

# 基于空间滤波方法的中国省际人口迁移驱动因素

古恒宇<sup>1</sup>, 沈体雁<sup>1</sup>, 刘子亮<sup>2</sup>, 孟 鑫<sup>1</sup>

(1. 北京大学政府管理学院, 北京 100871; 2. 华南师范大学经济与管理学院, 广州 510006)

**摘要:** 人口迁移数据中往往存在较强的网络自相关性, 以往基于最小二乘估计的重力模型与迁移数据的拟合度较低, 而改进后的泊松重力模型仍存在过度离散的缺陷, 以上问题均导致既有迁移模型中的估计偏差。本文构建了特征向量空间滤波(ESF)负二项重力模型, 基于2015年全国1%人口抽样调查数据, 研究2010-2015年中国省际人口迁移的驱动因素。结果表明: ① 省际人口迁移流间存在显著的空间溢出效应, ESF能有效地提取数据中的网络自相关性以降低模型的估计偏差, 排序在前1.4%的特征向量即可提取较强的网络自相关信息。② 省际人口迁移流之间存在明显的过度离散现象, 考虑到数据离散的负二项重力模型更适用于人口迁移驱动因素的估计。③ 网络自相关性会导致模型对距离相关变量估计的上偏与大部分非距离变量估计的下偏, 修正后的模型揭示出以下驱动因素: 区域人口特征、社会网络、经济发展、教育水平等因素是引发省际人口迁移的重要原因, 而居住环境与公路网络等因素也逐渐成为影响人口迁移重要的“拉力”因素。④ 与既有研究相比, 社会网络因素(迁移存量、流动链指数)对人口迁移的影响日益增强, 而空间距离对人口迁移的影响进一步呈现弱化趋势。

**关键词:** 省际人口迁移; 空间滤波; 负二项重力模型; 驱动因素; 中国

DOI: 10.11821/dlxb201902002

## 1 引言

人口流动和迁移是影响中国未来数十年人口发展的关键议题。随着中国城市化进程的加快, 人口迁移成为各地区人口规模和结构变动的重要因素, 深刻影响着中国人口和城镇化的发展。在学界, 人口迁移问题长期以来受到人口学、地理学学者的关注。现有的人口迁移模型按照研究尺度可分为宏观模型与微观模型两类, 宏观迁移模型基于普查或相关统计数据, 着眼于区域总体人口或特定群体的迁移模式, 分析迁移过程的“推力”和“拉力”等驱动因素, 模拟迁移过程的变化, 甚至预测迁移流的大小。微观迁移模型主要关注个体或家庭的迁移行为, 试图解释潜在移民留在目前居住地或迁往其他地区的决策过程, 数据来源主要为普查或社会调查中的个体资料。由于微观数据较难获取等原因, 宏观迁移模型是学界研究的主流<sup>[1]</sup>。

重力模型(Gravity Model)是当前运用较多的宏观迁移模型, 最初的重力模型只含有迁入地人口规模、迁出地人口规模、迁移距离3个变量。而基于“推拉理论”改进的

收稿日期: 2017-11-09; 修订日期: 2018-11-29

基金项目: 国家社会科学基金重大项目(17ZDA055); 国家自然科学基金项目(71473008); 国家自然科学基金重大项目(71733001) [Foundation: National Social Science Foundation of China, No.17ZDA055; National Natural Science Foundation of China, No.71473008, No.71733001]

作者简介: 古恒宇(1994-), 男, 广东河源人, 博士生, 研究方向为空间人口学、区域分析与规划。

E-mail: henry.gu@pku.edu.cn

通讯作者: 沈体雁(1971-), 男, 湖北天门人, 教授, 博导, 研究方向为城市规划与区域经济。E-mail: tyshen@pku.edu.cn

拓展重力模型通过考虑一系列人口和经济社会指标作为人口迁移的驱动因素,提高了模型对人口迁移过程的解释能力。近年来,对重力迁移模型的研究主要集中在两方面:①模型形式拓展:将重力模型与前沿计量模型(如泊松模型)结合,通过一系列计量手段,分解模型中被解释变量的效应,分析模型的误差来源。Shen运用多层泊松迁移模型分析中国1985-2000年人口迁移的影响因素,并对人口迁移数量的增加进行分解,发现62.28%的增加是因为解释变量的变化,而37.72%的增加是因为模型参数的变化<sup>[2]</sup>。②研究对象拓展:基于重力模型研究具体化、细分后的各类人口迁移现象,如技术迁移、性别迁移。Liu等探讨了中国省际技术迁移和低技术迁移的影响因素,发现技术性迁移受到距离、失业率、外商直接投资的影响相对较小,而受到实际工资水平的影响相对较大<sup>[3]</sup>。

重力模型隐含着—个基本假设,即模型各变量是相互独立的,而通过迁移距离变量便可消除区域起始地和目的地间的空间依赖效应。然而该假设受到了一些学者的质疑<sup>[4-6]</sup>。Griffith等发现,加拿大人口流动中起始地的出行人口数量会受到周围人口的出行倾向的影响,而指向同一目的地的出行流量同样受目的地周围地区吸引力偏好的影响,即人口流动网络中存在空间溢出效应<sup>[6]</sup>。在迁移流层面,空间自相关性表征为基于拓扑格局的网络自相关效应(Network Autocorrelation),若其在模型中未能得到很好地处理,估计结果有偏<sup>[7]</sup>,则需要通过引入空间权重矩阵,将空间溢出效应纳入模型进行估计,从而得到空间计量交互模型。空间计量交互模型常采用空间滞后模型的形式<sup>[5]</sup>,通过引入空间滞后项对空间溢出效应进行估计,而对数据中由其他原因导致的空间溢出的估计并不充分。另一种处理方法试图通过某种算子“过滤”样本数据中的网络自相关效应,这种处理方法被称为空间滤波(Spatial Filtering)。空间滤波方法不受模型前提假设的限制,通过调整滤波器算子,通常能更彻底地降低误差项中的网络自相关效应。现有的空间滤波方法主要包括自回归线性法、Getis's G法、特征向量空间滤波法(Eigenvector Spatial Filtering, ESF)3种。比较而言,ESF不仅数据限制小,且更适用于处理模型内生性问题。近几年,ESF被逐渐运用在国外人口迁移的研究中。Chun等使用美国洲际公路年度移民数据,基于ESF构建的网络自相关线性模型和广义线性混合模型研究了美国洲际人口迁移驱动因素,结果表明对网络自相关性的处理提升了模型的拟合能力<sup>[8]</sup>。

在中国省际人口迁移方面已积累了大量的研究成果,研究热点主要集中在人口迁移的空间格局及其演变<sup>[9]</sup>、人口迁移的集聚测度<sup>[10]</sup>、人口迁移的影响因素<sup>[11]</sup>以及人口迁移与地区经济发展的关系<sup>[12]</sup>等方面。近年国内学者开始关注人口迁移中空间溢出效应的影响<sup>[1, 13-14]</sup>。例如,张红历等基于新经济地理学的相关成果,将市场潜能、预期收入等因子加入在传统的重力模型中,运用空间滞后模型和空间误差模型对中国省际人口迁移因素进行分析,发现市场潜能和预期收入对人口迁入有显著的促进作用,但对人口迁移的作用路径各异<sup>[13]</sup>。蒲英霞等基于中国2010年第六次人口普查数据,运用空间计量交互模型对中国省际人口迁移因素进行分析,通过空间滞后项技术捕捉了省际人口迁移中多种变量的多边效应机制,发现诸如空间距离、区域工资水平等变量在空间上均表现出一定的溢出效应<sup>[1]</sup>。

通过综述文献发现,国内相关研究主要采用空间计量交互模型的研究方法,而基于空间滤波的研究案例比较欠缺;此外,基于最小二乘法估计(OLS)的传统重力模型仍是主流,而基于最大似然估计(MLE)且与人口迁移数据拟合度更高的泊松模型(PGM)、负二项模型(NBGM)等模型的运用较少。本文基于2015年全国1%人口抽样调查数据,尝试将空间滤波方法运用在拓展重力模型中,并结合空间句法(Space Syntax)等分析手段,对中国2010-2015年省际人口迁移的驱动因素进行研究。

## 2 数据来源与变量选择

### 2.1 数据来源

本文使用的人口迁移数据来源于《2015年全国1%人口抽样调查资料》<sup>[15]</sup>截面长表数据中的“全国按现住地和五年前常住地分的人口”。研究对象包括全国31个省（自治区、直辖市），剔除缺失值后共926条迁移流（不包含港澳台地区和省内迁移流）。考虑到迁移人数的实际比例与数据的整数性质，因变量按照人口抽样比（1.55%）还原，并进行四舍五入取整处理。2010-2015年中国省际人口迁移出发—目的地（Origin-destination, O-D）流地理空间格局和拓扑格局如图1所示。

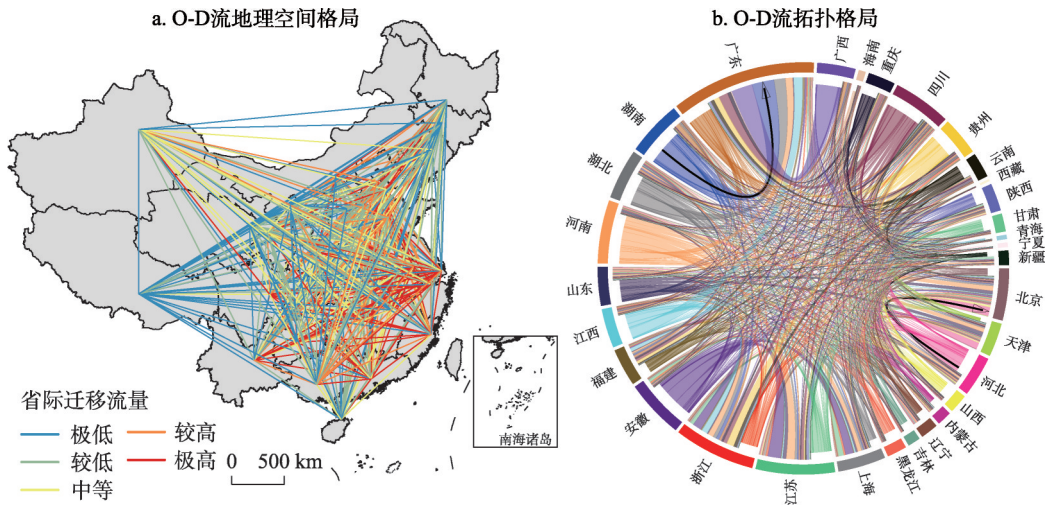


图1 2010-2015年中国省际人口迁移O-D流地理空间格局和拓扑格局图

Fig. 1 Geographical pattern and topological pattern of Chinese interprovincial migration flows from 2010 to 2015

### 2.2 变量选择及预处理

参考国内外相关理论及研究成果<sup>[9-14]</sup>，人口迁移驱动因素主要包括人口特征因素<sup>[9-10]</sup>、经济发展因素<sup>[11-14]</sup>、居住环境因素<sup>[12]</sup>、教育水平因素<sup>[9, 12]</sup>、迁移距离因素<sup>[14]</sup>等类别。考虑到中国近年来交通基础设施的大规模建设，以及既有文献中较为缺乏对以公路为主的交通基础设施的分析，本文在上述因素的基础上加入公路网络因素，共6类驱动因素进行研究。对所有模型中变量之间是否存在多重共线性进行检验，若自变量的方差膨胀因子（VIF）大于10，则存在明显的多重共线性，模型中未通过共线性检验的变量将被剔除，最终共选择迁入地及迁出地人口规模、迁出地性别比、迁移存量、流动链指数、迁出地外商直接投资占地区生产总值比、迁入地及迁出地近五年GDP平均增速、迁入地及迁出地人口密度、迁入地及迁出地人均住房面积、迁入地及迁出地平均受教育年限、迁入地局部接近度、迁入地全局接近度、迁入地全局穿行度、空间距离等18个变量。考虑到经济发展对人口迁移的影响具有滞后效应，2个经济发展变量均使用前一期的变量。同一区域作为迁出地和迁入地时选用相同解释变量，通过角标加以区分（如 $POP_i$ 、 $POP_j$ ）。变量描述、预期效应及数据来源如表1所示。其中，各省平均受教育年限根据胡鞍钢等<sup>[16]</sup>中的算法计算<sup>①</sup>。公路网络数据来源于2016年出版的《中国交通地图册》<sup>[17]</sup>中高速公

① 平均受教育年限=(大专以上人口数×16+高中人口数×12+初中人口数×9+小学人口数×6)/总人口数

表1 变量描述、预期效应与数据来源  
Tab. 1 Descriptions, expected effects and data sources of major explanatory

| 变量                      | 描述  | 预期效应         |              |
|-------------------------|---|--------------|--------------|
|                         |   | 迁出地 <i>i</i> | 迁入地 <i>j</i> |
| 人口特征因素：                 |   |              |              |
| 人口规模( <i>POP</i> )      | 2015年各省总人口(人) <sup>a</sup>  | +            | +            |
| 性别比( <i>SEX</i> )       | 2015年各省性别比 <sup>b</sup>   | +            |              |
| 社会网络因素：                 |   |              |              |
| 迁移存量( <i>FLOW</i> )     | 2005-2010年各省人口迁移存量(人) <sup>c</sup>  | +            | +            |
| 流动链指数( <i>MSTOCK</i> )  | 2010-2015年各省迁出人口在各迁出地的构成比( <i>i</i> 省流向 <i>j</i> 省的迁移人数/ <i>i</i> 省的总迁出人数) <sup>b</sup> |              |              |
| 经济发展因素：                 |   |              |              |
| 外商直接投资比例( <i>FDI</i> )  | 2010年各省外商直接投资占GDP比重 <sup>a</sup>  | -            |              |
| GDP增速 <i>GDPI</i>       | 2011-2015年各省平均GDP增长率 <sup>a</sup>   | -            | +            |
| 居住环境因素：                 |   |              |              |
| 人口密度( <i>DEN</i> )      | 2015年各省城镇人口密度(人/km <sup>2</sup> ) <sup>a</sup>  | -            | +            |
| 人均住房面积( <i>AREA</i> )   | 2015年各省城镇居民人均住房建筑面积(m <sup>2</sup> ) <sup>a</sup>                                       | -            | +            |
| 教育水平因素：                 |   |              |              |
| 平均受教育年限( <i>ASCHO</i> ) | 2015年各省平均受教育年限(年) <sup>b</sup>  | -            | +            |
| 公路网络因素：                 |   |              |              |
| 局部接近度( <i>C100</i> )    | 2015年各省公路网络平均局部接近度  |              | +            |
| 全局接近度( <i>C2000</i> )   | 2015年各省公路网络平均全局接近度  |              | +            |
| 全局穿行度( <i>B2000</i> )   | 2015年各省公路网络平均全局穿行度  |              | 未知           |
| 迁移距离因素：                 |   |              |              |
| 空间距离( <i>DIST</i> )     | 迁入地和迁入地省会城市之间空间距离(km)   | -            |              |

注：数据来源：a.《中国统计年鉴(2016)》<sup>[18]</sup>；b.《2015年全国1%人口抽样调查资料》<sup>[15]</sup>、《中国2010年第六次人口普查数据》<sup>[19]</sup>。

路、国道和省道图，在GIS平台中进行地理配准及矢量化处理后进行空间句法计算，并将线要素结果汇总至各省域面的结果作为解释变量。

3 研究方法

3.1 重力模型

**3.1.1 泊松重力模型(PGM)** 在研究在大规模迁移流时，传统基于最小二乘的估计方法与样本的拟合程度不高<sup>[20]</sup>。此外，由于各个离散空间单元的迁移人数具有非负整数性质，因而泊松模型被逐渐运用，并被证实更加符合人口迁移的实际情况<sup>[21]</sup>。假设迁移人数  $M_{ij}$  满足泊松（Poisson）分布：

$$\Pr(M_{ij}) = \frac{\exp(-\mu_{ij})\mu_{ij}^{M_{ij}}}{M_{ij}!}, \quad (M_{ij}=0, 1, \dots) \tag{1}$$

条件均值  $\mu_{ij}$  与一个具有多变量的外生函数相关：

$$\mu_{ij} = \exp(\alpha_0 + \sum_{q=1}^Q \alpha_{1q} \ln X_{iq} + \sum_{l=1}^L \alpha_{2l} \ln X_{jl} + \alpha_3 \ln D_{ij}) \tag{2}$$



式中:  $X_{iq}$  为迁出地  $i$  的第  $q$  个变量;  $\alpha_{iq}$  为迁出地中第  $q$  个变量的系数;  $X_{jl}$  为迁入地  $j$  的第  $l$  个变量;  $\alpha_{jl}$  为迁入地中第  $l$  个变量的系数。

**3.1.2 负二项重力模型(NBGM)** 泊松回归模型的基本假设为离散平衡(Equidispersion), 即因变量的方差与平均值相等。然而, 根据迁移数据的实际情况, 不仅每条迁移流间差异巨大, 而且同一迁移流的不同年份的数据也有较大不同, 离散平衡的假设往往不成立, 从而导致过度离散(Overdispersion)问题。若沿用泊松回归模型, 则将导致有偏估计。负二项回归模型通过引入参数  $\alpha$ , 以测度过度离散水平<sup>[3]</sup>。条件均值  $\mu_{ij}$  同(2)式, 但假设条件与泊松回归不同, 模型中方差应满足一个关于条件均值  $\mu_{ij}$ 、离散系数  $\alpha$  的函数:

$$\text{Var}(M_{ij}) = \mu_{ij} + \alpha \mu_{ij}^2 \quad (3)$$

迁移人数  $M_{ij}$  应满足泊松—伽马(Poisson-Gamma)混合分布:

$$\text{Pr}(M_{ij}) = \frac{\Gamma(M_{ij} + \alpha^{-1})}{M_{ij}! \Gamma(\alpha^{-1})} \left( \frac{\alpha^{-1}}{\alpha^{-1} + \mu_{ij}} \right)^{\alpha^{-1}} \left( \frac{\mu_{ij}}{\alpha^{-1} + \mu_{ij}} \right)^{M_{ij}} \quad (4)$$

式中:  $\Gamma$  表示标准的伽马分布函数;  $\alpha$  为过度离散系数,  $\alpha$  的数值越大, 离散水平越高, 当  $\alpha = 0$ , 即不存在过度离散时, 负二项模型退化为泊松回归模型。

### 3.2 特征向量空间滤波(ESF)

ESF 的思想本质上是从特定的空间结构(如中国省际人口迁移网络)中提取表达空间结构信息的特征向量, 并将筛选后的特征向量作为因变量中网络自相关性的控制变量加入模型进行回归。由于各特征向量一定程度上代表了潜在的网络自相关信息, 特征向量作为解释变量即相当于“过滤”了来源于空间结构的网络自相关性对模型估计的影响。

步骤具体如下, 首先构建一个转化后的空间权重矩阵:

$$(I - AA^T/n)S(I - AA^T/n) \quad (5)$$

式中:  $I$  是  $n \times n$  维单位矩阵;  $A$  表征数值 1 的  $n \times 1$  维矩阵;  $S$  为  $n \times n$  维空间权重矩阵。

当空间权重矩阵用于表征网络关系时, 考虑到省际人口迁移的实际数为  $n^2 - n$  条, 转化矩阵可被表达为:

$$(I_{n^2-n} - A_{n^2-n} A_{n^2-n}^T / n^2 - n) S^N (I_{n^2-n} - A_{n^2-n} A_{n^2-n}^T / n^2 - n) \quad (6)$$

式中:  $I_{n^2-n}$  是  $(n^2 - n) \times (n^2 - n)$  维单位矩阵;  $A$  是数值 1 的  $(n^2 - n) \times 1$  维矩阵;  $S^N$  是经过转换后的  $(n^2 - n) \times (n^2 - n)$  维的网络权重矩阵, 其原始表达如下:

$$S_{ij,kl}^N = \begin{cases} 1, & i=k \text{ 且 } S_{jl}=1; \text{ 或 } j=l \text{ 且 } S_{ik}=1 \\ 0, & \text{其他} \end{cases} \quad (7)$$

式中:  $S_{ij,kl}^N$  表达迁出地  $i$  至迁入地  $j$  间的迁移流与迁出地  $k$  至迁入地  $l$  间迁移流的关系;  $S_{ik}(S_{jl})$  为  $n \times n$  维空间权重矩阵  $S$  中的对应数值, 表示各省的相邻关系。

$S^N$  的构建可参照克罗内克算法公式<sup>[22]</sup>:

$$S_{ij,kl}^N = S \oplus S = I_n \otimes S + S \otimes I_n \quad (8)$$

根据省际人口迁移的实际情况, 原始的网络权重矩阵需剔除与省内迁移相关的元素。

特征向量中的各元素可代表对应地理单元上的空间结构信息, 基于各元素可计算各特征向量对应的莫兰指数(MC)<sup>[23]</sup>可以证明, 对应最大特征值的特征向量  $E_1$  同时对应最大的 MC; 对应排名第二特征值的特征向量  $E_2$  同时对应排名第二的 MC, 以此类推到  $E_n$ , 这反映出各特征向量具有排序性的性质<sup>[24]</sup>。此外, 各特征向量具备正交性(orthogonality)

的性质, 即它们之间是互不相关的 (uncorrelatedness)。总的来说, 各特征向量分别反映在特定的空间结构下网络自相关性的一种潜在的可能形式。而因变量中的网络自相关性的影响, 在一定程度上可以看作是各特征向量反映的网络自相关信息的线性组合<sup>[23]</sup>。在这个意义上, 在给定网络权重矩阵  $S^N$  时, 特征向量能表达因变量中潜在的网络自相关性的强弱。

### 3.3 空间句法 (Space Syntax)

本文使用空间句法计算各省份的公路网络特征, 并作为解释变量引入重力模型中, 以探究公路网络特征对省际人口迁移的影响。空间句法从非欧氏距离的建模角度分析空间和功能的关系, 其线段模型考虑到路网偏转角度和人口迁移的相关性, 通过对道路的建模分析, 揭示人们受公路网络影响下的迁移规律<sup>[25]</sup>。sDNA 模型是前沿的空间句法模型, 模型主要围绕接近度和穿行度两个变量展开分析<sup>[26]</sup>。

(1) 搜索半径 ( $R$ ): 表征计算某路段路网形态变量时考虑的空间范围, 搜索半径越大, 说明该路段路网形态变量是由较大范围内的路网特征计算而来, 根据中国省际人口迁移距离特征, 设定 100 km 与 2000 km 两个搜索半径, 以表征中微观尺度和宏观尺度下公路网络对省际人口迁移的影响。

(2) 接近度 ( $C$ ): 接近度代表某路网迁移到搜索半径内其余路网的难易程度, 接近度高的路网通常具有较高可达性, 对区域迁移流具有更大的吸引力。计算公式如下:

$$Closeness(x) = \sum_{y \in R_x} \frac{p(y)}{d(x, y)} \quad (9)$$

式中:  $p(y)$  为搜索半径  $R$  内节点  $y$  的权重,  $p(y) \in [0, 1]$ ;  $d(x, y)$  为节点  $x$  到节点  $y$  的最短拓扑距离;  $Closeness(x)$  为接近度。

(3) 穿行度 ( $B$ ): 穿行度用于衡量路网被搜索半径内迁移流通过的概率, 穿行度越高代表路网的通过性越强, 相应地便承载着较多的通过性迁移流。计算公式如下:

$$Betweenness(x) = \sum_{y \in N} \sum_{z \in R_y} OD(y, z, x) \frac{P(z)}{Links(y)} \quad (10)$$

式中:  $OD(y, z, x)$  为搜索半径  $R$  内通过节点  $x$  的节点  $y$  与  $z$  之间最短拓扑路径;  $Betweenness(x)$  为节点  $x$  的穿行度;  $Links(y)$  为对于元素  $y$  半径  $R$  内的节点总数。

### 3.4 特征向量空间滤波负二项重力模型框架 (ESF NBGM)

在 MATLAB 编程环境下, 依据公式 (7) 的方法构建网络权重矩阵 (由于海南省曾归广东省管辖, 定义广东和海南相邻), 并计算对应的特征向量及特征值。最终, 本文构建了基于特征向量空间滤波的负二项重力模型框架 (ESF NBGM):

$$\mu_{ij} = \exp \left( \alpha_0 + \sum_{q=1}^Q \alpha_{1q} \ln X_{iq} + \sum_{l=1}^L \alpha_{2l} \ln X_{jl} + \alpha_3 \ln D_{ij} + \sum_{k=1}^K \alpha_{4k} E_k \right) \quad (11)$$

式中: 新加入的  $E_k$  为经过挑选后的特征向量的第  $k$  个组合;  $\alpha_{4k}$  为特征向量变量中第  $k$  个系数; 其余变量已在公式 (2) 说明。

回归模型中, 特征向量的加入主要经过以下两个步骤: ① 模型将提取与  $MC_{\max}$  比值大于或等于 0.25 的  $MC$  所对应的特征向量; ② 采用逐步回归法 (向前), 选取显著表达网络自相关性的代理变量进入模型进行回归分析 ( $p \leq 0.01$ )。本文共构造 4 个模型以探究中国省际人口迁移驱动因素, 分别为泊松重力模型、ESF 泊松重力模型、负二项重力模型、ESF 负二项重力模型。

## 4 中国省际人口迁移驱动因素分析

### 4.1 中国省际人口迁移流间存在显著的空间溢出效应

通过分析发现中国省际人口迁移流特征向量空间滤波能很好地识别和提取网络权重矩阵中的网络自相关性,降低因数据中网络自相关性而导致的估计偏差。少数特征值较高的特征向量即可提取数据中网络自相关性较强的空间格局。通过合理设定阈值,模型可以达到不同的过滤效果。

2010-2015年中国省际人口迁移的 $MC$ 结果显示,中国省际人口迁移网络的 $MC$ 值为0.378 ( $p \leq 0.001$ ),数据中存在显著的空间溢出效应。对比负二项重力模型与ESF负二项重力模型的结果,赤池信息准则(AIC)与贝叶斯信息准则(BIC)可作为模型网络自相关提取能力的判别指标。泊松重力模型中加入的148个特征向量及负二项重力模型中,加入的10个特征向量,分别导致模型各自的 $AIC$ 值下降了1162719与70, $BIC$ 值则下降了1162633与15,同时,离散系数 $\alpha$ 下降至0.137。因此,空间滤波能较好抓取数据中的网络自相关性,从而减少因自相关带来的估计误差。

利用ESF过滤数据中的网络自相关性,将提取出来的930个特征向量按照特征值的大小关系进行排序。选取 $MC$ 与 $MC_{\max}$ 比值约为1、0.75、0.5、0.25的特征向量(记为 $E_1$ 、 $E_{13}$ 、 $E_{55}$ 、 $E_{190}$ ),并根据特征向量中各代表空间结构信息的元素,以“分位数”法(Quantile)进行分级地图表达。各特征向量中不同区位上的空间结构信息可构成该特征向量对应的网络自相关格局。图2中,暖色线段对应着具有较高空间结构信息的O-D流,冷色线段对应着具有较低空间结构信息的O-D流。图2a描述了 $E_1$ 、 $E_{13}$ 、 $E_{55}$ 、 $E_{190}$ 四个特征向量对应的网络自相关格局,其中, $E_1$ 对应的格局具有较高的网络自相关性, $MC$ 值达到1.108。随着 $MC$ 与 $MC_{\max}$ 比值的下降,特征向量对网络自相关性的表达也减弱,对于 $E_{190}$ 的空间格局,其 $MC$ 值降至0.282。为更清晰地观察特征向量对应的空间格局中网络自相关信息的强弱,从 $E_1$ 、 $E_{13}$ 、 $E_{55}$ 、 $E_{190}$ 四个特征向量中筛选出指向陕西省和河南省的空间结构信息O-D流,并对此进行图示表达。如图2b、2c所示,排序较高的特征向量 $E_1$ 对应的空间格局中,目的地相同且出发地相邻的O-D流数值更为接近,网络相关性较高,如广东省和湖南省指向陕西省的空间结构信息数值都较低;排序较低的特征向量 $E_{190}$ 对应的空间格局中,目的地相同却出发地相邻的O-D流差异性较大,网络自相关性较低,如广东省指向陕西省的空间结构信息数值较低,而湖南省较高。从图2可直观看出,在空间结构(网络权重矩阵)确定的情况下,通过不同的特征向量,能达到揭示网络自相关信息的目的。

为了检验各特征向量对网络权重矩阵中网络自相关格局的表达差异,将 $MC/MC_{\max}$ 的值与其对应的特征值排序进行位序—规模分析,如表2、图3所示,排名约在前20%的特征向量具有一定的提取网络自相关信息的能力,排名约在前1.4%的特征向量能揭示出较高的网络自相关格局。以上结果说明,少数排序较前的特征向量即可提取数据中较强的网络自相关信息,而将这部分特征向量作为解释变量进入模型进行回归,往往不会给模型增加过多的计算量。只要合理调整阈值,控制进入模型的特征向量个数,就能达到不同的过滤效果,这从一个侧面说明了ESF空间滤波的灵活性。

### 4.2 考虑到过度离散问题的负二项重力模型更具适用性

省际人口迁移流之间存在明显的过度离散现象,违背了泊松重力模型中对解释变量方差与均值相等的基本假设,考虑到数据离散程度的负二项重力模型更适用于省际人口迁移的计量分析。



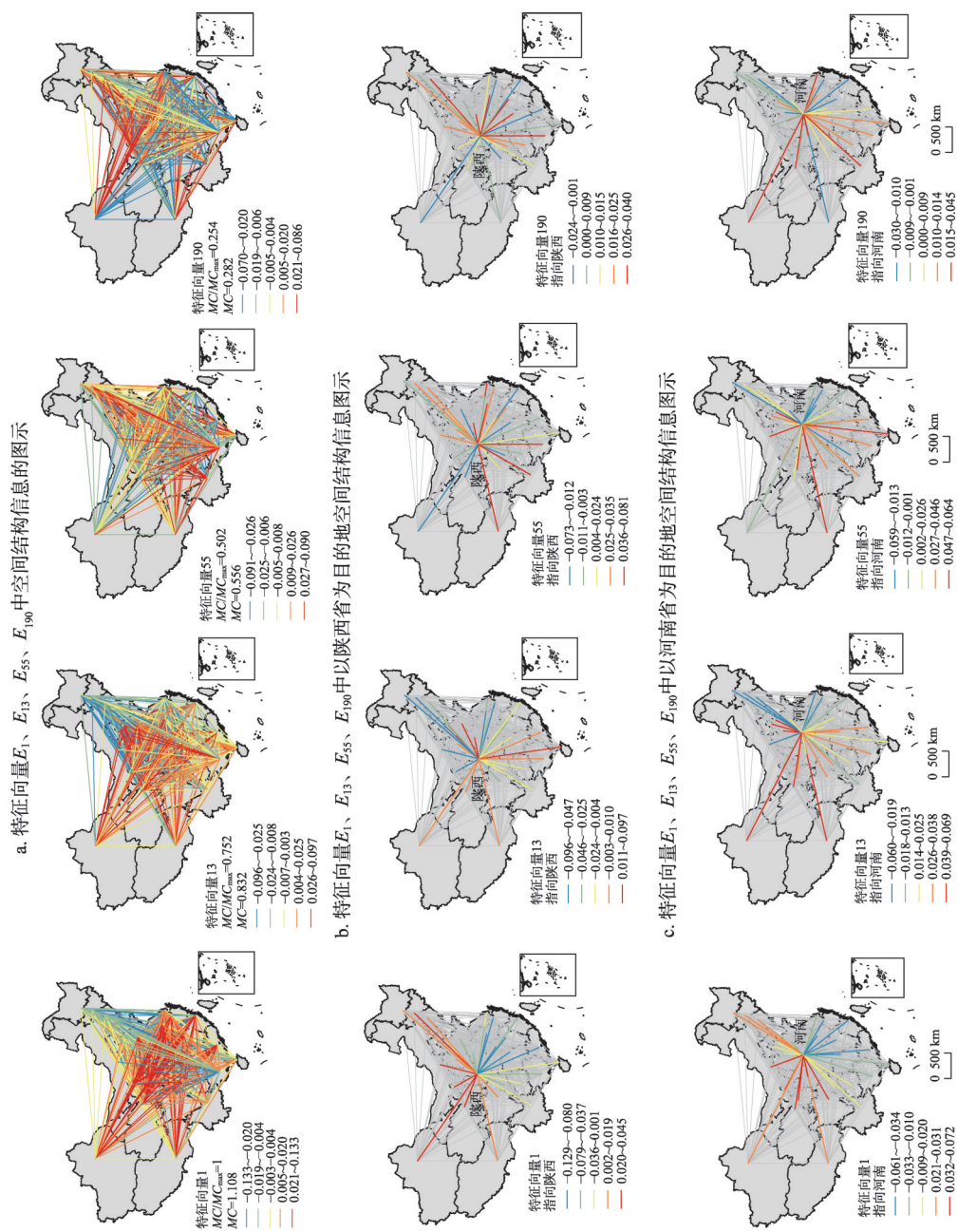


图2 各特征向量对网络权重矩阵中空间结构信息的提取差异

Fig. 2 Extraction of separate spatially structured components in the space weight matrix by eigenvectors



表2 特征向量对网络权重矩阵中网络自相关性的提取能力

Tab. 2 ESF's ability of capturing network autocorrelation in the space weight matrix

| 网络自相关信息提取能力 | $MC/MC_{\max}$ 取值范围                  | 特征向量个数 | 比例(%)  |
|-------------|--------------------------------------|--------|--------|
| 较高          | $MC/MC_{\max} \in (0.75, 0.1]$       | 13     | 1.398  |
| 中等          | $MC/MC_{\max} \in (0.5, 0.75]$       | 44     | 4.732  |
| 较低          | $MC/MC_{\max} \in (0.25, 0.5]$       | 134    | 14.408 |
| 极低          | $\{E = MC/MC_{\max}   E \leq 0.25\}$ | 739    | 79.462 |

对比泊松重力模型、负二项重力模型及其相应的空间滤波模型的计算结果（表3）。直观上，被解释变量的方差达到其均值的40万倍，与泊松回归要求方差与均值相等的假设不一致。负二项模型的离散系数 $\alpha$ 为0.158，其LR检验在千分之一水平上显著，强烈拒绝系数为0的原假设，同样证明数据中存在过度离散。本文借鉴 Hilbe<sup>[27]</sup>的方法，构建一个离散统计量 $D$ （ $\text{Pearsonchi}^2/\text{自由度}$ ）以测度泊松模型与负二项模型中数据的离散程度<sup>[27]</sup>。大样本情况下，若 $D$ 值超过1.05则代表数据的离散程度过大。结果显示，泊松重力模型（模型1）及ESF泊松重力模型（模型2）中 $D$ 值分别为2907.653、1809.796，统计量远超过阈值，继续沿用泊松模型会导致系数估计的过度显著现象。负二项重力模型（模型3）和ESF负二项重力模型（模型4）中 $D$ 值则分别为0.164、0.146，表明负二项模型能够很好地解决过度离散的问题。最后，对比各模型的对数似然值，发现模型3与模型4的对数似然值较模型1与模型2分别呈现出大幅度的提升，说明负二项模型拟合程度较泊松模型而言更高。通过上述对比，本文的分析应基于ESF负二项重力模型展开。

4.3 人口、社会、经济和教育等因素的影响机制分析

网络自相关性会导致距离相关变量估计的上偏与大部分非距离变量估计的下偏。通过空间滤波校正后的模型显示，区域人口特征、社会网络、经济发展、教育水平等因素对省际人口迁移产生重要影响，其中，迁出地性别比较高、流动链指数较大、迁移存量较多、迁入地GDP平均增速较快、迁入地平均受教育年限较长的省份对人口迁移数量的促进作用较大。

对比负二项重力模型与ESF负二项重力模型结果（表3），发现通过空间滤波处理后模型中4个距离相关变量（ $C100$ 、 $C2000$ 、 $B2000$ 、 $DIST$ ）的回归系数均下降，如局部接近度与迁移距离的下降幅度均超过0.05；另一方面，大部分非距离变量的回归系数均上升。由于网络自相关性体现了“相邻”的概念，若不对其进行处理，会夸大距离相关变量的影响以及弱化非距离变量的效应。通过空间滤波方法对数据中的网络自相关性进行过滤后，人口特征、社会网络、经济发展、教育水平仍是影响省际人口迁移的重要驱动因素。

4.3.1 人口特征因素 人口规模（POP）对省际人口迁移产生较大的正向影响，符合理论

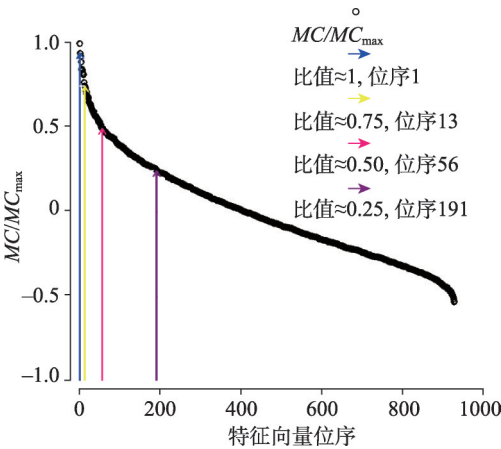


图3 特征向量及其对应的网络自相关性的位序—规模图

Fig. 3 Order-Scale pattern of eigenvectors and its autocorrelation

表3 中国省际人口迁移泊松重力模型、ESF 泊松重力模型、负二项重力模型、ESF 负二项重力模型结果

Tab. 3 Results of interprovincial migration in China, based on PGM, ESF PGM, NBGM and ESF NBGM

| 变量名称          | (1) 泊松重力模型 |       | (2) ESF 泊松重力模型 |       | (3) 负二项重力模型 |       | (4) ESF 负二项重力模型 |       |
|---------------|------------|-------|----------------|-------|-------------|-------|-----------------|-------|
|               | 系数         | P值    | 系数             | P值    | 系数          | P值    | 系数              | P值    |
| $POP_i$       | 0.248***   | 0.000 | 0.335          | 0.000 | 0.377***    | 0.000 | 0.382***        | 0.000 |
| $POP_j$       | 0.12***    | 0.000 | 0.139          | 0.000 | 0.189***    | 0.000 | 0.174***        | 0.000 |
| $SEX_i$       | 1.973***   | 0.000 | 2.725          | 0.000 | 0.723       | 0.089 | 1.152**         | 0.006 |
| $FLOW_{ij}$   | 0.689***   | 0.000 | 0.687          | 0.000 | 0.617***    | 0.000 | 0.627***        | 0.000 |
| $MSTOCK_{ij}$ | 0.565***   | 0.000 | 0.729          | 0.000 | 1.156***    | 0.000 | 1.201***        | 0.000 |
| $FDI_i$       | -0.033***  | 0.000 | -0.042         | 0.000 | -0.054***   | 0.001 | -0.072***       | 0.000 |
| $GDPI_i$      | -0.063***  | 0.000 | -0.104         | 0.000 | 0.106       | 0.323 | 0.143           | 0.174 |
| $GDPI_j$      | 1.057***   | 0.000 | 0.681          | 0.000 | 0.933***    | 0.000 | 0.995***        | 0.000 |
| $DEN_i$       | 0.073***   | 0.000 | 0.190          | 0.000 | 0.066       | 0.075 | 0.037           | 0.318 |
| $DEN_j$       | 0.016***   | 0.000 | 0.049          | 0.000 | 0.148***    | 0.000 | 0.151***        | 0.000 |
| $AREA_i$      | -0.051***  | 0.000 | -0.673         | 0.000 | -0.289      | 0.051 | -0.204          | 0.154 |
| $AREA_j$      | 0.847***   | 0.000 | 1.384          | 0.000 | 0.645***    | 0.000 | 0.577***        | 0.000 |
| $ASCHO_i$     | 0.155***   | 0.000 | -0.029         | 0.000 | 0.642***    | 0.000 | 0.781***        | 0.000 |
| $ASCHO_j$     | 0.744***   | 0.000 | 0.100          | 0.000 | 1.013***    | 0.000 | 0.986***        | 0.000 |
| $C100_j$      | 0.352***   | 0.000 | 0.472          | 0.000 | 0.328***    | 0.000 | 0.252***        | 0.001 |
| $C2000_j$     | -0.021***  | 0.000 | -0.016         | 0.000 | -0.027***   | 0.000 | -0.021***       | 0.001 |
| $B2000_j$     | 0.002***   | 0.000 | 0.000          | 0.002 | 0.002**     | 0.005 | 0.002**         | 0.008 |
| $DIST_{ij}$   | -0.051***  | 0.000 | -0.063         | 0.000 | -0.114***   | 0.000 | -0.090***       | 0.000 |
| Constant      | -10.722*** | 0.000 | -14.759***     | 0.000 | -5.197*     | 0.026 | -7.282***       | 0.001 |
| 特征向量数         | -          |       | 148            |       | -           |       | 10              |       |
| 对数似然值         | -1255010   |       | -673502        |       | -9576       |       | -9544           |       |
| AIC           | 2510057    |       | 1347338        |       | 18994       |       | 18924           |       |
| BIC           | 2515000    |       | 1352367        |       | 19288       |       | 19273           |       |
| $\alpha$      | -          |       | -              |       | 0.158       |       | 0.137           |       |
| D统计量          | 2907.653   |       | 1809.796       |       | 0.164       |       | 0.146           |       |
| 方差/均值         |            |       |                |       | 401503      |       |                 |       |
| MC            |            |       |                |       | 0.378***    |       |                 |       |

注：\*、\*\*和\*\*\*分别表示在5%、1%与0.1%的显著性水平上显著。

预期。平均地，人口规模对省际人口迁移区域人口规模每增加1%，会导致迁出流数量上升约0.38%，迁入流数量上升约0.17%。区域人口基数与迁出人口数量有直观上的较强正相关关系，其对人口迁移的“推力”作用相比“拉力”作用而言更加明显。迁出地性别比（SEX）对省际人口迁移产生极大的正向影响，回归系数为人口规模的3~7倍，迁出地性别比每增加1%，会导致迁出流数量上升约1.15%，这与中国省际人口流动迁移中仍以男性为主有关，此外，根据以往研究的结论<sup>[28]</sup>，务工经商是中国人口流动迁移的主要目的之一，而男性人口比例较大的省份往往拥有更多的外出务工人员。

**4.3.2 社会网络因素** 流动链指数（MSTOCK）对省际人口迁移存在显著的正向影响，回归系数为解释变量中最高，符合预期。流动链指数可以反映迁入地和迁出地之间的社会网络联系，省际流动链指数每增加1%，人口迁移数量上升约1.2%，显示出迁入某省份

的人群通过社会网络（地缘、血缘、同事朋友关系等）为后迁的人群提供便利，吸引更多的人口迁入，如亲友等社会网络联系是迁移人群获得就业信息的重要方式。迁移存量同样对省际人口迁移施加较大的正向作用，与预期相符。迁移存量（*FLOW*）每增加1%，人口迁移数量上升约0.63%，说明上一期人口迁移行为会对现期人口迁移产生一定跨期促进效应，先迁的人群可以为后迁的人群提供信息等方面的帮助，从而规避迁移风险。在中国注重“人情纽带”的文化背景下，社会网络因素对省际人口迁移的影响作用不可忽视。

**4.3.3 经济发展因素** 一方面，迁入地的地区人均GDP平均增速（*GDPI*）作为衡量区域经济发展及吸引力的重要指标，对塑造省际人口迁移格局产生重大的作用。地区人均GDP平均增速每增加1%，会导致迁入流数量上升约0.99%，说明人们往往选择人均GDP增速较快的省份作为迁移目的地。另一方面，迁出地的GDP平均增速对省际人口迁移不产生显著影响，与预期不符合，说明GDP增速对省际人口迁移的“拉力”作用不足。其次，计量结果显示出与以往研究不同的结论<sup>[29]</sup>，即各省FDI占比（*FDI*）对人口迁出产生的“推力”作用较小，市场化力量对迁出流的影响相对较弱。

**4.3.4 教育水平因素** 教育水平是影响人口迁移的重要变量，本文使用反映各等级综合教育水平的平均受教育年限（*ASCHO*）作为解释变量研究教育对省际人口迁移的影响。研究发现，平均受教育年限对迁入人口数量产生较大的正向影响，平均受教育年限增加1%，迁入流上升约0.99%，说明教育水平越高的区域对外来人口具有更大的吸引力。平均受教育年限对迁出流的作用则不符合理论预期，平均受教育年限增加1%，迁出流上升约0.78%，说明区域教育水平的提高一方面会吸引更多的外来人口，同时也会加速区域内部和区域间高素质人才的流动。

#### 4.4 居住环境与公路网络的影响机制分析

居住环境因素与公路网络因素对中国省际人口迁移产生着不可忽视的影响，这种影响主要体现在对迁入流的“拉力”作用上。人均居住面积较大、人口密度较高的省份往往能吸引更多的外省迁移人口。局部尺度公路网络可达性的提升对迁入流产生较大的影响，全局尺度公路可达性及穿行度对迁入流产生的影响较小，公路网络对中短距离人口迁移的影响大于远距离人口迁移。

**4.4.1 居住环境因素** 转型阶段中国的城镇化日益重视以人为本的导向，人口迁移也更加关注居住环境因素的影响，其中，人口密度（*DEN*）和人均居住面积（*AREA*）是衡量居住舒适程度的重要指标。研究发现，迁入省份的城市人口密度及人均住房面积对省际人口迁移产生显著的“拉力”作用，与预期相符。其中，人均住房面积对人口迁移的影响较大，人均住房面积增加1%，迁入流增加约0.58%；人口密度对迁入人口数量产生一定的影响，人口密度增加1%，迁入流上升约0.15%。这是因为人口密度高的区域在带来较便利的就业条件及公共服务的同时，同时带来诸如拥堵问题等负效应。而迁出省份的人口密度及人均居住面积对人口迁移均不产生显著影响，未符合预期。以上结果说明，尽管人们很关注目的地省份居住环境的改善，但跨省迁移行为对现住地居住环境变化的弹性较小。

**4.4.2 公路网络因素** 既有关关注交通对省际人口迁移影响的文献主要集中在轨道交通领域<sup>[30]</sup>，随着中国高速公路网络的加速建设，公路网络（包括高速公路、国道、省道）对人口迁移的影响作用日益增加。通过空间句法计算得到的两个尺度下公路网络接近度和穿行度均对省际人口迁移产生显著的影响。结果显示（图4），局部尺度公路网络接近度

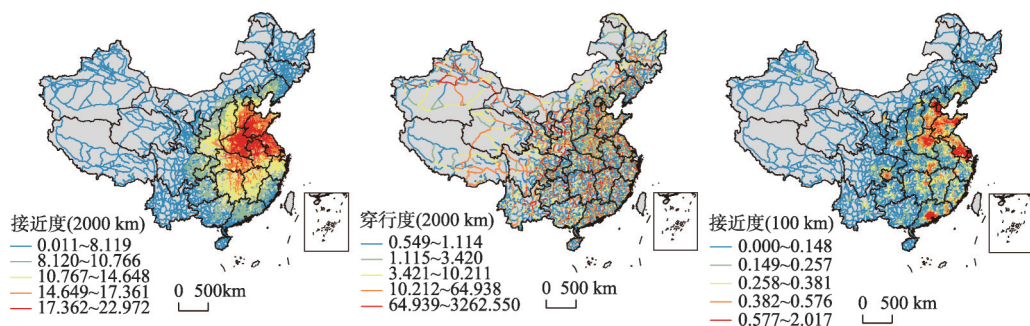


图4 2015年公路网络局部接近度、全局接近度和全局穿行度

Fig. 4 Results of C100, C2000 and B2000 in 2015

对省际人口迁移的影响作用较大,平均地,公路网络局部接近度(C100)每增加1%,迁入流上升约0.25%。另一方面全局尺度公路网络接近度(C2000)与穿行度(B2000)对省际人口迁移的影响作用极小。对于中近距离的迁移行为,公路可达性较高的省份会带来迁移成本的降低,因此对人口迁移产生较大的吸引力;而目的地省份公路网络可达性及穿行性的变化几乎不影响迁移距离在2000 km左右的远距离迁移行为。一个可能的原因是,对于远距离迁移的人口而言,航空、铁路为主要交通方式,而中近距离迁移的人口则较依赖公路交通网络。

#### 4.5 社会网络与迁移距离影响机制分析

社会网络因素(迁移存量、流动链指数)对省际人口迁移的影响日益增强,而空间距离对省际人口迁移的影响进一步呈现弱化趋势。既有研究中已涉及社会网络因素对省际人口迁移的影响,蒲英霞等<sup>[1]</sup>和Fan等<sup>[12]</sup>发现迁移存量对人口迁移产生一定的影响(系数约为0.4~0.8),劳昕等发现流动链系数对人口迁移的影响系数约为0.6~0.8<sup>[31]</sup>。本文发现社会网络因素对省际人口迁移的影响系数相对于其他因子而言较高,是塑造未来人口迁移格局的重要力量。此外,迁移距离作为重力模型的基础变量,在以往研究中表达为省会间空间距离<sup>[31]</sup>、省会间最短铁路距离<sup>[30]</sup>、省会间最短公路距离<sup>[32]</sup>等方式,已有文献发现迁移距离对省际人口迁移的影响呈现弱化趋势<sup>[32]</sup>。相比以往研究,在使用ESF方法处理网络自相关的影响后,空间距离(DIST)对人口迁移仅产生极其微弱的负向影响,回归系数为-0.09,说明随着地区的交通基础设施日益完善,以及互联网等通信技术的发展,物理距离对人口迁移产生的影响力逐渐减弱。

## 5 结论与讨论

### 5.1 结论

本文以2015年中国1%人口抽样调查数据中以现住地和五年前常住地划分的省际人口迁移为对象,构建基于特征向量空间滤波的负二项重力模型(ESF NBRM)研究了中国省际人口迁移的驱动因素,得出以下结论:

(1) 2010-2015年中国省际人口迁移流间存在显著的空间溢出效应,MC值达到0.378;图示分析和位序—规模分析均显示,ESF空间滤波方法能有效地提取数据中的网络自相关性,降低因数据中网络自相关性而导致的估计偏差,且少数排序较高(前1.4%)的特征向量即可提取数据中较高的网络自相关信息;空间滤波方法在进行O-D迁



移流建模时具有受数据限制小、计算灵活、对网络自相关性提取力强等优势,筛选后的特征向量可作为迁出地、迁入地遗漏变量的代理变量进入模型,通过合理设定阈值,模型可达到不同的过滤效果。

(2) 2010-2015年中国省际人口迁移流之间存在明显的过度离散现象,被解释变量的方差达到其均值的40万倍,违背了泊松重力模型中离散平衡(被解释变量的方差与均值相等)的基本假设。两个泊松模型的D统计量均远超阈值,证明了过度离散的存在。通过对负二项重力模型离散系数的LR检验及相应对数似然值的比较,说明考虑到数据离散的负二项重力模型更适用于省际人口迁移的计量分析。

(3) 数据中的网络自相关性会导致模型对距离相关变量( $C100$ ,  $C2000$ ,  $B2000$ ,  $DIST$ )估计的上偏与大部分非距离变量估计的下偏,通过空间滤波校正后的模型揭示出中国省际人口迁移的以下影响因素:区域人口特征、社会网络、经济发展、教育水平等因素对中国的省际人口迁移产生了重要影响,其中,迁入地性别比、GDP平均增速、平均受教育年限是重要的“拉力”因子,代表省份间社会网络联系的流动链指数、迁移存量同样对人口迁移产生极大的正向影响。居住环境因素与公路网络因素日益成为影响对中国省际人口迁移的“拉力”因子:人均居住面积较大、人口密度较高的省份能吸引更多的外省迁移人口;局部尺度公路网络可达性对迁入流产生较大正向影响,全局尺度公路网络对迁入流影响较小,公路网络对中短距离人口迁移的影响更大。

(4) 与既有研究结论相比,社会网络因素(迁移存量、流动链指数)对省际人口迁移的影响日益增大,是塑造未来人口迁移格局的重要力量,省际流动链指数和迁移存量每增加1%,人口迁移数量分别上升约1.2%与0.63%。使用ESF方法处理网络自相关的影响后,空间距离对人口迁移仅产生极其微弱的负向影响。

总体上看,中国2010-2015年间省际人口迁移的驱动因素与以往相比并未发生根本性的变化,在市场条件下,社会经济条件仍是影响人口迁移的首要原因。同时,随着中国新型城镇化的建设与道路基础设施的完善,城市环境以及城市公路网络交通可达性的提升对人口迁移产生了重要因素,而诸如迁移距离等变量对人口迁移的影响力逐渐减弱,以上因素可能导致中国省际迁移的明显局部性变化。因此,需通过制定合理的区域发展战略,重视地区环境生态、交通出行便利性以及地区间的社会文化联系对人口迁移的影响,适度控制省际人口迁移的流向和强度,促进区域经济的均衡发展。

## 5.2 讨论

本文仍存在以下不足之处:① 由于数据统计口径受限,本文仅对于一个时间截面的省际人口数据进行研究,缺乏对多个截面的对比分析;② 空间滤波方法虽然能够极大地过滤数据中的网络自相关性,但相比空间自回归模型,空间滤波方法的解释力较弱(不能对空间溢出效应的大小进行估计);③ 诸如土地利用等对省际人口迁移产生影响的变量,由于数据获取难度较大,并未纳入模型进行分析,从而导致遗漏变量的问题。

有鉴于此,有以下拓展方向:① 可基于更新的人口普查及抽查数据,进行多期面板实证研究;② 在模型中进一步考虑更多影响省际人口迁移的因素,并探索既有模型中不符合预期的因素的原因;③ 需拓展基于空间滤波的省际人口迁移实证研究,对比空间滤波和空间自回归两大类空间计量模型的优劣;④ 空间分层异质性(spatial stratified heterogeneity)是省际人口迁移网络的重要特征之一,可在处理人口迁移流网络自相关效应的同时,结合地理探测器等工具对省际迁移流的网络分异性以及各类型变量对人口迁移空间分异的影响程度进行探测<sup>[33-34]</sup>。

## 参考文献(References)

- [1] Pu Yingxia, Han Hongling, Ge Ying, et al. Multilateral mechanism analysis of interprovincial migration flows in China. *Acta Geographica Sinica*, 2016, 71(2): 205-216. [蒲英霞, 韩洪凌, 葛莹, 等. 中国省际人口迁移的多边效应机制分析. *地理学报*, 2016, 71(2): 205-216.]
- [2] Shen J. Explaining interregional migration changes in China, 1985-2000, using a decomposition approach. *Regional Studies*, 2015, 49(7): 1176-1192.
- [3] Liu Y, Shen J. Modelling skilled and less-skilled interregional migrations in China, 2000-2005. *Population, Space and Place*, 2017. Doi: 10.1002/psp.2027.
- [4] Griffith D A. Spatial structure and spatial interaction: 25 years later. *The Review of Regional Studies*, 2007, 37(1): 28-38.
- [5] Lesage J P, Pace R K. Spatial econometric modeling of origin-destination flows. *Journal of Regional Science*, 2008, 48(5): 941-967.
- [6] Griffith D A, Jones K G. Explorations into the relationship between spatial structure and spatial interaction. *Environment & Planning A*, 1980, 12(2): 187-201.
- [7] Griffith D A. Modeling spatial autocorrelation in spatial interaction data: Empirical evidence from 2002 Germany journey-to-work flows. *Journal of Geographical Systems*, 2009, 11(2): 117-140.
- [8] Chun Y, Griffith D A. Modeling network autocorrelation in space-time migration flow data: An eigenvector spatial filtering approach. *Annals of the Association of American Geographers*, 2011, 101(3): 523-536.
- [9] Zhu Y, Chen W. The settlement intention of China's floating population in the cities: Recent changes and multifaceted individual-level determinants. *Population Space & Place*, 2010, 16(4): 253-267.
- [10] Ding Jinhong, Liu Zhenyu, Cheng Danming, et al. Areal differentiation of inter-provincial migration in China and characteristics of the flow field. *Acta Geographica Sinica*, 2005, 60(1): 106-114. [丁金宏, 刘振宇, 程丹明, 等. 中国人口迁移的区域差异与流场特征. *地理学报*, 2005, 60(1): 106-114.]
- [11] Cai F, Wang D. Migration as marketization: What can we learn from China's 2000 census data? *China Review*, 2003, 3(2): 73-93.
- [12] Fan C C. Interprovincial migration, population redistribution, and regional development in China: 1990 and 2000 census comparisons. *Professional Geographer*, 2005, 57(2): 295-311.
- [13] Zhang Hongli, Liang Yinhe, Yang Weiqiong. Market potential, expected wages and interprovincial migration: Based on spatial econometrics model. *Journal of Applied Statistics and Management*, 2016, 35(5): 868-880. [张红历, 梁银鹤, 杨维琼. 市场潜能、预期收入与跨省人口流动: 基于空间计量模型的分析. *数理统计与管理*, 2016, 35(5): 868-880.]
- [14] Lian Lei. An empirical study on the spatial effects in the migration. *Population & Economics*, 2016(2): 30-39. [连蕾. 我国人口迁移过程中的空间效应实证研究. *人口与经济*, 2016(2): 30-39.]
- [15] 1% Population Sample Survey Office of the State Council, Population Statistics Department of National Bureau of Statistics. China's 2015 1% Population Sample Survey Data. Beijing: China Statistics Press, 2017. [国务院全国1%人口抽样调查领导小组办公室, 国家统计局人口和就业统计司. 2015年全国1%人口抽样调查资料. 北京: 中国统计出版社, 2017.]
- [16] Hu Angang, Li Chunbo. New poverty during the new century: Knowledge poverty. *Social Sciences in China*, 2001(3): 70-81. [胡鞍钢, 李春波. 新世纪的新贫困: 知识贫困. *中国社会科学*, 2001(3): 70-81.]
- [17] China Communications Press. China Traffic Atlas. Beijing: China Communications Press, 2016. [人民交通出版社. 中国交通地图册. 北京: 人民交通出版社, 2016.]
- [18] National Bureau of Statistics (NBS). China Statistical Yearbook (2016). Beijing: China Statistics Press, 2016. <http://www.stats.gov.cn/tjsj/ndsj/2016/indexch.htm>. [国家统计局. 中国统计年鉴(2016). 北京: 中国统计出版社, 2016. <http://www.stats.gov.cn/tjsj/ndsj/2016/indexch.htm>.]
- [19] Population Census Office under the State Council, Department of Population and Employment Statistics under the National Bureau of Statistics (NBS). Tabulation on the 2010 Population Census of the People's Republic of China (Book I). Beijing: China Statistics Press, 2012. [国务院人口普查办公室, 国家统计局人口和就业统计司. 中国2010年人口普查资料(上). 北京: 中国统计出版社, 2012.]
- [20] Shen J. Changing patterns and determinants of interprovincial migration in China 1985-2000. *Population Space & Place*, 2012, 18(3): 384-402.
- [21] Flowerdew R, Aitkin M. A method of fitting the gravity model based on the Poisson distribution. *Journal of Regional*

- Science, 1982, 22(2): 191-202.
- [22] Chun Y. Modeling network autocorrelation within migration flows by eigenvector spatial filtering. *Journal of Geographical Systems*, 2008, 10(4): 317-344.
- [23] Tiefelsdorf M, Boots B. The exact distribution of Moran's I. *Environment and Planning A*, 1995, 27(6): 985-999.
- [24] Getis A, Griffith D A. Comparative spatial filtering in regression analysis. *Geographical Analysis*, 2002, 34(2): 130-140.
- [25] Tao Wei, Gu Hengyu, Chen Hannan. Guangzhou's spatial distribution under the effect of the urban road network on the hotel industry. *Tourism Tribune*, 2015, 30(10): 99-108. [陶伟, 古恒宇, 陈昊楠. 路网形态对城市酒店业空间布局的影响研究广州案例. *旅游学刊*, 2015, 30(10): 99-108.]
- [26] Gu Hengyu, Shen Tiyan, Zhou Lin, et al. Measuring street layout's spatio-temporal effects on housing price based on GWR and sDNA model: The case study of Guangzhou. *Economic Geography*, 2018, 38(3): 82-91. [古恒宇, 沈体雁, 周麟, 等. 基于 GWR 和 sDNA 模型的广州市路网形态对住宅价格影响的时空分析. *经济地理*, 2018, 38(3): 82-91.]
- [27] Hilbe J. *Modeling Count Data*. New York: Cambridge University Press, 2014.
- [28] Liu Yanling, Feng Jian. Characteristics and impact factors of migration in China: Based on the analysis of the sixth census data. *Human Geography*, 2014, 29(2): 129-137. [刘晏伶, 冯健. 中国人口迁移特征及其影响因素: 基于第六次人口普查数据的分析. *人文地理*, 2014, 29(2): 129-137.]
- [29] Cai Fang, Wang Dewen. Population migration and mobility as marketization: Based on the fifth census data. *Chinese Journal of Population Science*, 2003(5): 11-19. [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(5): 11-19.]
- [30] Ma Wei, Wang Yahua, Liu Shenglong. Transportation infrastructure and population migration in China: An analysis based on gravity model. *China Soft Science*, 2012(3): 69-78. [马伟, 王亚华, 刘生龙. 交通基础设施与中国人口迁移: 基于引力模型分析. *中国软科学*, 2012(3): 69-78.]
- [31] Lao Xin, Shen Tiyan. An exploration into China's urban system evolution forecast based on intercity migration. *Population & Economics*, 2016(6): 35-47. [劳昕, 沈体雁. 基于人口迁移的中国城市体系演化预测研究. *人口与经济*, 2016(6): 35-47.]
- [32] Wang Guixin, Pan Zehan, Lu Yanqiu. China's inter-provincial migration patterns and influential factors: Evidence from year 2000 and 2010 population census of China. *Chinese Journal of Population Science*, 2012(5): 2-13. [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. *中国人口科学*, 2012(5): 2-13.]
- [33] Wang Jinfeng, Xu Chengdong. Geodetector: Principle and prospective. *Acta Geographica Sinica*, 2017, 72(1): 116-134. [王劲峰, 徐成东. 地理探测器: 原理与展望. *地理学报*, 2017, 72(1): 116-134.]
- [34] Li Jiaming, Lu Dadao, Xu Chengdong, et al. Spatial heterogeneity and its changes of population on the two sides of Hu Line. *Acta Geographica Sinica*, 2017, 72(1): 148-160. [李佳谔, 陆大道, 徐成东, 等. 胡焕庸线两侧人口的空间分异性及其变化. *地理学报*, 2017, 72(1): 148-160.]

## Driving mechanism of interprovincial population migration flows in China based on spatial filtering

GU Hengyu<sup>1</sup>, SHEN Tiyan<sup>1</sup>, LIU Ziliang<sup>2</sup>, MENG Xin<sup>1</sup>

(1. School of Government, Peking University, Beijing 100871, China; 2. School of Economics and Management, South China Normal University, Guangzhou 510006, China)

**Abstract:** According to previous studies, not only does the conditional gravity model based on ordinary least squares often bring about poor fitting of migration flows in reality, but also there exists overdispersion in the extended Poisson gravity model. Simultaneously, network autocorrelation usually exists in population migration data (e.g., the spatial interaction among migration flows). The problems mentioned above result in biased estimation. In order to capture network autocorrelation and deal with the issue of overdispersion, we build an eigenvector spatial filtering negative binomial gravity model (ESF NBGM) based on the data of 1% national population sample survey in 2015, to analyze the driving mechanism of interprovincial population migration flows in China. The results are as follows: (1) Positive spatial spillover effect exists in interprovincial population migration flows, and ESF can capture network autocorrelation in data, so as to reduce the estimated deviation of the model. Furthermore, eigenvectors ranking top 1.4% can properly interpret the spatial pattern of high network autocorrelation in data. (2) There exists overdispersion in China's interprovincial migration flows. Considering this problem, a negative binomial regression model is more suitable for the estimation of driving mechanism for population migration, together with statistical enhancement. (3) Network autocorrelation leads to overestimation of distance variables and underestimation of non-distance variables. The results of the improved model reveal that: chief factors the affect driving mechanism are regional population characters, social network, economic development and education level. Meanwhile, living environment and road network gradually become one of the most crucial pulling factors that influence migration flows. (4) Compared to previous studies, social network (i.e. migration stock) plays a more significant role in population migration flows, while the impact of spatial distance keeps weakening.

**Keywords:** interprovincial population migration flows; spatial filtering; negative binomial gravity model; driving mechanism; China