

基于贝叶斯估计的空间函数型自回归模型及其应用

杨炜明^{1,2}, 李明杰^{1,2}

- 重庆工商大学 数学与统计学院, 重庆 400067
- 经济社会应用统计重庆市重点实验室, 重庆 400067

摘要:目的 为了研究函数型数据中响应变量的空间相关性,根据现有研究方法,对具有空间依赖性的函数型数据进行研究,并提出其模型的贝叶斯估计方法。方法 以典型空间自回归模型为基础,根据函数响应变量的空间依赖性,假设响应变量和解释变量间存在内生关系,生成空间函数型自回归模型,通过主成分分析将模型中函数型部分变为离散型,然后在给定先验情况下计算模型中参数的完全条件后验分布,使用贝叶斯 MCMC 方法进行估计。结果 使用联合 Gibbs 采样和随机游动的 Metropolis-Hastings 算法对模型中参数进行估计,通过模拟研究发现:不同参数下模型的函数型系数以及其他参数的估计偏差和均方误差较小,由此验证了贝叶斯估计方法的有效性,同时将空间函数型模型用于重庆市主城区新房平均价格的实证分析,结果表明所提出模型的贝叶斯估计方法是有效的。结论 使用贝叶斯估计方法对模型中参数进行估计,在不同情况下函数型解释变量的估计效果一直都比较好,并且随着样本量的增大,其估计效果也越来越好,可以认为使用贝叶斯估计方法对空间函数型自回归模型进行估计是有效且可行的,同时通过实证分析说明重庆市主城区新房平均价格具有空间自相关性,而且会受到二手房挂牌量的影响。

关键词:函数型数据分析; 贝叶斯估计; Gibbs 采样; 随机游动的 Metropolis-Hastings 算法
中图分类号:O212.8 **文献标识码:**A **doi:**10.16055/j.issn.1672-058X.2024.0003.014

Spatial Function Autoregressive Model Based on Bayesian Estimation and Its Application

YANG Weiming^{1,2}, LI Mingjie^{1,2}

- School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, China
- Chongqing Key Laboratory of Applied Statistics for Economic and Social Applications, Chongqing 400067, China

Abstract: Objective To study the spatial correlation of response variables in functional data, according to existing research methods, functional data with spatial dependence were studied and the Bayesian estimation method was proposed.

Methods Based on the typical spatial autoregressive model and according to the spatial dependence of the functional response variables, the spatial function autoregressive model was generated by assuming that there was an endogenous relationship between the response variables and the explanatory variables. The functional components in the model were changed into discrete types by the principal component analysis, and then the complete conditional posterior distribution of the parameters in the calculation model was calculated under a given prior condition. The parameters in the model were estimated using the Bayesian MCMC method. **Results** The Metropolis-Hastings algorithm, which combined Gibbs sampling and random walk, was used to estimate the parameters in the model. Through simulation research, it was found that the functional coefficients of the model and the estimation deviations and mean square errors of other parameters under

收稿日期:2023-02-28 修回日期:2023-04-03 文章编号:1672-058X(2024)03-0104-09

基金项目:重庆市自然科学基金项目资助(CSTC2020JCYJ-MSXMX0394).

作者简介:杨炜明(1981—),男,湖南长沙人,博士,教授,从事经济统计、空间统计研究。

通讯作者:李明杰(1998—),男,重庆丰都人,硕士研究生,从事经济统计分析研究。Email:1728503770@qq.com.

引用格式:杨炜明,李明杰.基于贝叶斯估计的空间函数型自回归模型及其应用[J].重庆工商大学学报(自然科学版),2024,41(3):104—112.

YANG Weiming, LI Mingjie. Spatial function autoregressive model based on Bayesian estimation and its application[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2024, 41(3): 104—112.

different parameter models were small, which verified the effectiveness of the Bayesian estimation method. At the same time, the spatial function model was applied to the empirical analysis of the average price of new houses in the main urban area of Chongqing. The results show that the Bayesian estimation method of the proposed model is effective.

Conclusion The estimation effect of functional explanatory variables using Bayesian estimation method to estimate the parameters in the model is always better under different circumstances, and its estimation effect is also more and more effective as the sample size increases. Thus, it can be considered that using spatial function model to estimate the spatial function autoregressive model is effective and feasible. Meanwhile, the empirical analysis shows that the average house price of new houses in the main urban area of Chongqing has spatial autocorrelation, and it will be affected by the listing volume of second-hand houses.

Keywords: functional data analysis; Bayesian estimation; Gibbs sampling; random walk based Metropolis-Hastings algorithm

1 引言

近些年,随着科技的进步和发展,统计学研究中所需的数据在收集和存储时也变得灵活多样,研究者收集获取数据的方法以及数据的表现形式已经不同于往日,有的数据以图像的方式呈现,并随着时间变化呈现出函数型特征。由于这些复杂数据的曲线呈现出函数的特征,因此可以将它们看成关于时间或者空间的函数,在统计学上将它们称为函数型数据(Functional Data)。函数型数据在生活中越来越常见,如温度、天气、湿度、股票等。由于所研究的样本数据是经过一个函数过程而产生的,因此可以观测到这种特征的数据,在分析时需要把观测到的数据看成一个整体进行分析。与传统数据分析不同的是,函数型数据分析^[1]能够分析离散型和连续型数据,同时也能将有限维和无限维数据进行转换,具有更强的普适性。

通过研究函数型数据的一些模型和方法,比如函数型线性回归模型(FLM)和函数型主成分分析(FPCA),可以认识到高维数据的动态演化规律,解析复杂数据下的本质特征,对数据有更合理、更直观的研究和解释。关于函数型线性回归模型(FLM)的研究,James^[2]在2002年提出一种方法,该方法将广义线性模型(GLM)扩展到某些预测变量是来自曲线或函数观察值的情况,即函数逻辑回归和函数审查回归,并将它们用于解决标准数据缺失的问题;Cai等^[3]在2006年对函数型数据线性回归分析中斜率函数的估计方法进行了研究,他们发现在给定解释变量特定值的情况下,回归模型中响应变量预测值的收敛速度是由预测的平滑度和斜率函数之间的微妙相互作用决定的;Zhou等^[4]在2020年的研究中,对随机缺失响应的泛函偏线性模型分别构造了斜率函数、非参数分量和响应变量均值估计量,同时建立了估计量的收敛速度,并对此进行了模拟研究;Feng等^[5]在函数回归中引入函数单指数变化系数模型,并通过主成分分析和基函数逼近对

斜率函数和系数函数进行估计;Shi等^[6]在研究了函数线性分位数回归模型的充分性检验,并提出一种基于函数主成分分析的非参数U过程检验统计量,证明了在零假设下,检验统计量渐近地服从正态分布。在实际研究中发现,函数型数据在建模和分析时,除了需要考虑解释变量与被解释变量之间的关系,还需要考虑被解释变量具有空间依赖性的情况。

空间回归模型作为经典线性回归模型的推广,它能处理截面数据变量间的相互依赖性。在空间回归模型中,空间自回归模型解决空间依赖性是通过考虑被解释变量的空间自相关设置实现的。空间自回归模型作为空间计量经济学中比较重要的空间回归模型,可以应用于经济、环境、统计等领域。Lambert等^[7]于2010年提出包含空间自回归模型空间计量方法的两步有限信息最大似然估计法,最终得出随着空间自回归自相关和样本量的增加,空间自回归计数估计器在偏差方面比空间版本获得了增益;Cheng等^[8]研究了部分线性单指数空间自回归模型(PLSISARM)的广义矩估计方法,并通过蒙特卡洛模拟表明估值器在有限样本表现良好;Lam等^[9]在研究中通过指定空间权重矩阵,将稀疏矩阵调整成为模型的稀疏空间权重矩阵估计量,同时对股票数据的模拟和应用证实了这个程序的良好性能;Rineda-Rios等^[10]对空间自回归(SAR)模型使用最小二乘法和最大似然法作为参数的估计方法,并通过模拟研究评估所提出方法的性能;方丽婷^[11]在2019年的研究中将空间自回归模型与分位数相结合,提出在正态先验和双指数先验条件下空间自回归分位数回归模型的贝叶斯估计方法。对于空间计量模型,大多数学者使用极大似然估计和贝叶斯估计对模型未知参数进行估计。Qu等^[12]在对空间自回归模型(SAR)的研究中提出三种估计方法,分别是两阶段辅助变量法(2SIV)、准最大似然估计法(QMLE)和广义矩法(GMM),通过这些方法建立了估计渐近正态性以及

相合性,并使用蒙特卡罗方法研究了有限个样本模型的性质。

空间自回归模型和函数型数据分析在国内的研究中已经比较普遍,而对于同时具有空间结构和函数型特征的数据分析研究却相对较少。徐登可等^[13]对部分函数型空间滞后模型结合空间响应变量进行研究,提出了模型的贝叶斯估计,并对加拿大气温数据进行实证分析验证了贝叶斯方法的有效性;Shin 等^[14]在 2017 年结合函数型线性回归模型(FLM)在处理高维数据方面的优势和空间自回归(SAR)模型在处理空间依赖方面的优势,提出了空间函数型线性模型(Spatial Functional Linear Model, SFLM),并研究了 SFLM 的估计方法;Ternynck^[15]在 2014 年的研究中提出了函数型数据回归模型在含有空间效应时的估计方法;Ahmed 等^[16]提出了空间滞后模型(SLM)在函数型数据框架中的扩展(FSLM 模型),并通过最大似然估计方法对模型中的参数进行估计。

根据上述文献可知,由于函数型线性回归模型(FLM)和函数型主成分分析(FPCA)可以对函数型数据的特点进行全面剖析,所以函数型线性回归模型(FLM)和函数型主成分分析(FPCA)通常作为大多数学者研究函数型数据时的统计分析方式。而在空间计量方面的研究中,则以空间自回归模型(SAR)为例研究模型中参数的估计方法。对于具有空间结构的函数型数据研究较少,且研究时模型的参数估计一般采用极大似然估计和贝叶斯估计。因此,利用贝叶斯估计方法对具有空间依赖性的函数型数据进行分析,能够使模型中参数的统计推断更加精确。

本文以贝叶斯估计为基础,考虑函数型数据中函数响应变量的空间相关性,对具有空间依赖性的函数型数据进行研究,构建了空间函数型自回归模型,通过主成分分析以及 Gibbs 抽样和 Metropolis-Hastings 算法相结合的混合算法研究函数型空间自回归模型的贝叶斯估计,同时通过模拟研究验证贝叶斯估计方法的有效性,最后对重庆市主城区 2018-11—2022-10 的新房平均价格进行了实证分析。

2 模型设计与构建

空间自回归模型解决空间依赖性是通过加入因变量的空间自相关设置实现的,具体的空间自回归模型如下:

$$y = \rho W y + X \beta + \varepsilon$$

其中, $\varepsilon \sim N(0, \sigma^2 I_n)$ 。为了获取研究模型,需要考虑 n 个空间单元,它们属于固定的可数区域 $D \subset R^n$ 中,接下来从每个空间单元中观测得到一个响应变量 y_i 和一个函数型解释变量 $X_i(t)$, $t \in T$, 取它在希尔伯特空间

$L^2(T)$ 里面的值, $L^2(T)$ 表示所有定义在 T 上的平方可积函数构成的希尔伯特空间, T 是实数域 \mathbf{R} 中的紧区间,一般假设 $T = [0, 1]$ 。

根据空间自回归模型,假设 y_i 和 $X_i(t)$ 之间存在内生关系,就能得到在 T 上的函数型空间自回归模型:

$$y_i = \rho \sum_{j=1}^n w_{ij} y_j + \int_0^1 X_i(t) \beta(t) dt + \varepsilon_i$$

其中, $\beta(t)$ 是定义在 $[0, 1]$ 上的平方可积系数函数; ρ 是未知的空间自相关系数且 $\rho \in [0, 1]$; w_{ij} 是空间权重矩阵 \mathbf{W} 的第 (i, j) 个元素,它代表 i 与 j 之间的权重,当 $i = j$ 时, $w_{ij} = 0$ 。在模型中,假设误差项 $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, 2, \dots, n$; 令 $\mathbf{Y} = (y_1, y_2, \dots, y_n)^T$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n)^T$, 就能将模型写成矩阵形式:

$$\mathbf{Y} = \rho \mathbf{W} \mathbf{Y} + \int_0^1 X_i(t) \beta(t) dt + \boldsymbol{\varepsilon} \quad (1)$$

在这里 $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 I_n)$, $i = 1, 2, \dots, n$ 。

对于模型中的 $\int_0^1 X(t) \beta(t) dt$ 部分, 定义希尔伯特空间 $L^2(T)$ 上的一个标准正交基为 $\{\varphi_j, j \in \mathbf{N}\}$, 这样就能根据 K-L 定理可将 $X(t)$ 和 $\beta(t)$ 分解为

$$X(t) = \sum_{j=1}^{\infty} \xi_j \varphi_j(t), \beta(t) = \sum_{j=1}^{\infty} \gamma_j \varphi_j(t)$$

其中, ξ_j 是均值为 0, 方差 $E(\xi_j^2) = \lambda_j$ 的不相关随机变量, 而 ξ_j 和 γ_j 可表示为

$$\xi_j = \int_0^1 X(t) \varphi_j(t) dt, \gamma_j = \int_0^1 \beta(t) \varphi_j(t) dt$$

根据上面的分解, 将 $X(t)$ 和 $\beta(t)$ 代入 $\int_0^1 X(t) \beta(t) dt$, 就能得到

$$\begin{aligned} \int_0^1 X_i(t) \beta(t) dt &= \int_0^1 \left(\sum_{k=1}^{\infty} \xi_{ik} \varphi_k(t) \right) \left(\sum_{j=1}^{\infty} \gamma_j \varphi_j(t) \right) dt = \\ &= \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \xi_{ik} \gamma_j \left(\int_0^1 \varphi_k(t) \varphi_j(t) dt \right) = \\ &= \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \xi_{ik} \gamma_j \langle \varphi_k, \varphi_j \rangle = \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \xi_{ik} \gamma_j \delta_{kj} = \sum_{j=1}^{\infty} \xi_{ij} \gamma_j \end{aligned}$$

其中, $\langle \varphi_k, \varphi_j \rangle = \int_0^1 \varphi_k(t) \varphi_j(t) dt$ 表示 φ_k 和 φ_j 在希尔伯特空间 $L^2(T)$ 上的内积, δ_{kj} 是克罗内克尔符号。这样模型式(1)可以变为

$$\mathbf{Y} = \rho \mathbf{W} \mathbf{Y} + \sum_{j=1}^{\infty} \xi_{ij} \gamma_j + \boldsymbol{\varepsilon}$$

根据函数型主成分分析, 选取前 m 个主成分对模型进行分析, 上述模型可以近似等于

$$\mathbf{Y} \approx \rho \mathbf{W} \mathbf{Y} + \sum_{j=1}^m \xi_{ij} \gamma_j + \boldsymbol{\varepsilon}$$

令 $\mathbf{Z} = [\xi_{ij}]_{i \times j}$, $i = 1, 2, \dots, n$; $j = 1, 2, \dots, m$; $\boldsymbol{\gamma} =$

$(\gamma_1, \gamma_2, \dots, \gamma_m)^T$, 可将模型变为

$$Y \approx \rho WY + Z\gamma + \varepsilon$$

由于 $\varepsilon \sim N(0, \sigma^2 I_n)$, 则模型的似然函数为

$$L(\rho, \gamma, \sigma^2 | Y, Z) \propto |A| \cdot (\sigma^2)^{-\frac{n}{2}} \cdot \exp\left\{-\frac{B^T B}{2\sigma^2}\right\}$$

其中, $A = I_n - \rho W, B = AY - Z\gamma, I_n$ 是 n 维单位矩阵。

3 贝叶斯估计

在使用贝叶斯估计方法对模型中的未知参数进行估计时, 需要知道未知参数的先验分布。对于模型中未知参数的推断, 本文采用马尔科夫链蒙特卡罗算法 (MCMC) 对未知参数的后验分布进行抽样。模型推断前需对未知参数选取合适的先验分布。先假设所有参数是相互独立的, 根据 LeSage 和 Pace^[17] 的文章, 假设 $\gamma \sim N(\gamma_0, \Sigma_\gamma), \sigma^2 \sim IG(c, d)$, 其中超参数 $\gamma_0, \Sigma_\gamma, c, d$ 都是已知的, 对于参数 ρ 通常假设它的先验分布为均匀分布, 即 $\rho \sim U\left(\frac{1}{\xi_{\min}}, \frac{1}{\xi_{\max}}\right)$, 其中 ξ_{\min}, ξ_{\max} 分别为空间权重矩阵 W 的最小特征值和最大特征值。

3.1 完全后验条件分布

当忽略参数先验中的常数项时, 同时根据贝叶斯定理就能将模型的参数联合后验分布以比例关系的形式呈现如下:

$$\pi(\rho, \gamma, \sigma^2 | Y, Z) \propto L(\rho, \gamma, \sigma^2 | Y, Z) \pi(\rho) \pi(\gamma) \pi(\sigma^2)$$

代入具体的参数先验分布和似然函数得到:

$$\pi(\rho, \gamma, \sigma^2 | Y, Z) \propto |A| (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{B^T B}{2\sigma^2}\right\} \times (\sigma^2)^{-(c+1)} \exp\left\{-\frac{d}{\sigma^2}\right\} \times \exp\left\{-\frac{1}{2}(\gamma - \gamma_0)^T \Sigma_\gamma^{-1}(\gamma - \gamma_0)\right\}$$

因为要使用 MCMC 算法进行抽样, 需要知道每一个参数的完全条件后验分布。所以根据上面的先验分布, 可以推断出所有参数的完全条件后验分布, 再利用 MCMC 算法进行抽样。所求参数的完全条件分布如下:

$$\pi(\gamma | \rho, \sigma^2, Y, Z) \propto \exp\left\{-\frac{(AY - Z\gamma)^T (AY - Z\gamma)}{2\sigma^2}\right\} \times \exp\left\{-\frac{1}{2}(\gamma - \gamma_0)^T \Sigma_\gamma^{-1}(\gamma - \gamma_0)\right\} \times \exp\left\{-\frac{1}{2}(AY - Z\gamma)^T (\sigma^2 I_n)^{-1} (AY - Z\gamma) - \frac{1}{2}(\gamma - \gamma_0)^T \Sigma_\gamma^{-1}(\gamma - \gamma_0)\right\}$$

大概可以看出应该是一个多元正态分布, 所以需要对里面进行配凑, 将其配成一个多元正态分布。展开里面的多元正态分布项:

$$(AY - Z\gamma)^T (\sigma^2 I_n)^{-1} (AY - Z\gamma) =$$

$$\gamma^T \left(\frac{1}{\sigma^2} Z^T Z\right) - \gamma^T \left(\frac{1}{\sigma^2} Z^T A Y\right) - \left(\frac{1}{\sigma^2} Y^T A^T Z\right) \gamma +$$

$$Y^T A^T (\sigma^2 I_n)^{-1} A Y (\gamma - \gamma_0)^T \Sigma_\gamma^{-1} (\gamma - \gamma_0) =$$

$$\gamma^T \Sigma_\gamma^{-1} \gamma - \gamma^T \Sigma_\gamma^{-1} \gamma_0 - \gamma_0^T \Sigma_\gamma^{-1} \gamma + \gamma_0^T \Sigma_\gamma^{-1} \gamma_0$$

将它们合并变形就可以看出 γ 的完全后验条件分布是多元正态分布, 即

$$\pi(\gamma | \rho, \sigma^2, Y, Z) \sim N(\hat{\gamma}_0, \hat{\Sigma}_\gamma)$$

其中, $\hat{\gamma}_0 = (Z^T Z + \sigma^2 \Sigma_\gamma^{-1})^{-1} (Z^T A Y + \sigma^2 \Sigma_\gamma^{-1} \gamma_0), \hat{\Sigma}_\gamma = (Z^T Z + \sigma^2 \Sigma_\gamma^{-1})^{-1}$ 。

σ^2 的完全条件分布为

$$\pi(\sigma^2 | \rho, \gamma, Y, Z) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{(AY - Z\gamma)^T (AY - Z\gamma)}{2\sigma^2}\right\} \times (\sigma^2)^{-(c+1)} \exp\left\{-\frac{d}{\sigma^2}\right\} \times (\sigma^2)^{-\left(\frac{n}{2} + c + 1\right)} \exp\left\{-\frac{1}{\sigma^2} \left(\frac{(AY - Z\gamma)^T (AY - Z\gamma)}{2} + d\right)\right\}$$

很容易就能看出 σ^2 的完全后验条件分布是一个逆伽玛分布, 即

$$\pi(\sigma^2 | \rho, \gamma, Y, Z) \sim IG(\hat{c}, \hat{d})$$

其中, $\hat{c} = \frac{n}{2} + c, \hat{d} = \frac{(AY - Z\gamma)^T (AY - Z\gamma)}{2} + d$ 。

ρ 的完全条件分布:

$$\pi(\rho | \gamma, \sigma^2, Y, Z) \propto |A| \cdot \exp\left\{-\frac{(AY - Z\gamma)^T (AY - Z\gamma)}{2\sigma^2}\right\}$$

由于行列 $(A = I_n - \rho W)$ 中含有未知参数 ρ , 所以无法判断出 ρ 的完全后验条件分布具体是什么分布。

3.2 贝叶斯估计的 Metropolis-Hastings 算法

已知参数 γ 和 σ^2 的完全条件分布分别为多元正态分布和逆伽玛分布, 故对其使用 Gibbs 采样方法抽样获取随机数。参数 ρ 的完全条件分布不是一个具体的分布, 它的分布也比较复杂, 采用 Gibbs 抽取随机数比较困难, 所以本文选择随机游动的 Metropolis-Hastings 算法产生参数 ρ 的随机数。在具体实践中, 对函数型空间自回归模型的抽样是抽取参数序列 (ρ, γ, σ^2) 的随机数, 本文是将 Gibbs 抽样方法和 Metropolis-Hastings 算法相结合, 这种抽样方法是 Gilks 等^[18] 提出的, 称为 Metropolis-within-Gibbs 算法。

在对参数 ρ 使用随机游动的 Metropolis-Hastings 算法进行抽样时, 根据 LeSage 和 Pace^[17] 在 2009 年的研究, 选择正态分布作为提议分布。假设参数 ρ 目前值为 $\rho^{(t)}$, 经过 $t+1$ 次迭代, 从提议分布 $g(\rho^* | \rho^{(t)}) = f(|\rho^* - \rho^{(t)}|)$ 中产生 ρ^* , 转移过程为 $\rho^* = \rho^{(t)} + kv, v \sim N(0, 1), k$ 为转移参数, 则 ρ^* 的接受概率为 $\alpha(\rho^*, \rho^{(t)}) = \min\{1, R\}$, 其中:

$$R = \frac{\pi(\rho^* | \gamma, \sigma^2) g(\rho^{(t)} | \rho^*)}{\pi(\rho^{(t)} | \gamma, \sigma^2) g(\rho^* | \rho^{(t)})} = \frac{|I_n - \rho^* W| \cdot \exp\left\{-\frac{((I_n - \rho^* W) Y - Z\gamma)^T ((I_n - \rho^* W) Y - Z\gamma)}{2\sigma^2}\right\}}{|I_n - \rho^{(t)} W| \cdot \exp\left\{-\frac{((I_n - \rho^{(t)} W) Y - Z\gamma)^T ((I_n - \rho^{(t)} W) Y - Z\gamma)}{2\sigma^2}\right\}}$$

函数型空间自回归模型 MCMC 估计步骤如下:

步骤 1 给定参数的初始值 $(\rho^{(0)}, \gamma^{(0)}, \sigma^{2(0)})$, 设 $(\rho^{(t)}, \gamma^{(t)}, \sigma^{2(t)})$ 表示参数第 t 次迭代的结果。

步骤 2 从 $\pi(\gamma | \rho^{(t)}, \sigma^{2(t)})$ 中抽取 $\gamma^{(t+1)}$ 。

步骤 3 从 $\pi(\sigma^2 | \rho^{(t)}, \gamma^{(t+1)})$ 中抽取 $\sigma^{2(t+1)}$ 。

步骤 4 根据均值分布 $U(0, 1)$ 中生成随机数 v , 接

下来从转移过程 $\rho^* = \rho^{(t)} + kv$ 里面生成 ρ^* , 那么

$$\rho^{(t+1)} = \begin{cases} \rho^*, & v < \alpha = \min\{1, R\} \\ \rho^{(t)}, & \text{其他} \end{cases}$$

里面的接受概率需要根据得到的最新参数进行计算, 即 $(\rho^{(t+1)}, \gamma^{(t+1)}, \sigma^{2(t+1)})$, 代入得接受概率为

$$R = \frac{\pi(\rho^* | \gamma^{(t+1)}, \sigma^{2(t+1)}) g(\rho^{(t)} | \rho^*)}{\pi(\rho^{(t)} | \gamma^{(t+1)}, \sigma^{2(t+1)}) g(\rho^* | \rho^{(t)})} = \frac{|I_n - \rho^* W| \cdot \exp\left\{-\frac{((I_n - \rho^* W) Y - Z\gamma^{(t+1)})^T ((I_n - \rho^* W) Y - Z\gamma^{(t+1)})}{2\sigma^{2(t+1)}}\right\}}{|I_n - \rho^{(t)} W| \cdot \exp\left\{-\frac{((I_n - \rho^{(t)} W) Y - Z\gamma^{(t+1)})^T ((I_n - \rho^{(t)} W) Y - Z\gamma^{(t+1)})}{2\sigma^{2(t+1)}}\right\}}$$

步骤 5 令 $t=t+1$, 重复上述步骤 2 至步骤 4, 直到所得的马尔科夫链收敛。

令 $\delta = (\rho, \gamma, \sigma^2)$, 凭借上面的步骤生成的序列 $(\delta^{(0)}, \delta^{(1)}, \delta^{(2)}, \dots, \delta^{(N)})$ (假设迭代到 N 次收敛) 不仅是一个马尔科夫链, 而且还平稳分布收敛于 $\pi(\rho, \gamma, \sigma^2 | Y, Z)$ 。

4 数值模拟

下面通过数值模拟对上述构建的模型估计方法进行有效性检验。在对每个参数的估计结果进行模拟分析时, 计算出不同参数的后验均值和后验方差。设置空间权重矩阵:

$$W = (w_{ij})_{n \times n}, w_{ij} = 0.3^{|i-j|} I(i \neq j)$$

其中, $1 \leq i, j \leq n, w_{ij} = 0 (i=j)$ 。

在模拟研究时, 需要将空间权重矩阵进行准化处理, 样本量设置为 $n=200, n=300$ 和 $n=400$ 三种情况。

模型中自回归参数 ρ 设置不同值为 $\rho=0.3, \rho=0.5, \rho=0.7$, 模型方差取 $\sigma^2=0.25, \sigma^2=0.75, \sigma^2=1$ 。对于模型中函数型部分, 函数型系数定义为

$$\beta_1(t) = \sum_{j=1}^{50} b_j \varphi_j(t)$$

其中, $b_1=0.1, b_j=2\sqrt{2}(-1)^{j+1}j^{-2}, j \geq 2; \varphi_j(t) = \sqrt{2} \cos(j\pi t)$;

$\beta_2(t) = \sqrt{2} \sin(\pi t/2) + 3\sqrt{2} \sin(3\pi t/2)$ 。

$\beta_1(t)$ 所对应的变量 $X_i(t)$ 设置为

$$X_i(t) = \sum_{j=1}^{50} \zeta_{ij} a_j \varphi_j(t)$$

其中, $a_j = (-1)^{j+1} j^{-2\alpha}, \alpha = 1$ 和 $1.5; \varphi_j(t) = \sqrt{2} \cos(j\pi t)$,

ζ_{ij} 相互独立且同分布, 服从分布 $\zeta_{ij} \sim U[-\sqrt{3}, \sqrt{3}]$ 。

$\beta_2(t)$ 所对应的变量 $X_i(t)$ 设置为

$$X_i(t) = \sum_{j=1}^{50} \xi_j \varphi_j(t)$$

其中, ξ_j 是一个独立同分布的随机变量, 它满足:

$$\xi_j \sim N(0, ((j-0.5)\pi)^{-2})$$

$$\varphi_j(t) = \sqrt{2} \sin((j-0.5)\pi t)$$

这样就可以生成两个函数型解释变量。

对于模型中参数先验分布中超参数的选择, 本文取 $\gamma_0 = 0_m, \Sigma_\gamma = 10 \times I_m, c=1, d=2$; 对于截断参数 m , 这里根据徐登可等^[13]的研究, 使用函数型主成分分析方法获取, 即

$$m = \min\left\{k: \sum_{i=1}^k \hat{\lambda}_i / \sum_{i=1}^n \hat{\lambda}_i \geq 90\%\right\}$$

数据在进行数值模拟时根据下面的式子生成响应变量 Y 的样本点。

$$Y = (I_n - \rho W)^{-1} \left(\int_0^1 X_1(t) \beta_1(t) dt + \int_0^1 X_2(t) \beta_2(t) dt + \varepsilon \right)$$

其中, $\varepsilon \sim N(0, 1)$ 。

接下来利用所生成响应变量 Y 的样本点, 使用 Gibbs 采样和随机游动 Metropolis-Hastings 算法中的 MCMC 算法对模型中的参数进行贝叶斯估计, 即每个模拟数据在进行 2 000 次迭代后产生的贝叶斯估计值。在对研究的模型进行评估时, 主要对择参数 ρ 和 σ^2 取不同值进行对比分析, 并用参数的偏差和均方误差 (R_{MSE}) 来评估参数的估计效果。当模型中参数的估计偏差和均方误差值越小, 表示模型参数的估计效果越好; 反之, 当模型中参数估计偏差和均方误差值越大, 说明模型参数估算的效果越差。其中均方误差用来评估函数型系数 $\beta(t)$ 的估计效果, 具体计算公式为

$$R_{MSE} = \frac{1}{N} \sum_{i=1}^N (\hat{\beta}(t_i) - \beta(t_i))^2$$

其中, $t_i (i=1, 2, \dots, N, t_i \in [0, 1])$ 表示函数型系数 $\hat{\beta}(t)$ 计算的时间, $N=100$ 。同理, R_{MSE} 越小则表明估计

效果越好。

表 1—表 3 为分别取不同参数值进行横向和纵向对比贝叶斯估计的结果,其中 R_{bias} 表示经过 2 000 次迭代计算的未知参数贝叶斯估计值与给定值的偏差, R_{SD} 则表示是未知参数贝叶斯估计值的标准差。

表 1 $\rho=0.3$ 时模型参数的估计效果

Table 1 Estimation effect of model parameters when $\rho=0.3$

σ^2	n	ρ		σ^2	
		R_{bias}	R_{SD}	R_{bias}	R_{SD}
0.25	200	0.018 71	0.258 52	0.001 28	0.236 83
	300	0.006 31	0.318 05	0.001 02	0.237 67
	400	0.002 50	0.224 52	0.000 94	0.237 06
0.75	200	-0.017 24	0.609 52	0.001 30	0.737 03
	300	0.006 40	0.318 14	0.001 01	0.376 51
	400	0.002 49	0.224 62	0.001 78	0.245 84
1	200	-0.015 72	0.609 51	0.001 31	0.627 12
	300	0.006 32	0.317 95	0.001 03	0.364 82
	400	0.002 25	0.215 82	0.000 78	0.236 95

表 2 $\rho=0.5$ 时模型参数的估计效果

Table 2 Estimation effect of model parameters when $\rho=0.5$

σ^2	n	ρ		σ^2	
		R_{bias}	R_{SD}	R_{bias}	R_{SD}
0.25	200	0.010 34	0.181 42	0.001 16	0.134 52
	300	0.004 25	0.116 15	0.001 01	0.113 76
	400	0.001 54	0.127 32	0.000 76	0.087 06
0.75	200	-0.012 24	0.179 54	0.001 09	0.414 53
	300	0.003 24	0.108 44	0.000 94	0.174 68
	400	0.001 49	0.084 73	0.001 26	0.104 15
1	200	-0.011 73	0.109 31	0.000 83	0.324 17
	300	0.004 13	0.101 75	0.000 65	0.161 85
	400	0.001 25	0.065 82	0.000 78	0.061 65

表 3 $\rho=0.7$ 时模型参数的估计效果

Table 3 Estimation effect of model parameters when $\rho=0.7$

σ^2	n	ρ		σ^2	
		R_{bias}	R_{SD}	R_{bias}	R_{SD}
0.25	200	0.010 12	0.160 82	0.001 04	0.113 62
	300	0.003 13	0.107 16	0.000 84	0.101 46
	400	0.001 22	0.104 23	0.000 61	0.065 12
0.75	200	-0.011 35	0.159 64	0.001 02	0.224 25
	300	0.002 45	0.106 24	0.000 76	0.123 28
	400	0.001 16	0.071 42	0.000 46	0.101 67
1	200	-0.010 26	0.108 21	0.000 74	0.154 21
	300	0.003 02	0.103 35	0.000 54	0.110 68
	400	0.001 08	0.061 62	0.000 32	0.072 06

根据表 1—表 3 中未知参数的 R_{bias} 和 R_{SD} 值,可以看出用贝叶斯方法估计参数的效果比较理想,且估计准确性也较好。其中,随着参数 σ^2 的不断增大,估计偏差 R_{bias} 变得越来越小,未知参数的估计效果也是越来越好的;随着参数样本量 n 的不断增大,估计偏差 R_{bias} 的变化呈不稳定,既有增大也有变小,而 R_{SD} 却是不断变小的,所以也能说明估计的准确性在变高;对于空间位置参数 ρ ,无论它怎样变化, R_{bias} 和 R_{SD} 总是趋于平稳,说明空间位置参数 ρ 无法影响贝叶斯估计的有效性。综上,使用贝叶斯方法对函数型空间自回归模型进行估计,可以提高估计的准确性,同时也能证明估计的有效性。

表 4 和表 5 所展示的是不同位置参数下函数型系数 $\beta_1(t)$ 和 $\beta_2(t)$ 的贝叶斯估计结果,用均方误差 R_{MSE} 来判定其估计效果。

表 4 不同参数下函数型系数 $\beta_1(t)$ 的贝叶斯估计效果

Table 4 Bayesian estimation effect of functional coefficient $\beta_1(t)$ with different parameters

ρ	σ^2	R_{MSE}		
		$n=200$	$n=300$	$n=400$
0.3	0.25	0.553 8	0.436 5	0.241 6
	0.75	0.512 5	0.368 1	0.223 6
	1	0.463 2	0.314 6	0.217 3
0.5	0.25	0.534 7	0.436 6	0.242 0
	0.75	0.513 6	0.367 9	0.223 7
	1	0.452 8	0.314 5	0.217 2
0.7	0.25	0.531 8	0.433 2	0.241 5
	0.75	0.501 2	0.368 1	0.223 5
	1	0.463 2	0.314 6	0.216 9

表 5 不同参数下函数型系数 $\beta_2(t)$ 的贝叶斯估计效果

Table 5 Bayesian estimation effect of functional coefficient $\beta_2(t)$ with different parameters

ρ	σ^2	R_{MSE}		
		$n=200$	$n=300$	$n=400$
0.3	0.25	0.453 2	0.321 3	0.223 5
	0.75	0.422 1	0.287 1	0.210 6
	1	0.321 2	0.201 2	0.190 4
0.5	0.25	0.212 7	0.401 8	0.232 1
	0.75	0.315 9	0.321 4	0.209 2
	1	0.347 4	0.263 3	0.192 6
0.7	0.25	0.364 3	0.324 3	0.231 2
	0.75	0.324 2	0.246 4	0.199 3
	1	0.262 2	0.204 3	0.184 5

由表 4 和表 5 可知:随着模型中参数 σ^2 和 ρ 的不断增大,函数型系数 $\beta_1(t)$ 和 $\beta_2(t)$ 的均方误差 (R_{MSE}) 都是越来越小,说明模型中函数型系数的估计效果也是越来越好。

图 1—图 4 是函数型斜率系数 $\beta_1(t)$ 在不同参数条件下的贝叶斯估计结果与原函数型斜率系数的对比图。图 5—图 8 是函数型斜率系数 $\beta_2(t)$ 在不同参数条件下的贝叶斯估计结果与原函数型斜率系数的对比图。根据图 1—图 8 可以看出:无论是针对斜率函数 $\beta_1(t)$ 或是斜率函数 $\beta_2(t)$, 它们的估计效果都不错,并且 $\beta_1(t)$ 的估计效果随着样本量的增大而逐渐变好,而 $\beta_2(t)$ 的估计效果随着样本量的增大而逐渐变差。

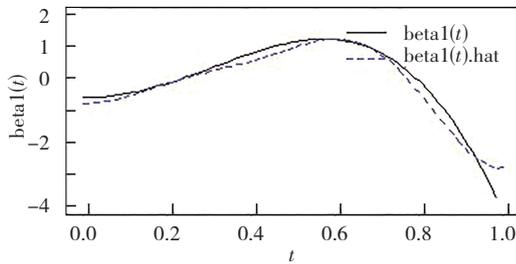


图 1 $\rho=0.3, \sigma^2=0.25, n=200$ 时的函数型系数 $\beta_1(t)$ 的贝叶斯估计结果

Fig. 1 Bayesian estimation results of functional coefficient $\beta_1(t)$ when $\rho=0.3, \sigma^2=0.25, n=200$

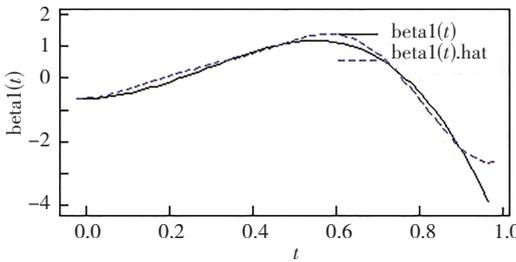


图 2 $\rho=0.5, \sigma^2=0.75, n=200$ 时的函数型系数 $\beta_1(t)$ 的贝叶斯估计结果

Fig. 2 Bayesian estimation results of functional coefficient $\beta_1(t)$ when $\rho=0.5, \sigma^2=0.75, n=200$

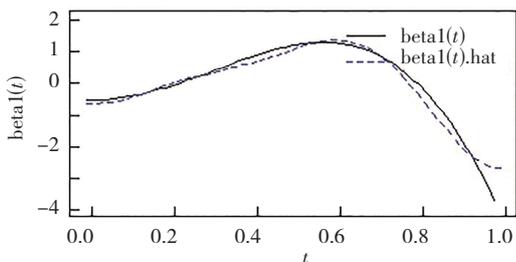


图 3 $\rho=0.7, \sigma^2=1, n=200$ 时的函数型系数 $\beta_1(t)$ 的贝叶斯估计结果

Fig. 3 Bayesian estimation results of functional coefficient $\beta_1(t)$ when $\rho=0.7, \sigma^2=1, n=200$

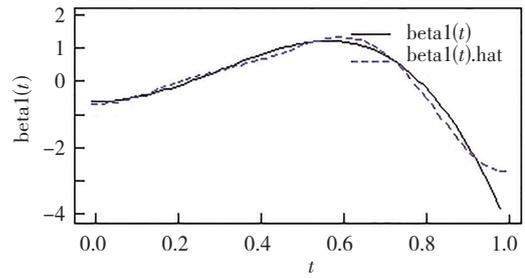


图 4 $\rho=0.5, \sigma^2=1, n=400$ 时的函数型系数 $\beta_1(t)$ 的贝叶斯估计结果

Fig. 4 Bayesian estimation results of functional coefficient $\beta_1(t)$ when $\rho=0.5, \sigma^2=1, n=400$

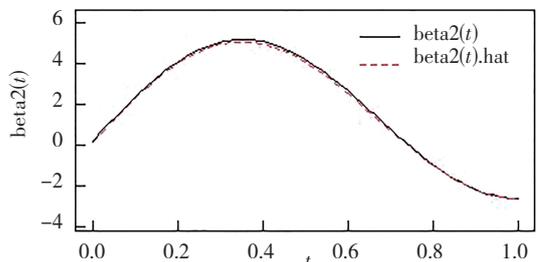


图 5 $\rho=0.3, \sigma^2=0.25, n=200$ 时的函数型系数 $\beta_2(t)$ 的贝叶斯估计结果

Fig. 5 Bayesian estimation results of functional coefficient $\beta_2(t)$ when $\rho=0.3, \sigma^2=0.25, n=200$

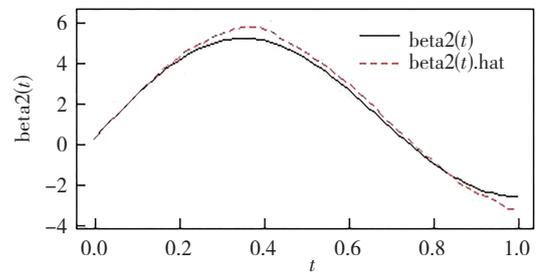


图 6 $\rho=0.5, \sigma^2=0.75, n=200$ 时的函数型系数 $\beta_2(t)$ 的贝叶斯估计结果

Fig. 6 Bayesian estimation results of functional coefficient $\beta_2(t)$ when $\rho=0.5, \sigma^2=0.75, n=200$

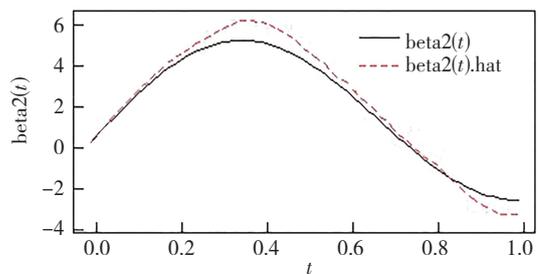


图 7 $\rho=0.7, \sigma^2=1, n=200$ 时的函数型系数 $\beta_2(t)$ 的贝叶斯估计结果

Fig. 7 Bayesian estimation results of functional coefficient $\beta_2(t)$ when $\rho=0.7, \sigma^2=1, n=200$

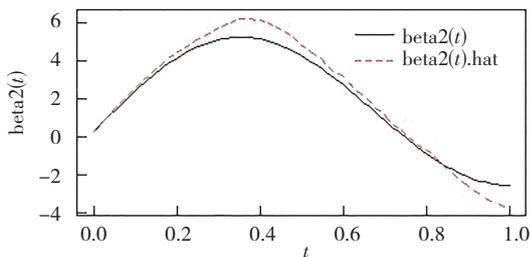


图 8 $\rho=0.5, \sigma^2=1, n=400$ 时的函数型系数 $\beta_2(t)$ 的贝叶斯估计结果

Fig. 8 Bayesian estimation results of functional coefficient $\beta_2(t)$ when $\rho=0.5, \sigma^2=1, n=400$

5 实证分析

下面将模型中函数型响应变量应用的实际数据进行估计,利用空间函数型自回归模型的贝叶斯方法对重庆市主城区 2018-11—2022-10 的月度平均房价数据进行分析。图形中的结果是通过每月平均房价计算得到的每年房价均价,图 9 展示的是 2021 年重庆市主城九大区新房的成交均价,颜色越深房价越高。

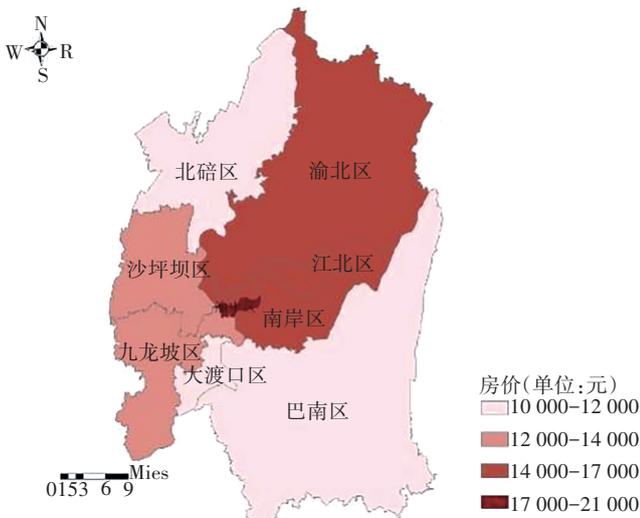


图 9 重庆市主城区新房平均价格空间分布图

Fig. 9 Spatial distribution of average house price of new houses in the main urban area of Chongqing

图 9 可以看出:房价的空间分布是不均匀的,在部分地理相邻的区域有着类似的房价,这说明房价是一个空间自相关变量。接下来使用莫兰指数来证实。根据重庆市主城 9 个区的经纬度,通过 GeoDa 软件,以 Rook 邻接方式(车相邻)生成空间权重矩阵,将得到的空间权重矩阵进行行标准化处理得到最终的空间权重矩阵 W ,然后使用 Stata 软件,利用下面的莫兰指数 (Moran's I) 公式计算空间相关性

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2}$$

其中, $n=9$; w_{ij} 表示空间权重矩阵 W 的 (i, j) 元素。莫兰指数结果见表 6。从表中可以看出:2018—2022 年,每年重庆市主城区房价的莫兰指数都大于 0,所以可以认为重庆主城区的房价存在空间相关性,即认为其存在空间效应。

表 6 2018—2022 年重庆市主城区房价莫兰指数

Table 6 Moran index of housing prices in main urban area of Chongqing from 2018 to 2022

年份	2018	2019	2020	2021	2022
Moran's I	0.091	0.083	0.093	0.133	0.105

为了研究二手房挂牌数量对新房房价的影响,这里假设被解释变量 Y 为 2018-11—2022-10 重庆市主城 9 个区的新房平均价格的对数,函数型解释变量 $X(t)$ 为 2018—2022 年重庆市主城 9 个区的月度二手房挂牌量曲线(数据来源于重庆链家网、安居客等房地产网站)。然后代入模型进行实证分析,其中构建的模型为

$$Y = \rho WY + \int_1^{48} X(t) \beta(t) dt + \varepsilon, t = 1, 2, \dots, 48$$

接下来根据获取的数据以及上述模型,对重庆市主城区新房价格进行模拟,同时通过本文提出的贝叶斯估计方法和先验信息对模型中的未知参数进行估计,得到空间自相关参数的估计为 $\hat{\rho} = 0.5863, \hat{\sigma}^2 = 0.8242$,而函数型系数的估计结果见图 10。从图中可以看出贝叶斯估计后的曲线与原本的函数型系数曲线几乎重合,说明贝叶斯估计的效果非常好,同时将它与重庆主城区新房房价进行对比,会发现重庆主城区二手房挂牌量在一定程度上能够影响新房房价,因为二手房价格可能更低且不会有烂尾风险,所以重庆主城区二手房挂牌量在空间上会对新房房价造成影响。

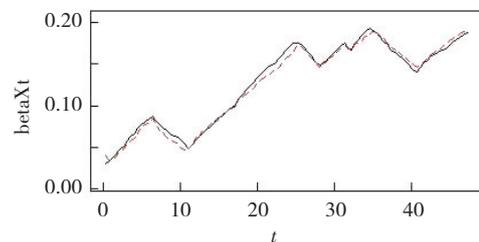


图 10 函数型自变量的斜率函数估计曲线 (注:红色虚线为估计值)

Fig. 10 Slope function estimation curve of functional independent variable (note: red dotted line is the estimated value)

同理,用偏差、标准差和均方误差来检验估计的有效性和准确性,在迭代 2 000 次后,空间自相关参数 ρ 估计结果的偏差和标准差分别为 0.012 47 和 0.268 5,而函数型系数 $\beta(t)$ 的估计均方误差 $R_{MSE} \approx 0.3216$,它

是比较小的,所以可以认为在对空间函数型自回归模型进行实证分析时,使用贝叶斯估计方法是不错的。

6 结 论

利用被解释变量与函数型解释变量在希尔伯特空间上的内生关系,使用 K-L 定理研究了空间函数型自回归模型的贝叶斯估计方法。即对模型中的未知参数采用 Gibbs 采样和随机游动的 Metropolis-Hastings 算法进行估计,模型的模拟研究表明:随着模型中参数 σ^2 和样本量 n 的不断增大,模型中未知参数的估计效果也越来越好;在不同情况下,函数型解释变量的估计效果一直较好,且随着样本量的增大,其估计效果变得越来越好。

对重庆市主城区新房价格进行实证分析,通过 Stata 软件计算出莫兰指数(Moran's I),证实了数据的空间相关性,这样就能使用空间函数型自回归模型对重庆市主城区新房价格进行建模和贝叶斯估计。结果显示:空间自相关参数以及模型方差的估计为 $\hat{\rho} = 0.5863$, $\hat{\sigma}^2 = 0.8242$,并且函数型系数的估计效果也非常不错。在使用贝叶斯估计方法后,模型中函数型系数的均方误差都较小,则可以认为使用贝叶斯估计方法对空间函数型自回归模型进行估计是有效且可行的,同时也能说明重庆市主城区新房平均价格具有空间自相关性,并且会受到二手房挂牌量的影响。这种建模方法在空间统计和空间计量经济学领域具有一定的创新性,同样也可以考虑将这种贝叶斯估计方法应用于其他同类型的模型和其他领域的研究。

参考文献(References):

- [1] RAMSAY J O, DALZELL C J. Some tools for functional data analysis[J]. *Journal of the Royal Statistical Society*, 1991, 53(3): 539—572.
- [2] JAMES G M. Generalized linear models with functional predictors[J]. *Journal of the Royal Statistical Society: Series B(Statistical Methodology)*, 2002, 64(3): 411—432.
- [3] CAI T T, HALL P. Prediction in functional linear regression[J]. *The Annals of Statistics*, 2006, 34(5): 2159—2179.
- [4] ZHOU J J, PENG Q Y. Estimation for functional partial linear models with missing responses[J]. *Statistics & Probability Letters*, 2020, 156: 108598—108610.
- [5] FENG S, TIAN P, HU Y, et al. Estimation in functional single-index varying coefficient model[J]. *Journal of Statistical Planning and Inference*, 2021, 214: 62—75.
- [6] SHI G, DU J, SUN Z, et al. Checking the adequacy of functional linear quantile regression model [J]. *Journal of Statistical Planning and Inference*, 2021, 210: 64—75.
- [7] LAMBERT D M, BROWN J P, FLORAX R J G M. A two-step estimator for a spatial lag model of counts: Theory, small sample performance and an application[J]. *Regional Science and Urban Economics*, 2010, 40(4): 241—252.
- [8] CHENG S, CHEN J. Estimation of partially linear single-index spatial autoregressive model [J]. *Statistical Papers*, 2021, 62: 495—531.
- [9] LAM C, SOUZA P C L. Estimation and selection of spatial weight matrix in a spatial lag model[J]. *Journal of Business & Economic Statistics*, 2020, 38(3): 693—710.
- [10] PINEDA-RÍOS W, GIRALDO R, PORCU E. Functional SAR models: With application to spatial econometrics [J]. *Spatial statistics*, 2019, 29: 145—159.
- [11] 方丽婷,李坤明.空间自回归分位数回归模型的贝叶斯估计[J]. *数量经济技术经济研究*, 2019, 36(9): 102—116.
FANG Li-ting, LI Kun-ming. Bayesian estimation of spatial autoregressive quantile regression model [J]. *Research on Quantitative Economy and Technology Economy*, 2019, 36(9): 102—116.
- [12] QU X, LEE L. Estimating a spatial autoregressive model with an endogenous spatial weight matrix[J]. *Journal of Econometrics*, 2015, 184(2): 209—232.
- [13] 徐登可,田瑞琴.函数型空间自回归模型的贝叶斯估计[J]. *高校应用数学学报 A 辑*, 2022, 37(3): 323—336.
XU Deng-ke, TIAN Rui-qin. Bayesian estimation of functional spatial autoregressive models[J]. *Journal of Mathematics*, 2022, 37(3): 323—336.
- [14] SHIN H. Partial functional linear regression [J]. *Journal of Statistical Planning and Inference*, 2009, 139(10): 3405—3418.
- [15] TERNYNCK C. Spatial regression estimation for functional data with spatial dependency[J]. *Journal de la Société Française de Statistique*, 2014, 155(2): 138—160.
- [16] ALASSANE A, CARBRAL E N. Bayesian estimation of the functional spatial lag model[J]. *Journal of Time Series Econometrics*, 2020, 12(2): 20190047—20190058.
- [17] LESAGE J P, PACE R K. Introduction to spatial econometrics[M]. Boca Raton: CRC Press, 2009.
- [18] GILKS W R, RICHARDSON S, SPIEGELHALTER D J. Markov Chain monte carlo in practice[M]. Boca Raton: CRC Press, 1995.

责任编辑:李翠薇