

第6卷·第3期

(总第25期)

Vol. 6 No. 3

经济学(季刊)

China Economic Quarterly

北京大学中国经济研究中心

北京大学出版社

2007.4

图书在版编目(CIP)数据

经济学(季刊). 第6卷. 第3期/林毅夫 姚洋主编. —北京: 北京大学出版社, 2007. 4

ISBN 978-7-301-12127-6

I. 经... II. ①林... ②姚... III. 经济学-丛刊 IV. F0-55

中国版本图书馆 CIP 数据核字(2007)第 064771 号

书 名: 经济学(季刊) 第6卷·第3期

著作责任者: 林毅夫 姚 洋 主编

责任编辑: 朱启兵

标准书号: ISBN 978-7-301-12127-6/F·1608

出版发行: 北京大学出版社

地 址: 北京市海淀区成府路 205 号 100871

网 址: <http://www.pup.cn>

电子邮箱: ceq@ccer.pku.edu.cn

电 话: 邮购部 62752015 发行部 62750672 编辑部 62758908
出版部 62754962

印 刷 者: 北京大学印刷厂

经 销 者: 新华书店

787 mm×1092 mm 16 开本 19.25 印张 371 千字

2007 年 4 月第 1 版 2007 年 4 月第 1 次印刷

国内定价: 48.00 元

International Price: US \$ 24.00

未经许可, 不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有, 侵权必究

举报电话 010-62752024 电子邮箱 fd@pup.pku.edu.cn

2005—2006 年度《经济学(季刊)》最佳论文奖

为了鼓励规范的经济学研究,《经济学(季刊)》特设立年度最佳论文奖,每年度从《经济学(季刊)》发表的文章中评选出最佳论文一篇,奖金为人民币5000元。获奖论文由本刊学术委员会在提名论文中以多数投票的形式决定。

2005—2006 年度《经济学(季刊)》最佳论文奖提名论文是(排名不分先后):

张晏、龚六堂:“分税制改革、财政分权与中国经济增长”(第5卷第1期)

唐国正:“股权二元结构下配股对股权价值的影响”(第5卷第2期)

高雷、何少华、黄志忠:“公司治理与掏空”(第5卷第4期)

2005—2006 年度《经济学(季刊)》最佳论文奖授予复旦大学经济学院张晏和北京大学光华管理学院龚六堂发表于《经济学(季刊)》第5卷第1期的“分税制改革、财政分权与中国经济增长”。

《经济学(季刊)》

二〇〇七年四月

经 济 学 (季 刊)

第 6 卷 第 3 期

(总第 25 期)

2007 年 4 月

目 录

论 文

我国资本回报率估测(1978—2006)

——新一轮投资增长和经济景气微观基础

..... CCER“中国经济观察”研究组 (723)

择优分配最优化原理及其应用 茅于軾 (759)

我国金融机构的存差:历史沿革和含义变化 汪 洋 (781)

我国制造业区域集聚程度决定因素的研究 路江涌 陶志刚 (801)

中国城市劳动力市场的一体化进程 杨 涛 盛柳刚 (817)

中国农村医疗保障制度:一项基于异质性个体决策行为的理论研究

..... 封 进 宋 铮 (841)

网上交易中的私人秩序

——社区、声誉与第三方中介 吴德胜 (859)

纵向关联市场间的价格传递

..... 王秀清 H. T. Weldegebriel A. J. Rayner (885)

高考录取机制的博弈分析 聂海峰 (899)

近代外国在华投资:规模与效应分析 杨德才 (917)

经济学的性质与哲学视角审视下的经济学

——一个基于经济思想史的理论回顾与展望 韦 森 (945)

综 述

本地市场效应:理论和经验研究的新近进展 钱学锋 梁 琦 (969)

“人以群分”:非市场互动和群分效应的文献评论 陆 铭 张 爽 (991)

我国资本回报率估测(1978—2006)

——新一轮投资增长和经济景气微观基础

CCER“中国经济观察”研究组*

摘要 本文估测我国改革开放时期工业企业资本回报率九个系列指标,并在此基础上系统观察我国资本回报率长期走势特点、资本回报率与经济景气短期波动关系、物价变动因素对资本回报率影响、不同类型企业资本回报率以及中外资本回报率比较等问题。本文发现的经验证据提示我国新一轮投资较快增长和经济景气具有相当程度微观基础,因而对反思和探讨宏观调控政策措施选择问题具有借鉴意义。

关键词 中国资本回报率, 中外资本回报率国际比较, 经济利润率

一、我国资本回报率讨论:问题的提出

过去近30年,我国改革开放带动经济起飞,取得举世瞩目成就。与东亚和其他经济体发展经验相类似,投资增长对推动我国经济成长发挥了关键作用,因而资本回报率变动构成观察经济运行效率的重要微观指标。我国经济起飞时期资本回报率水平和走势如何?资本回报率走势成因是什么?它们与经济成长其他方面结构特点存在什么内在联系?考察这些问题对理解我国经济转型发展的具体规律,对分析近年宏观经济运行表现和演变前景,对评估相关政策和干预措施的利弊得失,都具有认识借鉴意义。

从投资效率视角考察经济运行的需要,使学术界近年越发重视研究资本回报率。虽然人们对资本回报率作为投资评估指标的分析价值存在共识,然而在该指标数量水平和具体走势定量估测和判断上存在较多分歧。我国体制转型伴随经济统计指标体系不断调整改进,在资本存量和利润指标时间序列数据可获得性上存在特殊困难;即便存在相关指标数据,由于统计口径和方法演变,如何解读这些数据也会发生疑问。资本回报率研究展开过程伴随不同意见交流,并在不久前引发一场引起国内外媒体广泛关注的讨论。

2006年5月,世界银行中国代表处(下面简称世行)在一份有关中国经济运行的季度分析报告中简略评论到中国资本回报率上升数据(世界银行中

* 执笔人:卢锋;研究人员:宋国青、卢锋、唐杰、赵洪岩、刘鏊。CCER“中国经济观察”研究组此前完成一份题为“经济转型成长与资本回报率演变——我国改革开放时期资本回报率估测(1978—2005)”的研究报告初稿,本文是这份报告部分章节的修改稿。2007年元月11日CCER“中国经济观察”举办“中国资本回报率:事实、原因和政策含义”研讨会,笔者感谢参会人员对该报告初稿的评论和研讨。

国代表处, 2006), 该机构研究人员此前的专题报告观察并评论了我国企业利润增长对投资资金来源的影响 (世界银行中国代表处, 2005, Kuijs, 2005; 2006)。宋国青教授报告近年我国资本回报率上升情况, 讨论了改革开放以来资本回报率先降后升形态及其与经济景气变动的关系, 并由此对我国近年投资率过高观点提出反思和质疑 (宋国青, 2006)。梁红博士观察到近年工业企业回报率上升超过市场预期, 并考察企业投资与利润存留关系, 对中外企业回报率水平进行比较, 在此基础上提出中国投资高速增长具有可持续性 (梁红, 2006)。

主要针对上述世行报告内容, 单伟建博士在多篇文章和记者访谈中提出了分析和评论 (Shan, 2006a; 2006b; 2006c; 单伟建, 2006a; 2006b; 2006c; 全秋梅, 2006)。单伟建博士主要阐述了三个层面的观点。一是在肯定近年我国工业资本回报率增长的同时, 从不同方面分析有关统计数据可能高估资本回报率实际水平, 说明世行季报大大高估我国资本回报率。二是对企业投资较大部分来自利润留成观点提出质疑。三是评论世行报告内容对我国政府加强金融风险管理和宏观调控会产生负面政策影响。单文观点明快而言辞犀利, 率先在《远东经济评论》、《华尔街时报》等国外重要报刊发表后, 很快在国内外媒体上引发广泛回应、评论和跟踪报道。

世行报告作者 (Kuijs and Hofman, 2006a; 2006b; 2006c) 对单伟建博士观点给以回应, 认为单文误解了相关统计数据 and 经济学概念, 没有正确表达他们的观点。其他评论人员则采取不同立场。有的研究人员倾向于支持单伟建观点, 如罗奇先生在专题评论文章中表示同意单伟建博士提出的批评而不能接受世行报告观点 (Roach, 2006)。也有接受我国企业资本利润率强劲增长的判断, 认同我国宏观经济并未特别过热的观点 (赵晓, 2006)。还有评论文章概述这一问题讨论的重要意义并对不同观点做出了比较中性的评价 (肖耿, 2006; 胡鲁滨、胡蛟, 2006)。上述讨论使“资本回报率”这个通常“行而不远”的财务指标一时成为财经媒体上的热点词汇。¹

资本回报率争论实质涉及三个层面问题: 一是回报率增长速度有多快以及目前水平有多高? 二是给定资本回报率增长, 目前投资较快增长是否严重过热? 三是给定资本回报率和投资增长动因判断, 应当采用什么政策措施调节管理投资和总需求增长? 三者各有特定内涵, 然而第一层面问题具有特殊重要性。观察我国经济起飞时期资本回报率演变走势和现实水平, 对于理解我国经济成长展开机制和具体规律具有超越直接政策讨论的认识价值。另外

¹ 前几年讨论宏观经济形势也有研究和评论涉及从投资效率角度分析投资是否过热问题。有研究人员通过观察 ICOR 下降以及上市公司利润率下降等指标变动讨论投资过热判断 (许小年、肖倩, 2003)。也有研究人员考察我国上市公司和工业企业资产利润率等指标变动并与银行利率比较, 提出我国投资微观效率可能在不断改善的结论 (高善文, 2004)。

从三者关系看，仔细估测资本回报率是分析另外两个层面问题的基础。实际争论情况也显示，对我国资本回报率水平和增长情况“看淡”和“看重”观点的差异，与政策评论的不同倾向具有内在联系。

研究人员对我国资本回报率已提出初步度量。然而现有文献主要以形势分析、会议发言、报章评论或记者访谈等形式发表，对相关概念指标关系、具体数据内涵、历史变动及国际比较等问题受表达方式和研究目标限制难以细致考察，对资本回报率基本数据尚未进行系统整理、甄别和评估。要从学理层面寻求这一问题的共识基础，需要把资本回报率定量考察从媒体评论层面深入到学术研究层面，在深入分析相关数据和其他有关经验证据基础上，对我国资本回报长期演变真实图景求得比较符合实际的估测和判断。

以上述观察为背景，CCER“中国经济观察”课题组通过一组论文系统考察了我国改革开放时期资本回报率问题。本文侧重估测改革开放时期我国资本回报率数量水平并观察其变动的结构特点。全文分七节。第二节讨论不同资本回报率概念定义及其与相应度量指标的关系。第三节报告我国资本存量和资本回报时间序列数据，讨论说明整理估测资本回报数据序列涉及的若干问题。第四节估测九种资本回报率时间序列数据，观察不同经济类型企业回报率表现差异、资本回报率演变趋势及其与宏观景气变动关系等问题。第五节考察物价变动对会计资本回报率的影响，并通过调整物价因素作用估测真实经济回报率。第六节对中国、美国和日本资本回报率水平进行初步比较。第七节概括全文主要发现并评论其政策含义。

二、资本回报率及相关指标的定义和关系

为避免概念歧义，首先考察梳理相关概念定义及其对应统计指标关系，并讨论说明本文选择资本回报率估计方法和指标体系的思路和依据。本文不加区分地利用资本回报率、利润率、收益率等概念。资本回报率被定义为资本回报与创造回报所用资本直接度量值之间的数量比率关系。² 本文理解的资本回报率具有两方面常识性特点。计算回报率的资本回报是流量，资本是存量，因而资本回报率是流量与存量之间的比率。另外计算回报率的分子项资本回报与分母项资本存量，都利用直接度量统计数据测算。

依据数据来源不同，至少存在两类满足上述资本回报率概念的度量指标。一类是基于企业财务会计资本回报（如利润（profits before tax）、净利润（net profits）等）和资本存量（如净资产（net assets）或资产（asset）等）数据

² “根据2002年国家五部委修订《企业绩效评价操作细则》的指标解释：‘净资产收益率是企业一定时期内的净利润同平均净资产的比率。净资产收益率充分体现了投资者投入企业的自有资本获取净收益的能力，突出反映了投资与报酬的关系，是评价企业资本经营效益的核心指标’”（朱向东 2004，第58页）。

测算的资本回报率。另一类是基于国民收入账户统计体系中资本回报即“营业盈余 (operating surplus)”指标以及固定资产存量等指标计算的资本回报率。为行文便利,本文把第一类指标称为“以微观数据为基础的资本回报率 (micro-economic data based capital return)”或简称“微观资本回报率”,第二类称为“宏观数据为基础的资本回报率 (macro-economic data based capital return)”或简称“宏观资本回报率”。

两类资本回报率度量指标各有优点和局限。例如宏观资本回报率具有指标统计对象覆盖经济整体,资本存量经过物价调整因而度量结果更接近经济分析意义上真实回报率概念等优点。³微观资本回报率优点在于,指标定义结构简明清晰,便于直观解读,对指标值变动贡献因素可以通过分析相关经验证据加以追溯解析,指标度量值对经济基本面因素变动反应比较灵敏,企业财务月报统计能较快提供最新回报率数据信息等,另外我国资本回报率争论直接对象也是微观资本回报率。综合考虑,本文选用企业微观资本回报率指标作为估测研究对象。考虑到这一时期工业在我国经济中占据的特殊地位,也由于工业企业财务数据较为完整规范,本文进一步把考察估测对象限定为工业资本回报率。对会计回报率受物价变动因素影响问题,后面将专门调整并估测经济资本回报率。

依据常识,资本回报被定义为企业运用资本获得收入并扣除合同性成本费用后所形成的剩余。本文进一步把资本回报分解为资本所有者收益 (即利润) 和社会收益 (即相关政府税收) 两个部分。“资产 = 负债 - 权益”这个基本会计等式关系提供界定资本和资产存量的概念框架,会计损益表基本关系式“利润 = 收入 - 成本”给出定义资本回报指标的概念框架。国际通行会计体系定义了不同口径的利润指标,包括“毛利润 (gross profits)”、“营业利润 (operation profits)”、“税前利润 (profits before tax)”、“净利润 (net profits)”等。其中净利润指标直接衡量股东等资本所有者从资本运营中所获得的净回报,构成从资本所有者角度评估投资效益的最有实质意义的指标。从产权激励角度看,这一指标具有其他资本回报指标不可替代的功能。

不过净利润并非运用资本提供的全部剩余,资本剩余一部分以税收形式构成资本社会回报。企业所得税显然构成资本社会回报的组成部分。⁴此外还应考虑间接税以及企业补贴因素的影响。如果特定企业或特定行业整体获得补贴收入⁵,这会增加资本权益回报,因而考虑社会回报时应扣除这一转移收入。另外企业需要缴纳的交易税或增值税等间接税通常被企业和消费者共同

³ 白重恩、谢长泰和钱颖一(2007)提供了运用这一方法度量我国资本回报率的最新研究成果。

⁴ 我国财政部门对企业所得税定义为“国家参与企业利润分配,正确处理国家与企业分配关系的一个重要税种”(《中华人民共和国财政部税制税则司,1998,第55页)。该税种可以被理解为国家代表社会对于企业利用资本创造剩余的分享。

⁵ 如补贴利用某种节能材料,补贴到西部贫困地区投资,或者为控制失业对某些企业提供亏损补贴等。

负担，如能适当估计企业和消费者对间接税的负担比例，其中企业负担部分应被理解为资本社会回报的一部分。

上述理解可用下面关系式概括表达：

$$\begin{aligned}
 \text{资本总回报} &= \text{权益回报} + \text{社会回报} \\
 &= \text{净利润} + (\text{企业所得税} + \text{企业负担的间接税} - \text{净补贴收入}) \\
 &= \text{税前利润} + \text{企业负担间接税} - \text{净补贴收入} \quad (1)
 \end{aligned}$$

衡量资本回报率还需要资本存量数据。本文选择三种口径宽窄不同的指标。一是从资本所有者角度看，标准资本存量度量指标应是权益或股本（equity），相应获得权益或净资产回报率指标（return on equity, ROE）。二是从企业通过负债占用更大范围资源创造收入和回报的角度看，可以用资产替代权益定义资产回报率（return on assets, ROA）。三是考虑固定资产在资本构成中的特殊地位，还可以利用固定资产净值作为资本度量指标，定义固定资产回报率（return on capital, ROC）。

给定上述资本回报和资本存量度量指标，可以定义表1报告的九个资本回报率指标。其中净利润是资本所有者回报。总利润额指未扣除企业所得税的利润总额，即权益回报加上国家通过企业所得税参与资本剩余分配获得的部分社会回报。总回报是在利润总额基础上加上资本方承担的间接税，是资本权益和社会两重回报的度量指标，它与我国常规统计系统中利税总额指标的差异在于扣除了消费者承担的间接税部分。三个资本回报指标与权益、资产、固定资产净值等资本存量指标组合形成九个回报率指标，提供本文衡量资本回报率的指标体系。

表1 资本回报率若干指标定义

	权益	资产	固定资产净值
净利润	权益净利润率(ROE)	资产净利润率(ROA)	固定资产净利润率(ROC)
总利润	权益总利润率	资产总利润率	固定资产总利润率
总回报	权益总回报率	资产总回报率	固定资产总回报率

三、我国资本存量和资本回报数据和讨论

CCER“中国经济观察”研究组（2007）考察了我国工业财务和税制统计体制的演变情况，对资本回报率基本数据可获得性进行了评估。20世纪90年代初引入新企业会计制度以来，我国资本回报与资本存量数据已大体具有国际可比性以及历时可比性。20世纪80年代前后资本存量尚未系统采用资本权益和资产统计指标，仅有固定资产净值数据；这一时期利润数据还不够完整，不过可以获得与后来营业利润和税前利润大体对应的统计指标。税制改

革演变与本文考察主题最为相关,包括 80 年代初两步“利改税”以及 1994 年税制改革,可以利用有关企业所得税、工业品间接税数据信息来估计资本权益回报以外的社会回报。在考察资本回报率指标体系定义和我国相关统计系统演变的基础上,本节报告资本存量和资本回报数据,讨论和说明资本回报数据整理估测过程涉及的主要问题。

(一) 资本存量和资本回报数据报告

图 1 报告了中国工业企业资本存量三种基本指标数据。其中所有者权益和资产数据从 1993 年开始,固定资产净值覆盖 1978 年以来整个改革开放时期。权益与固定资产净值两个指标在 1993—2006 年共 14 年中同时有数据,数据形态显示二者绝对水平比较接近,变动轨迹关联度很高。如“权益/固定资产净值”比值最低点在 1995 年为 0.851,最高点在 2005 年为 1.106,14 年平均值约为 0.95。权益和固定资产净值两组数据相关系数约为 0.99。

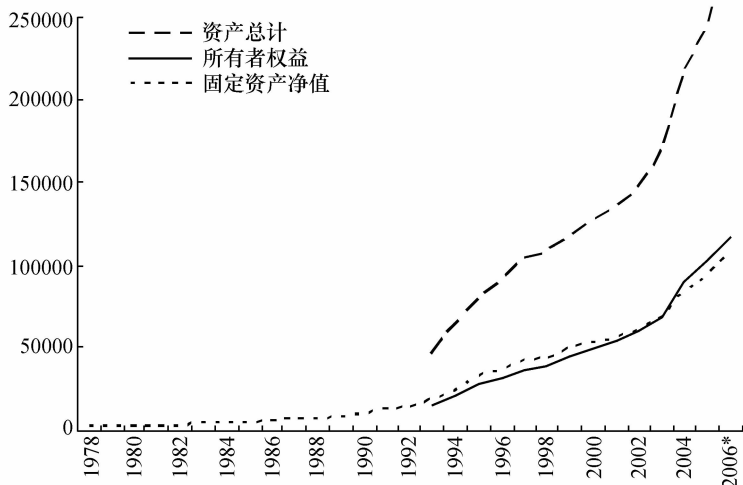


图 1 中国工业企业资本存量若干指标 (亿元, 1978—2006 年)

数据来源和说明: 固定资产净值 1978—1999 年数据以及权益和资产 1993—1999 年数据来自《中国工业交通能源 50 年统计资料汇编》(第 52、94、102 页)。2000—2005 年权益和资产数据来自《中国统计年鉴 2006》。2003 以前固定资产净值原始数据是年末数, 1998 年以后统计部门还报告“固定资产净值年平均余额”, 然而 2004 年以后不再报告固定资产净值年末值。2004—2005 年固定资产净值年末值采用下面方法估算: 1998—2003 年间“固定资产净值”与“固定资产净值年平均余额”比例值为 1.019, 考虑到该比例自 2001 年以来上升趋势和 2004—2005 年投资较快增长, 假设 2004—2006 年上述比例值为 1.04, 加上 2004—2005 年“固定资产净值年平均余额”推算这两年“固定资产净值”。1998—2005 年固定资产净值年平均余额数据来自《中国统计年鉴 2006》。2006 年资产、权益、固定资产净值根据《中国经济景气月报 (2006)》报告该年 10 月底数据折算, 折算公式是“2006 年资本存量 = 2005 年底资本存量 + (2006 年 10 月底资本存量 - 2005 年底资本存量) × 12/10。

与资本存量相比，资本回报数据更为复杂。表2报告了中国改革开放时期工业企业利税及相关指标数据。我国现行企业财务统计制度下，资本回报基本指标有两项：一是利润总额，即没有扣除企业所得税的税前利润，利润总额减去所得税得到净利润。二是利税总额，其中税金部分除了利润总额内含所得税外，还包括主营业务税收及附加以及增值税等间接税。

表2 中国工业企业利税、收入和成本(亿元, 1978—2006)

年份	主营业 务收入	主营业 务成本	主营业 务税金 及附加	产品销 售利润	利润 总额	实交 所得税	净利润	本年应 交增值 税	税金 总额	利税 总额	间接税 企业负 担部分	资本 总回报
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1978					599				317	916	97	696
1979					654				340	994	104	758
1980					692				367	1061	112	804
1981					682				397	1081	121	803
1982					704				435	1142	133	837
1983					772				461	1236	141	913
1984					852				530	1386	162	1014
1985	7899	6114			944				727	1664	222	1166
1986	8963				878				788	1665	240	1118
1987	10890	8839			1005				888	1893	271	1276
1988	14001	11268			1190				1099	2289	335	1525
1989	15847	13066			1000				1275	2276	389	1389
1990	16793	14220			560				1386	1946	423	983
1991	20597	17357			643				1591	2233	485	1128
1992	25866	21566			972				1828	2800	557	1530
1993	38084	29949	2321	4329	1602	511	1092		2321	3924	708	2310
1994	42399	33681	1067	6396	1797	573	1224		3139	4935	957	2754
1995	52936	42861	1057	7519	1635	521	1114	2331	3415	5050	1042	2677
1996	57970	47066	1180	7608	1490	525	965	2477	3657	5147	1116	2605
1997	63451	51634	1252	8084	1703	590	1113	2785	4037	5741	1231	2935
1998	64149	52798	1237	7493	1458	599	859	2827	4064	5522	1239	2698
1999	69852	57340	1308	7993	2288	658	1630	3106	4414	6702	1346	3634
2000	84152	68654	1434	10732	4393	678	3716	3685	5119	9513	1561	5955
2001	93733	77260	1554	11406	4733	1060	3673	4018	5572	10305	1699	6433
2002	109486	90244	1762	13442	5784	921	4863	4476	6238	12022	1902	7687
2003	143172	118638	2049	17295	8337	1056	7281	5488	7537	15874	2299	10636

(续表)

年份	主营业 务收入	主营业 务成本	主营业 务税金 及附加	产品销 售利润	利润 总额	实交 所得税	净利润	本年应 交增值 税	税金 总额	利税 总额	间接税 企业负 担部分	资本 总回报
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
2004	198909	167246	2616	22343	11929	1478	10451	6913	9529	21458	2906	14836
2005	248544	209863	2997	27449	14803	1806	12996	8521	11518	26321	3513	18316
2006	296418	252027	3579	31393	17636	2005	15632	9456	13035	30671	3976	21612

资料来源和说明：(1) 主营业务收入 1985—1999 年数据来自《中国工业交通能源 50 年统计资料汇编》2000—2005 年数据来自《中国统计年鉴 2006》，较早年份该指标名称为“产品销售收入”，依据同时有两种指标年份数据观察，产品销售收入与后来主营业务收入是同一指标。(2) 主营业务成本 1985—1991 年数据来自《中国工业经济统计年鉴 1992》，1992—1997 年数据来自历年《中国统计年鉴》，1998—2005 年数据来自《中国统计年鉴 2006》。该指标早先年份的名称为“产品销售成本”。(3) 主营业务税金及附加 1993—1997 年数据来自历年《中国统计年鉴》，1998—2005 年数据来自《中国统计年鉴 2006》。(4) 产品销售利润 1993—2002 年数据来自历年《中国统计年鉴》，2003—2005 年数据采用以下方法估计：先计算出 1993—2002 年间“主营业务收入 - 主营业务成本 - 主营业务税金及附加”与“产品销售利润”比例的平均值为 1.3，利用该比例值和 2003—2005 年“主营业务收入 - 主营业务成本 - 主营业务税金及附加”指标值计算产品销售利润。(5) 利润总额 1978—1999 年数据来自《中国工业交通能源 50 年统计资料汇编》2000—2005 年数据来自《中国统计年鉴 2006》。(6) 实交所得税 2000—2004 年数据来自历年《中国税务年鉴》，1995 年数据是第三次全国工业普查提供的利润总额减去净利润之差。1993—1994 年数据依据当年利润总额和 1995 年实交所得税占利润总额比率估算。1996—1999 年和 2005 年数据的估计方法是：先计算出 2000—2004 年实交所得税占增值税比例平均值为 21.2%，然后利用这一比例值和历年增值税估算。(7) 1995 年净利润数据来自第三次全国工业普查结果，其他年份数据是用当年利润总额减去实交所得税的估计数。(8) 本年应交增值税 1995—1997 年数据来自历年《中国统计年鉴》，1998—2005 年数据来自《中国统计年鉴 2006》。(9) 税金总额 1978—1999 年数据来自《中国工业交通能源 50 年统计资料汇编》，其中 1978—1984 年数据为全民所有制与集体所有制企业加总数。2000—2001 年数据来自历年《中国工业经济统计年鉴》。1998—2001 年份相关数据显示，税金总额不包括所得税。2002—2005 年数据根据主营业务税金及附加与本年应交增值税加总而得。(10) 利税总额 1977—1999 年数据来自《中国工业交通能源 50 年统计资料汇编》，2002—2003 年数据来自历年《中国工业经济统计年鉴》。1977—1999 年相关数据显示，利税总额为利润总额与税金总额之和。2000—2001 年和 2004—2005 年的利税总额是利润总额和税金总额加总而得的估计数。(11) 企业负担税金总额是间接税企业负担部分与实交所得税之和，间接税企业负担部分估计方法见正文讨论说明。(12) 总回报是间接税企业负担部分与利润总额之和。

企业统计口径 1998 年以前是“乡及乡以上独立核算工业企业”，此后调整为“国有及非国有规模以上工业企业”。口径调整使企业数量从 1997 年的 468506 个减少到 1998 年的 165080 个，不过统计对象实际经济规模由此引起的变动要小得多。如 1997 年“独立核算工业企业”和 1998 年“规模以上工业企业”工业增加值分别是 19835 亿元和 19422 亿元，所有者权益分别为 36629 亿元和 39445 亿元（参见《中国工业交通能源 50 年统计资料汇编》）。因而统计口径调整对利润率误差影响应比较有限。

2006 年营收、成本、主营税金、利润总额、增值税根据当年《中国经济景气月报》10 月底数据折算（乘以 12 再除以 10），销售利润、实交所得税、净利润、税金总额、间接税、总回报根据采用上述方法推算而得。

第二节讨论说明，资本社会回报大体等于企业所得税加上间接税中资本分担部分并减去净补贴。社会回报与净利润构成表 2 报告的资本总回报。图 2 报告净利润、税前利润、总回报、利税总额等四个资本回报数据。后面观察的利润率主要利用前三项资本回报指标。

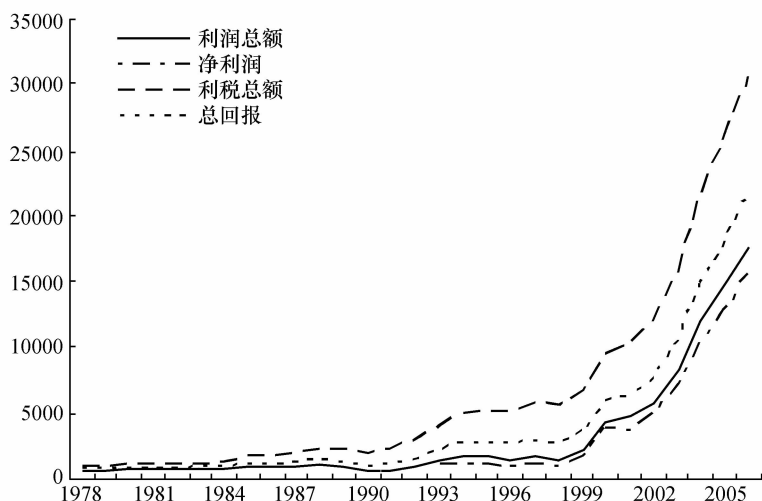


图2 中国工业企业若干资本回报数据 (亿元, 1978—2006年)

数据来源：来自表2 相关数据。

表2 还报告了产生上述基本回报指标的构成因素数据。不过由于利税指标与构成因素数据来源不尽相同，因而具体数据仅具有结构提示性，并不存在严格意义上的数量加总关系。大体而言，利润总额、税收总额、收入等指标数据比较完整，其他指标不同程度地缺少数据，有的指标仅有少数年份数据。分阶段观察，20 世纪90 年代中期以来数据比较完整，前期数据缺失率较高。

由于原始数据可获得性和统计数据口径差异等限制因素，获得表2 资本回报数据涉及一些估测处理过程。另外从有关争论情况看，对我国资本回报率走势判断的意见分歧主要发生在对我国资本回报数据是否存在系统性高估的疑问方面。为此对表2 部分数据的估计方法及有关质疑涉及的三点问题给以讨论说明。

(二) 资本社会回报的定量估测问题

依据第二节讨论，资本回报不仅包括资本所有者获得的权益回报，还包括通过所得税和资本承担间接税所贡献的社会回报，两重回报相加形成资本总回报。然而以统计部门提供的“利税总额”作为资本总回报的度量指标存在高估问题。因为企业税金所含间接税虽由企业缴纳，然而通常在不同程度上可以转嫁给消费者，如果把间接税一对一理解为资本社会回报则存在高估误差，适当方式应当是在间接税中分离出资本方承担部分作为资本社会回报，在此基础上估算资本总回报。

确定间接税中资本社会回报部分的关键是要估计间接税中资本方和消费者双方各自承担的比重。本文利用收入法 GDP 统计中资本盈余和劳动者报酬

相对比例估计企业间接税中资本剩余份额。可以从两个来源得到我国收入法 GDP 核算中营业盈余与劳动者报酬数据。第一个来源是全国投入产出表。其优势在于覆盖全国,统计指标具有更好的概念框架和逻辑关系作为基础;局限在于 1992 年以后投入产出表才有营业盈余和劳动者报酬数据,总共只能得到 5 个年份数据,这 5 年营业盈余相对比例简单平均值为 30.8%。⁶第二个来源是各省区收入法 GDP 核算中包含的营业盈余和劳动者报酬数额,将所有省区加总可以间接得到全国数据。这方面指标从 1993 年开始有数据,除 1995 和 2004 两年外,到 2005 年共有 11 个年份数据。用省区数据加总计算的营业盈余相对比例值在 1994—2003 年期间比较稳定,最高是 1993 年的 32.1%,最低是 1998 年的 26.3%,然而 2005 年上升到 41.7%,11 年简单平均值为 29.5%⁷,与上述全国数据(30.8%)很接近。本文以二者简单平均值(30.15%)作为工业部门间接税资本方承担比重估计值。将表 2 中历年不包含所得税的税金总额乘以 30.15%,得到间接税中企业负担部分估计值,并以此估算获得表 2 中资本总回报数据。

(三) 利润总额包含补贴收入问题

企业利润总额包含补贴收入,准确理解企业回报的经济含义需要考虑补贴问题。依据我国相关规定,“补贴收入是指企业按规定实际收到退还的增值税,或按销量或工作量等依据国家规定的补助定额计算并按期发放的定额补贴,以及属于国家财政扶持的领域而给与的其他形式的补贴”(CCH 亚洲公司,2005,第 174 页)。给定上述补贴定义,补贴与资本回报在定性意义上存在两重关系。

一是补贴与净利润即资本所有者回报具有正向关系。对于资本所有者而言,补贴与其他创利因素带来的资本回报具有等值性,在这个意义上企业通过补贴获得收入和利润具有真实性。二是补贴对资本社会回报产生高估影响。补贴一般来自政府财政收入,可以看作是国家代表社会从工业部门得到的负回报,因而衡量资本社会回报以及资本总回报时需要扣除净补贴部分,否则会产生高估偏差。

对我国转型期经济环境而言,更细致讨论补贴对资本社会回报影响还至少需要考虑两方面问题。一是补贴收入来源和补贴实现机制问题。如果补贴不是来源于政府财政收入,而是通过政府价格干预或其他干预措施,把部分企业税前利润转移给另一部分企业,则这类补贴代表资本回报在不同企业之

⁶ CCER“中国经济观察”研究组(2007)表 4-3 报告根据全国投入产出表计算的营业盈余占“营业盈余和劳动者报酬之和”比例值。

⁷ CCER“中国经济观察”研究组(2007)表 4-4 报告根据各省区收入法 GDP 分解数据计算的营业盈余占“营业盈余和劳动者报酬之和”比例值。

间转移，不对资本总回报带来高估影响。深入研究补贴与资本回报关系，需要考虑这类可能存在的企业间转移性补贴因素的影响。二是获得补贴利益企业对象的分布问题。我国企业分为国有、集体、私营、外资等不同类型，政府补贴可能主要发放给国有企业，私营及其他企业获得政府财政补贴可能性较小，因而如果区分企业类型加以研究，补贴对资本回报派生误差应主要表现为国有企业回报率高估。

具体定量估算历年补贴存在数据困难。我国统计部门报告了两类政府补贴数据。一是所谓“政策性补贴”，主要与政府对粮油等农产品价格和流通干预政策有关⁸，与本文讨论的工业利润所包含补贴因素关系不大。二是所谓亏损补贴，可能与工业利润存在较多关联。图3报告这两类补贴数据。

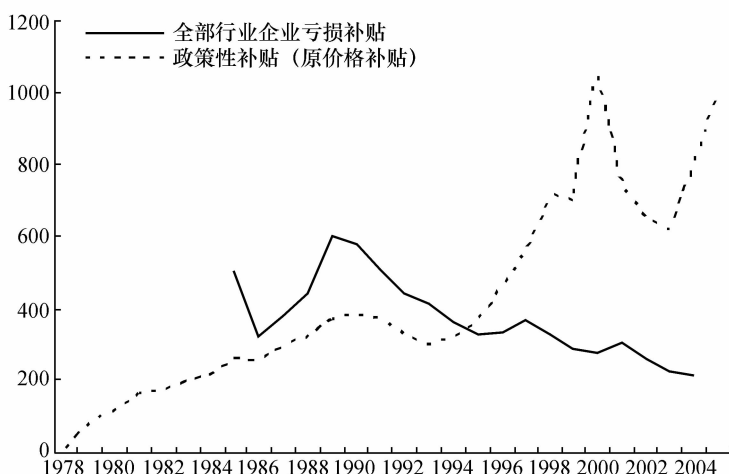


图3 我国两类财政补贴（亿元，1978—2005年）

数据来源和说明：全部行业企业亏损补贴数据全部来自《中国财政年鉴（2005）》第337页。政策性补贴1989—2005年数据来自《中国统计年鉴（2006）》，1978—1988年数据来自《中国统计年鉴（1996）》，指标名称为“价格补贴支出”。

依据上述观察，似乎有理由假设工业净补贴应是亏损补贴一部分。然而从2004年全国经济普查数据看，工业规模以上企业所得到净补贴与当年亏损补贴数不一致。2004年净补贴为509亿元，比上述亏损补贴数据高出一倍左右。受到数据限制，本文估算资本回报率时没有扣除净补贴。由于净补贴等于负税收，这一处理方法对企业当期利润率度量没有实质性扭曲，然而对资本社会回报和总回报会带来高估偏差。如2004年对社会总回报率约高估0.34个百分点，没有数据的年份无法实际计算，估计对总回报率高估应在0.3—

⁸ “政策性补贴支出”的定义是指“经国家批准、由国家财政拨给用于粮棉油等产品的价格补贴支出，主要包括粮、棉、油差价补贴，平抑物价和储备粮补贴，农业生产资料价差补贴，粮食风险基金，副食品风险基金，地方煤炭风险基金等”（《中国统计年鉴（2006）》第302页）。

0.5 个百分点。另外如上面讨论,这方面误差应主要体现为国有企业资本回报率高估。

(四) 投资收益对资本回报率影响问题

我国企业财务统计在计算利润时包括投资收益,作为营业利润和利润总额这两个利润指标之间的过渡因素之一。⁹企业利润核算考虑投资收益(亏损)是符合常识的方法,也是与国际通行企业会计规则相一致的财务统计方法。¹⁰在资本回报率讨论中,单伟建博士提出这一处理方法可能从两个方面高估资本回报:一是由于投资收益重复计算而高估了资本回报率(Shan, 2006a, p. 30; 单伟建, 2006a),二是对企业投资亏损和利润不对称的会计处理方法高估资本回报率(Shan, 2006a, p. 31)。上述质疑对准确估计我国资本回报率有启发意义,然而仔细考察发现这方面因素并不会显著高估资本回报率。

对第一点质疑可以给出直接回答:计算利润时对投资收益重复计算并不必然高估资本回报率。道理很简单,把企业间投资收益计入利润确实会重复计算利润,然而企业间投资意味着净资产所有方和利用方发生转移,因而净资产也会重复计算;由于资本回报率是资本收益与资本存量比率,分子分母同时重复计算并不必然导致比率值高估。¹¹

进一步讨论,给定企业间交叉投资这一市场环境背景,利润核算包括投资收益确实可能引入误差因素,然而误差方向可能低估也可能高估,因而是一个经验性问题,在一般讨论场合并不足以断定会发生高估利润。从常识上考量,如果不同企业存在不同方向误差,在企业数量足够大并且事件发生机制具有某种随机性情况下,绝大部分误差应能相互抵消,因而对总体利润率估计不至于发生显著影响。

第二点质疑是说投资收益为正,即存在利润时会重复计算,然而如果投资为负即发生亏损时不会重复计算,会计规则处理两类情况的不对称方法会引入资本回报高估效应。对这个比较细节的问题需要区分不同情况讨论。第一,如果投资企业对被投资企业拥有“控制”权,则投资企业和被投资企业会合并财务报表,这时亏损也会作为负投资收益计算,因而不会高估利润。第二,如果投资企业对被投资企业存在“共同控制或重大影响”,则投资企业对被投资企业的长期股权投资应采用权益法核算,那么亏损也会作为负投资收益计算,不会高估利润。第三,只有在投资企业对被投资企业没有“控制、

⁹ 另一个因素是上面讨论的补贴收入。

¹⁰ 国外会计损益表在营业利润(operation profit)与税前利润(earning before income taxes)之间存在所谓“其他收入或费用项目(other income or expense)”这一衔接性项目指标,统计内容被定义为“运营之外发生的收入和成本如红利和利息收入、利息支出、投资收益或亏损”(参见 Lyn and Ormiston, 2001, p. 98)。

¹¹ 单伟建博士似乎也同意这一分析。细心读者会发现,单文用一段话阐述利润统计包括投资收益会由于重复计算导致利润高估,后面接着用一句话提到上述净资产重复计算因素可能会抵消高估误差影响,实质上已经自问自答了这一质疑(Shan, 2006a, pp. 30—31)。

共同控制或重大影响”、投资企业对被投资企业的长期股权投资应采用成本法核算时，投资亏损才有可能由于没有重复计算，会在相应数量范围和程度上对当期利润带来高估影响。然而即便在这一特殊情况下，亏损仍以类似于“负盈余公积”的作用带来资本价值减少，因而从投资周期全程角度看，仍会通过总投资收益降低表现出来。也就是说，即便特殊情形下当期流量利润万一发生某种高估，由此发生的统计误差在较长时间窗口中仍会通过资本存量规模相应变动得以自我矫正。对于本文讨论主题而言，亏损没有重复计算假设下资本回报率的潜在高估影响，可以看作相对次要的二级甚至三级误差，在后续资本回报率估测研究中无需考虑。

四、资本回报率估测结果及其特点观察

前面定义了我国工业资本回报率指标体系，考察了我国企业会计和税收制度演变，报告并讨论了资本存量和资本回报数据。在此基础上，本节估测我国工业资本回报率，并观察其结构表现和长期走势等方面特点。

（一）资本回报率初步估测结果

利用上节的不同资本回报和资本存量数据，可直接计算九种资本回报率指标的估计结果。由于数据可获得性限制，不同指标时间长短跨度不同，有的覆盖改革开放整个时期，有的从 20 世纪 90 年代初开始可以获得连续数据。图 4 报告仅有 1993 年以来数据的七个资本回报率指标。它们走势形态很接近，都是存在先降后升的走势，不约而同地在 20 世纪 90 年代末止跌回升，七种回报率指标均上升了好几倍（2—6 倍之间）。比较 1998 年（低谷年份）

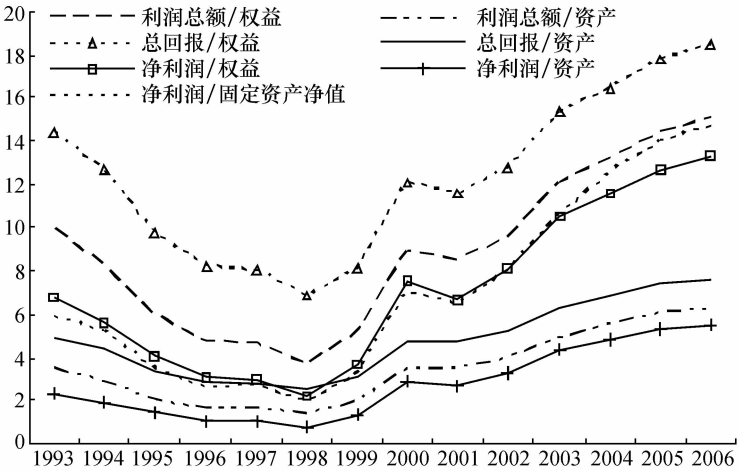


图 4 我国七种工业资本回报率估测（%，1993—2006 年）

数据来源：用图 1 和表 2 报告的资本存量和资本回报数据计算。

与 2006 年的水平,工业净资产净利润率从 2.2% 上升到 13.4%,总资产净利润率从 0.8% 上升到 5.5%;净资产税前利润率从 3.7% 上升到 15.1%,总资产税前利润率从 1.3% 上升到 6.2%;净资产总回报率从 6.8% 上升到 18.5%,总资产总回报率从 2.5% 上升到 7.6%。

图 5 报告了有 1978 年以来的数据的两个资本回报率指标:固定资产净值税前利润率和总回报率。从中可见,上述由降转升的走势,是我国资本回报率改革开放时期演变趋势的阶段性的表现。长期下降可以延伸到改革开放初期,两个回报率指标在 1978 年分别为 25% 和 29%,1993 年减少到 7.6% 和 11.6%,1998 年低谷时进一步下降到 3.3% 和 6.2%。与图 4 报告的七种回报率走势相类似,1998 年以后这两个资本回报率指标也开始强劲上升,到 2006 年分别上升到 16.5% 和 20.3%。

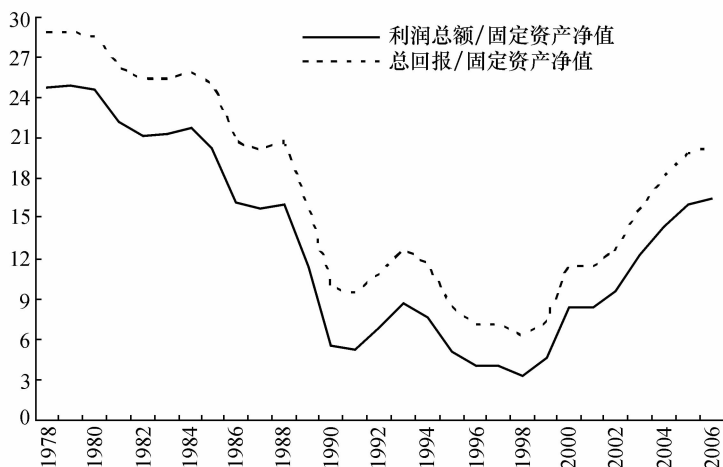


图 5 我国两种工业资本回报率估测 (%) , 1978—2006 年)

数据来源:用图 1 和表 2 报告的资本存量和资本回报数据计算。

上述资本回报率估测数据变动显示两个特点。一是不同资本回报率指标绝对水平相当接近。如“利润总额/固定资产净值”与“利润总额/权益”在 1993—2005 年期间平均值分别是 8.2% 和 8.4%, 仅相差 0.2 个百分点。2002 年及以前“利润总额/固定资产净值”总是低于或等于“利润总额/权益”, 但从 2003 年开始高于“利润总额/权益”。在 1995 年两个指标之间差距最大, 相当于“利润总额/固定资产净值”的 17.5%, 2001 年二者基本相等。二是不同指标在各有观察值时期走势具有明显相关性。表 3 显示了 9 个不同回报率指标的相关系数矩阵, 分为整个时期和 1993 年以来两种情况。所有指标序列的相关系数均在 0.95—0.99 以上。不同回报率指标绝对水平很是接近以及其变动形态高度相关, 使我们有可能克服数据可获得性限制, 间接推测某些关键资本回报指标部分缺乏数据年份的变动情况。

表 3 我国资本回报率若干指标相关性(1978—2006 年)

	净利润 /权益	净利润 /资产	净利润 /固定 资产	利润总额 /权益	利润总额 /资产	利润总额 /固定 资产	总回报 /权益	总回报 /资产	总回报 /固定 资产
净利润/权益	1.000	0.997	0.996	0.991	0.998	0.995	0.971	0.996	0.989
净利润/资产	0.997	1.000	0.998	0.979	0.996	0.991	0.950	0.989	0.981
净利润/固定资产	0.996	0.998	1.000	0.981	0.995	0.996	0.955	0.990	0.988
利润总额/权益	0.991	0.979	0.981	1.000	0.993	0.993	0.993	0.997	0.994
利润总额/资产	0.998	0.996	0.995	0.993	1.000	0.997	0.973	1.000	0.992
利润总额/固定资产	0.995	0.991	0.996	0.993	0.997	1.000	0.976	0.997	0.998
总回报/权益	0.971	0.950	0.955	0.993	0.973	0.976	1.000	0.984	0.986
总回报/资产	0.996	0.989	0.990	0.997	1.000	0.997	0.984	1.000	0.996
总回报/固定资产	0.989	0.981	0.988	0.994	0.992	0.998	0.986	0.996	1.000

数据来源和说明 数据来自图 4 和图 5。依据正文相关内容说明,“利润总额/固定资产”与“总回报/固定资产”指标数据时段为 1978—2006 年、其余指标数据时段为 1993—2006 年两种情况。

计算资本回报率涉及的另一个尚未讨论的问题是对资本存量采用期初¹²还是期末数据。从计算资本回报率的基本概念关系看,期初和期末资本存量数据选择各有利弊。上面计算的回报率采用期末资本存量数据。为比较不同指标选择对资本回报率估算结果的影响,图 6 分别用期初和期末资本存量计算资本回报率(权益税前利润率),结果显示期末期初净资产选择所得回报率绝对水平很接近,二者变动趋势也大体相同。然而也有两点显著区别。一是从绝对水平比较,利用期末资本存量计算的回报率较低,利用期初资本存量计算的回报率较高。二是个别年份有较大差别甚至方向不同。如采用期初资本存量计算的 2004 年回报率较高,但是 2005 年该指标值下降¹³;用期末资本存量计算资本回报率则增长比较稳定,没有上述起落波动现象。

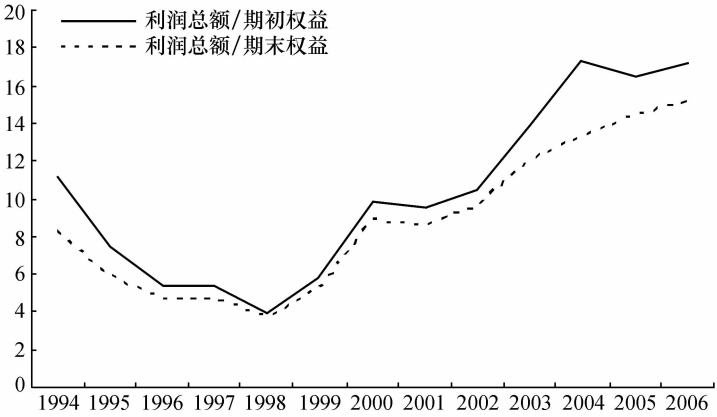


图 6 净资产期初期末选择对回报率计算结果影响(%, 1994—2006 年)

数据来源和说明：利用图 1 和表 2 报告的权益资本和利润总额数据计算。

¹² 本期期初数据实际上通常采用上期期末数据。
¹³ 如世行报告观察到“2005 年盈利状况改善的速度有所放慢”的判断可能是采用这一计算方法得到的结果(世界银行中国代表处 2006,第 17 页)。

(二) 不同类型企业资本回报率比较

上面报告了我国工业部门所有国有与规模以上非国有企业资本回报率。由于特定历史和体制转型背景,我国企业依据所有制差别划分为不同经济属性。图 7 报告了 1993 年以来国有及国有控股、三资、私营等三类企业净资产税前利润率的变动,由于数据限制,其中私营企业回报率从 1998 年开始计算。从中可见两类非国有企业资本回报率总体高于国有及国有控股企业。1993—2005 年国有企业利润率简单平均值为 6.45%,三资企业同期同一指标为 8.38%。1998—2005 年间国有企业利润率简单平均值为 7.62%,三资企业为 9.44%,私营企业为 13.86%,私营企业资本回报率比国有企业高出 81.9%。

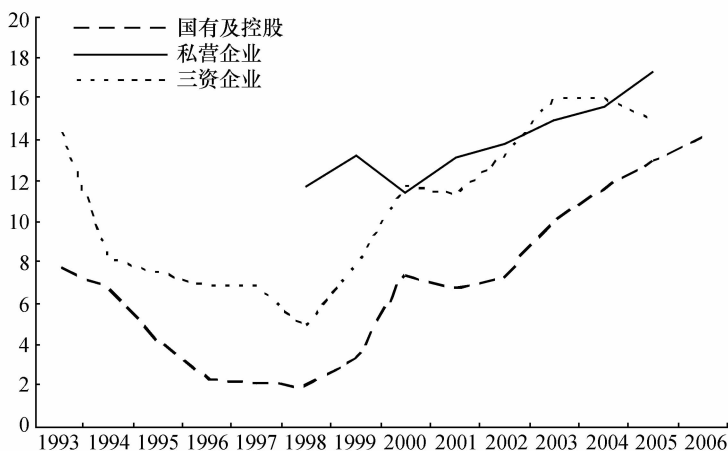


图 7 我国不同类型企业净资产税前利润率 (%)，1993—2006 年)

数据来源和说明: 各类企业 1993—1997 年净资产和利润总额数据来自 1994—1998 年《中国统计年鉴》, 1998—2005 年数据来自《中国统计年鉴 (2006)》。1993—1997 年间企业分类是: 国有经济、集体经济、股份制经济、外资、港澳台资, 图中 1993—1997 年的“国有及控股”数据是“国有经济”数据, 没有考虑股份制经济中国有控股部分。三资企业数据由外资和港澳台资加总而得。

图 8 报告了 1998—2005 年国有及国有控股、三资、私营等三类企业固定资产净值税前利润率数据。这一时期国有企业固定资产税前利润率简单平均值为 6.98%, 三资企业为 13.86%, 私营企业为 16.25%, 私营企业资本回报率比国有企业高出 1.33 倍。

图 7 和图 8 报告的数据还显示, 国有企业近期资本回报率相对增长速度超过非国有企业。如 1998—2005 年间, 私营企业净资产税前利润率增长 5.7 个百分点, 而国有企业净资产税前利润率增长 10.2 个百分点, 国有企业资本回报率增长幅度比私营企业高出近八成。国有企业资本回报率相对较快改善, 一方面说明推进国有企业改革以及国家剥离国有企业名下银行坏账等措施取

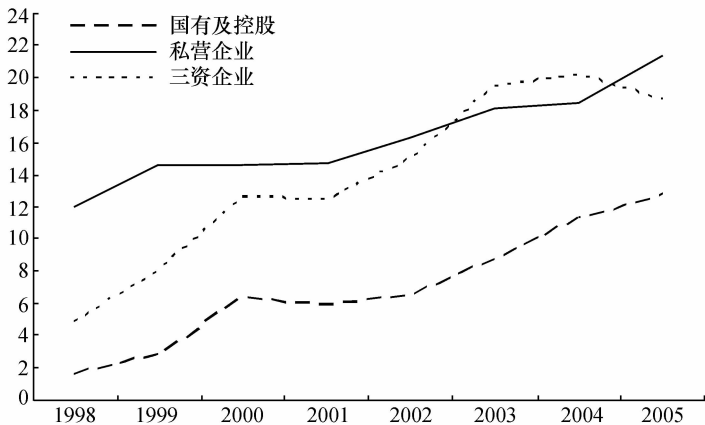


图 8 我国不同类型企业固定资产净值税前利润率（%，1998—2005 年）

数据来源和说明：利润总额数据来源与图 7 相同。固定资产净值数据是估计数，依据图 1 说明的估计方法，将固定资产净值年均余额乘以 1.04，固定资产净值年均余额来自《中国统计年鉴（2006）》。

得成效，另一方面又与现行有利于国有企业的垄断和准入限制政策有关。

从两类企业主要创利行业分布可以看出它们的盈利模式与保护政策关系的差异。图 9 和图 10 分别报告了 2005 年国有与私营企业利润来源前十大行业分布情况，显示两类企业利润来源具有不同特点。国有企业利润集中度较高，2005 年前十大利润行业占国有企业利润总额比重高达 95%，前四大行业比重为 74%；私营企业前十大利润行业占利润总额比重只有 61%，不仅远远低于国有企业同一指标，比国有企业前四大行业比重还相差 13 个百分点。

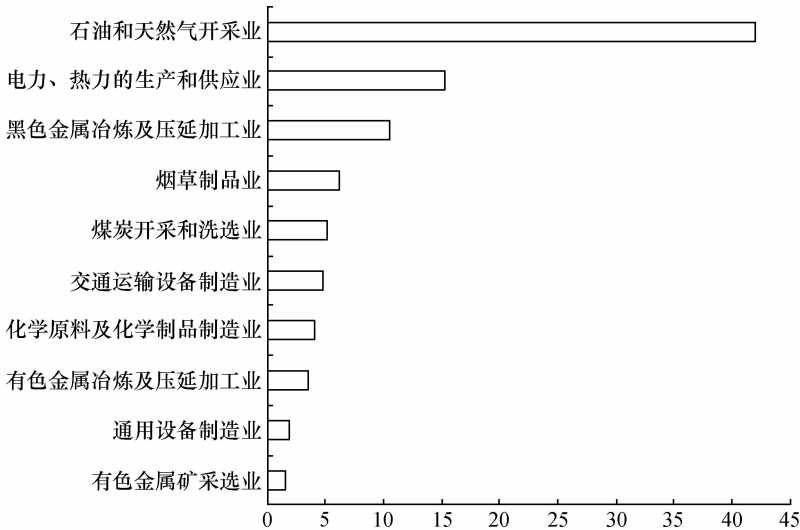


图 9 国有企业利润来源前十大行业分布（%，2005 年）

数据来源：《中国统计年鉴（2006）》。

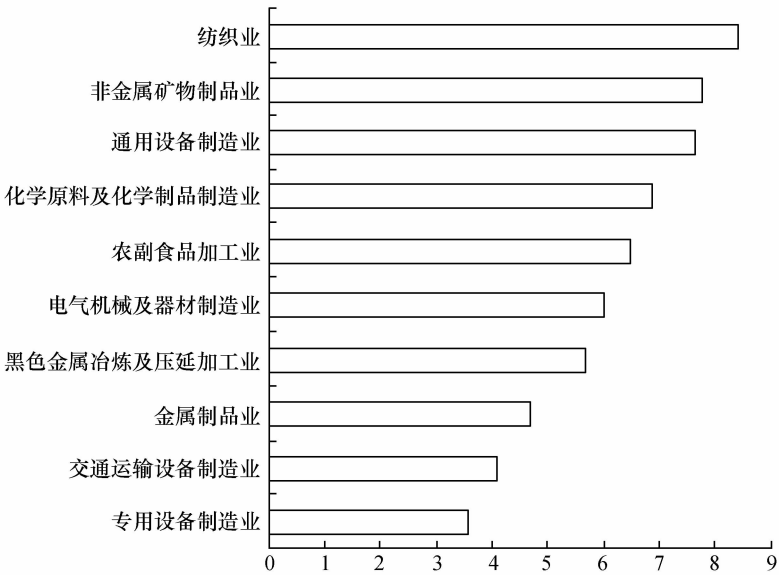


图 10 私营企业利润前十大行业分布 (% , 2005 年)

数据来源：《中国统计年鉴 (2006)》。

两类企业主要盈利行业来源和分布更是大相径庭。两类企业各自前六个最重要利润创造行业没有任何交叉。对国有企业盈利贡献最大的四个部门为“石油和天然气开采业”、“电力、热力的生产和供应业”、“黑色金属冶炼和延压加工业”以及“烟草制品业”。其中石油和天然气开采业和烟草制品业属于国家垄断行业。得益于近年石油价格大幅上涨等原因，2005 年石油和天然气开采业单个行业创造了所有国有工业企业利润总额的 42%。另外电力、热力的生产和供应业以及黑色金属冶炼及压延加工业，也是管制程度较高的行业或者近年宏观调控措施有助于国有企业在其中扩张的行业。与此相对照，私营企业利润来源最重要的六大行业包括“纺织业”、“非金属矿物制品业”、“通用设备制造业”、“化学原料及化学制品制造业”、“农副食品加工业”和“电气机械及器材制造业”，都属于竞争程度较高的行业。由此可见，我国国有企业近年盈利能力增长较快，在相当程度上仍依赖于行政管制政策的扶持和保护。

(三) 我国资本回报率长期走势特点

初步估测结果显示我国改革开放时期工业资本回报率演变呈现出超越经济景气周期的长期走势。经济运行的一般规律表现为资本回报率波动与宏观经济景气波动具有正向联系，通常是经济增长较快的景气时期利润率较高，反之亦然。后面观察显示，我国改革开放时期资本回报率表现同样具有这一属性。然而我国资本回报率变动的显著特征，在于经历了先下降后上升的长

期走势，似乎正在演绎一个“V”型历史轨迹。

从图 11 可以直观看出一次性趋势变动。固定资产总回报率指标从 20 世纪 70 年代末到 90 年代中期近 20 年在波动中呈现下降趋势，从 28% 左右下降到最低点 6.1%，下降幅度约为 80%。这一下降趋势在 1998 年开始发生逆转，从 1998 年的 6.2% 稳步上升到 2006 年的 20.3%，大约每年上升约 1.75 个百分点。税前利润率也有类似特点。图 11 还报告了利用 HP 滤波方法处理这两个回报率指标所提取的走势更为清晰鲜明的平滑趋势值。考虑前面提到的不同指标之间统计联系和量值接近关系，有理由推测其他指标估计值也应当具有类似长程走势。

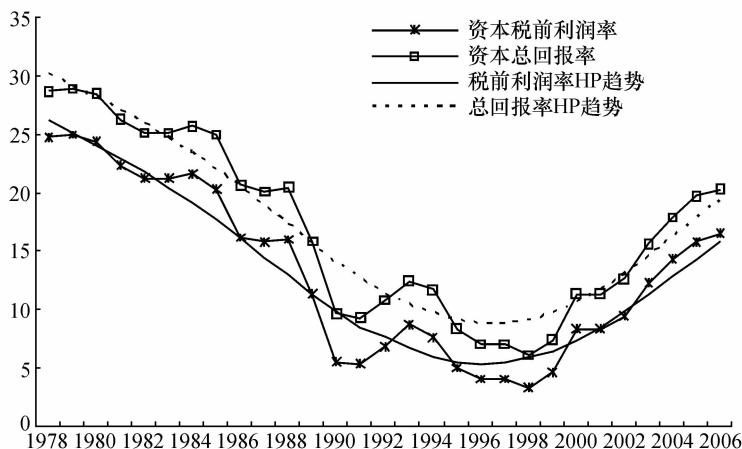


图 11 我国资本回报率长期演变趋势（%，1978—2006 年）

数据来源和说明：资本回报率是固定资产净值税前利润率和总回报率。数据来自图 5。HP 平滑系数取 100。

为了在国际比较意义上观察上述走势的特征性，图 12 同时报告了中国、美国、日本资本利润率 HP 平滑趋势值。虽然美国、日本资本回报率也经历了不同幅度升降，然而我国资本回报率在 1978—1997 年间下降幅度以及 1998 年以来回升强劲程度都来得更大。如美国从 1978 年的 24% 下降到 1989 年的 17%，日本从 1970 年的 22% 下降到 20 世纪 90 年代末的 9% 左右，但是中国从 1978 年的 24.7% 下降到 1998 年的 3.3%。上升情况是，中国到 2006 年上升到 16.5%，日本从 1955 年的 17% 上升到 1970 年的 22%，美国从 1989 年的 17% 上升到 1997 年的 19%。

我国工业资本回报在制度转型和经济起飞时期表现出先降后升的长期走势，在事实观察层面是一个极为特殊的现象，在分析层面可能是对理解中国经济成长内在规律具有重要解释意义的现象。我们在后续研究中将对其发生背景和根源进行专题考察。

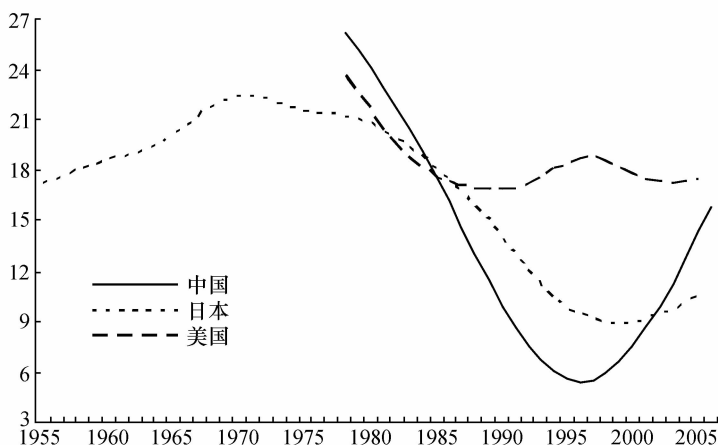


图 12 中美日税前利润率长期走势比较 (%)，1955—2006 年)

数据来源和说明：中国的趋势指标是工业企业权益税前利润率 HP 平滑趋势，美国和日本回报率趋势指标是制造业资本权益税前利润率 HP 平滑趋势值。中国 1993—2006 年净资产税前利润率由利润总额/所有者权益计算，中国 1978—1992 年净资产税前收益率由各年固定资产税前收益率与 1993—2005 年所有者权益/固定资产净值比值平均值估算，中国利润总额数据来自表 2，固定资产净值和所有者权益数据来自图 1。美国制造业税前利润率 (Annual rate of profit before tax on stockholders' equity) 截止年份为 2005 年，数据来自 U. S. Census Bureau—Quarterly Financial Report, 1996 Q1 & 2006 Q1；原始数据为季度数据，用各季度税前利润率平均值计算得到年利润率。日本制造业税前利润率由 Current profits/Equity 计算，Current profits, Equity 来自 Ministry of Finance—Financial Statements Statistics of Corporations by Industries, Quarterly；原数据为季度数据，年度利润率采用“该年度四个季度 Current profits 总和/第四季度 Equity”公式计算而得，截止年份为 2006 年。

(四) 资本回报率与经济景气变动关系

一般认为企业回报率变动是与经济景气波动相联系的变量。¹⁴为进一步观察我国资本回报率与宏观周期变量的统计联系，以税前利润率与实际 GDP 对其 HP 平滑趋势偏离作为各自短期波动度量，图 13 报告了固定资产税前利润率与 GDP 波动关系的散点数据，简单回归分析的拟合值 R^2 为 0.335，提示它们之间存在显著正向联系。固定资产总回报率与实际 GDP 短期波动也有类似关系。

为与外国比较，采用同样方法观察同期美国和日本制造业资本回报率波动与各自实际 GDP 波动的关系。图 14 报告了 1978—2005 年美国制造业税前利润率与实际 GDP 对其 HP 趋势偏离关系，与 GDP 关系的散点简单线性回归拟合 R^2 值为 0.323，与我国同一组变量回归拟合值很接近。

¹⁴ 如 Lucas (1977) 把企业利润与宏观经济景气走势一致性变动作为经济周期七大表现形态之一。

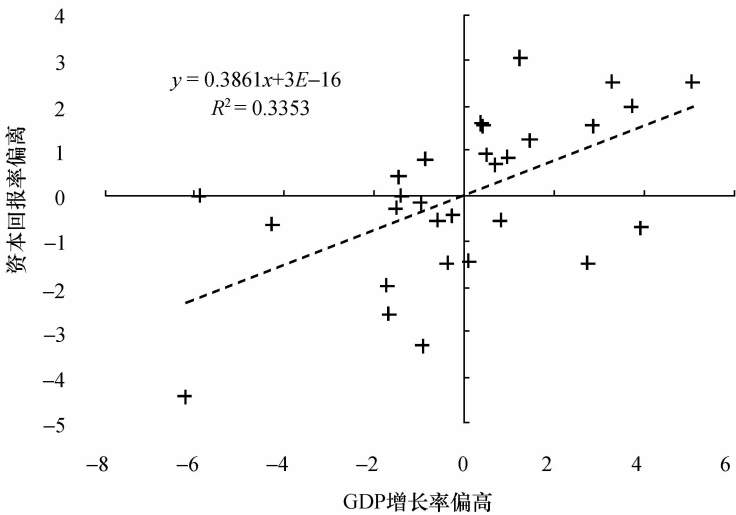


图 13 我国工业税前利润率波动与 GDP 增长率波动关系（%，1978—2006 年）

数据来源：税前利润率是税前利润对固定资产净值比率，数据来自图 5。1978—2005 年 GDP 增长率由《中国统计年鉴（2006）》中不变价格 GDP 指数计算，2006 年 GDP 数据来自国家统计局新闻，中华人民共和国国家统计局局长谢伏瞻：2006 年我国国民经济继续保持平稳较快发展（2007 年 1 月 25 日）。偏离指实际值对其 HP 平滑趋势的差距值。

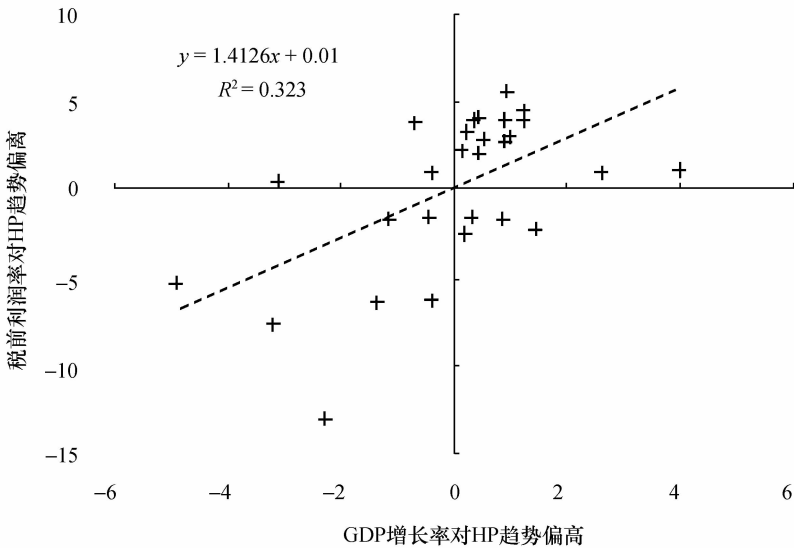


图 14 美国制造业税前利润率与 GDP 增长率波动关系（%，1978—2005 年）

数据来源和说明：美国制造业税前利润率数据来源见图 12 说明。美国 GDP 是 2000 年不变价 GDP，数据来自 Bureau of Economic Analysis—NIPA—Table 1. 1. 6. Real Gross Domestic Product ,Chained Dollars.

图 15 报告 1978—2005 年日本制造业税前利润率与 GDP 对其 HP 趋势偏离关系,与 GDP 关系的散点简单线性回归拟合 R^2 值为 0.53,说明日本资本回报率与经济景气变动联系更为紧密。

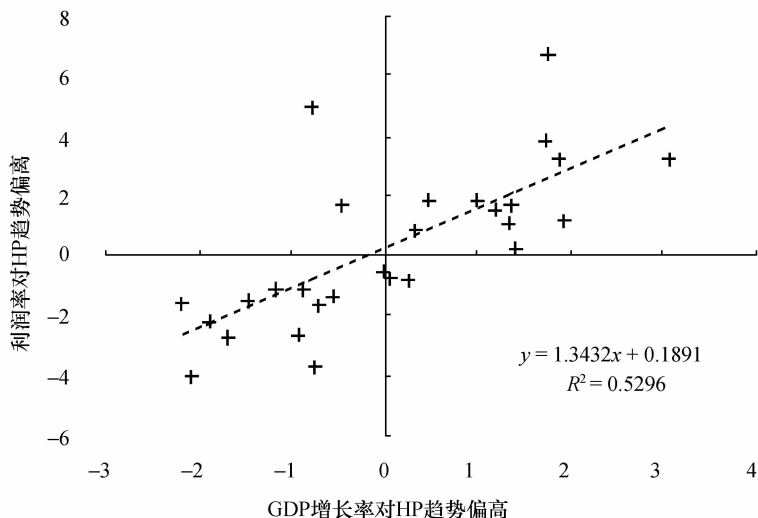


图 15 日本制造业税前利润率与 GDP 增长率波动关系 (% , 1978—2005 年)

数据来源和说明:日本制造业税前利润率数据来源见图 12 说明。日本 GDP 是 2000 年不变价 GDP,来自 World Bank—WDI—GDP (constant LCU)。

五、物价因素调整与经济利润率估测

在一个物价变动的环境中,会计上核算固定资产的方法会对资本回报率测算引入相关变量度量方法不一致的问题。因而从经济学分析角度看,会计利润率受到物价变动影响会高估或低估真实资本回报率。以固定资产作为资本存量指标,资本回报率等于利润和固定资产比率,利润等于收入与成本差额。会计核算收入和成本大部分项目(如销售收入、工资、新购买原料、管理费用等)都以当期价格衡量,如果所有变量都是当期名义值,所得利润率应是经济分析意义上的真实回报率。问题在于固定资产和某些成本项目并不以当期现价衡量,如固定资产通常以历史购置价格记账而并非以现价记账,成本项目中折旧支出受固定资产记账方法影响也不是用现价核算,因而资产价格变动会影响利润率估算结果。¹⁵

物价变动与利润率关系一般是:通货膨胀时期会计利润率高估真实利润

¹⁵ 另外计算利润时对存货的处理方式,也会由于不一定能够进行现价调整对资本回报率带来误差。不过目前会计制度改革通过公允价值方法正在调整这方面的影响。由于无法获得工业企业存货数据,所以本文将不考虑这一因素影响。

率，持续通货紧缩则可能低估真实利润率。虽然利用微观资本回报率通常并不专门考虑物价因素影响¹⁶，鉴于在本文考察期我国物价曾发生较大变动，某些年份发生过比较严重的通货膨胀，为避免高估我国真实资本回报率，有必要考察上一节会计资本回报率估测结果在多大程度上是物价变动等名义性因素影响的结果。CCER“中国经济观察”研究组（2007）分析了物价因素影响资本回报率作用机制，说明经济利润率与财务回报率的主要变量关系如下：

$$r'_t = \frac{\pi'_t}{K'_t} = \frac{\pi_t - \delta_t((1 + p_t)K'_{t-1} - K_{t-1})}{K_t + (1 - \delta_t)((1 + p_t)K'_{t-1} - K_{t-1})} \quad (2)$$

公式（2）中 r'_t 是经济利润率； K_t 和 π_t 分别表示会计报表记录的资本和利润； δ_t 和 p_t 分别表示折旧率和资本价格增长率； π'_t 表示经济利润； K'_t 表示经济资本，即用当期价格度量的资本。其中经济资本（ K'_t ）可由公式（3）估算：

$$K'_t = K'_0 \prod_{j=1}^t ((1 - \delta_j)(1 + p_j)) + \sum_{i=1}^t ((K_i - (1 - \delta_i)K_{i-1}) \prod_{j=i+1}^t ((1 - \delta_j)(1 + p_j))) \quad (3)$$

公式（3）中初始期经济资本（ K'_0 ）可由公式（4）估算：

$$K'_0 = \frac{g + \delta}{1 + g - (1 - \delta)(1 + p)} K_0 \quad (4)$$

其中 g ， p ， δ 分别表示会计资本增长率、固定资产价格增长率、折旧率在初始年份之前平均水平。

在建立物价因素调整模型基础上，CCER“中国经济观察”研究组（2007）利用各期会计资本（ K_t ）和会计利润（ π_t ）已有数据，整理和估测固定资产价格增长率（ p_t ）、折旧率（ δ_t ）、初始期经济资本（ K'_0 ）等相关指标数据，获得了我国改革开放时期工业资本真实回报率的估计结果。图 16 报告了利用 5%，10% 和 15% 三种资本折旧率估算的 1978—2005 年工业企业经济利润率，并与会计利润率进行了比较；图 17 报告用 10% 折旧率估算的 1978—2005 年会计利润率偏离经济利润率的绝对值和相对值。

调整结果显示通货膨胀对本文考察时期我国会计利润率产生相当重要影

¹⁶ 20 世纪 70—80 年代通货膨胀水平较高时国外学术界也有研究对财务核算利润率进行通货膨胀调整问题的文献（Bulow and Showen, 1983）。另外发达国家宏观经济学国民收入账户体系报告的营业盈余（operating surplus）和固定资本存量（capital stock）数据通常经过通货膨胀调整，如这类资料中“capital consumption adjustments”表示经过调整的折旧和资本存量指标，“inventory adjustments”表示经过调整的存货支出。

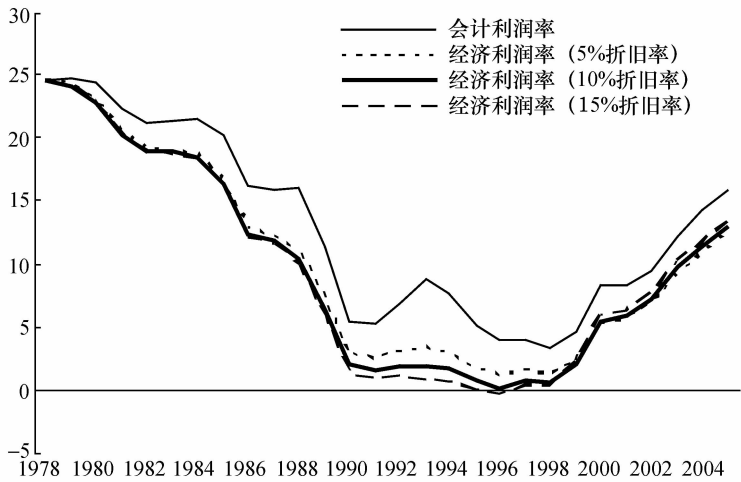


图 16 我国工业经济利润率与会计利润率比较 (%, 1978—2005 年)

数据来源和说明：会计利润率 = 利润总额 / 固定资产净值，数据来自图 5。经济利润率估算方法和数据来源见 CCER “中国经济观察” 研究组 (2007) 第 6—2 节讨论内容。

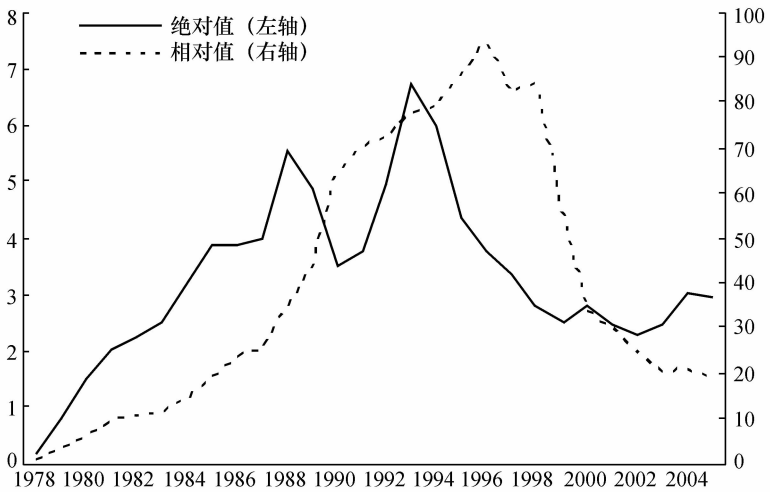


图 17 我国会计利润率偏离经济利润率绝对值与相对值 (%, 1978—2005 年)

数据来源：偏离绝对值是会计利润率减去经济利润率的差额，偏离相对值是绝对偏离值除以会计利润率所得比率值。会计利润率和经济利润率数据来自图 16，经济利润率是选用 10% 折旧率得到的估算结果。

响。物价涨幅最高的 20 世纪 90 年代中期某些年份，用财务会计数据计算的利润率相对经济利润率高估八至九成，也就是说账面会计利润率绝大部分来自物价上涨因素影响。随着晚近时期我国宏观经济运行出现通货紧缩和物价温和上涨局面，物价对资本回报率影响的绝对水平大幅下降，2001—2005 年物价影响绝对值下降到平均约为 2.6 个百分点；物价对会计利润率的相对影

响程度也持续下降，2005年会计利润率仅比经济利润率高出18.5%，与20世纪90年代中期情况已不可同日而语。

真实经济回报率估测结果显示，上节对我国资本回报率变动的基本观察仍然成立，其中某些判断甚至更为强化。一是资本回报率仍然呈现先降后升的长期走势，不过该指标变动在1990—1998年间经历一个“锅底形”低谷期，1998年前后在趋势变动中的拐点含义有所减弱。二是1998年以来我国资本回报率强劲增长判断不变，采用经济回报率指标后这一时期增长绝对值和比例值甚至有所提升。三是2005年真实税前利润率为13%，扣除企业实交所得税应仍在10%上下的较高水平。这个真实资本回报率大大高于目前以利率衡量的机会成本¹⁷，也会远远高于万亿美元外汇占款用本币计算的目前收益率和预期收益率，这一观察结果对思考我国近年投资增长的经济根源具有启示意义。

最后采用图13的方法观察经过物价调整后真实利润率与宏观景气变动的关系。图18报告的数据显示真实资本回报率仍对实际GDP波动具有显著正向关系，不过统计回归拟合程度有所降低。我们知道，1998年以来我国宏观经济运行经历了通货紧缩和景气高涨等不同周期阶段的变动，然而会计回报率和经济回报率都持续强劲增长，说明本文观察的晚近时期我国资本回报率变动，主要不是与宏观景气周期波动相联系的产物，而是代表了某种趋势性变动。如假定目前GDP增长高于潜在增长趋势1个百分点，利用真实资本回报率与实际GDP短期波动的回归系数，可粗略推测景气因素仅对真实资本回报率带来约0.27个百分点的提升作用，可见目前资本回报率观察值与趋势值差别很小。

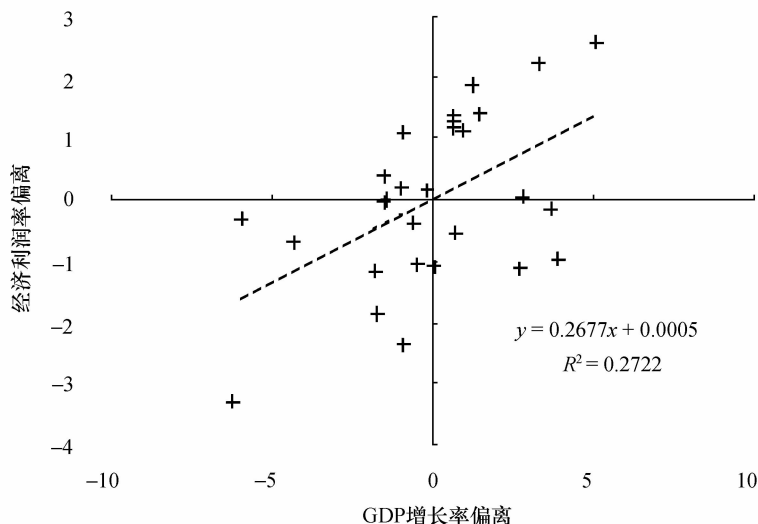


图18 我国工业经济利润率波动与GDP增长率波动关系（1978—2005年）

¹⁷ 数据来源和说明：估计方法与图13相同。GDP数据来源见图13说明。真实资本回报率数据来自图16。我国2005年中期存贷款名义利率为2.25%和3.58%。用1.8%消费物价增长率作为对存款利率的通货膨胀指数折算利率为0.44%。用4.9%工业品出厂价增长率作为通货膨胀指数，实际贷款利率为0.65%。

六、中外资本回报率初步比较

最近资本回报率争论涉及中外资本回报率比较问题。¹⁸受数据可获得性限制,本文仅利用日本和美国制造业会计资本回报率数据做初步观察比较。资本回报率国际比较的前提条件是回报率指标定义大体可比。上面对会计利润率与物价变动关系的研究提示,如果不同国家一般物价演变路径存在显著差异,会计利润率与真实回报率比较结果可能出现不一致问题。另外两国政府处于正储蓄还是负储蓄状态、政府财政支出结构都对两国资本回报率比较结果发生影响。¹⁹考虑不同国家会计资本回报率严格说来存在某些不可比因素,对下面的比较观察结果需要谨慎解读。

图 19 报告了中、美、日三国资本税前利润率数据,其中日本数据从 1955 年开始,大体覆盖日本战后经济快速追赶直至与最发达国家人均收入大体收敛整个时期。中国和美国数据从 1978 年开始。20 世纪 50 年代至 70 年代前后日本税前利润率波动很大,然而平均水平较高,约为 20%;进入 80 年代后持续下降,到 90 年代末低于 10%;近年出现回升走势,2005 年约为 12%。如前文观察,我国工业税前利润率在 80 年代前后持续下降,与日本同期利润率下降走势比较接近,但降幅更大。不过我国资本回报率近年强劲回升,2005 年已比日本同类指标高出 2.1 个百分点。

美国制造业税前利润率经历大幅波动,如 2001 年前后由于经济衰退和“9·11”事件的双重打击,利润率从 20% 左右下降到 4.5%。然而得益于比较灵活的企业制度和劳动力市场,以及外包对供应链和产业结构的不断重组作用,美国制造业利润率总体上保持了较高水平。我国资本回报率在改革初期与美国似乎差别不大,然而我国那时较高的账面利润率与计划经济时期的价格扭曲和工业垄断体制因素的遗留影响有关,并不意味着当时两国企业一般效率水平已不分轩輊。²⁰我国资本回报率到 20 世纪 90 年代中后期处于低谷时,与同期处于高位运行的美国回报率相比差距达 7—8 倍之多。然而晚近时

¹⁸ 如单伟建博士强调资本回报率“问题的关键在于,中国的资本是否得到了有效的配置与使用,从而能够实现与国际水平相当、甚至更高的资本回报率。换言之,以国际标准衡量,中国的资本是否得到了有效利用”(单伟建 2006c)。我国企业真实资本回报率“比 2005 年平均贷款利率水平仅仅高出 3 到 4 个百分点而已。这个投资回报率对利息率‘风险溢价’水平远远低于同年度中国香港和美国上市企业分别超过 6% 到 9% 的平均数,汗颜尚且不及,有什么值得大惊小怪的呢?”(单伟建 2006b)。

¹⁹ 两国政府部门处于正储蓄或负储蓄状态及其相对规模大小等因素,作为未来税收变动约束条件之一可能会间接影响两国企业会计利润率走势。资金流量表数据显示我国政府近年一直是正储蓄方,2000—2003 年政府储蓄占 GDP 比重从 5.7% 上升到 8.1%。美国政府在 20 世纪 90 年代后半期保持正储蓄,但是 2002—2005 年间为负储蓄,年均负储蓄占 GDP 比重约为 1%。另一方面我国政府支出相对美国有较大投入投入到基础设施以及其他有利于提升资本回报率的建设项目,这些因素对解读两国资本回报率也具有相应含义。

²⁰ 后续研究专题讨论资本回报率长期走势的具体原因时将进一步考察这一问题。

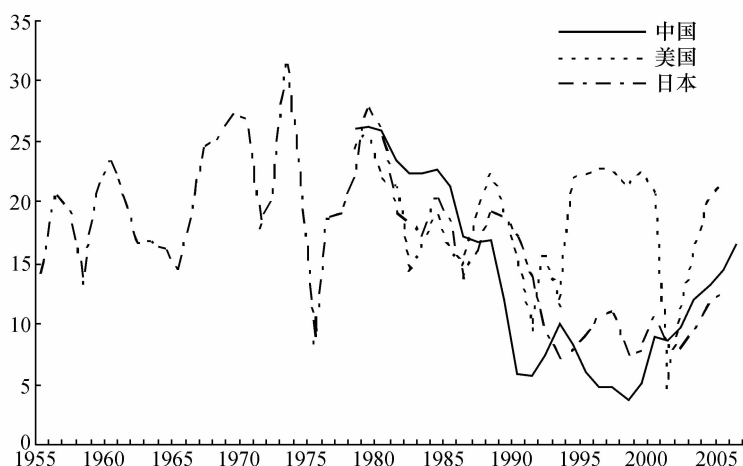


图 19 中、美、日三国税前利润率比较(%, 1955—2006 年)

数据来源和说明：利润率指所有者权益税前利润率。中国为国有及非国有规模以上工业企业利润率，1993—2005 年是权益资本税前利润率数据，1978—1992 年权益数据利用各年固定资产净值税前利润率与 1993—2005 年“所有者权益/固定资产净值”比率的平均值估算，原始数据来自图 1 和表 2。美国和日本的都是制造业净资产税前利润率数据，数据来源见图 12 说明。

期我国资本回报率强劲增长，与美国这一指标差距大幅缩小。2005 年我国工业资本回报率平均水平为 14.4%，比美国制造业回报率 21.4 低 7 个百分点，约三成。

图 20 和图 21 报告了日本和美国制造业税前利润率与我国工业不同类型企业同类指标的比较情况。图 20 的数据显示近年我国国有企业利润率与日本很接近，如 2001—2005 年日本制造业回报率简单平均值为 9.8%，我国国有企业为 9.7%。然而我国私营企业和三资企业利润率已显著胜过日本制造业企业回报率，2001—2005 年私营企业和三资企业利润率简单平均值分别为 14.9% 和 14.22%，比日本企业同一指标高出三四成。图 21 的数据显示我国私营企业追赶美国制造业资本回报率的成绩更为明显。在同时有数据的 1998—2005 年间，美国年均回报率为 17.2%，我国私营企业为 13.9%，二者相差 19.2%；2005 年美国与我国私营企业资本回报率分别为 21.4% 和 17.3%，差距也约为 19.2%；而我国国有企业与美国资本回报率差距较大，在同时有数据的 1993—2005 年间，国有企业回报率比美国低 64.8%。

图 22 报告了中国工业企业净利润率与美国制造业净利润率。美国数据从 1978 年开始，由于数据可获得性限制，中国数据从 1993 年开始。两组回报率走势比较形态大体与图 19 中的税前利润率类似，20 世纪 90 年代前中期中国工业净回报率持续下降到低谷，美国资本净回报率则高位运行，1998 年二者分别为 2.2% 和 15.8%，差距高达 6 倍左右。随着我国净利润率晚近时期快速增长，两国差距大幅缩小。到 2005 年中美两国净利润率分别为 12.6% 和

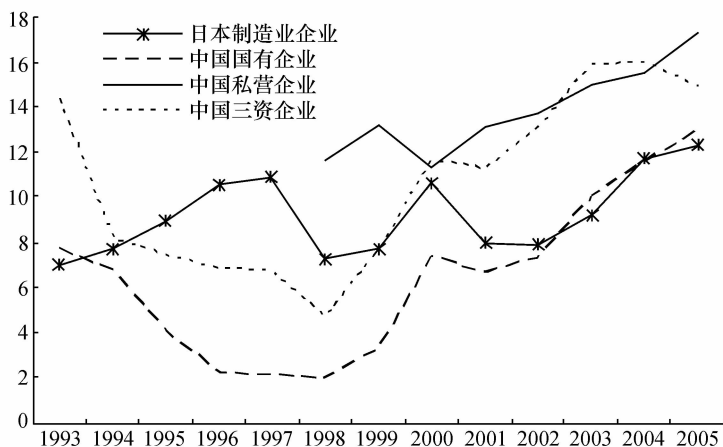


图 20 日本企业与中国不同类型企业税前利润率比较 (%)，1993—2005 年)

数据来源和说明：中国利润率指标为工业净资产税前利润率，数据来自图 7。日本为制造业净资产税前利润，数据来源与图 12 同一指标相同。

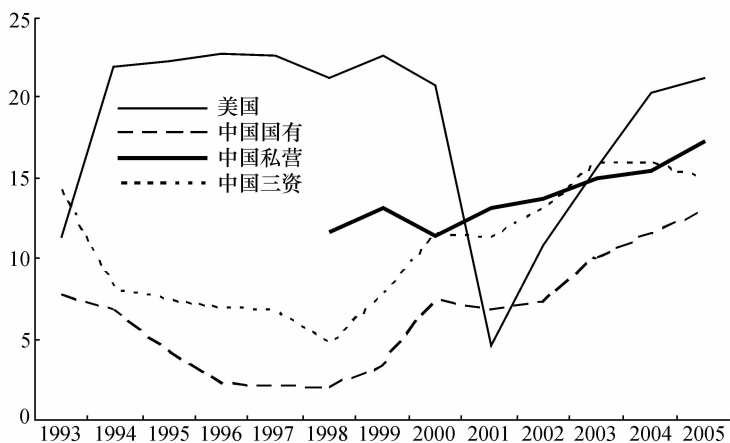


图 21 美国企业与中国不同类型企业税前利润率比较 (%)，1993—2005 年)

数据来源和说明：中国利润率指标为工业净资产税前利润率，数据来自图 7。美国为制造业净资产税前利润，数据来源与图 12 同一指标相同。

15.8%，我国比美国低 20% 左右。我国私营企业净利润率与美国同一指标差距更小。2005 年我国私营企业税前净利润率为 14.5%，比美国同年制造业税前净利润率 15.8% 仅低 1.3 个百分点，相差幅度不到一成。如果以同时有数据的 2000—2004 年五年计算，由于美国回报率在 2001 年和 2002 年大跌，我国私营企业净利润率简单平均值甚至显著高于美国水平。

上述经验证据表明，20 世纪 90 年代中期我国工业资本回报率处于低谷时，与美国等发达国家同类指标比较确实差距很大。然而我国资本回报率晚近时期持续强劲增长，目前已超过日本，并快速追赶和逼近美国制造业回报

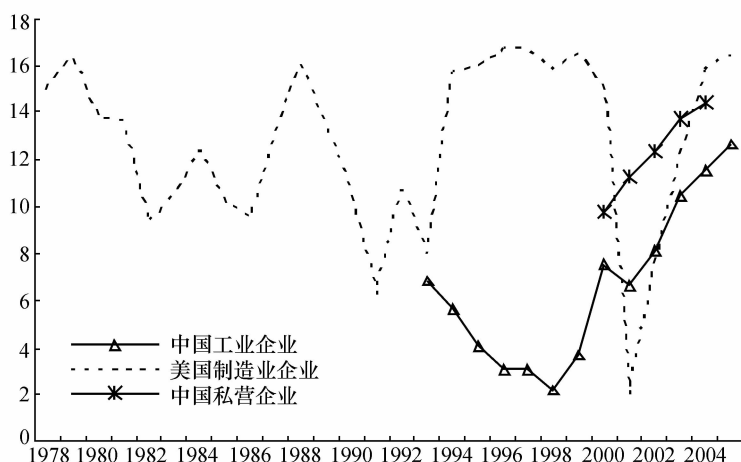


图22 美国企业与中国不同类型企业净利润率比较（%，1978—2005年）

数据来源和说明：中国指标为工业净资产净利润率。中国工业企业整体1993—2005年净资产净利润率数据见图4。2000—2004年中国私营企业净资产净利润率数据利用净利润和权益数据计算，其中净利润是利润总额减去实交所得税估算值，利润总额和权益数据来自《中国统计年鉴（2006）》，实交所得税来自历年《中国税务年鉴》。美国指标为制造业税后利润率（Annual rate of profit after tax on stockholders' equity）。美国数据来自 U. S. Census Bureau—Quarterly Financial Report, 1996 Q1 & 2006 Q1；原始数据为季度数据，用各季度税后利润率平均值计算得到年利润率。

率的水平。中外资本回报率阶段性的比较差异特点，应是中国经济转型追赶进程内在规律在相对绩效方面的具体表现。面对这一比较动态图景，用历史经验推论最新进展或有刻舟求剑之虞。有关资本回报率的讨论中，有评论认为我国资本回报率与外国比较仍如“一个增长的侏儒不能与巨人同日而语”²¹。从上述估测结果看，如果一定把外国资本回报率比作巨人，也许可以把目前我国资本回报率快速追赶解读为一个东方巨人成长的故事。

上一节对我国会计利润率进行了物价变动因素调整以获得真实经济利润率估计值。由于没有对美国和日本会计资本回报率进行类似调整，无法直接就经济利润率进行具体比较。真实利润率与会计利润率差异受资本折旧率、固定资产价格变动轨迹、投资增速等因素影响，这里仅依据三国以往相关物价指数变动情况，对中外会计资本回报率比较结果可能需要修正的问题做初步讨论推测。图23报告了中、美、日三国过去半个多世纪的固定资产和工业品价格指数。其中日本物价增长幅度最小。我国固定资产价格指数整个时期增长幅度低于美国，然而1978—2005年涨幅高于美国，过去十年前后又低于美国，呈现出两头低和中间高的比较格局。

上面对我国会计利润率进行物价变动调整的结果显示，近年我国经济利

²¹ “A growing dwarf is hardly a giant”(Shan, 2006b).

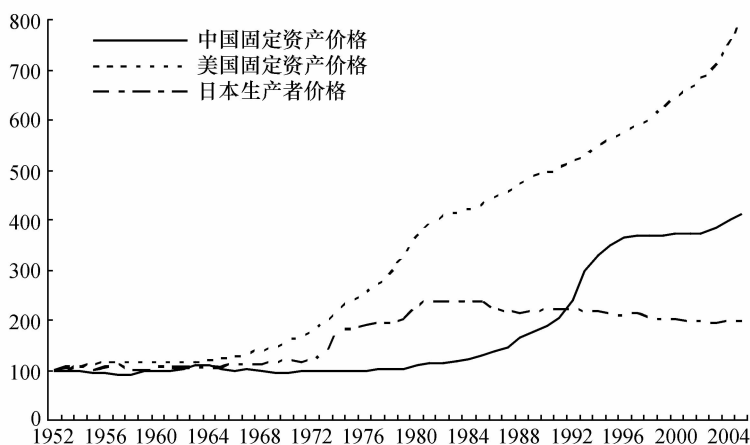


图 23 中国、美国与日本物价比较 (1952 = 100, 1952—2005 年)

数据来源和说明：中国 1953—1990 年固定资产投资价格指数采用固定资本形成平减指数，固定资本平减指数由固定资本形成总额与固定资本形成不变价指数计算，固定资本形成总额与不变价指数数据来自中国国家统计局国民经济核算司（1997，第 50 - 51 页）；1991—2005 年为固定资产投资价格指数，数据来自历年《中国统计年鉴》。美国资本价格定义是“Deflator of Fixed Assets and Consumer Durable Goods Net Stock”，数据是 2000 = 100 指数，数据来源为 Bureau of Economic Analysis。日本 1952—1970 年价格是批发价格指数（1970 = 100），来自《帕尔格雷夫世界历史统计（亚洲、非洲和大洋洲卷）》（第 3 版）第 970 页；1970—2005 年是生产者价格，来自 IMF-IFS 数据库“PPI/WPI (2000 = 100)”。

利润率比会计利润率大约低 2.6 个百分点。日本物价上涨幅度小得多，过去二十多年中一些年份处于通货紧缩状态，日本经济利润率可能与会计利润率不相上下，因而中国与日本会计利润率比较的结果需要结合两国物价变动情况做相应修正。然而即便假定日本会计和经济利润率大体相等，目前我国私营企业资本经济回报率估计仍已显著超过日本同类指标。考虑我国与美国固定资产价格相对增长两头低和中间高的情况，两国经济利润率比较情况应与上述会计利润率比较判断差别不大。综上所述，考虑物价调整因素后，我国真实资本回报率超过日本和逼近美国的结论应仍能成立。

七、本文的发现及其政策含义：结论性评论

我国新一轮经济景气增长表现出高增长低通胀的可喜组合特点，然而较快的投资增长又使人们对目前增长机制的合理性疑虑重重。常识告诉我们，利用行政手段动员资源高投资缺乏效率，最终会走向危机甚至崩溃，因而关注投资增长的经济合理性和可持续性条件是必要而正确的。不过也需要指出，高投资并不必然等于低效率。如果在市场经济环境下投资形成的资本能创造较大剩余，较高的投资则有可能增加微观主体和社会整体的跨期财富总量并

具有经济合理性。因而对投资率或投资增长的合理性问题，不仅要着眼其具体数量水平进行评估，还要结合资本回报率等微观指标加以辨识。

近来有关我国资本回报率的研究和讨论，对理解我国目前经济形势具有认识和政策意义。鉴于问题复杂性，也考虑目前发表研究文献大都用评论短文等形式发表，不便细致考察资本回报率涉及的诸多复杂因素，我们通过一组论文比较系统地研究这一问题，本文侧重估测改革开放时期我国资本回报率九个序列指标并观察其走势特点。为此本文梳理有关概念指标关系，厘清相关统计制度历史演变背景，细致整理我国工业企业资本回报、资本存量等指标时间序列数据，测算观察资本回报率演变轨迹和目前水平，并对有关争论所提出问题进行分析。

需要说明的是，我国体制转型伴随着经济统计指标体系的逐步调整改进，因而在资本存量和利润指标时间序列数据可获得性上存在困难。即便存在相关数据，解读其具体含义时也会因为统计口径和方法演变而发生疑义。因而本文对资本回报率进行定量研究时，不可避免地需要利用经济学和统计技术常识，对某些问题采取整理、猜测甚至假设的方法加以处理。有关这类估测处理的具体依据和假设，本文依据学术规范要求一一仔细说明。我们刻意选择把研究过程的“魔鬼细节”透明化，一方面是为便于读者更准确解读本文所得结论的含义及其局限性，另一方面也是为后续研究对本文观点提出批评或改进预留“接口”。

本文提供六点主要发现，具体如下：

第一，晚近时期我国工业资本回报率发生了真实而非虚构的强劲增长。经过仔细分析的经验证据显示，1998—2006年间9个资本回报率序列指标以很高统计相关性同时增长。1998年到2006年，以权益作为资本存量计算，净资产净利润率从2.2%上升到13.4%，税前利润率从3.7%上升到15.1%，总回报率从6.8%上升到18.5%；以资产作为资本存量计算，总资产净利润率从0.8%上升到5.5%，税前利润率从1.3%上升到6.2%，总回报率从2.5%上升到7.6%。计算总资产回报率所用的利润数据已扣除利息支出，2005年不扣除利息的总资产净利润率约为6%—7%。如以利率作为投资机会成本，2005年总资产净利润率比一年期贷款利率4.9%高出23%—43%，比当年企业和居民存款利率2.25%更是高出170%—210%。如考虑利息税因素，总资产净利润率比存款利率高出幅度更大。

第二，不同类型企业资本回报率存在显著差异。比较不同类型企业资本回报率的绝对水平，私营企业最高，三资企业其次，国有或国有控股企业（下面简称国有企业）较低。以净资产税前利润率为例，2005年私营企业为17.3%，三资企业为14.9%，国有企业为12.9%。不过比较1998年以来资本回报率的增长幅度，三类企业排序则正好相反：国有企业增幅最高，为10.9个百分点；三资企业其次，为10.2个百分点；私营企业较低，为5.7个百分点。

点。晚近时期国有企业资本回报率较快增长,说明国有企业改革政策发生成效。然而比较国有和私营企业主要盈利行业分布可以看出,国有企业盈利增长在相当程度上仍依赖于行政管制政策的扶持和保护。持续观察将来不同企业资本回报率的相对变动是一个重要研究问题。

第三,体制转型背景下资本回报率呈现先降后升走势。我国资本回报率晚近时期增长,承接此前二十年左右的趋势性下降,使该指标变动在整个改革开放时期大体展现出一个先降后升的长程走势。我国体制转型和经济发展过程内生出降低和提升资本回报率两种力量,上述走势应是这两种力量在不同发展阶段消长盈缩的产物。后续专题研究论文利用资本回报率贡献因素分解构架考察,发现前期价格自由化改革带动原料价格相对上涨、竞争环境下垄断利润耗散、企业放权让利、改革时期单位劳动成本上升等因素导致前期回报率大幅下降。晚近时期原料相对价格仍在上涨,然而相反方向作用的力量开始生效,如劳动生产率快速增长基础上单位劳动成本下降、管理费用和利息费用相对下降、资本运营效率水平提升等等,这些因素推动资本回报率在 1998 年前后止跌回升并强劲增长。

第四,调整物价因素后对资本回报率增长的判断仍然成立。对企业会计利润率进行物价因素调整的结果显示,通货膨胀对我国会计利润率产生了重要影响。通货膨胀最高的 1993 年前后,账面会计利润率绝大部分来自物价上涨因素影响。晚近时期物价对会计利润率影响绝对水平大幅下降,近年大体在 3 个百分点上下,2005 年账面利润率由此高估的幅度仅约为 18%。调整通货膨胀因素后,真实资本回报率在 1990—1998 年间经历“锅底形”低谷期,然而贯穿整个时期先降后升的大势形态依然存在。1998 年以来,真实回报率绝对和相对增加幅度更高,2005 年用固定资产净值作为资本存量计算的真实税前回报率约为 12.9%,扣除实际所得税负后真实回报率约为 10%,2006 年真实净资产净回报率也在 10% 上下。真实资本回报率大大高于目前以利率衡量的机会成本,也应远远高于万亿美元外汇占款用本币计算的目前收益率和预期收益率。

第五,资本回报率与经济景气波动存在显著联系。用税前利润率对其 HP 偏离作为资本回报率短期变动度量指标,用实际 GDP 对其 HP 趋势偏离作为经济景气短期变动度量指标,发现会计资本回报率以及经过物价变动因素调整的真实资本回报率与实际 GDP 波动都存在显著正向联系。晚近时期我国宏观经济运行经历了通货紧缩和景气高涨等不同周期阶段变动,同期会计和经济资本回报率持续强劲增长,说明我国资本回报率变动主要代表某种趋势性变动,而不仅仅是宏观景气周期变动的结果。如假定目前 GDP 增长高于潜在增长趋势 1 个百分点,利用真实资本回报率与实际 GDP 短期波动回归系数,可粗略推测景气因素仅对真实资本回报率带来约 0.27 个百分点的提升作用,可见目前资本回报率观察值与趋势值差别很小。

第六，我国工业资本回报率胜过日本，逼近美国。不同国家财务资本回报率严格来说存在某些不可比因素。由于所得数据限制，本文仅对中、美、日工业和制造业的资本回报率进行了初步比较。我国资本回报率经历了先降后升走势，20世纪90年代中后期该指标处于低谷时，远不及日本经济快速追赶阶段的制造业回报率水平，与美国同期制造业资本回报率相比差距更大。晚近时期我国资本回报率快速增长，目前会计利润率已超过日本，经过物价调整的经济回报率的比较情况可能有所不同，推测至少我国私营企业该指标已超过日本。我国工业会计利润率与美国制造业同一指标相比差距快速缩小，私营企业与美国回报率水平已相当接近。考虑中美有关物价动态比较情况，两国经济回报率与会计回报率比较情况应大体类似，因而真实资本回报率逼近美国的判断仍能成立。

本文的研究结果对理解目前经济景气增长和讨论宏观调控政策具有参考意义。首先本文对目前投资较快增长和新一轮经济景气补充提供了一个观察评估视角。流行观点倾向于认为高投资增长一定是低效率增长，倾向于相信投资增速超过某个数量水平则必然过高过热，倾向于假设近年投资增长主要动因是地方政府干预和要素市场扭曲，等等。这类看法都有一定道理，对实际经济表现也有一定解释作用。然而依据经济学的基本分析逻辑，市场经济环境下投资增长受到资本回报率等微观绩效变量制约或推动。全面理解我国投资增长和新一轮经济景气的发生机制，还需要对我国经济基本面因素进行更深入的实证分析。

如果我国真实资本回报率确实发生趋势性持续增长，那么从经济学投资理论视角看，应有理由假设与均衡分析意义上的合意资本存量比较，我国目前资本存量相对不足，因而企业通过较快投资，缩小现实资本存量与合意资本存量之间差距，可能具有较大程度的经济合理性。这样看来，目前我国投资快速增长以及高增长低通胀的宏观经济表现，可能具有相当程度的微观基础，因而实现中央最近提出的“又好又快”的经济发展方针确实具有现实可能性。从一个更广阔的视角思考，一国经济起飞的早期阶段人均资本存量很低，快速投资及其导入伴随的“嵌入式技术（embedded technology）”进步以及人力资本积累提升，不仅有可能构成持续增长的重要驱动因素，而且可能是经济追赶特定阶段的题中应有之义。因而把“投资驱动增长”无条件等同于“粗放低效增长”的流行看法是否准确，可能还要在理论依据和经验证据方面进一步探讨。

本文的研究结果对我国宏观经济政策也有借鉴意义。如果认定目前投资增长与宏观景气较大程度上是地方官员干预等扭曲性因素派生的结果，那么便会合乎逻辑地推论微观决策主体对利率、汇率、价格等市场性参数的变动较大程度上无动于衷，便会进一步合乎逻辑地推论目前投资较快增长在很大程度上难免需要采用产业政策、数量控制等侧重行政性手段的调节干预，而

市场性调节手段就只能退而发挥某种次要和辅助作用。反过来看,如果近年投资较快增长具有相当程度的经济基本面因素支持,如果目前经济景气在相当程度上是市场机制力量推动的结果,而地方官员干预和要素市场不完善等扭曲性因素仅发挥了相对次要和次生作用,那么宏观经济管理措施就有可能,并且应当更加重视采用与市场经济原则兼容一致的操作手段。

笔者愿意再次肯定政府在市场经济条件下进行总需求管理的必要性,重申即便对具有经济基本面支持的景气增长,政府也需要进行宏观调节。实质问题不在于一般意义上是否有必要进行宏观管理,而在于如何反思评估我国独特的宏观调控?在于从我国体制和政策较长时期演变进程看,是应更多利用市场性“调节手段”进行总需求管理,还是更多采用行政性“控制措施”对经济运行进行微观干预?是应深化改革使“宏观调控”逐步转变为“宏观调节”,还是诉诸形形色色的数量控制和行政干预,使“宏观调控”更接近“控制经济”。近年贯彻科学发展观取得巨大成就,然而如何防范大范围行政干预和寻租活动在宏观调控名义下乘势回潮,如何把推进关键领域改革议程与促进社会公平和谐措施有机结合起来,创造我国经济追赶新阶段所迫切需要的体制和政策环境,仍是下一步经济发展大政方针选择所需考量的重大问题。资本回报率作为经济运行的微观效率指标,为这方面探讨提供了值得关注的经验证据。

参 考 文 献

- [1] 白重恩、谢长泰、钱颖一,“中国的资本回报率”,《比较》2007年1月第28辑,第1—22页。
- [2] Bulow, J., and J. Showen, “Inflation, Corporate profits, and the rate of return to capital”, in Hall, R. (ed.), *Inflation: Causes and Effects*. Chicago: The University of Chicago Press, 1983.
- [3] CCH 亚洲公司,《中国企业会计制度》。北京:中信出版社 2005 年。
- [4] CCER“中国经济观察”研究组,“我国资本回报率(1978—2006)——新一轮投资增长和经济景气微观基础”,北京大学中国经济研究中心中文讨论稿, No. C2007002 2007 年 2 月 9 日。
- [5] 第三次全国工业普查办公室,《中华人民共和国 1995 年第三次全国工业普查资料汇编(综合、行业卷)》。北京:中国统计出版社,1997 年。
- [6] Fraser, L., and Ormiston, A., *Understanding Financial Statements*, 6th edition, Prentice-Hall, 2003.
- [7] 高善文,“投资效率并不差”,《新财富》2004 年第 6 期,第 10—23 页。
- [8] 国家统计局工业交通统计司,“新中国的工业统计”,《统计研究》2002 年第 7 期,第 37—40 页。
- [9] 黄勇峰、任若恩、刘晓生,“中国制造业资本存量永续盘存法估计”,《经济学(季刊)》2002 年第 1 卷第 2 期,第 377 页。
- [10] 胡鲁滨、胡蛟,“中国资本回报率 15% 还是 9%?——一场高度技术性的学术之争隐含大相径庭的宏观判断与政策建议指向”,《财经》2006 年第 20 期,第 68—69 页。
- [11] Kujis, L., “Investment and Saving in China”, World Bank Policy Research Working Paper 3636, June 2005.

- [12] Kuijs, L., "How will China's saving-investment balance evolve?" World Bank China Office Research Working Paper No. 5. May 5th, 2006.
- [13] Kuijs, L., and B. Hofman, "Letters to the Editor of The Wall Street Journal Asia", The Wall Street Journal Asia, September 6th, 2006a.
- [14] Kuijs, L., and B. Hofman, "Profits drive China's boom", Far East Economic Review, 2006b, 169(8), 39—43.
- [15] Kuijs, L., and B. Hofman, "企业利润推动中国繁荣之车",《华尔街日报(中文网络版)》2006(c)年10月12日。
- [16] 梁红, "中国投资的高速增长是可持续的",《高盛全球经济研究报告系列》第146号 2006年10月3日。
- [17] Lucas, R., "Understanding business cycles", in Brunner, K., and A. Meltzer (eds.), Stabilization of the Domestic and International Economy. North Holland: Amsterdam, 1977.
- [18] 齐舒畅, "我国投入产出表的编制和应用情况简介", 见国家统计局网站 http://www.stats.gov.cn/tjdt/gmjhs/t20030527_80228.htm.
- [19] 全秋梅, "专访 TPG 新桥投资集团合伙人单伟建:与世行论战'中国资本回报率'始末",《第一财经日报》2006年9月14日 A8 版。
- [20] Roach, S., "The Great Chinese Profits Debate", October 6th, 2006, <http://www.morganstanley.com/views/gef/archive/2006/20061006-Fri.html>.
- [21] 单伟建 a, "世界银行高估了中国企业的盈利水平",《华尔街日报(中文网络版)》2006年9月5日。
- [22] 单伟建 b, "中国企业的利润率之谜",《远东经济评论(中文网络版)》2006年9月6日。
- [23] 单伟建 c, "中国的低利润增长模式",《华尔街日报(中文网络版)》2006年11月24日。
- [24] Shan, W., "The World Bank's China Delusions", Far Eastern Economic Review, 2006a, 169(9), 29—32 (a revised version of the article is also published at The Wall Street Journal Asia, September 1, 2006).
- [25] Shan, W., "The Mystery of Chinese Profitability", The Wall Street Journal Asia, 2006b, 169(7).
- [26] Shan, W., "China's low-profit growth model", Far East Economic Review, 2006c, 169(9), 23—28.
- [27] 世界银行中国代表处,《国有企业分红:分多少?分给谁?》,未公开发表研究报告 2005年10月17日。
- [28] 世界银行中国代表处,《中国经济季报》2006年5月。
- [29] 宋国青, "中国投资率太低",《CCER 中国经济观察》2006年夏季(总第6期),第49—58页 2006年7月29日。
- [30] 肖耿, "中国资本回报率之谜",《南方周末》,"外部观察栏目" 2006年9月21日。
- [31] 许小年、肖倩, "投资过快增长,但通货膨胀还未到来——没有温度的高烧",《财经》2003年第17期,第44—45页。
- [32] 张军、吴桂英、张吉鹏, "中国省际物质资本存量估算:1952—2000",《经济研究》2004年第10期,第35页。
- [33] 张西天、石少梅、杜小燕, "中国会计核算制度模式的演进", <http://www.e521.com/library/11/1102/1113104639.htm> 2000年11月13日。
- [34] 赵晓, "中国的投资回报率高出奇 经济并未过热",《新青年·权衡》2006年10月13日。
- [35] 中国国家统计局国民经济核算司,《中国国内生产总值核算历史资料 1952—1995》。大连:东北财经大学出版社,1997年。

- [36] 中华人民共和国财政部,《工业企业财务制度》。北京:煤炭工业出版社,1993年。
- [37] 中华人民共和国财政部税制税则司,《中国税收制度》。北京:企业管理出版社,1998年。
- [38] 朱向东,“如何进一步完善‘净资产收益率’的计算方法”,《建材财会》,2004年第4期,第58—59页。
- [39] 邹至庄,《中国转型经济》。北京:中国人民大学出版社,2005年。

Measurements of China's Capital Return (1978—2006) : Microeconomic Underpinnings for the Recent Economic Boom in China

CCER China Economic Observer Research Group
(Peking University)

Abstract This paper measures 9 indicators of capital return for the Chinese industrial firms during the period of 1978—2006. It examines the issues such as the trend of profitability, differences of profitability between various categories of firms, impacts of inflation on the profitability calculated using the corporate accounting data, comparison of profitability among China, Japan and the US. Evidence found by the paper indicates the micro-economic underpinnings for the Chinese economic boom in recent years.

JEL Classification L25, O47, O12

择优分配最优化原理及其应用

茅于軾*

摘要 本文讨论在资源约束条件下如何分配使产出的效果极大。从最简单的两块土地分配化肥使增产的粮食极大化开始,在边际上不断调整,得出一般约束条件下的最优化方法,即拉格朗日乘子法。此优化原理更可以推广到动态过程,借用物理学中的捷降线问题,把落差看成资源,在逐步分配落差中使旅程的时间最短,进而把动态问题一般化,推导出变分法中的欧拉方程。择优分配原理可以将拉氏乘数法和欧拉方程都解释为资源的优化配置问题,从而给出了它们的经济学意义。本方法为一基本的优化原理,有广泛的应用可能。

关键词 资源配置,静态和动态优化原理,拉格朗日乘子法和欧拉方程的经济学解释

一、择优分配原理

在经济工作中经常遇到分配问题,所谓分配问题就是将供应有限的资源分给若干部门,以得到最佳的经济效果。这里的资源,可以是能源、资金、外汇、矿石、木材,甚至可以是人力、时间、土地等等。

为了判断一个分配方案是否优于另一个,必须有评价的标准。如果评价标准还没有建立起来,讨论优劣便没有任何意义。建立评价标准是经济学上的基本问题之一,本文不准备对这个问题加以讨论,而是假定已经具备了这个标准以后,讨论如何制定分配方案的方法。

如果只有一个部门参加分配,最佳分配量是使收益获得最大的量。如果收益对于分配的函数关系为已知,则通过微分方法可以求得切线为水平处的分配量 X^* 为最佳分配量,见图1。如果资源量少于 X^* ,则显然应将全部资源分配完,因为收益随着分配而增长。

从这里所举的例子我们可以看到,“收益”尽管在不同的情况下可以有不同的计量方法,例如铁矿石的分配的收益可以是产出的生铁量,电能的分配的收益可以是由于分配电能而得到的产值,资金的分配的收益可以是利润数等等,但所有这些收益均必须是纯收益,即扣除支出以后的收益。否则,对

* 天则经济研究所。作者通讯地址:北京市三里河南沙沟2-1-1,100045;电话:(010)68522129;E-mail:maoysh@mail.ied.ac.cn。本文为作者供职于中国能源研究会和铁道部科学研究院时完成的,最初发表于1981年8月,为一油印本。原文共12节两万余字,现在为了压缩篇幅,减为八节,1.3万字。压缩掉的四节为:五、择优分配与宏观经济;六、择优分配原理的若干应用;七、可靠性的最优分配;十二、利用Kuhn-Tucker最优性条件求收益函数。现在发表的除了节、公式、图表的编号改变外,其余一字未改。

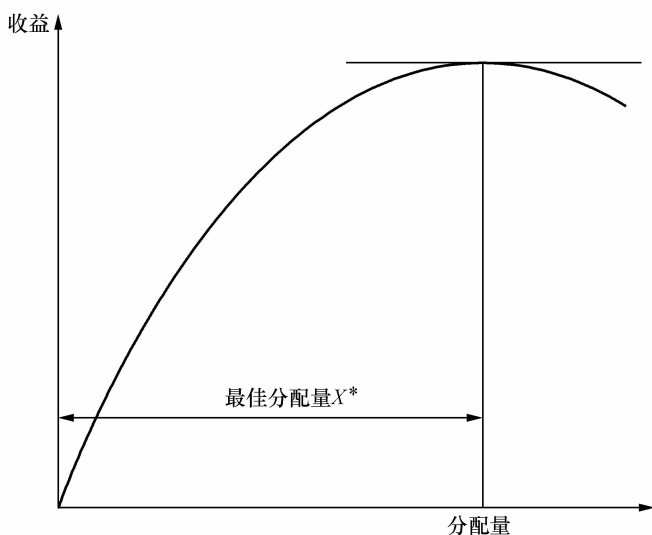


图 1 投入和产出——分配和收益

于亏损企业，分配量越多亏损越大，尽管毛收益也随着分配而增加。

当有几个部门参加分配时，显然不能用简单的微分方法求得最优解。考虑只有两个部门参加分配的情况。我们先定义收益 g 对于分配 x 的导数 $\frac{dg}{dx}$ ，称为边际收益率，或边际产出率。对于这两个部门各有其收益函数 $g_1(x_1)$ 及 $g_2(x_2)$ 。对应于任何一个分配方案，两个部门各有其边际收益率 $\frac{dg_1}{dx_1}$ 及 $\frac{dg_2}{dx_2}$ ，见图 2。如果这两个边际产出率不相等，这个分配方案必有改善的可能。设想对边际收益率较大的第一部门增加微量分配 dx ，同时对边际收益率较小的第二部门减少分配 dx ，则在总分配额 $x_1 + x_2 = X$ 不变的条件下，两部门的总收益可以获得增量 $\frac{dg_1}{dx_1}dx - \frac{dg_2}{dx_2}dx = \left(\frac{dg_1}{dx_1} - \frac{dg_2}{dx_2}\right)dx$ 。可见当 $\frac{dg_1}{dx_1} \neq \frac{dg_2}{dx_2}$ 时，分配方案可以继续改进，因此不是最优分配方案。仅当两部门具有同样的边际收益率时，分配方案才可能是最优的。

当有多于两个部门参加分配时，上面的结论同样成立。因为如果有一个最优分配方案，则任意两个部门之间均应具有同样的边际收益率。否则仅对这两个部门的分配加以调整就可以改善总收益量。任意两部门都有相同的边际收益率，只有当所有各部门的边际收益率相同时才有可能。因此，边际收益率相等是最优分配方案的必要条件。

现在考虑收益函数是凹函数时，通过逐步的微量调整以改善分配方案，最终会得到什么结果。图 1、图 2 中所绘的均系凹收益函数的情况，即对于同样的分配量 dx ，当分配数量 x 增多时，如由 $x = a$ 增加到 $x = b$ ， $a < b$ 收益的

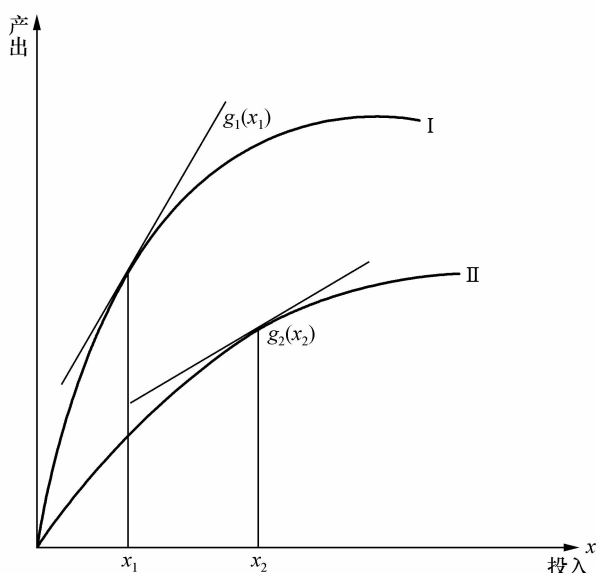


图2 边际产出率

增量 $\left. \frac{dg}{dx} \right|_{x=a} > \left. \frac{dg}{dx} \right|_{x=b}$ ，在改善分配效果时，是将 $\frac{dg}{dx}$ 较大的部门增加分配量， x 增加以后该部门的 $\frac{dg}{dx}$ 必减小；而 $\frac{dg}{dx}$ 较小的部门减少分配量后， $\frac{dg}{dx}$ 增大，这是凹函数的条件，改善分配的根据是 $\frac{dg_1}{dx_1} \neq \frac{dg_2}{dx_2}$ ，经过调整改善之后各部门的 $\frac{dg}{dx}$ 大的变小，小的变大。因此逐步调整的结果，最终将使所有部门的边际收益率均收敛于同一数值 λ 。即 $\frac{dg_i}{dx_i} = \lambda (i=1, 2, \dots, N)$ ， N 为参加分配的部门数。

如果我们将总资源量 X 分成若干小份 dx ，然后一份一份选择效果最佳的部门逐步分配下去，这就是一个择优分配的过程。此时最先得到分配的部门必是在原点附近收益函数最陡的部门。以后的分配必定是使 $\frac{dg}{dx}$ 最大的部门优先照顾，但因 dx 是微增量，因此结果一定是每一个分配方案都使各部门具有同样的边际收益率。 X 从零开始一直到全部分配完毕，历经了无数的分配方案，每个方案都有 $\frac{dg_i}{dx_i} = \lambda$ 的特点，而且它们都是最优分配方案，因此，等边际产出率同时又是最优分配方案的充分条件。因此可以得出结论：对凹的收益函数，各分配部门具有相等的边际产出率，是最优分配的充分必要条件。此时我们还假定了收益曲线是通过原点的，即分配为零时收益亦为零。

根据这个结论，我们很容易得到求各部门最优分配量的方法：任意确定一个切线的斜率即 $\frac{dg}{dx}$ ，对各部门的收益函数作平行切线，各切点所对应的横

坐标 x_i 即为第 i 部门应该分得的资源量。如果 $\sum x_i \neq X$, 可以改变 $\frac{dg}{dx}$ 的值, 当切线斜度变陡, 由于函数的凹性, 各个 x_i 均变小, 反之 λ 变小时各个 x_i 均变大。最后总可试凑使 $\sum x_i = X$ 。这个试凑过程在电子计算机上很容易实现。另一个求得最优分配方案的方法是: 由任意一个满足 $\sum x_i = X$ 的可行方案出发, 不断优化这个方案, 即减少 $\frac{dg}{dx}$ 小的部门的分配量, 改配给 $\frac{dg}{dx}$ 大的部门。最后各部门的 $\frac{dg}{dx}$ 均相等时即为最优分配方案。

如果收益函数是凸的, 逐步改善分配的结果不会收敛于等边际收益率。因为对于凸函数, $\frac{dg}{dx}$ 大的部门增加分配以后, $\frac{dg}{dx}$ 变得更大, 而小的变得更小。结果全部资源都将分配给一个部门。而且分配的结果与最初方案的选定有关, 因为全部资源最终都将分配给最初方案中 $\frac{dg}{dx}$ 最大的那个部门, 而不同的初始分配方案下, 各部门相对的 $\frac{dg}{dx}$ 的大小次序不同。然而凸函数求极小值时, 完全可以应用方才讨论的凹函数求极大值的理论。在经济问题中成本函数对于生产的各种要素 (投入的资本、劳力、应用的物料、生产进度安排等) 往往具有凸特性。使成本极小化就可应用成本对生产要素分配量的边际变化率相等的最优化条件, 来确定生产要素的最佳分配。

二、择优分配的对偶原理

在各个部门的收益函数已经完全确定的情况下, 一定数量的资源量 x 对应着一定的边际收益率 λ , 并且对应着一定的总收益 $\sum g_i(x_i^*) = G$ 。当 x 变化时, λ 和 G 也跟着变化。如果我们分别用 X 、 Λ 、 G 表示资源量、统一边际收益率和总收益量的实数集合, 则三个集合之间存在着——对应的映射关系, 如图 3。 X 到 Λ 的映射由 φ 表示, Λ 到 G 的映射由 ϕ 表示。——对应的关系由收益函数的凹性所决定。举例说, 如果资源量为 100 单位, 对几个部门分配得到彼此统一的边际收益率为 0.3, 此时总收益量为 70 个单位, 如果资源增为 101 单位, 由于函数的凹性, 边际收益率必将减少, 设减少为 0.29, 总收益量将因投入的资源总量有所增加而增加, 设增加为 71。这个例子说明了 X 中的任一个数都与 Λ 及 G 中唯一的某一个数相对应, 反之 Λ (或 G) 中的任一个数也与 X 和 G (或 X 和 Λ) 中唯一的数相对应。

这三个集合之间的对应关系, 启示了择优分配的对偶原理, 即某一个资源总量通过求出的统一边际收益率对应着一个最大的收益总量, 同时某一收

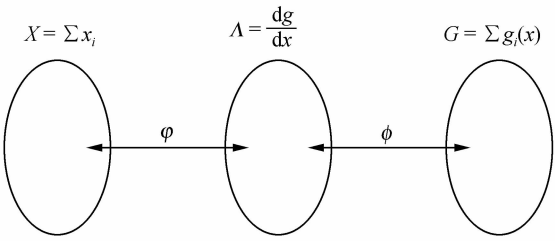


图 3 X、 Λ 、G 集合间的关系

益总量通过统一的边际收益率对应着一个最小的资源总量。同一个资源总量由于分配方案的不同，可以有不同的收益总量，但最大的收益总量只有一个，即对应应在 G 集合中的数，那是通过统一边际收益率求出的；同样，为达到一定的收益总量，可以耗用不同数量的资源，但最少的资源量只有一个，即对应应在 X 集合中的数，那也是通过统一边际收益率求出的。

为什么 X 对应着 G 集合所需的最少资源需要量呢？这可用上面所举的例子来加以证明。为了获得 71 个单位的总收益，最少的资源需要量不可能大于 101 个单位，因为从 φ 和 ϕ 的对应我们确实只用了 101 个总资源获得了 71 个总收益。同时最少的资源需要量又不可能小于 101 个单位，因为如果只用了少于 101 个资源也能获得 71 个收益，那就违背了上升段凹函数的定义，即仅当 x 增加时方始有 λ 的减小，在上升段 $\lambda \left(= \frac{dg}{dx} \right)$ 无论如何减小还是大于零的，即 g 的变化方向与 x 相同，它随 x 的增加而增加，随 x 的减小而减小，现在 x 既然从 101 减小， g 也必然从 71 减小。可见最少的资源需用量既不可能大于 101 也不可能小于 101，于是 101 就是为得到 71 个总收益所必需的资源量。因此证明了 x 集合中的每一个数对应着 G 集合中每一个数所需的最少资源量。但这个结论仅当资源分配量对应着收益函数上升段的情况是对的，如分配量超过了图 1 中的 X^* ，收益函数转入下降段，结论就不再适用。但实际经济工作中发生的问题中的收益函数都处于上升段。

三、统一边际收益率的含义

对于一定的资源量，对于一定的收益量，都有唯一的统一边际收益率与其对应。边际收益率的单位在能源分配问题中可能是每吨标准煤的工业产值；在铁矿石的分配问题中可能是每吨矿石产出的生铁量；在资金分配的问题中就是资金利润率，此时 λ 的含义是边际资本产出率，当它以年率的形式表示时，其倒数就是投资还本期。

由于 X、 Λ 、G 三个集合的对应关系，一定的边际资本产出率和一定的资本总额互相对应着。当资本充裕时对应的边际资本产出率较低，也就是投资

还本期可以放宽。在市场调节的情况下,资本供应充足将使银行利率降低。因此银行利率实际上和边际资本产出率相联系。一个企业如果从银行贷款从事经营,则贷款利率不得高于企业的资本产出率,否则不但本金永远无法偿还,连利息也付不起。

基本建设投资的最优分配,就应该用边际资本产出率或投资还本期来衡量。对应于一定的资金总额有一定的投资还本期,基建项目中投资还本期超过这个统一投资还本期的都属缓建项目。短于统一还本期的都属应建项目。按照这样一个极简单的原则分配投资,必定得出投资的最优分配方案。当然这只是从经济效果来衡量。重大投资项目的决定还要考虑政治、军事等其他因素。

上面所说的统一的投资还本期应按照资金多寡的情况每年调整。资金紧张时统一投资还本期短,即仅对获利最丰的项目投资。资金充裕时还可对获利次佳的部门投资。

正因为边际资本产出率对应着资金总额,所以一个国家从国外取得贷款的数量应与贷款的利息相关。利息率越高贷款额越少。这两者的数量应通过专门研究客观地确定,并作为国际间金融往来的重要参考资料。显然,贷款额不仅与利率有关,而且与参加分配的部门数 N 有关。 N 越大,即兴建的项目越多,在同样利率条件下贷款额可以增大,同时收益总量 G 也增大。开发生产门路、发展经济优势,就是增加 N 的具体措施。

收益总量对应着资源总量。在资金分配的情况下,总的利润额决定了应该准备的资金总额。如果资金不足此数,就不可能达到总的利润额;如果资金超过此数,利润虽能保证,但降低了资金利润率,保证不了投资效果。根据预定的目标估计需要的资源最低数量,是经济计划中的一项重要任务。

统一边际收益率 λ 又标志着资源的匮乏程度。 λ 越大说明这种资源短缺得越严重,因为 λ 大意味着投入资源可以得到的收益很多,或者意味着由于缺少资源损失的收益很多。如果用每吨煤影响多少产值来表示煤炭的短缺程度,这个数字应该是选择分配最后得到的统一边际收益率。如果煤矿附近发生着大量浪费,而在远离煤矿的工业区又严重地短缺,这两处的边际收益率相差很大,没有做到择优分配,就不存在统一的边际收益率,也就无从衡量煤的短缺程度。只有进行了择优分配,将浪费的煤改配给缺煤地区,才能最后估计出增加煤的产量是否还能增加产值,以及每吨煤能增加的产值是多少。

统一的边际收益率可用来衡量稀缺物资的经济价值。对于不受资源限制的产品,只要投入资本和劳力就可制造出来,但诸如石油,某些有色金属、热带作物等,由于受到资源及其他限制,它们的生产带有垄断性质。因此当对它们的需求量超过资源条件所能提供的数量时,它们的价格往往超过一般条件下确定的数值。在资本主义生产条件下这就是广义的级差地租发生的原因。在社会主义条件下如何正确度量这类稀缺物资的价值,并据此在技术经

济比较中作出判断，合理地分配和使用，必须借助于这类物资对各有关部门分配最终得到的统一边际收益率来计算。这个边际收益率就是该项物资的影子价格，以后我们将讨论它的数学定义和含义。

社会主义企业由于自然条件不同，在同样经营管理水平及同样努力程度下，产品成本有颇大的区别，这种苦乐不均的现象必须用征收资源税的方法来纠正。一般地说，统一边际收益率可用以确定在正确制定可比成本时所要考虑的资源税的税率。从这里我们也可以看到，在技术经济比较中对稀缺物资（如石油制品）的价值强调用生产成本来计算是不对的，应该用成本加上利润和资源税，或者直接按择优分配最终达到的统一边际收益率再加适当调整来计算。

边际收益率是两个微增量的比，用它来说明资源稀缺程度时，必须注意随着资源供应的增加，边际收益率将逐渐降低。说明能源匮乏就不能用每亿吨能源增产若干亿元，因为这样表示远远不是微增量之比。

四、经济活动中收益函数的凹性

择优分配结果得到统一边际收益率，仅当收益函数为凹函数时才有这个结论。事实上经济活动中的收益函数是否具有凹性？其原因是什么？有无例外情况？这是有待进一步讨论的。

首先，我们相信，在一般情况下收益函数具有凹性，即收益的增长比投入的增长慢。其根本的原因是收益取决于一系列的因素，其中的任一个因素受到某种限制不能跟着变化时，投资资源的效果将受它的限制，收益就逐渐减小。农作物的收成可随施肥量而增加，但当土地面积一定时，收成因受土地的限制而不可能与施肥量成正比增加。工厂的产量与能源供应量有关，但由于受工厂设备、劳力、空间等的限制，产量也不可能与能源供应量成正比地增加。空间、时间、人力、资源等的限制都是客观存在的，它们最终都会在经济活动中以某种形式表现出来。

其次，经济关系中的逐级分配现象也使收益函数出现凹性。如果 x_1 、 x_2 、 x_3 分别表示农、轻、重的投资分配，则对农业来说， x_1 又可分配于粮食、林业、科研、教育等。 $g_1(x_1)$ 所表示的农业投资的收益将是由上述各属于农业的各部门的收益所构成。于是将 x_2 再对粮食、林业、科研等农业内的各部门进行择优分配，先分配给收益最显著的部门，其次分配给收益稍次的部门。可见 x_1 增大时，由于最后分配的部门必是收益最差的部门，故 $g_1(x_1)$ 的增加在 x_1 越大时就越小，说明了 $g_1(x_1)$ 必然是凹的。不仅农业内部有逐级分配，工业、交通业亦均如此。甚至小到一个具体产品也有投资的逐级分配问题，例如花费一笔投资用于提高电视机某一零件的可靠性，这笔投资一定首先用于可靠性最敏感的某道工序。投资数很大时开始照顾到提高可靠性比较困难

的工序。这个例子说明了逐级分配现象带有普遍性。每一级的择优分配都促使上一级的收益函数呈凹性。

但是收益函数的凹性却不是没有例外的。例如由于协作的关系两个人完成的工作量大于一个人工作量的两倍。规模很大时,由于社会分工的发展,产出量的增加亦可能大于投入量的增加。由于经营管理方法的改进(例如采用了择优分配的方法等),或者由于工艺技术的改进,甚至可以在投入量不变的条件下增加收益。这些都会导致收益函数呈凸性。从人类生产力发展的总过程来看,改进生产组织和技术进步始终是最活跃的因素,但是在某一个特定时期生产方式相对地比较定型的时候,收益函数就表现出凹性。

在资本主义社会,由于企业和个人追逐利润,并由于市场调节的作用,分配过程会不断地优化改进。在资金分配的情况下,表现为资金利润率趋于统一的自发过程。这种现象说明了资金分配的收益函数确实是凹的,因为对于凸的收益函数,分配优化的结果不会收敛于统一边际收益率,这在前面已经证明过了。然而由于资本主义生产缺乏统一的计划调度,它的优化过程是在利益互相对立的条件下进行的。技术垄断和情报保密妨碍了最优条件的达到,实际的分配结果是在最优分配方案的周围振荡。由于实力巨大的垄断组织的出现,或者单纯由于不利条件的随机组合,这种振荡的振幅有时可能是很大的,因此可能造成严重的社会后果。在社会主义的生产条件下,从理论上讲我们可以不必依赖于市场和竞争的作用,只要通过计划的最优化就可实现最佳的经济效果。但事实上由于原始资料不可能十分详尽周全,数据的传递不可能毫无延误,最优化问题的数学计划也远未臻完善,我们还不得不依靠市场调节的反馈作用来补充计划的不足。计划最优化工作做得越差,我们将越多地依赖于市场调节,计划最优化做得越是详尽严密,市场调节的作用越小。最优化的计划调节在整个经济调节中所占的比例越高,说明经济活动中的随机现象越少。如果可以用熵来度量经济活动中的随机化程度,那么不断完善和加强计划调节的作用,使经济活动目标更明确,配合更协调,效果更完美,实际上就是一个不断减熵的过程。因此社会主义生产的高度发展是组织最优化作用逐渐加强的过程。相反,如果既不依靠计划的最优化,又不利用市场调节的反馈机制,经济工作必然发生严重的混乱,其后果将是灾难性的。回顾世界经济日益商品化依赖的历史,许多重大的经济危机就是由于上述两种现象同时存在而引起的,尽管发生这两个现象的原因可能是经济本身的,也可能是政治性或军事性的重大动荡。

五、择优分配原理和 Lagrange 乘数法

前面讨论的资源分配问题和可靠性分配问题等,都是在约束条件下求极值的问题。这类问题普遍地可以用 Lagrange 乘数法求得解决。但前文我们并

未依靠 Lagrange 乘数法, 同样求得了解决方法, 而且这种方法的经济意义十分明显。本节中我们将进一步证明, 用择优分配原理可以求得 Lagrange 乘数法的一般结果。

在资源分配中 Lagrange 乘数法的提法是:

求能使目标函数 $G(x_1, x_2, \dots, x_N) = \sum_{i=1}^N g_i(x_i)$ 取极大值的各个 $x_i (i = 1, 2, \dots, N)$, 同时满足约束条件 $\sum_{i=1}^N x_i = X$, 用 Lagrange 乘数法求解, 作函数 $L(x_1, x_2, \dots, x_N) = \sum_{i=1}^N g_i(x_i) + \lambda(X - \sum_{i=1}^N x_i)$, 令 L 函数取极值, 其必要条件是 $\frac{\partial L}{\partial x_i} = 0$, 即 $\frac{\partial g_i}{\partial x_i} = \lambda (i = 1, 2, \dots, N)$ 。

前面我们对择优分配作经济意义讨论的时候, 中心的思想是: 资源在各部门间互相作微量调整时 (但这种调整不能破坏规定的约束条件, 即 $\sum_{i=1}^N x_i = X$), 目标函数可以不断地改善。但在最优分配时此种调整已不能使目标函数得到改善。这是最优分配所应满足的必要条件。但资源分配量的微量调整的方向可正可负, 即既可能从甲部门给乙部门 dx , 也可以从乙部门给甲部门 dx 。两种调整方法都不能使目标函数增大, 故这两种调整的结果只能是使目标函数的改变为零。换言之, 从甲部门减少 dx 的分配量如果使目标函数改变 $-dG$, 则对乙部门增加 dx 的分配量必定使目标函数改变 $+dG$, 二者相加为零, 也就是目标函数的边际收益率大小相等且符号相反。这个条件对任意两个分配部门均应成立。当约束条件不是简单的 $\sum x_i = X$ 时, 上述推论依旧成立, 所不同的只是某一 x_i 减少 dx 时, 另一 x_k 不是增加 dx , 而是要根据约束条件的具体要求增加。

根据以上论证, 我们得到极其重要的择优分配的结果: 边际收益率相等为最优化的必要条件。具体表述如下: 一个约束条件下求最优化的问题, 可以视为对各个变量的分配问题。在达到最优分配时, 任意两个变量在遵循约束条件的情况下作相应的微量调整时, 对目标函数的边际收益率应大小相等, 符号相反。下面我们用这个最优化条件来证明 Lagrange 乘数法。

$$\text{目标函数 } G(x_1, x_2, \dots, x_N) \text{ 取极大,} \quad (1)$$

$$\text{约束条件 } R(x_1, x_2, \dots, x_N) = 0. \quad (2)$$

令任一个 x_j 改变 dx_j 另一个 x_k 改变 dx_k , 其余的 x_i 均不变, 约束函数 R 的改变量在忽略 dx 高阶项时可表示为:

$$\frac{\partial R}{\partial x_j} dx_j + \frac{\partial R}{\partial x_k} dx_k = 0, \quad (3)$$

(3) 式必须等于零, 否则约束条件被破坏。

从 (3) 式得

$$dx_j = - \frac{\frac{\partial R}{\partial x_k}}{\frac{\partial R}{\partial x_j}} dx_k, \quad (4)$$

(4) 式说明, 为了遵循约束条件, x_k 改变 dx_k 时 x_j 应该变多少。

x_j 、 x_k 作上述调整时, 它们对于目标函数的边际收益率为 $\frac{\partial G}{\partial x_j} dx_j$ 及 $\frac{\partial G}{\partial x_k} dx_k$, 二者大小相等, 符号相反, 故

$$\frac{\partial G}{\partial x_j} dx_j = - \frac{\partial G}{\partial x_k} dx_k, \quad (5)$$

将 (4) 式代入, 得

$$\frac{\frac{\partial G}{\partial x_j}}{\frac{\partial R}{\partial x_j}} = \frac{\frac{\partial G}{\partial x_k}}{\frac{\partial R}{\partial x_k}} = \lambda. \quad (6)$$

x_j 、 x_k 为任意选取的, 故有

$$\frac{\partial G}{\partial x_i} = \lambda \frac{\partial R}{\partial x_i}, \quad (i = 1, 2, \dots, N), \quad (7)$$

也可写为

$$\frac{\partial}{\partial x_i} (G - \lambda R) = 0, \quad (i = 1, 2, \dots, N). \quad (8)$$

(8) 式即为 Lagrange 条件。它共有 N 个, 加上约束条件 (2), 共可解出 N 个 x_i 及 λ 。但解得的结果为最优分配的必要条件, 仅当目标函数为凹时同时又为充分条件。或者目标函数要求为极小, 同时它又是凸函数时 (8) 式也是充分条件。

(6) 式又可写成

$$\frac{\partial G}{\partial R} = \lambda. \quad (9)$$

此式明确地表明了乘数 λ 的意义。它表明约束条件微量变化时, 目标函数的增量与约束条件增量的比, 或者目标函数的值依赖于约束条件的程度, 故称之为灵敏度系数。在经济问题中 λ 有其经济含义。如 R 表明某种资

数值 G 集合之间通过灵敏度系数 Λ 集合存在着对应关系。图 3 所讨论的是约束条件中的常数项变化而形成了 X 集合，事实上约束条件中任一系数变化都会有一个集合，而且都会通过与 Λ 的对应关系而最终与目标函数 G 集合有对应关系。 Λ 即为边际收益率。可见边际收益率对应着确定的目标函数值，不论约束条件是如何规定的。

六、择优分配原理在连续分配问题中的应用

前面讨论的分配问题，分配的部门是数得清的几个。当分配的部门无穷增多，每个部门的分配量趋近于零时，就成为连续分配问题。择优分配原理也可以应用于连续分配问题，而且可以得到更普遍的结果。连续分配问题的解在工程技术中有很广泛的应用。下面通过三个例子来说明：

1. 火车或汽车省功运行的速度分配规律。

已知火车运行的阻力为速度的凸单调递增函数。给定火车从甲站行驶到乙站的时间为 T ，问火车司机应如何控制速度，使行驶中克服阻力所做的功为极小。

现在的目标函数是 $\min \int_0^S R(v)ds$ ，约束条件是 $\int_0^S \frac{ds}{v} = T$ 。

式中 S 为两站间的距离。欲求速度与距离的关系 $V=V(s)$ 。

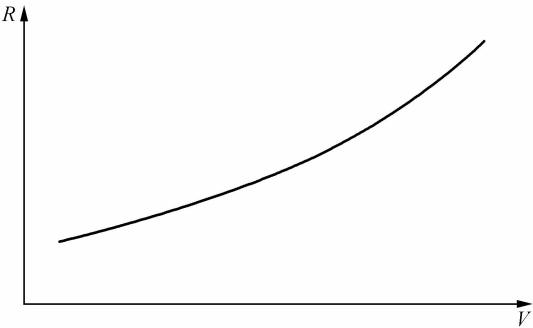


图4 火车阻力和速度的关系

现将 S 等分为 N 小段，每段长均为 ΔS 。于是现在的问题是如何将 T 分给这 N 个小段。每段分给的时间为 ΔT_i ，则 $\sum_{i=1}^N \Delta T_i = T$ 。每段的速度 $V_i = \frac{\Delta S}{\Delta T_i}$ ，分定了 ΔT_i ，就对应一定的 V_i 。从图 4 知道，由于各段的距离均同为 ΔS ，故每一段中 R 与 ΔT_i 的关系完全是同一条曲线。相对于图 2 现有 I、II、……许多条线，但这些线都重合为一条线。凸函数极小化的问题，完全可以用凹函数极大化的原理。在最优分配时边际收益率均相同，即平行切线的切点的横坐标对应最优分配量。现所有的 $g(x)$ 线均重合，平行切线的切点为同一个

点,说明各个小段应用同一个速度运行,即分配同样的 $\Delta T_i = \Delta T$ 。以上的讨论当 $N \rightarrow \infty$ 时同样成立,此时即转变为连续的分配问题。

2. 使成本为最低的生产进度安排

与上例相似的一个问题是:已知生产成本为生产速度的凸函数。现要求在一定的时间 T 内生产一定量的产品量 N ,问应如何安排生产进度,可使总的产品成本为极小。

本问题与上例是完全对应的。也是将时间 T 分配给 N 件产品。每件产品的成本与生产进度的关系就是类似于图 4 的凸函数曲线,而总成本为每件产品的成本之和。根据上例的分析我们得知,成本最低的进度安排是均匀生产,任何突击和会战都会使成本增加。无论成本与生产速度之间函数关系的具体形状如何,只要是凸函数,上述的结论总是成立的。当然,这里假定了其他一切生产条件均不变。

严格地说,本问题与上例是有区别的。上例中阻力只是速度的函数,与加速度无关。而生产成本则与生产的加速度有关,即重新调整生产进度时将发生额外的费用。因此上例完全用静态的分配模型求解是可以的,而本例则理应建立一个包括生产调整费用的模型。但本例所得的结论显然是正确的,因为均匀生产不需用任何调整所发生的费用。

3. 锅炉中燃烧率的最优分配问题

众所周知,锅炉的吸收热量是燃烧率(以每单位时间单位炉篦面积上燃掉燃料的重量计量)的凹函数。在整个炉篦面积上由于各点通风分布的不同,离开传热面的距离不同,而且产生的燃气的路径也不同,所以同样燃烧一定量的煤,在不同的位置燃烧,锅炉吸收的热量是不同的,换言之,锅炉所能吸收的热量 G 是燃烧率 u 及炉篦坐标位置 (x, y) 的函数:

$$g = g(u, x, y),$$

现在的问题是:在一定数量的总燃烧量 U

$$U = \iint u(x, y) dx dy \quad (19)$$

的约束条件下,求能使锅炉吸收的总热量 G

$$G = \iint g(u, x, y) dx dy \quad (20)$$

为极大的燃烧率的分布函数 $U = u(x, y)$ 。

本例所讨论的问题与前面讨论的最优分配问题有一基本的不同点。前面讨论的问题中收益函数 g 仅为分配量的函数,而本问题中收益函数 g 不仅为分配量 u 的函数,而且是 u 所在位置的坐标 (x, y) 的函数,因此 g 函数已不能用如图 1 中的平面曲线来表示,而是一个三维空间中的一个曲面。

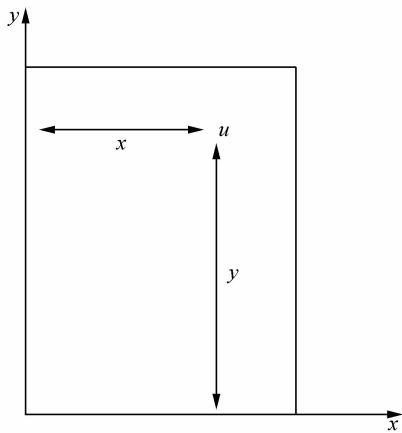


图5 燃烧率的最优分布

这类问题可以用具有等周条件的变分问题的模型来解决，而且结果是解一个退化为一般代数方程的微分方程。但我们基于择优分配原理的讨论可知，最优解必须满足边际产出率相等的条件。现在的边际产出率是 $\frac{\partial g(u \times y)}{\partial u}$ ，故最优解必须满足的条件是

$$\frac{\partial g(u \times y)}{\partial u} = \lambda. \tag{21}$$

λ 的值要根据约束条件 (19) 式来确定。 u 值越大，对应的 λ 越小。 U 即相当于前面讨论的总资源量的限制。

(21) 式有它的经济含义。它说明在实现最优分配时，在炉篋的任一块小面积 $dxdy$ 上投入少许煤量 du 都能使锅炉吸收的热量增加同样的数量 λ 。因此如果有某一小块 $\frac{\partial g}{\partial u}$ 稍大一些的话，就应将该坐标处投的煤 du 改投到该处。如此不断调整，由于 $g(u \times y)$ 函数的凹性，最后必须使得各处的燃烧率的边际产出率相等。现 $g(u \times y)$ 函数是已知的，故 (21) 式就是问题的解。这种形式的解要求资源 u 的分布是一个随坐标变化的函数，这与前面讨论的结果不同。进一步推广使择优分配原理的应用有更多方面的可能。

在连续分配问题中应用择优分配的概念，主要用于解决工程技术方面的问题。

七、动态问题中择优分配原理的应用

动态问题和静态问题的根本区别在于：动态问题的最优方案不但要考虑决策所得的收益，而且与规定的始终点的条件有关，当始终点的条件改变

时,最优方案跟着改变。一个动态问题往往可以在空间上或实践上分成若干阶段。每一小段采取不同的策略,不但影响到该小段中的收益,而且影响该小段终点的状态,而且因此又影响到下一小段的方案的选取。以国民经济规划为例,始点条件是规划开始时的经济参数,如固定资产数、人口等;规划的目的可能是五年计划内人民消费总额为极大;同时要求期末的固定资产、人口等达到某一预定目标。要解出最优的积累比例、人口的增长率等。显然,如果期末要求达到的固定资产越多,则规划期内人民消费总额将减少,但这对下一个规划期生活的改善创造了条件。因此积累率的选取不但要满足目标函数极大,还应保证期末的经济参数。这一类经济规划问题往往还有一系列约束条件,如资源的限制、部门间经济联系的限制等。经济规划是一个非常复杂的问题。下面我们将讨论动态问题中最简单的情况,即最速降线,或称捷降线的问题。这是古典变分法中的一个典型问题。它要求解出从始点 $P_0(x_0, y_0)$ 到终点 $P_N(x_N, y_N)$ 的一条曲线的方程,使质点沿该曲线依靠重力从始点滑行至终点所花的时间为最少,忽略摩擦阻力。

连接 P_0P_N 的直线是两点间最短的线,可是质点沿直线运行时间未必最短,因为质点未能在开始时获得较大的加速度。如果开始时先垂直向下,加速度虽然最大,行经的距离却加长。因此所求的线可能是介乎上述两种情况之间的某一条曲线。捷降线的原理在铁路编组站驼峰纵断面的设计中有所应用,但要增加一个相邻车辆车钩夹角的限制条件。

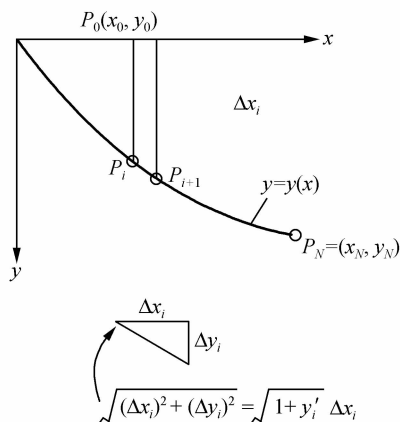


图 6 捷降线问题

从图 6 中设 $y=y(x)$ 为所求的捷降线,曲线经过 P_i 及 P_{i+1} 两点。根据物理学定理,质点到达 P_1 点时的速度为 $\sqrt{2gy_1}$, y_1 为 P_1 点的纵坐标。按照积分学的原理,质点从 P_0 到 P_N 所花的时间 G 为:

$$G = \int_{x_0}^{x_N} \frac{\sqrt{1 + y'^2}}{\sqrt{2gy}} dx, \quad (22)$$

或者写成和式

$$G = \lim_{\Delta x_i \rightarrow 0} \sum_{i=0}^{N-1} \frac{\sqrt{1 + y_i'^2} \Delta x_i}{\sqrt{2g \sum_{j=0}^i \Delta y_j}}, \quad (23)$$

更一般地考虑被积函数为一个一般的函数 $g(y, y')$, 则

$$G = \lim_{\Delta x_i \rightarrow 0} \sum_{i=0}^{N-1} g\left(y_0 + \sum_{j=0}^i \Delta y_j, y_i'\right) \Delta x_i \quad (24)$$

从择优分配的概念来考虑捷降线问题, 我们很自然地 will $y_N - y_0$ 视为持有的资源量, 因为 $y_N - y_0 = \Delta Y$ 为全部落差, 质点的速度是依靠落差而获得的。现在的问题是如何将落差 ΔY 分配给各个小段, 使收益函数 G 取极值。一旦对于各个小段 Δx_i 分得的落差 Δy_i 被确定, 整个曲线 $y = y(x)$ 就得到了。由于总的落差是有限的, 于是在 G 取得极值的同时, 必须满足约束条件:

$$\sum_{i=0}^{N-1} \Delta y_i = y_N - y_0 = \Delta Y, \quad (25)$$

同时 Δy_i 又与 y_i' 存在着关系:

$$\Delta y_i = y_i' \cdot \Delta x_i, \quad (i = 0, 1, \dots, N-1). \quad (26)$$

(26) 式也是必须满足的约束条件。

(24)、(25)、(26) 式已构成了从择优分配概念描述的捷降线的变分问题。但捷降线问题还不是变分问题的普遍的提法, 因为 (23) 或 (24) 式的被积函数或 g 函数中未出现 x , 即收益与分配部门所在的坐标位置 x 无关。在更一般的变分问题中被积函数或收益函数同时与 x, y, y' 有关。即应取极值的目标函数 G 为

$$G = \int_{x_0}^{x_N} g(x, y, y') dx, \quad (27)$$

边界条件为

$$\left. \begin{array}{l} x = x_0, y = y_0 \\ x = x_N, y = y_N \end{array} \right\}, \quad (28)$$

x_0, y_0, x_N, y_N 均为给定的数。

现在可以将变分法所要解决的问题从分配与收益的意义来理解, 并用择优分配的原理来解决。 $g(x, y, y') dx$ 为位于 x 的无穷多个待分配部门之一的收益函数。 y 为累计投入的资源量, 在问题中新投入的总资源量限制为 $y_N - y_0$, 待分配的部门的坐标位置在 x_0 至 x_N 之间。

将目标函数写成和式

$$G = \lim_{\Delta x_i \rightarrow 0} \sum_{i=0}^{N-1} g(x_0 + \sum_{j=0}^i \Delta x_j, y_0 + \sum_{j=0}^i \Delta y_j, y'_i) \Delta x_i. \quad (29)$$

约束条件 (25)、(26) 式依旧成立。

于是我们将无约束的变分问题转化为具有约束条件的广义的资源分配问题。

设按问题的物理意义或经济意义, 在所讨论的 x 和 y 的变化范围内这个和式是收敛的。现应用 Lagrange 乘数法, 构成新函数 H :

$$H = \lim_{\Delta x_i \rightarrow 0} \left[\sum_{i=0}^{N-1} g(x_0 + \sum_{j=0}^i \Delta x_j, y_0 + \sum_{j=0}^i \Delta y_j, y'_i) \Delta x_i + \sum_{i=0}^{N-1} \lambda_i (\Delta y_i - y'_i \Delta x_i) + \Delta \left(\sum_{i=0}^{N-1} \Delta y_i - \Delta y \right) \right]. \quad (30)$$

此处的每一个 λ_i 均不同, 它们随着 i 或者说随着 x 而变, 是 x 的函数。

为使 H 取得极值, 可以变化各个 Δy_i 来实现, 也可以变化 y'_i 来实现, 尽管由于 (26) 式的联系, 二者实质是相同的, 但在数学上确有不同的形式。我们将先后从这两方面来加以讨论。假定 $g(x, y, y')$ 对于 y 及 y' 的偏导数存在, 且 (30) 式为一致收敛。现在首先令变化 y_i 使 H 达到极值, 则有:

$$\frac{\partial H}{\partial y'_i} = \lim_{\Delta x_i \rightarrow 0} [g_{y'}(x_0 + \sum_{j=0}^i \Delta x_j, y_0 + \sum_{j=0}^i \Delta y_j, y'_i) \Delta x_i - \lambda_i \Delta x_i] = 0,$$

或

$$\lim_{\Delta x_i \rightarrow 0} g_{y'}(x_0 + \sum_{j=0}^i \Delta x_j, y_0 + \sum_{j=0}^i \Delta y_j, y'_i) - \lambda_i = 0, \quad (i = 0, 1, 2, \dots, N-1), \quad (31)$$

(31) 式对于任何一个中间点 x_i 处均成立, 因此可写为:

$$g_{y'}(x, y, y') = \lambda, \quad (32)$$

此处 λ 为 x 的函数。

其次再来变化 Δy_i , 使 H 取极值, 则有:

$$\frac{\partial H}{\partial \Delta y_k} = \lim_{\Delta x_i \rightarrow 0} \sum_{i=k}^{N-1} g_k(x_0 + \sum_{j=0}^i \Delta x_j, y_0 + \sum_{j=0}^i \Delta y_j, y'_i) \Delta x_i + \lambda_k + \Delta = 0, \quad (k = 0, 1, 2, \dots, N-1). \quad (33)$$

因为 H 中对 g 求和时足标等于及大于 k 的项内均包含有 Δy_k 。(33) 共有 N 个等式。取极限得:

$$\int_{x_k}^{x_N} g_y(x, y, y') dx + \lambda_k + \Lambda = 0, \quad (34)$$

假设被积函数 g_y 的原函数为 $R(x, y, y')$, 则 (34) 式可写成:

$$R(x_N, y_N, y'_N) - R(x_k, y_k, y'_k) + \lambda_k + \Lambda = 0. \quad (35)$$

(35) 式中第一项为常数, 第二项、第三项对于 $(x_0 < x_k < x_N)$ 中的任一点 x_k 均成立。将 (35) 式对 x 求导, 注意到 Λ 为常数, 得:

$$-g_y(x, y, y') + \frac{d\lambda}{dx} = 0, \quad (36)$$

将 (32) 式对 x 求导:

$$\frac{d}{dx} g_y(x, y, y') = \frac{d\lambda}{dx}, \quad (37)$$

将 (36) 式代入 (37) 式即得欧拉方程

$$g_y - \frac{d}{dx} g_y = 0. \quad (38)$$

至此, 我们按择优分配的概念同样得出了欧拉方程。择优分配原理不但可用于解决静态问题, 亦可用于解决动态问题。资源的限制可以从广义来理解, 等周条件、边界条件等都可以转化作资源限制来处理, 于是这个原理就可灵活地应用到许多提法不同的问题。择优分配对不同的最优化问题可以作出共同的解释, 我们认为它是最优化方法的一个普遍基础, 所以称它为择优分配量最优化原理。

(32) 式表明了收益函数 y' 的边际收益率 g_y 不再是一个常数, 而是一个 x 的函数 λ 。

上面的推导过程也提示了求欧拉方程数值解的一个方法。(34) 式可写为:

$$\int_{x_0}^{x_N} g_y(x, y, y') dx - \int_{x_0}^{x_k} g_y(x, y, y') dx + \lambda_k + \Lambda = 0, \quad (39)$$

于是

$$\lambda_k = \int_{x_0}^{x_k} g_y(x, y, y') dx - \int_{x_0}^{x_N} g_y(x, y, y') dx - \Lambda, \quad (40)$$

但 $\int_{x_0}^{x_k} g_y(x, y, y') dx$ 为常数, 可与 Λ 合并, 令

$$\int_{x_0}^{x_k} g_y(x, y, y') dx + \Lambda = \Lambda', \quad (41)$$

再将 (40) 式写成极限求和式, 得:

$$\lambda_k = \lim_{\Delta x_i \rightarrow 0} \sum_{i=0}^k g_y(x_0 + \sum_{j=0}^i \Delta x_j, y_0 + \sum_{j=0}^i \Delta y_j, y'_i) \Delta x_i - \Lambda', \quad k = 0, 1, 2, \dots, N-1 \quad (42)$$

将 (42) 式与 (26) 式、(31) 式联立, 当设定一个 Λ' 之后, 对于每一个 i 均可求出 Δy_i , $\Delta y'_i$ 及 λ_k 。逐步解到 $i = N-1$ 时, 每一个 Δy_i 均可求得。如果各个 Δy_i 能满足 (25) 式 $\sum_{i=0}^{N-1} \Delta y_i = \Delta Y$, 则说明设定的 Λ' 是合适的, 否则可另设 Λ' 的数值重新逐步求解, 直到 (25) 式满足为止。实际计算时可绘出 Λ' 与 $\sum_{i=0}^{N-1} \Delta y_i$ 的关系曲线, 从曲线上取 $\sum_{i=0}^{N-1} \Delta y_i = \Delta Y$ 处 Λ' 的值, 可减少试凑的计算工作量。对应每一个 Λ' 值计算得到的每一条 $y = y(x)$ 曲线都是问题的最优解, 所不同的只是终点条件不同。

八、欧拉方程的经济含义

在捷降线的研究中将落差 ΔY 视作资源, 将落差在横坐标中进行分配, 结果可以得出捷降线的解。一般地, 将纵坐标的改变视作资源用择优分配的原理推导欧拉方程时, 正如前面表述的, 还要借助于 Lagrange 乘数法。事实上, 根据择优分配的结果——边际收益率相等为最优化的必要条件, 我们可以直接得出欧拉方程。这个边际收益率相等的具体表述, 已在第五节中给出。现在我们将它应用到动态问题的最优解, 并以此直接得出变分法中的欧拉方程。

现讨论最基本的变分问题, 它仍由 (27) 及 (28) 式所规定。设在图 7 中 P_0 及 P_N 为给定的两个端点, 而 $y = y^*(x)$ 为能使 (27) 式达到极限的最优解, $P_i(x_i, y_i)$ 为极限曲线上任一点。目标函数 (27) 式可写成和式:

$$G = \lim_{\Delta x \rightarrow 0} \sum g(x, y, y') \Delta x. \quad (43)$$

任一通过 P_0 、 P_N 两点的曲线 $y = y(x)$ 上的任一点 P , 有坐标 x 及 y , 并有该处的 $y' = \frac{dy}{dx}$ 。将这三个数值代入 $g(x, y, y') \Delta x$ 可得出构成 G 的一个微单元。但仅在极值曲线 $y = y^*(x)$ 上才能使 (43) 式的和为极大 (小)。

参考图 7, 在极值曲线 $y = y^*(x)$ 上任一点 P_i 处, 设保留该点的坐标 (x_i, y_i) 不动, 而令 y'_i 作微量改变。此时在 P_i 点处的 G 的微单元 $g(x_i, y_i, y'_i) \Delta x$ 将产生相应的改变。当忽略高阶项时其改变量为:

$$g_{y'}(x_i, y_i, y'_i) \Delta x \Delta y'_i. \quad (44)$$

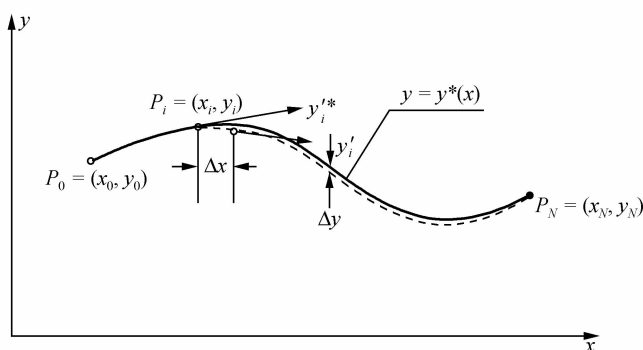


图7 从边际收益概念推导欧拉方程

这就是由于在 P_i 点处曲线的方向作了 $\Delta y'_i$ 的改变而对目标函数 G 的影响，或者说，这就是由于曲线切线方向改变而得到的边际收入。但由于在 P_i 处的方向改变，使得以后曲线的纵坐标跟着发生变化。原来横坐标递增 Δx ，纵坐标改变 Δy ，现在 Δy 本身将改变 $\Delta^2 y$ ，即 y 的二阶差分。故微单元的改变为：

$$g_y(x, y, y') \Delta x \Delta^2 y.$$

从 P 点以后每个微单元都有这样的改变量，故一共使目标函数 G 改变了

$$\int_{x_i}^{x_N} g_y(x, y, y') dx \Delta^2 y. \quad (45)$$

这是由于曲线在 P_i 处方向改变而不得不引起的一笔边际损失，因为约束条件是曲线最后必须通过 $P_N(x_N, y_N)$ 。

如果 $y = y^*(x)$ 是极值曲线，则我们可以断言，边际收入必定等于边际损失。如果二者不等，则通过这样的改变将有净的收益，亦即 $y = y^*(x)$ 还可以改善，它不是极值曲线。因此有

$$g_y(x_i, y_i, y'_i) \Delta x \Delta y'_i + \int_{x_i}^{x_N} g_y(x, y, y') dx \Delta^2 y = 0. \quad (46)$$

上述的极值条件在任一点处 x_i 均应满足，因此特定的点 $P(x_i, y_i)$ 可换为曲线上的任意点 $P(x, y)$ ，又因 $\Delta y' = \frac{\Delta^2 y}{\Delta x}$ ，故有：

$$g_y(x, y, y') \Delta^2 y + \int_x^{x_N} g_y(x, y, y') dx \Delta^2 y = 0. \quad (47)$$

对 x 微分，并注意到 x_N 为常数，得

$$\frac{d}{dx} g_y(x, y, y') - g_y(x, y, y') = 0, \quad (48)$$

即为欧拉方程。

从上述的推导过程我们可以察知，欧拉方程是有它的经济含义的。它说明，极值曲线具有这样的特性，在曲线的任一点稍微改变其方向所得的边际收益（损失），将恰为以后曲线位置改变所发生的边际损失（收益）所抵消；曲线无论在哪一点，朝哪个方向移动都不能改善目标函数的值。

参 考 文 献

Bellman, R., and S. Dreyfus, *Applied Dynamic Programming*. Princeton, N. J. : Princeton University Press, 1962.

Principle of Optimal Allocation and Its Applications

YUSHI MAO

(Unirule Institute of Economics)

Abstract The problem dealt with in this paper is how to allocate resources to achieve maximum effectiveness. Starting from a simple problem, to distribute chemical fertilizer on two plots of land to obtain maximum yield by continuous improvements on the margin, the method of Lagrange Multiplier can be deducted. This idea can be extended to dynamic problems. Borrowing the problem of Brachistochrone, to find a path descending along which a movable particle is to reach a given point in a vertical plane in the shortest time by the force of gravity, the same idea of resources allocation is applied to dynamic cases, where the height difference is treated as resources. Through further generalization, the Euler differential equation in calculus of variation can be derived. The Principle of Optimal Allocation is a fundamental principle of optimization and explains the economic meanings of both the Lagrange Multiplier method and the Euler equation in calculus of variation. This Principle has a wide area of applications.

JEL Classification B41, C02, C61

我国金融机构的存差：历史沿革和含义变化

汪 洋*

摘 要 1994年以来,我国银行体系的存差规模不断扩大,并成为国内学者关注的焦点。针对当前国内学者对金融机构“存差”概念的不同认识,本文旨在为分析这一问题提供合理的分析视角。研究表明:金融机构存差的机构口径差异是导致国内学者出现模糊认识的关键。本文从信贷管理体制历史沿革的角度对不同时期我国金融机构存差的含义进行了分析,解释了区域金融的视角下对存差扩大的不同理解。文章最后还讨论了我国金融机构存差的扩大与目前国内流动性过剩之间的关系。

关键词 存差,流动性,外汇占款

一、引 言

1994年我国金融机构新增存款10845亿元人民币,新增贷款7867亿元,开始出现了增量意义上的存差2978亿元。1995年我国金融机构存款余额为53862亿元,贷款余额为50538亿元,出现了存量意义上的存差3324亿元人民币。此后这一规模逐年扩大,与此同时我国金融机构的贷存比也逐年下降,并引起国内经济学家们的关注。2002年以来,这一规模迅速扩大,2006年年末金融机构人民币各项存款余额为33.54万亿元,各项贷款余额22.53万亿元,存差规模达到11.01万亿元人民币,占同期广义货币供应量 M_2 的比重达到32%。为什么存差规模不断扩大?它又会产生何种影响?这些问题已经构成我国学者近年来探讨的焦点问题之一(孙国峰,2002;胡海鸥、赵慈拉,2002;张红地,2003;张杰,2003;殷孟波、贺光明,2003;江其务,2003;伍志文等,2004;李德,2004;孙工声,2005;李斌,2006;张晓慧,2006;李宏瑾,2006;李扬、彭兴韵,2006;等等)。

在对这一现象的讨论当中,主要的分析思路包括以下几方面:一是将存差与我国宏观经济与金融的运行特点,如通货紧缩、惜贷、资金过剩以及储蓄大于投资、国际收支顺差所造成的外汇占款增加等问题联系起来,分析存差与这些因素之间的关系;二是将存差与商业银行的运行特点以及金融制度安排联系起来,如资金运用多元化、分支机构揽存的制度激励和对贷款的制

* 江西财经大学金融学院,江西财经大学金融发展与风险防范研究中心。通讯地址:南昌市双港路江西财经大学金融学院330013。电话:(0791)3816792。E-mail:wy0329@yahoo.com.cn。本文系江西省社科“十一五”(2006—2007)规划项目《我国金融机构存差问题的研究:以江西省为例》的阶段性成果,课题编号06YJ19。作者感谢匿名审稿人的修改意见,文中错误与疏漏之处由作者负责。

度约束以及邮政储蓄的只存不贷等等；三是从制度经济学的角度出发，认为我国政府对金融机构的控制造成了银行体系存差规模的扩大。当然，彼此关于存差影响因素的分析也不是完全冲突，只不过更强调其中的某一点或几点。除此之外，在对存差问题的讨论当中，不少学者还从区域金融的角度出发，探讨本地区金融机构存贷差的变化原因，继而扩展到对全国范围内存贷差变化的含义，如何田（2004）。在研究方法上，有两个特征非常明显，一是采用计量经济学的方法对存差规模的影响因素进行分析（伍志文等，2004；李宏瑾，2006），二是利用银行会计分录和金融机构资产负债简表作为文字性论述的辅助。

笔者以为，我国学者对存差问题的不同理解，基本源于两个方面。一是忽略存差的机构口径，有的甚至就根本没有涉及金融机构存差的口径问题，笼统地讨论存差的影响因素；更多的是混淆商业银行视角下的存差和包括商业银行和中央银行在内的整个银行体系视角下的存差口径。这种状况在 2006 年以来有所改变，张晓慧（2006）明确指出应该以整个银行体系视角下的存差分析为主，因为其宏观经济含义更为显著。二是混淆区域经济和全国经济视角下的存差问题，其本质是从本地区金融机构存差出现和变化的原因出发，得出若干政策建议。这些政策建议从全国来看不具有实践意义，也不能作为制定或改变货币政策的依据；但是基于全国视角的研究得出的结论从区域金融的角度看又缺乏足够的说服力，尤其是对某些经济欠发达的地区来说。

针对上述缺陷，本文将主要从以下三个方面，即从我国信贷管理体制变迁的历史视角、金融机构口径差异的视角和区域金融的视角，来探讨我国金融机构的存差问题，旨在揭示国内学者不同观点背后的深层次原因。文章最后还将银行存差与目前我国金融机构流动性过剩问题联系在一起讨论。本文的结构安排如下：第一部分是引言；第二部分则是从历史视角出发，探讨我国不同历史时期存（贷）差含义的实质；第三部分主要从机构口径的差异出发，探讨现阶段我国金融机构存差的影响因素；第四部分从区域金融的角度出发，分析我国若干经济欠发达地区的经济主管部门将存差扩大和金融机构对当地经济支持力度下降联系在一起的深层次原因；第五部分探讨我国金融机构存差和银行体系流动性“泛滥”之间的关系；最后是结论。

二、存差问题的分析视角之一：历史视角

（一）统存统贷信贷管理体制下的存贷差问题

改革开放之前，我国金融机构就只有中国人民银行；改革开放之后，恢复与成立各类金融机构，打破中国人民银行“大一统”的垄断局面是这一阶段最主要的特点。1984 年，中国人民银行转变为中央银行，在此之前，它既

承担中央银行的职能，又扮演商业银行的角色。在信贷管理体制上，则以中国人民银行的“统存统贷”信贷管理体制为核心。

1. 统存统贷管理体制下的主要特征：贷差

改革开放之前，中国人民银行的资金来源和运用的种类较少，在资金运用方仅有黄金外汇占款和各类贷款项目，资金来源方有各类存款和货币发行项目。

表1 中国人民银行资金来源运用表

黄金外汇占款	货币发行
贷款	各种存款
—工业贷款	—储蓄存款
—商业贷款	—企业存款
—.....	—财政存款

由于黄金外汇占款项目的年度变化非常小，在中国人民银行制定年度信贷计划和货币发行计划过程中，该项目往往可以忽略不计，因此有以下公式成立，

$$\text{各项贷款} = \text{各项存款} + \text{货币发行}, \quad (1)$$

如果从增量的角度看，有以下公式成立，

$$\text{新增各项贷款} = \text{新增各项存款} + \text{新增货币发行量}. \quad (2)$$

因此，从（2）式分析，必然有贷差出现，而不是出现存差；否则“货币发行”项目余额将下降，其经济含义是当年没有货币净发行，而是货币净回笼，这表明经济处于紧缩或萧条状态。所以说，从全国来看，贷差是一种必然现象。此时中国人民银行扮演的是中央银行和商业银行的双重角色，因此（2）式反映的是整个银行体系的存贷关系。

2. 存差是否有可能出现？

如果说贷差是我国信贷计划管理体制的主要特征，那么是否意味着存差就不会出现呢？答案是否定的。从特定时间段来分析，尤其是每年春节过后，全国货币净回笼基本上是一个周期性的特征，这就构成了我国金融机构在特定时期的存差出现。然而，这并不表明在更长的时间跨度内“货币发行”项目一直处于净回笼的状态，这可以从我国“货币发行”项目不断上升的事实当中得到证实。从更长的周期来看，因为“货币发行”所形成的流通中现金是不断增加的，所以在信贷计划管理体制下通常只会出现贷差。即使1998年之后我国中央银行完全取消信贷计划管理制度，从全国范围来看，春节之后的几个月内现金净回笼仍然是一种常态，但此时（增量意义上的）存差仅仅是年度内的一种周期性现象。

(二) 信贷差额分级管理体制下的存贷差问题

1. 存差和贷差的实质

自 20 世纪 70 年代末开始,我国开始实行金融体制的市场化改革。在信贷管理体制上,1979—1983 年中国人民银行实行了“统一计划,分级管理,存贷挂钩,差额控制”的管理方法,这是中国人民银行尚未全面向中央银行转变之前的一种信贷管理体制。与此前实行的统存统贷体制相比,信贷计划管理的特色基本没有变化,不过金融机构主动吸收存款的意义在日渐扩大。对基层的专业银行而言,这标志着以多存多贷为核心的市场化改革开始走上前台。

信贷差额管理体制的主要特征包括两方面:一是分级管理,二是差额控制。所谓分级管理,就是中国人民银行总行根据国家批准的信贷计划,通过逐级层层下达信贷差额计划来管理与核定各地信贷规模和现金投放;所谓差额管理,就是中国人民银行的上级行对下级行重点考核的是存贷差额额度,人民银行各分行在完成存款计划和不突破差额的前提下多存可以多贷。总行通过汇总各地的信贷差额保证全国总的信贷差额控制在计划范围内,并与现金投放计划相配套。这当中值得强调的是,在金融机构资产负债仍然以贷款和存款为主的情况下,由于各地经济和金融发展水平的差异,人总行对各地分支机构的信贷差额控制方法并不完全相同。通常对少数经济发达地区的信贷管理采取的是存差管理方法,而对经济欠发达地区的信贷管理则采取贷差管理方法。前者要求经济发达地区回笼现金,即年度贷款增量小于存款增量,这意味着经济发达地区即使有足够的存款来源也不能发放超额度的贷款;后者则意味着在经济欠发达地区贷款的投放要大于其吸收的存款,这表明欠发达地区即使没有足够的存款也能够发放贷款(尚明、吴晓灵、罗兰波,1992;张新泽、王毅,1999)。

从区域经济与金融(即分级管理)的角度来看,由于贷款是当时企业获得外部资金的主要来源,欠发达地区的贷差规模越大,其经济主管部门就会认为金融机构对当地经济发展的信贷支持力度就越大,因此全国大多数地区都将贷差的大小视为金融机构支持当地经济发展的力度。从此时开始,“存差”和“贷差”概念浮出水面,开始成为我国银行界的专业术语之一,日后逐渐成为国内经济学家关注的焦点。

2. 存贷差的政策意义

显然,在“大一统”银行体制和“统存统贷”的信贷管理体制下,从区域资金来源与运用的角度分析,也存在资金来源大于资金运用和资金来源小于资金运用两种情况,但当时存贷款均通过指标管理,存贷双方并不挂钩,存差(贷差)在当时并不具有现实的管理意义,更不具有对当地经济实施金融支持的政策含义。因此,只有从我国开始实施信贷差额管理体制开始,信

贷计划的分级管理才使得存差（贷差）指标有了现实意义。

必须指出的是，将贷差规模的大小作为当地金融机构支持本地经济发展力度的看法置于全国经济背景下则可能是完全错误的。因为从本地经济来看，贷差的扩大是信贷投放的增加；如果放大到全国范围来看，贷差的扩大通常意味着全国信贷规模的膨胀，其结果往往与通货膨胀相联系。由于本地经济主管部门只关注当地经济发展速度，对全国的通货膨胀并不负任何责任；人总行在当时不仅负有促进经济增长的责任，还肩负控制通货膨胀的职责，其调控方法就是通过控制贷款投放总额和存贷差额来实现的。如果为刺激国内经济增长放任贷差规模的任意扩大，必然导致全国的通货膨胀。所以说，在信贷差额管理体制下，存差仅仅是一个周期性和区域性的金融现象，其出现源于中国人民银行的信贷差额分级管理体制。事实上，此时存差仅仅是少数经济发达地区的信贷特征，并不为经济学家们所关注。

（三）实贷实存信贷管理体制下的存贷差问题

20 世纪 80 年代，中国金融体制改革中意义最为深远的莫过于中国人民银行开始履行中央银行的职能，我国的银行体系从过去的大一统银行体系向二级银行体系转变。在调控方法上，开始引进西方发达国家通常采用的法定存款准备金制度，这就使得中国人民银行必须开始尝试二级银行制度下的基础货币供给问题。该信贷管理体制下的主要特征表现为：

第一，再贷款是中国人民银行注入基础货币的最主要途径。与西方国家中央银行注入基础货币的主流方法——通过购买国债的公开市场操作方式——不同，在中国国债市场刚刚起步、外汇占款规模太小的现实约束情况下，再贷款是当时中国人民银行向整个银行体系注入基础货币唯一的可行途径。为了适应新的中央银行体制，中国人民银行采取了新的信贷资金管理办法，即“统一计划、划分资金、实贷实存、相互融通”。其核心是“实贷实存”，从中央银行宏观间接调控金融机构的角度来解释，就是中央银行通过再贷款途径向当时的专业银行提供基础货币，这就是我们通常所说的“上贷下存”的方法，即中央银行贷款给专业银行，专业银行立即将资金存入中央银行。

对于专业银行而言，这一操作的结果就是在其资产负债表的资产方的“超额存款准备金”增加，负债方的“向中央银行借款”增加（见表 2）；对中央银行而言，其资产负债表的变化就是资产方的“对金融机构再贷款”增加，负债方的“金融机构超额存款准备金”增加。

表 2 “上贷下存”信贷管理模式在专业银行和中央银行资产负债表上的反映

中央银行		专业银行	
资产	负债	资产	负债
再贷款 +	超额存款准备金 +	超额存款准备金 +	向中央银行借款 +

第二,整个银行体系的信贷特征仍然以贷差为主。

在“实贷实存”体制实施的初期,我国中央银行资产方的“外汇占款”数额比较小(基本可以忽略),因此中央银行资产负债表大致可以简化为以下公式

$$\text{再贷款} = \text{货币发行} + \text{存款准备金}. \quad (3)$$

通常,货币发行是不断增长的,因此必然有再贷款大于存款准备金(包括法定存款准备金和超额存款准备金两部分),这意味着专业银行全部的流动性都通过中央银行再贷款途径解决,形成商业银行的“超贷”现象。这也可以从商业银行资产负债表得到反证。它可以简化为以下公式

$$\text{库存现金} + \text{存款准备金} + \text{各项贷款} = \text{各项存款} + \text{再贷款} + \text{资本金}. \quad (4)$$

假定忽略资本金项目,由于再贷款余额要大于存款准备金余额,因此商业银行的各项贷款必然大于各项存款,在整体上形成商业银行视角下的贷差。因此,贷差规模的大小也就可以视为中央银行对当地经济信贷支持力度的强弱。

如果将中央银行资产负债表和商业银行资产负债表合并,形成整个银行体系的资产负债表,将会有以下公式成立

$$\text{各项贷款} = \text{流通中现金} + \text{各项存款}. \quad (5)$$

(5)式和(1)式完全一致(“流通中现金”项目由“货币发行”扣减“库存现金”得到)。这表明实贷实存信贷管理体制下贷差依然是该制度下最为主要的信贷特征。因此,我们可以认为:即使是在二级银行体制下,即使商业银行需要交纳法定存款准备金,存差也并不是一个必然现象。这很大程度上取决于中央银行注入基础货币的主要方式。

第三,从区域信贷管理的角度看,贷差仍然具有金融机构支持当地经济发展的意义。

与信贷差额管理体制一样,实贷实存的信贷管理体制仍然保留了信贷计划管理和分级管理两个特征,其差异体现为实贷实存的信贷管理体制形成了对专业银行的资金硬约束,也就是专业基层行如果没有资金来源,即使有贷款额度也不能放款;如果有足够的资金来源,但没有贷款额度,该行可以通过同业拆借将资金贷给其他金融机构。当然,这一管理体制也存在为国内学者所诟病的“资金、计划两张皮”的弊端。尽管二级银行制度形成了,但是中国人民银行通过层层下达信贷计划的管理模式并没有改变,这就使得存(贷)差在区域信贷管理上仍然具有探讨的意义。与信贷差额控制体制一样,在企业仍然主要通过银行贷款进行融资的条件下,某一地区的贷差规模或指标仍然具有金融机构对当地经济发展的信贷支持力度的政策含义。顺着这一思路,贷差规模的下降乃至存差的出现,从区域金融管理的角度来看,不考

虑发行股票、债券融资等方式，我们就可以视为金融机构支持当地经济的力度在下降。与信贷差额管理体制相比，这一政策含义基本没有变化。因此，我们对金融机构存贷差的分析，就必须从全国和区域两个角度来分析才是完整的。

（四）1994 年人民币汇率并轨改革后存贷差的新变化

1994 年我国实行外汇管理体制变革，人民币汇率并轨刺激了当年的外贸顺差，资本与金融账户也出现大规模顺差，国际收支的双顺差使得我国外汇储备迅速扩大。与此同时，存差问题从过去一个局部的区域金融现象开始演变为一个全国性的金融现象，并逐渐引起国内学者的关注。1994 年全国新增存贷款出现存差 2978 亿元，1995 年全国存贷款余额出现存差 3324 亿元人民币，此后存差问题一直困扰着国内经济学者。

为什么我国会在 1994 年开始出现增量意义上的存差现象呢？为什么又和人民币汇率并轨的年份如此巧合呢？笔者以为，我国中央银行注入基础货币渠道的变化决定了金融机构存差的出现是一种必然现象。如前所述，实贷实存体制实施之初，中央银行再贷款是货币当局注入基础货币的主渠道，如果货币当局注入基础货币的主渠道发生变化，就必然影响存贷差的规模。1994 年人民币汇率并轨改革对货币当局的直接结果就是在强制结售汇制度下外汇储备的迅速增加，达到 304 亿美元，按 1 美元等于 8.6212 元人民币的年平均汇价计算，这相当于中央银行注入了 2621 亿元的基础货币。由于 1994 年通货膨胀率创历史最高，达到 21.7%，中国人民银行为抑制通货膨胀，必须采取相应的对冲方式来控制基础货币的投放。由于强制结售汇制度下外汇储备增加的被动性，货币当局只有通过收回再贷款来收缩金融机构的流动性。这意味着货币当局注入基础货币的渠道发生了根本性的变化，即从再贷款途径转变为外汇占款途径。

就中央银行而言，（3）式变化为

$$\text{再贷款} + \text{外汇占款} = \text{货币发行} + \text{存款准备金}. \quad (6)$$

就整个银行体系而言，（5）式转变为

$$\text{外汇储备} + \text{各项贷款} = \text{流通中现金} + \text{各项存款}. \quad (7)$$

如果以增量的形式来看，（7）式可以改为

$$\text{新增外汇储备} + \text{新增贷款} = \text{新增流通中现金} + \text{新增存款}. \quad (8)$$

当年新增外汇储备大于年度流通中现金的增加值时，整个银行体系新增存贷款出现存差就是正常现象了。伴随着增量存差的出现，存贷款余额出现存差则是迟早的事情了。因此，我国 1994 年新增存贷款在全国范围内出现存差的根本原因在于当年的人民币汇率并轨改革导致的外汇储备高速增长。

1995 年出现存量意义的存差,则是 1994 年外汇储备增长趋势的延续而已。因此,可以认为 1994—1996 年我国金融机构开始出现存差,乃至存差不断扩大完全是中央银行资产多样化的结果,与商业银行无涉。由此得到的推论则是我们对金融机构存贷差的分析必须从两个角度来分析,一是商业银行角度,二是包括中央银行在内的整个银行体系的角度,如此才可能完整和准确。

上述分析还可以澄清两点模糊认识:第一,1994 年我国金融机构存差的出现恰恰是外汇储备增加后的结果,而不是因为存差是用于收购外汇储备的主要资金来源。第二,1994 年我国出现增量意义上的存差并不是由于当年开始实施的商业银行资产负债比例管理的缘故。在我国推行的这一办法当中,其中规定贷存比必须小于或等于 75%,因此有学者认为是因为这一比例规定的执行我国才出现存差的。事实上,四大行在资产负债比例管理的初期,实行的是增量上的资产负债比例管理,从总量上看贷存比仍然远高于 75%。之所以出现存差,主要是因为外汇占款迅速扩大的缘故。

(五) 1997 年下半年通货紧缩之后存贷差的变化

在 1997 年下半年我国经济陷入通货紧缩之后,存差规模不断扩大,不少学者认为这是金融机构“惜贷”的表现之一;有的学者则认为存差是储蓄增长相对过快、贷款增长相对过缓的结果;甚至还有经济学家认为正是因为金融系统的存差才支持了政策当局积极的财政与货币政策。2002 年以来,存差的扩大则被视为国内流动性过剩、资金闲置的表现。我们到底应该如何来认识我国金融机构存差的扩大?

1. 中央银行资产负债表的变化

众所周知,1997 年 7 月亚洲金融危机正式爆发,此后不久中国出现了通货紧缩,这种情形一直持续到了 2002 年。在经济增长乏力的背景下,我国政府推出了稳健的货币政策和积极的财政政策(实际上就是扩张性的货币政策与财政政策)。具体到货币政策方面,在人民币汇率面临贬值压力、资本大量外逃和外汇储备增加额远低于经常账户顺差额的大背景下,我国中央银行注入基础货币的主渠道发生了新的变化,从 1994—1996 年间的外汇占款转变为以购买国债、对资产管理公司(AMC)的债权为主,这就使得(6)式变为

$$\begin{aligned} & \text{再贷款} + \text{外汇占款} + \text{对政府债权} + \text{对 AMC 的债权} \\ & = \text{货币发行} + \text{存款准备金} \end{aligned} \quad (9)$$

2002 年以来,我国中央银行又重新回到了 1994—1996 年以外汇占款为注入基础货币主渠道的模式。

2. 商业银行资产负债表的变化

与此同时,我国中央政府面对亚洲其他国家遭受金融危机冲击的前车之鉴,也采取了一系列控制金融风险的措施,如成立四家资产管理公司,加大

对四大国有商业银行不良资产的处置力度；对商业银行基层分支机构贷款额度加以限制；撤销若干欠发达地区的分支机构等等。各家商业银行总行在控制分支机构贷款的背景下，加大了对国债、政策性金融债券、企业债券以及 AMC 发行的债券等债券资产的投资力度，这就使得（4）式变为

$$\begin{aligned} & \text{库存现金} + \text{存款准备金} + \text{各项贷款} + \text{各类债券} \\ & = \text{各项存款} + \text{再贷款} + \text{资本金}, \end{aligned} \quad (10)$$

将（9）和（10）式合并之后，得到整个银行体系的资产负债公式

$$\begin{aligned} & \text{外汇储备} + \text{对政府债权} + \text{对 AMC 的债权} + \text{各项贷款} \\ & = \text{流通中现金} + \text{各项存款} + \text{资本金}, \end{aligned} \quad (11)$$

其中，对政府债权包括中央银行购买的国债以及对政府的透支和商业银行持有的国债两部分；对 AMC 的债权包括中央银行对 AMC 的债权和商业银行对 AMC 的债权两部分。

（11）式移项之后，可以得到

$$\text{存差} = \text{外汇储备} + \text{对政府债权} + \text{对 AMC 的债权} - \text{流通中现金} - \text{资本金}.$$

这表明，影响存差的因素既有来自中央银行方面的，也有来自商业银行方面的。为了更清楚地揭示存差的影响因素，下面我们将从不同的金融机构口径来详细分析。

三、存差问题的分析视角之二：机构口径差异

在进一步讨论存差问题之前，我们首先必须清楚地界定“存差”的基本概念。通常，对于“存差”的解释，国内主流的观点是仅仅从金融机构存款余额（增加额）和贷款余额（增加额）相对比的角度来分析，认为存差是前者大于后者的差额（胡海鸥、赵慈拉，2002；张晓慧，2006）；也有经济学家认为存差是“扣除准备金后的金融机构存款余额大于贷款余额的差额”（何田，2004；李斌，2006）；还有经济学家将存差分为封闭经济条件下的存差和开放经济条件下的存差（张红地，2003），强调 1994 年之后我国国际收支顺差导致外汇占款增加是存差出现的重要经济背景。然而，不少经济学家在讨论存差问题时，并没有明确指出金融机构的口径问题。也就是说，在存差问题分析当中，研究者没有明确金融机构的口径仅仅是商业银行，还是包括中央银行在内的整个银行体系。张晓慧（2006）的研究和我国中央银行的《货币政策执行报告》很好地弥补了这一缺憾，明确指出包括中央银行在内的整个金融机构口径下的存差概念更有探讨的意义。

笔者以为，从机构口径对存差问题的分析大体可以包括三个方面。第一个方面是从不同机构组成的系统口径下分析存差的含义。因为即使探讨宏观

意义上的存差问题更有价值,我们也不能忽略对微观意义下存差影响因素的讨论。这可以分为两个层面:一是不包括中央银行,仅仅从商业银行的角度来分析的金融机构存差口径;二是包括中央银行在内的金融机构存差口径。不少学者出现逻辑上的不一致的关键在于其分析不能始终如一地从某一固定的口径出发来讨论。第二个方面是在金融机构信贷收支平衡表当中,对由于项目设置不完整或不明确导致的问题进行分析。由于银行概览与金融机构信贷收支平衡表机构口径相同,因此第三个方面是分析这两张统计报表下存差指标的差异。

(一) 不同金融机构组合下存差的含义

在讨论金融机构的存差问题时,为何金融机构的口径问题如此重要呢?下面我们以存款准备金项目为例来予以解释。如果以包括中央银行在内的金融机构口径来分析存差,就不应该考虑存款准备金项目,因为该项目在中央银行资产负债表和商业银行资产负债表的报表合并过程中被抵消了。如果仅仅以商业银行的口径来分析,存款准备金就是其资产方的项目之一,因此可以作为解释我国商业银行口径下存差规模发生变化的原因之一。其他的项目,如中央银行对商业银行的再贷款以及其发行的中央银行票据均可以做同样的理解。

我国中央银行公布的《2005 年第四季度中国货币政策执行报告》以及张晓慧(2006)都对我国的存差问题进行了讨论,同时张晓慧强调了存差口径的重要性,并认为我国金融机构出现存差主要有四个方面的原因:银行持有非金融机构发行的债券总量不断增长;近年来商业银行不良贷款剥离与核销力度加大;金融体系外汇资产增长较快和现金需求增长趋缓。除了最后一点外,笔者以为前面三点因素有进一步讨论的必要。

1. 银行持有非金融机构发行的债券因素

中国人民银行货币政策分析小组(2006)强调商业银行购买企业债和国债在其资产运用当中的比例在增加,这使得金融机构的存差规模增加。上述论断无疑是正确的。然而,我国商业银行不仅仅增加对非金融机构发行的债券——企业债和国债——的持有,在 1998—2002 年我国出现通货紧缩期间,我国商业银行就大量投资政策性银行发行的金融债券,2002 年以来,还大量持有中央银行发行的中央银行票据。以上项目余额的增加与金融机构存差之间的关系如何呢?

首先,从商业银行的角度来看,商业银行的资产除了贷款之外,近年来还增持了企业债、国债、金融债和中央银行票据,负债方仍然以存款为主,这势必造成商业银行贷存比的下降,商业银行视角下(不包括政策性银行)的存差必然增加。因此,不仅仅只有企业债和国债会导致商业银行存差扩大,而且金融债和中央银行票据也会导致同样的效果。

其次，从包括政策性银行和商业银行在内的金融系统分析，政策性银行发行的金融债（在其负债方）和商业银行持有的金融债（在其资产方）在合并报表时将会被抵消掉。因此，金融债券不构成该系统下存差扩大的因素。

再次，从包括中央银行、政策性银行和商业银行在内的金融系统分析，中央银行发行的中央银行票据（在其负债方）和商业银行持有的中央银行票据（在其资产方）在合并报表时将会被抵消掉。因此，中央银行票据不构成该系统下存差扩大的因素。

《2005 年第四季度中国货币政策执行报告》采用 2005 年和 1995 年我国商业银行有价证券及投资占资金运用比例的变化来说明我国金融机构（实际上是指整个银行体系）存差的变化原因。笔者以为：商业银行有价证券及投资占比的提高不仅有商业银行投资非金融机构发行债券的因素，也有商业银行投资金融债券和中央银行票据的因素，因此尽管后者不构成整个银行体系存差扩大的理由，但却是商业银行存差扩大的因素。《报告》分析的是包括中央银行在内的金融机构存差问题，在论证过程中却使用商业银行口径下的数据，其纰漏是显而易见的。

2. 商业银行不良贷款剥离与核销因素

亚洲金融危机之后，我国政府日益重视金融机构的不良贷款问题，并加大了对国有商业银行不良贷款剥离与核销的力度。所谓不良贷款的剥离，就是使国有商业银行资产方的不良贷款转变成等额的优质资产；所谓不良贷款的核销，就是国有商业银行用其所有者权益抵消其不良贷款带来的损失。

商业银行不良贷款的核销在其资产负债表上的变化如表 3 所示。

表 3 国有商业银行资产负债表

单位：亿元人民币

资 产		负 债	
不良贷款	- 1000	资本金	- 1000

注：不良贷款下降 1000 亿元仅仅是假设的例子。

对国有商业银行而言，不良贷款的核销仅仅使得资产方的不良贷款下降，而负债方的存款没有变化，因此这不仅会使得商业银行视角下的存差扩大，而且会使得整个银行体系下的存差扩大。

我国国有商业银行不良贷款的剥离与 1999 年前后成立的四家资产管理公司（AMC）关系密切。由于财政部只拨付每家资产管理公司 100 亿元的资本金，所以四家资产管理公司并无足够的资金来按账面价值收购国有商业银行巨额的不良贷款，其资金缺口主要是通过直接向对口的国有商业银行发行债券（假定为 8000 亿元）和向中央银行借款（假定为 6000 亿元）来弥补资金缺口的（见表 4）。

表 4 资产管理公司剥离四大行 14000 亿元不良贷款有关各方资产负债表的变化

单位 亿元人民币

中央银行				国有商业银行				资产管理公司			
资产		负债		资产		负债		资产		负债	
对 AMC 债权	6000	AMC 存款	6000	超额准备	6000			不良贷款	14000	向央行借款	6000
			- 6000	AMC 债券	8000					债券发行	8000
		超额准备	6000	不良贷款	- 14000						

注：本表忽略了对资产管理公司资本金账户的讨论。

资产管理公司剥离四大行 14000 亿元不良贷款的过程结束后，中央银行负债方的国有商业银行超额准备金将增加 6000 亿元，是因为资产管理公司用中央银行提供的资金向国有商业银行按名义价格购买了 6000 亿元的不良贷款，因此国有商业银行超额准备金会增加。对国有商业银行而言，其资产方不良贷款的剥离同样使得其贷款余额下降 14000 亿元，其他资产，如 AMC 发行的债券、超额准备金等，将会相应地分别增加 8000 和 6000 亿元。在存款不发生变化的情况下，这会使得商业银行的贷存比上升以及存差规模扩大。

如果将资产管理公司、国有商业银行视为一个系统，合并其资产负债表，此时国有商业银行不良贷款的剥离并没有使得该系统的不良贷款余额下降，因此该系统的存差数额并不会下降。然而，该系统资产方（国有商业银行）超额准备金将会增加 6000 亿元，负债方 AMC 向央行借款增加 6000 亿元。如果上述系统的资产负债表再和中央银行资产负债表合并，那么包括国有商业银行、中央银行和资产管理公司在内的整个系统的资产负债表各个项目互相抵消，最终各项目余额不发生任何变化（见表 5）。

表 5 AMC、国有商业银行和中央银行不同组合构成的系统的资产负债表

单位 亿元人民币

AMC 和国有商业银行				AMC、国有商业银行和中央银行				国有商业银行和中央银行			
资产		负债		资产		负债		资产		负债	
超额准备	6000	AMC 向央行借款	6000	超额准备	6000	超额准备	6000	对 AMC 债权	6000		
				AMC 债券	8000	AMC 存款	6000	AMC 债券	8000		
				对 AMC 债权	6000		- 6000	不良贷款	- 14000		
				不良贷款	- 14000	向央行借款	6000				
				不良贷款	14000	债券发行	8000				

《中国货币政策执行报告》之所以认为商业银行不良贷款剥离是我国银行系统存差扩大的原因，是由于资产管理公司不属于我国银行概览的机构统计范围，即不是广义货币供应量的统计机构对象，更不是金融机构信贷收支平衡表的机构统计范围。因此，当以上业务发生时，仅仅合并国有商业银行和中央银行的资产负债表，该系统资产方的不良贷款下降了 14000 亿元，而负债方的存款没有任何变化，因此从这一口径统计得到的结论就是金融系统的存差扩大了。

因此，在商业银行的视角下，不良贷款的剥离和核销都会使得资产方的贷款余额下降，负债方的存款不变，因此其存差将会扩大。在银行概览的视角下，也会出现同一现象。然而在银行概览和资产管理公司报表合并的视角下，只有不良贷款的核销才会使得存差扩大。

3. 金融体系外汇资产增长较快

我国货币当局认为金融体系外汇资产增长较快是存差扩大的原因之一。这其中既有我国中央银行外汇储备增加尤为迅猛的因素，又有商业银行外汇资产增加的因素，不过前者是金融体系外汇资产增加的主导因素。对中央银行而言，2002年以来我国外汇储备的快速增加使得其成为中央银行注入基础货币的主渠道，即资产方的外汇储备增加，负债方的超额准备金增加；对商业银行而言，其结果是资产方的超额准备金增加，负债方的存款余额同时增加。因此，从整个金融体系来看，中央银行负债方的超额准备金和商业银行资产方的超额准备金互相抵消，最终体现为外汇资产扩大的同时使得国内经济主体的存款也迅速上升，其结果必然是银行体系存差规模的扩大（见表6）。这表明中央银行注入基础货币的主要途径变化为外汇储备渠道，是我国近年来存差规模大幅度上升的重要因素。

表6 外汇储备增加在各金融机构资产负债表上的变化

单位：亿元人民币

中央银行		商业银行		银行概览	
资产	负债	资产	负债	资产	负债
外汇储备 1000	超额准备 1000	超额准备 1000	企业存款 1000	外汇储备 1000	企业存款 1000

综上所述，从我国广义货币供应量 M_2 所对应的金融系统（银行概览）来看，存差规模的扩大有两点原因：一是和货币当局注入金融系统流动性的方式密切相关，从过去的再贷款方式过渡到目前的外汇储备和向金融资产管理公司提供的融资方式；二是和商业银行资产的多元化有关，如购买非金融机构发行的债券。从商业银行本身来看，存差的扩大和商业银行资产方资产的多元化密切相关，如各类债券的购买、外汇资产的持有、贷款的剥离和核销、存款准备金因素等等。

国内不少经济学家将存差现象归结为中国政府对金融系统的控制（李斌，2006；张杰，2003），如李斌（2006）认为“外汇占款和中央银行利用外汇注资核销国有银行不良贷款可以解释绝大部分的存差”，这实际上是金融系统（包含中央银行在内）视角下的存差概念，但是却将存差定义为“扣除准备金后的金融机构存款余额大于贷款余额的差额”，这一定义则是基于商业银行的角度。张杰（2003）在其文章中也并没有明确国有银行的具体口径仅仅是国有商业银行还是包括中国人民银行以及国有商业银行在内的所有机构，同时，在分析当中却仅仅利用商业银行资产负债表来阐述其对存差的认知。

(二) 金融机构信贷收支平衡表项目设置的不足

如果我们将金融机构信贷收支平衡表的“金银占款”、“外汇占款”与“在国际金融机构资产”项目相加,再扣减“对国际金融机构负债”项目,归类为“对外净资产”项目。根据该表就可以得出以下公式

$$\text{存差} = \text{对外净资产} + \text{有价证券及投资} - \text{金融债券} - \text{流通中现金} - \text{其他}.$$

(12)

因此,我们可以根据(12)式大体得出存差的经济含义变化,这是否准确呢?下面我们来详细分析。

表 7 金融机构信贷收支平衡表与银行概览的对比

金融机构信贷收支平衡表		银行概览	
资金来源	资金运用	资产	负债
各项存款	各项贷款	国外净资产	货币和准货币
企业存款	短期贷款	国内信贷	货币
财政存款	工业贷款	对政府债权(净)	流通中现金
机关团体存款	商业贷款	对非金融部门债权	活期存款
储蓄存款	建筑业贷款	对其他金融部门债权	准货币
农业存款	农业贷款		定期存款
信托存款	其他短期贷款		储蓄存款
其他存款	中长期贷款		其他存款
金融债券	信托贷款		外币存款
流通中现金	其他类贷款		债券
对国际金融机构负债	有价证券及投资		央行债券
其他	金银占款		实收资本
	外汇占款		其他(净)
	在国际金融机构资产		

注:金融机构信贷收支平衡表的机构范围包括:中国人民银行、中国工商银行、中国农业银行、中国银行、中国建设银行、国家开发银行、中国进出口银行、中国农业发展银行、交通银行、中信实业银行、光大银行、华夏银行、广东发展银行、深圳发展银行、招商银行、浦东发展银行、福建兴业银行、民生银行、烟台住房银行、蚌埠住房银行、城市商业银行、城市信用社、农村信用社、财务公司、信托投资公司、租赁公司、邮政储汇局。银行概览,就是除保险公司、证券公司之外其他金融机构资产负债表的合并报表。转引自中国人民银行的《中国人民银行统计季报》。

1. 有关不良资产的处理并未明确反映在金融机构信贷收支平衡表当中

1999 年我国先后成立了四家金融资产管理公司,用于收购四大行的不良资产。在这个过程当中,中国人民银行对这四家金融资产管理公司提供了巨额的融资,这个内容在金融机构信贷收支平衡表中并没有用单独的科目反映出来,而是进入了该表的“其他”项目当中。同样,按照(12)式来分析存差的经济含义显然不足以全面揭示这一影响因素。当然,张晓慧(2006)的分析则跳出了这一窠臼。

2. 对中央汇金投资公司的债权也未明确反映在金融机构信贷收支平衡表

当中

2003年12月，中央汇金投资公司在北京成立。随后，中央宣布通过汇金公司向中国银行和中国建设银行注入450亿美元外汇储备作为两家银行的资本金。此后，汇金公司还注资中国进出口银行、中国银河金融控股公司、中国工商银行、申银万国、国泰君安等机构，而汇金公司主要是依靠中央银行提供资金来源，尽管汇金公司并不是金融机构信贷收支平衡表统计口径范围内的金融机构，但这一关系却应该反映在该表当中。事实上，这一内容并没有明确地采用一个科目反映在金融机构信贷收支平衡表中，而是笼统地反映在“其他”项目中。对中央汇金投资公司的债权是中央银行资产之一，根据前面的分析，这也应该是我国金融机构存差扩大的因素之一。

以上分析表明，在我国对存差问题的主流分析当中，采用的是金融机构信贷收支平衡表的口径，但是由于该表在项目设置方面过于简略，对于准确揭示存差的经济含义增添了人为障碍。因此，即使机构口径明确，但也必须对报表内项目进行适时调整，以反映经济现实的变化。

（三）银行概览下的存差分析

由于银行概览和金融机构信贷收支平衡表所涵盖的金融机构一致，因此李扬、彭兴韵（2006）将银行概览口径下存差和金融机构信贷收支平衡表口径下的存差进行对比分析。如果将银行概览中的“货币与准货币”剔除“流通中现金”之后视为各项存款，将“对非金融部门债权”视同为贷款，这一简化是否合适呢？笔者以为，这种对比分析必须要进行以下调整才更有意义。

第一，由于金融机构信贷收支平衡表的“各项存款”当中有“财政存款”一项，而这一项内容并不进入银行概览的“货币与准货币”项目，因此要扣减这一项目余额。

第二，将“对非金融部门债权”视同为贷款也存在问题。由于金融机构信贷收支平衡表的资金运用方是按照资产种类来分类的，银行概览是按照业务对象来分类的，所以“对非金融部门债权”不仅有对非金融部门（即企业）的贷款，还包括对企业债券的投资，因此对比分析必须采用该项目下的贷款类子账户的数据。

总之，尽管从机构口径的角度看，金融机构信贷收支平衡表与银行概览基本相同，但是由于具体项目设置的不同，只有在调整项目口径的差异之后，这种比较分析才更有意义。

综上所述，对于存差的分析，我们首先要确立存差概念所涵盖的机构口径范围，其次，在机构口径明确的基础上，报表所反映的项目也不能过于简化，才可能准确揭示存差变化的深层次原因。展望未来，人民币汇率的波动幅度将会更大，我国中央银行也不可能继续大规模地收购外汇储备，到那时中国人民银行必将寻求其他的注入基础货币的方式；与此同时，商业银行市

市场化改革的进一步推进也将会导致其资产和负债项目的多样化,在这种状态下,金融机构存差的意义还将有新的变化。

四、存差问题的分析视角之三:区域金融差异

中国人民银行货币政策分析小组(2006)以广州和北京为例,来证明存差出现背后的多种因素。笔者认为:从区域金融的视角分析各地区的存差问题,虽然不具有宏观的政策意义,但是存差规模与当地金融机构对本地经济的支持力度存在的某种相关性仍然使得该问题具有深入探讨的必要。

(一)区域金融视角下存差的影响因素

地区存差的实质就是某一地区金融机构信贷收支平衡表视角下的存差。不少经济欠发达地区,尤其是县域经济的经济主管部门认为存差规模的增加意味着当地金融机构对本地区经济支持力度的下降,其原因和 1996 年以来各家商业银行全面推行授权授信管理制度不无关系。这一制度的实施是商业银行采取一级法人制度、强化内部统一管理和风险管理的基础性工作。从资产和负债两方面来看,上述做法对其分支机构的影响体现在以下两点:

一是资金管理逐渐向全额集中管理过渡,这是实现在统一法人制度下的资金管理。该制度实施的初期,下级行向上级行缴存二级存款准备金,扩大上级行的资金来源,集中统一向大型项目上集中。伴随着中国建设银行、中国银行和中国工商银行的上市改制,各家银行的资金管理向全额集中管理过渡的趋势更为明显。其特征表现为:基层分支机构每吸收一笔存款,都上存省分行,并按照内部资金利率获取利息收入;每发生一笔资产业务都从省分行借入资金,并按照内部利率支付利息(刘振海,2006)。因此从商业银行分支机构的视角分析,不需要再寻求资金的平衡。

二是加强了资产的风险管理。各商业银行总行的措施体现在以下几方面:第一,强调一线直接贷款人的责任,如贷款第一责任人制度、贷款责任终身追究制度等等,将职工的个人收入与资产风险挂钩,使得各分支机构在贷款业务方面较以前更为谨慎,这一“惜贷”现象在 1998—2002 年我国通货紧缩时期表现得尤为突出;第二,在资金统一调配的条件下,总行增加了对债券资产的投资,如中央银行票据、金融债券、企业债等等,相对收缩贷款的发放,尤其是在经济欠发达地区的贷款投放;第三,信贷审批权上收,贷款项目准入门槛提高。这对于防范基层分行信用风险的发生无疑具有积极作用,然而对于欠发达地区,尤其是某些县域经济体而言,能够达到信贷标准的贷款项目并不多,从而通过贷款方式获得的资金也日渐下降。

（二）欠发达地区视角下的存差扩大

尽管实现资金的统一管理和加强资产风险管理代表了商业银行资金管理的发展方向，但是从分支机构，尤其是欠发达地区的金融机构来看，其贷款的功能在弱化，贷存比不断下降，并逐渐演变成上级行的储蓄机构。从区域金融的角度看，欠发达地区的商业银行分支机构一方面资金上存，另一方面贷款被严格控制，其出现存差的规模扩大就是情理之中的事情了。这逐渐打破了长期以来我国发达地区金融机构出现存差，欠发达地区金融机构出现贷差的信贷特征。

如果说地方经济主管部门在分析本地区金融机构的存差规模时，只顾及本地区存贷款的变化，而忽略从全局的视角分析，如外汇占款因素、全国信贷增加总额等因素，的确会得出诸如政府对经济支持力度下降等以偏概全的结论。然而，从纵向的角度看，商业银行总行对资产风险的关注，强化分支机构贷款发放的管理和约束，一定程度上都抑制了欠发达地区的贷款投放，加之通过企业上市发行股票或者发行企业债券融资等方式对欠发达地区来说近乎水中花、镜中月，外汇占款分布又极不平均，所以我们不能完全否认某些欠发达地区存差的扩大是当地金融机构以贷款方式对本地经济支持力度下降的信号之一。当然，笔者也不认为因此就可以得出若干政策建议，如加大欠发达地区金融机构的贷款发放以改变存差扩大的现实。本节分析仅仅试图说明为什么某些欠发达地区的经济主管部门会将本地金融机构的存差扩大与金融支持的力度挂钩，这背后与我国商业银行管理方式的改变、人民币汇率预期升值背景下的外汇储备大规模增加以及长期以来欠发达地区经济主管部门对本地区贷差的思维定势都不无关系。当然，针对具体地区的存差影响因素，我们还有详细分析的必要。

五、金融体系存差的扩大与其流动性的“泛滥”

伴随着我国金融机构存差规模的扩大，国内金融体系流动性也出现了“泛滥”，表现为货币市场利率长期保持在低水平、银行信贷超常规增长等等。这两者出现一定的同步性，使得不少学者认为我国金融机构存差的扩大导致了国内金融体系流动性的泛滥。笔者则认为：这两者不是因果关系或者反向的因果关系，它们只是均源于我国2002年以来外汇储备的高速增长。

2002年以来，人民币汇率处于巨大的升值压力当中，我国货币当局为避免人民币汇率升值，主动进入银行间外汇市场买进外汇，因此外汇储备大规模增加。即便是2005年7月21日我国货币当局实施人民币汇率改革，并且主动将人民币兑美元汇率升值2%，但是此后人民币汇率的波动范围仍被控制在很小的范围内。由于人民币汇率的升值幅度远低于国际社会的预期，人民币

汇率的升值压力巨大。为防止其升值过快,中国人民银行仍然被迫大量买入外汇,这就使得我国的外汇储备以极快的速度在增加。货币当局在大量收兑外汇储备的同时,商业银行获得的超额准备金也大量增加(即货币当局主动投放了相应的人民币),由此造成金融体系流动性的“泛滥”。尽管中国人民银行采用连续提高法定存款准备金率和大规模发行中央银行票据的方式来收回金融体系的流动性,但仍然不足以完全对冲因外汇储备造成的流动性过剩。

表 8 2001 年以来我国外汇储备的增加额

单位:亿美元

年度	2001	2002	2003	2004	2005	2006
增加额	466	742	1168	2067	2089	2473

注 2003 年的增加额不含向国有商业银行注资的 450 亿美元。

资料来源:国家外汇管理局。

从整个银行体系来看,外汇储备的增加表明中央银行注入基础货币的模式重复了 1994—1996 年人民币汇率并轨之后货币当局注入基础货币的模式,即(8)式所反映出来的事实。如前所述,现阶段中央银行注入基础货币渠道以外汇占款途径为主,不论是整个金融体系还是商业银行系统本身,所对应的存差规模出现大幅度增加就是大势所趋了。所以说,现阶段我国金融机构存差的扩大和金融体系的流动性泛滥存在内在的联系,其本质源于人民币汇率巨大升值压力下的外汇储备快速增加。

六、结 论

综上所述,存差这一术语源于我国信贷差额管理体制下的分层管理方法,它从过去的区域金融特征转变为目前的宏观金融特征,其规模不断扩大,并日益引起国内经济学家们和货币当局的关注。然而,对存差现象的解释却众说纷纭。在国内经济与金融形势日益复杂的背景下,存差的影响因素日益复杂,既有中央银行的因素,也有商业银行自身的因素;就其背后的经济含义来看,它既不是商业银行的资金是否紧张还是过剩的指标,也不代表中央银行对经济支持力度的下降。即使存差扩大反映了我国储蓄大于投资的事实,但是其涵盖的范围远不止这一点。正如有的经济学家建议的那样,存差概念应该淡出我国金融理论与实践(李扬、彭兴韵,2006;胡海鸥、赵慈拉,2002)。本文分别从信贷管理制度的历史沿革、金融机构口径差异和区域金融三个视角,对存差现象展开研究,旨在澄清国内学者对存差认识上的模糊之处。研究结果表明:

第一,从整个银行体系来看,存贷差的变化与中央银行注入基础货币的渠道以及商业银行资产的多样化密切相关。

第二,在某些欠发达地区金融机构投融资方式仍然以存贷款方式为主的

条件下,存差规模的扩大可以视为当地金融机构对其经济支持力度下降的信号,但这并不可以类推到全国。

第三,目前银行体系存差规模大幅度增加和我国金融机构流动性过剩均源于2002年以来人民币汇率巨大的升值压力导致的外汇储备超高速增长。

参 考 文 献

- [1] 郭浩,“如何认识银行的存贷差”,《中国金融》2003年第4期,第52—53页。
- [2] 郭浩,“我国存款货币银行的存贷差问题分析”,《国际金融研究》2002年第2期,第50—55页。
- [3] 何田,“银行存差:去向解析与问题探索”,《金融研究》2004年第4期,第136—142页。
- [4] 胡海鸥、赵慈拉,“‘存差’概念应该淡出我国金融理论与实践”,《上海经济研究》2002年第7期,第61—65页。
- [5] 江其务,“关于中国金融系统‘存差’的金融分析”,《财贸经济》2003年第4期,第22—26页。
- [6] 李斌,“存差、金融控制与铸币税——兼对我国‘M2/GDP过高之谜’的再解释”,《管理世界》2006年第3期,第29—39页。
- [7] 李宏瑾,“中央银行、信用货币创造与‘存差’”,《金融研究》2006年第10期,第8—22页。
- [8] 李扬、彭兴韵,“对存差的再认识”,《中国金融》2006年第24期,第34—36页。
- [9] 刘振海,“当前商业银行资金信贷管理的新变化”,《中国金融》2006年第14期,第57—58页。
- [10] 尚明、吴晓灵、罗兰波,《银行信用管理与货币供应》。北京:中国人民大学出版社,1992年版。
- [11] 孙工声,“关于银行存差现象的分析报告”,《中国金融》2005年第14期,第28—30页。
- [12] 孙国峰,“关于当前银行‘存差’问题的思考”,《财贸经济》2002年第10期,第33—35页。
- [13] 伍志文、鞠方、赵细英,“我国银行存差扩大成因的实证分析”,《财经研究》2004年第4期,第27—38页。
- [14] 殷孟波、贺向明,“存款激励与贷款约束——四万亿存差生成的制度分析”,《财经科学》2003年第4期,第12—16页。
- [15] 曾康霖,“怎样看待商业银行的存贷差额”,《金融研究》2000年第3期,第71—74页。
- [16] 张红地,“对于我国金融机构存差问题的认识”,《中国金融》2003年第13期,第31—32页。
- [17] 张杰,“国有银行的存差:逻辑与性质”,《金融研究》2003年第6期,第1—13页。
- [18] 张晓慧,“正确认识当前的‘存差’问题”,《中国金融》2006年第4期,第29—31页。
- [19] 张新泽、王毅,“‘存差’误解与贷款增长”,《经济研究》1999年第11期,第47—53页。
- [20] 中国人民银行货币政策分析小组,《2005年第四季度中国货币政策执行报告》,中国人民银行网站 www.pbc.gov.cn。

The Deposit-Loan Gap in China : Historical Evidence and Implications

YANG WANG

(Jiangxi University of Finance & Economics)

Abstract Since 1994 , the gap between loans and deposits in Chinese financial institutions has increased gradually to a point that catches the attention of economists and policy makers. This paper aims at offering a new and proper perspective on the gap. The author believes that differences among different financial institutions are crucial to the analysis of the deposit-loan gap. With a review of the evolution of the Chinese banking system and credit management system , we have discussed the changed implications of the deposit-loan gap. In the last section , we also link the gap to excessive liquidities in deposit institutions , and find that excess foreign exchange reserves are the underlying cause for the deposit-loan gap and excessive liquidities.

JEL Classification E52 , E58 , G21

我国制造业区域集聚程度 决定因素的研究

路江涌 陶志刚*

摘要 本文分析了中国近年来(1998—2003)制造业行业集聚的趋势以及影响行业集聚的微观基础。研究表明样本期间中国制造业行业集聚呈上升态势。我们利用面板数据和动态估计过程模型,一定程度上控制了现有实证研究文献中存在的一些可能造成结果偏差的缺陷。我们发现:地方保护主义在很大程度上限制了中国制造业的区域集聚。同时,溢出效应、运输成本和自然禀赋也是影响行业集聚的重要因素。

关键词 地方保护, 产业集聚, 微观经济基础

一、引言

行业集聚是经济学研究中的一个古老而又常新的话题。经济学家对行业集聚的分析可以追溯到亚当·斯密(1776)关于劳动分工的阐述。之后,理论经济学家对行业集聚微观经济基础进行了系统的研究。主要的发展轨迹包括 Marshall (1920) 将集聚产生的原因分为知识溢出效应(knowledge spillover)、劳动力市场发育(labor pooling)和投入共享(input sharing); Ohlin (1935) 则着重强调自然资源禀赋的重要性; Krugman (1991) 从集聚可能带来的规模经济角度来解释同行业集聚的现象。相对于行业集聚微观经济基础的理论研究而言,这个领域的实证研究发展较为滞后。近期的重要文献包括: Ellison and Glaeser (1999) 和 Kim (1999) 考察并发现了自然禀赋优势的重要作用; Audretsh and Feldman (1996) 分析了行业集聚对创新的影响,并发现集聚程度较高的行业的创新活动更多,从而支持了行业集聚的知识溢出效果; Holmes (1999) 发现了投入共享和行业集聚的正相关关系; Rosenthal and Strange (2001) 则综合了现有理论和实证研究的主要方面,对行业集聚的微观经济基础进行了更为全面的实证分析。¹

* 路江涌,清华大学经济管理学院企业战略与政策系,清华大学中国与世界经济研究中心;陶志刚,香港大学经济及工商管理学院。通讯作者及地址:路江涌,清华大学经济管理学院,100084;电话:(010) 62792726;E-mail: lujy3@sem.tsinghua.edu.cn。作者感谢清华大学中国与世界经济研究中心研究基金的支持,感谢匿名评审人的审稿意见。当然,文责自负。

¹ Rosenthal and Strange (2004) 以及 Duranton and Puga (2004) 分别对行业集聚微观基础的实证和理论研究进行了详细的文献综述。

近些年来,一些经济学家开始对行业集聚的其他微观基础方面进行分析。例如,Ades and Glaeser (1995)研究了政治稳定和集聚的关系;Holmes (1998)则按照是否有劳动法的标准把美国各州分成鼓励商业行为(pro-business)和限制商业行为(anti-business)两类,并发现一个明显的现象,即从限制商业行为的州进入鼓励商业行为的州的情况下,制造业的活动显著增加。关于中国的研究更多关注的是地方保护的程度和对行业集聚的作用。Young (2000)和 Poncet (2003)分别利用经济部门的数据和省份间投入产出关系的数据分析发现,中国各地区的地方保护程度在 20 世纪 90 年代有所增加。与此相反,另一些研究表明地方保护程度在改革开放后整体上逐步下降。相关研究包括:Fan and Wei (2006)通过对 1990 年初以来中国城市间物价趋同模式的研究发现,在这一时期,中国主要城市的物价越来越符合“一价定律”,即相同商品的价格趋同。这从一个侧面反映了地方保护程度的下降。国务院发展研究中心在 2003 和 2004 年对中国地方保护程度进行了两次调查。直接的证据表明,超过 70% 的受访者认为所在地的地方保护程度较 10 年或 20 年前明显下降(李善同等,2004;Tan and Zhang, 2005)。白重恩等(2004)是现有关于地方保护程度对行业集聚影响的最直接的研究。在这个研究中,行业内国有企业比例和行业利税率被用来代表地方政府保护该行业的积极性。研究发现那些国有企业比例较高的行业和平均利润率较高的行业发生行业集聚的可能性较低,这支持了地方政府保护某些行业的积极性降低行业集聚程度的假说。唐正清(2005)采用中国产业 2002 年横截面数据,发现基于企业创新的技术溢出并未起作用,相反阻碍了产业集聚趋势;同时政府对于产业资本的直接投资不能促进产业集聚,而通过市场机制发生的科研投入与产业集聚正相关。

以往对中国各行业集聚程度的研究所基于的数据大都是分类较粗的行业和地区层次的数据,而这些数据限制了我们对更细分的行业和地区间差异性的研究。Ellison and Glaeser (1997)对传统的、基于行业和地区层次数据的行业集聚指标进行了改进(后文简称 EG 系数),利用企业层次的数据,考虑了不同行业内结构的差异,从而增加了不同行业间的可比性。利用涵盖中国大部分工业产值的企业层次数据,我们首先分析了基于中国不同细分程度行业和区域组合计算的行业集聚程度从 1998 年至 2003 年的趋势,发现行业集聚程度在此期间稳健增长。接下来,我们着重分析了影响中国行业集聚程度的因素。具体来讲,我们在省、市和县三个区域层次用 3 位行业代码行业的特征来解释这些行业集聚程度。在行业特征方面,我们依据现有理论和实证文献,控制尽量多的因素(包括知识外溢、劳动力市场发育、投入共享、地方保护、自然禀赋优势、运输成本和规模经济等)。在计量模型方面,我们采用动态 GMM 模型来控制回归分析中可能存在的不可测度变量(unobserved varia-

bles) 和内生性问题 (endogeneity)。² 我们的研究表明：在中国的制造业中，地方保护因素和外部性在各个区域层次回归中显示稳健的正向作用；我们还发现自然禀赋优势和运输成本对行业集聚程度有增强作用；此外，结果表明，使用 EG 系数表示行业集聚时，规模经济的因素不再显著影响行业集聚程度。

本文以下部分结构如下：第二部分考察 1998 年至 2003 年间行业集聚的程度和趋势；第三部分分析影响行业集聚的因素。第四部分是全文的结论。

二、近年中国各制造业行业集聚程度和趋势

(一) 数据

本文使用的主要数据来自国家统计局工业企业统计数据库，涵盖 1998 年至 2003 年采掘业、制造业和电力、煤气和水生产和供应业的全部国有和年销售收入 500 万人民币以上的非国有企业。数据库包括分析行业集聚所需的区域代码、行业代码、产值等。我们在本文中集中分析制造业的行业集聚情况，而把采掘业以及电力、煤气和水生产和供应业排除在外，主要是因为这些行业的集聚程度在很大程度上依赖于自然资源的分布，对其进行分析的经济意义不大。在计算行业集聚程度时，我们主要依据产值数据。其原因主要是中国不同所有制企业存在不同程度的冗员现象，而各行业企业的所有制构成不同，可能造成分析结果不可比。为了使不同年份的行业集聚可比，我们使用了以不变价格表示的企业产值，而对于那些没有报告不变价格产值的企业，我们计算其所在的 4 位代码行业的价格指数，并对企业产值进行调整。

行业集聚研究中最重要数据项是企业的行业和区域。我们在使用原始数据中企业的行业和区域信息之前，对这些信息做了大量的整理工作。具体来讲，首先，我们将企业填报的区域代码和企业所在县区文字信息进行核对，由于这些文字信息基于企业填报的详细地址，因此在准确性方面应该优于企业的区域代码。其次，我们用 1999 行政区划代码体系 (GB/T 2260-1999) 作为基准，对样本年度内中国的行政区划代码进行了统一。³ 最后，由于中国在 2003 年实行了新的行业代码分类 (GB/T 4754-2002)，取代了旧的行业代码分类 (GB/T 4754-1994)。为了使 2003 年的数据和以前年度统一，我们利用新旧代码对比表和企业的产品信息将 2003 年度企业数据中的行业代码转换为旧

² Hanson (2001) 以及 Combes and Overman (2004) 提出关于行业集聚和区域专业化分工影响因素的实证研究中普遍存在三个主要的计量缺陷，即存在不可测度变量、内生性问题和缺失变量。根据我们的文献搜索，现有文献中对这些计量缺陷进行有效控制的研究只有白重恩等 (2004)。

³ 通过这两个步骤，我们对样本中近 100 万个企业/年度中的 14.9 万 (或者说 14.28%) 个观测值的区域代码做了调整。

的行业代码。

(二) 区域和行业的选择

由于数据可得性的限制，现有关于行业集聚的研究通常局限于较大的区域和较粗的行业分类，这在一定程度上忽略了更加细分的区域和行业之间的异质性。例如，小汽车制造业和自行车制造业同属交通工具制造业，但是，显然这两种行业的差异性很大，其区域集聚的状况、原因和影响都会有一定的差别。从区域角度而言，一个重要问题是区域的选择问题 (Modifiable Area Unit Problem, MAUP)。图 1 列示某行业在两种区域定义下的不同情况：如果我们定义第一种情况下的区域单位为县级，那么，在总计 16 个单元格中有 10 个单元格存在该行业，各为该行业总体的 10%；在第二种情况下，我们把区域单位合并为市级区域。那么，该行业存在于 3 个单元格中的 2 个，而且在其中一个单元格中达 90%。如果我们计算基于两种情况下区域划分的行业集聚，显然第二种情况下行业的区域集聚情况更加明显，而这两者间的差异是由人为划定区域造成的。

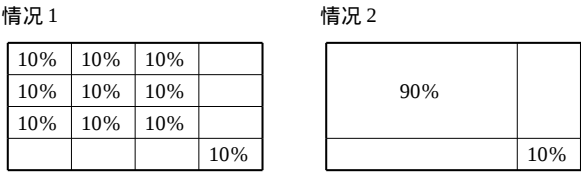


图 1 区域单位选择对行业集聚计算结果的影响

解决或减少区域选择问题的影响有两种可供选择的方法。其一是利用企业精确位置的信息，基于企业间准确的距离测算行业集聚，从而有效地去除了区域边界的影响 (Duranton and Overman, 2005)。但是，使用这种方法的一个主要的困难是企业精确位置的信息往往很难获得。另一种常用的方法是利用不同区域级别计算多组行业集聚指标，以观察基于不同区域级别结果的稳健性 (Rosenthal and Strange, 2001; Devereux, Griffith and Simpson, 2004)。在本文中，我们使用了后一种方法。

(三) 行业集聚系数的选择

现有关于中国行业集聚的研究大都是基于基尼 (Gini) 系数或 Hoover 系数，而此类系数是基于行业层次数据来计算的，从而不能够反映行业内部结构对行业集聚程度的影响。为了解决这个问题，Ellison and Glaeser (1997) 设计了 γ 系数，调整了不同行业内部企业几种程度的差异，从而使得计算结果在行业间、国家间和不同时间上更加可比。目前，国际上基于 γ 系数对行业集聚的研究很多，但是，对中国的相关研究还很少。⁴本文中，我们利用企业

层次产值数据计算了中国各制造业行业 1998—2003 年以 γ 系数表示的行业集聚情况，并着重解释行业间集聚情况差异的原因。

$$\gamma \text{ 系数的计算公式是 } \gamma_i = \frac{G_i - (1 - \sum_r x_r^2)H_i}{(1 - \sum_r x_r^2)(1 - H_i)}$$
 其中 G_i 是行业 i 在 r 个区

域内的基尼系数 $G_i = \sum_r (x_r - s_r)^2$ x_r 是区域 r 所有行业产值占全国所有行业产值的比例 s_r 是行业 i 在区域 r 的产值占该行业全国产值的比例 而 H_i 是行业 i 的赫芬达尔系数 $H_i = \sum_i z_i^2$ 其中 z_i 是企业 i 的产值占行业 i (包括 1 ... i 个企业) 产值的比例。

(四) 中国制造业集聚趋势 (1998—2003)

表 1 列示了基于 2 位、3 位和 4 位代码行业以及省、市和县级区域计算的 γ 系数的加权平均值，这里的权重是各行业的产值。表 1 表明各个行业和区域组合的 γ 系数在 1998—2003 年间呈稳健的上升趋势。此外，表 1 的最后两栏列示的 γ 系数最大和最小的行业表明了区域选择问题 (MAUP) 的存在。以烟草加工业为例，在县和市级区域层次计算的结果表明该行业是 γ 系数最小的行业，但是，在省级区域层次计算的结果则表明该行业是 γ 系数最大的行业。这说明了烟草加工业主要集中于几个省份，但是在这些省份内却广泛分布，造成了在市级和县级层次上的集聚程度较小。

表 1 中国制造业的 γ 系数(1998—2003)

行业和区域	1998	1999	2000	2001	2002	2003	2003				
	加权平均值						最小值	最大值	标准差	γ_i 系数最小的行业	γ_i 系数最大的行业
2 位代码行业											
县	0.0023	0.0025	0.0030	0.0034	0.0037	0.0042	0.0003	0.0097	0.0024	烟草加工业	电子及通信设备制造业
市	0.0061	0.0066	0.0070	0.0083	0.0089	0.0103	-0.0036	0.0289	0.0067	烟草加工业	电子及通信设备制造业
省	0.0206	0.0213	0.0220	0.0243	0.0265	0.0296	0.0024	0.0685	0.0187	木材加工及竹藤棕草制品业	烟草加工业

⁴ 国际上 Alecke et al. (2006), Devereux et al. (2004), Maurel and Sedillot (1999) 以及 Rosenthal and Strange (2001) 分别对德国、英国、法国和美国的行业集聚进行了 γ 系数的测算。国内的一个例外是路江涌、陶志刚 (2006) 利用企业层次职工人数数据用 γ 系数估算了中国 1998—2003 年的行业集聚情况。但是，一个值得注意的问题是，我国不同所有制企业中可能存在不同程度的冗员问题，从而造成所有制构成比例不同的行业之间可比性下降。

（续表）

行业 and 区域	1998	1999	2000	2001	2002	2003	2003				
	加权平均值						最小值	最大值	标准差	γ_i 系数 最小的行业	γ_i 系数 最大的行业
3 位代码行业											
县	0.0058	0.0057	0.0066	0.0072	0.0077	0.0086	- 0.0072	0.1351	0.0174	塑料家具制造 业	电气机械修理业
市	0.0127	0.0135	0.0146	0.0164	0.0174	0.0191	- 0.0312	0.2005	0.0296	其他 饮料制造 业	水产品加工业
省	0.0386	0.0391	0.0400	0.0426	0.0453	0.0491	- 0.1532	0.4119	0.0738	电子设备及通 信设备修理业	丝绢纺织业
4 位代码行业											
县	0.0087	0.0080	0.0092	0.0100	0.0109	0.0121	- 0.0319	0.3510	0.0368	潜水装备制造 业	锡冶炼业
市	0.0170	0.0180	0.0194	0.0213	0.0233	0.0248	- 0.1776	0.4920	0.0562	潜水装备制造 业	航标器材制造业
省	0.0485	0.0488	0.0500	0.0518	0.0577	0.0613	- 0.8997	1.1382	0.1207	潜水装备制造 业	摩托车修理业

三、制造业区域集聚程度的决定因素

（一）行业集聚的微观基础

我们主要从外部性、地方保护主义、运输成本、规模经济和自然资源禀赋等几个方面讨论行业集聚的微观基础。首先，我们从集聚所产生的外部性入手。Marshall（1920）将集聚产生的原因分为知识溢出、劳动力市场发育和投入共享。知识溢出是一个难以直接度量的现象，在本文中，我们使用行业新产品占总产出的比例度量可能的知识溢出。⁵和通常使用的研发投入占企业收入比例相比，新产品占总产出的比例和企业专利指标类似，可能更好地反映了创新能力的积累（Feldman et al.，2002）。除了知识溢出，溢出效应还可能来自信息共享。Lovely et al.（2005）研究了企业总部集聚和企业出口的关系，并发现：在行业出口信息比较难以获得的情况下，企业的总部呈现更为集中的趋势；而相对于出口比例较小的行业，那些出口比例较大的行业总部更为集中。本文中，我们使用行业出口占行业总产出的比例来度量可能来自出口方面信息共享的重要性。劳动力市场发育对生产力的促进作用在于某行业可能因为其需要的专门人才在一些地区的集聚更容易找到合适的人才。通常情况下，专业人才和需要这些人才的行业的相互作用会提高行业的集聚程度。由于比较细的行业层次上对专业人才

⁵ 新产品指采用新技术原理、新设计构思，生产的全新产品或在结构、材质、工艺等某一方面比老产品有明显改进，从而显著提高了生产性能或扩大了使用功能的产品。包括在全国范围内第一次研制、生产的国家级新产品和在省、自治区、直辖市、计划单列市范围内第一次研制、生产的地区级新产品。

需求的数据在中国很难找到,⁶我们利用工资数据计算了行业相对工资水平,用来衡量行业对专业人才的需求。在计算行业相对工资水平时,我们对地区差异进行了调整,即首先计算某行业在某地区的相对工资水平,然后将各地区该行业相对工资水平进行加权平均,得出全国该行业的相对工资水平。

具体计算公式如下：
$$\text{Wage_premia}_i = \frac{\sum_r \frac{\text{Wage}_{ir}}{\text{Wage}_r} \times \text{Emp}_{ir}}{\sum_r \text{Emp}_{ir}}$$
，其中 Wage_i 和

Wage_{ir} 是全国和某地区的一个 3 位代码行业职工的平均工资,而 Emp_{ir} 是该地区该行业职工人数。为了排除国有企业中可能存在的未在工资中反映的职工福利的影响,我们在计算中未包括国有企业。为了检验投入共享对集聚的影响,我们使用了行业中间投入中外购投入占企业总产出的比例作为测度指标。Holmes (1999) 用相同的指标研究发现,行业对外购投入依赖程度和该行业的集聚程度呈正相关关系。但是,由于只有面板数据, Holmes (1999) 没有能够考察两者间的因果关系。本文利用面板数据考察行业对外购投入依赖程度对行业集聚程度的影响。

近年来,地方保护主义在中国的学术界日益受到重视,这方面的研究逐渐涌现出来。周黎安 (2004) 分析了中国地方保护主义和重复建设问题长期存在的原因;徐现祥、李郇 (2004) 以长三角城市群为样本,发现 1990—2002 年间地方市场分割对区域协调发展的阻碍作用已经下降了近 50%;李善同等 (2004) 通过大范围的调查说明中国地方保护主义趋势逐渐下降的趋势;白重恩等 (2004) 以及胡向婷、张璐 (2005) 分别分析了地方保护主义对行业区域分布和地区产业结构的影响;黄玖立、李坤望 (2006) 则发现 80 年代由于新旧工业基地的轮替,地区专业化和分布的不平衡性稍有退化;而 90 年代地区专业化则显著加深,分布的不平衡性和集中程度明显加速。本文中,我们把地方保护引入计量模型,考察地方保护对行业集聚的影响。我们沿用与白重恩等 (2004) 类似的指标来衡量地方政府对行业保护的积极性,但与白重恩等 (2004) 不同的是,我们的指标更为细化,从而能更精确地衡量地方政府通过保护某些行业可能获得的利益。我们使用两个指标,其一是某行业地方税收占企业收入的比例,其二是某行业国有资本所拥有的产出占行业总产出的比例。地方税收是地方种税收和中央地方共享种税收中归地方所有部分之和。我们使用的数据中有四项指标和税收有关:应交增值税、应交企业所得税、产品销售税金及附加以及管理费用中的税金。我们依据中国税收制度中的相关规定计算各个行业对地方税收的贡献。第一,增值税的

⁶ Rosenthal and Strange (2001) 使用了具有本科、硕士和博士学位员工的比例来衡量行业对专业人才的需求。

25% 为地方收入；第二，某些行业、所有制形式和隶属关系企业的企业所得税属中央税种（如中央企业缴纳的所得税和外资企业所得税等）的，我们予以剔除；第三，产品销售税金及附加中的资源税比例较大，而我们没有单独的资源税的数据，因此，我们把成为资源税课税对象的行业从计量分析中予以剔除（烟、酒及酒精、化妆品、护肤护发品、贵重首饰、鞭炮、焰火、汽油、柴油、汽车轮胎、摩托车、小汽车行业）；第四，管理费用中核算的税金比例较小，且多为地方税种，对此不做进一步处理。由于地方政府从国有企业中获得的收益不仅仅局限于国有企业的税收，我们用行业内国有资本所拥有的产出占行业总产出的比例来衡量地方政府在某行业的其他收益。

运输成本通常被认为是影响行业集聚的重要因素，然而，没有关于行业运输成本的调查，在中国度量行业层次的运输成本存在很多困难。本文中，我们首先利用 Rosenthal and Strange (2001) 中使用的存货占销售比例的指标衡量行业产品的易腐败性。这个指标的合理性在于产品的易腐败性和产品单位距离的运输成本之间存在一定程度的正相关关系。与 Rosenthal and Strange (2001) 略有区别，我们使用完工产品占销售的比例，因为完工产品比存货更加符合要测度的经济关系。⁷ 然而，我们也意识到产品的易腐败性只是影响产品运输成本的一个侧面，因此，我们还使用产品销售成本（不含广告费）来度量产品到达消费者手中的成本。产品销售成本（不含广告费）包括销售人员的工资、差旅费、运输费用等。

实现规模经济是行业发生集聚的一个重要原因，但是，由于 γ 系数在构建时已经考虑了行业的结构特征（赫芬达尔系数），因此，一些相关文献在分析行业集聚原因时不再单独考虑规模经济的作用（Dumais et al., 1997；Rosenthal and Strange, 2001）。然而，由于赫芬达尔系数和规模经济并不是一个等价的概念，一些学者认为仍然应该控制规模经济因素（Alecke et al., 2006）。在本文中，我们用行业内企业平均产值作为行业规模的指标。

在现有的文献中，关于自然资源禀赋对行业集聚的影响存在着争论。有研究表明自然资源禀赋是影响行业集聚最重要的因素（Kim, 1999；Ellison and Glaeser, 1999）。也有学者认为由于自然禀赋的重要性依赖于运输成本的高低，而在 20 世纪以来，实际运输成本的下降幅度达到 90%（Glaeser and Kohlhase, 2004），因此，自然禀赋的重要性也可能有所下降。在实际操作中，行业层次自然禀赋的重要性比较难以衡量。一个可行的方法是依赖于投入产出表，但是，中国目前公开的比较详细的最新投入产出表是 1997 年的，这样，我们就难以在面板数据的分析中考察自然禀赋的作用。因此，我们只是在不控制行

⁷ 以食品制造业为例，完工产品占销售的比例与产品的易腐败性之间存在正相关关系。例如，在 2003 年，食品制造业中 6 个 3 位代码行业的完工产品占销售的比例分别为罐头食品制造业(0.145)、发酵制品业(0.058)、其他食品制造业(0.054)、调味品制造业(0.049)、糕点、糖果制造业(0.040)和乳制品制造业(0.034)。

业固定效应的回归分析中考察了自然禀赋的作用。我们从投入产出表中计算了 3 个衡量行业对资源依赖程度的指标：农产品占行业投入的比率、矿产品占行业投入的比率以及电、燃气和水占行业投入的比率。

在实证分析中，我们使用 3 位代码的行业进行分析，其原因是：首先，企业填报的行业代码在年度之间可能因为各种原因发生变化。除了企业经营范围变化方面的原因，也可能是由于企业的经营范围覆盖多个行业代码，而造成各年度对主要行业代码的确定不一致。我们使用 3 位代码的行业可以一定程度上避免这种因素的影响；其次，我们用来衡量企业对自然资源依赖程度所使用的投入产出表中对行业的分类最接近于 3 位代码行业；最后，我们测算行业对地方税收贡献时需要把一些是资源税课税对象的行业剔除，而这些行业大都是 3 位代码的行业。我们在表 2 中列示了因变量和自变量的定义和统计值，在表 3 中列示了各变量间的相关系数。

表 2 变量定义和统计值

变量	定义	样本量	均值	标准差	最小值	最大值
γ 系数 (3 位行业, 基于 3 位行业代码和县级区域) 的 γ 系数		1025	0.0084	0.0165	- 0.0192	0.2567
γ 系数 (3 位行业, 基于 3 位行业代码和市级区域) 的 γ 系数		1025	0.0170	0.0271	- 0.0917	0.3038
γ 系数 (3 位行业, 基于 3 位行业代码和省级区域) 的 γ 系数		1025	0.0449	0.0674	- 0.4498	0.4169
国有产值比例	行业内企业国有资本相应产值之和/行业总产值	1025	0.2308	0.2065	0	1
地方税收贡献率	行业地方税收/行业销售收入	1025	0.0281	0.0278	0.0021	0.3929
行业相对工资	$\frac{\sum_i \frac{Wage_{ir}}{Wage_r} \times Emp_{ir}}{\sum_i Emp_{ir}}$	1024	1.0518	0.2641	0.4500	3.4842
外购中间投入比例	行业外购中间投入/行业总产值	1025	0.7638	0.0461	0.3973	0.8963
完工产品比例	行业完工产品/行业总产值	1025	0.0897	0.0467	0.0091	0.4366
销售费用比例	行业销售费用(不含广告费)/行业总产值	1025	0.0401	0.0275	0.0004	0.2602
新产品比例	行业新产品/行业总产值	1025	0.0728	0.0897	0	0.6430
出口比例	行业出口产值/行业总销售产值	1025	0.2217	0.1933	0	0.8409
企业平均规模	行业总产值/行业内企业数	1025	68.22	136.50	5.27	1407.54
农产品投入的比率	某行业来自农产品行业投入系数之和	1025	0.0966	0.1749	0	0.8086
矿产品投入的比率	某行业来自矿产品行业投入系数之和	1025	0.0548	0.1258	0.0001	0.7180
电、燃气和水投入的比率	某行业来自电、燃气和水行业投入系数之和	1025	0.0295	0.0269	0.0020	0.2089

注 行业变量在 3 位代码行业基础上计算。

(二) 计量方法和实证结果

1. 计量方法

Hanson (2001) 对区域集聚实证分析文献的评述中指出: 现有研究中估计方法的缺陷很大程度上限制了人们对集聚经济效果相关研究的接受程度。而主要的缺陷是现有研究没能够较好地处理不可观测的特征、内生性问题和集聚效应的多重来源问题。无独有偶, Combes and Overman (2004) 也强调了类似问题的普遍性。造成这类问题的主要原因是现有研究使用的多是截面数据, 例如, Rosenthal and Strange (2001) 虽然对美国各制造业集聚的原因做了细致的实证分析, 但是, 他们同时指出: 由于使用的是截面数据, 他们的计量结果中除了自然资源和运输成本等相对外生的变量外, 其他变量只是描述了行业其他特性和集聚之间的一个均衡状态。

接下来我们分别对 Hanson (2001) 提出的不可观测的特征、内生性和集聚效应的多重来源等三个问题进行讨论。对于不可观测的特征, 通常的做法是基于面板数据使用固定效应模型, 这样可以把各时间点间不变的不可观测特征去除。我们首先给出一个基本的固定效应模型: $\gamma_{it} = \alpha\gamma_{i,t-1} + \beta X_{it} + \phi_i + \varphi_t + \varepsilon_{it}$, 其中 γ_{it} 和 $\gamma_{i,t-1}$ 分别是行业 i 在 t 年和 $t-1$ 年的集聚程度, X_{it} 是行业 i 在 t 年的一组特征变量, ϕ_i 是行业 i 的固定效应, φ_t 是 t 年的固定效应, 而 ε_{it} 是扰动项。在上式中, 由于行业的集聚是一个缓慢的调整过程, 所以我们使用了 Arellano and Bond (1991) 建立的动态估计过程。在这个估算过程中, 行业特殊系数在上述方程式的一阶差分中被消去, 得到的方程中使用因变量和事先给定变量的各阶滞后值, 用外生变量的差值作为工具变量, 从而在一定程度上减轻了内生性问题。在 X_{it} 中, 我们尽量控制现有理论中影响行业集聚的因素, 包括外部性、地方保护主义、运输成本和规模经济等各个方面。这些变量帮助我们控制集聚效应的多重来源问题。

由于我们只有 1 年的自然资源禀赋的数据, 我们把自然资源禀赋要素引入随机效应模型: $\gamma_{it} = \alpha\gamma_{i,t-1} + \beta X_{it} + \eta N_i + \varphi_t + \varepsilon_{it}$, 这里的 N_i 代表着农产品投入的比率、矿产品投入的比率和电、燃气和水投入的比率。

2. 实证结果

我们在表 4 中报告了基于动态估计过程模型的县级、市级和省级行政区域三种情况下的实证结果。结果显示国有产值比例和地方税收贡献率对行业集聚有负的影响, 说明国有产值比例高和对地方税收贡献率大的行业更倾向于分散而不是集聚, 这个结果和白重恩等 (2004) 的发现相一致。关于潜在外部性对集聚的促进作用, 我们的结果和理论预期也大体一致。具体表现为: 虽然只是在省级区域层次上显著, 但在各个区域层次上相对工资水平对行业

表 3 相关系数表

γ 系数 (3 位行业、县)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
国有产值比例	1												
地方税收贡献率	-0.186	1											
行业相对工资	-0.065	0.398	1										
外购中间投入比例	0.082	0.340	0.322	1									
完工产品比例	0.087	-0.187	-0.331	-0.256	1								
销售费用比例	-0.084	0.227	-0.036	-0.070	-0.060	1							
新产品比例	-0.048	-0.061	0.116	0.145	-0.277	0.070	1						
企业平均规模	-0.054	0.237	-0.036	0.305	0.023	0.057	0.135	1					
出口比例	-0.059	0.370	0.451	0.365	-0.114	-0.201	-0.140	0.183	1				
农产品投入的比率	0.186	-0.480	-0.275	-0.214	0.148	-0.133	-0.216	-0.054	-0.082	1			
矿产品投入的比率	0.025	0.079	0.135	-0.057	-0.063	0.103	0.153	-0.214	-0.022	-0.169	1		
电、燃气和水投入的比率	0.053	0.223	0.035	0.061	0.178	-0.096	-0.120	-0.167	0.251	-0.245	-0.189	1	
	-0.122	0.262	-0.011	-0.063	0.080	-0.012	-0.060	-0.147	0.045	-0.270	-0.246	0.490	1

注：黑体表示的相关系数在 5% 的水平上显著 (2-tailed)。

集聚均有正向影响，从而支持了劳动力市场发育会提高行业集聚程度的假说。关于投入要素共享的假说，我们发现外购中间投入比例的系数为正，并且在县级和市级区域层次上分别在 15% 和 5% 的统计水平上显著。这在一定程度上证实了 Holmes（1999）的发现，并且进一步说明存在投入要素共享到行业集聚的因果关系。回归结果表明关于技术溢出的新产品比例的系数虽然为正，但只是在省级区域层次上显著。而出口比例的系数则在所有区域层次上都显著为正，这和 Lovely 等（2005）所发现的有关出口信息共享的结果相一致。我们在分析中预期商品的易腐败程度会使得企业更倾向于接近消费者，从而降低企业集聚的程度。虽然在前文中我们在食品制造业中发现了与理论预期相一致的证据，但是表 4 的回归结果表明完工产品比例的系数在三组回归中符号不一致，并且都不显著。这也许说明完工产品比例并不能很好地代表运输成本和适用于所有行业。相反地，销售费用比例较好地符合了理论预期，其在三组回归中符号均为负，并且在统计上显著。⁸此外，表 4 中企业平均规模在统计上不显著，这说明在我们的研究中，由于使用了 γ 系数，规模经济的作用已经被比较充分地考虑。关于 Arellano and Bond（1991）动态估计过程结果的一致性和其中使用的工具变量的合理性，我们分别通过二次自相关的检验和 Sargan 检验来确定。二次自相关的检验表明不存在回归方程参差项间的二次自相关，而 Sargan 检验则表明回归中使用的工具变量是合适的（ p 值在 0.161 和 0.245 之间）。

表 4 基于动态估计过程模型的县级、市级和省级区域回归结果

自变量	县级	市级	省级
滞后一期的 γ 系数	- 0.2613 *** (0.0464)	- 0.2380 *** (0.0479)	0.0479 (0.0472)
国有产值比例	- 0.0375 *** (0.0114)	- 0.0642 *** (0.0147)	- 0.2157 *** (0.0221)
地方税收贡献率	- 0.4849 *** (0.1765)	- 0.5561 ** (0.2252)	- 1.5295 *** (0.3325)
行业相对工资	0.0095 ** (0.0038)	0.0140 *** (0.0049)	0.0078 (0.0074)
外购中间投入比例	0.0347 + (0.0230)	0.0578 ** (0.0293)	0.0126 (0.0420)
新产品比例	0.0134 (0.0173)	0.0120 (0.0222)	0.0602 * (0.0335)
出口比例	0.0393 ** (0.0182)	0.0550 ** (0.0236)	0.0682 * (0.0360)
完工产品比例	- 0.0033 (0.0278)	0.0308 (0.0357)	- 0.0276 (0.0537)

⁸ 完工产品比例和销售费用比例的相关系数为 0.07。

销售费用比例	- 0.1198 ⁺ (0.0802)	- 0.1855 [*] (0.1041)	- 0.2921 [*] (0.1590)
--------	-----------------------------------	-----------------------------------	-----------------------------------

(续表)

自变量	县级	市级	省级
企业平均规模	- 1.42E-05 (1.58E-05)	- 4.91E-06 (2.03E-05)	- 3.89E-06 (3.08E-05)
常数项	- 0.0003 (0.0005)	- 0.0003 (0.0006)	- 0.0046 *** (0.0010)
年份哑变量	Yes	Yes	Yes
样本量	637	637	637
Sargan Test (Pr > χ^2)	0.2613	0.512	0.1546
AR(1) (Pr > z)	0.2509	0.1911	0.0337
AR(2) (Pr > z)	0.1606	0.173	0.2453

注：***, **, * 和 + 分别表示在 1%, 5%, 10% 和 15% 水平上显著。括号内为标准差。

我们在表 5 中报告了考虑了自然禀赋优势的随机回归模型的结果。表 5 中关于溢出效应、地方保护主义、运输成本和规模经济的结果和基于动态估计过程模型的回归结果基本一致。关于自然禀赋优势,我们发现:农产品投入的比率和矿产品投入的比率显著地正向影响行业集聚,而电、燃气和水投入的比率的影响为负,并只在县级区域层次上显著。这里,如何解释电、燃气和水投入的比率的负向影响比较困难,可能的解释是电、燃气和水的比较容易实现区域间的运输,农产品和矿产品的运输成本较高。

表 5 基于随机效应模型的计量结果

自变量	县级	市级	省级
国有产值比例	- 0.0114 ** (0.0051)	- 0.0150 ** (0.0074)	- 0.0543 *** (0.0148)
地方税收贡献率	- 0.1648 * (0.0946)	- 0.2084 + (0.1323)	- 0.0647 (0.2541)
行业相对工资	0.0064 *** (0.0022)	0.0096 *** (0.0030)	0.0156 *** (0.0054)
外购中间投入比例	0.0519 ** (0.0147)	0.0968 *** (0.0196)	0.1433 *** (0.0361)
新产品比例	0.0051 (0.0090)	0.0111 (0.0128)	0.0704 *** (0.0247)
出口比例	0.0154 *** (0.0054)	0.0257 *** (0.0081)	0.0453 *** (0.0166)
完工产品比例	- 0.0037 (0.0164)	0.0379 * (0.0233)	- 0.0191 (0.0448)
销售费用比例	- 0.0060 (0.0306)	- 0.0154 (0.0445)	- 0.0598 (0.0883)
企业平均规模	- 1.14E-05 (9.02E-06)	- 8.98E-06 (1.31E-05)	0.0811 *** (0.0258)
农产品投入的比率	0.0084 (0.0063)	0.0308 *** (0.0107)	0.1156 *** (0.0432)
矿产品投入的比率	0.0313 *** (0.0103)	0.0459 *** (0.0176)	0.1020 (0.1941)

(续表)

自变量	县级	市级	省级
电、燃气和水投入的比率	- 0.0845 ** (0.0472)	- 0.0698 (0.0803)	- 0.0032 (0.0034)
常数项	- 0.0028 ** (0.0015)	- 0.0074 *** (0.0021)	- 0.0017 (0.0032)
年份哑变量	Yes	Yes	Yes
样本量	959	959	959
R ²	0.1201	0.1489	0.1083

注：***，**，* 和 + 分别表示在 1%，5%，10% 和 15% 水平上显著。括号内为标准差。

四、结 论

本文从实证的角度检验了中国近年来制造业行业集聚的趋势和影响集聚的微观基础。通过使用企业层次的数据，我们借鉴了国际上近期关于行业集聚指标的发展，首先发现与现有大多数研究相一致的结论，即中国近年来行业集聚程度呈上升趋势。在此基础上，我们基于相关理论和现有的实证研究，利用动态估计过程模型对可能影响中国行业集聚的要素进行了计量分析。由于使用了面板数据、动态估计方面和尽可能全面的变量，我们比较有效地控制了相关实证研究中普遍存在的不可观测的行业特征、内生性和集聚效应的多重来源等问题。从实证结果方面，分析证实了地方保护主义因素对行业集聚的影响。此外，我们的结果稳健地支持了知识外溢、信息外溢、劳动力市场发育和投入要素共享对行业集聚的影响。最后，我们发现对自然要素的依赖也在一定程度上促进了中国制造业的集聚。

参 考 文 献

[1] Ales, A., and E. Glaeser, "Trade and Circuses, Explaining Urban Giants", Quarterly Journal of Economics, 1995, 110(1), 195—227.

[2] Alecke, B., C. Alsleben, F. Scharr, and G. Untiedt, "Are There Really High-tech Clusters? The Geographic Concentration of German Manufacturing Industries and Its Determinants", Annals of Regional Science, 2006, 40(1), 19—42.

[3] Arellano, M., and S. Bond, "Some Tests of Specification for Panel Data, Monte Carlo Evidence and An Application to Employment Equations", Review of Economic Studies, 1991, 58(2), 277—297.

[4] Audretsch, D., and M. Feldman, "R&D Spillovers and the Geography of Innovation and Production", American Economic Review, 1996, 86(3), 630—640.

[5] 白重恩、杜颖娟、陶志刚、全月婷, "地方保护主义及产业地区集中度的决定因素和变动趋势", 《经济研究》2004 年第 4 期 第 29—40 页。

[6] Combes, P., and H. Overman, "The Spatial Distribution of Economic Activities in the European Union", in Henderson, J., and J. Thisse(eds.), Handbook of Regional and Urban Economics, Vol. 4. North-Holland: Amsterdam, 2004.

- [7] Devereux, M., R. Griffith, and H. Simpson, "The Geographic Distribution of Production Activity in the UK", *Regional Science and Urban Economics*, 2004, 34(5), 533—564.
- [8] Dumais, G., G. Ellison, and E. Glaeser, "Geographic Concentration as a Dynamic Process", NBER Working Paper No. 6270. 84, 1997.
- [9] Duranton, G., and H. Overman, "Testing for Localisation Using Micro-Geographic Data", *Review of Economic Studies*, 2005, 72(4), 1077—1106.
- [10] Duranton, G., and D. Puga, "Micro-foundations of Urban Agglomeration Economies", in Henderson, J., and J. Thisse(eds.), *Handbook of Regional and Urban Economics*, Vol. 4. North-Holland: Amsterdam, 2004.
- [11] Ellison, G., and E. Glaeser, "Geographic Concentration in U. S. Manufacturing Industries, A Dartboard Approach", *Journal of Political Economy*, 1997, 105(5), 889—927.
- [12] Ellison, G., and E. Glaeser, "The Geographic Concentration of Industry, Does Natural Advantage Explain Agglomeration?" *American Economic Review*, 1999, 89(2), 311—316.
- [13] Fan, C., and X. Wei, "The Law of One Price, Evidence from the Transitional Economy of China", *Review of Economics and Statistics*, 2006, 88(4), 682—697.
- [14] Feldman, M., I. Feller, J. Bercovitz, and R. Burton, "University-Technology Transfer and the System of Innovation", in Feldman, M., and N. Massard (eds.), *Institutions and Systems in the Geography of Innovation: Economics of Science, Technology, and Innovation*. Boston, MA: Kluwer Academic Publishers, 2002.
- [15] Glaeser, E., and J. Kohlhase, "Cities, Regions and the Decline of Transport Costs", *Papers in Regional Science*, 2004, 83(1), 197—228.
- [16] Hanson, G., "Scale Economies and the Geographic Concentration of Industry", *Journal of Economic Geography*, 2001, 1(3), 255—276.
- [17] Holmes, T., "The Effect of State Policies on the Location of Manufacturing: Evidence from State Borders", *Journal of Political Economy*, 1998, 106(4), 667—705.
- [18] Holmes, T., "Localization of Industry and Vertical Disintegration", *Review of Economics and Statistics*, 1999, 81(2), 314—325.
- [19] 胡向婷、张璐, "地方保护主义对地区产业结构的影响——理论与实证分析", 《经济研究》, 2005 年第 2 期, 第 102—112 页。
- [20] Kim, S., "Regions, Resources, and Economic Geography, Sources of U. S. Regional Comparative Advantage, 1880—1987", *Regional Science and Urban Economics*, 1999, 29(1), 1—32.
- [21] Krugman, P., *Geography and Trade*. Cambridge, Massachusetts: MIT Press, 1991.
- [22] 李善同、侯永志、刘云中、陈波, "中国国内地方保护问题的调查与分析", 《经济研究》, 2004 年第 11 期, 第 78—84 页。
- [23] Lovely, M., S. Rosenthal, and S. Sharma, "Information, Agglomeration, and the Headquarters of U. S. Exporters", *Regional Science and Urban Economics*, 2005, 35(2), 167—191.
- [24] 黄玖立、李坤望, "对外贸易、地方保护和中国的产业布局", 《经济学(季刊)》, 第 5 卷第 3 期, 第 731—760 页。
- [25] Marshall, A., *Principles of Economics*. New York: MacMillan, 1920.
- [26] Maurel, F., and B. Sedillot, "A Measure of the Geographic Concentration in French Manufacturing Industries", *Regional Science and Urban Economics*, 1999, 29(5), 575—604.
- [27] Ohlin, B., *Interregional and International Trade*. Cambridge: Harvard University Press, 1935.
- [28] Poncet, S., "Measuring Chinese Domestic and International Integration", *China Economic Review*, 2003, 14(1), 1—21.

- [29] Rosenthal, S., and W. Strange, "The Determinants of Agglomeration", *Journal of Urban Economics*, 2001, 50(2), 191—229.
- [30] Rosenthal, S., and W. Strange, "Evidence on the Nature and Sources of Agglomeration Economies", in Henderson, J., and J. Thisse(eds.), *Handbook of Regional and Urban Economics*, Vol. 4. North-Holland: Amsterdam, 2004.
- [31] Smith, A., *An Inquiry into the Nature and Causes of the Wealth of Nations*. London: W. Strahan, and T. Cadell, 1776.
- [32] Tan, K., and X. Zhang, "Incremental Reform and Distortions in the Product and Factor Markets in China", Working Paper, 2005.
- [33] 唐正清, "中国产业集聚实证研究", 北京大学 2005 年博士论文。
- [34] 徐现祥、李郁, "市场一体化与区域协调发展", 《经济研究》, 2005 年第 12 期, 第 57—67 页。
- [35] Young, A., "The Razor's Edge, Distortions and Incremental Reform in the People's Republic of China", *Quarterly Journal of Economics*, 2000, 115(4), 1091—1135.
- [36] 周黎安, "晋升博弈中政府官员的激励与合作——兼论我国地方保护主义和重复建设问题长期存在的原因", 《经济研究》, 2004 年第 6 期, 第 33—40 页。

Determinants of Industrial Agglomeration in China : Evidence from Panel Data

JIANGYONG LU
(Tsinghua University)

ZHIGANG TAO
(University of Hong Kong)

Abstract This paper investigates recent trends and micro-foundations of the spatial concentration of China's manufacturing industries using representative firm-level data. It is found that the overall industrial agglomeration in China has increased steadily in recent years. It is also found that, besides the usual determinants of Marshallian externalities, transaction costs, and natural advantages, local protectionism plays an important role in determining China's industrial agglomeration.

JEL Classification L11, R12, R30

中国城市劳动力市场的一体化进程

杨 涛 盛柳刚*

摘 要 本文从区域劳动经济学的视角分析城市间劳动力市场的发育和融合。我们介绍了分析区域劳动力市场的理论框架与国际经验,并利用住户调查数据发现,1988—1995年间,随着工资和教育回报率的上升,它们在城市间的差异有所扩大;虽然区域间的工资收敛并不显著,但受教育程度较低的工人工资收敛速度快于受教育程度高的技术型工人。针对这些实证结论,我们提出了促进区域劳动力市场统一的配套改革措施,以及建立全国性的以身份证为核心的大学生就业市场的建议。

关键词 区域劳动力市场,教育回报,工资收敛

一、引 言

中国经济改革开放二十多年来,劳动力的结构布局和行业分配发生了深刻的变化。国家统计局的数据显示,1978年,绝大部分劳动力滞留在国有企业和农村公社中;到90年代末,大约1/3的农村劳动力已转移到非农经济中,大约3/5的城镇劳动力在非国有企业里就业;目前,大约有一亿农民工进城务工。劳动力流动成为中国改革开放以后促进经济增长的主要因素之一。世界银行(1997a)研究发现,1978—1995年之间,劳动力从低生产率的农业流向高生产率的非农产业,对年均9.4%的GDP增长率的贡献占到了1.5%。蔡昉、王德文(1999)在此基础之上考虑了区域间劳动力的转移,发现改革以来GDP的高速增长中有21%来自于劳动力的流动。

中国的劳动力市场改革已经进行了二十多年,但是,区域间的劳动力市场是否出现了整合?区域间工资是否出现了收敛?已有的研究侧重城乡间劳动力的流动和收入差距,而忽视了对改革过程中城市间劳动力市场的静态特征和动态演变过程的研究。本文有两个研究目的:第一,介绍西方区域劳动力市场经济学的分析框架;第二,利用这一方法研究中国改革过程中城市劳

* 杨涛,美国弗吉尼亚理工大学经济系,清华大学中国与世界经济研究中心;盛柳刚,美国弗吉尼亚理工大学经济系。通讯作者及地址:Dennis Tao Yang, Department of Economics, Virginia Polytechnic Institute and State University, 3033 Pamplin Hall (0316), Blacksburg, VA 24061, USA, E-mail: dyang@vt.edu; 中国与世界经济研究中心,清华大学经济管理学院舜德楼南101, 100084。Wilson Sheng, Department of Economics, Virginia Polytechnic Institute and State University, 3033 Pamplin Hall (0316), Blacksburg, VA 24061, USA, E-mail: sheng@vt.edu。本文系作者于2006年夏季访问清华大学中国与世界经济研究中心期间的研究成果。作者感谢参加清华大学中国与世界经济研究中心和中国社会科学院人口与劳动经济研究所研讨会的人员所提出的有益建议,感谢两位匿名审稿人的宝贵意见。当然,作者对文中所有错误和遗漏负责。

动力市场的发育和融合过程。

区域劳动力市场 (local labor market) 是劳动经济学的一个研究领域, 它试图描述和解释不同地区劳动力市场的工资、就业、失业率、劳动力流动等主要指标的截面差异和动态演变路径。一个经典的模型是 Topel (1986) 提出的动态区域均衡模型, 其核心思想是, 如果允许劳动力自由流动, 个人将选择到未来收入折现值最高的地区就业, 那么区域间同质劳动力的工资差异就会由于劳动力自由流动而缩小。¹ 他发现美国各州工资率对州际劳动力市场的冲击很敏感, 而且不同流动性的劳动力群体工资弹性不同。Blanchard and Katz (1992) 通过分析美国各州的就业人口增长率、失业率的动态路径进一步发现, 在美国区域劳动力市场受到外部负面冲击之后, 劳动力流动是区域劳动力市场重新回到均衡的主要调整机制。Borjas (2001) 发现, 最具有流动性的墨西哥移民加速了美国州际间工资收敛的速度。

区域劳动力市场经济学方法可以帮助我们分析中国改革进程中城市间劳动力市场的静态差异, 理解市场融合和发育的动态过程和均衡机制。如果同类劳动者跨区域的工资不断收敛, 则说明劳动力市场的效率有所提高, 反之, 如果存在抑制劳动力自由流动的制度性障碍, 则会造成劳动力市场的分隔, 阻碍生产要素价格的一体化。目前从区域劳动力市场的视角分析中国劳动力市场的研究比较少, 本文旨在抛砖引玉, 指出一个新的研究方向, 并补足我们对改革进程中城乡移民、劳动力部门间流动的认识。

为了考察改革进程中区域间劳动力市场的融合程度, 我们至少需要两个时点的数据。因此我们采用了中国社科院和西方学者共同收集的 1988 和 1995 年的两次城市住户调查 (CHIP) 数据。数据来自 10 个省份多个城市, 1988 年含 67 个城市, 1995 年含 62 个城市, 其中有 29 个城市为两次调查覆盖, 因此组成了一个两年的城市面板数据。我们的重点是分析城市间工资差距和教育回报率的截面差异, 并分析在这一时期内城市间工资收敛和教育回报率的分布状况。

我们采用明瑟收入函数估计了样本中各个城市普通劳动力真实工资水平和教育回报。数据显示平均真实工资在这一期间上升了 40%, 平均教育回报也由 3.5% 上升到 5.4%。实证结果同时表明这两个指标的城市间差异几乎扩大了一倍, 这说明在中国并没有出现各城市间要素价格均等化。另外, 我们利用收入收敛模型, 分析了 29 个城市的面板数据, 得到与国际经验不同的两个结论: 第一, 发达国家区域间劳动力市场虽然收敛速度较慢, 但收敛是显著的, 而中国尽管劳动力市场改革成效显著, 已有相当规模的劳动力流动, 但是收敛并不显著。第二, 在发达国家和其他发展中国家, 一般具有受教育水平较高的劳动者流动性比较大, 市场融合度高, 而中国则不同。从收敛系

¹ 由于发达国家城市化水平比较高, 因此主流经济学中的“区域”主要是指城市之间。在 Topel (1986) 一文中, “local labor market”也有局部或本地劳动力市场的涵义。

数来看，各城市间受教育水平较低的非技术工人的收敛速度是受教育水平较高工人的两倍，这说明中国受教育水平高的劳动力流动性反而比较差。

我们认为，中国城市劳动力市场存在分割迹象主要来源于抑制劳动力流动的制度性障碍：区域封闭的户籍制度以及附着于户籍制度之上的本地居民优先的就业政策、社会保障和教育培训机会、难以转移的养老金、住房管理体制以及形形色色的地方保护。我们探讨了这些制度障碍的原因和未来改革的方向，并针对受教育程度较高的劳动力流动性差的现象，提出了一个政策建议——建立全国统一的大学生就业市场，尝试用身份证取代户口本，取消住房、养老金、教育培训机会等与户口的挂钩，帮助这部分新进入劳动力市场的就业人口摆脱诸多制度的约束，并逐步推广，以期最终取消户籍制度。

文章第二部分介绍了区域劳动力市场经济学的一般分析框架，以及对研究中国劳动力市场的启示和意义；第三部分描述了两次城市住户调查的数据，包括城市 and 个人的特征；第四部分估计了城市间的工资结构和教育回报，以及七年间工资的收敛状况；第五部分讨论了中国区域劳动力市场分割背后的制度原因，探讨了未来改革的方向；第六部分提出并讨论了建立全国统一的大学生就业市场的建议。

二、区域劳动力市场经济学的方法论

区域劳动力市场经济学 (local labor market economics) 既研究局部劳动力市场的工资确定、就业、失业率、劳动力流动等的截面差异，也分析这些变量在区域劳动力市场间的动态演变路径。作为劳动力经济学的一个领域，它强调局部劳动力市场的特征对整体劳动力就业和工资的区域分布的影响，而且力图解释外生冲击发生之后区域劳动力市场的调整机制。

国内比较熟悉的区域劳动力市场经济学的一个视角是区域之间的收入差距研究。Barro and Sala-i-Martin (1991, 1992) 开创了研究区域间人均收入收敛情况的先河。² 他们的研究在新古典经济增长基本模型中边际递减假设之上，将跨国人均收入收敛的研究方法应用于一个国家或者共同体内各个区域的人均收入收敛研究。他们发现，1880—1988 年美国各州和 1950 年以来的西欧 73 个地区间都存在着人均收入收敛的情况。虽然收敛的速度是缓慢的，年均为 2% 左右，但是统计上是显著的。同时，劳动力的自由流动对条件收敛速度的贡献率达到 1/3 左右。

许多学者按照这一思路检验了改革开放以来中国各省区之间的人均 GDP 和人均收入的收敛情况，Jian et al. (1996) 发现，80 年代人均收入明显收

² 他们研究的是人均收入，而不是工资本身。本文不特意区分工资与收入的差别。

敛,但从 1990 年开始,各地区的收入差距开始扩大。林毅夫等(1998)研究发现,1978 年以来人均 GDP 差异在东部、中部和西部地区内部减小,而东中西部地区间的差异却在扩大。蔡昉等(2001)发现劳动力市场的扭曲影响中国地区间经济增长的条件收敛。刘强(2001)的研究结果显示,大规模劳动力的区际迁移是中国地区间经济增长收敛的重要诱发因素。

收入收敛研究视角可以说明贫穷地区是否能够比富裕地区发展得更快,以及哪些因素会影响地区间的收敛。但是,这一视角无法解释局部劳动力市场在工资、就业、失业率和劳动力流动之间的巨大差异,也无法解释区域间劳动力市场特征对工资等变量在区域分布上的影响。Topel(1986)的动态区域均衡模型则为我们从微观上分析这些问题提供了一般性的分析框架。

动态区域均衡模型试图分析各地区劳动力市场特征及其演变对整体劳动力就业和工资的空间(地理上)分布的影响,这些特征包括各地区的失业率、工资与收入、劳动力净流入和经济增长等因素。其核心思想是,如果劳动力允许自由流动,个人将选择到未来收入贴现值最高的地区就业,那么区域间的同质劳动力间的工资差异就会由于同质劳动力自由流动而缩小。在这种情况下,我们就可以从两个角度来考察劳动力市场的特征对就业和工资的影响。第一,各地区劳动力市场的差异和演变将引起工资调整和劳动力的流动;第二,当一个外生冲击影响一个地区的劳动力需求时,该地区和其他地区的劳动力市场就会根据区域的特点进行调整,使之重新达到均衡,其中劳动力流动和工资的动态调整路径是我们感兴趣的课题。

Topel(1986)利用美国 1977—1979 年的现住人口调查(Current Population Surveys)数据发现,工资并不像人们想象的那样没有弹性,相反,工资率对各地区(州)的劳动力市场差异很敏感。一个对某地区的外生正冲击将增加该地区的劳动力需求,从而提高工资率。然而预期未来收入增加时,由于外地劳动力的流入,将会导致当前工资率的下降。这样就有必要区分冲击是长期的还是短期的。当外生冲击是短期时,由于预期未来收入不会上升,外地劳动力流入有限,因此该地区工资率的上升将高于永久外生冲击带来的工资率上升。另一个重要发现是,工资率的灵活度在流动性不同的劳动力人群中分布并不均匀,缺乏流动性的劳动力工资具有更高的弹性。

Blanchard and Katz(1992)通过分析美国各州的就业人口增长率、失业率的动态路径,进一步研究了美国区域劳动力市场(州)在外部负面冲击之后的调整机制。他们发现,外生冲击对就业增长率有着持久的影响,而对失业率则只有短期的影响。而且,当一个负面冲击发生时,劳动力流动是区域劳动力市场重新回到均衡的主要调整机制,而促使劳动力迁移的主要原因是失业率的上升,而不是当地工资的下降。他们的发现也支持了 Barro and Sala-i-Martin(1991, 1992)的结论。

Borjas(2001)研究了最具有流动性的移民群体对美国州际间工资差距的

影响。他指出移民作为自由选择的群体，进入美国后再迁移的成本低、流动性大，比美国本土人对区域间的工资差距更为敏感，因此他们的流动对于缩小区域间的工资差距具有显著的作用。Borjas 研究发现，移民加速了区域间工资收敛的速度，需要 17.3 年便可消除 50% 的区域间的工资差距，而且移民对美国本土经济的贡献也比以前的经验研究要高得多。

就目前中国的现状而言，区域劳动力市场经济学对于我们分析中国的区域劳动力市场至少具有以下几个方面的重要意义：

第一，中国各地区要素禀赋结构不同，改革步伐不一致，造成了各地区劳动力市场的工资和教育回报的巨大差距。但是劳动力自由流动在理论上可以缩小工资差距。如果没有出现收敛，我们则可以进一步分析哪些因素阻碍了劳动力的自由流动，比如户籍制度、住房的商品化程度、养老保险的可携带性以及地方保护等，并且由此我们可以提出一般性的政策建议。

第二，Topel 指出了工资弹性在不同流动性的劳动力人群中具有异质性，流动性强的群体的工资弹性比较低，对区域市场的冲击的敏感性不如流动性弱的群体，因此流动性强的群体的工资水平取决于地理上更广阔的市场。Borjas (2001) 提出流动性较高的移民对于促进地区间工资收敛具有显著的作用，这说明在这一群体内的工资差距相对较小，或者说收敛速度较快。因此，我们可以通过检验中国区域间流动性程度不同的群体的工资收敛状况，来验证劳动力迁移是否是区域间工资调整的机制。

第三，在美国等发达国家和其他发展中国家，具有较高学历和技术水平的人，由于获取信息的能力强，因此流动性比较大，而中国劳动力流动主要体现在受教育程度较低的农民工上。本文的研究发现，各地区间受教育水平较低的工人工资的收敛速度是受教育水平较高工人的两倍左右。收敛速度的差异间接反映了不同教育水平的群体的流动性程度，也就是说，不同教育水平的群体在目前我国劳动力市场中所受到的流动约束是不同的。这些实证结果对进一步改革中国劳动力市场具有重要的启示意义，它强调政策应该具有针对性。受此启发，我们根据受教育水平较高的工人区域间工资缺乏收敛的事实，提出了建立高校毕业生全国统一就业市场的对策。

三、数 据

本文采用的数据来自于中国社科院和西方学者合作的 1988 年和 1995 年的两次城市住户调查（见 Riskin et al., 2001）。两年间的样本和问卷设计具有可比性。1988 年调查包括 9009 家城市住户，涵盖了 10 个省份，1995 年增加了四川省，包括 6931 家城市住户。为了保证可比性，我们采用两次调查都覆盖的 10 个省份，放弃了四川省。在本文采用的样本中，1988 年包含 17074 个个体，1995 年包含 9627 个个体。这些个体是企业或单位的长期职工和长期

合同工；我们排除了退休职工、学生、临时工、私有企业主和个体户，他们的收入与具有稳定职位的职工缺乏可比性。

我们定义的收入包括所有类型的工资：基本工资、浮动工资、合同工资、加班费、奖金和兼职收入，以及津贴、辛劳补助和其他收入。虽然关于收入的纪录比较丰富，但并不理想。比如，问卷并没有包括医疗保险和养老金的价值，而且，企业提供住房并没有作为收入计算在内。在本文中，我们采用月度收入作为衡量收入的指标。

表 1 描述了我们所选样本中分省工人的月度收入分布状况。第（1）和第（5）列列出了按照 1988 年的不变价格计算并根据国家消费价格指数调整之后的真实收入。七年中平均收入从 1927 元上升到 2812 元，增加了 45.9%。表 1 显示各省之间的工资差异很大，而且这一差异随着时间的推移而扩大。1988 年收入最高的广东省人均月度收入比收入最低的河南省高 80%；到了 1995 年，收入最高的广东省和收入最低的甘肃省相比，差距扩大到了 144%。

表 1 分省样本分布和工人的真实收入

	1988 年				1995 年				1988—1995 年
	月度平 均收入 (1)	调整后的 月度收入 (2)	城市 个数 (3)	劳动者 人数 (4)	月度平 均收入 (5)	调整后的 月度收入 (6)	城市 个数 (7)	劳动者 人数 (8)	面板数据 城市个数 (9)
全样本	1927 (1034)	1827 (912)	67	17074	2812 (1712)	2574 (1412)	62	9627	29
北京	2052 (734)	2124 (760)	1	833	3797 (1643)	3383 (1464)	1	802	1
陕西	1667 (1220)	1555 (1138)	8	1830	2198 (1119)	2004 (1021)	8	1032	4
辽宁	1844 (618)	1899 (636)	4	1842	2487 (1232)	2468 (1223)	4	1185	3
江苏	1906 (625)	1887 (619)	8	2231	3026 (1472)	2810 (1367)	9	1199	5
安徽	1758 (1061)	1888 (1140)	8	1693	2195 (1098)	2293 (1148)	6	792	3
河南	1555 (709)	1492 (680)	8	1912	2120 (1061)	2154 (1079)	8	912	4
湖南	1767 (728)	1737 (716)	9	1887	2652 (1179)	2377 (1057)	7	1167	0
广东	2800 (1598)	2039 (1164)	8	1968	5089 (3040)	3844 (2296)	8	869	5
云南	2027 (848)	1931 (807)	9	1765	2563 (1082)	2418 (1021)	8	1079	4
甘肃	1907 (1088)	1906 (1088)	4	1113	2084 (927)	2037 (906)	3	590	0

注 第(1)列的平均收入以 1988 年的价格计算，第(5)列中 1995 年的真实收入是由国家消费价格指数调整产生，第(2)列和第(6)列汇报了根据分省价格指数调整后的按 1988 年不变价格计算的可比收入。括号中的值是标准差。

然而，使用社会消费品价格指数来计算真实工资存在一个问题。即使对同一商品，其价格在不同的省份也可能是不同的，其价格差反映了省际之间的交通成本和贸易障碍（Brandt and Holz, 2006）。遗憾的是，由于国家统计局公布的数据不全，我们无法得到不同年度不同城市商品的可比价格。为了对生活成本进行调整，我们采用了一组 Brandt and Holz（2006）设计的分地区的价格指数，这一指数是依据一揽子常用商品的价格计算，商品覆盖了 10 大类消费品和服务中的 64 种。表 1 中的第（2）和第（6）列陈述了根据这一价格指数调整后的按照 1988 年不变价格计算的可比平均收入。从整个样本来看，这一调整后的平均收入基本上低于第（1）列和第（5）列按照国家消费品价格指数调整的平均收入，而且前者的标准差也要小于后者。这一发现支持了如下结论，由于有不可贸易品的存在，不同地区的收入和价格往往成正相关（Kravis and Lipsey, 1988）。基于同样的理由，在调整后的数据中，最富和最穷省份的收入之比缩小了：1988 年为 42%，1995 年为 92%。以 1988 年可比价格计算，那两年国家正式职工的平均收入分别是 1747 元和 2413 元，与分地区价格指数调整后的数据接近，但与未调整的数据差距较大。因此，我们以下的分析均采用调整后的数据。

两次城市住户调查覆盖了不同地区多个城市，1988 年包括 67 个城市，1995 年包括 62 个城市。平均来看，在这两次调查中每一城市分别有 255 个和 155 个观测值。表 1 中的最后一列给出了两次调查均覆盖的 29 个城市在不同省份的分布。在此基础上，我们可以构造一个两年的、以城市为单位的面板数据。这些城市分布在中国各地，因此可以把它们看作具有不同制度环境的区域性劳动力市场。这一数据结构允许我们估计不同城市的工人收入和教育回报，从而观察城市间的差异和跨时间的变化。此外，我们利用这一面板数据可以检验这七年间地区劳动力市场的收入收敛情况。

表 2 描述了关键变量的统计特征。表的上半部分是影响收入的个人特征。七年之间劳动力的受教育程度增加了 1.1 年，而工作经验保持稳定。城市住户调查将受教育水平根据学历分成七类。为了估计教育的回报率，必须将这七类转换成相应的受教育年限。根据已有的研究和中国的教育体系，我们进行如下转换：大学及其以上学历定为 16 年，大专和职业技术学校为 14 年，高中 12 年，初中 9 年，小学 6 年，三年及其以上小学水平转换为 4 年，三年以下小学水平为 2 年。根据传统的做法，我们把年龄减去受教育年限和学龄前 6 年，得到相应的工作经验。从其他的个人特征来看，两年中少数民族比例分别占 3.7% 和 4.6%，党员的比例从 24% 上升到了 26.1%。

表 2 变量描述

变量名	1988 年		1995 年	
	均值	标准差	均值	标准差
Log(收入)	7. 4	0. 4	7. 7	0. 7
劳动者个人特征：				
受教育程度(年)	10. 7	2. 9	11. 8	2. 6
工作经验(年)	20. 5	11. 0	20. 9	10. 1
男性劳动者比率	52. 8	49. 9	53. 6	49. 9
党员比率	24. 0	42. 7	26. 1	43. 9
少数民族比率	3. 7	18. 9	4. 6	21. 0
城市经济环境：				
受过初中以上教育的职工比率	49. 3	50. 0	66. 2	47. 3
国营和地方公有企业职工比例	79. 4	40. 4	83. 3	37. 3
开放度(外资与合资企业职工比率)	0. 3	5. 4	0. 8	8. 9
通过市场渠道找到工作的比率	—		13. 0	33. 6
信息设施(家庭拥有电话的比率)	4. 9	21. 7	41. 8	49. 3

表 2 的下半部分列出了影响区域劳动力市场工资结构和教育回报的城市经济环境的指标。这些指标揭示了七年间中国区域劳动力市场的一些特征。我们利用有初中以上学历的工人占劳动力的比例来衡量技术工人的供给，这一比例从 1988 年的 49. 3% 上升到了 1995 年的 66. 2%。同时，尽管 1995 年在 外资和合资企业就职的工人比例仍然很低，但这一比例几乎是 1988 年的三倍。以安装电话的家庭比例作为衡量基础设施的指标，该比例从 4. 9% 上升到了 41. 8%。另一个表明劳动力市场开放程度的变量是工人是否通过市场化的渠道寻找工作。在 1995 年，有 13% 的职工是通过非公共机构靠自己找到了工作。遗憾的是，1988 年的调查没有收集这一显示区域劳动力市场发育程度的指标。

四、工资结构和教育回报

（一）计量模型

明瑟方程被广泛应用于估计人力资本和收入之间的关系。我们采用以下明瑟函数的一个扩展作为估计教育回报的基本模型。

$$\log Y_i = \beta_0 + \beta_1 S_i + \beta_2 E_i + \beta_3 E_i^2 + \sum_{j=1}^3 \gamma_j X_{ij} + \mu_i , \tag{1}$$

其中 Y_i 是工人 i 的月收入， S_i 是教育程度， E_i 是工作经验， E_i^2 是工作经验的平方项， μ_i 是扰动项， $E(\mu_i) = 0$ 。变量 X_{ij} 代表影响个人收入的其他三个指标：性别、党员和少数民族。

在方程(1)中, β_1 衡量了整个样本平均的教育回报。经典的人力资本理论认为教育有利于提高个人收入, 因此我们预期 $\beta_1 > 0$ 。常常与工作经验相联系的工作培训, 应该和收入有正相关的关系, 因此我们预期 $\beta_2 > 0$; 如果工作经验报酬边际递减, 那么 $\beta_3 < 0$ 。其他变量诸如性别、党员和民族可能对收入也有独立于教育的作用, 但如果这些变量与教育相关, 将它们包含在方程(1)当中则会系统性地改变对 β_1 的估计。

最小二乘法被广泛应用于估计人力资本回报的计量模型。但是由于最小二乘法忽略了个人能力变量, 而个人能力与教育程度有正相关的关系, 因此估计的教育回报率可能有偏。传统上有三种办法来处理这种遗漏变量产生的偏差: 工具变量、固定效应模型、直接衡量能力(参见 Card, 1999)。但是以上三种办法都对数据的要求非常高。

我们承认上述计量上的问题非常重要, 但是由于城镇住户调查没有足够系统的数据可以让我们矫正遗漏能力变量导致的偏差, 因此我们仍然采用最小二乘法。同时, 由于这两次调查所涉及的被调查者一般是在 1978 年改革前接受教育, 而改革前的教育体系非常僵化, 个人几乎没有根据自身比较优势选择受教育程度的自由, 因此选择性偏差的问题对这一群体可能并不严重。而且, 由于大多数对转型国家的教育回报率的研究并没有控制遗漏能力变量导致的偏差, 因此采用最小二乘法可以与这些已有的研究进行比较。

方程(1)暗含着两个假设, 即普通劳动力的收入, 也就是方程的截距项, 和教育回报率在不同城市间都是相同的。虽然方程(1)能够给出整个样本的平均教育回报, 但是接下来的数据分析显著地拒绝了上述两个假设。为了放松这两个约束, 使得普通劳动力的收入和教育回报在城际之间有差异, 我们采用更加灵活的模型:

$$\log Y_i = \sum_{k=1}^n \beta_{ek} D_{ik} + \sum_{k=1}^n \beta_{sk} S_{ik} D_{ik} + \beta_2 E_i + \beta_3 E_i^2 + \sum_{j=1}^3 \gamma_j X_{ij} + \mu_i, \quad (2)$$

这里 β_{ek} 是城市 k 中普通劳动力的对数化收入³, β_{sk} 是城市 k 的教育回报率, D_{ik} 是表示观察值所属城市的虚拟变量, 如果工人 i 在城市 k , 则 $D_{ik} = 1$, 反之 $D_{ik} = 0$ 。因此, 方程(2)可以为 1988 年和 1995 年分别估计 67 个和 62 个城市水平的参数: 各个城市的普通劳动力的收入和教育回报率。我们将使用这些估计得到的参数来看工资结构和教育回报率的城际间差异和动态变化特征。

(二) 计量结果

表 3 列出了明瑟收入方程的估计结果。第 (1)、(2)、(4) 和 (5) 列汇

³ 本文中, 普通劳动力的对数化收入是指剔除教育回报率和经验收益后的收入。因此, 这个收入可以理解为教育与经验为零的基本劳动收入。

报了方程 (1) 的估计结果：所有城市的平均教育回报率和其他个人特征对收入的影响。第 (3) 和第 (6) 列估算了不同城市普通劳动力的收入和教育回报率，但我们并没有在表中直接给出这 258(=67 × 2 + 62 × 2)个估计值；在之后的研究中，我们会仔细分析这些估计值。因为个体观察值分别来自不同群体，如家庭和不同所有制企业，因此可能存在异方差。我们所有的回归都使用了 Huber-White 标准差来去除异方差的影响。非常高的 F 值拒绝了所有参数为零的假设。

表 3 明显收入方程的估计

解释变量	1988 年			1995 年		
	(1)	(2)	(3)	(4)	(5)	(6)
教育	0. 0389 (0. 0010)	0. 0323 (0. 0010)	—	0. 0718 (0. 0029)	0. 0606 (0. 0031)	—
教育 × 城市	—	—	有	—	—	有
工作经验	0. 0418 (0. 0013)	0. 0414 (0. 0014)	0. 0446 (0. 0020)	0. 0753 (0. 0037)	0. 0745 (0. 0038)	0. 0700 (0. 0037)
工作经验的平方项	- 0. 0005 (0. 0000)	- 0. 0006 (0. 0000)	- 0. 0006 (0. 0000)	- 0. 0014 (0. 0001)	- 0. 0015 (0. 0001)	- 0. 0014 (0. 0001)
男性	—	0. 0969 (0. 0061)	0. 1109 (0. 0074)	—	0. 1569 (0. 0141)	0. 1647 (0. 0133)
党员	—	0. 0772 (0. 0068)	0. 0828 (0. 0082)	—	0. 1125 (0. 0160)	0. 1309 (0. 0153)
少数民族	—	0. 0105 (0. 0135)	0. 0431 (0. 0232)	—	- 0. 0472 (0. 0302)	- 0. 0193 (0. 0336)
城市的虚拟变量	—	—	有	—	—	有
常数项	6. 4296 (0. 0184)	6. 4591 (0. 0192)	—	6. 0116 (0. 0520)	6. 0801 (0. 0548)	—
R ²	0. 2690	0. 2886	0. 4072	0. 1553	0. 1743	0. 2991
F 值	1520. 78	942. 07	72. 73	391. 997	248. 52	32. 33
自由度	(317033)	(617030)	(13916897)	(39623)	(69620)	(1289498)

注 括号中的值是 Huber-White 标准差 ,它可以修正不同形式的异方差。

比较 1988 年和 1995 年的回归结果，我们发现了几个有意思的结论。第一，在这七年间，所有城市的平均教育回报率有较大幅度的上升，从 1988 年的 3. 2% 和 3. 9% 分别上升到 1995 年的 6. 1% 和 7. 2%。同时，收入的性别差距从 9. 7% 和 11. 1% 提高到了 15. 7% 和 16. 5%。党员与非党员的差距从 7. 7% 和 8. 3% 上升到了 11. 3% 和 13. 1%。⁴ 从整体来看，少数民族工人的收入与汉

⁴ 工作经验的回报率并没有呈现上涨趋势。依据表 2 中汇报的平均工作经验年限计算，经验的回报率的波动范围由 1988 年的 1. 68%—2. 13% 下降到 1995 年的 1. 15%—1. 68% (工作经验的平方项为负说明了工作经验报酬的边际递减性质)。这个结果与改革前强调工作经验和改革后强调工作能力的事实是一致的。

族工人的收入没有太大的差异，估计值比较小而且不显著。

以上关于1988年到1995年间教育回报率上升、收入的性别差距扩大和党员回报率结论和已有的研究基本一致。⁵工人个人特征的收益率差距扩大和中国收入不平等急剧上升的事实是紧密相连的。在所有拥有可比数据的国家中，改革开放以来中国收入不平等上升最快（世界银行，1997b；Yang，1999）。工资差距可以被分解为个人特征的不同和个人特征的回报率的差异。因此，我们的研究发现，个人特征回报率的变化是中国在转型中收入差距扩大的一个重要原因。

如果将性别、党员和民族三个变量加入到我们的基本模型，1988年和1995年的教育回报率就会下降。这说明受教育程度和这些个人特征是相关的。在这两年，教育和男性比例、党员比例有正相关性，1995年，受教育程度与少数民族比例是负相关的。值得注意的是，党员回报率上升的结论，拒绝了经济改革以来政治资本对收入的相对重要性应该下降的假说。我们的发现似乎支持另一种观点：党员的身份是一种有益于职业的投资，它向潜在的雇主发送信号来传递本来不可观测的个人品质，由此提高了个人的收入。

综上所述，从已有的证据来看，由于人力资本的回报率上升、收入性别差距扩大和党员与非党员间的收入差异扩大，中国城市的工资结构趋于发散，而且有更多的工资差异没有被这些解释变量解释，因为 R^2 从1988年到1995年显著地下降了。

虽然方程（1）描述了全国范围内城市住户收入的结构变化，但是它对教育回报率的估计是基于不同地区普通劳动力收入相等和教育回报率相等的假设之上。相关的F检验显著地拒绝了这两个假设：F值在1988年和1995年分别为1.81，6.31和2.00，4.81。因此，我们采用更为灵活的方程（2）来估计不同城市的普通劳动力收入以及教育的回报率。

表4给出了1988年和1995年估计得到的中国城市普通劳动力的对数化收入和教育回报率的分布情况。对比两年的结果可以看出，这些估计值随着时间的推移而趋于分散。在这七年间，普通劳动力平均的对数化真实收入从6.36下降到了6.18，因为检验两个均值相等的t值为3.21，其差距在统计上是显著的。按城市平均的教育回报率从3.5%上升到了5.4%；检验两个均值相等的t值为6.2，说明教育回报率有了显著上升。结合之前党员、性别差异扩大的结论，我们发现在这七年间，城市中平均主义的薪酬制度已经很大程度上被基于个人特征的薪酬体制所取代。普通劳动力的收入和教育回报率在城市间的差异扩大了：前者的范围由1988年的（5.64，6.98）扩大到了1995年的（4.73，7.14），而后的范围也由（0.01，0.07）扩大到

⁵ 如李实、丁赛（2003）Zhang et al.（2005）Gustafsson and Li（2000）Ho et al.（2002）。

(- 0.02 , 0.13)。⁶其他衡量不平等的指标也表现出收入和教育回报差异变大的趋势。在这期间，普通劳动力的对数化收入的标准差、变异系数和基尼系数都翻了一番以上。从这些指标来看，教育回报率的变化情况也大致相同。

表 4 城市普通劳动力工资与教育回报率的分布

	Log(普通劳动力工资)		教育回报率	
	1988 年	1995 年	1988 年	1995 年
均值	6.357	6.180	0.035	0.054
最低值	5.641	4.729	0.009	- 0.018
最高值	6.975	7.136	0.070	0.134
标准差	0.230	0.482	0.012	0.029
变异系数	0.036	0.078	0.343	0.542
Gini 系数	0.020	0.043	0.191	0.302
城市个数	67	62	67	62

图 1 显示了 1988 年和 1995 年城市间普通劳动力收入的分布状况。横轴是估计所得的普通劳动力的对数化收入，纵轴是观察得到的各个城市的平均对数化收入。1988 年的散点图相对集中在一起，这与当时政府保持各地区收入水平相当的事实是一致的。1995 年的散点图的均值比较高，表明在改革中工资已有显著提高；更重要的是，普通劳动力的对数化收入更加分散，产生了区域间的工资差异。进一步分析，散点图向上倾斜说明较高平均收入的城市给缺乏工作经验的非技术工人更高的报酬。

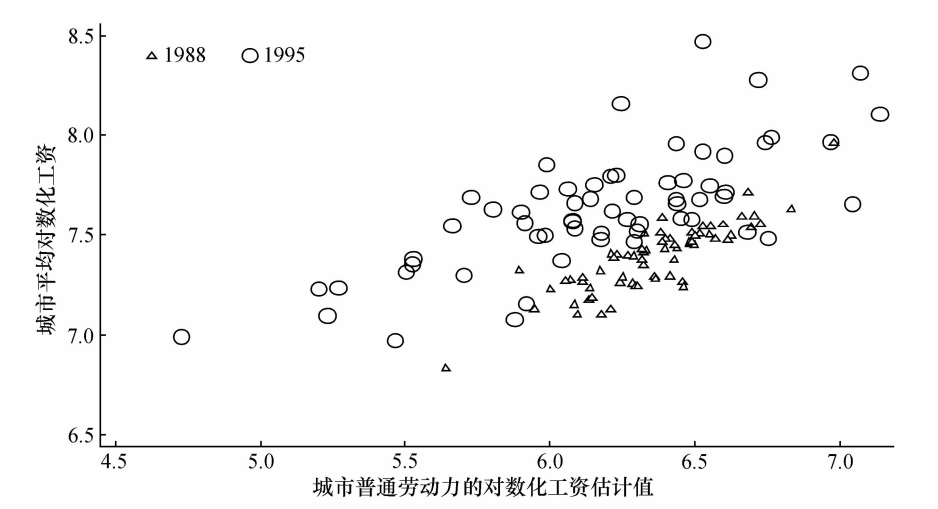


图 1 中国城市 1988 年与 1995 年估计所得普通劳动力的对数化工资

图 2 显示了 1988 年和 1995 年不同城市间教育回报率的分散程度。横轴是不同城市的估计的教育回报率，纵轴是观察得到的各个城市的平均对数化收

⁶ 在 1995 年，有两个城市的教育回报率为负值，但在统计上均不显著。

入。1988 年的散点图聚集在一个狭窄的范围内，而 1995 年的散点图几乎遍布图 2。转型期间这一教育回报率分布的动态变化是与逐渐放宽计划体制下刚性工资的举措是一致的。但是数据并没有显示城市的收入水平和教育回报率的关系。比如 1988 年，较高的教育回报率同时出现在收入水平高和收入水平低的省份。同样的，在 1995 年，两者并没有明显的关系。

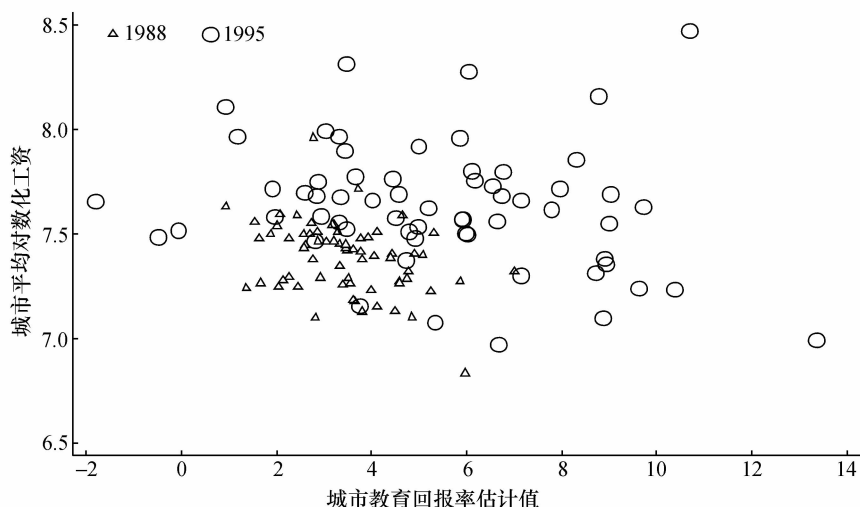


图2 中国城市 1988 年与 1995 年教育回报率的分布情况

从国际比较的角度来看，中国各城市间的教育回报率差异是相当惊人的。本文中各城市的教育回报率是使用一致的估计方法和可比的数据计算得到的。1988 年和 1995 年教育回报的变异系数分别为 0.343 和 0.542。Trostel et al. (2002) 使用微观数据估计了世界上 28 个国家的教育回报率。因为数据来源的背景不同，包括不同的市场环境和高度隔绝的各国劳动力市场，我们会预料跨国的教育回报率差异会比中国区域间的差异大很多。结果却不是这样，Trostel et al. 估算的跨国教育回报率的变异系数大约在 0.57 和 0.59 之间，与 1995 年中国城市间的变异系数大致相同。由此可见，在 90 年代中期，中国各城市间远远没有实现要素价格均等化。

（三）收入有否收敛？

在区域劳动经济学的框架内，如果存在市场竞争，并给定劳动力需求不变⁷，职工的自由流动是引发收入收敛、达到高效率劳动力市场的重要机制。区域间收入和教育回报的差异将吸引劳动力从收入较低的城市迁移到收入较高的城市，因此缩小城市间的收入差距。这一过程会改善劳动力的配置效率，

⁷ 技术进步、投资和产业结构变化等因素都可能影响劳动力需求，但是我们不在本文分析这些因素对局部劳动力市场的影响。

使整体经济获益。在美国,劳动力流动和区域间的市场竞争使得相对工资在各州之间稳定地收敛(参见 Barro and Sala-i-Martin, 1991, 1992; Blanchard and Katz, 1992)。这些学者的研究表明,虽然各州工资和收入的收敛速度比较慢,但是非常显著。

由于中国城市间收入差距巨大,如果城市间没有阻碍劳动力自由流动的政策和制度障碍,我们应该可以预料城市间收入收敛的速度很快。为了检验 1988—1995 年间收入是否收敛,我们采用如下的计量模型:

$$\log w_{k,1995} - \log w_{k,1988} = \alpha + \beta \log w_{k,1988} + \varepsilon_k, \quad (3)$$

这里 $w_{k,t}$ 是城市 k 在 t 年的收入, β 是收敛系数。 $\beta < 0$ 说明一个城市的工资增长率与这个城市的初始工资水平存在负相关,即贫穷城市的收入增长比富裕的城市快,城市间的收入趋于收敛。对美国各州之间年度收入收敛的研究发现,收敛系数 β 的基准是 -0.02 左右,也就是说年均 2%。因此,消除州际之间 50% 的工资差距大概需要 30—35 年。如果中国的区域劳动力市场开始融合,那么我们可以预料在这七年中收敛系数应该显著地小于零。

图 3 的纵轴画出了 1988—1995 年 29 个城市的对数化收入的年均增长速度,横轴是 1988 年的对数化收入。图中的直线是线性回归得到的拟合线,斜率为 -0.033, t 值为 -1.03, R^2 为 0.038。这些结果显示,尽管收敛系数为负符合我们的期望,但在 1988 年至 1995 年间中国的城市之间并没有出现显著的收入收敛趋势。虽然劳动力已经有了迁移的动机,但是导致均衡的机制——劳动力自由流动并没有有效的运作,从而消除巨大的城市间收入差距。这一发现与如下观点是一致的:直到 90 年代中期劳动力市场仍旧存在限制劳

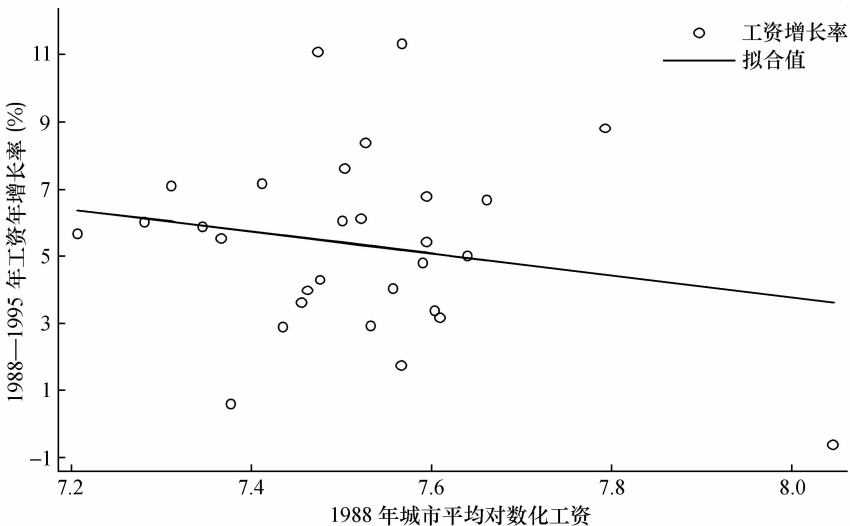


图 3 中国城市间的收入收敛状况

动力自由流动的法律和制度障碍，从而阻碍了劳动力在区域间的最优配置。Johnson（2003）使用统计局的调查数据发现，1990—2000 年间中国省际的劳动力迁移率只是同时期美国州际劳动力迁移率的 1/4。由于中国区域劳动力市场之间的显著非均衡状态，一个更具有比较意义的是美国农业劳动力大量向工业转移的时期；那时美国劳动力的迁移率是中国 20 世纪 90 年代的十倍左右。

继后我们考察了收入收敛速度在不同受教育程度的群体间是否存在差异。我们将劳动力分为四组：最高教育程度分别为小学毕业、初中毕业、高中毕业以及大专以上教育。图 4 给出了分组的 1988—1995 年 29 个城市的对数化收入的年均增长速度，横轴是 1988 年的对数化收入。我们发现与上述情况相同，收敛系数为负但是并不显著，因此并没有很强的证据支持不同区域劳动力市场间所有层次的技术工人收入趋同的现象。一个有意思的现象是：从收敛系数的绝对值来看，最高教育程度为小学毕业的工人收入的收敛速度（- 0.073）要显著高于其他三组教育层次的收敛速度（- 0.038，- 0.039 和 - 0.031），表明低教育程度工人的收入在城市间的收敛速度更快。虽然从统计上看，这些估计值并不显著，其 t 值分别为 1.41，1.03，1.39 和 0.84，但低教育组收敛系数显著性程度较高，已接近于传统的 10% 显著性水平。这些结果与 1988—1995 年间大量农民工进城的事实是一致的（参见 Fleisher and Yang，2006；Cai et al.，2004）。换言之，如果农民工在进城打工时系统地比

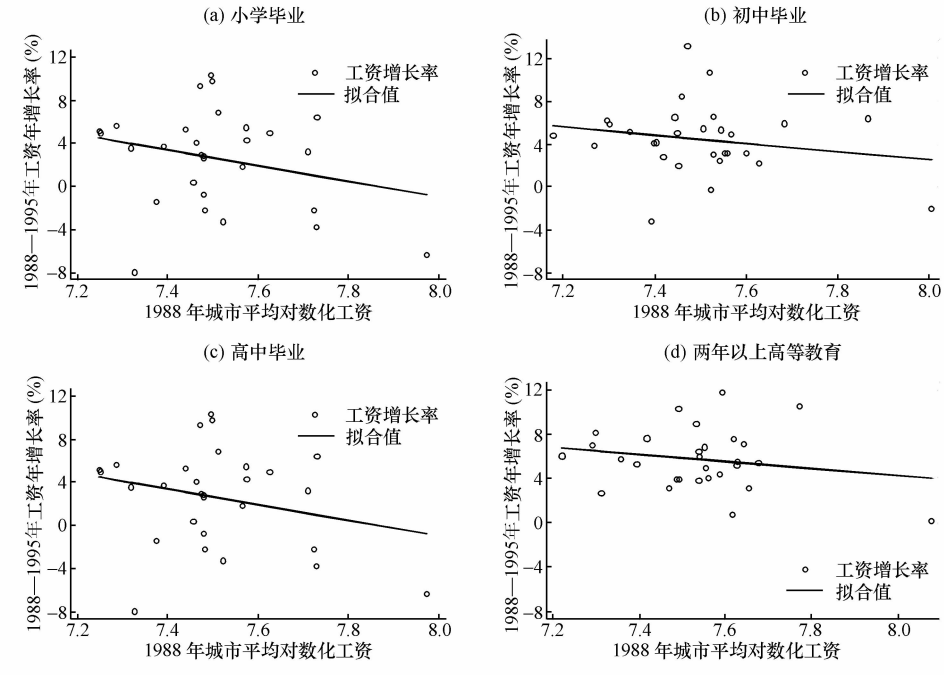


图 4 中国城市间按工人受教育程度分组的收入收敛状况

较目的地的收入差距,并做出相应选择,那么农民工进城就会促成低教育工人收入在城际间的收敛。在给定区域劳动力市场需求不变的前提下,以上实证结果也说明,技术工人在城市间迁移可能遭受到更多的制度障碍。技术工人一般在城市内有比较稳定的工作,因此迁移时需要考虑户口、住房、社会保障和子女受教育的问题;而农民工从农村出来,相对来说这些因素造成的阻碍比较小。

最后需要指出,收入收敛的参数不显著也可能是由于样本量小造成的。我们的分析旨在抛砖引玉,希望更多的学者关注、研究城市间的工资收敛状况。

五、劳动力市场分割的制度原因和相应的政策建议

我们的研究发现,在 1988—1995 年间,中国各城市间的工人收入和教育回报差异扩大了。从理论上讲,工资和教育回报的差异是促使劳动力迅速、大规模自由流动的内在动力,但是事实上这一差异并没有引发显著的工资收敛。同时,与其他国家相比,中国的教育回报在改革后虽然增长显著,但是仍然比较低。⁸我们的研究部分地解释了这一现象:直至 90 年代中期,中国区域劳动力市场之间仍旧存在市场分割,高教育的技术劳动者不能充分自由流动,这种市场扭曲使得人才没有配置到效率最高的地方,因此导致较低的教育回报,而个人则没有充分的动机投资教育。最近的一些调查和研究发现 (Fleisher and Yang, 2006; Cai et al., 2004), 在过去几年内中国劳动力市场的功能和效率并没有显著的改进。消除区域劳动力市场分割和教育回报的差异,需要进一步深化劳动力体制改革,我们简要分析这些制度障碍及其背后的动因,并提出相应的政策建议。

(一) 继续放宽对户口迁移的限制,就业政策应与户籍制度脱钩

户籍制度作为我国行政管理体制的重要组成部分,在计划经济时代是隔绝城乡之间、区域之间劳动力自由流动的主要障碍。改革以来的一系列措施已经有效地降低了户籍制度对劳动力流动性的负面作用。因此,客观地讲,户籍制度本身已经不再是限制劳动力流动的主要障碍,问题在于和户籍制度挂钩的一系列歧视性的就业、社会保障和教育培训等方面的政策。

首先是就业政策的歧视。城市就业政策仍然将本地城市人口的就业问题摆在优先地位。地方劳动局采用各种奖惩手段,鼓励用人单位雇用本地劳动力,限制和排斥外来劳动力。Cai (2003) 对北京的户籍制度做过案例研究,

⁸ Psacharopoulos (1994) 研究表明世界平均的教育回报率为 10.1%, 亚洲为 9.6%, 中低收入国家的教育回报率在 11.2% 与 11.7% 之间。

北京市对外地务工人员有严格的数量控制和工种限制，且随着城市就业形势的恶化而趋于加强。在整个90年代后期宏观经济通货紧缩的形势下，北京的劳动就业管理一方面加强下岗情况的统计工作，实施再就业工程，减少下岗；另一方面，更加严格地限制外地民工就业，明确要求企业优先雇用本地劳动力，同时将雇用外地民工与雇用下岗职工挂起钩来。这种歧视性的就业政策使得农民工主要集中于技术要求低、体力劳动为主的行业：如建筑业、制造业和批发零售餐饮服务行业。这与我们的研究结果是一致的，由于受教育水平低的劳动力流动的阻碍比较小，因此城市间非技术工人收入收敛速度较快。

其次是社会保障的歧视。目前城镇居民普遍享受着养老、医疗、失业、生育和工伤等五项法定保险，而没有户口的外地人口社会保障程度极低。社会保障体系基本上采用属地化管理，大部分社会保障措施都与户籍制度挂钩，因此，从社会保障的覆盖范围来看，上述五类社会保障基本上优先照顾拥有当地户口的城镇居民，没有户口的外来人口很难享受和参与到现有的社会保障体制中来。在城市中务工的农民工，基本上没有医疗保险和工伤保险，也没有养老金，他们只是趁着年轻在城市中出卖自己的劳动力。目前中央政府和地方各级政府已经开始努力扩大社会保障的覆盖面。

再次是教育培训机会的歧视。这一问题不仅表现在外地人口本身教育培训的机会较少，而且突出地表现在流动人口子女的受教育问题，子女的上学问题是流动人口面临的最大困难之一，流动人口子女的辍学、失学、超龄的问题都比较突出。流动人口一般读不起城市公立学校，他们一般选择城市打工子弟学校和家乡的学校。显然，城市义务教育体系排斥流动人口子女的现状，没有缓解这一制约因素，反而严重阻碍了劳动力的自由流动。

户籍制度改革一方面要继续放宽对户口迁移的限制，另一方面需要重点改革与户籍制度挂钩的就业政策、社会保障和教育培训机会。

第一，原则上来讲应该放宽落户的条件，按照稳定的住房、职业或者生活来源给予落户。第二，逐步放松对外来人口就业的职业限制，允许他们从事技术水平更高的职业。外地人口职业的分散化以及允许他们从事技术水平更高的职业，不仅可以吸引高素质的人才进入本地市场，而且可以提高外地人口的收入，将有利于解决社会保障和教育培训方面的歧视，有利于扩大社会保障的覆盖面，有利于让外地人口分享教育培训机会。第三，将民办学校合法化，推动各城市兴办民办学校，为外地人口子女在城市接受教育创造机会，改变他们父母与子女长期分离的局面。

（二）做实个人账户，提高养老金的可携带性

养老保险是影响劳动力自由流动的另一个重要因素。显然如果职工的养老保险无法在单位之间、区域之间顺利转移，那么劳动力流动的成本就会很高。养老金在地区之间转移困难，一个原因是各地区养老金管理条例的不一

致,比如目前全国各地单位缴费的标准和养老金的支付水平并不一样。但更重要的是,由于社会统筹账户入不敷出,个人账户的资金常常被挪用来填补养老金的空缺,剩下的缺口则由中央政府和地方政府按比例分担。在这种情况下,一个职工跨省就业,即使养老金账户实现了转移,资金并没有转移过来。由于账户已经迁出,迁出地政府不愿为其支付养老金,而迁入所在地的政府则因为需要负责其全部养老金而加重其财政负担,故也不愿意支付。因此目前的养老保险基本是各省区各自为政,在省际实现养老金的转移还是颇为困难的。

要提高养老金的可转移性,一个办法是继续提高统筹的层次,建立全国统筹的养老保险体系,在全国层面上实现现收现支,由中央政府负责缺口。但是由于养老金缺口庞大,同时随着老龄化问题日益严峻,让中央政府独自承担恐怕难以为继。如果让地方政府参与,由于各省财政实力差异悬殊和养老金缺口不同,省级统筹比全国统筹更容易发挥地方政府的积极性,因此恐怕难以实现全国统筹。目前来看,要提高养老金的可转移性,首先要做的和能做的则是做实个人账户,允许个人账户转移的同时有资金转移。2005年出台的《决定》强调做实个人账户,目的是解决个人空账运行的问题,但如何在做实个人账户的同时,提高养老金在区域间的转移性,则需要更多具体的政策,如全国统一的个人账户。其次,在做实个人账户并允许转移的基础上,再来考虑解决社会统筹账户的转移,可以根据各省不同的缴费水平和支付水平,将职工在迁出地的社会统筹账户的资金折算成迁入地的社会统筹账户资金。最后,要拓宽养老金的投资渠道,使得养老金有较高的回报率,这样才能吸引更多的人参加养老保险,从而缩小养老金的缺口,加快积累的速度。

(三) 进一步推进住房改革,将流动人口的住房需求纳入改革日程之中

住房改革是劳动力自由流动的先决条件。因为如果没有市场化的房地产出售和出租,职工无法轻易地与为他提供住房的工作单位解除劳动合同关系,因此跳槽将是非常困难的。目前住房改革解除了职工长期依赖企业提供住房的状况,而且大幅度地提高了人均居住面积,但是住房对于劳动力流动的障碍仍然存在。这主要体现在拥有户籍的城市居民和没有户籍的流动人口的住房的选择集合是有明显界限的。住房改革之后的城市居民在购买房屋时至少有三种选择:以市场价格购买拥有完整产权的商品房、以市场价格租房或者以较低的价格购买产权不完整经济适用房(Wang and Murie, 2000)。很多城市居民倾向于第三种选择,中央政府也积极鼓励地方政府兴建经济适用房,为中低收入的家庭提供住房。由于户籍制度的原因,流动人口基本上还是为主流的住房体制所排斥。吴维平、王汉生(2002)调查了京沪两地的流动人口的住房选择,发现流动人口有限的住房选择包

括：企业提供的宿舍、租房、工棚、自购商品房。比例最高的住房选择是租房和工棚。住房改革的歧视性并不仅仅体现在住房选择上，也体现在城市居民和流动人口的居住条件上。流动人口的住房狭小、拥挤不堪、缺水少厕。根据吴维平、王汉生（2002）的调查显示，京沪两地流动人口的人均住房面积是城市居民的三分之一。虽然住房改革允许外地人口购买商品房，但由于商品房的价格很高，超出了绝大部分流动人口的承受能力，流动人口很难从市场化的住房改革中获益。

大部分的流动人口都认为住房是他们在城市中常住的最大困难。因此住房事实上已成为劳动力流动的巨大障碍。然而现有的城市住房改革一直忽视了流动人口的住房需求，将他们排斥在住房体系改革之外。这一做法无疑是不利于劳动力在区域间的自由流动的。今后的城市流动人口管理法规应该着力于改善流动人口的生活状况。住房改革不应该将流动人口排斥在外，政府的安居工程应该将流动人口的住房需求考虑在内，允许流动人口申请廉租房，用于交换低廉价格的旧单元房的二级房屋市场应向流动人口开放；一些以前只针对当地居民的工作和教育培训机会应向流动人口敞开大门；尝试建立流动人口的住房公积金制度。

六、一个可行建议：建立高校毕业生 全国统一就业市场

本文第二部分关于区域劳动力市场研究的文献研究表明劳动力自由流动是促进区域劳动力工资和教育回报收敛的均衡机制，而 Borjas（2001）研究说明，一部分对区域间工资差异比较敏感的人群能够自由流动就能有效地促进区域间工资收敛。从我们的实证研究结果来看，受教育水平低的职工工资收敛速度较快和大量民工进城的事实是一致的，这也说明了流动中的民工作为均衡的动力，促进了区域间技术水平低的工人的工资收敛。我们的研究也发现受教育水平较高的职工工资收敛速度比较慢，因此如果存在一支受教育水平较高的均衡的力量，可能将有效地促进这一教育水平的职工的收入收敛并提高劳动力市场的效率。在中国目前的劳动力市场上，恰恰存在着这么一支队伍，他们受教育水平很高，没有太大的迁移成本，但却对市场的工资差异最为敏感，而且他们的规模正在逐步壮大，这支队伍就是普通高等学校的应届毕业生。拆除户籍的栅栏，使用身份证替代户口本，从而建立全国统一的大学生就业市场，不仅有利于提高大学生的就业率，有利于提高劳动力市场的匹配效率，而且这种以身份证为核心的新体系可以逐步推广到新增劳动力，并为渐进地在全国范围内取消户籍制度积累经验。

普通高等学校包括本科院校和高职（专科）院校，这些院校的毕业生是中国目前受教育水平最高的群体，同时他们一般都很年轻。美国劳动力经济

学研究的一个基本结论是职工的流动性与年龄呈反比关系,年龄越大,流动性越小。从成本来讲,高校毕业生在选择城市就业时,沉没成本相对已就业的年轻人要低,他们没有养老金转移的束缚,也没有住房公积金转移的障碍。从激励的角度来看,年轻人选择在哪个城市立足时,需要考虑的是一生的预期收入,他们未来时间长,对未来的生活更为看重,因此对区域间的工资差异是非常敏感的。从规模上来看,由于自 1999 年以来的高校扩招,这一群体的规模在迅速上升。教育部全国教育事业统计公报显示,普通高校毕业生从 2002 年的 133.73 万人迅速上升到 2004 年的 239.12 万人,2006 年为 413 万,可以预计这一增长趋势将保持若干年。

建立全国性的高校毕业生就业市场,目前主要的障碍在于户籍制度和地方保护。户籍制度的阻碍作用主要体现在两方面:第一,西部地区大学毕业生要进入东部地区就业很困难;第二,一旦离开读书所在地城市到其他城市就业,要想返回则非常困难,因此大部分学生会倾向于一次就业先选户口,二次择业,这造成了企业和员工的低效率匹配,也造成了人力资本的极大浪费。

为了解决大学生就业问题,学术界和政府已经认识到要放松户籍制度对应届大学生就业的束缚,逐步开始推动大学生跨地区就业。但是如果改革只是在应届大学生就业这一特殊时点放松户籍约束,一旦他们找到工作,就又进入了以户籍制度为核心的老系统,那么改革的作用将微乎其微,甚至比短期内不放松户籍都严重。因为大学生能够认识到如果不在应届毕业时落实中意城市的户口,将来会更加困难,也就是说购买户口的价格在应届毕业时是最低的,错过了这个机会户口的价格就会上升,因此大学生宁愿专业不对口,宁愿零工资就业,也要拿到中意城市的户口。这样一来,户口成为了应届毕业生就业时最重要的衡量因素,由此必然导致就业匹配效率的下降。

第二个障碍是地方保护。由于高校毕业生越来越多,毕业生就业问题成了各级地方政府的政治任务。各级地方政府为了优先保证本地高校毕业生的就业问题,要求各个机关和单位尽量优先照顾本地生源,比如,各级政府机关和国有单位的招聘一般都要求本地户口;同时给用人单位在引进人才上分配户口指标,严格控制外地人才的进入。有些地方甚至列出院校名单,不在名单之上的毕业生则很难在该地获得户口或引进指标。

下面从操作层面探讨建立全国统一的高校毕业生就业市场。

首先,使用身份证替代户口本。身份证与户口本的差异在于前者不带属地化管理的特点,要消除户籍制度对全国形成统一劳动力市场的阻碍作用,根本之道在于用身份证管理体制取代户籍管理体制。前面的讨论说明对新进入劳动力市场的应届大学毕业生,不能只是在短期内放松户籍制度,而必须以身份证取代户口本,将相关的住房和养老、医疗、教育等社

会保障体制与身份证挂钩，允许住房公积金、养老金在区域间的转移，这样使得大学生出城就业容易，回城就业也不难，从而提高以大学生为主体的受教育程度较高的劳动力在城市间的对流，促进劳动力市场的整合和提高工作匹配的效率。户籍制度改革牵一发而动全身，由于历史的原因，要解决存量问题很难，但是改革可以从流量着手。我们认为建立全国性的统一的大学生就业市场可以作为户籍制度改革的一个突破口。应届大学生作为新增劳动力的一个群体，我们的建议相当于将这部分群体作为试点，以身份证取代户口本，将相关的住房和养老、医疗、教育等社会保障体制与身份证挂钩。这一建议与目前各个城市引进人才做法的不同之处在于，人才引进之后就进入了户籍制度为核心的系统，而我们的建议则主张“新人新办法”，对新增就业的大学生尝试采用身份证为核心的系统，这一试点可以为对其他受教育水平高的劳动力放松户籍控制积累经验，未来也可以将试点推广到新增劳动力和全部人口。

其次，改变政府激励机制，消除地方政府实施政策歧视的动机。地方政府优先照顾本地生源主要是为了保证他们的就业率，但是这种做法却不利于本地的经济发展，尤其是会阻碍受教育水平较高的人才。所以中央政府必须转变衡量地方政府在促进高校毕业生的工作绩效的指标，不能单一地以本地生源就业率为指标，而应该以促进新增就业为目标。地方政府应该取消对企业的户口配额控制，让企业自主招聘员工。

有人会担心，一旦放开户籍制度对大学生的限制，大学生将纷纷涌向大城市，从而造成大城市高失业率。我们认为出现这种情况的可能性不大。一是雇主仍然会择优录用，不可能雇佣所有的人；二是因为当我们消除了对大城市户口的保护，就减少了户口的含金量，求职者将会有动力到外地去找工作。在20年前，渐渐放开城乡劳动力市场时，也有人担忧会造成城市失业，出现贫民窟，事实上并没有出现这个现象，因为在城市找不到工作，农民工就会回家发展。高素质的劳动力有更大的选择余地，他们不会放弃外地不错的工作，而到北京、上海等大城市流落街头。放开户籍制度之后，出城容易，进城也不难，就会打消毕业生到全国各地就业的隐忧，从而会提高劳动力市场的效率。

最后，建立全国性的以身份证为核心的高校毕业生就业市场刻不容缓。未来十五年是中国“人口红利”时期，之后中国将逐渐进入老龄化社会，因此在今后一段时间内，新增劳动力的规模可以说是“空前绝后”的，而大学应届毕业生将成为新增劳动力的一个主要组成部分，“十一五”期间预计有2500万以上的高校毕业生需要就业。错过了这一历史机遇，要通过“流量”解决“存量”问题的渐进改革方法将会非常困难。

参 考 文 献

- [1] Barro, R., and X. Sala-i-Martin, "Convergence across States and Regions", *Brookings Paper of Economic Activities*, 1991, (1), 107—158.
- [2] Barro, R., and X. Sala-i-Martin, "Convergence", *Journal of Political Economy*, 1992, 100(2), 223—251.
- [3] Blanchard, O., and L. Katz, "Regional Evolutions", *Brookings Paper of Economic Activities*, 1992, (1), 1—61.
- [4] Brandt, L., and C. Holz, "Spatial Price Differences in China: Estimates and Implications", *Economic Development and Cultural Change*, 2006, 55, 43—86.
- [5] Borjas, G., "Does Immigration Grease the Wheels of the Labor Market?" *Brookings Papers on Economic Activity*, 2001, (1), 69—119.
- [6] 蔡昉、王德文, "中国经济增长的可持续性与劳动贡献", 《经济研究》, 1999 年第 10 期, 第 62—68 页。
- [7] 蔡昉、王德文、都阳, "劳动力市场扭曲对区域差距的影响", 《中国社会科学》, 2001 年第 2 期, 第 30—37 页。
- [8] Cai, F., A. Park, and Y. Zhao, "The Chinese Labor Market", paper presented at a conference on China's Economic Transition: Origins, Mechanisms, and Consequences. Pittsburgh, Nov. 2004.
- [9] Cai, F., "Reform of labor policy in China: A Perspective of Political Economy", Working Paper Series No. 28, Institute of Population and Labor Economics, Chinese Academy of Social Sciences, 2003.
- [10] Card, D., "The Causal Effect of Education on Earnings", in Orley A., and C. David (eds.), *Handbook of Labor Economics*, Vol. 3A. New York: Elsevier Science, 1999.
- [11] Fleisher, B., and D. Yang, "Problems of China's Rural Labor Markets and Rural-Urban Migration", *The Chinese Economy*, 2006, 39(3), 6—25.
- [12] Gustafsson, B., and S. Li, "Economic Transformation and the Gender Earnings Gap in Urban China", *Journal of Population Economics*, 2000, 13(2), 305—329.
- [13] Ho, S., X. Dong, P. Bowles, and F. MacPhail, "Privatization and Enterprise Wage Structures during Transition: Evidence from Rural Industry in China", *Economics of Transition*, 2002, 10(3), 659—688.
- [14] Jian, T., J. Sachs, and A. Warner, "Trends in Regional Inequality in China", *China Economic Review*, 1996, 7(1), 1—21.
- [15] Johnson, G., "Provincial Migration in China in the 1990s", *China Economic Review*, 2003, 14(1), 22—31.
- [16] Kravis, I., and L. Robert, "National Price Levels and the Prices of Tradables and Non-tradables", *American Economic Review*, 1988, 78(2), 474—478.
- [17] 李实、丁赛, "中国城镇教育收益率的长期变动趋势", 《中国社会科学》, 2003 年第 6 期, 第 58—72 页。
- [18] Li, S., "Housing Tenure and Residential Mobility in Urban China: a Study of Commodity Housing Development in Beijing and Guangzhou", *Urban Affairs Review*, 2003, 38(4), 510—534.
- [19] 林毅夫、蔡昉、李周, "中国经济转型时期的地区差距分析", 《经济研究》, 1998 年第 6 期, 第 3—10 页。
- [20] 刘强, "中国经济增长的收敛性分析", 《经济研究》, 2001 年第 6 期, 第 70—77 页。
- [21] 吕绍青、张守礼, "流动儿童教育——逐渐进入视野的研究课题", 《战略与管理》, 2001 年第 4 期。

- [22] Psacharopoulos, G., "Returns to Investment in Education : a Global Update", *World Development*, 1994, 22(9), 1325—1343.
- [23] Riskin, C., R. Zhao, and S. Li, "Introduction to the Retreat from Equality : Highlights of the Findings", in : Riskin, C., R. Zhao, and S. Li (eds.), *China's Retreat from Equality : Income Distribution and Economic Transition*. New York : M. E. Sharpe, 2001.
- [24] Topel, Robert H., "Local Labor Markets", *Journal of Political Economy*, 1986, 94 (3, part 2) : S111—43.
- [25] Trostel, Philip, Walker, Ian, Woodlley, Paul, "Estimates of the Economic Return to Schooling for 28 Countries", *Labour Economics*, 2002, 9, 1—16.
- [26] Wang, Y., and A. Murie, "Social and Spatial Implications of Housing Reform in China", *International Journal of Urban and Regional Research*, 2000, 24(2), 397—417.
- [27] World Bank, *China 2020 : Sharing Rising Incomes*. Washington D. C : World Bank, 1997a.
- [28] World Bank, *China 2020 : Development Challenges in the New Century*. Washington, D. C. : World Bank, 1997b.
- [29] 吴维平、王汉生, "寄居大都市 京沪两地流动人口住房现状分析", 《社会学研究》2002年第3期, 第92—110页。
- [30] Yang, D., "Urban-based Policies and Rising Income Inequality in China", *American Economic Review Papers and Proceedings*, 1999, 89(2), 306—310.
- [31] Yang, D., "Determinants of Schooling Returns during Transition : Evidence from Chinese Cities", *Journal of Comparative Economics*, 2005, 33(2), 244—264.
- [32] Zhang, J., Y. Zhao, A. Park, and X. Song, "Economic Returns to Schooling in Urban China : 1988 to 2001", *Journal of Comparative Economics*, 2005, 33(4), 730—752.
- [33] Zhao, Y., and J. Xu, "China's Urban Pension System : Reforms and Problems", *Cato Journal*, 2002, 21(3), 395—414.

Wages and Returns to Education in Chinese Cities

DENNIS T. YANG

(Virginia Tech and Tsinghua University)

LIUGANG SHENG

(Virginia Tech)

Abstract This paper uses 1988 and 1995 data from a national household survey to document wages and returns to education for a large number of cities during China's economic transition. Between these two years, average real wages and returns to education increased, yet their spatial dispersions across cities widened dramatically. Although there was no sign of significant wage convergence in Chinese cities during this period, the wage convergence rate of non-skilled labor is higher than that of skilled labor. Based on these empirical results, we propose to set up an integrated na-

tional labor market for college graduates.

JEL Classification J31 , O15 , P21

中国农村医疗保障制度：一项基于 异质性个体决策行为的理论研究

封进 宋铮*

摘要 2003年初新型农村医疗保障制度在全国范围内试点。本文研究以下三个问题：这个自愿型的医疗保障体系的参与率有多高？保障体系自身是否可能实现收支平衡？人头税形式的缴费方式是否会使穷人受损而富人获利？为此，我们构建了一个异质性个体的消费—医疗支出决策模型，在拟合农村消费—医疗支出行为的基础上计算了农村医疗保障制度的福利效应。结果表明，在一定条件下，现行制度可以实现收支平衡，参与率可以在90%以上，健康状况较差的穷人是最大的受益者。

关键词 农村医疗保障制度，参与率，福利效应

一、引言

2003年初国务院提出《建立新型农村合作医疗的意见》，在全国范围内试点新型农村合作医疗制度，计划到2010年实现在全国建立基本覆盖农村居民的新型农村合作医疗制度的目标。现行制度的基本原则是自愿参加、多方筹资、以收定支。每年农户以家庭为单位按每人每年缴纳“合作医疗费”，同各级政府的补助一起形成合作医疗基金。与一般的公共医疗保障制度相比，这个保障制度的特点可以被归纳为以下三点：第一，自愿参加；第二，缴费偏低，政府补贴；第三，人头税型的缴费方式。¹

由此引出的疑问是这种自愿型医疗保障体系的参与率有多高，缴费偏低的体系自身是否可能实现收支平衡，人头税形式的缴费方式是否会使穷人受损而富人获利？前两个疑问涉及逆向选择和道德风险。就逆向选择而言，收入较低而健康较好的个体可能不愿意参加医疗保障体系，造成参与率和缴费总额的下降。就道德风险而言，由于保障体系降低了医疗支出的有效价格，体系内的个体可能增加医疗支出，造成缴费偏低的医疗保障体系无法实现收

* 复旦大学经济学院。通讯作者及地址：封进，上海邯郸路220号复旦大学经济学院200433；电话：13764006419；E-mail: jfeng@fudan.edu.cn。本研究得到国家自然科学基金的资助（项目编号：70573024），特此致谢。感谢复旦大学经济学院985项目的支持。感谢“中国女学者培训项目”的支持。感谢匿名审稿人的建议。感谢李玲、钟甫宁和 Philip H. Brown 的评论和建议。

¹ 根据国务院的规定，中国城镇职工医疗保障制度下的缴费率为本人工资的2%（用人单位缴纳按职工工资总额的6%缴纳）。根据规定每人每年缴费不低于10元，2003年实际缴费平均水平为14.6元（胡善联，2004）。而10元的农村医疗保险费支出仅相当于2002年农民人均纯收入的0.38%。

支平衡,或实现平衡的医疗保障制度只能提供较低的费用共付比例。²此外,逆向选择和道德风险也在相互影响。如果选择加入保障体系的个体的医疗价格弹性较大,就会加剧道德风险问题。而在道德风险的作用下,实现平衡的保障制度只能提供较低的费用共付比例,这又可能使得更多的个体退出保障体系,加剧逆向选择问题。第三个问题涉及平等和资源的再分配。由于患病人群和低收入人群不重合,一般说来富人会比穷人更多地利用卫生资源,因此人头税可能造成穷人补贴富人的情况(朱玲,2000)。

经验研究可以回答前两个疑问。不过,由于现行的中国农村医疗保障制度始于2003年,相关的数据还非常有限。而根据历史数据进行推测,使用一般的计量模型直接估计个体的决策行为可能遭遇类似于“卢卡斯批判”的问题,即当决策环境发生变化以后,个体的决策函数也可能发生相应的变化。因此,本文从考察中国农村个体的消费—医疗选择入手,根据历史数据估计个体的效用函数和决策环境,然后再预测个体在现行的保障制度下的决策行为。使用这一方法的另一个好处是我们可以计算现行保障制度对于个体的福利影响,并在此基础上探讨最优医疗保障制度问题。估计结果表明,中国农民对于健康的相对风险规避系数远高于消费。此外,从1990年到2000年,医疗相对价格上涨了5倍以上,而2000年至2002年间的医疗相对价格比较稳定。在较高的健康相对风险规避系数的作用下,医疗价格弹性不足,个体的实际医疗支出存在一定的刚性,实际医疗支出的下降幅度小于医疗价格的上升幅度,造成名义医疗支出与医疗价格正相关。

基于上述估计结果,我们模拟了现行医疗保障制度下中国农民的个体决策行为和福利变化,并对前面提出的三个疑问做出了回应。首先,这个保障体系的参与率可以达到92%。其次,只要把医疗支出的共付比例控制在50%左右,目前的保障制度是可以实现自我平衡的。最后,这个保障体系的主要受益者是收入较低而健康也较差的个体,它符合医疗保障制度向病人和穷人倾斜的基本要求。那么为什么现行的中国农村医疗保障制度没有出现这些似乎显然的问题呢?本文的分析表明,较少的缴费金额、较高的风险规避性、较低的医疗支出倾向和政府财政补贴是导致逆向选择问题较小的主要原因。实际医疗支出的刚性使得穷人的医疗支出倾向显著高于富人,而且由于穷人的收入偏低,遭遇健康负向冲击以后可能需要负债,这一财富效应会引起较大的福利损失。综合以上两点,穷人在保障体系中有较高的相对受益。我们的估算表明,穷人这部分较高的相对受益超过了因人头税型的缴费方式造成的相对损失。

本文的以下内容由五部分组成。第二部分简单回顾了中国农村医疗保障

² 本文中医疗保障制度的共付比例是指医疗支出中由合作医疗分担的比例。

制度的历史变迁，并总结了农村个体消费—医疗选择行为的基本特征。第三部分给出了一个简单的异质性个体决策模型，并定义了实现收支平衡的强制型和自愿型医疗保障制度。第四部分估计了模型的参数。基于估计结果，我们在第五部分中模拟了现行的中国农村医疗保障制度，考察了参与率、收支平衡性和福利影响等方面的问题。此外，我们还对如何设计最优医疗保障制度进行了初步的探讨。第六部分是全文的总结。

二、中国农村消费—医疗支出 决策行为的基本特征

中国农村的合作医疗制度建立于1955年，到20世纪70年代中期，合作医疗制度覆盖了超过90%的大队。这一制度被世界银行称为“低收入发展中国家举世无双的成就”（世界银行，1993）。然而70年代末、80年代初这一制度全面瓦解，到80年代末，只有5%的村还存在合作医疗制度。农村生产从集体化变为个体化被认为是这一制度瓦解的主要原因，乡村管理部门由此失去了分配产出的权利，合作医疗基金入不敷出，医务人员队伍不稳定。另一方面，这一制度自身缺乏明确的原则和科学的规划也是原因之一，市场化改革使人们的认识发生变化，乡村集体不再承担对医疗体系的组织和动员工作。在没有了政治压力后，很多人开始逃避缴费，加剧了合作医疗制度的财务困难（Feng et al., 1995；朱玲，2000）。从90年代初开始不少地方政府进行了恢复农村合作医疗的努力，但到1998年也只有约9.5%的农村人口有医疗保障，90%的农村居民没有任何形式的医疗保险（Liu et al., 2002）。

没有医疗保障的中国农民的医疗支出是比较低的，1997年人均医疗支出仅为66元，到2003年也只有115元左右，不到城镇人均医疗支出水平的1/4。³决定医疗支出的因素很多，但收入和医疗相对价格显然是其中的关键因素。⁴就收入的影响而言，根据农业部农村固定观察点调查数据，即“全国农村社会经济典型调查数据”，我们将观察户依据各年人均收入水平进行五等分，形成五个组，并把最低收入组的收入正规化为1。在同一年份，不同收入水平的农户的医疗支出倾向有较大的差异。虽然人均收入高的农户人均医疗支出水平比较高，但图1显示，医疗支出倾向则相反，即人均收入越低的组，医疗支出倾向越高。

³ 资料来源：《中国统计年鉴》。

⁴ 这里的医疗相对价格并不仅仅表示药品、医疗服务等方面的相对价格，它是一个更为广义的概念，包含了由于制度因素所引起的有效医疗价格的变化，比如由于村级医疗设施的消失造成的农民就诊费用的上升。

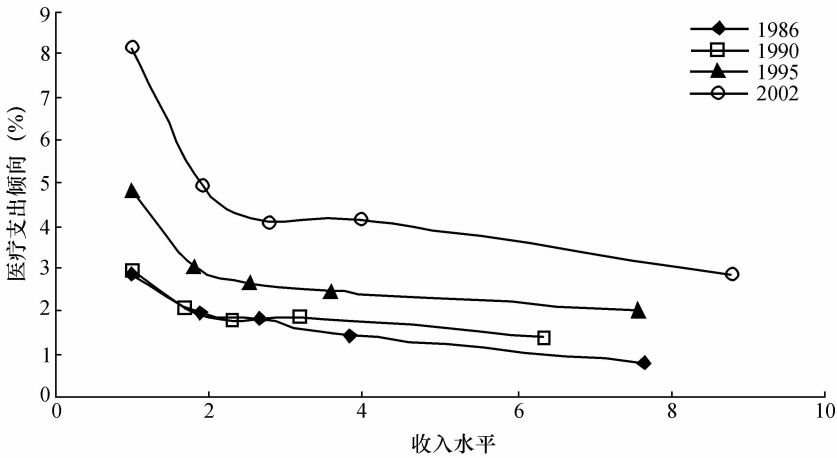


图 1 医疗支出倾向 (%)

资料来源：根据“全国农村社会经济典型调查数据”整理。

另一方面，图 1 显示随着时间的推移，医疗支出倾向曲线整体上移，这一现象就需要用收入以外的其他因素加以解释。医疗价格上涨是一个不争的事实（表 1），如果医疗支出缺乏价格弹性，医疗费用的增长就是一个必然的结果。在 90 年代，平均一次门诊费从 1990 年的约 10 元上涨到 1999 年的 79 元，年均增长 24.5%，平均一次住院费从 1990 年的 473 元上涨到 1999 年的 2891 元，年均增长 22.25%，而同期以现价衡量的农民人均纯收入的年均增长率仅为 13.88%。可见，医疗费用的增长幅度已经大大超过了同期农民收入的增长幅度。因此，中国农民医疗支出的特征事实是农户的医疗支出倾向随收入递减，各收入水平的农户的医疗支出倾向均随时间呈现上升趋势。这两个特征事实反映出农村医疗支出缺乏收入弹性和价格弹性。⁵

表 1 中国医疗费用增长与农民人均收入比较(现价)

年份	1985	1990	1995	1999	1990—1999 年的 增长率(%)
农民人均纯收入(元,下同)	397.6	686.31	1577.74	2210.34	13.80
平均每一次门诊医疗费	—	10.9	29.6	79	24.53
平均每一出院者住院医疗费	—	473.3	1273.0	2891.1	22.25

资料来源：引自陈佳贵主编的《中国社会保障发展报告》(1997—2001)。社科文献出版社 2001 年版。

与医疗支出相对照，农村居民对其他物品的消费同样表现出一定的刚性（图 2）。随着收入的增加，消费倾向逐渐下降，表明消费的增长幅度落后于

⁵ 我们另一项用 CHNS(中国健康和营养调查数据)的研究表明，农民医疗需求收入弹性为 0.3，价格弹性为 0.7(Feng et al., 2007)。

收入的增长幅度。相应地，农民的储蓄则表现出比较高的收入弹性。尤其值得注意的是最低收入组的消费倾向大于1，表明农民必须通过动用储蓄或负债来满足必要的消费水平。另一方面，消费倾向曲线随时间表现为不断下移。一个可能的解释是在医疗价格迅速上涨的时期，消费相对于医疗的价格有所下降。由此可以推断农民的消费价格弹性也是不足的。⁶下面我们将通过一个简单的异质性个体决策模型来拟合上述特征事实，并以此为基础来考察中国农村医疗保障制度的参与率、收支平衡性及其对异质性个体的福利影响。

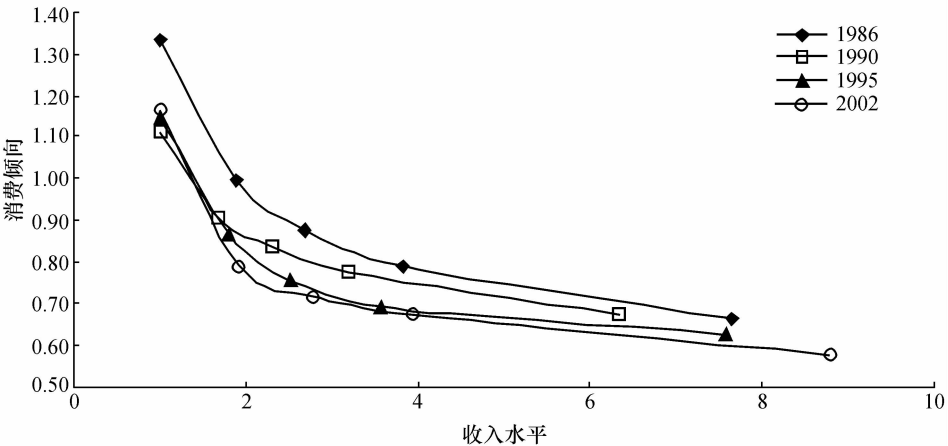


图2 消费倾向

资料来源：根据“全国农村社会经济典型调查数据”整理。

三、一个医疗支出决策模型

Grossman（1972）首先研究了人们对健康的需求以及由此引起的消费—医疗支出行为。基于他的研究，⁷我们构建一个单期模型，分析农户短期的消费—医疗决策问题。农户的决策通常以家庭为单位进行，这里我们将一个农户家庭抽象为一个个体，用农户家庭的人均状况作为这个个体的特征。个体的效用不仅取决于消费水平，还受到健康状况的影响。由于患病会带来负的效用，医疗支出可以提高健康水平，从而直接影响消费者的效用。与传统模型不同，我们引入了个体能力和初始健康水平这两个刻画个体特征的变量，以体现个体决策在这两个维度上的差异，从而考察医疗保障制度通过异质性个体的消费—医疗支出选择所产生的包括参与率、收支平衡性和福利效应等

⁶ 农民的消费行为还可能受到其他很多因素的影响，诸如教育、养老等，在农村社会保障制度严重缺乏的情况下，农民保持较高的储蓄倾向是一种理性的选择。

⁷ 在Grossman的模型中，健康既是消费品又是投资品，健康的投资性主要表现在减少疾病带来的时间损失，通过增加市场性或非市场性活动时间，获得更多的收益。

在内的一系列结果。

假设个体的能力 $\omega(i)$ 和期初的健康水平 $H_0(j)$ 外生给定, 用 i 区分个体能力的差异, φ 是 $\omega(i)$ 的概率密度函数, 用 j 区分个体初始健康状况的差异, φ 是 $H_0(j)$ 的概率密度函数, $i=1, \dots, I, j=1, \dots, J$ 。 ω 独立于 H_0 。存在一个医疗保障制度。如果这个制度是自愿型的, 个体需要选择是否加入。之后, 他会遭遇一个可能改变其健康水平的外生冲击 (可以理解为生病或康复)。用 H_0 和 $\tilde{H}$ 分别表示冲击前后健康水平的向量, $\tilde{H}$ 服从一个马尔可夫过程

$$\tilde{H} = \Pi \cdot H_0, \quad (1)$$

其中 Π 为转移矩阵, 可以表述为:

$$\Pi = \begin{bmatrix} \pi(1, 1) & \dots & \pi(1, j) & \dots \\ \vdots & \vdots & \vdots & \vdots \\ \pi(k, 1) & \vdots & \pi(k, j) & \vdots \\ \vdots & \dots & \dots & \pi(J, J) \end{bmatrix},$$

元素 $\pi(k, j)$ 是个体健康水平由 $H_0(j)$ 变化至 $\tilde{H}(k)$ 的概率, 用 $\tilde{\varphi}$ 表示 $\tilde{H}(k)$ 的分布, $k=1, \dots, J$ 。冲击发生后, 个体参加劳动并获得收入。劳动收入主要取决于能力 $\omega(i)$, 但健康状况也可能影响个体的收入, 因此我们假设收入服从

$$W(i, k) = \omega(i) + \lambda \tilde{H}(k), \quad (2)$$

其中 $\lambda > 0$ 表示健康对于收入的回报率。个体根据 $\tilde{H}(k)$ 和 $W(i, k)$ 以及是否有医疗保障进行消费 $C(i, k)$ 和医疗支出 $h(i, k)$ 。用 $H(i, k)$ 表示期末的健康水平, 有

$$H(i, k) = h(i, k) + \tilde{H}(k), \quad (3)$$

即医疗支出可以改善健康状况。上述事件发生的时间顺序如图 3 所示。



图 3 时间顺序

(一) 最优医疗支出

我们从后向前求解这一模型。首先来看给定 $\tilde{H}(k)$ 和 $W(i, k)$ 以及参加医疗保障体系的情况下个体的最优消费—医疗支出选择。假设效用函数是加性可分的。为了反映农民可以通过动用储蓄或负债来满足消费和医疗支出, 除了消费和健康以外, 我们在效用函数中加入财富。效用函数满足如下形式

$$U(i, k) = \frac{C(i, k)^{1-\sigma} - 1}{1-\sigma} + \alpha \frac{H(i, k)^{1-\gamma} - 1}{1-\gamma} + \beta \frac{\exp(-\theta A(i, k))}{-\theta}, \quad (4)$$

效用函数由三部分构成。前两部分是消费和健康的效用，它们服从常见的CRRA形式，其中 $\sigma > 0$ 和 $\gamma > 0$ 分别表示消费和健康的相对风险规避系数。第三部分表示财富的效用，它服从CARA形式，其中 $\theta > 0$ 是财富的绝对风险规避系数。 $\alpha > 0$ 和 $\beta > 0$ 分别反映了健康和财富的相对权重。⁸

除了(2)和(3)式外，个体决策还要受到如下条件的约束：

$$C(i, k) + (1 - \tau)Ph(i, k) + A(i, k) = RA_0(i, k) + W(i, k) - T, \quad (5)$$

$$h(i, j) \geq 0, \quad (6)$$

(5)式为预算约束，其中 P 为医疗商品和服务相对于其他消费品的价格， τ 为保障体系对医疗费用的共付比例， $\tau \in [0, 1]$ ， $A_0(i, k)$ 为个体期初的财富， R 体现财富的总收益率， T 为医疗保障制度所要求的保险费。根据中国农村医疗保障制度的实际情况，我们假设 T 为人头税。在没有医疗保障制度或个体没有选择参加医疗保障体系的情况下， $\tau = T = 0$ 。由于医疗支出非负，所以需要加上(6)式。值得一提的是，这个模型允许负债。 $A(i, k) < 0$ 反映了农民会借钱对消费和医疗支出进行融资。

这个最优问题的一阶条件为：

$$C(i, k)^{-\sigma} = \beta \exp(-\theta A(i, k)), \quad (7)$$

$$\alpha(\tilde{H}(k) + h(i, k))^{-\gamma} = \beta(1 - \tau)P \exp(-\theta A(i, k)) + \mu, \quad (8)$$

其中 μ 为库恩-塔克乘子，当 $h > 0$ 时 $\mu = 0$ 。利用(2)、(3)、(5)、(6)、(7)和(8)式，可以发现一些非常直观的比较静态结果：最优医疗支出水平 Ph 与个体能力 ω 正相关，与保险费 T 和健康水平 $\tilde{H}$ 负相关。

(二) 参加医疗保障体系的选择

解出最优消费—医疗行为以后，我们可以得到个体参加以及没有参加医疗保障体系的间接效用函数，分别用 $\tilde{V}^1(i, k)$ 和 $\tilde{V}^0(i, k)$ 表示。根据(1)式就可以得到个体在遭遇健康冲击前对于参加或不参加医疗保障体系的效用期望 $V^0(i, j)$ 和 $V^1(i, j)$ 。

$$V^l(i, j) = E[\tilde{V}^l(i, k)] = \sum_k \pi(k, j) \tilde{V}^l(i, k), \quad l = 0, 1. \quad (9)$$

如果医疗保障制度遵循自愿参加原则，个体选择参加自愿型医疗保障制度的条件为⁹：

⁸ 假设财富效用服从CARA形式的原因在于个体可能选择负债，即 $A(i, k) < 0$ ，这违反了CRRA型效用函数的规定。

⁹ 我们假设在预期效用相等的情况下个体也会选择参加医疗保障体系。

$$V^j(i, j) \geq V^0(i, j). \quad (10)$$

根据 (10) 式, 每一个个体决定他是否参加医疗保障体系, 由此可以得到愿意参加医疗保障制度的个体在能力 ω 和初始健康状况 H_0 上的分布, 分别用 ϕ^1 和 φ^1 表示。

(三) 医疗保障制度的平衡

我们先来考察强制型医疗保障制度, 即个体必须支付保险费 T , 享受补贴 τ 。如前文所述, 中国目前试行的农村医疗保障制度的原则是多方筹资, 政府根据个体支付保险费的总和 T 来制定补贴 G 。假设政府遵循的补贴规则为 $G = F(T)$, 把人口数量正规化为 1, 医疗保障制度的预算平衡为:

$$T + F(T) = \sum_i \sum_k \tau Ph(i, k) \phi(\omega(i)) \bar{\varphi}(\bar{H}(k)). \quad (11)$$

(11) 式的含义为, 在既定的政策参数下, 不同的个体行为选择各自的医疗支出 $h(i, k)$, 在已知个体的分布后, 就可以得到医疗保障制度的总支出, 即 (11) 式的左边。(11) 式的右边是医疗保障制度的总收入, 由征收的医疗保险费和政府补贴构成。在保险费和政府补贴既定的情况下, 就需要通过调整医疗支出共付比例 τ , 并考虑它们对个体决策的影响, 从而最终满足 (11) 式的要求。更为严格地, 我们给出如下定义:

定义 1 一个在给定医疗相对价格 P 、保险费 T 和政府补贴函数 $F(\cdot)$ 下实现强制型医疗保障制度收支平衡的均衡由 $\{C(i, k), h(i, k), A(i, k), \tau\}$ 构成, 它们满足:

(1) 给定 P 、 τ 和 T , 个体根据 (2)、(3)、(5) 至 (8) 式求解 $C(i, k)$ 、 $h(i, k)$ 和 $A(i, k)$ 。

(2) 给定所有的 $h(i, k)$, 医疗支出共付比例 τ 满足 (11) 式。

由于 $h(i, k)$ 没有解析解, 在下文中我们将采用不动点迭代的方法求解均衡下的 τ 。具体说来, 首先猜测均衡 τ 为 τ_0 , 然后解出 $h_0(i, k)$ 。把 $h_0(i, k)$ 代入 (11) 式可以得到一个新的 τ_1 。再根据 τ_1 解出 $h_1(i, k)$, 如此循环, 直到 $\tau_n = \tau_{n+1}$ 。

就数学形式而言, 自愿型医疗保障制度的收支平衡与强制型制度类似, 我们只需要将 (11) 式中的 T 、 ϕ 和 $\bar{\varphi}$ 替换为 T 、 ϕ^1 和 φ^1

$$T + F(T) = \sum_i \sum_j \tau Ph(i, k) \phi^1(\omega(i)) \varphi^1(H_0(j)), \quad (12)$$

其中 T 是参加保障体系的个体缴纳的保费总和 $T = \sum_i \sum_j T \phi^1(\omega(i)) \varphi^1(H_0(j))$,

$h(i, k)$ 是能力为 $\omega(i)$ 、期初健康水平为 $\tilde{H}(j)$ 的个体在遭遇健康冲击以后的医疗支出。相应地，定义 1 需要被修正为

定义 2 一个在给定医疗相对价格 P 、保险费 T 和政府补贴函数 $F(\cdot)$ 下实现自愿型医疗保障制度收支平衡的均衡由 $\{C(i, k), h(i, k), A(i, k), \tau\}$ 构成，它们满足：

- (1) 给定 P 、 τ 和 T ，个体根据 (2)、(3)、(5) 至 (8) 式求解 $C(i, k)$ 、 $h(i, k)$ 和 $A(i, k)$ 。
- (2) 给定 $C(i, k)$ 、 $h(i, k)$ 和 $A(i, k)$ ，个体根据 (9) 和 (10) 式选择是否参加医疗保障体系，进而得到参加医疗保障体系的个体在能力 ω 和初始健康状况 H_0 上的分布 ϕ^1 和 φ^1 。
- (3) 给定所有的 $h(i, k)$ 和分布 ϕ^1 和 φ^1 ，医疗支出共付比例 τ 满足 (12) 式。

求解自愿型医疗保障制度均衡的方法与前文类似。具体说来，首先猜测均衡 τ 为 τ_0 ，然后解出 $h_0(i, k)$ 、 ϕ_0^1 和 φ_0^1 。把 $h_0(i, k)$ 、 ϕ_0^1 和 φ_0^1 代入 (12) 式可以得到一个新的 τ_1 。再根据 τ_1 解出 $h_1(i, k)$ 、 ϕ_1^1 和 φ_1^1 ，如此循环，直到 $\tau_n = \tau_{n+1}$ 、 $\phi_n^1 = \phi_{n+1}^1$ 和 $\varphi_n^1 = \varphi_{n+1}^1$ 。

四、参数估计

我们的目标是在个体理性选择的基础上考察农村医疗保障制度的参与率、收支平衡性及其福利效应。由于医疗保障制度的变化会改变个体决策环境，所以我们不能直接估计消费函数和医疗支出函数。一个可行的方法是对模型的参数作校准 (calibration)，然后模拟医疗保障制度的福利影响。但是由于此前估计中国个体偏好和医疗相对价格的研究很少，因此需要校准的参数很多，校准的标准本身也就成为一个需要讨论的问题。为了避免这些争议，本文采用以模拟为基础的计量方法，对模型的未知参数进行估计 (参见 Duffe and Singleton, 1993 ; Adda and Cooper, 2003)。需要估计的参数有个体偏好参数 $\sigma, \gamma, \theta, \alpha$ 和 β ，以及各年的医疗相对价格 $P(t)$ 。估计的基本方法参见附录。估计的结果如表 2 所示。

表2 参数估计值

个体偏好参数	相对价格
$\sigma = 1.651$	$P_{1986} = 0.077$
$\gamma = 8.547$	$P_{1990} = 0.084$
$\theta = 0.772$	$P_{1995} = 0.176$
$\alpha = 1.743$	$P_{2000} = 0.444$

$\beta = 0.594$

$P_{2001} = 0.442$

$P_{2002} = 0.441$

注 根据“全国农村社会经济典型调查数据”(1986、1990、1995、2000—2002)校准。

表 2 表明， $\gamma/\sigma = 5.18$ ，即个体对于健康的风险规避程度明显大于消费。另一个结果是医疗相对价格的演进。在 1986 年至 1990 年间，医疗相对价格基本保持平稳。随着市场化进程的加快，从 1990 年到 2000 年，医疗相对价格上涨了 5 倍以上，这与表 1 中的数据是一致的。2000 年以后，相对价格非常稳定，保持在 0.44 左右。最后，由于 γ 比较大，医疗价格弹性比较小（等于 - 0.316），个体的实际医疗支出存在一定的刚性，实际医疗支出的下降幅度小于医疗相对价格的上升幅度，造成名义医疗支出与医疗相对价格正相关。

有三点需要说明。首先，图 4 给出了上述估计的拟合效果，其中实线和

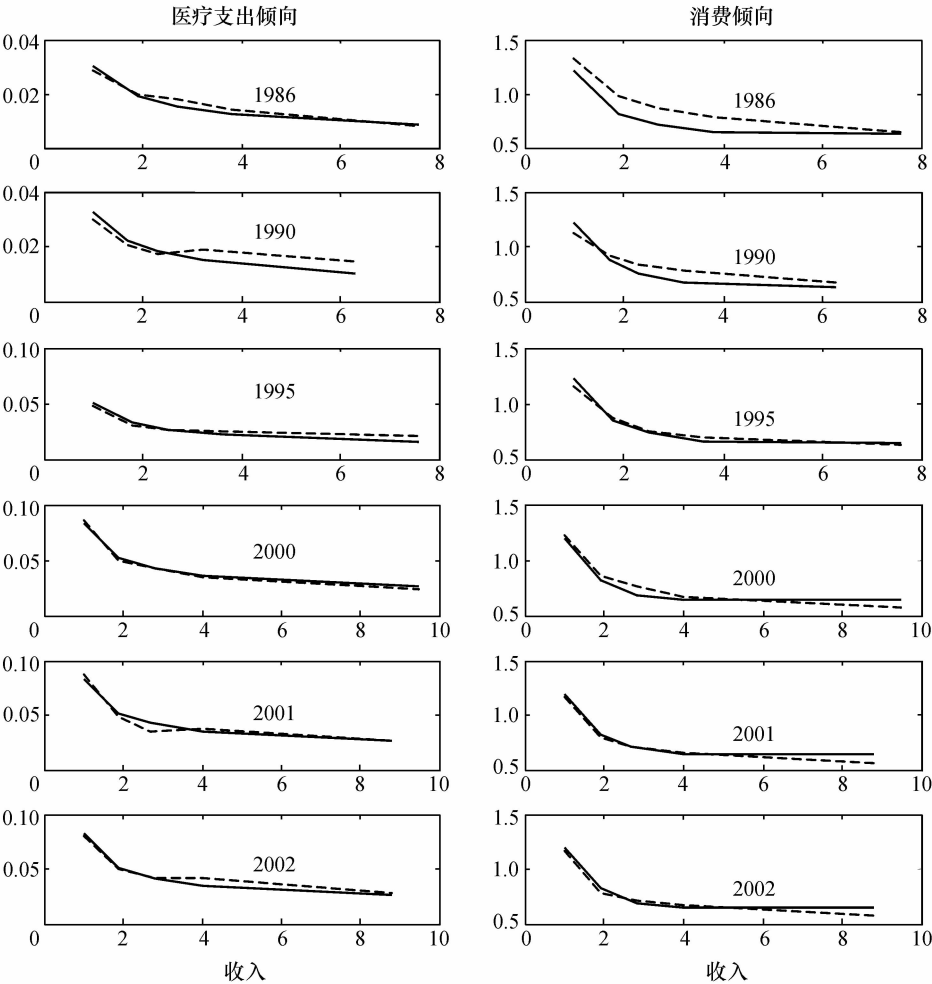


图 4 拟合效果

虚线分别表示拟合值和实际值。可以看出，拟合的误差 ($R^2 = 0.928$) 很大程度上来自于对 1986 年和 1990 年消费倾向的估计。但是，我们的模型可以很好地拟合最近几年中国农民的消费—医疗支出行为，这为下文分析医疗保障制度的一系列影响提供了比较坚实的基础。其次，估计的相对价格变化可能并不完全等价于医疗真实价格的变化。由于各年的相对价格 $P(t)$ 是模型中唯一随时间可变的参数，它实际上还包含了模型所不能刻画而又随时间变化的决策环境的信息。最后，一般说来，非线性最小二乘法对于初始猜测值的要求很高，搜寻结果可能并不唯一。我们对初始猜测值做了简单的摄动 (Perturbation)，发现各类初始猜测值均能收敛到表 2 给出的估计值，因此我们认为这里的搜寻结果至少是局部唯一的。

五、医疗保障制度的参与率、平衡性和福利效应

在参数估计的基础上可以模拟实现平衡的医疗保障制度的参与率、共付比例和福利效应。为了更好地模拟医疗保障制度的现实影响，我们采用 2002 年的医疗相对价格作为模拟参数。计算期望效用 V^0 和 V^1 还需要设置转移矩阵 Π 。令 $\pi(i, j) = \rho$ 和 $\pi(i, j) = (1 - \rho)/(J - 1)$ ，对于 $i \neq j$ 。 ρ 表示没有遭遇健康冲击的概率，我们令 $\rho = 0.8$ 。敏感性分析表明不同的 ρ 对结论的影响不大。然后我们考虑中国现行的农村医疗保障制度。按照 2002 年最低收入组的收入水平（人均 782 元），我们取 T 为最低收入组收入水平的 $1/78$ 。设政府补贴规模为总的保险费的 2 倍，定义 1 和 2 中的政府补贴函数为 $G = 2T$ 和 $G = 2T'$ 。

首先考察强制型医疗保障制度。根据定义 1，实现医疗保障制度收支平衡的 $\tau = 0.531$ 。没有医疗保障制度的预期效用等价于 V^0 ，因此实行强制型医疗保障制度对于个体的福利效应就等于 $V^1 - V^0$ ，用 ΔV 表示，具体的数值参见表 3。可以看出，实行强制型医疗保障制度的最大受益者是收入较低且健康较差的个体。但是，这个保障制度会损害收入较低和健康较好个体的福利。对于高收入者而言，医疗保障制度的影响不大。如果医疗保障制度是自愿的，表 3 中左上方福利受到损失的个体不会选择参加医疗保障体系。这时，收缴的保险费会下降。根据定义 2，实现自愿型医疗保障制度收支平衡的 $\tau = 0.509$ ，参与率为 92%。随着福利受损个体的退出，与强制型制度相比，共付比例和参加保障体系的个体福利改善的幅度均有所下降。

表3 医疗保障制度的福利效应

健康状况 收入水平	低		中低		中等		中高		高	
	强制	自愿	强制	自愿	强制	自愿	强制	自愿	强制	自愿
低	0.068	0.065	0.040	0.038	0.014	0.013	-0.001	—	-0.001	—

中低	0.051	0.049	0.033	0.031	0.016	0.015	0.002	0.000	0.005	0.000
中等	0.040	0.038	0.027	0.026	0.015	0.014	0.003	0.003	0.001	0.001
中高	0.029	0.027	0.021	0.020	0.013	0.012	0.006	0.005	0.001	0.001
高	0.011	0.010	0.009	0.008	0.007	0.006	0.005	0.004	0.002	0.002

在此基础上，我们进行一些比较静态分析。首先，将缴费金额增加 1 倍，即 T 为最低收入组收入水平的 1/39，在强制型制度下， $\tau = 0.749$ ；在自愿型制度下， $\tau = 0.654$ ，参与率下降到 76%。表 4 的结果显示，在强制型制度下，受益最大的依然是收入较低且健康状况较差的人，但缴费金额提高后，福利受损的人数将会增加，这些人是收入较低而健康较好的人。因此，在自愿型制度下，这些人必然会退出医疗保障制度，而他们退出的结果是医疗保障体系可以提供的共付比例下降，如此，又将导致另一些人因福利受损而退出这一体系。最后的结果是大部分健康状况很好的人都将退出保障体系，实现医疗保障制度收支平衡的共付比例也有明显的下降。

表 4 缴费金额增加一倍后的福利效应

健康状况 收入水平	低		中低		中等		中高		高	
	强制	自愿	强制	自愿	强制	自愿	强制	自愿	强制	自愿
低	0.086	0.076	0.051	0.042	0.019	0.010	- 0.000	—	- 0.000	—
中低	0.065	0.058	0.042	0.036	0.020	0.014	0.001	—	0.001	—
中等	0.050	0.045	0.034	0.030	0.019	0.015	0.005	0.001	0.002	—
中高	0.030	0.033	0.026	0.023	0.016	0.014	0.007	0.004	0.002	—
高	0.014	0.013	0.011	0.010	0.008	0.008	0.006	0.005	0.003	0.003

然后我们把政府对于个体的财政补贴减少为 10 元，即补贴函数修改为 $G = T$ 和 $G = T'$ 。在强制型制度下， $\tau = 0.433$ ；在自愿型制度下， $\tau = 0.388$ ，参与率下降到 80%。表 5 的结果显示，受益最大的依然是收入较低且健康状况较差的人。如果取消财政补贴，自愿型制度下的 $\tau = 0.261$ ，参与率会进一步下降到 72%。

表 5 财政补贴减少一半后的福利效应

健康状况 收入水平	低		中低		中等		中高		高	
	强制	自愿	强制	自愿	强制	自愿	强制	自愿	强制	自愿
低	0.054	0.047	0.031	0.027	0.010	0.008	- 0.002	—	- 0.002	—
中低	0.041	0.036	0.026	0.022	0.012	0.010	- 0.001	—	- 0.001	—
中等	0.032	0.028	0.021	0.019	0.011	0.010	0.002	0.001	0.000	—
中高	0.023	0.020	0.016	0.014	0.010	0.009	0.004	0.003	0.001	0.000
高	0.009	0.008	0.007	0.006	0.005	0.005	0.004	0.003	0.002	0.002

（一）讨论

如前所述，现有关于中国农村医疗保障制度的讨论主要集中在以下三点。第一，农民是否有激励参加遵循自愿原则的医疗保障制度，或者说这

个保障体系的参与率究竟有多大；第二，目前的缴费金额和政府补贴能否实现保障体系的自我平衡；第三，人头税型的缴费方式是否会使富人得益更多，从而进一步加剧中国农村的不平等。我们的估算结果对以上三个问题做出了回应。首先，只有小部分收入较低而健康较好的农民不会选择参加医疗保障体系。具体说来，在每人每年缴费仅为低限 10 元时，我们估计这个保障体系的参与率可以达到 92%。其次，即使个体和政府只需缴纳 10 元和 20 元的保险费，只要把医疗支出的共付比例 τ 控制在 50% 左右，目前的医疗保障制度是可以实现自我平衡的。最后，这个保障体系的主要受益者是收入较低而健康也较差的个体，它符合医疗保障制度向病人和穷人倾斜的基本要求。

前人提出的这三个疑问是很直观的。那么为什么在我们的模型中目前试行的医疗保障制度没有出现这些问题呢？首先，由于缴费金额很低，即便就最低收入组而言，保险费也仅占收入的 1.3%，参加保障体系的成本很小。而考虑到可能遭遇的健康冲击以及由此产生的医疗支出的不确定性，在较高的风险规避系数的作用下，参加保障体系可以显著降低风险，提高预期效用。表 4 的模拟结果表明如果缴费金额上升一倍，参与率就会下降到 76%，逆向选择问题就会比较突出。但这仅仅是问题的一个方面。其次，如果总和医疗支出很大，实现保障制度收支平衡的共付比例会比较小，保障体系的风险规避效应也就随之减少。不过中国农民医疗支出比较小，因此即便在保险费很低的情况下，经过政府财政的补贴，实现保障制度收支平衡的共付比例依然可以保持在比较高的水平。表 5 表明，如果取消补贴，参与率会下降到 72%。因此，我们认为较少的缴费金额、较高的风险规避性、较低的医疗支出倾向和政府财政补贴是导致现行医疗保障制度逆向选择问题较小的主要原因。

随之而来的是道德风险问题，即引入医疗保障制度以后，共付比例 τ 使得个体医疗支出的有效价格 $P(1 - \tau)$ 下降，这可能诱使个体提高医疗支出，增加保障体系的负担，从而降低实现保障制度收支平衡的共付比例。所幸的是，我们的估计表明，医疗价格弹性较小，实际医疗支出 h 存在着一定的刚性，所以当 $P(1 - \tau)$ 下降以后， h 的增加幅度小于 $P(1 - \tau)$ 下降的幅度，实现保障制度平衡的共付比例因此并不会显著减少。最后，相对于富人而言，人头税型的缴费方式的确增加了穷人的负担，而且数据表明富人的医疗支出也比穷人多。但是， h 的刚性使得穷人的医疗支出倾向显著高于富人，而且由于穷人的收入偏低，遭遇健康负向冲击以后可能需要负债，这一财富效应会引起较大的福利损失。综合以上两点，穷人在保障体系中有较高的相对受益。我们的估算表明，穷人这部分较高的相对受益超过了因人头税型的缴费方式造成的相对损失。

(二) 关于最优医疗保障制度的一些思考

一个有趣的问题是我们能否在某种政府补贴规则下找到“最优”的保险费。首先,最优性需要满足帕累托条件,因此我们只考虑自愿型医疗保障制度。其次,我们需要一个总福利函数,即个体期望效用的加权总和。如果用选择参加医疗保障体系个体的分布 ϕ^1 和 φ^1 作为权重,总福利的增进 $\mathfrak{N}$ 等于

$$\mathfrak{N} = \sum_i \sum_j \Delta V(i, j) \varphi^1(\omega(i)) \phi^1(H_0(j)). \quad (14)$$

图 5 给出了 $\mathfrak{N}$ 对于保险费 T 的函数形状。图 5 中有 6 个区域。当 T 较小时 (区域 1), 所有人都愿意参加医疗保障体系, 而且 $\mathfrak{N}$ 是 T 的单调递增函数。在政府补贴与个人缴费成正比的补贴规则下 (参见图 6), 这是一个显然的结果。但是, 当 T 增加到一定程度以后 ($T > 0.0107$, 区域 2), 收入较低且健康较好的个体不会选择参加医疗体系, 因此收缴的保险费、实现医疗保障制度收支平衡的共付比例都有所下降, 导致总福利的增进 $\mathfrak{N}$ 出现了一个明显的下降。但是, 只要选择参加医疗体系的个体数量不变, 在区域 2 中 $\mathfrak{N}$ 依然是 T 的单调递增函数, 而且只要 T 足够大, 区域 2 中的 $\mathfrak{N}$ 可以超过区域 1 中 $\mathfrak{N}$ 的最大值。根据我们的估计 ($T = 0.013$), 目前试行的中国农村医疗保障制度就落在区域 2 的中部。随着 T 的增加, 上述情况依然成立, 如图 5 和图 6 所示。由此可见, 虽然增加保险费可能改善个体的总和福利, 但是却需要面对财政补贴上升和医疗体系参与率下降的问题。我们计划在以后的研究中对最优保险费乃至最优政府补贴规则做更加深入的研究。

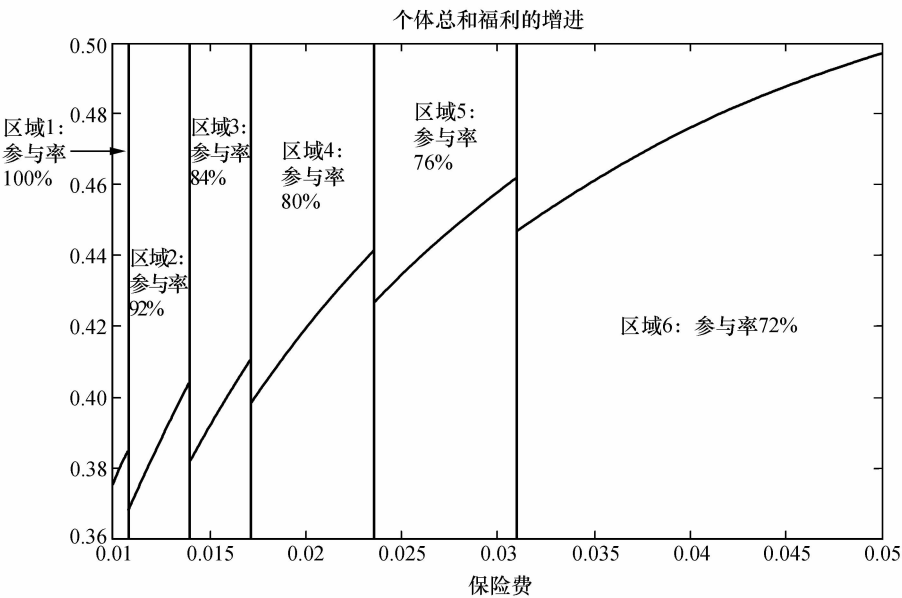


图5 总福利改进与参与率

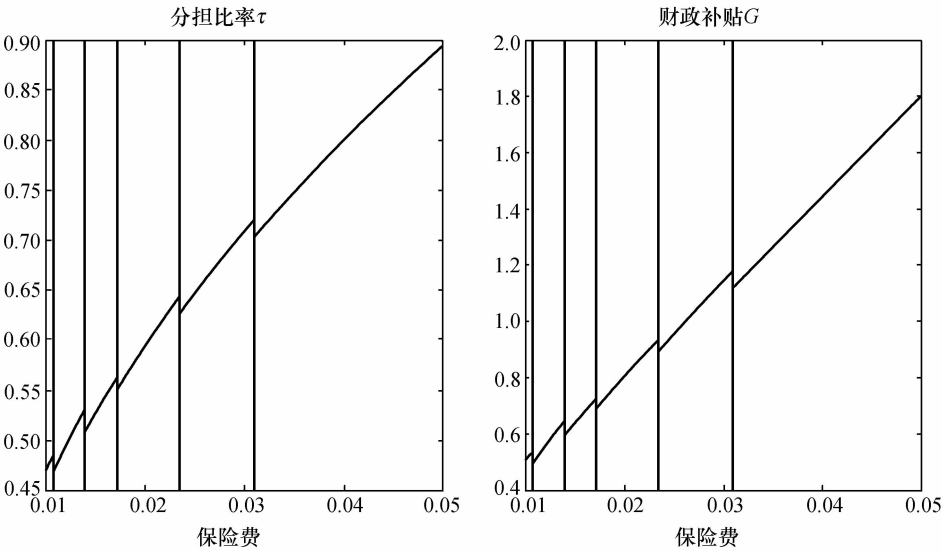


图6 保险费、共付比例和政府补贴

六、总 结

本文构建了一个异质性个体的消费—医疗支出决策模型，并在拟合中国农村消费—医疗支出行为的基础上估计了中国农民的效用函数和医疗相对价

格。在对实现收支平衡的医疗保障制度进行严格定义以后,我们计算了现行中国农村医疗保障制度对于参与率、费用共付比例以及福利效应等三方面的影响。我们的计算结果表明,在现行的制度设计和缴费水平下,参与率可以在 75% 以上,最高可以达到 92%,健康状态较差的穷人是这个保障体系最大的受益者。较少的缴费金额、较高的风险规避性、较低的医疗支出倾向和政府财政补贴导致逆向选择并不显著,道德风险则受到了较低的医疗价格弹性的限制,而穷人较高的医疗支出倾向则使得他们可以从保障体系中获得更大的利益。

值得注意的是,很多调查发现,现实中农民参与医疗保障制度的积极性并不高。有文献报告在推行医疗保障制度的过程中平均参与率只有 56% 左右(胡善联,2004),这与本文估算的高于 75% 的参与率似乎并不一致。需要指出的是现实中农村合作医疗一般只补贴大病,这将加重逆向选择,降低合作医疗的吸引力,降低参与率。仅仅补贴大病带来的另一问题是不利于农民在健康出现问题时及时治疗,导致小病拖成大病,反而引起更多的医疗支出,也影响到合作医疗自身的资金平衡。本文的模型在医疗保险补贴上未区分大病和门诊费用,结果表明这样的设计可以减轻逆向选择问题,而且医疗保障体系的共付率可以达到 50% 左右。此外,除了本文模型所刻画的经济、健康因素以外,农民的选择还受到很多其他因素的影响。较为突出的一个因素是长期以来农民对集体经济组织怀有一定的不信任感,担心乡村干部会以权谋私或合作医疗资金被挪用,因而用不参与来对抗(曾庆义等,2003)。

现行的中国农村医疗保障制度按地区统筹,而中国地区之间的情况差别很大,本文可能无法反映由这种差异所造成的医疗保障制度对于不同地区的不同影响。但是,我们的模型和结论并不依赖于收入的绝对水平。如果各地区的收入分布比较近似,统筹范围的变化不会改变基本的结论。此外,各个地区试行的费用分担规则也并非都是线性的,比如“大病保障”规则或“报销上限”规则都会对个体的医疗支出行为产生影响。今后,我们将在获得分地区的个体数据和费用分担规则的基础上对农村医疗保障制度做更为细致的评价。

另一个未来研究的方向是延续第五节中关于最优医疗保障制度的讨论。现有的文献往往在收入可观测或同质个体的假设基础上寻找最优的医疗保障制度(比如 Blomqvist and Horn, 1984; Blomqvist, 1997)。而我们则计划在收入不可观测的异质性个体模型中寻找“最优”的保险费、费用分担规则和政府补贴政策,因为这样的“最优”制度在中国农村更加易于实施。

附录 参数估计方法

正文中参数估计的具体方法如下:我们假设个体初始的财富 $A_0 = 0$ 。个人能力 ω 和冲

击后的健康水平 $\tilde{H}$ 服从均匀分布。¹⁰ 令 $J=5$, $\tilde{H}=(0.7 \ 0.9 \ 1.1 \ 1.3 \ 1.5)$ 。敏感性分析表明放松这些假设并不会对结论产生本质影响（限于篇幅，我们略去了敏感性分析结果）。¹¹ 令 $I=5$, ω 的取值依据全国农村社会经济典型调查数据。¹² 用 W^* 表示实际观测到的收入（把最低收入组的收入正规化为 1）， $\omega=W^*-\lambda\tilde{H}$ 。由于模型中的收入 $W=\omega+\lambda\tilde{H}$ ，经过实际收入的调整以后， λ 的取值与下文中的参数估计和福利分析结果无关。如前文所述，中国农村在过去二十多年间基本没有医疗保障制度，所以我们令 $\tau=T=0$ 。

需要估计的参数有个体偏好参数 $\sigma \ \gamma \ \theta \ \alpha$ 和 β ，以及各年的医疗相对价格 $P(t)$ 。估计的基本方法如下。给定 $\omega(i)$ 、 $\tilde{H}(k)$ ，个体的医疗支出倾向 $z(i, k)=Ph(i, k)/W(i, k)$ 和消费倾向 $c(i, k)=C(i, k)/W(i, k)$ 可以被表示为参数向量 $v=(\sigma \ \gamma \ \theta \ \alpha \ \beta)$ 和相对价格 $P(t)$ 的函数，即 $z(i, k)=f_{i, k}(v, P(t))$ 和 $c(i, k)=g_{i, k}(v, P(t))$ 。在 $\omega(i)$ 给定情况下按 $\tilde{H}$ 平均的各年医疗支出倾向和消费倾向分别为 $\bar{z}_t(i)=\sum_k f_{i, k}(v, P(t))\bar{\varphi}(\tilde{H}(k))$ 和 $\bar{c}_t(i)=\sum_k g_{i, k}(v, P(t))\bar{\varphi}(\tilde{H}(k))$ 。用 $z_t^*(i)$ 和 $c_t^*(i)$ 表示实际观测的各年分组数据，我们根据下面的非线性最小二乘法估计参数向量 v 和相对价格 $\{P(t)\}_{t=1}^T$

$$\min_{v, \{P(t)\}_{t=1}^T} \sum_t \sum_i ((z_t^*(i) - \bar{z}_t(i))^2 + \mu(t)(c_t^*(i) - \bar{c}_t(i))^2),$$

其中 $\mu(t)=(\bar{z}_t^*/\bar{c}_t^*)^2$ ， $\bar{z}_t^*$ 和 $\bar{c}_t^*$ 是 $z_t^*(i)$ 和 $c_t^*(i)$ 的均值。由于 $z_t^*(i)$ 和 $c_t^*(i)$ 的数值相差很大（前者接近于 0 而后者接近于 1），我们使用 $\mu(t)$ 调整这两个变量的量纲。估计的结果如表 2 所示。

参 考 文 献

[1] Adda, J., and R. Cooper, Dynamic Economics: Quantitative Methods. Cambridge: The MIT Press, 2003.

[2] Blomqvist, A., "Optimal Non-linear Health Insurance", Journal of Health Economics, 1997, 16(3), 303—321.

[3] Blomqvist, A., and H. Horn, "Public Health Insurance and Optimal Income Taxation", Journal of Public Economics, 1984, 24(3), 353—371.

[4] Cutler, D., "Public Policy for Health Care", NBER Working Paper, 1996, No. 5591.

[5] Duffie, D., and K. Singleton, "Simulated Moments Estimation of Markov Models of Asset Prices", Econometrica, 1993, 61(4), 929—952.

¹⁰ 第五节中对于转移矩阵 II 的设置保证了 H 与 $\tilde{H}$ 同分布。

¹¹ $\tilde{H}=(0.7 \ 0.9 \ 1.1 \ 1.3 \ 1.5)$ 的取值带有一定的随意性，这是因为 $\tilde{H}$ 的均值和健康的权重 α 无法被相互独立地估计出来， $\tilde{H}$ 的均值越大， α 的估计值就越小。不过，在我们所进行的各种敏感性分析中， $\tilde{H}$ 的均值及其分布假设对本文的主要结论不会产生重大影响（因为其主要作用都被 α 的估计值所反应）。我们还可以直接估计 $\tilde{H}$ 的上界和下界，这样除了相对价格以外，需要估计的参数就变成了 7 个。估计的结果与这里的设置差别不大，对于本文的基本结论也不构成影响。

¹² 农村调查数据按收入五等分。

- [6] Feng, J., B. Qin, and Y. Yu, "Wealth, Education and Demand for Medical Care: Evidence from Rural China", Paper presented in 2007 Midwest Economics Association Annual Conference, US.
- [7] Feng, X., S. Tang, G. Bloom, M. Segall, and X. Gu, "Cooperative Medical Schemes in Contemporary Rural China", *Social Science and Medicine*, 1995, 41(8), 1111—1185.
- [8] Grossman, M., "On the Concept of Health Capital and the Demand for Health", *Journal of Political Economy*, 1972, 80(2), 223—255.
- [9] 胡善联, "全国新型农村合作医疗制度的筹资运行状况", 《中国卫生经济》2004 年第 9 期, 第 24—25 页。
- [10] Liu Y., K. Rao, and S. Hu, People's Republic of China: Towards Establishing A Rural Health Protection System, Asian Development Bank Publication Stock No. 090902, 2002.
- [11] 曾庆义、李新民、张绍军, "农民自愿还需政府组织引导——对建立新型农村合作医疗制度有关问题的思考", 《中国卫生经济》2003 年第 11 期, 第 39—40 页。
- [12] 朱玲, "政府与农村基本医疗保健保障制度选择", 《中国社会科学》2000 年第 4 期, 第 89—99 页。

Health Care System in Rural China : A Quantitative Approach Based on Heterogeneous Individuals

JIN FENG ZHENG SONG

(Fudan University)

Abstract This paper aims at providing some responses to the following three hotly debated issues regarding China's new rural cooperative medical system (CMS), which was launched at the beginning of 2003. Firstly, how many people would join the system voluntarily? Secondly, can the system be self-balanced? And thirdly, would the lump-sum tax benefit the rich more than the poor? We build a decision model with heterogeneous agents and compute the implications of the CMS and find under certain conditions, the system can be self-financed and sustained and the rate of participation could be higher than 90%. Moreover, it is the unhealthy poorer that benefit more from the CMS.

JEL Classification C15, I38, I28

网上交易中的私人秩序 ——社区、声誉与第三方中介

吴德胜*

摘 要 网络扩大了市场范围,减少了信息收集成本和信息传播成本,使买家和卖家更容易配对。但是,网上 C2C 交易中也存在着严重的囚徒困境问题。交易网站以及其他第三方中介建立了信用评价等一系列机制来解决网上交易的治理。本文主要探讨在法律制度不完善或法律不能很好发挥作用的情况下,交易方以及追求利润最大化的第三方中介如何建立起私人秩序来保证交易的顺利进行。这些私人秩序在一定程度上起到了替代法律这种公共秩序的作用。

关键词 网上交易,私人秩序,社区,声誉,第三方中介

引 言

随着信息技术的飞速发展,因特网的普及,电子商务浪潮席卷全球各地,网上购物这一新的消费模式也正融入人们的生活。美国 eBay 公司(www.ebay.com)成立于 1995 年,经过十余年的发展,已成为全球最大的 C2C(consumer to consumer)网上交易场所。截至 2006 年 6 月,eBay 注册用户数已超过了 2 亿,在全球 33 个国家和地区都开展了业务,全球范围内每时每刻有将近 5000 万件商品在 eBay 上出售。2005 年 eBay 的交易总额超过 443 亿美元。在 eBay 上交易的商品种类繁多,收藏品、计算机及配件、娱乐产品、汽车,甚至票务等商品,都能在 eBay 上找到。

1998 年 8 月,两位创业者模仿 eBay 的商业模式,创立了易趣网(www.eget.com)。2000 年 1 月根据 CNNIC 第五次“中国互联网络发展状况统计调查”显示,易趣成为中国最受欢迎的拍卖网站。2002 年 3 月,eBay 公司注资 3000 万美元,与易趣结成战略合作伙伴关系。2003 年 6 月,eBay 向易趣追加 1.5 亿美元的投资,全资收购了易趣,易趣也改名为 eBay 易趣(www.ebay.com.cn)。¹淘宝网(www.taobao.com)则由国内 B2B(企业对企业)网站——

* 南开大学公司治理研究中心。通讯地址:天津市南开大学公司治理研究中心,300071;电话:(022)23498746;E-mail:xdshengwu@yahoo.com.cn。本文为教育部哲学社会科学研究重大课题攻关项目(03JZD0018)和自然科学基金重点项目(70532001)的阶段性成果。作者感谢李维安教授的指导,匿名审稿人和主编姚洋教授提出的中肯而有价值的修改意见,以及李飞跃提出的有益评论。作者同时也感谢南开大学数学学院刘玉坤博士在模型上给予的帮助。当然,本文文责自负。

¹ 本文这里对名称做如下说明:eBay 指 eBay 公司的全球交易网站(www.ebay.com),eBay 易趣或易趣则指 eBay 公司的中文网站(www.ebay.com.cn),即原易趣网。2006 年 12 月 20 日,eBay 宣布与 Tom 公司共建合资公司,eBay 易趣将改名为 Tom 易趣。

阿里巴巴投资于2003年5月10日创办,凭借免费和本土化的策略,淘宝网在2005年已经超过了易趣,成为国内最大的C2C交易网站,而易趣仍然保持发展态势,但其市场份额已经从第一降至第二。据统计,2005年中国网上购物的人数达到2000万人,C2C成交额达135.05亿元,其中通过网上支付购物款的比例近50%。²

相比网下交易,网上C2C交易的优势有:第一,因特网减少了信息收集成本和信息传播成本,买家和卖家可以更容易配对,从而减少了交易成本。第二,因特网扩大了市场范围,消费者可以从网上搜寻到更便宜的商品,卖家也能面向全国甚至全世界销售商品,这就意味着交易者可以从网上C2C交易中获得更多的剩余。第三,网上C2C交易市场存在着网络外部性:一个卖家是否在网站上销售商品取决于其他卖家和买家的决策,买家是否在这个市场上购物又取决于这个网站上卖家的数目以及卖家登录商品的种类和数量,这样就形成了一个正反馈机制。亚马逊、当当网等B2C(business to consumer)网站每增加一位用户,客户关系只增加一次,网站只能向新用户出售产品。但对像eBay、淘宝网这样的网站来说,增加的每一位新用户都有可能与已有的用户建立交易关系。网上C2C交易吸引交易者的另外一个优势是,绝大多数国家目前对网上C2C交易实施免税政策,同样的商品,网上比网下便宜。

但是网上交易³也有其与生俱来的缺陷。买家和卖家在交易时并不见面,对货物的质量和特征没有网下购物的亲身体验,只能依靠卖家在网上对货物所做的描述,但这种描述是无法代替亲身体验的。网上交易还存在支付风险,当买家通过网络支付时,网络黑客可能会盗用其银行卡。也许网上交易面临的最大缺陷是交易过程中的信息不对称。在交易发生前,买家和卖家无法了解对方过去的交易历史。即使买家和卖家就交易达成一致,买家也担心如果他先付款,卖家会不发货或发送的货物与描述的不一致,卖家也担心如果他先发货,可能会收不到货款。这种逆向选择和双向道德风险问题严重阻碍了网上交易的发展。

与电子商务相关的法律也不完善,除非发生欺诈行为,司法部门一般很少介入网上交易中的各种争端。即使司法部门可以介入,由于交易条款很难得到证实,执行成本也非常高。

针对网上交易的上述问题,包括交易网站在内的第三方中介建立了信用评价系统和其他交易机制,如第三方托管(escrow)、第三方支付、在线争议解决等交易机制。

² 数据来源于中国互联网信息中心发布的第16次、17次和18次《中国互联网络发展状况统计报告》。

³ 由于网上B2B和B2C交易中囚徒困境问题相对较轻,理论分析的意义不是很突出,再加上网上B2C、B2C和C2C交易有融合的趋势,因此本文在这里笼统地用网上交易来代替网上C2C交易。本文之所以不使用网上拍卖这个概念是因为在易趣和淘宝网上绝大部分交易采用的是固定价格销售方式,竞价销售所占的比重很小。即使在美国eBay上,竞价销售所占的比例也在下降。

信用评价系统在网上交易中发挥着基础性的作用。信用评价系统鼓励买家和卖家在交易完成后对对方进行信用评价,这样信用评价系统就对买家和卖家的交易历史进行了概括。其他买家在交易发生之前都可以观察到卖家的信用评价,然后决定是否从卖家那里购买商品。因此,信用评价中的一些指标就可以衡量交易者的声誉,通过信用评价系统,声誉机制就可以发挥作用。网上交易的迅速发展,欺诈发生的概率非常小,⁴说明信用评价系统以及其他交易机制可以很好地保证交易方的合作行为,保障交易的顺利进行。

由于法律和基础制度的不同,中国的网上交易与美国 eBay 上的交易存在一些差异。eBay 把自己的功能严格限定为买家和卖家提供一个网上交易场所及其他非常重要的附带服务,如支付工具(PayPal)和通讯工具(Skype)。⁵相反,中国的交易网站则承担了更多的功能,比如淘宝网提供的支付工具——支付宝就兼有第三方托管的功能,淘宝网还承担争议解决,即执行中介的功能,而在美国,这类服务由网站以外的其他第三方中介提供。淘宝网还首先建立了卖家商盟制度。

因此,本文提出以下问题:在法律缺失或法律不能很好发挥作用的情况下,网上交易社区中卖家的声誉是如何发挥作用的?其他各种交易机制又起到何种作用?中国和美国网上交易出现差异的制度原因又是什么?

本文的思路类似于 Milgrom, North and Weingast (1990) 对中世纪欧洲集市贸易的兴起以及法商(law merchant)作用所做的分析和 Greif (1993) 对中世纪地中海马格里布(Maghribi)犹太人商人社区内部雇佣关系的研究。不同的是,网上交易中交易者的人数更多,而且交易者都是匿名的,重复交易所占的比例更少,⁶这就对信息的传递和惩罚的实施提出了更高的要求。

本文运用随机匹配博弈(random matching game)模型对网上交易中信用评价系统的作用进行分析,假设信用评价存在噪声,即信息传递是不完全的。模型的结论是:只要贴现因子足够大,买家观察卖家在最近时间里收到的信用评价,如果卖家有欺骗行为,买家就对其实施集体惩罚策略,这种集体惩罚策略可以激励卖家选择诚实行为。因此,本文的结论是:通过信用评价系统,声誉机制可以很好地发挥作用。本文进一步分析在线争议解决、第三方托管以及商盟等机制的作用:这些机制可以起到补充信用评价系统的作用。另外,由于中国不存在一个完善的全社会信用体系,网上交易与网下交易被割裂开来,网上交易中的欺骗记录对欺骗者网下的交易和生活并不会造成多大的影响,这导致中国的网上交易采取了一些与美国不同的交易机制。这样,

⁴ 据 eBay 的报道,每 40000 笔交易中可能有一起欺诈事件(Baron 2002)。

⁵ 由于网上交易中买卖双方沟通的重要性,淘宝网建立了专门用于淘宝网上交易的聊天工具——淘宝网旺旺。eBay 也收购了全球最大的即时通讯工具 Skype。引入即时通讯工具不但可以避免不必要的争议发生,还可以加快购物的速度。

⁶ 根据 Resnick and Zeckhauser (2002) 所做的一项调查,89% 的交易都是一次性的。

本文就说明了在法律和社会信用体系等公共秩序缺失的情况下，网上交易中自发的私人秩序和有组织的私人秩序可以起到替代公共秩序的作用。

本文的结构安排如下：第一部分介绍了与交易的治理机制有关的理论，以及目前对网上交易进行分析的理论和实证文献。第二部分通过一个随机匹配博弈模型来说明网上交易中自发的私人秩序，即建立在信用评价系统上的声誉机制如何发挥作用。第三部分分析了网上交易中有组织私人秩序所起的作用。第四部分简要地分析了中国和美国网上交易差异的制度原因。最后是本文的结论、不足以及进一步的研究方向。

一、文献综述——交易的治理机制

许多商业交易都存在囚徒困境 (prisoner's dilemma) 的问题。卖家可能通过提供低质量的产品来欺骗下游厂商，买家也有可能在卖家交付产品后拒绝支付货款。最优的结果是双方都坚持合作行为，即卖家提供高质量的商品，买家支付货款。但是，每一方都面临欺骗对方的诱惑。如果交易只发生一次，用博弈论的术语说，在一个静态的博弈中，博弈方的道德风险会导致一个无效率的纳什均衡，即卖家提供低质量的商品，买家拖欠货款。

(一) 自发的私人秩序——双边关系型契约

有许多治理机制可以打破囚徒困境的低效率均衡。一是依赖法律的正式契约 (formal contract)。如果交易方事先在合同中详细地描述产品的特征和质量、付款的细则，未来状态以及各种状态下双方的权利和责任，以及对欺骗行为的惩罚，那么法庭将会根据条款来执行合同，只要合同足够清晰，法律的执行是有效的，那么交易方将不会违约，合作结果的博弈均衡就可以达成。

但是，契约可能是不完备的 (incomplete)，法庭可能无法验证私人签订的法律条款。一个可以解决囚徒困境的市场治理机制是关系型契约 (relational contract) 或非正式契约 (informal contract)。⁷所谓关系型契约就是由契约的未来价值所维持的非正式的安排 (Baker, Gibbons and Murphy, 2002)，关系型契约可分为双边关系型契约 (bilateral relational contract) 和多边关系型契约 (multilateral relational contract)。双边关系型契约是指仅发生在两方之间重复的契约关系，多边关系型契约则指维持网络或社区成员合作行为的非正式契约安排，有的文献称其为社区规范 (Aoki, 2001)。关系型契约在商业交易中广泛存在。纵向的供应链关系和横向的战略联盟之间常常涉

⁷ 关系型契约的概念最早出现在社会学和法学文献中 (Macaulay, 1963; Macneil, 1978, 1985)，Williamson (1979) 把关系型契约的思想引入到交易成本经济学中。Baker, Gibbons and Murphy (1994, 2002) 正式地将不完备契约分析方法引入到关系型契约中，并用关系型契约来分析组织内部和外部的契约安排。

及长期、非正式的契约安排,企业利用关系型契约来协调无法预见的或然事件(contingencies)。我国农业产业过程中企业与农户之间就农产品生产和销售形成的分包制,由于很难由法律来实施,也是一种比较典型的关系型契约。

根据标准重复博弈中的无名氏定理(folk theorem)(Rubinstein, 1979; Fudenberg and Maskin, 1986),只要交易方有足够的耐心,博弈重复的阶段足够长,维持合作带来未来收益流的现值就有可能大于违约得到的短期收益,那么交易方将维持合作行为,双边的关系型契约就可以发挥作用。关系型契约不需要法庭或其他第三方来支持,它是自我执行的(self-enforcing)。

(二) 自发的私人秩序——多边关系型契约

但是,有时候交易方同另一方只进行一次或数次交易,在交易重复次数很少的情况下,交易又是如何治理的?比如,参与人有一方是固定的,而另一方是不固定的,一个例子是火车站旁的商店与乘客之间的关系,由于乘客与商店的交易是一次性的,商店就有可能提供价高质次的商品,乘客预期到这一点就尽量不在火车站旁的商店购买商品。还有一种更复杂的情况是,买家和卖家好像是从总体中随机抽出进行匹配的,一方同另一方重复交易的概率趋向于零。博弈论中把这种类型的交易称之为随机匹配博弈(random matching game)。当交易是随机匹配时,双边博弈中的惩罚机制就不再有效,合作行为就很难达成,交易又是如何治理的?

多边关系型契约或社区规范如果要发挥作用,就需要比双边关系型契约更严格的条件。一个条件是信息的迅速传播。在火车站旁商店的例子中,只要商店欺骗乘客的信息可以在乘客中低成本低地传递,就可以将所有的乘客看成一个顾客,参与人一方固定的交易多边关系型契约就等同于双边关系型契约,双边关系型契约下的惩罚机制就可以发挥作用,由于担心声誉⁸受到损害,商店可能会提供高质量的商品。

但是,提供违约信息是一项公共产品,对在某一交易中遭受欺骗的交易方来说,将对方违约的信息传递给其他交易方对整个团体的收益要大于对他的收益,单个交易方缺乏提供违约信息的激励。因此,多边关系型契约要发

⁸ 目前,对声誉的模型化处理有两种思路。第一种思路是在无限次重复博弈的框架内进行处理,把声誉看成是完全信息重复博弈中的均衡策略。例如,在一个重复囚徒困境中,“冷酷”策略的均衡,即“一直合作直到对手背叛,并且从此以后背叛”,可以解释成描述了这样的情况:每个人都有“合作”的声誉,但是这个声誉在他第一次背叛的时候就会消失。第二种思路是在不完全信息的有限博弈的框架内处理,即 KMRW 模型(Kreps, Milgrom, Roberts and Milson, 1982),假设博弈参与人的类型是不完全信息,这样,每个参与人的声誉就可以概括为他的对手对其类型的信念,例如,中央银行有坚持执行货币政策的声誉。对声誉的具体介绍见 Fudenberg and Tirole(1991, Chapter 9)。博弈论专家通常认为不完全信息的处理方式更严格、合理。但是,第一种思路却更直观,不像第二种思路需要假设博弈参与人有不同的类型和策略,在多个参与人的博弈中,大多数博弈论应用文献,比如 Greif(1993)和 Clay(1997)一般将完全信息重复博弈中的均衡策略也理解为声誉,这样博弈参与人的历史行为就与其收益联系在一起。

挥作用的另一个条件是交易方可以有效地实施集体制裁 (collective punishment), 否则交易方的策略就是不可置信的。但是在缺乏第三方中介的关系型契约中, 很难保证交易方不与违约方进行交易。如果不能对违约行为进行惩罚, 重复博弈的合作结果均衡就无法达成。这也解释了为什么没有第三方中介参与的多边关系型契约仅发生在特定的以血缘、地缘或种族为纽带的小社区中, 如 Greif (1993) 文中的马格里布犹太人社区, 民间金融盛行的浙江台州。在这些社区中, 人际交往有助于社区内部的信息传递, 如果社区成员有违约行为, 或没有对违约方进行惩罚, 整个社区将通过联合抵制对其实施集体惩罚。McMillan and Woodruff (2000) 把双边关系型契约以及这种仅依赖社区内部的声誉机制的多边关系型契约称为自发的私人秩序 (spontaneous private order)。

(三) 有组织的私人秩序

如果关系型契约所处的网络成员是流动的, 或者在一个较大、匿名的社区中, 没有地缘、血缘或种族关系的支持, 那么信息收集和传递的成本是非常高的, 集体惩罚也是不可能的, 声誉机制和自发的私人秩序就不再起作用。为了突破小社区的限制, 就必须依靠有第三方中介参与的有组织的私人秩序 (organized private order)。

博弈文献把标准重复博弈中的无名氏定理扩展到随机匹配博弈中。Kandori (1992)、Ellison (1994) 和 Okuno-Fujiwara and Postlewaite (1995) 等人的研究表明: 即使社区内部信息无法完全传递, 合作结果的博弈均衡也可以达到, 而且在某种程度上也是稳健的 (robust)。

除了上述博弈文献, 还有相当多的文献对经济史以及现代经济中商人社区内部的声誉机制的作用进行了分析。Greif (1993) 对中世纪地中海马格里布犹太人商人社区内部代理关系和 Clay (1997) 对墨西哥时代的加利福尼亚商人团体内部代理关系的研究表明, 只要社区规模不是很大, 信息的传递比较通畅, 只依靠声誉机制就可以保证交易的顺利进行。

但是, 如果社区规模比较大, 没有地缘或血缘等社会关系的支持, 信息收集和传递的成本是非常高的, 这时就必须依赖第三方中介来发挥作用。商会、行业协会等第三方中介所起到的第一个作用就是充当信息中介 (information intermediary), 给社区内部的交易者提供违约的信息。第三方中的另一个作用是充当执行中介 (enforcement intermediary), 实施统一的惩罚、裁决和剩余分配等 (McMillan and Woodruff, 2000)。这种由第三方中介支持的声誉机制可以保证交易的顺利进行, Milgrom, North and Weingast (1990) 对中世纪欧洲集市贸易中法商、Bernstein (1992) 对钻石行业中非正式契约关系、Gambetta (1993) 对意大利西西里岛黑手党、McMillan and Woodruff (2000) 对转型国家交易的私人执行机制以及 Bernstein (2001) 对棉花行业中第三方

中介作用的分析就说明了这一点。

(四) 自发的私人秩序、有组织的私人秩序与公共秩序

当交易条款只能被交易双方观察到,但不能被第三方证实或证实成本非常高时,关系型契约或自发的私人秩序才能发挥作用。有组织的私人秩序,如第三方执行中介的出现则使契约条款可以被证实。自发的私人秩序的优势是契约的执行不必依赖第三方,契约的执行成本较低,其缺陷是契约易受外界环境的影响,契约的稳定性因而较差,作用范围也较窄。相比自发的私人秩序,有组织的私人秩序的优点是第三方中介可以证实契约条款,使非正式的契约安排变得明晰和完备,增加了交易的“可缔约性”(contractibility),契约条款因而可由第三方中介来执行,但第三方执行中介的引入同时带来一定的契约成本。私人第三方中介相对于法庭的优势是,私人中介拥有专业知识和更多的信息,可以较低的成本来做出事实判断。一项争议对法庭可能是无法证实的,但对私人仲裁来说,可能是可证实的(Dixit, 2003a)。但私人第三方执行中介不具备法庭的强制力。

自发的私人秩序、有组织的私人秩序与公共秩序之间存在一定的互补性和替代性。一般来说,在法律制度运行良好的国家,私人秩序更多地起到减轻法律调查和执行过程中交易成本的作用;在法律制度不完美的国家或在法律不能很好发挥作用的情况下,私人秩序更多地起到替代法律的作用。

(五) 网上交易社区中的私人秩序和公共秩序

与一般商业交易相比,网上交易的复杂性在于:网上交易是非频繁的,不是重复交易,非常接近博弈论中的随机匹配博弈,特定买家与卖家之间交易重复发生的概率非常小,接近于零,这导致标准重复博弈模型中的个人惩罚机制(personal punishment)无法实施,仅依靠双边关系型契约不足以维持交易者的合作行为。如果信息的收集和传递成本很高,那么声誉机制也无法发挥作用。

虽然法律这种公共秩序对网上交易中的欺诈行为起到一定的遏制作用,但是,对于交易中发生的一般争议,法庭的证实成本相当高,因此,法庭不能很好地解决网上交易引发的各种争端。网上社区中交易者数以万计,不像以血缘、地缘或种族为纽带的社区,可以有效地传递信息并实施驱逐等惩罚措施,仅仅依靠自发的私人秩序可能无法保证网上交易的顺利进行。

以盈利为目的的第三方中介可以起到保障交易顺利进行的作用。第三方中介这种有组织的私人秩序在交易中主要充当信息中介和执行中介,传递交易者的有关信息,解决交易中出现的各种争议。C2C 交易网站,如 eBay、淘宝网就是网上交易中最重要第三方中介,它们建立了网上交易平台和双向信用评价系统,并提供了第三方支付工具。第三方托管机构、在线争议解决

机构、商盟则是网上交易中另外几种比较重要的第三方中介。

网上交易的自发的私人秩序和有组织的私人秩序相互补充。网上交易的声誉机制以一种成本较低的方式来保证交易者的合作行为，在线争议解决机制、第三方托管和商盟等机制则起到补充声誉机制的作用。

Lucking-Reiley (2000) 首先对美国 eBay 的发展、eBay 上的交易机制进行了详细介绍。Baron (2002) 和 Dellarocas (2003) 通过模型分析了网上交易中声誉机制如何通过信用评价系统发挥作用。还有一类文献 (McDonald and Slawson, 2002; Melnik and Alm, 2002; Resnick and Zeckhauser, 2002; Bajari and Hortacsu, 2003; Livingston, 2005; Houser and Wooders, 2006; Lucking-Reiley, Bryan, Prasad and Reeves, 2006) 通过采集 eBay 上的交易数据或采用实验的方法 (Bolton, Katok and Ockenfels, 2004) 研究网上声誉机制是否发挥作用，着重于卖家声誉 (正面的评价、负面评价或二者兼有) 对拍卖商品价格的影响。总的来说，这些实证文献的结论是卖家声誉对拍卖商品价格有显著的影响。李维安、吴德胜和徐皓 (2006) 以及周黎安、张维迎、顾全林和沈懿 (2006) 分别对淘宝网和易趣上卖家声誉的作用进行了验证。

尽管不少文献对网上交易的声誉机制做了理论和经验分析，但是很少有文献考虑第三方中介 (第三方托管、在线争议解决机构和商盟) 在网上交易中所起到的作用。另外，上述分析主要针对美国 eBay，由于中国的法律制度、社会信用体系不完善，中国的网上 C2C 交易有许多不同于 eBay 的特点，这也需要 we 进行分析，找出背后的深层原因。

二、网上交易中自发的私人秩序——声誉机制

本节主要介绍 eBay 和国内 C2C 交易网站，主要是易趣和淘宝网上的信用评价系统，探讨建立在信用评价系统上的声誉机制如何发挥作用。

(一) 网上交易中的信用评价系统

信用评价系统或反馈机制 (feedback system) 由 eBay 首先创立，并为其他 C2C 和 B2B 网站所模仿。信用评价系统提供了一个信息收集、传递的平台。eBay、易趣和淘宝网上信用评价的基本原则是：买家和卖家每成功交易一笔，就可以对交易对象作一次信用评价。评价分为“好评”、“中评”和“差评”三种。⁹当交易顺利完成时，交易者可以给予对方好评，并给出类似如下的评价：“非常好的卖家，极力推荐”、“商品很好，发货及时”；当发货较慢、商品与网上描述有较小的差别时，买家一般给予卖家中评，评价的内容

⁹ 易趣和淘宝网将信用评价分为好评、中评和差评，美国 eBay 则把反馈相应地称为：正面 (positive)、中性 (neutral) 和负面 (negative)。

可能是“商品与网上描述的颜色有些不一致”；当卖家交付的物品与网上描述有显著差距，或卖家在交易过程中态度不好时，买家可以给予差评，并给予类似如下的评价：“商品有破损，与网上严重不符，要求退货时态度恶劣”。好评加一分，中评没有影响，差评扣一分。将交易者从所有的好评、中评和差评中收到的分数累加起来就得到交易者的信用度，其中好评数在交易次数中所占的比例就是卖家的好评率。如果交易者的信用度低于一定的值（eBay 上为 -4），系统就会自动终止用户在网上交易的权利。

美国，许多交易者既是买家同时也是卖家，但是在易趣和淘宝网上绝大多数交易者买家和卖家的身份是分离的。相对销售商品，购买商品更容易获得好评。因此，在中国，许多卖家通过购买价值较小的商品来获得较高的信用评价。为了避免这种行为，淘宝网将交易者作为买家和卖家收到的信用评价分开，使信用评价系统更能准确地传递交易者的信息。

淘宝网要求交易完成以后，交易双方必须进行信用评价，如果买家或卖家（一般是买家）在规定的时间内没有进行评价，系统自动地给予对方好评。卖家认为买家给予的差评不合理也可提交淘宝网仲裁，避免了买家以威胁给予差评来敲诈卖家。另外，eBay 和淘宝网都制定一些规则防止卖家用不真实的交易来“炒作”信用。总的来说，网上交易的信用评价系统目前已经比较成熟，可以很好地衡量交易者的声誉。

信用评价系统的作用在于它自动地给每一个交易者建立了一个简介（profile）。从交易者简介中可以观察到买家和卖家积累起来的声誉：信用度、好评率、好评数、中评数、差评数、差评率，历史交易中交易对手给予的评价的内容，在最近的时间里（1 周、1 个月及 1 年中）买家或卖家收到的好评、中评和差评的数目。如附录 C 中的图 1 和图 2 所示。交易者简介就相当于博弈论中博弈方的历史信息。信用评价系统使得交易者简介成为公共信息，所有买家在交易发生之前都能观察到卖家的简介，然后决定是否从卖家那里购买商品。

如果信用评价系统不存在，信息传递的速度将非常慢，不诚实的卖家即使欺骗了买家，在相当长的一段时期内也不会被其他买家发现。这样，卖家选择欺骗得到的暂时收益将大大高于选择诚实可以得到的长期收益，卖家就没有激励去建立长期声誉，网上交易市场规模就会很小，甚至市场根本就不存在。

只要信息传递是通畅的，网上交易社区内部的声誉机制就能发挥作用。但是，在社区内部传递违约信息是有成本的，社区越大，成本越高。在网上交易中，很难想象如果由一个机构专门负责信息收集和传递将产生多高的成本。现实中，eBay、易趣和淘宝网上的信息发布是由交易者做出的，而信息收集和传递则是由信用评价系统自动完成的，eBay 等网站只负责核实交易者信息的准确性。因此，双向信用评价系统通过分散化的方式（uncentralized）

来处理信息,交易者和网站需要付出的信息收集和传递成本非常低。相比 Milgrom et al. (1990)、Greif (1993) 等文献中的信息传递机制,信用评价系统中的交易者简介以更充分、更概括,但又不损失效率的方式在买家和卖家中传递信息 (Dellarocas, 2003)。

(二) 模型：信用评价不存在噪声

网上交易的经验证据显示,卖家的声誉相对买家的声誉在网上拍卖中更重要,网上交易的支付流程也要求买家付款后卖家再发货,因此,卖家更有可能出现道德风险问题。为了分析的简化,本文假设网上交易中只有卖家存在机会主义行为,因此这是一个单边囚徒困境问题 (one-sided prisoner's dilemma): 卖家可以提供高质量或低质量的商品,买家决定购买或不购买。遵循博弈模型中的术语,当卖家提供高质量商品时,可以认为其在博弈中选择了诚实行为;当卖家提供低质量商品时,认为其选择了欺骗。卖家选择欺骗可以得到 d , 选择诚实可以得到 c , $d > c$ 。贴现因子为 δ 。

假设一个完全同质产品的网上交易市场存在众多的卖家和买家。为了分析的简化,进一步假设卖家出售商品的价格相同,因此,买家只关注卖家收到的信用评价,而不关注卖家商品的价格。假设卖家有两种类型: 诚实卖家和机会主义卖家。¹⁰ 诚实卖家和机会主义卖家所占的比例分别为 x 和 $1 - x$ 。大部分卖家都是机会主义卖家,因此 $x < 1/2$ 。诚实卖家永远都选择诚实行为,机会主义卖家在决策前会进行成本收益计算: 如果选择诚实的长期收益大于选择欺骗的短期收益,那么选择诚实;否则选择欺骗。

在淘宝网上,卖家的交易者简介列出了卖家在最近 1 周、最近 1 个月和最近 6 个月内收到的好评、中评和差评数目,¹¹ 因此,买家可以清楚地了解卖家在最近的时间里是否有欺骗行为。如果观察到卖家在最近收到差评,买家就不会同该卖家进行交易。本文下面将证明: 如果买家遵循这种没有协调的集体惩罚策略,由于声誉的考虑,在一定的条件下,卖家就会选择诚实行为。

命题 1 当信用评价不存在噪声时,如果

$$\frac{d - c}{c} \leq \frac{\delta(1 - \delta^T)}{1 - \delta} \quad (1)$$

成立,下面的策略组合就是一个子博弈完美均衡。

买家策略: 如果卖家在最近 T 时段里收到过差评,则不从该卖家那里购买商品; 卖家策略: 卖家不会欺骗买家。

¹⁰ 本文这里没有像 Tirole (1996)、Dixit (2003b) 等文献那样假设交易者有三种类型: 诚实者、不诚实者和机会主义者。因为不诚实卖家没有耐心,总会选择欺骗,但 eBay、易趣和淘宝网都规定卖家的信用度一旦低于一个临界值,就禁止其在网上进行交易。因此,假设卖家只有两种类型是一个合理的简化。

¹¹ eBay 和易趣列出了卖家在过去的 1 个月、6 个月和 12 个月内得到的好评、中评和差评的数目。

证明见附录 A。

定义 $(d - c)/c$ 为卖家欺骗的收益率, 因此, 只要商品欺骗的收益率不是非常大, 建立在信用评价系统上的声誉机制就可以保证卖家的诚实行为。

但是, 由于网上交易不像网下交易买卖双方可以当面交流, 对商品也有亲身体验, 有时尽管卖家尽其所能向买家提供合适的商品或服务, 但由于沟通的不便, 买家却对卖家产生了误解, 认为卖家有意欺骗, 并给予卖家一个差评。eBay、易趣和淘宝网提供的卖家交易者简介中往往可以看到卖家和买家就争议各执一词, 有时很难判断究竟是卖家还是买家应该为争议负主要责任。因此, 假设信用评价有可能出现噪声比较合理, 卖家即使选择了诚实行为, 仍有可能被买家误解, 从而在其交易者简介中留下差评。

(三) 模型: 信用评价存在噪声

假设卖家选择诚实行为但买家认为其选择了欺骗的概率为 α , 即信用评价出现噪声的概率为 α , 如果 α 接近于 $1/2$, 那么信用评价系统发挥的作用很小。本文这里假设 α 较小。进一步假设如果卖家选择了欺骗, 买家不会错误地认为其选择了诚实, 即买家对卖家欺骗行为的判断不会出现差错。

命题 2 即使信用评价存在噪声, 当下式

$$\frac{d - c}{c} \leq \frac{(1 - \alpha)\delta(1 - \delta^T)}{1 - (1 - \alpha)\delta - \alpha\delta^{T+1}} \quad (2)$$

满足时, 下面的策略组合仍是一个子博弈完美均衡。

买家策略: 如果卖家在最近的 T 时段里收到过差评, 则不从该卖家那里购买商品; 卖家策略: 卖家不会欺骗买家。

证明见附录 B。

对诚实的卖家来说, 只有信用评价出现噪声时, 他才会收到差评, 因此诚实卖家的好评率将为 $1 - \alpha$ 。在均衡时, 机会主义卖家也会选择诚实行为, 诚实卖家和机会主义卖家的好评率就没有差异。在现实中, 卖家的好评率都很高, 几乎都在 95% 以上,¹² 这说明出于长期声誉的考虑, 机会主义卖家也选择了诚实行为。

从上式可以看出, 均衡条件取决于 α 和 δ 。噪声出现的概率 α 越小, 贴现因子 δ 越大, 上述条件越容易满足。

在 α 、 δ 和 T 相同的情况下, 不等式 (2) 的右边小于不等式 (1) 的右边。¹³ 与信用评价不存在噪声相比, 在信用评价存在噪声时, 声誉机制要发挥

¹² 好评率的高低还与卖家出售商品的种类有关。有些商品, 比如虚拟点卡, 销售过程比较简单, 大部分卖家的好评率都在 99% 以上。

作用需要更严格的条件。换一种表述方式,令 δ^* 和 δ^{**} 分别为使两个不等式等号成立的贴现因子。因为(1)式和(2)式右边是 δ 的增函数,故 $\delta^{**} \geq \delta^*$,在信用评价存在噪声时,只有当贴现因子较高时,声誉机制才能发挥作用。因此,信用评价的噪声减弱了声誉机制的作用。

当 $\alpha = 0.02$, $T = 30$, $\delta = 0.99$ 时,不等式(1)为 $(d - c)/c \leq 25.8$,不等式(2)为 $(d - c)/c \leq 16.7$ 。对于网上交易的一般商品来说,欺骗的收益率不是很大,(2)式的约束可以保证卖家选择诚实行为。

所以,只要欺骗的收益率不是非常大或贴现因子足够大,即使信用评价存在噪声,如果买家按照如下策略来选择卖家:观察卖家的交易者简介,如果交易者简介显示卖家在最近的一段时间里有差评,就不选择该卖家,那么卖家将选择诚实行为,建立在信用评价系统上的声誉机制就能发挥作用。

(四) 进一步的讨论及不足

本文的思路类似于 Greif (1993)。不同的是, Greif 是在完全信息无限重复博弈的框架内处理声誉的,将商人的历史行为理解为声誉。本文则是在不完全信息重复博弈的框架内处理声誉的,将声誉理解成买家对卖家类型的信念。¹⁴

本文假设买家通过观察卖家的交易者简介,根据卖家最近一段时间内是否收到差评来选择卖家,而且考虑到了噪声的影响,本文的假设及结论比较符合网上交易的实际情况。

网上交易是如此丰富,以至于不能用一个简单的模型说明所有的问题。近期有过欺骗行为的卖家有可能通过降价来吸引买家,鉴于同时引入信用评价和价格过于复杂,本文没有考虑这种情况。不过,随着网上交易社区的扩大,买家和卖家的数目越来越多,交易的商品种类也越来越多,每一种商品将有更多的卖家,也将面临更多的替代商品,这一方面导致市场竞争强度增加,商品价格的差异性会减少,买家更关注卖家的信用评价,卖家出售商品的价格相同这一假设将更接近现实。另一方面,交易也更频繁,这意味着贴现因子 δ 变大,更接近于1。在其他参数不变的情况下,网上交易的发展将扩大声誉机制的作用范围。

¹³ 不等式(2)右边的分母大于不等式(1)右边的分母,且(2)式右边的分子小于(1)式右边的分子,故(2)式的右边小于(1)式的右边。

¹⁴ Greif 的处理方法存在一些不足。当不存在一个具有强制力的执行中介时,如果一个代理人-商人欺骗某一个委托人-商人后,其他委托人-商人并没有动力却惩罚这个代理人,因此就不存在一个集体惩罚机制。为了构建一个子博弈完美均衡, Greif 引入了代理人雇佣市场,并假设所有代理人的保留效用和能力相同,由于有欺骗历史的商人未来被雇佣的概率较低,在代理人参与约束的限制下,有欺骗历史的商人所要求的工资就高于没有欺骗行为的商人,这样其他委托人-商人就不会雇佣有欺骗行为的商人。因此, Greif 虽然没有假设商人具有不同的类型,但引入了代理人雇佣市场,并增加了所有商人是同质的这一不合理的假设。

另外,本文虽然没有像一般的重复博弈模型那样假设买家的策略是触发战略(trigger strategy)或针锋相对策略(tit-for-tat),但是本文中买家的集体惩罚策略仍存在一个协调的问题:如果买家不能就惩罚期长度达成共识,就有买家同正处于惩罚期的卖家进行交易,那么上面的策略组合就不是一个子博弈完美均衡。但是,淘宝网卖家交易者简介列出了卖家在最近1周、最近1个月和最近6个月内是否收到过差评。因此,买家可以根据卖家销售商品的价值的大小来选择惩罚的时间长度。对于价值较小的商品,选择1周的惩罚期就可以激励卖家选择诚实行为;对于一般商品或价值较高的商品,买家可以分别实施期限为1个月或半年的集体惩罚。交易者简介实际上自动地起到了协调的作用。

网站不但建立了交易平台,还建立了一个网上社区,通过社区论坛(BBS)上的交流,大多数买家对惩罚期的长度可以形成比较一致的认识,这样也能部分解决协调问题。即使买家对惩罚期的长度没有形成一致的认识,假设 $T^{\min}$ 为所有惩罚期中的最小值,那么买家仍然可以实施长度为 $T^{\min}$ 的集体惩罚。

三、网上交易中有组织的私人秩序

本文第二部分主要分析了网站作为信息中介的作用,即提供了信用评价系统。信用评价系统对买卖双方的交易历史进行简要地描述,提供了交易方欺骗的历史信息,买家因而可以对在最近的时间里欺骗行为的卖家实施集体惩罚,声誉机制从而可以发挥作用。尽管网站介入了信用评价系统的建立,起到了信息中介的作用,但网站并没有收集和传播信息,这些功能都是系统自动完成的,信用评价系统主要是依靠声誉机制在发挥作用。因此,可以认为信用评价系统更接近自发的私人秩序。

已有的文献大多集中于网上交易社区中的声誉机制是否有效,而没有考虑第三方中介作为有组织的私人秩序在网上交易中所起到的作用。这一节进一步分析eBay、易趣和淘宝网上在线争议解决机制、第三方托管以及商盟制度对声誉机制的补充作用。

自发的私人秩序的优势是契约的执行不必依赖第三方,契约执行成本较低,其缺陷是契约易受外界环境的影响,契约的稳定性较差,作用范围也较窄。如果欺骗的收益率非常高,即(2)式不满足,声誉机制就不能发挥作用。第三方托管等机制可以降低卖家欺骗的收益率,使价值较高(欺骗的收益率因而也较高)的商品也能在网上交易。在线争议解决机制则可以使无法证实的契约条款变得可以证实,提高了契约的完备程度。当引入第三方中介的成本较低时,网上交易中的有组织私人秩序就可以起到补充自发的私人秩序的作用。网上交易的声誉机制以一种成本较低的方式来保证交易者的诚

实行为，在线争议解决机制、第三方托管等机制则起到补充声誉机制的作用。

（一）网上交易的争议解决机制

1. 法律及执法部门

对于网上交易中出现的一般争议，法律无法解决或解决成本非常高。但是对于网络欺诈，则必须依靠法律的强制力。尽管欺诈发生的概率非常低，据 eBay 的报道，每 40000 件交易中有 1 件是欺诈。但是，相对网下交易，网上交易出现欺诈的概率非常高。根据 FBI 的报告，2001 年 42% 的欺诈投诉与网上拍卖有关。¹⁵对于网上欺诈行为，eBay 成立了专门的网上交易欺诈调查小组对 eBay 上的交易进行监视，并在出现欺诈行为后配合联邦贸易委员会（Federal Trade Commission，FTC）等相关部门进行调查。

虽然法律这种公共秩序对网上交易中的欺诈行为起到一定的遏制作用，但是，对交易中发生的一般争议，由于契约是不完备的，法庭可能无法验证契约条款，因此，法律制度不能很好地解决网上交易中出现的各种争端。

2. 网站提供的争议解决机制

除了网络欺诈，网上交易中的纠纷大多是物品未收到或物品与卖家在网上所做的描述显著不符。¹⁶当出现“物品未收到或与描述显著不符”（Item Not Received or Significantly Not as Described）时，eBay 建议交易者通过“物品未收到或与描述显著不符”处理流程来进行沟通，使用信用卡支付的用户需要先通知信用卡公司，在美国，信用卡公司给予了消费者很好的保护，发卡公司通常提供防止在线欺诈的保护措施。“物品未收到或与描述显著不符”处理流程是一个买家和卖家交流的标准平台，许多问题仅仅是由误解造成的，通过相互沟通可能就能解决问题。如果通过“物品未收到或与描述显著不符”处理流程无法解决争议，eBay 建议交易者将争议提交中立的在线争议解决机制（Online Dispute Resolution，ODR），绝大多数争端都可以通过这两种方式得到解决。eBay 并不介入争议的解决，如果买卖双方仍无法达成一致，使用 PayPal 进行支付的用户可以申请贝宝买家保护项目（PayPal Buyer Protection）或贝宝卖家保护（PayPal Seller Protection）的补偿。用户每半年只能申请三次补偿。

3. 私人第三方仲裁——在线争议解决机制

在线争议解决主要包括在线沟通（online negotiation）、在线调解（online mediation）和在线仲裁（online arbitration）三种方式。在线争议解决是自愿的，只有当买家和卖家都同意将争议提交私人第三方中介时，在线争议解决

¹⁵ Baron, David P., "Private Ordering on the Internet: The eBay Community of Traders", Business and Politics, 2002, 4(3), p. 271.

¹⁶ 除了网上欺诈行为，网上交易出现的争端绝大多数是以上两种。

方才可以介入到争议中。SquareTrade.com 是 eBay 推荐其用户使用的在线争议解决网站, 它为 eBay 上的交易者提供在线沟通和在线调解这两种争议解决方式。

当买家和卖家同意将争议提交 SquareTrade 后, 争议进入在线沟通期。SquareTrade 为买家和卖家提供一个密码保护的网页, 争议双方在网页上传递信息, 交换各自的观点和请求。网站实质上起的是一个提供场所、加强沟通、促进交流的作用。很多争议通过彼此沟通便很快得到解决。SquareTrade 的在线沟通对所有用户都是免费的。

当买家和卖家无法通过在线沟通解决争端, 争议进入到在线调解 (online mediation) 阶段。争议由具有相关的专业知识和技能的调解员来调解。调解员帮助争议双方分析争议焦点, 明确双方的利益, 帮助当事人之间达成解决方案。在整个争议解决过程中, 调解员并不评价争议双方的对错, 也不由调解员做出调解协议。在线调解是在线争议解决中使用最多、最重要的一种方式。在线调解是收费的, 对一项 eBay 上的在线争议调解 SquareTrade 的收费是 19.5 美元, 其中的 15 美元由 eBay 代替用户支付。

在线争议解决机制的一个作用是扩展了交易的“可缔约性”, 即提高了契约的完备程度, 使契约变得明晰 (explicit) 和正式 (formal)。网上交易中的契约条款是法庭无法证实的或证实成本非常高, 但是私人仲裁机构可以利用其专业知识, 以较低的成本做出事实判断。一项争议对法庭可能是无法证实的, 但对私人仲裁来说, 可能就是可证实的 (Dixit, 2003a)。由于私人仲裁机构比法庭掌握更多的信息和专业知识, 在美国的司法惯例中, 法官往往遵循私人仲裁机构的裁决, 因此第三方仲裁也具有一定的强制力。由于正式契约的作用范围大于关系型契约, 在线争议解决机制的引入扩大了交易机会。

买家和卖家可以通过一些措施降低交易条款的证实成本, 比如, 交易者可以保留交易过程中双方的聊天记录、发货或付款的凭证以及截图, 一旦出现纠纷后, 这些都可以作为以后私人仲裁机构或法庭裁决的证据。

在线争议解决机制的另一个作用是帮助交易者取消不合理的负面评价。对 eBay 用户来说, 如果认为交易对手给予的信用评价不合理, 他可以求助于 eBay 的“物品未收到或与描述显著不符”处理流程, 或直接提交给 SquareTrade。作为 eBay 认可的在线调解商, SquareTrade 可以帮助交易者取消不合理的中评或差评。因此, 在线争议解决机制降低了信用评价出现噪声的概率 α , 使信用评价更能反映交易者的声誉。(2) 式的右边随 α 的降低而增大, 在线争议解决机制扩大了信用评价系统的作用范围, 增强了网上交易的声誉机制。

由于第三方市场中介发育的不完善, 中国的网上交易还没有引入在线争议解决机制, 交易网站充当了争议解决机构。

（二）第三方托管

网上交易的惯例一般是要求买家先付款，因此买家网上购物面临的最大风险就是付款后收不到货物。第三方托管机构作为受托方起到了代收代付的作用，极大地降低了卖家欺诈的概率。第三方托管的基本流程是，买家在网上把货款汇给第三方托管机构，第三方托管机构收到货款之后通知卖家发货，买家收到货物之后再通知第三方托管机构，第三方托管机构才把钱转到卖家的账户上。第三方托管机构会追踪货物的托运直至发送到买家。www.escrow.com是eBay推荐其用户使用的一家网上第三方托管机构。由于每一笔网上交易第三方托管机构要付出几乎相同的服务，因此第三方托管服务具有规模经济，货物的价值越大，第三方托管机构收取的佣金比例就越低。对价值在5000美元以下的商品，www.escrow.com标准的托管服务收取的佣金率为3.25%，但不低于50美元。

在eBay发展的早期，交易的商品大多是价值较小的各种收藏品、计算机及其配件。第三方托管的出现降低了卖家欺骗的收益 d ，进而降低了欺骗的收益率 $(d - c)/c$ ，使得价值较高（欺骗的收益率也较高）的商品，如汽车也能在网上交易。第三方托管这种有组织的私人秩序很好地补充了网上交易的声誉机制。

（三）商盟

卖家商盟制度是淘宝网的一个创新。淘宝网鼓励卖家组织起来建立商盟，卖家的自我管理可以减少淘宝网付出的成本。目前在淘宝网上有两类商盟，一类是行业商盟，另一类是区域商盟，主要是城市商盟。

商盟可以起到以下作用：众多卖家组织成商盟，尤其是行业商盟可以形成外部效应，从而使商盟中的每一个商家从中获利；小卖家组织成城市商盟可以增加其谈判力量，降低商品配送费用和采购成本；通过加入商盟，小卖家不但可以汲取经验教训，获得商业信息，还可以找到归属感；商盟还可以起到协调的作用，防止成员之间展开恶性的价格竞争。

最重要的是，商盟还可以起到承诺作用。单个卖家通过建立起商盟，并以商盟形成的集体声誉作为抵押（hostage）向买家承诺商盟成员不会有欺骗行为。商盟成员的欺骗行为将损害整个商盟的集体声誉（collective reputation）¹⁷，买家因而可以对整个商盟实施惩罚。通过建立商盟，商盟成员就向买

¹⁷ 集体声誉之所以会出现是因为卖家过去的行为是无法完全观测到的（Tirole 1996, Winfree and McCluskey 2005）。例如，卖家欺骗买家，然后辩解是双方沟通出现差错造成的。eBay和淘宝网提供的卖家的交易者简介中往往可以看到卖家和买家就争议各执一词，有时很难判断究竟是卖家还是买家应该为争议负主要责任。如果单个卖家过去的行为是可完美观测的（perfectly observed），那么卖家就有激励去维护个人声誉（individual reputation），集体声誉就没有必要。

家发出一个明确的承诺：本商盟成员提供的商品都是高质量的；否则，所有的成员都会受到牵连；如果某一成员提供了低质量的商品，商盟将对其实施严厉惩罚。

为了使向买家发出的承诺是可置信的 (credible)，商盟制定了严格的入盟程序和规章制度。另外，与其他第三方执行中介相比，商盟成员之间比较熟悉，因此商盟拥有更多的信息，可以更清晰、以更低成本对争议进行裁决。例如，武汉商盟就承诺：在商品确认签收后的 3 天内，如果商品无损坏情况，商盟包退换。商品有任何问题，买家可以直接向商盟管理部门投诉。

商盟以其集体声誉作为抵押向买家发出可置信承诺，因此，加入商盟的卖家更值得买家信任。

对淘宝网来说，商盟可以增加淘宝社区的凝聚力，吸引更多的卖家到淘宝网上进行交易。商盟可以使卖家自我管理、互相监督来打击欺诈行为。目前，商盟运作已进入良性循环，欺诈行为一旦发生，买家举报后商盟内部会对欺诈者先行予以处罚，然后交给淘宝网管理员处理。商盟从而起到补充信用评价系统的作用，减少了交易网站的付出，增加了网上交易中买家与卖家之间的信任。李维安、吴德胜和徐皓 (2006) 验证了淘宝网上商盟制度所起到的对声誉机制的补充作用。为了支持商盟的发展，淘宝网在促销、推荐上对各商盟进行支持。

鉴于商盟在网上交易中所起的作用，易趣也于 2006 年底成立了具有类似功能的易趣卖家公社。易趣网这样描述成立卖家公社的目的：“易趣卖家公社建立的目的就在于增加卖家队伍的凝聚力以将各方的优秀卖家组织起来，共谋福利，互相帮助、互相监督，互相交流销售经验。帮助 eBay 易趣制定更加符合市场规范的诚信安全政策。”¹⁸

对商盟制度如何起到补充声誉机制的作用还需要做进一步的正式分析。

四、中美网上交易的制度差异

中国的网上交易有许多不同于美国的特点，这一点体现在英文拍卖网站 eBay 与 eBay 控股的中文网站易趣，以及易趣与阿里巴巴旗下的淘宝网之间的差异上。本文认为主要的原因是中国缺乏一个完善的全社会信用体系，仅依靠声誉机制不足以遏制不诚实卖家的欺骗行为。

(一) 支付体系的不同

美国 eBay 上使用的支付工具——PayPal 只是一个第三方支付工具，并不

¹⁸ <http://forums.ebay.com.cn/forum.jspa?forumID=400000038>.

兼有第三方托管的功能，而淘宝网推出的“支付宝”则同时具备了第三方支付和第三方托管的功能。¹⁹

eBay 的 PayPal 线上支付平台在欧美国家使用率非常高。PayPal 的交易流程是：当买家将货款汇到其 PayPal 上的账户后，PayPal 会直接转到卖家的账上，交易速度非常快，也避免了不同银行之间转账的麻烦。如果买家出现货物未收到或与描述不一致，再通过各种争议解决方法将货款追回。在欧美发达国家，由于有完善的信用卡系统作为在线支付的保障，如果用户在网上交易中欺骗了对方，很容易留下不良的信用记录，影响其网下的经济交往和生活，因此卖家收到货款而不发货的概率非常小，PayPal 和买家承担的风险也很小。不过，2005 年 7 月 eBay 将 PayPal 引入到易趣交易平台后，并没有受到交易者的欢迎。在中国，由于不存在一个统一的信用征集机构，交易者可以轻易地改变银行账户，如果没有其他保障机制，卖家收到货款而不发货的概率非常高，这正是 PayPal 在中国无法推行的主要原因。

易趣的竞争对手淘宝网则根据中国的实际情况推出了具备第三方托管功能的交易平台“支付宝”。支付宝的交易流程是：买家将货款汇给支付宝，支付宝通知卖家可以发货；卖家发货，买家签收后通知支付宝货物已收到，支付宝再将货款转移到卖家在支付宝上的账号上。从支付流程上来说类似于 PayPal 的支付模式，不同之处在于支付宝同时具有第三方托管的功能，相当于 eBay 上的 PayPal + escrow.com，即“第三方支付 + 第三方托管”。在中国，网上交易者不相信贝宝，这迫使易趣不得不模仿淘宝网推出了第三方托管工具——安付通，而在美国，eBay 只推荐交易金额较高的商品（1000 美元以上）使用第三方托管机构 escrow.com。由于缺乏一个统一的社会信用体系，普通商品的交易也必须使用第三方托管，而第三方托管服务是有成本的，因此，与美国相比，中国网上交易要支付更多的交易成本。

（二）网站在网上交易中承担的角色不同

淘宝网与 eBay 在网上交易中承担的角色不同。在美国，eBay 只负责提供网上交易场所和支付工具，通过双向信用评价系统提供交易者的信息，吸引更多的用户加入到 eBay 社区中，并不介入双方的交易和争议的解决，交易产生的风险、损失、责任由买卖双方自行承担，其功能类似于 Dixit（2003b）文中的信息中介。在中国，淘宝网、易趣不得不建立起一整套网上交易机制，从提供网上交易场所、支付工具到第三方托管等交易机制，不但充当了信息中介的角色，而且还充当了 Dixit（2003b）文中执行中介的角色。当出现了物品未收到或与描述显著不符时，如果买家和卖家不能就争端达成一致，淘宝

¹⁹ 淘宝网还制定了支付宝交易纠纷处理规则，交易过程中发生的争议一般根据支付宝交易纠纷处理规则来解决。支付宝实际上代替淘宝网起到了第三方执行中介的作用。

网通过其制定的支付宝交易纠纷处理规则来解决争端。在美国,上述执行中介的角色则由其他第三方中介,如 SquareTrade 来充当。淘宝网充当执行中介可能会影响其公正性,也可能会给其带来法律风险。由于承担了过多的市场功能,易趣和淘宝网至今仍未实现盈利。另外,淘宝网还建立了商盟制度,商盟的自我管理可以减轻淘宝网在执行功能上的付出。这一点在上一节已进行了介绍。

(三) 社会信用体系在网上交易中的作用

中国和美国网上交易最大的不同在于美国建立了一套非常成熟的信用体系。美国政府通过立法和监管将信用体系纳入法律范畴,个人征信则以市场化运作为主。信用服务公司——信用局 (credit bureau) 对消费者信用进行评估并提供个人信用服务。信用局主要从银行、信用卡公司、公用事业公司和零售商等渠道了解消费者付款记录的最新信息。交易者在网上交易中的不良记录会影响其在网下的声誉。这样网上交易与网下交易就形成了一个关联博弈 (linkage game), 网下的惩罚机制就加强了网上交易的惩罚力度,降低了交易者的违约诱惑 (Aoki, 2001)。简单地说,卖家选择欺骗虽然得到了很高的暂时收益 d , 但由于其网下的信用有不良记录, 社会信用体系将对其实施一段时间的惩罚 p , 从而加强了对网上交易中不诚实行为的惩罚力度。在美国, 全社会信用体系作为公共秩序形成了网上交易的制度基础设施, 从而减少了网上交易参与者付出的成本。

由于不存在一个完善的全社会信用体系, 中国的网上交易与网下交易被割裂开来, 网上交易中的不良记录对欺骗者网下的交易和生活并不会造成很大的影响。为了保证网上交易的顺利进行, 易趣和淘宝网就必须承担更多的功能, 建立起一系列的交易机制。由于法律和社会信用体系这两种公共秩序的缺乏, 中国的网上交易不得不更多地依赖有组织的私人秩序。因此, 有组织的私人秩序就起到了替代公共秩序的作用。

五、结论、不足及未来的研究方向

网上交易中存在非常严重的囚徒困境问题, 在法律这种公共秩序缺失的情况下, 交易网站以及其他第三方中介建立了信用评价系统、第三方支付、第三方托管、在线争议解决、商盟等机制来减轻网上交易中囚徒困境问题。随着网上交易的迅速发展, 违约以及欺诈发生的概率非常小, 说明上述机制可以很好地发挥作用。

本文首先运用随机匹配博弈模型对网上交易中信用评价系统的作用进行分析。本文的结论是: 建立在信用评价系统上的声誉机制可以很好地发挥作用。本文进一步分析在线争议解决、第三方托管以及商盟等机制的作用, 这

些机制可以起到补充声誉机制的作用。这样，本文就说明了在法律这种公共秩序缺失的情况下，自发的私人秩序和有组织的私人秩序减轻了网上交易中囚徒困境问题，从而起到了替代法律制度的作用。

与美国相比，中国缺乏一个完善的全社会信用体系，网上交易与网下交易被割裂开来，网上交易中的欺骗记录对欺骗者网下的交易和生活并不会造成多大的影响。社会信用体系的缺乏导致中国的交易网站承担了比美国 eBay 更多的角色，建立了一系列 eBay 没有的交易机制。这样，本文也说明了有组织的私人与社会信用体系这种公共秩序之间存在着替代关系。

因此，本文的结论是：在公共秩序缺失的情况下，网上交易中自发的私人秩序和有组织的私人秩序可以很好地发挥作用，起到替代公共秩序的作用。

本文也存在一些不足。本文没有像 Milgrom et al. (1990) 和 Dixit (2003b) 那样建立一个包含第三方中介的博弈模型，而是在声誉模型的基础上对第三方托管、在线争议解决、商盟等第三方中介和社会信用体系的作用进行了简单的定性分析。这主要是因为网上交易内容是如此丰富，第三方中介所起的作用各不相同，以至于很难用一个统一的模型来说明网上交易中各交易机制的作用。Milgrom et al. 对中世纪欧洲集市里法商的作用做了高度的抽象后才建立了一个包含第三方中介的模型。网上交易中，如果要在一个统一的模型中分析第三方中介的作用，势必要做非常抽象的假设，但这会忽略很多有价值的内容。

在线争议解决机制、商盟、第三方托管本身就是一系列有意义的制度，值得单独进一步的分析。例如，在线争议解决机制作为第三方仲裁提高了网上交易契约的完备程度，买家就可以选择在交易前与卖家签订一个比较正式的契约，规定商品的质量。因此，就需要考虑引入在线争议解决机制后正式契约与关系型契约（非正式契约）之间的关系。²⁰ 本文只是笼统地说明了在线争议解决机制的引入可以起到补充信用评价系统的作用，没有对第三方执行中介的作用做详细的分析。对在线争议解决机制、商盟、第三方托管等第三方中介所起的作用以及其与信用评价系统之间的关系进行分析将是下一步的工作。

附录 A 命题 1 的证明

证明：要证明上述策略组合是子博弈完美均衡只需证明卖家在每一期都不会选择欺骗，买家不会偏离上面的集体惩罚措施。

²⁰ 非正式契约或关系型契约与正式契约的关系也是最近契约理论研究的前沿问题，一些理论、实证以及实验文献对非正式契约与正式契约的关系进行了研究 (Brown, Falk and Fehr, 2004; Dixit, 2003a; Lazzarini, Miller, and Zenger, 2004; Pearce and Stacchetti, 1998; Poppo and Zenger, 2002; Schmidt and Schnitzer, 1995)。

在第 t 期, 如果卖家选择诚实, 并在随后各期都选择诚实, 则每期都可以得到收益 c 。如果卖家欺骗了买家, 那么在第 t 期可以获得欺骗的收益 d , 随后的 T 期, 其他买家都观察到这一点, 不再同其进行交易, 惩罚一直持续到第 $t+T$ 期。在第 $t+T+1$ 期, 卖家如果选择诚实, 则随后每期可以得到 c 。定义卖家选择诚实得到的贴现到第 t 期的收益为 V_c , 则 $V_c = c/(1-\delta)$ 。定义卖家选择欺骗得到的贴现到第 t 期的预期收益为 V_d , $V_d = d + \delta^{T+1}c/(1-\delta)$ 。当且仅当 $V_c \geq V_d$, 即下式满足时, 卖家才会选择诚实行为:

$$\frac{d-c}{c} \leq \frac{\delta(1-\delta^T)}{1-\delta}. \quad (1)$$

不等式的右边是关于 δ 的递增函数, 因此存在唯一的 δ^* 使等号成立。在 d 、 c 、 T 和 α 既定的情况下, 当 $\delta \geq \delta^*$ 时, 上式成立。

接下来验证买家不会偏离上面的策略。对买家来说, 如果他观察到一个处于惩罚期的卖家, 因为信用评价没有噪声, 诚实的卖家也不会选择欺骗, 因此他会断定这个卖家是一个机会主义者, 从而不会同该卖家进行交易, 转而去寻找诚实买家。因此, 买家不会偏离上面的集体惩罚策略。

所以, 在 (1) 式满足的情况下, 上述策略组合就是一个子博弈完美均衡。证毕。

附录 B 命题 2 的证明

证明: 要证明上述策略组合是子博弈完美均衡只需证明卖家在每一期都不会选择欺骗, 买家不会偏离集体惩罚。

在第 t 期, 如果卖家选择诚实, 当期可以获得收益 c 。由于信用评价存在噪声, 有 $(1-\alpha)$ 的概率买家认为其选择了诚实行为, 在第 $t+1$ 期, 卖家的收益流与第 t 期相同。有 α 的概率买家认为卖家选择了欺骗, 所有的买家将对其实施时间为 T 的集体惩罚。惩罚持续到第 $t+T$ 期结束, 卖家在 $t+T+1$ 期面临与第 t 期相同的收益流。

卖家选择欺骗的收益流与信用评价不存在噪声时相同。如果卖家欺骗了买家, 那么在第 t 期可以获得收益 d , 随后的 T 期, 其他买家都观察到这一点, 不再同其进行交易。集体惩罚持续到第 $t+T$ 期。卖家随后的收益流与第 t 期相同。

定义卖家选择诚实行为得到的贴现到第 t 期的预期收益为 V_c^n , 则 $V_c^n = c + (1-\alpha)\delta V_c^n + \alpha\delta^{T+1}V_c^n$, 解 V_c^n , 得到

$$V_c^n = \frac{c}{1 - (1-\alpha)\delta - \alpha\delta^{T+1}}.$$

定义卖家选择欺骗得到的贴现到第 t 期的收益为 V_d^n , $V_d^n = d + \delta^{T+1}V_c^n$ 。当且仅当 $V_c^n \geq V_d^n$, 即下式满足时,

$$\frac{d-c}{c} \leq \frac{(1-\alpha)\delta(1-\delta^T)}{1 - (1-\alpha)\delta - \alpha\delta^{T+1}}, \quad (2)$$

卖家才会选择诚实行为。不等式的右边是关于 δ 的递增函数, 因此存在唯一的 δ^{**} 使等号成立。即存在一个 δ^{**} , 在 d 、 c 、 T 和 α 既定的情况下, 当 $\delta \geq \delta^{**}$ 时, 上式成立。

对买家来说, 如果他观察到一个处于惩罚期的卖家, 他会根据贝叶斯推断, 判断该卖

家是机会主义卖家还是诚实卖家。因为诚实卖家只有在信用评价出现噪声时才会被误认为是机会主义者，而 α 是一个很小的概率，在一次交易中，卖家收到了差评，那么买家有足够的理由认为其是一个机会主义者，因此买家不会选择该卖家，而会选择其他最近没有收到差评的卖家。

较为正式的证明如下：

用 O 表示机会主义（Opportunistic）卖家，H 表示诚实（Honest）卖家；D 表示欺骗（Deception）；C 表示合作（Cooperation）；N 表示差评（Negative），P 表示好评（Positive）。 $\Pr(N|O)$ 和 $\Pr(N|H)$ 分别表示机会主义卖家和诚实卖家收到差评的概率， $\Pr(N)$ 为卖家收到差评的概率， $\Pr(O)$ 和 $\Pr(H)$ 分别表示机会主义卖家和诚实卖家所占的比例， $\Pr(O) = 1 - x$ ， $\Pr(H) = x$ 。

卖家收到差评后买家认为其属于机会主义卖家的概率为：

$$\Pr(O|N) = \frac{\Pr(N|O)\text{Prob}(O)}{\Pr(N)} = \frac{\Pr(N|O)(1-x)}{\Pr(N)}.$$

卖家收到差评后买家认为其属于诚实卖家的概率为：

$$\Pr(H|N) = \frac{\Pr(N|H)\Pr(H)}{\Pr(N)} = \frac{\alpha x}{\Pr(N)}.$$

机会主义者收到差评的概率等于买家认为机会主义者会选择欺骗的先验概率 $\Pr(D|O)$ 加上机会主义卖家选择诚实但却收到差评的概率 α ，即 $\Pr(N|O) = \Pr(D|O) + \alpha$ 。

先验概率 $\Pr(D|O)$ ，即机会主义卖家选择欺骗的概率大于零，且 $1 - x > x$ ，这样 $\Pr(O|N) > \Pr(H|N)$ 。如果卖家收到差评，买家断定他是机会主义者，不同他进行交易。这就证明了买家不会偏离集体惩罚机制。

所以，在（2）式满足的情况下，上述策略组合就是一个子博弈完美均衡。证毕。

附录 C 一个卖家的交易简介

下面的两张图是淘宝网上一个卖家的交易者简介。交易者简介包含两部分，一是交易者的信用评价，二是交易者收到的评价的具体内容，图 2 显示的是卖家 angelskysky 收到的“来自买家的评价”的一部分内容。



图 1 交易者的信用评价

收到的所有评价				来自买家的评价	来自卖家的评价	给别人的评价
所有评价	所有好评	所有中评	所有差评			
评价内容				宝贝名称/评价人		
好评	详情	2008.05.11 23:11		中国移动充值卡面值30元的重庆地...	买家：369yishui	49.5元
好评	详情	hao... 2008.05.11 14:04		《东之》网易30点快冲4.75一张可...	买家：玉环章钢锋	4.75元
好评	详情	2008.05.11 01:29		梦幻西游30点的快冲可以寄售也可...	买家：hay5887	4.75元
好评	详情	好卖家，交易迅速，态度又好好！	2008.05.10 23:29	魔兽世界300点快冲12.99	买家：秋露儿	12.99元
好评	详情	2008.05.10 23:18		网易150点直冲，可以寄售！	买家：ppookkop	12.99元

图 2 交易者历史上收到的（部分）评价的内容

参 考 文 献

[1] Aoki , M. , Towards a Comparative Institutional Analysis. Cambridge , MA : MIT Press , 2001.

[2] Bajari , P. , and A. Hortacsu , “ Winner’s Curse , Reserve Prices and Endogenous Entry : Empirical Insights from eBay Auctions ” , Rand Journal of Economics , 2003 , 34 (2) , 329—355.

[3] Baker , G. , R. Gibbons , and K. Murphy , “ Subjective Performance Measures in Optimal Incentive Contracts ” , Quarterly Journal of Economics , 1994 , 109 (4) , 1125—1156.

[4] Baker , G. , R. Gibbons , and K. Murphy , “ Relational Contracts and the Theory of the Firm ” , Quarterly Journal of Economics , 2002 , 117 (1) , 39—83.

[5] Baron , D. , “ Private Ordering on the Internet : The eBay Community of Traders ” , Business and Politics , 2002 , 4 (3) , 245—274.

[6] Bernstein , L. , “ Opting Out of the Legal System : Extralegal Contractual Relations in the Diamond Industry ” , Journal of Legal Studies , 1992 , 21 (1) , 115—121.

[7] Bernstein , L. , “ Private Commercial Law in the Cotton Industry : Creating Cooperation through Rules , Norms , and Institutions ” , Michigan Law Review , 2001 , 99 , 1724—1790.

[8] Bolton , G. , E. Katok , and A. Ockenfels , “ How Effective Are Electronic Reputation Mechanisms ? An Experimental Investigation ” , Management Science , 2004 , 50 (11) , 1587—1602.

[9] Brown , M. , A. Falk , and E. Fehr , “ Relational Contracts and the Nature of Market Interactions ” , Econometrica , 2004 , 72 (3) , 747—780.

[10] Clay , K. , “ Trade without Law : Private-Order Institutions in Mexican California ” , Journal of Law , Economics , and Organizations , 1997 , 13 (1) , 202—231.

[11] Dellarocas , C. , “ The Digitization of Word of Mouth : Promise and Challenges of Online Feedback Mechanisms ” , Management Science , 2003 , 49 (10) , 1407—1424.

[12] Dixit , A. , “ Arbitration and Information ” , Princeton University Working Paper , 2003 (a).

[13] Dixit , A. , “ On Modes of Economic Governance ” , Econometrica , 2003 (b) , 71 (2) , 449—481.

[14] Ellison , G. , “ Cooperation in the Prisoner’s Dilemma with Anonymous Random Matching ” , Review of Economic Studies , 1994 , 61 (3) , 567—588.

[15] Fudenberg , D. , and E. Maskin , “ The Folk Theorem for Repeated Games with Discounting and Incomplete Information ” , Econometrica , 1986 , 54 , 533—554.

- [16] Fudenberg, D., and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991.
- [17] Gambetta, D., *The Sicilian Mafia: the Business of Protection*. Cambridge, MA: Harvard University Press, 1993.
- [18] Greif, A., "Reputation and Coalition in Medieval Trade", *Journal of Economic History*, 1989, 49(4), 857—882.
- [19] Greif, A., "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition", *American Economic Review*, 1993, 83(3), 525—548.
- [20] Houser, D., and J. Wooders, "Reputation in Auctions: Theory, and Evidence from eBay", *Journal of Economics and Management Strategy*, 2006, 15(2), 255—548.
- [21] Johnson, S., J. McMillan, and C. Woodruff, "Courts and Relational Contracts", *Journal of Law, Economics, and Organization*, 2002, 18(1), 221—277.
- [22] Kalyanam, K., and S. McIntyre, "Return on Reputation in Online Auction Markets", Santa Clara University Working Paper, 2001.
- [23] Kandori, M., "Social Norms and Community Enforcement", *Review of Economic Studies*, 1992, 59(1), 63—80.
- [24] Klein, B., and K. Leffler, "The Role of Market Forces in Assuring Contractual Performance", *Journal of Political Economy*, 1981, 89(4), 615—641.
- [25] Kreps, D., P. Milgrom, J. Roberts, and R. Wilson, "Rational Cooperation in the Finitely Repeated Prisoners' Dilemma", *Journal of Economic Theory*, 1982, 27(2), 245—252.
- [26] Lazzarini, S., G. Miller, and T. Zenger, "Order with Some Law: Complementarity versus Substitution of Formal and Informal Arrangements", *Journal of Law, Economics, and Organization*, 2004, 20(2), 261—298.
- [27] 李维安、吴德胜、徐皓, "网上交易中的声誉机制——来自淘宝网的证据", 第六届(2006)经济学年会 武汉大学 2006年12月16—17日。
- [28] Livingston, J., "How Valuable is a Good Reputation? A Sample Selection Model of Internet Auctions", *Review of Economics and Statistics*, 2005, 87(3), 453—465.
- [29] Lucking-Reiley, D., "Auction on the Internet: What's Being Auctioned, and How?" *Journal of Industrial Economics*, 2000, 48(3), 227—253.
- [30] Lucking-Reiley, D., D. Bryan, N. Prasad, and D. Reeves, "Pennies from eBay: the Determinants of Price in Online Auctions", Working Paper, 2006.
- [31] Macaulay, S., "Non Contractual Relations in Business: A Preliminary Study", *American Sociological Review*, 1963, 28(1), 55—67.
- [32] Macneil, I., "Contracts: Adjustments of Long-term Economic Relations under Classical, Neoclassical, and Relational Contract Law", *Northwestern University Law Review*, 1978, 72(192), 854—906.
- [33] Macneil, I., "Relational Contract: What We Do and Do Not Know", *Wisconsin Law Review*, 1985, 10(3), 483—526.
- [34] McDonald, C., and V. Slawson, "Reputation in an Internet Auction Market", *Economics Inquiry*, 2002, 40(3), 633—650.
- [35] McMillan, J., and C. Woodruff, "Private Order under Dysfunctional Public Order", *Michigan Law Review*, 2000, 98(8), 2421—2458.
- [36] Melnik, M., and J. Alm, "Does a Seller's Reputation matter? Evidence from eBay Auctions", *Journal of Industrial Economics*, 2002, 50(3), 337—349.

- [37] Milgrom, P., D. North, and B. Weingast, "The Role of Institutions in the Revival of Trade: The Law Merchant, Private Judges, and the Champagne Fairs", *Economics and Politics*, 1990, 2(1), 1—23.
- [38] Okuno-Fujiwara, M., and A. Postlewaite, "Social Norms and Random Matching Games", *Games and Economic Behavior*, 1995, 9(1), 79—109.
- [39] Pearce, D., and E. Stacchetti, "The Interaction of Implicit and Explicit Contracts in Repeated Agency", *Games and Economic Behavior*, 1998, 23(1), 75—96.
- [40] Poppo L., and Zenger T., "Do Formal Contracts and Relational Governance Function as Substitutes or Complements", *Strategic Management Journal*, 2002, 23(8), 707—725.
- [41] Resnick, P., and R. Zeckhauser, "Trust among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System", in Baye, M. (ed.), *The Economics of the Internet and E-Commerce: Advances in Applied Microeconomics*, Vol. 11. Greenwich, CT: JAI Press, 2002.
- [42] Rubinstein, A., "Equilibrium in Supergames with the Overtaking Criterion", *Journal of Economic Theory*, 1979, 21(1), 1—9.
- [43] Schmidt, K., and M. Schnitzer, "The Interaction of Explicit and Implicit Contracts", *Economic Letters*, 1995, 48(2), 193—199.
- [44] Tirole, J., "A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality)", *Review of Economic Studies*, 1996, 63(1), 1—22.
- [45] Winfree, J., and J. McCluskey, "Collective Reputation and Quality", *American Journal of Agricultural Economics*, 2005, 87(1), 206—213.
- [46] Williamson, O., "Transaction-Cost Economics: The Governance of Contractual Relations", *Journal of Law and Economics*, 1979, 22(2), 233—261.
- [47] 周黎安、张维迎、顾全林、沈懿, "信誉的价值:以网上拍卖交易为例", 《经济研究》2006 年第 12 期, 第 81—91 页。

Private Order in Online Trade: Community, Reputation and Third-party Intermediaries

DESHENG WU
(Nankai University)

Abstract In contrast to off-line trade, online trade reduces information collecting and disseminating costs, and makes sellers and buyers match easier. However, severe prisoners' dilemma exists in online C2C trade. Online auction sites and other third-party intermediaries established online feedback systems, online dispute resolution mechanism and third-party escrow to reduce traders' opportunistic behavior. This paper explores how traders and third-party intermediaries establish

private order to support online trade under Dysfunctional Public Order. These private orders can substitute legal system in online trade.

JEL Classification D44 , D82 , L86

纵向关联市场间的价格传递

王秀清 H. T. Weldegebriel A. J. Rayner^{*}

摘要 本文运用均衡移动模型研究农业生产者与食品零售商之间的纵向价格传递关系。模型建立在如下假设之上：即农产品（食品）营销企业在零售环节拥有寡占卖方力量、在农产品收购环节拥有寡占买方力量、营销企业所需要的农产品投入与其他营销投入之间存在一定替代关系并拥有规模报酬可变的技术。结果表明，市场力量和规模报酬对价格传递的影响十分复杂，一方面取决于农产品供给函数和食品需求函数的具体形式，另一方面还取决于外生冲击作用下农产品收购环节与食品零售环节市场力量变化的相对幅度。

关键词 价格传递，规模报酬，市场力量

一、引言

在纵向关联的市场之间，某一环节的价格变化会通过产业关联传递到其他环节并导致整个产业链的福利再分配。2003年10月下旬中国再次发生农产品价格普遍上涨，甚至又一次引起人们对通货膨胀的担心（张永军，2003；王秀清、钱小平，2004）。媒体、学者和消费者普遍注意到这样一种现象，即农产品及食品零售价格的上涨幅度远远高于农民所实际感受到的农产品收购价格上涨幅度，也就是说出现了所谓的“农产品价格放大效应”（辛贤、谭向勇，2000）。而另一方面，以禽流感 and 疯牛病为代表的食品安全事件直接对消费者的食品需求造成冲击，相关食品零售价格迅速下降并传递到畜产品收购价格，对畜牧业生产带来严重影响。同样，相对于食品零售价格下降而言，畜产品收购价格的下降幅度更大。这种纵向关联市场间价格变动的不对称性源于多种因素的综合作用。例如，从农产品收购环节到食品零售环节价格上涨幅度的放大就可能受到以下因素的影响。第一，食品市场的不完全竞争。在一个不完全竞争的市场上，拥有市场力量的企业常常具备一定能力来操纵市场价格，从而使价格传递偏离完全竞争的轨道。就农业产业链而言，在农产品收购环节农民常常不得不面对买方寡占力量，而在食品零售环节广大消

^{*} 王秀清，中国农业大学经济管理学院，H. T. Weldegebriel, A. J. Rayner, University of Nottingham, Nottingham, UK. 通讯作者及地址：王秀清，北京市海淀区圆明园西路二号院，100094；电话：(010) 62732626；E-mail: wangxq@cau.edu.cn。作者首先感谢两位匿名审稿人提出的宝贵修改意见。美国奥本大学 Henry Kinnucan 为中国农业大学研究生开设的 EDM 专题课使本文中作者初次接触均衡移动模型并获益匪浅，为本文研究奠定了方法论基础。在研究和论文撰写过程中，中国农业大学辛贤提供了文献和数学方法等方面的支持，刘拥军提出了宝贵的修改建议，在此一并致谢。

费者又不得不面对零售商的卖方寡占力量。市场力量的存在与变化直接影响到农业产业链条两端乃至营销企业的福利。第二,农产品流通领域的技术特征。在农产品加工、批发和零售过程中,是否存在规模经济,农产品和其他投入品之间是否可以替代以及替代程度如何,都会影响农产品收购价格发生变化时营销企业的决策安排,从而最终影响零售价格。第三,营销投入品的供给弹性。交通运输业发展的相对滞后常常导致在市场之间难以及时调度农产品。短期内营销投入品供给缺乏弹性也容易导致农产品价格上涨或下降幅度的加剧。第四,价格传递滞后。针对价格变化,营销企业和农户都可能需要一定的反应和调整时间。因此,零售环节发生的价格变化可能要经过一定的滞后期才能传递到农产品收购环节,反之亦然。第五,农产品供给弹性和食品零售需求弹性。针对食品零售价格的变化,农产品收购价格的变化幅度与农产品供给弹性密切相关。反之,针对农产品收购价格的变化,食品零售价格的变化幅度与食品零售需求弹性密切相关。第六,替代品。农产品收购价格变化导致以之为原料的食品零售价格变化。但是零售价格变化的幅度要受到其他食品替代程度以及价格水平的影响。第七,库存调整。食品加工营销企业常常会通过调节库存来适应市场价格的变化以获得最大的利润,因此库存量和库存计划也会对市场间的价格传递产生一定的影响。但是,究竟这些因素如何影响价格传递、影响程度如何,尚未引起国内学术界的足够重视。而从国际学术界研究进展来看,虽然从 70 年代起就有大量文献对价格传递问题进行理论分析或经验检验,但是仍然存在很大的不足,需要进一步的理论探索。

学术界有关价格传递问题的研究分为两大类。第一类是依据时间序列数据进行计量经济检验。这类研究得益于格兰杰因果检验方法和误差修正模型的发展,出现了大量文献。起初,这些方法主要用于检验纵向相关的两个市场之间是否存在整合关系,从而判断市场一体化的程度,甚至用于反垄断实践中的市场界定。近年来又广泛应用于价格传递的对称性检验。但是,由于缺乏理论的框架,该方法并不能很好地解释究竟是什么原因导致价格传递的不对称。第二类着重从理论上推导纵向关联市场间的价格传递关系,确定价格传递弹性的模型。这类研究建立在有关纵向关联市场上企业行为的基本假设之上。本文属于这类研究方法,因此对第一类的文献不做详细的总结与评论。

最早从理论上研究农产品价格传递问题的是美国学者 Gardner (1975)。在这篇开创性文献中, Gardner 采用 Muth (1964) 开发的均衡移动模型 (Equilibrium Displacement Model) 分析了完全竞争市场条件下的农产品价格传递,在营销企业存在规模报酬不变技术的假设前提下分别推导出两个价格传递弹性公式。一个用于反映从农产品收购环节向食品零售环节的价格传递,另一个反映从食品零售环节向农产品收购环节的价格传递。多年来,在该理论模型

影响下，一大批学者运用美国农产品市场的实际数据进行了价格传递弹性模拟。在中国最早研究农产品价格传递问题的是辛贤、谭向勇（2000），在这篇文献中他们以 Gardner 模型为基础分析了中国从生猪收购到猪肉零售市场间的价格传递，发现中国存在明显的农产品价格放大效应。但是，完全竞争假设显然不符合各个国家的农产品市场实际。Holloway（1991）和 McCorriston 等（1998）的研究解除了完全竞争假设，建立了不完全竞争条件下从农产品收购向食品零售的价格传递模型。这些模型假设在食品零售环节存在卖方寡占力量，而在农产品收购环节依然存在完全竞争，食品营销企业依然具有规模报酬不变的技术特征。McCorriston 等（2001）进一步在最新的研究中解除了上述规模报酬不变假设，分析了规模报酬可变情况下的价格传递，解释了规模经济对农产品价格传递的影响，并建立了相应的价格传递弹性公式。把农产品收购环节不完全竞争因素考虑进来的是 Weldegebriel（2004）。Weldegebriel 的研究同时把农产品收购环节存在的不完全竞争特征和食品零售环节存在的不完全竞争特征考虑进来，借鉴 Azzam and Pagoulatos（1990）检验美国肉类加工产业卖方寡占力量和买方寡占力量以及 Kinnucan（2003）确定不完全竞争条件下最优广告密度的分析手段，建立了较为完整的从农产品收购到食品零售的价格传递模型。至此，价格传递的理论模型终于可以通过与完全竞争情况下的模型对比来揭示各个环节市场结构及其变化对价格传递弹性的影响，从而更接近经济运行的实际。但是，在严格的规模报酬不变假设下，该模型还不能够很好地揭示规模经济对价格传递的影响，也不能解释规模经济和市场力量交叉作用对价格传递的影响。另一方面，自 Gardner（1975）在完全竞争市场假设前提下给出从食品零售到农产品收购的价格传递弹性公式以来，一直没有新的文献能够在研究从食品零售到农产品收购的价格传递时考虑不完全竞争和规模经济的影响。本文希望通过解除规模报酬不变假设，来进一步揭示规模经济与市场力量共同作用下从农产品收购到食品零售的价格传递特征以及从食品零售到农产品收购的价格传递特征。本文接下来的安排如下：第二节给出本文的理论模型，分别推导得到从农产品收购到食品零售以及从食品零售到农产品收购的价格传递弹性公式；第三节以农产品收购到食品零售的价格传递为例，根据新的弹性公式集中分析规模经济和市场力量共同作用对价格传递的影响；第四节是本文的结论。

二、模 型

（一）理论框架

本文采用均衡移动分析方法来研究纵向关联市场间的价格传递。基本假设前提如下：在某一农产品的纵向产业链上，众多农业生产者把所生产的农

产品卖给下游的营销企业, 这些营销企业在农产品收购环节具备一定的买方寡占力量。营销企业利用所购买的农产品和其他营销投入品生产出可供消费者直接消费的食品, 并将这些食品在零售市场直接卖给消费者。营销企业生产同质产品、在食品零售环节从事数量竞争并具有一定的卖方寡占力量 (Rogers and Sexton, 1994)。也就是说, 为了分析方便, 我们把从农产品收购到食品零售之间的加工批发环节视为黑箱, 不考虑黑箱内的价格传递。这相当于假设农民和消费者之间是若干个一体化的农产品营销企业。营销企业生产中只有两种投入品, 一种是从农民手中购买的农产品, 另一种是营销投入品, 这两种投入之间存在替代的可能性 (例如在农产品运输时既可以选择铁路也可以选择公路, 相对于农产品与营销投入品的组合比例发生变化)。这里的营销投入品相当于各种营销投入品的组合 (包括加工流通过程中的劳动力、包装和运输等) 或标准化, 假设营销投入品市场是完全竞争的。在着重考察农产品收购和食品零售环节市场力量对价格传递影响时做这样的假设不会影响分析的结论。以上假设跟 Weldegebriel (2004) 的文献完全一致。不同的是, 本文假设在营销企业生产过程中存在着可变的规模报酬技术, 而不仅仅是规模报酬不变。在上述基本假设前提下, 纵向关联市场的初始均衡可以由以下六个公式来表示:

食品 (农产品加工品) 的零售需求函数如 (1) 式所示。式中 R 代表食品零售价格, Q 代表食品的零售需求量, N 代表影响食品零售需求的外生变量 (代表诸如禽流感、疯牛病等能够影响食品零售需求的冲击源)。

$$R = h(Q, N). \quad (1)$$

营销企业生产函数如 (2) 式所示。式中 A 和 M 分别代表农产品和营销投入品的投入量。生产过程中只有两种投入, 而资本投入在短期内固定不变。为了揭示规模报酬的影响, 假设营销企业生产函数为 ρ 次齐次。 $\rho > 1$ 代表规模报酬递增, $\rho = 1$ 代表规模报酬不变, $\rho < 1$ 代表规模报酬递减。

$$Q = f(A, M). \quad (2)$$

农产品和营销投入品的供给函数分别如 (3) 和 (4) 式所示。式中 P 和 W 分别代表农产品和营销投入品的市场价格。 Z 为影响农产品供给的外生变量 (代表诸如自然灾害等能够影响农业生产的供给冲击源)。

$$P = k(A, Z), \quad (3)$$

$$W = g(M). \quad (4)$$

营销部门的农产品需求函数和营销投入品需求函数分别如 (5) 和 (6) 式所示。根据营销企业利润最大化的一阶条件推导出每个企业的要素需求函数, 然后按照各个企业的市场份额加权求和得到营销部门总体的要素需求函数。具体推导过程可以参考 Cowling and Waterson (1976) 和 Bhuyan and Lopez (1997)

等文献。式中 f_A 和 f_M 分别代表农产品和营销投入品的边际产出； η 、 ε 分别代表食品零售需求弹性和农产品供给弹性； θ 和 φ 分别代表食品零售环节和农产品收购环节营销企业的猜测弹性； θ/η 和 φ/ε 则分别代表营销部门在食品零售市场上的卖方寡占力量和农产品收购市场上的买方寡占力量。

$$R\left(1 + \frac{\theta}{\eta}\right)f_A = P\left(1 + \frac{\varphi}{\varepsilon}\right), \quad (5)$$

$$R\left(1 + \frac{\theta}{\eta}\right)f_M = W. \quad (6)$$

(二) 外生冲击下的均衡移动

上述初始均衡市场在外生冲击下会发生一系列反应，从而移动到新的均衡。外生冲击下的均衡移动可以通过对(1)至(6)式求全微分来进行分析。在外生冲击来自农产品供给时(即 $d\ln N=0$, $d\ln Z \neq 0$)，对初始系统求全微分，并用对数形式表示变量的变化率，可以得到以下六个新的方程(而有关外生冲击来自食品需求方面的分析过程放在附录中)：

$$d\ln Q = \eta d\ln R, \quad (7)$$

$$d\ln Q = \alpha d\ln A + \beta d\ln M, \quad (8)$$

$$d\ln P = \frac{1}{\varepsilon} d\ln A + t d\ln Z, \quad (9)$$

$$d\ln W = \gamma d\ln M \quad \text{或} \quad d\ln W = \frac{1}{e_m} d\ln M, \quad (10)$$

$$d\ln P = \frac{1+\mu}{1+\delta} d\ln R + \frac{1}{(1+\delta)} \left[(\rho-1) - \frac{1+\sigma(\rho-1)}{\sigma\rho} \beta \right] d\ln A \\ + \frac{1+\sigma(\rho-1)}{(1+\delta)\sigma\rho} \beta d\ln M, \quad (11)$$

$$d\ln W = (1+\mu) d\ln R + \frac{1+\sigma(\rho-1)}{\sigma\rho} \alpha d\ln A \\ + \left[(\rho-1) - \frac{1+\sigma(\rho-1)}{\sigma\rho} \alpha \right] d\ln M. \quad (12)$$

上述公式中， α 和 β 分别代表农产品和营销投入品的产出弹性($\alpha + \beta = \rho$)； ρ 为规模报酬； ε 代表农产品供给弹性； $\gamma (= 1/e_m)$ 为营销投入品供给弹性的倒数； σ 代表农产品与营销投入间的替代弹性； t 代表农产品供给方面的外生冲击所导致农产品收购价格变化的弹性；而 s 则代表食品零售需求方面的外生冲击导致的食品零售价格变化的弹性； $\mu = -\varpi\theta/(\eta + \theta)$ ，反映食品零售价格因外生冲击而变化时卖方寡占力量的变化； $\varpi = \partial \ln \eta / \partial \ln R$ ，代表食品需求弹性随食品零售价格变化而变化的程度； $\delta = -\lambda\varphi/(\varepsilon + \varphi)$ ，反映农产品收购价格因外生冲击而变化时买方寡占力量的变化； $\lambda = \partial \ln \varepsilon / \partial \ln P$ ，代表农产

品供给弹性随农产品收购价格变化而变化的程度。在市场力量和规模报酬共同作用条件下,农产品和营销投入品的价值份额分别由以下两个式子表示:

$$\psi_A = \frac{(1 + \varphi/\varepsilon)}{\rho(1 + \theta/\eta)} \times S_A = \frac{\alpha}{\rho} \text{ 和 } \psi_M = \frac{1}{\rho(1 + \theta/\eta)} \times S_M = \frac{\beta}{\rho} \quad \psi_A + \psi_M = 1. \text{ 式中 } S_A, S_M$$

分别代表 A 和 M 在完全竞争且规模报酬不变情况下的成本份额。

由农产品供给(Z)或食品需求(N)方面的外生冲击所导致六个内生变量(Q, A, M, R, P, W)的短期变化率可以通过求解上述方程组而获得。把公式(7)、(9)和(11)分别代入公式(8)、(10)和(12)并经过一些变换,可以得到以下三个系统方程:

$$\begin{aligned} & -\eta \mathrm{dln} R + \alpha \varepsilon \mathrm{dln} P + \frac{\beta}{\gamma} \mathrm{dln} W = \alpha \varepsilon \mathrm{dln} Z, \\ & \frac{1 + \mu}{1 + \delta} \mathrm{dln} R + \left[\frac{\sigma \rho (\rho - 1) - [1 + \sigma (\rho - 1)] \beta}{(1 + \delta) \sigma \rho} \varepsilon - 1 \right] \mathrm{dln} P + \frac{[1 + \sigma (\rho - 1)] \beta}{(1 + \delta) \sigma \rho \gamma} \mathrm{dln} W \\ & \quad = \frac{\sigma \rho (\rho - 1) - [1 + \sigma (\rho - 1)] \beta}{(1 + \delta) \sigma \rho} \varepsilon \mathrm{dln} Z, \\ & (1 + \mu) \mathrm{dln} R + \frac{[1 + \sigma (\rho - 1)] \alpha \varepsilon}{\sigma \rho} \mathrm{dln} P + \left[\frac{\sigma \rho (\rho - 1) - [1 + \sigma (\rho - 1)] \alpha}{\sigma \rho \gamma} - 1 \right] \mathrm{dln} W \\ & \quad = \frac{[1 + \sigma (\rho - 1)] \alpha}{\sigma \rho} \varepsilon \mathrm{dln} Z. \end{aligned}$$

运用克莱姆法则 (Cramer's rule) 求解上述联立方程组,可以得到外生冲击所导致的价格零售价格变化率 (dlnR)、农产品收购价格变化率 (dlnP) 和营销投入品价格变化率 (dlnW)。根据这些变化率可以推导得出价格传递弹性。

第一类,因农产品供给方面的外生冲击 (即 $\mathrm{dln} N = 0, \mathrm{dln} Z \neq 0$) 所导致的从农产品收购到食品零售的价格传递,其价格传递弹性定义为 $\tau = (\mathrm{dln} R / \mathrm{dln} P)$ 。具体表达式为:

$$\begin{aligned} \tau^{P \rightarrow R} &= \frac{\alpha \rho (1 + \delta) (1 + \sigma \gamma)}{(\rho + \alpha \sigma \gamma) [(1 + \mu) \rho + \eta (\rho - 1)] - \eta \beta \gamma} \\ &= \frac{\rho \psi_A (1 + \delta) (1 + \sigma \gamma)}{(1 + \psi_A \sigma \gamma) [(1 + \mu) \rho + \eta (\rho - 1)] - (1 - \psi_A) \eta \gamma}. \end{aligned}$$

第二类,因食品需求方面的外生冲击 (即 $\mathrm{dln} Z = 0, \mathrm{dln} N \neq 0$) 所导致的从食品零售到农产品收购的价格传递,其价格传递弹性定义为 $\tau = (\mathrm{dln} P / \mathrm{dln} R)$,具体表达式为 (分析过程见附录):

$$\begin{aligned} \tau^{R \rightarrow P} &= \frac{\rho (1 + \mu) (1 + \sigma \gamma)}{(1 + \delta) [(\alpha + \rho \sigma \gamma) - \sigma (\rho - 1) \beta] - \varepsilon [(\rho + \alpha \sigma \gamma) (\rho - 1) - \beta \gamma]} \\ &= \frac{(1 + \mu) (1 + \sigma \gamma)}{(1 + \delta) [(\psi_A + \sigma \gamma) - \sigma (\rho - 1) (1 - \psi_A)] - \varepsilon [(1 + \psi_A \sigma \gamma) (\rho - 1) - (1 - \psi_A) \gamma]}. \end{aligned}$$

三、市场力量和规模报酬对价格传递的影响

显然，价格传递弹性受到市场力量和规模报酬等因素的影响。为了揭示市场力量和规模报酬的作用，我们讨论上述价格传递弹性公式的几种变形。有关讨论主要集中在因农产品供给方面的外生冲击所导致的从农产品收购到食品零售的价格传递，而类似的分析过程也可以用于从食品零售到农产品收购的价格传递。限于篇幅，本文略去了后一种价格传递的讨论。

（一）完全价格传递（即价格传递弹性等于1）

当营销部门具有规模报酬不变的技术特性（ $\rho = 1$ ）、营销投入品具有无限的供给弹性（ $\gamma = 0$ ）、营销过程中农产品的产出弹性达到某一关键值时（ $\alpha = (1 + \mu)/(1 + \delta)$ ），无论是从农产品收购到食品零售的价格传递，还是从食品零售到农产品收购的价格传递，价格传递弹性均等于1。也就是说，在上述条件得以满足情况下，能够出现价格变化在纵向关联市场之间的完全传递。

（二）价格传递被阻止

农产品供给方面的外生冲击导致农产品收购价格变化。如果随着农产品收购价格上升而营销企业买方寡占力量以相同幅度弱化，或者随着农产品收购价格下降而营销企业买方寡占力量以相同幅度增强（即 $\delta = -1$ 或 $\lambda = (\varepsilon + \varphi)/\varphi$ ），那么农产品收购环节的价格变化就不会传递给食品零售环节（即价格传递弹性等于零）。

类似地，食品需求方面的外生冲击导致食品零售价格变化。如果随着食品零售价格上升而营销企业卖方寡占力量以相同幅度弱化，或者随着食品零售价格下降而营销企业卖方寡占力量以相同幅度增强（即 $\mu = -1$ 或 $\varpi = (\eta + \theta)/\theta$ ），那么食品零售环节的价格变化就不会传递给农产品收购环节（即价格传递弹性等于零）。

（三）规模报酬的影响

规模报酬对价格传递弹性的影响，可以通过对规模报酬参数 ρ 求偏导数来考察。从农产品收购到食品零售的价格传递，其弹性对规模报酬的偏导数如下所示：

$$\frac{\partial \tau}{\partial \rho} = \frac{-\psi_A \eta (1 + \delta) (1 + \sigma \gamma) [(1 + \psi_A \sigma \gamma) + (1 - \psi_A) \gamma]}{\{(1 + \psi_A \sigma \gamma) [(1 + \mu) \rho + (\rho - 1) \eta] - (1 - \psi_A) \eta \gamma\}^2}$$

$$\Rightarrow \begin{cases} > 0, & \forall \delta > -1, \\ = 0, & \forall \delta = -1, \\ < 0, & \forall \delta < -1. \end{cases}$$

显然, 规模报酬对价格传递弹性的影响是不确定的。由于分母、 ψ_A 、 γ 和 σ 等参数为正, 而食品需求弹性 η 为负, 因此偏导数的符号主要取决于参数 δ 的大小。也就是说, 规模报酬对价格传递弹性的影响跟农产品收购环节买方寡占力量的变化幅度密切相关。

如果随着农产品收购价格上升而营销企业买方寡占力量以相同幅度弱化, 或者随着农产品收购价格下降而营销企业买方寡占力量以相同幅度增强 (即 $\delta = -1$), 那么规模报酬对价格传递弹性就丝毫没有影响。

如果随着农产品收购价格上升而营销企业买方寡占力量以相对较低的幅度弱化, 或者随着农产品收购价格下降而营销企业买方寡占力量以相对较低的幅度增强 (即 $\delta > -1$), 那么规模报酬对价格传递弹性就会有显著的正影响。随着规模报酬参数值 ρ 的增加, 价格传递弹性 τ 也相应扩大; 反之, 如果随着农产品收购价格上升而营销企业买方寡占力量以相对较大的幅度弱化, 或者随着农产品收购价格下降而营销企业买方寡占力量以相对较大的幅度增强 (即 $\delta < -1$), 那么规模报酬对价格传递弹性就会有显著的负面影响。随着规模报酬参数值 ρ 的增加, 价格传递弹性 τ 则相应减小。因此, 及时监测伴随价格变化而发生的市场力量变化, 对反垄断实践而言至关重要。

(四) 市场力量的影响

解析市场力量对价格传递弹性的影响, 需要一个可供比较的参照系。Gardner (1975) 的文献提供了一个很好的参照系: 纵向关联市场是完全竞争的 (即 $\theta = \varphi = 0$), 营销企业拥有规模报酬不变的技术 (即 $\rho = 1$), 其价格传递弹性公式如下所示:

$$\tau^c = \frac{\alpha(1 + \sigma\gamma)}{1 + \alpha\sigma\gamma - \eta\beta\gamma} = \frac{S_A(1 + \sigma\gamma)}{(1 + S_A\sigma\gamma) - (1 - S_A)\eta\gamma}.$$

用参照系的价格传递弹性公式除以一般公式得到如下比值:

$$\frac{\tau^c}{\tau} = \frac{[\rho(1 + \theta/\eta) + (1 + \varphi/\varepsilon)S_A\sigma\gamma][(1 + \mu)\rho + \eta(\rho - 1)] - (1 - S_A)\eta\gamma}{\rho(1 + \delta)(1 + \varphi/\varepsilon)[(1 + S_A\sigma\gamma) - (1 - S_A)\eta\gamma]}.$$

为了集中分析市场力量的影响, 我们假设营销投入品供给具有无限弹性 (即 $\gamma = 0$, 该假设并不影响分析的结论), 由此可以得到如下简化式:

$$\frac{\tau^c}{\tau} = \frac{(1 + \theta/\eta)[(1 + \mu)\rho + (\rho - 1)\eta]}{(1 + \phi/\varepsilon)(1 + \delta)}.$$

首先分解规模报酬的影响。令 $\theta = \varphi = 0$ (因而 $\mu = \delta = 0$), 我们可以得到

与 McCorrison et al. (2001) 类似的公式：

$$\frac{\tau^c}{\tau} = \rho + \eta(\rho - 1) \Rightarrow \begin{cases} > 1, & \text{若 } \rho > 1, \eta > -1 \text{ 或 } \rho < 1, \eta < -1, \\ = 1, & \text{若 } \rho = 1 \text{ 或 } \eta = -1, \\ < 1, & \text{若 } \rho > 1, \eta < -1 \text{ 或 } \rho < 1, \eta > -1. \end{cases}$$

与完全竞争情况相比，价格传递弹性既可能放大也可能缩小，食品需求弹性和规模报酬起着决定性作用。与完全竞争的情况相比，当食品需求富有弹性时（即 $\eta < -1$ ），规模报酬递增（ $\rho > 1$ ）将会导致价格传递弹性放大，而规模报酬递减（ $\rho < 1$ ）将会导致价格传递弹性缩减。与此相反，当食品需求缺乏弹性时（即 $\eta > -1$ ），规模报酬递增（ $\rho > 1$ ）将会导致价格传递弹性缩减，而规模报酬递减（ $\rho < 1$ ）将会导致价格传递弹性放大。

接下来分解市场力量的影响。令 $\rho = 1$ ，我们可以得到如下公式：

$$\frac{\tau^c}{\tau} = \frac{(1 + \theta/\eta)}{(1 + \varphi/\varepsilon)} \times \frac{(1 + \mu)}{(1 + \delta)} = \frac{1 + (1 - \varpi)\theta/\eta}{1 + (1 - \lambda)\varphi/\varepsilon}.$$

显然，价格传递弹性偏离完全竞争状态的程度取决于两方面因素的共同作用。一是既存的农产品收购环节买方市场力量与食品零售环节卖方市场力量，二是在外生冲击作用下上述两种市场力量的变化程度。在不知道农产品供给和食品需求函数具体形式的情况下，我们很难确定市场力量及其变化究竟导致价格传递弹性放大还是缩减。如果营销企业在与对手竞争过程中把市场力量控制在接近于零的状态，那么价格传递弹性的放大或缩小则主要取决于外生冲击时市场力量的变化。如果收购环节买方寡占力量的增加幅度大于零售环节卖方寡占力量的增加幅度，价格传递弹性将会放大，反之则缩小。为了能够获得一些更直接的信息，我们接下来讨论两种经常被用于计量经济检验的特殊函数形式，即线性函数和固定弹性函数。

如果农产品供给和食品需求都是线性函数，那么上述公式则可以表示如下：

$$\frac{\tau^c}{\tau} = \frac{1 + (1 - \varpi)\theta/\eta}{1 + (1 - \lambda)\varphi/\varepsilon} = \frac{1 + \theta}{1 + \varphi} \Rightarrow \begin{cases} > 1, & \forall \theta > \varphi, \\ = 1, & \forall \theta = \varphi, \\ < 1, & \forall \theta < \varphi. \end{cases}$$

在线性函数下，价格传递弹性偏离完全竞争的程度取决于纵向关联市场间两类猜测弹性大小的对比。当食品零售环节猜测弹性大于农产品收购环节猜测弹性时，价格传递弹性相对完全竞争而缩减；当两种猜测弹性相等时，价格传递弹性保持完全竞争的水平；当农产品收购环节猜测弹性大于食品零售环节猜测弹性时，价格传递弹性相对完全竞争而放大。

如果农产品供给和食品需求都是固定弹性函数，则 $\varpi = \lambda = 0$ ，从而 $\mu = \delta = 0$ 。这等价于在外生冲击作用下两个环节的市场力量依然保持不变。因此，

价格传递弹性偏离完全竞争的程度则主要取决于原有的市场力量对比。其表达式如下：

$$\frac{\tau^c}{\tau} = \frac{1 + (1 - \varpi)\theta/\eta}{1 + (1 - \lambda)\varphi/\varepsilon} = \frac{1 + \theta/\eta}{1 + \varphi/\varepsilon} \leq 1.$$

因为 $\theta/\eta \leq 0$ 且 $\varphi/\varepsilon \geq 0$ ，所以上述比值始终不会大于 1。这意味着在固定弹性函数或者市场力量不发生变化的情况下，农产品收购或食品零售中任何一个环节拥有市场力量或具备不完全竞争的市场特性，那么价格传递弹性都会相对完全竞争而放大。沿着 Weldegebriel (2004) 的分析思路，我们在表 1 给出了线性函数和固定弹性函数四种不同组合的分析结果。

表 1 市场力量对价格传递弹性的影响($\rho=1$ $\eta=0$)

农产品供给函数与食品需求函数特征	与参照系的价格传递弹性比较			
	买方寡占 ($\varphi \neq 0$ $\theta = 0$)	卖方寡占 ($\theta \neq 0$ $\varphi = 0$)	买方寡占和卖方寡占并存 ($\varphi \neq 0$ $\theta \neq 0$)	
线性农产品供给与线性食品需求	+	-	$\theta > \varphi$	-
	+	-	$\theta = \varphi$	0
	+	-	$\theta < \varphi$	+
固定弹性农产品供给与固定弹性食品需求	+	+	$\delta = \mu = 0$	+
固定弹性食品需求与线性农产品供给	+	+	$(1 + \varphi) > (1 + \theta/\eta)$	+
线性食品需求与固定弹性农产品供给	+	-	$\theta > (\varphi/\varepsilon)$	-
	+	-	$\theta = (\varphi/\varepsilon)$	0
	+	-	$\theta < (\varphi/\varepsilon)$	+

注 相对于完全竞争的参照系而言，- 表示弹性缩减，+ 表示弹性放大 ρ 表示弹性不变。

表 1 显示，在规模报酬不变、营销投入供给弹性无限大的情况下，市场力量对价格传递弹性具有显著的影响。如果食品需求具有固定不变的弹性，那么无论是买方寡占、卖方寡占，还是买方寡占与卖方寡占并存，都将导致价格传递弹性相对完全竞争情况而放大。如果食品需求曲线呈线性，那么买方寡占的价格传递弹性相对完全竞争而放大、卖方寡占的价格传递弹性相对完全竞争而缩减，而买方寡占与卖方寡占并存的影响十分复杂。在线性食品需求与线性农产品供给相组合的市场上，食品零售环节的猜测弹性大于（小于）农产品收购环节的猜测弹性将导致价格传递弹性相对完全竞争而缩减（放大），而猜测弹性相等则具有与完全竞争相同的效果。也就是说，在线性食品需求与线性农产品供给相组合的市场上，如果农产品收购环节与食品零售环节的猜测弹性一致，那么两个环节的市场力量相互抵消，从而也能达到完全竞争情况下的价格传递弹性。这意味着即使存在市场力量，也不一定会偏离完全竞争的轨道。反过来，我们也不能因为发现价格传递弹性与完全竞争水平一致而否定市场力量的存在。同理，在线性食品需求与固定弹性农产品供给相组合的市场上，食品零售环节的猜测弹性大于（小于）农产品收购环节的市场力量将导致价格传递弹性相对完全竞争而缩减（放大），而食品零售环节猜测弹性等于农产品收购环节的市场力量则具有与完全竞争相同的

效果。

（五）规模报酬和市场力量的共同影响

规模报酬和市场力量交叉作用的影响十分复杂，在不了解农产品供给和食品需求函数具体形式的情况下，很难给出一般性结论。为此，我们继续考察线性函数和固定弹性函数不同组合情况下市场力量与规模报酬交叉作用对价格传递弹性的影响效果。分析主要集中在买方寡占与卖方寡占并存的情况，结果如表 2 所示。在规模报酬不变、农产品供给和食品需求均呈线性函数情况下，如果农产品收购环节的猜测弹性等于食品零售环节的猜测弹性，或者在规模报酬不变、农产品供给呈固定弹性而食品需求呈线性函数情况下，食品零售环节的猜测弹性恰好等于农产品收购环节的市场力量，那么都可以导致纵向关联市场上两个环节的市场力量作用相抵，从而达到与完全竞争相同的效果。但是，如果规模报酬可变，就无法达到完全竞争的效果。规模报酬递增时，价格传递弹性相对完全竞争参照系而缩减，规模报酬递减时相对完全竞争参照系而放大。表 2 显示，在规模报酬不变情况下，市场力量的作用效果相对容易确定，而规模报酬可变情况下市场力量与规模报酬交叉作用的效果很难确定。从总体趋势来看，规模报酬递增与市场力量相结合容易导致价格传递弹性相对完全竞争而缩减（尽管也存在放大的可能），而规模报酬递减与市场力量相结合容易导致价格传递弹性相对完全竞争而放大（尽管也存在缩减的可能）。

表 2 市场力量和规模报酬对价格传递弹性的影响($\gamma=0$)

农产品供给函数与食品需求函数特征	与参照系的价格传递弹性比较			
	买方寡占和卖方寡占相结合($\varphi \neq 0 \ \theta \neq 0$)			
		$\rho=1$	$\rho>1$	$\rho<1$
线性农产品供给与线性食品需求	$\theta > \varphi$ 且 $\theta \leq -\eta$	-	-	?
	$\theta = \varphi$ 且 $\theta \leq -\eta$	0	-	+
	$\theta < \varphi$ 且 $\theta \leq -\eta$	+	?	+
固定弹性农产品供给与固定弹性食品需求	$\delta = \mu = 0 \ \theta \leq -\eta$	+	?	?
固定弹性食品需求与线性农产品供给	$\theta \leq -\eta$	+	?	?
线性食品需求与固定弹性农产品供给	$\theta > \varphi/\varepsilon$ 且 $\theta \leq -\eta$	-	-	?
	$\theta = \varphi/\varepsilon$ 且 $\theta \leq -\eta$	0	-	+
	$\theta < \varphi/\varepsilon$ 且 $\theta \leq -\eta$	+	?	+

注 相对于完全竞争的参照系而言，- 表示弹性缩减，+ 表示弹性放大 0 表示弹性不变，? 表示结果不确定。

四、结 论

本文研究了市场力量和规模报酬对纵向关联市场间价格传递的影响。所得出的价格传递弹性模型适用于以下市场特征：农产品收购环节存在买方寡

占力量、食品零售环节存在卖方寡占力量、生产同质产品并从事数量竞争的一体化营销企业拥有规模报酬可变的技术、农产品和营销投入之间存在一定替代关系。

研究结果表明,市场力量和规模报酬对价格传递的影响十分复杂,其结果一方面取决于农产品供给函数和食品需求函数的具体形式,另一方面还取决于外生冲击作用下农产品收购环节与食品零售环节市场力量变化的相对幅度。在不了解农产品供给函数和食品需求函数具体形式情况下,很难为竞争政策的制定给出确定性的结论。尽管在大多数情况下,市场力量的存在常常导致价格传递偏离完全竞争的轨道,但在农产品营销企业具有规模报酬不变的技术特征情况下,仍然存在收购环节买方寡占力量与零售环节卖方寡占力量相互抵消从而使价格传递接近完全竞争效果的可能性。我们既不能因为价格传递接近完全竞争的效果而否认市场力量的存在,也不能因为存在市场力量而断然认定价格传递必然偏离完全竞争的效果。

有关农产品供给函数和食品需求函数的四种不同组合分析结果表明,从总体趋势来看,规模报酬递增技术与市场力量相结合容易导致价格传递弹性相对完全竞争而缩减(尽管也存在放大的可能),而规模报酬递减技术与市场力量相结合容易导致价格传递弹性相对完全竞争而放大(尽管也存在缩减的可能)。

附 录

当外生冲击来自食品零售需求时, $d\ln Z = 0$, $d\ln N \neq 0$ 。对初始系统求全微分并用对数形式表示变量的变化率可以得到以下六个新的方程:

$$d\ln Q = \eta d\ln R + s d\ln N, \quad (A1)$$

$$d\ln Q = \alpha d\ln A + \beta d\ln M, \quad (A2)$$

$$d\ln P = \frac{1}{\varepsilon} d\ln A, \quad (A3)$$

$$d\ln W = \gamma d\ln M \quad \text{或} \quad d\ln W = \frac{1}{e_m} d\ln M, \quad (A4)$$

$$d\ln P = \frac{1 + \mu}{1 + \delta} d\ln R + \frac{1}{(1 + \delta)} \left[(\rho - 1) - \frac{1 + \sigma(\rho - 1)}{\sigma\rho} \beta \right] d\ln A \\ + \frac{1 + \sigma(\rho - 1)}{(1 + \delta)\sigma\rho} \beta d\ln M, \quad (A5)$$

$$d\ln W = (1 + \mu) d\ln R + \frac{1 + \sigma(\rho - 1)}{\sigma\rho} \alpha d\ln A + \left[(\rho - 1) - \frac{1 + \sigma(\rho - 1)}{\sigma\rho} \alpha \right] d\ln M. \quad (A6)$$

由食品需求(N)方面的外生冲击所导致六个内生变量(Q, A, M, R, P, W)的短期变化率可以通过求解上述方程组而获得。把(A1)、(A3)和(A5)式分别代入(A2)、(A4)和(A6)式并经过一些变换,可以得到以下三个系统方程:

$$- \eta d\ln R + \alpha \varepsilon d\ln P + \frac{\beta}{\gamma} d\ln W = s d\ln N,$$

$$\frac{1+\mu}{1+\delta} \text{dlnR} + \left[\frac{\sigma\rho(\rho-1) - [1+\sigma(\rho-1)]\beta}{(1+\delta)\sigma\rho} \varepsilon - 1 \right] \text{dlnP} + \frac{[1+\sigma(\rho-1)]\beta}{(1+\delta)\sigma\rho\gamma} \text{dlnW} = 0,$$

$$(1+\mu) \text{dlnR} + \frac{[1+\sigma(\rho-1)]\alpha\varepsilon}{\sigma\rho} \text{dlnP} + \left[\frac{\sigma\rho(\rho-1) - [1+\sigma(\rho-1)]\alpha}{\sigma\rho\gamma} - 1 \right] \text{dlnW} = 0.$$

运用克莱姆法则 (Cramer's rule) 求解上述联立方程组可以得到外生冲击所导致的食物零售价格变化率 (dlnR)、农产品收购价格变化率 (dlnP) 和营销投入品价格变化率 (dlnW)。根据这些变化率可以推导出价格传递弹性公式如下：

$$\tau^{R \rightarrow P} = \frac{\rho(1+\mu)(1+\sigma\gamma)}{(1+\delta)[(\alpha+\rho\sigma\gamma) - \sigma(\rho-1)\beta] - \varepsilon[(\rho+\alpha\sigma\gamma)(\rho-1) - \beta\gamma]}$$

$$= \frac{(1+\mu)(1+\sigma\gamma)}{(1+\delta)[(\psi_A + \sigma\gamma) - \sigma(\rho-1)(1-\psi_A)] - \varepsilon[(1+\psi_A\sigma\gamma)(\rho-1) - (1-\psi_A)\gamma]}$$

参考文献

- [1] Azzam, A., and E. Pagoulatos, "Testing Oligopolistic and Oligopsonistic Behavior: an Application to the US Meat Packing Industry", *Journal of Agricultural Economics*, 1990, 41(3), 362—370.
- [2] Bhuyan, S., and R. Lopez, "Oligopoly Power in the Food and Tobacco Industries", *Journal of Agricultural Economics*, 1997, 79(3), 1035—1043.
- [3] Cowling, K., and M. Waterson, "Price Cost Margins and Market Structure", *Economica*, 1976, 43(171), 267—274.
- [4] Gardner, B., "The Farm-Retail Price Spread in a Competitive Food Industry", *American Journal of Agricultural Economics*, 1975, 57(3), 399—409.
- [5] Goodwin, B., and M. Holt, "Price Transmission and Asymmetric Adjustment in the U. S. Beef Sector", *American Journal of Agricultural Economics*, 1999, 81(3), 630—637.
- [6] Weldegebriel, H., "Imperfect Price Transmission: Is market Power Really to Blame?", *Journal of Agricultural Economics*, 2004, 55(1), 101—114.
- [7] Holloway, G., "The Farm-retail Price Spread in an Imperfectly Competitive Food Industry", *American Journal of Agricultural Economics*, 1991, 73(4), 979—989.
- [8] Kinnucan, H., "Optimal Generic Advertising in an Imperfectly Competitive Food Industry with Variable Proportions", *Agricultural Economics*, 2003, 29(2), 143—158.
- [9] McCorriston, S., C. Morgan, and A. Rayner, "Processing Technology, Market Structure and Price Transmission", *Journal of Agricultural Economics*, 1998, 49, 185—201.
- [10] McCorriston, S., C. Morgan, and A. Rayner, "Price Transmission: The Interaction between Market Power and Returns to Scale", *European Review of Agricultural Economics*, 2001, 28(2), 143—159.
- [11] McCorriston, S., and I. Sheldon, "Trade Policy in Vertically Related Markets", *Oxford Economic Paper*, 1996, 48, 664—672.
- [12] Muth, R., "The Derived Demand Curve for a Productive Factor and the Industry Supply Curve", *Oxford Economic Papers*, 1964, 16(2), 221—234.
- [13] Rogers, R., and R. Sexton, "Assessing the Importance of Oligopsony Power in Agricultural Markets", *American Journal of Agricultural Economics*, 1994, 76(5), 1143—1150.
- [14] 王秀清, 钱小平, "1981—2000 年中国农产品价格上涨的波及效应", 《中国农村经济》2004 年第 2 期, 第 12—15 页。

[15] 辛贤、谭向勇,“农产品价格的放大效应研究”,《中国农村观察》2000年第1期,第52—57页。

[16] 张永军,“粮油价格全面上涨是新一轮通货膨胀的前兆吗?”《上海证券报(网络版)》2003年10月31日。

Price Transmission in Vertically Related Markets

XIUQING WANG

(China Agricultural University)

H. T. Weldegebriel A. J. Rayner

(University of Nottingham)

Abstract This paper models the vertical relation between retailers and suppliers in the food industry whereby retailers exercise the seller's power in their relations with consumers and the buyer's power in their relations with producers. The degree of price transmission, relative to the perfectly competitive benchmark, both from the farm to the retail sector assuming a supply shock and from the retail to the farm sector assuming a demand shock has been evaluated.

JEL Classification L11, Q13, Q10

高考录取机制的博弈分析

聂海峰*

摘要 本文分析了高考招生中考后知分报考录取机制下的志愿填报博弈。完全信息时这个显示偏好博弈只有唯一的纳什均衡结果,均衡是帕累托有效和公平的。但是,真实的偏好并不一定是每个考生的均衡策略,达到均衡结果需要参与人之间的协调。使用文献中的 Gale-Shapley 学生最优机制,真实偏好是学生的优势策略,结果也是帕累托有效、公平的。本文介绍的录取机制,对于高考录取制度的改革提供了一种可行的思路。

关键词 匹配, Gale-Shapley 学生最优机制, 优势策略

一、引言

在当前高考招生制度下,“高分低就”是一个一直困扰考生和招生机构的问题。通过统一考试被录取是学生进入大学的基本方式,招生学校根据考生的志愿和考试分数的高低决定考生是否被录取。但是由于各个学校先录取第一志愿的学生,相同志愿的考生按照分数高低排序后录取,只有当第一志愿录取不满的时候才招收第二志愿的学生。因此,当考生没有被第一志愿学校录取时,他可能因第二志愿学校已经招满学生而不能被录取,即使他的分数高于该校已经录取的学生。如果志愿填写不当,考生不但有丧失分数优先性的风险,甚至可能遭遇不能被任何大学录取的厄运。这种现象被称作“高分低就”或者“高分落榜”。这在高考录取中并不鲜见。据2005年7月26日的《中国青年报》报道,甘肃省招办统计,在当年甘肃省高考第一批次录取过程中,就有八百多名高分考生落榜。

为了避免考生“高分落榜”,招生部门尝试了许多办法。以填报时间为例,越来越多的省份采用知分填报志愿的方式。2004年有10个省市采用知分填报,到了2006年就有14个省市。除此之外,江苏省2005年还同时采用了“平行院校志愿”的录取机制,对“第一志愿优先”的录取机制进行了重大变革。尽管在一定程度上缓解了“高分低就”现象,但是这些尝试都处在分散的试验阶段,缺乏一个完整的理论体系。因而,如果对高考录取制度建立一个模型,这必将有利于我们更好地理解 and 解决“高分低就”和“高分落

* 中山大学岭南学院。通讯地址:广东省广州市新港西路135号中山大学岭南学院,510275;电话:(020)84110652;E-mail: niehf@mail.sysu.edu.cn。作者感谢2005年制度经济学年会上各位年会参与人的评论,也非常感谢陈晓阳和匿名评审人的意见,这些意见使文章增色不少。当然,所有文责由作者承担。

榜”现象。

本文正是提供了这样一个分析框架,从博弈论的角度分析高考录取机制的运作。我们分析了知分填报时“第一志愿优先”的录取机制衍生的考生志愿填报博弈。当考生的偏好是所有考生的共同知识时,这个显示偏好博弈的纳什均衡结果是帕累托有效、公平的,没有“高分低就”现象发生。因而,在知分填报完全信息时,考生间的策略互动可以消除“高分低就”现象。这个结果在一定程度上支持各省逐渐采用考后出分填报方式的合理性。

虽然在均衡中考生可以避免“高分低就”,但是考生的真实偏好可能不是考生的均衡策略,寻求最优策略和协调到均衡并不是一件容易的事情。因而,人们不禁要问:对于高考录取是否存在一个录取机制,使得录取的结果是帕累托有效、公平的,并且考生在填报志愿时填报自己对学校的真实偏好是他的优势策略?Gale-Shapley 学生最优机制就满足这些性质。

对于高考录取问题,在 Gale-Shapley 学生最优机制下,真实显示偏好是每个学生的优势策略,不需要和其他人的策略进行协调。考生真实显示偏好时得到的结果也正是高考录取机制下考生博弈的均衡结果。同时, Gale-Shapley 学生最优机制也尊重考生的分数进步,如果考生的高考成绩提高,录取他的大学在他的偏好次序上不会下降。在这个机制下,不论填报志愿是在考前、知分还是估计分数填报,真实偏好都是学生的均衡策略。Gale-Shapley 学生最优机制使得考生无需费心进行策略操纵,也避免了“高分低就”的损失。

钟笑寒、程娜和何云帆(2004)最早对中国高考录取机制进行研究,讨论了完全信息时分数不确定时不同填报志愿方式对于不同能力考生录取结果的影响。他们的分析表明,高考的录取结果对填报时间是敏感的,分数不确定对于低能力考生的填报有遏制作用。聂海峰(2006)讨论了填报志愿时间不同时高考录取机制产生“高分低就”的原因,与知分填报时均衡中没有“高分低就”现象发生不同,在考前填报时会出现没有纯策略均衡的情况,“高分低就现象”不可避免。聂海峰(2006)中的例子也表明,当考生的偏好是考生的私人信息时,知分填报时也存在均衡时发生“高分低就”的可能。

对招生录取机制的研究属于文献中的大学招生模型,这类文献滥觞于 Gale and Shapley (1962) 的文章。和高考录取中只有学生是积极参与人不同,在 Gale-Shapley 大学招生模型中,学生和学校都是积极的参与人,每个学生对所有大学有一个偏好排序,每个大学也对所有学生有偏好排序。他们证明,对于学校和考生任意的偏好,都存在学生和大学的组合,使得录取的结果是稳定的,并且提出了达到稳定匹配的 Gale-Shapley 机制。此后, Gale-Shapley 机制得到了广泛的研究,在 Roth and Sotomayor (1990) 中有全面的总结。

Gale-Shapley 机制作为录取分配机制在最近的文献中得到越来越多的研究,并且在实践中已有广泛的应用。土耳其也是通过标准化考试来进行大学录取的。但和中国不同,它通过一个类似中央招生办公室的机构,对全体学

生按照志愿和报考专业需要的单科成绩排名来统一录取。Balinski and Sönmez (1999) 研究了土耳其的大学招生制度, 指出用 Gale-Shapley 学生最优机制可以更好地录取学生。Abdulkadiroglu and Tayfun Sönmez (2003) 首次从机制设计的角度研究了美国公立学校的择校制度。在择校问题中, 学校所在地的法律使得所有学生在每个学校中都有一个优先顺序, 不同的学校中学生的排序不同, 学生的志愿顺序和在学校的优先顺序决定他是否被录取。他们研究的“波士顿机制”和高考的录取机制类似, 也有因第一志愿不当而丧失排序优势的风险。他们提出 Gale-Shapley 机制是一个改进的替代机制。Sönmez and Ergin (2006) 研究了“波士顿机制”下的各种博弈。Sönmez and Chen (2006) 利用试验经济学的方法分析了“波士顿机制”, 发现这个机制下学生会操纵自己申报的偏好, 与 Gale-Shapley 学生最优机制相比有较大的效率损失。在 Roth 等博弈论专家的咨询协助下, 波士顿和纽约的公立学校择校机制改革为 Gale-Shapley 学生最优机制 (Atila Abdulkadiroglu, Parag A. Pathak and Alvin E. Roth, 2005; Atila Abdulkadiroglu, Parag A. Pathak, Alvin E. Roth and Tayfun Sönmez, 2005, 2006)。除了择校机制外, 在真实的世界中, Gale-Shapley 学生最优机制也被应用于美国和英国住院医师和医院的匹配 (Roth, 1991; Roth and Peranson, 1997, 1999)。Roth (2002) 以 Gale-Shapley 机制的应用为例描述了在市场设计时的问题和方法。

本文其他部分的组织如下: 第二部分是对中国高考招生问题的描述; 第三部分刻画了高考博弈均衡结果的性质; 第四部分论述了 Gale-Shapley 学生最优机制和这个机制在高考录取中的性质; 最后是结论。

二、高考招生问题

(一) 高考录取机制

高考录取的基本要素是学生的考分、学生的志愿和学校的招生计划。志愿是学生向招生负责机构申报的对大学的偏好。在不同的省份, 志愿递交的时间和全国统一招生考试的时间并不一致。统一招生考试由法定统一时间进行, 志愿递交时间各省自行安排。目前共有三种制度安排: 考前填报, 这时考生没有参加招生考试; 估分填报, 在招生考试结束后的 1—2 日内填报; 知分填报, 每个考生知道自己的分数和所有考生的分数分布。在 2006 年, 北京、上海、吉林、广东等 4 个省(市) 为考前填报, 浙江、江苏、四川等 14 个省市是知分填报, 其他的地区是估分填报。除了江苏省外, 其他知分填报的省份的录取机制都保留着“第一志愿优先”的特点, 我们把这种“第一志愿优先”的机制称为高考录取机制。这其中以浙江省的录取机制最为典型, 因此, 在本文中的高考录取机制就是浙江省采用的录取机制。

高考录取是由省招生办公室和学校共同完成的。学校制定并公布招生计划,学生报名参加统一招生考试,分数公布后向省招生办公室填报自己的志愿。¹ 根据全部考生的分数和全部学校的招生计划人数,各省招办划定最低录取分数线。只有分数高于最低录取分数的考生才能参与录取。招生办公室根据考生的志愿向学校投递考生的档案,学校根据招生计划和考生的考分排名录取。这个录取机制可以描述如下:

第一轮 考虑所有考生的第一志愿。根据考生的志愿,招生办公室将考生分配到相应的学校。各个学校根据考生的分数排序,按照分数从高到低录取。如果考生数超过招生计划数,则只录取招生计划的人数,其他考生被拒绝。如果考生数小于招生计划数,所有考生都被录取。

第二轮 接下来考虑所有没有被录取的考生的第二志愿。招生办公室根据考生第二志愿把考生送到相应的学校。如果考生该志愿的学校已经招满人数,则该考生被拒绝;若学校仍有空余名额,根据第二志愿考生的成绩排序,从高到低一次一个录取,直到用完所有名额或者所有考生都被招收。

第三轮 考虑所有还没有被录取的考生。在这一轮,把所有没有被录取的考生按照分数从高到低排序,依次考虑每个学生的志愿。首先对于分数最高的学生,按照他的志愿顺序从高到低依次考虑,如果这个学校还有空余名额,则考生被录取,同时学校的招生名额减少一个,如果学校此时恰好用完了招生名额,则停止招生;如果志愿学校没有空余名额,就考虑考生的下一个志愿学校;如果考生所有的志愿都被考虑,所有志愿学校都已经停止招生,则这个考生落榜。然后处理排在其后面的考生的志愿。这样顺次下去直到所有考生都被考虑或者所有招生学校的名额都用完,录取程序停止。

这里描述的录取机制和实际执行的录取机制相比,我们没有区分学校志愿和(学校,专业)志愿。在实际填报志愿中,考生的志愿包括学校排序和学校内的专业排序,并且有是否服从学校内专业调剂的选项。这种志愿的填报方式,使得考生对学校 and 专业的志愿成为字典序形式的偏好。在实际操作中,由招生办公室执行学校的录取,而由学校来执行专业的录取。当学校专业只录取有志愿的学生时,专业可以看作是一个独立的学校;当学校按照分数高低排序后分配专业招收的时候,学校志愿和(学校,专业)志愿的执行是一致的。因此我们不区分学校和(学校,专业),统一称为学校。

考生的分数虽然决定了他在一所学校中的排序,但在高考录取机制下这个排序是内生的,是由考生的分数和他填报的志愿共同决定的,考生可以通

¹ 在实际申报的志愿中,只允许包含有限数目的学校,这个限制相对于参与录取的学校来说是一个非常巨大的限制,这种数目限制也可能会引起无效率。在本文的分析中假设不存在这个数目限制。随着计算机和网络填报方式的推广,这个数目限制可以轻易地取消。另一方面由于交易成本的存在,考生偏好中能包含的学校数目可能也是有限的,但是具体的数目可能是一个经验的问题。

过操纵他的志愿来操纵他的排序。因而，考生填写的志愿不一定反映他对大学真实的偏好，而只是其分数有优势的学校。如何首选第一志愿是每个考生及其家庭都极为关注的事情，整个志愿填报过程成了他们寻求最优操纵策略的博弈过程。

(二) 高考录取模型

为了分析高考录取机制下的志愿填报博弈，我们引入一些符号。一个高考招生问题由以下几部分构成：

(1) 所有考生的集合 $S = \{s_1, s_2, \dots, s_n\}$ 。

(2) 所有学校的集合 $C = \{c_1, c_2, \dots, c_m\}$ 。

(3) 一个招生计划人数向量 $q = (q_{c_1}, q_{c_2}, \dots, q_{c_m})$ ，其中 q_{c_i} 是学校 c_i 的计划招生人数。

(4) 每个学生高考录取分数的向量 $f = (f_{s_1}, f_{s_2}, \dots, f_{s_n})$ ，其中 f_{s_i} 是学生 s_i 的高考分数和规定的加分后的录取分数。我们假设考生的考分没有相同的²，对任意的 $s, \bar{s} \in S$ ，都有 $f_s \neq f_{\bar{s}}$ ，所有的考生的录取分数就得到一个完全的排序。

(5) 学生的严格偏好集合 $P_S = \{P_{s_1}, P_{s_2}, \dots, P_{s_n}\}$ 。每一个学生 s_i 都在集合 $C \cup \{c_0\}$ 上有一个严格的偏好 P_{s_i} ，这里 c_0 表示不上学的选择，我们规定 $q_{c_0} = |S|$ ，表示不上大学的选择可以容纳所有的学生。我们用集合 R 表示 $C \cup \{c_0\}$ 上所有严格偏好序的集合。对任何 $s_i \in S$ ， R_{s_i} 是由 P_{s_i} 得到的非严格偏好，即对于 $c, \bar{c} \in C \cup \{c_0\}$ ， $c R_{s_i} \bar{c}$ 当且仅当 $c P_{s_i} \bar{c}$ 或者 $c = \bar{c}$ 。如果 $c R_{s_i} c_0$ ，学校 c 对于学生 s_i 是可接受的。这样，如果考生 s_i 的偏好为 $P_{s_i} : c_1 - c_0$ ，表示学生只选择学校 c_1 ，其他学校都不可接受，我们省略了学生不可接受学校的排序。对所有的 $s \in S$ ，我们用 $P_{-s} = \{P_{\bar{s}}\}_{\bar{s} \in S \setminus \{s\}}$ 表示除去学生 s 外其他所有学生的偏好组。

由于不能使每个学生都被自己最偏好的大学录取，因此高考录取制度根据学生的偏好和考试成绩选择能够进入大学的学生，把学校入学资格分配给学生。我们把录取结果称为匹配。一个匹配 (match) 就是把招生学校分配到所有考生的一种方案，方案中每一个考生最多能够得到一个大学的入学名额，每个大学招收的人数不超过它的招生计划。形式上，一个匹配就是一个函数 $\mu : S \rightarrow C \cup \{c_0\}$ ，满足对所有的 $c \in C \cup \{c_0\}$ ，有 $|\mu^{-1}(c)| \leq q_c$ ，即每个大学招生不超过它的计划。如果 $\mu(s) = c_0$ ，就表示学生 s 没有被任何大学录取。我们把所有匹配的集合记为 M 。学生 s 的偏好 P_s 是定义在 $C \cup \{c_0\}$ ，我们可以把它扩展到匹配的集合 M 上：学生 s 偏好匹配 μ 超过匹配 η ，当且仅当 $\mu(s) P_s \eta(s)$ 。在下文中，我们仍用 P_s 表示这个扩展的偏好。

接着我们定义一些匹配可能满足的性质。对于所有的 $s \in S$ ，都有 $\mu(s)$

² 在实际中考生的总分相同时，可以使用一定的规则再排序，如单科成绩优先等方式。

$R_s c_0$, 则匹配 μ 是个人理性的。也就是说, 个人理性的匹配中所有的学生得到的学校都是可接受的。如果对任意的 $s \in S$ 和 $c \in C$, 若 $c P_s \mu(s)$ 就有 $|\mu^{-1}(c)| = q_c$, 匹配 μ 就被称为是没有浪费的。在没有浪费的匹配中, 如果学生更偏好的学校没有录取他, 那么这所学校必定已经招满了学生。对于两个匹配 μ 和 η , 如果所有的学生都不偏好 μ 甚于 η , 而有些学生偏好 η 甚于 μ , 则 η 帕累托优于 μ 。用符号来表示, η 帕累托优于 μ , 就是对所有的 $s \in S$, 都有 $\eta(s) R_s \mu(s)$, 且存在 $s_i \in S$ $\eta(s_i) P_{s_i} \mu(s_i)$ 。如果不存在匹配帕累托优于匹配 η , 匹配 η 就是帕累托有效的。可以看到, 一个匹配是帕累托有效的, 必然满足个人理性并且没有浪费。

在高考录取中, 分数高低决定了学生在学校录取时的排序, 是一个公平的指标, 我们自然希望录取的结果有一种分数面前的公平性。Balinski and Sönmez (1999) 分析土耳其的招生制度时, 引入了公平的概念:

匹配 μ 被称为是公平的 (fair), 如果对任意的 $s \in S$ 和 $\hat{c} \in C \cup \{c_0\}$, 若 $\hat{c} P_s \mu(s)$ 必然有 $|\mu^{-1}(\hat{c})| = q_{\hat{c}}$, 且对任意的 $\bar{s} \in \mu^{-1}(\hat{c})$, 有 $f_{\bar{s}} > f_s$ 。

在公平的匹配中, 一个学生不能被他喜欢的学校录取, 是由于这个学校用完了招生名额, 并且已经录取的学生的成绩都高于这个学生, 这个性质反映了学生在分数面前的公平性。因而, 公平的匹配就是没有“高分低就”现象发生的匹配。

在高考招生问题中, 对于录取结果的公平性和帕累托有效性之间的关系, 我们有如下的定理, 定理的证明放在了附录中。

定理 1 一个匹配若满足公平性、个人理性, 必然是帕累托有效的匹配。

高考录取的目的就是给定五元组 $\{S, C, q, f, P_s\}$, 通过统一的录取程序达到学生和学校的一个匹配。在下面的分析中, $\{S, C, q, f\}$ 都是固定的, 因而给定一个学生的偏好组 P_s , 就得到了大学招生的基本要素, 我们称其为一个高考招生问题。我们把所有学生偏好组 P_s 的可能集合记为 P 。一个录取机制 (mechanism) 就是一个系统的程序 $T: P \rightarrow M$, 对于每个高考招生问题 $P_s \in P$, 得到一个学生和学校的匹配。如果一个机制对每个高考招生问题给出的匹配都是帕累托有效的匹配, 这个录取机制就是帕累托有效的机制。如果它对每个高考招生问题给出的匹配都是公平的匹配, 这个录取机制是公平的机制。如果一个录取机制需要每个学生报告他对学校的偏好, 这个机制就是直接机制。高考录取机制就是一个直接机制, 我们在下面将要看到, 这个机制不是公平的机制。但是当考生操纵自己的偏好时, 这个机制下显示偏好博弈的纳什均衡总是公平的匹配。

三、高考志愿填报博弈

在这一节里, 我们将分析高考录取机制下的考生志愿填报博弈。由于高

考录取机制是一个直接机制,给定考生真实偏好 P_s , 每个考生的策略空间 M_{s_i} 就是 $C \cup \{c_0\}$ 上的所有严格偏好的集合 R^3 。给定每个考生 s 的一个策略 $Q_s \in R$, 就得到全体考生的策略组合 $Q_s = \{Q_{s_1}, Q_{s_2}, \dots, Q_{s_n}\}$, 则 Q_s 的全体集合为 $P = \prod_{i=1}^n M_{s_i}$ 。允许混合策略存在时, 每个考生 s 的混合策略是 $\Delta(M_s)$, 表示 $C \cup \{c_0\}$ 上的所有严格偏好集合上的概率分布, 这时全体考生策略集合为 $M = \prod_{i=1}^n \Delta(M_{s_i})$ 。在这一节的分析中假设 P_s 是所有考生的共同知识, T 表示高考录取机制。在知分填报时, 学校招生计划和考生分数都已经确定, 是所有考生的共同知识, 高考志愿填报博弈就是一个显示偏好博弈 $\{S, (M, T), P_s\}$ 。

这个博弈的顺序如下: 首先, 学校招生计划和考生的考试分数确定, 考生进入填报志愿阶段, 全部考生同时行动。每个考生 s 向招生办公室报告他的志愿 $Q_s \in R$, Q_s 不一定是他的真实偏好 P_s , 所有人的志愿组合 $Q_s = \{Q_{s_1}, Q_{s_2}, \dots, Q_{s_n}\}$ 。接着, 根据考生填报的志愿 Q_s , 招生办公室和学校根据 $\{S, C, q, f, Q_s\}$ 按照高考录取机制进行录取, 得到最后的匹配 $T[Q_s]$ 。

我们定义这个博弈的一个纯策略纳什均衡为一个志愿组合 $Q_s^* = \{Q_{s_1}^*, Q_{s_2}^*, \dots, Q_{s_n}^*\}$, 对所有的 $s_i \in S$, 对任意的 $Q_{s_i} \in R$, 都有 $T[Q_s^*](s_i) R_{s_i} T[Q_{s_i}, Q_{s_{-s_i}}^*](s_i)$ 。这里 R_{s_i} 是学生 s_i 的真实偏好, $T[Q_s^*](s_i)$ 表示学生 s_i 在志愿组 Q_s^* 时得到的录取学校, $T[Q_{s_i}, Q_{s_{-s_i}}^*](s_i)$ 是给定其他人的志愿组合 $Q_{s_{-s_i}}^*$, 学生 s_i 填报 Q_{s_i} 时按照高考录取机制得到的录取学校。

在高考录取这个显示偏好博弈中, 每个人都寻求最优的策略, 努力使他的高考分数发挥最大的作用。在高考的实际招生中, 我们看到会发生“高分低就”的现象, 是否在这个博弈中没有均衡结果导致了这些现象。我们有如下结果:

定理 2 在高考志愿填报博弈 $\{S, (M, T), P_s\}$ 中存在纳什均衡。

证明 由于每个考生的纯策略是对所有学校和不上学这个选择的一个严格偏好, 由于学校的数目是有限的, 因而严格偏好的种类也是有限的。同时, 考生的数目是有限的。因此, 这是一个有限博弈。根据纳什定理, 任何有限博弈都存在均衡 (可能是混合策略均衡)。

虽然这个博弈存在均衡, 但是我们更加关心是否存在纯策略的均衡, 混合策略的均衡不是一个很好的结果。在下一节我们将证明, 确实存在纯策略的纳什均衡。我们先用下面的例子说明浙江机制录取的过程和这个机制下的

³ 这意味着学生申报的志愿不受限制, 可以包含任意数目的可接受的学校。考生在申报志愿时不需要申报所有学校的一个完全排序, 只需申报他可接受的学校即可, 考生的志愿就是他申报的可接受学校的排序。

显示偏好博弈。

例 1 假设有三个考生(i_1, i_2, i_3)和两所学校(c_1, c_2) c_0 表示不上学。每个学校的招生名额 $q_{c_i} = 1$, 考生的成绩和真实偏好如下:

$$\begin{array}{lll} i_1 & 90 & c_1 - c_2 - c_0 \\ i_2 & 85 & c_1 - c_2 - c_0 \\ i_3 & 80 & c_2 - c_0 - c_1 \end{array}$$

如果考生都真实地申报自己的志愿, 在高考机制下的录取结果是:

第一轮, 考虑所有考生的第一志愿, 考生 i_1, i_2 投档到学校 c_1 , 考生 i_3 被投档到学校 c_2 。学校按照考分从低到高录取。学校 c_1 录取考生 i_1 , 考生 i_2 被拒绝; 考生 i_3 被学校 c_2 录取。两所学校都用完了招生名额。

第二轮, 考虑没有被录取的考生的第二志愿。但是所有的学校都已经录取满额, 录取结束, i_2 高考落榜。

这样, 考生真实申报志愿的录取结果是 $\mu_1 = \begin{bmatrix} i_1 & i_2 & i_3 \\ c_1 & c_0 & c_2 \end{bmatrix}$ 。给定考生的真实偏好, 我们可以看到匹配 μ_1 不是公平的, 考生 i_2 偏好学校 c_2 , 他的分数也比 c_2 录取的考生 i_3 高。

这表明高考录取机制不是公平的机制, 考生真实申报志愿有可能“高分低就”。

在这个例子中, 真实申报“高分落榜”使得 i_2 有动机操纵他申报的志愿。当其他人申报真实志愿时, i_2 操纵他的偏好填报 $c_2 - c_1 - c_0$, 这时录取的结果

是 $\begin{bmatrix} i_1 & i_2 & i_3 \\ c_1 & c_2 & c_0 \end{bmatrix}$ 。可以验证, 志愿组合 $\begin{bmatrix} i_1 & x_1 - c_2 - c_0 \\ i_2 & x_2 - c_1 - c_0 \\ i_3 & x_3 - c_0 - c_1 \end{bmatrix}$ 是一个纳什均衡。

在这个简单的高考博弈中, 所有的纯策略纳什均衡为志愿组合

$\begin{bmatrix} i_1 & x_1 - c_2 - c_0 \\ i_2 & x_2 - c_1 - c_0 \\ i_3 & x_{i1} - c_{i2} - c_{i3} \end{bmatrix}$, 其中 i_3 的志愿 $c_{i1} - c_{i2} - c_{i3}$ 是 (c_1, c_2, c_0) 的一个任意排列。所

有这些纳什均衡的结果都是 $\begin{bmatrix} i_1 & i_2 & i_3 \\ c_1 & c_2 & c_0 \end{bmatrix}$, 这个结果是公平的匹配。

上面的例子表明, 在知分报考的填报博弈中, 博弈的纯策略均衡结果是唯一的, 并且是公平的匹配, 这个特点在一般的情形下也是成立的。

定理 3 在高考博弈 $\{S, (M, T), P_S\}$ 中, 纯策略纳什均衡结果是唯一的: 令 $A = \{\mu | \mu = T[Q^*], Q^* \in M, Q^*$ 是纯策略纳什均衡 $\}$, 则 $|A| = 1$ 。

有关定理 3 的证明详见附录, 这里叙述一下证明的主要思路: 首先, 作

为一个录取博弈的纳什均衡，均衡结果是个人理性、没有浪费和公平的。其次，我们用归谬法证明，公平性使得均衡只有一个纯策略的纳什均衡结果。这个定理的证明也表明，在知分填报的显示偏好博弈中，博弈的纯策略均衡结果是唯一的，是公平的。

我们前面的例子表明真实偏好可能并不是均衡策略，虽然均衡结果是唯一的，但是可能存在多个均衡。因此在这个博弈中达到均衡结果需要考生间的协调。每年各省参加高考人数有几十万，整个高考志愿填报就成了一个成千上万人参加的巨大博弈。对每个家长来说，如何寻找自己的最优策略，并不是一件轻而易举的事情。在实际报考中也存在协调的努力，各学校、各市都举办考前志愿预报来协调本地区考生之间的志愿，避免报考志愿的过度集中。这些协调的规模与实际招生的规模相比很小，每个考生也仍然可能很难调整到均衡的策略。是否存在一种招生机制，使得每个学生的真实偏好是他的均衡策略呢？

仍然以例1中的情形为例，我们介绍下面的录取程序，在所有考生真实申报自己的偏好的情况时，可以达到高考录取机制下显示偏好博弈的均衡结果。

第一轮 所有考生向可接受的学校中最偏好的学校提出申请，考生 i_1, i_2 投档到学校 c_1 ，考生 i_3 投档到学校 c_2 。学校按照考分从高到低录取。考生 i_1 被学校 c_1 预录取，考生 i_2 被拒绝；考生 i_3 被学校 c_2 预录取。两所学校都用完了招生名额。

第二轮 考虑上一轮被拒绝的考生，向没有拒绝过他的可接受的学校中最偏好的学校提出申请。考生 i_2 向学校 c_2 申请。学校 c_2 考虑新的考生 i_2 和上一轮预录取的考生 i_3 ，考生 i_3 被拒绝， i_2 被学校 c_2 预录取。学校 c_1 本轮没有申请的考生。

第三轮 考虑上一轮被拒绝的考生。考生 i_3 没有可接受的学校，停止申请。本轮没有学校拒绝申请，学校录取程序结束。这时，在学校预录取名单上的考生被正式录取，没有被任何学校录取的考生落榜。最后的录取结果

是：
$$\begin{bmatrix} i_1 & i_2 & i_3 \\ c_1 & c_2 & c_0 \end{bmatrix}。$$

在所有考生真实显示偏好的情况下，这个录取程序达到了和高考录取机制博弈均衡时同样的招生结果。这个机制和高考机制的不同之处，在于每一轮学校的录取都不是最终录取，这使得考生申报偏好时不用担心丧失分数的优势。这个机制正是 Gale-Shapley 学生最优机制，我们在下面将说明在这个机制下每个考生真实显示自己的偏好是他的优势策略。

四、Gale-Shapley 学生最优机制

Gale-Shapley 学生最优机制最早是在大学招生模型中提出来的 (Gale and Shapley, 1962)。大学招生模型由 $\{S, C, q, P_S, P_C\}$ 组成, 包括一组学生的集合 $S = \{s_1, s_2, \dots, s_n\}$ 和一组学校的集合 $C = \{c_1, c_2, \dots, c_m\}$, 每个学校有固定的招生人数向量 $q = (q_{c_1}, q_{c_2}, \dots, q_{c_m})$ 。每一个学生 s 对所有学校 C 和不上学的选择 c_0 有一个严格的偏好 P_s , 全体学生的偏好集合 $P_S = \{P_{s_1}, P_{s_2}, \dots, P_{s_n}\}$, 同时, 每个学校 c 也有一个对全体学生 S 的严格偏好 P_c ⁴, 全部学校的偏好集合 $P_C = \{P_{c_1}, P_{c_2}, \dots, P_{c_m}\}$ 。对于任何 $c \in C$, 我们用 $P_{-c} = (P_c)_{c \in C \setminus \{c\}}$ 表示其他学校的偏好, R_s 和 R_c 表示分别 P_s 和 P_c 相应的非严格偏好。

高考和大学招生模型有着一个重要的不同之处: 在大学招生模型中, 学生和学校都是积极参与人, 他们同时可以操纵自己的偏好。而在高考录取中, 只有学生是积极行动的参与人, 大学按照学生的考分高低录取学生, 几乎没有自己的选择余地。但是, 给定一个高考招生问题 $\{S, C, q, f, P_S\}$, 我们可以通过考生的分数构造大学的偏好, 把高考问题转化为相应的大学招生模型 $\{S, C, q, P_S, P_C\}$, 使得大学招生模型中的概念和方法可以应用在高考招生问题中。

学校偏好 P_c 的构造方法如下: 对于每个大学 $c \in C$, 它的偏好 P_c 满足: 对所有的 $s_1, s_2 \in S$, $s_1 P_c s_2$, 当且仅当 $f_{s_1} > f_{s_2}$, 并且对所有的 $s \in S$, $s P_c s_0$ 。这里, s_0 表示一个学生也不录取的选择。

这样一来, 在高考招生问题对应的大学招生模型中所有的学校对考生有相同的偏好。大学招生模型中录取的结果也是学校和考生的匹配。我们前面在高考招生问题中定义的匹配和个人理性的概念仍适用于大学招生模型。在大学招生模型中, 一个非常重要的概念是稳定性 (stable)。一个学生和学校的组合 $(s, c) \in S \times C$ 阻碍 (block) 了一个匹配 μ : 如果 $c P_s \mu(s)$, 并且 $|\mu^{-1}(c)| < q_c$; 或者 $c P_s \mu(s)$, 并且存在 $\tilde{s} \in \mu^{-1}(c)$ $s P_c \tilde{s}$ 。如果匹配 η 是个人理性的, 并且不会被任何学生和学校的组合阻碍, 那么这个匹配是稳定的。由稳定性的定义我们可以看到稳定性和公平有很大的关联, 一个匹配是稳定的当且仅当它是个人理性的、没有浪费和公平的。由我们前面的证明可知, 完全信息时高考显示偏好博弈的纳什均衡结果是一个稳定的匹配。

Gale and Shapley (1962) 证明, 在大学招生模型中稳定匹配总是存在的, 并且在所有的稳定匹配中, 存在一个匹配对于学生来说是帕累托最优的, 一

⁴ 由于学校可以录取多个人, 因此学校也对不同集合的学生存在偏好, 关于学校的偏好讨论参见 Roth (1985)。

个匹配对于所有学校帕累托最优。他们提出了达到学生（学校）最优稳定匹配的机制，这个机制通过延迟接受学生（学校）申请来达到。我们在前文已经看到了学生最优机制如何运作，这里将它概述如下：

Gale-Shapley 学生最优机制

第一轮 所有的学生向他可接受的大学中最偏好的学校申请。每一所大学 c 从所有申请的学生中留下它偏好排序上最好的 q_c 个学生，其他的学生都被拒绝，进入下一轮申请。被留下的 q_c 个学生记入预录名单。如果申请的学生不足 q_c 个，就预录所有的学生。

第二轮 所有被拒绝的学生向他可接受并且没有拒绝过他的大学中最偏好的学校申请。每一所大学 c 考虑所有本轮申请的学生和上一轮留在预录名单上的学生，从中留下最好的 q_c 个学生，拒绝其他学生；如果学生不足 q_c 个，就留下所有的学生。被拒绝的学生进入下一轮申请，留下的学生记入预录名单。

.....

第 K 轮 所有在上一轮被拒绝的学生向他可接受并且没有拒绝过他的大学中最偏好的学校申请。如果没有可接受的学校，这个学生就停止申请。每一所大学 c 考虑本轮所有申请的学生和上一轮留在预录名单上的学生，从中留下最好的 q_c 个学生，拒绝其他学生；如果学生不足 q_c 个，就留下所有的学生。被拒绝的学生进入下一轮申请，留下的学生记入预录名单。

当没有学生的申请被拒绝的时候，这个录取程序停止。由于学生和学校的数目有限，这个过程一定会在有限步后停止，所有留在预录名单上的学生就被学校正式录取。如果一个学生被所有的学校拒绝，那么这个学生就落榜了。

在高考问题对应的特殊大学招生模型中，通过 Gale-Shapley 学生最优机制可以达到学生最优的稳定匹配。由于所有学校的偏好都和考生分数排序相同，这时只存在唯一的稳定匹配。为了说明这一点，我们证明每一个稳定的匹配都可以作为高考录取博弈的纳什均衡结果。

定理 4 每一个稳定的匹配都可以作为一个高考录取博弈的均衡结果。

证明 设 μ 是一个稳定的匹配，我们考虑如下的策略组合 $Q = (Q_s, Q_{-s})$ ，学生 s 的策略 Q_s 是把他的匹配 $\mu(s)$ 作为第一志愿，其他的学校都不可接受。这时，通过高考机制我们必然有 $\mu(s) = T[Q](s)$ 。对于学生 s 如果存在一个策略 $\tilde{Q}$ 使得 $c = T[\tilde{Q}, Q_{-s}](s) P_s \mu(s)$ ，这里 P_s 是学生 s 的真实偏好，必然有 $|\mu^{-1}(c)| < q_c$ ，或者存在 $\tilde{s} \in \mu^{-1}(c) P_c \tilde{s}$ ，但这和匹配 μ 稳定矛盾。

这个定理也提供了一种构造高考招生博弈中纯策略均衡的方法。由于稳

定匹配总是存在的,因此我们也证明了在高考志愿填报博弈中总是存在纯策略纳什均衡的。结合我们证明的填报博弈中均衡结果的唯一性,我们证明了在高考问题对应的大学招生模型中稳定的匹配是唯一的。通过高考填报博弈的均衡和 Gale-Shapley 学生最优机制,都可以达到这个稳定匹配。但是和高考填报博弈相比, Gale-Shapley 学生最优机制在高考录取中更多的优势是:

定理 5 (Dubins and Freedman, 1981; Roth, 1982) Gale-Shapley 学生最优机制是抗操纵的 (strategy-proof)。

在 Gale-Shapley 学生最优机制下,每个人的真实偏好就是他的优势策略,不论其他人如何填报,考生填报自己真实的偏好得到的结果不比填报任何其他志愿来得差,这就是无操纵 (Strategyproof) 的含义。虽然在这个机制下均衡不唯一⁵,但是由于 Gale-Shapley 学生最优机制的无操纵性,给定其他人的均衡策略,考生使用自己真实偏好和他的均衡策略得到的结果是一样的。给定其他人的任何策略组合,考生的真实偏好总是他的最优反应之一,因而考生不需要协调,可以不必关心其他人填报的策略。在 Gale-Shapley 学生最优机制下,得到的匹配总是公平的,考生也不需要担心由于填报志愿不当而丧失了考分的优先性。Roth (1989) 证明,即使考生对其他人和学校的偏好存在不确定时,在 Gale-Shapley 学生最优机制下每个考生申报真实偏好仍然是他的均衡策略。这样一来,即使是在考前或者估分填报时,使用 Gale-Shapley 学生最优机制仍使所有考生真实填报自己的志愿构成一个均衡,填报志愿十分简单,考分真正地发挥分配的作用。

在高考招生中,分数面前的公平性是实际操作中的基本要求,考分高的学生有更大的优先性录取到他偏好的学校,我们希望录取的机制是一个公平的机制。在所有的公平的机制中:

定理 6 (Gale and Shapley, 1962; Balinski and Sönmez, 1999) Gale-Shapley 学生最优机制帕累托优于任何公平机制。

这个定理的含义是给定一个招生问题,按照 Gale-Shapley 学生最优机制录取得到的结果帕累托优于任何公平匹配的机制得到的结果。如果我们希望最后录取结果没有“高分低就”现象,那么 Gale-Shapley 学生最优机制得到的结果是所有分数公平匹配中帕累托最优的。

定理 7 (Alcalde and Barberà, 1994) 在高考招生问题中, Gale-Shapley 学生最优机制是唯一满足个人理性、公平、没有浪费和抗操纵的学生录取机制。

⁵ 感谢匿名审稿人指出这一点。

这个定理刻画了在高考招生问题中 Gale-Shapley 学生最优机制的唯一性。只有 Gale-Shapley 学生最优机制才做到使得录取结果满足个人理性，没有浪费和分数公平，并且考生真实申报自己的学校偏好是优势策略。

分数的公平性的含义，一方面是指在高考录取过程中，一个学生的考分是他能否被学校录取的指标；另一方面，录取机制也应当尊重学生考分的进步，当一个学生的考分提高时，他所能达到的学校不会比他分数低时所能达到的学校还差。这样的机制就是尊重考分的机制。

定理8 (Balinski and Sönmez, 1999) 唯一的尊重学生考分进步，并且满足个人理性，没有浪费和公平的学生录取机制是 Gale-Shapley 学生最优机制。

以上这些定理表明，如果高考制度仍然是用统一考试和分数来决定录取，Gale-Shapley 学生最优机制具有的许多性质优于当前的高考录取机制。

在 Gale-Shapley 大学招生模型中，与学生最优机制相应的还有大学最优机制，这个机制和学生最优机制的不同在于每一轮中由大学发出邀请，考生决定是否接受学校的邀请。Gale and Shapley (1962) 证明学校最优机制也可以达到稳定的匹配。

Gale-Shapley 学校最优机制

第一轮 所有的学校向它偏好排序上前 q_c 个学生发出入学邀请。收到邀请的学生从中选择他最偏好的大学，拒绝其他大学的邀请。

第二轮 所有被拒绝的大学依据被拒绝的名额向它可接受并且在其偏好排序上未被邀请过的学生发出邀请。收到邀请的学生考虑本次所有的邀请和上一轮保留的邀请，从中选择他最偏好的学校，拒绝其他的邀请。

.....

第 K 轮 所有在上一轮被拒绝的学校根据被拒绝的名额向它可接受并且在其偏好排序上未被邀请过的学生发出邀请。收到邀请的学生考虑本次所有的邀请和上一轮保留的邀请，从中选择他最偏好的学校，拒绝其他的邀请。

当没有学校的邀请被拒绝时，录取程序就停止。由于学生和学校的数目有限，这个过程一定会在有限步后停止，所有学校的邀请就被学生正式接受，学生被该学校正式录取。如果学生没有收到任何学校的邀请，那么这个学生就落榜了。

在高考录取对应的大学招生问题中，所有学校的偏好都相同，只有一个稳定的匹配，因此学校最优机制得到的结果和考生最优机制录取的结果是一样的。在一般的大学招生模型中，这两个机制下的结果是不同的。由于学校的偏好是考生的考分决定的排序，所有学校会偏好同样的考生，所有被拒绝的学校总是给同样的学生发送邀请，因而 Gale-Shapley 学校最优机制等价于下面的机制：

分数公平系列独裁机制

先把所有考生按分数从高到低排序,然后从高分到低分一次一个依次录取。第一个考生从所有招生学校中,选择他最偏好的学校,被选学校的招生名额减少一个;如果一个学校用完招生名额,就停止招生;如果没有考生可接受的学校,他选择不上学。然后考虑下一个学生。一般来说,第 k 个考生从在他选择时仍招生的所有学校中选择他最偏好的学校,被选学校的招生名额相应地减少一个,如果学校用完招生名额,就停止招生;如果没有可接受的学校,他选择不上学;然后考虑下一个学生。这样一直下去,直到所有的考生都被考虑或者所有的学校的招生名额用完,招生停止,剩下的考生落榜。

在这个机制下,考生按照分数高低排序后依次选择,最后的录取结果是公平的,这个机制是帕累托有效,并且是抗操纵的(Zhou, 1990)。这个机制在高考的具体录取操作中已经被应用了,就是江苏省 2005 年的“平行院校志愿”机制:

“平行院校志愿的投档原则是‘分数优先,遵循志愿’。具体地说,在某个批次投档时,计算机投档系统按照考生的高考 5 门科目总分从高到低顺序逐个检索、投档。当检索到某个考生时,再依次检索该考生所填报的 A、B、C 或 A、B、C、D、E 志愿。只要以上被检索的 3 所或 5 所院校志愿中一经出现符合投档条件的院校即向该院校投档,由高校决定该生录取与否及所录取专业;如该生填报的志愿经检索后均不符合投档条件,该生则无法被投档,但该生此时也已享受了本批次平行院校志愿的检索、投档机会,计算机将自动进行下一个考生的检索、投档。”

如果不考虑考生志愿中的学校数目限制⁶,“平行院校志愿”机制就是我们上面描述的系列独裁机制。因而,“平行院校志愿”机制也等价于 Gale-Shapley 学生最优机制。关于 Gale-Shapley 学生最优机制对传统的高考录取机制可能造成的改进,虽然我们没有明确的数据对比,但是从江苏省的政府工作总结中对于“平行院校志愿”机制的评价可以看到部分的效果⁷:

“……去年,我省实行平行志愿填报方式,志愿填报率明显提高、高分落榜考生明显减少、志愿有效率明显增强、降分院校和降分幅度明显降低,受到考生、家长、学校和社会的广泛赞誉。今年将继续实行平行志愿的方式,同时将进一步完善操作办法。”

⁶ 考生只能填写三个学校,而在每个批次招生的学校上百个,数目限制可能会造成拥挤导致无效率。

⁷ 来自《我省部署 2006 年普通高校招生工作》,见 <http://www.jszs.net/zhaokaodongtai/184357.shtml>。

五、结 论

高考录取机制“第一志愿优先”的性质，致使录取机制不是一个公平机制，也不是一个抗操纵的机制。考生为了避免“高分低就”，使得高考志愿填报成了一个几十万考生参加的协调博弈。与当前的高考录取机制相比，Gale-Shapley 学生最优机制具有一系列的优良性质，而且这个机制也已经在实践中被广泛使用。江苏省 2005 年采用的“平行院校志愿”机制正是 Gale-Shapley 学生最优机制的一种特殊形式。这个录取制度的出现提供了制度自发演进的一个例子，印证了制度的变革可以减少无效率的原理。

“波士顿机制”与高考录取机制类似，也是“第一志愿优先”，Atila Abdulkadiroglu, Parag A. Pathak, Alvin E. Roth 和 Tayfun Sönmez (2006) 分析了这个机制下学生填报的志愿数据。他们发现并不是所有的考生和家长都意识到或者有能力在这个机制下操纵自己的偏好，一些家长没有避免明显的错误。和经济理论中的均衡分析不同，在实际中可能并不是所有家长都有同样的能力和资源计算出均衡策略。因而录取机制无操纵也是一种公平的要求，使所有考生在同一平台上竞争。

本文分析了知分报考时的情形，只是把分数作为确定录取顺序的公平分配指标。使用分数的理由之一是分数反映了能力，分数如何反映能力这一点在比较报考时间安排时显得更加重要了。钟笑寒、程娜和何云帆 (2004) 得到了初步的结论，但是他们的结论依赖分数只是能力反映的假设。另一方面，高分也和家庭教育和家庭投资有关。因而，需要经验研究考生高考分数的决定因素和能力之间的关系，这对于理解高考分数录取也有重要的启示作用。

对于是否按照考分标准来录取学生也存在争论，一些院校也在探索自主确定招生标准。当允许学校对学生有不同的排序时，高考录取机制就变成了“波士顿机制”，而这个机制与 Gale-Shapley 学生最优机制相比是有缺陷的 (Abdulkadiroglu and Sönmez, 2003)。Gale-Shapley 学生最优机制中也允许不同的学校对考生有不同的偏好，在这个情形下仍然是抗操纵的。如果高考录取仍然实行集中录取，Gale-Shapley 学生最优机制既可以允许学校有不同的评价标准，同时也提供给学生足够的选择空间。这个机制可以是高考录取机制改革方案的一个备选项。

本文只是理论的分析，从理论上比较了不同机制的效果。作为一个经验比较，一个工作就是可以利用考生实际填报的志愿数据比较 Gale-Shapley 学生最优机制和实际录取结果，虽然这些偏好不一定是考生真实偏好，但是可以提供考察机制变换对于不同考生的影响的一个可能基准。另一个常见的担心是由于高考录取人数规模太大，Gale-Shapley 学生最优机制的适用性可能受到限制。由于这个算法是多项式时间的，因此在实际应用时可以分成两阶段来

完成：第一阶段是考生和学校分别形成并报告自己的偏好，第二阶段是利用 Gale-Shapley 学生最优机制得到学生和学校的匹配。实际的高考录取也持续 3 个月的时间，随着现在计算机能力的提高，得出结果可能并不需要太多的时间。当然，机制的实际设计执行可能需要更多的考虑。⁸

本文的分析中忽略了志愿填报中包含的学校数目的限制，假设考生申报的志愿中可以包含任意数目的学校。志愿数目限制的存在使得考生在填报志愿时必须选择填报的学校，这种量的限制可能引起质的变化。Antonio Romero-Medina (1998) 已经分析了存在数量限制时的 Gale-Shapley 学生最优机制下的博弈，一个值得进一步研究的方向就是分析存在数量限制时高考录取机制下的博弈，比较本文中的结论是否成立。

附 录

定理 1 的证明 给定个人理性、公平的匹配 μ ，假设它不是帕累托有效的，则存在一个匹配 η ，对于任意的 $s \in S$ ，都有 $\eta(s) R_s \mu(s)$ ，并且存在 $s_{i_1} \eta(s_{i_1}) P_{s_{i_1}} \mu(s_{i_1})$ ，由偏好的严格性，我们知道 $\eta(s_{i_1}) \neq c_0$ 。令 $\eta(s_{i_1}) = c_{i_1}$ ，则由匹配 μ 的公平性，可以知道 $|\mu^{-1}(c_{i_1})| = q_{c_{i_1}}$ ，并且对于任意的 $\bar{s} \in \mu^{-1}(c_{i_1})$ ，都有 $f_{\bar{s}} > f_{s_{i_1}}$ 。由于在匹配 η 中， $\eta(s_1) = c_{i_1}$ ，因此必然存在 $s_{i_2} \in \mu(c_{i_1})$ ，满足 $\eta(s_{i_2}) P_{s_{i_2}} \mu(s_{i_2})$ 。同样的方法，我们可以证明，对任意 $\eta(s_{i_k}) P_{s_{i_k}} \mu(s_{i_k})$ ，我们都可以找到 $s_{i_{k+1}}$ ，满足 $\eta(s_{i_{k+1}}) P_{s_{i_{k+1}}} \mu(s_{i_{k+1}})$ ，并且 $f_{s_{i_{k+1}}} > f_{s_{i_k}}$ 。由于学生的人数是有限的，这是不可能的。这样，我们就证明了任何一个人理性、公平的匹配都是帕累托有效的匹配。

定理 3 的证明 (1) 如果一个匹配 μ 是均衡的结果，必然是个人理性的。

不然，存在 $\bar{s} \in S$ ，有 $c_0 P_{\bar{s}} \mu(\bar{s})$ 。这里 $\mu = T[Q_{\bar{s}}^*, Q_{-\bar{s}}^*]$ ， $Q_{\bar{s}}^*$ 是在均衡时学生 $\bar{s}$ 的策略，他的真实的偏好是 $P_{\bar{s}}$ ， $Q_{-\bar{s}}^*$ 是除学生 $\bar{s}$ 外其他所有人的均衡策略。给定其他人的策略 $Q_{-\bar{s}}^*$ 不变，学生 $\bar{s}$ 的策略变为偏好 $\hat{Q}_{\bar{s}}$ ，偏好 $\hat{Q}_{\bar{s}}$ 把选择 c_0 作为学生 $\bar{s}$ 的第一志愿。在偏好组 $(\hat{Q}_{\bar{s}}, Q_{-\bar{s}}^*)$ 下，高考机制的结果必然有 $c_0 = T[\hat{Q}_{\bar{s}}, Q_{-\bar{s}}^*](\bar{s})$ ，这样我们就得到 $c_0 P_{\bar{s}} \mu(\bar{s})$ 。这和匹配 μ 是均衡匹配结果相矛盾。

(2) 如果一个匹配 μ 是均衡的结果，必然是公平的。

如果匹配 μ 不是公平的，那么存在 $\bar{s} \in S$ 和 $\hat{c} \in C$ ，有 $\hat{c} P_{\bar{s}} \mu(\bar{s})$ ，或者 $|\mu^{-1}(\hat{c})| < q_{\hat{c}}$ 或者存在 $\bar{s} \in \mu^{-1}(\hat{c})$ ，有 $f_{\bar{s}} < f_{\hat{c}}$ 。这里 $\mu = T[Q_{\bar{s}}^*, Q_{-\bar{s}}^*]$ ， $Q_{\bar{s}}^*$ 是在均衡时学生 $\bar{s}$ 的策略，他的真实的偏好是 $P_{\bar{s}}$ ， $Q_{-\bar{s}}^*$ 是除学生 $\bar{s}$ 外其他所有人的策略。给定其他人的策略 $Q_{-\bar{s}}^*$ 不变，学生 $\bar{s}$ 的策略变为偏好 $\hat{Q}_{\bar{s}}$ ，偏好 $\hat{Q}_{\bar{s}}$ 把学校 $\hat{c}$ 作为学生 $\bar{s}$ 的第一志愿。在偏好组 $(\hat{Q}_{\bar{s}}, Q_{-\bar{s}}^*)$ 下，高考机制的结果必然有 $\hat{c} = T[\hat{Q}_{\bar{s}}, Q_{-\bar{s}}^*](\bar{s})$ 。这样我们就得到 $\hat{c}(\bar{s}) P_{\bar{s}} \mu(\bar{s})$ ，这

⁸ Atilla Abdulkaduroglu, Parag A. Pathak, Alvin E. Roth (2005) Atilla Abdulkaduroglu, Parag A. Pathak, Alvin E. Roth and Tayfun Sönmez (2005) 总结了他们在美国公立学校录取中应用 Gale-Shapley 学生最优机制时的经验和结果。

和 $\mu = T[Q_s^*, Q_{-s}^*]$ 是纳什均衡结果矛盾。因此均衡结果必然是公平的。

(3) 如果 $|A| \neq 1$, 则存在两个均衡结果 $\mu_1 = T[Q^1]$, $\mu_2 = T[Q^2]$, 并且 $\mu_1 \neq \mu_2$ 。必然存在 $s_0 \in S$, 有 $\mu_1(s_0) \neq \mu_2(s_0)$, 不妨设 $\mu_1(s_0) P_{s_0} \mu_2(s_0)$, 这是根据偏好的严格性得来的。

令 $m_0 = \mu_1(s_0)$, 由于 μ_2 是公平的, 我们必然有对于任意的 $\bar{s} \in \mu_2^{-1}(m_0)$, $f_{\bar{s}} > f_{s_0}$, 并且 $|\mu_2^{-1}(m_0)| = q_{m_0}$ 。由于 μ_1 是公平的, 我们必然对于任意的 $\bar{s} \in \mu_2^{-1}(m_0)$, 有 $\mu_1(\bar{s}) R_s m_0$, 不然就违反公平性。由于在 μ_1 中可以有 $q_{m_0} + 1$ 个人被学校 m_0 录取, 我们必然有 $s_1 \in \mu_2^{-1}(m_0)$, $f_{s_1} > f_{s_0}$ 而且 $m_1 = \mu_1(s_1) \neq m_0$ 。由于偏好是严格的, 我们必然有 $m_1 P_{s_1} m_0$; 这样我们得到 $\mu_1(s_1) P_{s_1} \mu_2(s_1)$ 且 $f_{s_1} > f_{s_0}$ 。对于任意的 $\mu_1(s_i) P_{s_i} \mu_2(s_i)$, 按照和上面同样的推理我们可以得到 $s_{i+1} \in \mu_2^{-1}(m_i)$, $f_{s_{i+1}} > f_{s_i}$ 并且 $\mu_1(s_{i+1}) P_{s_i} \mu_2(s_{i+1})$ 。

这个推理可以一直进行下去, 这和学生人数有限相矛盾, 因此前提不成立, 只有一个纳什均衡结果。

参考文献

- [1] Abdulkadiroglu, A., P. Pathak, and A. Roth, "The New York City High School Match", *American Economic Review*, 2005, 95(2), 364—367.
- [2] Abdulkadiroglu, A., P. Pathak, A. Roth, and T. Sönmez, "The Boston Public School Match", *American Economic Review*, 2005, 95(2), 368—371.
- [3] Abdulkadiroglu, A., P. Pathak, A. Roth, and T. Sönmez, "Changing the Boston School Choice Mechanism", 2006, NBER Working Paper 11965.
- [4] Abdulkadiroglu, A., and T. Sönmez, "School Choice: A Mechanism Design Approach", *American Economic Review*, 2003, 93(3), 729—747.
- [5] Alcalde, J., and S. Barberà, "Top Dominance and the Possibility of Strategy-proof Stable Solutions to Matching Problems", *Economic Theory*, 1994, 4(3), 417—435.
- [6] Balinski, M., and T. Sönmez, "A Tale of Two Mechanisms: Student Placement", *Journal of Economic Theory*, 1999, 84(1), 73—94.
- [7] Chen, Y., and T. Sönmez, "School Choice: An Experimental Study", *Journal of Economic Theory*, 2006, 127(1), 202—231.
- [8] Dubins, L., and D. Freedman, "Machiavelli and the Gale-Shapley Algorithm", *American Mathematical Monthly*, 1981, 88(7), 485—494.
- [9] Ergin, H., and T. Sönmez, "Games of School Choice under the Boston Mechanism", *Journal of Public Economics* 2006, 90(1—2), 215—237.
- [10] Gale, D., and L. Shapley, "College Admissions and the Stability of Marriage", *American Mathematical Monthly*, 1962, 69(1), 9—15.
- [11] 聂海峰, "填报高考志愿哪种方式最优?"《南方经济》2006年第6期,第75—90页。
- [12] Romero-Medina, A., "Implementation of Stable Solutions in a Restricted Matching Market", *Review of Economic Design* 1998, 3(2), 137—147.
- [13] Roth, A., "The Economics of Matching: Stability and Incentives", *Mathematical Operation Research*, 1982, 7(4), 617—628.
- [14] Roth, A., "The College Admissions Problems is not Equivalent to the Marriage Problem", *Journal of Economic Theory*, 1985, 36(2), 277—288.

- [15] Roth, A., "Two-Sided Matching with Incomplete Information about Others' Preferences", *Games and Economic Behavior*, 1989, 1(2), 191—209.
- [16] Roth, A., "A Natural Experiment in the Organization of Entry Level Labor Markets: Regional Markets for New Physicians and Surgeons in the U. K. ", *American Economic Review*, 1991, 81(3), 415—440.
- [17] Roth, A., "The Economist as Engineer: Game Theory, Experimentation, and Computation as Tools for Design Economics", *Econometrica*, 2002, 70(4), 1341—1378.
- [18] Roth, A., and E. Peranson, "The Effects of a Change in the NRMP Matching Algorithm", *Journal of American Medical Association*, 1997, (278), 729—732.
- [19] Roth, A., and E. Peranson, "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design", *American Economic Review*, 1999, 89(4), 784—870.
- [20] Roth, A., and M. Sotomayor, *Two-Sided Matching: A Study in Game Theoretic Modeling and Analysis*. London/New York: Cambridge Univ. Press, 1990.
- [21] 钟笑寒、程娜、何云帆, "花落谁家——高考志愿填报机制的博弈模型", 《经济学季刊》2004 年第 3 卷第 3 期, 第 763—778 页。
- [22] Zhou, L., "On a Conjecture by Gale about One-Sided Matching Problems", *Journal of Economic Theory*, 1990, 52(1), 123—135.

A Game-theoretical Analysis of China's College Admission Mechanism

HAIFENG NIE
(Sun Yat-sen University)

Abstract The College admission mechanism in China is a giant revealing preference game played by millions of students every year. Unfortunately, the true preferences of students are not their optimal equilibrium strategies. We prove that the equilibrium outcome is unique and Pareto efficient, but it is hardly a possible mission for students to coordinate to equilibrium. Fortunately, the admirable Gale-Shapley model of deferred-acceptance mechanism in college admission can be adapted to achieve the same outcome with more advantages. A great deal of efficiency would be achieved if the existing college admission mechanism had been replaced.

JEL Classification C78, D61, D78

近代外国在华投资：规模与效应分析

杨德才*

摘要 传统观点认为,近代外国在华投资不仅盈利极其丰厚,而且还严重抑制了中国本土企业的成长与发展,成为阻碍近代中国经济发展的一个重要因素。本文通过对近代外国在华投资的规模、利润率以及外国投资与日俱增背景下中资企业成长、发展情况的分析,发现传统观点是值得商榷的,文章认为近代外国在华直接投资的正效应是非常明显的,不仅没有成为中资企业发展的阻碍,而且还刺激并促进了中资企业的成长与发展。当然,至于外国间接投资的效应则需要进行分析。

关键词 近代外国投资,规模,效应

鸦片战争后,外国资本紧随坚船利炮来到了中国,外国在近代中国的投资帷幕便由此拉开。从投资者的角度来说,国外投资最基本的动机是资本所有者追求其资产组合价值或资产净值最大化的愿望。资本逐利,追求价值的最大化本无可厚非,然而,近代外国在华的投资则是在列强强迫中国签订并接受不平等条约的背景下发生的,因而,这种外国投资的性质有了很大的变异。尽管近代中国政府允许外国在华直接投资办厂(即外国投资在华取得合法权)是在签订了《马关条约》之后,但事实上,西方列强在华的投资早在鸦片战争后便已开始。那么,近代外国在华规模究竟有多大?这些外国投资赢利情况怎样?它们对近代中国本土企业的成长与发展产生了怎样的影响?甚至,它们对近代中国经济的发展又产生了哪些深远的影响呢?本文将对这些问题逐一进行研究。

一、外国投资作用：意见与争论

所谓外国投资(Foreign Investment),是从投资的接受国来讲的,是指来自本国(或称为母国)之外的资本所从事的投资行为;而如果站在投资者所属的国家来讲,则又称为国外投资,是指从投资者所属的国家之外获得资产的行为。外国投资一般包括外国直接投资(Foreign Direct Investment)和外国间接投资(Foreign Indirect Investment)两种基本形式,其中最为主要的是外国直接投资(简称FDI)。直接投资一般涉及工厂和土地等生产资料的所有

* 南京大学经济学系。通讯地址:江苏省南京市汉口路22号南京大学经济学系 210093;电话:(025) 83592024;E-mail: ydc@nju.edu.cn。本文为第五届中国经济学会入选论文,感谢匿名审稿人的中肯而又富有启发性的评论。笔者文责自负。

权,如兴办工厂、开发房地产等活动,而间接投资则主要涉及债券、银行贷款和证券(股票)等金融资产,不直接参与生产。如果股票所有权使股东控制了厂商的经营活动,股票投资也可以认为是直接投资。

赫伯特·G. 格鲁贝尔认为,造成外国直接投资的动机是很复杂的,除了追求资产价值的最大化外,还有其他的动机。这些动机可能牵涉到外在性或市场的不完全性有关系,而通过在国外投资、成为跨国公司,它们会顺次被内在化或者被消除掉。另外,厂商之所以要从事外国直接投资,还在于它们要控制产品质量和保持商业秘密。外国直接投资最后一个主要原因涉及政府政策造成的扭曲。¹

外国投资行为虽然是与西方列强的对外侵略活动同步发生的(甚至更早),但对外国投资行为的经济研究则是在 20 世纪中叶以后,关于外国投资对东道国经济发展的影响,争论颇多,难以统一:一派持支持赞成的态度,另一派则持反对批评的观点。²

持支持赞成态度的认为,外国投资可以补充东道国可用资源供给,如储蓄、外汇、政府财政收入和人力资本技能等与获得增长和发展目标应有的资源水平的差距。例如,在一个简单的“储蓄-投资缺口”模型中,哈罗德-多马模型提出国内储蓄率 s 与产出增长率 g 之间存在直接关系,用方程式表示为 $g = s/k$,其中 k 是国内资本产出比率。如果希望的国民产出增长率 g 定为每年 7%,而资本-产出比率是 3,则必需的年储蓄率是 21% (因为 $s = gk$)。比如,如果国内所能够动员的储蓄率只有国内生产总值的 16%,则可以说存在 5% 的“储蓄缺口”。如果一国可以用外国投资弥补这个缺口,它就可能很好地达到预定的增长率。持支持赞成态度的还认为,外国投资除了为东道国提供金融资源和新工厂外,而且还提供了“一揽子”所需资源,包括管理经验、企业家能力和技术技能,这可以通过培训计划和在工作中学习的方式被转移到当地国家。他们认为,这种知识、技能和技术的转移对接受国而言是富有生产效益的。

对外国投资持反对批评的观点,原因是多方面的。在经济方面,外国投资并不能弥补东道国经济发展的资本缺口,尽管外国投资会提供资本,它们也可能降低储蓄和投资率。这是由于它们会通过东道国签订排他性的生产协定而窒息竞争,不能将大量的利润再投资,并以高利率贷款给东道国。外国投资提供的管理、企业家技能、思想、技术等对促进当地这些资源的发展可能是微不足道的,而且可能由于外国投资所创办的公司(尤其是那些进行国外投资的跨国公司)垄断了国内市场从而窒息了国内企业的发展,从实

¹ 伊特韦尔等编,《新帕尔格雷夫经济学大辞典》第二卷,陈岱孙主编译,经济科学出版社 1996 年版,第 432 页。

² 迈克尔·P. 托达罗,《经济发展》,黄卫平等译,中国经济出版社 1999 年版,第 527—530 页。

际上阻碍了这些资源的发展。此外，外国投资所创办的公司可能会压抑国内企业家和运用它们的先进技术等，抑制当地小企业的出现和发展，从而损害东道国的经济。最后，从政治的角度看，极端的情况是外国投资所创办的公司甚至可能用回扣的直接手段贿赂最高阶层的政府官员或其他重要人员，从而在各个阶层的政治决策中能够发挥相当大的影响，阻碍东道国经济效益、社会福利最大化的实现。

上述关于外国投资作用的争论，是针对一个主权完整的东道国而言的。虽然不能肯定地说对一个主权完整的国家而言外国投资尽是正面的积极效应，但是至少外国投资对东道国经济发展所产生的正效应是主要的。而在近代，外国投资的大量进入显然不是在中国主权完整的状态下发生的，它们受到领事裁判权、协定关税等特权的庇护，那么，享有特权的外国投资对近代中国经济发展究竟产生了怎样的影响，就需要客观、具体地分析。

二、外国直接投资及其热潮

（一）甲午战争前的外国直接投资

直接投资最基本的特征就是直接从事生产活动。外国在近代中国的投资首先是从直接投资开始的。从直接投资的规模、方向和影响上来看，近代外国在华投资明显地以甲午战争为界线，分为前后两个时期。

鸦片战争以后，随着西方列强对华侵略范围的不断扩大和它们获得的特权越来越多，它们在华因直接投资而设立的各种工厂、企业也越来越多，涉及的行业主要是航运业、船舶修造业、农副产品加工业、服务性轻小工业和公用事业等。其中在鸦片战争后到19世纪60年代中期期间，外国投资的重点是航运业和船舶修造业，成立了许多后来比较著名的船厂和轮船公司，如祥生船厂、香港黄埔船坞公司、耶松船厂、旗昌轮船公司、怡和洋行、太古洋行等。从60年代到80年代期间，外国投资相对集中于农副产品加工业，投资兴办了许多砖茶厂、缫丝厂、轧花厂、制糖厂等。在80年代后至1894年，外国投资的重点是公用事业和杂项工业，如在租界内开办自来水、自来火、电光公司等，还投资兴办酿酒、制药等工业。这样，到甲午战争前，外国在华直接投资设立的工业企业，据估计，共有191个³，其中船舶修造业和出口商品加工业116个，印刷厂13个，食品加工业23个，公用事业8个，化学工业11个，建筑及杂项工业各10个。到1894年，据估计，外国在华工业

³ 赵德馨(1988)，第76页。关于外国投资到甲午战争前在华创办的工业企业准确数字，争论颇多。据孙毓棠研究，到甲午战争前，外国投资在华创办的工业企业计88个(孙毓棠、汪敬虞主编，《中国近代工业史资料》第一辑，第一册，科学出版社1957年版，第242—247页)。还有人认为其数量应为103个(费正清编，《剑桥中国晚清史》，下册，中国社会科学院历史研究所编译室译，中国社会科学出版社1993年版，第40页)。

资本总额近 2000 万元，投资总额近 2800 万元（约折合 1424.5 万美元）；航运业投资总额约为 1334.3 万美元。具体如表 1 所示。

表 1 甲午战争前外国直接投资估计(1894 年) 单位 :万元

业 别		资本额	投资额
工业投资	船舶修造业	494.3	811.2
	砖茶制造业	400.0	560.0
	机器缫丝业	397.2	517.3
	进出口加工业	149.3	193.8
	一般轻工业	379.3	491.0
	公用事业	152.3	204.5
合计		1972.4	2777.8 (约为 1424.5 万美元)
航运业投资			1334.3 万美元
总计			2758.8 万美元

资料来源：(1) 孙毓棠(1981) ,第 116 页；(2) 许涤新、吴承明(1990) ,第 133 页。

从表 1 可以看出，甲午战争前，外国在华直接投资的产业主要集中于船舶航运业和农副产品加工业，其目的是直接为进出口贸易服务。这一时期，外国在华投资之所以相对集中于上述两个产业，是因为当时这两个产业的市场比较大，每个产业中的具体的每个工厂投资规模则不需要太大，但利润率却很高，利润比较丰厚。⁴ 表 2 是甲午战争前某些年份外国在华直接投资创办的部分企业的赢利状况。

表 2 甲午战争前在华外国直接投资创办企业的获利状况

企业名称	年代	资本	利润	利润率	股票附息率
香港黄埔船坞公司	1886	1562500 元	375049 元	24.0%	16%
	1893	1562500 元	322386 元	20.6%	14%
祥生船厂	1891.5—1892.4	800000 两	139903 两	17.5%	12%
	1893.5—1894.4	800000 两	138964 两	17.4%	12%
耶松船厂	1892.7—1893.6	750000 两	101128 两	13.5%	9%
	1893.7—1894.6	750000 两	157259 两	21.0%	11%
美查兄弟有限公司	1889.11—1890.10	275000 两	32471 两	11.8%	10%
	1892.11—1893.10	275000 两	15796 两	5.7%	—
老德记药房	1892	120000 元	21422 元	17.9%	7%
新上海制冰厂	1893	36000 两	5138 两	14.3%	9%
上海大英自来火房	1874	150000 两	23121 两	15.4%	11%
	1889	200000 两	43476 两	21.7%	12%
	1894	200000 两	49456 两	24.7%	12%
上海自来水公司	1884	119639 镑	4885 镑	4.1%	4%
	1890	143926 镑	55630 两	9.1%	8%
	1894	144000 镑	71355 两	10.7%	9%

资料来源：历年《北华捷报》载各企业营业报告与股东会报告。转见孙毓棠(1981) ,第 118 页。

⁴ 关于外国在华投资的赢利状况，学术界的观点并不完全一致。有的（尤其是我国的老一辈经济史学家）认为很高，有的（如我国新生代经济史学研究者和一些国际著名的从事中国经济史学研究的学者）认为其实并不是很高。究竟是高还是低，在本文后面的部分将作出分析。

（二）甲午战争后的外国直接投资

这一时期，外国在华投资的规模有明显增加，其直接投资涉及的产业主要有：制造业、矿冶业、铁路航运业和房地产业等。投资的范围也由甲午战争前集中于东南部沿海和租界而不断地向内地拓展，几乎遍及各地。投资的目的在于生产就地销售的商品。外国在华产业资本 1913 年较 1894 年增长二十余倍，1920 年较 1894 年增长近三十倍，⁵从而形成了外国投资中国的热潮。

制造业方面的直接投资主要投资于纺织、面粉、卷烟、造船及农副产品加工等，在华直接设厂制造生产。根据汪敬虞先生的研究，在 1895—1913 年间，外国直接投资创办了 115 个工厂企业，投资总额为 4807.8 万元。其中，英国 24 个，投资总额为 1069.3 万元；日本 36 个，投资总额为 1868.9 万元；俄国 37 个，投资总额为 907.6 万元；德国、法国和美国分别为 6 个、4 个和 6 个，各自的投资总额分别为 361.9 万元、258.1 万元和 311.0 万元；另外还有其他国家的企业 2 个，投资总额为 31.0 万元。⁶至于 1895—1913 年间外国直接投资在制造业内部各产业间的分布状况，如表 3 所示。

表3 制造业外国直接投资分布状况(1895—1913年)

	家数	投资额(万元)	占制造业比重(%)
造船工业	7	289.5	6.02
纺织工业	16	1251.5	26.03
其中 棉纺织	12	1082.6	22.52
缫丝	4	168.9	3.51
食品工业	69	2355.6	49.0
其中 面粉	43	1035.9	21.55
卷烟	4	383.2	7.97
榨油	11	536.2	11.15
其他工业	23	911.2	18.95

资料来源：汪敬虞，《中国近代工业史资料》第2辑，上册，中华书局1962年版，第7—11页。

外国投资兴办的制造业中，资本主要集中于棉纺织业和面粉业，两者投资额占整个制造业投资总额的 44.07%。棉纺织业又主要以投资办纱厂为主，1897—1914 年间，仅英、美、德、日四国在华纱厂投资额就达 916.6 万元。其中著名的纱厂有英国的怡和纱厂、老公茂纱厂和日本的日信纱厂、上海纺绩公司等。面粉业以俄商和日商的直接投资为主，截止到 1913 年，外国在华投资所办的面粉厂达 43 家，投资总额达 1035.9 万元。其中俄商工厂 35 家、日商工厂 6 家，各自的投资额分别为 700.2 万元、264.7 万元。

在 1895—1913 年间，外国在华用于矿冶业方面的投资总额达 4996.9 万

⁵ 赵德馨(1988)，第116页。

⁶ 许涤新、吴承明(1990)，第548页。

元，设立厂家计 32 个。用于矿冶业方面的直接投资总额超过制造业，使得矿冶业成为甲午战争之后外国直接投资最多的产业，这其中又以英国商人投资最多，达 3793.0 万元。据研究，外国投资在中外煤矿投资总额中所占比重，1906 年为 84.8%，1913 年为 79.6%，1919 年为 60.6%。在煤矿开采方面，1913 年中国煤总产量为 1288 万吨，其中机械采煤 767.8 万吨，而外国资本控制的 7 大煤矿（开滦、焦作、抚顺、本溪湖、华德、井陉、临城）在煤总产量和机械采煤量中分别占 46.8% 和 78.4%。外国投资在近代中国铁矿业与冶铁业中，几乎达到了完全垄断的程度，其中在 1900—1925 年间，生铁与铁矿产量接近 100% 都被外资控制着。

甲午战争后，以铁路航运业为主的运输业是在外国直接投资的所有产业中增长最快、比重最大的产业。据统计，外国在华运输业的投资由 1894 年的 13.3 百万美元增加为 1914 年的 335.6 百万美元，增长 24 倍，其所占外国在华投资总量的份额由 1894 年的 12.2% 上升到 34.9%。⁷ 外国对中国航运业的直接投资在甲午战争前即已开始，甲午战争后继续增长，1894 年前，在华外国轮船公司有 14 家，1895—1913 年新设 14 家，1914—1930 年又设立 15 家。在中国各通商口岸进出中外轮船总吨位中，外轮 1897 年占 76.8%，1916 年占 77.6%，1926 年占 80.0%。1911 年，在长江轮运总吨位中，中国最大的轮船公司招商局与外国太古、怡和、日清三大轮船公司相比较，只占 16.2%。然而，这一时期在铁路航运业中增长最快的并不是航运业，而是对铁路业的投资。截止到 1914 年，外国直接投资于铁路航运业的总资本额为 335.6 百万美元，而其中铁路业投资即达 292 百万美元。关于外国在华铁路投资的具体情况，吴承明先生作了非常深入的研究，其研究结论大致如表 4 所示。

表 4 1914 年外国在华铁路直接投资估计

单位：里程（公里）、投资估计（百万美元）

国别	铁路名称	建筑年代	里程	投资估计
俄	中东铁路	1898—1903	1721	189.3
德	胶济铁路	1899—1904	446	15.0
法	滇越铁路	1903—1909	464	32.0
日	南满铁路	1899—1903	1105	49.0
英	广九铁路（英段）	1907	36	6.7
合计			3772	292.0

资料来源：吴承明，《帝国主义在旧中国的投资》，人民出版社 1955 年版，第 168 页。

外国人在中国占有土地和建造房屋也是一种直接投资，这种直接投资行为同样在甲午战争前即已发生，只是在甲午战争后增长更为迅速而已。据仅仅对英、美、法、日、德五国在华房地产业投资的估计，1902 年投资总额约

⁷ 许涤新、吴承明（1990），第 529 页、540 页。

为 79.6 百万美元，而至 1914 年则已达 225.3 百万美元，为 1902 年的 2.83 倍，增长还是比较快的。

外国在近代中国的直接投资，在甲午战争前多以中小资本为主，许多外国中小资本借助于不平等条约的保护，试图通过在华的直接投资、获取高额利润，迅速度过资本积累阶段，寻求更大的扩张，因此，甲午战争前，从总体上看，直接投资的规模并不很大。而甲午战争后，外国在华直接投资的总量增长很快，其投资特点也由起先的中小资本唱主角渐变为具有垄断性质的大公司（如英国的怡和洋行、日本的南满铁道株式会社等）来投资扩张，因而增长很快。由于垄断生产经营、多元化经营、规模经营，一般获利程度都不错，平均利润率大致在 10%—15% 之间。

三、外国间接投资与日俱增

外国在近代中国的投资，从总量上来看，甲午战争前以直接投资为主，而甲午战争后虽然直接投资额比战前增长了很多，但在外国投资总量中其所占的比重则是下降的，更主要地是以间接投资为重。吴承明先生在对甲午战争前外国在华投资的估计中，其中直接投资的有：航运业、工业、贸易业、其他商业等，它们各自占外国投资总额的比重分别为：11.17%、11.93%、35.12%、4.55%，即直接投资占总投资的 62.77%。而间接投资有：银行业、⁸保险业、政府贷款等，其各自占外国投资总额的比重对应的是：23.52%、5.02%、8.69%，间接投资占总投资额的 37.23%。⁹ 间接投资量明显小于直接投资量。而甲午战争后，这种情况发生了非常大的变化。

甲午战争前外国在华的间接投资首先是从设立银行开始的。鸦片战争后，外国在华设立的第一家银行是东方银行（初名 Oriental Bank，1851 年改称 Oriental Banking Corporation）分行，¹⁰是由英国人设立的。期后，英国人又设立了有利银行（原为 Chartered Bank of Asia，1858 年与 Mercantile Bank of India，London and China 合并，改组为 Chartered Mercantile Bank of India，London and China）和麦加利银行（Chartered Bank of India，Australia and China，亦称渣打银行）分行。这期间最著名的、发展最快的银行是由几个在华的英国大商人 1865 年发起成立于香港的汇丰银行，该银行原定注册资本为 500 万港元，而实际注册资本则不到一半，然而，经过短暂的发展，其拥有资本额迅速增长，1883 年增资至 750 万元、1890 年增资至 1000 万元；同时它还在汉口（1868

⁸ 关于外国在东道国银行业的投资，究竟应看作直接投资、还是间接投资，学术界的观点并不一致。本文中采用的是现代外国投资的相关理论及主要观点，将其视为间接投资的一部分。

⁹ 许涤新、吴承明（1990），第 133 页。

¹⁰ 关于 Oriental Bank、Oriental Banking Corporation，中文有译为丽如银行、金宝银行和东亚银公司的。

年)、天津(1880 年)和北京(1885 年)开设了分行。¹¹后来在中国开设银行的还有法、德、日等国。截止到 1894 年,在中国共有 8 家外资银行,它们的资本及投资额估计如表 5 所示。

表 5 甲午战争前外资银行在华投资估计(1894 年)

银行名称	资本额	在华投资额估计		
		比重	投资额	折合美元
呵加刺	1000000 镑	1/2	500000 镑	2450000
有利	7500000 镑	1/4	1875000 镑	9187500
麦加利	800000 镑	1/3	350000 镑	1715000
汇丰	15000000 港元	全部	15000000 港元	7500000
德华	5000000 两	全部	5000000 港元	3500000
大东惠通	300000 镑	1/3	100000 镑	490000
中华汇理	2219000 港元	全部	2329000 港元	1164500
横滨正金	4500000 日元	1/2	4094000 日元	2087940
合计				28094940

资料来源:许涤新、吴承明(1990),第 127 页。

甲午战争前在外国间接投资中还有对保险业的投资和对中国政府的放贷。按吴承明先生的估计,到 1894 年止,在华设有机构的保险公司共有 9 家,投资总额合计 5992800 美元。¹²同一时期,外资银行或洋行还通过向中国政府放贷的形式进行间接投资,以获取利息。依据徐义生统计,自 1853 年至 1894 年共有贷款 46 笔,到 1894 年尚有 7 笔未清偿,仅这 7 笔未清偿的放贷,估计折合即达 10380349 美元,¹³可见当时通过放贷向中国进行间接投资的数量也并不很小。

甲午战争后,从外国对华投资的构成来看,间接投资有不断上升的趋势,在外国投资总额中的份额明显增长许多。有论者认为,从 1894 年到 1913 年间,中国长期资本流入量中外国对中国政府的贷款占据很大份额,1895、1912、1913 年,外国对中国的贷款约占长期资本流入量的 78%,而这三年外国对中国的直接投资只占 20% 左右。¹⁴目前,关于甲午战争后外国在华间接投资占这一时期外国总投资的比重,学术界争论还颇多,但是,不管间接投资具体的比重究竟是多少,有一点却是可以肯定的,即外国间接投资的份额上升是比较迅速的。

甲午战争后直到第一次世界大战,中国政府总是面临着赔款抑或财政正常开支的沉重压力,为了维持政府的正常运转,中国政府决定大举外债,向外国借款,因而,也就为外国的间接投资提供了市场,使得当时以向中国政府提供贷款方式进行的间接投资的数额,占到了外国在华投资总额的一半以

¹¹ 孙毓棠,《抗戈集》,中华书局 1981 年版,第 146—147 页。
¹² 许涤新、吴承明(1990),第 128 页。
¹³ 许涤新、吴承明(1990),第 132 页。
¹⁴ 王利华,《近代外人对华投资的影响因素剖析》,《南开经济研究》,1997 年第 2 期。

上。根据徐义生的统计，从甲午战争到第一次世界大战为止，中国历届政府向外国举借的贷款总额（单位：库平银两）分别如下：¹⁵

	笔数	贷款额	实收额
清政府所贷(1894—1911)	112	1203825452.94	660535961.08
北洋政府及南方政府(1911—1914)	77	565986935.89	347300073.28
合计	189	1769812388.83	1007836034.36

另据吴承明先生研究统计，从19世纪60年代中期至中华人民共和国建立前，外国通过向中国历届政府提供贷款（这里仅指军事财政贷款）方式进行的间接投资总额即达2500.3百万美元，¹⁶具体情况如下（单位：百万美元）：

时 期	国外贷款总额	年均贷款额
1865—1902	286.4	7.5
1903—1914	238.3	19.9
1915—1930	304.1	19.0
1931—1936	35.8	6.0
1937—1941	635.9	127.2
1942—1948	999.8	142.8
合计	2500.3	29.8

对铁路业及其他实业的间接投资。根据宓汝成的研究，截止到1914年底，外国对近代中国铁路建设进行间接投资（即贷款）修筑的铁路里程为4846公里，占当时铁路总里程的50.65%。具体而言，甲午战争后由外国间接投资修筑的铁路主要有：京奉铁路（英）、京汉铁路（英、法）、粤汉铁路（美、英、德、法）、津浦铁路（英、德）、吉长铁路（日）、广九铁路（英）等。投资国家除了美、英、德、法、日外，还有俄、比、荷等国，而以英国最多。第一次世界大战后，日本超过英国。从1898年至1927年，外国对中国的铁路贷款重要的有63笔。在对中国铁路进行间接投资的同时，外资还对当时的其他实业进行了间接投资，如煤矿、铁矿、通讯、航运等。发生在清政府覆亡前的除铁路以外的著名实业间接投资有：1899年的萍乡煤矿借款、1902年的汉阳铁厂借款、1908年的邮传部借款和汉冶萍借款、1910年的井陉煤矿借款等。到北洋政府时期，外国对中国实业的间接投资依然继续着，但与晚清时期有一个共同的特征，即在外国间接投资的总额中，绝大多数的对中国间接投资的份额是以中国政府军事、财政借款的方式发生的，实业性借款在北洋政府时期外国间接投资总额中占有1/4左右的比重。如整个北洋政

¹⁵ 这里所列出的贷款额系指有明确记载、可以统计到的数字，而一些资料不详的则不包括在内（徐义生（1962），第52、240页）。

¹⁶ 吴承明（1985），第47页。

府时期,实业借款额共计只占当时借款总额的23%左右,而在贷款的实收总额中也只占29%左右。¹⁷

在甲午战争后与日俱增的间接投资热潮中,外国的在华银行扮演了十分重要的角色,对间接投资的持续上升起了推波助澜的作用。正因为如此,甲午战争后,外国在华银行业发展很快。1895—1914年间,新设的西方银行约有十家,分支机构四十余处,尚有日本在华银行和中日合资银行,加上分支机构近三十处。这一时期在外国对华进行间接投资中最为活跃的银行是汇丰银行与华俄道胜银行。这样,到1914年,包括银行业、保险业在内的外国金融业在中国的投资估计总额为75572000美元。¹⁸

四、利润率、资本成本与资本吸收能力¹⁹

资本逐利、增殖的本性决定了资本的流动方向。在西方近代工业文明已经比较发达的时候,中国则还停留在传统落后的农业文明里,生产力低下、商业经济落后、社会贫困、人口众多,但所有这些在投资者的眼里却尽是商机,因为与此相对应的必然是资本短缺而投资空间大、劳动力与资源丰富而廉价、市场广阔而盈利机会多,也就是说,在这样的地区进行投资将会有—个非常不错的预期利润率。预期利润率的高低成为影响资本流向的最为重要的因素。此外,资本成本、资本的吸收消化能力、社会政治的稳定程度等都影响到资本的流动。

虽然投资于中国市场的预期利润率很高,但资本的流动及其变动趋势显然更受到实际利润率的影响,实际利润率成为外国投资额和投资方向的最最终决定因素。更何况,在近代中国从事投资的外国资本绝大多数是民间私人投资呢?私人投资对实际利润率高低更是十分敏感。仅以1913—1914年为例,在英国对华投资总额中,属于民间资本的投资总额即达1610百万美元,数量很大。²⁰再如美国,在其对外的投资总额中,长期的情况是民间投资的份额超过政府或国家投资的份额,即使到20世纪初,这种状况也没有改变,如1914年、1919年、1924年和1930年,政府或国家投资的份额仅占美国所有对外投资总额的30%、27.8%、27.8%和20%,国家投资份额不断下降,民间投资

¹⁷ 北洋时期的总借款额为1336996790.11银元,其中铁路借款为232326594.97银元,占总借款额的17.34%,航运电信借款额为38913570.60银元,占总借款额的2.90%,工矿借款总额为38135413.58银元,占总借款额的2.85%。在上述总借款额中,北洋政府实收到贷款976414254.08银元,上述三大实业实际收到的贷款分别为214264993.52、35538675.19、34399378.45银元,各占北洋政府实收到贷款总额的21.89%、3.63%、3.51%(徐义生(1962),第242—243页)。

¹⁸ 许涤新、吴承明(1990),第536页。

¹⁹ 本节分析参考了王利华先生发表在《南开经济研究》1997年第2期上的《近代外人对华投资的影响因素剖析》一文的部分内容。

²⁰ Svedberg, P., "The Portfolio-Direct Composition of Private Foreign Investment in 1914 Revisited", *Economic Journal*, 1978, 88(352), 763—777.

份额却不断上升。²¹虽然近代中国的潜在市场比较大，潜在需求也不小，但是，从已有的资料可以看出，外国在华直接投资所办的一些企业其实际利润率却并不高，远远低于预期利润率。如表6所示。

表6 1872—1932年中国几个行业外企利润率各区间分布状况(%)

利润占净值的% [*] (区间)	外贸行业	保险业	银行业	运输业	公用事业	制造业	其他	全部投资
亏损	5.0	0.8	1.7	7.3	3.2	10.4	6.8	5.0
0—4.99	10.0	3.3	6.8	9.1	4.8	18.3	22.7	10.3
5—9.99	45.0	10.6	37.3	40.0	32.3	20.0	34.1	25.9
10—14.99	20.0	16.3	32.2	23.6	32.3	21.7	15.9	22.6
15—19.99	15.0	17.9	13.6	18.2	14.5	12.2	6.8	14.4
20—24.99	0.0	20.3	1.7	0.0	9.7	7.0	6.8	9.0
25—29.99	0.0	13.0	1.7	0.0	3.2	4.3	2.3	5.2
30 以上	5.0	17.9	5.1	1.8	0.0	6.1	4.5	7.5
总数	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
0—10	60.0	14.7	45.8	56.4	40.3	48.7	63.6	41.2
0—15	80.0	31.0	78.0	80.0	72.6	70.4	79.5	63.8
序号	20	123	59	55	62	115	44	478

说明：^{*}利润是根据现代会计方法重新计算的，如折旧和坏账被视作花费，净值不仅包括实收资本，而且还包括所有结余。公开发表的利润数据未将折旧扣除，资本只包括实收资本，如表2孙毓棠先生使用公布的利润和资本数据。再如孙毓棠先生计算出，1879年上海煤气公司的利润率为23.4%（见孙毓棠，《中国近代工业史资料》第1卷，科学出版社1957年版，第181页），而根据现代会计方法计算的结果却是14%。

资料来源：（1）《北华捷报》，1873～1933年；（2）王利华，《近代外人对华投资的影响因素剖析》，《南开经济研究》，1997年第2期。

由表6可知，1872—1932年间在华直接投资的外国企业所获利润率大多在5%—20%之间，其中，约41%的企业利润率在10%以下，64%左右的企业利润率低于15%，对华直接投资利润率达到25%或在其上的大约只占总数的13%。值得注意的是，制造业的利润率极低，其半数左右的利润率低于10%。根据表7，可以算出外国企业在中国的加权平均利润率为13.38%。而1908—1937年英国国内直接投资的利润率并不比在中国从事直接投资的利润率低多少，有的年份甚至还比中国高。美国情形稍有不同，1900—1914年美国国内制造业直接投资的利润率较低，而1919—1928年其制造业的利润率却有大幅度提高，1919年甚至高达18.3%，如表7所示。

²¹ U. S. Department of Commerce, Historical Statistics of the United States, Part 2, pp. 868—869. 转见宫崎犀一等(1990)第63页。

表 7 英国与美国国内直接投资的利润率(%)

年份	英国		美国 (制造业)	年份	英国		美国 (制造业)	年份	英国		美国 (制造业)
	所有 股本	普通股			所有 股本	普通股			所有 股本	普通股	
1900			8.34	1912	10.88	14.47	6.20	1926	10.81	14.06	12.4
1901			7.72	1913	11.53	16.08	6.16	1927	10.84	14.07	9.5
1902			8.25	1914	10.37	13.72	5.60	1928	10.74	13.77	11.0
1903			7.57	1915	12.39	17.46		1929	10.56	13.44	
1904			6.45	1916	12.93	17.86		1930	8.68	10.55	
1905			7.09	1917	12.88	17.43		1931	5.73	6.18	
1906			8.34	1918	13.67	18.62		1932	5.72	6.41	
1907			7.90	1919	14.60	19.92	18.3	1933	6.69	7.92	
1908	6.83	8.38	6.07	1920	13.24	17.58	12.3	1934	8.04	9.80	
1909	7.74	9.40	6.89	1921	6.81	7.75	2.9	1935	9.05	11.51	
1910	9.35	11.97	7.26	1922	9.05	11.18	10.2	1936	10.62	13.66	
1911	9.95	12.82	7.10	1923	9.94	12.67	11.2	1937	12.10	15.83	
				1924	10.68	13.77	10.0				
				1925	11.37	14.91	12.1				

资料来源：(1) Hargreaves Parkinson, “British Industrial Profits”, The Economist, Dec. 17, 1938. pp. 597—603.
(2) E. I. Epstein and R. A. Gordon, “Profits of Selected American Industrial Corporations, 1900—1914”, Review of Economic Statistics, Feb. 1939, pp. 122—128.
(3) Ralph C. Epstein, Industrial Profits in the United States, New York, 1934, p. 56.
(4) 王利华,《近代外国对华投资的影响因素剖析》,《南开经济研究》,1997 年第 2 期。

根据表 7 可以计算出，1908—1937 年间英国全部资本的平均利润率为 10.13%，普通资本的平均利润率为 13.11%；1900—1928 年间美国制造业投资的平均利润率为 8.68%。对照它们在华直接投资所获得的利润率或平均利润率，近代外国在华直接投资的实际利润率并不像预期的那样高，与在其国内投资相比也未见得高出很多。

在实际的投资中，投资的实际利润率低于预期利润率是经常发生的一种很正常的经济现象，不正常的是，一些外国投资者发现，其在华投资获得的实际利润率甚至也并不比在他自己的国家投资所获得的利润率高，这对他们后来决定资本的流向影响很大。如 1920—1938 年间驻华英美烟公司的账面利润率在 14% 左右，²²而同期前后美国卷烟工业年均利润率亦达 13.9%。²³

资本成本对资本的流向也有重要的影响。一般而言，资本成本的高低受资本的可获得程度、市场分割或市场一体化程度、汇率风险、货币状况和国

²² 上海社会科学院经济研究所编：《英美烟公司在华企业资料汇编》第四册，中华书局 1983 年版，第 1597 页。
²³ E. B. Alarfer, Economics of American Industry, New York, 1957, p. 636.

别风险（包括政治局势、政府政策、文化差异等）等因素的影响。中国的投资环境不太理想，可供投资用的资金不多（即资本可获得程度不高），资本市场发育不良且处于分割状态，货币单位繁杂，政治动荡不安，与西方存在很大的文化差异等等，所有这些因素都促使资本成本提高，从而降低了资本收益率，最终导致较低的投资额。这里仅以资本的可获得程度、市场分割和市场一体化程度对资本成本的影响为例，说明资本成本对投资决策的影响。

资本成本等于权益资本和债务资本的加权平均成本，即

$$K = k_e[E/(E + D)] + k_d[(1 - T)D/(D + E)]$$

上式中 K 表示企业资本的加权平均成本或资本成本， k_e 表示税后权益资本成本， k_d 表示税前债务资本成本， T 表示边际税率， E 表示权益资本的市场价值， D 表示债务资本的市场价值。此式表明，资本成本不仅受权益资本成本的影响，而且也受债务资本成本的影响。这两种资本成本都受资本市场各种风险的影响。

国内企业的资本可获得程度是由国内资本市场的供给和需求决定的，国内企业按其最佳资本结构（即资本成本最低时的债务资本和权益资本的比例）筹集资金，但由于筹集的资本仅限于国内市场，当筹集的资本绝对额超过一定水平时，边际资本成本就会上升。进行国际投资的企业其资金来源不限于某一国内市场，它们除在母国市场筹资外，还在国际资本市场和所在国资本市场筹资。若国际资本的可获得程度高，企业不仅能在资本结构不变的条件下增加集资能力，而且还能筹集到成本比国内企业更低的资本。

因此，资本可获得程度和资本市场一体化程度越高，资本成本就越低，资本预算越多，投资额就会增加；反之，较低的资本可获得程度和资本市场被分割必然带动资本成本上升，减少资本预算，投资额也就必然会降低。

近代中国由于未能形成较为成熟的资本市场、由存款积蓄转化而来的资本不多、将十分有限的储蓄转化为投资的机制也很不健全，加之中国实行银本位而西方发达国家实行金本位，使得中国货币无法纳入国际货币体系，从而致使在近代中国的市场上筹资的成本一直很高。此外，近代中国的政治高风险、中西文化的差异以及中国普遍的排外情绪都增加了外国在华投资的资本成本，甚至有时有些外国直接投资的企业也会出现在亏损状态下经营的情形。²⁴

除了上述的实际利润率、资本成本的影响外，一国对资本的吸收能力的大小也直接关系到资本的盈利程度和后续资本的流动方向。一般来讲，影响资本吸收能力大小的主要因素有：人口的数量、消费需求、人力资本质量、

²⁴ 侯继明，《关于外人在华投资的“压迫论”（1895—1937年）》，载《中国近代经济史研究资料》第5辑，上海社会科学院出版社1986年版，第111—129页。

生产力水平、民族文化传统等。

人口数量与消费需求对投资的影响主要表现在产品的市场需求和产品市场实现等方面。古典经济学家马尔萨斯、西斯蒙第等对此早有论述。人口的多少直接关乎产品市场需求的多少,两者一般是成正比的。也正因为这个原因,在中国的大门被打开之后,外国的资本与商品蜂拥而入。同样,在历史上,以英国为代表的西方发达国家的绝大多数国外投资投向了美洲,是因为在美洲的开发过程中,有大量的移民人口进入了美洲,从而创造了极大的市场需求。近代中国虽然有四万万之众,但并未产生很大的市场需求。

有人将近代中国未产生巨大市场需求、进而未引致外国投资规模迅速上升的原因,部分归结为伴随着西方坚船利炮未能形成一个向中国移民的热潮。我们认为,这个分析有失偏颇。事实上,移民也是一种投资,移民热潮之所以能形成是因为移民移往的国家或地区能够提供更多、更好的创业与赢利机会。近代中国之所以未形成外国人移往中国的热潮,是因为中国不具备这样的条件或机会。没有移民的国家或地区,其市场需求的扩大可以借助于内在的原居民的消费需求的扩大而扩大。一国居民消费需求的扩大,受到诸如人均国民收入、居民可支配收入、边际消费倾向等因素的影响,近代中国消费需求不旺,显然与中国的贫弱以及居民可支配收入低关联在一起的。没有旺盛的市场需求,意味着产品市场价值实现的难度加大,从而影响到投资者生产能力的利用和发展,直至影响其赢利和投资意愿。

人力资本质量对资本的吸收能力大小影响最为直接,低水平的人力资本质量直接影响着生产力水平的发挥与发展,相反,高质量的人力资本却可以产生良好的“溢出”效应并获得一个增长“余值”,进而产生示范效应,促进投资的进一步增长。近代中国不论是在甲午战争前还是在其后,人力资本的状况不仅没有得到多大改善,反而因社会的动荡而每况愈下,这样的人力资本状况决定了其对外国投资的吸收能力将是十分有限的。

一个民族的文化传统也影响着资本的吸收能力。中国几千年一以贯之的农本思想、小农意识、自给自足的自然经济等,使得中国文化传统对外来势力具有很强的排斥性,既影响了外国投资设厂,又影响了市场扩大。

实际利润率、资本成本以及资本吸收能力的大小、高低对外国直接投资的影响十分巨大,在这些因素共同作用下,外国在华投资的方向作出了比较大的调整,即间接投资在总投资中的份额持续上升。²⁵这是因为有利于外国在

²⁵ 关于近代外国在华投资的比例,即直接投资与间接投资孰重孰轻的问题,学术界的观点并不一致,有不少学者认为直接投资的份额更多一些,如雷麦认为,直接投资占总投资的份额 1902 年为 63.9%、1914 年为 67.3%;吴承明认为,直接投资占总投资的份额 1902 年为 65.1%、1914 年为 66.3%。但也有不少学者的观点正好与他们相反,认为间接投资的份额更多。本文作者将外国的借款视为间接投资,因而,也持间接投资份额更多一些的观点。我们认为,之所以出现判断观点上的相左,是因为每个学者在分析时候的统计口径不一致而造成的。

华进行间接投资（主要是借款）的因素更多，主要有：

利率高。以借款为主的外国间接投资，几乎全部是以高利率为条件的。外国向中国提供的贷款，年利率一般是4%—5%，铁路借款则有不少高达7%，而到第一次世界大战期间，更是高达8%—10%。这样的利率，即使在当时的国际金融市场上，也属于比较高的。

回扣大。除了比较高的利息收入外，回扣高也是外国间接投资不断增长的一个重要原因。高额的回扣也构成了外国投资者的利润。当时回扣小者为5.875%、大者为17%。如1896年的“英德借款”1600万镑是按94%实付，债券分两部分发行，1000万镑于该年3月按98.75%发售，600万镑于同年9月按99%发售；发行银行获得77.5万镑的现实利润（即回扣）。再如“善后大借款”，总额为2500万镑，按84%付款，债券按90%发行，发行的银行团净赚150万镑，手续费约30万镑。²⁶

风险小。直接投资由于绝大多数是自己投资、自己经营，生产出来的产品能否迅速实现其价值或有多少能实现其价值，却不是由自己决定的，而是由市场来决定，因而存在一定的市场风险。而间接投资，尤其是借款，由于是以国家信用做担保、由当时中国最为稳定的几个大的税种的税收来做抵押担保，因而，风险很小。风险小而收益高，外国投资者又何乐而不为呢？

成本低。由于直接投资需要调研、组织、设计和管理，所以相应地需要支付各项费用，致使生产成本中非直接生产性成本不断上升。而间接投资因为并不涉及生产与管理，因而，省去了直接投资所需支付的一切交易成本、管理成本和监督成本等，有利于投资总收益的提高和增长。

利润率、资本成本与资本吸收能力对直接投资的抑制，利息与回扣、风险与成本对间接投资的诱致，使得外国投资的结构在甲午战争后发生了明显变化：直接投资份额下降，间接投资份额上升。但是，这种变化也不是绝对的，不论是直接投资还是间接投资的变动都是非线性的、无规律的，呈现不规则的变动，究其原因，恐怕与近代中国尤其是民国建立之后的中国现实国情关联在一起的。由于近代中国是个政治动荡不安的国家，而且市场又被严重分割，极高的政治风险和不良的市场状况决定了外国对华投资变动的可能趋势。

五、外国投资增长中的中资企业发展

在外国投资的不断增长中，中资企业发展状况又呈现怎样一种情景呢？对外国投资增长背景下的中资企业发展情景的分析，有助于全面认识外国投资的性质与效应。事实上，国内绝大多数经济史学家将近代与日俱增的

²⁶ 许涤新、吴承明(1990)，第567—568页。

外国投资定性为经济侵略，并将其视为阻碍近代中国经济发展的因素之一。²⁷而很少或几乎没有学者认为外国投资只是一种经济行为，是资本逐利增殖本性的外在表现，其对近代中国的经济发展也有正效应。实际上，应对近代外国在华投资的性质与效应做出具体分析，不可概而论之，否则，将很难辨明是非曲直。

首先分析外国在华投资总额占其国际投资总额的比例。到第一次世界大战结束之前，英国无疑是世界上最为强大的国家，其对外投资构成了当时国际投资的最主要部分，如 1885 年、1900 年和 1913 年，主要国家对外投资额分别为 12000、22000 和 44000 百万美元，其中英国分别占 7000、11000 和 18300 百万美元；其次是法国，其对应年份的投资额分别为 3500、5600 和 8700 百万美元²⁸。既然英法两国是当时世界上对外投资最多的国家，那么，只要分析这两个国家对外投资在世界范围内地区的分布，便可对其在华投资的性质做出一个一般性的、浅层次的判断。表 8 为 19 世纪末至 20 世纪初英法两国对外投资的地区投资额分布状况。

表 8 19 世纪末至 20 世纪初英法对外投资的地区分布

单位 100 万美元 %		
地区	英国(1896 年)	法国(1902)
欧洲	1250(10.4)	4100(70.6)
北美洲	2250(18.8)	150(2.6)
美国	1550	100
加拿大	700	50
拉丁美洲	1250(10.4)	650(11.2)
大洋洲	1600(13.3)	n. (n.)
亚洲	1700(14.2)	200(3.4)
印度、锡兰	1450	...
中国	150	150 *
非洲	1900(15.8)	700(12.1)
其他地区	2050(17.1)	...
合计	12000(100)	5800(100)

注：* 包括法属殖民地和菲律宾的大约 3000 万美元。

资料来源 转见宫崎犀一等(1990) 第 58 页。

根据表 8 可以计算出，1896 年英国对华投资数量只占其对外投资总额的 1.25%，1902 年法国对华投资数量也只占其对外投资总额的 2.07%；到 1914 年第一次世界大战爆发前，英国、法国、德国和美国等几个主要国家对华的

²⁷ 近代外国投资被定性为经济侵略，持这种观点的学者有：吴承明（许涤新、吴承明，1990）、汪敬虞（汪敬虞主编，《中国近代经济史》，人民出版社 2000 年版）、郑友揆（1984）等。而国外的研究者则持相反的观点，认为并非仅仅是经济侵略，如侯继明（Hou Chi-ming, *Foreign Investment and Economic Development in China, 1840—1937*. Cambridge, Mass: Harvard University Press, 1965）、费正清（费正清编：《剑桥中国晚清史》，中国社会科学出版社 1993 年版）等。

²⁸ Brown, M. B., *Economics of Imperialism*, Penguin, 1974, p. 171.

投资额分别为 600、150、250 和 50 百万美元，各占其对外投资总额的 3%、1.66%、4.31% 和 1.43%。²⁹ 从 19 世纪末到 20 世纪 40 年代前，上述几个国家对外投资主要分布如下：英国集中在北美洲和亚洲、法国和德国集中在欧洲和美洲、美国集中在北美洲和拉丁美洲。英国在亚洲进行国际投资的分布具体状况如表 9 所示。

表 9 19 世纪末至 20 世纪 40 年代前英国在亚洲国际投资分布

地区	单位：百万美元 %		
	1896 年	1914 年	1938 年
印度、锡兰	1450(85.29)	1850(52.11)	3050*(35.47)
中国	150(8.82)	600(16.90)	850(9.88)
亚洲	1700(100.00)	3550(100.00)	8600***(100.00)

注：* 包括缅甸在内。*** 还包括大洋洲在内。

资料来源：根据宫崎犀一等(1990)第 58—60 页表格数字计算而得。

根据表 9 也能看出，英国在亚洲地区进行的国际投资中，其对中国的投资所占份额一直不是很高，最高的时候也未超过 20%。一般来讲，要通过投资对一个国家进行经济侵略，则侵略国应集中其国际投资的绝大多数份额投资于被侵略国，否则，要形成经济侵略将是比较困难的。是不是由于中国经济的存量过分薄弱、是不是由于中国近代工业文明太少，而只需要较少的投资就可以形成经济侵略并控制中国的经济命脉呢？这需要进行进一步的探讨后方能回答，但如果对照英国在印度的投资情况，将更有利于问题的深入分析和思考。

其次分析外国在华投资的结构与行业分布。正像前面分析所指出的一样，外国投资的结构分为直接投资与间接投资，如果将外国金融机构向中国政府提供的贷款视为投资，那么，甲午战争后外国在华的间接投资比例是不断上升的。就这些以贷款形式出现的间接投资而言，由于附有十分苛刻的政治条件，使得中国政府在获得这些借款后，实际上跌入了别人早已设计好的陷阱，被人左右，任人宰割。从这个角度而言，以贷款为主要方式的外国间接投资具有很强的欺骗性、奴役性和侵略性。这一点是不容置疑的！

但是，从另一个角度来看，在近代中国的外国投资中，不论是直接投资还是间接投资，其投资的主体始终是外国的民间资本而非外国的国家或政府资本，甚至许多附有苛刻条件的借款也是由外国民间资本扮演主角的。既然是民间资本成为了外国投资的主体，而民间资本除了最具有资本的原初特性（即逐利、增殖）外，还有一个特点就是分散性与盲目性。分散性与盲目性这一特点决定了民间资本的投资不可能是有意识的、有目的地去对一个国家进行经济侵略，其在东道国某些行业投资的相对集中，很可能是因为这个行业的产品或服务有着非常强烈的市场需求或严重的供给不足，即有很好的市场利润前景。在市场利润这个杠杆的引导下，资本迅速在某些行业聚集，形成

²⁹ 宫崎犀一等(1990)第 59 页。

对一国某些重要行业或部门的控制与垄断。如果外国的投资及其聚集，仅此而已，这仍只能视为是市场的自发行为而非目的性很强的经济侵略。当然，必须要看到的是，近代外国投资的大量进入是在不平等条约保护下并享有许多经济政治特权的背景下发生的，这就使得外国投资的性质有了很大的变异，因此，从这个角度而言，近代外国在华的投资具有一定的侵略性。

表 10 为近代外国在华的外资厂矿和中外合办厂矿的情况，从这个表中大致可以看清外国投资在华的行业、地区分布，由此可以对外国在近代中国的投资是否形成经济侵略、进而控制中国经济命脉做出自己的判断。

表 10 1895—1913 年设在中国的外资厂矿和中外合办厂矿的国别、产业、地点和创办资本的规模

		厂 矿 数				创办资本(1000 中国元)			
		外资 (A)	中外 合办(B)	A + B	总数 的%	外资 (A)	中外 合办(B)	A + B	总数 的%
国 别	英国	28	9	37	27.21	25465	24216	59681	48.16
	法国	4	2	6	4.41	2986	1609	4515	4.38
	德国	9	3	12	8.82	6534	1147	7681	7.45
	日本	35	14	49	36.03	19829	6501	26330	25.53
	俄国	9	8	17	12.50	3866	2782	6648	6.44
	美国	6	2	8	5.88	2551	689	3240	3.14
	其他	5	2	7	5.15	3028	1950	4978	4.83
	合 计	96	40	136	100.00	64259	38894	103153	100.00
行 业	采矿	10	22	32	23.53	18533	31436	49969	48.44
	工程与造船	7	0	7	5.15	2895	0	2895	2.81
	电力与自来水	16	3	19	13.97	10772	742	11514	11.16
	纺织	12	4	16	11.76	10325	2190	12515	12.13
	食品	34	5	39	28.68	15126	2022	17148	16.62
	其他	17	6	23	16.91	6608	2504	9112	8.83
	合 计	96	40	136	100.00	64529	38894	103153	100.00
地 点	江苏	33	11	44	32.35	18501	5121	23622	22.90
	直隶	7	6	13	9.56	4645	24309	28954	28.07
	湖北	8	2	10	7.35	3576	1609	5185	5.03
	山东	5	1	6	4.41	3503	92	3595	3.49
	满洲	37	16	53	38.97	18680	5332	24012	23.28
	其他	6	4	10	7.35	15354	2431	17785	17.24
	合 计	96	40	136	100.00	64259	38894	103153	100.00
创 办 资 本 规 模	不到 100000	2	5	7	5.15	120	288	408	0.40
	100000—249999	32	15	47	34.56	5021	2213	7234	7.01
	250000—499999	30	7	37	27.21	11623	3132	14755	14.30
	500000—999999	17	6	23	16.91	12887	4088	16975	16.46
	1000000—4999999	14	5	19	13.97	20622	7541	28163	27.30
	超过 5000000	1	2	3	2.21	13986	21632	35618	34.53
	合 计	96	40	136	100.00	64259	38894	103153	100.00

资料来源：（1）根据汪敬虞编：《中国近代工业史资料，1895—1914 年》第二辑第一册，中华书局 1962 年版，第 2—13 页计算。（2）转见费正清编，《剑桥中国晚清史》，下册，中国社会科学院历史研究所编译室译，中国社会科学出版社 1993 年版，第 41—42 页。

表10清楚地显示出，甲午战争后直到第一次世界大战爆发前，外国在华投资行业比较多的为采矿业和食品业，采矿业占外国在华设厂总数和投资总额的对应份额分别为23.53%与48.44%、食品业则为28.68%与16.62%。除掉外国直接投资设厂和中外合资办厂外，其余的份额应该是属于包括官办资本企业在内的中资企业的，这样算来，在甲午战争后，采矿业中有76.47%的企业和51.56%的投资额、食品业中有71.32%的企业和83.38%的投资额是属于中资企业的，外国投资在最为集中的两个产业上都未达到绝对垄断的地步，由此可见，外国投资即使有经济侵略，也没有曾经论述和想象的那么严重。

最后来看外国投资不断增长背景下的中资企业发展情况。学者侯继明认为，随着外国人在1895—1937年间增加了其工业投资，“中国经济的现代部分出现了华人企业与外国企业长期共存的情况……（而这种）共存……表明，从长远看，（华人企业）能够与（外国企业）发展一样快。”³⁰学者高家龙在研究了1890—1930年英美烟公司与南洋烟草公司的竞争、发展情况后，认为这两家公司不仅仅是竞争对手，而且都是非常成功的大企业，是由于两者的直接竞争导致了它们的成功。³¹甲午战争后，由于外国人取得了在华投资办厂的合法权，从而导致了在许多产业中都存在着中外企业竞争的情况：棉纺市场上有华人申新纱厂与日本内外棉株式会社之间的竞争；纺织机械市场上有华人大隆机器厂与两个美国制造业者——萨科—洛厄尔机器厂和维定机器厂之间的竞争；火柴市场上有大中华火柴股份有限公司与两个外国的公司——瑞典火柴公司和日本铃木会社之间的竞争；图书和出版市场上有两个华人出版社——商务印书馆和中华书局与英人兆祥洋行之间的竞争；肥皂、制碱与肥料市场上有华人永利制碱公司和五洲皂药厂与英国利华兄弟公司、卜内门洋碱有限公司和帝国化学工业公司之间的竞争；水泥市场上有华人广东水泥厂与在香港的英国青洲英坭公司之间的竞争；钢铁市场上有华人汉阳铁厂与日本在鞍山和本溪的南满株式会社之间的竞争；烟草市场上有华人的南洋兄弟烟草公司与外国的英美烟公司之间的竞争；等等。中资企业与外资企业之间的竞争几乎遍及当时中国的各个产业与行业。

虽然创办企业和发展企业的市场环境充满了竞争和不确定性，但是市场尤其是企业发展的活力也正来源于此。如果没有竞争，企业就将失去发展的动力。表面上看，似乎没有外资企业的进入和竞争，中资企业就可以从从容容地去发展，并能最大化地从事资本积累而逐步扩大，但事实上，在外资企业进入还不是很多的甲午战争前，中资企业并没有发展起来。更有，如果没

³⁰ Hou Chi-ming, *Foreign Investment and Economic Development in China, 1840—1937*. Cambridge, Mass.: Harvard University Press, 1965, pp. 138—139.

³¹ 高家龙：《中国的大企业——烟草工业中的中外竞争》，商务印书馆2001年版，第318页。

有外资企业的进入及其对中国利权的攫取，中国政府能否进行允许中国人办企业的制度变迁还不可预测！几千年一以贯之的“重农抑商”制度安排决定了，中资企业即使没有外资企业的市场竞争与压力也将很难发展起来。中国的历史实际上早已为此做出了最好的注释。反观甲午战争后市场竞争日益加剧的情景下，反而是近代史上中资企业成长与发展比较好的时期（如表 11 所示），实在值得深思！

表 11 1895—1913 年开业的中资制造业和采矿业的企业数和创办资本(按行业)

资本单位 1000 中国元

行业	企业数	占总数的%	创办资本	占总数的%	平均资本
煤矿	42	7.65	14508	12.06	345
金属采矿和冶炼	39	7.10	7565	6.29	194
金属加工	15	2.73	2787	2.32	186
公用事业(水电)	46	8.38	21600	17.96	470
水泥	3	0.55	2620	2.18	873
砖瓦	12	2.19	651	0.54	543
陶器	7	1.28	772	0.64	110
玻璃	10	1.82	3429	2.85	343
火柴	26	4.74	3444	2.86	132
蜡烛、肥皂	18	3.28	805	0.67	45
轧花	3	0.55	280	0.23	93
纺纱	19	3.46	10454	8.69	550
织布、印染	27	4.92	1261	1.05	47
缫丝	97	17.67	11584	9.63	119
毛织	7	1.28	5215	4.34	745
麻织	4	0.73	1000	0.83	250
其他织布	6	1.09	732	0.61	122
碾米	9	1.64	1012	0.84	113
制面	53	9.65	8622	7.17	163
榨油	28	5.10	4752	3.95	170
纸烟、烟草	20	3.64	1378	1.15	69
其他食品	15	2.73	3111	2.59	207
造纸	14	2.55	5929	4.93	424
印刷	6	1.09	1160	0.97	193
鞣革	11	2.00	4608	3.83	419
其他	12	2.19	1009	0.84	84
合计	549	100.00	120288	100.00	219

资料来源 根据汪敬虞编，《中国近代工业史资料，1895—1914 年》第二辑第二册，第 869—920 页计算。

从表 11 来分析，中资企业不仅行业分布十分分散而且资本规模都不大。若从企业数量来看，中资企业最为集中的行业依次是缫丝、面粉、煤矿和公用事业（水电）；而若从资本规模来看，中资企业资本集聚最多的行业又依次是公用事业（水电）、煤矿、缫丝和纺纱。综合来看，在企业数和创办资本这两方面，中资企业的大部分集中在纺织业等轻型制造业（共 160 个，占企业总数的 29.14%；资本额共 30246000 中国元，占总资本的 25.14%）和食品加工业（计 125 个，占企业总数的 22.76%；资本额共 18875000 中国元，占总资本的 15.69%）这类行业。采矿业占总企业数的 14.75%，占总资本的 18.35%。对照表 10，如果仅从企业数量与资本数量分布而言，中资企业与外资企业呈现了类似的特征。如果从投资的总规模和企业平均资本来比较，则中资企业与外资企业的差别似乎也不是很大。具体如表 12 所示。

表 12 1895—1913 年中资、外资*企业规模比较

创办资本规模	企业数量及各占份额				资本数量及各占份额				平均资本规模	
	企业数		占总数的%		创办资本		占总数的%			
	中资	外资	中资	外资	中资	外资	中资	外资		
10000—49999	214		38.98		5899		4.90		28	
50000—99999	89	2	16.21	2.08	7052	120	5.86	0.19	79	60
100000—249999	104	32	18.94	33.33	17732	5021	14.74	7.81	170	157
250000—499999	85	30	15.48	31.25	29901	11623	24.86	18.09	352	387
500000—999999	40	17	7.29	17.71	27980	12887	23.26	20.06	700	758
1000000—4999999	17	14	3.10	14.58	31725	20622	26.37	32.09	1866	1473
5000000 以上		1		1.04		13986		21.77		13986
合计	549	96	100.00	100.00	120288	64259	100.00	100.00	219	669

注：* 外资企业不包括中外合资企业。

资料来源 根据费正清编，《剑桥中国晚清史》，下册，中国社会科学院历史研究所编译室译，中国社会科学出版社 1993 年版，第 41—42、48 页表格资料计算而得。

综合表 10、表 11 和表 12 来看，在甲午战争后外国投资绝对量明显增多的情况下，中资企业发展的速度和规模是非常快的，外国投资对中资企业发展的挤压之势似乎也并未形成，或即使已经形成但并不十分严重，而且，这种挤压似乎更应该理解为构建在不平等条约基础之上的市场竞争。中国的经济史学家们认为这种形式的中外竞争是不平等的，并列出了外资企业较中资企业拥有的有利条件，如易于获得外资、尖端技术、企业人才、管理效率以及政治和外交特权等以证明这种观点。但国外的学者则不赞成此观点，认为中外竞争是平等的，他们也承认外资企业可能拥有上述各种优势，但他们又认为这些优势被中资企业所拥有的有利条件抵消了。侯继明认为，外资企业

具有“更多的资本，更大的资本密集程度，先进的技术，不受官方干涉”等优势，但中资企业也独享某些优势，如“了解本地情形、民族主义和企业易于转向”。³²帕金斯认为，外资企业可能“资本成本较低……但是这一优势被华人企业中更廉价的劳动力成本和更大的劳动强度所抵消。”³³吉尔伯特·罗兹曼认为，外国投资有利于中资企业的仿效行为发生，并且认为，“任何企业家只要能有效地制造并销售其产品，就能获取利润。外国企业家在现代技术、信贷以及对国际市场行情的掌握方面，优于中国企业家，但中国企业家在雇佣劳动力、获得本地产销信息方面同样也优于外国企业家。中外企业家究竟谁最有优势，实际上要取决于变幻莫测的市场情况。”³⁴关于竞争平等与否的争论虽然难以达成一致，但甲午战争后中资企业的蓬勃发展却是有目共睹的事实。基于以上的分析，似乎可以得出一个一般性的结论：在外国投资与日俱增的情形下，中资企业的发展不仅未受到太大的约束和阻滞，而且在外资企业扩张的冲击下也成长迅速。

六、近代外国在华直接投资的效应

外国间接投资由于绝大部分是以借款的形式出现的，这样的外国投资加重了中国经济发展的灾难，具有很强的掠夺性，对近代中国经济发展没有任何的正效应可言，因而，不再予以分析。至于以铁路借款等“实业借款”形式出现的间接投资，虽然也具有一定的负面效应，但由于这些间接投资对近代中国基础设施的改善和供给、对中国实业的发展有一定的促进作用，所以，将其连同外国直接投资的效应一并加以分析。

外国直接投资是紧随坚船利炮来到中国的，由外国直接投资而创设的外资企业，是中国大地上出现的第一批使用机器生产的近代工业文明。尽管外国直接投资对近代中资企业的正常成长有一些影响，“尽管有外资企业和进口商品的竞争”，但也应该看到，“中国现代部门的工业仍在扩展”，³⁵外国在华投资在客观上对中国近代生产力的产生和发展，对中国早期工业化的启动和发展，对中资企业的成长和壮大，对市场的扩大和市场制度的建设等也是有推动力的，³⁶因而，具有积极的正面效应。

³² Hou Chi-ming, *Foreign Investment and Economic Development in China, 1840—1937*. p. 154.

³³ Perkins, Dwight H., *China's Modern Economy in Historical Perspective*. Stanford University Press, 1975. p. 45.

³⁴ 吉尔伯特·罗兹曼(1988),第423页。

³⁵ 吉尔伯特·罗兹曼(1988),第420页。

³⁶ 陈谦平等:《近代英国在华直接投资评析》,载吴景平等主编:《近代中国的经济与社会》,上海古籍出版社2002年版。

1. 资本的溢出效应

外国投资增加了中国社会的资本流量，有利于缓解近代中国社会资本的短缺。由于资本严重不足，导致中国资金市场求大于供，银行放款利率居高不下。根据日本人的调查，1910年中国银行放款利率一般在10%左右，最高达20%。³⁷这时工厂融通资金的主要来源是钱庄。辛亥革命以前，上海钱庄放款年利率大抵在7.2%—12%左右，银根紧张时甚至在20%左右。上海以外的各埠钱庄放款利息“多数比上海为高”。³⁸1920年左右，中国一些城市放款利率又有所上升。从北洋政府财政部档案材料看，当时京、津一带资金短缺，“所有各项利率均较前增高”。其中以“救济农工”相标榜的农工银行，1917年以来放款利率最高为18%，而到1921年时这一上限已升至21.6%。³⁹高利率一方面侵蚀原有的产业资本，另一方面又吸引社会资金流向流通领域，而使生产领域资本形成更加不足。这样形成一个“低资本形成—低收入—低储蓄能力—高利率—低资本形成”的恶性循环。而从资金的需求者，即生产企业角度来看，甲午战争后近代工业发展最快的是棉纺织业，其中纺纱业的发展又超过织布业，其原因之一是创办纺纱厂“资本可以稍轻”，其规模可以“随其资本之大小而设立”，⁴⁰对于资本实力微薄的实业家来说，选取纺纱厂作为投资对象，就不难理解其苦衷。与此形成鲜明对照的是同一时期英国银行贷款利率的低下。当时在华英国银行的放款利率一般在4%—8%之间，⁴¹而英国国内银行的放款利率在1900—1908年间更低，平均不过3.4%。⁴²大量价格低廉的外国资本的流入为中国“现代化经济部门筹措了很大一部分资金”，⁴³这些资金对近代中国经济发展的促进作用是不言自明的。

2. 技术与管理技能的正外部效应

所谓技术的正外部效应主要是指投资主体在从事生产或其他经济行为时，并非有意转让或传播其技术，而是在其经济行为中自然输出技术及其引起的技术提高。它的主要形式有：在外国投资企业受过训练并有经验的人员流动、外国投资者的竞争压力迫使当地生产厂家提高自己的技术水平、仿效外国投资者的新产品和新工艺等。所谓管理技能的正外部效应是指东道国可以借助外国所设公司的人员流动，使东道国有关人员获得管理经验；借助东道国企

³⁷ 严中平(1955)，第144页。

³⁸ 严中平(1955)，第146—147页。

³⁹ 中国第二历史档案馆编：《中华民国史档案资料汇编》，第三辑，金融（一），江苏古籍出版社1991年版，第420—424页。

⁴⁰ 张国辉(1988)。

⁴¹ 张继凤(1982)。

⁴² 转见严中平(1955)，第161页。

⁴³ 侯继明：《外国投资和经济现代化》，载张仲礼主编：《中国近代经济史论著选译》，上海社会科学院出版社1987年版，第52页。

业模仿与之竞争的外国投资者的管理技术,提高组织技能。吉尔伯特·罗兹曼说:外国直接投资“能使中国企业家学会新的技术,借到资金且仿效外国企业的一套经营管理程序”。⁴⁴

3. 进口替代效应的增强

鸦片战争以后,洋货进口不断增加,机制品市场日益扩张。在洋货行销的刺激下,中国的仿洋机制品工业产生了。洋货多销一分,利益多外流一分,于是商战、实业救国思潮掀起,民族替代工业纷纷崛起。而外国在华直接投资则进一步增强了这种进口替代效应。以卷烟业为例:卷烟是近代中国进口之大宗商品,民国以后进口值增长较快,从民国初年的每年一千多万元增长至 20 世纪 30 年代的六千多万元;尤其是英美烟公司在华经营的成功,刺激了国货卷烟工业的规模也不断扩大,从民国初年的十余家,发展到 20 世纪 30 年代的一百多家,成为中国近代民族工业的一大支柱产业。其他如肥皂、造纸、电灯、酸碱、铅笔、化妆品、电扇、甘油等国货工业都是以进口替代的形式切入市场而发展起来的。据 1936 年《中国经济年鉴》记载,中国近代工业有食品、棉纺织、碱、酸、盐等 18 个主要行业,其中 14 个是进口替代行业,占主要行业总数的 77.8%。再对国货工业的地域分布进行分析,这一情况将会看得更加清楚:口岸地区,特别是上海,是外国在华直接投资的主要地区,而民族工业的产值也在 50% 左右,是近代民族工业无可争议的经济中心。进口替代民族工业的发展是与外国直接投资设厂的刺激有着密切关联的。

4. 交通等基础设施建设的带动效应与溢出效应

基础设施供给的多少直接影响到一国的经济发展,基础设施的建设与改善不仅可以降低经济发展的交易成本、运输成本,而且还可以获得积聚效应、规模效应。中国传统的运输工具是木帆船,而随着外国投资航运业,以蒸汽为动力的轮船开始在中国出现,从而引发了中国航运业的一场革命:汽船开始逐步取代帆船。据研究,到 19 世纪 60 年代,帆船与汽船各半;至 90 年代,帆船运载量只占总吨数的 1/30,并继续呈现不断减低的趋势。轮船在运载量和速度上均优于帆船,对市场的扩大以及运输成本的下降是很有利的。再如铁路建设,从 1903 年至 1907 年,外国直接投资兴建的铁路计 3772 公里,外国间接投资(即中国政府借款)修建的铁路为 4846 公里,在外国投资的带动下,中国政府也开始自筹资金增加铁路的供给,这样,到 1913 年,中国建成通车的铁路里程达 9854 公里,改变了二十多年前中国无一寸铁路的局面。⁴⁵ 铁路的

⁴⁴ 吉尔伯特·罗兹曼(1988),第 423 页。

⁴⁵ 安格斯·麦迪森(1999),第 77 页。

建设与通车，既给铁路沿线的经济发展带去了商品经济观念、扩大了市场，又方便了货物的运输、节约了成本。由于铁路建设对经济发展的成效显著，此后的中国政府继续利用内外资进行建设，以增加铁路的供给。据史载，到20世纪30年代，国内共建筑铁路16972公里，其中13353公里归当时中国政府所有。“在这13353公里铁路的投资中，45.5%是外国贷款的，由六七个国家在近四十年内提供的贷款总额累计达5400万英镑。”⁴⁶此外，外国投资在租界内和通商口岸城市对自来水、煤气、电力等公共基础设施的建设与供给，也带动了中国政府对其他“自开商埠”及非通商口岸以至所有城市公共基础设施的建设与供给，促进了城市本身以及近代城市化的发展。

至于外国投资基础设施建设的溢出效应更加明显。由于外国投资的地域相对集中，使得其基础设施产品的供给也相对集中于以上海为代表的几个大城市。这些大城市基础设施因不断得到外国投资而日益改善，也为中资企业的发展提供了便利而可以乘机“搭便车”，对中资企业外部成本的降低至关重要。由于这个原因，所以可以看到，近代兴起的中资企业绝大多数都集中在以上海为主的几个大城市，而这正是外国投资基础设施的溢出效应有利于中资企业发展的表证。当然，还应该看到中资企业上海等地相对集中的另一个重要原因，即安全感的影响。由于“中国的企业家们感到条约口岸是一个能够提供商业安全和有效生活方式的世外桃源，在这些避风港里，他们通过对制造业的投资不仅能保持自己的财富，而且享有更大的自由去支配这些财富”，⁴⁷所以导致了中资企业在地理空间上的相对集中。

关于外国投资的效应问题，吉尔伯特·罗兹曼的这段论述也是很有启发的。“成千上万的中国地方企业所使用的工具依然主要是依靠人力而不是资本。他们通过经纪人和商人组成的网络交换产品，有一整套的城镇和乡村各级市场，乡村市场是定期逢集的。而手工业生产始终与农田耕作的忙闲相关。中国企业家不得不去等待良好的条件来组织以工厂为基础的生产，这些条件包括：适宜的技术、与其他行业（房地产、金融、商业）一样有吸引力的利润、需有充足的低利率贷款以购买机器设备。以上条件只有条约口岸才具备，西方人在那里也有开办银行或其他企业的自由。有这些自由的口岸市场和大规模的外国公司的存在，中国企业才会有刺激和榜样组织起来，否则什么也谈不上。”⁴⁸

虽然外国直接投资具有非常明显的示范、溢出与积聚等效应，但众所周知，外国直接投资所产生的这些效应能否发挥出来、对东道国的民族产业和

⁴⁶ 郑友揆(1984)第53页。

⁴⁷ 吉尔伯特·罗兹曼(1988)第420页。

⁴⁸ 吉尔伯特·罗兹曼(1988)第422页。

企业能否产生带动作用,主要取决于东道国民族企业的投资能力、投资冲动。一位外国学者在比较中、日、俄三国近代经济发展历史时认为,三国都有吸引、利用和控制外资的问题,日、俄在“利用那些可能成为剥削他们的人这方面较为成功”,而中国由于“较弱的政治结构”,造成很大的损失。⁴⁹所谓中国“较弱的政治结构”,当然是指近代中国的半殖民地化,但也包括担负领导近代中国经济发展重任的软弱的中国政府。由于软弱的中国政府没有承担起领导经济发展的重任,不仅导致了外资利用的无序混乱,而且还导致稀缺资源的长期非优化配置,从而造成近代中国产业市场的结构严重失衡。而日本、俄国由于有一个强大的政府,其通过产业政策、投资政策等经济制度设计安排,不仅充分地利用了外国投资来促进本国经济的快速发展,而且还迅速地实现了工业化,摆脱了也一度如中国似的任人欺凌的境地。近代,由于中国主权的不完整,由于缺少一个强大的中央政府,尽管外国在华的投资数量不少,但中国并未借此发展起来,像日本、俄国那样较为成功地利用外国投资的好运始终没有在近代中国降临。

尽管外国投资未能阻止近代中国本土企业的成长,尽管外国投资具有多方面的正效应,但是,在分析近代外国在华投资效应时,则不能不注意外资的进入是在不平等条约的庇护下发生的,更不能不注意外资的进入及其对东道国资源的肆意采伐是一种赤裸裸的强权侵略。近代外国在华投资其本意是为了实现其在中国赢利的长期化,并在每次列强侵华战争中扮演着极不光彩的角色,推波助澜,助纣为虐。然而,即使面临着极其不利的政治、经济环境,中资企业在外国投资的刺激下,依然似滚滚长江东逝水般而不可遏制地成长并壮大起来了,这是外国投资者所没有想到的,因此,从这个角度而言,近代外国在华投资在客观上又推动了中国经济的进步。

参 考 文 献

- [1] 安格斯·麦迪森,《中国经济的长远未来》,楚序平、吴湘松译。北京:新华出版社,1999年。
- [2] 陈谦平等,“近代英国在华直接投资评析”,载吴景平等主编《近代中国的经济与社会》,第239—256页。上海:上海古籍出版社,2002年。
- [3] Elvin, M., “中日俄近代早期经济增长的比较”,《经济学术资料》,1982年第5期,第47—51页。
- [4] 高家龙,《中国的大企业——烟草工业中的中外竞争》。北京:商务印书馆,2001年。
- [5] 宫崎犀一等,《近代国际经济要览·16世纪以来》,陈小洪等译。北京:中国财政经济出版社,1990年。
- [6] Hou, C., *Foreign Investment and Economic Development in China, 1840—1937*. Cambridge, Mass.: Harvard University Press, 1965.

⁴⁹ Elvin, Mark(1982)。

- [7] 侯继明,“外国投资和经济现代化”,载张仲礼主编《中国近代经济史论著选译》,第50—78页。上海:上海社会科学院出版社,1987年。
- [8] 吉尔伯特·罗兹曼,《中国的现代化》,国家社会科学基金“比较现代化”课题组译。南京:江苏人民出版社,1988年。
- [9] 迈克尔·P.托达罗,《经济发展》,黄卫平等译。北京:中国经济出版社,1999年。
- [10] Perkins, D., *China's Modern Economy in Historical Perspective*. Stanford: Stanford University Press, 1975.
- [11] 上海社会科学院经济研究所,《英美烟公司在华企业资料汇编》第四册。北京:中华书局,1983年。
- [12] 孙毓棠,《抗戈集》。北京:中华书局,1981年。
- [13] Svedberg, P., “The Portfolio-Direct Composition of Private Foreign Investment in 1914 Revisited”, *Economic Journal*, 1978, 88(352), 763—777.
- [14] 王利华,“近代对外对华投资的影响因素剖析”,《南开经济研究》,1997年第2期,第66—72页。
- [15] 吴承明,《中国资本主义与国内市场》。北京:中国社会科学出版社,1985年。
- [16] 许涤新、吴承明,《中国资本主义发展史》第2卷。北京:人民出版社,1990年。
- [17] 徐义生,《中国近代外债统计资料(1853—1927)》。北京:中华书局,1962年。
- [18] 严中平,《中国棉纺织史稿》。北京:科学出版社,1955年。
- [19] 伊特韦尔等,《新帕尔格雷夫经济学大辞典》第二卷,陈岱孙主编译。北京:经济科学出版社,1996年。
- [20] 张国辉,“中国棉纺织业1895—1927年的发展和不发展”,载《中国社会科学院经济研究所集刊》第10集,第188—223页。北京:中国社会科学出版社,1988年。
- [21] 张继凤,“略论百年来我国利率的变迁”,《金融研究》,1982年第4期,第48—51页。
- [22] 赵德馨,《中国近代国民经济史教程》。北京:高等教育出版社,1988年。
- [23] 郑友揆,《中国的对外贸易和工业发展(1840—1948)》。上海:上海社会科学院出版社,1984年。

Foreign Investment in Modern China : the Size and the Effect

DECAI YANG
(Nanjing University)

Abstract The received wisdom is that foreign investment in modern China was an important factor that blocked Chinese economic development in the early twenty century. This paper analyzes the size and profit margin of foreign investment in modern China. It also analyzes the development of domestic enterprises in the years when foreign investment increased steadily. We find that the positive effects of foreign direct investment to domestic enterprises are highly significant. Foreign investment was not a factor that blocked China's economic development but a factor contributing to

the growth of domestic enterprises. However ,the effect of foreign indirect investment needs more analysis.

JEL Classification E22 , N01 , O16

经济学的性质与哲学视角审视下的经济学 ——一个基于经济思想史的理论回顾与展望

韦 森*

摘 要 基于对经济思想史上一些相关文献的理论回顾,本文认为,在
中国经济与中国经济学理论互动演变的当下格局中,是对经济学的性质、意义、
任务以及它到底是一门什么样的学问这样一些元经济学问题进行理论反思和讨
论的时候了。第二小节的文献追溯发现,在经济学说史上,伴随着由“政治经
济学”向“经济学”的转变,经济学家们对经济学的本质和任务的理解也发生
了很大变化。第三小节讨论了经济学能否成为一门价值中立的科学问题。研究
发现,任何经济学派的理论观点都会潜含着某些价值判断和伦理原则在其中,
经济学家作为一个“学术人”所面临的问题,首先是如何透过自己的文化信念
和价值观,去发现人类社会中的某些“自然秩序”。第四小节讨论了经济学能
否成为像物理学、化学等自然科学那样一种高度形式化的“精密科学”这一问
题,并对目前主流经济学中流行的唯科学主义的研究偏向做了一些反思性的评
论。本文最后所得出的结论是:为了中国的理论经济学在未来的健康成长,应
该鼓励不同经济学流派和研究路径之间的竞争、对话和多元化发展。

关键词 政治经济学, 经济学, 经济学方法论, 实证主义, 唯科学主义

“理性永远存在,但它并不永远存在于理性的形式之中。”

——卡尔·马克思

一、引言:从豪斯曼的《经济学的哲学》谈起

自1978年改革开放以来,中国经济已经保持了近30年的高速增长。同
时,自同年召开的十一届三中全会以来,中国学术思想界也经历了一个近30
年的思想解放过程。伴随着经济高速增长、社会进步和思想解放过程,中国
的经济学理论也有了长足的进步。不但从整体上来说,中国的经济理论研究
已经大致完成了一个从计划经济时代以苏联政治经济学教科书为基本框架和

* 韦森,复旦大学经济学院。通讯地址:上海邯郸路220号,复旦大学经济学院200433;电话:(021) 65643525。E-mail:liweisen@fudan.edu.cn。本论文是笔者所承担的2005年教育部一般项目“现代经济学的语言与修辞:一个马克思主义经济学与西方经济学的综合比较视角”的部分研究成果。笔者感谢黄有光、孙广振、陈志武、文贯中、洪永森、姚洋、史晋川、叶航、周业安、李文溥、董志强、宋铮、杭行、陈钊等诸位的评论和批评意见。我的学生皮建才、梁捷、方钦、席天扬、杨荷、周子贤、丁建峰以及李华芳等也提出了诸多修改意见,尤其是指出了我的打字错误。他们的一些意见已经采纳进本文的最后定稿之中。当然,本文的所有的观点和依然存在的纰漏由作者全部负责。

范本的单一经济学话语体系向一种当代经济学和古典经济学研究共存且诸多门类经济学科共同繁荣发展的时期转型；在各大学经济学院和商学院的实际教学过程中，也实际发生并正在试行着一种多元化课程设置的动态演变格局。中国经济学理论和中国高校经济学教育的双重进步和同步发展，不但使一批又一批的中国高校经济学专业的博士和硕士等专业理论人才进入教学或科研单位并逐渐形成了中国经济学中新生代的主流研究力量，而且也有成千上万的毕业生进入了一些国外著名大学攻读研究生课程，并有一大批从国外获经济学博士学位的中青年经济学家回国执教，或在各科研机构 and 实际经济部门从事经济理论研究。更为可喜的，这一新生代的中青年经济学人，开始以现代经济学的研究方法和语言与国际经济学界的同仁们进行交流，并开始在国际上的外文经济学学术刊物上发表文章。

改革开放近 30 年来，中国的经济总量增长是巨大的，同时中国经济学理论在整体上所取得的进步也是巨大的。然而，这双重的巨大成就，并不能掩盖在中国社会现实中以及在中国经济学理论发展过程中同时存在的一些问题。正如经历了 30 年高速经济增长的中国社会正在走向何方这样一个巨大的理论和现实问题有待学术界共同探讨和思考一样，中国理论经济学也同样面临着未来道路如何走这样一个重大问题。另外，这里不得不说，正如改革开放以来近 30 年的中国高速经济增长巨大成就可能部分遮蔽了中国理论界观察社会问题的视界一样，当下中国的经济繁荣，也似乎致使忙忙碌碌的中国经济学人没时间来反思经济学到底是一门什么样的学问这样一个“形而上学”问题了。然而，现在的问题可能是这样：只有大致理解经济学到底是怎样的一门学问了，中国经济学界方能较合宜地理解经济学人自身在经济社会发展中的确当位置和角色，也才能进一步较清楚地把握我们这个社会的未来走向。一句话，在当前中国经济繁荣、社会进步与经济理论互动发展的格局中，从未来制度变迁和经济理论发展的现实需要来看，是在整体上反思经济学的本质和任务到底是什么的时候了。

2007 年 1 月，北京世纪文景出版公司翻译出版了美国经济学家豪斯曼编的《经济学的哲学》（丁建峰译）文集。这部文集的编者丹尼尔·豪斯曼（Daniel M. Hausman）现为美国威斯康星大学麦迪逊分校的哲学教授，并且已经担任在国际甚有影响的《经济学与哲学》杂志的主编十多年了。由这位行内专家所精心收编的文集，荟萃了近现代经济思想史上一些理论大师的经典之作，覆盖了经济学诸领域中的一些“元经济学”（meta-economics）的深层问题。在豪斯曼《经济学的哲学》中译本的封底，中国经济学家汪丁丁教授曾表达了一个非常重要的理论判断：“为了让经济学走向 21 世纪，我们应当知道它在 20 世纪的整体状况。对经济学的整体把握，就是经济学的哲学。”笔者非常同意这一见解。这篇文献综述性文章，自不敢言是对经济学的发展作任何整体性的把握，而仅仅是基于笔者一点非常有限的文献阅读，就理论经

济学中几个相互关联着的重要问题，做些文献梳理，并附带增加自己的一点“边际评论”。这里，笔者仅挑出三个主要问题来，就教于经济学界的先辈、同侪和青年同仁。这三个问题分别是：经济学的本质（nature）与任务是什么？经济学能否成为一门价值中立的学科？经济学能否成为像物理学和化学那样一种高度形式化和“技术化”的“精密科学”？

二、经济学的本质与任务

可能自“政治经济学”（the political economy）和“经济学”（economics）这些术语出现以及理论经济学学科形成以来，¹经济学是什么似乎就成了经济学家们的一个挥之不去的问题。较早，经济学家们多是从经济学的定义去把握经济学的本质和任务的。譬如，在《国富论》中，斯密（Smith, 1776/1930, vol. II, p. 177）曾把政治经济学视作为“研究国民财富性质和原因的一门学问”。斯密（Smith, 1776/1930, vol. I, p. 395）还认为，“作为政治家和立法者的一门科学的政治经济学，有两个不同目标：首先，为人民提供充足的收入或生计，或者更确切地说，使人民能给自己提供这样的收入和生计；其次，为国家和社会提供公共服务所需的充分收入。概言之，其目的在于富国裕民”。正因为斯密本人在写《国富论》时心目中把经济学视作为写给政治家和立法者读的一门学问，在他的话语解释（discourse）中，到处不无建议君主与政府应当在治理社会时如何行事的治国方略和政策建议。换句话说，斯密恰恰是在古汉语“经邦济世”和“经世济民”的“经济”意义上来看待经济学的性质与任务的。用施蒂格勒（1986, ch. 18）的话来说，在斯密等英国古典经济学家看来，经济学的任务是劝说（preach）人类社会尤其是国家的主权者（the sovereign——一译“君主”）应该采取什么样的合宜治理方式和制度形式，来使一国的经济繁荣，人民致富。有了这一明确的学术目标，在《国富论》中，斯密不仅讨论了一些经济和贸易政策，而且还较广泛讨论了政治、法律甚至外交等方面的问题。譬如，除了提出“看不见的手”这一著名的经济学原理外，斯密在《国富论》中还讨论了保证这种看不见的手能

¹ 从经济学说史上来考证，最早是法国重商主义的代表人物孟克列钦（Antoyne de Montchrétien, 1575—1621）最早使用“政治经济学”一词，他1615年和1616年分别发表了“献给国王和王后的政治经济学”和“论政治经济学”两篇文章，最早把“政治”与“经济”联合起来使用，比斯图亚特早了150年。另据澳大利亚悉尼大学著名经济学说史学家格罗内维根（Grenewegen, 1987）考证，在英语世界中，最早使用“政治经济学”作为其书名的，是英国经济学家斯图亚特勋爵（Sir James Steuart）1767年，他出版了自己的经济学著作《政治经济学原理研究》。但1875年，西方一位经济学家麦克劳德（H. D. MacLeod, 1875）在“政治经济学是什么？”一文中，主张将“政治经济学”改为“经济学”，并将其定义为“探讨支配可交换物品数量的科学”。在20世纪上半叶，西方一位著名经济学家坎南（Cannan, 1929, p. 44）认为，尽管麦克劳德提出了这一主张，但是，直到马歇尔（Alfred Marshall）之后——尤其是在马歇尔的《经济学原理》在19和20世纪之交对西方经济学界的巨大影响下，人们才逐渐接受了“经济学”这个概念。到了20世纪20年代，西方经济学界就比较通用“经济学”而不是“政治经济学”了。

得以良好运作的自由制度原则及其法律保障问题。斯密 (Smith, 1776/1930, p. 184) 指出: “一些特惠或限制制度, 一经完全废除, 最明白和最单纯的自然和自由制度就会自动建立起来。每一个人, 在他不完全违反正义的法律时, 都应给予完全的自由, 让其依照自己的方式去追求自己的利益, 并以其产业和资本与其他任何人以及其他阶层进行竞争。”依照这一理念, 斯密还对社会的收入分配提出了一个非常重要但却在很大程度上被人忽视了的观点: 各阶层人民的收入分配, 也应该是由市场竞争来决定。斯密 (Smith, 1776/1930, vol. II, p. 152) 说: “君主应该给予各阶层子民以公正和公平的待遇, 如仅仅为了促进某一阶层的利益, 而损害另外一个阶层的利益, 显然与此相违。”斯密对市场经济进步意义的确信, 甚至到了这样一种程度: 只要给定充分的自由竞争和市场秩序的自然成长空间, 市场诚信体制也会随着交易的扩大而慢慢产生出来。譬如, 在 1762 年到 1763 年撰写的《法理学讲义》中, 斯密 (Smith, 1978, p. 538) 就曾明确指出: “只要商业在某一国家兴起, 就总会带来人们正直和守时 (probity and punctuality) 的习惯。这些美德在未开化的国家里几乎是不存在的”。斯密还接着举例到, 在当时欧洲各国当中, 荷兰人最注重做生意, 同时也是最遵守诺言的人。²

对斯密本人对自发市场秩序运行之合理性的这种强烈信念, 就连比较推崇自由市场体制的芝加哥学派中一位重要思想家施蒂格勒 (George Stigler) 也似乎感到难能完全接受, 并好像略有微词。在 1980 年在哈佛大学所做的题为“经济学抑或伦理学”的著名坦纳人类价值讲座 (the Tanner Lectures on Human Values) 的讲演稿中, 施蒂格勒曾把斯密比喻为那种自由市场体制合理性理念的“传道者” (preacher), 并认为在《国富论》中到处都可以找到这种“传道”、“说教”或言“劝说”的例子。照施蒂格勒看来, 在斯密之后, 这种基于经济学家个人价值和伦理判断的说教就比较少了, 他还举例到, 在李嘉图 (David Ricardo) 的《政治经济学与赋税原理》以及在穆勒的《政治经济学原理》中, 这种“说教”就不多了, 而到了马歇尔的《经济学原理》中, 就根本没有“说教”了 (参见 Stigler, 1986, 中译本, 392 页)。施蒂格勒还认为, 从那之后, 经济学家们就变得“圆滑老练”起来, 用今天的话说, 人们不再像斯密那样关心人类社会到底是如何运作的这类“经邦济世”的大学问了, 而是把话尽量说得婉转巧妙, 从而只是做些解释现实经济现象的“实际功夫”。如果说施蒂格勒的这一观察无误的话, 我们今天似可以进一步认为, 在经济思想史上, 斯密对经济学性质和任务的解说和理解, 在后来的

² 现在看来, 斯密在 18 世纪的这一理论猜测, 似乎有普遍意义, 因为, 斯密的这一判断, 似乎也被当代中国市场化进程的历史实践所证实了。改革开放以来, 中国东南沿海地区尤其是浙江省的市场化程度比较高, 而这些省份和地区的市场诚信程度和诚信体制建设也比市场化程度相对落后的中国西部和内地省份要高一些。这应该是一个可以通过实证检验来证明的事实。

经济学家中在很大程度上被“修改”(revised——这里可不是在“修正”意义上说的,而且也很难说是被“修歪”)了。尤其是到了另一位英国古典经济学家、功利主义伦理学家和法学家边沁(Jeremy Bentham)那里,经济学的性质和任务被明显“功利化”了。照边沁(Bentham, pp. 1793—1795, p. 223)看来,“政治经济学既可被当作一门科学,也可被当作一项技艺,但当科学被运用于指导技艺时,两者可视为同一回事”³。

在认识经济学的任务和性质时,我们必须省悟到,斯密之所以像一个为市场经济运作之合理性大声呼喊的传道人那样写出《国富论》、《道德情操论》以及《法理学讲义》等经典著作,完全与他所处的时代密切相关。在18世纪后半期,英国的市场经济体制才刚刚萌生,重商主义的国家干预政策还在大行其道,保护市场运作的法治和宪政民主体制还没完全建立起来,更重要的是,当时的英国还处在工业革命的黎明前夜中。在此情况下,斯密把自己的经济学的任务确定为一种为政治家和立法者所写的一门学问,并决意要把自由市场体制的理念传道给(preach)世人,现在看来是完全可以理解的。然而,在《国富论》出版后的一个多世纪中,法治化的市场经济体制在英国、荷兰、法国、德国等欧洲国家相继生成了。正如施蒂格勒(参见 Stigler, 1986, 中译本, 418 页)所注意到的那样,“自《国富论》出版以后的一个多世纪里,经济发展的步伐一直以一种前所未有的速率和规模向前迈进。西方世界的生产技术、经济体制、生活方式,乃至政治制度,都发生了持久而意义深远的变化”。19世纪西方世界的兴起,西方国家中法治化的市场经济运作体制的生成,标志着西方社会从整体上已经完成了其现代化过程,因而,像斯密那样为呼唤法治化市场经济体制的生成和降临的“市场经济理论传道人”的历史使命已随之结束了,接着,经济学的现实任务和历史使命也随着社会的进步和历史变迁而发生了根本性的变化。这样,到边沁和约翰·穆勒(John Stuart Mill)所处的19世纪中后期,西方的社会历史已经不再需要经济学家们像斯密那样为呼唤出来“市场经济”而进行呐喊了,而是较多地要求经济学家们去注重探究市场运行中人们的行为方式并较功利地研究和教导人们如何在市场经济体制中作出自己的最佳选择,这就是可以理解的。⁴由此我们今天也可以判断,如果我们把经济学认作为从财富创造和社会福利增进的视角来探析人类社会是如何运作的这样一种话语体系的话,其任务和意义亦会随着人类社会的历史演进进程而不断发生变化,但是,经济学本身所展

³ 边沁的这一定义当时就遭到另一位经济学家西尼尔(Nassau W. Senior)的批评,说边沁这种把科学与技术混为一谈的做法将损害经济学的发展(参见 Senior, 1836/1938, p. 3)。

⁴ 这里必须指出,不能因为西方市场经济体制在斯密的《国富论》出版一百多年后基本形成了,就认为斯密这部伟大著作中关于市场经济运行的基本原理就过时了和不重要了。即使在当代社会,如果任何经济学人在读斯密的《国富论》时还不能为他在两百多年前的一些理论洞见所震撼和感动,那他本人就该自问一下自己是否真得算是一个“经济学人”了。

示的市场运行的一些基本原理,应该是经济学本身所永远秉有的,否则,经济学也就失去了它本身是有关人类经济社会运作基本原理的一门学问这一特殊性质了。

在 19 世纪后半期,处在自亚当·斯密所沿袭下来的英国古典经济学的学术传统中,以及在英国市场经济业已形成的历史背景中,约翰·穆勒(John S. Mill)对经济学的任务以及现实的和历史的意义的理解,较斯密来说就有了一些“修改”,并且,穆勒对经济学本质的理解,较斯密而言也发生了一些变化。譬如,在《政治经济学的定义及其方法》这一经济学说史上的经典名篇中,穆勒(参见 Hausman, 1994, 中译本, 53—55 页)曾对政治经济学下了这样一个定义:“政治经济学是探讨某些社会现象规律的科学,这些规律产生于人们生产财富的联合行动之中,就其所涉及的现象而言,不会由于人们对其他目标的追求而被修正。”接着,穆勒还指出,在确定政治经济学的上述定义时,他试图将其本质刻画为一门抽象科学,并认为其方法是先验方法,即建立在人们嗜好财利、为追求财富而精心选择手段的假设基础之上(这一点分别为后来的新古典主流经济学和奥地利学派的一位重要经济学家米塞斯,即 Ludwig von Mises, 所承传下来,下面我们还会谈到这一点)。从穆勒的这一经典名篇的这些论述中,我们发现,穆勒对经济学家以及经济学的理论任务的理解较斯密已经发生某些变化。与斯密那样把经济学理解为为政治家和立法者所写的一门经邦济世、经世济民的大学问⁵不同,穆勒较倾向于把理论经济学视作为研究人是如何追求财富且建立在某些有关人的行为模式的一些先验假定基础之上的“精密科学”。今天我们也从中可以发现,穆勒的一些论断,实际上与马歇尔之后西方新古典主流学派的建立在理性经济人基础上的抽象推理精神,在许多方面是相契的,或者说穆勒的经济学思想,已经蕴含了当代主流经济学的理论建构理路和分析精神的萌芽。

在西方经济思想史上,马歇尔(Alfred Marshall)以及其经济理论,可谓是一个承前启后的转折点。不但自马歇尔之后,“政治经济学”这一理论经济学的通称逐渐为“经济学”一词所代替,而且人们对经济学性质和任务的理解,也发生了很大变化。在《经济学原理》中,马歇尔(Marshall, 1920, p. 1)曾对经济学的性质做了一个折中的把握:“政治经济学或经济学是一门研究人类一般生活事务的学问”,“一方面它是研究财富的学科,另一方面,也是更重要的方面,它是研究人的学科的一部分”,或者概言之,“它研究在个人与社会活动中与获取和使用物质福利必需品最密切相关的那一部分”。对马歇尔

⁵ 斯密的这一理论精神被后来的马克思所强烈地继承下来了。但是,二人的区别在于,在马克思那里,政治经济学已经不再是为政治家和立法者所写的“经济学”学问了,而是向世人“劝说”一个理想的美好社会应该是怎样的“理论论证”和“构想”。由于讨论马克思主义政治经济学的本质与任务已经属于另外的专门研究话题,并在实质上已属于另外一种经济学的话语体系,故本文就不再展开讨论了。

的这些论述，我们今天似可以做这样的解读：经济学作为研究人类一般生活事务的一门学科，既要研究财富的创造和增殖问题，也要研究人的经济行为，但目的却只有一个，即探究和论述人类社会如何才能最佳地创造和使用财富。

马歇尔之后，尤其是在第二次世界大战之后，新古典经济学繁盛发展起来，并逐渐成为了当代经济学的主流。当代主流经济学家对经济学本质和任务的理解，较早被英国伦敦经济学院的一位老牌经济学家罗宾斯（Lionel Robbins）较全面和系统地予以阐述出来。在1932年出版的《经济科学的本质和意义》的这本小册子（1935年出版第二版）中，罗宾斯对经济学的本质、任务、意义以及研究方法等进行了专门的探讨。按照罗宾斯（Robbins, 1948，参见中译本，16—20页）的理解，经济学的本质是研究资源稀缺条件下人类在配置资源方面是如何行为的一门形式化的社会科学。罗宾斯认为，在错综复杂的现实世界中，当人们进行多种目标的选择时，经济学可以帮助人们理解自己选择的意义，以及如何做选择，故“它提供了人们合理行动的一种技术”，因为，“我们要想做出完全合理的选择，就必须先知道自己偏好些什么”（同上，123页）。

罗宾斯对经济学本质和意义的这种理解和把握，为后来的绝大多数当代经济学家所赞同，并在此精神下不断进行理论创新和学术建构，从而逐渐形成了当代经济学的宏大理论世界。之后，一些主流经济学家对经济学的本质和任务也做过一些具体的描述，但与罗宾斯的理解已经没有什么实质性的差别。譬如，在“经济学抑或伦理学”中，施蒂格勒（Stigler, 1986，参见中译本，393页）曾明确指出：“经济学的主要任务一直是以大众可以接受的方式来解释实际经济现象，而且近200年来我们能始终诚心诚意地尽力完成这一使命，尽管并不总能取得极大的成功。”施蒂格勒的这一见解，实际上不过是复述了弗里德曼（Milton Friedman）在20世纪50年代就提出的一些基本见解。譬如，在其著名的《实证主义方法论》一文中，弗里德曼（Friedman, 1953）曾强调指出，实证经济学在原则上是独立于任何价值判断的，“作为一种实证科学，经济学是一种被尝试接受的、关于经济现象的概括体系，用以对条件变化的结果作出预测”（参见Hausman, 1994，中译本，174页）。应该说，经济学作为解释社会现实中经济现象的一门实证科学，今天已经不再是弗里德曼、施蒂格勒和芝加哥学派的几个经济学家的主张和信念了，而是变成了当代主流经济学家或言世界上绝大多数经济学家所接受和信奉的基本立场。然而，从上述经济思想史的简短梳理中，我们已经大致看出，这种实证经济学或言“实证主义精神下的理论经济学”，显然已经不是经济学鼻祖亚当·斯密眼中原来的样子了，或者说在某种程度上已经偏离了经济学的亚当·斯密传统。就此而论，哈耶克（F. A. von Hayek）认为，只有从门格尔

(Carl Menger) 到他本人一脉承传下来的奥地利经济学分析道路⁶, 才代表了经济学理论的亚当·斯密传统。现在看来, 哈耶克的这一见解不无道理, 或者至少说是完全可以理解的。

概言之, 自亚当·斯密的《国富论》出版以来, 经济学家们对经济学的本质、任务和意义的理解和把握, 已经发生了很大变化。在此, 笔者不想再对经济学的性质和任务做任何新的定义性把握, 而只想在这里提出这样一个现实问题: 在面临着 21 世纪初叶中国经济继续进行着的一种非常独特的市场深化过程的历史格局和大势中, 中国当下和未来的理论经济学的历史使命是什么? 中国的经济学家们又将能够做些什么? 是一股脑地全部挤入当代新古典主流经济学的话语体系中对当下社会现实中的一些现象(包括中国社会转型期的一些过渡性的“特殊现象”)做些形式化的实证研究? 还是部分回归到经济学的亚当·斯密传统, 走出一条中国经济学理论发展的自身道路? 或至少走一段自己的独立发展路程?

三、经济学是否能成为一门价值中立的学科?

从上一节的分析中, 我们已经知道, 作为研究和解释人类社会经济活动现象的一门知识体系, 经济学的任务会随着不同历史阶段上的人类社会发展的不同格局和不同现实要求而发生改变, 并在世界历史上实际上经历了从斯密的那“经邦济世”的大学问, 到马歇尔的经济学的双重任务, 再到弗里德曼的对广义的市场现象进行实证研究这样一种经济分析史历史的转变过程。然而, 尽管经济学的任务随人类社会的历史变迁而变化了, 但经济学的本质和意义(significance)却并没有发生多少实质性的变化。我这里是说, 尽管随着西方社会中法治化市场经济体制的成长和成熟, 经济学家们不大再需要

⁶ 在 1937 年发表的一篇关于“经济学与知识”的论文中, 哈耶克(见 Hayek, 1949)似乎倾向于把经济学理解为关于“选择的纯粹逻辑”(pure logic of choice)。这是哈耶克早年的思想。到了晚年, 哈耶克则认为, 由于“经济”(economy)一词的含义含混不清, 他主张用一个希腊词的 Catallaxy(邓正来主张把它翻译为“偶合秩序”)来替代它。相应地, 哈耶克则根据一位在 19 世纪初一位英国经济学家和枢机主教华特理(Archbishop Whately, 1831)在他的《政治经济学导论》一书中所提出的一个概念“catallactics”(含义为“研究市场交易现象的科学”), 主张以此概念来取代当代充满实证主义和经验主义精神的且高度形式化的“economics”。在哈耶克的后期著作中, 他对研究自发市场交易秩序生成和运作机理的这种“交易的科学”做过较详细的分析和阐述。米塞斯本人也赞成采用华特理的这一“catallactics”概念, 并且在他的《人类行动》(Mises, 1949)和《经济学的认识论问题》(Mises, 1960)等著作中分别探讨了 this 概念含义。但是, 与哈耶克的晚年思想的区别在于, 对经验主义和实证主义一样深恶痛绝的米塞斯则明确地把经济学和这种研究市场现象的交易的科学理解为有关人类经济行动的一门先验科学, 即他所要建构的“praxeology”(人类行动的科学)的一部分。照米塞斯(Mises, 1960, EPE, I. 2. 1)看来, “the basic principles of economics are not empirical but a priori”(经济学的基本原理不是经验的, 而是先验的)。在《经济学的认识论问题》一书中, 米塞斯(Mises, 1960, 参见中译本, 12 页)还更明确地说: “努力得出普遍正确知识的人类行动的科学是一个理论体系, 它迄今为止最为精心地构建的一个分支是经济学。在其所有的分支中, 这门科学是先验的, 而不是经验的。正如逻辑学和数学一样, 它不是得自于经验, 它先于经验。它现在就像过去一样, 是行动和事实的逻辑。”

为通过市场交易过程而配置资源的合理性 (raison d'être) 和优点进行“说教” (preaching) 了, 而主要致力于研究通过市场运行的“机理”是怎样的这样的“现实的”问题, 但经济学的目的和意义似乎只有一个: 那就是向人们展示, 通过形成什么样的社会秩序和制度安排, 人类社会才能最有效地创造和最佳地使用财富, 从而增加个人、群体、国家乃至整个人类社会的福祉。

经济学的本质和意义是如此, 但是在完成这一理论任务时, 由于不同经济学家们观察社会问题的视角不同, 特别是由于经济学家们使用不同的经济学术语和解释话语体系, 就会得出完全不同的结论, 并且会导致完全不同的政治诉求和经济主张, 于是乎, 好像经济学家们就如熊彼特 (Joseph Schumpeter) 所见的那样均落入了不同的意识形态⁷之中。这里不妨举一个简单的例子。如果使用源自英国古典经济学的劳动价值论的有关术语和话语体系, 就很自然会得出在任何市场经济条件下正常的商品生产中均存在着剩余价值的创造与占取问题, 因而“剥削”问题是一个无论如何也绕不过去的理论死结。相反, 如果采取“边际效用价值论”概念及其相关联着的分析话语体系, 人们则会达致现代经济学中的欧拉定理 (the Euler Theorem): 完全竞争下的所有生产要素投入的收益均等于其边际产出, 因而如若不存在垄断, 就不存在剥削问题。正因为这一点, 许多人认为, 劳动价值论是维护劳动阶级利益的, 边际效用价值论则是为资产所有者阶层的剥削做辩护的 (参见 Schumpeter, 1949, p. 352)。单从这一简单的例子中, 我们就可以发现, 在经济学的话语解释中, 使用不同概念、术语和不同“话语系统”, 就会产生对现实世界的不同看法, 从而发生了经济学中不同“意识形态”的分野。

于是, 这里就出现了这样一个根本性的问题: 作为现实经济过程之“镜像反映”的经济学中的概念、原理、定理以及古典经济学家所喜欢使用的“经济规律” (economic law, 或翻译为“经济法则”) 到底是指一种什么样的“自成一类的存在” (sui generis)⁸? 为什么观察同一个现实经济过程, 经济学家们有时会得出完全不同的结论? 对这些问题的回答, 显然已经远远超出了理论解释中经济学家们个人的价值判断以及任何一个经济学人在现实观察和进行经济理论思考时的知识结构中所潜含着的意识形态背景问题了, 而必然牵涉到经济学的本体论 (ontology) 和认识论 (英文为 “epistemology”, 或译“知识论”) 等一系列复杂问题, 并且说到底与经济学的语言这一最深层问题密切相关着。由于经济学的语言问题极其复杂, 我们只能留作为以后的研究

⁷ 由于意识形态 (英文为 “ideology”)。美国著名的华裔思想家林毓生教授主张把它译为“意带牢结” 这是个非常不错的建议) 问题太过复杂 加上有西方学者如丹尼尔·贝尔 (Daniel Bell) 等认为现代世界已经到了“ideology”终结的时代了 在以下的讨论中 我们毋宁用“文化信念” (cultural beliefs) 或“价值观”来代替之。

⁸ 在《货币、货币哲学与货币数量论》一文中, 韦森 (2004, 67 页) 曾认为: “人类生活世界中的经济现象界是经由人类社会活动和交往过程而形成的一种法国社会学家涂尔干 (Émile Durkheim) 和德国社会学家西美尔 (Georg Simmel) 所说的那种‘自成一类的存在’ (sui generis)”。

任务而这里暂且不再深究,但目前我们至少可以猜测到,在经济学说史上所形成的诸多的流派,原来只是一些大家共同使用某套经济学术语并以此进行推理和交流而形成的一些不同的“语言共同体”!因而,只有从语言学和语言哲学的视角历史地和理论地考察和审视一些基本经济术语与现实对象性的关系,并进而对经济学的语言进行哲学本体论的考察,才能期望对这一现象发生的原因能有一个较深层的理解(但不可能期望所有问题全得到解答)。⁹由于从经济学的语言和修辞的视角梳理经济学中一些基本概念的生成和演变过程太过复杂,已经远远超出了目前这篇文章所能深入讨论的可能,在本文下面的讨论中,我们这里只考察经济学是否是一门价值中立的科学这样一个相对简单的问题。

应该说,自 20 世纪初以来,经济学家们从古典的“political economy”话语体系逐渐转向了“economics”的话语体系,随之,经济学家们的主要理论诉求,就是要构建像“physics”(物理学)等自然科学那样一种关于人类经济社会运行内在法则的带有“ics”词缀形式的“economics”的“科学”(下一小节我们会专门讨论这一问题)。自从这一转变发生以来,大多数经济学家一方面尽量避开伦理判断,另一方面则公开声言要尽量做到“价值中立”。值得注意的是,不但像芝加哥学派的主要思想家弗里德曼那样公开表明,要把经济学构建成一门实证科学,其“最终目标是建立一套‘理论’或‘假说’,来对尚未观察到的事物作出有效且有意义……的推测”(见 Hausman, 1994, 中译本 151 页),而且像米塞斯这样的奥地利学派的旗手,也公开声言“经济学作为一门科学对价值判断是中性的”(参见 Mises, 1960, 中译本 36 页),并且,在他的《人类行动》一书的导言中,米塞斯还明确指出:“如果像道德监察官那样带着某种任意的标准和主观价值判断的赞同或不赞同的态度来研究社会事实,那将是徒劳无益的”(参见 Mises, 1949, p. 2)。应该说,当代大多数西方经济学家均赞同米塞斯的这一观点。譬如,为了说明经济学的理论分析应该是价值中立的,诺贝尔经济学奖得主之一索洛(Robert M. Solow)教授就曾具体举例到,像国民收入、失业率、价格水平、利率等等这些经济学术语以及其所涵指的现实经济对象性,与社会制度无关,它们“对资本主义和社会主义同样适用”(见 Hausman, 1994, 中译本 208—209 页)¹⁰。

⁹ 按照当代一位美国大哲学家奎因(Willard V. O. Quine)的语言分析本体论哲学,任何理论都有某种指谓的概念框架,没有任何理论可以独立地离开某种概念(语言)框架而得以讲说和辨认。依照奎因的这种语言分析哲学的理路,可以认为,对经济学说史上任何一个流派的学术观点的评价,以及对经济学本身“客观性”(韦伯语)和“科学性”的评估,可能最终都要追溯到经济学的语言分析中去,这是本文作者目前的一个基本判断。

¹⁰ 这里应该指出,尽管索洛提出一些经济术语可以超越经济学家个人的价值判断以及意识形态之争而适用于任何社会,但是他还是观察到,经济学的任何结论都很难逃脱经济学家们的意识形态背景。在对“经济学”的“科学性”和“超越性”进行上述辩护之后,索洛最后还是承认:“经济学可能确立的任何结论(即使经过了最激烈的批判审视之后)都必须建立在关于经济制度的哲学和意识形态讨论的一套基本规则之上”(见 Hausman, 1994, 中译本 220 页)。

另一方面，也有一些西方学者对经济学家们的上述立场和理论诉求大为怀疑。譬如，1974年诺贝尔经济学奖得主、瑞典经济学家缪尔达尔（Gunnar Myrdal, 1969, p. 74）在其《社会研究中的客观性》一书中就曾指出：“事实上，没有任何社会科学或社会研究的分支是‘道德中性的’（amoral）或‘政治中性的’（apolitical）。没有任何社会科学是‘中性的’或简单来说‘事实的’，实际上在这些词的传统意义上来说也不可能是‘客观的’。任何研究永远是并且在逻辑上必然是建立在道德和政治评价的基础之上的，研究者必须明确表明自己的价值观。”另外，作为《经济学与哲学》杂志主编的豪斯曼和麦克佛森（Michael S. McPherson）也认为，尽管经济学家们相信存在着与伦理学不相干的经济科学原则，尽管他们相信伦理学无助于他们的实证经济分析，但是“如果经济学家们拒绝关注伦理问题，那么他们将不会知道应该提出什么问题”。根据这一判断，豪斯曼和麦克佛森在《经济学、理性和伦理学》一文中还主张，“即使纯粹的实证经济学，经济学家们也必须考虑所研究的社会和他们自己的道德规范”（参见 Hausman, 1994, 中译本, 223—226页）。基于这一判断，豪斯曼和麦克佛森（同上, 232页）还用带着不无讽刺的口吻说：“经济学家们并不费心研究道德哲学，但却[常常]自我感觉拥有道德权威”。因为，照当代西方一些主流经济学家们看来，帕累托最优显然是最理想的。照此逻辑，任何帕累托增进，也显然就是一个道德上的改善。由于按照新古典主流经济学的推理，只有完全竞争才能达到帕累托最优，因而，这些实证经济学家们实际上假定完全竞争才是最符合道德的。由此看来，这些声称价值中立的经济学家，并不是没有个人价值判断（value-free），而是无时无刻不在做着功利主义的价值判断（即从增进社会效率一维来观察社会问题）。照此看来，这些实证经济学家们，逃脱了市场运行中的价值判断了吗？显然没有。另外，豪斯曼和麦克佛森（Hausman, 1994, 中译本, 231页）还曾举例道：如果社会存在一种状态R，在这一状态下，有成千上万的人沦为饿殍，但却仍无法通过损害某些人的利益（哪怕是极其微小的利益）来改善这千万人的利益，那R无疑仍然是一个帕累托最优状态。但这种状态是否符合道德？这一状况是否应该改变？主流经济学家又该如何回答这一问题？或者就根本拒绝回答而把这类问题留给政治哲学家和伦理学家去回答？

退一步讲，即使当代主流经济学家们从他们的职业分工角度坚持拒绝研究或不回答上述问题，他们也仍然绕不开经济学的价值判断和伦理立场问题。在《经济学与伦理学》的小册子中，韦森（2002, 84—95页）曾论证到，如果从“双头”（duopoly）合谋定价的思路来论证“独占”（monopoly）的定价原理，就会发现，如果独占者可以做到“完美差别定价”（perfect price discrimination），完全垄断（即市场独占）与完全竞争在社会总福利的最大化上应该没有任何差别。因为，独占者的“完美差别定价”，可以像完全竞争一样

达到无任何“无谓社会福利净损失”(dead-weight loss)的市场均衡状态(帕累托最优)。但与完全竞争的市场安排不同,在一独占企业能完全垄断市场并且能做到“完美差别定价”的市场安排中,所有由“消费者剩余”所表示的“福利增进”都变成了独占企业的“垄断利润”了。那么,既然同样可以达到市场均衡的帕累托效率,为什么大多数主流经济学家坚持认为垄断就是不好的,是不符合道德的;完全竞争就是好的,就是符合道德的?难道他们就没有先入之见的道德判断在其中?仔细思考这个问题,将会把新古典主义经济学家们这些“假道德中立者”(pseudo-moral-neutral)逼到一个非考虑市场安排的道德含蕴不可的墙角。从这一视角来考察问题,我们就会发现,以新古典主义为主流的所谓“科学主义”的经济学,并非没有伦理判断和道德基础。其道德基础恰恰在于从边沁、穆勒、西季威克(Henry Sidgwick)到埃奇沃思(Francis Y. Edgeworth)的功利主义伦理学,即认为,只有完全竞争的市场才是“最公正”因而才是“最道德的”。因为,新古典经济学家们相信,只有在完全竞争的市场中,才能达到“帕累托最优”,才会无“无谓社会福利净损失”问题。但是,如上所述,新古典主流经济学家们能逃脱完全垄断(市场独占)加完美差别定价能做到与完全竞争一样的市场均衡效率的道德评判么?

到这里,我们也许才能真正理解凯恩斯在 1938 年 7 月 4 日致哈罗德(Roy Harrod)信中的所说的一段话的意思:“我认为经济学在本质上是一门道德科学,而不是自然科学。这也就是说,它必须运用内省和价值判断”(参见 Hausman, 1994, 中译本, 253 页)。

这里应该指出,通过上述推理,我们逼出了一些自称是价值中性(value-neutral)的经济学家们自己的价值判断和伦理立场,并不是说功利主义伦理学没有道理,也不是说功利主义伦理学不无可取之处,恰恰相反,笔者认为,人类社会经济活动的理性基础在很大程度上是建立在功利主义伦理学所发现的一些基本原则之上的。经济学作为一门社会科学,显然离不开功利主义的经济和伦理的“计算”,但把功利主义作为它的唯一基础,显然也会带来许多问题。因为,功利主义内秉着许多它自己无法摆脱的理论困境,且人类社会毕竟也有超越功利主义考虑的美德和良善。譬如,从静态上考虑,给吸毒、贩毒和性产业等社会现象开绿灯,在一定时点上肯定是一种帕累托增进。经济学家们是否由此就主张对这些社会现象都应该发给“准许”的牌照?另外,人类也有超越功利计算的自由、权利、正义、公平、平等等其他基本伦理价值。如果按照功利主义伦理学的思路一竿子插到底,在人类社会中将会出现许多怪异现象,甚至从某种程度上会动摇人类文明社会的基础。更何况,仅从功利主义伦理学的一些价值判断出发,既可以导致自由市场竞争是最合乎道德的判断,也可能同样推出集权的中央计划经济的合理性——这关键取决于对人类的天性如何认识和把握。

经济学不能超越伦理学而做到价值中立，这并不意味着经济学要回到过去的意识形态之争中去。¹¹ 经济学的伦理基础和道德判断问题的复杂性，与中国市场经济体制成长过程中出现的诸多社会问题一起，正急切地呼唤着中国经济学界重视对福利经济学和社会选择理论问题的研究。中国的经济学、政治学、法学、伦理学、社会学以及整个哲学社会科学的进步，也要求中国的学人对这些学科交叉领域中构成人类社会运作之深层理性基础的一些基本价值进行评判和审视。这无疑是未来中国社会经济发展所提出的一项非常紧迫的时代理论要求。

四、经济学能否成为像物理学和化学那样一种高度形式化的“精密科学”？

在《经济学的哲学》这本文集的“导言”中，豪斯曼（Hausman，1994，1—2页）一开始就重新提出这样一个问题：“经济学是一门科学吗？”他还具体解释道，许多经济学家之所以常常被人提起这个问题，原因在于，“现在，不但一般大众怀疑经济学家，就连经济学家自己也在怀疑自己”。¹² 为什么人们常常问这样一个问题？究其原因，可能与事关经济学命运和任务的一个最深层问题密切相关：有着自由意志和各自行动原则的人的行动是否能使经济学达致像物理学和几何学那样一种高度形式化的“精密科学”？

从经济学说史来看，并不是在20世纪以后经济学家们才致力于把经济学作为一门“科学”来建构。其实，早在《国富论》中，亚当·斯密就在很多地方把经济学称作为一种“science”了。之后，从李嘉图到马克思，从边沁、穆勒到马歇尔，经济学家们都似乎致力于把自己的经济学理论作为一种科学来构建，以至于像米塞斯那样自认为传承了康德先验哲学传统的奥地利学派的思想家，也在自己的后半生中致力于把他的人类行动学变成一种高

¹¹ 在谈到亚当·斯密的工作时，熊彼特（Schumpeter，1949，p. 352）曾指出：“就亚当·斯密而言，令人感兴趣的实际上不是他没有意识形态的偏袒，而是这些偏袒无害”。熊彼特还具体解释道，当斯密写《国富论》时，他曾全部把同情给予了劳动者，而对土地所有者和资本家则采取了一个完全旁观者的态度，认为后者是必要的恶。由此熊彼特认为，斯密的意识形态偏袒，并没有损害他的经济学的“科学性”。当熊彼特说这话时，显然有他的价值判断在其中。经济学家不能超越意识形态，或如美国著名学家鲍尔丁（Kenneth Boulding，1970）所见那样，经济学家不可能超越他们所在的“文化”和“亚文化”中一些信念进行理论思维和价值判断——正如人难以超越语言进行思维一样，但这并不一定就完全是负面的。这里的问题的实质在于，经济学家作为一个“学术人”（homo academicus），其经济学的分析是否能透过他所处社会的文化信念并超越自己潜在的价值观（或言超越熊彼特所认为的那种永远挥之不去且永远伴随着我们的种种“意识形态”）去发现人类社会运行的“自然秩序”或言“天道”（providence）。

¹² 从学说史来看，早在19世纪末20世纪初，老牌的美国制度经济学家凡勃伦（Thorstein Veblen）就开始提出经济学是不是科学这一问题了，并于1898年就以“为什么经济学还不是一门演化的科学？”的话题作过文章（参 Eichner，1983，p. 205）。

度公理化的科学¹³。现在看来,只有思想深邃且见解独到的哈耶克似乎是个例外(参见哈耶克的《科学的反革命》一书,见 Hayek, 1952)。

这里应该特别指出,尽管古典经济学家们一开始就致力于把自己的经济学作为一种社会科学来建构,但与自 20 世纪初以来那种在唯科学主义 (scientism) 精神指导下大多数经济学家朝着一种高度形式化和技术化的方向拼命发展还是有很大区别。这里应该首先指出,毫无疑问,自 20 世纪初以来,尤其是自第二次世界大战以来,当代经济学在任何一个领域中都取得了巨大的进步,并极大地推进了人们对各种社会经济现象的认识和理解。然而,在当代经济学在各个领域大踏步前进和繁盛发展的同时,也产生了一些非常令人值得深思和反省的现象,那就是在唯科学主义的盛行的当代社会中,经济学“莫名其妙地”且“自发地”走向了一个高度形式化和技术化的演化发展路径,以至于不仅几乎所有经济学家们都致力于把自己的研究和著述用数学语言来表述,用数学模型来构建,而且还衍生出了一套固定的文章写作格式,或言几乎所有经济学文章都按照一个套路——即先选定一个题目,再想出几条假设,然后构建模型,进行数学推导和证明,最后证毕而达致结论——来演练。没有这种高度形式化和技术化的论证和证明形式,任何经济学论文都会被蔑视是为“非科学的”,是一种“storyteller”,从而从事这类研究和以这种“storyteller”形式写文章的学者,在国际学术刊物上几乎不能发表出文章,不能出版书籍,进而不能继续保留大学的教职,更难以争得在国际学术讨论会上发言的机会。

对当代经济学中普遍流行的这种纯技术化和形式化的套路,西方一位经

¹³ 譬如在《经济学的认识论》中,米塞斯(Mises, 1960, 参见中译本, 129 页)曾提出,像自然科学一样,一旦建立起来了关于人类行动的原理的体系,就能像自然科学、逻辑学和数学那样建立起一种人类行动学。为了达致他的人类行动的科学,与对理性主义(rationalism)以及人的理性能力抱有怀疑态度的哈耶克有很大不同,米塞斯则反过来求助于“理性主义”。他认为(Mises, 1960, 参见中译本, 67 页),理性主义在人类行动的领域中引起了两个革命性的变化:一方面,“迄今为止,历史学一直是唯一的有关人的行动的科学,理性主义把批判方法引入了历史学,使这门科学摆脱了对编年史和过去历史著作中流传下来的那些东西的幼稚的附属地位,并教导它不仅要用新的资料如文档、铭文等,而且要对所有来源进行批判性的审查”;另一方面,“理性主义的另一个伟大成就是建立起人类行动的科学理论,即目的在于确定人类行为普遍和正确规律的科学”。米塞斯还接着指出,这门关于人类行动的科学,应当首先在名义上归功于法国的实证主义哲学家和社会学家孔德(August Comte),而其基础,却是为 18 世纪到 19 世纪初的实证主义思想家——首先包括经济学家——所发展起来的。米塞斯还认为,这类人类行动科学中的经济学,是到目前为止“社会学的一门最精良的分支”。在 1933 年为这本书所写的德文版序言中,米塞斯(Mises, 1960, 参见中译本, 5 页)还明确表示:“本书的目的,就是确立[人类行动学]这样一门科学的逻辑合理性,它的研究对象为普遍正确的人类行动的法则。即无论行动者所处的地方、时间、种族、国籍和阶级如何,这都可以宣布为正确的规律”。但是到了晚年,米塞斯才好像对这种当今世界上盛行的唯科学主义有所警惕(是否曾反过来受哈耶克的一些著作的影响?)。比如,在 20 世纪 60 年代为这本书所写的英文版序言中,米塞斯(Mises, 1960, 参见中译本, 1 页)一上来就明确指出:“我们时代所普遍接受的认识论学说,并不承认在自然科学研究的事件领域与作为经济学和历史学研究对象的人类行动领域之间存在根本性的差别。人们充满了关于‘统一科学’的混乱思想,这就是必须根据牛顿物理学研究质量与运动时所依据的方法来研究人类行为。根据这种所谓的研究人类问题的‘实证’方法,他们计划建立‘社会工程’,这种新技术可以使未来有计划社会的‘经济沙皇’能以一种工程师利用技术处理无生命的物质的方式来处理活生生的人”。米塞斯接着追加道:“这些做法完全歪曲了人类行动科学的每一个方面”。

济学家 Warren J. Samuels (1998, p. 354) 曾评论道：“经济学家们使用数量化的技术是如此惯常，以至于根本不考虑它们的知识论和技术的局限。这类研究，一些要达致某种真实的结果，一些则欲达致某些确定性的结论，而另一些则仅仅要达致某些最优解。但是，大多数人在做这类研究时，并没有意识到演绎、归纳和某一特定计量技术的局限。”这种对人类的经济行为以及现实的经济过程仅做纯数理推导、证明和实证检验的“演练套路”，在一些制度经济学家眼中，“不仅仅是理论分析的一种滥用 (misuse)，而且还有一种导致使理性运用无多大意义 (trivialization) 的潜在危险，这一世风很快使经济学变成了一门技术，而不再是一门社会科学” (Kapp, 1968, p. 7)。事实上，早在 20 世纪 30 年代，凯恩斯就曾意识到这种经济学分析纯技术化和完全数量模型化倾向的潜在危险，并在《就业、利息和货币通论》一书中尖锐地指出：“在近来的‘数理’经济学中，只能代表拼凑之物的部分实在是太多了；这些部分的不精确程度正与他们赖以成立的假设条件是一样的。假设条件使那些作者能在矫揉造作和毫无用处的数学符号中，忘掉现实世界的复杂性和相互依赖性” (Keynes, 1936, p. 298)。在当代纯形式化和高度技术化的主流经济学在国际上仍然繁盛发展且初被引进到当今中国社会的态势中，重读凯恩斯在 20 世纪 30 年代说过的这句话，真有一种醍醐灌顶的感觉！

正是鉴于这样一个事实——即几乎所有经济学家们都挤入（或言被迫赶入）当代主流经济学中这样一个高度形式化和技术化的“窄门”或“栈道”中畸形发展，在 1992 年召开的美国经济学年会上，由英国经济学家 Geoffrey Hodgson（今年担任国际制度经济学学会会长）、芬兰经济学家 Uskali Mäki 和美国经济学家 Donald McCloskey 所组织发起，并由四位诺贝尔经济学纪念奖得主 Franco Modigliani, Paul Samuelson, Herbert Simon 和 Jan Tinbergen 以及其他四十余位国际知名经济学家所共同签署了一份“为提倡多元化和严密的经济学的呼吁书”。从这份简短的呼吁书的用辞中，可以看出，这四十余位经济学家已经深切地痛感到目前高度形式化和技术化的主流经济学方法所给当代经济理论研究带来的潜在问题和危险，从而强烈呼吁要鼓励不同经济学流派和研究进路之间的竞争和多元化发展。¹⁴可能也正是因为当代经济学中出现了这种高度形式化分析技术独家垄断的格局，另一方面人们又对这种主要靠数学

¹⁴ 这份“A Plea for A Pluralistic and Rigorous Economics”的三段原文如下：“We the undersigned are concerned with the threat to economic science posed by intellectual monopoly. Economists today enforce a monopoly of method or core assumptions, often defended on no better ground than that it constitutes the ‘mainstream’. Economists will advocate free competition, but will not practice it in the marketplace of ideas.” “Consequently, we call for a new spirit of pluralism in economics, involving critical conversation and tolerant communication between different approaches. Such pluralism should not undermine the standards of rigor; not a less, rigorous science.” “We believe that the new pluralism should be reflected in the character of scientific debate, in the range of contributions in its journals, and in the training and hiring of economists” (American Economic Review, Papers and Proceedings, 1992, p. xxv).

推理和模型建构的经济学分析套路的意义和潜在问题存疑甚多,自 20 世纪中后期以来,在国际经济学界不断有人发出“经济学到底是否是科学?”这样的疑问。

经济学到底是否是科学?在回答这一问题之前,显然应该稍微梳理以下两个问题:

首先,什么是“科学”(英文为“science”)¹⁵?按照《现代汉语词典》的解释,“科学”是“反映自然、社会和思维等客观规律的分科的知识体系”。“科学”一词的这种汉语定义法,显然把“科学”视为一种非常神圣的和庄严的东西了,因为科学被明确规定为“反映客观规律的知识体系”。但是,从词源学上来看,在西方文字中“science”一词的原初含义,并非如此“高贵”。譬如,在《论经济学的历史与方法》一书中,奈特(Knight, 1956, p. 121)就曾指出,“当我们提到社会科学一词时,在起源和词源学意义上对我们来说,‘科学’或多或少地只是知识的一个同义词”¹⁶。如果按奈特对“science”这样最“宽容”和“谦恭”的理解来把握“科学”和“社会科学”,而不是像在现代中文中那种赋予“科学”以“神圣”和“庄严”的含义那样,那么,把经济学视为一门社会科学,应该是没问题的。¹⁷另外值得注意的是,最新版《牛津英语大词典》只是对“science”做了这样的界定:“the intellectual and practical activity encompassing the systematic study of structure and behaviour of physical and natural world through observation and experiment”。如果按此定义,所有社会科学,原来并不是真正的“科学”,而只是在一种在对“科学”一词的“转借”和“隐喻”(metaphor)意义上来说的,即“the scientific study of human society and social relationships”。照此定义,说经济学是“科学”¹⁸,那也只能从词的转借和隐喻意义上理解了。然而,从英语“科学”的这种原本含义中,我们恰好理解了当代主流经济学家欲把经济学变成“科学”之努力的最初缘由。那些想把经济学变成一门实证科学的经济学

¹⁵ 从词源学上来考证,在明万历年间,利玛窦和徐光启最初把拉丁文的“scientia”翻译为“格物穷理之法”。把“science”翻译为中文的“科学”(据考证,宋代学者陈亮最早使用“科学”一词时,原意为“科举之学”),只是近代的事。据汪晖(2005, 105—106 页)考证,最早把“science”翻译为“科学”的,是日本明治维新时期思想家西周。汪晖认为,在 1874 年,西周在《明六杂志》上第一次把“science”翻译为“科学”。另外,学界也有另一种说法,认为是日本近代著名启蒙思想家福泽瑜吉(1835—1901)最早从“分科之学”意义上把“science”翻译为“科学”的。西文“science”一词这种汉译法,在中日甲午战争前后由日本传入中国。

¹⁶ 按照奈特的见解,只要从知道什么、能否知道以及如何知道的视角反思我们是怎样获取社会知识的,最终必定会走向哲学追问(同上)。

¹⁷ 从词源学上来说,奈特的这种定义符合英文“science”的原初含义,因为这个英文词来自拉丁语的“scientia”,其本来的含义就是“知识”和“学问”的意思。笔者甚至揣摩,当亚当·斯密在《国富论》中提到经济学时常常说它是一门“science”时,大致也是在这种拉丁语原意上使用它的——尽管他很看重经济学,并把它视为“政治家和立法者所应该了解的一门学问”。做这种推测的理由是,在斯密写作《国富论》时,现代科学还未真正在西方社会中兴起,故还谈不到后来流行的“唯科学主义”问题。

¹⁸ 照奈特(Knight, 1956, p. 151)看来,这种“科学”是大写的“Science”,而这种大写的“Science”与科学实证论密切相关。

家们，不正是想通过对人类社会中经济现象的“观察”、“实验”以及与之相关联的“(自然)科学(式)地”“推理”和“证明”，来把经济学做得一种像物理学、化学那样一种实证的“分科的知识体系”¹⁹。

其次，在对“科学”一词的含义有了上述把握之后，我们还要大致了解当代经济学中所铺天盖地弥漫着的“科学主义”精神，方能理解为什么这么多经济学家都挤进“实证经济学”的窄门中繁盛发展这一现象。从辞义上来说，目前人们所说的“科学主义”(scientism，亦译为“唯科学主义”)，实际上是一个泛称。就社会科学的思想方法论而言，“科学主义”是指与从19世纪中期以来所出现的实证主义(positivism，包括logical positivism或称logical empiricism)、工具理性主义(instrumental rationalism或instrumentalism)、实用主义(pragmatism)、证伪主义(falsificationism)以及其他科学哲学学派有关的一种泛化了的哲学观。其中，实证主义是唯科学主义的核心理念，并且，实证主义也是当代唯科学主义的最初理论渊源。从哲学史来看，现代科学主义或实证主义是由法国哲学家和社会学家孔德(Auguste Comte, 1798—1857)开始而发展起来的。在1843年出版的《实证哲学教程》中，孔德提出了他影响后世深远的实证主义哲学纲领(参见Comte, 1865)。按照孔德的实证主义哲学观，人类社会的发展，全都要经历三个历史阶段：神学阶段、形而上学阶段和实证阶段。根据这一划分，孔德认为，在他之后，哲学的主要任务就是要彻底摆脱传统的形而上学哲学观，而用实证主义哲学观来代替之。孔德还认为，要达到这个目的，首要的就是要用自然科学的方法来改造哲学和社会科学。依照孔德的实证哲学思路，只有当一门学科完完全全地建立在观察到的事实基础之上，才能真正成为一门实证科学。依照孔德的这种实证主义哲学观，在当代唯科学主义中逐渐形成了三个基本理论信念：(1)自然科学和数学的方法应该被应用于包括哲学、社会科学、人文科学在内的一切研究领域；(2)只有(自然)科学的方法才能富有成效地获取知识；(3)自然科学的知识和研究方法是唯一正确的，是人类知识的典范，因而可以推广运用到研究人类社会的所有问题，包括哲学、伦理学、人类学和宗教学等等，而经济学的研究更是不言而喻地应该首先包括在其中。这种唯科学主义，到了波普尔(Karl Popper)的证伪主义那里，又进一步得到了“科学的”和“哲

¹⁹ 在《实证经济学方法论》的这一当代经济学说史上的经典名篇中，弗里德曼一上来就指出，是老凯恩斯(John Neville Keynes，约翰·梅纳德·凯恩斯的父亲)在《政治经济学的范围与方法》中最早把经济学认定为是“一门独立的实证科学”的(参见Hausman, 1994，中译本，148页)。在老凯恩斯的那本在近现代经济思想史上颇具影响的小册子中，他还区分了“实证科学”和“规范科学”。照他看来，“一门实证科学……是关于‘是什么’的系统知识，而一门规范科学或称为规则科学……则是讨论‘应该是什么’之标准的系统性知识体系”。老凯恩斯还认为，“此二者的混淆甚为普遍，并成为很多谬误的根源，贻害甚大”(Keynes, 1891, pp. 34—36, 46)。在目前科学主义似乎一统“经济学江湖”的态势中，也许有人会问这样一个问题：科学还有“规范的”么？不是自维也纳学派的逻辑实证主义尤其是波普尔的证伪主义以来，所有的规范问题研究——如思辨哲学、伦理学、宗教神学等等——不都被贴上了“玄学”(形而上学)的标签而被排除在“科学”的范围之外了？

学的”强化。按照波普尔的证伪主义和他所提出的科学发现的逻辑,凡是不能被证实或反过来被证伪的理论判断,均是非科学的,或言应该被归结为形而上学问题而被排除在“科学”之外。

由孔德所开始而发展起来的这种实证主义和科学主义的精神,自 20 世纪之后在美国哲学、经济学和社会科学各领域大行其道,并构成了当代主流经济学的最深层的思想方法论基础,以至于像弗里德曼这样的当代泰斗式大经济学家,都会得出这样著名且为人们所常常津津称道的结论:“一个理论越是‘简单’,它对某个特定领域内的现象进行推测的初始知识就越少;一个理论越是‘富于成果’,它对现象做的推测就越精确,而且该理论所能推测的范围也越大,在它启发下更深入的研究路线也就越多”(见 Hausman, 1994, 中译本, 153 页)。结果,在这铺天盖地的实证主义和科学主义精神的浸染下,当代绝大多数经济学家都做起这样一种“想出一个研究题目→做出几个假设→构建模型→进行数学推导证明→得出结论→再进行理论推测或提出政策建议”的“科学论证逻辑”功夫来了²⁰。于是,就有了当代经济学中这种千家万户均演练“实证分析”和“科学解释”的理论格局。

这里需要特别指出的是,尽管当代经济学的分精神和思想方法论基础是实证主义的科学主义,而这种科学主义则是从近代欧洲启蒙精神心态(the enlightenment mentality)中萌发并成长出来的,但是,在当代国际经济学界的激烈生存竞争生态中,世界上绝大多数经济学人很少有时间去回顾并审视一下从近代启蒙运动的科学精神到孔德的原初实证论、再到卡尔纳普(Rudolph Carnap)的新实证主义的演变历史和潜在的问题,很少有人去花时间阅读从 20 世纪 30 年代的维也纳小组的逻辑实证主义到卡尔·波普尔的科学发现的逻辑,到托马斯·库恩(Thomas S. Kuhn)的科学革命的结构理论,再到拉卡托斯(Imre Lakatos)的科学研究纲领的方面的哲学文献,更没有多少人去啃读从皮尔斯(Charles S. Peirce)、詹姆斯(Henry James)、杜威(John Dewey)这些早期美国实用主义哲学家的著作到当代美国新实用主义的大哲学家奎因(Willard van O. Quine)和戴维森(Donald Davison)的深邃艰深的理论文献了。当今国际经济学界的生存生态,好像是一种颇为残酷的达尔文式的“物竞天(刊)择”。在一种“publish or perish”的“物竞刊择”的进化选择机制中,面临着找到一份大学教职或取得某大学永久教席的巨大生存压力,一个个忙忙碌碌的经济学人,哪有时间去阅读这些思辨难懂的哲学文献?哪有精

²⁰ 如果稍微读一点实用主义的哲学文献,就会发现,在当代经济学中所流行的这种经济学作文著书的演练套路,连美国实用主义和工具主义哲学家杜威(John Dewey, 1859—1952)所提出的“思想五步说”都做不到。杜威的思想五步说是:发现疑难,产生问题,提出假说,推理演绎,证实假说。在当今国际经济学界的生存竞争生态中,很多经济学人(当然不是全部)不是为现实和理论中的一些“真问题”(serious problems)做文章,而是多为发表写文章。结果,许多人的研究实际上只有四步:(挖空心思)想出个题目,提出假说,推理演绎,证实(或证伪)假说。——这就是当今一些经济学人进行“科学研究”的现实逻辑!

力去管顾经济学的本体论、认识论、方法论和伦理学这些玄学问题？只要能做出模型来，只要能写出形式漂亮且高度技术化的论文来，且只要写出的论文能发表在国际英文刊物上，管它什么是科学！管它什么是理论研究的现实意义！又何必管它窗外洪水滔天！

对于在国际经济学界生存竞争恶性循环机制的压力下这一流行的理论演练套路，西方一位甚关注经济学发展的哲学家罗森伯格（Alexander Rosenberg）教授曾在一篇《经济学是什么——如果它不是科学》的论文中做了一个整体的把握，并对经济学的认识论和方法论基础做了非常到位的评估。照罗森伯格看来，在这种几乎国际经济学界人人都在演练着的经济学研究套路的核心，是在近代科学中已经取得了巨大成功的牛顿力学和达尔文进化论所共同内含着的一种“研究策略”，他把这种“研究策略”称作为“极值策略原理”：在牛顿力学体系中，“系统永远都在最大化或者最小化一个反映系统中可能的力学状态的变量”；而在达尔文的自然选择理论中，“这个策略假定自然环境最大化适存度这一变量，……在生物学里，假定最大化适存度导致了影响种群中基因变异因素地发现，更主要的是，它导致了遗传学规律的发现”。罗森伯格接着指出：“极值策略是重要的方法论策略，因为它能非常有效地被避免证伪，这使它成为了研究纲领的‘核’，把那些原应成为反常或反例的现象变为了理论的新预测，或把它们作为领域拓展和精度深化的新机遇。相应地，也许可以认为，经济学家之所以钟爱这种理论，与其说反映了他们的自满，不如说反映了他们在方法论上根深蒂固的保守性：由于这个策略在力学与生物学的广泛领域内取得了很大成功，没有道理不试着用它来解释人类行为”（见 Hausman, 1994, 中译本, 336—337 页）。罗森伯格对当代经济学中广为流行的这种科学主义研究套路的思想和研究者心态的这番评论，真可谓是深中肯綮、入木三分！

然而，现在的一个问题是：这种“唯科学”、“唯实证”——或曰“唯发表”、“唯引证”的研究道路就是“科学的”？经由这种研究道路就能使经济学达致像物理学、化学、数学那样高度形式化和精密的大写了的“科学”？

对于这个问题，国际上有不少学者已经发表过许多深邃和独到的见解了，我们不再增加任何新评论。这里，仅让我们回顾一下一些国际著名学者是如何看待这个问题的。

在 20 世纪 50 年代以前，芝加哥经济学派的创始人奈特就在一系列文章和著作中对经济学和其他社会科学的元问题以及社会科学与自然科学研究的区别做出过许多深邃的评论，且今天读来仍然觉得发人深思。譬如，在为奈特 70 大寿所专门汇编的《论经济学的历史与方法》纪念文集中，就收入了奈特在这方面的许多重要文章。譬如，在《社会科学》、《社会的因果律》以及《经济学中的真理是什么》等论文中，奈特（Knight, 1956）曾一再指出，用自然科学方法和分析技术来研究和处理社会科学中的问题，这本身就会带来

许多问题。照奈特看来,这其中的一个主要问题是,社会科学(包括经济学)所研究的人的行为和行动,与自然界的“惰性物体”(inert objects)甚至与生物和其他动物的运动和“行为”是有很区别的。其主要区别在于,人类不仅有依照自己的理性计算进行经济活动的先天禀赋,而且是有着自己的自由意志并按照自己的道德标准、价值观和情感判断来行事的,并且,人会有爱和恨,且反复无常和自我矛盾,还会受“劝说”、“告诫”、“引诱”、“强制”、“欺骗”、“威胁”等等的影晌,因而,作为一个社会的又是单个的行为者,人常常会改变自己的行动和决策。况且,现实中的人是各种各样的,不宜也不应该把他们全划归为一个抽象的类型,从而像对自然界的“惰性物体”一样来进行理论处理。基于上述看法,奈特(Knight, 1956, pp. 124—125)认为,“人类社会是有意识、有目的的无数个人的联合,因而不能被当作一个机体那样而被分解为许多部分来进行分析,……就此而言,个人对社会而言,在逻辑上是先在的(prior)。在社会研究中,一个更为要命的事实是,……在一个社会中,总有出于每人的个人本质(individual nature)而非社会本质而产生的问题”。奈特(Knight, 1956, p. 122)甚至发现,在对人的实证主义和实用主义的研究处理方法中,有一个摆脱不了的内在矛盾:“如果从实证主义的视角来看待人,人可以根本就不行动;如果从实用主义的视角来观察人,人却不能完全自我行动”。考虑到上述种种问题,奈特(Knight, 1956, pp. 175—176)指出:“更为重要的是,在给定的变化条件下预测人的行为和预测自然物体(physical objects)运动的根本区别在于,后者不会非理性地或带有情感地行动,也不会犯错误,不会‘改变主意’(或多或少与他们的反应模式相关);而作为人类,则显然会这么做。与自然物相对照而言的人类的这一特征,……显然使那些作为科学家的经济学家们处在一个非常尴尬的境地,并且似乎经常使他们无计可施”。由此,奈特得出这样一个结论:人类行为的这些独特性,使那些基于先前行为观察而做实证主义预测并试图从中推导出抽象经济法则的做法陷入一个无法摆脱的困境之中,因此,社会科学(包括其中的经济学)是根本不能像自然科学那样来做的。

除奈特外,曾被人们广泛认为引发了当代认识论革命的一位大哲学家和化学家迈克尔·波兰尼(Michael Polanyi, 1962, p. 88)也曾在一系论著中批评并嘲笑当代社会研究中那种唯科学主义者的“现代主义的思想方法论”,认为他们只不过建立了“一种关于有效意义的过于空想而不切实际的标准(quixotic standards),如果严格执行这些标准,相当于我们自愿变成了低能儿”。照波兰尼看来,现代唯科学主义相信,人类的知识是确定无疑的,是独立于形而上学、道德律和个人信念的(personal conviction)。但是,波兰尼却认为,这种唯科学主义,实际上所带给人们的,只不过是一套科学家们自己的形而上学、道德观念和个人信念,只不过使其改了个名字而叫做“科学方法”而已,因而这种现代科学主义,即不能、也不应该带给我们它所承诺的

东西 (Polanyi, 1962, 1964)。波兰尼还认为, 科学知识和其他个人知识, 并没有多大的差异, 因而, 如果欲使科学知识变得与众不同, 而不是把它发展得更好, 那就会把科学推向一条死路 (McCloskey, 1983, p. 488)。

1983年, 一位叫 Alfred S. Eichner 的论者曾编辑出版了一本书名就叫《为什么经济学还不是一门科学》的文集。这本文集中曾收入了 James A. Sweaney 和 Robert Premus 两位作者的文章, 题目为《现代经验主义与经济学中的量子跃迁 (quantum-leap) 推理》。在这篇文章中, Sweaney 和 Premus 曾对作为现代经济学的方法论基础的逻辑实证主义 (亦称“逻辑经验主义”) 做了这样反思性的批评: “很显然, 在作为一种科学而不是讲故事的经济学中的形式主义的失败, 是试图将逻辑实证主义应用于其根本行不通的领域。社会科学与自然科学在许多方面有根本性的差别, 这些差别导致把自然科学方法应用于社会科学时将受到非常大大局限”。这两位作者还认为, 经济学家们对逻辑实证主义方法论的误用, 甚至反过来导致了本来有一定道理和合适用场的逻辑实证主义本身声名狼藉 (参见 Eichner, 1983, p. 42)。这本文集还收入了 Ray E. Canterbury 和 Robert J. Burkhardt 两位作者共同撰写的另一篇题为《我们问经济学是否是一门科学的意思是什么》的文章。在这篇文章中, Canterbury 和 Burkhardt 认为, 在回答经济学是否是一门科学时, “经济学家们实际上并没有遵循他们的规范化诫命, 因而即使按照经济学家们自己的标准, 经济学也不是一门科学。然而, 由于经济学本身有一套范式 (paradigm) 和一些虔诚的从业者, 故在独断的库恩科学哲学的意义 (依照其标准, 须有一套单一的统盖一切的范式, 且该领域大都赞同这一范式) 上来说, 经济学倒是一门科学”。

通过回顾国际上一些学者对经济学是否是一门科学这一问题的讨论, 我们今天可以大致认为, 作为一种对人类社会经济现象进行“科学的研究”而长期演化下来的一门知识体系, 经济学无疑是一门“社会科学”, 并且在 20 世纪与所有其他社会科学门类相比较, 经济学确实已经取得了巨大的理论进展和丰硕的研究成果。然而, 即使我们认定经济学是一门社会科学——从这种隐喻意义上我们可以认为它已经是一种“科学”了——但这并不能遮盖得了这门社会科学的现状中仍然潜含着的种种奇异现象和大量问题。本文所要再次提请学界反思的仍然是奈特和波兰尼所早就反复提出的老问题: 经济学理论研究到底能否依照实证主义方法像物理学、化学、生物学和数学的研究那样来做? 换句话说, 经济学家们欲把经济学建构成像物理学、化学和数学那样一种高度形式化和公理化的“精密科学”是否可能? 并且在多大程度上已经取得了成功? 对于这些问题, 我们今天的一些初步印象是, 用这种唯科学主义的和高度形式化的方法来进行社会问题的研究, 不但有可能使一些经济学人 (当然不是全部) 的学术探讨和理论建构之努力的结果可能是“拣了芝麻, 丢了西瓜”——我这里是说由于太注重形式和技术, 从而或多或少地

忽略了经济学的“经邦济世”的历史的使命和现实的意义。因此,如果说目前这种太注重技术和高度形式化的经济理论研究不是积重难返或者说误入了歧途的话,至少它要达致像物理学和化学那样一种高度形式化和技术化的“精密科学”,还有很长的路要走,甚至这一目标能否最终达到,今天也仍然值得怀疑。或者,经济学理论研究的高度形式化和技术化确实做到了,其论证也高度精密化了,但这些理论推演所产生的成果,却已经不再是斯密意义上的“经济学”了,而只不过是那些自认为是“经济学人”的人们的智力游戏之结晶的一种新的形而上学的符号体系。

五、余论:经济学内部不同研究进路之间的 竞争才能产生富有意义的经济学

在 20 世纪 60 年代,萨缪尔森 (Samuelson, 1963, p. 1772) 曾经发现,在一个学科内部的“研究成就”(fruitfulness)与“对这个学科进行方法讨论的偏好(propensity)”之间,有一种“非常强的负相关关系”(a strong inverse relationship)。萨翁此见,非常值得玩味和深思。有经济学名家的箴言在此,我们今天似乎不应过多地忧虑并花太多精力从经济学方法论的视角来讨论经济学自身发展的一些“形而上学”问题。然而,考虑到当今中国社会以及中国经济学自身发展中的诸多奇异现象以及可能出现的问题,现在讨论并反思一下本文第二、三、四小节所讨论的三个元经济学问题,也许不无助益。1922 年,在为一位英国经济学家罗伯逊的《货币》一书所写的序言中,凯恩斯 (J. M. Keynes, 1922) 曾指出:“经济学不是教条,而是一种方法,一种智慧工具,一种思维方式,它有助于掌握它的人得出正确的结论。”如果像凯恩斯这样以某种“隐喻”式的话语分析视角来审视经济学的性质和任务的话,我们可以从一种相对“humble”但又不无信心地心态来展望经济学理论在中国的未来发展:正如市场竞争会产生效率一样,只有经济学内部诸多流派和研究进路的多元化发展,才最有利于未来中国经济学理论的成长。固守一隅,或只认为某种方法才是“最科学的”,或只朝着一个方向头也不回地死钻牛角尖,这只会导致中国经济学的畸形发展,甚至会在某种程度上致使整个中国经济学理论的发展误入歧途。如果“经济学产品”像物质商品和劳务一样也有一个“市场”的话,只有充分的市场竞争,只有“经济学产品”的多元化生产,才会有未来的“经济学市场”的繁荣。从此视角来看,无论是“经济学市场”中的“人为垄断”,还是如施蒂格勒 (Stigler, 1951) 在谈到“斯密困境”(a Smithian dilemma) 时所发现的那种由市场竞争本身所(可能)自发产生的“经济学市场”的“自然垄断”,都有可能会导致中国的未来经济学理论进入一种新制度经济学家们所喜欢说的那种“锁入”(lock-in) 状态。这不是结论,而只是一个提醒。

参 考 文 献

- [1] Boulding, K., "Is Economics Cultural-Bound?" *American Economic Review (Papers and Proceedings)*, 1970, 60(2), 406—411.
- [2] Cannan, E., *A Review of Economic Theory* London: P. S. King & Son, 1929.
- [3] Comte, A., *A General View of Positivism*, trans. from French by J. Bridges. London: Trubner, 1865.
- [4] Durkheim, É., *De la Division du Travail Social*, Paris: Alcan. Eng. trans., *The Division of Labour in Society*, trans. by W. D. Halls, London: Macmillan, 1902 (1984). 中译本, 涂尔干:《社会分工论》, 渠东译。北京: 三联书店 2000 年。
- [5] Eichner, A., "Why Economics is not yet a Science", in Eichner A. (ed.), *Why Economics is not yet a Science*. Armonk, N. Y.: M. E. Sharpe, 1983, 205—241.
- [6] Friedman, M., *Essays on Positive Economics*, Chicago: Chicago University Press, 1953.
- [7] Groenewegen, P., "Political Economy and Economics", in Eatwell J. et al (eds.), *The New Palgrave: A Dictionary of Economics*, vol. 3. London: Macmillan, 1987.
- [8] Hausman, D., *The Philosophy of Economics: An Anthology*, 2nd ed., Cambridge: Cambridge University Press, 1994. 中译本 豪斯曼,《经济学的哲学》, 丁剑峰译。北京: 世纪文景文化传播有限公司, 2007 年。
- [9] Hayek, F., *Individualism and Economic Order*. London: Routledge & Kegan Paul, 1949.
- [10] Hayek, F., *The Counter-Revolution of Science: Studies on the Abuse of Reason*. Indianapolis: Liberty Press, 1952.
- [11] Hayek, F., *Law, Legislation and Liberty*. London: Routledge & Kegan Paul, 1982.
- [12] Keynes, J. M., "Introduction to Cambridge Economic Handbooks", in Sir Robertson, D., *Money*, London: Nisbet, 1922.
- [13] Keynes, J. N., *The Scope and Method of Political Economy*. London: Macmillan, 1891.
- [14] Knight, F., 1956, *On History and Methods of Economics*. Chicago: Chicago University Press, 1956.
- [15] MacLeod, H., "What is Political Economy?" *Contemporary Review*, 1875, 25, 871—893.
- [16] Marshall, A., *Principles of Economics*, 8th ed.. London: Macmillan, 1920.
- [17] McCloskey, D., "The Rhetoric of Economics", *Journal of Economic Literature*, 1983, 21(2), 481—517
- [18] Mises, L., *Human Action: A Treatise on Economics*. London: W. Hodge, 1949.
- [19] Mises, L., *Epistemological Problems of Economics*, translated by George Reisman, Princeton, N. J.: Van Nostrand, 1960. 中译本 米塞斯,《经济学认识论问题》, 梁小民译。北京: 中国经济科学出版社, 2001 年。
- [20] Myrdal, G., *Objectivity in Social Research*. New York: Pantheon, 1969.
- [21] Polanyi, M., *Personal Knowledge: Towards a Post-Critical Philosophy*. Chicago, IL: University of Chicago Press, 1962.
- [22] Polanyi, M., *Science, Faith and Society*. Chicago: University of Chicago Press, 1964.
- [23] Redman, D., *Economics and the Philosophy of Science*. Oxford: Oxford University Press, 1993.
- [24] Robins, L., *An Essay on the Nature and Significance of Economic Science*. London: Macmillan, 1948. 中译本 罗宾斯,《经济科学的性质和意义》, 朱泱译。北京: 商务印书馆 2000 年。
- [25] Samuelson P., "Discussion on Ernest Nagel's 'Assumptions in Economic Theory'", *American Economic Review (Papers and Proceedings)*, 1963, 53(2), 231—236.
- [26] Schumpeter, J., 1949, "Science and Ideology", *American Economic Review*, 1949, 39(2), 345—359.
- [27] Schumpeter, J., *History of Economic Analysis*. New York: Oxford University Press, 1954.
- [28] Senior, N., *An Outline of the Science of Political Economy*. London, G. Allen & Unwin, 1836/1938.

- [29] Smith, A., *An Inquiry into the Nature and Causes of the Wealth of Nations*, 5th ed. . London, Methuen & co., 1776/1930 .
- [30] Smith, A., *Lectures on Jurisprudence*. Cambridge : Cambridge University Press, 1978.
- [31] Stigler, G., "The Division of Labor is Limited by the Extent of the Market", *Journal of Political Economy*, 1951, 59(3), 185—193.
- [32] Stigler, G., *The Essence of Stigler*, ed. by Kurt Leube & Thomas G. Moore. Stanford, CA :Hoove Institution Press, 1986. 中译本 施蒂格勒,《施蒂格勒精萃》,吴珠华译。北京 :商务印书馆,1999 年。
- [33] 汪晖,“科学话语共同体和新文化运动的形成”,《学术月刊》2005 年第 7 期,第 104—113 页。
- [34] 韦森,《经济学与伦理学:探寻市场经济的伦理维度与道德基础》。上海 :上海人民出版社,2002 年。
- [35] 韦森,“货币、货币哲学与货币数量论”,《中国社会科学》2004 年第 4 期,第 61—67 页。
- [36] Whately, R., *Introductory Lectures on Political-Economy*. London :B. Fellowes, 1831.

The Nature of Economics in a Perspective of Philosophy

WEISEN LI

(Fudan University)

Abstract Based on a review of relevant literature in the history of economic thought, this paper shows that it is time to make some reflections on the nature, role and significance of economics in the current situation of the dynamic growth of the Chinese economy and the theoretical economics in China. Section Two finds that along with transformation from “political economy” to “economics”, economists’ understanding of the nature and role of economics also changed. Section Three discusses the question whether economics could become a social science of value-free or not and finds that any school in history of economic thought may have some value judgment and ethical foundation in its theoretical discourse. As a result, the top priority for economists is to try to find the “natural order” or “providence” in human economic and social reality. Section Four discusses the possibility whether economics could become a rigorous science like physics, chemistry and many other highly formalized natural sciences. It also makes some reflection on the propensity of scientism of the mainstream economists. Section Five reaches the conclusion that for the healthy growth of economic theory in China, competitions and dialogues between different schools and thus pluralism in economics should be encouraged.

JEL Classification A11, A12, B40

本地市场效应：理论和经验研究的新近进展

钱学锋 梁琦*

摘要 本地市场效应是指,在一个存在报酬递增和贸易成本的世界中,那些拥有相对较大国内市场需求的国家将成为净出口国。它源自新古典贸易理论解释复杂贸易实践的理论局限和新贸易理论经验证据匮乏的困境,在探寻两种理论范式相对解释力的过程中应运而生,并随着空间经济学的迅猛崛起而成为新经济地理范式(New Economic Geography Paradigm)的核心特征之一。本文在回顾本地市场效应的理论背景和基本理论模型的基础上,对其理论模型和经验证据的新近发展作了系统的梳理并进行了简要的评论。

关键词 本地市场效应,空间经济学,贸易

Krugman (1980)指出,在一个存在报酬递增和贸易成本的世界中,那些拥有相对较大国内市场需求的国家将成为净出口国。此即“本地市场效应”(Home Market Effects)¹。它源自新古典贸易理论解释复杂贸易实践的理论局限和新贸易理论经验证据匮乏的困境,在探寻两种理论范式相对解释力的过程中应运而生,并随着空间经济学的迅猛崛起而成为新经济地理范式(New Economic Geography Paradigm)的核心特征之一。同时,由于它还还为不同规模的国家提供了显著的福利含义,因此也成为经济一体化和全球化时代诸多国家政策讨论的中心(Chung, 2002)。

一、理论背景

众所周知,国际贸易理论主要沿着两种理论范式发展,即完全竞争、规模报酬不变的新古典框架(Perfect Competition and Constant Returns to Scale,

* 钱学锋,中南财经政法大学工商管理学院,南京大学经济学院;梁琦,中山大学管理学院,南京大学经济学院。通讯作者及地址:钱学锋,中南财经政法大学工商管理学院,430073;电话:13487077155;E-mail:xfqian@126.com。本文是国家社会科学基金项目“中国制造业分工与地区经济增长研究”(04BJL052)的研究成果,作者感谢匿名审稿人的建设性评论。

¹ 关于 Home Market Effects 的译法问题,目前国内主要存在两种:一是“本地市场效应”,二是“本土市场效应”。梁琦(2004)最早译为“本地市场效应”(p. 96);梁琦在译著《空间经济学:城市、区域与国际贸易》中也将其翻译成“本地市场效应”(p. 28)和“国内市场效应”(p. 68)。而张帆、潘佐红(2006)、何茵、李东、姚洋等(2006)则译为“本土市场效应”。我们认为,在国际贸易中,翻译成“本土市场效应”或“国内市场效应”是比较合适的,它是指一国规模报酬递增的行业由于存在超常需求而产生的向国外出口的情况。但翻译成“本地市场效应”,其应用范围将更为广泛。因为在一国内部次级区域之间也可能由于这种效应而产生贸易,这时如果翻译成“本土市场效应”似乎就不太准确。而翻译成“本地市场效应”就不存在这个问题,相对于国外可以说“本地”,相对于国内不同区域也可以说是“本地”。因此,我们觉得翻译成“本地市场效应”将使 Home Market Effects 不仅适用于国际贸易,也适用于国内不同区域之间的贸易。

以下简称 CRS-PC) 和垄断竞争、规模报酬递增的新贸易理论框架 (Monopolistic Competition and Increasing Returns to Scale, 以下简称 IRS-MC)。在 CRS-PC 框架下, 国家间的相对技术和要素禀赋的差异决定了贸易和专业化的模式; 而在 IRS-MC 框架下, 决定国家间贸易和专业化模式的则是递增的规模报酬。但这并不意味着二者不能相容, 它们可以共存于同一个模型中, 一些部门属于 CRS-PC 框架, 而另一些部门则呈现 IRS-MC 的特征 (Trionfetti, 1999)。然而, 不同理论框架下所蕴涵的经济政策含义和福利分配则是完全不同的, 因此, 探寻识别两种理论范式的途径并尝试量化它们分别与各种贸易和专业化模式的相关性就显得尤为重要 (Br  hart and Trionfetti, 2005)。

传统上, 主要存在三类研究尝试间接地测度主要贸易理论范式的相对解释力 (Trionfetti, 1999; Br  hart and Trionfetti, 2005)。第一类研究属于产业组织的经验文献。这类研究直接测度规模经济但很少有支持性证据²; 第二类研究关注产业内贸易并将其作为 IRS-MC 范式重要性的证据 (Greenaway and Milner, 1986; Leamer and Levinsohn, 1995)。然而, 产业内贸易作为 IRS-MC 范式重要性的证据却是不确定的, 因为有研究表明 (Falvey and Kierzkowski, 1987; Davis, 1995), 产业内贸易也可能在拓展的 CRS-PC 框架内发生; 第三类研究则试图发掘大量可靠的引力方程的实证成果以支持 IRS-MC 范式。确实有研究表明 (Helpman, 1987) 引力方程直接对应于 IRS-MC 模型。但是, 引力方程也可以从许多其他类型的模型中推导出来 (Davis and Weinstein, 2001; Feenstra, Markusen and Rose, 2001)。由此可见, 上述三类研究都不能很好地为 IRS-MC 范式提供充分的证据, 从而区分两类理论范式的相对解释力³。因此, 直接构建区分两类理论范式的理论模型并采用统计方法或计量方法推断出可识别的假设就显得尤为迫切。这类工作以 Davis and Weinstein (1996, 1999, 2003) 为起点。他们基于 Krugman (1980) 的 IRS-MC 模型发展了一个识别标准: 超常需求 (Idiosyncratic Demand)。它体现了一个大于 1: 1 的专业化模式, 从而引起本地市场效应。特别地, Davis and Weinstein (1996) 指出, 在比较优势模型中, 对某种商品的超常需求将导致它的进口; 而在报酬递增模型中, 拥有超常需求的国家将会成为生产区位并出口该商品。由于本地市场效应不可能出现在 CRS-PC 模型中, 因而这一特征可以作为实证检验两种范式的基础。正是因为本地市场效应的这种特殊性质, 大量的实证检验本地市场效应的文献自 Davis and Weinstein (1996) 以后开始出现。⁴

然而, 实证检验需要得到理论模型的支撑。尽管 Krugman (1980) 在模

² 相关的文献综述可参见 Scherer and Ross (1990) 和 Tybout (1993)。

³ 也难怪 Krugman (1994) 在评论新贸易理论的经验证据时无奈地问道: “世界贸易的多少可以由报酬递增来解释, 而不是比较优势? 这是一个很难精确回答的问题。但很显然, 精确答案就在那里, 只不过我们不知道而已。”

型中证明了本地市场效应的存在,但由于他模型中的严格假设条件的限制,并不意味着本地市场效应的普遍存在性。因此,本地市场效应的普遍存在性需要得到进一步的理论证明。而肇始于20世纪90年代初期的空间经济学的迅猛发展则恰恰为本地市场效应提供了强大的理论基础。基于IRS-MC框架,空间经济学主要分析在集聚力和分散力的相互作用之下企业空间区位的选择。尽管企业空间集聚的机制多种多样,但无论在何种机制下,都呈现出不同程度的本地市场效应或本地市场放大效应(Home Market Magnification Effects)(Baldwin et al., 2003):即区位需求的一个细小外生变化将使需求扩大区产生一个更大比例的产业再定位。因此,在某种程度上可以认为,本地市场效应是空间经济模型中的核心特征之一(Davis and Weinstein, 1996; Fujita, Krugman and Venables, 1999; Schumacher, 2003; Okubo and Rebeyrol, 2006)。

二、基本理论模型

Krugman (1980) 和 Helpman and Krugman (1985) 描述了一个单一生产要素、两国、两部门模型。一个部门在 CRS-PC 框架下生产同质产品,不存在运输成本;另一个部门在 IRS-MC 框架下生产水平差异化产品,该部门存在固定成本和不变的边际成本,运输成本采用冰山型,即运输 τ 单位产品只有 1 单位到达目的地。效用函数是 C-D 函数和 CES 函数的复合形式。如果 μ 是本国相对于外国生产的产品数量, $\phi < 1$ 是贸易自由度, λ 是本国相对于外国的需求比例。Krugman (1980) 指出⁵:

$$\mu = \frac{\lambda - \phi}{1 - \phi\lambda}. \quad (1)$$

如果本国和外国有着相等的需求规模 ($\lambda = 1$), 则本国和外国生产的产品数量相等 ($\mu = 1$); 但如果本国对 IRS-MC 部门产品需求稍微上升 ($\lambda > 1$), 则本国相对于外国将生产更大比例的 IRS-MC 部门产品 ($\mu > 1$) 从而出现本地市场效应并成为该产品的净出口国。

Davis and Weinstein (1996, 1999, 2003) 对 (1) 式求导得到:

$$\frac{\partial \mu}{\partial \lambda} = \frac{1 - \phi^2}{(1 - \phi\lambda)^2} > 1. \quad (2)$$

⁴ 这里必须指出 Justman (1994) 的贡献, 他较早地强调了区位需求对于本地生产的重要性。他侧重检验的是本地需求和生产之间的相关性, 虽然得到了二者之间的正向关系, 但这并不意味着产生本地市场效应。因为正向关系可能是大于 1 也可能是小于 1, 本地市场效应中生产对需求的反应是大于 1 的。

⁵ 必须说明的是, 在 Krugman (1980) 中, Krugman 并没有采用这一符号。这里为了使定义更清晰, 我们采用 Davis and Weinstein (1996, 1999, 2003) 根据 Krugman (1980) 改进的符号表示。

从而进一步证明了 Krugman (1980) 模型中本地市场效应的存在。Head and Mayer (2004) 指出, 当两国具有对称结构时 ($\lambda = 1$), 我们就得到了经典的本地市场效应 (Classic HME):

$$\frac{\partial \mu}{\partial \lambda} = \frac{1 + \phi}{1 - \phi} > 1. \quad (3)$$

为什么在 IRS-MC 框架下会出现本地市场效应呢? Krugman (1980) 只是简单地分析了企业有定位于大的市场实现规模经济和减少运输成本的需要。Davis and Weinstein (1996, 2003) 则进行了具体的说明: 在贸易成本存在的情况下, 厂商有定位于需求较大市场的激励。这同时意味着越来越多的厂商将离开较小的市场, 因为过高的贸易成本使得它们无法通过出口的方式将产品输入较大的市场。因此, 超常需求将促使厂商不断进入较大市场追求剩余从而产生本地市场效应。但是在 CRS-PC 框架下, 需求的增加仅在价格上升的情况下才能引起供给的增加, 由于无贸易成本, 这将引起外国对本地市场的出口增加。因此, 超常需求将由本地的额外供给和增加进口来得到满足, 本地市场效应并不会出现。

三、模型的拓展

然而, Krugman (1980) 也承认, 由于严格的假设条件, 他模型中的本地市场效应并不意味着普遍存在性, 模型需要在多个方面进行拓展。比如, 考虑存在运输成本的农业部门、异质企业、企业策略性行为、贸易成本、跨国公司、多国框架等假设条件下, 本地市场效应还会继续存在吗? 或者它将发生什么样的变化? Davis and Weinstein (1996) 则简明扼要地指出, 尽管并不存在一个理论模型融入了所有这些假设条件, 但为了实现经验检验, 一个完备的空间经济理论则必须至少包括报酬递增和运输成本。事实上, 近年来, 与本地市场效应相关的空间经济理论模型除了通过修改或放松 Krugman (1980) 和 Helpman and Krugman (1985) 的模型而在多个方向上进行拓展之外, 还开始考察了本地市场效应的福利含义。

(一) 本地市场效应与农业部门

现有文献的一个普遍性假设是存在一个在 CRS-PC 框架下生产和自由贸易的同质产品部门 (农业部门)。该部门存在一方面可以保证要素价格均等化, 并提高模型的可解性; 另一方面它可以弥补 IRS-MC 部门的贸易不平衡, 从而导致国际专业化 (Crozet and Trionfetti, 2006)。然而, 自由贸易的 CRS-PC 部门虽比较简化却不太符合现实。正如 Head and Mayer (2004) 评述的: “CRS 部门可能并不存在零贸易成本或者吸收所有贸易不平衡的能力”。这就意味

着,如果放松 CRS-PC 部门零贸易成本的假设,本地市场效应还会存在吗?

Davis (1998) 发现,同质产品部门的零贸易成本假设并非无关紧要的。他在 Krugman (1980) 和 Helpman and Krugman (1985) 模型的基础上,假定农业部门同样存在一个冰山型运输成本 γ 。特别的,如果同质产品部门的运输成本与制造品部门的相同 ($\gamma = \tau > 1$), 那么 $\mu = \lambda$, 即本地市场效应消失。⁶

Yu (2005) 对 Davis (1998) 的模型进行了进一步的拓展,假定同质产品没有贸易,并且用 CES 函数取代 C-D 函数作为上层效用函数⁷: $U = (C_x^\rho + C_y^\rho)^{1/\rho}$, $\rho \in (-1, 1)$ 。 $\eta = 1/(1 - \rho)$ 是差异产品 C_x 和同质产品 C_y 之间的替代弹性。他证明,当 $\eta > 1$ 时, $\mu > \lambda$; 当 $\eta = 1$ 时, $\mu = \lambda$; 当 $\eta < 1$ 时, $\mu < \lambda$ 。这说明,本地市场效应的存在、消失或者逆向本地市场效应 (Reverse HME, 指 IRS 部门生产份额或企业份额的变化小于需求或支出变化的份额,但二者之间仍为正向关系) 的出现,取决于差异产品和同质产品之间的替代弹性是大于、等于还是小于 1。

Zeng and Kikuchi (2006) 认为,尽管 Davis (1998) 和 Yu (2005) 提醒人们农业部门应该引起重视,但在他们的模型中假定农业部门过高的运输成本可能导致该部门不存在贸易,而且农业部门被标准化处理。这也不太符合现实。因此, Zeng and Kikuchi (2006) 的目的是弄清在国家间农业部门存在贸易的情况下农业部门和本地市场效应的关系。传统上,这种关系由于农业部门被标准化而隐藏。他们在两个方面重新构建了农业部门: 第一,将农业不作为标准化部门,从而可以识别农业部门对本地市场效应的影响; 第二,假定两国生产的农产品是差异化的,这样从外国进口大量的农产品将不会改变本国农产品的价格。他们发现,本地市场效应可以表示为 $\mu - \lambda = \Delta \times (\lambda - 1/2) > 0$ 。 Δ 是差异产品运输成本 τ 和农产品运输成本 γ 的函数且为正, $\lambda \in (1/2, 1]$ 。因此,本地市场效应并没有因为农业部门运输成本的存在而消失。 Zeng and Kikuchi (2006) 进一步认为, Davis (1998) 和 Yu (2005) 的结论只能在两国之间不存在同质的农产品贸易时才能成立。

尽管 Davis (1998)、Yu (2005) 和 Zeng and Kikuchi (2006) 的结论有所不同,这主要取决于他们所选用的模型和假设条件。但至少说明, Krugman (1980) 模型所得出的结论并不能广为适用。

(二) 本地市场效应与异质企业

现有文献假设 IRS-MC 部门的企业有着相同的成本函数,即都存在一个固定成本和不变的边际成本。这意味着国家间的企业是同质的。近年来空间经

⁶ Krugman and Venables (1999) 对 Davis (1998) 的结论作了回应,他们认为,如果允许规模报酬不变部门的产品差异化,即便两个部门的贸易成本相同,本地市场效应将会重新出现。

⁷ 用 CES 函数代替 C-D 函数作为上层效用函数,将使得用于差异产品的支出内生化。

济学的一个新的研究路径是关于企业异质性贸易模型,这类模型也涉及本地市场效应,目前主要有两个分支:一是以 Melitz (2003) 为代表的没有区位变化的自由进入模型 (Free-entry Model without Delocation)⁸;二是以 Baldwin and Okubo (2004) 为代表的没有自由进入的区位变化模型 (Delocation Model without Free-entry)。这两个分支的共同特征在于都是在 D-S 垄断竞争模型中考虑了企业边际成本差异 (企业异质性),且所有模型都发现了本地市场效应。但必须指出的是,这类模型不仅仅关注本地市场效应,实际上,这一研究路径重点在于考察关于企业生产率的异质性对国际贸易的影响 (Okubo and Rebeyrol, 2006)。因此,这些模型中所发现的本地市场效应只不过是顺带的副产品。

基于 Baldwin and Okubo (2004), Okubo and Rebeyrol (2006) 根据市场规模引入进入沉没成本或规制成本 (Market Entry Sunk Costs or Regulation Costs) 构建了一个新的异质企业贸易模型。当引入规制成本时,他们发现:

$$(S_n - 0.5) = (S_E - 0.5) \left(\frac{1 + \phi}{1 - \phi} \right) \left(1 - \frac{\sigma}{\mu} \frac{2\beta((S_n - S_n^2) + \phi/(1 - \phi)^2)}{(1 + \phi)} \right). \quad (4)$$

(4) 式中左边为本国相对生产份额,右边第一项为本国相对需求规模,第二项即为 (3) 式中的经典本地市场效应,第三项括号中的第二项为规制成本。显然,规制成本作为一种离心力将会弱化经典的本地市场效应。原因在于,规制成本与市场规模正相关,相应的,大国意味着较高的规制成本,而小国意味着较小的规制成本。这同时意味着规制成本的差异有利于小国。这说明大国可能主动接受有关规制方面的国际协定,而小国则不太愿意接受并坚持自己单方面的规制。而且,如图 1 (a) 所示,与经典的线性本地市场效应不同,此处的本地市场效应呈现出非线性,当规制成本足够大时,甚至会出现逆向本地市场效应和反向本地市场效应 (Anti-HME, 即在 IRS-MC 部门,生产份额的变化和需求份额的变化之间负相关)。

当进一步引入企业异质性时,由于规制成本的存在,高生产率的企业将有能力抵消规制成本的影响,定位于大国享受较大的市场规模,而低生产率的企业由于无法承受较高的规制成本将不得不定位于小国。因此,企业异质性触发了双向空间排序 (Two-way Spatial Sorting): 高生存率的企业从小国移向大国而低生产率的企业从大国移向小国,并引致平均生产率的差异。因此,与存在规制成本的同质企业模型相比,企业异质性可能减小离心力,反而促进集聚力,并增强本地市场效应 (如图 1 (b))。

⁸ 这一分支的模型还有 Falvey et al. (2004), Helpman et al. (2003), Baldwin and Robert-Nicoud (2004) 以及 Baldwin and Forslid (2004)。

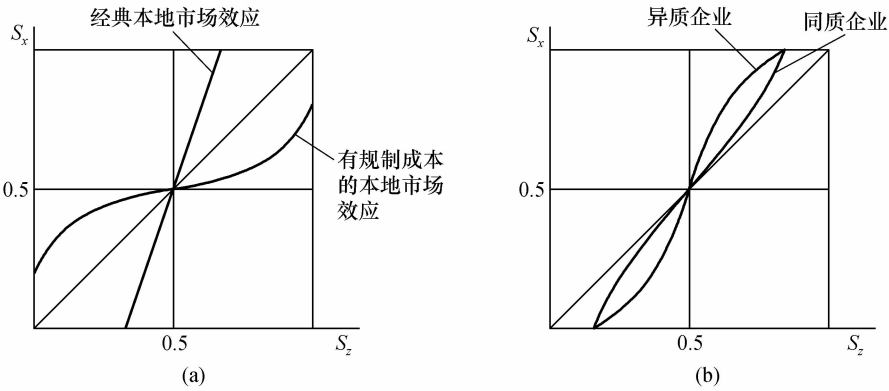


图1 规制成本、异质企业与本地市场效应

资料来源：Okubo and Rebeyrol (2006)。

(三) 本地市场效应与企业策略性行为

在 IRS-MC 框架下，企业之间应该存在策略性行为。但遗憾的是，目前绝大部分文献在考察本地市场效应时都忽略了这一点。Head, Mayer and Ries (2002) 弥补了这一不足。他们在一个统一的框架下考虑了一个两阶段的博弈。首先，IRS 部门的企业将选择一个区位，然后选择价格或产出。当企业选择价格时，Head, Mayer and Ries (2002) 采用的是 Helpman-Krugman (1985) 模型和 Ottaviano-Tabuchi-Thisse (2002) 模型，前者采用 C-D 和 CES 复合效用函数和冰山型运输成本，后者采用拟线性效用函数和线性运输成本；当企业选择产出时，他们分别采用 Brander (1981) 模型和 Markusen-Venables (1988) 模型，前者考虑了企业间同质产品的古诺竞争，后者则考虑了国家产品差异 (National Product Differentiation) 的古诺竞争。

他们发现，当采用 Helpman-Krugman (1985) 模型时，结论简单明确，一国 IRS 部门的企业份额是消费者份额的线性函数，且斜率大于 1。即存在本地市场效应。当采用 Ottaviano-Tabuchi-Thisse (2002) 模型时，本地市场效应可以表示为：

$$\frac{dS^*}{dx} = \frac{2(2(1 - \omega) - \tau)}{\tau\theta N} > 4. \tag{5}$$

这说明，当在模型中采用拟线性效用函数和线性运输成本时将出现强烈的本地市场效应。当采用企业间同质产品古诺竞争的 Brander (1981) 模型时，也出现了较强的本地市场效应：

$$\frac{dS^*}{dx} = \frac{2(1 - \omega - \tau/2)}{\tau N} > 2. \tag{6}$$

但显然小于采用 Ottaviano-Tabuchi-Thisse (2002) 模型时的本地市场效应

(见图 2)。最后,当采用国家产品差异古诺竞争的 Markusen-Venables (1988) 模型时,企业份额和消费者份额之间呈现非线性关系,更为重要的,国别差异产品替代弹性较小时:

$$\frac{dS^*}{dx} = \frac{(N+2)\tau(2-\tau)}{2N[(\tau-1)^2+1]} < 1. \quad (7)$$

这说明出现了逆向本地市场效应:大国拥有较少比例的企业份额并成为 IRS 部门的净进口国。

因此,Head, Mayer and Ries (2002) 的结论表明,在 Krugman (1980) 和 Helpman and Krugman (1985) 模型中出现本地市场效应的几个假设并非必须的。首先,产品差异化并非必须的,因为同质产品的 Brander 模型展现了本地市场效应;其次,放松复合效用函数和冰山型的运输成本本地市场效应也是稳定的;最后,本地市场效应并不依赖于缺少价格反应的 D-S 模型。

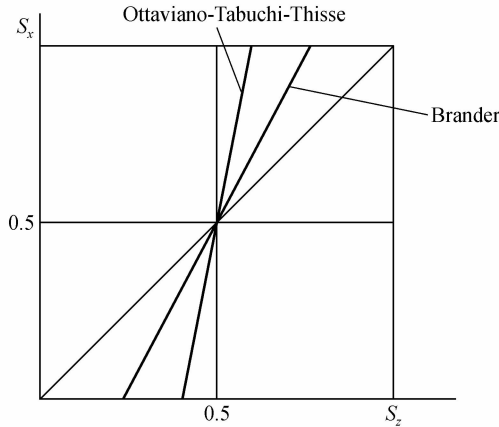


图 2 Brander 和 Ottaviano-Tabuchi-Thisse 模型中的本地市场效应

资料来源:Head, Mayer and Ries (2002)。

(四) 本地市场效应与贸易成本

贸易成本对本地市场效应的产生至关重要。在经典的本地市场效应中,贸易成本的增大将弱化本地市场效应。但这是假定劳动的工资弹性为零(完全弹性)的结果。Head and Mayer (2004) 考察了不完全劳动工资弹性时贸易成本对本地市场效应的影响。他们基于 Fujita et al. (1999) 中的本地市场效应,将其改写为:

$$d\lambda^*/d\theta = M/(1 + (1 + (M^2 - 1)\sigma)/\eta). \quad (8)$$

当劳动工资弹性 $\eta \rightarrow \infty$ 时, $d\lambda^*/d\theta = M = (1 + \phi)/(1 - \phi) > 1$, 即经典的本地市场效应;但当劳动工资弹性较小时,它将导致大国市场的高工资从而

弱化甚至出现逆向本地市场效应。⁹例如，当 $\phi \rightarrow 0$ ，只要 η 为不完全劳动工资弹性，总有 $d\lambda^*/d\theta = 1/(1 + 1/\eta) < 1$ 。因此，本地市场效应与贸易成本之间的单调递减关系仅当完全劳动工资弹性时才成立（图3）。

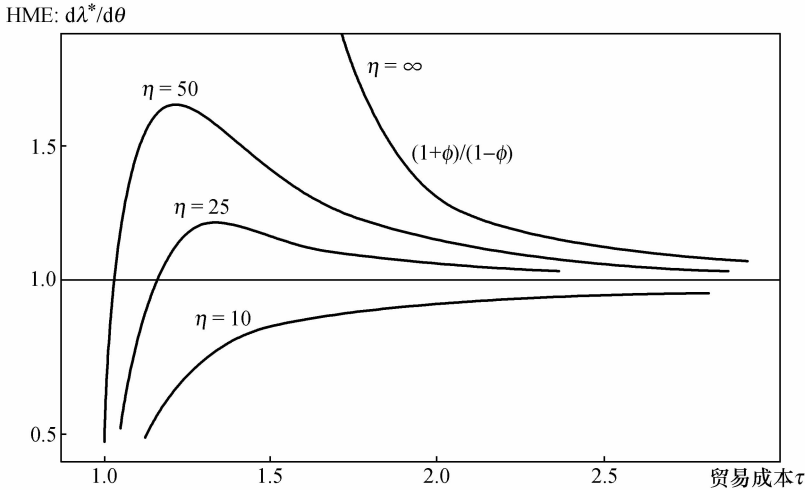


图3 贸易成本、劳动工资弹性与本地市场效应

资料来源：Head and Mayer（2004）。

另外，Crozet and Trionfetti（2006）放弃了 CRS-PC 部门，构建了一个新的两部门模型：一个 IRS-MC 部门，一个 CRS-PC-A 部门（国家产品差异化模型）。他们从另外一个角度考察了贸易成本变化对本地市场效应的影响。他们发现，IRS-MC 部门存在本地市场效应，但 CRS-PC-A 部门本地市场效应消失了。进一步的，他们又发现：

$$\frac{d(dS_N/dS_L)}{d\phi_A} > 0. \tag{9}$$

这意味着 CRS-PC-A 部门贸易成本的增加（ ϕ_A 下降）弱化了 IRS-MC 部门的本地市场效应。另一方面：

$$\frac{d(dS_N/dS_L)}{d\phi} > 0, \quad \frac{d(dS_A/dS_L)}{d\phi} < 0. \tag{10}$$

这说明，两个部门的贸易成本增加将弱化 IRS-MC 部门的本地市场效应，但增强了 CRS-PC-A 部门生产对于需求的反应程度。显然，Crozet and Trionfetti（2006）关于贸易成本对 IRS-MC 部门本地市场效应的影响与经典情形下相同。

⁹ 这与 Head and Mayer（2001）的结论一致，他们也认为逆向本地市场效应的原因在于不完全的劳动工资弹性。

(五) 本地市场效应与跨国公司

跨国公司可以通过在国外设厂的方式在不发生运输成本的情况下利用较高的外国需求,从而破坏了本地市场效应的驱动力。但目前绝大多数研究忽略了跨国公司大量存在的现实。Larch (2003) 证明,在跨国公司存在的情况下,本地市场效应可以通过另外的渠道,即利润汇回而不是产业间贸易重新出现。Larch 首先在垄断竞争的市场结构下引入水平型跨国公司,该公司在两个国家有一个总部和一个工厂;并且假定贸易和 FDI 并存,存在资本、技能型劳动和非技能型劳动三种要素。然后推导出了本地市场效应的公式:

$$\text{HME} = \frac{\frac{(n_i + m_i + m_j)x_{ii} + n_i x_{ij}}{(n_j + m_i + m_j)x_{ji} + n_j x_{ji}}}{\frac{\phi}{1 - \phi}}, \quad (11)$$

式中,分子衡量的是本国相对生产比例,分母是三种要素衡量的本国相对需求规模。其中, $n_i(n_j)$ 代表 $i(j)$ 国本土企业的数量, $m_i(m_j)$ 代表在 $i(j)$ 国有总部的跨国公司的数量, x 代表产出(需求), ϕ 是要素禀赋衡量的国家规模。但由于该式并不具备解析解,Larch 采用数值模拟的方式分几种情形进行了讨论:(1) 如果仅有本国企业且两个部门的贸易成本相同,那么不存在本地市场效应(Davis, 1998)。如果允许两个部门存在不同的贸易成本,基于差异产品的贸易将导致本地市场效应,它与差异产品部门的需求份额负相关,而与差异产品的替代弹性正相关;(2) 如果仅有水平型跨国公司,将不存在本地市场效应;(3) 如果本国企业和跨国公司同时存在,那么将会出现本地市场效应,但它源于利润汇回的渠道而不是产业间贸易。而且,它同贸易成本以及差异产品部门的需求份额正相关,同差异产品的替代弹性负相关;(4) 如果两种类型的企业同时存在,但当差异产品部门的需求份额非常低或者差异产品的生产变得更技能型劳动密集型和较低的资本密集型时,且额外被用于跨国公司的技能型劳动足够少,那么逆向本地市场效应将会出现。Larch (2003) 结论的主要经验含义在于可以判断哪些部门跨国公司更重要,哪些部门本国企业更重要。

Toulemonde (2005) 也发展一个两部门(CRS-PC 和 IRS-PC)非对称规模的两国模型以考察跨国公司对本地市场效应的影响。在模型里,企业必须考虑它需要多少工厂在何处进行生产。Toulemonde 发现,在运输成本较高时,所有的企业都将是跨国公司型的,跨国公司将在两个国家同时生产并以当地工厂满足各自市场的需求。因此,一国生产水平将会与其市场规模等比例增长,不存在本地市场效应;当运输成本较低时,本地市场效应重新出现;而且随着运输成本的降低,更多的企业定位于大国,跨国公司将在一个国家进

行生产,本地市场效应得到强化。然而,Toulemonde也发现,当每个国家都存在单一工厂企业和跨国公司时,小国的生产将会增加。因此,在这种情况下,即使大国获取了更多的贸易自由化的利益,两国的生产都增加了。但是当贸易成本低到所有的跨国公司集中生产时,贸易自由化仅仅使大国获益。

(六) 本地市场效应与多个国家

Krugman (1993) 指出,在一个多国经济中,本地市场效应没有直接的定义,因为没有有一个显而易见的基准来衡量不完全竞争部门企业超比例生产的存在。但正式在多国框架下考察本地市场效应的是 Behrens et al. (2004)。Behrens et al. (2004) 认为,难以在多国框架下构建一个本地市场效应假设的原因在于,国家作为生产区位的吸引力既取决于相对国内市场规模(吸引力)也取决于同其他外国市场的接近程度(准入性)。因此,在多国框架下要想对本地市场效应进行一般性的精确定义,就必须区分动态本地市场效应(Dynamic HME)与静态本地市场效应(Static HME)。

在多国框架下,动态本地市场效应要求: $\sum_j \frac{\partial \lambda_i^*}{\partial \theta_j} \frac{d \theta_j}{d \theta_i} \frac{\theta_i}{\lambda_i^*} > 1$ 。很显然,

在这种情况下由于无法确定需求比例 θ , 因而无法对本地市场效应进行定义。Behrens et al. (2004) 已经证明,除非国家间的贸易成本是两两对称的,否则在多国框架下动态定义不成立。原因在于一旦考虑到第三国效应(Third Country Effects),一国需求的增加将对应于一个小于 1:1 的产出增加,因为其他国家可能由于较好的准入性而吸引了一些企业。极端的,需求的增加甚至可能导致产业份额的下降(反向 HME)。因此,在多国框架下,只有当真实的产业分布根据准入性的影响进行调整之后才能观察到本地市场效应。而

在多国框架下,静态本地市场效应要求: $\frac{\lambda_1^*}{\theta_1} \geq \frac{\lambda_2^*}{\theta_2} \geq \dots \geq \frac{\lambda_M^*}{\theta_M}$ 。在这种情况下不存在国家需求份额的交叉现象(Leap-frogging),国家的需求份额根据经济规模来排序,小国拥有较少的垄断竞争部门企业,大国则拥有较多的垄断竞争部门企业。而且,静态市场效应并不受第三国市场准入性的影响,因此,在控制准入性后总能发现本地市场效应。

进一步的,Suedekum (2006) 认为, Behrens et al. (2004) 只是指出了在多国框架下动态本地市场效应被颠覆的理论可能性,但是,他们并没有归纳出什么样的第三国效应将消除本地市场效应。在一个三国模型中,Suedekum 证明,当且仅当本国需求份额额外生增加时,动态本地市场效应可以从两国推广至三国模型;如果外部发生一个需求向拥有较好准入性国家的转移,那么将对本国产业份额造成负面影响;而当两个外部国家具有同等的准入性时,外部需求转移并不影响本国的产业份额;对于一个给定的转向较好准入

性国家的需求,本地市场效应出现的条件是本国需求增加足够大。但当某个外部国家与本国存在一个同时的需求增加,只要这个外部国家具有更好的市场准入性,那么本地市场效应将消失。应该说,Suedekum (2006)的工作比较好地阐明了第三国效应影响本地市场效应的机制,是对 Behrens et al. (2004)的进一步补充和完善。当然,Suedekum (2006)也承认,在更多数量国家存在的情况下考察这一影响是一件更为复杂而艰巨的任务。

(七)本地市场效应的福利含义

由于本地市场效应是企业区位选择的重要因素,对于不同的国家有着不同的影响。因而许多研究已经开始关注本地市场效应的福利含义。目前,这类研究主要分为两类。

第一类研究着重于两国间的横向福利比较。Krugman (1980)指出,为使呈现本地市场效应的 IRS-MC 部门实现贸易平衡,就要求小国保持一个相对低的要素价格。这就意味着与大国的贸易自由化可能降低小国的工资。Davis (1998)则担心本地市场效应可能会使小国出现“去工业化”(De-industrialised)现象。Fujita et al. (1999)认为,在某种程度上,本地市场效应使工人在大国受益将出现一个移民的累积循环过程从而导致中心-外围模式。此外,Head, Mayer and Ries (2002)、Toulemonde (2005)等的研究也表明,似乎本地市场效应只会使大国受益,小国受损。然而,情况并非完全如此。Head and Ries (2001)运用国家产品差异化不变规模报酬模型,发现出现了逆向本地市场效应:小国可能成为净出口国。而且,贸易自由化将会促使小国制造业增加产出份额,原因在于小国企业增加了对大国市场的准入而获益。Chung (2002)通过将非同位偏好(Nonhomothetic Preferences)融入模型,考察在相对市场规模的条件下,富有的小国(单位资本收入高但 IRS-MC 产品市场绝对小)是否仅仅因为它的规模劣势而受损。如果答案是肯定的,贸易对小国不利,那么在推进贸易自由化和深度一体化的同时必须考虑某些协调机制。但他发现,与同位偏好假设下本地市场效应以小国的损失为代价不同,非同位偏好的本地市场效应模型提供了一个相互受益的情形。如果两国有不同的单位资本收入,需求模式由于非同位偏好而呈现出非对称,进一步的,生产的区位分布将反映这种非对称。另外,Hanson and Xiang (2004)、Yu (2005)等的研究也表明,所谓的“去工业化”现象并不一定会发生或者要弱得多。

第二类研究则主要关注本地市场效应的全球福利含义。这类研究以 Ottaviano (2001)为代表,主要考察本地市场效应引致的国际专业化的效率。Ottaviano 在 Krugman (1980)模型的基本框架下,构建了两要素、以标准化产品衡量的运输成本以及非对称的拟线性二次型效用函数模型,发现垄断竞争部门趋向于无效率地集聚于大国。而且,企业的市场势力越强、贸易成本越

低,集聚于大国的企业越多。然而,过度集聚可能取决于成本和需求函数的具体形式,但无效率却相对稳定。当垄断竞争企业选择区位时,它只考虑自己的利润而不充分考虑其区位决定对部门内其他企业和总消费者剩余的影响。因而最终的结果将是无效率的区位选择和缺乏效率的国际专业化。

总的来看,由于采用的模型不同,本地市场效应对不同规模国家的福利含义得出的结论也有很大的差异,因此还有待进一步深入考察;至于本地市场效应的全球福利含义在很大程度上也受制于模型的假设条件,而且这类研究目前还比较少,因此也不能轻易做出判断。

四、经验证据

Head and Mayer (2004) 指出,检验基于 Krugman (1980) 模型的本地市场效应必须面对有关模型和数据的三个方面的困难。第一,如何在检验本地市场效应的过程中分离要素禀赋 (H-O 模型) 对贸易模式的影响? 第二,如何处理多个部门之间的关系? 第三,如何测度多国框架下的需求份额? 此外,什么样的数据才能用来衡量需求差异?¹⁰ 这或许是直到 16 年后 (Davis and Weinstein, 1996) 才开始对 Krugman (1980) 的本地市场效应概念进行经验检验的原因所在。然而,近年来,理论模型的不断拓展和数据处理技术的日益完善丰富了经验检验的手段,许多实证检验文献开始陆续出现。这些文献主要从两个方向上检验本地市场效应,即检验经典的本地市场效应和检验修正的本地市场效应。

(一) 检验经典的本地市场效应

这一方向主要考察本地需求和供给 (生产) 之间的关系,并试图分离本地市场效应和要素禀赋对贸易模式的不同影响。主要文献有 Davis and Weinstein (1996, 1999)、Schumacher (2003) 等。

基于一个含有本地市场效应的报酬递增和要素禀赋的融合模型, Davis and Weinstein (1996, 1999) 对方程进行估计:

$$X_g^{nc} = \alpha_g^n + \beta_1 \text{SHARE}_g^{nc} + \beta_2 \text{IDIODEM}_g^{nc} + \Omega_g^n V^c + \varepsilon_g^{nc}, \quad (12)$$

式中, X_g^{nc} 衡量国家 c 产业 n 产品 g 的产出; SHARE_g^{nc} 是国家 c 产业 n 产品 g 相对于世界其他国家的产出水平; IDIODEM_g^{nc} 衡量超常需求, 其系数 β_2 将决定是否在本地市场效应; V^c 是要素禀赋向量。这些指标需要通过投入-产出

¹⁰ Head and Mayer (2004) 特别批评了一种显而易见的方法, 即所谓的“显性消费”或“国内吸收”方法。这种方法将国内需求等同于国内生产加进口再减去出口。然而, 这种方法要求贸易数据和生产数据具有可比性。但由于不同的机构采用不同的分类法来采集数据, 贸易数据和生产数据在各国范围内更缺乏可比性。

数据进行计算。根据 Davis and Weinstein (1996, 1999) 的理论模型, 当 $\beta_2 = 0$ 时, 表明贸易模式由比较优势决定; 当 $\beta_2 \in (0, 1]$ 时, 属于存在贸易成本的比较优势情形; 而当 $\beta_2 > 1$ 时, 本地市场效应出现, 它将决定贸易模式。运用该方程, Davis and Weinstein (1996) 首先检验了 OECD 国家的制造业生产结构, 结果并不支持本地市场效应。要素禀赋解释了 90% 的生产模式, 本地市场效应仅为 5%。紧接着, Davis and Weinstein (1999) 又运用该方程考察了日本的区域生产结构, 发现 19 个制造业部门中有 8 个支持本地市场效应。这说明, 相比国际生产结构而言, 本地市场效应对于解释一国区域层次的产业结构更为有效。为什么本地市场效应在区域层次上要强于国家层次上? Davis and Weinstein (1999) 认为原因可能有两个: 第一个是贸易成本, 无论是运输成本还是贸易壁垒, 区域之间肯定低于国家之间。低贸易成本意味着在相对小的市场上保护较少; 第二个在于区域间要素流动性要强于国家之间。较强的要素流动性将强化本地市场效应。

此外, 采用与 Davis and Weinstein (1996, 1999) 类似的方法, Domeque et al. (2005) 对西班牙 17 个地区 9 个制造业部门 1965—1995 的样本进行检验后也发现了本地市场效应; 张帆和潘佐红 (2006) 根据中国 1997 年 31 个省 (市、区) 19 个产业的投入-产出数据, 发现本地市场效应在决定中国区域间生产和贸易的类型上起着显著的作用。

Schumacher (2003) 认为现有文献无论是在理论上还是经验上只考虑了总收入 (需求) 而忽视了单位资本收入 (资本劳动比) 对本地市场效应的影响。因此, Schumacher (2003) 的目的在于考察这二者对比较优势的共同影响并将两者区别开来。他主要采用引力方程方法估计下列方程:

$$\ln X_{aij} - \ln X_{aji} = (\beta_1^a - \beta_3^a)(\ln Y_i - \ln Y_j) + (\beta_2^a - \beta_4^a)(\ln y_i - \ln y_j). \quad (13)$$

这意味着两国的出口/进口比率 $\frac{X_{aij}}{X_{aji}}$ 取决于两国的 GNP (总需求) 比率 $\frac{Y_i}{Y_j}$

和单位资本收入比率 $\frac{y_i}{y_j}$ 。进一步的, 考虑显示性比较优势指数:

$$\begin{aligned} & (\ln X_{aij} - \ln X_{aji}) - (\ln X_{oij} - \ln X_{oji}) \\ &= [(\beta_1^a - \beta_3^a) - (\beta_1^o - \beta_3^o)](\ln Y_i - \ln Y_j) \\ &+ [(\beta_2^a - \beta_4^a) - (\beta_2^o - \beta_4^o)](\ln y_i - \ln y_j). \end{aligned} \quad (14)$$

结合这两个方程, (1) 如果两国具有相同的单位资本收入 (要素禀赋相同), 那么贸易模式由相对规模决定, 当衡量出口收入弹性与进口收入弹性差值的 $\beta_1^a - \beta_3^a$ 为正时, 则存在本地市场效应; (2) 如果具有相同的规模, 那么贸易模式由要素禀赋决定, $\beta_2^a - \beta_4^a$ 衡量了传统比较优势的大小。

最后, Schwmacher (2003) 运用 22 个 OECD 国家 25 个三位数 ISIC 产业

的实证结果表明,许多制造业部门出现了本地市场效应,这些部门可能是资本密集型的也可能是劳动密集型的。这意味着,低收入国家源自相对要素禀赋的比较优势(在劳动密集型部门)可能会被高收入国家的较大经济规模抵消;另一方面,高收入国家在某些资本密集型产业的传统比较优势可能因为它们的较大经济规模而得以强化。Schumacher (2003)认为出口收入弹性大于进口收入弹性将会出现本地市场效应这一结论与 Melchior (1998)和 Feenstra et al. (2001)的结论类似,但 Schumacher (2003)同时认为他们在没有考虑传统比较优势的情况下检验本地市场效应,可能会扭曲结果从而使得他们的结论在某种程度上反映的是要素密集度。

(二) 检验修正的本地市场效应

理论模型的拓展丰富了本地市场效应的定义和内涵,相应的,一些文献在对经典的本地市场效应进行修正的基础上重新进行了实证检验。这些文献又至少包括三类:第一类文献以本土偏向效应(Home Biased Effects)重新定义本地市场效应,然后进行检验,如 Trionfetti (2001)和 Brühart and Trionfetti (2005);第二类文献在考察市场准入性影响的情况下进行检验,如 Davis and Weinstein (2003)、Hanson and Xiang (2004)和 Behrens et al. (2004);第三类文献则试图验证逆向本地市场效应和非线性本地市场效应,如 Head and Ries (2001)和 Crozet and Trionfetti (2006)。

1. 以本地偏向效应定义的本地市场效应

在 Helpman and Krugman (1985)模型的基础上,Trionfetti (2001)引入了“本地偏向的需求”(Home Biased Demand)¹¹这一新的概念作为识别两种贸易范式标准,指出本地偏向的需求份额与 IRS-MC 部门的产出正相关,与 CRS-PC 部门不相关,从而将本地市场效应定义为产出对于本地偏向的超常需求的不成比例的反应。这样,本地产出份额与本地超常需求和本地偏向的超常需求正相关。但 Trionfetti (2001)进一步证明,即便在本地偏向的超常需求存在的情况下,本地超常需求并不必然在 IRS-MC 部门引起本地市场效应,它有可能反映的只是贸易成本。Trionfetti (2001)对以下方程进行估计:

¹¹ Trionfetti (2001)列举了大量经验证据说明了本地偏向需求的普遍性。Davis and Weinstein (1996)提醒人们本地市场效应和本地偏向(home bias)这一概念有明显的区别。后一概念是指本地消费者对本地产品有着相对较强的需求偏好。在消费上的本地偏向往往只是因为贸易壁垒的结果。而且,本地偏向不含有任何有关市场结构或规模报酬的假设。相反,本地市场效应的出现是由于贸易成本和报酬递增互动的结果,它也是垄断竞争框架下的一个典型发展。Brühart and Trionfetti (2005)进一步指出,二者的共同点是都检验需求的地理分布与产出的地理分布之间的关系,但关键的差异在于本地偏向的需求从本地需求中分离出来且不受市场规模和贸易成本的影响,从而具有很强的稳定性,而本地市场效应则对这些很敏感。

$$\frac{n_{Si}}{N_s} = \beta_0 + \beta_1 \frac{E_{Si}}{\sum_i E_{Si}} + \beta_2 \frac{E_{hSi}}{\sum_i E_{hSi}} + \mu_{Si} , \tag{15}$$

式中，左边是本地产出比例， $\frac{E_{Si}}{\sum_i E_{Si}}$ 和 $\frac{E_{hSi}}{\sum_i E_{hSi}}$ 分别反映本地的超常需求和本地

偏向的超常需求。 β_1 和 β_2 这两个系数非常重要（见表 1）。根据 Trionfetti (2001)， β_2 将是主要的识别标准。如果 β_2 为正且统计上显著，那么说明该部门是 IRS-MC 部门，存在本地市场效应；而如果 β_2 为零或不显著地异于零，那么该部门将是 CRS-PC 部门。另一方面， β_1 不能作为主要的识别标准，它与贸易成本相关。如果它为正，说明该部门存在贸易成本；如果它为零，则该部门不存在贸易成本。最后，采用 8 个国家的投入-产出数据进行检验，发现 7 个部门与 IRS-MC 部门相关，存在本地市场效应；而另外 10 个部门则与 CRS-PC 部门相关。

Br hart and Trionfetti (2005) 采用了与 Trionfetti (2001) 相同的识别标准进行经验检验，即本地偏向的需求将影响 IRS-MC 部门的国际专业化，但不会影响 CRS-PC 部门的区位。但不同的是，Br hart and Trionfetti (2005) 在理论上进一步证明了本地偏向需求作为识别标准的稳定性¹²，而且在具体检验方法上他们分为两个步骤：首先采用引力模型估计本地偏向的需求，然后再运用类似于 Trionfetti (2001) 的估计方程对 6 个 OECD 国家 18 个制造业部门的 4 年数据进行面板估计。他们在 5 个部门发现了本地市场效应。

表 1 Trionfetti(2001)中的系数含义

β_1	贸易成本	β_2	理论范式
+	相关	+	IRS-MC
+	相关	0	CRS-PC
0	不相关	+	IRS-MC
0	不相关	0	CRS-PC

2. 考虑市场准入性影响下的本地市场效应

Davis and Weinstein (2003) 指出，Davis and Weinstein (1996) 对 OECD 国家的经验研究结论不支持本地市场效应的原因可能在于过分简化了模型中的地理因素。不同的国家存在不同的市场准入程度，这一因素显然没有被考虑到。因此，基于 Davis and Weinstein (1996, 1999)，他们首先通过引力方程估计出距离对需求的影响，从而将市场准入因素考虑进模型中，超常需求将由本地需求和邻国需求两部分构成。在经过这样的技术处理之后，在 OECD 国家重新找到了支持本地市场效应的证据。

¹² 倍重引力模型方法的好处在证明仅可运用引力方法性模型和贸易成本偏差还不完全能够论证资源性部还是资源禀赋本地偏向需求难以测度指标都是稳定的。

Hanson and Xiang (2004) 不满足于考察产出和需求间关系时可能产生的多重共线性以及使用引力模型估计总量数据可能丢失那些本可以在双边贸易中产生本地市场效应的产业变量的缺陷, 采用倍差引力模型方法 (Difference-in-difference)¹³ 检验本地市场效应。首先, 选择那些在出口市场上有着同等贸易壁垒的国家对; 其次, 选择两组产业, 一组属于高运输成本和强规模经济, 另一组属于低运输成本和弱规模经济; 最后, 考察是否像本地市场效应隐含的那样: 在每一国家对中, 大国倾向于出口相对高运输成本、强规模经济的商品。他们的估计方程如下:

$$\ln \left(\frac{S_{mjk}/S_{mhk}}{S_{ojk}/S_{ohk}} \right) = \alpha + \beta \ln \left(\frac{Y_j}{Y_h} \right) + \phi (X_j - X_h) + \theta \ln \left(\frac{d_{jk}}{d_{hk}} \right) + \varepsilon_{mojkh}, \quad (16)$$

式中, 左边是用倍差法衡量的贸易流量: 两国向共同的第三国市场出口的高运输成本、强规模经济产品与低运输成本、弱规模经济产品的比率的差值; 右边第二项是两国相对市场规模, 由两种方法进行衡量: 第一种方法仅考虑两国的 GDP, 第二种方法采用两国的市场潜力, 这就考虑到邻国市场的准入性。如果系数 β 大于零, 则存在本地市场效应。右边第三项考虑两国的生产成本差异, 这就能将本地市场效应与传统比较优势对贸易模式的影响分离开来。Hanson and Xiang (2004) 的经验结论是, 本地市场效应存在且依赖于产业的运输成本。对于高运输成本产业, 国家的市场规模决定国家出口; 对于中间运输成本产业, 邻国的市场规模非常重要。本国市场规模加邻国市场规模将决定国家出口。这说明, 对于这些产业, 出口生产也可能集中在小国, 前提是小国有较大的邻国从而增加了其产品的有效需求。因而, Hanson and Xiang (2004) 的结论也表明贸易成本下降导致的小国去工业化效应要弱得多。

Behrens et al. (2004) 在多国框架下将本地需求分解为相对国内市场规模 (吸引力) 和同其他外国市场的接近程度 (准入性), 即:

$$\lambda^* = \beta W \lambda^{\text{size}} + (1 - \beta) \lambda^{\text{hub}}. \quad (17)$$

λ^{size} 代表相对国内市场规模, λ^{hub} 衡量了准入性, β 体现了贸易成本, 显然贸易成本越小准入性对本地需求的影响越大。这意味着要想获得真实的本地需求信息需要对准入性进行过滤:

$$\lambda^{\text{size}} = (\beta W)^{-1} [\lambda^* - (1 - \beta) \lambda^{\text{hub}}]. \quad (18)$$

据此, Behrens et al. (2004) 对两组各 20 个国家 (一组 OECD 国家和一组发展中国家及新兴工业化国家) 25 个三位数 ISIC 部门分别进行了记号检验 (Sign Test, 基于本地市场效应的预期是否大于 1/2) 和排序检验 (Rank Test, 在控制准入性之后, 一国的相对产出份额是否与其需求份额一致), 发现了强烈支持本地市场效应的证据, 尤其是在 OECD 国家之间。

3. 检验逆向本地市场效应和非线性本地市场效应

Head and Ries (2001) 在国家产品差异化不变规模报酬模型中发现出现了逆向本地市场效应, 进而采用美加 1990—1995 之间 6 年双边三位数制造业贸易的面板模型证实了他们的结论。

Crozet and Trionfetti (2006) 的两部门模型预期所有部门产出和需求之间存在非线性关系。为了证实这一结论, 他们构造了以下估计方程:

$$S_s = \alpha_1 S_D + \alpha_2 S_D \cdot |S_D| \quad (19)$$

S_s 和 S_D 分别代表产出和需求, 系数 α_1 和 α_2 将分别反映是否存在本地市场效应和线性本地市场效应 (见表 2)。最后, 25 个国家、25 个部门 7 年的面板数据模型实证结果表明, 16 个部门强烈地支持了非线性本地市场效应。

表 2 Crozet and Trionfetti(2006)的系数含义

	非线性 HME	线性 HME
IRS-MC	$\alpha_1 \geq 1 \quad \alpha_2 > 0$	$\alpha_1 > 1 \quad \alpha_2 = 0$
CRS-PC	$\alpha_1 \leq 1 \quad \alpha_2 < 0$	$\alpha_1 < 1 \quad \alpha_2 = 0$

五、简要评论及启示

自从 Krugman (1980) 提出本地市场效应的概念之后, 由于它可以作为区别两种贸易理论范式的识别标准并具有显著的福利含义, 因而其理论模型在多个方向上进行了拓展, 经验证据也被越来越多的文献所提供。从理论的角度而言, 一方面, 许多理论模型都发现了本地市场效应, 这说明经典本地市场效应所赖以存在的基础假设并非不可或缺的; 另一方面, 许多模型也发现了逆向本地市场效应或反向本地市场效应, 这又说明本地市场效应的出现与否严重地依赖模型的假设条件本身。另外, 福利含义的研究表明, 本地市场效应对不同规模国家的影响也很难界定。与此同时, 经验研究的证据也是不确定的, 一些检验需要对本地市场效应的定义进行修正之后才能得到良好的估计结果, 而另一些研究则直接支持了逆向本地市场效应。而且, 选取的区域层级不同, 结果也可能截然相反。另外, 大多数经验研究的样本来自发达的工业化国家, 至今仍缺少对发展中国家的研究。因此, 无论是理论研究还是经验检验, 都有待进一步的深入。

长期以来, 中国的比较优势被解释为低廉的劳动力成本, 果真如此, 那么显而易见的是, 这种优势将难以持久。尽管从理论上来说, 中国的劳动力可以无限供给, 但近年来出现的“民工荒”现象已经开始凸现出中国劳动力供给的结构性不足。与此同时, 劳动力工资成本不断上升已经成为必然的趋势, 这在某种程度上将使得中国与东南亚其他国家相比的劳动力优势不复存在。而且, 空间经济学的新近研究表明, 在国际产业转移的背景下, 劳动力

比较优势可能成为一种向心力，使参与国际分工的国家仍能保住自身的劳动密集型产业，但这也同时意味着大量的产业部门将被锁定在价值链的低端，从而成为产业升级的障碍。因此，以单纯的劳动力比较优势参与国际分工与贸易，将不利于中国对外贸易的可持续发展。但如果某些产业存在本地市场效应，那将意味着中国比较优势的新的源泉，国内需求和规模经济非常重要。其政策涵义在于，政府正确识别产业区位选择的因素（要素禀赋还是本地需求）将最大限度地提高资源的配置效率。因此，在国内区域层次和双边甚至多边层次上检验中国产业部门是否存在本地市场效应就显得尤为重要。

改革开放以来，中国的产业逐步由中西部地区向东部沿海地区转移，由此使得东部沿海地区成为产业集聚中心。产业集聚、FDI 集聚和对外贸易集聚共同缔造了东部沿海地区三大集聚中心的地位，同时，这也是中国区域经济发展最为显著的空间特征。有关研究表明，中国地区消费差距指数最大为13倍，衡量收入分配不平等程度的基尼系数为0.44。¹⁴中国区域经济发展的不平衡已成为构建社会主义和谐社会的重要障碍。对于广大的中西部地区而言，由于缺乏产业集聚和FDI的集聚，再加上地理区位的劣势和过高的贸易成本，使其很难通过对外贸易来拉动地区的经济发展。因此，注重本地区的内部市场需求和国内其他区域的需求，逐步培育相关产业的本地市场效应，或许是实现经济发展的一个现实选择。而就国家推动中西部发展的战略举措，如西部大开发、振兴东北和中部崛起，除了在政策和资金上向中西部地区倾斜之外，旨在培育本地市场发展的政策以及消除区域壁垒的政策也不容忽视。

参 考 文 献

- [1] Baldwin, R., and R. Forslid, "Trade Liberalization with Heterogenous Firms", CEPR Discussion Papers 4635, 2004.
- [2] Baldwin, R., R. Forslid, P. Martin, G. Ottaviano, and F. Robert-Nicoud, *Economic Geography and Public Policy*. Princeton NJ: Princeton University Press, 2003.
- [3] Baldwin, R., and T. Okubo, "Heterogeneous Firms, Agglomeration and Economic Geography: Selection and Sorting", CEPR Discussion Papers 4602, 2004.
- [4] Baldwin, R., and F. Robert-Nicoud, "The Impact of Trade on Intra-industry Reallocations and Aggregate Industry Productivity: A Comment", NBER Working Papers 10718, 2004.
- [5] Behrens, K., A. Lamorgese, G. Ottaviano, and T. Tabuchi, "Testing the 'Home Market Effects' in a Multi-Country World: A Theory-Based Approach", CEPR Discussion Papers 4468, 2004.
- [6] Brander, J., "Intra-Industry Trade in Identical Commodities", *Journal of International Economics*, 1981, 11(1), 1—14.
- [7] Br  hart, M., and F. Trionfetti, "A Test of Trade Theories when Expenditure is Home Biased", CEPR Discussion Papers 5097, 2005.

¹⁴ 参见《中国经济时报》2006年9月4日：《中国奇迹PK日本奇迹》。

- [8] Chung , C. , “ Nonhomothetic Preferences and the Home Market Effect : Does Relative Market Size Matter ? ” [http : / / www. biz. uiowa. edu / econ / seminars / Spring03 / cchung. pdf](http://www.biz.uiowa.edu/econ/seminars/Spring03/cchung.pdf) , 2002.
- [9] Crozet , M. , and F. Trionfetti , “ Trade Costs and the Home Market Effect ” , [http : / / www. univ- cefi. fr / IMG / pdf / TradeCostHME_ Febuary7_1_. pdf](http://www.univ-cefi.fr/IMG/pdf/TradeCostHME_Febuary7_1_.pdf) , 2006.
- [10] Davis , D. , “ Intra- industry Trade : A Heckscher- Ohlin- Ricardo Approach ” , *Journal of International Economics* , 1995 , 39 (3 — 4) , 201 — 226.
- [11] Davis , D. , “ The Home Market , Trade and Industrial Structure ” , *American Economic Review* , 1998 , 88 (5) , 1264 — 1276.
- [12] Davis , D. , and D. Weinstein , “ Does Economic Geography Matter for International Specialisation ? ” NBER Working Papers 5706 , 1996.
- [13] Davis , D. , and D. Weinstein , “ Economic Geography and Regional Production Structure : An Empirical Investigation ” , *European Economic Review* , 1999 , 43 (2) , 379 — 407.
- [14] Davis , D. , and D. Weinstein , “ An Account of Global Factor Trade ” , *American Economic Review* , 2001 , 91 (5) , 1423 — 1453.
- [15] Davis , D. , and D. Weinstein , “ Market Access , Economic Geography and Comparative Advantage : An Empirical Test ” , *Journal of International Economics* , 2003 , 59 (1) , 1 — 23.
- [16] Dixit , A. , and J. Stiglitz , “ Monopolistic Competition and Optimum Product Diversity ” , *American Economic Review* , 1977 , 67 (3) , 297 — 308.
- [17] Domeque , N. , C. Fillat , and F. Sanz , “ The Home Market Effect in Spanish Industry : An Empirical Analysis ” , [http : / / www. etsg. org / ETSG2005 / papers / domeque. pdf](http://www.etsg.org/ETSG2005/papers/domeque.pdf) , 2005.
- [18] Falvey , R. , D. Greenaway , and Z. Yu , “ Intra- industry Trade between Asymmetric Countries with Heterogeneous Firms ” , University of Nottingham Research Paper No. 2004 / 05 , 2004.
- [19] Falvey , R. , and H. Kierzkowski , “ Product Quality , Intra- Industry Trade and (In) perfect Competition ” , in Kierzkowski , H. (ed.) , *Protection and Competition in International Trade*. Oxford University Press , 1987 , 143 — 161.
- [20] Feenstra , R. , A. Rose , and J. Markusen , “ Using the Gravity Equation to Differentiate Among Alternative Theories of Trade ” , *Canadian Journal of Economics* , 2001 , 34 (2) , 430 — 47.
- [21] Fujita , M. , P. Krugman , and A. Venables , *The Spatial Economy : Cities , Regions , and International Trade*. Cambridge , MA : MIT Press. 1999.
- [22] Greenaway , D. , and C. Milner , *The Economics of Intra- Industry Trade*. Oxford : Blackwell , 1986.
- [23] Hanson , G. , and X. Chong , “ The Home Market Effect and Bilateral Trade Patterns ” , *American Economic Review* , 2004 , 94 (4) , 1108 — 1129.
- [24] 何茵、李东、姚洋 , “ FDI 存在 ‘ 本土市场效应 ’ 吗 ? ” , 第六届中国经济学年会论文 2006。
- [25] Head , K. , and T. Mayer , “ The Empirics of Agglomeration and Trade ” , CEPR Working Paper 3985 , also in Henderson , J. and J- F. Thisse (eds) , *Handbook of Urban and Regional Economics*. Amsterdam : North Holland , 2004.
- [26] Head , K. , T. Mayer , and J. Ries , “ On the Pervasiveness of Home Market Effects ” , *Economica* , 2002 , 69 (275) , 371 — 390.
- [27] Head , K. , and J. Ries , “ Increasing Returns Versus National Product Differentiation as an Explanation for the Pattern of US- Canada Trade ” , *American Economic Review* , 2001 , 91 (4) , 858 — 876.
- [28] Helpman , E. , “ Imperfect Competition and International Trade : Evidence from 14 Industrial Countries ” , *Journal of the Japanese and International Economies* , 1987 , 1 (1) , 62 — 81.
- [29] Helpman , E. , and P. Krugman , *Market Structure and Foreign Trade*. Cambridge (Mass) : MIT Press , 1985.
- [30] Helpman , E. , M. Melitz , and S. Yeaple , “ Export versus FDI ” , NBER Working Papers 9439 , 2003.

- [31] Holmes, T., and J. Stevens, "Does Home Market Size Matter for the Pattern of Trade?", *Journal of International Economics*, 2005, 65(2), 489—505.
- [32] Justman, M., "The Effect of Local Demand on Industry Location", *Review of Economics and Statistics*, 1994, 76(4), 742—753.
- [33] Krugman, P., "Scale Economies, Product Differentiation, and the Pattern of Trade", *American Economic Review*, 1980, 70(5), 950—959.
- [34] Krugman, P., "The Hub Effect: or, Threeness in International Trade", in Ethier, W., E. Helpman, and J. Neary (eds.), *Theory, Policy and Dynamics in International Trade*. Cambridge: Cambridge University Press, 1993.
- [35] Krugman, P., "Fluctuations, Instability, and Agglomeration", NBER Working Papers 4616, 1994.
- [36] Krugman, P., and A. Venables, "How Robust Is the Home Market Effect?" Mimeo, MIT and LSE, 1999.
- [37] Larch, M., "The Home Market Effect in Models with Multinational Enterprises", http://www.ecomod.net/conferences/ecomod2004/ecomod2004_papers/128.pdf, 2003.
- [38] Leamer, E., and J. Levinsohn, "International Trade Theory: The Evidence", in Grossman, G. and K. Rogoff (eds.), *Handbook of International Economics*, 3. New York: Elsevier, 1995.
- [39] 梁琦,《产业集聚论》。北京:商务印书馆,2004年。
- [40] Markusen, J., and A. Venables, "Trade Policy with Increasing Returns and Imperfect Competition: Contradictory Results from Competing Assumptions", *Journal of International Economics*, 1988, 24(3—4), 299—316.
- [41] Medin, H., "Firms' Export Decisions: Fixed Trade Costs and the Size of the Export Market", NUI Working Paper No. 618, 2001.
- [42] Melchior, A., "Market Size Effects as an Explanation of the Gravity Relationship in International Trade", Paper to the Conference on International Trade and Market Structures, 2—3 April 1998 in Paris/Le Mans.
- [43] Melitz, M., "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity", *Econometrica*, 2003, 71(6), 1695—1725.
- [44] Okubo, T., and V. Rebeyrol, "Home Market Effect and Regulation Costs: Homogeneous and Heterogeneous Firm Trade Models", HEI Working Paper No. 02, 2006.
- [45] Ottaviano, G., "Home Market Effects and the (in) Efficiency of International Specialization", Mimeo, Bocconi University, 2001.
- [46] Ottaviano, G., T. Tabuchi, and J-F. Thisse, "Agglomeration and Trade Revisited", *International Economic Review*, 2002, 43(2), 409—436.
- [47] Scherer, F., and D. Ross, *Industrial Market Structure and Economic Performance*. Boston, MA: Houghton Mifflin Company, 1990.
- [48] Schumacher, D., "Home Market and Traditional Effects on Comparative Advantage in a Gravity Approach", DIW Discussion Paper 344, 2003.
- [49] Suedekum, J., "The Home Market Effect in a Three-Country Model", <http://www.uni-konstanz.de/suedekum/shadow.pdf>, 2006.
- [50] 藤田昌久、保罗·克鲁格曼、安东尼·J. 维纳布尔斯,《空间经济学:城市、区域与国际贸易》,梁琦主译。北京:中国人民大学出版社,2005年。
- [51] Toulemonde, E., "Does Production Fall in a Small Country When Trade Costs Decrease? Home Market Effect in the Face of Multinationals", <http://www.etsg.org/ETSC2005/papers/toulemonde.pdf>, 2005.
- [52] Trionfetti, F., "On the Home Market Effect: Theory and Empirical Evidence", <http://www.cepr.org/meets/wkcn/2/2290/Papers/trionfetti.pdf>, 1999.
- [53] Trionfetti, F., "Using Home-Biased Demand to Test for Trade Theories", *Weltwirtschaftliches Archiv*, 2001, 137, 404—426.

- [54] Tybout, J., "Internal Returns to Scale as a Source of Comparative Advantage : the Evidence", *American Economic Review*, 1993, 83(2), 440—444.
- [55] Weder, R., "Comparative Home-Market Advantage : An Empirical Analysis of British and American Exports", *Review of World Economics*, 2003, 139(2), 220—247.
- [56] Yu, Z., "Trade, Market Size, and Industrial Structure : the Home Market Effects Rvisited", *Canadian Journal of Economics*, 2005, 38(1), 255—272.
- [57] Zeng, D., and T. Kikuchi, "Home Market Effect and the Agricultural Sector", <http://www.rieb.kobe-u.ac.jp/coe/seminar/pdf/Zeng.pdf>, 2006.
- [58] 张帆、潘佐红, "本土市场效应及其对中国省间生产和贸易的影响", 《经济学(季刊)》2006年第5卷第2期, 第307—328页。

The Home Market Effect : Recent Developments in Theoretical and Empirical Research

XUEFENG QIAN

(Zhongnan University of Economics and Law)

QI LIANG

(Nanjing University)

Abstract The home market effect refers to the assessment that countries with larger home markets will be a net exporter in a world with increasing returns and trade costs. It arises as a dynamic research topic against the theoretic limitation of the neoclassic trade theory to explain complex trade facts and the difficulty of finding empirical evidence for the new trade theory. It combines the explanatory powers of the two theoretic paradigms and becomes one of the key elements of the new economic geography. After a review of the theoretic background and the basic theoretic model of the home market effect, this paper summaries and discusses the recent developments of its theoretic extensions and empirical research.

JEL Classification F10, O24, R10

“人以群分”： 非市场互动和群分效应的文献评论

陆 铭 张 爽*

摘 要 由于存在着非市场的互动, 市场的均衡结果是人们按不同的收入和社会地位而分层, 从而出现“群分效应”, 在现实中又突出地表现为居住区的分割, 这对于低收入阶层的收入、人力资本积累和公共品消费有着显著的负面影响。政府应在开展实证研究的基础上, 有针对性地采取一些经济和社会政策减少群分效应和居住区分割的负面影响, 促进经济发展和社会和谐。

关键词 非市场互动, 群分效应, 居住区分割, 社会经济学

“近朱者赤, 近墨者黑。”

——中国古语

“物以类聚, 人以群分。”

——中国古语

“长期以来, 对于社会相互影响在经济学里应构成什么样的领域是合适的, 经济学家们非常矛盾。”

——Charles F. Manski (2000)

一、引 言

在传统的经济学中, 价格理论是整个经济学理论的核心, 在价格的形成过程中, 不同的相关行为人通常都是独立地对市场信号作出反应, 人和人之间通过市场价格而间接地相互联系, 相应地, 仅对市场信号起反应的代表性行为成为经济学模型中的行动主体。相比之下, 非市场互动(non-market interaction)则长期处于经济学研究的边缘地带,¹ 尽管“人类学家和社会学家

* 陆铭, 复旦大学经济学系、就业与社会保障研究中心和中国社会主义市场经济研究中心; 张爽, 复旦大学经济学系和就业与社会保障研究中心。通讯作者及地址: 陆铭, 上海市邯郸路220号复旦大学经济学院200433; 电话: (021)65642064; E-mail: minglu73@263.net。本文在复旦大学报告过, 作者感谢陈钊、丁维莉、桂勇、纪月梅、李猛、李维森、王永钦、姚洋、张军、张晏等的讨论, 作者感谢复旦大学经济学院985工程项目、来自教育部的对全国优秀博士论文作者的研究资助和上海市“曙光学者”项目的研究资助。

¹ “互动”实际上是指行为人之间的相互作用, 翻译成“互动”是为了行文的简练。在许多文献里, non-market interaction 也被表述为 social interaction, 即社会互动。我们在这里使用非市场互动, 是为了强调这一机制与市场机制的区别。

反复地告诉经济学家文化、规范和社会结构的重要性,经济学家却没听,主要是因为这些其他学科没有发展出强有力的技术来分析行为所受到的社会影响”(Becker and Murphy, 2000)。这一状况在最近 20 年,特别是最近 10 年的经济学研究中正在发生显著的变化,经济学的理论和实证研究越来越重视非市场互动的形成机制及其影响。²关于非市场互动的研究使得人们对很多问题有了全新的认识,已经形成了“社会经济学”(Social Economics)这样一个前沿研究领域。社会经济学的兴起主要不是源于经济学对于“社会问题”的关注,而是因为经济学在其理论构建中引入了行为人之社会互动,极大地拓展了经济学的研究范式。在公共政策方面,小幅度的公共政策可以通过社会互动而放大其效应,这就是所谓的“社会乘数”(social multiplier)(Becker and Murphy, 2000)³,社会乘数的存在有助于提高公共政策的有效性。

诺贝尔奖得主 Gary S. Becker 和克拉克奖得主 Kevin M. Murphy 的著作《社会经济学:社会环境中的市场行为》被认为是一部社会经济学的代表性著作,广泛地涉及了社会经济学的各种问题,包括社会资本、习惯与社会互动、群分(sorting)与婚姻市场、社区的分割与融合、产品质量的升级、身份与不平等,以及时尚、规范和价值观的建模(Becker and Murphy, 2000)。需要说明的是,经济学中有关非市场作用的研究本身就是一个发展迅速且非常庞杂的领域,举例来说,博弈论中社会规范的演进、劳动经济学中家庭成员之间的互动、甚至宏观经济增长理论中人力资本在时间和空间维度上的溢出效应都与非市场互动有关。研究非市场互动的文献还与心理经济学和社会学紧密相联, Manski (2000) 对这些文献作过一个较为全面的学说史方面的梳理,讨论了与非市场互动有关的经济学和社会学的关系,以及在非市场互动的实证经济学研究中可能遇到的一些问题。Scheinkman (2006) 对关于非市场互动的理论和实证研究进行了简明而完整的综述。Zanella (2004) 是一篇非常好的有关非市场互动的经济学建模的综述,可以帮助读者很好地了解非市场互动的引入如何发展了经济学理论模型,对理论感兴趣的读者一定要看一下。与这些已有的文献所不同的是,本文将重点讨论基于非市场互动的群分效应(sorting)⁴的产生及其影响。我们将借助于一个一般均衡的理论框架将有关非市场互动对于教育、住房和劳动力市场的影响进行综合,由此揭示出群分效应是如何作为市场均衡的结果而产生的。通过这一综述,我们试图帮助读者了解国际上关于非市场互动和群分效应的研究的最新进展。鉴于国际上

² 有人将博弈论作为经济学家研究社会互动的开始,也许是因为经济学家不想承认自己长期以来忽视了人类行为中的社会互动。但是,本文所说的社会互动的含义要比博弈论里的互动要广泛得多。

³ Scheinkman (2006) 给出了这样一个“社会乘数”的定义:它是一个比率,其分子是一个参数的变化所引起的平均反应,分母是当人们忽略同伴行为变化时,一个参数变化引起的平均反应。

⁴ 在文献中,“sorting”是指在个人的理性选择之下,均衡的结果是类型相同(或接近)的人聚集在一起,而不同类型的人之间处于隔离状态。我们将其翻译为“群分”,是想借助于“物以类聚,人以群分”的俗语突出其基于理性选择的含义。

有关群分效应的研究的突破有不少是实证研究，因此，我们也对相关实证研究中遇到的主要问题进行了一些评论，并结合有关群分效应的研究给出了一些实例。

本文是这样安排的：在第二节里，我们将通过一个用文字表述的一般均衡理论框架对已有的有关非市场互动和群分效应的研究进行综合，在这个一般均衡理论里，教育市场、住房市场和劳动力市场的均衡将同时决定。其中，特别重要的是，我们将通过这一模型来展现群分效应是如何作为一种市场均衡结果而产生的。第三节是一个与计量经济学有关的部分。第四节将综述已有的实证研究所得到的结论，并对不同结论之间存在的联系和差异进行讨论。在此基础上，我们将针对现有的研究讨论此类研究的方向和有待解决的问题，以及已有的研究所具有的公共政策含义，并且将结合中国经济转型中出现的问题提出此类研究对于中国经济和社会发展的实际意义。第五节是全文的总结。

二、非市场互动和群分效应：一般均衡理论

（一）非市场互动的概念

非市场互动是指不同行为人之相互影响，而这种影响又不是直接在市场上通过价格的变化而产生的。举例来说，在最简单的消费品市场的供给需求均衡模型里，一个消费者收入提高了，就会增加这种商品的需求（假设商品是正常商品），于是就会抬高市场价格，而如果其他条件都不发生变化，价格上升就会导致其他消费者的需求量下降。这种消费者之间的相互影响是通过市场的价格机制来产生的，即使不同的消费者之间距离非常远，这种影响也可以借助于价格机制来产生。传统的经济学理论集中于研究价格机制，其中不同的行为人之相互影响完全通过市场来实现，这样一来，空间的概念便不重要。但是，事实上，空间是非常重要的，人与人的相互影响在很多时候都是在相同的空间下（例如同样的居住区、学校和工作单位）产生的，其发生过程也没有价格机制的直接作用，这就是经济学家们所关注的非市场互动。也正因为如此，非市场互动也被称为“社会规范（social norms）”、“邻居效应（neighborhood effect）”、“遵从效应（conformity）”、“模仿（imitation）”、“传染（contagion）”、“时尚（epidemics）”、“从众效应（bandwagons）”、“羊群效应（herd behavior）”（Manski, 2000）。⁵

非市场互动本身是一个非常宽泛的概念，我们不妨将其大致划分为三大

⁵ 社会学家们也对非市场的相互作用进行了大量的研究。Manski (2000) 认为，由于社会学中的概念庞杂，使得相关研究进展非常慢，也正因为如此，本文不讨论纯粹社会学的相关研究。

类。第一类非市场互动纯粹是通过偏好和心理的因素起作用的,一个行为人的决策除了基于他自己的偏好以外,还取决于他的决策与他的比较人群的决策的偏离度 (Akerlof, 1997; Akerlof and Kranton, 2000; Crane, 1991; Glaeser and Scheinkman, 2001)。第二类非市场互动与信息有关,一个人往往会模仿他周围人的行为,这往往是因为他缺乏足够的信息来独立做决策 (Bikhchandani, Hirshleifer and Welch, 1992; Bernheim, 1994; Jones, 1984; Sah, 1991)。⁶ 第三类非市场互动是人和人之间的相互作用直接影响到相关行为人的表现。例如,一个人的社会关系网络是非常重要的资源,个人的联系有助于找工作、得到建议和心理上的支持、甚至获得贷款。⁷ 更重要的是,人和人之间的互动会产生人力资本积累的外部性,从而表现出同群效应 (peer effect)。例如,在学校教育中,一个人的同学的平均水平越高,他就会受到某种正面的影响,从而获得更好的成绩。中国的古话中所谓“近朱者赤,近墨者黑”本质上说的就是这个道理。同群效应是人们在自己的生活经验中都能感受到的,但它究竟是不是存在,却是在实证研究中非常难以证明的,对这之中的难点和技术问题,我们将在下一节中加以讨论。

(二) 群分效应的一般均衡经济学

非市场互动会对人的行为产生一系列的影响,近年来有关非市场互动的一个重要的研究方向就是群分效应的产生。在接下来的讨论中,我们将群分效应的产生机制从简单到复杂区分为三个层面:第一个层面是,群分效应直接产生于人与人之间偏好的不同;在第二层面,我们引入对行为人的生产力有影响的“同群效应”,并且构建一个地方公共品提供的一般均衡理论分析框架,展开对于房地产市场、教育市场和劳动力市场的讨论,展现群分效应是如何作为市场均衡的一般结果而存在的;在第三个层面,我们进一步在地方公共品提供的一般均衡理论框架中引入公共品提供者之间的竞争。在这个理论分析中,我们将说明这样的命题:只要存在人与人之间偏好的不同,非市场互动就能够产生群分效应;而在引入房地产市场、教育市场和劳动力市场的一般均衡分析,特别是引入与群分效应相联系的人与人之间的相互影响(尤其是对生产力的影响)之后,群分效应将被加剧;最后,当存在地方公共

⁶ Miller et al. (1993) 以形式化的方式发展了相关的“风尚理论”(fashion theory) 并研究了消费的同群效应(peer effect)、求异效应(distinction effect)和渴望效应(aspiration effect)。消费的同群效应是指消费者的行为受社会身份相近的人群的影响,求异效应是人们想在消费上区别于社会身份低于自己的人,而渴望效应是指消费者想模仿社会身份高于自己的人的行为。他的分析还假定个体行为人的信息总是不足的,信息的传播是导致行动上的潮流效应的成因。

⁷ Granovetter (1995) 认为,工作往往是通过在找工作前很久就已经形成的社会关系而找到的。文献里(包括中文文献)有大量关于社会关系网络的研究,已经超出了本文的范围。Ioannides et al. (2004) 综述了大量有关社会关系网络对于人的找工作行为的影响,相当多的研究发现社会关系网络有利于提高个人找工作的质量和工作所获得的收入。

品供给者之间的相互竞争的时候，以地方公共品供给单位（如学校）数量和差异性度量的群分效应将进一步加剧。在展开详细的理论分析之前，需要说明的是，我们所说的“地方公共品”是指一定地理范围内⁸的公共产品，只有这个地理范围之内的居民才可以享受，而其他居民则无法享受。地方公共品的提供者是基层（地方）政府，这对应于英文中的“local government”的概念，这个概念并不包含对于政府级别和管辖范围的限定，而只是说这一级政府提供着只有当地居民才可以享受的地方公共产品。在后文中，我们将以教育作为地方公共品的实例，阐述教育消费中的群分效应的产生，以及在人力资本积累过程中的非市场互动。之所以以教育为例，是基于三点原因：第一，教育（特别是中小学教育）是最为有代表性的地方公共品，特别是在按居住地和学区入学的制度下，教育消费的地理范围限制非常明显。第二，教育是人力资本的积累过程，对受教育者在劳动力市场上的表现具有直接的影响。第三，教育支出在居民消费和政府支出中占有非常高的比重（在发达国家更是如此），因此，有关教育的经济学研究也正在成为国际经济学界的研究热点和前沿。

人与人之间的偏好差异是导致群分效应的最为简单而直接的原因。这方面原因往往与心理学和社会学方面的因素有关。例如，人们在消费中表现出的同群效应（peer effect）、求异效应（distinction effect）和渴望效应（aspiration effect）就属于纯粹心理方面的非市场互动（Miller et al., 1993）。⁹纯粹偏好（心理）的原因可能直接导致群分效应。举例来说，在美国，不同种族的人群（白人、黑人、亚裔等）往往是相对聚居在一起的，这其中的一部分原因就可能是由偏好导致的。在实证研究中，很难区分这种种族聚居的现象的来源，它可能来源于住房市场的歧视（例如最近美国对中国移民的歧视），也可能来源于对于邻居中种族构成的直接偏好（例如最近美国的中国移民喜欢和其他华人住在一起），还有可能来源于不可观察的社区质量中的一些有种族特点的方面（例如对于华人聚居区的中国商店的偏好）（Bayer, 2005）。但是，如果不考虑与群分效应相伴随的其他结果，那么，仅仅与偏好有关的群分效应就不会成为经济学研究的焦点问题。而经济学家们之所以关注群分效应，是因为群分效应对于住房市场、教育市场和人们在劳动力市场上的表现有着错综复杂的影响。在接下来的分析中，我们将一步步由浅入深地进行群

⁸ 根据分析对象的不同，我们所说的地理范围可以指一个省内、一个市内、一个区内，甚至一个街道、一个居住小区。

⁹ 在传统微观经济学里，Leibenstein(1950)早在1950年就提出由外部行为引起的消费需求可以分为从众效应（bandwagon effect）、虚荣效应（snob effect）和凡勃伦效应（Veblen effect）。从众效应（bandwagon effect）指个人消费随他人消费增加而增加，以显得时髦；虚荣效应（snob effect）指个人消费随他人消费增加而减少，以显得与众不同；凡勃伦效应（Veblen effect）指个人消费随着价格的增加而增加，以显示财富。前两种效应实际就反映了非市场互动对消费的影响。

分效应的产生及其影响的理论分析。首先,我们先在基于偏好的群分效应中考虑住房市场,将群分效应与居住区分割(residential segregation)联系起来,然后,我们考虑一个包含住房市场和教育市场的一般均衡模型,并引入教育中的同群效应,将群分效应的产生放在这个一般均衡的模型中加以讨论。最后,我们再引入地区公共品供给的竞争,将讨论地方政府之间和公共品提供单位(如学校)之间的竞争对于群分效应产生的影响,通过这样的讨论,我们将在一般均衡模型中纳入对地方公共财政的分析。

群分效应的一个直接影响就是对于住房市场、居住区分割和城市形态的影响。接下来,我们先不考虑地方公共品的提供,而只是在基于偏好的群分效应的产生机制中引入住房市场,我们将发现,群分效应将与收入联系起来,并且产生基于收入的居住区分割。不妨假设有两个完全相同的社区,而人们具有与同收入水平的人共同居住的心理和偏好。假设有一个随机的因素,导致某一个新来的富人选择社区 A 居住下来,他的进入将导致这个社区的房价上升,而房价的上升将使得边际上的低收入者选择搬离这个社区,搬进社区 B,同时,也将使边际上的高收入者在喜欢与富人住在一起的心理下选择搬进社区 A。于是,社区 A 的房价进一步上升,而社区 B 的房价则有所下降,如此反复的结果就是,最终社区 A 将成为高收入者聚居的“富人区”,而社区 B 则逐渐成为低收入者聚居的“贫民窟”。Epplé and Sieg (1999) 引入了住房市场,分析了居民对于居住地的选择,他们假定家庭的效用取决于家庭的偏好、公共品消费量、住房商品消费量和其他私人消费品数量。¹⁰ 他们证明,在均衡状态下住房市场和城市形态出现以下特征:(1) 边界无差异性,即处于不同社会之间的边界上的居民对于居住在两个相邻的社区是无差异的;(2) 分割特征,不同收入的家庭按其偏好处于严格的分割状态,不同社区的居民收入之间没有相互的重叠;¹¹ (3) 递增的消费品组合,一个社区比另一个社区房价高,并且公共品消费量多,当且仅当其居民的家庭收入更高。Bayer et al. (2004, 2005) 所讨论的住房市场均衡也是建立在居民对于邻居的偏好的基础上的,他们还用数值模拟的方式讨论了这样的住房市场均衡和居住区的分割对于不同人群的影响,对这些结果,我们将在后文中进行评介。

接下来,我们来考虑一个简单的包含住房和教育市场的一般均衡模型,并且在这个模型中引入对于行为人的生产力有直接影响的“同群效应”,例如在中学教育中,学生的成绩显著受到其同学的平均成绩的影响(Ding and Lehrer, 2005)。¹² 在这个模型中,考虑两个行为主体,一是消费者,二是以学校

¹⁰ 尽管他们的模型中也有公共产品,但在这个公共品的消费中没有考虑同群效应。

¹¹ 按收入水平表现出的群分效应是最为直接的,但在收入之外,群分效应也可能与职业和社会地位等因素有关。

¹² 教育中的同群效应被很多实证研究所证实,Ding and Lehrer (2005) 发现,在中国的高中教育中也显著地存在着非线性的同群效应。

为例的地方公共品供给单位。地方的公共产品只有在一定地理范围之内的居民才能消费。仍然假定消费者的效用取决于其扣除住房价格和教育支出后的净收入，以及家庭所获得的教育质量，而教育质量则取决于家庭的禀赋（例如孩子的能力）和学校学生的总体水平。不难理解，由于优质教育资源（如好的师资和生源）的稀缺性，不同的家庭将在市场上通过竞争的方式来获取优质教育资源。在不附加任何其他条件的情况下，收入较高的家庭将更有能力获得优质教育资源，如果学校的招生是以居住地为条件的话，那么，高收入家庭就会在好学校的周围聚居，而这将抬高好学校周围的房价，于是，边际上的低收入家庭将选择搬离好学校所在的社区，从而失去获得优质教育资源的机会。实际上，通过住房市场的运作，好学校的质量就在当地房价中被“资本化”了。Epplé and Romano（1998，2000）研究了教育市场的运作和学校教育的群分效应，他们证明，在均衡状态下，学校之间将形成严格的分等级状态，高收入家庭将更有能力获取优质教育资源。Rothstein（2006）构造了一个包含多社区的数学模型，模拟了 Tiebout 型群分效应的产生，在他的模型中，居民的效用取决于其扣除租金后的净收入和社区的质量，而社区的质量又是社区居民平均收入的增函数。尽管社区平均收入是通过什么样的机制影响到社区的质量的并没有在模型中加以讨论，但不难理解，如果收入反映了家庭的人力资本拥有量，而且同群效应存在的话，那么，高收入家庭的聚居将通过当地的学校教育中的同群效应提高社区的质量。他证明：在模型的任何均衡下都存在完全按收入进行的人群分隔，最高收入的一组家庭居住在高质量的社区，并且付最高的租金；收入第二高的一组家庭居住在质量和租金都第二高的社区；等等。社区之间按租金、质量和居民收入进行排序的结果是一致的。同时，他还证明，这种产生 Tiebout 型群分效应的均衡总是存在的。

不难理解，与上文所述的纯粹基于偏好的机制相比，在引入影响生产力的同群效应后，对于优质生源的同群效应较为偏好，且收入水平相对较高的家庭将选择搬入教育公共品质量较高的居住区，同群效应对这部分家庭来说是选择聚居的收益。因此，相比只是由于偏好原因导致的群分效应，在公共品提供模型中，学校教育质量将在当地的房价上被“资本化”，群分效应将被加剧，居民将形成按收入划分的聚居和相应的居住区的分隔。¹³

接下来，我们在上面的一般均衡分析框架中考虑地方政府间的竞争，并

¹³ 值得加以讨论的是，在 Rothstein（2006）的模型中，他采取的社区质量函数是家庭对同群效应的评价和学校治理绩效的线性可加函数，也就是说，同群效应和学校治理绩效在决定家庭对社区质量的评价时是相互替代的，在这一假定下，他的数值模拟结果显示，家庭对同群效应的评价提高将减少人群按社区质量进行的群分效应。之所以得到这个不符合直觉的结论，完全与作者对于函数形式的假设有关。而从经济学的理论来看，将同群效应视为学校治理绩效的替代是完全没有道理的。恰恰相反，同群效应应该被作为决定学校治理绩效的“投入品”。

讨论引入公共品供给方的竞争后对于群分效应的影响。在引入了地方间政府的竞争之后,上述这个一般均衡分析框架在本质上就是 Tiebout (1956) 提出的财政分权和公共品提供模型。在这个模型中,假设地方政府比中央政府具有信息优势,也就是说,地方政府比中央政府更加了解当地居民的偏好和公共品生产过程中的相关信息。将地方的公共品提供交给地方政府去做,而不是由中央政府直接来提供地方的公共品,实际上就是实行了财政分权体制,这一体制有三个优点。首先,它增加了地方政府提高公共品生产效率的激励,如果是中央财政来提供地方公共品,那么,地方政府完全可能通过夸大公共品生产成本的方式来增加中央对于该地方的财政投入。其次,地方财政提供公共品可以减少搭便车 (free rider) 行为,在中央财政体制下,各地方政府都会尽力少投入地方公共品,而等待中央补贴,这实际上就是地方政府之间相互搭便车的行为。最后,不同的人群对于地方公共品的偏好是不一样的,地方政府可以更为灵活地面对当地居民的需要,提高公共品的生产和配置效率。由于居民可以自由地选择居住地,实际上就形成了一种“用脚投票”的机制,这种机制形成了地方政府之间的竞争,如果一个地方的政府不提供优质的公共品服务,那么就会使得当地居民 (尤其是富裕的家庭) 搬走,这直接地影响到当地的房价和人力资源水平,即使在很短的时期里,也足以对地方政府的财政收入产生影响。在上述这个简单的模型当中,分散决策和地方政府间的竞争可以实现公共资源的有效配置,这就是 Tiebout (1956) 提出的财政分权和公共品提供模型的核心思想。在这个模型中,我们可以看到,由于存在消费者对于地方公共品的偏好的差异,以及地方政府之间的竞争,消费者可以通过选择居住地来实现偏好与地方公共品之间的匹配,同时也给地方公共品的供给者提供了提高公共品生产效率的激励。这个过程就是文献中通常所说的 Tiebout 群分 (Tiebout sorting) 的过程。在中国,一些较高收入家庭通过买房子住到好学校附近,或者通过买户口的方式迁移到教育较发达的城市或地区,为孩子获取优质教育资源和提高上大学的机会创造条件,都是 Tiebout 群分的现象。

在研究群分效应的文献中,还在另外一个意义上使用“竞争”这个词,即,将公共品提供单位的数量 (如学校和社区的数量) 增加作为供给者之间的竞争加剧。Rothstein (2006) 模拟了社区数量与群分效应之间的关系,结果发现,社区数量的增加的确会加剧人群在学校的治理绩效 (effectiveness) 方面的群分效应。在 Epple and Romano (1998) 的研究中,他们在教育市场竞争的模型中引入了对于教育券 (voucher) 的分析,引入教育券的作用是加强消费者对于学校的选择,其结果是,与不存在教育券的市场相比,存在教育券的均衡教育市场上,私有学校将增加,总体上的群分效应将更为明显。与此同时,Epple and Romano (1998) 证明,在一个同时存在公立学校和私立学校的教育市场上,如果学校间的竞争充分的话,至少存在着两种机制可以改

善教育的机会公平，是有利于减少群分效应对低收入家庭的负面影响的。第一，由于受教育者之间存在着同学相互影响的同群效应，私立学校有激励去竞争那些好学生，提高本校学生的平均质量，其竞争的结果就是学校将有激励实行价格歧视，一方面给予好学生奖学金，另一方面对高收入家庭的子女收取较高的学费，这就会大大提高低收入家庭子女获取优质教育资源的机会。第二，可以引入教育券来加强消费者的选择，这也能提高低收入家庭子女获取优质教育资源的机会。如果教育券在不同学生之间是没有差别的（flat-rate），那么，实行教育券的结果是使得低收入家庭的高能力子女获取优质教育资源的机会提高，成为受益最多的社会群体。但是，这也使得公立学校的好学生更多地流失，公立学校学生的平均质量下降，留在公立学校的学生所享有的同群效应也会下降。同时，一部分收入相对较高但孩子能力并不太强的家庭为了保持让他们的孩子能够仍然获取较高的同群效应，也会选择离开公立学校，但其进入私立学校的代价是必须支付更高的学费。Epple and Romano（2002）进一步讨论了其他可能的教育券机制对教育市场的影响，他们证明，如果将获取教育券的数额与学生的能力相联系，让高能力的学生获得更多的教育券，那么，低收入家庭的高能力子女将获益更多。在他们的数值模拟结果中，引入了教育券后有95%的人比没有教育券的时候获得了更高的福利。他们还进一步考虑了附加学费限制的情况，即要求学校只能将教育券作为学费，既不能收更高的学费，也不能少收费。这时，学校之间的竞争仍然存在，竞争的结果是私立学校将有激励招收任何类型的学生，学校间的差异将消失，在均一教育券之下部分学校的好学生流失的问题也就相应消失了。

上述一般均衡理论的分析框架所得出的结论是：（1）只要存在居民的偏好差异，就可能导致群分效应，如果在居民偏好差异的基础上引入住房市场和居民对于居住地的选择的话，群分效应的表现将是居住区的分割。（2）在引入地方公共品质量在住房市场上的“资本化”和公共品消费中的同群效应后，将加剧群分效应和居住区的分割。（3）进一步引入地方政府之间的竞争后，Tiebout 群分的出现将导致群分效应进一步加剧。而公共品供给单位之间的竞争和消费者选择的加大则可能产生两种结果，一方面，以公共品提供单位和差异性为度量的群分效应将加剧，另一方面，如果竞争的加剧是以消费者获得更大的选择权来实现的（例如以教育券的方式），那么，低收入家庭的好孩子将从中获益，群分效应对低收入家庭的好孩子的负面影响将减小。值得特别强调的是，在教育市场上引入教育券这样的机制，可以提高低收入家庭的好孩子获得优质教育资源的可能性，对增加全社会的教育产出也是有利的，这时，教育的公平与效率是可以兼顾的（丁维莉、陆铭，2005；陆铭、蒋仕卿，2006）。

三、非市场互动和群分效应的计量经济学

上一节中我们通过一个一般均衡理论框架讨论了群分效应的产生,这一讨论仍然是理论层面的。而非市场互动和群分效应之所以重要,除了相关研究在理论上突破在传统经济学以价格机制为核心的研究范式以外,经济学家还特别关心与非市场互动和群分效应有关的实证研究,这些实证研究为在经济学理论中引入非市场互动提供了基础。Jones (2002)在对 Becker 和 Murphy 所著《社会经济学:社会环境中的市场行为》的书评里提到,此书提出的包含社会互动的理论框架还需要进一步的实证研究加以证实。尽管用计量模型来识别社会互动仍然存在技术上的困难,但是,对于长期以来怀疑社会互动的重要性的人们来说,也只有规范的实证检验才能够说服这些怀疑的声音。同时,实证研究也对教育和劳动力市场等方面的公共政策制定有着重大的影响。这一节中,我们先从概念上区分三类不同的非市场互动,然后从方法上讨论有关的计量经济学问题,对于实证研究讨论的核心问题及相关发现我们将在下一节中总结讨论。对于非市场互动的计量经济学的一个更为技术性的研究,可以参阅 Brock and Durlauf (2000)。

(一) 识别非市场互动的计量模型

1. 提出计量模型

Manski (1993) 提出,识别非市场互动的计量模型如下:

$$y = \alpha + \beta E[y | x] + E[z | x] \gamma + z' \eta + x' \delta + \varepsilon. \quad (1)$$

在(1)式中, y 代表个人的产出(例如学生的学习成绩); $z \in R^K$,代表可观察的个人特征, K 表示有 K 种可观察的个人特征。 $x \in R^J$,代表个人所属的组群的特征, J 代表有 J 种可观察的组群特征。上标表示 K 种个人特征或 J 种组群特征的列向量的转置。个人的产出 y 受到四类可观察变量的影响:个人的特征 z ,同组成员的平均产出 $E[y|x]$ 、同组成员的平均特征 $E[z|x]$ 、所属群体的特征 x 。 ε 代表不可观察的个人特征(如个人能力)或组群特征。

2. 三种不同的非市场互动

Manski (1993) 指出,上式中存在三种不同的市场互动。

(1) 内生的互动(endogenous interaction): 行为人的行为随着他处的人的行为而变化,在(1)式中表现为 y 随 $E[y|x]$ 的变化而变化, β 就度量了内生的互动的效应。这实际上是用同组的其他成员的表现解释一个特定成员的表现,强调成员相互之间的影响和互动。大多数实证研究都采用同组成员的平均表现对个人产出的影响来识别同群效应。

(2) 情境的互动 (contextual interaction): 情境的互动是指一个人的行为与他所处的群体的外生特征有关。在 (1) 式中表现为 y 随 $E[z|x]$ 的变化而变化, γ 就度量了情境的互动的效应。这实际上是用组的成员的共同特征来解释一个人的行为。

(3) 关联效应 (correlated effects): 同组的行为人行为一致的原因在于组的成员拥有相似的个人特征或者面临着相似的制度环境。在 (1) 式中表现为 y 随 z 的变化而变化。 δ 就度量了关联效应。

在以上三种非市场互动中, 内生的互动和情境的互动区分了社会成员被他们所处的社会环境所影响的不同机制, 而关联效应则指的是非社会性的现象。Manski (2000) 举了教育的例子来说明这三个概念的区别。如果一个学生的成绩与他所处的中学、宗教组别或者其他参照组的平均成绩相关, 那么就发生了内生的互动。在教育经济学中所研究的同群效应就是一种内生的互动。如果学生的成绩是与他所处的组别的社会经济构成 (如性别、宗教、收入、教育等) 相关的, 那么, 就存在着情境的互动。而关联效应指的是指学生的成绩相近, 是因为他们拥有共同的教师或者因为他们有相似的家庭背景。

对上述三种非市场互动的区分具有重要的公共政策含义。例如, 当为部分学生提供额外的补习的时候, 如果学生之间存在内生的互动, 那么, 补习将首先提高接受补习的那些学生的成绩, 然后又会通过内生的互动再影响到其他没有接受补习的学生的成绩, 这时, “社会乘数” 效应就出现了, 而情境的互动和关联效应则不存在这种反馈机制。也正因为如此, 经济学家们往往特别关注内生的互动, 相比之下, 社会学家们则相对更为关注情境的互动。

在实证研究中, 如何将内生的互动与其他两种非市场互动区别开来, 即如何在 (1) 式中有用地估计 β , 仍然是具有挑战性的问题。

将 (1) 式写为 (2) 式:

$$E[y | x, z] = \alpha + \beta E[y | x] + E[z | x]' \gamma + z' \eta + x' \delta. \quad (2)$$

接下来, 在 (2) 式两边都对 z 加总, 实际就是对两边在个人层面进行加总, 设组内的人数为 n , 得到在组群层面的 (3) 式:

$$\sum_{i=1}^n E[y | x, z] = n\alpha + n\beta E[y | x] + \gamma \sum_{i=1}^n E[z | x] + \eta \sum_{i=1}^n z' + nx' \delta, \quad (3)$$

(3) 式再写为 (4) 式 (注意, 在方程左边进行组群层面的加总时, 个人特征就不起作用了):

$$nE[y | x] = n\alpha + n\beta E[y | x] + nE[z | x]' \gamma + nE[z | x]' \eta + nx' \delta, \quad (4)$$

对 (4) 式两边化简, 就得到 (5) 式, 这时就形成了社会均衡条件 (Social

Equilibrium Condition)¹⁴。对它的直观解释是：通过对个人层面的加总，由此得到的组内成员的平均产出不再受到个人特征的影响。

$$E[y|x] = \alpha + \beta E[y|x] + E[z|x]\gamma + E[z|x]\eta + x'\delta. \quad (5)$$

当 $\beta \neq 1$ 时：(5) 式的唯一解可写为 (6) 式：

$$E[y|x] = \frac{\alpha}{1-\beta} + E[z|x] \frac{\gamma + \eta}{1-\beta} + x' \frac{\delta}{1-\beta}. \quad (6)$$

由 (6) 式可见， $E[y|x]$ 是 $[1, E[z|x], x]$ 的线形组合，即组内成员的平均产出受到组内成员的平均特征和组群特征的影响。在这种情况下，很难直接估计出 β ，并且把内生互动与情境互动、关联效应区别开。

最后，将 (6) 式代入 (2) 式，就能够得到简约方程 (7) 式：

$$E[y|x,z] = \frac{\alpha}{1-\beta} + E[z|x] \frac{\gamma + \beta\eta}{1-\beta} + z'\eta + x' \frac{\delta}{1-\beta}. \quad (7)$$

当 $1, E[z|x], x, z$ 之间不存在线性相关关系时，就能够识别出 (7) 式中的 $\frac{\alpha}{1-\beta}$ 、 $\frac{\gamma + \beta\eta}{1-\beta}$ 、 η 和 $\frac{\delta}{1-\beta}$ 。值得注意的是，在这个简约式里，仍然无法有效地识别出根据 (1) 或 (2) 式所定义的内生的互动、情境的互动和关联效应这三种不同的社会互动效应，因此，要识别出这几种效应，特别是内生的互动效应，还必须依赖于我们将在 (三) 里讨论的那些手段来对 (1) 式进行估计。在那之前，我们还是先来看一下直接对 (1) 式进行估计可能存在的问题。

(二) 非市场互动的实证研究中的估计偏误

在非市场互动问题的实证研究中，经济学家们通常使用的是微观数据，而不是加总的宏观数据，而在微观数据分析中，最容易存在的问题就是各种各样的估计偏误。我们在 (一) 中分析的基础上，来讨论几种常见的计量偏误，并结合我们所关心的问题来举一些实例。

1. 联立性问题 (simultaneity problem) 或影像问题 (reflection problem)

联立性问题是指出于解释变量和被解释变量存在相互影响而导致的估计偏误。在研究非市场的相互作用的过程中，这种联立性问题往往具体表现为影像问题 (Manski, 1993)。也就是说，被解释变量和解释变量就好像是一个人和他在镜子里的影子一样是同时运动的，我们难以判断这种被观察到的相关性是因为影子的运动“导致”还是“反映”了照镜子的人的运动。在 (2) 式里，影像问题就表现为： $E[y|x,z]$ 会受到 $E[y|x]$ 的影响；而另一方面上，

¹⁴ 从数学公式上看，社会均衡条件就是当 $\beta \neq 1$ 时，组内成员的平均产出 $E[y|x]$ 的解是唯一的。

$E[y|x]$ 也会影响 $E[y|x,z]$ 。在这种情况下,会导致对 $E[y|x]$ 的系数 β 的估计偏误。

举例来说,在研究教育的同群效应的时候,一个班级的平均质量实际上也受到被解释的某个学生的质量的影响,这就好比班级的质量是学生的质量的影子,两者是同时运动的,不能简单地认为是班级的质量决定了某个学生的质量。Manski (2000) 简要地总结了一些之前的文献中(包括 Brock and Durlauf, 2000; Manski, 1993, 1997; Moffit, 1999) 提出的一些克服影像问题来识别非市场互动的�影响的方法,其中包括:(1) 如果研究者事先知道组的平均值对于被解释的个体值的作用的滞后期的话,可以使用滞后的组平均值而不是即期值作为解释变量;(2) 如果事先知道组的平均值与被解释变量之间的非线性的关系的话,那么可以使用组平均值的非线性形式作为解释变量;(3) 使用组的其他特征(例如中间值)而不是均值来作为解释变量,这要求研究者得事先对组的行为模式有相应的了解;(4) 如果可以找到外生的工具变量,并且这个变量直接影响一部分组的成员的行为变量,而不影响所有成员的结果的话,那么,这个工具变量也可用来识别非市场互动的影响。

2. 选择性偏误 (selection bias)

读者不妨回忆一下,如果我们希望利用普通最小二乘法 (OLS) 得到一个方程的无偏估计,那就需要要求模型的解释变量与模型的残差不相关。如果模型的解释变量是在某些未观察的因素的作用下通过行为人的选择而决定的,那么模型估计的结果就可能产生偏误。将 (1) 式写为 (8) 式:

$$y = \alpha + \beta E[y|x] + E[z|x]\gamma + z'\eta + x'\delta + \mu + \phi. \quad (8)$$

(1) 式中的未观察因素 ε 包含两个部分, μ 和 ϕ 。 μ (可能是个人特征或组群特征) 会同时影响个人选择进入这个组群和个人的产出 y ; 而 ϕ 为其他不影响个人对组群选择的未观察因素。当存在 μ 这样的未观测因素时,由选择带来的遗漏变量偏误就产生了。这意味着由 μ 产生的对 y 的影响被包括在了与 μ 相关的解释变量的系数里。

例如,在教育同群效应的研究中,通常需要用班级的平均成绩来代表同学的质量,同群效应是不是显著存在就取决于一个学生的成绩是不是显著地与其他同学的平均成绩正相关。但是,作为度量一个学生群体的平均质量的指标并不是随机产生的,而通常是由家长的偏好决定的。正如我们在上一节中所说的那样,在教育仅服务于一个特定地理范围内的家庭的市场上(按地段或学区入学),家长将选择搬迁到好学校的附近,以使得自己的孩子获得好学校的教育资源的可能性。也就是说,一个孩子是不是在一个平均水平较高的环境里学习本身可能是家长的选择结果 (Ding and Lehrer, 2005)。同样道理,如果我们想研究社区的分割对于一个人的人力资本积累和劳动力市场表现的影响,那么,一些度量社区质量的指标,如居住区内的平均收入

水平、平均受教育水平、人口构成等,也是被研究对象进行选择后得到的结果。Cutler and Glaeser (1997)指出,居住区的分割程度也可能是被选择的,最为成功的人可能会选择居住在社区分割程度较低的城市,而最不成功的人则可能选择居住在社区分割程度较高的城市。

选择性偏误本质上是一种特殊的遗漏变量的偏误 (omitted variable bias)。遗漏变量是指在计量模型中没有放入可能对被解释变量产生影响的变量,如果这些遗漏变量与计量模型的解释变量之间是不相关 (正交) 的,那么,对于那些被包括在计量方程中的变量系数来说,遗漏变量并不会造成模型估计的偏误。但是,在实际的计量估计中,遗漏变量往往与模型的一个或几个变量相关,这时就会存在缺失变量的偏误。¹⁵例如在教育生产函数的估计中,通常认为教师的质量是教育绩效的重要因素,但在很多教育生产函数的估计中,由于数据的局限,往往只能控制教师的受教育年限和教学经验,Ding and Lehrer (2005)的研究中使用了来自中国的数据,由于他们的数据中有基于教师教学质量的评估和晋升数据,因此可以在学生成绩的决定方程中更为完整地控制教师质量的指标。另一个例子是,在居住区的选择模型中,价格是影响居民是否选择居住在一个地区的影响因素,价格越高,居民选择居住在这个地区的可能性越低。但是,在估计这个居住区的选择模型的时候,价格可能反映了一些未能观察到的住房特征,高价格的住房可能具有某种未能在数据中被包括的优点,如果是这样的话,就可能造成在存在遗漏变量的情况下使价格对于居住区选择的影响偏向于零。同样道理,一个城市的居住区分割程度可能反映了一些缺失的城市特征 (Cutler and Glaeser, 1997); 作为居住区选择的解释变量的社会人口特征变量 (包括邻居的种族、收入和教育水平、学校的质量) 也往往与未能观察到的住房和周围环境的质量有关,这也会导致估计的偏误 (Bayer et al., 2005)。

3. 漏损偏误 (attrition bias)

在有关非市场互动的研究中,研究者往往希望通过动态的跨期的数据来看非市场互动的影 响。尽管最初的观察样本是随机的,但在动态的过程中,却可能出现样本的漏损,如果这个漏损不是随机产生的 (通常都不是随机的),那么,如果直接对数据进行分析,就可能产生漏损偏误,这实际上是一种特殊的选择性偏误。举例来说,在教育经济学中,一个争论热点话题就是小规模 的班级是不是有利于提高教学效果。在美国田纳西州的小班实验的数据中就存在类似的情况,除了漏损偏误,这个 小班数据还存在退出和替换所造成的偏误,每年大约有 10% 的人变换了班级类型,在三年级之前,超过一半在 幼儿园时期参加项目的学生在三年级之前离开了项目。这些样本的变化

¹⁵ 很多计量经济学的教科书对此都有非常清楚的解释,Davidson and Mackinnon (2004) 的解释比较形象。

完全可能不是随机产生的，如果不对这些样本变化加以考虑，就可能导致估计的偏误，对漏损偏误进行了考虑之后，小班教学并不对学生成绩产生显著影响，这与之前的研究得到的发现形成了鲜明的对比（Ding and Lehrer, 2004）。

（三）克服与非市场互动有关的估计偏误的实证方法

对于上一小节中讨论的实证研究中可能出现的偏误，读者不难发现，尽管其产生的原因有着细微的差别，但从统计学的角度来说，这些偏误产生的原因本质上都是源于在计量模型中未能正确地考虑一些可能与解释变量相关的因素（包括未观察到的因素）的影响。下面我们讨论已有的实证研究是如何克服偏误的。

1. 自然实验（Natural Experiment or Randomized Experiment）

我们提到，导致非市场互动难以被识别的一个重要原因在于，人们进入这个组群可能是有选择的。因此，有可能存在一些未观察到的因素会影响人们对于组群的选择和组群内的非市场互动，造成估计上的偏误。Hoxby（2000）和 Winston（1999）提出，家庭对于居住地和学校的选择在很大程度上取决于他们对学校内同群效应的预计。在这种情况下，如果能够采用自然实验对于人群进行随机分组，将有效地克服由选择带来的偏误。因为，人们如何进入该组群是外生的，不会受到遗漏的个人或组群特征的影响。Katz et al.（2001）以及 Ludwig et al.（2001）研究了非市场互动对于青少年犯罪的影响，他们都使用了基于美国政府在1994年开始在四个大城市实施的住房流动计划（Moving To Opportunity，以下简称MTO计划）的数据。该计划针对贫困率较高地区的居民，将他们随机地分为三组：实验组，比较组和控制组。实验组能够得到在私有住房市场上的住房券，该住房券只能用于购买低贫困率地区的住房；比较组也能够获得住房券，但是没有任何限制；控制组没有任何住房券。在这样的自然实验设计下，人们进入某个居住地很少地受到个人选择的影响，能较好地克服内生性问题，由此得到的结论也比较有说服力。基于该住房流动计划的一系列实证研究有非常重要的公共政策含义，我们将在下一节中进行讨论。Sacerdote（2001）在对于同学之间的同群效应的研究中也采取了自然实验的方法。Dartmouth College的新生入学时，被随机地安排所在的寝室和室友，这样就能够避免可能与个人特征相联系的对于室友的选择。该研究发现，由此产生的同群效应对于学生的学习成绩和社团参与都有重要影响。当然，实行完全随机的自然实验有非常大的难度。因此，一些实证研究也采取了准随机的实验（Quasi-randomized experiment），但是，准随机实验的外生性也是需要仔细推敲的。例如，Rosenbaum（1991）对于同群效应的研究采用了基于芝加哥政府实行的Gautreaux计划的数据，4000户低收入的黑人家庭自愿进入该计划，政府给予他们住房补贴，鼓励他们搬到白人居住的地

区。但是,准随机实验与随机实验一个重要的区别在于:在准随机实验中,人们进入该实验是有选择的,那么由选择带来的偏误仍然难以完全避免。因此,Duncan (1991)指出,进入 Gautreaux 计划的黑人家庭也是有选择的,在此样本上得到的结论可能是有偏的,而且,该结论是否对于所有低收入家庭都具有普遍性也值得怀疑。

2. 工具变量法

对于可能存在的偏误来说,一个有效的克服方法就是找到解释变量的有效的工具变量。一个有效的工具变量应该在统计上符合两个基本的条件:第一,当模型的所有变量(包括内生变量)被控制以后,这个工具变量将不对被解释变量产生影响;第二,当模型的其他解释变量被控制后,模型中的内生变量应该能够足够显著地被工具变量解释。如果这两个条件得到满足,说明工具变量本身不直接影响被解释变量(条件一),但它通过影响“被工具的”内生变量而对被解释变量有间接影响(条件二)。计量经济学可以估计出工具变量如何影响“被工具的”内生变量,然后在此基础上估计“被工具的”内生变量如何再影响被解释变量,这样一来,就可以将解释变量的内生性造成的估计偏误给“剥离”出去。刚才我们所说的是工具变量在统计上应该满足的性质,但是,说起来容易,做起来难,如果要找到有效的工具变量并不是一件容易的事,因为在上述两个条件中,第二个是可以直接检验的,而第一个条件是无法被严格检验的,因此,不同的研究者往往对于一个工具变量是不是严格外生的持有不同的看法。¹⁶从近十年的研究来看,历史和地理的因素经常成为经济学家们选择工具变量的考虑对象,因为,如果一个历史或地理的变量通过统计上的检验的话,那么通常人们能够接受这样的变量是严格外生的。¹⁷

在 Manski (1993) 提出对三种非市场互动的区别之前,对于非市场互动的工具变量的选择还存在误区。Case and Katz (1991) 用组群的外生特征来作为组群产出的工具变量,发现同群效应对于青少年犯罪有显著影响。但是,我们从 Manski (1993) 的研究中得知,组群的外生特征本身也会影响个人的表现,是一种情境的互动。因此,这一工具变量的使用会导致估计的偏误。

接下来,我们举几个实例来讨论在关于非市场互动的实证研究中,什么样的工具变量是好的工具变量。在有关非市场互动和群分效应的实证研究中,诸如社区的质量、学校之间的竞争度之类的变量都被认为是内生的,这在上

¹⁶ 在最近这些年里,实证研究中一个非常有趣的现象就是,如果一项研究找到了令人信服的“聪明”而“巧妙”的工具变量,那么这样的研究发表在好杂志的可能性就大大提高。但是,即使是在顶级杂志上发表的文章,也仍然可能在工具变量的选择方面被人批评。

¹⁷ 在对于制度的工具变量选择中,经济学家们也非常依赖于历史和地理的变量(Acemoglu et al., 2001),而这些工具变量是不是有效恰恰是争论的焦点所在(Glaeser et al., 2004),因为在理论上很难排除工具变量还可能通过影响除了“被工具变量”之外的其他变量而间接地影响被解释变量,甚至在理论上可能直接影响被解释变量。

面的讨论中已经加以说明了。Hoxby (2000) 研究了公立学校之间的竞争是不是对提高学校的生产力有利的问题。在她的研究中, 她用一定地理面积内的学区数量来度量公立学校之间的竞争和 Tiebout 选择的程度, 为了避免这一变量的内生性导致的偏误, 她使用相应地区内河流的数量来作为工具变量。她的思想是, 河流的数量是城市里“自然的”学区边界。在 18 世纪和 19 世纪的时候, 当学区的边界在美国被确定的时候, 一个非常重要的考虑因素就是学生们去上学需要花费的时间, 而自然的障碍显然显著地影响了学生上学的时间。尽管随着交通运输和基础设施等各方面条件的改善, 自然的障碍对于学生上学时间的影响已经大大下降了, 但是, 由于自然障碍对最初的学区边界有着显著的影响, 而最初的学区边界又对今天的学区边界有着关键的影响, 所以, 可以用自然状况来作为学区数量的工具变量。Hoxby 的这篇文章虽然正式发表于 2000 年, 但其在 NBER 的工作论文实际上发表于 1994 年。在文章问世以后, 她使用的工具变量被很多人奉为经典, Cutler 和 Glaeser (1997) 沿用了河流数量作为居住区分割程度的工具变量。¹⁸

上面我们纯粹是从统计学的角度来讨论工具变量的有效性的, 这是不是就是我们在选择工具变量时需要考虑的全部问题呢? 不是的。如果在经济学理论上工具变量应该与内生变量有正相关关系, 但在检验的时候却发现两者之间存在着难以解释的负相关关系, 那也不行。此外, 对于两个同样通过统计检验的工具变量来说, 仍然可能存在超过统计意义上的, 在经济学意义上的“好”与“坏”的差别的。Hoxby (2000) 所使用的工具变量尽管可以被接受为有效的工具变量, 但是, 地理的自然界限与学校之间的竞争度之间的关系并没有直接的经济含义。更进一步地说, 由于工具变量的有效性本身表明工具变量对于被解释变量有着间接的影响, 所以, 如果我们能从这一逻辑中推演出一些经济变量之间的关系, 甚至一些政策变量对于经济系统的间接影响的话, 那就更好了。Hoxby (2000) 所使用的工具变量虽然已经是非常好了, 但仍然不是一个“足够好”的工具变量。

与此相对照的是 Urquiola (2005) 的工作, 他的研究是检验择校行为是不是导致了群分效应, 这也需要寻找一个度量学校可供选择的程度的指标。在以前的文献中, 学校的可供选择程度的指标通常就是学区的数量或者学校的数量, 文献里对于是学区间的竞争还是学区内部学校间的竞争导致群分效应是存在争论的 (Clotfelter, 1999; Alesina, Baqir and Hoxby, 2004; Hoxby, 2000)。Urquiola (2005) 的研究策略是, 他使用了一个固定效应模型, 并且将观察变量分为两组, 控制学区的哑变量。在这个模型中, 他用了一个地区

¹⁸ 但是, 对于 Hoxby (2000) 的研究也有一些批评, Rothstein (2004) 对 Hoxby (2000) 中工具变量的使用方法等一系列问题提出了质疑, 而 Hoxby (2005) 则对其批评进行了非常严厉的反驳。

内部的中小学校数量差异来识别学区可选择程度对群分效应的影响。其中的原理是,一个地区内的不同学区有不同的中小学校数量,中学通常具有较高的固定成本,因而规模经济性比较强,而小学的固定成本低,没有很强的规模经济性。当入学率上升时,由于中学有更强的规模经济性,通常小学的数量增加更多,因此,一个地区内的中小学之间的数量差异就反映了该地区的学区的可选择程度。正如作者自己指出的那样,虽然不能确切地说中小学校数量的差异是完全随机的,但是,由于这种差异主要是由中小学之间的固定成本和规模经济差异导致的,因此,这个变量能够代理学区可选择程度,而与一个地区的特征相关性较低。可能也正是因为因此,作者在他的研究中没有使用其他的工具变量,但由于其解释变量能够让人信服地接受为外生于模型的变量,因此,其结果也是较为可信的。更重要的是,中小学校数量的差异来源于中小学之间的固定成本差异这一点本身具有非常强的经济学含义。从这一意义上来说,找到好的工具变量并不是实证研究唯一可用的研究策略。如果可以在经济学理论的基础上选择巧妙的解释变量,并且能够借助于经济学理论说明其外生性,这不失为一种好的研究策略。

3. 估算社会乘数

以上两种方法运用比较普遍,它们都考察了组群的平均值对于个人行为或表现的影响,以此来识别内生的非市场互动,这种方法就存在组群的平均值本身是内生变量的问题。除此之外,一些较新的研究(Glaeser and Scheinkman, 2003; Glaeser, Sacerdote and Scheinkman, 2003)也用估算社会乘数的方式来直接度量社会互动效应的大小。将社会乘数在理论上的定义运用到实证上时,一个重要的直觉是:用加总数据回归得到的系数(比如教育对收入的影响系数)不同于用个人数据回归得到的系数,因为用加总数据得到的系数已经包含了人与人之间非市场互动产生的效应(Becker and Murphy, 2000)。Glaeser, Sacerdote and Scheinkman (2003)提出了两种从实证上识别社会乘数的方法:第一,用加总数据得到的系数与用个人数据得到的系数之比就衡量了社会乘数;第二,首先对个人数据进行回归,然后得到基于个人数据的被解释变量的预测值,再将其加总后用来回归加总数据,这个变量的系数就度量了社会乘数。估算社会乘数的时候,由于没有将组群的平均值这样的变量放在个人数据的回归方程的右边,回避了这个变量本身可能是个内生变量的问题。Glaeser, Sacerdote and Scheinkman (2003)将以上方法运用到了三项实证研究中,分别考察了同群效应对大学里的社团参与、犯罪率和收入水平的影响。他们的研究有力地证实了社会乘数的存在。

四、群分效应、社区分割和收入分配

伴随着理论的发展,关于非市场互动与群分效应的实证研究也不断向前

推进，这些研究大大增进了人们对于非市场互动的了解，同时也使人们认识到，经济学对相关问题的研究还刚刚开始。在众多的实证研究中，我们的综述将集中在四个核心问题上：（1）择校行为与群分效应的产生；（2）居住区分割效应的动态发展及其对收入分配的影响；（3）公共政策对形成居住区分割和群分效应的影响；（4）试图减少群分效应和社会分割的公共政策是否有效：以 MIO 计划为例。我们将在已有文献的基础上分别对这些问题进行梳理，并对中国的相关问题进行讨论。

（一）我们知道了什么？

首先，一系列的文献讨论了择校行为与群分效应的产生。在这类研究中存在争议的焦点问题是：择校行为是不是会导致群分或分层（stratification），更具体地说，学区间的竞争度（或可供选择程度）是否会影响孩子所在学校的同学的质量（Urquiola, 2005）。在之前的争论中，Clotfelter（1999）认为学区数量的增加会影响孩子们在学校里的同学的质量，从而加剧学区间的群分效应；而 Alesina, Baqir and Hoxby（2004）和 Hoxby（2000）则认为，这一论点没有考虑通过按地段入学产生作用的学区内部的群分效应，他们认为，是学区内的学校数量，而不是学区的数量决定了孩子的同学的同质程度。用计量经济学的语言来说，引起上述分歧的主要原因在于，当用学区数量来度量学区间竞争度时，会存在遗漏变量的内生性问题，因为学区内的群分效应没有得到控制。这一问题在 Urquiola（2005）的研究中得到了较好的解决，正如我们在上一节中提到，他使用地区内部的中小学学校数量差异来度量学区的可选择程度，当中小学校数量差异较大时，说明这个地区内的学区之间的竞争度也较高，这就是一个比较令人信服的外生变量。他发现，学区间的竞争度会影响孩子的同学的质量，并减少私立学校的入学率。由此可见，学区间的竞争度将影响公立学校的学生的构成和分布。因此，择校行为是产生群分效应的一种重要机制。

其次，居住区分割效应的动态发展及其对收入分配的影响也受到实证研究的密切关注。已有研究证实，居住区分割对于犯罪率、教育产出、劳动力市场产出等各个方面都会产生影响（Glaeser et al., 1996；Cutler and Glaeser, 1997）。但是，这一影响是否在长期存在，还是有待于回答的问题。在 Cutler and Glaeser（1997）基于 1990 年美国城市数据的研究发现中，种族间的居住区分割会对黑人的教育和劳动力市场的产出都存在显著的负面效应。在分割严重的地区，黑人的高中毕业率、就业率和收入都更低。由于居住区分割可能存在内生性问题，该研究为城市的居住区分割这一解释变量选择了两组工具变量。第一组是 1962 年市镇级地方政府的数量和州政府财政来源中政府间转移支付所占比重，其理论基础在于：首先，市镇级地方政府的数量能够通过 Tiebout 机制影响分割，当一个地区的地方政府数量越多时，地方政府的税

率和公共品提供的差异性也越大,促进了分割的产生;其次,当州政府¹⁹的财政来源中政府间转移支付所占的比重较高时,将降低这一地区的政府实行分割政策的激励,使分割程度比较低。第二组是用相应地区内的河流数量来代表自然障碍,这是沿用了 Hoxby (2000) 的方法。值得注意的是,由于这两组工具变量都对居住区分割有长期的影响,通过它们来识别居住区分割的效应实际上就揭示了分割的长期效应的存在。而 Oreopoulos (2003) 却怀疑分割效应在长期存在。这项研究基于多伦多市的公有住房计划,由此形成的居住区分割较少地受到个人选择的影响,从而有比较强的外生性。为了观察长期效应,他们采用了个人样本从儿童到成人阶段的长期跟踪数据。实证结果发现,在长期,居住区分割对于劳动力市场的产出并没有显著影响。但是,该研究仅针对了多伦多一个城市的数据,由此得到的结论是否具有普遍性还存在疑问。从理论上来说,在居住区分割严重和人口流动率低的地区,分割产生的效应会长期保持;而在居住区分割并不严重和人口流动率高的地区,分割的效应会随着人口的流动而减弱,因此很难在长期内得以观察。所以,用多伦多的数据发现的分割不存在长期效应这一结论,很可能与这个城市从总体上人口流动比较充分有关。Miguel et al. (2006) 发现,在印度尼西亚的工业化进程中,劳动力向外地的流动会降低当地的相互合作。这从一个侧面说明,频繁的人口流动将减少人们之间的非市场互动,进而可能减轻原社区的非市场互动在长期发生作用的程度。此外,人与人之间的非市场互动是长期存在的,如果人口流动性强,人就可以与不同的人发生互动,就可能减少其年幼时的居住环境的影响,但也有可能幼年时的社区环境的影响不会随着成年后环境的改变而显著减弱。这些都是有待实证研究回答的具有挑战性的问题。

Cutler and Glaeser (1997) 的研究还为我们理解居住区分割对收入分配的影响提供了重要启示。他们发现,居住区分割对于黑人的负面影响尤其显著,而对白人却不显著。他们还进一步提出了三种机制来对此进行解释:第一,居住区分割减少了黑人与其他有较高受教育水平的人群之间的互动;第二,分割尤其会伤害年轻的黑人,因为社会给他们的父母塑造了负面的形象;第三,分割增加了黑人寻找工作的难度。Bayer et al. (2005) 的研究则表明,因种族因素导致的群分和居住区分割显著地减少了黑人,特别是高收入、高受教育水平的黑人家庭的公共品消费。其中的道理很简单,对于黑人家庭来说,除非他们的收入非常高,可以买得起白人社区的房子,否则,他们只能住在黑人社区,即使在收入上他们足以消费更多的公共品,他们也不得不与较低收入的黑人家庭一起去消费较低水平的公共品。因此,居住区分割可能主要

¹⁹ 这里之所以不用市级政府的财政来源中转移支付所占的比例,是因为市级的这一指标与当地的黑人群体贫困程度很相关,不是一个外生的指标。而州政府的这一指标更多地反映了该州政府的政治特征。

影响社会中的弱势群体。那么，当分割加剧时，弱势群体受到的负面影响更大，他们的人力资本水平、收入水平和公共品消费水平都相对更低，这样无疑会扩大整个社会的收入差距。这种效应是非常显著的福利损失，而且在代际之间也会有持续的影响。

再次，在居住区分割的效应得到证实之后，经济学家们试图通过实证研究来发现分割的形成机制。其中一个非常重要的问题是：政府的公共政策如何影响居住区分割的形成？对此问题的回答，能够帮助我们理解，市场机制和政府力量分别在居住区分割中扮演怎样的角色。Cutler (1999) 的研究发现，在美国，政府对于黑人的歧视政策对种族间的居住区分割有重要影响。在 20 世纪 50 年代以前，这种歧视表现为政府直接地用法律手段限制黑人进入白人的居住地。而在 20 世纪后半期，随着政府逐渐取消直接的歧视政策，种族间的居住区分割却并没有显著减轻。自由竞争的住房市场成为居住区分割产生的主要机制，典型的表现是：白人为了与黑人分割，需要支付更高的房价来与白人聚居在一起。Becker and Murphy (2000) 从理论上指出，虽然政府不再能通过法律手段实行直接的歧视政策，但政府采取的一些间接的歧视政策仍然在影响着居住区分割。例如，通过限制白人的居住地规模，阻止了黑人对于白人社区中高质量公共品的享用，由此也会加剧种族间的居住区分割。这提醒我们需要更加关注政府运用的某些公共政策，它们很有可能通过间接的方式影响分割和社会福利。当然，这还需要实证研究的进一步检验。

群分效应还可能通过身份认同和社会流动性的下降而得以加强。很显然，群分效应可能会扩大整个社会的收入差距。当群分效应和社会分割较严重时，拥有较高收入和社会地位的群体之间的非市场互动还有可能对这个群体产生积极影响，而低收入和社会地位的群体之间的非市场互动则更多地产生相互间消极的影响，这时，收入差距将进一步扩大，社会流动性将下降。对于社会的流动性来说，我们有必要提到 Akerlof (2000) 对于身份认同的经济学分析，他认为，不同身份的社会群体的成员会表现出对这一群体行为的遵循，这一视角很好地解释了黑人的高犯罪率和低收入水平。而群分效应和居住区的分割会通过同群效应强化人们的身份认同，尤其会强化弱势群体的身份认同，这将降低整个社会的流动性。一个典型的例子是印度的种姓制度，在一项基于实验的研究中，一些自愿参加实验的初中男生被要求解迷宫题，当他们的种姓不被公开时，不同种姓的学生的表现没有显著差异，而在种姓被公布的情况下，低种姓的学生的表现就显得更差了。造成这一差异的原因在于：低种姓的学生的身份被公开后，他们预期到自己将受到歧视待遇，因此会降低努力程度。作者还进行了一系列的实验设计来证实这个机制的存在。这个实验说明，由历史上的种姓制度造成的人群差异将影响人们的信念和预期，从而形成不同的社会经济表现 (Hoff and Pandey, 2004)。在印度，人们严格

按照种姓阶层的高低聚居,由此形成的分割导致印度的社会流动性非常低,种姓阶层之间持续存在巨大的收入差距,并将影响到社会的现代化进程(Munshi and Rosenzweig, 2006)。最近,印度政府开始重奖不同种姓的居民之间的相互通婚,但这项促进社会融合的政策仍然面临着巨大的困难。

最后,美国政府也开始逐渐意识到社会分割产生的负面影响,尤其是长期存在于美国社会的种族分割。Marmaros and Sacerdote (2006)用 Dartmouth College 随机安排新生寝室的数据研究了友谊的形成。他们发现,种族分割使黑人与白人之间的友谊很难形成,但是,如果将黑人和白人安排在同一寝室或者同一班级,将有效地促进黑人与白人之间友谊的形成,虽然空间上的接近仍然不能使黑人和白人之间的友谊就像白人和白人之间的友谊一样。这一发现带来的政策含义是,缩小种族间的地理差距有利于促进非市场互动,减少分割。自从 20 世纪 90 年代以来,美国政府开始实施一些试图减少群分效应和社会分割的公共政策,其中最具有代表性的就是我们在上文提到的 MTO 计划,随机地给低收入家庭提供住房券。但是,该计划是否能够有效地减少社会分割?在学界却还存在争议。Katz et al. (2001)以及 Ludwig et al. (2001)用该计划的数据研究了非市场互动的的影响,他们都发现,在该计划实行了两三年之后,青少年犯罪率下降了。而且,Ludwig et al. (2001)还发现,在短期内,同群效应提高了低收入家庭的孩子的学习成绩。但是,Sanbonmatsu et al. (2006)研究了进入该计划四到七年的家庭样本,却发现,在长期里,同群效应并没有显著影响低收入家庭的孩子的学习成绩。Sanbonmatsu et al. (2006)对此提出了重要的解释:穷人用住房券搬进富人社区,但却并没有真正地 and 富人邻居达到融合,富人社区对穷人仍然存在歧视,限制了穷人享用富人社区的公共品,尤其是高质量的学校。在这种情况下,一些穷人把孩子送回穷人社区上学,或者干脆搬离富人社区。所以,从长期的实证结果来看,穷人的孩子并没有享受到富人社区的同群效应的积极影响。由此可见,此项旨在减少分割的公共政策在长期内并没有达到预期的效果。我们能够从这个例子中得到关于公共政策制定的重要启示,仅仅缩小穷人和富人间在地理上的差距还不够,只有真正消除对于穷人的歧视,使他们能够享受到高质量的公共品和同群效应的积极影响,才能够有效地从分割走向融合。

(二) 中国的群分、社会分割和公共政策

将对于群分和社会分割的讨论放到中国的背景下,能够帮助我们理解中国社会结构的变化,也对于中国转型时期的公共政策制定有重要意义。

在中国,最为明显的社会分割莫过于城乡之间。在传统的计划经济体制下,政府为保护城市居民的利益和保障工业部门的资本积累,制定了以户籍制度为代表的城市倾向政策,限制了农村劳动力向城市的流动,这导致中国

社会长期处于城乡二元分割的格局。改革开放以来，对于城乡劳动力流动的直接限制逐渐放宽，但是，由于在政策制定的过程中农民利益长期被忽视，城乡分割的局面并没有得以根本的改变，中国的城市化水平仍然较低，城乡收入差距在持续地扩大（蔡昉等，2004；陆铭、陈钊，2004；Lu and Chen，2006；陈钊、陆铭，2006）。在市场化改革的进程中，政府对于农村居民的间接歧视仍然以各种形式存在，农民工在城市受到来自社会保障、义务教育、公共服务和权益保护等各个方面的歧视待遇，也极大地限制了城乡间的劳动力流动（Zhao，1999a，1999b；蔡昉等，2004）。在城市倾向的经济政策下，流动到城市后的农村居民面临更高的生活成本，所以劳动力的流动依然难以缓解城乡收入差距（Yang，1999；Yao，2000），同时，由于存在着城市居民对于农村移民的歧视，在移民规模越来越大的同时，由歧视引起的社会冲突越来越多，城市为此付出的成本越来越高，而这又使得城市居民和农村移民间的分割程度越来越高（陈钊、陆铭，2006）。

而在经济转型的过程中，另一种新的社会分割趋势开始在城市内部出现，由于城市居民的收入差距扩大，而住房市场又逐渐发育成熟，与户籍制度相关的对劳动力流动的限制逐渐减少，城市居民按照收入水平和社会身份的不同形成了居住区的分割。在中国很多大中城市，高收入群体聚居的“富人区”已经比较普遍地存在，而低收入者聚居的“贫民窟”也开始局部地出现。在群分效应的产生过程中，教育市场起到了非常重要的作用，一方面，教育市场和住房市场的一般均衡模型同样适用于中国，这也使得我们为什么可以看到广泛存在的择校现象（Tsang，2001；丁维莉、陆铭，2005；陆铭、蒋仕卿，2006）。我们可以预见，随着市场化程度的不断提高，教育市场、住房市场和劳动力市场进一步发育，人口流动和居住地选择更为自由，中国城市内的群分效应和居住区分割将成为市场均衡的一般结果存在。特别是在按片划分入学地的制度下，居住区分割和教育质量的关系就更为紧密了。

无论是城乡分割还是城市内的居住区分割，都将对社会和谐产生负面影响。城乡分割造成的最直接的影响是，城乡居民间的收入差距持续扩大（蔡昉等，2004；陆铭、陈钊，2004；Lu and Chen，2006；陈钊、陆铭，2006）。而且，城乡分割还形成了城里人与农民两种不同的社会身份，这种身份认同在城市政府对于农民直接或间接的歧视政策下得到强化，阻碍了城乡居民间的社会流动性的提高。城乡分割和对农村流动人口的歧视，还会增加城市的社会冲突和资源在非生产性消耗，对城市部门的经济增长不利。群分效应、社会分割以及与之相关的收入差距还可能对经济持续增长产生负面影响，我们的研究证实，城乡收入差距的扩大在短期内和长期内都对经济增长不利（陆铭等，2005；Wan et al.，2006）。

从实证研究的进展来看，运用中国数据开展的有关非市场互动的研究还刚刚起步。相对较多的研究集中在资本市场的“羊群效应”和社会互动上，

例如最近李涛 (2006a ; 2006b) 关于股市上的社会互动的研究。但是在那些与群分效应有关的社会互动的研究方面, 基于中国的数据的研究却几乎仍然是空白。在教育方面, Ding and Lehrer (2005) 证实了中学教育中存在非线性的同群效应。在社会资本方面, 经济学家和社会学家们共同发现了它对于中国劳动力市场的重要影响 (参见 Knight and Yueh, 2002 ; 桂勇等, 2004 ; 及这些研究的文献综述)。我们自己的研究发现, 在中国农村地区, 同村的其他家庭的公共信任能够显著地减少本家庭的贫困和劳动力流动, 而且, 总体上来说社区层面的社会资本对于减少农村贫困的作用并不会随着市场化的推进而降低 (陆铭、张爽, 2006 ; 张爽、陆铭、章元, 2006)。在更为广阔的研究主题里, 中国经济学家还有非常多的工作需要做。事实上, 在中国, 人们对很多具有重大学术价值和公共政策含义的问题还缺乏了解, 比如说: 中国的择校行为对学校提高教育质量是否有积极作用? 中国居民的收入是否持续地受到他们小时候的生活环境的影响? 中国城市的群分效应和居住区分割是否影响到了人们的收入、人力资本积累和行为模式? 在城市化的进程中, 农村移民是否因为受到城市居民的歧视而影响到他们对城市政府的信任, 并且聚居在一起, 影响到城市社会的和谐? 这些都有待于实证研究去回答。²⁰

当社会分割造成的负面影响已经不可避免时, 政府应及时给予重视, 否则, 一旦社会分割严重到一定程度, 将给经济和社会的和谐发展带来巨大的障碍, 并且也使得相应教育均等、缓解贫困等的公共政策实施效果大打折扣。那么政府能够做些什么来减轻社会分割呢? 政府至少能够做到以下三点: 第一, 政府应该逐渐减少, 直至取消对于农村居民的直接或间接的歧视政策, 促进城乡从分割走向融合。第二, 在城市内部, 政府也要避免有可能会加剧居住区分割的公共政策, 例如成片地提供廉租房。由于选择廉租房的通常是穷人, 他们聚居产生的同群效应将加剧穷人与富人居住区之间的分割。政府可以考虑将廉租房在空间上适当加以分散, 有利于低收入群体融入社会。²¹ 第三, 由于分割主要对弱势群体产生负面影响, 政府应该更多地关注弱势群体的利益, 加大对于低收入居民的社区的公共品提供的投入, 减轻群分和居住区分割对他们带来的损失。此外, 政府还应该在一些对群分效应和社会分割产生影响的政策出台时作谨慎的考虑, 避免因为缺乏科学的论证, 使得政策的效果反而可能加剧群分效应和社会分割。比如说, 中国实行的是基础教育按片入学的政策, 与之相伴随的是大量的择校现象。事实上, 择校现象本身就是学校之间存在竞争的表现, 应该采取一些政策 (如促进学校间竞争、给

²⁰ 一些社会学家的案例研究也可以为我们了解居民间的社会互动和身份认同提供信息。例如, 张敏 (2006) 报告了温州基督教徒们之间的相互影响和身份认同, 以及他们与非教徒之间的“人以群分”和矛盾冲突。

²¹ 法国在这方面做得比较好。

低收入家庭的好孩子提供奖学金和教育券等)缓解择校行为和群分效应的负面影响,但不应简单地禁止择校。2006年8月4日《新华每日电讯》的一则消息说:“由于江西省教育厅连续接到部分学校的电话或书面材料反映,一些普通高中尤其是部分省优秀重点中学跨市、县(区)违规招生,并向家长和学生承诺支付一定‘奖学金’、减免学杂费,有的甚至用‘重金收购’的方式‘买学生’。为此,江西省教育厅下发紧急通知,在坚决制止普通高中违规招生行为的同时,将违规招收的学生劝退回原地入学。”直接禁止择校或者不允许学校竞争好生源的结果可能是:一方面,学校间的竞争被人为地削弱,教育资源的利用效率将难以保证;另一方面,一些低收入家庭的好孩子可能因此而失去上好学校的机会,而高收入家庭却只能通过买房子来让子女进好学校。这样一来,高收入家庭其实并没有停止择校,而是将择校费交给了房地产商(或好学校周围房子的业主),而不是学校,于是,好学校附近的房价将可能因为政府禁止择校的政策而进一步上升,群分效应和居住区分割将可能因此而加剧,低收入群体将反而因此而受损(陆铭、蒋仕卿,2006)。

五、总 结

如果说20世纪贝克尔给经济学带来的是一个“帝国主义”时代的话,他在世纪之交出版的著作《社会经济学:社会环境中的市场行为》更多地体现了经济学向社会学家的学习和借鉴。在社会经济学的兴起过程中,经济学终于突破了将人看作是市场上独立行动的个体的范式,使得人在经济学的视野里真正地成为了社会的人。在经济学理论中,人和人之间的相互影响直接进入了理论模型,空间的重要性得到了经济学家的高度重视,这即使不是经济学的一场革命,也至少是极大的发展。

由于存在着非市场的互动,市场的均衡结果是人们按不同的收入和社会地位而分层,从而出现“群分效应”,在现实中又突出地表现为居住区的分割,这对于低收入阶层的收入、人力资本积累和公共品消费有着显著的负面影响,对经济、政治和社会的现代化也有阻碍作用。政府应在开展实证研究的基础上,有针对性地采取一些经济和社会政策减少群分效应和居住区分割的负面影响,促进经济发展和社会和谐。

参 考 文 献

- [1] Acemoglu, D., S. Johnson, and J. Robinson, “The Colonial Origins of Comparative Development: An Empirical Investigation”, *American Economic Review*, 2001, 91(5), 1369—1401.
- [2] Akerlof, G., “Social Distance and Social Decisions”, *Econometrica*, 1997, 65(5), 1005—1027.

- [3] Akerlof, G., and R. Kranton, "Economics and Identity", *Quarterly Journal of Economics*, 2000, 115 (3), 715—753.
- [4] Alesina, A., B. Reza, and C. Hoxby, "Political Jurisdictions in Heterogeneous Communities", *Journal of Political Economy*, 2004, 112(2), 348—396.
- [5] Bayer, P., R. McMillan, and K. Rueben, "Residential Segregation in General Equilibrium", NBER Working Papers 11095, 2005.
- [6] Bayer, P., R. McMillan, and K. Rueben, "An Equilibrium Model of Sorting in an Urban Housing Market", NBER Working Papers 10865, 2004.
- [7] Becker, G., and K. Murphy, *Social Economics: Market Behavior in a Social Environment*. Harvard University Press, 2000.
- [8] Bernheim, D., "A Theory of Conformity", *Journal of Political Economy*, 1994, 102(5), 841—877.
- [9] Bikhchandani, S., D. Hirshleifer, and I. Welch, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades", *Journal of Political Economy*, 1992, 100(5), 992—1026.
- [10] Brock, W., and S. Durlauf, "Interaction-Based Models", in Heckman, J. and E. Leamer (eds.), *Handbook of Econometrics*, vol. 5. Amsterdam: North-Holland, 2000, 3297—3380.
- [11] 蔡昉、都阳、王美艳,《劳动力流动的政治经济学》。上海:上海三联书店,上海人民出版社 2004 年。
- [12] Case, A., and L. Katz, "The Company You Keep: The Effect of Family and Neighborhood on Disadvantaged Families", NBER Working Papers 3705, 1991.
- [13] 陈钊、陆铭,《从分割到融合:城乡经济增长和社会融合的政治经济学》,复旦大学工作论文, 2006 年。
- [14] Clotfelter, C., "Public School Segregation in Metropolitan Areas", *Land Economics*, 1999, 75(4), 487—504.
- [15] Cutler, D., and E. Glaeser, "Are Ghettos Good or Bad?" *Quarterly Journal of Economics*, 1997, 112 (3), 826—872.
- [16] Cutler, D., E. Glaeser, and J. Vigdor, "The Rise and Decline of the American Ghetto", *Journal of Political Economy*, 1999, 107(3), 455—506.
- [17] Davidson, R., and J. Mackinnon, *Econometric Theory and Methods*. New York: Oxford University Press, 2004.
- [18] Ding, W., and S. Lehrer, "Estimating Dynamic Treatment Effects from Project STAR", Working Paper, School of Policy Studies, Queen's University, 2004.
- [19] Ding, W., and S. Lehrer, "Do Peers Affect Student Achievement in China's Secondary Schools?" *Review of Economics and Statistics*, 2005, forthcoming.
- [20] 丁维莉、陆铭,《教育的公平与效率是鱼和熊掌吗——基础教育财政的一般均衡分析》,《中国社会科学》2005 年第 6 期,第 47—57 页。
- [21] Duncan, G., and S. Raudenbush, "Neighborhoods and Adolescent Development: How Can We Determine the Links?" in Booth, A., and A. Crouter (eds.), *Does It Take a Village? Community Effects on Children, Adolescents, and Families*. State College, Penn.: Pennsylvania State University Press, 2001, 105—136.
- [22] Epple, D., and H. Sieg, "Estimating Equilibrium Models of Local Jurisdictions", *Journal of Political Economy*, 1991, 107(4), 645—681.

- [23] Epple, D., and R. Romano, "Competition between Private and Public Schools, Vouchers, and Peer-Group Effects", *American Economic Review*, 1998, 88(1), 33—62.
- [24] Epple, D., and R. Romano, "Educational Vouchers and Cream Skimming", NBER Working Papers 9354, 2002.
- [25] Glaeser, E., and J. Scheinkman, "Measuring Social Interactions", in Durlauf, S. and P. Young (eds.), *Social Dynamics*. Boston, MA: MIT Press, 2001, 83—102.
- [26] Glaeser, E., and J. Scheinkman, "Non-market Interactions", in Dewatripont, M. L. Hansen, and S. Turnovsky (eds.), *Advances in Economics and Econometrics: Theory and Applications*. Cambridge, MA: Cambridge University Press, 2003, 339—369.
- [27] Glaeser, E., R. La Porta, F. Lopez-de-Silanes, and A. Shleifer, "Do Institutions Cause Growth?" NBER Working Papers 10568, 2004.
- [28] Glaeser, E., B. Sacerdote, and J. Scheinkman, "Crime and Social Interactions", *Quarterly Journal of Economics*, 1996, 111(2), 507—548.
- [29] Glaeser, E., B. Sacerdote, and J. Scheinkman, "The Social Multiplier", *Journal of the European Economic Association*, 2003, 1(2—3), 345—53.
- [30] Granovetter, M., *Getting a Job*, 2nd edition. Chicago, IL: University of Chicago Press, 1995.
- [31] 桂勇、陆德梅、朱国宏, "经济转型、关系强度与求职行为——一项关于失业群体的实证研究", 《世界经济文汇》2004年第2期, 第21—33页。
- [32] Hoff, K., and P. Pandey, "Belief Systems and Durable Inequalities: An Experimental Investigation of Indian Caste", World Bank Policy Research Working Paper No. 3351, 2004.
- [33] Hoxby, C., "Does Competition among Public Schools Benefit Students and Taxpayers?" *American Economic Review*, 2000, 90(5), 1209—38.
- [34] Hoxby, C., "Competition among Schools: A Reply to Rothstein (2004)", NBER Working Papers 11216, 2005.
- [35] Ioannides, Y., and L. Loury, "Job Information Networks, Neighborhood Effects, and Inequality", *Journal of Economic Literature*, 2004, 42(4), 1056—1093.
- [36] Jones, S., *The Economics of Conformism*. New York, NY: Basil Blackwell, 1984.
- [37] Jones, S., "Book Review on Social Economics", *Journal of Economic Literature*, 2002, 40(2), 575—577.
- [38] Katz, L., A. Kling, and J. Liebman, "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment", *Quarterly Journal of Economics*, 2001, 116(2), 607—654.
- [39] Knight, J., and L. Yueh, "The Role of Social Capital in the Labor Market in China", Oxford University, Department of Economics Discussion Paper, 2002. (中文版, "社会资本在中国劳动力市场中的作用" 载李实、佐藤宏主编, 《经济转型的代价——中国城市失业、贫困、收入差距的经验分析》。北京: 中国财政经济出版社 2004年, 第310—346页。)
- [40] Leibenstein, H., "Bandwagon, Snob, and Veblen Effects in the Theory of Consumer Demand", *Quarterly Journal of Economics*, 1950, 64(2), 183—207.
- [41] 李涛 a, "社会互动、信任与股市参与", 《经济研究》2006年第1期, 第34—45页。
- [42] 李涛 b, "社会互动与投资选择", 《经济研究》2006年第8期, 第45—57页。
- [43] 陆铭、陈钊, "城市化、城市倾向的经济政策与城乡收入差距", 《经济研究》2004年第6期, 第50—58页。

- [44] Lu, M., and Z. Chen, "Urbanization, Urban-Biased Policies and Urban-Rural Inequality in China: 1987—2001", *Chinese Economy*, 2006, 39(3), 42—63.
- [45] 陆铭、陈钊、万广华, "因患寡, 而患不均: 中国的收入差距、投资、教育和增长的相互影响", 《经济研究》2006 年第 12 期, 第 4—14 页。
- [46] 陆铭、蒋仕卿, "反思教育产业化的反思——有效利用教育资源的理论与政策", 复旦大学工作论文 2006 年。
- [47] 陆铭、张爽, "离开了土地, 却未离开家乡? ——中国农村的公共信任与劳动力流动", 复旦大学工作论文 2006 年。
- [48] Ludwig, J., P. Hirschfeld, and G. Duncan, "Urban Poverty and Juvenile Crime: Evidence from a Randomized Housing-Mobility Experiment", *Quarterly Journal of Economics*, 2001, 116(2), 665—679.
- [49] Manski, C., "Economic Analysis of Social Interactions", *Journal of Economic Perspectives*, 2000, 14(3), 115—136.
- [50] Manski, C., "Identification of Endogenous Social Effects: The Reflection Problem", *Review of Economic Studies*, 1993, 60(3), 531—542.
- [51] Manski, C., "Identification of Anonymous Endogenous Interactions", in Author, B., S. Durlauf, and D. Lane (eds.), *The Economy as an Evolving Complex System II*. Mass.: Addison-Wesley, 1997.
- [52] Marmaro, D., and B. Sacerdote, "How do Friendships Form?" *Quarterly Journal of Economics*, 2006, 121(1), 79—119.
- [53] Miguel, E., P. Gertler, and D. Levine, "Does Industrialization Built or Destroy Social Networks", *Economic Development and Culture Change*, 2006, 54(2), 524—528.
- [54] Miller, C., S. McIntyre, and M. Mantrala, "Towards Formalising Fashion Theory", *Journal of Marketing Research*, 1993, 30, 142—151.
- [55] Moffit, R., "Policy Interventions, Low-Level Equilibria, and Social Interactions", Working Paper, Department of Economics, Johns Hopkins University, 1999.
- [56] Munshi, K., and M. Rosenzweig, "Traditional Institutions Meet the Modern World: Caste, Gender and Schooling Choice in a Globalizing Economy", *American Economic Review*, 2006, 96(4), 1225—1252.
- [57] Oreopoulos, P., "The Long-Run Consequences of Living in a Poor Neighborhood", *Quarterly Journal of Economics*, 2003, 118(4), 1533—1575.
- [58] Rosenbaum, J., "Black Pioneers-Do Their Moves to the Suburbs Increase Economic Opportunity for Mothers and Children?" *Housing Policy Debates*, 1991, 2(4), 1179—1213.
- [59] Rothstein, J., "Does Competition Among Public Schools Benefit Students and Taxpayers? A Comment on Hoxby (2000)", NBER Working Papers 11215, 2004.
- [60] Rothstein, J., "Good Principals or Good Peers? Parental Valuation of School Characteristics, Tiebout Equilibrium, and the Incentive Effects of Competition among Jurisdictions", *American Economic Review*, 2006, 96(4), 1333—1350.
- [61] Sacerdote, B., "Peer Effects with Random Assignment: Results for Dartmouth Roommates", *Quarterly Journal of Economics*, 2001, 116(2), 681—704.
- [62] Sah, R., "Social Osmosis and Patterns of Crime", *Journal of Political Economy*, 1991, 99(6), 1272—1295.
- [63] Sanbonmatsu, L., J. Kling, G. Duncan, and J. Brooks-Gunn, "Neighborhoods and Academic Achievement: Results from the Moving to Opportunity Experiment", NBER Working Papers 11909, 2006.

- [64] Scheinkman, J. , “ Social Interaction ” , The New Palgrave , 2nd edition , 2006 , forthcoming.
- [65] Tiebout , C. , “ A Pure Theory of Local Expenditures ” , Journal of Political Economy , 1956 , 64 (5) , 416—424.
- [66] Tsang , M. , “ School Choice in the People’s Republic of China ” , Working Paper , Teachers College Columbia University , 2001.
- [67] Urquiola , M. , “ Does School Choice Lead to Sorting ? Evidence from Tiebout Variation ” , American Economic Review , 2005 , 95 (4) , 1310—1326.
- [68] Wan , G. , M. Lu , and Z. Chen , “ The Inequality-Growth Nexus in the Short and Long Runs : Empirical Evidence from China ” , Journal of Comparative Economics , 2006 , 34 (4) , 654—667.
- [69] Winston , G. , “ Subsidies , Hierarchy and Peers : The Awkward Economics of Higher Education ” , Journal of Economic Perspectives , 1999 , 13 (1) , 13—36.
- [70] Yang , D. , “ Urban-Biased Policies and Rising Income Inequality in China ” , American Economic Review , 1999 , 89 (2) , 306—310.
- [71] Yao , S. , “ Economic Development and Poverty Reduction in China over 20 Years of Reform ” , Economic Development and Cultural Change , 2000 , 48 (3) , 447—474.
- [72] Zanella , G. , “ Social Interactions and Economic Behavior ” , Working Paper , Department of Economics , University Wisconsin-Madison , and University of Siena , 2004.
- [73] 张敏 , “ 基督徒身份认同——浙江温州案例 ” 载张静主编《身份认同研究 : 观念态度、理据》。上海人民出版社 2006 年 , 第 85—146 页。
- [74] 张爽、陆铭、章元 , “ 社会资本的作用随市场化进程减弱还是加强 ? ——来自中国农村贫困的实证研究 ” , 《经济学(季刊)》2007 年第 6 卷第 2 期 , 第 539—560 页。
- [75] Zhao , Y. , “ Leaving the Countryside : Rural-to-Urban Migration Decisions in China ” , American Economic Review , 1999a , 89 (2) , 281—286.
- [76] Zhao , Y. , “ Labor Migration and Earning Differences : The Case of Rural China ” , Economic Development and Culture Change , 1999b , 47 (4) , 767—782.

Non-Market Interaction and Sorting : A Literature Review

MING LU SHUANG ZHANG
(Fudan University)

Abstract When non-market interaction exists , the result of sorting as a market equilibrium shows that , people are stratified by income level and social status. A typical kind of sorting is residential segregation in the real world. Sorting has significantly negative effects on income , human capital accumulation and public good consumption for people with low-income. Based on empirical studies , government should make some public policies to mitigate the negative effect of sorting and residential segregation , thus to promote economic development and social harmony.

JEL Classification Z13 , J15 , J18