

# 地区间收入差距对农村劳动力迁移的影响

——基于第五次全国人口普查数据的研究

王格玮\*

**摘要** 本文利用中国第五次全国人口普查数据,对农村劳动力迁移决策的影响因素进行了经验分析,发现地区间人均收入差距和迁移距离对劳动力迁移有显著的影响。通过对收入差距变动的模拟,我们发现1985—1990年至1995—2000年期间增加的模拟农村移民数量只约有40%,与实际移民数量一倍以上的增加幅度有很大差距。这说明1990年代移民的大幅增加还与户籍制度改革等其他因素有关。

**关键词** 劳动力迁移,收入差距, nested logit

## 一、引言

自20世纪80年代以来,中国人口迁移的总量一直在上升,这在学术界基本上已是共识。几次权威的人口普查和抽样调查的数据,虽然由于问卷口径不一致使得可比性欠佳<sup>1</sup>,但从中还是可以发现非常明显的增长的趋势。表1显示了第四次和第五次人口普查的人口迁移总量,可以发现十年间移民的数量有了一个很大的增长,其中省际移民量的增长幅度几乎达到了150%。

表1 第四次与第五次全国人口普查的人口迁移量

单位:万人

	本县市区以外移民	省际移民
1985—1990	3413	1107
1995—2000	6165	2747
增长率	80.6%	148.1%

备注:全部移民的年龄都在五周岁以上。

资料来源:王格玮(2004a),五普数据均按四普相似口径进行调整。

如此巨大的移民流量对人们的经济生活造成了很大影响。尤其是农村向城镇迁移的农民,一方面为城镇输送了大量的廉价劳动力,推动了当地经济的发展,另一方面所造成的社会治安和损害城市形象问题也成为城镇居民所

\* 北京大学中国经济研究中心。通讯地址:北京市北京大学中国经济研究中心,100871; E-mail: gwwang@hotmail.com。感谢赵耀辉、何英华、赵冬,本人从他们的建议和论文中受到很多启发,林毅夫、赵忠、姚洋与 Alber Park 的批评指导也使我受益良多,在此一并致谢。当然,文责自负。

<sup>1</sup> 批评来自周皓(2002),Zhao(2003),杨雪(2004)等等。王格玮(2004a)详细讨论了四次全国性人口调查关于人口迁移规模的可比性,并试图将第四次全国人口普查(四普)和第五次全国人口普查(五普)所反映的1985—1990年和1995—2000年的人口迁移量调整到相近的口径予以比较。

关注的焦点。所以,当民间“迁移自由”的呼声日益高涨、地区间经济水平差距持续扩大时,这些因素将导致移民量如何变动,必然成为一个颇具意义的研究方向。

近年来,中国地区间的收入差距问题受到了广泛关注。李实(2002), Kanbur and Zhang(2003), Lin *et al.*(2004)等研究从几种地区间(包括城乡和沿海—内地)的不平等指数中均发现,自1985年来中国地区间收入差距呈持续扩大态势。中国人口迁移的主要流向是由内地向沿海、农村向城镇(王格玮,2004b),第五次全国人口普查显示1995—2000年中国跨县、市、区移民中有50%以上是务工经商型,这个比例在省际移民中更是达到了70%。<sup>2</sup>这说明地区间收入差距对中国人口迁移必然有一定的影响。

90年代间,中国人口的整体教育水平有了很大水平的提高,初中以上教育程度占五岁以上人口的比例由第四次普查时的36.6%提高到第五次普查时的50.8%。教育水平的提高减少了人们的迁移心理成本,增加了就业成功机会,也将大大提高人口的流动性。同时,户籍制度改革在各地区也在不断地进行着,移民的条件比以前又有所放松。此外,移民网络的加强、交通和通讯费用的降低以及高校扩招和收费的增加<sup>3</sup>,这都是中国国内迁移规模可能有如此增长的理由。十年间移民数量的大幅增加必然是多种因素共同造成的,这些因素今后的变化也都将会影响到21世纪初中国人口迁移的趋势。因此,我们有必要对影响现阶段中国人口迁移的各因素进行深入的量化分析,这样才能准确地预测今后中国的人口迁移规模。

本文的主要目的是利用第五次全国人口普查微观数据,建立微观个体选择模型,从个人效用最大化的角度,对农村劳动力迁移决策的各因素进行分析。人口迁移因素分析的理论和经验研究都已经十分丰富,但是以往大家所应用的线性回归和非线性多元选择模型都有缺陷。因此,笔者在利用最新的数据进行因素分析时,对多元选择模型的方法进行了修正,对各种可能的设定进行了详细而深入的探讨。文中将会牵涉到收入变量的设定,收入弹性的计算以及收入差距变动的模拟,并将解答诸如城乡收入差距继续拉大,是否会对城镇带来沉重的农村移民压力之类大家所关注的问题。

本文的结构安排如下:第二部分将回顾人口迁移因素研究的经验模型的发展,介绍本文具体的分析框架和模型设定,并对数据和回归中的样本选择作简要说明。在第三部分我将讨论模型中变量的选择。然后在第四部分中对经验结果进行解释和分析。第五部分是文章的小结,包括基本结论、贡献所在以及存在的问题。

<sup>2</sup> 本文内关于第五次全国人口普查的数据资料如未说明均来源于原始抽样数据,关于数据的说明见后文。

<sup>3</sup> 高校收费增加造成农村家庭子女入学困难,很多家长会因此被迫外出打工以求赚得更多的收入,这个想法来源于何英华(2004)。

## 二、人口迁移因素的经验分析框架

### （一）文献综述

#### 1. 第一阶段——线性回归模型

在 20 世纪 60 年代和 70 年代，研究者基本上都采用古典线性回归模型，分析各地区经济和地理因素对各地间（place-to-place）移民量的影响（Greenwood, 1975）。在这些模型中，地区间的移民量或各地的净移民量被用作因变量，自变量包括影响迁移的各种因素，如距离、地区平均收入等等。因此，这种模型实际上是一个加总（aggregate）宏观计量模型。它有几个缺点：（1）无微观基础。线性回归模型实际上可以通过引力模型转化而来，但引力模型本身只是社会学家和人口学家的一个概念模型。Niedercon and Bechdolt（1969）提出了一个效用最大化模型试图来解决引力模型的微观基础问题，但这篇文章将地区内的每个人都看成是同质的，并且将他们的迁移收益看作是迁移次数和迁入地人口的函数。而且，如此推导出的引力模型中，迁移量是指一段时间内地区间总的迁移量，即迁移人次，这对数据的要求也相当高。Fields（1979）将 MNL（multinomial logit）模型加以改造，也发展出了一个所谓的双对数线性模型。但这与用频率代替概率，而将 logit 模型转化成线性模型估计的方法一样，舍更有效的最大似然法不用，并无多大意义，而且反而要去解决模型设定所导致的异方差问题。（2）需要解决内生性问题。由于人口迁移中的劳动力流动在理论上应该使地区间的经济出现收敛现象，并在发达国家已经得到经验结果的支持（Barro and Sala-i-Martin, 1995），这样地区的平均收入、失业率等经济指标就不是外生的，就需要用联立方程模型来解决内生性问题。这就导致结果更不可信。（3）难以估计人口特征对迁移的影响。年龄、性别等人口特征是人口自身的因素，不同特征的人往往具有不同的迁移倾向，如 Sjaastad（1962）就表明年纪越轻的人迁移倾向就越高。这些特征对迁移决策的重要性往往要大于某些宏观变量。我们虽然也可以根据个人特征分组进行回归，但还是不能获知个人特征与迁移行为的关系。

不过，线性回归模型至今仍有着广泛应用<sup>4</sup>。这一方面是因为其简单直观，另一方面是因为线性模型的使用已经有了相当长的时间，新的研究与以往的结果进行对比非常方便。Lin *et al.*（2004）就利用双对数线性模型用中国 1990 年和 2000 年人口普查（1985—1990 和 1995—2000）与 Greenwood（1969）对美国 1960 年人口普查（1955—1960）的回归结果进行了比较。由

<sup>4</sup> 近年来中文文献对中国人口迁移因素的分析基本上都集中于线性回归模型，例见王桂新（1997），严善平（1998），段成荣（2000, 2001），朱农（2001），蔡昉、王德文（2003）。Chan *et al.*（1997）利用 1990 年全国人口普查数据对户口与非户口迁移进行了分析，使用的也是线性回归模型。

于模型简单,变量也易于统一,所以两篇文章结果的可比性很好。

## 2. 第二阶段——多元选择(非线性)模型

人口迁移是一种个人行为,地区的迁移量只是个人迁移的总和。所以,微观计量才是研究人口迁移更好的选择,也就是说,经验模型应该分析每个人的迁移行为而不是一个地区的迁移量。但是,对个人迁移进行定量分析不同于对其他如工资等连续变量的分析,迁移是一种选择行为,有两种方法来对之加以量化。一种是分析迁与不迁或是对迁入地的选择,一种是分析迁移次数,无论怎样分析,迁移行为只能用离散的变量来度量。当我们将迁移作为因变量时,就需要应用复杂的离散因变量(discrete-dependent variable)模型,所以,对人口迁移影响因素的研究一直受制于计量方法的发展,往往只能在理论计量的某个环节取得重大突破后,经验研究随之也迈进一步。

在20世纪60年代,probit和logit这两种二元(binomial)选择模型已经发展成熟。但仅将决策人的选择设定为“迁移”和“不迁移”会产生很多问题,尤其是对中国来说,各地之间的经济、地理和人文状况差别很大,迁移到此省和迁移到彼省很难把它们当成“迁移”这一个选择。一般的迁移二元选择模型,会加入决策者在原居住地的收入,以观察迁移前收入对迁移决策的影响<sup>5</sup>。而事实上,真正对决策起作用的应该是每个人所面临的收入差(Sjaastad, 1962),或可以说成预期迁移所能得到的收入与迁移前的收入两者的现值之差。在二元选择模型中,是很难在模型中加进迁入地的收入的。一方面,对于不迁移的人,我们无从推断他们的意愿迁入地究竟是哪里,这样就不能获知他们的迁移收入;另一方面,我们也不能直接地用每个移民在迁入地的收入来代表他们迁移前做决策时所面对的迁移收入,这是因为,移民的迁移收入是一个自选择的结果,内生于他的迁移决策。另外,对于迁移影响很大的迁移距离,由于类似于收入变量的缺陷也不能加入到二元选择模型中去。

McFadden(1973, 1981)相继发展了MNL和NL(nested logit)这两种多元选择模型,大大扩展了对经济人的选择行为进行分析的空间。在二元选择中选择单一的缺陷由此得到解决,决策者的选择集可以设定为迁入到各个省乃至更细。由于省级行政区域内的经济和地理情况比较相似,我们就可以运用多元选择模型来分析省级地区收入和地区间距离对人口迁移决策的影响。NL模型是MNL模型的一个扩展,解决了著名的IIA问题(无关选择独立性)。它假设决策是在一个嵌套(nested)的选择集中进行的,允许把相似的选择置于不同的子集(nest)中,以便与其他的选择区别开来。如果某个子集中某些选择还有更进一步的相似关系,还可以继续归并,这样就有了一个多层嵌套选择集合。对每个子集中选择的相异程度,都可以估计出一个相异性

<sup>5</sup> 二元选择模型在中国人口迁移研究中的应用例见Liang and White(1997),庞丽华(2001),Zhu(2002)等文章。

参数 (dissimilarity parameters), 它必须落在  $[0, 1]$  区间内。

直接利用 MNL 模型来分析人口迁移的文献已有很多。如 Liang and White (1997) 对 1987—1992 年间中国人口省际迁移的研究和 Zhao (1999) 对四川农村移民职业选择的研究, 是将 MNL 模型应用到中国人口迁移研究的典型文献。Falaris (1987) 用二层 NL 模型估计委内瑞拉男性人口的国内迁移, 是笔者所检索到最早的一篇用 NL 模型分析人口迁移的文章。之后 NL 模型在迁移因素分析中还是很少, 如 Newbold (1996) 用二层 NL 模型估计了加拿大国外入境移民在本国内部的迁移, Knapp *et al.* (2001) 同样用二层 NL 模型分析了美国国内迁移对城市中心和市郊的选择。Ma and Liaw (1997) 将 NL 模型应用于中国的人口迁移分析, 利用 1987 年中国 1% 人口抽样调查数据, 对 17—29 岁的人口在 1985—1987 这三年间的省际迁移和城、镇、农村间迁移进行了全面的研究。这篇文章的分析方法在中国人口迁移的研究中称得上是凤毛麟角。

## (二) 本文的分析框架

若欲对中国人口迁移作一个全面分析, 必须在模型中包含三种类型的选择: 迁与不迁、迁往何省、迁往城镇与迁往农村。迁与不迁是人口迁移中最基本的问题, 其重要性一般都要大于迁入地的选择问题, 如果只将移民作为总体 (如 Liang and White, 1997), 就不能对迁与不迁的行为进行因素分析。将各省作为不同的迁移选择是考虑到中国作为一个地理面积很大的国家, 到各地的迁移行为有着不同的特征。同时, 以省为单位来收集数据也会带来很大的便利。至于城镇和农村的划分, 则是进一步考虑到同省的城镇和农村经济环境相差巨大, 户籍制度也使得农业户口移民在城镇居住要受到很多的限制。另外, 引入城镇和农村的选择也使我们得以对影响城镇迁移和农村迁移的因素进行分析。当然, 过多地引入选择会使回归的运算量成指数级地增长, 选择集的大小与计算能力的限制是需要我们来权衡的一对矛盾。

本文的基本思路是, 假设样本中的每个人都要在这五年中作一个迁移决策<sup>6</sup>, 这时, 他面临着 61 个不同的选择: 不迁移、迁往 30 个 (不包括西藏) 省 (自治区和直辖市) 的城镇以及迁往 30 个省的农村, 这里所说的城镇和农村都是指居住地。这就等于我们假设, 对所有的人来说, 每个省内的所有城镇是无差异的, 每个省内的所有农村也是无差异的。每个省的城镇和农村都被抽象为单一的选择<sup>7</sup>, 各个选择具有不同的特征, 如预期获得的收入, 就业

<sup>6</sup> 本文经验分析中所指的 1995—2000 年的人口迁移均定义为五年间最后一次移动是由现常住地外的县市区移入的人口, 关于人口迁移定义的讨论详见王格玮 (2004b), “最后一次移出地”与“五年前常住地”就省级范围来说近乎一致。五普中“五年前常住地”这一指标提供的信息太少, 一是省内迁移的迁移距离无法界定, 二是没有迁出地的城乡类型信息。

<sup>7</sup> 其中也包括该人在本省内部跨地区迁移这一个选择。

机会, 距离(代表了交通成本和心理成本)等等。这样, 所有的人都有一个相同的选择集合, 只是他们从这些选择中获得的效用不同, 每个人在决策时最大化自己的效用。

IIA 这个 MNL 模型的基本假设在本文劳动力迁移的研究框架下不能成立, 因为迁移到 A 地和迁移到 B 地这两个选择相对于不迁来说更为相似。另外还可以举出一些例子, 比如迁移到东部沿海某两个地区和迁移到西部某地区相比, 由于我们不能完全将东部沿海地区的特征控制到变量中来, 这样它们的相对近似就是对 IIA 的一个反驳。事实上, IIA 作为一种理想假定, 不可能在多元选择设定中完全成立, 如果各选择较为独立, 一般都不用考虑。只是在人口迁移的地域选择中, 有一些明显的特征违背了 IIA, 所以 NL 模型是较 MNL 模型更为理想的方法。

本文将 61 个选择分为三个子集(见图 1): “不迁移”单独作为一个子集, 迁移到 30 个省的城市作为一个子集, 迁移到 30 个省的农村作为一个子集。前面已指出, 迁与不迁是性质完全不同的两种选择, 势必要将它们划分开来。在 60 个迁移的选择中, 迁移到城市和迁移到农村这两种选择有较大不同, 而相对来说, 各自的 30 个选择更为相似。

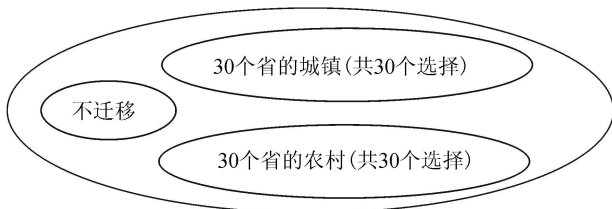


图 1 NL 模型二层嵌套选择集设定

### (三) 数据及研究对象说明

本文的经验分析所采用的微观数据为 2000 年第五次人口普查的 0.95% 按户抽样数据。整体数据包括 1180111 个样本点, 70 多个变量, 比如基本人口特征<sup>8</sup>(年龄, 性别, 婚姻状况)、户口状况、五个迁移变量(最近一次迁移时间, 迁出地, 迁出地类型, 迁移原因以及五年前常住地)以及受教育程度等等。本文从整体数据中又筛选出了包含 518049 个样本点的子样本, 将研究对象集中于中国年龄为 15 周岁到 64 周岁之间的农村劳动力<sup>9</sup>。

<sup>8</sup> 第五次人口普查的标准时间为 2000 年 11 月 1 日 0 时, 人口普查员登记的人口及其各种特征都是反映那个时间点上的情况。所以, 从普查数据中得到的各项指标在本文中如无另外说明均为登记人口在标准时间的信息。

<sup>9</sup> 另外也作了一些技术性的删节, 但数量很小——(1) 删去了迁移前或期末居住在西藏自治区的人; (2) 删去了普查时正在国外工作或学习的人; (3) 删去了从港、澳、台以及国外迁入的人。对于每个移民在迁移时是否为劳动力我们无法确定, 因此本文对劳动力的界定并不是绝对精确的。

划分农村和城镇人口有两个标准, 一是户籍, 二是地域。从表 2 中可以看出, 虽然大多数的劳动力户口性质与居住地类型相一致, 但也有 15.1% 的农业户口劳动力居住于城镇, 3% 的非农业户口劳动力居住于农村。通常人们所指的农村人口和城镇人口事实上都是按户口类型划分。因此本文所提到的农村劳动力均指其户口性质为农业, 而与居住地无关, 即农业户口劳动力。

表 2 全国人口分常住地和户口类型的分布 %

	居住于城镇	居住于乡村
农业户口	12.5	60.3
非农业户口	23.3	3.8

备注: 移民的居住地均指迁出地。

资料来源: 第五次全国人口普查 0.95% 按户抽样数据。

### 三、变量的选取与设定

#### (一) 宏观变量

本文的模型选择了三个宏观变量<sup>10</sup>: 地区人均收入、迁移距离和本省、邻省迁移虚拟变量。地区收入变量采用了 1995 年(决策期初) 各省城镇人均可支配收入和各省农村人均纯收入<sup>11</sup>, 并进行了对数化。本文中省际迁移距离取为各省省会间的最短铁路距离, 而省内迁移的距离则根据 Ma & Liaw (1997) 的做法无论远近均取为省的半径(将省视为一个圆, 用面积算得) 再乘以 0.88<sup>12</sup>。对距离变量也采取对数形式, 由于不迁移是用零距离来识别的, 本文对所有距离加上 1 (公里) 后再取对数。人口迁移还牵涉到文化、习俗、语言等一系列影响因素。鉴于省内和邻省迁移量十分大, 非距离所完全能够解释, 故引入本省或邻省迁移的虚拟变量, 强调在其他变量被控制住的情况下(包括距离), 移民优先选择本省或者邻省。关于三个宏观变量的描述性统计见表 3。

表 3 宏观解释变量描述性统计 %

	均值	标准差	最小值	最大值
地区间距离(公里)	1905	1038	40	6313
本邻省虚拟变量	0.178	0.383	0	1
地区人均收入(元)	2979	1633	880	7439

资料来源: 国家统计局(1996), 铁道部(1997)。

<sup>10</sup> 之所以没有像其他文献那样加入大量的宏观变量, 主要是基于回归估计的考虑。NL 模型的最大似然估计仍是用 Newton-Raphson 方法迭代递归, 当似然函数收敛于一个值时, 才能得到稳定的最终结果。

<sup>11</sup> 该数据来源于国家统计局(1996)。

<sup>12</sup> 实际上, 海南省省会海口到广州的距离是 620 公里, 到广西壮族自治区首府南宁的距离是 510 公里, 另外也考虑了过海的因素(见何英华, 2004)。该数据来源于(铁道部, 1997)。

使用地区人均收入变量,就相当于将个人期望收入写成地区人均收入和个人特征的简化式,这只是一个折衷的办法。这样就等于假设一个人在各地区的期望收入都是地区收入和个人特征以相同的函数形式所决定。这个假设与现实有些差异。尤其在中国,各地的劳动力市场并不是完全开放的,存在着各种管制和地方保护,由于地区间的差异,对工资乃至收入的决定就不是相同的。以地区收入替代个人收入对农村劳动力迁移因素的分析影响不大,但可能对城镇劳动力迁移回归结果的解释造成一些难度。另外,在使用地区收入后,个人收入对迁移的影响会部分体现在个人特征中。这样个人特征因素的影响中就杂合了各种迁移收益和成本剔除我们所能控制的因素之后的效应。如受教育程度的差异既包含对个人收入的影响,也包含所遭受劳动力市场管制的不同。在简化式中,这些影响是不能细分的。

MNL 和 NL 模型隐含着—个经验特征:将某个选择特征变量在各个选择中同时减去相同的值,将不会影响模型的估计结果,即回归中用到的信息都只是选择特征变量的相对量。本文的研究重点放在地区间收入差距对迁移决策的影响,NL 模型的经验特征就正好符合需要。它使我们不必像在二元选择模型中那样去设定收入差距,即使我们应用的是各地区的收入绝对量,估计的系数反映的却是收入相对量的影响。本文对收入变量进行了对数化,因此所代表的收入差距实际上应该是相对收入差距,即各地间的收入比值。

## (二) 个人特征变量

受教育程度是人口迁移决策中的一个十分重要的变量。教育作为一项人力资本投资使受教育者具有更高的工作技能和信息获取能力,同时教育还会改变一个人的观念,尤其是对农村劳动力来说,教育会更大地降低迁移的心理成本。对于中国来说,还有关键的一点就是城乡的分割。城镇为了控制移民数量,往往会出台一系列法规限制外来劳工的就业,这些限制一般来说随着教育程度的升高会有所降低。所以,受教育程度的回归结果还暗含了城镇就业限制对各种教育程度人口的差异。在本文模型中,我们把农村劳动力的教育程度分为三类:小学以下、初中、高中以上。

根据 Sjaastad (1962) 的迁移人力资本投资理论,年龄对迁移决策者的选择有重要影响,大量文献及五普数据均有结果表明农村移民量在低年龄段呈上升趋势,在 20—24 岁年龄段上达到峰值并随后下降,图像类似于—条倒 U 型曲线。因此,本文模型中除年龄变量外还添加了年龄的二次项,以求能反映这种特征。

劳动力的婚姻状况被分为三类:未婚、已婚但五年决策期内不曾结婚(下简称为前已婚)、五年决策期内有结婚经历(下简称为五年间结婚)<sup>13</sup>。这

<sup>13</sup> 对于比例很小的再婚、离婚、丧偶几类人由于很难确定五年决策期内有无结婚,均作相应调整归入未婚或五年间有结婚经历类。



样分类的目的是为了更好地反映婚迁的现象。

### （三）参照组（Reference Group）

由于引入了个人特征变量，其中婚姻状况和教育程度都是类别变量，回归时只能取某一类人为参照组，观察其他类人相对于参照组来说迁移决策有何不同。在本文模型中，参照组取为前已婚及小学教育程度者<sup>14</sup>。这样，本文的模型中就有两种相对效应：城镇与农村迁移相对于不迁移，以及非参照群组相对于参照组。

## 四、经验结果分析

### （一）具体回归<sup>15</sup>的样本选择

有文献（如 Zhu, 2002）发现男性和女性之间的迁移倾向有着结构性的不同<sup>16</sup>，因此本文将这两类人分开回归<sup>17</sup>。非线性回归模型中，因为变量的边际效应的大小并不仅仅取决于各自的回归系数，不同样本间的回归结果不能直接比较系数。不过，我们可以通过具体的计算加以解决，最终还是可以把两类人作迁移决策时各因素的影响共同比较。

对每种个人特征我们都可以估计城镇迁移和农村迁移两个系数。在实际回归中，本文发现年龄对于迁移到城镇和迁移到农村这两个决策的影响差别并不大。还有，未婚者相对于前已婚者迁移的倾向更高，但迁移到城镇的差别和迁移到农村的差别是不大的，为防止结果过于繁杂，本文对年龄和未婚这两个变量都只估计一个迁移效应。

在下面，本文将着重对农村男性劳动力的回归结果进行分析和解释，而对于女性劳动力较类似的结果，分析将比较简略。

### （二）农村男性劳动力的回归结果

农村男性劳动力占全部劳动力的 41.1%，在全部移民中也占 38.8%，是迁移研究中最受关注的一部分群体。在务工经商型移民中，农村男性劳动力<sup>18</sup>

<sup>14</sup> 事实上，如参照组个人特征这样的年轻已婚者是极少的，本文只是把他们当作一个基准，如果要看年轻未婚者的系数，只需要加上未婚者相对于已婚者的回归系数就可以了。

<sup>15</sup> 本文所有 NL 模型的估计使用的均是 Stata 7 中的 `nlogitrum` 命令，而不是其内含的有缺陷的 `nlogit` 命令。关于这两个命令的详细说明以及对 NL 模型性质的相关探讨，请参考 Heiss(2002)和 Koppelman & Wen(1998)。Nlogitrum 程序可通过在 Stata 中运行“`net describe st0017, from(http://www.stata-journal.com/software/sj2-3)`”获得。

<sup>16</sup> 即不能只用性别变量来区分，需要所有的变量都重新回归。

<sup>17</sup> 由于在本文模型中，每个决策者面临着 61 个选择，在具体回归时对计算机内存的要求相当高，这样就使得我们不可能使用全部样本进行回归。根据尝试，样本量在 1 万—2 万间是最好的选择，这样得到的结果已经非常显著。

<sup>18</sup> 为简洁起见，在不引起歧义的情况下，本部分文章有时将“农村男性劳动力”简称为“农村劳动力”或“劳动力”。

更要占到 51%。与因其他动机迁移的人相比, 务工经商型移民再迁移的倾向与频率会更高。因此, 我们重点来分析这一群体的经验结果(表 4 第 I 列)。

表 4 农村劳动力回归结果

变量	男性劳动力 NL 模型	女性劳动力 NL 模型	男性劳动力 MNL 模型
距离	-0.1564*** (0.0272)	-0.2413*** (0.0351)	-0.9028*** (0.0571)
本邻省虚拟变量	0.1718*** (0.0338)	0.2958*** (0.0489)	1.0798*** (0.1032)
地区收入	0.5673*** (0.0920)	0.8277*** (0.1101)	1.8099*** (0.0891)
年龄[迁移]	0.0767*** (0.0267)	0.2373*** (0.0381)	0.0805*** (0.0269)
年龄平方项[迁移]	-0.0018*** (0.0004)	-0.0046*** (0.0006)	-0.0019*** (0.0004)
未婚[迁移]	0.4275*** (0.1187)	1.4026*** (0.1495)	0.4492*** (0.1190)
五年间结婚[城镇迁移]	0.2748** (0.1246)	0.3613** (0.1484)	0.2921*** (0.1259)
五年间结婚[农村迁移]	0.4659*** (0.1679)	1.9513*** (0.1434)	0.5007*** (0.1687)
初中[城镇迁移]	0.5647*** (0.1026)	0.4538*** (0.1036)	0.6020*** (0.1034)
高中及以上[城镇迁移]	1.1314*** (0.1373)	1.0430*** (0.1786)	1.2552*** (0.1387)
初中[农村迁移]	-0.1166 (0.1364)	-0.1592 (0.1266)	-0.1580 (0.1371)
高中及以上[农村迁移]	0.2579 (0.2165)	0.3807 (0.2442)	0.3028 (0.2172)
常数项[城镇迁移]	-3.8507*** (0.4963)	-6.1092*** (0.6354)	-2.5157*** (0.6081)
常数项[农村迁移]	-4.1368*** (0.5034)	-6.3109*** (0.6444)	-1.2172** (0.6068)
相异性参数[城镇迁移]	0.1313*** (0.0219)	0.1856*** (0.0259)	—
相异性参数[农村迁移]	0.2676*** (0.0462)	0.2898*** (0.0416)	—
Log likelihood	-5676.91	-4776.82	-5886.43
样本量	13780	12118	13780
抽样比率	5%	5%	5%

备注 1: 上标为 \*\*\* 表示在 1% 的意义上显著, \*\* 表示在 5% 的意义上显著, \* 表示在 10% 的意义上显著。

备注 2: 参照组为前已婚、小学教育程度、15 岁。参照选择为“不迁移”。

资料来源: 第五次全国人口普查 0.95% 按户抽样数据。

### 1. 地区收入影响分析

三个非个人特征变量: 距离、地区收入和本邻省虚拟变量的回归结果与理论预期完全相符。距离的系数为负, 显示这个变量是迁移成本的一个合适度量。本邻省虚拟变量的系数为正, 说明即使在控制距离, 即迁移距离相同

的情况下, 移民还是倾向于省内迁移或是邻省迁移。

地区收入的系数显著大于零, 说明收入对各选择的边际效用为正, 收入差距对迁移倾向有正的影响。对于“不迁移”这个选择, 原居住地收入越高, 决策者越倾向于不迁移; 如果是某个“迁移”的选择, 则意味着该地的平均收入越高, 决策者越倾向于迁移到该地。另外, 我们还可以看到收入的系数值大于距离的系数(绝对)值, 这说明收入的边际效应要大于距离。但也应注意到在现实中, 各地间收入差距超过一倍的并不多, 而距离多则两千多公里, 少则数百公里, 所以很远的距离给迁移选择带来的负效应很有可能超出两地间收入差距所带来的正效应, 回归的结果说明的只是边际的影响。总之, 地区收入和距离是农村劳动力迁移目的地选择最重要的变量之一, 这点是毋庸置疑的。

在 MNL 和 NL 模型中, 各变量对选择概率的边际效应或弹性并不是单一不变的, 而是一个与迁移概率本身有关的函数(见表 5)。迁移概率本身又是一个由变量与回归系数组成的函数, 所以各变量对选择概率的边际效应归根到底要取决于所有变量和回归系数的数值, 这是与古典线性回归模型不同的地方。本文为衡量各因素对迁移决策的影响, 采取了一种加权平均的方法, 对于每一地区人口内部由于年龄、婚姻状况、教育程度的差异而产生的各种不同回归结果, 都用每一种人的实际数量作为权重进行加权平均。

表 5 NL 模型对数形式变量的直接与间接弹性

直接弹性( $X_n$ 对 $P_n$ )	间接弹性( $X_n$ 对 $P_m$ )
$n$ 不在某子集中 $(1 - P_n)\beta$	$N$ 与 $m$ 不在同一子集中 $-P_n\beta$
$n$ 在子集 $N_k$ 中 $\left[ (1 - P_n) + \left( \frac{1}{\tau_k} - 1 \right) (1 - P_{n k}) \right] \beta$	$N$ 与 $m$ 不在同一子集 $N_k$ 中 $- \left[ P_n + \left( \frac{1}{\tau_k} - 1 \right) P_{n k} \right] \beta$

(1) 迁移概率的决策地收入弹性。回归结果显示地区收入变量的系数为正值, 说明决策地收入越高, 决策者越倾向于不迁移, 弹性必为负值。根据表 5 还可知道, 迁移的决策地收入弹性与预测迁移概率成反比, 因此高收入地区劳动力的弹性<sup>19</sup>更大。图 2 显示了各地收入对该地决策者迁移概率的弹性, 即使随着收入增加, 弹性有增大的趋势, 但幅度并不是很大, 基本上在  $-0.525$  至  $-0.55$  之间。当其他地区收入不变, 而决策地收入增加 1%, 迁移概率将下降 0.5% 多一些, 不到 1%, 所以是弹性不足。收入对迁移概率的影响较之于其他变量来说更大, 但通过观察决策地收入与迁移概率的弹性关系可知迁移行为还是具有一定的稳定性。

<sup>19</sup> 论及大小的时候均指绝对值, 下同。

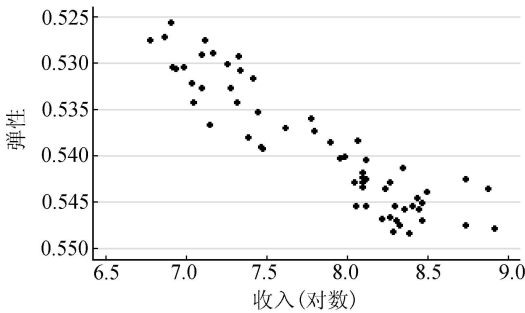


图2 各地区农村男性劳动力迁移选择的决策地收入弹性

以上我们分析的都是决策地的收入与迁移倾向的关系。如果决策地的收入上升1%，而其他地区收入不变，这就意味着决策者面对的收入差距变动幅度也为1%<sup>20</sup>。这与决策地收入不变，而外界收入同时上升1%对迁移决策产生的效果是完全一样的。这里的收入差距变动相当于决策地与全国其他地区之间的收入比值均发生变动，而收入差距的变动还可能还有其他形式，下面将分析另外一种特例。

(2) 迁移倾向与其他地区收入的关系。我们进一步来观察当某个地区收入变化，同时保持其他地区收入不变时，所有中国各地区农村男性劳动力迁移决策所受到的影响。根据回归结果，如果只是  $j$  选择的地区收入上升，其他选择的地区收入不变，那么除了  $j$  选择所在地的决策者的迁移倾向可能会下降外，其他地区的决策者都将更倾向于迁移。

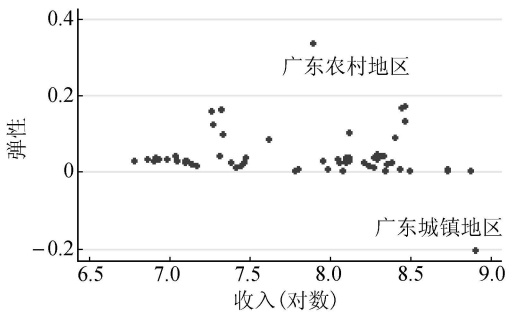


图3 各地区农村男性劳动力迁移选择对广东城镇人均收入的弹性

举一个具体的例子：如果广东省城镇人均收入发生变化，而全国其他地区人均收入都不变，各地的农村男性劳动力的迁移倾向将会如何变动？图3显示了这个例子的结果，我们发现除了原来就在广东城镇居住的农村男性劳动力外，其余地区的人都显示出了正的迁移倾向<sup>21</sup>。居住于广东城镇的劳动力

<sup>20</sup> 如果仅某地区的收入发生改变1%，而其他地区收入不变，因为相对收入差距指的是收入比值，所以收入比值的变动也为1%。在这种意义上，迁移的收入弹性可以看做某种收入差距的迁移弹性。

<sup>21</sup> 不一定是迁移到广东城镇，而是由于广东城镇人均收入发生变化，使全国的收入差距发生了全面的变化。

的迁移弹性在  $-0.2$  左右，而不是上面所计算的  $-0.5$  左右，这是因为在本例中，一个劳动力面临着居住地及其余广东城镇人均收入的同时变化，不仅决策地收入变化了，外部一个地区的收入也发生了变化，所以迁移倾向并没有下降得那么多。在除广东城镇地区之外的农村劳动力中，虽然都表现了正的弹性，但却有一些地区的弹性异常地高，尤其是广东农村地区劳动力，他们的收入迁移弹性达到了  $0.3$  以上。这是因为在这个例子中，各地劳动力的迁移弹性受具体地理位置的影响，其中与广东城镇距离的远近起到了决定性的作用，因此另外一些弹性较高的地区都是湖南、江西等广东的邻省。

(3) 收入差距的区间变动对迁移倾向的影响。在 NL 模型这样的非线性模型中，边际效应不是恒定不变的，也即点弹性和区间弹性是不一样的。从点弹性上我们可以获知变量的边际影响，从区间弹性上我们可以获知变量的区间变动所带来的影响。下面我所做的收入差距的变动模拟，实际上就是收入的区间变动对迁移倾向的影响。

对城乡收入差距，有很多计算方法，我们这里不涉及具体的城乡收入差距指数，而将各城镇收入同增加  $1\%$ ，而农村收入不变这种情况，定义为城乡收入差距扩大了  $1\%$ ，其他的情况也可类似定义。通过对城乡收入差距变化的模拟，我们可以计算各地劳动力的预测迁移概率会发生多大的变化。为了更有针对性，我将对象限定为决策地为农村的农村男性劳动力，并观察他们迁移、迁移到城镇概率的变化。结果表明，迁移到城镇的概率最小的上升了  $0.23$  个百分点，最大的上升了  $0.47$  个百分点，全国的总加权平均为  $0.38$  个百分点。换算成增幅，迁移到城镇的概率最小上升了  $8.16\%$ ，最大上升了  $9.99\%$ ，全国的总加权平均为  $9.56\%$ 。中国农村男性劳动力的数量为  $2$  亿  $9000$  万，如果迁移概率平均上升  $0.38$  个百分点，则迁移到城镇可能增加的数量将会有  $110$  万，要占  $1995$ — $2000$  年这五年间农村男性劳动力移民的  $5.76\%$ ，占普查时已经居住在城镇的农村男性劳动力的  $1.88\%$ 。据 Lin *et al.* (2004) 按消费价格指数调整的城乡收入比例计算， $1985$ — $1995$  年十年间中国城乡收入差距扩大了  $37\%$ （比值从  $1.82$  增到  $2.50$ ），所以根据我的估计结果，这个期间城乡收入差距的扩大，所带来的农村男性劳动力移民量的增加在  $35\%$  左右<sup>22</sup>。

从上面的分析可以获知，地区间收入差距虽然是劳动力迁移决策的重要因素，但收入对迁移概率的弹性还是弱于单位弹性。通过城乡收入差距变动的模拟，发现所造成农村劳动力移民量的增加并不大<sup>23</sup>，与实际农村移民数量的增加幅度有很大差距。所以我们有必要用  $1985$  年各地的真实收入模拟出

<sup>22</sup> 需要说明的是，我的估计均是根据预测概率所作的估计。预测概率本身是一种拟合概率，也即平均化了不可观测因素的迁移倾向，所预测的地区人口迁移概率是一个平均迁移概率。这样的预测与事实必然会有出入，不可能是非常准确的估计。所以，我们在这里的结果也只能看做为一个参考性结果。

<sup>23</sup> 后文显示收入对女性农村劳动力迁移倾向的影响大于男性，但也是弹性不足。

1985—1990年间的各地农村劳动力迁移概率,以将省际收入差距的变动同时考虑进来。经计算,模拟的1985—1990年间全国各地农村男性劳动力的迁移概率在 $[0.035, 0.065]$ 之间,十年间农村男性劳动力迁移概率因收入变化而上升了1.5个百分点左右,上升比例也仅约有40%。这说明,90年代移民量的大幅增加还有其他因素。一方面,80年代末农民纷纷走出家园,外出打工,由此带来的丰厚回报带动了更多的人向发达地区迁移。随着移民网络的逐步加强,农民迁移的信息成本不断降低(Zhao, 2001),迁移观念也在不断更新。因此,农村劳动力迁移数量的增加并不是随着户籍制度的改革而一步增加到城镇所能控制的限度,而是一个渐进的过程。另一方面,随着社会各界对迁徙自由的呼吁,各地区尤其是大城市虽然对农村劳动力迁移仍有很多控制,但还是不断地摸索着一种“双赢”模式,即放宽农村劳动力向城镇迁移的限制,同时城镇也能从中获利。所以,制度因素在十年间又有了很大的转变,这也是中国移民数量增长的一个重要因素。

20世纪90年代中国地区间收入差距的变化不能完全解释移民量的增长,但这丝毫不能削弱收入差距对迁移行为的重要性。无论是信息、观念因素还是制度因素都只不过延缓或阻碍了迁移数量的增长,归根到底,地区间巨大的收入差距现实才是十年间移民数量大幅增长的根本原因。

## 2. 个人特征影响分析

由于模型中有两种相对效应,个人特征变量系数的解释要相对复杂一些。具体来说,由于本文的参照群组设定为小学教育程度和前已婚者,那么估计出来的城镇和农村迁移常数项的数值应该是他们的迁移倾向,而其他群组相对于参照组的迁到城镇或农村的倾向差别则反映在个人特征变量的系数中。这样如果要知道其他群组自己本身迁到城镇、迁到农村以及不迁移这几种行为的倾向及其差别就应该将个人特征变量的系数与参照群组的城镇和农村迁移常数项结合起来考虑。

结果表明,教育程度越高,劳动力越倾向于迁移。但是,教育程度高所带来迁移倾向的增加,主要是反映在更倾向于迁移到城镇上:初中的城镇迁移系数为0.5647,高中以上的迁移系数为1.1314。教育程度的提高对迁移到农村的影响则都不显著<sup>24</sup>。如本文变量选择部分所言,教育程度对迁移倾向的影响暗含了现中国城乡之间的分割,农村劳动力在城镇就业受到事实上的不平等,而这种不平等的界定是受教育程度。城镇对低教育农村劳动力的限制大大降低了他们迁往城镇的意愿,这一点充分地在本文的结果中得到验证。

年龄系数为正以及年龄平方项的系数为负,证实了本文预先估计的倒U型曲线的存在。在本文设定的15—64岁范围内,从15岁开始,迁移倾向随年龄的上升而上升,但到了某一个年龄峰值,迁移倾向又开始下降了。用模

<sup>24</sup> 本文中所提及的“显著”均指统计意义上显著地不等于零。

型的结果计算，年龄的迁移倾向峰值为 21 岁，与事实完全一致。

对婚姻状况的回归结果与我们先前的猜测完全相符。前已婚者的迁移倾向要比未婚者低得多 (0.4275)，而在五年间结过婚的人迁移到城镇的倾向要比未婚者低 (0.2748 < 0.4275)，但迁移到农村的倾向却要比未婚者高 (0.4659 > 0.4275)；另外，他们迁移到城镇的倾向比前已婚者还是要高 (0.2748)。虽然本文将同县市区移动排除在迁移的定义之外，但大量的农村跨县市甚至跨省婚姻依然存在。这些结果都说明婚姻是迁移的一个重要因素。当然，农村婚迁大部分是农村到农村的迁移，所以五年间结婚者比未婚者去城镇的倾向要小。

### 3. 相异性参数：对模型设定合理性的验证

表 4 第 I 列还显示了城镇和农村的相异性系数，即 IV 的系数，分别是 0.1373 和 0.2676。这两个系数显著地大于 0，自然也显著地小于 1，所以根据我们模型设定部分的讨论，通过 NL 模型对选择集进行嵌套分层的做法是合理的。这两个系数还表明，相对于其他的选择，城镇迁移的 30 个选择<sup>25</sup>或是农村迁移的 30 个选择都有较高的相似性，而且，城镇迁移的选择比农村迁移的选择相似程度更高。由农村男性劳动力移民分迁移到城镇和迁移到农村的迁移原因比例 (表 6)，我们可以看出，迁移到城镇的劳动力有 91% 都是务工经商型，相当的一致，而迁移到农村者则只有 74% 是务工经商型，动机有所分散。这样，影响迁往城镇的不可观测因素对他们来说就有可能比较相似，影响迁移农村的不可观测因素相关性就因之而稍弱。

表 6 男性农村劳动力移民迁移原因分目的地分布

迁移原因	迁移到城镇	迁移到农村
务工经商	91.0	74.7
工作调动	1.4	2.4
分配录用	0.3	0.9
学习培训	0.2	0.1
拆迁搬家	0.7	2.9
婚姻迁入	1.0	5.4
随迁家属	1.0	1.2
投亲靠友	1.6	5.0
其他	2.8	7.3
总计	100	100

资料来源：第五次全国人口普查 0.95% 按户抽样数据。

### 4. 与线性和 MNL 模型回归结果的比较

由于本文使用的五普数据较新，模型也较为特殊，迄今很难找到可以用以比较的文献结果。所以我自己用线性和 MNL 模型，对农村男性劳动力以与

<sup>25</sup> 前面已经论及，这些相似性实际上是“影响各个选择决策的不可观测因素的相似性”。为简略起见，均以“选择的相似性”代替之。

上面 NL 模型相同的变量进行回归, 然后比较这几个模型结果之间的差异, 并分析造成偏差的原因。

线性模型与 Lin *et al.* (2004) 相同:

$$\ln M_{ij} = \alpha + \beta_1 \ln Y_{ij} + \beta_2 \ln D_{ij} + \beta_3 C + \epsilon_{ij}. \quad (1)$$

其中  $M_{ij}$  是五年间  $i$  省到  $j$  省的净移民比率 (由  $i$  省到  $j$  省的净移民除以  $i$  省的相应人口),  $Y_{ij}$  是  $i$  省和  $j$  省在决策期初的收入差距 (由 1995 年  $j$  省的人均收入除以  $i$  省的人均收入),  $D_{ij}$  是  $i$  省和  $j$  省的距离 (同省之间则取省内迁移距离),  $C$  是本邻省虚拟变量。有两点需要说明。一是样本较 NL 模型缩减为全部劳动力省际移民, 这是由于线性回归是个加总模型, 只能分析移民量与各因素之间的相关关系。二是很多省之间的移民量为零, 这是因为我们使用的是抽样数据, 且又细分为农村男性劳动力, 而实际的 870 (30×29) 个迁移方向中有 359 个为零, 故在回归中将些观察值舍去。

OLS 回归结果列于表 7, 三个变量以及常数项均在 0.1% 的水平下显著。其中收入差距与距离的系数分别为 0.684<sup>26</sup> 和 -0.394, 因为采用了双对数模型, 所以系数本身就代表着各自的弹性。用线性模型的结果与 NL 模型对比, 收入差距的弹性较为接近, 后者的收入弹性为 0.54, 略小于前者。收入差距的弹性大于距离的弹性, 也与 NL 模型的结果相吻合。

表 7 农村男性劳动力迁移线性模型 OLS 回归结果

变量	回归结果
收入差距	0.6838 (0.1121)
距离	-0.3940 (0.1121)
本邻省虚拟变量	0.5602 (0.1694)
常数项	-4.6995 (0.8373)
$R^2$	0.15
$N$	511

备注: 所有系数均在 0.1% 水平上显著不为零。

资料来源: 第五次全国人口普查 0.95% 按户抽样数据。

由于线性回归模型的缺陷, 其结果必然有一定的偏差。首先, 线性回归模型的样本仅为移民, 而且其收入差距仅为实际迁入地与迁出地之间的收入比例, 这两方面的正选择性必将导致 OLS 回归系数偏高 (upward biased)。其次, 线性模型不能考虑到移民的人口结构, 只能立于一个全国平均的角度上

<sup>26</sup> Lin *et al.* (2004) 以全国人口作为样本, 并加入其他一些控制变量后的收入差距回归系数为 0.595。



来分析，而事实上各省移民的教育等人口特征结构是有差别的，这也会导致回归结果产生偏差。

MNL 模型采用 (1) 式进行回归。其设定惟一与前面的 NL 模型不同之处就是不采取分层嵌套结构，所需估计的系数少了相异性参数。之前的 NL 模型已经验证了中国人口迁移的迁入地选择中 IIA 假设不能成立，这就意味着城镇变量和农村变量的变动对每个人迁移倾向的影响不同（见表 5）。现在我们来分析决策地收入对迁移概率的弹性。表 4 第 III 列显示了 MNL 模型的回归结果。经计算，决策地收入对迁移概率的平均弹性为  $-1.707$ ，与 NL 模型的结果相差很大。表 5 对此有清楚的解释，在 MNL 模型中，某地区收入变化对其他迁移倾向的影响是相同的，然而在 NL 模型中，如果两选择处于同一嵌套子集内，某选择因素的变化会引起另一选择倾向的巨大变化，这是因为城镇和农村迁移的相异性参数都十分小。而在 MNL 模型中，就相当于把这些影响都平均化了，因此出现了很大偏差。

### （三）农村女性劳动力的回归结果

上面我们着重对农村男性劳动力样本的回归结果作了详尽分析。农村女性劳动力占全部劳动力的 36.2%，在全部移民中占 34.8%，在务工经商型移民中占 37.7%。对女性农村劳动力回归的结果显示于表 4 的第 II 列。由于各变量的边际效应和弹性大小不完全反映在系数上，还要取决于各组人本身的迁移概率，所以各变量在各组人口的系数在理论上不能直接横向比较。但是，由于各组人的迁移概率实际上是十分相近的，弹性的大小基本上可以由系数决定，这一点将会从下面的具体计算中得到印证，这就为组间变量效应的比较提供了一种最便利的方式。

我们先来看组内结果。农村女性劳动力回归结果的显著性与男性十分相似。除教育对迁移到农村影响不大外，其余变量的系数都显著地不等于零。年龄的迁移倾向峰值为 25.8 岁，大于男性。婚姻状况对迁移影响的差别更大，未婚者比前已婚者迁移倾向要高出很多，五年间结婚者城镇迁移倾向略高于前已婚者，迁移到农村的倾向不仅高于前已婚者，而且还比未婚者高出许多，这反映了女性农村劳动力婚迁比例高，而婚迁又集中在农村到农村迁移的特征。女性农村劳动力迁移的教育特征与男性完全一致，教育程度的上升提高了迁移到城镇的倾向，但迁移到农村的倾向在不同教育程度人口间的差别均不显著。收入对女性劳动力迁移的影响同样要大于距离很多，本邻省迁移虚拟变量也非常显著。

女性劳动力的收入变量系数大于男性 ( $0.8277 > 0.5673$ )。各地女性农村劳动力的决策地收入弹性在  $-0.78$  到  $-0.81$  之间，明显地要大出男性 ( $-0.5$ ) 许多，这说明收入对女性迁移决策的边际影响要更大。关于收入差距的区间变动对迁移概率的影响，农村女性劳动力在城乡收入差距上升 10% 时

城镇迁移概率上升的百分比区间为 [4.6, 6.4], 明显低于与男性的 [8.1, 10]。

## 五、结 论

本文利用中国第五次全国人口普查数据, 对农村劳动力迁移决策的影响因素进行了经验分析。在计量上采用了 McFadden 的 NL (nested logit) 模型, 避免了多元选择 MNL 模型中过强的 IIA 假设, 并用 FIML (full information maximum likelihood) 方法进行估计, 以求获得更有效的结果<sup>27</sup>。

与以往文献相比较, 本文的分析有四个特点。

第一, 用最新数据揭示了影响中国农村劳动力迁移的关键因素在现阶段的表现。Lin *et al.* (2004) 通过比较 1985—1990 年和 1995—2000 年这两个时期的省际迁移特点发现, 1995—2000 年的人口迁移与地区间收入差距等经济变量的相关程度较十年前要高得多。这样, 建立微观效用模型, 从个人效用最大化角度出发来研究人口迁移才有意义。同时由于劳动力迁移规模的迅速扩大, 劳动力迁移的因素与影响也变成了一个极具潜力的研究方向。

第二, 对多元选择 NL 模型如何与人口迁移选择行为相联系进行了深入探讨, 提出了一个一般性研究框架。以往研究人口迁移因素多使用线性加总回归模型, 近来也有一些微观选择模型的应用, 但这些模型本身都有缺陷。线性回归模型寻求的是地区间迁移量与各地区变量之间的联系, 是一个加总模型。既缺乏微观基础, 也难以解决样本和移民的选择性问题。由于迁移是一个正选择的行为, OLS 估计值要比真实值偏大, 而本文通过用 NL 模型的结果与之比较, 证明了这一点。多元选择 MNL 模型的欠缺在于过强的 IIA 假设, 即各选择 (的不可观测因素) 都是独立的, 而在一个人口迁移这样的选择行为中, 本文的 NL 模型回归结果证明各迁移目的地之间必然存在着不同的相似程度, 如城镇迁移和农村迁移就是差别很大的两类选择。这样, 我们发现 MNL 模型的回归结果也存在着很大的偏差。多元选择模型有着一些很特殊的经验特征, 而人口迁移作为一个复杂的选择行为, 两者的结合必然产生很多值得推敲的地方。以往将迁移置入多元选择框架的研究从未就这一点进行过考虑, 而本文则着力在此处进行扩展。先是提出了一个包含不迁移及迁移到各省城镇和农村的迁移决策选择集。无论从将研究范围扩大到农村与城镇间迁移, 还是从中国国内迁移地区分割的合理性与精确性来考虑, 这都是一个最为理想的分析框架。

第三, 提出了整合分析多元选择模型回归结果的思路。多元选择模型是

<sup>27</sup> 以往估计人口迁移选择的 NL 模型, 受软件和计算机能力所限, 一般都采用不有效的 LIML (limited information maximum likelihood) 方法进行估计, 如 Falaris (1987) 和 Ma and Liaw (1997)。

非线性模型，这样变量的边际效应和弹性将随本身以及所有其他变量的变化而变化，所以本文提出了一个加权平均方法，将人口特征因素予以平均，而保留地区差异进行结果的评估。这里的意义也在于，地理因素是外生的，是每个迁移决策者必然面对的。通过对地理位置控制的放松，我们可以通过各回归结果依地区之间的分布来考察地理位置和各其他经济因素对迁移的影响孰大孰小。总之，加权平均方法使我们可以任意整合所需要的结果，便于分析更细的，如个别地区经济因素变化而引起全国人口迁移规模的变动这样的现象。

第四，涵盖了广泛的人口群体。Ma and Liaw (1997) 研究的是 17 至 29 岁的全国青年人口，Liang and White (1997) 则包含了 15—64 岁的农村到城镇省际移民。本文的样本则涵盖了所有 15—64 岁的全国农村劳动力，总体达到 5.45 亿。

本文囿于计算设备，未能对 NL 模型在中国人口迁移如何分层进一步探讨。在变量的选择上，只选取了对迁移具有决定性的基本变量。而这些变量的基本结果在前人的文献中都已出现，而对大家比较感兴趣的影响中国劳动力迁移的因素，诸如户籍管制<sup>28</sup>、移民网络以及地区间生活（物价）水平差异等，因为无法取得数据，事实上全部置入了模型的不可能观测因素。这一方面可能对结果造成偏差，另一方面也说明本文未能揭示出劳动力迁移与特殊经济现象的联系。这也是我所认为的今后人口迁移因素分析的主要方向所在，而本文的多元选择模型分析框架完全可以继续沿用或加以发展。

下面是对回归结果的总结以及简要评述。在回归结果中，我们发现个人特征的不同会带来迁移倾向显著的差异。尤其是教育程度，还暗含着个人收入的升水和城镇根据这项特征所施加的就业限制。受教育程度越高，城镇迁移倾向会有很大程度地上升。但对迁移到农村，教育程度的不同没有对决策造成显著的影响。在农村移民中，年龄分布呈一倒 U 形曲线，迁移倾向的年龄峰值一般在 20—25 岁左右。婚姻状况同样很大地影响着迁移倾向，五年前已婚者的迁移倾向是最低的。而女性五年间有结婚经历者的迁移倾向很高，尤其是迁往农村的倾向，要高于未婚者。但在是否迁往城镇的决策中，未婚者更容易选择迁移。

地区间人均收入差距、迁移距离和习俗都对农村劳动力迁移有显著的影响。一般地，决策者倾向于迁往经济水平更发达，迁移成本更低以及本邻省的地区。与迁移距离相比，地区收入对迁移选择的弹性更大。面对其他地区经济状况的变化，决策者的迁移倾向也要受到迁移距离很大的影响，邻近地区人口的迁移倾向变化幅度要超过其他地区人口迁移倾向的变化幅度很多。

农村劳动力迁移概率对地区收入（差距）的弹性不足。通过对收入差距

<sup>28</sup> 何英华(2004)对此有深入的探索。

变动的模拟,我们发现1985—1995年增加的模拟移民数量并不大,与实际移民数量的增加幅度有很大差距。这说明,90年代移民量的大幅增加还有如移民网络的加强以及户籍制度改革的持续等其他因素。同样地,模拟城乡收入差距扩大10%约将增加移民35%,平均到每年仅有3%左右。这是与2000年后中国劳动力市场渐趋成熟,地区间农村劳动力移民数量增幅减缓的趋势相同的。因此我们可以预期,如果城乡分割、户口歧视的现状保持不变,仅地区间收入差距的继续扩大,并不能使得中国劳动力尤其是农村劳动力迁移量还能有10年前那样的大幅上升。地区间的收入差距在事实上推动着巨大的潜在迁移量,无非是户籍制度以及结合教育程度等因素造成的城镇行业管制限制了农村移民的生活和就业水平,才使得移民量在一定程度上被各地政府有序调控,这两者正是一推一拉的两种力量。既然推力是如此之大,那么各地政府如果要维持城乡人口迁移的现状,也势必要继续实行严格的户籍制度管理以将意欲迁移的人拉回去。由此,户籍管理制度的改革终其根本是内生于各地的经济状况的<sup>29</sup>,只有城乡经济都较发达的地区才有动力进行改革。各地政府只是为了本地经济的发展才愿意输入一定数量的劳动力,在移民政策上往往要加上很多的条条框框(Zhao, 2003),而不肯让渡一点城镇自己的利益。从这个角度来说,缩小地区间包括城乡和沿海内地的收入差距,对加速中国户籍制度的改革进程也有重要意义。

男性和女性劳动力回归结果的显著性基本一致,在数值上有所不同。男性劳动力的迁移收入弹性要比女性小,但地区收入的区间变动会对男性劳动力的迁移倾向产生更大的影响。

## 参 考 文 献

- [1] Barro, R. J. and X. Sala-i-Martin, *Economic Growth*. McGraw Hill, 1995.
- [2] 蔡昉、王德文,“作为市场化的人口流动——第五次全国人口普查数据分析”,《中国人口科学》,2003年第5期,第11—19页。
- [3] Chan, K. W., T. Liu and Y. Yang, “Hukou and Non-Hukou Migrations in China: Comparisons and Contrasts”, *International Journal of Population Geography*, 1999, 5, 425—448.
- [4] 段成荣,“影响我国省际人口迁移的个人特征分析——兼论“时间”因素在人口迁移研究中的重要性”,《人口研究》,2000年第7期,第14—23页。
- [5] 段成荣,“省际人口迁移迁入地选择的影响因素分析”,《人口研究》,2001年第1期,第56—61页。
- [6] Falaris, E. M., “A Nested Logit Migration Model with Selectivity”, *International Economic Review*, 1987, 28(2), 429—443.
- [7] Fields, G. S., “Place-to-place Migration: Some New Evidence”, *Review of Economics and Statistics*, 1979, 61(1), 21—32.

<sup>29</sup> 对户籍管理制度产生的内生性在林毅夫等(1994)中已经有了很明确的结论。

- [8] 国家统计局,《中国统计年鉴,1995》。北京:中国统计出版社,1996年。
- [9] Greenwood, M. J., “Research on Internal Migration in the United States: A Survey”, *Journal of Economic Literature*, 1975, 13, 397—433.
- [10] 何英华,“户籍制度松紧程度的一个衡量”,北京大学中国经济研究中心硕士论文,2004年。
- [11] Heiss, F., “Specification(s) of Nested Logit Models”, Mannheim Research Institute for the Economics of Aging University Mannheim, Discussion Paper, 2002.
- [12] Kanbur, R. and X. Zhang, “Fifty Years of Regional Inequality in China: A Journey Through Central Planning, Reform and Openness”, Paper for the UNU/WIDER Project Conference on Spatial Inequality in Asia, March, 2003.
- [13] Knapp, T. A., N. E. White and D. E. Clark, “A Nested Logit Approach to Household Mobility”, *Journal of Regional Science*, 2001, 40(1), 1—22.
- [14] Koppelman, F. S. and C. Wen, “Alternative Nested Logit Models: Structure, Properties and Estimation”, *Transport Research-B*, 1998, 32, 289—298.
- [15] 李实,“中国个人收入分配研究回顾与展望”,北京大学“社会公平与公正研讨会”会议论文,2002年7月。
- [16] Liang, Z. and M. White, “Market Transition, Government Policies, and Inter-provincial Migration in China: 1983—1988”, *Economic Development and Cultural Change*, 1997, 45(2), 321—339.
- [17] Lin, J. Y., G. Wang and Y. Zhao, “Regional Inequality and Labor Transfers in China”, *Economic Development and Cultural Change*, 2004, 52(2), forthcoming.
- [18] 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,上海人民出版社,1994年。
- [19] Ma, Z. and K. Liaw, “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.
- [20] McFadden, D., “Conditional Logit Analysis of Qualitative Choice Behavior”, in Zarembka, P., ed. *Frontiers in Econometrics*, 105—135. New York: Wiley, 1973.
- [21] McFadden, D., “Econometric Models of Probabilistic Choice”, in Manski, C. F. and D. McFadden, eds, *Structural Analysis of Discrete Data With Econometric Applications*, 198—272. Cambridge, Mass.: MIT Press, 1981.
- [22] Niedercorn, J. H. and B. V. Jr. Bechdolt, “An Economic Derivation of the ‘Gravity Law’ of Spatial Interaction”, *Journal of Regional science*, 1969, 9, 273—282.
- [23] Newbold, K. B., “Internal Migration of the Foreign-born in Canada”, *International Migration Review*, 1996, 30(3), 728—747.
- [24] 庞丽华,“多层次分析方法在人口迁移研究中的应用——省际劳动力迁移的多层次分析”,《中国农村观察》,2001年第2期,第11—17页。
- [25] Sjaastad, L. A., “The Costs and Returns of Human Migration”, *Journal of Political Economy*, 1962, 70 (Supplement), 80—93.
- [26] 铁道部,《全国铁道旅客时刻表,1997—1998年》。北京:中国铁道出版社,1997年。
- [27] 王格玮,“四次全国性人口调查的迁移信息和统计结果的分析和比较:1982—2000”,手稿,北京大学中国经济研究中心,2004年(a)。
- [28] 王格玮,“1995—2000年中国人口迁移规模:对第五次全国人口普查数据的深入考察”,手稿,北京大学中国经济研究中心,2004年(b)。
- [29] 王桂新,“中国人口迁移弹性与劳动市场转变”,《南方人口》,1997年第2期,第43—47页。

- [30] 严善平, “中国九十年代地区间人口迁移的实态和机制”, 《人口与经济》, 1998年第3期, 第3—13页。
- [31] 杨雪, “再论人口‘迁移’的概念与统计口径界定”, 《西北人口》, 2004年第1期, 第19—23页。
- [32] Zhao, Y., “Labor Migration and Earnings Differences: The Case of China”, *Economic Development and Cultural Change*, 1999, 47(4), 767—782.
- [33] Zhao, Y., “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.
- [34] Zhao, Z., “Migration, Labor Market Flexibility, and Wage Determination in China: A Review”, Working paper, E2003007, China Center for Economic Research, Peking University, 2003.
- [35] 周皓, “我国人口迁移研究的回顾、总结与讨论”, 《人口与经济》, 2002年第1期, 第56—59页。
- [36] 朱农, “中国四元经济下的人口迁移——理论、现状和实证分析”, 《人口与经济》, 2001年1期, 第44—52页。
- [37] Zhu, N., “The Impacts of Income Gaps on Migration Decisions in China”, *China Economic Review*, 2002, 13, 213—230.

## The Effects of Regional Income Gap on Migration: New Empirical Results from the 2000 Census

GEWEI WANG  
(*Peking University*)

**Abstract** Using the 2000 census data, this paper employs a choice-based model to analyze the determinants of labor migration in China. Empirical results show that regional income gap and distance have significant but unequal effects on labor migration and migration is more sensitive to income gap than to distance. Simulation shows that the effect of the income gap from 1985—1990 to 1995—2000 only accounts for forty percent of the actual migration volume. This shows that changes of other factors, such as the gradual liberalization of the Hukou system, played an important role on the change of migration volume.

**JEL Classification** C25, O15, R23