



No. C2004016

2004-10

1995-2000 年户口迁移和非户口迁移：  
描述与分析

何英华\*

No. C2004016

2004 年 10 月 8 日

---

\* Graduate Student, Department of Economics, Columbia University. Email: [yh2165@columbia.edu](mailto:yh2165@columbia.edu). 感谢赵耀辉教授给予的悉心指导、沈明高教授和赵忠教授的宝贵意见；与王格玮、赵冬以及其他同学的众多讨论也令我获益匪浅。

1995-2000 年户口迁移和非户口迁移：  
描述与分析

何英华<sup>\*</sup>

No. C2004016      2004 年 10 月 8 日

**摘要：**利用 2000 年人口普查抽样数据，本文具体描述了户口制度下 1995-2000 年省际户口和非户口移民的特征，发现他们在多个方面存在的根本差异。在此基础上，本文利用各地间迁移模型（Place-to-place Migration Model）对户口移民和非户口移民进行计量分析，发现了户口迁移和非户口迁移的不同机制：一些经济因素可较好的解释非户口迁移，而对户口移民的作用很多不同于预期。

**关键词：**户口制度，迁移，各地间迁移模型。

**JEL 分类：**J18, O15, R19

---

<sup>\*</sup> Graduate Student, Department of Economics, Columbia University. Email: [yh2165@columbia.edu](mailto:yh2165@columbia.edu). 感谢赵耀辉教授给予的悉心指导、沈明高教授和赵忠教授的宝贵意见；与王格玮、赵冬以及其他同学的众多讨论也令我获益匪浅。

# 1995-2000 年户口迁移和非户口迁移：描述与分析

何英华

**摘要：**利用 2000 年人口普查抽样数据，本文具体描述了户口制度下 1995-2000 年省际户口和非户口移民的特征，发现他们在多个方面存在的根本差异。在此基础上，本文利用各地间迁移模型（Place-to-place Migration Model）对户口移民和非户口移民进行计量分析，发现了户口迁移和非户口迁移的不同机制：一些经济因素可较好的解释非户口迁移，而对户口移民的作用很多不同于预期。

**关键词：**户口制度，迁移，各地间迁移模型。

## *Hukou and Non-Hukou Migrations in China: 1995-2000*

HE Yinghua

*(China Center for Economic Research, Peking University)*

### **Abstract:**

Using the 2000 census data, the paper compares and contrasts the *hukou* and non-*hukou* migration in China during 1995 through 2000. With an adjusted Place-to-Place Migration Model, the determinants of the two types of migrants are investigated. The inter-provincial migrants are further divided into rural labor migrants and rural-urban labor migrants. The empirical results show that some economic variables have a great influence on the non-*hukou* migrants, but have little on the *hukou* migrants. This indicates the household registration system have an impact on the migration. The paper then conducts a test of structural difference (similar to the Chow test), and the result show that the equations of *hukou* and non-*hukou* migration are structurally different.

**JEL Classification:** J18, O15, R19

## 一、引言

2003年8月1日起，国务院颁布的《城市生活无着的流浪乞讨人员救助管理办法》正式施行，同时，施行了21年的《城市流浪乞讨人员收容遣送办法》宣告终结。这一转变的直接原因是一个大学生——孙志刚的意外死亡（陈峰，王雷，2003）；更重要的，群众对于当前户籍制度的强烈改革愿望是促成这一改变的根本原因。

追根溯源，现行的户籍制度起源于建国初期；1958年1月9日，新中国颁布了第一部户籍制度——《中华人民共和国户口登记条例》，其中包括常住、暂住、出生、死亡、迁出、迁入、变更等7项人口登记制度。它以法律形式严格限制农民进入城市，限制城市间人口流动，城乡分离的“二元经济模式”由此而生。1975年修正的《宪法》颁布，取消了关于公民迁移自由的条文。

户籍制度是很多学者关注的对象，这方面的研究已有很多。对于户籍（户口）制度的较为详细的描述，有 Cheng 和 Selden（1994）以及 Chan 和 Zhang（1999），后者描述了中国大陆曾经实行过的各种户口制度，并且着重讨论了户口制度下，乡村到城市的迁移问题。

对于户籍制度的产生，林毅夫等（1994）提出了很有解释力的理论：制度选择内生于发展战略；建国初期重工业优先的发展战略使得中国形成了计划经济体制，而户籍制度是计划经济的重要组成部分；限制人口流动，对于保证城市“充分就业”和福利不外溢至关重要。至于改革后户籍制度得以延续，蔡昉等（2001）提出，这是由于城市利益集团的影响，他们还利用计划迁移数量决定的计量分析、以及北京市就业保护政策的演变实例证实了他们的假说。

蔡昉等（2001）所讨论的计划迁移指的是：地方政府批准户口变动为依据的迁移。也就是说，迁移者在更换居住地的同时，得到了迁移目的地的户口——这些包含于我们所定义的户口迁移：一个迁移者迁移到了一个目的地，同时拥有当地的户口<sup>1</sup>。反之，如果迁移者没有目的地的户口，那么她/他就是非户口移民。户口制度下非户口迁移的行为更是值得关注。因为生活中日益司空见惯的农民工，正是非户口移民，其数量也大大超过户口移民。以1995-2000年间的省际移民来看，总迁移人数高达3364万人，其中84.7%为非户口移民<sup>2</sup>。

针对改革后中国国内移民，目前已有较多的文献，Chan, Liu 和 Yang（1999）作了一个较好的总结。实证分析方面，大家一般所用到的数据是全国人口普查数据（1990年第四次普查和2000年第五次普查）、人口抽样数据（1987年，1995年等）以及一些专门的移民调查（参见 Mallee, 1996; Chan, Liu 和 Yang 1999）。这些数据都存在着一个共同的问题——虽然它们提供了较为详细的移民人口学特征信息，但是基本的大的调查中都没有微观工资数据或者收入数据，这使得我们在研究时只能利用宏观经济方面的数据。

另外，在现有的文献中，大多数移民研究没有深入探讨户口制度——这个在中国移民问题中无法忽视的因素。如，Liang 和 White（1997）的研究中，他们利用1988年2%计划生育调查的10%抽样数据，分析了省际迁移，他们所定义的移民是1983-1988年间发生过省际迁移的人口，并且把那些没有迁入地户口的迁移人口定义为暂时的省际移民（temporary migrants）。但在计量分析中，他们并没有区分移民是否暂时，仅在分析迁移人口选择哪个省作为迁入地时，在模型中加入了暂时移民分别和乡镇企业、工业产出以及人口的交叉项。这种方法忽视了户口迁移和非户口迁移之间可能存在的截然不同的内在机制。

在户口制度下，将迁移划分为户口迁移和非户口迁移的研究，有 Chan, Liu 和 Yang

<sup>1</sup> 这样定义户口迁移，要注意区分两种移民，第一种是迁移到目的地，同时把户口转到了目的地；第二种是迁移到目的地，但是没有转换户口，因为户口原本就在目的地——这第二种户口移民也就是回流的移民，也是许多农民工日后的选择。从严格意义上讲，只有第一种户口移民才是户口移民，但是大多数关于移民的数据——包括我们将用到的2000年人口普查数据，没有提供信息让我们区分这两种户口移民。因此，我们只能采用文中较为宽松的户口移民的定义。关于数据中移民回流的问题，第二部分还有相关讨论。

<sup>2</sup> 数字根据2000年人口普查0.95%抽样数据折算而来。移民的定义详见第二部分。

(1999)。他们利用 1990 年人口普查抽样数据, 首先对户口迁移和非户口迁移作了一个详细而清晰的描述, 总结了户口移民和非户口移民在个人特征、流向和集中程度上的相同点和不同点; 然后, 他们利用各地间迁移模型 (Place-to-place Migration Model) 分析了户口迁移和非户口迁移的决定因素。本文将借鉴这一方法, 但更注重从经济学的角度来分析 1995-2000 年间户口移民和非户口移民的决定因素。

同时, 在讨论一个地区户籍管制时, 也要注意, 现行的户籍管理制度可分为两方面: (1) 发放户口的管制——即政府决定哪些人可以得到户口; (2) 对于非户口居民的歧视和管制, 如暂住证、暂住费、就业限制等制度。两个方面相辅相成, 前者直接决定了哪些人可以逃脱第二方面的管制; 而后者存在体现出户口的价值所在, 才使得前者的存在有意义。

本文的剩余部分安排如下: 第二部分介绍数据和移民定义, 具体比较了省际户口迁移和非户口迁移的异同; 并且进一步地在总移民中划分出农村劳动力移民和农村到城镇劳动力移民, 来具体分析这几类移民中户口迁移与非户口迁移的异同。第三部分运用各地间迁移模型, 分析各类移民中户口移民和非户口移民的决定因素, 并且检验了户口移民和非户口移民是否具有结构性的不同。第四部分作了简要总结, 并且提出了本文一些可以改进的地方。

## 二、描述性分析

### (一) 数据和移民定义

本文所利用的数据是 2000 年人口普查 0.95% 抽样数据, 调查的标准时间是 2000 年 11 月 1 日零时。这一数据提供了非常详细的个人信息 (但很遗憾, 这一数据也没有微观工资或是收入数据), 主要有年龄、教育、婚姻状况、现住地信息; 在迁移信息方面, 我们可以知道此人何时迁来本地, 从何地迁来本地居住<sup>3</sup> (必须是 1995 年以后的移民); 同时, 我们也知道迁移者的户口状况, 根据我们的研究目的, 数据中户口状况可以区分为两种<sup>4</sup>: 1、居住在本地, 户口在本地; 2、居住在本地, 户口在外地。

本文中定义的移民包括下列三类人: (1) 年龄在 5 周岁以上, 户口状况是第 1 种, 并且到本地<sup>5</sup>来居住的时间是 1995 年 10 月 31 日之后; (2) 年龄在 5 周岁以上, 户口状况是第 2 种, 到本地来居住的时间是 1995 年 10 月 31 日之后, 并且在本地居住半年以上; (3) 年龄在 5 周岁以上, 户口状况是第 2 种, 在本地居住不满半年, 但离开户口登记地半年以上。

根据移民的不同户口状况, 第 (1) 类定义为户口移民<sup>6</sup>, 第 (2) 和第 (3) 类定义为非户口移民。这样的移民定义, 可以排除一些短期的流动人口, 比如探亲访友、旅游和出差人士, 而这些人并不是我们所要研究的对象。

除了对迁移时间的规定之外, 我们还在迁移地域上对移民作了分类, 迁出地和目的地都在同一省内, 即为省内移民; 如果迁出地是外省, 那么她/他就是省际移民。除了总的移民之外, 我们还关心从农村迁出的移民, 特别是农村劳动力的迁移行为, 所以定义农村劳动

<sup>3</sup>2000 年普查中有两项迁出地的信息: (1) 从何地来本地居住, 反映最后一次迁移的情况; (2) 五年前的常住地。本文所用的是 (1), 因为这样有更多的信息可以利用, 如更详细的地址, 迁出地的类型 (城镇还是乡村) 等。Chan, Liu 和 Yang (1999) 所用的是 (2), 他们研究的 1990 年人口普查数据没有提供第一项信息。

<sup>4</sup> 原数据提供的户口状况分为 5 种: (1) 居住本乡镇街道, 户口在本乡镇街道; (2) 居住本乡镇街道半年以上, 户口在外乡镇街道; (3) 在本乡镇街道居住不满半年, 离开户口登记地半年以上; (4) 居住本乡镇街道, 户口待定; (5) 原住本乡镇街道, 现在国外工作学习, 暂无户口。我们的研究不考虑向国外移民, 因此不考虑 (5) 这类人。同时, 文中的第 1 种户口状况的人包括 (1)(4) 两种人——因为我们可以认为, 户口待定的人正在办理本地户口 (当然有一定误差); 文中的第 2 种户口状况包括 (2)(3) 两类人。

<sup>5</sup> 正如脚注 4 所阐述的, 这里的本地与外地的区分是以乡镇街道划分的, 并且时间长度为半年; 这与较早的数据不太相同, 如 1990 年人口普查的本地外地划分是县 (市) 一级的, 而且时间长度为一年。因此, 两次普查反映的情况不能直接比较。

<sup>6</sup> 关于这样定义户口移民的一个潜在问题, 请参阅本文脚注 1 的讨论。

力移民为年龄在 15-64 岁之间、迁出地类型是乡村<sup>7</sup>的迁移者。更进一步，定义农村到城镇劳动力移民为年龄在 15-64 岁之间、迁出地类型是乡村、目的地类型是城镇<sup>8</sup>的迁移者。相应的，定义城镇劳动力移民为 15-64 岁之间、迁出地类型是城镇、具有非农业户口的迁移者。

## （二）省内省际移民与全国人口的个人特征比较

现行的户籍管理制度下，一方面政府控制着可获得当地户口的人数，另一方面制订各种“法规”和“规定”来限制非户口居民在当地的权利；通过歧视非户口居民，户口的作用得到体现。在表 1 中，我们简要列出户口居民和非户口居民主要区别，特别是一些非户口居民受歧视的领域。可以说，非户口居民在居住地受到各个方面的歧视：经济、政治，有时甚至人身安全也没有保障（甚至有像孙志刚那样的惨剧！）。

在这样两方面的管制下，个人在做迁移决策时，必然也会将两种迁移——户口迁移和非户口迁移的巨大差异考虑在内。也可以预见，这两类迁移会呈现不同的图景。表 2 给出了 1995-2000 年间各类移民与全国人口个人特征对比。表格的第二列给出的是全国 5 岁以上人口的各项特征，作为对比的参照组。表格的第三、四、五列分别是所有省内省际移民、省内省际户口移民和非户口移民的特征。最后的两列是省际户口移民和非户口移民的特征。

从表 2 中，可以发现以下几点：

（1）从性别看，总移民（省内省际移民）和总的户口移民中，男性比例都低于全国人口中男性比例，特别是总户口移民中，男性比例为 45.68%，为各类人口中最低。可能的解释是，在中国，结婚之后一般都是女方迁至男方所在地，所以总移民中，女性迁移比例要高于男性迁移比例；另一方面，婚迁一般容易换户口，因此户口移民中女性比例也较高。但在省际移民中，男性比例都较高，且高于全国人口比例。一个可能的解释是，女性迁移对距离要比男性敏感，并且距离越大，婚迁的可能就越小。

（2）从年龄分布看，移民一般集中在 15-39 岁之间，尤其是 20-29 岁的人比例最高。这一年龄更换居住地，最有可能产生终生的影响，这与 Chan, Liu 和 Yang（1999）等研究的发现相一致。另外，省际移民的年龄分布更加集中，15-39 岁的省际户口移民比例为 85.29%，省际非户口移民中这一比例为 83.62%。

（3）从婚姻状况看，各类移民中已婚的比例均低于全国水平。其中，女性的已婚比例大大高于男性的已婚比例；户口移民中女性已婚的比例要高于非户口移民中女性已婚的比例；这从另一方面说明了婚迁对于女性的重要性。而男性的分户口状况的婚姻状况分布正好与女性相反——户口移民中男性未婚比例最高（64.43%）。

（4）从现住地类型看，各类移民主要迁往城市，比例都在 50% 以上；而小城镇户口吸引力不大——总户口移民中迁往镇的只有 19.41%，省际户口移民迁往镇的只有 9.88%。另外，省际户口移民迁往农村的比例（36.76%）在各类移民迁往农村的比例中最高。

（5）从迁出地类型看，户口移民从街道迁出的比例较高，因为大学生毕业后工作，容易获得户口；非户口移民中，从农村迁出的移民占绝大部分（3/4 以上）。户口管制目的之一在于严格控制农村到城市的迁移，同时，户口发放对于教育程度要求很高，因此农村迁出的移民很难获得户口。

（6）从受教育程度看，移民的教育水平总体上要高于全国平均水平。而户口移民的教育水平是各类移民中最高的：总的户口移民中大学专科以上学历的有 27.74%，省际户口移民中这一比例高达 36.62%（并且有 28.72% 在大学本科以上），全国人口的这一比例仅为 3.94%。在两类非户口移民中，小学和初中是主要的教育水平，两者总比例约为 70%。这体现了户

<sup>7</sup> 本数据中，迁出地类型分为：乡、镇的居委会、镇的村委会、街道；我们将乡和镇的村委会划分为农村，将镇的居委会和街道划分为城镇。

<sup>8</sup> 很遗憾，迁入地类型的划分没有迁出地类型那样的详细，它只分为城市、镇、乡村，没有进一步的说明；我们把城市和镇划分为城镇。

口发放对于教育水平的挑选。

(7) 从劳动力比例<sup>9</sup>看, 省际非户口移民中比例最高 (89.61%), 户口移民中最低 (略高于 50%), 与全国水平相比, 低了约 20 个百分点。这是因为户口移民中有很一部分是学生 (特别是大学生), 而他们不属于劳动力。这表明非户口移民因为工作因素而迁移的比例较高。

(8) 户口移民在总的移民中所占比例很小, 全部省内省际移民中, 户口移民只占总移民的 29.34%; 省际移民中, 户口移民只占 15.18%。

综上所述, 户口移民与非户口移民存在不容忽视的差别; 另外, 省际移民与总的移民相比, 趋势基本一致, 且在一定程度上, 省际移民只是放大了总移民的一些特征。为与其他已有文献 (如 Liang 和 White, 1997, Ma 和 Liaw, 1997; Chan, Liu 和 Yang, 1999) 保持一定的可比性, 下面的讨论将集中在省际移民上。为了更好的了解我们关心的农村移民问题, 下文将细分各类移民, 分别进行讨论。

### (三) 各类移民的迁移原因比较

表 3 给出了总的省际移民、省际农村劳动力移民、省际农村到城镇劳动力移民和省际城镇劳动力移民的迁移原因分布。

表 3 中一个显著特征是: 非户口移民中“务工经商”的比例非常高, 两类农村移民中这一比例高达 85.68% 和 87.48%; 这一比例最低的是省际城镇劳动力移民, 也高达 50.67%。

除省际城镇劳动力移民外, 各类户口移民中, 比例最高的两项原因是“婚姻迁入”和“学习培训”, 这与户口管制一致, 因为户口政策规定能户口迁移的人主要是大学生和配偶。引人注目的是, 省际农村到城镇劳动力户口移民中, “婚姻迁入”仅占 12.42%, 是各类户口移民中比例最低的; 而“学习培训”的比例高达 58.39%, 为各类户口移民中比例最高。这跟我们现实的观察一致——农村人要想获得一个城市户口, 最主要的出路还是考大学; 城镇居民的农村配偶要想获得城镇户口, 并非易事。

省际城镇劳动力移民户口中, “学习培训”的比例依然是最高的, 占 47.57%; 但更引人瞩目的是“分配录用”和“工作调动”比例 (15.72% 和 14.83%), 高于其他各类移民。这一方面反映了大学生毕业 (各高校都在城镇) 后的迁移情况, 他们可以较容易的获得当地户口 (城镇劳动力户口迁移者中, 大专学历的占 14.38%, 大学本科以上学历的占 50.02%, 教育水平大大高于其他类型的移民), 另一方面也反映了户口管制政策对于城镇户口迁移者的控制——工作调动还带有浓重的计划经济色彩。

总的来说, 在迁移原因上, 非户口移民主要因为经济因素而自发迁移 (“务工经商”), 而户口移民的途径还带有计划调配的色彩 (“工作调动”、“分配录用”)。将移民按照迁出地进行细分后, 这样的差异更加明显。

### (四) 各类移民的行业、职业分布和失业率

大多数迁移者为了更高的收入而背井离乡 (Greenwood, 1969, 1975; Harris 和 Todaro, 1970), 总想得到更好的工作。但非户口移民在很多方面受到歧视, 又不能很容易的得到当地户口——这两方面的管制下, 必然会对户口和非户口移民的行业、职业分布产生影响。

按照 2000 年行业年平均工资排序, 分析移民的行业分布, 我们可以发现:<sup>10</sup>

(1) 省际农村劳动力移民的行业分布基本类似于省际农村到城镇劳动力移民; 只是前者有更多的户口移民从事农业 (69.64%), 从事制造业的人数较少 (10.46%)。这是因为农村移民中有约 1/2 的户口移民目的地是农村 (表 3 显示, 50.54% 的农村户口移民目的地是农村)。

<sup>9</sup> 劳动力定义为: 正在工作或是正在找工作的 15-64 岁的人。因此劳动力比例等于劳动力总数比上 15-64 岁总人数。请注意, 这与之前定义的劳动力移民不同, 劳动力移民定义只是规定了年龄在 15-64 岁之间。

<sup>10</sup> 限于篇幅, 文中没有列出各类移民具体的行业和职业分布表, 如有需要, 可向作者索取。

(2) 非户口移民大多分布在制造业、建筑业, 以及批发零售贸易和餐饮业——在这 3 个行业就业的人数超过 55% ;

(3) 对于平均工资最高的 6 个行业, 户口移民在这些行业就业比例基本都显著高于非户口移民, 并且大多数行业高于全国平均水平; 同时, 像国家机关、政党机关和社会团体这样政府完全控制的行业, 户口移民的比例更是大大高于非户口移民。

(4) 从农村到城镇户口移民中, 有 23.02% 从事农林牧渔业, 这个比例如此之高, 可能的原因是我们对于迁移目的地没有非常详细划分, 特别是未能区分出镇的村委会和镇的居委会。

再从各类移民的职业分布看, 对于各类非户口移民, 主要从事的职业是: 商业、服务业人员, 生产、运输设备操作及有关人员, 建筑材料生产、加工人员及其它——三类职业就业总计超过 75% (城镇劳动力非户口移民稍少, 约 60%)。而各类户口移民中专业技术人员比例明显高于相应的非户口移民——这反映出政府发放户口时, 对技术水平的要求。

最后, 从各类移民的失业率<sup>11</sup>分析, 户口移民的失业率大大高于非户口移民的失业率; 非户口移民的失业率很低, 且低于全国水平, 这可能是由于非户口移民的流动性大, 而且对于就业岗位不挑剔 (也不存在可挑剔的余地), 因此大多能找到工作, 实在找不到工作的, 早已换地方或回乡了; 户口移民失业率较高 (高于全国水平), 这一方面是他们对失业的承受力较高——包括在当地的社会保障体系中, 另一方面, 能够得到当地户口说明他们在当地有一定的资源可以利用 (如人际关系等), 因此可以承担成本, 慢慢等待找到好工作。

### (五) 各类移民的流向分析

图 1 描述了省际户口移民和非户口移民最大的 30 个流向<sup>12</sup>。由于非户口移民占总移民的大多数 (84.73%), 因此总移民的流向基本与其类似。

图 1(a) 描述的省际户口移民数量占总户口移民的 23%。与图 1 (b) 相比较, 图中一个显著的特征是迁移方向分散, 距离较短 (一般都是邻省迁移)。另外, 山东与东北三省的迁移联系明显, 这与之前的研究结果一致 (如 Chan, Liu 和 Yang, 1999; Ma 和 Liaw, 1997)。还有一个引人注意的现象是: 从江苏、上海和浙江迁移到安徽的户口移民数量不少; 从广东迁移到四川、湖南和江西的户口移民也有很多。对于这种现象的一个推测是移民回迁——这些人以前到外地打工而且没有更换户口, 现在回到了户口所在地。数据中没有关于移民回迁的直接数据, 但从其他一些信息来看, 回迁的可能性最大。<sup>13</sup>

图 1(b) 省际非户口移民最大的 30 个流向包括了总非户口移民的 60%。从图上看出, 东南沿海 (江浙沪闽粤, 特别是粤) 如同一个巨大的引力中心, 把邻省的移民吸引过去。其他的迁移目的地主要还有北京 (北方的经济中心, 全国的文化、政治中心) 和新疆 (边疆的一个资源丰富、地广人稀的自治区)。从迁出地看, 上述沿海省市的内陆邻省是最主要的移民来源; 另一个重要来源是四川 (人口稠密的西部省)——虽然距离主要迁移目的地较远, 但从该省迁出的移民依然为数不少。

考虑到图 1(a) 中反映的移民回迁问题, 并且占移民绝大多数 (78.75%) 的农村移民是更

<sup>11</sup> 根据 2000 年人口普查数据, 各类移民失业率如下表:

	全国人口 (15 岁)			全部省际移民		省际农村到城镇劳动力移民		省际城镇劳动力移民	
	全部	镇	城市	户口	非户口	户口	非户口	户口	非户口
失业率	2.55	6.3	9.29	4.71	1.71	12.58	1.85	6.71	3.9

<sup>12</sup> 图中所画出的迁移方向不一定只有 30 个, 有几个图超过 30 个迁移方向; 这是因为对有几类移民的数量排序后, 人数第 30 多的迁移方向有并列的几个。

<sup>13</sup> 以江苏到安徽的户口移民为例: 样本中总的户口移民是 52 人; 从户口状况看, 88% 是农业户口; 从现住地类型看, 有 72% 是农村; 从迁出地类型看, 有 60% 从城镇迁出; 从教育状况看, 有 90% 是初中或以下的; 从迁移原因看, 投亲靠友和务工经商各占 28%, 婚姻迁入仅有 15%。从一个较为发达的城镇, 迁移到较为落后的农村, 具备了当地的农业户口, 而且婚姻迁入不是主要原因, 唯一合理的解释是移民的回迁。



值得关注的对象,因此应对农村移民进行单独分析。图 2 描述了农村劳动力移民的迁移方向。

图 2(a) 描述的农村省际户口移民数量占农村户口移民的 27%。与省际农村劳动力非户口移民(图 2(b))相比较,户口移民目的地也很分散,邻省迁移较多。这种现象的一个理由是婚迁对于距离比较敏感,而婚迁是该类户口移民的主要原因(占 39.94%)。再与总的省际户口移民(图 1(a))比较,省际农村劳动力户口移民的迁移距离要稍大一些;并且图 1(b)中的移民回迁现象,在图 2(b)中没有体现出来——这也说明了“回迁”的推测是正确的,因为那些被怀疑是回迁的移民基本不是从农村出来的。

图 2(b)给出的省际农村劳动力非户口移民占该类移民的 65%。与图 1(b)比较,两幅图几乎完全一致——因为省际非户口移民主要是农村移民,而农村移民大多数也是非户口移民。

除以上两副图所描述的差异外,各类移民的流向也存在着集中程度的差异,这在图 3 得到反映——迁移方向最集中的是农村非户口移民,其次是总的非户口移民/农村劳动力移民(图中给出的是两条线,但是这两类移民集中程度基本一致,所以两条线基本重合了);而迁移方向最不集中的是总的省际户口移民。

上述分析揭示了户口移民与非户口移民的差异;为更具体地考察两类移民的不同机制,下文将把这两类移民分开,从计量经济学的角度考察两者的差异。

### 三、省际迁移的计量分析

#### (一) 各地间迁移模型 (place-to-place migration model)

各地间迁移模型研究的是各地区之间迁移量由哪些因素决定。这一模型与其他模型的不同之处在于,它关注的是总迁移量 (gross migration),而不是地区之间的净迁移量。<sup>14</sup>

文献中的各地间迁移模型有许多的函数形式,有线性,也有双对数(double-logarithmic)形式<sup>15</sup>。本文将采用的函数形式是双对数形式。这种形式的模型可以从条件 logit 模型 (conditional logit model) (McFadden, 1974)推导而来, Schultz(1982)利用了这一模型研究了人口迁移问题。

条件 Logit 模型中,一个在  $i$  地的个人迁移到  $j$  地的概率可以表示为:

$$P_{ij} = e^{Z_{ij}} / \sum_j e^{Z_{ij}}, i \neq j \quad (1)$$

$$\sum_j P_{ij} = 1, \forall i \quad (2)$$

并且  $Z_{ij}$  是迁出地和迁入地变量 ( $X_i$  和  $X_j$ ), 以及两地之间距离  $D_{ij}$  的对数线性函数, 可以表示为:

$$Z_{ij} = \mathbf{a} + \sum_m \mathbf{l}_m \ln X_{mi} + \sum_m \mathbf{g}_m \ln X_{mj} + \mathbf{d} \ln D_{ij}, i \neq j \quad (3)$$

对于不迁移的人, 概率为:

<sup>14</sup>Fields(1979)总结了这类模型的几点好处:(1)在完全了解地区之间得迁移流动之后,我们可以在此基础上计算地区间的净迁移量,而反过来则不可行;(2)这一模型与微观行为模型相对应;(3)这一模型使得我们能够研究特定迁出地和目的地组合;(4)可以让我们了解居住在不同地方的人们所面对的其他选择的个数;(5)可以揭示可能存在的迁出地和迁入地变量作用不对称问题(这里的不对称问题指的是:迁出地和迁入地相对应的变量——如迁出地的失业率和迁入地的失业率——对迁移产生的影响程度不同,反映在计量结果中,两个变量系数的绝对值可以不同)。

<sup>15</sup>各种函数形式的讨论和实证的结果比较可以参阅 Fields (1982)。

$$P_{ii} = e^{Z_{ii}} / \sum_j e^{Z_{ij}} \quad (4)$$

注意到迁移和不迁移存在着本质的区别，从更换居住地的成本来看，我们可以认为，不迁移（认为从本地迁到本地）到迁移，成本的变化是不连续的。我们在用相同的变量来衡量这两个选择的时候，必然不能要求两者的系数是相同的。<sup>16</sup>因此，

$$Z_{ii} = a^* + \sum_m l_m^* \ln X_{mi} + \sum_m g_m^* \ln X_{mi} = a^* + \sum_m (l_m^* + g_m^*) \ln X_{mi} \quad (5)$$

结合上面（1）—（5），可以得到：

$$\ln(P_{ij} / P_{ii}) = Z_{ij} - Z_{ii} = \tilde{a} + \sum_m \tilde{l}_m \ln X_{mi} + \sum_m \tilde{g}_m \ln X_{mj} + \tilde{d} \ln D_{ij} \quad (6)$$

系数上的一弯（~）表示是由（3）和（5）结合而来的系数。另外，由于  $P_{ij}$  的变化比  $P_{ii}$  要大得多<sup>17</sup>，因此（6）可以近似化简为：

$$\ln m_{ij} = \tilde{a} + \sum_m \tilde{l}_m \ln X_{mi} + \sum_m \tilde{g}_m \ln X_{mj} + \tilde{d} \ln D_{ij} + e_{ij} \quad (7)$$

其中， $m_{ij}$  是从  $i$  省迁到  $j$  省的移民比例； $e_{ij}$  是扰动项，可以解释为测量误差，以及其他一些无法衡量、但是与等式右边其他变量不相关的因素。<sup>18</sup>

在较早的研究中，如 Greenwood(1969)，很多模型在设定时加上了变量作用对称的限制，即  $X_{mi}$  和  $X_{mj}$  中相对应的变量系数符号相反，绝对值相等（即在实际回归中采用两个变量的比率）。这样的设定可以解释为：在  $i$  地某个指标上升了 1% 对于迁移行为的影响与  $j$  地相对应的指标下降 1% 的影响大小是一样的。与经典的推力-引力模型（push-pull model）相联系，这一假定相当于推力与引力发生的机制是一样的。而在（7）中，这一假定已经得到了放松。

结合户口制度，移民可划分为户口移民和非户口移民（第二部分的分析反映了两者之间的巨大差异），与第二部分一样，根据迁出地和迁入地的不同，还将细分出几类移民。

## （二）关于应变变量 $m_{ij}$ 的一点讨论

对于（7）式中的应变变量  $m_{ij}$ ，它从  $i$  省迁到  $j$  省的移民比例；由于数据的不全，并且从西藏迁出和迁入西藏的人数很少，我们将西藏排除在外。这样，每个省的迁移人口有 29 个选择，所有可能的省际迁移方向（迁出省—迁入省的组合）有  $(30 \times 29) 870$  个。但是，我们的数据是 2000 年人口普查 0.95% 抽样数据，不可避免的，有多个迁移方向的移民人数是零（省际总移民中，有 114 个迁移方向人数为零，细分的省际移民有更多方向迁移人数为零）。

在我们的回归方程（7）中， $m_{ij}$  需要取对数后进行回归，但是  $m_{ij}$  又存在一些数值为零

<sup>16</sup> 可以参阅 Schultz(1982)年的讨论，他认为，这样的模型其实包含了两步的决策——是否迁移，决定迁移后选择哪个目的地；这样的话，就需要构造一个 Nested Logit 模型进行估计，对于中国国内迁移问题运用这一方法的研究，可以参阅 Ma 和 Liaw(1997)和 Wang(2003)。

<sup>17</sup> 由于中国的人口众多，除少数几个省外，各省的人口都在千万以上，非移民数量很大，比例比较稳定。在数据中（西藏除外），1995-2000 年未进行省际迁移的比例最高为广东省（99.38%），最低为江西省（93.28%），均值为 97.48%，标准差为 1.56%（仅为均值的 1.60%）。而各省之间的总迁移比例  $P_{ij}$  最高是湖南到广东（4.05%），最低是零（有 114 个），均值是 0.088%，标准差是 0.278%（为均值的 314.61%）。因此可以认为， $P_{ij}$  的变化比  $P_{ii}$  要大得多。可以参阅 Fields(1977,1982), Chan, Liu 和 Yang (1999)的讨论。

<sup>18</sup> （7）式的所有变量都取了对数形式，Schultz(1982)论证了这种形式的几点好处：（1）期望工资假说提出了工资与就业率之间乘积关系的互动，这可以通过对数形式在计量模型中设定；（2）当迁移的成本主要是机会成本时，期望收入可以用来近似估计迁移回报；（3）通过这种形式的函数设定，可以比较 conditional logit 模型和其他对数形式的非 logit 模型；（4）在实证研究中，这种形式的设定能够得到更高的 R-square（拟和优度）。对于第（4），本文的研究也证实了这一点。亦可参阅 Fields（1977,1982）的总结。

的样本点，这给模型估计带来了一个难题。Chan, Liu 和 Yang (1999)年的研究中，也有同样的问题。他们的处理方法是放弃了那些为零的迁移方向，只对不为零的各个迁移方向的  $m_{ij}$  做了回归。很明显，这是一个有偏的估计；在  $m_{ij} > 0$  的条件下，(7) 的条件期望可以表示为：

$$E(\ln m_{ij} | m_{ij} > 0) = \tilde{\alpha} + \sum_m \tilde{\Gamma}_m \ln X_{mi} + \sum_m \tilde{\beta}_m \ln X_{mj} + \tilde{\delta} \ln D_{ij} + E(e_{ij} | m_{ij} > 0) \quad (8)$$

只有满足  $E(e_{ij} | m_{ij} > 0) = 0$  的条件下，只对不为零的各个迁移方向的  $m_{ij}$  回归估计才是无偏的；但是， $m_{ij}$  与  $e_{ij}$  必定相关， $E(e_{ij} | m_{ij} > 0) \neq 0$ 。

Fu 和 Gabriel (2002)应用 GMM (广义矩估计) 方法来解决这一问题，他们认为，从每个省迁移出来的人一般只会选择少数几个目的地省，因此观察到很多迁移方向的迁移率很小甚至为零；他们也认为不能将为零的迁移方向舍弃，更重要的是：系数的估计对这些很小(甚至为零)的迁移率很敏感。因此，他们运用了更加一般化(也是更复杂的)GMM 进行估计。

本文将不采用 GMM 进行估计；对于 Fu 和 Gabriel 的论证有几点值得进一步思考的地方：(1) 关于每个省移民目的地的选择问题，“只会选择少数几个目的地省”的结论有些仓促。这可能与他们所用的数据有关，他们的数据是 1995 年 1% 人口抽样调查数据。这个调查的设计很有可能遗漏移民。<sup>19</sup> 赵耀辉等 (2003) 比较了 1990 年人口普查数据和 1995 年人口抽样调查数据中的移民数量，发现 1995 年的各类移民反而少于 1990 年的移民数量。<sup>20</sup> (2) 我们要研究的是迁移率有哪些因素决定，较小的迁移率对于系数的影响大，这正是我们模型的应有之义。(3) GMM 可以得到无偏的估计，但是其应用的复杂性也让我们考虑：是否有“更经济”的方法？

本文对这一问题的处理是在样本计算得出的迁移率  $m_{ij}^*$  上加上一个正的微小量  $\mathbf{j}$ <sup>21</sup> 使得回归式中的  $m_{ij} = m_{ij}^* + \mathbf{j}$ 。<sup>22</sup> 这样对结果不会产生很大的偏差，而益处在于我们可以利用那些原来不能利用的信息了。

根据移民定义，我们计算  $m_{ij}$  具体的方法是：先得出样本中该类省际移民的具体数目

<sup>19</sup>1995 年人口 1% 抽样调查的抽样方法是：分层、多阶段、整群抽样——首先按照地区平均抽样确定调查小区，然后按照居委会名册进行入户调查。而一般的移民，居住地不是很稳定，而且一般也倾向于不去居委会登记。详见全国人口抽样调查办公室 (1997)。

<sup>20</sup> 1990 年人口普查中的移民定义要比 1995 年抽样数据的定义严格，所以，如果抽样无偏的话，前者倾向于低估移民。详见查瑞传 (1996)，全国人口抽样调查办公室 (1997)，赵耀辉等 (2003)。

<sup>21</sup> 在实际回归时，取  $\mathbf{j} = 1 \times 10^{-22}$ ，这一取值是较为任意的。正如后面讨论的，从实证的结果来看，对于任意  $0 < \mathbf{j} < 1 \times 10^{-10}$ ，对于加上  $\mathbf{j}$  后回归的结果没有什么大的差别，具有一定的稳健性 (robustness)。

<sup>22</sup> 这样处理的理由如下：(1) 人口普查本身存在着一定的漏登率，2000 年人口普查的漏登率是 1.81%。如果漏登是完全随机的，那么对我们的结果不会有什么影响。但是一个移民被遗漏登记的概率高呢，还是一个长期居民被遗漏的概率高？显然，前者会更高，特别是农民工之类流动性很强的移民。因此，在样本计算出来的迁移率上加上一个正的微小量  $\mathbf{j}$ ，可以看作是对这种系统性偏差的一定纠正。(2) 单纯从随机测量误差的角度来看，对于所有样本迁移率  $m_{ij}^* = 0$  的迁移方向，它们的测量误差一定是负的；而且，我们也没有理由相信另外所有的  $m_{ij}^* > 0$  中所包含的测量误差期望为正，这样，总的测量误差期望为零的假设很难满足； $\mathbf{j}$  有助于调整这一类的误差。(3) logit 模型的理论基础来看， $m_{ij}$  衡量的是从  $i$  省迁到  $j$  省的概率；这一概率为零意味着  $i$  省的人毫无可能迁移至  $j$  省，但从实际来看，很难相信这一概率为零。从另一方面看，在离散选择模型中，概率很小也意味着观察到的  $m_{ij}$  很有可能为零。因此，用微小的概率来代替观察到的零概率，更符合实际情况，也不会有大的偏差。(4) 从实证的结果来看，对于任意  $0 < \mathbf{j} < 1 \times 10^{-10}$ ，加上不同  $\mathbf{j}$  后回归的结果没有什么大的差别。

$M_{ij}$ ，然后折算成全国实际数目，再比上 1995 年迁出省的总人口数  $pop95_i$ ，即  $m_{ij} = j + M_{ij} / (9.5 \times 10^{-4} \times pop95_i)$ 。另外，由于我们还将对移民作进一步细分，对于农村劳动力移民，计算时分母上的 1995 年迁出省的总人口数变成 1995 年迁出省的农村人口数<sup>23</sup>；对于城镇劳动力移民，在分母上人口数是 1995 年迁出省的城镇人口数<sup>24</sup>。

### (三) 自变量的选择与定义<sup>25</sup>

在迁移模型中，对于那些因素起着决定作用，Greenwood(1975)做了一个较为完善的综述。下文我们将结合中国的实际情况，确定回归方程(7)中所应包括的自变量。

#### 距离

有很多理由让我们把距离包括在迁移模型中，这些理由可以总结为两条：(1) 距离增加会导致迁移的成本上升，包括交通成本和心理成本；(2) 距离增加导致迁移者可以获得的迁移目的地信息减少。<sup>26</sup> 本文衡量的从  $i$  省到  $j$  省的距离  $D_{ij}$  以两省省会间的铁路距离衡量<sup>27</sup>，这是因为中国境内长途旅行的首要交通工具还是火车。<sup>28</sup>

除此之外，还应注意另一种情况：假设如果黑龙江到吉林的距离等于河南到山东的距离，而四省其它各因素都完全相同，但地理位置按照实际不变，那么可以预期，黑龙江到吉林的移民要多于河南到山东的移民。因为河南移民面临的其他选择的平均距离要小于黑龙江移民所面临的。<sup>29</sup> 因此，我们引入另一个距离变量  $AD_{ij} = \sum_{n \neq j} D_{in} / 28$ ，衡量的是从  $i$  省到除  $j$  省之外其他各省的平均距离。<sup>30</sup> 理论上预期，模型中  $D_{ij}$  的系数为负， $AD_{ij}$  的系数为正。

#### 收入水平

获得更高的收入，是很多移民的迁移动机。经济学家也非常关注收入对于移民影响的方向和程度 (Sjaastad, 1962)；Greenwood (1975) 综述了收入与移民的关系。本文将采用的收入指标是 1996 年中国统计年鉴中迁出省  $i$  1995 年农村人均收入  $RI_i$ ，迁出省  $i$  和迁入省  $j$  城镇人均收入  $UI_i$  和  $UI_j$ <sup>31</sup>。没有利用迁入省  $j$  的农村人均收入  $RI_j$ ，一方面是因为迁移的目的地绝大部分是在城镇地区 (70% 左右，见表 2)；另一方面，加入  $RI_j$  后可能造成多重共线性问题，而  $UI_j$  的高低在很大程度上也就反映了  $RI_j$  的高低。

分别采用农村和城镇人均收入，而不是全部人口的人均收入，这是因为统计年鉴关于收入的数据都是农村与城镇分开的，要获得全部人口的人均收入，只能根据相应人口比例折算；另外，对于占省际移民绝大多数 (78.75%) 的农村移民，有一个重要的替代选择是迁移到本省的城镇，因此，本省城镇收入应对移民决策有明显的负影响。

<sup>23</sup> 更严格的说，这里应该是 1995 年迁出省的农村 10-59 岁人口数，但是考虑到各省人口年龄分布基本类似，我们以 1995 年迁出省的农村人口数替代。

<sup>24</sup> 同样的，这里也没有做年龄的划分，参阅脚注 23。

<sup>25</sup> 以下讨论中，1997 年之前重庆市的数据根据四川省的数据相应折算，同样，四川省的数据也作相应折算。

<sup>26</sup> 一个全面而且较为技术性的讨论，可以参阅 Schwartz (1973)。

<sup>27</sup> 海南与其他各省的距离为广州到其他各省的距离加上 620 公里。实际上，海南省省会海口到广州的距离是 620 公里，到广西省会南宁的距离是 510 公里，在计算中，我们综合考虑了跨海的因素。

<sup>28</sup> 相关讨论可参阅 Chan, Liu 和 Yang(1999)。

<sup>29</sup> 类似讨论见于 Fields(1977)对美国国内移民的研究。

<sup>30</sup> 依据同样的论点，我们也应该把关于其他各个变量的其他各省平均值 (如其他各省市的平均收入等) 放入回归式中；但实际计算可以发现，其他变量在各省市的差距并不是很大，除去两个省，剩下的 28 个省平均之后，变化很小。因此，本文不考虑其他变量的其他各省平均值。

<sup>31</sup> 同一省内的农村收入和城镇收入存在较强的相关关系， $RI_i$  和  $UI_i$  相关系数为 0.70，但是在我们的研究中，没有造成足够大的多重共线性问题，因此我们保留这两个变量；同时  $RI_j$  和  $UI_j$  相关性也很强，并且在加入该省高校招生人数时，与  $RI_j$  相关也很高 (相关系数 0.65)，这造成  $RI_j$  的 VIF (Variance Inflation Factor) 在各个回归中都是最高的 (在 6.3-9.3 之间)，一般认为 VIF 超过 10 时，多重共线性问题比较严重，但是考虑到  $RI_j$  并不是很重要的变量，同时也为避免可能的多重共线性问题，我们没有在回归中用到  $RI_j$ 。

## 失业率和 GDP 增长率

根据 Todaro (1969) 和 Harris 和 Todaro (1970) 的研究, 移民对于迁入地的选择考虑的是期望工资, 因此不仅要考虑工资水平, 还要考虑失业率 (也就是就业机会的高低)。在中国考虑失业率的问题, 还有更深层次的意义。正如之前已经讨论的户籍制度中, 其中一方面就是对于非户口居民的管制和歧视, 在各地政府出台这些管制和歧视措施时, 一个堂而皇之的名义就是保护下岗工人的就业机会;<sup>32</sup> 另外, 很多地方用来补助下岗职工的大部分资金来自于非户口居民 (暂住人口) 缴纳的暂住费等管理费用。<sup>33</sup>

但是, 对于如何衡量失业率, 存在着较大的问题。首先, 用哪个时期的失业率来衡量? 如用 1995 年以后的数据, 会产生联立偏差 (simultaneous bias), 因为失业率与迁移率是同时决定的。但如采用 1995 年或这之前的失业率, 那么就无法反映下岗工人大量出现的现象——这是在 1995 年开始在全国范围出现的。<sup>34</sup> 其次, 对于中国大陆各省市的失业率, 没有较好的统计指标——公开发布的失业率只是城镇登记失业率, 不能真实的反映就业情况。

综合上述利弊, 本文根据 1995 年 1% 人口抽样调查数据计算失业率, 定义总非农劳动力 = (在工作的人数 + 在找工作的人数 - 农业就业人数)<sup>35</sup>, 失业率  $U_i$  和  $U_j$  等于该省在找工作的人数比上总非农劳动力。 $U_i$  和  $U_j$  主要衡量的是非农产业的就业情况, 预期  $U_i$  对移民率有正的影响,  $U_j$  对移民率有负的影响。

同样根据预期工资的理论, 本文将 GDP 增长率也考虑在内。由于迁移到目的地后, 移民一般会居住一个较长的时期, 因此对于该地的发展前景的预期会有很重要的作用。GDP 增长率的增长率一方面可以衡量收入水平的增长, 另一方面也可以代表就业机会的增长。考虑到联立偏差, 这里的增长率用 1990-1995 年平均年增长率  $G_i$  和  $G_j$  衡量。

## 人口因素

迁出地和迁入地的人口数量, 对于迁移的影响, 已有很多讨论, 如: Schultz (1982) 基于引力模型的讨论, 他认为人口数量可以影响迁移, 但是它反映的不仅仅是人口数量大小的影响。Chan, Liu 和 Yang (1999) 对此也作了讨论, 在他们的研究中, 发现迁出地的人口数量对于迁移率有负的影响, 但是他们没有把迁入地的人口数量放入模型中分析。

人口众多可能反映很多问题, 如资源紧张、生活拥挤等, 同时也意味着有较多的经济机会, 因为人口众多意味着市场较大。因此, 人口数量是个重要的变量, 虽然它所体现的并非“纯粹的”人口数量问题; 基于此本文仍将考虑人口数量对于迁移的影响。

Fields (1982) 没有包括这个变量, 理由是 (美国) 人口数量受到前期移民的很大影响, 而前期的移民由前期的经济状况决定, 前期的经济状况与当期的经济状况密切相关。但这在中国并不是问题, 正如 Chan, Liu 和 Yang (1999) 所论述的, 中国各省人口受到移民的影响很小, 因为人口基数实在太大了。

但是另有一个因素需要考虑在内, 中国各省市的面积差异巨大——新疆 166 万平方公里, 而上海仅 0.63 万平方公里。而真正人口数量产生的影响是人们的集中程度。因此要衡量的是人口密度。具体的, 人口密度  $P_i$  和  $P_j$  分别是迁出省  $i$  和迁入省  $j$  1995 年的人口密度。

## 乡村人口人均耕地面积

移民中绝大部分是农村移民; 农民进城打工, 一方面是农业回报太低, 另一方面也是人均耕地太少, 造成隐性失业严重。另外, 中国现行的 (1995-2000 年间施行的) 土地制度下, 没有当地农村户口就没有权利参加土地承包 (见表 1)。基于这两个原因, 可以预期乡村人口人均耕地面积越大, 移民率越低; 户口移民也会减少。同时, 迁入地的乡村人口人均

<sup>32</sup> 一个例子可以参阅蔡昉等 (2001) 对于北京市政府不同时期对于外地劳动力的就业歧视政策的总结。

<sup>33</sup> 北京大学中国经济研究中心课题组, 1998, 1999; 赵耀辉, 徐建国, 2001。

<sup>34</sup> 北京大学中国经济研究中心课题组, 1998, 1999。

<sup>35</sup> 由于农业就业人口占绝大多数 (60% 左右), 而农业的隐性失业无法衡量, 将农业就业人口计入总劳动力后, 得到的就业率会偏低, 并且各省市的差距不能很好的反映。

耕地面积，还在一定程度上反映了迁入地人口压力和农村过剩劳动力水平。

本文所用的各省市耕地数据是 1996 年农业普查得到的耕地面积，<sup>36</sup>乡村人口是 1995 年各省市乡村人口数；前者除以后者，得到乡村人口人均耕地  $AL_i$  和  $AL_j$ 。乡村人口人均耕地和人口密度存在一定的相关性<sup>37</sup>，但是，两者衡量的并非同一回事。而且，在控制了乡村人口人均耕地之后，人口密度越高，可以认为该地区城镇（非农行业）人口越多。<sup>38</sup>

### 移民存量

移民存量对于迁移的作用，主要是体现在移民网络的作用上；这一作用在其他国家已经被证实对移民有显著的影响（Greenwood, 1969），也有研究表明了在中国有很重要的影响（如 Rozelle 等，1999）；Chan, Liu 和 Yang (1999)也发现了移民网络的重要作用，但他们所考虑的移民网络以省到省的前一期移民存量来代表，这是一种较粗略的方法，因为这种移民存量所反映的不是单纯的移民网络问题；同时，这样处理可能存在内生性问题，因为前一期的移民数量受前一期经济状况的影响，而前一期经济状况与当期的经济状况密切相关（Hall, 1970; Fields, 1979）；Zhao (2001)从更加微观的角度考察了移民网络对于迁移的影响——移民网络以该村的早期迁出移民数量代表，迁出地的缩小可以进一步的“提纯”移民网络的作用。

限于数据的可得性，同时由于研究对象是跨省移民，本文也用省际移民存量来衡量移民网络问题。考虑到移民存量的内生性，我们不采用 1990-1995 年的迁移量，而采用 1985-1990 年间从  $i$  省迁入  $j$  省的移民率  $MS_{ij}$ （从  $i$  省迁入  $j$  省的移民数量除以  $i$  省 1985 年人口数），这样可以部分避免内生性问题。<sup>39</sup>这一指标由 1990 年人口普查 1% 抽样数据计算而来。

另外，我们将对移民进行细分为户口移民和非户口移民、农村移民等；由于户口制度造成的城乡分割，农村移民和非户口移民的迁移行为一般都是非正式的渠道（“务工经商”等），而城镇移民和户口移民一般是通过正式（官方）的渠道（如“分配录用”、“工作调动”和“学习培训”等）。从现实看，农民工迁移的信息多来自于早期的农民工，而与其他迁到同一地方的户口移民和城镇移民没有什么联系。<sup>40</sup>因此，在分析各类不同移民时，移民存量也调整为相应的类别（从  $i$  省迁入  $j$  省的该类移民数量除以  $i$  省 1990 年该类人口数）。<sup>41</sup>可以预期，移民存量对于迁移率有正的作用；相对于户口移民，对非户口移民的影响会更显著。

### 高校招生人数

从上文的描述中可看出，大学生（包括新生和毕业生）是跨省移民的重要一部分，特别是 1999 年后，开始了高校扩招。对于农村移民来说，要想进行户口迁移，考上大学是最主要的途径。因此，我们需要考虑高校招生对于迁移的影响。

迁出省  $i$  的高校招生人数比率  $CS_i$  对于本省迁出移民的影响可以体现在两个方面：（1）本省招生比率高，该省的学生上大学的机会高<sup>42</sup>，出去打工的人可能会少一些，但是不会影响出省上大学的人数，因为外省在本省的招生指标是相对稳定的；（2）招生比例高，毕业生人数也相应的高，而毕业生工作时可能迁往全国各省（特别是外省生源的毕业生），因此，迁出地招生比例高会提高省际迁出率，特别是户口移民的迁出率（因为大学毕业生的“分配

<sup>36</sup> 衡量的是 1996 年 10 月 31 日时点耕地面积，见中国统计年鉴 2000 年。

<sup>37</sup> 两者相关系数-0.60，但是从其他检验多重共线性的统计量（如 VIF）来看，问题不大。

<sup>38</sup> 当然这里最好有一个衡量城镇人口密度的指标，但目前只能找到一个建成区面积；而在建成区的定义上各地标准不尽相同，也不能够充分反映城镇地区的面积；因此，在这里没有采用这一指标。

<sup>39</sup> 采用 1985-1990 年移民存量的另一个理由是，我们可以根据 1990 年人口普查数据得出分类的移民存量；而对于 1990-1995 年迁移存量，由于缺少原始数据，不能够细分各类移民存量，并且正如上文讨论的，后者更有可能遗漏移民。

<sup>40</sup> 比如，可以认为，某省农村到北京来上大学的人与同一乡来北京打工的民工之间，基本没有直接联系。

<sup>41</sup> 这里移民存量的分类没有做年龄上的进一步限定，只定义 1990 年移民为 5 周岁以上人口（2000 年已经 15 岁）；因此，对应于 1995-2000 年的各类移民，移民存量的类别只有：省际户口移民和非户口移民、省际农村户口移民和农村非户口移民，省际农村到城镇户口移民和农村到城镇非户口移民，省际城镇户口移民和非户口移民。在计算移民存量时，分母上的该类人口总数即上述几类人口的总数。

<sup>42</sup> 从目前情况来看，各高校主要在所在地省市招生。

录用”一般很容易获得当地户口)。综合两方面的因素,预期本省高校招生比率  $CS_i$  对于非户口移民迁出率的影响是负的,但影响不会很大,因为上大学的人数占迁移人口很小一部分;预期本省高校招生比率  $CS_i$  对于户口移民迁出率的影响为正。迁入省  $j$  的高校招生人数比率  $CS_j$  可以代表迁出省  $i$  的人到  $j$  省上大学的机会,  $CS_j$  越高,机会越多,迁移率——特别是户口迁移率也就越高。

另一方面,高校招生比例也体现了该省高等教育事业的发达程度,这也一定程度上表现了该省文化水平的高低。因此在解释  $CS_i$  和  $CS_j$  的作用时,需要多方面的考虑。<sup>43</sup>

本文采用的高校招生人数是 2000 年各省市高校的招生人数占 2000 年人口总数的百分比。1999 年,中国的高校普遍进行了扩招,但这种扩招基本也是在原有招生规模上成比例的扩大,因此 2000 年的数据仍然可以反映各省招生人数的相对大小,并且是外生变量。

### 其他变量

其他一些研究中国迁移问题的文献中,经常把乡镇企业就业变量、直接外国投资(FDI)以及农业占经济的比重考虑在模型中(如 Liang 和 White, 1997; Chan, Liu 和 Yang, 1999)。他们主要考察的是乡镇企业对于吸收农村过剩劳动力的作用,而 FDI 也是创造就业就会的重要方式。但归根结底,移民关心的是就业机会和收入水平,而不是乡镇企业或者 FDI 的多少。另外,农业占经济的比重衡量的是一个地区工业化的程度,在控制了收入水平、就业机会等变量后,这也不会对迁移决策起显著的作用。基于以上原因,本文不采用这些变量。

应变量和所有自变量的定义、记号,以及数据来源,在表 4 做了一个总结。

### (四) 模型估计与结果

对于各类移民,我们回归(7)式以估计各个系数。由于是加总数据,我们构造模型的一个内涵假设是每个人(移民)都是同质的,因此每个省到省的迁移流也是同质的,这样(7)中包含的扰动项  $e_{ij}$  符合经典线性回归的假设:  $E(e_{ij} | X) = 0$  ( $X$  是模型中所包含的自变量),方差协方差矩阵  $Var(e | X) = \mathbf{s}^2 I$ 。但是,事实上,每个人是异质的,除了模型中所包含的自变量外,迁移的决策受到很多其他因素的影响,而加总数据中,扰动项  $e_{ij}$  可以看

作是  $e_{ij} = \sum_{n_{ij}=1}^{N_{ij}} e_{n_{ij}} / N_{ij}$  ——是  $N_{ij}$  个迁移者扰动项的均值。因此,直接回归(7)存在着异方

差的问题。对此用 Goldfeld-Quandt 检验<sup>44</sup>,结果显示迁移人数越少的方差越大,拒绝了同方差假设。根据这一检验结果,我们对回归进行了加权,权重为该迁移方向的人数。<sup>45</sup>对加权

<sup>43</sup>同时,由于  $CS_i$  和  $CS_j$  与当地的收入水平存在正相关的关系,正如上文所讨论的,舍弃迁入地农村人均收入  $RI_j$  后,可以避免多重共线性问题。

<sup>44</sup> 参阅 Goldfeld 和 Quandt (1965), 或者 Greene(2003)的简要介绍。检验的方法是:按照移民数量先对 870 个样本点由少到多排序,分别回归数量最少的  $n$  个和数量最多  $n$  个样本点 ( $n < 435$ ), 分别得到残差平方和  $e_1'e_1$  和  $e_2'e_2$  (在本研究中  $e_1'e_1 > e_2'e_2$ )。构造  $F[n-K, n-K] = [e_1'e_1 / (n-K)] / [e_2'e_2 / (n-K)]$ , 在同方差的假设下(即  $H_0: F[n-K, n-K]=1$ ), 统计量  $F[n-K, n-K]$  服从自由度为  $[n-K, n-K]$  的 F 分布。我们的检验中,取  $n=335$ , 各个回归中  $e_1'e_1$  基本都是  $e_2'e_2$  的 1000 倍左右;因此,拒绝同方差的假设。

<sup>45</sup> Chan, Liu 和 Yang (1999)的研究中也用了加权回归,但他们并没有从异方差的角度来说明;他们用的权重是该迁移方向的移民数量占总移民的比例,这在本质上与本文用的权重(直接用该迁移方向的移民数量)是完全一致的,因为只有相对的比重才起作用。另外,由于有些迁移方向的移民数量为零,这里我们仍然采用前面的方法,在移民数量上加上一个较小量,实际回归时的较小量为 0.0001;跟前面类似,加上的量小于 0.001 后,对结果没有大的影响;但在这里不宜加上太小的量,因为过于小的权重相当于忽略了那些移民数量为零的迁移方向。

后的回归进行 Goldfeld-Quandt 检验，结果显示不存在异方差问题。<sup>46</sup>

表 5 报告了总的省际移民、省际农村移民和省际农村到城镇移民的回归结果，每类移民分别分为户口移民和非户口移民。下文将对各类移民分别进行分析。<sup>47</sup>

### 总的省际移民

总的省际移民分别分为：总省际移民，户口和非户口移民三个小类进行了计量分析。首先可以发现的是，非户口移民模型的拟合优度 (R square)(0.816) 要大大高于户口移民模型的拟合优度 (0.494)，这说明所选的变量可以更好的解释非户口移民。<sup>48</sup>另外，对于各类移民率，距离有负的作用，到其他各省的平均距离有正的作用，移民存量（移民网络）有正的作用；这些对于三小类移民都是一致的，跟我们理论的预期也一致。

三小类移民对于人口密度的反应方向也是相同的，迁出地和迁入地人口密度的增加都倾向于减少迁移率。由于已经控制了乡村人口人均耕地因素，所以，可以认为人口密度的增加意味着城镇地区（非农行业）的人口增加；这一方面是如引力模型所说的，人口多意味着经济机会也多，另一方面也会造成资源的紧张。但移民对于迁出地和迁入地资源和机会可利用的程度不同——这正是变量作用不对称的根源。在迁出地（家乡），移民一般都会有人际关系等一些无形“资产”，而这些她/他在迁入地一般都不具备；因此，相比较而言，在迁出地（家乡）人们一般能够更有效的利用各种机会——包括人口众多带来的机会，而资源紧张就不会造成太大的困难；在迁入地，人生地不熟，人口众多带来的机会不能像在家乡那样有效利用，人口密度大更多的意味着资源的紧张。因此，迁出地和迁入地人口密度的增加都会减少移民率。另外，从系数的大小来看，非户口移民对于人口密度的反应要强于户口移民的反应（两个系数在一定程度上都是非户口移民模型中大一些），而对于每类移民，对于迁入地人口密度的反应要强于迁出地人口密度的反应。

乡村人口人均耕地面积的系数在三个回归中存在很大的差异：迁出地  $AL_i$  对于户口移民没有影响，但是对于非户口移民有负的影响；而迁入地  $AL_j$  对户口移民有负的影响，对非户口移民没有影响；由于省际移民中主要是非户口移民（占 84.82%），因此，乡村人口人均耕地面积对于总移民影响与其对非户口移民的影响是一致的。对此的一个解释是：户口移民主要是“学习培训”和“婚姻迁入”（51.72%），这些移民基本不会考虑迁出地耕地面积的多少（如：“学习培训”——上大学，迁户口，跳出“农门”，成为一个城市人，这是农家子女求之不得的事，放弃一点耕地是非常愿意的）。相反的，对于非户口移民，迁出地  $AL_i$  较大意味着他们在家乡的隐性失业情况不是很严重，因此迁移出去的人也就少一些；由于没有得到迁入地的户口，因此也不能正常承包迁入地的耕地，迁入地  $AL_j$  对他们没有影响；同时，迁入地  $AL_j$  对非户口移民的影响不显著，这反映出迁入地农业人口隐性失业对于移民的影响不明显。但是迁入地  $AL_j$  对户口移民的副作用，目前还没有很好的解释，当然也可以看作户口制度对于迁移的扭曲作用。

人均收入的各个变量，对户口移民和非户口移民有截然相反的影响：对于户口移民，只有迁出地城镇人均收入  $UI_i$  有影响，但是符号是正的，与预期的相反；其他的收入变量没

<sup>46</sup> 本文在这里仅考察了异方差的问题，没有对自相关（autocorrelation）问题进行详细讨论。实际上，自相关也是有可能存在的，比如说迁移至同一地区的人，会受到一些共同的、但没有在自变量中反映的因素影响；另一方面，同一地区迁移出来的人，又有可能受到这类因素的影响；这样，给定  $i$ ， $e_{in}$ ， $\forall n$  都是相关的；给定  $j$ ， $e_{mj}$ ， $\forall m$  也是相关的；那样就要用广义最小二乘法（GLS）来估计，但之前又要估计  $e_{ij}$  的方差协方差矩阵——这会是一个很复杂，而且结果并不一定很好的方法，因为估计  $e_{ij}$  的方差协方差矩阵本身就是一件比较麻烦的事情。不考虑自相关问题，直接用最小二乘法时，估计依然是无偏的，但是不一定是有效的（Greene, 2003）。基于以上讨论，我们不考虑自相关问题。

<sup>47</sup> 初稿中包括了城镇移民的计量分析结果，限于篇幅，为在此报告。

<sup>48</sup> 在更细的分类移民分析中，这样的差异也是存在的。



有显著的作用。如果认为对于更高收入的追求是自发迁移的标志，那么可以认为，户口迁移中自发的因素不是很大，甚至被抹煞了。对于非户口移民（总省际移民与之类似），与预期一致，迁出地农村人均收入  $RI_i$  和城镇人均收入  $UI_i$  对迁移率有负的影响，并且从系数大小看，后者对与移民的影响较大，这说明农村移民的首选还是迁移到本省的城镇；迁入地城镇人均收入  $UI_j$  对于迁移率有正的影响，并且系数较大（4.833）——对更高收入的追求是非户口迁移的主要原因之一。

对于 GDP 增长率的反应，各类移民也存在差别。迁入地 GDP 增长率  $G_j$  对于各类移民的影响都是正的，但是影响的程度不一致——对于非户口移民的影响（系数 1.845）要大于对户口移民的影响（系数 0.925）——这也反映出非户口移民对于经济因素的更为敏感。但是对于迁出地的 GDP 增长率  $G_i$ ，户口移民的反应为负（与预期一致），但非户口移民的反应为正。对于后者，可能的解释是：一方面，GDP 增长率  $G_i$  受到 GDP 基数的很大影响，一些 GDP 基数很低的省增长率较高；另一方面，有些地区，如浙江等，通过对外的迁移，一定程度上也促进了 GDP 的增长。其中的机制，还有待于进一步的分析。

失业率对于户口移民和非户口移民的影响，正如预期的一样——迁出地失业率  $U_i$  对迁移率有正的影响，迁入地失业率  $U_j$  对迁移率有负的影响。但是，将这两类移民加总后，迁入地失业率  $U_j$  对迁移率影响不显著，这也再一次提醒我们：不区分户口移民和非户口移民，可能会造成估计的偏差。

最后，高校招生人数对于户口移民有正的影响，而对非户口移民没有影响。这与我们的预期基本一致，因为本省招生人数增加，大学毕业生也增加，这样毕业分配时，迁移到外地的（特别是回到生源地的）户口迁移人数也会增加，因此本省高校招生比率  $CS_i$  对迁移率有正的影响；同时，迁入省的高校招生比率  $CS_j$  增加意味着在迁出省招生人数上升，户口移民也将增加。对于总移民和非户口移民，这两个指标没有影响，这一方面是因为跨省大学新生人数不多<sup>49</sup>，另一方面可以看出，在控制了其他经济因素后，高校招生比率反应的文化因素对于移民并没有显著影响。

### 结构性差异检验

从第二部分的描述我们已经看出户口移民和非户口移民在许多特征存在巨大差异，而上述两类移民的回归结果也有明显不同，在这里我们将进一步地进行计量检验：户口移民和非户口移民是否存在结构性的差异？

参考结构性变化检验的思路，<sup>50</sup>我们可以检验省际户口移民和省际非户口移民是否存在结构性的差异——即是否两个回归中每个自变量的系数都是不同的。得到统计量  $F[17,1706]=73.45^{51}$ ，P 值小于 0.00001，因此，拒绝零假设，结论：省际户口移民和省际非户口移民存在结构性区别。同样也可以把省际农村劳动力移民、省际农村到城镇劳动力移民分为户口移民和非户口移民，也进行结构性检验，结果都显示两类移民存在结构性的不同。

### 农村劳动力移民和农村到城镇劳动力移民

<sup>49</sup> 《中国统计年鉴 2001》显示，2000 年全国在校本专科大学生人数为 5,560,900 人，但估计其中只有不超过 20% 左右为外省生源，因为各地招生主要集中在本省。

<sup>50</sup> 结构性检验原用于检验一定时点后经济结构是否发生根本变化；其目的等价于检验两个子样本划分的时间是否与经济结构无关。本文检验的是以户口状况划分的两类移民是否存在结构性差别；若户口状况对于迁移无关（如，按移民生日的单双号划分），两类移民应无结构性区别；若户口状况对于迁移有影响（如，教育程度对迁移的影响），那么两类移民就会有结构性差别。不可否认，这样的检验是粗略的，因为户口状况本身是迁移决策的一部分。

<sup>51</sup> 具体的检验方法是：首先分别对省际户口移民和非户口移民回归，得到残差平方和  $e_1'e_1$  和  $e_2'e_2$ ；然后对受限制（即两类移民回归系数相同）的模型进行回归，得到残差平方和  $e_*'e_*$ ，统计量的计算公式是：

$F[17,1706] = [(e_*'e_* - e_1'e_1 - e_2'e_2) / 17] / [(e_1'e_1 + e_2'e_2) / (870 \times 2 - 17 \times 2)]$ ，在零假设（没有结构性差别）下，该统计量服从自由度为 [17,1706] 的 F 分布。

在农村到城镇劳动力移民的回归分析中包括迁入地  $AL_j$ ，主要是考虑到：(1) 迁入地农业隐性失业对于迁移的影响；(2) 控制人口密度后， $AL_j$  增加等价于城镇人口增加。从结果来看， $AL_j$  增加倾向于减少移民，可能的原因是移民回避人口过多的城镇。

从表 5 同样可看出户口移民与非户口移民之间的巨大差别，而且这几类移民与总的户口移民和非户口移民的特征基本一致；当然，我们也可以发现一些这几类移民的独特之处：

(1) 移民存量（移民网络）对于农村到城镇劳动力户口移民没有影响。正如上文所提到的，农村到城镇的户口移民主要是上大学；我们用的移民存量是 1985-1990 年间的该类移民，当时的这些人主要是早期知青等计划性迁移的回流(Chan, Liu 和 Yang 1999)；因此这一结果情理之中的。

(2) 迁出省的收入（包括农村人均收入和城镇人均收入）对这两大类移民中的户口移民没有影响。这一结果验证了我们在上文的讨论——农村（到城镇）劳动力户口移民主要是“学习培训”的大学生，在控制了高校招生人数后，收入不应起作用。因为同样的原因，农村到城镇劳动力户口移民对于迁入省的 GDP 增长率  $G_j$  和失业率  $U_j$  不敏感；迁出地的失业率  $U_i$  对于两类户口移民也没有影响。

(3) 农村到城镇劳动力户口和非户口移民不受到迁入地 GDP 增长率  $G_j$  的影响。考虑到农村和城镇之间收入的巨大差异（2000 年城镇和农村人均收入之比为 2.8，参阅施新政，2003），增长率没有影响亦属正常。另外，对于增长率的不敏感在一定程度上反映了非户口移民迁移行为不是一种很长期的行为，因为在长期行为中，增长率的预期会有显著的作用。

(4) 迁出省高校招生比率  $CS_i$  对农村到城镇劳动力非户口移民有正的影响——从直觉来看，本省高校的招生人数对于本省农村劳动力迁移到外省打工并没有直接的联系，结果如此的可能原因是这一指标反映了其他的因素；比如，农村上大学的新生增加，导致农村在校人数（包括中学和大学）增加，学费上涨较快，迫使家人出门打工；另外，最近几年不要户口的大学毕业生增加<sup>52</sup>，可能也是原因之一。

## 四、结论和进一步的研究

本文初步研究了 1995-2000 年间国内户口迁移和非户口迁移，首先具体描述了 1995-2000 年省际移民的特征，特别比较了户口移民和非户口移民的异同点，发现他们在多个方面存在的根本差异，如，个人特征、迁移原因、行业和职业分布、迁移方向等。另外，针对农村劳动力移民、农村到城镇劳动力移民和城镇劳动力移民作了专门的分析。在此基础上，我们利用各地间迁移模型对户口移民和非户口移民分别作了计量分析。主要的发现是：一些经济因素可以较好的解释非户口迁移，而对户口移民的作用很多不像预期的那样；反映出户口制度对于户口移民还有很大程度的决定作用。

本文分析的是加总数据，而迁移决策还取决于很多的个人特征（Greenwood, 1975）；特别是户口迁移与非户口迁移，我们可以看到，他们在个人特征上就存在着很大的差异。在目前的户口制度之下，迁移决策是一个两方决定的结果：个人决定是否迁移，是否想要迁移目的地的户口；而政府决定是否发放户口。这样，就产生了三种人——非移民、户口移民和非户口移民。而在本文的分析中，我们没有涉及到户口移民和非户口移民之间的联系，而实际上，两者之间是存在着联系的——可能每个人都想要拿到户口，但是很多人没有成功，所以只好成为非户口移民。

另外，如何衡量户口制度，户口制度是否真正起到了减少移民进入的作用？对于这两个大家关注的问题，本文没有解决；希望进一步的研究能给我们提供解答。

<sup>52</sup> 表 2 显示，省际非户口移民中有 4.04% 的人教育程度为大专或以上，折合成全国人口数，约 1,151,613 人。

## 参考文献

- [1] 北京大学中国经济研究中心课题组：“上海：城市职工与农村民工的分层与融合”，《改革》，1998年第4期。
- [2] 北京大学中国经济研究中心课题组：“南京：福利惯性下的劳动力市场”，《改革》，1999年第4期。
- [3] 蔡昉，都阳，王美艳，户籍制度与劳动力市场保护，经济研究，2001（12）。
- [4] Chan Kam Wing, Ta Liu and Yunyan Yang, 1999. "Hukou and Non-Hukou Migration: Comparisons and Contrasts", *International Journal of Population Geography*, 5: pp.425-448.
- [5] Chan, Kam Wing and Li Zhang (1999), "The Hukou System and Rural-Urban Migration in China: Processes and Changes." *The China Quarterly* 160: 818-55.
- [6] 陈峰，王雷，“被收容者孙志刚之死”，南方都市报，2003年4月25日。
- [7] Cheng, Tiejun and Mark Selden, "China's Hukou System." *China Quarterly*, 1994, Pp. 644-668.
- [8] Fields, Gary. S., "Place-to-place migration: some new evidence", *The Review of Economics and Statistics*, Volume 61, Issue 1, 1979 (Feb.), 21-32.
- [9] Fields, Gary. S., "Place-to-place migration in Colombia." *Economic Development and Cultural Change*, 1982, 30(3): 539-558.
- [10] Greenwood, M. J., "An Analysis of the Determinants of Geographic Labor Mobility in the United States." *Review of Economics and Statistics* 1969, 51:189-194.
- [11] Greenwood, M. J., "Research on internal migration in the United States." *Journal of Economic Literature* 1975, 13: 397-433.
- [12] Hall, Robert E., "Why if the Unemployment Rate So High at Full Employment?" *Brookings Papers on Economic Activity* 1970, 3: 369-420.
- [13] Harris, J.R. and M.P. Todaro, "Migration, Unemployment and Development: A Two Sector Analysis", *American Economic Review*, 1970, Vol. 60, pp. 126-142.
- [14] Liang Zai and Michael White, "Market Transition, Government Policies, and Interprovincial Migration in China: 1983-1988." *Economic Development and Cultural Change*. 1997, 45(2): 321-39.
- [15] 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。
- [16] 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.
- [17] Mallee, H., "In Defense of Migration: Recent Chinese Studies on Rural Population Mobility." *China Information*, 1996, 10(3/4): 108-140.
- [18] 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，1997。
- [19] Rozelle, Scott, Li Guo, Minggao Shen, Amelia Hughart, and John Giles "Leaving China's Farms: Survey Results of New Paths and Remaining Hurdles to Rural Migration," *The China Quarterly*, June 1999.
- [20] 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
- [21] Schwartz, Aba, 1973. "Interpreting the Effect of Distance on Migration," *Journal of Political Economy*, Vol. 81 (5): 1153-69.
- [22] 施新政，“中国城乡收入差距分析”，北京大学中国经济研究中心硕士论文，2003。
- [23] 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>）
- [24] Todaro, Michael P., "A Model of Labor Migration and Urban Unemployment in Less Developed Countries", *The American Economic Review*, Volume 59, 1969, Issue 1: 138-148.
- [25] Wang, Gewei, "Income Inequality and Migration in China", *Unpublished*, CCER, Peking University, 2003.
- [26] 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，1996。
- [27] Zhao, Yaohui, "The Role of Migrant Networks in Labor Migration: The Case of China", paper presented at the International Conference on Urbanization in China, Xiamen, China, June 2001.
- [28] 赵耀辉，徐建国，“我国城镇养老保险体制改革中的激励机制问题”，*经济学季刊*，2001年第1卷第1期。
- [29] 赵耀辉等，“中国人口迁移与收入差距——一个基于五普资料的研究”，未发表文稿，2003。
- [30] 中国国家统计局，《中国统计年鉴》北京：中国统计出版社，相应年份。

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

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

表 2、1995-2000 年间各类移民与全国人口个人特征对比

特征	5 岁以上 全国人口	省内省际移民			省际移民	
		总计	户口	非户口	户口	非户口
<u>性别</u>						
男	51.2	49.79	45.68	51.48	52.91	52.83
女	48.8	50.21	54.32	48.68	47.09	47.17
<u>年龄(年)</u>						
5--14	19.00	7.46	8.19	7.15	4.64	6.58
15--19	8.43	19.30	24.01	17.34	15.46	17.40
20--29	16.91	40.01	42.23	39.09	46.86	51.66
30--39	19.90	18.77	13.14	21.11	22.97	14.56
40--49	14.61	7.31	6.46	7.66	6.14	5.13
50--59	9.72	3.85	3.16	4.14	2.37	2.36
60--64	3.67	1.26	1.00	1.36	0.68	0.94
>=65	7.77	2.04	1.81	2.13	0.88	1.37
<u>婚姻状况 (&gt;=15 岁)</u>						
未婚	20.16	43.12	48.44	40.93	49.48	44.24
已婚	79.84	56.88	51.56	59.07	50.52	55.76
男未婚	23.59	45.76	58.87	41.06	64.43	42.96
男已婚	76.41	54.24	41.13	58.94	35.57	57.04
女未婚	16.65	40.55	40.00	40.80	32.79	45.67
女已婚	83.35	59.45	60.00	59.20	67.21	54.33
<u>现住地类型</u>						
市	23.39	58.99	59.5	58.78	53.36	53.03
镇	13.23	21.69	19.47	22.61	9.88	21.66
乡	63.38	19.32	21.03	18.62	36.76	25.32
<u>迁出地类型</u>						
城市街道		19.06	28.14	15.33	37.95	9.51
镇居委会		11.3	16.51	9.16	13.62	6.27
村		69.64	55.36	75.51	48.43	84.22
<u>受教育程度 (&gt;=6 岁)</u>						
扫盲班及以下	10.24	3.47	2.3	3.96	2.56	3.16
小学	38.24	19.82	14.54	22.01	17.66	23.04
初中	35.83	40.43	24.76	46.95	26.99	56.07
高中	8.29	13.02	10.85	13.92	7.72	9.98
中专	3.46	10.63	19.8	6.82	8.44	3.71
大学专科	2.6	6.77	12.93	4.2	7.9	2.58
大学本科及以上	1.34	5.87	14.81	2.14	28.72	1.46
<u>劳动力比例 (&gt;=15 岁)</u>						
	76.90	71.23	53.59	78.47	55.96	89.61
样本中人口数	1,112,963	88,697	26,020	62,677	4,847	27,080

数据来源：2000 年人口普查 0.95% 抽样数据。

表 3 : 1995-2000 年各类移民迁移原因对比

迁移原因	省际移民		省际农村 劳动力移民		省际农村到城镇 劳动力移民		省际城镇 劳动力移民	
	户口	非户口	户口	非户口	户口	非户口	户口	非户口
务工经商	9.2	77.03	7.78	85.68	5.12	87.48	4.48	50.67
工作调动	8.38	1.72	3.09	0.71	4.93	0.75	14.83	9.96
分配录用	7.46	0.54	2.95	0.17	5.31	0.19	15.72	3.66
学习培训	31.2	2.04	29.3	0.87	58.39	1.13	47.57	10.47
拆迁搬家	2.19	0.54	2.2	0.29	1.33	0.19	2.07	1.44
婚姻迁入	20.52	2.94	39.94	2.89	12.42	1.28	4.42	4.34
随迁家属	9.2	7.23	5.3	3.86	6.73	3.97	5.48	6.12
投亲靠友	5.56	4.39	5.3	2.56	4.27	2.32	2.8	8.04
其它	6.27	3.57	4.13	2.98	1.52	2.69	2.63	5.3
样本中人数	4,847	27,080	2,133	21,700	1,055	15,924	1,787	2,923

数据来源：2000 年人口普查 0.95% 抽样数据。

表 4、变量定义和数据来源\*

变量名称和记号	变量定义	数据来源
迁移率 $m_{ij}$	1995-2000 年间从 $i$ 省迁入 $j$ 省的各类移民率 (从 $i$ 省迁入 $j$ 省的该类移民数量除以 $i$ 省 1995 年相应人口数)	2000 年人口普查 0.95% 抽样数据
距离 $D_{ij}$	迁出省 $i$ 省会到迁入省 $j$ 省会的铁路距离。	中国地图册铁路里程表
到它省平均距离 $AD_{ij}$	$AD_{ij} = \sum_{n \neq j} D_{in} / 28$ ;从 $i$ 省省会到除 $j$ 省之外其他各省省会的平均距离。	中国地图册铁路里程表
移民存量 $MS_{ij}$	1985-1990 年间从 $i$ 省迁入 $j$ 省的各类移民率 (从 $i$ 省迁入 $j$ 省的该类移民数量除以 $i$ 省 1985 年相应人口数)	1990 年人口普查 1% 抽样数据
人口密度 $P_i$	迁出省 $i$ 1995 年的人口密度 (该省人数除以该省面积)	中国统计年鉴 1996 年
人口密度 $P_j$	迁入省 $j$ 1995 年的人口密度 (该省人数除以该省面积)	中国统计年鉴 1996 年
人均耕地 $AL_i$	迁出省 $i$ 乡村人口人均耕地 (耕地面积/乡村人口数)	中国统计年鉴 1996 年, 2001 年
人均耕地 $AL_j$	迁入省 $j$ 乡村人口人均耕地 (耕地面积/乡村人口数)	中国统计年鉴 1996 年, 2001 年
农村人均收入 $RI_i$	迁出省 $i$ 1995 年农村人均收入	中国统计年鉴 1996 年
城镇人均收入 $UI_i$	迁出省 $i$ 1995 年城镇人均收入	中国统计年鉴 1996 年
城镇人均收入 $UI_j$	迁入省 $j$ 1995 年城镇人均收入	1996 年中国统计年鉴
GDP 增长率 $G_i$	迁出省 $i$ 1990-1995 年 GDP 平均年增长率	中国统计年鉴 1991 年, 1996 年
GDP 增长率 $G_j$	迁入省 $j$ 1990-1995 年 GDP 平均年增长率	中国统计年鉴 1991 年, 1996 年
失业率 $U_i$	迁出省 $i$ 1995 年非农行业失业率: 在找工作的人/(在找工作的人数+在找工作的人数-农业就业人数)	1995 年 1% 人口抽样调查数据
失业率 $U_j$	迁入省 $j$ 1995 年非农行业失业率: 在找工作的人/(在找工作的人数+在找工作的人数-农业就业人数)	1995 年 1% 人口抽样调查数据
高校招生比率 $CS_i$	1995 年 $i$ 省高校的招生人数占 1995 年该省人口总数的百分比	中国统计年鉴 1996 年
高校招生比率 $CS_j$	1995 年 $j$ 省高校的招生人数占 1995 年该省人口总数的百分比	中国统计年鉴 1996 年

\*1997 年前重庆市的数据根据四川省的数据相应折算, 同样, 分析中四川省的数据也作了相应的折算。

表 5、省际分类移民率回归结果：双对数形式（样本点=870）

自变量	总的省际移民						省际农村劳动力移民				省际农村到城镇劳动力移民			
	总省际移民		户口移民		非户口移民		户口移民		非户口移民		户口移民		非户口移民	
	系数	标准差	系数	标准差	系数	标准差	系数	标准差	系数	标准差	系数	标准差	系数	标准差
距离 $D_{ij}$	-0.832***	(0.049)	-0.736***	(0.042)	-0.968***	(0.049)	-0.643***	(0.043)	-0.948***	(0.050)	-0.707***	(0.045)	-0.861***	(0.049)
到它省平均距离 $AD_{ij}$	0.585**	(0.229)	1.303***	(0.201)	0.855***	(0.240)	1.756***	(0.204)	0.903***	(0.250)	1.937***	(0.216)	1.536***	(0.226)
移民存量 $MS_{ij}$	0.198***	(0.015)	0.073***	(0.009)	0.099***	(0.009)	0.026***	(0.004)	0.091***	(0.008)	0.002	(0.002)	0.009***	(0.003)
人口密度 $P_i$	-0.110*	(0.064)	-0.123***	(0.046)	-0.159**	(0.073)	-0.016	(0.052)	-0.166**	(0.084)	-0.137***	(0.049)	-0.238***	(0.058)
人口密度 $P_j$	-0.504***	(0.045)	-0.422***	(0.044)	-0.562***	(0.046)	-0.441***	(0.044)	-0.550***	(0.046)	-0.356***	(0.052)	-0.476***	(0.049)
人均耕地 $AL_i$	-0.330***	(0.090)	-0.095	(0.087)	-0.362***	(0.093)	-0.037	(0.092)	-0.315***	(0.097)	-0.212**	(0.093)	-0.370***	(0.090)
人均耕地 $AL_j$	0.024	(0.098)	-0.164**	(0.081)	0.04	(0.107)	-0.103	(0.086)	0.117	(0.108)	-0.311***	(0.084)	-0.195*	(0.100)
农村人均收入 $RI_i$	-0.291*	(0.151)	0.048	(0.146)	-0.540***	(0.157)	0.006	(0.150)	-0.559***	(0.158)	-0.116	(0.152)	0.307**	(0.155)
城镇人均收入 $UI_i$	-1.084***	(0.263)	0.654***	(0.224)	-1.557***	(0.288)	0.345	(0.248)	-1.800***	(0.303)	-0.062	(0.250)	-0.504*	(0.264)
城镇人均收入 $UI_j$	4.295***	(0.153)	0.237	(0.166)	4.833***	(0.164)	-0.015	(0.174)	4.889***	(0.162)	0.588***	(0.179)	4.090***	(0.167)
GDP 增长率 $G_i$	0.105	(0.307)	-0.967***	(0.258)	0.662**	(0.334)	-1.274***	(0.275)	0.789**	(0.350)	-1.262***	(0.281)	-1.086***	(0.328)
GDP 增长率 $G_j$	1.501***	(0.262)	0.925***	(0.242)	1.845***	(0.277)	1.510***	(0.245)	2.075***	(0.276)	0.429	(0.284)	0.282	(0.293)
失业率 $U_i$	0.512***	(0.105)	0.164*	(0.086)	0.565***	(0.115)	0.094	(0.089)	0.585***	(0.117)	0.13	(0.092)	0.682***	(0.108)
失业率 $U_j$	-0.148	(0.102)	-0.383***	(0.087)	-0.198*	(0.110)	-0.344***	(0.089)	-0.169	(0.112)	-0.01	(0.097)	-0.233**	(0.104)
高校招生比率 $CS_i$	-0.133	(0.099)	0.283***	(0.080)	-0.122	(0.108)	0.270***	(0.092)	-0.118	(0.116)	1.117***	(0.087)	1.206***	(0.099)
高校招生比率 $CS_j$	-0.031	(0.071)	0.215***	(0.066)	0.039	(0.074)	0.245***	(0.071)	0.076	(0.075)	0.456***	(0.072)	-0.061	(0.072)
常数项	-28.809***	(2.064)	-12.681***	(1.854)	-31.444***	(2.196)	-14.006***	(1.855)	-31.143***	(2.265)	-7.333***	(2.086)	-29.511***	(2.098)
R-square	0.816		0.494		0.816		0.446		0.816		0.524		0.703	

注：应变量为  $i$  省到  $j$  省该类移民率；所有变量取对数后回归，并且以该类移民的数量作为权重；下标  $i$  表示迁出省特征，下标  $j$  表示迁入省特征。

\*\*\*, \*\*, \* 分别表示在 1%, 5%, 10% 水平下显著不为零。





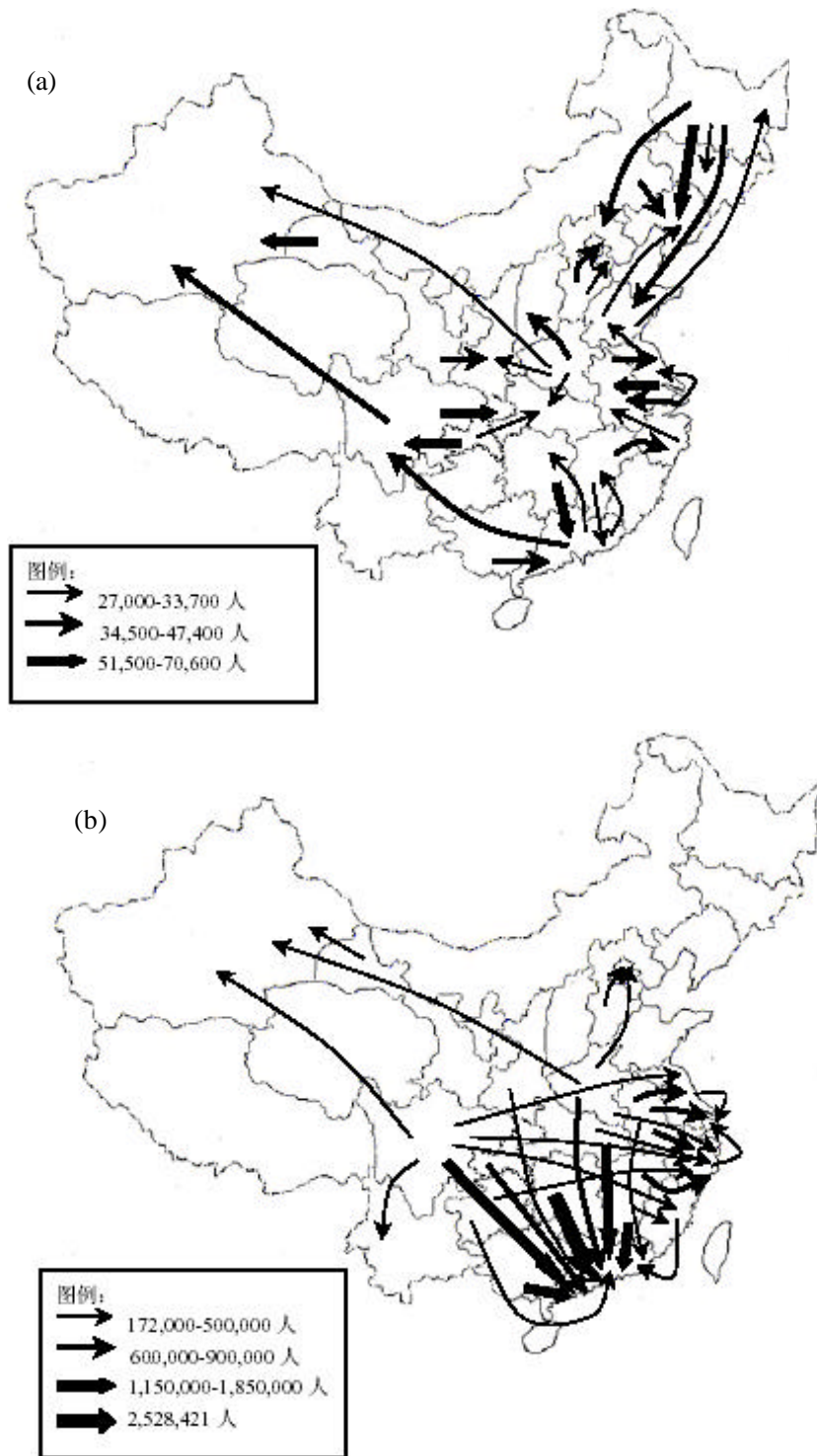


图 1 1995-2000 年最大的 30 个省际迁移人口流向示意图:

(a) 户口移民, (b) 非户口移民。

数据来源: 2000 年全国人口普查 0.95% 抽样数据。

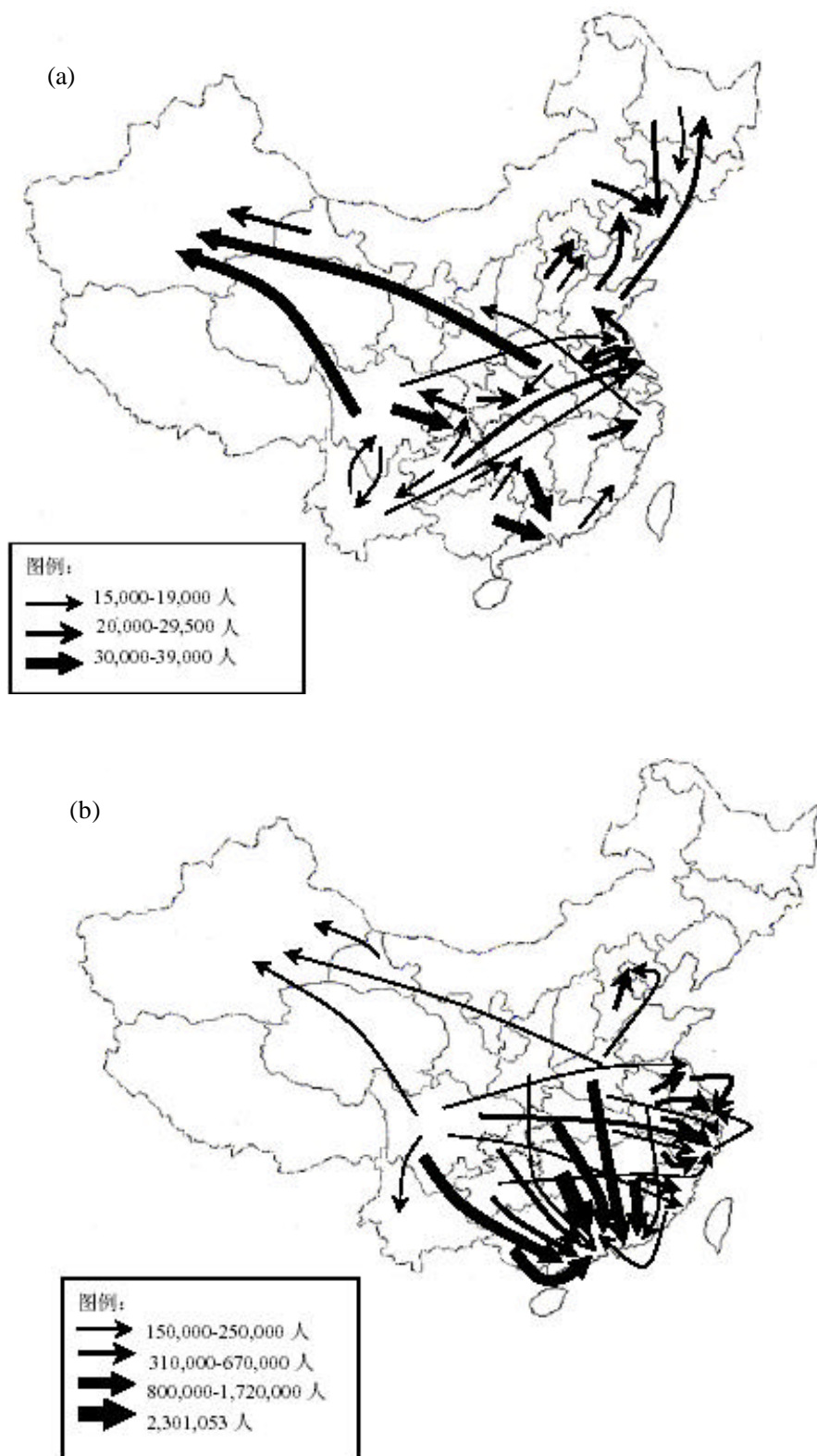


图 2 1995-2000 年最大的 30 个省际农村 15-64 岁迁移人口流向示意图：  
 (a) 户口移民，(b) 非户口移民。

数据来源：2000 年全国人口普查 0.95% 抽样数据。

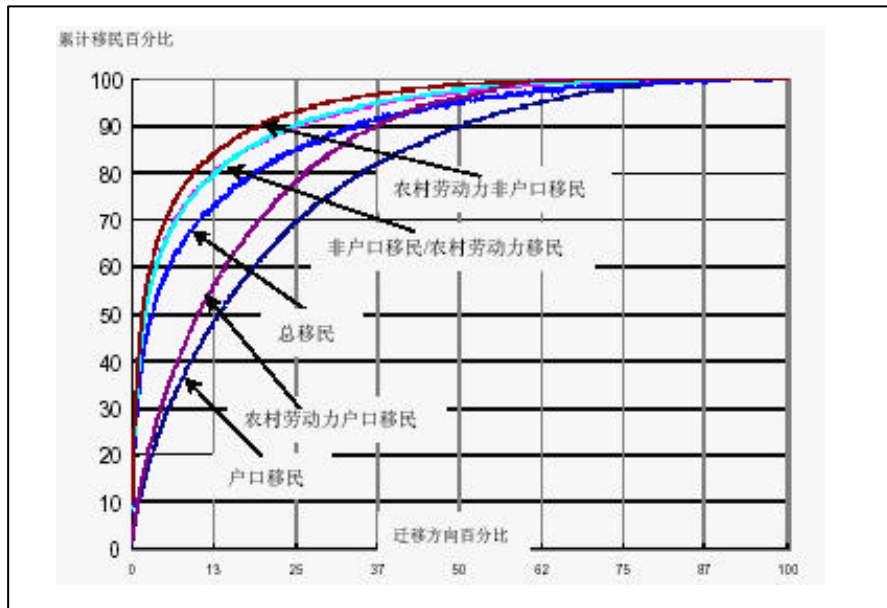


图3 迁移方向和移民数量百分比

注：除西藏以外，理论上的迁移方向有 870 个 (30 × 29)。但实际上有很多迁移方向人数为零。总移民不为零的迁移方向共有 756 个，本图中的迁移方向百分比分母都归一为 756。

# 北京大学中国经济研究中心

## 中文讨论稿系列

### 目 录

No.C2004016	1995-2000 年户口迁移和非户口迁移：描述与分析.....何英华
No.C2004015	地方保护和市场分割：从发展战略的角度考察 ..... 林毅夫 刘培林
No.C2004014	金融创新与长期经济增长..... 施建淮
No.C2004013	中国城市居民预防性储蓄及预防性动机强度：1999-2003 ..... 施建淮 朱海婷
No.C2004012	通过公共产品的供给调控房地产市场.....汪 浩 王小龙
No.C2004011	从 2000 人口普查看我国家庭与老年人居住安排的变化 ..... 曾 毅 王正联
No.C2004010	以晚育为杠杆，平稳向二孩政策过渡.....曾 毅
No.C2004009	我国长期米价研究（1644-2000）.....卢 锋 彭凯翔
No.C2004008	连锁超市“通道费”.....汪 浩
No.C2004007	东亚金融货币合作：短期，中期和长期.....施建淮
No.C2004006	产出不确定，信息不对称与双重危机发生机制 ..... 施建淮 郭美新
No.C2004005	产品内分工：一个分析框架.....卢 锋
No.C2004004	信用体系、金融改革与中国经济发展.....林毅夫
No.C2004003	中国城市居民收入分布的变化：1988-1999 .....万定山
No.C2004002	与林老师对话：经济学方法论篇.....中国经济研究中心
No.C2004001	融资、地价与楼盘价格趋势..... 平新乔 陈敏彦
No.C2003033	要素禀赋、专业化分工、贸易的理论与实证：

- 与杨小凯、张永生商榷  
 ..... 鞠建东 林毅夫 王勇
- No.C2003032 政府保护的动机与效果——一个实证分析  
 .....平新乔
- No.C2003031 “柳暗花明之路”还是空中楼阁？  
 ——就我国中小企业融资改革方向与汤敏博士商榷  
 ..... 陆挺
- No.C2003030 我国农产品贸易趋势和结构变动.....卢锋 雷蕾
- No.C2003029 政策性负担、道德风险与预算软约束  
 .....林毅夫 李志赞
- No.C2003028 经济发展的比较优势战略理论  
 ——兼评《对外开放新时期的对外贸易战略与贸易政策评论》  
 .....林毅夫 孙希芳
- No.C2003027 中国的国有企业与金融体制改革.....林毅夫 李志赞
- No.C2003026 金融改革与农村经济发展.....林毅夫
- No.C2003025 信息、非正规金融与中小企业融资.....林毅夫 孙希芳
- No.C2003024 与林老师对话:发展战略篇之三  
 .....林毅夫 中国经济研究中心 2001 级研究生
- No.C2003023 与林老师对话:发展战略篇之二  
 .....林毅夫 中国经济研究中心 2001 级研究生
- No.C2003022 与林老师对话:发展战略篇之一  
 .....林毅夫 中国经济研究中心 2001 级研究生
- No.C2003021 假冒生产对专利制度的伤害.....平新乔 尹静
- No.C2003020 中国地区(制造业行业)间的技术溢出分析.....尹静 平新乔  
 .....林毅夫 刘培林
- No.C2003018 银行监管的七大政策及其有效性.....张俊喜
- No.C2003017 从中国农民医疗保健支出行为  
 看农村医疗保健融资机制的选择.....平新乔
- No.C2003016 中国国有企业的代理成本估算.....平新乔 范瑛 郝朝艳
- No.C2003015 再谈卫生保健市场中市场与政府的作用  
 .....海闻 王健 赵忠 石光
- No.C2003014 退耕还林:私人承包与政府规制.....王小龙
- No.C2003013 俄国学界评林毅夫,蔡昉和李周著  
 《中国的奇迹:发展战略与经济改革》  
 .....Vadim Shabalin,D.Sc Yakov Pevzer
- No.C2003012 跨界治理:台资参与昆山制度创新的个案研究  
 .....柏兰芝 潘毅
- No. C2003011 工资合约,灰色收入和职业生涯考虑.....范瑛 平新乔
- No. C2003010 农村卫生服务体系探讨  
 .....海闻 王健 陈秋霖 赵忠 侯振刚
- No. C2003009 中国能否通过在农村创造非农工作职位  
 来转移大部分农业劳动力.....D. 盖尔 约翰逊
- No. C2003008 政策性负担与企业的预算软约束--来自中国的实证研究  
 .....林毅夫 刘明兴 章奇