

# 经济学(季刊)

*China Economic Quarterly*

## 论文

周绍杰等：中国城市居民的家庭收入、消费和储蓄行为

马小勇 白永秀：中国农户的收入风险应对机制  
与消费波动

周云波：城市化、城乡差距及全国居民总体收入  
差距的变动

郭继强 陆利丽：工资差异均值分解的一种新改进

刘雅南 邵宜航：政府竞争与区域经济差异

王贤彬等：地方官员更替与经济增长

杨剑侠等：消费者的上瘾行为和运营商定价

龚 强：消费者谈判能力与厂商标价策略

李晓义 李建标：不完备市场的多层次治理

肖俊极 李国栋：非价格策略与消费者购买行为

行伟波 李善同：本地偏好、边界效应与市场一体化

鲁晓东 李荣林：区域经济一体化、FDI与国际生产转移

王自锋：汇率水平与波动对外国直接投资的影响

程培堃等：FDI对国内投资的挤出(入)效应

符 淼：地理距离和技术外溢效应

吴延兵 刘霞辉：人力资本与研发行为

雷国雄 陈 恩：一个拟生物演化制度模型

## 综述

赵红军：李约瑟之谜

北京大学中国经济研究中心

北京大学出版社

2009·7



**图书在版编目(CIP)数据**

经济学:季刊. 第8卷. 第4期/林毅夫,姚洋主编. —北京:北京大学出版社,  
2009.7

ISBN 978-7-301-15471-7

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(2009)第 116493 号

**书 名: 经济学(季刊) 第8卷·第4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 978-7-301-15471-7/F·2245

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](mailto: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 29 印张 564 千字

2009 年 7 月第 1 版 2009 年 7 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](mailto:fd@pup.pku.edu.cn)

## 2007—2008 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。获奖论文由本刊三位主编在提名论文中以多数投票的形式决定。

2007—2008 年度(第 7 卷)《经济学(季刊)》最佳论文奖提名论文是(排名不分先后):

罗雨泽、朱善利、陈玉宇、罗来军:“外商直接投资的空间外溢效应”(第 7 卷第 2 期)

佐藤宏、李实:“中国农村地区的家庭成分、家庭文化和教育”(第 7 卷第 4 期)

刘永平、陆铭:“放松计划生育政策将如何影响经济增长”(第 7 卷第 4 期)

**2007—2008 年度(第 7 卷)《经济学(季刊)》最佳论文奖授予日本一桥大学经济系佐藤宏和北京师范大学经济与工商管理学院李实合作发表于《经济学(季刊)》第 7 卷第 4 期的“中国农村地区的家庭成分、家庭文化和教育”。**

# 经济学(季刊)

第8卷 第4期

(总第34期)

2009年7月

## 目 录

### 论 文

- 中国城市居民的家庭收入、消费和储蓄行为：  
一个基于组群的实证研究 ..... 周绍杰 张俊森 李宏彬 1197
- 中国农户的收入风险应对机制与消费波动：  
来自陕西的经验证据 ..... 马小勇 白永秀 1221
- 城市化、城乡差距以及全国居民总体收入差距的变动  
——收入差距倒U形假说的实证检验 ..... 周云波 1239
- 工资差异均值分解的一种新改进 ..... 郭继强 陆利丽 1257
- 政府竞争与区域经济差异 ..... 刘雅南 邵宜航 1281
- 地方官员更替与经济增长 ..... 王贤彬 徐现祥 李 邨 1301
- 运营商利用消费者的上瘾行为定价了吗  
——来自中国网络游戏产业的经验证据 ... 杨剑侠 陈宏民 包 兴 1329
- 消费者谈判能力与厂商标价策略  
——模型及基于实验经济学的检验 ..... 龚 强 1383
- 不完备市场的多层次治理  
——基于比较制度实验的研究 ..... 李晓义 李建标 1407
- 非价格策略与消费者购买行为  
——以会员积分制为例 ..... 肖俊极 李国栋 1435
- 本地偏好、边界效应与市场一体化  
——基于中国地区间增值税流动数据的实证研究 ... 行伟波 李善同 1455
- 区域经济一体化、FDI与国际生产转移：  
一个自由资本模型 ..... 鲁晓东 李荣林 1475
- 汇率水平与波动程度对外国直接投资的影响研究 ..... 王自锋 1497
- FDI对国内投资的挤出(入)效应：  
产业组织视角 ..... 程培堃 周应恒 殷志扬 1527
- 地理距离和技术外溢效应  
——对技术和经济集聚现象的空间计量学解释 ..... 符 森 1549
- 人力资本与研发行为  
——基于民营企业调研数据的分析 ..... 吴延兵 刘霞辉 1567
- 制度变迁：一个拟生物演化模型 ..... 雷国雄 陈 恩 1591
- 综 述
- 李约瑟之谜：经济学家应接受旧解还是新解？ ..... 赵红军 1615

# China Economic Quarterly

Vol. 8 No. 4

July, 2009

## CONTENTS

### Papers

- Income, Consumption and Saving Behavior of Chinese Urban Households;  
A Cohort Analysis ..... Shaojie Zhou, Junsen Zhang and Hongbin Li 1197
- Chinese Farmers' Income Risk Coping Mechanisms and Consumption  
Fluctuations: An Empirical Study from Shanxi Province  
..... Xiaoyong Ma and Yongxiu Bai 1221
- Urbanization, Urban-Rural Income Gap and Overall Income Inequality  
in China: An Empirical Test of the Inverse-U Hypothesis ... Yunbo Zhou 1239
- A New Approach to Ameliorating the Average Value Decomposition of  
Wage Differentials ..... Jiqiang Guo and Lili Lu 1257
- Intergovernmental Competition and Regional Disparities  
..... Yanan Liu and Yihang Shao 1281
- Provincial Governors' Turnovers and Economic Growth:  
Evidence from China ..... Xianbin Wang, Xianxiang Xu and Xun Li 1301
- Have Service Providers Taken Advantage of Consumers' Addictions in Their  
Pricing Strategies? Evidence from Chinese Online Games  
..... Jianxia Yang, Hongmin Chen and Xing Bao 1329
- Consumers' Bargaining Power and Firms' Price-Posting Strategy:  
A Model and Its Experimental Tests ..... Qiang Gong 1383
- Multi-level Governance of Incomplete Markets: A Study Based on  
Comparative Institutional Experiments ..... Xiaoyi Li and Jianbiao Li 1407
- Non-Price Strategy and Consumers' Purchase Behavior: Evidence from  
A Frequency Reward Program ..... Junji Xiao and Guodong Li 1435
- Home Biases, Border Effects and Market Integration: An Empirical Study  
Based on Inter-provincial VAT Survey ... Weibo Xing and Shantong Li 1455
- Regional Integration, FDI and International Production Location:  
A Footloose Capital Model ..... Xiaodong Lu and Ronglin Li 1475
- The Effects of Exchange Rate Levels and Volatility on FDI ... Zifeng Wang 1497
- Trade-off between FDI and Domestic Investment: From the Industrial Organization  
Perspective ..... Peigang Cheng, Yingheng Zhou and Zhiyang Yin 1527
- Geographical Distance and Technological Spillover Effects: A Spatial  
Econometric Explanation of Technological and Economic  
Agglomeration Phenomena ..... Miao Fu 1549
- Human Capital and R&D Activity: Evidence from Chinese Private  
Enterprises ..... Yanbing Wu and Xiahui Liu 1567
- Institutional Change: A Bionic Evolutional Model ... Guoxiong Lei and En Chen 1591

### Review Papers

- The Needham Puzzle: Should Economists Accept Old Hypotheses  
or New Ones? ..... Hongjun Zhao 1615

## 中国城市居民的家庭收入、消费和储蓄行为： 一个基于组群的实证研究

周绍杰 张俊森 李宏彬\*

**摘要** 近年来,中国的高储蓄率越来越引起关注。本文利用中国城镇居民住户调查数据对城市家庭的收入、消费以及储蓄率进行基于组群分析的实证研究。实证结果表明,中国城市家庭的储蓄行为具有独特性,家庭储蓄率不断提高。本文从两个方面对此做出解释:其一,各个组群的家庭消费增长慢于收入增长;其二,年轻组群的高储蓄倾向及其在样本中随时间的比重不断提高。此外,本文也分析了养老金收入对于年老组群家庭收入和储蓄的影响。

**关键词** 储蓄率, 组群效应, 组群分析

### 一、引言

在过去的三十年中,中国通过实施改革开放政策实现了由计划经济向市场经济的转型,并实现了长期的高速增长。中国的高速经济增长与中国的高投资率密切相关。根据表1,中国的全社会固定资产投资率自20世纪80年代末以来一直保持着较高的水平,2003年达到近50%。与世界其他国家相比,中国的投资率无疑是非常高的。现有的研究也确认了中国的高储蓄率问题。Kraay (2000)发现中国的储蓄率明显高于世界其他国家;Modigliani and Cao (2004)也发现中国90年代初期的家庭和公司的储蓄率明显高于部分发达国家在60、70年代经济处于高速增长时期。从家庭部门来看,城乡居民家庭银行存款在2003年超过10万亿元,相当于当年GDP的88%,而城乡居民人民币储蓄存款余额的年增加额在1988—1996年和2000—2003年期间也分别具有明显的上升趋势。如果把居民家庭持有的债券和股票计算在内,居民家庭的金融资产持有水平就更高了。于是,就产生了如下的问题:中国的储蓄率为什么如此之高?

\* 周绍杰,清华大学公共管理学院;电话:(010)62797446;E-mail:zhoushaojie@tsinghua.edu.cn;张俊森,香港中文大学经济系;电话:(00852)26098186;E-mail:jszhang@cuhk.edu.hk;李宏彬,清华大学经管学院经济系;电话:(010)62792924;E-mail:lihongbin@sem.tsinghua.edu.cn。通信作者及地址:周绍杰,清华大学公共管理学院318室,100084。本文作者非常感谢两位匿名审稿人以及《经济学(季刊)》主编姚洋教授对本文所提出的有益修改建议。文中错误和疏漏之处完全由作者负责。

表1 国内生产总值、固定资产投资率以及城乡居民储蓄统计(1988—2003)

年份	全社会固定资产投资率 (%)	城乡居民人民币 储蓄存款年底 余额(亿元)	城乡居民人民币 储蓄存款年 增加额(亿元)	城乡居民人民币储蓄 存款年底余额 /GDP(%)
1988	31.17	3 822.2	740.8	25.60
1989	26.08	5 196.4	1 374.2	30.73
1990	24.36	7 119.8	1 923.4	38.39
1991	25.88	9 241.6	2 121.8	42.75
1992	30.33	11 759.4	2 517.8	44.15
1993	37.74	15 203.5	3 444.1	43.90
1994	38.13	21 518.8	6 315.3	46.02
1995	35.10	29 662.3	8 143.5	50.72
1996	34.41	38 520.8	8 858.5	56.74
1997	33.92	46 279.8	7 759	62.15
1998	36.65	53 407.5	7 615.4	68.17
1999	36.26	59 621.8	6 253	72.65
2000	37.01	64 332.4	4 976.7	71.91
2001	39.04	73 762.4	9 457.6	75.80
2002	42.83	86 910.6	13 233.2	82.64
2003	49.93	103 617.3	16 631.6	88.27

数据来源:国家统计局,《中国统计年鉴2000》;国家统计局,《中国统计年鉴2005》。

在解释中国总体储蓄率方面,以往的研究分别从经济增长(Qian, 1988; Modigliani and Cao, 2004)、人口(Modigliani and Cao, 2004)、企业部门的高储蓄(Kujis, 2005; Chamon and Prasad, 2007)以及预防性储蓄动机(Meng, 2004)等方面进行了解释。事实上,与其他国家相比,中国的高储蓄率不仅与家庭部门的高储蓄率有关,更与企业部门和政府部门的储蓄有关(Kujis, 2005)。从部门划分来看,家庭部门储蓄只是总体储蓄的一部分。因此,如果单纯地从家庭部门储蓄理解中国的总体储蓄率,将会产生理解上的偏差,而以往从家庭、企业和政府三部分分解的视角研究中国储蓄问题也证实了这一点(Qian, 1988; Kraay, 2000; Kujis, 2005; Chamon and Prasad, 2007)。此外,总体的高储蓄也与国民收入分配有关。从1980年以来的历史数据来看,家庭收入占国民收入的比重具有下降趋势,不断下降的居民家庭收入份额抑制了居民消费需求的提升,在一定程度上解释了储蓄率总体的上升趋势(Aziz and Cui, 2007)。

总体来说,关于中国的储蓄问题,以往的研究主要集中于宏观研究方面,利用微观数据的实证研究则相对较少,尤其对居民家庭的收入模式、支出模式以及储蓄率的动态变化缺乏深入分析,这包括如下方面的问题:在居民家庭收入随经济增长不断提高的背景下,居民家庭收入、消费以及储蓄率的生命周期模式是什么?上述变量在代际之间是否具有显著差异?这种代际之间的系统性的差异如何解释?中国居民家庭的储蓄率模式是否与其他国家存在显著差异?如何从家庭部门的视角,理解中国储蓄率的提高?国家的收入分



配政策（例如，养老金改革）是否对家庭的收支模式和储蓄率具有显著影响？

本文利用中国城镇住户调查数据（Chinese Urban Household Survey, CUHS）对上述问题进行实证研究。下文的结构安排如下：第二部分介绍基于组群分析（cohort analysis）的实证方法，第三部分和第四部分分别是模型和计量设定，第五部分是数据描述和组群构造，第六部分为实证结果，第七部分解释中国城市家庭的储蓄模式，最后一个部分是本文的结论。

## 二、组群分析方法

作为标准的消费理论，生命周期假说（Life Cycle Hypothesis）在家庭终生效用最大化以及信贷市场完备的假设下得出家庭倾向于平滑其终生消费的结论。在家庭收入呈现出一个驼峰形状（hump-shaped）的条件下，消费平滑使得家庭资产持有表现出生命周期模式，即家庭资产呈现出驼峰形状，这与家庭储蓄率随年龄的动态变化密切相关。家庭储蓄率的生命周期曲线是理解家庭消费行为和家庭资产动态变化的重要依据。因此，在实证研究中，可以通过家庭储蓄率的生命周期特征来检验家庭消费是否遵循了生命周期假说的预测。

在实证研究中，对家庭部门的调查数据为研究家庭储蓄行为的生命周期特征提供了素材。世界上大多数国家对家庭的经济调查都是采取了样本轮换的做法，而不是对一个特定家庭做长期观察<sup>1</sup>，形成了具有连续时间序列特性的独立截面数据（time series of independent cross-sectional data，后文称为 TSICS 数据）。利用此类数据可以跟踪某一特定“组群”（cohort）随时间变化的平均行为。这样，在研究家庭经济行为的生命周期特征时，尽管我们无法做到对一个特定家庭的经济行为作终生观测，但是可以通过分析“一类家庭”平均行为特征的方法来分析家庭经济行为，这种方法在实证研究中被称为组群分析。这一技术是通过样本的组群来构造合成面板数据（synthetic panel data）。

一般来说，组群的定义是根据观测样本所具有的固定特征，该固定特征不会随时间发生变化，例如，个体的出生年份和性别。在实证研究中，组群通常是由个体的出生年份定义的，而在以家庭为观测对象的研究中，组群的定义通常是根据家庭户主的出生年份。在给出组群的定义后，“组群值”（cohort value）是该组群内所有观测值的均值。在实证研究中，该方法首次被 Browning *et al.*（1985）采用，Verbeek and Nijman（1992）分析了这种方法的适用条件。<sup>2</sup>尽管组群分析方法是因为缺乏真实的面板数据，但是，利用合成

<sup>1</sup> 一般来说，对一个家庭经济行为的调查如果期限过长，会对被调查家庭造成负担，从而不利于家庭真实经济行为信息的搜集。

<sup>2</sup> 通过利用真实面板数据以及利用合成面板数据的实证结果进行比较，他们发现合成面板数据使用的有效性取决于两个条件：其一，每个组群内的观测个体的数量应该足够大（100个、200个观测个体）；其二，每一个组群的组群均值应当具有足够的随时间的变化性（time variation）。

面板数据的实证研究并不必然地导致实证结果的不可信。相反,这种方法相对于面板数据而言具有两种优势:可处理样本损失(sample attrition)以及测量误差(errors-in-variables)问题<sup>3</sup>(Deaton, 1985)。事实上,组群分析方法扩展了 TSICS 数据的使用,在实证研究中,合成面板数据是宏观研究和微观研究的桥梁,可以同时分析与“商业周期”(the “business” cycle)相关的变化和与生命周期相关的变化(Browning *et al.*, 1985)。

尽管合成面板数据提供了特定组群在特定年龄阶段的经济行为,然而,如果要估计家庭经济行为的生命周期特征,则需要对组群之间的系统性差异做出调整。在实证研究中,组群之间的差异被定义为组群效应(cohort effect),受到广泛关注(Shorrocks, 1975; Deaton and Paxson, 1989, 1993, 2000; Gokhale *et al.*, 1996; Attanasio *et al.*, 1999; Jappelli, 1999; Gibson and Scobie, 2001; Fernández-Villaverde and Krueger, 2002)。通常来说,组群效应是使用 TSICS 数据做实证分析不可忽视的重要方面。例如, Jappelli (1999) 比较了利用截面数据估计出的消费行为的年龄曲线(age profile)和利用经过组群效应调整的消费行为的年龄曲线,发现组群效应具有显著的影响。事实上,利用 TSICS 数据做实证分析时,如果不控制组群效应,组群效应将会被混同到所估计的年龄曲线中,从而造成估计上的误差。因此,组群分析的一个重要使命就是在估计家庭经济行为的年龄曲线时控制组群效应的影响。

在组群分析中,控制组群效应的方法是把要分析的变量(例如,家庭收入、家庭消费和家庭储蓄率)进行分解,以控制组群效应的影响,同时也分解出年龄效应(age effect)以及年代效应(year effect)(Deaton and Paxson, 1993; Deaton, 1997; Gibson and Scobie, 2001)。在对家庭消费和家庭财富的基于组群效应的分解中,大多数实证结果都得出一个倒“U”形的家庭储蓄率或者家庭财富持有的生命周期模式(Shorrocks, 1975; Attanasio *et al.*, 1999; Jappelli, 1999; Fernández-Villaverde and Krueger, 2002)。

### 三、理论模型:基于组群效应分解

尽管当前关于组群分解的实证研究已经不少(Deaton and Paxson, 1993; Deaton, 1997; Gibson and Scobie, 2001; Attanasio *et al.*, 1999; Jappelli, 1999; Fernández-Villaverde and Krueger, 2002),但是这些实证研究均没有给出一个理论模型来解释分解工作的经济原理。本文将利用一个理论模型给出

<sup>3</sup> 合成面板数据的构造是根据每年调查的新样本,并不依赖于样本是否具有面板数据特征,因此,样本损失问题不存在。此外,使用样本均值的方法可以降低奇异观测个体的影响,使得测量误差减少。

解释，它也是计量设定的理论基础。

对于家庭收入的分解，本文从分析个体的工资着手。根据经济学理论，个体的工资是由个体的人力资本决定的（包括个体的教育背景、工作经历等方面的因素）。此外，从动态的角度来看，个体的工资随经济增长或者社会生产率的提高（用每个劳动力的GDP来衡量）而提高。结合以上两点，我们可以把个体在给定年份的工资定义为给定年份的社会生产率和个体的人力资本的乘积：

$$y_{i,t}(h_{i,t}) = p_t \times g_t(h_{i,t}), \quad (1)$$

其中， $y_{i,t}$ 是个体 $i$ 在 $t$ 年的工资； $p_t$ 是 $t$ 年时的生产率； $h_{i,t}$ 是一个变量向量，以衡量个体 $i$ 的人力资本（例如性别、教育背景、工作经历等）以及其他影响个体工资的因素（例如个体所在企业的特征和所在的地区）； $g_t(h)$ 是 $t$ 年的工资决定方程。我们可以把社会生产率定义为所有劳动力的加权平均工资。这样，如果我们假定 $t$ 年时人力资本的概率密度函数为 $f_t(h)$ ， $t$ 年的社会生产率可表示为

$$p_t = \int_{\Omega_{i,t}} y_t(h) \times f_t(h) dh. \quad (2)$$

把方程（1）代入方程（2）中并消去 $p_t$ ，可以得出如下恒等式：

$$1 = \int_{\Omega_{i,t}} g_t(h) \times f_t(h) dh, \quad (3)$$

方程（3）表明工资决定方程 $g_t(h)$ 的形式内生地由人力资本分布状况 $f_t(h)$ 决定。

我们把生产率的增长定义为

$$(p_{t+1} - p_t)/p_t = r_{t+1}, \quad (4)$$

其中， $r_{t+1}$ 是社会生产率在 $t+1$ 年的增长率。为了分析方便，我们假定在劳动力市场没有发生显著的结构变化（例如劳动参与率以及平均每个劳动力的劳动供给）的条件下， $r_{t+1}$ 视为与该国的GDP的增长率相等。方程（4）的连续形式可以写为

$$p'(t) = p(t) \times r(t). \quad (5)$$

解此微分方程，

$$p(t) = p(t_0) \times \exp \left[ \int_{t_0}^t r(t) dt \right], \quad (6)$$

其中 $t_0$ 是起始年份（或者是参考年份）。方程（6）表明， $t$ 年的生产率， $p(t)$ ，可以分解为两部分：其一， $t_0$ 年时的生产率， $p(t_0)$ ；其二， $t_0$ 年至 $t$

年的生产率增长,  $\exp\left[\int_{t_0}^t r(t)dt\right]$ 。这样, 根据方程(6)和方程(1), 工资方程可改写成个体年龄和出生年份的函数:

$$y(a_i, c_i) = p(c_i) \times \exp\left[\int_{c_i}^{c_i+a_i} r(\tau)d\tau\right] \times g[h_i(a_i, c_i)], \quad (7)$$

其中,  $c_i$  为个体  $i$  的出生年份,  $a_i$  为个体  $i$  的年龄, 其对数形式为

$$\text{Ln}[y(a_i, c_i)] = \text{Ln}[p(c_i)] + \int_{c_i}^{c_i+a_i} r(\tau)d\tau + \text{Ln}[g(h_i(a_i, c_i))]. \quad (8)$$

由方程(8)可知, 个体的工资可以被分解为两个部分:  $\text{Ln}[p(c_i)]$  是随年龄不变的部分 (age-invariant part), 只与个体的出生年份有关;  $\int_{c_i}^{c_i+a_i} r(\tau)d\tau + \text{Ln}[g(h(a, c))]$  是随年龄变化的部分 (age-variant part)。其中,  $\int_{c_i}^{c_i+a_i} r(\tau)d\tau$  是从  $c_i$  年到  $c_i+a_i$  年之间社会生产率的增长;  $\text{Ln}[g(h(a_i, c_i))]$  表示出生于  $c_i$  年的个体随年龄变化的生命周期特征。一般来说, 个体在工作年龄的时候赚取劳动收入, 而到退休的时候赚取养老金收入。对于个体的养老金收入, 它一般与个体退休前的劳动收入水平相关。因此, 这个分析框架可以扩展到对养老金收入的分析。

家庭收入是家庭成员收入的总和, 包括劳动收入和养老金收入。为了分析方便, 本研究把家庭收入定义为户主的收入与一个转移因子的乘积。转移因子定义为家庭有收入成员 (包括依靠工资收入的家庭成员和依靠养老金收入的家庭成员) 的数目与家庭有收入成员的平均收入与户主收入的比率的乘积, 其表达式为

$$d(a, c) = fe(a, c) \times rw(a, c), \quad (9)$$

其中  $d(a, c)$  为转移因子;  $fe(a, c)$  为家庭有收入成员的数目;  $rw(a, c)$  为家庭有收入成员的平均收入与户主收入的比率, 它表示了户主在一个家庭中的经济地位以及家庭的收入结构。这样, 家庭收入可以被定义为

$$hy(a, c) = y(a, c) \times d(a, c), \quad (10)$$

$hy(a, c)$  为家庭收入。结合方程(8)和方程(9), 家庭收入可以写成如下形式,

$$\begin{aligned} \text{Ln}[hy(a_i, c_i)] &= \text{Ln}[p(c_i)] + \text{Ln}[fe(a_i, c_i)] \\ &+ \text{Ln}[rw(a_i, c_i)] + \int_{c_i}^{c_i+a_i} r(t)dt + \text{Ln}\{g[h(a_i, c_i)]\}. \end{aligned} \quad (11)$$

在方程(11)的右边,  $\text{Ln}[p(c)]$  是由家庭户主的出生年份决定的, 不随时间变化, 被定义为组群效应。方程(11)右边的其他四项均是随年龄变化的。此外, 考虑到人口转型问题,  $fe(a, c)$  的年龄曲线在各组群间存在差异,

可以被视为组群变量和年龄变量的交叉项。把最后三项合并为一项，定义为  $hw(a_{i,t}, c_i)$ ：

$$hw(a_i, c_i) = \int_{c_i}^{c_i+a_i} r(\tau) d\tau + \text{Ln}[rw(a_i, c_i)] + \text{Ln}[g(h(a_i, c_i))]. \quad (12)$$

$hw(a_i, c_i)$  可以被定义为年龄效应 (age effect)。

在实证研究中，由于在给定一年中组群变量和年龄变量之间存在固定关系，利用单一的截面数据不可能进行上述分解工作。此外，由于样本存在时间序列特性，分解工作也涉及一个随年份变化的变量，我们把它定义为年代效应 (year effect)。年代效应反映的是短期外部冲击的影响（例如，利率以及通货膨胀率的短期变化），反映经济周期的影响。这样，结合方程 (11) 和方程 (12)，可以把家庭收入的分解方程写成如下形式：

$$\begin{aligned} \text{Ln}[hy(a_{i(t)}, c_{i(t)})] = & \text{Ln}[p(c_{i(t)})] + \text{Ln}[fe(a_{i(t)}, c_{i(t)})] \\ & + hw(a_{i(t)}, c_{i(t)}) + ys_t + \zeta_{i(t)}, \end{aligned} \quad (13)$$

其中，下标  $i(t)$  表明家庭  $i(t)$  在  $t$  年出现在样本中； $ys_t$  是年代效应； $\zeta_{i(t)}$  表示家庭  $i(t)$  的其他特征 (idiosyncratic term)，在回归分析中视为误差项。这样，根据方程 (13)，家庭收入可以被分解为组群效应、家庭有收入者的数目、年龄效应（包括生产率的增长、个体人力资本的生命周期特征以及家庭收入结构）以及年代效应。

类似的，在一些简单以及合理的假设下，家庭消费也可以分解为包括年龄效应、组群效应和年代效应的若干部分。首先，经济增长赋予不同组群的家庭不同的终生资源。我们可以用家庭收入的现值来衡量一个典型家庭的终生资源：

$$w(c_i) = \sum_{t=c_i}^T \frac{hy_i(t)}{(1+r)^{t-c_i}}, \quad (14)$$

其中  $r$  为利率。这样，一个家庭的终生资源可以表达为

$$w(c_i) = p(c_i) \sum_{t=c_i}^T \frac{d(t, c_i) \exp\left[\int_{c_i}^t r(t) dt\right] \times g_i(h_{i,t})}{(1+r)^{t-c_i}}. \quad (15)$$

根据方程 (15) 中的  $p(c_i)$ ，我们可以看到一个家庭的终生资源也具有组群特征。

其次，根据生命周期假说，家庭消费相对于家庭终生资源存在一个生命周期模式。这样，两者之间的关系可以通过如下方程表示：

$$hc(a_i, c_i) = w(c_i) \times p(a_i, c_i), \quad (16)$$

$p(a_i, c_i)$  为家庭消费的年龄变化曲线，它与人口变量相关。此外，考虑到人口

转型的存在, 家庭消费的生命周期模式应当在各个组群中有所不同。在理论模型中, 我们把  $p(a_i, c_i)$  分解为两个部分:

$$p(a_i, c_i) = fs(a_i, c_i) \times f(a_i, c_i), \quad (17)$$

其中,  $fs(a_i, c_i)$  为家庭人口数量, 它也是衡量人口转型最重要的变量;  $f(a_i, c_i)$  可以被定义为在控制了家庭人口数量之后的家庭消费的年龄曲线。这样, 方程 (16) 可以被修改为

$$hc(a_i, c_i) = w(c_i) \times fs(a_i, c_i) \times f(a_i, c_i). \quad (18)$$

与分解家庭收入方程类似, 我们也可以把方程 (18) 写成对数方程形式, 并加一个随年份变化的短期冲击变量  $cs_t$  (与方程 (13) 中的  $ys_t$  类似), 以及表明个体特性的变量  $\xi_{i(t)}$  (与方程 (13) 中的  $\zeta_{i(t)}$  类似)。这样, 家庭消费方程也可以写成如下形式:

$$\begin{aligned} \text{Ln}[hc(a_{i(t)}, c_{i(t)})] = & \text{Ln}[w(c_{i(t)})] + \text{Ln}[fs(a_{i(t)}, c_{i(t)})] \\ & + \text{Ln}[f(a_{i(t)}, c_{i(t)})] + cs_t + \xi_{i(t)}. \end{aligned} \quad (19)$$

这样, 家庭消费也可以被分解为四个部分: 与组群变量相关的部分 ( $\text{Ln}[w(c_{i(t)})]$ , 定义为家庭消费的组群效应)、家庭人口规模、随年龄变化的部分 ( $\text{Ln}[f(a_i, c_i)]$ , 定义为家庭消费的年龄效应) 以及与年份有关的变量 ( $cs_t$ , 定义为年代效应)。

关于家庭储蓄率, 常规的定义为

$$sr_i = 1 - hc_i / hy_i, \quad (20)$$

其中,  $sr_i$  为家庭储蓄率。在实证研究中, 如果储蓄率不高, 方程 (20) 可以被定义为如下形式:

$$sr_i \approx \text{Ln}(hy_i) - \text{Ln}(hc_i). \quad (21)$$

这样, 家庭储蓄率可以被定义为方程 (13) 与方程 (19) 的差值, 家庭储蓄率在控制了人口变量以后 (家庭收入者的数目以及家庭人口数目) 也可以被分解出组群效应、年龄效应和年代效应。

在以上的模型中, 尽管家庭收入的组群效应在数量上与生产率的增长是相对应的, 然而如果从代际之间的差异来讲, 组群效应包含其他内涵。例如, 出生于不同年代的个体, 由于在成长过程中社会环境的差异, 他们在风险态度、折现因子以及偏好方面有差异, 进而对消费行为产生影响 (Ryder, 1965)。年龄效应描述了个体消费行为随着年龄变化的特征, 这一特征在各个组群之间相似, 这是因为个体的人力资本在一生的变化模式以及一个家庭一生中的人口事件的发生顺序均具有明显的生命周期特征。年代效应被假定为衡量商业周期的影响 (例如通货膨胀以及利率), 可以被视为短期的冲击。

#### 四、计量设定

上述模型为计量设定提供了分析框架。以往的实证研究中涉及几种分解方法，而 Deaton 的方法由于其灵活性和便利性是被广泛采用的，本文因此采用该方法。Deaton (1997) 认为，当数据的样本量大的时候，最好还是让数据本身去决定各种效应的估计结果。<sup>4</sup>该计量模型为

$$Z_{i,c_{j-k+1},a_j,t_k} = \alpha_{c_{j-k+1}} + \beta_{a_j} + \delta_{t_k} + BX_{i,c_{j-k+1},a_j,t_k} + \varepsilon_{i,c_{j-k+1},a_j,t_k}, \quad (22)$$

其中  $Z_{i,c_{j-k+1},a_j,t_k}$  为被解释变量（例如家庭收入的对数、家庭消费的对数以及家庭储蓄率）； $\alpha_{c_{j-k+1}}$  为组群效应； $\beta_{a_j}$  为年龄效应； $\delta_{t_k}$  为年代效应； $X_{i,c_{j-k+1},a_j,t_k}$  为其他控制变量的向量变量； $B$  为  $X_{i,c_{j-k+1},a_j,t_k}$  的估计系数。

Deaton 分解方法涉及两个假设。第一个假设是，不存在年龄效应、组群效应以及年代效应之间的交互影响，假定这三种效应是对称的。应当说，该假设是很强的。因为在之前的理论模型中，我们可以看到年龄变量和组群变量共同出现在某些项中（例如  $fe$  和  $fs$  中），因而它们并非相互独立。在本文的计量设定中，我们考虑了代际之间人口转型对家庭收入和消费的影响，从而加入人口变量作为控制变量。实际上，考虑到人口转型，可以把人口变量视为组群变量以及年龄变量之间的交互影响项。本研究与其他研究相比的一个差异就是，其他研究的分解工作并没有把人口变量加入到控制变量中。

第二个假设关于识别问题（identification problem），这是因为组群变量、年龄变量和年份变量三者之间存在固定关系。本文采用 Deaton (1997) 的方法设定识别问题：第一，把增长归到年龄效应和组群效应之中；第二，年代效应被定义为周期性的波动或者商业周期效应，假定年代效应的均值为零并与时间趋势正交。实际上，Deaton 的设定与本文的模型设定是一致的。根据对识别问题所设定的约束条件，在计量分析中需要对时间变量进行转换。假定年代效应为  $\delta_t$ ， $t=1, 2, \dots, T$ ，那么 Deaton 所假设的年代效应与时间趋势的正交关系可以表示如下：

$$\sum_{t=1}^T \delta_t \times d_t = 0 \quad \text{和} \quad \sum_{t=1}^T \delta_t \times t \times d_t = 0, \quad (23)$$

$d_t$  为虚拟变量，当年代为  $t$  时它等于 1，反之为 0。根据方程 (23)，我们可以得出  $\delta_1$  和  $\delta_2$  的解为

<sup>4</sup> 正如 Deaton 所评论的那样，如果数据的时间序列太短，使用这种方法是非常危险的。这是因为，在一个短的时间序列中，很难把短期冲击所造成的波动和长期的趋势分开。只有当时间序列充分长时，才有可能把商业周期的因素从长期的时间趋势变动中分解出来。

$$\delta_1 = \sum_{t=3}^T (t-2) \times d_t \times \delta_t \quad \text{和} \quad \delta_2 = \sum_{t=3}^T (1-t) \times d_t \times \delta_t. \quad (24)$$

把  $\delta_1$  和  $\delta_2$  代入方程 (22), 可重新获得年代效应的系数  $\delta_t$ ,  $t=3, \dots, T$ , 我们就可以得出转换后的年代虚变量 (排除第一年和第二年, 从第三年到第  $T$  年):

$$d_t^* = d_t - [(t-1)d_2 - (t-2)d_1], \quad (25)$$

$d_t^*$  为修改后的虚拟变量。 $d_t^*$  的系数就是  $t$  年时的年代效应,  $t=3, \dots, T$ 。而第一年和第二年的时间效应可以通过方程 (24) 计算出来。

在计量分析中, 各组虚拟变量设定如下: (1) 组群虚拟变量排除了出生最早的组群 (被设定为组群虚拟变量的参考组); (2) 年龄虚拟变量排除了最年轻的年龄 (被设定为年龄虚拟变量的参考组); (3)  $T-2$  个年代虚拟变量 (根据方程 (25) 转换)。年龄效应、组群效应和年代效应就是对应的虚拟变量的系数。此外, 控制变量  $X_{i,c_j-k+1,a_j,t_k}$  针对不同被解释变量而不同, 具体指定如下: 当被分解的变量为家庭收入时,  $X_{i,c_j-k+1,a_j,t_k}$  为家庭有收入者的数目; 当被分解的变量为家庭消费时,  $X_{i,c_j-k+1,a_j,t_k}$  为家庭人口数目; 当被分解的变量为家庭储蓄率时,  $X_{i,c_j-k+1,a_j,t_k}$  包括家庭有收入者的数目和家庭人口数目。

## 五、数据及组群构造

### (一) 数据

本文使用的数据是由国家统计局收集的中国城市家庭调查数据, 时间跨度为 1988—2003 年。样本来自六个省区, 包括北京 (华北)、辽宁 (东北)、浙江 (华东)、广东 (华南)、四川 (西南) 和陕西 (西北)。这六个省区在地理上具有代表性。本文实证研究的分析对象分别为家庭收入、家庭消费和家庭储蓄率。根据 CUHS 的数据手册, 家庭收入被定义为家庭可支配收入, 它是家庭总收入减去个人收入所得税的差值。家庭消费是家庭生活费方面的支出, 包括食品、服装、交通、通信、娱乐、教育、医疗以及其他方面。除了这些变量, 计量分析还使用人口变量, 例如家庭有收入的人数和家庭人口数目。表 2 是 CUHS 数据在各年份以及在各地区的样本分布, 各时间截面上的样本量均超过 3000。在本文的计量分析中, 由于部分样本记录有误或者记录不完整, 所使用的样本量略有减少。



表2 各调查年份的样本量

年份	北京	辽宁	浙江	广东	四川	陕西	总计
1988	200	699	450	600	850	450	3 249
1989	100	700	450	650	550	550	3 000
1990	200	700	500	550	800	550	3 300
1991	99	750	500	600	850	550	3 349
1992	500	1 000	550	550	1 100	500	4 200
1993	500	1 000	550	600	735	500	3 885
1994	500	995	550	599	750	500	3 894
1995	500	1 000	550	600	750	500	3 900
1996	500	1 000	550	600	750	500	3 900
1997	500	1 000	550	600	750	500	3 900
1998	500	1 000	550	600	750	500	3 900
1999	500	1 000	550	500	750	500	3 800
2000	500	1 000	550	600	1 100	500	4 250
2001	500	1 000	550	600	650	500	3 800
2002	999	3 250	1 686	1 568	1 406	1 400	10 309
2003	995	3 600	2 133	1 524	1 649	1 450	11 351

## (二) 组群构造

遵循一般的方法，我们根据家庭户主的出生年代定义一个家庭的“组群”。在计算组群均值时，如果由组群和年份所定义的单元的样本量太小，那么所计算出的年龄曲线（age profile）有可能具有很大的波动性，从而偏离该组群实际年龄曲线。通常的做法是为每一个组群选择一个“带宽”，以增加每一单元内的样本数量。例如，对于生于1945年的组群，可以使用出生于从1944年到1946年的样本的均值作为出生于1945年的组群的均值。然而，带宽设置需要平衡单元内的观测个体的数量与组群内部的异质性（intra-cohort heterogeneity）之间的关系。一般来说，带宽设置越大，组群内部的异质性也越高，尤其是对处于经济高速增长的国家。本研究选择5年的带宽来计算每一个组群的年龄曲线。这样，每一个组群与其相邻组群最多有11年的交叠。

此外，实证研究应当注意样本选择问题（sample selection problem）。一般来说，户主是家庭的主要收入者以及家庭经济问题的主要决策者。在样本的采集过程中，有可能产生样本选择问题，尤其对那些户主非常年轻以及非常老的样本。一方面，富有的家庭户主相对于贫穷的家庭户主可能会寿命更长（Shorrocks, 1975）。而另一方面，从富裕家庭中出来的个体，其结婚的时间可能也较早。因此，在实证研究中，我们排除了那些户主非常年轻以及非常年老的观测个体，以减少样本选择问题。中国的人均预期寿命为71.8年。这样，本论文中的家庭户主年龄分布在25—71岁之间。另外，1988—2003年期间的样本中，最老的样本的家庭户主出生于1920年，1988年为68岁；最年轻样本的家庭户主出生在1974年，在2003年为29岁。在分解工作中，有