

部分线性可加空间滞后分位数回归模型的 贝叶斯估计——基于自由节点样条的研究

陶长琪¹, 徐玉婷²

(1. 江西财经大学统计与数据科学学院, 南昌 330013; 2. 南昌师范学院数学与信息科学学院, 南昌 330032)

摘要 部分线性可加空间滞后模型的研究目前几乎都建立在均值回归的估计之上. 均值回归研究协变量对响应变量在条件分布均值位置的影响, 并不能反映两者在条件分布尾部的关系, 容易造成信息的遗漏. 为了克服这一缺陷, 本文构建了部分线性可加空间滞后分位数回归模型 (PLASLQRM), 在采用自由节点样条拟合非参数函数的基础上, 借助逆跳马尔可夫链蒙特卡罗算法对其进行了贝叶斯估计, 形成了基于自由节点样条的贝叶斯分位数回归方法 (BFQ). 为检验 BFQ 方法的有效性, 将该方法与基于惩罚样条的贝叶斯分位数回归方法 (BPQ), 基于自由节点样条的贝叶斯均值回归方法 (BFM) 和广义矩方法 (GMM) 进行了模拟比较. 结果显示, BFQ 方法与其它三种方法相比更不易受异常值的影响, 性能更稳定. 并且, 在误差分布呈尖峰厚尾和偏斜特征时, 该方法对参数部分的估计和非参数部分的拟合都更具优势. 最后, 选取省域碳排放量为实证研究对象, 运用 BFQ 方法分析各类因素对其线性和非线性的影响, 进一步验证了该方法在实际问题中估计参数和非参数的能力.

关键词 部分线性可加空间滞后模型; 分位数回归; 贝叶斯估计; 自由节点样条

Bayesian Estimation of Partially Linear Additive Spatial Lag Quantile Regression Models with Free-Knot Splines

TAO Changqi¹, XU Yuting²

(1. School of Statistics and Data Science, Jiangxi University of Finance and Economics, Nanchang 330013, China; 2. School of Mathematics and Information Science, Nanchang Normal University, Nanchang 330032, China)

收稿日期: 2023-06-15

基金项目: 国家自然科学基金 (71973055, 71773041, 72163008); 南昌师范学院博士科研基金 (NSBSJJ2023013)

Supported by National Natural Science Foundation of China (71973055, 71773041, 72163008); Doctoral Research Foundation of Nanchang Normal University (NSBSJJ2023013)

作者简介: 陶长琪, 博士, 教授, 博士生导师, 研究方向: 计量经济学方法与应用, E-mail: tcq_822@163.com; 通信作者: 徐玉婷, 博士, 副教授, 研究方向: 分位数回归、计量经济学方法与应用, E-mail: 121037798@qq.com.

Abstract At present, the research of partial linear additive spatial lag models is almost based on the estimation of mean regression. However, mean regression studies the influence of covariates on the mean position of response variables in conditional distribution, which can't reflect the relationship between them in the tail of conditional distribution, which is easy to cause the omission of information. In order to overcome this defect, this paper constructs a partial linear additive spatial lag quantile regression model (PLASLQRM). On the basis of fitting nonparametric functions with free-knot splines, Bayesian estimation is carried out with the help of reversible jump Markov chain Monte Carlo algorithm, and a Bayesian quantile regression method (BFQ) based on free-knot splines is formed. In order to test the effectiveness of BFQ method, this method is simulated and compared with Bayesian quantile regression method based on P-splines (BPQ), Bayesian mean regression method based on free-knot splines (BFM) and generalized moment method (GMM). The results show that compared with the other three methods, BFQ method is less vulnerable to extreme values and has more stable performance. Furthermore, when the error distribution is characterized by sharp peak, thick tail and skew, this method has more advantages in the estimation of parametric part and the fitting of nonparametric part. Finally, the provincial carbon emission is selected as the empirical research object, and the influence of various factors on its linearity and nonlinearity is analyzed by using BFQ method, which further verifies the ability of this method to estimate parametric and nonparametric functions in practical problem.

Keywords partially linear additive spatial lag model; quantile regression; Bayesian estimation; free-knot splines

1 引言

在空间计量经济学模型中, 空间滞后模型 (SLM) 受到了学界的广泛关注 (李龙飞, 2021), 也出现了许多估计 SLM 的方法, 如拟极大似然估计法 (Lee, 2004; Liu et al., 2020)、广义矩估计法 (Kelejian and Prucha, 1999; Lee and Liu, 2010) 和广义经验似然估计法 (Jin and Lee, 2019) 等. 然而, 这些方法通常假设变量间的函数关系为线性的, 少有研究者在建立模型前查验是否为非参数或半参数模型. 实际上, 经典的线性模型备受推崇的原因在于线性形式便于估计和理解, 学界很少对线性关系是否存在进行讨论. 然而, 在很多经济问题中, 个体偏好、政策制度等因素变化会导致经济结构发生改变, 使得经济变量对响应变量产生非线性的影响, 若仍然采用线性模型来刻画变量间的非线性关系将出现模型设定偏误问题. 出于模型可解释性以及降维的考虑, 半参数模型被很多学者 (Krisztin, 2018) 引入到空间计量经济学框架中, 其中, 部分线性可加空间滞后模型 (PLASLM) 成为了各种半参数模型中最常用的模型之一. Du et al. (2018) 运用工具变量设计了针对 PLASLM 的广义矩估计法 (GMM). 谢琰等 (2018) 先通过样条逼近可加函数, 再运用截面拟似然的方法对 PLASLM 进行估计. 罗国旺和吴密霞 (2020) 采用 sieve 两阶段最小二乘法估计 PLASLM. Li and Kang (2022) 运用级数近似未知光滑函数, 并提出了两阶段最小二乘法解决高维 PLASLM 的变量选择问题.

绝大多数文献采用上述频率派方法估计 PLASLM, 事实上, 传统的频率派方法常常会面

临目标函数非凸或不可导的问题, 在这种情况下很难得到理论上的解析解, 需要借助数值优化来求解, 很多时候还需要解决非线性优化问题, 求解也将受到挑战. 与频率派方法相比, 贝叶斯方法不仅能利用样本信息还能利用先验信息, 从而在小样本环境下也有更好的估计效果, 且无需最大化任何函数, 拥有更宽松的估计条件. 因此越来越多学者 (方丽婷和李坤明, 2019) 运用贝叶斯方法估计空间计量模型. Chen et al. (2021) 在运用样条逼近可加函数的基础上使用了贝叶斯方法估计 PLASLM, 且是基于惩罚样条实施的贝叶斯估计. 样条逼近光滑函数的效果由节点的个数和位置决定, 节点过少会出现欠拟合, 节点过多则会出现过拟合. 节点的选取至今没有统一的标准, 常见的做法是根据经验选取大量等间距排列的点序列作为节点序列. 为防止出现过拟合, Eilers and Marx (1996) 通过在目标函数中添加样条系数的一阶或二阶差分作为惩罚项来控制节点个数从而形成了惩罚样条. 但是, 惩罚参数的选择会影响模型估计的效果, 且最初的节点个数和位置仍需人为设定, 具有较大的主观性. 为此, Chen and Chen (2021) 将节点个数和位置作为待估参数进行贝叶斯估计, 即基于自由节点样条进行贝叶斯估计 (BFM), 有效解决了节点个数和位置的选取问题.

上述关于 PLASLM 的估计方法都为均值回归方法, 均值回归只关注协变量对响应变量均值水平的影响, 而研究者可能更关心协变量对响应变量条件分布两端的影响, 例如研究房屋价格的影响因素时, 观察人员可能更关注的是高房价的成因, 因此, 仅应用均值回归进行实证分析可能会造成一些重要信息的遗漏. 分位数回归则不同, 它是关于响应变量所有条件分位数的回归, 能全面地反映协变量对不同分位水平响应变量的影响, 尤其能充分刻画具有厚尾特征的因变量整体特征. 而且, 分位数回归不容易受异常值的影响, 对模型的随机误差分布也没有严格的约束, 估计结果相对稳健. 胡亚南等 (2021) 构建了部分线性可加空间滞后分位数回归模型 (PLASLQRM), 并得到了 PLASLQRM 的有效估计. 但是, 一方面, 当前利用分位数回归对部分线性可加空间滞后模型 (PLASLM) 进行分析的研究仍相对较少, 更多地是对参数模型进行分位数回归估计 (李坤明和方丽婷, 2018); 另一方面, 包括胡亚南等 (2021) 在内的研究仍是采用频率派估计方法对 PLASLQRM 进行估计, 仍将面临作为目标函数的损失函数因不可导或不够平滑而难以求解的问题. 因此, 本文在 Chen and Chen (2021) 研究的基础上, 尝试将部分线性可加空间滞后模型和分位数回归相结合, 将分位数回归作为均值回归的稳健替代构建 PLASLQRM, 并基于自由节点样条进行贝叶斯估计, 解决样条节点的选取问题, 以便更好地刻画协变量与响应变量之间的参数或非参数关系.

本文的边际贡献: 一是运用贝叶斯技术将部分线性可加空间滞后模型进行了分位数回归, 弥补现有研究缺乏贝叶斯技术应用于空间分位数回归模型的不足. 二是对部分线性可加空间滞后分位数回归模型设计了基于自由节点样条的贝叶斯估计方法, 使估计方法具有空间自适应性, 解决了样条节点个数和位置的选取问题, 减少了人为的主观干扰, 确保了估计结果的可靠性.

2 模型和方法

2.1 部分线性可加空间滞后分位数回归模型

部分线性可加空间滞后分位数回归模型的一般形式如模型 (1) 所示:

$$\begin{cases} y_i = \rho_\tau \sum_{l \neq i} w_{il} y_l + x_i^\top \alpha_\tau + \sum_{j=1}^p g_{\tau j}(z_{ij}) + \varepsilon_i, & i = 1, \dots, N \\ Pr(\varepsilon_i \leq 0 | x_i, z_i) = \tau, \end{cases} \quad (1)$$

其中 y_i 表示响应变量 Y 的第 i 个观测值, $Y = (y_1, y_2, \dots, y_N)^\top$, $x_i = (x_{i1}, x_{i2}, \dots, x_{iq})^\top$ 表示协变量 X 第 i 个 q 维观测向量, $X^\top = (x_1, x_2, \dots, x_N)^\top$, $z_i = (z_{i1}, z_{i2}, \dots, z_{ip})^\top$ 表示协变量 Z 的第 i 个 p 维观测向量, $Z^\top = (z_1, z_2, \dots, z_N)^\top$, ρ_τ 是反映空间自相关程度的参数, w_{il} 表示 $N \times N$ 维空间权重矩阵 W 的第 i 行第 l 列元素, 且 W 中的对角线元素均为零, $\alpha_\tau = (\alpha_{\tau 1}, \alpha_{\tau 2}, \dots, \alpha_{\tau q})^\top$ 为待估的参数向量, $\tau \in (0, 1)$ 表示分位点. $g_{\tau j}(\cdot)$ 是未知的一元非参数光滑函数, 为确保非参数光滑函数具有可识别性, 通常要求其满足 $\sum_{i=1}^N g_{\tau j}(z_{ij}) = 0$, $j = 1, 2, \dots, p$, 随机误差向量 $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N)^\top$, ε_i 为独立于 (x_i, z_i) 的随机误差, 其在给定 (x_i, z_i) 条件下的 τ 分位数为零. 根据 Yu and Moyeed (2001) 的做法, 假定模型中随机误差 ε_i 服从非对称拉普拉斯分布 (ALD), 记为 $\varepsilon_i \in \text{ALD}(0, \sigma, \tau)$, 此时其概率密度函数为:

$$f(\varepsilon_i | 0, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left(-\tilde{\rho}_\tau\left(\frac{\varepsilon_i}{\sigma}\right)\right),$$

其中 0 为位置参数, σ 为尺度参数, τ 为偏度参数, $\tilde{\rho}_\tau(\mu) = \mu(\tau - I(\mu \leq 0))$ 称为损失函数 (loss function), $I(\cdot)$ 为指示函数. ALD 能用于贝叶斯分析的一个关键性质是其可以表示成正态分布和指数分布的混合形式 (Kozumi and Kobayashi, 2011), 即 $\varepsilon_i = \tilde{k}_1 \xi_i + \sqrt{\tilde{k}_2 \sigma \xi_i} \cdot \eta_i$, 其中 ξ_i 与 η_i 相互独立, $\xi_i \in \exp(\frac{1}{\sigma})$, $\eta_i \in N(0, 1)$, $\tilde{k}_1 = \frac{1-2\tau}{\tau(1-\tau)}$, $\tilde{k}_2 = \frac{2}{\tau(1-\tau)}$. 为简洁起见, 本文后续内容略去参数 ρ , α 和光滑函数 $g_j(\cdot)$ 的下标 τ .

模型 (1) 含有未知的非参数光滑函数 $g_j(\cdot)$, $j = 1, 2, \dots, p$, 显然无法直接对模型进行估计. 为便于估计, 本文先采用 B 样条基函数来逼近 $g_j(\cdot)$, 将其表示成 B 样条基函数的线性组合形式再进行后续估计.

给定节点位置 $\zeta_j = (\zeta_{j1}, \zeta_{j2}, \dots, \zeta_{jk_j})^\top$, $a_j = \zeta_{j0} < \zeta_{j1} < \dots < \zeta_{jk_j} < \zeta_{j(k_j+1)} = b_j$ 和 k_j 个内节点, a_j, b_j 为边界节点, 即 $a_j = \min_{1 \leq i \leq N}(z_{ij})$, $b_j = \max_{1 \leq i \leq N}(z_{ij})$, 选取次数为 m_j 的 B 样条基函数 $B_{jl}(\cdot)$, $l = 1, 2, \dots, K_j$, 将光滑函数 $g_j(\cdot)$ 表示成如下线性组合形式:

$$g_j(z_{ij}) = \sum_{l=1}^{K_j} B_{jl}(z_{ij}) \beta_j = B_i^\top(z_{ij}) \beta_j,$$

其中 $B_i(z_{ij}) = (B_{j1}(z_{ij}), B_{j2}(z_{ij}), \dots, B_{jK_j}(z_{ij}))^\top$ 表示由节点位置决定的样条基向量, 样条基的个数为 $K_j = 1 + m_j + k_j$, $\beta_j = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jK_j})^\top$ 表示样条系数向量. 通过对 $g_j(\cdot)$ 的样条逼近, 可将模型 (1) 转化为“线性”模型 (2):

$$y_i = \rho \sum_{l \neq i} w_{il} y_l + x_i^\top \alpha + B_i^\top(z_i) \beta + \varepsilon_i, \quad i = 1, \dots, N, \quad (2)$$

其中 $B_i(z_i) = (B_i^\top(z_{i1}), B_i^\top(z_{i2}), \dots, B_i^\top(z_{ip}))^\top$, $\beta = (\beta_1^\top, \beta_2^\top, \dots, \beta_p^\top)^\top$ 均表示维数为 $K \times 1$ 的向量, $K = \sum_{j=1}^p K_j$. 进一步可得到模型 (2) 的矩阵形式:

$$Y = \rho W Y + X^\top \alpha + B^\top(Z) \beta + \varepsilon,$$

其中 $B^\top(Z)$ 表示维数为 $N \times K$ 的矩阵, 其第 i 行为 $B_i^\top(z_i)$. 同时, 记 $1_N = (1, 1, \dots, 1)^\top$ 为 $N \times 1$ 维向量, $\tilde{B}_j = (B_1(z_{1j}), B_2(z_{2j}), \dots, B_N(z_{Nj}))^\top$ 为 $N \times K_j$ 维矩阵, 根据光滑函数的样条逼近形式, 模型的可识别条件 $\sum_{i=1}^N g_j(z_{ij}) = \sum_{i=1}^N \sum_{l=1}^{K_j} B_{jl}(z_{ij})\beta_j = 0$ 等价于 $1_N^\top \tilde{B}_j \beta_j = 0$, 记 $Q_j = 1_N^\top \tilde{B}_j$, 因此, 模型可识别的约束条件等价于 $Q_j \beta_j = 0, j = 1, 2, \dots, p$.

2.2 贝叶斯估计方法

贝叶斯方法进行估计的过程就是模拟参数后验分布的过程, 模拟采样的方法决定算法的效率. 一方面, 为了解决模型 (1) 的估计问题, 本节构建了 Metropolis-Hastings 算法 (M-H) 和吉布斯算法 (Gibbs) 相结合的贝叶斯分位数回归方法来对模型参数的后验分布实施采样, 采样方法有效且易于实施; 另一方面, 样条基函数逼近光滑函数时, 通常人为设定节点的个数和位置, 这种主观设定会造成函数估计出现欠拟合或过拟合, 本研究将节点的个数和位置作为待估参数, 采用逆跳马尔可夫链蒙特卡罗 (RJMC MC) 算法在维数可变的空间内对节点的个数和位置进行采样, 有效解决了节点个数和位置的选取问题, 实现了节点的选取“自由”. RJMC MC 算法 (Green, 1995) 是种更适用于一般状态空间的 M-H 算法, 与普通的 M-H 算法需要固定参数维度空间不同, 它可以在不同维度的参数子空间内跳跃采样. 借助 RJMC MC 算法这种性能, 本文借鉴 Chen and Wang (2018) 的方法对节点序列在移动中仅能改变一个节点位置的传统做法进行改进, 使节点序列能根据空间状态调整所有节点的位置, 从而使算法在确定节点位置时具有空间自适应性, 促进了算法的快速收敛, 提升了算法的运行速度. 因此, 对模型 (1) 基于自由节点样条进行贝叶斯估计具有一定的实践意义.

在对模型进行具体估计之前, 首先构建模型的贝叶斯分析框架.

记 $A(\rho) = I_N - \rho W$, $\tilde{Y} = A(\rho)Y = (\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_N)^\top$, 并将 ALD 的混合形式代入模型 (2), 则模型 (2) 可改写为:

$$\tilde{y}_i = x_i^\top \alpha + B_i^\top(z_i)\beta + \tilde{k}_1 \xi_i + \sqrt{\tilde{k}_2 \sigma \xi_i} \cdot \eta_i, \quad i = 1, \dots, N. \quad (3)$$

模型 (3) 的似然函数为:

$$L(\tilde{Y}|\rho, \theta, \sigma, \xi) = \left(\prod_{i=1}^N (2\pi \tilde{k}_2 \sigma \xi_i)^{-1/2} \right) \exp \left(- \sum_{i=1}^N \frac{(\tilde{y}_i - x_i^\top \alpha - B_i^\top(z_i)\beta - \tilde{k}_1 \xi_i)^2}{2\tilde{k}_2 \sigma \xi_i} \right).$$

其中 $\xi = (\xi_1, \xi_2, \dots, \xi_N)^\top$, $\theta = (\alpha^\top, \beta^\top)^\top$, $\eta = (\eta_1, \eta_2, \dots, \eta_N)^\top$. 由于 $Y = A(\rho)^{-1} \tilde{Y}$, 则模型 (2) 的似然函数可表示为:

$$L(Y|\rho, \theta, \sigma, \xi) = |A(\rho)| \left(\prod_{i=1}^N (2\pi \tilde{k}_2 \sigma \xi_i)^{-1/2} \right) \exp \left(- \sum_{i=1}^N \frac{((A(\rho)Y)_i - x_i^\top \alpha - B_i^\top(z_i)\beta - \tilde{k}_1 \xi_i)^2}{2\tilde{k}_2 \sigma \xi_i} \right).$$

进一步得出模型 (2) 的似然函数对应的矩阵形式为:

$$L(Y|\rho, \theta, \sigma, \xi) = |A(\rho)| (2\pi \tilde{k}_2 \sigma)^{-\frac{N}{2}} |\text{diag}(\xi)|^{-\frac{1}{2}} \exp \left\{ - \frac{1}{2\tilde{k}_2 \sigma} (A(\rho)Y - \phi(X, Z)\theta - \tilde{k}_1 \xi)^\top \times \right. \\ \left. (\text{diag}(\xi))^{-1} (A(\rho)Y - \phi(X, Z)\theta - \tilde{k}_1 \xi) \right\}, \quad (4)$$

其中 $\phi(X, Z) = (X^\top, B^\top(Z))$, $\text{diag}(\xi) = \text{diag}(\xi_1, \xi_2, \dots, \xi_N)$.

1. 先验设定

为进行下一步的贝叶斯估计, 接下来设定各参数的先验分布.

1) 空间相关系数 ρ 的先验分布. 通常设 ρ 服从均匀分布 $U(\tilde{\lambda}_{\max}^{-1}, \tilde{\lambda}_{\min}^{-1})$, $\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max}$ 分别为行标准化空间权重矩阵 W 的最大、最小特征值. 可见, ρ 的先验分布 $\pi(\rho) \propto 1$.

2) 参数向量 α 的先验分布. 设定 α 服从均值为 0, 协方差矩阵为 $\varphi_\alpha \sigma I_q$ 的多元正态分布. φ_α 为 α 的先验超参数, 其服从超参数为 $\frac{\gamma_\alpha}{2}$ 和 $\frac{s_\alpha^2}{2}$ 的逆伽马分布 (IG).

3) 参数向量 β_j 的先验分布. 设定 β_j 服从均值为 0, 协方差矩阵为 $\varphi_j \sigma I_{K_j}$ 的多元正态分布且满足可识别约束条件 $Q_j \beta_j = 0$. φ_j 为 β_j 的先验超参数, 其服从超参数为 $\frac{\gamma_j}{2}$ 和 $\frac{s_j^2}{2}$ 的逆伽马分布 (IG).

4) 内节点个数 k_j 和位置 ζ_j 的先验分布. 设定 k_j 服从均值为 λ_j 的泊松分布, 由于 k_j 个内节点 $a_j = \zeta_{j0} < \zeta_{j1} < \dots < \zeta_{jk_j} < \zeta_{j(k_j+1)} = b_j$ 将区间 $[a_j, b_j]$ 分成 $k_j + 1$ 个子区间, 子区间长度与总体区间长度的比值为 $\frac{\zeta_{jh} - \zeta_{j(h-1)}}{b_j - a_j}$, $h = 1, 2, \dots, k_j + 1$, 通常设定该比值服从狄利克雷分布 $D(1, 1, \dots, 1; 1)$, 因此内节点位置 $\zeta_j = (\zeta_{j1}, \zeta_{j2}, \dots, \zeta_{jk_j})^\top$ 可看作区间 $[a_j, b_j]$ 上的顺序统计量, 其先验分布应为:

$$\pi(\zeta_j | k_j) = \frac{k_j!}{(b_j - a_j)^{k_j}} \Delta_j, \Delta_j = I\{a_j = \zeta_{j0} < \zeta_{j1} < \dots < \zeta_{jk_j} < \zeta_{j(k_j+1)} = b_j\}.$$

5) 辅助参数 ξ_j 的先验分布. ξ_j 服从参数为 $\frac{1}{\sigma}$ 的指数分布, σ 为其超参数, 设 σ 服从超参数为 $\frac{\gamma_\sigma}{2}$ 和 $\frac{s_\sigma^2}{2}$ 的逆伽马分布.

2. 联合后验密度及条件后验分布

基于自由节点样条进行贝叶斯估计之前还需要得到各未知参数的条件后验分布, 根据各个参数先验分布相互独立的设定以及先验分布和似然函数的表达式 (4), 再由贝叶斯定理和联合后验分布, 可得所有参数的联合后验密度的核如下:

$$\begin{aligned} p(\rho, \theta, k, \zeta, \sigma, \xi, \varphi | Y) &\propto |A(\rho)| (2\pi \tilde{k}_2 \sigma)^{-\frac{N}{2}} |\text{diag}(\xi)|^{-\frac{1}{2}} \frac{1}{\tilde{\lambda}_{\min}^{-1} - \tilde{\lambda}_{\max}^{-1}} \times \\ &\exp \left\{ -\frac{1}{2\tilde{k}_2 \sigma} (A(\rho)Y - \phi(X, Z)\theta - \tilde{k}_1 \xi)^\top (\text{diag}(\xi))^{-1} \times \right. \\ &\left. (A(\rho)Y - \phi(X, Z)\theta - \tilde{k}_1 \xi) \right\} \left(\frac{1}{\sigma} \right)^N \exp \left\{ -\frac{1}{\sigma} 1_N^\top \xi \right\} \sigma^{-\frac{\gamma_\sigma}{2} - 1} \times \\ &\exp \left\{ -\frac{s_\sigma^2}{2\sigma} \right\} (2\pi \varphi_\alpha \sigma)^{-\frac{q}{2}} \exp \left\{ -\frac{\alpha^\top \alpha}{2\varphi_\alpha \sigma} \right\} \varphi_\alpha^{-\frac{\gamma_\alpha}{2} - 1} \exp \left\{ -\frac{s_\alpha^2}{2\varphi_\alpha} \right\} \times \\ &\prod_{j=1}^p (2\pi \varphi_j \sigma)^{-\frac{K_j}{2}} \exp \left\{ -\frac{\beta_j^\top \beta_j}{2\varphi_j \sigma} \right\} I\{Q_j \beta_j = 0\} \varphi_j^{-\frac{\gamma_j}{2} - 1} \exp \left\{ -\frac{s_j^2}{2\varphi_j} \right\} \times \\ &\prod_{j=1}^p \frac{k_j!}{(b_j - a_j)^{k_j}} \Delta_j \frac{\lambda_j^{k_j}}{k_j!} \exp\{-\lambda_j\}, \end{aligned} \quad (5)$$

其中 $k = (k_1, \dots, k_p)^\top$, $\zeta = (\zeta_1^\top, \dots, \zeta_p^\top)^\top$, $\varphi = (\varphi_\alpha, \varphi_1, \dots, \varphi_p)^\top$, 由联合后验密度的核表达式 (5), 可得到各个参数的全条件后验分布密度函数.

1) 对于 ρ , 其条件后验分布密度函数正比于下式:

$$\pi(\rho|\theta, k, \zeta, \sigma, \xi, \varphi, Y) \propto |A(\rho)| \exp \left\{ -\frac{1}{2\tilde{k}_2\sigma} (A(\rho)Y - X^\top \alpha - B^\top(Z)\beta - \tilde{k}_1\xi)^\top \times \right. \\ \left. (\text{diag}(\xi))^{-1} (A(\rho)Y - X^\top \alpha - B^\top(Z)\beta - \tilde{k}_1\xi) \right\}. \quad (6)$$

2) 对于 ξ_i , 其条件后验分布密度函数具有以下特征:

$$\pi(\xi_i|\rho, \theta, k, \zeta, \sigma, \varphi, Y) \propto \xi_i^{-1/2} \exp \left\{ -\frac{1}{2}(\delta_i \xi_i^{-1} + \tilde{k}_3 \xi_i) \right\},$$

其中 $\delta_i = (y_i - \rho \sum_{l \neq i} w_{il} y_l - x_i^\top \alpha - B_i^\top(z_i)\beta)^2 / \tilde{k}_2\sigma$, $\tilde{k}_3 = \frac{\tilde{k}_1^2}{\tilde{k}_2\sigma} + \frac{2}{\sigma}$, 上式表明 ξ_i 的条件后验分布密度正比于广义逆高斯分布 (GIG) 的核, 因此, 其条件后验分布为:

$$\pi(\xi_i|\rho, \theta, k, \zeta, \sigma, \varphi, Y) \propto \text{GIG}(1/2, \delta_i, \tilde{k}_3), \quad i = 1, \dots, N. \quad (7)$$

3) 对于 $\varphi = (\varphi_\alpha, \varphi_1, \dots, \varphi_p)^\top$ 中的 φ_α 和 $\varphi_j, j = 1, \dots, p$, 条件后验分布分别为:

$$\pi(\varphi_\alpha|\rho, \theta, k, \zeta, \sigma, \xi, Y) \propto \text{IG} \left(\frac{q}{2} + \frac{\gamma_\alpha}{2}, \frac{\alpha^\top \alpha}{2\sigma} + \frac{s_\alpha^2}{2} \right), \quad (8)$$

$$\pi(\varphi_j|\rho, \theta, k, \zeta, \sigma, \xi, Y) \propto \text{IG} \left(\frac{K_j}{2} + \frac{\gamma_j}{2}, \frac{\beta_j^\top \beta_j}{2\sigma} + \frac{s_j^2}{2} \right). \quad (9)$$

在给定 ρ, ξ, φ 的条件下, 由式 (5) 可得 θ, k, ζ, σ 的联合后验分布为:

$$\pi(\theta, k, \zeta, \sigma|\rho, \xi, \varphi, Y) \propto \sigma^{-\frac{q+K}{2}} |\Xi|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2\sigma} (\theta - \tilde{\theta})^\top \Xi (\theta - \tilde{\theta}) \right\} I\{Q\beta = 0\} \times \\ \sigma^{-\frac{\gamma_\sigma + 3N}{2} - 1} \left(\frac{s_\alpha^2 + S^2}{2} \right)^{(\gamma_\sigma + 3N)/2} \exp \left\{ -\frac{1}{\sigma} \left(\frac{s_\alpha^2 + S^2}{2} \right) \right\} \times \\ |\Xi|^{-\frac{1}{2}} \left(\frac{s_\alpha^2 + S^2}{2} \right)^{-\frac{\gamma_\sigma + 3N}{2}} \prod_{j=1}^p \left(\frac{\lambda_j \varphi_j^{-\frac{1}{2}}}{b_j - a_j} \right)^{k_j} \Delta_j \exp\{-\lambda_j\}, \quad (10)$$

其中 $\Xi = \text{diag}(J^{-1}) + \frac{1}{\tilde{k}_2} \phi^\top(X, Z)(\text{diag}(\xi))^{-1} \phi(X, Z)$, $\text{diag}(J^{-1}) = \text{diag}\{\varphi_\alpha^{-1} I_q, \varphi_1^{-1} I_{K_1}, \dots, \varphi_p^{-1} I_{K_p}\}$, $Q = (Q_1, \dots, Q_p)$, $\tilde{\theta} = \frac{1}{\tilde{k}_2} \Xi^{-1} \phi^\top(X, Z)(\text{diag}(\xi))^{-1} (A(\rho)Y - \tilde{k}_1\xi)$, $S^2 = 2 \cdot 1_N^\top \xi + \frac{1}{\tilde{k}_2} (A(\rho)Y - \tilde{k}_1\xi)^\top (\text{diag}(\xi))^{-1} (A(\rho)Y - \tilde{k}_1\xi) - \tilde{\theta}^\top \Xi \tilde{\theta}$.

4) 对于 σ , 从联合后验分布 (10) 中可得其条件后验分布为逆伽马分布:

$$\pi(\sigma|\rho, \theta, k, \zeta, \xi, \varphi, Y) \propto \text{IG} \left(\frac{\gamma_\sigma + 3N}{2}, \frac{s_\sigma^2 + S^2}{2} \right). \quad (11)$$

5) 对于 $\theta = (\alpha^\top, \beta^\top)^\top$, 从式 (10) 中可得其条件后验分布为有约束的多元正态分布:

$$\pi(\theta|\rho, k, \zeta, \sigma, \xi, \varphi, Y) \propto N(\tilde{\theta}, \Xi^{-1}\sigma) I\{Q\beta = 0\}. \quad (12)$$

其中参数 α 服从协方差矩阵为 $\Sigma_\alpha = \Xi_\alpha^{-1}\sigma$, $\Xi_\alpha = \varphi_\alpha^{-1}I_q + \frac{1}{k_2}X(\text{diag}(\xi))^{-1}X^\top$, 均值向量为 $\mu_\alpha = \frac{1}{k_2}\Xi_\alpha^{-1}X(\text{diag}(\xi))^{-1}(A(\rho)Y - B^\top(Z)\beta - \tilde{k}_1\xi)$ 的多元正态分布, 即:

$$\pi(\alpha|\rho, \beta, k, \zeta, \sigma, \xi, \varphi, Y) \propto N(\mu_\alpha, \Sigma_\alpha).$$

参数 β_j 服从协方差矩阵为 $\Sigma_j = \Xi_j^{-1}\sigma$, $\Xi_j = \varphi_j^{-1}I_{K_j} + \frac{1}{k_2}\tilde{B}_j^\top(\text{diag}(\xi))^{-1}\tilde{B}_j$, 均值向量为 $\mu_j = \frac{1}{k_2}\Xi_j^{-1}\tilde{B}_j^\top(\text{diag}(\xi))^{-1}(A(\rho)Y - X^\top\alpha - \sum_{m \neq j}^p \tilde{B}_m\beta_m - \tilde{k}_1\xi)$ 的多元正态分布, 即,

$$\pi(\beta_j|\rho, \alpha, \beta_{-j}, k, \zeta, \sigma, \xi, \varphi, Y) \propto N(\mu_j, \Sigma_j), \quad j = 1, \dots, p.$$

此外, β_j 还需满足可识别条件 $Q_j\beta_j = 0$. 参照 Panagiotelis and Smith (2008) 的做法, 只需从多元正态分布 $N(\mu_j, \Sigma_j)$ 中抽取候选值 β_j^* , 并用式 (13) 对 β_j^* 进行调整, 便可获得满足可识别条件的 β_j , 式 (13) 的表达式为:

$$\beta_j = \beta_j^* - \Sigma_j Q_j^\top (Q_j \Sigma_j Q_j^\top)^{-1} Q_j \beta_j^*. \quad (13)$$

6) 对于 $k = (k_1, \dots, k_p)^\top$ 和 $\zeta = (\zeta_1^\top, \dots, \zeta_p^\top)^\top$, 从式 (10) 中可知节点的个数 k_j 和位置 ζ_j , $j = 1, \dots, p$ 的联合后验分布正比于:

$$\pi(k_j, \zeta_j|\rho, \alpha, \beta, k_{-j}, \zeta_{-j}, \sigma, \xi, \varphi, Y) \propto |\Xi|^{-\frac{1}{2}} \left(\frac{s_\sigma^2 + S^2}{2} \right)^{-\frac{(\gamma_\sigma + 3N)}{2}} \left(\frac{\lambda_j \varphi_j^{-\frac{1}{2}}}{b_j - a_j} \right)^{k_j} \Delta_j \exp\{-\lambda_j\}. \quad (14)$$

3. 抽样设计

从所有参数的条件后验分布可以看出, 参数 ρ 的条件后验分布并不是常见的密度分布函数, 直接进行抽样模拟存在较大的困难. 为解决这一问题, 本文采用 Chen and Chen (2021) 的做法, 对参数 ρ 实施 M-H 抽样, 即, 取位置参数为 ρ , 尺度参数为 δ_ρ , 在区间 $(-1, 1)$ 上的截尾柯西分布为建议分布, 从中抽取 ρ 的候选值 ρ^* , 并计算候选值 ρ^* 的接受率 R_ρ , 若来自均匀分布 $U(-1, 1)$ 的随机数大于接受率 R_ρ , 则接受候选值 ρ^* 作为 ρ 的抽样值, 否则拒绝该候选值. 其中, δ_ρ 为调整参数, R_ρ 的计算方法为:

$$R_\rho = \min \left\{ 1, \frac{\pi(\rho^*|\theta, k, \zeta, \sigma, \xi, \varphi, Y)}{\pi(\rho|\theta, k, \zeta, \sigma, \xi, \varphi, Y)} \times C_\rho \right\},$$

$$C_\rho = \frac{\arctan((1 - \rho)/\delta_\rho) - \arctan((-1 - \rho)/\delta_\rho)}{\arctan((1 - \rho^*)/\delta_\rho) - \arctan((-1 - \rho^*)/\delta_\rho)}.$$

同时, 由于节点的个数 k_j 和位置 ζ_j 是变化的, 因此本文采用 RJMCMC 算法在维数可变的空间内对 k_j 和 ζ_j 进行抽样. 按照基于自由节点样条贝叶斯估计的一般做法, 在每一次的抽样中, RJMCMC 算法通常采用三种方式改变节点个数与位置: 增加一个节点 (Birth), 删除一个节点 (Death) 和移动一个节点 (Move). 假定当前 k_j 个节点以概率 b_{k_j} 增加一个新节点, 以概率 d_{k_j} 删除一个已存在的节点, 则移动一个已有节点的概率为 $1 - b_{k_j} - d_{k_j}$. 借鉴 Poon and Wang (2013) 的做法, 设概率 $b_{k_j} = 0.4 \min\{1, \lambda_j/(k_j + 1)\}$, $d_{k_j} = 0.4 \min\{1, k_j/\lambda_j\}$, 显

然节点的状态转移满足细致平衡条件 (detailed balance condition).

三种改变节点方式的详情如下:

增加一个节点, 是指从边界节点构成的均匀分布 $U(a_j, b_j)$ 中随机选取一个数作为节点添加到已有节点中并重新排序形成新的节点位置 ζ_j^* 作为候选节点位置, 此时节点的个数为 $k_j^* = k_j + 1$, 候选节点位置的接受率为:

$$R_j^b = \min \left\{ 1, \frac{1}{\sqrt{\varphi_j}} \Lambda \right\},$$

其中 $\Lambda = (\frac{|\Xi|}{|\Xi^*|})^{1/2} (\frac{s_\sigma^2 + S^2}{s_\sigma^2 + S^{*2}})^{(\gamma_\sigma + 3N)/2}$, Ξ^* 和 S^* 指 Ξ 和 S 在位置 ζ_j^* 处的取值.

删除一个节点, 是指从已有的节点位置 ζ_j 中随机选中一个节点后再去除该节点, 将剩余节点构成的新节点位置 ζ_j^* 作为候选节点位置, 此时节点的个数为 $k_j^* = k_j - 1$, 候选节点位置的接受率为 $R_j^d = \min\{1, \sqrt{\varphi_j} \Lambda\}$.

移动一个节点, 是指从已存在的节点中随机选取一个节点进行移动以改变其位置, 但这种移动方法每次仅能改变一个节点的位置, 其它节点位置仍保持不变. 为使每次迭代都能改变所有节点的位置, 使节点位置的选取具有空间自适应性, 以便加快算法的收敛速度, 借鉴 Chen and Wang (2018) 的方法, 任选一个 k_j 维的方向向量 $c_j = (c_{j1}, \dots, c_{jk_j})^\top$, 并从位置参数为 0, 尺度参数为 δ_{ψ_j} , 且在区间 (ψ_{j1}, ψ_{j2}) 上的截尾柯西分布中抽取数 ψ_j , 将节点位置 ζ_j 沿向量 $\psi_j c_j$ 平移, 最终得到重新排序后的候选节点位置 $\zeta_j^* = \zeta_j + \psi_j c_j$, 此时节点的个数为 $k_j^* = k_j$, δ_{ψ_j} 为调整参数, 区间 (ψ_{j1}, ψ_{j2}) 的定义为:

$$\Omega_j = \{\psi_j : \zeta_j^* = \zeta_j + \psi_j c_j, a_j < \zeta_{ji}^* < b_j, i = 1, 2, \dots; k_j\} = (\psi_{j1}, \psi_{j2}).$$

候选节点位置 ζ_j^* 的接受率为:

$$R_j^m = \min \left\{ 1, \frac{\arctan(\psi_{j2}/\delta_{\psi_j}) - \arctan(\psi_{j1}/\delta_{\psi_j})}{\arctan((\psi_{j2} - \psi_j)/\delta_{\psi_j}) - \arctan((\psi_{j1} - \psi_j)/\delta_{\psi_j})} \Lambda \right\}.$$

其他参数的条件后验分布均为标准的分布函数, 可以用 Gibbs 抽样法直接进行抽样. 因此, 根据上述各待估参数的条件后验分布, 对模型 (1) 基于自由节点样条进行贝叶斯估计的步骤如下:

步骤 1 给定待估参数的初值 $\rho^{(0)}, \theta^{(0)}, k^{(0)}, \zeta^{(0)}, \sigma^{(0)}, \xi^{(0)}, \varphi^{(0)}$, 假设 $\rho^{(t)}, \theta^{(t)}, k^{(t)}, \zeta^{(t)}, \sigma^{(t)}, \xi^{(t)}, \varphi^{(t)}$ 是第 t 次迭代的结果;

步骤 2 从条件后验分布式 (6) 中运用 M-H 抽样法生成 $\rho^{(t+1)}$;

步骤 3 从条件后验分布式 (11) 中运用 Gibbs 抽样法生成 $\sigma^{(t+1)}$;

步骤 4 从条件后验分布式 (14) 中运用 RJMCMC 算法生成 $k^{(t+1)}$ 和 $\zeta^{(t+1)}$;

步骤 5 从条件后验分布式 (12) 中运用 Gibbs 抽样法生成 $\theta^{(t+1)}$ 再运用式 (13) 得到调整后的 $\beta^{(t+1)}$;

步骤 6 从条件后验分布式 (8) 和 (9) 中运用 Gibbs 抽样法生成 $\varphi^{(t+1)}$;

步骤 7 从条件后验分布式 (7) 中运用 Gibbs 抽样法生成 $\xi^{(t+1)}$;

步骤 8 每次用本期生成的数据代入下一期生成新的数据, 重复步骤 2 至 7, 直至收敛.

3 数值模拟

3.1 数据出现异常值时四种方法的比较

实际数据难免会出现异常值, 如何有效处理含异常值的数据是学界关注的话题之一. 该数值模拟旨在阐述, 当数据受到异常值影响时, 与其它方法相比, 本文所提方法受异常值的影响较小, 模型的稳定性具有较大改进.

按如下模型生成模拟数据:

$$Y = \rho WY + X^\top \alpha + g_1(Z_1) + g_2(Z_2) + \varepsilon_\tau, \quad (15)$$

其中 $Y = (y_1, y_2, \dots, y_N)^\top$, $X^\top = (x_1, x_2, \dots, x_N)$, $x_i = (x_{i1}, x_{i2})^\top$, 且 x_{i1}, x_{i2} 均服从标准正态分布, $Z_1 = (z_{11}, z_{21}, \dots, z_{N1})^\top$, $Z_2 = (z_{12}, z_{22}, \dots, z_{N2})^\top$, 且 z_{i1}, z_{i2} 相互独立分别服从均匀分布 $U(0, 1)$ 和 $U(-1, 1)$, $i = 1, \dots, N$. 为便于理解本文提出的方法取简单的光滑函数 $g_1(Z_1) = Z_1$, $g_2(Z_2) = Z_2^2$, $N = 100$. $\varepsilon_\tau = \varepsilon - F^{-1}(\tau)$, F 为 ε 的分布函数, 此时从 ε 中减去 $F^{-1}(\tau)$ 可使得随机误差项的 τ 分位数为 0. $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)^\top$, 此次模拟取随机误差 ε_i 服从标准正态分布, 基于 Rook 相邻构造空间权重矩阵 W , 相应的变量系数设为 $\rho = 0.5$, $\alpha = (\alpha_1, \alpha_2)^\top = (1, -1)^\top$.

再按如下机制生成受异常值影响的模拟数据:

找到数据模型 (15) 中响应变量 Y 的 0.05 和 0.95 分位数, 分别记为 $Y_{0.05}$, $Y_{0.95}$, 对 Y 中所有大于 0.95 分位数的点 y 加上 10, 所有小于 0.05 分位数的点 y 减去 10, 即, 若 $y > Y_{0.95}$, 则 $y_{\text{new}} = y + 10$; 若 $y < Y_{0.05}$, 则 $y_{\text{new}} = y - 10$, Y 中其它数值保持不变, 由此得到含异常值的响应变量作为新的响应变量 Y_{new} 进行模型估计.

接下来对含异常值的数据模型 (15) 运用基于自由节点样条的贝叶斯分位数回归方法 (BFQ), 基于惩罚样条的贝叶斯分位数回归方法 (BPQ), 基于自由节点样条的均值回归方法 (BFM) 和 Du (2018) 提出的广义矩方法 (GMM) 进行估计. 前三者采用 MCMC 算法进行计算, 后者利用工具变量法进行计算.

四种方法均采用三次 B 样条基函数来近似光滑函数, BFQ 方法的先验超参数设为 $\gamma_\sigma = s_\sigma^2 = 1$, $\gamma_\alpha = \gamma_j = 1$, $s_\alpha^2 = 0.005$, $s_j^2 = 0.005$, $\lambda_j = 2$, $j = 1, \dots, p$, 调整 δ_ρ 和 δ_{ψ_j} 的值使对应参数的接受率在 25% 左右, 并从各参数的先验分布中抽取初值. BPQ 方法采用二阶随机游走作为惩罚, 取 $k_1 = 22$, $k_2 = 26$ 个等间距内节点, 其它先验超参数的设置与 BFQ 相同. BFM 方法中参数 σ^2 取逆伽马分布 $IG(1, 1)$ 作为先验, 其它先验超参数的设置与 BFQ 一致. GMM 中内节点个数为 $N^{1/5}$ 的整数部分.

考虑 τ 取 0.1、0.5、0.9 三个分位点, 分位数回归方法 BFQ 和 BPQ 在三个分位点均迭代 6000 次, 舍弃前面 4000 次, 取后面 2000 个抽样样本的均值作为参数的估计值. 均值回归方法 BFM 和 GMM 只能在 0.5 分位点处进行估计. 关于光滑函数的估计效果, 采用平均绝对误差 (MADE) 来衡量:

$$\text{MADE}_j = \frac{1}{200} \sum_{i=1}^{200} |\hat{g}_j(\tilde{z}_{ij}) - g_j(\tilde{z}_{ij})|, \quad \text{MADE} = \frac{1}{p} \sum_{j=1}^p \text{MADE}_j,$$

其中, $\{\tilde{z}_{ij}\}_{i=1}^{200}$ 表示区间 $[a_j, b_j]$, $j = 1, \dots, p$, 上呈等间距排列的 200 个格点, $\hat{g}_j(\tilde{z}_{ij})$ 表示函数的拟合值, $g_j(\tilde{z}_{ij})$ 表示函数的真值.

表 1 和表 2 分别给出了数据未受异常值影响以及受异常值影响时四种方法的估计结果. 从表 1 可以看出, 当数据未受异常值影响时, 在 0.5 分位点处, 均值回归估计方法 BFM、GMM 的参数估计和函数估计的结果均比分位数回归方法 BFQ、BPQ 的估计结果更具优势. 但是, 当数据受异常值污染时, 一方面, 从参数的估计结果来看, 在 0.5 分位点处, 除 α_2 的估计值外, 本文所提 BFQ 方法的参数估计值偏离真值程度最小, 标准误也相对较小, 参数估计的精度较高, 估计结果最理想, 说明当数据受异常值影响时, BFQ 方法比其它方法更不易受异常值的干扰, 性能更稳定. 同时, 分位数回归方法 BFQ 和 BPQ 的参数估计结果在均值和标准误上要优于均值回归方法 BFM 和 GMM 的估计结果, 这表明当数据受异常值干扰时分位数回归比均值回归方法对异常值的敏感度更低, 抗干扰性更强. 此外, 在 0.1、0.5 和 0.9 三个分位点处, BFQ 方法的参数估计均值和标准误总体上都优于 BPQ 方法, 说明虽然两者都是采用贝叶斯分位数回归方法处理含异常值的数据, 但是基于自由节点样条进行的估计明显比基于惩罚样条进行的估计更优越.

另一方面, 从非参数部分的估计结果来看, 在 0.5 分位点处, BFQ 方法的 MADE 值最小, GMM 方法的 MADE 值最大, 说明跟另外三种方法相比本文所提 BFQ 对非参数光滑函数的总体拟合效果最好, 而 GMM 方法的拟合效果最不理想. 与参数部分的估计结果相类似的是: 在 0.5 分位点处, 分位数回归方法 BFQ 和 BPQ 的 MADE 值要小于均值回归方法 BFM 和 GMM 的估计值; 在 0.1、0.5 和 0.9 三个分位点处, BFQ 方法的 MADE 值都小于 BPQ 方法的估计值, 说明在光滑函数的拟合中, 基于自由节点样条进行的估计依然比基于惩罚样条进行的估计更具优势. 值得关注的是, 本文所提 BFQ 方法在 0.1、0.5 和 0.9 三个不同分位点处的 MADE 值之间的差距与 BPQ 方法相比相对较小, 表明 BFQ 方法的估计结果相对来说比较稳定波动较小, 对曲线的估计和其它方法相比具有较高的稳定性.

从上述结果可以看出本文所提 BFQ 方法的估计效果最好, 原因在于: 四种方法中 BFQ 和 BPQ 属于分位数回归, BFM 和 GMM 属于均值回归, 分位数回归通过最小化加权的残差绝对值之和来估计回归参数, 而均值回归利用残差平方和最小来获得参数的估计. 与加权的残差绝对值之和相比, 残差平方和更容易受异常值干扰, 使得均值回归在出现异常值时产生更大偏差, 从而导致均值回归方法 BFM 和 GMM 的估计效果不如分位数回归方法 BFQ. 同时, BFQ 和 BPQ 虽然都属于分位数回归方法且都采用样条逼近光滑函数, 但是拟合过程中节点的个数和位置的选取存在较大差异. BPQ 方法出于简便的需要选取了大量等间距排列的点序列作为节点序列, 节点个数和位置由人为设定, 这种主观设定容易造成函数估计出现欠拟合或过拟合. BFQ 方法则将节点个数和位置视为待估参数, 采用逆跳马尔可夫链蒙特卡罗算法在维数可变的空间内对节点的个数和位置进行采样, 使节点位置和个数的选取具有空间自适应性, 从而提高了光滑函数的拟合效果, 最终使得 BFQ 方法比 BPQ 方法具有更好的模拟结果.

表 1 数据未受异常值影响时四种方法的估计结果比较

项目	估计	BFQ			BPQ			BFM	GMM
		$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.5$	$\tau = 0.5$
ρ	均值	0.268	0.448	0.847	0.169	0.278	0.850	0.519	0.548
	标准误	0.147	0.042	0.184	0.172	0.169	0.180	0.042	0.151
	置信区间	[-0.020, 0.556]	[0.366, 0.530]	[0.485, 1.209]	[-0.171, 0.509]	[-0.053, 0.609]	[0.495, 1.205]	[0.437, 0.601]	[0.249, 0.847]
α_1	均值	0.966	0.941	0.924	0.955	0.938	0.894	0.944	1.011
	标准误	0.147	0.042	0.184	0.172	0.169	0.180	0.042	0.151
	置信区间	[0.384, 1.548]	[0.666, 1.216]	[0.450, 1.398]	[0.322, 1.588]	[0.523, 1.353]	[0.359, 1.429]	[0.723, 1.165]	[0.769, 1.253]
α_2	均值	-1.025	-0.922	-0.896	-1.049	-0.926	-0.881	-0.917	-0.938
	标准误	0.232	0.219	0.228	0.317	0.265	0.253	0.105	0.114
	置信区间	[-1.478, -0.572]	[-1.351, -0.492]	[-1.343, -0.449]	[-1.673, -0.424]	[-1.448, -0.404]	[-1.380, -0.382]	[-1.122, -0.712]	[-1.164, -0.712]
$g_1(\cdot)$	MADE ₁	0.316	0.315	0.327	0.328	0.345	0.333	0.227	0.242
$g_2(\cdot)$	MADE ₂	0.302	0.281	0.305	0.359	0.324	0.308	0.234	0.194
	MADE	0.309	0.298	0.316	0.343	0.335	0.321	0.231	0.218

表 2 数据受异常值影响时四种方法的估计结果比较

项目	估计	BFQ			BPQ			BFM	GMM
		$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.5$	$\tau = 0.5$
ρ	均值	0.246	0.563	0.775	0.082	0.228	0.879	0.820	0.798
	标准误	0.099	0.059	0.066	0.059	0.155	0.061	0.085	0.117
	置信区间	[0.042, 0.434]	[0.438, 0.675]	[0.653, 0.909]	[-0.005, 0.215]	[-0.104, 0.520]	[0.748, 0.966]	[0.687, 0.953]	[0.567, 1.029]
α_1	均值	1.218	1.093	1.126	1.684	1.142	1.181	1.538	1.893
	标准误	0.280	0.177	0.262	0.394	0.218	0.303	0.286	0.288
	置信区间	[0.706, 1.804]	[0.754, 1.440]	[0.640, 1.658]	[0.960, 2.470]	[0.736, 1.567]	[0.581, 1.786]	[0.900, 1.940]	[1.270, 2.516]
α_2	均值	-1.449	-1.473	-1.459	-1.133	-1.202	-1.409	-1.574	-2.148
	标准误	0.341	0.249	0.310	0.390	0.278	0.347	0.390	0.389
	置信区间	[-2.135, -0.799]	[-1.958, -0.996]	[-2.083, -0.873]	[-1.932, -0.400]	[-1.738, -0.628]	[-2.122, -0.728]	[-2.193, -0.783]	[-2.859, -1.435]
$g_1(\cdot)$	MADE ₁	0.396	0.328	0.340	1.557	0.398	0.885	0.202	0.489
$g_2(\cdot)$	MADE ₂	0.318	0.350	0.434	1.068	0.358	0.604	0.566	0.619
	MADE	0.357	0.339	0.387	1.313	0.378	0.745	0.384	0.554

3.2 不同误差分布下四种方法的比较

在估计模型时, 采用了样条逼近法对可加函数进行近似处理, 将原模型转化为经典的线性空间自回归模型再进行估计. 线性空间自回归模型通常假设随机误差服从正态分布, 而用

样条逼近法对可加函数进行近似替代所产生的模型设定误差都将归结到随机误差项中, 这在很大程度上将导致模型误差不再服从正态分布的假设, 进而造成估计结果存在偏差. 此次模拟将对模型处于不同误差假设下四种方法的表现进行比较研究, 进一步验证本文所提 BFQ 方法的可行性以及优越性.

按如下模型生成模拟数据:

$$Y = \rho WY + X^\top \alpha + g_1(Z_1) + g_2(Z_2) + \varepsilon_\tau, \quad (16)$$

其中 $Y = (y_1, y_2, \dots, y_N)^\top$, $X^\top = (x_1, x_2, \dots, x_N)$, $x_i = (x_{i1}, x_{i2})^\top$, 且 x_{i1}, x_{i2} 均服从标准正态分布, $Z_1 = (z_{11}, z_{21}, \dots, z_{N1})^\top$, $Z_2 = (z_{12}, z_{22}, \dots, z_{N2})^\top$, 且 z_{i1}, z_{i2} 相互独立分别服从均匀分布 $U(-1, 1)$ 和 $U(0, 1)$, $i = 1, \dots, N$. 光滑函数设为 $g_1(Z_1) = \sin(\pi Z_1)$, $g_2(Z_2) = 4Z_2(1 - Z_2^2)$, 构造维度为 $N \times N$ 的 Case 空间权重矩阵 $W = I_r \otimes \Gamma_m$, $\Gamma_m = (1_m 1_m^\top - I_m)/(m - 1)$, 1_m 表示维数为 $m \times 1$, 元素全为 1 的向量, $N = r \times m$, $\otimes$ 表示 Kronecker 乘积, 相应的变量系数设为 $\rho = 0.5$, $\alpha = (1, -1)^\top$. 考虑 $N = 100$ 以及 $N = 400$ 两组样本容量, 对应的权重矩阵 W 设置为 $(r, m) = \{(5, 20), (5, 80)\}$. $\varepsilon_\tau = \varepsilon - F^{-1}(\tau)$, F 为 ε 的分布函数, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)^\top$, 此次模拟随机误差 ε_i 服从四种不同的分布, 分别是正态分布 $N(0, 1)$, 卡方分布 $\chi^2(1)$ 、学生分布 $t(1)$ 和柯西分布 $\text{Cauchy}(0, 0.5)$. 显然, 随着误差项由正态分布向卡方、学生以及柯西分布过渡, 数据集逐渐呈现偏斜和厚尾的分布态势.

在 $\tau = 0.1$ 、 $\tau = 0.5$ 、 $\tau = 0.9$ 三种分位点处, 对两组样本分别在不同的误差分布假设下运用四种方法进行估计. 采用与数值模拟 3.1 相同的设置, BFQ、BPQ 和 BFM 方法均抽样迭代 6000 次, 舍弃前 4000 次, 取后 2000 个抽样样本的均值作为参数的估计值, 且四种方法均重复 20 次.

表 3 和表 4 给出了四种误差分布下四种方法的估计模拟结果, 为了节省篇幅, 只给出了部分模拟结果. 参数的估计效果采用均方误差 (MSE) 来衡量:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\tilde{\Theta}^i - \Theta)^2,$$

其中, Θ 表示参数真值, $\tilde{\Theta}^i$ 表示第 i 次模拟的估计值, n 表示重复的次数. 光滑函数除了采用 MADE 来度量单次的估计效果, 还采用积分根均方误差 (SSE) 和积分平方偏差 (Bias) 来度量估计效果:

$$\text{SSE}(\hat{g}_j) = \left\{ \int E[\hat{g}_j(z) - g_j(z)]^2 dz \right\}^{1/2}, \quad \text{Bias}(\hat{g}_j) = \int [E\hat{g}_j(z) - g_j(z)]^2 dz.$$

其中, 期望依据样本的经验估计来计算, 积分采用区间 $[a_j, b_j]$, $j = 1, \dots, p$ 内呈等间距排列的 200 个格点上的黎曼和来近似, $\hat{g}_j(z)$ 表示函数的拟合值, $g_j(z)$ 表示函数的真值. 表 3 和表 4 中的均值和标准误取重复模拟的平均值, MADE_1 、 MADE_2 和 MADE 值取重复模拟的中位数.

首先, 在相同误差分布下对不同样本组间四种方法的估计结果进行比较. 根据表 3 和表 4, 从相同误差分布和分位点处样本容量分别为 100 和 400 的估计结果来看, 随着样本容量的

表 3 样本量 100 时学生分布 T 和柯西分布 C 下四种方法的估计结果比较

分布	项目	估计	BFQ			BPQ			BFM	GMM
			$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.5$	$\tau = 0.5$
T	ρ	均值	0.216	0.507	0.571	0.607	0.554	0.654	0.668	1.190
		标准误	0.076	0.034	0.043	0.042	0.036	0.038	0.041	1.879
		MSE_{ρ}	0.350	0.002	0.043	0.103	0.081	0.094	0.170	2.878
	α_1	均值	0.908	0.950	0.905	0.810	0.897	0.884	0.423	1.720
		标准误	0.256	0.171	0.231	0.267	0.178	0.279	0.687	2.671
		MSE_{α_1}	0.081	0.020	0.045	0.144	0.142	0.107	2.208	7.821
	α_2	均值	-0.898	-0.904	-0.959	-0.764	-0.831	-0.763	-0.829	-0.341
		标准误	0.260	0.169	0.233	0.274	0.196	0.278	0.786	2.884
		MSE_{α_2}	0.065	0.030	0.061	0.162	0.117	0.198	1.841	8.936
	$g_1(\cdot)$	$Bias_1$	0.009	0.007	0.001	0.002	0.020	0.062	0.009	4.734
		SSE_1	0.524	0.459	0.483	0.920	0.867	0.917	2.485	7.777
		$MADE_1$	0.447	0.389	0.413	0.640	0.627	0.681	0.694	3.952
	$g_2(\cdot)$	$Bias_2$	0.011	0.033	0.001	0.173	0.068	0.075	0.359	1.025
		SSE_2	0.392	0.351	0.370	0.886	0.872	0.878	1.810	7.831
		$MADE_2$	0.322	0.296	0.315	0.611	0.608	0.660	0.717	3.332
		MADE	0.386	0.351	0.369	0.627	0.609	0.678	0.701	3.767
C	ρ	均值	0.445	0.493	0.429	0.532	0.514	0.655	0.629	0.753
		标准误	0.076	0.077	0.069	0.084	0.067	0.072	0.049	0.748
		MSE_{ρ}	0.071	0.002	0.073	0.125	0.036	0.073	0.053	0.683
	α_1	均值	0.869	0.960	0.924	0.867	0.921	0.909	0.788	0.702
		标准误	0.237	0.172	0.218	0.274	0.211	0.196	0.249	1.321
		MSE_{α_1}	0.026	0.012	0.016	0.034	0.024	0.052	0.250	2.697
	α_2	均值	-0.959	-0.964	-0.975	-0.936	-0.950	-0.903	-0.814	-1.264
		标准误	0.229	0.175	0.212	0.279	0.203	0.199	0.254	1.324
		MSE_{α_2}	0.026	0.018	0.021	0.053	0.041	0.033	0.148	3.121
	$g_1(\cdot)$	$Bias_1$	0.029	0.017	0.004	0.001	0.004	0.005	0.339	0.137
		SSE_1	0.436	0.394	0.417	0.790	0.700	0.774	0.741	4.021
		$MADE_1$	0.373	0.344	0.365	0.705	0.638	0.659	0.673	1.312
	$g_2(\cdot)$	$Bias_2$	0.037	0.008	0.032	0.175	0.011	0.038	0.020	0.924
		SSE_2	0.315	0.292	0.312	0.597	0.572	0.598	0.709	3.674
		$MADE_2$	0.258	0.243	0.257	0.481	0.484	0.486	0.497	1.536
		MADE	0.324	0.290	0.316	0.593	0.584	0.589	0.592	1.465

增大,一方面,四种方法参数部分估计的 MSE 值总体上都在减小,参数估计值也越来越接近真值,估计的精度也越来越高,说明参数部分的估计是收敛的;另一方面,非参数部分估计的 $Bias_1$ 、 $Bias_2$ 、 SSE_1 、 SSE_2 和 $MADE$ 值也在变小,说明随着样本容量的增大光滑函数的拟合程度也越来越高.总的来看,随着样本容量的增大,四种方法的估计结果都呈收敛态势,满足一致性.

其次,在同一组样本容量下对不同误差分布间四种方法的估计结果进行比较.由表 3 和

表 4 样本量 400 时正态分布 N 和卡方分布 χ^2 下四种方法的估计结果比较

分布	项目	估计	BFQ			BPQ			BFM	GMM
			$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.1$	$\tau = 0.5$	$\tau = 0.9$	$\tau = 0.5$	$\tau = 0.5$
N	ρ	均值	0.297	0.386	0.306	0.240	0.276	0.128	0.415	0.469
		标准误	0.170	0.162	0.156	0.176	0.226	0.152	0.115	0.168
		MSE_ρ	0.228	0.032	0.219	0.275	0.098	0.220	0.041	0.031
	α_1	均值	1.039	0.985	1.034	1.040	0.984	1.020	1.001	0.991
		标准误	0.127	0.093	0.128	0.140	0.107	0.146	0.048	0.051
		MSE_{α_1}	0.006	0.004	0.004	0.007	0.006	0.005	0.001	0.004
	α_2	均值	-0.963	-0.978	-0.985	-1.007	-1.002	-0.969	-0.997	-0.989
		标准误	0.133	0.095	0.132	0.149	0.108	0.144	0.047	0.051
		MSE_{α_2}	0.009	0.003	0.004	0.006	0.004	0.005	0.002	0.001
	$g_1(\cdot)$	$Bias_1$	0.000	0.013	0.017	0.001	0.000	0.000	0.000	0.001
		SSE_1	0.251	0.236	0.242	0.702	0.624	0.753	0.161	0.132
		$MADE_1$	0.219	0.203	0.219	0.634	0.616	0.652	0.136	0.099
	$g_2(\cdot)$	$Bias_2$	0.001	0.004	0.011	0.009	0.014	0.089	0.040	0.004
		SSE_2	0.175	0.144	0.165	0.533	0.494	0.529	0.121	0.125
		$MADE_2$	0.145	0.120	0.134	0.462	0.414	0.420	0.091	0.082
		MADE	0.182	0.158	0.174	0.554	0.539	0.567	0.110	0.101
χ^2	ρ	均值	0.352	0.648	0.366	0.764	0.746	0.776	0.844	0.728
		标准误	0.070	0.036	0.013	0.094	0.075	0.009	0.013	0.907
		MSE_ρ	0.023	0.041	0.117	0.083	0.064	0.127	0.138	1.358
	α_1	均值	1.014	0.993	1.027	0.971	0.979	1.049	1.060	1.033
		标准误	0.112	0.029	0.146	0.161	0.103	0.144	0.036	0.216
		MSE_{α_1}	0.001	0.001	0.001	0.002	0.002	0.006	0.003	0.021
	α_2	均值	-0.980	-0.991	-1.016	-0.965	-0.981	-1.026	-1.040	-0.975
		标准误	0.115	0.080	0.152	0.168	0.104	0.139	0.037	0.286
		MSE_{α_2}	0.001	0.003	0.004	0.006	0.006	0.008	0.004	0.019
	$g_1(\cdot)$	$Bias_1$	0.006	0.006	0.006	0.003	0.001	0.001	0.000	0.013
		SSE_1	0.212	0.192	0.325	0.730	0.689	0.719	0.207	0.208
		$MADE_1$	0.184	0.159	0.290	0.623	0.611	0.635	0.175	0.160
	$g_2(\cdot)$	$Bias_2$	0.004	0.002	0.005	0.003	0.011	0.016	0.003	0.019
		SSE_2	0.143	0.132	0.163	0.484	0.462	0.520	0.133	0.207
		$MADE_2$	0.120	0.094	0.132	0.422	0.405	0.424	0.102	0.155
		MADE	0.151	0.135	0.205	0.523	0.508	0.558	0.141	0.152

表 4 可知, 不管是 100 还是 400 的样本, 在 0.5 分位点处, 当误差服从标准正态分布时, 从参数部分的均值、标准误和 MSE 值以及非参数部分的 $Bias_1$ 、 $Bias_2$ 、 SSE_1 、 SSE_2 和 MADE 值, 均值回归方法 BFM 和 GMM 的估计结果总体上要优于分位数回归方法 BFQ 和 BPQ 的估计结果, 表明当误差分布满足正态分布这一条件时, 均值回归方法比分位数回归方法更有效. 然而, 当误差分布逐渐向厚尾的学生和柯西分布转变时, 无论是参数估计还是非参数拟合, 分位数回归方法 BFQ 和 BPQ 的估计结果均超越了均值回归方法 BFM 和 GMM, 说明

分位数回归方法比均值回归方法更适合于处理尖峰厚尾的数据. 特别的, 当误差服从厚尾的学生、柯西以及偏态的卡方这三种非正态分布时, BFQ 方法估计出的参数和非参数部分评价指标整体上都小于其它方法, 这进一步说明四种方法中 BFQ 方法不仅在处理尖峰厚尾数据时具有优越性, 而且在处理出现偏斜态势的数据时也有其它方法无法比拟的优势. 同时, 在 0.1、0.5 和 0.9 三个分位点处, 不管误差服从何种分布, BFQ 方法的参数和非参数部分的估计结果都比 BPQ 方法的表现更加突出, 说明基于自由节点样条进行贝叶斯分位数回归估计总是比基于惩罚样条进行的估计效果更好.

值得注意的是, 在同一组样本容量下, 在 0.5 分位点处, 不论是何种样本, 总体上来看均值回归方法 BFM 和 GMM 估计出的均值, 标准误, MSE、Bias₁、Bias₂、SSE₁、SSE₂ 和 MADE 值在不同误差分布之间的波动较大, 而分位数回归方法 BFQ 和 BPQ 的估计结果在不同误差分布间的差距很小, 表明误差的非正态分布对均值回归造成了很大干扰, 这也是均值回归为保证估计的有效性通常设立“误差服从正态分布”这一苛刻条件的原因之一. 相比较而言, 分位数回归估计结果中 MADE 等指标的值在各种误差分布间趋于稳定, 这意味着不管误差项的实际分布为何种分布, 都能运用分位数回归方法进行估计, 总体的估计结果基本保持一致, 这与贝叶斯分位数回归的奠基人 Yu and Moyeed (2001) 创立贝叶斯分位数回归方法时, 无需对误差项做任何特定分布假设的宗旨不谋而合. 因此, 本文所提 BFQ 方法对误差分布极其宽松的条件设置为部分线性可加空间滞后模型提供了一个更加普适和有效的估计方法.

4 省域碳排放影响因素的实证研究

当前, 中国兑现碳减排承诺面临巨大压力, 如何充分挖掘省域经济发展的碳减排潜力成为学界研究的热点问题. 从已有的文献来看, 关于碳排放的问题大部分采用均值回归的方法进行研究. 省域间碳排放的巨大差距使得碳排放数据容易偏离正态分布, 以往的均值回归方法并不能很好地适用于这种非正态的碳排放数据, 从而使得均值回归估计的结果出现较大偏误. 基于上述考虑, 本文利用所提出的方法对省域碳排放的各类线性和非线性影响因素进行分析, 不仅能充分利用碳排放数据的尾部信息, 而且能得到比均值回归更稳健的实证结果, 以此说明本方法为实际问题的研究提供了一个更有力的分析工具.

4.1 被解释变量的选取

选取碳排放量的自然对数作为被解释变量, 文中所指的碳排放量是指燃烧原煤、洗精煤、焦炭、焦炉煤气、原油、汽油、煤油、柴油、燃料油、液化石油气、炼厂干气、天然气、其他洗煤等 13 种一次能源所排放的二氧化碳量. 由于统计年鉴中无二氧化碳排放量数据, 为更准确估计其排放量, 根据《2006 年 IPCC 国家温室气体排放清单指南》提供的能源消耗碳排放测度方法测算中国各省域的碳排放量:

$$CO_2 = \sum_{i=1}^n EC_i \times NCV_i \times CEF_i \times COF_i \times \frac{44}{12}.$$

其中 EC_i 为第 i 种能源消耗量, 用能源平衡表中各能源总消费量来表示, 单位: 万吨; NCV_i 为第 i 种能源的平均低位发热值, 单位: kJ/kg; CEF_i 为第 i 种能源的单位平均低位发热值

的碳排放系数, 单位: kg/GJ; COF_i 为第 i 种能源的碳氧化因子; 44 和 12 分别表示二氧化碳分子和碳元素的摩尔量。

4.2 解释变量的选取

参考国内外研究成果, 结合省域碳排放数据的特点, 从线性和非线性两方面的因素入手, 选取以下五个因素作为解释变量:

能源消费结构——能源消费总量中煤炭消费量越高, 碳排放总量越大, 因此能源消费结构是影响碳排放的重要因素之一。为了分析能源消费结构对省域碳排放的影响, 以煤炭消费量占能源消费总量的比重来衡量能源消费结构, 预期其回归系数符号为正。

产业结构——产业结构中以工业为主的第二产业的占比越大, 越对碳排放有直接的正向影响。因此, 采用第二产业增加值占地区国内生成总值的比重来衡量产业结构, 并预期其回归系数符号为正。

技术进步水平——采用科技经费支出占财政支出的比重来衡量技术进步水平。如果技术进步以环保需求为导向具有“绿色偏向”, 则有利于节能减排, 但如果技术进步以扩大生成规模促进经济增长为导向, 则对碳排放的抑制作用并不显著 (邵帅等, 2019)。因此, 技术进步水平的回归系数符号预期无法确定。

环境规制——碳排放是生产及消费过程中的一种外部性行为, 市场机制对该行为的调控是失灵的, 需要政府制订一系列环境政策来限制。因此, 环境规制也显著影响着碳排放, 且这种影响存在“绿色悖论”和“倒逼减排”两种情形, 甚至环境规制与碳排放有可能呈现倒“U”型的曲线关系 (张华和魏晓平, 2014; 李菁等, 2021), 还有可能呈现倒“N”型曲线关系 (李治国等 (2022))。本文采用现行有效的地方性环保法规和规章总数来度量环境规制, 且预期环境规制与碳排放之间存在复杂的非线性关系。

城镇化水平——城镇化进程的加快, 一方面会扩大城市规模增加能源消费需求, 引致碳排放增加; 另一方面当城镇化达到较高水平时, 环保技术进步和能源效率的提高有利于节能减排, 城镇化与碳排放最终呈现倒“U型”曲线关系 (邵帅等 (2019))。随着碳排放水平的增大, 城镇化也可能呈现抑制作用, 表现出强制减排效应, 两者关系曲线逐渐呈现倒“V型”特征 (邵帅等, 2022)。本文采用城镇人口占各省年末常住人口的比重来衡量城镇化水平, 且预期城镇化水平与碳排放之间呈现非线性特征。

4.3 实证分析

1) 样本的选取

本文采用中国 2017 年 30 个省市区 (港澳台地区和西藏自治区因数据缺失而被剔除) 的数据作为样本进行实证研究。一次能源消费数据来源于《中国能源统计年鉴》, 各地区第二产业增加值、地区国内生成总值、科技经费支出、财政支出、城镇人口和各省年末常住人口数据来源于国家统计局, 现行有效的地方性环保法规和规章总数来源于《中国环境年鉴》等。

2) 空间相关性的检验

资源禀赋和基础设施条件的相似性, 使得邻省间的经济交流日益密切, 这可能导致省域碳排放存在空间关联。为了检验空间相关性, 根据省域空间是否相邻构造文献中常用的二值空间权重矩阵, 即 Rook 相邻权重矩阵 $W = (w_{ij})_{n \times n}$, 其中 w_{ij} 是权重矩阵 W 的元素, 当两

省地理相邻时 $w_{ij} = 1$, 不相邻时 $w_{ij} = 0$, 且权重矩阵用于参数估计时均进行行标准化处理. 在估算出省域碳排放量的基础上, 应用 stata 软件计算 2017 年省域碳排放的全局 Moran's I 指数. 结果显示, 全局 Moran's I 指数为 0.246 大于 0, 且在 0.01 水平下显著, 表明邻近省域的碳排放出现一定程度的集聚效应, 呈现显著的正向空间关联性.

3) 模型的选择

从已有的文献来看, 环境规制和城镇化水平都与碳排放之间存在明显的非线性关系, 若直接采用简单的线性模型对碳排放的影响因素进行分析将产生较大偏差. 同时, 基于碳排放具有空间溢出效应的上述论证, 并考虑到能源消费结构、技术进步水平和产业结构均与碳排放存在一定的线性关系, 故采用部分线性可加空间滞后模型对数据进行分析, 以便最大限度满足模型的前提假设.

此外, 经计算可知, 被解释变量的密度曲线偏度为 2.210, 峰度为 9.417, 呈明显的右偏分布, 且峰度值远大于正态分布的峰度 3, 具有鲜明的尖峰厚尾特征. 被解释变量的尖峰厚尾和偏态分布与常见的均值回归模型的正态分布假设相去甚远, 因此应考虑采用对误差分布更稳健的分位数回归方法进行建模分析.

建立关于省域碳排放的部分线性可加空间滞后模型:

$$Y = \rho WY + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + g_1(Z_1) + g_2(Z_2) + \varepsilon, \quad (17)$$

其中 Y , X_1 , X_2 , X_3 , Z_1 , Z_2 分别为省域碳排放的对数、产业结构、技术进步水平、能源消费结构、环境规制和城镇化水平; ρ , α_1 , α_2 , α_3 为参数部分待估计的回归系数, $g_1(\cdot)$, $g_2(\cdot)$ 为非参数部分待拟合的光滑函数. 为了做进一步的比较研究, 采用与数值模拟 3.1 相同的先验分布超参数和设置, 在 0.4、0.5 和 0.6 分位点处运用 BFQ、BPQ、BFM 和 GMM 四种方法对模型 (17) 进行估计, 其中, 三种贝叶斯估计法均抽样迭代 10000 次, 舍弃前 4000 次, 取后 6000 个观测值的均值作为参数的估计值.

四种方法的参数部分估计结果 (表 5) 显示, 首先, 在 0.5 分位点处, 各省域间碳排放存在显著的正向空间依赖性, 表明各省域间的碳排放存在明显的相互模仿和策略性竞争效应, 该结论也进一步验证了 Moran's I 指数的检验结果; 其次, 产业结构、技术进步水平和能源消费结构对省域碳排放均呈现出正向的驱动作用, 说明 2017 年中国整体的技术进步水平仍是以提高生产率为导向的, 节能减排技术开发和应用的时滞性使得技术进步对碳排放没有发挥出应有的抑制效应, 而产业结构和能源消费结构对省域碳排放的影响则符合预期; 第三, 虽然三个影响因素的参数估计值并不相等, 但从影响程度来看, 四种方法的估计结果均表明能源消费结构对碳排放的影响最大, 产业结构次之, 技术进步水平影响最小, 可见煤炭占比过高的能源消费结构才是阻碍碳减排的关键因素.

综上所述, 在被解释变量呈现尖峰厚尾和偏态分布的基础上, 一方面考虑到四种方法估计结果经济意义的总体一致性, 另一方面考虑到分位数回归方法对误差分布和全局估计的稳健性以及自由节点样条比惩罚样条估计的优越性, 所以, 此处选用结果更为稳健和出色的 BFQ 方法对实证数据进行研究, 即构建省域碳排放有关的部分线性可加空间滞后分位数回归模型, 并采用基于自由节点样条的贝叶斯方法进行估计.

表 5 省域碳排放影响因素模型中四种方法的参数估计结果

项目	估计	BFQ			BPQ			BFM	GMM
		$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.4$	$\tau = 0.5$	$\tau = 0.6$	$\tau = 0.5$	$\tau = 0.5$
ρ	均值	0.870	0.886	0.899	0.813	0.846	0.864	0.899	0.660
	标准误	0.058	0.063	0.062	0.112	0.113	0.104	0.056	0.255
α_1	均值	0.334	0.378	0.454	0.366	0.401	0.481	0.421	0.674
	标准误	0.163	0.161	0.171	0.247	0.246	0.267	0.896	0.899
α_2	均值	0.256	0.272	0.373	0.108	0.089	0.047	0.106	0.187
	标准误	0.334	0.328	0.334	0.397	0.412	0.413	0.358	0.542
α_3	均值	0.545	0.597	0.605	0.501	0.531	0.614	1.016	2.186
	标准误	0.091	0.093	0.097	0.115	0.111	0.119	0.598	0.970

5 结论

研究采用贝叶斯方法对部分线性可加空间滞后分位数回归模型进行了全面分析, 其中, 运用了自由节点样条来拟合非参数函数, 并将节点个数和位置作为待估参数借助逆跳马尔可夫链蒙特卡罗算法对其进行了贝叶斯估计, 形成了基于自由节点样条的贝叶斯分位数回归方法 (BFQ). 为验证 BFQ 方法的有效性, 将其与 BPQ、BFM 和 GMM 方法进行了模拟比较. 结果显示: 一方面, 当数据受异常值影响时, BFQ 方法与其它三种方法相比更不易受异常值的干扰, 性能更稳定; 另一方面, 当误差分布呈尖峰厚尾和偏斜态势时, BFQ 方法的估计效果也要优于其它三种方法. 最后, 以省域碳排放量为研究对象, 对影响其的线性和非线性因素进行实证研究, 来进一步检验 BFQ 方法在实际问题中对参数和非参数部分的估计能力. 实证结果表明, BFQ 方法适用于处理具有尖峰厚尾和偏态特征的省域碳排放数据, 刻画出了各因素对省域碳排放的线性和非线性影响, 达到了预期的估计效果.

参 考 文 献

- 方丽婷, 李坤明, (2019). 空间滞后分位数回归模型的贝叶斯估计 [J]. 数量经济技术经济研究, 36(9): 102–116.
- Fang L T, Li K M, (2019). Bayesian Estimation of Spatial Lag Quantile Regression Model[J]. Journal of Quantitative & Technological Economics, 36(9): 102–116.
- 胡亚南, 王金天, 田茂再, (2022). 高维部分线性可加的空间分位回归模型 [J]. 数理统计与管理, 41(5): 843–857.
- Hu Y N, Wang J T, Tian M Z, (2022). High-dimensional Partially Linear Additive Spatial Quantile Regressive Model[J]. Journal of Applied Statistics and Management, 41(5): 843–857.
- 李菁, 李小平, 郝良峰, (2021). 技术创新约束下双重环境规制对碳排放强度的影响 [J]. 中国人口·资源与环境, 31(9): 34–44.
- Li J, Li X P, Hao L F, (2021). Impact of Dual Environmental Regulations on Carbon Emission Intensity under the Constraint of Technological Innovation[J]. China Population, Resources and Environment, 31(9): 34–44.
- 李坤明, 方丽婷, (2018). 空间滞后分位数回归模型的工具变量估计及参数检验 [J]. 统计研究, 35(10): 103–

- 115.
- Li K M, Fang L T, (2018). Instrumental Variables Estimation and Parametric Tests of Spatial Lag Quantile Regression Model[J]. Statistical Research, 35(10): 103–115.
- 李龙飞, (2021). 空间计量经济学中的空间自回归模型 [J]. 计量经济学报, 1(1): 36–65.
- Lee L F, (2021). The Spatial Autoregression Model in Spatial Econometrics[J]. China Journal of Econometrics, 1(1): 36–65.
- 李治国, 杨雅涵, 赵园春, (2022). 地方政府竞争促进了地区碳排放强度吗?[J]. 经济与管理评论, 38(2): 136–146.
- Li Z G, Yang Y H, Zhao Y C, (2022). Does Local Government Promote Regional Carbon Emission Intensity?[J]. Review of Economy and Management, 38(2): 136–146.
- 罗国旺, 吴密霞, (2020). 部分线性可加空间自回归模型在约束条件下的统计推断 [J]. 数理统计与管理, 39(6): 1000–1009.
- Luo G W, Wu M X, (2020). Statistical Inference for Partially Linear Additive Spatial Autoregressive Models under Restricted Conditions[J]. Journal of Applied Statistics and Management, 39(6): 1000–1009.
- 邵帅, 范美婷, 杨莉莉, (2022). 经济结构调整、绿色技术进步与中国低碳转型发展——基于总体技术前沿和空间溢出效应视角的经验考察 [J]. 管理世界, 38(2): 46–69.
- Shao S, Fan M T, Yang L L, (2022). Economic Restructuring, Green Technical Progress, and Low-carbon Transition Development in China: An Empirical Investigation Based on the Overall Technology Frontier and Spatial Spillover Effect[J]. Journal of Management World, 38(2): 46–69.
- 邵帅, 张可, 豆建民, (2019). 经济集聚的节能减排效应: 理论与中国经验 [J]. 管理世界, 35(1): 36–60.
- Shao S, Zhang K, Dou J M, (2019). Effects of Economic Agglomeration on Energy Saving and Emission Reduction: Theory and Empirical Evidence from China[J]. Journal of Management World, 35(1): 36–60.
- 谢琨, 刘磊, 曹瑞元, (2018). 一种新的空间计量模型——部分线性可加自回归模型及其应用 [J]. 数理统计与管理, 37(2): 235–242.
- Xie L, Liu L, Cao R Y, (2018). A New Type of Spatial Econometric Model — Partial Linear Additive Spatial Autoregressive Model and Its Application[J]. Journal of Applied Statistics and Management, 37(2): 235–242.
- 张华, 魏晓平, (2014). 绿色悖论抑或倒逼减排——环境规制对碳排放影响的双重效应 [J]. 中国人口·资源与环境, 24(9): 21–29.
- Zhang H, Wei X P, (2014). Green Paradox or Forced Emission-reduction: Dual Effect of Environmental Regulation on Carbon Emissions[J]. China Population, Resources and Environment, 24(9): 21–29.
- Chen Z Y, Wang H B, (2018). Latent Single-index Models for Ordinal Data[J]. Statistics and Computing, 28(3): 699–711.
- Chen Z, Chen J, (2021). Bayesian Analysis of Partially Linear Additive Spatial Autoregressive Models with Free-knot Splines[J]. Symmetry, 13(9): 1635.
- Chen Z, Chen M, Xing G, (2021). Bayesian Estimation of Partially Linear Additive Spatial Autoregressive Models with P-Splines[J]. Mathematical Problems in Engineering, 2021: 1777469.
- Du J, Sun X, Cao R, et al, (2018). Statistical Inference for Partially Linear Additive Spatial Autoregressive Models[J]. Spatial Statistics, 25: 52–67.
- Eilers P H C, Marx B D, (1996). Flexible Smoothing with B-splines and Penalties[J]. Statistical Science, 11(2): 89–121.

- Green P J, (1995). Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination[J]. *Biometrika*, 82(4): 711–732.
- Jin F, Lee L F, (2019). GEL Estimation and Tests of Spatial Autoregressive Models[J]. *Journal of Econometrics*, 208(2): 585–612.
- Kelejian H H, Prucha I R, (1999). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model[J]. *International Economic Review*, 40(2): 509–533.
- Kozumi H, Kobayashi G, (2011). Gibbs Sampling Methods for Bayesian Quantile Regression[J]. *Journal of Statistical Computation and Simulation*, 81(11): 1565–1578.
- Krisztin T, (2018). Semi-parametric Spatial Autoregressive Models in Freight Generation Modeling[J]. *Transportation Research Part E: Logistics and Transportation Review*, 114: 121–143.
- Lee L F, (2004). Asymptotic Distributions of Quasi-maximum Likelihood Estimators for Spatial Autoregressive Models[J]. *Econometrica*, 72(6): 1899–1925.
- Lee L, Liu X, (2010). Efficient GMM Estimation of High Order Spatial Autoregressive Models with Autoregressive Disturbances[J]. *Econometric Theory*, 26(1): 187–230.
- Li T, Kang X, (2022). Variable Selection of Higher-order Partially Linear Spatial Autoregressive Model with a Diverging Number of Parameters[J]. *Statistical Papers*, 63(1): 243–285.
- Liu T, Xu X, Lee L, (2022). Consistency without Compactness of the Parameter Space in Spatial Econometrics[J]. *Economics Letters*, 210: 110224.
- Panagiotelis A, Smith M, (2008). Bayesian Identification, Selection and Estimation of Semiparametric Functions in High-dimensional Additive Models[J]. *Journal of Econometrics*, 143(2): 291–316.
- Poon W Y, Wang H B, (2013). Bayesian Analysis of Generalized Partially Linear Single-index Models[J]. *Computational Statistics & Data Analysis*, 68: 251–261.
- Yu K, Moyeed R A, (2001). Bayesian Quantile Regression[J]. *Statistics & Probability Letters*, 54(4): 437–447.