

人口流迁

基于嵌套 Logit 模型的中国省际人口 二次迁移影响因素分析

刘建波 王桂新 魏 星

【内容摘要】 本文主要根据中国第五次人口普查 1% 抽样数据, 采用嵌套 Logit 模型, 考察和分析 1990 年代后期中国省际人口二次迁移的主要影响因素。通过对一般迁移、再迁移和回归迁移的考察分析发现, 无论是个人因素还是区域经济因素, 都对其迁移行为具有重要影响: 再迁移者明显受到潜在目的地区域环境因素的吸引; 而回归迁移者对出生地的经济状况敏感程度较弱, 但个人因素对其迁移选择有较大影响。

关键词: 迁移; 再迁移 (Onward Migration); 回归迁移 (Return Migration); 嵌套 Logit 模型

【作者简介】 刘建波, 复旦大学人口研究所硕士研究生; 王桂新, 复旦大学人口研究所教授; 魏星, 女, 复旦大学人口研究所博士研究生。上海: 200433

1 引言

作为人类活动的重要组成部分, 人口的迁移流动古已有之, 但这一过程经常会受到诸如就业、收入等区域环境因素的影响, 在人口迁移规模、强度、频度、参与的人群等方面亦表现出不同的特征, 中国省际人口迁移发展状况就是最好的佐证。新中国成立以来, 省际人口迁移(其他人口迁移流动亦然) 受到户籍制度的制约以及相关政策的影响, 其规模、流向、强度等均呈大起大落的无序变动。而且在相当长的时间里, 区域人口迁移的规模和频度被控制在极低的水平。经济体制改革以来, 中国人口迁移日趋活跃, 迁移规模逐渐增大, 在 1980 年代持续增强, 进入 1990 年代后虽有几年出现钝化态势, 但经过数年的能量累积, 到 1990 年代后期旋即又以更猛的势头“井喷”而出, 人口迁移规模急剧扩张, 与此前 20 年的迁移水平相比, 几乎呈指数函数陡增(王桂新、刘建波, 2003)。

长期以来, 人们多将人口迁移视作社会现象, 从国家(社会)的宏观层面上进行研究, 在这种宏观的研究模式下, 人口迁移仅仅是一种社会过程。但是归根结底来讲, 人口迁移毕竟是个人的选择行为和活动过程, 它与社会宏观结构有关, 却又不能简单等同于社会过程。因此, 从微观角度出发的研究模式往往将人口迁移定义为个人行为选择的集合。尽管这两种研究思维往往会导出殊途同归的结论, 但是在很多情况下, 研究角度的不同决定了它们所能揭示的内涵也将相差甚远。

宏观层次的迁移研究认为, 人口迁移是由地区间的就业机会、收入水平以及人均拥有的社会资源等方面的差别所造成的; 将人口迁移作为区域差异的调控因素来关注; 其迁移方向是从乡村地区流向就业机会较多、收入水平较高和社会资源相对更丰富的城市地区。但是, 宏观层次的迁移研究不易解释逆向迁移流的存在意义, 同时也难以解释那些发生在就业、工资等方面差别不大的地区间的迁移现象; 况且在社会制度没有相应变革的情况下, 人口在区域之间的自由迁移能否消除(或减少)区域差异也是值得怀疑的。Kao- Lee Liaw 发现在加拿大人口迁移极其频繁, 但是地区间一直保持着巨大的差异, 人口迁移在传统意义上对区域差异的调控、均衡作用并没有体现出来(Kao- Lee Liaw, 1990)。

这种理论和现实情况的差异看起来像是一种“悖论”。其实, 我们应当看到人口迁移是一个高度

“个人化”的过程, 不同的人群有不同的迁移偏好, 不同的迁移者对目的地有不同的选择: 迁移行为不仅仅取决于各个地区的经济水平及其差异, 还受到个人因素和其他区域因素的作用。

与宏观研究方法不同, 微观角度对迁移的研究, 是从每个迁移者的迁移倾向入手研究, 这样的微观研究可以比较清楚的解释人们迁移的动因。Kao- Lee Liaw 和 Ledent 在 1987 年提出用离开本地的概率和迁移目的地选择概率两个层次的嵌套 Logit 模型来考察迁移者的迁移动因, 1988 年他们将迁移者的个人因素纳入到这个模型里面(Kao- Lee Liaw, 1990)。本文也将遵循这一思路, 构建一个多层次的嵌套 Logit 模型, 来研究中国 1990 年代后期的中国省际迁移选择模式, 考察个人因素和区域环境因素对迁移者选择行为的影响。我国第五次人口普查的数据为这种方法提供了必要的技术支持。人口普查登记了每个人的居住地变动情况, 从中得到的每一条迁移记录都是单个迁移者的迁移行为, 而不是宏观意义上的总迁移量或者净迁移量, 这才使得对具体的迁移行为有感性的接触。^①

以往的人口迁移研究多关注调查时期内(如 5 年内)常住地是否改变即是否发生迁移, 到调查时点呈现的最后一次迁移的规模、流向、强度, 以及移民属性特征等问题, 而往往忽略在调查时期内发生的多次迁移及其迁移路径。但是人们通常更希望尽可能复原并了解实际发生的人口迁移行为、迁移路径及迁移模式。本文力图对这个问题进行一些初步探讨。为了更清楚描述人们的迁移行为, 本文将经历过多次迁移的人群分离出来, 仔细研究这些高流动性的人口究竟有什么特征。由于数据的限制, 我们现在最多只能知道每个迁移者生命中所经历的三个地点: 出生地、5 年前居住地和现住地; 根据迁移者的这三个地点属性, 我们定义以下两类迁移行为:

回归迁移(Return Migration): 迁移者经历了多次迁移以后, 最终回到自己的家乡(出生地), 本文把这样的迁移称之为回归迁移; 在数据上表现为该迁移者现在居住在出生的省份, 而 5 年前居住地却在外省。

再迁移(Onward Migration): 迁移者在 5 年之前至少迁移过一次(从出生地到 5 年前居住地), 但在最近 5 年之中又迁往另外一个目的地(现在居住地)的迁移。事实上, 回归迁移是一种特殊的, 即最近 5 年内迁往的另一个目的地为出生地的再迁移类型。但本文以后分析的再迁移仅指在最近 5 年内又迁往出生地以外另一个目的地的迁移。

为叙述方便起见, 我们把所有发生回归迁移和再迁移的移民统称为二次迁移者, 从出生地迁往 5 年前居住地的迁移称为其第一次迁移, 回归迁移或再迁移称为其第二次迁移。

有学者在研究国外的人口二次迁移现象后发现, 不仅再迁移人口要受经济因素影响, 回归迁移人口对此也很敏感(K. Bruce Newbold, Kao- Lee Liaw, 1995)。那么中国的情况是否与国外的研究经验相符, 这也是本文要关注的重点。

2 人口迁移多层次嵌套 Logit 模型的构建

在本文的模型里, 我们假定所有在 1990 年代后期的生存人口是潜在迁移者(potential migrant, 简称 PM), 如果他们当时不是居住在自己出生的省份, 那么就定义为潜在的二次迁移者。^② 为了构建模型, 我们将迁移者的迁移决策过程分解为以下三个层次: 第一层次, 他们选择是否要迁移离开本地; 第二层次, 如果他们选择了要离开, 那么进一步选择是回到自己的故乡(出生地)还是继续前往另一个陌生的地方; 第三层次, 如果这些二次迁移者选择的不是回到自己的出生地, 那么他们将选择要迁入新的目的地。

那些潜在迁移者的迁移偏好是由若干个人因素和社会环境因素共同作用的结果。对于一个出生

① 本文引用的人口数据主要来自全国第五次人口普查 1% 光盘数据。

② 忽略死亡迁移人口的影响, 只取在 1995 年之前发生过迁移, 而 2000 年人口普查时还生存的迁移人口作为研究对象。

在 l 省, 1995 年时居住在 i 省, 其个人因素为 a 的 PM 而言, 他的迁出概率被定义为 $P_{D|i, a, l}$, 选择留在原地的概率为 $P_{S|i, a, l}$; 在第二层次的考虑中, 选择回归的概率为 $P_{R|i, a, l, D}$, 选择再迁移的概率为 $P_{O|i, a, l, D}$; 在第三层次中, 选择 j 省作为迁移目的地的概率为 $P_{j|i, a, l, O} (j \neq i)$ 。这些概率均为在给定个人和区域信息下的条件概率。

进一步, 假设这几个选择都是基于潜在迁移者自身的效用水平来决定的。和前面假设相一致, 在三层次的选择模型里, 定义每个潜在迁移者选择迁向 j 省得到的效用用以下随机变量表示:

$$U_{j|i, a, l} = V_{j|i, a, l} + E_{f(i, j)|i, a, l} + E'_{g(i, j, l)|i, a, l} + E''_{j|i, a, l} \quad (1)$$

(1) 式中 $V_{j|i, a, l}$ 为非随机部分, 剩下三项均为随机变量, 代表了三个层次的选择。其中 $f(i, j)$ 、 $g(i, j, l)$ 定义如下:

$$f(i, j) = \begin{cases} S(\text{Stay}), & i = j \\ D(\text{Depart}), & i \neq j \end{cases}, \quad g(i, j, l) = \begin{cases} S(\text{Stay}), & i = j \\ R(\text{Return}), & i \neq j, j = l \\ O(\text{Oward}), & i \neq j, i \neq l \end{cases}$$

基于效用最大化的假设, 我们认为每个潜在迁移者会比较每一个可选目标: 是迁移还是留在本地, 是迁回自己的出生地还是迁往一个新的地方。通过比较, 从中挑选一个期望效用最高的目标, 并以此来决定自己的选择行为。

根据模型的三个层次, 本文将这个嵌套 Logit 模型分解为三个子模型。

首先, 第三层次目的地选择的子模型为:

$$P_{j|i, a, l, O} = \frac{\exp(V_{j|i, a, l} / \mu_3)}{\sum_{k \neq i, l} \exp(V_{k|i, a, l} / \mu_3)} \quad (2)$$

上式中的 μ_3 是一个随机参数, 代表 $E''_{j|i, a, l}$ 和 $E''_{k|i, a, l}$ 所有的不确定性。 $V_{j|i, a, l} = CZ_{j|i, a, l}$, C 是一个行向量参数, 表示各个因素对个人效用影响的权重。 $Z_{j|i, a, l}$ 为包含了所有可观测因素的列向量, 这里的因素包括各个可供选择的目的地经济水平、就业状况和其他区域环境因素, 以及这些因素和个人因素的交互作用。

其次, 在选择再迁移还是回归迁移的第二层子模型里, 其选择概率为:

$$P_{R|i, a, l, D} = \frac{\exp(V_{j|i, a, l} / \mu_2 - \mu_3 I_{O|i, a, l} / \mu_2)}{1 + \exp(V_{l|i, a, l} / \mu_2 - \mu_3 I_{O|i, a, l} / \mu_2)} \quad (3)$$

与 μ_3 一样, (3) 式中的 μ_2 也是一个未知的随机参数, 代表这个子模型的不确定性。 $I_{O|i, a, l} = \text{Ln}[\sum_{k \neq i, l} \exp(V_{k|i, a, l} / \mu_3)]$ 包括了所有可供选择的再迁移目的地对二次迁移者的吸引作用, $V_{l|i, a, l} / \mu_2 = BY_{l|i, a, l} + B_0$, $Y_{l|i, a, l}$ 为一个列向量, 包含那些地区所有的可观测变量值, B 为它的行向量参数。

再次, 第一层子模型的迁出概率定义如下:

$$P_{D|i, a, l} = \frac{\exp(A_0 + \mu_2 I_{D|i, a, l} - V_{j|i, a, l})}{1 + \exp(A_0 + \mu_2 I_{D|i, a, l} - V_{j|i, a, l})} \quad (4)$$

其中, $I_{D|i, a, l} = \text{Ln}[\sum_{k \neq i, l} \exp(V_{k|i, a, l} / \mu_2) + \exp(\mu_3 I_{O|i, a, l} / \mu_2)]$, 表示除目前居住地以外所有地区对潜在迁移者的吸引作用。 $V_{j|i, a, l} = AX_{j|i, a, l}$, $X_{j|i, a, l}$ 为一个列向量, 包含了除目前居住地以外所有地区的所有可观测变量值, A 为它的行向量参数。

本模型的拟合优度 (Goodness of Fit) 主要用 Nagelkerke R^2 来表示。 Nagelkerke 修改了 Cox & Snell R^2 , 令 $R^2 = \frac{R_{cs}^2}{R_{MAX}^2}$, 其中 $R_{cs}^2 = 1 - \left(\frac{L(0)}{L(\beta)}\right)^2$, $R_{MAX}^2 = 1 - (L(0))^2$, $L(0)$ 为方程式中只包含常数项时的似然比, $L(\beta)$ 为当前模型的似然值。 Nagelkerke R^2 与线性模型中的 R^2 相似, 用来反映由回归方程解释的变异的百分比。

3 一般迁移者和二次迁移者的差异对比

2000 年人口普查资料显示,在 1995~2000 年 5 年间,我国省际迁移人口规模达 3398.12 万人,年平均规模约为 679.6 万人。在我们使用的 1% 抽样样本里,总人口为 1180111 人,迁移者(所有迁移者) 32909 人,再迁移者 1051 人,回归迁移者 1716 人。迁移人口占总人口的比重约等于再迁移人口(或回归迁移人口)占迁移人口的比重。回归迁移尽管只是再迁移的一种特殊形式,但是这类迁移人口总数已经超过了其他再迁移人口的规模,说明回归迁移是一种非常重要的二次迁移模式,有必要作单独分析。这些迁移人口的简单属性如表 1 所示。

表 1 一般迁移者和二次迁移者属性特征的简单对比(%)

		总人口	一般迁移者	再迁移者	回归迁移者
性别	合计	100.0	100.0	100.0	100.0
	男性	51.2	52.8	51.2	67.2
	女性	48.8	47.2	48.8	32.8
教育程度	合计	100.0	100.0	100.0	100.0
	较低	28.3	19.6	12.5	19.3
	中等	44.5	65.3	60.7	60.2
	较高	3.7	4.5	18.6	15.0
	在校学生	23.6	10.6	8.2	5.4
婚姻状况	合计	100.0	100.0	100.0	100.0
	未婚	20.2	45.0	37.9	36.3
	初婚	71.2	52.6	57.0	57.8
	再婚	2.2	1.2	1.8	2.3
	离婚	0.9	0.5	1.1	1.7
	丧偶	5.6	0.7	2.3	1.9
年龄	合计	100.0	100.0	100.0	100.0
	15 岁以下	23.2	7.5	4.0	5.2
	15~60 岁	66.6	91.0	90.1	89.0
	60 岁以上	10.2	1.5	5.9	5.8
户口	合计	100.0	100.0	100.0	100.0
	农业	75.2	78.4	52.8	49.4
	非农业	24.8	21.6	47.2	50.6
居住地类型	合计	100.0	100.0	100.0	100.0
	城市	23.4	53.0	64.5	36.1
	镇	13.2	19.9	19.8	14.0
	乡村	63.4	27.1	15.7	49.9
迁出地类型	合计	—	100.0	100.0	100.0
	乡	—	38.6	22.4	10.3
	镇居委	—	7.4	12.6	16.4
	镇村委	—	40.1	22.5	17.0
	街道	—	13.9	42.5	56.3
迁移原因	合计	—	100.0	100.0	100.0
	经济	—	75.3	72.6	54.6
	非经济	—	24.7	27.4	45.4

注:①受教育水平变量中,“较低”包括未上过学、扫盲班和小学,“中等”包括初中、高中和中专,“较高”包括大学专科、大学本科和研究生。

②迁移原因变量中,“经济”原因包括务工经商、工作调动、分配录用和学习培训,“非经济”原因包括拆迁搬家、婚姻迁入、随迁家属、投靠亲友和其他原因。

从表 1 中可以看出迁移人口和非迁移人口、一般迁移人口和二次迁移人口的差异:

首先,各种类型的迁移人口都比总人口更具有“活力”,二次迁移人口在某些方面较之一般迁移人口更胜一筹。如从年龄上看,迁移的人群中年轻人的比重较高,抚养比较低;从婚姻状况看,迁移人口的未婚比例和单身比例明显高于总人口;在教育程度方面,迁移人口的平均受教育年限更长,二次迁移人口在高学历方面的优势尤其明显。

其次,比较再迁移、回归迁移和一般迁移之间的差别可以发现,二次迁移者有很多特征不同于一般的迁移者,而且回归迁移者的特点更鲜明。如与一般迁移者携带农村户口出来闯荡生活不同,二次迁移者中非农业户口的比例很高,而且有一大半人原来是居住在街道的。回归迁移者相对更可能选择迁回乡村,而再迁移者较一般迁移者而言更多的涌向城市。回归迁移者还有一些自身的特点,如其男性比例明显高于其他人群,迁移目的在很大程度上是非经济原因(45.4%)。那么究竟是哪些因素在促使这些迁移人口多次更换居住地,以下将根据前面构建的嵌套 Logit 模型来分析个人因素和区域因素对二次迁移者迁移决定的影响。

4 基于多层次嵌套 Logit 模型的二次迁移影响因素分析

4.1 个人变量和区域环境变量的选择

潜在迁移者的效用不仅仅取决于个人因素,还受他们所处环境因素的影响。因此,我们在构建模型时不仅要考虑代表性个人因素变量,而且也要考虑区域环境因素变量。经过分析,构建模型时初步选择了以下变量:

个人因素变量:性别、年龄、教育、婚姻、就业状况、户口、家庭人口总数、居住地类型、迁出地类型;

区域经济因素变量:各省市区的平均工资、从业人员总数、迁移传统、各省 GDP 总量、人均 GDP、农村人均纯收入、资本形成总额、固定资产投资、存货增长、居民平均消费总额、城乡消费对比、各省市区的总人口等等,以及这些变量的增长率。

对以上变量作进一步分析,发现有些变量之间存在较强的相关性,可能会造成共线性影响,因此在构建模型时剔除那些和主要变量相关性过高的其他变量(相关系数大于 0.8 或者小于-0.8)。由于这个原因,各省市区的 GDP 总量和人均 GDP 都没有进入我们构建的模型。^①

4.2 样本数据的筛选

在以下的参数估计中,本文选择将 20~44 岁的年轻人作为研究对象,其目的有二:第一,这个年龄段的人群最具有迁移偏好,分年龄迁移率在这里达到峰值,这样就避免了其他年龄段因为数据量过小而带来的统计上的误差;第二,年轻人的迁移动因几乎都是与经济因素有关,我们将进一步研究各地的就业状况、经济水平等区域条件是如何与个人因素交互作用的,同时也更好的理解各地经济政策对人口迁移的影响和制约。

4.3 模型参数的估计

4.3.1 子模型 1:潜在迁移者的迁出概率参数估计

根据迁出概率的子模型,我们从普查数据提取所有 5 年前居住地并非出生地的潜在二次迁移者样本,剔除一些缺失变量值的样本,共 40381 个数据,使用 Logistic 回归进行估计。为了说明个人因素和区域因素分别如何作用于迁出概率,这个子模型分三次分别拟合只有个人因素的情况、个人因素加区域环境因素的情况以及两种因素交互作用对迁出概率的影响,具体分析结果如表 2 所示。

从表 2 中可以看出,这个子模型的拟合程度比较好,解释能力也比较强。在模型 1 中,只有个人因素变量进入了模型,结果显示年龄、有无配偶、是否就业、户口性质、家庭人口总数和个人受教育程度都有显

^① 数据来源:1997~2001 年中国统计年鉴,中国统计出版社。考虑 1996 年统计年鉴没有重庆市的数据,因此在文章的分析里我们只用了 4 年的平均值。

著的影响。非农业人口的迁移概率较低;年龄和有无配偶(有配偶= 1,无= 0)也会影响迁移的选择。而家庭人口越多,潜在迁移者则越有可能迁往外地;教育程度越高,迁移的动力也会增加。对于就业,我们的数据是反映迁移后的就业状况,因此不能反映迁移前的状况,较为合理的解释应该是年轻人的失业率都比较高,而迁移人口相对较容易找到工作(他们有更大的欲望或动力去寻找工作)。

表 2 个人因素和区域环境因素对迁移者迁出概率的影响

	模型 1	模型 2	模型 3
个人因素			
常数项	5.09**	10.76**	10.14**
年龄	-0.11**	-0.12**	-0.12**
配偶	-0.52**	-0.58**	-0.59**
就业	0.14**	0.15**	0.14**
户口	-1.52**	-1.85**	-1.56**
家庭人口总数	0.08**	0.07**	0.07**
受教育程度	0.07**	0.19**	0.23**
区域因素			
就业增长率		-0.18**	0.04
农村人均纯收入		-0.07**	-0.04**
城乡消费对比		-0.18**	-0.20**
居民平均消费		-0.00**	-0.00**
固定资产投资增长率		-0.02**	-0.02**
农村消费增长率		0.09**	0.14**
交互项			
户口 × 农村消费增长			-0.04**
教育 × 就业增长率			-0.05**
教育 × 农村收入			-0.01
Nagelkerke R ²	0.321	0.512	0.514

注:* 代表在 0.05 水平上显著;** 代表在 0.01 水平上显著。

模型 2 加入了区域环境因素的变量,使整个模型的解释能力明显提高(Nagelkerke R² 由 0.321 变为 0.512)。除了受教育程度变量的影响力有明显上升外,其他个人变量对迁出概率的影响几乎不变。区域环境因素变量中,就业增长率、农村人均纯收入、城乡消费对比、居民平均消费总额、固定资产投资增长率对迁出概率有负向的影响,其中当地就业机会的增长和农村人均纯收入的增加使当地潜在二次迁移者外迁的概率明显降低。固定投资的增长和农村消费能力的上升也有类似的作用,不过这个作用与前者相比明显薄弱。城乡消费对比有相对较强的负向作用,说明城市相对消费水平越高,该地区的迁出概率就越低,考虑到本文研究的二次迁移者大部分是从城镇迁出出去的,这是因为城市的生活环境越好,当地人口越有可能留在本地。农村消费增长率对外迁概率有较弱的正向作用,可能的解释是农村消费增长速度较快,使得城镇的相对生活成本增长较缓,因此对潜在迁移者的“推力”作用较小。

模型 3 加入个人因素变量和区域因素变量的交互作用,尽管这个作用对迁出概率的影响统计上显著不为 0,但是影响力却是比较小,而且对模型拟合程度的贡献也不大。

4.3.2 子模型 2: 二次迁移者的回归概率参数估计

这个子模型可使用的数据为 1924 个,同样使用 Logistic 回归来估计。和上个子模型相似,这次也是分三次分别拟合只有个人因素的情况、个人因素加区域因素的情况以及两种因素交互作用对回归概率的影响。如表 3 所示,这个子模型拟合的效果虽然不够理想,但是仍然能够从中得出有意义的结论。

表 3 个人因素和区域环境因素对二次迁移者回归概率的影响

	模型 1	模型 2	模型 3
常数项	3.30**	3.93**	3.94**
个人因素			
年龄	-0.03**	-0.04**	-0.05**
性别	-0.90**	-0.85**	-0.89**
家庭人口总数	-0.08**	-0.09**	-0.09**
受教育程度	-0.22**	-0.29**	-0.25**
迁移原因	-1.19**	-1.09**	-1.07**
迁出地类型 ^①	0.53**	0.47**	0.46**
区域因素			
资本形成总额		+0.00**	+0.00**
城乡消费对比增长率		0.02*	-0.02
就业增长率		0.22**	0.21**
区域总人口		-0.00**	-0.00**
交互项			
性别×城乡对比增长率			0.02
教育×城乡对比增长率			-0.02**
年龄×城乡对比增长率			+0.00*
Nagelkerke R ²	0.214	0.313	0.329

注：* 代表在 0.05 水平上显著；** 代表在 0.01 水平上显著。

进入模型 1 的个人因素变量包括年龄、性别、家庭人口总数、受教育程度、迁移原因、迁出地类型。除了迁出地类型以外，其他变量对回归概率都呈反向关系。从回归分析的结果来看，男性相对女性更愿意回归迁移，而以经济原因发生迁移的人群则更愿意到另外一个地方去闯荡。另外，受教育程度和迁出地类型也会明显的影响迁移者的回归迁移愿望，受教育年限越长，迁移者越有动力迁往其他的省份而不是自己的出生地，迁移前住在居委或街道的人往回迁的意愿高于居住在乡村的人。

进入模型 2 的区域环境变量对迁出概率的影响都比较弱，只有“故乡”就业增长率能够较明显的促进迁移者选择回归的概率。城乡消费对比的增加和该地区资本形成对此有微弱的正面作用，而故乡的人口总量对于外地迁移者选择回归有很弱的阻碍影响。

这里，个人因素和区域环境因素的交互作用对模型的拟合贡献比子模型 1 强，但是它们对回归迁移概率的影响较弱，只有个人受教育程度和城乡居民消费对比增长的交互作用有一定影响，这个影响在统计上显著。

从这个子模型的拟合情况来看，个人因素变量对二次迁移者回归迁移的影响比区域环境因素更大一些。

4.3.3 子模型 3: 再迁移目的地选择概率的参数估计

为了简化模型，我们将迁移目的地选择范围从 31 个省市区缩小到 3 个地带，即中央政府提出西部大开发战略以后新划分的三大地带^②。表 4 显示了总人口和各类迁移人口在三大地带上的分布，子模型 3 考察了再迁移人口对东、中、西三地带目的地选择的影响因素。

① 对迁出地类型的取值做了一些修改，使之成为一个有序数组，从低到高依次表示从乡村到街道。

② 其中东部地带包括北京、天津、河北、辽宁、上海、江苏、浙江、福建、山东、广东、海南 11 个省、直辖市；中部地带包括黑龙江、吉林、山西、安徽、江西、河南、湖北、湖南 8 个省；西部地带包括重庆、四川、贵州、云南、广西、西藏、陕西、甘肃、青海、宁夏、新疆、内蒙古共 12 个省、自治区和直辖市。

表 4 迁移人口现住地三大地带分布 (%)

现住地	总人口	一般迁移	再迁移	回归迁移
东部地带	39.8	74.9	71.6	43.7
中部地带	32.2	9.6	9.3	27.0
西部地带	28.0	15.5	19.0	29.3
合计	100.0	100.0	100.0	100.0

这次参数估计共有 752 个可用样本,使用多分类 Logistic 回归对目的地选择概率进行参数估计。结果拟合的非常好,模型的解释能力很强,如表 5 所示。

表 5 个人因素和区域经济因素对再迁移者目的地选择概率的影响

目的地= 中部地带		目的地= 西部地带	
截距	58.09**	截距	173.41**
个人因素		个人因素	
受教育程度	- 0.08	受教育程度	- 0.54**
迁移原因	0.74	迁移原因	2.35**
区域经济因素		区域经济因素	
Log(从业人员数)	5.69**	Log(从业人员数)	4.24*
居民消费总额增长率	- 45.35**	居民消费总额增长率	- 11.56
Log(居民消费总额)	- 3.55**	Log(居民消费总额)	- 14.84**
Log(资本形成总额)	- 9.79**	Log(资本形成总额)	- 12.44**
对照组:目的地= 东部地带			
Log likelihood = - 107.065		Pseudo R ² = 0.81	

注:* 代表在 0.05 水平上显著;** 代表在 0.01 水平上显著。

表 5 的模型说明了再迁移者选择某个地带作为迁入目的地的概率是如何受个人因素和区域经济因素影响的。可以看出,受教育水平较高的再迁移者,在比较东部和中部、东部和西部地带时,选择迁向东部的概率会高一些,而比较中、西部时则会倾向于迁向中部省区。一般认为,中国经济发展水平和活跃程度,东部优于中部,中部优于西部,即由东向西呈梯度分布。从这个意义上说,受教育水平较高的迁移者在选择迁目的地时更容易受区域之间经济发展差异的影响。或换言之,受教育水平较高的迁移人口在做再迁移决策时,区域经济因素将是其主要考虑因素。

当某个迁移者的再迁移原因为经济原因时,他更有可能选择西部而不是东部地带。对于不同目的地就业岗位的相同增长,再迁移者选择中西部地带的概率会偏高一些。这是因为东部就业岗位已经很多,增加相同数量的就业岗位所增加的百分比很低,这个“利好”信息已难给再迁移者带来强烈的刺激。但是面对相同的消费总额和资本总额增长时,再迁移者将会偏向于选择东部地带,这说明再迁移者相对更看重消费和资本的增长。

另外,从这个子模型还可以看出,区域环境因素对再迁移者的迁移选择影响显著,而且影响力也非常大,可见区域经济发展差异对再迁移决策的影响超过迁移者个人因素的影响。

5 主要结论

长期以来人口迁移是学者们研究的热点问题之一。但是由于受研究资料的限制,使对人口多次迁移这一更复杂的人口迁移现象研究不多。本文在有限数据的基础上,利用嵌套 Logit 模型对 1990 年代后期中国省际人口二次迁移进行了定量考察和分析,得出了一些很有意义的结果。

本文研究发现,和国外研究经验相似,个人的迁移选择不仅受区域环境因素的影响,而且也受个人因素的影响。再迁移者在选择迁入目的地时明显受到潜在目的地经济因素的吸引;回归迁移者受出生地经济因素的影响较小,而是受个人因素的影响较大。但是无论从哪个角度来考察迁移行为,都不能只看到迁入地、迁出地之间经济发展的差异,而应该综合考察区域环境因素和个人因素的影响,

这也是本文开头提及的“迁移悖论”的一个解释。

本文根据嵌套 Logit 模型分析结果显示, 中国省际二次迁移及其影响因素主要表现出以下几个特征:

第一, 二次迁移者受教育程度高于一般迁移者, 非农户口比重也明显高于一般迁移者。因此相比之下, 受教育程度较高、非农业人口的迁移活性更强, 更容易发生二次迁移。回归迁移者男性比重明显超过一般迁移者和再迁移者, 而且他们更有可能选择迁往乡村。

第二, 潜在迁移者对是否迁移的决策, 受个人因素和区域环境因素的双重影响。就个人因素而言, 单身的年轻人更容易选择迁移, 相反年龄较大、“拖家带口”的人不容易迁移。而且, 在中国长期存在的城乡二元结构体系之下, 农业人口的迁移欲望相对更强。就区域环境因素而言, 城乡消费差异对迁移选择的影响相对较大, 而农村收入、消费以及城镇居民消费虽然对迁移决定有影响但是影响较弱。考虑区域因素变量以后, 受教育水平变量对迁移决定的影响明显增强。可能受教育程度较高的人更容易受区域差异的影响, 而且有更强烈的迁移欲望。

第三, 二次迁移者对是否回迁的决策, 主要受个人因素的影响, 而区域环境因素的影响较小。在个人影响因素中, 性别、受教育程度以及迁出地类型的影响较强, 如男性、受教育程度较低的人口、原住地是城市的人口更容易回迁, 而年龄、所在家庭人口数量的影响相对较小。

第四, 再迁移者对迁移目的地选择, 则主要受区域环境因素的影响, 而个人因素的影响较小。再迁移者更愿意选择迁向经济发展水平较高的东部地带, 而且受教育程度较高的迁移人口这种欲望更强烈。

需要指出的是由于本文使用的数据所限, 无法得到迁移者在多次迁移时点具有的可变属性特征, 而是以第五次人口普查时点前最后一次迁移时迁移者所具有的各种特征来替代迁移时点的某些特征, 因此不能准确地考察和分析多次迁移的过程及其影响因素。另外, 人口迁移是一个非常复杂的过程, 影响迁移决定的因素也多种多样、错综复杂, 加之区域环境变量统计的局限, 为筛选模型影响变量带来很大困难, 因此本研究构建的嵌套 Logit 模型对 1990 年代后期中国省际人口二次迁移决策的解释能力不是很强。如何建立更有说服力、解释能力更强的迁移决策模型有待进一步深入研究与探讨。

参考文献:

- 1 王桂新, 刘建波. 1990 年代后期我国省际人口迁移区域模式研究. 市场与人口分析, 2003; 4
- 2 K. Bruce Newbold, Kao- Lee Liaw. 1995. Return and Onward Migrations in Canada, 1976- 1981: An Explanation Based on Personal and Ecological Variables. The Canadian Geographer, No. 1
- 3 Kao- Lee Liaw. 1990. Joint Effects of Personal Factors and Ecological Variables on the Interprovincial Migration Pattern of Young Adults in Canada: A Nested Analysis. Geographical Analysis, Vol. 22, No. 3
- 4 鹿丽华. 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; 2

Determinants of Inter- provincial Migration in China: Nested Logit Models

Abstract: This paper applies three- level nested logit models to the 1% sample data of China's fifth census to examine factors affecting interprovincial migration in China in the late 1990s. It turns out that both personal factors and regional variables are important to migrants' decision: the onward migrants are obviously affected by the regional economy while the return migrants are less sensitive to regional economic differences but their personal factors are rather influential.

Keywords: Migration; Onward Migration; Return Migration; Nested Logit Model

Authors: Liu Jianbo, Wang Guixin, and Wei Xing, Institute of Population Research, Fudan University, Shanghai, 200433.