

## 户籍制度松紧程度的一个衡量

何英华\*

**摘要** 本文构建了户籍制度下的个人迁移决策模型。在此基础上,本文衡量了户籍制度的松紧程度,同时估计了个人是否要户口的意愿。文章通过排除变量解决了部分可观察性问题,并用两步估计法纠正了选择性偏差。文章的计量部分从2000年人口普查的抽样数据中选取了大专毕业及以上学历的男性劳动力样本,估计了户籍制度松紧程度和个人意愿水平。

**关键词** 户籍制度,部分可观察性,选择性偏差

### 一、引 论

户籍制度是中国的一项重要制度。以1958年1月9日颁布的《中华人民共和国户口登记条例》为开端,户籍制度已经延续了近50年。不可否认,户籍制度影响到国人生活的方方面面;有人甚至认为,户籍制度在中国造成了国中之国的情形(a *de facto* passport mechanism)。

户籍制度的相关文献已有很多。研究的问题主要有户籍制度形成的历史渊源(Cheng和Selden, 1994)和逻辑基础(林毅夫等, 1994),改革之后维持的原因(蔡昉等, 2001)以及制度的变迁(Chan和Zhang, 1999; 蔡昉等, 2001),户口迁移和非户口迁移的不同机制比较(Chan等, 1999; 何英华, 2003),户籍制度变化的决定因素(蔡昉等, 2001)。从这一项制度的本意和逻辑基础来说,户籍制度是为了控制人们的迁移,它是改革之前的赶超战略得以维持的配套政策(林毅夫等, 1994);而改革之后,它则是城镇居民特殊待遇得以维持的重要手段(蔡昉等, 2001)。因此,在论及户籍制度对中国社会的影响时,我们必然要问:户籍制度对于中国国内迁移的影响到底是什么?很遗憾,目前没有一项正式的研究能够回答这个问题,就算能给我们一些答案,也只是一些定性的分析。<sup>1</sup>其中最主要的困难是我们没有衡量户籍制度松

\* 北京大学中国经济研究中心, 100871; E-mail: he\_yinghua@hotmail.com。感谢赵耀辉的悉心指导,同样感谢John Giles、James Heckman、林毅夫、朴之水(Albert Park)、沈明高、沈艳、姚洋、张车伟、赵忠等的宝贵意见。与王格玮、赵冬以及劳动经济学组的其他研究者的众多讨论亦令我获益匪浅。

<sup>1</sup> 在发达国家,关于制度对迁移的影响的研究也比较少(Greenwood, 1997)。一个例子是Holtz-Eakin(1993)所研究的不可携带的健康保险对工人流动的影响。针对我国,制度对迁移的影响是大家关注的焦点。除了户籍制度外,土地制度也是研究的重点,如Yao(2001b),他利用tobit模型揭示了土地分配制度与迁移的关系。绝大部分中国移民问题的研究没有明确包括户籍制度的影响,如Liang和White(1997)等。

紧程度的一个指标。

因此,首要的问题是构造一个衡量户籍制度松紧程度的指标。本文试图估计户籍制度的松紧程度,而在此之前,有必要对户籍制度做一个简要分析、讨论一下户籍制度如何影响迁移行为。

户籍制度对移民的影响主要反映在一纸户口上。每个人有各自的户口所在地,通常由父母的户口所在地决定。而在迁移时,户口并不能随意地随着每个人的居住地的变化而变化。如果一个迁移者想要得到户口,他必须得到当地政府的批准。通常,只有计划迁移的人(包括国有部门的招工和工作调动,毕业生分配就业等)能获得当地户口。改革开放后,户籍制度逐渐改革,出现了“自理口粮”、“小城镇户口”和“蓝印户口”等等,但是要想得到当地户口,仍然有购买住房和职业方面的要求(参见 Chan 和 Zhang, 1999)。因此,迁移者的户口状态可以分为两类:户口在本地(户口移民)和户口在外地(非户口移民)。如果居住在某地,是否有当地户口不影响个人的各种福利和权利,那么户籍制度也就没有任何实际意义。但目前的实际情况并非如此,一个非户口移民会受到当地政府的种种歧视(可参见蔡昉等(2001),所讨论的北京市对非户口移民的歧视政策)。

表1 户口居民和非户口居民对比

	户口居民	非户口居民
户口状况	户口在本地	户口在外地
在居住地享有的权利	完全享有	部分享有
社会保障	在社保体系之中	一般在社保体系之外,同时取决于就业单位
就业	享有就业特权,特别是公有部门	就业受到限制,特别是在公有部门
工作类型	长期、正式的工作	临时、非正式的工作,或者是自我雇佣
工资	正常,并且享受较高的福利待遇	工资报酬低于同种工作的户口居民,不能享受正常的福利待遇
纳税和交费	缴纳法定税费	除户口居民应缴纳的税费之外,还需要办理暂住证、交纳暂住费等
教育	完全享有受教育的权利	交纳一定的费用:“借读费”、“赞助费”等
(农村)耕地	有权从集体正常承包(须为农业户口)	无权要求集体承包耕地,可以从其他各人转包
政治权力	完全享有	不完全,如基本不能参加社区正常的公共事务管理

资料来源:Chan 和 Zhang(1999),蔡昉等(2001),宋洪远(2001)。

如表1所示,户口移民和非户口移民在居住地所享有的权利有很大差别。<sup>2</sup>因为户口移民已经取得当地的户口,因此他们在当地的权利与本地非移民没有区别。而非户口移民则由于没有当地户口,在很多地方只能享受部分

<sup>2</sup> 例如,可参见 Yao(2001a)在四个沿海村庄调查基础上所做的描述和分析。

权利。比如，他们不能参加一些社会保障项目；不能够在一些岗位就业——尤其是国有部门，而很多只能在非正式部门从事临时工作；即使他们与当地居民从事一样的工作，他们的报酬也低一些。一项在上海的调查显示，移民（主要是非户口移民）的生产效率比当地工人高 50%，而企业雇佣移民的成本仅为雇佣当地居民的 1/5（北京大学中国经济研究中心，1998）。在南京，雇佣一个正式合同工的成本是雇佣一个临时工的两倍（北京大学中国经济研究中心，1999），而临时工一般也都是非户口移民。除了收入低，非户口移民还需要缴纳额外的费用，如暂住费<sup>3</sup>；如果没有当地户口的学龄儿童要在当地入学，通常需要缴纳“赞助费”或者“借读费”等。除此之外，非户口移民还经常遭受政府机关滥用职权之苦，动辄要缴纳各种费用，甚至付出更加昂贵的成本。<sup>4</sup>如果是在农村的非户口移民，他们不能承包当地的土地，最多只能从其他村民那里转包（张曙光，施贤文，2003）。在政治权利方面，非户口移民也只有他们户口所在地的投票权，因此无法在其居住地选举出自己的代言人，更无法参加当地的公共事务管理（Yao, 2001a）。

因此，我们可以将户籍制度分解为两个方面：（1）政府决定发放户口的标准；（2）政府实行歧视非户口移民的政策。歧视政策提高了迁移成本，而前者决定了什么样的人可以获得户口、避免受歧视。当我们说某地的户籍制度较紧，所指的是获得当地户口很困难，并且在当地的非户口移民受到很大的歧视。从逻辑上来说，地方政府如果决定要歧视非户口移民，必定要使他们难以逃脱这种歧视——让他们难以获得当地户口；另一方面，如果不想很多人获得当地户口，那必定是为了保护户口移民的一些“特权”，而非户口移民不能分享这些权力，即受到歧视。<sup>5</sup>因此，户籍制度两个方面的关系是正相关的——发放户口的标准越高，非户口移民受到的歧视越严重；反之亦然。<sup>6</sup>

从上述讨论中可以看出，想要对户籍制度松紧程度做一个衡量，存在两个可能的途径：（1）通过各地户籍制度的规定（包括哪些人可以获得户口，以及对非户口移民的歧视政策）来衡量；（2）通过实证分析，控制个人特征后，估计他在各地（各个时期）获得户口的概率，以此作为户籍制度的衡量。因

<sup>3</sup> 如，在北京移民的平均月工资为 533 元（Zhao, 1999），据笔者了解，当时政府规定的暂住费为 120 元一年。

<sup>4</sup> 如，2003 年，广东省佛山市禅城区一些村委每月收取 20 元的暂住费，而规定的暂住费是每月 2.5 元（付艳霞，赵相林，2003）。又比如，在深圳的非户口移民孙志刚（大学毕业生），因为行政部门的滥用职权而死亡（陈峰，王雷，2003）。

<sup>5</sup> 当然也存在这样的情况：要获得某地的户口很困难，但是当地对非户口移民的歧视并不是太多——不用缴纳暂住费等杂费，中国农村大多是这样的情况。但因为没有户口，非户口移民不能直接承包当地的土地，不能获得建房的宅基地，这些不能获得的权利也是一种歧视。

<sup>6</sup> 当然我们不能排除政策存在自相矛盾的地方，如，实行计划生育的同时实行一些鼓励生育的政策（如农村耕地按照人口调整）。这个问题可能需要进一步的研究，但在本文中，先假定政策自治，户籍制度两个方面存在正相关的关系。在这个假定下，本文用户口的标准来代表户籍制度的松紧程度。需要指出的是，如果这个假定不成立，本文的方法依然可以部分衡量户籍制度松紧程度，因为发放户口的标准毫无疑问是户籍制度中的一个重要部分。

为户籍制度的两个方面存在正相关,所以我们只要衡量了其中一方面,就可以知道整个户籍制度的情况。

但第一个途径基本上是不可能的。户籍制度涉及到多个政府部门,如公安部门、劳动和社会保障部门、教育部门等。要想查阅一个县、一个市的户籍制度资料,或许是可行的。但是,我们的研究不能只局限在一县一市的范围内,因为户籍制度的松紧程度是一个相对水平,因为其他(特别是周边的)各个县市的户籍制度会直接对其产生影响。而要查阅全国或者某省各个地区的户籍制度,这根本就无法做到,因为相关的资料不仅卷帙浩繁,而且没有统一格式,也没有统一部门来管理。更为致命的是,文件上的规定并不一定就是实际执行中的标准。

第二个途径一直以来也是不可行的,主要是由于数据存在问题。现有的数据其实不适合研究移民问题(Zhao, 2003),更不适合研究户籍制度。因为要估计个人在各地获得户口的概率,必然需要个人户口变动的纪录,很遗憾,目前没有包含这类信息的数据。包括本文将使用的2000年人口普查数据,其中也没有户口变动的信息。<sup>7</sup>然而,本文试图通过这个途径,衡量横截面上各地(各省市)户籍制度的相对松紧程度。为了使数据缺陷造成的影响最低,本文抽取了特定的子样本(大专毕业及以上的城镇男性劳动力),并假设了其中户口移民的原户口所在地。这在后文亦有讨论。

为了描述户籍制度下的个人迁移决策,本文建立了一个行为模型(Behavioral Model)。个人在考虑是否迁移时,将是否愿意、是否能够获得目的地的户口考虑在内。如果迁移到某地,个人的户口状态由个人的意愿和政府的户籍管制同时决定。因此,我们可以认为,在户籍制度下,个人迁移模型中包括了一个户口状态的联合决策模型。

如果个人迁移到某地,他可能是户口移民(如果他自己愿意要一个户口,并且政府批准了他的申请),也可能是非户口移民(或许是他自己不想要当地户口;或许是他想要,但是政府没有批准他的申请);并且在两种户口状态下,个人的收益(效用)有很大差别。个人通过比较他在各地的期望效用而决定迁移;通常这种迁移决策可以通过离散选择模型来估计(如, Ma 和 Liaw, 1997; Falaris, 1987)。但是,涉及到户籍制度时,问题变得复杂一些。上文已经提到,我们不知道移民成为非户口移民是因为他自己不想要当地户口,还是他想要、但政府没有批准他的申请,即联合决策模型只有部分可观察(partial observability)。<sup>8</sup>另外,我们只能看到某人迁移到某地后的户口状

<sup>7</sup> 毫无疑问,这一缺陷影响了本文结论的有效性,但本文试图将这种影响降到最低。

<sup>8</sup> Abowd 和 Farber(1982)建立了一个类似的存在部分可观察性的联合决策模型,他们研究的是工人是否参加工会。他们认为,工人是否参加工会同时受到工人和工会的影响;他们的分析中应用了较强的假设——个人决策和工会决策的不可观测因素是独立的。Maddala(1983, Section 11.6)对此做了一个简要的评论。

态，而不能看到他在全国各地的可能结果（这些是反事实的结果——counterfactual outcomes），因此，在利用移民样本估计户口状态决定模型时，存在着选择性偏差，并且，这种选择性来自于个人和政府两个方面（selectivity bias with multiple criteria）。<sup>9</sup>

如果存在排除变量（exclusive variable）——影响一方的决策中，而不影响另一方的决策，那么联合决策模型的部分可观察性就可以得到解决——模型可以识别。在本文的户口决定模型中，个人在考虑是否需要当地户口时，会衡量放弃原来户口的成本，因为，要重新拿回原来的户口也是件不容易的事情。所以，个人会比政府更多地考虑原户口所在地的特征，相关的变量就可以作为排除变量。

选择性偏差在文献中经常涉及。有关个人离散选择中的选择性偏差，Lee（1983）提供了一个纠正方法，Falaris（1987）利用这一方法纠正了移民工资的选择性偏差。本文亦将采用这个方法——首先估计迁移模型的简化式（reduced form），计算纠正偏差的因子；然后估计纠正了选择性偏差的（户口状态）联合决策模型。得到这些估计的参数之后，我们可以控制个人的特征，计算他在各地得到户口的概率，并以此作为户籍制度松紧程度的一个衡量。

利用 2000 年人口普查抽样数据，通过以上程序，户口状态联合决策模型的估计结果显示：对于个人，学历越高，从城市迁出或迁到城市，就更想得到户口；对于政府，如果失业率越低、老年人口比例越低，或者户口申请者学历越高、年纪越轻，那么政府就更有可能发放户口给他们。根据这些结果，文章构建了户籍制度松紧程度的一个衡量——控制了个人特征后，各省城市地区发放户口的概率。这一指标显示，上海、海南、云南和天津等省市的户籍制度较紧，黑龙江、新疆、青海、宁夏和甘肃等省区的户籍制度较松，而北京的松紧程度为中等水平。

本文的剩余部分安排如下：第二部分建立了户籍制度下的迁移决策模型，包括其中的户口状态联合决策模型；第三部分介绍了文章的数据和变量，并讨论了模型识别问题。第四部分通过估计大专以上城镇户口的男性劳动力子样本，用它建立了户籍制度松紧程度的衡量。最后一部分，即第五部分是一个简要总结。

## 二、户籍制度下的个人迁移决策模型

户籍制度下的个人迁移决策如图 1 所示，个人决策受到政府发放户口标准的影响。当个人进行迁移决策时，他会考虑他获得目的地户口的概率，因为这会影响到他的期望效用；而这个概率由个人和政府同时决定。在这样的情况

<sup>9</sup> Maddala（1983，Section 9.6）对于 multiple criteria for selectivity 的相关模型做了简要总结。

形下,个人迁移决策会有三种结果:不迁移、成为户口移民和成为非户口移民。

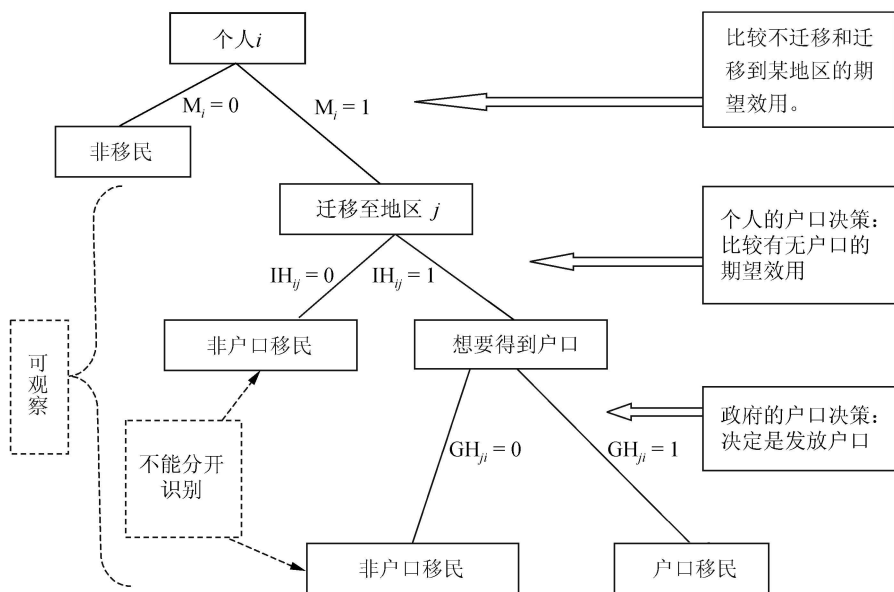


图1 户籍制度下的迁移决策

注:  $IH_{ij}=0$  表示个人  $i$  无意在地区  $j$  得到户口,  $IH_{ij}=1$  表示个人想要得到当地户口。  $GH_{ji}$  表示地区  $j$  政府的户口决策,  $GH_{ji}=0$  表示不给个人  $i$  发放户口,  $GH_{ji}=1$  表示不给个人  $i$  发放户口。我们可以观察到那些迁移到某地的移民的户口状态。

假设个人决定是否想要户口的标准是:<sup>10</sup>

$$y_{ij}^I = X_{ij}^I \beta^I - \epsilon_{ij}^I. \quad (1)$$

其中,  $y_{ij}^I$  是个人  $i$  的隐含指标 (latent index), 可以理解为个人  $i$  在地区  $j$  的期望效用; 上标  $I$  表示个人决策;  $X_{ij}^I$  是影响个人效用的一系列变量, 包括个人特征、迁出地和迁入地的特征, 而  $\epsilon_{ij}^I$  是残差项。这样, 个人是否申请户口的决策可以表示为:

$$IH_{ij} = 1, \quad \text{if } y_{ij}^I > 0, \quad \text{or } \epsilon_{ij}^I < X_{ij}^I \beta^I, \quad (2)$$

$$IH_{ij} = 0, \quad \text{if } y_{ij}^I \leq 0, \quad \text{or } \epsilon_{ij}^I \geq X_{ij}^I \beta^I. \quad (3)$$

其中,  $IH_{ij}=1$  表示个人  $i$  想要得到地区  $j$  的户口;  $IH_{ij}=0$  则表示不想得到该

<sup>10</sup> 可能会存在疑问,为什么在政府歧视非户口移民的同时,还会有一些人不要户口。原因同样在于政府的歧视政策,并且户口不能随心所欲地更换。当个人得到新的户口时,他必须放弃原有的户口,而一旦放弃,原来的户口也不容易拿回来。因此,个人会比较两地的户籍制度以及其他相关的福利。一个明显的例子是,“西部大开发”中,很多西部省份为了吸引高技术劳动力,纷纷推出激励措施,其中有一项就是东部移民可以保留原来的户口,但在西部享受户口居民的同等待遇。另外,北京大学学生就业手册上也有类似的规定——支援西部开发的毕业生可以将户口保留在原籍,甚至保留在学校。

户口。

类似的，我们可以将地区  $j$  的政府是否给个人  $i$  发放户口的标准写为：

$$y_{ji}^G = X_{ji}^G \beta^G - \epsilon_{ji}^G. \quad (4)$$

其中，上标  $G$  表示政府决策； $X_{ji}^G$  是影响政府户口决策的一列变量，包括个人特征、迁出地和迁入地的特征，而  $\epsilon_{ji}^G$  是残差项。政府的具体决策标准是：

$$GH_{ji} = 1, \quad \text{if } y_{ji}^G > 0, \quad \text{or } \epsilon_{ji}^G < X_{ji}^G \beta^G, \quad (5)$$

$$GH_{ji} = 0, \quad \text{if } y_{ji}^G \leq 0, \quad \text{or } \epsilon_{ji}^G \geq X_{ji}^G \beta^G. \quad (6)$$

此处， $GH_{ji} = 1$  表示政府  $j$  愿意发放当地户口给个人  $i$ ；反之，则  $GH_{ji} = 0$ 。在控制了个人特征之后，各地  $GH_{ji} = 1$  的概率就可以作为各地户籍制度松紧程度的衡量。

但是，我们不能将个人和政府的户口决策分开，最后观察到的只是移民的户口状态——户口在本地，或者是户口在外地；也就是说，我们只能观察到个人最终有没有得到户口，即联合决策的结果（ $IH_{ij} \cdot GH_{ji}$ ）——这就是所谓的部分可观察性问题（partial observability）。

我们可以进一步假设，个人和政府决策中的残差符合二元联合正态分布，即：

$$(\epsilon_{ij}^I, \epsilon_{ji}^G) \sim \text{Normal}(0, 0, 1, 1, \rho_{IG}). \quad (7)$$

$\rho_{IG}$  是  $\epsilon_{ij}^I$  和  $\epsilon_{ji}^G$  的相关系数。<sup>11</sup>

这样，个人  $i$  得到地区  $j$  户口的概率就是：

$$\begin{aligned} \text{Prob}(H_{ij} = 1) &= \text{Prob}(\epsilon_{ij}^I < X_{ij}^I \beta^I, \epsilon_{ji}^G < X_{ji}^G \beta^G) \\ &= \Phi_2(X_{ij}^I \beta^I, X_{ji}^G \beta^G, \rho_{IG}). \end{aligned} \quad (8)$$

其中  $\Phi_2(\cdot, \cdot, \rho_{IG})$  是二元正态分布的累积分布函数，且相关系数为  $\rho_{IG}$ 。个人  $i$  在地区  $j$  没有户口的概率是：

$$\begin{aligned} \text{Prob}(H_{ij} = 0) &= 1 - \text{Prob}(H_{ij} = 1) \\ &= 1 - \Phi_2(X_{ij}^I \beta^I, X_{ji}^G \beta^G, \rho_{IG}). \end{aligned} \quad (9)$$

给定个人  $i$  的个人特征，地区  $j$  发放户口的概率就可以表示户籍制度松紧程度：

<sup>11</sup> 这里对 Abowd 和 Farber(1982)的模型做了放松，他们的文章中，两个残差是独立的。假设两个残差相关是有一定依据的：如移民在迁入地有家人或亲戚，那么他可能会更愿意要一个户口，因为他可能长期在当地定居；另一方面，政府给他户口的概率也会更高，因为这时候他更容易达到发放户口的标准（可以参见 Chan 和 Zhang, 1999）。

$$\text{Prob}(\text{GH}_{ij} = 1) = \Phi_1(\epsilon_{ji}^G < X_{ji}^G \beta^G). \quad (10)$$

这里的  $\Phi_1(\cdot)$  是一元正态分布的累积分布函数。只要能将这个联合决策模型估计出,我们就可以对户籍制度松紧程度做一个衡量。

可以看到,这里的概率并不是实际户口移民占该地总移民的比例。实际的比例存在着选择性偏差,因为在政府挑选合适的人发放户口的同时,个人对户籍制度作出了反应。(8),(9)和(10)三个方程并不能直接估计出来,因为我们只能观察到那些真正迁移的人的户口状态,即  $M_{ij} = 1$  的那些人( $M_{ij} = 1$ 表示个人  $i$  迁移到地区  $j$ ),见图1。因此,我们需要调整户口状态中的选择性偏差。

如图1所示,户籍制度下的迁移决策受到获得当地户口概率的影响。假设个人  $i$  在考虑居住在地区  $j$  时,期望效用可以表示为:

$$U_{ij} = \text{Prob}(H_{ij} = 1) V_{ij}^H + \text{Prob}(H_{ij} = 0) V_{ij}^N, \quad (11)$$

$$V_{ij}^H = Z_{ij}^H \gamma^H + \eta_{ij}^H, \quad (12)$$

$$V_{ij}^N = Z_{ij}^N \gamma^N + \eta_{ij}^N. \quad (13)$$

$H_{ij} = 1$ 表示  $i$  在地区  $j$  拥有当地户口;反之,  $H_{ij} = 0$ 。上标  $H$  和  $N$  分别表示户口移民和非户口移民。 $V_{ij}^H$ 和  $V_{ij}^N$ 分别表示个人有当地户口( $H_{ij} = 1$ )和没有当地户口( $H_{ij} = 0$ )时的效用。 $Z_{ij}^H$ 和  $Z_{ij}^N$ 分别是影响户口居民和非户口居民( $i$ )的个人特征和地区  $j$  的地区特征。

将(11),(12)和(13)与(8)和(9)结合在一起,考虑到这些方程中变量的重复问题,我们可以将个人  $i$  在地区  $j$  的效用函数简化式写为:

$$U_{ij} = X_{ij} \beta + \mu_{ij}. \quad (14)$$

$X_{ij}$ 是个人特征和地区特征向量;为了进一步建立 Nested Logit 模型,我们进一步假设残差  $\mu_{ij}$ 服从 generalized extreme value (GEV) 分布。而选择迁移到地区  $j$  的决策标准是:

$$U_{ij} > \max U_{il}, \quad l \neq j. \quad (15)$$

在当前二元结构下,中国的农村和城镇存在巨大的差别,从户籍制度的角度来说,更是如此;<sup>12</sup>因此,我们可以认为农村和城镇存在结构性的差别。考虑到这一点,我们将各个地区划分为两类——农村地区和城镇地区;在迁移决策中,移民首先在三个选择中决策:不迁移、迁移到城镇地区和迁移到农村地区。如果此人选择了迁移到城镇或者农村,他在第二层决策中选择一个省

<sup>12</sup> 在中国农村地区,除非是因为他/她与当地人结婚,一般很难得到一个当地农村的户口。而城镇地区获得户口的途径要多一些,如在当地买房、投资,以及大学生毕业分配等。Chan 和 Zhang (1999)对此有较为详细的讨论。



作为他的迁移目的地（图2）。<sup>13</sup>这样，效用函数（14）可以改写为：

$$U_{ijk} = X_{ijk}\beta + Z_k\gamma + \mu_{ijk}. \quad (16)$$

其中  $j$  表示省份， $k$  表示地区的类型——农村或者城镇， $X_{ijk}$  是地区  $jk$  的地区特征和个人  $i$  的个人特征， $Z_k$  是类型  $k$  地区的共同特征。

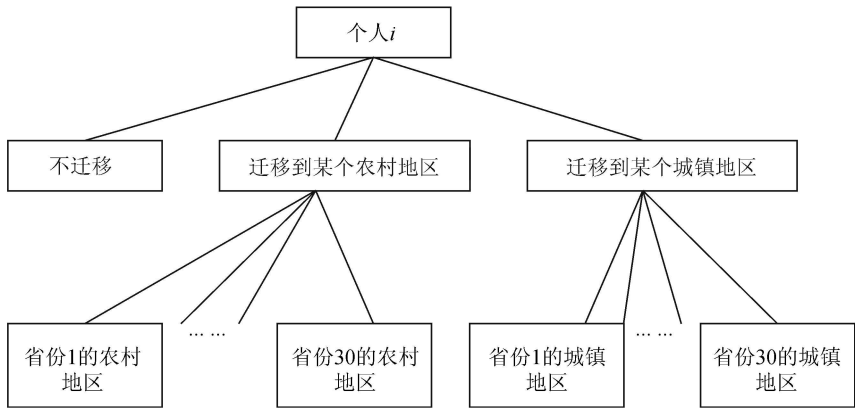


图2 Nested Logit 迁移决策模型简化式的结构

注：西藏没有包括在分析中，因此总共有30个省、市、自治区；省内迁移也包括在上述的结构中。

我们以  $M_{ijk} = 1$  表示个人  $i$  选择迁移到  $j$  省的  $k$  地区，并且记  $P_{ijk} \equiv \text{Prob}(M_{ijk} = 1)$  为迁移到该地区的概率。McFadden (1981) 证明  $P_{ijk} = P_{ij|k} \cdot P_k$ ，并且这个离散选择模型可以用 Nested Logit 方法来估计。两个概率—— $P_{ij|k}$  和  $P_k$  可以表示为：

$$P_{ij|k} = \frac{\exp(X_{ijk}\beta)}{\sum_{m=1}^{J_k} \exp(X_{imk}\beta)}, \quad (17)$$

$$P_k = \frac{\exp(Z_k\gamma + \tau_k I_k)}{\sum_{n=1}^3 \exp(Z_n\gamma + \tau_n I_n)}. \quad (18)$$

其中  $I_k = \ln\left(\sum_{m=1}^{J_k} \exp(X_{imk}\beta)\right)$  是类型为  $k$  的地区的 inclusive value， $\tau_k$  是 inclusive value 的系数。如果  $\tau_k = 1$ ，这个 Nested Logit Model 就等同于 Conditional Logit Model，其中假设了各个选项之间是独立不相关的（independence of irrel-

<sup>13</sup> 这一结构实际上是王格玮(2004)所建立的，他详细讨论和比较了中国迁移问题中各种 nested logit 模型的模型设定问题。同样，Ma 和 Liaw (1997)也利用 nested logit 模型研究了中国的迁移问题。

evant alternatives, IIA)。<sup>14</sup>在下文中,为了简单起见,我们在不至于混淆的地方将下标  $k$  略去。

为了调整选择性偏差,我们需要计算  $\text{Prob}(H_{ij} = 1 | M_{ij} = 1)$  (即在迁移至地区  $j$  的条件下,个人获得当地户口的概率)以及  $\text{Prob}(H_{ij} = 0 | M_{ij} = 1)$ 。在 Nested Logit 模型中,我们假设  $\mu_{ijk}$  服从 GEV 分布, Lee (1983) 介绍了在此情况下调整选择性偏差的一般方法。首先,  $\mu_{ijk}$  可以转化为服从标准正态分布的随机变量:

$$\epsilon_{ij}^M \equiv \Phi^{-1}(F(\mu_{ij})). \quad (19)$$

这里的  $\Phi_1^{-1}(\cdot)$  是标准正态分布累积分布函数的反函数。通过这个转换,条件于迁移 ( $M_{ij} = 1$ ), 个人  $i$  在地区  $j$  获得户口的概率是:

$$\text{Prob}(H_{ij} = 1 | M_{ij} = 1) = \text{Prob}(\epsilon_{ij}^I < X_{ij}^I \beta^I, \epsilon_{ji}^G < X_{ji}^G \beta^G | \epsilon_{ij}^M < \Phi^{-1}(P_{ij})). \quad (20)$$

其中的  $P_{ij} = \text{Prob}(M_{ij} = 1)$  是个人  $i$  选择迁移到  $j$  地区的概率,在估计了迁移决策模型的简化式之后,这个概率可以计算得出。

由于向量  $(\epsilon_{ij}^I, \epsilon_{ij}^G, \epsilon_{ij}^M)$  中所有元素的边际分布都是标准正态分布,因此它们的联合分布是三元标准正态分布。而它们的相关系数矩阵可以表示为:

$$\Sigma = \begin{pmatrix} 1 & \rho_{IG} & \rho_{IM} \\ \rho_{IG} & 1 & \rho_{GM} \\ \rho_{IM} & \rho_{GM} & 1 \end{pmatrix}. \quad (21)$$

其中,  $\rho_{IG}$  是  $\epsilon_{ij}^I$  和  $\epsilon_{ji}^G$  的相关系数,  $\rho_{IM}$  是  $\epsilon_{ij}^I$  和  $\epsilon_{ij}^M$  的相关系数,  $\rho_{GM}$  是  $\epsilon_{ji}^G$  和  $\epsilon_{ij}^M$  的相关系数。这样,方程 (20) 可以写为:

$$\begin{aligned} \text{Prob}(H_{ij} = 1 | M_{ij} = 1) &= \frac{\text{Prob}(\epsilon_{ij}^I < X_{ij}^I \beta^I, \epsilon_{ji}^G < X_{ji}^G \beta^G, \epsilon_{ij}^M < \Phi^{-1}(P_{ij}))}{\text{Prob}(\epsilon_{ij}^M < \Phi^{-1}(P_{ij}))} \\ &= \frac{\Phi_3(X_{ij}^I \beta^I, X_{ji}^G \beta^G, \Phi^{-1}(P_{ij}); \Sigma)}{P_{ij}}. \end{aligned} \quad (22)$$

此处  $\Phi_3(\cdot, \cdot, \cdot, \Sigma)$  是相关系数矩阵为  $\Sigma$  的三元正态分布累积分布函数。

这样,我们可以将户口状态联合决策模型,即方程 (8) 和 (9),用极大似然的方法来估计。似然函数可以写为:

$$\ln L = \sum_{H_{ij}=1} \ln[\text{Prob}(H_{ij} = 1 | M_{ij} = 1)] + \sum_{H_{ij}=0} \ln[\text{Prob}(H_{ij} = 0 | M_{ij} = 1)]$$

<sup>14</sup> Greene (2000, Section 19.7.4) 有更详细的讨论。

$$\begin{aligned}
 &= \sum_{H_{ij}=1} \ln \left[ \frac{\Phi_3(X_{ij}^L \beta^L, X_{ij}^G \beta^G, \Phi_1^{-1}(P_{ij}); \Sigma)}{P_{ij}} \right] \\
 &+ \sum_{H_{ij}=0} \ln \left[ 1 - \frac{\Phi_3(X_{ij}^L \beta^L, X_{ij}^G \beta^G, \Phi_1^{-1}(P_{ij}); \Sigma)}{P_{ij}} \right]. \quad (23)
 \end{aligned}$$

估计出方程(23)后,可以利用这些估计的系数,按照方程(10)来建立户籍制度松紧程度的指标。

### 三、数据、模型识别和变量

#### (一) 数据

本文所使用的主要数据是2000年人口普查的0.95%随机抽样。其中包括在普查标准时间(2000年11月1日零时)的个人基本信息,如性别、年龄、教育程度等。关于迁移,我们知道个人在1995—2000年间最后一次迁移的情况,包括迁出地和目的地,以及这两地的类型:农村、镇或者城市。

由于存在户籍制度,移民的定义变得稍微有些复杂。本文的移民定义为五岁以上、居住地与五年前不同(至少是跨县、市、区)的人。如果该移民的户口在现居住地,那么他就是户口移民;如果户口不在现居住地,他就是非户口移民。为了排除短期旅游探亲的人,我们需要对非户口移民的居住时间作一个限定——非户口移民必须在当地居住半年以上,或者未满半年、但离开户口所在地半年以上。

对于非户口移民,我们知道他们的户口所在地,分析中以此作为他们的迁出地。但对于户口移民,普查数据中没有他们转换户口的信息,因此我们只能根据户口移民最后一次迁移的信息,以及他们出生地和五年前(1995年11月1日前)的常住地信息来假设出他们原来的户口所在地。直接把户口移民的最后一次迁移的迁出地作为他原来的户口所在地,这是不适当的,因为在1995—2000年间,个人可能发生多次迁移行为,但户口不一定每次都发生转换,甚至可能根本没有发生转换。比如,有很多外出打工的农民工一般都不会转换他们的户口,在外辛苦了较长一段时间之后,他们大多回到故乡;而这些人可能被算作是户口移民,因为他们的居住地发生了变化,并且现居住地和户口所在地在同一个地方。利用与本文同样的数据,何英华(2003)发现了这一现象,尤其是迁移到农村的移民中,移民回迁的现象比较普遍。这些回迁移民的行为与一般的移民存在巨大的差异,因此,在分析时需要对这些人特别处理。同时利用户口移民最后一次迁移的信息,以及出生地和五年前(1995年11月1日前)的常住地信息,这样可以尽量避免这个问题。<sup>15</sup>

<sup>15</sup> 幸运的是,很多户口移民(特别是我们将要抽取的大专以上学历的男性劳动力样本中)只有很少一部分最后一次迁移的迁出地不是五年前的常住地。下文有更详细的讨论。

## (二) 模型识别与变量

方程(8)和(9)所描述的户口状态联合决策模型中,  $\beta^I$  和  $\beta^G$  的识别需要至少有一个变量在  $X_{ij}^I$  或者  $X_{ji}^G$  中, 并且不在另一个中。

由于存在选择性偏差, 即只有在个人迁移至某地之后 ( $M_{ij}=1$ ), 我们才能观察到他的户口状态, 因此需要通过估计方程(23)来纠正这一偏差。在这个情况下, 模型的识别需要更强的条件。条件于  $M_{ij}=1$ ,  $\beta^I$  和  $\beta^G$  的识别还需要额外的一个条件: 需要至少有一个变量在迁移模型的简化式中(即  $X_{ij}$  中), 但是不在户口状态联合决策模型中(即  $X_{ij}^I$  和  $X_{ji}^G$  中)。

当个人在决定是否想要获得当地的户口时, 他会比较迁出地和迁入地的特征。因为一旦他获得了迁入地的户口, 他必须放弃迁出地的户口, 而户口一旦放弃, 重新申请需要成本, 并且有风险。但是, 政府在考虑个人的户口申请时, 更关注当地的地区特征和个人特征; 即使关注一些个人迁出地的特征(如是否来自大城市等), 政府也不会像个人那样关注那么多的迁出地特征。这样, 将一些迁出地特征作为排除变量(exclusive variable), 第一个识别条件就可以得到满足。

对于第二个条件, 我们知道迁移受到距离的影响,<sup>16</sup> 当我们把户籍制度下的个人迁移决策中的一部分——户口决策抽象出来单独考察后, 我们可以认为, 距离不影响个人的户口决策——即距离包括在  $X_{ij}$  中, 而不在  $X_{ji}^I$  中, 更不在  $X_{ji}^G$  中(政府在考虑个人的户口申请时, 不会考虑个人迁移的距离)。这里可能存在一些疑问, 因为在我们的日常观察中, 人们在考虑是不是获得户口时, 跟是否想要长久居住在当地相联系, 这必然又跟离老家的远近有关系——比如, 在其他因素相同的情况下, 一个老家在浙江的人可能更愿意在上海获得户口, 并在上海长期定居, 而不是在北京, 因为北京离浙江距离较远。<sup>17</sup> 但是, 我们也可以将这种影响归结为距离对迁移的影响, 而把个人户口决策单纯限定于比较两个户口的含金量。因此, 距离可以作为纠正选择性偏差的排除变量(exclusive variable), 这样方程(23)就可以识别。

在迁移决策简化式的决策树中(图2), 个人在第一层次的决策中有三个选择: 不迁移、迁移到一个农村地区 and 迁移到一个城镇地区; 在第二层次, 如果个人选择了迁移(到农村或者城镇), 他会选择一个明确的地点, 本文所指的是各个省。在第二层次, 个人的决策受到下列因素的影响: 收入水平<sup>18</sup>、GDP增长率、失业率, 以及是否是跨省迁移。同时, 我们也考虑其他一些可

<sup>16</sup> 更详细的讨论可以参见 Schwartz(1973)和 Greenwood(1975)。

<sup>17</sup> 这反映了最近移民研究的新趋势, 即研究家庭决策, 而不仅仅限定在个人层次上的分析。但很遗憾, 我们的数据并不能让我们做基于家庭层次的迁移决策分析。

<sup>18</sup> 这里的收入水平是个人居住地的省级人均收入, 并不是个人层次的收入数据; 这是中国人口普查数据中的一个最大的缺陷。如果有微观收入数据, 很多分析可以深入一些。

能影响户籍制度松紧程度的因素，如城镇地区的人口密度<sup>19</sup>、政府开支占 GDP 的比例、老年（年龄在 65 岁以上）人口比例等等。<sup>20</sup>

为了避免联立偏差（simultaneity bias），几乎所有变量都是 1995 的数值。只有城镇地区人口密度的数据是 1996 年的，因为在 1996 年之前，没有相关的省级数据。

在第一层次的选择中，个人的选择受决定于一些个人特征：教育程度、年龄和年龄平方，以及婚姻状况。

表 2 变量定义和数据来源

变量名	变量定义	数据来源
URINC95	1995 年城镇家庭人均可支配收入(对数值)	中国统计年鉴, 1996
RUINC95	1995 年农村家庭人均净收入(对数值)	中国统计年鉴, 1996
GR	1990 年至 1995 年年均真实 GDP 增长率(乘以 100)	中国统计年鉴, 1991—1996
UNEM95	1995 年非农业失业率(乘以 100)	1995 年全国 1% 人口抽样调查资料
ER95	1995 年 64 岁以上老年人口比例(乘以 100)	中国统计年鉴, 1996
DAPC	1996 年城镇地区人均建成区面积(平方公里/万人)	中国统计年鉴, 1997
POPGR	1990 年至 1995 年年均人口增长率(乘以 100)	第四次人口普查抽样数据和 1995 年全国 1% 人口抽样调查资料
GEGDP95	1995 年政府支出占当地 GDP 的比例(乘以 100)	中国统计年鉴, 1996
DIS	两省间距离为省会间铁路里程的对数值, 省内迁移的距离为 $0.88 \times \sqrt{\text{该省面积}/\pi}$ 的对数值。	全国铁道旅客时刻表
AL	1996 年农村人口人均耕地面积。	中国统计年鉴, 1997
IN_PROV	虚拟变量, 若移民为省内迁移, 则 IN_PROV = 1。	
BTS	虚拟变量, 如果该地区是京津沪之一, 则 BTS = 1。	
AGEUR	AGEUR = (2000 年 11 月 1 日时个人的年龄) × (地区为城镇)	
AGESQUR	AGESQUR = (AGEUR 的平方)	
AGERU	AGERU = (2000 年 11 月 1 日时个人的年龄) × (地区为农村)	
AGESQRU	AGESQRU = (AGERU 的平方)	
EDU1—EDU3	教育程度的虚拟变量 若为大专毕业, 则 EDU1 = 1 若为本科毕业, 则 EDU2 = 1 若为研究生毕业或以上, 则 EDU3 = 1	
OC/DC	迁出地/迁入地是城市	
OBTS/DBTS	迁出地/迁入地是京津沪之一	

注：若无特殊说明，则数据来源是 2000 年人口普查数据。

<sup>19</sup> 本文中，城镇地区的人口密度是该省城镇人口人均建成区面积。农村地区的对应变量是农村人口人均耕地面积。正如我们所知，在全国各地的农村地区，户籍制度松紧程度大体一致，因此，只有当农业工作是个人选择的重要组成部分时，农村人均耕地面积才包括在分析中。

<sup>20</sup> 显然，即使户籍制度不存在，这些影响户籍制度的因素也会影响迁移决策。但由于我们估计的是迁移决策的简化式，因此，这样做并没有问题。

由于农村和城镇之间存在巨大的差异,我们允许一些变量的影响在农村和城镇不同。比如,在估计收入水平的的影响时,模型中该变量和农村与城镇的虚拟变量构成了两个交叉项(interacted term)。

所有变量的定义和数据来源在表2中作了一个总结。

#### 四、估计户籍制度松紧程度

本部分将建立户籍制度松紧程度的指标——即给定个人特征后,政府给户口申请者发放户口的概率。首先我们抽取了一个大专程度以上男性劳动力样本来估计迁移决策的简化式,并且得到了调整选择性偏差的因子,然后,我们估计了户口状态联合决策模型。

在中国城镇地区,户口发放标准中的一项重要限制是教育程度。如果一个迁移者的教育程度在高中或以下,那么他基本上不可能得到户口,除非是因为婚姻或者其他特殊原因。因此,为了衡量各省的户籍制度松紧程度的差异,我们将抽取教育程度较高的劳动力样本。同时,对于大专毕业和本科毕业生的人,有一些是成人教育学历,而我们知道这些学历与正规大学的本专科毕业生大不相同,同时,考虑到女性劳动力的迁移受到更多非经济因素的影响,因此我们最后抽取的是年龄在15—64岁的、居住地(或迁出地)是城镇、有非农业户口、大专(非成人教育)毕业以上的男性劳动力。<sup>21</sup>

在这个样本中,总共有10882个人,其中1763个,即16.2%是移民。在这些移民中,32.84%是跨省移民,64.38%是户口移民(见表3)——对比总移民中的户口移民比例(26%<sup>22</sup>),这一比例是相当高的。由于我们不知道户口移民的原户口所在地,对此我们需要假设。我们首先考察的是移民最后一次迁移的信息,看其最后一次迁移的迁出地是否与五年前——1995年11月1日之前的常住地相同;若是,则认为他的原户口所在地就是最后一次迁移的迁出地;若不相同,再参考其出生地信息,确定其原户口所在地。从表3中可看出,只有3.11%的户口移民最后一次迁移的迁出地不是五年前常住地。另外,由于我们把迁移的出发地和目的地抽象为30个省中的城镇和农村地区,所以,这样假设他们的原户口所在地不会有大的问题。

<sup>21</sup> 现在或五年前居住在西藏和国外的人排除在样本外,这样,下面的分析中共有30个省(市、自治区)。

<sup>22</sup> 数据来源:2000年人口普查0.95%抽样数据。

表3 用以衡量户籍制度松紧程度的样本(城镇大专以上男性劳动力)基本情况

特征	百分比	特征	百分比
男性	100	居住地(移民的迁出地)类型	
		镇	21.24
教育程度(非成人学历)		城市	78.76
大专毕业	54.96		
本科毕业	41.84	样本中的移民	16.2
研究生毕业及以上	3.2	移民中的省际移民	32.84
		移民中的户口移民	64.38
2000年11月1日的年龄		移民迁出地显然非5年前常住地的比例 <sup>b</sup>	3.11
	15—25		12.57
	26—35	移民迁入地类型	
	36—45		
	46—55	村	7.66
	56—64	镇	13.95
		城市	78.39
样本数量	10882		

数据来源:2000年人口普查数据0.95%样本。

注:1. 所有教育程度都是非成人教育学历,并且是完成学业的。

2. 2000年人口普查中,关于迁移的信息有两项。一个是最后一次迁移,我们知道移民的迁出地类型;另一个是根据个人的五年前常住地,但我们只知道该地是在本省还是外省,不知道该地的类型。对于非户口移民,我们知道他们现在的户口所在地;对于户口移民,我们也只知道他们现在的户口所在地,不知道他们原来的户口所在地。此处,我们假设户口移民的原户口所在地即为其迁出地。移民在五年中可能发生多次迁移,其最后一次迁移也不一定发生户口转换。为避免这种情况,我们仔细考察两项迁移信息——首先看两项信息是否一致;若不一致则参考个人的出生地,确定其原户口所在地。此处的比例(3.11%)所指的是三个地区(最后一次迁移的迁出地、五年前常住地和出生地)都不相同的户口移民。对于这些人,我们假设他们原户口所在地是五年前常住地,并且地区类型与最后一次迁移的迁出地一样。

### (一) 简化式的估计

我们首先估计了Nested Logit迁移模型的简化式,估计的结果在附录1给出。模型I至模型IV包括了不同的变量,以检验估计的稳健性(robustness)。

虽然是简化式的估计结果,但我们还是可以对估计的结果做一个简要的讨论。结果显示,大多数系数的符号和显著程度与理论预期相一致。<sup>23</sup>收入水平(URINC95和RUINC95)对迁移有正的影响,失业率有负的影响,这与Todaro(1969)以及Harris和Todaro(1970)预期的一致。GDP增长率(GR)系数符号也为正。

上文已经提到,我们假定三个人口变量与政府的户籍制度相关。当然,我们不能说这些变量对迁移没有直接的影响,例如,引力模型(Gravity Model)认为人口数量对迁移决策有重要影响(例见Schultz,1982)。人口增长率(POPGR)的系数为正,但Fields(1979)认为,各地的劳动力市场是相对稳

<sup>23</sup> 准确地说,讨论变量的影响时,要看它的边际效应(marginal effect):在Nested Logit模型中,变量的边际效应符号与其系数的符号一致。下文对此有更详细的讨论。

定的,因此分析中包括人口增长率可能会导致联立偏差。因此在模型 IV 中,没有包括人口增长率,结果显示系数的估计值基本没有变化。

其他一些地区特征,如距离的对数值、省内迁移的虚拟变量和京津沪的虚拟变量,也包括在模型中。京津沪的虚拟变量包括在分析中,原因在于京津沪是直辖市<sup>24</sup>,很多方面与一般的省不同,特别是考虑了户籍制度之后。

选择不迁移、迁移到农村或者城镇,这些受到个人特征的影响。教育程度、年龄、年龄平方、婚姻状况,以及农村和城镇的固定效应(fixed effect)被包括在其中。

我们尝试了两种教育程度的分类:分为大专(EDU1)、本科(EDU2)和研究生及以上(EDU3)三类;或者分为大专(EDU1)和大专以上两类。从结果来看,本科(EDU2)和研究生及以上(EDU3)的系数没有显著差异;因此,在模型 II 和模型 IV 中,它们被归结为同一类。对于婚姻状况,模型 III 中的结果表明,它只对迁移到农村有显著影响,并且,模型中是否包括婚姻状况,对其他系数的估计值没有大的影响。另外,婚姻的决策在很多时候与迁移决策是同时作出的,因此有可能导致联立偏差。我们的样本是男性样本,而在中国一般来说是妻随夫居住,所以,婚姻状况没有显著影响是符合实际情况的。

基于上述讨论,下文中我们将使用两组估计系数:教育分为大专和大专以上两类,婚姻状况包括或不包括在模型中(模型 I 和 III)。

## (二) 户口状态联合决策模型的估计

估计了简化式之后,我们可以计算纠正选择性偏差的因子;然后通过估计方程(23),我们可以获得户口状态联合决策模型的无偏估计。我们用两组估计值(附录 1 中的模型 II 和模型 III)来估计方程(23)。从第二部分的讨论中我们已经知道,移民的户口决策受到其迁出地和迁入地差异的影响<sup>25</sup>,个人特征也有重要作用。

方程(23)的两组估计结果在表 4(a)和(b)给出;表 4(a)中是个人户口决策的估计值,表 4(b)是政府户口决策的估计值;其中,模型 I 中没有包括个人的婚姻状况,模型 II 则包括了婚姻状况变量。

<sup>24</sup> 重庆也是直辖市,但是它从四川省分出的时间是 1997 年,并且,重庆市所辖面积大、人口多,更类似于一个省,而与京津沪有较大差别。因此,本文将重庆当作是一个省,而不是像京津沪那样。

<sup>25</sup> 更确切地说,我们在模型中包括了迁出地和迁入地相关变量的差。这种设定其实需要假设迁入地和迁出地的相同变量对迁移的影响是对称的,相关的讨论可以参见 Fields (1979)。



表 4(a) 户口状态联合决策模型：个人决策部分(N=1628)

变量	模型 I		模型 II	
	系数	标准差	系数	标准差
常数项	0.116	(79.722)	0.224	(4.882)
AGE	0.047	(1.255)	0.015	(0.277)
AGESQ	-0.001	(0.013)	-0.001	(0.003)
EDU1	-0.183	(0.265)	-0.594**	(0.328)
URINC95 的差异	-0.018	(2.006)	-0.038	(1.668)
GR 的差异	-0.088	(0.129)	-0.074	(0.121)
UNEM95 的差异	-0.029	(0.066)	-0.031	(0.098)
GEGDP95 的差异	-0.117	(0.086)	-0.191*	(0.073)
ER95 的差异	-0.021	(0.164)	0.042	(0.230)
DAPC 的差异	0.002	(1.244)	0.017	(0.825)
POPGR 的差异	-0.013	(0.460)	0.016	(0.509)
DC	0.699	(0.550)	0.674*	(0.121)
OC	0.659***	(0.394)	0.490	(0.502)
DBTS	0.463	(0.855)	0.254	(2.034)
OBTS	0.866***	(0.506)	0.671	(1.503)
IN_PROV	0.550	(0.807)	0.726***	(0.444)
MARR	—	—	-0.257	(0.573)
$\epsilon^I$ 和 $\epsilon^G$ 的相关系数	0.029	(0.896)	0.231	(0.333)
$\epsilon^I$ 和 $\epsilon^M$ 的相关系数	-0.041	(0.277)	0.106	(0.193)
对数似然值	-954.092		-946.904	

注：1. 因变量是移民的户口状态——户口移民还是非户口移民，所估计的极大似然函数即方程(23)。本模型是存在两个选择标准的联合决策模型，并且纠正了其中的选择性偏差(见图2)。这里所给出的估计值是个人户口决策部分，其中所指的“差异”是该变量的迁出地水平减去迁入地水平。估计中所用到的样本是迁移到城镇的大专毕业及以上男性劳动力移民。

2. \*, \*\*, \*\*\*: 系数分别在 1%, 5% 和 10% 水平上显著不为零。

表 4(b) 户口状态联合决策模型：政府决策部分(N=1628)

变量	模型 I		模型 II	
	系数	标准差	系数	标准差
常数项	0.069	(2.881)	0.079	(2.966)
AGE	-0.052**	(0.024)	-0.056	(0.203)
AGESQ	0.001**	(0.000)	0.001	(0.002)
EDU1	-0.407*	(0.087)	-0.219	(0.256)
URINC95	0.446	(0.392)	0.446	(0.534)
GR	0.028	(0.032)	0.031	(0.030)
UNEM95	-0.060*	(0.021)	-0.061***	(0.036)
GEGDP95	-0.033	(0.025)	-0.018	(0.035)
ER95	-0.279*	(0.067)	-0.300*	(0.083)
DAPC	0.155	(0.378)	0.653	(0.558)
POPGR	-0.140	(0.204)	-0.334	(0.287)
DC	-0.034	(0.258)	-0.202	(0.189)
OC	0.254	(0.181)	0.260	(0.367)
DBTS	0.115	(0.256)	0.130	(0.950)
OBTS	0.105	(0.118)	-0.118	(0.303)
IN_PROV	0.540***	(0.290)	0.530	(0.372)
MARR	—	—	0.061	(0.278)
$\epsilon^G$ 和 $\epsilon^I$ 的相关系数	0.029	(0.896)	0.231	(0.333)
$\epsilon^G$ 和 $\epsilon^M$ 的相关系数	0.033	(0.154)	0.053	(0.216)
对数似然值	-954.092		-946.904	

注：1. 因变量是移民的户口状态——户口移民还是非户口移民，所估计的极大似然函数即方程(23)。本模型是存在两个选择标准的联合决策模型，并且纠正了其中的选择性偏差(见图2)。这里所给出的估计值是政府户口决策部分。估计中所用到的样本是迁移到城镇的大专毕业及以上男性劳动力移民。

2. \*, \*\*, \*\*\*: 系数分别在 1%, 5% 和 10% 水平上显著不为零。

在模型 I 中, 35 个系数中只有 8 个是显著的 (10% 的显著性水平上)。在个人的户口决策中, 迁出地为城市 (OC) 和迁出地是京津沪之一 (OBTS), 可以增加个人想要获得户口的意愿<sup>26</sup>; 而其他的变量没有显著影响。在政府的户口决策中, 个人年龄的影响是 U 形的, U 形的最低点大约是 26 岁; 对于教育程度更高、或者是省内迁移 (IN\_PROV) 的人, 政府更有可能发放户口; 而失业率 (UNEM95) 以及老年人口比例 (ER95) 则会降低政府发放户口的概率。其余变量对政府的户口决策没有显著影响。

在模型 II 中, 总共 37 个变量有 6 个是显著的 (10% 的显著性水平上)。在个人户口决策中, 如果个人教育程度较低 (大专毕业), 那么他想要户口的意愿就相对较低; 同样, 如果当地的政府支出占当地 GDP 比例 (GEGDP95) 相对高于个人的迁出地水平, 这一意愿也会降低; 如果个人迁移至城市 (DC)、或者是省内迁移 (IN\_PROV), 那么他想要户口的意愿会相对较高。在政府户口决策模型中, 我们同样发现, 失业率 (UMEM95) 和老年人口比例 (ER95) 对发放户口有负的影响。

比较模型 I 和模型 II 的估计结果, 相同变量的系数符号基本相同, 数值大小也接近。这表明婚姻状态这个变量没有导致严重的联立偏差问题。<sup>27</sup>

表 5 模型设定的 Hausman 检验

	零假设	统计量: $\chi^2(37)$
Hausman 检验 I	$\epsilon^I$ 和 $\epsilon^M$ 的相关系数等于零 $\epsilon^G$ 和 $\epsilon^M$ 的相关系数等于零 即, 不存在选择性偏差——迁移决策与个人和政府的户口决策无关	87.885 Prob > $\chi^2(37) = 0.000$
Hausman 检验 II	$\epsilon^I$ 和 $\epsilon^G$ 的相关系数等于零 $\epsilon^I$ 和 $\epsilon^M$ 的相关系数等于零 $\epsilon^G$ 和 $\epsilon^M$ 的相关系数等于零 即, 三个决策——迁移决策、个人和政府的户口决策——都无关。	592.43 Prob > $\chi^2(37) = 0.000$

( $\epsilon_{ij}^I$ ,  $\epsilon_{ji}^G$ ,  $\epsilon_{ij}^M$ ) 之间的三个相关系数都不显著; 我们利用 Hausman 检验来验证模型设定, 检验结果在表 5 给出。两个零假设——( $\epsilon_{ji}^G$ ,  $\epsilon_{ij}^M$ ) 之间和 ( $\epsilon_{ij}^I$ ,  $\epsilon_{ij}^M$ ) 之间的相关系数为零, ( $\epsilon_{ij}^I$ ,  $\epsilon_{ji}^G$ ,  $\epsilon_{ij}^M$ ) 之间的三个相关系数为零, 都被拒绝。因此我们可以认为, 确实有必要调整选择性偏差。

我们得到政府户口决策模型的估计值后, 就可以利用这些系数来预测各省政府发放户口的概率, 同时可以控制个人想要户口的意愿和个人的其

<sup>26</sup> 确切地说, 每个变量对于个人户口决策的影响应该是该变量的边际效应。我们知道, 标准正态分布累积分布函数  $\Phi(x\beta)$  的导数是  $\phi\beta(x\beta)$ , 其中  $\phi(\cdot)$  是标准正态分布密度函数。因此, 各个变量的边际效应与其系数的符号一致。同样, 讨论政府的户口决策模型时, 我们也只要看变量系数的符号。

<sup>27</sup> 更确切的回答这个问题, 需要进行 Hausman 检验。

他特征。本文预测概率时，给定个人是一个 25 岁、大专毕业并且从外省迁移来的移民；各个省发放户口的概率相对高低在图 3 中给出。利用模型 I（未包括婚姻状态）和模型 II（包括婚姻状态）的两组估计值，我们构造了两组概率——SHRS I（利用模型 I 的估计值）和 SHRS II（利用模型 II 的估计值），可以看到两组概率没有显著差异。图中显示，上海发放户口的概率最低，表明上海的户籍制度最紧；黑龙江、青海、宁夏和新疆发放户口的概率很高，即户籍制度较松。<sup>28</sup>

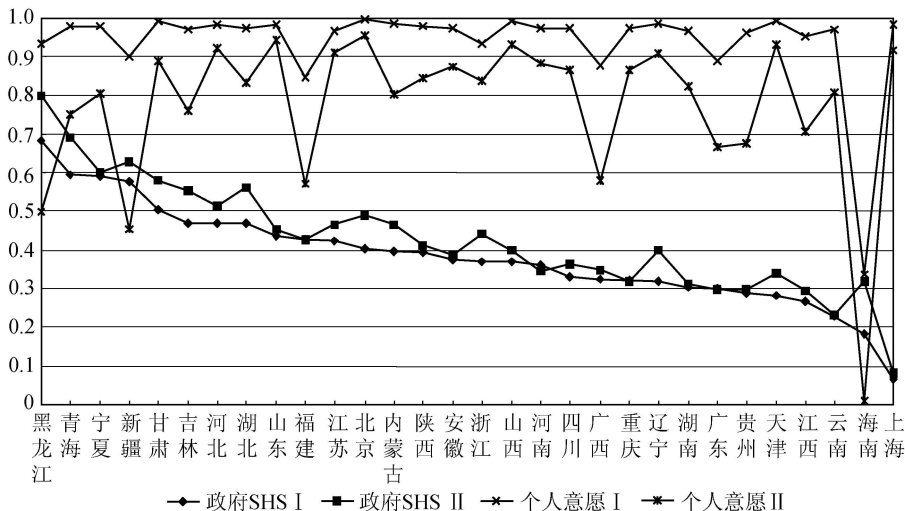


图 3 预测的中国大陆 30 省（市、自治区）户籍制度松紧程度和个人获得当地户口的意愿：按照 SHSI 排列

注：1. 西藏没有包括在分析中，因此总共有 30 个省、市、自治区。表中所给出的政府 SHS 预测值各地政府给大专毕业、年龄 25 岁、从外省迁来的未婚移民发放户口的概率。计算个人意愿时，假设个人原户口所在地的特征是全国平均水平，个人为大专毕业、年龄 25 岁、从外省迁来的未婚男性。政府 SHS I 和个人意愿 I 计算时利用了没有包括婚姻状况变量的估计值，政府 SHS II 和个人意愿 II 计算时利用了包括婚姻状况变量的估计值。

2. 政府 SHS I：均值为 0.384，标准差为 0.129；政府 SHS II：均值为 0.427，标准差为 0.147。个人意愿 I：均值为 0.940，标准差为 0.120；个人意愿 II：均值为 0.774，标准差为 0.197。

一个有意思的发现是北京户籍制度并不像想像中的那么紧，只处于中等水平。一般认为，大城市（如北京和上海）的户籍制度较紧，但从表 4（b）结果中可看出，京津沪虚拟变量（DBTS）系数符号为正——在其他变量相同的情况下，京津沪更有可能发放户口。在北京，有许多政府部门以及国有企业，在这些单位就业，户口问题较容易解决。<sup>29</sup>

<sup>28</sup> 户籍制度最松和最紧的一些省份的相关变量值在附录 3 中给出。

<sup>29</sup> 对于农村移民，他们几乎不可能在国有部门得到一份工作。但北京发放户口的概率仍然可以反映对非户口移民的歧视程度，作为中国的首都，北京户籍制度的滥用程度相对较低。

给定个人特征,图3同时给出了个人在各省获得户口的意愿,个人意愿也通过两套估计值(包括或不包括婚姻状况变量)来预测。结果显示,个人在京津沪获得户口的意愿较高,而在海南、新疆、黑龙江、广西等地的意愿较低。

个人意愿普遍较高(意愿 I 均值为 0.940, 意愿 II 均值为 0.774),大大高于政府发放户口的概率(SHS I 均值为 0.384, SHSII 均值为 0.427)。这表明,在各地移民获得户口的概率主要还是由政府决定。

表6 预测的户籍制度松紧程度指标与实际各类移民中户口移民比例的相关性

各省城市各类移民中的户口移民比例	与 SHS I 的相关系数	与 SHS II 的相关系数
SHS II/SHS I	0.967	0.967
城镇大专以上学历劳动力移民	-0.121	-0.167
城镇大专以上学历女性劳动力移民	0.076	-0.023
农村男性劳动力移民	0.267	0.165
农村女性劳动力移民	0.290	0.185
城镇大专以上学历男性劳动力省际移民	-0.094	-0.129
城镇大专以上学历女性劳动力省际移民	-0.002	-0.063
农村男性劳动力省际移民	0.421	0.427
农村女性劳动力省际移民	0.152	0.087
劳动力总移民	0.046	-0.014
城镇大专以上学历劳动力总移民	-0.029	-0.117
城镇大专以上学历劳动力省际移民	0.150	0.069
农村劳动力总移民	0.300	0.264
农村劳动力省际移民	0.013	-0.050
各省城市户籍迁入率 <sup>a</sup>	0.325	0.223
各省城市户籍迁入率(外省迁入) <sup>a</sup>	0.151	0.178

注:1. 除各省城市户籍迁入率和各省城市户籍迁入率(外省迁入)外,所有户口移民比例都从2000年人口普查抽样数据中计算。

2. a: 数据来源中华人民共和国公安部(1996、1997和1998)。各省城市户籍迁入率计算方法是:当年迁入各省城市地区户籍的人口除以当年该省城市人口,移民包括省内移民和外省移民。各省城市户籍迁入率(外省迁入)只包括了外省移民。

至于户籍制度松紧程度指标的有效性,我们可以对比这些指标和其他可能的指标。有人可能会认为,迁移到城市的各类移民中户口移民中的比例可以衡量当地的户籍制度。上文已经考虑这种衡量,因为存在选择性偏差,所以这种衡量是有偏的。表6给出了SHS I和SHS II与各类移民中的户口移民比例的相关系数,其数值在-0.094至0.421之间。可以看出,各个指标之间存在较大的差异。同时,表6的最后两行给出了本文构建的指标和各省城市户籍迁入率之间的相关系数,这两个相关系数也不大。因此我们可以认为,通过上述修正选择性偏差的程序,构建的指标已经消除了这些偏差。

## 五、结论和可改进之处

本文构建了户籍制度下的迁移决策模型，其中个人的户口状态由个人和政府双方联合决定，迁移时个人考虑户籍制度松紧程度。在此基础上，本文尝试地衡量了户籍制度松紧程度，即在给定个人特征的情况下，政府发放户口的概率可以衡量户籍制度松紧程度。

这个模型主要涉及了两个问题：户口状态联合决策模型中的部分可观察性（partial observability）和有多个标准的选择性偏差（selectivity bias with multiple criteria）。

部分可观察性是因为我们只能观察到迁移者最后的户口状态，而不知道个人是否想要，政府是否批准了户口申请。要识别这样的联合决策模型，需要至少有一个排除变量，它在个人或政府的户口决策中，但不在另一方的决策中。因为个人会比政府更多地考虑迁出地的特征，因此，这个识别条件可以得到满足。同时，由于我们只能看到某人迁移到某地后的户口状态，因此在估计移民户口状态联合决策模型时，还存在选择性偏差——这些移民是政府和个人选择之后的结果。要纠正选择性偏差，同时识别联合决策模型，这还需要一个识别条件——有一个排除变量只在迁移模型中，而不在户口决策模型中。我们把户口决策抽象为迁移者比较他在当地和迁出地的户口价值以及在当地作为户口移民和非户口移民的效用，因此，距离不在户口状态联合决策模型中，但距离会影响迁移，因此在迁移决策中。

选择性偏差是因为移民和政府都进行了选择——政府选择了在一定程度上的户籍制度的前提下，移民对此作出反应，最后迁移至某地。为了纠正选择性偏差，本文采用了两步法来估计：首先用 Nested Logit 方法估计了一个迁移模型的简化式，并且得到了纠正选择性偏差的因子；然后在纠正了选择性偏差的同时，估计了移民户口状态的联合决策模型。

我们利用从 2000 年人口普查的抽样数据中选取了大专毕业及以上学历（非成人教育学历）的男性劳动力样本，通过上述方法，估计了户口状态联合决策模型。结果显示：对于个人，学历越高，从城市迁出或迁到城市，就更想得到户口；对于政府，如果失业率越低、老年人口比例越低，或者户口申请者学历越高、年纪越轻，那么政府就更有可能发放户口给他们。根据这些估计值，文章构建了户籍制度松紧程度的一个衡量——控制了个人特征后，各省城镇地区发放户口的概率。这一指标显示，上海的户籍制度最紧，黑龙江、青海和宁夏户籍制度最松，而北京的户籍制度处于中等水平。另一方面，个人在京津沪获得户口的意愿较高，而在海南、新疆、黑龙江、广西等地的意愿较低。

受限于数据等问题，本文存在着一些需要改进之处。首先是户籍制度

指标的有效性。我们需要假设政府发放户口的概率与其歧视非户口移民政策存在反向关系,这仍然需要证据来验证。户籍制度指标估计的稳健性(robustness)仍然需要进一步验证。

其次是估计中的联立偏差问题。在我们估计中,一个潜在的假设是政府首先确定了户籍制度的松紧程度,然后个人对此作出反应。而实际情况可能是政府的户籍制度松紧程度随着移民的多少而变化着——即两者的决策是联立决定的。由于我们利用了城镇高学历男性劳动力样本来估计和构造户籍制度松紧程度指标,而各地区一般都不存在高学历劳动力过剩的现象,因此这一问题应已有所减轻。

再次是一些数据问题。2000年人口普查数据中,没有户口移民的原户口所在地信息,我们假设了户口移民的原户口所在地,数据中也没有微观工资数据,这必然对估计的有效性造成了影响。

最后,本文的分析只局限在个人层次上,但迁移事实上是一种家庭决策;当涉及到户籍制度时,家庭的作用更大。比如,在当前户籍制度下,农村移民在城市长期居住非常困难,因此一般会以个人暂时在城市工作的形式进行迁移。在分析中,考察劳动力迁移形式(暂时或长久迁移、家庭或个人迁移)的异同,将有助于我们加深对户籍制度的理解。

#### 附录1 Nested Logit 迁移决策模型简化式的估计

	模型 I		模型 II		模型 III		模型 IV	
	系数	标准差	系数	标准差	系数	标准差	系数	标准差
最底层的选择——选择迁移到某省								
URINC95	0.464*	(0.083)	0.467*	(0.084)	0.520*	(0.086)	0.513*	(0.084)
RUINC95	0.650*	(0.084)	0.642*	(0.085)	0.663*	(0.087)	0.461*	(0.084)
ER95	0.150*	(0.029)	0.152*	(0.030)	0.152*	(0.030)	0.110*	(0.028)
GR	0.117*	(0.017)	0.117*	(0.017)	0.115*	(0.017)	0.140*	(0.015)
UNEM95	-0.074*	(0.015)	-0.074*	(0.015)	-0.074*	(0.015)	-0.068*	(0.014)
DAPC	-0.172	(0.260)	-0.184	(0.260)	-0.213	(0.263)	0.028	(0.248)
POPGR	0.430*	(0.106)	0.431*	(0.107)	0.433*	(0.108)	—	—
GEGDP95	0.109*	(0.018)	0.109*	(0.018)	0.108*	(0.018)	0.119*	(0.016)
DIS	-0.839*	(0.051)	-0.839*	(0.051)	-0.837*	(0.053)	-0.825*	(0.053)
BTS	-0.080	(0.153)	-0.086	(0.154)	-0.107	(0.156)	-0.400*	(0.123)
IN_PROV	2.258*	(0.127)	2.258*	(0.128)	2.263*	(0.133)	2.264*	(0.131)
第一层选择——不迁移、迁移到农村地区或迁移到城镇地区								
EDU1UR	—	—	-0.419*	(0.054)	-0.418*	(0.054)	-0.418*	(0.054)
EDU1RU	—	—	0.604*	(0.208)	0.611*	(0.209)	0.619*	(0.209)
EDU2UR	0.396*	(0.055)	—	—	—	—	—	—
EDU2RU	0.694*	(0.133)	—	—	—	—	—	—
EDU3UR	-0.597*	(0.209)	—	—	—	—	—	—
EDU3RU	-0.866	(1.018)	—	—	—	—	—	—

(续表)

	模型 I		模型 II		模型 III		模型 IV	
	系数	标准差	系数	标准差	系数	标准差	系数	标准差
AGERU	-0.473*	(0.067)	-0.225*	(0.066)	-0.275*	(0.077)	-0.278*	(0.077)
AGESQRU	0.005*	(0.001)	0.005*	(0.001)	0.005*	(0.001)	0.005*	(0.001)
AGEUR	-0.012***	(0.007)	0.237*	(0.006)	0.154*	(0.007)	0.152*	(0.007)
AGESQR	0.000**	(0.000)	0.000**	(0.000)	0.000*	(0.000)	0.000*	(0.000)
MARRU	—	—	—	—	-0.013	(0.233)	-0.009	(0.233)
MARUR	—	—	—	—	0.179*	(0.049)	0.172*	(0.050)
常数项×Urban	-1.988*	(0.125)	-1.675*	(0.124)	-1.870*	(0.123)	-1.743*	(0.131)
常数项×Rural	6.798*	(1.135)	6.205*	(1.162)	5.875*	(1.330)	5.736*	(1.329)
<i>Inclusive Value</i> 的系数								
不迁移	0.327*	(0.063)	0.319*	(0.063)	0.299*	(0.064)	0.330*	(0.065)
迁移到城镇	0.395*	(0.073)	0.399*	(0.073)	0.395*	(0.074)	0.414*	(0.075)
迁移到农村	0.217*	(0.079)	0.212*	(0.079)	0.164**	(0.082)	0.269*	(0.100)
对数似然值	-7904.020		-7905.360		-7901.485		-7909.313	
样本大小:(61×10882) 663802								

注:1. 这里估计的模型是两层次的 nested logit 模型,其结构如图 2 所示。西藏没有包括在分析中,因此总共有 30 个省、市、自治区。模型 I-IV 中包括了不同的变量组合。

2. \*, \*\*, \*\*\*: 系数分别在 1%, 5% 和 10% 水平上显著不为零。括号中为标准差。

## 附录 2 政府户口决策模型中一些省份的变量值

省份	URINC95(+)	GR(+)	GEGDP95(-)	UNEM95(-)	ER95(-)	DAPC(+)	POPGR(-)
黑龙江	8.124	20.161	8.667	10.990	6.360	0.801	0.664
青海	8.108	16.020	17.423	7.336	5.250	0.722	0.727
宁夏	8.126	18.416	13.547	6.334	5.840	0.791	2.080
新疆	8.334	22.055	11.551	10.480	6.790	0.781	1.738
甘肃	8.056	15.177	14.709	5.615	6.460	0.947	1.423
北京	8.738	19.892	11.069	3.987	10.810	0.683	0.155
广东	8.914	26.599	9.767	7.058	11.500	0.762	1.740
贵州	8.277	16.588	13.543	9.653	8.620	0.496	1.212
天津	8.503	21.352	10.143	4.837	11.690	0.734	0.568
江西	8.125	20.118	9.156	8.896	10.090	0.586	1.055
云南	8.315	18.897	19.483	5.613	8.900	0.493	0.968
海南	8.470	25.876	11.639	21.497	10.240	1.092	1.563
上海	8.881	23.691	10.558	5.248	16.000	0.447	-0.306
均值	8.327	21.364	10.014	6.806	9.635	0.716	1.070
标准差	0.244	3.890	3.509	3.459	2.240	0.164	0.481
最小值	7.960	13.690	4.917	2.595	5.250	0.447	-0.306
最大值	8.914	29.711	19.483	21.497	16.000	1.092	2.080

注:变量名旁边括号内的符号表示该变量对政府发放户口的影响。这些省份已经按照户籍制度由松到紧排列,以北京为界,前五个是户籍制度最松的地区,后七个是户籍制度最紧的地区。

## 参考文献

- [1] Abowd, John and Henry Farber, "Job Queues and the Union Status of Workers", *Industrial and Labor Relations Review*, 1982, 35(3), 354—367.
- [2] 北京大学中国经济研究中心“城市劳动力市场”课题组,“上海:城市职工与农村民工的分层与融合”,《改革》,1998年第4期,第99—100页。
- [3] 北京大学中国经济研究中心“城市劳动力市场”课题组,“南京:福利惯性下的劳动力市场”,《改革》,1999年第4期,第85—92页。
- [4] Cai, Fang, "Institutional Barriers in Two Processes of Rural Labor Migration in China", Working Paper Series No. 9, Institute of Population Studies, Chinese Academy of Social Sciences, 2001.
- [5] 蔡昉、都阳、王美艳,“户籍制度与劳动力市场保护”,《经济研究》,2001年第12期,第41—49页。
- [6] Chan Kam Wing, Ta Liu and Yunyan Yang, "Hukou and Non-hukou Migration: Comparisons and Contrasts", *International Journal of Population Geography*, 1999, 5, 425—448.
- [7] Chan, Kam Wing and Li Zhang, "The hukou System and Rural-Urban Migration in China: Processes and Changes", *The China Quarterly*, 1999, 160, 818—855.
- [8] 陈峰、王雷,“被收容者孙志刚之死”,《南方都市报》,2003年4月25日。
- [9] Cheng, Tiejun and Mark Selden, "China's Hukou System", *China Quarterly*, 1994, 139, 644—668.
- [10] Falaris, Evangelos, "A Nested Logit Model with Selectivity", *International Economic Review*, 1987, 28(2), 429—443.
- [11] Fields, Gary. S., "Place-to-place Migration: Some New Evidence", *The Review of Economics and Statistics*, 1979, 61(1), 21—32.
- [12] 付艳霞、赵相林,“五元卫生费为何难执行?”《佛山日报》,2003年9月8日。
- [13] Greene, William, *Econometric Analysis* (Fifth Edition). FT Prentice Hall, 2003.
- [14] Greenwood, M. J., "Research on Internal Migration in the United States", *Journal of Economic Literature*, 1975, 13, 397—433.
- [15] Greenwood, M. J., "Internal Migration In Developed Countries", in M.R. Rosenzweig and O. Stark eds., *Handbook of Population and Family Economics*, Chapter 12. Elsevier Science, 1997.
- [16] 国家统计局,《中国统计年鉴》。北京:中国统计出版社,相关年份。
- [17] Harris, J.R. and M. P. Todaro, "Migration, Unemployment and Development: A Two Sector Analysis", *American Economic Review*, 1970, 60, 126—142.
- [18] 何英华,“户口移民和非户口移民:1995—2000”,未发表文稿,北京大学中国经济研究中心,2003年。
- [19] Holtz-Eakin, Douglas, "Is Health Insurance Crippling the Labor Market?" *Policy Briefs*, The Jerome Levy Economics Institute, December, 1993.
- [20] Lee, Lung-Fei, "Generalized Econometric Models with Selectivity", *Econometrica*, 1983, 51(2), 507—512.
- [21] Liang, Zai and Michael White, "Market Transition, Government Policies, and Interprovincial Migration in China: 1983—1988", *Economic Development and Cultural Change*, 1997, 45(2), 321—339.
- [22] 林毅夫、蔡昉、李周,《中国的奇迹:发展战略与经济改革》。上海:上海三联书店和上海人民出版社,1994年。
- [23] Ma, Z. and Liaw, K., "Explaining Hierarchical and Interprovincial Migrations of Chinese Young Adults by Personal Factors and Place Attributes: A Nested Logit Analysis", *Mathematical Population Studies*, 1997, 6(3), 217—239.



- [24] Maddala, G. S., *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press, 1983.
- [25] McFadden, Daniel L., “Econometric Models of Probabilistic Choice”, in C. F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data With Econometric Applications*, Cambridge, Mass.: MIT Press, 1981, 198—272.
- [26] 全国人口抽样调查办公室,《1995年全国1%人口抽样调查资料》。北京:中国统计出版社,1997年。
- [27] Schultz, T. P., “Lifetime Migration within Educational Strata in Venezuela: Estimates of a Logistic Model”, *Economic Development and Cultural Change*, 1982, 30(3), 559—594.
- [28] Schwartz, Aba, “Interpreting the Effect of Distance on Migration”, *Journal of Political Economy*, 1973, 81(Sep.-Oct.), 1153—1169.
- [29] 宋洪远,“关于农村劳动力流动的政策问题分析”,天则双周学术讨论会,2001年11月30日。  
(<http://www.unirule.org.cn/symposium/c205.htm>).
- [30] 铁道部编,《全国铁道旅客时刻表》,1997—1998年。北京:中国铁道出版社,1997年。
- [31] Todaro, Michael P., “A Model of Labor Migration and Urban Unemployment in Less Developed Countries”, *The American Economic Review*, 1969, 59(1), 138—148.
- [32] 王格玮,《中国人口迁移的因素分析:一个基于 nested logit 多元选择模型的框架》,硕士学位论文,北京大学中国经济研究中心,2004年。
- [33] 项飏,《跨越边界的社区——北京“浙江村”的生活史》。北京:生活·读书·新知三联书店,2000年。
- [34] Yao, Yang, “Social Exclusion and Economic Discrimination: The Status of Migrations in China’s Coastal Rural Area”, Working paper E2001005, China Center for Economic Research, Peking University, 2001a.
- [35] Yao, Yang, “Egalitarian Land Distribution and Labor Migration in Rural China”, Working paper E2001007, China Center for Economic Research, Peking University, 2001b.
- [36] 查瑞传等主编,《中国第四次全国人口普查资料分析》。北京:高等教育出版社,1996年。
- [37] 张曙光、施贤文,“市场分割、资本深化和教育深化”,未发表文稿,2003年。
- [38] Zhao, Yaohui, “Labor Migration and Earnings Differences: The Case of Rural China”, *Economic Development and Cultural Change*, 1999, 47(4), 767—782.
- [39] Zhao, Zhong, “Migration, Labor Market Flexibility, and Wage Determination in China: A Review”, Working Paper, E2003007, China Center for Economic Research, Peking University, 2003.
- [40] 中华人民共和国公安部编,《中华人民共和国全国分县市人口统计资料》,1995年度,1996年度,1997年度。北京:中国人民公安大学出版社,1996、1997、1998年。

## Measuring the Strictness of the Household Registration System

YINGHUA HE  
(Peking University)

**Abstract** This paper develops a migration model under China’s household registration system and tries to measure the strictness of the system. The estimation involves partial observability

and selectivity biases with multiple criteria. The paper uses variable exclusion to identify the model and thus solves the problem of partial observability. It then adopts a 2-step strategy to correct the selectivity bias. A sample of urban male with university or higher degrees is drawn from the 2000 population census data, and is used for the estimation.

**JEL Classification** C25, C81, J68