

城市化与人口流迁

中国跨区域人口迁移流空间结构研究 ——描述、模型表达与预测

曾永明

(江西财经大学 经济学院, 江西 南昌 330013)

摘 要: 关于人口跨区域流动的空间结构研究与数理统计明显不足, 尤其是鲜见对人口迁移流量数据自身运行规律和数据内生化过程的统计研究。描述、模型化表达与预测是人口迁移流空间结构研究的三个不应分割的系统过程。鉴于此, 将人口迁移流空间结构的描述、模型化表达和预测的“基本程式”整合, 采用乘法分量模型、对数线性模型和双边比例调整法分别来描述、模型化表达和预测我国人口迁移流空间结构特征。实证过程着眼于人口统计学的视角, 应用 2011—2017 年中国流动人口动态监测调查数据和 2000—2020 年三次人口普查数据等资料, 以长江经济带 11 省份和中国东部、中部、西部和东北四大区域之间人口迁移流为例, 基于人口迁移流矩阵的内生规律和数据自我生成过程来全景式刻画研究我国跨区域人口迁移流空间结构。主要结果显示, 列联表及其内生的乘法分量模型可以很好地描述具有二维结构的人口迁移流; $[O D A O A]$ 四参数的非饱和对数线性模型可以刻画中国跨区域人口流动空间结构的主要特征, 拟合程度达到 82%; 人口迁移流预测意义重大但探索不足, 双边比例调整法可作为人口迁移流预测的良好工具。研究主要目的是尝试引入和搭建国内人口迁移流空间结构研究的体系框架或基本程式, 推进我国人口空间统计或空间人口学的研究与应用。

关键词: 人口迁移流; 空间结构; 数据内生过程; 人口空间统计; 研究程式

中图分类号: C922 **文献标识码:** A **文章编号:** 1000-4149 (2022) 06-0058-19

DOI: 10.3969/j.issn.1000-4149.2022.00.039

收稿日期: 2022-04-06; 修订日期: 2022-08-03

基金项目: 国家自然科学基金地区科学基金项目“邻里效应对流动人口相对贫困的影响机制与治理策略研究——以赣南革命老区流出劳动人口为例”(72064018); 教育部人文社会科学研究青年基金项目“长江经济带流动人口时空格局演变及其对区域均衡发展的作用机制研究”(18YJC790006); 国家自然科学基金地区科学基金项目“数字乡村建设对脱贫地区农户收入增长与差距的双重效应研究: 影响机理与双赢策略”(72263014)。

作者简介: 曾永明, 经济学博士, 江西财经大学经济学院副教授。

一、引言

2020年中国流动人口达到3.76亿，比2010年增长69.3%^①，引起广泛关注。国家统计局局长宁吉喆在第七次全国人口普查发布会上表示，我国经济社会持续发展，为人口的迁移流动创造了条件，人口流动趋势更加明显，流动人口规模进一步扩大。人口迁移作为人口学过程之一和人口转变的重要部分^[1-2]，因其特殊的空间属性，显著区别于生育、死亡过程。不过，年龄结构可以是人口学的中心概念（central concept），但空间结构（比如人口迁移流）并不是^[3]。事实上，人口迁移流动是规范人口学（formal demography）研究的基本组成部分，然而在中国生育制度的政策影响、人口老龄化及健康中国建设背景下，对于生育、死亡（健康）的研究比人口迁移的影响要更深远。不同于生育、死亡过程一般仅影响独立的某区域，人口迁移流发生在一个多区域系统内：每个区域迁出人口均迁往多个其他区域，形成一个交互的空间“流结构”和系统^[1]，因此对于人口迁移流空间结构的概念定义和测量依然比较模糊，模型化表达更是不足^[3]，对人口迁移流空间结构的深化研究也从未停止。

虽然莱文斯坦（Ravenstien）、刘易斯（Lewis）等大批学者研究了人口迁移的基本规律、生成动因等^[4-5]，为人口迁移理论研究作出巨大贡献，但从量化的人口统计学角度来描述或测量、模型化表达和预测人口迁移流空间结构的历史并不长。事实上，对人口迁移流空间结构研究的历史梳理后发现，对人口迁移流的描述、模型化表达和预测是人口迁移研究的三个阶段，也是不应该分割的三个系统过程或基本程式。“描述”应该是最为基础的阶段，它是对人口迁移流空间结构和迁移流的认知过程。人口迁移流不同于人口规模、人口年龄等向量形式，其表现出矩阵形式和交互结构，对其描述过程也就变得复杂，统计学意义上的平均值、中位数、标准差等不再适用，这也限制了早期关于人口迁移流的深化认知和研究，更阻碍了现实问题的发现与应用。早期由于跨区域人口迁移流量数据统计的不完整，导致数据缺失问题，描述都难以实现，也影响后续模型化和预测过程^[6]。因此，早期关于人口迁移流空间结构的研究主要是集中在如何定义、测量和补数据上^[7]。夏洛克（Shryock）在描述人口迁移流空间结构研究上做了早期的努力，定义和设置了移民区域偏好指数^[8]。克莱顿（Clayton）比较早地定义了人口迁移流空间结构，认为移民在来源地和目的地之间的交互流动产生的联系即为迁移流空间结构，并将该定义应用在美国跨州移民的识别研究中^[9]。多位学者将移民效应（migration efficiency）的概念广泛应用到人口迁移结构描述中^[10-11]。最近拉莫（Raymer）等在整合前人研究基础上，采用乘法分量模型（Multiplicative Components Model）作为描述人口迁移结构的一般分析框架^[12-14]，该模型简单并易于理解，已用于多个地区或国家的研究，在描述人口迁移流空间结构研究领域产生了较大影响。

在人口迁移流空间结构的描述性研究中，移民数据不同于存量数据，它以矩阵或对偶形式出现，数据要求更高，以致数据通常缺失而不完整，即便是在当前时代，国际移民矩阵数据也不完善，也就是说，对国际间移民的数据都未能精准掌握，而且估计出完整可行的移民矩阵并不容易。阿贝尔和山德尔（Abel & Sander）提出了一种在保持已知的净移民规模基

① 数据来源：国务院第七次人口普查领导小组办公室。2020年第七次全国人口普查主要数据[M]。北京：中国统计出版社，2021：80-81。

础上从移民存量数据着手的估计方式，估算了全球 196 个国家间 1990—2010 的移民流数据，成果发表于 *Science* 杂志^[15]，可见补数据对于人口迁移研究的重要性。拉莫也对北欧国家之间的移民缺失数据进行了弥补^[6]。相比于国际移民数据，国家内部人口迁移流矩阵因人口普查或抽样调查得到相对完整的数据，因此对于国内人口迁移流空间结构的研究更多的是放在后两个阶段上：模型化表达和预测。

一个人口学概念如果能够通过简化通用的理论模型表达出来，对其接受认可和实际应用均是有意义的。比如人们熟知的人口平衡方程、生命表或预期寿命、总和生育率测算等。人口迁移流空间结构并没有一个广泛接受的数学表达形式或模型，不过其特殊性在于，人口迁移链接了迁出地和迁入地多个区域，因此空间交互模型被认为是表达人口迁移流空间结构的最理想方式^[3]。但具体到实证分析中，重心模型、熵最大化模型或信息最小化模型、对数线性模型也是常见的，不过其实这三者本质结构是一致的。当前多数学者采用对数线性模型来分析和解释人口迁移流空间结构，因为人口迁移流量表可以看作一个迁出地—迁入地双向列联表，表中的每个单元格就是人口迁移量，而对数线性模型有一个良好的统计理论和方法论机制将列联表分析和离散多变量分析框架完美地对接起来^[16]，而且对数线性模型可以纳入广义线性模型中，其相对机动的变量分布设置形式使其能更好匹配人口迁移流量数据特征，比如泊松分布、二项分布等^[17-18]。基于对数线性模型或重心模型等对中国跨区域人口迁移流空间结构的研究并不少^[2,19]，但相对于国外类似研究有两点密切关联的差异，一是对人口迁移流数据内生规律和衍生形式利用和分析不足，即从人口统计学视角切入的基础性研究不多，主要是对内生数据之外的经济社会因素讨论较多；二是局限于以迁入地、迁出地的区域效应和物理距离为主，对区域交互效应、年龄结构效应、性别效应及区域一年龄等交互效应拓展不足。

关于人口议题相关的预测大多是围绕总量或趋势视角，比如人口总量、城镇化和老龄化等，也有个别特殊的视角，比如人口分布预测^[20-21]，但是关于人口迁移流的预测则不常见，因为它是“一张网”而不是“一个点”的预测，难度大幅提升。既有关于“流”的预测研究中，基于双边比例调整（bi-proportionally adjustment）方法的人口迁移流预测在国外有多位人口学者做了成功的尝试^[3,16,22]。双边比例调整方法的基本原理是寻找一个合适的矩阵使其满足边际约束条件并与基期矩阵的分布函数保持一致。双边比例调整方法在国际贸易方面的研究文献较多，涉及基于区域间贸易流或投入产出表进行替代效应和制造效应的调整^[23-24]。根据国际贸易相关理论，徐国祥和陈海龙首次对我国省区间人口投入产出表做了编制^[25]，其中也应用到双边比例调整方法，但仅是作为估计值的调整方法，未应用到预测。汪子龙等基于多源数据并采用带有时间序列的线性拟合模型对人口迁移流进行了预测^[26]，但实证对象并非国内而是欧盟国家之间的人口迁移流。可以说，国内关于人口迁移预测实属鲜见，虽然对于拥有庞大流动人口的中国进行人口迁移流预测非常有必要。

梳理国外研究发现，关于人口迁移流空间结构的描述、模型表达与预测三个阶段或三个系统过程的研究虽然比较零散，但基本形成了一般化的研究框架，在长期的探索过程中对各个子议题的研究有了比较有效的解决方式甚至是“标准程式”。遗憾的是国内对于这一框架

的梳理还不足，理论研究较为稀缺，实证也较为罕见，尤其是对人口迁移流空间结构本身运行过程的分析还不足。显然，从人口统计学的角度来研究人口迁移流动的内生规律并基于人口迁移流矩阵自我数据生成过程的研究有很大价值。鉴于此，本文目的就在于将人口迁移流空间结构的描述、模型化表达和预测整合到一起，并作为一个标准程式，基于人口迁移流矩阵自我数据生成过程来研究我国跨区域人口迁移，期望形成关于中国人口迁移流空间结构研究的一个“基本程式”，也期待未来更多学者尤其国内学者对中国人口迁移问题进行研究，特别是从“流”数据内生过程出发，在“基本程式”上进行发展创新，推进国内空间人口学、人口统计学等研究，拓展人口迁移流的应用并对现实问题作出应有响应。本文具体研究内容包括：基于 2011—2017 年中国流动人口动态监测调查数据（China Migrants Dynamic Survey, CMDS）和最近三次的人口普查和抽样调查数据，并以长江经济带 11 省份和中国东部、中部、西部和东北四大区域之间人口迁移流为例，采用乘法分量模型来描述国内人口迁移流空间结构特征，通过对数线性模型来表达国内人口迁移流空间结构特征，应用双边比例调整法来预测国内人口迁移流空间结构特征，尝试通过描述、模型化表达和预测来建立一个国内人口迁移流空间结构的研究程式。

二、数据来源与样本处理

中国流动人口动态监测调查（CMDS）是由国家卫生健康委员会（原国家卫生和计划生育委员会）组织的在流入地对流动人口进行的连续断面调查，自 2009 年起每年进行监测调查。该调查以在流入地居住一个月以上、非本市（县、区）户籍的 15—59 周岁流动人口为调查对象（2015 年后不再设置年龄上限，即调查 15 周岁以上所有年龄段流动人口），采取分层、多阶段、与规模成比例的 PPS 方法进行抽样。调查区域涵盖全国 31 个省（市、自治区）和新疆生产建设兵团（不包含香港、澳门和台湾地区）。CMDS 数据有比较详细的关于个体来源地（户籍地）和目的地（现居地）的空间属性，是分析流动人口空间结构的良好样本，还能结合年龄、性别等进行灵活分析。事实上，对 CMDS 的开发研究大多数限于个体经济、社会问题的应用，而忽视该数据库调查对象本身强烈的空间属性。本文从人口迁移流空间结构视角对这一数据库进行空间挖掘，充分挖掘其在人口空间统计上的应用价值。

需要说明的是，尽管 CMDS 始于 2009 年，但 2009 和 2010 年的调查方案与之后的调查有所差异，2011 年后开始启用新的流动人口动态监测调查方案，调查体系更成熟，样本量更大，代表性也更好，因此本文起始年设为 2011 年。同时，因分析框架一致并限于篇幅，本文不分析所有年份，最终选择 2011、2014 和 2017 年样本进行实证分析。为了全面分析全国人口迁移流的空间结构，本应该选择全国 31 个省（市、自治区）为佳，但鉴于迁出地—迁入地 OD 结构（Origin-Destination）列联表过大和篇幅所限，文章以长江经济带 11 个省级单元为例，既有代表性，也能在文中列表分析。另外，加入年龄后，迁出地—迁入地—年龄 ODA 结构（Origin-Destination-Age）相关图表和模型参数更加复杂，因此进一步缩减到以中国东部、中部、西部和东北地区四个大区进行分析。当然，案例区域的选择并不影响研究的基本程式、基本结论和意义所在，而且全国任何区域、省级或市级间的人口迁移流空间结构

分析完全同理。

具体实证分析中，本文实际上仅有一个变量或研究指标，即人口迁移流量，其他相关变量是通过人口迁移流量矩阵内生而来。本文以流动人口户籍地作为迁出地（O）、以现居地作为迁入地（D）将样本进行汇总并建立起全国 31 个省份（剔除来源地为港澳台及国外样本，现居地不含港澳台及国外样本，无需处理）之间的人口迁移流量矩阵，并汇总为东部、中部、西部和东北四大区域之间的人口迁移流量矩阵。当考虑年龄结构时，则每个年龄段（比如每 5 岁一组）构建一个人口迁移流量矩阵（性别或户籍同理，限于篇幅本文实证中未作这类分析）。最终构建了长江经济带 11 省份 2011 年、2014 年 $11 \times 11 \times 9$ 的人口迁移流量矩阵和 2017 年 $11 \times 11 \times 14$ 的人口迁移流量矩阵（9 和 14 分别为每 5 岁一组的年龄组个数，2015 年前调查对象为 15—59 岁，2015 年之后为 15 岁以上所有年龄段流动人口），以及中国四大区域 2011 年、2014 年 $4 \times 4 \times 9$ 的人口迁移流量矩阵和 2017 年 $4 \times 4 \times 14$ 的人口迁移流量矩阵进行实证分析。另外，本文在人口迁移流的预测研究中，也用到了 2000、2010 和 2020 年的三次人口普查数据。

三、中国跨区域人口迁移流空间结构的描述

1. 人口迁移流空间结构描述方法

人口迁移流量是一种计数（counts）数据，它可以从属性数据分析（categorical data analysis）视角中的列联表切入^[27]，其中以双向或二维列联表最为基础。人口迁移流的基本结构是双向列联表，包括迁出地和迁入地（origin by destination，OD），多维列联表还可以引入年龄、性别等人口学属性。

（1）双向列联表与人口迁移流空间结构基本描述。如表 1 所示，人口从来源地 i 迁移到目的地 j ，计两地间的人口迁移量为 n_{ij} ，从某来源地流出的边际总量（marginal totals）记为 n_{i+} ，流入某目的地的边际总量记为 n_{+j} ，而迁移总量记为 n_{++} 。需要指出的是，区域内部的流动不计入测算，即当 $i=j$ 时， $n_{ij}=0$ ，因为该流量不属于跨区域流动。人口迁移流列联表可以分解为几个独立的分量，并构成乘法分量模型^[3]：

$$n_{ij} = (T)(O_i)(D_j)(OD_{ij}) \tag{1}$$

其中，总分量 T （相当于表 1 中 n_{++} ）表示总体迁移水平或总效应（overall effects）；迁出地分量 $O_i(n_{i+}/n_{++})$ 表示来源地的相对推力或迁出效应，迁入地分量 $D_j(n_{+j}/n_{++})$ 表示目的地的相对拉力或迁入效应， O_i 和 D_j 构成主效应（main effects）；迁出地—迁入地交互分

表 1 人口迁移流的 OD 双向列联表结构

迁出地 O_i	迁入地 D_j				(边际) 总量
	1	2	3	4	
1	$n_{11} (=0)$	n_{12}	n_{13}	n_{14}	n_{1+}
2	n_{21}	$n_{22} (=0)$	n_{23}	n_{24}	n_{2+}
3	n_{31}	n_{32}	$n_{33} (=0)$	n_{34}	n_{3+}
4	n_{41}	n_{42}	n_{43}	$n_{44} (=0)$	n_{4+}
(边际) 总量	n_{+1}	n_{+2}	n_{+3}	n_{+4}	n_{++}

量 OD_{ij} 定义为 $n_{ij}/[(T)(O_i)(D_j)]$ ，表示观测迁移量与期望迁移量的比率（ratio of observed migration to expected migration），也有学者将其解释为迁出地与迁入地之间的物理或社会距离^[13] 或吸引系数^[26]，该值越大则两地间的联系越强。当交互分量大于 1 时，表示观测值多于期望值，反之当交互分量小于 1 时，表示观测值低于期望值。

(2) 引入其他人口学特征的多维结构描述。式 (1) 是最基本的人口迁移流描述方式，这类模型还可以引入出生地、年龄、性别^[13-15] 等，其中年龄在移民过程中影响较大，移民本身是一个年龄选择（age-selective）过程^[28]，因此，为了清晰展示各参数估计过程且考虑篇幅问题，本文仅引入年龄进行拓展分析（其他同理）。则模型表达形式为：

$$n_{ijx} = (T)(O_i)(D_j)(A_x)(OD_{ij})(OA_{ix})(DA_{jx})(ODA_{ijx}) \quad (2)$$

其中， A 指年龄（age）， x 表示 5 岁的年龄分组，其他参数同上。显然，式 (2) 相对于式 (1) 复杂很多，模型增加了年龄项，因此包含了三个二维交互项和一个三维交互项，总共有八个参数，称为饱和模型（简记 ODA 模型）。尽管如此，模型参数的意义和解释并未发生变化，依然相对简单，总效应 T 、主效应 O_i 和 D_j 、迁出地—迁入地交互效应 OD_{ij} 是一致的：

$$T = \sum_{ijx} n_{ijx} \quad (3)$$

$$O_i = \sum_{jx} n_{ijx}/T \quad (4)$$

$$D_j = \sum_{ix} n_{ijx}/T \quad (5)$$

$$OD_{ij} = \sum_x n_{ijx}/[(T)(O_i)(D_j)] \quad (6)$$

新增的相关年龄效应参数为：

$$A_x = \sum_{ij} n_{ijx}/T \quad (7)$$

$$OA_{ix} = \sum_j n_{ijx}/[(T)(O_i)(A_x)] \quad (8)$$

$$DA_{jx} = \sum_i n_{ijx}/[(T)(D_j)(A_x)] \quad (9)$$

$$ODA_{ijx} = n_{ijx}/[(T)(O_i)(D_j)(A_x)(OD_{ij})(OA_{ix})(DA_{jx})] \quad (10)$$

A_x 、 OA_{ix} 、 DA_{jx} 、 ODA_{ijx} 分别定义为年龄主效应、迁出地—年龄交互效应、迁入地—年龄交互效应和迁出地—迁入地—年龄交互效应。以上所有参数都需要满足相应约束条件：

$$\sum_i O_i = \sum_j D_j = A_x = 1 \quad (11)$$

$$\frac{\sum_x (A_x \sum_i OA_{ix})}{m} = \frac{\sum_x (A_x \sum_j DA_{jx})}{m} = \frac{\sum_i (O_i \sum_i OA_{ix})}{k} = \frac{\sum_j (D_j \sum_j DA_{jx})}{k} = 1 \quad (12)$$

其中， m 为区域个数， k 为年龄分组数。

2. 人口迁移流空间结构描述的实证分析

(1) 双向列联表：长江经济带省际人口迁移流描述。为了解释和描述各乘法分量，如

表 2 和表 3 所示，基于 2017 年 CMDS 数据，以长江经济带 11 个省级单元之间的跨区域人口迁移流为例进行分析。表 2 显示，在调查的长江经济带 56913 个样本中跨省人口迁移流总量为 23660 人。从边际总量来看，安徽迁出人口最多（6579 人），浙江迁入人口最多（6869 人），其中安徽迁往江苏的流量最大（2118 人）。表 3 是各乘法分量，为了解释各乘法分量数据生成过程，以江苏迁往上海的流量 1216 人为例，该流量值分解为四个乘法分量：

$$n_{21} = (T)(O_2)(D_1)(OD_{21}) = (23660)(0.072)(0.219)(3.25) = 1216$$

或者：

$$OD_{21} = n_{21} / [(T)(O_2)(D_1)] = 1216 / [(23660)(0.072)(0.219)] = 3.25$$

其中，下标 2 表示第二行的江苏，下标 1 表示第一列的上海。这些分量的解释比较简单，总分量 23660 人即长江经济带的跨省迁移人口总量，迁出地分量 0.072 表示 7.2% 的人口是从江苏迁出的，迁入地分量 0.219 表示 21.9% 的人口迁往上海，最后比较重要的交互分量 3.25 表示区域间的联系强度，数据意义为实际观测值与期望值的比率，即相对每 10 个期望迁移流，大约有 32 个实际迁移流。显然，交互分量 3.25 显著大于 1，显示了两者之间较

表 2 长江经济带人口迁移流联表：观测流量 人

迁出	迁入											合计
	上海	江苏	浙江	安徽	江西	湖北	湖南	重庆	四川	贵州	云南	
上海	0	21	8	3	6	1	1	3	1	1	1	46
江苏	1216	0	245	55	39	37	9	33	17	24	32	1707
浙江	390	203	0	64	102	61	46	33	37	79	135	1150
安徽	2009	2118	1939	0	121	107	33	32	21	96	103	6579
江西	333	226	1098	34	0	82	85	33	44	91	155	2181
湖北	324	241	550	20	118	0	124	101	75	89	183	1825
湖南	181	99	483	16	153	123	0	48	58	425	539	2125
重庆	160	125	378	5	25	260	33	0	233	239	319	1777
四川	415	432	752	14	77	78	41	981	0	417	572	3779
贵州	93	112	1001	6	41	18	20	148	20	0	308	1767
云南	61	87	415	4	10	4	5	29	56	53	0	724
合计	5182	3664	6869	221	692	771	397	1441	562	1514	2347	23660

表 3 长江经济带人口迁移流联表：乘法分量 人

迁出	OD_{ij}											合计
	上海	江苏	浙江	安徽	江西	湖北	湖南	重庆	四川	贵州	云南	
上海	0.00	2.95	0.60	6.98	4.46	0.67	1.30	1.07	0.92	0.34	0.22	0.002
江苏	3.25	0.00	0.49	3.45	0.78	0.67	0.31	0.32	0.42	0.22	0.19	0.072
浙江	1.55	1.14	0.00	5.96	3.03	1.63	2.38	0.47	1.35	1.07	1.18	0.049
安徽	1.39	2.08	1.02	0.00	0.63	0.50	0.30	0.08	0.13	0.23	0.16	0.278
江西	0.70	0.67	1.73	1.67	0.00	1.15	2.32	0.25	0.85	0.65	0.72	0.092
湖北	0.81	0.85	1.04	1.17	2.21	0.00	4.05	0.91	1.73	0.76	1.01	0.077
湖南	0.39	0.30	0.78	0.81	2.46	1.78	0.00	0.37	1.15	3.13	2.56	0.090
重庆	0.41	0.45	0.73	0.30	0.48	4.49	1.11	0.00	5.52	2.10	1.81	0.075
四川	0.50	0.74	0.69	0.40	0.70	0.63	0.65	4.26	0.00	1.72	1.53	0.160
贵州	0.24	0.41	1.95	0.36	0.79	0.31	0.67	1.38	0.48	0.00	1.76	0.075
云南	0.38	0.78	1.97	0.59	0.47	0.17	0.41	0.66	3.26	1.14	0.00	0.031
D_j	0.219	0.155	0.290	0.009	0.029	0.033	0.017	0.061	0.024	0.064	0.099	23660

强的联系度。其具有强联系度（交互分量大于 2）的有 16 个，比如上海—安徽、湖北—湖南、重庆—四川等，这些强联系度区域的共同特点是属于边界省份，在物理空间上相连，这一结果也符合人口迁移的地理临近规律^[4]。从交互分量看，长江经济带省际间人口迁移流具有明显的空间选择性，高强度的流量发生在少部分空间，多数省份之间迁移流并不大。

（2）多维列联表：引入年龄的中国四大区域人口迁移流描述。加入年龄后，ODA 结构相关图表和模型参数更加复杂，限于篇幅，文章以中国东部、中部、西部和东北地区四个大区进行分析。趋势上则以 CMDS 中 2011、2014 和 2017 年三期的数据进行时间序列上的分析。依据前面提及的 ODA 饱和模型，包含八个参数。首先，总效应分量即跨区域流动总人口分别为 42937、68009、55605，因为该分量是一个绝对数值，然而每年 CMDS 抽样的样本数不一致，因此趋势性的比较意义不大，不过这并不影响后续其他分量的趋势性分析（其他分量是相对数值，与样本量规模的具体值无关）。图 1 显示了区域主效应，其中迁出地主效应以中部为主，超过 55% 的跨区域迁入人口来自中部，西部也达到 26%；迁入地主效应在东部具有绝对优势，占比达到 70%；而西部和东北迁出地和迁入地主效应基本相当。综合比较发现，东部地区是中国主要迁入地，中部地区是主要迁出地。从时间趋势来看，波动非常微小，一个较为明显的趋势是东部的迁入地效应有所降低，而西部的迁入地效应有所增加，这符合最近中国人口迁移的新特点：西部人口就近迁移的趋势有所增加。人口回流和就近流迁表现出较强动力，从目前人口流动的范围来看，长距离，特别是跨省流动人口所占的比重呈现下降趋势，省内流动人口的地位上升明显^[29]，传统的人口流动模式正被逐渐打破。

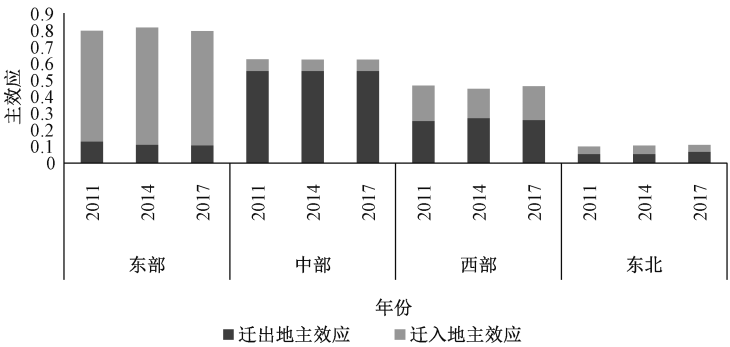


图 1 迁出地主效应分量 O_i 和迁入地主效应分量 D_j 演变趋势

对于年龄主效应 A_x ，需要说明的是，CMDS 调查均设定为 15 周岁以上流动人口，因此本文无法分析 15 岁以下人口的迁移结构。同时，在 2015 年以前，最高年龄设定为 59 周岁，即 60 岁以下，而 2015 年及之后不设上限，因此，本文分析的 2011、2014 年仅为 15—59 周岁，2017 年为 15 周岁及以上，如图 2 所示。从曲线结构可以看出，中国人口迁移的年龄特征在 2011—2017 年基本一致，呈现明显的倒“U”型特征，其中 25—40 岁是人口迁移的主要群体。莱文斯坦的经典人口迁移法则中“年龄律”并未发生明显改变，中青年依然是人口迁移的核心群体。

从时间趋势来看，年龄波峰从 2011 年的 35—39 岁年龄段转到 2014 和 2017 年的 25—29

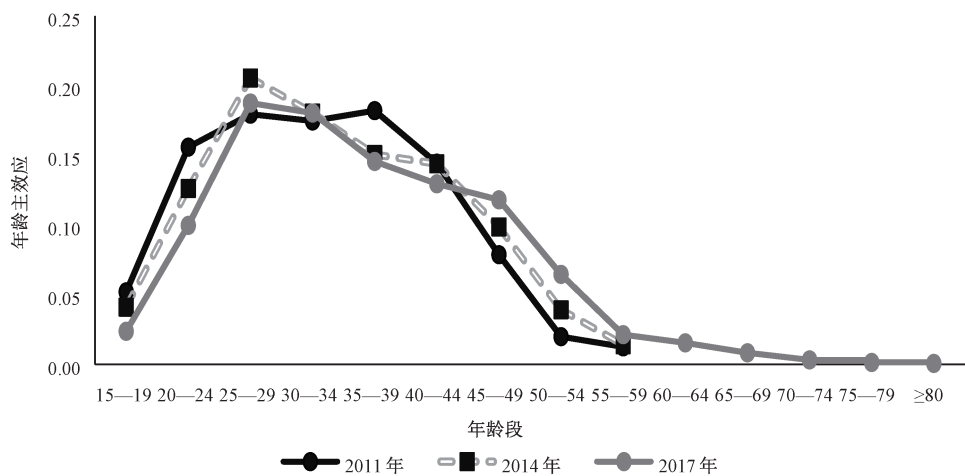


图 2 年龄主效应分量 A_x 演变趋势

岁年龄段，说明这一时期中国人口迁移有从老一代向新生代转移的历史转变特征，年龄结构转向年轻化；而且 2011 年 25—29、30—34、35—39 岁三个年龄段分布较为均衡，而后两个时期有明显的独立波峰特征。25 岁以下的群体跨区域迁移比例有明显的下降趋势，原因可能包括，一方面是中国整体人口结构的变化，即生育下降的累计效应，另一方面是流动人口中的就近就业创业的比例在提升，跨区域流动则有所放缓^[29]。45 岁及以上群体迁移比例有明显的上升趋势，一方面可能是老龄化的宏观背景所导致，另一方面可能是因为当前家庭化迁移的比例在上升，更多比例的中老年群体随子女迁移到流入地。

以上三个主效应的分析比较好理解，他们是基于独立的比例测算，而在交互分量中，各交互分量是通过各主效应的综合比例测算得到，表示的是观察值与期望值的偏离，一定程度上也可以理解为两者之间的“联系强度”。迁出地—迁入地交互分量 OD_{ij} 参数结果如图 3 所示，各分量在时间趋势上相对较为稳定，但空间差异较大，前者东部—中部、东部—西部和东部—东北有最为明显的强度， OD_{ij} 均在 2 以上，尤其前者超过 4，即实际观察到从东部流

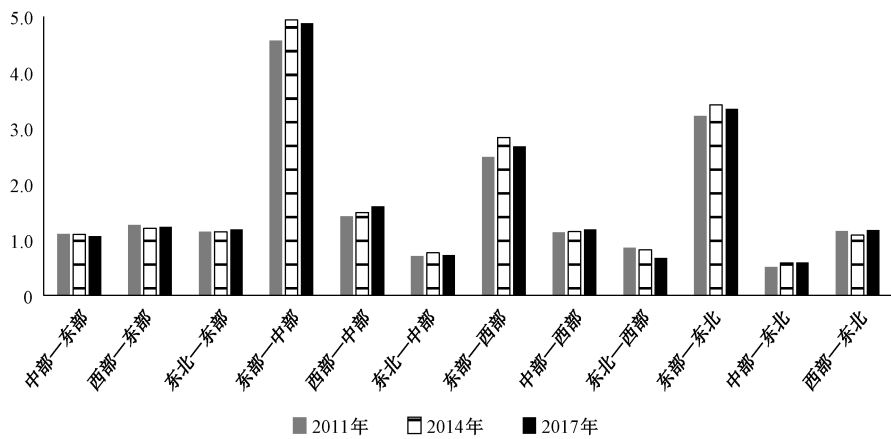


图 3 迁出地—迁入地交互分量 OD_{ij} 演变趋势

入中部的人口是期望的 4 倍。而东北—中部、东北—西部和中部—东北 OD_{ij} 均低于 1，即观测迁移人口要低于期望迁移人口。剩下的几个分量则基本趋近于 1，即观测迁移人口与期望迁移人口相当。

迁出地一年龄交互分量 OA_{ix} 和迁入地一年龄交互分量 DA_{jx} 参数结果如图 4 所示，该分量在四大区域的基本特征有明显的差异，均与图 2 中的年龄主效应分量的倒“U”型特征存在较大差异。总体上 2011—2017 年的交互分量 OA_{ix} 三条曲线和交互分量 DA_{jx} 在每个区域的基本波动趋势是相似的，尤其是在 15—59 岁的年龄段，即说明该分量特征的稳定，也说明抽样保持了较为良好的连续性。60 岁以后变化较大，可能是受样本限制的影响，仅 2017 年有数据，但这不影响基本特征分析。

在东部地区， OA_{ix} 的走势基本是随着年龄上升迁出规模上升，而 DA_{jx} 走势大致随着年龄上升迁入规模下降。可能是由于东部对于年轻人口的吸引力更大，对于中老年群体表现出一定的排斥，而且年轻群体对于东部大城市的适应力要强于中老年群体^[30]，因此中老年群体迁出比例大而迁入比例小，青年群体则相反。

在中部地区， OA_{ix} 与 DA_{jx} 趋势也大致相反，前者随着年龄缓慢下降，后者随着年龄缓慢上升。具体来说，中部作为全国主要流出群体来源地，在 50 岁以前各年龄段比例基本相当，外出的年龄选择效应较弱，到了 50 岁之后，同样因老化导致的就业匹配下降使得外出比例下降，而迁入的比例上升。这里的“迁入”其实可能是“被迫回流”^[31]，中老年群体选择就近迁移。

在西部地区， OA_{ix} 与 DA_{jx} 趋势同东部相似，但趋势曲线幅度更大。事实上，同中部一样，西部是全国主要流出群体另一来源地，外出各年龄群体比例相当，走势比较稳定。不过在 15—19 岁年龄段，其迁出比例明显更高，可能的原因是西部该年龄段（大致是高中阶段）就学率不高，远低于东部高中普及率^[32]，直接外出务工比例较大。 DA_{jx} 陡峭的趋势说明，对于西部中老年群体，迁入或回流表现出非常显著的动力。

在东北地区， OA_{ix} 与 DA_{jx} 均表现出随着年龄增加而上升的趋势，而且两者在前期都比较稳定地上升，但前者在中老年阶段有非常显著的提升，尤其是老年阶段迁出东北的比例较大。这可能是“候鸟迁移”的表现，老年群体冬季避寒迁出、夏季返回避暑成为东北的典型特征^[33]，因此东北中老年群体的迁移力表现出比其他年龄段更显著的特征。

四、中国跨区域人口迁移流的对数线性模型表达

1. 对数线性模型表达的基本原理

乘法分量模型的结构是乘积形式，如果取对数则可以变换为对数线性模型（Log-linear model）表达形式，因此，实际上乘法分量模型和对数线性模型在本质上是一致的。将式（2）取对数得到：

$$\log(n_{ij}) = \lambda + \lambda_i^O + \lambda_j^D + \lambda_x^A + \lambda_{ij}^{OD} + \lambda_{ix}^{OA} + \lambda_{jx}^{DA} + \lambda_{ijx}^{ODA} \quad (13)$$

其中，参数 λ 上标 O 和 D 分别指示迁出地 i 和迁入地 j ，显然模型（2）和（13）是一

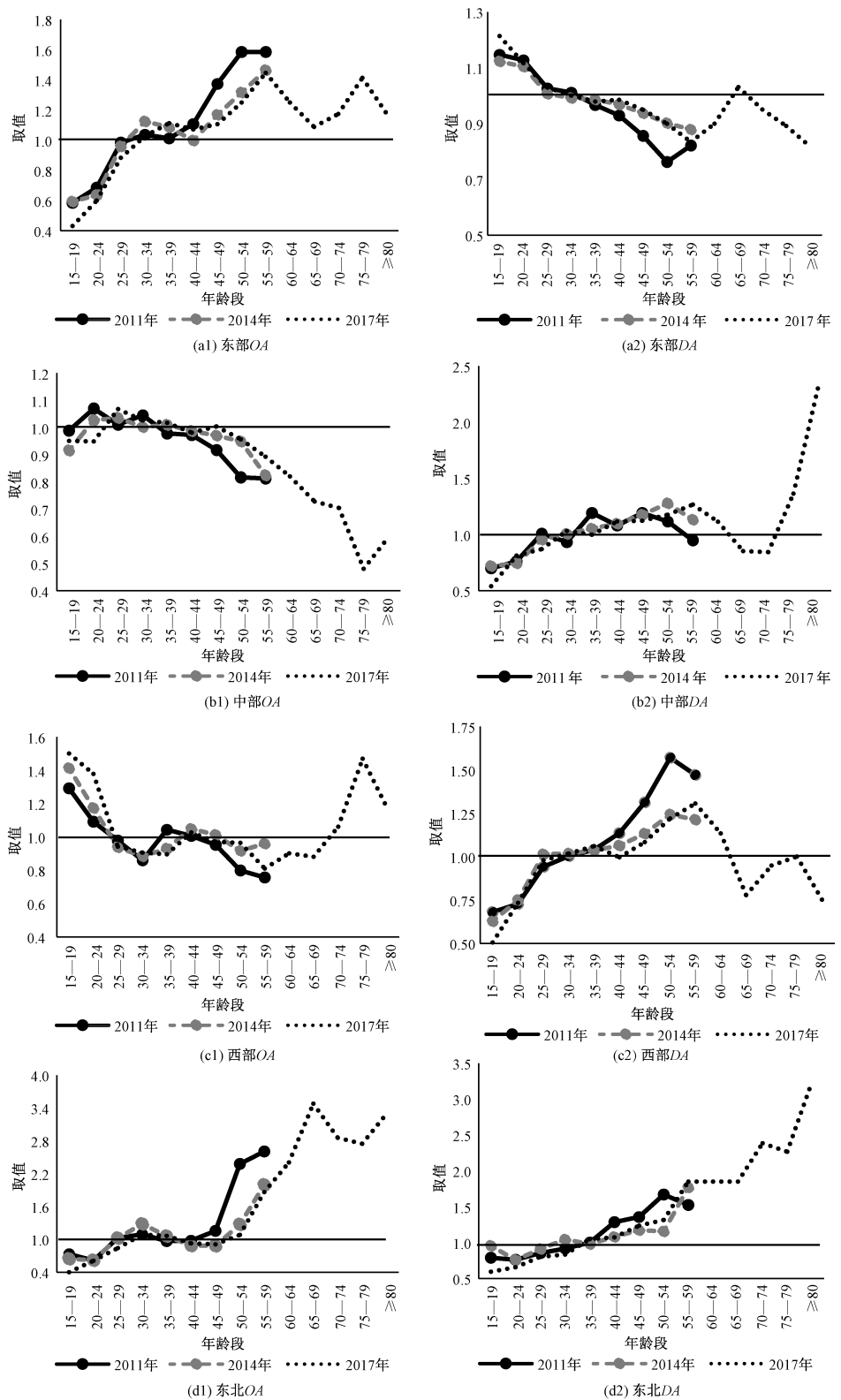


图 4 迁出地—年龄交互分量 OA_{ix} 、迁入地—年龄交互分量 DA_{jx} 演变趋势

致的，都为对数线性模型或其变形，前述以乘法分量模型开始分析是为了方便理解并引入对数线性模型。当分解测算参数和描述空间结构特征时用乘法形式较为合适，当拟合模型参数并表达空间结构时用对数线性模型形式较为妥当。对数线性模型有一套完备的统计建模框架，可归属于广义线性回归类型^[27]。对数线性模型的参数与前述乘法分量模型的各分量基本有一致的含义和解释，不再赘述。

模型（13）包含了全部八个参数，称为饱和模型，模型（13）的简化形式称为不饱和模型，比如不考虑最后三项（其他形式同理），可得到如下模型：

$$\log(\hat{n}_{ijx}) = \lambda + \lambda_i^O + \lambda_j^D + \lambda_x^A + \lambda_{ij}^{OD} \quad (14)$$

简化形式的模型主要目的在于以更少的参数来拟合和表达区域人口迁移流，通过各种简化形式的模型拟合与比较找出相对简单的形式来表达人口流动的空间结构，同时也能找出最为关键的参数或分量。不过任何形式的简化模型都是非饱和的，即参数有所缩减，因此，通常会增加一个补偿量（偏移量，offset）到非饱和模型中，因此模型（14）通常为：

$$\log(\hat{n}_{ijx}) = \lambda + \lambda_i^O + \lambda_j^D + \lambda_x^A + \lambda_{ij}^{OD} + \log(\widetilde{n}_{ijx}) \quad (15)$$

其中， $\log(\widetilde{n}_{ijx})$ 为补偿量，通常有两种方式来构建补偿量，一是以对角线为 0、非对角线为 1 的结构化矩阵^[16]；二是引入辅助信息，比如历史普查或抽样调查数据^[3,14]，相比距离测算等，历史人口迁移数据能够更好地估计当前的流动性，历史辅助信息也通常用来捕捉人口迁移流的空间形态。

需要指出的是，人口迁移流量是非负的计数数据，通常用泊松分布、二项分布等，它们比正态分布能更好拟合人口迁移流数据，这在人口迁移的重心模型实证中有很多文献做了充分的应用^[17-19]，本文以泊松分布进行实证拟合。同时模型（15）属于广义线性回归模型范畴，最小二乘法估计存在偏差，一般用极大似然估计能得到无偏估计。本文将拟合饱和模型及各类组合的非饱和模型来识别主要参数，找出既相对简化又能最大限度拟合人口迁移流矩阵的模型及结构。其中模型的拟合优度可以通过极大似然比统计量来判定：

$$G^2 = 2 \sum n_{ijx} \ln(n_{ijx} / \hat{n}_{ijx}) \quad (16)$$

该统计量越小拟合越好。模型的拟合优度还可以通过伪 R^2 （pseudo R^2 ）进行判定，该值越大则拟合越好，本文同时给出两种结果。

2. 实证分析：对数线性模型拟合、比较与关键分量表达

本文以最近的第七次全国人口普查省际之间的人口迁移流重新整理得到中国四大区域迁移流作为补偿量进行计算，并放弃人口迁移流正态分布的假定，采用泊松分布进行拟合。由于各个分量的组合类型多达数十种，限于篇幅，选择六个代表性的拟合模型进行分析，结果如表 4 所示^①。拟合模型（1）—（6）分别为 $[O D OD]$ 、 $[A]$ 、 $[O D A]$ 、 $[O D A OA]$ 、 $[O D OD OA DA ODA]$ 、 $[O D A OD OA DA ODA]$ 六个组合模型估计结果。从拟合系数来看，除了模型（5）的 ODA 分量不显著外其他所有模型参数均显著通过 1% 统计检验，应该说通

① 限于篇幅，没有报告每个模型拟合系数、标准差等参数，有需要可联系作者索取。

过数据自生过程得到的分量在拟合人口迁移流量时能得到比较好的效果。不过，从拟合优度来看，六个模型的伪 R^2 差距比较大，最差的为 0.0752，最优的为 0.8189，通过比较，说明有影响拟合优度的关键参数。为此，本文给出更多模型的拟合优度伪 R^2 ，并与 G^2 一起来分析影响拟合优度的主要分量，结果报告于表 4。

表 4 人口迁移流量拟合的代表性模型

分量	模型 (1)	模型 (2)	模型 (3)	模型 (4)	模型 (5)	模型 (6)
O	0.884 *** (0.028)		0.954 *** (0.025)	1.052 *** (0.025)	0.123 *** (0.032)	1.005 *** (0.028)
D	-1.288 *** (0.019)		-1.241 *** (0.017)	-1.255 *** (0.017)	-1.295 *** (0.020)	-1.261 *** (0.019)
A		14.390 *** (0.073)	14.390 *** (0.073)	14.544 *** (0.074)		14.639 *** (0.074)
OD	-0.037 *** (0.007)				-0.008 *** (0.007)	-0.023 *** (0.007)
OA				0.5476 *** (0.017)	0.287 *** (0.012)	0.523 *** (0.017)
DA					-0.012 *** (0.017)	0.367 *** (0.026)
ODA					-0.000 (0.000)	0.000 *** (0.000)
常数 (T)	-9.66 *** (0.022)	-11.59 *** (0.011)	-11.27 *** (0.015)	-11.88 *** (0.025)	-10.05 *** (0.048)	-12.28 *** (0.066)
观测值	224	224	224	224	224	224
极大似然值	-30818	-9063	-6574	-6133	-30542	-6035
pseudo R^2	0.0752	0.728	0.8027	0.8159	0.0834	0.8189

注：*、**、*** 分别表示 10%、5%和 1%显著水平上显著。

同样，限于篇幅仅给出部分拟合优度结果，尽管未穷尽所有参数可能的组合模型，但是表 5 给出了 14 个主要的组合模型拟合优度，不影响分析和基本结论。通过比较 14 个模型的极大似然比统计量 G^2 和伪 R^2 ，发现年龄主效应或者说人口迁移的年龄结构是中国跨区域人口迁移流的最大影响因素^①，仅考虑年龄结构主效应时 G^2/df 为 77.2，伪 R^2 为 0.728。同时，其他模型中，当加入年龄主效应时 G^2/df 越来越小，伪 R^2 越来越大，当不加入时则相反，因此年龄主效应应该是影响拟合优度的最关键因素。比较时还发现， OA

表 5 多元分量组合模型的拟合优度统计量

模型参数组合	G^2	G^2/df	pseudo R^2
$O D OD$	60658	275.2	0.0752
$O D OA$	60115	273.3	0.0833
$O D DA$	60684	275.8	0.0748
$O D A$	12170	55.3	0.8027
$O D A OA$	11452	52.3	0.8159
$O D A DA$	11901	54.3	0.8067
$O D A OD$	12214	55.8	0.8135
A	17149	77.2	0.728
$A OD$	16925	76.6	0.7314
$A OA DA$	16130	73.3	0.7433
$A OD OA DA$	15947	72.8	0.746
$A OA DA ODA$	16086	73.5	0.744
$O D OD OA DA ODA$	60106	276.9	0.834
$O D A OD OA DA ODA$	11092	51.4	0.8199

① 该结论是基于仅考虑迁出地、迁入地和年龄三个因素的情形下，如果进一步考虑出生地、性别、户籍等可能结论有所变化，前面已述，鉴于篇幅和描述的复杂性暂未考虑其他因素。

也是一个较为重要的影响参数，在保持 A 的情形下，叠加 OA 也对拟合优度有较为明显的影响。通过拟合优度可知，表中 $[O D A OD OA DA ODA]$ 为最优组合，但比较发现， $[O D A]$ 、 $[O D A OA]$ 、 $[O D A DA]$ 、 $[O D A OD]$ 组合与最优相差甚微，因此可用简化的但拟合优度高的模型来指征中国跨区域人口流动的空间特征，比如 $[O D A OA]$ 四个参数基本可以刻画中国人口迁移结构。

五、中国跨区域人口迁移流矩阵的预测

1. 预测方法介绍：RAS 预测模型

前面已述，关于人口议题相关的预测大多是围绕总量或趋势视角，比如人口总量、城镇化或老龄化等，而关于人口迁移流矩阵预测在国内还比较罕见。既有研究中，基于双边比例调整方法的人口迁移流矩阵预测有较多的成功尝试，也有国内学者预测了国外人口迁移流。鉴于此，本文结合国内外学者成果，以双边比例调整方法来预测我国跨区域人口迁移流矩阵，并以长江经济带 11 省份进行实证分析来说明该方法的应用。

依据奈尔 (Nair) 的研究^[22] 人口迁移流的双边比例调整基本方程是：

$$N(t) = r_i \cdot N(t-1) \cdot s_j \quad (17)$$

其中， $N(t)$ 为预测期的人口迁移流矩阵， $N(t-1)$ 为基期的人口迁移流矩阵。设 $n_{ij}(t)$ 是 $N(t)$ 中的元素，表示从区域 i 到 j 的人口迁移量； $n_{ij}(t-1)$ 为基期 $N(t-1)$ 矩阵中人口迁移量元素。 r_i 和 s_j 为行、列的平衡因子，因此该方法习惯称为 RAS 法，其核心问题是寻找合适的 r_i 和 s_j 即可作出预测。根据乘法分量模型 (1) 相关边际总量的定义可知：

$$\sum_i n_{ij}(t) = n_{+j}(t) = s_j \sum_i r_i \cdot n_{ij}(t-1) \quad (18)$$

$$\sum_j n_{ij}(t) = n_{i+}(t) = r_i \sum_j s_j \cdot n_{ij}(t-1) \quad (19)$$

变换即可得到：

$$s_j = n_{+j}(t) / \sum_i r_i \cdot n_{ij}(t-1) \quad (20)$$

$$r_i = n_{i+}(t) / \sum_j s_j \cdot n_{ij}(t-1) \quad (21)$$

因此，如果能先期给定或预测边际总量 $n_{+j}(t)$ 和 $n_{i+}(t)$ 的值，则通过迭代方法使上面两式收敛即可解得 r_i 和 s_j ，也即能预测出 $n_{ij}(t)$ 。因此，当前问题已转换为边际总量的预测，而该预测可借鉴拉莫的处理^[14]，比如：

$$n_{+j}(t) = n_{+j}(t-1) + \frac{[(n_{+j}(t-2) - n_{+j}(t-3))] + [n_{+j}(t-1) - n_{+j}(t-2)]}{2} \quad (22)$$

其中， $n_{+j}(t-2)$ 、 $n_{+j}(t-3)$ 为前两期、前三期的列边际总量值，行边际总量值预测同理。如果考虑年龄，只需对每个年龄段的迁移矩阵做预测，原理是一致的。依此基于 RAS 方法的人口迁移流矩阵预测即可实现。

2. 实证分析：长江经济带省际间人口迁移流预测

本文以长江经济带 11 省份进行实证分析，其中 2020 年为基期，2010 年为前两期，2000 年为前三期，要预测的是 2030 年长江经济带 11 省份之间的人口迁移流矩阵。2020 年

人口迁移流量矩阵见表 2 和表 3，限于篇幅 2010 和 2000 年人口迁移流量未给出。不过预测 2030 年数据之前，先给以上预测模型做检验，看是否能有效作出预测。这里以 2010 年的迁移流矩阵预测 2020 年迁移流矩阵作为检验，此时预测年（2020 年）的观测值及行、列边际总量值是已知的，因此，可用 RAS 方法直接预测，并进行观测值与预测值的比较，结果报告于表 6。表 6 中的 Panel A 部分是 2020 年长江经济带跨省人口迁移流矩阵预测值，Panel B 部分是实际值与预测值之比，比值趋近 1 则说明实际值与预测值较为吻合，即预测较为准

表 6 长江经济带 2020 年跨省人口迁移流矩阵预测检验及 2030 年预测矩阵												
迁出	迁入											
	上海	江苏	浙江	安徽	江西	湖北	湖南	重庆	四川	贵州	云南	n_{i+}
Panel A	2020 年预测值（万人）											
上海	0.0	11.3	5.4	4.2	2.0	1.2	0.7	0.8	1.5	0.3	0.5	28.1
江苏	179.0	0.0	49.3	23.9	6.3	9.2	4.4	4.5	7.4	2.7	3.1	289.8
浙江	44.3	31.0	0.0	10.6	9.5	11.4	6.8	5.7	9.7	4.7	7.3	141.0
安徽	268.6	311.5	285.7	0.0	10.8	12.7	5.4	3.5	5.7	2.9	3.9	910.8
江西	44.9	28.6	170.9	6.5	0.0	10.9	13.7	3.4	6.0	3.0	5.4	293.3
湖北	30.3	34.9	81.0	6.3	8.0	0.0	19.9	7.9	10.6	4.5	6.8	210.2
湖南	19.2	20.0	76.3	4.8	11.8	16.9	0.0	5.7	8.5	14.3	18.0	195.4
重庆	19.6	18.1	61.9	2.4	2.4	19.6	4.1	0.0	52.3	14.0	21.3	215.7
四川	56.2	68.6	135.2	6.9	6.4	13.6	9.4	106.5	0.0	27.9	50.2	481.0
贵州	21.8	47.6	267.5	8.5	10.2	7.0	13.8	24.7	17.2	0.0	44.6	462.9
云南	9.1	21.6	65.0	6.6	2.9	4.4	5.6	9.5	24.1	7.8	0.0	156.6
n_{+j}	693.2	593.1	1198.1	80.8	70.2	106.9	83.9	172.1	143.1	82.2	161.2	3384.8
Panel B	2020 年实际值与预测值之比											
上海	0	1.37	0.84	0.61	0.67	0.81	0.82	0.66	0.76	1.09	1.03	1.00
江苏	1.00	0	0.97	1.10	0.83	0.89	1.01	0.86	1.01	0.95	1.34	1.00
浙江	1.16	0.99	0	0.92	0.95	0.76	0.86	0.64	0.85	1.09	1.15	1.00
安徽	0.90	0.99	1.10	0	0.87	1.06	1.03	1.03	1.18	0.99	1.18	1.00
江西	1.12	1.12	0.92	0.99	0	1.03	0.91	1.11	1.22	1.27	1.47	1.00
湖北	1.38	1.01	0.69	1.22	1.26	0	1.11	1.18	1.23	1.17	1.42	1.00
湖南	1.24	0.81	0.94	1.01	1.02	1.02	0	1.09	1.19	1.07	1.02	1.00
重庆	0.96	0.95	0.94	1.28	1.34	0.92	1.37	0	1.07	0.97	1.03	1.00
四川	0.92	0.91	1.00	1.18	1.12	1.08	1.01	1.06	0	0.95	1.04	1.00
贵州	0.87	1.05	1.06	0.85	0.92	1.41	0.97	0.89	0.94	0	0.75	1.00
云南	1.52	1.20	1.07	0.71	1.19	1.10	0.79	0.67	0.70	0.87	0	1.00
n_{+j}	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Panel C	2030 预测值（万人）											
上海	0.0	20.2	6.0	3.6	1.8	1.3	0.8	0.8	1.5	0.5	0.6	37.1
江苏	235.1	0.0	64.1	36.8	7.1	10.7	6.0	5.5	10.1	3.3	4.8	383.5
浙江	62.5	37.6	0.0	12.7	11.4	10.4	7.3	4.8	10.3	6.1	9.2	172.4
安徽	319.6	409.2	424.0	0.0	12.9	17.6	7.6	5.1	9.1	3.8	5.4	1214.4
江西	63.6	41.1	205.4	8.7	0.0	14.2	16.2	5.2	9.5	4.8	9.1	377.8
湖北	55.1	46.8	76.0	10.9	13.9	0.0	30.0	13.3	17.7	6.9	11.5	282.1
湖南	32.0	22.0	98.8	6.9	16.8	23.0	0.0	9.1	14.0	20.2	22.2	265.1
重庆	26.1	24.0	82.5	4.7	4.7	24.9	8.1	0.0	79.3	18.5	27.3	299.9
四川	64.5	78.3	173.5	11.0	9.3	18.3	12.2	153.7	0.0	32.6	58.6	611.8
贵州	26.3	70.2	404.5	10.8	13.6	13.7	19.3	33.5	23.1	0.0	41.9	656.9
云南	19.6	36.9	100.2	7.1	5.0	6.8	6.4	9.8	24.4	9.5	0.0	225.7
n_{+j}	904.5	786.2	1635.1	113.1	96.4	140.9	113.9	240.8	199.0	106.1	190.7	4526.8

确。从该比值来看，大部分预测比较准确，其中江苏—上海、江苏—四川、湖北—江苏、四川—湖南等区域间迁移流量预测几乎完全一致。当然也有差距较大的预测值，比如重庆—湖南迁移流的实际值是预测值的 1.37 倍，上海—安徽迁移流的实际值是预测值的 0.61 倍，但这种情况并不多。

总体上看，RAS 方法能较为准确地预测区域间人口迁移流矩阵。因此，本文用中国人口普查 2000、2010 和 2020 年三期数据预测 2030 年人口迁移流矩阵，其中基期迁移流矩阵为 2020 年人口迁移流矩阵，得到长江经济带 2030 年跨省迁移总规模约为 4526.8 万人。通过总规模与预测得到的每个省份迁入比和迁出比测算即可得到 11 个省份的行、列边际总量，然后基于 RAS 方法即可预测得到 2030 年长江经济带跨省人口迁移流矩阵，结果报告于表 6 的 Panel C 部分。比如预计 2030 年江西到上海的人口迁移流量约为 63.6 万人，而上海到江西的人口迁移流量约为 1.8 万人。通过 RAS 预测，能够预知未来区域间人口迁移流的交互情况，比起传统的区域人口迁移总量预测有更加丰富的意义，深化了区域人口结构变化、空间交互联系研究的范畴。

六、结论与讨论

人口迁移流动是规范人口学研究的重要组成部分，但在中国生育制度的政策影响、老龄化及健康中国建设背景下，对生育、死亡（健康）的研究比人口迁移的影响要更深远。不过在中国即将进入人口负增长时期，人口迁移流动却表现出更加剧烈的脉动。流动人口规模急剧增长、空间结构日益复杂、跨区域交互流动剧增，已成为我国人口形势中最引人注目的特点之一。不过，关于中国的人口迁移流动研究更多地是作为经济社会运行中的一个外生因素，对其本身运行过程的分析还不足，因此，抛开相关性或因果关系的视角，从人口统计学的角度来研究人口迁移流动的内生规律依然有很大价值空间。

本文从人口统计学视角着手，基于人口迁移流量自我数据生成过程来研究我国跨区域人口迁移流的空间结构议题，重点在于分解人口迁移矩阵的主要成分并进行模型表达，基于功能选择性较强的广义线性模型进行模型拟合并寻找关键人口迁移的结构要素，同时跳出人口总量预测等传统视角，对我国人口迁移流矩阵或者说空间结构进行了预测。研究将人口迁移流空间结构的描述、模型表达和拟合预测纳入一个框架体系，通过整合三阶段统一框架的全景式刻画研究，形成了该议题分析的“基本程式”。应该说，本文深化了中国人口迁移研究的范畴，在研究视角和人口统计应用研究上有一定的创新性和拓展性。基于实证研究内容，得到的结论和讨论主要有以下三方面。

首先，列联表是人口迁移流空间结构的良好描述方式，对乘法分量模型的解释具有很好的作用，更重要的是能表达相关乘法分量的数据生成过程。乘法分量模型有很好的包容性，能将空间结构（迁出地、迁入地，甚至出生地等）、年龄结构、性别结构等均纳入分析框架，尽管增加分量类型会提升模型表达复杂程度，但其意义依然相对简单而丰富，尤其是交互效应分量对理解区域人口迁移流的内在结构非常有效。比如迁出地—迁入地交互效应分量所表征的区域间联系强度、迁出地—一年龄交互效应分量和迁入地—一年龄交互效应分量所反映

的空间一年龄差异等，更加深化了对于实证中长江经济带 11 省份和中国东部、中部、西部和东北四大区域的迁移流空间结构，也对于解释现实问题提供了新的有效视角。

其次，对数线性模型与乘法分量模型本质上是一致的，其中对数线性结构能基于功能选择性较强的广义线性模型原理进行模型拟合，对于寻找影响人口迁移流量的内生关键分量具有重要意义。结合人口迁移流量数据更加符合泊松分布、二项分布等而非正态分布的事实，本文实证分析发现中国四大区域人口迁移流的关键分量是年龄主效应 A ，其次迁出地年龄交互效应 OA 的影响也较为明显。综合比较，可用 $[ODA OA]$ 的简化模型来指征中国跨区域人口流动的空间特征，这四个参数基本可以刻画中国人口迁移结构，拟合程度达到 82%。

最后，关于人口迁移流矩阵预测在国内还比较罕见，本文基于国外较为成熟的双边比例调整法（又称 RAS 法）首次应用人口普查数据对我国长江经济带 11 省份之间的跨区域人口迁移流进行了预测检验，发现 RAS 法在人口迁移流矩阵预测上效果较好，并对 2030 年该区域内人口迁移流矩阵进行了预测。相比于传统人口总量的预测，通过 RAS 预测，能够预知未来区域间人口迁移流的交互情况，比起传统总量预测有更加丰富的意义，深化了区域人口结构变化、空间交互联系研究的范畴。

综合以上研究结论，人口迁移流的空间结构描述、数据内生过程和模型表达能丰富人口统计学理论，其基本原理能够进行延展分析，不局限于区域层级和人口学属性特征，比如大区、省级、地市级之间的人口迁移流，年龄、性别甚至婚姻、户籍等也均可纳入分析。本文的实证分析主要是案例解释，相关研究过程完全可以移植、延展。本文研究也有一定的政策意义，比如人口迁移流矩阵预测对于流动人口的数量分析、区域治理、协调发展、资源配置和规划提供一定参考。总之，人口迁移流的深化研究还有广阔空间，人口空间统计或空间人口学等理论、实证和政策研究还任重道远。

参考文献：

[1] STILLWELL J, DARAS K, BELL M, et al. The image studio: a tool for internal migration analysis and modelling [J]. Applied Spatial Analysis and Policy, 2014, 7 (1): 5-23.

[2] SHEN J. Error analysis of regional migration modeling [J]. Annals of the American Association of Geographers, 2016, 106 (6): 1253-1267.

[3] ROGERS A, WILLEKENS F, LITTLE J, et al. Describing migration spatial structure [J]. Papers in Regional Science, 2002, 81: 29-48.

[4] RAVENSTEIN E. The laws of migration [J]. Journal of the Royal Statistical Society, 1889, 52 (2): 241-305.

[5] LEWIS W A. Economic development with unlimited supply of labor [J]. The Manchester School, 1954, 22 (2): 139-191.

[6] RAYMER J. The estimation of international migration flows: a general technique focused on the origin - destination association structure [J]. Environment and Planning A, 2007, 39 (4): 985-995.

[7] STILLWELL J, DUKE-WILLIAMS O, DENNETT A. Technologies for migration and commuting analysis: spatial interaction data applications [M]. Hershey: Information Science Publishing, 2010: 1-30.

[8] SHRYOCK H S. Population mobility within the United States [R]. Community and Family Study Center, University of Chicago, 1964.

[9] CLAYTON C. The structure of interstate and interregional migration: 1965-1970 [J]. Annals of Regional Science, 1977, 11

- (1): 109-122.
- [10] PLANE D A. A systemic demographic efficiency analysis of U. S. interstate population exchange, 1935-1980 [J]. *Economic Geography*, 1984, 60 (4): 294-312.
- [11] SHEN J. Modelling regional migration in China: estimation and decomposition [J]. *Environment and Planning A*, 1999, 31 (7), 1223-1238.
- [12] RAYMER J, ANDREI R. Using age and spatial flow structures in the indirect estimation of migration streams [J]. *Demography*, 2007, 44 (2): 199-223.
- [13] RAYMER J, BONAGUIDI A, VALENTINI A. Describing and projecting the age and spatial structures of interregional migration in Italy [J]. *Population, Space and Place*, 2006, 12 (5): 371-388.
- [14] RAYMER J, BIDDLE N, CAMPBELL P. Analyzing and projecting indigenous migration in Australia [J]. *Applied Spatial Analysis & Policy*, 2017, 10 (2): 211-232.
- [15] ABEL G J, SANDER N. Quantifying global international migration flows [J]. *Science*, 2014, 343 (6178): 1520-1522.
- [16] WILLEKENS F. Log-linear modelling of spatial interaction [J]. *Papers in Regional Science*, 1983, 52 (1): 187-205.
- [17] FLOWERDEW R, LOVETT A. Fitting constrained Poisson regression models to interurban migration flows [J]. *Geographical Analysis*, 1988, 20 (4): 288-307.
- [18] LIU Y, SHEN J. Spatial patterns and determinants of skilled internal migration in China, 2000-2005 [J]. *Papers in Regional Science*, 2014, 93 (4): 749-771.
- [19] SHEN J. Changing patterns and determinants of interprovincial migration in China 1985-2000 [J]. *Population, Space and Place*, 2012, 18 (3): 384-402.
- [20] MALCOLM O A. Simulating the spatial distribution of population and emissions to 2100 [J]. *Environmental Resource Economic*, 2008, 39: 199-221.
- [21] 曾永明. 中国省际人口迁移的地缘效应与驱动机制: 男女有别吗? [J]. *人口研究*, 2017 (5): 40-51.
- [22] NAIR P S. Estimation of period-specific gross migration flows from limited data: bi-proportional adjustment approach [J]. *Demography*, 1985, 22 (1): 133-142.
- [23] 刘亚清, 李林鹏, 吴振信. 投入产出表更新调整方法及应用研究 [J]. *中国管理科学*, 2015 (S1): 706-710.
- [24] 王涛, 苏雅, 王晴晴. 中国省际贸易矩阵的估计与应用 [J]. *统计研究*, 2019 (4): 62-72.
- [25] 徐国祥, 陈海龙. 中国省区间人口投入产出表编制方法创新及实证研究 [J]. *统计研究*, 2019 (4): 5-18.
- [26] 汪子龙, 王柱, 於志文, 等. 多源数据跨国人口迁移预测 [J]. *浙江大学学报 (工学版)*, 2019 (9): 1759-1767.
- [27] AGRETI A. *Categorical data analysis (third edition)* [M]. Hoboken, New Jersey: John Wiley & Sons, Inc., 2013: 113-117.
- [28] BERNARD A, BELL M, CHARLES-EDWARDS E. Internal migration age patterns and the transition to adulthood: Australia and Great Britain compared [J]. *Journal of Population Research*, 2016, 33 (2): 1-24.
- [29] 林李月, 朱宇, 柯文前. 城镇化中后期中国人口迁移流动形式的转变及政策应对 [J]. *地理科学进展*, 2020 (12): 2054-2067.
- [30] 杨菊华. 空间理论视角下老年流动人口的社会适应 [J]. *社会学研究*, 2021 (3): 180-203, 229-230.
- [31] 林李月, 朱宇, 柯文前. 区域协调发展背景下流动人口回流的空间意愿及其政策含义 [J]. *地理研究*, 2021 (5): 1515-1528.
- [32] 刘丽群, 李汉学. 区域性推进高中阶段教育普及的战略定位与攻坚策略 [J]. *中国教育学刊*, 2020 (10): 34-40.
- [33] 王郅强, 赵昊骏. “候鸟式”养老群体的公共服务供需矛盾分析——以三亚市为例 [J]. *行政论坛*, 2019 (2): 103-109.

Research on the Spatial Structure of China's Interregional Migration Flow:
Description, Model Expression and Prediction

ZENG Yongming

(School of Economics, Jiangxi University of Finance and Economics,
Nanchang 330013, China)

Abstract: The research on the spatial structure and mathematical statistic of the cross-regional population flow is obviously insufficient, especially the statistical research on the operation law of the population migration data itself and the internalization process of the data is very rare. In this article, description, model expression and prediction are three separate systematic processes which should not be separated in the study of spatial flow structure of population migration. In view of this, this paper integrates the “basic formula” of describing, modelling and predicting the spatial structure of migration flows; multiplicative component model, log-linear model and bi-proportional adjustment method are used to describe, model and predict the spatial structure characteristics of migration flow in China. From the perspective of demographics, the empirical process applies the data of China Migration Dynamics Survey from 2011 to 2017 and the three census data from 2000 to 2020, and takes the population migration between the 11 provincial areas in the Yangtze River Economic Belt and the four major regions of China's east, central, west and northeast flows as an example. Based on the endogenous law of the migration flow matrix and the data self-generation process the spatial structure of cross-regional population migration flow in China is studied comprehensively. The main results show that the contingency table and its endogenous multiplicative component model can well describe the population migration flow with a two-dimensional structure. The unsaturated log-linear model with four-parameter $[O D A O A]$ can describe the cross-regional main characteristics of the population flow structure in China, with a fitting degree reaches 82%. The prediction of migration flow is of great significance, but the exploration is insufficient. The two sided proportional adjustment method can be used as a good tool for the population migration flow prediction. The main purpose of the article is to introduce and build a system framework or basic program for the study of the spatial structure of population migration, and to promote the research and application of spatial statistics or spatial population in China.

Keywords: migration flows; spatial structure; data self-generation process; demographic spatial statistics; research formulas

[责任编辑 刘爱华]