

中国城市居民收入分布的变化：1988—1999年

万定山*

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情况进行分解，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。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，其中省份和人力资本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的影响。

关键词 收入分布，禀赋分布，参数结构

一、引言

20多年来的中国经济改革无疑已经取得了重大成功。与改革所取得的巨大成就相伴随的，是中国居民个人收入分布的显著变化。

中国的城市经济改革始于20世纪80年代中期，至90年代初中期为一段落，从90年代中后期开始，更为激进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。如果把关于不平等问题的研究看作是对改革政策的评估，那么，有意义的问题至少应包括两个方面：改革的结果如何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)用10省城市住户调查数据算得1987年收入分布基尼系数为0.26；对较具代表性的1988、1995年，国家统计局计算的两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)；社科院的计算结果则分别是0.233和0.332(城镇家庭人均可支配收入口径，李思勤，1999)。这些研究的基本结论相近，即自20世纪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呈持续上升趋势，而进入90年代以

* 博时基金管理有限公司。通信地址：中国深圳市福田区深南大道7088号招商银行大厦29层，博时基金管理有限公司，518040；电话：(0755)83169999-1230；E-mail: wands@boshi.com.cn。感谢导师赵忠以及赵耀辉一直以来的悉心指导和关心，感谢Martin Fournier的有益建议以及宋晓青的诸多帮助，同样感谢姚洋及两位匿名审稿人中肯且极富建设性的评论。文章的完成直接受益于赵耀辉和香港中文大学张俊森提供的数据库。同北京大学中国经济研究中心劳动经济学小组诸位同学的讨论亦令我受益匪浅。文中一切错误都由作者负责。

后,这一趋势进一步加速。

与总体趋势同样重要的是各类人群的绝对变化。研究表明(赵人伟、李实,1999; Meng, 2001),尽管个体之间的差异巨大,但1988至1995年间的改革为绝大部分人带来了福利增加。然而,进入20世纪90年代后期以来,处于收入分布最低5个百分点的人其境遇却在绝对地变差。这些研究为目前国内一些学者的观点提供了佐证:中国的城市经济改革以20世纪90年代中期为分水岭,此前的改革是帕累托改进式的,而之后的改革则更多是财富的重新分配。

个人收入分布的显著变化引起了政府及学术界的广泛关注。为了探究收入分布变化背后的原因,学者们已展开了广泛的研究。就内容而言,各种分析中所讨论的主要因素包括:非国有部门的快速发展、企业改革与改制、个人特征、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)、垄断等。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。

以上这些研究基本上是描述性的。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 年的三次调查数据,在方法论上她借鉴了 Fields (1998) 的分解方法,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 年以及 1995—1999 年,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的 5% 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。

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 年和 1995 年调查数据的分析,他们发现人力资本变量是能够识别的最重要的逆平等化因素,其次是省份变量。

这些文献是对中国城市居民收入分布变动的初步分析,他们的共同不足在于:缺乏一个统一的框架以分析个人或家庭收入的决定过程。

在本文中,我们希望进一步深入探讨的问题包括三个方面:从个人最优化模型的角度出发,考虑外生性约束条件的变化、行为方式的变化分别在多大程度改变了居民收入分布,这样区分的涵义在于:(1) 约束条件是我们部分可控的;(2) 克服经典 Mincer 工资估计方程的内生性问题,将个人就业选择

行为模型化，从而检验在控制不同因素的情况下个人特征等因素对收入分布的影响是否将显著地不同；(3) 分离个人特征的分布及其对应报酬率对收入分布变动的两种不同影响（如人力资本对收入的影响），从而更准确地指出政策建议的方向。这就需要我们构建一个完整的经济模型系统。与以前的研究相比，本文具有以下几个重要特点：

我们的模型更具结构性而非先验设定。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经济学理论模型为基础，这一理论模型是一般均衡性质的，同时，我们的模型以家庭效用函数模型为基础来构建参与选择及教育选择行为方程。

我们首次采用了微观模拟的方法来对中国的微观收入数据进行分析。与模型设定相结合，对于整个收入决定过程中的各种因素——主要是人群禀赋和各种参数结构，我们所采用的方法能够在保持其他各种相关因素不变的情况下，单独考察某个因素对收入分布差别的定量影响，从而评估其相对重要性。这一方法来自 BFL (2002)，我们在其基础上尝试使用 Lee (1983) 的方法来纠正自选择偏差问题，从而对其作出改进。

从数据质量来看，我们的数据来自国家统计局城调总队的年度家户调查系统，其良好的代表性及调查方法的成熟性提高了数据的可信度。

文章结构如下：第二部分是对数据及 1988—1999 年间中国城市居民收入分布变化总体趋势的简要描述；第三部分阐述本文所用模型，包括一般性框架、模型设定和具体估计方法；第四部分是对模拟方法和步骤的说明；第五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。

二、中国城市居民收入分布变化的总体趋势

(一) 数据

本文所用数据来自国家统计局城调队城市家计调查 (UHS)。由于自 1992 年起 UHS 只涵盖北京、辽宁、浙江、广东、四川和陕西六省，为使各年数据一致，我们仅选用六省 1988—1999 间 12 年的数据。这确实减少了样本量，但并未削弱其代表性，对于一些基本统计量的结果也影响不大。对数据的进一步说明，详见附件。

(二) 1988—1999 年间收入分布变动总体趋势

1. 收入指标

对收入概念的界定一直存有争议。在我们使用的 12 年 UHS 数据中，前四年个人收入指标是生活费收入，从 1992 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某种低估，

主要包括城市居民得到的各种福利如医疗、住房、教育、养老等,但数据的缺陷应不会给不平等程度的估测带来严重偏差,福利分配的主要依据是等级制度,本身即具强烈的平均主义色彩。对收入指标内容的详细说明见附录。

2. 总体趋势

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名义月可支配收入。¹我们选取的样本是所有样本,未作统计上的样本剔除。

由表1可见,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,上升幅度超过50%。特别地,进入20世纪90年代以来,不平等程度急速上升,1992、1993、1994三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,1994—1995年间基尼系数有所下降,这与国家统计局计算的结果是一致的。²表1中其他各指标所反映的事实与基尼系数基本相同。

表2给出了按分位数组(quantile group)计算的结果,以进一步分析1988—1999年间收入分布中不同百分点人群状况的相对变动。

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)。

¹ 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)。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。

² 国家统计局公布的1994年和1995年城市居民基尼系数分别为0.30和0.28。

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

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，这表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，1993 和 1994 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。

在对居民收入分布变动的总体情况有所了解后，一个更重要的问题是，究竟何种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分解的不平等测度。可以证明，如果加上“良好”性质的约束，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）。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，这一方法无法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。为解决这些问题，需要使用更细致的分解方法。

三、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 和 Lustig (1998) 的世界银行研究建议。这一方法实质上是著名的 Oaxaca-

Blinder 分解公式的推广。³ 与工资或所得⁴ 相比, 收入概念内涵更丰富, 其决定也更复杂。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, 如果考虑的是家庭收入的决定, 那么家庭结构也应同时被看作是选择的结果。

(一) 分解方法: 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:

$$f(y) = \iiint F(y, T) dT.$$

具体来说, 考虑某个家庭在某一时期的收入决定函数:

$$y_{it} = Y[\omega_{it}, v_{it}(\omega_{it}, \mu_{it}; \gamma_t), \epsilon_{it}; \lambda_t; \beta_t], \quad (1)$$

y_{it} 表示第 i 个家庭在第 t 时期的收入, ω_{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 ω_{it} 、结构性选择参数 γ_t 以及随机扰动项 μ_{it} 等共同决定。 ϵ_{it} 代表不可观察的家庭特征, λ_t 是参与选择的参数结构, β_t 则是工资方程的参数结构, 这些变量都是向量。相应地, 第 t 时期收入的分布可表示为:

$$D_t = D_t[\omega_t, v_t(\omega_t, \mu_t; \gamma_t), \epsilon_t; \lambda_t; \beta_t]. \quad (2)$$

这里 ω_t 、 v_t 、 μ_t 、 ϵ_t 是由 t 时期收入分布中所有人的相关向量构成的矩阵。从上式出发, 借助模拟方法可对不同收入分布作全面比较, 并评估各影响因素的相对重要性。

(二) 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:

$$y = \sum_{i=1}^n (L_i \omega_i + y_{i0}), \quad (3)$$

³ 参见 Oaxaca(1973)和 Blinder(1973)。

⁴ 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。

其中，个人 i 的工资率 w_i 为：

$$w_i = f(X_i, \epsilon_i), \quad (4)$$

个人 i 的工作时间 L_i 为：

$$L_i = \text{Sup}(0, g(Z_i, \eta_i)). \quad (5)$$

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，其包含变量可重复但不必定相同， ϵ_i 、 η_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，它决定了个人 i 是选择不工作（劳动时间为 0）还是工作，这一方程的设定是简化式的，因为工资率没有进入。

（三）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。由于缺少关于个人工作时间的详细数据，我们无法估计（4）、（5）方程，而以下式作为替代：

$$y_h = \frac{1}{n_h} \left[\sum_{i=1}^{n_h} \sum_{j=1}^J (I_{hi}^j y_{hi}^j + y_{hi0}) \right], \quad (6)$$

h 表示家庭， n_h 是家庭成员个数， j 是参与选择，共 J 个，在本文中具有选择能力的家庭成员面临四个选择：家务劳动或待业在家⁵、国有企业工作、集体企业工作或其他所有制企业工作⁶。家庭收入等于成员收入总和，而各成员收入包括挣得收入和非劳动收入。⁷ y_{hi}^j 表示第 i 个人在第 j 选择中所能挣到的收入， I_{hi}^j 是 i 选择 j 的示性函数。 y_{hi0} 是第 i 个人的非挣得收入，包括财产性收入、转移收入和家庭副业生产收入。

从（6）式出发，一个家庭的总体收入由以下一组方程决定：

$$\log y_i^j = \alpha_j + X_i \beta_j + \epsilon_{ij}, \quad (7)$$

如果 i 的选择为 j ，则

⁵ 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，笔者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。

⁶ 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产业、第二产业、第三产业三类，以及家务劳动者和待业者。根据这种分类方法，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。

⁷ 变量说明详见附录。

$$P_{ij} = \text{Prob}(U_{ij} > U_{ik}) = \text{Prob}(Z_i \lambda_j + \eta_{ij} - Z_i \lambda_k - \eta_{ik} > 0),$$

$$\text{for all } k \neq j. \quad (8)$$

(7)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, 取自然对数形式。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。 X_i 是第 i 个人的个人特征构成的向量, ϵ_{ij} 是随机扰动项, 在这里被理解为不可观察到的个人特征, α_j 、 β_j 是在第 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。⁸ 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、行业虚拟变量、省份和家庭角色虚拟变量以及逆 Mill 比率。^{9、10}

(7)式是多元 Logit 个人参与选择方程, 采用简化式。 P_{ij} 是第 i 个人选择第 j 个选择的概率, $U_{ik} = Z_i \lambda_k + \eta_{ik}$ 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, λ_k 是与第 k 个选择对应的参数结构, η_{ik} 表示 i 选择第 k 个选择得到的随机效用, 服从双指数分布。 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, 以及省份和家庭角色虚拟变量。可计算得到 P_{ij} :

$$P_{is} = \text{Pr}\{j = s\} = P^s(Z_i, \lambda) = \frac{e^{Z_i \lambda_s}}{e^{Z_i \lambda_s} + \sum_{j \neq s} e^{Z_i \lambda_j}}. \quad (9)$$

进一步地, 再采用多元 Logit 模型对个人教育程度选择行为进行估计:

$$\text{MLE}(e_i | \omega_i) \rightarrow \gamma, \quad (10)$$

e_i 是第 i 个人的教育程度, ω_i 是一系列完全外生的家庭和个人特征变量。¹¹ 在 ω_i 中, 当事人个人的特征包括年龄、地区和性别, 家庭的特征是成年人的个数。¹² γ 是回归所得的参数。

⁸ 方程(8)是 hedonic 形式的工资方程, 参见 Rosen (1967), 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。

⁹ 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。

¹⁰ 1988 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, 1999 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: 1999 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。

¹¹ 就我们所选定的 ω_i 变量而言, 这种决定关系可能并不具有很强的因果性解释, 纯统计意义上的解释更多, 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估计。从这一点上来说, 第五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。

¹² 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数相对于现有的样本数量而言过大, 因此放弃。

对个人非挣得收入，可采用 Tobit 形式的方程来估计：

$$\text{Tobit}(y_{i0} | e_i, \text{age}_i, \text{region}_i, \text{sex}_i, n_{ah}) \rightarrow \delta, \quad (11)$$

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。 δ 是回归所得的参数。

(四) 模型的估计

需要估计的模型有 (7)、(8)、(10) 和 (11)，其中对工资方程 (7) 的估计是重点。

Lee (1983) 给出了多元选择情形下纠正自选择偏差的一般性方法，在此我们沿用这种方法。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被修正为：

$$\log y_i^s = \alpha_s + X_i \beta_s - \sigma_s \rho_s \phi(J(Z_i \lambda_s)) / F(Z_i \lambda_s) + \omega_i^s, \quad (12)$$

其中， σ_s 是工资方程 (7) ϵ_{is} 的标准差， ρ_s 是 ϵ_{is} 与参与选择方程扰动项 η_{is} 的相关系数， ω_i^s 是工资方程被修正后得到的新随机扰动项，其条件期望为 0，条件方差相异。 $J = \Phi^{-1} \cdot F$ ， 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 $\xi_i^s = \max_{j \neq s} U_{ij} - \eta_{is}$ 的累积分布函数。¹³ $\phi(J(Z_i \lambda_s)) / F(Z_i \lambda_s)$ 是逆 Mill 比率。

需要注意的是，方程 (7) 和 (8) 中设定的工资方程和参与选择方程都是简化式的。这完全是出于可操作性的考虑。

对 (10) 和 (11) 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去解释，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。¹⁴

对参与选择方程 (8) 的估计分性别进行，原因在于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。¹⁵ 对非挣得收入方程 (11) 的估计是用全部样本进行的。¹⁶

¹³ $F(\xi_i^s) = \exp(\xi_i^s) / [\exp(\xi_i^s) + \sum_{j \neq s} \exp(X_j \beta_j)]$.

¹⁴ 在今后的课题中，我们将对此方法进行进一步研究。

¹⁵ 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，由于后面要做模拟，所以未采用这一分类估计方式。

¹⁶ 对各计量模型的估计样本及模拟样本的说明因篇幅所限略去，感兴趣的读者可与作者联系。

四、模 拟

在得到估计结果之后,应进行模拟以得到所需要的反事实的分布。以下说明模拟步骤。

第一步是想看工资方程系数,也即人力资本要素价格对两年间收入分布差别的影响。为此,可将估计的1999年工资方程的系数代入1988年工资方程中,其他所有因素保持不变。进一步地,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 ω_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 ω_i^s 的分布就可以了,在正态分布假设下,只需对 ω_i^s 作倍数调整: $\widehat{\omega}_i^s = \frac{\sigma_\tau}{\sigma_t} \omega_i^s$ 。 ω_i^s 是 t 时期第 i 个人在 s 选择下工资方程中的扰动项, σ_τ 和 σ_t 则分别是 τ 年和 t 年扰动项对应的标准差。

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1999年参与选择方程的参数结构代入1988年的参与选择方程,这样可以得到模拟的 $Z_i\lambda_j$, U_{ij} 中的另一部分是随机效用项 η_{ij} ,对它的模拟是唯一的困难所在。我们要模拟样本中所有具备选择能力的人群在每种选择下的效用 U_{ij} ,但必然有部分人其随机效用项是无法直接观察到的,这是因为我们所能直接观察到的只是各种选择中目前已经在位的选择者,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,¹⁷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, η_{ij} 的累积分布函数为:

$$F(\eta_{ij} | pc = j) = F(\eta_{ij} | Z_i\lambda_j + \eta_{ij} > \max_{s \neq j}(Z_i\lambda_s + \eta_{is})), \quad (13)$$

而未被选择的选择 k 对应的随机扰动项的分布为:

$$F(\eta_{ik} | pc = j) = F(\eta_{ik} | Z_i\lambda_j + \eta_{ij} > Z_i\lambda_s + \eta_{is}), \quad k \neq j. \quad (14)$$

在得到全部的 $Z_i\lambda_j$ 以及 η_{ij} 后,可以得到模拟的 U_{ij} ,对每个人所有的 U_{ij} 进行

¹⁷ 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1989—1999年、或1988—1998年的数据组合重复所有操作,看结论的变动是否显著。

比较，就可以得到模拟的这个人的参与选择。

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该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。

五、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 12 年间中国城市居民收入分布变动进行分解的结果列在表 3 中，全部结论均根据家庭人均月收入模拟结果计算得到。我们对 1999 年的原始个人月收入数据进行了通货膨胀平减处理，以转换成 1988 年的可比价格。¹⁸

表 3 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	115.6	0.229	0.091	0.090	0.101
1999	233.1	0.345	0.199	0.205	0.264
1. $\alpha\beta$	223.8	0.412	0.282	0.303	0.429
性别	111.2	0.240	0.100	0.100	0.110
年龄及 年龄平方	97.8	0.236	0.097	0.096	0.105
经验及 经验平方	88.7	0.251	0.107	0.109	0.122
教育程度	108.1	0.245	0.104	0.104	0.115

¹⁸ 价格指数数据来自《中国市场统计年鉴》公布的居民生活费用价格指数。我们选用的是分省市指数，1992 年选用的是分省总指数。1989 年的数据由《中国统计年鉴》查得。

(续表)

		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)
	行业	87.8	0.249	0.106	0.107	0.120
	省份	137.9	0.377	0.234	0.251	0.341
	逆 Mill 比率	99.2	0.254	0.113	0.111	0.123
	常数项	359.2	0.267	0.139	0.125	0.134
2.	$\alpha\beta\sigma$	230.9	0.425	0.301	0.327	0.492
	性别	114.2	0.257	0.115	0.116	0.133
	年龄及 年龄平方	100.4	0.252	0.112	0.112	0.127
	经验及 经验平方	91.1	0.266	0.121	0.123	0.141
	教育程度	111.1	0.261	0.119	0.119	0.138
	行业	90.0	0.264	0.119	0.121	0.140
	省份	142.3	0.391	0.253	0.277	0.401
	逆 Mill 比率	101.7	0.268	0.126	0.125	0.144
	常数项	370.7	0.287	0.158	0.145	0.165
3.	λ	93.4	0.301	0.161	0.162	0.164
4.	$\lambda\alpha\beta$	211.0	0.424	0.305	0.319	0.458
5.	$\lambda\alpha\beta\sigma$	208.2	0.423	0.305	0.322	0.488
6.	γ	71.0	0.414	0.194	0.317	0.295
7.	$\gamma\alpha\beta$	222.7	0.447	0.337	0.371	0.616
8.	$\gamma\alpha\beta\sigma$	209.0	0.422	0.295	0.313	0.434
9.	$\lambda\gamma$	93.1	0.307	0.173	0.168	0.170
10.	$\lambda\gamma\alpha\beta$	207.2	0.429	0.314	0.329	0.474
11.	$\lambda\gamma\alpha\beta\sigma$	207.1	0.426	0.310	0.327	0.494
12.	δ	185.5	0.158	0.041	0.042	0.046
13.	$\delta\lambda\gamma\alpha\beta\sigma$	276.7	0.331	0.176	0.200	0.292

表3中包括了根据家庭人均月收入计算的各种收入分布测度指标,其中家庭人均月收入属于水平指标,其他则属于分散度指标。

我们对表3结果进行详细分析。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11.6个百分点,上升幅度约50%。¹⁹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表3中标号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质的回报率与1999年相同,那么1988年的基尼系数将上升到0.412,约上升18个百分点,超过1999年真实基尼系数6个点。这一结论表明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

¹⁹ 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3给出的结果中,已对1999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。

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的存在。²⁰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。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。供需双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较高的回报率，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人力资本变量。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15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。分析表明，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，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。

在重要性上次于省份虚拟变量系数的依次是经验、行业及教育变量，它们对两年收入分布基尼系数的影响分别为2.2、2.0、1.6个百分点。因此，以经验和教育为内涵的人力资本因素的系数对1988年基尼系数有约4个百分点的影响。这一作用完全由人力资本回报率的改变所导致。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%，这里的结论相似。Meng (2001) 用家庭劳动力的平均年龄及其平方、平均教育年数及其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。Meng (2001)

²⁰ 这一观点来自周其仁(2003)的一次谈话。

区分了两个不同阶段：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，这将导致内生性偏差。

行业差距的变化将使得 1988 年的基尼系数上升 2 个百分点，约占两年基尼系数总差别的 17%。这表明行业差距扩大了。

其他因素如性别、年龄及年龄平方对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 8.6% 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 比率的影响也在 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，其中省份虚拟变量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 115.6 元增加到 137.9 元，证明 12 年间各地区收入水平普遍提高；教育虚拟变量系数的重要性其次，对应的模拟结果是 108.1 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 88.7 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。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。²¹ 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

²¹ 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和不可观察个人特征回报率的变化带来的影响。

征的影响也有与上述相似的结论。

表 4 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	876	64	837
国有企业	4794	3876	3076	3830
集体企业	1284	824	435	857
其他所有制企业	181	1032	3033	1084

注：模拟结果 1 是对 1988 年数据采用 1999 年参与选择参数得到的结果；结果 2 是首先采用 1999 年教育选择参数得到模拟的 1988 年人群教育程度分布，然后采用 1988 年参与选择系数而得到；结果 3 与结果 1 的不同在于第二步的参与选择参数也采用了 1999 年的。表 5 结果则全部反之。

表 5 对 1999 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	415	130	101	131
国有企业	4627	5616	2607	5605
集体企业	918	1005	264	1015
其他所有制企业	895	104	3883	104

在表 3 中，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 1988 年的参与选择行为与 1999 年一样，那么 1988 年的收入基尼系数将上升约 7 个百分点，升幅比较大，其中原因可以从模拟得到的参与选择结果来分析（表 4 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，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 的结果表明改变参与选择行为参数后，得到的 1988 年真实平均家庭人均月收入较实际水平有所下降，这主要是由于模拟结果 1 中选择家务劳动或待业的人数大大增加了。²² 4 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 1988 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。

²² 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。

表 6 教育选择模拟结果

教育程度	1988				1999			
	原始		模拟		原始		模拟	
	人数	百分比	人数	百分比	人数	百分比	人数	百分比
大学本科及大专	830	7.85	1706	16.13	2049	17.39	569	4.83
中专	840	7.94	370	3.5	1212	10.28	357	3.03
高中	2048	19.36	672	6.35	2681	22.75	1537	13.04
初中	3472	32.83	1505	14.23	3511	29.79	3612	30.65
小学及其它	3387	32.02	6324	59.79	2333	19.79	5711	48.46
总计	10577	100	10577	100	11786	100	11786	100

在表 3 中, 标号 6、7、8 行是考虑教育选择内生化的模拟结果。如果 1988 年的教育选择参数结构与 1999 年相同, 那么收入的基尼系数将上升到 0.414。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。表 6 是教育选择的模拟结果, 从中可以看到, 如果 1988 年的教育结构与 1999 年相同, 那么拥有大学本科及大专文化、小学及其他文化程度的人的比例将上升很多, 而处于中间部分的人群却大量减少了, 这意味着 12 年间我国城市居民的教育分布具有两极化的趋势。从表 4 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, 这也是与表 3 中标号 3 的结果相比, 标号 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。标号 7、8 的结果则表明, 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。

表 3 中标号 9、10、11 是以上两步的综合。比较标号 3 与标号 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 4 中模拟结果 3 与模拟结果 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。

表 3 中标号 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 7 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底来自哪一方面，可能是财产收入，也可能是转移收入，对非挣得收入进一步的细分研究应该可以弄清根源所在。表3中标号13的结果则表明，1988和1999两年间收入基尼系数约11个百分点的差别中，有约10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。

六、结论及政策涵义

本文首先构建了一系列结构性计量模型，并在此基础上通过微观模拟的方法来分析1988—1999年间中国城市居民收入分布的变动。

研究结果表明，工资或所得方程的参数结构对收入分布的影响很大，其中省份和人力资本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，各地区之间存在的差异仍然巨大，对这一部分差异的解释值得继续研究。人群不可观察特征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。参与选择行为的转变导致了就业结构的剧烈变化（表4模拟结果1），从而使得收入分布的分散度上升。更为重要的是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，或者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7个百分点。

总之，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和参与选择方式）的结构性变化，

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,这是经历转型的国家面临的一个自然过程。

相比之下,地区因素、人力资本(主要是经验和教育回报率)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,这表明各地之间的差别并不完全由劳动力的素质及其对应回报的差别所造成,对此,政府应该在缩小地区差异上作出更多的努力,如实施适当的财政和税收政策倾斜,正如在改革初期对待沿海开放城市那样;加强落后地区基础设施建设;大力改革低效的地方政府,降低交易成本等。经验、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,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。从各国经济发展的历史经验来看,高层次教育所获回报率的相对上升将难以避免,我们的社会正变得“赢家通吃”。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。

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,曾被形容为异军突起的乡镇企业其经营状况每况愈下,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(尤其是内地的农村)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,如果作综合性考察的话,可以发现从20世纪40年代末期开始一直到20世纪90年代,各国收入分配的趋势很难改变,像中国和东欧国家这种发生较大变化的现象都是社会剧

烈变动的产物。²³这也许可以成为令我们稍感欣慰的理由。而另一方面，政府在防止地区差异进一步扩大和鼓励支持人力资本投资上仍将任重道远。

附录 变量说明²⁴

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 + 离退休再就业人员收入 + 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 - 记账补贴；

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。

行业虚拟变量：

工业：包括工业和农业，因样本中从事农业的人数很少，所以与工业放在一起；

流通部门：包括交通运输业、邮电通讯业、商业、饮食业、物资供销和仓储业；

服务部门 1：为生产和生活服务的部门，包括金融、保险业、地质勘查业、房地产、公用事业、居民服务业，咨询服务业和综合技术服务业，农、林、牧、渔、水利服务业和水利业，公路、内河（湖）航道养护业等；

服务部门 2：为提高科学文化水平和居民素质服务的部门，包括教育、文化、广播电视、科学研究、卫生、体育和社会福利事业等；

服务部门 3：为社会公共需要服务的部门，包括国家机关、政党机关、社会团体以及军队和警察等。

参 考 文 献

- [1] Bourguignon, F., "Decomposable Income Inequality Measures", *Econometrica*, 1979, 47(4), 901—920.
- [2] 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 DE-CRA, Mimeo*, 1998.
- [3] Bourguignon, F., M. Fournier and M. Gurgand, "Fast Development with a Stable Income Distribution: Taiwan, 1979—1994", *Review of Income and Wealth*, June 2001, 47(2), 1—25.

²³ 参见 Bourguignon, Fournier and Gurgand(2001)对台湾地区的研究。

²⁴ 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。

- [4] Bourguignon, F., Francisco H. G. Ferreira and Phillippe G. Leite, "Beyond Oaxaca-Blinder: Accounting for Differences in Household Income Distributions Across Countries", *The World Bank Working Paper*, 2002.
- [5] Bourguignon, F., Francisco H. G. Ferreira and Nora Lustig, "MIDD: The Microeconomics of Income Distribution Dynamics", *Mimeo*, 2001.
- [6] Cowell, Frank A., *Measuring Inequality*, 2nd edition. London: Prentice Hall, 1995.
- [7] Cutler, David M. and Lawrence F. Katz, "Macroeconomic Performance and the Disadvantaged", *Brookings Papers on Economic Activity*, 1991, 2, 1—74.
- [8] Cutler, David M. and Lawrence F. Katz, "Rising Inequality? Changes in the Distribution of Income and Consumption in the 1980's", *American Economic Review*, 1992, 82, 546—551.
- [9] Deaton, Angus, *The Analysis of Household Surveys*. Baltimore: John Hopkins University Press, 1997.
- [10] Deaton, Angus and John Muellbauer, *Economics and Consumer Behavior*. New York: Cambridge University Press, 1980.
- [11] Ehrenberg, Ronald G. and Robert S. Smith, *Modern Labor Economics: Theory and Public Policy*, 6th edition. Reading, MA: Addison-Wesley, 1997.
- [12] Greene, William H., *Econometric Analysis*, third edition. Upper Saddle River: Prentice-Hall International, 1997.
- [13] 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.
- [14] Heckman, James J., "Sample Selection Bias as a Specification Error", *Econometrica*, Jan. 1979, Vol. 47, 153—162.
- [15] Kakwani, Nanak C., *Income Inequality and Poverty: Methods of Estimating and Policy Applications*. New York and Oxford: Oxford University Press, 1980.
- [16] 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.
- [17] Khan, Azizur, Keith Griffin, Carl Riskin and Renwei Zhao, "Household Income and its Distribution in China", *China Quarterly*, 1993, 132, 1029—1061.
- [18] Krueger, Dirk and Fabrizio Perri, "Does Income Inequality Lead to Consumption Inequality? Evidence and Theory", *NBER Working Paper*, 2002.
- [19] Lee Lung-Fei, "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables", *International Economic Review*, Jun. 1978, Vol. 19, 415—433.
- [20] ———, "Generalized Econometric Models with Selectivity", *Econometrica*, March 1983, Vol. 51, 507—512.
- [21] Litchfield, Julie A, "Inequality: Methods and Tools", *Mimeo*, World Bank, 1999.
- [22] Maddala, G. S., *Limited-Dependent and Qualitative Variables*. New York: Cambridge University Press, 1983.
- [23] Meng, Xin, "Economic Restructuring and Income Inequality in Urban China", Project Paper, 2001.
- [24] Trost, Robert P. and Lung-Fei Lee, "Technical Training and Earnings: A Polychotomous Choice Model with Selectivity", *Review of Economics and Statistics*, Feb. 1984, Vol. 66, 151—156.
- [25] 阿玛蒂亚·森, "评估不平等和贫困的概念性挑战", 《经济学(季刊)》, 2003年第2卷第2期, 第257—270页。

- [26] 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年第3期,第21—31页。
- [27] 陈宗胜,“中国居民收入分配差别的深入研究”,《经济研究》,2000年第7期,第68—71页。
- [28] 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)》,2002年第1卷第3期,第563—574页。
- [29] 菲歇尔,《利息理论》,陈彪如译。上海:上海人民出版社,1999年。
- [30] 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》。北京:中国财政经济出版社,1999年。
- [31] 国家统计局,《中国统计年鉴》。北京:中国统计出版社,1990年。
- [32] 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。
- [33] 国家统计局贸易物资统计司,《中国市场统计年鉴》。北京:中国统计出版社,1991—2000年。
- [34] 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年第8期,第3—10页。
- [35] 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》。北京:中国财政经济出版社,1999年。
- [36] 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》。北京:中国财政经济出版社,1999年。
- [37] 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年第7期,第72—76页。
- [38] 李实,“中国个人收入分配”,《经济学(季刊)》,2003年第2卷第2期,第378—404页。
- [39] 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》。北京:中国社会科学出版社,1994年。
- [40] 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年第3期,第32—40页。
- [41] 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年第4期,第42—51页。
- [42] 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》。北京:中国财政经济出版社,1999年。
- [43] 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》。北京:中国财政经济出版社,1999年。
- [44] 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。
- [45] 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》。北京:中国财政经济出版社,1999年。
- [46] 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。
- [47] 赵人伟、格里芬,《中国居民收入分配研究》。北京:中国社会科学出版社,1994年。
- [48] 赵人伟、李实,“中国居民收入差距的扩大及其原因”,《经济研究》,1997年第9期,第19—28页。
- [49] 赵人伟、李实、李思勤,《中国居民收入分配再研究》。北京:中国财政经济出版社,1999年。
- [50] 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年第1卷第3期,第575—588页。

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

DINGSHAN WAN

(Boshi Fund Management Co. Limited)

Abstract This paper combines micro-econometric and micro-simulation methods to study changes in household income distribution in urban China between 1988 and 1999. It aims at exploring the quantitative effects of various determinants, especially endowments and structural parameters, on incomes distribution. After controlling selection biases, we find that changes in structural parameters of the wage equation greatly influence income distribution, among which dummies of province and education are the most significant. Changes in the distribution of unobservable characteristics of people provide little explanation power. Parameters of both the occupational and educational choice equations, especially the later, greatly affect income distribution.

JEL Classification C15, D31, I31