

不确定性视角下的中国省际人口迁移机制分析

蒲英霞^{1,2,3}, 武振伟¹, 葛莹⁴, 孔繁花¹

(1. 南京大学地理与海洋科学学院, 南京 210023; 2. 江苏省地理信息技术重点实验室, 南京 210023; 3. 江苏省地理信息资源开发与利用协同创新中心, 南京 210023; 4. 河海大学地球科学与工程学院, 南京 211100)

摘要: 人口迁移过程具有内在的不确定性。贝叶斯模型平均方法(BMA)为不确定性问题提供了行之有效的解决方案。然而, 当前该方法多用于线性回归模型在变量选择时出现的模型不确定性问题, 很少用于空间建模。本文以2010—2015年中国省际人口迁移流为例, 将BMA方法应用于空间OD模型, 在考虑网络空间结构的基础上选取迁出地和迁入地各7个解释变量及距离因素, 利用马尔可夫链—蒙特卡罗模型综合方法(MC³)进行模型抽样, 以后验模型概率为权重计算相应变量的迁出地、迁入地和网络效应等, 定量分析不确定性背景下省际人口迁移影响因素和空间机制。结果表明: ① BMA模型估计结果更为稳健可靠。与单一模型相比, BMA中变量效应估计的90%可信区间明显缩小, 不确定性程度显著降低, 结果更为精确; ② 区域经济社会发展对省际迁移至关重要。经模型空间抽样后, 迁出地人口规模和GDP、迁入地教育水平和迁移存量等的变量后验包含概率大于90%; ③ 网络效应在省际迁移过程中不可忽视。所有变量的网络效应占总体效应的40%以上, 其中工资、城镇化率、教育和迁移存量等的网络效应(绝对值)大于各自的迁出地和迁入地效应; ④ 若不考虑迁移建模中的不确定性, 绝大多数区域经济社会变量对省际迁移的影响会被高估。

关键词: 人口迁移; 空间OD模型; 贝叶斯模型平均; 后验模型概率; 变量后验包含概率; 网络效应; 中国

DOI: 10.11821/dlxb202112007

1 引言

作为自然科学和哲学领域最早提出的一个范畴, 不确定性及由此产生的“黑天鹅”问题已被经济学、地理学、社会学和心理学等众多领域所关注^[1-8]。在人口迁移过程中, 由于地区经济社会发展所面临的复杂性日益突出, 同时迁移个体自身的文化知识、判断能力以及对目的地信息了解的局限性等, 加剧了迁移过程的风险和不确定性^[9-12]。当前随着空间自相关概念的引入, 空间溢出效应为区域人口迁移过程提供了更为合理的解释^[13-20]。然而, 人口迁移建模中的不确定性问题, 包括如何选择解释变量、如何设定模型结构、如何定义网络空间结构等, 直接关系到区域人口迁移影响因素和动力机制分析的稳健性和可靠性^[21]。若忽视这些问题, 可能会对主导人口迁移过程的关键因素做出不恰当的分析判断, 进而影响科学制定区域人口发展规划、国土空间规划以及其他经济社会宏观战略决策等^[22-23]。

收稿日期: 2020-10-23; 修订日期: 2021-05-20

基金项目: 国家自然科学基金项目(41771417, 41771029); 江苏高校优势学科建设工程; 江苏省地理信息资源开发与利用协同创新中心资助项目 [Foundation: National Natural Science Foundation of China, No.41771417, No.41771029; Priority Academic Program Development of Jiangsu Higher Education Institutions; Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application]

作者简介: 蒲英霞(1972-), 女, 山东日照人, 博士, 副教授, 主要从事GIS与空间数据分析集成、区域人口迁移建模与复杂地理计算。E-mail: yingxiapu@nju.edu.cn

贝叶斯模型平均方法 (Bayesian Model Averaging, BMA) 为降低模型不确定性问题提供了一套切实可行的解决方案。Draper 最早于 1995 年提出了 BMA 思想, 即在“真”模型未知的情形下, 通过构建一个两两互斥的模型空间 $M = \{M_1, M_2, \dots, M_k\}$, 以每个模型的后验概率为权重对所有可能出现的模型进行加权平均, 得到一个尽可能“平均”的结果, 以此表征“真”模型, 进而避免逐步回归方法得到的单一“最优”模型所面临的不确定性问题^[24-25]。在马尔可夫链—蒙特卡罗方法 (Markov Chain Monte Carlo, MCMC) 的基础上, 马尔可夫链—蒙特卡罗模型综合方法 (Markov Chain Monte Carlo Model Composition, MC³) 应运而生, 为 BMA 的实现提供了技术支撑^[26-27]。目前, BMA 方法以贝叶斯统计理论为基础, 因能够克服人为主观遴选解释变量造成的信息损失等优势, 在经济增长、金融管制和水资源管理等领域逐步得到应用^[28-32], 但主要集中在经典线性回归模型的变量筛选与预测分析等方面, 缺乏针对空间模型的相关处理。目前, 研究人员已将 BMA 方法拓展到空间回归模型中, 并以美国 1990—2000 年州际人口迁移为例进行实证研究^[33]。然而, 随着空间计量模型的发展, 单纯的系数估计已不能很好地解释自变量变化对因变量产生的影响, 取而代之则是考虑空间溢出的效应估计^[34]。

本文以 2010—2015 年中国省际人口迁移流为例, 在考虑迁移流之间相互依赖关系的基础上, 选取迁出地和迁入地常住人口规模、地区生产总值 (GDP)、城镇职工平均工资、城镇化率、每千人医疗卫生机构床位数、每十万人拥有大专及以上学历人数和迁移存量等区域经济社会变量以及时间距离因素, 利用 BMA 和马尔可夫链—蒙特卡罗模型综合方法 (MC³) 对模型空间进行抽样, 比较变量的后验包含概率, 获得重要的影响因素, 并根据后验模型概率对所有可能的模型进行加权平均, 综合考虑人口迁移建模中变量选择所面临的不确定性问题, 为中国省际人口迁移动力机制分析提供更为稳健和可靠的结果。为更好地解释区域经济社会变量对人口迁移的影响, 本文在对模型系数估计的基础上, 进一步计算变量的迁出地效应、迁入地效应和网络效应等, 对人口迁移的影响因素和空间动力机制进行了定量分析。基于贝叶斯统计理论分析框架, 顾及模型不确定性问题的 BMA 技术也可为商品流、信息流等其他要素流动现象提供有益的分析 and 借鉴^[35]。

2 研究方法

2.1 贝叶斯模型平均

贝叶斯模型平均 (BMA) 在模型的拟合效果、解释力和预测精度等方面都优于单个模型, 已成为当前解决模型不确定性问题的流行技术^[36-37]。基于贝叶斯原理, BMA 方法将模型本身和模型参数都看作具有某种统计分布形式的随机变量, 以所有可能出现的模型组成模型空间, 通过抽样技术计算每个模型的后验概率, 以此为权重对所有模型进行加权平均。假定未知且不可观测的“真”模型由模型空间 $M = \{M_1, M_2, \dots, M_k\}$ 按照某种概率分布生成, 即:

$$p(\theta|D) = \sum_{i=1}^k p(M_i|D)p(\theta|M_i, D) \quad (1)$$

式中: θ 为待估参数向量; D 为观测样本数据; M_i 为模型空间中第 i 个模型; k 为模型数量; $p(M_i|D)$ 为模型 M_i 的后验概率; $p(\theta|M_i, D)$ 是模型 M_i 中参数向量 θ 的后验概率密度; $p(\theta|D)$ 为不同模型条件下参数向量 θ 的后验密度分布的加权平均。

根据贝叶斯原理, 模型 M_i 的后验概率 $p(M_i|D)$ 的计算公式如下:

$$p(M_i|D) = \frac{p(D|M_i)p(M_i)}{\sum_{j=1}^k p(D|M_j)p(M_j)} \quad (2)$$

$$p(D|M_i) = \int p(D|\theta_i, M_i) p(\theta_i|M_i) d\theta_i \quad (3)$$

式中： $p(D|M_i)$ 为模型 M_i 的似然函数； $p(M_i)$ 为模型 M_i 的先验概率； θ_i 为模型 M_i 的参数向量； $p(\theta_i|M_i)$ 表示模型 M_i 所对应的参数向量 θ_i 的先验概率分布。参数向量 θ 的后验方差与其后验均值一样，也是加权平均形式。

2.2 马尔可夫链—蒙特卡罗模型综合方法

对含有 K 个解释变量的线性回归模型而言，如何对模型空间 (2^K) 进行抽样，寻找后验概率较高的模型是实现 BMA 方法的关键。目前，主流的抽样算法是马尔可夫链—蒙特卡罗模型综合方法 (MC³)，即在马尔可夫链—蒙特卡罗 (MCMC) 的基础上，采用取舍方法保留后验概率较高的模型，当抽样次数达到一定数量后能够保证抽样基于所有模型，且结果收敛^[27]。具体地，假定当前马尔可夫链处在模型 M_i 的位置上，则下一步移动到新模型 M_j 的概率是：

$$l = \min \left\{ 1, \frac{p(D|M_j)p(M_j)}{p(D|M_i)p(M_i)} \right\} \quad (4)$$

下一步仍继续停留在当前模型 M_i 位置上的概率为 $1-l$ 。模型 M_j 是在一个新的模型空间上通过均匀分布方法生成的。对于一个含有 2^K 个模型的空间，有：

$$p(M_j) = 2^{-K}, j = 1, 2, \dots, 2^K, \text{ 且 } \sum_{j=1}^{2^K} p(M_j) = 1 \quad (5)$$

引入或删除某个解释变量的先验概率为 $1/2$ ，事先对变量和模型的选择均无任何偏好。

2.3 空间 OD 模型

在人口迁移过程中，某区域要素变化不仅影响本区域的迁入与迁出，还通过溢出效应对周边地区及整个迁移系统产生影响。空间 OD 模型是在重力模型的基础上，以因变量和/或自变量的空间滞后形式为补充的一类空间计量相互作用模型。其中，空间自回归模型 (Spatial Autoregressive Model, SAR) 作为一种全局溢出模型最为常见，其具体形式如下^[15, 34]：

$$y = \rho W y + \alpha \mathbf{1}_N + X_o \beta_o + X_d \beta_d + \gamma g + \varepsilon, \varepsilon \sim N(0, \sigma^2 V) \quad (6)$$

$$V_{ii} = v_i, i = 1, 2, \dots, n; V_{ij} = 0, i \neq j$$

式中：因变量迁移流 y 是一个 $N \times 1$ ($N = n^2$) 的列向量； $\mathbf{1}_N$ 是一个 $N \times 1$ 的单位列向量； α 为截距项系数； X_o 和 X_d 分别表示迁出地和迁入地的自变量向量，可通过一个 $n \times K$ 阶变量矩阵 X 的克罗内克积 (\otimes) 得到，即 $X_o = X \otimes I_n$ ， $X_d = I_n \otimes X$ ； β_o 和 β_d 分别表示迁出地和迁入地变量的系数向量； g 表示一个 $N \times 1$ 阶距离列向量； γ 为参数； ε 为一个 $N \times 1$ 阶扰动项， $\varepsilon \sim N(0, \sigma^2 V)$ ； V 为一对角阵表示的异方差，元素为 (v_1, v_2, \dots, v_n) ， v_i 服从独立 $\chi^2(r)/r$ 分布， σ^2 服从逆伽马分布； W 为一行标准化的 $N \times N$ 阶网络权重矩阵； $W y$ 是空间滞后因变量； ρ 是网络自相关强度。

模型参数 (β, σ^2, ρ) 的先验分布对模型后验概率分布具有实质性影响。

$$\pi(\beta) \sim N(c, T), \pi(\sigma^2) \sim IG(a, b), \pi(\rho) \sim U(1/\lambda_{\min}, 1/\lambda_{\max}), \pi(r/v_i) \sim iid \chi^2(r)$$

式中：参数 β 是 α 、 β_o 、 β_d 和 γ 的组合，其先验信息存在大量不确定性（或无信息型先验分布），通常设为： $c = 0$ ， $T = I_{2K+1} \times 10^{12}$ (K 为变量数， $2K$ 为迁出地和迁入地变量数)； σ^2 是无信息型先验分布， $a = b = 0$ ； λ_{\min} 和 λ_{\max} 分别为网络权重矩阵 W 的最小和最大特征根。采用 Gibbs 抽样方法对 β 和 σ^2 进行抽样，Metropolis-Hastings (M-H) 方法对 ρ 进行抽样^[38]。

2.4 效应估计

空间 OD 模型的系数估计不再表示区域要素变化对迁移流产生的影响。其因变量对

自变量的偏导数形式为变量的总体效应(式(7)),从中分解出直接效应(迁出地/迁入地效应)和间接网络效应,从不同角度描述区域要素变化对迁移流产生的多重影响^[15]。考虑到各效应的矩阵维数较大,不利于直观显示,以下通过计算其标量形式予以简化。

总体效应的标量式为 $te = (1/n^2)l_n' \times TE \times l_n$ 。其中, TE 是一个 $n^2 \times n$ 阶矩阵,具体如下:

$$TE = \begin{pmatrix} \partial Y_1 / \partial X_1^r \\ \partial Y_2 / \partial X_2^r \\ \vdots \\ \partial Y_n / \partial X_n^r \end{pmatrix} = (I_{n^2} - \rho W)^{-1} \begin{pmatrix} Jd_1 \beta_d^r + Jo_1 \beta_o^r \\ Jd_2 \beta_d^r + Jo_2 \beta_o^r \\ \vdots \\ Jd_n \beta_d^r + Jo_n \beta_o^r \end{pmatrix} \quad (7)$$

式中: Jd_i 是一个 $n \times n$ 阶零矩阵,且第 i 行等于 l_n' ; Jo_i 是一个 $n \times n$ 阶零矩阵,且第 i 列等于 l_n 。

迁出地效应的标量式为 $oe = (1/n^2)l_n' \times OE \times l_n$, 其中 $n^2 \times n$ 阶矩阵 OE 如下:

$$OE = (I_{n^2} - \rho W)^{-1} \begin{pmatrix} \tilde{Jo}_1 \beta_o^r \\ \tilde{Jo}_2 \beta_o^r \\ \vdots \\ \tilde{Jo}_n \beta_o^r \end{pmatrix} \quad (8)$$

式中: \tilde{Jo}_i 是 $n \times n$ 阶矩阵 Jo_i 中元素 (i, i) 为零,调整目的是分离出内部效应。

类似地,迁入地效应的标量式为 $de = (1/n^2)l_n' \times DE \times l_n$, 其中 $n^2 \times n$ 阶矩阵 DE 如下:

$$DE = (I_{n^2} - \rho W)^{-1} \begin{pmatrix} \tilde{Jd}_1 \beta_d^r \\ \tilde{Jd}_2 \beta_d^r \\ \vdots \\ \tilde{Jd}_n \beta_d^r \end{pmatrix} \quad (9)$$

式中: \tilde{Jd}_i 是 $n \times n$ 阶矩阵 Jd_i 中元素 (i, i) 为零,调整目的同上。

网络效应的标量式为 $ne = (1/n^2)l_n' \times NE \times l_n$, 其中 $n^2 \times n$ 阶矩阵 NE 如下:

$$NE = (I_{n^2} - \rho W)^{-1} \begin{pmatrix} \tilde{Jn}_1(\beta_d^r + \beta_o^r) \\ \tilde{Jn}_2(\beta_d^r + \beta_o^r) \\ \vdots \\ \tilde{Jn}_n(\beta_d^r + \beta_o^r) \end{pmatrix}, \quad \tilde{Jn}_i = J_i - Jd_i - Jo_i \quad (10)$$

式中: J_i 是值全为1的 $n \times n$ 阶矩阵。

3 中国省际人口迁移机制分析

3.1 数据来源

2010年以来,中国经济和人口同时进入了“新常态”^[39-40],经济增长放缓和产业结构升级优化,同时人口低水平增长和人口素质显著提高,对中国人口迁移基本路径和空间格局产生了深刻影响^[41]。本文以中国31个省(自治区、直辖市)的省际人口迁移流为研究对象(暂未包括港澳台地区和省内迁移),综合考虑现有文献研究成果^[42-46],从人口因素、经济因素和社会因素等角度,选择迁出地/迁入地的常住人口规模、地区生产总值(GDP)、城镇职工平均工资、城镇化率、每千人医疗卫生机构床位数(医疗水平)、每十万人拥有大专以上人口数(教育水平)、迁移存量和时间距离等变量表征人口迁移影响因素(表1)。其中,2010—2015年省际人口迁移数据来自《2015年全国1%人口抽样调查

表 1 模型解释变量描述
Tab. 1 Description of model explanatory variables

变量名称	变量描述(单位)	预期迁出地/迁入地效应
<i>O_POP/D_POP</i>	迁出地/迁入地 2010 年常住人口总数(万人)	+/+
<i>O_GDP/D_GDP</i>	迁出地/迁入地 2010 年地区生产总值(亿元)	-/+
<i>O_Wage/D_Wage</i>	迁出地/迁入地 2010 年城镇职工平均工资(元)	-/+
<i>O_Urban/D_Urban</i>	迁出地/迁入地 2010 年城镇化率(%)	+/+
<i>O_HB/D_HB</i>	迁出地/迁入地 2010 年每千人医疗卫生机构床位数(张)	+/+
<i>O_Edu/D_Edu</i>	迁出地/迁入地 2010 年每十万人拥有大专及以上学历人数(人)	+/+
<i>O_Flow/D_Flow</i>	2005—2010 年间地区迁出/迁入总量(人)	+/+
<i>Distance</i>	省会城市之间 2010 年铁路运输最短时间(s)	-

资料》“全国按现住地和 5 年前常住地分的人口”(抽样比为 1.55%)^[47],社会经济要素采用 2010 年数据,来自《中国统计年鉴》(2011)^[48],迁移存量来自《2010 年第六次全国人口普查资料》“全国按现住地和 5 年前常住地分的人口”(抽样比为 9.55%)^[49],距离采用 2010 年省会城市之间铁路运输最短时间,来自《中国铁路旅客列车时刻表》(2010)^[50]。所有变量在运算中均取自然对数,便于后面对应效应估计进行弹性解释。

弹性原指某一物质抵抗外来冲击具有的稳定程度,现在也用来衡量一个因变量对另一个自变量变化的敏感程度,即自变量变化 1% 所引起因变量变化的百分比^[51]。在当前人口迁移领域,预期人口迁移活动随着人口增长而持续活跃,即人口规模的迁出地效应和迁入地效应皆为正;地区 GDP 和工资水平分别表征区域宏观和个体微观经济状况,经济发展会促进外来人口迁入而降低本地人口迁出,预期迁出地效应为负,迁入地效应为正;城镇化是人口迁移的结果,也会反过来进一步促进人口迁移,预期迁出地和迁入地效应均为正;医疗水平和教育水平分别表征地区公共资源和人力资本质量,预期迁出地和迁入地效应均为正。随着时间距离的增长,省际人口迁移趋于下降,预期距离变量的符号为负。

网络结构主要考虑迁出地和迁入地的空间依赖关系。首先根据公共边界原则构建一个 31×31 阶空间权重矩阵 C (定义海南省和广东省为邻居);其次通过克罗内克积分别得到迁出地网络权重矩阵 W_o ($W_o = C \otimes I_n$) 和迁入地网络权重矩阵 W_d ($W_d = I_n \otimes C$),同时去除内部流所在的行与列,使之变成一个 930×930 阶矩阵。最后,综合考虑上述两种情形,即 $W = W_o + W_d$ 。

3.2 MC³抽样结果

在模型抽样过程中,本文共进行 30 万次抽样,得到 2615 个唯一的模型。其中,146 个模型的后验概率大于 0.01%,概率总和达 99.5%;58 个模型的后验概率大于 0.1%,概率总和为 96.5%。表 2 是按后验概率大小排序得到的前 10 个模型及各变量的后验包含概率。其中,“1/0”分别表示模型包含/不包含该变量,最后一行表示模型后验概率。这 10 个模型的后验概率显著异于其先验概率 (2^{-15}),概率总和超过 76%,其中排名第 1 的模型后验概率为 0.405。

由表 2 可知:① 不同变量在模型空间中出现的后验包含概率大小不同。其中, O_POP 、 O_GDP 、 D_Edu 、 D_Flow 和 $Distance$ 等变量的包含概率均超过 90%,说明迁出地人口与 GDP、迁入地教育水平与迁移存量以及时间距离对于省际迁移的影响不可忽视; O_Flow 、 D_HB 等变量的包含概率在 25% 以下,表明在研究时段内迁出地存量和迁入地医疗卫生等因素对省际迁移决策的影响较小。② 同一种类型的迁出地变量和迁入地

表2 空间OD模型中后验概率最大的10个模型

Tab. 2 The 10 highest posterior probability among spatial Origin-Destination (OD) models

变量	后验概率模型										后验包含概率
	10	9	8	7	6	5	4	3	2	1	
<i>O_POP</i>	1	1	1	1	1	1	1	1	1	1	0.924
<i>O_GDP</i>	1	1	1	1	1	1	1	1	1	1	0.925
<i>O_Wage</i>	0	1	0	0	0	0	0	1	1	0	0.456
<i>O_Urban</i>	1	1	1	0	1	1	1	1	0	1	0.797
<i>O_HB</i>	1	0	1	0	1	1	0	1	0	1	0.696
<i>O_Edu</i>	1	1	0	1	0	0	1	0	1	0	0.459
<i>O_Flow</i>	0	0	0	0	0	0	0	0	0	0	0.238
<i>D_POP</i>	0	0	1	0	1	1	0	0	0	0	0.459
<i>D_GDP</i>	0	0	0	0	1	1	0	0	0	0	0.446
<i>D_Wage</i>	1	1	1	1	1	0	1	1	1	1	0.839
<i>D_Urban</i>	1	1	1	1	1	0	1	1	1	1	0.824
<i>D_HB</i>	0	0	0	0	0	0	0	0	0	0	0.205
<i>D_Edu</i>	1	1	1	1	1	1	1	1	1	1	0.923
<i>D_Flow</i>	1	1	1	1	1	1	1	1	1	1	0.962
<i>Distance</i>	1	1	1	1	1	1	1	1	1	1	0.926
后验概率	0.017	0.019	0.019	0.019	0.026	0.049	0.058	0.063	0.095	0.405	

变量发挥着不同的作用。例如，对于迁出地而言，人口规模和地区GDP具有较好的解释力；而对于迁入地，地区教育水平和迁入存量则更为关键，反映了“人往高处走”的迁移规律。这些因素在迁移中到底发挥多大作用，需要进一步计算各自的迁出地效应、迁入地效应和网络效应等。

3.3 系数估计分析

以后验概率大于0.1%的58个模型（概率总和超过96%）代表整个模型空间。经MCMC抽样得到的样本具有一定的自相关性。为了确保抽样的有效性，降低样本之间的自相关程度，需要对抽样进行分组，使得不同组之间的均值近似独立^[52]。图1a表明，滞后30阶以上的样本间的自相关系数可忽略不计。为保险起见，将抽样次数设定为60000，并舍弃前10000次，将每批次抽样间隔设为100，则每个模型可得到完全独立的500个系数抽样记录。图1b表明分批次抽样后，样本之间的自相关系数几乎都在2个标准差范围内波动，极大克服了抽样间的自相关性，模型系数估计更加有效。

经BMA之后，模型系数估计的均值和90%可信区间（Credible Interval）如表3所示。从表3可以看出，人口迁移明显受到周边迁移环境的影响，且不能忽视（ ρ 显著为正）。迁移主方向是从农业剩余劳动力大省迁往就业机会和劳动力市场较大的省份，人口增长对迁出影响较大。区域GDP或工资水平增长则会降低人口迁移增幅，符合预期。对迁出地而言，GDP的影响要大于工资水平；而对迁入地而言，工资变化影响更大些。城镇化率和医疗卫生水平提升有利于人口外迁；不同的是，城镇化率较高地区会抑制外来人口增长幅度。高素质人才增长将促进人口流动，相比较而言，对迁入地的影响更大。迁移存量的影响主要表现在迁入地。

由于存在空间依赖关系，要想真正量化区域经济社会要素变化对迁移流变动产生的影响，需要将整个人口迁移网络联系起来。

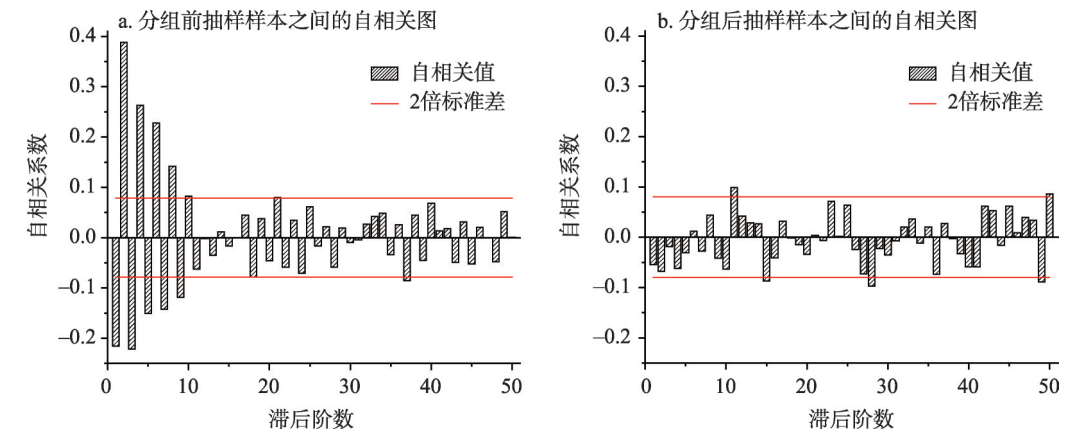


图1 分组前后抽样样本之间的自相关图
Fig. 1 Autocorrelation coefficients before and after sampling in batches

表3 BMA 加权后的空间OD模型各系数估计的后验均值与90%可信区间							
Tab. 3 Coefficient estimates and 95% credible intervals of origin and destination variables after BMA							
变量	Lower 5%	均值	Upper 95%	变量	Lower 5%	均值	Upper 95%
Const	3.329	3.664	4.041	Distance	-0.311	-0.283	-0.254
O_POP	1.057	1.165	1.282	D_POP	0.099	0.111	0.122
O_GDP	-0.887	-0.759	-0.648	D_GDP	-0.108	-0.095	-0.080
O_Wage	-0.029	-0.019	-0.010	D_Wage	-0.300	-0.274	-0.246
O_Urban	0.050	0.106	0.168	D_Urban	-0.392	-0.352	-0.308
O_HB	0.091	0.113	0.139	D_HB	0.000	0.001	0.003
O_Edu	0.064	0.074	0.085	D_Edu	0.363	0.404	0.449
O_Flow	-0.005	-0.003	-0.001	D_Flow	0.609	0.639	0.671
ρ	0.579	0.616	0.652				

注：Lower 5%和Upper 95%分别表示参数估计结果的90%可信区间下限和上限,后表同。

3.4 效应估计分析

根据模型后验概率与系数估计结果，按照公式 (7)~(10) 分别计算各变量的总体效应、迁出地效应、迁入地效应和网络效应（表4）。同时对单一模型（即解释变量100%出现在空间OD模型中）也进行了效应估计，作为比较的基准（表5）。考虑到区域经济社会因素的变化同时对迁出流和迁入流产生影响，下面不再区分迁出地和迁入地变量。总体上，除了迁移存量外，单一模型中变量的总体效应（绝对值）均大于BMA模型，这意味着若忽视变量选择中的不确定性，会高估区域经济社会因素对省际人口迁移的影响。更进一步，那些被高估的变量主要表现在迁入地效应和网络效应。经BMA后，工资水平、城镇化率、医疗状况、教育水平和迁出存量等的迁出地效应由不显著变为显著，表明这些因素共同对省际迁移产生了实质性影响。

3.4.1 总体效应 人口规模的总体效应在所有变量中居首位，表明人口压力是省际迁移活动的首要因素。平均地，若某地区人口规模相对增长1%，从长期来看，整个省际迁移流量的增幅将达到3.254%，其中迁出流增幅1.712%，迁入流增幅0.156%，其他周边区域迁移流增幅1.386%。这意味着地区人口增长不仅会产生更多迁出流，也会通过集聚效应吸引更多外来人口，与现有研究结果相一致^[53-54]。不同的是，GDP、工资水平和城镇化率

表 4 BMA 加权后空间 OD 模型中各解释变量的效应分布
Tab. 4 Effects distribution of explanatory variables in spatial OD models after BMA probability weighting

变量	总体效应			迁出地效应			迁入地效应			网络效应		
	Lower 5%	均值	Upper 95%	Lower 5%	均值	Upper 95%	Lower 5%	均值	Upper 95%	Lower 5%	均值	Upper 95%
POP	2.281	3.254	3.809	1.511	1.712	1.909	0.137	0.156	0.175	1.128	1.386	1.729
GDP	-2.611	-2.172	-1.775	-1.307	-1.116	-0.917	-0.144	-0.125	-0.107	-1.201	-0.931	-0.737
Wage	-0.870	-0.732	-0.621	-0.056	-0.040	-0.025	-0.322	-0.291	-0.257	-0.520	-0.402	-0.325
Urban	-0.847	-0.609	-0.405	0.051	0.141	0.231	-0.428	-0.371	-0.315	-0.523	-0.379	-0.272
HB	0.224	0.295	0.386	0.127	0.166	0.206	0.003	0.005	0.008	0.087	0.124	0.175
Edu	1.023	1.191	1.395	0.106	0.125	0.144	0.378	0.430	0.482	0.521	0.635	0.793
Flow	1.392	1.592	1.841	0.018	0.025	0.032	0.631	0.680	0.726	0.732	0.887	1.093

表 5 全要素空间 OD 模型(单一模型)各变量的效应分布
Tab. 5 Effects distribution of explanatory variables in the spatial OD model with all elements (single model)

变量	总体效应			迁出地效应			迁入地效应			网络效应		
	Lower 5%	均值	Upper 95%	Lower 5%	均值	Upper 95%	Lower 5%	均值	Upper 95%	Lower 5%	均值	Upper 95%
POP	2.862	4.609	6.832	1.270	2.012	2.842	0.178	0.519	0.863	1.061	2.077	3.566
GDP	-4.281	-2.764	-1.456	-1.682	-1.090	-0.546	-0.764	-0.398	-0.089	-2.270	-1.275	-0.618
Wage	-5.202	-3.267	-1.628	-1.060	-0.496	0.060	-1.539	-1.064	-0.563	-3.037	-1.707	-0.901
Urban	-4.327	-1.776	0.282	-0.597	0.314	1.273	-1.572	-1.013	-0.474	-2.400	-1.077	-0.045
HB	-0.145	1.383	3.118	-0.188	0.577	1.269	-0.243	0.176	0.571	-0.090	0.630	1.536
Edu	1.680	3.431	5.558	-0.067	0.596	1.258	0.617	1.060	1.495	0.834	1.775	3.072
Flow	0.764	1.300	1.918	-0.245	-0.020	0.183	0.466	0.585	0.710	0.429	0.735	1.154

的提升对整个迁移网络的增幅具有负向作用,这与2008年金融危机以来中国正处于经济结构调整和产业转型期有较大关系^[41]。迁移存量对于整个迁移网络的发展不可小觑,存量越多社会网络关系越复杂,人们获取就业、居住等的信息渠道越畅通,越能够有效降低迁移成本与风险^[55]。地区教育水平居第4位,是影响省际迁移的重要因素。平均而言,若某地区大专以上学历人口数量相对增长1%,则整个省际迁移流量增幅为1.191%,其中迁往教育科技水平相对发达地区是主流方向。医疗卫生水平反映了地区公共服务的供给优势,其提升在一定程度上会促进人口迁移。

3.4.2 迁出地效应 区域要素变化不仅直接改变自身迁移流,还通过周边地区进一步反馈到自身区域。通过比较各迁出地变量的系数(表3)与迁出地效应(表4),可以发现各个变量的反馈效应对迁出地效应的贡献。总体上,人口规模对省际迁出的影响最大。若当前某区域人口规模相对增长1%,则跨省迁出人口数量平均增幅长期将达1.712%。而地区GDP则有所不同,若某区域GDP相对增长1%,其跨省迁出人口增幅平均降低1.116%。地区经济发展将会降低本地人口外迁增长速度,表明生存和发展仍是人口迁移的主因。值得注意的是,地区城镇职工工资增长也会降低人口跨省外迁的增速,但作用相对有限。随着区域经济发展和城镇化水平的提高,整个地区教育水平和医疗状况不断改善,增强了人们对外界环境的适应能力,进一步促进人口外迁^[56-58]。类似地,迁出存量对跨省迁移的影响也相对较小,但由单一模型中的不显著变得显著。

3.4.3 迁入地效应 通过比较迁入地变量的系数(表3)和迁入地效应(表4),可发现变量的反馈作用对迁入地效应的贡献。此时,迁移存量和教育水平成为影响省际迁移的重要因素。平均地,若某地区迁入存量相对增长1%,从长期来看,外来跨省迁入人口增幅将达0.68%;若某地区大专以上学历人口相对提高一个百分点,则跨省迁入人口长期平均增幅为0.43%。同时,人口规模产生的集聚作用也会吸引外来人口,但相对于迁移存量和人口质量的提升,人口规模的影响相对较小。新古典主义经济理论认为,地区经济发展较好、工资水平较高地区更容易吸引移民,其城镇化水平也较高^[59-60]。然而,在中国经济和人口“新常态”背景下,经济结构转型和优化升级、经济增长明显放缓以及劳动年龄人口减少等,使得GDP、工资水平和城镇化发展在不同程度上降低了当前及未来省际迁移的增长幅度,表现出明显不同于过去的发展趋势^[41, 61]。

3.4.4 网络效应 最新的人口迁移理论认为,网络效应是形成大规模人口迁移的主要原因之一,特别是在迁移网络形成之前^[62-63]。表4表明,各变量的网络效应占总体效应的40%以上,甚至超过各自的迁出地效应或迁入地效应,进一步证实了网络效应在省际人口迁移过程中扮演着非常重要的角色,且不容忽视。以人口规模为例,若某地区人口相对增幅1%,除了本地区的迁出流和迁入流相应持续增长外,其他周边区域跨省迁移量的平均增幅长期将达到1.386%,远大于对本地迁入流的影响。而教育水平的网络溢出效应甚至超过了其迁入地效应和迁出地效应之和。若地区大专以上学历人口数量相对增长1%,则周边跨省迁移人口的平均增幅长期将达0.635%,充分反映了高素质人才在省际迁移中的关键作用。需要指出的是,网络效应和其他效应一样,均是一种长期效应,在区域社会经济因素发生变化后,迁移网络系统随着时间的推移而达到再平衡状态时对人口迁移产生的累积综合性结果。

4 结论与讨论

本文从变量选择的不确定性入手,基于贝叶斯模型平均方法(BMA),对经济和人

口“新常态”背景下的2010—2015年中国省际人口迁移影响因素和动力机制进行了定量分析。在综合考虑迁出地和迁入地空间依赖关系的基础上,利用马尔可夫链—蒙特卡罗模型综合方法(MC³)对区域经济社会要素(人口规模、GDP、工资水平、城镇化率、医疗水平、教育水平和迁移存量)以及距离进行概率分析,对后验概率大于0.1%的58个模型进行加权平均,基于系数估计结果计算变量的迁出地效应、迁入地效应和网络效应等,得到以下结论:

(1)当前中国省际人口迁移主要由迁出地效应和网络效应推动,迁入地效应为辅。BMA分析结果表明,迁出地人口和GDP的变量后验包含概率在90%以上,其迁出地效应和网络效应(绝对值)远大于其相应的迁入地效应;迁入地教育水平和迁移存量的变量后验包含概率也大于90%,其迁入地效应虽居各变量前列,但仍低于其相应的网络效应。这是普通重力模型所无法解释的,也为中国供给侧结构性改革提供新的证据。

(2)地区人口数量、质量以及迁移存量在省际迁移中发挥关键作用。当前人口规模变动产生的影响最为显著,各地区人口势能差为大规模迁移活动提供了潜在动力。同时,人口质量提升也注入不竭动力,高素质人才集聚地区经济发展和社会综合实力较强,更能吸引外来人口。通过亲属、朋友等社会关系网络,迁移个体能够获得相对准确信息,减少迁移成本。

(3)在经济发展“新常态”背景下,GDP和工资等变量的相对增速总体上对省际迁移的增幅产生负向效应。随着区域协调发展战略的深入实施和东部沿海地区经济结构性减速,劳动密集型产业向中西部有序转移,省际迁移增速呈现放缓迹象。然而,新型城镇化不仅有赖于人口迁移,在迁出地通过政府公共产品的投入和消费质量升级进一步促进迁移活动。

(4)区域经济社会因素通过空间结构对整个迁移网络起到了共同演化的作用。区域要素变化对从该区域出发或抵达的人口迁移流产生直接影响,进一步通过空间结构关系影响周边地区,对其产生网络溢出效应,同时该区域反过来又作为周边地区的邻居,对区域自身迁移流产生一定的反馈作用。经BMA后,区域各要素的效应估计皆显著,表明对人口迁移网络的发展与演化产生共同的影响,这是不考虑模型不确定性情形下无法观测到的。

本文利用BMA方法,对人口迁移建模中变量选择方面的不确定性问题进行了讨论,一定程度上提高了模型估计结果的稳健性和可靠性。然而,本文仍存在以下问题有待研究:①变量选择需要进一步优化。本文在解释变量选择时主要基于现有文献,而当前对2010年以来省际迁移影响因素和动力机制方面的研究相对较少,因此在变量选择方面仍存在一定的局限性;②其他模型不确定性问题。目前采用固定的网络空间结构($W_0 + W_d$)和空间自回归模型(SAR)在重力模型中的扩展,仅对变量选择中的不确定性进行了考虑。其他网络权重矩阵设定方式以及空间杜宾模型、空间误差项模型等扩展形式,都有待于进一步设定和检验;③时间结构的设定。空间OD模型本质上是一个横截面模型,不能显式表达时间维度。因此,模型估计结果是区域要素变化达到新的平衡状态时对系统产生的累积影响,是一个长期均衡值。今后将进一步考虑时空耦合关系^[64];④不同空间尺度的应用。省际人口迁移在国家和区域人口再分布中具有重要意义。然而,相对于省内迁移,其数量和规模均较小,地域空间关系对于省际迁移的影响也会小于地市级或县级尺度人口迁移。下一步将会考虑省内人口迁移,进一步分析更细空间尺度和邻接关系对人口迁移的影响。

参考文献(References)

- [1] Keynes J M. A Treatise on Probability. London: Macmillan, 1921.
- [2] Knight F H. Risk, Uncertainty and Profit. Chicago: University of Chicago Press, 1921.
- [3] Wu Lun, Yu Hailong, Gao Zhenji, et al. The frame of GIS uncertainty and methods of GIS data uncertainty. *Geography and Territorial Research*, 2002, 18(4): 1-5. [邬伦, 于海龙, 高振纪, 等. GIS不确定性框架体系与数据不确定性研究方法. *地理学与国土研究*, 2002, 18(4): 1-5.]
- [4] Taleb N N. The Black Swan: The Impact of the Highly Improbable. New York: Random House, 2007.
- [5] Ji Xiaolei. Uncertain Age: Global Recession and Chance for China. Guangzhou: Guangdong Tourism Press, 2012. [季晓磊. 不确定时代: 全球衰退与中国机会. 广州: 广东旅游出版社, 2012.]
- [6] Hansen L P. Nobel lecture: Uncertainty outside and inside econometric models. *Journal of Political Economy*, 2014, 122: 5. DOI: 10.1086/678456.
- [7] Cheng Changxiu, Shi Peijun, Song Changqing, et al. Geographic big-data: A new opportunity for geography complexity study. *Acta Geographica Sinica*, 2018, 73(8): 1397-1406. [程昌秀, 史培军, 宋长青, 等. 地理大数据为地理复杂性研究提供新机遇. *地理学报*, 2018, 73(8): 1397-1406.]
- [8] Sakai Y J M. Keynes versus F.H. Knight: Risk, Probability, and Uncertainty. Singapore: Springer Nature Singapore Pte Ltd., 2019.
- [9] Wang Xiaoqin, Tian Zhengang. Uncertainty and its impacts on rural-urban migration in China. *China Agricultural Economy*, 2001(7): 16-20. [汪小勤, 田振刚. 论我国城乡人口迁移中的不确定性及其影响. *中国农村经济*, 2001(7): 16-20.]
- [10] Bijak J. Forecasting International Migration in Europe: A Bayesian View. London: Springer Science Business Media, 2011.
- [11] Williams A M, Baláz V. Migration, risk and uncertainty: Theoretical perspectives. *Population, Space and Place*, 2012, 18: 167-180.
- [12] Bijak J, Bryant J. Bayesian demography 250 years after Bayes. *Population Studies*, 2016, 70(1): 1-19.
- [13] LeSage J P, Pace R K. Spatial econometric modeling of origin-destination flows. *Journal of Regional Science*, 2008, 48(5): 941-967.
- [14] Fischer M M, Wang J. Spatial Data Analysis: Models, Methods and Techniques. Heidelberg: Springer, 2011.
- [15] LeSage J P, Thomas-Agnan C. Interpreting spatial econometric origin-destination flow models. *Journal of Regional Science*, 2015, 55(2): 188-208.
- [16] Lian Lei. An empirical study on the spatial effects in the migration. *Population and 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.]
- [17] 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.]
- [18] 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. *数量统计与管理*, 2016, 35(5): 868-880.]
- [19] Gu Hengyu, Shen Tiyan, Liu Zixiang, et al. Driving mechanism of interprovincial population migration flows in China based on spatial filtering. *Acta Geographica Sinica*, 2019, 74(2): 222-237. [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. *地理学报*, 2019, 74(2): 222-237.]
- [20] Shi Bo, Zhang Xinyue. Technology accumulation, spatial spillover and population migration. *China Population, Resources and Environment*, 2019, 29(2): 156-165. [师博, 张新月. 技术积累、空间溢出与人口迁移. *中国人口·资源与环境*, 2019, 29(2): 156-165.]
- [21] Parent O, LeSage J P. Spatial econometric model averaging//Fischer M M, Getis A. *Handbook of Applied Spatial Analysis*. Berlin: Springer-Verlag, 2010: 35-460.
- [22] Morgan M G, Henrion M, Small M. Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis. Cambridge: Cambridge University Press, 1990.
- [23] The State Council, People's Republic of China. National plan on population development during the 2016-2030 period. http://english.www.gov.cn/policies/latest_releases/2017/01/25/content_281475551306587.htm, 2017-01-25. [国务院. 国家人口发展规划(2016—2030年). http://www.gov.cn/zhengce/content/2017-01/25/content_5163309.htm, 2017-01-25.]
- [24] Draper D. Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society*, 1995, 57(1): 45-97.

- [25] Ando T. Bayesian Model Selection and Statistical Modelling. Boca Raton: CRC Press, 2010.
- [26] Chib S. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 1995, 90(432): 1313-1321.
- [27] Madigan D, York J, Allard D. Bayesian graphical models for discrete data. *International Statistical Review*, 1995, 63(2): 215-232.
- [28] Fernández C, Ley E, Steel M F J. Model uncertainty in cross - country growth regressions. *Journal of Applied Econometrics*, 2001, 16(5): 563-576.
- [29] León-González R, Montolio D. Growth, convergence and public investment: A Bayesian model averaging approach. *Applied Economics*, 2004, 36(17): 1925-1936.
- [30] Tsai FT-C. Bayesian model averaging assessment on groundwater management under model structure uncertainty. *Stochastic Environmental Research & Risk Assessment*, 2010, 24(6): 845-861.
- [31] Horvath R. Research & development and growth: A Bayesian model averaging analysis. *Economic Modelling*, 2011, 28(6): 2669-2673.
- [32] Feldkircher M, Horvath R, Rusnak M. Exchange market pressures during the financial crisis: A Bayesian model averaging evidence. *Journal of International Money and Finance*, 2014, 40(2): 21-41.
- [33] LeSage J P, Parent O. Bayesian model averaging for spatial econometric models. *Geographical Analysis*, 2007, 39(3): 241-267.
- [34] LeSage J P, Fischer M M. Spatial regression-based model specifications for exogeneous and endogenous spatial interaction//Patuelli R, Arbia G. *Spatial Econometric Interaction Modelling*. Switzerland: Springer International Publishing, 2016: 15-36.
- [35] The Communist Party of China (CPC) Central Committee and the State Council, People's Republic of China. Guideline on improving the market-based allocation mechanism of production factors. http://english.www.gov.cn/policies/latestreleases/202004/10/content_WS5e8faa09c6d0c201c2cc0922.html, 2020-04-10. [中共中央、国务院. 关于构建更加完善的要素市场化配置体制机制的意见. http://www.gov.cn/zhengce/2020-04/09/content_5500622.htm, 2020-04-09.]
- [36] Hoeting J A, Madigan D, Raftery A E, et al. Bayesian model averaging: A tutorial. *Statistical Science*, 1999, 14(4): 382-401.
- [37] Wang Liang. Overview and prospects of Bayesian model averaging. *Journal of Technical Economics & Management*, 2016(3): 19-23. [王亮. 贝叶斯模型平均方法研究综述与展望. *技术经济与管理研究*, 2016(3): 19-23.]
- [38] LeSage J P, and Pace R K. *Introduction to Spatial Econometrics*. Boca Raton: CRC Press, Taylor & Francis Group, 2009.
- [39] Zheng Jingping. The new normal and suggestions of China's economy. *China Development Observation*, 2014(11): 42-44. [郑京平. 中国经济的新常态及应对建议. *中国发展观察*, 2014(11): 42-44.]
- [40] Li Jianmin. China's new normal in population and economy. *Population Research*, 2015, 39(1): 3-13. [李建民. 我国的人口新常态与经济新常态. *人口研究*, 2015, 39(1): 3-13.]
- [41] Migrant Population Division under National Health and Family Planning Commissions, P. R. China. *China's Floating Population Development Report (2017)*. Beijing: China Population Publishing House, 2017. [国家卫生和计划生育委员会流动人口司. *中国流动人口发展报告(2017)*. 北京: 中国人口出版社, 2017.]
- [42] Wang Guixin, Pan Zehan. China's floating population spatial distribution and influencing factors: Evidence from year 2020 population census of China. *Modern Urban Research*, 2013, 28(3): 4-11. [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. *现代城市研究*, 2013, 28(3): 4-11.]
- [43] Liu Shenglong. Factors analysis of interprovincial migration in China. *The Journal of Quantitative & Technical Economics*, 2014, 31(4): 83-98. [刘生龙. 中国跨省人口迁移的影响因素分析. *数量经济技术经济研究*, 2014, 31(4): 83-98.]
- [44] 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.]
- [45] Zhang Yaojun, Cen Qiao. Spatial patterns of population mobility and determinants of inter-provincial migration in China. *Population Research*, 2014, 38(5): 54-71. [张耀军, 岑俏. 中国人口空间流动格局与省际流动影响因素研究. *人口研究*, 2014, 38(5): 54-71.]
- [46] Gu Hengyu, Liu Ziliang, Shen Tiyan, et al. Modelling interprovincial migration in China from 1995 to 2015 based on an eigenvector spatial filtering negative binomial model. *Population, Space and Place*, 2019, 25(8): e2253. DOI: 10.1002/psp.2253.

- [47] Population Census Office under the State Council, Department of Population and Employment Statistics under the National Bureau of Statistics (NBS). Tabulation on the 2015 1% Population Sampling Survey of the People's Republic of China. Beijing: China Statistics Press, 2016. [国务院人口普查办公室, 国家统计局人口和就业统计司. 2015年全国人口1%人口抽样调查资料. 北京: 中国统计出版社, 2016.]
- [48] National Bureau of Statistics (NBS). China Statistical Yearbook (2011). Beijing: China Statistics Press, 2012. [国家统计局. 中国统计年鉴(2011). 北京: 中国统计出版社, 2012.]
- [49] 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. 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. 北京: 中国统计出版社, 2012.]
- [50] Transport Bureau of the Ministry of Railway (TBMR). China Railway Timetable 2010. Beijing: China Railway Publishing House, 2010. [铁道部运输局. 全国铁路旅客列车时刻表. 北京: 中国铁道出版社, 2010.]
- [51] Marshall A. Principles of Economics. Zhu Zhitai, Chen Liangbi, trans. Beijing: The Commercial Press, 2005. [马歇尔. 经济学原理. 朱志泰, 陈良璧, 译. 北京: 商务印书馆, 2005.]
- [52] LeSage J P. The theory and practice of spatial econometrics. Toledo: University of Toledo. 1999.
- [53] Shen J F. Explaining interregional migration changes in China, 1985-2000: Using a decomposition approach. *Regional Studies*, 2015, 49(7): 1176-1192.
- [54] Shen J F. Modelling interregional migration in China in 2005-2010: The roles of regional attributes and spatial interaction effects in modelling error. *Population, Space and Place*, 2017, 23(3): 1-14. DOI: 10.1002/psp.2014.
- [55] Fan C C. Modeling interprovincial migration in China, 1985-2000. *Eurasian Geography and Economics*, 2005, 46(3): 165-184.
- [56] Gennaioli N, La Porta R, Lopez-de-Silanes F, et al. Human capital and regional development. *The Quarterly Journal of Economics*, 2013, 128(1): 105-164.
- [57] Bauernschuster S, Falck O, Heblich S, et al. Why are educated and risk-loving persons more mobile across regions? *Journal of Economic Behavior & Organization*, 2014, 98: 56-69.
- [58] Faggian A, Modrego F, McCann P. Human capital and regional development//Capello R, Nijkamp P. (Eds.) *Handbook of Regional Growth and Development Theories: Revised and Extended* (2nd). Northampton: Edward Elgar Publishing Ltd, 2019: 149-171.
- [59] Sjaastad L A. The costs and returns of human migration. *Journal of Political Economy*, 1962, 70(5): 80-93.
- [60] Todaro M P. A model of labor migration and urban unemployment in less developed countries. *The American Economic Review*, 1969, 59(1): 138-148.
- [61] Zou Xiangjiang, Wu Dan. Is Chinese migration population decreasing? An analysis based on 2015 1% population survey data. *Jiangxi Social Sciences*, 2017, 37(5): 235-244. [邹湘江, 吴丹. 我国流动人口减少了吗? 基于2015年1%人口抽样调查数据的分析. *江西社会科学*, 2017, 37(5): 235-244.]
- [62] Epstein G S. Herd and network effects in migration decision-making. *Journal of Ethnic and Migration Studies*, 2008, 34(4): 567-583.
- [63] De Haas H. The internal dynamics of migration processes: A theoretical inquiry. *Journal of Ethnic and Migration Studies*, 2010, 36(10): 1587-1617.
- [64] Zhu Yu, Lin Liyue. Studies on the temporal processes of migration and their spatial effects in China: Progress and prospect. *Scientia Geographica Sinica*, 2016, 36(6): 820-828. [朱宇, 林李月. 中国人口迁移流动的时间过程及其空间效应: 回顾与展望. *地理科学*, 2016, 36(6): 820-828.]

Analyzing the spatial mechanism of interprovincial migration in China under uncertainty

PU Yingxia^{1,2,3}, WU Zhenwei¹, GE Ying⁴, KONG Fanhua¹

(1. School of Geography and Ocean Science, Nanjing University, Nanjing 210023, China; 2. Jiangsu Provincial

Key Laboratory of Geographic Information Science and Technology, Nanjing 210023, China;

3. Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and

Application, Nanjing 210023, China; 4. School of Earth Science and Engineering, Hohai University,

Nanjing 211100, China)

Abstract: Population migration process has an innate uncertainty with the increasing complexity of regional socioeconomic development. Bayesian model averaging (BMA) provides a feasible solution to the uncertainty of linear regression models. However, model uncertainties are seldom considered in spatial modeling. To reduce the uncertainties in migration modeling, this paper incorporates BMA approaches with spatial origin-destination (OD) models to quantify the spillover mechanism of interprovincial migration in China, 2010-2015. Specifically, we specified network dependence for migration flows and selected origin/destination's population size, gross domestic product (GDP), real wage, urbanization rate, the number of beds in health facilities per 1000 persons, the number of people over college level per 100000 persons in 2010 and migration stocks between 2005 and 2010 as well as railway travel time between provincial capitals in 2010 as explanatory variables. Among 2615 unique models based on 300000 samples using Markov chain Monte Carlo model combination (MC³), 58 models with posterior probability greater than 0.1% were chosen to estimate explanatory variables' origin effects, destination effects, and network effects. Some findings are as follows: (1) BMA model estimates are more robust and reliable. Compared with results from the single spatial origin-destination (OD) model, the widths of 90% credible interval of different explanatory variables are markedly reduced, indicating the degree of model uncertainty has been greatly decreased. (2) Regional population size, quality, and migration stocks have a significant influence on interprovincial migration processes. After model sampling, the inclusion probabilities of population size and GDP at origins, education level and migration stocks at destinations as well as distance variable are beyond 90%. (3) Network effects of most variables are statistically significant, accounting for more than 40% of their corresponding total effects. Moreover, the spillover effects of real wage, education level, and migration stocks are even greater than their corresponding origin and destination effects. (4) The impacts of most explanatory variables on interprovincial migration would be overestimated without considering uncertainties in modeling migration processes.

Keywords: population migration; spatial OD models; Bayesian model averaging (BMA); posterior model probability; posterior probability of variable inclusion; network effects; China