

doi: 10.16104/j.issn.1673-1891.2024.03.009

约束条件下线性可加空间自回归模型方法研究

刘苏兵^{1,2}

(1. 安徽工程大学数理与金融学院, 安徽 芜湖 241000; 2. 安徽机电职业技术学院公共基础教学部, 安徽 芜湖 241002)

摘要: [目的] 针对空间依赖性问题, 提出约束条件下线性可加空间自回归模型的统计推断方法。[方法] 先对线性可加空间自回归模型增加线性约束条件。通过 B 样条基函数近似求解未知的平滑函数, 并代入回归模型, 生成包含 B 样条基函数的空间自回归模型, 对此模型的损失函数进行最小化, 获取约束条件下模型的稀疏估计, 通过 SCAD 惩罚函数无偏估计较大系数。在 10 种约束条件下, 研究了线性可加空间自回归模型估计的渐近性质, 得到在最优收敛速度下进行斜率系数的估计时, 函数项部分不会改变参数估计的渐近分布特性。[结果] 实验结果表明, 该方法不仅在理论上具有可行性, 而且在实践中易于操作。相比非约束估计, 约束稀疏估计更为优秀。[结论] 在交通数据统计分析中, 该方法具有实用性和有效性。

关键词: 约束条件; 线性可加; 空间自回归模型; 统计推断方法; 稀疏估计; 渐近性质

中图分类号: O212.1 **文献标志码:** A **文章编号:** 1673-1891(2024)03-0069-09

Research on Linear Additive Spatial Autoregressive Model Method Under Constraints

LIU Subing^{1,2}

(1. School of Mathematics-Physics and Finance, Anhui University of Technology, Wuhu 241000, Anhui, China; 2. Public Basic Teaching Department, Anhui Mechanical and Electrical Vocational Technical College, Wuhu 241002, Anhui, China)

Abstract: [Objective] To propose a statistical inference method for linear additive spatial autoregressive models under constraint conditions to address spatial dependency issues. [Method] First, linear constraints are added to the linear additive spatial autoregressive model. By approximating the unknown smoothing function using B-spline basis functions and substituting it into the regression model, a spatial autoregressive model containing B-spline basis functions is generated. The loss function of this model is minimized to obtain a sparse estimate of the model under constraint conditions. The SCAD penalty function is used to estimate the larger coefficients without bias. The asymptotic properties of linear additive space autoregressive model estimation were studied under 10 constraint conditions, and it was found that when estimating the slope coefficient at the optimal convergence speed, the asymptotic distribution characteristics of the parameter estimation were not affected by the function term. [Result] The experimental results show that this method is not only feasible in theory, but also easy to operate in practice. Compared to unconstrained estimation, constrained sparse estimation is

收稿日期: 2024-01-11

作者简介: 刘苏兵(1987—), 男, 安徽滁州人, 讲师, 硕士, 研究方向: 应用数学、经济统计, e-mail: 1446780008@qq.com。

superior. [Conclusion] This method is practical and effective in traffic data statistical analysis.

Keywords: constraint condition; linearly additive; spatial autoregressive model; statistical inference method; sparse estimation; asymptotic property

0 引言

统计推断是统计学中的一种重要方法,它是基于对样本数据的分析来推测总体的性质和特征,在许多领域中,由于受限于成本、时间或可行性等因素^[1],无法直接对总体进行全面调查。此时,通过样本数据来进行统计推断,可以帮助人们降低这种不确定性,了解总体的特征和规律。基于历史数据和当前状况的统计推断,可以帮助人们预测未来的趋势和可能性。例如,通过分析过去几年的销售数据,可以预测未来的销售情况,从而制定相应的策略^[2];通过统计推断,可以发现存在的问题和不足之处,从而进行改进和完善;通过分析产品质量检测数据,可以发现产品的缺陷,进而改进生产工艺或流程^[3]。因此,统计推断作为一种重要的方法论,在各个领域中都发挥着重要的作用。

统计推断的方法有很多,比如 Chen 等^[4]提出的 p 样条部分线性加性空间自回归模型的贝叶斯统计推断方法。此方法可以全面评估协变量对空间数据响应的线性与非线性影响。利用贝叶斯 p 样条,结合马尔可夫链蒙特卡罗法,开发出贝叶斯估计法 PLASARM。同时,设计 Gibbs 采样器来探索未知参数的联合后验分布,并针对实际问题进行精确地统计推断。这种方法需要知道先验概率,且先验概率往往取决于假设,而假设的模型有很多种,这就可能导致推断效果的不佳。Lustosa 等^[5]提出不同模型数据规模下的机制统计推断方法,将统计原理融入数据处理中,不仅可以简化操作和优化方案,还可以简化推断过程。通过强调数据建模、数据集合设计、可视化以及细致的数据重组,有助于提高数据统计的准确性

和数据选择的合理性,进而对经济预测等实际问题进行准确的统计推断。此方法对数据质量要求高,如果数据存在异常值、缺失值等问题,会影响统计推断的准确性。张洁等^[6]提出一种采用自激励模型进行统计推断,此方法基于最小二乘法 (conditional least squares, CLS) 和连续机器学习 (continuous machine learning, CML) 方法的参数估计,有效解决了整值时间序列分段现象的问题,并对预测问题进行了深入研究。为了更精确地描述时间序列特性,该方法扩大了 NBTINAR 模型的要求。针对整值门限自回归过程,研究其基础属性与非参数估计课题,获取参数的精确估计以及区间估计,这也为实际统计问题提供了推断依据。但此方法是基于 CLS 和 CML 方法的参数估计,在实际应用中,模型选择和参数估计可能存在主观性和不确定性。赵志文等^[7]提出周期随机系数自回归模型的方法进行统计推断,此方法研究多元随机系数自回归模型,并采用非参方法进行参数统计推断;同时,考虑到实际数据中的协变量影响,提出了带有协变量的随机系数自回归模型,并讨论了模型的参数估计等问题;此外,考虑到协变量数据的测量误差,提出了相应的模型,并利用经验似然方法对多元一阶随机系数自回归模型的参数进行估计和假设检验,同时检验自回归系数是否恒为常数。这种方法为解决实际问题提供了有效的统计推断方法,但此方法计算复杂度较高,经验似然方法涉及复杂的统计计算,对于大规模数据或复杂模型可能会变得计算量大且耗时。

约束条件下线性可加空间自回归模型的统计推断方法推动了统计推断领域的发展,为解决实际问题提供了新的思路和方向,可应用于各种

实际问题,如经济预测、城市规划等,为决策者提供有力支持。约束条件有助于模型更好地拟合数据,从而提高预测的准确性,也可以降低模型的复杂度,使模型更易于理解和应用。通过对约束条件的设定和应用,可以更好地洞察数据的内在结构和规律,从而作出更有效的决策。因此本文提出约束条件下线性可加空间自回归模型的统计推断方法。

1 约束条件下线性可加空间自回归模型

1.1 构建线性可加空间自回归模型

本次用 $\{Y_i, X_i, Z_i\}_{i=1}^n$ 描述观测获得的高维空间数据, X_i, Y_i, Z_i, n 分别用于描述 a 维的协变量、响应变量、 b 维的协变量、空间单元的个数,且 X_i 和 Y_i 呈线性关系, Z_i 和 Y_i 呈不确定形态的非线性关系。 n 个空间单元的相邻关系通过空间权重矩阵 W 描述,由此考虑部分线性可加的空间自回归模型表述为式(1)^[8]:

$$Y = \delta WY + X\sigma + \sum_{j=1}^b p_j(Z_j) + e \quad (1)$$

式中: $\delta, p_j(\cdot), \sigma$ 分别用于描述空间自回归系数、未知的平滑函数、 $\alpha \times 1$ 维的系数向量。设定在 $1 < j < b$ 时, $F[p_j(Z_j)] = 0$ 成立,来确保非参数部分的可辨识性。 e 用于描述随机扰动项,符合 $F(e|X, Z) = 0$,并且是有边界的方差。上述模型中的空间滞后项反映空间上的相互依赖关系。参数部分明确描述协变量和响应变量之间的线性关系,系数意义清晰。非参数部分则灵活处理协变量和响应变量间不确定的具体关系形态^[9-10]。考虑到空间数据的异质性,不对扰动项进行正态性或同方差的假设。

在考虑部分线性可加的空间自回归模型中,其参数都是在某些约束限制下进行选取^[11],本文选取线性约束作为模型的约束方式,具体用式

(2)表述。

$$S\sigma = l \quad (2)$$

式中: S 和 l 分别用于描述行向量线性无关的矩阵 $m \times n$ 以及 $m \times 1$ 维的向量。通过线性约束,加快参数估计速度^[12-13]。

1.2 通过B样条函数近似未知的平滑函数

考虑部分线性可加的空间自回归模型中,未知的平滑函数通过组合的B样条基函数来近似求解。下面详细描述,设定 c_j 与 d_j 是正整数, $g_j = c_j + d_j$,把 $0 \sim 1$ 区间分成 $d_j + 1$ 个子区间,用 $h_j = 0 < u_{j,1} < \dots < u_{j,d_j} < 1 = h_j$ 描述子区间的内部节点。由 c_j 阶和 d_j 个节点数构成的样条函数空间用 α_j 描述^[14]。在B样条函数空间 α_j 中选择一组基函数,用 $\{B_{j,m}(\cdot)\}_{m=1}^{g_j}$ 描述,未知的平滑函数通过样条函数来替代,由(3)表达。

$$p_j(\cdot) = \sum_{m=1}^{g_j} B_{j,m}(\cdot) \chi_{j,m}, \quad j = 1, 2, \dots, b \quad (3)$$

式中: $\chi = (\chi_{1,1}, \dots, \chi_{1,g_1}, \dots, \chi_{b,1}, \dots, \chi_{b,g_b})^T$ 。

令 $B(z_{i,t})$ 为:

$$B(z_{i,t}) = [B_{1,1}(z_{i,t,1}), \dots, B_{1,g_1}(z_{i,t,1}), \dots, B_{b,1}(z_{i,t,b}), \dots, B_{b,g_b}(z_{i,t,b})]^T \quad (4)$$

将式(3)代入式(1),得到(5)式:

$$Y = \delta \tilde{W}Y + X\sigma + B(Z)\chi + Dn(\mu) + \lambda \quad (5)$$

式中: $Y = (Y_1^T, \dots, Y_N^T)^T, Y = (y_{i,1}, \dots, y_{i,T})^T$ 。

$$\tilde{w} = w \otimes J_T,$$

J_T 用于描述阶数是 T 的单位矩阵。

$$B(Z) = \left\{ [B(Z_1)]^T, \dots, [B(Z_N)]^T \right\}^T,$$

$$B(Z_i) = \left\{ B(Z_{i,1}), \dots, B(Z_{i,1}) \right\}^T,$$

$$D = [-1_{N-1}, J_{N-1}]^T \otimes 1_T,$$

1_T 和 \otimes 分别用于描述维数为 T 且元素都是1的向量以及Kronecker积。 $\eta = (\eta_2, \dots, \eta_N)^T, e_i = (e_{i,1}, \dots, e_{i,T})^T$ 。

通过以上过程可以进一步增强模型对空间

数据的拟合能力,捕捉更复杂的空间模式和关系。

1.3 线性可加空间自回归模型稀疏估计实现

在分位回归框架下,对式(5)进行稀疏估计。为简化空间回归模型的复杂性,采用特征向量线性组合对空间滞后项近似表达,并通过B样条基函数近似未知的平滑函数。然而,这种近似可能带来模型设定误差,将其纳入随机扰动项中。由于随机扰动项可能不满足正态性以及同方差性,难以确保均值回归估计量的优良性质。因此,采用分位回归,其稳健性更佳。高维数据会展现出异质性,协变量对响应变量的影响随位置变化。分位回归能全面揭示这种关系,并合理解释协变量效应的稀疏性。

1.3.1 目标函数

当分位水平相异时,式(1)的非参数可加部分中,协变量对响应变量的影响会有所不同,这可以通过观察 $P_j(Z_j)$ 值是否为零来体现。在式(5)中,选择合适的基函数组合,以便筛选出非参数部分的关键协变量。特定的样条信息定义各协变量,并通过样条基函数进行展开。因此,对变量的筛选本质上就是对样条信息的筛选。将惩罚的目标函数最小化处理,来获取稀疏估计,用式(6)表达。

$$Lb(\chi, \sigma, \eta) = \frac{1}{n} \sum_{i=1}^n \delta_\tau[Y_i - F_i\chi - X_i\sigma - B(U_i)^T\eta] + \sum_{j=1}^g b_\mu(\|\eta_j\|_2) \tag{6}$$

式中: $\delta_\tau(u) = u[\tau - J(U < 0)]$, $\tau, b_\mu(\cdot)$ 分别用于描述分位数损失函数的具体形式、在 0~1 区间取值的分位水平、惩罚函数,控制参数 μ 用于调节模型的复杂度。惩罚函数仅有惩罚非参数部分的 B 样条基函数的系数; F_i 代表条件分布函数。

1.3.2 惩罚函数

设定 $b_\mu(\varphi)$ 为非减和凹,且满足 $\varphi \in [0, +\infty)$,

处于 $(0, +\infty)$, 区间内存在连续的导数,用 $b_\mu(\varphi)$ 描述。本文采用非凸惩罚函数 SCAD 进行惩罚估计,其惩罚函数为:

$$b_\mu(\varphi; r) = \begin{cases} \mu|\varphi|, & |\varphi| \leq \mu \\ \frac{-\varphi^2 + 2r\mu|\varphi| - \mu^2}{2(r-1)}, & \mu < |\varphi| \leq r\mu \\ \frac{(r+1)\mu^2}{2}, & |\varphi| > r\mu \end{cases} \tag{7}$$

式中: $r \geq 2, \mu \geq 0$ 。在 $(-\infty, 0) \cup (0, \infty)$ 中, SCAD 惩罚函数既连续又可微,其在原点呈现奇异特性,从而达成稀疏性估计的目的。在 $|\varphi| > r\mu$ 时导数为 0, 函数值表现为常数,防止在模型中对较大系数进行过度惩罚。用(8)式表达当 $\varphi \in [-r\mu, r\mu]$ 时的导数公式:

$$b'_\mu(\varphi) = \mu \left[1(\varphi \leq \mu) + \frac{(r\mu - \varphi)_+}{(r-1)\mu} 1(\varphi > \mu) \right] \tag{8}$$

式中: $(r\mu - \varphi)_+ = \max\{x, 0\}$, 本文中指定 $r=3.5$ 。SCAD 惩罚回归具有生成稀疏解的能力,且能无偏估计较大系数。

1.4 线性可加空间自回归模型估计的渐近性分析

参数空间用 $\Theta = A \times B$ 表示,用 λ_0 表示斜率系数,当协变量 D_i, X_i, Z_i 被指定时,条件分布函数、条件密度函数分别用 $F_i(\cdot)$ 以及 $f(\cdot) = f_{e|D_i, X_i, Z_i}$ 表示。以下假设是推导约束条件下线性可加空间自回归模型估计的渐近性质的前提条件。

假设 1: 不等式满足 $E\|X\|^4 \leq w, E\psi_j^4 \leq w\rho_j^2, (j \geq 1)$ 。 w 是取值各异的常量。 Ψ, ρ 分别表示不相关的函数型主成分得分和协方差函数矩阵对应的特征值。

假设 2: 具有使 $w^{-1}j^{-a} < \rho_j < wj^{-a}, \rho_j - \rho_{j+1} > w^j j^{-a-1}, (j \geq 1)$ 成立的 $a > 1$ 。

假设 3: 不等式 $|\chi_j| < wj^{-b}, (j > 1)$ 成立的前提条件是 $b > a/2 + 1$ 。

假设 4: 分位数水平 τ 指定值,参数空间 Θ 的

内节点用 $[\vartheta(\tau), \chi(\tau)]$ 表示。

假设5: 随机误差项 e_i 存在有界的密度函数 $f_i(\cdot)$, 而且在接近零的地方连续可微, 其一阶导数是有界的, 全部 i 都满足。

假设6: 参数空间 Θ 中, 处于点 $(\vartheta, \chi)'$ 位置的 $\partial E[I(Y < M\vartheta + V'\chi)\tilde{X}]/\partial(\vartheta, \chi')$ 是满秩的。

假设7: 参数空间 Θ 的映射 $(\vartheta, \chi) \mapsto E\left\{\left[\tau - I(Y < M\vartheta + V'\chi)\right]\tilde{X}\right\}$ 是单射的。

假设8: 界定 $\Omega = \text{diag}\{f_i[\Psi_i(\tau)]\}$, 式中, $\psi_i(\tau) = M_i\vartheta(\tau) + V_i'\chi(\tau) + K_i'\varsigma(\tau)$, $\varsigma(\tau)$ 是工具矩阵 K 对应的系数。 $V = (\chi', \varsigma)'$, 且 $J_v = \lim_{n \rightarrow \infty} \frac{1}{n} \tilde{X}'\Omega$, 式中, $\tilde{X} = [V, K]$ 。正定矩阵为 $J_\vartheta = \lim_{n \rightarrow \infty} \frac{1}{n} \tilde{X}'\Omega M$ 和 $S = \lim_{n \rightarrow \infty} \frac{\tau(1-\tau)}{n} \tilde{X}'\tilde{X}$ 。矩阵 J_v^{-1} 的公正分割用 $[\bar{J}_\chi, \bar{J}_\varsigma]$ 表示, 而且 $H = \bar{J}_\varsigma' A \bar{J}_\varsigma$, 故矩阵 J_v 与 $J_\vartheta H J_\vartheta$ 是可逆的。

假设9: $\max_i \|Y_i\| = \max_i \|X_i\| = U(\sqrt{n})$ 。

假设10: 矩阵 $I - \vartheta W$ 是满秩的, 矩阵 W 与 $(I - \vartheta W)^{-1}$ 两者的行之和以及列之和全部一致有界, 以上成立的前提条件是 $|\vartheta| < 1$ 。

假设1至假设3的条件在函数型性模型内是传统的假设。假设1为通常的矩假设条件。假设2条件内, 参数 a 决定了核函数曲线的平滑幅度, 也反映了斜率系数估计的难度。假设2中的第2个假设条件确保特征函数能被估计出来, 为斜率系数的准确估计提供充分的依据。假设3度量了函数 $\lambda(\tau, t)$ 曲线的光滑幅度。条件假设4至假设10为空间分位数自回归模型所需要满足的标准条件。当条件假设1至假设10均得到满足时, 指定 $\tau \in (0, 1)$, 则其存在:

$$\sqrt{n}[\hat{\vartheta}(\tau) - \vartheta(\tau)] \xrightarrow{a} N(0, RSR') \quad (9)$$

式中: $R = (J_\vartheta' H J_\vartheta)^{-1} J_\vartheta' H$, $H = J_\varsigma' A J_\varsigma$ 。矩阵 \bar{J}_χ 和 \bar{J}_ς 的维数分别是 $\dim(\chi) \times \dim(\chi, \varsigma)$ 和 $\dim(\varsigma) \times \dim(\chi, \varsigma)$ 。在条件假设8里规定了矩阵 S , 矩阵 J_v^{-1} 的公正分割用 $[\bar{J}_\chi, \bar{J}_\varsigma]'$ 表示。

在 $\dim(\varsigma) = \dim(\vartheta)$ 成立时, $\hat{\vartheta}(\tau)$ 的渐近方差不受矩阵 $A(\vartheta)$ 的影响, $\hat{\vartheta}(\tau)$ 的渐近方差的简洁方式为 $J_\vartheta^{-1} S J_\vartheta^{-1}$ 。假如 $\dim(\varsigma) > \dim(\vartheta)$, 则 $\hat{\vartheta}(\tau)$ 的渐近方差会受到矩阵 $A(\vartheta)$ 选择结果的影响。 $\hat{\varsigma}(\vartheta, \tau) = \bar{J}_\varsigma' S J_\varsigma'$ 的渐近方差的逆选取矩阵 $A(\vartheta)$, 所以获得 $\sqrt{n}[\hat{\vartheta}(\tau) - \vartheta(\tau)]$ 的渐近方差, 用式(10)表示。

$$\omega_\vartheta^2 = \left[J_\vartheta' \bar{J}_\varsigma' (\bar{J}_\varsigma' S \bar{J}_\varsigma)^{-1} \bar{J}_\varsigma J_\vartheta \right]^{-1} \quad (10)$$

在理论上, 如果满足一定的条件, 选用高效的工具变量以及适当的得分函数权重矩阵, 可以实现参数 $\hat{\vartheta}$ 的估计更加准确可靠。

如果条件假设1至假设10成立, 设定截断参数 m 符合 $m \in n^{1/(a+2b)[15]}$, 由此可得式(11):

$$\left\| \hat{\lambda}(\tau, t) - \lambda_0(\tau, t) \right\|^2 = U_p \left[n^{1-2b/(a+2b)} \right] \quad (11)$$

斜率系数的估计在最优收敛速度下进行, 并且函数项部分不会改变参数估计 $\hat{\vartheta}$ 的渐近分布特性。

2 实验分析

为了深入探究某市交通状况随时间的演变趋势, 并验证约束条件下线性可加空间自回归模型方法在城市交通分析中的实用性和有效性, 本次选取该市1995—2020年的交通数据作为实验数据。这些数据涵盖了交通流量、行驶速度等关键指标, 为分析城市交通状况提供了丰富的数据支撑。实验数据主要来源于该市交通管理局、统计局等相关政府部门以及公开的市场数据。数据包括每日的交通流量、行驶速度、交通拥堵指

数等详细数据。在收集到原始数据后,根据研究目的和模型要求,对数据进行筛选和整合。具体步骤包括剔除异常值、缺失值填充、数据标准化等,以确保数据的准确性和可靠性。同时,还对数据进行了时间序列分析,以识别出可能存在的季节性或周期性模式。尽管本文主要关注线性回归矩阵的均值和标准差随时间的变化趋势,但为更好地理解和分析交通状况的空间分布和相互影响,也对交通数据进行了空间化处理,包括构建空间权重矩阵等。

根据研究目的和数据的可用性,选择交通流量、行驶速度等关键变量作为模型的解释变量和因变量。同时,为控制其他影响交通状况的因素,加入城市经济、人口规模等辅助变量。将数据按照时间顺序进行排序,并构建时间序列模型来分析交通流量和行驶速度等变量的变化趋势。通过计算线性回归矩阵的均值和标准差,可以更清晰地了解这些变量随时间的变化情况。

在设定好实验数据背景后,采用本文所提出的统计推断方法对该市的交通数据进行实证分析,由式(12)的线性可加空间自回归模型在随机误差服从标准正态分布时生成数据。

$$Y = \delta WY + \sigma_1 X_1 + \sigma_2 X_2 + g_1(V_1) + g_2(V_2) + e \quad (12)$$

式中： $\sigma_0 = (1.1)^T$, $X_1 = (X_{11}, X_{12}, \dots, X_{1N})^T$, $X_2 = (X_{21}, X_{22}, \dots, X_{2N})^T$, 其中 X_{1N} 和 X_{2N} 服从标准正态分布。
 $V_1 = (V_{11}, V_{12}, \dots, V_{1N})^T$,
 $V_2 = (V_{21}, V_{22}, \dots, V_{2N})^T$, V_{1N} 、 V_{2N} 二者互不相关,在 $[0, 1]$ 区间服从标准均匀分布。 $g_1(t) = 0.5 - t$, $g_2(t) = \sin(2\pi t)$ 因空间权重矩阵中存在 P 个与不相关的区块,且各个区块存在 O 个单体,相同区块的元素权重一致,故 $W = I_p \otimes B_o$, $B_o = (L_o L_o^T - I_o) / (O - 1)$, L_o 、 \otimes 分别代表 O 长度的全部元素向

量是 1 以及 Kronecker 积。选取 $P=30, 60, O=15, 30$, 分位数水平 $\tau=0.1, 0.4, 0.8$, 分别代表高、中、低 3 种水平。空间滞后参数 $\delta=0.1, 0.4, 0.8$, 针对空间依赖的强、中、弱 3 种联系。结果如表 1 所示。

表 1 随机误差服从标准正态分布的推断结果

分位数	δ	O	P	MSE (δ)	MSE (σ_1)	MSE (σ_2)	MAD (g_1)	MAD (g_2)
0.1	0.1	15	30	13.33	8.56	8.92	16.40	16.83
		15	60	11.99	5.91	5.89	11.57	11.49
		30	30	13.48	5.98	5.91	11.43	11.70
		30	60	11.76	4.17	4.28	8.66	8.74
	0.4	15	30	10.93	8.77	8.62	16.29	16.56
		15	60	9.01	5.91	6.13	11.48	11.59
		30	30	11.23	6.06	6.08	11.28	11.67
		30	60	9.15	4.29	4.24	8.60	8.71
	0.8	15	30	5.63	9.10	8.99	16.57	16.55
		15	60	3.89	6.03	6.24	11.50	11.71
		30	30	5.81	6.19	6.07	11.40	11.60
		30	60	3.73	4.16	4.23	8.68	8.81
0.4	0.1	15	30	10.93	8.77	8.62	16.29	16.56
		15	60	11.24	4.32	4.53	8.36	8.52
		30	30	12.93	4.41	4.32	8.36	8.51
		30	60	11.24	3.07	3.03	6.33	6.42
	0.4	15	30	10.32	6.62	6.31	11.91	12.28
		15	60	8.47	4.49	4.61	8.37	8.65
		30	30	10.67	4.26	4.55	8.29	8.51
		30	60	8.51	3.09	3.18	6.28	6.44
	0.8	15	30	5.03	6.64	6.76	11.86	12.33
		15	60	3.76	4.48	4.67	8.42	8.56
		30	30	5.09	4.68	4.55	8.36	8.68
		30	60	3.70	3.05	3.16	6.36	6.54
0.8	0.1	15	30	13.22	8.60	8.50	16.42	16.58
		15	60	12.71	6.07	6.17	11.54	11.66
		30	30	13.49	6.20	5.79	11.35	11.42
		30	60	11.99	4.19	4.29	8.74	8.77
	0.4	15	30	11.67	8.59	8.83	16.49	16.73
		15	60	9.71	6.07	6.18	11.53	11.61

续表

分位数	δ	O	P	MSE ($\hat{\sigma}$)	MSE (σ_1)	MSE (σ_2)	MAD (g_1)	MAD (g_2)
0.8		30	30	11.49	6.36	6.42	11.42	11.64
		30	60	9.75	4.16	4.21	8.82	8.82
	15	30	6.40	9.00	9.08	16.32	16.60	
		60	4.67	6.27	6.32	11.40	11.59	
	30	30	6.37	5.98	6.48	11.53	11.65	
		60	4.34	4.31	4.11	8.78	8.77	

表 1 中的偏差、均方误差以及平均绝对偏差的数值均已扩大了 100 倍。验证次数 560 次, MSE 值考虑了模型的线性特征。MSE($\hat{\sigma}$) 规定成 $(1/560) \sum_{i=1}^{560} (\hat{\sigma}_i - \sigma_o)^T (\hat{\sigma}_i - \sigma_o) 560$, MSE($\hat{\delta}$) 规定成 $(1/560) \sum_{i=1}^{560} (\hat{\delta}_i - \delta_o)^2$ 。平均绝对偏差 MAD 统计指标是非参数可加函数估计标准, 如式(13) 所示。

$$MAD[\hat{g}(v)] = (1/560) \sum_{j=1}^{560} \int_0^1 |\hat{g}(v^j) - g(v^j)| hv(13)$$

式中:数据集重复运算 j 次的平均值用 v^j 表示。通过分析表 1 中的数据, 可以明确地看到, 随着样本数量的递增, 均方误差 MSE 和平均绝对偏差 MAD 都在逐渐减小。这一趋势表明, 所采用的估计方法具有稳健性, 能够有效地处理服从标准正态分布的误差项, 这在现实问题的数据分析和统计推断中尤为重要。由于实际数据往往受到多种因素的影响, 误差项的分布往往是不确定的。因此, 方法的稳健性是保证分析结果可靠性的关键。此外, 该方法不仅理论依据坚实, 而且操作简便, 能够快速得出分析结果, 解决实际问题的实用性增强。综上所述, 本文所提出的方法不仅在理论上具有可行性, 而且在实践中易于操作, 为数据的统计推断提供了有效的工具。

对比约束与非约束下的 2 种稀疏估计, 取 $a=80, 160$, 验证次数 560 次, 参数部分考虑 95% 下的平均覆盖率 ACP、置信区间长度 ALC、误差 Bias 以及均方根误差 RMSE, 结果如表 2 所示。

表 2 约束与非约束条件下估计对比结果

样本数	有约束估计				无约束估计					
	Bias	RMSE	ACP	ALC	Bias	RMSE	ACP	ALC		
$\delta=0.1$	$a=80$	δ	-0.003 4	0.059 5	94.09	0.221 2	-0.004 6	0.063 7	91.47	0.222 6
		σ_1	0.000 8	0.094 9	92.95	0.341 0	-0.000 8	0.113 3	90.12	0.394 6
		σ_2	-0.000 7	0.076 0	91.37	0.273 6	-0.000 9	0.111 7	88.35	0.3937
	$a=160$	δ	-0.000 6	0.040 2	93.63	0.152 7	-0.000 6	0.042 9	93.45	0.155 6
		σ_1	0.000 1	0.064 2	92.87	0.239 0	-0.000 5	0.075 3	92.58	0.278 3
		σ_2	-0.000 7	0.050 3	93.24	0.193 0	-0.002 6	0.079 1	93.12	0.279 7
$\delta=0.4$	$a=80$	δ	-0.000 8	0.041 7	94.12	0.145 8	0.001 4	0.044 9	91.89	0.151 2
		σ_1	-0.000 5	0.095 6	93.32	0.340 1	0.002 2	0.113 8	93.12	0.396 2
		σ_2	-0.000 7	0.076 7	92.19	0.275 6	-0.001 8	0.114 4	92.35	0.402 2
	$a=160$	δ	-0.000 1	0.025 8	92.89	0.099 2	0.000 7	0.028 8	91.89	0.104 0
		σ_1	-0.000 3	0.062 4	92.96	0.239 2	-0.000 3	0.075 5	92.87	0.279 9
		σ_2	-0.001 1	0.051 4	93.04	0.192 1	-0.001 9	0.275 1	91.96	0.281 9
$\delta=0.8$	$a=80$	δ	-0.001 2	0.015 5	92.14	0.058 2	0.000 4	0.019 0	93.45	0.064 0
		σ_1	0.000 4	0.065 6	93.24	0.339 1	0.001 2	0.112 2	92.45	0.398 4
		σ_2	0.000 6	0.073 8	92.48	0.272 6	-0.000 2	0.114 0	94.32	0.404 5
	$a=160$	δ	0.001 2	0.009 8	93.68	0.040 2	-0.001 6	0.013 0	91.86	0.145 2
		σ_1	-0.000 8	0.062 1	91.98	0.239 8	-0.002 1	0.074 8	92.78	0.282 6
		σ_2	-0.000 5	0.050 5	94.07	0.193 0	0.002 1	0.077 5	91.68	0.288 4

从表2可以清晰地看到,随着样本量的增加,参数无论是约束还是非约束估计,其误差 Bias 的绝对值都逐渐趋近于0。这表明这2种估计方法在样本量增大的情况下都能得到越来越准确的结果,从而验证了它们的相合性。当约束条件成立时,参数约束稀疏估计的RMSE明显小于非约束估计,这意味着在约束条件下,本文估计方法能够更精确地捕捉到数据的内在规律。另外,虽然2种估计方法所得到的参数置信区间都随着样本量的增加而接近实际的95%置信区间水平,但约束稀疏估计的区间长度ALC明显短于非约束估计。说明本文方法在给出相同置信水平的同时,能够提供更精确的区间估计。综上所述,本文提出的约束稀疏估计相较于非约束估计在各个方面都表现得更优秀,因此在实际应用中具有更高的实用价值。

选取数据集中该市1995—2020年的交通数据。给出随时间变动的线性回归矩阵 $X(t)$ 的均值以及标准差的离散分布图与拟合曲线描绘图,如图1所示。

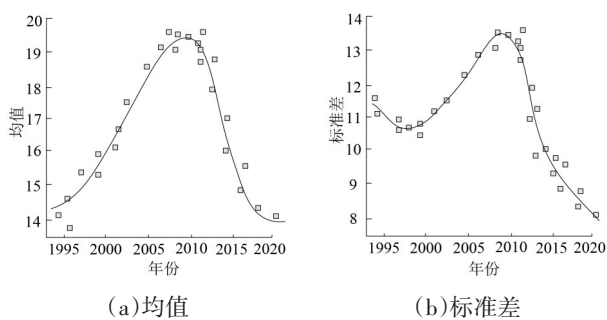


图1 均值与标准差离散分布与拟合曲线图

如图1所示,在深入分析该市1995—2020年的交通数据时,利用统计分析方法,分析线性回归矩阵的均值和标准差随时间的变化趋势。首先,从线性回归矩阵的均值变化来看,自1995年起,该市的交通流量和行驶速度等相关指标呈现稳步增长态势。据统计,1995年该市的日均交通

流量与2010年该市的日均交通流量的相比,这一指标已经增长了约1.4倍。这一显著增长主要得益于城市经济的飞速发展、人口的不断增加以及道路交通网络的持续完善。然而,从2010年开始,线性回归矩阵的均值却出现了下降趋势。通过对比数据,可以发现自2010至2020年,该市的日均交通流量虽然仍保持在较高水平,但增速明显放缓,且在某些年份甚至出现了小幅下降。这一变化可能归因于市政府近年来实施的一系列交通管理措施,如推广公共交通、限制私家车进入市中心等,有效缓解了交通拥堵问题,提高了城市交通效率。

与此同时,本文还关注了线性回归矩阵的标准差随时间的变化。数据显示,在2008年左右,标准差达到了峰值。结合历史背景,发现这一年正值全球经济危机,该市的经济发展受到了严重影响,导致许多企业倒闭、员工失业,进而影响了人们的出行需求和交通状况。交通流量的波动加剧,使得标准差达到了最高点。然而,在度过经济危机后,标准差开始逐渐下降。这表明随着经济的复苏和交通管理措施的进一步完善,该市的交通状况逐渐趋于稳定。特别是近年来,随着智能交通系统的推广和大数据技术的应用,市政府能够更加精准地掌握交通状况,及时采取措施应对交通拥堵和事故等问题,从而进一步降低了交通流量的波动。

综上所述,通过深入分析该市1995—2020年的交通数据,可以清晰地看到交通状况随时间的演变趋势。这不仅验证了本文提出的分析方法的实用性和有效性,还为政策制定者和规划者提供了有力的数据支持。

3 结论

为解决统计推断大规模数据集时面临的稳

定性和效率问题,本文采用一种创新的约束条件下线性可加空间自回归模型的统计推断方法。这种方法结合线性可加空间自回归模型和约束条件,通过约束条件下稀疏估计来优化模型参数。经过深入的理论研究,证明此估计的渐进性质,为其有效性提供坚实的理论基础。通过实际数据验证,本文所提出的方法表现出显著的优势。与传统的非约束估计相比,约束稀疏估计在

各个方面都展现出更高的性能。这不仅体现在模型的预测精度上,还体现在模型的稳定性和泛化能力上。这种方法为数据的统计推断提供更加高效和稳定的工具,有助于解决众多经济、交通等领域的实际问题。下一步,将深入研究如何将惩罚参数引入模型中,以进一步完善模型。这将有助于扩大模型的应用范围,使其能够处理更为复杂的实际问题。

参考文献:

- [1] 冯三营,和文琦.局部平稳部分线性面板数据模型的统计推断[J].统计与决策,2022,38(3):34-39.
- [2] 白仲林,孙艳华.一种协整时间序列的动态因果效应估计与推断方法[J].统计研究,2021,38(10):134-150.
- [3] 韩玉,田宝成,王书鹏.时变方差自回归模型的联合估计函数方法推断[J].北京航空航天大学学报,2022,48(5):756-761.
- [4] CHEN Z, CHEN M, XING G. Bayesian estimation of partially linear additive spatial autoregressive models with P-splines[J]. Mathematical Problems in Engineering, 2021(1):1-14.
- [5] LUSTOSA D M, MILO A. Mechanistic inference from statistical models at different data-size regimes[J]. ACS Catalysis, 2022, 12(13):7886-7906.
- [6] 张洁,张玉,董小刚.自激励广义二项门限自回归模型的统计推断[J].吉林大学学报(理学版),2023,61(2):275-284.
- [7] 赵志文,徐圣楠.周期随机系数自回归模型的统计推断[J].统计与决策,2022,38(4):50-54.
- [8] 丁飞鹏.固定效应部分线性单指标空间自回归面板模型的二次推断函数估计[J].高校应用数学学报A辑,2023,38(4):407-426.
- [9] 王慧,吴茜茜.空间自回归模型中系数变量及误差项的贝叶斯估计[J].合肥工业大学学报(自然科学版),2021,44(9):1291-1296.
- [10] 胡亚南,王金天,田茂再.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,2022,41(5):843-857.
- [11] 刘高生,柏杨,余平.部分函数型线性空间自回归模型的假设检验[J].数学学报(中文版),2023,66(2):239-252.
- [12] 陈秀平,蔡光辉.部分线性可加分位数回归模型的NG估计量[J].统计与决策,2021,37(18):24-27.
- [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.
- [14] 李云霞,王心愉.函数型部分线性单指标空间自回归模型[J].高校应用数学学报A辑,2023,38(2):151-165.
- [15] 任帅,程文慧,周洁.混合广义部分线性加性模型的参数估计[J].广西师范大学学报自然科学版,2022,40(1):108-124.