^① 数据可以免费从世界银行相关网址下载：<http://www.worldbank.org/research/projects/bank-regulation.htm>。

塞尔资本协定”的第三个支柱就是市场化监管。Flannery (2001) 给出了市场化监管的精确定义。第四维度是明确的存款保险制度涵盖范围。Garcia (1999) 对此有个很好的理论综述。存款保险制度比较特殊，因为它有一些在银行设立时就需要满足的要求，而它往往在银行的日常经营或事后出现问题时发挥作用。因此，与第一维度的政府拥有银行程度相似，“有效银行监管的核心原则”也把明确的存款保险制度列为一种特殊的监管安排。

以上四维划分尽管涵盖面广，但是可能忽略了一些特别重要或是理论上更为关注的特定监管措施。因此，参考 Barth 等人 (2002) 的研究，我们还定义了三个窄维度的监管措施，包括直接监管的先决条件、事先监管和政府授权非政府机构进行监管的力度。^① 在此我们简单描述有关变量如下：

(1) 政府拥有银行的程度：我们引入两个互相替代的变量。

①GB95：1995 年给定国家中前 10 大银行的国有资产比例。^②

②GOVBANK：给定国家中所有大股东是政府（50% 以上的资产为政府所有）的银行资产的比例。

(2) 政府直接监管银行的力度：如上所述，我们把它细分为四个子维度。

①PRECOND：宽维度的直接监管的先决条件，包括监管机构的独立性和灵活性。

PRECONDN：窄维度的直接监管的先决条件，仅包括监管机构的灵活性。

②EXANTE：宽维度的事先监管，包括以下五方面的措施：

a) BUSISCOPE：对银行从业范围的限制；

b) ENTRY：银行申请营业执照所需提交的相关文件；

c) OWNERSHIP：银行所有制结构的限制；

d) INITIALK：初始的资本充足率要求；

e) INVESTMENT：银行能否拥有非金融公司。

EXANTIEN：窄维度的事先监管，仅包括 BUSISCOPE 和 ENTRY。

③ONGOING：日常监管，即政府对银行日常经营的监管，包括以下三方面的措施：

a) OVERALLK：总的资本充足率要求；

b) RISKMANA：风险管理要求；

c) CORRECTIVE：监管机构的及时纠错能力。

④EXPOST：事后监管，即政府在银行经营出现问题时的监管，包括以下三方面的措施：

a) SANCTION：对银行管理人员的惩罚措施；

b) DECINSOL：监管机构中止银行经营和宣布银行破产的能力；

c) RESTRUC：监管机构重组银行的能力。

(3) 政府授权非政府机构监管银行的力度：宽维度（EMPOWERMENT）的定义包括以下五方面的措施：

①ACCOUNTING：银行的会计准则；

① 由于篇幅所限，具体的变量定义不在附录中列出，感兴趣的读者可以直接向作者索取。

② 数据来自于 La Porta 等人 (2002)。

②SUBDEBT：是否允许银行资本金包括从属债务；

③AUDITING：对外部审计的要求；

④ ACCURACY：信息公布的准确性；

⑤DISCLOSURE：对公众的信息披露要求。

EMPOWERMENTN：窄维度的政府授权非政府机构进行监管的力度，仅包括 ACCOUNTING。

(4) 明确的存款保险制度涵盖范围：简写为 DIP。

有关变量赋值，我们的处理方法如下：由于问卷中的绝大多数问题的答案是属于“是/否”，如果该国存在问卷特定问题所涉及的监管措施，我们给相应的变量赋“1”，反之赋“0”。对于剩下的五个答案不属于“是/否”型的问题，如果该国对问题中涉及的监管措施有着最为严格的规定，我们取“4”或“3”，反之，我们取“1”来指代最为松散的规定，中间值“3”或“2”表示温和的规定。^①

与 Barth 等人 (2002) 一致，我们在构造以上有关银行监管变量时，采用了两种方法。首先，一个非常直观的方法是把有关变量包括的所有单个问题的答案赋值直接加总，但是这种问题与问题之间的相等权重假设可能受到置疑。于是我们引入了第二种变量构造法，即第一主因子法 (First Principal Component Method)。这种方法的优点在于可以自动寻找变量之间的最优权重。在以下的分析中，我们采用的是第一主因子法构造的银行监管变量。为了检验研究结果的稳定性，我们也使用了简单加总法构造的银行监管变量进行分析，发现主要结论完全一致。^②

(二) 金融发展水平数据

在评价不同的监管维度以及不同的监管模式的影响时，我们引入了金融发展水平，而不是比较直观的银行发展水平作为考量标准。这是因为，Levine (2002) 发现只有金融发展水平和经济增长稳定地正相关，而在单独的银行或资本市场的发展与经济增长之间不存在稳定的关系。

与 Levine (2002) 相同，我们引入三个变量来衡量金融发展水平^③。首先，FINAACTI，即金融业活动，用来衡量股票市场和银行业的总体活动能力。其次，FINASIZE，即金融业规模，用来衡量股票市场和银行业的总体规模。再次，FINAEFFI，即金融业效率，用来衡量股票市场和银行业的总体效率。

三、相关分析^④

总结我们的发现如下：第一，两种不同的国有银行比例 GB95 和 GOVBANK 之间存在

① 由于篇幅所限，具体的变量赋值方法不在附表中列出，感兴趣的读者可以直接向作者索取。

② 由于篇幅所限，具体结果不在附表中列出，感兴趣的读者可以直接向作者索取。

③ 由于篇幅所限，具体的变量定义不在附表中列出，感兴趣的读者可以直接向作者索取。

④ 由于篇幅所限，具体的相关分析结果不在附表中列出，感兴趣的读者可以直接向作者索取。

着很强的显著的正相关性，从而支持了我们在测量政府拥有银行的程度时可以交替使用以上两个变量。第二，国有银行比例（GB95）和宽维度的政府直接的事先监管显著正相关，说明国有银行受到政府事先的严格监管，例如设置严格的市场准入限制，以及规定非常单一的银行从业范围，这同 Barth 等人（2002）的发现相吻合。第三，国有银行比例（GOVBANK）与政府授权非政府机构监管银行的力度显著负相关，说明政府拥有银行妨碍了市场力量对银行的监督，这与 Barth 等人（2002）的发现一致。第四，政府直接的事先监管与政府授权非政府机构监管银行的力度显著负相关，而与政府直接的事后监管显著正相关。我们还发现政府授权非政府机构监管银行的力度与政府直接的事后监管显著负相关。因此，政府倾向于同时进行较强或较弱的直接的事先和事后监管，而这些直接的监管措施限制了市场力量在监管银行中的作用。第五，政府直接监管的先决条件以及政府直接的日常监管都与其他任何监管维度之间不存在明显的相关性。第六，在全部三组有宽维度和窄维度不同定义的变量之间都存在着显著的正相关性，从而支持我们对它们的交替使用。最后，与 Barth 等人（2002）的发现一致，明确的存款保险制度涵盖范围同其他任何监管维度之间都不存在明显的相关性，这可能说明有些国家的存款保险制度的设计不很科学，缺乏相应的市场力量进行监管，从而无法克服存款保险制度带来的银行经营动机的可能扭曲（Demirgüç-Kunt 和 Kane, 2002）。

对于三组不同的测度金融发展水平的变量，我们发现三者之间存在着显著的正相关性，这也同 Levine（2002）的发现一致。

四、三种银行监管模式

在这一节，基于以上的四种银行监管维度的不同组合，我们讨论了世界上三种主要的商业银行监管模式。为了方便今后更加清晰地讨论不同模式对金融发展水平的影响，在面对宽或是窄维度的变量选择时，我们使用了后者。附表 1 给出了相关发现的描述。

首先，许多国家在解决如何监管银行的问题时，主要依赖政府直接拥有银行，典型的国家有印度和中国。以 50% 作为分界点，在我们的样本中有 33 个国家的国有银行比例（用 GB95 或 GOVBANK 衡量）高于 50%，占到整个样本的 28%。我们把这种主要依赖政府直接拥有银行的监管模式称为印度 - 中国型。比较各国监管机构的灵活性与整个样本的中位数，我们可以进一步把印度 - 中国型细分为印度型和中国型，前者的监管机构灵活性更高，而后者正好相反。在我们的样本中有 6 个国家属于印度型，26 个国家属于中国型，分别占到总样本的 5% 和 22%。对于越南，我们缺乏相应的监管机构灵活性的数据。

其次，在那些没有主导性的国有银行的国家中，一部分国家强调政府对银行的直接监管，而忽视市场力量的作用。具体地说，这些国家倾向于依赖政府强有力的直接的事先或事后监管（相应的变量值大于总样本的中位数），而较少发动非政府的力量来监管银行（相应的变量值小于或等于总样本的中位数），典型的国家有南非、菲律宾以及墨西哥。在我们的样本中有 43 个国家呈现出这种监管模式，占到总样本的 36%。我们把这种监管模式称为南非 - 菲律宾 - 墨西哥型。类似地，比较各国监管机构的灵活性与整个样本的中位数，我们进一步把南非 - 菲律宾 - 墨西哥型细分为南非型和菲律宾 - 墨西哥型，前者的监

管机构灵活性更高，而后者正好相反。样本中有 5 个国家属于南非型，37 个国家属于菲律宾 - 墨西哥型，分别占到总样本的 4% 和 31%。对于冈比亚，我们缺乏相应的监管机构灵活性的数据。

再次，在没有主导性的国有银行的国家中，还有一些国家的监管模式与南非 - 菲律宾 - 墨西哥型截然相反，它们主要依赖非政府力量来监管银行，而较少使用政府直接监管。具体而言，这些国家倾向于依靠非政府机构即市场力量来监管银行（相应的变量值大于总样本的中位数），而较少依赖政府直接的事先和事后监管（相应的变量值小于或等于总样本的中位数），典型的国家有德国、美国、瑞士和法国。我们的样本中共有 24 个国家呈现出这种监管模式，占到总样本的 21%，我们称之为德国 - 美国 - 瑞士 - 法国型。同样，比较各国监管机构的灵活性与整个样本的中位数，我们进一步把德国 - 美国 - 瑞士 - 法国型分为德国 - 美国型和瑞士型，前者的监管机构灵活性更高，而后者正好相反。样本中有 9 个国家属于德国 - 美国型，14 个国家属于瑞士型，分别占到总样本的 8% 和 12%。对于法国，我们缺乏相应的监管机构灵活性的数据。

此外，在意大利、列支敦士登和比利时三国，国有银行比例较低，然而政府的直接监管（事先或者事后）和非政府机构的监管都很强（相应的变量值均大于总样本的中位数）。

最后，还有 15 个国家或者数据缺失，或者监管模式难以归类，它们占总样本的 13%。

总结以上可以发现，尽管世界各国在监管其商业银行时会采取不同的监管维度的组合，其呈现出的监管模式却可以分为三大类：①印度 - 中国型，其特点是极度依赖政府直接拥有银行；②南非 - 菲律宾 - 墨西哥型，其特点是强调政府对银行的直接监管，而忽视市场力量的作用；③德国 - 美国 - 瑞士 - 法国型，其特点是主要依赖非政府力量来监管银行，而较少使用政府直接监管措施。以上三种模式概括了当今世界主要的银行监管模式，其样本覆盖率达到了 85%。

五、不同监管模式的影响

这一节集中讨论不同的银行监管维度以及模式对金融发展水平的影响。与 Levine (2002) 相同，我们分别使用了三组指标来测量金融发展水平，即金融业活动、金融业规模以及金融业效率。如前所述，宽维度的银行监管变量可能会混淆一些特别重要的银行监管措施的效果 (Barth 等, 2002)，为了避免这种可能性的产生，并且比较不同宽窄维度变量的效果，在回归分析中我们分别考虑了宽维度和窄维度的银行监管变量。

需要指出的是，作为自变量的银行监管维度的收集时间是 1998 ~ 2000 年，而作为因变量的金融发展水平取自 1995 ~ 1997 年^①，这样自变量的时间先于因变量的时间，可能导致银行监管措施的内生性，进而影响回归分析的准确性。我们认为，有两个原因可以打消以上疑虑。首先，正如原始数据的收集者 Barth 等人 (2002) 所言，绝大多数银行监管措施在近 20 年间未曾发生变化，因此相对金融发展水平而言，银行监管措施非常稳定；其次，银行监管措施对金融业发展产生影响的时间，相对于反向关系，即金融业发展对银行

① 我们无法收集到相应的 2000 年之后的金融发展水平数据。

监管措施产生影响的时间要短得多。因此，集中在一个较短的时间范围内，因果关系只能是从银行监管措施到金融业发展水平，从而支持了我们的分析框架。

相应的回归方程式如下：

$$Y_i = \alpha + \sum_{j=1}^k \beta_j X_{ij} + \epsilon_i$$

其中： i 表示特定国家； j 表示不同的监管维度。

附表 2 和 3 分别给出了包含宽维度和窄维度的银行监管变量的多元回归结果。

比较附表 2 和 3，我们有如下发现：政府直接拥有银行对以金融业规模和金融业效率衡量的金融发展水平有着显著的负面影响，这与 La Porta 等人（2002）关于国有银行的发现一致。政府直接的事后监管对以金融业规模衡量的金融发展水平有着显著的负面影响，而明确的存款保险制度涵盖范围有助于提高金融业效率。

集中于窄维度的银行监管措施，我们进一步发现：监管机构的灵活性有助于扩大金融业规模；政府直接的事先监管对以金融业活动和金融业效率衡量的金融发展水平有着显著的负面影响^①；政府在监管银行方面对非政府机构的授权有助于提高金融业效率。

此外，我们发现政府直接的日常监管对金融发展水平没有显著的影响。

总结以上关于不同银行监管维度对金融发展水平影响的结论，对应我们归纳的世界上三种主要的银行监管模式，我们发现：在促进金融业发展方面，①由于在德国 - 美国 - 瑞士 - 法国型的监管模式下，国有银行比例较低，政府直接监管力度较弱，而主要依靠市场力量监管银行，因而表现最佳；②在南非 - 菲律宾 - 墨西哥型的监管模式下，虽然国有银行比例也较低，但是政府直接监管的力度较强，而对市场力量的依赖较弱，因此表现次之；③在印度 - 中国型的监管模式下，占主导地位的国有银行导致其表现最差。

最后，有关窄维度的银行监管措施的分析还表明，特定的监管措施有时在影响金融发展水平方面要比综合的监管手段更为重要。

六、中国商业银行监管模式的一个简单的国际比较

在这一节通过比较中国与世界平均水平以及德国和美国的商业银行监管模式^②，我们更加详细地描述了中国商业银行监管模式的特点，从而为中国商业银行监管改革提供了一些可操作意见。

附表 4 总结了有关的比较结果。发现如下：

第一，中国的国有银行比例很高。中国的政府拥有银行比例（99.45%）是样本中最高的，远高于世界平均水平（29.1%）^③，也远高于德国（36.36%）和美国（0）。

① 政府直接的事先或事后监管对金融发展水平的显著的负面影响，与自由银行制度（free banking）促进了金融以及经济发展的历史证据相吻合。有关自由银行制度的描述，可以参考 Dowd（1992）。

② 根据上一节的讨论，德国和美国的商业银行监管模式在推动金融发展方面表现最佳。

③ 我们采用样本中位数来表示平均水平。由于我们比较的是各种变量的排列顺序，从统计学角度分析，样本中位数比平均值更为准确。

第二，中国的银行监管机构的独立性不强^①。尽管中国的银行监管机构的灵活性(-0.4127)^②与世界平均水平相同，但是其宽维度的先决条件(-0.3381)要远低于世界平均水平(0.3243)。而德国和美国在上述两方面都高于中国(1.3652, 1.4570)。

第三，中国对银行的事先监管也很严格。中国的宽维度的事先监管(3.5881)是样本中最严格的，远高于世界平均水平(-0.1348)，而窄维度的事先监管(0.6525)也高于世界平均水平(0.0116)。与中国相比，德国和美国在上述两方面都较低(在德国，-2.0393，-3.5234；在美国，0.7411，-0.1435)。

第四，中国政府在建立严格的银行会计制度方面表现最佳。以窄维度的政府授权非政府机构进行监管的力度为例，中国(1.3114)与德国持平，是样本中最高的，远高于美国(0.4085)以及世界平均水平(0.4085)。

最后，由于中国目前仍然实行隐含的存款保险制度，相应的明确的存款保险制度涵盖范围(-1.0031)要远低于世界平均水平(-0.5036)、德国(0.7080)以及美国(5.0591)。

总结以上结果，尽管严格的银行会计制度有助于中国金融业的发展，但是中国的商业银行监管主要依赖政府拥有银行，监管机构的独立性很低^③，同时又设置严格的事先监管要求，以及缺乏明确的存款保险制度，这些都不利于中国金融业的发展。与世界平均水平以及德国和美国的比较，更加突出了以上中国商业银行监管模式的不足，尤其是表现在最高的国有银行比例和最强的事先直接监管。

七、结论及政策含义

基于世界上118个国家在1998~2000年间的银行监管的横截面数据，本文回答了以往文献所忽略的两个问题：第一，当今世界有几种主要的银行监管模式？第二，不同的监管模式对金融发展水平有什么样的影响？此外，通过与世界平均水平以及德国和美国的比较，我们还具体分析了中国商业银行监管模式的特点。

我们发现，银行监管措施在国家间的不同使用表现在以下四个维度上：①政府拥有银行的程度；②政府直接监管银行的力度；③政府授权非政府机构进行监管的力度；④明确的存款保险制度涵盖范围。不同国家在监管银行时可能采用上述四种维度的不同组合，因而世界各国的商业银行监管呈现出三种主要模式：①印度-中国型，其特点是极度依赖政府直接拥有银行；②南非-菲律宾-墨西哥型，其特点是强调政府对银行的直接监管，而忽视市场力量的作用；③德国-美国-瑞士-法国型，其特点是主要依赖非政府力量来监管银行，而较少使用政府直接监管措施。

分析不同的监管维度对金融发展水平的影响，结果表明，灵活的监管机构、非政府机构的监督，以及涵盖面广泛的存款保险制度都对金融业的发展有正面影响，而政府直接拥

① 2003年中国银监会成立，在此之前中央人民银行负责监管商业银行。

② 负值的出现是由于我们使用了第一主因子法构造银行监管变量，以下相同。

③ 银监会的建立可能会增强监管机构的独立性，这有待于今后进一步的研究检验。

有银行，以及事先或事后的政府直接监管都不利于金融业的发展。进而，总结不同的监管模式对不同监管维度的依赖性，得到以下结论：在促进金融业发展方面，德国 - 美国 - 瑞士 - 法国型的监管模式最佳，南非 - 菲律宾 - 墨西哥型次之，而印度 - 中国型最差。

比较中国与世界平均水平以及德国和美国的商业银行监管模式，我们发现：尽管在中国最为严格的银行会计制度有助于推动金融业的发展，但是最高的国有银行比例、较低的监管机构的独立性、最为严格的事先监管要求以及缺乏明确的存款保险制度，这些都对中国金融业的发展有着负面影响。

本文对当前中国银行业改革有着重要的政策含义。首先，政府应当鼓励民营银行的设立，进而逐步减轻对国有银行的依赖；其次，银监会的运行要特别注意保持其监管的独立性；再次，银行监管机构应该逐步放松直接管制^①，例如适当降低市场准入标准，放宽银行经营范围；最后，在适当的时机应当把中国当前隐含的存款保险制度明确地外部化。

参考文献

- [1] Barth, James R., Gerard Jr., Caprio, and Ross Levine, *The Regulation and Supervision of Banks around the World: A New Database*, World Bank Working Paper, 2001.
- [2] Barth, James R., Gerard Jr., Caprio, and Ross Levine, *Bank Regulation and Supervision: What Works Best?* World Bank Working Paper, 2002.
- [3] Basel Committee, *Core Principles for Effective Banking Supervision*, 1997.
- [4] Basel Committee, *The New Basel Capital Accord*, 2002.
- [5] Beck, Thorsten, Asli, Demirgüç-Kunt, and Ross Levine, *A New Database on Financial Development and Structure*, World Bank, 1999.
- [6] Beck, Thorsten and Ross Levine, *Stock Markets, Banks, and Growth: Panel Evidence*, *Journal of Banking and Finance*, forthcoming, 2002.
- [7] Caprio, Gerald Jr., *Banking on Crisis: Expensive Lessons from Recent Financial Crisis*, World Bank Working Paper, 1998.
- [8] Demirgüç-Kunt, Asli, and Edward J. Kane, *Deposit Insurance Around the Globe: Where Does it Work?* *Journal of Economic Perspective*, 2002, Vol. 16, pp175 ~ 195.
- [9] Dowd, Kevin, *Introduction: The Experience of Free Banking*, In Kevin Dowd (ed.), *The Experience of Free Banking*, Routledge: London and New York, 1992.
- [10] Flannery, Mark J., *The Faces of "Market Discipline"*, *Journal of Financial Services Research*, October/December, 2001, pp107 ~ 119.
- [11] Freixas Xavier, and Jean-Charles Rochet, *Microeconomics of Banking*, The MIT Press, Cambridge, Massachusetts, London, England, 2001.
- [12] Garcia, Gillian G. H., *Deposit Insurance: A Survey of Actual and Best Practices*, Working Paper 99/54, IMF, 1999.
- [13] Herring, Richard J. and Anthony M. Santomero, *What is Optimal Financial Regulation?* Wharton Financial Institutions Center Working Paper, 2002, pp0 ~ 34.
- [14] La Porta, Rafael, Florencio, Lopez-de-Silanes, and Andrei Shleifer, *Government Ownership of Banks*,

^① 刘志君等人(2002)的研究发现放松银行管制可以推动中国宏观经济增长。

Journal of Finance, 2002, Vol.57.

- [15] Levine, Ross, Bank-Based or Market-Based Financial Systems: Which is Better? Journal of Financial Intermediation, 2002.
- [16] 赵志君, 马跃, 郭益耀, 曾澍基, 姚兆锋, 刘树成: 《银行业放松管制的理论分析与宏观效果——兼论中国加入 WTO 对大陆和香港的影响》, 《经济研究》, 2002 年第 6 期, 第 14~22 页。

附表 1 世界上主要的商业银行监管模式

	商业银行监管模式		国家或地区	观测值	样本比例
GB95 或 GOVB ANK ≥ 50%	PRECONDN > 总样本的中位数		阿根廷、孟加拉、不丹、印度、罗马尼亚、斯洛伐克	6	5%
	PRECONDN ≤ 总样本的中位数		阿尔巴尼亚、奥地利、白俄罗斯、巴西、保加利亚、布隆迪、中国、捷克、埃及、希腊、冰岛、以色列、牙买加、哈萨克斯坦、莱索托、马尔代夫、波兰、俄罗斯、卢旺达、斯洛文尼亚、斯里兰卡、中国台湾、土耳其、土库曼斯坦、委内瑞拉、南斯拉夫	26	22%
	数据缺失		越南	1	1%
GB95 和 GOVB ANK < 50%		PRECONDN ≤ 总样本的中位数	爱尔兰、立陶宛、南非、圣多美和普林西比、英国	5	4%
	EXANTEN 或 EXPOST > 总样本的中位数, 并且 EMPOWERMENTIN ≤ 总样本的中位数	PRECONDN ≤ 总样本的中位数	亚美尼亚、阿塞拜疆、巴林、玻利维亚、波斯尼亚 - 黑塞哥维纳、博茨瓦纳、英属维尔京群岛、柬埔寨、格鲁吉亚、加纳、直布罗陀、危地马拉、圭亚那、匈牙利、印尼、日本、约旦、肯尼亚、吉尔吉斯斯坦、黎巴嫩、马其顿、马拉维、马耳他、毛里求斯、墨西哥、摩洛哥、尼日利亚、阿曼、秘鲁、菲律宾、卡塔尔、萨尔瓦多、西萨摩亚、汤加、瓦努阿图、赞比亚、沙特阿拉伯	37	31%
		数据缺失	冈比亚	1	1%
		PRECONDN > 总样本的中位数	阿鲁巴、澳大利亚、加拿大、智利、爱沙尼亚、德国、巴拿马、葡萄牙、美国	9	8%
	EXANTEN 和 EXPOST ≤ 总样本的中位数, 并且 EMPOWERMENTIN > 总样本的中位数	PRECONDN ≤ 总样本的中位数	开曼群岛、克罗地亚、塞浦路斯、丹麦、芬兰、科威特、卢森堡、中国澳门、摩尔多瓦、荷兰、西班牙、瑞士、瑞典、特立尼达 - 多巴哥	14	12%
		数据缺失	法国	1	1%

续上附表

	商业银行监管模式		国家或地区	观测值	样本比例
GB95 和 GOVB ANK <50%	EXANTEN 或 EXPOST > 总样本的中位数, 并且 EMPOWERMENTIN > 总样本的中位数	PRECONDN ≤ 总 样本的中位数	意大利、列支敦士登	2	2%
		数据缺失	比利时	1	1%
	其他	PRECONDN > 总 样本的中位数	格恩西岛、所罗门群岛	2	2%
		PRECONDN ≤ 总 样本的中位数	洪都拉斯、塔吉克斯坦、泰国、特克斯和凯科 斯群岛	4	3%
数据缺失		韩国、拉脱维亚、马来西亚、纳米比亚、尼泊 尔、新西兰、波多黎各、塞舌尔、新加坡	9	7%	

附表 2 商业银行监管维度(宽维度)对金融发展水平的影响

因变量	自变量										
	FINAACTI	GB95	GOVBANK	PRECOND	EXANTE	ONGOING	EXPOST	EMPOWERMENT	DIP	常数项	观测值
	- 0.04 (- 1.70)		0.31 (0.60)	- 1.05 (- 1.52)	- 0.33 (- 0.73)	- 0.45 (- 0.91)	0.21 (0.33)	0.31 (1.08)	- 3.10** (- 2.60)	27	0.35
		0.01 (0.25)	0.3 (0.53)	- 0.74 (- 1.25)	- 0.05 (- 0.10)	- 0.74 (- 1.51)	0.04 (0.05)	0.4 (1.36)	- 4.71*** (- 4.47)	28	0.2287
FINASIZE	GB95	GOVBANK	PRECOND	EXANTE	ONGOING	EXPOST	EMPOWERMENT	DIP	常数项	观测值	R ²
	- 0.02*** (- 3.35)		0.06 (0.64)	- 0.11 (- 0.64)	- 0.03 (- 0.24)	- 0.24* (- 1.76)	- 0.2 (- 1.36)	- 0.02 (- 0.22)	0.18 (0.59)	26	0.5538
		- 0.01 (- 0.66)	- 0.004 (- 0.02)	- 0.04 (- 0.26)	0.03 (0.19)	- 0.33* (- 1.91)	- 0.21 (- 1.11)	0.05 (0.50)	- 0.36 (- 1.20)	27	0.3049
FINAEFFI	GB95	GOVBANK	PRECOND	EXANTE	ONGOING	EXPOST	EMPOWERMENT	DIP	常数项	观测值	R ²
	- 0.04* (- 1.73)		0.35 (0.84)	- 0.71 (- 1.04)	0.02 (0.04)	- 0.17 (- 0.40)	0.43 (- 0.40)	0.3 (1.27)	1.2 (1.20)	31	0.3044
		- 0.02 (- 0.78)	0.28 (0.64)	- 0.52 (- 0.91)	0.2 (0.43)	- 0.33 (- 0.69)	0.31 (0.48)	0.40* (1.81)	0.04 (0.04)	31	0.2054

注: 回归采用最小二乘法, 括号中的数据为 t 检验结果, * (**, ***) 表明回归系数的显著性水平达 90% (95%, 99%)。

附表3 商业银行监管维度(窄维度)对金融发展水平的影响

因变量	自变量										
	GB95	GOVBANK	PRECONDN	EXANTEN	ONGOING	EXPOST	EMPOWERMENTIN	DIP	常数项	观测值	R ²
FINAACTI	- 0.04 (- 1.70)		0.33 (0.61)	- 0.64 (- 1.45)	- 0.06 (- 0.18)	- 0.13 (- 0.30)	0.91 (1.32)	0.28 (1.07)	- 2.82** (- 2.71)	33	0.3928
		- 0.03 (- 1.21)	0.26 (0.43)	- 0.82* (- 1.82)	0.08 (0.18)	- 0.32 (- 0.69)	0.67 (1.05)	0.42 (1.61)	- 3.98*** (- 4.85)	34	0.2979
FINASIZE	- 0.02*** (- 3.05)		0.27* (1.95)	- 0.06 (- 0.74)	- 0.08 (- 0.84)	- 0.21* (- 1.76)	- 0.24 (- 1.40)	- 0.03 (- 0.61)	0.21 (0.82)	32	0.5366
		- 0.02** (- 2.19)	0.2 (1.07)	- 0.18 (- 1.37)	- 0.03 (- 0.20)	- 0.26* (- 1.98)	- 0.32 (- 1.71)	0.03 (0.37)	- 0.12 (- 0.47)	33	0.428
FINAEFFI	- 0.04* (- 1.94)		- 0.02 (- 0.05)	- 0.79* (- 1.84)	0.03 (0.12)	0.01 (0.03)	1.07* (1.99)	0.3 (1.34)	1.41 (1.59)	36	0.4167
		- 0.03 (- 1.68)	- 0.08 (- 0.17)	- 0.88** (- 2.59)	0.2 (0.64)	- 0.19 (- 0.46)	0.83 (1.67)	0.39* (1.71)	0.34 (0.46)	36	0.3741

注:回归采用最小二乘法,括号中的数据为t检验结果,* (**, ***)表明回归系数的显著性水平达90%(95%, 99%)。

附表4 中国银行监管模式的国际比较

	中国	德国	美国	世界的中位数	观测值
政府拥有银行(GB95)	99.45%	36.36%	0	29.1%	69
宽维度的先决条件(PRECOND)	- 0.3381	1.4570	1.4570	0.3243	104
窄维度的先决条件(PRECODN)	- 0.4127	1.3652	1.3652	- 0.4127	113
宽维度的事先监管(EXANTE)	3.5881	- 2.0393	0.7411	- 0.1348	105
窄维度的事先监管(EXANTEN)	0.6525	- 3.5234	- 0.1435	0.0116	118
窄维度的政府授权非政府机构监管银行(EMPOWERMENTIN)	1.3114	1.3114	0.4085	0.4085	113
明确的存款保险制度(DIP)	- 1.0031	0.7080	5.0591	- 0.5036	114

联合外汇干预的实证研究*

陈浪南 黄 洵

一、相关研究述评

中央银行的外汇干预是通过相应的传导机制或传导途径来实现的。在资本完全自由流动条件下，外汇干预有两条传导途径：资产组合途径（portfolio-balance channel）和信号或预期途径（signaling channel）。本文将外汇干预通过资产组合途径对汇率产生的影响称为外汇干预的资产组合效应；相应地，将外汇干预通过信号或预期途径对汇率产生的影响称为外汇干预的预期效应。

（一）资产组合效应的实证研究

对外汇干预的资产组合效应的实证研究，根据所采用的模型的不同，可分为三类：①对传统的资产组合模型进行参数估计和检验；②对加入均值-方差最优化约束条件的资产组合模型进行参数估计和检验；③对包含资产组合模型的宏观经济模型进行参数估计和检验。尽管不同学者采用的模型有所不同，但是模型的框架是建立在资产组合模型的基础上。相应地，相当大部分研究均涉及对风险溢价假定（1-1）式的检验。

$$rp = i - i^* - \frac{ES_{t+1} - S_t}{S_t} = f\left(\frac{B}{w}\right) \quad (1-1)$$

式中： rp 表示本国资产的风险溢价； i 表示本国利率； i^* 表示外国利率， S 表示即期汇率（本币/外币）； ΔS^e 表示本币预期升（贬）值率，且 $\Delta S^e = \frac{ES_{t+1} - S_t}{S_t}$ ； w 表示本国私人部门财富总量； B 表示本国债券。在对预期汇率的处理上，通常假设理性预期。因而对（1-1）式的检验实际上是对资产不完全替代和理性预期的联合假设（joint hypothesis）的检验。尽管大多数研究的结果是拒绝联合假设，但仍不能据此得出资产不完全替代的结论，因为人们无法排除因理性预期不成立而导致联合假设不成立的可能性。这是资产组合效应实证研究中有待解决的一个主要问题。

第一类实证研究直接估计资产组合模型中的资产需求函数或其逆转形式〔式（1-1）〕。学者除了要对上述有关预期汇率进行处理外，还受到中央银行外汇干预数据可得性的限制。由于大多数国家的中央银行不愿向外界披露其外汇干预的详细数据，学者通常以

* 本文为 AFDP (2003) 研究课题和教育部社科“十五”课题成果之一。中山大学王美今教授对该文提出多处修改意见，在此表示感谢。原载：《经济研究》，2004年第5期。

外汇储备的变动来代表中央银行干预行为的发生和干预量大小。但是，外汇储备的变动并不全是由外汇干预引起的，因此在估计时必然存在偏差。研究者们采用了不同的方法来解决这些问题。Dankner 等人（1987）曾对美元/马克、美元/日元、美元/加拿大元的资产需求函数及其逆转形式进行检验。他们将美元视为外币，采用了 20 世纪 70 年代中后期马克债券和日元债券存量的月度数据、加元债券的季度数据作为本币资产存量的值。在对汇率预期的处理上，他们采用了静态预期和理性预期两种假设。在运用两阶段最小二乘法之后，他们发现资产完全替代和理性预期的联合假设在美元/马克和美元/加元的情况下通不过检验。但他们同时认为，欲说明冲销性干预对汇率是否有确定性影响，尚缺乏充分证据。Rogoff（1984）则对 1973 年 3 月至 1980 的 12 月美元对加元的汇率及相应的资产存量的月数据进行研究。他采用理性预期的假设，并分别运用普遍最小二乘法和两阶段最小二乘法对（1-1）式进行参数估计和检验，结果是相对资产存量系数的估计值不仅不显著，而且符号错误。因此 Rogoff 认为相对资产供给的变动与风险溢价之间并不存在明确的正相关关系。Loopesko（1984）的研究则得出相反的结果。由于可获得美联储每日外汇干预的数据，Loopesko 以央行累计外汇干预量取代国内资产存量变动，并且对（1-1）式进行修正。在对美国和其他 G-7 国家货币汇率进行研究后，Loopesko 发现联合假设在大多数情况下均不能通过检验，并且央行累计外汇干预量参数的估计值是显著的。因此 Loopesko 认为，冲销性干预通过资产组合渠道影响汇率的效应是存在的。Ghosh（1992）则采用 1980~1988 年间美元对马克汇率的月数据分别对汇率决定的货币模型和资产组合模型进行研究，他试图运用资产组合模型来解释现实汇率偏离由货币模型决定的汇率的部分。Ghosh 的结论是冲销性干预的资产组合效应是存在的，但是很微弱。

第二类实证研究的方法是由 Frankel（1982）率先提出。Frankel 在传统的资产组合模型中引入均值 - 方差最优化约束。均值 - 方差最优化约束可表述为：投资者在期末真实财富 w_{t+1} 的均值和方差之间进行权衡，以实现财富预期效用最大化。其中 w_{t+1} 可表示为资产组合结构和实际回报率的函数：

$$w_{t+1} = w_t + w_t x_t' r_{t+1} + w_t (1 - x_t' \tau) r_{t+1}^s \quad (1-2)$$

式中： w_t 是期初真实财富； x_t 向量表示投资者资产组合中除美元资产以外其他资产各占的份额； r 向量是对应于各种资产的实际回收率； τ 是单位列向量，则投资者的目标函数为：

$$\max F[E(w_{t+1}), \text{Var}(w_{t+1})] \quad (1-3)$$

将（1-3）式对 x_t 求导并令其为零，可得到一形式与（1-1）式相似的方程。Frankel 的研究同样拒绝了资产完全替代和理性预期的联合假设，但仍无法就相对资产供给与风险溢价之间的联系给出实证上的支持。Frankel 和 Engel 在 1984 年对 Frankel 早期所做的研究加以改进。他们放宽了通胀率预先给定这一假设，并运用最大似然法对模型中的均值 - 方差最优化约束进行检验，但结果仍令人失望。Lewis（1988）在上述研究的基础上引入通货膨胀风险和资产市场扰动这两个不确定性因素，但结果仍与前人的研究无多大区别。Engel 和 Rodriguez（1989）试图解释为何 Frankel 的模型得不到实践数据的支持。他们首先将真

实财富的方差与相关宏观经济数据相联系（如美国的货币供应、石油价格等），然后令方差的形成符合时间序列 ARCH 过程。但他们的模型也不为实际数据所支持。以上研究对预期的处理均为采用理性预期假设。Dominguez 和 Frankel（1993）则另辟蹊径，他们采用市场参与者汇率预期的调查数据（survey data）取代一般的以实际汇率变动率代替预期汇率变动率的预期处理方法，并设计了一个两方程模型以区分冲销性干预的资产组合效应和预期效应。他们得出冲销性干预具有资产组合效应的结论。

第三类研究围绕着包含资产组合模型的宏观经济模型进行。Obstfeld（1983）对德意志联邦银行 1975 ~ 1981 年间的外汇干预进行实证研究。他构建的宏观经济模型包括货币供给、货币需求，德国居民对本国债券的需求以及外国投资者对这类债券的需求。Obstfeld 发现，模型中只有外国投资者对国内债券需求方程中的风险溢价的参数估计值统计上显著。他的结论是冲销性干预的资产组合效应是很有限的。Kearney 和 Mac Donald（1986）则研究了英格兰银行的外汇干预。他们发现格兰银行总是进行冲销性干预，试图消除其干预对基础货币的影响，他们建立了与 Obstfeld 相似的宏观经济模型，对该模型估计的结果是，英国居民和外国居民对英国债券的需求函数中，参数的估计值的符号是正确的，但大多统计上不显著。为了检测英格兰银行外汇干预的有效性，他们还运用模型进行了两次政策操作模拟。第一次模拟英格兰银行的非冲销性干预，第二次模拟冲销性干预。模拟的结果是非冲销性干预引起英镑较大幅度的瞬间升值，冲销性干预的效果则要小得多。Blundell-Wignall 和 Masson（1985）也研究了德国的外汇干预。他们的宏观经济模型中包含了多恩布什的汇率超调模型。他们采用完全信息最大似然法对模型进行估计，结果是风险溢价系数尽管很小，但统计上显著。

纵观上述关于冲销性干预的资产组合效应的实证研究，可以看到，冲销性干预的资产组合效应很少获得实证上的支持。

（二）预期效应的实证研究

对外汇干预的预期效应的实证研究，大体上遵循两条思路进行：①以穆萨（Mussa）1981 年关于外汇干预预期途径的理论为基础，检测央行的外汇干预能否作为未来货币政策变动的信号。②在资产组合模型的框架下，直接考察外汇干预对汇率预期的影响。

Klein 和 Rosengren（1991a）首先对外汇干预能否预示未来货币政策变动进行了研究。他们考察了 1985 年广场会议协议至 1987 年股市崩溃期间，美联储每日外汇干预数据（从财经新闻报导中获得）与再贴现率变动之间的关系。他们发现外汇干预并不能预示央行货币政策的变动。Dominguez（1992）利用了 1977 年 2 月至 1981 年 2 月间的数据对美联储和德意志联邦银行的外汇干预进行研究。首先，她将外汇干预对每周货币供给扰动进行回归。回归的结果表明外汇干预能传递未来货币供给变动的信息。其次，她将风险溢价对外汇干预进行回归，发现干预能影响市场汇率预期。Lewis（1992）研究了 1985 ~ 1990 年间美联储外汇干预与美国货币政策变量之间的关系，发现二者存在密切关系。Lewis 为了减少干预未被完全冲销的可能，将干预数据滞后一个时期，却发现此时外汇干预不能预示除了未来联邦基金利率之外其他货币政策变量的变动。在与 Kaminsky 合作的另一篇论文中，Lewis（1992）同样对 1985 ~ 1990 年间美联储的外汇干预进行研究。他们的结论是：货币

政策走向与外汇干预所预示的方向是相反的。

遵循第二种思路进行预期效应实证研究的代表人物有 Humpage, Dominguez 和 Frankel 等人。Humpage 在其 1989 年的研究中假定外汇干预只通过预期途径影响汇率。他将即期汇率对滞后两天的即期汇率、外汇干预和滞后的利率差进行回归。Humpage 认为, 滞后即期汇率可作为预期汇率的替代, 利率差则可反映货币政策的实际变动。当滞后即期汇率的参数估计值为 1 且外汇干预变量的参数估计值显著不为零时, 表明外汇干预可通过预期途径影响汇率。Humpage 对 1984 年 8 月至 1987 年 10 月间美元/日元和美元/马克的汇率分别进行研究, 结果是外汇干预与未来汇率波动不存在显著的相关关系。此后 Dominguez 等人也从事过类似的研究, 但他们均未将外汇干预的资产组合效应与预期效应区分开来, 只是强调预期效应或像 Humpage 那样假设外汇干预只具有预期效应, 直至 Dominguez 和 Frankel 在 1993 年的一篇论文中做出这样的尝试。他们的研究与以往的研究相比, 有以下不同: ①他们采用 MMS 提供的调查数据 (survey data) 所反映的市场汇率预期作为方程中预期汇率变动率的值。这种方法尽管可以克服以实际汇率变动率代替预期汇率变动率产生的问题, 但也带来新的问题。首先, 有关汇率预期的数据来自对市场参与者的调查, 但调查的结果并不是市场参与者的实际经济决策。其次, 所有被调查对象的预期对市场行为的影响不一定是同等程度的, 使得调查实证对市场行为的反映并不真实。②他们设计了一个两方程模型, 试图运用两个方程来分别检测资产组合效应和预期效应。他们利用 1982 ~ 1984 年间美联储和德意志联邦银行提供的每日干预数据和 MMS 提供的调查数据, 采用工具变量法对两个方程进行参数估计和检验, 结果是外汇干预的资产组合效应和预期效应在统计上显著。

近年来西方学术界对外汇干预的实证研究比较侧重于预期效应的实证研究。综观这些研究的结果, 可以发现, 外汇干预的预期效应在实证上得到的支持比资产组合效应来得多。

(三) 联合干预的实证研究

1985 年广场会议协议之后的 80 年代中后期, 西方七国央行的联合干预实践引起了学术界的注意。研究的主要内容是联合干预是否对汇率产生更大影响, 或者说联合干预是否比非联合干预更有效。Loopesko 在 1984 年研究外汇干预资产组合效应的论文中, 同时对联合干预是否比非联合干预更有效进行了研究。Loopesko 对联合干预所下的定义是: 两个以上国家的货币当局在同一天进行的方向相同的外汇买卖操作。相应地, 非联合干预是一国单独进行的或两个以上国家进行的方向相反的外汇买卖操作。Loopesko 的结论是联合干预与非联合干预的效果无甚差别。Dominguez (1990) 则发现, 在某些时段上, 联合干预比非联合干预更有效。此外, Humpage 和 Osterberg (1992)、Eijffinger 和 Gruijters (1992) 等人也对联合干预进行了研究。从研究的结果来看, 半数的研究认为联合干预比非联合干预更有效, 另半数的研究则认为二者不存在明显差异。

(四) 评价

基于上述实证研究, 学术界对外汇干预持有的较普遍的看法是: 冲销性干预的预期效

应可能比资产组合效应更为重要。也就是说，冲销性干预很可能是通过预期途径，而不是资产组合途径对汇率产生影响的。但是，如果没有国内经济政策相应调整的支持，冲销性干预对汇率的影响是短暂的。

笔者认为，学术界对外汇干预所做的实证研究工作，存在以下不足：①研究者主要对 20 世纪 90 年代以前，尤其是 80 年代中后期发生的外汇干预进行研究。对 20 世纪 90 年代以来国际外汇市场动荡化加剧背景下的外汇干预的实证研究几乎无人涉及。②研究者均认为外汇干预的影响不可能持久，但对外汇干预的影响究竟能延续多长时间，均未给出度量的方法。有鉴于此，笔者试图通过对近期美国、日本中央银行的联合外汇干预进行实证研究，以期对弥补上述不足有所助益。

二、干预分析的实证模型

笔者拟采用 Box 和 Diao (1975) 的干预分析模型来从实证上考察中央银行的联合外汇干预对汇率的影响。时间序列经常受诸如罢工、促销等突发事件或经济政策变化的影响，人们将这类外部事件称为干预。Box 和 Diao 于 1975 年提出的干预分析模型为研究干预事件对时间序列的影响提供了有力的数学工具。该模型的形式是：

$$Y_t = y_t + N_t \quad (2-1)$$

式中：

- (1) Y_t 是被研究的序列；
- (2) y_t 表示干预事件对 Y_t 序列的影响，也称为干预输出；

$$y_t = \frac{\omega(B)B^b}{\delta(B)} \xi_t \quad (2-2)$$

其中： ξ_t 表示干预输入； B 是滞后算子，即 $B^i \xi_t = \xi_{t-i}$ ； $\omega(B)$ 是 B 的 s 次多项式， $\omega(B) = \omega_0 - \omega_1 B - \dots - \omega_s B^s$ ，它测量干预输入最初的预期的影响； $\delta(B)$ 是 B 的 r 次多项式， $\delta(B) = 1 - \delta_1 B - \dots - \delta_r B^r$ ，它测量干预输出或干预影响的持久程度。 b 代表干预输出的时间延迟，即干预事件在时刻 T 发生，其影响在时刻 $T+b$ 才开始显现。

- (3) N_t 表示 Y_t 剔除干预影响后的噪声序列，可用 ARIMA(p, d, q) 模型表示，即：

$$N_t = \frac{\theta(B)}{\varphi(B)} \alpha_t \quad (2-3)$$

其中： α_t 为白噪声序列； $\theta(B)$ 是滑动平均的滞后算子多项式； $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ ； $\varphi(B) = (1 - B)^d \phi(B)$ ， d 示差分阶数， $\phi(B)$ 是自回归算子多项式， $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ 。

- (4) 干预输入 ξ_t 有两种常用的表现形式：一种是阶跃函数形式，它表示干预事件在时刻 T 发生后，其影响一直持续下去：

$$\xi_t = S_t^{(T)} = \begin{cases} 0 & t < T \\ 1 & t \geq T \end{cases}$$

另一种是脉冲函数形式，它表示干预事件在时刻 T 发生，只产生短暂影响：

$$\xi_t = P_t^{(T)} = \begin{cases} 0 & t = T \\ 1 & t \neq T \end{cases}$$

由于脉冲函数可由阶跃函数的一阶差分产生, 即 $P_t^T = (1 - B) S_t^T$, 因而干预输入的阶跃函数形式和脉冲函数形式可相互转化。式 (2-2) 中, 干预输出的表现形式是由 ξ_t , $\omega(B)$, $\delta(B)$ 的形式和 b 的取值共同决定的。 $\delta(B)$ 通常取较简单的形式: $\delta(B) = 1 - \delta B$; 若假定 $b = 1$, 则输出形式 $y_t = \frac{\omega_0 B}{1 - \delta B} S_t^{(T)}$, 其中 $0 \leq \delta \leq 1$ 。当 $\delta = 0$ 时, $y_t = \omega_0 B S_t^{(T)}$, 表示干预的影响在时刻 $T+1$ 才显现, 此后该影响保持不变; 当 $\delta = 1$ 时, $y_t = \frac{\omega_0 B}{1 - B} S_t^{(T)}$, 表示干预的影响自时刻 $T+1$ 起呈无限线性增长; 当 $0 < \delta < 1$ 时, 则干预的影响自时刻 $T+1$ 起呈非线性的增长。

同理, 当干预输出为 $y_t = \frac{\omega_0 B}{1 - \delta B} P_t^{(T)}$ ($0 \leq \delta \leq 1$) 时, 若 $\delta = 1$, $y_t = \frac{\omega_0 B}{1 - B} P_t^{(T)}$, 表示实际影响状况与 $y_t = \omega_0 B S_t^{(T)}$ 相同; 若 $0 < \delta < 1$, 表示一个脉冲式的突发变化在时刻 T 发生, 在时刻 $T+1$ 达到水平 ω_0 , 然后开始衰减, 衰减的速度由 δ 值决定, δ 值越小, 干预的影响衰减越快; 若 $\delta = 0$, 则干预的影响在时刻 $T+1$ 后立即消失。

(5) 综合所述, 干预分析模型为:

$$Y_t = \frac{\omega(B)B^b}{\delta(B)} \xi_t + \frac{\theta(B)}{\varphi(B)} \alpha_t \quad (2-4)$$

由于干预输入 ξ_t 在干预事件发生前取值为零, 因此干预分析模型中 N_t 序列实际上就是干预事件发生前 Y_t 的序列, 即 $N_t = Y_t$ ($t < T$)。也就是说, 干预分析模型将干预事件发生后的 Y_t 序列视为由原先未受干预影响的 Y_t 序列和干预输出两个序列组成。运用干预分析模型进行研究的目的, 就是要区分干预事件发生后的 Y_t 观察值的变化可由原先的 Y_t 序列解释的部分和应由干预影响解释的部分。首先应对干预事件发生前的 Y_t 序列 ($t < T$) 进行识别, 得出其噪声模型的拟合形式。在此基础上对干预输出 y_t 的形式进行估计和检验, 然后再对总体模型进行诊断检验。在模型的总体检验通过之后, 可根据 $\omega(B)$ 和 $\delta(B)$ 的估计形式得到所需的结论。

三、近期日美联合外汇干预的实证检验

笔者选用的样本是从 1996 年 8 月 15 日至 1999 年 1 月 6 日期间国际外汇市场上日元/美元汇率的每日收盘价。自 1995 年末日元/美元汇率开始贬值以后, 日、美两国央行对日元一路下跌采取不干预的默许态度, 直至 1998 年 6 月 15 日日元/美元汇率跌至七年来最低价, 两国央行才在 6 月 17 日联合干预。笔者采用干预分析模型来考察这一次日、美两国央行对日元/美元汇率进行的联合干预的效果, 是基于以下考虑: ①能够在掌握联合干预发生时间的情况下进行研究; ②通过估计并分析干预输出 y_t 的形式, 以得到外汇干预是否对汇率产生影响以及该影响能否持久的结论。

(一) 数据的收集与处理

本研究所采用的汇率数据来源于路透社汇率报价系统。所取样本共包含 631 期数据，其中 1998 年 6 月 17 日前共有 477 期数据，即 $T=478$ 。再假定这一“干预”影响为脉冲形式，即：

$$\xi_t = P_t^{(T)} = \begin{cases} 1 & T = 478 \\ 0 & T \neq 478 \end{cases}$$

(二) 建模步骤

第一步对干预发生前的 Y_t 序列进行识别。首先计算 Y_t 前 477 个观察值的样本自相关函数和偏相关函数（表 3-1），从中可以发现干预发生前的 Y_t 序列是非平稳的，需要进行差分平稳化处理。令 $Z_t = (1 - B) Y_t$ ($t < 478$)， Z_t 的样本自相关函数和偏相关函数（表 3-2）显示 Z_t 是一个白噪声序列，因而干预发生前的 Y_t 序列可以随机游走模型来表示。由于 N_t 序列与干预发生前的 Y_t 序列特征相同，因此 $(1 - B) N_t = \alpha_t$ ，即 N_t 可用 ARIMA(0, 1, 0) 模型表示，其中 $p=0$, $d=1$, $q=0$ 。为了检验该模型是否恰当，采用修正的 Q 统计量，有：

$$Q = n(n+2) \sum_{k=1}^{15} (n-k)^{-1} r_a^2(k) \quad (3-1)$$

对其残差估计值 a_t 进行诊断检验。Ljung 和 Box 等人 (1978) 证明，Q 近似服从自由度为 $(k - p - q)$ 的 χ^2 分布。式中 $n = N - d$ ，N 为样本容量，在此 $N=477$ 。d 为差分阶数，因此 $n=476$ 。取残差的前 15 个自相关估计值 $r_a(k)$ ， $k=1, 2, \dots, 15$ 。将 n 值及 $r_a(k)$ 的值（即表 3-2 中 Z_t 的自相关函数估计值）代入式 (3-1)，可得 $Q=9.199$ 小于 $\chi_{0.05}^2(15) = 24.996$ 。因此将 N_t 序列识别为随机游动模型是妥当的。

表 3-1 干预发生前 Y_t 序列的样本自相关函数和偏自相关函数值

K	1	2	3	4	5	6	7	8
自相关	0.983	0.964	0.946	0.929	0.913	0.898	0.883	0.868
偏自相关	0.983	-0.060	0.027	0.003	0.032	0.007	-0.020	0.017
K	9	10	11	12	13	14	15	
自相关	0.854	0.841	0.826	0.812	0.798	0.784	0.770	
偏自相关	0.026	-0.002	-0.040	0.007	0.005	-0.011	-0.010	

表 3-2 Z_t 序列的样本自相关函数和偏自相关函数数值

K	1	2	3	4	5	6	7	8
自相关	- 0.041	0.019	- 0.020	0.029	0.016	- 0.011	- 0.031	- 0.022
偏自相关	- 0.041	0.018	- 0.019	0.027	0.019	- 0.011	- 0.032	- 0.025
K	9	10	11	12	13	14	15	
自相关	- 0.007	0.070	- 0.060	0.018	0.016	0.055	- 0.049	
偏自相关	- 0.009	0.079	- 0.052	0.013	0.013	0.048	- 0.048	

第二步对干预输出的形式进行识别。通过观察样本数据，发现日、美两国央行采取联合外汇干预行动的当天市场汇率发生较为明显的变化，即干预对汇率的影响不存在输出延迟，故取 $b=0$ 。由于干预输入的取值具有确定性，对 $\omega(B)/\delta(B)$ 结构的识别无法以预白噪声化技术为基础。 $\omega(B)$ 多项式的次数 s 和 $\delta(B)$ 多项式的次数 r 只能从较高阶开始逐步试算。首先对 $s=2, r=2$ 的情形进行参数估计，五个待估计参数 $\omega_0, \omega_1, \omega_2, \delta_1$ 和 δ_2 并不能都通过显著性检验。此后逐个剔除通不过检验的参数，并重新对模型进行估计，最后的估计结果是：只有 ω_0 和 δ_1 的估计值在 1% 的显著性水平上通过检验（括号中的数值是标准差）：

$$\omega_0 = - 6.5571(1.0519), \quad \delta = 0.1724(0.0454)$$

所以，拟合模型：

$$Y_t = \frac{- 6.5571}{1 - 0.1724B} \xi_t + \frac{1}{1 - B} \alpha_t \quad (3-2)$$

其中：

$$\xi_t = P_t^{(T)} = \begin{cases} 1 & T = 478 \\ 0 & T \neq 478 \end{cases}$$

最后，再利用残差估计值对模型进行诊断检验。取残差的前 15 个自相关估计值 $r_{aa}(k)$ ，按下式计算 \tilde{Q} 统计量，有：

$$\tilde{Q} = m(m+2) \sum_{k=1}^{15} (m-k)^{-1} r_{aa}^2(k) \quad (3-3)$$

式中： $m = N - u - d - p$ ，其中 N 是样本容量，在此 $N = 631$ ， $u = \max(r, b + s)$ ，其中 $r = 1$ ， $b = 0$ ， $s = 0$ ，因此 $u = 1$ ；又由于对 N_t 序列拟合的模型为 $ARIMA(0, 1, 0)$ ，因此 $m = 629$ 。将值及 r_{aa} 值代入式 (3-3)，计算出 \tilde{Q} 值为 17.232 小于 $\chi_{0.05}^2(15) = 24.996$ 。因此模型 (3-2) 是恰当的，联合干预是脉冲形式的假设得到了证实。

四、结论与启示

模型中 ω_0 的估计值统计上显著，表明 1998 年 6 月 17 日，美、日联合外汇干预对日元/美元汇率产生了一定的影响。 δ_1 的估计值较小且显著不为零，表明此次干预的影响虽不是瞬间即逝，但衰减较快，并没有持续太长时间。事实上，日本自 1997 年出现经济衰

退以后，国内经济持续低迷，迟迟未出现复苏的迹象。此外，日本国内银行体系的坏账问题也一直得不到解决。在日本国内经济的状况未能好转的情况下，外汇干预对汇率影响的短暂性是预期中的事。换句话说，日元汇率变动的决定性因素，不是外汇干预，而是日本国内经济、国际经济的发展状况和国际资本流动。

至此，我们可以得出这样的结论：联合干预与非联合干预在效果上没有重大差别，外汇干预能影响汇率，但其影响很短暂。对一国货币当局而言，在国际外汇市场动荡加剧的背景下，努力提高干预汇市的能力以对付投机冲击固然是重要的，但是实现汇率政策目标乃至经济运行的外部均衡，决不可仅寄希望于外汇干预。如果货币当局的干预目标得不到国内经济基本面和国际经济环境的支持，那么干预行动将难以获得预期的效果。

本文的分析及其结论对我国央行的外汇干预又有怎样的启发和借鉴意义呢？应当指出的是，本文的分析是以资本自由流动为前提的，而我国当前对资本项目仍实行较为严格的管制，国内资本市场与国外资本市场基本处于分割状态。但是也必须看到，逐步放开资本项目管制、实现人民币自由兑换是大势所趋。从未雨绸缪的意义上说，我们可以得到以下几方面的启示：

(1) 央行外汇干预的有效性。一方面，不能单纯依靠外汇干预来实现稳定人民币汇率的目标，尤其是不能将其看成是维持人民币汇率稳定的根本手段。不可否认，近年来我国中央银行的外汇干预对维持人民币汇率的稳定发挥了重要的作用，但这与我国对资本项目实行较为严格的管制是分不开的。一旦资本项目管制放松，国内外资本相互流动，人民币汇率受到各种因素影响的可能性将大大加强。当国内经济出现较大的失衡或缺陷时，央行的外汇干预就有可能失效，1997年亚洲金融危机中东南亚国家央行的汇率干预就是一个很好的例子。因此，实现人民币汇率的相对稳定，关键在于确保国内经济的持续、均衡发展，保证金融体系的健康运行。另一方面，当人民币实行自由兑换后，国内资本市场与国际资本市场实现一体化，即使国内经济正常运行，来自外部的因素仍有可能造成人民币汇率的不稳定。因此，维持人民币汇率的稳定，增强我国央行的外汇干预能力和确保国内、国外经济的均衡发展同样重要。

(2) 央行外汇干预的合理性。从我国央行在银行间外汇交易市场上干预的实践看，其干预目标是确保现行人民币汇率的稳定。由于我国目前还没有人民币均衡汇率的确定机制，央行在外汇市场上的干预缺乏均衡汇率的指示器，实际买卖时只能以现行汇率的稳定为目标，而不管现行市场汇率是否合理。在不考虑其他因素的情况下，就我国目前的汇率形成机制，目前我国银行间外汇交易市场是一个不完全的市场，其形成的汇率存在不合理性。笔者认为，我国央行如要使其干预目标趋于合理性，首要的任务是建立人民币均衡汇率的确定机制，以该机制确定的均衡汇率作为外汇干预的主要依据。

(3) 外汇干预传导的途径。央行外汇干预的传导途径主要有资产组合途径和预期途径。就我国的现实看，随着金融体制改革的深化和资本项目管制的逐步放开，我国外汇市场的供求量将增大，真实性外汇交易所占比重将减少，央行通过干预改变外汇市场容量的能力将下降，预期途径可能成为央行外汇干预的主要传导途径。因此，我国应加强对预期途径的研究，为将来放松资本项目管制时央行有效干预汇市进行理论上的准备。因此，我国应加强对资本项目放松管制条件下外汇干预传导途径的研究，以提高央行的外汇干预能

力。

(4) 央行外汇干预的制度环境性因素。目前我国央行外汇干预在某种程度上受不完善的制度性因素所制约。除了前述汇率形成机制不合理外，货币市场不发达也导致央行无法灵活运用其他国家最常用的冲销工具——公开市场业务。此外，利率非市场化则使央行不能充分利用利率这一最主要的间接干预手段来配合直接干预。因此，要提高我国央行的外汇干预能力，还应改进当前外汇干预的制度环境和作用条件。

(5) 外汇干预的技巧。目前我国央行可使用的冲销工具较少，可以借鉴国外做法，设立类似的中央银行证券以作为外汇干预的冲销工具。此外，在干预时机与力度的选择上，应借鉴国外央行的经验，在汇率变动初期，人们尚未形成普遍预期时开始干预，之后逐步降低干预力度。再则，为提高干预效果，央行可根据实际需要采用间接干预手段等等。因此，我国央行应借鉴国外央行外汇干预的做法，提高外汇干预的技巧。

参考文献

- [1] Box, G. E. P. and Tiao, G. C., *Intervention Analysis with Applications to Economic and Environmental Problems*, *Journal of American Statistical Association*, March 1975, Vol.70.
- [2] Cukierman Alex, *Central Bank Behavior and Credibility: Some Recent Theoretical Developments*, *Federal Reserve Bank of St Louis Review*, 1987, Vol.68.
- [3] Danker, D. J., Hass, R. A., Henderson, D. W., Symansky, S. A. and Tryon, R. W., *Small Empirical Models of Exchange Market Intervention*, *Journal of Policy Modeling*, 1987, Vol.9, pp143 ~ 173.
- [4] Dominguez, k., *Market Responses to Coordinated Central Bank Intervention*, *Carnegie Rochester Conference Series on Public Policy*, 1999, Vol.32, pp121 ~ 164.
- [5] Dominguez, k. and Frankel, J. A., *Foreign Exchange Intervention: An Empirical Assessment*, In Jeffrey A. Frankel, ed., *On Exchange Rates*, Cambridge, MA: MIT Press, 1993.
- [6] Edison, H. T., *The Effectiveness of Central Bank Intervention: A Survey of Literature After 1982*, *Special Papers in International Economics*, Princeton University, New Jersey, 1993.
- [7] Ghosh, *Is It Signaling? Exchange Intervention and The Dollar-deutschemark Rate*, *Journal of International Economics*, 1993, Vol.32, pp 201 ~ 220.
- [8] Kaminsky, G. L. and Lewis, K. K., *Does Foreign Exchange Intervention Signal future monetary policy?* Board of Governors of the Federal Reserve, Washington D. C., 1992.
- [9] Klein, M. and Rosengren, E., *Foreign Exchange Intervention as a Signal of Monetary Policy*, *New England Economic Review*, May/June, 1991a, pp39 ~ 50.
- [10] Klein, M. and Rosengren, E., *What Do We Learn From Foreign Exchange Intervention?* Working Paper, Clark University, 1991b.
- [11] Loopesko, *Relationships among Exchange Rates, Intervention, and the Interest Rates: An Empirical Investigation*, *Journal of International Money and Finance*, 1984, Vol.3, pp257 ~ 277.
- [12] Mussa, M., *The Role of Official Intervention*, Group of Thirty, Occasional Papers 6, 1981.
- [13] Osterberg, W. P. and Humes, R., *The Inaccuracy of Newspaper Reports of US Foreign Exchange Intervention*, *Federal Reserve Bank of Cleveland, Economic Review*, 1993, Vol.4, pp25 ~ 33.
- [14] Rogoff, k., *On the Effects of Sterilized Intervention: An Analysis of Weekly Data*, *Journal of Monetary Economics*, 1984, Vol.13.

中国 A 股与 B 股的市场分割性检验^{*}

邹功达 陈浪南

一、市场分割研究理论述评

股票市场分割 (market segmentation) 检验是对两个或多个市场的股票价格是否遵循统一的定价模式进行假设检验。实证结论支持统一定价假设则称市场是一体化 (integration) 的, 否则称市场是分割的。所谓统一的定价, 指的是这样一种情况, 即投资者在不同市场上投资于相似的金融工具, 获得的经风险调整后的预期收益率是相等的。

国外文献对股票市场分割研究的论述颇丰, 自 20 世纪 70 年代就有了有关这方面的研究。早期的研究是试图通过两个市场指数之间的相关关系来区分两个市场是一体化的还是分割的; 但 Adler & Dumas (1975, 1983), Solnik (1974b) 的研究表明, 市场指数之间的协方差不能证明市场是一体化还是分割的。

Solnik (1977) 就国际资本市场是一体化还是分割的问题进行了论述, 并且对市场分割的检验作了理论探讨, 阐述了对国际市场的一体化或分割性进行实证检验的困难。他认为, 对市场分割的有效检验方法似乎应该是先假定一种引起市场分割的不完美的市场型态, 然后去研究这种型态对投资组合最优化和资产定价的特定影响。

Stehle (1977) 以 Sharpe-Lintner CAPM 模型为基础, 假定投资者具有对数效用函数, 推导出了检验市场分割与市场一体化假设的两类模型。然后, 运用上述模型, 以美国纽约股票交易所上市的股票月度数据为样本, 检验美国市场与包括比利时、加拿大、法国、德国、意大利、日本、荷兰、瑞士、英国和美国在内的世界股票市场之间分割或一体化的关系, 结果发现美国市场与世界市场之间既拒绝分割也拒绝一体化。

Stehle 首次以 CAPM 模型为基础对市场分割与一体化检验方法进行革新, 使市场分割性研究由以前分析指数相关性向前推进了一大步, 为后来学者的同类研究提供了方法论指导。其不足之处是对市场的假设要么是完全分割要么是完全一体化, 没有在模型中设置一种中间情况, 即中等程度的分割。

Vihang Errunza 和 Etienne Losq (1985) 检验了世界资本市场的“中等程度的市场分割” (mild segmentation) 假定。他们按照 Solnik (1977) 建议, 引进了一种不完美的市场型态——中等程度的市场分割。这种型态假定有一批投资者由于政府的限制不能交易某些证券构成的证券子集 (a subset of securities), 而其他的投资者可以不受限制地投资于任何证券,

^{*} 本文为国家自然科学基金项目 (79800010 和 79870039)、教育部社科基金以及中加教育合作课题 (CCUIPP) 成果之一。原载:《经济研究》, 2002 年第 4 期。

于是这一证券子集就拥有一个“超额的”(super)风险溢价。

在中等程度市场分割的假定下, Vihang Errunza 和 Etienne Losq 首先推导出一个定价模式,接着在定价模式的基础上演绎出检验模式,然后利用从美国随机选取的股票以及九个非发达国家中选取的具有大交易量的股票 1976~1980 年的月收益率数据作为样本进行实证检验,检验结果没有与中等程度市场分割假定相违背,因此可以认为中等程度的市场分割得到支持。

Errunza 和 Etienne Losq 首次在市场分割性研究中假设中等程度分割情况,而在此之前研究的要么是完全分割,要么是完全一体化。但是这种方法要求选取的样本股的交易要非常活跃,成交量要大,这样就限制了该方法的适用性。

Philippe Jorion 与 Eduardo Schwartz (1986) 采用 CAPM 模型研究股票市场指数的分割问题,他们认为,如果市场是一体化的,则世界市场指数具有均值-方差有效性(mean-variance efficient),这样,只有被 CAPM 定价过的风险才是相对世界市场指数的系统风险;另一方面,完全分割的市场就意味着只有国内因素,即国内系统风险才进入资产定价模型。他们以 CAPM 为理论模型,采用类似于 Stehle (1977) 推导出来的检验模式,运用极大似然估计方法,考察了加拿大股票市场与整个北美市场一体化还是分割的问题。Jorion 和 Schwartz 利用 1968~1982 年多伦多股票交易所上市的 749 只股票的月收益率(选取样本时要求每只股票月观察值达 60 个以上)及整个北美股票市场指数进行实证检验,结果发现市场一体化假设遭到拒绝,而市场分割假定通过假设检验。其后,他们进一步将样本分为两个子样本,一个包括同时在加拿大与美国上市的股票(即交叉上市),另一个不包括交叉上市的股票,实证结果表明两个子样本都拒绝市场一体化假定,从而说明两个市场是分割的。

Jorion 和 Schwartz 的检验,在理论模式上沿袭前人的成果,检验模式在 Stehle 的方法上有所推进,尤其改进了对交易不活跃股票的 Beta 系数的估计,其结论为后人研究提供对比基础。

Mustafa N. Gultekin, N. Bulent Gultekin 和 Alessandro Penati (1989) 运用多因素资产定价模型(multi-factor asset pricing model),利用日本 1977 年 1 月 1 日至 1984 年 12 月 31 日股票周收益率数据,研究了日本与美国股票市场的一体化问题。1980 年 12 月日本颁布了《外汇外贸管理法》,完全取消资本管制。他们分别以该法颁布之前的四年及之后的四年作为子样本研究两个国家的资产定价模式,结果发现在该法颁布之前的四年内日本与美国股票市场的风险定价是不同的,而在该法颁布之后的四年内风险定价相同。这就说明两国的股票市场一体化程度随着市场自由化的深化而增强,进而表明市场分割是由政府的资本管制政策引起的。

Gultekin, Gultekin 和 Penati 首次在市场分割研究领域采用多因素模型,丰富了市场分割研究的方法论;但是这一方法有一个最根本的缺陷,就是纳入模型的因素的选取带有较大主观性,这势必影响研究结论的可靠性。

Usha R. Mittoo (1992) 运用了 CAPM 与多因素资产定价两种模式,重新检验了 1977~1986 年期间(这一期间相对没有资本管制)加拿大与美国股票市场分割性。在这两种模式下,检验结果显示这一时期两国股票市场的一体化程度不断加强:1977~1981 年两个市

场是分割的，与 Jorion 和 Schwartz (1986) 对 1968 ~ 1982 年两个市场的检验结论相似；而 1982 ~ 1986 年则支持市场一体化假定。此外，他们运用多因素资产定价模式，考察了加拿大上市公司在美国交易所或 NASDAQ 交叉上市的股票及没有在美国上市而只在加拿大本土上市的股票，检验结果表明交叉上市的股票满足一体化假定，而没有交叉上市的股票更倾向于市场分割假定。

Mittoo 采用 CAPM 模型进行检验时，其方法与 Jorion 和 Schwartz (1986) 相类；采用多因素定价模型检验的方法与 Gultekin, Gultekin 和 Penati (1989) 相似。他的突出贡献是分别采用两种模式检验同样两个市场的分割性，这样就检查了这两个模式进行市场分割检验的效果大小，结果发现这两种方法的检验力很相近，多因素模式检验力稍强些，但是多因素方法有我们前文提到过的因素选取有主观性的缺陷。

John Y. Campbell 和 Yasushi Hamao (1992) 运用实证研究方法，提供了美国与日本的长期资本市场一体化的证据。他们利用美国与日本的证券组合的每月超过美国国库券 (treasury bill) 利率的超额收益率的可预见性来研究两个国家的长期资本市场的一体化问题。在 1971 ~ 1990 年期间，两个国家可采用相似的变量 (包括股息 - 价格比和利率) 来预测超额收益。此外，在 80 年代，采用美国的这些变量有助于预测日本股票的超额收益，这就说明两个国家的预期超额收益是同向变动的，从而表明两个国家的长期资本市场趋向一体化。

Campbell 和 Hamao 为市场分割提供了一个新的检验方法，但这一方法只适用于该文研究的这种特殊案例中，其推广有很大难度。

国内研究情况：从公开发表的文献看，国内在这方面的研究目前还是空白，从而本文的研究没有前人的研究可供参考，启迪思维；尽管有国外的相关文献可供参考，但毕竟国外的市场与我国股票市场有着巨大的差别，所以本文对我国股票市场分割研究在国内的该领域研究算是开一个头，不免有差错及考虑不周之处。

二、A、B 股市场分割的实证检验设计

下文先进行 B 股市场与整个股票市场之间的一体化 (分割性) 检验。

(一) 样本选取

为了比较不同时期市场一体化或分割性的变化，我们选取两个不同的样本期进行研究。

第一个样本：取 1994 年 1 月 15 日前上市的 B 股为样本股，以 1994 年 1 月至 2000 年 6 月的交易收盘价为研究对象。这样的样本股上海有 20 只，深圳有 19 只，其上市公司都是双重挂牌 (dual listing) 公司。在检验过程中， β 分别以间隔双周及一个月的收益率计算，用以比较 β 计算间隔期之不同对检验结果的影响。

第二个样本：取 1997 年 6 月 30 日前上市的 B 股为样本股，以 1997 年 7 月至 2000 年 6 月的交易收盘价为研究对象。这样的样本股上海有 41 只，其中只发行 B 股而无 A 股的公司有 5 家；深圳有 47 只，其中只发行 B 股而无 A 股的公司有 9 家。由于样本期短，而估

计 β 至少需要 60 个收益率观察值, 所以 β 以间隔双周的收益率计算。

(二) 检验模式的选择与调整^①

给定 Black 的 CAPM 模型:

$$E(R_i) - R_F = \gamma_0 + \gamma_1 \beta_i \quad (1)$$

式中: $E(R_i)$ 为资产 i 的期望收益率; R_F 为无风险收益率; γ_0 为零- β 资产收益率与无风险收益率之差; $\gamma_1 = E(R_M) - R_F - \gamma_0$, R_M 为市场组合收益率。

如果 A、B 股市场一体化, 则综合市场指数具有均值-方差有效性, 即股票定价完全由综合市场组合收益率 R_t 决定, B 股指数收益率 R_B 不进入资产定价模型, 也就是相对于 B 股市场组合收益率 R_B 的系统风险 β_i^B 对资产定价没有任何解释力, 但是, 很显然不能通过一个对 R_B 的单元回归来加以检验, 因为 B 股市场组合收益率与综合市场组合收益率正相关, 而对两个市场组合收益率的多元回归则会由于存在多重共线性而无法进行。

因此, 我们必须将 B 股市场组合收益率中与综合市场组合收益率不相关的部分分离出来, 我们取 B 股指数收益率代表 B 股市场组合收益率, 取综合指数收益率代表综合市场组合收益率, 这样通过作 B 股指数收益率对综合指数收益率的投影就可将不相关部分分离出来。

$$R_B - R_F = \alpha_0 + \alpha_1(R_t - R_F) + V_{B-1} \quad (2)$$

式中: V_{B-1} 为回归残差项, 其含义是 B 股指数收益率中与综合指数收益率无关的那一部分, $E(V_{B-1}) = \text{cov}(V_{B-1}, R_t) = 0$ 。

于是市场一体化的检验模式为:

$$E(R_i) - R_F = \gamma_0^i + \gamma_1^i \beta_i^1 + \gamma_2^{B-1} \beta_i^{B-1} \quad (3)$$

式中: $E(R_i)$ 为股票 i 的期望收益率, R_F 为无风险收益率; $\gamma_0^i, \gamma_1^i, \gamma_2^{B-1}$ 为常数; β_i^1 为相对于综合市场指数收益率的系统风险; β_i^{B-1} 为相对于残差 V_{B-1} 的系统风险。

如果市场完全一体化, 则 $\gamma_2^{B-1} = 0$ 。

一只股票 i 的风险收益率可分为三部分: 第一部分与综合市场收益率完全相关, 第二部分与残差完全相关, 第三部分与这两个因素都无关。于是有:

$$R_{it} - R_{Ft} = \alpha_i + \beta_i^1 R_{it} + \beta_i^{B-1} V_{B-1,t} + e_{it}^1 \quad (4)$$

式中: $\alpha_i = E(R_{it}) - R_{Ft} - \beta_i^1 E(R_{it})$, 于是有:

$$R_{it} - R_{Ft} = E(R_{it}) - R_{Ft} + \beta_i^1 (R_{it} - E(R_{it})) + \beta_i^{B-1} V_{B-1,t} + e_{it}^1 \quad (5)$$

将式 (3) 代入式 (5), 且由式 (1) 可得 $\gamma_1^i = E(R_i) - R_F - \gamma_0^i$, 于是有:

$$R_{it} - R_{Ft} = \gamma_0^i (1 - \beta_i^B) + \gamma_2^{B-1} \beta_i^{B-1} + \beta_i^1 (R_{it} - R_{Ft}) + \beta_i^{B-1} V_{B-1,t} + e_{it}^1 \quad (6)$$

检验市场一体化, 就是检验式 (6) 中 $\gamma_2^{B-1} = 0$ 是否成立。

对于市场分割检验模式的推导与上述市场一体化检验模式的推导过程相似, 只需将各

^① 市场一体化(分割)检验模式最早由 Stehle (1977) 推导出来, 后经 Jorion 和 Schwartz (1986) 以及 Mittoo (1992) 改造, 运用于加拿大证券市场与世界证券市场之间的一体化检验。本文将他们的检验模式加以修正, 用于中国国内 B 股市场与整个股票市场之间的一体化(分割)性检验。

公式中变量上标或下标 B 与 I 的位置互换即可。

于是，对于市场分割只需检验在式 (7) 中， $\gamma_2^{I-B} = 0$ 是否成立。

$$R_{it} - R_{Ft} = \gamma_0^B(1 - \beta_i^B) + \gamma_2^{I-B}\beta_i^{I-B} + \beta_i^B(R_{Bt} - R_{Ft}) + \beta_i^{I-B}V_{I-B,t} + e_{it}^B \quad (7)$$

(三) 估计方法的推导与选择

第一步，将样本股按 β^I/β^B 或 β^B/β^I 排序后分组。

分别采用式 (8) ~ (11) 估计样本股相对综合指数收益率与 B 股指数收益率的系统风险 β^I 与 β^B 。由于 B 股交易比较不活跃 (infrequent trading)，为了减少交易不活跃股票的系统风险度量误差，我们采用 Dimson (1979) 提出来的方法^①，估计滞后两期 β_{-2} 、滞后一期 β_{-1} 、同期 β_0 、提前一期 β_1 ，然后将它们加总，得到 β 。

$$R_{it} = \alpha_i^I + \sum_{k=-2}^1 \beta_{ik}^I R_{I,t+k} + \epsilon_{it} \quad (8)$$

$$\beta_i^I = \sum_{k=-2}^1 \beta_{ik}^I \quad (9)$$

$$R_{it} = \alpha_i^B + \sum_{k=-2}^1 \beta_{ik}^B R_{B,t+k} + \epsilon_{it} \quad (10)$$

$$\beta_i^B = \sum_{k=-2}^1 \beta_{ik}^B \quad (11)$$

采用两种方法分组：

β^I/β^B ：先按 β^I 大小排序，将样本股等分为 n_1 组，再对每组股票按 β_i^B 大小排序，每组等分为 n_2 小组，这样共得到 $N = n_1 n_2$ 个组，每组有 $\frac{\text{样本股数}}{N}$ 只股票。

β^B/β^I ：先按 β_i^B 大小排序，将样本股等分为 n_1 组，再对每组股票按 β_i^I 大小排序，每组等分为 n_2 小组，这样共得到 $N = n_1 n_2$ 个组，每组有 $\frac{\text{样本股数}}{N}$ 只股票。

计算各组的平均收益率：

$$R_{pit} = \frac{\sum R_{it}}{N} \quad (i = 1, 2, \dots, N) \quad (12)$$

式中： R_{pit} 是第 i 个投资组合在 t 时刻的收益率； $\sum R_{it}$ 表示组合中所有股票在 t 时刻的收益率之和。

第二步，求残差序列 $V_{B-I,t}$ 。

$$R_{Bt} - R_{Ft} = \alpha_0 + \alpha_1(R_{It} - R_{Ft}) + V_{B-I,t} \quad (13)$$

采用普通最小二乘法 (OLS) 对式 (13) 进行估计，得到残差 $V_{B-I,t}$ 。

第三步，求股票组合相对综合指数收益率与 B 股指数收益率的系统风险 $\beta_{pi}^I, \beta_{pi}^B$ 。

$$R_{pit} = \alpha_{pi} + \sum_{k=-2}^1 \beta_{pik}^I R_{I,t+k} + \sum_{k=-2}^1 \beta_{pik}^{B-I} V_{B-I,t+k} + e_{pit} \quad (i = 1, 2, \dots, N) \quad (14)$$

① 使用 Dimson 方法估计的 β 误差减少，参见：Dimson, Risk Measurement When Shares are Subject to Infrequent Trading, 1979.

$$\beta_{pi}^I = \sum_{k=-2}^1 \beta_{pk}^I \quad (15)$$

$$\beta_{pi}^{B-1} = \sum_{k=-2}^1 \beta_{pk}^{B-1} \quad (16)$$

对方程组 (14) 采用普通最小二乘法 (OLS) 进行估计^①, 求出 β_{pk}^I 及 β_{pk}^{B-1} ($k = -2, -1, 0, 1$), 然后根据式 (15) 与式 (16) 计算出 $\beta_{pi}^I, \beta_{pi}^{B-1}$ 。

第四步, 求出 γ_0^I 的初始估计值。

对式 (14) 两边求期望, 再与式 (3) 比较可得:

$$\alpha_{pi} = R_f(1 - \beta_{pi}^I) + \gamma_0^I(1 - \beta_{pi}^I) \quad (17)$$

使用从方程组 (14) 回归中得到的 N 个残差序列, 计算这些方程的同期协方差阵的估计量 Σ , 然后运用广义最小二乘法 (GLS) 对式 (17) 进行回归分析, 估计出:

$$\gamma_0^I = \frac{\alpha'_{pi} \sum^{-1} (1_N - \beta_{pi}^I)}{(1_N - \beta_{pi}^I) \sum^{-1} (1_N - \beta_{pi}^I)} \quad (18)$$

式中: $\alpha_{pi}, \beta_{pi}^I$ 采用式 (14) 估计所得估计值, 1_N 是元素为 1 的 N 阶向量。

第五步, 求出 γ_2^{B-1} 的初始估计值。

由式 (3) 知:

$$R_{pit} - R_{Ft} = \gamma_0^I + \gamma_1^I \beta_{pi}^I + \gamma_2^{B-1} \beta_{pi}^{B-1} + \epsilon_{pit}$$

以前面估计出来的 $\beta_{pi}^I, \beta_{pi}^{B-1}$ 作为已知值, 运用面板数据 (Panel-data) 采用最小二乘法 (OLS) 拟合回归方程:

$$R_{pit} - R_{Ft} = b_0 + \gamma_1^I \beta_{pi}^I + \gamma_2^{B-1} \beta_{pi}^{B-1} \quad (19)$$

$$i = 1, 2, \dots, N; t = 1, 2, \dots, T_i$$

可得 γ_2^{B-1} 的初始值 γ_2^{B-1} 。

第六步, 估计 γ_2^{B-1} 的最终值。

由式 (6) 知:

$$R_{pit} - R_{Ft} = \gamma_0^I(1 - \beta_{pi}^I) + \gamma_2^{B-1} \beta_{pi}^{B-1} + \beta_{pi}^I(R_{it} - R_{Ft}) + \beta_{pi}^{B-1} V_{B-1, it} + e_{pit} \quad (20)$$

这是一个非线性方程, 我们采用泰勒级数展开式 (Taylor series expansion) 将非线性因子 $\gamma_0^I \beta_{pi}^I$ 与 $\gamma_2^{B-1} \beta_{pi}^{B-1}$ 线性化^③:

$$\gamma_0^I \beta_{pi}^I \approx \gamma_0^I \beta_{pi}^I + \gamma_0^I (\beta_{pi}^I - \beta_{pi}^I) + \beta_{pi}^I (\gamma_0^I - \gamma_0^I) = \gamma_0^I \beta_{pi}^I + \gamma_0^I \beta_{pi}^I - \beta_{pi}^I \gamma_0^I \quad (21)$$

$$\begin{aligned} \gamma_2^{B-1} \beta_{pi}^{B-1} &\approx \gamma_2^{B-1} \beta_{pi}^{B-1} + \gamma_2^{B-1} (\beta_{pi}^{B-1} - \beta_{pi}^{B-1}) + \beta_{pi}^{B-1} (\gamma_2^{B-1} - \gamma_2^{B-1}) \\ &= \gamma_2^{B-1} \beta_{pi}^{B-1} + \gamma_2^{B-1} \beta_{pi}^{B-1} - \gamma_2^{B-1} \beta_{pi}^{B-1} \end{aligned} \quad (22)$$

① 在 Gibbons 的多元回归模型 (MVRM) 中对式 (14) 的估计采用 OLS 分别对每个方程进行估计。参见: Gibbons, M. R., Multivariate Tests of Financial Models: A New Approach, 1982.

② Black, Jensen 和 Scholes (1972) 曾采用普通最小二乘法估计。参见: Black, F., M. C. Jensen and M. Scholes, the Capital Asset Pricing Model: Some Empirical Finding, 1972.

③ Gibbins (1982) 在将 Black 版的 CAPM 的非线性限制 $a_i = r(1 - B_i)$, 线性化时也是采用泰勒级数展开式。参见: Gibbons, M. R., Multivariate Tests of Financial Models: A New Approach, 1982.

将式 (21) 与式 (22) 代入式 (20) 可得线性化后的方程：

$$R_{pit} - R_{Ft} - \gamma_0^I \beta_{pi}^I + \gamma_{pi}^{B-1} \beta_{pi}^{B-1} = \gamma_0^I (1 - \beta_{pi}^I) + \gamma_2^{B-1} \beta_{pi}^{B-1} + \beta_{pi}^I (R_{it} - R_{Ft} - \gamma_0^I) + \beta_{pi}^{B-1} (\gamma_2^{B-1} + V_{B-1,t}) + e_{pit}^I \quad (23)$$

对式 (23) 采用系统估计法——似无关回归 (SUR) 方法进行回归, 可估计出系数 γ_0^I , β_{pi}^I , γ_2^{B-1} 与 β_{pi}^{B-1} , 查看回归系数 γ_2^{B-1} 的 T 统计检验值判断它的统计显著性。

市场分割检验的实证估计方法与上述方法类似。

三、实证检验的过程及结果分析

(一) 估计过程

变量取值 上海 B 股指数采用上证 B 股指数 (代码 1a003), 上海综合指数采用上证指数 (代码 1a001); 深圳 B 股指数采用深成 B 股 (代码 2a03), 深圳综合指数采用深圳成分指数 (代码 2a01)。以 B 股指数为参照系, 综合指数取与之对应交易日的数据, B 股个股以 B 股指数交易日为参照, 指数有交易 B 股无交易的, 补充个股收盘价 (取停牌前一天的收盘价代替), 然后采用对指数数据直接求收益率, 对个股价格复权后再求收益率。无风险收益率取央行公布的个人储蓄三个月定期存款利率。

1. 上交所市场一体化 (分割性) 检验

(1) 样本为 1994 年 1 月 15 日前上市的 B 股股票, 以及同期价格指数与无风险收益率, 样本期为 1994 年 1 至 2000 年 6 月, 采用月收益率计算系统风险。

我们从 20 只满足条件的股票中选取了 16 只作为样本股, 计算出 β_i^I 与 β_i^B ; 然后先按 β_i^I (β_i^B) 大小将 16 只股票等分为 2 组, 接着对每组按 β_i^B (β_i^I) 大小再等分为 2 小组, 这样得到 4 个股票组合, 每个组合包括 4 只股票, 求得每个组合 78 期收益率。接下来就按前述估计方法依步估计。

(2) 样本为 1994 年 1 月 15 日前上市的 B 股股票, 样本期为 1994 年 1 月至 2000 年 6 月, 采用双周收益率计算系统风险。

选取与上同样的 16 只股票作为样本, 分组方法也相同, 因为收益率计算间隔变短, 于是求得 162 期收益率。

(3) 样本为 1997 年 6 月 30 日前上市的 B 股股票, 样本期为 1997 年 6 月至 2000 年 6 月, 采用双周收益率计算系统风险。

从 41 只满足条件的股票中选取 36 只作为样本股, 计算出 β_i^I 与 β_i^B ; 然后先按 β_i^I (β_i^B) 大小将 36 只股票等分为 3 组, 接着对每组按 β_i^B (β_i^I) 大小再等分为 3 个小组, 这样得到 9 个股票组合, 每个组合包括 4 只股票; 再求得每个组合 74 期收益率。接下来就按前述估计方法依步估计。

(4) 样本为 1997 年 6 月 30 日前上市仅发行 B 股的公司股票, 样本期为 1997 年 6 月至 2000 年 6 月, 采用双周收益率计算系统风险。

这样的股票共有 5 只, 对其不进行分组, 直接进行一体化 (分割) 性检验。

2. 深交所市场一体化（分割性）检验

(1) 样本为 1994 年 1 月 15 日前上市的 B 股股票，样本期为 1994 年 1 月至 2000 年 6 月，采用月收益率计算系统风险。

从 19 只满足条件的股票中选取 16 只股票作为样本，分组方式与上交所第 1 种情形相同，每个组合的收益率也是 78 期。

(2) 样本为 1994 年 1 月 15 日前上市的 B 股股票，样本期为 1994 年 1 月至 2000 年 6 月，采用双周收益率计算系统风险。

与第 1 种情形相同，只是每个组合的收益率期数增加为 161 期。

(3) 样本为 1997 年 6 月 30 日前上市的 B 股股票，样本期为 1997 年 6 月至 2000 年 6 月，采用双周收益率计算系统风险。

从 47 只满足条件的股票中选取 45 只作为样本，计算出 β_i^I 与 β_i^B ；然后先按 β_i^I (β_i^B) 大小将 45 只股票等分为 3 组，接着对每组按 β_i^B (β_i^I) 大小再等分为 3 个小组，这样得到 9 个股票组合，每个组合包括 5 只股票；再求得每个组合 73 期收益率。接下来就按前述估计方法依步估计。

(4) 样本为 1997 年 6 月 30 日前上市仅发行 B 股的公司股票，样本期为 1997 年 6 月至 2000 年 6 月，采用双周收益率计算系统风险。

满足条件的股票共有 9 只，不对其分组，直接依估计方法进行一体化（分割）性检验。

（二）实证结果分析

按上述估计方法进行实证分析，我们得到的实证结果如表 1 与表 2 所示。

从上交所的情况看，采用月收益率计算系统风险进行检验的结果表明，1994 年 1 月以前上市的 B 股在 1994~2000 年间与整个股市具有较强的一体化。如表 1 所示， γ_2^{B-1} 的估计值都是不显著的，所以我们不能拒绝 B 股与整个股市的一体化；但 γ_2^{1-B} 的估计值在 β^I/β^B 排序方式下在 5% 显著性水平下，在 β^B/β^I 排序方式下在 1% 显著性水平下显著，所以我们拒绝 B 股与整个股市之间的分割性原假设。

在采用双周收益率进行检验时，发现 1994 年 1 月以前上市的 B 股在 1994~2000 年间与整个股市之间的一体化是不稳定的，对排序方式敏感。在 β^I/β^B 排序方式下，既不能拒绝一体化假设也不能拒绝分割性假设，因为表 1 中的实证结果表明， γ_2^{B-1} ， γ_2^{1-B} 的估计值都不显著；但在 β^B/β^I 排序方式下，既拒绝完全一体化也拒绝完全分割， γ_2^{B-1} 的估计值在 5% 的显著性水平下显著，而 γ_2^{1-B} 的估计值在 1% 显著性水平下显著。

1997 年 7 月以前上市的 B 股在 1997~2000 年 6 月期间的一体化较强。从表 1 中的结果可知，无论采用哪种方式排序，的估计值都不显著，所以我们不能拒绝一体化假设；而 γ_2^{1-B} 的估计值都在 1% 的显著性水平下显著，因此我们拒绝分割性假设。

只发行 B 股的公司检验结果表明，既不能拒绝一体化假设也不能拒绝分割性假设，因为 γ_2^{B-1} ， γ_2^{1-B} 的估计值都不显著。

表 1 上海证券交易所市场一体化(分割性)检验结果

样 本	样本期	收益率 间隔期	排序方式	一体化检验		分割性检验	
				γ_0^1	γ_2^{B-1}	γ_0^B	γ_2^{1-B}
1994 年前上市 的双重挂牌公 司	1994 ~ 2000	一个月	β^1/β^B	- 0.0318 (- 2.67)	- 0.0091 (- 0.80)	- 0.0333 (- 2.35)	0.0514 ** (1.95)
			β^B/β^1	- 0.0128 (- 0.88)	- 0.0149 (- 1.11)	- 0.0427 (2.43)	0.0553 *** (3.41)
	1994 ~ 2000	双 周	β^1/β^B	- 0.0330 (- 3.77)	- 0.0053 (- 0.58)	- 0.0475 (- 4.25)	0.0034 (0.21)
			β^B/β^1	- 0.0116 (- 0.99)	- 0.0272 ** (- 2.40)	- 0.1408 (- 9.55)	- 0.1322 *** (- 7.33)
1997 年 7 月 前上市的双重挂 牌公司	1997 ~ 2000	双 周	β^1/β^B	- 0.0045 (- 0.75)	- 0.0030 (- 0.80)	0.0348 (4.13)	- 0.0530 *** (- 5.79)
			β^B/β^1	- 0.0104 (- 1.86)	0.0002 (0.0396)	- 0.0021 (- 0.28)	- 0.0374 *** (- 3.51)
仅发行 B 股 的公司	1997 ~ 2000	双 周	无排序	0.0340 (0.77)	- 0.0276 (- 1.07)	- 0.0536 (- 2.68)	0.0016 (0.17)

注：一体化检验的零假设为 $\gamma_2^{B-1} = 0$ ，分割性检验的零假设为 $\gamma_2^{1-B} = 0$ 。*** 表示在显著性水平为 1% 时拒绝零假设；** 表示在显著性水平为 5% 时拒绝零假设；* 表示在显著性水平为 10% 时拒绝零假设。

综上所述，采用月收益率的一体化强于采用双周收益率；1997 ~ 2000 年期间 B 股与整个股市的一体化比 1994 ~ 2000 年期间强，这表明随着时间的推移一体化有加强之趋势；双重上市公司 B 股与整个股市的一体化强于只发行 B 股的公司。

从深交所的情况看，与上交所不同，采用双周收益率比采用月收益率的一体化强。如表 2 所示，采用月收益率进行检验，1994 年 1 月以前上市的 B 股在 1994 ~ 2000 年间与整个股市既不能拒绝完全一体化也不能拒绝分割性假设。因为从表 2 中的回归结果可以发现，无论是采用 β^1/β^B 还是 β^B/β^1 排序， γ_2^{B-1} 与 γ_2^{1-B} 的估计值都是统计不显著。当采用双周收益率进行分割性检验时，在两种排序方式下， γ_2^{B-1} 与 γ_2^{1-B} 的估计值都在 1% 显著性水平下显著，这样我们就可以拒绝分割性原假设；而采用双周收益率进行一体化检验时，在两种排序方式下， γ_2^{B-1} 与 γ_2^{1-B} 的估计值都不显著，故不能拒绝一体化原假设。

表 2 深圳交易所 A、B 股市场一体化(分割性)检验结果

样 本	样本期	收益率 间隔期	排序方式	一体化检验		分割性检验	
				γ_0^1	γ_2^{B-1}	γ_0^B	γ_2^{1-B}
1994 年前上市 的双重挂牌公 司	1994 ~ 2000	一个月	β^1/β^B	- 0.0357 (- 0.86)	0.0113 (0.31)	- 0.0405 (- 1.01)	- 0.0044 (- 0.19)
			β^B/β^1	- 0.0336 (- 0.99)	0.0061 (0.31)	- 0.0264 (- 0.67)	0.0067 (0.35)
	1994 ~ 2000	双 周	β^1/β^B	- 0.0359 (- 2.68)	- 0.0014 (- 0.19)	- 0.1193 (- 4.79)	0.1642*** (6.58)
			β^B/β^1	- 0.0278 (- 2.46)	- 0.0039 (- 0.59)	- 0.1128 (- 4.35)	0.1387*** (5.98)
1997 年 7 月 前 上市的双重挂 牌公司	1997 ~ 2000	双 周	β^1/β^B	- 0.0227 (- 4.62)	- 0.0052 (- 0.54)	0.0025 (0.31)	0.0691*** (7.39)
			β^B/β^1	- 0.0210 (- 4.07)	- 0.0004 (- 0.05)	- 0.021 (- 0.28)	- 0.0374*** (- 3.51)
仅发行 B 股 的公司	1997 ~ 2000	双 周	无排序	0.0120 (0.65)	0.0669*** (4.88)	- 0.0259 (- 2.20)	- 0.0023 (- 0.26)

注：一体化检验的零假设为 $\gamma_2^{B-1} = 0$ ，分割性检验的零假设为 $\gamma_2^{1-B} = 0$ 。*** 表示在显著性水平为 1% 时拒绝零假设；** 表示在显著性水平为 5% 时拒绝零假设；* 表示在显著性水平为 10% 时拒绝零假设。

采用双周收益率对 1997 年 7 月以前上市的 B 股在 1997 ~ 2000 年 6 月期间的一体化进行检验，我们发现，无论采用哪种方式排序 γ_2^{B-1} 与 γ_2^{1-B} 的估计值都不显著，如表 2 所示，所以我们不能拒绝一体化假设；而进行分割性检验时， γ_2^{B-1} 与 γ_2^{1-B} 的估计值都在 1% 的显著性水平下显著，因此我们拒绝分割性假设，从而认为 1997 年 7 月前上市的 B 股在 1997 ~ 2000 年期间具有较强的一体化。

只发行 B 股的公司检验结果表明，拒绝一体化假设但不能拒绝分割性假设，如表 2 所示， γ_2^{B-1} 的估计值在 1% 的显著性水平下统计显著，而 γ_2^{1-B} 的估计值不显著。

综上所述，采用月收益率的一体化弱于采用双周收益率；采用双周收益率检验，1994 ~ 2000，1997 ~ 2000 年期间 B 股与整个股市的一体化都较强；双重上市公司 B 股与整个股市的一体化强于只发行 B 股的公司，只发行 B 股的公司拒绝一体化假设。

四、结论

从上面实证结果分析可知，上海与深圳两个交易所 B 股与整个股市具有较强程度的一体化，但是一体化对计算收益率的间隔期敏感，因此我们不能得出 B 股与整个股市完

全一体化的结论，尤其是只发行 B 股的公司与整个股市的一体化程度更差些。但尽管如此，实证结果表明， γ_2^{1-B} 的估计值显著非零的次数远多于 γ_2^{B-1} ，这就是说更倾向于拒绝分割性假设。这样，我们认为，总的看来，B 股市场与整个股市在相当程度上是一体化的（这可能与我们的直觉很不一样），由于 A 股市场是构成整个股市的最主要部分，因此 B 股与 A 股一体化程度较高。

参考文献

- [1] Alexander, Gorgon J., Cheol S. Eun, and S. Janakiramanan, Asset Pricing and Dual Listing on Foreign Capital Markets: A Note, *Journal of Finance*, 1987, Vol.42, pp151 ~ 158.
- [2] Bekaert, Geert, and Campbell Harvey, Time Varying World Market Integration, *Journal of Finance*, 1995, Vol.50, pp403 ~ 444.
- [3] Black, F., M. C. Jensen and M. Scholes, The Capital Asset Pricing Model: Some Empirical Tests, In M. C. Jensen (ed.), *Studies in the Theory of Capital Markets* (Praeger, New York, N. Y), 1972.
- [4] Campbell, John, and Yasushi Hamao, Predictable Stock Returns in the United States and Japan, *Journal of Finance*, 1992, Vol.47, pp43 ~ 69.
- [5] Dimson, E., Risk Measurement When Shares Are Subject to Infrequent Trading, *Journal of Financial Economics*, 1979, Vol.7, pp197 ~ 226.
- [6] Errunza, Vihang Etienne Losq, International Asset Pricing under Mild Segmentation: Theory and Test, *Journal of Finance*, 1985, Vol.40, pp105 ~ 124.
- [7] Eun, Cheol S., and S. Janakiramanan, A Model of International Asset Pricing with a Constraint on the Foreign Equity Ownership, *Journal of Finance*, 1986, Vol.41, pp897 ~ 914.
- [8] Fama, E. F., and J. D. Macbeth, Risk, Return and Equilibrium: Empirical Tests, *Journal of Political Economy*, 1973, Vol.81, pp607 ~ 636.
- [9] Foerster, Stephen R. and G. Andrew Karolyi, the Effects of Market Segmentation and Investor Recognition on Asset Prices: Evidence from Foreign Stocks Listing in the United States, *Journal of Finance*, 1999, Vol.54, pp981 ~ 1013.
- [10] Gultekin, N. Bulent, Mustafa Gultekin, and Alessandro Penati, Capital Controls and International Capital Market Segmentation: Evidence from Japanese and American Stock Markets, *Journal of Finance*, 1989, Vol.44, pp849 ~ 869.
- [11] Hietala, Pekka T., Asset Pricing in Partially Segmented Markets: Evidence from the Finnish Market, *Journal of Finance*, 1989, Vol.44, pp697 ~ 718.
- [12] Jorion, Phillippe and Eduardo Schwartz, Integration versus Segmentation in the Canadian Stock Market, *Journal of Finance*, 1986, Vol.41, pp603 ~ 616.
- [13] Kadlec, Gregory B. and John J. McConnell, The Effect of Market Segmentation and Illiquidity on Asset prices: Evidence from Exchange Listings, *Journal of Finance*, 1994, Vol.49, pp611 ~ 636.
- [14] Merton, Robert C., Presidential Address: A Simple Model of Capital Market Equilibrium with Incomplete Information, *Journal of Finance*, 1987, Vol.42, pp483 ~ 510.
- [15] Mittoo, Usha, Additional Evidence on Integration in the Canadian Stock Market, *Journal of Finance*, 1992, Vol.47, pp2035 ~ 2054.
- [16] Scholes, M. and J. Williams, Estimating Betas from Non-synchronous Data, *Journal of Financial Economics*, 1977, Vol.5, pp309 ~ 328.

-
- [17] Solnik , Bruno H. , *European Capital Markets* (Lexington Books , Lexington , Mass) Solnik , Bruno H. , *The International Pricing of Risk : An Empirical Investigation of the World Capital Market Structure* , *Journal of Finance* , 1974a , Vol.29 , pp365 ~ 378.
- [18] Solnik , Bruno H. , *An Equilibrium Model of the International Capital Market* , *Journal of Economic Theory* 1974b , Vol. 8 , pp500 ~ 524.
- [19] Solnik , Bruno H. , *Testing International Asset Pricing : Some Pessimistic Views* , *Journal of Finance* , 1977 , Vol.32 , pp503 ~ 512.
- [20] Stehle , Richard , *An Empirical Test of the Alternative Hypotheses of National and International Pricing of Risky Assets* , *Journal of Finance* , 1977 , Vol.32 , pp493 ~ 502.
- [21] Stoll , Hans R. , and Robert E. Whaley , *Transaction Costs and the Small Firm Effect* , *Journal of Financial Economics* , 1983 , Vol.12 , pp57 ~ 79.
- [22] Stulz , Rene , *On the Effect of Barriers to International Investment* , *Journal of Finance* , 1981 , Vol.36 , pp383 ~ 403.
- [23] Wheatley , Simon M. , *Some Tests of International Equity Integration* , *Journal of Financial Economics* , 1988 , Vol.21 , pp177 ~ 212.
- [24] Gibbons , M. R. , *Multivariate Tests of Financial Models : A New Approach* , *Journal of Financial Economics* , 1982 , Vol.10 , pp3 ~ 27.

信息披露机制对私营企业融资决策的影响^{*}

王宣喻 储小平

一、信息不对称下的私营企业融资决策：一个假说

信用制度，作为人类大规模合作秩序的扩展，其主要的功能就是将分散在广大的互不相识的人群中的有关投资机会的信息汇集起来（汪丁丁，1997）。但是，资本市场却要承担由此而引发的融资风险。资本市场为了识别和分散配置因信息不完全而带来的信贷风险，自发地生成了一种由低到高的层级结构以作为自我保护机制。无论是已经实现金融自由化的国家，还是继续实行金融抑制的发展中国家，其资本市场总是分层的（周其仁等，1998）。在制度层面上，资本市场的这种层级结构跨越了非正式的金融制度和正式的金融制度的连接点，形成了一种涵盖直接融资和间接融资两种资金交易方式，并在地理意义上按融资规模由小到大向外扩展的倒金字塔式的空间结构（王宣喻，储小平，2002），如图1所示。

不同层级的资本市场，分别对应于不同的融资规模和允许使用的资金交易方式，依靠各自适用的制度框架所提供的技术支持以计算处理因信息不完全而带来的潜在的放贷风险。资本市场的层级越高，融资规模越大，贷款人需要借款人提供的信息量越多，融资风险也就越大。但是在实践中，资本市场的信息不完全并不能由金融体系和金融机制的创新活动而得到完美解决，在每一层级的资本市场上，融资双方拥有的信息都是不可能完全对称的。一般说来，融资双方存在的非对称信息大致可分为两种：一种是有关借款者的个人信息；另一种是有关企业经营方面的信息（王宣喻，储小平，2002）。个人信息包括借款人的人品素质、个人能力、风险偏好等方面的信息，企业经营方面的信息包括会计信息、资本运营、技术创新与产品开发、管理控制、产权分配、投资决策、市场营销、生产情况、利润分配、机构设置、人事变动、税收缴纳等方面的信息。个人信息是一种人格化的信息，它只能被有限范围内的家人、亲戚朋友、同学、同事以及左邻右舍的街坊等熟人所掌握和了解。对于陌生人而言，这种人格化的信息是不可观测的，或者要付出高昂的成本才能被观测到。“不可保证性”是人格化信息的根本特点，借款人很难向别人证明自己不是骗子，即使赌咒发誓保证自己是诚实守信的人，投资者也未必会相信。企业经营信息则与企业家的个人素质、能力和风险偏好无关，它是一种可以以数据和书面资料形式反映的非人格化的信息。在大多数情况下，企业经营信息都是一种法律承诺，是具有法律

^{*} 本文系国家自然科学基金资助项目（79970049）的阶段性成果。原载：《经济研究》，2002年第10期，发表时内容有修改。

效力的，因此，这种信息并不专门面向和自己熟识的人群与金融机构，信息一经公开，其他利益相关者就能够根据自己的需要任意加以利用。在正式的金融制度下，资本市场能够量化分析与评估潜在的信贷风险，主要的依据就是这些具有“可计算性”的企业经营方面的信息。企业如想在资本市场上顺利地进行融资活动，就必须解决这两个方面的信息不对称给贷款人带来的资金风险评估问题。

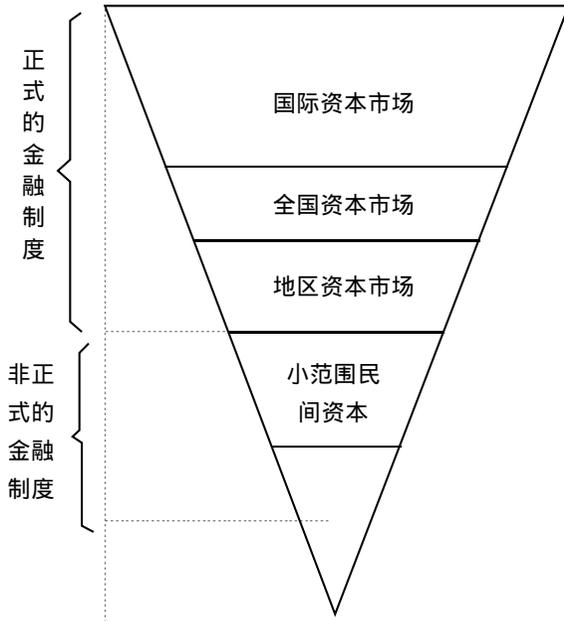


图1 资本市场的层级结构

但是，尽管资本市场都需要解决来自借款申请者个人信息和企业经营信息这两个方面的信息不对称问题，然而，这两种信息在决定信贷决策时所起的作用是大不相同的。在非正式的金融制度下，民间借贷的依据是对借款人的熟识和了解，依靠个人社会关系网络所提供的人格化信息来评估资金安全程度。这种人格化的融资方式基本上是不需要借款人很正式地提供有关企业经营方面的信息的，而且借贷双方很少签订正式的借款合同，有时候甚至完全是“口头合同”，道德准则、风俗习惯、帮规会规、价值观念等非正式的制度规则达成了自发执行契约的限制。而在正式的金融制度下，由于申请借款的人的范围大大扩展，贷款人对大多数借款人的了解都很有有限，而人格化信息的不可保证性又很难让贷款人相信借款人的自我表白或第三方的转述，所以，当融资规模趋大时，贷款人将降低人格化信息在风险评估模型中的影响力，而主要依据企业经营管理方面的非人格化信息来计算信贷风险。在贷款人一整套的信贷风险评估模型中，尤其是在自动信贷评级的计算机系统软件中，有关借款人的个人信息只是作为一个侧面了解的变量，仅作参考。银行和非银行的金融机构以及证券交易所在对借款人进行资信调查时，都会要求企业按照一定的程序和格式报送各种资料，提供符合条件的抵押品和寻求有相应实力的担保人，借贷双方需要签订正式的借款合同，契约履行的第三方监督机制交由国家成文的法律来承担。一般地，资本

市场的层级越高，融资规模越大，国家法律规定借款人必须披露的信息就越详细、越规范。

但是，交易双方在谈判过程中，拥有信息优势的一方总是倾向于输出对自己有利的信息，而控制住对自己不利的信息。在资本市场上，借款人同样也会坚持这一原则：向贷款人提供的信息不会给企业的经营管理带来麻烦和风险。因此，理性的借款人在计算融资成本时，不仅仅考虑资金成本（利息）和手续费、印花税、工本费、咨询费、差旅费、谈判支出等各项交易费用，还会考虑公开或泄露机密信息可能会给企业带来损失的风险成本。

至此，本文提出如下假说：一方面，由于我国目前正处在社会转型期，经济、政治、法律、社会文化等各个方面的体制转换和制度变迁还没有最后完成，产品市场无序竞争，要素市场无序流动，社会信用低下，法制不健全，权力腐败盛行，行政管理效率低下，政府政策缺乏长期一致性，私营企业的外部环境很不稳定。另一方面，我国的私营企业是在极其艰难的环境下冒着巨大的政治风险顽强地成长壮大的。由于长期地被排斥在国有经济体制之外，很难利用正式的制度安排获取企业发展所必需的资源，因此私营企业往往会通过一些变通的手段来获得支持，如戴红帽子、假合资、与国有企业联营等等。这么做的后果是造成私营企业产权模糊，资产归属不明确；有的私营企业甚至蓄意侵吞国家财产，化公为私，但是又不能通过法律程序合法化。此外，很多私营企业内部管理不规范，经营指导思想不正确，没有正式的会计核算制度，瞒报收入，偷漏税款，甚至走私制假牟取暴利。在这种情况下，将经营管理信息控制在企业内部有助于降低企业的经营风险。因此，在不存在制度歧视和政策倾斜的前提下，如果不考虑资金利率和手续费、印花税等交易成本的差异，那么，影响私营企业融资秩序和融资渠道选择的主要因素，将是各层级资本市场信息披露机制给私营企业带来的风险成本的大小。层级越低的资本市场，要求企业披露的经营信息就越少、越不规范，企业融资的风险成本就越小；层级越高的资本市场，要求企业披露的经营信息就越多、越规范，企业融资的风险成本就越大。当私营企业完全垄断企业经营信息时，企业的融资成本最小；此时，在各个资本市场上，相等资金的融资成本是相同的，图2的曲线将变成一条在纵轴截距为 k 的水平线。特别地，如果私营企业连有关企业主人方面的信息也完全垄断，那么，企业将不发生借贷行为，资金周转只依赖企业自身的经营收入。在实践中，完全垄断企业经营信息和个人信息的情况几乎不存在；但是考虑到实际的融资风险，私营企业最优的融资决策仍将是尽可能地依靠自身积累解决资金问题，或是在较低层级的资本市场如民间借贷市场上进行资金融通活动。

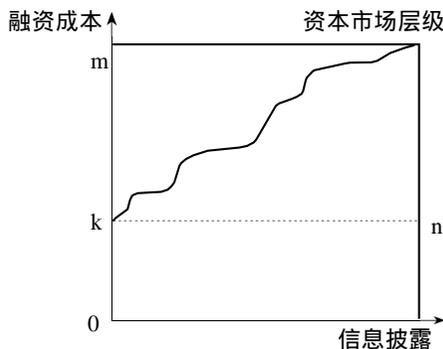


图2 信息披露与私营企业的融资成本

图2说明：下横轴的信息披露指的是企业经营信息，主要包括会计信息、资本运营、利润分配、生产情况、机构设置、管理控制、产权分配、投资决策、市场营销、产品开发与技术创新、人事变动、税收交纳情况等等；上横轴的资本市场层级指的是小范围民间资本市场、地区资本市场、全国资本市场和国际资本市场。

二、私营企业融资问题的问卷调查

我们问卷调查和深度访谈的对象仅限于私营企业主。调查设计采取意愿探问和现状调查相结合的方式，请私营企业主就企业融资问题和企业内部治理与管理这两方面的问题作出回答。课题组共发放问卷136份，回收108份，回收率为79.41%；有效问卷106份，有效率为98.15%。有效样本中，潮汕地区（主要是潮州和汕头）61份，珠三角地区（主要是广州和深圳）26份，浙江及湖北等私营企业比较发达的地区19份。

对私营企业进行实证研究的难度一直是学术界所公认的。由于文化传统、政治体制、社会环境等方面的原因，中国人大多信奉财不外露、树大招风的经营理念。对私营企业而言，能不让别人知道底细就尽量不让人摸清，这也是私营企业长期适应外部环境的结果。所以，我们的问卷调查和访谈无法进行严格的概率抽样，也没有采取邮寄、电话联系等方式，而是通过主持人的社会关系网络直接给比较熟悉的私营企业主发放问卷，并再三说明这次问卷调查是纯粹的学术研究，没有任何的官方背景，研究者将对了解到的所有情况进行严格保密，希望这样能够减少他们的顾忌和疑虑。但即便如此，我们的回收率还是不到80%。尤其是在私营企业遍布的潮州市庵埠镇，我们总共发放了19份问卷，却只回收了1份，而问卷发放的地点还是在彼此觥筹交欢的晚筵上。而在回收的有效问卷中，将“贵公司的资产额”一栏空下未填的共34份，占全部106份有效问卷的32.08%；将“贵公司的主要产品”一栏空下的有14份，占全部有效问卷的13.21%；将“贵公司的法人代表”一栏空下的有18份，占全部有效问卷的16.98%；将“贵公司的员工人数”一栏空下的有16份，占全部有效问卷的15.09%；将资产额、员工人数、主要产品和公司法人代表4栏中任意空下2栏的有22份，占全部有效问卷的20.75%；将以上4栏全部空下的有14份，占全部有效问卷的13.21%。由此可见，在现阶段，私营企业对控制企业经营方面的信息的愿望是非常强烈的。

1. 私营企业到底存不存在融资难的问题

有关“贵公司对于解决日常资金问题的看法”一题，回答情况如下：

答 案	非常容易	很容易	容 易	一 般	较 难	很 难	非常难
份 数	8	2	24	44	20	6	2
比例 (%)	7.55	1.89	22.64	41.51	18.87	5.66	1.89

可以看出，对日常资金来源明显感到容易的私营企业占32.08%，感到不困难的占73.59%，而对日常资金的使用明显感到困难的私营企业仅占26.42%。在感到困难的私营

企业中，程度也有所区别，真正感到很难和非常难的私营企业只有 7.55%。

有关“贵公司对于解决大规模投资或上马大项目所需资金问题的看法”一题：

答 案	非常容易	很容易	容 易	一 般	较 难	很 难	非常难
份 数	4	0	12	24	30	14	2
比例 (%)	4.65	0	13.95	27.91	34.88	16.28	2.33

注：对此题的回答共 86 人。

对解决大规模投资的资金问题明显感到容易的私营企业占 18.60%，感到不困难的占 46.51%，明显感到困难的占 53.49%。但是，感到很困难和非常困难的私营企业只有 18.61%，和感到容易的企业基本持平，而绝大多数的私营企业都没有感受到资金瓶颈的强烈约束。

2. 私营企业的资金来源

有关“贵公司日常的资金来源”一题：

答 案	企业自有资金	银行贷款	国家政策 性贷款	其他单 位借款	亲戚朋 友借款	海外金融 机构贷款	其 他
份 数	86	46	10	16	20	0	6
比例 (%)	81.13	43.40	9.43	15.09	18.87	0	6.02

很明显，私营企业流动资金来源非常广泛，不仅依靠自身积累（81.13%），同时也能够从银行和政府部门获得支持（52.83%）；不仅可以依靠亲戚朋友借款（18.87%），也能够从其他单位借款周转（15.09%），同时还有其他途径筹措资金（6.02%）。这也许能够解释私营企业为何不会感到日常资金周转的困难。值得注意的是，私营企业依靠亲戚朋友和其他单位借款的比例高达 33.96%，再加上企业自有资金，其比例已远远超过依靠银行和政府部门提供的资金支持。

有关“贵公司固定资产和大规模投资的资金来源”一题：

答 案	自有 资金	银行 贷款	国家政策 性贷款	其他单位 借款	亲戚朋 友借款	海外金融 机构贷款	发行股 票证券	风险资 金投入	其 他
份 数	80	48	2	14	10	2	4	0	6
比例 (%)	75.47	45.28	1.89	13.21	9.43	1.89	3.77	0	5.66

和流动资金来源一样，私营企业的固定资金来源也呈现出多渠道性，但主要是依靠企业自有资金和银行贷款。此外，亲戚朋友和其他单位借款的作用仍然不可忽视，其比例为

22.64%。企业自有资金和民间借贷在解决固定资金来源问题上作用也仍然远远大于银行的作用。

有关“您估算一下，贵公司从银行等金融机构获得的贷款支持大约占所需资金的比例”一题：

答 案	10% 以下	10% ~ 20%	20% ~ 30%	30% ~ 40%	40% ~ 50%	50% ~ 60%	60% ~ 70%	70% 以上
份 数	30	18	20	10	4	4	2	4
比例 (%)	32.61	19.57	21.74	10.87	4.35	4.35	2.17	4.35

注：此题实际回答人数为 92 人。

另外还有 9 份问卷，答题者专门在“其他”栏里注明“从来未贷过款”。这里，我们可以很清楚地看出，虽然银行贷款是私营企业资金来源之一，但是在利用比例上是比较小的，73.92%的私营企业只能贷到所需资金的 30% 以下，其余的资金缺口还得依靠企业通过别的途径解决。特别让人吃惊的是，调研企业中竟有 9.78% 的企业表示从未从金融机构贷过款！

3. 朋友间的资金借贷关系

有关“您有没有经常借些钱给朋友周转？如有，那么您为何愿意这么做”一题：

答 案	朋友有财 产抵押	朋友有能 力赚钱	朋友为人 诚实守信	朋友有可行 的计划书	朋友有公司 或不动产	私人交情 不好拒绝	朋友的身 份地位	其 他
份 数	2	14	26	4	0	30	2	10
比例 (%)	2.27	15.91	29.55	4.55	0	34.09	2.27	11.36

注：本题实际回答人数为 88 人。

所有对本题给出回答的人基本都可视为民间借贷的参与者，可见这一比例之高 (83.02%)，这还不包括拒绝回答此题的人中可能还会存在的参与者。由此可以判断，我国民间借贷资本市场的规模是十分惊人的。私人借贷中，财产抵押和正式的计划书基本不起作用，借贷双方的人情面子和对借款人的了解是决定信贷关系的决定性因素。

有关“朋友间的资金互助，一般会签正式合同”一题：

答 案	会，并要有 担保人	看情况，数 额大的要	看情况，老 关系不要	随便写张 借条即可	口头合同，全 凭个人信誉	其 他
份 数	14	38	10	22	16	2
比例 (%)	14.74	40	10.53	23.16	16.84	2.11

注：本题实际回答人数为 95 人。

明确表示不需要签正式合同的占 40%，超过明确要签正式合同的比例（14.74%）。视具体情况而定的比例为 50.53%，说明民间借贷在防范资金借贷风险方面也有一套自发生成的机制。需要说明的是，民间借贷中“正式合同”的概念，就是要有一个熟人做担保人，在一张借款人写下的借据上按手印或签字，以个人的信誉和承诺作为还款担保。这张借据并不是附有各种条款的打印或复印的合同文本，在形式上不拘一格，甚至可以是一张旧烟盒纸。

有关“您从朋友处借款，朋友会不会计利息”一题：

答 案	以国家贷款利息为依据	一般不计息	低于国家贷款利息	高于国家贷款利息	看情况，数额小的不计息	看情况，关系好的不计息	其 他
份 数	16	22	2	22	22	18	0
比例（%）	16.67	22.92	2.08	22.92	22.92	18.75	0

注：本题实际回答人数为 96 人。

在私人借贷中，明确表示不计息的只有 22.92% 的比例，绝大多数的资金拆借都是付息的，并且有 22.92% 的拆借是高于国家利率水平的。这说明民间的资金互助是以牟利为主要目的的。同时，由于人情关系，私人拆借也体现了灵活性，如根据双方关系的亲近程度和资金数额的大小，债权人将视情况决定要不要借款人支付利息。

4. 民营企业与银行等金融机构的关系

有关“您现在从银行等金融机构贷款，您认为它们为什么愿意贷款给您”一题：

答 案	公司规模与实力	公司以往的信用	足够的财产抵押	贷款用途说明与计划书	老板为人诚实可信	老板的经营能力	与银行的私人关系	其 他
份 数	48	46	60	16	20	24	16	8
比例（%）	45.28	43.40	56.60	15.09	18.87	22.64	15.09	7.55

从企业主的角度看，银行等金融机构愿意借出资金，主要是因为足够的财产抵押、公司的规模与实力以及公司以往的信用。同时，企业主的个人信息（能力与诚信因素）也是银行贷款重要的约束条件。这可以说明，企业主也大致可以判定银行等金融机构评估资金风险的主要依据。

有关“您从银行等金融机构贷款时，愿不愿意按银行的要求提供所有的信息”一题：

答 案	愿意, 只要能贷到款	不愿意, 因为这涉及企业机密	看情况, 只要不涉及企业机密就提供	只要银行对所有企业一视同仁就提供	所有报送银行的信息都要经过筛选	其 他
份 数	20	10	40	20	6	0
比例 (%)	20.41	10.20	40.82	20.41	6.12	0

注：本题实际回答人数为 98 人。

明确表示向银行报送信息是有条件的私营企业的比例总计为 77.55%，而表示无条件地按银行等金融机构要求报送所有信息的企业仅占 20.41%，有 50% 以上的企业不愿意因为贷款而公开机密信息。另外，有 2 份问卷特地说明：“这是银行要求的，不提供不行。”这些都表明大多数私营企业在向金融机构提出借款请求时，一般都会综合考虑因为融资而给企业带来的风险。

从以上的统计结果可以得出：①私营企业目前并不存在强烈的资金约束，或者说，私营企业只是在正式的金融制度下存在着融资困难问题。这和我们深入企业与私营企业主实地访谈所得出的结论是一致的。②私营企业的资金来源渠道很广，但企业主要是依靠自有资金进行周转。同时，民间借贷在调剂企业资金余缺方面也发挥着重要作用。银行贷款是私营企业资金来源的一个部分，但是所占比例很小。③民间借贷作为非正式的金融制度，手续简便，很少需要订立正式的借款合同，借款人不需要提交有关企业经营方面的信息，契约监督的力量主要靠道德传统、人情面子和个人的信誉。此时，私营企业的融资成本仅包括利息成本。如果借款数额不大，借款人和贷款人关系又较好，那么利息成本是非常低廉的。因此，如果私营企业资金需求不大，通过民间借贷融通资金就是私营企业最优融资决策。④私营企业不愿按银行的要求提供所有信息，才是私营企业在正式的金融制度下融资困难的主要原因。银行等金融机构在不能获得足够的信息以评估资金风险时，就会要求企业提供抵押和担保以使潜在的风险物化固定。公司规模实力必须通过资料才能确切验证，它本身就是一种不对称信息。以往的信用关系虽然有助于减少双方的信息不对称，但在我们的实地访谈中，银行则表示更看重抵押品和符合条件的担保人，而抵押品的度量和担保人的信息又成为新的非对称信息需要予以解决。

三、两大资本市场要求企业的信息披露

（一）商业银行要求企业的信息披露

在正式的金融制度下，企业向金融机构申请贷款是要按照一定程序和规定进行的。首先，企业必须在一家金融机构开立账户，和金融机构建立起资金往来关系；然后，企业向金融机构正式提交贷款申请报告，说明申请事由、借款资金用途、借款期限、还款资金来源等等；金融机构将对企业的资信情况进行详细的调查，要求企业提供所有的与借款内容相关的资料，经查实后进行风险等级评估；最后，银企双方签订正式的借款合同，银行将贷款划到企业早先开立的账户上。企业在这一过程中处于被动的地位，它必须配合银行的

调查工作，接受会计师事务所、审计事务所甚至律师事务所的清查与审核，并保证提交的各种资料具有法律效力。

以下是汕头市商业银行、中国银行、工商银行、建设银行、广东发展银行汕头分行等 5 家银行要求借款人报送的各种资料，经笔者实地收集并整理后，绘出一张企业必须披露的经营信息简表（见表 1）。

表 1 银行要求企业报送的信息

经营信息	报送资料	备注
会计信息	借款人近三年和最近一期会计报表。企业近三个月存款账户流水账复印件	担保人上年度和最近一期会计报表。上年度财务审计报告
投资决策	项目计划书	
市场营销	商品交易合同或协议书	
技术创新与产品开发	新产品计划书。新技术的先进性、适用性、可行性、新颖性证明	
生产情况	近三年主要产品产量、质量、工艺装备水平变化、产销率等	
利润分配		了解
资本运营	如有合并、分立、合伙、合资股权变动情况，提前 30 天通知银行	
产权分配	已年审的营业执照、法人代码证。股东情况与验资报告	抵押质押物所有权或使用权证明文件与公证书。担保人营业执照、企业代码
管理控制	法人代表身份证复印件、签字样式。公司章程、合资合同与成立批文。授权签字文件与被授权人身份证复印件及其签字样式。董事会文件	担保人法人代表身份证复印件。担保人内部管理结构。了解管理人员水平
机构设置	公司组织结构介绍材料	
人力资源		了解
税收缴纳	已年审的税务登记证	

可以看出，每一个方面的经营信息对私营企业来说都是很重要的商业机密。尤其是要求企业报送会计报表，最容易给私营企业的资产带来难以估量的风险。我国的私营企业在很长一段时期内都是没有会计报表或没有正式的会计报表的，全靠老板心里一本账核算成本与收益，后来在税务机关和工商管理部门的压力下，才开始建一些简单的流水账。由于

经营手段的灵活变通性很强，一些收入和支出是不能在账面上反映的，因此有的私营企业甚至使用两套账，对付政府部门的是一套账，自己进行会计核算的则是另外一套账。同时，我们在实际调研中发现，绝大多数私营企业的财务部门都是由家族成员控制，外人一般是无法进入的。在企业经营管理不规范的情况下，这种全部垄断会计信息的做法是有利于私营企业的积累与发展壮大的。不过，这些企业编制的会计报表是经不起会计师事务所的审计的，而且，一旦被审计，企业就会有许多问题暴露出来，甚至有可能导致企业的破产和倒闭。因此，私营企业宁愿损失巨大的投资收益，也不愿冒泄露会计信息的风险。此外，由于中国一贯的政治体制和意识形态的主导思想，私有财产一直没有得到国家根本大法《宪法》的承认，财不外露也不失是预防万一的好方法。据笔者的访谈了解，私营企业贷款一般都需要足额抵押或有银行信赖的担保人，而财产抵押和担保都需要对企业财产进行彻底的清点和价值评估，因此一些私营企业的家底就会暴露出来，企业中一些来路不明的财产就难以自圆其说。另一方面，就社会环境来看，中国没有在民众中培养起尊重产权的观念，社会信任资源极度稀缺，要素市场和产品市场无序竞争，企业法制意识淡薄，假冒伪劣现象严重，机会主义倾向明显，企业的投资决策、技术创新、市场营销、人事变动、资本运营、管理控制等方面的信息一旦被竞争对手掌握，就有可能被仿冒或利用，企业也会遭到激烈的攻击与报复，而制假者和恶意竞争者却不能遭到相应的惩罚。对私营企业而言，将这些信息全部垄断是有利的，除非披露这些信息带来的收益能够大于潜在的风险成本。

笔者除了收集5家银行的资料外，还对汕头市商业银行某支行的詹行长、广东发展银行汕头分行信贷部经理肖女士、中国银行的胡先生进行了实地访谈。据这三位业内人士介绍，在具体的操作中，银行不仅要求企业提供有关企业经营方面的各种资料和信息，以供信贷风险评估软件或模型评定风险等级，而且还要到企业中实际进行观察，直观地了解企业的规模、职工人数、员工士气、工艺设备、企业精神、机构设置、经营状况、主营收入等等各方面的情况，尤其要与企业主个人进行面对面的交流，以判断他（她）的知识水平、人品素质、社会经验、管理能力、经营思想、价值观念、抱负和追求，给企业领导人打出印象分。这些有关个人的信息虽然是表外信息，但是非常重要，银行在做出信贷决策时通常会参考。这表明，即使是在正式的金融制度下，资本市场也还需要解决有关借款个人方面的信息不对称问题，只是其重要性已经有所降低。

（二）上市公司的信息披露

我国的国家级资本市场主要是上海证券交易所和深圳证券交易所。一直“千呼万唤不出来”的创业板市场可以看作是“准国家级”的资本市场，其地位大致相当于美国交易所（AMEX）。对这两个资本市场的监管，主要由国务院中国证券监督管理委员会（以下简称“中国证监会”）负责。由于上市公司的资本社会化范围广，资金筹集的后果影响深远，因此中国证监会依据《中华人民共和国证券法》的有关条款，对拟上市公司的申请程序要进行严格核准，并详细规定了上市公司辅导报告、招股说明、中期报告、年度报告和发生重大事件时的临时公告等各个阶段信息披露的格式和内容（见表2），明确要求上市公司必须满足“持续信息公开”原则（《中华人民共和国证券法》第三节）。

表2 上市公司各阶段信息披露

报告期 公告内容	公开募股	中期报告	年度报告	备注
会计信息	发行人最近三年并当年中期经审计的会计报表主要数据,并提供详细说明	出具注册会计师明确表达意见的会计报表。采用数据列表方式提供报告期末和上年末主要财务指标,并提供详细说明	同中期报告。但要标明会计政策或会计估计的变化或合并范围	1. 所有信息必须按照规定的标准格式和内容披露。并有相应的披露时间。 2. 上市公司在公开募股以前,应经1年期的辅导时间。其间每2个月应报送一次辅导报告,内容包括发行人基本情况、财务报表及其详细说明、股权结构等。 3. 凡遇以下重大事件发生,均需发布临时公告: (1) 公司经营方针和经营范围发生重大变化; (2) 公司的重大投资行为和重大财产购置; (3) 公司订立重要合同,而该合同可能对公司的资产、负债、权益和经营成果产生重要影响; (4) 公司发生重大债务和未能清偿到期重大债务的违约情况; (5) 公司发生重大亏损或遭受超过净资产10%以上的重大损失; (6) 公司生产经营的外部条件发生重大变化;
投资决策	正在进行或计划进行的投资项目,包括对设备、厂房、土地及研究开发项目的投资。有全体董事签字的募集资金运用的可行性分析	应介绍募集投资项目的投入情况、进度以及收益情况,并解释未达进度的原因	同中期报告	
市场营销	主要市场及其市场占有率情况、销售额、销售方式(包括海外市场)。市场发展计划。关联交易的说明	主要业务范围及其经营情况。分行业介绍。其他业务活动介绍。详细介绍重大关联交易	股份公司是否独立拥有采购和销售系统。详细介绍主营业务和其他业务情况。详细说明关联交易	
产品开发与技术创新	新技术、新项目、新产品开发的有关情况			
生产情况	产品品种、生产能力、原料供应、自然资源损耗情况。生产经营发展战略	生产经营环境变化对生产经营成果有严重影响的须明确说明	主要产品、主营业务收入情况。生产能力、技术水平变化	
利润分配	发行人股利分配的各项政策	公司中期拟订的利润分配预案、公积金转增股本预案、公司上年度各项利润分配方案的执行情况	制定利润分配的详细方案	

续上表

报告期 公告内容	公开募股	中期报告	年度报告	备注
资本运营	详细说明过去三年的重大改组、变更、并购、招配股的情况	报告期内发生的收购、兼并、资产重组的情况。重大关联交易。报告期内发行新股、送配股情况	已发行股票债券情况。企业购并、重组等详细情况。托管、承包、租赁情况	(7) 公司的董事长、1/3 以上的董事或经理发生变动； (8) 持有公司 5% 以上股份的股东，其持有股份情况发生较大变化； (9) 公司减资、合并、分立、解散及申请破产的决定； (10) 涉及公司的重大诉讼，法院依法撤销股东大会、董事会决议； (11) 法律、行政法规规定的其他事项。
产权分配	发行股票前后股权变化情况。重要股东和关联企业情况	持股超过 5% 以上的股东详细情况	持有公司股份最多的前 10 名股东名单和持股数额	
管理控制	以方框图披露发行人内部管理结构。关联企业以及发行人对其他企业的持股情况	上市公司与控股股东在人员、资产、财务上的“三分开”情况。提交股东大会审议的重要事项	重点说明“三分开”情况。董事、监事、经理及有关高级经理人员简介及持股情况	
机构设置	以方框图表示发行人组织结构。公司章程修改情况	机构重大改变情况	机构重大改变情况	
人力资源	董事会构成、监事会构成。人员扩充计划	聘任、解聘会计师事务所情况。重大的人事变动	重大的人事变动	
税收缴纳				

资料来源：根据《证券法》和《中国证券监督管理委员会公告》中的有关规定整理。

如果不存在违规操作，如此透明的信息公开实际上已经将企业置于社会监督和政府监督的双重监督之下，企业的控制权和管理权将不再由企业家个人或家族垄断。这种情况对于私营企业来说，将是一个极大的挑战。首先，持续信息披露要求企业经营管理实现规范化、科学化，这将对企业主个人提出重大挑战。私营企业主文化水平普遍较低（全国私企 1995 年抽样数据），一般而言，他们对现代企业制度的理解并不是很透彻，学习先进的管理思想和管理技术的能力也有限。其次，经营管理规范化、科学化将对私营企业的传统管理方法和经营哲学提出挑战。机会主义倾向将被纠正，短视行为将被制止，利益相关者的意见将被尊重。但这些要求和做法，私营企业在短期内是不可能达到和实现的，私营企业需要一段时间进行转型，并要有社会环境的改造予以配合才能消化历史遗留的问题。再次，企业上市意味着公司有可能会被别的公司购并或接管，而一旦被购并或接管，私营企业主将丧失企业的名义所有权（尽管财产所有权有可能会增加），这对视企业为自家财产的中国人来说，是不可饶恕的失败。因此，即使将上市的机会向所有符合条件的私营企业

全部开放，也不会有很多的私营企业立即提出上市申请，除了那些改制比较成功的私营企业。

通过比较我国地区级资本市场和国家级资本市场的信息披露规定，我们可以看出：
①国家级资本市场要求的信息披露比地区级资本市场的要求更具体更详细。尤其是公开发行证券公司的信息披露，中国证监会的规定更为严格。
②国家级资本市场的要求是持续信息公开，并保证所有股民都可以于公告发布之后在特定地点随时查看，企业经营信息真正变成了公共品。但是企业向银行等金融机构提供的资料并不能让所有人随意查看，而且企业也没有义务对银行持续提供信息，银行等金融机构对贷款使用的监督力度较小。
③资本市场的层级越高，非人格化融资的特点就越明显。在地区级的资本市场上，金融机构还要实地了解企业主个人的人品素质、能力气魄等方面的人格化信息，但是在国家级的资本市场上，证交所基本上剔除了人格化信息的影响，主要依靠企业提供的非人格化信息进行风险评估。这种主要依靠非人格化的信息进行的融资活动就称为非人格化融资。此外，我们还可以推论，在国际资本市场上，企业需要提供的信息应该会更好、更详尽，尽管我们没有直接的资料可以证明。

四、结论与建议

本文旨在检验一个有关私营企业融资决策的假说，即在不存在制度歧视和政策倾斜的前提下，如果不考虑资本市场资金拆借的利息成本和执行成本，那么影响私营企业融资决策的主要因素将是各层级资本市场要求的信息披露机制。信息披露机制目的在于解决融资双方之间存在的两个方面的信息不对称问题：有关借款人的个人信息与企业经营管理方面的信息。但是在现阶段，正式的金融制度要求私营企业公开或披露的有关企业经营管理方面的信息，将会给企业带来很大的经营风险。因此，私营企业的理性决策将是努力控制住企业的经营管理信息，在较低层级的资本市场上依靠人格化信息的传递而建立起来的私人信任进行人格化融资活动。

必须指出，以上的结论是建立在较严格的假设基础上的。在实践中，控制企业经营信息虽然可以减少资产风险，但它是与借款资金的收入相比较的，如果潜在收入大于因公开经营信息而带来的风险成本，那么私营企业将愿意向银行和证券市场提出融资申请。这时，低层级资本市场的收益 - 成本差小于高层级资本市场的收益 - 成本差。私营企业愿意通过银行进行非足额融资，是因为银行在具体操作中规定不同借款规模的企业提供的信息是不同的。借款 10 万和借款 100 万，银行对企业的资信调查程度是不一样的。而且，有一定规模的私营企业一般都会谋求转型，向规范化管理方向迈进，披露一部分信息并不会给企业资产带来很大风险。所以，私营企业的资金来源中才有一部分银行等金融机构的信贷支持，而不是全部通过民间借贷筹措资金。

此外，现实中各资本市场的融资成本差异较大，尤其是公开募股上市，各种费用不是一般规模企业所能够承担的。这些需要从会计账户中直接扣除的成本支出对私营企业融资渠道的选择影响也很大。

综合考虑本文的分析，本文对如何解决一度成为热点话题的私营企业“融资难”问题

提出以下建议：

第一，私营企业“融资难”问题，主要是指难以通过正式的金融制度进行资金融通活动，而这又是因为资本市场所规定的信息披露机制会给企业带来经营风险。所以，对私营企业发展进行金融支持的重点不在于出台各项对私营企业的政策扶持，如建立私营企业融资担保制度等，关键在于如何改进和完善信贷风险评估的指标体系。考虑到私营企业的实际情况，适当调整一些指标的测度方法，以降低私营企业披露经营信息可能带来的风险成本。金融机构应该在信贷行为发生前就对一些有发展潜力的私营企业进行观察和评估，利用长期累积的信息正确评估企业的信用等级；而实际的资金拆借发生时就可以根据这些累积的数据分析做出信贷决策，以避免对一些敏感信息采取正面获取的方式。

第二，民间借贷市场是现阶段私营企业解决资金问题的主要来源之一，这是金融监管部门必须承认的事实。这一层级的资本市场规模庞大，涉及面极其广泛。因此，国家金融体系完善的重点不是排斥民间借贷在金融活动中的影响，而是因势利导，想方设法将这些游离在经济体系外的资金纳入到金融体系中，如建立地区性的资本市场、成立私募基金、允许民间银行开业等，为资金短缺企业开辟新的融资渠道。考虑到历史和其他方面的因素，在引导过程中应区别对待存量资金和流量资金。对于存量资金，有关部门无须过多地关心其来源，只要它能够按正常程序浮出水面即可；而对于流量资金，则应该严格审查其来源的合法性，以免为“洗黑钱”提供便利条件。

第三，私营企业担心企业经营信息的披露会给企业带来经营风险，这一方面是因为历史和社会环境的原因，另一方面也是由于一些私营企业主自身的原因，如机会主义倾向和短视行为。因此，私营企业必须树立可持续发展的经营理念，规范自己的经营行为，引进科学的管理方法，建立有效的内部治理机制，逐步消化历史遗留问题，按现代企业管理方法整理和报送各种资料，将信息披露可能带来的经营风险降到最低，这样才有可能从低层级资本市场的人格化融资走向高层级资本市场的非人格化融资。当然，政府应该为私营企业规范经营行为创造一个宽松的社会环境。

参考文献

- [1] 王宣喻，储小平：《资本市场的层级结构与信息不对称下的私营企业融资决策》，《上海经济研究》，2002年第4期。
- [2] 林毅夫，李永军：《中小金融机构发展与中小企业融资》，《经济研究》，2001年第1期。
- [3] 汪丁丁：《回顾“金融革命”》，《经济研究》，1997年第12期
- [4] 张维迎：《企业理论与国有企业改革》，北京大学出版社1999年版。
- [5] 张杰：《民营经济的金融困境与融资次序》，《经济研究》，2000年第4期。
- [6] 周其仁，张居衍，戴勇：《设立并逐步开放全国性的为中小企业融资服务的资本市场》，《经济学消息报》，1998年1月2日。

中国股票市场波动率的高频估计与特性分析^{*}

黄后川 陈浪南

一、引言与先前该领域研究述评

近 20 年来，对波动率模型的研究已成为金融经济学领域研究的重要内容之一。自 Engle 于 1982 年提出 ARCH 模型以来，经济学界已经发表了数千篇关于条件异方差或波动率的论文。特别是最近十年，Andersen 等学者提出用高频分时数据估计波动率的方法，这种方法可以得到比较准确的波动率估计值，称为“已实现波动率”（Realized Volatility）。以此为基础，众多学者在波动率的特性和预测两方面进行了更深入的研究，大大拓展了这个研究领域。

Andersen, Bollerslev, Diebold, Ebens (1999, 2001a, 2001b) 等金融经济学家对这种高频估计方法以及“已实现波动率”的特性与预测进行了一系列研究，他们得出了如下几个主要结论（计算的波动率都是日波动率）：

(1) 如果价格遵循普通的扩散过程，用此方法计算的已实现波动率，是无偏的。而且，当高频数据的时间间隔趋近于 0 时，已实现波动率的测量误差也趋于 0。因此可以把已实现波动率当作一个观测值，它没有经典算法所带来的时间滞后。

(2) 通过对外汇市场和道·琼斯工业股票的实证研究，发现：① 股票市场中，正收益对后续波动率的影响不如负收益明显，即波动率具有杠杆效应。② 已实现波动率的对数具有明显的长期记忆特性。③ 虽然已实现波动率明显向右倾斜，但已实现波动率的对数呈现正态分布。④ 虽然原始的收益率数据有明显的高峰和大尾巴，但收益率除已实现波动率呈现正态分布。⑤ 股票市场的波动率与相关度呈相同方向运动，降低了资产组合分散化在高波动率时的作用。

(3) 依据 (2) 中的结论，用体现长期记忆的分数综合 - 移动平均自回归（Auto Regression Fractional Integrated Moving Average, ARFIMA）方法可以得到更好的波动率预测。使用正态 - 对数正态混合分布可以得到很好的概率密度和分位数估计（例如 VaR）。

已实现波动率的一个重要用途是作为对以前各类模型进行评价的基准。它的另一个更重要的用途是用于检验波动率的各种特性，并对未来波动率进行预测，因为已实现波动率可以直接当作波动率的一种观测值，因此可以采用一般的时间序列方法，无须像 ARCH 模型一样通过模拟收益率序列得到内含的波动序列。

^{*} 本文是国家自然科学基金课题（79800010、70042005）、上交所 2002 年联合研究计划课题、教育部社科“十五”课题（01jb790026）及 2002 年厦门大学校级课题成果之一。原载：《经济研究》，2003 年第 2 期。

许多经济学家,如 Engle, Ng (1993) 等,都通过实证研究指出了信息(一般用前期收益度量)与波动率之间的不对称性,即好信息与坏信息对未来波动率的影响程度不同,通常,负收益率相对于正收益率而言对未来波动率的影响更大些。对该现象有两种解释:第一种称为“杠杆效应”,即一个大的负收益增加了财务和操作杠杆,反过来提升了股票收益的波动率(Christie, 1982);第二种解释认为,市场风险贴水是波动率的增函数,大的负收益由于存在一个波动率反馈效应,因而对波动率的影响比正收益更大些。

波动率的一个重要特性是它的持续性,即市场波动一般会持续一段时间,随着时间的推移而慢慢消失。通常,经济学家们把时间序列分为确定性的趋势和移动平均自回归随机过程两个部分。判别确定性趋势的一种常用方法是测试时间序列是否存在一个单位根。Granger (1966) 发现大多宏观或金融变量都存在着自相关系数慢速衰减现象,随机游走或 ARIMA 只是具备这种现象的特例之一。

本文旨在应用高频数据估计中国股票市场的已实现波动率。在计算各种频率的已实现波动率的基础上,本文构造了一种较为精确的估计波动率的方法,可以更好地平衡测量误差与微观结构误差,并进一步考察中国股市波动率的不对称和长期记忆特性。

二、研究方法

1. 波动率的高频估计

使用高频交易数据(如 5 分钟、1 天)来计算低频波动率(如 1 天、1 月)的理论模型如下:

假定 n 维对数价格向量 p_t 遵从多维连续的正态发散过程:

$$dp_t = \mu_t dt + \Omega_t dW_t$$

其中: Ω_t 为 $n \times n$ 维的强平稳的正定的扩散矩阵; W_t 是标准的布朗运动。则收益 $r_{t+h,h} = \ln p_{t+h} - \ln p_t$ 服从以下条件概率分布:

$$r_{t+h,h} / \sigma \{ \mu_{t+\tau}, \Omega_{t+\tau} \}_{\tau=0}^h \sim N \left(\int_0^h \mu_{t+\tau} d\tau, \int_0^h \Omega_{t+\tau} d\tau \right)$$

其中: $\int_0^h \Omega_{t+\tau} d\tau$ 为综合的方差。根据二次变换理论,当 $\Delta \rightarrow 0$ 时,有:

$$\sum_{j=1}^N \dots \dots (h/\Delta) r_{t+j,\Delta} \cdot r'_{t+j,\Delta} - \int_0^h \Omega_{t+\tau} d\tau \rightarrow 0$$

由此构造估计量:

$$\sigma^2 = \sum_{j=1}^N \dots \dots (h/\Delta) r_{t+j,\Delta} \cdot r'_{t+j,\Delta} \quad (1)$$

它是无偏的,而且当 $\Delta \rightarrow 0$ 时,测量误差也趋于 0。

2. 时间频率与波动率估计

假定某一日的资产收益率为 r , 如果把此日内的交易期间分成 N 份,可以得到包含 N 个期间段的收益率序列 $\{r_1, r_2, \dots, r_n\}$, 那么对应数据频率 N 的波动率估计值如下: $\sigma_N^2 = \sum_{i=1}^N r_i^2$ 。为简便起见,我们比较 $N=1$ 和 $N=n>1$ 时波动率的估计值:

$$\sigma_1^2 = r^2 = \left(\sum_{i=1}^n r_i \right)^2 = \sum_{i=1}^n r_i^2 + \sum_{i=1}^n \sum_{j=1, j \neq i}^n r_i r_j = \sum_{i=1}^n r_i^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n r_i r_j$$

$$= \sigma_n^2 + 2 \sum_{i=1}^n \sum_{j=i+1}^n r_i r_j = \sigma_n^2 + 2 \sum_{i=1}^n \sum_{j=1}^n r_i r_{i+j} \quad (2)$$

即：低频波动率估计值 = 高频波动率估计值 + 2 * 高频收益率在低频期间内的自协方差之和。

另外， $\sigma^2 = \sum_{i=1}^n r_i^2$ 估计误差可由下式表示：

$$E\left(\sum_{i=1}^n r_i^2 - \sigma^2\right)^2 = \frac{\sigma^4}{n} (K_1 - 1) \left(1 + 2 \sum_{i=1}^{n-1} \frac{n-i}{n} \rho_i\right) \quad (3)$$

其中： K_1 是序列 r_i 的峰度； ρ_i 是序列 r_i^2 的第 i 个自相关系数（Karatzas, Shreve, 1988）。根据 (3) 式，容易看出，估计误差随着数据频率的增高而降低，如果要得到准确的波动率估计值，最好选取最高的数据频率和最小的时间间隔。

然而，在个股价格的高频时间序列中，价格上下跳跃意味着高频收益率序列的自相关系数是个负数，根据 (2) 式，因为 $E(r_i r_{i+j}) < 0$ ，所以高频估计值会比低频估计值来得大。由 (3) 式，估计误差不仅与频率 n 有关，也与序列 r_i^2 的自相关系数 ρ_i 有关，选取极高频数据来估计波动率，往往会由于 ρ_i 的显著增大而产生较大误差（微观结构磨擦误差），因此，最佳的估计频率应当在测量误差与微观结构误差间进行权衡。

3. 已实现波动率的不对称特性

我们研究上证 A、B 指数“已实现方差的对数”和收益率之间的不对称特性，所采用的模型是以下的折线方程：

$$y = \omega - k_1(x - x_0)I_{x-x_0}^- + k_2(x - x_0)I_{x-x_0}^+ \quad (4)$$

其中： y 为波动率； x 为上期收益率； $I_{x-x_0}^-$ 在 $x - x_0 < 0$ 时取 1，其他情况时取 0； $I_{x-x_0}^+$ 在 $x - x_0 \geq 0$ 时取 1，其他情况时取 0。 ω ， x_0 ， k_1 ， k_2 为待估参数，采用非线性最小平方估计进行模拟。当折线的两个斜率不等，负收益对应的斜率 k_1 显著大于正收益对应的斜率 k_2 时，即表明已实现波动率存在明显的不对称性。

4. 已实现波动率的长期记忆特性

波动率的一个重要特性是它的持续性，即市场波动一般会持续一段时间，随着时间的推移而慢慢消失。我们可以通过分数综合自回归移动平均模型（Fractional Integrated Autoregressing Moving Average, ARFIMA）更好地模拟具备这种现象的时间序列。

长期记忆或是短期记忆过程可以用统一的 ARFIMA 模型刻划，ARFIMA (p, d, q) 模型的形式如下：

$$\phi(L)(1-L)^d y_t = \theta(L)\epsilon_t \quad (5)$$

其中： y_t 是可观察的序列（如通货膨胀率）； ϵ_t 为一平稳的噪声序列； $\phi(L)$ ， $\theta(L)$ 分别为 L 的 p 阶、 q 阶多项式； d 是一个分数，表示对序列进行某种形式的差分。 $d=0$ 时，ARFIMA (p, d, q) 模型退化为 ARMA (p, q)，当 $d=1$ 时，就是 ARIMA 模型。

有许多方法可以检测和估计长期记忆过程的存在与否以及代表自相似程度的参数，我们使用修正的周期图方法估计分数综合参数。周期图方法又称为 GPH 估计方法，是 Geweke, Porter-Hudak (1983) 提出的一种非参数周期图回归方法。

首先，需要计算时间序列在各频率下的周期图： $I(\lambda) = |\sum_{j=1}^n X_j e^{j\lambda}|^2 / (2\pi N)$ ，其中

i 为虚数单位, $\lambda = 2\pi k/N$ 为傅立叶频率。因为 $I(\lambda)$ 是频谱密度的一个估计量, 所以长期记忆过程的 $I(\lambda)$ 的对数应当与 $|\lambda|^{-2d}$ 的对数成正比, 构造以下线性方程:

$$\ln(I(\lambda_k)) = \beta_0 + \beta_1 \ln(\sin^2(\lambda_k/2)) + \eta$$

估计出 β_1 后, 根据 $d = -\beta_1$ 可以计算出分数综合的参数。若分数综合参数的估计值在统计上显著, 则说明波动率序列具有长期记忆性。

三、实证研究与分析

1. 样本数据

本文使用的原始数据为上证 A、B 股指数在 2000 年 1 月 4 日至 2002 年 3 月 25 日期间内的每笔交易数据。其中上证 A 股指数数据共计 252510 条, 包含 519 个交易日, 平均每日 485 条; 上证 B 股指数数据共计 199106 条, 包含 513 个交易日, 平均每日 388 条。期间内有异常的原始数据如表 1 所示。

表 1 有异常的高频(每笔交易)数据列表

日期	数据	异常原因
2000 - 03 - 27 至 2000 - 03 - 31	上证 A 指, 上证 B 指	缺失数据
2001 - 02 - 20 至 2001 - 02 - 23	上证 B 指	暂停交易
2001 - 02 - 26 至 2001 - 02 - 27	上证 B 指	暂停交易
2001 - 05 - 08	上证 A 指, 上证 B 指	缺失数据
2001 - 11 - 15	上证 A 指, 上证 B 指	缺失数据

由于数据缺失, 少了 7 个交易日的每笔交易数据, 另外, 因为上证 B 指在 2000 年 2 月 20 日至 2000 年 2 月 27 日之间暂停交易, 所以上证 B 指比上证 A 指要少 6 个交易日的数据。

鉴于数据缺失的交易日较少, 对于这些交易日, 我们把它们直接从时间序列中扣除。另外, 为了研究上证 A 指、上证 B 指之间的协方差和相关度, 我们从上证 A 指中除去对应的 6 个交易日, 再计算协方差和相关度序列, 因此协方差和相关度序列的长度为 513。

2. 最优时间间隔与波动率估计

由于现实当中的价格并不是严格服从正态发散过程的, 因此在估计已实现波动率时, 高频数据的时间间隔并不是越小越好, 用极高频数据估计波动率往往会因为微观磨擦而产生较大误差。Torben G. Andersen 等人在对道·琼斯 30 只工业股票做实证分析时, 直接选取了 5 分钟数据, 而在其他文献中他们又使用“标记图”方法得出较佳时间间隔。然而, 用于估计“已实现波动率”的最佳频率是与市场微观特性相关的, 在不同市场、不同时期可能会有所不同。因此, 对所有时期的数据样本全部选取一个数据频率, 可能存在问题。

图 1 和图 2 分别是典型的个股和指数的分时走势图。在个股走势图中, 高频价格序列存在着明显的锯齿, 表明价格经常在短期内上下跳跃, 其主要原因就是市场微观结构导致

的交易磨擦；图 2 中所示的指数高频率序列图与个股走势图的情形相反，股指序列相当于多只股票价格序列的加权累计，经过加权累计，单只股票的价格跳跃对股指影响很小，因此股指序列中看不到单只股价序列中的跳跃现象，指数序列相对比较平滑。然而，两种时间序列都与严格的正态发散过程有一定差距，因而在用于估计波动率时会产生估计误差。

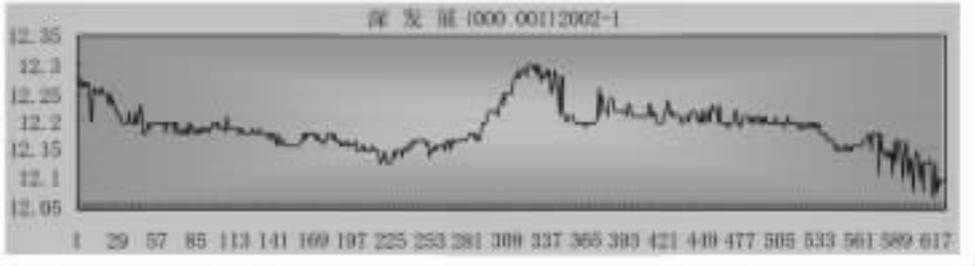


图 1 深发展个股分时走势图 (2002 - 01 - 04)

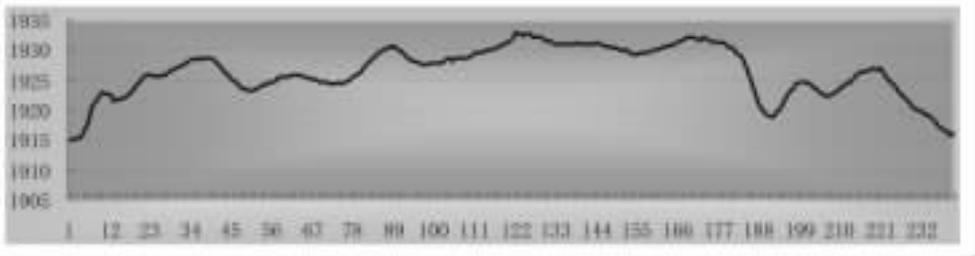


图 2 上证 A 股指数走势图 (2000 - 04 - 03)

同样，平滑的指数的高频时间序列也不适于使用极高频的数据来估计波动率，如果价格曲线在较长时间段内保持向上或向下，就相当于时间段内高频收益率序列的自协方差是个正数，根据公式 (2)，采用极高频的数据会显著降低波动率的估计值，根据公式 (3)，使用极高频的数据也会由于 ρ_i 的显著增大而产生了较大的误差，因此也需要寻找恰当的数据频率，以平衡常规的测量误差与微观结构误差。

如图 3 所示，采用时间间隔为 1 分钟的高频数据时，波动率估计值最低，随着时间间隔的增大（数据频率降低），波动率估计值逐渐上升，但同时常规测量误差也增加（估计值上下明显波动），就图中曲线而言，比较适宜的时间间隔是 10~17 分钟。

为尽可能减少微观结构误差和常规测量误差，我们首先对直接高频估计曲线计算两次移动平均值： $\sigma_t^2 = \left(\sum_{i=1}^5 \sigma_{t-3+i}^2 \right) / 5$ ， $\sigma_t'^2 = \left(\sum_{i=1}^5 \sigma_{t-3+i}'^2 \right) / 5$ ，得到移动平均估计曲线。显然，经过移动平均处理后，估计值的上下波动明显减小，常规测量误差显著减小。因为时间间隔较小时（数据频率较高），波动率估计值存在明显的负偏差，我们截去移动平均估计曲线的上升部分，取曲线向下走的第一个结点作为最终估计值，例如图 3 中，最佳时间间隔为 16 分钟，估计值为 0.00003373。类似的方法也可应用到单只股票的高频数据上。

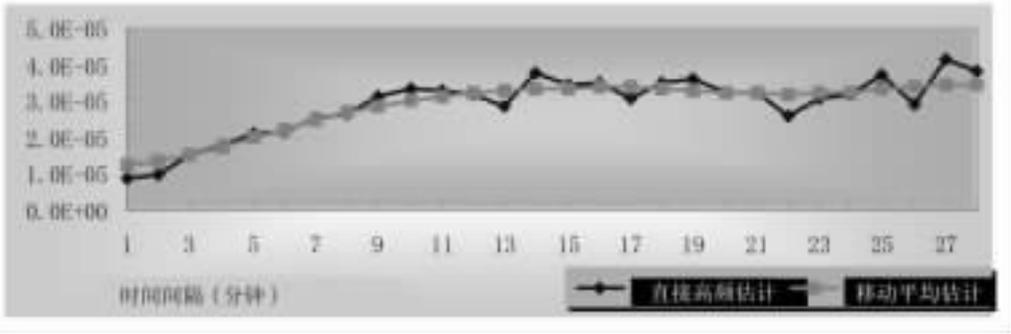


图3 上证A股指数的波动率高频估计 (2001 - 01 - 02)

如图4所示，采用时间间隔为1分钟的高频数据时，波动率估计值最高，随着时间间隔的增大（数据频率降低），波动率估计值逐渐下降，但同时常规测量误差也增加（估计值上下明显波动），就图中曲线而言，比较适宜的时间间隔是10~15分钟。同样，首先对直接高频估计曲线计算移动平均值，得到移动平均估计曲线，显然，经过移动平均处理后，常规测量误差得以大大降低。因为时间间隔较小时（数据频率较高），波动率估计值存在明显的正偏差，我们截去移动平均估计曲线的持续下降部分，取曲线向上走的第一个结点作为最终估计值，例如图4中，最佳时间间隔为15分钟，估计值为0.00002686。

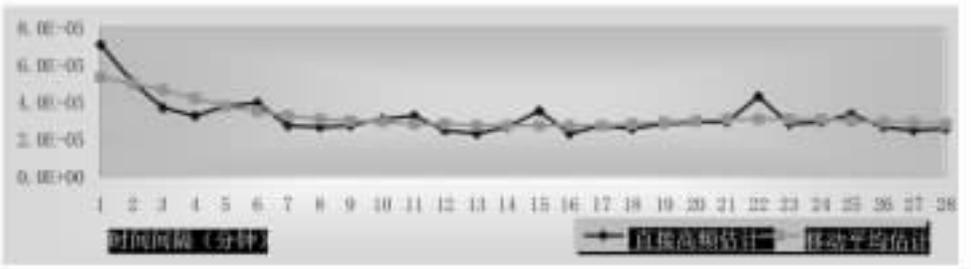


图4 单只股票的波动率高频估计 (深发展, 000001, 2002 - 01 - 04)

根据前面的波动率高频估计方法，我们计算得到的上证A指方差、上证B指方差序列。

如图5、图6所示，上证A指方差（波动率）在2000年1~3月经历了一次高峰，而后在2000年5月至2001年7月相对较低，2001年7月后又进入一个波动较大的时期。上证B指方差（波动率）在2001年2月之前较低，此后一段时期波动幅度相对较大。从图中可看出，两个方差（波动率）都有明显的前后相关性或持续性，高波动率和低波动率都集中分布在某些特定的时期，而且，上证B指波动率要比上证A指大一些。

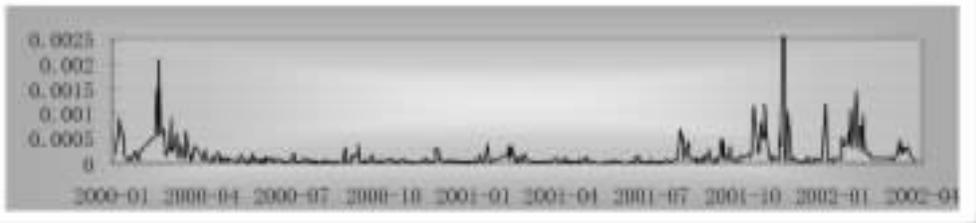


图5 上证 A 股指数的高频估计方差序列

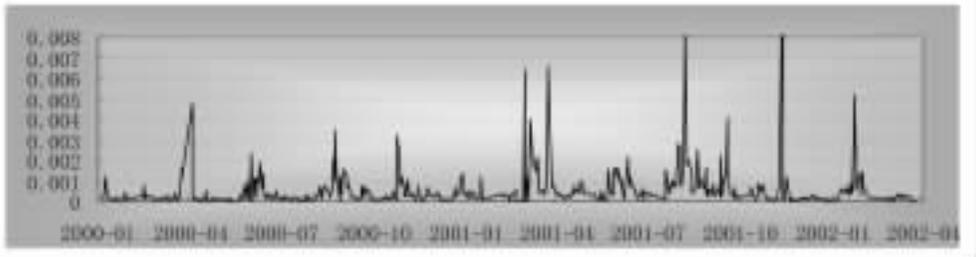


图6 上证 B 股指数的高频估计方差序列

如图7所示,上证A、B指数相关度在2000年1月至2001年2月之间较低,一般都在0.6以下,均值为0.3477;2001年3月之后相关度显著增高,一般都在0.6之上,均值为0.7809。

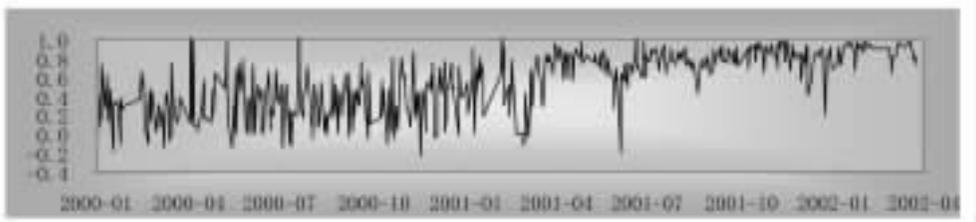


图7 上证 A、B 股指数的高频估计相关度序列

3. 已实现波动率的不对称特性

我们研究上证A、B指数“已实现方差的对数”和收益率之间的不对称特性,所采用的模型是(4)式的折线方程:

$$y = \omega - k_1(x - x_0)I_{x-x_0}^- + k_2(x - x_0)I_{x-x_0}^+ \quad (4)$$

其中: y 为波动率; x 为上期收益率; $I_{x-x_0}^-$ 在 $x - x_0 < 0$ 时取 1, 其他情况时取 0; $I_{x-x_0}^+$ 在 $x - x_0 \geq 0$ 时取 1, 其他情况时取 0。采用非线性最小平方估计进行拟合。应用上证 A 股、B 股的对数方差、收益率及相关度数据, 对该模型拟合的结果列于表 2 中。

表2 上证A指、B指对数方差与收益率，相关度与上期收益率之和的折线模拟

	样本数	ω	x_0	k_1	k_2	R^2	F
A指与上期收益率	518	- 10.218	0.001883	66.255	57.917	0.3360	87.052
B指与上期收益率	512	- 8.7237	- 0.00565	35.853	15.44	0.1669	34.0467
相关度与上期收益率之和	512	0.52006	0.0175	1.6286	0.70276	0.0192	3.3199

如表2所示，折线方程(4)式能够较好地模拟上证A、B指数对数方差与收益率的相关关系，对上证A指模拟的 R^2 达到0.3以上，对上证B指模拟的 R^2 为0.167，均较显著。其次，折线的两个斜率不等，负收益对应的斜率 k_1 显著大于正收益对应的斜率 k_2 ，表明上证A指、B指均有显著的不对称性。再次，上证A指、B指的折线端点 x_0 接近于0，不对称性只在于正负收益的斜率不同。最后，相关度确实有一定的不对称性，然而，拟合的 R^2 仅有0.0192，说明不对称性不显著。

4. 已实现波动率的长期记忆特性

波动率的一个重要特性是它的持续性，即市场波动一般会持续一段时间，随着时间的推移而慢慢消失。我们可以通过ARFIMA模型拟合具备这种现象的时间序列。

我们应用修正的周期图方法估计ARFIMA模型的分数综合参数。首先，需要计算时间序列在各频率下的周期图： $I(\lambda) = |\sum_{j=1}^n X_j e^{j\lambda}| / (2\pi N)$ ，其中 i 为虚数单位， $\lambda = 2\pi k/N$ 为傅立叶频率。所以长期记忆过程的 $I(\lambda)$ 的对数应当与 $|\lambda|^{-2d}$ 的对数成正比，即有：

$$\ln(I(\lambda_k)) = \beta_0 + \beta_1 \ln(\sin^2(\lambda_k/2)) + \eta$$

估计出 β_1 后，根据 $d = -\beta_1$ 可以计算出分数综合的参数。若分数综合参数的估计值在统计上显著，则说明波动率序列具有长期记忆性。

通常，用周期法计算得到的 $\ln(I(\lambda_k))$ 序列波动很大，我们对此序列进行移动平均(MA5)后，再用线性回归得到分数综合参数。如表3所示，线性估计的 R^2 都比较大，三个分数综合参数估计值都在0~0.5之间，表明三个时间序列均有显著的长期记忆特征。上证B指波动率序列的分数综合参数为0.28，低于上证A指波动率以及A、B指相关度的参数估计，说明上证B指波动率序列的持续性相对较低。

表3 上证A、B指波动率、相关度的长期记忆检验表

	样本数	Beta	R^2	d
上证A指波动率	22	- 0.36544	0.9128	0.36544
上证B指波动率	22	- 0.28042	0.9487	0.28042
A、B指相关度	22	- 0.4453	0.94422	0.4453

四、结论

针对中国股票市场波动率,本文进行了有关高频估计、特性(分布、不对称、长期记忆)以及模拟与预测等等各方面的研究。

首先,我们发现股票指数与个股的高频交易数据中的微观磨擦影响正好相反,使用极高频的数据会大大增加个股的波动率估计值,相反却会大大降低指数的波动率估计值。因为测量误差与微观磨擦误差在各个时期有较大变化,所以直接选用5分钟或15分钟交易数据并不可取,凭主观进行判断的“标记图”方法也有一定偏差,本文中构造了一种较为精确的估计波动率的方法,得到了较为准确的上证A、B指数的波动率序列。

其次,基于已实现波动率,本文研究了中国股市波动率的不对称性和长期记忆特性。在波动率的不对称特性方面,上证A、B指均有明显对上期收益和本期收益的不对称性,折线回归可以很好地模拟这种不对称性;在波动率的长期记忆特性方面,上证A、B指波动率及其相关度都有明显的持续性和长期记忆特征,上证B指波动率持续性较弱,上证A指波动率和两个指数的相关度持续性较强些。

参考文献

- [1] Andersen T. G., Bollerslev T., Diebold F. X., and Labys P., The Distribution of Exchange Rate Volatility, *Journal of the American Statistical Association*, Working Paper 6961, 2001a.
- [2] Andersen T. G., Bollerslev T., Diebold F. X., and Labys P., Exchange Rate Returns Standardized by Realized Volatility are (Nearly) Gaussian, *Multinational Finance Journal*, 2001b.
- [3] Andersen T. G., Bollerslev T., Diebold F. X. and Ebens H., The Distribution of Realized Stock Return Volatility, *Journal of Financial Economics*, 2001.
- [4] Andersen T. G., Bollerslev T., Diebold F. X., and Labys P., Modeling and Forecasting Realized Volatility, 2001.
- [5] Baillie R. T., Bollerslev T. and Mikkelsen H. O., Fractional Integrated Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 1996, Vol. 74, pp3 ~ 30.
- [6] Bollerslev T., Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics*, 1986, Vol. 31, pp307 ~ 327.
- [7] Bollerslev T. and Mikkelsen H., Modeling and Pricing Long-Memory in Stock Market Volatility, *Journal of Econometrics*, 1996, Vol. 73, pp151 ~ 184.
- [8] Ebens H., Realized Stock Volatility, The Johns Hopkins University, 1999.
- [9] Engle R. F. and T. Bollerslev, Modelling the Persistence of Conditional Variances, *Econometric Reviews*, 1998, Vol. 5, pp1 ~ 50.
- [10] Engle R. F. and V. K. Ng., Measuring and Testing the Impact of News on Volatility, *Journal of Finance* 1993, Vol. 48, pp1749 ~ 1778.
- [11] Geweke J. and Porter-Hudak S., The Estimation and Application of Long Memory Time Series Models, *Journal of Time Series Analysis*, 1983, Vol. 4, pp221 ~ 238.
- [12] Henry O., Modelling the asymmetry of stock market volatility, *Applied Financial Economics*, 1998, Vol. 8, pp145 ~ 153.

- [13] Hosking, J. R. M. , Fractional Differencing, *Biometrika*, 1994, Vol.68, pp165 ~ 176.
- [14] Sowell F. , Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics*, 1992, Vol.53, pp165 ~ 188.
- [15] 陆懋祖：《高等时间序列经济计量学》，上海人民出版社 1998 年版。

中国证券投资基金市场择时能力的非参数检验*

牛 鸿 詹俊义

一、引言

投资者投资于基金实际上就是委托投资专家即基金经理代替自己把资金投资于证券市场。基金经理是否真的能使投资者的回报更高呢？这可以从基金运作的业绩来对基金经理的能力进行判断。

一般地，我们把基金经理的能力分为两种：一种称之为选股能力（Selectivity），也就是基金经理能够预测某只股票将来价格走势的能力；另一种称之为市场择时能力（Market Timing），也就是基金经理能够预测整个股票市场价格走势的能力。假设基金经理都能够对其既定的预测做出正确的决策，则基金的业绩就是基金经理选股能力和择时能力或者两者组合的直接表现。因此，研究了基金的择时能力，也就部分地评价了基金的业绩。如果市场上升时，基金上升得更多（或更快）；市场下降时，基金下降得更少（或更慢），那么我们就说基金具有较好的择时能力。反之，则基金具有较差的择时能力。

对市场择时能力的研究文献，在国外已经是汗牛充栋，但就其方法而言，则主要是运用参数方法进行研究的，而对于非参数方法却涉及很少，且大多数都很难在实际中运用。本文运用 Jiang（2003）提出的一种用事后收益率即可进行非参数检验的方法对基金的择时能力进行研究，并同参数检验进行对比。

二、文献回顾和模型的建立

（一）参数检验

Treynor & Mazuy（1966）在基金业绩评价的发展史上首次对市场择时能力进行研究，他们在其论文中提出了一个检验的方法，用表达式表示就是：

$$r_{i,t+1} = \alpha_i + \beta_{i,t} r_{m,t+1} + \gamma_i [r_{m,t+1}]^2 + \epsilon_{i,t+1} \quad (1)$$

式中： $r_{i,t+1}$ 为第 i 只基金在 $t+1$ 时期的收益率； α_i 为常数项； $\beta_{i,t}$ 为基于 t 时期的信息而决定的 $t+1$ 时期的风险系数； $r_{m,t+1}$ 为 $t+1$ 时期市场收益率； $\epsilon_{i,t+1}$ 为随机误差项。所有的收益率都是相对于无风险资产收益率而言的，也即我们这里所说的收益率实际都是超额

* 原载：《管理世界》，2004 年第 10 期。

收益率，后文中如果没有特别说明，则都是这么表示的。若 γ_i 是正的且显著，则表示如果市场上升，基金将上升得更多，如果市场下降，基金将下降得更少，也就是基金具有很好的择时能力，因此 γ_i 是择时能力的度量指标。Admati 等（1986）论证了上面这个式子。他们假设基金经理在 t 时期观察到一个私人信号 y_t ，它等于下一期市场实际的收益率加上一个独立的扰动项：

$$y_t = r_{m,t+1} + \eta_t \quad (2)$$

如果基金经理有很好的择时能力，则与此相对应就是扰动项的方差是有限的，即 $\sigma_\eta^2 \ll 1$ 。若基金经理有不变的绝对风险厌恶 [即 $R'_A(W) = 0$ ，其中 W 表示资产价值] 偏好并想最大化期望效用，则投资组合的 β 值是信号的一个线性函数，根据上面两个假设，就可以推导得到式（1）。由此可见，TM 模型 [即式（1）] 的条件是相当严格的。

Henriksson & Merton（1981）在 Merton（1981）的基础上，定义了基金经理的择时能力：

$$\Delta = \Pr(r_{m,t+1} > 0 \mid r_{m,t+1} > 0) + \Pr(r_{m,t+1} < 0 \mid r_{m,t+1} < 0) - 1 \quad (3)$$

式中： $r_{m,t+1}$ 是经理在 t 时期对 $t+1$ 时期实际收益率 $r_{m,t+1}$ 的预测值。如果有较好的择时能力，则 Δ 会大于 0。他们只是分两种情况来讨论：如果 $r_{m,t+1} > 0$ 也即当预测到下一时期的市场收益率比无风险资产收益率高的时候，设定一个较高的 β 值；如果 $r_{m,t+1} < 0$ ，也即预测下一个时期比无风险资产收益率低的时候，设定一个较低的 β 值。根据式（3）的假设，可以得到：

$$r_{i,t,j} = \alpha_i + \beta_{i,t} r_{m,t+1} + \gamma_i \max(0, -r_{m,t+1}) + \varepsilon_{i,t+1} \quad (4)$$

式（4）中： $\max(0, -r_{m,t+1})$ 实际上就是一份看跌期权的价值，其履约价格即为无风险资产收益率。因此，基金经理实施市场择时策略的“收益”就是 γ_i 份这样的期权，如果 γ_i 是正显著的，则 $\gamma_i \max(0, -r_{m,t+1})$ 就是基金经理超常的择时能力的回报；如果是负的，则是损失。故 γ_i 就是衡量基金经理择时能力的指标。由此可见，HM 模型 [即式（4）] 只是考察了市场收益率大于或者小于无风险收益率这两种情况，因此其条件也是相当严格的。

上面两种模型是检验择时能力的经典的方法，当然还有其他一些参数检验方法，例如 Fama & French（1993）提出的三因子模型，对 TM 模型和 HM 模型都增加 SMB（小盘股收益率减去大盘股收益率之差）和 HML（高账面市值比股票收益率减去低账面市值比股票收益率之差）这两个解释变量。我们在这里就不再详细讨论，下面主要谈谈参数方法检验择时能力存在的一些缺点和不足。

首先，无论是 TM 模型还是 HM 模型，两者都是基于 CAPM（Capital Asset Pricing Model，资本资产定价模型）的。当市场达到均衡时，所有基金和市场都应该在同一条 SML（Securities Market Line，证券市场线）上，如果存在显著的 α_i 或者显著的 γ_i 则表明市场不能解释一切，这与 CAPM 是不相容的，因此，这两种参数方法在理论上都有一定的缺陷。其次，由上面对 TM 和 HM 模型的分析可知，两个模型所要求的条件相对来说是很严格的，条件越严格，则结论的真实性就越值得怀疑。最后，很多参数方法的研究表明，选股能力和市场择时能力之间存在很严重的负相关性。Henriksson（1984）对 1968 年 2 月至 1980 年 6 月的 116 只开放式基金进行研究，结果发现 59 只 α_i [为式（4）中的估计值，下

同]为正的基金中有 49 只 γ_i 为负, 而 57 只 α_i 为负的基金中有 34 只 γ_i 为正。可以看出, α_i 和 γ_i 之间存在较为严重的负相关, 它所表明的是选股能力强的基金经理其市场择时能力反而差, 而择时能力强的则其选股能力不理想! 这确实难以令人信服。沈维涛、黄兴李 (2001) 对 1995 年 5 月 10 日之前上市的 10 只封闭式基金在 1999 年 5 月 14 日至 2001 年 3 月 23 日的择时能力进行研究, 结果在 TM 模型中负相关的基金有 4 只, HM 模型中负相关的基金有 6 只, 似乎看不出 α_i 和 γ_i 有明显负相关的关系。但是, 张新、杜书明 (2002) 对 1999 年底之前设立的 22 只封闭式基金在 1999 年 12 月 31 日至 2000 年 9 月 28 日的择时能力进行研究, 结果是 TM 模型中研究的所有基金都是负相关! 而在 HM 模型中, 则有 21 只基金存在负相关。可以看出, α_i 和 γ_i 之间存在较严重的负相关。因此, 这些模型的设定在某种程度上是值得怀疑的。笔者在下面介绍一种可以避开上述问题、弥补上述不足的检验择时能力的方法——非参数检验。

(二) 非参数检验

由于参数检验方法存在着上面所说的问题和不足, 因此我们试图寻找一种替代的检验方法。Merton (1981) 证明, 若 $\Pr(r_{m,t+1} > 0 | r_{m,t+1} > 0) + \Pr(r_{m,t+1} < 0 | r_{m,t+1} < 0) > 1$, 则表明择时是理性的。若市场择时者 (即基金经理) 对市场的预测是可以被 (外部) 检验者观察到的, 利用 Merton (1981) 上述成果, Henriksson & Merton (1981) 提出了一种非参数检验方法, 这种方法不必假设市场收益率的分布和单个证券定价的方式。但是在现实中, 大多数情况下检验者只能得到基金的历史收益率, 而不能得到基金经理的预测情况。Henriksson 使用基金资产组合中股票的比例作为市场择时预测的一个代理来检验择时能力。他研究的结论是没有找到具有较好择时能力的证据。但是, 这种方法的一个致命之处是: 如果基金经理的预测期间跟这些比例数据可以获得的期间不对应, 那么这种方法就存在很大的度量误差。汪光成 (2002) 就是利用这种方法对我国的 2001 年底之前上市的 33 只基金在 1998 年 6 月 30 日至 2001 年 6 月 30 日的持现比例 (应该是指现金与债券占基金净值的比例) 进行非参数检验的, 检验结果是没有明显的择时能力。但是正如上面所述, 它的缺陷是很明显的。为了解决 Henriksson 上面所提到的度量误差, Henriksson (1984) 运用基金历史收益率与消极管理的基金收益率作对比进行非参数检验。但是, 这种方法也是有问题的, 而且不适用于我国当前的这种情况。综上所述, 上面所提到的这些非参数检验方法都很难在现实中得到很好的运用。

Jiang (2003) 提出了另外一种非参数检验的方法。下面简要介绍这种方法。

正如在式 (1) 中所设定, 一个想择时的经理将是内生移动 $\beta_{i,t}$ 的。设 $\beta_{i,t} = \Phi(r_{m,t+1})$, 其中 $r_{m,t+1} = E(r_{m,t+1} | Z_t)$ 是经理基于 t 时期他所能获得的信息集 Z_t 而做出的对下一个时期市场收益率的预测, 同时合理假设经理不会犯系统性错误, 因而经理的预测是无偏的。对于响应函数 $\Phi(\cdot)$, 不必要求是线性的, 只要求它是 $r_{m,t+1}$ 的非减函数即可。经理的择时能力依赖于信号 Z_t 的相关性和准确性。如果预测正确的概率超过预测不正确的概率的值为 0, 也即下面的参数:

$$v = \Pr(r_{m,t_1+1} > r_{m,t_2+1} | r_{m,t_1+1} > r_{m,t_2+1})$$

$$\begin{aligned}
 & - \Pr(r_{m,t_1+1}) < r_{m,t_2+1} \mid r_{m,t_1+1} < r_{m,t_2+1}) \\
 & = 2\Pr(r_{m,t_1+1}) > r_{m,t_2+1} \mid r_{m,t_1+1} > r_{m,t_2+1}) - 1 \quad (5)
 \end{aligned}$$

v 取 0 时, 对于两个不同的时期 t_1 和 t_2 , 基金经理是缺乏择时能力的。极限的情况是: 如果预测总是完美的, 也即 $r_{m,t+1} = r_{m,t+1}$, 则达到它的上界 1; 相反的情况是完全预测错误了, 则 v 等于 -1。因此, $v \in [-1, 1]$ 就是基金经理择时能力的度量。信息集 Z_i 越准确, v 的值就越高。

如果投资者的择时和选股的信息是独立的, 并且有非减绝对风险厌恶系数的话, Grinblatt & Titman (1989b) 证明了当市场变得更加好时, 投资者总是增加他们的风险系数 Beta, 也即 $\partial\beta_i / \partial r_{m,t+1} > 0$ 。把这个结论跟 (5) 结合起来, 则有:

$$\begin{aligned}
 & \Pr(\beta_{t_1} > \beta_{t_2} \mid r_{m,t_1+1} > r_{m,t_2+1}) - \Pr(\beta_{t_1} < \beta_{t_2} \mid r_{m,t_1+1} < r_{m,t_2+1}) \\
 & = 2\Pr(\beta_{t_1} > \beta_{t_2} \mid r_{m,t_1+1} > r_{m,t_2+1}) - 1 \quad (6)
 \end{aligned}$$

当且仅当经理拥有超常信息时, 上面的概率才会大于 0。在任何市场收益率的范围内, 如果基金经理择时正确的话, 基金的收益率平均是: 当市场收益率较高时, 它应该上升得更高; 当市场收益率较低时, 它应该损失更少。如果是这样, 那么基金收益率具有凸性的概率应该高于具有凹性的概率。但是, 对某一基金的所有收益率来说可以不必是整体凸性或者整体凹性的 (即某一基金可以是局部凸性或者是局部凹性的)。也不必局限于 TM 等模型的参数形式。根据式 (6) 可以构造一个非参数的统计量。

对于从样本中取出任意三个时期的收益率 $\{r_{i,t_j}, r_{m,t_j}\}_{j=1}^3$, 使 $r_{m,t_1} < r_{m,t_2} < r_{m,t_3}$, 有超常信息的经理平均上应该在 $[r_{m,t_2}, r_{m,t_3}]$ 保持高于在 $[r_{m,t_2}, r_{m,t_1}]$ 的 Beta 值。因为在任何点上风险系数 Beta 是基金收益率相对于市场收益率的斜率, 因此有下面的参数:

$$\begin{aligned}
 \theta = & \Pr[(r_{i,t_3} - r_{i,t_2}) / (r_{m,t_3} - r_{m,t_2}) > (r_{i,t_2} - r_{i,t_1}) / (r_{m,t_2} - r_{m,t_1})] \\
 & - \Pr[(r_{i,t_3} - r_{i,t_2}) / (r_{m,t_3} - r_{m,t_2}) < (r_{i,t_2} - r_{i,t_1}) / (r_{m,t_2} - r_{m,t_1})] \quad (7)
 \end{aligned}$$

式中: θ 是经理择时能力的度量指标。当 $\theta = 0$ 时, 表明没有很好的择时能力, 预测正确的概率跟错误的概率是相等的, 在这个假设下, Beta 同市场收益率不相关。由数学上可知, 具有三阶核的 U 统计量是 θ 的一个自然选择:

$$\theta_n = \binom{n}{3}^{-1} \sum_{r_{m,t_1} < r_{m,t_2} < r_{m,t_3}} \left\{ 1 \left(\frac{r_k - r_j}{r_{m,k} - r_{m,j}} > \frac{r_j - r_i}{r_{m,k} - r_{m,i}} \right) - 1 \left(\frac{r_k - r_j}{r_{m,k} - r_{m,j}} < \frac{r_j - r_i}{r_{m,k} - r_{m,i}} \right) \right\} \quad (8)$$

式中: n 是样本容量; $1(\cdot)$ 是指标函数 [也就是在 $1(q)$ 中, $q > 0$ 时其值取 1, $q = 0$ 时其值取 0, $q < 0$ 时其值取 -1], 根据 U 统计量的性质, θ_n 是 θ 的一个 \sqrt{n} -一致和渐进正态的估计量 (Abrevaya & Jiang, 2003), 也即:

$$\sqrt{n}(\theta_n - \theta) \xrightarrow{d} N(0, \sigma_{\theta_n}^2) \quad (9)$$

同时, 它也是对同一参数 θ (基金正确预测下一时期市场收益率的概率) 的所有无偏估计量中的最小方差估计量。

令 $Z_t \equiv (r_{i,t_j}, r_{m,t_j})$, 则 θ_n 的核函数可以表示为:

$$h(z_i, z_j, z_k) = 1[(r_k - r_i)/(r_{m,k} - r_{m,j}) > (r_j - r_i)/(r_{m,j} - r_{m,i}) | r_{m,i} < r_{m,j} < r_{m,k}] - 1[(r_k - r_i)/(r_{m,k} - r_{m,j}) < (r_j - r_i)/(r_{m,j} - r_{m,i}) | r_{m,i} < r_{m,j} < r_{m,k}] \quad (10)$$

Abrevaya & Jiang (2003) 使用杰克刀 (Jackknife) 方法推导了 θ_n 的渐进标准误差:

$$\theta_{\theta_n}^2 = [9/n \sum_{i=1}^n \binom{n-1}{2}^{-1} \sum_{j < k, j \neq i, k \neq i} h(z_i, z_j, z_k) - \theta_n]^2 \quad (11)$$

在 $H_0: \theta = 0$ 下, 根据式 (9) 可得 θ_n 的标准差为 $\sqrt{\sigma_{\theta_n}^2/n}$, $\sqrt{n}\theta_n/\sqrt{\sigma_{\theta_n}^2} \xrightarrow{d} N(0, 1)$, 因此若 $\sqrt{n}|\theta_n|/\sqrt{\sigma_{\theta_n}^2} > z_{\alpha/2}$, 则表明 θ_n 在显著性水平 α 下拒绝 H_0 , 若此时某基金 $\theta_n > 0$, 则表明该基金具有较好的择时能力, 如果是相反 ($\theta_n < 0$), 则基金具有较差的择时能力。

三、样本与数据

(一) 样本期选择

本文研究的基金为 2001 年之前上市交易的 33 只封闭式基金。样本期为 2001 年 1 月 5 日至 2003 年 8 月 29 日。为什么样本期要从 2001 年的 1 月 5 日开始呢? 这主要有两个考虑: 一是可以研究尽可能多的基金 (达到 33 只, 接近封闭式基金总只数的 2/3); 二是考虑到 2000 年 5 月 18 日证监会发布了《关于调整证券投资基金认购新股事项的通知》, 取消了基金在新股配售方面的特权, 但是, 基金先前申购的部分新股在一定时期内其流通还是受到限制的, 为了完全剔除新股配售特权的影响, 样本期的起始点应该选择在所有这些流通受限的股票可以流通的时间之后。基于上述两个考虑, 我们可以更加全面真实地比较基金的业绩与市场基准之间的关系。

(二) 数据及基金周收益率

33 只基金的每周五基金单位净值和样本期内分红数据均来自上海聚源数据库。在样本期内共有 127 个有交易的周五, 因此, 每只基金有 127 个单位净值数据。而对于在一周中有交易, 但是周五没有交易的该周数据, 笔者不拟采用, 造成的影响实际是很小的, 毫不影响我们的检验结论。

本文基金的周收益率定义如下:

$$R_t = (NAV_t + D_t - NAV_{t-1})/NAV_{t-1} \quad (12)$$

式中: R_t 为基金在第 t 周的收益率 (没有减去无风险收益率); NAV_t 为第 t 周末基金净资产; NAV_{t-1} 为第 $t-1$ 周末基金净资产; D_t 为基金在第 t 周分发给基金投资者的现金分红。由上式可知, 每只基金共有 126 个周收益率数据。

(三) 市场基准收益率

众所周知, 我国基金投资股票实际上是投资于深圳 A 股和上海 A 股两个股票市场

流通的股票，而不是所有的股票。由于深圳 A 股指数和上海 A 股指数都是以全部的股本来计算，而不是仅按流通股本部分计算，因此有理由认为这两个指数并不能真正反映基金股票投资市场的实际情况。鉴于这种情况，笔者选用中信指数作为深沪两市股票指数复合的代理，该指数选择深沪两市各行业上市公司 A 股股票中流通市值占该行业前 60% 的若干股票作为样本股，按流通股本加权构造的。另外，我国规定了各基金要至少有 20% 的资产净值投资在国债上，很多研究者便把投资于股票市场资产比例处理为 80%。笔者对所选的 33 只基金在 2001 ~ 2003 年三年中期的持股比例进行统计，2001 年为 66.98%，2002 年为 68.64%，2003 年为 62.85%，平均的持股比例为 66.16%，远低于 80%，因此笔者构造了一个市场基准收益率：

$$R_{m,t} = R_{equity,t} \times 70\% + R_{bond,t} \times 30\% \quad (13)$$

式中： $R_{m,t}$ 为第 t 周复合市场收益率； $R_{equity,t}$ 为中信指数第 t 周收益率； $R_{equity,t}$ 是中信国债指数第 t 周收益率。中信指数和中信国债指数数据均来源于中信指数分析工具。

(四) 无风险利率

考虑到我国的实际情况，本文采用 1 年期定期存款利率作为无风险利率，并按每年 52 周折算为周无风险利率，也就是在 2002 年 2 月 21 日（不含）之前的周利率为 2.25%/52，之后的为 1.98%/52，不考虑利息税。

四、实证研究

(一) 参数检验

参数检验包括 TM 模型和 HM 模型的检验，见表 1 和表 2

表 1 TM 模型回归参数表

序号	代号	简称	α (%)	β	γ	$t(\gamma)$	$Porb(\gamma)$	R^2
1	184688	开元	0.0952	0.8716	-0.0103	-0.6516	0.5159	0.6180
2	184689	普惠	0.1720	0.8836	-0.0434	-1.1013	0.2729	0.2107
3	184690	同益	0.1462	0.8493	-0.0248	-2.5555	0.0118	0.8020
4	184691	景宏	-0.0311	1.0523	-0.0055	-0.1717	0.8640	0.3646
5	184692	裕隆	0.1038	0.8454	-0.0195	-1.6072	0.1106	0.7203
6	184693	普丰	0.0939	0.9388	-0.0250	-0.8332	0.4064	0.3402
7	184695	景博	0.0361	0.9384	-0.0207	-0.9301	0.3541	0.4837
8	184696	裕华	0.0714	0.6761	0.0005	0.0490	0.9610	0.6784
9	184698	天元	0.0797	0.8126	-0.0211	-1.9358	0.0552	0.7467

续上表

序号	代号	简称	α (%)	β	γ	t (γ)	Porb (γ)	R ²
10	184699	同盛	0.1458	0.8676	- 0.0214	- 2.4670	0.0150	0.8407
11	184701	景福	0.0463	1.0285	- 0.0112	0.5145	0.6078	0.5430
12	184702	同智	0.1315	0.7976	- 0.0157	- 1.9127	0.0581	0.8325
13	184703	金盛	0.2876	0.9149	- 0.0471	- 4.4915	0.0000	0.8018
14	184705	裕泽	0.1595	0.9251	- 0.0114	- 1.1272	0.2619	0.8161
15	184708	兴科	0.1561	0.8153	- 0.0100	- 0.7986	0.4261	0.6902
16	184710	隆元	0.2027	0.9316	- 0.0370	- 2.9414	0.0039	0.7442
17	184718	兴安	0.0730	0.8581	- 0.0091	- 0.9285	0.3550	0.8014
18	500001	金泰	0.0970	0.9358	- 0.0323	- 3.7944	0.0002	0.8650
19	500002	泰和	0.1917	0.9579	- 0.0216	- 2.3945	0.0182	0.8564
20	500003	安信	0.1658	0.8109	- 0.0269	- 2.5609	0.0117	0.7590
21	500005	汉盛	0.1286	0.8441	- 0.0474	- 4.3607	0.0000	0.7629
22	500006	裕阳	0.0895	0.9092	- 0.0215	- 1.9631	0.0519	0.7853
23	500007	景阳	0.1267	0.9293	- 0.0154	- 0.5242	0.6011	0.3475
24	500008	兴华	0.0508	0.7953	- 0.0048	- 0.1949	0.8458	0.3551
25	500009	安顺	- 0.0259	0.6925	- 0.0249	- 1.1214	0.2643	0.3398
26	500010	金元	0.0832	0.9395	- 0.0162	- 1.3951	0.1655	0.7768
27	500011	金鑫	0.0850	0.9223	- 0.0163	- 2.2480	0.0264	0.8952
28	500015	汉兴	0.0692	0.9154	- 0.0468	- 4.5385	0.0000	0.8075
29	500016	裕元	0.1041	0.7674	- 0.0036	- 0.3220	0.7480	0.7180
30	500018	兴和	0.1762	0.9426	- 0.0087	- 0.3327	0.7399	0.4088
31	500021	金鼎	0.1505	0.9667	- 0.0272	- 2.7705	0.0065	0.8368
32	500025	汉鼎	0.0299	0.8349	- 0.0243	- 2.3699	0.0193	0.7777
33	500035	汉博	0.1361	0.7083	- 0.0442	- 3.9242	0.0001	0.6802

表 2 HM模型回归参数表

序号	代号	简称	α (%)	β	γ	t (γ)	Porb (γ)	R ²
1	184688	开元	0.1245	0.8182	- 0.0980	- 0.5477	0.5849	0.6176
2	184689	普惠	0.1496	0.7569	- 0.2154	- 0.4801	0.6320	0.2044
3	184690	同益	0.2179	0.7198	- 0.2378	- 2.1483	0.0337	0.7991
4	184691	景宏	- 0.1767	1.1330	0.1668	0.4596	0.6466	0.3656
5	184692	裕隆	0.1288	0.7649	- 0.1442	- 1.0446	0.2982	0.7170
6	184693	普丰	0.0541	0.8839	- 0.0877	- 0.2569	0.7977	0.3368
7	184695	景博	0.0416	0.8670	0.1248	- 0.4929	0.6229	0.4811
8	184696	裕华	0.1095	0.6521	- 0.0487	- 0.3994	0.6903	0.6788
9	184698	天元	0.1506	0.6957	- 0.2156	- 1.7446	0.0836	0.7452
10	184699	同盛	0.2197	0.7476	- 0.2216	- 2.2446	0.0266	0.8394
11	184701	景福	- 0.0109	1.0306	0.0144	0.0583	0.9536	0.5420
12	184702	同智	0.1813	0.7125	- 0.1567	- 1.6751	0.0965	0.8314
13	184703	金盛	0.4149	0.6748	- 0.4397	- 3.6041	0.0005	0.7913
14	184705	裕泽	0.1999	0.8605	- 0.1193	- 1.0408	0.3000	0.8158
15	184708	兴科	0.1429	0.7914	- 0.0389	0.2726	0.7856	0.6888
16	184710	隆元	0.3289	0.7255	- 0.3807	- 2.6585	0.0089	0.7411
17	184718	兴安	0.0875	0.8184	- 0.0715	- 0.6396	0.5236	0.8007
18	500001	金泰	0.2055	0.7570	- 0.3299	- 3.3900	0.0009	0.8621
19	500002	泰和	0.2655	0.8374	- 0.2226	- 2.1693	0.0320	0.8552
20	500003	安信	0.2401	0.6727	- 0.2532	- 2.1101	0.0369	0.7550
21	500005	汉盛	0.2694	0.5938	- 0.4599	- 3.6570	0.0004	0.7531
22	500006	裕阳	0.1779	0.7793	- 0.2415	- 1.9504	0.0534	0.7852
23	500007	景阳	0.0616	0.9233	0.0016	0.0048	0.9962	0.3460
24	500008	兴华	- 0.0191	0.8270	0.0678	0.2418	0.8093	0.3552
25	500009	安顺	0.1247	0.5092	- 0.3455	- 1.3778	0.1708	0.3432
26	500010	金元	0.1531	0.8393	- 0.1866	- 1.4215	0.1577	0.7769
27	500011	金鑫	0.1452	0.8283	- 0.1741	- 2.1119	0.0367	0.8947
28	500015	汉兴	0.2265	0.6561	- 0.4786	- 4.0347	0.0001	0.8015
29	500016	裕元	0.0928	0.7633	- 0.0049	- 0.0390	0.9690	0.7177
30	500018	兴和	0.1072	0.9609	0.0445	0.1504	0.8807	0.4083
31	500021	金鼎	0.2241	0.8280	- 0.2540	- 2.2621	0.0254	0.8335
32	500025	汉鼎	0.0984	0.7091	- 0.2306	- 1.9709	0.0510	0.7747
33	500035	汉博	0.2489	0.4876	- 0.4034	- 3.0956	0.0024	0.6662

表 1 和表 2 中, $\text{Prob}(\gamma) = [1 - \text{Prob}(x \leq |\gamma|)] \times 2$, 当该栏的值小于 0.1 (0.05), 我们就说该对应的基金在 10% (5%) 显著性水平下是显著的。由表 1 和表 2 可知, TM (HM) 模型 $\gamma < 0$ 的基金有 32 (28) 只, 在 10% 和 5% 显著性水平下显著的基金数分别达到 16 (16) 和 13 (12) 只, 也就是说, 无论是 TM 模型还是 HM 模型, 检验结果都表明绝大多数基金经理具有较差的择时能力而不是具有较好的择时能力, 而且他们当中具有较差择时能力的基金有近一半还是很显著的! 另外, TM 模型和 HM 模型中 α 与 γ 之间存在负相关的基金只数分别为 30 和 31 只, 也即几乎所有基金经理的选股能力与择时能力都是负相关的。这个结论很难令人信服, 从而使得我们对其结论的可靠性提出疑问。

(二) 非参数检验

非参数检验是利用 Jiang (2003) 提出的方法进行检验的, 检验结果见表 3。

表 3 非参数检验数据表

序号	代号	简称	θ (%)	θ 标准差 (%)	$\sqrt{n} \theta /\sqrt{\sigma_\theta^2}$	$\text{Porb}(\sqrt{n} \theta /\sqrt{\sigma_\theta^2})$
1	184688	开元	- 0.2138	8.3441	0.0256	0.9796
2	184689	普惠	- 5.6467	8.4777	0.6661	0.5053
3	184690	同益	- 3.4575	8.4953	0.4070	0.6840
4	184691	景宏	0.4737	8.2402	0.0575	0.9541
5	184692	裕隆	- 3.1601	8.0709	0.3915	0.6954
6	184693	普丰	- 3.8280	8.2684	0.4630	0.6434
7	184695	景博	- 4.3250	8.0394	0.5380	0.5906
8	184696	裕华	- 3.3948	8.3047	0.4088	0.6827
9	184698	天元	- 2.1413	8.4258	0.2541	0.7994
10	184699	同盛	- 3.8808	8.3402	0.4653	0.6417
11	184701	景福	0.4479	8.3748	0.0535	0.9573
12	184702	同智	- 2.9493	8.3427	0.3535	0.7237
13	184703	金盛	- 6.7312	8.1811	0.8228	0.4106
14	184705	裕泽	- 3.2842	8.0954	0.4057	0.6850
15	184708	兴科	1.3512	8.4390	0.1601	0.8728
16	184710	隆元	- 4.2464	8.0937	0.5247	0.5998
17	184718	兴安	2.5886	8.1759	0.3166	0.7515
18	500001	金泰	- 6.8246	7.9943	0.8537	0.3933
19	500002	泰和	- 2.0498	8.0617	0.2543	0.7993
20	500003	安信	2.4233	8.7857	0.2758	0.7827

续上表

序号	代号	简称	θ (%)	θ 标准差 (%)	$\sqrt{n} \theta /\sqrt{\sigma_{\theta n}^2}$	Prob ($\sqrt{n} \theta /\sqrt{\sigma_{\theta n}^2}$)
21	500005	汉盛	- 7.6571	8.2847	0.9243	0.3553
22	500006	裕阳	- 4.0326	8.0304	0.5022	0.6155
23	500007	景阳	- 1.5810	8.2017	0.1928	0.8471
24	500008	兴华	- 1.4962	8.1496	0.1836	0.8543
25	500009	安顺	- 3.0077	8.6911	0.3461	0.7293
26	500010	金元	- 1.6707	8.1305	0.2055	0.8372
27	500011	金鑫	- 3.2965	7.8150	0.4218	0.6732
28	500015	汉兴	- 8.6083	8.2438	1.0442	0.2964
29	500016	裕元	1.5060	8.3318	0.1808	0.8565
30	500018	兴和	- 1.7045	8.3351	0.2045	0.8380
31	500021	金鼎	- 4.8000	8.3729	0.5733	0.5664
32	500025	汉鼎	- 0.1825	8.1287	0.0225	0.9820
33	500035	汉博	- 5.6823	7.8486	0.7240	0.4691
平均	—	—	- 2.7595	1.4359	1.9218	0.0546
加权	—	—	- 2.7804	1.4345	1.9382	0.0526

表 3 中, $\text{Prob}(\sqrt{n}|\theta|/\sqrt{\sigma_{\theta n}^2}) = [1 - \text{Prob}(x \leq \sqrt{n}|\theta|/\sqrt{\sigma_{\theta n}^2})] \times 2$, 当该栏的值大于 0.1 (0.05), 我们就说对应的基金在 10% (5%) 显著性水平下是不显著的。由表 3 可知, 从整体上看, 基金具有较差的择时能力 (等权平均的 $\theta = -2.7595\%$, 用标准差的倒数加权平均的 $\theta = -2.7804\%$), 而且无论是等权平均或者是不等权平均, 均在 10% 显著性水平下是显著的。从单个基金来看, $\theta > 0$ 的基金有 6 只, $\theta < 0$ 的基金有 27 只。在 $\alpha = 10\%$ 的显著性水平下, 无论是 $\theta > 0$ 还是 $\theta < 0$, 没有一只基金是显著的, 也就是说没有一个基金经理表现出显著很好的择时能力 (θ 显著大于零), 也没有一个基金经理表现出显著很差的择时能力 (θ 显著小于零)。这个结果跟我们运用参数方法检验的结果有很大的区别。参数检验的结论是在个体上很多基金经理存在显著较差的择时能力。正如我们在参数检验中所指出的一样, 由于参数检验中 α 与 γ 之间严重的负相关, 从而使得我们怀疑其实证结果的正确性, 而非参数检验不存在这样的问题。因此, 我们相信非参数检验的结果更加可信。

五、结论与启示

本文运用 Jiang (2003) 提出的非参数方法对我国 33 只封闭式证券投资基金的择时能力进行实证研究, 并把研究的结果跟参数方法 (本文只采用 TM 和 HM 这两个参数模型)

的结果相对比。研究表明非参数方法可以克服参数方法过于严格的假设条件以及避开选股能力与择时能力之间存在严重负相关性这一难以解释的问题。因此,非参数方法的实证结果应该更加具有说服力。非参数实证研究的结果表明,对于我们所研究的 33 只基金来说,虽然基金经理更多地表现出较差的择时能力这一跟国外、国内相类似的结论(本文中基金整体上在 10% 显著性水平下是负显著的),但是,没有一个基金经理表现出显著很好的择时能力或者是显著很差的择时能力。

上述结论对于投资者和基金经理来说都具有非常重要的意义,给予我们的启示是:①就市场择时能力而言,基金在整体上或者是个体上都不能实现投资者所期望的专家投资,基金的表现并没有比市场好;②当前基于择时能力的基金评级指标是失真的,存在误导的可能;③对于基金管理人来说,由于基金在个体上既没有显著较好的市场择时能力,也没有显著较差的市场择时能力,因而基金经理应该把更多的时间和精力花在单个证券的选择上。实际上,这也是当前我国基金管理公司研究和关注的重点。只有努力提高证券选择的能力,才能使自己的基金在同行中鹤立鸡群、傲视群雄。

参考文献

- [1] 沈维涛,黄兴李:《我国证券投资基金业绩的实证研究与评价》,《经济研究》,2001 年第 9 期。
- [2] 王聪:《证券投资经济绩效评估模型分析》,《经济研究》,2001 年第 9 期。
- [3] 汪光成:《基金的市场时机把握能力研究》,《经济研究》,2002 年第 1 期。
- [4] 魏先华,朱世武,梁衡义:《中国基金经理能正确把握市场时机吗?》,《世界经济》,2003 年第 6 期。
- [5] 张新,杜书明:《中国证券投资基金能否战胜市场?》,《金融研究》,2003 年第 1 期。
- [6] Abrevaya, Jason, and Wei Jiang (姜纬), Pairwise-slopes Statistics for Testing Curvature, Working Paper, 2003.
- [7] Admati, Anat R., Sudipto Bhattacharya, Paul Pfleiderer, and Stephen A. Ross, On Timing and Selectivity, Journal of Finance, 1986, Vol.41, pp715 ~ 730.
- [8] Fama, E., and K. French, Common Risk Factors, in the Returns on Stocks and Bonds, Journal of Financial Economics, 1986, Vol.33, pp3 ~ 56.
- [9] Grinblatt, Mark, and Sheridan Titman, Portfolio Performance Evaluation: Old Issues and New Insights, Review of Financial Studies, 1896, Vol.3, pp393 ~ 421.
- [10] Henriksson, Roy D., Market Timing and Investment Performance: An Empirical Investigation, Journal of Business, 1984, Vol.57, pp73 ~ 96.
- [11] Henriksson, Roy D., and Robert C. Merton, On Market Timing and Investment Performance. II. Statistical Procedures for Evaluating Forecasting Skills, Journal of Business, 1981, Vol.54, pp513 ~ 533.
- [12] Ippolito, Richard A., On Studies of Mutual Fund Performance, 1962 - 1991, Financial Analysts Journal Jan/ Feb, 1993, Vol.93, pp42 ~ 50.
- [13] Jiang, Wei (姜纬), A Nonparametric Test of Market Timing, Journal of Empirical Finance, 2003, Vol. 10, No. 4, pp 399 ~ 425.
- [14] Merton, Robert C., On Market Timing and Investment Performance. I. An Equilibrium Theory of Value for Market Forecasts, Journal of Business, 1981, Vol.54, pp363 ~ 406.

- [15] Simons , Katerina , Risk-adjusted Performance of Mutual Funds , New England Economic Review Sep/Oct , 1998 , Vol.98 , pp33 ~ 48.
- [16] Treynor , Jack L. and Kay K. Mazuy , Can Mutual Funds Outguess the Market ? Harvard Business Review , 1966 , Vol.44 Issue 4 , pp131 ~ 136.

中国货币市场运行：内生性、调整时滞与动态*

王 曦 舒 元

我国目前实行固定利率制度、特殊的汇率制度及外汇市场安排。在这种特殊的金融制度下，我国货币市场具有特殊的运行特征及含义。规范地阐述其特征和含义，正是本文的研究目的。

一、理论基础与比较静态

这里先阐述现有制度下货币需求、货币市场均衡、国际收支决定和内生货币供给的机制及其含义，再把它们综合成为一个统一的比较静态分析框架，为后面的动态分析提供理论基础。

（一）经济转型中的货币需求

王曦（2001）在评价了传统货币需求理论之后，以凯恩斯对货币需求动机的分类作为基本框架，在交易和预防性货币需求中吸取随机货币需求理论的要旨，在投机性动机中综合托宾对于风险的规范性分析和弗里德曼对于不同财富形式的强调，规范地分析了我国货币需求的微观基础，推导出了经济转型时期我国货币需求的理论模型，并进行了实证分析。其分析表明，在证券市场出现以后，我国实际货币需求（L）函数为：

$$L = f(Y + \Omega + iF_p + i \pm \pi^e - O) \quad (1)$$

其中：“+”、“-”和“±”分别表示L对该解释变量的偏导数为正、负和不定；Y， Ω ， F_p ，i和 π^e 分别代表实际国民收入、非国有经济的发展程度、价格自由化进程、利率和通货膨胀预期；O为其他因素的影响，包括证券市场交易量、收益率的预期和方差等。

式（1）表明，我国的货币需求是对于实际余额的需求，与物价的绝对水平无关。

（二）固定利率下的货币市场均衡与商品价格

我国实行固定利率制，在调节货币市场失衡时，教科书式的利率机制无效。失衡只能通过商品价格来调整，物价的变动成为失衡的调节机制。

考虑货币市场均衡条件： $M_s/P = L(i, o_1)$ 。其中， M_s 和P分别是初始的名义货币供给和价格水平；L是实际货币需求，它是利率i和其他变量 o_1 的函数。假定初始状态

* 本研究得到□川博士论文重大创新项目基金的资助，感谢中山大学王美今教授等提出的宝贵意见。原载：《经济学》（季刊），2003年第4期。

时货币市场处于均衡，均衡利率是 i_0 。如果名义货币供给增加， M_s/P 增加至 M'_s/P ，货币市场出现超额货币供给。如果利率是可变的， i_0 会变动到新的利率 i' 以实现新的均衡；但是如果利率是固定的，并且 o_1 是外生的^①，则调节非均衡就只能通过物价水平的变动来进行。此时为了维持货币市场均衡，物价水平必须由 P 上升到 P' ，使得实际货币供给回到初始水平，即 $M'_s/P' = M_s/P$ ，以维持固定利率。以上分析同样适用于 M_s 减少或外生冲击 o_1 引起 L 变动的情况，这里不再赘述。

对货币市场均衡条件 $M_s/P = L$ 两边取自然对数，有货币市场均衡方程：

$$\ln(P) = \ln(M_s) - \ln(L) \quad (2)$$

(三) 物价水平、国际收支与外汇储备

国际收支包括经常账户和资本账户。经常账户方面最主要的是贸易收支项目，它取决于国内外收入水平和实际有效汇率。其中，收入水平在短期内可以视为外生给定；贸易收支与实际有效汇率的关系取决于马歇尔 - 勒纳条件。有证据显示，马歇尔 - 勒纳条件在我国得到满足（课题组，1999），因此贸易收支顺差额与实际有效汇率（REER）水平是正相关的。实际有效汇率是指考虑到物价水平变动并以多边贸易或竞争力指标加权的汇率指标，作为一个简化的处理，这里定义 $REER = E_r \cdot P_f/P$ 。其中 E_r 是以直接标价法表示的本国汇率； P_f 和 P 分别是国外和本国物价水平。

在资本账户方面，我国对间接资本流动实行严格的管制，因此对于外汇供给比较重要的子项目是外商直接投资（FDI）和我国对外直接投资（DFI）。关于 FDI，国外学者已经多有论述（Dunning, 1993；Chen Chunlai, 1997），它取决于国内效率工资、市场潜力、FDI 存量、经济增长、世界主要市场的地理距离等非汇率、非物价因素。关于 DFI，可以反过来论述。

以 σ 表示除了 REER 以外所有其他因素对国际收支的影响，综合以上分析，有我国国际收支顺差（BP）的决定公式为：

$$BP = g((E_r \cdot P_f/P) + \sigma)$$

外汇储备（ R_p ）是国际收支顺差的累积，则根据上式，有：

$$R_p = \sum_{i=0}^{\infty} BP_{t-i} = \sum_{i=0}^{\infty} g((E_{r,t-i} \cdot P_{f,t-i}/P_{t-i}) + \sigma_{t-i})^{\textcircled{2}}$$

为了方便地发析，我们可以把所有滞后因素都合并到其他因素中，则有：

$$R_p = h((E_r \cdot P_f/P) + \sigma')$$

其中： σ' 为其他因素。设 R_p 具有对数形式：

$$\ln R_p = \alpha \ln(E_r \cdot P_f/P) + o = \alpha \ln(E_r \cdot P_f) - \alpha \ln P + o \quad (3)$$

① 在式（1）中，利率是固定的，制度性因素（ Ω 和 F_p ）明显是外生的；在局部分析的框架中， Y 和 O 是外生的； π^e 取决于价格水平的持续变化，与一次性价格冲击没有关系，因而也是外生的。

② 在 1994 年以前，我国实行双轨制的汇率制度和外汇留成制度，因此国际收支和外汇储备的决定方程会有所变化。但是从我国外汇储备的实际变化上看，我国大规模的国际收支顺差是在 1994 年以后发生的。这里的简化处理应该不会对最终结论产生重要影响。

其中： $\alpha > 0$ ，代表外汇储备对实际有效汇率的弹性； o 为其他因素影响。可见，我国外汇储备与国内物价水平成反比，并受到其他因素的影响。

(四) 外汇市场运行与内生货币供给

货币供给 (M_s) 等于货币乘数 (mm) 乘以基础货币 (B)，即 $M_s = mm \cdot B$ 。

1994年，我国对外汇管理体制进行了重大改革，开始实行以结售汇制为主要特征的“有管理的浮动汇率制”。陈平和王曦(2002)分析了该制度下外汇交易市场经济主体的微观行为基础，并据此建立了我国的外汇需求和供给函数；进而分析了市场的均衡和非均衡态势，讨论了人民币汇率稳定运行态势后面的微观行为原因。分析发现：①人民币汇率市场非均衡是外汇市场的主要特点。②对中央银行行为的分析表明，非均衡的经常性和信息的缺乏使得当前汇率制度类似于“盯住汇率制”。③经常性的失衡使得冲销性干预极为困难，在现有的债券市场规模下，大规模的冲销性干预也是不可能的^①。

这些结论意味着：①人民币事实上的固定汇率特征使得我们可以把式(3)中的 E_r 视为一个常数。事实上，在1995年7月以后，人民币/美元汇率一直保持在 8.27 ~ 8.33 (人民币/美元) 之间。②在结售汇制下，国际收支差额直接表现为货币发行的变动，外汇占款发行使得我国的货币供给具有很强的内生性。

对货币创造方程 $M_s = mm \cdot B$ 两边取对数，有以下的内生货币供给方程：^②

$$\ln(M_s) = \ln(mm) + \ln(B) = \ln(mm) + \ln(R_p) - \ln[R_p/B] \quad (4)$$

(五) 货币 - 商品 - 外汇市场的联合比较静态分析

综合以上分析，可以推导出我国货币市场均衡时的货币供给量、价格水平和国际储备。

为了简便起见，分别以 m, l, p 和 b 代表 $\ln M_s, \ln L, \ln P$ 和 $\ln R_p$ ；记 $c = \alpha \ln(E_r \cdot P_f) + o$ ；并以 k 和 x 分别代表 $\ln(mm)$ 和 $\ln(R_p/B)$ 。由式(2)、(3)和(4)，得到我国货币市场比较静态的联立方程：

$$\begin{cases} m - p - l = 0 & (5) \\ b = c - \alpha p & (6) \end{cases}$$

$$m = k + b - x \quad (7)$$

其中：式(5)为均衡方程，其他为决定方程。解这个联立方程，得到货币市场均衡时的货币供给量、价格水平和外汇储备为：

① 关于“均衡”与“非均衡”两个概念，在经济学中历来有不同的理解。一种看法大致认为，在给定的资源、信息与制度约束下，经济主体能够实现其最优化行为，这就应该理解为“均衡”。这方面的代表是所谓“一般均衡”分析方法。另一种看法更加传统些，它强调市场供需的对比。认为在给定的价格变量（如商品价格、利率和工资等）下，市场供给和需求量不一致就是“非均衡”，否则为“均衡”。本文的“均衡”和“非均衡”概念主要指市场供需而言。

② 后文在处理中相当于把 BP/B 和 mm 作为外生变量。在1996年以后货币当局的资产负债表上， BP/B 一直维持在 0.3 ~ 0.4 之间，这样做基本上是合理的。另外， mm 取决于现有制度下公众、商业银行的行为以及中央银行的政策取向，因此 mm 也可以视为外生变量。

$$\begin{aligned} m^* &= (k + c + \alpha l - x)/(1 + \alpha) \\ p^* &= (k + c - l - x)/(1 + \alpha) \\ b^* &= [c - \alpha(k - x - l)]/(1 + \alpha) \end{aligned} \quad (8)$$

以上均衡总量的公式可见，任何影响 k , c , α , l 和 x 的因素都会反映在货币市场上。如果我们试图进行进一步的分析，可以把式 (1) 和式 (3) 中更深层次的决定因素代入到式 (8) 内，分别分析它们对均衡总量的影响，限于篇幅，这里不再赘述。

二、货币市场动态分析 - 调整时滞与均衡动态

比较静态分析给出了货币市场均衡时的总量水平。但货币市场的实际运行是一个动态的过程。要分析货币市场的实际运行，必须考虑到更现实的情况，这就是商品、货币和外汇市场的调整时滞及其影响。以下分析先考虑最简单的情况，即除非有所说明，所有的外生变量都保持固定不变；然后再考虑更现实的情形。

(一) 调整时滞与调整方程

由式 (5)，有 $p = m - l$ 。它表示在给定的名义货币供给和实际货币需求下，物价应该达到的水平，以后简称为物价局部均衡水平。在固定利率制下，价格水平成为是货币市场失衡的调节机制。但是价格水平的变动不是一蹴而就的，存在调整时滞。考虑经济中存在超额货币供给的情况，如果厂商不知道货币超额供给的信息，他就不可能一次性地把商品价格调节到均衡水平，而是会随着市场状况进行逐步的调整；另一方面，即使厂商知道货币超额供给的信息，并且预期商品价格水平将会上升到新的均衡水平，他也不可能迅速地调节价格水平。这是因为，经济中存在众多竞争性的厂商，如果单个厂商一次性地将价格提高到位而其他厂商并不跟随，该厂商的产品将没有销路。因此，厂商在提价时必然要考虑其他厂商的行为，相互观望的调整过程决定了价格水平的调整存在时滞。

无论如何，厂商调整价格水平总是有所依据的，这个依据就是经济中货币超额供给的程度。容易想象，如果经济中货币超额供给的量，价格水平调整的速度就快；超额供给小，价格水平调整就慢。这意味着价格水平具有下面的渐进调整机制：

$$p = \lambda[(m - p) - l] = \lambda[m - (p + l)] \quad (0 < \lambda < 1) \quad (9)$$

其中： $p = dp/dt$ ，是对数价格水平的变化速度或绝对价格水平增长率； $(m - p)$ 为实际货币供给对数； $(p + l)$ 为名义货币需求对数。上式的含义在于，价格水平根据实际超额货币供给 $[(m - p) - l]$ 或名义超额货币供给 $[m - (p + l)]$ 进行调整，渐进地趋进局部均衡水平。

在货币供给方面，合并式 (6) 和 (7)，有： $m = k + c - x - \alpha p$ 。它表示在给定的价格水平和外生变量下，货币供给应该达到的水平，以后简称为货币供给局部均衡水平。货币创造是指基础货币会通过“存款 \rightarrow 贷款 \rightarrow 派生存款 \rightarrow 派生贷款 \rightarrow ……”的机制创造货币，这也需要一个过程，因此货币供给实际水平不会马上达到局部均衡水平。二者的差距越大，货币创造过程中的当期派生存款和派生贷款就越多，货币创造就越快；差距越小，货币创造就越慢。则货币供给具有下面的渐进调整机制：

$$\dot{m} = \lambda[(k + c - x - \alpha p) - m] \quad (10)$$

其中： $\dot{m} = dm/dt$ ，为对数货币供给的变化速度或货币供给的增长率；设调整系数也是 λ 。^①上式表示，货币供给根据局部均衡水平和实际水平的差额进行调整，渐进地趋进局部均衡水平。

(二) 货币市场均衡动态

把式(9)和(10)联立，得到我国货币市场运行的动态微分系统：

$$\begin{cases} \dot{p} = \lambda[(m - p) - 1] = \lambda[m - (p + 1)] & (9) \\ \dot{m} = \lambda[(k + c - x - \alpha p) - m] & (10) \end{cases}$$

下面我们利用相图(Phase Diagram)技术分析这个系统，其好处在于简便和直观。对于该系统更为正式的分析，可以利用联立微分方程，详见附录。两种方法分析的结论是一致的。

做 $p - m$ 相图，第一步是绘出局部均衡条件，见图1。对于方程(9)，局部均衡时 p 不变。此时有 $\dot{p} = 0$ 或 $m - p = 1$ ，这也是比较静态分析中的货币市场均衡条件，表示价格的调整到位，它是向右上方倾斜的直线。对于方程(10)，局部均衡时 m 不变，则 $\dot{m} = 0$ 或 $m + \alpha p = k + c - x$ ，这也是比较静态分析中的货币创造方程，表示货币创造完成，它是向右下方倾斜的直线。这两条局部均衡直线相交于 A，即为一般均衡点或稳态(Steady State)点，表示两个局部均衡条件同时满足，A 也就是比较静态分析中的均衡点。两条直线把平面分割成 I、II、III 和 IV 四个象限。做相图的第二步是给出非稳态点的运动方向。在 $m - p = 1$ 的右下方的点， $m - p - 1 > 0$ ，则根据式(8)， $\dot{p} > 0$ ，在 I 和 IV 象限画出向上的箭头；在 $m - p = 1$ 的左上方， $m - p - 1 < 0$ ， $\dot{p} < 0$ ，在 II 和 III 象限画出向下的箭头。类似地， $\dot{m} = 0$ 右上方的点， $\dot{m} < 0$ ；左下方的点， $\dot{m} > 0$ 。因此在 I 和 II 象限中画出左箭头；在 III、IV 象限中画出右箭头。至此，平面中各点的运动方向已定。

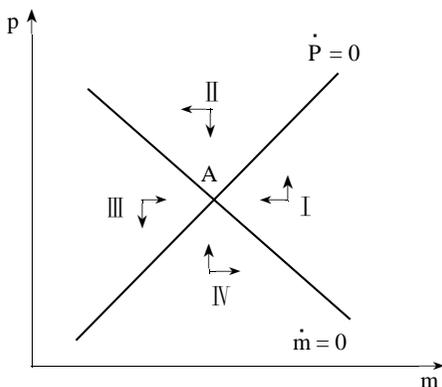


图1 相图 - 局部均衡条件与调整方向

① 为了方便起见，我们引入该假设。可以证明，只要货币供给和价格水平的调整系数相差不大，就不会影响下面分析的定性结论。这个假定也是有依据的：可以证明，如果两个调整系数相差很大，在货币供给和货币需求方面的持续正向冲击下，不会引起图4和5的回型和尖峰型运动，这与图6的 (m, p) 实际运行形态不符。当然，如果能够实证评估出调整系数是最理想的。但是本文第一部分的分析表明，在 l 、 b 和 nm 的决定因素中含有诸多的制度变量，现有的数据量根本无法满足实证估计的要求。

下面考虑单次冲击下价格水平和货币供给的联合运动。图 2 表示了一次货币供给方面冲击的动态影响。假定在初始状态 (m_0, p_0) ，货币市场处于均衡状态。此时如果 k, c 或 x 发生了一次性的变化使得 $(k + c - x)$ 增加，货币市场即遭受一次货币供给方面的冲击。此时，直线 $m=0$ 向右上方移动，它与货币市场局部均衡直线 $p=0$ 相交于新的稳态点 (m^*, p^*) 。但是 (m_0, p_0) 不会以直线运动到 (m^*, p^*) ，按照相图规定的方向，在 (m_0, p_0) ，由于 $p=0$ 和 $m>0$ ， m 将脱离直线 $p=0$ ，运动到象限 IV；类似地，在象限 I， p 将脱离直线 $m=0$ ；……。另外，由于在每一个象限，总有 m 或 p 的方向是朝向新稳态点的，因此 (m, p) 的运动轨迹不会发散或循环。这个过程持续下去， (m, p) 的运动轨迹将是一条螺旋曲线， (m_0, p_0) 渐进地并且振荡地趋进 (m^*, p^*) 。货币需求方面冲击的影响可以类推，它也是一条螺旋线。

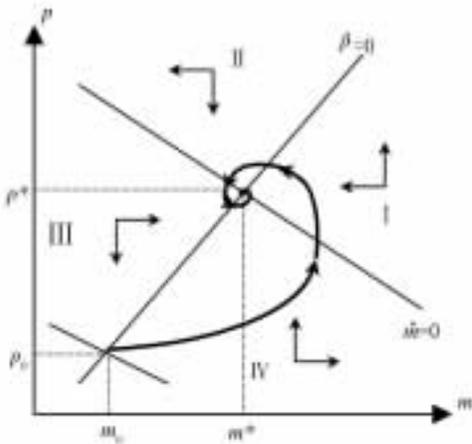


图 2 单次冲击下的动态调整

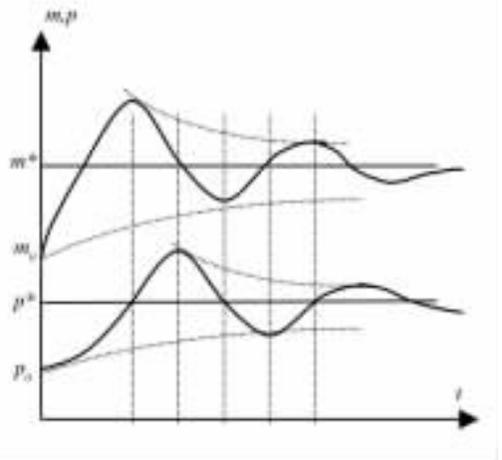


图 3 货币供给与价格的时间轨迹

我们也可以单独绘出 m 和 p 的时间轨迹。附录中的正式分析表明，它们将以递减的正弦波渐进地趋近于新的稳态水平，并且二者具有相同的周期 $T = 2\pi\sqrt{\alpha}/(\lambda\alpha)$ ，周期的长度与外汇储备的汇率弹性和调整系数有关，与其他因素和冲击的大小无关。对于图 2 的情形， m 与 p 的时间轨迹见图 3。

(三) 更现实的调整

货币市场遭受单一冲击的情形不足以解释现实的情况。现实中，我国货币市场同时遭受到货币供给和货币需求不间断的冲击。^① 总体看来，我国的名义货币供给将主要因为国际收支持续顺差而持续增加，因为其他因素而有所波动；实际货币需求因为经济持续增长而增长，因为其他因素而有所波动。在这些因素的共同作用下，我国货币市场的运行会呈现更复杂的形态。

^① 见后文中对于冲击经常性的论述。

考虑一个更加现实的调整过程，在这个过程中，货币市场先后遭受到两次货币供给冲击和一次货币需求冲击。图 4 中，A 点为初始状态。此时一次货币供给冲击使得货币创造局部均衡曲线由 m_0 移动到 m_1 ，系统新的稳态点为 m_1 与货币市场局部均衡线 l_0 的交点， (m, p) 将沿着 AB 螺旋线进行运动；当到达 B 点后，系统遭受到一次货币需求冲击， l_0 移动到 l_1 ，新的稳态点为 l_1 与 m_1 的交点， (m, p) 将沿着 BC 螺旋线进行运动；当到达 C 点以后，又一次货币供给冲击使得 m_1 上移到 m_2 ， (m, p) 新的运动开始。在这个过程中， (m, p) 在 B 点形成了一个回型，在 C 点形成了一个峰。如果引起 l_0 移动到 l_1 的货币需求冲击发生得较快，则在 B 点不足以形成回型时 (m, p) 就开始向右下方移动，这就是图 5 的情形。此时，在 B 点形成了一个尖峰，在 C 点形成了一个小峰。

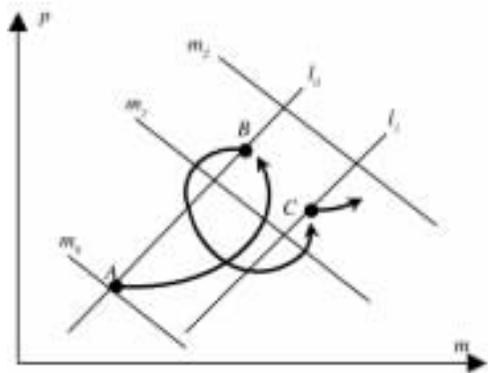


图 4 更现实的调整——回型

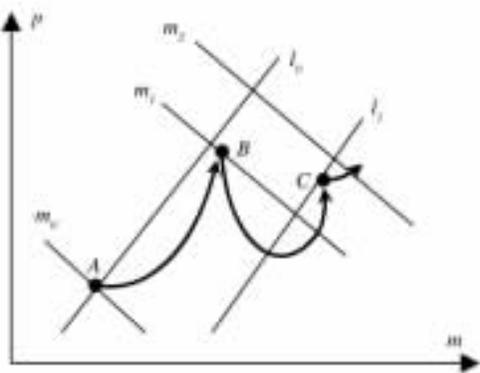


图 5 更现实的调整——峰型

通过模拟可以发现，不同的冲击组合会使 (m, p) 组合的运动呈现复杂的形态，但都以图 4 和图 5 的回型和峰型最为典型。深究 (m, p) 各种形态的决定因素与本文分析的关联不大，这里不再赘述。

图 6 给出了 1996 年 6 月至 2001 年 6 月以月份数据描绘的我国不同口径的供给和价格水平之间的实际组合运动。^① 从图中可见，我国货币市场的实际运行与理论预测是相符的。其中， $m_0 - p$ 的实际轨迹表现出明显的回型运动； $m_2 - p$ 表现出明显的峰型运动； $m_1 - p$ 的情况处于 m_0 和 m_1 之间。

^① 数据来源，《中国人民银行统计季报》各期。其中，价格水平采用定基比消费物价指数，根据同比和环比消费物价指数核算，数据备索。

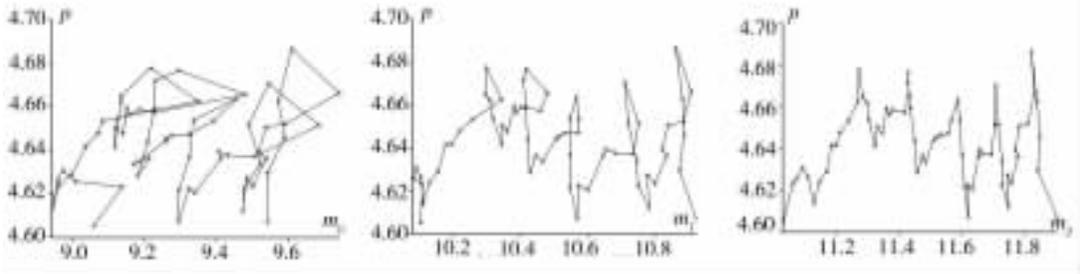


图6 我国货币供给 - 价格水平的实际运行： m_0 ， m_1 和 m_2

三、货币供给 - 价格的因果关系：结论与检验

式 (9) 和式 (10) 微分动态系统的结构还暗示着：货币供给和价格水平是互为因果的关系。

以 Δp_t 和 Δm_t 替代式 (9) 和 (10) 中的 p 和 m ；把其他各项都加上下标“ $t - 1$ ”，可得出差分动态系统：

$$\begin{cases} \Delta p_t = \lambda [(m_{t-1} - p_{t-1}) - l_{t-1}] \\ \Delta m_t = \lambda [(k_{t-1} + c_{t-1} - x - \alpha p_{t-1}) - m_{t-1}] \end{cases}$$

动态系统表示，当期的 p 根据上期的超额货币供给进行调整；当期的 m 根据上期货币供给局部均衡水平和现实水平的差额进行调整。再把 $\Delta p_t = p_t - p_{t-1}$ 和 $\Delta m_t = m_t - m_{t-1}$ 代入到差分系统中，经整理，有：

$$\begin{cases} p_t = \lambda (m_{t-1} - l_{t-1}) + (1 - \lambda) p_{t-1} \\ m_t = \lambda (k_{t-1} + c_{t-1} - x_{t-1} - \alpha p_{t-1}) - (1 - \lambda) m_{t-1} \end{cases}$$

可见，滞后的 m 对 p 有解释力；滞后的 p 对 m 有解释力。这也就是说，运用包含了 m_t (或 p_t) 的信息就能够更好地预测 p_{t+1} (或 m_{t+1})，这刚好意味着， m 和 p 是戈兰杰因果 (Granger Causality) 意义上的互为因果的关系。

货币供给 - 物价的互为戈兰杰因果关系是我们的一个理论推论，它在另一方面也提供了检验理论模型合理性的一个实证方法。下面我们用戈兰杰因果检验 (Granger, 1969) 来验证这个理论预测。限于篇幅，这里只对 m_2 和 p 进行分析，可以验证，对于 m_0 和 m_1 ，结论一致。

戈氏检验首先回归 K 阶水准项 ADL (自回归分布滞后) 过程的无限制方程：

$$y_t = \sum_{i=1}^m \alpha_i y_{t-i} + \sum_{i=1}^m \beta_i x_{t-i} + \varepsilon_t$$

其中： ε 为扰动项。再回归水准项 AR (自回归) 过程的限制方程：

$$y_t = \sum_{i=1}^m \alpha_i y_{t-i} + \varepsilon_t$$

再构造 F 统计量： $F(m, N - 2m - 1) = (N - 2m - 1) (RSS_r - RSS_{ur}) / m (RSS_{ur})$ ，检验其显著性，以检验假设“ X 不是 Y 的 Granger 原因”。其中， RSS_{ur} 和 RSS_r 分别代表无

限制方程和限制方程的残差平方和。

进行戈氏因果检验的前提是时间序列必须是平稳的，因此要对变量进行平稳性检验。

从 1996 年 1 月到 2001 年 6 月， m 和 p 共有 66 个数据，基本符合大样本分析的要求。首先利用 X11 技术对 m_2 和 p 进行季节性调整，然后对消除季节性波动后的时间序列进行 ADF 检验 (Augmented Dickey-Fuller Test, 见 Harris, 1995), 检验结果见表 1。

表 1 变量单位根检验结果

	m_2	Δm_2	$\Delta \Delta m_2$	p	Δp
有常数项、有趋势项	- 3.03 (10)	- 3.25 (6)*	- 4.3 (10)**	- 4.85 (1)**	- 4.41 (2)**
有常数项、无趋势项	- 1.48 (9)	- 2.95 (8)**	- 3.96 (10)**	- 4.39 (1)**	- 4.32 (2)**
无常数项、无趋势项	2.74 (9)	- 0.88 (10)	- 4.03 (10)**	0.60 (1)	- 4.47 (2)**

注：① ADF 检验的临界值来自 McKinnon 的计算，在 Eview 软件中直接给出。

② ** 和 * 分别表示在 5% 和 10% 的水平上显著。

③ 括号中的数表示滞后的阶数，滞后的最大阶数根据 Schwert 标准计算： $\text{int} [12 (66/100) 1/4] = 10$ ，实际选择的滞后阶数采用最大化修正后的 R^2 标准。

由表 1 可见， m_2 不是平稳序列； Δm_2 未通过无常数项、无趋势项的检验，但通过了其他两项备选检验； $\Delta \Delta m_2$ 通过了所有三项检验。因此， m_2 可能是一阶单整序列 $I(1)$ ，但一定是二阶单整序列 $I(2)$ 。 p 未通过无常数项、无趋势项的检验，但通过了其他两项备选检验； Δp 通过了所有三项检验。因此， p 可能是平稳序列，一定是一阶单整序列 $I(1)$ 。

可见， $\Delta \Delta m_2$ 和 Δp 一定是平稳的，根据前述程序检验二者之间的戈兰杰因果关系，见表 2。

表 2 $\Delta \Delta m_2$ 与 Δp 的因果关系检验

零假设	滞后阶数 (k)	观察值数量	F (k, N - 2k - 1)	P 值
$H_0: \Delta p$ 不是 $\Delta \Delta m_2$ 的戈兰杰原因	9	55	2.28**	0.038
$H_0': \Delta \Delta m_2$ 不是 Δp 的戈兰杰原因	6	58	2.06*	0.077

注：N=66，为样本总数；** 和 * 分别表示在 5% 和 10% 的水平上显著。

戈兰杰因果关系检验结果说明， H_0 非真，因此我们拒绝 H_0 ，接受 H_1 ，即 Δp 是 $\Delta \Delta m_2$ 的戈兰杰原因； H_0' 也非真，因此我们拒绝 H_0' ，接受 H_1' ，即 $\Delta \Delta m_2$ 是 Δp 的戈兰杰原因。则 $\Delta \Delta m_2$ 和 Δp 是互为戈兰杰因果的关系，进而 m_2 和 p 一定也是互为戈兰杰因

果关系^①，这证实了前面的理论预测。

四、我国货币市场的运行特征与含义

(一) 理论和实证分析表明

(1) 特殊的制度导致了货币市场特殊的“价格 - 国际收支 - 货币供给”动态调节机制。

(2) 货币市场均衡是渐进稳定的；在调整过程中，经济总量呈现周期性振荡，并且图 2 至 5 表明，货币供给和价格水平在向均衡调节的振荡过程中，会经常性地超过稳态水平，存在着超调 (Overshooting) 现象。

(3) 在价格水平 - 货币供给的联合运动中，对于不同口径的货币，由于冲击发生频率、强度以及调整时滞的不同，货币 - 价格存在回型和峰型两种典型运动形态。

(4) 货币供给和价格是互为因果的关系。对于相关的研究而言，这意味着，利用短期数据 (如月份和季度数据) 对货币和价格进行理论建模和实证分析，必须考虑滞后变量的影响，否则会引起模型的设定错误 (Specification Error)。

(二) 货币市场的这种特殊运行特征意味着具有多方面的含义

1. 我国货币市场会遭受经常性的冲击，货币市场失衡是普遍现象

由式 (8) 可知，任何引起货币需求、货币供给的内部和外部冲击都会引起市场均衡总量的变化。由于引起这些冲击的根源是复杂多样和经常性的，我国货币市场也会遭受经常性的冲击。

在式 (1) 的货币需求函数中，影响其趋势的因素主要包括实际收入、价格自由化进程和非国有经济的发展等经济和制度因素。我国持续的经济增长、价格自由化进程的逐步推进和非国有经济的逐步壮大，使得实际货币需求出现增长的趋势。这意味着货币市场将承受持续的正向货币需求冲击。另外，在短期，证券市场交易、预期通货膨胀率、利率和我国 GDP 的季节性波动，从而货币需求也会有经常性的波动。货币市场因此将承受经常性的短期冲击。

货币供给冲击来源于基础货币和货币乘数，前者又主要来源于国际收支的变动。

在国际收支方面，改革开放以来，我国国际收支总体上持续顺差，但是在不同的年份有所不同。尤其是最近十年，我国的外汇储备增长迅速，现在已经超过 2000 亿美元。在结售汇制度下，外汇占款内生化了货币供给，货币市场遭受持续的国际收支冲击。这方面的一个典型例子就是 1997 年末发生的亚洲金融危机。另外，由式 (3)，外汇储备的变动

^① 变量 ΔX 是 ΔY 的戈兰杰原因，可简化表示为： $\Delta Y_t = a\Delta X_{t-j} + bZ_t + \epsilon_t$ ，其中 a 显著不为 0； j 是正整数； Z_t 是其他解释变量。该式经过差分变换，有： $Y_t = Y_{t-1} + aX_{t-j} + aX_{t-j-1} + Z_t + \epsilon_t$ 。这表示滞后的 X 对 Y 有解释力，即 X 也是 Y 的戈兰杰原因。这个结论可以推论到：如果 $\Delta\Delta X$ 和 ΔY 在戈兰杰意义上互为因果，则 X 与 Y 也必然在戈兰杰意义上互为因果。

还受到其他国家价格水平、效率工资市场潜力、经济增长等诸多因素的影响。它们也会影响货币市场。

在货币乘数方面，20世纪90年代中期以后，我国中央银行的独立性增强，金融调控方式实现由直接调控向间接调控的根本性转变；银行商业化管理体制逐渐建立。市场化的经济和金融体制已经初步建立，传统意义上的货币乘数机制开始发挥作用，运用传统分析方法分析货币乘数成为可能。王曦（2002）通过对影响货币乘数各个经济主体微观行为基础的分析，建立了我国货币乘数理论模型：

$$mm = g(\alpha - \delta + i + \beta + r_1 + r_2 - \varphi + r - o)$$

其中： α 为公众对流动性的偏好； δ 为存款流动性系数； β 为商业银行盈利性偏好； φ 为贷款安全性系数； r_1 和 r_2 分别为贷款和债券利率； r 为法定存款准备金比率； o 为其他存款比率。可见，货币乘数受到诸多因素的影响。它并不是一个常数，1996年以后， mm_1 在1~1.5之间有所波动， mm_2 有轻微的上升趋势。我国的货币乘数和国际收支具有季节性波动的特点。

可见，在货币需求和货币供给复杂和持续的冲击下，我国货币市场必然处于动荡之中，货币市场失衡是一个普遍现象。

2. 冲击的作用具有波及性、永久性和持续性

(1) 波及性：在固定利率制度下，货币市场的失衡只能通过商品市场的价格来调节；商品市场价格水平的变动会引起国际贸易条件的变化；国际收支的变动又通过结售汇制引起了货币供给的内生变动。在这个过程中，局部市场的冲击不能在局部消化，必然波及到其他市场，冲击具有波及性。

(2) 永久性：在式(8)中，把货币供给、物价水平和国际收支对解释变量求导，可以得出它们对各种冲击的反映。例如，对于国际收支冲击 dc ，有 $dm^*/dc = dp^*/dc = db^*/dc = 1/(1+\alpha)$ 。这表明，市场的联动对冲击会以有一定的消化，但冲击的效果在各个市场的联动过程中不能被完全消除，冲击的作用是永久的。仍以国际收支冲击为例，根据式(7)， dc 对外汇储备的直接影响是 dc ，但最终影响是 $dc/(1+\alpha)$ ，冲击被消化 $\alpha dc/(1+\alpha)$ 。

(3) 持续性：由微分动态系统可知，在存在价格和货币创造的调整时滞的情况下，就在特殊制度安排下形成了特殊的“价格-国际收支-货币供给”动态调节机制。冲击的调节不会一蹴而就，而是一个渐进的过程。在理论上，这个调节永远不可能完成（见附录的证明）。可见，冲击的作用是动态的和持续的。这也进一步证明，我国货币市场失衡是普遍的现象。

3. 特殊的货币与汇率制度安排导致了低效的市场调节机制

在固定利率制度和特殊的汇率制度下，各个市场的冲击不能在局部消化，必然波及到其他市场；经济转型的过程使得冲击的来源更加广泛，在内部、外部的冲击下，货币-商品-外汇市场承受着持续而复杂的冲击；制度的僵化使得冲击不能被完全消化，最终将反映在未来的经济运行之中；最后，在非均衡向均衡的调节过程中，货币供给和价格水平将会呈现超调。现有制度下的频繁冲击造成了宏观经济不寻常的振荡，引起市场信号的紊乱，扭曲市场在资源配置中调节作用，对经济是十分有害的。

1997年亚洲金融危机之后，竞争国家出口价格下降，由于人民币汇率僵硬，我国贸易顺差减少。在结售汇制度下，货币投放量逐年下滑，1997、1998和1999年，我国 M_2 的增长率分别为17.32%，15.34%和14.74%，这相当于执行了紧缩性货币政策。之后，由于僵硬的利率机制，对外贸易冲击波及其他经济部门，物价增幅逐年下降，1998年之后出现了通货紧缩，1997、1998和1999年，我国的同比商品零售价格指数分别为100.8，97.4和97.1。之后，出口的减少和通货紧缩使得实际GDP增长率也逐年下降，依次为8.6%，7.8%和7.1%，经济紧缩的迹象逐步明显化。我国经济蒙受了巨大的损失。

可见，目前的金融制度及其特殊“价格-国际收支-货币供给”机制在调节冲击方面是低效率的。

4. 货币政策很难有效地调节冲击，甚至也是低效的

理论上，当市场机制不能进行有效调节时，货币当局可以介入其中。但货币当局要能够准确、及时地调整冲击，需要有三个条件：一是能够识别冲击的来源和作用方向；二是能够判断冲击的力度；三是能够根据冲击的方向和力度制定相应的货币政策以消除冲击的影响。在我国，货币当局要满足这三个条件，几乎是不可能的。

首先，我们的分析表明，我国的货币需求和国际收支受到诸多的经济和非经济因素的影响。冲击来源和作用方式的复杂性使得货币当局无法准确识别冲击的根源，超调的存在使得货币当局很难识别冲击作用的方向。因此，货币当局在识别冲击方面是低效的。

其次，要量化冲击的作用也十分困难。尤其是其中的制度性冲击，我们虽然能够在理论上知道制度性冲击的作用方向，但是量化其影响是十分困难的。

最后，货币当局几乎不可能准确地调节冲击。即使货币当局能够准确地识别原始冲击的来源、作用方向和作用力度，并制定出相应的货币政策（例如通过冲销政策）以消除冲击的影响。但货币当局抵消冲击的货币政策实际上又对货币市场强加了一次新的冲击，而新的冲击必然会带来新一轮的振荡和超调。除非货币当局能够准确认识到原始冲击和货币政策冲击的振荡周期和振荡幅度，并能够使货币政策冲击刚好在每一次振荡中恰好抵消了原始冲击（这几乎是不可能的），否则，两次冲击很有可能在某些阶段叠加起来，造成更大的振荡。另外，我国债券市场不发达，也不可能进行大规模的冲销性干预。可见，货币政策在调节冲击时是低效的。

五、结语——改革的方向

在固定利率制和结售汇制的特殊制度背景下，本文进一步考虑我国货币和商品市场的调整时滞，建立了我国货币市场的动态系统理论模型，描述了我国货币市场的运行态势。本文论证了我国价格与货币总量的“超调”现象，分析了冲击作用的特点，指出了制定和实行货币政策的困难性。本文还从理论和实证两个方面论证了货币供给与价格之间的戈兰杰因果关系。

研究表明，特殊制度导致的特殊“价格-国际收支-货币供给”机制是低效的，而货币政策又不能弥补市场机制的不足，这样就只能从根本上改革汇率与利率制度。在汇率和利率中，政策当局至少要市场化其中的一个。

如果改革结售汇制并实行汇率市场化，此时，由于固定利率制，经济中仍然存在价格的调整。但特殊的内生货币供给机制和“价格 - 国际收支 - 货币供给”机制将被消除，也就不存在价格与货币存量的联动。如果存在超额货币供给，价格水平仍将上升，但价格水平也只会单一地沿着图 1 中的物价局部均衡曲线进行调节，“超调”将被消除。类似地，如果实行利率市场化，虽然仍然存在国际收支引起的内生货币供给机制，但货币市场失衡可以通过利率机制进行调节，货币市场失衡对商品市场的价格冲击将被隔绝，“超调”也将消除。当然，理论上，同时实行汇率和利率市场化的效果会更好。^①

可见，实行汇率与利率市场化之后，特殊的“价格 - 国际收支 - 货币供给”机制将被消除，也就不再存在“超调”现象。另外，由于消除了货币与商品市场的联动机制，冲击作用的“经常性”和“波及性”将有所缓解；制定和执行货币政策的困难将减少。消除了“超调”，价格能够更好地发挥对资源配置的指导作用；另外，当局可以更有效地制定并实行货币政策，这都将意味着社会福利的提高。

随着我国刚刚加入 WTO，实行汇率和利率市场化这一改革要求显得更加急迫。可以预见，在加入 WTO 以后，随着我国同世界经济更快速地融为一体，国际收支方面和国内经济的冲击将在更广的范围内、以更高的强度和频率发生，因此，我国的货币需求和外汇供给将会具有更大的波动性，如果不进行市场化改革，我国宏观经济将呈现更频繁和更强烈的振荡。

附录：动态系统的正式分析

利用消元法解式 (9) 和 (10) 的微分方程组，可以得到只含有 m 和 p 的两个二阶微分方程：

$$\begin{aligned} m' + 2\lambda m + \lambda^2(1 + \alpha)m &= \lambda^2(k + c - x + \alpha l) \\ p' + 2\lambda p + \lambda^2(1 + \alpha)p &= \lambda^2(k + c - x + l) \end{aligned}$$

这里 m 和 p 的特征方程是一样的，这取决于动态系统的线性特征。以 r 代表特征根，则 m 和 p 共同的特征方程是：

$$r^2 + 2\lambda r + \lambda^2(1 + \alpha) = 0$$

① 经济学中一般认为，如果局部市场的变量（如商品价格、利率、汇率和工资等）具有充分的灵活性，则在局部市场失衡中，这些变量将担负起调节失衡的主要任务。例如，通常认为，浮动汇率制下，国际收支失衡主要通过汇率变动来进行自动调节；浮动利率制下，货币市场失衡将主要通过利率的变动进行调节，等等。当然，如果我们把整个经济作为一个整体来考虑，各个局部市场总会存在这样或那样的联动机制。但是就本文研究的问题而言，我们已经表明，改革汇率与利率制度，至少会消除本文阐述的市场总量“超调”和市场联动现象。

在以利率机制为基础的货币市场失衡调节中，一个重要的问题就是货币需求的利率敏感性问题。在式 (1) 中，利率对实际货币需求的方向是不定的。这是因为：①我国目前特殊的活期存款利息制度；②央行在调整利率时向来对各种利率同升同降。相关的理论分析可以参见王曦 (2002)。实行利率自由化以后，央行不再能控制所有利率，利率结构也不可能再保持稳定。货币需求的利率敏感性将加强。目前金融制度下我国货币需求与不同利率的关系的理论分析与实证检验，是笔者正在研究的一个课题。

从大方向上看，实行汇率与利率制度改革是必然的趋势，但这里就存在一个孰先孰后的改革进程安排问题。就此，笔者正在进行相关的研究。由于这个问题不是本文的重点，这里不再赘述。

该特征方程的根 r_1 和 r_2 是： $-\lambda \pm \lambda \sqrt{-\alpha}$ 。

可见，该特征方程具有复数根，因此 m 和 p 齐次方程的通解为：

$$m(t) = \exp(-\lambda t) [C_1 \sin(\lambda \sqrt{\alpha} \cdot t) + C_2 \cos(\lambda \sqrt{\alpha} \cdot t)]$$

$$p(t) = \exp(-\lambda t) [C_3 \sin(\lambda \sqrt{\alpha} \cdot t) + C_4 \cos(\lambda \sqrt{\alpha} \cdot t)]$$

其中， C_1 ， C_2 ， C_3 和 C_4 是由初始条件决定的常数。再利用待定系数法，可以解出特解（稳态解）为： $m^* = (k + c - x + a) / (1 + \alpha)$ ； $p^* = (k + c - x + 1) / (1 + \alpha)$ 。则非齐次微分方程的通解为：

$$m(t) = \exp(-\lambda t) [C_1 \sin(\lambda \sqrt{at}) + C_2 \cos(\lambda \sqrt{at})] + m^*$$

$$p(t) = \exp(-\lambda t) [C_3 \sin(\lambda \sqrt{at}) + C_4 \cos(\lambda \sqrt{at})] + p^*$$

联系到本问题的特殊性质， m 和 p 的解有以下特点：

- (1) 由于 $-\lambda < 0$ ，该动态系统是渐进稳定的。
- (2) 由于调整具有指数渐进模式，在理论上，完全的调节永远不可能完成。
- (3) 在向稳态点 (m^*, p^*) 过渡的过程中， $m(t)$ 和 $p(t)$ 将呈现正弦波的振荡形式。
- (4) m 与 p 具有相同的振荡周期 T ， $T = 2\pi\sqrt{\alpha} / (\lambda\alpha)$ 。
- (5) 当外生变量存在连续的冲击时，根据 m 和 p 的解， (m, p) 组合点存在多种运动形态，但以图 4 和图 5 中所描述的回型和峰型两种形态最为典型。这很容易利用计算机来模拟。

参考文献

- [1] 陈平，王曦：《人民币汇率的非均衡分析与汇率制度的宏观效率》，《经济研究》，2002 年第 6 期。
 - [2] 王曦：《经济转型中的货币需求和货币流通速度》，《经济研究》，2001 年第 10 期。
 - [3] 王曦：《中国转型经济总需求分析：微观基础与总量运行》，中山大学博士论文，2002 年。
 - [4] 中国社会科学院经济研究所宏观课题组：《贸易、资本流动与汇率政策》，《经济研究》，1999 年第 9 期。
 - [5] Chen Chunlai, The Location Determinants of Foreign Direct Investment in Developing Countries, Chinese Economies Research Center, School of Economics, The University of Adelaide, Working Paper, 1997.
 - [6] Dunning, J., Multinational Enterprises and the Global Economy, Addison-Wesley, Wokingham, England, 1993.
 - [7] Granger, C. W., Investigating Casual Relations by Econometric Models and Cross-Spectral Method, In Dale J. Poirier (ed.), The Methodology of Econometrics, 1969, Vol.1, pp263 ~ 277.
 - [8] Engle, R. F. and C. W. Granger, Cointegration and Error Correction: Representation, Estimation and Testing, Econometrica, 1987, Vol.55, pp251 ~ 276.
- Harris, R., Using Cointegration Analysis in Econometric Modeling, Prentice Hall, 1995.

中国股市涨跌停绩效的经验分析及政策建议*

陈平 龙华

证券市场微观交易制度的目标可以归结为流动性、有效性、稳定性和透明性。针对各国证券市场，尤其是新兴市场格外关注的稳定性目标，各国当局在微观制度的设计上都制定了各种有助于稳定股票价格波动的措施，其中使用得最多同时也是直接对价格进行限制的两种是涨跌幅限制（Daily Price Limit）和交易暂停（Trading Halt）。

一、对价格限制制度的争议

从制度上对证券价格波动加以直接或间接限制的初衷自然是为了消除价格的异常波动，减少虚拟经济泡沫以及避免恐慌性股市崩溃。但从国外经济学的研究和各国（尤其是经济发达国家）证券市场对价格限制措施的实际情况来看，其作用仍处在被置疑、考证和优化的过程中。学术界和经济界的争论主要分为赞成和反对两派。

赞成者认为价格限制从如下几个方面对稳定市场起作用：①涨跌幅限制能够在市场剧烈变动时将每日价格波动维持在一定幅度内，对制度的制定者而言是一种成本较低的降低不确定性及波动性的方法。②价格限制有助于信用交易制度的实施，将信用交易者的风险控制一定范围内并进而保证结算体系的安全。中国的证券市场目前并不存在信用交易制度，无买空卖空，涨跌停板还不具有这方面的意义。③股价达到限价时，可为非理性交易（跟风、恐慌性抛售）提供回归理性、消化信息的时间，防止恐慌及过度反应。④交易暂停措施有助于各方提防非正常事件的发生，使有关信息有效地传递给市场参与者，有助于达到一个半强有效式市场。⑤有些采用做市商制度的证券市场，做市商既是经纪人，又是自营商。如果市场突然出现买卖失衡，做市商有义务平衡和创造市场。若市况急剧变化，做市商难免无力履行其职责，因而有必要暂停交易。中国的证券市场属于指令驱动系统（区别于报价驱动系统），不存在做市商制度，价格限制对做市商的保护作用尚且谈不上。

对价格限制的反对意见则可以归结为：①价格限制妨碍了市场机制的正常运作，减少了流动性交易者的交易意愿，降低市场流动性，并阻碍均衡价格的形成，导致市场运作无效率。②停市及涨跌幅限制只是对市场走势暂时加以抑制，并不能阻止其运动方向，在信息不明的情况下，市场焦虑的情绪可能不减反增，会加剧市场波动。③一般而言，噪音交易者（Noise Trader）多半采取正向回馈交易策略（Positive Feedback Strategies），即当股价上涨时买进，而在股价下跌时卖出，一般散户会视涨跌停为买卖信号；而主力大户基于这种认识，可能操纵涨跌停来误导噪音交易者。④停牌后，部分人为套现不得不抛售其他股

* 原载：《世界经济》，2003年第2期。

票，易引起连锁反应，停牌股票复牌后，累积多时的抛盘有可能全数涌出，跌幅会更大。

⑤有人从监管者职能考虑，担心监管者一旦作出短暂的干预（交易暂停）后，有可能被迫采取更严厉、更广泛的措施，如长时间停市、停止衍生品种的交易、强迫上市公司回购及直接出资托市等，这将给市场运作带来负面影响。而交易所作为证券市场参与者的一方，无权也没有能力对市场是否理性作出判断，市场在一段时间内的非理性是正常的，交易所不应干预。

二、证券市场价格限制理论与经验分析：文献回顾

20世纪70年代末80年代初，经济学界开始从理论或经验分析的角度来探讨价格限制的微观传导机制。从文献回顾来看，价格限制正面作用的证据似乎不足。

Ma, Rao 和 Sears (1989) 检验了涨跌停对期货和国债市场的影响，发现价格在达到涨跌停后会有反转 (Price Reversal) 现象，且价格波动幅度会降低，直至回到原始水平。Greewald 和 Stein (1991) 认为涨跌停及断路措施有助于信息的传播，能使市场暂时保持平静。他们在1991年的一项研究中发现一个包括市价和限价指令的连续交易系统必然会牺牲一定程度的信息效率，在市场波动不大的情况下，信息效率损失有限；而在市场波动很大的情况下，效率问题会恶化并且扩散，市场断路措施以牺牲一定的即时性为代价，却可以使指令包含更大的信息集。Hopewell 和 Schwartz (1978) 发现限价措施使新信息能有效反映在股价上，但股价会有过度反应。

相比之下，关于价格限制负面作用的证据则相对更为充分。Lee, Ready 和 Seguin (1994) 对纽约证券交易所 (NYSE) 的交易暂停制度进行了研究。他们定义了“假暂停”样本（即在不存在交易暂停制度时，按照若干假设已达到暂停条件而由于制度的原因未受到暂停的约束的样本），并通过与“真暂停”样本比较，得出交易暂停同时增加了价格波动性与交易量，区别在于暂停导致的交易量增加能够延续到暂停日起的第三日，而放大的价格波动在暂停后若干小时内就衰微下去了。King, Pownall 和 Waymire (1991) 对 NYSE 与公司信息披露有关的交易暂停样本进行了经验检验，发现此类暂停的发生通常是由具有很高信息含量的非常规消息引发的，这些消息导致了价格发现的推迟和更加的不确定。Kim 和 Rhee (1997) 用 1989~1992 年的东京证券交易所上市股票数据检验涨跌幅限制对股票市场的影响，检验结果支持涨跌幅限制的反对意见：涨跌停阻碍了股票均衡价格的实现，干扰了交易流动性，并且对价格波动也只是起了暂时的硬性抑制，波动具有外溢性。Lauterbach 和 Ben (1993) 以 1987 年 10 月全球股市崩溃风潮中的以色列特拉维夫股票交易所为研究对象，因为该交易所的市场断路措施是有选择性的——当市场达到某种限制，不是所有的股票都暂停交易。他们对市场断路措施的检验发现：第一，交易暂停和价格限制对股市下跌的程度几乎没有作用，而仅仅是使股市崩溃时期的波动变得平滑一些。第二，有微弱的证据表明，交易暂停有助于减小超额供给的缺口。Subrahmanyam (1994) 将数学引入模型市场断路机制，根据一系列严格的假设，将模型推导从单市场跨期模型扩展到双市场跨期模型，得出：在单市场模型中，“断路”使价格波动性增加；在双市场模型中，一个市场的“断路”会使该市场的价格波动转移到另一个市场，并且，流动性也发生了转

移,所以,“断路”机制是以价格波动的转移和流动性的降低为代价的。

基于分析对象、采样时间和地点的差异、分析方法的区别、影响因素取舍的不同,上述研究难有统一定论,但都不同程度地支持了限制措施具有助长涨跌、干扰交易的作用这一结论。

三、中国股市涨跌停制度绩效经验分析

目前中国的股市在稳定市场方面主要采用了涨跌停板和股票停牌制度,但是,由于触发股票停牌的原因多种多样,在采样、归类上不便统计,所以本文拟对中国股市涨跌停制度绩效进行经验分析,通过对沪市 1997~2000 年的个股数据的测算,采用“事件研究法”(Event Study)^①,对股票价格波动外溢说(Volatility Spillover Hypothesis)、价格发现推迟说(Delayed Price Discovery Hypothesis)和交易干扰说(Trading Interference Hypothesis)进行检验。

(一) 涨跌停数据的甄别与样本集的确定

我们定义涨停样本集为 A,如下所示:

$$A : C_t + 0.005 \geq 1.1C_{t-1} \quad (3-1)^{\textcircled{2}}$$

式中: C_t 为股票 t 日收盘价; C_{t-1} 为股票 t-1 日收盘价。比较样本集 B 为涨幅达到 9% 但不到涨停的样本集,检验样本集 C 为涨幅达到 8% 但不到 9% 的样本集。作为比较样本, B 集中的个股涨幅非常接近 A 样本的涨幅,在对上述三个假说的检验中,通过比较 A、B 集的差异与 B、C 集的差异,可能得到有意义的结论。

$$B : 1.1C_{t-1} - 0.005 > C_t \geq 1.09C_{t-1} \quad (3-2)$$

$$C : 1.09C_{t-1} > C_t \geq 1.08C_{t-1} \quad (3-3)$$

公式(3-1)、(3-2)、(3-3)为价格向上运动的情况,如果价格向下运动,则 A、B、C 的定义分别为:

$$A : C_t - 0.005 \leq 0.9C_{t-1} \quad (3-4)$$

$$B : 0.9C_{t-1} + 0.005 < C_t \leq 0.91C_{t-1} \quad (3-5)$$

$$C : 0.9C_{t-1} < C_t \leq 0.92C_{t-1} \quad (3-6)$$

以沪市 1997~2000 年 4 年间的上市 A 股个股日线数据^③为取样区间(包括 1997 年后在沪市上市的股票)。由于在接下来的检验中使用了 10 日交易链研究方法,所以 2000 年 12 月

① “事件研究法”主要用于分析某一特定事件发生后证券市场价格或交易量等的反应。它最初是法玛、费雪、詹森和罗尔在发表于 1969 年 2 月《国际经济学评论》上的《股票价格对新信息调整》一文中提出来的,后被学术界广泛采用。

② 因为价格涨跌的最小单位为 0.01 元,就存在涨跌幅度不会恰好等于 10% 的情况,所以证券交易所规定,当价格上下波动虽然还不到 10%,但是再涨(跌)多 0.01 元就超过 10% 时,以相差 0.01 元的两个价格中最接近波动 10% 的那个价格为涨跌停价。用方程式表达为 $\frac{C_t + 0.01}{C_{t-1}} - 1.1 \geq 1.1 - \frac{C_t}{C_{t-1}}$, 推导得 $C_t + 0.005 \geq 1.1C_{t-1}$

③ 本文使用的股票价格和交易量的数据,如未另加说明,均下载于 <http://stock.21cn.com/sjxz.html>。

在沪市上市的股票若当年的交易日不足 10 天，则被剔除。因此，有效的股票只数为 572 只。对 A、B、C 集进行统计，结果如表 1 所示。

表 1 样本数据统计结果

价格向上运动			价格向下运动		
	全部	剔除连续涨停		全部	剔除连续跌停
样本集 A	4219	3144	样本集 A	2171	2026
1997	713	571	1997	810	762
1998	677	550	1998	453	442
1999	1356	984	1999	576	527
2000	1473	1039	2000	332	295
样本集 B	(n = 1185)		样本集 B	(n = 0)	
样本集 C	(n = 1610)		样本集 C	(n = 0)	

从表 1 中可以看出：①涨跌停限制更多的是价格的向上运动，因为 4 年中涨停的次数几乎是跌停的两倍。②涨停的样本数基本上逐年递增，与此相反，跌停样本数逐年递减。在涨跌停制度设立之初的 1997 年，跌停的发生甚至比涨停还多，可到了 2000 年，涨停数已经是跌停数的 4.44 倍。这与股市大势的发展相匹配，1997~2000 年沪市上证指数从 919 点上升到 2073 点，总体上处于上升通道。③统计结果中一个出人意料的发现是，在价格向下运动中，样本集 B 和 C 为空。这说明了一个非常重要的事实，就是价格下跌到一定程度（不到 8%，甚至可能是 6%，7%）后，就没有什么挽留的力量了，已经属于恐慌性抛售，直至跌到 10% 由跌停板打住。那么股价在下跌到 10% 之前，有一段跌幅是不连续的，我们知道市场效率除了蕴涵价格的信息含量之外还蕴涵着价格的连续，这给中国股市是否有必要改进价格限制制度增强价格连续性提供了思考空间。

（二）价格波动外溢假说检验

Fama (1989) 指出，如果价格发现的过程被中断，可能导致的后果是波动性的提高。Lehmann (1989) 认为，严重的供求失衡导致了价格运动到涨跌停板，从而将无法实现的交易转移到后继的交易日中。这就是价格波动外溢假说。

为了检验该假说，我们使用 10 日交易数据链，即 $t = -4, -3, -2, -1, 0, 1, 2, 3, 4, 5$ 。其中，第 0 日为样本集 A、B、C 中的股票分别达到定义值的那一天。

令：

$$r_{t,j} = \ln \frac{C_{t,j}}{C_{t-1,j}} \quad (3-7)$$

其中： $C_{t,j}$ 为 j 股票 t 日收盘价。

$$V_{t,j} = (r_{t,j})^2 \quad (3-8)$$

$$V_t = \left(\sum_{j=1}^n V_{t,j} \right) / n$$

对于不同样本集， $n = N_A, N_B, N_C$ 。 (3-9)

以 $V_{t,j}$ 来衡量价格波动，我们对 A、B、C 样本集的每只股票在 10 日交易链中的每一天计算该值，再对各集的每一天计算平均数。如果 A 股票在涨(跌)停日后经历的波动性大于其他两个股票亚类 B 与 C，并且 A、B 的比较结果异于 B、C 集的比较结果，则可以认为价格波动外溢说得到支持。

考虑到价格向下运动的 B、C 集为空，无法与 A 集进行比较和检验，本文只对涨停情况的价格波动外溢进行检验。此外，由于连续涨停的特殊性，即连续涨停的情况，每一天都可视为 $t=0$ ，那么就会出现涨停前的日子也出现涨停，增加统计中 0 日前的价格波动，使结果出现偏差。并且，如果在统计检验中对连续涨停情况不加区分，同一个涨停点既可能 $t=0$ ，也可能 $t=-1$ 或者 $t=1$ ，这样对同一样本有反复使用的情况。确定如何使用连续涨停样本在方法上和编程中都是个现实的困难，因此，检验中将连续涨跌停样本也加以剔除。

对价格波动的比较结果如表 2 所示。

表 2 价格波动外溢假说检验结果

交易日	V_t (A集)	A、B 比较	V_t (B集)	B、C 比较	V_t (C集)
- 4	1.888		1.557		2.155
- 3	1.874		1.801		1.956
- 2	2.005		1.865		2.473
- 1	5.428		2.614		2.865
0	14.580	≫	8.236	≫	6.603
1	3.920		2.361	≫	1.784
2	2.327	>	2.876	>	2.140
3	2.469		2.742	≫	1.861
4	3.038	>	2.558	≫	1.757
5	2.169	>	2.070		1.712

注：①表中三栏 V 值都按计算的原值乘以 10^3 以使结果更加直观，而不影响比较的效果。②为检验上述比较在统计意义上的显著性，本表通过采用 Wilcoxon 统计量来对两个数集进行秩检验的方法来衡量比较结果的显著性。即，以 A 集各股在 1 日的波动幅度和 B 集各股在 1 日的波动幅度为例，对分别包括 N_A 个 $V_{1,jA}$ 值和 N_B 个 $V_{1,jB}$ 值的两个数组进行非参数的秩检验，看接受假设： $V_{1,jA}$ 随机地大于 $V_{1,jB}$ 具有多大的置信水平。如果显著水平为 0.01，表中显示比较结果为 \gg ；如果显著水平为 0.05，表中显示比较结果为 $>$ 。

从表 2 中可以看出：①0 日的 V_0 值，A 显著地大于 B，B 显著地大于 C。这是由 A、B、C 的定义所致。②从 B、C 比较来看，1 至 4 日 B 的 V 值都显著大于 C 的 V 值，可是，对 A 与 B 进行比较，结果却是模糊的。从均值的意义上看，在 1 日和 - 1 日， V_A 都

较大程度地大于 V_B ，可是假设 V_A 随机地大于 V_B 却都通不过显著性检验。③从 1 日后的结果看，2、4、5 日 A 的波动幅度是大于 B 的，但是与 B、C 比较对照没有特别之处，没有理由认为对于涨停样本，价格波动会外溢到 1 日以后。

对于该假说的检验并不很成功，究其原因，最主要的因素就是剔除了连续涨停样本。价格限制导致波动外溢，极端的表现就是连续触及价格限制。如果加入连续涨停样本并以第 1 个涨停日为 0 日，就会增加 1 日为涨停的情况，那么，在 A、B 秩比较中，前者大于后者的显著性会增强，检验的结果会更明确。由于编程的困难，本文未能增加连续涨停样本，这是个遗憾。

综上，在剔除连续涨停样本的情况下，只能在均值的意义上表明股票在经历涨停之前 1 日和后 1 日的波动幅度明显高于未达涨停的股票（从图 1 中可以清楚地看到），但是对于价格波动外溢说的检验结果是不显著的。

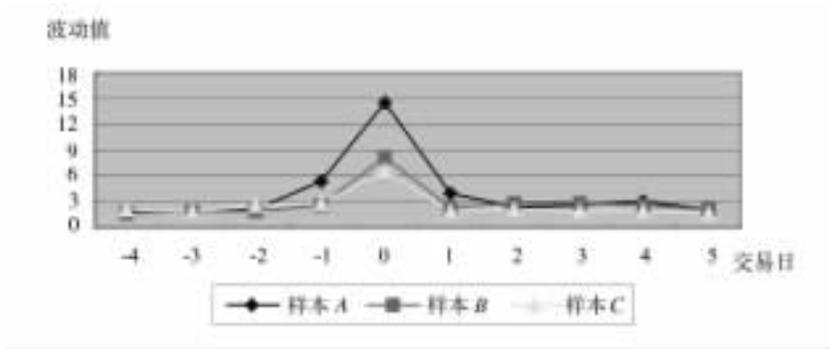


图 1 价格波动外溢假说检验结果

(三) 价格发现推迟假说检验

有学者指出，价格发现被推迟也是涨跌停制度实施的成本之一。当价格运动到涨跌停，交易可能被干扰，从而阻挡了价格在当天运动到均衡水平的可能。因此，达到涨跌停的股票不得等到日后才能继续向其真实价格行进。这种观点即价格发现推迟说。与此矛盾的一种假说是价格反转说，即股票价格在经历单日的剧烈波动后倾向于反向运动。

为验证该假说，对 A、B、C 三种的股票，计算如下两种收益： $r(O_0 C_0)$ 和 $r(C_0 O_1)$ 。

$$r(O_0 C_0) = \ln(C_0 / O_0)$$

$$r(C_0 O_1) = \ln(O_1 / C_0)$$

$r(O_0 C_0)$ 为股票从 0 日开盘到 0 日收盘时的收益； $r(C_0 O_1)$ 为股票从 0 日收盘到 1 日开盘时的收益。

O_t, C_t 分别代表开盘价、收盘价，其下标数字代表第 t 日。由于收益可能为正、负和零，所以上述两种收益有 9 种组合： $[+, +], [+, -], [+, 0], [0, +], [0, -], [0, 0], [-, +], [-, -], [-, 0]$

价格发现推迟说认为涨停股票应经历正的隔夜收益 $r(C_0 O_1)$ ，而跌停股票则经历负的隔夜收益 $r(C_0 O_1)$ 。当然，股票总会经历价格连续或反转，如果每日价格波动对日后价格无影响，那么价格变动是随机的，也就是说，对于在 0 日经历任何价格变化的股票来说，它在第 1 日经历价格连续和价格反转的概率是差不多的。所以，如欲支持价格发现推迟说，我们期望观察到对于涨跌停股票（A 样本），价格连续多于价格反转，并且多于 B、C 样本的价格连续情况。与此同时，B、C 经历类似的价格连续。根据 Roll（1989）的研究，如果价格连续，说明涨跌停的信号阻碍了理性交易或者信息交易。否则，按照过度反应行为假说，价格剧烈波动后更趋向于价格反转。

对于涨停样本， $[+, +]$ 和 $[0, +]$ 为价格连续情况，其中 $[0, +]$ 意味着股票在 0 日以涨停价开盘； $[+, -]$ ， $[0, -]$ ， $[-, +]$ ， $[-, -]$ ， $[-, 0]$ 五种情况为价格反转； $[+, 0]$ 和 $[0, 0]$ 为价格隔夜无变化的情况。相应地，对于跌停样本， $[-, -]$ 和 $[0, -]$ 为价格连续情况，其中 $[0, -]$ 意味着股票在 0 日以跌停价开盘； $[-, +]$ ， $[0, +]$ ， $[+, -]$ ， $[+, +]$ ， $[+, 0]$ 五种情况为价格反转； $[-, 0]$ 和 $[0, 0]$ 为价格隔夜无变化的情况。

对表 1 中剔除连续涨停的价格向上运动的所有样本计算 $r(O_0 C_0)$ 和 $r(C_0 O_1)$ 并进行统计，结果如表 3 所示。

表 3 价格发现推迟假说检验结果一（全部样本统计结果）

价格运动	P_A	P_B	P_C	$P_A - P_B$ (z 值)
价格上涨	1.00 (3144)	1.00 (1185)	1.00 (1610)	
价格连续	0.78 (2462)	0.52 (613)	0.49 (793)	0.26 (29.53)
价格反转	0.15 (462)	0.36 (422)	0.36 (586)	- 0.21 (- 24.89)
价格不变	0.07 (220)	0.13 (150)	0.14 (231)	0.05 (- 10.01)

注：① P_A 表示在样本集 A 中各种价格运动情况所占样本个数的比例，以百分比表示，以此类推。2~4 列括号中的数字表示发生某种价格运动情况的样本数目，第 5 列中括号中的数字为 $P_A - P_B$ 的 z 值。② 由于四舍五入的缘故，分列的概率之和有可能不等于 1。

从表 3 中可以看出：① 即便 A 股票与 B 股票经历了非常接近的上漲幅度，A 发生的价格连续的概率明显地高于 B^①，百分比差距为 26%。与此同时，样本 B 与 C 的价格连续概率相差无几。这意味着价格限制阻碍了价格发现的过程，因此能够支持价格发现推迟假说。② 中国股市的数据并不支持价格过度反应行为说，因为与 78% 的价格连续相比，15% 的价格反转显得比较单薄。再对 B 集与 C 集进行比较，作为价格单日涨幅亦相当高的 B 样本与 C 样本，经历了非常类似的价格反转，均为 36%，而这个数字是 A 样本的两

① 为了验证结果是否显著，本文使用了一种标准的非参数二项检验，即计算 z 统计量来回答如下问题：样本集 A 是否比样本集 B 经历更多的价格连续？ $z = (C_A - P_B N_A) / [(P_B (1 - P_B) N_A)^{1/2}]^{1/5}$ 。 C_A 表示在样本集 A 中分别经历价格连续、价格反转和价格不变的样本个数。如果样本量足够大，z 统计量是服从正态分布的。

倍有余，所以，如果没有涨跌停限制，价格过度反应反而更容易得到纠正。

在表 3 的统计中可能存在的一个问题是，涨停样本 A 是以 0 日的最高成交价收盘的，而样本 B 和 C 在 0 日则有可能不是以最高成交价收盘，即可能存在收盘前价格从波动幅度最高处回落的情形。如果考虑到两种情况提供给交易者的信息可能存在差异，以下将对样本集 A、B 和 C 的股票在最高价收盘的情况重新进行统计，分别以符号 A^* 、 B^* 和 C^* 表示这三种样本。经统计，在价格上涨的情况下， $N_{A^*} = 3053$ ， $N_{B^*} = 261$ ， $N_{C^*} = 247$ 。以样本 A^* 、 B^* 和 C^* 重新对价格发现推迟假说进行检验，结果如表 4 所示。从表 4 中可以看出，以上两点结论得到了更强的支持。

表 4 价格发现推迟假说检验结果二（价格在 0 日最高价收盘的股票样本统计结果）

价格运动	P_A^*	P_B^*	P_C^*	$P_A^* - P_B^*$ (z 值)
价格上涨	1.00 (3053)	1.00 (261)	1.00 (247)	
价格连续	0.79 (2426)	0.51 (132)	0.48 (119)	0.28 (32.77)
价格反转	0.14 (417)	0.36 (93)	0.36 (89)	- 0.22 (- 25.72)
价格不变	0.07 (210)	0.14 (36)	0.16 (39)	- 0.07 (- 11.34)

(四) 交易干扰假说检验

Lauterbach 和 Ben (1993) 认为交易暂停对股票交易流动性的干扰明显是这种制度实施的成本之一。Fama (1989) 认为如果价格限制妨碍了交易，那么当天的股票流动性会降低，从而有可能导致接下来几天交易量的增加。Lehmann (1989) 对此提供了一种替代的解释。他认为交易指令的不平衡引起交易不足，进而导致价格到达其限制幅度，在接下来的几天，缺少耐心的投资者会在并非最理想的价格买进或者卖出，或者有耐心的投资者一直等到价格达到理想水平再交易，于是指令的不平衡得到纠正。两种情形都意味着在涨跌停后的几天，交易量会增加，因此对市场交易造成干扰。

为检验该假说，仍然使用 10 日交易数据链。如欲支持交易干扰说，则期望看到股票在涨跌停日后几天内的交易量明显放大，以比较样本 B 和检验样本 C 作为参照。

为使不同的股票的交易量具有可比性，使用换手率 (Turnover Ratio) 表示交易量，其定义如下：

$$\text{换手率} = \text{股票 } j \text{ 第 } t \text{ 日交易量} / \text{股票 } j \text{ 第 } t \text{ 日流通总股数}$$

用符号表示为：

$$TR_{t,j} = TVOL_{t,j} / SOUT_{t,j} \quad (3-10)$$

其中， $TR_{t,j}$ 为股票 j 第 t 日的换手率 (Turnover Ratio)； $TVOL_{t,j}$ 为股票 j 第 t 日的交易量 (Trading Volume)； $SOUT_{t,j}$ 为股票 j 第 t 日的流通盘 (Shares Outstanding)。

交易量的日变化率用公式表达为：

$$TC_{t,j} = \ln(TR_{t,j} / TR_{t-1,j}) \times 100 \quad (3-11)$$

由于公共来源中的股票 K 线数据不包含换手率，只有成交量与成交金额，又因为 $\frac{TR_{t,j}}{TR_{t-1,j}}$

$$= \frac{TVOL_{t,j}/SOUT_{t,j}}{TVOL_{t-1,j}/SOUT_{t-1,j}} = \frac{TVOL_{t,j}}{SOUT_{t-1,j}} \cdot \frac{SOUT_{t-1,j}}{SOUT_{t,j}}$$
 而 $\frac{SOUT_{t-1,j}}{SOUT_{t,j}}$ 表示股票 j 相连两天的流通盘之比, 我们知道除了 t 日为股票的除权交易日时前一日与当日的流通盘之比不为 1 之外, 相连两天的流通盘大小是一样的, 所以有:

$$\begin{aligned}
 TC_{t,j} &= \ln(TR_{t,j}/TR_{t-1,j}) \times 100 \\
 &\approx \ln(TVOL_{t,j}/TVOL_{t-1,j}) \times 100 \quad (3-12)
 \end{aligned}$$

除非当选取的样本的 10 日中恰好包括除权交易日, 否则这种近似丝毫不影响结果。根据经验, 一只股票一年中的除权交易日非常有限, 甚至可能一年中不发生任何送、配股和增发, 因此, 10 日链中包含除权交易日发生的概率并不大。^①

我们剔除连续涨停样本, 并且仅对价格向上运动的样本做分析。对于样本 A 在 10 日交易链中的每一天, 计算 $\sum_{j=1}^n TC_{t,j}/n$ ($n = N_A$; $t = -4, -3, -2, -1, 0, 1, 2, 3, 4, 5$), 得出涨停日 (0 日) 前 4 日后 5 日交易变化率的平均值。类似地, 对于样本 B 与 C, 亦可计算 10 日交易链中的交易量变化率, 结果如表 5 所示。

表 3-5 交易干扰假说检验结果

交易日	样本 A (涨停 %)	A、B 比较	样本 B (上涨 %)	B、C 比较	样本 C (上涨 %)
- 4	2.084		- 1.033		- 1.698
- 3	2.379		4.458		3.612
- 2	12.062		8.511		9.140
- 1	13.717		11.013		11.262
0	80.427		79.930		74.743
1	12.728	≫	- 10.971	>	- 15.115
2	- 52.338	≪	- 33.183	<	- 29.095
3	- 13.675	≪	- 8.070		- 11.396
4	- 13.287		- 11.812		- 11.537
5	- 8.674		- 7.773		- 7.106

注: 为检验上述比较在统计意义上的显著性, 本表通过采用 Wilcoxon 统计量来对两个数集进行秩检验的方法来衡量比较结果的显著性。即以 A 集各股在 1 日的交易量变化率和 B 集各股在 1 日的交易量变化率为例, 对分别包括 N_A 个 $TC_{1,jA}$ 值和 N_B 个 $TC_{1,jB}$ 值的两个数组进行非参数的秩检验, 看接受假设: $TC_{1,jA}$ 随机地大于 $TC_{1,jB}$ 具有多大的置信水平。对于显著水平为 0.01 和 0.05, 表中分别显示比较结果为 \gg 和 $>$ 或 \ll 和 $<$ 。

从表 5 中可以看出: ① 0 日三种股票的交易量都比前一日激增, 分别从 13.72%, 11.01%, 11.26% 的增加率上升到 80.43%, 79.93%, 74.74%。而且, 作为涨幅最为接近的样本 A 和样本 B, 交易量的增加非常接近, 超过 B、C 的接近程度。② 在 1 日, 样本 A

① 必须补充说明, 在除权日, 某些市场主体为了填权, 有可能在除权当日拉升股价, 所以除权日股票价格上涨幅度较高是可能的, 甚至可能达到涨停。我们在对样本做随机抽样, 检验样本 10 日链中包含除权日的可能性时发现, 在随机摘取的 20 个涨停样本中, 仅有一个样本在 10 日链中包含除权日。可想而知, 简化过程会导致一旦除权日发生在 10 日链中, 计算的当日交易量变化率会比实际的数值大, 但由于样本量足够大, 经过平均后这种误差可以忽略, 所以有理由认为这种简化是可行的。

的交易量继续显著上升 (12.73%)，与此同时，B 和 C 的交易量则显著下降 (分别为 -10.97% 和 -15.12%) 的。样本 A 和 B 在 0 日经历了非常接近的涨幅和交易量增加率，但是没有受到价格限制的样本 B 和 C 在 1 日的交易量明显下降说明大部分 B、C 股票的交易者在 0 日都能建立合意的股票头寸。相比之下，由于涨停对供求运动平衡的阻碍作用，部分 A 样本股票的交易者不得不等到 1 日才能在合意的价格进行交易。^③从 1 日后的交易量变化率来看，A、B、C 三种股票的交易量逐日减少，其中以 A 的下降速度最快，在第 2 日下降了 52.34%，远高于 B、C 的 33.18% 和 29.10%。这说明涨停对合意交易的干扰基本上只是导致交易外溢到第 1 日。

(五) 结论

根据以上分析，对于涨停股票，价格波动在涨停前 1 日和后 1 日的波动幅度相对样本 B 和 C 更大，但是由于剔除连续涨停样本，这种结果不显著 (价格波动外溢说)；价格连续发生的概率显著地高于其他两类样本，从而推迟了价格发现的时间 (价格发现推迟说)；相对于其他两类样本，交易量在 0 日后收缩，涨停样本交易量发生了外溢，对 0 日合意的交易形成了干扰 (交易干扰说)。对中国股市涨跌停制度效率的质疑基本上得到支持。

四、中国股票市场的特征

(一) 市场信息分布特征

经过 13 年的成长，我国的股市仍处于初期阶段，信息的完整性、分布均匀性和时效性与发达国家相比还存在较大差距。我国公开发表的信息比较有限，特别是企业信息披露的规范程度还很低。股市中由信息操纵而实现市场操纵 (或称价格操纵) 的行为普遍存在。从中国的现实情况看，没有理由认为信息缺陷导致的市场预期趋向一致而实现股价涨停或跌停，会在价格限制再次解除前实现信息的完整、准确和对市场的充分渗透，这从控制信息的主体的动机和中国股市有关信息传播的制度条件和媒介条件来看都不大可能实现。

(二) 个人理性、总体理性与现实的距离

对所有投资者都是理性的假设对于任何国家的资本市场来说都过于简化。个人理性首先意味着所有投资者都知道正确的定价模型。在 EMH 内在的线性关系下，投资者对信息的解读包括对符号的解读和对量的解读，这一条件不要说在中国这样一个新兴的证券市场，就是在最成熟的发达国家证券市场，也是不成立的。其次，个人理性还包含着投资者独立做出决策的假定，但在现实中，投资者往往出于对其他投资者的行动的估计而决策。中国股市的机构化程度比发达国家的股票市场要低得多，机构投资者太少，散户占绝大多数。在存在信息缺陷的情况下，市场散户认为机构投资者掌握比他们更多的信息，甚至可能认为市场的大多数投资者比自己聪明，所以其投资策略很容易“随大流”。这种“模仿传染机制”描述了中国股市普遍存在的“从众行为” (Herd Behavior)。

当基于噪音而不是信息的交易占据市场的优势地位导致价格达到涨跌停，过多的噪音交易反而会减少股票价格中的信息含量。在这种市场条件下，如果有信息交易者（通常是机构投资者和资金大户）基于自己的理性判断对所持有的股票予以变现时，首先要承担噪音交易者带来的风险，即占据市场主导地位的噪声交易使价格进一步偏离内在价值，而不会像理性交易者所预期的那样发生逆转；其次，中国股市中的信息交易者在可能的情况下往往对信息采取垄断态度，即并不采取有利于增加股票价格信息含量的交易，反而基于对噪音交易者的估计，采取助长噪音的交易，以达到谋取暴利的目的。市场操纵导致了市场总体理性水平的下降。信息所有者的这种策略使得涨跌停除了有限的一段时间之外很难提供别的东西使投资者向“优秀的分析家”蜕变，而交易者的狂热或沮丧往往会由于涨停或跌停而加剧。

（三）涨跌停的临界趋势效应

在中国的股市，投资者无法搜集所有甚至是基本的信息，或许也不知道如何解读所有已知的信息，所以资本市场混沌理论认为的人们更可能对趋势作出反应可能更为贴近中国现实。按照这一观点，人们可能在收到信息时并不对其作出反应，待趋势变化得到证实后，再采取行动。涨跌停很可能被视为证实某种趋势的临界变化值，这种强烈的临界趋势效应导致人们怀有“价格尚未涨够或跌够”的想法，从而导致价格进一步顺应趋势。具体地说，涨停的临界趋势效应的主要作用是使卖盘不适当地减少。因此，只要市场主力能够或敢于封住涨停板，卖盘会大幅度减少。不仅如此，在个股出现过涨停板以后，股价的第一、二次回调总会吸引大批买盘进场抢筹，这无形之中就为市场主力操纵股价提供了方便。跌停的临界趋势效应则使买盘迅速萎缩，使人无法再行卖出，并且还会对明天的走势产生强烈的负面影响。卖盘无法在当天兑现，便会集中到第二天去，这样，跌停起到了助跌作用。

五、政策思考

从中国股市的现实情况和本文的分析结果看，中国股市的价格限制制度有待改进。当然，在一个新兴的、市场规模偏小、散户众多、投资者欠成熟、市场操纵行为盛行的证券市场，政策当局的干预是不可缺少的。

（一）放宽涨跌限制幅度

放宽涨跌限制幅度的原因可归结为：

（1）对涨跌幅进行限制的目的在于减缓市场价格波动的剧烈程度，尤其在投机过度和交易过热的市况下。可是，从本文经验分析的结果看，现行 10% 的涨跌幅既不能挽住价格的走势（涨跌停日后价格更倾向于继续涨跌），还降低了市场效率（涨停的实施是以推迟价格发现和干扰交易为代价的）。中国股市中存在大量流通市值较小的股票，尽管与散户的资金总量比，机构投资者的资金量占市场投资总额的比重较小，但机构投资者却能够比较容易地实现对小盘股的价格操纵，将价格拉到涨跌停，通过临界趋势效应强化对价格

走势的误导。所以，放宽涨跌限制幅度从某种意义上说能够增加市场操纵的难度，从而有利于市场发展。

(2) 由于投资者普遍存在风险厌恶，即从心理上对相同程度的上涨和下跌的承受能力是不一样的，处于套牢盘中的投资者倾向于固守潜在损失，这在一定程度上能够阻止价格下跌。可是，分析表明，股票跌至跌停前几乎没有过渡区域，价格失去一定的连续性。从本文结果看，至少是从 8% 的跌幅起再往下跌就完全属于恐慌性的了。价格变化不连续是市场不理性的比较极端的表现。那么，将跌停幅度放大，从短期的意义看，可以加大市场操纵困难，从政策暗示的作用上增强投资者的心理承受能力，从而有可能增强价格连续性；从长期的意义看，可提请投资者认识股票市场的风险，有助于市场的机构化发展。与世界各国相比，我国股票市场的机构化程度还很低，散户投资比例占 90% 以上。在成熟的股票市场，尤其是在没有涨跌限制的股票市场，散户资金的集中程度要高得多。从市场理性化发展的长远看，机构化是必要的，它能够增强市场整体理性和价格连续性。

(3) 从横向比较来看，我国 10% 的涨幅限制在国际上已属偏小。按照股票的价格高低设定涨跌幅金额限制的，换算成百分比的话，则涨跌幅从最低的 6.67% 到最高的 30% 不等，并且个股价格越高，可涨跌的百分比幅度越低。此外，韩国为 15%，泰国为 30%，吉隆坡单日最大涨幅可能达到 69%、跌幅达 51%，^① 菲律宾为涨幅 50%、跌幅 40%，在法国巴黎连续交易主板市场单日最大涨幅可达到 21.25%，跌幅最大能达到 18.75%。

(4) 中国证券市场对风险较高的股票已实行特别处理，在一定程度上减小了价格波动过于剧烈的隐患。所以，对原定 PT、ST 股票 5% 的涨跌幅限制可维持不变。

综上所述，本文建议将涨跌停幅度放宽，并且从价格连续性考虑，可区分上下午盘，上午盘的涨跌幅为前一日收盘价的 10%，下午盘的涨跌幅为上午盘收盘价的 10%，这样一来，全日最大涨跌幅分别为前一日的 21% 和 19%，从单日看，具有轻微的非对称性。

(二) 改进交易暂停制度

交易暂停是一个较为笼统的概念，既包括个股的交易暂停，亦包括股票板块甚至整个市场的交易暂停。我们知道，中国股市的股票交易停牌制度仅仅是针对个股而言，并且不能在盘中暂停交易，而只能在收市后，停牌时间的长短也没有具体规定，取决于上市公司何时公告。只在收市后才使用的交易暂停谈不上是完整意义的交易暂停，因为个股如果在盘中发生异常情况就可能缓不应急，所以应该在加强对股市交易情况进行实时监控的规则条件和技术条件下，考虑在盘中进行个股交易暂停。交易所不对停牌股票恢复交易公告的信息质量予以审查，就不排除敷衍、搪塞的现象，所以应考虑增加对公告信息进行实质性审查的措施。

中国的交易停牌制度设计中没有考虑到对整个市场实施暂停，即“市场断路”措施。在对个股涨跌幅限制放宽之后，股票指数单日的波动幅度也相应加大了，而随着股市市值的日益扩大，引入市场断路机制具有一定的必要性。尽管中国股市最近几年未曾经历过恐

^① 因吉隆坡交易所的交易分上下午盘，上午盘的涨跌幅为前一日收盘价的 30%，下午盘的涨跌幅为上午盘收盘价的 30%。

慌性下跌的浪潮，但防患于未然还是必要的。考虑到符号经济日益脱离实质经济的运动频率和幅度、两者之间的互动性的矛盾，以及随着整个经济的开放程度的扩大输入股灾的可能，我国对股市运行趋势应保持警惕。1987年的全球股市风潮就是一个例子。1987年10月纽约股票市场发生了历史上最大的一次崩盘事件并最终酿成全球股灾。也就是在那之后，国外一些证券交易所才开始实施市场交易暂停。

参考文献

- [1] Fama, Eugene, Business Conditions and Expected Returns on Stocks and Bonds, *Journal of Financial Economics*, 1989, Vol.25, pp23 ~ 49.
- [2] Greenwald, Bruce C., and Jeremy C. Stein, Transactional Risk, Market Crashes, and The Role of Circuit Breakers, *Journal of Business*, 1991, Vol.64, pp443 ~ 462.
- [3] Hopewell, Michael H., and Arthur L. Schwartz, Temporary Trading Suspensions in Individual NYSE Securities, *Journal of Finance*, 1978, Vol.33, pp1355 ~ 1373.
- [4] Kim, Kenneth A., and S. Ghon Rhee, Price Limit Performance: Evidence from the Tokyo Stock Exchange, *Journal of Finance*, 1997, Vol.52, pp885 ~ 901
- [5] King, Ronald, Grace Pownall, and Gregory Waymire, Corporate Disclosure and Price Discovery Associated with NYSE Temporary Trading Halts, *Contemporary Accounting Research*, 1991, Vol.8, pp509 ~ 531.
- [6] Lauterbach, B., and U. Ben-Zion, Stock Market Crashes and the Performance of Circuit Breakers: Empirical Evidence, *Journal of Finance*, 1993, Vol.48, pp1909 ~ 1925.
- [7] Lee, C., J. Ready, and J. Seguin, Volume, Volatility, and New York Stock Exchange Trading Halts, *Journal of Finance*, 1994, Vol.49, pp183 ~ 214.
- [8] Lehmann, Bruce N., Commentary: Volatility, Price Resolution, and the Effectiveness of Price Limits, *Journal of Financial Services Research*, 1989, Vol.3, pp205 ~ 209.
- [9] Ma, Rao, and Sears, Limit Moves and Price Resolution: The Case of the Treasury Bond Futures Markets, *Journal of Futures Markets*, 1989, Vol.9, No.4, pp321 ~ 335.
- [10] Roll, R., Price Volatility, International Market Links, and Their Implications for Regulatory Policies, *Journal of Financial Services Research*, 1989, Vol.3, pp211 ~ 246.
- [11] Subrahmanyam, Avandhar, Circuit Breakers and Market Volatility: A Theoretical Perspective, *Journal of Finance*, 1994, Vol.49, pp237 ~ 254.

养老金投资规模、投资者保护与 股票市场发展*

——论经合组织国家的经验与中国的选择

李 涛 李 红

一、引言

当今世界范围内，尤其是在发达国家，养老金规模都非常庞大，如经合组织国家(OECD)金融资产的相当部分都是由养老金持有的，其中美国的养老金持有该国35.1%的金融资产，而在英国这个比例高达40.1% (Becht et al., 2002)。如何通过投资实现养老金增值，是世界各国政府以及私营养老金机构面临的一个主要问题。以往对养老金投资的研究主要集中在养老金购买股票的超额收益，如Wahal (1996)和Del Guercio and Hawkins (1999)等，以及养老金投资的“羊群行为”方面，如Stein (2001)和Scharfstein and Stein (1990)等。然而，他们忽视了另外两个有关养老金投资的重要问题，即养老金投资规模对投资者保护程度和股票市场发展水平的影响。

首先，养老金购买上市公司的股票或债券后，能够成为一个强而有力的机构投资者，对投资者保护程度产生影响 (Shleifer and Vishny, 1997; Becht et al., 2002)，我们称之为养老金投资的“投资者保护论”。一方面，作为机构投资者的养老金可以提高投资者保护程度。这表现为：①通过持有相当比例的公司股票或债券，机构投资者能够有效地保护自身的利益。Shleifer and Vishny (1997)指出，在缺乏足够的法律保障投资者利益的情况下，选择成为大的投资者从而获得对公司的实质控制权可以避免其利益受损。与一般的中小投资者相比，机构投资者较高的股权或债权比例使其更加积极主动地搜集公司相关信息并监督经理人员，从而减轻了公司内部人基于信息不对称所做出的自利行为。机构投资者还可以投票任免公司管理人员，对公司管理层实施实质性约束 (Shleifer and Vishny, 1986)。②机构投资者还可以保护中小投资者利益。机构投资者对公司经理人员积极有效的监督所带来的公司业绩改善也会使中小投资者获益，而中小投资者在监督内部人时往往会“搭便车”，其监督成本较低，因此比较机构投资者与中小投资者得自监督公司经营管理的净收益，有可能后者更高。此外，LLSV^① (2002)指出，由于大股东对中小股东的侵犯是有成本的，因此随着大股东现金收益权的增加，它们与中小股东的利益会趋于一致。机构投资者在保护自身的同时，也保护了中小投资者。另一方面，机构投资者可能侵犯中小投资者利益，降低投资者保护程度。Shleifer and Vishny (1997)认为，当机构投资者的投票权超过其现金

* 感谢中国人民大学经济学院陈享光教授和张宇教授的宝贵建议。原载：《金融研究》，2004年第5期。

① LLSV是四位作者的名字缩写，他们分别是La Porta, Lopez-de-Silanes, Shleifer, Vishny。下同。

收益权时，他们会采取一些利己但损害中小投资者利益的行为，如大量派发特别股息，或者以低价与其控制的其他公司进行关联交易。以意大利为例，Zingales (1994) 发现享有优先投票权的股票能产生超额回报。由于这些股票往往由机构投资者持有，这说明了机构投资者可能通过侵犯中小投资者而获得超额收益。

其次，养老金购买上市公司股票，也会影响股票市场发展水平，我们称之为养老金投资的“股市发展论”。作为机构投资者的养老金对股票市场发展的推动作用可以体现在它的下列功能上：清算结算、积累资金、转移资源、风险管理、风险控制、金融创新以及激励基金经理人 (Bodie, 1990; Davis, 1996; Vitas, 1998)。以七国集团 (G7) 为例，Davis (2000) 详细探讨了包括养老金在内的机构投资者对金融发展以及金融稳定的作用，发现机构投资者规模对一国的金融资产总规模有着积极影响。

为了弥补以往养老金投资研究的不足，本文考察了养老金投资规模对投资者保护程度和股票市场发展水平的影响。以经合组织国家为例，我们发现，养老金投资规模对投资者保护程度没有显著的影响，而对股票市场发展水平有着显著的推动作用。控制了其他制度性因素，实证结果相当稳定。这支持了养老金投资的“股市发展论”，而否定了养老金投资的“投资者保护论”。针对中国社保基金入市问题，我们认为需要澄清两点认识：①从大力推动股票市场发展角度分析，有关机构应当积极推动社保基金入市，这有助于扩大中国股票市场规模，增强其流动性；②社保基金入市不应当承担加强市场上投资者保护的责任，如何更好地保护包括社保基金在内的广大投资者利益，这需要相关的立法和监管机构加以研究解决。

本文的结构如下：第二节介绍了我们使用的数据^①；第三节和第四节分别检验了养老金投资规模对投资者保护程度和股票市场发展水平的影响；第五节讨论了中国的社保基金入市问题；最后一节做了总结。

二、数据样本

本文的数据样本包括以下五个部分：

第一，经合组织国家的养老金投资规模数据来自 OECD (2003) 提供的有关经合组织 17 个成员国在 1992 年的养老金投资规模统计数据。相应变量记作 PENSION1992，等于 1992 年各国养老金持有的金融资产与其国内生产总值之比^②。

第二，经合组织国家的投资者保护数据来自 LLSV (1998) 和 Beck et al. (2000)。由于投资者包括股东和债权人，因此投资者保护程度指标也包括股东保护程度指标和债权人保护程度指标。①股东保护程度：数据来自 LLSV (1998)，包括两个指标：“一股一票制”和“反经理人权力”。前者记作 OSOV，如果一国的公司法或商法规定，上市公司普通股的投票权是一股一票，赋值为 1，反之为 0。一股一票制的实行有助于保护投资者利益。样本中卢森堡的 OSOV 为缺省值，样本观测值为 16。后者记作 ANTIDI，是一个测量股东

① 有关回归变量的相关分析结果，感兴趣的读者可以向作者索取。

② 需要指出的是，为了统一计数单位，我们将本文中所有涉及比例形式的变量都以小数表示。

保护程度的综合性指标,赋值范围在 0 至 6 之间,赋值越大,股东保护程度就越高。样本中冰岛和卢森堡的 ANTIDI 为缺省值,样本观测值为 15。②债权人保护程度:数据来自 LLSV (1998) 和 Beck et al. (2000),包括两个指标:“LLSV 的债权人权力”和“Beck 的债权人权力”。前者记作 LLSVCR,是 LLSV (1998) 根据世界各国公司法或商法的有关规定构造的测量债权人保护程度的综合性指标,赋值范围在 0 至 4 之间。后者记作 LLBCR,是 Beck et al. (2000) 在 LLSV (1998) 基础上,重新设计的一个测量债权人保护程度的综合性指标,赋值范围在 -2 至 1 之间。两个指标的赋值越大,债权人保护程度就越高。样本中冰岛、卢森堡以及挪威的 LLSVCR 和 LLBCR 均为缺省值,样本观测值都是 14。

第三,经合组织国家的股票市场发展水平数据来自世界银行 1999 年完成的一项有关世界各国金融发展与金融结构的调查研究。与 Beck et al. (1999) 一致,在测量股票市场发展水平时,我们考虑了三个指标:①股票市场规模:记作 SMC GDP,等于 1993~1994 年间一国股票市场总市值与国内生产总值之比的平均值;②股票市场流动性,记作 SMVGDP,等于 1993~1994 年间一国股票市场总交易额与国内生产总值之比的平均值;③股票市场效率,记作 TURNOVER,等于 1993~1994 年一国股票市场总交易额与总市值之比的平均值。以上变量的观测值都是 17。

第四,考虑到投资者保护程度和股票市场发展水平还可能受到其他因素的影响,我们在分析中引入了以下控制变量:①经济发展水平:记作 GDPP,等于 1993~1994 年间各国以 1995 年美元不变价计算的人均国内生产总值的自然对数值的平均值。数据来自 World Bank (2001),样本观测值为 17。经济发展水平可能影响到投资者保护程度 (LLSV, 1998) 和股票市场发展水平 (Beck et al., 2000)。②法律体系:我们将一国的法律体系划分为普通法、法国大陆法、德国大陆法和斯堪的那维亚大陆法等四种,分别记作 ENGLISH、FRENCH、GERMAN 和 SCANDINAVIAN。如果一国的法律体系属于其中的某一种,赋值为 1,反之为 0。数据来自 LLSV (1999),样本观测值都为 17。一国的法律体系也会对其投资者保护程度 (LLSV, 1998) 和股票市场发展水平 (Beck et al., 2003) 产生影响。③文化背景:在测量以宗教信仰为代表的文化背景时,我们采用了各国信仰基督教、天主教、穆斯林教以及其他宗教(包括无宗教信仰)的人口比例,分别记作 PROTESTANT、CATHOLIC、MUSLIM 和 OTHER。数据来自 LLSV (1999),样本观测值都是 17。一国的文化背景也会对其投资者保护程度 (Stulz and Williamson, 2003) 和股票市场发展水平 (Beck et al., 2003) 产生影响。④法律执行程度:根据 LLSV (1998) 和 Beck et al. (2000),我们引入了两组不同的法律执行程度指标,即 LLSVEN 和 LLBEN。二者都是反映一国法律执行程度的综合性指标,赋值范围都在 0 至 10 之间,赋值越大,法律执行程度越高。样本中冰岛、卢森堡以及挪威的 LLSVEN 和 LLBEN 都为缺省值,样本观测值为 14。一国的法律执行程度也会影响投资者保护程度 (LLSV, 1998) 和股票市场发展水平 (Beck et al., 2000)。⑤会计标准:记作 ACCOUNTING,等于国际财务分析与研究中心发布的《国际会计与审计趋势报告》就 1990 年各国公司的年度报告中 90 个会计细目所做的包含或忽略的会计审计处理方法的评估结果。数据来自于 LLSV (1998),赋值越大,会计标准越严格。样本中冰岛、卢森堡以及挪威的 ACCOUNTING 为缺省值,样本观测值为 14。会计标准也会影响投资者保护程度 (LLSV, 1998) 和股票市场发展水平 (Beck et al., 2000)。

最后，在投资者保护程度分析中，考虑到养老金投资规模的内生性可能，即投资者保护程度可能影响养老金投资规模（Beck and Levine, 2003），我们引入了劳动力人口比例作为工具变量。这是因为：养老金投资规模会受到该国劳动力人口比例的影响（科林·吉列恩等，2002）；而投资者保护程度和劳动力人口比例之间没有必然的联系。相应变量记作 LABORRATIO，等于 1992 年一国劳动力人口在总人口中的比重，数据来自 World Bank (2001)，样本观测值为 17。

三、养老金投资规模对投资者保护程度的影响

这一节中，我们首先讨论了检验养老金投资的“投资者保护论”的回归模型，然后分析了有关的实证结果。

（一）回归模型

$$OSOVI_i (\text{ANTIDI}_i, \text{LLSVCR}_i, \text{LLBCR}_i) = \alpha_0 + \alpha_1 * \text{PENSION1992}_i (\text{LABORRATIO}_i) + \alpha_2 * \text{GDPP}_i + f (\text{LLSVEN}_i (\text{LLBEN}_i), \text{ACCOUNTING}_i, \text{FRENCH}_i, \text{GERMAN}_i, \text{SCANDINAVIAN}_i, \text{CATHOLIC}_i, \text{MUSLIM}_i, \text{OTHER}_i) + \epsilon_i$$

其中，下标 i 表示不同的国家； $f()$ 表示回归分析中控制变量的组合； ϵ 是残差项。在分析中，我们分别使用了 OSOV 和 ANTIDI 来测量股东保护程度，LLSVCR 和 LLBCR 来测量债权人保护程度。考虑到 PENSION1992 的内生性可能，我们引入了 LABORRATIO 作为工具变量，并使用了工具变量回归方法。由于 LLSVEN 和 LLBEN 高度正相关，在回归分析中我们对其分别使用。考虑到共线性问题，分析中去掉了 ENGLISH 和 PROTESTANT。

（一）实证结果

表 1 给出了有关养老金投资规模对投资者保护程度影响的实证结果，总结如下。

首先，观察养老金投资规模对股东保护程度的影响，我们发现：控制了养老金投资规模内生性可能，不论是以一股一票制还是反经理人权力作为股东保护程度的指标，养老金投资规模对于股东保护程度都没有显著影响。可能的解释是，作为机构投资者的养老金，一方面会积极主动地收集公司信息，监督公司管理人员，减轻由于信息不对称导致的委托代理问题，防止损害股东利益的内部人行为产生（Shleifer and Vishny, 1997），增强了股东保护程度；而另一方面也可能会利用对公司的实质控制权来侵犯中小股东的利益（Shleifer and Vishny, 1997），降低了股东保护程度。综合正反两方面的影响，二者效果可能接近，因而养老金投资规模不会显著影响股东保护程度。此外，在以反经理人权力测量股东保护程度时，与 LLSV (1998) 一致，与普通法国家相比，法国大陆法国家的股东保护程度较低，如回归式 (3) 和 (4)，可能是由于普通法强调保护个体利益，而法国大陆法强调保护政府利益，容易出现机构投资者“俘获”监管机构或者二者串谋，损害中小股东利益的情况（Beck et al., 2002）。最后，在信仰其他宗教或无宗教信仰人口比例较高的国家，以反经理人权力测量的股东保护程度较基督教国家为高。

表 1 以投资者保护程度为因变量的工具变量回归结果

	股东保护程度				债权人保护程度			
	OSOV		ANTIDI		LLSVCR		LLBCR	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PENSION1992	0.039 (0.06)	0.124 (0.22)	5.228 (1.86)	5.108 (1.95)	- 1.218 (- 0.43)	- 1.535 (- 0.55)	- 1.550 (- 1.85)	- 1.587 (- 1.80)
LLSVEN	- 0.141 (- 0.73)		0.735 (0.79)		- 0.446 (- 0.36)		- 0.369 (- 0.65)	
LLBEN		- 0.268 (- 1.16)		1.123 (1.15)		- 0.355 (- 0.22)		- 0.482 (- 0.75)
ACCOUNTING	- 0.006 (- 1.09)	- 0.006 (- 1.05)	- 0.059 (- 1.86)	- 0.059 (- 2.10)	- 0.005 (- 0.13)	- 0.005 (- 0.13)	- 0.024 (- 1.27)	- 0.024 (- 1.36)
FRENCH	0.126 (0.42)	0.079 (0.27)	- 3.880** (- 3.35)	- 3.787** (- 3.47)	0.696 (0.50)	0.823 (0.62)	- 0.305 (- 0.51)	- 0.305 (- 0.56)
GERMAN	0.209 (1.02)	0.134 (0.64)	- 2.771 (- 1.68)	- 2.530 (- 1.58)	1.129 (0.70)	1.164 (0.69)	0.904 (1.10)	0.828 (0.99)
SCANDINAVIAN	0.143 (0.26)	0.229 (0.50)	1.351 (- 0.39)	1.130 (0.35)	0.800 (0.27)	0.665 (0.24)	0.461 (0.36)	0.505 (0.41)
CATHOLIC	- 0.119 (- 0.13)	0.167 (0.20)	5.105 (- 1.21)	4.111 (1.00)	- 2.381 (- 0.56)	- 2.373 (- 0.54)	- 1.020 (- 0.69)	- 0.669 (- 0.47)
MUSLIM	0.018 (0.00)	- 5.776 (- 0.20)	- 158.817 (- 1.80)	- 132.908 (- 1.82)	189.086 (1.60)	178.462 (1.30)	190.680** (4.15)	178.939** (3.86)
OTHER	0.852 (1.20)	0.918 (1.55)	6.271* (2.41)	6.067* (2.53)	- 0.699 (- 0.20)	- 0.747 (- 0.22)	0.163 (0.13)	0.222 (0.18)
GDPP	0.226 (0.62)	0.383 (1.01)	- 0.977 (- 0.57)	- 1.416 (- 0.86)	- 0.001 (0.00)	- 0.185 (- 0.07)	0.091 (0.08)	0.199 (0.18)
constant	- 0.766 (- 0.34)	- 1.250 (- 0.59)	7.009 (0.90)	8.034 (1.18)	6.383 (0.53)	7.554 (0.61)	3.310 (0.68)	3.242 (0.70)
obs	14	14	14	14	14	14	14	14
R ²	0.9012	0.9151	0.9171	0.9257	0.6813	0.6816	0.9129	0.9138

注：我们使用了工具变量回归方法，引入 LABORRATIO 作为工具变量。***（**，*）表明回归系数的显著性水平达到 99%（95%，90%）。括号中是相应的 t 检验结果。

其次，观察养老金投资规模对债权人保护程度的影响，我们同样发现：控制了养老金投资规模的内生性可能，养老金投资规模对于债权人保护程度都没有显著影响。原因相似，作为机构投资者的养老金投资于证券市场，会对债权人保护产生正反两方面的影响，两相抵消，净效果不是很明显。此外，与 Stulz and Williamson (2003) 不同，我们发现穆斯林国家的债权人保护程度 (LLBCR) 比基督教国家更高，这可能反映了穆斯林国家对合乎教义的债权关系保护更强 (Inglehart, 2000)。

总结以上发现，养老金投资规模对投资者保护程度并没有显著影响，因此我们的分析并不支持养老金投资的“投资者保护论”。

四、养老金投资规模对股票市场发展水平的影响

这一节中，我们首先讨论了检验养老金投资的“股市发展论”的回归模型，然后分析

了有关的实证结果。

（一）回归模型

$$SMCGDP_i (SMVGDP_i, \text{TURNOVER}_i) = \alpha_0 + \alpha_1 * \text{PENSION1992}_i + \alpha_2 * \text{GDPP}_i + f(\text{LLSVEN}_i (\text{LLBEN}_i), \text{ACCOUNTING}_i, \text{FRENCH}_i, \text{GERMAN}_i, \text{SCANDINAVLAN}_i, \text{CATHOLIC}_i, \text{MUSLIM}_i, \text{OTHER}_i) + \varepsilon_i$$

模型的基本框架与投资者保护程度的回归模型相似。不同的是：我们分别使用了 SMCGDP、SMVGDP 和 TURNOVER 来测量股票市场发展水平以及普通的多变量回归分析方法^①。

（二）实证结果

表 2 给出了有关养老金投资规模对股票市场发展水平影响的实证结果，总结如下。

首先，观察养老金投资规模对股票市场规模的影响，我们发现：养老金投资规模对股票市场的规模有着显著的推动作用。回归式（9）和（10）中 PENSION1992 的回归系数都为 1.171，显著性水平为 99%。其经济含义是：如果一国的养老金投资规模提高 1%，那么该国的股票市场规模就会提高 1.171%。这支持了 Bodie（1990）、Davis（1996，2000）和 Vittas（1998）的论述，作为机构投资者的养老金可以通过发挥清算结算、积累资金、转移资源、风险管理、风险控制、金融创新以及激励基金经理人等功能，扩大股票市场的融资规模。与普通法国家相比，法国大陆法国家的股票市场规模较低，而斯堪的那维亚大陆法国家的股票市场规模较高。这可能是由于，与普通法相比，法国大陆法对私人利益的保护较弱，损害了私人投资的意愿，进而损害了股市扩张，如 Beck et al.（2000）；斯堪的那维亚大陆法则强调社会规范对私人利益的保护（Coffee，2001），而灵活的社会规范可以及时适应市场变化，扩大了股市规模。此外，与 Beck et al.（2003）不同，与基督教国家相比，天主教国家的股市规模更大，这可能是由于法国大陆法和天主教人口比例之间较高的正相关性（0.8325）将较多的对股市规模的负面影响赋予了法国大陆法变量，进而出现了天主教人口比例的正回归系数。最后，与基督教国家相比，信仰其他宗教或者无宗教信仰的人口比例较高的国家的股市规模较大。

其次，观察养老金投资规模对股票市场流动性的影响，我们发现：养老金投资规模对股票市场流动性也有着显著的推动作用。以回归式（12）为例，PENSION1992 的回归系数为 0.937，显著性水平为 99%。其经济含义是：如果一国的养老金投资规模提高 1%，那么该国的股票市场流动性就会提高 0.937%。这同样支持了 Bodie（1990）、Davis（1996，2000）和 Vittas（1998）的论述，作为机构投资者的养老金可以通过向股市提供各种服务，提高股市流动性。与普通法国家相比，法国大陆法国家的股票市场流动性较低，而德国和斯堪的那维亚大陆法国家的股票市场流动性较高。原因与上相似：法国大陆法对私人产权的保护程度较低，损害了私人投资意（Beck et al.，2000），减低了股市的流动性；德国大

^① 这是由于股票市场发展水平变量的考察时间是 1993~1994 年，而 PENSION1992 的考察时间是 1992 年，因此这里不存在养老金投资规模的内生性可能。

表 2 以股票市场的发展水平为因变量的多变量回归结果

	SMCGDP		SMVGDGP		TURNOVER	
	(9)	(10)	(11)	(12)	(13)	(14)
PENSION1992	1.171*** (10.62)	1.171*** (12.39)	0.958*** (12.95)	0.937*** (13.45)	0.146 (2.30)	0.102 (1.72)
LLSVEN	0.041 (0.46)		- 0.179 (- 1.93)		- 0.419*** (- 7.25)	
LLBEN		0.084 (1.23)		- 0.200 (- 1.86)		- 0.523** (- 4.72)
ACCOUNTING	0.003 (0.88)	0.002 (1.03)	0.007 (2.19)	0.007 (2.05)	0.014*** (7.61)	0.014** (4.31)
FRENCH	- 0.476** (- 4.88)	- 0.464** (- 4.20)	- 0.324** (- 3.73)	- 0.312* (- 2.49)	0.097 (1.60)	0.105 (1.02)
GERMAN	0.168 (1.84)	0.196 (2.16)	0.351** (3.29)	0.336* (2.43)	0.496*** (6.12)	0.426* (2.75)
SCANDINAVIAN	0.765** (4.34)	0.759** (4.56)	0.574** (3.97)	0.594** (3.67)	- 0.201 (- 1.88)	- 0.152 (- 0.90)
CATHOLIC	1.384*** (6.34)	1.325*** (6.74)	0.667** (4.09)	0.815** (5.24)	- 1.102*** (- 8.91)	- 0.719** (- 5.69)
MUSLIM	12.290 (1.21)	13.514 (1.65)	18.373* (2.39)	12.594* (2.41)	18.479** (3.59)	5.068 (1.29)
OTHER	1.103*** (7.24)	1.104*** (8.23)	0.250 (2.16)	0.289 (2.06)	- 1.175*** (- 10.45)	- 1.100*** (- 7.95)
GDPP	- 0.035 (- 0.20)	- 0.091 (- 0.71)	- 0.063 (- 0.35)	- 0.059 (- 0.30)	- 0.124 (- 1.11)	- 0.038 (- 0.19)
constant	- 0.877 (- 0.90)	- 0.702 (- 0.98)	1.416 (1.59)	1.569 (1.83)	5.288*** (10.41)	5.348*** (7.42)
obs	14	14	14	14	14	14
R ²	0.9719	0.9742	0.9711	0.9601	0.9960	0.9894

注：*** (** , *) 表明回归系数的显著性水平达到 99% (95% , 90%)。括号中是相应的 t 检验结果。

陆法较高的适应性使得其更注意保护私人产权 (Beck et al., 2002), 提高了股市的流动性; 斯堪的那维亚大陆法对社会规范的强调有助于保护私人产权 (Coffee, 2001), 也提高了股市的流动性。同样, 与 Beck et al. (2003) 不同, 与基督教国家相比, 天主教国家的股市流动性更强, 这也是由法国大陆法和天主教人口比例之间较高的正相关性所致。而穆斯林国家较高的股市流动性同样反映了其教义对于认可的金融活动的完全支持 (Inglehart, 2000)。

最后, 观察养老金投资规模对股票市场效率的影响, 我们发现: 养老金投资规模对股票市场效率没有显著的影响。由于养老金投资规模既可以扩大股市规模, 也可以提高股市流动性, 如果这二者效果接近, 作为其比值的股市效率则不会受到养老金投资规模的显著影响。严格的执法程度会降低股市效率, 这与 Beck et al. (2000) 的发现相反, 可能是由于严格执法限制了股票市场的投机性买卖, 进而表现为股市效率的降低。与 Beck et al.

(2000)一致,严格的会计标准有助于提高股市效率,这反映了严格准确的会计信息有助于股票交易的顺利进行,从而提高了股市效率。与普通法国家相比,德国大陆法国家的股市效率较高,这同样反映了德国大陆法较高的适应性能够提高股市效率。此外,与基督教国家相比,天主教或其他国家的股市效率越低,这可能是由于天主教对等级、政府管制的强调,妨碍了自由交易的发展,降低了股市效率(Beck et al., 2003)。

总结以上发现,养老金投资规模对于股票市场的规模和流动性有着显著的推动作用,因此我们的分析支持了养老金投资的“股市发展论”。

五、中国的社保基金入市问题

中国的社会保障制度正处在一个转型与变革的关键时期(郑功成,2002),特别是随着2000年全国社会保障基金的建立,社保基金入市问题成为各方关注的焦点(李绍光,2001)。以往研究强调了股票市场发展对社保基金的影响,如张红地(2003)。而本文的分析则为社保基金入市问题提供了一个新的视角,即社保基金入市可以推动股票市场发展,这反过来会促进社保基金的成长。

表3比较了中国与经合组织成员国的养老金投资规模以及股市的规模和流动性^①。

表3 中国与其他国家的比较

国家	养老金投资规模	股票市场规模	股票市场流动性
澳大利亚	21.30%	62.69%	26.54%
比利时	2.50%	34.58%	5.44%
加拿大	31.30%	55.78%	27.85%
丹麦	14.40%	30.38%	17.20%
德国	2.90%	22.10%	19.19%
冰岛	40.80%	10.96%	0.95%
意大利	3.10%	14.55%	9.14%
日本	12.70%	67.56%	23.12%
韩国	3.20%	40.40%	69.40%
卢森堡	17.20%	143.09%	5.22%
荷兰	76.00%	59.87%	36.06%
挪威	4.50%	22.75%	7.35%
葡萄牙	2.90%	14.88%	5.92%
瑞典	1.60%	54.90%	33.25%
瑞士	59.60%	102.79%	78.86%
英国	52.70%	113.48%	68.06%
美国	50.00%	73.65%	51.23%
中国	0.85% a	37.43% ^①	62.78% ^②

注:①相应的变量时间为2001年,数据来自社保基金年度报告。②相应的变量时间为2002年,数据来自《中国统计摘要2003》和《广州统计年鉴2003》。

① 由于社保基金在中国成立不久,中国的上述数据分别来自2001和2002年,而经合组织的数据则来自1992~1994年。

我们发现：首先，中国当前的社保基金投资规模远低于经合组织国家 20 世纪 90 年代初的水平。中国社保基金投资规模在 2001 年仅为 0.85%，而 1992 年经合组织国家中养老金投资规模最低的瑞典也达到了 1.60%，是中国的近 2 倍，而同期养老金投资规模最高的荷兰达到了 76.00%，是中国的近 90 倍。因此，中国的社保基金规模很低，发展空间很大。其次，虽然中国的股市流动性（62.78%）稍高，大概相当于 1993~1994 年 17 个经合组织国家中第 4 名（前三名分别为瑞士、韩国、英国，对应值分别是 78.86%，69.40%，68.06%），但是股市规模（37.43%）很低，还不及经合组织国家 20 世纪 90 年代初以中位数计的平均水平（瑞典，54.90%），因此，中国的股票市场也有待发展。

二者综合考虑，我们认为加大中国社保基金投资规模可以扩大股市规模，提高其流动性，为养老金增值提供更大的空间。但是，考虑到养老金投资规模对投资者保护的无效性结论，我们主张社保基金入市不应当背负提高市场上对投资者保护程度的任务，而只须考虑如何在风险较高的股票市场上保值增值。如何更好地保护包括社保基金在内的广大投资者利益，这是有关监管和立法机构需要研究解决的紧迫问题。

六、结论和政策含义

以经合组织国家为例，我们分析了养老金投资规模对投资者保护程度和股票市场发展水平的影响，并探讨了中国的社保基金入市问题。结果显示：养老金投资规模对投资者保护程度没有显著影响，但对股票市场发展水平却有着显著的推动作用。多个控制变量的引入也支持了以上的稳定发现。因此本文支持了养老金投资的“股市发展论”，而否定了养老金投资的“投资者保护论”。

考虑到中国当前的社保基金入市问题，我们认为应当特别需要澄清两点认识：从大力推动股票市场发展的角度分析，有关机构应该积极推动社保基金入市，这有助于扩大中国股市规模，增强其流动性；但是，社保基金入市不应当承担加强市场上投资者保护的责任，如何更好地保护包括社保基金在内的广大投资者利益，这需要相关的立法和监管机构加以研究解决。

参考文献

- [1] 科林·吉列恩，约翰·特纳，克利夫·贝雷，丹尼斯·拉图利普：《全球养老保障——改革与发展》，杨燕绥译，中国劳动社会保障出版社 2002 年版。
- [2] 李绍光：《养老金的投资及管理——国内理论探讨与实践综述》，《经济社会体制比较》，2001 年第 1 期。
- [3] 张红：《进一步加快我国社会保障基金进入资本市场步伐》，《管理世界》，2003 年第 6 期。
- [4] 郑功成：《加入 WTO 与中国的社会保障改革》，《管理世界》，2002 年第 4 期。
- [5] Becht, Macro, Patrick, Bolton, and Alisa, Röell, Corporate Governance and Control, NBER Working Paper 9371, 2002.
- [6] Beck, Thorsten, Asli, Demirgüç-Kunt, and Ross, Levine, Law, Endowments, and Finance, Journal of Financial Economics, 2003, Vol. 70, pp137 ~ 181.
- [7] Beck, Thorsten, Asli, Demirgüç-Kunt, and Ross, Levine, Law and Finance: Why Does Legal Origin Matter?

Journal of Comparative Economics, 2002.

- [8] Beck, Thorsten, Asli Demirgüç-Kunt, and Ross, Levine, *A New Database on Financial Development and Structure*, World Bank, 1999.
- [9] Beck, Thorsten, and Ross, Levine, *Legal Institutions and Financial Development*, Working Paper, 2003.
- [10] Beck, Thorsten, Ross, Levine, and Norman, Loayza, *Financial Intermediary Development and Growth: Causes and Causality*, *Journal of Monetary Economics*, 2000, Vol.46, pp31 ~ 77.
- [11] Bodie, Z., *Pension Funds and Financial Innovation*, *Financial Management*, Autumn, 1990.
- [12] Coffee, John C., *Do Norms Matter? A Cross-Country Examination of the Private Benefits of Control*, Working Paper, Columbia Law School, 2001.
- [13] Davis, E. P., *Implications of the Growth of Institutional Investors for the Financial Sectors*, Discussion Paper PI- 0001, The Pensions Institute, Birkbeck College, University of London, UK, 2000.
- [14] Davis, E. P., *The Role of Institutional Investors in the Evolution of Financial Structure and Behavior*, Financial Markets Group and ESRC Working Paper, 1996.
- [15] Del, Guercio D., and J., Hawkins, *The Motivation and Impact of Pension Fund Activism*, *Journal of Financial Economics* 1999, Vol.52, pp293 ~ 340.
- [16] Inglehart, Ronald, *Culture and Democracy*, In eds. L. E. Harrison and S. P. Huntington, *Culture Matters*, Basic Books, New York, NY, 2000.
- [17] La Porta, R., Lopez-de-Silanes, F., Shleifer, A., and R. W., Vishny, *Investor Protection and Corporate Valuation*, *Journal of Finance*, 2002, Vol.57, pp1147 ~ 1170.
- [18] La Porta, R., Lopez-de-Silanes, F., Shleifer, A., and R. W., Vishny, *The Quality of Government*, *Journal of Law, Economics, and Organization*, 1999, Vol.15, pp222 ~ 279.
- [19] La Porta, R., Lopez-de-Silanes, F., Shleifer, A., and R. W., Vishny, *Law and Finance*, *Journal of Political Economy*, 1998, Vol.15, pp1113 ~ 1155.
- [20] OECD, *OECD in Figures: Statistics on the Member Countries*, September, 2003
- [21] Scharfstein, D. S., and J. C., Stein, *Herd Behavior and Investment*, *American Economic Review*, 1990, Vol.80, pp465 ~ 479.
- [22] Shleifer, Andrei, and Robert W., Vishny, *Large Shareholders and Corporate Control*, *Journal of Political Economy*, 1986, Vol.94, pp461 ~ 488.
- [23] Shleifer, Andrei, and Robert W., Vishny, *A Survey of Corporate Governance*, *Journal of Finance*, 1997, Vol.52, pp737 ~ 783.
- [24] Stein, Jeremy C., *Agency, Information and Corporate Investment*, in eds. George Constantinides, Milt Harris, and René Stulz, *Handbook of Economics of Finance*, Amsterdam: North-Holland, 2001.
- [25] Stulz, René, and Rohan, Williamson, *Culture, Openness, and Finance*, *Journal of Financial Economics*, 2003, Vol.70, pp313 ~ 349.
- [26] Vittas, D., *Institutional Investors and Security Markets: Which Comes First? The World Bank Research Group, The ABCD LAC Conference, June 28 ~ 30, 1998, San Salvador, El Salvador.*
- [27] Wahal, S., *Public Pension Fund Activism and Firm Performance*, *Journal of Financial and Quantitative Analysis*, 1996, Vol.31, pp1 ~ 23.
- [28] World Bank, *World Development Indicator 2001*, The World Bank Group, Washington DC, 2001.
- [29] Zingales, Luigi, *The Value of the Voting Right: A Study of the Milan Stock Exchange Experience*, *Review of Financial Studies*, 1994, Vol.7, pp125 ~ 148.

投资基金业的跨界活动与障碍*

李仲翔 李仲飞 陆 军

一、投资基金业的跨界活动

跨界活动 (cross-border activity) 是指投资基金或投资基金管理公司 (统称为投资基金业) 跨越国家或司法管辖区的业务活动。

投资基金业的跨界推销的主要方式是通过外国基金的直接销售以及国内基金使用外国基金管理机构。基金管理人还开发出了一些变通的结构模式来克服对基金经营者想在本国之外推销基金的烦琐的法律、监管和其他限制。

另外,有一些多边的协定来辅助跨界活动,例如欧盟 UCITS 指导原则,适用于欧盟成员国。在没有这种多边协定的情况下,基金经营者在跨界活动将面临好多障碍。

(一) 跨界活动的类型

1. 外国基金在境内的直接销售

在本国之外销售基金的最直接的方式是直接向外国投资者销售基金。大多数司法管辖至少要求外国监管机构提供与本国监管机构同样或类似的投资者保护机制。另外,基金单位的销售或推销诉诸于不同的监管和法律规定和限制。

2. 外国基金管理公司在境内开展咨询活动

国内的基金管理人将部分功能或职责委托给第三者或寻求第三者的帮助,如在外国司法管辖的基金管理机构。一个想向外国机构提供服务的本国基金管理机构可以通过在外国建立分支机构或直接提供服务而不需要具体出现在外国司法管辖区内。这种选择取决于本国监管机构与外国监管机构的的关系、外国监管机构对基金管理机构的法律限制等。

在这方面,合作是非常重要的。母国监管机构在评估基金管理机构的适合性、与国内基金的业务安排以及遵守母国监管机构法律的程度方面可以向外国监管机构提供帮助。

3. 其他跨界活动

至于其他方式的跨界活动,出于成本和资源的考虑,不需要实质性的母国监管机构和外国监管机构卷入监管活动。

这种基金操作的一般方式是,基金建立一种结构,可以将一部分或全部资产投资到一个外国基金或一组选择的外国基金,而实际上不需要把自己建立成为一个国内基金。这种基金的结构可以帮助基金的操作者克服税收、文化和官僚方面的原因造成的困难。投资者

* 原载:《国际金融研究》,2003年第2期。

倾向于投资当地的基金，而不是直接投资到外国基金。在这一意义上，这种结构的建立被看作是推进外国基金投资的推销工具。

（二）跨界活动的程度

1. 外国基金的直接推销

目前从一个司法管辖区到另一个司法管辖区的跨界活动差别很大。例如，在中国香港地区和瑞士，外国基金的数量超过当地的数量。香港地区可以接受任何被认为与其对投资者保护相当的国外基金。美国只有 19 家外国基金，主要来自加拿大，最后一个是在 1973 年批准的。1994 年，澳大利亚放宽了对外国基金的政策，但截止到 1996 年，只有几个外国基金在那销售。

自从引入 UCITS 指南、寻求在欧盟成员国内协调一致的基金结构以来，欧盟成员国基金的跨界活动一直在增加。UCITS 在一个成员国注册后可以在其他成员国销售，只需遵守其他成员国对与推销和广告的规则要求有关的条件。这是卢森堡作为离岸基金迅速发展的主要因素。但非欧盟成员国在这些成员国内的跨界活动非常少。

2. 外国基金管理机构的使用

外国基金管理机构对国内基金提供帮助的程度差别很大，尽管多数取决于如何定义“管理”和“咨询”。在很多国家如奥地利、瑞士和欧盟国家，管理公司必须是在本国监管机构（home regulatory authority）控制下的当地实体。然而，基金也可以使用外部的基金管理机构或委托投资职责，假如管理公司的最终的责任没有被影响或者本国监管机构的监管没有降低。这样，监管机构在批准外国基金管理机构方面的卷入程度有很大的不同，从需要事先批准（法国和意大利）到除了披露没有任何要求（如澳大利亚、德国、卢森堡和荷兰）。

（三）基金跨界推销的障碍

跨界推销活动仍然存在着法律和监管方面的障碍。监管机构通常不直接负责税收问题。但税收是一个障碍；另一个是贸易壁垒。

1. 监管和法律问题

（1）基金的认可或批准。

虽然大多数国家的监管机构有权认可外国基金，但几乎所有的国家和其他司法管辖区立法禁止外国基金的推销，除非在外国监管机构具有可比的投资者保护水平。其他国家有限制外国基金推销的严格法律，实际上使注册和批准外国基金成为不可能。例如在西班牙，还没有非欧盟国家能满足其监管机构对具有可比性的投资者保护的要求。

美国要求只有证券与交易委员会（SEC）能够对外国基金进行监管和执法，而且必须是与公众利益和投资者保护一致的。这意味着 SEC 要看外国基金的结构、组织和经营是否与美国“投资公司法”一致，SEC 的官员是否可以对外国基金在出现问题时采取执法行动。由于外国监管体系与美国投资公司法的要求不同，目前大多数外国基金不愿意或不能重新构造他们的基金来与投资公司法一致。另外，任何发行者在美国销售其股份必须遵守销售活动所在州的法律和联邦法律

(2) 外国监管机构对基金的调查。

像加拿大的安大略省 (Ontario)、法国、德国、意大利、日本、卢森堡、墨西哥、西班牙和瑞典的司法管辖都有对外国监管机构进入他们的管辖区进行调查的限制。在许多国家和地区如香港，虽然没有对外国监管机构进行调查的法律障碍，但外国监管机构必须先得到当地监管机构的允许。在荷兰，允许外国监管机构由荷兰监管机构代表其参与调查。在美国，关于外国监管机构进入美国进行检查，除了通知美国的司法部外，没有法律障碍。

然而，有些司法管辖区禁止外国监管机构在其管辖区进行外国基金的调查和监督。例如，在卢森堡和瑞士，外国监管即不允许单独也不允许与当地监管机构联合进行调查或检查。在法国和日本，外国监管机构的调查和检查也被禁止。在澳大利亚，外国监管机构的正式询问只有通过共同的立法才行。

也有对外国监管机构的调查限制在只有被调查的基金管理机构或相关人事同意才行。在加拿大的安大略省，外国监管机构只有拥有相关信息的居民的协助下才能进入该省进行调查。如果该人拒绝自愿地提供信息，外国机构就不能进行调查，除非再征求该省证券委员会 (OSC) 的协助。

(3) 外国监管机构对基金的日常检查和监督。

在一些司法管辖区，只有在与当地的监管机构联合的情况下才能进行检查和监督。当外国监管具有清楚可辨的利益关系时，英国才允许外国监管机构进行联合检查。在其他司法管辖区，如德国，必须具有互惠的协议时才能进行。在其他国家，外国监管机构进行检查不需要获得当地监管机构的允许。如美国，外国监管不需要得到 SEC 的允许。但要求向美国司法部提供监督的通知，然后通知美国国务院。

(4) 主权问题。

外国监管机构对当地基金的检查活动实际上与主权问题相联。如果基金被所在国的监管机构监管而不是被外国机构监管，问题就产生了，即是否允许外国监管应当进入本国司法管辖区进行基金的调查和检查，特别是没有当地监管机构出现的情况下。因此监管机构面临着如何平衡主权考虑来发展互惠机制为跨界活动和进行充足的监管提供方便。

当外国监管机构希望对与基金有关的某一特定问题进行调查或监督时，而当地监管机构对这一问题却并不感兴趣，或者认为不是特别恰当时，就会产生主权问题。有些国家通过下列方式解决这个问题：或者是让外国监管机构进行检查或调查活动（如果当地监管机构在场的话），或者允许外国监管机构在当地监管机构的驻地审阅相关的文件。

(5) 交换信息的能力。

国内的保密法也会限制本国监管机构对外国监管机构的信息释放与提供。例如，除非满足特殊的条件，澳大利亚国内法限制保密信息的释放。

2. 影响跨界活动的其他因素

(1) 资源。监管机构应当解决的一个问题是如何有效地利用有限的监管资源。完善监管机制，最大限度地改善跨界监管。有些监管机构和其他监管机构建立了信息交换程序来改善合作和效率。例如，英国和澳大利亚在调查活动和证据收集方面可以代表对方提供相互协助。

(2) 损失的恢复。许多投资者在决定是否投资于外国基金时考虑的是挽回基金产生的损失的能力。这是与投资者保护有关的问题。它可能是希望在境外销售其基金的基金经营者的一个考虑。

在澳大利亚这样的国家可能不是那么显著。该国有立法规定，外国基金的促销人必须向澳大利亚监管机构展示以下之一：①某一澳大利亚法庭在外国司法管辖区有判决的认可以及执法的相互谅解和安排，或者，有监管方面的要求来为投资者挽回在外国司法管辖区所欠的损失提供方便。②遵守澳大利亚公司法，包括公司注册。

(3) 文化问题。另一个可能影响跨界活动的因素是文化障碍。文化的差别可能影响当地的投资者对外国基金的吸引力。一个国家的投资者可能对另一个国家在该国法律下建立的投资工具感到不舒服或者对他们的结构不熟悉。在一定程度上主辅基金 (Master-feeder) 在当地销售，迎合了当地投资者和监管者的需要。

二、解决对跨界活动障碍的通行做法

为了有效地监管跨界活动，传统上，监管机构采用了许多方法来改善监管。这些方法主要是集中在建立信息交换和监管合作的机制上。

1. 谅解备忘录

为了保护国内市场不受来自本国司法管辖之外的欺诈和非法行为影响，好多国家签订了条约、双边协定或谅解备忘录。谅解备忘录不是可执法的协定，只是相关方的意向性声明；主要集中在信息交换的能力、程序，以便为刑事或民事活动提供证据。

2. 共同协助协定

有些国家的立法允许在一定的条件下，当外国监管机构要对某一个具体的公司或人员进行检查以获得证据时，帮助外国监管机构获得相关的信息。这种协助克服了当外国监管机构没有权利在另一个国家对账册、记录等进行检查时，获得证据或传唤证人出庭等障碍。

澳大利亚已经做了这样的立法 (the Mutual Assistance in Business Regulation Act, 1992)。美国、英国和香港地区也进行了类似的立法。

3. 为合作和分享信息的非正式安排

为了从另一监管机构获得信息或寻求协助，非正式的合作交流是很普遍的，但主要依赖于个人的交往，而不是监管机构之间的官方正式关系。但这种合作是不稳定和不充分的。

参考文献

- [1] Investment Company Act of 1940 and Investment Advisers Act of 1940, 1998 edition, RR Donnelley Financial.
- [2] Protecting Investors: A Half Century of Investment Company Regulation, Division of Investment Management United States Securities and Exchange Commission, May 1992.
- [3] 2000 Mutual Funds and Investment Management Conference, March 26 ~ 30, 2000, Co-sponsored by the Federal Bar Association and the ICI Education Foundation.

- [4] ICI Mutual Annual Report , 1998.
- [5] Gramm - Leach - Bliley Act , October 26 , 1999 , U. S. Congress.
- [6] Financial Supervision and the Internet , MAS (Monetary Authority of Singapore) managing Director 's speech. Copyright 2000 Times Business Publications , Business Times (Singapore) , April 12 , 2000.
- [7] 1999 Annual Report , Financial Service Authority , the U. K.
- [8] 1999 Annual Report , Singapore Monetary Authority.

影响港元汇率的货币政策因素分析*

林 江 夏育松

一、引言

对于香港这样一个小规模开放经济而言，汇率在经济生活中扮演着极其重要的角色。香港的对外贸易总值是本地生产总值的 2.4 倍，其银行体系总存款的 50% 是外币存款，而香港股票市场交易额的三分之一也为外资主导。在这种环境因素下，香港经济的稳定与港元汇率的稳定息息相关。在联系汇率制下，香港因为维持了港元汇率的长期稳定而逐步发展成为一个国际金融中心。然而在 1997 年的亚洲金融危机中，港元汇率虽然承受住了巨大的贬值压力，但是香港经济仍在相当一段时期内遭受了较为严重的打击：资本大量外流，股指巨幅下跌，资产市场价格大幅回落，消费降低，经济甚至出现负增长。在维护港元汇率稳定的过程中，香港经济的发展对汇率的稳定也变得越来越敏感。因此，研究影响港元汇率稳定的有关因素，找出维护港元汇率稳定的对策，仍然是十分重要的命题。

汇率是货币的外在价值，汇率政策通常是货币政策的一部分，而货币政策也是经典汇率理论模型中的重要内容。香港在这方面尤为突出：汇率在香港经济中的重要作用，就决定了汇率政策在货币政策的重要性。事实上，香港在采取货币政策维护汇率稳定时从来都是不遗余力的。在与美元挂钩的联系汇率制度确立以前，香港缺乏有力的货币政策工具，货币政策在维护汇率稳定时的作用非常有限；而以联汇制为核心货币政策，无论是在政策目标上还是政策工具上都有了巨大改变，港元汇率由此获得了空前的稳定。

货币政策涉及利率、货币供应及通货膨胀等多项经济因素，但并不是每一个都与港元汇率密切相关，故在接下来的部分中，本文将在分析联系汇率下香港货币政策与港元汇率的相关关系基础上，发掘与港元汇率变动高度相互影响的因素变量，并采用实际数据进行计量分析，得出实证结果。

二、香港的货币政策机制

（一）联系汇率制

1983 年 10 月 15 日，香港采用美元联系汇率制。按照此制度，发钞银行要发行港币，必须以 1 美元 = 7.8 港元的固定汇率向政府的外汇基金（现金融管理局）缴纳相应数量的

* 原载：《国际金融研究》，2003 年第 10 期。

美元，外汇基金则向发钞银行签发负债证明书，并承诺按 1 比 7.8 的汇率赎回证明书，发钞银行也同意按照相同的汇率向其他银行提供纸币或从其他银行收回纸币。但在外汇市场里，港元汇率仍然是自由浮动的，不过联系汇率制的内在“套戥”机制会使之向官方的固定汇率靠拢。例如，当市场上港元和美元交易的汇率高于 7.8 的水平，银行就会通过发钞银行把它们持有的港元纸币按照固定汇率兑现为美元，然后在市场上以市场汇率卖出美元，换回港元，从而赚取差价利润。银行套汇活动的后果是，流通中的港元纸币数量下降，银行体系的银根抽紧，港元利率将会上升，从而吸引资金流入，美元供给增加。套汇以及紧随其后的套息活动会影响港元与美元之间的相对供求，推动市场汇率至固定汇率水平，直到套利机会消失。

（二）货币政策的内容

联汇制下的货币政策与联汇制本身是密不可分的。联汇制本身是一种货币发行局制度，发钞银行每发行 7.8 港元，必须在外汇基金无息存入 1 美元。港元的发行有了百分之百的外汇保证，这无疑会极大地稳定和维护外界对港元的信心。

在货币政策的目标问题上，香港的选择非常明确——以保持货币的外在价值，即港元汇率的稳定作为惟一目标。其合理性在于：在香港，外来因素占主导地位，本地经济的增长会受到当局无法控制的外部因素的制约，外资和外贸增长变化的货币表现同样很难为金融当局所控制，以调控利率或货币供应量从而抑制通货膨胀作为货币政策目标是不切实际的。而在“套戥”机制的作用和政策工具的辅助下，稳定汇率的政策目标则较容易实现，具有很强的操作性。

实现政策目标的基本手段是“套戥”机制，但现实中“套戥”机制的运作存在诸多局限。首先，7.8 的官方固定汇率只对银行适用。只有银行才能够与外汇基金进行交易并进行套汇活动，而其他市场参与者都被排除在外。其次，银行进行套汇的能力是有限的，因为银行必须维持一定程度的现金比率以应付顾客的提存要求。再者，在以 1 比 7.8 的汇率与外汇基金直接或间接地进行交易时，银行只能使用港元现钞，而并非港元存款。如果当市场汇率出现大幅波动时，自发的“套戥”将无法维持汇率的稳定，此时外汇基金必须在外汇市场对港元进行直接买卖干预，以影响港元资金的供应和调节利率，并最终实现汇率的稳定。但是银行及其客户的正常业务活动，例如外币资产买卖和同业拆借，也会影响银行同业拆借市场上的资金供应以及同业拆借利率水平，如果这种影响不为金融当局所控制的话，就很可能与外汇基金为稳定汇率而进行的干预行动方向相左，从而令干预的实际效果大打折扣。这时候运用货币政策工具进行干预就是必需的。

1988 年 7 月，港府与作为中央结算所管理银行的汇丰银行达成新会计安排。由于香港的每一家持牌银行都是在汇丰银行或与汇丰银行有业务往来的结算银行开设账户，故外汇基金就此与汇丰银行达成协议：汇丰银行必须在外汇基金开设一个账户，该账户的港元余额不得小于整个银行体系存放在汇丰银行的结算余额。汇丰银行承诺其在银行同业市场上进行活动所产生的结算净额不得超过汇丰银行在外汇基金的港元账户余额，否则，汇丰必须向外汇基金支付惩罚性利率。这样，金融当局就能有效地控制银行同业银根和间接地

影响银行同业拆借利率，这种功能类似于法定准备金政策。^①

1990年3月，外汇基金开始发售外汇基金票据，其目的在于银行同业市场的流动资金和同业拆借利率水平。当银行拥有大量资金，银行就会把其剩余的资金注入银行同业拆借市场，这样市场上港元供给增加，利率下降，从而会产生使港元汇率下调的压力。这时金融当局可以向银行发售外汇基金票据，吸纳部分流动资金，从而减少了市场上的港元供给，拉高利率，缓解港元汇率下调的压力。同时，外汇基金票据的息率高低也会对同业拆借利率的高低产生影响。外汇基金票据计划事实上使得金融当局拥有了公开市场操作的政策工具。

为了能更好地控制银行同业银根，港府于1992年5月推出了流动资金调节机制。在此机制下，银行若在每日票据结算前发现头寸不足需要隔夜融资时，可就该银行所持有的外汇基金票据与外汇基金签订隔夜回购协议回。相反，若某银行有剩余的隔夜资金，也可按拆入利率把该款存入外汇基金。外汇基金的拆入拆出利率实际上成为银行同业隔夜拆借利率的上下限。外汇基金有权随时取消或变动拆入拆出利率，以调节银行的流动性水平和影响银行同业市场的拆借利率，从而确保港元汇率的稳定。

1993年，香港金融管理局成立，负责履行中央银行的职能。1998年9月，在经受了亚洲金融危机考验之后，金融管理局为了巩固联系汇率制度，推出了7项技术性措施。这些措施集中表现为两点：其一为兑换保证。金管局明确保证，银行可以按7.75港元兑1^②美元的固定汇率，将其在结算账户的港元兑换成美元。银行按固定汇率兑换美元不必再经过发钞银行，而且不再受现钞交易的限制，这样不仅港元稳定的信心得到增强，也扫除了一些“套戥”机制运作的障碍。其二是改革流动资金调节机制，更名为贴现窗。流动资金调节机制下的拆入利率被撤销，拆出利率被定为基本利率，银行可重复使用其持有的外汇基金票据签订回购协议以获取隔夜流动资金而不受惩罚。

三、货币政策因素与港元汇率的关系

丁伯根法则指出：要实现多个宏观经济目标，当局必须拥有等于或多于目标数的政策工具。货币政策实质上是单一的工具，所以它不能同时被赋予超过一个的政策目标，特别是这些目标之间至少在短期内可能存在冲突。因此，金融当局只能在汇率、利率和货币供应量三个目标之间，选择其中之一来加以控制。

既然香港以稳定汇率作为其货币政策的惟一目标，那么它就不能再通过调控利率或货币供应量来抑制通货膨胀，而一般来说，长期的高通货膨胀会对汇率造成贬值的压力。通货膨胀意味着香港的商品价格上升，出口商品的竞争力下降，从而出口到美国的商品数量减少，同时通胀使得美国商品的价格相对下降，导致进口增加。这样，美元的供给下降而

^① 1996年12月，香港开始采用即时支付系统，所有持牌银行都直接在金融管理局开立结算账户进行结算，新会计安排因而被取消。在新系统下，对货币市场操作的影响变得更为直接和有效。

^② 7.75是当时金管局的干预汇率水平。从1998年11月26日后起，兑换保证适用汇率从7.75的水平每天自动调整一个基点（即0.0001港元），直至调整到7.80的水平为止。

对美元的需求上升，港元汇率就面临贬值的压力。但事实上这种情况并未发生，图 1 和图 2 显示，香港在 1986~1992 年期间的通胀持续走高，但同期的汇率却没有下调的趋势。原因在于，通胀促使贬值发生的前提是：通胀中贸易品的价格上升。但是，香港通胀的压力却主要缘于非贸易品或服务方面^①，故香港的通胀虽然远高于同期美国的水平，却并没有对汇率的稳定造成威胁。

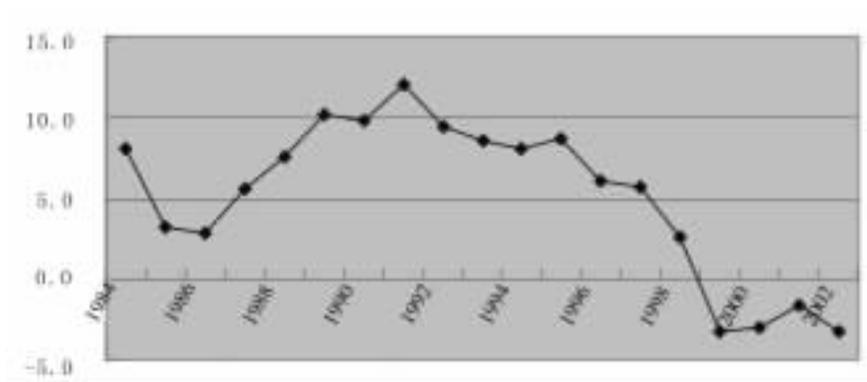


图 1 香港通货膨胀率

数据来源：历年的《香港经济年鉴》

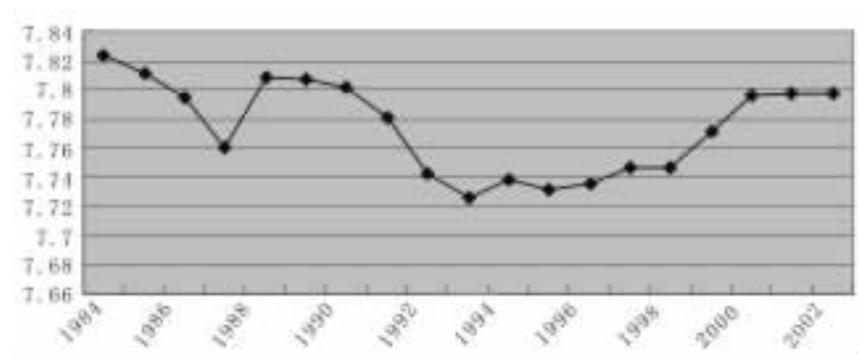


图 2 港元对美元汇率

数据来源：同图 1

反过来，政策目标上的取舍也并非造成香港通货膨胀的上升。首先，虽然货币的发行机制是与联系汇率制相联系的，但是货币的供应量却并不由香港金融当局控制。一方面，随着经济的增长和收入的增加，公众对现钞的需求也增加。面对这种情况，发钞银行只要向外汇基金无息存入足够的美元就能发行新的港元以满足需求。理论上，联系汇率制因为“套戥”机制作用可以控制货币的发行量在一定范围内变化，但现实中“套戥”机制却由于前面提到的原因而很少发生作用。另一方面，虽然金融当局有新会计安排、外汇基金票据买卖和流动资金调节机制的政策工具可以运用，但仅限于影响银行同业市场的流动资金

^① 1994~1995 年度香港政府收支预算案报告将香港的高通胀归咎于“土地短缺，房产价格不断上升，劳工短缺，生产与服务成本上升”，见《香港大辞典》第 220 页。

水平，这远不同于中央银行对整个信贷规模的控制，所以货币供应量的走势是游离于缺少中央银行的香港金融当局的掌握之外的。因而即使失控的货币供应量是通货膨胀的成因之一，罪魁祸首也不是稳定汇率的货币政策目标。其次，虽然香港是一个非常外向型的经济实体，几乎所有消费品均依赖进口，进口价格在决定本地物价水平的过程中扮演十分重要的角色，但是实际上香港的进口价格年均上升的速度非常低，很多年份都达不到 2%，故此，因港元与美元挂钩，香港因而产生输入型通货膨胀的假设在事实上是不存在的。

金融当局不能控制的货币供应量也不会影响汇率的稳定。因为持续快速增长的货币供应量的一般后果是通货膨胀的产生，从而导致汇率面临贬值的压力。但是从上面的分析来看，这个结论在香港不成立。而银行同业拆借利率却能充当汇率变化的晴雨表。因为外汇市场上 80%^① 的交易是银行同业间的交易，港元对美元的汇率实际上由同业市场港元与美元之间的相对供求决定的，而同业拆借利率的高低不仅反映同业市场上港元资金的供求水平，也推动供求向均衡运动。纯粹市场中的同业拆借利率自发地反映和调节汇率的变动，而在有干预的情况下，这种作用更为强有力。

联系汇率制的内在“现金套戥”机制很少实际发生作用，所以联系汇率制在稳定汇率方面的成功实际反映了金融当局干预的成功。无论是外汇基金在外汇市场上进行买卖，还是货币政策工具的运用，实质上都体现了当局对银行同业市场的干预，对整个同业市场的资金供应量和同业拆借利率水平的控制，其最终作用都是重新调节港元和美元之间的相对供给与需求，使汇率恢复稳定。在这一过程中，同业拆借利率的变化包含了当局的干预，因为不仅当局能直接控制同业拆借利率，而且同业间流动资金规模在当局的控制下最终也会促使同业拆借利率发生变化。成功的干预意味着同业拆借利率的变化会促使汇率的调整；而汇率的一个不稳定变化也会迎来当局的一个快速干预，信号就是同业拆借利率的调整。汇率与同业拆借利率之间似乎存在反馈关系。这里要提到银行存款利率，当港元汇率面临下调压力时，提高港元的存款利率会吸引人们将美元换成港元以获取更高的收益，美元的供给增加，需求下降，从而缓解港元汇率下调的压力。但是存款利率的调整速度和力度却远不及同业拆借利率，而且存款利率在相当一段时间内受银行同业公会的利率上限规则限制（注：解除的说明），所以存款利率的变化与汇率的变化可能相当不匹配，甚至在统计上没有太大关系。

四、实证分析

在这一部分，我们将采用香港的实际数据，运用 Granger 因果关系检验来验证上一部分的分析结果。所谓 Granger 因果关系，是指如果时间序列 x 的滞后值的使用有助于时间序列 y 的预测，那么就称 x 是 y 的 Granger 原因，记为 $x \rightarrow y$ 。如果 $x \rightarrow y$ ，同时 $y \rightarrow x$ ，则两个变量之间存在反馈关系。二元 Granger 因果关系的检验通常采用如下的二元向量自回归形式：

^① 钱荣堃，陈平等主编：《国际金融》，南开大学出版社。

$$x_t = \alpha_0 + \sum_{i=1}^m \alpha_i x_{t-i} + \sum_{i=1}^n \beta_i y_{t-i} + \varepsilon_t$$

$$y_t = \gamma_0 + \sum_{i=1}^m \gamma_i x_{t-i} + \sum_{i=1}^n \delta_i y_{t-i} + \eta_t$$

检验的零假设是 $\beta_i = 0$ ($i = 1, 2, \dots, n$) [或 $\delta_i = 0$ ($i = 1, 2, \dots, m$)], 基于该零假设构造的 F 统计量如果大于 5% 置信水平下的临界值, 我们就应该拒绝该零假设, 认为 x 是 y 的 Granger 原因, 即 $x \rightarrow y$ (或 $y \rightarrow x$)。如果 β 和 δ 都显著的异于 0, 则 x 与 y 之间存在反馈关系。

Granger 因果关系检验要求平稳的时间序列, 所以在这之前我们必须对所有的时间序列进行 ADF 检验, 来检验它们是否存在单位根。检验的过程是进行如下回归:

$$\Delta x_t = \rho_0 + \rho_1 t + \rho_2 x_{t-1} + \sum_{i=1}^m \delta_i \Delta x_{t-i}$$

并作假设检验: $H_0: \rho_2 = 0$; $H_1: \rho_2 < 0$ 。如果接受而拒绝, 则说明序列 x 存在单位根, 是非平稳的; 否则说明 x 是平稳的。 t 表示时间趋势, 也可以根据序列的实际情况不将截距和时间趋势加入回归。回归中加入 n 个滞后项是为了使残差项为白噪声。对于非平稳时间序列, 还需要检验其高阶差分得平稳性。如果序列的 d 阶差分是平稳的, 则称此序列是 d 阶单整的。

我们主要采用 1984 年 1 月至 2002 年 12 月的月度数据, 各个变量对应的名称见表 1。其中, 由于数据缺失的缘故, 香港通货膨胀率由甲类物价指数增长率代替, 样本期间为 1990 年 1 月至 2002 年 12 月。数据来源于《香港经济年鉴》、《香港大辞典》、香港特区政府统计处网页以及香港金融管理局网页。

表 1 数据变量名

港元对美元汇率	EX	货币供应量	M2
通货膨胀率	INF	3 个月定期存款利率	R3
隔夜同业拆借利率	HIBOR_N	隔月同业拆借利率	HIBOR_M

表 2 给出了表 1 中所有变量的平稳性检验结果。其中只有隔夜和隔月的同业拆借利率是平稳的, 而港元汇率、货币供应量、通胀率和 3 个月定期存款利率都存在单位根, 但是它们的一阶差分都是平稳的。接下来, 我们对所有的平稳数据进行二元 Granger 因果关系检验。

由于 Granger 因果关系检验的结果对设定的滞后期数很敏感, 故在检验之前我们先行对每一组变量构造二元 VAR 系统, 通过对不同滞后期的 VAR 系统的 AIC 和 SC 标准进行权衡比较, 得出最优的滞后期数。大部分 VAR 系统的最优滞后期数为 3, 所以我们选择 3 为滞后期数进行 Granger 检验。

表 2 ADF 检验结果

变量	ADF 统计量	检验形式 (C, T, L)	变量	ADF 统计量	检验形式 (C, T, L)
EX	- 1.973461	C, N, 1	Δ EX	- 12.03792**	C, N, 1
M2	- 2.208338	C, T, 1	Δ M2	- 12.26585**	C, T, 1
INF	- 0.967024	C, N, 1	Δ INF	- 9.171490**	C, N, 1
R3	- 2.732268	C, N, 1	Δ R3	- 8.778390**	C, N, 1
HIBOR_N	- 4.298388**	C, N, 1	HIBOR_M	- 4.298388**	C, N, 1

注：检验形式 C, T 和 L 分别表示单位检验方程包括截距、时间趋势和滞后阶数, N 指不包括 C 或 T。Δ 表示一阶差分。** 表示在 5% 的置信水平下显著。

检验的结果 (见表 3) 与我们在上一部分的分析结论几乎完全一致。货币供应量、通胀与港元汇率之间均不存在 Granger 因果关系, 这说明 M2 和 INF 的变化都不会对 EX 的变化产生影响。而 R3 不影响 EX 也证实了我们的怀疑, 定期存款利率很难随汇率的变动而及时调整, 同时 EX 不影响 R3 也说明定期存款利率的变动只要受其他因素的制约。只有隔夜拆借利率最能反映汇率的变动, 而且两者之间是一种反馈关系。相比之下, 隔月拆借利率易受汇率变动的影响, 但在促使汇率调整方面却没什么作为。

表 3 Granger 因果关系检验结果

零假设 H_0	F 统计量	结 论
Δ M2 不是 Δ EX 的 Granger 原因	0.33973	接受 H_0
Δ EX 不是 Δ M2 的 Granger 原因	1.76957	接受 H_0
Δ INF 不是 Δ EX 的 Granger 原因	0.11458	接受 H_0
Δ EX 不是 Δ INF 的 Granger 原因	1.09535	接受 H_0
Δ R3 不是 Δ EX 的 Granger 原因	0.80953	接受 H_0
Δ EX 不是 Δ R3 的 Granger 原因	1.42885	接受 H_0
HIBOR_N 不是 Δ EX 的 Granger 原因	2.72823**	HIBOR_N \rightarrow Δ EX
Δ EX 不是 HIBOR_N 的 Granger 原因	6.64723**	Δ EX \rightarrow HIBOR_N
HIBOR_M 不是 Δ EX 的 Granger 原因	1.08143	接受 H_0
Δ EX 不是 HIBOR_M 的 Granger 原因	2.15166*	Δ EX \rightarrow HIBOR_M

注：** 表示在 5% 的置信水平下显著; * 表示在 10% 的置信水平下显著。

月度数据反映中期的变化趋势, 而汇率和同业拆借利率都是高频率变化的, 因而我们进一步采用每日数据进行分析 (见表 4)。每日数据来源于香港金融管理局的网页, 并且根据交易日的差别得到了调整。结果是, 无论是隔夜拆借利率还是隔月拆借利率, 它们与汇率变化之间的相互影响都得到了检验结果的强有力支持。短期高频数据的检验结果说

明，同业拆借利率和汇率变化彼此之间存在非常灵敏的反应。更重要的是，汇率面对同业拆借利率能迅速产生调整反应，这非常有利于当局控制同业拆借利率来稳定汇率。

表4 每日数据的检验结果

ADF 检验					
变量	ADF 统计量	检验形式 (C, T, L)	变量	ADF 统计量	检验形式 (C, T, L)
EX	- 2.569441	C, N, 5	Δ EX	- 32.41588**	C, N, 5
HIBOR-N	- 10.66279**	C, N, 5	HIBOR-M	- 5.443657**	C, N, 5
Granger 因果检验					
零假设 H_0		F 统计量	结论		
HIBOR_N 不是 Δ EX 的 Granger 原因		1.9538**	HIBOR_N \rightarrow Δ EX		
Δ EX 不是 HIBOR_N 的 Granger 原因		48.8797**	Δ EX \rightarrow HIBOR_N		
HIBOR_M 不是 Δ EX 的 Granger 原因		5.86734**	HIBOR_M \rightarrow Δ EX		
Δ EX 不是 HIBOR_M 的 Granger 原因		23.7060**	Δ EX \rightarrow HIBOR_M		

值得注意的是，在最近的三年里，香港一直处于通货紧缩的境况，货币供应也开始下降，趋于收缩，而同期的港元汇率也非常接近 7.8 的水平。通过对 2000 年 1 月至 2002 年 12 月的子样本进行检验（结果从略），我们发现它们之间仍然不存在因果关系。看来 1998 年亚洲金融危机过后，港元汇率就恢复了高度稳定，但危机对香港经济的打击却很大，消费随之下挫，通货出现较长时间的紧缩。总而言之，通货因素不是影响港元汇率的因素。

五、结论与展望

我们考察了以联系汇率制为核心的香港货币政策，并着重分析了可能影响港元汇率的传统货币政策所涉及的几个因素。在香港的特殊经济和政策环境下，通货膨胀的压力和货币供应量的独立增长均不是导致港元汇率不稳定的因素，同业拆借利率与港元汇率之间却表现出有力的反馈关系，而同业拆借利率正是受香港货币政策工具控制的。对于稳定汇率这个目标来说，香港的货币政策是非常合适的，而且也把握住了维持汇率长期稳定的关键因素。但问题还远没有结束。有效的货币政策虽然成功地维护了港元汇率的稳定，这并不意味着联系汇率制就高枕无忧了。我们的分析从侧面反映了这样的问题，影响港元汇率稳定的并非通胀、货币供应等这些本地内部因素，而主要是外在的不稳定因素所造成的港元信心危机。在过去，港元因美元的贬值而受益，而近来美元汇率的不断下跌迹象是否会再次为香港经济的全面复苏提供机会？如果亚洲金融危机重现，我们是否能继续以现有的政策来坚守港元的稳定？香港与内地今后的经济交流和融合会将港元与人民币之间的关系引向何方？等等。这些都是我们今后需要深入研究的课题。

参考文献

- [1] 林江著：《过渡时期香港的金融管理》，四川人民出版社 1995 年版。
- [2] 林江著：《香港的金融机构与金融市场》，四川人民出版社 1993 年版。
- [3] 魏巍贤：《人民币汇率决定模型的实证分析》，《系统工程理论与实践》，2000 年 3 月。
- [4] 刘晓勇：《谈谈香港的货币政策》，《经济体制比较》。
- [5] 姜波克，陆前进编著：《汇率理论和政策研究》，复旦大学出版社 2000 年版。
- [6] 钱荣堃，陈平等主编：《国际金融》，南开大学出版社。
- [7] 刘泽生等主编：《香港大辞典》，广州出版社 1994 年版。
- [8] 陆懋祖著：《高等时间序列计量经济学》，上海人民出版社。
- [9] 《金融管理局年报》（1996~2001）香港金融管理局。
- [10] 《金融管理局季报》（1996~2001）香港金融管理局。
- [11] Sheng Andrew, The Linked Exchange Rate System: Review and Prospects, Hong Kong Monetary Authority Quarterly Bulletin, May 2000.
- [12] Greenwood John, Advantages of the Current Linked Rate System for Hong Kong Dollar: A Restatement, HKCER Letters, May 1999, Vol. 15.
- [13] Richard Yan-Kiho, Robert Hanay Scott and Kite Ann Wong, The Hong Kong Financial System, Oxford University Press, Hong Kong, 1998.
- [14] Y. C. Jao, F. H. King, Historical Perspective and Contemporary Analysis of Hong Kong, 1999, pp135 ~ 136.
- [15] Taylor, J., The Rules of Exchange Rates in Monetary Policy Rules, American Economic Review, May 2001.

金融结构历史演变的初步考察*

劳 平

在整个 20 世纪，经济学家关于金融结构与经济发展关系的研究，始终围绕着两种基本结构——银行主导型和市场主导型金融结构——在推动金融和经济发展上的优劣之分的争论而展开。然而，这样的争论把重点放在国别间的横向比较，对一国金融结构的历史演变缺乏应有的关注。基于这样的考虑，本文试图从历史的角度出发，考察一国金融结构的演变轨迹，以便更全面地对金融结构与经济发展之间的关系展开分析。本文分两部分：首先简要介绍已有的横向比较研究成果，接着结合美国的情况，从节约交易成本的角度纵向考察一国金融结构的演进轨迹，并得出有关结论。

一、对金融结构与经济发展关系的横向研究的简要介绍

金融体系的基本功能是促进金融资源从盈余部门转移到不足部门，即通过金融资源的动员和转移，实现收入向储蓄和储蓄向投资的转化。金融结构反映了实现金融资源动员和转移的各类市场、各种机构和交易工具在整个金融体系中的相对规模、地位及其对经济增长所起的作用。有两种金融结构最具普遍性：一种主要通过金融中介体来实现金融资源的动员和转移，被称为银行主导型金融结构，以德国和日本为代表；另一种主要依靠证券市场来聚集和配置金融资源，被称为市场主导型金融结构，以英国和美国为代表。在金融结构和经济增长与发展的关系上，无论是历史学家、经济学家还是政策制定者都对银行主导型和市场主导型两类金融结构给予了足够的关注。研究主要围绕着两种金融结构孰优孰劣而展开，以横向比较研究为主。

（一）银行主导论

银行主导论者认为，银行主导型金融体系在以下几方面占有优势：首先，金融中介体可以降低投资者获得及处理企业和管理者信息的成本，因而有利于资源配置和企业控制 (Diamond, 1984; Boyd and Prescott, 1986)。其次，金融中介体降低了投资者资产分散化的交易成本，有利于风险共担。不仅如此，金融中介机构还可以减少跨期风险 (Allen and Gale, 1999) 以及流动性风险 (Diamond and Dybvig, 1983; Bencivenga and Smith, 1991)。再次，通过节约与储蓄的聚敛和信息不对称相关的成本，金融中介体促进了储蓄的流动和集中 (Sirri and Tufano, 1995; Lamoreaux, 1995)。

但是，对银行主导型金融体系持批评态度的学者指出它的主要弱点是不利于创新性企

* 原载：《国际金融研究》，2003 年第 6 期。

业和项目的发展。在标准化的环境下，银行在收集和处理企业信息方面具有优势，但在非标准化（即涉及创新过程的不确定性）的环境下，银行往往不如证券市场有效率（Allen & Gale, 1999）。另外，由于银行对企业的特殊地位，银行对企业的监控可能倾向于非效率。

（二）市场主导论

市场主导论则强调功能完备的证券市场尤其是股票市场在激励投资者获取信息、行使对企业的监控、风险管理和促进资本的流动等方面都发挥着重要的作用。股票市场规模越大、流动性越强，当事人就越容易通过证券交易使其获得的有价值的信息迅速转变为收益，从而促进市场参与者努力去收集企业的信息。股票市场的接管压力也有助于实施对企业融资之后的财务监控。在风险管理方面，投资者可以构造证券组合以实现风险分散化的目标，同时二级市场的存在降低了流动性风险（Levine, 1991）。Levine 和 Zervos 1998 年的论文特别指出是市场的流动性而非规模对长期增长有重要作用。股票市场上有关会计准则及信息披露等方面的规定也鼓励资源流向股票市场。

对市场主导型金融结构的批评主要集中在信息外部性和企业监控两方面。在一个有效市场上，个别投资者不愿花费时间和资源去搜集信息，如果其他投资者也能搭便车从这些信息中获利的话（Stiglitz, 1985）。不仅如此，内部人和外部人之间的信息不对称降低了并购威胁的效力，使市场监控企业的作用下降。流动性市场也为投资者提供了方便的退出选择，因而削弱了他们监控企业的动力（Shleifer and Vishny, 1986）。

（三）金融服务论

在讨论中，一些学者不赞成把银行主导型和市场主导型金融结构对立起来的观点，提出了金融服务论。金融服务论的观点认为，银行也好，证券市场也好，其基本功能都是为企业和产业的发展提供金融服务。金融服务主要体现在储蓄的动员和转移、项目评估、企业监控和风险管理等方面。首要的是金融服务的可得性及质量，而非金融体系的结构。

金融服务论进一步指出，在提供金融服务上，银行和市场两者之间是互补而非替代的关系。例如，证券市场的存在减少了企业对银行的过分依赖，同时使投资者获得一条退出通道。

（四）法律主导论

法律主导论与金融服务论相近，由以 La Porta 为首的四位美国经济学家和法学家提出。他们同样反对单方面推崇银行主导或市场主导的金融结构，但是与金融服务论不同的是，他们没有停留在金融体系所提供的服务层面，而是深入到一国法律体系的渊源及其特征与各国的金融发展及金融结构之间的联系上。他们认为，一方面，一国金融结构的差异无法解释企业成长和经济增长；另一方面，金融结构的差异与各国法律体系的渊源有着密切的联系。英国普通法（Common law，又称英美法）更重视小股东的权利，而法国民法（Civil law，罗马法的主要流派）则较为强调债权人的权利。因此，市场主导型金融结构拥有普通法法律传统的可能性更大，而银行主导型结构的国家拥有法国民法法律渊源的概率更高。

（五）小结

上述研究以横向为主，对金融结构和经济发展之间的关系，不仅在理论上从不同角度进行了较为深入的剖析，也获得了一定的实证支持。然而，正如 Levine 在其 1997 年的论文中指出的那样：“在对金融体系的演化和功能有清晰的认识之前，我们不可能对长期经济增长有完整的理解。”因而要对金融结构和经济发展的关系有全面的认识，还必须考察金融结构的历史演变。

二、对金融结构历史演变的初步考察

（一）相关研究

随着对金融结构与经济发展关系的深入探讨，金融结构的演进也逐渐被关注。事实上，在跨国比较研究中，也已透露出金融结构演变的某些线索。例如，DemirgucKunt 和 Levine (1999) 及 Levine (2000) 的研究表明，在有较高收入水平的国家，证券市场相对于银行部门更加活跃和更有效率，并且保险公司、养老基金、共同基金以及其他非银行金融中介倾向于变得更大。这不禁让人思考，是否随着经济的发展，金融体系沿着从银行主导型向市场主导型的路径演进？

早在 1966 年，Patrick 就从金融发展与经济增长之间的因果关系出发，提出了金融发展“供给导向”和“需求导向”之间的区别，为金融结构的演变提供了一个有用的参考框架。Patrick 认为，“供给导向”的金融发展先于对金融服务的需求，因而对经济增长有着自主的积极影响，对动员那些阻滞在传统部门的资源，使之转移到能够促进经济增长的现代部门，并确保投资于最有活力的项目方面，起基础性的作用。“需求导向”的金融发展是实际经济部门发展的结果。市场的不断拓宽和产品的不断增长，要求更有效地分散风险以及更好地控制交易成本，因此金融发展在经济增长过程中起了一个更好的推动作用。Patrick (1996) 提出的假设是“供给导向”的金融发展对早期的经济发展有着支配的作用，特别是当它能更为有效地包含技术创新的投资提供资金时更是如此。一旦经济发展趋于成熟，“需求导向”的金融发展就该发生作用了。另外，发展中国家同发达国家之间的差距越大，则越有可能遵循“供给导向”的金融发展模式，这一假设被格申克龙 (1962) 的研究结果以及经历了后工业化的一些国家（例如日本）的实践所证明。

Claudia Dziobek 和 John K. Garrett (1998) 提出了金融系统的趋同问题，美国转向全能型银行，而德国加强了其市场导向的融资，这两种金融系统看起来好像要趋同于一种共同的中间模式。大部分关于金融系统的文章都或含蓄或明确地提出，金融市场与全能银行不存在兼容性问题，这更增加了出现趋同的可能性。

国内首先对金融结构演变进行系统研究的是白钦先教授。白教授在其专著《比较银行学》中，对金融体系结构变迁的演进规律进行了归纳和总结，提出一国金融结构的演变实际上是一个“金融倾斜及其逆转”的过程。金融体系的核心功能是如何将资金从盈余部门转移到短缺部门，实现由储蓄向投资的转化。这一过程当中存在两种融资方式：间接融资

和直接融资。就世界各国金融业发展的历史而言，一般是间接金融发展在前，直接金融发展在后，并且在业务总量和市场份额方面，在相当长的历史时间内，间接金融的业务总量及市场份额大大超过了直接金融。白教授将间接金融与直接金融的这种不平行和不对称发展称为金融倾斜。随着人类社会生产力发展水平的不断提高、科学技术日新月异的变革以及经济金融的充分作用与高度发展，传统的以间接金融为主的金融格局出现了新的变化，即直接金融的发展速度大大超过了间接金融，以致在绝对量和市场份额方面逐渐接近乃至超过间接金融。间接金融和直接金融在发展速度和相对份额上发生的这种反向变化被称金融倾斜的逆转。

唐寿宁、王晋斌在《投资者选择与金融系统演变》中指出，银行主导型和市场主导型金融结构的划分很大程度上只注意到了金融系统的外在形式，还不能揭示决定不同金融系统模式的内在因素。金融系统模式的演变实际上是通过投资者对日常金融产品、金融中介的选择进行的。正是由于投资者在不同种类金融产品、不同风格金融中介之间的流动，不断地改变着各类金融中介在金融系统中的各自位置，从而推动着金融系统模式的不断演变。依据投资者的风险感受度以及对风险管理的需求，他们把金融系统划分为政府主导型和中介主导型，而中介主导型又依据具体的风险管理方式划分为保守中介（主要指银行）主导型和自由中介（主要指市场）主导型。

上述研究均在一定程度上从不同的角度揭示了一国金融结构的演进规律。Patrick (1966) 的研究虽然不是直接针对这个问题，而是重点放在揭示金融发展和经济增长之间可能存在的因果关系上，但他所提出的“供给导向”和“需求导向”的金融发展为解释金融结构的变迁提供了很好的框架，因而是很有益的。不过，它也是不完全的，因为他没有把两种情况的互补性考虑进去，事实上，“供给导向”的金融发展可以加速经济增长，而“需求导向”的金融发展，也不仅仅是金融体系被动适应实际部门的发展需求，实际的经济增长使得金融体系能够完成自身的自由发展。因为实际收入的持续增加，为昂贵的和应用日益广泛的金融中介的建立提供了支持。

Claudia Dziobek 和 John R. Garrett (1998) 提出金融系统的趋同问题值得关注。金融结构的变迁是否朝着同一方面？如果是，这个相同的方向又是什么？Claudia Dziobek 和 John R. Garrett 认为是市场和银行的调和。白钦先教授将金融结构的变迁归结为“金融倾斜及其逆转”的过程，实际上也是持金融结构的演进会趋同的观点，只不过白教授认为趋同的是市场主导型的金融结构。唐寿宁、王晋斌着重研究了中国金融系统的演化，认为目前中国的金融系统呈现出政府 - 中介混合型模式的特征，今后更大的可能是在现有的政府 - 中介混合主导型的模式上深化演化。笔者认为，仅仅从投资者选择的观点去考虑金融结构的演进是不够的，它的积极意义在于提供了微观角度的解释，但忽视了经济和金融之间的互动，并且可以将之归入 Patrick 所提出的“需求导向”的金融发展当中。

（二）对金融结构演进轨迹的初步解释

综合上述研究，笔者认为，可以以交易成本为突破口，对金融结构的演进轨迹给出一个初步解释。金融系统的演进是各种因素综合作用的结果，但是其中应该存在着某种内在因素，推动着金融系统向着某一方向发展。尽管在不同国家或同一国家的不同阶段，金融

系统的演变或者受到政府强有力的干预，如德国、日本和大部分发展中国家经历的那样，也可能金融系统自然演变的成分更大，如英国和美国，但是，长期来看，金融系统的演变很可能朝着节约交易成本的方向前进。

如前所述，金融体系的基本功能主要体现在储蓄的动员和转移、项目评估、企业监控和风险管理等方面。而这些功能的发挥对应着一定的交易成本。金融结构的演进取决于交易成本的大小。进一步看，交易成本主要与以下几个因素有关。

(1) 产业结构。Allen (1992) 提出，银行只能提供“一元审查”机制，而市场能提供“多元审查”机制，即金融市场通过多种机制（市场价格、交易量、并购企图等等）来提供多重检查。由此可见，当最优的决策规则难以形成并且对公司如何运作存在争议时（比如说，在信息交替更新很快的情况下），金融市场将配置资本。Allen 的观点与英国 19 世纪（当时英国第一个完成工业革命）股票市场的重要地位相吻合，也符合美国（当时美国是第一个完成后工业革命的国家）最依赖股票市场的事实。借鉴 Allen 的观点，笔者认为，在一国发展相对熟悉的传统工业时，银行的“一元审查”机制所需的交易成本相对较低，银行的发展将优先于金融市场的发展；而在发展有争议的新兴市场时，市场的重要性将会增加。这就解释了为什么绝大部分不发达国家以及后发工业国首先发展银行主导型的金融结构，因为银行主导型金融体系能为传统工业提供便宜的资金来源，这类似于 Patrick (1966) 提出的“供给导向”的金融发展。这是金融结构的初级阶段，通常表现为银行主导型的结构。

(2) 人均 GDP。人均 GDP 是衡量一国经济发展阶段的重要指标。当经济发展到一定的阶段，随着新兴产业的兴起、产品的多样化和全球化以及人均财富的积累，资金盈余者和短缺者更加细分，需求呈现多样化的特征，经济对金融提出了更高的要求，金融体系在推动经济增长和发展的同时，也必须适应经济发展阶段的要求，趋向于更加自由的分权的市场主导的金融结构。相对于这样的经济发展阶段，市场主导型的金融结构更能节约交易成本。这类似于 Patrick (1966) 提出的“需求导向”的金融发展。这是金融结构演变的第二阶段，银行和市场都在发展，但市场的成分相对加重。

(3) 金融业自身的发展特点。银行在经济和金融发展初期呈现出较强的规模经济效益，而当银行扩张到一定程度时，将进入规模不经济阶段。换言之，银行的长期成本曲线呈 U 形。当银行规模增长到一定程度时，必然会出现银行和市场的融合。正如大量研究所显示的那样，银行和市场各有千秋，它们之间应当是互补而非替代的关系。它们的融合将有助于降低交易成本，提供更优质的金融服务。这将是金融结构演变的第三阶段。

(4) 政策法规。政策法规是影响金融体系的外部力量。这里，笔者不同意法律主导论将法律因素视为解释各国金融结构差异的最重要因素的观点。政府的有关政策虽然能在一定程度上加速或减缓金融系统演进的速度，但是长期来看，政府维持较高的交易成本，金融系统的成本太大，金融体系的演进最终沿着内在因素的驱动前进。也就是说，政府对本国金融系统的设计也应该遵循金融体系自身的发展规律。

这样，笔者以交易成本为突破口，将一国金融结构的演进大致描述为从银行主导型到市场主导型再到相互融合的阶段。以美国为例，1930 年以前，贷款类资产的比重大于证券类和基金类，金融结构为银行主导型，为第一阶段；1940~1960 年，贷款类资产的比重

大大低于证券类，市场的增长势头超过了银行，为第二阶段；1960年以后为银行和市场相融合的第三阶段，贷款类资产虽然仍低于证券类资产，但比重有所回升，基金类资产比重稳步上升。不仅如此，随着金融资产品种的丰富，货币类资产占金融资产总额的比重从1900年的48%一直下降到2000年的12%。以上对美国百年以来金融结构变迁轨迹的划分虽然略显粗糙，仍需结合美国具体的经济金融状况作进一步的细分，也仍需对世界其他国家展开研究，但它起码表明，把银行和市场纯粹对立起来的观点是不可取的，并且在金融结构的演变过程中，政府的作用有一定的局限性，最终银行或证券市场发挥主导作用，取决于哪种金融结构能为相应的经济发展阶段提供较低的交易成本。

参考文献

- [1] Ross Levine, Bank-Based or Market-Based Financial Systems: Which is Better? World Bank Working Paper, mimeo, 2000.
- [2] Patrick H. T., Financial Development and Economic Growth in Underdeveloped Countries, Economic Development and Cultural Change, 1966, Vol. 14, pp174~189.
- [3] Asli Demirguc-Kunt and Ross Levine, Bank-Based and Market-Based Financial Systems: Cross-Country Comparisons, World Bank Working Paper, mimeo, 1999.
- [4] Claudia Dziobek and John R. Garrett, Convergence of Financial Systems and Regulatory Policy Challenges in Europe and in the United States, 1998. 转引自北京奥尔多投资研究中心主编:《金融系统演变考》,中国财政经济出版社2002年版,第299~321页。
- [5] Allen, F., Stock Markets and Resource Allocation in Capital Markets and Financial Intermediation, In C. Mayer and X. Vives (eds), Cambridge University Press, Cambridge, 1992.
- [6] 白钦先主编:《比较银行学》,河南人民出版社1989年版。
- [7] 唐寿宁,王晋斌:《投资者选择与金融系统演变》,见:北京奥尔多投资研究中心主编:《金融系统演变考》,中国财政经济出版社2002年版,第1~21页。
- [8] [荷]尼尔斯·赫米斯,罗伯特·伦辛克主编:《金融发展与经济增长——发展中国家(地区)的理论和经验》,余昌森等译,经济科学出版社2001年版。
- [9] 张捷等著:《金融结构与经济发展》,中国科学文化出版社,香港教科文出版有限公司2001年版。

中国股票市场收益非线性相关结构的经验分析*

何兴强

一、引言

在 20 世纪 70 年代,关于时间序列的普遍分析工具,是以 Box-Jenkins 方法为代表的 ARMA 模型,其基础是时间序列数据遵循传统的线性关系。

近 20 年来,关于时间序列非线性相关结构的研究取得了巨大进展。Engle 在 1982 年首创了自回归条件异方差 (ARCH) 模型, Bollerslev (1986) 在此基础上又创立了广义自回归条件异方差 (GARCH) 模型。此后, ARCH 和 ARCH - M 族模型在理论上不断完善,被广泛运用于分析各类经济金融时间序列,现已发展成描述时间序列非线性相关结构的典型方法之一。

在对股票市场的实证研究方面, ARCH 和 GARCH 模型首先被成功地运用到分析美国股票市场,然后,这种方法迅即被广泛应用于分析其他股票市场 (Bollerslev et al., 1992)。近年来,学者们也展开了利用 ARCH 和 ARCH - M 族模型分析我国股票市场的研究:魏巍贤、周晓明 (1999) 利用 GARCH、QGARCH 和 GJR 模型,以 1992 ~ 1998 年上证综合和深证成分周收盘价为对象,分析了我国股票市场的波动特征;刘国旗 (2000) 也运用 GARCH、QGARCH 和 GJR 模型研究了 1992 ~ 1998 年上证综合和深证成分周收盘指数的波动;汤果、何晓群、顾岚 (1999) 利用 FIGARCH 模型对上海和纽约股市的方差持久性进行了探讨和比较;张思奇、马刚、冉华 (2000) 利用 ARMA - GARCH - M 模型研究了上证 A 股综合指数;陈浪南、黄杰鲲 (2002) 利用 GJR - M 模型研究了中国股票市场的非对称性波动。何兴强 (2003) 曾以 GARCH 模型为基础研究过上证指数收益和波动性的星期效应,并运用 EGARCH、EGARCH - M、GJR 和 GJR - M 模型,探讨了中国股票市场的非对称性波动和风险 - 收益权衡。但是,就普遍而言,人们在分析过程中对 ARCH 和 ARCH - M 族模型的选用都具有一定的随意性。而 ARCH 和 ARCH - M 族模型的选用,意味着时间序列的非线性结构服从不同的类型。

我们可以注意到,从模型选择 (对于给定的一个数据样本集合,在备择模型集合中选择哪一个作为该样本的生成模型) 的角度,以上研究主要是从利用可以概括多个备择模型

* 本文为作者博士学位论文《中国股票市场可预测性研究》第二章的主要内容,在第三届中国青年经济学者论坛上被选为交流论文。本文作者感谢王则柯教授和王美今教授的指导。本文的研究得到国家自然科学基金重点项目 (10131030) 和高等学校全国优秀博士学位论文作者专项资金 (200267) 的资助。原载:《世界经济》,2004 年第 8 期。

的“广义模型”进行拟合的方式，来探讨模型选择问题的。关于股票市场收益非线性相关性的研究，国内管理科学与工程领域的学者也作了许多工作，这里不再枚举。在运用“广义模型”途径探讨模型选择方面，柯珂、张世英（2003）的文献在国内比较典型，他们通过拟合 30 余个模型来探讨模型选择的问题。

几乎在 ARCH 模型诞生的同时，关于时间序列非线性结构的统计检验也取得了巨大进步。McLeod 和 Li 在 1983 年创立的 McLeod-Li 检验，是诊断时间序列非线性相关性最早的统计检验。1987 年，Brock, Dechert 和 Scheinkman 又创建了 BDS 检验。特别地，BDS 检验对于序列不相关但却不独立的情况表现出良好的检验效果（Brock, Hsieh and Lebaron, 1992）。但上述两种检验还不能甄别时间序列非线性相关结构的类型：不能识别时间序列的非线性相关究竟是确定性相关还是随机性相关，也不能区分序列的随机非线性相关性究竟是积式相关还是和式相关。Hsieh 检验在一定程度上填补了这一空白，可以用来识别序列随机非线性相关的积式相关或和式相关性（Hsieh, 1989）。目前，利用统计检验方法诊断我国股票市场收益非线性相关的存在性和结构的研究还相对匮乏：戴国强、吴祥林（1999）曾尝试利用 BDS 检验诊断我国股市收益非线性相关结构的存在性，目前我们还尚未发现利用 Hsieh 检验甄别积式相关或和式相关的研究。

本文主要利用统计检验，研究中国股票市场收益非线性相关结构的存在性和基本类型，并试图进一步对考察中国股票市场收益的非线性相关问题时所采用的模型进行选择。本文通过假设检验来探讨模型选择这一做法，在国内还是很少的。

文章的研究思路是：首先利用 ARMA 模型“过滤”股市收益序列的线性相关成分，然后针对“过滤”后的股市收益序列进行统计检验。McLeod - Li 和 BDS 检验表明，中国股票市场收益存在显著的非线性相关性。在此基础上，Hsieh 检验表明股市收益的随机非线性相关性主要表现为积式相关、和式相关性不显著。最后，我们再对积式相关代表性模型（ARMA - GARCH）的拟合残差序列进行 McLeod - Li 和 BDS 检验，检验结果进一步证实了 Hsieh 检验的结论，并通过拟合和式相关的代表性模型 ARMA - GARCH - M 对统计检验结果进行验证。中国股票市场收益存在显著的非线性相关结构，股市收益的随机非线性相关性主要表现为积式相关。这也说明我国股票市场的投机成分明显，不存在显著的风险 - 收益权衡。

本文的结构安排是：第二部分描述随机非线性模型；第三部分叙述诊断时间序列非线性相关结构的几个检验，包括 McLeod-Li, BDS 和 Hsieh 检验；第四部分是数据说明和基本统计分析；第五部分对股市收益非线性相关的存在性和结构进行统计检验；第六部分在对 ARMA - GARCH 拟合残差序列进行 McLeod - Li 和 BDS 检验的基础上，进一步拟合了和式相关的代表性模型 ARMA - GARCH - M，对统计检验的结果进行验证；最后是实证结论。

二、随机非线性模型

当股市收益源于一个线性随机过程时，人们通常可以用 ARMA 模型描述它的演变。Box, Jenkin 和 Reinsel（1994）指出，任何一个不带常数项的稳定线性随机过程可以用一个 ARMA (p, q) 模型任意逼近：

$$r_t = \sum_{j=1}^p \phi_j r_{t-j} + \sum_{k=0}^q \theta_k u_{t-k} \quad (1)$$

其中：新息 u_t 代表白噪声：对任意 t 和 $j \neq 0$ ， $E(u_t) = 0$ ， $\text{var}(u_t) = \sigma^2$ ， $\text{cov}(u_t, u_{t-j}) = 0$ 。

人们普遍使用 Box-Jenkin 方法诊断 ARMA 模型，该方法主要依托于模型残差序列的自相关和偏自相关函数：如果残差序列不存在显著的自相关，则可认为已经挑选了一个合意的 ARMA 模型。此时，从模型残差的自相关和偏自相关函数角度，我们不能拒绝残差序列服从白噪声的假定。但如果平方残差序列表现出系统演变，则新息 u_t 不服从 IID 分布，不能视为白噪声。

扩展可以用 ARMA 模型任意逼近的线性模型，关于收益序列一般化的随机非线性模型通常可表示为：

$$r_t = f(u_t, u_{t-1}, u_{t-2}, \dots) \quad (2)$$

式中： r_t 表示股市收益；新息 u_t 服从独立同分布 (IID)。在实际应用时需要对上述一般化模型赋予一定的限制。下面的非线性模型是一种应用广泛的形式 (Campbell, Lo and Mackinlay, 1997)：

$$r_t = g(u_{t-1}, u_{t-2}, \dots) + u_t h(u_{t-1}, u_{t-2}, \dots) \quad (3)$$

其中：函数 $g(\cdot)$ 代表条件均值： $E(r_t | \Omega_{t-1}) = g(u_{t-1}, u_{t-2}, \dots)$ ； $h^2(\cdot)$ 代表条件方差： $\text{var}(r_t | \Omega_{t-1}) = h^2(u_{t-1}, u_{t-2}, \dots)$ 。

Ω_{t-1} 代表 $t-1$ 时刻的信息集。方程(3)提供了区分积式相关与和式相关的依据。积式相关意味着条件方差具有非线性结构， $h^2(\cdot)$ 是一个非线性函数。和式相关意味着条件均值具有非线性结构， $g(\cdot)$ 为一个非线性函数。

设条件方差可以用一个 GARCH(p, q) 过程描述，则经过 ARMA 模型“过滤”后的股市收益 e_t 服从条件正态分布，均值为零： $E(e_t | \Omega_{t-1}) = 0$ ，条件方差为：

$$\text{var}(e_t | \Omega_{t-1}) = \sigma_t^2 = \alpha_0 + \sum_{j=0}^q \alpha_j e_{t-j}^2 + \sum_{k=0}^p \gamma_k \sigma_{t-k}^2 \quad (4)$$

模型 (4) 实际上是非线性模型 (3) 在 $g(\cdot) = 0$ ， $h(\cdot) = \sigma_t$ 时的特例。金融时间序列通常可用 GARCH (1, 1) 过程描述。

GARCH 模型只反映了收益序列的积式非线性相关性，如果还要刻画收益序列的和式非线性相关性，可以把 GARCH 模型进一步拓展为 GARCH-M 模型。在 GARCH-M 模型中，股市收益 e_t 可以表示为：

$$e_t = \beta_0 + \beta_1 \sigma_t + u_t \sigma_t \quad (5)$$

该模型是非线性模型 (3) 在 $g(\cdot) = \beta_0 + \beta_1 \sigma_t$ 和 $h(\cdot) = \sigma_t$ 时的特例。根据 GARCH-M 模型的设计，(5) 式中参数 β_1 是风险-收益权衡，或投资者相对风险规避系数的度量。

三、非线性相关性检验方法

(一) McLeod - Li 检验

在 ARMA 残差 e_t 不存在自相关性时,平方残差序列仍可能显著自相关。Granger 和 Andersen (1978) 指出,可利用 ARMA 平方残差的自相关函数来诊断时间序列的非线性相关性:如果平方残差序列显著自相关,则原时间序列具有非线性结构。

McLeod 和 Li (1983) 以 ARMA 平方残差序列的自相关函数为基础,在残差序列正态白噪声的零假设条件下,构建了诊断序列非线性相关性的 McLeod - Li 统计量:

$$Q_{ec}(k) = n(n+2) \sum_{j=1}^k r_{ec}^2(j)/(n-j) \quad (6)$$

其中: n 是 ARMA 残差序列的数据个数; $r_{ec}(j)$ 是平方残差序列的自相关函数:

$$r_{ec}(j) = \frac{\sum_{t=j+1}^n (e_t^2 - \sigma^2)(e_{t-j}^2 - \sigma^2)}{\sum_{t=1}^n (e_t^2 - \sigma^2)^2} \quad j = 1, 2, \dots, k \quad (7)$$

McLeod - Li 统计量 $Q_{ec}(k)$ 收敛于自由度为 k 的 χ^2 分布。

蒙特卡罗实验表明,当样本规模在 50 到 200 之间时,McLeod - Li 检验倾向于过多接受零假定 (McLeod and Li, 1983)。随着样本规模的增大,该统计量的检验效果会得到改善,但正态分布假设仍然比较关键 (Cromwell, Labys and Terraza, 1994)。同时, $Q_{ec}(k)$ 统计量中参数 k 的选择不惟一。

(二) BDS 检验

BDS 统计量也可以用来诊断 ARMA “过滤”后的股市收益序列是否存在非线性相关性。BDS 检验的零假设是独立同分布 (IID),备择假设没有具体化。BDS 统计量的定义如下:

首先,根据序列 $\{e_t\}_{t=1}^T$ 构造一个 m 维嵌入空间 $e_t^m = (e_t, e_{t+1}, \dots, e_{t+(m-1)})$,这个空间常称为序列的 m - 历史。记 $T_m = T - m + 1$,则嵌入空间中的向量可构成 $T_m(T_m - 1)/2$ 对不同的组合。在此基础上相关积分定义为:

$$C_m(\epsilon, T) = \frac{2}{T_m(T_m - 1)} \times \sum_{1 \leq t < s \leq T_m} I_\epsilon(e_t^m, e_s^m) \quad (8)$$

其中: I_ϵ 为指示函数:

$$I_\epsilon(e_t^m, e_s^m) = \begin{cases} 1 & \text{如果 } \|e_t^m, e_s^m\| < \epsilon \\ 0 & \text{其他} \end{cases}$$

式中: $\|\cdot\|$ 表示上确界范数: $\|e_t^m, e_s^m\| = \max\{|e_t - e_s|, \dots, |e_{t+(m-1)} - e_{s+(m-1)}|\}$

相关积分 $C_m(\epsilon, T)$ 的取值在 0 到 1 之间,直观含义是嵌入空间中两点之间距离小于的概率,反映了连续 m 个观察值的集中程度。例如,如果对任意 t 和 s ,向量 e_t^m 和 e_s^m 之间的距离

都大于或等于 ϵ 则相关积分 $C_m(\epsilon, T)$ 取到最小值 0。

在相关积分的基础上, BDS 统计量定义为 (Brock et al., 1996):

$$BDS(m, \epsilon) = \sqrt{T_m} [C_m(\epsilon, T) - C_1(\epsilon, T)^m] / \sigma_m(\epsilon, T) \quad (9)$$

在独立同分布条件下, 当 $T_m \rightarrow \infty$ 时, $BDS(m, \epsilon)$ 收敛于标准正态分布 $N(0, 1)$ (Brock et al., 1996, 定理 2)。其中, $\sigma_m(\epsilon, T)$ 满足:

$$\begin{aligned} \frac{1}{4} \sigma_m^2 = & m(m-2)K_1^{2m-2}(K_2 - K_1^2) + K_2^m - K_1^{2m} \\ & + 2 \sum_{j=1}^{m-1} [K_1^{2j}(K_1^{m-k} - K_1^{2(m-j)}) - mK_1^{2m-2}(K_2 - K_1^2)] \end{aligned}$$

K_1, K_2 的一致估计为:

$$K_1 = \frac{1}{T_m^2} \sum_{s=1}^{T_m} \sum_{t=1}^{T_m} I_\epsilon(e_s^m, e_t^m); \quad K_2 = \frac{1}{T_m^3} \sum_r \sum_{s=1}^{T_m} \sum_{t=1}^{T_m} I_\epsilon(e_r^m, e_s^m) I_\epsilon(e_s^m, e_t^m)$$

BDS 统计量显著为正或负都意味着非 IID 行为。例如, 如果两个 m -历史 $e_t^m = (e_t, e_{t+1}, \dots, e_{t+(m-1)})$ 和 $e_s^m = (e_s, e_{s+1}, \dots, e_{s+(m-1)})$ 相距“近”的概率, 大于 e_t 与 e_s 相距“近”的概率的 m 次方, 则 $BDS(m, \epsilon)$ 统计量为正, 即如股价波动“群集”这类模式比随机游走模式发生的频率高。

在实际应用中需要适当选取 ϵ 和嵌入空间维数 m 。其中 ϵ 是判断白噪声的一个标准: ϵ 取值太小会导致过多接受零假设, ϵ 取值太大则会高估嵌入空间中向量间的贴近程度 (Scheinkman and LeBaron, 1989)。对大样本 ($T > 500$) 情形, Brock, Hsieh 和 LeBaron (1992) 建议的取值为序列标准差的 0.5, 1.0 和 1.5 倍。在实证研究中, ϵ 的取值通常在序列标准差的 0.5 至 2 倍之间。

嵌入空间维数 m 是序列复杂程度的一个度量: 一个混沌过程为有限维, 而一个随机游走过程则为无限维。在实际应用中, 人们主要关心的是低维 (复杂程度较低的) 混沌, 高维混沌通常可视为和随机游走等价。Hsieh (1989) 建议嵌入空间维数 m 可以选择在 2 到 10 之间。但当 m 取值过大时, 相关积分的“真实值”可能被低估。Brock 等 (1991) 建议 m 的选取满足 $T_m/m \geq 200$, 这意味着只有在样本规模很大时才会出现取值过大。Barnett 等 (1995) 建议 m 的取值不要超过 8。综合上述讨论, 我们把嵌入空间维数 m 限制在 2 到 5 之间。

(三) Hsieh 检验

为了识别序列随机非线性相关的积式相关或和式相关性, Hsieh (1989) 考察了 ARMA 残差的条件期望。条件期望等于 0:

$$E(e_t | r_{t-1}, \dots, r_{t-h}, e_{t-1}, \dots, e_{t-h}) = 0 \quad (10)$$

意味着随机非线性相关性为积式相关, 非线性相关结构通过影响方差间接影响均值。相应地, 和式非线性相关意味着:

$$E(e_t | r_{t-1}, \dots, r_{t-h}, e_{t-1}, \dots, e_{t-h}) \neq 0 \quad (11)$$

直观含义是非线性相关性会直接影响预期收益。

Hsieh 检验具有明显的直观含义：如果视 ARMA 残差 $\{e_t\}$ 为滞后项 $r_{t-1}, \dots, r_{t-h} - t-h, e_{t-1}, \dots, e_{t-h}$ 的一个函数 f , 则函数 f 在零点附近的二次泰勒展开式中会出现诸如 $r_{t-j}r_{t-k}, r_{t-j}e_{t-k}$ 和 $e_{t-j}e_{t-k}$ 之类的项。这样, 积式相关意味着 e_t 和任何这类项之间不相关, 而和式相关则意味着 e_t 至少和其中某一项相关。

这样, 积式相关意味着残差序列 $\{e_t\}$ 的标准化三阶矩必然等于零。即：

$$\rho_{eee}(j, k) = E\left(\frac{e_t e_{t-j} e_{t-k}}{\sigma_e^3}\right) \quad (12)$$

必然等于零。和式相关意味着 (12) 式定义的期望值不等于零。其中 σ_e 是 ARMA 残差序列的标准差。在此基础上, Hsieh 统计量 (1989) 定义为：

$$H = \sqrt{\text{Tr}_{eee}(j, k)} / s_r(j, k) \quad (13)$$

其中：

$$r_{eee}(j, k) = \frac{1}{T} \sum e_t e_{t-j} e_{t-k} / \left(\frac{1}{T} \sum e_t^2\right)^{1.5}; \quad s_r^2(j, k) = \frac{1}{T} \sum e_t^2 e_{t-j}^2 e_{t-k}^2 / \left(\frac{1}{T} \sum e_t^2\right)^3$$

分别是三阶矩 $\rho_{eee}(j, k)$ 的一致估计和 $r_{eee}(j, k)$ 的方差估计。在积式相关条件下, Hsieh 统计量收敛于标准正态分布 $N(0, 1)$ 。蒙特卡罗试验反映出 Hsieh 统计量在积式相关与和式相关的识别方面有良好的鉴别力 (Hsieh, 1989)。

四、数据说明和基本描述统计

本文使用的数据是上证综合收盘指数、深证成分和综合收盘指数。样本区间为 1993 年 1 月 4 日至 2002 年 6 月 19 日。对每周股价水平的衡量采用了 Lo 和 MacKinlay (1988) 提出的方式。在金融实证研究中, 这种衡量周股价水平的方式经常被采用 (Aggarwai, Inclan & Leal, 1999; Darrat & Zhong, 2000)。具体地, 每周股价水平 P_t 以周三的股价指数代表, 如果周三股价指数缺损, 则用周二的股价指数代表, 如果周二的股价指数也缺损, 则用周四的股价指数代表, 如果周二和周四的股价指数都缺损, 则本周的股价水平视为缺损值处理。对缺损值的处理方法是以前一周的股价水平代表本周的股价水平。采用对数收益衡量方式, 股市周收益 $r_t = \log(P_t/P_{t-1})$ 。本文的实证工作在 Eview 3.1 环境下进行, 关于文中涉及的几个检验在 Eview 3.1 环境中编写程序完成。

股市周收益的基本描述统计如表 1 所示。均值和中位数反映出上证股市的平均收益高于深证股市, 标准差揭示出上证股市的波动也相应更大。偏度统计量反映出沪深股市周收益序列的分布密度函数向右偏, 上证股市收益密度函数向右偏的程度更高。峰度统计量揭示出沪深股市都呈现出显著的狭峰特征, 且上证股市更为显著。不论是对上证还是对深证股市, J-B 统计量都反映出周收益序列不服从正态分布。表 1 中的 Ljung-Box 统计量, 表明沪深股市周收益序列都具有显著的自相关性。

表 1 基本描述统计和 Ljung-Box 统计量

	上证综指	深证成指	深证综指
样本区间	1993 年 1 月 4 日至 2002 年 6 月 19 日		
观察值数	494	494	494
均值	0.001366	0.0005728	0.001278
最大值	0.3466	0.3139	0.3465
最小值	- 0.2467	- 0.2935	- 0.3168
标准差	0.0558	0.0514	0.0542
偏度	0.8128***	0.3750***	0.1993*
峰度	10.86***	9.56***	11.10***
J-B	226.29***	183.47***	175.16***
Q(5)	13.670**	11.323**	7.106
Q(10)	17.400*	22.518**	20.217**
Q(20)	29.265*	33.133**	30.903*
Q(30)	41.215*	48.947**	46.408**
Q(40)	49.740	68.083***	65.108***
Q(50)	59.819	77.471***	73.469**

注：偏度 (Skewness) : $S = \frac{1}{T} \sum_{t=1}^T \left(\frac{r_t - \bar{r}}{\sigma} \right)^3$ ，峰度 (Kurtosis) : $K = \frac{1}{T} \sum_{t=1}^T \left(\frac{r_t - \bar{r}}{\sigma} \right)^4$ ，J-B = $\frac{T-n}{6} (S^2 + \frac{1}{4} (K-3)^2)$ 服从 $\chi^2(2)$ ，滞后 s 阶的 Ljung-Box 统计量 $Q(s) = T(T+2) \sum_{k=1}^s \rho_k^2 / (T-K)$ 渐进服从分布 $\chi^2(2s)$ 。其中： r_t 代表股市收益序列；T 为样本观察值数； σ 代表估计标准差；n 代表估计参数个数； ρ_k 为第 k 阶自相关函数；*，**，*** 分别表示在 10%，5% 和 1% 水平下统计显著。

五、非线性相关性的实证检验

在检验股市收益的非线性相关性之前，我们先利用 ARMA 模型“过滤”股市收益的线性相关成分。根据 Box-Jenkins 方法的思想，在诊断 ARMA 模型时需要综合考虑两个方面：一是 AIC 信息准则^①，二是 Ljung-Box 统计量。具体来说，我们运用 AIC 信息值最小的 ARMA 模型来“过滤”股市收益的线性相关成分。对股市收益 r_t 的 ARMA 拟合如表 2 所示。 $Q_r(k)$ 统计量反映出股市收益的线性相关成分已经被成功消除。但是，不论是上证综合还是深证成分和综合收益，ARMA 模型“过滤”后的股市收益序列的 McLeod-Li 统计量 $Q_{rr}(k)$ 都表现为统计显著，这说明股市收益序列存在显著的非线性相关性。

① AIC 信息准则 (Akaike Information Criterion) 对应的统计量为 $-2l/T + 2n/T$ ，其中 n 为模型估计参数的个数，l 为模型的对数似然率，T 为观察值数。

表2 ARMA模型和McLeod-Li检验

	上证综指	深证成指	深证综指
ARMA模型	ARMA(7,3)	ARMA(4,3)	ARMA(6,5)
R ²	0.1248	0.0802	0.0722
AIC	-3.0648	-3.1460	-3.0258
Qr(5)	0.1219	0.8673	0.6413
Qr(10)	6.24	10.58	8.31
Qr(20)	16.81	26.61	17.61
Qr(30)	30.69	42.29	30.60
Qr(40)	39.08	63.20**	50.30
Qr(50)	51.77	73.05**	59.12
Qrr(5)	103.9***	44.61***	61.45***
Qrr(10)	120.20***	87.17***	102.0***
Qrr(20)	165.80***	105.40***	128.80***
Qrr(30)	178.0***	117.10***	153.50***
Qrr(40)	191.0**	121.10**	161.80**
Qrr(50)	208.3**	125.0**	165.80**

注：ARMA(p,q)模型中p和q分别表示自回归和滑动平均阶；Qr(k)表示ARMA(p,q)模型残差的滞后k阶的Ljung-Box统计量，Qrr(k)表示ARMA(p,q)模型平方残差的滞后k阶的McLeod-Li统计量；*，**，***分别表示在10%，5%和1%水平下统计显著。

关于ARMA模型“过滤”后股市收益非线性相关性的BDS检验结果如表3所示。不论是对上证综合还是对深证成分和综合收益，BDS统计量都表现出一定的显著性，拒绝ARMA残差服从独立同分布的假设。显著的BDS统计量说明股票市场收益具有非线性相关结构，进一步强化了McLeod-Li检验的结论。

由于目前还没有好的方法甄别股市收益非线性相关性的确定性或随机性本质，并且本文侧重寻找描述股票收益非线性相关结构的随机模型，因此，我们重点诊断随机非线性相关的积式相关或和式相关性。表4给出了滞后1至4阶的Hsieh检验结果。因为Hsieh统计量H收敛于标准正态分布N(0,1)，因此，Hsieh检验的显著性可以根据标准正态分布确认。在表4中，Hsieh统计量在绝对数量上都很小，在5%和1%显著水平下都不能拒绝三阶矩等于零的假设，这说明股市收益序列不具有显著的额外的和式非线性相关性。Hsieh检验表明股市收益的随机非线性相关性主要表现为积式相关。

表 3 ARMA 残差非线性相关性的 BDS 检验

ε/SD	m	上证综合	深证成分	深证综合
0.5	2	0.8481	1.5174	1.1246
0.5	3	1.5614	2.1254*	1.6597
0.5	4	2.3266*	2.7591*	2.1457*
0.5	5	3.2204*	3.4743*	2.4367*
1.0	2	0.5830	1.0161	0.7934
1.0	3	1.0014	1.3681	1.1098
1.0	4	1.4346	1.6077	1.3527
1.0	5	1.9233	1.9216	1.5924
1.5	2	0.3520	0.5507	0.4212
1.5	3	0.5544	0.7137	0.5603
1.5	4	0.7401	0.8045	0.6513
1.5	5	0.9439	0.9389	0.7582

注：* 表示在 5% 水平下统计显著。

表 4 ARMA 残差的 Hsieh 检验

滞后阶 (j, k)	上证综合	深证成分	深证综合
(1, 1)	0.1035	0.1373	0.1269
(1, 2)	- 0.3051	0.3420	0.5974
(1, 3)	- 0.2663	0.3455	0.3143
(1, 4)	0.4058	- 0.2341	- 0.4070
(2, 2)	- 0.1080	0.1779	0.1270
(2, 3)	- 0.4352	- 0.5880	- 1.0131
(2, 4)	- 0.3946	- 0.7889	- 0.3776
(3, 3)	- 0.1198	0.0626	0.1585
(3, 4)	0.2812	0.7788	1.3517
(4, 4)	0.0820	- 0.1041	- 0.1162

注：*，**，*** 分别表示在 10%，5% 和 1% 水平下统计显著。

六、随机非线性相关结构的模型考察

股市收益积式相关代表性模型 ARMA-GARCH (1, 1) 的拟合结果见表 5。比较表 2 容易看到，ARMA-GARCH 模型的 AIC 统计量更小，标准残差的 Ljung-Box 统计量 $Q_s(k)$ 也

更小,并且在统计上不显著。这都反映出 ARMA-GARCH(1,1)比单纯的 ARMA 模型更适合刻画我国股票市场收益的演变。

表 5 中还反映了对 ARMA-GARCH 标准残差的 McLeod-Li 和 BDS 检验结果,以及和式相关代表性模型的拟合结果。对 ARMA-GARCH 标准残差的 McLeod-Li 统计量 $Q_{SS}(k)$ 在统计上不显著,表明股市收益的随机非线性相关性已经被 ARMA-GARCH 模型成功消除,也即股市收益存在显著的积式非线性相关性、和式相关性不显著。关于 ARMA-GARCH 标准残差的 BDS 统计量也统计不显著,再次证实了上述 McLeod-Li 检验的结果。对上证综合、深证成分和综合收益情形,表 5 中关于和式相关代表性模型的 ARMA-GARCH(1,1)-拟合中,参数 β_1 在统计上都不显著,又进一步证实了上述关于 ARMA-GARCH 标准残差的 McLeod-Li 和 BDS 检验结果。

因此,模型考察表明股市周收益序列的随机非线性相关性主要表现为积式相关、和式相关性不显著,ARMA-GARCH(1,1)模型比 ARMA-GARCH(1,1)-M 更适合描述我国股市收益的演变。这进一步证实了文章第五部分关于股市收益非线性相关的存在性和基本结构的统计检验结果,并揭示出中国股票市场不存在显著的风险-收益权衡、投机成分明显。

七、实证结论

在利用 ARMA 模型消除股市收益序列线性相关成分的基础上,我们针对“过滤”后的股市收益序列分别进行 McLeod-Li, BDS 和 Hsieh 检验,以诊断股市收益非线性相关结构的存在性和基本类型。然后,我们对积式相关代表性模型的拟合残差进行 McLeod-Li 和 BDS 诊断,并通过拟合和式相关的代表性模型对统计检验结果进行验证。研究发现的主要结论是:

(1) 中国股票市场收益存在显著的非线性相关性。

(2) 我国股票市场收益的随机非线性相关性显著表现为积式相关,不存在显著的和式相关结构。

(3) ARMA-GARCH 模型比 ARMA-GARCH-M 模型更适合描述我国股票市场收益的演变。我国股票市场不存在显著的风险-收益权衡、投机成分明显。

(4) 在利用随机非线性随机模型进一步研究我国股票市场的数量特征时,在 ARCH 和 ARCH-M 族模型之间,应该主要考虑以 ARCH 族模型为基础。

本文的研究尚存一些局限。比如,在本文研究的数据样本这十年间,中国股票市场经历了若干次比较大的变化,股票市场收益的非线性相关结构可能存在动态的变化。又如,本文采用的几个检验都有较严格的适用前提,实际数据在满足检验前提方面可能存在缺陷。这些问题有待于进一步深入研究。

表5 ARMA-GARCH(1,1)拟合、McLeod-Li、BDS检验和ARMA-GARCH(1,1)-M验证
股市收益的ARMA-GARCH(1,1)拟合

股市收益的ARMA-GARCH(1,1)拟合				股市收益的ARMA-GARCH(1,1)-M拟合				
	上证综指	深证成指	深证综指		上证综指	深证成指	深证综指	
α_0	9.44E-05***	0.000161***	0.000147***		8.93E-05**	0.000155***	0.000122***	
α_1	0.166290***	0.236650***	0.247474***		0.178981***	0.221039***	0.223781***	
γ_1	0.799597***	0.725511***	0.727841***		0.798198***	0.739998***	0.756960***	
β_1					0.005793	0.127385	0.144105	
AIC	-3.3155	-3.3217	-3.2883		-3.286210	-3.3136	-3.2461	
ARMA-GARCH(1,1)标准残差的Ljung-Box统计量				ARMA-GARCH(1,1)标准残差的BDS检验				
	上证综合	深证成分	深证综合	ϵ/SD	m	上证综合	深证成分	深证综合
Qs(5)	0.9417	7.171	8.780	0.5	2	-0.002082	0.1099	0.1972
Qs(10)	6.028	14.72	14.07	0.5	3	0.02463	0.3568	0.1832
Qs(20)	13.211	22.40	21.93	0.5	4	-0.002945	0.2957	-0.03039
Qs(30)	24.88	34.23	34.03	0.5	5	0.004943	0.1237	-0.2636
Qs(40)	33.33	50.31	46.54	1.0	2	0.04723	0.09626	0.1440
ARMA-GARCH(1,1)标准残差的McLeod-Li检验				1.0	3	0.1209	0.2098	0.1726
				1.0	4	0.07626	0.1560	0.1132
Qss(5)	2.651	3.617	3.460	1.0	5	0.04525	0.08235	0.03343
Qss(10)	3.821	5.782	5.584	1.5	2	0.01967	0.05445	0.08308
Qss(20)	9.796	9.145	9.576	1.5	3	0.05913	0.09425	0.09166
Qss(30)	13.34	10.04	15.15	1.5	4	0.03040	0.06073	0.04991
Qss(40)	19.66	21.44	23.06	1.5	5	0.01627	0.02841	0.008007

注：* * 和 * * * 分别表示在5%和1%水平下统计显著。

参考文献

- [1] 陈浪南, 黄杰鲲:《中国股票市场波动非对称性的实证研究》,《金融研究》,2002年第5期。
- [2] 戴国强, 吴祥林:《金融市场微观结构理论》,上海财经大学出版社1999年版。
- [3] 何兴强:《上证指数收益和波动性的星期效应检验》,《中山大学学报》(哲社版),2003年第6期。
- [4] 何兴强, 孙群燕:《中国股票市场的杠杆效应和风险收益权衡》,《南方经济》,2003年第9期。
- [5] 柯珂, 张世英:《分整增广GARCH-M模型》,《系统工程学报》,2003年第2期。
- [6] 刘国旗:《非线性GARCH模型在中国股市波动预测中的应用研究》,《统计研究》,2000年第1期。
- [7] 汤果, 何晓群, 顾岚:《FIGARCH对股市收益长记忆性的实证分析》,《统计研究》,1999年第7期。
- [8] 魏巍贤, 周晓明:《中国股票市场波动的非线性GARCH预测模型》,《预测》,1999年第5期。
- [9] 张思奇, 马刚, 冉华:《股票市场风险、收益与市场效率》,《世界经济》,2000年第5期。

- [10] Aggarwai, R. ; Inclan, C. and Leal, R. , Volatility in Emerging Stock Markets , *Journal of Finance and Quantitative Analysis* , 1999 , Vol.34 , No.1 , pp33 ~ 55.
- [11] Barnett, W. A. ; Gallant A. R. ; Hinrich, M. J. ; Jungeilges, J. A. ; Kaplan, D. T. and Jensen, M. J. , Robustness of Nonlinearity and Chaos Tests to Measurement Error, Inference Method, and Sample Size , *Journal of Economic Behavior and Organization* , 1995 , Vol.27 , pp301 ~ 320.
- [12] Bollerslev, T. ; Chou, R. Y. and Kroner, K. F. , ARCH Modelling in Finance : A Review of Theory and Empirical Evidence , *Journal of Econometrics* , 1992 , Vol.52 , pp5 ~ 59.
- [13] Bollerslev, T. , Generalized Autoregressive Conditionally Heteroscedasticity , *Journal of Econometrica* , 1986 , Vol.31 , pp307 ~ 327.
- [14] Box, G. E. P. ; Jenkins, G. M. and Reinsel, G. C. , *Time Series Analysis, Forecasting and Control* , 3th ed. , Holden-Day, San Francisco , 1994.
- [15] Brock, W. A. ; Dechert, W. D. and Scheinkman, J. , A Test for Independence Based on the Correlation Dimension , Unpublished Manuscript, Madison : University of Wisconsin , 1987.
- [16] Brock, W. A. ; Hsieh, D. A. and LeBaron, B. , *Nonlinear Dynamics, Chaos, and Instability : Statistical Theory and Economic Evidence* , Cambridge, Massachussets , 1992.
- [17] Brock, W. A. ; Scheinkman, J. A. ; Dechert, W. D. and LeBaron, B. , A Test for Independence on the Correlation Dimension *Economic Review* , 1996 , Vol.15 , No.3 , pp197 ~ 235.
- [18] Brock, W. A. ; Hsieh, D. A. and LeBaron, B. , *Nonlinear Dynamics, Chaos and Instability* , MIT Press , Cambridge , MA , 1991.
- [19] Campbell, J. Y. ; Lo, A. W. and Mackinlay, A. C. , *The Econometrics of Financial Markets* , Princeton University Press , Princeton New Jersey , 1997.
- [20] Cronwell, J. B. ; Labys, W. C. and Terraza, M. , *Univariate Tests for Time Series Models* , Sage , Thousand Oaks , CA , 1994.
- [21] Darrat, A. F. and Maosen Zhong , On Testing the Random-Walk Hypothesis : A Model-Comparison Approach , *The Financial Review* , 2000 , Vol.35 , pp105 ~ 124.
- [22] Engle, R. F. , Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation , *Econometrica* , 1982 , Vol.50 , pp987 ~ 1008.
- [23] Granger, C. W. J. and Andersen, A. P. , *An Introduction to Bilinear Time Series Models* , Vandenhoeck & Ruprecht , Gottingen , 1978.
- [24] Hsieh, D. A. , Testing for Nonlinear Dependence in Daily Foreign Exchange Rate , *Journal of Business* , 1989 , Vol. 62 , pp339 ~ 368.
- [25] Lo, A. W. and A. C. MacKinlay , Stock Market Prices Does Not Follow Random Walks : Evidence from A Simple Specification Test , *Review of Financial Studies* , 1988 , Vol.1 , pp41 ~ 66.
- [26] McLeod, A. J. and Li, W. K. , Diagnostics Checking ARMA Time Series Models Using Squared Residual Autocorrelations , *Journal of Time Series Analysis* , 1983 , Vol.4 , pp269 ~ 273.
- [27] Scheinkman, J. A. and LeBaron, B. , Nonlinear Dynamics and Stock Returns , *Journal of Business* , 1989 , Vol.62 , pp311 ~ 337.

第六部分

数量经济学

泰勒规则在中国的协整检验^{*}

陆 军 钟 丹

一、引言

《中国人民银行法》已经明确我国货币政策的最终目标是“保持币值的稳定，并以此促进经济增长”。而应采用什么政策工具和中介目标来实现最终目标一直是备受关注和争议的问题。1993年，人民银行首次向社会公布货币供应量指标，这表明在经历了20世纪90年代初金融失控局面的冲击后，人民银行意识到利用信贷规模和现金投放作为政策工具的货币政策调控机制的缺陷，开始逐步转向以货币供应量为货币政策的中间目标。1996年起，人民银行采用货币供应量M1和M2作为货币政策的调控目标，1998年，央行取消了信贷规模控制，进一步确立了货币供应量作为中间目标的地位。然而，几年来调控货币供应量的实践表明，作为中介目标的货币供应量在指标可控性、可测性和与最终目标关联度上，都不令人满意。不仅货币供应量的目标值无法实现，其测量口径引起许多争议，而且货币供应量的变动与物价、经济增长率的变动趋势也出现了不一致（夏斌、廖强，2001）。甚至有学者提出货币供应量已不宜作为当前我国货币政策的中间目标，应借鉴泰勒规则，采用利率指标。在2002年举办的“货币政策目标与操作”国际研讨会上，有专家指出，从中国经济金融的发展以及理论逻辑出发，利率比货币供应量可能更适宜作为货币政策的中介目标（王煜、张红地、余明，2002）。因此，我们需要做好从货币供应量目标向利率目标转移的理论研究的准备。

就利率作为货币政策的中介目标而言，目前国内已经开始进行探索性的研究，谢平、罗雄（2002）对泰勒规则在中国的实证研究就是其中最具有代表性的一篇。他们首次运用历史分析法和反应函数法检验泰勒规则在中国是否成立，指出泰勒规则可以很好地衡量中国货币政策，能够为中国货币政策提供一个参照尺度。在当前我国利率市场化改革逐步深入推进的背景下，他们的研究提供了今后中国货币政策操作的一个新方向——借鉴泰勒规则，采用利率目标，在理论和实践上都有重要的启示。然而他们对泰勒规则的研究还存在一些不够完善之处。例如，他们用历史分析法和反应函数法得到的结论实际上存在矛盾之处。一方面，他们直接采用泰勒教授根据美国经济数据提出的传统泰勒规则形式，运用历史分析法计算我国货币政策中利率的泰勒规则值并作为随后检验和分析的基础；而另一方面，他们估计得到的反应函数却说明了我国的货币政策是不稳定的，并非按照传统的泰勒规则操作。造成这一矛盾之处的主要原因在于他们运用历史分析法时直接采用传统的泰勒

^{*} 原载：《经济研究》，2003年第8期。

规则形式。事实上，如果用计量方法，根据我国的具体数据估计得到泰勒规则的具体形式，可能可以更好地说明我国的情况。此外，他们对泰勒规则实际操作中存在的确定性也欠缺深入的讨论，尤其是忽略了货币政策时滞对泰勒规则操作的影响。

鉴于此，本文尝试在谢平、罗雄（2002）一文的研究基础上，进一步探讨在我国应用泰勒规则的可行性。具体而言，本文从泰勒规则出发，利用协整理论检验我国银行间同业拆借利率与经济增长和通货膨胀水平之间是否存在稳定的长期关系，并估计我国泰勒规则的具体形式。考虑到货币政策的时滞问题，本文引入前瞻性泰勒规则，用预期通货膨胀率缺口代替原有的通货膨胀率缺口作为估计变量，对泰勒规则作进一步修正。

文章的第二部分将简要回顾与探讨传统的泰勒规则，介绍泰勒规则的含义和国外相关文献的实证结果；第三部分提出实务上运用泰勒规则所遭遇的问题，并就中国现有的资料寻求可能的解答；第四部分以 Johansen 协整分析方法来检验国内银行同业拆借利率的长期走势是否符合泰勒规则，同时估计我国泰勒规则的具体形式；第五部分引入前瞻性泰勒规则，对传统的泰勒规则予以修正；第六部分是结论与改进方向。

二、传统泰勒规则的回顾与探讨

1993年，在“Macroeconomic Policy in a World Economy: from Econometric Design to Practical Operation”这篇文章里，泰勒（John Taylor）教授使用多国的理性预期模型，通过7个工业国家数据的模拟分析，发现利率规则最有利于央行保持产出与物价稳定，其中利率规则是指根据产出和物价水平与设定目标值之间的差距来调节利率的货币政策规则。随后在“Discretion versus Policy Rules in Practice”（Taylor, 1993）一文中，Taylor教授提出了一个简单的代表性利率规则，并用实证分析证明这一规则很切合1987~1992年美国联邦资金利率的走势，这一规则后被称为“泰勒规则”，具体形式如下：

$$fir = rstar + p + 0.5ygap + 0.5(p-ptarget)$$

其中：fir为联邦基金利率；rstar为实际均衡利率（泰勒教授根据美国资料设定为2%）；p为平均通货膨胀率（本季度与上三个季度通胀率的平均值）；ygap为实际总产出缺口，以 $100(Y - Y^*)/Y^*$ 表示，其中Y为实际国内生产总值（GDP），而 Y^* 为潜在产出（按照2.2%的同比增长率增长的潜在GDP）；ptarget为设定的通货膨胀率目标值（泰勒教授设定为2%）。

泰勒规则的基本含义在于，货币当局在执行货币政策时遵循以下反馈规则，即根据通货膨胀率和总产出的实际值与目标值之间的差距来调节联邦基金利率的走势。虽然泰勒规则的形式很简单，但它既可以很好地反映货币当局的两个主要目标，即在短期内维持实际产出在其潜在产出水平附近，同时在长期内有效地控制通胀率，而且还可以概括现实中影响货币政策的基本因素（Judd and Rudebusch, 1998）。此外，还有学者认为，产出缺口调整因素考虑了未来通胀的可能性，所以提出的政策建议具有一定前瞻性。

后续研究泰勒规则的文献，对该法则的运用提出了各种问题、扩展以及修正，并形成了一系列将联邦基金利率的目标值设定为以下两个变量：通货膨胀率以及实际产出各自与目标值偏离的函数的利率规则，即所谓的“泰勒类型规则”（Taylor-type rules），这类规则

与原始的泰勒规则有以下不同：

首先是通胀缺口与产出缺口的权重设定不同，即对两个调整系数的设定不同。泰勒教授根据当时的一些研究成果和基于简化的考虑，将这两个调整系数均先验设定为 0.5，见 (1) 式。但许多学者认为，若用计量方法估计这两个系数，可以更切合货币政策的历史情况。而且，由于这两个系数反映了货币政策对通胀和产出偏离目标值的反应程度，它们必须根据各国实际情况和货币当局的偏好加以修正。此外，根据 Andrew Levin, Volker Wieland and John C. Williams (1998) 的研究显示，调整系数越大，模型越稳定。

其次是使用数据的时间选择上有所不同。原始泰勒规则中使用的是当期数据，即当期的联邦基金利率必须按照当期的产出缺口以及通胀缺口来调整。这其实是假设央行在设定本期的联邦基金利率时，就已经掌握了本期通胀和实际产出的数据。然而这假设是不现实的，因为一般本季度产出和通胀的数据最快都只能在下季度才统计出来。考虑到实际操作上的可行性，有学者提出应使用滞后一期的产出和通胀数据，即本期的联邦基金利率根据上一期的产出和通胀缺口来设定。

再次是“泰勒类型规则”可能会考虑联储平滑利率的行为 (Kozicki, 1999)。许多学者都注意到联储有平滑利率走势的偏好，这意味着货币政策对通胀与产出缺口的反应与调整不是在一个季度内一次完成，而是在几个季度内逐渐完成。考虑联储平滑利率行为的“泰勒类型规则”，一般假设本期的联邦基金利率除了取决于产出与通胀缺口因素外，还取决于上期的联邦基金利率。

最后是对通胀和产出缺口的测量方法不同。衡量通胀可以采取很多种指标，包括消费物价指数、扣除能源和必需品的核心消费物价指数和 GDP 平减指数等，而对潜在产出的估计也有许多种方法，由此就有许多种衡量产出缺口的的方法，对此，本文的第三部分会作进一步讨论。Kozicki (1998) 对通胀率和产出缺口采用不同的测量方法，分别估计泰勒规则，发现不同指标得出的联邦基金利率对产出与通胀缺口的反应程度有显著差异。

三、估计泰勒规则时存在的不确定性及其可能解答

在实际估计时，泰勒规则中各组成部分的设定与衡量都存在许多不确定性。具体而言，有以下几方面：一是实际总产出缺口的估算；二是通货膨胀率的均衡值或目标值的设定；三是实际均衡利率值的确定。在这一部分，我们会结合我国的实际数据，逐一讨论这些不确定性。

(一) 潜在产出和实际总产出缺口的估算

潜在产出是指当经济体系的生产资源均充分利用的情况下所能达到的最大产出。在文献中，估算潜在产出的方法主要有两大类：一是对实际产出的时间序列进行分解的方法，即将实际产出分成潜在产出和围绕潜在产出波动的产出缺口两部分，包括 Wharton 学派的潜在产出线估算方法、Hodrick-Prescott 滤波方法（简称 HP 滤波）、Beveridge-Nelson 方法和状态空间-卡尔曼滤波方法 (state space-Kalman filter) 等等。二是利用生产函数估计潜在产出。生产函数法假定产出遵循固定的生产函数模式，随生产要素投入量的变化而增减。

而按生产函数法估算潜在产出的最大优点是其背后有经济理论的支持，因而在文献中被大量采用。本文将采取这种方法来估计潜在产出。

将生产函数设定为：

$$Y = f(L, K, A)$$

其中：Y, K, L 和 A 分别表示产出、固定资本存量、劳动者人数和技术进步因素。在实际估算生产函数时会遇到以下三个问题：一是资本存量的实际资料欠缺；二是充分就业劳动量的估算；三是技术进步因素 A 的估算。

1. 关于资本存量的估算

由于资本存量的实际资料欠缺，我们用全国的固定资产净值来代替资本存量，再根据式子 $K_{t+1} - K_t = I_t - D_t$ 估算，其中 K_{t+1}, K_t 分别代表 $t+1$ 年和 t 年的固定资产净值； I_t 代表 t 年的固定资产投资； D_t 代表 t 年的固定资产折旧。

2. 关于充分就业劳动量的估算问题

一般认为，失业率为自然失业率的就业量就是充分就业劳动量。而自然失业率受人口、经济制度和社会阶层变化等非经济因素的影响，估计上十分困难。本文以统计年鉴中提供的“经济活动人口”作为充分就业的劳动量，即将自然失业率视为零。当然，这只是一个简化的处理，因为自然失业率为零的可能性很低。

3. 关于技术进步因素 A 的处理

产出的增加有赖于要素投入量的增加，但同时也依赖于许多其他非要素因素，例如教育普及、研究开发及金融体制的变革等等，这些因素一般被概括为技术进步因素。对这一因素的处理，本文参照 OECD (1994) 的处理方法，用两阶段估计法。

假设我国的生产函数为柯布 - 道格拉斯生产函数，由于使用的季度数据存在明显的季节性变化，所以在回归方程中加入 D_1, D_2, D_3 三个季节虚拟变量。

$$D_1 = \begin{cases} 1 & \text{一季度} \\ 0 & \text{其他} \end{cases} \quad D_2 = \begin{cases} 1 & \text{二季度} \\ 0 & \text{其他} \end{cases} \quad D_3 = \begin{cases} 1 & \text{三季度} \\ 0 & \text{其他} \end{cases}$$

估计以下式子：

$$\log(Y/L) = C_0 + C_1 \times \log(K/L) + C_3 \times D_1 + C_4 \times D_2 + C_5 \times D_3 \quad (3)$$

然后将上式的残差项以 Hodrick- Prescott 滤波 (HP-filter) 的方法加以平滑化所产生的新数列作为 A_t 项，即：

$$A_t = \text{HP-filter}\{\log(Y/L) - [C_0 + C_1 \log(K/L) + C_3 \times D_1 + C_4 \times D_2 + C_5 \times D_3]\} \quad (4)$$

进而将 A_t 项代入，估计下式：

$$\log(Y/L) = C_0^* + C_1^* \log(K/L) + C_2^* A_t + C_3^* \times D_1 + C_4^* \times D_2 + C_5^* \times D_3 \quad (5)$$

最后，用充分就业的劳动量 (L^*) 代替劳动就业量，就可以得到潜在产出 (Y^*)。即：

$$Y^* = \exp\{\log(L^*) + C_0^* + C_1^* [\log(K) - \log(L^*)] + C_2^* A_t + C_3^* \times D_1 + C_4^* \times D_2 + C_5^* \times D_3\} \quad (6)$$

图 1 列出由生产函数方法估计的潜在 GDP 估计值和实际产出值，而对应的实际产出缺口走势图则见图 2。关于估计我国生产函数时使用的数据、具体过程及结果的说明请看附录 1。

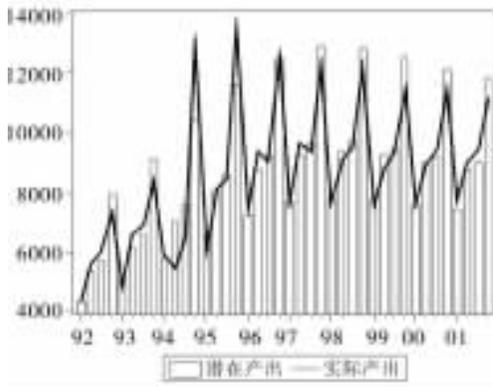


图1 潜在 GDP 估计值和实际产出值

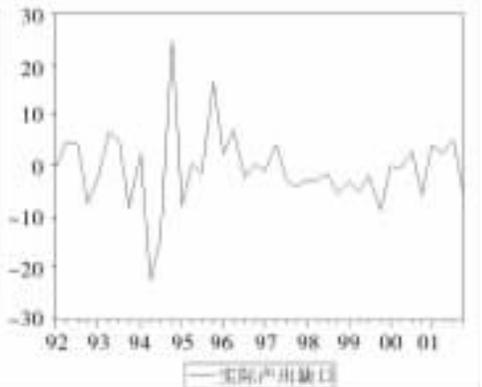


图2 实际产出缺口估计值

(二) 关于通货膨胀率均衡值或目标值的设定

泰勒规则以“维持通货膨胀稳定”为政策目标，因此，运用泰勒规则时，自然会遇到两个问题：一是以何种物价指数代表通货膨胀指标；二是所欲稳定的目标通货膨胀水平的确定。第一个问题比较简单，一般常用消费者物价指数（CPI）衡量通货膨胀水平。关于第二个问题，文献中确定通胀率均衡值的方法有以下几种：第一种方法是简单使用一个较长时期的平均通货膨胀率作为通胀率的均衡值；第二种方法操作上也很简单，即直接采用中央银行公布的通胀目标的平均值。

我们这里使用“潜在物价指数”法，引入“潜在物价指数（ P^* ）”作为通货膨胀的目标值。潜在物价指数基于 Fisher 的交易方程式，具体指在给定货币供给量和均衡货币流通速度（ V^* ）下，能维持潜在产出（ Y^* ）正常交易活动的均衡物价水平，即：

$$P^* = M V^* / Y^*$$

其中： M 以广义货币 M_2 表示； V^* 为均衡货币流通速度下； Y^* 为潜在产出。理论上， V^* 应是产出、物价、利率乃至制度改变等因素的函数，很难准确估计。简单起见，我们用每年实际值的年平均值（ v ）视作当年均衡值（ V^* ）。即每年的 $v = \text{当年的名义 GDP} / \text{当年末的广义货币}$ 。

将 M_2 、 V^* 及前面由生产函数法估算得到的潜在产出（ Y^* ）代入（7）式，则可求得 P^* 的估计值。而 P^* 的同比增长率就代表了目标通货膨胀率（PSTAR）水平。

(三) 关于实际均衡利率的设定

关于实际均衡利率的设定也有许多不同的方法。方法一是利用潜在 GDP 的同比增长率估算。Taylor 教授（1993）就是采用这一方法，他设定实际均衡利率等于潜在 GDP 的同比增长率，似乎颇符合美国的情况。方法二是利用回归方程，估计泰勒规则的形式，那么回归方程的截距项就是实际均衡利率。然而这一方法缺乏理论依据，因为估计得到的泰勒规则形式只能反映以往的货币政策制定者对通胀和产出可能作出的反应（Gramlich, 1998）。方法三是 Gramlich（1998）和 Stuart（1996）等学者提出的，用指数化的公债（index-linked

bond) 利率作为实际均衡利率的替代变量。第四种方法认为实际均衡利率等于资本的边际生产力 (marginal productivity of capital), 通过估算资本的边际生产力确定实际均衡利率。

还有一种求实际利率的方法, 它依据费雪方程式, 资产的实际报酬率 (实际利率) 等于扣除预期通胀率的名义报酬率。如 Barro (1992) 提出以三个月的国库券利率扣除预期通货膨胀率 (以实际 CPI 指数的年增长率代替) 作为短期的实际均衡利率。本文也采用这一办法, 具体是用银行一年期定期存款利率扣除实际 CPI 年增长率作为实际均衡利率。

四、泰勒规则的实证分析

在这一部分, 我们以计量经济学的协整理论 (Cointegration Theory) 为基础检验泰勒规则, 即通过实证检验我国的银行同业拆借利率与实际均衡利率、通胀率、通胀缺口及产出缺口这五个变量间是否存在“协整”关系或长期均衡关系, 说明泰勒规则在中国是否成立。

(一) 协整理论的简单介绍

协整理论是研究分析非平稳时间序列的一个重要方法。Engle and Granger (1987) 指出, 如果两个或两个以上的非平稳时间序列 (含有单位根的时间序列) 的线性组合能构成平稳的时间序列, 则称这些非平稳时间序列是协整的, 称得到的平稳的线性组合为协整方程, 可以认为协整方程的存在说明这些变量 (即非平稳的时间序列) 之间存在长期的均衡关系。所以, 如果通过检验, 证明泰勒规则内的各个变量之间的确存在协整关系, 就能说明泰勒规则是成立的, 且可以作为货币政策的操作规则。本文将采用 Johansen 提出的协整检验方法来检验变量之间的协整关系。

(二) 计量模型的设立和数据说明

我们将拟检验的模型设定为简单的泰勒规则模式:

$$\text{INTER} = a_1 \text{RSTAR} + a_2 \text{CPI} + a_3 \text{YGAP} + a_4 \text{PGAP} \quad (8)$$

其中: INTER 为银行同业拆借利率^①; RSTAR 是实际均衡利率, 以一年期定期存款利率扣除 CPI 的年增长率代替; CPI 为消费者物价指数的年增长率^②; YGAP 代表潜在产出缺口, 即 $100(Y - Y^*)/Y^*$, 其中, Y 代表实际 GDP, Y^* 用第二部分的生产函数法估算得到; PGAP 代表通货膨胀缺口, 即 $\text{CPI} - \text{PSTAR}$, 其中 PSTAR 表示目标通货膨胀率, 用上一部分第 (二) 点介绍的方法, 估算“潜在物价指数” P^* 的年增长率得到。

在我们要建立的向量自回归模型中, 内生变量向量组 $y_t = (\text{INTER}, \text{RSTAR}, \text{CPI}, \text{YGAP}, \text{PGAP})$, 外生变量 x_t 只包括常数项。

① 1992 ~ 1995 年选取上海融资中心同业拆解利率, 引自谢平、罗雄:《泰勒规则及其在中国货币政策中的检验》,《经济研究》, 2002 年第 3 期。1996 ~ 2001 年选取人民银行各期季报公布的 7 天的同业拆解利率。

② 将中国人民银行统计季报提供的消费物价指数 (上年同期 = 100) 的月度数据作简单平均得到季度的消费物价指数数值。

从结果上看, 5 个变量间确实存在长期均衡关系, 传统泰勒规则适合描述我国银行同业拆借利率的长期走势。每个估计符号都符合理论要求, 而且产出缺口的反应系数的估计值为 0.497, 与传统泰勒规则非常接近, 然而通胀缺口的调整系数却只有 0.089, 与传统泰勒规则差别较大。我们知道, 泰勒规则中通货膨胀缺口调整因素和产出缺口调整因素的权重可以体现短期内央行面对经济成长和物价稳定两难局面时的权衡结果, 反映中央银行执行货币政策的偏好, 即经济成长与物价稳定孰重的决策尺度 (Kozicki, 1999)。所以, 这在一定程度上可以说明人民银行相对更重视保持产出的稳定。当然, 由于我们使用从 1992 年到 2001 年的季度数据, 其中包含的 1993~1994 年的恶性通货膨胀数据对估计结果也会有一定影响。

五、前瞻性泰勒规则

(一) 前瞻性泰勒规则的含义

前面传统的泰勒规则对通货膨胀率缺口的衡量, 是直接以当期实际通货膨胀率扣除物价上涨的最终目标 (或长期均衡值) 表示, 计算上相当方便。只是中央银行在制定货币政策时, 同期或过去的实际通货膨胀率并非理论实务上关注的重心, 事前的预期的通货膨胀率才是斟酌政策取向的关键。好的货币政策在认为未来物价有通货膨胀压力时就应该及时调整。况且, 由于货币政策从制定、实施到影响实体经济需要一段较长的时间, 即货币政策存在所谓的时滞问题, 因此, 若将泰勒规则作进一步修正, 在传统模型中引入预期因素, 改用预期通货膨胀率缺口作为估计变量, 使之成为前瞻性的政策决策方式, 应是比较合乎实际的做法, 还有助于央行更好地调控同业拆借利率。

以预期为基础的前瞻性泰勒规则的具体形式如下:

$$i_t = r^* + \pi_t + \mu_1^* (y_t - y^*) + \mu_2^* (E_t \pi_{t+1} - \pi^*) \quad \mu_1, \mu_2 > 0$$

与原始的泰勒规则模型 (1) 式比较, 原来的通货膨胀率 π 被预期的通货膨胀率 $E_t(\pi_{t+1} | I_t)$ 代替, 其中 $E_t(\pi_{t+1} | I_t)$ 表示以 t 期的信息集 (Information set) I_t 来预测 $t+1$ 期的通货膨胀率, i_t 即 r_{t+1} 。所以在前瞻性泰勒规则中, 目标利率不是取决于已经实现的或“事后”的通货膨胀率, 而是取决于根据已有信息而作出的对未来通货膨胀率的预期。

(二) 预期通货膨胀率的估算

要估算预期的通货膨胀率, 最简化的方式是利用公众或专业机构对未来通货膨胀率的看法, 作为预期的通货膨胀率。但是, 目前我国并无此项问卷调查资料与预期指标。在文献中, 估算预期通货膨胀率的方法有以下两大类: 一是直接估计的方式, 主要依据 Lucas 供给曲线所导出的预期修正的菲利普斯曲线 (expectation-augmented Phillips curve)。如 Brayton & Mauskopf (1985), Rudebusch & Svensson (1998) 的实现通货膨胀率加权平均模型以及 Hallman, Porter & Small (1989, 1991) 的均衡物价模型 (Pstar Model)。二是间接估计的方式。根据费雪方程式 ($R_t = r^* + \pi_t^e$) 估计隐含的预期通货膨胀率。如 Higo (1999) 分

别以稳态的均衡经济增长率、潜在产出增长率及资本报酬率的估算值作为实际利率的替代变量，进而间接估计预期的通货膨胀率。

本文将采用直接估计的方式，引入菲利普斯 - 奥肯曲线，估算预期的通胀率，具体方法如下：

附加预期的菲利普斯曲线可以用下式表示：

$$\pi_t = f(u_t) + \pi_t^* = \pi_t^* + b(u_t - u^*)$$

奥肯法则指的是美经济学家奥肯所观察到的在失业率与商品市场的超额需求之间存在负向关系这一实证法则 (Oken, 1970, 1974)。这法则可表示为：

$$u = u^* - a^* \text{ygap}_t \quad (\text{其中 } \text{ygap}_t = \frac{Y_t - Y_t^*}{Y_t^*} \times 100)$$

其中： u 与 u^* 表示实际与自然失业率。 Y 是实际产出， Y^* 是潜在产出（即失业率等于自然失业率时的产出）。

将奥肯法则和菲利普斯曲线结合起来，我们可以得到菲利普斯 - 奥肯曲线：

$$\pi_t = \pi_t^* + ab^* \text{ygap}_t$$

由于未来的通货膨胀率与去年同期 (π_{t-1}) 和上一季度 (π_{t-1}^q) 的实际通货膨胀率关系最为密切，所以估计预期通货膨胀率时，上式中包括 π_{t-1} 和 π_{t-1}^q 两项。

实际的估计式子及结果如下：

$$\begin{aligned} \pi_t &= \beta_1 \text{ygap}_t + \beta_2 \pi_{t-1} + \beta_3 \pi_{t-1}^q \\ \pi_t &= 0.256 \text{ygap}_t - 0.256 \pi_{t-1} + 1.21 \pi_{t-1}^q + e \\ &(0.713) \quad (-4.67) \quad (21.9) \end{aligned}$$

所以预期通胀率 $\pi_t^e = -0.256 \pi_{t-1} + 1.21 \pi_{t-1}^q$

(三) 对前瞻性泰勒规则的协整检验

按照前面第四部分介绍的方法，只是将通胀缺口改为预期通胀缺口 ($\text{PGAP_EX} = \pi_t^e - \text{PSTAR}$)，进行 Johansen 协整检验，结果如表 4 所示。

表 4 协整检验前瞻性泰勒规则估计结果

零假设： 协整向量的数目	特征值	迹统计量	临界值	
			5% 显著水平	1% 显著水平
0**	0.643	67.50	59.46	66.52
至多 1 个	0.360	32.47	39.89	45.58
至多 2 个	0.306	17.30	24.31	29.75
至多 3 个	0.077	4.88	12.53	16.31
至多 4 个	0.062	2.16	3.84	6.51

注：* * 表示在 1% 显著性水平拒绝零假设。

检验结果表明，这 5 个变量间存在一个协整方程，见下式：

$$\text{INTER} = 1.697^* \text{RSTAR} + 1.373^* \text{CPI} + 0.131^* \text{PGAP} - \text{EX} + 0.509^* \text{YGAP}$$

(0.187) (0.074) (0.076) (0.221)

既然通过协整检验，协整方程个数为 1，而且各个系数的符号均符合理论要求，说明前瞻性的泰勒规则同样适用于描述我国银行同业拆借利率。对比前面检验泰勒规则得到的协整（ $\text{INTER} = 1.527^* \text{RSTAR} + 1.316^* \text{CPI} + 0.497^* \text{YGAP} + 0.089^* \text{PGAP}$ ），我们发现，方程前瞻性泰勒规则的物价及产出缺口的反应系数的估计值均略高，而产出缺口的系数估计值仍与传统泰勒规则十分接近，预期通胀缺口的系数估计值仍与传统泰勒规则相差较远。

（四）两种规则短期误差纠正模型预测误差的比较

前面已证明估计变量之间存在长期均衡关系，那么，对于银行同业拆借利率的短期动态调整，可以求得递延 1 期的误差修正模型。表 5 列出泰勒规则与前瞻性泰勒规则对应的误差纠正模型对实际同业拆借利率拟合精度的指标，包括平均绝对误差、均方误差和 Theil 不等系数。观察结果，发现两个模型的预测误差无明显差异，说明通胀率是否具有“向前看”（forward-looking）并不影响拟合优度，对预测误差改善效果有限。这结论与 Fuhrer（1997）相同^①。

表 5 误差纠正模型的预测误差的比较

预测误差	泰勒规则模型	前瞻性泰勒规则模型
平均绝对误差 (MAE)	0.329	0.325
均方误差 (RMSE)	0.484	0.486
Theil 不等系数	0.026	0.027

六、结论与改进方向

（一）结论

自从 1996 年 6 月起，我国金融机构可根据市场资金供求状况，自行确定同业拆借市场上的拆借利率。尽管目前中央银行的货币数量控制主要以货币供应量为操作目标，但是透过再贴现率、法定存款准备金率或公开市场操作等政策工具的调节，银行体系的净超额准备的变动仍将间接影响银行同业拆借利率的高低。换句话说，同业拆借市场上的资金价格在理论上仍然可以反映央行货币政策的取向，而本文的总体实证结果也支持这点结论。具体而言，Johansen 协整检验结果表明变量间存在长期均衡关系且协整方程个数为 1，而且各个估计系数的符号都符合理论，所以实证表明传统及前瞻性泰勒规则都可用于描述我国银行同业拆借利率的走势，而银行同业拆借利率可以作为央行货币政策的参考依据。

^① Fuhrer（1997）实证结论指出通胀率是否具有前瞻性（forward-looking）并不影响拟合优度，但预期符合前瞻性时，可以使模型合理解释冲击的影响，有助于模型的动态稳定。

(二) 改进方向

1. 通货膨胀率目标值的设定

在推导通货膨胀目标值时,为求简便,仅假设货币流通速度的当年均衡值等于每年实际值的年平均值(V)。但是,理论上 V^* 是产出、物价、利率甚至是制度改变等因素的函数。因此,未来在估计通货膨胀目标值时,应对 V^* 加以深入的探讨。

2. 对产出缺口的估计

本文采用生产函数法估计潜在产出,进而估算产出缺口,虽然该方法有理论支持,然而因种种原因,我们在估计过程中不可避免地涉及一些假设,例如估计资本存量涉及的假设以及对充分就业人口的假设等,对于这些假设是否可能导致较大的偏差,还需要进一步的验证。另外,或许可以尝试一些新的更好的估计产出缺口的方法,并进一步比较各种不同方法得到的产出缺口是否会对最终泰勒规则的估计结果产生显著性的影响。

3. 对泰勒规则的进一步修正和讨论

泰勒规则在假定货币当局对利率控制具有独立性的前提下运作,然而实际中这点很难做到。Athanasios Orphanides (1997)指出,在固定汇率制度下,中央银行操纵银行同业拆借利率的自由度会受汇率固定而降低,因此,在固定汇率制度下,泰勒规则有修正的必要。我国目前实际上实行的仍然是钉住美元的固定汇率制度,所以在进一步研究泰勒规则在中国的具体应用时,必须寻求方法考虑这一因素。另外,在我国目前利率市场化程度低、实际经济部门对利率变动反应迟钝的情况下,如何具体使用泰勒规则还存在许多值得进一步讨论的问题。

附录 1 我国生产函数的估计

我们使用的数据主要来自中国统计年鉴和中国人民银行 1996~2001 年的各期季报,样本数据包括 1992 年第 1 季度到 2001 年第 4 季度,共 40 个样本点。

Y : 实际国内生产总值(用 1992 年不变价格表示)。由于没有季度的 GDP 平减指数,我们统一用年度的 GDP 平减指数将季度 GDP 的名义值转化为实际值。另外,这里的 GDP 表示季度发生值。

L : 使用年末全国劳动从业人员的数量^①。

L^* : 年末经济活动人口数^②。

K : 固定资产净值。

具体估算固定资产净值时需注意以下几点:

(1) 期初(1978 年)我国的资本存量价值的假设。根据中国统计年鉴,我国 1978 年年底国有企业固定资产的净值为 3201.4 亿元,由于当时国有经济的比重占了大部分(假设为 80%),则可以合理估计当

① 年鉴上没有提供季度数据,这里假设每季的从业人口和经济活动人口都等于当年年末的从业人口和经济活动人口。

② 由于年鉴上没有提供 2000 年和 2001 年经济活动人口的数据,我们根据以下公式和已知的数据进行估算:经济活动人口 = 城镇经济活动人口 + 农村经济活动人口 = (城镇就业人口 + 城镇失业人员 + 国有企业下岗职工) + 农村从业人口 = 从业人口 + 城镇失业人员 + 国有企业下岗职工。

2000 年经济活动人口 = (21274 + 595 + 657) + 49876 = 72402 万人

2001 年经济活动人口 = 73025 + 894 + 515 = 74434 万人

时全国的固定资产净值 $3201.4 \times 5/4 = 4001.75$ 亿元。

(2) 折旧率的假设。年鉴上有国有企业 1978 ~ 1992 年的固定资产折旧率，大部分都在 5% 左右，所以假设固定资产的年折旧率为 5%，而季度折旧率为 1.25%。

(3) 由于统计年鉴从 1991 年才开始提供固定资产投资价格指数这一指标，所以，对于 1992 年以前的年度固定资产投资数值，我们使用全国工业品出厂价格指数将固定资产投资的数据转换成用 1992 不变价格表示的数值，而对于 1993 年开始的固定资产投资的季度数据，则用固定资产投资价格指数将其折算为按 1992 年不变价格表示的数值。

用最小二乘法估计对 (3) 式和 (5) 式，结果如下 (括号中的数值代表 t 统计量)：

$$\log(Y/L) = -1.92 + 0.334 \log(K/L) - 0.485 \times D_1 - 0.310 \times D_2 - 0.272 \times D_3$$

$$(-43.4) (5.65) \quad (-8.31) \quad (-5.32) \quad (-4.68)$$

$$\text{Adjusted R-squared} = 0.744 \quad \text{DW} = 0.86$$

$$\log(Y/L) = -1.921 + 0.344 \log(K/L) + 1.327At - 0.486 \times D_1 - 0.311 \times D_2 - 0.273 \times D_3$$

$$(-68.7) (9.1) \quad (7.3) \quad (-13.2) \quad (-8.45) \quad (-7.4)$$

$$\text{Adjusted R-squared} = 0.898 \quad \text{DW} = 2.17$$

用 (L^*) 代替上式的 (L) 就可以得到潜在产出 (Y^*) 的估计值。

参考文献

- [1] 谢平，罗雄：《泰勒规则及其在中国货币政策中的检验》，《经济研究》，2002 年第 3 期。
- [2] 夏斌，廖强：《货币供应量已不宜作为当前我国货币政策的中间目标》，《经济研究》，2001 年第 8 期。
- [3] 李维刚：《泰勒规则、联储货币政策及我国货币调控问题的思考》，《国际金融研究》，2001 年第 6 期。
- [4] 王宇：《货币政策理论与操作的最新发展》，《世界经济》，2000 年第 3 期
- [5] 王煜，张红地，余明：《“货币政策目标与操作”国际研讨会综述》，《中国金融》，2002 年第 8 期。
- [6] Bennett T. McCallum, Recent Developments in the Analysis of Monetary Policy Rules, Homer Jones Memorial Lecture, March 11, 1999.
- [7] Edward M. Gramlich, Monetary Rules, Member of Board of Governors of the Federal Reserve System at the Eastern Economic Association, New York, February 27, 1999.
- [8] Hallman, J. J., R. D. Porter, and D. H. Small, M2 per Unit of Potential GNP as an Anchor for the Price Level, Staff Study No. 157, Board of Governors of the Federal Reserve System, Washington, DC, 1989.
- [9] Higo, M., What Can Inflation Expectations and Core Inflation Tell Us about Monetary Policy in Japan? IMES Discussion Paper Series 99 - E - 22, 1999.
- [10] John P. Judd and Glenn D. Rudebusch, Taylor's Rule and the Fed: 1970 - 1997, Federal Reserve Bank of San Francisco, Economic Review, 1998, Vol.3.
- [11] Kozicki Sharon, How Useful Are Taylor Rules for Monetary Policy? Economic Review, Second Quarter, 1999, Federal Reserve Bank of Kansas.
- [12] Michael Woodford, Pitfalls of Forward-Looking Monetary Policy, American Economic Review, 2000, Vol.90, No.2, pp100 ~ 104.
- [13] Nicoletta Batini and Andrew Haldane, Monetary Policy Rules and Inflation Forecasts, Bank of England Quarterly Bulletin: February 1999.
- [14] Taylor, John B., Discretion versus Policy Rules in Practice, Carnegie-Rochester Conference Series on Public

Policy , 1993 , pp195 ~ 214.

- [15] Taylor , John B. , *Macroeconomic Policy in a World Economy : From Econometric Design to Practical Operation* , New York : W. W. Norton , 1993.
- [16] Taylor , John B. , *A Historical Analysis of Monetary Policy Rules* , In Taylor , John B. , ed , *Monetary Policy Rules* , University of Chicago Press , 1999.

技术交易规则与超常收益研究^①

陈浪南 王艺明

一、引言

技术分析始于 19 世纪末的道氏理论，在流行超过一个世纪后，至今仍被广泛应用于金融资产收益预测，其间经历了形态分析、指标分析及自动交易规则等不同发展阶段。尽管有很多现代金融学者不甚赞同技术分析，但也有不少学者研究发现，技术分析是从市场价格中获得有用信息的有效方法。

实际上，学术界对技术交易规则的研究兴趣，主要源于技术交易规则与有效市场假说的对立。根据 Malkiel (1992)，如果股市是弱式有效的，那么技术分析将是徒劳的，因为在此情况下，历史交易的信息都已反应在股价中，再利用这些已被反应的信息显然不能获得超常收益。相应地，学术界对技术分析的研究也多集中在技术交易规则能否获得超常收益上。事实上，已有大量研究文献直接或间接地证明技术交易规则对金融资产收益具有实际预测能力。然而，在评价技术规则时经常遇到，却很少被直接指出的一个重要问题是数据窥察 (data-snooping)。所谓数据窥察，指当给定数据集被多次用于推断或模型选择时，所获得的任何令人满意的结果可能仅基于偶然，而非模型自身拥有的实际预测能力。

数据窥察未必是特定研究者努力的结果。它的产生可能源于历史上曾经被考虑过的整个规则集合上存在的微妙的生存偏倚 (survivorship bias)。随着时间的推移，投资者会对广泛的技术交易规则，包括不同规则的数千种不同参数设置进行试验。偶然取得较好历史收益的规则会受到较多重视，而不成功的规则则逐渐被忘记。这样，在经过很长的样本阶段后，只有很小的规则集合生存下来，这些技术规则的历史交易记录，则作为它们在投资决策中的价值的证明。如果考虑了足够多的规则，即使在很大的样本中，某些规则的成功也可能仅靠运气，而它们本身并不具有对资产收益的预测能力。因此，仅基于生存子集的推断可能会产生误导，因为它没有考虑整个初始规则集合，该集合中多数规则可能表现不佳。

本文旨在通过新实证技术——White (2000) 真实性检验试算法 (reality check bootstrap methodology) 的应用，拓展对技术交易规则的研究。尽管试算法 (bootstrap) 作为一种再抽样技术不是新的实证技术，但本文所采用的真实性检验试算法可以实现对数据窥察的精确调整，从而能确定技术规则所取得的超常收益，是基于其自身预测能力，还是仅靠碰运

^① 本文是国家自然科学基金课题 (79800010 和 78970039) 以及中加教育合作课题 (CCUIPP) 成果之一。原载：《经济研究》，2001 年第 12 期。

气。

二、文献回顾

学术界和实务界所共同关心的一个问题是，技术交易规则能否帮助投资者获得超常收益。目前多数研究认为技术规则能提供有价值的交易信号。其中一项较全面且有重要影响的研究是 Brock, Lakonishok & LeBaron (1992, 以下简称 BLL)。他们将 26 种技术规则应用于超过 90 年的道琼斯指数，发现它们显著优于持有现金的基准模型，并且所考察各类规则的各种参数设置都优于基准。BLL 的实证结论或者说明股市不是弱式有效的（与许多研究者既有的观念相悖），或者说明风险溢价（risk premium）即使在短期内（日间水平上）也存在显著波动。

BLL 在其研究中阐述了数据窥察问题，他们指出，“可以设计出多种移动平均模型，并且，毫无疑问，其中某些会产生较好的结果。然而，数据窥察的危险也是巨大的。”他们的做法是对原始数据拟合各种模型，并对残差再抽样，建立大量试算法样本，目的是检验结论的统计显著性和稳健性。然而，BLL 无法“对所有规则进行综合检验。该检验必须考察不同规则结果间的相关性”^①。BLL 发现，所考察的各规则都能够获得超常收益，但其研究的主要缺陷在于对数据窥察偏倚无能为力。他们指出，“技术分析发现各种伪形态的可能性无法被分散。尽管没有根治数据窥察偏倚的办法，我们可以减轻该问题（的危害）：^①通过报告所有交易策略的结果；^②通过应用非常长的数据序列，从 1897 年到 1986 年的道琼斯指数；^③强调结论对各种非重叠子阶段的稳健性。”

其他学者在评价技术规则业绩时也针对数据窥察采取了多种措施。Levich & Thomas (1993) 考察移动平均和过滤规则在外币期货市场的应用时，将试算法应用于期货的原始收益数据，发现某些技术规则有利可图。Olser & Chang (1995) 应用试算法检验头肩顶形态在外汇市场的获利能力时，也有类似发现。关于数据窥察，Levich & Thomas 指出，“显然还可以分析其他过滤尺度和移动平均长度，以及其他技术模型。（但）应避免该类数据提炼（data-mining）行为。”

总结前人研究可以得到以下结论：第一，存在技术规则能够获得超常收益的证据。第二，该实证证据的有效性受到数据窥察偏倚的影响。第三，处理数据窥察的方法是仅考察规则的小子集，以避免产生数据窥察偏倚。但正如前文的论述，有理由相信该策略事实上未必有效。历史上成功的技术规则也最容易引起研究者的注意，因为它们最为教科书或金融出版物所推崇。也即，尽管单个研究者可能行为谨慎，并且不对技术规则进行广泛试验，但整个金融界已经成为有效的“过滤器”，因此原则上必须考察投资者曾经考虑过的所有技术模型。

与前人对数据窥察所采取的避免或减轻的做法不同，White (2000) 发展的真实性检验允许数据提炼的存在，但通过考察所有技术交易规则收益间的相关性，能够实现

^① BLL 通过计算所有 26 种技术交易规则的平均收益，实际上考察了 26 种规则范围内的数据窥察。但这样应视为从 26 种规则中随机选择的期望收益，而非最优规则的收益。

窥察的精确调整,给出拒绝零假定(最优规则的预测能力不优于基准模型)的置信度,从而能确定技术规则所取得的超常收益,是基于其自身实际预测能力,还是仅靠碰运气,在该问题上取得重要突破。

White的方法还有其他广泛的应用,因为数据窥察的危险在金融研究和经济研究的其他领域也存在,如股票收益的可预测性(Foster, Smith & Whaley, 1997)、汇率和利率建模、横截面资产定价模型的因子和“异象”识别(Lo & MacKinlay, 1990),以及其他没有准确参数设定和函数形式的模型检验(指存在数据依赖,并且有意义的样本外试验难以进行)。

三、研究方法

(一) White 真实性检验 p 值

在Diebold & Mariano (1995)和West (1996)的研究基础上,White (2000)发展了检验给定模型(最优规则)在调整数据窥察效应后是否优于基准模型的方法,其思想是评价全体模型集合的业绩估计量分布。White真实性检验基于 1×1 的技术交易规则收益估计量:

$$f = n^{-1} \sum_{t=R}^T f_{t+1} \quad (1)$$

其中: l 是交易规则数; n 是预测期间,从 R 到 T ,即 $T = R + n - 1$, $f_{t+1} = f(\chi_t, \beta_l)$ 是 $t+1$ 期的收益估计量。在本文中,由不同参数设置的交易规则($\beta_k, k = 1, \dots, l$)直接产生收益并用于估计。

在本文样本中, n 为1600。 R 等于251,因为技术规则要求前250日的数据以产生交易信号。为评价各规则业绩,用下标 k 标识各规则。 $f_{k,t+1}$ 计算公式为:

$$f_{k,t+1} = \ln[1 + y_{t+1} s_k(\chi_t, \beta_k)] - \ln[1 + y_{t+1} s_0(\chi_t, \beta_0)] \quad (\text{其中 } \chi_t = \{X_{t-1}\}_{k=0}^R) \quad (2)$$

式中: X_t 是原始股票价格序列; $y_{t+1} = (X_{t+1} - X_t)/X_t$, $s_k(\cdot)$ 和 $s_0(\cdot)$ 是“信号”函数,根据系统参数 β_k 和 β_0 将价格信息序列转变为市场头寸^①。信号函数的取值范围是:1表示多头;0表示中性头寸(即退出市场);-1表示空头。下面我们将应用平均收益标准评价交易规则。本文检验是否存在超常收益技术规则的自然零假定是,最优交易规则的业绩不优于基准模型,也就是说:

$$H_0: \max_{k=1, \dots, l} \{E(f_k)\} \leq 0 \quad (3)$$

拒绝该零假定说明最优技术交易规则的业绩优于基准模型。White (2000)证明,该零假定的检验可以通过对 $f_{k,t+1}$ 的估计值应用Politis & Romano (1994)的平稳试算法(stationary bootstrap)实现。该方法可应用于严平稳和弱相关的时间序列,通过再抽样产生平稳的伪时间序列(pseudo-time series)。

我们应用再抽样版本的 $\bar{f} = n^{-1} \sum_{t=R}^T f_{t+1}^*$ 推导真实性检验p值以检验零假定(即最优

① 对于具有定义在期末财富上的对数效用函数的风险厌恶型消费者而言,具有最高累积收益的交易规则即为最优交易规则。

模型的预测能力不优于基准)。再抽样统计量的计算方法为：

$$\bar{f}^* = n^{-1} \sum_{t=R}^T f_{t+1}^*, f_{t+1}^* = f(Z_{\theta(t+1)}, \beta) \quad (t = R, \dots, T) \quad (4)$$

其中： $\theta(t)$ 是根据平稳试算法产生的随机数。首先选择一个平滑参数 $q = q_n$, $0 < q_n \leq 1$, $q_n \rightarrow 0$, 当 $n \rightarrow \infty$ 时, $nq_n \rightarrow \infty$, 算法过程为：

(1) 令 $t = R$, 从 $\{R, \dots, T\}$ 中随机、独立和均匀地选择一值赋予 $\theta(t)$ 。

(2) 令 $t = t + 1$ 。如果 $t > T$, 则算法停止。否则, 生成随机变量 U , 该变量在 $[0, 1]$ 上均匀分布, 且与其他任何变量相独立。①如果 $U < q$, 从 $\{R, \dots, T\}$ 中随机、独立和均匀地选择一值赋予 $\theta(t)$ 。②如果 $U \geq q$, 令 $\theta(t) = \theta(t-1) + 1$ 以扩展数据块；如果 $\theta(t) > T$, 令 $\theta(t) = R$ 。

(3) 重复步骤 (2)。

因此, 平稳试算法从原始数据中再抽样出不同长度的数据块, 数据块的长度服从几何分布, 平均长度为 $1/q$ 。大的 q 值适应于弱相关的原始数据, 而小的 q 值则适应于相关性较高的原始数据。考虑到股票日收益的弱相关性, 本文所使用的 q 值是 0.1, 对应数据块的平均长度为 10。

对技术交易规则收益的再抽样可产生 \bar{f}_k 的 B 个试算法值, 记为 $\bar{f}_{k,i}^*$, 其中 i 表示第 i 个试算法样本。令 $B = 500$, 并构造以下统计量：

$$\bar{V}_1 = \max_{k=1, \dots, l} \{\sqrt{n}(\bar{f}_k)\}, \bar{V}_{1,i}^* = \max_{k=1, \dots, l} \{\sqrt{n}(\bar{f}_{k,i}^* - \bar{f}_k)\} \quad (5)$$

我们将 \bar{V}_1 与 $\bar{V}_{1,i}^*$ 的 (上侧) 分位数相比较就可以得到零假定的 White 真实性检验 p 值, 即, 将 $\bar{V}_{1,i}^*$ ($i = 1, \dots, B$) 按升序排列得到 $\bar{V}_{1,(1)}^*, \bar{V}_{1,(2)}^*, \dots, \bar{V}_{1,(B)}^*$, 找到 M 使 $\bar{V}_{1,(M)}^* < \bar{V}_1 < \bar{V}_{1,(M+1)}^*$, 真实性检验 p 值的计算方法为 $p = 1 - M/B$ 。

通过遍历所有 l 个交易规则, 应用所有 l 个交易规则的最大值, White 真实性检验 p 值引入数据窥察效应。计算得到的 White 真实性检验 p 值是拒绝零假定 (即最优交易规则的业绩不优于基准模型) 的置信度。

(二) 技术交易规则集合

由于评价最优规则时数据窥察效应的大小是由交易模型回报间的相关性决定的, 因此设定适当的交易规则集合有重要意义。本文采用与 Ryan, Allan & Halbert (1999) 相同的技术交易规则集合。该集合中的技术规则在样本期间均被广泛采用, 包括过滤规则、移动平均、阻力支撑、通道突破 (channel breakouts) 和 OBV 平均 (on-balance volume average) 的总共 7846 种不同参数设置^①。对各类规则均选择广泛的参数设置以期覆盖投资者曾考虑过的各类模型。当然, 事实上无法完全做到这一点, 但本文采用规则集合已显著大于前人的研究。

① 这一点对本文结论的经济解释很重要, 只有当所考察的交易规则在样本期间广泛流行时, 存在超常收益交易规则才能得到市场无效率或事后风险溢价波动的结论。假定我们可以发现某个交易规则显著优于基准模型, 但是基于样本期过后方可获得的技术 (如神经网络), 由于投资者在样本期内无法获得上述技术, 因此这无法作为市场弱式无效的证据。

注意到 BLL 所考察的包括简单、固定移动平均和交易范围突破等三种规则的 26 种不同参数设置，应用由上述常用简单规则所组成的小子集的目的旨在减轻数据窥察所造成的影响。因此，在本文的实证研究中也计算了 BLL 规则集合的实证结果作为比较，以期考察上述策略是否能够真正奏效。类似 BLL，本文在平均收益标准下的基准规则是“零”系统，即总是退出市场， S_0 总是等于零。

值得考虑的一个重要问题是，本文所采用的规则集合的跨度空间是否显著大于 26 个 BLL 规则。为研究该问题，我们计算 BLL 规则集合的收益方差矩阵 (26×26)，再从全体规则集合中随机地选择 474 个规则，加入 26 个 BLL 规则，总共组成 500 个规则，计算其收益方差矩阵 (500×500)^①。收益矩阵所使用数据是上证指数 (1994 年 1 月 1 日至 2000 年 6 月 30 日)。对该两矩阵运用主成分分析，分别得到特征值集合。非零特征值越多，规则集合的有效跨度就越大。

图 1 报告了两收益方差矩阵的特征值分布。特征值依降序排列在横轴上，而纵轴显示

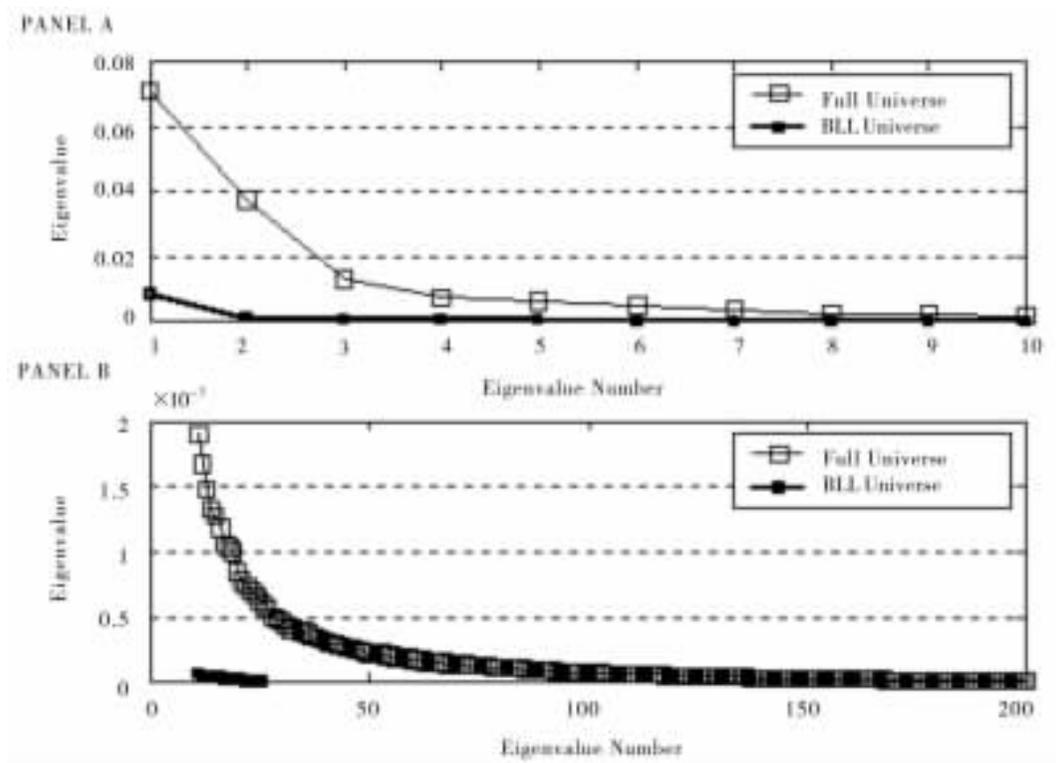


图 1 BLL 与全体规则集合的跨度：收益方差矩阵的特征值分布

注：该图报告了依降序排列的 BLL 收益方差矩阵特征值 (BLL Universe, 共 26 个) 和全体规则收益方差矩阵特征值 (Full Universe, 前 200 个)。Panel A 报告了前 10 个最大特征值，Panel B 报告了 11~200 个最大特征值 (BLL 规则有 26 个特征值)。特征值依降序排列在横轴上，纵轴显示特征值的大小

① 由于计算能力的限制，所以只选择全体交易规则的子样本。收益方差矩阵的计算应用 MATLAB 软件的 COV 和 EIG 函数。

特征值大小。如图所示，全体规则收益方差矩阵的非零特征值数明显大于 BLL 矩阵。在 BLL 矩阵的特征值集合中，大于 1×10^{-5} 的特征值有 20 个；而在全体规则的特征值集合中，大于 1×10^{-5} 的特征值有 225 个。对全体规则集合的不同随机样本进行多次试验，都得到相似结论。从而可以确信，全体规则集合的跨度空间显著大于 BLL 集合。强调所使用的规则集合具有足够大跨度空间的重要性在于，本文对数据窥察的调整仅对跨度空间内的数据窥察进行。

四、实证检验与分析

(一) 真实性检验 p 值与名义 p 值

本文所考察的样本股票主要是 1994 年 1 月 1 日以前在沪深上市的股票，以及上证指数和深证成指。首先从中选择出具有代表性的样本数据，为此，我们选取了上证指数、深证成指，此外在两市中各行业分别选取一只股票（共 11 只，参见表 1）。股指及各样本股票都以每日收盘价和成交量作为分析资料。本文所处理数据范围是 1994 年 1 月 1 日至 2000 年 6 月 30 日，共有约 1590 个数据。计算收益率时所有个股均经过除权处理。考虑到

表 1 平均收益标准下最优技术交易规则的业绩

样本股票	BLL 交易规则集合			全体交易规则集合		
	平均收益	White p 值	名义 p 值	平均收益	White p 值	名义 p 值
上证指数	7.31%	0.572	0.068	39.90%	0.664	0.014
青岛啤酒	1.96%	0.708	0.196	34.56%	1.000	0.012
新世界	- 9.56%	0.920	0.486	39.33%	0.920	0.142
浦东金桥	- 5.95%	0.908	0.484	47.22%	0.922	0.022
爱建股份	5.88%	0.616	0.214	57.19%	0.816	0.028
原水股份	- 6.29%	0.998	0.604	48.03%	0.974	0.042
深证成指	26.04%	0.406	0.230	49.45%	0.536	0.024
深发展 A	39.67%	0.254	0.130	70.22%	0.316	0.014
深万科 A	43.53%	0.224	0.028	69.93%	0.448	0.004
深安达 A	- 4.55%	0.978	0.738	77.60%	0.704	0.042
深物业 A	- 12.74%	0.520	0.148	43.54%	0.988	0.082
深中华 A	18.08%	0.792	0.558	60.54%	0.872	0.014
新都酒店	7.24%	0.768	0.256	56.09%	0.978	0.112

注：该表报告了各股票指数和样本股票的最优技术交易规则的业绩估计量（年平均收益）和 White 真实性检验 p 值及名义 p 值。名义 p 值是将试算法仅应用于最优交易规则，从而忽略数据窥察效应。

所使用的技术规则中包括成交量指标，因此同时也对成交量数据除权^①。表 1 报告了各股票指数和样本股票的最优技术交易规则的业绩估计量（平均年收益^②）和 White 真实性检验 p 值及名义 p 值。

名义 p 值是将试算法仅应用于最优交易规则，从而忽略数据窥察效应。因此，两 p 值之差就代表业绩估计量中数据窥察偏倚的大小。

表 1 中的名义 p 值表示忽略数据窥察条件下，最优规则不优于基准模型的概率；而 White 的 p 值则表示调整数据窥察效应后的 p 值，即在考虑到可能存在的数据窥察效应并对其进行调整后，最优交易规则不优于基准规则的概率。

根据表 1 中的名义 p 值，在一般的置信度（10%）下，我们可以拒绝上证指数和深万科的 BLL 最优交易规则不优于基准的零假定。但在调整数据窥察效应后，对任何一只样本股指和股票都无法拒绝零假定，各只样本股票和股指都存在明显的数据窥察效应，新世界、浦东金桥、原水股份和深安达等四只股票的 BLL 最优交易规则不优于基准模型的概率非常高（高于 90%）。

在一般置信度（10%）下，除新世界和新都酒店外，我们可以拒绝其他样本股指或股票的最优交易规则（选自全体规则集合）不优于基准模型的零假定。但在调整数据窥察效应的影响后，对任何一只股票指数或样本股票都无法拒绝零假定，各只股票指数和样本股票都存在相当大的数据窥察效应。另外，全体规则集合中最优规则的平均收益比 BLL 最优规则的平均收益高得多。

（二）数据窥察效应的动态演化情况

我们还可动态地考察数据窥察效应的形成过程。在现有规则集合中新增一种交易规则时，如果边际规则不优于原规则集合中的最优规则，则检验零假定的 p 值将上升，这基于以下事实：最优规则选自更大的规则集合。这里的“更大”是指规则集合的有效跨度增大了。如果边际规则提高了最大业绩统计量，则检验零假定的 p 值将下降，因为更好的业绩使最优模型（边际规则）自身包含有价值经济信息的可能性增大。图 2 演示了数据窥察效应在全体交易规则集合中的动态演化情况。Panel A 报告全体规则集合中第 n 个（ $n = 1, \dots, 7846$ ，在横轴上表示）交易规则的平均年收益（current average return，应用于深圳成指，用散点表示）和规则集合 $i = 1, \dots, n$ 的最优平均年收益（best average return）。Panel B 同步显示规则集合 $i = 1, \dots, n$ 调整数据窥察效应后的 p 值（current p -value）。当规则数 n 为 98, 2058, 6071 和 6190 时，规则集合 $i = 1, \dots, n$ 的最优平均年收益提高，调整数据窥察效应后的 p 值同步减少，这与前文的分析相符。当最优平均年收益不变时，调整数据窥察效应后 p 值的变化取决于规则集合的有效跨度，如果边际规则显著增大规则集合的有效跨度， p 值会显著增加。

① 所有除权数据资料均来自分析家证券投资分析系统。

② 这里指年平均收益，其计算方法是将样本期内的平均日收益乘以 252（每年平均交易日数），样本期内的平均日收益等于总收益除以样本期内的日数。

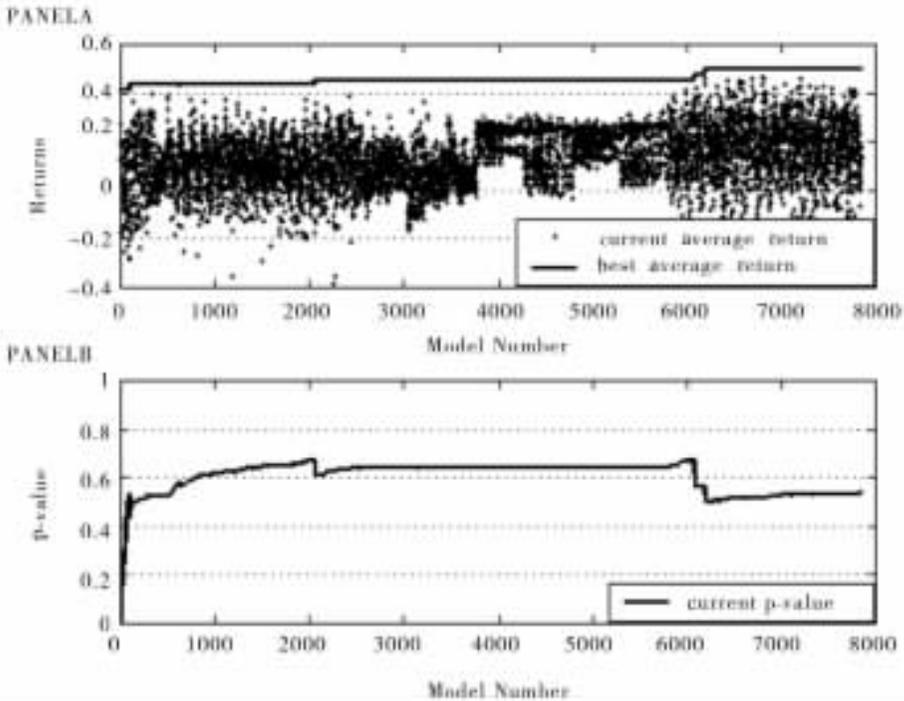


图2 平均收益最优规则的经济与统计业绩：深证成指

注：Panel A 报告全体规则集合中第 n 个 ($n = 1, \dots, 7846$, 在横轴上表示) 交易规则的平均年收益 (current average return, 应用于上证指数, 用散点表示) 和规则集合 $i = 1, \dots, n$ 的最优平均年收益 (best average return)。Panel B 同步显示规则集合 $i = 1, \dots, n$ 调整数据窥察效应后的 p 值 (current p-value)

值得注意的是, 上述动态演化情况仅反映本文所进行的特定试验中 p 值的动态调整^①, 但这仍然有助于我们理解数据窥察效应在现实世界中的传播情况。

(三) 对股票组合的实证检验

根据前文的实证结论, 我们发现各样本选自全体规则集合的最优技术交易规则确实能取得优于基准规则的收益, 但在调整数据窥察效应后, 多数规则无法在一般置信度 (10%) 下拒绝零假定。即便为避免产生数据窥察偏倚, 仅考察规则的小子集, BLL 规则集合, 对多数样本, 由于最优规则的收益仅略高于甚至略低于基准模型, 从而使真实性检验 p 值上升, 仍然无法拒绝零假定。

投资者希望从广泛发展的技术分析方法中找到确实具有预测能力 (能够拒绝零假定) 的技术规则。但如果仅将技术交易规则集合应用于个股, 显然无法达到该目的。其中一个重要原因是本文的样本期较短。BLL 指出, “通过应用非常长的数据序列”, 可以有效地避免数据窥察效应的产生。

数据窥察可视为一种数据依赖现象, 即取得超常收益的模型的预测能力依赖于特定的

① 图2中的相邻技术交易规则主要是同种技术交易规则的不同参数设置。

数据集合,反之,具有实际预测能力的模型应能够在多个股票上获得超常收益。为此,我们可将技术规则在多个股票上的平均收益作为其业绩估计量。通过计算每期收益的算术平均,可以减少偶然性因素的影响,从而使各交易规则的收益序列更能反映其实际预测能力。最终将使整个规则集合的收益方差矩阵的有效跨度减少,得到较小的 White 真实性检验 p 值。

本文应用样本股票构造 9 种投资组合,以考察数据窥察效应在股票投资组合中的变化情况。为降低组合内股价序列间的相关系数,我们先对所有的样本进行聚类分析^①,所考察的股票包括 1994 年以前在深沪两地上市的所有股票序列,共有约 230 组数据,依 Pearson 相关程度划分为 45 组,组内股票有较高相关性,组间股票的相关性较低,从每组中随机地选择一个股票作代表,分别放入 9 个投资组合中。各组合所包括的股票参见表 2。

表 2 股票组合的构成情况

组合	样本股票
组合 I	青岛啤酒、方正科技、兴业房产、轻工机械、上实联合
组合 II	组合 I、大众交通、永生股份、胶带股份、轮胎橡胶、申能股份
组合 III	组合 II、乐山电力、望春花、飞乐音响、华晨集团、青鸟天桥
组合 IV	组合 III、西南药业、国泰股份、东新电碳、辽源得亨、深万科
组合 V	组合 IV、深达声、深华源、深科技、丽珠集团、陕解放
组合 VI	组合 V、甬成功、云南白药、中天企业、豫白鸽、深华新
组合 VII	组合 VI、华北制药、鞍山信托、宏盛科技、耀皮玻璃、津劝业
组合 VIII	组合 VII、万象集团、益民百货、成商集团、天鹅股份、甬城隍庙
组合 IX	组合 VIII、凌桥股份、界龙实业、四川长虹、浙江创业、上柴股份

注:该表报告了组合 I ~ 组合 IX 的构成情况,从组合 II 到组合 IX,每种组合均在原组合基础上新增 5 个样本股票,因此组合 IX 由 45 个样本股票组成。

表 3 报告了各股票组合的最优技术交易规则的业绩估计量(年平均收益)和 White s 真实性检验 p 值及名义 p 值。分别选自 BLL 交易规则集合(共 26 个规则)和全体交易规则集合(共 7846 个规则)。平均收益的计算方法是:将持有资金平均投资于组合内的各只样本股票,交易规则分别根据各只股票的历史收益序列产生交易信号并得到技术规则收益,加总后平均。

随着组合内股票数的增加,全体规则集中最优规则的真实性检验 p 值和名义 p 值均不断下降,两 p 值之差也不断缩小,说明数据窥察效应在不断减小,至组合 VIII 和组合 IX,已经可以在一般的置信度(10%)下拒绝零假定,且最优规则的平均年收益保持在 26% 左右,说明该方法可以有效帮助投资者选择具有实际预测能力的交易规则,且不必担心由参数搜索引致的数据窥察问题。

^① 应用 SPSS 软件的 PROXIMITIES 函数实现本文中的变量聚类分析,聚类方法(cluster method)为“furthest neighbor”,测量尺度(measure)为“pearson correlation”。

表 3 平均收益标准下最优规则的业绩：组合 I ~ 组合 IX

样本股票	BLL 交易规则集合			全体交易规则集合		
	平均收益	White's p 值	名义 p 值	平均收益	White's p 值	名义 p 值
组合 I	0.09%	0.752	0.384	26.98%	0.958	0.074
组合 II	- 5.87%	0.856	0.718	23.31%	0.826	0.034
组合 III	- 4.32%	0.862	0.744	26.12%	0.634	0.056
组合 IV	- 2.21%	0.802	0.598	24.66%	0.606	0.050
组合 V	- 2.03%	0.840	0.540	25.49%	0.430	0.044
组合 VI	- 1.62%	0.844	0.524	27.03%	0.204	0.020
组合 VII	- 2.17%	0.892	0.572	26.64%	0.154	0.022
组合 VIII	- 3.03%	0.920	0.700	26.94%	0.084	0.008
组合 IX	- 2.83%	0.934	0.698	26.34%	0.060	0.004

注：该表报告了各组合的最优技术规则的业绩估计量和 White's 真实性检验 p 值及名义 p 值。名义 p 值是将试算法仅应用于最优交易规则，从而忽略数据窥察效应。平均收益指年平均收益，计算方法是：先计算每期收益，将持有资金平均投资于组合内的各只股票，交易规则分别按各只股票的历史收益序列产生交易信号并得到技术规则收益，加总后平均；接着计算样本期内平均收益，再乘以 252。

相反地，由于 BLL 规则集合中的最优规则无法取得优于基准模型的收益，其真实性检验 p 值和名义 p 值均较高，因此无法拒绝零假定。

五、结论

本文应用新实证技术——White 真实性检验 p 值，通过考察交易规则收益间的相关性，帮助研究者控制在计算投资业绩统计显著性时可能出现的数据窥察偏倚。该方法在金融领域将有广泛的应用，因为金融收益预测的信息和决策规则始终是金融领域关注的中心，然而常常被忽视的问题是，这些规则的预测能力可能源于大量研究者对具有预测能力的模型设定的广泛搜索。而本文所采用的方法提供了控制数据窥察偏倚的技术，通过计算 White 真实性检验 p 值，可以得到拒绝零假定（最优业绩规则在调整数据窥察效应后不优于基准）的置信度。

本文的方法对投资者选择有效的投资策略也有很大帮助。假定对大量决策规则进行试验后，投资者发现某种优于基准策略且能取得很高收益的规则，下一步要做的就是确定其业绩在多大程度上是数据窥察的后果，在多大程度上基于规则自身的实际预测能力。由于所考察的规则收益间存在复杂的相关关系，要实现上述目标是很困难的，而本文提供的试算法则有助于该问题的解决。

本文的实证研究结论可以概括为以下几方面：

(1) 本文分别应用全体规则集合和 BLL 的 26 种技术规则进行对比试验，发现各个股

样本选自全体规则集合的最优规则能够取得优于基准的收益，但在调整数据窥察效应后，多数规则无法在一般置信度下拒绝零假定。说明参数搜索过程中存在明显的数据窥察效应，特别是在样本期较短的情况下更为严重。

(2) 多数个股样本选自 BLL 规则集合的最优规则收益较低，仅略高于或略低于基准模型，从而使真实性检验 p 值上升，无法拒绝零假定。说明通过应用技术规则的小子集以避免数据窥察产生，其有效性值得怀疑。

(3) 研究证明，应用技术规则在多个股票上的平均收益作为其业绩估计量，可以使各规则的收益序列更能反映其实际预测能力，从而能够通过参数搜索发现具有实际预测能力（能够拒绝零假定）的技术规则。该方法特别适用在样本期较短且可获得多个平行股票样本的情况下使用。

参考文献

- [1] Brock, William, Josef Lakonishok, and Blake LeBaron, Simple technical trading rules and the stochastic properties of stock returns, *Journal of Finance*, 1992, Vol.47, pp1731 ~ 1746.
- [2] Diebold, Francis X., *Elements of Forecasting* (South-Western College Publishing, Cincinnati, Ohio), 1998.
- [3] Foster, F. Douglas, Tom Smith, and Rober E. Whaley, Assessing goodness-of-fit of asset pricing models: The distribution of the maximal, *Journal of Finance*, 1997, Vol.52, pp591 ~ 607.
- [4] Levich, Richard, and Lee Thomas, III, The significance of technical trading - rule profits in the foreign exchange market: A bootstrap approach, *Journal of International Money and Finance*, 1993, Vol.12, pp 451 ~ 474.
- [5] Lo, Andrew W., and A. Craig MacKinlay, *A Non-Random Walk Down Wall Street* (Princeton University Press, Princeton, NJ), 1999.
- [6] Lo, Andrew W., and A. Craig MacKinlay, Data-snooping biases in tests of financial asset pricing models, *Review of Financial Studies*, 1990, Vol.3, pp431 ~ 467.
- [7] Osler, Carol L., and P. H. Kevin Chang, Head and shoulders: Not just a flaky pattern, *Federal Reserve Bank of New York Staff Report*, 1995, Vol.4, pp 1 ~ 65.
- [8] Ryan Sullivan, Allan Timmermann and Halbert White, Data-snooping, technical trading rule performance, and the bootstrap, *Journal of Finance*, 1999, Vol.5, pp 1647 ~ 1691.
- [9] Politis, Dimitris, and Joseph Romano, The stationary bootstrap, *Journal of the American Statistical Association*, 1994, Vol.89, pp1303 ~ 1313.
- [10] Taylor, Stephen, Trading futures using a channel rule: A study of the predictive power of technical analysis with currency examples, *Journal of Futures Markets*, 1994, Vol.14, pp215 ~ 235.
- [11] West, Kenneth D., Asymptotic inference about predictive ability, *Econometrica*, 1994, Vol.14, pp1067 ~ 1084.
- [12] White, Halbert, A reality check for data snooping, *Econometrica*, 2000, Vol.68, pp1097 ~ 1126.

Skiba 的一个未解决问题*

韦华全 王燕鸣

一、引言

本文之群恒为有限， U 表示超可解群类。

如果存在 G 的子群 K ，使得 $G = HK$ 且 $H \cap K = 1$ ，则有限群 G 的子群 H 称为在 G 中可补 (complemented)，子群 K 称为 H 在 G 中的补 (complement)。

有限群的某些子群的补的存在性给群的结构提供了有用的信息。例如，Hall 在他的著名的系列论文中证明了，有限群可解当且仅当每个 Sylow 子群有补^[1]。他刻画了每个子群都有补的群^[2]，称这样的群为可补群 (complemented groups)，并且证明它们恰恰是 Sylow 子群皆为初等交换群的超可解群。

最近几年，考查 Sylow 子群的极大、极小子群对有限群结构的影响是一个人们感兴趣的课题。譬如，文献 [3] 引入了弱于可补子群的 c -可补子群的概念，作为应用，证明了有限群 G 超可解当且仅当 G 的每个 Sylow 子群的极大子群在 G 中 c -可补。有关 c -可补极小子群的若干结果见文献 [4, 5]。在最新版本的《The Kourovka Notebook》中，Skiba 提出了如下有关有限群 Sylow 子群极大子群是否有补的未解决问题 (见文献 [6] 问题 15.81)：

设 G 为有限非超可解群。是否存在 G 的非循环的 Sylow 子群 P ，使得 P 的某个极大子群在 G 中无真补？

本方将给出以上问题的一个肯定回答。事实上，我们利用群系理论，得到了比原问题更一般且深刻的结果。作为推论，直接给出了 Skiba 问题的肯定回答。我们也避开了使用奇阶定理和其他深刻定理，改进了文献 [3] 的一些技巧。

主要结果如下：

定理 1.1 设 F 是含 U 的饱和群系， G 是有限非 F -群且有一个正规子群 H ，使得 G/H 属于 F ，则存在 H 的非循环的 Sylow 子群 P ，使得 P 的某个极大子群在 G 中没有任何 c -补。

定理 1.1 的等价定理如下：

定理 1.1' 设 F 是含 U 的饱和群系， H 是有限群 G 的正规子群，使得 G/H 属于 F 。如果 H 的每个非循环 Sylow 子群 P ，使得 P 的某个极大子群在 G 中没有任何 c -补。

* 国家自然科学基金 (批准号：10271119) 和博士点基金及广东省自然科学基金资助项目。原载：《中国科学 A 辑数学》，2004 年第 34 卷第 3 期。

取 F 为超可解群类。设 $H = G$ 并以较强的可补假设取代 c -可补假设。Skiba 问题的肯定回答成为定理 1.1' 的直接推论。

对偶地考虑极小子群的情形，获得：

定理 1.2 设 F 是含 U 的饱和群系，假设 G 是有限非 F -群且有一个正规子群 H ，使得 G/H 属于 F ，则广义 Fitting 是子群 $F^*(H)$ 有非循环的 Sylow 子群 P ，使得 P 的某个极小子群或 4 阶循环子群在 G 中没有任何 c -补。

群类 F 称为群系，如果：(i) $G \in F$ 及 $N \triangleleft G$ 蕴含 $G/N \in F$ ；(ii) 若 $N_1, N_2 \triangleleft G$ ，使得 $G/N_1, G/N_2 \in F$ ，则 $G/(N_1 \cap N_2) \in F$ 。一个群系称为饱和的，如果 $G/\Phi(G) \in F$ 蕴含 $G \in F$ 。

二、预备知识

定义 2.1 (文献 [3] 定义 1.1) 若存在 G 的子群 K ，使得 $G = HK$ 且 $H \cap K \leq H_G = \text{Core}_G(H)$ ，则群 G 的子群 H 称为在 G 中 c -可补，称 K 为 H 在 G 中的一个 c -补。

若存在 G 的正规子群 K ，使得 $G = HK$ 且 $H \cap K \leq H_G$ ，则子群 H 称为在 G 中 c -正规^[7]。由定义 2.1 易知，可补子群或 c -正规子群必定是 c -可补子群。文献 [3] 的两个例子表明其逆不成立。

引理 2.2 (文献 [3] 引理 2.1) 设 G 是有限群，则：

- (1) 若 H 在 G 中 c -可补，且 $H \leq M \leq G$ ，则 H 在 M 中 c -可补。
- (2) 设 $N \triangleleft G$ 且 $N \leq H$ ，则 H 在 G 中 c -可补当且仅当 H/N 在 G/N 中 c -可补。
- (3) 设 π 是一个素数集合， H 是 G 的 π -子群，而 N 是 G 的正规 π' -子群。如果 H 在 G 中 c -可补，那么 HN/N 在 G/N 中 c -可补。如果 N 还正规化 H ，那么其逆也成立。
- (4) 设 $H \leq G$ 且 $L \leq \Phi(H)$ 。如果 L 在 G 中 c -可补，那么 $L \triangleleft G$ 且 $L \leq \Phi(G)$ 。

引理 2.3 (文献 [5] 定理 4.5) 设 F 是一个含 U 的饱和群系， H 是群 G 的一个可解正规子群，使得 G/H 属于 F 。如果 $F(H)$ 的每个 Sylow 子群的极大子群在 G 中 c -可补，那么 G 属于 F 。

引理 2.4 (文献^[5]定理 4.1) 设 \mathcal{F} 是一个含 \mathcal{U} 的饱和群系， H 是群 G 的一个可解正规子群，使得 G/H 属于 \mathcal{F} 。如果 $F(H)$ 的每个 Sylow 子群的极大子群在 G 中 c -可补，那么 G 属于 \mathcal{F} 。

引理 2.4 (文献 [5] 定理 4.1) 设 \mathcal{F} 是一个含 \mathcal{U} 的饱和群系， H 是群 G 的可解正规子群，使得 G/H 属于 \mathcal{F} 。如果 $F(H)$ 的每个极小子群及 4 阶循环子群在 G 中 c -可补，那么 G 属于 \mathcal{F} 。

设 G 是一个群。 G 的广义 Fitting 子群 $F^*(G)$ 是 G 的惟一极大拟幂零正规子群。它是 Fitting 子群 $F(G)$ 的自然推广，它的定义及重要性质可见文献 [8]。在此我们给出下述基本事实以便随后应用。

引理 2.5 设 G 为群， N 是 G 的子群。

- (1) 如果 N 在 G 中正规，那么 $F^*(N) \leq F^*(G)$ ；

(2) 若 $G \neq 1$, 则 $F^*(G) \neq 1$; 事实上, $F^*(G)/F(G) = \text{Soc}(F(G)C_G(F(G)))/F(G)$;

(3) $F^*(F^*(G)) = F^*(G) \geq F(G)$; 如果 $F^*(G)$ 可解, 那么 $F^*(G) = F(G)$ 。

引理 2.6 (文献 [9] 引理 3.2) 设 \mathcal{F} 是含 \mathcal{C} 的饱和群系, H 是群 G 的正规子群, 使得 G/H 属于 \mathcal{F} . 如果 $F^*(H)$ 的每个极小子群及 4 阶循环子群在 G 中 c -正规, 那么 G 属于 \mathcal{F} .

引理 2.7 (文献 [5] 引理 2.8) 设 M 是群 G 的极大子群, P 是 G 的正规 p -子群, 使得 $G = PM$, 这里 p 是素数, 则:

(1) $P \cap M$ 是 G 的正规子群。

(2) 若 $p > 2$ 且 P 的极小子群皆在 G 中正规, 则 M 在 G 中的指数是 p 。

三、定理的证明

为证明定理 1.1 并使之易于理解, 我们引入一个关键命题。称群 G 有超可解型 Sylow 塔, 如果 $|G| = p_1^{a_1} p_2^{a_2} \dots p_k^{a_k}$, 其中 $p_1 < p_2 < \dots < p_k$, 使得 Hall 子群 $P_k, P_k P_{k-1}, \dots, P_k P_{k-1}, \dots, P_2$ 在 G 中正规 (其中的 P_i 为 G 的 Sylow p_i -子群)。

命题 3.1 设 \mathcal{F} 是含 \mathcal{C} 的饱和群系, H 是群 G 的正规子群, 使得 G/H 属于 \mathcal{F} . 如果 H 的每个非循环 Sylow 子群的极大子群在 G 中 c -可补, 那么 H 有超可解型 Sylow 塔 (为了叙述简便, 下面所述的 Sylow 塔均指超可解型 Sylow 塔)。

证 假若命题不真并设 G 为极小阶反例。特别地, H 没有 Sylow 塔, 则有:

(a) $H = G$ 。

否则 $H < G$. $H/H \in \mathcal{F}$, 且由引理 2.2 (1), H 的任一非循环 Sylow 子群的极大子群皆在 H 中 c -可补, 所以 H 满足命题的假设. G 的极小选择蕴含 H 有 Sylow 塔, 矛盾, 因而 $H = G$ 。

(b) 对 G 的每个素数幂阶的非平凡正规子群 L , G/L 有 Sylow 塔。

不妨设 L 的阶为 r^k , 并设 R 是 G 的 Sylow r -子群, 我们只需证 G/L 满足命题的条件. 设 $\bar{Q} = QL/L$ 是 G/L 的一个非循环 Sylow q -子群. 可假定 Q 是 G 的 Sylow q -子群. 显然 Q 非循环. 若 $r = q$, 可设 $L \leq R = Q$, 则 \bar{Q} 的每个极大子群 (如果存在, 否则讨论在空集里皆为真) 具有形状 Q_1/L , 其中 Q_1 是 Q 的极大子群. 由假设, Q_1 在 G 中 c -可补, 故由引理 2.2 (2), Q_1/L 在 G/L 中 c -可补. 现设 $r \neq q$, 并设 \bar{Q}_1 是 \bar{Q} 的一个极大子群. 不失一般性, 可设 $\bar{Q}_1 = Q_1 L/L$ 而 Q_1 是 Q 的极大子群. 因 Q_1 在 G 中 c -可补, 据引理 2.2 (3), $Q_1 L/L$ 在 G/L 中 c -可补, 这样 G/L 满足命题的条件. 由 G 之极小性, G/L 有 Sylow 塔。

(c) $F(G) = O_r(G)$, 其中 r 为 $|G|$ 的某个素因子, 并且 $\Phi(G) = 1 = \Phi(F(G))$ 。

因为 Sylow 塔群类构成饱和群系 (见文献 [10]), 所以由 (b) 知道至多有一个素数 r , 使得 $O_r(G) \neq 1$, 因而 $|F(G)|$ 至多有一个素因子 (或许 $F(G) = 1$)。此时, 因 $\Phi(G) \leq F(G)$, $\Phi(G)$ 或是单位元群或是素数幂阶子群. 在后者, (b) 蕴含 $G/\Phi(G)$ 有 Sylow

塔从而 G 亦然，矛盾。

现设 p 是 $|G|$ 的最小素因子。考虑下列两种情形：

(d) $O_p(G) \neq 1$ 。设 N 是 G 的含于 $O_p(G)$ 的极小正规子群。由 (c)， $F(G) = O_p(G)$ 是初等交换 p - 群。由 (b)， G/N 有 Sylow 塔从而 G 可解。由饱和群系的性质， N 是 G 的惟一极小正规子群。再由 (c)，存在 G 的极大子群 M ，使得 $G = NM$ 且 $N \cap M = 1$ 。易知 $N \leq F(G)$ ，且由 $F(G)$ 为初等交换知 $F(G) \cap M$ 是 G 的正规子群，故 $F(G) \cap M = 1$ ，而 $F(G) = N$ 。设 r 为 $|M|$ 的最大素因子，并设 M_r 为 M 的 Sylow r - 子群，则 M_r 是 M 的正规子群，因为 $M \cong G/N$ 有 Sylow 塔，因此 NM_r 是 G 的正规子群。显然， $p < r$ 。因 N 不含于 M_r ，由 M 在 G 中的极大性，我们有 $N_G(M_r) = M$ 。设 G_p 是 G 的一个 Sylow p - 子群，则 $G_p M_r = G_p N M_r$ 是 G 的子群且是 Hall 子群。这样 $G_p M_r$ 的每个非循环 Sylow 子群的极大子群在 $G_p M_r$ 中有一个 c - 补。若 $G_p M_r < G$ ，则由 G 之极小性， $G_p M_r$ 有 Sylow 塔。这就得 M_r 在 $G_p M_r$ 中正规，故 $NM_r = N \times M_r$ ，从而 $M_r \leq C_G(N) = C_G(F(G)) \leq F(G) = N$ ，矛盾。现设 $G = G_p M_r$ 。假如 G_p 循环，那么 G 为 p - 幂零， M_r 是 G 的正规子群，导致矛盾，故设 G_p 非循环。显然， $M = (G_p \cap M)M_r$ 且 $G_p \cap M < G_p$ 。现设 G_1 是 G_p 的包含 $G_p \cap M$ 的一个极大子群，则 L 不含于 G_1 ，否则 $G_p = G_p \cap (NM) = N(G_p \cap M) \leq G_1$ ，矛盾，因此 $(G_1)_G = 1$ 。既然 G_1 在 G 中 c - 可补， G_1 在 G 中有补子群 K_1 。现 $|K_1|_p = p$ ，故 K_1 有正规 p - 补 K_{1p} ，在这种情形下， K_{1p} 实为 G 的 Sylow r - 子群。由 Sylow 定理， G 中有元素 g ，使得 $K_{1p}^g = M_r$ 。进一步， $K_1^g \leq N_G(K_{1p}^g) = N_G(M_r) = M$ 。 K_{1p} 被 K_1 正规化，故 g 可看做 G_1 的元。这样有 $G = G_1 K_1^g = G_1 M$ 。最后 $G_p = G_1 (G_p \cap M) = G_1$ ，也是矛盾。

(e) $O_p(G) = 1$ 。如果 G 的 Sylow p - 子群 G_p 循环，那么 G 有正规 p - 补，记为 K ，由于 K 的 Sylow 子群的极大子群实际上是 G 的 Sylow 子群的极大子群，所以 K 满足命题的假设，故由 G 的极小性 K 有 Sylow 塔。在这种情形下，因为 p 是 $|G|$ 的最小素因子， G 也有 Sylow 塔，与选择矛盾环。现假设 G_p 非循环。设 G_0 是 G_p 的极大子群，则由命题的假设存在子群 K_0 ，使得 $G = G_0 K_0$ 且 $G_0 \cap K_0 = 1$ 。由于 K_0 的 Sylow q - 子群的极大子群也是 G 的 Sylow q - 子群的极大子群（对 $p \neq q$ ）或是 1 （当 $p = q$ 时），故 K_0 满足命题的假设， G 之极小性蕴含 K_0 有 Sylow 塔。现设 q 是 $|K_0|$ 的最大素因子（从而也是 $|G|$ 的最大素因子）， Q_0 是 K_0 的 Sylow q - 子群，并设 K_{0q} 是 K_0 的一个 Hall q - 子群，则 Q_0 在 K_0 中正规且 $K_0 = K_{0q} Q_0$ 。记 $N = N_G(Q_0)$ ，则 $K_0 \leq N$ ，故 $G = G_0 K_0 = G_p N$ 。如果 $G_p \cap N = G_p$ ，那么 $N = G$ ，即 Q_0 在 G 中正规。由 (b)， G/Q_0 有 Sylow 塔；结果 G 有 Sylow 塔，导致矛盾，故 $G_p \cap N < G_p$ 。设 G_1 是 G_p 的包含 $G_p \cap N$ 的一个极大子群，又设 K_1 是 G_1 在 G 中的补。类似的讨论可知 K_1 有 Sylow 塔。设 Q_1 是 K_1 的 Sylow q - 子群， K_{1q} 是 K_1 的 Hall q - 子群，则 Q_1 是 K_1 的正规子群且 $K_1 = K_{1q} Q_1$ 。由于 Q_1 及 Q_0 都是 G 的 Sylow q - 子群，由 Sylow 定理， G 中存在元素 g ，使得 $Q_1^g = Q_0$ 。但 Q_1 被 K_1 正规化，故 g 可看做 G_1 的元。这样我们有 $K_1^g \leq N_G(Q_1^g) = N_G(Q_0) = N$ ，从而 $G = G_1 K_1^g = G_1 N$ 。进一步 $G_p = G_1 (G_p \cap N) = G_1$ ，矛盾。

这就是说反例不存在。定理证毕。

定理 1.1 的证 设 q 是 $|H|$ 的最大素因子, Q 是 H 的 Sylow q -子群. 由命题 3.1, Q 在 H 中正规, 当然 Q 也在 G 中正规. 往证 H/Q 的每个非循环 Sylow 子群的极大子群在 G/Q 中 c -可补. 设 RQ/Q 是 H/Q 的一个非循环的 Sylow r -子群, 其中 R 是 H 的 Sylow r -子群且 $r \neq q$, 则 R 非循环, 故 R 的每个极大子群在 G 中 c -可补. 由引理 2.2 (3), RQ/Q 的每个极大子群在 G/Q 中 c -可补, 从而由归纳法得 G/Q 属于 \mathcal{S} . 此时, 如果 Q 循环, 那么 Q 的每个子群在 G 中正规; 如果 Q 非循环, 那么由定理的假设, Q 的每个极大子群在 G 中 c -可补. 最后由引理 2.3, G 属于 \mathcal{S} .

定理 1.2 的证 假设定理不真, 则 $F^*(H)$ 的非循环 Sylow 极小子群及 4 阶循环子群在 G 中 c -可补. 我们有:

(1) $F^*(H) = F(H) \neq 1$. 首先, $F^*(H)$ 必定是 p -幂零的, 这里 p 是 $|F^*(H)|$ 的最小素因子. 事实上, 假定 $F^*(H)$ 非 p -幂零, 则 $F^*(H)$ 含有一个极小非 p -幂零子群 N . 由 Itô (参见文献 [10] IV, 定理 5.4) 和 Schmidt (参见文献 [10] III, 引理 5.2) 的结果, N 有正规 Sylow p -子群 N_p 和循环 Sylow q -子群 N_q , 使得 $N = [N_p] N_q$. 再者, 当 $p > 2$ 时 N_p 的方指数是 p ; 当 $p = 2$ 时 N_p 的方指数至多是 4. 设 x 是 N_p 的一个元, 则 $o(x) = p$ 或 4. 既然 N_p 非循环 (否则 N 为 p -幂零) 且 x^p 的阶至多是 2, 由引理 2.2 (1), x^p 在 N 中 c -可补, 即存在子群 K , 使得 $N = x^p K$ 且 $x^p \cap K \leq x^p N$. 所以对某整数 i 及 K 的某元 y , $x = (x^p)^i y$, 即 $x^{1-p^i} = y$. 进一步地, $x^p \leq x = x^{1-p^i} = y \leq K$, 故 $x^p = x^p N \leq x N$. 另一方面, 由于 x 在 N 中 c -可补, 故存在子群 L , 使得 $N = x L$ 且 $x \cap L \leq x N$. 注意到 $x^p L$ 是一个群, 若 $x^p L = N$, 则 $x = x^p (x \cap L) \leq x N$, 即 $x = x N$. 此时若进一步有 $x N_q = N$, 则 $x = N_p$ 循环, 矛盾. 所以 $x N_q$ 是 N 的真子群, 从而是幂零的. 这表明 N_p 正规化 N_q , 因为 x 是 N_p 的任意元, 从而 N_q 是 N 的正规 p -补, 导致矛盾. 这样 $x^p L$ 必定是 N 的真子群, 当然 $x^p L$ 幂零. 注意到 $|N : x^p L| = p$ 且 p 是 $|F^*(H)|$ 的最小素因子, 故 $x^p L$ 是 N 的一个正规子群. 易知 N_q 是 $x^p L$ 的特征子群, 所以 N_q 在 N 中正规, 这与 N 的假设矛盾. 这样 $F^*(H)$ 必定是 p -幂零群. 进一步提出 $F^*(H)$ 有超可解型 Sylow 塔. 特别地, $F^*(H)$ 可解并由引理 2.5 (2) 和 (3) 得 $F^*(H) = F(H) \neq 1$.

(2) H 非可解且 $\Phi(H) \neq F(H)$. 由 (1) 易知 $F(H)$ 的每个极小子群及 4 阶循环子群在 G 中 c -可补. 若 H 可解, 则由引理 2.4, G 属于 \mathcal{S} , 这与定理的假设矛盾, 所以 H 非可解. 假若 $\Phi(H) = F(H)$, 则由引理 2.2 (4), $F^*(H) = F(H)$ 的每个极小子群及 4 阶循环子群在 G 中正规. 由引理 2.6 的一个应用即得 H 超可解. 特别地, H 可解, 矛盾.

(3) 存在某个 $Q_p(H)$ 和 H 的某极大子群 M , 使得 $Q_p(H) \cap M$ 在 $Q_p(H)$ 中极大. 即然 $\Phi(H) \neq F(H)$, 存在某个 $Q_p(H)$ 及 H 的某个极大子群 M , 使得 $Q_p(H) \not\leq M$, 故 $H = O_p(H)M$. 为证 $Q_p(H) \cap M$ 在 $Q_p(H)$ 中极大, 只需处理下面两种情形:

① $p > 2$. 设 x 是含于 $O_p(H)$ 的一个极小子群但在 H 中非正规, 则 $O_p(H)$ 非循环, 故由引理 2.2 (1), x 在 H 中 c -可补, 即存在 H 的子群 K , 使得 $H = x K$ 且 $x \cap K = 1$. 进一步, K 在 H 中极大且由引理 2.7 (1) 知 $Q_p(H) \cap M$ 在 H 中正规. 又

$Q_p(H) = x (Q_p(H) \cap K)$ 。若 $Q_p(H) \cap K \leq M$ ，则 $H = Q_p(H)M = x M$ 满足 $x \cap M = 1$ 。这就推出 $|Q_p(H) : Q_p(H) \cap M| = |H : M| = |x| = p$ ，所以 $Q_p(H) \cap M$ 在 $Q_p(H)$ 中极大。若 $Q_p(H) \cap K \not\leq M$ ，则 $H = (Q_p(H) \cap K)M$ ，这里 x 不属于 $Q_p(H) \cap K$ 。同样的讨论可设 $Q_p(H) \cap K$ 的极小子群皆在 H 中正规。据引理 2.7 (2)， $|Q_p(H) : Q_p(H) \cap M| = |H : M| = p$ ，故 $Q_p(H) \cap M$ 在 $Q_p(H)$ 中极大。

② $p=2$ 。设 $\pi(H) = \{p_1, p_2, \dots, p_n\}$ 是 $|H|$ 的素因子集， M_{p_i} 是 M 的 Sylow p_i -子群，其中 $i=1, 2, \dots, n$ 且 $p_1=2$ 。容易知道 $O_2(H)M_2 = H_2$ 是 H 的 Sylow 2-子群。现设 P_1 是 H_2 的包含 M_2 的极大子群，并记 $P_2 = P_1 \cap O_2(H)$ ，则 $P_1 = P_2 M_2$ 。此外， $P_2 \cap M_2 = O_2(H) \cap M_2$ ，故 $|O_2(H) : P_2| = |O_2(H)M_2 : P_2 M_2| = |H_2 : P_1| = 2$ ，即 P_2 是 $O_2(H)$ 的极大子群。又对每个 $i \neq 1$ ，由引理 2.2 (1) 及引理 2.4， $O_2(H)M_{p_i}$ 超可解，故 $O_2(H)M_{p_i} = O_2(H) \times M_{p_i}$ 。进一步， $P_2 M_{p_i}$ 成群，其中 $i=1, 2, \dots, n$ ，因此 $P_2 M_{p_1}, M_{p_2}, \dots, M_{p_n} = P_2 M$ 也成群。即然 $|O_2(H) : P_2| = 2$ 且 $P_2 \cap M = Q_p(H) \cap M$ ，故 $P_2 M < O_2(H)M = H$ 。由 M 在 H 中的极大性知 $P_2 M = M$ ，故 $P_2 \leq M$ 。这样 $O_2(H) \cap M = P_2 \cap M = P_2$ ，从而 $O_2(H) \cap M$ 是 $O_2(H)$ 的极大子群。

(4) 最后的矛盾。由 (3)， $H = Q_p(H)M$ 。记 $L = Q_p(H) \cap M$ 。由引理 2.7 (1)， L 在 H 中正规且 $H/L = (Q_p(H)/L)(M/L)$ 。设 $C_{H/L}(Q_p(H)/L) = C/L$ ，则 C 是 H 的正规子群。由引理 2.5 (1)，又 $F^*(C) \leq F^*(H)$ 。又 $F^*(C)$ 的非循环 Sylow 子群的极小子群及 4 阶循环子群在 G 中 c -可补，若 $C < H$ ，则由归纳法得 C 可解。因为 $Q_p(H)/L$ 是 H/L 的 p 阶正规子群，所以 $H/C \cong (H/L) / (C/L)$ 循环，于是 H 可解，与 (2) 矛盾。若 $C = H$ ，则 $Q_p(H) \leq Z(H/L)$ 。这就推出 $Q_p(H)/L$ 中心化 M/L ，故 M/L 是 H/L 的正规子群。进而 M 在 H 中正规并由归纳法得 M 可解。再由 H/M 可解又得 H 可解，矛盾。

上述的矛盾说明定理的结论必定成立。

参考文献

[1] Hall P. , A characteristic property of soluble groups , J. London Math Soc. , 1937 , Vol.12 , pp198 ~ 200.
 [2] Hall P. , Complemented groups , J. London Math Soc. , 1937 , Vol.12 , pp201 ~ 204.
 [3] Wang Y. , Finite groups with some subgroups of Sylow subgroups c -supplemented , J. Algebra , 2000 , Vol.224 , pp467 ~ 478.
 [4] Ballerster-Bolinches A. , Wang Y. , Guo X. , c -supplemented subgroups of finite groups , Glasgow Math J. , 2000 , Vol.65 , pp383 ~ 389.
 [5] Wang Y. , Wei H. , Li Y. , A generalization of Kramer's theorem and its applications , Bull Austral. Math Soc. , 2002 , Vol.65 , pp467 ~ 475.
 [6] Mazurov V. D. , Khukhro E. I. , Unsolved Problems in Group Theory , The Kourovka , No.15 , Novosibirsk , 2002.
 [7] Wang Y. , c -normality of groups and its properties , J. Algebra , 1996 , Vol.180 , pp954 ~ 965.
 [8] Huppert B. , Blackburn N. , Finite Groups III , Berlin : Springer-Verlag , 1982.
 [9] Wang Y. , The influence of minimal subgroups on the structure of finite groups , Acta Math , Sinica , English Series , 2000 , Vol.16 , pp63 ~ 70.
 [10] Huppert B. , Endliche Gruppen I. , Berlin/Heidelberg/New York : Springer-Verlag , 1983.

基于跳跃过程的指数期权模型*

杨智元 陈浪南

一、引言

期权定价理论常用于描述标的资产运动变化过程的随机模型是连续时间和连续样本轨道的扩散过程。许多研究者如 Black & Scholes (1973)、Merton (1973)、Rubinstein (1983)、Zhang (1995) 等人均采用了扩散过程研究期权定价问题。

最早对股票指数期权进行研究的是 Evnine & Rudd (1985)。一般地,对指数期权进行的研究大多采用扩散过程描述标的资产运动变化、利用修正的 Black & Scholes 公式进行实证分析。

应该指出,如果简单利用 Black & Scholes 公式对指数期权进行定价,则会存在理论上的不一致性。一般地,如果把股票指数看作一个随机过程,假设股指遵从几何布朗运动,则可利用 Black & Scholes 公式进行定价。然而,由于对数正态分布的算术加权平均不再是对数正态分布,这样,如果个股的价格是对数正态分布的,则股票指数的“价值”就不应是对数正态分布。换句话说,股票指数按几何布朗运动变化与个股按几何布朗运动变化的假设是相互矛盾的。

Bailey & Stulz (1989) 给出了一个不同于 Black & Scholes 分析框架的指数期权定价模型。他们采用一般均衡分析研究股票指数期权定价问题。他们的研究对象是一个具理想市场、连续交易的简单生产经济。

Bailey & Stulz 首先给出了指数的波动率为常数情形下的股票指数期权定价公式。然后他们把波动率为常数的情形推广到波动率为随机的情形并给出相应的定价公式。

Bailey & Stulz 的模型同样是采用扩散过程模型研究期权定价问题,他们的模型推广了 Merton (1973) 的随机利率模型和 Hull & White (1987) 的随机波动率模型。在波动率或者利率是随机的情形下,利用 Black & Scholes 公式对期权进行定价时将产生显著的偏差。他们的研究能够对这一误差进行一定程度的修正,并且其随机利率模型可描述利率的均值回复现象。然而,一般均衡分析所得出的模式是不容易进行实证检验的,在应用上也有一定的难度。而且,他们的模型在理论上有一不足之处:模型中的随机利率不能给出符合现实的利率期限结构。

描述标的资产运动变化过程的连续时间随机模型还有一类是不连续样本轨道的跳跃过

* 本文为国家自然科学基金项目 (79870039) 以及中加教育合作课题 (CCUIPP) 1999 成果之一。原载:《经济研究》,2001 年第 2 期。

程。不过，按 Runggaldier (1997)，采用随机点过程研究金融问题的仅有 Babbs & Webber (1993)，Bjork (1995)，Bjork, Kabanov & Runggaldier (1995)，Bjork, Masi, Kabanov & Runggaldier (1995)，Jarrow & Madan (1995a)，Jarrow & Madan (1995b)，Lindberg, Orzag & Perraudin (1995)，Shirakawa (1991) 等^①，再加上早期研究股票期权定价的 Cox & Ross (1975)，Cox & Ross (1976) 以及 Merton (1976) 等。相比于被大量采用的扩散过程模型，跳跃过程方面的研究还显得很不足。

应用扩散过程进行研究的最大优点是能够对证券价格变动给予较好的模拟^②。另外，在利用扩散过程研究期权定价时，所得到的定价方程往往是一个可转化为热传导方程的二阶线性抛物形方程，能够利用标准的数学物理方程理论进行求解。而且所得的价格公式往往较为简捷。相比之下，跳跃过程对证券价格变动则拟合得较差。而且，利用跳跃过程构建模型经常需要加入许多限制，这使得模型与现实会有较大的偏差。此外，定价方程的求解一般也比较复杂，有时甚至没有解析解^③。

尽管使用跳跃过程进行研究有其不足之处。然而，跳跃运动过程有其特殊性：在一定的条件下，按跳跃运动过程变化的总体，其各个个体的变动可以相互独立。这一特殊性是扩散过程所无法描述的。因此，分析在跳跃过程之下的期权定价在理论和实践上均具有重要意义。笔者准备在这一部分对 Cox & Ross (1975) 以及 Cox & Ross (1976) 的一个模型进行修正。导致本修正模型的动因是：笔者认为可以利用此修正的模型对股票指数期权进行定价。

在 1975 年，Cox & Ross 研究了基于跳跃过程的期权定价问题。

$$dS/S = \mu dt + \begin{cases} \lambda dt & k-1 \\ 0 & 1-\lambda dt \end{cases}$$

Cox & Ross (1976) 则分析了一般的跳跃过程：

$$dS = \mu x dt + \begin{cases} \lambda x dt & k-1 \\ 0 & 1-\lambda x dt \end{cases}$$

其中： x 为状态变量，Cox & Ross 设定 $x = S$ 。

Cox & Ross 对 $\mu(S)$ 、 $\lambda(S)$ 、 $k(S)$ 的形式给予不同的设定，分析了各种不同的跳跃过程

① Runggaldier 所指的研究是指利率理论以及利率期权方面的研究。参见 Runggaldier (1997) 第 114 ~ 115 页。不过，据笔者所知，其他种类期权定价方面的研究也基本上不采用跳跃过程模型 [美式期权方面的研究可参见 Salquek (1997)；新型期权方面的研究可参见 Zhang (1997)；另外，Gemmill (1993) 对各种种类期权的定价研究有充分介绍]

② 扩散过程模型对一些经济数据的时间序列模拟得十分不错，特别是模拟股票和债券的价格变动。参见陈舜：《期权定价理论及其应用》第 13 页。

③ 构建跳跃过程模型时一般要引入一些限制，如 Cox & Ross (1975, 1976) 的模型假设股票价格只会发生跳跃变化，而且只发生正的跳跃；Merton (1976) 的模型假设跳跃是非系统风险，只需把扩散过程的风险消除，则资产组合的收益为无风险收益；标的资产的价格一般不会按这些过程变化。Cox & Ross (1976) 的文章研究了三个跳跃模型。他们给出了第一个模型的解。对于第二个模型，他们认为可以用风险中性定价求解。他们给出了标的资产价格变量的概率密度函数，利用此概率密度函数可以进行估计，但无法得出简捷的解 (closed form solution)，见 Cox & Ross (1975, 1976) 第 159 页。第三个模型如果含有 μ ，则同样没有简捷的解。

的性质。在 $\lambda(S) = \lambda \cdot S$ 的情形下，由于组成总体的各个个体是独立变化的，这一特性无法描述股票的变化，Cox & Ross 集中精力研究与跳跃过程的极限相对应的扩散过程的情形。在利用 Kolmogorov 向后方程得出极限过程之后，他们给出一些定价公式^①。

1976 年，Merton 研究了基于跳跃扩散过程的股票期权的定价问题。

$$dS/S = \alpha - \lambda k dt + \sigma dW + dq = \alpha - \lambda k dt + \sigma dW + \begin{cases} \lambda dt & k - 1 \\ 0 & 1 - \lambda dt \end{cases}$$

应当指出的是，Merton 的标的资产模型由跳跃过程和扩散过程混合而成，而且模型中的 k 为随机变量，这一模型推广了 Cox & Ross (1975) 的跳跃模型。不过，Merton 对跳跃部分未加以定价，而是假设跳跃部分为非系统风险，无需风险溢价。因此，对于分析纯粹跳跃的运动过程，这一方法是不可行的。

二、标的资产价值运动过程的假设

我们首先对市场环境以及市场参与者做一个假设：假设市场是理想市场（资产可无限细分、无税、无交易费用、允许卖空）；市场是连续运作的，交易连续不断地进行；无风险利率 r 为常数；市场参与者是价格接受者，他们的参与无法改变价格的变动过程；市场参与者对证券价格的变化有着完全一致的认识；所有市场参与者可按无风险利率借入或贷出资金。

在某一时刻，考察一个由 N 个个体组成的总体。假设每一个个体的价值为 1，每一个个体按如下过程变化：

(1) 在区间 $(t, t + \Delta t)$ 内，个体将确定性增长 $\alpha \Delta t$ 。

(2) 除了确定性的增长之外，个体将按 Poisson 过程变化：在区间 $(t, t + \Delta t)$ 内，个体发生一个跳跃变化（增加 1 个单位价值）的概率是 $(\lambda \Delta t) + o(\Delta t)$ ；个体发生多于一个跳跃变化的概率是 $o(\Delta t)$ ；个体不发生跳跃变化概率是 $1 - (\lambda \Delta t) + o(\Delta t)$ 。

另外，个体之间发生跳跃与否相互独立。

个体发生变化过程如下所示：

$$\Delta G = \alpha \Delta t + \begin{cases} \lambda \Delta t & 1 \\ 0 & 1 - \lambda \Delta t \end{cases}$$

这里， ΔG 为个体价值的增加量。

考虑总体的变化：首先，个体不发生跳跃的概率为 $1 - (\lambda \Delta t)$ ，由于每一个个体发生跳跃与否相互独立，故总体不发生跳跃（每一个个体都不发生跳跃）的概率为：

$$[1 - (\lambda \Delta t)] [1 - (\lambda \Delta t)] \cdot \dots \cdot [1 - (\lambda \Delta t)] = [1 - (\lambda \Delta t)]^N = 1 - N(\lambda \Delta t) + o(\Delta t)$$

^① Cox & Ross 给出当 $dS = \mu S dt + \sigma \sqrt{S} dW$ 以及 $dS = \mu S dt + \sigma dW$ 时的期权定价公式。就 $\lambda(S) = \lambda \cdot S, k(S) = 1 + k - 1$ 的情形，Cox & Ross 利用 Feller (1966) 关于纯生过程的结论，给出一个定价公式。参见 Cox & Ross (1976)。

因此，总体发生跳跃的概率为 $1 - [1 - (\lambda\Delta t)]^N = \lambda N\Delta t + o(\Delta t)$ 。

应当指出，只要有一个或多个个体发生跳跃，我们就说总体发生跳跃。换句话说，我们这里所说的总体发生跳跃并没有限制要多少个个体发生跳跃。当 N 个个体中有 b_1 个发生跳跃时，总体跳跃的幅度为 b_1 ；当 N 个个体中有 b_2 个发生跳跃时，总体跳跃的幅度为 b_2 。因此，总体发生跳跃的幅度应是一个随机变量。

现在我们假设标的资产价值运动过程如下：

$$dS = \alpha S dt + \begin{cases} \lambda S dt & k-1 S \\ 1 - \lambda S dt & 0 \end{cases} \quad 2.1$$

这里， q 是 Poisson 过程；跳跃以概率 $\lambda S dt$ 发生，当跳跃发生时， q 的取值为 1，如跳跃不发生，则 q 取值为 0； λ 表示 Poisson 过程的强度，它是一个常数； k 是一个固定的常数； α 是无跳跃发生时的预期收益率。

分析跳跃发生时的情形，假定在时刻 t 资产的价值为 S_t ，当跳跃发生时，资产的价值除了按无跳跃发生时的情形预期增长之外，另外，在一个瞬间，价值的增加量 $\Delta S_j = (k-1)S_t$ ，故价值由 S_t 拉升到 $S_t + \Delta S_j = S_t + (k-1)S_t = kS_t$ 。

在离散情形下：

$$\Delta S = \alpha S \Delta t + \begin{cases} \lambda S \Delta t & k-1 S \\ 1 - \lambda S \Delta t & 0 \end{cases} \quad 2.2$$

考察股票价格指数，股票价格指数由各成分股票加权平均而成。一般而言，个股是随机变动的，且个股之间变动相对独立。

我们准备利用 (2.1) 式所设定的运动过程描述股票价格指数。假设股票价格指数由各成分股票简单加总构成，个股价值总和 S 代表股指。假设个股随机变动，且个股之间变动相互独立。跳跃事件的发生（可看作某个信息的到达）将引起某些个股增值，从而引起股指的变化。

在一定程度上，我们可以利用这种运动过程描述股票指数的变化^①。

应该注意，这个过程与 Cox & Ross (1975) 的运动过程形式相近，但所描述的过程完全不同。Cox & Ross 的过程要求每一个个体的变化完全相关，一个个体变化时，每一个个体都同时发生变化。但这里所给的过程则要求每一个个体的变化完全相互独立：一个个体变化时，其他个体可能发生变化，也可能不发生变化。

^① 不过，股票指数一般是加权平均而不是简单加总。另外，这里假设跳跃的幅度为 $(k-1)S$ ， k 为一个固定的常数，股指的变动也不会如此有规则。而且，股指也不可能只发生正的跳跃。比较真实的跳跃过程应该假设 k 为一个随机变量，跳跃按生灭过程给出。

三、期权定价方程

假设有一个基于上述标的资产的期权，其价格 C 依赖于 S 和 t ，且关于 S 和 t 连续二次可微。

假设现在为时刻 t ，分析 C 的变化，过了一小段时间 Δt 后：

$$\Delta C = C(S_t + \Delta t, t + \Delta t) - C(S_t, t) \quad (3.1)$$

利用 Taylor 公式，省略 Δt 的高阶无穷小量，得：

$$\begin{aligned} \Delta C &= \frac{\partial C}{\partial t} \Delta t + \frac{\partial C}{\partial S} [\alpha S \Delta t + (k-1)S \Delta q] + o(\Delta t) \\ &= \left(\frac{\partial C}{\partial t} + \alpha S \frac{\partial C}{\partial S} \right) \Delta t + \frac{\partial C}{\partial S} (k-1)S \Delta q \end{aligned} \quad (3.2)$$

现在分析 $\frac{\partial C}{\partial S} (k-1)S \Delta q$ 项。正如在普通扩散过程中，可以把 $\frac{\partial C}{\partial S} \Delta S$ 看成 S 的变化引起 C 的变化一样， $\frac{\partial C}{\partial S} (k-1)S \Delta q$ 可看成跳跃发生引起 C 的变化。

假设跳跃发生，则在一个瞬间， $\Delta S_j = (k-1)S$ ，标的资产价格从 S 变为 $S + \Delta S_j = kS$ 。因此，当 $\Delta t \rightarrow 0$ 时，有：

$$\frac{\partial C}{\partial S} (k-1)S \Delta q = \Delta C(t) \Delta q \rightarrow [C(kS, t) - C(S, t)] \Delta q \quad (3.3)$$

这里， $\Delta C(t)$ 是在时刻 t 发生跳跃时，期权价格的变化。所以有：

$$\Delta C = \left(\frac{\partial C}{\partial t} + \alpha S \frac{\partial C}{\partial S} \right) \Delta t + [C(kS, t) - C(S, t)] \Delta q \quad (3.4)$$

我们采用构造瞬时无风险资产组合的办法来确定期权定价方程。假设所有投资者都认为股票价格指数将按 (2.1) 变化。构造瞬时无风险资产组合 V ，假设该组合由 n 份标的资产和 m 份期权组成，即：

$$V = n \cdot S + m \cdot C \quad (3.5)$$

$$\Delta V = n \cdot \Delta S + m \cdot \Delta C \quad (3.6)$$

$$V = n\alpha S \Delta t + n(k-1)S \Delta q + m \left(\frac{\partial C}{\partial t} + \alpha S \frac{\partial C}{\partial S} \right) \Delta t + m [C(kS, t) - C(S, t)] \Delta q \quad (3.7)$$

要使组合瞬时无风险，首先，风险项的系数必须为 0，即：

$$n(k-1)S + m [C(kS, t) - C(S, t)] = 0 \quad (3.8)$$

其次，预期收益率必须为无风险利率 r ，即：

$$\Delta V = rV \Delta t \quad (3.9)$$

由 (3.5)、(3.7)、(3.8)、(3.9) 四式，得：

$$n(\alpha - r)S = m \left(rC - \frac{\partial C}{\partial t} - \alpha S \frac{\partial C}{\partial S} \right) \quad (3.10)$$

由 (3.8) 和 (3.10) 两式，得：

$$(rk - \alpha)C(S, t) + (\alpha - r)C(kS, t) - \alpha S(k-1) \frac{\partial C}{\partial S} = (k-1) \frac{\partial C}{\partial t} \quad (3.11)$$

这就是对应于股票指数变化过程 (2.1) 的期权定价方程。

四、期权定价公式

方程 (3.11) 是一个差分微分方程, 笔者目前未能找到求解的方法。而且, 由于方程含有 α , 不独立于风险偏好, 无法应用风险中性定价原理进行定价。

为了应用风险中性定价方法, 我们假设 $\alpha = 0$, 即标的资产运动过程无漂移率。则方程变为:

$$rkC(S, t) - rC(kS, t) = (k - 1) \frac{\partial C}{\partial t} \quad (4.1)$$

由于方程 (4.1) 不含 α , 独立于风险偏好, 故可以应用风险中性定价原理进行定价。

我们分析基于过程 (2.1) 的欧式股票指数期权 C 。假设期权的到期日为 T , 执行价格为 X 。

在风险中性世界里, $E\left\{\frac{C_T}{C_t}\right\} = \frac{1}{C_t} E\{C_T\} = e^{r(T-t)}$, 故:

$$C_t = e^{-r(T-t)} E\{C_T\} = e^{-r(T-t)} E\{\max(0, S_T - X)\} \quad (4.2)$$

在 $\alpha = 0$ 时, 对应于 (2.1) 式, 我们有:

$$E\{\max(0, S_T - X)\} = \sum_{j \geq X} (j - X) P(S_T = j) = \sum_{j \geq X} (j - X) P\left(Y = \frac{j}{(k-1)S_t}\right)$$

化简上式, 得:

$$\begin{aligned} E\{\max(0, S_T - X)\} &= S_t e^{r(T-t)} \sum_{j \geq \left[\frac{X}{(k-1)S_t} + 1\right]} B(1 + 1/(k-1)) + 1, e^{-r(T-t)} \\ &\quad - X \sum_{j \geq \left[\frac{X}{(k-1)S_t} + 1\right]} B(1/(k-1)), e^{-r(T-t)} \end{aligned}$$

这里, $B(j; x, q) = \binom{j-1}{x-1} q^x (1-q)^{j-x}$, $\binom{j-1}{x-1} = \frac{\Gamma(j)}{\Gamma(x)\Gamma(j-x+1)}$, $\Gamma(\cdot)$ 是 Γ 函数 $[y]$

是不超过 y 的最大整数部分。

结合 (4.2) 式, 可得欧式看涨指数期权的价格为:

$$\begin{aligned} C_t &= S_t \sum_{j \geq \left[\frac{X}{(k-1)S_t} + 1\right]} B(1 + 1/(k-1)) + 1, e^{-r(T-t)} \\ &\quad - X e^{-r(T-t)} \sum_{j \geq \left[\frac{X}{(k-1)S_t} + 1\right]} B(1/(k-1)), e^{-r(T-t)} \end{aligned}$$

五、模型的优缺点

在对股票价格指数变化过程加以一些限制之后 (只发生正的跳跃变化, k 为常数, 跳跃强度为 λS) 笔者推得了基于股票价格指数的期权定价方程。在对股票价格指数变化过程再加一个限制之后 ($\alpha = 0$), 笔者求得了基于股票价格指数的欧式看涨期权价格。

一般地,如果假设组成股票指数的个股按几何布朗运动变化,个股的价格为对数正态分布。由于对数正态分布的算术加权平均不再是对数正态分布,如果采用经典的 Black & Scholes 定价分析方法,把股票价格指数看作一个随机过程,假设股指遵从几何布朗运动,利用 Black & Scholes 公式进行定价,则会存在理论上的不一致性。

本模式抓住了组成股票价格指数的个股运动变化相对独立的特点,从另一个角度描述了股票价格指数的运动变化过程,从而避免了描述股票价格指数与描述个股价格的理论不一致性。然而,本模式也存在着一些不足之处,如跳跃只是纯生过程,不确定性只会引起资产的增值,而不会减少资产的价值;还有,跳跃变化幅度比较规则(k 只是一个常数,而不是随机变量);此外,笔者所得到的定价公式是在股票指数无漂移率的情况下求出的。

对于股票指数,一个比较真实的描述应该利用生灭过程进行。股指可能发生正的跳跃,也可能发生负的跳跃。此外,跳跃的幅度应该是随机的,个股之间的变化应该是相对独立的,而不能完全独立或完全相关。

然而,由于数学处理的复杂性,到目前为止,在期权定价理论中,利用比较一般的生灭过程进行的研究还很少见。这将是期权定价研究者今后的一个努力方向。

参考文献

- [1] Bailey, W., and Stulz, R., The Price of Stock Index Options in a General Equilibrium Model, *Journal of Financial and Quantitative Analysis*, 1989, Vol.24, No.1, pp1 ~ 12
- [2] Bharucha-Reid, A. T.: 《马尔柯夫过程论初步及其应用》,杨纪珂,吴立德译,上海科学技术出版社 1979 年版。
- [3] Black, F. and Scholes, M., The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 1973, Vol.81, No.3, pp637 ~ 654.
- [4] Cox, J. C. and Ross, S. A., The Pricing of Options for Jump Processes, Rondey L. White Center Working Paper No.2 ~ 75, 1975 (University of Pennsylvania, Penn).
- [5] Cox, J. C. and Ross, S. A., The Valuation of Options for Alternative Stochastic Processes, *Journal of Financial Economics*, 1976, Vol.3, pp145 ~ 166.
- [6] Eynine, J. and Rudd, A., Index Options: The Early Evidence, *Journal of Finance*, 1985, Vol.40, pp743 ~ 756.
- [7] Gemmill, G., *Option Pricing: An International Perspective*, McGraw-Hill, 1993.
- [8] Hull, J., *Options, Futures, and Other Derivatives*, Prentice-Hall, 1997.
- [9] Merton, R. C., Theory of Rational Option Pricing, *Bell Journal of Economics and Management*, 1973, Vol.4, pp141 ~ 183.
- [10] Merton, R. C., Option Pricing When Underlying Stock Returns Are Dis-continuous, *Journal of Financial Economics*, 1976, Vol.3, pp125 ~ 144.
- [11] Runggaldier, W., *Financial Mathematics*, Springer Verlag, 1997.

安全第一准则下的动态资产组合选择*

李仲飞 姚 京

一、引言

现代资产组合理论中的投资者的行为一般描述为对风险和期望收益的权衡，这类问题的研究一般都追溯到 Markowitz 在 1952 年的先驱性工作^[1]。这之后研究工作大都是遵循均值风险的分析框架建立与现实更吻合的资产组合模型，如将 Markowitz 标准模型扩展成动态模型以及选用更为合理的风险指标来度量风险。由于动态均值 - 方差问题求解的种种困难^[2]，这块领域长期无人问津，直至近期才出现文献 [3] 对多阶段情形，文献 [4] 以及文献 [5] 对连续时间情形建立了均值 - 方差型模型，并推导了显式的最优解和有效边界。近年来也出现了许多不同的风险概念，如绝对偏差、半方差以及近年来流行起来的 VaR (Value-at-Risk) 等，本文考虑的是 Roy (1952) 首次提出的安全第一 (safety-first) 准则^[6]。Roy 认为投资者为了避免某些“灾难”水平的损失发生，在资产配置时会力图使这种水平发生的概率最小，这个思想和之后的一些变形形成了所谓的安全第一准则。尽管安全第一准则是一个对资产组合和风险管理十分有意义的指标，但对它的研究并不如方差或 VaR 那样受到关注，涉及如何将安全第一准则应用于动态资产组合选择问题近乎空白。本文考虑了连续时间情况下安全第一准则的最优资产组合选择问题，所获结果可方便地应用于动态投资决策与管理的实践中。

二、动态资产组合选择模型

考虑一个标准的 Black-Scholes 型金融市场。假设市场上存在 $n + 1$ 种资产 (或证券) $i = 0, 1, \dots, n$ ，所有的资产都在计划期 $[0, T]$ 内连续交易，其中资产 $i = 0$ 为无风险的债券，它的价格过程 $P_0(t)$ 服从普通微分方程 (确定性的)：

$$dP_0(t) = P_0(t)r dt, t \in [0, T], P_0(0) = 1$$

其中： r 为无风险资产的利率，这里假设它为常数。其余 n 种为风险资产，第 i ($i = 1, 2, \dots, n$) 种风险资产的价格过程 $P_i(t)$ 是一个特殊 Itô 的过程，满足如下随机微分方程：

$$dP_i(t) = P_i(t)(b_i dt + \sum_{j=1}^n \sigma_{ij} dB_j(t)), t \in [0, T], P_i(0) = p_i$$

* 资助项目：国家自然科学基金 (10171115)；高等学校全国优秀博士学位论文专项资金 (200267)；教育部人文社会科学研究“十五”规划项目 (01JJA630009)；广东省自然科学基金 (011193)；广东省哲学社会科学规划项目 (02M13)。原载：《系统工程理论与实践》，2004 年第 1 期。

其中： b_i 为风险资产 i 的瞬时期望回报率， $B(t) = (B_1(t), \dots, B_n(t))$ 为 n 维标准布朗运动。一般来说， b_i 和 σ_{ij} ($i, j = 1, 2, \dots, n$) 依赖于经济在时间 t 的状态，但本文假定它们均为常数。记 $b = (b_1, \dots, b_n)$ ， $\sigma = (\sigma_{ij})_{n \times n}$ 。为简便起见，我们还假定矩阵 σ 是可逆的且 $b \geq r1$ ，这里 $1 = (1, 1, \dots, 1) \in R^n$ 。

在任何时刻 t ，投资者进行资产组合决策。设 $\pi_i(t)$ 表示时刻 t 投资在资产 i 上财富的份额。记 $\pi(t) = (\pi_1(t), \dots, \pi_n(t)) \in R^n$ ，则 $\pi_0(t) = 1 - \pi(t)1$ 。资产组合过程 $\pi(t)$ 称为一个资产组合投资策略。在本文中，我们不考虑交易费和税赋，且假设资产组合投资策略 $\pi(t)$ 是自融资的。于是：

$$dW^r(t) = W^r(t)\{((1 - \pi(t)1)r + \pi(t)b)dt + \pi(t)\sigma dB(t)\}, W^r(0) = \omega$$

这里， $\omega > 0$ 是投资者的初始财富。

下面我们考虑常数再调整 (best constant-rebalanced) 资产组合投资策略的情况。所谓的常数再调整资产组合投资策略，是一个不随时间变化的向量： $\pi(t) = \pi = (\pi_1(t), \dots, \pi_n(t))$ ， $t \in [0, T]$ 。但应该注意，这样一个投资策略仍需进行动态交易，因为在每个时间点 t 都要进行头寸的调整，以保证投资在每个资产上的财富比例保持不变。

根据标准的 Ito 积分以及事实 $E[e^{\beta_1(t)}] = e^{\beta^2/2}$ (这里 E 是数学期望算子)，我们得到财富过程 $W^r(t)$ ， $t \in [0, T]$ 的如下显式公式^[5]：

$$W^r(t) = \omega \exp((\pi'(b - r1) + r - \|\pi'\sigma\|^2/2)t + \pi'\sigma B(t)) \quad (2.1)$$

$$E[W^r(t)] = \omega \exp((\pi'(b - r1) + r)t) \quad (2.2)$$

$$\text{Var}[W^r(t)] = \omega^2 \exp(2(\pi'(b - r1) + r)t)(\exp(\|\pi'\sigma\|^2 t) - 1) \quad (2.3)$$

这里， $\|\cdot\|$ 表示 R 空间中的欧氏模。

Roy 的安全第一准则是最小化终端财富 $W^r(T)$ 低于事先选定的水平 C 的概率^[6]，即：

$$\min_{\pi} P(W^r(T) < C)$$

这里， $C > 0$ 。 C 的值从投资者的角度可视为“灾难”水平，其大小反映了投资者的风险厌恶程度；而终端财富低于“灾难”水平的概率反映了风险的大小，它与人们通常对风险的看法一致。

三、最优动态投资策略

当时 $C \leq \omega \exp(rT)$ ，显然 (P) 的最优解为 $\pi^* = 0$ ，并且最优值为 0。

下面考虑 $C > \omega \exp(rT)$ 的情形。由于 $\xi := \frac{\pi'\sigma B(T)}{\|\pi'\sigma\| \sqrt{T}}$ 服从标准正态分布，我们有：

$$\begin{aligned} P(W^r(T) < C) &= P((\pi'(b - r1) + r - \|\pi'\sigma\|^2/2)T + \pi'\sigma B(t) < \ln(C/\omega)) \\ &= P\left(\xi < \frac{\ln(C/\omega) - (\pi'(b - r1) + r - \|\pi'\sigma\|^2/2)T}{\|\pi'\sigma\| \sqrt{T}}\right) \\ &= \Phi\left(\frac{\ln(C/\omega) - (\pi'(b - r1) + r - \|\pi'\sigma\|^2/2)T}{\|\pi'\sigma\| \sqrt{T}}\right) \end{aligned}$$

其中： Φ 是标准正态分布的分布函数。由于 Φ 是严格单调递增的，求解问题 (P) 等价于

求解问题：

$$(P') \quad \min_{\pi} \frac{\ln(C/\omega) - (\pi'(b - r1) + r - \|\pi'\sigma\|^2/2)T}{\|\pi'\sigma\| \sqrt{T}}$$

显然该问题又等价于如下二层优化问题：

$$(P'') \quad \min_{\epsilon \geq 0} \min_{\pi \in \Pi(\epsilon)} \frac{\ln(C/\omega) - (\pi'(b - r1) + r - \|\pi'\sigma\|^2/2)T}{\|\pi'\sigma\| \sqrt{T}}$$

其中： $\Pi(\epsilon) = \{\pi : \|\pi'\sigma\| = \epsilon\}$ 。我们先对给定的 $\epsilon \geq 0$ ，求解里层优化问题，它可写成如下等价形式：

$$\min_{\pi \in \Pi(\epsilon)} \frac{\ln(C/\omega) - (\pi'(b - r1) + r - \|\pi'\sigma\|^2/2)T}{\|\pi'\sigma\| \sqrt{T}}$$

求解该问题又等价于求解如下问题：

$$\max_{\pi} \pi'(b - r1) \quad \text{s.t.} \quad \pi'(\sigma\sigma')\pi = \epsilon^2$$

利用 Lagrange 方法，立即得到该问题的惟一最优解：

$$\pi_{\epsilon}^* = \epsilon \frac{(\sigma\sigma')^{-1}(b - r1)}{\|\sigma^{-1}(b - r1)\|}$$

带入计算得到 $\pi_{\epsilon}^*(b - r1) = \epsilon\theta$ ，其中 $\theta = \|\sigma^{-1}(b - r1)\|$ 。于是，问题 (P'') 化为：

$$\min_{\epsilon \geq 0} \frac{\ln(C/\omega) - (\epsilon\theta + r - \|\epsilon\|^2/2)T}{\epsilon\sqrt{T}} = \frac{\ln(C/W) - rT}{\epsilon\sqrt{T}} + \frac{\epsilon}{2}\sqrt{T} - \frac{\theta}{\sqrt{T}}$$

利用微分法立即得到该问题的最优解：

$$\epsilon^* = \sqrt{2(\ln(C/W) - rT)/T}$$

综上，我们已求得原问题 (P) 的最优解：

$$\pi^* = \epsilon^* \frac{(\sigma\sigma')^{-1}(b - r1)}{\|\sigma^{-1}(b - r1)\|} \quad (3.1)$$

其中 (规定 $a^+ = \max\{a, 0\}$)：

$$\epsilon^* = 2(\ln(C/W) - rT)^+ / T \quad (3.2)$$

当 $C > \omega \exp(rT)$ 时，我们有如下结论：

(1) 最优投资策略对应的风险 (即终端财富 $W^*(T)$ 低于给定水平的概率) 是：

$$\alpha := P(W^*(T) < C) = \Phi \sqrt{2(\ln(C/\omega) - rT)} - \theta\sqrt{T} \quad (3.3)$$

它是计划期 T 的递减函数，也就是说，长期投资风险更小。这一结果符合人们的一般认识。

(2) 最优投资策略对应的期望终端财富是：

$$\mu := E[W^*(T)] = \omega \exp(\epsilon^* \theta T + rT) \quad (3.4)$$

这里， ϵ^* 通过 (3.2) 式决定于 C。当 C 进一步满足：

$$C \geq \omega \exp(rT + (-r\theta\sqrt{2T} + \sqrt{2T r^2 \theta^2 + 4rT})/4)$$

时，它是计划期 T 的递增函数，也就是说，长期投资能产生更高的期望终端财富。这一结果与我们观察到的现象是相符合的。

(3) 最优投资策略对应的风险和期望终端财富满足如下关系式：

$$\alpha = \Phi\left(\frac{\ln(\mu/\omega) - rT}{\theta\sqrt{T}} - \theta\sqrt{T}\right) \quad (3.5)$$

这可通过 (3.2) ~ (3.4) 式得到。这个函数关系与投资者的偏好无关，它是一个期望终端财富的递增的、凹的函数。也就是说，高收益总是伴随着高风险，但边际风险随着财富的增加而递减。

现在我们举一个数值例子。设初始财富 $\omega = 1000$ ，市场上只有一种股票（即 $n = 1$ ）和一种无风险债券，它们的回报率分别为 $b = 0.1$ 和 $r = 0.05$ ，股票的标准差为 $\sigma = 0.2$ 。取问题 (P) 事先设定的“灾难”水平 C 为 5 年期的债券收益。让计划期在上变化，就可以得到不同的最优策略及相应的期望终端财富。这些期望终端财富随的变化情况画在图 1 的左图中，我们同时还给出了纯股票策略和纯债券策略的期望终端财富图 1：三种投资策略的比较，左图中 C 为常数，右图中 C 为 T 的函数作为 T 的函数图形。可以看出安全第一准则控制下的最优策略的期望终端财富随着计划期的增加以一个递减的速度增加，最终收敛到纯债券策略。这告诉我们同样大小的 C 在不同的计划期所起的作用不同，这同时也可以从 (3.2) 式中看出。我们可以设想 C 具有“时变”（time-varying）的特征，即投资者会针对不同的计划期相应地调整 C 的水平，如保证满足结论 (2) 中的要求。在图 1 的右图中，我们给出了这类情况，分别考虑了 $C(T) = \exp(5rT)$ 和 $C(T) = \exp(rT + 0.1r^2 T)$ 两种情形。可以看出随着计划期的增加，期望终端财富以一个递增的速度增加，即为结论 (2) 的实现。这同时也意味着投资者将更高比例的份额投资于风险资产，和文献 [5] 的结论是相似的。

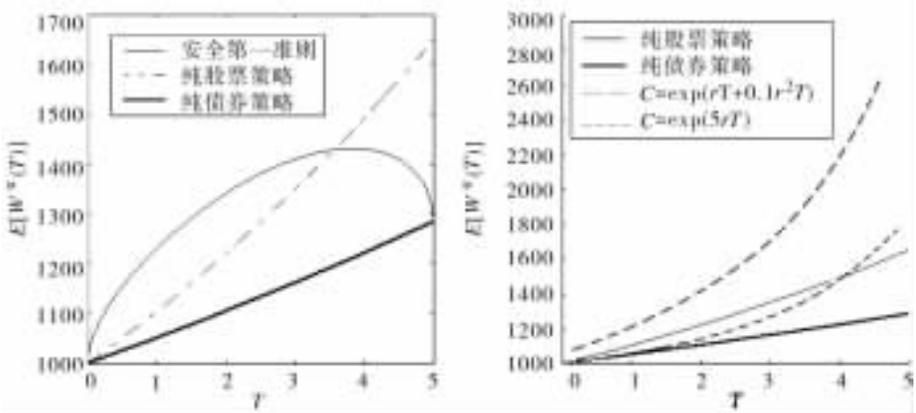


图 1 三种投资策略的比较 (左图中 C 为常数 右图中 C 为 T 的函数)

四、与均值 - 方差模型的联系以及经济解释

为了与 Markowitz 均值 - 方差问题的最优资产组合投资策略比较，考虑如下问题：

$$\max_{\pi \in R^n} E[W^*(T)] \quad \text{s.t.} \quad \text{Var}[W^*(T)] \leq C_s$$

这里， $C_s > 0$ 是方差 $\text{Var}[W^*(T)]$ 的一个给定水平。

Emmer, Kluppelberg 和 Korn (2001) 已经解出上述均值 - 方差问题的最优解形式^[5], 在此基础上我们可以得到均值 - 方差问题的有效边界的显式表达式:

$$\sigma^2[W^*(T)] = E[W^*(T)] \left[\exp\left(\frac{\ln(E[W^*(T)]/\omega) - rT}{\|\sigma^{-1}(b - r1)\|^2 T}\right) - 1 \right]$$

其中: $E[W^*(T)] \geq \omega \exp(rT)$ 。令 $m := E[\ln W^*(T)] = \ln \omega + (\pi'(b - r1) + r - \|\pi'\sigma\|^2/2)T$ 及 $s := \sigma[\ln W^*(T)] = \|\pi'\sigma\| \sqrt{T}$, 由 (2.1) ~ (2.3) 式可以得出:

$$\begin{cases} E[W^*(T)] = \exp(m + s^2/2) \\ \text{Var}[W^*(T)] = \exp(2m + s^2)(\exp(s^2) - 1) \end{cases}$$

将上述关系代入有效边界的表达式, 我们可以得到:

$$m = s \|\sigma'(b - r1)\| \sqrt{T} - s^2/2 + rT + \ln \omega, m + s^2/2 \geq rT + \ln \omega$$

很明显, 经过上述变化后可行的组合仍然在有效边界的下方。图 2 给出了在新平面上的有效边界, 注意到这里是在 (s, m) 平面上表示在 $(\sigma[W^*(T)], E[W^*(T)])$ 意义下的有效边界, 因此它的形状与常见的有效边界有所不同。类似第三节的计算, 我们可将问题 (P') 转化为:

$$\max_k k = \frac{-\ln C + m}{s}$$

其中: k 越大表示终端财富 $W^*(T)$ 小于 C 的概率越小, “风险” 越小。回到 (s, m) 平面上, 我们可以发现 $k = (-\ln C + m)/s$ 代表着一条斜率为 k , 截距为 $\ln C$ 的直线, 即 $m = k \cdot s + \ln C$ 。至此比较资产组合在安全第一指标下的有效性可以直观地理解为比较资产组合和 $(0, \ln C)$ 连线的斜率, 斜率越大即意味着“风险” 越小。这条线有明确的经济含义: ① 这条直线上的所有组合的终端财富小于所选择水平 (C) 的概率相同 (均为 $\Phi(-k)$), 即“风险” 水平相同, 因此这条线实际是在安全第一准则意义下的无差异曲线; ② 在这条直线上方的所有组合的终端财富小于 C 的概率均小于 $\Phi(-k)$, 在这条直线下所有组合的终端财富小于 C 的概率均大于 $\Phi(-k)$, 这意味着左上方的组合是较优的组合或斜率越大的无差异曲线代表越小的“风险” 水平。至此问题 (P) 可以理解为寻找均值 - 方差模型的有效边界中和 $(0, \ln C)$ 连线的斜率最大的组合, 不难发现在有效边界是凹的情况下个体的最优风险资产组合即为切点组合 (位于图 2 中的 A 点)。注意到在 Markowitz 标准模型中难以得出这样的结论: 标准模型中有效边界是一条直线, 无差异曲线与有效边界相切的情形只发生在它们重合的时候^[7]。

从对图 2 的分析中我们还可以看出 C 的选择在一定程度上反映了投资者对风险的偏好程度。当 C 取值较大时, 表示投资者对风险有较大容忍力, 切点 A 离无风险组合 (同时也是最小方差组合) 较远, 是一个波动性较大的资产组合, 相当于风险厌恶程度比较低的投资者在均值 - 方差有效边界上选择相应的组合头寸。同理, 选取较小的 C 等价于遵循更为保守的策略, 当 $C < \omega \exp(rT)$ 时, 最优的资产组合策略是将所有财富均投资于无风险资产, 此时无风险组合 $(0, \ln \omega + rT)$ 与 $(0, \ln C)$ 连线的斜率最大 (为正无穷), 有 $P(W^*(T) \leq C) = 0$ 。

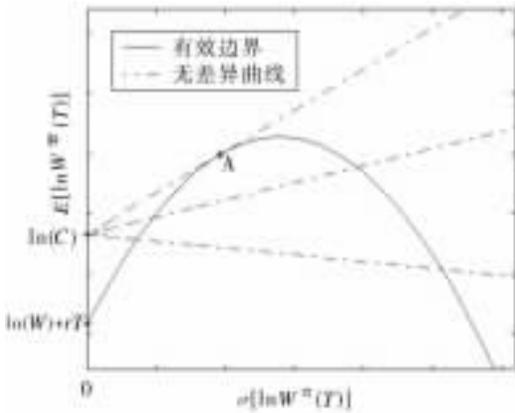


图2 无差异曲线与有效边界相切在 A 点
(其中所用的参数与图 1 同)

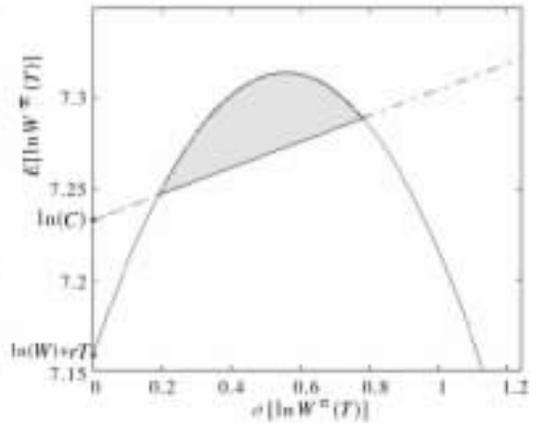


图3 阴影区域为加入约束后的可选择的资产组合
(其中所用的参数与图 1 同)

尽管在实践操作中安全第一准则还没有得到正式的推广，但安全第一准则描述风险的思想却非常适合用于监管或作为金融机构内部风险管理的目标，如监管机构可以要求金融机构承诺财富小于某一确定水平（即为 C ）的概率不能超过一个很小的概率水平（但不一定要求达到最小的概率水平）。注意到这里的要求和 VaR 的定义很相似，而事实上在很多情况下安全第一准则和 VaR 都是一个问题在不同方面的表述。根据前文的分析，监管机构施加的这个约束相当于一过 $(0, \ln C)$ 与有效边界相交的直线，金融机构只有在这条直线上方的区域选择相应的资产组合（见图 3）。从图 3 中可以发现波动率相当大的资产组合均被排除了，这说明安全第一准则有利于防止金融机构过度投机；另一方面，无风险组合附近的一些极端保守的投资策略也被排除了，这实际上是排除了那些将较大的概率水平集中在低收益或损失的区域，从成本（包括机会成本）的角度出发，这也可视作为一种风险的存在形式。综上可以看出安全第一准则不像方差一样局限在资产的波动性上，还反映了资产损益的分布情况，是一个综合性的指标。

五、结束语

本文在 Black-Scholes 金融市场设置下，导出了安全第一准则意义下的最优常数再调整资产组合投资策略的显式解析表达式，并与 Markowitz 均值 - 方差问题的最优资产组合投资策略进行比较，还结合数值例子对我们的概念与结论做了经济学解释。不仅本文的结果可方便地应用于动态投资管理与决策的实践中，而且本文的求解方法也可用于讨论一些类似的动态投资决策问题，如 Capital-at-Risk 型、Value-at-Risk 型问题。

参考文献

- [1] Markowitz H., Portfolio selection, The Journal of Finance, 1952, Vol.7, pp77~91.
 [2] Chen A., Jen C., Zions S., The optimal portfolio revision policy, Journal of Business, 1971, Vol.44,

pp51 ~ 61 .

- [3] Li D. , Ng W. L. , Optimal dynamic portfolio selection : Multiperiod mean-variance formulation , *Mathematical Finance* , 2000 , Vol.10 , pp387 ~ 406 .
- [4] Zhou X. Y. , Li D. , Continuous-time mean-variance portfolio selection : A stochastic LQ framework , *Applied Mathematics and Optimization* , 2000 , Vol.42 , pp19 ~ 33 .
- [5] Emmer S. , Kluppelberg C. , Korn R. , Optimal portfolios with bounded capital-at-risk , *Mathematical Finance* , 2001 , Vol.11 , pp365 ~ 384 .
- [6] Roy A. D. , Safety-first and the holding of assets , *Econometrica* , 1952 , Vol.20 , No.3 , pp431 ~ 449 .
- [7] Pyle D. H. , Turnovsky S. J. , Safety-first and expected utility maximization in mean-standard deviation portfolio analysis , *The Review of Economics and Statistics* , 1970 , Vol.52 , No.1 , pp75 ~ 81 .

我国寿险业粗死亡率的实证研究*

王艺明 陈浪南

一、绪论

死亡率是人类面临死亡风险的概率，而生命表则是死亡率按年龄归纳成的统计表。目前生命表被广泛应用于保险业，成为各保险公司计算保险费和责任保证金的基础。生命表中各年龄级别的死亡率是否能真正反映保单死亡率，直接关系到寿险公司是否能计算出合理的保险费率并正常地运营，以及保户缴交保险费的高低。目前我国寿险公司计算死亡率的基础为 1990~1993 年中国人寿保险业经验生命表 (China Life Insurance Mortality Table, 1990~1993)，至今已有近 10 年。由张揖平等 (1998) 的分析可知：寿险业的粗死亡率有逐年下降的趋势，这可能是因为卫生条件改善、医疗技术进步、个人对健康的重视及核保技术提高等因素所造成。然而他们使用 Logistic 模型所预估的粗死亡率在高年龄层有低估的现象。

本文的主要目的是应用具有转折点的 Logistic 模型，即在转折点前后参数不同的 Logistic 模型，对我国经验生命表所反映的保单粗死亡率作实证研究。目前，国内该领域的研究尚属空白。虽然，转折点的模型在文献上已有很多探讨，但是多数局限于以下两种情形：一种是假设独立同分布情况下，如 Hinkley & Hinkley (1970) 假设二项式分布、Hinkley (1970) 假设正态分布；另一种是假设传统的回归模型情况下，如 Broemeling & Tsurumi (1987) 和 Shaban (1980)。至于具有转折点的 Logistic 模型的讨论，目前仅见于张揖平 (1999)。

本文将以中国人寿保险业经验生命表 (1990~1993) 中，保险年龄为 15~74 岁男性和女性的保险死亡人数 (actual deaths)、暴露数 (exposed to risk) 和粗死亡率 (crude mortality rate) 等数据为研究对象，应用具有转折点的 Logistic 模型作实证分析。

二、研究模型与方法

张揖平 (1998) 发现无论男性或女性的死亡率大约在 15 岁以后均和年龄成一非线性递增关系，虽然许多分析死亡率的研究均采用修匀 (graduation) 技巧，可参考 London (1985) 的讨论，但大部分的修匀技巧均强调数值计算和平滑 (smooth) 的假设，有关参数

* 本文为国家社会科学基金课题 (03AJY008) 和厦门大学校级课题的成果之一。原载：《统计研究》，2004 年第 4 期。

模型优于一般着重平均修匀方法的原因可参考 Shaban (1980)。为此, 本文在比较分析的过程中采用的参数模型为有转折点的 Logistic 模型。

(一) 研究模型

令反应变量 (response variable) Y_i 服从二项式分布, 参数为 n_i 和 p_i , $i = 1, 2, \dots, m$ 其中 Y_i 表示死亡人数, n_i 表示暴露数, p_i 表示平均粗死亡率, Y_i/n_i 即表示粗死亡率。本文主要探讨 1990~1993 年中国寿险业男性与女性粗死亡率的预估模型。由张揖平 (1998) 可知: 若使用 Logistic 模型预估粗死亡率, 虽然大部分年龄层均预估不错, 但在高年龄层均有低估的现象, 因此本文采用具有转折点的 Logistic 模型解决该问题, 也即转折点前后参数不同的 Logistic 模型。假设影响粗死亡率的解释变量 (explanation variable) x_i 表示年龄, τ 为转折点参数, d_i 表示虚拟变量 (dummy variable), 令:

$$d_i = \begin{cases} 0 & \text{如果 } x_i \leq \tau \\ 1 & \text{如果 } x_i > \tau \end{cases} \quad (1)$$

其 Logistic 模型为:

$$\text{logit}(p_i) = \log(p_i/(1 - p_i)) = \beta_0 + \beta_1(x_i - \tau) + \beta_2(x_i - \tau)d_i \quad (2)$$

也即当年龄小于或等于 τ 时, 斜率为 β_1 , 当年龄大于 τ 时, 斜率为 β_2 , 且当年龄为 τ 时, $\text{logit}(p_i) = \beta_0$ 。因此(2)式可视为连续转折情况。(2)式经过计算, p_i 可表示为:

$$p_i = \frac{\exp(\beta_0 + \beta_1(x_i - \tau) + \beta_2(x_i - \tau)d_i)}{1 + \exp(\beta_0 + \beta_1(x_i - \tau) + \beta_2(x_i - \tau)d_i)} \quad (3)$$

其中: τ, β_0, β_1 和 β_2 均为未知参数。

(二) 检验假设与方法

在具有转折点的 Logistic 模型下, Y_i 服从二项式分布, 参数为 n_i 和 p_i , $i = 1, 2, \dots, m$ 其中 p_i 满足(2)式, 当 $\beta_2 = 0$ 表示转折点不存在, 则 $\beta_2 \neq 0$ 表示转折点确实存在。如果要检验是否有转折点存在, 考虑:

$$H_0: \beta_2 = 0 \quad \text{v. s.} \quad H_1: \beta_2 \neq 0$$

本文采用两种检验方法: 改进的似然比检验 (modification of the likelihood ratio test) (James, James & Siegmund, 1992) 和改进的沃尔德检验 (modified of the Wald's test)。

1. 改进的似然比检验

若 H_1 成立, 则似然函数 (likelihood function) 为:

$$L_{H_1}(\tau, \beta_0, \beta_1, \beta_2) = \prod_{i=1}^m (n_i, y_i)^r p_i^{y_i} (1 - p_i)^{m - y_i} \quad (4)$$

因此:

$$\begin{aligned} \log H_{H_1}(\tau, \beta_0, \beta_1, \beta_2) &= \sum_{i=1}^m \{ \log(n_i, y_i)^r + y_i \log(p_i) + (n_i - y_i) \log(1 - p_i) \} \\ &= \sum_{i=1}^m \{ \log(n_i, y_i)^r + y_i \log(p_i/(1 - p_i)) + n_i \log(1 - p_i) \} \\ &= \sum_{i=1}^m \{ \log(n_i, y_i)^r + y_i \eta_i + n_i \log(1 + \exp(\eta_i)) \} \end{aligned}$$

(5)

其中： $\eta_i = \beta_0 + \beta_1(x_i - \tau) + \beta_2(x_i - \tau)d_i$ (6)

若 H_0 成立，则：

$$\log L_{H_0}(\beta_0, \beta_1) = \sum_{i=1}^m \left\{ \log \left(\frac{n_i}{y_i} \right) + y_i(\beta_0 + \beta_1(x_i - \tau)) + n_i \exp(1 + \log(\beta_0 + \beta_1(x_i - \tau))) \right\} \quad (7)$$

因此，如果假设转折点 τ 已知，则似然比检验 (likelihood ratio test) 为：

$$\lambda(\tau) = \max_{\beta_0, \beta_1, \beta_2} \log L_{H_1}(\tau, \beta_0, \beta_1, \beta_2) - \max_{\beta_0, \beta_1} \log L_{H_0}(\beta_0, \beta_1) \quad (8)$$

其中： $\max_{\beta_0, \beta_1, \beta_2} \log L_{H_1}(\tau, \beta_0, \beta_1, \beta_2)$ 和 $\max_{\beta_0, \beta_1} \log L_{H_0}(\beta_0, \beta_1)$ 可经 Fisher-scoring 的迭代 (iterative) 方法求得 (Collett, 1991)。当 H_0 成立时 $2\lambda(\tau)$ 其分布近似 $\chi^2(1)$ 因此，如果 α 为显著性水平，当：

$$2\lambda(\tau) \geq \chi_{\alpha}^1(1) \quad (9)$$

成立时，表示有足够证据拒绝 H_0 ，其中 $\chi_{\alpha}^1(1)$ 为 $\chi^2(1)$ 右尾的第 100α 百分位。如果转折点 τ 为未知，则改进的似然比检验统计量为：

$$\max_{\tau_0+1 \leq \tau \leq m-\tau_0} \lambda(\tau) = \max_{\tau_0+1 \leq \tau \leq m-\tau_0} \max_{\beta_0, \beta_1, \beta_2} \log L_{H_1}(\tau, \beta_0, \beta_1, \beta_2) - \max_{\beta_0, \beta_1} \log L_{H_0}(\beta_0, \beta_1) \quad (10)$$

其中： τ_0 一般取为 0.05m，而只在 $\tau_0 + 1$ 和 $m - \tau_0$ 中寻找 $\lambda(\tau)$ 的极大值，其主要目的是希望转折点不会发生于边界附近。当 $2\max_{\tau_0+1 \leq \tau \leq m-\tau_0} \lambda(\tau)$ 足够大时，表示有足够证据拒绝 H_0 。可应用 Boole 不等式 (Bain & Engelhardt, 1992) 求 $2\max_{\tau_0+1 \leq \tau \leq m-\tau_0} \lambda(\tau)$ 的 p-value 的上界，p-value 定义为：

$$P_{H_0}(2\max_{\tau_0+1 \leq \tau \leq m-\tau_0} \lambda(\tau) \geq 2\max_{\tau_0+1 \leq \tau \leq m-\tau_0} \lambda_{\text{obs}}(\tau)) \quad (11)$$

其中： $\lambda_{\text{obs}}(\tau)$ 表示将观察值代入 $\lambda(\tau)$ 中，根据 Boole 不等式，有以下结果：

$$\begin{aligned} P_{H_0}(2\max_{\tau_0+1 \leq \tau \leq m-\tau_0} \lambda(\tau) \geq 2\max_{\tau_0+1 \leq \tau \leq m-\tau_0} \lambda_{\text{obs}}(\tau)) \\ &= P_{H_0} \left\{ \bigcup_{i=\tau_0+1}^{m-\tau_0} (2\lambda(i) > 2\max_{\tau_0+1 \leq \tau \leq m-\tau_0} \lambda_{\text{obs}}(\tau)) \right\} \\ &\leq \sum_{i=\tau_0+1}^{m-\tau_0} P_{H_0}(2\lambda(i) \geq 2\max_{\tau_0+1 \leq \tau \leq m-\tau_0} \lambda_{\text{obs}}(\tau)) \\ &\leq \sum_{i=\tau_0+1}^{m-\tau_0} P_{H_0}(2\lambda(i) \geq 2\lambda_{\text{obs}}(i)) = P_U \end{aligned} \quad (12)$$

如果 P_U 小于给定的显著水平 α ，则拒绝 H_0 。

2. 改进的沃尔德检验

如果转折点 τ 假设为已知，其沃尔德检验统计量为：

$$W(\tau) = \{\beta_2(\tau)\}^2 / \text{Var}(\beta_2(\tau)) \quad (13)$$

其中： $\beta_2(\tau)$ 为 β_2 的最大似然估计 (maximum likelihood estimator)； $\text{Var}(\beta_2(\tau))$ 为 $\beta_2(\tau)$ 的方差的估计量。当 H_0 成立时， $W(\tau)$ 的分布近似 $\chi^2(1)$ 因此，若 α 为显著水平，当：

$$W(\tau) \geq \chi_{\alpha}^1(1) \quad (14)$$

成立时,表示有足够证据拒绝 H_0 ,其详细结果可参考 SAS (1996)。若转折点 τ 为未知,其改进的沃尔德检验统计量为:

$$\max_{\tau_0+1 \leq \tau \leq m-\tau_0} W(\tau) \quad (15)$$

τ_0 也取为 $0.05m$ 。而其 p-value 也可用类似改进的似然比检验法做出,由 Boole 不等式求出其 p-value 的上界作为判断标准。

(三) 参数估计

当 β_2 显著不为 0 时,也即转折点确实存在,因此需要估计 τ 和其余参数 β_0 , β_1 和 β_2 。如果转折点 τ 假设为已知, β_0 , β_1 和 β_2 的估计量 $\beta_0(\tau)$, $\beta_1(\tau)$ 和 $\beta_2(\tau)$ 可以利用最大似然法求得,也即:

$$\log L_{H_1}(\tau, \beta_0(\tau), \beta_1(\tau), \beta_2(\tau)) = \max_{\beta_0, \beta_1, \beta_2} \log L_{H_1}(\tau, \beta_0, \beta_1, \beta_2) \quad (16)$$

如果转折点 τ 为未知,可先利用检验的结果求得 τ 的估计量,例如,改进的最大似然比检验法为求 τ_L 使得:

$$\lambda(\tau_L) = \max_{\tau_0+1 \leq \tau \leq m-\tau_0} \lambda(\tau) \quad (17)$$

其中: $\lambda(\tau)$ 定义于(8)式。因此 β_0 , β_1 和 β_2 的估计量可定义为 $\beta_0(\tau_L)$, $\beta_1(\tau_L)$ 和 $\beta_2(\tau_L)$ 。而改进的沃尔德检验法为求 τ_W 使得:

$$W(\tau_W) = \max_{\tau_0+1 \leq \tau \leq m-\tau_0} W(\tau) \quad (18)$$

其中: $W(\tau)$ 定义于(13)式,因此 β_0 , β_1 和 β_2 的估计量可定义为 $\beta_0(\tau_W)$, $\beta_1(\tau_W)$ 和 $\beta_2(\tau_W)$ 。

三、实证结果与分析

我们主要应用具有转折点的 Logistic 模型对我国 1990~1993 观察年度 15~74 岁男性和女性的粗死亡率数据作实证分析,应用上文介绍的方法,根据下列步骤:①取 $m = 60$, $\tau_0 = 0.05m = 3$;②对 $\tau = 4, \dots, 57$,分别计算 $\lambda(\tau)$ 和 $W(\tau)$,其中 $\lambda(\tau)$ 和 $W(\tau)$ 分别定义于(8)和(13)式,并且求得对应的 p-value,该部分可以通过 SAS 软件计算得到;③通过比较得到的 $\max_{4 \leq \tau \leq 57} \lambda(\tau)$ 和 $\max_{4 \leq \tau \leq 57} W(\tau)$,再由(12)式,可求得检验统计量 $\max_{4 \leq \tau \leq 57} W(\tau)$ 的 p-value 的上界;以同样方法可求得检验统计量的 p-value 的上界;④显著水平 α 取为 0.01,若 p-value 的上界小于 0.01,则有足够证据显示 β_2 不为 0,也即转折点显著存在,并且由(17)和(18)式可求得转折点发生在何处。

经过计算可以得到:就男性而言,使用改进的似然比检验,检验统计量 $\max_{4 \leq \tau \leq 57} \lambda(\tau) = 563.54$,其 p-value 的上界小于 0.01,使用改进的沃尔德检验,检验统计量 $\max_{4 \leq \tau \leq 57} W(\tau) = 427.05$,其 p-value 的上界也小于 0.01;就女性而言,使用改进的似然比检验,检验统计量 $\max_{4 \leq \tau \leq 57} \lambda(\tau) = 173.22$,其 p-value 的上界小于 0.01,使用改进的沃尔德检验,其检验统计量 $\max_{4 \leq \tau \leq 57} W(\tau) = 183.60$,其 p-value 的上界也小于 0.01。因此无论男性或女性的粗死亡率均有显著转折情况发生。而利用上文的参数估计方法估计未知

参数其结果如表 1 所示, 其中 MLRT 表示由改进的似然比检验法求出的估计值, MWT 表示由改进的沃尔德检验法求出的估计值, 由此可知男性的粗死亡率在 31 ~ 32 岁有一个明显的转折点发生, 而女性的粗死亡率在 28 ~ 29 岁有一个明显的转折点发生。

表 1 不同方法的参数估计

	男 性		女 性	
	MLRT	MWT	MLRT	MWT
τ	32	31	29	28
β_0	- 7.7898	- 7.7898	- 6.6231	- 7.8715
β_1	- 0.0029	- 0.0029	0.0058	- 0.0375
β_2	0.0930	0.0930	0.0930	0.1061

四、结论与今后研究方向

本文主要应用具有转折点的 Logistic 模型建立 1990 ~ 1993 观察年度中国 15 ~ 74 岁男性和女性的粗死亡率模型, 并应用改进的似然比检验法和改进的沃尔德检验法检验是否有转折点, 如果有转折点发生则估计转折点发生在何处, 并求出其余未知参数, 用以建立粗死亡率模型。结论为男性的粗死亡率在 31 ~ 32 岁有一个明显的转折点发生, 而女性的粗死亡率在 28 ~ 29 岁有一个明显的转折点发生。

就粗死亡率而言, 本文假设至多只有一个转折点, 文献上虽有一些研究假设不止一个转折点, 甚至转折点个数未知 (Yao, 1978; Meelis, Bressers & Haccou, 1991), 其作法大致为先假设至多一个转折点, 如果检验出有显著转折点时, 再由转折点前后分为两段数据, 分别讨论是否再有转折情况发生, 重复该步骤, 直到每段数据都显示无转折情况 (Srivastava & Worsley, 1986), 但文献上均是在一些较简单的模型下讨论, 如果要推广到具有转折点的 Logistic 模型将是重要而且复杂的问题, 可作为进一步研究的方向。

参考文献

- [1] Bain, L. J. and Engelhardt, M., Introduction to Probability and Mathematical Statistics, PWS - KNET Publishing Company, Boston, 1992.
- [2] Broemeling, L. D. and Tsurumi, H., Econometrics and Structural Change, Marcel Dekker, New York, 1987.
- [3] Collett, D., Modelling Binary Data, Chapman and Hall, New York, 1991.
- [4] Hinkley, D. V., Inference about the Change-Point in a Sequence of Random Variables, Biometrika, 1970, Vol.57, pp1 ~ 17.
- [5] Hinkley, D. V. and Hinkley, E. V., Inference about the Change-Point in a Sequence of Binomial Variables, Biometrika, 1970, Vol.57, pp477 ~ 488.
- [6] James, B., James, K. L., and Siegmund, D., Asymptotic Approximations for Likelihood Ratio Tests and Confidence Regions for a Change-Point in the Mean of a Multivariate Normal Distribution, Statistica Sinica, 1992, Vol.2, pp69 ~ 90.
- [7] London, D., Graduation: The Revision of Estimates, Actex, Connecticut, 1985.

- [8] Meelis, E. , Bressers , M. , and Haccou , P. , Non-parametric Testing for the Number of Change Point in a Sequence of Independent Random Variables , J. Statist. Comput. Simul. , 1991 , Vol.39 , pp129 ~ 137.
- [9] SAS , SAS/STAT Software : Changes and Enhancements through Release 6.11 , SAS Institute Inc. , North Carolina , 1996.
- [10] Shaban , S. A. , Change-Point Problem and Two-Phase Regression : an Annotated Bibliography , International Statistical Review , 1980 , Vol.48 , pp83 ~ 93.
- [11] Srivastava , M. S. and Worsley , K. J. , Likelihood Ratio Tests for a Change in the Multivariate Normal Mean , Journal of the American Statistical Association , 1986 , Vol.81 , pp199 ~ 204.
- [12] Yao , Y. C. , Estimating the Number of Change-Point via Schwarz ' Criterion , Statistics and Probability Letters , 1978 , Vol.6 , pp181 ~ 189.
- [13] 张揖平 , 林逢彰 , 陈彦良 : 《台湾地区寿险业粗死亡率的探讨》 , 见 : 《第二届海峡两岸财经与商学研讨会论文集》 , 第 127 ~ 141 页。
- [14] 张揖平 : 《寿险业粗死亡率之研究》 , 《东吴经济商学学报》 , 第 77 ~ 88 页。

股市政策对个体证券投资者交易行为的影响^{*}

——行为金融理论对我国股市低迷现状的解释

孙建军 王美今

一、引言

2001年6月12日财政部发布《减持国有股筹集社会保障基金管理暂行办法》后，股市应声下跌，而后低迷的状况延续至今，上证综指跌幅达41.3%。期间，国有股减持暂停，佣金下调、QFII准入等利好政策也相继出台，但市场并没有出现许多人预期的反弹与高涨。面对不断下跌和间或反弹的股市，人们不禁会问：国有股减持暂停、一些利好政策出台到底对个体证券投资者行为有没有影响？为此，本文从行为金融分析的角度，引入过度自信投资者来研究我国股市政策对个体证券投资者行为的影响。基于这些影响，解释股市低迷的现状^①。

证券市场的运行是众多投资者（理性与非理性投资者）“合力”的结果，对证券市场宏观层面的解释必须从市场微观主体的类型与行为出发。国外对理性投资者与过度自信投资者的投资者行为的研究比较成熟。Benos（1998），Odean（1998），Kyle与Wang（1997）分别在不同的假定下论证了市场上因过度自信投资者的存在，交易量与价格波动加剧，价格的信息传递功能增强。Hirshleifer和Luo（2001）论证了理性投资者与过度自信投资者群体规模之间存在长期均衡。这些研究为本文考察我国股市中理性投资者与过度自信投资者的交易行为及其群体规模提供了良好的理论基础。国内学者对我国“政策市”特征下投资者交易行为的相关研究才刚起步^②。李心丹（2002）用沪深两市投资者的交易数据进行了实证分析，发现股市政策对交易者的交易频率有明显的影响，得出了我国个体证券投资者存在过度自信的结论。该结论为本文引入过度自信投资者的合理性提供了现实依据。

本文从证券市场微观主体的结构与行为入手，假定理性投资者与过度自信投资者均是追求效用最大化的风险厌恶者。同时，针对我国政策市的特征，将政策性变量纳入分析框架。本文论证了：若股市政策不明朗、不稳定，过度自信投资者比理性投资者买入或卖出更多的股票；而且若股市政策明朗、稳定，过度自信投资者与理性投资者均要比股市政策不明朗、不稳定条件下买入或卖出更多的股票；过度自信投资者对政策的自信程度会影响

* 本文是国家自然科学基金项目的阶段性研究成果（项目批准号：70273060）。原载：《数量经济技术经济研究》，2004年第6期。

① 王开国（2002）从我国股市内在的、周期性的脆弱性角度阐释了股市目前的低迷。认为股市政策引发内在的脆弱性，内在的脆弱最终导致市场失去平衡能力。

② 成思危等（2003）用股市周期、政策周期与经济周期之间的互动关系论证了我国股市“政策市”特征。

这两类投资者的规模^③。具体而言,当过度自信投资者自信程度加深时,其群体规模会变大,理性投资者群体规模变小;反之,其群体规模会变小,理性投资者群体规模变大。

本文的结构安排如下:第二部分构造风险厌恶的理性投资者与过度自信投资者静态模型,分析追求财富期望效用最大化的两类投资者在股票买卖上的差异,求解他们的期望利润差额,奠定进行动态分析的基础;第三部分构造一个理性投资者与过度自信投资者的动态模型,揭示过度自信投资者在市场上自信程度的变化对两个投资者群体规模的影响;第四部分是结论与建议。

二、静态模型

假设我国证券市场是一个单期的竞争性市场,由两种类型的证券构成:一种是无风险的债券,对该债券的单位支付为常数 1;另一种证券为股票,股票的初始价格为 P ,在期内,其价格为 θ ($\theta \sim N(\bar{\theta}, \sigma_{\theta}^2)$)。同时假设市场上存在两种类型的投资者:理性投资者与过度自信投资者,这两类投资者的总人数保持不变。理性投资者与过度自信投资者都收到关于股票的信号 s ($s = \theta + \epsilon$, θ 与 ϵ 相互独立),并且都知道 θ 的分布。理性投资者能正确地理解 ϵ 的分布, $\epsilon \sim N(0, \sigma_{\epsilon}^2)$,过度自信投资者不能正确地理解 ϵ 的分布,认为 $\epsilon \sim N(0, \sigma_{\epsilon_0}^2)$,其中 $\sigma_{\epsilon_0}^2 < \sigma_{\epsilon}^2$ 。这里 $\sigma_{\epsilon_0}^2$ 与 σ_{ϵ}^2 均大于 0,表明了理性投资者与过度自信投资者都认识到他们收到的信号是不完美的, $\sigma_{\epsilon_0}^2 < \sigma_{\epsilon}^2$ 表明了过度自信投资者对股票信号的过度自信。在我国,股市政策往往不明朗和不稳定,因而政策的变化对股票收益的影响非常明显^④。本文同时考虑两类投资者对期内政策的理解。用 z 表示政策变化对每股收益的冲击。从历次重大政策的出台看,政策制定者并非将维持股市的资源配置效率放在首位,因此可以假定 z 与 s , θ 及 ϵ 相互独立。理性投资者能正确理解 z 的分布, $z \sim N(0, \sigma_z^2)$,过度自信投资者不能正确理解 z 的分布, $z \sim N(0, \sigma_{z_0}^2)$,其中 $\sigma_{z_0}^2 < \sigma_z^2$ 。这里 $\sigma_{z_0}^2$ 与 σ_z^2 均大于 0,表明理性投资者与过度自信投资者都认识到他们不能确切地把握未来的政策, $\sigma_{z_0}^2 < \sigma_z^2$ 表明了过度自信投资者自己对政策的过度自信。

理性投资者与过度自信投资者基于自己对期初信号的理解构造资产组合,使他们在期末财富的期望效用最大化。财富用 ω_i 表示($i = r$ 表示理性投资者; $i = o$ 表示过度自信投资者)。投资者的效用函数假定为 $U(\omega_i) = -e^{-a\omega_i}$,^⑤ $a > 0$ 。可以证明, a 表示绝对对风险厌恶系数。两类投资者在期末的财富由期初财富(ω)和投资于两种证券的收益构成。由于单位无风险债券的支付为常数 1,所以对投资者 i ($i = r, o$)而言,期末的财富 $\omega_i = \omega + x_i(\theta - P)$

③ 对政策的自信,本文定义为过度自信投资者或者提前获得了关于股市政策的私人信息而表现出的过度自信;或者股市政策出现后过度自信投资者对自己理解与处理政策的能力过度自信。

④ 金晓斌、唐利民(2001)的统计数据说明:1992~2000年初,政策性因素是造成股市异常波动的首要因素,占总影响的46%。在这8年的市场剧烈波动中,涨跌幅超过20%的有16次,其中政策因素8次。施东晖(2001)的实证研究列举了1992~2000年初沪市52次异常波动,由政策因素引起的波动共有30次,约占60%。该文将每月的股价波动性和当月出台的政策变量进行回归结果显示:当月出台的政策对该月的股价波动具有显著的影响。

⑤ 该效用函数有良好的一阶与二阶导数性质,Grossman与Stiglitz(1980)也运用了该效用函数作为信息交易者的效用函数。

+ $x_i z$ 里, x_i 表示投资者 i 对股票的需求量 ($x_i > 0$ 表示买入; $x_i < 0$ 表示卖出)。由于 θ 与 z 是相互独立的正态随机变量, 故期末财富也服从正态分布, 两类投资者的期望效用函数因而可以表示成均值 - 方差的效用函数^①。这样, 对投资者 i ($i = r, o$) 而言, 其对股票的需求量是下列规划问题的解:

$$\max_{x_i} E_i(\omega_i | s) - \frac{a}{2} \text{Var}_i(\omega_i | s) \quad (1)$$

$$\text{s.t. } \omega_i = \omega + x_i(\theta - P) + x_i z \quad (2)$$

由 θ 和 ε 是相互独立的正态分布变量知:

$$E_i(\theta | s) = \bar{\theta} + \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_i^2}(\theta + \varepsilon - \bar{\theta})$$

$$\text{Var}_i(\theta | s) = \frac{\sigma_\theta^2 \sigma_i^2}{\sigma_\theta^2 + \sigma_i^2} \quad (3)$$

其中: $\sigma_i^2 = \sigma_{er}^2$ (当 $i = r$ 时); $\sigma_i^2 = \sigma_{eo}^2$ (当 $i = o$ 时)。

先将 (2) 式代入 (1) 式, 后代入 (3) 式及 Z_i 的分布 (利用到 z 与 θ 的相互独立性), 由极值条件可以得到该规划问题的解:

$$x_i = \frac{\bar{\theta} + \eta_i(\theta - \bar{\theta} + \varepsilon) - P}{av_i}$$

$$\text{其中: } \eta_i = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_i^2}; v_i = \frac{\sigma_\theta^2 \sigma_i^2}{\sigma_\theta^2 + \sigma_i^2} + \sigma_z^2 \quad (4)$$

因为 $\eta_o > \eta_r$, $v_o < v_r$, 所以 $|x_o| > |x_r|$ 。该不等式表明: 在我国“政策市”环境下, 由于过度自信投资者对股票的认识上存在着较高的条件均值与较低的条件方差, 过度自信投资者比理性投资者买入或卖出更多的股票。

这两类投资者的期望利润 π_i ($i = o, r$) 为:

$$\pi_i = E(x_i(\theta - P) + x_i z) = \frac{\bar{\theta} + \eta_i \sigma_\theta^2 - 2P\bar{\theta} + P^2}{av_i} \quad (5)$$

理性投资者与过度自信投资者之间的期望利润差:

$$\begin{aligned} \Delta &= \pi_r - \pi_o \\ &= \frac{((\bar{\theta} - P)^2(\sigma_\theta^2 + \sigma_{er}^2) + \sigma_\theta^4)(\sigma_\theta^2 \sigma_{eo}^2 + \sigma_\theta^2 v_o^2 + \sigma_{eo}^2 v_o^2)}{a(\sigma_\theta^2 \sigma_{er}^2 + \sigma_\theta^2 v_r^2 + \sigma_{er}^2 v_r^2)(\sigma_\theta^2 \sigma_{eo}^2 + \sigma_\theta^2 v_o^2 + \sigma_{er}^2 v_o^2)} \\ &\quad - \frac{((\bar{\theta} - P)^2(\sigma_\theta^2 + \sigma_{eo}^2) + \sigma_\theta^4)(\sigma_\theta^2 \sigma_{er}^2 + \sigma_\theta^2 v_r^2 + \sigma_{er}^2 v_r^2)}{a(\sigma_\theta^2 \sigma_{er}^2 + \sigma_\theta^2 v_r^2 + \sigma_{er}^2 v_r^2)(\sigma_\theta^2 \sigma_{eo}^2 + \sigma_\theta^2 v_o^2 + \sigma_{eo}^2 v_o^2)} \end{aligned}$$

三、动态过程

为了研究理性投资者与过度自信投资者在长期中的情况, 本文把对两类投资者的分析纳入到一个动态演化过程中。在这个过程中, 理性投资者与过度自信投资者行为的动态演

① [美] Hal R. Varian:《微观经济学》(高级教程), 周洪等译, 经济科学出版社 1997 年版, 第 202 ~ 203 页。

化取决于他们各自的投资策略所带来的相对利润或损失的变化情况。这是合理的,因为股市上现有的投资者倾向于使用已经被实践证明能够使扩大利润或减少损失的投资策略,而新加入到股市的投资者也倾向于模仿这些策略。对投资者 i 而言,投资者 $j(i, j \in \{r, o\}, i \neq j)$ 的利润逐渐扩大或损失逐渐减少,一部分投资者 i 会加入到投资者 j 的行列中来。因此,在下一期投资者 j 的人数会增加。具体而言,令理性投资者占总投资者的人数比例为 λ , $0 \leq \lambda \leq 1$, 其遵循如下的动态演化过程:

$$\lambda_{t+1} = \lambda_t + f(\pi_t - \pi_o) = \lambda_t + f(\Delta)$$

其中: $f(\cdot): (-\infty, +\infty) \rightarrow [-1, 1]$ 是一个连续函数,且满足以下的性质: ① 若 $\Delta = \pi_t - \pi_o$ 逐渐减少,即相对于过度自信投资者而言,理性投资者的相对利润逐渐减少或相对损失逐渐加大,则 $f(\cdot) < 0$; ② 若 $\Delta = \pi_t - \pi_o$ 逐渐增加,即相对于过度自信投资者而言,理性投资者的相对利润逐渐增加或相对损失逐渐减少,则 $f(\cdot) > 0$ 。

以上的动态演化方程与 Hirshleifer 和 Luo (2001) 分析过度自信投资者能在长期中生存下来使用的动态演化方程相似。这种类型的动态演化过程建立在理性投资者与过度自信投资者总体平均结果的差异之上。例如,在每期末,投资者 A 随机地遇到另一个投资者 B。如果投资者 B 是与投资者 A 相反的一种类型(即一个是理性投资者,则另外一个就是过度自信投资者),投资者 B 在这一时期内所使用的投资策略使他自己相对于投资者 A 利润增加或损失减少,则投资者 A 会有兴趣使用投资者 B 的投资策略。否则,投资者 A 不改变自己的策略。基于此,本文提出如下命题:

命题 对所有的正参数 $(a, \sigma_\theta^2, \sigma_{\epsilon t}^2, \sigma_{\epsilon o}^2, \sigma_\pi^2, \sigma_\alpha^2)$ 而言, ① 当股市政策趋于明朗、稳定时,理性投资者与过度自信投资者均会买入或卖出更多的股票; ② 若过度自信投资者对股市政策自信程度加深,则其群体规模会变大,理性投资者群体规模变小。反之,其群体规模会变小,理性投资者群体规模变大。

证明: ① 用 v_i' 与 x_i' 分别表示政策明朗、稳定后的相应变量。若政策趋于明朗、稳定,则 σ_α^2 渐小。因为 $\lim_{\sigma_\alpha^2 \rightarrow 0} v_i = \frac{\sigma_\theta^2 \sigma_i^2}{\sigma_\theta^2 + \sigma_i^2} \equiv v_i' < v_i$, 所以 $|x_i| = \left| \frac{\bar{\theta} + \eta_i(\theta + \epsilon - \bar{\theta}) - P}{av_i} \right| < \left| \frac{\bar{\theta} + \eta_i(\theta + \epsilon - \bar{\theta}) - P}{av_i'} \right| \equiv |x_i'|$ 。所以,若政策趋于明朗、稳定,股市上的理性投资者与过度自信投资者均会买入或卖出更多的股票。② 因为由 (6) 式知, $\frac{\partial \Delta}{\partial v_o} > 0$, 所以 $\frac{\partial \Delta}{\partial v_o} =$

$\frac{\partial \Delta}{\partial v_o} \frac{\partial v_o}{\partial v_o} = 2v_o \frac{\partial \Delta}{\partial v_o} > 0$ 。该结论说明了 v_o 与 Δ 同方向变化。若过度自信者对股市政策的自信程度加深,即 σ_α^2 渐小,则 v_o 渐小, Δ 也渐小。由 (7) 式及其性质知, λ_{t+1} 减小, $1 - \lambda_{t+1}$ 增加。同理可证,当过度自信投资者变得不自信时,则其投资者比例降低,而理性投资者比例增加。证毕。

四、结论与建议

若政策不明朗、不稳定,由静态模型结论知,过度自信投资者对股票的需求量或抛售

量会比理性投资者大。这一结论说明：过度自信投资者通常会推动股市大起大落。因为牛市买入的信号多，过度自信投资者对股票的需求会超出理性，进一步推动股市向泡沫的方向转化；在熊市，卖出的信号多，过度自信投资者对股票的抛售会超出理性，进一步推动股市向萧条的方向转化。政策不明朗、不稳定情况下的熊市，过度自信投资者对股票的抛售超出理性，从而会抵销部分利好消息对股市的拉升。这一点可能是国有股减持暂停、相继利好政策出台却无助于改变股市低迷的原因之一。也由此可见，在股市狂跌后国有股减持暂停是合理的。否则，过度自信投资者的抛售行为会进一步推动萧条的形成。

若股市政策明朗稳定，股市政策所带来的风险会减小，理性投资者与过度自信投资者均会增加对股票的需求。这一点已经由定理的第一个结论得到了证明。这一结论说明国有股的减持必须具备两个必要条件：一是政策的明朗稳定；二是在时机上选择买入信号较多的牛市。因为在这两个条件下，理性投资者与过度自信投资者均会加大对股票的需求量，从而为国有股减持扩容创造价格稳定的空间。2001年6月牛市（上综指2450点左右）之时，国有股减持在时机上的选择是恰当的，但政府决策层与管理层没能给投资者明朗、稳定的向股民让利的政策信号，是股市应声下跌的一个非常重要的原因。因为在牛市，国有股以流通股股价而非协议价减持，这样就给了投资者一个政府想尽可能多地从股市上套现的“相机（牛市）抉择”政策信号。

由定理的第二个结论知，长期中，过度自信投资者对股市政策的自信程度，能决定两类投资者的规模，这个结论是有意思的。过度自信投资者的信心程度实际上是一种情绪。这种过度自信的情绪具有传染性，能传播给理性投资者，从而使过度自信投资者增加，理性投资者减少。其情绪的传播机制实际上是：投资过程中情绪能够影响过度自信投资者相对于理性投资者获利或损失的大小，相对获利或损失的大小又会使投资者改变投资策略，从而改变这两个投资者群体各自占总投资者的比例。2001年6月国有股减持政策出台，过度自信投资者的自信受挫，这种情绪的传播致使目前股市理性投资者比例上升，市场上理性投资者人数增多。因此，尽管后来有利好政策出台，但股票需求比理性投资者要多的过度自信投资者的规模在变小，故市场还是显得疲软。这一点可能是国有股减持暂停、相继利好政策出台却无助于改变股市低迷的原因之二。

近年来，股市政策不稳定也不明朗。政府决策与管理层既竭力稳定市场（如市场下跌时不断推出利好政策，召开打气座谈会，发表积极评论等），又不愿轻易放弃自己既定的政策目标（如为国企和经济发展筹集资金，推进规范化、市场化、国际化等），采用了含糊和实用的态度。这样，即使利好消息不断，两类投资者对股票的需求还是受到相当程度的限制，投资者尤其是过度自信投资者对政策的自信受挫，理性投资者人数增多，从而阻碍了股市爬升。这种不明朗、不稳定的股市政策应当尽早尽快放弃，转而向建立和遵守“游戏规则”的方向转变。具体而言，决策者与管理层应当把精力放在上市公司规则的建立与执行，对上市公司行为进行监督，以及上市公司退出机制的建立与执行上来，使投资者对政府股市政策的稳定与明朗形成一个良好的预期。全流通的政策应明确承诺让利于民，从而向投资者传递这样一个信号：政府推动全流通并不是纯粹为了套现，其主要目的还是为了重建资本市场的资源配置功能。全流通的政策应当在流通日期、方案上有连续稳定的步骤、计划，这样，实现全流通才具备良好的微观行为基础。

参考文献

- [1] 成思危主编：《诊断与治疗——揭示中国的股票市场》，经济科学出版社 2003 年版。
- [2] 曹凤岐，李学彬主编：《趋势——北大资本高级论坛》，新世界出版社 2002 年版。
- [3] 李心丹，王冀宁：《中国个体证券投资者交易行为的实证研究》，《经济研究》，2002 年第 11 期。
- [4] 金晓斌，唐利民：《政策与股票投资者博弈分析》，海通证券 2000 年研究报告。
- [5] 施东晖：《中国股市微观行为理论与实证》，上海远东出版社 2001 年版。
- [6] 王开国：《政策诱因与中国证券市场的内在不稳定性——转轨过程中的新兴市场的特征》，《经济研究》，2002 年第 12 期。
- [7] [美] Hal R. Varian：《微观经济学》（高级教程），周洪等译，经济科学出版社 1997 年版。
- [8] Benos, A. S. , Overconfident Speculators in Call Markets : Trade Patterns and Survival , Journal of Financial Market , 1998 , Vol.1 , pp353 ~ 383 .
- [9] Grossman. S. J. , Stiglitz. J. E. , On the Impossibility of Informationally Efficient Markets , The American Economic Review , 1980 , Vol.6 , pp393 ~ 408 .
- [10] Hirshleifer , D. , Luo , G. Y. , On the Survival of Overconfident Traders in a Competitive Securities Market , Journal of Financial Market , 2001 , Vol.4 , pp73 ~ 84 .
- [11] Kyle , A. S. , Wang , F. A. , Speculation Duopoly with Agreement to Disagree : Can Overconfidence Survive the Market Test ? Journal of Financial , 1997 , Vol.52 , pp2073 ~ 2090 .
- [12] Odean , T. , Volume , Volatility , Price and Profit When All Trades Are above Average , Journal of Finance , 1998 , Vol.53 , pp1934 ~ 1987 .
- [13] Paul G. Hoel , Introduction to Mathematical Statistics , New York , 1962 , p200 .

EaR 风险度量与动态投资决策*

李仲飞 汪寿阳

一、引言

投资组合选择和风险管理一直是投资金融领域重要而核心的问题。1952年 Markowitz 提出了用方差来度量风险、用均值和方差的最佳权衡来选择资产组合的均值 - 方差方法。该方法在后来引发了理论扩展与实际应用两方面的大量工作，目前已经成为现代金融理论的基础。

然而，经典的均值 - 方差方法面临着两大挑战：第一个挑战是多阶段或连续时间均值 - 方差模型的建立。Markowitz 的模型是单阶段或静态的。后来虽然有大量的文献研究动态模型，如 Merton (1968), Samuelson (1969)，但几乎都是最大化终端财富或消费的某种形式的时间可加效用函数。还有文献（如 Chen, Jen & Zions, 1971）报道说求解动态均值 - 方差模型会遇到巨大的困难。因此，均值 - 方差型动态模型的研究长期以来是一个空白领域。直到最近，才有文献 Li 和 Ng (2000) 对多阶段情形，Zhou 和 Li (2000) 以及 Emmer, Kluppelberg 和 Korn (2001) 对连续时间情形建立了均值 - 方差型模型，并推导出了显式的最优解和有效边界。第二个挑战是更合理的风险概念的提出。传统的风险概念是建立在投资收益的不确定性基础上的，用随机收益的方差（或标准差）来度量。方差具有良好的数学特性，但其缺陷也已成为人们所认识，就连 Markowitz 本人也承认：除了方差之外，也存在多种风险度量方法，其中理论上最完美的度量法应属于半方差（Semi-Variance）。为了找到既具备理论的完善性、计量的方便性，又符合实际的风险度量新方法，金融学术界与投资界的研究者和实业家进行了大量的研究，目前已提出绝对偏差、半方差、不足概率（Shortfall Probability）、安全第一准则（Safety-First）、一致风险（Coherence Risk）、受限期望损失（Limited Expected Loss）、在险价值（Value at Risk）、在险资本（Capital at Risk）以及基于底侧风险（Downside Risk）概念的 Lower Partial Moment 等多种风险度量方法。其中，在险价值（Basak and Shapiro, 2001）是最近提出并广为使用的一个风险度量，其重要性与日俱增，因为许多金融机构和市场监管部门已将其作为控制市场风险的主要方法。基于在险价值概念，Emmer, Kluppelberg 和 Korn (2001) 引入了一种在险资本（Capital at Risk）的概念，用其代替方差来度量风险，建立了动态投资组合选择的均值 - 在险资本模型，并推

* 资助项目：国家自然科学基金项目（10171115）、教育部人文社会科学研究“十五”规划项目（01JA630009）、广东省自然科学基金项目（011193）、广东省哲学社会科学规划项目（02MI3）。原载：《数量经济技术经济研究》，2003年第1期。

导出了解析形式的最优解和有效边界。

另外,近年来在许多证券组合优化问题的研究中提出了常数再调整证券组合(Constant-Rebalanced Portfolio)投资策略的概念,例如,Cover (1991),Emmer, Kluppelberg 和 Korn (2001),Helmbold 等人(1998)。所谓常数再调整证券组合投资策略,是这样一种投资策略,它保持投资在各资产上的财富份额不随时间变化。但常数再调整证券组合投资策略并不意味着不交易,因为资产价格在随时间变化,为保持财富比例不变就得不断调整头寸。

本文引入另一种风险度量——在险收益(Earnings at Risk),用它替代方差,建立连续时间动态投资组合选择的均值-在险收益模型。在 Black-Scholes 型金融市场的假设下,给出了最优常数再调整证券组合投资策略和有效边界的显式表达式,它们容易计算,可方便地应用于动态投资决策与管理的实践中。我们也试图给出概念与结论的经济学解释。

二、金融市场模型与风险度量

考虑一个标准的 Black-Scholes 型金融市场。假设市场上存在 $n+1$ 种资产(或证券) $i = 0, 1, \dots, n$, 所有的资产都在计划期 $[0, T]$ 内连续交易,其中资产 $i = 0$ 为无风险的债券,它的价格过程 $P_0(t)$ 服从普通微分方程(确定性的):

$$dP_0(t) = P_0(t)r dt, \quad t \in [0, T], \quad P_0(0) = 1$$

其中: r 为无风险资产的利率,这里假设它为常数。其余 n 种为风险资产,第 i ($i = 1, 2, \dots, n$) 种风险资产的价格过程 $P_i(t)$ 是一个特殊的 Ito 过程,满足如下随机微分方程:

$$dP_i(t) = P_i(t) \left(b_i dt + \sum_{j=1}^n \sigma_{ij} dB_j(t) \right), \quad t \in [0, T], \quad P_i(0) = P_i$$

其中: b_i 为风险资产 i 的瞬时期望回报率; $B(t) = (B_1(t), \dots, B_n(t))$ 为 n -维标准布朗运动,它的每个分量表示市场中外生的、独立的不确定性因素; σ_{ij} 是影响第 i 种资产的第 j 个不确定性因素的瞬时标准差。一般来说, b_i 和 σ_{ij} ($i, j = 1, 2, \dots, n$) 依赖于经济在时间 t 的状态,但本文假定它们均为常数。记 $b = (b_1, \dots, b_n)$, $\sigma = (\sigma_{ij})_{n \times n}$ 。为简单起见,我们还假定矩阵 σ 是可逆的且 $b \geq r1$, 这里 $1 = (1, 1, \dots, 1) \in \mathbb{R}^n$ 。

在任何时刻 t , 投资者进行资产组合决策。设 $\pi_i(t)$ 表示时刻 t 投资在资产 i 上财富的份额。记 $\pi(t) = (\pi_1(t), \dots, \pi_n(t)) \in \mathbb{R}^n$, 则 $\pi_0(t) = 1 - \pi(t)1$ 。资产组合过程 $\pi(t)$ 称为一个资产组合投资策略。设时刻 t 的财富为 $W^r(t)$, 则时刻 t 持有资产 i 的份数为:

$$N_0(t) = W^r(t)(1 - \pi(t)1)/P_0(t), \quad i = 0$$

$$N_i(t) = W^r(t)\pi_i(t)/P_i(t), \quad i = 1, \dots, n$$

因此,我们有 $W^r(t) = \sum_{i=0}^n N_i(t)P_i(t)$ 。

我们不考虑交易费和税赋,且假设资产组合投资策略 $\pi(t)$ 是自融资的。于是有:

$$dW^r(t) = \sum_{i=0}^n N_i(t)dP_i(t)$$

$$\begin{aligned}
&= \left\{ rN_0(t)P_0(t) + \sum_{i=1}^n b_i N_i(t)P_i(t) \right\} dt + \sum_{i=1}^n N_i(t)P_i(t) \sum_{j=1}^n \sigma_{ij} dB_j(t) \\
&= W^r(t) \{ ((1 - \pi(t))r + \pi(t)b)dt + \pi(t)\sigma dB(t) \}
\end{aligned}$$

$$W^r(0) = \omega$$

这里, $\omega > 0$ 是投资者的初始财富。

下面我们限制在常数再调整证券组合范围内讨论问题。正如引言所述, 常数再调整证券组合投资策略每时每刻都要进行头寸的调整, 以使得投资在每个资产上的财富比例保持不变。因此常数再调整证券组合投资策略是一个不随时间变化的向量: $\pi(t) = \pi = (\pi_1, \dots, \pi_n)$, $t \in [0, T]$ 。应该注意, 这样一个投资策略仍然需要进行动态交易, 因为在每个时间点投资比例要再调整以回到向量 π 。

根据标准的 Ito 积分以及事实 $E[e^{sB_j(t)}] = e^{s^2/2}$ (这里 E 是数学期望算子), 我们得到财富过程 $W^r(t)$, $t \in [0, T]$ 的如下显式公式 (参见 Emmer, Kluppelberg & Korn, 2001):

$$W^r(t) = \omega \exp((\pi'(b - r1) + r - \|\pi'\sigma\|^2/2)t + \pi'\sigma B(t)) \quad (2.1)$$

$$E[W^r(t)] = \omega \exp((\pi'(b - r1) + r)t) \quad (2.2)$$

$$\text{Var}[W^r(t)] = \omega^2 \exp(2(\pi'(b - r1) + r)t) (\exp(\|\pi'\sigma\|^2 t) - 1) \quad (2.3)$$

这里, $\|\cdot\|$ 表示 R_n 空间中的欧氏模。

给定数 $\alpha \in (0, 1)$, 初始财富 ω 和投资计划期 T , 我们用 $\rho_0(\pi)$ 表示证券组合投资策略 π 对应的终端财富 $W^r(T)$ 的 α -分位数, 即:

$$P(W^r(T) \leq \rho_0(\pi)) = \alpha \quad (2.4)$$

这里, $P(\cdot)$ 表示概率。用 $\rho(\pi)$ 表示证券组合投资策略 π 的终端财富 $W^r(T)$ 发生在其 α -分位数 $\rho_0(\pi)$ 之下的平均值, 即:

$$\rho(\pi) = E[W^r(T) | W^r(T) \leq \rho_0(\pi)] \quad (2.5)$$

定义 2.1 证券组合投资策略 π 的在险收益 (Earnings at Risk, EaR) 是:

$$\text{EaR}(\pi, \omega, T) := E[W^r(T)] - \rho(\pi)$$

在不至于混淆时, 我们将 $\text{EaR}(\pi, \omega, T)$ 简单地记为 $\text{EaR}(\pi)$ 。

从定义我们可以发现, 在险收益实际上是用两个资产价值水平之间的差距来度量风险的, 即为: 证券组合投资策略的终端财富在其一定置信水平之下的平均值水平与其总平均值水平之间的距离。因此, 风险收益具有简洁的含义和直观的价值判断。这一明显优点使得资产组合的风险能够化为一个可以与财富相匹配的数字, 有利于经营管理目标的实现。更重要的是, 与方差不同, 风险收益侧重于对影响投资绩效的目标不足边的管理, 更接近于投资者对风险的真实心理感受。

令 z_α 为标准正态分布的 α -分位数。记 Φ 为标准正态随机变量的概率。

由于 $\pi'\sigma W(T)/[\|\pi'\sigma\|\sqrt{T}]$ 是一个标准正态随机变量, 根据公式 (2.1)、(2.4) 及 (2.5) 我们可以将 ρ_0 和 ρ 显式地表示为:

$$\rho_0(\pi) = \omega \exp((\pi'(b - r1) + r - \|\pi'\sigma\|^2/2)T + z_\alpha \|\pi'\sigma\| \sqrt{T}) \quad (2.6)$$

$$\rho(\pi) = \omega \exp((\pi'(b - r1) + r)T) \frac{\Phi(z_\alpha - \|\pi'\sigma\| \sqrt{T})}{\alpha} \quad (2.7)$$

于是可得到在险收益的如下显式表达式：

$$EaR(\pi) = \omega \exp((\pi'(b - r1) + r)T) \left[1 - \frac{\Phi(z_t - \frac{\|\pi'\sigma\| \sqrt{T}}{\alpha})}{\alpha} \right] \quad (2.8)$$

为避免本文结果需进行多种子情形的讨论，本文总是做如下假设：

假设 2.1 $\alpha < 0.5$ 从而 $z_t < 0$ 。

下面给出在险收益的极值性质，其证明见附录，

命题 2.1

$$(i) \sup_{\pi \in R^n} EaR(\pi) = \begin{cases} \omega e^{rT} & \text{若 } b = r1 \\ + & \text{否则} \end{cases}。$$

(ii) $\min_{\pi \in R^n} EaR(\pi) = 0$ 且最小值仅在纯债券策略下达到。

性质 (i) 表明，在一个风险中性的市场里，风险收益有上界且为纯债券策略产生的终端财富；而在一个风险非中性的市场里，风险收益无上界。性质 (ii) 表明，纯债券策略的在险收益为零，其他策略的在险收益均大于零，风险收益的这个行为与方差极为相像。

三、动态证券组合选择模型、最优策略与有效边界

Markowitz 的均值 - 方差模型是在给定证券组合投资策略的终端财富的方差水平的条件下最大化证券组合投资策略的期望终端财富，即求解如下优化问题：

$$(P) \quad \max_{\pi \in R^n} E[W^*(T)] \quad \text{s.t.} \quad \text{Var}[W^*(T)] \leq C$$

这里， $C > 0$ 是方差 $\text{Var}[W^*(T)]$ 的一个给定水平。

类似地，本文考虑这样的最优证券组合投资策略选择问题，它在给定证券组合投资策略的在险收益水平的条件下最大化证券组合投资策略的期望终端财富。更确切地说，我们求解如下优化问题：

$$(P) \quad \max_{\pi \in R^n} E[W^*(T)] \quad \text{s.t.} \quad EaR(\pi) \leq C$$

其中： C 是在险收益 $EaR(\pi)$ 的一个给定水平，投资者可根据其愿望事先设定。由于在险收益的全局最小值是零 [命题 2.1 (ii)]，这里我们设 $C > 0$ 。

均值 - 在险收益问题 (P) 的最优常数再调整证券组合投资策略可由如下定理来求得 (其证明见附录)：

定理 3.1 设 $b \neq r1$ 。问题 (P) 的惟一最优解是：

$$\pi^* = \varepsilon^* \frac{(\sigma\sigma')^{-1}(b - r1)}{\|\sigma^{-1}(b - r1)\|} \quad (3.1)$$

最优期望终端财富是：

$$E[W^*(T)] = \omega \exp((r + \varepsilon^* \|\sigma^{-1}(b - r1)\|)T) \quad (3.2)$$

其中： ε^* 是方程：

$$\varepsilon\theta T + rT + \ln\left(1 - \frac{\Phi(z_t - \varepsilon\sqrt{T})}{\alpha}\right) = \ln \frac{C}{\omega} \quad (3.3)$$

的惟一非负解, 这里 $\theta = \|\sigma^{-1}(b - r1)\|$ 。

为了理解为什么方程 (3.3) 有惟一解 (同时也为了使该定理的证明简化), 我们指出 (根据微积分容易证明) 方程 (3.3) 左端的函数:

$$g(\varepsilon) := \varepsilon\theta T + rT + \ln\left(1 - \frac{\Phi(z_\alpha - \varepsilon\sqrt{T})}{\alpha}\right) \quad (3.4)$$

在 $(0, +\infty)$ 上是严格单调递增的, 且:

$$\lim_{\varepsilon \rightarrow 0} g(\varepsilon) = -\infty, \quad \lim_{\varepsilon \rightarrow +\infty} g(\varepsilon) = +\infty$$

我们看到, 最优期望终端财富的值只是通过模 $\|\sigma^{-1}(b - r1)\|$ 依赖于股票, 而对股票的个数没有明显的依赖性。这样, 我们总是可以选取适当的市场参数值 b 和 σ 使得由多个股票组成的市场和仅由一个股票和一个债券组成的市场有相同的模 $\|\sigma^{-1}(b - r1)\|$, 从而在这两个市场的最优投资的效果就没有什么差异。因此, 上述定理可以解释成为一种共同基金定理。

定理 3.1 还告诉我们, 问题 (P) 的最优解使得其约束条件成为等式, 即 $EaR(\pi^*) = C$ 。因此, 最优投资策略 π 对应的期望终端财富 $\mu := E[W(T)]$ 和在险收益 $EaR := EaR(\pi)$ 满足方程组:

$$\begin{cases} \mu = \omega \exp((r + \varepsilon \|\sigma^{-1}(b - r1)\|)T) \\ \varepsilon\theta T + rT + \ln\left(1 - \frac{\Phi(z_\alpha - \varepsilon\sqrt{T})}{\alpha}\right) = \ln \frac{EaR}{\omega} \end{cases}$$

其中: $\varepsilon \geq 0$, 经化简得到:

$$\begin{cases} \varepsilon = \frac{\ln(\mu/\omega) - rT}{\theta T} \\ EaR = \omega \exp(rT + \varepsilon\theta T) \left(1 - \frac{\Phi(z_\alpha - \varepsilon\sqrt{T})}{\alpha}\right) \end{cases}$$

将第一个方程带入第二个方程得到 EaR 与 μ 的关系式:

$$EaR(\mu) = \mu \left[1 - \frac{1}{\alpha} \Phi\left(z_\alpha - \frac{\ln(\mu/\omega) - rT}{\|\sigma^{-1}(b - r1)\| \sqrt{T}}\right)\right] \quad (3.5)$$

这里, $\mu \geq \omega \exp(rT)$ 。这个函数在均值 - 在险收益平面上的图形就是均值 - 在险收益问题的有效边界。

例 1 设初始财富 $\omega = 1000$, 市场上有一种股票 (即 $n = 1$) 和一种无风险债券, 它们的回报率分别为 $b = 0.1$ 和 $r = 0.05$, 股票的标准差为 $\sigma = 0.2$, 令 $\alpha = 0.05$ 则 $z_\alpha = -1.64$, $\theta = 0.25$ 。取问题 (P) 事先设定的风险收益水平 C 为计划期 $T = 5$ 的纯股票策略 $\pi = 1$ 的风险收益 $EaR(1, 1000, 5)$ 。让计划期 T 在 $(0, 5]$ 上变化, 就得到不同的最优策略及相应的期望终端财富。这些期望终端财富随 T 的变化情况见图 1, 用实线表示。同时我们还画出了纯股票策略以及债券策略的期望终端财富作为 T 的函数的图形, 分别用虚线和粗实线表示。我们看到, 有风险控制的最优策略的期望终端财富甚至超过了无风险控制纯股票策略的期望终端财富。只要我们看看相应的证券组合就明白其中的道理了。在图 2 中我们画出了最优策略和纯股票上的投资份额作为计划期 $T(0 < T \leq 5)$ 的函数的图形。当 $T < 5$ 时, 最优策略总是含有债券的空头头寸。

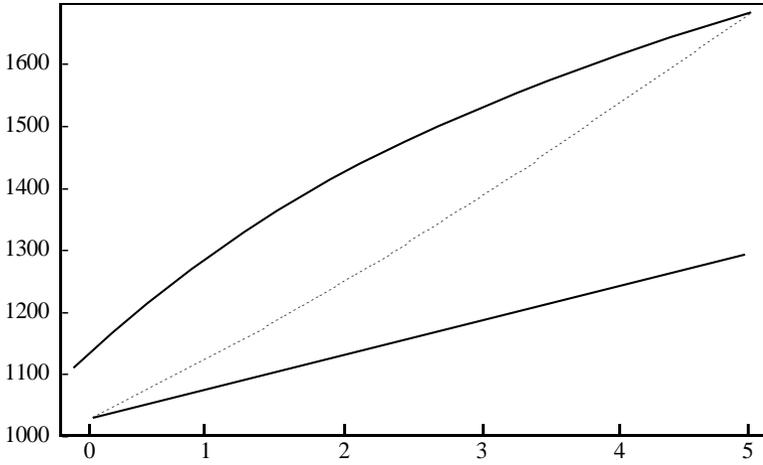


图 1 不同投资策略的期望终端财富作为计划期 T ($0 < T \leq 5$) 的函数的图形

为了比较，我们可类似地给出均值 - 方差问题 (P) 的最优解，参见 Emmer, Klüppelberg 和 Korn (2001)。

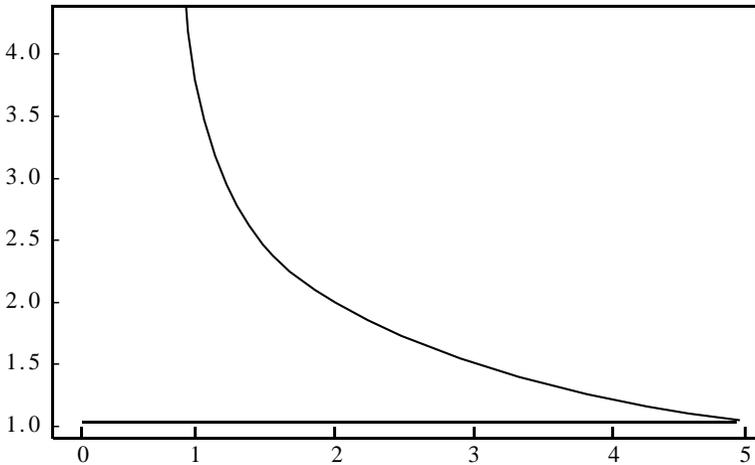


图 2 最优投资策略在股票上的投资份额作为计划期 T ($0 < T \leq 5$) 的函数的图形

定理 3.2 (P) 的最优解是：

$$\pi = \epsilon \frac{(\sigma\sigma')^{-1}(b - r1)}{\|\sigma^{-1}(b - r1)\|} \tag{3.6}$$

其中： ϵ 是如下方程的惟一正解：

$$\|\sigma^{-1}(b - r1)\| \epsilon T + rT - \frac{1}{2} \ln\left(\frac{C}{\omega h + 2(\exp(\epsilon^2 T) - 1)}\right) = 0 \tag{3.7}$$

相应的期望终端财富为：

$$E[W^*(T)] = \omega \exp((r + \varepsilon \|\sigma^{-1}(b - r1)\|)T) \quad (3.8)$$

可以验证, 问题(P)的最优解使得其约束条件成为等式, 即 $\text{VaR}(\pi^*) = C$ 。据此并利用(3.7)和(3.8)式, 我们可以得到均值 - 方差问题(P)的有效边界的显式表达式:

$$\text{Var}(\mu) = \mu^2 \left[\exp\left(\frac{(\ln(\mu/\omega) - rT)^2}{\|\sigma^{-1}(b - r1)\|^2 T}\right) - 1 \right], \quad \mu \geq \omega \exp(rT) \quad (3.9)$$

作为定理 3.1 和 3.2 的推论, 即使给定的风险收益水平, 方差水平 [以及在险资本水平, 见 Emmer, Klüppelberg & Korn (2001)] 相同, 均值 - 在险收益问题和均值 - 方差问题 (以及均值 - 在险资本问题) 的最优常数再调整证券组合投资策略也不同, 不过它们只相差一个常数倍数 (即它们作为向量是平行的)。这意味着, 只要这些不同风险度量的水平选得不同且适当, 可使这些不同的风险度量产生的最优常数再调整证券组合投资策略相同, 从而终端财富的期望收益相同。这进一步说明, 均值 - 在险收益问题和均值 - 方差问题 (以及均值 - 在险资本问题) 的有效边界是不同的。

如果投资者能够事先说明他希望的期望终端财富 μ , 则我们可首先通过方程(3.2)或(3.8)求出最优的 ε , 然后将这个 ε 代入 (3.1) 或 (3.6) 式, 就得到最优常数再调整证券组合投资策略。

最后, 我们再举一个数值例子来说明均值 - 在险收益有效边界同均值 - 方差有效边界的差异。

例 2 令市场参数同前例, 即 $n = 1, w = 1000, T = 5, r = 0.5, \alpha = 0.5, b = 0.1, \sigma = 0.2, z_\alpha = -1.64, \beta = 0.25$ 。将这些数据代入(3.5)和(3.9)式, 就可画出均值 - 在险收益问题和均值 - 方差问题的有效边界, 分别见图 3 和图 4。在这两幅图中, 横轴都表示均值, 纵轴分别表示在险收益和方差。图形清楚地表明, 当期望的终端财富增加时, 风险不管是用在险收益还是用方差度量均随之增加, 或者说高收益总是伴随着高风险, 这一点与我们对现实的观测是相符的。另一方面, 在均值 - 在险收益有效边界中, 财富的边际风险随着财富的增加而递减; 而在均值 - 方差有效边界中, 财富的边际风险随着财富的增加而递增, 这一点是两种有效边界最重要的不同了。

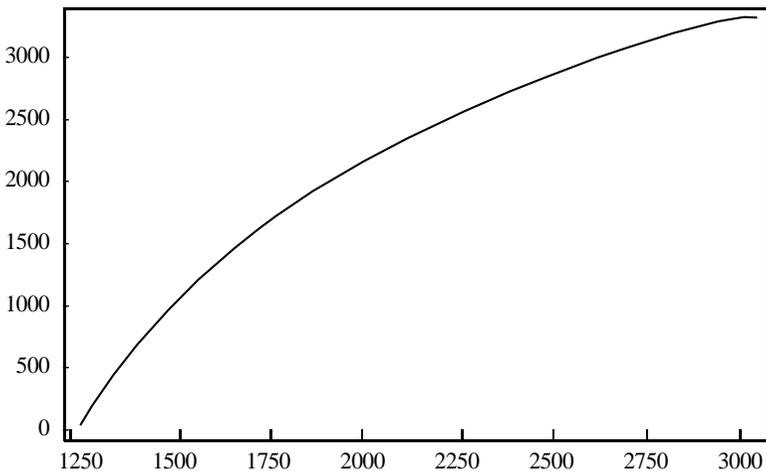


图 3 均值 - 在险收益问题的有效边界

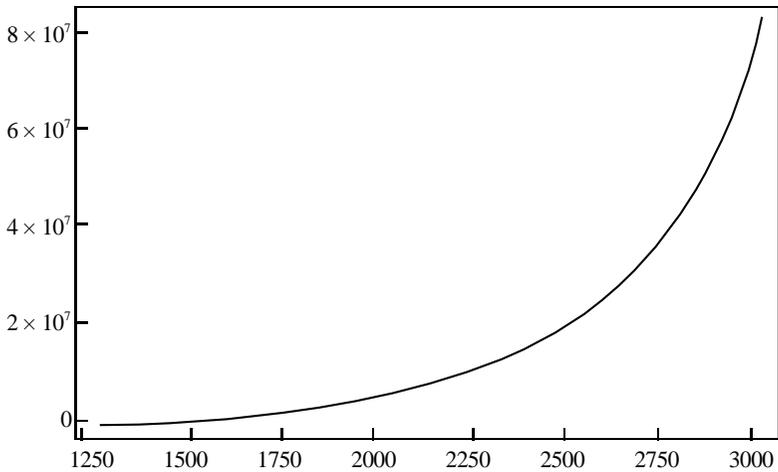


图 4 均值 - 方差问题的有效边界

四、结束语

本文引入了在险收益的概念，以它作为一种新的风险度量，来替代传统的风险度量 - 方差。类似于 Markowitz 的均值 - 方差模型，我们建立了均值 - 在险收益模型。在 Black-Scholes 设置下，我们给出了均值 - 在险收益意义下的最优常数再调整证券组合投资策略的分析表达式以及有效边界的显式表达式。还结合数值例子对我们的概念与结论做了经济学解释。不仅本文的结果可方便地应用于动态投资管理与决策的实践中，而且本文的思路与方法也可用于讨论一些类似的动态投资决策问题，如 Capital at Risk 型，Value at Risk 型和 Safety-First 型问题。

附 录

命题 2.1 的证明：如果 $b = r1$ 则结论显然成立。现在我们设 $b \neq r1$ 。

(i) 将表达式 (2.8) 重新写为：

$$EaR(\pi) = \begin{cases} we^{f(\pi)} & \text{若 } \|\pi' \sigma\| > 0 \\ 0 & \text{若 } \|\pi' \sigma\| = 0 \end{cases} \quad (1)$$

这里：

$$f(\pi) = (\pi'(b - r1) + r)T + \ln\left(1 - \frac{\Phi(z_\pi - \|\pi' \sigma\| \sqrt{T})}{\alpha}\right) \quad (2)$$

对任意给定的 $\epsilon > 0$ ，考虑如下优化问题：

$$\max_{\pi} f(\pi) \quad \text{s.t.} \quad \|\pi' \sigma\| = \epsilon \quad (3)$$

问题 (3) 的约束条件定义了一个椭球，在其边界上目标函数等于：

$$f(\pi) = (\pi'(b - r1) + r)T + \ln\left(1 - \frac{\Phi(z_\pi - \epsilon \sqrt{T})}{\alpha}\right)$$

因此，求解问题 (3) 等价于求解问题：

$$\max_{\pi} \pi'(b - r1) \quad \text{s.t.} \quad \pi'(\sigma\sigma')\pi = \epsilon^2 \quad (4)$$

利用 Lagrange 方法, 立即得到该问题的惟一最优解:

$$\pi_{\varepsilon}^* = \varepsilon \frac{(\sigma\sigma')^{-1}(b - r1)}{\|\sigma^{-1}(b - r1)\|} \quad (5)$$

代入计算得到:

$$f(\pi_{\varepsilon}^*) = \varepsilon\theta T + rT + \ln\left(1 - \frac{\Phi(z_{\varepsilon} - \varepsilon\sqrt{T})}{\alpha}\right)$$

这里, $\theta = \|\sigma^{-1}(b - r1)\|$, 显然有:

$$\lim_{\varepsilon \rightarrow +} f(\pi_{\varepsilon}^*) = +$$

这导致结论 (i)。

(ii) 根据 (2.8), 对任意 $\pi \neq 0$ 有 $\text{EaR}(\pi) > 0 = \text{EaR}(0)$ 。又由于矩阵 σ 被假设为可逆的, 即得结论。

定理 3.1 的证明, 对任意 $\varepsilon \geq 0$, 定义集合:

$$\begin{aligned} \Pi(\varepsilon) &= \{\pi : \|\pi'\sigma\| = \varepsilon, \text{EaR}(\pi) \leq C \\ &= \left\{ \pi : \|\pi'\sigma\| = \varepsilon, \text{we}^{(b-r1)T} \left[1 - \frac{\Phi(z_{\varepsilon} - \varepsilon\sqrt{T})}{\alpha} \right] \leq C \right\} \\ &= \left\{ \pi : \|\pi'\sigma\| = \varepsilon, (b-r1)\pi T \leq \ln \frac{C}{W} - rT - \ln\left(1 - \frac{\Phi(z_{\varepsilon} - \varepsilon\sqrt{T})}{\alpha}\right) \right\} \end{aligned}$$

这里我们规定, 当 $\varepsilon = 0$ 时 $\ln\left(1 - \frac{\Phi(z_{\varepsilon} - \varepsilon\sqrt{T})}{\alpha}\right) = -$ 。再定义集合:

$$\begin{aligned} \Xi &= \{\varepsilon : \varepsilon \geq 0, \Pi(\varepsilon) \neq \emptyset\} \\ &= \left\{ \varepsilon : \varepsilon \geq 0, \ln \frac{C}{W} - rT - \ln\left(1 - \frac{\Phi(z_{\varepsilon} - \varepsilon\sqrt{T})}{\alpha}\right) \geq -\varepsilon\theta T \right\} \\ &= \Xi_1 \cup \Xi_2 \end{aligned}$$

其中: $\theta = \|\sigma^{-1}(b - r1)\|$, 且:

$$\begin{aligned} \Xi_1 &= \left\{ \varepsilon : \varepsilon \geq 0, \ln \frac{C}{W} - rT - \ln\left(1 - \frac{\Phi(z_{\varepsilon} - \varepsilon\sqrt{T})}{\alpha}\right) \geq \varepsilon\theta T \right\} \\ \Xi_2 &= \left\{ \varepsilon : \varepsilon > 0, -\varepsilon\theta T \leq \ln \frac{C}{W} - rT - \ln\left(1 - \frac{\Phi(z_{\varepsilon} - \varepsilon\sqrt{T})}{\alpha}\right) < \varepsilon\theta T \right\} \end{aligned}$$

于是, 问题 (P) 的可行解集为 $\bigcup_{\varepsilon \in \Xi} \Pi(\varepsilon)$ 。因此, 问题 (P) 可等价地写成:

$$(P) \quad \max_{\varepsilon \in \Xi} \max_{\pi \in \Pi(\varepsilon)} E[W^r(T)]$$

注意到目标函数的表达式 (2.2), 该问题与如下问题同解:

$$(P') \quad \max_{\varepsilon \in \Xi} \max_{\pi \in \Pi(\varepsilon)} (b - r1)\pi T$$

首先, 对每个给定的 $\varepsilon \in \Xi$ 我们来求解问题 $\max_{\pi \in \Pi(\varepsilon)} (b - r1)\pi T$, 亦即:

$$(P_{\varepsilon}) \quad \begin{cases} \max (b - r1)\pi T \\ \text{s.t.} \quad \pi'(\sigma\sigma')\pi = \varepsilon^2 \\ (b - r1)\pi T \leq \ln \frac{C}{W} - rT - \ln\left(1 - \frac{\Phi(z_{\varepsilon} - \varepsilon\sqrt{T})}{\alpha}\right) \end{cases}$$

当 $\varepsilon \in \Xi_1$ 且 $\varepsilon > 0$ 时, 应用 Kuhn-Tucker 条件 (参见 Avriel, 1976), 得到的惟一最优解:

$$\pi_{\varepsilon}^* = \varepsilon \frac{(\sigma\sigma')^{-1}(b - r1)}{\|\sigma^{-1}(b - r1)\|} \quad (6)$$

从而 (P_{ε}) 的最优目标值是:

$$(b - r1)\pi_{\varepsilon}^* T = \varepsilon\theta T \quad (7)$$

当 $\varepsilon = 0 \in \Xi_1$ 时, 显然问题 (P_ε) 的最优解与最优目标值亦分别具有(6)和(7)的形式(因为 α 非奇异)。当 $\varepsilon \in \Xi_2$ 时, 超平面:

$$(b - r_1)\pi T = \ln \frac{C}{W} - rT - \ln \left(1 - \frac{\Phi(z_\varepsilon - \varepsilon\sqrt{T})}{\alpha} \right)$$

与椭圆 $\|\pi^* \sigma\| = \varepsilon$ 的每个交点 π_ε 都是 (P_ε) 的最优解并具有目标值

$$(b - r_1)\pi_\varepsilon T = \ln \frac{C}{W} - rT - \ln \left(1 - \frac{\Phi(z_\varepsilon - \varepsilon\sqrt{T})}{\alpha} \right) \quad (8)$$

综上, 我们已经求得问题的 (P_ε) 最优目标值:

$$h(\varepsilon) := \begin{cases} \varepsilon\theta T & \text{若 } \varepsilon \in \Xi_1 \\ \ln \frac{C}{W} - rT - \ln \left(1 - \frac{\Phi(z_\varepsilon - \varepsilon\sqrt{T})}{\alpha} \right) & \text{若 } \varepsilon \in \Xi_2 \end{cases} \quad (9)$$

这样, 问题 (P) 的最优解可通过求解如下问题来获得:

$$\max h(\varepsilon) \quad \text{s.t.} \quad \varepsilon \in \Xi \quad (10)$$

现在我们来求解问题 $\max_{\varepsilon \in \Xi_1} h(\varepsilon)$, 亦即:

$$(P_\varepsilon^1) \quad \begin{cases} \max \varepsilon\theta T \\ \text{s.t.} \quad \ln \frac{C}{W} - rT - \ln \left(1 - \frac{\Phi(z_\varepsilon - \varepsilon\sqrt{T})}{\alpha} \right) \geq \varepsilon\theta T \\ \varepsilon \geq 0 \end{cases}$$

利用式 (3.4) 定义的函数 $g(\varepsilon)$, 上面的问题可写成:

$$(P_\varepsilon^1) \quad \max \varepsilon\theta T \quad \text{s.t.} \quad g(\varepsilon) \leq \ln \frac{C}{W}, \quad \varepsilon \geq 0$$

由于 $g(\varepsilon)$ 在 $(0, +\infty)$ 上是严格递增的且值域为 $(-\infty, +\infty)$, 因此如上问题有惟一最优解 ε^* , 它是方程:

$$g(\varepsilon^*) = \ln \frac{C}{W} \quad [\text{即方程(3.3)}]$$

的解, 这样, 对所有的 $\varepsilon \in \Xi_1, \varepsilon \neq \varepsilon^*$, 有 $h(\varepsilon^*) = \varepsilon^* \theta T > \varepsilon \theta T = h(\varepsilon)$ 。现令 $\varepsilon \in \Xi_2$ 则由 Ξ_2 的定义, 一定有 $g(\varepsilon) > \ln \frac{C}{W} = g(\varepsilon^*)$ 。因此, 由 $g(\varepsilon)$ 的严格单调递增性, 有 $\varepsilon > \varepsilon^*$ 。定义函数:

$$\Psi(\varepsilon) := \ln \frac{C}{W} - rT - \ln \left(1 - \frac{\Phi(z_\varepsilon - \varepsilon\sqrt{T})}{\alpha} \right)$$

它在 $(0, +\infty)$ 上是严格单调递减的, 且:

$$\Psi(\varepsilon) := \ln \frac{C}{W} - \varepsilon\theta T - g(\varepsilon), \quad h(\varepsilon) := \begin{cases} \varepsilon\theta T & \text{若 } \varepsilon \in \Xi_1 \\ \Psi(\varepsilon) & \text{若 } \varepsilon \in \Xi_2 \end{cases}$$

于是 $\Psi(\varepsilon) < \Psi(\varepsilon^*) = \ln \frac{C}{W} - \varepsilon^* \theta T - g(\varepsilon^*) = \varepsilon^* \theta T$ 即 $h(\varepsilon) < \varepsilon^* \theta T = h(\varepsilon^*)$ 。因此, ε^* 也是 $h(\varepsilon)$ 在整个 Ξ 上的惟一最大值点, 进而 π^* 是 (P) 的惟一最优解。将 π^* 代入 (2.2) 就可得到 (3.2)。

参考文献

- [1] Avriel M., *Nonlinear Programming: Analysis and Methods*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1976.
- [2] Basak S., Shapiro A., *Value-at-Risk-based risk management: Optimal policies and asset prices*, *The Review of Financial Studies*, 2001, Vol. 14, pp371 ~ 29.
- [3] Chen A., Jen C., Zions S., *The optimal portfolio revision policy*, *Journal of Business*, 1971, Vol. 44, pp 51 ~ 61.

-
- [4] Cover, T. M. , Universal portfolios , *Mathematical Finance* , 1991 . , Vol.1 , pp 1 ~ 29.
- [5] Emmer S. , Kluppelberg C. , Korn R. , Optimal rtfolios with bounded capital-at-risk , *Mathematical Finance* , 2001 , Vol.11 , pp365 ~ 384.
- [6] Helmbold D. P. , Schapire R. E. , Singer Y. , Warmuth M. K. , On-Line Portfolio Selection Using Multiplicative Updates , *Mathematical Finance* , 1998 , Vol.8 , pp325 ~ 347.
- [7] Li D. , Ng W. L. , Optimal dynamic portfolio selection : multiperiod mean-variance formulation , *Mathematical Fiance* , 2000 , Vol.10 , pp247 ~ 256.
- [8] Markowitz H. , Portfolio selection , *The Journal of Finance* , 1952 , Vol.7 , pp77 ~ 91.
- [9] Merton R. C. , Lifetime portfolio selection under uncertainty : the continuous-time model , *Review of Economics and Statistics* , 1969 , Vol.51 , pp239 ~ 246.
- [10] Samuelson P. A. , Lifetime portfolio selection by dynamic stochastic proqramming , *The Review of Economics and Statistics* , 1969 , Vol.51 , pp239 ~ 246.
- [11] Zhou X. Y. , Li D. , Continuous-time mean-variance portfolio selection : A stochastic LQ framework , *Applied Mathematics and Optimization* , 2000 , Vol.42 , pp19 ~ 33.

保险精算中单损失和双损失环境 的一些重要理论再辨析*

杨智元 王艺明 陈浪南

生存模型的估计是精算理论的重要组成部分，广泛应用于包括人寿保险、企业养老金计划、国民社会保障计划、医学研究、国民医疗保障计划在内的各个领域。从 20 世纪 80 年代中国开始精算教育以来，我国的精算教育和精算职业已经得到很大的发展，但我们也不应该回避这样一个现实：与西方发达国家相比，我们在精算方面的研究仍存在很大差距。《数量经济技术经济研究》2002 年第 9 期刊登张涤新的文章《保险精算中单损失和双损失环境的一些重要理论辨析》（以下简称“张文”），拜读之后，深有感触。应该肯定的是，敢于对世界权威的保险精算著作提出质疑和挑战，而不是盲目崇拜权威，这对于引进的西方理论有更加精确的理解、更加清楚的认识，确实有着积极的作用。

通读张文之后，我们特意找来《生存模型》及一些相关的书籍，就相关的章节细读数遍。经过深入研究后，却总觉得有骨鲠在喉、不吐不快之感。因此特意写下如下文字，抛砖引玉。谬误之处，恳请各位同行批评指正。

一、有关两个“误区”的不同看法

（一）“误区 1”是否是误区？误在何处？

张文引用了译著中“如果随机事件死亡和撤出是独立的，则 ${}_t p_x^{\bar{}} = \exp(-\int_0^t \mu_{x+s}^{(\tau)} ds)$ $= {}_t p_x^{(d)} {}_t p_x^{(\omega)}$ ”。但接在这段文字的后面是这样一句张文中未引用的话“读者应当了解这个独立性在基本多重损失理论中并不总是假定成立的（见 Benjamin and Pollard [9] 或 Bowers, et al. [10]）”。张文为何忽略了这句话，我们不得而知。不过，我们可以看到，如果按照张文的意思，则 London 是在自己打自己的嘴巴，就是说 London 认为“独立性”是公式（4）（张文中的公式编号，下同）成立的必要条件，但又说 Benjamin 等人和 Bowers 等人可以不用这一条件也能推导出公式（4）。也就是说“独立性”是公式（4）成立的必要条件又不是必要条件。因为无法获得英文版的原著，我们无法做进一步的判断。但我们认为，作为一本曾经被北美精算师学会为准会员资格考试所指定的教材和参考书^②，应该不

* 原载：《数量经济技术经济研究》，2004 年第 3 期。

② 由于北美精算师考试体系 2000 年起有重大变革，《生存模型》一书目前并非指定的教材和参考书。

至于犯这么初等的错误。

应该指出，公式 (4) 成立与否确实不依赖于死亡和撤出两随机事件是否独立。那么，London 为何在书中提到独立性假设呢？我们再看看后面 London 对生存模型所作的估计。事实上，假如以 D_i 表示第 i 个人在区间 $(x, x + 1)$ 死亡的随机变量，以 W_i 表示第 i 个人在区间 $(x, x + 1)$ 撤出的随机变量。在双损失环境中，假如死亡 ($J = d$) 和撤出 ($J = w$) 两事件不独立，死亡可能会影响撤出，撤出也可能会影响死亡，则 D_i 和 W_i 的分布将不是下述形式：

$$D_i = \begin{cases} 1 & \text{概率为 } s_{i-r_i} q_{x+r_i}^{(d)} \\ 0 & \text{概率为 } 1 - s_{i-r_i} q_{x+r_i}^{(d)} \end{cases} \quad W_i = \begin{cases} 1 & \text{概率为 } s_{i-r_i} q_{x+r_i}^{(w)} \\ 0 & \text{概率为 } 1 - s_{i-r_i} q_{x+r_i}^{(w)} \end{cases} \quad (1)$$

而这样一来，双损失环境下的基本矩关系：

$$E[D_x] = \sum_{i=1}^n s_{i-r_i} q_{x+r_i}^{(d)} = d_x \quad E[W_x] = \sum_{i=1}^n s_{i-r_i} q_{x+r_i}^{(w)} = w_x$$

将不会成立（这里， D_x 表示所有 n_x 个人在 $(x, x + 1]$ 中死亡的随机变量， d_x 是观察到的死亡数， W_x 表示所有 n_x 个人在 $(x, x + 1]$ 中撤出的随机变量， w_x 是观察到的撤出数，见 London 的《生存模型》第 120，121 页）。

但是，死亡 ($J = d$) 和撤出 ($J = w$) 两事件独立则可以保证 D_i 和 W_i 的分布如上述的形式 (1)。这样，基本矩关系自然成立，London 才有可能进行后面的矩估计。

此外，在双损失环境下的极大似然估计中，独立性的假设将使得 q_x 的估计比较简单。这一点可参见 London 的《生存模型》第 163 页，“如果独立性假设不存在，那么就假设了一个相关模型，这样 q_x 的估计就很复杂了。关于这方面的讨论，有兴趣的读者可参阅 Robinson 著作 [52]”。可见独立性的假设是与生存模型的估计联系在一起的。

这里还应该提到的是，由于 Bowers 等人的《精算数学》并没有牵涉到双损失环境下生存模型的估计，因此他们没有要求死亡 ($J = d$) 和撤出 ($J = w$) 两事件独立也是顺理成章的。

当然，我们这里也应该指出，London 这种跳跃幅度极大的写作风格，可能更适合于写作专业期刊上的学术论文。作为一本让读者了解掌握精算学基本原理的书，前因后果没有交待清楚的写作方式，确实是不太合适的，这应该可以算是一个小“误区”吧！

(二) “误区 2” 是否完全错误？错在哪里？

张文引用了译著中这样一段话：“我们用 ${}_t p_x^{(d)}$ 表示的恰恰正是以前用 ${}_t p_x$ 表示的同一概念。我们只是简单地加上了一个记号以强调‘对应死亡的生存’的想法。类似地，我们现在用 $q_x^{(d)}$ 来表示 $(x, x + 1]$ 中死亡的概率，注意 $q_x^{(d)}$ 表示的恰恰是本书以前用较简单的 q_x 表示的同一概念。”然后“论证”了“London 的理解和解释是完全错误的”。

我们认为，这一“误区”事实上牵涉到究竟是在双损失环境或在单损失环境下讨论 ${}_t p_x^{(d)}$ (${}_t p_x$) 的含义这一问题。在双损失环境下，由于分析问题的框架完全不同于单损失环境，这两个符号对应的肯定不是同一概念，而不是张文中的“可能不是同一概念”，更无需进行所谓“严格的证明”。细心的读者可以在译著中发现，London 注明“该理论更为详细的内容，请读者查阅 Jordan [32] 第 14 章或 Bowers, et al. [10] 第 9 章”（见《生存模型》第 103 页）。

而在 Bowers, et al. 的《精算数学》第 251 页中, Bowers 等人明确指出: “符号 ${}_tq_x^{(j)}$ 表示在年龄 x 与 $x + 1$ 之间由原因 j 引起的损失概率, 它与 ${}_tq_x^{(j)}$ 不同。而 ${}_tp_x^{(j)}$ 与 ${}_tp_x^{(j)} - x$ 不同, 也不一定是一个生存函数, 因为并不要求 $\lim_{t \rightarrow \infty} {}_tp_x^{(j)} = 0$ 。”这样一来, 结合张文的意思, 如果是在双损失环境下讨论这个问题, London 又是自己在间接地批驳自己: 一方面认为 ${}_tp_x^{(d)}$ 与 ${}_tp_x$ 表示同一概念, 另一方面又认为 ${}_tp_x^{(d)}$ 与 ${}_tp_x$ 表示不同的概念, 因为 ${}_tp_x$ 是一个生存函数, 而 ${}_tp_x^{(j)}$ 不一定是一个生存函数。

不过, 为了不至于断章取义, 这里有必要提一提的是, 在译著中, 紧挨着张文引用的“我们 ${}_tp_x^{(d)}$ 用表示的恰恰正是以前用 ${}_tp_x$ 表示的同一概念……”这段话的前面, 还有一段张文没有引用的话“此处的每个概率在单损失环境中都有它们各自的意义”。我们看看在单损失环境中这两个符号表示的究竟是什么意思。按 London 书中第 104 页的关系式:

$${}_tp_x^{(d)} = \exp\left(-\int_0^t \mu_{x+u}^{(d)} du\right)$$

在单损失环境中, 由于只考虑死亡这一损因, $\mu_{x+u}^{(d)}$ 事实上就是 μ_{x+u} , 而 ${}_tp_x$ 满足关系式 ${}_tp_x = \exp\left(-\int_0^t \mu_{x+u} du\right)$ 是不容争议的。在单损失环境中, ${}_tp_x^{(d)}$ 与 ${}_tp_x$ 表示同一概念应该没有什么争议。

此外, London 还指出“以强调‘对应于死亡的生存’”, 这是否意味着只考虑死亡这一单损失因素, 是否是在单损失环境下探讨两个符号的同一性, 仁者见仁, 智者见智, 相信读者会自己加以判断。

二、双损失环境中一个重要公式辨析

公式 ${}_tp_x^{\tau} = \exp\left(-\int_0^t \mu_{x+s}^{(\tau)} ds\right) = {}_tp_x^{(d)} {}_tp_x^{(w)}$ 是双损失环境中的一个重要公式。这一公式可以从 $\mu_{x+t}^{(\tau)} = \mu_{x+t}^{(d)} + \mu_{x+t}^{(w)}$ 直接得到。那么 ${}_t\mu_{x+t}^{(\tau)} = \mu_{x+t}^{(d)} + \mu_{x+t}^{(w)}$ 是否无条件成立呢? 事实上, 张文中也注意到了“……被保险人死亡的同时不可能作出撤出缴付保险费, 死亡随机事件和撤出随机事件互斥”。由于两个损因互斥, ${}_t\mu_{x+t}^{(\tau)} = \mu_{x+t}^{(d)} + \mu_{x+t}^{(w)}$ 自然成立, 无需任何其他条件。

狭义意义上讲, 双损失环境即死亡和撤出。在这样的双损失环境下, 两个损因是互斥的, 公式 ${}_tp_x^{\tau} = \exp\left(-\int_0^t \mu_{x+s}^{(\tau)} ds\right) = {}_tp_x^{(d)} {}_tp_x^{(w)}$ 的成立是情理之中的。但从广义意义上说, 多重损失并不仅限于死亡和撤出。正如张文中所说的“在研究人的寿命终止时间与各种交通事故的关系时, ……分别表示乘坐飞机失事、乘坐海轮失事或其他以外交通事故。上述例子均涉及保险精算中的多重损失环境”。确实, 在精算研究中, 有时需要讨论同一种事故引起的不同种类伤害的概率, 如乘坐火车失事时引起眼睛失明、引起耳聋、引起四肢残废等等事件发生的概率。在公共健康计划的应用中, 可能有许多并非互不相容的损失(死亡)原因, 如心血管疾病引起死亡、癌症引起死亡、消化道疾病引起死亡、意外事故引起死亡等。在这些情形下, 由于这些事件并非互不相容事件, 被保险人可能在乘坐火车失事引起眼睛失明的同时引起耳聋, 并且引起四肢残废。在一个事件发生时, 另外的事件可能

同时发生。在这种情形下，公式 $\mu_{x+t}^{(\tau)} = \mu_{x+t}^{(d)} + \mu_{x+t}^{(w)}$ 就需要一些相应的限制条件才能成立。这种多重损失环境将引致一些不同的结论，公式 ${}_t p_x^{\tau} = \exp(-\int_0^t \mu_{x+s}^{(\tau)} ds) = {}_t p_x^{(d)} {}_t p_x^{(w)}$ 并非无条件成立。以下我们就这种广义情形对上述公式进行推广。不过，应该特别指出的是，这里并不牵涉到 London 或者张涤新谁对谁错、是否存在错误的问题，我们这里讨论的框架是不相同的。

由于广义的双损因并非非此即彼的形式，我们需要对危险率的定义进行一点修正。我们称 $J \in A$ 为由于原因 A 引起的损失， $J \in B$ 为由于原因 B 引起的损失， $J \in (A \cap B)$ 为由于原因 A 和 B 共同引起的损失，记：

$$\mu_t^{(A)} = \lim_{\Delta t \rightarrow 0^+} \frac{P(t < T \leq t + \Delta t, J \in A | T > t)}{\Delta t}$$

$\mu_t^{(A)}$ 称为原因 A 对应的危险率。则我们有如下性质：

性质 1 在广义的双损因环境 A, B 中 $\mu_{x+t}^{(\tau)} = \mu_{x+t}^{(A)} + \mu_{x+t}^{(B)} - \mu_{x+t}^{(A \cap B)}$

证明 首先，由于 $P(t < T \leq t + \Delta t | T > t) = P(t < T \leq t + \Delta t, J \in A | T > t) + P(t < T \leq t + \Delta t, J \in B | T > t) - P(t < T \leq t + \Delta t, J \in A \cap B | T > t)$ ，因此，
$$\frac{P(t < T \leq t + \Delta t | T > t)}{\Delta t} = \frac{P(t < T \leq t + \Delta t, J \in A | T > t)}{\Delta t} + \frac{P(t < T \leq t + \Delta t, J \in B | T > t)}{\Delta t} - \frac{P(t < T \leq t + \Delta t, J \in A \cap B | T > t)}{\Delta t}$$
，所以，

$$\mu_{x+t}^{(\tau)} = \mu_{x+t}^{(A)} + \mu_{x+t}^{(B)} - \mu_{x+t}^{(A \cap B)}$$

由性质 1，我们可以看到，在广义的双损因环境下，相应的公式应该是：

$${}_t p_x^{\tau} = \exp(-\int_0^t \mu_{x+s}^{(\tau)} ds) = {}_t p_x^{(A)} {}_t p_x^{(B)} {}_t p_x^{(A \cap B)}$$

此外，我们有如下的性质 1'：

性质 1' 在广义的双损因环境 A, B 中，若损因 A 和 B 是互不相容的，则有：

$$\mu_{x+t}^{(\tau)} = \mu_{x+t}^{(A)} + \mu_{x+t}^{(B)}$$

由于损因 A 和 B 互不相容， $P(t < T \leq t + \Delta t, J \in A \cap B | T > t) = 0$ ，这一性质是显而易见的。因此，公式 ${}_t p_x^{\tau} = {}_t p_x^{(A)} + {}_t p_x^{(B)}$ 在这种情形下是成立的。

参考文献

[1] Lee, E. T.: 《生存数据分析的统计方法》，陈家鼎等译，中国统计出版社 1998 年版。
 [2] 鲍尔斯等 (Bowers, et al.): 《精算数学》，余跃年等译，上海科学技术出版社 1996 年版。
 [3] [美] 蒋庆琅: 《随机过程原理与生命科学模型》，方积乾译，上海翻译出版公司 1987 年版。
 [4] 李秀芳: 《寿险精算实务》，南开大学出版社 2000 年版。
 [5] 伦敦 (D. London): 《生存模型》，陈子毅译，上海科学技术出版社 1996 年版。
 [6] 张涤新: 《保险精算中单损失和双损失环境的一些重要理论辨析》，《数量经济技术经济研究》，2002 年第 9 期。
 [7] 周江雄等: 《生命表的构造理论》，南开大学出版社 2001 年版。

基于 VaR 的金融资产配置模型*

姚 京 李仲飞

一、引言

Value at Risk (VaR) 是 20 世纪 90 年代发展起来的风险管理方法,由于一些大的金融监管机构的推广应用,这种方法逐渐成为全行业衡量风险的标准。监管当局和金融机构对 VaR 技术的青睐,极大地激发了学术界的研究兴趣。除了不断地完善测量方法,一些学者开始讨论在 VaR 约束下的资产组合选择问题。资产组合选择问题一般可以追述到马可维茨 1952 年提出的均值 - 方差模型^[1],由于他首次提出一个系统的处理风险与收益的方法而被授予 1990 年的诺贝尔经济学奖。这之后许多学者对均值 - 方差模型提出了改进和批评,其中一个主要的途径是使用新的风险指标来代替方差来建立新的均值风险模型。发展新的风险指标以及树立更加一般化的均值 - 风险框架成为现代金融理论的重要内容。VaR 作为一种风行的风险指标,它对资产组合选择的理论和实践可能产生的影响不容忽视。与此类似的问题可以追述到 1952 年 Roy^[2] 出的安全 - 首要问题 (Safety-First)。Pyle 和 Turnovsky^[3] 使用几何方法分析了安全 - 首要问题的三种形式与均值 - 方差模型的联系。文献 [4] 分析了将 VaR 作为风险管理目标的单期模型。在国内,吴世农、陈斌^[5] 就马可维茨的“方差风险”度量方法、哈洛的 LPM 方法和 VaR 方法三者 in 资产选择的应用效率的高低进行了比较和实证分析,得出马可维茨的“方差风险”度量方法为最差,LPM 方法居中,VaR 方法最具效率的结论。李健^[6] 对上述文章进行了批评,指出“吴文在分析方法和分析结论均存在一些问题”,并认为三种指标并没有直接的可比性,不能简单地放在一个坐标平面上进行比较,但没有进一步对几种指标进行深入比较。

本文仿照 Pyle 和 Turnovsky 的框架,对均值 - 方差模型和均值 - VaR 模型进行较为详细的比较。首先对一般假设下的均值 - VaR 模型进行比较详细的分析,再进一步考察存在无风险资产、负债和脱离正态分布假设时模型的情况。

二、均值 - VaR 模型

(一) 均值 - VaR 模型的组合边界和有效边界

首先考虑市场上没有无风险资产的情况。考虑一个存在 $n(n \geq 2)$ 种风险资产的、资产

* 基金项目:国家自然科学基金项目(10171115)、高等学校全国优秀博士学位论文作者专项资金项目(200267)、教育部人文社会科学研究“十五”规划项目(01JAG30009)、广东省自然科学基金项目(011193)、广东省哲学社会科学规划项目(02M13)。原载:《中国管理科学》,2004 年第 1 期。

交易无摩擦的经济,并且假设该经济认同个体的卖空行为,资产的收益率服从多元正态分布。这里仍沿用均值-方差模型中所使用的两个基本假设:① n 种风险资产的随机收益的协方差矩阵(Σ)是正定矩阵;② n 种风险资产的随机收益是非共线性的。

设投资者拥有 1 单位财富,他所面临的决策是根据偏好选择所持有的头寸: $x = (x_1, x_2, \dots, x_n)$ 。设 r_x 为组合 x 的收益。给定置信水平 α ,则 VaR 可以定义为 $\text{Prob}(r_x \leq -\text{VaR}_\alpha) = 1 - \alpha$ 。将变量 r_x 标准化为 $(r_x - E[r_x]) \cdot \sigma[r_x]$,设标准化变量的累积分布函数为 F ,则有:

$$\text{Prob}(r_x \leq -\text{VaR}_\alpha) = F\left(\frac{-\text{VaR}_\alpha - E[r_x]}{\sigma[r_x]}\right)$$

由于 F 是单调函数, VaR_α 也可以表达为 $\text{VaR}_\alpha = -F^{-1}(1 - \alpha) \cdot \sigma[r_x] - E[r_x]$,该 VaR 值不仅依赖于 α ,而且依赖于 r_x ,我们记为 $\text{VaR}_\alpha[r_x]$ 。而在正态分布的假设下,有 $F^{-1}(\alpha) = -F^{-1}(1 - \alpha)$,上式还可以进一步简化为 $\text{VaR}_\alpha[r_x] = z_\alpha \sigma[r_x] - E[r_x]$,其中 z_α 为标准正态分布的 α 分位数,即 $F^{-1}(\alpha)$ 。可以看出 z_α 对应着风险厌恶系数,即对置信度的选择在一定程度上反映了投资者对风险的偏好程度。在这里风险相对于方差或标准差而言有了新的定义,它不但包含资产客观上的性质(如标准差)的信息,也是风险厌恶系数(主观上)的函数,有效地将投资者的偏好考虑在内。Baumol^[7]提出的风险测量指标 $E[r_x] - k\sigma[r_x]$ 即为 VaR 的相反数。此时均值-VaR 模型可以表示为:

$$\min_x \text{VaR}_\alpha[r_x] = z_\alpha \sigma[r_x] - E[r_x]$$

$$\text{s.t. } E[r_x] \geq r, \sum x_i = 1$$

其中: r 为给定的期望收益水平。很容易看出在给定置信度水平($\alpha > 50\%$)下, z_α 严格大于零。另外,可以证明不等式约束在最优解处是紧的。因此均值-VaR 模型的解与标准的均值-方差模型是一样的,具有相同的边界组合。这样很多在均值-方差框架下的结果可以直接借用。众所周知,均值-方差模型的组合边界在均值-标准差平面上的关系可以表示为双曲线:

$$\frac{\sigma^2[r_x]}{1/C} - \frac{(E[r_x] - A/C)^2}{D/C^2} = 1$$

其中: $A = 1' \sum^{-1} \mu$, $B = \mu' \sum^{-1} \mu$, $C = 1' \sum^{-1} 1$, $D = BC - A^2$; 1 为分量均为 1 的向量, μ 为 n 种风险资产的期望收益向量。

利用 VaR 和标准差之间的关系可以得出均值-VaR 模型的组合边界,它也可以表达为均值-VaR 平面中的一条双曲线:

$$\frac{(\text{VaR}_\alpha[r_x] + E[r_x])^2}{z_\alpha^2/C} - \frac{(E[r_x] - A/C)^2}{D/C^2} = 1$$

图 1 给出了在均值-VaR 中置信度水平分别为 95%, 99%, 99.9% 时的组合边界,它们都对应着标准均值-方差模型的组合边界。特别要指出的是对于一个对称的分布而言, $\alpha \rightarrow 1/2$ 时有 $z_\alpha \rightarrow 0$,从而由定义有 $\lim_{\alpha \rightarrow 0.5} \text{VaR}_\alpha[r_x] = -E[r_x]$,即此时均值-VaR 的组合边界退化成为一条过原点的直线,它的斜率为 -1。

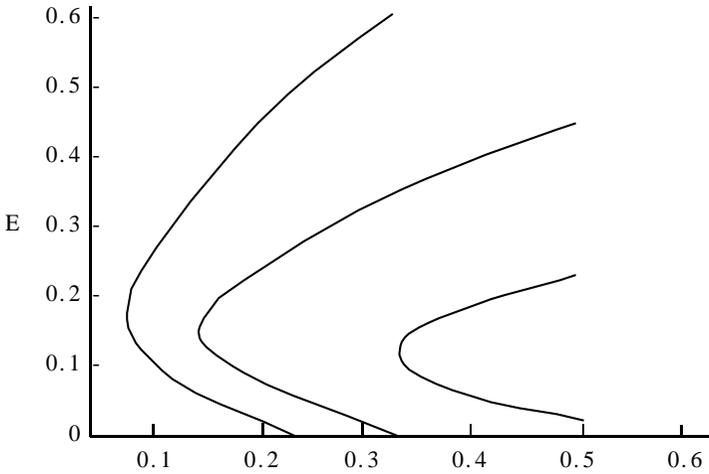


图1 三条组合边界从左到右分别对应着 95% , 99% , 99.9% 的置信度水平

与均值 - 方差模型一样, 在每条双曲线上存在着一个在所有组合中 VaR 最小的组合, 其坐标可以通过标准的微分性质确定。对组合边界方程两端求导并令 $dVaR_{\alpha}[r_x]dE[r_x] = 0$, 可以解得 $E[r_x] = \frac{A/D + VaR/\sigma^2}{C/D - 1/\sigma^2}$ 将此式代入组合边界方程中可解得最小 VaR 组合的均

值和 VaR 分别为: $E := \frac{D(-A + \sqrt{z_{\alpha}^2 C - D}) + z_{\alpha}^2 AC}{C(z_{\alpha}^2 C - D)}$, $V := \frac{-A + \sqrt{z_{\alpha}^2 C - D}}{C}$ 。这里有两点

值得注意: ① 最小 VaR 组合并没有落在双曲线的顶点上, 因为这条双曲线的对称轴并不是和坐标轴平行的, 同样地, 双曲线的渐近线斜率的绝对值也是不相等的(见图 2); ② 从上述表达式可以看出只有当 $N(\sqrt{D/C}) < \alpha$ (N 为标准正态分布的分布函数) 时候, 最小 VaR 组合才存在, 因此在均值 - VaR 的框架下最小 VaR 组合(文献 [4] 给出了另一种证明)也决定于置信度水平的选择。同样, 投资者不会选择期望收益率低于最小 VaR 组合收益率的组合, 因此双曲线的上半部分(在最小 VaR 组合以上的部分)即为有效边界。

为了能将 VaR 和方差两种风险指标在同一个平面中比较, 下面引入一个新的概念。由上述的定义可以看出 VaR 可以看作是给定置信度水平下均值 $E[r_x]$ 和标准差 $\sigma[r_x]$ 的函数, 则对某一水平的 VaR 值 V_c , 方程 $f(E[r_x], \sigma[r_x]) = V_c$ 确定一条无差异曲线(在这条线上所有的点的 VaR 都相等) 本文称为等 VaR 线。可以看出在均值标准差的平面中等 VaR 线是一条直线, 它的斜率为 $dE[r_x]/d\sigma[r_x] = z_{\alpha}$, 对于某一置信度水平来说是常数, 它在纵轴上的截距为 $-V_c$ 。图 3 给出了在均值 - 标准差平面中引入等 VaR 线时的情景, 给定置信度水平的同时, 等 VaR 线的斜率也同时确定, 等 VaR 线表现为一组平行线。由于 $\alpha > 50\%$, 这组平行线的斜率是正的, 因此在这组平行线当中可以找到一条(惟一的一条)均值 - 标准差组合边界的切线, 切点组合即为最小 VaR 组合(图中的 N 点), 截距即为组合 N 的 VaR 值相反数 $-VaR(N)$ 。可以看出只有标准差下的有效边界组合才可能是 VaR 有效边界组合, 在图 3 中标准差下的有效组合 A 落在 N 的上方, 标准差下的无效组合 B 在 N 的下方。

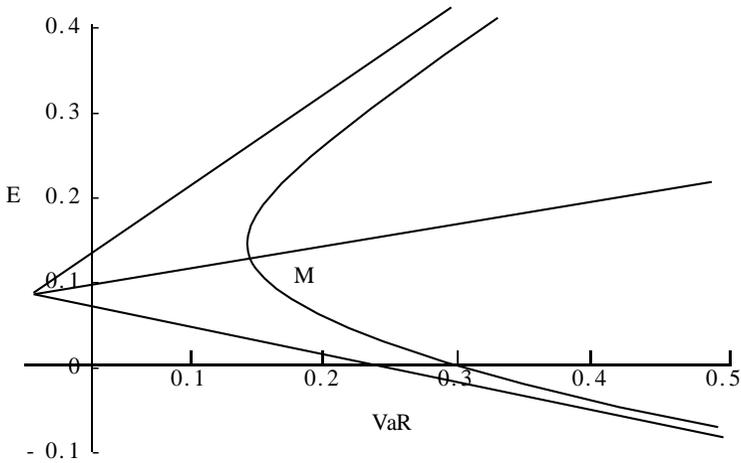


图 2 M 为最小方差组合

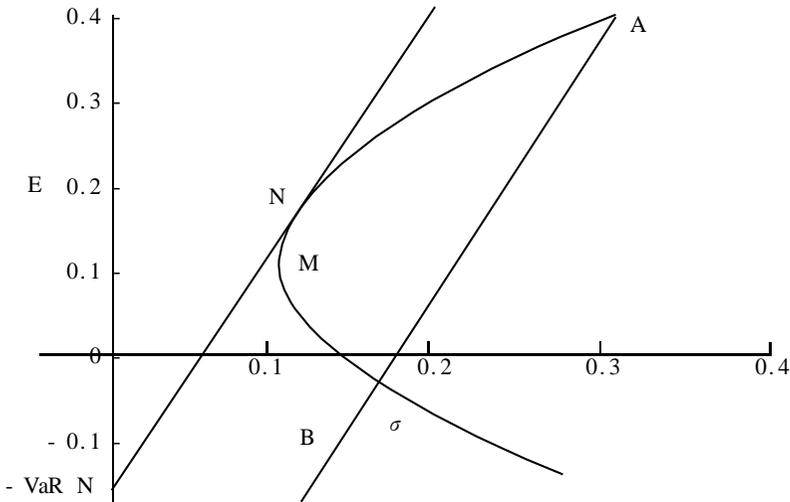


图 3 切点组合 N 对应着最小 VaR 组合

通过对图 3 的分析还可以得出以下几条性质：①当 $\alpha \leq N(\sqrt{D/C})$ 时，等 VaR 线的斜率小于等于渐近线的斜率，不存在最小 VaR 组合，这和上文中的结论是一样的。②若存在最小 VaR 组合，最小 VaR 组合将落在均值 - 标准差框架下的有效边界上，因此它同时是均值 - 标准差的有效组合；反过来，虽然最小方差组合也落在均值 - VaR 框架下的组合边界上，但从图 3 中可以看出它在最小 VaR 组合的下方，因此最小方差组合并不是 VaR 条件下的有效组合，VaR 有效边界是标准差有效边界的子集。③当 $\alpha \rightarrow 1$ 时，等 VaR 线的斜率趋于无穷大，最小 VaR 组合趋近于最小方差组合，同时 VaR 有效边界收敛于标准差有效边界。

Basak 和 Shapiro^[8]的分析结果指出基于 VaR 的风险管理可能导致资产面临更大的风险，在股票市场不景气的时候使股票波动率变得更大，这样金融机构可能承担更大的损失，并

以此解释 1998 年股市不景气时,一些银行、对冲基金发生的巨大损失。本文的结论也可得到类似的结果。当期望收益非常小的时候,或者在经济处于萧条阶段,所有的投资者都表现为相对大的风险规避行为,此时的所选择的有效组合接近于各自标准下的最小风险组合。但是以 VaR 为风险管理目标同时也意味着即使选择最小 VaR 组合,它的标准差与最小方差组合尚有一些距离,因此此时在 VaR 框架下所选择的组合表现为较大的波动性。由上文的分析还可以发现置信度 α 选择的越小, VaR 的组合边界与标准差的组合边界的差别越大,而且差别主要体现在最小方差组合和最小 VaR 组合之间的一段距离。这样如果在两种模型之间进行转换,可能出现所选择的组合出现巨大的变动。

(二) 引入无风险资产

这一节引入无风险资产,投资者可以以无风险利率 r_f 对其进行借贷。根据上节一样的程序,均值 - 方差组合边界为:

$$\sigma^2[r_x] = (E[r_x] - r_f)^2 / H$$

其中: $H = Cr_f^2 - 2Ar_f + B$,由此我们可以得出均值 - VaR 框架下的组合边界的方程:

$$\left(\frac{\text{VaR}[\alpha, r_x] + E[r_x]}{z_\alpha} \right)^2 = \frac{(E[r_x] - r_f)^2}{H}$$

上式同样可以表示成两条射线:

$$\text{VaR}_\alpha[r_x] = \begin{cases} \frac{E[r_x](z_\alpha - \sqrt{H}) - z_\alpha r_f}{\sqrt{H}} & \text{当 } E[r_x] \geq r_f \\ \frac{E[r_x](z_\alpha + \sqrt{H}) + z_\alpha r_f}{\sqrt{H}} & \text{当 } E[r_x] < r_f \end{cases}$$

但要注意的是和上节的分析一样,这两条射线的斜率的绝对值同样是不等的。与均值 - 方差下的组合边界不同,引入无风险资产后, VaR 条件下的组合边界的端点落在第二象限内。从 VaR 的定义可以发现:当完全投资于无风险资产时 $\text{VaR}_\alpha = -r_f$ 为负值,因此在第二象限内,坐标为 $(-r_f, r_f)$ 。此外,在 VaR 条件下,引入无风险资产前后的组合边界的关系仍然存在,例如,在 $r_f \leq E[r_{\text{mp}}]$ 的时候,组合的两个边界相切(见图 4)。

此处只讨论 $r_f \leq E[x_{\text{mp}}]$ 的情况,其他结论可通过类似的分析得出。从坐标 $(-r_f, r_f)$ 可以引一条与组合边界相切的直线,可以解得切点组合的坐标为: $E_m = \frac{Ar_f - B}{-A + r_f C}$, $\text{VaR}_m = \frac{Ar_f + z_\alpha \sqrt{Cr_f^2 - 2Ar_f + B} - B}{-A + r_f C}$,切线的斜率为: $\frac{E_m - r_f}{\text{VaR}_m + r_f} = -\frac{H}{H - z_\alpha \sqrt{H}}$,和有效边界的斜率相等,即切线与有效边界是重合的,所得的切点组合可视为 VaR 条件下的市场组合,这里

要特别引起注意的是: $E_m = \frac{Ar_f - B}{-A + r_f C} = \frac{A}{C} - \frac{D}{C^2(r_f - A/C)}$,即两种风险指标在各自的条件下所导出的市场组合是一样的。这表明当投资者都使用均值 - VaR 模型进行资产选择时 CAPM 仍然成立,这个结论可以用于拓宽 CAPM 的结论。进一步,当一部分投资者使用马可维茨的均值 - 方差模型,另一部分投资者使用均值 - VaR 模型进行资产选择时,切点组合仍然不变, CAPM 也还是成立的。此时,引入无风险资产后的有效边界也就成了 VaR 条

件下的资本市场线 (CML), 可计算其公式为 :

$$\frac{E[r_x] - E_m}{\text{VaR}[r_x] - \text{VaR}_m} = \frac{E_m - r_f}{\text{VaR}_m + r_f}$$

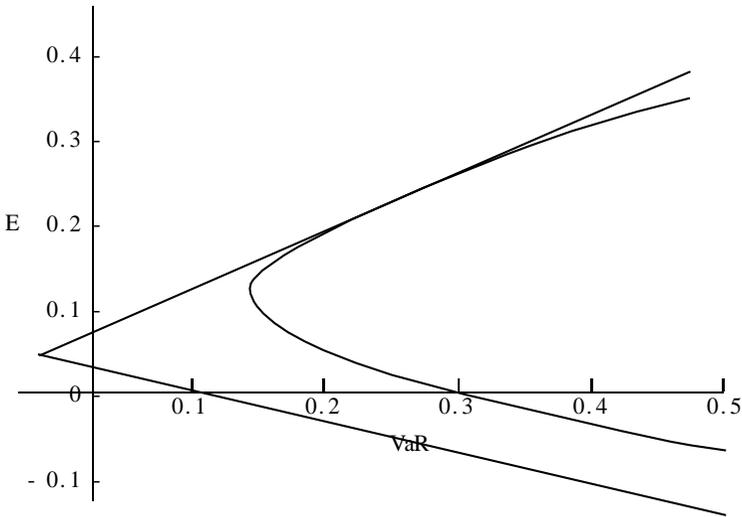


图 4 有效边界和组合边界存在惟一切点

引入无风险资产后同样可以使用等 VaR 线 (见图 5) 得出一些有用的结论。与前面的归纳类似, 我们有: ①当 $\alpha \leq N(\sqrt{H})$ 时, 等 VaR 线的斜率小于等于标准差有效边界的斜率, 不存在最小 VaR 组合; ②若存在最小 VaR 组合, 最小 VaR 组合将落在均值 - 标准差有效边界上, 在图中表现为等 VaR 线过标准差有效边界的顶点, 因此它同时是最小方差组合, 此时最小方差组合同时也是 VaR 下的有效组合; ③同②中的分析, 若存在最小 VaR 组合, VaR 有效边界上的组合也是标准差有效边界上的组合, 反过来同样成立。

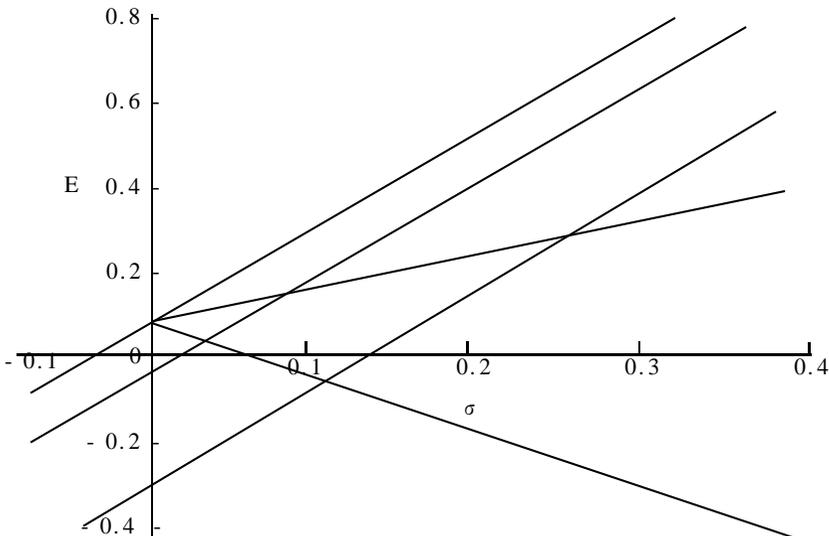


图 5 等 VaR 线与存在无风险资产时的组合边界

三、引入负债

这里我们将模型进一步扩展到存在负债的情形。假设负债的初值为 L_0 ，期末价值为随机变量 L_1 ，其收益率为 $R_L = (L_1 - L_0)/L_0$ 。风险资产组合的收益为随机变量 R_x （同上文中的 r_x ）。设资产的初值为 A_0 ，则此时期初的净资产为 $S_0 = A_0 - L_0$ ，期末净资产为 $S_1 = A_0(1 + R_x) - L_0(1 + R_L)$ ，正则化后的收益为 $R_s = (S_1 - S_0)/A_0 = R_x - R_L/f_0$ ，其中 $f_0 = A_0/L_0$ ，为初期的负债资产比例。则在存在负债的情况下，马可维茨模型为：

$$\begin{aligned} & \min_x \sigma^2 [R_x - R_L/f_0] \\ & \text{s.t. } E[R_x - R_L/f_0] \geq r, \sum x_i = 1 \end{aligned}$$

Keel 和 Muller^[9]得出的有效组合之间的关系式为：

$$\begin{aligned} E[R_x - R_L/f_0] &= E[R_{x^{\min}} - R_L/f_0] + \lambda E[R_z] \\ \sigma^2 [R_x - R_L/f_0] &= \sigma^2 [R_{x^{\min}} - R_L/f_0] + \lambda^2 \sigma^2 [R_z] \end{aligned}$$

其中： $x^{\min} = \frac{\sum^{-1} 1}{1' \sum^{-1} 1} + \left[\sum^{-1} \gamma - \frac{1' \sum^{-1} \gamma \sum^{-1} 1}{1' \sum^{-1} 1} \right]$ ， $\gamma = (\gamma_1, \dots, \gamma_N)$ 为资产与负债的协

方差向量，即 $\gamma_i = \text{Cov}(R_i, R_L)/f_0$ ， $z = \sum^{-1} \mu - \frac{1' \sum^{-1} \mu}{1' \sum^{-1} 1} \sum^{-1} 1$ 。

这样我们可以得出，在前述的基本假设之下，净资产的收益（ $E[R_x - R_L/f_0]$ ）和标准差（ $\sigma[R_x - R_L/f_0]$ ）之间仍然存在双曲线的关系：

$$\sigma^2 [R_x - R_L/f_0] = \sigma_{\min}^2 + \frac{C}{D} (E[R_x - R_L/f_0] - \mu_{\min})^2$$

其中： $\mu_{\min} = A/C + G - AF/C - E[R_L/f_0]$ ； $\sigma_{\min}^2 = (F - 1)^2/C - M + \sigma^2(R_L/f_0)$ ； A, B, C, D 与前同； $M = \gamma \sum^{-1} \gamma$ ， $F = \gamma \sum^{-1} 1$ ， $G = \gamma \sum^{-1} \mu$ 。很容易发现净资产组合边界的渐近线与马氏组合边界的渐近线的斜率相同，相互平行。一个很明显的结论是：净资产的风险不能由资产和负债的简单加减所得，这在风险管理的实践中有重要的指导意义。

同样将目标函数变换为 $\text{VaR}_\alpha [R_x - R_L/f_0] = z_\alpha \sigma [R_x - R_L/f_0] - E[R_x - R_L/f_0]$ 后所得问题和上述问题是同解的，在 $(E[R_x - R_L/f_0], \text{VaR}_\alpha [R_x - R_L/f_0])$ 平面上仍然为双曲线：

$$\left(\frac{\text{VaR}_\alpha [R_x - R_L/f_0] + E[R_x - R_L/f_0]}{z} \right)^2 = \sigma_{\min}^2 + \frac{C}{D} (E[R_x - R_L/f_0] - \mu_{\min})^2$$

图 6 和图 7 分别给出了在标准差和 VaR 两种标准下加入负债带来的组合边界的变动。在两种标准下，加入负债后的组合边界都出现了不同程度的右移。表明在同样的预期收益下，资产组合的风险变大了。但具体组合边界的移动情况还需要视 $(\mu_{\min}, \sigma_{\min})$ 的具体取值而定。

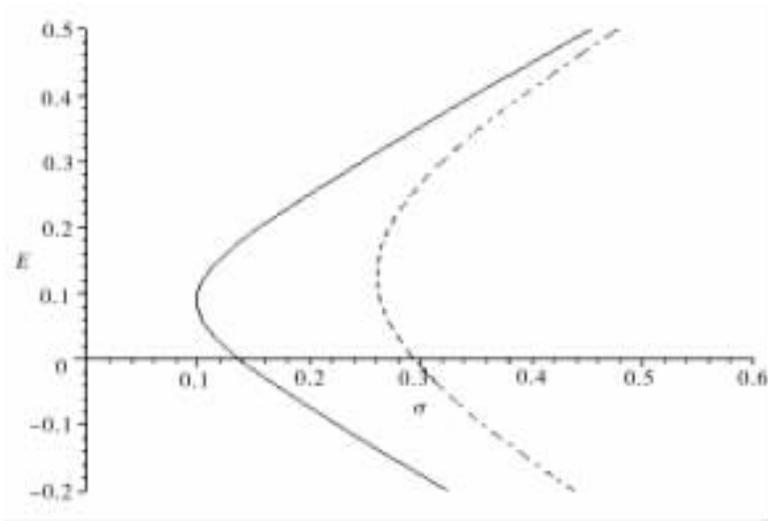


图 6 虚线为存在负债情形，实线为马可维茨的组合边界

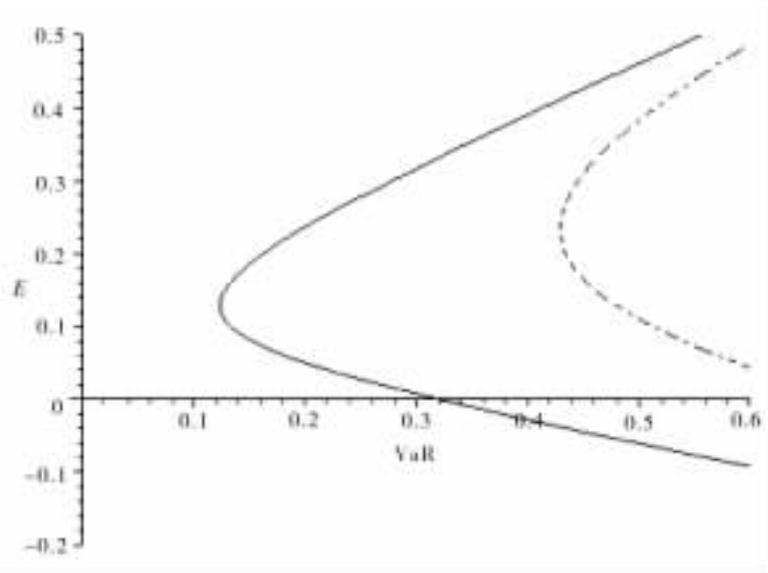


图 7 虚线为存在负债情形，实线为不存在负债时的均值 - VaR 组合边界

四、非正态分布情形

前文基本上是在正态分布的基础上进行论述的，但从本文所使用的定义中可以看出本文的方法可以推广到一些由两个参数决定的分布族，其中比较有代表性的有椭球等高分布族。椭球等高分布族中包含了许多常用的分布，正态分布、 t 分布、混合正态分布等等都在这个分布族内。假设资产收益服从某种椭球等高分布，其均值向量为 μ ，散度（dispersion）矩阵 Q 是一个正定矩阵，则组合收益 r_x 可以标准化为：

$$\frac{r_x - x' \mu}{(x' Q x)^{1/2}} \sim G(K)$$

其中： $G(K)$ 是一个确定的分布函数，具体到正态分布假设下即为标准正态分布。从上式可以看出，只要经过简单的变换（如将协方差矩阵用散度矩阵代替），就可以将前文的结论推广到椭球等高分布族的假设下。文献 [10] 分析了在椭球等高分布假设下的均值 - 散度模型。

当收益率的分布存在有限的均值和方差时候，还可以利用 Chebyshev s 不等式得出 VaR 的上界^[2]，由 $\text{Prob}(z_\alpha \sigma[r_x] \leq |r_x - E[r_x]|) \leq (1/z_\alpha)^2$ ，可得 $\text{VaR}_\alpha[r_x] \leq \frac{\sigma_x}{\sqrt{1-\alpha}} - E[r_x]$ ，

然后以上界代替精确的 VaR 值作为上述优化问题的目标函数。在一般情况下 $z_\alpha < \frac{1}{\sqrt{1-\alpha}}$ ，对上界求解所得到的有效组合仍然是标准差的有效组合的子集，当然同时在正态分布假设下得到的有效边界又是上界求解所得到的有效边界的子集。

考虑置信度为 95% 时，在正态分布时 $z_\alpha = 1.64$ ，用 Chebyshev s 不等式得出的上界为 $u_\alpha = \frac{1}{\sqrt{1-0.95}} = 4.47$ 。图 8 给出了两种情况下的组合边界。

除了使用参数分布的假设，实践中另一种常用的确定分布的方法是使用历史数据去得到所谓的“经验”（empirical）分布，如历史模拟法。这样得出的分布可以脱离任何分布的假设，但所得到的分布是不连续的，同时由于 VaR 的设置方式，它所推导出的均值 - VaR 组合边界并不像正态分布假设下的均值 - VaR 的组合边界一样是凸的，而是存在大量的局部极值，这给求解过程带来了很大的麻烦，此时只能使用数值方法来计算出组合边界。图 9 中所展现的组合边界是使用模拟数据计算出的，使用历史模拟法所得的组合边界上存在着大量的局部极值，使用一般的算法寻找切点组合时很容易陷入局部极值。从图 9 中也可以看出，使用历史分布得出的均值 - VaR 的组合边界虽然不是凸的，但可以近似看作是凸的，这意味着可以使用平滑技术（如样条）对原数据进行逼近，之后使用一般的算法求解切点组合时就不用担心陷入局部极值了。

通过将模型向非正态分布的推广，我们可以发现有效边界的位置对于分布假设是敏感的。不恰当分布假设以及分位量的估计误差会造成管理者和监管者的错觉。对于同一个组合，较大的分位量即意味着有较大的 VaR 值，表现为较大的风险。这样潜在高风险的市场可能表现出低风险的现象，比如某个外汇的历史数据可能具有很小的波动率，但同时面临着非常高的贬值风险。此时基于 VaR 的风险管理体系也可能含有促使金融机构持有更大的风险头寸或创造逃避监管的机会的因素，这在一定的条件下甚至会造成整个金融市场的不稳定。当然也可能出现相反的情况。一般而言，机构管理者对所持有的资产有更多的信息，他们按照所持有的信息选择有效组合头寸达到收益最大化。但这种头寸选择在监管者（具有不太完全的信息）眼中可能有太大的风险。而事实情况是在新的复杂的金融市场环境下，金融监管当局对及时、有效地从外部监管金融机构的风险状况和决策感到越来越力不从心，以巴塞尔银行监督管理委员会为代表的国际金融监管当局开始强调金融机构建立起有效的风险内部控制机制，并同意而且是鼓励金融机构采用其内部模型来计量风

险，从而使得外部监管与内部控制更加紧密地结合起来。

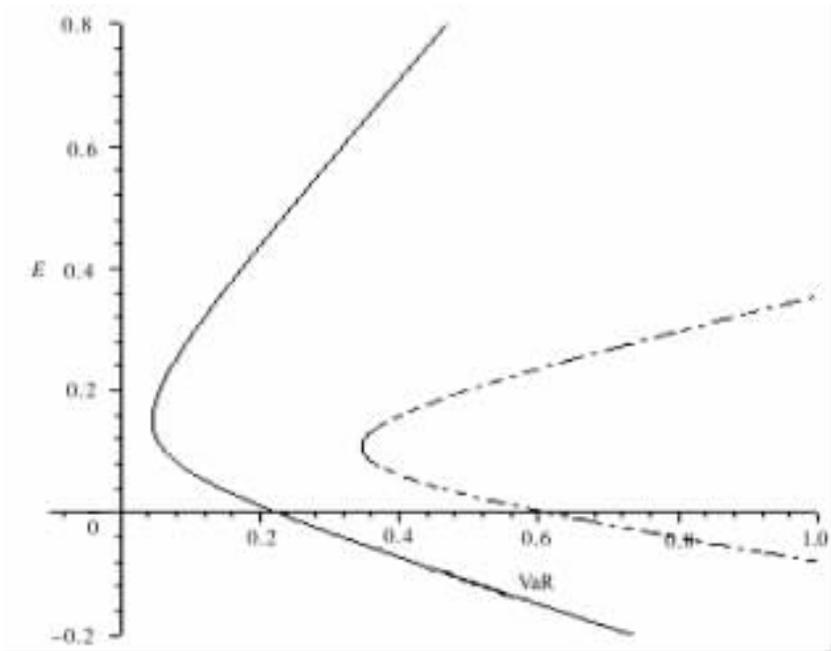


图 8 从左到右分别为正态分布与上界两种情形下的组合边界

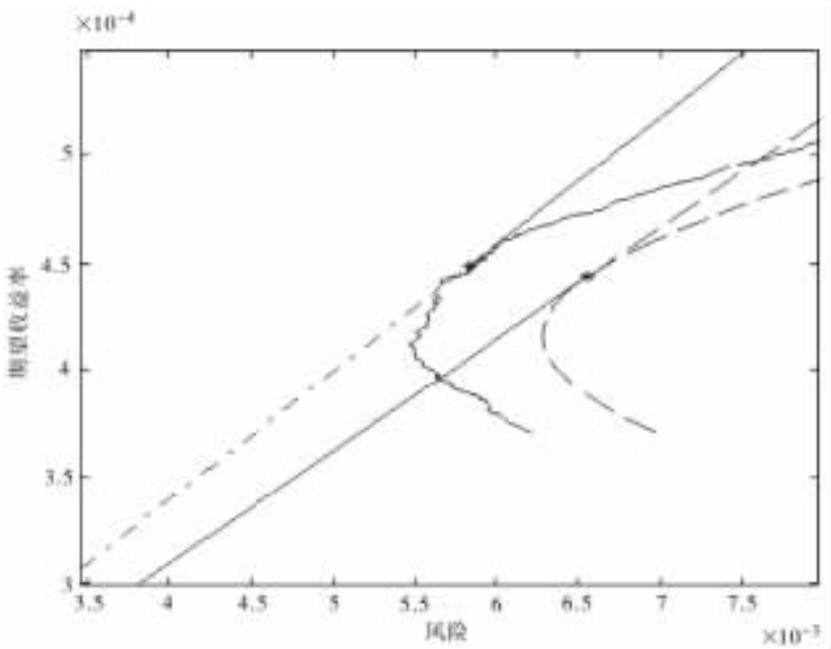


图 9 从左到右分别为使用“经验”分布和正态分布所计算的组合边界

五、结束语

本文刻画了使用 VaR 代替标准差作为风险指标的均值 - VaR 模型。我们得出的结论是：均值 - VaR 模型与均值 - 方差模型有许多相似之处，但即使是在正态分布的假设下，均值 - VaR 模型和均值 - 方差模型也是不尽相同的，均值 - VaR 模型组合边界的位置受所选择置信度的影响，特别是在选择了过于小的置信度水平时，得不出组合边界的解。引入无风险资产后，VaR 的有效边界和标准差的有效边界代表同样的组合，且有相同的市场组合，这意味着两种模型在一定条件下是兼容的。此时资本资产定价模型（CAPM）仍然成立，其应用范围得到了拓宽。引入负债后，组合边界发生了变化，但渐近线的斜率仍然保持不变。

需要指出的是，VaR 技术是一种非常复杂的方法，VaR 计算方法的选用以及历史数据的处理方式等都将显著地影响到决策过程的复杂程度和最终的精度。在使用 VaR 时，对它的假设条件和计算方法的了解是非常必要的。本文只考虑了单期模型的情形，向多期和连续时间情况下的扩展值得进一步研究。

参考文献

- [1] Markowitz H. , Portfolio selection , Journal of Finance , 1952 , Vol.7 , No.1 , pp7 ~ 91.
- [2] Roy A. D. , Safety-first and the holding of assets , Econometrica , 1952 , Vol.20 , No.3 , pp431 ~ 449.
- [3] Pyle D. H. , Turnovsky S. J. , Safety-first and expected utility maximization in mean-standard deviation portfolio analysis , The Review of Economics and Statistics , 1970 , Vol.52 , No.1 , pp75 ~ 81.
- [4] Alexander G. , Baptista A. , Economic Implications of Using a Mean-VaR Model for Portfolio Selection : A Comparison with Mean-Variance Analysis , Journal of Economic Dynamics & Control , 2002 , Vol.26 , pp1159 ~ 1193.
- [5] 吴世农, 陈斌:《风险度量方法与金融资产配置模型的理论和实证研究》,《经济研究》,1999年第9期,第30~38页。
- [6] 李健:《收益率非规则分布条件下有效风险度量方法的寻找——与吴世农、陈斌二位先生商榷》,《经济研究》,2000年第1期,第66~70页。
- [7] Baumol W. J. , An expected gain-confidence limit criterion for portfolio selection , Management Science , 1963 , Vol.10.
- [8] Basak S. , Shapiro A. , Value-at-Risk based risk management : Optimal policies and asset prices , Review of Financial Studies , 2001 , Vol.14 , pp371 ~ 405.
- [9] Keel A. , Muller H. H. , Efficient portfolios in the asset liability context , Astin Bulletin , 1995 , Vol.25 , No.1 , pp33 ~ 48.
- [10] Ortobelli S. , Rachev S. , Safety first analysis and stable paretian approach to portfolio choice theory , Mathematical and Computer Modelling , 2001 , Vol.34 , pp1037 ~ 1072.

黎曼流形上不变集的分形维数估计*

瞿成勤 周作领

一、引言

设 $K \subset \mathbb{R}^n$ 是欧氏空间中 C^1 -映射的紧不变集, Douady 和 Oesterlé^[1]第一次给出了这种不变集 Hausdorff 维数的上界, 这就是著名的 Douady-Oesterlé 型估计. 为了研究 Hilbert 空间中无穷维动力系统生成的不变集的分形性质, Temam^[2]利用 Douady-Oesterlé 技巧给出了不变集的 Hausdorff 维数和分形维数的估计, 证明了: 如果不变集 X 的 m 个整体 Lyapunov 指数的和被函数 $f(m)$ 控制, 那么 $\dim_H X \leq d_0$, 且 $\dim_F X \leq c \cdot d_0$ ($c > 1$), 这里 $f(d_0) = 0$. 然而在实际应用中所考虑的映射通常是作用在实体位相空间上 (见文 [3~8]), 而实体位相同空间一般都是流形, 例如二阶数字滤波器是任意数字滤波器得以实现的基本结构单元^[5]. 通常的具有两个补充非线性故障数字滤波器是以黎曼流形上依赖于参数的分段光滑微分方程为其模型, 而此系统对某个特定参数有类似分形的不变集. 事实上, 在黎曼流形上存在大量的分形不变集, 为了研究这种不变集的分形性质, 文 [9~11] 利用 Douady-Oesterlé 技巧给出了黎曼流形上 C^1 -映射不变集 Hausdorff 维数和分形维数的估计, 得到了与 Temam^[2]同样的结果, 即: 如果不变集 X 的 m 个整体 Lyapunov 指数和被函数 $f(m)$ 控制, 那么 $\dim_H X \leq d_0$, 且 $\dim_F X \leq c \cdot d_0$ ($c > 1$), 这里 $f(d_0) = 0$. 本文发展文 [12] 的技巧, 在 $f(m)$ 是凹函数的情形下证明了这两种维数有同样的上界, 从而给出了黎曼流形上不变集的分形维数更为精细的估计. 利用此估计和映射的非单映射度给出黎曼流形上非单映射不变集的分形维数的上界, 在某些情形下改进了文 [13] 中定理 3.1, 并减弱了 Douady-Oesterlé 定理中关于映射光滑性的假设.

二、基本定义和记号

我们首先回顾一些基本定义和记号. 设 (X, ρ) 是一度量空间, $K \subset X$ 是一紧子集, ϵ 是一实数, 记 $N_K(\epsilon)$ 为覆盖 K 半径为 ϵ 的闭球的最少个数, 即:

$$N_K(\epsilon) = \min \left\{ k : K \subset \bigcup_{i=1}^k B(u_i, \epsilon), u_i \in X \right\}$$

定义:

* 基金项目: 国家自然科学基金资助项目、晨兴中心资助课题. 原载:《数学学报》, 2004 年第 1 期.

$$\dim_F(K) = \limsup_{\epsilon \rightarrow 0} \frac{\log N_K(\epsilon)}{\log(1/\epsilon)}$$

是 K 的分形维数

为了利用奇异值函数，我们首先回顾奇异值的定义。设 E, E' 是两个 n -维欧几里德空间，它们的内积分别是 $(\cdot, \cdot)_E$ 和 $(\cdot, \cdot)_{E'}$ 。设 $L^*: E' \rightarrow E$ 是 L 的伴随算子，即对任意 $x \in E$ 和 $y \in E'$ ，有 $(Lx, y)_E = (x, L^*y)_{E'}$ 。称半正定算子 $(L^*L)^{\frac{1}{2}}$ 的特征值是算子 L 的奇异值，记为 $\alpha_1(L) \geq \dots \geq \alpha_n(L)$ 。对任意 $j \in \{0, 1, \dots, n\}$ ，定义：

$$w_j(L) = \begin{cases} \alpha_1(L) \dots \alpha_j(L), & j > 0 \\ 1, & j = 0 \end{cases}$$

对任意 $d \in (0, n]$ ，定义：

$$w_d(L) = w_{[d]}(L)^{1-s} w_{[d]+1}(L)^s$$

是算子 L 的奇异值函数，这里 $[d]$ 是 d 的整数部分， $s = d - [d]$ 类似地，对椭球我们也有上述定义，设 E 是 n -维欧几里德空间， u_1, \dots, u_n 是 E 的标准正交基， ϵ 是椭球，其半轴长分别是 $\alpha_1(\epsilon) \geq \dots \geq \alpha_n(\epsilon) > 0$ ，即：

$$\epsilon = \left\{ \alpha_1 u_1 + \dots + \alpha_n u_n \in E : (\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n, \left(\frac{\alpha_1}{\alpha_1(\epsilon)} \right)^2 + \dots + \left(\frac{\alpha_n}{\alpha_n(\epsilon)} \right)^2 \leq 1 \right\}$$

对任意 $j \in \{0, 1, \dots, n\}$ ，定义：

$$w_j(\epsilon) = \begin{cases} \alpha_1(\epsilon) \dots \alpha_j(\epsilon), & j > 0 \\ 1, & j = 0 \end{cases}$$

对任意 $d \in (0, n]$ ，定义：

$$w_d(\epsilon) = w_{[d]}(\epsilon)^{1-s} w_{[d]+1}(\epsilon)^s$$

引理 1.1^[11] 设 $L: E \rightarrow E'$ 和 $L': E' \rightarrow E''$ 的两个线性算子，则对任意 $d \in [0, n]$ ，有 $w_d(LL') \leq w_d(L) w_d(L')$ 。

引理 1.2^[9] 设 ϵ 是一椭球，其半轴长分别是 $\alpha_1(\epsilon) \geq \dots \geq \alpha_n(\epsilon) > 0$ ，那么对任意 $\eta > 0$ ，存在椭球 $\epsilon' = \left(1 + \frac{\eta}{\alpha_n(\epsilon)}\right) \epsilon$ ，使得 $\epsilon + B_\eta(0) \subset \epsilon'$ 。这里 $B_\eta(0)$ 是中心在原点、半径为 η 的闭球。

引理 1.3^[14] 设 ϵ 是一椭球，其半轴长分别是 $\alpha_1(\epsilon) \geq \dots \geq \alpha_n(\epsilon) > 0$ ，那么对任意 $r > 0$ ， ϵ 可由 $2^l \frac{\omega_1(\epsilon)}{r^l}$ 个半径为 $r\sqrt{l+1}$ 的闭球覆盖。这里 l 由下式定义：

$$l := \begin{cases} 0, & r > \alpha_1(\epsilon) \\ m, & \alpha_{m+1}(\epsilon) < r \leq \alpha_m(\epsilon), m \in \{1, \dots, n-1\} \\ n, & r \leq \alpha_n(\epsilon) \end{cases}$$

三、主要定理

设 (M, g) 是 n -维光滑黎曼流形， $U \subset M$ 是一开子集， $\varphi: U \rightarrow M$ 是一 C^1 -映射，

$K \subset U$ 是关于 φ 的紧不变集, 那么切映射 $d_u\varphi: T_uM \rightarrow T_{\varphi(u)}M$ 是一线性算子. 记:

$$\bar{\omega}_d(p) := \sup_{u \in K} \omega_d(d_u\varphi^p) \quad \underline{\alpha}_n(p) := \inf_{u \in K} \alpha_n(d_u\varphi^p) \quad (2.1)$$

那么极限 $\lim_{p \rightarrow \infty} \frac{\bar{\omega}_d(p)}{p}$ 存在, 记为 $q(d)$.

定理 2.1 设 (M, g) 是 n -维光滑黎曼流形, $U \subset M$ 是一开子集, $\varphi: U \rightarrow M$ 是一 C^1 -映射, $K \subset U$ 是关于 φ 的紧不变集. 如果存在一凹函数 $f(t)$, 使得 $f(t) \geq q(t)$, 则 $\dim_F(K) \leq d$. 这里 d 满足 $f(d) = 0$.

证明 由于 $f(t)$ 是凹函数, 且 $q(0) = 0$, 因此, 我们能选取一线性函数 $g(t)$, 使得 $g(d) = 0$, 且 $f(t) \leq g(t)$ ($t \geq 0$), 从而我们只要在 $f(t) = -\alpha t + b$ ($\alpha, b > 0$) 情形下证明定理即可.

由 (2.1) 式知, 对任意 $\epsilon > 0$, 存在一正常数 $c_d := c_d(\epsilon)$, 使得:

$$\bar{\omega}_d(p) \leq c_d e^{(q(d)+\epsilon)p} \quad \text{for } p = 1, 2, \dots \quad (2.2)$$

取 $h \in (0, \alpha)$, D 是一正实数, 且 $D > \max_{0 \leq d \leq n} (d + f(d)/h) = b/h$, 则对任意 $d \in [0, n]$, 有 $-h(D-d) + f(d) < 0$, 从而当 $p \rightarrow \infty$ 时, 有 $\bar{\omega}_d(p) e^{-hp(D-d)} \rightarrow 0$.

因此, 可选取 p 足够大, 使得 $\sigma := e^{-hp} < \frac{1}{4\sqrt{n+1}}$, 且:

$$2^{n+2D}(n+1)^{D/2} \max_{0 \leq d \leq n} \bar{\omega}_d(p) \sigma^{D-d} < 1 \quad (2.3)$$

由于 $K \subset U$ 是紧的, 因此存在一开子集 $V \subset M$ 和一紧子集 $K \subset M$, 使得 $K \subset V \subset K \subset U$. 从而对上面选定的 p 和任意 $\eta \in (0, \underline{\alpha}_n(p))$, 可选取 $r_1 (> 0)$ 足够小, 使得对任意 $u, v \in V$, $g(uv) \leq r_1$, 有:

$$\| \tau_{\varphi^p(u)}^{\varphi^p(v)} d_v \varphi^p - d_u \varphi^p \| \leq \eta$$

这里, $\| \cdot \|$ 是算子范数; $g(\cdot)$ 表示 M 上任意两点之间的测地距离; T_u^v 表示由平行移动诱导的两个切空间 T_uM 和 T_vM 之间的等距.

设 $\exp_u: T_uM \rightarrow M$ 表示点 $u \in M$ 处的切映射, 由于 \exp_u 是光滑映射, 且 $\|dO_u \exp_u\| = 1$, 因此存在 $r_u > 0$, 使得对任意 $v \in B_{r_u}(O_u)$, 有 $\|d_v \exp_u\| \leq 2$. 这里 O_u 表示切空间 T_uM 的原点. 由于 V 是含在一个紧集内, 因此存在 $r_2 > 0$, 使得对任意 $u \in V$ 和任意 $v \in B_{r_2}(O_u)$, 有:

$$\|d_v \exp_u\| \leq 1 \quad (2.5)$$

由于 $V \subset \bar{K}$ (紧), 因此存在 $\alpha > 0$, 使得 $\sup_{u \in V} \alpha_1(d_u\varphi^p) < \alpha$, 且可选取实数:

$$r_0 \leq \min \left\{ r_1, \frac{r_2}{4\sqrt{n+1} + 2\alpha} \right\} \quad (2.6)$$

足够小, 使得任意含有 K 中点的球 $B_{r_0}(u)$ 都完全含在 V 中. 设 $\epsilon \in (0, r_0)$. 由于 K 是紧的, 因此存在有限个点 $u_j \in V$ ($j = 1, \dots, N_K(\epsilon)$), 使得 $K = \bigcup_{j=1}^{N_K(\epsilon)} B_\epsilon(u_j) \cap K$ 且 $\varphi^p(K) = \bigcup_{j=1}^{N_K(\epsilon)} \varphi^p(B_\epsilon(u_j) \cap K)$. 可对微映射 φ^p , 由 Taylor 公式知, 对任意 $v \in B_\epsilon(u_j)$, 我们有 [9]

$$\| \exp_{\varphi^p(u_j)}^{-1} \varphi^p(\nu) - d_{u_j} \varphi^p(\exp_{u_j}^{-1} \nu) \| = \sup_{\omega \in B_r(u_j)} \| \tau_{\varphi^p(u_j)}^{\varphi^p(\omega)} d_w \varphi^p \tau_{u_j}^{\omega} - d_{u_j} \varphi^p \| \cdot \| \exp_{u_j}^{-1}(\omega) \|^2$$

结合 (2.4) 式, 得:

$$\varphi^p(B_{\epsilon}(u_j)) \subset \exp_{\varphi^p(u_j)}(d_{u_j} \varphi^p(B_{\epsilon}(O_{u_j})) + B_{\eta}(\epsilon))$$

由于 $\epsilon_j = d_{u_j} \varphi^p(B_1(O_{u_j}))$ 是 $E = T_{\varphi^p(u_j)} M$ 中的椭球, 由引理 1.2, 得:

$$\varphi^p(B_{\epsilon}(u_j)) \subset \exp_{\varphi^p(u_j)}(\epsilon_j + B_{\eta}(O_{\varphi^p(u_j)})) \subset \exp_{\varphi^p(u_j)}(\epsilon'_j)$$

这里, $\epsilon'_j = \epsilon_j (1 + \frac{\eta}{\alpha_n(\epsilon)})$. 引理 1.2 蕴含了 $\omega_d(\epsilon'_j) \leq (2\epsilon)^d \omega_d(d_{u_j} \varphi^p)$ 设 $r = 2\epsilon \in \sigma$, 由引理 1.3 得 ϵ'_j 能被 $N_{\epsilon'_j}(\sqrt{1+1}r)$ 个半径为 $\sqrt{1+1}r$ ($\leq \sqrt{1+1}r$) 的闭球覆盖.

这里:

$$N_{\epsilon'_j}(\sqrt{1+1}r) \leq 2^1 \frac{\omega_1(\epsilon'_j)}{r^1} \leq 2^n \max_{1 \leq l \leq n} \frac{\bar{\omega}_l(p)}{\sigma^l}$$

因此, ϵ'_j 能被 $N_j := N_{\epsilon'_j}(\sqrt{n+1}r)$ 个半径为 $\sqrt{n+1}r$ 的闭球 B_{ij} 覆盖, 这里:

$$N_{\epsilon'_j}(\sqrt{n+1}r) \leq 2^n \max_{1 \leq l \leq n} \frac{\bar{\omega}_l(p)}{\sigma^l}$$

对任意 $i = 1, \dots, N_j$ 如果 $B_{ij} \cap \epsilon'_j \neq \emptyset$, 由 $\sigma < 0$ 和 (2.6) 式得 $B_{ij} \subset B_{r_2}(O_{\varphi^p(u_j)})$. 利用 (2.5) 式得 $\exp_{\varphi^p(u_j)}(B_{ij}) \subset B_M(\omega_{ij}, 2\sqrt{n+1}r)$, 因此有:

$$K \subset \bigcup_{j=1}^{N_k(\epsilon)} \bigcup_{i=1}^{N_j} B_M(\omega_{ij}, 2\sqrt{n+1}r)$$

且:

$$N_K(2\sqrt{n+1}r) \leq 2^n N_K(\epsilon) \max_{1 \leq l \leq n} \frac{\bar{\omega}_l(p)}{\sigma^l}$$

记 $s = 4\sqrt{n+1}\sigma$, 由 (2.3) 和 (2.7) 式, 得:

$$N_K(s\epsilon) \leq 2^n N_K(\epsilon) \max_{1 \leq l \leq n} \frac{\bar{\omega}_l(p)}{\sigma^l} \leq 2^{n+2D} (n+1)^{D^2} (\max_{1 \leq l \leq n} (\bar{\omega}_l(p) \sigma^{D-1})) s^{-D} N_K(\epsilon) \leq S^{-D} N_K(\epsilon)$$

由于 $s < 1$, 因此对任意整数 k , 我们有 $N_K(s^k \epsilon) \leq s^{-kD} N_K(\epsilon)$. 对任意 $\delta > 0$, 设 k 满足 $s^{k+1} \epsilon < \delta \leq s^k \epsilon$, 则 $N_K(\delta) \leq (1/\delta)^D (\epsilon/S)^D N_K(\epsilon)$. 因此 $\dim_F(K) \leq D$, 从而 $\dim_F(K) \leq b/h$. 令 $h \rightarrow a$, 我们有 $\dim_F(K) \leq b/a$. 这就完成了定理 2.1 的证明.

四、分形维数估计

Lyapunov 指数在动力系统中有非常重要的应用 (见文 [2, 15]), 通常, 人们利用 Lyapunov 指数代替奇异值来对不变集的分形维数进行估计.

设 K 是关于 φ 的紧不变集. 定义:

$$\mu_1 = \lim_{p \rightarrow +} \frac{1}{p} \log \bar{\omega}_1(p), \mu_j = \lim_{p \rightarrow +} \frac{1}{p} \log \frac{\bar{\omega}_j(p)}{\bar{\omega}_{j-1}(p)} \quad \text{for } j = 2, \dots, n$$

定理 3.1 设 (M, g) 是 n -维光滑黎曼流形, $U \subset M$ 是一开子集, $\varphi: U \rightarrow M$ 是一

C^1 -映射, $K \subset U$ 是关于 φ 的紧不变集, 如果 $\mu_1 \geq \dots \geq \mu_n$, 且存在某一正整数 $m \in \{1, \dots, n-1\}$, 使得:

$$\mu_1 + \dots + \mu_{m+1} < 0 \quad (3.1)$$

则 $\dim_F(K) \leq m + \frac{(\mu_1 + \dots + \mu_m)_+}{|\mu_{m+1}|}$. 这里, $(\cdot)_+$ 表示实数的正部.

证明 对任意 $j = 1, \dots, n$ 我们有 $\mu_1 + \dots + \mu_j = \lim_{p \rightarrow \infty} \frac{1}{p} \log \bar{\omega}_j(p)$, 因此对任意实数 $d = [d] + s \geq 1$ 有:

$$\lim_{p \rightarrow \infty} \frac{1}{p} \log \bar{\omega}_d(p) \leq \lim_{p \rightarrow \infty} \left(\frac{1-s}{p} \log \bar{\omega}_{[d]}(p) \right) + \lim_{p \rightarrow \infty} \left(\frac{s}{p} \log \bar{\omega}_{[d]+1}(p) \right) \leq \mu_1 + \dots + \mu_{[d]} + s\mu_{[d]+1}$$

设 $f(t) := \mu_1 + \dots + \mu_{[t]} + s\mu_{[t]+1}$. 这里, $s = t - [t]$. 显然, $f(t)$ 是一分段线性函数且是凹的. 记 $d = m + \frac{\mu_1 + \dots + \mu_m}{|\mu_{m+1}|}$, 则 $f(d) = 0$. 因此由定理 2.1, 得 $\dim_F(K) \leq m + \frac{(\mu_1 + \dots + \mu_m)_+}{|\mu_{m+1}|}$. 这就完成了定理 3.1 的证明.

注 1 Heineken 和 Reitmann^[10]证明了 $\dim_F(K) \leq \max_{j=1, \dots, m} \left(m + \frac{(\mu_1 + \dots + \mu_j)_+}{|\mu_{m+1}|} \right)$. 因此我们的结果在 $\mu_1 \geq \dots \geq \mu_n$ 时改进了文 [10] 中的定理 2.2

五、非单映射不变集的分形维数

事实上, 人们在具体的物理和技术系统的应用中所考虑的映射时常是非常映射, 且不满足 Douady-Oesterlé 定理中关于映射光滑性的假设 (见文 [16] 中例 2 和 3). 对于那样的非单映射, 文 [16] 第一次利用非单映射度给出了不变集的 Douady-Oesterlé 型 Hausdorff 维数估计.

设 (M, g) 是 n -维光滑黎曼流形, $U \subset M$ 是一开子集, $\varphi: U \rightarrow M$ 是一映射. 处理不可微映射的一种方式, 假设:

(C1) 存在有限个紧子集 $K \subset M$, $K_1 \subset M, \dots, K_\kappa \subset M$, 使得 $K = \bigcup_{i=1}^\kappa K_i$, $\varphi(K_i) = K$, $i = 1, \dots, \kappa$.

(C2) 每个映射 $\varphi|_{K_i}$ ($i = 1, \dots, \kappa$) 在 K_i 的一个开邻域 $U_i \subset M$ 内都能延拓成 C^1 -微分同胚.

下面, 我们称 (C2) 中的 $\varphi_i := \varphi|_{K_i}$ 是部分映射.

设 $D_m = \{ (i_1, \dots, i_m) : 1 \leq i_j \leq \kappa, 1 \leq j \leq m \}$, 对任意 $\tau = (i_1, \dots, i_m) \in D_m$, 记:

$$\varphi_\tau^{-1} = \varphi_{i_1}^{-1} \circ \dots \circ \varphi_{i_m}^{-1}, w_d(\tau, m) = \sup_{u \in K} \omega_d(d_u \varphi_\tau^{-1}), w_d(m) = \sup_{\tau \in D_m} w_d(\tau, m)$$

且:

$$g(d) = \lim_{p \rightarrow \infty} \frac{\log w_d(p)}{p} \quad (4.1)$$

定理 4.1 设 (M, g) 是 n -维光滑黎曼流形, $U \subset M$ 是一开子集, $\varphi: U \rightarrow M$ 是一映射。且满足 (C1) 和 (C2)。如果存在一凹函数 $f(t)$, 使得 $f(t) \geq g(t)$, 则 $\dim_F(K) \leq d$ 。这里, d 满足 $f(d) = -\log k$ 。

为了证明定理 4.1, 我们需要下述引理:

引理 4.1^[13] 设 (M, g) 是 n -维光滑黎曼流形, $U \subset M$ 是一开子集, $\varphi: U \rightarrow M$ 是一映射, 且满足 (C1) 和 (C2), 那么对任意自然数 p , 映射 φ^p 也满足 (C1) 和 (C2)。

定理 4.1 的证明 类似于定理 2.1 的证明, 我们可以假设 $f(t) = -at + b$ 。这里, $a, b > 0$ 。

由 (4.1) 式知, 对任意 $\epsilon > 0$, 存在一正常数 $c_d := c_d(\epsilon)$, 使得:

$$w_d(p) \leq c_d e^{(\epsilon(d) + \epsilon)p} \quad (4.2)$$

取 $h \in (0, a)$, D 是一正实数, 且 $D > \max_{0 \leq d \leq n} (d + \frac{f(d) + \log k}{h}) = \frac{b + \log k}{h}$, 则对任意 $d \in [0, n]$, 有 $-h(D - d) + \log k + f(d) < 0$, 从而当 $p \rightarrow \infty$ 时, 有 $w_d(p) \kappa^p e^{-hp(D-d)} \rightarrow 0$ 。因此, 可选取 p 足够大, 使得 $\sigma := e^{-hp} < \frac{1}{4\sqrt{n+1}}$, 且:

$$2^{n+2D} (n+1)^{D^2} \max_{0 \leq d \leq n} w_d(p) \kappa^p \sigma^{D-d} < 1 \quad (4.3)$$

由于每个部分映射 $\varphi|_{K_i}$ ($i = 1, \dots, \kappa$) 在 K_i 的一个开邻域 $U_i \subset M$ 内都能延拓成 C^1 -微分同胚, 由引理 4.1 知, 对每个映射 $\varphi_\tau^{-1} = \varphi_{i_1}^{-1} \circ \dots \circ \varphi_{i_p}^{-1}$, 存在 K 的一个开邻域 V_{i_1, \dots, i_p} , 使得 $\varphi_{i_1}^{-1} \circ \dots \circ \varphi_{i_p}^{-1}: V_{i_1, \dots, i_p} \rightarrow M$ 是一 C^1 -映射。令 $K_\tau = \varphi_\tau^{-1}(K)$, 类似于定理 2.1 的证明, 存在 $r_\tau > 0$, 使得对任意 $\epsilon \in (0, r_\tau)$, 我们有:

$$N_{K_\tau}(s \in) \leq 2^n N_K(\epsilon) \max_{1 \leq l \leq n} \frac{w_l(p)}{\sigma^l} \leq 2^{n+2D} (n+1)^{D^2} (\max_{1 \leq l \leq n} w_l(p) \sigma^{D-1}) s^{-D} N_K(\epsilon)$$

这里, $s = 4\sqrt{n+1}\sigma$ 。记 $r_0 = \min\{r_\tau : \tau \in D_p\}$, 对任意 $\epsilon \in (0, r_0)$, 由 (4.3) 式, 得:

$$N_K(s \in) \leq 2^{n+2D} (n+1)^{D^2} (\max_{1 \leq l \leq n} w_l(p) \kappa^p \sigma^{D-1}) s^{-D} N_K(\epsilon) \leq s^{-D} N_K(\epsilon)$$

类似于定理 2.1 的证明, 我们有 $\dim_F(K) \leq D$ 因此 $\dim_F(K) \leq \frac{b + \log k}{h}$ 。令 $h \rightarrow a$,

我们有 $\dim_F(K) \leq \frac{b + \log k}{a}$ 。这就完成了定理 4.1 的证明。

记:

$$\Omega_{d,k} := \left(\max_{i=1, \dots, k} \sup_{u \in K} \alpha_1(d_u \varphi_i^{-1}) \right) \dots \left(\max_{i=1, \dots, k} \sup_{u \in K} \alpha_{[d]}(d_u \varphi_i^{-1}) \right) \left(\max_{i=1, \dots, k} \sup_{u \in K} \alpha_{[d]+1}(d_u \varphi_i^{-1})^{d-[d]} \right)$$

$$w_{d,k} = \max_{i=1, \dots, k} \left(\sup_{u \in K} \alpha_1(d_u \varphi_i^{-1}) \dots \alpha_{[d]}(d_u \varphi_i^{-1}) \alpha_{[d]+1}(d_u \varphi_i^{-1})^{d-[d]} \right)$$

显然, 对任意 d , 我们有:

$$w_{d,k} \leq \Omega_{d,k} \quad (4.4)$$

推论 4.1 设 (M, g) 是 n -维光滑黎曼流形, $U \subset M$ 是一开子集, $\varphi: U \rightarrow M$ 是一映射, 且满足 (C1) 和 (C2)。如果存在 d , 使得:

$$\Omega_{d, \kappa} < \frac{1}{\kappa} \quad (4.5)$$

则 $\dim_F(K) \leq d$

证明 由于:

$$g(d) = \lim_{p \rightarrow \infty} \frac{\log w_d(p)}{p} \leq \log \Omega_{d, \kappa} = \log \left(\max_{i=1, \dots, k} \sup_{u \in K} \alpha_1(\varphi_i^{-1}) \right) + \dots \\ + \log \left(\max_{i=1, \dots, k} \sup_{u \in K} \alpha_{[d]}(\varphi_i^{-1}) \right) + (d - [d]) \log \left(\max_{i=1, \dots, k} \sup_{u \in K} \alpha_{[d]+1}(\varphi_i^{-1}) \right)$$

记 $f(t) = \log \Omega_{t, \kappa}$, 由于 $\alpha_1(\varphi_i^{-1}) \geq \dots \geq \alpha_n(\varphi_i^{-1})$ for $i = 1, \dots, k$, 因此 $f(t)$ 是一段线性函数且是凹的。令 d_0 是方程 $f(t) = -\log \kappa$ 的解, 由定理 4.1 得 $\dim_F(K) \leq d_0$, 假设 (4.5) 式蕴含了 $d \geq d_0$, 因此 $\dim_F(K) \leq d$ 。

注 2 在文 [13] 中 Franz 证明了, 若 d 满足 $\omega_{d, k} < \frac{1}{k}$, 则 $\dim_H(K) \leq d$ 。一般而言 $\dim_H(K) \leq \dim_F(K)$, 因此, 推论 4.1 在 (4.4) 式取等号的情形下改进了文 [13] 中的定理 3.1。

例 4.1 (Julia 集) 非单映射的一个著名例子是复平面上生成 Julia 集的二次多项式。下面, 我们利用定理 4.1 给出 Julia 集的分形维数估计。现在考虑具体形成 $\varphi: C \rightarrow C_\varphi (z) = z^2 + c$ 的多项式的 Julia 集, 这里 $c \in C$ 是一参数, 如果 $|c| > \frac{1}{4} (5 + 2\sqrt{6})$, 由文 [17] 知, 存在 K_1, K_2 , 使得:

$$K = K_1 \cup K_2 \subset \{z \in C : |z| \leq \sqrt{|2c|}\}, \quad K_1 \cap K_2 = \emptyset$$

且:

$$\sup_{z \in K} \alpha_i(d_z \varphi_{K_j}^{-1}) \leq \frac{1}{2\sqrt{|c|} - \sqrt{|2c|}} \quad i, j = 1, 2$$

因此有:

$$g(d) = \lim_{p \rightarrow \infty} \frac{\log w_d(p)}{p} \leq -d \log 2 \sqrt{|c| - \sqrt{|2c|}}$$

记 $f(d) = -d \log 2 \sqrt{|c| - \sqrt{|2c|}}$ 则 $g(d) \leq f(d)$, 由定理 4.1 我们有:

$$\dim_F(K) \leq \frac{\log 2}{\log 2 \sqrt{|c| - \sqrt{|2c|}}}$$

例 4.2^[18] 设 $\Gamma = \{k(2\pi i) : k \in Z\}$, $\pi: R^2 \rightarrow R^2/\Gamma$ 是标准投影映射, 则 $\{(\tau_{1\tau(U)}^{-1}, \tau(U)) : U \subset R^2 \text{ 是开集}, \tau: U \rightarrow R^2/\Gamma \text{ 是单射}\}$ 是 R^2/Γ 的一个光滑坐标图, 因此, R^2/Γ 是一个 2-维光滑黎曼流形。记:

$$M = R^2/\Gamma = R \times S^1 = \{(u, \nu) : u \in R, \nu = e^{i\theta}, \theta \in [0, 2\pi)\}$$

$$A_1 = \{(u, \nu) \in M : u \in [0, \frac{1}{3}]\}, \quad A_2 = \{(u, \nu) \in M : u \in [\frac{2}{3}, 1]\}$$

这里 S^1 是单位圆。定义:

$$\varphi(u, v) = \begin{cases} (3u, v), & (u, v) \in A_1 \\ (3u - 2, v), & (u, v) \in A_2 \end{cases}$$

记 $K = \bigcap_{i=1}^{\infty} \varphi^{-i}(A_1 \cup A_2)$, $K_1 = K \cap A_1$, $K_2 = K \cap A_2$, 则对任意 $x = (u, v) \in K$, 有:

$$\alpha_1(d_x \varphi_{K_j}^{-1}) = 1, \alpha_2(d_x \varphi_{K_j}^{-1}) = \frac{1}{3}, \quad j = 1, 2$$

因此有:

$$g(d) = \lim_{p \rightarrow 0} \frac{\log w_d(p)}{p} \leq (1-d) \log 3$$

由定理 4.1, 我们有 $\dim_F(K) \leq 1 + \frac{\log 2}{\log 3}$ 。由于 $K = C \times S^1$, 这里 C 是三分 Cantor 集, 因此

$$\dim_F(K) = 1 + \frac{\log 2}{\log 3}。由此可见上面的估计是精细的。$$

致谢 衷心感谢文志英教授给予作者的宝贵意见和建议。

参考文献

- [1] Douady A., Oesterlé J., Dimension de Hausdorff des attracteurs, C. R. Acad. Sci. Paris Ser. A., 1980, Vol. 290, pp1135 ~ 1138.
- [2] Temam R., Infinite-dimensional dynamical systems in Mechanics and Physics, New York, Berlin: Springer, 1988.
- [3] Afraimovich V. S., On the Lyapunov dimension of invariant sets in a model of active medium, In: Methods of qualitative theory of differential equations, Gorki State University, Gorki, 1986, pp19 ~ 29 (Russian).
- [4] Belykh V. N., Models of Discrete systems of phase synchronization, In: Systems of phase synchronization (eds.: V. V. Shakhgildyan and L. N. Belyustina), Moscow: Radioi Svyaz, 1982, pp161 ~ 176 (Russian).
- [5] Kocarev L., Chua L. O., On chaos in digital filters: Case $b = -1$, IEEE Trans Circuits and Systems (II), Analog and Digital Signal-Processing, 1993, Vol. 40, pp404 ~ 407.
- [6] Leonov G. A., Reitmann V., Smirnova V. B., Non-local methods for pendulum-like feed-back systems, Teubner-Texte zur Mathematik, Bd 132 B G Teubner, Stuttgart, Leipzig 1992.
- [7] Noack A., Hausdorff dimension estimates for time-dependent control systems, ZAMM, 1997, Vol. 77, pp891 ~ 899.
- [8] Reitmann V., Schnabel U., HAUSDORFF dimension estimates for invariant sets of piecewise smooth maps, ZAMM, 2000, Vol. 88, pp623 ~ 632.
- [9] Boichenko V. K., Franz A., Leonov G. A., Reitmann V., Hausdorff and fractal dimension estimates for invariant sets of non-injective maps, Z. Anal. Anw., 1998, Vol. 17, pp207 ~ 223.
- [10] Heineken W., Reitmann V., Fractal dimension estimates for flow-invariant sets on manifolds, Preprint, 59/98, DFG-Schwerpunktprogramm "Dynamik: Analysis, Effiziente Simulation und Ergodentheorie", 1998.
- [11] Noack A., Reitmann V., Hausdorff dimension of invariant sets of time-dependent vector fields, ZAA, 1996, Vol. 15, pp457 ~ 473.
- [12] Chepyzhov V. V., Ilyin A. A., A note on the fractal dimension of attractors of dissipative dynamical systems, Nonlinear Anal Theory Methods Appl., 2001, Vol. 44, pp811 ~ 819.
- [13] Franz A., Upper and lower Hausdorff Dimension estimates for invariant sets of K -1-endomorphisms, Math.

- Nachr. , 2001 , Vol.223 , pp23 ~ 32.
- [14] Franz A. , Hausdorff dimension estimates for non-injective maps using the cardinality of the pre-image sets , Nonlinearity , 2000 , Vol.13 , pp1425 ~ 1438.
- [15] Constantin P. , Foiasb C. , Global Lyapunov exponents , Kaplan-Yorke formulas and the dimension of the attractors for 2D Navier-Stokes equations , Comm. Pure Appl. Math. , 1985 , Vol.38 , pp1 ~ 27.
- [16] Mirle A. , Hausdorff dimension estimates for invariant sets of k -1-maps , Preprint , 25/95 , DFG-Schwerpunktprogramm "Dynamik : Analysis , effiziente Simulation und Ergodentheorie" , 1995.
- [17] Falconer K. J. , Fractal geometry : Mathematical foundations and applications , Chichester : John Wiley & Sons , 1990.
- [18] Franz A. , Hausdorff dimension estimates for invariant sets with an equivariant tangent bundle splitting , Nonlinearity , 1998 , Vol.11 , pp1063 ~ 1074

一类广义 Sierpinski 海绵的填充测度*

朱智伟 周作领 罗俊

一、引言

确定分形集合的 Hausdorff 测度与填充测度是一个十分困难的问题。迄今为止, 仅对少数的自相似集能够得到 Hausdorff 处测度的准确值, 例如三分 Cantor 集 (见 [2]) 与某些 Sierpinski 地毯 (见 [5])。最近, [6] 得到了一类广义 Sierpinski 海绵的 Hausdorff 测度的准确值。本文将确定该类海绵的填充测度值。

给一集合 $E \subset \mathbb{R}^m$, 集合 E 的一个 δ - 填充指的是球心为集合 E 中的点, 并且直径不超过 δ 的一个互不相交的可数闭球族。对于 $s \geq 0$, 集合 E 的 s - 维预填充测度定义为:

$$P_0^s(E) = \inf_{\delta > 0} \{ P_\delta^s(E) \}$$

其中: $P_\delta^s(E) = \sup \left\{ \sum_{B_i \in \mathcal{R}} |B_i|^s : \mathcal{R} \text{ 是 } E \text{ 的一个 } \delta\text{- 填充} \right\}$, $|B_i|$ 表示球 B_i 的直径。 E 的

s - 维填充测度定义为:

$$P^s(E) = \inf \left\{ \sum_{i=1}^{\infty} P_0^s(E_i) : E \subset \bigcup_{i=1}^{\infty} E_i \right\}$$

利用上述定义的填充测度, E 的填充维数 $\text{dimp}(E)$ 可由下列式子确定:

$$\text{dimp}(E) = \inf \{ s \geq 0 : P^s(E) = 0 \} = \sup \{ s \geq 0 : P^s(E) = \infty \}$$

关于填充测度与填充维数的基本性质可见 [2, 4]。下面给出关于 Sierpinski 海绵的定义。

设整数 $m \geq 2$, $0 < \lambda \leq 2^{-m}$ 。 $I_0^m \subset \mathbb{R}^m$ 为 m - 维单位立方体, 易见 I_0^m 有 $2m$ 个顶。在 I_0^m 上挖去除位于其顶点处边长为 λ 的 2^m 个小 m - 维立方体 (其中位于顶点处的 m 个面与 I_0^m 的相应面重合) 外的其余部分, 得到的集合的闭包记为 $I_1^m(\lambda)$ 。对: $I_1^m(\lambda)$ 中的每一个小立方体重复上述过程, 得到 2^{2m} 个边长为 λ^2 的小 m - 维立方体组成的集合, 将其闭包记为 $I_2^m(\lambda)$ 。无限重复上述过程, 得到 $I_0^m \supset I_1^m(\lambda) \supset \dots \supset I_n^m(\lambda) \supset \dots$ 。非空集合 $S^m(\lambda) = \bigcap_{n=0}^{\infty} I_n^m(\lambda)$, 称为广义 Sierpinski 海绵。

显然, $S^m(\lambda)$ 是一个自相似集, 其 Hausdorff 维数与填充维数相等, 即 $s(\lambda) = \text{dimH}(S^m(\lambda)) = \text{dimp}(S^m(\lambda))$, 其中 $s(\lambda) = -\log 2^m / \log \lambda$ 。由文 [6] 知道 $S^m(\lambda)$ 的 Hausdorff 测度为 $\sqrt{m^{s(\lambda)}}$ 。本文将确定 $S^m(\lambda)$ 的填充测度, 得到如下结果:

* 基金项目: 国家自然科学基金、教育部博士点基金、广东省自然科学基金、中山大学高等学术研究中心及肇庆学院青年基金资助项目。原载:《数学年刊 A 辑》, 2004 年第 1 期。

定理 1.1 设 $P^{s(\lambda)}(S^m(\lambda))$ 表示 $S^m(\lambda)$ 的 $s(\lambda)$ -维填充测度, 则 $P^{s(\lambda)}(S^m(\lambda)) = 2^m \cdot 2^{s(\lambda)} (1-\lambda)^{s(\lambda)}$, 其中 $s(\lambda) = -\log 2^m / \log \lambda$.

推论 1.1 对于 $m=2, \lambda=1/4, S^m(\lambda)$ 为文 [5] 中所讨论的 Sierpinski 地毯, 此时 $P^1(S^2(1/4)) = 6$.

二、相关的记号和引理

对任何整数 $n \geq 1$, 包含于 $I_n^m(\lambda)$ 中边长为 λ^n 的 m -维立方体的个数为 2^{nm} . 每个这样的小立方体称为 $I_n^m(\lambda)$ 中的基本立方体, 记为 $\Theta_n(\lambda)$. 按如下方式引入一个直角坐标系: 取 I_0^m 的一个顶点为原点 O , 由原点发出的 m 条边为坐标轴. 记这些坐标轴为 X_1, X_2, \dots, X_m . I_0^m 的通过原点的对角线称为主对角线, 记为 I_0^m .

设 $n \geq 1$, 在立方体 $\Theta_n(\lambda)$ 中有 2^{m-1} 条对角线, 其中仅有一条与 I_0^m 的主对角线平行, 称该对角线为 $\Theta_n(\lambda)$ 的主对线, 并记为 L_n^m .

此外, 设集合 A 为 I_0^m 的 2^m 个顶点构成的集合, 则 $A = \{(x_1, x_2, \dots, x_m) : x_i = 0 \text{ 或 } 1, 1 \leq i \leq m\}$. 将集合 A 分成 $m+1$ 不相交的子集合, $A = \bigcup_{k=0}^m A_k$, 其中 $A_k = \{(x_1, x_2, \dots, x_m) \in A : k = \text{Card}\{i : x_i = 1\}\}$. 显然, 对任何 k , 集合 A_k 包含有 I_0^m 的 C_m^k 个顶点.

引理 2.1 (见 [6]) 设 $\vec{y} = (y_1, y_2, \dots, y_m) \in A, \Theta_1(\lambda)$ 为 $I_1^m(\lambda)$ 中以 \vec{y} 为顶点的基本立方体. 则 $\Theta_1(\lambda)$ 的 2^m 个顶点 (z_1, z_2, \dots, z_m) 满足:

$$\begin{cases} z_i = 1 \text{ 或 } 1 - \lambda, & y_i = 1 \\ z_i = 0 \text{ 或 } \lambda, & y_i = 0 \end{cases}$$

此外, 位于 $\Theta_1(\lambda)$ 主对角线上的两个顶点的坐标分别满足:

$$z_i = \begin{cases} 1 - \lambda, & y_i = 1 \\ 0, & y_i = 0 \end{cases}, \quad z_i = \begin{cases} 1, & y_i = 1 \\ \lambda, & y_i = 0 \end{cases}$$

如果顶点 $\vec{y} \in A_j, j \in \{0, 1, 2, \dots, m\}$, 设 P_1 及 Q_1 是位于 $\Theta_1(\lambda)$ 主对角线上的两个点, 由引理 2.1 则有:

$$|OP_1| = (1-\lambda)\sqrt{j}, \quad |OQ_1| = \sqrt{j + (m-j)\lambda^2}$$

为了得到 $S^m(\lambda)$ 的填充测度的准确值, 还需用到下面的引理:

引理 2.2 (见 [3]) 设 $E \subset R^m$ 是一个 Borel 集, μ 是一有限的 Borel 测度, $0 < c <$

(a) 如果对所有 $x \in E$, 有 $\liminf_{r \rightarrow 0} \frac{\mu(B(x, r))}{r^s} \geq c$ 则有 $P^s(E) \leq \frac{2^s(E)}{c}$.

(b) 如果对所有 $x \in E$, 有 $\liminf_{r \rightarrow 0} \frac{\mu(B(x, r))}{r^s} \leq c$ 则有 $P^s(E) \geq \frac{2^s(E)}{c}$.

设 $s > 0$, 定义点 x 关于有限测度 μ 及填充测度 P^s 的下球密度分别为:

$$\underline{D}^s(\mu, x) = \liminf_{r \rightarrow 0} \frac{\mu(B(x, r))}{(2r)^s} \tag{2.1}$$

$$\underline{D}^s(P^s, x) = \liminf_{r \rightarrow 0} \frac{P^s(B(x, r))}{(2r)^s} \quad (2.2)$$

对于上述 (2.2) 定义的下球密度, 根据 [4, 推论 4.2], 当 $m=2, s=1$ 时, 如果 $P^s(E) < \infty$, 则 $\underline{D}^s(P^s, x) = 1$ 对 P^s -几乎所有 $x \in E$ 成立。下面给出一个更一般的结果。

引理 2.3 设 $s > 0, E \subset \mathbb{R}^m$ 为 Borel 集。如果 $P^s(E) < \infty$, 则 $\underline{D}^s(P^s, x) = 1$ 对 P^s -几乎所有 $x \in E$ 成立。

证 设 μ 为一有限 Borel 测度, $c = \inf \{ \underline{D}^s(\mu, x) : x \in E \}$ 。由引理 2.2 (a) 得到:

$$\mu(E) \geq P^s(E) \cdot \inf \{ \underline{D}^s(\mu, x) : x \in E \} \quad (2.3)$$

对于 $\forall k > 1$, 令 $D_k = \{x \in E : \underline{D}^s(P^s, x) \geq 1 + 1/k\}$ 。当 $\mu = P^s|_E$ 时, 对集合 D_k 使用不等式 (2.3), 得到:

$$P^s(D_k) \geq (1 + 1/k) \cdot P^s(D_k)$$

因此 $P^s(D_k) = 0$, 从而 $\underline{D}^s(P^s, x) \leq 1$ 对 P^s -几乎所有 $x \in E$ 成立。

另一方面, 令 $c = \sup \{ \underline{D}^s(\mu, x) : x \in E \}$ 。利用引 2.2 (b), 得到:

$$\mu(E) \leq P^s(E) \cdot \sup \{ \underline{D}^s(\mu, x) : x \in E \} \quad (2.4)$$

从而对 $\forall k > 1$, 令 $G_k = \{x \in E : \underline{D}^s(P^s, x) \leq 1 - 1/k\}$ 。当 $\mu = P^s|_E$ 时, 对集合 G_k 使用不等式 (2.4), 得出:

$$P^s(G_k) \leq (1 - 1/k) \cdot P^s(G_k)$$

因此 $P^s(G_k) = 0$, 从而 $\underline{D}^s(P^s, x) \geq 1$ 对 P^s -几乎所有 $x \in E$ 成立。至此, 引理 2.3 得证。

对任意 Borel 集 $E \subset \mathbb{R}^m$, 定义一个以 $S^m(\lambda)$ 作为支撑的概率测度 μ 为:

$$\mu(E) = \frac{P^{s(\lambda)}(E \cap S^m(\lambda))}{P^{s(\lambda)}(S^m(\lambda))} \quad (2.5)$$

引理 2.4 $\mu, \underline{D}^{s(\lambda)}(\mu, x)$ 如上 (2.5), (2.1) 定义, 则有 $P^{s(\lambda)}(E) = \frac{1}{\underline{D}^{s(\lambda)}(\mu, x)}$ 对 μ -几乎所有 $x \in S^m(\lambda)$ 成立。

证 利用下球密度定义及 (2.5), 得到:

$$\underline{D}^{s(\lambda)}(\mu, x) \cdot P^{s(\lambda)}(S^m(\lambda)) = \underline{D}^{s(\lambda)}(P^s(\lambda), x)$$

因此由引理 2.3 得到 $\underline{D}^{s(\lambda)}(\mu, x) \cdot P^{s(\lambda)}(S^m(\lambda)) = 1$ 对 μ -几乎所有 $x \in S^m(\lambda)$ 成立。从而引理 2.4 得证。

广义 Sierpinski 海绵 $S^m(\lambda)$ 可以看作是迭代函数系 $\{f_1, f_2, \dots, f_{2^m}\}$ 的吸引子, 其中对任何 $i = 1, 2, \dots, 2^m, f_i: \mathbb{R}^m \rightarrow \mathbb{R}^m$ 为相似压缩映射, 压缩比均为 λ 。则有:

$$S^m(\lambda) = \bigcup_{n=1}^{2^m} f_n(S^m(\lambda))$$

对任给的整数 $k \geq 0$, 定义:

$$I_k = \{(i_1, i_2, \dots, i_k) : i_j \in \{1, 2, \dots, 2^m\}, j = 1, 2, \dots, k\}$$

$$I = \{(i_1, i_2, \dots) : i_j \in \{1, 2, \dots, 2^m\}, j = 1, 2, \dots\}$$

则有 $f_i(I_0^m) \subset I_0^m$ 对任何 $i \in \{1, 2, \dots, 2^m\}$ 成立, 并且 $I_k^m(\lambda) = \bigcup_{I_k} f_{i_1} \circ f_{i_2} \circ \dots \circ f_{i_k}(I_0^m)$, 其中的并取遍所有的 $(i_1, i_2, \dots, i_k) \in I_k$, 对任何 $(i_1, i_2, \dots, i_k) \in I_k$ 有:

$$f_{i_1} \circ f_{i_2} \circ \dots \circ f_{i_k}(I_0^m) \subset f_{i_1} \circ f_{i_2} \circ \dots \circ f_{i_{k-1}}(I_0^m)$$

容易看出集合序列 $f_{i_1} \circ f_{i_2} \circ \dots \circ f_{i_k}(I_0^m)$ 关于 k 是递减的。注意到该集合序列的交是一个单点集, 即:

$$\{x_{i_1 i_2 \dots}\} = \bigcap_{k=0} f_{i_1} \circ f_{i_2} \circ \dots \circ f_{i_k}(I_0^m)$$

则 $x_{i_1 i_2 \dots} \in S^m(\lambda)$, 并且 $S^m(\lambda) = \bigcup \{x_{i_1 i_2 \dots}\}$

对于 $k \geq 1$ 及 $p \geq 1$, 按如下方式构造一个集合 $F_{p,k}$ 。对任何整数 $k \geq 1$, 令 $f_1^k = \underbrace{f_1 \circ f_1 \circ \dots \circ f_1}_k$, 定义:

$$F_{p,k} = \bigcup_{n=p} \bigcup_{(i_1, i_2, \dots, i_n) \in I_n} f_{i_1} \circ f_{i_2} \circ \dots \circ f_{i_n}(f_1^k(I_0^m)) \quad (p \geq 1) \quad (2.6)$$

引理 2.5 $\mu, F_{p,k}$ 如 (2.5), (2.6) 中定义。则对任何 $k \geq 1, p \geq 1$, 有 $\mu(F_{p,k}) = 1$ 。进而对任何 $k \geq 1$ 有 $\mu(\bigcap_{p \geq 1} F_{p,k}) = 1$ 。此外还有 $\mu(\bigcap_{p \geq 1} (\bigcap_{p \geq 1} F_{p,k})) = 1$ 。

证 由 $F_{p,k}$ 的定义, 有:

$$F_{p,k} = \bigcup_{(i_1, i_2, \dots, i_n) \in I_{p,k}} f_{i_1} \circ f_{i_2} \circ \dots \circ f_{i_p}(A)$$

其中 $A = \bigcup_{n=0} \bigcup_{(i_1, i_2, \dots, i_n) \in I_n} f_{i_1} \circ f_{i_2} \circ \dots \circ f_{i_n}(f_1^k(I_0^m))$ 。易知 $\mu(F_{p,k}) = 2^{mp} \cdot \lambda^{p \cdot s(\lambda)} \mu(A) = \mu(A)$ 。因此只需证明 $\mu(A) = 1$ 即可。

显然有 $A \supset \bigcup_{q=0} A_q$, 其中 $A_0 = f_1^k(I_0^m)$, 当 $q \geq 1$ 时, 有:

$$A_q = \bigcup_{(i_1, i_2, \dots, i_{qk}) \in I_{qk}} f_{i_1} \circ f_{i_2} \circ \dots \circ f_{i_{qk}}(f_1^k(I_0^m))$$

按如下方式定义集合 A_q 的子集 B_q ($q \geq 0$) 为 $q \geq 1$ 时,

$$B_q = \bigcup_{\substack{(i_1, i_2, \dots, i_{qk}) \in I_{qk} \\ (i_{sk+1}, \dots, i_{(s+1)k}) \neq (1, \dots, 1) \\ 0 \leq s \leq q-1}} f_{i_1} \circ f_{i_2} \circ \dots \circ f_{i_{qk}}(f_1^k(I_0^m))$$

则有 $A \supset \bigcup_{q=0} A_q \supset \bigcup_{q=0} B_q$, 并且下列两条性质成立:

(i) 对任何非负整数 q, q' , 且 $q \neq q'$, 有 $B_q \cap B_{q'} = \emptyset$ 成立。

(ii) 对于 $q \geq 0$, 有 $\mu(B_q) = (1 - \frac{1}{2^{km}})^q \cdot \frac{1}{2^{km}}$ 成立。

对于 (i), 不失一般性假设 $q' > q$, 则存在一个整数 $t \geq 1$, 使得 $q' = q + t$, 由 $B_{q'}$ 的定义得到:

$$B_{q'} = \bigcup_{\substack{(i_1, i_2, \dots, i_{(q+t)k}) \in I_{(q+t)k} \\ (i_{sk+1}, \dots, i_{(s+1)k}) \neq (1, \dots, 1) \\ 0 \leq s \leq (q+t)-1}} f_{i_1} \circ f_{i_2} \circ \dots \circ f_{i_{(q+t)k}}(f_1^k(I_0^m))$$

因为 $(i_{qk+1}, i_{qk+2}, \dots, i_{(q+1)k}) \neq (1, \dots, 1)$ 则有 $B_q \cap B_{q+t} = \emptyset$, 注意到 $B_{q'} \subset B_{q+t}$, 所以 $B_q \subset B_{q'} = \emptyset$ 。

对于 (ii), 由 B_q ($q \geq 0$) 的定义易知 B_0 仅包含 $I_k^m(\lambda)$ 中的一个基本立方体, B_1 含有 $I_{2k}^m(\lambda)$ 中的 $2^{km} - 1$ 个基本立方体, B_2 含有 $I_{3k}^m(\lambda)$ 中的 $(2^{km} - 1)(2^{km} - 1)$ 个基本立方体. 利用归纳法得到, B_q 包含有 $I_{(q+1)k}^m(\lambda)$ 中的 $(2^{km-1} - 1)^q$ 个基本立方体. 因此:

$$\mu(B_q) = (2^{km} - 1)^q \cdot \frac{1}{2^{(q+1)km}} = \left(1 - \frac{1}{2^{km}}\right)^q \cdot \frac{1}{2^{km}}$$

所以 $\mu(B_q) = \left(1 - \frac{1}{2^{km}}\right)^q \cdot \frac{1}{2^{km}}$, 从而 (ii) 是成立的.

从 (i) (ii) 得到 $\mu\left(\bigcup_{q=0}^{\infty} B_q\right) = \sum_{q=0}^{\infty} \left(1 - \frac{1}{2^{km}}\right)^q \cdot \frac{1}{2^{km}} = 1$, 所以 $\mu(A) = 1$, 即 $\mu(F_{q,k}) = 1$. 对任何 $k \geq 1$ 及 $p \geq 1$ 成立. 注意到集合序列 $\{F_{p,k}\}_{p \geq 1}$ 是递减的, 由此得出对任何 $k \geq 1$, 有 $\mu\left(\bigcap_{p \geq 1} F_{p,k}\right) = \lim_{k \rightarrow \infty} \mu(F_{p,k}) = 1$, 进一步可以看出集合序列 $\left\{\bigcap_{p \geq 1} F_{p,k}\right\}_{k \geq 1}$ 仍然是递减的, 所以 $\mu\left(\bigcap_{k \geq 1} \left(\bigcap_{p \geq 1} F_{p,k}\right)\right) = \lim_{k \rightarrow \infty} \mu\left(\bigcap_{p \geq 1} F_{p,k}\right) = 1$. 从而引理 2.5 得证.

三、定理的证明

引理 3.1 对于 $n \geq 0$, 令 $V_n(\lambda)$ 为基本立方体 $I_n^m(\lambda)$ 的所有顶点构成的集合, 则对任何 $x \in V_n(\lambda)$, 有 $\underline{D}^{s(\lambda)}(\mu, x) = \frac{1}{2^m \cdot 2^{s(\lambda)} (1 - \lambda)^{s(\lambda)}}$.

证 容易看出 x 是边长为 λ^n 的基本立方体的一个顶点, 则当取 $r = \lambda^n - \lambda^{n+1}$ 时, 球 $B(x, r)$ 仅包含 $I_{n+1}^m(\lambda)$ 的一个基本立方体, 因此:

$$\frac{\mu(B(x, r))}{(2r)^{s(\lambda)}} = \frac{2^{-(n+1)m}}{(2 \cdot \lambda^n - 2 \cdot \lambda^{n+1})^{s(\lambda)}} = \frac{1}{2^m \cdot 2^{s(\lambda)} (1 - \lambda)^{s(\lambda)}}$$

注意到上述推导用到了 $\lambda^{s(\lambda)} = \frac{1}{2^m}$, 并且当 $n \rightarrow \infty$ 时有 $r \rightarrow 0$, 所以:

$$\underline{D}^{s(\lambda)}(\mu, x) \leq \frac{1}{2^m \cdot 2^{s(\lambda)} (1 - \lambda)^{s(\lambda)}}$$

为证明相反的不等式也成立, 由于 $S^{m(\lambda)}$ 是一个自相似集, 只需对 $x \in V_0(\lambda)$ 及 $\sqrt{m} < r \leq \sqrt{m}$ 的情形进行证明即可. 将点 x 取为坐标原点, I_0^m 的从 x 出发的 m 条边为坐标轴.

(1) 如果 $\lambda \sqrt{m} < r \leq 1 - \lambda$, 球 $B(x, r)$ 至少包含 $I_1^m(\lambda)$ 的一个基本立方体, 则:

$$\frac{\mu(B(x, r))}{(2r)^{s(\lambda)}} = \frac{2^{-m}}{2^{s(\lambda)} (1 - \lambda)^{s(\lambda)}} = \frac{1}{2^m \cdot 2^{s(\lambda)} (1 - \lambda)^{s(\lambda)}}$$

(2) 如果 $1 - \lambda < r \leq \sqrt{1 + (m-1)\lambda^2}$, 则存在整数 $n \geq 1$, 使得:

$$\sqrt{(1 - \lambda + \lambda^{n+1})^2 + (m-1)\lambda^{2(n+1)}} < r \leq \sqrt{(1 - \lambda + \lambda^n)^2 + (m-1)\lambda^{2n}}$$

令 $P_{1,j} = (x_1, x_2, \dots, x_m)$, $1 \leq j \leq m$ 其中 $x_j = 1 - \lambda$ 且当 $i \neq j$ 时, $x_i = 0$, $1 \leq i \leq m$. 则 $P_{1,1}, P_{1,2}, \dots, P_{1,m}$ 分别是 $I_1^m(\lambda)$ 中 m 个基本立方体的 m 个顶点, 并且点 $P_{1,j}$ 位于坐标轴

X_j 上。令 $\Theta_{n+1,j} (1 \leq j \leq m)$ 表示 $I_{n+1}^m(\lambda)$ 中以点 $P_{1,j}$ 作为其顶点之一的基本立方体，点 $Q_{n+1,j} = (y_1, y_2, \dots, y_m)$ 表示 $\Theta_{n+1,j}$ 的距离原点最远的顶点，则 $y_j = 1 - \lambda + \lambda^{n+1}$ ，并且当 $i \neq j$ 时，有 $y_i = \lambda^{n+1}$ 。因此：

$$|OQ_{n+1,j}| = \dots = |OQ_{n+1,m}| = \sqrt{(1 - \lambda + \lambda^{n+1})^2 + (m - 1)\lambda^{2(n+1)}}$$

这表明球 $B(x, r)$ 包含 $I_1^m(\lambda)$ 的一个基本立方体和至少 $I_{n+1}^m(\lambda)$ 的 m 个基本立方体，故：

$$\begin{aligned} & \frac{\mu(B(x, r))}{(2r)^{s(\lambda)}} \\ & \geq \frac{\frac{1}{2^m} + \frac{m}{2^{m(n+1)}}}{[2\sqrt{(1 - \lambda + \lambda^{n+1})^2 + (m - 1)\lambda^{2n}}]^{s(\lambda)}} \\ & = \frac{1 + \frac{m}{2^{nm}}}{2^m \cdot 2^{s(\lambda)} (1 - \lambda)^{s(\lambda)} \sqrt{[1 + 2 \cdot \frac{\lambda^n}{1 - \lambda} + m \cdot (\frac{\lambda^n}{1 - \lambda})^2]^{s(\lambda)}}} \\ & \geq \frac{1}{2^m \cdot 2^{s(\lambda)} (1 - \lambda)^{s(\lambda)}} \cdot \frac{\sqrt{1 + \frac{2m}{2^{nm}} + \frac{m^2}{2^{2nm}}}}{\sqrt{[1 + 2 \cdot \frac{2^{-nm}}{1 - 2^{-m}} + m \cdot (\frac{2^{-nm}}{1 - 2^{-m}})^2]^{s(\lambda)}}}} \quad (0 < \lambda \leq \frac{1}{2^m}) \\ & \geq \frac{1}{2^m \cdot 2^{s(\lambda)} (1 - \lambda)^{s(\lambda)}} \cdot \frac{\sqrt{1 + \frac{2m}{2^{nm}} + \frac{m^2}{2^{2nm}}}}{\sqrt{[1 + 2 \cdot \frac{2^m}{2^{nm} (2^m - 1)} + m \cdot (\frac{2^m}{2^{nm} (2^m - 1)})^2]^{s(\lambda)}}}} \quad (0 < s(\lambda) \leq 1) \\ & \geq \frac{1}{2^m \cdot 2^{s(\lambda)} (1 - \lambda)^{s(\lambda)}} \cdot \frac{\sqrt{1 + 2 \cdot \frac{m}{2^{nm}} + \frac{m^2}{2^{2nm}}}}{\sqrt{[1 + 2 \cdot \frac{2m}{2^{nm} (2^m - 1)} + m \cdot (\frac{1}{2^{nm}})^2 \cdot \frac{2^{2m}}{(2^m - 1)^2}]^{s(\lambda)}}}} \end{aligned}$$

注意到当 $m > 1$ 时，显然有 $\frac{2^m}{2^m - 1} \leq m$ 及 $\frac{2^{2m}}{(2^m - 1)^2} \leq m$ ，所以：

$$\frac{\mu(B(x, r))}{(2r)^{s(\lambda)}} \geq \frac{1}{2^m \cdot 2^{s(\lambda)} (1 - \lambda)^{s(\lambda)}} \cdot \frac{\sqrt{1 + \frac{2 \cdot m}{2^{nm}} + \frac{m^2}{2^{2nm}}}}{\sqrt{1 + \frac{2 \cdot m}{2^{nm}} + \frac{m \cdot m}{2^{2nm}}}} = \frac{1}{2^m \cdot 2^{s(\lambda)} (1 - \lambda)^{s(\lambda)}}$$

对所有 $n \geq 1$ 成立。

(3) 如果 $\sqrt{1 + (m - 1) \cdot \lambda^2} < r \leq \sqrt{m}$ ，则从上述(2)的分析中得到，立方体 $\Theta_{1,j} (1 \leq j \leq m)$ 包含于球 $B(x, r)$ 中。此外，球 $B(x, r)$ 还包含有 $I_1^m(\lambda)$ 的位于原点处的基本立方体，所以：

$$\frac{\mu(B(x, r))}{(2r)^{s(\lambda)}} \geq \frac{\frac{m+1}{2^m}}{2^{s(\lambda)} \cdot \sqrt{m}^{s(\lambda)}} \geq \frac{m+1}{2^m \cdot 2^{s(\lambda)} \sqrt{m}^{s(\lambda)}}$$

$$\begin{aligned}
&= \frac{1}{2^m \cdot 2^{s(\lambda)}} \frac{1}{(1-\lambda)^{s(\lambda)}} \cdot \frac{(m+1)(1-\lambda)^{s(\lambda)}}{\sqrt{m^{s(\lambda)}}} \\
&\geq \frac{1}{2^m \cdot 2^{s(\lambda)}} \frac{1}{(1-\lambda)^{s(\lambda)}} \cdot \frac{m \cdot (1 - \frac{1}{2^m})^{s(\lambda)}}{\sqrt{m^{2(\lambda)}}} \quad (1 < \lambda \leq \frac{1}{2^m}) \\
&\geq \frac{1}{2^m \cdot 2^{s(\lambda)}} \frac{1}{(1-\lambda)^{s(\lambda)}} \cdot \left[\frac{m(1 - \frac{1}{2^m})}{\sqrt{m}} \right]^{s(\lambda)} \quad (1 < s(\lambda) \leq 1) \\
&= \frac{1}{2^m \cdot 2^{s(\lambda)}} \frac{1}{(1-\lambda)^{s(\lambda)}} \cdot \left[(2^m - 1) \frac{\sqrt{m}}{2^m} \right]^{s(\lambda)} \geq \frac{1}{2^m \cdot 2^{s(\lambda)}} \frac{1}{(1-\lambda)^{s(\lambda)}}
\end{aligned}$$

(注意当 $m > 1$ 时, 有 $2^m - 1 \geq \frac{2^m}{\sqrt{m}}$)。由此便证明了引理 3.1。

引理 3.2 μ 是 (2.5) 定义的概率测度, 则对 μ -几乎所有的 $x \in S^m(\lambda)$, 有:

$$\underline{D}^{s(\lambda)}(\mu, x) = \frac{1}{2^m \cdot 2^{s(\lambda)} (1-\lambda)^{s(\lambda)}}$$

证 根据引理 2.5 只需对 $x \in S^m(\lambda) \cap (\bigcap_{k \geq 1} (\bigcap_{p \geq 1} F_{p,k}))$ 进行证明即可, 其中 $F_{p,k}$ 由 (2.6) 所定义。因此对任何 $p \geq 1$ 及 $k \geq 1$, 有 $x \in F_{p,k}$ 。由 $F_{p,k}$ 的定义, 对任何 $k \geq 1$, 存在 $n_p \geq p$, 使得存在一个 $y_{n_p} \in V_{n_p}$, 满足 $\text{dist}(x, y_{n_p}) \leq \lambda^{n_p+k} \sqrt{m}$ 。取 $r_p = \lambda^{n_p} - \lambda^{n_p+1} - \lambda^{n_p+k} \sqrt{m}$ (注意到由于 $0 < \lambda \leq 2^{-m}$, 所以 $r_p > 0$) 则有 $B(x, r_p) \subset B(y_{n_p}, \lambda^{n_p} - \lambda^{n_p+1})$, 这表明:

$$\begin{aligned}
\frac{\mu(B(x, r_p))}{(2r_p)^{s(\lambda)}} &\leq \frac{\mu(B(y_{n_p}, \lambda^{n_p} - \lambda^{n_p+1}))}{(2r_p)^{s(\lambda)}} \\
&= \frac{2^{-(n_p+1)m}}{2^{s(\lambda)} (\lambda^{n_p} - \lambda^{n_p+1} - \lambda^{n_p+k} \sqrt{m})^{s(\lambda)}} = \frac{1}{2^m \cdot 2^{s(\lambda)} (1-\lambda - \lambda^k \sqrt{m})^{s(\lambda)}}
\end{aligned}$$

注意到 $x \in S^m(\lambda) \cap (\bigcap_{k \geq 1} (\bigcap_{p \geq 1} F_{p,k}))$, 并且当 $p \rightarrow \infty$ 时, $r_p \rightarrow 0$, 得到对任何 $k \geq 1$, 有:

$$\underline{D}^{s(\lambda)}(\mu, x) \leq \frac{1}{2^m \cdot 2^{s(\lambda)} (1-\lambda - \lambda^k \sqrt{m})^{s(\lambda)}}$$

由于 $x \in S^m(\lambda) \cap (\bigcap_{p \geq 1} (\bigcap_{k \geq 1} F_{p,k}))$, 利用引理 2.5, 令 $k \rightarrow \infty$, 对 μ -几乎所有 $x \in S^m(\lambda)$, 有:

$$\underline{D}^{s(\lambda)}(\mu, x) \leq \frac{1}{2^m \cdot 2^{s(\lambda)} (1-\lambda)^{s(\lambda)}}$$

下面证明相反的不等式成立, 由于集合 $S^m(\lambda)$ 为自相似集, 所以只需对 $\lambda \sqrt{m} < r \leq \sqrt{m}$ 进行考虑即可。点 x 显然属于 $I_1^m(\lambda)$ 的某一个基本立方体中, 记该立方体为 Θ_1^0 。

(1) 如果 $\lambda \sqrt{m} < r \leq 1 - \lambda$, 由于 $0 < \lambda \leq 2^{-m}$, 球 $B(x, r)$ 至少应包含基本立方体 Θ_1^0 。因此:

$$\frac{\mu(B(x, r))}{(2r)^{s(\lambda)}} \geq \frac{2^{-m}}{2^{s(\lambda)} \cdot (1-\lambda)^{s(\lambda)}} = \frac{1}{2^m \cdot 2^{s(\lambda)} (1-\lambda)^{s(\lambda)}}$$

(2) 如果 $1 - \lambda < r \leq \sqrt{1 + (m-1)\lambda^2}$, 则存在整数 $n \geq 1$, 使得:

$$\sqrt{(1 - \lambda + \lambda^{n+1})^2 + (m - 1)\lambda^{2(n+1)}} < r \leq \sqrt{(1 - \lambda + \lambda^n)^2 + (m - 1)\lambda^{2n}}$$

由于 $0 < \lambda \leq 2^{-m}$ 球 $B(x, r)$ 显然包含立方体 Θ_1^0 。下面分析除了立方体 Θ_1^0 外, 还有多少个 $I_{n+1}^m(\lambda)$ 的基本立方体包含于球 $B(x, r)$ 中。

为方便起见, 先给出一些记号。注意到 Θ_1^0 的一个顶点与 I_0^m 的一个顶点重合, 我们选定该顶点作为坐标原点 O , 坐标轴即为 I_0^m 的 m 条边, 记这些坐标轴为 X_1, X_2, \dots, X_m 。对于 $x \in R^m$, 定义函数 $h_j(x) = x + B_j (j = 1, 2, \dots, m)$, 其中向量 $B_j = (b_1, b_2, \dots, b_m)$ 满足 $b_j = 1 - \lambda$, 而当 $i \neq j$ 时, $b_i = 0$ 。

设整数 $k \geq 1$, 假定 Θ_k^0 是 $I_k^m(\lambda)$ 的位于原点处的基本立方体(即 Θ_k^0 的一个顶点就是坐标原点)。立方体 Θ_k^0 包含有 $I_{k+1}^m(\lambda)$ 的 2^m 个基本立方体, 记这 2^m 个立方体为 $\Theta_{k+1}^0, \Theta_{k+1}^1, \Theta_{k+1}^2, \dots, \Theta_{k+1}^{2^m-1}$ 。不妨设 Θ_{k+1}^0 是位于原点处的基本立方体, 则余下的 $2^m - 1$ 的立方体 $\Theta_{k+1}^1, \Theta_{k+1}^2, \dots, \Theta_{k+1}^{2^m-1}$ 将位于 $(m - 1)$ - 维超平面 $x_i = \lambda^k - \lambda^{k+1}$ 与 $x_i = \lambda^k (i = 1, 2, \dots, m)$ 之间。

下面利用归纳法进行讨论。

(i) 如果 $x \in \Theta_{n+1}^0$, 令 $\Omega_{n+1}^j = h_j(\Theta_{n+1}^0) (j = 1, 2, \dots, m)$, 则 $\Omega_{n+1}^1, \Omega_{n+1}^2, \dots, \Omega_{n+1}^m$ 为 $I_{n+1}^m(\lambda)$ 的 m 个基本立方体, 并且易知这些立方体包含于球 $B(x, r)$ 中。

(ii) 如果 $x \in \Theta_n^0 / \Theta_{n+1}^0$, 则 x 必定属于基本立方体 $\Theta_{n+1}^1, \Theta_{n+1}^2, \dots, \Theta_{n+1}^{2^m-1}$ 其中的某一个。不失一般性, 仅考虑当 x 属于位于超平面 $x_i = \lambda^n - \lambda^{n+1}$ 与 $x_i = \lambda^n$ 之间的某一基本立方体的情形。假设 $\Theta_{n+1}^1, \Theta_{n+1}^2, \dots, \Theta_{n+1}^{2^m-1}$ 是位于超平面 $x_i = \lambda^n - \lambda^{n+1}$ 与 $x_i = \lambda^n$ 之间的 2^{m-1} 个基本立方体, 而且 Θ_{n+1}^1 有一条边与坐标轴 X_1 重合。令 $\Omega_n^1 = h_1(\Theta_n^0)$, 将证明球 $B(x, r)$ 至少包含 $\Omega_n^1 \cap I_{n+1}^m(\lambda)$ 中的 m 个基本立方体。利用集合 $(\bigcup_{j=1}^{2^{m-1}} \Theta_{n+1}^j) \cup \Omega_n^1 \cap I_{n+1}^m(\lambda)$ 的对称性, 仅需对 $x \in \Theta_{n+1}^1$ 进行考虑即可。

设 M 为 Θ_n^0 的所有顶点构成的集合: $M = \{(x_1, x_2, \dots, x_m) : x_i = 0 \text{ 或 } \lambda^n, 1 \leq i \leq m\}$ 。现将 M 分成 $m + 1$ 个不相交的子集, $M = \bigcup_{k=0}^m M_k$, 其中 $M_k = \{(x_1, x_2, \dots, x_m) \in M : k = \text{Card}\{i : x_i = \lambda^n\}\}$ 。令 P_1, P_2, \dots, P_m 是集合 M_1 中的 m 个顶点, 并且 Θ_{n+1}^i 是以 $P_i (i = 1, 2, \dots, m)$ 作为其顶点之一的基本立方体。设其中的 Θ_{n+1}^1 是边与坐标轴 X_1 重合的立方体。令 $\Omega_{n+1}^i = h_i(\Theta_{n+1}^1) (i = 0, 1, 2, \dots, m)$, 则 $\Omega_{n+1}^0, \Omega_{n+1}^1, \dots, \Omega_{n+1}^m$ 就是 $\Omega_n^1 \cap I_{n+1}^m(\lambda)$ 中的 $m + 1$ 个基本立方体。设 P_{n+1}^1 是 Θ_{n+1}^1 的距离原点最近的顶点, 而 $Q_{n+1}^0, Q_{n+1}^1, \dots, Q_{n+1}^m$ 分别是 $\Omega_{n+1}^0, \Omega_{n+1}^1, \dots, \Omega_{n+1}^m$ 中的距离原点最远的顶点。则:

$$|P_{n+1}^1 Q_{n+1}^0| = \sqrt{(1 - \lambda + \lambda^{n+1} - \lambda^n + \lambda^{n+1})^2 + (m - 1)\lambda^{2(n+1)}} < r$$

以及

$$\begin{aligned} |P_{n+1}^1 Q_{n+1}^2| &= |P_{n+1}^1 Q_{n+1}^3| = \dots = |P_{n+1}^1 Q_{n+1}^m| \\ &= \sqrt{(1 - \lambda - \lambda^n + 2\lambda^{n+1})^2 + \lambda^{2n} + (m - 2)\lambda^{2(n+1)}} \\ &= \sqrt{(1 - \lambda + \lambda^{n+1})^2 + (m - 1)\lambda^{2(n+1)} + \eta_{n+1}} \end{aligned}$$

其中:

$$\begin{aligned} \eta_{n+1} &= (\lambda^n - \lambda^{n+1})^2 - 2(1 - \lambda + \lambda^{n+1})(\lambda^n - \lambda^{n+1}) + \lambda^{2n} - \lambda^{2(n+1)} \\ &= 2\lambda^n(2\lambda + \lambda^n + \lambda^{n+2} - 1 - \lambda^2 - 2\lambda^{n+1}) \end{aligned}$$

由于 $0 < \lambda \leq 2^{-m}$, $m \geq 2$, $n \geq 1$, 则可得出 $\eta_{n+1} < 0$, 所以:

$$|P_{n+1}^1 Q_{n+1}^2| = |P_{n+1}^1 Q_{n+1}^3| = \dots = |P_{n+1}^1 Q_{n+1}^m| < r$$

上述讨论表明立方体 $\Omega_{n+1}^0, \Omega_{n+1}^2, \Omega_{n+1}^3, \dots, \Omega_{n+1}^m$ 均包含于球 $B(x, r)$ 中。也就是说 球 $B(x, r)$ 至少包含 $I_{n+1}^m(\lambda)$ 中的 m 个基本立方体。

(iii) 一般地, 如果 $x \in \Theta_k^0 / \Theta_{k+1}^0 (1 \leq k \leq n - 1)$ 则 x 必定属于立方体 $\Theta_{k+1}^1, \Theta_{k+1}^2, \dots, \Theta_{k+1}^{m-1}$ 其中之一。不失一般性, 仅需考虑当 x 属于超平面 $x_1 = \lambda^k - \lambda^{k+1}$ 与 $x_1 = \lambda^k$ 之间的某一立方体的情形。假设 $\Theta_{k+1}^1, \Theta_{k+1}^{2k+1}, \dots, \Theta_{k+1}^{2^{m-1}}$ 是位于超平面 $x_1 = \lambda^k - \lambda^{k+1}$ 与 $x_1 = \lambda^k$ 之间的 2^{m-1} 个基本立方体, 并且立方体 Θ_{k+1}^1 的一条边与坐标轴 x_1 重合。令 $\Omega_k^1 = h_1(\Theta_k^0)$ 将证明球

$B(x, r)$ 至少包含 $\Omega_k^1 \cap I_{k+1}^m(\lambda)$ 中的 m 个基本立方体。利用集合 $(\bigcup_{j=1}^{2^{m-1}} \Theta_{k+1}^j) \cup \Omega_k^1 \cap I_{k+1}^m(\lambda)$ 的对称性, 仅需对 $x \in \Theta_{k+1}^1$ 的情形进行讨论即可。

注意到立方体 $\Omega_{k+1}^1 = h_1(\Theta_{k+1}^0)$ 应包含于 Ω_k^1 中。令 P_{k+1}^1 是 Θ_{k+1}^1 的距离原点最近的顶点, 而 Q_{k+1}^1 是 Ω_{k+1}^1 的距原点最远的顶点, 则集合 $\Theta_{k+1}^1 \cup \Omega_{k+1}^1$ 的直径为 $|P_{k+1}^1 Q_{k+1}^1|$, 有:

$$\begin{aligned} |P_{k+1}^1 Q_{k+1}^1| &= \sqrt{(1 - \lambda - \lambda^k + 2\lambda^{k+1})^2 + (m - 1)\lambda^{2(k+1)}} \\ &= \sqrt{(1 - \lambda + \lambda^{n+1})^2 + (m - 1)\lambda^{2(n+1)} + \eta_{k+1}} \end{aligned}$$

其中:

$$\begin{aligned} \eta_{k+1} &= (\lambda^k - 2\lambda^{k+1} + \lambda^{n+1})^2 - 2(1 - \lambda + \lambda^{n+1})(\lambda^k - 2\lambda^{k+1} + \lambda^{n+1}) \\ &\quad + (m - 1)(\lambda^{2(k+1)} - \lambda^{2(n+1)}) \\ &= \lambda^{2k} + 6\lambda^{k+1} + 3\lambda^{2(k+1)} + 2\lambda^{n+2} + m\lambda^{2(k+1)} - 2\lambda^k - 4\lambda^{2k+1} - 4\lambda^{k+2} \\ &\quad - 2\lambda^{n+1} - m\lambda^{2(n+1)} \end{aligned}$$

由于 $0 < \lambda \leq 2^{-m}$, $m \geq 2$, $n \geq 1$, 所以有:

$$\eta_{k+1} < [(3 + m)\lambda^{2(k+1)} - 4\lambda^{2k+1}] + (2\lambda^{n+2} - 2\lambda^{n+1}) + \lambda^k(\lambda^k + 6\lambda - 2) < 0$$

因此 $|P_{k+1}^1 Q_{k+1}^1| < r$ 。这表明球 $B(x, r)$ 包含立方体 Θ_{k+1}^1 。从而 $B(x, r)$ 包含 $I_{k+2}^m(\lambda)$ 的 2^m 个基本立方体。

上述表明 球 $B(x, r)$ 包含立方体 Θ_1^0 以及至少 m 个 $I_{n+1}^m(\lambda)$ 的基本立方体, 则:

$$\frac{\mu(B(x, r))}{(2r)^{s(\lambda)}} \geq \frac{\frac{1}{2^m} + \frac{m}{2^{(n+1)m}}}{2^{s(\lambda)} \sqrt{[(1 - \lambda + \lambda^n)^2 + (m - 1)\lambda^{2n}]^{s(\lambda)}}}$$

余下的讨论与引理 3.1 证明中的 (2) 相同, 所以:

$$\frac{\mu(B(x, r))}{(2r)^{s(\lambda)}} \frac{1}{2^m \cdot 2^{s(\lambda)}(1 - \lambda)^{s(\lambda)}}$$

(3) 如果 $\sqrt{1 + (m - 1)\lambda^2} < r \leq \sqrt{m}$ 。容易看出立方体 Θ_j^0 与 $h_j(\Theta_j^0) (j = 1, 2, \dots, m)$ 均包含于球 $B(x, r)$ 中, 则球 $B(x, r)$ 至少包含 $I_1^m(\lambda)$ 的 $m + 1$ 个基本立方体, 所以:

$$\frac{\mu(B(x, r))}{(2r)^{s(\lambda)}} \geq \frac{\frac{m+1}{2^m}}{2^{s(\lambda)} \cdot \sqrt{m^{s(\lambda)}}}$$

余下的证明与引理 3.1 证明中的 (3) 相同, 从而:

$$\frac{\mu(B(x, r))}{(2r)^{s(\lambda)}} \frac{1}{2^m \cdot 2^{s(\lambda)} (1-\lambda)^{s(\lambda)}}$$

引理 3.2 得证。

定理的证明 从引理 2.4 及引理 3.2 立得。

致谢 感谢文志英教授、瞿成勤博士和贾保国博士为本文提出很好的建议。

参考文献

- [1] Falconer, K. J., The Geometry of Fractal Sets, Cambridge University Press, 1986.
- [2] Falconer, K. J., Fractal Geometry-Mathematical Foundations and Applications, John and Sons, New York, 1990.
- [3] Falconer, K. J., Techniques in Fractal Geometry, John Wiley and Sons Ltd., 1997.
- [4] Taylor, J. & Tricot, C., The packing measure of rectifiable subsets of the plane, Math. Camb. Phil. Soc, 1986, Vol.99, pp285 ~ 296
- [5] 周作领, 吴敏: 《一个 Sierpinski 地毯的 Hausdorff 测度》, 《中国科学 A 辑》, 1999 年第 2 期, 第 138 ~ 144 页。
- [6] Zhou Zuoling, Wu Min & Zhao Yanfen, The Hausdorff measure of a class of generalized Sierpinski sponges, Chinese Journal of Contemporary Mathematics, 2001, Vol.22, No.1, pp55 ~ 64; 《数学年刊》, 2001 年第 1 期, 第 57 ~ 64 页。

三分 Cantor 集自乘积的 Hausdorff 测度的估计*

贾保国 周作领 朱智伟

一、引言与定理

计算与估计分形集的 Hausdorff 维数与测度是分形几何研究的重要内容之一。一般地说,计算分形集合的 Hausdorff 维数,特别是计算 Hausdorff 测度是非常困难的。在满足开集条件下,自相似集的 Hausdorff 维数已被完全确定(参见文 [1])。但就在这种情形, Hausdorff 测度的计算与估计,除直线上 Cantor 集及其种种变形之外,也几乎无任何结果(参见文 [2~4])。文 [5] 首先得到维数为 1 的一种 Sierpinski 地毯的 Hausdorff 测度的准确值,文 [6] 又把这个结果推广到维数不大于 1 的情形。到目前为止,尚无任何维数大于 1 的分形的 Hausdorff 的测度被确定(含给出计算表达式)。本文研究了三分 Cantor 集 C 的自乘积集 $C \times C$ (维数大于 1)的 Hausdorff 测度,得到了 $H^s(C \times C)$ 的一个较好的估计。

在平面 R^2 上取单位正方形 E_0 ,在 E_0 上除了保留四个角上边长为 $\frac{1}{3}$ 的小正方形外删去其余部分的内部,得到四个边长为 $\frac{1}{3}$ 的小正方形的集合记为 E_1 ,对 E_1 中的每个正方形重复上述过程得到的集合记为 E_2 ,无限重复上述过程,得到 $E_0 \supset E_1 \supset \dots \supset E_n \supset \dots$ 。非空集合 $C \times C = \bigcap_{n=0} E_n$,即为 Cantor 集 C 的自乘积集,它是由四个相似比为 $\frac{1}{3}$ 的相似压缩函数生成的自相似集且满足开集条件。因此, $C \times C$ 的 Hausdorff 维数 $s = \dim_H(C \times C)$ 满足 $4 \times (\frac{1}{3})^s = 1$ 。即 $1 \leq s = \log_3 4 \leq 2$ 。本文的主要结果是:

定理 三分 Cantor 集的自乘积集 $C \times C$ 的 Hausdorff 测度满足不等式:

$$1 \leq H_{\log_3 4}(C \times C) \leq 1.502879.$$

二、相关的命题及引理

为了证明本文的定理,我们叙述两个命题和引理。

命题 1 设 $s = \log_3 4$, 则:

* 基金项目:国家自然科学基金资助项目(10041005)、国家教育部博士点资助项目(1999055810)。原载:《数学学报》,2003年第4期。

(1) 当 $0 \leq x \leq 1$ 时, $(1-x)^s + x^s \geq \frac{2}{2^s}$ 。

(2) 对任何 $x_1, x_2, x_3, x_4 \in (0, 1)$, 有:

$$\left(\frac{x_1 + x_2}{2}\right)^s = \frac{1}{2}(x_1^s + x_2^s) \cdot \left(\frac{x_1 + x_2 + x_3 + x_4}{4}\right)^s \leq \frac{x_1^s + x_2^s + x_3^s + x_4^s}{4}$$

证明 (1) 令 $f(x) = (1-x)^s + x^s$, $f'(x) = -s(1-x)^{s-1} + sx^{s-1}$, $f''(x) = s(s-1)(1-x)^{s-2} + s(s-1)x^{s-2} \geq 0$ 。因此 $f(x)$ 在 $[0, 1]$ 上是凹函数, 且当 $x = \frac{1}{2}$ 时, $f(x)$ 取最小值 $f\left(\frac{1}{2}\right) = \frac{2}{2^s}$, 从而结论 (1) 成立。

由于 $g(x) = x^s$ 在 $(0, 1)$ 上是凹函数, 从而结论 (2) 是显然的。证毕。

命题 2^[4] 设 $F = C \times C$, $U \subset R^n$ 为可测集合且 $|U| > 0$, 则 $H^s(F \cap U) \leq |U|^s$ 。

E_n 由 4^n 个边长为 $\frac{1}{3^n}$ 的正方形组成, 每个这样的正方形称为基本正方形, 记为 Ω_n 。

在 E_n 上定义分布函数 μ , 满足:

$$\begin{cases} \mu(E_0) = 2^{\frac{s}{2}} \\ \mu(E_n) = \frac{1}{4^n} \times 2^{\frac{s}{2}}, n = 1, 2, 3, \dots \\ \mu(E_0 - C \times C) = 0 \end{cases}$$

则 μ 为 E_0 上一个测度且 μ 为 $C \times C$ 上的一个质量分布。

如图 1 建立直角坐标系, E_0 的主和斜对角线方程分别为 $y = x, y + x = 1$ 。设 G 为平行于 E_0 的斜对角线且与 E_0 相交的直线, 记原点到 G 的距离 $d((0, 0), G) = g \geq 0$ 。记 G 与 x 轴, y 轴围成的三角形为 Δ_g 。

引理 1 $\mu \Delta_g \geq \frac{1}{2} \cdot \frac{9}{4^s} g^s, 0 \leq g \leq \frac{\sqrt{2}}{3}$ 。

证明 设 $0 < g \leq \frac{\sqrt{2}}{3}$, 即存在正整数 n , 使 $\frac{\sqrt{2}}{3^{n+1}} < g \leq \frac{\sqrt{2}}{3^n}$ 。如图 2, 考虑如下 6 种情况:

(1) 若 $|OA_1| < g \leq |OA_2|$, 即 $\frac{\sqrt{2}}{3^{n+1}} < g \leq \frac{\sqrt{2}}{3^{n+1}} + \frac{\sqrt{2}}{3^{n+3}} = \frac{\sqrt{2}}{3^{n+1}} \cdot \frac{10}{9}$, 则 $g^s \leq \frac{2^{\frac{s}{2}}}{4^{n+1}} \cdot \frac{10^s}{9^s}$,

且存在 $\Omega_{n+1} \in E_{n+1}$ 使 $\Delta_g \supset \Omega_{n+1}$, 从而有:

$$\mu(\Delta_g) \geq \frac{2^{\frac{s}{2}}}{4^{n+1}} \geq \frac{10^s}{9^s} \cdot g^s = \frac{16}{10^s} \cdot g^s \geq \frac{1}{2} \cdot \frac{9}{4^s} \cdot g^s$$

(2) 若 $|OA_2| < g \leq |OA_3|$, 即 $\frac{\sqrt{2}}{3^{n+1}} \cdot \frac{10}{9} \leq g \leq \frac{\sqrt{2}}{3^{n+1}} + \frac{\sqrt{2}}{3^{n+2}} = \frac{\sqrt{2}}{3^{n+1}} \cdot \frac{4}{3}$, 则 $g^s \leq \frac{4^s}{4} \cdot \frac{2^{\frac{s}{2}}}{4^{n+1}}$,

且存在 $\Omega_{n+3}^{(1)}$ 和 $\Omega_{n+3}^{(2)}$ 在 E_{n+3} 中, 使 $\Delta_g \supset \Omega_{n+1} \cup \Omega_{n+3}^{(1)} \cup \Omega_{n+3}^{(2)}$, 从而有:

$$\mu(\Delta_g) \geq \frac{2^{\frac{s}{2}}}{4^{n+1}} + \frac{2 \cdot 2^{\frac{s}{2}}}{4^{n+3}} = \frac{2^{\frac{s}{2}}}{4^{n+1}} \cdot \frac{9}{8} \geq \frac{9}{8} \cdot \frac{4}{4^s} \cdot g^s \geq \frac{1}{2} \cdot \frac{9}{4^s} \cdot g^s$$

(3) 若 $|OA_3| < g \leq |OA_4|$, 即 $\frac{\sqrt{2}}{3^{n+1}} \cdot \frac{4}{3} < g \leq \frac{1}{2} \cdot \frac{\sqrt{2}}{3^n}$, 则 $g^s \leq \frac{1}{2^s} \cdot \frac{2^{\frac{s}{2}}}{4^n} = \frac{4}{2^s} \cdot \frac{2^{\frac{s}{2}}}{4^{n+1}}$, 且

存在 $\Omega_{n+2}^{(1)}$ 和 $\Omega_{n+2}^{(2)}$ 在 E_{n+3} 中, 使 $\Delta_g \supset \Omega_{n+1} \cup \Omega_{n+2}^{(1)} \cup \Omega_{n+2}^{(2)}$, 从而有:

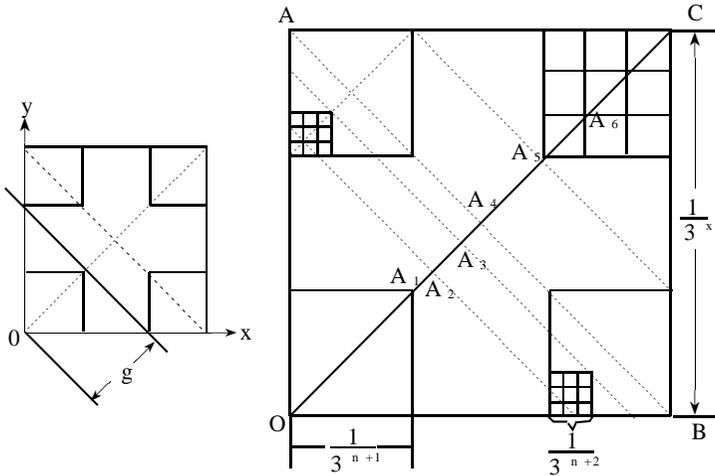


图 1

图 2

$$\mu(\Delta_g) \geq \frac{2^{\frac{s}{2}}}{4^{n+1}} + \frac{2 \cdot 2^{\frac{s}{2}}}{4^{n+2}} = \frac{2^{\frac{s}{2}}}{4^{n+1}} \cdot \frac{3}{2} \geq \frac{3}{2} \cdot \frac{2^s}{4} \cdot g^s \geq \frac{1}{2} \cdot \frac{9}{4^s} \cdot g^s$$

(4) 若 $|OA_4| < g \leq |OA_5|$, 即 $\frac{1}{2} \cdot \frac{\sqrt{2}}{3^n} < g \leq \frac{1}{2} \cdot \frac{\sqrt{2}}{3^n} + \frac{1}{2} \cdot \frac{\sqrt{2}}{3^{n+1}} = \frac{\sqrt{2}}{3^n} \cdot \frac{2}{3}$, 则 $g^s \leq \frac{2^{\frac{s}{2}}}{4^{n+1}} \cdot 2^s$, 且 Δ_g 包含 E_0 的斜对角线 $y + x = 1$ 以下的三角形 OAB , 从而有:

$$\mu(\Delta_g) \geq \frac{2 \cdot 2^{\frac{s}{2}}}{4^{n+2}} \geq \frac{2}{2^s} \cdot g^s \geq \frac{1}{2} \cdot \frac{9}{4^s} \cdot g^s$$

(5) 若 $|OA_5| < g \leq |OA_6|$, 即 $\frac{\sqrt{2}}{3^n} \cdot \frac{2}{3} < g \leq \frac{\sqrt{2}}{3^n} \cdot \frac{2}{3} + \frac{\sqrt{2}}{3^{n+2}} = \frac{\sqrt{2}}{3^{n+1}} \cdot \frac{7}{3}$, 则 $g^s \leq \frac{2^{\frac{s}{2}}}{4^{n+1}} \cdot \frac{7^s}{4}$, 且 Δ_g 包含三个不相交的 Ω_{n+1} , 从而有:

$$\mu(\Delta_g) \geq \frac{3 \cdot 2^{\frac{s}{2}}}{4^{n+1}} \geq \frac{3 \cdot 4}{7^s} \cdot g^s \geq \frac{1}{2} \cdot \frac{9}{4^s} g^s$$

(6) 若 $|OA_6| < g \leq |OC|$, 即 $\frac{\sqrt{2}}{3^{n+1}} \cdot \frac{7}{3} < g \leq \frac{\sqrt{2}}{3^n}$, 则 $g^s \leq \frac{2^{\frac{s}{2}}}{4^n}$, 且 Δ_g 包含三个不相交的 Ω_{n+1} 和一个 Ω_{n+2} , 从而有:

$$\mu(\Delta_g) \geq \frac{3 \cdot 2^{\frac{s}{2}}}{4^{n+1}} + \frac{2^{\frac{s}{2}}}{4^{n+2}} = \frac{2^{\frac{s}{2}}}{4^n} \cdot \frac{13}{16} \geq \frac{13}{16} g^s \geq \frac{1}{2} \cdot \frac{9}{4^s} g^s$$

证毕。

引理 2 对任何可测集 V , 有 $\mu(V) \leq 2^{\frac{s}{2}} |V|^s$.

证明 不妨设 $V \subset E_0$, 否则用 $V \cap E_0$ 代替。

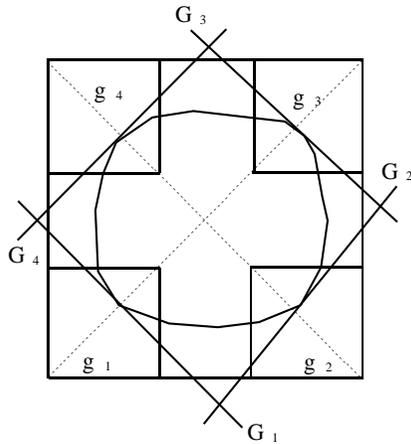


图 3

(1) 若 V 与 E_1 中四个基本正方形相交, 作平行于主和斜对角线的直线 G_1, G_2, G_3, G_4 使 $V \subset \Omega_V$, 且 V 与 Ω_V 的四边均相交 (见图 3), 其中 Ω_V 为由直线 G_1, G_2, G_3, G_4 围成的矩形。记距离:

$$d((0,0), G_1) = g_1, d((1,0), G_2) = g_2, d((1,1), G_3) = g_3, d((0,1), G_4) = g_4$$

则 $|V| \geq \sqrt{2} - g_1 - g_3, |V| \geq \sqrt{2} - g_2 - g_4$, 从而 $|V| \geq \sqrt{2} - \frac{1}{2}(g_1 + g_2 + g_3 + g_4) = \sqrt{2}[1 - \frac{\sqrt{2}}{4}(g_1 + g_2 + g_3 + g_4)]$, 由于 $\frac{\sqrt{2}}{4}(g_1 + g_2 + g_3 + g_4) \leq \frac{\sqrt{2}}{4} \cdot 4 \cdot \frac{\sqrt{2}}{3} < 1$, 从而由命题 1 和引理 1 知:

$$\begin{aligned} \frac{2^s}{2} |V|^s &\geq \frac{2^s}{2} \cdot 2^{\frac{s}{2}} \left[1 - \frac{\sqrt{2}}{4} (g_1 + g_2 + g_3 + g_4) \right]^s \geq \frac{2^s}{2} \cdot 2^{\frac{s}{2}} \left[\frac{2}{2^s} - 2^{\frac{s}{2}} \left(\frac{g_1 + g_2 + g_3 + g_4}{4} \right)^s \right] \\ &\geq 2^{\frac{s}{2}} - \frac{4^s}{2} \left(\frac{g_1^s + g_2^s + g_3^s + g_4^s}{4} \right) \end{aligned}$$

及

$$\mu(V) \leq \sum_{k=1}^4 \left[\frac{2^{\frac{s}{2}}}{4} - \mu(\Delta_{g_k}) \right] = 2^{\frac{s}{2}} \sum_{K=1}^4 \mu(\Delta_{g_k}) \leq 2^{\frac{s}{2}} - \frac{1}{2} \cdot \frac{9}{4^s} (g_1^s + g_2^s + g_3^s + g_4^s)$$

$$\text{故 } 2^{\frac{s}{2}} |V|^s - \mu(V) \geq \frac{2^s}{2} |V|^s - \mu(V) \geq \left(\frac{1}{2} \cdot \frac{9}{4^s} - \frac{4^s}{8} \right) (g_1^s + g_2^s + g_3^s + g_4^s) \geq 0$$

(2) 若 V 与 E_1 中三个基本正方形相交, 而与另一个不相交, 不妨设 V 与位于左上角的基本正方形不相交 (见图 3), 则 $|V| \geq \sqrt{2} - g_1 - g_3 = \sqrt{2} \left[1 - \frac{\sqrt{2}(g_1 + g_3)}{2} \right]$ 。由命题 1 知, $2^{\frac{s}{2}} |V|^s \geq 2^s \left[1 - \frac{\sqrt{2}(g_1 + g_3)}{2} \right]^s \geq 2^s \left[\frac{2}{2^s} - 2^{\frac{s}{2}} \left(\frac{g_1 + g_3}{2} \right)^s \right] \geq 2 - \frac{2^s \cdot 2^{\frac{s}{2}}}{2} (g_1^s + g_3^s)$ 。由引理 1 知, $\mu(V) \leq 3 \cdot \frac{2^{\frac{s}{2}}}{4} - [\mu(\Delta_{g_1}) + \mu(\Delta_{g_2}) + \mu(\Delta_{g_3})] \leq 3 \cdot \frac{2^{\frac{s}{2}}}{4} - \frac{1}{2} \cdot \frac{9}{4^s} (g_1^s + g_2^s + g_3^s)$ 。

由于 $g_1^s + g_3^s \leq 2 \cdot \frac{2^{\frac{s}{2}}}{3^s} = \frac{2^{\frac{s}{2}}}{2}$, 从而 :

$$\begin{aligned} 2^{\frac{s}{2}} |V|^s - \mu(V) &\geq 2 - 3 \cdot \frac{2^{\frac{s}{2}}}{4} + \left(\frac{1}{2} \cdot \frac{9}{4^s} - \frac{2^s \cdot 2^{\frac{s}{2}}}{2} \right) (g_1^s + g_3^s) + \frac{1}{2} \cdot \frac{9}{4^s} g_2^s \\ &\geq 2 - 3 \cdot \frac{2^{\frac{s}{2}}}{4} + \left(\frac{1}{2} \cdot \frac{9}{4^s} - \frac{2^s \cdot 2^{\frac{s}{2}}}{2} \right) \cdot \frac{2^{\frac{s}{2}}}{2} + \frac{1}{2} \cdot \frac{9}{4^s} g_2^s \\ &\geq 0.8384 - 0.832 > 0 \end{aligned}$$

(3) 若 V 仅与 E_1 中两个基本正方形 Ω_1, Ω_1 相交。当 Ω_1, Ω_1 位于对角线上, 不妨设在主对角线上(参见图3) 则 $|V| \geq \sqrt{2} - g_1 - g_3$, 由(2)和引理1得 $2^{\frac{s}{2}} |V|^s \geq 2 - \frac{2^s \cdot 2^{\frac{s}{2}}}{2} (g_1^s + g_3^s)$,

$\mu(V) \leq 2 \cdot \frac{2^{\frac{s}{2}}}{4} - [\mu(\triangle_{g_1}) + \mu(\triangle_{g_3})] \leq \frac{2^{\frac{s}{2}}}{4} - \frac{1}{2} \cdot \frac{9}{4^s} (g_1^s + g_3^s)$ 。由于 $g_1^s + g_3^s \leq 2 \cdot \frac{2^{\frac{s}{2}}}{3^s} = \frac{2^{\frac{s}{2}}}{2}$, 从而 :

$$\begin{aligned} 2^{\frac{s}{2}} |V|^s - \mu(V) &\geq 2 - \frac{2^{\frac{s}{2}}}{2} + \left(\frac{1}{2} \cdot \frac{9}{4^s} - \frac{2^s \cdot 2^{\frac{s}{2}}}{2} \right) (g_1^s + g_3^s) \geq 2 - \frac{2^{\frac{s}{2}}}{2} + \left(\frac{1}{2} \cdot \frac{9}{4^s} - \frac{2^s \cdot 2^{\frac{s}{2}}}{2} \right) \cdot \frac{2^{\frac{s}{2}}}{2} \\ &\geq 1.225 - 0.833 > 0 \end{aligned}$$

当 Ω_1, Ω_1 位于 E_0 中同一条边上, 如图4至图7所示, 记 $a = \min\{x : (x, y)\} \in (\Omega_1 \cup \Omega_1) \cap V\}$; $b = \max\{x : (x, y)\} \in (\Omega_1 \cup \Omega_1) \cap V\}$ 则 $0 \leq a \leq \frac{1}{3}, \frac{2}{3} \leq b \leq 1$, 由直线 $x = a, x = \frac{1}{3}, y = 0, y = \frac{1}{3}$ 围成的矩形记为 Ω_a , 由直线 $x = b, x = \frac{2}{3}, y = 0, y = \frac{1}{3}$ 围成的矩形记为 Ω_b , 则 $|V| \geq b - a$, 且 $\mu(V) \leq \mu(\Omega_a) + \mu(\Omega_b)$

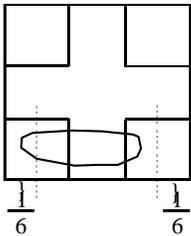


图4

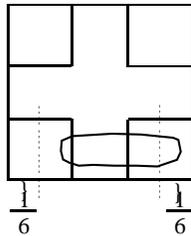


图5

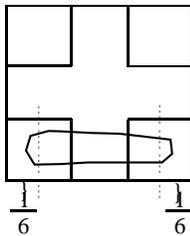


图6

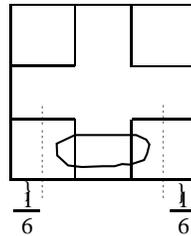


图7

(1) 当 $0 \leq a \leq \frac{1}{6}, \frac{2}{3} \leq b \leq 1$ 或 $\frac{1}{6} \leq a \leq 1, \frac{5}{6} \leq b \leq 1$ 时(见图4, 图5), 我们有 :

$$|V|^s \geq (b - a)^s \geq \left(\frac{1}{2} \right)^s, \mu(V) \leq \mu(\Omega_a) + \mu(\Omega_b) \leq \frac{2^{\frac{s}{2}}}{4} + \frac{2 \cdot 2^{\frac{s}{2}}}{16} = \frac{3 \cdot 2^{\frac{s}{2}}}{8}$$

故 $2^{\frac{s}{2}} |V|^s - \mu(V) \geq 2^{\frac{s}{2}} \cdot \left(\frac{1}{2}\right)^s - \frac{3 \cdot 2^{\frac{s}{2}}}{8} = 2^{\frac{s}{2}} \left(\frac{1}{2^s} - \frac{3}{8}\right) \geq 2^{\frac{s}{2}} (0.41667 - 0.375) > 0$.

(2) 当 $0 \leq a \leq \frac{1}{6}, \frac{5}{6} \leq b \leq 1$ 时(见图 6), 我们有:

$$|V|^s \geq (b - a)^s \geq \left(\frac{2}{3}\right)^s = \frac{2^s}{4}, \mu(V) \leq \mu(\Omega_a) + \mu(\Omega_b) \leq 2 \cdot \frac{2^{\frac{s}{2}}}{4}$$

故 $2^{\frac{s}{2}} |V|^s - \mu(V) \geq \frac{2^s \cdot 2^{\frac{s}{2}}}{4} - \frac{2^{\frac{s}{2}}}{2} = 2^{\frac{s}{2}} \left(\frac{2^s}{4} - \frac{1}{2}\right) > 0$.

(3) 当 $\frac{1}{6} \leq a \leq \frac{1}{3}, \frac{2}{3} \leq b \leq \frac{5}{6}$ 时(见图 7), 我们有:

$$|V|^s \geq (b - a)^s \geq \left(\frac{1}{3}\right)^s = \frac{1}{4}, \mu(V) \leq \mu(\Omega_a) + \mu(\Omega_b) \leq 4 \cdot \frac{2^{\frac{s}{2}}}{16} = \frac{2^{\frac{s}{2}}}{4}$$

故 $2^{\frac{s}{2}} |V|^s - \mu(V) \geq \frac{2^{\frac{s}{2}}}{4} - \frac{2^{\frac{s}{2}}}{4} = 0$

(4) 若 V 仅与 E_1 中一个基本正方形 Ω_1 相交, 则 $\mu(V) = \mu(V \cap \Omega_1)$ 。当 $V \cap \Omega_1$ 仅与 E_2 中一个基本正方形 Ω_2 相交, 则 $\mu(V) = \mu(V \cap \Omega_2)$ 。当 $V \cap \Omega_1$ 与 E_2 中二个、三个、四个基本正方形相交, 则由(1) ~ (3) 知 $\mu(V) = \mu(V \cap \Omega_1) \leq 2^{\frac{s}{2}} |V \cap \Omega_1|^s \leq 2^{\frac{s}{2}} |V|^s$ 。由数学归纳法知: 或者 $\mu(V) \leq 2^{\frac{s}{2}} |V|^s$ 或者存在 $\Omega_n \in E_n$, 使得 $\mu(V) = \mu(V \cap \Omega_n) \leq \mu(\Omega_n) = \frac{2^{\frac{s}{2}}}{4^n}$ 。

令 $n \rightarrow \infty$, 得 $\mu(V) = 0 \leq 2^{\frac{s}{2}} |V|^s$ 。证毕。

三、定理的证明

定理的证明 由质量分布原理 [1] 和引理 1 知:

$$2^{\frac{s}{2}} H^s(C \times C) \geq \mu(C \times C) = 2^{\frac{s}{2}}$$

从而 $H^s(C \times C) \geq 1$ 。下面估计 $H^s(C \times C)$ 的上界。

建立如图 8 所示的直角坐标系, 以正方形 E_0 的一个顶点 A 为坐标原点, 在正方形 E_0 的边上截取线段 $\overline{AE} = \overline{AF} = 2x$, $\overline{BG} = \overline{BH} = 2x$, $\overline{CI} = \overline{CJ} = 2x$, $\overline{DK} = \overline{DL} = 2x$, 记八边形 $EFGHIJKL$ 为 U_x 。取 $x = \frac{1}{3^3} + \frac{1}{3^6} + \frac{1}{3^9} + \frac{1}{3^{12}}$, 则除开 E_3 中的 4 个小正方形, E_6 中的 8 个小正方形, E_9 中的 16 个小正方形以及 E_{12} 中的 32 个小正方形外, 其余的小正方形都被 U_x 覆盖, 从而由命题 2 得:

$$\begin{aligned} & \left[1 - 4 \cdot \left(\frac{1}{4^3} + \frac{2}{4^6} + \frac{4}{4^9} + \frac{8}{4^{12}} \right) \right] H^s(C \times C) \\ & \leq H^s((C \times C) \cap U_x) \leq |U_x|^s \leq (\sqrt{1 + (1 - 4x)^2})^s \end{aligned}$$

即:

$$H^s(C \times C) \leq \frac{(\sqrt{1 + (1 - 4x)^2})^s}{1 - 4 \cdot \left(\frac{1}{4^3} + \frac{2}{4^6} + \frac{4}{4^9} + \frac{8}{4^{12}}\right)}$$

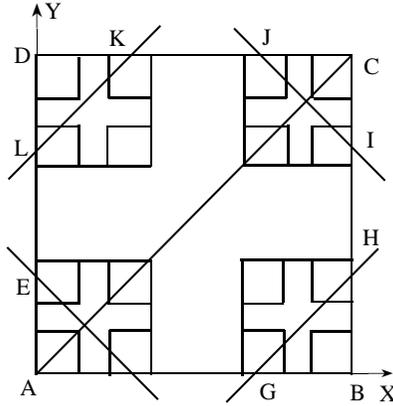


图 8

故 $H^s(C \times C) \leq 1.502879$ 。证毕。

致谢 作者对瞿成勤博士仔细审阅本文原稿表示感谢。

参考文献

- [1] Falconer K. , Fractal Geometry-Mathematical Foundations and Applications , New York : John Wiley and Sons , 1990.
- [2] Zhou Z. L. , The Hausdorff measure of the koch curve and Sierpinski , Prog. Nat. Sci. , 1997 , Vol.7 , No.4 , pp405 ~ 409.
- [3] Zhou Z. L. , Hausdorff measure of Sierpinski gasket , Science of Chiana , Ser. A. , 1997 , Vol.27 , No.6 , pp491 ~ 496.
- [4] Zhou Z. L. , Hausdorff measure of Self-similar sets-koch curve , Science of China , Ser. A. , 1998 , Vol.28 , No.2 , pp103 ~ 107.
- [5] Zhou Z. L. , Wu. M. , Hausdorff measure of a Sierpinski Carpet , Science of China , Ser. A. , 1999 , Vol.29 , No.2 , pp138 ~ 144.
- [6] Zhou Z. L. , Hausdorff measure of generalized Sierpinski Carpets , Approx. Th. & Appli. , 2000 , Vol.16 , No.2 , pp13 ~ 18.

自相似集的 Hausdorff 测度与连续性*

罗俊 周作领

一、引言

以 $\|x - y\|$ 表示欧氏空间 R^n 中两点 x, y 之间的距离, 以 $\|U\| = \sup\{\|x - y\| : x, y \in U\}$ 表示集合 $U \subset R^n$ 的直径. 设 $F \subset R^n$, 对 $s \geq 0, \delta > 0$, 令 $H_\delta^s(F) = \inf \sum_i \|U_i\|^s : F \subset (\cup_i U_i), \|U_i\| < \delta\}$ 则 $H^s(F) = \lim_{\delta \rightarrow 0} H_\delta^s(F)$ 称作集合 F 的 s 维 Hausdorff 测度, $\dim F = \inf\{s : H^s(F) = 0\} = \sup\{s : H^s(F) = \infty\}$ 称作 F 的 Hausdorff 维数. 有时, 我们将集合 F 的 $\dim F$ 维 Hausdorff 测度简称为 F 的“Hausdorff 测度”. 容易验证, $H^s(\cdot)$ 是 R^n 上的度量外测度, 从而是 Borel 测度^[1]. 通常, Hausdorff 维数可以看成对集合大小或不规则程度的某种量度, 而 Hausdorff 维数的定义与探讨均涉及 Hausdorff 测度. 然而, 除去满足开集条件的自相似集等特殊的集合之外, 计算和估计分形的 Hausdorff 维数是非常困难的问题, 而计算分形的 Hausdorff 测度尤其如此. 实际上, 除去线段上的例子和一些平凡的情形, 即使满足开集条件的自相似集, 其 Hausdorff 测度之准确值也难以得到. 到目前为止, 除去平凡情形, Hausdorff 维数大于 1 的自相似集的 Hausdorff 测度准确值问题没有一个得以解决. 比如, 对 $0 < a \leq \frac{1}{2}$, 定义 R^2 上的四个相似压缩如下:

$$f_{a,1}(x) = ax, f_{a,2}(x) = ax + \begin{pmatrix} 1-a \\ 0 \end{pmatrix}, f_{a,3}(x) = ax + \begin{pmatrix} 1-a \\ 1-a \end{pmatrix}, f_{a,4}(x) = ax + \begin{pmatrix} 0 \\ 1-a \end{pmatrix}$$

则 $\{f_{a,k} : 1 \leq k \leq 4\}$ 的吸引子 T_a 称作广义 Sierpinski 地毯, 它的 Hausdorff 维数 $\dim T_a = \frac{\ln 4}{-\ln a}$. 当 $\dim T_a \leq 1$ ($a \leq \frac{1}{4}$) 时, T_a 的 ($\frac{\ln 4}{-\ln a}$ 维) Hausdorff 测度等于 $2^{-\frac{\ln 2}{\ln a}}$, 因而连续依赖于参数 a ^[2,3]; 当 $\dim T_a = 2$ ($a = \frac{1}{2}$) 时, $T_a = [0, 1] \times [0, 1]$, 其 Hausdorff 测度等于 $\frac{4}{\pi}$; 当 $1 < \dim T_a < 2$ ($\frac{1}{4} < a < \frac{1}{2}$) 时, T_a 的 Hausdorff 测度准确值目前尚未得到. 一个自然的问题是, 当 $\frac{1}{4} < a \leq \frac{1}{2}$ 时, 自相似集 T_a 的 Hausdorff 测度是否也依赖于参数 a ? 后文的定理 3.4 将对此给以肯定回答.

本文考虑 R^n 自相似集 $T = T(f_1, \dots, f_m)$, 假如 T 满足强分离条件, 即 $f_i(T) \cap f_j$

* 基金项目: 国家自然科学基金 (10041005)、数学天元青年基金 (10226031)、广东省自然科学基金 (011221)、中山大学青年教师启动基金 (34100-1131206) 资助项目. 原载: 《数学学报》, 2003 年第 46 卷第 3 期.

$(T) = \Phi (i \neq j)$ 。那么,对每一 $\epsilon > 0$,必存在 $\delta > 0$,若 $D = D (g_1, \dots, g_m) \subset R^n$ 为自相似集且 $\sup \{ \| f_k (x) - g_k (x) \| : \| x \| \leq 1, 1 \leq k \leq m \} < \delta$,则 T 的 Hausdorff 测度相差不超过 ϵ 。也就是说,当 g_k 充分接近 $f_k (1 \leq k \leq m)$ 时, D 与 T 的 Hausdorff 测度也充分接近。

二、概念与基本结论

(一) 自相似集

对任意 $U, V \subset R^n$,以 $\| U - V \| = \inf \{ \| x - y \| : x \in U, y \in V \}$ 表示两集合之间的距离。设 f 为 R^n 上的压缩映射,若存在常数 $c \in (0, 1)$,使得 $\| f(x) - f(y) \| \leq c \| x - y \|$ 对所有 $x, y \in R^n$ 成立,则称 f 为 R^n 上的压缩映射;若 $\| f(x) - f(y) \| = c \| x - y \|$ 对所有 $x, y \in R^n$ 成立,则称 f 为 R^n 上的相似压缩, c 称作 f 的相似比。若 $f_1, \dots, f_m (m \geq 2)$ 为 R^n 上的压缩映射,则它们组成一个迭代函数系统 $\{f_1, \dots, f_m\}$;此时,必存在 R^n 中惟一的非空紧集 $T = T (f_1, \dots, f_m)$ 满足 $T = \bigcup_k f_k (T)$ (见文 [4])。这里,集合 T 称作 $\{f_1, \dots, f_m\}$ 的吸引子;若 f_1, \dots, f_m 都是相似压缩,则 T 称作由它们生成的自相似集。若存在非空开集 V ,使得 $f_i (V) \cap f_j (V) = \Phi (i \neq j)$ 且 $[\bigcup_k f_k (V)] \subset V$,则称 f_1, \dots, f_m 或自相似集 $T (f_1, \dots, f_m)$ 满足开集条件。设 $T = T (f_1, \dots, f_m)$ 为满足开集条件的自相似集, f_k 的相似比为 c_k ,则 T 的 Hausdorff 维数是满足 $\sum_k c_k^s = 1$ 的惟一正数 s , T 的 (s 维) Hausdorff 测度是正有限的,并且存在惟一的概率测度 μ_T ,使得 $\mu_T (B) = \sum_k c_k^{\dim T} \mu_T \circ f_k^{-1} (B)$ 对所有 Borel 集 $B \subset R^n$ 都成立 (见文 [5])。我们将 μ_T 称作支撑在 T 上的自然质量分布。对于满足开集条件的自相似集 T 和其上自然质量分布 μ_T ,下述结论成立:

- (1) 对于任意 Borel $B \subset R^n$,必有 $\mu_T (B) = \frac{H^{\dim T} (B \cap T)}{H^{\dim T} (T)}$, $\| B \| \dim T \geq H^{\dim T} (B \cap T)$;
- (2) $H^{\dim T} (T) = \inf \{ \sum_i \| U_i \| \dim T : T \subset \bigcup_i U_i \}$;
- (3) $H^{\dim T} (T) = \inf \{ \sum_i \| U_i \| \dim T : T \subset \bigcup_i U_i \}$,其中 U_i 为开集;
- (4) $H^{\dim T} (T) = \inf \{ \sum_i \| U_i \| \dim T : T \subset \bigcup_i U_i \}$,其中 U_i 为闭集。

由此,可以推出下述引理:

引理 1.1 设 $T \subset R^n$ 为满足开集条件的自相似集, μ_T 为其上自然质量分布,则 T 的 Hausdorff 测度等于 $\beta (T) = \inf \{ \frac{\| V \| \dim T}{\mu_T (V)} \}$,其中 V 取遍满足 $\mu_T (V) > 0$ 的闭集。

若自相似集 $T = T (f_1, \dots, f_m) \subset R^n$ 满足强分离条件,令 $\alpha (T) = \inf \{ \frac{\| V \| \dim T}{\mu_T (V)} \}$,其中 V 取遍满足 $\mu_T (V) > 0$ 和 $\| V \| \geq \min \{ \rho (f_i (T), f_j (T)) : i \neq j \}$ 的闭集。对闭集 V ,若 $\mu_T (T) > 0$ 且 $\| V \| < \min \{ \rho (f_i (T), f_j (T)) : i \neq j \}$,令 k 为最大的正整

数, 使得存在 $i_1, \dots, i_k \in \{1, \dots, m\}$, 满足 $(V \cap T) \subset f_{i_1} \circ \dots \circ f_{i_k}(T)$. 令 $g = f_{i_1} \circ \dots \circ f_{i_k}$, 则:

$$\frac{\|V \cap T\|_{\mu_T(V)}^{\dim T}}{\|V\|_{\mu_T(V)}^{\dim T}} = \frac{(c_{i_1} \dots c_{i_k})^{-\dim T} \|V \cap T\|_{\mu_T(V)}^{\dim T}}{(c_{i_1} \dots c_{i_k})^{-\dim T} \mu_T(V)} = \frac{\|g^{-1}(V \cap T)\|_{\mu_T(g^{-1}(V \cap T))}^{\dim T}}{\mu_T(g^{-1}(V \cap T))}^{\dim T}$$

其中: c_{i_j} 为 $f_{i_j} (1 \leq j \leq m)$ 的压缩比. 由于 $\|g^{-1}(V \cap T)\| \geq \min\{\rho(f_{i_1}(T), f_{i_2}(T)), \dots, \rho(f_{i_{k-1}}(T), f_{i_k}(T))\}$, 利用引理 1.1 可得下述引理 1.2.

引理 1.2 若 T 为满足强分离条件的自相似集, 则 T 的 Hausdorff 测度等于 $\alpha(T)$.

(二) 关于一致上半连续

设 (X, d) 为度量空间, 对 $r > 0, x \in X$, 以 $V(x, r)$ 表示 X 中以 x 为圆心 r 为半径的开球. 设 $f: X \rightarrow Y$ 为从 (X, d) 到度量空间 (Y, ρ) 的映射, 若对于任意 $\epsilon > 0$, 必然存在 $\delta > 0$, 使得任意 $x \in M, f(V(x, \delta)) \subset V(f(x), \epsilon)$, 则称 f 在 X 上一致连续.

定义 1.3 设 $f: X \rightarrow R$ 为度量空间 (X, d) 上的函数, 若对于任意 $\epsilon > 0$, 必存在 $\delta > 0$, 使得任意 $x \in X, f(V(x, \delta)) \subset (f(x) - \epsilon, f(x) + \epsilon)$, 则称 f 在 X 上一致上半连续; 若对于任意 $\epsilon > 0$, 必存在 $\delta > 0$, 使得任意 $x \in X, f(V(x, \delta)) \subset (f(x) - \epsilon, f(x) + \epsilon)$, 则称 f 在 X 上一致下半连续.

定理 1.4 一致下半连续性与一致连续性等价, 同样, 一致上半连续性与一致下半连续性等价.

证明 设 $f: M \rightarrow R$ 为一致下半连续映射. 若 f 非一致连续, 则有 $\epsilon > 0$ 和两个点列 $\{x_n\}, \{y_n\} \subset M$, 使得 $d(x_n, y_n) > \frac{1}{n}$ 且 $f(x_n) > f(y_n) - \epsilon$. 对上述 $\epsilon > 0$, 选取 $\delta > 0$, 使得 $f(B(x, \delta)) \subset (f(x) - \epsilon, f(x) + \epsilon)$ 对所有 $x \in M$ 都成立, 那么, 对于一切整数 $k > \frac{1}{\delta}$, 都有 $y_k \in B(x_k, \delta)$ 和 $f(y_k) > f(x_k) + \epsilon$. 矛盾. 证毕.

三、Hausdorff 测度的连续性

记 Ω 为 R^n 中所有非空紧子集的集合, $\rho: \Omega \times \Omega \rightarrow R$ 为其上 Hausdorff 度量. 以下恒设 $T = T(f_1, \dots, f_m)$ 为满足强分离条件的自相似集, $\omega = \frac{1}{4} \min\{\|f_i(T) - f_j(T)\| : i \neq j\}, X = \{z \in R^n : \|\{z\} - T\| \leq \omega\}$ 其中, 对每个相似压缩 f_k 存在 $n \times n$ 阶正交矩阵 A_k , 点 $a_k \in R^n$ 和常数 $p_k \in (0, 1)$ 满足 $f_k(x) = p_k A_k(x) + a_k$.

由 Hausdorff 测度和维数在相似映射下的变化特征, 我们不妨假定 $|X| = 1$ 且原点在 X 中.

选取常数 $\tau > 0$, 使得:

(1) $[p_k - \tau, p_k + \tau] \subset (0, 1)$ 对 $1 \leq k \leq n$ 均成立;

(2) 对任意相似压缩 $g_k(x) (1 \leq k \leq n)$, 若 $\max_{1 \leq k \leq n} \sup\{\|f_k(x) - g_k(x)\| : \|x\| \leq$

$1\} \leq \tau$ 则 $[\cup g_k(X)] \subset X$ 且 $\min\{\|g_i(X) - g_j(X)\| : i \neq j\} \geq \omega$ 。

恒记 Φ 为一切形如 $\{g_1, \dots, g_m\}$ 的迭代函数系统的集合, 其中, 每个 g_k 均为相似压缩, 对任意 $\{g_k\} \{h_k\} \in \Phi$, 令 $\phi(\{g_k\} \{h_k\}) = \max_{1 \leq k \leq n} \sup\{\|f_k(x) - g_k(x)\| : \|x\| \leq 1\}$, 则 ϕ 为 Φ 上的一个度量。可以验证, 对于任意迭代函数系统 $\{g_k : 1 \leq k \leq m\} \in \Phi$, 其吸引子必然是满足强分离条件的自相似集。由此, 可以得到下列引理 2.1。

引理 2.1 设自相似集 D, E 分别迭代函数系统 $\{g_k : 1 \leq k \leq m\} \{h_k : 1 \leq k \leq m\} \in \Phi$ 的吸引子, 那么, 对任意正数 ϵ , 存在 $\delta > 0$, 使得当 $\phi(\{g_k\} \{h_k\}) \leq \delta$ 时, $|\dim D - \dim E| < \epsilon$, 即, 当迭代函数系统 $\{g_1, \dots, g_m\} \{h_1, \dots, h_m\} \in \Phi$ 在度量 ϕ 下充分接近, 它们的吸引子的 Hausdorff 维数也充分接近。

引理 2.2 设迭代函数系统 $\{g_k : 1 \leq k \leq m\} \{h_k : 1 \leq k \leq m\} \in \Phi$, 概率测度 μ_d, μ_E 分别为它们的吸引子 D, E 上的自然质量分布。那么, 对所有正数 ϵ, δ , 存在正数 δ_1 , 使得对任意闭集 $Y \subset X$ 及其 δ -邻域 $Y_\delta = \{z : \|\{z\} - Y\| \leq \delta\}$, 不等式 $\mu_D(Y_\delta) \geq \mu_E(Y) - \epsilon$ 总成立。

证明 选取正整数 N , 使得对任一迭代函数系统 $\{g_k\} \in \Phi$ 和任一正整数序列 $k_1, \dots, k_N \in \{1, \dots, m\}$, 都有 $\|g_{k_1} \circ \dots \circ g_{k_N}(X)\| < \frac{1}{2}\delta$ 。对 $\delta > 0$, 任意取定 $\delta_2 \in (0, \frac{1}{6}\delta(1-p))$, 其中 $p = \max\{p_k + \tau : 1 \leq k \leq m\}$, 那么, 当 $\phi(\{g_k\} \{h_k\}) < \delta_2$ 时, $\rho(g_k(Y), h_k(Y)) < \frac{1}{2}(1-p)\delta$ 对任一紧致子集 $Y \subset X$ 和 $1 \leq k \leq m$ 都成立。从而, 对任意 $j < 0$ 和 $k_1, k_2, \dots, k_j \in \{1, \dots, m\}$, 下式总成立:

$$\rho(g_{k_j} \circ \dots \circ g_{k_1}(X), h_{k_j} \circ \dots \circ h_{k_1}(X)) < \frac{\delta}{2}$$

假如 μ_d, μ_E 分别为 $\{g_k : 1 \leq k \leq m\} \{h_k : 1 \leq k \leq m\}$ 的吸引子 D, E 上的自然质量分布, 那么, 由引理 2.1, 对 $\epsilon > 0$, 可选取 $\delta_1 \in (0, \delta_2)$ 使得当 $\phi(\{g_k\} \{h_k\}) < \delta_1$ 时, $|\mu_D[g_{k_N} \circ \dots \circ g_{k_1}(X)] - \mu_E[h_{k_N} \circ \dots \circ h_{k_1}(X)]| < m^{-N}\epsilon$ 对所有有限序列 $k_1, \dots, k_N \in \{1, \dots, m\}$ 都成立。于是, 对任一闭集 $Y \subset X$ 有:

$$\mu_E(Y) \leq \sum \{\mu_E \{h_{k_N} \circ \dots \circ h_{k_1}(X)\} : h_{k_N} \circ \dots \circ h_{k_1}(X) \cap Y \neq \emptyset\}$$

因 $\{h_{k_N} \circ \dots \circ h_{k_1}(X) : h_{k_N} \circ \dots \circ h_{k_1}(X) \cap Y \neq \emptyset\}$ 中所有元素之并集包含于集合 Y 的 $\frac{\delta}{2}$ -邻域, 故 $\{g_{k_N} \circ \dots \circ g_{k_1}(X) : h_{k_N} \circ \dots \circ h_{k_1}(X) \cap Y \neq \emptyset\}$ 中所有元素之并集必然包含于 Y 的 δ -邻域 Y_δ , 从而 $\mu_D(Y_\delta) \geq \mu_E(Y) - \epsilon$ 。类似的讨论可得 $\mu_E(Y_\delta) \geq \mu_D(Y) - \epsilon$ 。证毕。

定理 2.3 对任一迭代函数系统 $\{g_k\} \in \Phi$ 的吸引子 D , 令 $H(\{g_k\}) = H^{\dim D}(D)$, 则 $H : (\Phi, \phi) \rightarrow R$ 是一致上半连续映射, 从而是一致连续映射。

证明 记 λ 为 Φ 中所有迭代函数系统的吸引子的 Hausdorff 维数之下确界, 并取定 $\eta = \frac{1}{2}\omega^n < 1$ 。对任一 $\epsilon \in (0, \eta)$, 选取 $\delta_1 \in (0, \frac{1}{6}\epsilon\eta^2)$, 使得:

(i) 对任意 $y, z \in [\omega, 1]$ 和 $s, t \in [\lambda, n]$, 当 $|y - z| < 2\delta_1, |s - t| < \delta_1$ 时, $|y^s -$

$$z^j | < \frac{1}{6} \eta \epsilon ;$$

(ii) 对任意 $y \in [\eta, 1]$ 和 $z \in [\omega^n, 1]$, $\frac{z}{y - \delta_1} < \frac{z}{y} + \frac{1}{3} \epsilon$ 总成立。

由引理 2.1, 可选取 $\delta_2 \in (0, \delta_1)$, 使得对任意迭代函数系统 $\{g_k : 1 \leq k \leq m\} \setminus \{h_k : 1 \leq k \leq m\} \in \Phi$ 及其吸引子 D, E , 当 $\phi(\{g_k\} \setminus \{h_k\}) < \delta_2$ 时, $|\dim D - \dim E| < \delta_1$ 。

由引理 2.2, 可选取 $\delta \in (0, \delta_2)$, 使得对任意迭代函数系统 $\{g_k : 1 \leq k \leq m\} \setminus \{h_k : 1 \leq k \leq m\} \in \Phi$ 及其吸引子 D, E 上的自然质量分布 μ_D, μ_E , 当 $\phi(\{g_k\} \setminus \{h_k\}) \leq \delta$ 时, 不等式 $\mu_D(U) \geq \mu_E(V) - \delta_1$ 对任意闭集 $V \subset X$ 及其 $\delta_2 -$ 邻域 U 都成立。

对前述 $\epsilon > 0$, 取闭集 V , 使 $\|V\| \geq \omega$ 且 $\frac{1}{\mu_E(V)} \|V\|^{\dim E} < H^{\dim E}(E) + \frac{1}{3} \epsilon$ 。此时 $\mu_E(V) > \frac{1}{2} \omega^{\dim E} \geq \eta$, 于是:

$$\begin{aligned} H^{\dim D}(D) &\leq \frac{1}{\mu_D(U)} \|U\|^{\dim D} < \frac{1}{\mu_E(V) - \delta_1} \|V\|^{\dim E} + \frac{1}{3} \epsilon \\ &< \frac{1}{\mu_E(V)} \|V\|^{\dim E} + \frac{2}{3} \epsilon < H^{\dim E}(E) + \epsilon \end{aligned}$$

其中, 第二个不等号之成立是由(i)和 $\mu_E(V) - \delta_1 > \frac{1}{2} \eta$ 得到, 第三个不等号之成立是由(ii)得到。

综上所述, 我们对任一 $\epsilon > 0$, 可找到 $\delta > 0$, 使得对任意 $\{g_k : 1 \leq k \leq m\} \setminus \{h_k : 1 \leq k \leq m\} \in \Phi$ 及其吸引子 D, E , 当 $\phi(\{g_k\} \setminus \{h_k\}) < \delta$ 时, $H^{\dim D}(D) < H^{\dim E}(E) + \epsilon$ 。因此, 由 ϵ 的任意性, 映射 $H : (\Phi, \phi) \rightarrow \mathbb{R}$ 是一致上半连续的。证毕。

四、广义 Sierpinski 地毯

对 $a \in (0, \frac{1}{2}]$ 以及任意 $x \in \mathbb{R}^2$, 设映射 $f_{a,k}(x) (1 \leq k \leq 4)$ 如前文所定义, T_a 为迭代函数系统 $\{f_{a,k}\}$ 的吸引子, 即压缩比为 a 的广义 Sierpinski 地毯。令 $H(a)$ 为 T_a 的 $\dim T_a$ 维 Hausdorff 测度, 则由引理 2.3, 映射 $H : (0, \frac{1}{2}] \rightarrow \mathbb{R}$ 在开区间 $(0, \frac{1}{2})$ 内连续。这一节将讨论映射 H 在 $\frac{1}{2}$ 处的连续性。

设 $\{N_k : k \geq 1\} \subset \mathbb{R}^n$ 为非空紧子集序列, 则该序列的上极限 $\overline{\lim}_{k \rightarrow \infty} N_k$ 和下极限 $\underline{\lim}_{k \rightarrow \infty} N_k$ 定义如下: 点 x 属于 $\underline{\lim}_{k \rightarrow \infty} N_k$ 当且仅当任一包含 x 的开集 V 均与除有限个 N_k 外的其余所有紧子集 N_k 相交, 点 x 属于 $\overline{\lim}_{k \rightarrow \infty} N_k$ 当且仅当任一包含 x 的开集 V 均与无限个 N_k 相交。若一紧子集的上极限和下极限相等, 则称该序列收敛, 此时上极限和下极限通称为极限, 记 $\lim N_k$ 。注意, $\lim N_k$ 可能是空集。

我们已经知道, 欧氏空间中任一非空紧子集序列必有收敛子序列^[6]。此外, 不难验证, 若

一收敛的非空紧子集序列包含于一有界集合中,则其极限为非空紧子集,并且该序列在 Hausdorff 度量 ρ 下收敛到这一极限。

对任意 $a \in (0, \frac{1}{2}]$, 记 $\Delta_{a,0} = \{[0, a] \times [0, a]\}$ 对 $k \geq 1$, 令 $\Delta_k = \{T_{a,i_1} \circ \dots \circ T_{a,i_k}([0, 1] \times [0, 1]) : i_1, \dots, i_k \in \{1, 2, 3, 4\}\}$, 易见 Δ_k 中元素为 4^k 个边长为 a^k 的小正方形。以下恒设 $s(a) = \dim T_a, \mu_a$ 为 T_a 上的自然质量分布。

引理 3.1 设 $a \in (0, \frac{1}{2}]$, 闭集 $V \subset [0, 1] \times [0, 1]$ 若 $\|V\| < a^2$, 则存在闭集 $U \subset [0, 1] \times [0, 1]$, 使得 $\|U\| > \|V\|$ 且 $\frac{1}{\mu_a(U)} \|U\|^{s(a)} \leq \frac{1}{\mu_a(V)} \|V\|^{s(a)}$ 。

证明 不妨假定 V 与 $\Delta_{a,1}$ 中至少两个小正方形相交。记 $W_i (1 \leq i \leq 8)$ 为 $\Delta_{a,2}$ 中的小正方形, 如图 1 所示。假定 V 与小正方形 $Q_1, \dots, Q_j (2 \leq j \leq 4)$ 相交, 那么, 这 j 个小正方形或者在 $\{W_3, W_4, W_5, W_6\}$ 中, 或者在 $\{W_4, W_6, W_7, W_8\}$ 中。

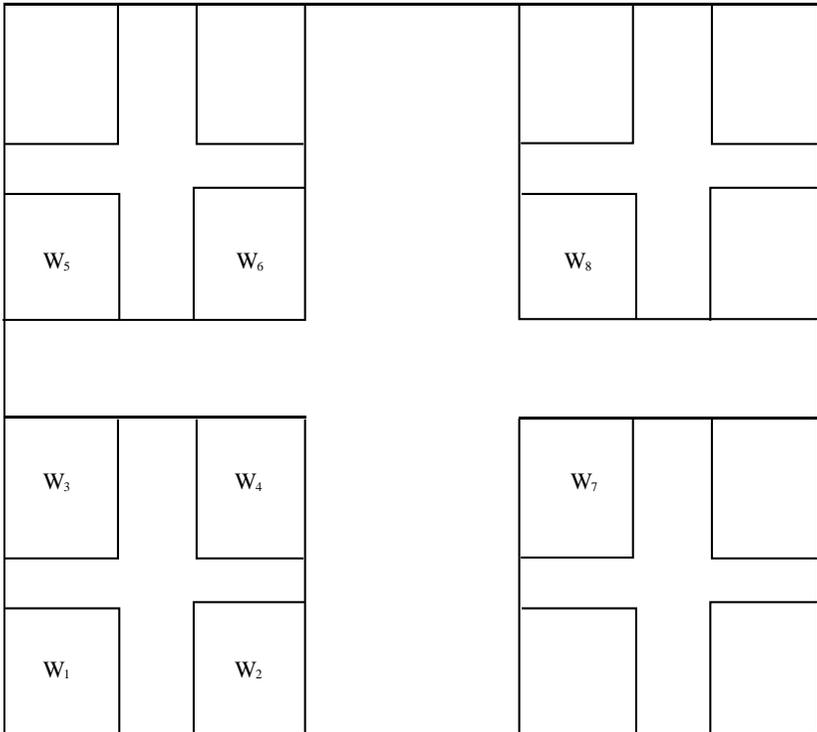


图 1 Δ_2 中 8 个小正方形的示意图

选取平面上的 8 个平移映射 $h_k (1 \leq k \leq 8)$, 其中 $h_k(W_{k+2}) = W_k (1 \leq k \leq 4), h_5(W_4) = W_1, h_6(W_7) = W_2, h_7(W_6) = W_3$ 且 $h_8(W_8) = W_4$ 。

映射 $g : (\cup_3^6 W_k) \rightarrow (\cup_1^4 W_k)$ 定义如下: 当 $x \in W_{k+2} (1 \leq k \leq 4), g(x) = h_k(x)$ 。映射 $h : (W_4 \cup W_6 \cup W_7 \cup W_8) \rightarrow (\cup_1^4 W_k)$ 定义如下: 当 $x \in W_4, h(x) = h_5(x)$; 当 $x \in W_7,$

$h(x) = h_6(x)$ 当 $x \in W_6$, $h(x) = h_7(x)$ 当 $x \in W_7$, $h(x) = h_8(x)$.

若 $\{Q_1, \dots, Q_j\}$ 在 $\{W_3, W_4, W_5, W_6\}$ 中, 令 $V_1 = g(V \cap T_a)$ 若 $\{Q_1, \dots, Q_j\}$ 在 $\{W_4, W_6, W_7, W_8\}$ 中, 令 $V_1 = h(V \cap T_a)$. 令 $U = f_{a,1}^{-1}(V_1)$, $\|U\| > \|V\|$ 且 $\frac{1}{\mu_a(U)} \|U\|^{s(a)} \leq \frac{1}{\mu_a(V)} \|V\|^{s(a)}$. 证毕。

引理 3.2 对任意 $a \in [\frac{1}{4}, \frac{1}{2}]$, $H(a) = \inf\{\frac{1}{\mu_a(V)} \|V\|^{s(a)}\}$ 其中 V 取遍 $[0, 1] \times [0, 1]$ 中满足 $\mu_a(V) > 0$ 和 $\|V\| \geq a^2$ 的闭集。

证明 只需验证 对任意闭集 $V \subset [0, 1] \times [0, 1]$ 若 $\mu_a(V) > 0$ 且 $\|V\| < a^2$ 则存在闭集 $V \subset [0, 1] \times [0, 1]$ 使得 $\|U\| \geq a^2$ 且 $\frac{1}{\mu_a(U)} \|U\|^{s(a)} \leq \frac{1}{\mu_a(V)} \|V\|^{s(a)}$.

假定有一闭集 $V \subset [0, 1] \times [0, 1]$ 满足 $\mu_a(V) > 0$ 和 $\|V\| < a^2$ 并且对任意闭集 $V \subset [0, 1] \times [0, 1]$ 若 $\|U\| \geq a^2$ 则 $\frac{1}{\mu_a(U)} \|U\|^{s(a)} > \frac{1}{\mu_a(V)} \|V\|^{s(a)}$. 令:

$$\gamma = \sup\left\{\alpha \in (0, a^2) \text{ 存在闭集 } U \text{ 使得 } \|U\| = \alpha \text{ 且 } \frac{1}{\mu_a(U)} \|U\|^{s(a)} \leq \frac{1}{\mu_a(V)} \|V\|^{s(a)}\right\}$$

选取直径单调增加的闭子集序列 $U_k \subset [0, 1] \times [0, 1]$, 使得 $\lim_{k \rightarrow \infty} \|U_k\| = \gamma$ 且

$$\frac{1}{\mu_a(U_k)} \|U_k\|^{s(a)} \leq \frac{1}{\mu_a(V)} \|V\|^{s(a)}. \text{不妨假设非空紧子集序列 } \{U_k\} \text{ 收敛到极限集 } U \text{ 则}$$

$$\frac{1}{\mu_a(U)} \|U\|^{s(a)} \leq \frac{1}{\mu_a(V)} \|V\|^{s(a)}. \text{若 } \gamma = a^2 \text{ 则 } \|U\| = \gamma = a^2 \text{ 与反证假设矛盾, 若}$$

$\gamma < a^2$, 由引理 3.1, 存在闭集 $W \subset [0, 1] \times [0, 1]$, 使得 $\|W\| > \|U\| = \gamma$ 且

$$\frac{1}{\mu_a(W)} \|W\|^{s(a)} \leq \frac{1}{\mu_a(U)} \|U\|^{s(a)} \leq \frac{1}{\mu_a(V)} \|V\|^{s(a)}. \text{矛盾. 证毕.}$$

由引理 3.2 不难得到下述定理:

定理 3.3 映射 $H: (0, \frac{1}{2}] \rightarrow \mathbb{R}$ 连续.

致谢 感谢清华大学数学系王志英教授、武汉大学数学系吴军教授和饶辉副教授与笔者进行的有益讨论, 其中引理 3.1 的结论与证明均来源于饶辉副教授的建设性建议。

参考文献

- [1] Falcomer K. J., Fractal geometry, New York: John Wiley and Sons, 1990.
- [2] Zhou Z. L., Wu M., Hausdorff measure of a sierpinski Carpet, Science in China, Ser. A., 1999, Vol.28, No.2, pp138~144 (in Chinese).
- [3] Zhu Y. C., Luo J., Hausdorff measure of generalizde sierpinski Carpets, Approx. Th. & Appli., 2000, Vol.16, No.2, pp13~18.
- [4] Hutchinson J. E., Fractals and Self-similarity, Indian Univ. Math. J., 1981, Vol.30, No.4, pp713~747.
- [5] Zhou Z. L., Hausdorff measure of Self-similar sets-koch curve, Science in China, Ser. A, 1998, Vol.28, No.2, pp103~107 (in Chinese).
- [6] Whyburn G. T., Topological analysis, Princeton University Press, 1958.

Cantor 集的自乘积集的 Hausdorff 测度的下界*

贾保国 周作领 朱智伟

一、引言

确定分形集的 Hausdorff 维数与测度是分形几何研究中的重要内容之一。一般地说,要计算分形集的 Hausdorff 维数尤其是其测度是相当困难的,满足开集条件的自相似集的 Hausdorff 维数已经确定出来^[1],但对其 Hausdorff 测度的计算却仍然是极其困难的^[2,3,4]。文[5]中研究了 Cantor 集 C 的自乘积集 $C \times C$ (维数大于 1) 的 Hausdorff 测度,得到了 $H(C \times C)$ 的一个估计:

$$1 \leq H^{\log_3 4}(C \times C) < 1.50288$$

本文进一步研究了 $H^{\log_3 4}(C \times C)$ 的下界,证明了

$$H^{\log_3 4}(C \times C) \geq 1.48329$$

从上面两个不等式,得到 $H(C \times C)$ 的估计误差小于 $1.50288 - 1.48329 = 0.01959$ 。

二、三分 Cantor 集 C 的自乘积集 $C \times C$ 的 Hausdorff 测度

在平面 R^2 上取单位正方形 E_0 ,在 E_0 上除了保留四个角上边长为 $\frac{1}{3}$ 的小正方形外删去其余部分,得到四个边长为 $\frac{1}{3}$ 的小正方形的集合记为 E_1 ;对 E_1 中的每个正方形重复上述过程,得到的集合记为 E_2 ;无限重复上述过程,得到:

$$E_0 \supset E_1 \supset E_2 \supset \dots \supset E_n \supset \dots$$

非空集合 $C \times C = \bigcap_{n=0} E_n$ 称为 Cantor 集 C 的自乘积集。它是由四个压缩比为 $\frac{1}{3}$ 的压缩函数生成的自相似集且满足开集条件,因此 $C \times C$ 的 Hausdorff 维数 $s = \dim_H(C \times C)$ 满足 $4 \times (\frac{1}{3})^s = 1$,即 $1 < s = \dim_H(C \times C) = \log_3 4$ 。

定理 2.1 Cantor 集的自乘积集 $C \times C$ 的 Hausdorff 测度满足不等式:

$$H^{\log_3 4}(C \times C) \geq 1.48329$$

* 基金项目:国家自然科学基金(10041005)、教育部博士点基金(1999055810)、中山大学高等学术研究中心基金(01M2)和广东省自然科学基金(011221)资助的项目。原载:《数学年刊 A 辑》,2003 年第 5 期。

命题 2.1^[1] 设 $s > 1$, 则对任意 $x_i \in (0, 1) (i = 1, 2, \dots, n)$, 有 $(\frac{1}{n} \sum_{i=1}^n x_i)^s \leq$

$$\frac{1}{n} \sum_{i=1}^n x_i^s.$$

对 $n \geq 1$, E_n 由 4^n 个边长为 $\frac{1}{3^n}$ 的正方形组成, 每个这样的正方形称为基本正方形,

记为 Ω_n , 在 $C \times C$ 上定义分布函数 μ , 满足:

$$\begin{cases} \mu(E_0) = 2^{\frac{s}{2}} \\ \mu(\Omega_n) = \frac{1}{4^n} \cdot 2^{\frac{s}{2}} \quad n = 0, 1, 2, \dots \\ \mu(E_0 - C \times C) = 0 \end{cases}$$

则 μ 为 E_0 上一个测度且 μ 为 $C \times C$ 上的一个质量分布。

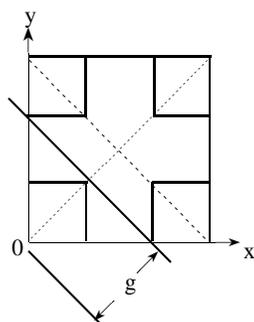


图 1

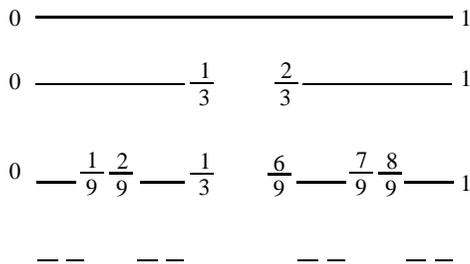


图 2

图 1 建立直角坐标系, E_0 的主和斜对角线方程分别为 $y = x$, $y + x = 1$. $C \times C$ 与 $[0, 1]$ 区间的交是三分 Cantor 集, $E_n \cap [0, 1]$ 由 2^n 个长为 $\frac{1}{3^n}$ 的闭区间组成 (见图 2),

每个这样的闭区间称为基本区间, 记为 I_n . 在 $E_n \cap [0, 1]$ 上定义分布函数 μ_1 , 满足:

$$\begin{cases} \mu([0, 1]) = 1 \\ \mu(I_n) = \frac{1}{2^n} \quad n = 1, 2, 3, \dots \\ \mu([0, 1] - C) = 0 \end{cases}$$

则 μ 为 $[0, 1]$ 上一个测度且 μ_1 为 C 上的一个质量分布. $\mu([0, x])$ 就是著名的 Cantor 函数, 它在 $[0, 1]$ 上单调增加且连续。

引理 2.1 对任意 $x \in (0, 1)$, 则 $\mu_1([0, x]) \geq \frac{2}{2^s} x^s$, 其中 $s = \log_3 4$.

证 根据 $\mu_1([0, x])$ 与 $\frac{2}{2^s} x^s$ 的连续性, 只需对 x 取 $I_n (n > 0)$ 的端点的值证明即可。容易

知道 $\frac{\mu_1([0, x])}{x^s}$ 只可能在基本区间 $I_n (I_n \neq [0, \frac{1}{3^n}])$ 的左端点 x_n 取得最小值。下面用数学

归纳法证明 : 当 $n = 1$ 时 , 有 :

$$x_1 = \frac{2}{3} \cdot \frac{\mu_1([0, x_1])}{x_1^s} = \frac{\frac{1}{2}}{\left(\frac{2}{3}\right)^s} = \frac{2}{2^s}$$

假设当 $n = k$ 时 , 引理成立 , 即对基本区间 I_k ($I_k \neq [0, \frac{1}{3^k}]$) 的左端点 x_k 都有 $\mu_1([0, x_k]) \geq \frac{2}{2^s} x_k$ 。当 $n = k + 1$ 时 , 有 :

(i) 若 $I_{k+1} = [\frac{2}{3^{k+1}}, \frac{1}{3^k}]$, 即 $x_{k+1} = \frac{2}{3^{k+1}}$, 则 :

$$\frac{\mu_1([0, x_{k+1}])}{x_{k+1}^s} = \frac{\frac{1}{2^{k+1}}}{\left(\frac{2}{3^{k+1}}\right)^s} = \frac{2^{k+1}}{2^s} \geq \frac{2}{2^s}$$

(ii) 若 $x_{k+1} = x_k$, 则由归纳假设知 , 引理成立。

(iii) 若 $x_{k+1} = x_k + \frac{2}{3^{k+1}}$, 则 $\mu_1([0, x_{k+1}]) = \mu_1([0, x_k]) + \frac{1}{2^{k+1}}$, 故 :

$$\mu_1([0, x_{k+1}]) - \frac{2}{2^s} \cdot x_{k+1}^s = \mu_1([0, x_k]) + \frac{1}{2^{k+1}} - \frac{2}{2^s} x_{k+1}^s \geq \frac{2}{2^s} x_k^s + \frac{1}{2^{k+1}} - \frac{2}{2^s} \left(x_k + \frac{2}{3^{k+1}}\right)^s$$

令 $f(x) = \frac{2}{2^s} x^s + \frac{1}{2^{k+1}} - \frac{2}{2^s} \left(x + \frac{2}{3^{k+1}}\right)^s$, 则当 $x \in [0, 1]$ 时 , $f'(x) \leq 0$, 从而 $f(x)$ 在 $[0, 1]$ 上是单调递减的。注意到 $x_k \leq 1 - \frac{1}{3^k}$, 由上式和拉格朗日中值定理 , 存在 $\zeta \in [1 - \frac{1}{3^k}, 1 - \frac{1}{3^{k+1}}]$, 使得 :

$$\begin{aligned} \mu_1([0, x_{k+1}]) - \frac{2}{2^s} \cdot x_{k+1}^s &\geq \frac{2}{2^s} \left(1 - \frac{1}{3^k}\right)^s + \frac{1}{2^{k+1}} - \frac{2}{2^s} \left(1 - \frac{1}{3^{k+1}}\right)^s \\ &= \frac{1}{2^{k+1}} - \frac{2}{2^s} \zeta^{s-1} \cdot \frac{2}{3^{k+1}} \geq \frac{1}{2^{k+1}} \left[1 - \frac{4s}{2^s} \cdot \left(\frac{2}{3}\right)^{k+1}\right] \geq \frac{1}{2^{k+1}} \left[1 - \frac{4s}{2^s} \cdot \frac{4}{9}\right] > 0 \end{aligned}$$

所以当 $n = k + 1$ 时 , 定理成立。

推论 2.1 对任意 $x \in \left(0, \frac{1}{3}\right]$, 则 $\mu_1([0, x]) \geq \frac{4s}{2^s} x^s$ 。

证 完全类似于上面引理的证明 , 从略。

设 G 为平行于 E_0 的斜对角线且与 E_0 相交的直线 , 记原点到 G 的距离 $d((0, 0), G) = g \geq 0$, 记 G 与 x 轴 , y 轴围成的三角形为 \triangle_g (见图 1)。

引理 2.2 $\mu(\triangle_g) \geq 0.9836 g^s$, $0 \leq g \leq \frac{\sqrt{2}}{3}$ 。

证 对 $0 \leq g \leq \frac{\sqrt{2}}{3}$, 存在自然数 k , 使得 $\frac{\sqrt{2}}{3^{k+1}} < g \leq \frac{\sqrt{2}}{3^k}$, 从而有 $\frac{\sqrt{2}}{3} < 3^k g \leq \sqrt{2}$ 。容易看出 $\mu(3^k \triangle_g) = 4^k \mu(\triangle_g)$, 并且 :

$$\frac{\mu(\Delta_{3^k \cdot g})}{(3^k \cdot g)^s} = \frac{4^k \mu(\Delta_g)}{4^k \cdot g^s} = \frac{\mu(\Delta_g)}{g^s}$$

因此只需证明, 当 $\frac{\sqrt{2}}{3} < g \leq \sqrt{2}$ 时, 引理是正确的即可。

任取自然数 n , 将 $A_1 B$ 分成 $2 \cdot 3^n$ 等份, 分点记为 $A_1, A_2, A_3, \dots, A_{2 \cdot 3^n}, A_{2 \cdot 3^n + 1} = B$ (见图 3)。易知:

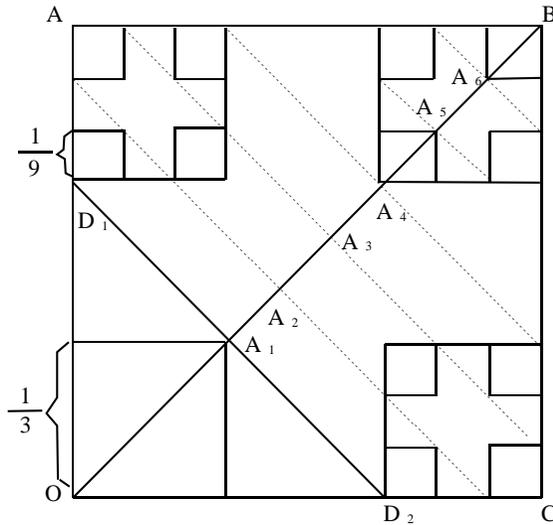


图 3

$$|AA_{i+1}| = \frac{1}{2 \cdot 3^n} \cdot \frac{2\sqrt{2}}{3} = \frac{\sqrt{2}}{3^{n+1}} \quad (i = 1, 2, \dots, 2 \cdot 3^n)$$

记:

$$g_{i,n} = |OA_i| = \frac{\sqrt{2}}{3} + \frac{(i-1)\sqrt{2}}{3^{n+1}} = \frac{\sqrt{2}}{3} \left(1 + \frac{i-1}{3^n} \right) \quad (i = 1, 2, \dots, 2 \cdot 3^n)$$

由于 $\frac{\sqrt{2}}{3} < g \leq \sqrt{2}$, 故存在 $i \in \{(i = 1, 2, \dots, 2 \cdot 3^n)\}$, 使得 $g_{i,n} < g \leq g_{i+1,n}$ 。从而:

$$g^s \leq g_{i+1,n}^s = \frac{\sqrt{2^s}}{3^s} \left(1 + \frac{i}{3^n} \right)^s = \frac{\sqrt{2^s}}{4} \left(1 + \frac{i}{3^n} \right)^s$$

$$\mu(\Delta_g) \geq \frac{\sqrt{2^s}}{4} + \frac{l_n^i \sqrt{2^s}}{4^{n+1}} = \frac{\sqrt{2^s}}{4} \left(1 + \frac{l_n^i}{4^n} \right) \geq \left(\frac{3^n}{3^n + i} \right)^s \left(1 + \frac{l_n^i}{4^n} \right) g^s$$

其中: l_n^i 表示在直线 G_i 和 $D_1 D_2$ 之间所包含的 $\frac{1}{3^{n+1}}$ -基本正方形 Ω_{n+1} 的个数, G_i 平行于 $D_1 D_2$ 且过等分点 A_i 。如图 3 中, $n=1, l_1^1=0, l_1^2=2, l_1^3=6, l_1^4=8, l_1^5=9, l_1^6=11$ 。记

$$h(n, i) = \left(\frac{3^n}{3^n + i} \right)^s \left(1 + \frac{l_n^i}{4^n} \right) = \frac{4^n + l_n^i}{(3^n + i)^s}。则:$$

$$\frac{\mu(\Delta_g)}{g^s} \geq h(n, i), g_{i,n} < g \leq g_{i+1,n} \quad (i = 1, 2, \dots, 2 \cdot 3^n)$$

从而,对任给的自然数 n ,有:

$$\inf_{\frac{\sqrt{2}}{3} < g \leq \sqrt{2}} \frac{\mu(\Delta_g)}{g^s} \geq \min_{1 \leq i \leq 2 \cdot 3^n} \{h(n, i)\}$$

当 $n = 1$ 时,容易算出:

$$h(1, 1) = \frac{4}{4^s} \geq 0.69557597, h(1, 2) = \frac{6}{5^s} \geq 0.78731578$$

$$h(1, 3) = \frac{10}{6^s} \geq 1.04251440, h(1, 4) = \frac{12}{7^s} \geq 1.02987812$$

$$h(1, 5) = \frac{13}{8^s} \geq 0.94269286, h(1, 6) = \frac{15}{9^s} \geq 0.9375$$

因此:

$$\inf_{\frac{\sqrt{2}}{3} < g \leq \sqrt{2}} \frac{\mu(\Delta_g)}{g^s} \geq \min_{1 \leq i \leq 2 \cdot 3^n} \{h(n, i)\} = \frac{4}{4^s} \geq 0.69557597$$

同理可得:

$$\text{当 } n = 2 \text{ 时, } \inf_{\frac{\sqrt{2}}{3} < g \leq \sqrt{2}} \frac{\mu(\Delta_g)}{g^s} \geq \min_{1 \leq i \leq 2 \cdot 3^n} \{h(n, i)\} = \left(\frac{9}{11}\right)^s \cdot \frac{9}{8} \geq 0.87333576;$$

$$\text{当 } n = 3 \text{ 时, } \inf_{\frac{\sqrt{2}}{3} < g \leq \sqrt{2}} \frac{\mu(\Delta_g)}{g^s} \geq \min_{1 \leq i \leq 2 \cdot 3^n} \{h(n, i)\} = \left(\frac{27}{31}\right)^s \cdot \frac{9}{8} \geq 0.94233021;$$

$$\text{当 } n = 4 \text{ 时, } \inf_{\frac{\sqrt{2}}{3} < g \leq \sqrt{2}} \frac{\mu(\Delta_g)}{g^s} \geq \min_{1 \leq i \leq 2 \cdot 3^n} \{h(n, i)\} = \left(\frac{81}{91}\right)^s \cdot \frac{288}{256} \geq 0.97039446.$$

容易证明:当 n 增加时, $\min_{1 \leq i \leq 2 \cdot 3^n} \{h(n, i)\}$ 随之增大.故极限 $\lim_{n \rightarrow \infty} \min_{1 \leq i \leq 2 \cdot 3^n} \{h(n, i)\}$ 存在.

为得到函数 $\frac{\mu(\Delta_g)}{g^s}$ 在区间 $[\frac{\sqrt{2}}{3}, \sqrt{2}]$ 上更好的下界,须对更大的 n 计算 $\min_{1 \leq i \leq 2 \cdot 3^n} \{h(n, i)\}$.

将 $A_i A_{i+1}$ 分成 3 等份,容易得出下面的递推公式:

$$I_{n+1}^{3i+1} = 3I_n^i + I_n^{i+1}, J_{n+1}^{3i+2} = I_n^i + 3I_n^{i+1} \quad (i = 1, 2, \dots, 2 \cdot 3^n; n = 1, 2, \dots)$$

借助于上面的递推公式和简单的数值计算,得:

$$\text{当 } n = 5 \text{ 时, } \min_{1 \leq i \leq 2 \cdot 3^5} \{h(5, i)\} \approx 0.97891182;$$

$$\text{当 } n = 6 \text{ 时, } \min_{1 \leq i \leq 2 \cdot 3^6} \{h(6, i)\} \approx 0.98216360;$$

$$\text{当 } n = 7 \text{ 时, } \min_{1 \leq i \leq 2 \cdot 3^7} \{h(7, i)\} \approx 0.98317099;$$

$$\text{当 } n = 8 \text{ 时, } \min_{1 \leq i \leq 2 \cdot 3^8} \{h(8, i)\} \approx 0.98351435;$$

$$\text{当 } n = 9 \text{ 时, } \min_{1 \leq i \leq 2 \cdot 3^9} \{h(9, i)\} \approx 0.98362401;$$

$$\text{当 } n = 10 \text{ 时, } \min_{1 \leq i \leq 2 \cdot 3^{10}} \{h(10, i)\} \approx 0.98365950.$$

...

上面的所有数都精确到八位小数.从而有:

$$\inf_{\frac{\sqrt{2}}{3} < g \leq \sqrt{2}} \frac{\mu(\Delta_g)}{g^s} \geq \lim_{n \rightarrow \infty} \min_{1 \leq i \leq 2 \cdot 3^n} \{h(n, i)\} \geq \min_{1 \leq i \leq 2 \cdot 3^{10}} \{h(10, i)\} > 0.9836$$

注 2.1 易知, 当 $g = \frac{\sqrt{2}}{3} \left(1 + \frac{1}{9}\right)$ 时, $\mu(\Delta_g) = \frac{\sqrt{2}^s}{4} \left(1 + \frac{1}{8}\right)$ 且:

$$\frac{\mu(\Delta_g)}{g^s} = \frac{9}{8} \cdot \left(\frac{9}{10}\right)^s \approx 0.984947$$

这说明引理 2.2 中的常数 0.9836 与 $\inf_{\frac{\sqrt{2}}{3} < g \leq \sqrt{2}} \frac{\mu(\Delta_g)}{g^s}$ 是比较接近的。

引理 2.3 对任何可测集合 V , 有 $\mu(V) \leq 1.044 |V|^s$.

证 不妨设 $V \subset E_0$, 否则用 $V \cap E_0$ 代替。

(1) 若 V 与 E_1 中四个基本正方形相交, 作平行于主和斜对角线的直线 G_1, G_2, G_3, G_4 使 $V \subset \Omega_V$, 且 V 与 Ω_V 的四边均相交 (见图 4) 其中 Ω_V 为由直线 G_1, G_2, G_3, G_4 围成的矩形。

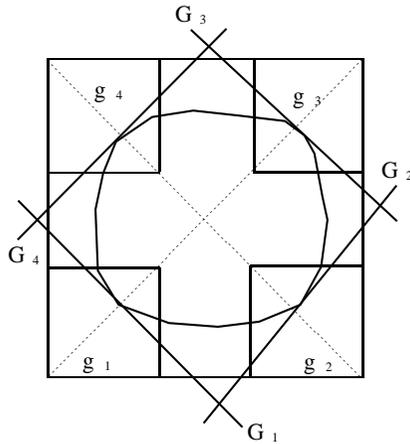


图 4

记距离:

$$d((0, 0), G_1) = g_1, d((1, 0), G_2) = g_2, d((0, 1), G_3) = g_3, d((0, 1), G_4) = g_4$$

则 $|V| \geq \sqrt{2} - g_1 - g_3, |V| \geq \sqrt{2} - g_2 - g_4, |V| \geq \sqrt{2} - \frac{1}{2}(g_1 + g_2 + g_3 + g_4)$ 。因此:

$$1.044 |V|^s \geq 1.044 \sqrt{2}^s \left[1 - \frac{\sqrt{2}}{4}(g_1 + g_2 + g_3 + g_4)\right]^s$$

由引理 2.2 得:

$$\mu(V) \leq \sum_{k=1}^4 \left[\frac{\sqrt{2}^s}{4} - \mu(\Delta_{g_k}) \right] = \sqrt{2}^s - \sum_{k=1}^4 \mu(\Delta_{g_k}) \leq \sqrt{2}^s - 0.9836(g_1^s + g_2^s + g_3^s + g_4^s)$$

令:

$$f(g_1, g_2, g_3, g_4) = 1.044 \sqrt{2}^s \left[1 - \frac{\sqrt{2}}{4}(g_1 + g_2 + g_3 + g_4)\right]^s + 0.9836(g_1^s + g_2^s + g_3^s + g_4^s) - \sqrt{2}^s$$

其中 $0 < g_i \leq \frac{\sqrt{2}}{3}$ ($i = 1, 2, 3, 4$)。对于 $f(g_1, g_2, g_3, g_4)$ 求偏导数并令 $f'_{g_1} = 0, f'_{g_2} = 0, f'_{g_3} = 0, f'_{g_4} = 0$ 得到 $f(g_1, g_2, g_3, g_4)$ 的稳定点 $(g_1^0 + g_2^0 + g_3^0 + g_4^0)$ 其中:

$$(g_1^0 + g_2^0 + g_3^0 + g_4^0) = \left[\sqrt{2} + \left(\frac{3.9344}{1.044\sqrt{2^s}\sqrt{2}} \right)^{\frac{1}{s-1}} \right]^{-1} \approx 0.10681788$$

令 $D_1 = \{(g_1, g_2, g_3, g_4) | 0 \leq g_1 + g_2 + g_3 + g_4 \leq \frac{4}{3}\sqrt{2}\}$ 。下面证明 $f(g_1, g_2, g_3, g_4)$ 在闭区域 D_1 上的最小值大于零。由于:

$$f(g_1^0, g_2^0, g_3^0, g_4^0) \geq 1.31486140 + 0.23397162 - 1.54856292 = 0.0002701 > 0$$

在 D_1 的边界上, 当 $g_1 + g_2 + g_3 + g_4 = \frac{4}{3}\sqrt{2}$ 时, 由命题 2.1 得到:

$$\begin{aligned} & f(g_1, g_2, g_3, g_4) \\ &= 1.044\sqrt{2^s} \left[1 - \frac{\sqrt{2}}{4}(g_1 + g_2 + g_3 + g_4) \right]^s + 0.9836(g_1^s + g_2^s + g_3^s + g_4^s) - \sqrt{2^s} \\ &\geq \frac{1.044}{4}\sqrt{2^s} + 0.9836 \times 4 \left(\frac{g_1 + g_2 + g_3 + g_4}{4} \right)^s - \sqrt{2^s} \\ &= \left(\frac{1.044}{4} + 0.9836 - 1 \right) \sqrt{2^s} > 0 \end{aligned}$$

当 $g_1 + g_2 + g_3 + g_4 = 0$ 时, $g_1 = g_2 = g_3 = g_4 = 0$ 。故 $f(0, 0, 0, 0) = (1.044 - 1)\sqrt{2^s} > 0$ 。因此:

$$1.044|V|^s - \mu(V) \geq f(g_1, g_2, g_3, g_4) > 0$$

(2) 若 V 与 E_1 中三个基本正方形相交, 而与另一个不相交, 不妨设 V 与位于左上角的基本正方形不相交 (见图 4), 则:

$$|V| \geq \sqrt{2} - g_1 - g_3, 1.044|V|^s \geq 1.044[\sqrt{2} - g_1 - g_3]^s$$

由引理 2.2 得:

$$\mu(V) \leq 3 \cdot \frac{\sqrt{2^s}}{4} - [\mu(\Delta_{g_1}) + \mu(\Delta_{g_2}) + \mu(\Delta_{g_3}) + \mu(\Delta_{g_4})] \leq 3 \cdot \frac{\sqrt{2^s}}{4} - 0.9836(g_1^s + g_2^s + g_3^s)$$

因此:

$$1.044|V|^s - \mu(V) \geq 1.044[\sqrt{2} - g_1 - g_3]^s + 0.9836(g_1^s + g_3^s) + 0.9836g_2^s - 3 \cdot \frac{\sqrt{2^s}}{4}$$

令 $f(g_1, g_3) = 1.044[\sqrt{2} - (g_1 + g_3)]^s + 0.9836(g_1^s + g_3^s) - 3 \cdot \frac{\sqrt{2^s}}{4}$ 其中 $0 < g_i \leq \frac{\sqrt{2}}{3}$ ($i = 1, 3$)。对 $f(g_1, g_3)$ 求偏导数并令 $f'_{g_1} = 0, f'_{g_3} = 0$ 。可以得到 $f(g_1, g_3)$ 的稳定点 (g_1^0, g_3^0) 其中:

$$g_1^0 = g_3^0 = \sqrt{2} \left[2 + \left(\frac{0.9836}{1.044} \right)^{\frac{1}{s-1}} \right]^{-1} \approx 0.50571669$$

令 $D_2 = \{(g_1, g_3) | 0 \leq g_1 + g_3 \leq \frac{2}{3}\sqrt{2}\}$ 。下面证明 $f(g_1, g_3)$ 在闭区域 D_2 上的最小值

大于零。因为 $(g_1^0, g_3^0) \notin D_2$, 所以 $f(g_1, g_3)$ 在 D_2 上的最小值点应位于 D_2 的边界上, 类似于情形(1)的讨论, 由命题 2.1 容易证明在 D_2 的界上, $f(g_1, g_3)$ 大于零。故:

$$1.044|V|^s - \mu(V) \geq f(g_1, g_3) > 0$$

(3) 若 V 仅与 E_1 中两个基本正方形 Ω_1 和 Ω_1 相交。当 Ω_1 和 Ω_1 位于对角线上, 不妨设主对角线上(见图 4), 则类似情形(2), 有 $1.044|V|^s - \mu(V) \geq f(g_1, g_3) > 0$ 。

当 Ω_1 和 Ω_1 位于 E_0 中同一条边上, 如图 5 所示, 记:

$$a = \min\{x : (x, y) \in V \cap (\Omega_1 \cup \Omega_1)\}, b = 1 - \min\{x : (x, y) \in V \cap (\Omega_1 \cup \Omega_1)\}, \\ \Omega_a = \{(x, y) | x \leq a, (x, y) \in \Omega_1\}, \Omega_b = \{(x, y) | x \leq b, (x, y) \in \Omega_1\}$$

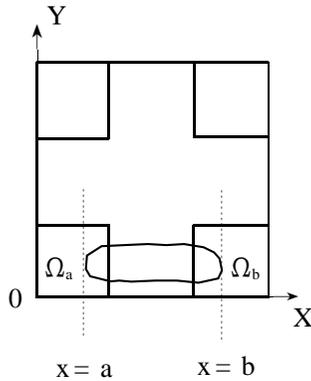


图 5

则 $|V| \geq 1 - a - b$, $1.044|V|^s \geq 1.044[1 - a - b]^s$ 。由推论 2.1 和 $0 \leq a, b \leq \frac{1}{3}$, 得到:

$$\mu(\Omega_a) \geq \frac{2\sqrt{2^s}}{\left(\frac{2}{9}\right)^s} a^s = \frac{2}{\sqrt{2^s}} a^s, \mu(\Omega_b) \geq \frac{2}{\sqrt{2^s}} b^s$$

因此:

$$\mu(V) \leq 2 \cdot \frac{\sqrt{2^s}}{4} - [\mu(\Omega_a) + \mu(\Omega_b)] \leq \frac{\sqrt{2^s}}{2} - \frac{2}{\sqrt{2^s}}(a^s + b^s)$$

故:

$$1.044|V|^s - \mu(V) \geq 1.044[1 - a - b]^s + \frac{2}{\sqrt{2^s}}(a^s + b^s) - \frac{\sqrt{2^s}}{2}$$

令 $f(a, b) = 1.044[1 - a - b]^s + \frac{2}{\sqrt{2^s}}(a^s + b^s) - \frac{\sqrt{2^s}}{2}$ 。对 $f(a, b)$ 求偏导数并令 $f_a = 0, f_b = 0$, 可得到 $f(a, b)$ 的稳定点 (a^0, b^0) , 其中:

$$a^0 = b^0 = \left[2 + \left(\frac{2}{1.044\sqrt{2^s}} \right)^{\frac{1}{s-1}} \right]^{-1} \approx 0.23509874$$

令 $D_3 = \{(a, b) | 0 \leq a + b \leq \frac{2}{3}\}$ 。类似于情形(1)的证明方法, 易证 $f(a, b)$ 在闭区域

D_3 上的最小值大于零。因此 $1.044|V|^s - \mu(V) \geq f(a, b) > 0$ 。

(4) 若 V 仅与 E_1 中一个基本正方形 Ω_1 相交, 则 $\mu(V) = \mu(V \cap \Omega_1)$ 。若 $V \cap \Omega_1$ 仅与 E_2 中一个基本正方形 Ω_2 相交, 则 $\mu(V) = \mu(V \cap \Omega_2)$ 。若 $V \cap \Omega_1$ 仅与 E_2 中二个、三个、四个基本正方形相交, 则由上面的情形(1)、(2)、(3)可得:

$$\mu(V) = \mu(V \cap \Omega_1) \leq 1.044|V \cap \Omega_1|^s \leq 1.044|V|^s$$

由数学归纳法, 或者 $\mu(V) \leq 1.044|V|^s$ 或者存在 $\Omega_n \subset E_n$, 使得:

$$\mu(V) = \mu(V \cap \Omega_n) \leq \mu(\Omega_n) = \frac{\sqrt{2}^s}{4^n}$$

令 $n \rightarrow \infty$ 得到 $\mu(V) = 0 \leq 1.044|V|^s$ 。

定理的证明 由质量分布原理^[1]和引理 2.3 得到 $1.044H^s(C \times C) \geq \mu(C \times C) =$

$\sqrt{2}^s$ 。从而 $H^s(S) \geq \frac{\sqrt{2}^s}{1.044} \geq 1.48329$ 。

参考文献

- [1] Falconer, K. J., Techniques in fractal geometry, John and Sons, New York, 1997.
- [2] Zhou Zuoling & Wu Min, The Hausdorff measure of Sierpinski carpet, Sci. China A., 1998, Vol.42, pp673 ~ 680.
- [3] Zhou Zuoling, The Hausdorff measure of self-similar sets-the Koch curve, Sci. China A., 1997, Vol.41, pp723 ~ 728.
- [4] Zhou Zuoling & Feng Li, A new estimate of the Hausdorff measure of the Sierpinski gasket, Nonlinearity, Vol.13, pp479 ~ 491.
- [5] Jia Baoguo, Zhou Zuoling & Zhu Zhiwei, An estimate of the Hausdorff measure of the Cartesian product of the middle third Cantor set with itself, Acta Math Sinica, 2003, Vol.46, No.6 (in Chinese).