

中国省际流动人口流入地 宏观选择机制的稳定性

——兼论重力模型的理论扩展与实证检验

周 皓, 刘文博

(北京大学 社会学系, 北京 100871)

【摘要】对流动人口流入地宏观选择机制及其稳定性的讨论有助于深入理解我国省际人口流动的空间模式及未来发展趋势。文章使用2000-2020年间历次人口普查、小普查及相关统计年鉴数据,结合迁移网络理论和新古典经济学宏观理论,在统计检验多种重力模型的基础上,利用分层泊松模型考察社会网络、劳动力规模和地区经济发展差异在省际流动人口流入地选择机制中的作用及其在时间上的稳定性;并在中国情境下对经典的存量重力模型进行理论扩展和实证检验。研究发现:流出地与流入地的迁移社会网络对迁移流的规模与方向起到显著的正向影响,这种作用在历时变化中呈现出极强的稳定性;作为人口迁移流的源头,流出地劳动力人口规模对人口迁移流呈现显著且一致的正向影响;经济因素对人口迁移流的影响具有相对稳定性,地区间的经济发展差异仍然是我国人口迁移流动的重要原因;流出地人口的流动强度对人口迁移流发挥着极强的持续稳定性影响,显现出流出地流动文化在供应端的决定作用。鉴于社会网络在流入地选择机制中的重要作用,传统的存量重力模型应被扩展为社会网络-重力模型;鉴于流出地劳动力规模在省际人口迁移流中的基础作用,我国省际人口迁移流的可持续问题须引起足够的重视。

【关键词】流动人口;流入地选择;宏观选择机制;重力模型;稳定性

【中图分类号】C922 **【文献标识码】**A doi:10.16405/j.cnki.1004-129X.2023.02.007

【文章编号】1004-129X(2023)02-0080-19

【收稿日期】2022-12-02

【基金项目】2019年国家哲学社会科学基金一般项目:中国流动人口调查的比较研究(19BRK039)

【作者简介】周 皓(1972-),男,浙江宁波人,北京大学社会学系教授;

刘文博(1993-),男,河北唐山人,北京大学社会学系博士研究生。

一、引言

第七次人口普查显示中国流动人口规模剧增,但总体上我国省际人口流动模式自20世纪90年代以来保持了相对稳定;^[1-3]且一直表现为强者恒强的特征。^[3-4]流入地选择机制是流动者基于个体特征、结合流入地特征而做出的综合性的理性决策;^[5]宏观机制在其流入地选择过程中占据着主要地



位。在社会经济长足发展、公共服务均等化政策实施多年之后,人口迁移流动模式稳定性的背后是否意味着流入地宏观选择机制的稳定性呢?反过来,这种选择机制的稳定性是否又预示未来我国省际人口迁移流模式的稳定性?会有何发展趋势呢?这是现实给我们提出的理论问题。

重力模型(Gravity model of migration)一直是人口迁移流动宏观选择机制研究的重要基础模型,是解释宏观人口迁移流流向、流量的重要解释模型,它以简约的形式被广泛应用于国际人口迁移、甚至国际贸易研究中,具有广泛而深远的意义。近年来虽然在统计分析方法上有了较大的进展,^[6-9]但从理论上,自杨云彦等^[10-11]和 Fan^[12]先后将社会网络理论引入重力模型,并提出存量模型(Stock model)以后并未有真正的进展,多数研究只是增改自变量予以不同的解释。

对流入地宏观选择机制稳定性的考察需要利用多轮数据进行历时性比较分析。目前国内多数研究使用一次或几次普查的数据讨论迁移流宏观选择机制问题;^{[6][12][13-14]}但关注点是选择机制的变化,而非稳定性,且不同时期的研究选择了不同的变量,难以对相同变量的作用进行一致性检验。对于流入地宏观选择机制稳定性的考察仍需利用多期数据、在相同的模型设置条件下予以进一步检验。

除此以外,当前许多研究通常忽略样本问题。国内对宏观迁移流的研究通常使用5年前常住地标准筛选样本。但该样本可能存在的混合样本属性会导致最终的分析结果可能是多种混合样本的平均效应,而非纯粹的流动人口效应;^[15]且迁移人口与流动人口是两种完全不同社会涵义的迁移行为,^[15]其迁入地的宏观选择机制本身也存在显著差异。^[10]若使用纯粹的户口标准对流动人口进行定义,则会使样本中同时包含存量与流量样本,对于那些在流入地居住了5年以上的存量流动人口,用普查时点(或普查时点前5年、或普查时点与前5年之间的时间中点)的社会经济指标进行因果推断时,会由于变量间时间顺序问题而无法揭示其真正的因果关系。因此,上述两种标准下的流动人口样本都可能存在一定的问题。本文将户口标准下5年内的流量流动人口为分析样本,避免由样本带来的干扰。

基于上述问题,本文将利用2000年以来的三次普查和两次小普查数据,以迁移网络理论和新古典经济学宏观理论为基础,在中国情境之下讨论经典重力模型的理论扩展,并利用实证数据检验这种理论扩展的可行性与稳健性,借以检验中国省际流动人口流入地宏观选择机制的稳定性。

二、文献综述与研究假设

(一)相关迁移理论

Massey等有关人口迁移理论的综述与评论一直是人口迁移流动研究的重要文献基础。^[16]该文综述了当时已有的经济学、社会学、地理学视角下的相关理论。新古典经济学宏观理论将迁移视为二元经济结构下,现代工业部门对农业部门剩余劳动力的吸纳,不同部门在经济发展和工资水平上的差异是影响迁移决策与迁移流持续的关键。在两地间经济差异消失的同时,迁移流也将停滞。由于迁移者总是处于低端劳动力市场,二元劳动力市场理论认为次级劳动力市场对低端劳动者的需求导致了外来移民的进入,进而使迁移流得以持续。在此基础上,社会资本形成、人力资本形成和市场整合这三者的相互促进过程是两国间迁移流持续的重要原因。^[17]

社会学视角之下,移民网络和组织结构是迁移流持续发展的另外两种重要原因。前者是一系列基于亲属、友谊以及同乡关系形成的人际纽带,可以将迁入地移民与迁出地的潜在移民联系起来,通

过减少迁移成本、降低风险,促使产生更多的移民行为,^[18]由此促进两地间迁移流的不断持续。后者则是在迁移社会网络基础上扩展而成的各种组织逐渐在制度上保持稳定,构成了移民可以利用的另一种社会资本,进而帮助移民潮持续发展。^[16]不同的迁移机制在不同时期发挥着支配作用:由收入最大化主导的移民盛行于20世纪70年代;80年代中期分散风险成为大多数移民的主要动力;90年代移民网络则成为迁移的重要原因。^[19]而累积因果理论将迁移流的持续扩展为同时考察流出地与流入地,并认为每一次国际移民的发生都会改变后续移民决策的社会环境,如收入分配、土地分配、农业组织、人力资本的地区分配以及工作的社会意义等,使得后续的移民行为更可能发生。^[20]上述理论从某种意义上可以被视为分别从流入地和流出地两个视角下的讨论与考察。此外,近年来移民政策成为国际迁移人口迁移目的地选择的关键因素,^[21]移民更倾向于进入对其接纳度更高、劳动力市场准入条件更好,以及能获得国籍和永久居留权的国家。^[22]

(二)重力模型及其发展

两地间人口迁移流的研究主要包括两个方面:一是对迁移流本身的研究(如迁移流流入-流出-净迁移流等流量和效应的分解);二是对迁移流的建模,如利用距离、人口及流出地与流入地的各种因素对迁移流的解释。^[6]但不论哪个方面,重力模型是分析基础。经典的重力模型认为两地间人口迁移流量(规模)与迁入地和迁出地的人口规模成正比,与两地的距离成反比:

$$M_{ij} = \frac{kP_i P_j}{D_{ij}^s} \text{ 或者: } \ln(M_{ij}) = a_0 + a_1 P_i + a_2 P_j + a_3 D_{ij} + \mu_{ij} \quad (1)$$

M_{ij} 表示流出地*i*和流入地*j*之间的迁移流量; P_i 和 P_j 分别为流出地和流入地的人口规模; D_{ij}^s 表示两地之间的距离, s 表示距离的形式; k 为常数。

该模型由于与理论的一致性、简约的形式与估计、较好的拟合度而被广泛应用于地区间或国际人口迁移研究,甚至被称为迁移研究中的“老龄巨星”(Ageing superstar)。^[8]已有许多研究很好地综述了重力模型在人口迁移流动研究中的应用和问题。^{[6-8][22]}本文不再赘述。

近年来,对重力模型的改进可以归结为理论解释能力与统计方法两个方面的改进。理论解释的改进主要是基于推拉理论对重力模型进行拓展,通过将一系列人口和经济社会指标作为人口迁移的驱动因素,提高模型对人口迁移过程的解释能力。^[23]其中,移民社会网络能够为后续者提供迁入地信息、降低迁移风险与迁移成本,且移民在流入地具有集聚性,流出地与流入地的家庭成员间的联系在迁移决策中起到了比其他因素更为重要的作用,^[24]因此,应将现有的迁移流与该迁移流历史相连接,以使模型更有解释能力,通用的方法是在模型中加入存量移民变量(可用前一期从流出地到流入地的迁移流规模或比例);否则模型结果可能会高估其他变量的真实作用。^[12]这说明理论视角上需要将迁移社会网络与重力模型相结合;在模型形式上,至少需要将上式转换为:

$$\ln(M_{ij}) = a_0 + a_1 P_i + a_2 P_j + a_3 D_{ij} + a_4 STOCK_{ij} + \mu_{ij} \quad (2)$$

$STOCK_{ij}$ 通常被表示为从*i*地流入*j*地的流量占*i*地总流出规模的比例。

此外,还可借鉴国际贸易理论,将国家是否相邻、国家间以往的殖民历史以及语言等^[25]因素扩充到自变量中,以增强模型的解释能力。

统计方法的改进主要通过解决传统重力模型中存在的单边作用问题、对数形式的影响及异方差性,^[25]以使统计结果更加稳健,^{[7-9][22-23][25-26]}并加以扩展(PPML-rdhf和结构分层PPML等)。或者通过计量手段分解模型中被解释变量的效应,分析模型的误差来源。^{[6][23]}其中,单边作用问题系指在仅考



虑流出地或流入地某一侧的变量作用时会由于存在整群效应(Clustering)而导致估计量标准误的偏差。^[27-28]同时,迁移流并非仅仅是仅仅单边或双边的问題,还存在着内部的空间相关性和多个迁移流之间不同形式的相关性问題。^[9]增加固定效应(即PPML-rdhf等模型)、分层泊松模型^[6]或扩展的交互分类分层模型^[9]和空间滤波方法^[23]等可以控制并解决这种偏差问題。此外,对数线性形式要求所有案例(迁移流)都具有非零值;但并非所有的双边关系都会存在一定量的迁移流,因此,SOLS回归^[29-30]和分层泊松模型^{[6][26]}被逐步引入。

上述对重力模型改进的两种路径是相辅相成的。但近年来由于迁移理论本身并未能有较深刻的突破,因此,增强重力模型的理论解释能力只是相对简单地增加或改变解释变量而无法做到真正的突破。相反,在数据可获得性的基础上,重力模型的统计方法得到了相对较好的改进与发展。

(三)流入地宏观选择机制的中国研究

有关中国省际人口迁移流动的流向、分布等问題已有众多研究成果。^{[14][31-33]}总体来看,地理学更关注空间分布与流场等(包括空间相关性等),人口学更关心结构与规模问題;而经济学则更关心迁移与经济的关系。在此不再赘述。本文将仅讨论流入地宏观选择机制的中国研究。

国内有关流动人口流入地选择机制的研究通常以推拉理论为分析框架,考察流出地、流入地及中间因素三个维度。在研究不断深入的过程中,研究关注点日趋全面,从仅关注经济因素到综合考察经济、社会、地理因素的影响,研究方法也不断改进。

早期研究主要聚焦经济层面,认为人口流动主要为了谋求更多经济机会、寻找更好的发展空间和更高的经济回报。^[13]不同地区在经济规模(以GDP或GNP为指标)、外商直接投资、^{[10][13]}产业结构、固定资产投资、工资水平^[34]和以失业率或城镇登记失业率为代表的就业机会等经济因素是影响流入地选择的重要因素。

除就业与工资外,劳动力还会为了更好的公共服务而流动。^[34]这些社会因素包括地区公共财政支出、基础教育资源和医疗服务资源,^[34]地区的社会资本、建成区绿化覆盖率、每百人公共图书馆藏书、医院卫生院数量、公共服务设施等。从流出地看,农业人口比重(农业劳动力占总劳动力的比重)和人地比(农业劳动力人口与耕地面积之比)^{[6][10]}等流出地的人口结构与自然环境等同样会影响人口的流出。

两地间的迁移网络可降低迁移的成本和风险,带来更高的流迁意愿。^[35]杨云彦等首次利用中国数据,结合迁移网络理论,在重力模型中加入1982-1986年间的迁移量,检验了社会网络在迁移流向中的作用。同时,该文还检验了户口迁移和非正式迁移(人口流动)之间的不同作用机制。^[10-11]Fan在总结国外研究成果的基础上提出了存量模型(Stock model),^[12]且在后续研究中被广泛引用;^{[6][36-37]}也有研究认为由于迁移网络的存在,后续的迁移人口通常会迁往前期移民的迁入地,因此,省际迁移流量存在着历史时间相关性或路径依赖。^[37]迁移社会网络对于人口流动网络的演化有着极为重要的影响且影响力在不断增强,验证了人口流动网络的积累性效应。^[38-39]

尽管空间距离等恒定因素的影响相对弱化,^[40]但基于重力模型的流出地与流入地之间的空间距离仍被认为是最重要的中间障碍因素之一。^[41]在具体研究中,省际人口流动的距离通常会被定义为两省省会城市间的距离,但却可以有不同的操作化,如空间地理直线距离、交通距离(多数使用铁路交通距离,也有使用铁路距离与高速公路距离的平均值^[37])以及由经济差异引出的经济距离等;同时也可采用不同的数学形式(如加幂次方、或取对数等)。近来的研究发现距离因素的阻碍作用在逐渐

减弱,^[42]而两地间在制度、文化等方面的邻近性则会给流入地选择带来正向影响。^{[41][43]}

上述总结体现出我国人口流动的流入地选择机制研究正从纯粹以经济为主导、忽略其他因素,转变为以经济为主导、关注多重社会属性的共同作用,“预期收入”的涵义应该扩展为同时包括社会因素与经济收入的“预期回报”。^[5]

(四)流动人口流入地选择机制的历时变化与稳定性

对于流动决策机制的变化问题,有些研究使用两次普查数据进行讨论,^{[4][12][42]}有些则使用了三次普查的数据;^[6]结果发现迁入地城镇收入水平等经济因素、^[4]就业结构和人文发展指数等吸引作用明显增强,^[42]而空间距离等恒定因素的影响则相对弱化。^{[4][42]}许多研究认为东部沿海地区的经济快速增长是影响人口迁移流向的重要原因。^{[27][44-47]}其中,省际迁移流量的增加中有 62.28% 是由于解释变量取值的变化,而另外 37.72% 则是由于模型系数变化所致。这表明流入地的人口结构与社会经济的变化占据着主导地位。^[6]

从历时角度看,杨云彦等对 1990 年四普的研究,^[10]Fan 对 1990 年四普和 2000 年五普的研究,^[12]Shen 对 1990、2000 和 2010 年普查的研究,^[6]虽然每个研究只是利用一次或几次普查数据,但将这些研究按照时间顺序串接起来,似乎可以被视为中国省际人口迁移流动流入地宏观选择机制,特别是迁移网络理论与重力模型在中国实证应用的一致性检验。但由于各研究采用了不同的控制变量,可能导致不同的结果。这就需要在相同的模型设置条件下(包括方法和控制变量等),检验研究自变量的作用是否会在不同时期的结果中产生差异或变化。

如果再比较上述这些研究所对应的宏观因素的作用(即模型中各变量的系数及显著性等)可以发现除了重力模型中的三个基本变量以外,其他变量如 GDP、文盲率、流出地或流入地的人均工资、迁出地人口增长率等其他各种指标均未表现出完全的一致性。这种不一致性既可能是因为时代环境的改变(如发展水平的差异、甚至发展的阶段性差异),也可能是因为与时代环境相关联的影响机制的改变。正如 Shen 在其文中指出相同模型对于 1985-1990 年和 1995-2000 年样本的解释能力要强于 2005-2010 年的样本,^[6]这从另一方面说明进入 21 世纪以来,中国迁移模式与过程可能发生了一定的变化,因而还需要进一步加强研究。当然,这种一致性检验还需要同时考虑模型设定误差问题,即如何平衡与评价简约模型的稳定性与一致性和扩展模型中扩展变量的不一致问题。毕竟模型设定可能会带来分析结果的偏差。

(五)简单评述

上述文献回顾表明我国省际人口迁移流的相关研究正逐渐转向更为深入的机制分析,研究内容不断扩展与深入,研究方法不断改进。但目前仍有以下问题值得思考:

其一,重力模型的理论扩展。尽管 Greenwood 已经指出如果不考虑迁移流的历史,并将其与社会网络理论相结合,那么估计结果可能会是有偏的。但是,到目前为止,国内外许多研究在利用重力模型讨论人口迁移流时,仍然未能关注迁移流历史这个变量,而只是以加入各种社会经济变量、提高理论解释能力为主。即使部分研究加入了迁移流历史的变量,通常或是从流出地视角出发,以前一期从 A 省到 B 省的流动人口占 A 省流出人口的比例为基础,而忽略了流动者在流入地中建立的社会网络;或者仅从流入地考虑而忽略流出地,未能两者兼顾。由此忽略重要变量而导致分析结果不能满足无偏性和有效性。当然,这一问题既与迁移社会网络理论和流动人口在流入地的集聚相关,也与流出地与流入地的推拉及其在模型中的设置有关。



其二,模型设定问题。人口迁移流动都会涉及流出地和流入地,正因如此,推拉理论被视为各种人口迁移流动理论的基础,正如重力模型分别从流出地与流入地两方面考虑人口变量。在简约模型的基础上,研究者通常希望能够增加自变量以提高理论解释能力。但模型设定就会碰到两个典型的问题:一是用于描述流入地和流出地各种社会经济特征的自变量是否必须相同?即是否必须形成对称模型;二是如果采用相同的自变量,那么,在对数化处理时,流入地与流出地的相同变量是各自以对数形式加入,还是以两个原始变量相除(以表示两地之间的差异)后再取对数形式加入^①?已有研究有各种不同的处理方法。单边与双边问题涉及检验迁出地推力因子和迁入地拉力因子的相对重要性。通常人们倾向于认为移民对迁出地的情况更加了解,因此迁出地的作用更大。但多项研究发现迁入地和迁出地经济状况对迁移的影响是不对称的,一般拉力因子所起的作用更强,这与以往的推测正好相反。关于这种不对称性,最常见的解释是迁移成本收益和移民选择性,^[10]因此,单边与双边问题既是方法问题,更是理论视角问题。不同的模型设定对应着不同的理论解释。

目前分层泊松模型逐步得到重视与发展,有研究将流出地省份作为层-2分析单位,那么,是否可以将流入地作为层-2分析单位呢?两种层-2分析单位的选择是否会影响分析结果呢?而且,在以流出省为层-2分析单位时,流出地的特征变量应该被视为层-2单位的特征,而层-1模型中就无法包括流出地的特征变量;反之亦然。那么,是否可以认为不同的层-2分析单位决定了不同的分析视角?当然,有研究引入了扩展的交互分类泊松分层模型,并讨论人口迁移流之间的各种关系以及这种方法在估计结果上的优越性。^[9]从模型的拟合结果看确实有一定的提高,但模型的估计系数和相应的标准误的改进则相对有限,且其模型设定仍然是双边的。从这个角度看,扩展的交互分类泊松分层模型是否还有可能被扩展呢?

其三,样本问题。除上述统计方法以外,分析样本同样也是重要的分析基础。我国的迁移流动人口包含了迁移人口与流动人口这两种有着完全不同社会含义的空间移动行为人口,且两类人口在迁移原因^[46]与宏观迁移选择机制^[10]等方面都存在着显著的差异。因此,在分析过程中应该注意区分这两类人口。以往的许多研究通常直接使用5年前常住地样本,而忽略这种样本的混合属性,^[15]进而导致分析结果是这些混合属性样本的真实原因的平均结果。近年来对我国人口迁移流动的讨论更多以流动人口代替所有的迁移流动人口,而忽略了迁移人口。这种忽略既会严重低估中国人口迁移流动的实际水平,也无法全面揭示我国人口迁移流动的真实原因(不论是宏观的还是微观的)。

其四,迁移流是否能够以及何以能够持续。虽然直至七普我国迁移流动人口的规模仍然在持续增加,但规模的增加源自总人口规模的增加与迁移率的提高,特别是年龄别流动率峰值的大幅提高及其后年龄别流动率的相对提高。^[48]而人口迁移流动的持续发展关键在于劳动年龄人口,特别是迁移流动率最强的20至30岁的年轻人口。在我国人口生育水平持续低迷多年以后,人口年龄结构老化的同时,劳动力年龄结构也在同步老化。而且,年轻一代与老一代的流动人口在迁移原因与结果等各方面都存在着极大的差异。那么,在结构老化、社会变迁与代际交替等多重影响之下,未来我国的人口迁移流动是否能够持续以及何以能够持续将是一个战略发展的问题。它不仅牵涉经济相

① 相除后加入模型与各自加入模型,一方面理论含义不同,即相除表示流入地与流出地在社会经济发展水平上的差异,以差异解释迁移流;另一方面则是模型含义或统计含义,即相除表示对数化以后,流出地与流入地的社会经济变量有相同的作用。

对发达地区的未来持续发展,而且也涉及现在作为流出地的中西部地区和农村地区的未来持续发展问题。

其五,迁移流的扩散问题。七普数据显示人口迁移流动的流入地和流出地同时存在着扩散的情况。从流入地看,尽管我国人口迁移流动的流入地以三大圈加一小圈为主,且浙江、江苏等东南省份和京津冀地区自2005年以来就表现为重要的流入地,但七普数据表明这些地区的流入人口规模进一步增加,且浙江、江苏等省份的流入规模和吸引力等有较大提升并有可能超过广东。从流出地看,各省流出率提高的同时,流出地亦不再集中于传统的几个流出省份,而呈现出逐步漫延扩散的趋势。而且海外华人在地理空间上已显现出逐步扩散的过程。这是否意味着人口的流动性也会如生育行为一样会呈现出扩散的趋势?^[49]这种扩散性是否会随着社会经济的发展,以及人口迁移流动程度的加剧而呈现出更为强烈的扩散机制?是否会像中国式的家庭或家庭关系那样呈现出圈层泛溢之势呢?其机制与机理问题尚需深入思考与检验。

上述这些理论、方法与现实问题都需要从各种不同视角予以讨论。在呼唤地理空间视角回归的同时,多学科多视角的讨论与研究才能促进对人口迁移流动空间分布的理解与把握。本文将仅针对所提及的部分问题予以讨论。

(六)研究假设

基于上述内容,本文将结合社会网络理论与重力模型,扩展存量模型(Stock model);并利用多轮普查数据和分层泊松模型,针对纯粹的省际流动人口,在统计控制多种变量的条件下,检验迁移网络在宏观迁移流中的持续影响作用;并将在此基础上,检验省际流动人口流入地宏观选择机制在时间上的稳定性。

首先,存量模型从流出地视角表明迁移网络的重要性。但迁移人口有集聚效应,即在流入地,后续的迁移流动人口总是首先集聚到已有迁移流动人口相对较多的地方,然后再逐步扩散。因此,流入地流动人口的集聚性所带来的社会网络同样会影响迁移流的方向与流量。这种集聚性从理论上要求将重力模型与流入地迁移社会网络相关联,是对存量模型的一种扩展;方法上则是基于对称模型的变量扩展。因此,与存量模型相对应,本文认为前期某省流动人口在流入地的集聚效应(迁移社会网络)对后续迁移流方向与流量具有重要的影响作用。这是假设一。

其次,重力模型将人口总规模视为基础变量,两地间的流量与两地的人口规模成正比。在流出地供给角度,流动人口的年龄别流动率表明劳动年龄人口规模才是流动人口的重要来源,而并非是总人口。虽然劳动年龄人口的规模与总人口规模在中国情境之下是基本一致的,但在实证分析中,流出地人口规模更确切的操作化定义应该是流出地的劳动年龄人口规模。因此,本文的第二个假设是:流出省份的劳动年龄人口规模可以替代流出地总人口规模,正向影响省际人口迁移流规模。

第三,尽管随着时间推移、全国与各省区的经济发展,我国的人口流动模式会产生一定的变化如Shen所提及的2010年六普所体现的新变化以及上文提及的扩散性等,^[6]且社会经济“预期回报”而非纯粹的“预期收入”成为吸引流动人口的重要因素,^[5]但总体上,省区间经济发展水平上的差异仍然是省际流动人口的重要流动原因。因此,本文的第三个假设是:经济发展的地区差异在当前我国省际人口迁移流动中仍然起到正向的影响作用。

除上述三个假设外,本文的另一个目标是检验宏观选择机制的稳定性,本文虽然未专门列出假设,但后续分析中会从历时维度去检验这些宏观选择机制的稳定性。上述三个假设所对应的稳定性



既是指在控制其他可能变量条件下的稳定作用,同时也指同一因素在历次普查或小普查数据所对应时间段上的作用稳定性检验^①。前者的检验需要通过各模型增加控制变量予以完成,而后者则需要利用多次普查或小普查数据进行检验。

三、数据、变量与方法

(一)数据来源

本文将利用2000年五普、2010年六普和2020年七普数据以及2005年和2015年的两次1%人口抽样调查(简称小普查)数据,共五期数据。相关的社会经济数据均源自《中国统计年鉴》。考虑因果关系的时间顺序,本文将选择普查时点与普查前5年的中间年份所对应的社会经济指标,如2015-2020年的模型(即2020年模型)将使用2017年的经济社会变量,2010-2015年的模型(即2015年模型)将使用2012年的经济社会变量,其余年份以此类推。使用中间年份的好处在于,类似于年中时点,它是一个平均效应,并将时点指标转换为类似于时期指标。

(二)变量测量与描述

1. 因变量

本研究的因变量是省际人口迁移流的流量。首先,本文的流动人口系指:现居住在本地、但户口登记地在外省,且已离开户口登记地半年以上、5年及以内的人口。这个定义与通常的省际流动人口的定义之间最大的区别在于时间。对于流动时间超过5年的流动人口,不论是选择普查年份,还是普查时点5年前的地区社会经济特征来分析其流入地选择机制,都会导致因果关系推断中的原因与结果在时间顺序上的颠倒。这也是本文选择5年内新增流动人口(即相当于流量流动人口)的主要原因。其次,为保证样本的纯粹性,在数据处理中以户口标准加时间标准挑选出相应的流动人口样本;并据此计算得到除香港、澳门、台湾之外的我国31个省级行政区划单位之间的迁出-迁入矩阵(OD矩阵),剔除流入与流出相同省份的组合后将矩阵向量化,每年共有930个流出-流入组合(由于西藏数据的缺失,因此2000年时仅有900个案例);5次普查合计4650个案例。但在建模时仍以历年的930个案例为分析样本。同时,为检验社会网络-存量模型,本文将前一期的OD矩阵作为后一期的解释变量,为此,本文还计算了1995年小普查的迁移矩阵用于2000年五普的分析模型。当然,由于小普查的样本规模本身就相对较小,可能会导致其OD矩阵中有部分单元格的流动人口规模相对较小,甚至为0。这是需要注意的。

2. 核心自变量

本研究的核心自变量包括经典重力模型的三要素(流出地人口规模、流入地人口规模、两地间的空间距离)和流动人口的社会网络。流入省的人口规模利用常住人口总量进行测量,流出省则以劳动力规模代替总人口,以此反映劳动力规模对人口迁移流的影响。流入与流出的省际距离将采用两省省会之间地理坐标的直线距离进行测量。迁移社会网络将从流出地与流入地两个角度分别测量:流出省的社会网络,也即流出地的流动人口存量,参考已有文献,^{[6][10][12]}将其定义为:从*i*省到*j*省的流量占*i*省流出人口总量的比例。流入地社会网络是指从*i*省到*j*省的流动人口占*j*省流动人口总量

^① 本文的稳定是从统计角度看,同一变量的系数显著性和作用方向在不同年份的分析结果中保持一致,并不是指系数的绝对值大小,且本文无意比较其系数所表示的作用大小。



的比例。

3. 控制变量

本研究的控制变量包括流入省与流出省的经济差异及社会发展特征。由于不同的社会经济特征对流出与流入的影响作用并不完全相同,本文采用了非对称模型。

表1 变量描述性统计(均值/百分比)

变量	2000	2005	2010	2015	2020
因变量					
流动人口规模	30.219 (146.674)	96.974 (457.361)	174.844 (592.186)	50.259 (160.220)	44.009 (118.269)
自变量					
流入地总人口对数	17.188 (0.904)	17.237 (0.889)	17.267 (0.869)	17.316 (0.846)	17.350 (0.835)
流出地劳动力人口对数	16.800 (0.914)	16.892 (0.890)	16.950 (0.869)	17.021 (0.845)	17.026 (0.828)
省际距离对数	7.067 (0.613)	7.067 (0.613)	7.067 (0.613)	7.067 (0.613)	7.067 (0.613)
流出地社会网络(%)	3.122 (6.592)	3.226 (6.881)	3.226 (6.986)	3.226 (6.277)	3.226 (5.663)
流入地社会网络(%)	3.122 (5.920)	3.226 (5.393)	3.226 (5.580)	3.226 (4.787)	3.226 (4.598)
GDP差异	1.317 (1.151)	1.371 (1.298)	1.306 (1.077)	1.181 (0.720)	1.163 (0.677)
第三产业比例差异	1.025 (0.230)	1.020 (0.207)	1.028 (0.251)	1.036 (0.288)	1.019 (0.203)
工资差异	1.058 (0.367)	1.084 (0.456)	1.071 (0.418)	1.040 (0.306)	1.045 (0.323)
登记失业率差异	1.173 (0.793)	1.063 (0.423)	1.032 (0.274)	1.057 (0.387)	1.052 (0.357)
流入地人均教育经费(元)	243.590 (184.839)	445.688 (402.416)	748.135 (397.566)	1 803.461 (604.218)	2 845.289 (926.012)
流入地工伤保险参保率(%)	4.645 (9.532)	4.843 (6.143)	10.206 (10.009)	14.152 (9.255)	16.626 (9.986)
流出地人口流出率(%)	23.648 (14.464)	3.463 (2.010)	38.430 (21.250)	5.875 (2.226)	36.988 (11.805)
是否邻省(是=1)	0.142 (0.349)	0.142 (0.349)	0.142 (0.349)	0.142 (0.349)	0.142 (0.349)

资料来源:根据历年《中国统计年鉴》整理得到。

注:括号内为标准误;流出地或流入地社会网络变量由于四舍五入的关系而呈现相同的数据,其实不论是均值还是方差在历年都有一定的差异且内部分布也有变化。



在经济特征方面,本文主要考察流入与流出地经济发展差异对迁移流的影响,选择地区生产总值(GDP)、产业结构、平均工资水平和登记失业率四个指标。其中,产业结构被定义为第三产业所占比例,由第三产业增加值/地区生产总值计算得出;工资水平为城镇非私营单位就业人员平均工资;登记失业率为城镇登记失业人员的失业率。地区间经济发展差异通过流入省与流出省指标相除得到。

社会特征方面,通过人均教育经费和工伤保险参保率考察流入地的社会发展水平对流动人口的吸引作用。人均教育经费=国家财政性教育经费/常住人口总量;工伤保险参保率=年末参加工伤保险人数/常住人口总量。在流出省特征方面,本文使用前一期普查中流出省的总流出率为测量变量,以考察各省的流出强度(流动性或迁移流动的文化习惯等)对后续迁移流的影响。此外,考虑空间邻近性的影响作用,模型还加入了“是否邻省”变量,即流入省与流出省在地理边界上是否相邻。

从上述变量选择可以看出本文对流出地与流入地特征的考察各有侧重:流入地侧重于社会经济民生等方面;流出地则侧重于经济和流动性两个方面。各变量的描述统计结果如表1所示。

(三)分析方法

本文将基于Shen所提的分层泊松模型,^[6]从流出地与流入地两个方面考察。该模型至少可以部分解决因共同的流出地或流入地而带来的迁移流相关问题。尽管扩展的交互分类分层模型可以得到更为精确的标准误估计量,^[9]但改进后的结果并未有实质性的改变。从模型简约的角度,本文仍然选择了分层泊松模型。

建模过程:首先依据重力模型,以流入地的总人口规模、流出地的劳动力规模和距离为自变量建立模型;其次,在此基础上加入流出地社会网络变量;第三步,加入流入地社会网络变量;第四步,再加入相同的社会经济变量,以流入省与流出省的社会经济变量之比值作为自变量,构建对称模型,以便在控制其他变量条件下检验社会网络-存量模型的适用性;第五步,进一步加入流入省与流出省的特征变量,构建非对称模型,作进一步的检验;最后考察历年模型之间的异同,检验宏观选择机制的稳定性。

层-2分析单位可选择流出地或流入地,但不同的层-2分析单位将得到不同的估计结果,因此,本文将分别以流出地和流入地为层-2分析单位建立模型。其实以流入地为层-2分析单位的模型是以流入地为视角,所有流入地特征应该被视为层-2变量而无法加入层-1中。反之亦然。为了能够展示两种不同视角下的差异,本文的模型中加入相同的自变量,从另一方面强调研究视角的重要性。某种意义上这些比较可以被视为稳健性检验。

四、分析结果

限于篇幅,本文将只列出最终模型结果,而不再列出详细的分析过程。最终结果如表2和表3所示。其中表2是以流出地为层-2分析单位的结果;表3则是以流入地为层-2分析单位的结果。

(一)模型整体检验

首先需要检验分层模型的必要性,即检验分层泊松模型与泊松模型之间是否具有显著性差异。表中最后一行的 $LR\ test\ chi2_c$ 结果表明分层泊松模型与一般泊松模型之间存在着显著的差异。因此分层模型是完全有必要的。其次,模型的整体检验表明至少有一个自变量是显著地不等于零。但本文中的模型检验更关心的是模型中的自变量是否对迁移流规模有显著的影响作用,而不关心这些因素对迁移流规模的预测效果。这与已有研究的目的有所不同。



表2 以流出地为层-2分析单位的分层泊松模型结果

	2000	2005	2010	2015	2020
流入地总人口对数	-0.377 7*** (0.024 8)	-0.535 5*** (0.012 4)	0.302 6*** (0.009 9)	0.378 9*** (0.026 7)	0.816 4*** (0.025 8)
流出地劳动力人口对数	0.994 9*** (0.133 7)	1.059 8*** (0.102 9)	0.842 2*** (0.133 8)	0.971 2*** (0.095 2)	0.983 9*** (0.130 4)
省际距离对数	-0.187 7*** (0.017 7)	-0.637 8*** (0.010 9)	-0.385 7*** (0.007 7)	-0.289 7*** (0.014 2)	-0.308 5*** (0.014 0)
流出地社会网络	0.036 0*** (0.000 9)	0.037 4*** (0.000 4)	0.041 8*** (0.000 3)	0.039 2*** (0.000 6)	0.034 1*** (0.000 7)
流入地社会网络	0.048 1*** (0.001 1)	0.034 1*** (0.000 7)	0.039 2*** (0.000 5)	0.056 4*** (0.001 0)	0.047 8*** (0.001 1)
GDP差异	0.167 7*** (0.012 3)	0.146 8*** (0.005 5)	0.027 5*** (0.006 0)	0.419 0*** (0.019 2)	0.071 0*** (0.017 4)
第三产业比例差异	-1.189 9*** (0.105 4)	-1.215 2*** (0.072 9)	-1.511 2*** (0.022 5)	-0.574 7*** (0.043 6)	-3.186 2*** (0.066 4)
工资差异	24.351 5*** (0.397 6)	22.184 0*** (0.310 0)	11.820 1*** (0.277 1)	6.633 6*** (0.482 4)	-0.564 1* (0.228 0)
登记失业率差异	0.018 1 (0.030 7)	-0.324 9*** (0.025 9)	0.759 9*** (0.032 6)	0.521 6*** (0.048 9)	0.503 8*** (0.030 8)
流入地人均教育经费	0.001 1*** (0.000 1)	-0.000 3*** (0.000 0)	0.002 8*** (0.000 0)	0.000 3*** (0.000 0)	0.000 5*** (0.000 0)
流入地工伤保险参保率	-0.017 1*** (0.001 2)	0.050 4*** (0.001 2)	-0.046 8*** (0.000 9)	0.017 9*** (0.001 7)	0.068 6*** (0.001 2)
流出地人口流出率	0.021 0* (0.008 3)	0.066 4 (0.044 5)	0.019 9*** (0.005 5)	0.041 8 (0.035 5)	0.009 9 (0.009 1)
是否邻省(是=1)	0.144 8*** (0.024 9)	-0.200 3*** (0.013 5)	-0.036 8*** (0.009 3)	-0.058 6*** (0.017 6)	0.028 4 (0.018 9)
常数项	-8.204 6*** (2.246 4)	-1.138 2 (1.695 6)	-15.248 1*** (2.219 2)	-20.364 2*** (1.672 1)	-26.336 4*** (2.290 8)
<i>lvar</i> (<i>_cons</i> [<i>orig</i>])	0.396 1*** (0.102 5)	0.203 3*** (0.052 2)	0.369 6*** (0.094 2)	0.176 4*** (0.046 2)	0.350 5*** (0.090 4)
<i>N</i>	900	930	930	930	930
<i>LL</i>	-9 381.243 7	-20 172.060 4	-21 091.964 9	-8 951.979 0	-6 355.358 7
<i>chi</i> ²	52 330.602 9	211 090.699 6	284 458.406 6	76 342.040 2	59 547.412 5
<i>BIC</i>	18 864.523 3	40 446.648 7	42 286.457 6	18 006.485 7	12 813.245 1
<i>chi</i> ² _c	5 512.724 8	10 247.587 6	27 014.456 8	4 196.820 6	4 006.412 2

注:括号内为标准误;*P<0.05,**P<0.01,***P<0.001。



表3 以流入地为层-2分析单位的分层泊松模型结果

	2000	2005	2010	2015	2020
流入地总人口对数	0.340 5 (0.188 2)	0.160 0 (0.156 2)	0.007 6 (0.150 3)	0.455 2 (0.402 3)	0.422 3** (0.135 4)
流出地劳动力人口对数	0.783 8*** (0.016 7)	0.760 3*** (0.009 8)	0.747 5*** (0.006 5)	0.816 9*** (0.011 9)	0.685 7*** (0.012 0)
省际距离对数	-0.873 8*** (0.018 1)	-0.507 9*** (0.010 1)	-0.497 0*** (0.007 3)	-0.288 6*** (0.014 2)	-0.421 4*** (0.014 0)
流出地社会网络	0.000 9 (0.000 5)	0.017 7*** (0.000 3)	0.021 8*** (0.000 2)	0.023 4*** (0.000 5)	0.017 1*** (0.000 5)
流入地社会网络	0.055 1*** (0.001 0)	0.052 6*** (0.000 6)	0.047 8*** (0.000 4)	0.066 3*** (0.001 0)	0.054 1*** (0.001 0)
GDP 差异	0.065 1*** (0.010 5)	0.040 5*** (0.004 1)	0.120 9*** (0.004 1)	0.281 0*** (0.015 3)	0.117 0*** (0.015 2)
第三产业比例差异	0.307 3*** (0.060 4)	1.091 2*** (0.055 3)	0.770 3*** (0.024 9)	-0.612 4*** (0.039 3)	-0.550 8*** (0.054 2)
工资差异	9.270 1*** (0.698 2)	7.492 5*** (0.475 8)	15.858 0*** (0.400 5)	6.197 0*** (0.612 1)	5.884 7*** (0.407 3)
登记失业率差异	0.012 5 (0.022 4)	-0.280 4*** (0.022 6)	-0.794 2*** (0.023 8)	0.191 4*** (0.023 7)	-0.239 8*** (0.023 1)
流入地人均教育经费	0.003 0* (0.001 2)	-0.000 1 (0.000 5)	-0.001 2 (0.000 9)	0.000 2 (0.000 8)	0.000 0 (0.000 2)
流入地工伤保险参保率	-0.013 1 (0.023 1)	0.106 4** (0.033 3)	0.053 1 (0.033 2)	0.032 9 (0.039 3)	0.031 5** (0.010 7)
流出地人口流出率	0.018 9*** (0.000 5)	0.083 1*** (0.002 5)	0.007 1*** (0.000 2)	0.071 8*** (0.003 1)	0.018 0*** (0.000 5)
是否邻省(是=1)	-0.128 2*** (0.023 6)	-0.205 4*** (0.012 0)	-0.096 5*** (0.008 6)	-0.137 2*** (0.016 9)	-0.045 3* (0.017 9)
常数项	-13.296 5*** (3.348 4)	-11.307 1*** (2.745 3)	-6.309 6* (2.840 5)	-18.623 2* (7.745 4)	-14.360 4*** (2.611 5)
<i>lvar</i> (<i>_cons</i> [<i>des</i>])	0.577 1*** (0.149 9)	0.543 3*** (0.138 5)	0.286 7*** (0.072 9)	0.880 9 (.) ⁺	0.131 8*** (0.033 9)
<i>N</i>	900	930	930	930	930
<i>LL</i>	-6 925.189 5	-13 497.555 8	-20 897.537 7	-8 743.300 1	-6 157.887 7
<i>chi</i> ²	39 624.008 7	124 369.734 6	209 287.674 4	61 224.475 4	42 091.193 1
<i>BIC</i>	13 952.415 0	27 097.639 4	41 897.603 3	17 582.292 8	12 418.303 3
<i>chi</i> ² _{<i>c</i>}	10 424.833 2	23 596.596 9	27 403.311 1	4 614.178 4	4 401.354 1

注:括号内为标准误;* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$;⁺此模型最终未收敛,因此未给出标准误。

(二)重力模型三要素的影响

流出地的劳动力规模和距离对省际迁移流的影响与重力模型中是完全一致的。表现为流出地

劳动力规模越大,迁移流的规模也就越大;距离越远,迁移流的规模相对越小。这说明流出地的劳动力规模可能是真正影响迁移流规模的原因之一。但流入地总人口规模在模型中却表现出很有意思的情况。在以流出地为层-2分析单位时,2000年五普和2005年小普查的两个模型中,该变量呈现出显著的负向作用;在2010年六普及其后的两次普查中,它呈现出显著的正向作用,与传统重力模型的结果是一致的。在以流入地为层-2分析单位时,前四次普查的模型中该变量均不显著,仅在2020年七普中呈现出显著的正向作用。如果再比较仅包含重力模型三要素的模型和再加入社会网络变量的模型,流入地总人口规模都呈现出显著的正向作用。只有在加入控制变量以后才会发生变化。这说明在不控制其他变量的情况下,流入地总人口规模确实起着重要的作用;但加入社会经济控制变量后,其作用不再是正向的,即地区间社会经济发展水平的较大差异才可能是真正促进人口迁移流动的重要原因;即便是人口规模相对较小的流入地,若其经济发展水平相对较高,也能够吸引相对更大的迁移流。

(三) 社会网络的影响

首先,除了2000年流入地模型中的流出地社会网络变量呈现出不显著的正向作用,在其余各年份的各模型中(包括未列出的各个阶段性模型)都呈现出显著的正向作用。以流出地模型中2020年样本的估计结果为例,流出地社会网络变量的系数为0.034 1,表示流出地社会网络每增加一个单位,相应的迁移流规模会扩大为原来的 $\exp(0.034 1)=1.035$ 倍;而流入地的社会网络如果增加一个单位,相应的迁移流规模则会扩大为原来的 $\exp(0.047 8)=1.049$ 倍。不论是流出地的社会网络,还是流入地的社会网络,都可能是后续迁移流持续与扩展的重要因素。其次,从作用的稳定性来看,不论是基础的重力模型,还是加入控制变量后的各个模型中,也不论是各个不同时期的普查数据中,流入地和流出地的社会网络因素都呈现出显著的正向作用。这表明在控制了其他变量的条件下,社会网络因素都会对省际人口迁移流流向与规模产生显著的影响作用且正向效应不变。这在一定程度上表明该变量的作用符合因果关系的条件,即两者关系不会因为第三个变量的存在而不存在。再次,不论是流出地模型,还是流入地模型中,也不论是在对称还是非对称模型中,社会网络因素的作用也相对比较稳定。且如果不包括社会网络因素,则不论是模型拟合度(chi2)还是bic指标,都会由于忽略重要变量而与完全模型之间存在显著差异。最后,社会网络对省际人口迁移流的绝对影响作用在历时变化中基本稳定。比如在流出地模型中,该变量在2000年时系数为0.036 0,2010年时为0.041 8,2020年时为0.034 1,其绝对值基本相差不大。综合来看,迁移人口社会网络是人口迁移流宏观影响因素中不可忽略的重要因素之一;如果重力模型中不包含这一变量,则模型结果可能存在忽略重要变量而带来偏差。另一方面,流入地的社会网络同样也是迁移流的重要影响因素,它既表明迁移人口在流入地的集聚效应,也反映了流入地社会网络对于后续迁移流持续的影响作用。因此,原有的存量模型有必要被扩展为社会网络-重力模型。

(四) 控制变量的影响

从上述结果可以看到,经济因素的四个变量均呈现出显著的作用。首先,不论是流入地还是流出地的模型中,两地间GDP之比均呈现出显著的正向影响,即两地间GDP的差异越大,流出或流入的迁移流规模也就越大。而且这一变量在历年的模型中都呈现出相同的显著正向性,说明其作用机制之稳定。其次,在流出地模型中,平均工资在前四次的普查或小普查中呈现出正向的显著效应,即平均工资差异越大,迁移流规模也就越大;但在2020年七普中却呈现出反向的显著性,即差异越小,



流量越大。在流入地模型中,历年均呈显著的正向作用。这种结果可能有几种不同的原因:一是测量问题。省级平均工资数据本身就是多种不同类型的数据,本文选用了其中一种,可能会因此影响估计结果。二是影响机制的变化。正如Shen所指出的,机制的变化和指标的变化同时会影响人口迁移流的规模。^[6]平均工资差异历来是影响人口迁移流的重要因素,但省际平均工资的差异也会随着社会经济发展而产生变化,从而影响人口迁移流的规模与方向。同时,人口迁移流动的来源地、流动方向等都有了一定的新的变化,从而使得总体上表现出来的选择机制也会呈现出一定的变化。而且,公共服务均等化实施多年,迁移流动人口中有许多人不再是纯粹为了经济收益进行流动,而是基于个体特征结合宏观社会经济状况(特别是流动者个体能够感知的经济收益)而做出的理性选择。综合来看,工资差异影响作用的变化背后既有迁移流动空间分布的转变,也有水平的转变和作用转变,亦有影响因素的转变。最后从流出地模型看,自2010年以来,失业率差异越大,越会形成较大的迁移流。而流入地模型中的失业率则呈现为显著的负向作用,即流入地的失业率越低,越能够吸引更大的迁移流。经济方面的指标尽管在历次普查中并不是完全一致的,但至少GDP和失业率两个变量呈现的作用与以往的研究结论是相同的。这说明控制变量起着应有的作用。

再看社会因素。一是在流出地模型中,除2005年以外,流入地的人均教育投入均呈现出显著的正向作用,即教育投入越多,流入规模也就越大。但从流入地模型看,流入地的人均教育投入并未呈现出显著的作用。这一方面说明流出地的推力和流入地的拉力并不是对等的,对流动人口而言,流出是必然的;但流入哪里则是由流入地的拉力决定的。这说明事实上流出地与流入地模型不仅仅只是模型设置问题(即重力模型应采用非对称的设置),更关键的是研究视角的改变。正如推拉模型所示,人口迁移流动必然涉及流出地、流入地和中间因素三个方面,相关研究必须同时关注这两个地点之间的异同。二是从流入地的拉力来看,已不再仅仅是经济收入,越来越多的研究表明社会因素在人口迁移流中起着越来越重要的作用。这一点同样可以由工伤保险参保率得到佐证。在流出地模型中,工伤保险参与率在2000年时为显著的负向,即参保率越低,迁移流规模越大;但随着时间的推移,2015年和2020年时该变量的作用呈现出显著的正向作用,即工伤保险参与率越高,迁移流规模也就越大。三是流出省份的流出率。本研究选择流出率变量是为体现它不同于存量变量的影响作用。存量变量是指某个迁移流占该省流出人口的比例。而本文的流出率是指前一次普查中该省的流出人口占流出省总人口的比例,某种意义上可以体现流出省份的流动普遍性,甚至可以表示流出省的迁移流动文化,或者是对流动的接受程度。在流入地模型中,流出省的流出率越高,迁移流规模越大。这表明迁移流规模从根本上受制于流出总人口的规模或比例,同时也说明流出省的流出文化或流动的普遍接受程度同样是重要基础。其实,从人口流动决策过程看,首先是决定是否流动,然后再讨论流到哪里去。而流出率则正表明潜在的流出人口规模。四是邻近性。邻近性是地理空间研究中关注的一个角度,通常认为邻近省份的人口迁移流也会越大。但在本文的结果中并未得到检验。如在流入地模型中,邻近性都呈现为显著的负向;同样的,在流出地模型中,除2000年和2020年外,它在其他普查样本中也呈现出显著的负向。这说明越是相邻省份,其迁移流规模相对越小。本研究在建模过程中也曾尝试将该变量删除,如果不包括该变量,则模型整体将无法收敛。因此,从模型建构角度看,它是一个重要变量。但结果与以往研究呈现出相反的方向。这种情况的产生原因可能有几个:一可能是模型加入的距离变量本身已经解释了因变量很大部分的变化;二是从整体角度看,全国省际人口迁移流中,有较大规模迁移流的相邻省份相对还是较少,而非邻近省份之间同样

存在相对较大规模的迁移流,同时,迁移流规模较大的相邻省份的反向迁移流也相对较小,从而使邻近性受到挑战。

五、结论与讨论

(一)研究结论

本文以2000-2020年间历次人口普查、小普查及相关统计年鉴数据为基础,利用改进的重力模型(分层泊松模型),检验了迁移社会网络、流出地劳动力规模 and 经济发展水平的地区差异在省际流动人口流入地宏观选择机制中的作用,以及这些因素的作用在时间维度上的稳定性。主要分析结果如下:

第一,前期的迁移社会网络,不论是在经典重力模型中,还是加入各种控制变量后的完全模型中,也不论是流出地社会网络的存量迁移流还是流入地的迁移社会网络,对当前人口迁移流的规模与方向都起着显著的影响作用。且这种作用在历时变化中呈现出极强的一致性。

第二,作为流出地人口的代替变量,流出地的劳动力人口规模呈现出一致的正向作用。尽管劳动力人口规模与总人口规模成正比,但在模型中呈现的显著正向作用至少表明它是影响人口迁移流的重要因素。流入地总人口规模的反向作用和流入地劳动力人口规模的不显著作用反过来说明了流出地劳动力人口规模作为人口迁移流源头的基础作用。

第三,尽管有部分经济变量在模型中呈现出不同于以往研究的结果,但在起到控制作用的同时,也体现了一定的稳定性,表明经济因素,特别是地区间经济发展水平的差异仍然是当前我国人口迁移流动的重要原因。

第四,在人口迁移流动强度逐步增加的社会环境之下,流出地人口的流动强度对人口迁移流的影响作用表明流出地在供应端的决定性作用。即潜在的流出人口规模越大,迁移流的规模也就越大。其作用表现出极强的稳定性。

(二)讨论

本文检验了文初提出的研究假设,并得到上述结论。但有关中国省际人口迁移流的研究仍然有许多问题值得进一步深入讨论。

其一,我国省际人口迁移流的流入地宏观选择机制的稳定性问题。本文最初的设计目的是基于我国人口迁移流动模式的稳定性而检验这种宏观机制的稳定性。严格来讲,机制的稳定性应该是各种社会经济因素对人口迁移流的影响作用在历年的模型中都呈现出一致的显著性和方向。但本文结果并不支持这种严格意义上的稳定性,虽然上述结论说明省区间经济发展水平的差异仍然是省际人口迁移流动的重要原因,而且非正式的迁移社会网络是人口流动的重要影响因素。在这个意义上,本文结果只能部分地支持机制稳定性的假设。这就涉及如何理解稳定性。首先,正如Shen对人口迁移流的解释,社会经济发展水平及其作用都会随时间而发生变化,最终的模型是两者共同作用的结果。^[6]本文结果表明省区间经济发展水平的差异仍然是人口迁移流动的重要基础,虽然是上述两种变化的共同结果,但至少也表明了一定的稳定性。其次,相同因素在不同时代会有不同的社会背景与含义,会起到不同的作用。如果说改革开放初期到20世纪90年代,中国省际人口迁移流动的重要原因来源于流出地艰苦的生存环境和严重的农村劳动力剩余(亦有人称之为隐性失业),表面上看是以提高经济收入为目的,其背后却是因流出地的艰难生活而不得不流出;那么到了改革开放多



年、社会经济得到充分发展的 21 世纪,省际人口流动的根本原因尽管表面上仍然为了提高经济收入,但却不再是流出地生存环境压力之下的被迫流出(这一点在模型预处理中有所体现,即如果选用 90 年代研究文献中所用的变量,如人均耕地面积等生存环境变量,在当前的数据和模型中均不显著),而是已经改变成为追求美好生活而做出的理性决策。因此,同一因素在不同的社会背景之下扮演着不同的角色,起着不同的作用。再次,各种因素在选择机制中的相对作用也会随时间而发生变化。随着社会发展进程与流动人口内部的世代更替,流动人口的流入地选择不再是纯粹的预期收入最大化,而是预期回报最大化;^[5]在经济发展水平或经济收入相差不大的情况下,社会因素在其中起到了越来越重要的作用。因此,机制的稳定性与新变化在某种意义上是共存的,并不能因新变化就否定了机制的稳定性;稳定性也不意味着机制一成不变。需要辩证地看待这种稳定性与新变化。

其二,省际流动人口迁移流持续问题。本文揭示了流出地劳动力规模是影响人口迁移流动的重要基础源泉。从现实情况看,多年以来中国人口的生育一直处于极低水平,未来流出省区的劳动力规模将逐步下降、劳动力年龄结构将逐步老化。在此背景下,未来我国省际人口迁移流是否仍然能够持续?如何能够保证迁移流的持续?这不仅涉及流入省区的社会经济发展活力,也影响流出省区的人口结构、劳动力结构。当前的省际人口迁移流动必然会同时对流入省和流出省产生深远的影响。而且,尽管七普数据表明人口迁移流动规模似乎仍然是在不断扩大,但如果从流动人口的年龄结构来看,这种规模的扩大主要来自人口结构本身(年龄别流动率最高、流动强度最大的 20-30 岁年龄段所对应的人口规模相对较大)和流动人口的年龄别流动率的改变(如年龄在 30-50 岁之间人口流动率的提高)这两者共同作用的结果。^[48]即使未来年龄别流动率有所上升,但各年龄段人口规模的逐步下降仍然无法避免流动人口规模的下降。因此,如何防范与避免以往珠三角地区劳动力短缺的经历,并最大程度降低劳动力短缺带来的经济发展问题等等的现实问题需要未雨绸缪。

其三,推拉力的作用机制问题。在整个讨论中,本文将中国当下的省际人口迁移流都置于推拉理论框架之下,不论是理论探讨还是模型设置。但作为推拉理论框架之下的流出地的推力和流入地的拉力,在摆脱了贫困压力和生存压力、全国社会经济得到长足发展的当下,到底哪个更可能引领中国省际人口迁移流的方向与规模呢?以往研究在考虑流动人口的决策时,通常是将流动决策与流入地选择分别考察,而且总认为流动决策与流动方向是综合的有机体。但潜在流动者的流动决策机制与流向选择机制是否可能存在类别差异,从而使不同的人群对推力和拉力的感知与决策是不同的?如某些人可能是在看到或者预期到流入地可能带来的各种收益与回报(不论是经济上的还是事业发展或子女教育等)才决定流动,从整体过程看是先有去向预期后的迁移决策。而另一类则可能是已经决定了要流出,但尚未决定流向哪里,在各种迁移社会网络的支持下得到了一些或者是充分的信息然后再决定流向。显然这两种类型的决策机制完全不同。后一种情况更适用于本文所讨论的研究对象即 5 年内新增的流动人口。对他们而言,流出地的作用并不是重要的(因为已经决定要流出了,而且也说明用流出率表示的流出地迁移流动文化的重要性),而流入地的各种特征(包括社会网络)才可能是重要的。一方面说明流入地迁移社会网络在其流动决策过程中的重要性,另一方面也说明他们的流向决策(由此形成的省际人口迁移流向与规模)更重要的是在于流入地的拉力,而非流出地的推力。当然,在流入地的选择过程中,流动者的个体与家庭特征(如是否有孩子或老人需要照顾等)可能同时作为阻力影响流入地选择决策,从而形成就近或远距离流动。

这一启示从理论上表明新时代流动人口流入地的选择机制更可能依赖于社会网络;除了亲情与

社会关系以外,迁移社会网络还能后续移民提供流入地的各种信息(就业、居住等),从而加强链式迁移流动。因此,重力模型的扩展需要特别强调社会网络的因素。从方法看,在利用重力模型建模时,除了应加入社会网络变量以外,有两点需要考虑:一是分层模型应以流出地作为层-2分析单位,考察流入地的吸引力而不是流出地的推力(因为流出地的特征是决定其是否流动,而不决定其流向);二是变量设置需要考虑流入地的各种绝对条件水平(包括经济与收入、社会环境、用工条件等),而不是流出地的特征;因此,模型设置也应该是非对称的。这从理论上为方法提供了依据,解决了层-2分析单位的选择与层-1模型中变量选择问题。本文的分析结果与上述讨论结果基本相同。

本文将重力模型与社会网络理论相结合,将存量-重力模型扩展为社会网络-重力模型,并检验了流出地劳动力规模在省际人口迁移流中的基础源泉作用,部分检验了宏观流入地选择机制的稳定性。这些可以被视为本文的主要贡献。当然文中存在的变量选择等模型设定问题、稳定性与作用分解等仍有待进一步深入的讨论与研究。

【参考文献】

- [1] 段成荣,赵畅,吕利丹.中国流动人口流入地分布变动特征(2000-2015)[J].人口与经济,2020,(1):89-99.
- [2] 周皓.中国人口流动模式的稳定性及启示——基于第七次全国人口普查公报数据的思考[J].中国人口科学,2021,(3):28-41.
- [3] 王桂新.中国人口流动与城镇化新动向的考察——基于第七次人口普查公布数据的初步解读[J].人口与经济,2021,(5):36-55.
- [4] 王桂新,潘泽瀚,陆燕秋.中国省际人口迁移区域模式变化及其影响因素——基于2000和2010年人口普查资料的分析[J].中国人口科学,2012,(5):2-13.
- [5] 周皓,刘文博.流动人口的流入地选择机制[J].人口研究,2022,(1):37-53.
- [6] Shen J. Explaining Interregional Migration Changes in China, 1985-2000, Using a Decomposition Approach[J]. Regional Studies, 2015, 49(7):1176-1192.
- [7] Egger P H, Tarlea F. Multi-way Clustering Estimation of Standard Errors in Gravity Models[J]. Economics Letters, 2015, 134:144-147.
- [8] Poot J, Alimi O, Cameron M P, et al. The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science[J]. Investigaciones Regionales, 2016, 36:63-86.
- [9] Zhang X N, Wang WW, Harris R, et al. Analysing Inter-provincial Urban Migration Flows in China: A New Multilevel Gravity Model Approach[J]. Migration Studies, 2020, 8(1):19-42.
- [10] 杨云彦,陈金永,刘塔.中国人口迁移:多区域模型及实证分析[J].中国人口科学,1999,(4):20-26.
- [11] Chan K W, Liu T, Yang Y. Hukou and Non-hukou Migrations in China Comparisons and Contrasts[J]. International Journal of Population Geography, 1999, (5):425-448.
- [12] Fan C C. Modeling Interprovincial Migration in China, 1985-2000[J]. Eurasian Geography and Economics, 2005, 46(3):165-184.
- [13] 巫锡炜,郭静,段成荣.地区发展、经济机会、收入回报与省际人口流动[J].南方人口,2013,(6):54-61.
- [14] 王桂新,潘泽瀚.我国流动人口的空间分布及其影响因素——基于第六次人口普查资料的分析[J].现代城市研究,2013,(3):4-11.
- [15] 周皓.中国迁移流动人口的统计定义——人口普查视角下的分析[J].中国人口科学,2022,(3):17-30.
- [16] Massey D S, Arango J, Hugo G, et al. Theories of International Migration: A Review and Appraisal[J]. Population and



- Development Review, 1993, 19(3): 431-466.
- [17] Massey D S, Espinosa K E. What's Driving Mexico-U.S. Migration? A Theoretical, Empirical, and Policy Analysis[J]. The American Journal of Sociology, 1997, 102(4): 939-999.
- [18] Massey D S. The Social and Economic Origins of Immigration[J]. The Annals of the American Academy of Political and Social Science, 1990, 510(1): 60-72.
- [19] Garip F. Discovering Diverse Mechanisms of Migration: The Mexico-US Stream 1970-2000[J]. Population and Development Review, 2012, 38(3): 393-433.
- [20] Stark O, Taylor J E, Shlomo Y. Remittances and Inequality[J]. The Economic Journal, 1986, 96(1): 722-740.
- [21] Crawley H, Hagen-Zanker J. Deciding Where to Go: Policies, People and Perceptions Shaping Destination Preferences [J]. International Migration, 2019, 57(1): 20-35.
- [22] Beine M, Bertoli S, Fernández-Huertas M J. A Practitioners' Guide to Gravity Models of International Migration [J]. World Economy, 2016, 39(4): 496-512.
- [23] 古恒宇, 沈体雁, 刘子亮, 等. 基于空间滤波方法的中国省际人口迁移驱动因素[J]. 地理学报, 2019, (2): 222-237.
- [24] Zavodny M. Welfare and the Locational Choices of New Immigrants[J]. Economic Review, 1997, Q II: 2-10.
- [25] Lewer J J, Van den Berg H. A Gravity Model of Immigration[J]. Economics Letters, 2008, 99(1): 164-167.
- [26] Shen J. Modelling Regional Migration in China: Estimation and Decomposition[J]. Environment & Planning A, 1999, 31(7): 1223-1238.
- [27] Redding S, Venables A J. Economic Geography and International Inequality[J]. Journal of International Economics, 2004, 62(1): 53-82.
- [28] Rose A K, Van Wincoop E. National Money as a Barrier to International Trade: The Real Case for Currency Union[J]. The American Economic Review, 2001, 91(2): 386-390.
- [29] Wang Z K, Winters L A. The Trading Potential of Eastern Europe[J]. Journal of Economic Integration, 1992, 7(2): 113-136.
- [30] Eichengreen B, Irwin D A. Trade Blocs, Currency Blocs and the Reorientation of World Trade in the 1930s[J]. Journal of International Economics, 1995, 38(1): 1-24.
- [31] Qi W, Abel G J, Liu S. Geographic Transformation of China's Internal Population Migration from 1995 to 2015: Insights from the Migration Centerline[J]. Applied Geography, 2021, 135: 102564.
- [32] 王桂新. 中国省际人口迁移变化特征——基于第七次全国人口普查数据的分析[J]. 中国人口科学, 2022, (3): 2-16.
- [33] Liang Z, White M J. Market Transition, Government Policies, and Interprovincial Migration in China: 1983-1988[J]. Economic Development & Cultural Change, 1997, 45(2): 321-339.
- [34] 夏怡然, 陆铭. 城市间的“孟母三迁”——公共服务影响劳动力流向的经验研究[J]. 管理世界, 2015, (10): 78-90.
- [35] 姚先国, 刘湘敏. 劳动力流迁决策中的迁移网络[J]. 浙江大学学报(人文社会科学版), 2002, (4): 125-131.
- [36] 沈建法. 中国 1995-2000 年省间人口迁移模拟与迁移因素对迁移规模贡献的分解[J]. 亚热带资源与环境学报, 2010, (4): 1-9.
- [37] Fan X, Liu H, Zhang Z, et al. The Spatio-Temporal Characteristics and Modeling Research of Inter-Provincial Migration in China[J]. Sustainability, 2018, 10(3): 618.
- [38] 杨云彦. 改革开放以来中国人口“非正式迁移”的状况[J]. 中国社会科学, 1996, (6): 59-73.
- [39] 盛广耀. 中国省际人口流动网络的演化及其解释[J]. 中国人口·资源与环境, 2018, (11): 1-9.
- [40] 朱宇, 丁金宏, 王桂新, 等. 近 40 年来的中国人口地理学——一个跨学科研究领域的进展[J]. 地理科学进展, 2017, (4): 466-482.
- [41] 刘涛, 卓云霞, 王浩晶. 邻近性对人口再流动目的地选择的影响[J]. 地理学报, 2020, (12): 2716-2729.



- [42] 李怡涵. 1985-2010年中国省际人口迁移的空间区域分布特征及影响因素研究[D]. 兰州:兰州大学,2015.
- [43] 卓云霞,刘涛,古维迎. 多维邻近性与城-城流动人口的流入地选择——基于嵌套Logit模型的实证分析[J]. 地理科学,2021,(7):1210-1218.
- [44] Cai F, Wang D. Migration as Marketization: What Can We Learn from China's 2000 Census Data?[J]. Chinese Journal of Population Science, 2003, 3(2):73-93.
- [45] Li S. Population Migration and Urbanization in China: A Comparative Analysis of the 1990 Population Census and the 1995 National One Percent Sample Population Survey[J]. International Migration Review, 2004, 38(2):655-685.
- [46] Sun M, Fan C C. China's Permanent and Temporary Migrants: Differentials and Changes, 1990-2000[J]. The Professional Geographer, 2011, 63(1):92-112.
- [47] Shen J. Increasing Internal Migration in China from 1985 to 2005: Institutional versus Economic Drivers[J]. Habitat International, 2013, 39(1):1-7.
- [48] 周皓. 我国流动人口年龄别流动率模式及其演变[J]. 华东师范大学学报(哲学社会科学版), 2023, (1):185-201.
- [49] Wu X, Yu D, Zhang Y, et al. Low Fertility Spread in China: A Blended Adaptation and Diffusion Explanation[J]. Population, Space and Place, 2022, 28(6):e2555.

[责任编辑 李新伟]

The Stability of Macro-Selection Mechanism of Destination of Interprovincial Floating Population in China

——Theoretical Extension and Empirical Test of Gravity Model

ZHOU Hao, LIU Wenbo

(Department of Sociology, Peking University, Beijing, 100871, China)

Abstract: The macro selection mechanism and its stability of floating population's destination is of great significance to reveal the spatial pattern and future development trend of China's interprovincial population flow. Based on the data of population censuses, minor censuses and relevant statistical yearbooks from 2000 to 2020, this paper examines the role of migration social network, labor force size and the level of economic development in the mechanism of migration destination selection of inter-provincial migrants, and their temporal stability by using hierarchical Poisson model. The classical gravity model is theoretically extended and empirically tested in the Chinese context. The results show that the migration social network has a significant effect on the scale and direction of migration flow, and this effect shows a strong consistency in the diachronic change. The size of the labor population in origin provinces has a significant and consistent positive impact on the migration flow and becomes the source of the migration flow. The influence of economic factors on migration flow is relatively stable, and the difference of regional economic development is still an important reason for population migration flow in China. The rate of outflow also plays a strong role in the continuous stability of migration flow. Based on the above findings, this paper emphasizes the role of social networks in the selection mechanism of inflow places, and extends the traditional stock-gravity model to the social network-gravity model. This paper also reveals the basic source function of labor force scale in the interprovincial migration flow, and discusses whether and how the interprovincial migration flow can be sustained in China.

Key Words: Floating Population, Destination Choice, Macro Selection Mechanism, Gravity Model, Stability