

中国城市居民收入分布的变化：1988-1999

万定山

北京大学中国经济研究中心

No. C2004003 2004年4月5日

中国城市居民收入分布的变化：1988-1999

万定山*

北京大学中国经济研究中心

2004年4月2日

摘要 本文主要运用微观计量的方法对1988-1999年间中国城市居民收入分布的变动进行分解，以定量地分析各种因素——人群禀赋和参数结构对收入分布差别的影响，同时对各种因素的相对重要性作出评估。对工资或所得方程的估计考虑了对选择性偏差可能带来的影响进行剔除。结论表明，工资或所得方程的参数结构对收入分布的影响很大，其中省份和教育虚拟变量的影响最为突出。人群不可观察特征之分布的变化对收入分布的影响不大。教育选择方程参数的改变对收入分布的影响非常大，其次是参与选择方程行为参数。但在先改变了教育方程参数的情况下，参与选择行为参数的改变给收入分布带来的额外影响不大。如果在1988年人们能以和1999年相同的方式（参数结构）获得非挣得收入，那么收入基尼系数也将下降很多。综合分析表明，主要是经济体制的激烈转型导致了收入分布的大幅变化。

关键词 收入分布，禀赋分布，参数结构，模拟

JEL 分类 C15, D31, I31, J13, J22

* 北京大学中国经济研究中心，100871；Email：dunning_wan@hotmail.com。感谢导师赵忠教授以及赵耀辉教授一直以来的悉心指导和关心，感谢赵耀辉教授、香港中文大学张俊森教授毫无保留的数据提供，同样感谢宋晓青的诸多帮助。同 CCER 劳动经济学小组诸位同学的讨论令我受益匪浅。文中一切错误都是我的。

Changes in Household Income Distribution in Urban China: 1988-1999

Dingshan WAN[†]

China Center for Economic Research
Peking University

April 2, 2004

Abstract: This paper combines microeconomic and micro-simulation methods to study changes in household income distribution in urban China between 1988 and 1999. We are aimed to explore quantitative effects of various factors on incomes distribution, primarily personal endowments distribution and structural parameters. After controlling the selection bias, we find that the changes in structural parameters of wages or earnings equations greatly influence income distribution, among which the parameters of province and education are the most important ones. Changes in distribution of personal unobservable characteristics have little effect. Parameters of both occupation choice and education choice equation change income distribution greatly, especially the later. However, the former causes little additional changes in income distribution if the later has been changed first. Furthermore, if people in 1988 can get unearned income in the same way as in 1999, the Gini coefficient of income will be decreased by a large extent. Based on our analysis, we believe that the dramatic transition of economic system has caused the great changes in income distribution.

Keywords: Income Distribution, Endowments Distribution, Structural Parameters, Simulation

JEL Classification Codes: C15, D31, I31, J13, J22

[†] China Center for Economic Research, Peking University, Beijing 100871, China. .Email: dunning_wan@hotmail.com. I would like to thank Zhong Zhao and Yaohui Zhao for their persistent guidances, and Yaohui Zhao and Junsen Zhang for their generosity in offering all the data. Xiaoqing Song has helped me a lot too. And undoubtedly I have benefited a lot from discussions with members of labor group at CCER. All errors in the paper are mine.

一、引言

20 多年来的中国经济改革无疑已经取得了重大的成功。这不仅表现在一系列的定量经济指标如 GDP 总额、GDP 增长率等上，更重要的是体现在各种制度的建立、规范和逐步完善上——正是这些制度构筑了现代市场经济正常运行的基础。改革的一个最重要的结果是，我们的社会变得更加开放，多样性的存在被得到允许。

与改革所取得的巨大成就相伴随的，是中国个人收入分布的显著变化。这也许是允许多样性存在的一个必然结果，因为人们的禀赋本就迥异。但个人收入分布的这种显著变化还是引起了广泛的关注，无论政府或是学术界。其中缘由是显而易见的，过度的收入分配不平等会成为社会不稳定的潜在根源，进而阻碍甚至严重损害经济发展；即使从纯粹经济学的角度来看，收入分配与经济增长的关系也是一个很重要的问题。对此，首先需要的是对不平等这一问题本身的研究，人们感兴趣的问题包括，目前的收入不平等程度到底有多大？改革开放以来收入分布的变化趋势及其具体时间路径（timing）是怎样的？是什么因素导致了收入分布的变化等等。本文希望能够利用现有的数据，对这些问题作细致的分析。与前文相比，本文的不同之处在于：

在方法论上，本文采用了简化式（reduced form）模型加微观模拟（simulation）的方法，所有与家庭收入决定相关的计量模型都以一个关于单个家庭收入决定的理论模型为基础，这一理论模型是一般均衡性质的，模型中的单个家庭收入由全部家庭成员的收入所决定，而每个成员的收入则被视作是一系列个人选择行为的结果。这一方法的好处在于，对于整个收入决定流程中的各种感兴趣因素——主要是人群禀赋和各种参数结构，我们都可以在保持其他各种相关因素不变的情况下（*ceteris paribus*），单独考察某个因素对两个收入分布差别的定量影响，从而也就可以评估各种因素的相对重要性。这一方法来自 BFL（2002），与他们不同的是，在对工资或所得方程的估计中，我们考虑了用 Lee（1983）的方法来对选择性偏差可能带来的影响进行剔除，

微观模拟的结果表明，工资或所得方程的参数结构对收入分布的影响很大，其中省份和教育虚拟变量的影响最为突出。在即使控制了人群个人特征分布及其回报率结构差异的情况下，地区效应对不平等程度变化的贡献仍然令人瞩目，这一点值得我们的关注。人群不可观察特征之分布的变化对收入分布的影响不大。教育选择方程参数的改变对收入分布的影响非常大，其次是参与选择方程行为参数。但在先改变了教育方程参数的情况下，参与选择行为参数的改变给收入分布带来的额外影响不大。如果在 1988 年人们能以与 1999 年相同的方式（参数结构）获得非挣得收入，那么收入基尼系数也将下降很多。

文章的结构如下:第二部分是关于改革开放以来中国收入分布变化研究的一个简要文献综述。第三部分是 1988-1999 年间中国城市居民收入分布变化的总体趋势描述。第四部分介绍本文所用的微观计量模型,包括一般性框架、模型的设定和具体估计方法。第五部分是对模拟方法和步骤的说明。第六部分是对收入分布模拟所得结果的分析。最后一部分是对文章的小结,包括基本结论、存在的问题、可以扩展的研究方向、以及政策涵义等。

二、相关文献

迄今为止,关于中国收入不平等问题的讨论已经相当多,我们主要关注这方面的经验研究。就经验研究而言,很重要的一点是,数据本身的质量将直接影响到研究结论的质量及可信度。由于在目前的中国普通研究者比较难以获得高质量的微观数据,因此不少相关研究都是以加总数据为基础来展开分析的,这大大限制了研究的范围及深入程度,以这类数据为基础进行的分析通常都局限于地区间收入不平等或是城市与农村间不平等问题。进入 90 年代以来,国际上流行的趋势是,以家计调查微观数据为基础的经验研究越来越得到广泛的认可,¹这主要是因为两个方面的原因:首先家户调查微观数据包括了更多关于决策者个人行为的信息,其次近几十年来微观计量理论获得了迅速发展,这些都使得更深入的研究成为可能。就中国的实际情况而言,由国家统计局的两支调查总队所分别进行的城市和农村家计调查已经积累了丰富的微观数据,利用这些数据也已产生了一些研究成果,这些研究主要是由国家统计局的一批研究者所作出的。由于中国的官方调查数据不对普通研究者开放,因而应用这些数据可能进行的研究拓展也受到了很大的限制,这无疑非常令人遗憾。简单的理由在于,研究者的分散化将提高研究的质量和效率,因为独特的知识分散于我们每个人的心灵,这也是上个世纪的社会主义大论战中哈耶克用来反驳社会主义理论的主要知识论依据。在中国收入分配问题的研究上,另一支作出了令人瞩目贡献的研究队伍来自中国社会科学院经济研究所,以及英国、美国和瑞典一批对中国问题感兴 68

趣的经济学家,鉴于对中国官方数据质量的深刻怀疑,他们甚至不辞艰苦分别在 1988、1995、1999 三次组织了独立于国家统计局的家计调查,以取得真实的一手资料。²他们的研究成果大都收集在两本专著中。³国内学者中其他比较有影响的研究包括陈宗胜(1997,2000,

¹ Deaton (1997) 是关于家户调查数据搜集及分析方法的一个全面介绍。

² 对两套数据各种指标内涵的详细比较、以及基于各自数据所得基本结论的差别,详见卡恩、李思勤(1999)。

³ 赵人伟、格里芬,《中国居民收入分配研究》,中国社会科学出版社,1994;赵人伟、李实、李思勤,《中国居民收入分配再研究》,中国财政经济出版社,1999。

2002) 等。⁴

由于数据上的限制，本文主要分析中国城市居民收入不平等的问题，⁵因此下面的文献回顾也主要关注这方面的研究。⁶

1、城市居民收入分布：总体趋势

中国的城市经济改革始于 80 年代中期，至 90 年代初中期为一段落，一般认为从 90 年代中后期开始更为激烈的改革措施得到推行。与此对应，有关中国城市居民收入分布变动的研究文献也主要关注这两个不同时段的情况。如果把不平等问题的研究看作是对改革政策的评估，那么这一评估感兴趣的问题至少应包括两个方面：改革的结果怎样地在不同的人群间进行分布，以及改革是否使得社会中绝大多数人群的福利增加。

对于城市改革以来居民的收入分布变动总体趋势，学者们进行了很多研究。Ahmad 和 Wang (1991) 利用 10 省城市住户调查数据，计算得到 1987 年收入分布基尼系数为 0.26。国家统计局的计算结果则表明，1988、1995 年城镇居民货币收入基尼系数分别为 0.23、0.28 (任才方、程学斌，1996)。⁷社科院计算得到 1988 和 1995 年城镇家庭人均可支配收入基尼系数分别为 0.233 和 0.332 (卡恩、李思勤，1999)。各研究结论在具体数值上稍有差别，这可能部分地源于他们定义了不同的收入指标，⁸但反映的基本趋势却相同。一个基本的定性结论是，自 80 年代中期改革起步以来，中国城市居民收入分配不平等程度一直呈上升趋势，而且这种趋势在进入 90 年代以后有所加剧。

与总体趋势同样重要的问题是各人群的绝对变化情况。研究表明 (赵人伟、李实，1999；Meng, 2001)，在 1988 至 1995 年间，改革为所有人带来了福利的改善，尽管程度不一，但进入 90 年代后期以来，处于收入分布最低 5 个百分点 (percentile) 的人其境遇绝对地变坏了。这些研究也为目前国内一些学者的观点提供了佐证，他们认为，中国城市的经济改革以 90 年代中期为分水岭，在此以前的改革是帕累托改进式的，而此后的改革则更大程度上是财富的重新分配。

2、城市居民收入分布：变动的原因

⁴ 陈宗胜的研究大部分都是基于天津市的数据所作出的，这限制了他的研究结论的意义。

⁵ 本文对城市 (urban) 和城镇作了区分，因为文中用到的数据来自城市家户调查，而非城镇。

⁶ 李实 (2003) 对有关中国个人收入分配的研究文献作了较详细的综述。

⁷ 文章没有对用来计算基尼系数的收入指标作详细说明，这使得我们无法比较各研究所使用收入指标内涵差别所在。

⁸ 社科院研究者们采用的个人可支配收入与国家统计局使用的人均收入之不同主要在于，前者不仅包括了各种货币性收入，而且还包括了房屋实物补贴、其他净补贴、自有房屋的估算租金价值，详见卡恩、李思勤 (1999) 的讨论。

这方面的文章较多。⁹就内容而言,各种分析中所讨论的主要贡献因素包括:非国有部门的快速发展、企业改革与改制(国有企业改制、企业内部分配体制的改变);城镇居民收入的主要来源是工资,职工工资分配的不平等程度的变化会直接影响收入的分布;对城镇职工工资函数的估计可以发现职工个人特征与工资差异之间的关系,比较引人注目的是城镇职工性别差异的扩大,教育水平、技术水平等人力资本因素对工资分布变动的的影响;行业和所有制差距;农村劳动力进城;财产(包括金融资产和自有房产);行业垄断等等。赵人伟、李实(1997)将各种可能对收入分配产生影响的因素划分为三类:经济增长和发展、经济改革和体制变迁、经济政策及其变化。此外,学者们对城市居民收入地区间差异扩大的问题也进行了很多研究。Knight、赵人伟和李实(1999)利用1988和1995的抽样调查数据,对城市居民收入分布的地区差异进行了分析。林毅夫、蔡昉和李周(1998)则基于以省、直辖市和自治区为观察值的加总数据对地区收入差距问题进行了初步研究。

以上这些研究基本上只是描述性的。从研究方法来看也大同小异:利用收入分布不平等测度指数对几个感兴趣时间点的收入差距进行分解,比较不同时间点上的分解结果,以考察各种因素对收入分布变动的贡献。通常用到的工具是基尼系数和 Theil 指数。

Meng(2001)使用了社科院1988、1995和1999的三次调查数据,但她在方法论上前进了一步,她借鉴了Fields(1998)的分解方法,其对Shorrocks(1982)的改进在于,对收入的分析是基于回归方程的,因而可以运用计量工具部分解决上面简单方法面临的困境。她试图对两个不同改革阶段上城市居民收入分布变化的本质及原因作出区分:1988-1995以及1995-1999,虽然在这两个阶段收入分布都变得更加不平等,但在第一阶段所有人群都绝对变好了,而地区差异的扩大是导致不平等程度上升的主要因素;在第二阶段处于收入分布最底端5%的人群其状况绝对地变坏了,使不平等程度扩大的主要原因是以企业破产、下岗增加为特征的经济重构。Meng(2001)所讨论的问题与本文有相近之处,后文还将作详细比较。

奈特、宋丽娜(1999)运用了与Meng(2001)类似的方法对中国城镇工资差距扩大的原因进行分解,通过对社科院1988和1995调查数据的分析,他们发现人力资本变量是能够识别的最重要的非均等化因素,其次是省份变量。

这些文献是对中国城市收入分布变动的初步分析,他们的一个共同不足在于缺乏一个统

⁹ 古斯塔夫森、李实(1999), Knight and Li(2001), 赖德胜(1999), Li(2000), 李实、李文彬(1994), 李实(1997), 李实、魏众、古斯塔夫森(2000), 李实、赵人伟(1999), 李实、赵人伟和张平(1999a、1999b), Meng(2001), 奈特、宋丽娜(1999)等。

一的框架以分析个人或家庭收入的决定过程，这将是本文所要解决的问题。

三、中国城市居民收入分布变化的总体趋势

1、数据

本文所用的数据来自国家统计局城调队 1988-1999 年间的城市家户调查 (Urban Household Survey, or UHS)。由于从 1992 年起这一调查只涵盖北京、辽宁、浙江、广东、四川和陕西六省，为了保持各年数据之间的一致性 (consistency)，我们只使用了这六省 12 年的数据。这确实减少了样本量，但并不降低样本的代表性，同样，样本量的这种选择对于一些基本统计量的结果影响也不大，这一点将在后面提到。关于数据的进一步说明，详见附录。

2、1988-1999 年间总体趋势

2.1 消费还是收入？

近年来，人们开始对以收入数据为基础来估计不平等程度的做法提出越来越多的怀疑和批评。对不平等程度进行估测的主要动机在于我们拥有这样的信念，即社会中每个人的福利状况都是同等重要的，应该得到关注，但收入——尤其是经验研究中所经常使用的当期收入指标，是否可以真实地反映出社会成员的福利状况呢？显然，以当期收入作为福利状况的代理 (proxy) 变量至少在两个很重要的方面存在严重不足：如果当期收入的变动中有相当大部分是由于收入的暂时性变动，而非永久性变动所导致，那么当期收入的变动就不能很好地反映当事人福利状况的变化，¹⁰在这种情况下，以当期收入来判断当事人一生所能支配的经济资源显然是不可靠的。另外，如果存在一个较发达的正式或是非正式的信贷市场，那么当事人可以通过借贷行为来平滑 (smooth) 自己的消费。¹¹对当期收入指标一个比较好的替代是当期消费指标，从经济学的基础知识中我们知道，作为消费者决策行为的预算约束条件，收入与价格进入间接效用函数 (indirect utility function)，而消费作为当事人决策的结果进入直接效用函数 (direct utility function) 因此从理论上说消费指标是个人福利状况更好的近似。事实上，国外已经有不少学者开始尝试从当期消费的角度去重新看待不平等程度，¹² Cutler 和 Katz (1991a, 1991b) 对美国的研究表明，用收入和消费计算的不平等程度的变化趋势基本一致，然而，Slesnick (1993, 2001) 却发现，两者并非一致，用消费数据计算得到的不平等程度不仅在绝对量要小，而且其在时间上的变动并不显著。因此，在这一点上还远未

¹⁰ 这体现了弗里德曼持久收入假说的精神。

¹¹ Krueger and Perri (2002)。

¹² 相关文献有：Cutler and Katz (1991, 1992)，Johnson and Shipp (1991)，Mayer and Jencks (1993)，Slesnick (1993, 2001)，Deaton and Paxson (1994)，Dynarski and Gruber (1997)，以及 Blundell and Preston (1998) 等。

取得共识。

虽然从消费的角度来重新看待不平等问题确是一件很有意义的事情,尤其是在国内目前尚未看到这方面研究的情况下,但本文还是准备从收入指标的角度进行分析。这样做的主要原因在于,对消费中最重要的耐用消费品和住房部分,折算各期消费流比较困难,这里既有数据可得性的问题也有方法的问题,因此要作出令人信服的估计不那么容易。这也许是另一篇文章需要解决的问题。

另外,就评估不平等程度所应依据的概念基础而言,也有学者提出应当对不平等概念的丰富内涵重新评估,有影响的包括森(Sen)所倡导的“作为能力剥夺”(the deprivation of capabilities)的贫困概念等。¹³

2.2 怎样的收入指标?

对收入概念的界定本身也存在着不少争议。在我们使用的1988-1999共12年间的UHS数据中,前四年个人收入的指标是生活费收入,从92年起变为个人可支配收入,二者都只包含了个人得到的货币收入,因此对城市居民的实际收入会有一定程度的低估,这也是一些研究者对官方数据质量感到怀疑进而定义新的收入指标的原因所在。低估的部分主要是城市居民得到的各种福利如医疗、住房、教育、养老等,但数据上的这一缺陷应不会对不平等程度的估测造成严重偏离,因为各种福利的分配主要是以等级制度为基础进行的,这种分配本身带有比较多的平均主义色彩,所以对不平等程度这样的相对测度指标应该不会有很大的影响。当然,如果我们感兴趣的问题是城乡居民之间的收入分布差别,或者全国范围内的居民收入分配差别,那么就应该对数据的低估问题给予足够的重视。另外,直观的感受使我们有理由相信,十几年来城市居民所享受福利的不平等变化没有其他收入的不平等变化大,因此忽略这部分收入将高估收入的不平等程度。

在进行的几次全国住户收入抽样调查中,中国社会科学院的研究者们定义了与国家统计局不同的收入指标,主要是将房屋实物补贴、其他净补贴收入以及自有房屋的估算租金也包括在内,对租金价值的计算采用居民本人的主观估计值。¹⁴正是基于这样的估计,他们认为在1988-1995年间城镇住房的私有化导致了城镇收入分配状况的急剧恶化(卡恩、李思勤,1999)。但这一结论值得怀疑,一个很实际的理由是,由这种估算的主观性所带来的偏差未必就小于原数据的缺陷所导致的偏差。对收入指标内容的详细说明请见附录一。

2.3 总体趋势

¹³ 感兴趣的读者可以参见森(2003)对这一问题的简要评述。

¹⁴ 房屋租金归算本身即是一个很重要的但并未取得一致的研究课题。

迄今为止，经济学家和统计学家们已经发展出了多种对收入不平等程度进行测度的工具。表 1 给出了 1988-1999 年间中国城市居民收入不平等程度的计算结果，计算的依据是家庭人均月可支配收入，¹⁵这一指标是名义的，未考虑通货膨胀的影响，¹⁶事实上不平等指标本质上是一个相对的概念，通货膨胀应该不会对它们产生影响。同时，我们选取的样本是所有样本，未作统计上的样本剔除。¹⁷除了传统的基尼系数外，还包括 Theil (0)指数、变动系数等，其中平均对数偏差 (mean log deviation)就是广义 Theil (0)指数，而变动系数 (coefficient of variation)平方的一半等于广义 Theil (2)指数，Theil (a)系列指数的一个特点是，a 越大该指数对于分布顶端的差异的敏感性越大，a 越小则对于分布底端的差异敏感性越大。通过对不同指数的观察，我们可以得到收入分布中不同部分人群收入的相对变动情况。¹⁸

通过表 1 可以看到，1988-1999 年间中国城市居民收入不平等程度呈明显上升趋势，12 年间基尼系数上升了约 12 个百分点，如果以 1988 年为基底，上升幅度达到 50%还要多，这个速度是很快的。进入 90 年代以来，不平等程度的变化速度有所上升，92、93、94 三年间基尼系数上升了近 8 个百分点，占 12 年间全部上升值的三分之二。同时，94-95 年间基尼系数有所下降，这与国家统计局公布的结果是一致的。¹⁹表 1 中其他各指标所反映的事实与基尼系数基本相同。

值得指出的是，Park 等 (2003) 采用同样的数据对城市居民个人工资分布变动的分析得到了相似的结论，1988 至 1999 年间城市居民个人工资分布变化的时间路径与收入分布完全相同，两年的工资分布基尼系数分别为 0.244 和 0.359。与我们的家庭人均收入指标相比，他们的工资指标中扣除了财产性收入和转移收入，因此，这两部分收入对收入分布趋势变化的影响应该是中性的。

上述总体指数反映的是全部样本人群的情况，但显然每个人在这一期间的状态变化可能都各不相同。为了进一步分析 1988-1999 年间收入分布中不同百分点 (percentile) 人群状况的相对变动，需要对总体不平等指数作更细致的分解。表 2 给出了 1988-1999 年间中国城

¹⁵ 家庭人均月可支配收入=家庭全年总可支配收入/(家庭人口数*12)。UHS 数据由两部分组成：家庭总数据及家庭个人数据，其中每个家庭最多可以包含有八个该家庭成员的个人信息。家庭全年总可支配收入在数据中有现成对应项，家庭人口数也有，在对数据的检验过程中，我们发现有小部分家庭接受调查成员数与家庭期末人口数指标不符 (包括大于或小于)，为防止谬误，对这一部分样本作了删除。

¹⁶ 对各年收入指标进行通货膨胀调整于结果的影响不大，所得到的不平等程度变动的趋势及具体选择 (timing) 大致相同，但各年结果在数值上有所上升。同时，对 88-91 年间利用所有样本计算得到各相应的指标，与只利用六省样本计算得到的结果相比，均有小幅下降，但并不改变我们的基本结论。关于通货膨胀调整所使用价格指数数据及方法，详见附录数据说明。

¹⁷ 习惯的做法是依照某一标准对样本进行删节，收入在标准值以下的样本被删掉。这样做是为了避免分布的一些极端值对分布的过度影响，不过本文对这种做法持保留意见。

¹⁸ 关于收入分布各种测度指标的性质和解释，参见 Cowell (1995) 或 Kakwani (1980)。

¹⁹ 国家统计局公布的 1994 和 1995 年城市居民基尼系数分别为 0.30 和 0.28

市居民收入不平等程度的变化按分位数组 (quantile group) 计算的结果, 分位数组是收入分配研究中经常用到的一个概念, 它可以被视作是离散形式的洛伦茨曲线 (Lorenz curve)。²⁰

从表 2 可以看出, 在这 12 年间只有处于收入分布最顶端三个分位数组中的人群其相对状况变好了。与 1988 年相比, 到 1999 年为止, 只有 8、9、10 三个组群中的人其收入和占全部收入的比重上升, 而且越位于顶端的人群上升的幅度越大, 分别是从 11.86% 上升到 12.49%, 从 13.82% 上升到 15.31%, 以及从 19.88% 上升到 26.12%, 这说明收入分布的分化比较严重。对境况相对变坏的人——尤其是位于分布最底端的三组人群而言, 93 和 94 年是收入分布变动最大的两年, 这一结论与观察总体指数得出的结论一致。位于收入分布中间的几个组群从总的来看状况恶化了, 但变化的趋势比较缓和, 未发生大的跳动。²¹

到目前为止, 我们对 1988-1999 年间中国城市居民收入分布变动的大致情况已有所了解, 结论与相关的研究基本相同。在得到基本的现象之后, 一个更重要的问题是, 究竟什么因素导致了收入分布的变动? 传统的对收入分布变动原因进行分析的方法是使用可以分解的不平等测度 (measure), 可分解性 (decomposability) 是评价不平等测度功用的一个很重要的标准 (Sen, 2003)。²²可以证明, 如果加上“良好”性质的约束, Theil 指数及更广义的 Theil 系列指数是唯一满足可分解性的一类不平等测度 (Shorrocks, 1980, 1984)。Theil 指数的特殊性在于, 它可以以子人群为基础进行分解。而基尼系数的一个有用性质则在于, 利用它可以将总收入差距表示 (分解) 为各分项收入差距的函数, 从而对收入分布的变动从构成结构方面进行分析。Kanbur 和 Zhang (2001) 使用 Theil 指数对中国的加总数据进行了分析。这些简单的方法和工具确实有一定的功效, 但它们都存在着方法论意义上的缺陷: 首先, 对某一因素使用广义 Theil 指数进行分解时, 不能控制住其他因素, 这使得我们无法得到对几个因素相对重要性的定量估计。其次, 就某些因素而言, 这一方法不能有效区分某一因素对收入分布的影响到底来自该因素自身分布的改变, 还是该因素所对应的回报 (参数结构) 的改变。²³为解决这些问题, 需要使用更细致的分解方法。本文大体上沿用了 Bourguignon、Ferreira、Leite (2002) (下面简称 BFL) 的方法, 但在有一点上作出了改进, 下面会详细说明。

²⁰ 参看文献同上 12。

²¹ 这里同样采用了未经通货膨胀调整的家庭人均月可支配收入数据作为计算的基础。

²² 可分解性意味着, 如果将赚取收入的人群样本分解为几个子人群, 那么这一人群样本的不平等测度能够被表示为子人群组内 (within) 与子人群之间 (between) 对应不平等测度的某种形式的和。(Bourguignon, 1979)

²³ BFL (2002)。

四、微观计量的分解：一般性框架、模型设定和估计

BFL (2002) 使用的分解方法直接来自于 Bourguignon、Ferreira、Lustig (1998) 的世界银行研究建议。这一研究项目的主要目的，是对东亚和拉丁美洲收入分布的动态变动进行微观分析，整个研究的理论基础是一个关于单个家庭收入决定的一般均衡模型，分析所采用的技术是建立在微观计量基础上的模拟方法。这一方法实质上是著名的 Oaxaca-Blinder 分解公式的推广。²⁴我们知道，到目前为止，在关于工资或所得 (earnings) 分布差别的经验研究中，各种各样的分解方法已经得到了广泛的应用，虽然具体的方法各有不同，包括参数方法、半参数或非参数方法、核估计 (kernel estimator) 方法、以及风险函数方法 (hazard function) 等，²⁵但隐含在这些方法背后的思想却是统一的，即运用模拟的方法得到感兴趣的反事实的 (counterfactual) 分布。虽然以上这些方法不能简单地被运用于关于收入分布差异的研究，但其背后的思想却是可以被借鉴的。与工资或所得 (earnings)²⁶相比，收入 (income) 的决定显然要更复杂一些，这源于收入这个概念本身包含了更多的东西，每个经济当事人的收入都应被看作是均衡的结果。在各种既定的约束条件下，理性的当事人将其有限的资源——时间、精力等有效地分配于各个竞争性 (competing) 的选择之间，以取得一定的收入，在给定的市场条件 (价格等) 下这一收入可以为当事人带来最大化其效用的消费束。“收入是一连串的事件。”(菲歇尔, 1999) 就收入分布的研究而言，当事人的参与选择行为应该被内生化的，如果考虑的是家庭收入的决定，那么同时家庭结构也应该被看作是选择的结果。相比之下，在对工资分布差异的研究中我们只需考虑工资本身。

1、分解方法：一个一般性的框架

我们的目标是要比较两年的收入分布，为此可先考虑收入 y 和家庭可观察特征 T 的联合分布 $F(y, T)$ ，于是收入的边缘分布可以表示为： $f(y) = \iint F(y, T) dT$ 。²⁷

具体来说，考虑某个家庭在某一时期的收入决定函数，其在第 t 时期的收入可被看作是由他的禀赋和一系列参数结构所决定的，表示为：

$$y_{it} = Y \left[w_{it}, v_{it} (w_{it}, m_{it}; g_t), e_{it}; I_t; b_t \right] \quad (1)$$

²⁴ 参见 Oaxaca (1973) 和 Blinder (1973)。

²⁵ 参数方法如 Juhn、Murphy、Pierce (1993)，半参数方法如 Dinardo、Fortin、Lemieux (1996)，核估计方法如 Dinardo 等 (1996)，风险函数方法如 Donald 等 (2000)。

²⁶ 因个体经营者的净收益也包括在后面定义的挣得收入中，因此“工资”的说法并不准确，英文中的 earnings 包括 Salary or wages、business profits、gains from investments 三项，因此要略准确一些。

²⁷ 参见 BFL (2002)。

这里 y_{it} 表示第 i 个家庭在第 t 时期的收入， w_{it} 表示家庭可观察特征中完全外生的部分如地区、成员性别等， v_{it} 则表示家庭可观察特征中非完全外生的部分如成员教育程度等， v_{it} 被看作是家庭成员选择的结果，它由完全外生的禀赋 w_{it} 、结构性选择参数 g_t 以及随机扰动项 m_t 等一起决定，之所以要对部分家庭可观察特征内生化的，主要是为了后面收入分布差别的分析，这样各收入分布之间的差别可以归结为一些完全外生的变量以及一系列结构参数的差异。 e_{it} 代表不可观察的家庭特征， l_t 是参与选择的参数结构，而 b_t 则是工资方程的参数结构，这些变量都是向量。可观察到的与不可观察到的家庭特征，加上随机扰动项以及全部参数结构等一同决定了家庭的收入。相应地，第 t 时期收入的分布可表示为：

$$D_t = D_t [w_t, v_t(w_t, m_t; g_t), e_t; l_t; b_t] \quad (2)$$

这里 w_t 、 v_t 、 m_t 、 e_t 是由 t 时期收入分布中所有人的相关向量构成的矩阵。

从上面的表达式出发可以对两个不同的收入分布进行全面的比较。具体来说，两个收入分布的差别的产生可能来自于两个方面：禀赋分布的差异以及参数结构的差异，如果要观察某一因素的变化可能会对收入分布产生的影响，就必须借助于模拟的方法得到反事实的收入分布。以价格效应为例，现在要考察 1988 和 1999 年两年间的收入分布的差别，感兴趣的问题是这一差别在多大程度上是由劳动力市场上各人力资本要素的回报率所决定，要得到这一结果，对应的方法就应该是，通过将估计得到的 1999 年各人力资本要素价格的回报率代入 1988 年的工资或所得方程，而保持其他所有因素不变，再将这样得到的收入分布与 1988 年的原始收入分布作比较，就可以观察到价格效应对收入分布差别的影响，用公式可以表示为：

$$P_{01} = D_1 [w_0, v_0(w_0, m_0; g_0), e_0; l_0; b_1] - D_0 [w_0, v_0(w_0, m_0; g_0), e_0; l_0; b_0] \quad (3)$$

其中 D_0 是 1988 年的原始收入分布， D_1 是模拟得到的 1988 年的收入分布，二者的不同唯一来源于价格参数的改变，它由 b_0 变成了 b_1 。同样道理，可以得到各种结构性选择参数、结构性选择参数与价格参数、以及各种细分的价格参数的变动对于收入分布的定量影响，从而对他们的相对重要性作出评估。

2、家庭收入模型

假设某个家庭中处于工作年龄的成员个数为 n ，则该家庭的总收入可以表示为：

$$y = \sum_{i=1}^n (L_i w_i + y_{i0}) \quad (4)$$

其中，个人 i 的工资率 w_i 为：

$$w_i = f(X_i, \mathbf{e}_i) \quad (5)$$

个人 i 的工作时间 L_i 为：

$$L_i = \text{Sup}(0, g(Z_i, \mathbf{h}_i)) \quad (6)$$

在上述方程中， y_{i0} 是个人 i 的非劳动（非挣得）收入， X_i 、 Z_i 均为 i 的个人可观察特征向量，分别与工资率和劳动时间的决定有关，其包含变量可以重复但不一定相同， \mathbf{e}_i 、 \mathbf{h}_i 是 i 的不可观察特征，分别决定工资和劳动时间。 f 是工资率决定函数， g 则可以看作是一个关于个人劳动时间供给的隐性函数（latent function），它决定了个人 i 是选择不工作（劳动时间为 0）还是工作。

3、计量模型设定

在实际的研究中，对上述家庭收入模型的估计受到现有数据的制约。由于没有关于个人工作时间的详细数据，我们无法估计（5）（6）中的工资率方程和劳动时间供给方程，而只能代之以总和的工资工程和离散形式的就业参与选择方程。

具体地说，单个家庭的人均收入由下式决定：

$$y_h = \frac{1}{n_h} \left[\sum_{i=1}^{n_h} \sum_{j=1}^J (I_{hi}^j y_{hi}^j + y_{hi0}) \right] \quad (7)$$

h 表示一个家庭， n_h 是家庭成员的个数， j 表示参与选择集，共有 J 个，在本文中具有选择能力的家庭成员面临的选择共有四个：家务劳动或待业在家²⁸、国有企业工作、集体企业工作或其他所有制企业工作²⁹³⁰。家庭的收入由每个成员的收入和决定，而每个成员的收入由

²⁸ 从经济学意义上讲，家务劳动与待业应该有根本性的不同，因为代表了不同的工作意愿，本文曾尝试将待业在家作为一个独立的选择，但由于接受调查者中参与家务劳动和待业在家的样本量过少，影响了后面的模型估计，因此放弃；同时，调查数据也未区分失业和待业两种状态。

²⁹ 这里其他所有制不仅包括了传统意义上的其他各种经济类型单位，主要是联营经济、股份制经济、外商和港、澳、台经济等，还包括城镇个体经营者和城镇个体被雇者，严格来说城镇个体经营者应该属于经济学意义上的厂商而非普通被雇佣者，他从事生产活动得到的全部报酬是企业利润，从这一点上来说，用下面（8）式中的工资或所得方程来估计个体经营者的收入并不合适，不过由于整体样本中城镇个体经营者人数很少，这样处理对结论的影响应该不大。

³⁰ 另一种对选择进行分类的方法也被尝试过，即按照就业者所在的行业来分类，习惯的做法是分为第一、二、三产业三类，以及家务劳动者和待业者。这种分类方法下，以下各方程估计结果在统计显著意义上的差别并不大，因此略去。

挣得收入和非劳动收入（或非挣得收入 unearned income）两部分构成。³¹上式中， y_{hi}^j 表示第*i*个人在第*j*个选择中所能挣到的收入， I_{hi}^j 是一个示性函数（indicator function），如果第*i*个人作出了第*j*个选择 I_{hi}^j 就为 1，否则为 0。 y_{hi0} 是第*i*个人所得的非挣得收入，主要包括财产性收入、转移收入和家庭副业生产收入三项。

从（7）式可知，要得到家庭收入，需要决定的变量包括 y_{hi}^j 、 I_{hi}^j 和 y_{hi0} 。相应地，一个家庭的总体收入是由以下一组方程决定的：

$$\log y_i^j = \mathbf{a}_j + X_i \mathbf{b}_j + \mathbf{e}_{ij}, \text{ 如果 } i \text{ 的选择为 } j \quad (8)$$

$$P_{ij} = \Pr ob(U_{ij} > U_{ik}) = \Pr ob(Z_i \mathbf{l}_j + \mathbf{h}_{ij} - Z_i \mathbf{l}_k - \mathbf{h}_{ik} > 0) \dots \text{for all } k \neq j \quad (9)$$

（8）式是工资或所得方程， y_i^j 是第*i*个人在第*j*个选择中所能挣得的收入，在这里能挣得正收入的选择有三个：国有企业工作、集体企业工作以及其他所有制企业工作。按照同类文献的惯例， y_i^j 以自然对数形式进入方程。 X_i 是第*i*个人的个人特征构成的向量， \mathbf{e}_{ij} 是随机扰动项，在这里被理解为不可观察到的个人特征， \mathbf{a}_j 、 \mathbf{b}_j 是第*j*个选择下工资或所得方程的参数结构，即个人特征的市场价格。³²这里， X_i 具体包括：常数项、性别、年龄及年龄平方、经验和经验平方、教育程度虚拟变量、年龄与教育乘积项、省份和家庭角色虚拟变量、以及逆 Mill 比率。³³³⁴³⁵在对各种选择下工资或所得方程的估计中，有一个很重要的问题是如何纠正自选择偏差（self-selection bias）的影响，这也是方程（8）中附加了条件的原因所在。自选择偏差的问题广泛存在于经济学的各项经验研究之中，这一现象产生的根本原因在于，如果要承认有经济学作为一门科学的存在，那么就必须承认所观察到的个人行为是理性的经济人在约束条件下作最大化（maximize utility）的结果，于是，我们所得到的数

³¹ 变量说明详见附录一。

³² 方程（8）是 hedonic 形式的工资方程，参见 Rosen（1967，JPE），或更简单地，Ehrenberg and Smith（1997）。另外， X_i 中的一部分如省份虚拟变量严格来说不能被称为人力资本。

³³ 对年龄与经验的相关系数估计表明二者相关性不大。具体的变量说明参见附录。

³⁴ 88 年数据中包括六种教育程度：大学本科、中专、高中、初中、小学和文盲，99 年数据中增加了大学专科一项，由于后面作模拟时要求各变量的含义（包括虚拟变量）必须相同，同时也由于文盲人口在样本中很少，因此将教育虚拟变量分为五类：99 年数据中大学专科与中专归为一类，两年数据中的小学和文盲都归为一类。

³⁵ Inverse Mill's ratio，对这一变量的解释在后面的估计说明中。

据常常都是非随机性的，需要专门的计量方法来处理。³⁶在多元选择情形下，自选择偏差的纠正较二元选择情形下更复杂一些，Lee(1983)给出了对这类问题的一般性解决办法，³⁷ Lee (1978, 1984) 是对这一方法的具体运用。在后面的估计中我们将用到这一方法。

(9) 是多元 Logit 形式的个人参与选择 (participation choice) 方程。为避免使得后面的模拟程序过于复杂，在这里我们采用了简化式的模型设定，也即各种可能选择下的工资没有进入参与选择方程。 P_{ij} 是第 i 个人选择第 j 个选择的概率， $U_{ik} = Z_i \mathbf{l}_k + \mathbf{h}_{ik}$ 是 i 选择第 k 个选择得到的效用， Z_i 是 i 的个人特征向量， \mathbf{l}_k 是与第 k 个选择对应的参数结构， \mathbf{h}_{ik} 表示 i 选择第 k 个选择得到的随机效用，³⁸服从双指数分布 (double exponential distribution)。 Z_i 包括：常数项、年龄和年龄平方、小孩个数、成年人个数、教育程度虚拟变量、年领与教育乘积项、以及省份和家庭角色虚拟变量。具体来说，可以计算得到 P_{ij} 为：

$$P_{is} = \Pr\{j = s\} = P^S(Z_i, \mathbf{I}) = \frac{e^{Z_i \mathbf{l}_s}}{e^{Z_i \mathbf{l}_s} + \sum_{j \neq s} e^{Z_i \mathbf{l}_j}} \quad (10)$$

根据 (2) 式，还需要对一部分的家庭可观察特征进行内生化处理，在这里考虑家庭成员的个人教育程度特征作为前面提到的 v_{it} ，可以将其视为选择的结果，同样采用多元 Logit 形式的选择方程对个人教育程度的选择行为进行估计：

$$ML_E(e_i | w_i) \rightarrow \mathbf{g} \quad (11)$$

e_i 是第 i 个人的教育程度， w_i 是一系列完全外生的家庭和个人特征变量，它们共同决定了 e_i 。³⁹在 w_i 中，当事人个人的特征包括：年龄、组群 (cohort) 地区和性别，家庭的特征是成年人的个数。⁴⁰ \mathbf{g} 是回归所得的参数。

³⁶ 在大样本情形下，随机性假设不再是必须的，但实际研究中更多的情形只是有限样本。参见 Greene (1997)。

³⁷ 这一模型是对二元选择情形下 Heckman 模型的推广。

³⁸ Maddala (1983)。

³⁹ 就我们所选定的 w_i 变量而言，这种决定关系可能并不具有很强的因果性 (causality) 解释，纯统计意义上的解释可能更多，但首先这样做并不影响我们的结论，因为我们的目的只是要找出对个人教育选择有解释力的因素，而非要对个人教育选择行为作严格经济学意义上的因果关系分析。其次，这一做法也可看作是对教育选择行为的一个后验的 (ex post facto) 估计。从这一点上来说，第六部分中由改变教育选择方程参数得到的模拟结果行为意义上的解释也要弱一些。

⁴⁰ 就家庭某个成员尤其是子女的教育选择而言，家庭其他成员如父母亲的一些个人特征——无论可观察到的 (如教育) 或不可观察到的，可能都会产生影响，甚至会有一定的因果性，最好的做法是应该去看看这些变量的影响。但由于数据中比较有意义的个人特征如教育只有教育程度的定性描述，没有具体的教育年限，我们只能以虚拟变量的形式加入教育变量，再考虑到家庭其他成员的个数，这样会使得解释变量的个

以上讨论的都是个人挣得收入的决定。对个人非挣得收入，可以采用 Tobit 形式的方程来估计，选用这一形式估计方程的理由是，数据中的非挣得收入主要包括财产性收入、转移性收入和家庭副业生产收入三项，由于样本中家庭副业生产收入为正的人数非常少，非挣得收入主要是来自财产性收入和转移性收入，对转移性收入，可以近似认为是满足一定条件才可以得到，那些不满足条件的人为 0。⁴¹

$$Tobit(y_{i0} | e_i, age_i, region_i, sex_i, n_{ah}) \rightarrow \mathbf{d} \quad (12)$$

其中 y_{i0} 是第 i 个人的非挣得收入， e_i 、 age_i 、 $region_i$ 、 sex_i 、 n_{ah} 分别表示其教育程度、年龄、地区、性别和家庭成年人人数。 \mathbf{d} 是回归所得的参数。

4、模型的估计

需要估计的模型有四个：(8)、(9)、(11) 和 (12)，其中对工资或所得方程 (8) 的估计是重点。

对于多元选择情形下工资或所得方程中的自选择偏差问题，Lee (1983) 提出了一种可行的纠正方法，这一方法与 Heckman (1974) 提出的两步法在精神上是一致的，即可以把自选择偏差看作是一种遗漏了相关变量的模型设定错误 (specification error)，要纠正这一误设所导致的偏差，需要对原模型进行两步估计：第一步先估计参与选择方程，并根据估计的结果计算得到逆 Mill 比率；第二步，将逆 Mill 比率作为遗漏变量添加到工资或所得方程中，作为新增的解释变量，然后用 OLS 方法回归就可以得到理论上一致的结果。根据 Lee (1983) 的结论，在 s 选择被选择的情况下，与 s 选择相对应的工资或所得方程应该被修正为：

$$\log y_i^s = \mathbf{a}_s + X_i \mathbf{b}_s - \mathbf{s}_s \mathbf{r}_s f(J(Z_i \mathbf{I}_s)) / F(Z_i \mathbf{I}_s) + \mathbf{w}_i^s \quad (13)$$

这里 \mathbf{s}_s 是原工资方程 (8) 中扰动项 \mathbf{e}_{is} 的标准差， \mathbf{r}_s 是 \mathbf{e}_{is} 与参与选择方程扰动项 \mathbf{h}_{is} 的相关系数， \mathbf{w}_i^s 是工资或所得方程被修正后得到的新的随机扰动项，它的条件期望为 0，也即：

$E(\mathbf{w}_i^s | s \text{ 被选择}) = 0$ ，条件方差相异。 $J = \Phi^{-1} \circ F$ ，其中 Φ 为标准正态累积分布函数， F 是被重新定义的随机扰动项 $\mathbf{x}_i^s = \max_{j \neq s} U_{ij} - \mathbf{h}_{is}$ 的累积分布函数。⁴² $f(J(Z_i \mathbf{I}_s)) / F(Z_i \mathbf{I}_s)$ 就是所谓逆 Mill 比率，先对参与选择方程 (9) 进行估计，得到参数结构 \mathbf{I}_s 的估计值，根据这一估

数相对于现有的样本数量而言过大。因此放弃。

⁴¹ 对非挣得收入的处理还有另外一种方法，这一方法的思路是先对某一年的非挣得收入样本进行排序，得到每个观察值在整个分布中的相对位置，然后根据这个值进行模拟。详见 BFL (2002)。

⁴² $F(\mathbf{x}_i^s) = \exp(\mathbf{x}_i^s) / \left[\exp(\mathbf{x}_i^s) + \sum_{j \neq s} \exp(X_i \mathbf{b}_j) \right]$

计值计算得到逆 Mill 比率，然后对 (13) 进行估计就可以得到所需要的工程方程系数的估计值，以及 w_i^s 的标准差的估计值。

需要注意的是，方程 (8) 和 (9) 中设定的工资方程和参与选择方程都是简化式 (reduced form) 而非结构式 (structural form) 形式的。在我们的模型设定中，参与选择方程的个人特征项 Z_i 中没有包括各种选择下可能得到的工资收入，然而在真实世界中，就单个决策者的参与决策而言，各种可能的选择下能够获得的潜在工资或所得显然具有重要意义。采用简化式的工资方程和选择方程带来的一个后果是，对于观察到的决策者参与选择行为方式的改变，我们将无法有效区分它在多大程度上是由工资率的变动所导致，以及在多大程度上确实源于决策者本身行为方式的改变。如果采用结构式的方程设定，这一问题确实可以得到解决，但会带来一些估计上的复杂性。Mcfadden (1978) 指出，在随机扰动项服从双指数分布的假设下，多元 Logit 模型可以由决策者追求随机效用最大化的假设推出。于是，如果把 (8) 和 (9) 设定为结构式的方程，就相当于各种可能选择之下的收入进入了效用函数，在这种情况下，(8) 和 (9) 应该被作为联立方程同时估计，因为此时 y_i^j 已经内生化了，内生化了带来了估计和识别上的困难，通过对参与选择模型施加一些合适的限制条件，可以解决模型识别的问题。⁴³ 此外，对于同时含有定性变量和受限变量的联立方程的估计和识别问题，Lee (1978) 也给出了一些讨论。但本文并不准备采用结构式形式的参与选择方程，这主要是考虑到后面进行模拟时的可操作性，避免过于繁杂。

对 (11) 和 (12) 的估计比较简单，存在的问题是回归得到的 Pseudo R^2 偏小，尤其是采用 Tobit 模型对非挣得收入进行估计的结果不好，这说明模型的解释力不强，主要的原因可能在于选用的解释变量不够，正如前面所提到的，对这一模型解释变量的选择，更多地是从统计意义上解释，因果性意义上的解释比较弱。从这一点来看，后面各项由模拟得到的最终结果中，包含了非挣得收入模拟的两项可靠性要差一些，一个可行的方法是采用前面提到的另一种方式对非挣得收入作重新模拟，然后比较两种方法所得的结果是否一致，通过这种方式可以判断结论的稳健性 (robustness)。⁴⁴

各计量模型的估计样本详见附录二。对参与选择方程 (9) 的估计是分性别进行的，这样做的理由是考虑到男性和女性的参与选择决策行为可能存在系统性的差别。⁴⁵ 对非挣得收入方程 (12) 的估计是用全部样本进行的，这是由于我们采用了删节形式的 (censored) Tobit 模型，在这种模型中，个人的非挣得收入被认为由某个内生的指标 (latent index) 所决定，而这一指标又取决于一系列个人特征，因此我们能够观察到的只是取得正的非挣得收入的那

⁴³ 参见 Bourguignon、Ferreira and Lustig (1998) 技术附录。

⁴⁴ 本文尚未作这一步。

⁴⁵ 另一种方式是按照家庭成员在家庭中的角色来分别估计，因为家长、配偶及子女的参与选择行为可能会存在系统性的差别。但由于样本量较少，在这种估计方式下很多变量都将由于观察值过少而被自动丢弃 (dropped)，由于后面要做模拟，所以未采用这一分类估计方式。

部分人。全部的计量模型估计结果都列在文章后面的附录三中。

五、模拟

在得到估计结果之后，重要的是要进行模拟以得到所需要的各种反事实的分布，对这些分布进行分析可以得到各项因素——禀赋以及参数结构或它们的任意组合对收入分布差别的定量影响。

具体的模拟方法已经在上节中予以说明。以下具体说明各个模拟步骤。

第一步是想看工资或所得方程的系数，也就是人力资本要素的价格对两年间收入分布差别的影响。为此，可以将估计得到的 99 年工资方程的系数（可以是某一个系数也可是全部系数）代入 88 年工资方程中，其他所有因素保持不变。这样做的一个假设是，两年之间样本人群的不可观察特征的分布未发生改变，但事实可能并不是这样的，为此，需要考察人群不可观察特征分布的系统性改变对两年之间收入分布差别的影响。工资方程中的 w_i^s 可以被理解成包括了一切不可观察的人群特征的影响，因此只要改变 w_i^s 的分布就可以了。在纠正了选择性偏差的工资或所得方程（13）中，扰动项 w_i^s 的分布通常被假设为服从正态分布，由于 w_i^s 的条件期望从而无条件期望为 0，因此两年之间的 w_i^s 分布的改变可由标准差的改变来刻画，因此，只需对 w_i^s 作倍数调整就可以了：

$$\square w_{it}^s = \frac{s_t}{s_t} w_{it}^s \quad (14)$$

w_{it}^s 是 t 时期第 i 个人在 s 选择下工资方程中的扰动项， s_t 和 s_t 则分别是 t 年和 t 年扰动项对应的标准差。这些值都容易从方程（13）中估计得到。

第二步是想看参与选择方程的参数结构对两年间收入分布差别的影响。为此只要将估计得到的 99 年参与选择方程的参数结构代入 88 年的参与选择方程，这样可以得到模拟的 $Z_i \mathbf{1}_j$ ， U_{ij} 中的另一部分是随机效用项 h_{ij} ，对它的模拟是唯一的困难所在。由于我们要模拟样本中所有人群在每种选择下的效用 U_{ij} ，因此必然有部分人的随机效用项是无法直接观察到的，我们所能直接观察到的只是各种选择中目前正在位的选择者，其它潜在选择者的随机效用项是无法直接观察到的。可行的办法有两种：一是先根据每个选择中可观察到的那部分人的数据，得到扰动项的经验分布，然后从该经验分布中随机抽取；二是直接算出扰动

项的真实分布，然后根据这一分布随机抽取，⁴⁶以此作为模拟的扰动项。这里采用后一种方法，由于已经假设参与选择方程中的随机扰动项服从双指数分布，可计算得到在第 i 个人选择 j 的情况下， h_{ij} 的累积分布函数为：⁴⁷

$$F(h_{ij} | pc = j) = F(h_{ij} | Z_i \mathbf{I}_j + h_{ij} > \max_{s \neq j} (Z_i \mathbf{I}_s + h_{is})) \quad (15)$$

而未被选择的选择 k 对应的随机扰动项的分布为：

$$F(h_{ik} | pc = j) = F(h_{ik} | Z_j \mathbf{I}_j + h_{ij} > Z_s \mathbf{I}_s + h_{is}) \cdots k \neq j \quad (16)$$

在得到全部的 $Z_i \mathbf{I}_j$ 以及 h_{ij} 后，可以得到模拟的 U_{ij} ，对每个人所有的 U_{ij} 进行比较，就可以得到模拟的这个人的参与选择。⁴⁸

第三步是前两步的综合。唯一的不同点在于，由于进行参与选择模拟后很多人的选择行为可能已经改变，因此这部分人在新选择下的工资方程的扰动项是不可观察的，为了实现模拟的目的，首先要根据每个选择下在位样本的数据得到改选择下扰动项的经验分布（主要是标准差），然后根据这一经验分布随机抽取新进入者在该选择下的扰动项。

第四步是加入教育程度选择行为的影响。对教育选择的模拟与第二步在方法上是一样的。同样，可以将第一步和这一步综合以考察教育选择参数结构与工资方程参数对收入分布差异的联合影响。

第五步是第二步与第四部的综合。目的在于分析教育选择参数结构、工资方程参数及参与选择参数对收入分布差异的联合影响。

第六步是考察非挣得收入分布的改变对两年间收入分布差异的影响。只考虑非挣得收入一项的改变对收入分布差异将产生的影响。

最后一步是考虑全部参数同时改变对两年间收入分布差异的影响。同时也可由此看到全部因素——禀赋和参数结构对两年间收入分布差异的综合解释力。

六、1988-1999 年间中国城市居民收入分布变动的分解

用模拟方法对 1988-1999 共 12 年间中国城市居民收入分布变动进行分解的结果列在文

⁴⁶ 应该记住的是，模拟的结果在一定程度上将受到随机变量抽取结果的影响，如果样本量很大，对扰动项随机抽取的结果会逼近其真实分布，在样本量有限的情况下，一个解决的办法是重复抽取多次，不过我们发现结果变动并不大。另一个增强文章结论稳健性的方法，是可以对 89-99 年、或 88-98 年的数据组合重复所有操作，看结论的变动是否显著。

⁴⁷ 详见 Bourguignon, Fournier and Gurgand (2001)。

⁴⁸ 这里未采用期望形式的模拟结果，而是得到一个确定的选择，这样做主要是为了方便，因为在后面的教育程度选择行为的模拟中，采用期望值可能会得到非整数，由于我们的数据中教育程度值是定性的而非具体的教育年限，非整数的教育程度将失去意义。这一解释来自 Ferreira 对作者的 email 回复，在此谨致谢意。

章的附表中。表 3 是根据个人月均收入模拟结果计算得到，表 4 由家庭人均月收入⁴⁹模拟结果计算得到，对 1999 年的原始个人月均收入数据作了通货膨胀指数平减，转换成 1988 年的人民币。⁵⁰需要指出的是，直接以个人月均收入为基础来计算不平等指数并不那么合理，虽然每个家庭成员的收入一般都是个人活动所得，但很多支出或消费行为却是以家庭为基础来进行决策的，家庭成员掌握的经济资源中至少有部分是可以共享的。这里之所以同时也列出了根据个人月均收入模拟计算得到的结果，主要是为了对比。更合理的作法是以家庭（household）为单位来分析不平等程度，由于每个家庭都是异质的（这不仅由于不同家庭人数不同，还因为不同年龄和精神状态的人其基本需求也不同），因此只有先对异质的家庭进行转换才可以进行比较，国外学者常用的转换标准是家庭均等比例（household equivalence scale），这一比例可根据基本营养需求或其他标准计算得出。⁵¹由于家庭均等比例的计算涉及到价值判断，难以取得共识，所以这里我们直接采用人均的算法，这样做的一个隐含假设是：同一家庭中不同的人的基本需求是相同的。

先来看表 3。表 3 给出根据个人月均收入计算的收入分布测度指标，包括水平（level）指标和分散度（spread）指标两个方面，平均个人月均收入属于水平指标，分散度指标包括基尼系数、广义 Theil（0）指数、广义 Theil（1）指数以及广义 Theil（2）指数，对广义 Theil（a）系列指数的性质已在前面简要说明。可以看到，无论是基尼系数的数量值还是两年之间基尼系数的相对差异，以个人月均收入为基础得到结果都较前面第三节的结果要大，对广义 Theil（a）系列指数这一点也成立，这说明以家庭为基础来计算收入不平度程度有均等化的作用，正如前文所言，这主要是因为家庭成员之间可以共用部分经济资源，从保险的意义上来说，一个家庭在收入和消费平滑上的作用，就象是一个小型的信贷市场。

从模拟的结果来看，工资或所得方程参数结构的差异解释了两年间收入分布差别的相当大部分，在 1988 和 1999 两年收入基尼系数 17 个百分点的差别中，有 12 个多百分点由这些参数的差别造成，其中，最有解释力的是教育程度⁵²和省份虚拟变量的系数。在考虑了两年

⁴⁹ 前面第三节中计算不平等指数时我们所用的是可支配收入指标，这里我们采用的是税前收入指标，这样做主要是因为对税收的模拟没有较好的办法，由于所得税是累进的，简单的回归可能会导致较大的偏差，另外，目前中国所得税的征管也很不完善，这主要是因为缺乏完善的收入监管和申报制度。

⁵⁰ 这里的平减所使用的价格指数是由《中国市场统计年鉴》所公布的居民生活费用价格指数，这一年鉴公布的是环比指数，将 89-99 共 11 年的环比指数累乘，可以得到 1999 年相对于 1988 年的价格指数，各省的收入数据用该省对应的价格指数调整。除 92 年外，各年年鉴都公布了分省且分城市和农村的总指数，我们选用的是城市指数，92 年选用的是各省总指数。89 年的数据由《中国统计年鉴》查得。另外，从 93 年起，这一指数的名称改为“居民消费价格指数”。我们发现价格调整对最终结论影响很小，这可能主要是因为 12 年间各省之间的相对价格变化并不大。另外，这些公布的指数也是通过 UHS 数据计算所得。

⁵¹ 参见 Slesnick（2001）。

⁵² 这一结论与 BFL（2002）对巴西与美国收入分布差异的分析结果相同。

间工资或所得方程参数结构的差异后,个人不可观察特征分布的变化对两年间收入分布差别的影响较小,只有1个百分点左右。参与选择行为以及教育程度选择行为的参数对收入分布差异的影响也较大,分别是近6个和7个百分点。但这两种参数对收入分布差别的联合解释能力并非各自解释能力的简单加总,即两种参数对两年间收入分布差别的影响是不可加的(non-additive)。非挣得收入参数结构的改变对两年间收入分布差别有一个降低的作用,这也是各种参数中唯一对收入分布差别有负向作用的量。

以上是对以个人月收入为基础作模拟所得结果的简要评述。下面对表4结果进行详细分析。表4是以家庭人均月收入为基础得到的结果。

总体来看,1999年家庭人均月收入的基尼系数比1988年上升了约13.2个百分点,上升幅度超过50%。⁵³从绝对水平来看,与1988年相比1999年真实家庭人均月收入也增加了一倍多。

工资或所得方程参数结构对两年间收入分布差别的影响如标号1与2之间对应的各行所示。如果1988年劳动力市场上各种个人特征的回报率同于1999年,那么1988年的基尼系数将上升到0.421,约有19个多百分点上升,超过了1999年真实基尼系数6个百分点。这一结论的重要涵义是:工资或所得方程的参数结构对两年间收入分布差别的影响巨大。对此的一个可能解释是,由于1988年时中国的市场化改革尚在起步阶段,劳动力市场的市场化程度很低,至少有三个方面的重要束缚造成了劳动力的不能自由流动:企业缺乏充分的自由雇工的权利、城市房屋租赁市场尚未放开以及粮食户口本的存在。⁵⁴由于不存在竞争性的劳动力市场,各种个人特征的回报率不等于其边际产品价值,或者说,个人特征的回报率被人为压低和平均化了的。这种状况到1999年已大大改善,多种经济成份尤其是私营企业以及各种形式的外资企业迅速发展,企业微观层次的改革使其自主权增加,劳动力市场变得更加开放等。供需方自由选择权力的增加使劳动力市场更加统一和更具竞争性,那部分拥有较高人力资本但在以前因为制度的束缚而不能取得应有收入的人,现在则拥有更多的机会去劳动力市场上按照其人力资本的边际产品价值挣得回报率,这种分化导致了不平等程度的增加。

在工资方程的各种参数中,对两年间收入分布差别影响最大的是省份和教育程度虚拟变量。如果采用1999年省份虚拟变量的参数,1988年的基尼系数将上升近16个百分点,超过了1999年真实的基尼系数水平,这是控制了两年间一系列个人特征及其回报率结构差异

⁵³ 表1中的结果显示二者的差别为11.5个百分点。两个结果略有差别,这主要是因为表4给出的结果中,已对99年的原始收入数据进行了通货膨胀调整。两表中1988年的结果也稍有差别,主要是因为在进行方程估计时,对原样本略有调整,详见附录中的数据说明。

⁵⁴ 这一观点来自周其仁(2003)的一次谈话。

后得到的“纯”影响。我们的样本包括了六个省的数据，每个省都是各自所在区域的代表，陕西代表了西北地区，辽宁代表了东北地区，四川是西南地区的典型，直觉的生活体验告诉我们 12 年来这些地区尤其西北、东北地区与沿海的差距拉大很快，而这里的分析则表明，即使抽象掉个体特征分布及其回报率差异所带来的影响，12 年间中国城市的地区间差异仍然大大上升。Meng (2001) 定义了与本文相近的家庭可支配收入概念，与本文可比性较强，采用与本文相同六省的数据作为样本得到的结论表明，地区效应可以解释 1988-1995 年间基尼系数变动值的 35.26%，仅次于经济转型因素，在 1995-1999 年间地区效应的解释力则降低为 0.73%。以往相关文献都是考察“毛”的地区差异，我们的研究表明了即使控制了人的特征及其回报率结构差异后依然存在着“纯”的地区差异。这可能是由于各地区的特征，例如宏观环境、社会观念和劳动力市场状况等等都不一样，因此自然会有地区差异。

在重要性上仅次于省份虚拟变量系数的是教育程度虚拟变量的系数，它的改变会使 1988 年基尼系数上升约 6 个百分点。这一作用完全由教育回报率的变化所导致，⁵⁵因为我们已经控制了个体教育程度分布本身的变化。教育回报率对不平等程度的这种影响与我们的直觉相一致。奈特、宋丽娜 (1999) 的估计显示，教育回报率的上升可以解释 1988 - 1995 年间城市职工工资差异扩大幅度的 10.1%，这里的结论则表明，教育回报率的变化对 1988-1999 年间城市居民收入分布的差别有更大的解释力。除了研究对象上的不同外，导致研究结论差别的原因可能来自多方面，首先本文采用了更长时间区间的的数据，其次所用方法不同也会导致结论精确性上的差异，此外还有一个潜在的解释是，与前一区间相比，1995-1999 年间教育回报率的变化更激烈地扩大了城市居民收入分布的差别，这是可能的，因为通常都认为 1995-1999 年间中国发生了更为激进的改革。教育程度与年龄交叉项系数的改变对收入分布差异的影响不大，在工资方程回归结果中这些项也都不显著，表明年龄的回报率伴随教育程度的改变并没有显著的不同。经验及经验平方的影响约为 2 个百分点。若按照通常作法将教育和经验视作人力资本的主要构成要素，则同样人力资本因素对于两年间收入分布差别的解释力很强。Meng (2001) 用家庭劳动力的平均年龄及平方、平均教育年数及平方来测度家庭人力资本存量，结论表明在 1995 - 1999 年间人力资本因素可以解释城市居民家庭收入差距扩大的 3.09%，而在 1988-1995 年间则达到 22.50%，居第三位。⁵⁶ Meng (2001) 区分

⁵⁵ 宋晓青 (2003) 用与本文同样的数据分析了 1988-1999 年间中国城市教育回报率的变化。

⁵⁶ Meng (2001) 的研究采用了中国社会科学院经济研究所课题组分别在 1988、1995、1999 年调查所得的数据，上面提到的结论是用北京、江苏、辽宁、河南、四川、甘肃六省数据可得，与本文可比性较强。运用全部十一省数据得到的结果分别为 23.65%、13.53%。虽然在具体数值上有所不同，但对于省份和教育虚拟变量参数在各因素中的相对重要性，本文与 Meng (2001) 的结论是一致的。

了两个不同阶段：1988-1995 的温和改革时期以及 1995-1999 的激进改革时期，这使得与她文章结论的比较存在困难。但 Meng (2001) 与本文在技术处理上的一些重要不同同样值得指出：在工资或所得方程的估计中，我们考虑了剔除选择性偏差可能带来的影响；Meng (2001) 对收入的回归以真实人均家庭可支配收入作为被解释变量，本文则以真实个人可支配收入作为被解释变量，并以此为基础作加总平均；此外，Meng (2001) 将家庭成员的就业状况、工作单位是否盈利、以及就业部门等理解为经济转型因素，并以此作为收入方程的解释变量，这会产生内生性 (endogeneity) 偏差。

其他因素如性别、年龄及年龄平方对两年间收入分布变化的影响不大，均约为一个百分点，占两年间基尼系数全部变动值的 7.6% 左右，说明性别差距并未有大幅扩大的趋势，这一结论与奈特、宋丽娜 (1999) 的研究近似。此外，逆 Mill 比率和常数项的影响也大概都在 2-3 个百分点之间。

以上是从分布的分散度特征来看。从分布的水平特征来看，结论相似。工资或所得方程的参数结构解释了两年间收入水平差别的绝大部分，其中教育程度系数对收入水平差别的影响最大，采用 1999 年的系数将使 1988 年的真实家庭人均月收入从 116.9 元增加到 161.8 元，证明 12 年间城市教育回报率有大幅上升；省份虚拟变量系数的重要性其次，对应的模拟结果是 140.2 元。上述结论说明这两个变量的系数对收入分布的分散度和水平同时有着重大影响。值得注意的是经验变量系数的模拟结果，采用 1999 年的系数将使 1988 年的真实家庭人均月收入骤降至 88.8 元，原因在于三种所有制经济中经验的回报率在 1999 年均已较 1988 年大幅下降，相比于国有制和集体所有制经济而言，经验的回报率在其他所有制经济中下降得最快 (附表 5、6)。工资制度的改革可能是导致经验回报率下降的原因之一，企业微观经营层次的自由化改革赋予了企业在工资决定上的更大自主权，与从前的工龄制度相比，以业绩定薪酬的工资制度因为更加有效率而被广为采用。

考虑人群不可观察特征分布的改变后 (标号 2 与 3 之间对应的各行)，得到的结论与上面基本相同，只是各种情形下的基尼系数数量水平稍有上升，这或者是由于个人不可观察特征项如能力等分布的分散度上升，或者是由于市场对这些特征给予的回报率更加真实地反映了个体之间存在的差别。⁵⁷后一种解释的可能性存在，因为自由化改革已使得劳动力市场更具有竞争性，能更有效地发现个人能力这样的隐藏信息并对之合理定价。对收入分布水平特

⁵⁷ 严格来说，工资或所得方程中的扰动项应该被看作是“不可观察个人特征所获得报酬——也即数量与价格的乘积”，但由于不可观察性，我们无法区分不可观察个人特征本身分布改变带来的影响，以及不可观察个人特征回报率的变化带来的影响。

征的影响也有与上相似的结论。

标号 3、4、5 行是考虑参与选择行为后的模拟结果。可以看出，如果人们在 88 年的参与选择行为与 99 年一样，那么 88 年的收入基尼系数将上升约 7 个百分点，这个幅度比较大，其中原因可以从模拟得到的参与选择结果来分析（表 5 中模拟结果 1），在模拟结果中，选择在国有企业和集体企业工作的人大幅减少，而选择在其他所有制企业工作的人大量增加，由于国有企业和集体企业分配制度上较浓厚的平均主义倾向，而其他所有制企业（主要是私有制和外商投资企业）更接近于按贡献分配，因此，参与选择行为改变导致收入基尼系数较大幅度上升的结果便不难理解，这种上升完全来自决策者参与选择行为方式的变化。另外，3 表明改变参与选择行为参数后，得到的 88 年真实平均家庭人均月收入较实际水平有所下降，这主要由于模拟结果 1 中选择家务劳动或待业的人数大大增加了。⁵⁸4 和 5 的结果表明，如果还加入工资或所得方程参数的影响，模拟得到的 88 年收入基尼系数还会有大幅度的上升，这再次表明工资方程参数对两年间收入分布的差别有很强的解释作用。需要指出的是，4 和 5 中的结果显然不能分别由 1、2 的结果和 3 的结果累加而成，这一点由方程（2）也很容易直接推得。

6、7、8 行是考虑教育选择内生化的模拟结果。如果 88 年的教育选择参数结构与 99 年相同，那么收入的基尼系数将上升到 0.412。这里面包括了两方面的影响，在参数改变导致教育程度选择（分布）的变化后，收入的分布将通过两种机制被影响到：一是通过工资或所得方程直接改变收入，二是通过改变参与选择间接改变收入，详见附录三的模拟流程图。从表 5 中模拟结果 2 可以看出，在教育选择参数发生改变，从而人群教育程度的分布改变的情况下，模拟得到的参与选择与原始参与选择有更大的差别：选择在国有和集体企业就业的人数进一步大幅下降，选择在其他所有制就业的人数则急剧上升，几乎占到总样本数的一半，这也是与 3 的结果相比，6 中收入的基尼系数大幅上升的原因，证明人们的教育选择行为方式将对收入不均等程度产生重要影响。引入 1999 年的教育选择系数也会导致 1988 年贫困程度的大幅上升。7、8 则表明，如果已经改变了教育选择的参数，那么工资方程系数的额外影响并不大。

9、10、11 是以上两步的综合。比较 3 与 9 可知，在已经改变教育选择系数的前提下，改变参与选择参数带来的额外影响很小（比较表 5 中模拟结果 3 与 1），但此时工资或所得方程参数结构对收入分布差别的影响却很大。

⁵⁸ 对待业的分析本不应简单等同于家务劳动，但由于数据限制，只能将二者看作同质的选择，这可能会导致结果上的一些偏差。

12 是考虑非挣得收入改变的结果。如果采用 1999 年的非挣得收入决定参数,1988 年的收入基尼系数会下降约 6 个百分点。由于文章定义的非挣得收入包括财产收入、转移收入和家庭副业生产收入(其中主要是前两项),我们尚不清楚非挣得收入导致的基尼系数下降到来自哪一方面,可能是财产收入,也可能是转移收入,⁵⁹对非挣得收入进一步的细分研究应该可以弄清根源所在。13 则表明,1988 和 1999 两年间收入基尼系数约 13 个百分点的差别中,有约 11 个百分点是由于工资或所得方程参数、参与选择行为参数、教育选择参数以及人群不可观察特征之分布的变化所引起,由于本文尚未考察人群完全外生可观察特征 w_i 之分布的改变对收入分布差别的影响,所以可以预见 1988 和 1999 两年间收入分布差别的绝大部分都可以由人群禀赋分布和参数结构的改变来解释。

七、结论及政策涵义

本文主要分析 1988-1999 年间中国城市居民收入分布的变动。在此之前,已有不少国内和国外文献对中国收入分布的变动作过研究,研究的内容包括:收入分布变动的总体趋势、城市居民收入分布的地区差异、城市居民收入分布变动的原因等。这些研究是对中国收入分布问题的初步研究,在深度上有所欠缺,这或者是由于数据的原因,或者是由于方法上的原因。

本文主要运用微观计量的方法对收入分布的变动进行分解,以定量地分析各种可能因素对收入分布变动的影响,这一方法来自 BFL(2002),其主要贡献在于,可以在保持其他所有相关因素不变的前提下,单独分析某个因素对收入分布差别的定量影响,进而可以对各种因素的相对重要性作出评估。文章所用的计量模型是建立在一个关于家庭收入决定的一般均衡模型基础之上的,但为了简化模型的识别、估计以及后面的模拟,我们采用了简化式的模型设定。与 BFL(2002)不同的是,我们在对工资或所得方程的估计上考虑了自选择偏差可能带来的影响。对工资方程估计的结果表明,自选择偏差确实存在(参见附表)。运用估计得到的结果,可以用微观模拟的方法具体分析各种因素——人群禀赋分布和各种参数结构对两年间收入分布差别的影响,这一模拟是通过得到反事实的分布来实现的。

模拟结果表明,无论是以家庭人均月收入或是个人月收入为基础来进行分析,得到的结论都是一致的。工资或所得方程的参数结构对收入分布的影响很大,其中省份和教育虚拟变量的影响最为突出。即使控制了人群特征分布及其回报率结构差异的影响,各地区之间存

⁵⁹ 财产收入包括利息、红利以及其他财产租金收入,转移收入包括离退休金、价格补贴、赡养收入、赠送收入、亲友搭伙费、记账补贴、出售财物收入及其他等。

在的差异仍然巨大，对这一部分差异的解释值得继续研究，它可能是由于各地区的特征，例如宏观环境、社会观念和劳动力市场状况等等都不一样所造成。人群不可观察特征之分布的变化对收入分布的影响不大。参与选择方程行为参数的改变对收入分布的影响较大，这里的一个重要涵义是，12 年来人们的就业选择行为已经发生了深刻的变化。这种变化可能源自人们价值观念的改变，意识形态在社会中的作用大大弱化，取而代之的是多元化的追求经济利益的价值取向，这导致了人们行为方式包括就业选择方式的转变，以前被绝大部分人视作就业第一选择甚至当作一种信仰来追求的国有企业铁饭碗不再吃香。参与选择行为的转变导致了就业结构的剧烈变化（表 5 模拟结果 1），从而使得收入分布的分散度上升。更为重要的教育选择方程参数的改变，它将通过两种途径影响收入的分布，即直接改变教育禀赋从而改变收入，以及通过改变教育禀赋进而改变参与选择行为从而改变收入。需要注意的是，在先改变了教育方程参数的情况下，参与选择行为参数的改变给收入分布带来的额外影响不大。如果在 1988 年人们可以和 1999 年相同的方式（结构参数）获得非挣得收入，那么 1988 年的收入不平等程度将下降 6 个百分点。由于没有更深入的研究，目前尚不清楚这种促使收入不均等程度下降的动力究竟来自何方。这也是本文可以继续努力完善的地方。

另外，由于数据可得性上的限制，本文主要关注的是中国城市居民收入分布差别的问题。鉴于巨大的城乡差距的存在，立足于全国范围（同时包括城乡）之上的研究将可能具有更加重大的政策意义。

总而言之，文章的结论表明，1988-1999 年间中国城市居民收入分布的大幅变动在一定程度上是经济体制转型的必然结果，但这并不表明政府已无需作为。

人们行为方式——主要是教育选择方式、参与选择方式的结构变化，反映了社会价值取向和思想观念的转变，即从过去的意识形态主导到现在的多元化社会价值观得到认可，这是任何经历转型的国家都将面临的一个自然过程。

相比之下，地区因素、人力资本——主要是教育回报率的变化对收入分布变化的显著影响则值得更多的关注。在剔除人群个人特征分布及其回报率结构差异影响后，地区差异对不平等程度的“纯”的影响仍然非常显著，这意味着各地之间的差别并不完全由劳动力的素质的差别所造成，因此政府应该在缩小地区差异上作出更多的努力，例如实行适当的财政、税收政策倾斜，就像在改革初期对待沿海开放城市那样；加强落后地区基础建设；大力改革低效的地方政府，降低交易成本等。教育回报率对不平等程度的显著影响则表明，政府应该鼓励和支持人力资本投资，例如加强教育投入，提高社会平均教育水平等，以防止人群教育分布分散度的进一步扩大，因为从各国经济发展的历史和经验来看，高等教育相对回报率的

上升将是一个不可避免的趋势，我们的社会正越来越成为一个“赢家通吃”(winner-take-all)的社会。与此相应的一个值得关注的问题是，随着经济体制转型的进一步深化，在未来的一段时期内中国的教育回报率可能进一步变化，这种变化将对居民收入分布产生怎样的和多大的影响，值得跟踪和研究。

另外，如果考虑到中国存在的巨大城乡差距，加强农村人力资本投资的问题将会显得更加迫切。20世纪90年代中后期以来，曾经被视作异军突起的乡镇企业经营状况每况愈下，1994年的财政分权制改革则使得乡镇一级政府在现有的五级政府体制中处于尴尬的境地，加上官员腐败等因素一起导致了乡镇财政状况的恶化，农村地区（尤其内地）的九年制义务教育不仅无法实行，而且农村居民还必须为乡村教师的工资支出以及其他名目繁多的教育行政支出买单，子女的教育费用已经在事实上成为农村家庭的一项重大支出，这使得他们不堪重负。在这种情形下，不少觉得无望的农村父母更倾向于子女辍学去沿海打工，而不是选择长期性的人力资本投资。从长期来看，农村年轻一代令人堪忧的教育状况将会对他们今后的收入状况带来更加不利的影响。对现实的改革需要从多方面着手，包括改革政府体制如减少行政层级、鼓励发展地方工业、以及大幅度增加对农村地区的基础教育投入等。

在理解了以上这些之后，对于过去12年中中国城市居民收入分布的大幅变化，我们应该有一个更为清醒的认识和判断，一方面，持续的社会制度的激烈转变是很少可见的。正如邹恒甫（2001）所指出的，如果作综合性考察的话，可以发现从上个世纪40年代末期开始一直到90年代，各国收入分配的趋势很难改变，像中国和东欧这种发生变化的现象都是社会剧烈变化的产物。⁶⁰这也许可以成为令我们稍感欣慰一些的理由。而另一方面，政府在防止地区差异进一步扩大、鼓励支持人力资本投资上仍将任重而道远。

⁶⁰ 参见 Bourguignon, Fournier and Gurgand (2001) 对台湾的研究。

参考文献

- Bourguignon, F., "Decomposable Income Inequality Measures." *Econometrica*, Vol.47, Issue 4, July 1979, 901-920.
- Bourguignon, F., Anne-Sophie Robilliard and Sherman Robinson, "Representative versus Real Households in the Macro-economic Modeling of Inequality." *mimeo*, 2002.
- Bourguignon, F., Francisco H.G. Ferreira and Nora Lustig, "The Microeconomics of Income Distribution Dynamics in East Asia and Latin America." *A research proposal, World Bank DECRA mimeo*, 1998.
- Bourguignon, F., M. Fournier and M. Gurgand, "Fast Development with a Stable Income Distribution: Taiwan, 1979-1994." *Review of Income and Wealth*, 47, NO.2, June 2001, 1-25.
- Bourguignon, F., Francisco H. G. Ferreira and Phillippe G. Leite, "Beyond Oaxaca-Blinder: Accounting for Differences in Household Income Distributions Across Countries." *NBER Working Paper*, 2002.
- Bourguignon, F., Francisco H.G. Ferreira and Nora Lustig, "MIDD: the Microeconomics of Income Distribution Dynamics." *mimeo*, 2001.
- Cowell, Frank A., *Measuring Inequality*, 2nd edition, London, Prentice Hall, 1995.
- Cutler, David M. and Lawrence F. Katz, "Macroeconomic Performance and the Disadvantaged." *Brookings Papers on Economic Activity*, 1991, (2), 1-74.
- Cutler, David M. and Lawrence F. Katz, "Rising Inequality? Changes in the Distribution of Income and Consumption in the 1980's." *American Economic Review*, Vol. 82, Papers and Proceedings, May 1992, 546-551.
- Deaton, Angus, *The Analysis of Household Surveys*, John Hopkins University Press, 1997.
- Deaton, Angus and John Muellbauer, *Economics and Consumer Behavior*, Cambridge University Press, 1980.
- Ehrenberg, Ronald G. and Robert S. Smith, *Modern Labor Economics: Theory and Public Policy*, 6th edition, Addison-Wesley, 1997.
- Greene, William H., *Econometric Analysis*, third edition, Prentice-Hall, 1997.
- Gustafsson, B. and Shi Li, "Inequality in China at the End of the 1980s: Locational Aspects and Household Characteristics." *Asian Economic Journal*, March 1998, 35-63.
- Heckman, James J., "Sample Selection Bias as a Specification Error." *Econometrica*, Vol. 47, Jan. 1979, 153-162.
- Juhn, Chinhui, Kevin M. Murphy and Brooks Pierce, "Wage Inequality and the Rise in Returns to Skill." *Journal of Political Economy*, Vol. 101, No. 3, 410-442.
- Kakwani, Nanak C., *Income Inequality and Poverty: Methods of Estimating and Policy Applications*, Oxford University Press, 1980.
- Kanbur, Ravi and Xiaobo Zhang, "Fifty Years of Regional Inequality in China: A Journey through Revolution, Reform and Openness." *Mimeo*, Cornell University and International Food Policy Research Institute, April, 2001.
- Khan, Azizur, Keith Griffin, Carl Riskin and Renwei Zhao, "Household Income and its Distribution in China." *China Quarterly*, 1993, 132, 1029-1061.
- Krueger, Dirk and Fabrizio Perri, "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory." *NBER Working Paper*, 2002.
- Lee Lung-Fei, "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables." *International Economic Review*, Vol. 19, Jun. 1978, 415-433.

- , “Generalized Econometric Models with Selectivity.” *Econometrica*, Vol. 51, March 1983, 507-512.
- Litchfield, Julie A. “Inequality: Methods and Tools.” *Text for World Bank’s Web Site on Inequality, Poverty, and Socio-economic Performance*.
- Maddala, G.S., *Limited-Dependent and Qualitative Variables*, 1983.
- Meng, Xin, “Economic Restructuring and Income Inequality in Urban China.” Project paper, 2001.
- Park, Albert, Xiaoqing Song, Junsen Zhang and Yaohui Zhao, “The Growth of wage inequality in Urban China, 1988 to 1999.” Draft paper, 2003.
- Slesnick, Daniel T., *Consumption and Social Welfare: Living Standards and Their Distribution in the United States*, Cambridge University Press, 2001.
- Trost, Robert P. and Lung-Fei Lee, “Technical Training and Earnings: A Polychotomous Choice Model With Selectivity.” *Review of Economics and Statistics*, Vol. 66, Feb. 1984, 151-156.
- 阿玛蒂亚·森, “评估不平等和贫困的概念性挑战”, 《经济学(季刊)》, 第2卷第2期, 2003年1月。
- 陈宗胜, “中国城市居民收入分配差别现状、趋势及影响因素——以天津市为案例”, 《经济研究》, 1997年3月。
- 陈宗胜, 《中国居民收入分配差别的深入研究》, 《经济研究》, 2000年7月。
- 陈宗胜、周云波, “城镇居民收入差别及制约其变动的某些因素”, 《经济学(季刊)》, 第1卷第3期, 2002年。
- 菲歇尔, 《利息理论》, 陈彪如译, 上海人民出版社, 1999年12月。
- 古斯塔夫森、李实, “中国变得更加不均等吗?”, 载赵人伟、李实、李思勤主编《中国居民收入分配再研究》, 中国财政经济出版社, 1999年。
- 国家计委综合司课题组, “90年代我国宏观收入分配的实证研究”, 《经济研究》, 1999年11月。
- 国家统计局, 《中国统计年鉴》, 中国统计出版社, 1990年。
- 国家统计局城市社会经济调查总队, 《中国城市住户调查手册》, 1996年。
- 国家统计局贸易物资统计司, 《中国市场统计年鉴》, 中国统计出版社, 1991-2000各期。
- 国家统计局农调总队课题组, “城乡居民收入差距研究”, 《经济研究》, 1994年12月。
- 国务院研究室课题组, “关于城镇居民个人收入差距的分析和建议”, 《经济研究》, 1997年8月。
- 卡恩、李思勤, “中国的收入和不均等”, 载赵人伟、李实、李思勤主编《中国居民收入分配再研究》, 中国财政经济出版社, 1999年。
- 赖德胜, “教育、劳动力市场与收入分配”, 载赵人伟、李实、李思勤主编《中国居民收入分配再研究》, 中国财政经济出版社, 1999年。
- 李实, “对收入分配研究中几个问题的进一步说明”, 《经济研究》, 2000年7月
- 李实, “中国个人收入分配”, 《经济学(季刊)》, 第2卷第2期, 2003年1月。
- 李实、李文彬, “中国居民教育投资的个人收益率的估计”, 载赵人伟、格里芬主编《中国居民收入分配研究》, 中国社会科学出版社, 1994年。
- 李实、尚列, “国有大中型企业间职工收入差距的分析”, 《经济研究》, 1993年3月。
- 李实、赵人伟, “中国居民收入分配再研究”, 《经济研究》, 1999年4月。
- 李实、赵人伟、张平, “中国经济转型与收入分配变动”, 《经济研究》, 1998年4月。
- 李实、赵人伟、张平, “中国收入分配变动的理论解释和经验分析”, 载赵人伟、李实、李思勤主编《中国居民收入分配再研究》, 中国财政经济出版社, 1999年。

李实、赵人伟、张平，“中国经济改革过程中的收入分配变动”，载赵人伟、李实、李思勤主编《中国居民收入分配再研究》，中国财政经济出版社，1999年。

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。

奈特、宋丽娜，“中国的经济增长、经济改革和收入差距的扩大”，载赵人伟、李实、李思勤主编《中国居民收入分配再研究》，中国财政经济出版社，1999年。

任才方、程学斌，“从城镇居民收入看分配差距”，《经济研究参考》，1996年第157期。

宋晓青，“1988年到1999年中国城镇教育回报的上升：需求和供给分析”，北京大学中国经济研究中心硕士毕业论文，2003年。

赵人伟、格里芬，《中国居民收入分配研究》，中国社会科学出版社，1994年。

赵人伟、李实，“中国居民收入差距的扩大及其原因”，《经济研究》，1997年9月。

赵人伟、李实、李思勤，《中国居民收入分配再研究》，中国财政经济出版社，1999年。

赵耀辉、李实，“中国城镇职工实物收入下降的原因分析”，《经济学（季刊）》，2002年3月。

附录一：变量说明⁶¹

生活费收入 = 全民所有制职工工资 + 集体所有制职工工资 + 职工从工作单位得到的其他收入 + 个体经营劳动者收入（包括个体雇主和自营者的生产经营性净收益和个体被雇者收入） + 被聘用或留用的离退休人员收入 + 其他就业者收入 + 其他劳动收入 + 财产性收入 + 转移性收入

可支配收入 = 国有经济单位职工收入 + 集体经济单位职工收入 + 其他各种经济类型单位 + 个体经营者的净收益 + 个体被雇者收入 + 离退休在就业人员收入 + 其他就业者收入 + 其他劳动者收入 + 财产性收入 + 转移性收入 - 个人所得税 - 记帐补贴

非劳动（非挣得）收入 = 财产性收入 + 转移收入

劳动（挣得）收入 = 可支配收入（或生活费收入） - 非劳动（非挣得）收入

⁶¹ 国家统计局城市社会经济调查总队，《中国城市住户调查手册》，1996。

附录二：对各方程估计样本与模拟样本的说明

非挣得收入

估计：所有人。

模拟：所有人。

教育选择

估计：1999年，就选择能力来说，应是所有80年代以前出生的人；但有不少80年代出生的人 earnings 也为正数，为了后面模拟，采用90年代以前出生的人为样本。相应的，1988年估计样本为所有80年代以前出生的人。

模拟：1999年样本中所有80年代前出生的人。

工资或所得方程

估计：所有挣得工资（总收入减去非挣得收入）为正数的人，应该也就是所有有能力参与选择的人中，除家务劳动者和待业人员之外其他人。但数据重两者不一致。采用交集为样本。

模拟：所有参与选择（或模拟参与选择）为后三者的人。

参与选择

估计：所有有能力作出参与选择的人，被排除在外的是离退休再就业人员、离退休人员、丧失劳动能力者、待分配者、在校学生、待升学者、其他非就业者（主要指16岁以下非在校生，以及60岁以上男性和55岁以上女性中非离退休或离退休再就业者）。

模拟：同上。

附录三：各种表格

表 1——1988-1999 年间中国城市居民收入不平等程度的变化：总体指数

年份	相对对数偏差	变动系数	对数标准差	基尼系数	Theil 指数	平均对数偏差
1988	0.159	0.451	0.420	0.227	0.089	0.087
1989	0.173	0.514	0.446	0.248	0.108	0.103
1990	0.166	0.498	0.424	0.238	0.101	0.095
1991	0.171	0.527	0.427	0.243	0.107	0.098
1992	0.182	0.587	0.456	0.260	0.125	0.112
1993	0.210	0.635	0.535	0.297	0.156	0.147
1994	0.229	0.680	0.580	0.322	0.179	0.172
1995	0.226	0.655	0.566	0.316	0.170	0.164
1996	0.231	0.668	0.582	0.323	0.178	0.173
1997	0.228	0.667	0.578	0.321	0.176	0.171
1998	0.232	0.678	0.592	0.326	0.183	0.178
1999	0.245	0.716	0.623	0.342	0.201	0.197

注：相对对数偏差 (relative log deviation), 对数标准差 (standard deviation of logs)。

表 2——1988-1999 年间中国城市居民收入不平等程度的变化：分位数组指数

分位数组	分位数组内变量数值占总变量数值的比重											
	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
1	4.39	4.39	4.46	4.54	4.47	3.62	3.15	3.32	3.19	3.22	3.06	2.88
2	6.16	6.13	6.14	6.1	5.62	5.16	4.67	4.73	4.65	4.6	4.54	4.3
3	7.17	6.93	7.03	6.96	7.08	6.28	5.74	5.89	5.63	5.8	5.7	5.51
4	7.93	7.74	7.77	7.76	7.28	6.97	6.84	6.75	6.7	6.68	6.73	6.48
5	8.7	8.47	8.55	8.37	8.45	7.94	7.82	7.87	7.87	8.02	7.75	7.58
6	9.58	9.26	9.36	9.22	9.33	9.16	8.98	9.06	9	8.89	9.01	8.94
7	10.5	10.2	10.27	10.33	9.99	10.3	10.44	10.39	10.71	10.58	10.62	10.38
8	11.86	11.52	11.58	11.56	11.51	12.03	12.21	12.31	12.06	12.35	12.24	12.49
9	13.82	13.74	13.82	13.74	13.86	14.57	15.18	15.15	15.19	15.21	15.17	15.31
10	19.88	21.63	21.02	21.43	22.4	23.97	24.97	24.53	24.99	24.63	25.18	26.12

表 3——个人月收入模拟得到的不平等：采用 1999 年的系数

	个人月均 收入均值	基尼系数	广义 Theil 指数 (0)	广义 Theil 指数 (1)	广义 Theil 指数 (2)
1988	111.8	0.503	0.252	0.505	0.443
1999	200.7	0.673	0.075	0.880	1.139
1、	219.4	0.626	0.464	0.741	0.978
性别	107.1	0.514	0.256	0.521	0.459
年龄及年龄平方	112.3	0.518	0.265	0.526	0.466
经验及经验平方	85.6	0.505	0.222	0.508	0.447
教育程度	157.0	0.556	0.339	0.583	0.568
教育与年龄乘积	82.4	0.511	0.223	0.516	0.464
省份	133.8	0.597	0.368	0.677	0.839
逆 Mill 比率	96.1	0.526	0.259	0.536	0.484
常数项	202.9	0.532	0.351	0.552	0.491
2、	227.3	0.637	0.490	0.777	1.121
性别	110.5	0.529	0.278	0.547	0.517
年龄及年龄平方	115.8	0.533	0.287	0.552	0.524
经验及经验平方	88.2	0.521	0.242	0.532	0.498
教育程度	162.2	0.569	0.362	0.611	0.640
教育与年龄乘积	84.9	0.526	0.243	0.540	0.517
省份	138.6	0.610	0.394	0.714	0.981
逆 Mill 比率	99.0	0.539	0.279	0.559	0.538
常数项	209.8	0.549	0.375	0.582	0.562
3、	99.6	0.562	0.332	0.622	0.568
4、	206.5	0.664	0.569	0.851	1.293
5、	204.4	0.661	0.562	0.841	1.260
6、	95.9	0.573	0.292	0.652	0.591
7、	228.2	0.661	0.561	0.856	1.580
8、	209.4	0.638	0.498	0.771	1.124
9、	99.4	0.563	0.336	0.625	0.568
10、	205.1	0.665	0.572	0.869	1.614
11、	201.6	0.659	0.559	0.834	1.226
12、	177.3	0.366	0.121	0.237	0.217
13、	266.0	0.532	0.343	0.510	0.743

表 4——家庭人均月收入模拟得到的不平等：采用 1999 年的系数

	家庭人均月 收入均值	基尼系数	广义 Theil 指数 (0)	广义 Theil 指数 (1)	广义 Theil 指数 (2)
1988	116.9	0.229	0.091	0.090	0.101
1999	258.9	0.361	0.217	0.230	0.312
1、	229.2	0.421	0.295	0.315	0.445
性别	110.7	0.240	0.100	0.100	0.110
年龄及年龄平方	116.0	0.243	0.103	0.102	0.113
经验及经验平方	88.8	0.248	0.104	0.105	0.117
教育程度	161.8	0.290	0.145	0.142	0.163
教育与年龄乘积	85.3	0.245	0.102	0.103	0.115
省份	140.2	0.387	0.245	0.265	0.364
逆 Mill 比率	99.6	0.255	0.113	0.111	0.123
常数项	207.8	0.252	0.115	0.110	0.120
2、	237.6	0.435	0.318	0.345	0.525
性别	114.1	0.259	0.118	0.118	0.137
年龄及年龄平方	119.6	0.262	0.120	0.120	0.140
经验及经验平方	91.5	0.265	0.120	0.121	0.140
教育程度	167.1	0.307	0.165	0.163	0.199
教育与年龄乘积	87.9	0.262	0.118	0.119	0.138
省份	145.3	0.402	0.269	0.295	0.438
逆 Mill 比率	102.5	0.271	0.129	0.128	0.148
常数项	215.0	0.274	0.135	0.132	0.154
3、	94.6	0.300	0.159	0.159	0.164
4、	216.2	0.439	0.330	0.343	0.500
5、	214.7	0.437	0.330	0.348	0.535
6、	71.6	0.412	0.209	0.316	0.296
7、	238.9	0.453	0.350	0.379	0.639
8、	218.9	0.428	0.308	0.327	0.473
9、	93.8	0.305	0.171	0.168	0.169
10、	215.8	0.440	0.338	0.361	0.616
11、	212.3	0.433	0.327	0.339	0.516
12、	183.3	0.165	0.044	0.045	0.050
13、	278.9	0.340	0.186	0.212	0.310

表 5——对 1988 年城市居民参与选择的模拟结果

可行选择	原始选择	模拟结果 1	模拟结果 2	模拟结果 3
家务劳动或待业	349	868	70	848
国有企业	4794	3932	3085	3885
集体企业	1284	803	433	815
其他所有制企业	181	1005	3020	1060

注：模拟结果 1 是对 88 年数据采用 99 年参与选择参数得到的结果；结果 2 是首先采用 99 年教育选择参数得到模拟的 88 年人群教育程度分布，然后采用 88 年参与选择系数而得到；3 与 1 的不同在于第二步的参与选择参数也采用了 99 年的。

附表 1——1988 年男性参与选择多元 Logit 回归方程

	国有制			集体所有制			其它所有制		
	系数	标准差	P 值	系数	标准差	P 值	系数	标准差	P 值
年龄	0.714	0.099	0.000	0.629	0.104	0.000	0.277	0.113	0.015
年龄平方	-0.009	0.001	0.000	-0.008	0.001	0.000	-0.004	0.001	0.004
小孩个数	-0.027	0.140	0.845	0.028	0.146	0.845	-0.171	0.176	0.332
成年人数	-0.118	0.238	0.622	-0.042	0.248	0.864	-0.258	0.336	0.443
教育程度									
大学本科	0.890	1.956	0.649	-2.153	2.201	0.328	-9.182	5.385	0.088
大专和中专	21.144	4.796	0.000	19.169	4.897	0.000	14.878	.	.
高中	-1.202	1.573	0.445	-2.405	1.620	0.138	-4.542	1.811	0.012
初中	-1.413	1.579	0.371	-1.580	1.617	0.329	-2.752	1.701	0.106
教育与年龄乘积									
大学本科	-0.002	0.046	0.960	0.011	0.052	0.833	0.128	0.084	0.125
大专和中专	0.053	0.094	0.569	0.048	0.097	0.621	0.118	.	.
高中	0.040	0.043	0.349	0.048	0.045	0.287	0.082	0.047	0.079
初中	0.068	0.045	0.132	0.056	0.046	0.228	0.079	0.047	0.092
省份									
辽宁	0.114	1.194	0.924	0.322	1.208	0.790	0.476	1.308	0.716
浙江	-1.885	1.061	0.076	-1.749	1.076	0.104	-1.119	1.177	0.342
广东	-1.553	1.108	0.161	-1.170	1.123	0.297	-0.343	1.223	0.779
四川	-2.238	1.054	0.034	-1.818	1.067	0.089	-1.903	1.179	0.107
陕西	-1.058	1.133	0.350	-0.714	1.147	0.534	-0.736	1.272	0.563
家庭角色									
家长	2.598	0.970	0.007	2.811	0.987	0.004	2.674	1.053	0.011
配偶	0.869	0.926	0.348	1.030	0.946	0.276	0.995	1.063	0.349
常数项	-7.143	2.292	0.002	-6.426	2.378	0.007	-1.618	2.583	0.531
Pseudo R2	0.1877								

附表 2——1999 年男性参与选择多元 Logit 回归方程

	国有制			集体所有制			其它所有制		
	系数	标准差	P 值	系数	标准差	P 值	系数	标准差	P 值
年龄	0.370	0.096	0.000	0.581	0.123	0.000	0.274	0.103	0.007
年龄平方	-0.004	0.001	0.000	-0.007	0.001	0.000	-0.003	0.001	0.002
小孩人数	0.335	0.159	0.035	0.285	0.185	0.124	0.364	0.168	0.030
成年人人数	-0.500	0.194	0.010	-0.074	0.215	0.731	-0.365	0.209	0.081
教育程度									
大学本科	1.020	2.995	0.734	0.382	3.631	0.916	1.133	3.084	0.713
大专和中专	0.995	2.741	0.717	0.684	3.452	0.843	1.938	2.865	0.499
高中	0.264	2.642	0.920	0.953	3.330	0.775	1.620	2.744	0.555
初中	-1.294	2.608	0.620	-0.249	3.279	0.939	-0.433	2.700	0.873
教育与年龄乘积									
大学本科	0.052	0.075	0.488	0.029	0.086	0.735	0.010	0.076	0.891
大专和中专	0.015	0.061	0.803	-0.003	0.075	0.969	-0.056	0.064	0.382
高中	0.005	0.056	0.934	-0.030	0.070	0.673	-0.052	0.058	0.367
初中	0.030	0.054	0.574	0.000	0.068	0.996	-0.005	0.055	0.926
省份									
辽宁	-1.280	0.470	0.007	-0.567	0.526	0.281	-0.397	0.483	0.411
浙江	-0.539	0.445	0.225	0.493	0.494	0.318	-0.657	0.462	0.155
广东	-1.150	0.458	0.012	-1.345	0.545	0.014	-2.114	0.507	0.000
四川	-1.390	0.443	0.002	-0.982	0.503	0.051	-1.504	0.465	0.001
陕西	-0.912	0.486	0.061	0.063	0.538	0.906	0.030	0.497	0.952
家庭角色									
家长	3.099	0.498	0.000	2.708	0.548	0.000	2.366	0.523	0.000
配偶	1.387	0.423	0.001	1.080	0.487	0.027	0.990	0.456	0.030
常数项	-4.321	3.020	0.152	-11.202	3.852	0.004	-3.603	3.172	0.256
Pseudo R2	0.1426								

附表 3——1988 年女性参与选择多元 Logit 回归方程

	国有制			集体所有制			其它所有制		
	系数	标准差	P 值	系数	标准差	P 值	系数	标准差	P 值
年龄	0.423	0.072	0.000	0.438	0.078	0.000	0.054	0.081	0.505
年龄平方	-0.006	0.001	0.000	-0.007	0.001	0.000	-0.002	0.001	0.074
小孩人数	-0.134	0.079	0.090	-0.094	0.083	0.257	0.022	0.118	0.855
成年人人数	0.053	0.168	0.754	0.116	0.173	0.501	0.184	0.235	0.434
教育程度									
大学本科	1.636	2.280	0.473	0.608	2.731	0.824	-5.839	4.682	0.212
大专和中专	21.023	1.222	0.000	19.028	.	.	14.378	3.058	0.000
高中	-0.488	1.230	0.692	-0.379	1.289	0.769	-5.189	1.503	0.001
初中	0.652	1.055	0.537	1.852	1.117	0.097	-2.442	1.200	0.042
教育与年龄乘积									
大学本科	0.068	0.053	0.199	0.015	0.067	0.821	0.137	0.077	0.075
大专和中专	0.089	0.062	0.149	0.067	0.067	0.316	0.169	.	.
高中	0.088	0.037	0.018	0.066	0.039	0.086	0.156	0.043	0.000
初中	0.038	0.025	0.125	-0.010	0.027	0.706	0.075	0.028	0.007
省份									
辽宁	0.195	0.473	0.681	-0.113	0.486	0.816	-0.619	0.648	0.340
浙江	-0.072	0.457	0.875	-0.189	0.468	0.686	-0.527	0.610	0.388
广东	-1.233	0.446	0.006	-1.007	0.456	0.027	-1.192	0.607	0.050
四川	-0.921	0.427	0.031	-0.891	0.438	0.042	-0.907	0.570	0.111
陕西	-0.040	0.494	0.936	-0.087	0.506	0.864	-0.384	0.663	0.563
家庭角色									
家长	2.666	0.463	0.000	3.099	0.478	0.000	2.286	0.617	0.000
配偶	0.921	0.355	0.009	1.538	0.374	0.000	0.784	0.525	0.135
常数项	-5.256	1.631	0.001	-6.537	1.760	0.000	-0.400	1.857	0.830
Pseudo R2	0.218								

附表 4——1999 年女性参与选择多元 Logit 回归方程

	国有制			集体所有制			其它所有制		
	系数	标准差	P 值	系数	标准差	P 值	系数	标准差	P 值
年龄	0.356	0.081	0.000	0.690	0.108	0.000	0.397	0.096	0.000
年龄平方	-0.005	0.001	0.000	-0.008	0.001	0.000	-0.005	0.001	0.000
小孩人数	-0.106	0.113	0.347	-0.187	0.134	0.163	-0.181	0.129	0.160
成年人人数	-0.366	0.153	0.017	-0.187	0.170	0.270	-0.078	0.163	0.631
教育程度									
大学本科	-0.152	2.108	0.943	1.918	2.606	0.462	0.878	2.260	0.698
大专和中专	1.188	1.706	0.486	3.958	2.308	0.086	1.965	1.927	0.308
高中	1.016	1.492	0.496	3.434	2.092	0.101	3.211	1.683	0.056
初中	-0.990	1.396	0.478	1.915	1.964	0.329	2.028	1.563	0.194
教育与年龄乘积									
大学本科	0.132	0.068	0.051	0.037	0.076	0.626	0.053	0.071	0.455
大专和中专	0.070	0.046	0.134	-0.037	0.058	0.520	-0.009	0.051	0.867
高中	0.029	0.035	0.415	-0.043	0.048	0.364	-0.057	0.040	0.151
初中	0.051	0.031	0.101	-0.018	0.043	0.676	-0.040	0.035	0.242
省份									
辽宁	-1.636	0.361	0.000	-0.804	0.408	0.049	-0.421	0.397	0.289
浙江	-1.260	0.354	0.000	0.418	0.394	0.289	-0.536	0.394	0.174
广东	-1.649	0.365	0.000	-1.644	0.438	0.000	-1.300	0.422	0.002
四川	-2.158	0.353	0.000	-1.202	0.399	0.003	-1.520	0.399	0.000
陕西	-0.750	0.405	0.064	0.184	0.448	0.682	0.832	0.434	0.055
家庭角色									
家长	2.197	0.382	0.000	1.589	0.421	0.000	1.273	0.411	0.002
配偶	0.568	0.305	0.063	0.000	0.349	0.999	-0.269	0.334	0.420
常数项	-4.150	2.009	0.039	-13.665	2.780	0.000	-7.091	2.334	0.002
Pseudo R2	0.169								

附表 5——1988 年分行业工资或所得回归方程（纠正自选择偏差）

	国有制			集体所有制			其它所有制		
	系数	标准差	P 值	系数	标准差	P 值	系数	标准差	P 值
男性	0.005	0.018	0.797	0.209	0.045	0.000	0.567	0.214	0.009
年龄	0.031	0.010	0.002	0.081	0.017	0.000	0.063	0.102	0.536
年龄平方	0.000	0.000	0.000	-0.001	0.000	0.000	-0.001	0.001	0.543
经验	0.050	0.005	0.000	0.051	0.008	0.000	0.083	0.035	0.021
经验平方	-0.001	0.000	0.000	-0.001	0.000	0.003	-0.003	0.001	0.004
教育程度									
大学本科	-0.585	0.157	0.000	1.213	0.443	0.006	(dropped)		
大专和中专	-0.480	0.161	0.003	0.229	0.419	0.585	-0.330	3.137	0.916
高中	-0.360	0.141	0.010	0.185	0.214	0.386	0.417	1.917	0.828
初中	-0.131	0.137	0.340	0.152	0.195	0.437	0.224	0.979	0.819
年龄与教育乘积									
大学本科	0.012	0.003	0.000	-0.026	0.010	0.013	0.002	0.019	0.934
大专和中专	0.007	0.003	0.034	0.002	0.010	0.828	-0.010	0.061	0.872
高中	0.008	0.003	0.019	0.001	0.006	0.892	-0.035	0.058	0.545
初中	0.002	0.003	0.587	-0.001	0.005	0.843	-0.019	0.020	0.356
省份									
辽宁	0.116	0.030	0.000	0.229	0.070	0.001	-0.826	0.576	0.155
浙江	-0.075	0.030	0.013	-0.080	0.068	0.240	-1.266	0.554	0.024
广东	-0.085	0.033	0.011	-0.307	0.071	0.000	-1.408	0.540	0.010
四川	-0.087	0.030	0.004	-0.081	0.066	0.224	-1.470	0.528	0.006
陕西	0.059	0.032	0.062	0.121	0.073	0.096	-0.796	0.592	0.181
逆 Mill 比率	-0.503	0.074	0.000	-0.052	0.134	0.696	0.514	0.678	0.450
常数项	6.691	0.231	0.000	5.107	0.390	0.000	5.627	1.368	0.000
Adj R-squared	0.340			0.342			0.223		

附表 6——1999 年分行业工资或所得回归方程（纠正自选择偏差）

	国有制			集体所有制			其它所有制		
	系数	标准差	P 值	系数	标准差	P 值	系数	标准差	P 值
男性	0.028	0.019	0.133	0.087	0.079	0.273	0.251	0.049	0.000
年龄	0.032	0.018	0.071	0.063	0.053	0.233	0.095	0.030	0.002
年龄平方	0.000	0.000	0.024	-0.001	0.001	0.115	-0.001	0.000	0.002
经验	0.020	0.008	0.012	0.029	0.018	0.111	0.014	0.008	0.077
经验平方	0.000	0.000	0.440	0.000	0.000	0.554	0.000	0.000	0.382
教育程度									
大学本科	0.219	0.428	0.609	0.712	0.936	0.447	1.639	0.698	0.019
大专和中专	0.140	0.430	0.745	-0.085	0.955	0.929	1.511	0.732	0.039
高中	0.325	0.427	0.447	0.372	0.915	0.685	2.030	0.690	0.003
初中	0.038	0.428	0.929	-0.526	0.885	0.553	1.148	0.668	0.086
年龄与教育乘积									
大学本科	-0.002	0.009	0.812	-0.011	0.021	0.603	-0.020	0.015	0.179
大专和中专	-0.004	0.009	0.672	0.006	0.021	0.776	-0.028	0.017	0.100
高中	-0.009	0.009	0.337	-0.008	0.020	0.680	-0.041	0.015	0.007
初中	-0.003	0.009	0.739	0.011	0.019	0.574	-0.022	0.014	0.118
省份									
辽宁	1.007	0.033	0.000	0.978	0.123	0.000	0.791	0.109	0.000
浙江	-0.153	0.031	0.000	-0.110	0.165	0.505	-0.463	0.099	0.000
广东	-0.549	0.031	0.000	-0.374	0.149	0.012	-0.626	0.132	0.000
四川	-0.148	0.030	0.000	-0.199	0.127	0.116	-0.525	0.107	0.000
陕西	0.417	0.038	0.000	0.310	0.135	0.022	0.228	0.123	0.064
逆 Mill 比率	-0.769	0.071	0.000	0.295	0.225	0.189	0.463	0.180	0.010
常数项	7.441	0.496	0.000	5.568	1.483	0.000	4.309	0.860	0.000
Adj R-squared	0.497			0.357			0.434		

注：附表 1-6 中，教育程度虚拟变量的基底（benchmark）都是小学文化程度或文盲；省份虚拟变量的基底是北京。

附表 7——1988 年教育程度选择多元 Logit 方程

	大学本科			大专和中专			高中			初中		
	系数	标准差	P 值	系数	标准差	P 值	系数	标准差	P 值	系数	标准差	P 值
男性	1.466	0.091	0.000	0.523	0.085	0.000	0.725	0.068	0.000	0.651	0.059	0.000
年龄	-0.044	0.006	0.000	-0.073	0.006	0.000	-0.034	0.005	0.000	-0.041	0.004	0.000
成年人数	-0.333	0.083	0.000	-0.202	0.076	0.008	-0.363	0.063	0.000	-0.242	0.052	0.000
出生年代												
1950-1959	0.440	0.154	0.004	-0.280	0.153	0.068	1.529	0.139	0.000	1.095	0.116	0.000
1960-1969	1.692	0.288	0.000	0.668	0.289	0.021	3.471	0.270	0.000	1.991	0.250	0.000
1970-1979	-5.608	0.472	0.000	-4.731	0.286	0.000	-0.965	0.223	0.000	-1.502	0.173	0.000
省份												
辽宁	0.259	0.204	0.204	0.206	0.210	0.326	0.058	0.163	0.723	-0.092	0.141	0.512
浙江	0.079	0.201	0.695	0.129	0.204	0.527	-0.107	0.158	0.501	-0.215	0.137	0.116
广东	-0.431	0.220	0.050	-0.088	0.216	0.685	-0.325	0.167	0.051	-0.209	0.142	0.140
四川	-0.127	0.199	0.522	0.068	0.201	0.735	-0.113	0.155	0.464	-0.158	0.133	0.233
陕西	0.279	0.212	0.188	0.364	0.215	0.090	0.139	0.168	0.409	-0.084	0.146	0.567
常数项	1.195	0.384	0.002	2.856	0.386	0.000	1.266	0.332	0.000	2.441	0.261	0.000
Pseudo R2	0.104											

注：出生年代虚拟变量的基底 (benchmark) 是“ 1950 年以前 ”；省份虚拟变量的基底是北京 (所有回归中都是)。回归是以具有小学文化程度的人及文盲为比较的基础 (base)。

附表 8——1999 年教育程度选择多元 Logit 方程

	大学本科			大专和中专			高中			初中		
	系数	标准差	P 值	系数	标准差	P 值	系数	标准差	P 值	系数	标准差	P 值
男性	1.210	0.077	0.000	0.553	0.082	0.000	0.742	0.071	0.000	0.611	0.066	0.000
年龄	-0.068	0.007	0.000	-0.054	0.007	0.000	-0.030	0.006	0.000	-0.043	0.005	0.000
成年人数	-0.445	0.076	0.000	-0.461	0.082	0.000	-0.432	0.071	0.000	-0.496	0.066	0.000
出生年代												
1950-1959	0.732	0.151	0.000	0.549	0.160	0.001	1.857	0.151	0.000	1.188	0.131	0.000
1960-1969	2.364	0.277	0.000	1.759	0.288	0.000	3.522	0.275	0.000	1.928	0.258	0.000
1970-1979	2.398	0.437	0.000	2.174	0.448	0.000	3.893	0.433	0.000	1.768	0.414	0.000
1980-1989	-4.679	0.348	0.000	-3.422	0.353	0.000	-0.683	0.313	0.029	-1.713	0.263	0.000
省份												
辽宁	-0.841	0.143	0.000	-0.595	0.153	0.000	-0.463	0.135	0.001	-0.970	0.131	0.000
浙江	-0.857	0.135	0.000	-0.751	0.146	0.000	-0.732	0.130	0.000	-0.311	0.120	0.010
广东	-0.824	0.153	0.000	-0.724	0.166	0.000	-0.500	0.146	0.001	-0.659	0.138	0.000
四川	-1.061	0.142	0.000	-0.787	0.151	0.000	-0.970	0.136	0.000	-0.808	0.127	0.000
陕西	-1.246	0.151	0.000	-0.970	0.162	0.000	-0.768	0.141	0.000	-0.660	0.131	0.000
常数项	4.638	0.445	0.000	3.937	0.466	0.000	2.231	0.421	0.000	4.047	0.357	0.000
Pseudo R2	0.103											

附表 9——非挣得收入 Tobit 回归方程

	1988			1999		
	系数	标准差	P 值	系数	标准差	P 值
男性	11.005	11.866	0.354	-283.584	48.419	0.000
年龄	14.476	2.271	0.000	-15.132	9.717	0.119
年龄平方	0.031	0.026	0.242	1.163	0.102	0.000
成年人人数	-50.953	10.682	0.000	-247.233	52.506	0.000
小孩个数	-66.287	6.416	0.000	-191.676	33.829	0.000
教育程度						
大学本科	30.732	23.631	0.193	164.156	90.174	0.069
大专和中专	17.884	23.346	0.444	452.641	96.247	0.000
高中	72.050	18.552	0.000	258.777	86.279	0.003
初中	59.717	16.481	0.000	241.667	79.920	0.003
省份						
辽宁	33.064	26.450	0.211	-857.474	81.246	0.000
浙江	-34.898	25.824	0.177	-421.501	70.333	0.000
广东	-89.879	27.452	0.001	-1893.918	97.162	0.000
四川	-70.516	25.346	0.005	-1807.084	85.681	0.000
陕西	-6.437	27.499	0.815	-1208.718	87.093	0.000
家庭角色						
家长	181.553	23.317	0.000	1408.960	101.988	0.000
配偶	17.021	23.490	0.469	525.965	102.670	0.000
常数项	-262.661	43.525	0.000	-1688.718	197.340	0.000
Pseudo R2	0.028			0.066		

附录四：模拟步骤流程图

注：下图中PC表示参与选择 (participation choice), UC表示不可观察的个人特征 (unobservable characteristics), 各参数下标中, 1表示99年, 0表示88年。

