

分类号 0212
U D C

密级 公开
编号 10741

兰州财经大学

LANZHOU UNIVERSITY OF FINANCE AND ECONOMICS

硕士学位论文

论文题目 空间变系数分位回归模型
的参数估计与应用

研究生姓名: 梁永玉

指导教师姓名、职称: 田茂再 教授

学科、专业名称: 统计学 数理统计学

研究方向: 复杂数据分析

提交日期: 2022年5月30日

独创性声明

本人声明所呈交的论文是我个人在导师指导下进行的研究工作及取得的研究成果。尽我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示了谢意。

学位论文作者签名： 梁永玉 签字日期： 2022.5.30

导师签名： 田茂再 签字日期： 2022.5.30

关于论文使用授权的说明

本人完全了解学校关于保留、使用学位论文的各项规定， 同意（选择“同意”/“不同意”）以下事项：

1.学校有权保留本论文的复印件和磁盘，允许论文被查阅和借阅，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文；

2.学校有权将本人的学位论文提交至清华大学“中国学术期刊（光盘版）电子杂志社”用于出版和编入CNKI《中国知识资源总库》或其他同类数据库，传播本学位论文的全部或部分内容。

学位论文作者签名： 梁永玉 签字日期： 2022.5.30

导师签名： 田茂再 签字日期： 2022.5.30

Parameter Estimation and Application of Spatial Varying Coefficient Quantile Regression Models

Candidate : LIANG Yongyu

Supervisor: TIAN Maozai

摘 要

空间结构数据分析与统计建模问题一直是许多科学领域研究的热点话题。简单的全局回归建模不仅忽略空间地理位置的数据信息,而且当协变量存在空间非平稳效应时无法满足研究的需要。空间变系数模型(SVCMs)是一种研究空间非平稳数据的有效工具,它可以用来研究空间依赖性和空间非平稳性质的影响关系。近年来,许多文献致力于探讨均值意义下的 SVCMs。然而,由于数据产生机制的复杂性,均值回归方法高斯误差的强制假设在实际生活中很难满足。分位回归估计方法会克服这一缺陷。所以将均值回归推广至分位回归很有必要。

首先,本文基于空间变系数非参数回归模型,研究了一种基于二元惩罚样条逼近的分位回归估计方法,该估计方法不仅可以处理具有复杂边界、不规则形状的空间区域,而且还展现出不同分位水平下的解释能力。在两种不同的情形下,分别给出了所提出估计量的理论性质,即收敛速率和渐近分布。对于参数的估计过程,提出一种基于交替方向乘子(ADMM)迭代算法实现模型的求解。对于模型拟合优度检验问题,提出了基于 Bootstrap 方法的拟合优度检验方法,并给出了检验实现算法。基于复合分位回归的良好性能,也提出了该模型的二元惩罚样条逼近的复合分位回归估计量和其实现算法。数值模拟结果显示所提出的估计方法比均值意义下更稳健。

其次,考虑到空间数据的平稳效应和非平稳效应共存,提出空间部分线性变系数模型(SPLVCMs),它是一种半参数模型,其线性部分和非参数部分分别表示空间协变量的平稳性和非平稳性影响。对于模型参数估计,提出了基于二元惩罚样条的非参数逼近分位回归估计方法,研究了其线性部分与非参数部分估计量的理论性质。针对模型实现,提出了基于 ADMM 的参数估计算法。不同情形下的仿真模拟显示所提出的模型和方法在非正态误差下有明显的优势,估计结果更加稳健、有效。

最后,利用佛罗里达州的失业率实际数据和我国空气质量数据验证本文所提出方法的应用价值。结果表明,所提出的模型能充分刻画协变量和响应变量之间的非平稳性影响关系,可为后续应用研究提供新的方法理论。

关键词: 空间统计模型 变系数与部分线性变系数 非参数与半参数 分位回归与复合分位回归 ADMM 算法 二元惩罚样条近似

Abstract

Analysis and statistical modeling of spatial structure data has always been a hot topic in many scientific fields. The simple global regression modeling not only ignores the data information of spatial geographic location, but also fails to meet the needs of research when the covariables have spatial non-stationary effects. Spatial Varying Coefficient Models(SVCMs) is an effective tool to study spatial nonstationary data. It can be used to study the relationship between spatial dependence and spatial nonstationary properties. In recent years, the research on SVCMs in the meaning of mean value is common. However, due to the complexity of data generation mechanism, the mandatory assumption of Gaussian error in mean regression method is difficult to satisfy in real life. The quantile regression estimation method can overcome this defect. Therefore, it is necessary to generalize mean regression to quantile regression estimation.

At First, this thesis introduces the spatial varying coefficient nonparametric regression model, and develops a quantile regression estimation method based on bivariate penalty spline approximation. This estimation method can not only deal with the spatial region with complex boundary and irregular shape, but also shows the interpretation ability under different quantile levels. In two different cases, the theoretical properties of the proposed estimator, the rate of convergence and

asymptotic distribution, are given respectively. For the estimation process of parameters, an iterative algorithm based on Alternating Direction Multiplier Method (ADMM) is proposed to solve the model. To test the goodness-of-fit of the model, a goodness-of-fit test method based on Bootstrap method is proposed in this thesis, and the test implementation algorithm is given. Based on the good performance of composite quantile regression, the thesis also presents the composite quantile regression estimator of bivariate penalty spline approximation of the model and its implementation algorithm. The numerical simulation results show that the proposed estimation method is more robust than the mean value.

Secondly, considering the coexistence of stationary effect and non-stationary effect of spatial data, the Spatial Partial Linear Varying Coefficient Models (SPLVCMs) is proposed. This model is a semi-parametric model, including linear part and non-parametric part, which represent the stationarity and non-stationarity effect of spatial covariables respectively. For the parameter estimation, a non-parametric approximation quantile regression estimation method based on bivariate penalty splines is proposed, and the theoretical properties of the linear part and the non-parametric part estimator are studied. A parameter estimation algorithm based on ADMM is proposed to realize the model. In the numerical simulation stage, the simulation results under different conditions show that the proposed method has obvious advantages under

non-normal error, and the estimation results are more robust and effective.

Finally, this thesis uses actual data of Florida unemployment rate and air quality data of China to verify the application value of the proposed method. The results show that the proposed models can fully describe the nonstationary influence relationship between covariates and response variables, which can provide a new method and theory for subsequent application research.

Keywords: Spatial statistical models; Varying coefficient and partial linear varying coefficient; Nonparametric and semi-parametric; Quantile regression and composite quantile regression; ADMM algorithm; Bivariate penalty spline approximation

目 录

1 引言	1
1.1 研究背景与意义.....	1
1.1.1 背景.....	1
1.1.2 研究意义.....	2
1.2 研究现状.....	3
1.3 本文研究内容与研究框架.....	5
1.4 重点与创新点.....	7
2 理论基础	9
2.1 变系数模型.....	9
2.2 分位回归与复合分位回归理论.....	10
2.3 Maxmin-Angle 三角剖分技术.....	11
2.4 基于三角剖分的二元样条.....	12
2.4.1 Benstein 基多项式与光滑性.....	12
2.4.2 二元样条逼近理论.....	15
2.5 交替方向乘子法.....	16
3 空间变系数分位回归模型估计	18
3.1 空间变系数模型.....	18
3.2 二元惩罚样条 QR 估计.....	18
3.2.1 估计.....	18
3.2.2 光滑参数的选取.....	20
3.3 估计量的理论性质.....	21
3.4 SVCMQR-ADMM 算法.....	26
3.5 Bootstrap 拟合优度检验.....	28
3.6 数值模拟.....	29
4 空间变系数复合分位回归模型估计	36
4.1 二元惩罚样条 CQR 估计.....	36

4.2 算法与实现	37
4.2.1 SVCMCQR-ADMM 算法	37
4.2.2 光滑参数的选取	39
4.3 数值模拟	39
5 空间部分线性变系数分位回归模型估计	44
5.1 空间部分线性变系数模型	44
5.2 偏二元惩罚样条 QR 估计	44
5.3 估计量的理论性质	46
5.4 算法与实现	49
5.2.1 SPLVCMQR-ADMM 算法	49
5.2.2 光滑参数的选取	51
5.5 数值模拟	51
6 实际数据应用	57
6.1 失业率数据分析	57
6.1.1 模型与评价	57
6.1.2 拟合优度检验	59
6.1.3 结果分析	59
6.2 空气质量数据分析	62
6.2.1 模型与评价	62
6.2.2 拟合优度检验	63
6.2.3 结果分析	64
7 总结与展望	65
7.1 全文总结	65
7.2 研究展望	66
参考文献	67
作者攻读硕士期间研究成果目录	71
后记	72

1 引言

1.1 研究背景与意义

1.1.1 背景

现如今,统计数据产生机制愈来愈复杂、多样,空间分布的复杂数据随处可见,如人口普查、经济领域、环境科学、流行病学等相关科学。这些数据的共同特点是样本量为 n 的观测数据通常来自 n 个不同的空间地理位置。对于这种数据的回归建模,最简单的方式就是建立全局视角下的空间回归模型,这种建模方式不考虑空间地理位置的影响,即假设空间分布的数据和协变量的固定效应(回归系数)在整个研究区域内具有空间平稳性,而对于感兴趣的响应和解释变量,利用简单多元线性回归模型进行相关问题的研究和分析。遗憾的是,这种简单统计建模不仅忽略了地理位置数据信息,而且当数据与空间位置相关时(这种情况非常常见),虽然大大节省了计算时间和成本,但得到的结果往往会与“真相”背道而驰。为了解决这一问题,一个自然的推广就是将空间结构嵌套入回归建模分析当中,即考虑不同的地理位置有不同的固定效应,模型回归系数推广为空间地理位置的二元函数,这种模型被称为空间变系数回归模型(SVCMs),它可以用来研究空间依赖性和空间非平稳性质的影响关系。

对于空间变系数模型中系数函数的处理,Brunsdon等(1996)利用局部光滑思想提出了地理加权回归(GWR)估计,它是对普通线性最小二乘(OLS)的扩展。基于核的方法主要特点在于它会根据样本点到回归点之间的距离大小给在某一个窗宽内的样本点赋予不同的权重,距离越小权重越大,然后对这些样本点进行加权平均。一个明显的缺点就是当样本量过高时,计算显得比较困难。对于非参数系数函数的另一种处理方式是正交级数序列逼近方法,该方法在一些规则区域内具有较好的结果。

然而,对于空间数据的研究,另一个必须要考虑的问题是地理区域的复杂性。现实中,数据产生区域往往会有尖锐突出的、不规则凹陷的复杂边界或者是存在内部空洞等的不规则区域(甚至是同时包含以上特征)。在这种情况下,以上方法

往往会造成空间复杂区域的“遗漏问题”。近年来, Wang 等(Lai & Wang, 2013; Mu 等, 2018; Kim & Wang, 2021)也考虑了该问题, 并且提出用一种基于三角剖分技术的二元惩罚 B-型样条逼近方法估计空间二元(如经纬度)系数函数。

上述一系列方法一个明显的约束就是它们都假设模型误差都存在而且必须是一个零均值同方差的高斯误差, 是均值估计的相关研究和探讨。这种假设在实际应用中很难满足。如果不考虑这一特点, 在非正态误差(尖峰厚尾, 或偏锋型误差)下, 甚至在不规范的正态分布误差下(混合正态分布, 污染数据, 存在较多的异常值等), 均值意义下的回归建模和一系列统计推断将接近失效。Koenker & Bassett(1978)提出分位回归(QR)在很多情况下会克服传统均值回归存在的缺陷, 使得模型估计更加稳健、有效。基于分位回归, Zou & Yuan(2008)提出了一种新的估计方法, 即复合分位回归(CQR)。并且 QR 和 CQR 在时变系数等一维变系数模型下已经有了大量的研究和应用。

所以, 本文将空间变系数模型在均值意义下的探讨推广至分位回归和复合分位回归, 提出空间变系数分位回归模型, 研究基于三角剖分技术的局部二元惩罚 B 型样条逼近的分位回归估计和复合分位回归估计。

有时候复杂的数据机制和研究问题的主次导致空间变系数模型太过绝对, 如果研究问题中同时存在平稳效应和非平稳效应, 全部使用变系数进行建模会导致不必要的计算压力, 所以本文将考虑进一步改进空间变系数分位回归模型, 推广至空间部分线性变系数分位回归模型。

1.1.2 研究意义

在空间数据统计分析的复杂性大背景下, 有关 SVCMS 稳健估计问题的讨论具有重要的理论意义和现实意义。

理论意义在于, 提出空间变系数非参数回归模型的一种基于二元惩罚样条逼近的分位回归估计方法, 该估计方法不仅可以处理具有复杂边界、不规则形状的空间区域, 而且还展现出不同分位水平下的解释能力。本文拓宽了空间变系数模型的应用领域, 不同于 Mu 等(2018)基于二元惩罚样条逼近对空间变系数模型的分析仅局限于对均值的探讨, 分位回归能够使所研究的问题得到更充分的刻画和描述。所提出的方法, 不仅保留了空间变系数模型的灵活性和可解释性, 还使估

计结果更加稳健有效。

现实意义在于,提出的方法可以为空间非平稳数据和空间半平稳数据的统计建模分析提供理论基础。在应用研究中,可以利用本文的方法得到更加稳健的估计结果。尤其在污染数据、异常数据以及非常态误差数据下应用价值尤为明显。

1.2 研究现状

空间结构数据分析如今扮演着越来越重要的角色。近几年,有大量文献致力于研究空间数据统计建模问题。例如具有微分正则化的空间回归模型 Ramsay(2002), Azzimonti 等(2015), Wilhelm & Sangalli(2016); 测地线低秩薄板样条 Wang & Ranalli(2007); 皂膜平滑技术 Wood 等(2008)。虽然这些方法在单一的双变量函数表示情况下具有良好的性能,但它们不允许协变量系数函数随地点变化。当数据表现出强空间依赖性时,往往会造成欠拟合等结果,导致很难刻画数据背后的“真相”。随着数据产生机制愈加复杂化,线性模型理论迅速发展,并延伸出了很多变体,其中之一就是将线性模型(LMs)推广至变系数模型(VCMs),其协变量系数会随着某一个光滑变量而发生变化,如 Hastie & Tibshirani(1993), Fan & Zhang(2008)。相比之下,VCM 不仅保留了 LM 的可解释性,而且变得更具灵活性。它不仅可以探索数据的动态特性,还可以减小模型拟合的偏差。在许多实际领域中,收集到的数据除了与时间等一元变量有关外,通常与地理位置有关,所以空间变系数模型(SVCMs)(Brunsdon 等, 1996); Fotheringham 等, 2002; 和 Mu 等, 2018)和广义空间变系数模型(GSVCMs)(Kim & Wang, 2021)自然就被提出,即假设模型中的协变量系数随着某个与空间地理位置相关的二维光滑向量的不同而存在显著差异。值得一提的是,SVCMs 几乎囊括了所有的半参数模型,如当系数为 0 或者为线性函数时,SVCMs 就变为线性模型(LMs); 当指标变量为单变量(如时间)时,变为变系数模型(VCMs)等。

对于 SVCMs 的估计,比较普遍的方法是基于局部核的方法。Brunsdon 等(1996)提出了一种地理加权回归(GWR)估计的方法,Wang 等(2008)提出了一种基于局部线性的 GWR 模型,局部线性估计将系数局部展开为空间位置的线性函数,然后用加权最小二乘法估计展开系数,Sun 等(2014)提出了局部多项式逼近的极大似然方法。基于核的方法主要特点在于它会根据样本点到回归点之间的距离大

小给在某一个窗宽内的样本点赋予不同的权重,距离越小权重越大,然后对这些样本点进行加权平均。一个明显的缺点就是当样本量过高时,计算显得比较困难。另一种对于空间域上变系数模型的处理方式为级数逼近方法。在现实中空间区域可能会有复杂的边界和内部空洞,为了解决复杂区域上的大型数据集估计问题, Kim & Wang(2021)提出了基于局部三角剖分技术的二元惩罚样条逼近方法(BPST)。该方法不仅解决了空间复杂区域的“遗漏问题”,还为非参数回归模型中的非参数分量提供了最优的收敛速率(Lai & Wang, 2013)。

然而上述估计方法均属于均值意义下的探讨,在面对非正态误差、污染数据时,估计结果往往不具备稳健性,进而导致一切基于模型的统计推断失效。Koenker & Bassett(1978)提出分位回归(QR)在很多情况下会克服传统均值回归存在的缺陷,使得模型估计更加稳健、有效。分位回归估计在一元变系数模型中也有广泛的发展和应用。如 Tang(2013),张园园等(2012),Guo(2012)。对于二元变系数模型的分位回归估计,Kong(2015)在二元响应分位回归研究中,提出了二元变系数分位回归估计,但假定系数关于空间恒定,研究时间依赖性。李家勋(2020)提出利用局部多项式展开的方法得到 SVCM 的分位回归估计。

基于分位回归,Zou & Yuan(2008)提出了一种新的估计方法,即复合分位回归(CQR)。CQR 可以同时得到多条分位曲线的估计,使得总体条件分布刻画得更加完整和系统,结果表明,无论误差分布如何,CQR 估计量与最小二乘估计量的相对效率均大于 70%。Guo, Tang & Tian 等(2013)利用 B-样条基函数逼近可加部分的非参数函数,提出了可加性部分线性模型的复合分位回归估计方法,并证明了估计量的大样本性质。Guo 等(2012)在其研究中说明了对于非正态误差,尤其是柯西误差,CQR 估计量可以比最小二乘估计量更有效,有时甚至可以任意地更有效,而对于正态误差 CQR 估计量几乎同样有效。然而,无论是空间变形系数模型的分位回归估计还是复合分位回归估计,在模型中都仅仅考虑了空间数据的非平稳性效应。如果模型中存在空间平稳效应的协变量,并将这些协变量的效应强制转化为非参数函数,不仅对拟合结果没有显著的作用,而且模型解释性也会变得复杂,模型运算时间也会大大增加。所以在模型中考虑平稳效应和非平稳效应的共存问题依然是值得进一步研究。

一个自然的推广就是在空间变系数模型中添加线性部分,转化为空间部分线

性变系数模型。部分线性变系数模型(PLVCMs),是由线性模型(LMs)和变系数模型(VCMs)结合的一种新的半参数回归模型,该模型同时保留了LMs的可解释性和VCMs的灵活性。在该模型中,变系数部分的核心技术在于协变量的系数会随着某一个光滑变量发生变化。近几年有大量的文献致力于此,如Ahmad等(2005)、Fan & Huang(2005)、张波和郭海兵(2015)。但上述文献都是基于均值意义下的探讨,即基于最小二乘和似然方法。这类方法明显的缺点是假设误差分布为正态分布。当数据中存在较多的离群点或者误差为尖峰厚尾分布时,统计推断将会失效。Koenker & Bassett(1978)提出的分位回归估计可以克服上述缺陷,它不假设误差分布的具体形式,也能得到一个稳健有效的结果。在分位回归框架下,对于部分线性模型也有诸多探讨,如Wang等(2009)、Shen & Liang(2018)、刘艳霞等(2021)。然而,上述模型均仅仅考虑了单变量的光滑变量,即认为系数非平稳性只与某个单变量有关。随着数据的复杂化,现实中收集到的数据往往与地理位置有关,一维变系数模型无法满足数据分析和建模的需要。所以本文考虑将部分线性变系数模型推广至空间部分变系数模型(SPLVCMs)。对于SPLVCMs的参数估计问题,目前还没有较为系统的讨论。

不难看出,国内外学者对均值意义下的空间变系数模型和一维变系数(时变)分位回归模型进行了大量的探讨和研究,对于空间变系数模型在分位回归估计和复合分位回归估计下的研究很少,Wang等(Lai & Wang, 2013; Mu等, 2018; Kim & Wang, 2021)基于三角剖分技术二元惩罚样条逼近思想对空间复杂区域数据的建模是很有意义的,但没有在分位回归建模理论基础上进行研究。此外,部分线性变系数模型虽然在一维情况下已经有了大量的研究,但这无法满足空间数据平稳性和非平稳性共存问题统计建模的研究需要。

1.3 本文研究内容与研究框架

本文致力于将空间变系数均值回归模型推广至分位回归模型,提出空间变系数模型的二元惩罚样条逼近分位回归估计方法,其估计量称为SVCM-BPBQR估计量。利用三角剖分技术将空间区域进行分解,形成若干个三角形;然后利用三角剖分上的二元惩罚样条逼近空间变系数函数,其实质就是将二元非参数函数进行参数化,并利用分位回归估计方法得到不同分位水平上的二元系数函数估计。

本文也讨论了二元惩罚样条逼近下分位回归估计量的理论性质,在假设存在样条逼近偏倚和不存在逼近偏倚的不同情形下讨论了估计量的收敛速率和渐近分布。

其次,针对复合分位回归的优良性能,提出了空间变系数模型的复合分位回归模型,并得到复合分位回归估计 $SVCM-BPBCQR$ 估计量。具体过程类似于分位回归估计过程,区别在于这里将会考虑多个分位点。

考虑到空间平稳效应的存在,首先讨论了空间变系数分位回归模型的拟合优度检验问题,提出一种基于 **Bootstrap** 方法的系数稳健性检验方法。其次,还将考虑协变量的空间平稳效应和空间非平稳效应,即考虑两种效应共存的空间部分线性变系数分位回归模型。提出其二元惩罚样条逼近下的分位回归 $SPLVCM-BPBQR$ 估计量,并讨论该估计量的理论性质。

其他创新点还体现在,对于参数估计计算提出了一种比单纯形法(Koenker, 2005)和内点法(Koenker & Park, 1996)更有效的 **ADMM** 算法,称为交替方向乘子法,该方法更适用于本文方法构造的凸优化问题。文中分别提出针对空间变系数分位回归和复合分位回归以及空间部分线性变系数分位回归的 **ADMM** 算法。

文章结构框架图如图 1 所示。依文章结构框架图,论文余下章节具体内容安排如下:第二章主要是与本文相关的理论基础知识,包括分位回归理论、复合分位回归理论,变系数模型与部分线性变系数模型,二维空间区域的三角剖分技术与三角形选取准则,二元张成的 **Benstein** 基多项式与逼近理论以及本文求解模型所用到的最优化算法交替方向乘子法(**ADMM** 算法);第三章至第五章内容是本文的主体部分,主要研究空间变系数模型的分位回归估计、复合分位回归估计和空间部分线性变系数模型的分位回归估计方法,分别提出针对不同模型或者估计方法的 $SVCM-BPBQR$ 、 $SVCM-BPBCQR$ 以及 $SPLVCM-BPBQR$ 三种估计量,研究估计量的大样本性质,并且提出基于 **ADMM** 算法的各估计量的求解算法,也给出了三种估计量的数值模拟结果。此外,这里还研究了基于空间变系数模型分位回归估计的拟合优度检验问题,提出了一种 **Bootstrap** 方法的系数平稳性检验算法。第六章是实际数据应用分析,这里以佛罗里达州的失业率数据集和我国空气质量数据集两个实际数据的应用分析,说明本文所提出方法的应用价值。第七章是研究总结和展望部分,简要总结本文的工作和得出的结论,并展望后续的研究工作切入点。

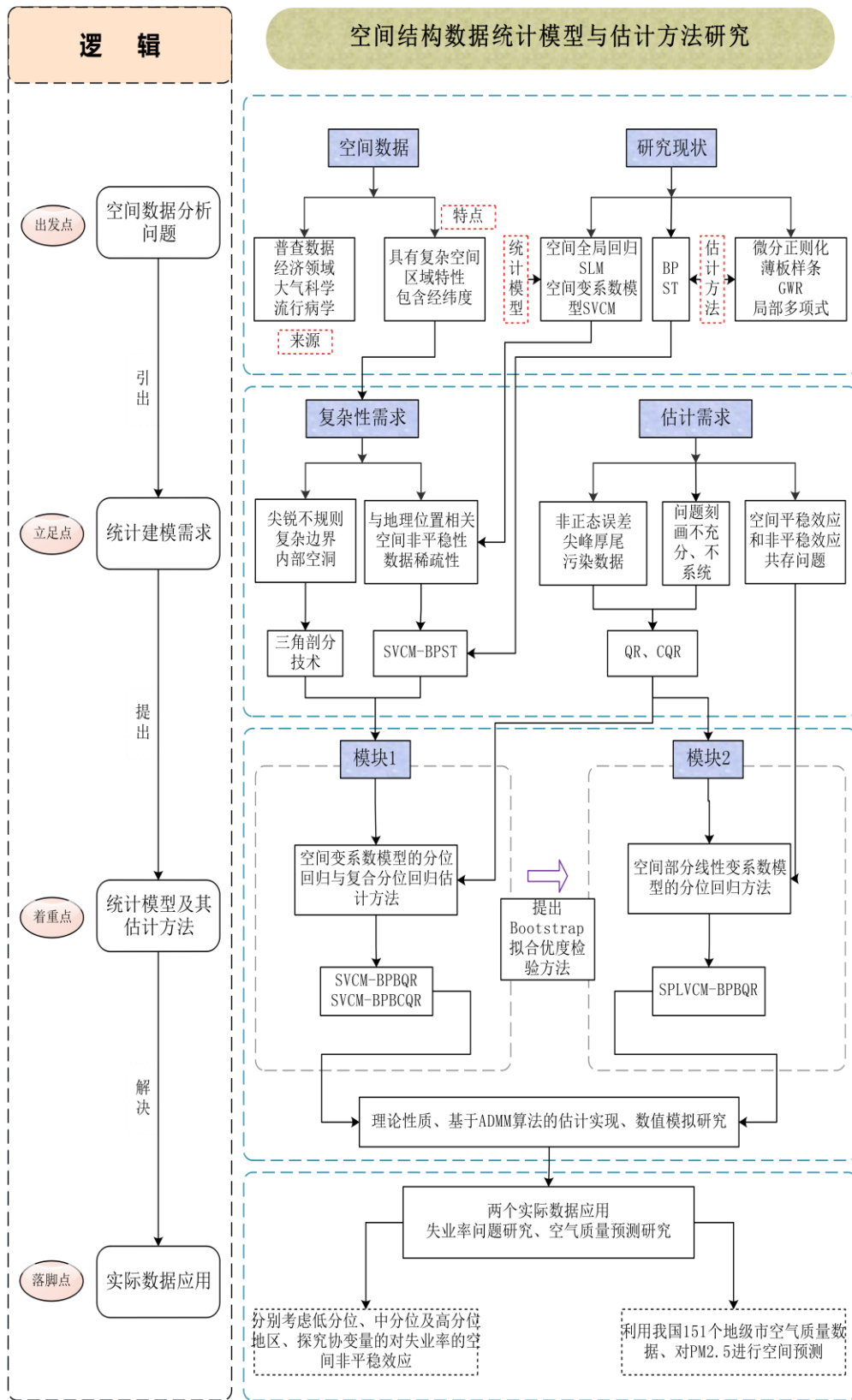


图 1 本文研究框架示意图

1.4 重点与创新点

根据 1.3 节研究内容框架和论文内容, 本文研究重点分为以下内容。一是在现有研究的基础上, 研究已有模型 $SVCMS$ 的稳健估计方法, 实现模型估计方法的推广, 使得该模型能够有效处理污染数据、厚尾数据等、而且使得估计出的参数能够更完整的刻画所研究的问题。二是考虑对 $SVCMS$ 进行推广, 增加模型的灵活性, 使其能够处理空间平稳效应和空间非平稳效应, 减少不必要的计算压力。三是考虑空间全变系数模型的拟合优度检验问题, 进而检验在空间数据分析中是否有必要利用空间变系数模型。其他研究要点有, 研究所提出的估计量的大样本性质、改进模型的求解算法, 使得模型求解快速有效。

创新点可归纳为以下几点:

(1) 研究稳健估计方法。提出空间变系数模型的二元惩罚样条逼近分位回归估计量 $SVCMS-BPBQR$, 复合分位回归估计量 $SVCMS-BPBCQR$;

(2) 推广 $SVCMS$ 。将经典的空间变系数模型推广为空间部分线性变系数模型 ($SPLVCMs$), 提出其分位回归估计量 $SPLVCM-BPBQR$;

(3) 讨论理论性质。讨论估计量的渐近性质包括收敛速度和渐近正态性, 并给出部分性质证明;

(4) 提出 **Bootstrap** 拟合优度检验方法。讨论空间变系数模型中在分位回归估计下的拟合优度检验问题;

(5) 改进模型求解算法。基于 **ADMM** 算法, 提出(1-2)中模型参数求解的最优化迭代算法。

2 理论基础

2.1 变系数模型

在统计建模和数据分析中,经典的线性模型对数据生成机制的错误描述可能会导致较大的偏差(Fan & Zhang, 2008)。而非参数建模对模型不做任何假设,当数据产生存在较大的偏差时,得到的估计量往往会带有更大的偏差,而且当维数过高时,还会存在“维数灾难”问题。基于上述问题,产生了许多新的方法,如投影追踪(Huber, 1985)、切片逆回归(Li, 1991)、单指标模型(Hardle & Stoker, 1990)、变系数模型(Hastie & Tibshirani, 1993)等。其中变系数模型基于实际需要被提出。由于传统的回归模型忽略了隐藏在数据的动态特性,拟合效果有时会大打折扣,更明智的做法可能是让常数参数按照特定的特征变化,从而产生变系数模型。假设由感兴趣的响应变量 Y 、一些解释变量 \mathbf{X} 和系数变化特征 U 构成的 n 个样本 $\{Y_i, \mathbf{X}_i, U_i\}_{i=1}^n$,变系数模型可以表示为

$$\mathbf{Y} = \mathbf{X}^\top \boldsymbol{\beta}(U) + \varepsilon. \quad (2.1)$$

其中 $\boldsymbol{\beta}(\cdot)$ 为一元系数函数向量,对于模型(2.1)的参数估计,比较常用的方法有光滑样条估计和局部多项式逼近(Hoover 等, 1998)、基函数逼近估计(Huang 等, 2002)以及局部线性估计(Fan & Zhang, 2008)。基于基函数逼近的估计将系数函数 $\beta(u)$ 通过基函数展开,即

$$\beta(u) \approx \sum_{l=1}^K \alpha_l B_l(u), \quad (2.2)$$

其中 $\{B_l, l = 1, \dots, K\}$ 表示基函数集合,常见的有多项式基函数、Fourier 基函数以及 B 样条基函数等。

变系数模型的优势在于它不仅吸收了非参数建模的灵活性,进而探讨数据的动态特性和非平稳性,而且还保留了参数模型的可解释性,自提出以来得到了广泛的应用。如今,变系数模型已经成功应用于各个科学领域,包括纵向、函数型和生存数据以及金融和经济数据分析。

2.2 分位回归与复合分位回归理论

回归模型,一直以来是研究感兴趣的响应变量与一些解释变量之间关系的有效工具。Koenker & Bassett(1978)提出的分位回归(QR),可以视为经典均值回归的推广。这表现在分位回归估计方法延拓了估计目标,从估计条件均值的最小二乘法扩展到估计条件分位函数。一般的条件分位函数估计方法就是最小化绝对误差的非对称加权和。其中,中位数回归估计可以是最小化绝对误差和。

首先,这里定义总体条件分位数。设有随机向量 (Y, X) ,其中 Y 在给定 $X = x$ 的条件下的累计分布函数定义为 $F_{Y|X=x}(y|x)$,则 $Y|X = x$ 的 $\tau(0 \leq \tau \leq 1)$ 阶条件分位数为

$$Q_\tau(Y|X = x) = \arg \inf \{y \in \mathbb{R}; F_{Y|X}(y|x) \geq \tau\}. \quad (2.3)$$

对于一般的回归模型

$$Y = g(X) + \varepsilon, \quad X \in \mathbb{R}^p, \quad (2.4)$$

响应变量 Y 的 τ 阶条件分位函数 $g_\tau(x)$ 满足 $\tau = Pr(Y \leq g_\tau(X)|X = x)$ 。若用“检验”函数表示,亦可以等价的定义为

$$g_\tau(x) = \arg \min_{a \in \mathbb{R}} E\{\rho_\tau(Y - a)|X = x\}, \quad (2.5)$$

其中 $\rho_\tau(u) = u\{\tau I(u \geq 0) - (1 - \tau)I(u \leq 0)\}$ 为分位回归参数估计的标准损失函数(又称为检验函数), $I(\cdot)$ 为示性函数。

从极大似然估计的角度,分位回归是由高斯似然到非对称 Laplace 密度的推广。对于一个基于模型(2.4)的样本 $\{X_i, Y_i\}_{i=1}^n$,如果现在假设模型误差 ε 有概率密度函数

$$f(\varepsilon) \propto \exp \left\{ - \sum_{i=1}^n \rho_\tau(Y_i - g(X_i)) \right\}, \quad (2.6)$$

则最大化联合似然函数与最小化检验函数是等价的。

基于此,一般可以从两个角度定义简单的线性分位回归。基于损失函数定义(Koenker & Bassett, 1978):

$$\min_{\beta \in \Theta} E\{\rho_\tau(\mathbf{Y} - \mathbf{X}^\top \beta)|\mathbf{X} = \mathbf{x}\}, \quad (2.7)$$

其中, Θ 为 β 的位置参数空间。基于非对称 Laplace 密度的定义(Yu & Moyeed, 2001):

$$f(\varepsilon) \propto \exp \left\{ - \sum_{i=1}^n \rho_{\tau}(Y_i - \mathbf{X}_i^{\top} \beta) \right\}. \quad (2.8)$$

相比于传统回归模型,分位回归的优势之一在于它刻画了感兴趣条件随机变量的各分位点随解释变量的变化情况,可以得到给定回归量的因变量的整体条件分布。优势之二是分位回归估计的参数结果被视为是稳健的,对响应变量观测值的离群点不敏感,所以当误差分布为易感染离群点分布时,往往比普通最小二乘更加有效。

复合分位回归(CQR)由 Zou & Yuan(2008)提出,与分位回归不同,它在同一个模型中考虑可多条分位回归曲线,并且设定各分位曲线上的参数相同。令 $\rho_{\tau_k}(u) = u[\tau_k - \mathcal{I}(u < 0)](k = 1, \dots, q)$ 为 q 个不同分位曲线的损失函数。记 c_{τ_k} 为 ε 的 $100\tau_k\%$ 分位数,假定 $\tau_1, \tau_2, \dots, \tau_q$ 满足 $0 < \tau_1 < \tau_2 < \dots < \tau_q < 1$,在一般情况下采用等间距分位,即 $\tau_k = k/(q+1)(k = 1, \dots, q)$ 。此外假定 ε 的密度函数在任何地方都存在,所以 c_{τ_k} 也是唯一确定的。以线性模型为例,通过最小化下面式子便可以得到复合分位回归估计

$$(\hat{c}_{\tau_1}, \dots, \hat{c}_{\tau_q}, \hat{\beta}_{CQR}) = \arg \min_{c_{\tau_1}, \dots, c_{\tau_q}, \beta} \sum_{k=1}^q \left\{ \sum_{i=1}^n \rho_{\tau_k}(Y_i - c_{\tau_k} - \mathbf{X}_i^{\top} \beta) \right\}. \quad (2.9)$$

由式(2.9)可以看出,复合分位回归和分位回归的不同之处就在于复合分位回归方法估计得到的系数 $\hat{\beta}_{CQR}$ 是考虑多个分位水平误差的结果,所以复合分位回归与任意分位水平的分位回归相比,会使的研究问题得到更充分和系统的刻画。

2.3 Maxmin-Angle 三角剖分

Lai & Schumaker(2007)提出了一种最大最小角(Maxmin-angle)准则,选择最小角度尽可能大的三角形剖分。Maxmin-Angle 三角剖分,主要通过将任意一个三角剖分中三角形的最小内角最大化,使三角形大小更加均匀。假设有4个顶点 $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_4$ 形成两个不同的三角剖分 T_1 和 T_2 ,每个三角剖分皆有2个三角形和

6 个内角。假设三角剖分 T_1 的 6 个内角中的最小角为 θ_1 ，三角剖分 T_2 的 6 个内角中的最小角为 θ_2 ，并且有 $\theta_1 > \theta_2$ 成立，则按照 Maxmin-Angle 原则，应该选择 T_1 作为更好的三角剖分。Maxmin-Angle 三角剖分的一大优势在于它能产生更加均匀的三角形。因为一般希望在进行三角剖分时，不存在较小的锐角和钝角。而且在进行三角剖分时，选择合适的三角形数量很重要。太少的三角形不能完全刻画系数函数特征，过多的三角形则需要大量的计算时间。在实际应用中，有很多软件包可以实现 Maxmin-Angle 三角剖分。一是使用 Delaunay 算法来进行三角剖分，参见 MATLAB 程序 *delaunay.m*，*dismesh2d.m* 或 MATHEMATICA 函数 *DelaunayTriangulation* (Kim & Wang, 2021)。二是用 DistMesh 是生成非结构三角网格和四面体网格，具体 DistMesh 生成器可参见 <http://persson.berkeley.edu/dismesh/>(Mu 等, 2017)。本文使用 *dismesh2d.m* 来生成三角剖分。再如 Lai & Wang (2019) 提供的 R 软件包 *Triangulation* 中的函数 *TriMesh(Pt, n, H = NULL)*，可建议使用 R 软件包 *Triangulation* 来实现空间区域的三角剖分。

2.4 基于三角剖分的二元样条

本节内容主要介绍三角剖分上的二元样条，包括三角形的重心坐标与 Benstein 基多项式、Benstein 基多项式的方向导数与三角形领边的光滑性，以及逼近理论。主要参考了 Lai & Schumaker(2007)以及 Zhou & Pan(2014)。

2.4.1 Benstein 基多项式与光滑性

假设有非退化的三角形 Δ ，其顶点记为 $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ (顶点记号为逆时针排序)，空间内任意的点 $\mathbf{u} \in \mathbb{R}^2$ 可由顶点 $\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3$ 线性表示为

$$\mathbf{u} = w_1 \mathbf{u}_1 + w_2 \mathbf{u}_2 + w_3 \mathbf{u}_3, \quad (2.10)$$

且满足条件 $w_1 + w_2 + w_3 = 1$ ，其中系数 (w_1, w_2, w_3) 称为点 \mathbf{u} 相对于三角形 Δ 的重心坐标，记为 \mathbf{w}_u 。约束条件 $w_1 + w_2 + w_3 = 1$ 保证了线性表示的唯一性。理论上 \mathbb{R}^2 内所有的点皆可以用式(2.10)表示，而这里只考虑三角形 Δ 内的点或者三角形 Δ 边上的点，即所有的系数 w_i 都是非负的。重心坐标的一个重要性质是它对

笛卡尔坐标的线性变换是不变的。由于二元样条和所有的模型方程随后建立在重心坐标上，它们对线性变换也是不变的。

给定一个三角形 Δ 和点 $\mathbf{u} \in \Delta$ 的重心坐标 $\mathbf{w}_u = (w_1, w_2, w_3)$ ，则关于三角形 Δ 的阶数为 $d \geq 1$ 的 Benstein 基多项式定义为

$$B_{\Delta;ijk}^d(\mathbf{u}) = \frac{d!}{i!j!k!} w_1^i w_2^j w_3^k, \quad i + j + k = d, \quad (2.11)$$

其中 $d \in \mathbb{Z}_+$ 为正整数， i, j, k 为非负整数。需要注意的是，重心坐标 (w_1, w_2, w_3) 都是笛卡尔坐标的线性函数。因此每个 $B_{\Delta;ijk}^d(\mathbf{u})$ 在笛卡尔坐标系中都是 d 次的多项式。令 \mathbb{P}_{Δ}^d 为定义在三角形 Δ 上所有阶数小于等于 d 的多项式空间，则 $B_{\Delta;ijk}^d(\mathbf{u}) \in \mathbb{P}_{\Delta}^d$ 。进一步，Benstein 基多项式的集合可以被定义为 $\mathbf{B}_{\Delta}^d = \{B_{\Delta;ijk}^d(\mathbf{u}) : i, j, k \geq 0, i + j + k = d\}$ ，该集合形成了空间 \mathbb{P}_{Δ}^d 的一组基底，且满足如下条件：

$$(1) \text{ 对于所有的 } \mathbf{u} \in \Delta, \quad 0 \leq B_{\Delta;ijk}^d(\mathbf{u}) \leq 1, \quad \sum_{i+j+k=d} B_{\Delta;ijk}^d(\mathbf{u}) = 1,$$

$$(2) B_{\Delta;ijk}^d(\mathbf{u}) \text{ 在点 } c_{ijk} = (i\mathbf{u}_1 + j\mathbf{u}_2 + k\mathbf{u}_3)/d \in \Delta \text{ 处有唯一的最大值。}$$

由于 \mathbf{B}_{Δ}^d 是多项式空间 \mathbb{P}_{Δ}^d 的一组基，所以任意的多项式 $\mathcal{P} \in \mathbb{P}_{\Delta}^d$ ，可以由 Benstein 基多项式唯一表示为

$$\mathcal{P}_{\Delta}(\mathbf{u}) = \sum_{i+j+k=d} \xi_{\Delta;ijk} B_{\Delta;ijk}^d(\mathbf{u}). \quad (2.12)$$

这种表示通常被称为 \mathcal{P} 相对于三角形 Δ 的 B-型表示。

重心坐标和 Benstein 基多项式非常有利于推导相邻三角形上多项式连续连接的条件。假设有两个三角形 $\Delta_1 = \langle w_1, w_2, w_3 \rangle$ ， $\Delta_2 = \langle w_4, w_3, w_2 \rangle$ ，则这两个三角形的邻边为 $e = \langle w_2, w_3 \rangle$ ，假设分别定义在三角形 Δ_1 和三角形 Δ_2 上的利用相应重心坐标的 Benstein 基多项式为 $\{B_{\Delta_1;ijk}^d\}_{i+j+k=d}$ 和 $\{B_{\Delta_2;ijk}^d\}_{i+j+k=d}$ 。考虑阶数为 d 的分别定义在三角形 Δ_1 和三角形 Δ_2 的多项式 $\mathcal{P}_{\Delta_1}(\mathbf{u})$ 和 $\mathcal{P}_{\Delta_2}(\mathbf{v})$ ，相应 B-型表示分别为

$$\mathcal{P}_{\Delta_1}(\mathbf{u}) = \sum_{i+j+k=d} \xi_{\Delta_1;ijk} B_{\Delta_1;ijk}^d(\mathbf{u}), \quad \mathcal{P}_{\Delta_2}(\mathbf{v}) = \sum_{i+j+k=d} \xi_{\Delta_2;ijk} B_{\Delta_2;ijk}^d(\mathbf{v}). \quad (2.12)$$

假设点 \mathbf{u} 位于邻边 e 上，则 $\mathbf{u} \in e$ 关于三角形 Δ_1 和三角形 Δ_2 的重心坐标 (w_1, w_2, w_3) 分别可以改写为 $(0, w_2, 1 - w_2)$ 和 $(0, 1 - w_2, w_2)$ ，则在此条件下式(2.12)可以改写为

$$\mathcal{P}_{\Delta_1}(\mathbf{u}) = \sum_{i+j+k=d} \xi_{\Delta_1;0jk} \frac{d!}{j!k!} w_2^j (1-w_2)^k, \quad \mathcal{P}_{\Delta_2}(\mathbf{v}) = \sum_{i+j+k=d} \xi_{\Delta_2;ijk} \frac{d!}{j!k!} (1-w_2)^j w_2^k.$$

基于此, 当且仅当 $\xi_{\Delta_1;0jk} = \xi_{\Delta_2;0kj}$ 时, 多项式 $\mathcal{P}_{\Delta_1}(\mathbf{u})$ 和 $\mathcal{P}_{\Delta_2}(\mathbf{v})$ 在邻边 e 上连续, 其中 $j, k \geq 0, j+k=d$ 。

下面给出 Benstein 基多项式方向导数的表达式, 多元光滑函数 $g(\mathbf{u})$ 在点 \mathbf{u} 关于方向 $\boldsymbol{\omega}$ 的导数定义为 $\nabla_{\boldsymbol{\omega}}^1 g$,

$$\nabla_{\boldsymbol{\omega}}^1 g = \left. \frac{\partial}{\partial t} g(\mathbf{u} + t\boldsymbol{\omega}) \right|_{t=0} = \lim_{t \rightarrow 0} \frac{g(\mathbf{u} + t\boldsymbol{\omega}) - g(\mathbf{u})}{t}, \quad (2.13)$$

以此类推, 阶数为 ζ 的高阶导数 $\nabla_{\boldsymbol{\omega}}^{\zeta} g$ 可以根据式(2.13)递归定义。

假设 $\boldsymbol{\omega} = \mathbf{a} - \mathbf{b}$, 其中 \mathbf{a}, \mathbf{b} 为两个点, 它们关于三角形 Δ 的重心坐标分别为 (a_1, a_2, a_3) 和 (b_1, b_2, b_3) 。则 $\boldsymbol{\omega}$ 关于三角形 Δ 的重心坐标是唯一的, 这里假设为 (c_1, c_2, c_3) , 并且有 $c_i = (a_i - b_i) (i=1, 2, 3)$, $c_1 + c_2 + c_3 = 0$ 。则 Benstein 基多项式的 $B_{\Delta;ijk}^d$ 的方向导数有如下表达式

$$\nabla_{\boldsymbol{\omega}}^1 B_{\Delta;ijk}^d(\mathbf{u}) = d\{c_1 B_{\Delta;(i-1)jk}^{d-1}(\mathbf{u}) + c_2 B_{\Delta;i(j-1)k}^{d-1}(\mathbf{u}) + c_3 B_{\Delta;ij(k-1)}^{d-1}(\mathbf{u})\}, \quad (2.14)$$

因此, 任意多项式 $\mathcal{P} \in \mathbb{P}_{\Delta}^d$, 经过式(2.12)的 B-型表示, 其方向导数为

$$\begin{aligned} \nabla_{\boldsymbol{\omega}}^1 \mathcal{P}_{\Delta}(\mathbf{u}) &= \sum_{i+j+k=d} \xi_{\Delta;ijk} \nabla_{\boldsymbol{\omega}}^1 B_{\Delta;ijk}^d(\mathbf{u}) \\ &= \sum_{i+j+k=d} \xi_{\Delta;ijk} d\{c_1 B_{\Delta;(i-1)jk}^{d-1}(\mathbf{u}) + c_2 B_{\Delta;i(j-1)k}^{d-1}(\mathbf{u}) + c_3 B_{\Delta;ij(k-1)}^{d-1}(\mathbf{u})\} \\ &= d \sum_{i+j+k=d} \xi_{\Delta;ijk}^{(1)} B_{\Delta;ijk}^{d-1}(\mathbf{u}), \end{aligned}$$

其中 $\xi_{\Delta;ijk}^{(1)}(\boldsymbol{\omega}) = c_1 \xi_{\Delta;(i+1)jk} + c_2 \xi_{\Delta;i(j+1)k} + c_3 \xi_{\Delta;ij(k+1)}$, 定义 $\xi_{\Delta;ijk}^{(0)}(\boldsymbol{\omega}) = \xi_{\Delta;ijk}$, 则有

$$\xi_{\Delta;ijk}^{(\zeta)}(\boldsymbol{\omega}) = c_1 \xi_{\Delta;(i+1)jk}^{(\zeta-1)} + c_2 \xi_{\Delta;i(j+1)k}^{(\zeta-1)} + c_3 \xi_{\Delta;ij(k+1)}^{(\zeta-1)}, \quad \zeta = 1, 2, \dots, d. \quad (2.15)$$

有了上述定义, $\mathcal{P} \in \mathbb{P}_{\Delta}^d$ 阶数为 ζ 的高阶导数 $\nabla_{\boldsymbol{\omega}}^{\zeta} \mathcal{P}(\mathbf{u})$ 为

$$\nabla_{\boldsymbol{\omega}}^{\zeta} \mathcal{P}(\mathbf{u}) = \frac{d!}{(d-\zeta)!} \sum_{i+j+k=d} \xi_{\Delta;ijk}^{(\zeta)}(\boldsymbol{\omega}) B_{\Delta;ijk}^{d-\zeta}(\mathbf{u}), \quad (2.16)$$

详细内容可参见 Lai & Schumaker(2007)。

现在重新考虑两个三角形 $\Delta_1 = \langle w_1, w_2, w_3 \rangle$, $\Delta_2 = \langle w_4, w_3, w_2 \rangle$, 邻边为 $e = \langle w_2, w_3 \rangle$, 阶数为 r 的多项式 $\mathcal{P}_{\Delta_1}(\mathbf{u})$ 和 $\mathcal{P}_{\Delta_2}(\mathbf{v})$ 。对于任意的 $\boldsymbol{\omega}$, 令 $\nabla_{\boldsymbol{\omega}}^{\kappa} \mathcal{P}(\mathbf{u})$ 为 \mathcal{P} 在点 \mathbf{u} 处关于方向 $\boldsymbol{\omega}$ 的 κ 阶导数。如果 $\mathcal{P}_{\Delta_1}(\mathbf{u})$ 和 $\mathcal{P}_{\Delta_2}(\mathbf{v})$ 所有方向各阶导数 ($\kappa < r$) 在 $e = \langle w_2, w_3 \rangle$ 的每一点上都一致, 即

$$\nabla_{\omega}^{\kappa} \mathcal{P}_{\Delta_1}(\mathbf{u}) = \nabla_{\omega}^{\kappa} \mathcal{P}_{\Delta_2}(\mathbf{u}), \forall \mathbf{u} \in e, \forall \omega, \kappa = 0, 1, \dots, r. \quad (2.17)$$

就称 $\mathcal{P}_{\Delta_1}(\mathbf{u})$ 和 $\mathcal{P}_{\Delta_2}(\mathbf{v})$ 在公共边 $e = \langle w_2, w_3 \rangle$ 上以 r 阶光滑连续。用多项式的 B-型表示式(2.12)重新表达, 即

$$\xi_{\Delta_2; \kappa j k} = \sum_{\alpha + \beta + \gamma = \kappa} \xi_{\Delta_1; \alpha(+\beta)(j+\gamma)} B_{\Delta_1; \alpha \beta \gamma}^{\kappa}(\mathbf{s}), \quad j + k = d - \kappa, \kappa = 0, \dots, r, \quad (2.18)$$

其中 \mathbf{s} 为 Δ_2 的顶点, 且不在邻边 ℓ 。前文中 $\xi_{\Delta_1; 0 j k} = \xi_{\Delta_2; 0 k j}$ 为 $\kappa = 0$ 的特殊情况。

2.4.2 二元样条逼近理论

对于任意形状的空间区域 $\Omega \in \mathbb{R}^2$, 这里利用三角形剖分技术将其划分为若干个小三角形, 然后利用 Benstein 基多项式逼近任意的二元函数(Kim & Wang, 2021)。设区域 Ω 可以分解为 $\Omega = \cup_{i=1}^N \Delta_i$, 其中 $\Delta_1, \Delta_2, \dots, \Delta_N$ 为 N 个互不重叠的三角形, 它们组成的集合 $T = \{\Delta_1, \Delta_2, \dots, \Delta_N\}$ 称为区域 Ω 的一个三角剖分。对于三角剖分 T , 这里指定一些特殊的记号: 记 $|T| = \max\{|\Delta|, \Delta \in T\}$, $r_T = \min\{r_{\Delta}, \Delta \in T\}$ 。其中 $|\Delta|$ 为 Δ 的最长边的模, r_{Δ} 为 Δ 的最大内接圆半径。假设三角形 $\Delta \in T$ 的顶点坐标为 $\mathbf{A}_i = (a_{1i}, a_{2i})^T, i = 1, 2, 3$ 。则三角形 Δ 可表示为 $\Delta = \langle \mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3 \rangle$, 空间内任意的点 $\mathbf{A} \in \mathbb{R}^2$ 可由顶点 $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ 线性表示为 $\mathbf{A} = w_1 \mathbf{A}_1 + w_2 \mathbf{A}_2 + w_3 \mathbf{A}_3$ 。其中 (w_1, w_2, w_3) 称为点 $\mathbf{A} \in \Delta$ 的重心坐标(Lai & Schumaker, 2007)且满足条件 $w_1 + w_2 + w_3 = 1$ 。

下面介绍 Benstein 基多项式和逼近理论。对于三角形 Δ , 阶数为 $d \geq 1$ 的 Benstein 基多项式定义为

$$B_{\Delta; ijk}^d(\mathbf{A}) = \frac{d!}{i!j!k!} w_1^i w_2^j w_3^k, \quad i + j + k = d, \quad (2.19)$$

式(2.19)可以由两个一元 Benstein 基多项式张量积得到。给定阶数 $d \in \mathbb{Z}_+$, 记 \mathbb{P}_{Δ}^d 为三角形 Δ 上所有阶数小于等于 d 的多项式空间, 则 Benstein 基多项式就是该空间的基。对于一个非退化的三角形 Δ 和系数 $\{\xi_{\Delta; ijk}\}$, 任意的多项式 $\mathcal{P} \in \mathbb{P}_{\Delta}^d$, 可以由 Benstein 基多项式唯一表示为

$$\mathcal{P}_{\Delta}(\mathbf{A}) = \sum_{i+j+k=d} \xi_{\Delta; ijk} B_{\Delta; ijk}^d(\mathbf{A}). \quad (2.20)$$

在近似二元函数之前，首先引出样条空间的定义(Kim & Wang, 2021)：给定 $0 \leq \nu < d$ 和三角剖分 T ，定义 T 上的阶数为 d 光滑度为 ν 的样条空间为

$$\mathbb{S}_d^\nu(T) = \{ \mathcal{P} \in \mathbb{C}_\nu(\Omega) : \mathcal{P}_{\Delta_i}(\mathbf{A}) \in \mathbb{P}_{\Delta_i}^d, \Delta_i \in T, i = 1, \dots, N \}, \quad (2.21)$$

式中， $\mathbb{C}_\nu(\Omega)$ 为 Ω 区域上的 ν 阶连续可微函数空间。则对于有 N_j 个三角形的三角剖分 $T_j, j = 0, 1, \dots, p$ ，即假设不同的待逼近函数均有不同的三角剖分。定义 $\{B_{j,m}\}_{m \in \mathcal{M}_j}$ 是一个样条空间 $\mathbb{S}_d^\nu(T_k)$ 上的二元 Benstein 基多项式集合。 \mathcal{M}_j 是三角剖分 T_j 上基函数的指标集，其基数为 $|\mathcal{M}_j| = N_j(d+1)(d+2)/2$ 则第 j 个二元函数 $f_j(\mathbf{A}) \in \mathbb{S}_d^\nu(T_j)$ 可以由 $\{B_{j,m}\}_{m \in \mathcal{M}_j}$ 逼近为

$$f_j(\mathbf{A}) \approx \sum_{m \in \mathcal{M}_j} \xi_{j,m} B_{j,m}(\mathbf{A}) = \boldsymbol{\xi}_j^\top \mathbf{B}_j(\mathbf{A}), \quad (2.22)$$

其中， $\mathbf{B}_j(\mathbf{A}) = \{B_{j,m}(\mathbf{A}), m \in \mathcal{M}_j\}^\top$ 为在位置 \mathbf{A} 上的二元基函数向量， $\boldsymbol{\xi}_j = \{\xi_{j,m}, m \in \mathcal{M}_j\}^\top$ 为与基函数向量所对应的样条系数向量。为了考虑三角剖分中各三角形公共边之间的光滑性，还需引入一个约束矩阵 \mathbf{H}_j ，满足条件 $\mathbf{H}_j \boldsymbol{\xi}_j = 0, j = 0, 1, \dots, p$ 。并且， \mathbf{H}_j 依赖于基函数的光滑参数和三角剖分的结构。

2.5 交替方向乘子法

交替方向乘子法(ADMM)是一种求解复杂优化问题的实现方法，该算法适用于求解典型的分布式凸优化问题。ADMM 的基本思想是将一个复杂的最优化问题分解为多个子优化问题。该算法在统计学领域常应用与罚回归问题，如 LASSO 惩罚回归等。ADMM 算法的一般思路，就是通过“分解-协调”过程，将大的全局目标函数分解为多个较小、较容易求解的局部子最优化问题。参数求解实现就是通过协调多个子问题的解进行迭代而得到全局目标函数的解。在将该算法应用到空间变系数分位回归模型、空间部分线性变系数分位回归模型之前，首先简要回顾一般的 ADMM 算法原理。

假设 $f(\mathbf{x}), g(\mathbf{y})$ 分别是关于 \mathbf{x}, \mathbf{y} 的实值凸性函数，存在一个可分离的最小化问题 $f(\mathbf{x}) + g(\mathbf{y})$ 有约束 $\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{y} = \mathbf{c}$ ，其中 \mathbf{A}, \mathbf{B} 是两个约束系数矩阵， \mathbf{c} 为常数向量。则 ADMM 问题的公式为

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{y}} \quad & f(\mathbf{x}) + g(\mathbf{y}) \\ \text{subject to} \quad & \mathbf{Ax} + \mathbf{By} = \mathbf{c} \end{aligned} \quad (2.23)$$

上述问题可以转化为增广的拉格朗日问题(Powell, 1967), 即

$$\mathcal{L}_h(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f(\mathbf{x}) + g(\mathbf{y}) + \mathbf{z}^\top (\mathbf{Ax} + \mathbf{By} - \mathbf{c}) + \frac{\hbar}{2} \|\mathbf{Ax} + \mathbf{By} - \mathbf{c}\|_2^2, \quad (2.24)$$

其中 \hbar 为调谐参数。现在令缩放因子 $u = (1/\hbar)\mathbf{z}$, $u_j = (1/\hbar)z_j$, 得到更加方便的增广拉格朗日乘子问题

$$\mathcal{L}_h^s(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f(\mathbf{x}) + g(\mathbf{y}) + \frac{\hbar}{2} \|\mathbf{Ax} + \mathbf{By} - \mathbf{c}\|_2^2 + \frac{\hbar}{2} \mathbf{u}_2, \quad (2.25)$$

上面式(2.25)的最优解可以通过下面的迭代机制得到

$$\begin{aligned} \mathbf{x}^{(t+1)} &= \arg \min_{\mathbf{x}} f(\mathbf{x}) + \frac{\hbar}{2} \|\mathbf{Ax} + \mathbf{By}^{(t)} - \mathbf{c} + \mathbf{u}^{(t)}\|_2^2 \\ \mathbf{y}^{(t+1)} &= \arg \min_{\mathbf{y}} g(\mathbf{y}) + \frac{\hbar}{2} \|\mathbf{Ax}^{(t+1)} + \mathbf{By} - \mathbf{c} + \mathbf{u}^{(t)}\|_2^2, \\ \mathbf{u}^{(t+1)} &= \mathbf{u}^{(t)} + \mathbf{Ax}^{(t+1)} + \mathbf{By}^{(t+1)} - \mathbf{c}, \end{aligned}$$

该迭代算法的停止条件将考虑原问题的残差项 $\epsilon_{pri}^{(t+1)}$ 和对偶问题的残差项 $\epsilon_{dual}^{(t+1)}$, 分别由 $\epsilon_{pri}^{(t+1)} = \mathbf{Ax}^{(t+1)} + \mathbf{By}^{(t+1)} - \mathbf{c}$ 和 $\epsilon_{dual}^{(t+1)} = \hbar \mathbf{A}^\top \mathbf{B}(\mathbf{y}^{(t+1)} - \mathbf{y}^{(t)})$ 给定。若下列条件

$$\|\epsilon_{pri}^{(t)}\|_2 \leq e_{pri} = \sqrt{p}e_{abs} + e_{rel} \max(\|\mathbf{Ax}^{(t)}\|_2, \|\mathbf{By}^{(t)}\|_2, \|\mathbf{c}\|_2),$$

$$\epsilon_{dual}^{(t+1)} \leq e_{dual} = \sqrt{n}e_{abs} + e_{rel} \|\mathbf{A}^\top \mathbf{z}^{(t)}\|_2,$$

同时满足, 则迭代终止。其中 p, n 分别为 \mathbf{c} 和 $\mathbf{A}^\top \mathbf{z}^{(t)}$ 的长度。 e_{abs}, e_{rel} 已知且根据经验给定。

基于 ADMM 算法在类似问题上的优越性, 下文将给出基于 ADMM 算法的 SVCM-BPBQR、SVCM-BPBCQR、SPLVCM-BPBQR 估计实现。有关该算法的起源和发展, 可参阅 Hestenes(1969), Boyd 等(2011)。

3 空间变系数分位回归模型估计

3.1 空间变系数模型

假设数据 $\{(Y_i, \mathbf{X}_i, \mathbf{S}_i)\}_{i=1}^n = \{Y_i, X_{i0}, X_{i1}, \dots, X_{ip}, S_{i1}, S_{i2}\}_{i=1}^n$ 是随机产生与总体 $(Y, \mathbf{X}, \mathbf{S})$ 联合分布的 n 个独立样本。其中 $Y \in \mathbb{R}^1$ 是感兴趣的响应变量， $\mathbf{X} = (X_0 = 1, X_1, \dots, X_p)^\top \in \mathbb{R}^{p+1}$ 为模型中包含常数项的 $p + 1$ 维的解释变量。 $\mathbf{S} = (S_1, S_2)^\top \in \mathbb{R}^2$ 是总体 (Y, \mathbf{X}) 在任意形状有界区域 $\Omega \subseteq \mathbb{R}^2$ 中数据的观测位置。空间变系数模型(SVCMs)可以表示为

$$\begin{aligned} Y_i &= \mathbf{X}_i^\top \boldsymbol{\beta}(\mathbf{S}_i) + \varepsilon_i \\ &= \mathbf{X}_i^\top \boldsymbol{\beta}(S_{i1}, S_{i2}) + \varepsilon_i, \end{aligned} \quad (3.1)$$

其中， $\boldsymbol{\beta}(\cdot) = \boldsymbol{\beta}(\cdot, \cdot) = (\beta_0(\cdot, \cdot), \beta_1(\cdot, \cdot), \dots, \beta_p(\cdot, \cdot))^\top$ 为二元空间变系数函数向量， $\varepsilon_i (i = 1, 2, \dots, n)$ 是随机误差项。需要说明的是，经典的均值回归常假定误差项 ε 服从正态分布，分位回归则不对 ε 的分布形式做任何假定，只假定误差项的100 τ %分位数为0。这样便可以得到模型参数的稳健估计。

3.2 二元惩罚样条 QR 估计

3.2.1 估计

分位回归模型的目标，就是估计感兴趣变量 Y 的 $\tau (0 < \tau < 1)$ 阶条件分位数。一般地，对于一个随机向量 $(Y, \mathbf{X}, \mathbf{S})$ ，在给定 $\mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s}$ 时，记随机变量 Y 的条件累计分布函数 $F_{Y|\mathbf{X}=\mathbf{x}, \mathbf{S}=\mathbf{s}}(y|\mathbf{x}, \mathbf{s})$ ，则条件随机变量 $Y|\mathbf{X}, \mathbf{S}$ 的 τ 阶条件分位数定义为

$$Q_\tau(Y|\mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s}) = \arg \inf \{y \in \mathbb{R} : F_{Y|\mathbf{X}=\mathbf{x}, \mathbf{S}=\mathbf{s}}(y|\mathbf{x}, \mathbf{s}) \geq \tau\}. \quad (3.2)$$

记 $\rho_\tau(u) = u[\tau - I(u < 0)]$ 为检验函数，在分位回归中，常常用来表示 Y 的 τ 阶条件分位数的损失函数。根据检验函数，利用下面式子就可得到 Y 的 τ 阶条件分位数的估计

$$\mathcal{Q}_\tau(\mathbf{x}, \mathbf{s}) = \arg \min_{a \in \mathbb{R}} E \{ \rho_\tau(Y - a) | \mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s} \}. \quad (3.3)$$

式(3.3)未考虑用其他变量解释 Y 。现在重新考虑本文的感兴趣的 SVCMSs, 即模型(3.1), 与之对应 $\mathcal{Q}_\tau(\mathbf{X}, \mathbf{S}) = c_\tau + \sum_{j=0}^p \mathbf{X}_j^\top \beta_{\tau,j}(\mathbf{S})$, 其中 $c_\tau = F^{-1}(\tau)$ 为 ε 的 $100\tau\%$ 分位数, 假定为 0。理论上如果要得到参数 $\beta(\cdot)$ 的分位回归估计, 则需要求解下面式子

$$\arg \min_{\beta_\tau(\cdot)} E \{ \rho_\tau [Y - \mathbf{X}^\top \beta_\tau(\mathbf{S})] \}, \quad (3.4)$$

其中 $\beta_\tau(\cdot) = \beta_\tau(\cdot, \cdot) = (\beta_{0,\tau}(\cdot, \cdot), \beta_{1,\tau}(\cdot, \cdot), \dots, \beta_{p,\tau}(\cdot, \cdot))^\top$ 。若已知样本 $\{(Y_i, \mathbf{X}_i, \mathbf{S}_i)\}_{i=1}^n$, 则公式(3.4)存在其经验形式, 将期望用平均值近似, 可以表示为

$$\arg \min_{\beta_\tau(\cdot)} \sum_{i=1}^n \rho_\tau \{ Y_i - \mathbf{X}_i^\top \beta_\tau(\mathbf{S}_i) \}, \quad (3.5)$$

一般称公式(3.4)或公式(3.5)为空间变系数分位回归模型。对于模型中的待估参数 $\beta_{j,\tau}(\mathbf{s}), j = 0, 1, \dots, p$, 利用公式(2.22)可以将其近似为

$$\beta_{j,\tau}(\mathbf{s}) \approx \boldsymbol{\xi}_{j,\tau}^\top \mathbf{B}_j(\mathbf{s}), \quad j = 0, 1, \dots, p, \quad (3.6)$$

为了简单化描述, 下文中省略参数的分位数标记 τ 。进一步, 公式(3.5)可改写为

$$\arg \min_{\boldsymbol{\xi}_0, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_p} \sum_{i=1}^n \rho_\tau \left\{ Y_i - \sum_{j=0}^p X_{ij} \boldsymbol{\xi}_j^\top \mathbf{B}_j(\mathbf{S}_i) \right\}. \quad (3.7)$$

如果直接求解公式(3.7), 可能会得到欠光滑的二元函数, 无法正确表达数据所隐含的实际情况, 究其原因有二: 一是没有考虑光滑约束(即两个相邻的三角形边界的光滑性), 二是没有考虑观测的稀疏性和不规则性(稀疏性体现在实际研究中获得的数据往往是空间区域的某些位置, 而不是全部)。对于后者, 可以对公式(3.7)添加一个粗糙惩罚来解决, 可称为惩罚样条近似法。具体形式如下:

$$\arg \min_{\boldsymbol{\xi}_0, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_p} \sum_{i=1}^n \rho_\tau \left\{ Y_i - \sum_{j=0}^p X_{ij} \boldsymbol{\xi}_j^\top \mathbf{B}_j(\mathbf{S}_i) \right\} + \frac{1}{2} \sum_{j=0}^p \lambda_j \mathcal{J}(\beta_j), \quad (3.8)$$

惩罚项具体形式如下, 对于任意的二元函数 $g(\mathbf{s}) = g(s_1, s_2)$ 有

$$\mathcal{J}(g) = \int \int_{\Omega} \{(\nabla_{s_1}^2 g)^2 + (\nabla_{s_2}^2 g)^2 + 2(\nabla_{s_1}^1 \nabla_{s_2}^1 g)\} ds_1 ds_2,$$

其中, $\nabla_{s_k}^{\zeta} g(\mathbf{s})$ 为 $g(\mathbf{s})$ 在区域 Ω 上任一点 (s_1, s_2) 方向 s_k , $k = 1, 2$ 上的 ζ 阶导数, λ_j , $j = 0, 1, \dots, p$ 为相应的非负光滑参数。对于式(3.8)中的惩罚项, 亦可以利用二元样条近似方法处理, 即

$$\mathcal{J}(\beta_j) \approx \mathcal{J}(\boldsymbol{\xi}_j^{\top} \mathbf{B}_j(\mathbf{S}_i)) = \boldsymbol{\xi}_j^{\top} \mathbf{P}_j \boldsymbol{\xi}_j, \quad (3.9)$$

其中, \mathbf{P}_j 对应 β_j 的分块对角惩罚矩阵。其次, 再考虑三角形邻边上的光滑性, 公式(3.8)又可以改写为

$$\begin{aligned} \arg \min_{\boldsymbol{\xi}_0, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_p} \sum_{i=1}^n \rho_{\tau} \left\{ Y_i - \sum_{j=0}^p X_{ij} \boldsymbol{\xi}_j^{\top} \mathbf{B}_j(\mathbf{S}_i) \right\} + \frac{1}{2} \sum_{j=0}^p \lambda_j \boldsymbol{\xi}_j^{\top} \mathbf{P}_j \boldsymbol{\xi}_j, \\ \text{s.t. } \mathbf{H}_j \boldsymbol{\xi}_j = \mathbf{0}. \end{aligned} \quad (3.10)$$

对于不同协变量对应的二元系数函数近似问题, 可以采用不同的三角剖分、Bernstein 基多项式展开阶数、光滑约束矩阵以及分块对角惩罚矩阵。这里为了方便理论陈述和模拟说明, 将上述条件统一。假设三角剖分 $T_j \equiv T$, 共有 N_T 个三角形, $\mathbf{B}_j(\mathbf{S}_i) \equiv \mathbf{B}(\mathbf{S}_i) = \{B_m, m \in \mathcal{M}\}^{\top}$, $|\mathcal{M}| = N_T(d+1)(d+2)/2$, $\mathbf{H}_j \equiv \mathbf{H}$, $\mathbf{P}_j \equiv \mathbf{P}$ 。除此之外, 为了式(3.10)变为无约束最优化问题, 采用 QR 矩阵分解处理 \mathbf{H} (秩为 ϱ), 详见 Kim & Wang(2021)。处理之后, 使得某些 $\boldsymbol{\xi}_j = \mathbf{Q}^* \boldsymbol{\xi}_j^*$ 能够保证光滑约束 $\mathbf{H} \boldsymbol{\xi}_j = \mathbf{0}$, 其中 \mathbf{Q}^* 来自

$$\mathbf{H}^{\top} = \mathbf{Q}\mathbf{R} = \left(\begin{array}{cc} \mathbf{Q}^{\varrho} & \mathbf{Q}^* \end{array} \right) \left(\begin{array}{c} \mathbf{R}_1 \\ \mathbf{R}_2 \end{array} \right), \quad (3.11)$$

其中 \mathbf{Q} 为正交矩阵, \mathbf{R} 为上三角矩阵, \mathbf{Q}^{ϱ} 是 \mathbf{Q} 的前 ϱ 列, $\mathbf{R}_2 = \mathbf{0}$ 。经过处理, 最终约束最优化问题(3.10)就可以转化为下面无约束最优化问题

$$\arg \min_{\boldsymbol{\xi}_0^*, \boldsymbol{\xi}_1^*, \dots, \boldsymbol{\xi}_p^*} \sum_{i=1}^n \rho_{\tau} \left\{ Y_i - \sum_{j=0}^p X_{ij} (\mathbf{Q}^* \boldsymbol{\xi}_j^*)^{\top} \mathbf{B}(\mathbf{S}_i) \right\} + \frac{1}{2} \sum_{j=0}^p \lambda_j (\mathbf{Q}^* \boldsymbol{\xi}_j^*)^{\top} \mathbf{P} \mathbf{Q}^* \boldsymbol{\xi}_j^*, \quad (3.12)$$

式(3.12)称为空间变系数复合分位回归模型的二元惩罚样条估计, 记为 SVCM-BPBQR 估计。假设 $\boldsymbol{\xi}^* = (\boldsymbol{\xi}_0^*, \boldsymbol{\xi}_1^*, \dots, \boldsymbol{\xi}_p^*)^{\top}$, 且 $\hat{\boldsymbol{\xi}}^* = (\hat{\boldsymbol{\xi}}_0^*, \hat{\boldsymbol{\xi}}_1^*, \dots, \hat{\boldsymbol{\xi}}_p^*)^{\top}$ 为公式

(3.12)的解, 则对于 $\hat{\xi}_j = \{\hat{\xi}_{j,m}, m \in \mathcal{M}\}^\top = \mathbf{Q}^* \hat{\xi}_j^*$, 可得到二元系数函数 $\beta_j(\mathbf{s})$ 的 SVCM-BPBQR 估计量, 表示为

$$\hat{\beta}_j(\mathbf{s}) = \hat{\xi}_j^\top \mathbf{B}(\mathbf{s}) = \sum_{m \in \mathcal{M}} \hat{\xi}_{j,m} B_m(\mathbf{s}), \quad j = 0, 1, \dots, p. \quad (3.13)$$

3.2.2 光滑参数的选取

对于给定的有限样本数据集, 选择恰当的惩罚参数是非常重要的。选择光滑参数 $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \dots, \lambda_p)^\top$, 一方面要考虑数据与模型之间的拟合优度, 还要考虑二元系数函数 $\beta_j(\mathbf{s})$ 的有效性, 一般可以利用交叉验证(CV)和广义交叉验证(GCV)方法来选取。其他方法还有 Fan & Li(2001)使用了五重交叉验证和广义交叉验证选择器。Wang 等(2007)在其研究中提出使用贝叶斯信息准则(BIC)。Chen & Chen(2008)提出了一种延拓贝叶斯信息准则(EBIC)等。考虑到 CV 会有数据欠光滑风险, 本文以 GCV 方法为例选取光滑参数, 即在网格上最小化下面 GCV 准则:

$$GCV(\boldsymbol{\lambda}) = \frac{n^{-1} \|\mathbf{Y} - \mathbf{L}(\boldsymbol{\lambda})\mathbf{Y}\|^2}{\{n^{-1} \text{tr}(\mathbf{I} - \mathbf{L}(\boldsymbol{\lambda}))\}^2}, \quad (3.14)$$

其中 $\mathbf{L}(\boldsymbol{\lambda}) = \mathbf{X}^*(\mathbf{X}^{*\top} \mathbf{X}^* + \boldsymbol{\lambda} \otimes \mathbf{Q}^{*\top} \mathbf{P} \mathbf{Q}^*)^{-1} \mathbf{X}^{*\top}$, $\text{tr}(\cdot)$ 表示矩阵的迹。

3.3 估计量的理论性质

本节研究 SVCM-BPBQR 的渐近性质。类似于 Andriyana 等(2014)的假设, 这里考虑两种情况下的渐近性质:

情形 1 设 $\beta_j(\cdot), j = 0, 1, \dots, p$ 均来自一般的 Sobolev 空间。则公式(2.22)的近似会造成偏倚。若要偏倚减小, 需要让 Benstein 基多项式的个数 $|\mathcal{M}_j| \rightarrow \infty$ 。即需要使三角形个数 $N_j \rightarrow \infty$ 。

情形 2 设 $\beta_j(\cdot), j = 0, 1, \dots, p$ 是 Benstein 基多项式 $\{B_m, m \in \mathcal{M}\}$ 张成的空间中的元素。则公式(2.22)的近似变成等式, 即不存在近似偏倚。

3.3.1 情形 1 的渐近性质

为了方便理论说明, 首先引入记号。令 $\mathbf{X}^* = (\mathbf{Q}^* \mathbf{X}_0^*, \mathbf{Q}^* \mathbf{X}_1^*, \dots, \mathbf{Q}^* \mathbf{X}_p^*)^\top$, $\mathbf{X}_j^* = (X_{1j} \mathbf{B}(\mathbf{S}_1), X_{2j} \mathbf{B}(\mathbf{S}_2), \dots, X_{nj} \mathbf{B}(\mathbf{S}_n))$, $\boldsymbol{\xi}^* = (\boldsymbol{\xi}_0^*, \boldsymbol{\xi}_1^*, \dots, \boldsymbol{\xi}_p^*)^\top$ 。记 $\|h\|_{\ell_2(\Omega)}^2 = \int_{\mathbf{x} \in \Omega} h^2(\mathbf{s}) dx_1 dx_2$ 为任意的二元函数 $h(\mathbf{x}) = h(x_1, x_2)$ 在闭区域 Ω 上的 ℓ_2 范数, $\|h\|_{\infty, \Omega} = \sup_{\mathbf{x} \in \Omega} |h(\mathbf{x})|$ 为无穷范数。

$$|h|_{\delta, \infty} = \max_{i+j=\delta} \|\nabla_{x_1}^i \nabla_{x_2}^j h(x_1, x_2)\|_{\infty},$$

$$W^{d+1, \infty}(\Omega) = \{h : |h|_{k, \infty, \Omega} < \infty, 0 \leq k \leq d+1\},$$

为标准的 Sobolev 空间。对于两个非负实值序列 a_n, b_n , 若存在常量 C_1, C_2 且 $0 < C_1 < C_2 < \infty$ 使得对于所有的 n , 有 $C_1 \leq a_n/b_n \leq C_2$ 成立, 则记为 $a_n \asymp b_n$ 。

其次引入一些假设, 使其保证理论性质能够成立:

(A1) 令 $f_{\mathbf{S}}(s)$ 为二维随机向量 $\mathbf{S} = (S_1, S_2)^\top$ 的联合密度函数, 并且 $f_{\mathbf{S}}(s)$ 连续、有上界, 有 $f_{\mathbf{S}}(s) > 0$ 成立。

(A2) 对于任意的 i , 随机向量 \mathbf{X}_i 的每个分量依概率有界。并且令 $E(\mathbf{X}^\top \mathbf{X} | \mathbf{S} = \mathbf{s}) = \mathbf{D}(\mathbf{s})$ 在任一点 \mathbf{s}_0 的领域内正定且连续, $\mathbf{D}(\mathbf{s})$ 的特征值 $\phi_0(\mathbf{s}) \leq \phi_1(\mathbf{s}) \leq \dots \leq \phi_p(\mathbf{s})$ 在远离 0 和 ∞ 处一致有界。

(A3) 误差 ε 有累计分布函数 $F_\varepsilon(\cdot)$ 和密度函数 $f_\varepsilon(\cdot)$ 。对于每一个 p 维向量 \mathbf{u} ,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \int_0^{u_0 + \mathbf{X}_i^* \mathbf{u}} \sqrt{n} [F_\varepsilon(a + t/\sqrt{n}) - F_\varepsilon(a)] dt \\ = \frac{1}{2} f_\varepsilon(a) (u_0, \mathbf{u}^\top) \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} (u_0, \mathbf{u}^\top)^\top. \end{aligned}$$

(A4) 系数函数 $\beta_j(\cdot), j = 0, 1, \dots, p$ 都来自于标准的 Sobolev 空间 $W^{d+1, \infty}(\Omega)$ 。

(A5) 三角剖分 T 是 π 拟均匀的。存在一个正常量 π , 使得 $|T|/r_T \leq \pi$ 。

(A6) 三角剖分 T 满足 $|T| \rightarrow 0, |T|^{-2} \log n/n \rightarrow 0$ 。

(A7) 对于任意 $g \in \mathbb{S}_d^y, \Delta \in T$, 存在一个与 g 和 Δ 独立的正常数 \mathbb{F}_1 , 并假设在三角形 $\Delta \in T$ 中最大观测数量 \mathbb{F}_2 , 使得对于所有的 $\Delta \in T$, 有

$$\mathbb{F}_1 \|g\|_{\infty, \Delta} \leq \{\sum_{\mathbf{S}_i \in \Delta, i=1, \dots, n} g(\mathbf{S}_i)^2\}^{1/2} \leq \mathbb{F}_2 \|g\|_{\infty, \Delta},$$

成立。其中, $\|g\|_{\infty, \Delta}$ 为 g 在三角形 Δ 上的无穷范数, 常量 \mathbb{F}_1 和 \mathbb{F}_2 满足 $\mathbb{F}_1/\mathbb{F}_2 = O(1)$ 。

条件(A1)和(A2)是非参数回归模型中的普遍假设, 如 Guo 等(2012), Mu 等(2018)。条件(A3)是分位回归文献中关于误差项的有关假设(Koenker, 2005)。条

件(A4)是对二元系数函数的限定和假设。条件(A5)限定空间区域三角剖分后的各个三角形尽可能均匀大小。条件(A7)保证了离散样条的存在性，并且限制了某个三角形上有太多的观测发生，(A5)和(A7)均来自 Mu 等(2018)。

定理 1 假设条件(A1-A7)成立，令 $\lambda_{\max} = \max\{\lambda_j, j = 0, 1, \dots, p\}$ ，并且假设 $\lambda_{\max} = o(n^{d/2(d+2)})$ ，如果三角剖分 T 的三角形的个数 N 和样本个数 n 满足关系 $N \asymp O(n^{1/(d+2)})$ ，则在一定的正则性条件下，对于任意的 $j = 0, 1, \dots, p$ ，当 $\lambda_{\max}/\sqrt{n} \rightarrow 0, n \rightarrow \infty$ ，SVCM-BPBQR 估计量 $\hat{\beta}_j(\cdot, \cdot)$ 是相合的，并且

$$\left\| \hat{\beta}_j(\cdot, \cdot) - \beta_j(\cdot, \cdot) \right\|_{\ell_2(\Omega)}^2 = O_p \left\{ n^{-(d+1)/(d+2)} \right\},$$

成立，并且这是 Stone(1982)中给出的最优收敛速率。

3.3.2 情形 2 的渐近性质

本小节建立估计量式(3.13)渐近分布。记 ξ_0^* 为 ξ^* 的真实值。在情形 2 的假设下，式(2.22)没有逼近偏差，因此这里假设三角剖分 T_j 中三角形个数 N_j 是固定的，则 $|\mathcal{M}_j|$ 是有限的。在情形 2 下，令 Y_1, Y_2, \dots, Y_n 是独立的随机变量，假定条件 τ 分位函数为 $q_\tau(Y_i | \mathbf{X}_i, \mathbf{S}_i) = q_\tau(Y_i | \mathbf{X}_i^*) = \mathbf{X}_i^{*\top} \boldsymbol{\xi}_\tau^*$ 。首先，引入假设：

(B1) 存在正定矩阵 \mathcal{D} 和 \mathcal{D}_1 ，使得

$$(I) \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i^{*\top} \mathbf{X}_i^* = \mathcal{D}$$

$$(II) \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f_{\epsilon, i}(\epsilon_i(\tau)) \mathbf{X}_i^{*\top} \mathbf{X}_i^* = \mathcal{D}_1(\tau)$$

(B2) $\max_{i=1, 2, \dots, n} \|\mathbf{X}_i^*\|_2 / \sqrt{n} \rightarrow 0$

条件(B1-B2)均来自 Koenker(2005)。

定理 2 若条件(A1-A3)、(A5-A7)、(B1-B2)成立。且存在 $\bar{\lambda} > 0$ ，使得当 $n \rightarrow \infty$ 时， $\lambda_{\max}/\sqrt{n} \rightarrow \bar{\lambda}$ ，则

$$\sqrt{n}(\hat{\boldsymbol{\xi}}^* - \boldsymbol{\xi}^*) \xrightarrow{d} \mathcal{N}(-\mathcal{D}_1^{-1} \nabla_{\mathcal{Q}_3}^1, \tau(1-\tau) \mathcal{D}_1^{-1} \mathcal{D} (\mathcal{D}_1^{-1})^\top),$$

其中 $\mathcal{Q}_3 = 1/2 \bar{\lambda} \mathbf{K} \sum_{j=0}^p \sum_{m \in \mathcal{M}} \theta_{jm} \text{sgn}(\xi_{jm}^*) |\xi_{jm}^*|$ 且 $\mathbf{K} = \mathbf{Q}^{*\top} \mathbf{P} \mathbf{Q}^*$ ， $\nabla_{\mathcal{Q}_3}^1$ 为 \mathcal{Q}_3 关于 $\boldsymbol{\theta} = \sqrt{n}(\boldsymbol{\xi}^* - \boldsymbol{\xi}_0^*)$ 的梯度向量。

证明 令

$$\mathcal{L}(\boldsymbol{\xi}^*) = \sum_{i=1}^n \rho_\tau \left\{ Y_i - \sum_{j=0}^p X_{ij} (\mathbf{Q}^* \boldsymbol{\xi}_j^*)^\top \mathbf{B}(\mathbf{S}_i) \right\} + \frac{1}{2} \sum_{j=0}^p \lambda_j (\mathbf{Q}^* \boldsymbol{\xi}_j^*)^\top \mathbf{P} \mathbf{Q}^* \boldsymbol{\xi}_j^*,$$

并考虑目标函数

$$Q_n(\boldsymbol{\theta}) = \mathcal{L}\left(\boldsymbol{\xi}^* + \frac{\boldsymbol{\theta}}{\sqrt{n}}\right) - \mathcal{L}(\boldsymbol{\xi}^*),$$

其中 $\boldsymbol{\theta} = (\boldsymbol{\theta}_0^\top, \boldsymbol{\theta}_1^\top, \dots, \boldsymbol{\theta}_p^\top)^\top \in \mathbb{R}^{|\mathcal{M}| \times (p+1)}$, $\boldsymbol{\theta}_j = (\theta_{j1}, \theta_{j2}, \dots, \theta_{j|\mathcal{M}|}) \in \mathbb{R}^{|\mathcal{M}|}$,

$\theta_{jm} = \sqrt{n}(\xi_{jm}^* - \xi_0^*)$ 。因为 $\hat{\boldsymbol{\xi}}^* = \arg \min_{\boldsymbol{\xi}^*} \mathcal{L}(\boldsymbol{\xi}^*)$ 及其凸性,

可以得到有 $\hat{\boldsymbol{\theta}} = \sqrt{n}(\hat{\boldsymbol{\xi}}^* - \boldsymbol{\xi}^*) = \arg \min_{\boldsymbol{\theta}} Q_n(\boldsymbol{\theta})$ 成立, 即 $Q_n(\boldsymbol{\theta})$ 的极限行为决定了 $\hat{\boldsymbol{\theta}}$ 的

渐近分布。为了简化记号令 $\mathbf{K} = \mathbf{Q}^{*\top} \mathbf{P} \mathbf{Q}^*$, 进一步有

$$\begin{aligned} Q_n(\boldsymbol{\theta}) &= \sum_{i=1}^n \left\{ \rho_\tau \left(Y_i - \mathbf{X}_i^{*\top} \boldsymbol{\xi}^* - \mathbf{X}_i^{*\top} \frac{\boldsymbol{\theta}}{\sqrt{n}} \right) - \rho_\tau \left(Y_i - \mathbf{X}_i^{*\top} \boldsymbol{\xi}^* \right) \right\} \\ &\quad + \frac{1}{2} \sum_{j=0}^p \lambda_j \mathbf{K} \left\{ \left\| \boldsymbol{\xi}_j^* + \frac{\boldsymbol{\theta}_j}{\sqrt{n}} \right\|_2^2 - \|\boldsymbol{\xi}_j^*\|_2^2 \right\} \end{aligned}$$

利用 Knight(1998)不等式,

$$\rho_\tau(u-v) - \rho_\tau(u) = -v\psi_\tau(u) + \int_0^v \{I(u \leq s) - I(u \leq 0)\} ds,$$

其中 $\psi_\tau(u) = \tau - I(u \leq 0)$, 进一步可以得到

$$\begin{aligned} Q_n(\boldsymbol{\theta}) &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{X}_i^{*\top} \boldsymbol{\theta} \psi_\tau(Y_i - \mathbf{X}_i^{*\top} \boldsymbol{\xi}^*) \\ &\quad + \int_0^{\frac{\mathbf{X}_i^{*\top} \boldsymbol{\theta}}{\sqrt{n}}} \{I(Y_i - \mathbf{X}_i^{*\top} \boldsymbol{\xi}^* \leq s) - I(Y_i - \mathbf{X}_i^{*\top} \boldsymbol{\xi}^* \leq 0)\} ds \\ &\quad + \frac{1}{2} \sum_{j=0}^p \lambda_j \mathbf{K} \left\{ \left\| \boldsymbol{\xi}_j^* + \frac{\boldsymbol{\theta}_j}{\sqrt{n}} \right\|_2^2 - \|\boldsymbol{\xi}_j^*\|_2^2 \right\} \\ &= Q_{1n}(\boldsymbol{\theta}) + Q_{2n}(\boldsymbol{\theta}) + Q_{3n}(\boldsymbol{\theta}). \end{aligned}$$

根据 Cramer-Wald 定理和 Lindeberg-Feller 中心极限定理, 利用假设, 有 $Q_{1n}(\boldsymbol{\theta}) \xrightarrow{d} -\boldsymbol{\theta} \mathbf{W}$, 其中 $\mathbf{W} \xrightarrow{d} \mathcal{N}(\mathbf{0}, \tau(1-\tau)\mathbf{D})$ 。

对于 $Q_{2n}(\boldsymbol{\theta})$, 令

$$Q_{2ni}(\boldsymbol{\theta}) = \int_0^{\mathbf{X}_i^{*\top} \boldsymbol{\theta} / \sqrt{n}} \{I(Y_i - \mathbf{X}_i^{*\top} \boldsymbol{\xi}^* \leq s) - I(Y_i - \mathbf{X}_i^{*\top} \boldsymbol{\xi}^* \leq 0)\} ds,$$

则有 $Q_{2n}(\boldsymbol{\theta}) = \sum_{i=1}^n E Q_{2ni}(\boldsymbol{\theta}) + \sum_{i=1}^n \{Q_{2ni}(\boldsymbol{\theta}) - E Q_{2ni}(\boldsymbol{\theta})\}$, 其中

$$\begin{aligned}
\sum_{i=1}^n E\mathcal{Q}_{2ni}(\boldsymbol{\theta}) &= \sum_{i=1}^n \int_0^{\frac{\mathbf{X}_i^{*\top}\boldsymbol{\theta}}{\sqrt{n}}} \{F_{\epsilon,i}(\epsilon_i + s) - F_{\epsilon,i}(\epsilon_i)\} ds \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^{\mathbf{X}_i^{*\top}\boldsymbol{\theta}} \left\{ F_{\epsilon,i}\left(\epsilon_i + \frac{t}{\sqrt{n}}\right) - F_{\epsilon,i}(\epsilon_i) \right\} dt \\
&= \frac{1}{n} \sum_{i=1}^n \int_0^{\mathbf{X}_i^{*\top}\boldsymbol{\theta}} \sqrt{n} \left\{ F_{\epsilon,i}\left(\epsilon_i + \frac{t}{\sqrt{n}}\right) - F_{\epsilon,i}(\epsilon_i) \right\} dt \\
&= \frac{1}{n} \sum_{i=1}^n \int_0^{\mathbf{X}_i^{*\top}\boldsymbol{\theta}} f_{\epsilon,i}(\epsilon_i) t dt (1 + o(1)) \\
&= \frac{1}{2n} \sum_{i=1}^n f_{\epsilon,i}(\epsilon_i) (\mathbf{X}_i^{*\top}\boldsymbol{\theta})^2 (1 + o(1)) \\
&= \frac{1}{2} \boldsymbol{\theta}^\top \mathcal{D}_{1,i} \boldsymbol{\theta} \xrightarrow{p} \frac{1}{2} \boldsymbol{\theta}^\top \mathcal{D}_1 \boldsymbol{\theta},
\end{aligned}$$

考虑方差

$$\begin{aligned}
&Var\{\mathcal{Q}_{2n}(\boldsymbol{\theta})\} \\
&= \sum_{i=1}^n Var\left\{ \int_0^{\mathbf{X}_i^{*\top}\boldsymbol{\theta}/\sqrt{n}} \left\{ I\left(Y_i - \mathbf{X}_i^{*\top}\boldsymbol{\xi}^* \leq s\right) - I\left(Y_i - \mathbf{X}_i^{*\top}\boldsymbol{\xi}^* \leq 0\right) \right\} ds \right\} \\
&\leq \sum_{i=1}^n E\left\{ \int_0^{\mathbf{X}_i^{*\top}\boldsymbol{\theta}/\sqrt{n}} \left\{ I\left(Y_i - \mathbf{X}_i^{*\top}\boldsymbol{\xi}^* \leq s\right) - I\left(Y_i - \mathbf{X}_i^{*\top}\boldsymbol{\xi}^* \leq 0\right) \right\} ds \right\}^2 \\
&\leq \frac{1}{\sqrt{n}} \max_{i=1,\dots,n} |\mathbf{X}_i^{*\top}\boldsymbol{\xi}^*| E\{\mathcal{Q}_{2n}(\boldsymbol{\theta})\} \\
&\leq \frac{1}{\sqrt{n}} \max_{i=1,\dots,n} \|\mathbf{X}_i^{*\top}\|_2 \|\boldsymbol{\xi}^*\|_2 \sum_{i=1}^n E\{\mathcal{Q}_{2ni}(\boldsymbol{\theta})\}
\end{aligned}$$

根据假设 (B2)，当 $n \rightarrow \infty$ 时，有 $1/\sqrt{n} \max_{i=1,\dots,n} \|\mathbf{X}_i^{*\top}\|_2 \|\boldsymbol{\xi}^*\|_2 \rightarrow 0$ ，而 $0 < \sum_{i=1}^n E\{\mathcal{Q}_{2ni}(\boldsymbol{\theta})\} < \infty$ ，故 $Var\{\mathcal{Q}_{2n}(\boldsymbol{\theta})\} = o_p(1)$ 。

对于 $\mathcal{Q}_{3n}(\boldsymbol{\theta})$ ，类似于 Andriyana 等(2014)，首先重新表述公式(3.12)的惩罚项，写为

$$\frac{1}{2} \sum_{j=0}^p \lambda_j (\mathbf{Q}^* \boldsymbol{\xi}_j^*)^\top \mathbf{P} \mathbf{Q}^* \boldsymbol{\xi}_j^* = \frac{1}{2} \sum_{j=0}^p \lambda_j \mathbf{K} \sum_{m \in \mathcal{M}} |\xi_{jm}^*|^2,$$

则

$$\mathcal{Q}_{3n}(\boldsymbol{\theta}) = \frac{1}{2} \sum_{j=0}^p \lambda_j \mathbf{K} \sum_{m \in \mathcal{M}} \left(\left| \xi_{jm}^* + \frac{1}{\sqrt{n}} \theta_{jm} \right|^2 - |\xi_{jm}^*|^2 \right),$$

回忆记号 $\lambda_{\max} = \{\lambda_j, j = 0, 1, \dots, p\}$ ，如果当 $n \rightarrow \infty$ 时有 $\lambda_{\max}/\sqrt{n} \rightarrow \bar{\lambda} \geq 0$ ，则

$$\mathcal{Q}_{3n}(\boldsymbol{\theta}) \rightarrow \frac{1}{2} \bar{\lambda} \mathbf{K} \sum_{j=0}^p \sum_{m \in \mathcal{M}} \theta_{j,m} \text{sgn}(\xi_{j,m}^*) |\xi_{j,m}^*| = \mathcal{Q}_3,$$

进一步, 令 $\nabla_{Q_3}^1$ 为 Q_3 关于 $\theta = \sqrt{n}(\xi^* - \xi_0^*)$ 的梯度向量, 通过简单运算可知 $\nabla_{Q_3}^1$ 独立于 θ 。

综上, 有 $Q_n(\theta) \xrightarrow{p} Q_0(\theta) = -\theta^\top \mathbf{W} + \frac{1}{2}\theta^\top \mathbf{D}_1 \theta + Q_3$ 。根据目标函数 $Q_n(\theta)$ 的凸性, 存在且惟一的

$$\hat{\theta} = \sqrt{n}(\hat{\xi}^* - \xi^*) \xrightarrow{p} \mathbf{D}_1^{-1} \mathbf{W} + \nabla_{Q_3}^1$$

进一步, 可得结论 $\sqrt{n}(\hat{\xi}^* - \xi^*) \xrightarrow{d} \mathcal{N}(-\mathbf{D}_1^{-1} \nabla_{Q_3}^1, \tau(1-\tau)\mathbf{D}_1^{-1} \mathbf{D} (\mathbf{D}_1^{-1})^\top)$ 。证毕。

上述定理 2 的渐近性质不严格假定误差分布, 下面假设误差分布是独立同分布的, 并不加证明的给出如下推论 2。

推论 2 在独立同分布误差模型以及定理 2 的条件下,

$$\sqrt{n}(\hat{\xi}^* - \xi^*) \xrightarrow{d} \mathcal{N}\left(-\mathbf{D}_1^{-1} \nabla_{Q_3}^1, \frac{\tau(1-\tau)\mathbf{D}}{f_{\epsilon_i}^2(\epsilon_i)}\right).$$

3.4 SVCMQR-ADMM 算法

本节研究如何求解 SVCM-BPBQR 估计量, 下面是其计算过程和算法实现。本文提出的估计量核心公式(3.12)是经典的凸优化问题。对此, 有很多成熟的最优化算法可以利用, 诸如单纯形法(Koenker, 2005)和内点算法(Koenker & Park, 1996)等。然而, 公式(3.12)的具有一定的特殊结构, 可以将其分解为两个子凸优化问题: 非光滑分位数损失函数和 L_2 型惩罚(Kong, 2015)。对于该类型的最优化目标, 可以利用更有效的交替方向乘数法(ADMM), 它将一个给定的可加可分凸优化问题分解为若干次凸优化问题。Pietrosanu(2020)重新表述了 LASSO 惩罚分位数回归, 并利用交替方向乘子法(ADMM)、最大化-最小化(MM)和坐标下降(CD)算法获得了相应的估计量, 说明了 ADMM 算法在并行性和可伸缩性方面的优势。

基于 ADMM 算法在类似的优化问题上具有一定优越性, 这里给出 SVCM-BPBQR 估计量的估计实现算法, 记为 SVCMQR-ADMM 算法。有关该算法的起源和发展, 可参阅 Hestenes(1969), Gabay & Mercier(1976), Boyd 等(2011)。

令 $\epsilon_i = \epsilon_i(\xi^*) = Y_i - \sum_{j=0}^p X_{ij}(\mathbf{Q}^* \xi_j^*)^\top \mathbf{B}(\mathbf{S}_i)$, $i = 1, 2, \dots, n$ 为分位回归估计的残差项, 记 $\mathcal{J}_\lambda(\xi^*) = 1/2 \sum_{j=0}^p \lambda_j (\mathbf{Q}^* \xi_j^*)^\top \mathbf{P} \mathbf{Q}^* \xi_j^* = 1/2 \lambda \otimes \xi^{*\top} \mathbf{Q}^{*\top} \mathbf{P} \mathbf{Q}^* \xi^*$, 其中 $\lambda = (\lambda_0, \lambda_1, \dots, \lambda_p)^\top$, 下面应用 ADMM 算法将正则分位回归目标函数(3.12)重新表述为下面带约束的凸优化问题:

$$\begin{aligned} \min_{\boldsymbol{\xi}^* \in \mathbb{R}^{|\mathcal{M}| \times (p+1)}} \quad & \sum_{i=1}^n \rho_{\tau}(\epsilon_i) + \mathcal{J}_{\lambda}(\boldsymbol{\xi}^*) \\ \text{subject to} \quad & \mathbf{X}^{*\top} \boldsymbol{\xi}^* + \boldsymbol{\epsilon} = \mathbf{Y} \end{aligned} \quad (3.14)$$

其中 $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$, $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^\top$, $\mathbf{X}^* = (\mathbf{Q}^* \mathbf{X}_0^*, \mathbf{Q}^* \mathbf{X}_1^*, \dots, \mathbf{Q}^* \mathbf{X}_p^*)^\top$, $\mathbf{X}_j^* = (X_{1j} \mathbf{B}(\mathbf{S}_1), X_{2j} \mathbf{B}(\mathbf{S}_2), \dots, X_{nj} \mathbf{B}(\mathbf{S}_n))$, $\boldsymbol{\xi}^* = (\xi_0^*, \xi_1^*, \dots, \xi_p^*)^\top$ 。

现在可以利用 ADMM 算法(Boyd 等, 2011)求解公式(3.14): 假设迭代到第 t 步, 则第 $t+1$ 步迭代过程为

$$\begin{aligned} \boldsymbol{\epsilon}^{(t+1)} &= \arg \min_{\boldsymbol{\epsilon} \in \mathbb{R}^n} \sum_{i=1}^n \rho_{\tau}(\epsilon_i) + \frac{\hbar}{2} \left\| \mathbf{Y} - \boldsymbol{\epsilon} - \mathbf{X}^{*\top} \boldsymbol{\xi}^{*(t)} + \frac{\mathbf{u}^{(t)}}{\hbar} \right\|_2^2 \\ \boldsymbol{\xi}^{*(t+1)} &= \arg \min_{\boldsymbol{\xi}^* \in \mathbb{R}^{|\mathcal{M}| \times (p+1)}} \frac{\hbar}{2} \left\| \mathbf{Y} - \boldsymbol{\epsilon}^{(t+1)} - \mathbf{X}^{*\top} \boldsymbol{\xi}^* + \frac{\mathbf{u}^{(t)}}{\hbar} \right\|_2^2 + \mathcal{J}_{\lambda}(\boldsymbol{\xi}^*) \\ \mathbf{u}^{(t+1)} &= \mathbf{u}^{(t)} + \hbar(\mathbf{Y} - \boldsymbol{\epsilon}^{(t+1)} - \mathbf{X}^{*\top} \boldsymbol{\xi}^{*(t+1)}), \end{aligned} \quad (3.15)$$

其中, μ 缩放后的为拉格朗日乘数, \hbar 是一个惩罚参数。Boyd 等(2011)建议取值 $\hbar = 1.2$ 。定义软门限算子 $\mathcal{S}_{\delta}: \mathbb{R}^m \rightarrow \mathbb{R}^m$ 为 $(\mathcal{S}_{\delta}(\mathbf{v}))_i = (v_i - a)_+ - (-v_i - a)_+$ 。下面利用该算子来简化公式(3.15)中 $\boldsymbol{\epsilon}$ 的更新。即

$$\boldsymbol{\epsilon}^{(t+1)} = \mathcal{S}_{\frac{1}{\hbar}} \left(\mathbf{Y} - \mathbf{X}^{*\top} \boldsymbol{\xi}^{*(t)} + \frac{\mathbf{u}^{(t)}}{\hbar} - (2\boldsymbol{\tau}_{n \times 1} - \mathbf{1}_{n \times 1})/\hbar \right), \quad (3.16)$$

对于 $\boldsymbol{\xi}^*$ 迭代, 可以将其视为一个带有 L_2 惩罚的最小二乘优化问题, 即岭回归问题。所以根据岭回归数值求解方法, 使用下面式子更新 $\boldsymbol{\xi}^*$

$$\boldsymbol{\xi}^{*(t+1)} = \left(\mathbf{X}^{*\top} \mathbf{X}^* + \lambda \otimes \mathbf{Q}^{*\top} \mathbf{P} \mathbf{Q}^* / \hbar \right)^{-1} \mathbf{X}^{*\top} \left(\mathbf{Y} - \boldsymbol{\epsilon}^{(t+1)} + \frac{\mathbf{u}^{(t)}}{\hbar} \right), \quad (3.17)$$

在公式(3.17)中, 注意到 $\mathbf{X}^{*\top} \mathbf{X}^* + \lambda \otimes \mathbf{Q}^{*\top} \mathbf{P} \mathbf{Q}^* / \hbar$ 的逆只需要计算一次, ADMM 迭代过程非常快速且有效(Kong, 2015)。

ADMM 算法利用原残差 $\boldsymbol{\epsilon}_{pri}$ 和对偶残差 $\boldsymbol{\epsilon}_{dual}$ (Boyd, 2011)定义 ADMM 算法迭代的终止条件。在第 k 次迭代, 原残差 $\boldsymbol{\epsilon}_{pri}^{(k)}$ 和对偶残差 $\boldsymbol{\epsilon}_{dual}^{(k)}$ 的计算公式为

$$\begin{aligned} \boldsymbol{\epsilon}_{pri}^{(k)} &= \mathbf{Y} - \mathbf{X}^{*\top} \boldsymbol{\xi}^{*(k)} - \boldsymbol{\epsilon}^{(k)} \\ \boldsymbol{\epsilon}_{dual}^{(k)} &= \hbar \mathbf{X}^{*\top} (\boldsymbol{\xi}^{*(k)} - \boldsymbol{\xi}^{*(k-1)}), \end{aligned} \quad (3.18)$$

相应地, 算法终止条件为

$$\|\boldsymbol{\epsilon}_{pri}^{(k)}\|_2 \leq e_{pri}, \quad \|\boldsymbol{\epsilon}_{dual}^{(k)}\|_2 \leq e_{dual}, \quad (3.19)$$

$$\begin{aligned} e_{pri} &= \sqrt{n}e_{abs} + e_{rel} \max \left\{ \|\mathbf{X}^{*\top} \boldsymbol{\xi}^{*(k)}\|_2^2, \|\boldsymbol{\epsilon}^{(k)}\|_2^2, \|\mathbf{Y}\|_2^2 \right\} \\ e_{dual} &= \sqrt{|\mathcal{M}|} \times pe_{abs} + e_{rel} \|\mathbf{X}^{*\top} \mathbf{u}^{(k)}\|_2^2, \end{aligned} \quad (3.20)$$

其中, e_{abs} 和 e_{rel} 为绝对误差限和相对误差限, 在实际计算中, 可以指定任意小的数, 这取决于对模型拟合精度的要求。本文采取 $e_{abs} = 10^{-4}$, $e_{rel} = 10^{-2}$ (Boyd 等, 2011)。

下面是 ADMM 算法执行过程:

Step1.待估参数初始化。利用惩罚迭代重加权最小二乘法 (Kim & Wang, 2021) 进行参数初始化 $\boldsymbol{\xi}^{*(0)}$, 并令 $\boldsymbol{\epsilon}^{(0)} = \mathbf{Y} - \sum_{j=0}^p \mathbf{X}_j (\mathbf{Q}^* \boldsymbol{\xi}_j^{*(0)})^\top \mathbf{B}(\mathbf{S})$, $\mathbf{u}^{(0)} = \mathbf{0}$;

Step2.迭代。假设已经迭代到第 $(t-1)$ 步, 利用式(3.15)求解 $(\boldsymbol{\xi}^{*(t)}, \boldsymbol{\epsilon}^{(t)}, \mathbf{u}^{(t)})$, 利用式(3.18)计算 $\boldsymbol{\epsilon}_{pri}^{(t)}, \boldsymbol{\epsilon}_{dual}^{(t)}$, 并判断条件式(3.19-3.20)是否成立;

Step3.若成立, 终止算法并输出 $\hat{\boldsymbol{\xi}}^* = \boldsymbol{\xi}^{*(t)}$; 否则返回 Step2, 继续下一次计算直到(3.20)成立或达到算法最大迭代次数。若后者发生, 判定为模型迭代未实现收敛。

3.5 Bootstrap 拟合优度检验

虽然在本文中对空间变系数模型进行了大量的探讨, 但是在实际应用中我们可能对二元空间系数函数是否真的随空间位置 $\mathbf{S} = (S_1, S_2)^\top$ 而发生变化也很感兴趣, 因为这可以让我们得到更加适合于特定数据的研究模型。基于此, 这里考虑以下拟合优度检验问题。为了检验模型(3.1)是否具有特定的参数形式(如线性回归模型), 本文参考 Cai 等(2000)、Mu 等(2018)的拟合优度检验方法, 提出了一种基于参数模型和非参数模型分位残差和(RSQ)比较的 Bootstrap 拟合优度检验。

构造假设检验

$$H_0: \beta_j(\mathbf{s}) = \beta_j(\mathbf{s}; \boldsymbol{\vartheta}) \text{ v.s. } H_1: \beta_j(\mathbf{s}) \neq \beta_j(\mathbf{s}; \boldsymbol{\vartheta}), \quad 0 \leq j \leq p. \quad (3.21)$$

其中, $\beta_k(\mathbf{s}; \boldsymbol{\vartheta})$ 为由一个未知参数向量 $\boldsymbol{\vartheta}$ 给定的函数族。

Cai 等(2000)和 Huang 等(2002)研究了基于比较原假设与备择假设的残差平方和的一元变系数模型的拟合优度检验。Guo 等(2012)在异方差一元变系数分位回

归中考虑了基于原假设和备择假设的分位残差和 RSQ，这里参考 Guo 等(2012)，提出对于二元空间变系数模型的拟合优度检验。

假设参数向量 $\boldsymbol{\theta}$ 的估计记为 $\hat{\boldsymbol{\theta}}$ ，则在原假设成立的情况下，分位残差和 RSQ 可以由以下式子计算

$$RSQ_0 = \sum_{i=1}^n \rho_{\tau} \left\{ Y_i - \sum_{j=0}^p X_{ij}^{\top} \hat{\beta}_j(\mathbf{S}_i, \hat{\boldsymbol{\theta}}) \right\}, \quad (3.22)$$

在备择假设下，分位残差和 RSQ 为

$$RSQ_1 = \sum_{i=1}^n \rho_{\tau} \left\{ Y_i - \mathbf{X}_i^{\top} \hat{\boldsymbol{\beta}}(\mathbf{S}_i) \right\}, \quad (3.23)$$

其中 $\hat{\boldsymbol{\beta}}(\mathbf{s}) = (\hat{\beta}_0(\mathbf{s}), \hat{\beta}_1(\mathbf{s}), \dots, \hat{\beta}_p(\mathbf{s}))^{\top}$ 是 SVCMM-BPBQR 估计量。则构造拟合优度检验统计量为

$$Q_n = \frac{RSQ_0 - RSQ_1}{RSQ_1} = \frac{RSQ_0}{RSQ_1} - 1, \quad (3.24)$$

并且对于较大的 Q_n 值，选择拒绝原假设。

在具体实施中，由于分位数回归不对误差分布进行具体假设，这里采用一种非参数 Bootstrap 方法，来得到检验统计量 Q_n 在原假设下的经验分布和检验 p 值：

Step1. 基于样本数据 $\{Y_i, \mathbf{X}_i, \mathbf{S}_i\}_{i=1}^n$ ，利用式 (3.13) 和式 (3.14) 得到 SVCMM-BPBQR 估计，然后利用式(3.22-24)计算观测统计量的值，记为 \check{Q}_n 。计算残差 $\hat{\epsilon}_i = Y_i - \mathbf{X}_i^{\top} \hat{\boldsymbol{\beta}}(\mathbf{S}_i)$, $i = 1, 2, \dots, n$ 和中心化残差 $\hat{\epsilon}_i^* = \hat{\epsilon}_i - \bar{\hat{\epsilon}}$ ，其中 $\bar{\hat{\epsilon}} = 1/n \sum_{i=1}^n \hat{\epsilon}_i$ 。定义新的响应

$$Y_i^* = \sum_{j=0}^p X_{ij}^{\top} \hat{\beta}_j(\mathbf{S}_i, \hat{\boldsymbol{\theta}}) + \hat{\epsilon}_i^*, \quad i = 1, 2, \dots, n;$$

Step2. 基于重构样本数据 $\{Y_i^*, \mathbf{X}_i, \mathbf{S}_i\}_{i=1}^n$ ，以有放回抽样的方式抽取 n 个样本，利用式(3.24)计算 Bootstrap 检验统计量 Q_n 的统计值 Q_n^* ；

Step3. 重复 Step2. 过程 B 次，得到 Bootstrap 检验统计量 Q_n 的 B 个样本 $\{Q_{n,b}^*\}_{b=1}^B$ ，并根据这些样本计算 Q_n 的经验分布 $F_{Q_n}^*$ ，令其上 α 分位点临界值为 $F_{Q_n, \alpha}^*$ ；

Step4.观察结果, 在给定显著性水平 α 的条件下, 如果一个 Q_n 的观测量 \check{Q}_n 落入拒绝域 $\mathcal{R} = \{\check{Q}_n > F_{\check{Q}_n, \alpha}^*\}$, 就拒绝原假设; 相应地, 该检验的 p 值可以由样本 $\{Q_{n,b}^*\}_{b=1}^B$ 中事件 $\{Q_{n,b}^* \geq \check{Q}_n, b = 1, 2, \dots, B\}$ 发生的频率进行估计, 即 $\hat{p} = \sum_{b=1}^B \mathcal{I}\{Q_{n,b}^* \geq \check{Q}_n\}/B$, 其中 $\mathcal{I}(\cdot)$ 为示性函数。

3.6 数值模拟

3.6.1 估计

在数值模拟中, 考虑 Wood 等(2008)构造的一种在矩形 $[-1, 3.5] \times [-1, 1]$ 内的修正后的马蹄形空间区域, 该矩形区域被均匀的分为 $N_s = 401 \times 901$ 个网格点, 并且函数只在马蹄形区域内取值。假设数据来自模型:

$$Y_i = \beta_0(\mathbf{S}_i)X_0 + \beta_1(\mathbf{S}_i)X_1 + \varepsilon_i, i = 1, 2, \dots, n, \quad (3.25)$$

其中, $\beta_0(\mathbf{s})$ 选取 Wood 等(2008)中的函数, $\beta_1(\mathbf{s})$ 选取 Kim & Wang(2021)中的函数, 图形展示见下文, 具体函数形式展示可参阅原文。假设协变量 $X_j \sim U(0, 1), j = 0, 1$ 。

为了说明 SVCM-BPBQR 估计方法的稳健性, 这里考虑 5 种误差情形: 正态分布 $N(0, 0.5^2)$ 、对数正态分布 $LN(0, 0.5^2)$ 、学生 t 分布 $t(3)$ 、拉普拉斯分布 $Laplace(0, 1, 1)$ 、柯西分布 $Cauchy(0, 0.5^2)$ 。分别代表误差分布为尖峰、厚尾或者非对称的情形。其中, 拉普拉斯分布误差利用 R 包 `GeneralizedHyperbolic` 中的函数 `rskewlap()` 产生。模拟分别考虑了小样本和大样本的情形: $n_1 = 800, n_2 = 2000$ 。对于每一种情形, 模拟 100 次。

此外, 这里采用的三角剖分 T (有 77 个三角形, 65 个顶点), 该三角剖分是计算速度和逼近效果之间的权衡, 详见图 8 及 Kim & Wang(2021), Kong(2015)。为了考虑样条展开阶数对估计的影响, 在模拟中考虑设定阶数 $d = 2$ 和 3 , 光滑度 $\nu = 1$, 光滑参数 (λ_0, λ_1) 通过广义交叉验证 GCV 准则进行选择。在模拟中, 考虑空间变系数模型惩罚迭代重加权最小二乘估计 SVCM-PIRLS(Kim & Wang, 2021)和本文方法 SVCM-BPBQR 的比较。SVCM-PIRLS 估计的实现利用 Kim & Wang(2021)附件中的 R 代码。为了便于比较各估计量的拟合情况, 这里计算估计量的平均积分均方误差 MISE 和平均积分绝对误差 MASE, 计算公式如下

$$\begin{aligned}
 MISE(\hat{\beta}_j) &= \frac{1}{\sum_{i=1}^{N_s} I\{\mathbf{s}_i \in \Omega\}} \sum_{\mathbf{s}_i \in \Omega, i=1,2,\dots,N_s} \left\{ \hat{\beta}_j(\mathbf{s}_i) - \beta_j(\mathbf{s}_i) \right\}^2, \quad j = 0, 1 \\
 MASE(\hat{\beta}_j) &= \frac{1}{\sum_{i=1}^{N_s} I\{\mathbf{s}_i \in \Omega\}} \sum_{\mathbf{s}_i \in \Omega, i=1,2,\dots,N_s} \left| \hat{\beta}_j(\mathbf{s}_i) - \beta_j(\mathbf{s}_i) \right|, \quad j = 0, 1,
 \end{aligned}
 \tag{3.26}$$

其中， \mathbf{s}_i 为矩形区域上的网格点。

表 1 $N(0, 0.5^2)$ 模拟结果

n	方法	d	$\beta_0(\mathbf{s})$		$\beta_1(\mathbf{s})$		Time
			MISE	MASE	MISE	MASE	
n_1	PIRLS	2	0.305(0.077)	1.075(0.151)	0.249(0.065)	0.988(0.139)	4.641
		3	0.269(0.070)	1.020(0.148)	0.223(0.062)	0.943(0.139)	12.041
	BPBQR	2	0.386(0.090)	1.225(0.161)	0.312(0.082)	1.124(0.163)	3.667
		3	0.317(0.087)	1.118(0.165)	0.275(0.080)	1.058(0.170)	7.388
n_2	PIRLS	2	0.160(0.035)	0.761(0.093)	0.138(0.030)	0.716(0.089)	9.208
		3	0.140(0.033)	0.713(0.093)	0.120(0.029)	0.671(0.092)	28.250
	BPBQR	2	0.215(0.043)	0.886(0.113)	0.180(0.041)	0.831(0.115)	6.134
		3	0.166(0.041)	0.795(0.116)	0.152(0.042)	0.771(0.120)	15.192

表 2 $LN(0, 0.5^2)$ 模拟结果

n	方法	d	$\beta_0(\mathbf{s})$		$\beta_1(\mathbf{s})$		Time
			MISE	MASE	MISE	MASE	
n_1	PIRLS	2	6.692(0.875)	6.348(0.444)	6.583(0.890)	6.346(0.453)	2.648
		3	6.636(0.865)	6.349(0.441)	6.543(0.888)	6.343(0.451)	9.321
	BPBQR	2	5.672(0.777)	5.849(0.414)	5.630(0.782)	5.859(0.826)	1.925
		3	5.621(0.778)	5.857(0.412)	5.592(0.779)	5.860(0.426)	5.373
n_2	PIRLS	2	6.379(0.553)	6.301(0.284)	6.420(0.584)	6.348(0.295)	7.103
		3	6.341(0.547)	6.302(0.283)	6.393(0.587)	6.347(0.295)	31.183
	BPBQR	2	5.416(0.470)	5.807(0.257)	5.417(0.470)	5.827(0.262)	3.750
		3	5.352(0.456)	5.797(0.252)	5.405(0.440)	5.835(0.245)	21.290

表 3 $t(3)$ 模拟结果

n	方法	d	$\beta_0(\mathbf{s})$		$\beta_1(\mathbf{s})$		Time
			MISE	MASE	MISE	MASE	
n_1	PIRLS	2	1.760(1.290)	2.661(0.614)	1.690(3.590)	2.917(0.803)	2.760
		3	1.810(2.510)	2.566(0.893)	1.87(5.06)	2.32(1.03)	12.050
	BPBQR	2	1.123(0.419)	2.139(0.409)	1.010(0.413)	2.015(0.408)	2.290
		3	1.047(0.426)	2.054(0.415)	0.995(0.440)	1.988(0.436)	5.650
n_2	PIRLS	2	0.844(0.341)	1.851(0.363)	0.688(0.414)	1.663(0.379)	7.390
		3	0.889(1.569)	1.772(0.622)	0.853(2.337)	1.634(0.679)	42.600
	BPBQR	2	0.531(0.155)	1.460(0.238)	0.488(0.174)	1.392(0.257)	5.250
		3	0.484(0.157)	1.391(0.251)	0.480(0.172)	1.382(0.256)	23.100

表 1-5 展示了所有误差分布不同样本量、不同阶数情形下两种估计量 SVCM-PIRLS 和 SVCM-BPBQR 数值模拟的结果, 汇总了 100 次模拟统计量 MISE、MASE 的均值和标准差。从表 1-5 可以看出, 随着样本量的增加, SVCM-PIRLS 和 SVCM-BPBQR 两种估计量都趋向于更好的结果, 即样本量越大, 估计量误差更小。并且惩罚样条展开阶数越高, 模型拟合的精度越高。表 1 是误差为标准正态分布的模拟结果, 在其他条件一致的情况下, PIRLS 的拟合表现要略好于 BPBQR。表 2-5 汇总的是误差分布为非正态分布时模型的拟合结果。可以看出, 在非正态分布误差情形下, BPBQR 估计量一致优于 PIRLS 估计量。尤其当误差分布为拉普拉斯分布和柯西分布时, 本文所提出的估计方法的稳健性优势更加明显。特别地, 误差分布为柯西分布时, BPBQR 估计量依然可以得到较好的估计结果, 而均值意义下的 PIRLS 估计方法趋于失效。对于误差分布为对数正态分布的情形, 虽然 BPBQR 估计量依然优于 PIRLS 估计量, 但两种估计量的偏差都比较大。这是由于假设误差项的 0.5 分位数为 0, 而实际误差项的 0.5 分位数越等于 1 造成的。

此外, 模拟还记录了 100 次模拟不同估计量在不同情形下的运行时间(表 1-5, Time), 所有的计算的运行均使用个人计算机, 其处理器为 Intel(R) Core(TM) i5-6200U CPU @ 2.30GHz, RAM=4 GB), 本文核心算法 ADMM 算法由 C++编写并在 R 中调用。可以看出, ADMM 估计算法平均运行时间皆比 PIRLS 算法的小得多。从展开阶数比较, 二阶惩罚样条近似花费的时间成本比三阶惩罚样条近似要少, 前者平均运行时间要少一半甚至更多。

表 4 Laplace(0, 1, 1)模拟结果

n	方法	d	$\beta_0(s)$		$\beta_1(s)$		Time
			MISE	MASE	MISE	MASE	
n_1	PIRLS	2	1.319(0.463)	2.338(0.447)	0.969(0.426)	2.003(0.429)	2.816
		3	1.116(0.432)	2.205(0.436)	0.891(0.421)	1.918(0.436)	9.370
	BPBQR	2	0.803(0.314)	1.787(0.348)	0.761(0.404)	1.729(0.403)	1.945
		3	0.733(0.319)	1.700(0.360)	0.725(0.397)	1.679(0.395)	4.810
n_2	PIRLS	2	0.706(0.222)	1.680(0.295)	0.528(0.171)	1.470(0.240)	7.550
		3	0.613(0.203)	1.577(0.286)	0.474(0.159)	1.390(0.250)	29.100
	BPBQR	2	0.419(0.127)	1.269(0.201)	0.364(0.126)	1.201(0.213)	4.670
		3	0.368(0.131)	1.196(0.215)	0.342(0.130)	1.165(0.215)	21.200

表 5 $Cauchy(0, 0.5^2)$ 模拟结果

n	方法	d	$\beta_0(s)$		$\beta_1(s)$		Time	
			MISE	MASE	MISE	MASE		
n_1	PIRLS	2	225961.6 (2178550.1)	49.679 (217.693)	244919.1 (2220139.2)	50.909 (221.743)	2.675	
		3	292606 (2439728)	57.3 (198.2)	494957 (3460993)	57.6 (214.8)	9.660	
		BPBQR	2	0.546 (0.194)	1.463 (0.261)	0.489 (0.205)	1.399 (0.273)	1.995
		3	0.469 (0.185)	1.356 (0.263)	0.450 (0.204)	1.337 (0.274)	4.853	
	n_2	PIRLS	2	32480.16 (208625.77)	46.285 (99.941)	41194.85 (347800.99)	55.894 (163.299)	8.044
			3	12377 (83612)	44.9 (85.9)	38583 (285099)	57 (143)	27.900
BPBQR			2	0.287 (0.069)	1.042 (0.139)	0.248 (0.069)	0.977 (0.146)	4.355
		3	0.166 (0.041)	0.795 (0.116)	0.152 (0.042)	0.771 (0.120)	15.192	

注：柯西误差下，在计算 BPBQR 估计结果的统计量的均值和标准差时，剔除了 2 次异常情况；PIRLS 统计量按原 100 次模拟结果计算。

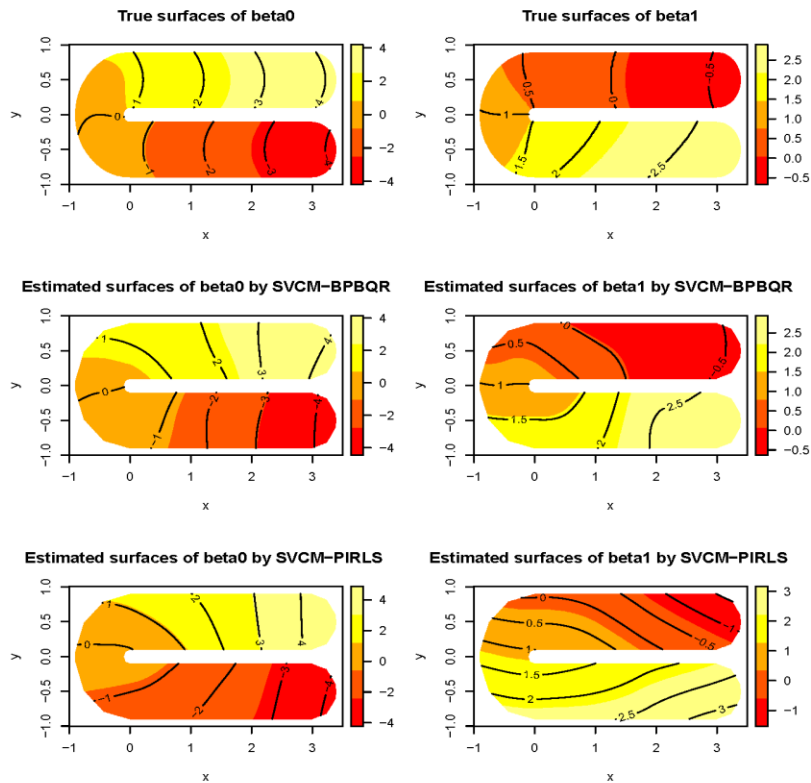


图 2 $Laplace(0, 0.5^2)$ 误差下的系数结果

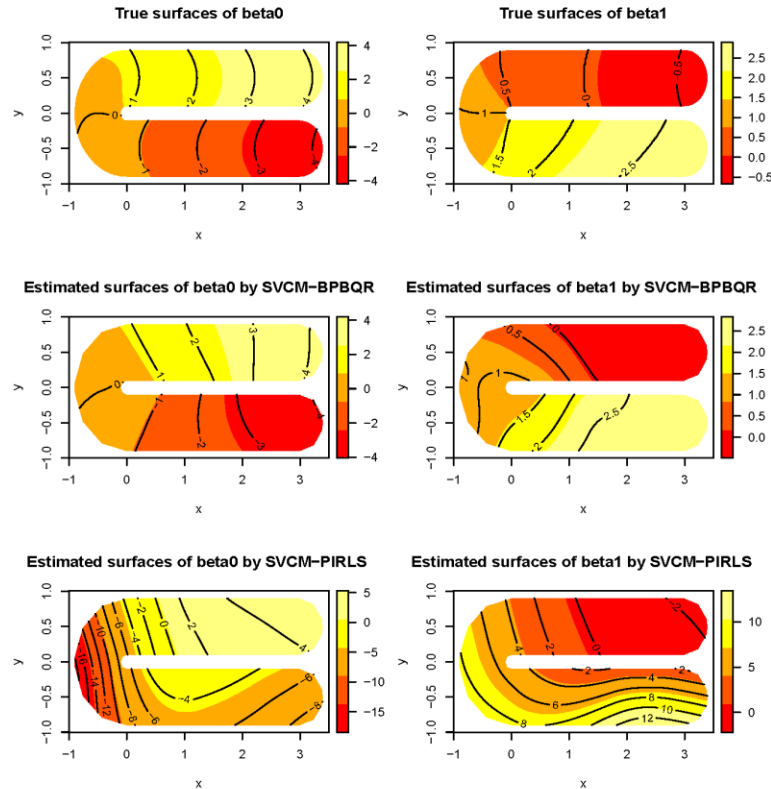


图 3 $Cauchy(0, 0.5^2)$ 误差下的系数结果

最后，绘制了其中某 1 次模拟估计的二元系数函数等高线图，其中图 2 为拉普拉斯分布误差下模拟数据的估计结果，图 3 为柯西分布误差下模拟数据的估计结果。图 2 和图 3 中，上左和上右分别为 $\beta_0(\mathbf{s}), \beta_1(\mathbf{s})$ 的真实图像，中左和中右为利用 BPBOR 方法对进行 $\beta_0(\mathbf{s}), \beta_1(\mathbf{s})$ 估计的结果，下左和下右为利用 PIRLS 方法对进行 $\beta_0(\mathbf{s}), \beta_1(\mathbf{s})$ 估计的结果。可以看出，在两种误差分布情形下，本文提出的方法估计效果要远远优于 PIRLS 估计。显然，拉普拉斯误差下对于 $\beta_1(\mathbf{s})$ 的估计和柯西误差下对于 $\beta_0(\mathbf{s}), \beta_1(\mathbf{s})$ 的估计 PIRLS 方法已经失效。而 BPBQR 估计也可以得到一个相对准确的结果。即基于本文方法的系数函数估计结果更接近“真相”并且更加稳健、有效。此外无论是图 1 或者图 2，两种方法对于 $\beta_0(\mathbf{s})$ 的估计要更好一些。

3.6.2 假设检验

为了说明第 3.5 节中拟合优度检验的有效性，本节执行了如下试验：

(1) 利用 3.6.1 小节模拟实例和相应的模型，并考虑如下检验：

$$H_0: \beta_j(\mathbf{s}) = \beta_j(\mathbf{s}; \boldsymbol{\vartheta}) \text{ v.s. } H_1: \beta_j(\mathbf{s}) \neq \beta_j(\mathbf{s}; \boldsymbol{\vartheta}), \quad 0 \leq j \leq 1. \quad (3.27)$$

利用本文所提出的拟合优度检验，在样本量为 800 的情况下做 100 次检验，对于每次检验，产生 100 个 Bootstrap 样本，设定检验显著性水平为 0.05。

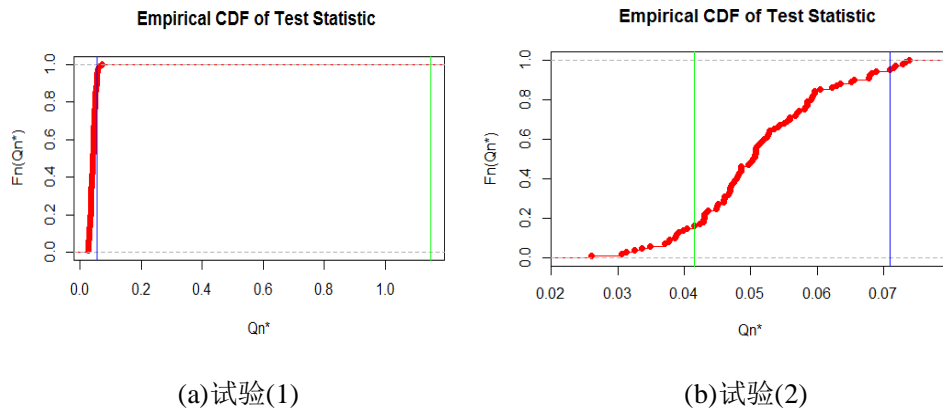


图 4 Q_n^* 检验统计量的 ECDF: 图中蓝色标记上 0.05 分数; 绿色标记观测统计量的值

经过模拟，发现该拟合优度检验可以准确的执行拟合优度检验，在 100 次模拟中，都显著的拒绝了原假设，且 p 值为 0.000。

(2)依然利用 3.6.1 小节模拟实例的数据，将模型假定为使原假设成立的全局回归。其他设定以试验(1)保持一致。模拟结果显示，在该实验中，各模拟均无法拒绝原假设，且 p 值为 0.840。

4 空间变系数复合分位回归模型估计

4.1 二元惩罚样条 CQR 估计

复合分位回归(CQR)是 Zou & Yuan(2008)提出的一种新方法。CQR 同时考虑了多条分位数曲线的信息,能够使感兴趣的问题刻画得更加全面。已有文献证明,在任何误差分布下,CQR 估计量对最小二乘估计量的相对有效性都要大于 70%。记 $c_{\tau_k} = F_{\varepsilon}^{-1}(\tau_k)$ 为 SVCM 中 ε 的 $100\tau_k\%$ 分位数。与 QR 估计不同的是,为了模型(3.1)的可识别性问题,即误差估计和常数估计的识别性问题,CQR 估计不再考虑常数项,即 $\mathbf{X} = (X_1, X_2, \dots, X_p)^\top \in \mathbb{R}^p$ 。并且这里假定 q 个分位水平 $\tau_1, \tau_2, \dots, \tau_q$ 满足 $0 < \tau_1 < \tau_2 < \dots < \tau_q < 1$, 并且考虑等间距分位,即 $\tau_k = k/(q+1) (k = 1, \dots, q)$ 。此外,假定 ε 的密度函数在任何地方都存在,所以 c_{τ_k} 也是唯一确定的,独立于协变量 $\{X_1, X_2, \dots, X_p\}$ 。不同于 q 个独立的分位回归模型,复合分位回归模型假设 q 个分位水平具有相同的协变量固定效应。

若已知样本 $\{(Y_i, \mathbf{X}_i, \mathbf{S}_i)\}_{i=1}^n$, 要得到空间变系数模型的复合分位回归估计,则需求解如下式子

$$\arg \min_{c_{\tau_1}, \dots, c_{\tau_q}, \beta_{\tau_k}(\cdot)} \sum_{i=1}^n \left\{ \sum_{k=1}^q \rho_{\tau_k} \left[Y_i - c_{\tau_k} - \mathbf{X}_i^\top \beta_{\tau_k}(\mathbf{S}_i) \right] \right\}, \quad (4.1)$$

其中, $\beta_{\tau_k}(\mathbf{S}_i) = (\beta_{1,\tau_k}(\mathbf{S}_i), \beta_{2,\tau_k}(\mathbf{S}_i), \dots, \beta_{p,\tau_k}(\mathbf{S}_i))^\top$, 为了简单表示,下文也省略系数函数的分位数标记 τ_k 。式(4.1)亦可称为空间变系数复合分位回归模型。对于模型中的待估参数 $\beta_j(\mathbf{s}), j = 1, 2, \dots, p$, 利用式(3.6)的近似,则式(4.1)改写为

$$\arg \min_{c_{\tau_1}, \dots, c_{\tau_q}, \xi_0, \xi_1, \dots, \xi_p} \sum_{k=1}^q \sum_{i=1}^n \rho_{\tau_k} \left\{ Y_i - c_{\tau_k} - \sum_{j=1}^p X_{ij} \xi_j^\top \mathbf{B}_j(\mathbf{S}_i) \right\}, \quad (4.2)$$

对于式(4.2),这里考虑与分位回归估计相同的光滑约束和空间区域观测数据的稀疏性和不规则性。经过对惩罚项的二元样条近似,使问题转化为求解

$$\begin{aligned} \arg \min_{c_{\tau_1}, \dots, c_{\tau_q}, \xi_0, \xi_1, \dots, \xi_p} & \sum_{k=1}^q \sum_{i=1}^n \rho_{\tau_k} \left\{ Y_i - c_{\tau_k} - \sum_{j=1}^p X_{ij} \xi_j^\top \mathbf{B}_j(\mathbf{S}_i) \right\} + \frac{1}{2} \sum_{j=1}^p \lambda_j \xi_j^\top \mathbf{P}_j \xi_j, \\ \text{s.t. } & \mathbf{H}_j \xi_j = \mathbf{0}. \end{aligned} \quad (4.3)$$

类似地，这里也考虑对所有系数函数执行相同的三角剖分和光滑约束等(详见本文 3.1 节内容)。经过对约束矩阵的 QR 分解等处理，最终约束最优化问题(4.3)转化为无约束惩罚问题

$$\arg \min_{c_{\tau_1}, \dots, c_{\tau_q}, \xi_0^*, \xi_1^*, \dots, \xi_p^*} \sum_{k=1}^q \sum_{i=1}^n \rho_{\tau_k} \left\{ Y_i - c_{\tau_k} - \sum_{j=1}^p X_{ij} (Q^* \xi_j^*)^\top B(S_i) \right\} + \frac{1}{2} \sum_{j=1}^p \lambda_j (Q^* \xi_j^*)^\top P Q^* \xi_j^*, \quad (4.4)$$

公式(4.4)称为空间变系数复合分位回归模型的二元惩罚样条估计，记为 SVCMM-BPBCQR 估计。假设 $\xi^* = (\xi_1^*, \xi_2^*, \dots, \xi_p^*)^\top$ ，且 $\check{\xi}^* = (\check{\xi}_1^*, \check{\xi}_2^*, \dots, \check{\xi}_p^*)^\top$ 为公式(4.4)的解，则对于 $\check{\xi}_j = \{\check{\xi}_{j,m}, m \in \mathcal{M}\}^\top = Q^* \xi_j^*$ ，则 $\beta_j(s)$ 的 SVCMM-BPBCQR 估计量为

$$\check{\beta}_j(s) = \check{\xi}_j^\top B(s) = \sum_{m \in \mathcal{M}} \check{\xi}_{j,m} B_m(s), \quad j = 1, 2, \dots, p. \quad (4.5)$$

4.2 算法与实现

4.2.1 SVCMMQR-ADMM 算法

本小节内容将分位回归 ADMM 算法拓展至复合分位回归 ADMM 算法。这里 \mathbf{Y} 的条件分位数由 $c_{\tau_k} + \sum_{j=1}^p X_j^\top \beta_j(\mathbf{S})$ 给定。通过二元惩罚样条近似，令 $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ ， $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^\top$ ， $\mathbf{X}^* = (Q^* \mathbf{X}_0^*, Q^* \mathbf{X}_1^*, \dots, Q^* \mathbf{X}_p^*)^\top$ ， $\mathbf{X}_j^* = (X_{1j} B(S_1), X_{2j} B(S_2), \dots, X_{nj} B(S_n))$ 。记

$$\epsilon_{ik} = \epsilon_{ik}(\xi^*) = Y_i - c_{\tau_k} - \sum_{j=0}^p X_{ij} (Q^* \xi_j^*)^\top B(S_i), i = 1, 2, \dots, n; k = 1, 2, \dots, q$$

为残差项，惩罚项 $\mathcal{J}_\lambda(\xi^*) = 1/2 \sum_{j=1}^p \lambda_j (Q^* \xi_j^*)^\top P Q^* \xi_j^* = 1/2 \boldsymbol{\lambda} \otimes \xi^{*\top} Q^{*\top} P Q^* \xi^*$,

其中 $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_p)^\top$ 。从分位回归到复合分位回归的推广，只需要容纳多个的条件分位数水平和截距项。为了得到 SVCMMs 复合分位回归 ADMM 算法(记为 SVCMMQR-ADMM 算法)，这里需要将每个分位水平的设计阵进行“叠加”并相应地调整响应变量，生成一个新的设计矩阵 $\mathbf{X}_{CQR}^* \in \mathbb{R}^{nq \times (q+p|\mathcal{M}|)}$ 。具体如下，

$$\mathbf{X}_{CQR}^* = \begin{pmatrix} (1 & 0 & 0 & \cdots & 0) & \mathbf{X}^* \\ (0 & 1 & 0 & \cdots & 0) & \mathbf{X}^* \\ (0 & 0 & 1 & \cdots & 0) & \mathbf{X}^* \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ (1 & 0 & 0 & \cdots & 1) & \mathbf{X}^* \end{pmatrix}_{nq \times (q+p|\mathcal{M}|)}, \quad \mathbf{Y}_{CQR} = \begin{pmatrix} \mathbf{Y} \\ \mathbf{Y} \\ \mathbf{Y} \\ \vdots \\ \mathbf{Y} \end{pmatrix}_{nq \times 1},$$

$$\mathbf{c} = \begin{pmatrix} (\mathbf{c}_{\tau_1})_{n \times 1} \\ (\mathbf{c}_{\tau_2})_{n \times 1} \\ (\mathbf{c}_{\tau_3})_{n \times 1} \\ \vdots \\ (\mathbf{c}_{\tau_q})_{n \times 1} \end{pmatrix}, \quad \boldsymbol{\tau}_{CQR} = \begin{pmatrix} (\tau_1)_{n \times 1} \\ (\tau_2)_{n \times 1} \\ (\tau_3)_{n \times 1} \\ \vdots \\ (\tau_q)_{n \times 1} \end{pmatrix}.$$

其中，在重构设计阵 \mathbf{X}_{CQR}^* 中， $(\mathbf{1} \ \mathbf{0} \ \mathbf{0} \ \dots \ \mathbf{0})$ 是一个 $n \times q$ 的矩阵，其每一行 $(1, 0, \dots, 0)^\top \in \mathbb{R}^q$ 。

现在应用 ADMM 算法，将正则分位回归(4.4)重新表述为凸优化问题：

$$\begin{aligned} \min_{\boldsymbol{\xi}_{CQR}^* \in \mathbb{R}^{q+|\mathcal{M}| \times (p+1)}} & \sum_{k=1}^q \sum_{i=1}^n \rho_{\tau_k}(\epsilon_{ik}) + \mathcal{J}_\lambda(\boldsymbol{\xi}^*) \\ \text{subject to} & \mathbf{X}_{CQR}^{*\top} \boldsymbol{\xi}_{CQR}^* + \boldsymbol{\epsilon} = \mathbf{Y}_{CQR} \end{aligned} \quad (4.6)$$

其中 $\boldsymbol{\xi}_{CQR}^* = (c_{\tau_1}, \dots, c_{\tau_q}, \xi_{01}^*, \dots, \xi_{0|\mathcal{M}|}^*, \dots, \xi_{p1}^*, \dots, \xi_{p|\mathcal{M}|}^*)^\top$ 。

在这里，依然可以利用 ADMM 算法的迭代机制(Boyd 等, 2011)求解(4.6): 假设迭代到第 t 步，则第 $t+1$ 步迭代机制为

$$\begin{aligned} \boldsymbol{\epsilon}^{(t+1)} &= \arg \min_{\boldsymbol{\epsilon} \in \mathbb{R}^{nq}} \sum_{k=1}^q \sum_{i=1}^n \rho_{\tau_k}(\epsilon_{ik}) + \frac{\hbar}{2} \left\| \mathbf{Y}_{CQR} - \boldsymbol{\epsilon} - \mathbf{X}_{CQR}^{*\top} \boldsymbol{\xi}_{CQR}^{*(t)} + \frac{\mathbf{u}^{(t)}}{\hbar} \right\|_2^2 \\ \boldsymbol{\xi}_{CQR}^{*(t+1)} &= \arg \min_{\boldsymbol{\xi}_{CQR}^* \in \mathbb{R}^{q+|\mathcal{M}| \times (p+1)}} \frac{\hbar}{2} \left\| \mathbf{Y}_{CQR} - \boldsymbol{\epsilon}^{(t+1)} - \mathbf{X}_{CQR}^{*\top} \boldsymbol{\xi}_{CQR}^* + \frac{\mathbf{u}^{(t)}}{\hbar} \right\|_2^2 + \mathcal{J}_\lambda(\boldsymbol{\xi}^*) \\ \mathbf{u}^{(t+1)} &= \mathbf{u}^{(t)} + \hbar \left(\mathbf{Y}_{CQR} - \boldsymbol{\epsilon}^{(t+1)} - \mathbf{X}_{CQR}^{*\top} \boldsymbol{\xi}_{CQR}^{*(t+1)} \right), \end{aligned} \quad (4.7)$$

相应地，类似于分位回归 ADMM 算法，这里用软门限算子和岭回归算法对式(4.7)中的迭代表达式进行化简。即

$$\begin{aligned} \boldsymbol{\epsilon}^{(t+1)} &= \mathcal{S}_{\frac{1}{\hbar}} \left(\mathbf{Y}_{CQR} - \mathbf{X}_{CQR}^{*\top} \boldsymbol{\xi}_{CQR}^{*(t)} + \frac{\mathbf{u}^{(t)}}{\hbar} - (2\boldsymbol{\tau}_{CQR} - \mathbf{1}_{nq \times 1}) / \hbar \right), \\ \boldsymbol{\xi}_{CQR}^{*(t+1)} &= \left(\mathbf{X}_{CQR}^{*\top} \mathbf{X}_{CQR}^* + \begin{pmatrix} \mathbf{0}_{q \times q} & \mathbf{0} \\ \mathbf{0} & \lambda \otimes \mathbf{Q}^{*\top} \mathbf{P} \mathbf{Q}^* \end{pmatrix} / \hbar \right)^{-1} \\ & \quad \mathbf{X}_{CQR}^{*\top} \left(\mathbf{Y}_{CQR} - \boldsymbol{\epsilon}^{(t+1)} + \frac{\mathbf{u}^{(t)}}{\hbar} \right). \end{aligned} \quad (4.8)$$

相应地，算法终止条件为

$$\begin{aligned}\boldsymbol{\epsilon}_{pri}^{(k)} &= \mathbf{Y}_{CQR} - \mathbf{X}_{CQR}^{*\top} \boldsymbol{\xi}_{CQR}^{*(k)} - \boldsymbol{\epsilon}^{(k)} \\ \boldsymbol{\epsilon}_{dual}^{(k)} &= \hbar \mathbf{X}_{CQR}^{*\top} \left(\boldsymbol{\xi}_{CQR}^{*(k)} - \boldsymbol{\xi}_{CQR}^{*(k-1)} \right),\end{aligned}\quad (4.9)$$

$$\left\| \boldsymbol{\epsilon}_{pri}^{(k)} \right\|_2 \leq e_{pri}, \quad \left\| \boldsymbol{\epsilon}_{dual}^{(k)} \right\|_2 \leq e_{dual}, \quad (4.10)$$

$$\begin{aligned}e_{pri} &= \sqrt{n}e_{abs} + e_{rel} \max \left\{ \left\| \mathbf{X}_{CQR}^{*\top} \boldsymbol{\xi}_{CQR}^{*(k)} \right\|_2^2, \left\| \boldsymbol{\epsilon}^{(k)} \right\|_2^2, \left\| \mathbf{Y}_{CQR} \right\|_2^2 \right\} \\ e_{dual} &= \sqrt{|\mathcal{M}|} \times pe_{abs} + e_{rel} \left\| \mathbf{X}_{CQR}^{*\top} \mathbf{u}^{(k)} \right\|_2^2,\end{aligned}$$

其中，上述公式中其他记号含义与分位回归 ADMM 算法一致。

4.2.2 光滑参数的选取

与分位回归类似，对于光滑参数向量 $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_p)^\top$ 的选取，这里依然采用 GCV 方法，即在一个网格上最小化如下 GCV 准则：

$$GCV(\boldsymbol{\lambda}) = \frac{n^{-1} \left\| \mathbf{Y} - \mathbf{L}(\boldsymbol{\lambda}) \mathbf{Y} \right\|^2}{\left\{ n^{-1} \text{tr}(\mathbf{I} - \mathbf{L}(\boldsymbol{\lambda})) \right\}^2}, \quad (4.11)$$

其中 $\mathbf{L}(\boldsymbol{\lambda}) = \mathbf{X}_{CQR}^* \left(\mathbf{X}_{CQR}^{*\top} \mathbf{X}_{CQR}^* + \begin{pmatrix} \mathbf{0}_{q \times q} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\lambda} \otimes \mathbf{Q}^{*\top} \mathbf{P} \mathbf{Q}^* \end{pmatrix} \right)^{-1} \mathbf{X}_{CQR}^{*\top}$ ， $\text{tr}(\cdot)$ 表示 \cdot 的迹。

4.3 数值模拟

在 CQR 估计数值模拟中，考虑与 3.6 节相同的马蹄形区域、数据生成模型和二元系数函数假定。为了说明 SVCM-BPBCQR 估计方法的稳健性，这里也考虑 5 种误差情形：混合正态分布 $MN(0, 0.9, 0.1, 0.5^2, 10^2)$ 、对数正态分布 $LN(0, 0.5^2)$ 、学生 t 分布 $t(3)$ 、拉普拉斯分布 $Laplace(0, 1, 1)$ 、柯西分布 $Cauchy(0, 0.5^2)$ 。其中，混合正态分布 $MN(\mu, p, q, \sigma_p^2, \sigma_q^2)$ 表示 $p100\%$ 的误差来自正态分布 $N(\mu, \sigma_p^2)$ 以及 $q100\%$ 的误差来自 $N(\mu, \sigma_q^2)$ ，且 $p + q = 1$ 。除此之外，模拟分别考虑了小样本和大样本的两种情形： $n_1 = 800, n_2 = 2000$ 。对于每一种情形，模拟 100 次。

这里将不再考虑惩罚样条展开阶数对分位回归估计的影响，在模拟中考虑设定阶数 $d = 2$ 。光滑度 $\nu = 1$ ，光滑参数 (λ_0, λ_1) 通过广义交叉验证 GCV 准则选择。参与模拟比较的估计方法有，空间变系数模型的惩罚迭代重加权最小二乘估计

SVCM-PIRLS(Kim & Wang, 2021)、本文提出的分位回归估计方法 SVCM-BPBQR、复合分位回归估计方法 SVCM-BPBCQR。并且对于 SVCM-BPBCQR 估计量,这里考虑 3 个分位水平 BPBCQR3, 6 个分位水平 BPBCQR5 和 9 个分位水平 BPBCQR9。为了便于比较各估计量的估计精度,计算了估计量的 MISE 和 MASE, 详见 3.6.1 节。

表 6-10 展示了所有误差分布不同样本量情形下估计量 SVCM-PIRLS 和 SVCM-BPBQR, 以及三种 SVCM-BPBCQR 估计量数值模拟的结果。可以看出, 样本量的增加会提高各种估计量的拟合精度。表 6 给出了误差为混合正态分布的数值模拟结果, 可以看到在其他条件一致的情况下, 复合分位回归 BPBCQR 的拟合表现相比于 PIRLS 优势明显, 并且也要优于本文的 BPBQR 估计量。表 7 给出了当误差分布为对数正态分布时模型的拟合结果。可以看出 BPBCQR 估计量一致优于 PIRLS 估计量和 BPBQR 估计量, 其优势在于不仅消除了误差分布均值不为 0 带来的巨大偏倚, 还使估计结果能够不受异常值的影响, 拟合精度达到了最优的水平。表 8 给出了误差分布为学生 t 分布时的模拟结果, 结果显示 BPBCQR 估计量依然表现出比 PIRLS 估计量和 BPBQR 估计量更好的结果。并且在 BPBCQR 估计量中, 考虑分位水平越多, 估计结果会趋于更精确。表 9-10 分别是在拉普拉斯分布和柯西分布误差下模拟结果, 可以看出在这两种误差下, 本章提出的复合分位回归估计量 BPBCQR 表现虽然优于 PIRLS, 但相比分位回归估计表现略差一些。本文也记录了 100 次模拟的平均计算时间, 可以看出随着分位水平的增多, 计算时间也将增加。

整体而言, 复合分位回归估计量 BPBCQR 在部分误差下, 其估计结果优于分位回归和均值回归。根据本文的三种 BPBCQR 估计量估计精度和估计时间, 在实际应用中如果数据量足够的大, 本文建议采用三个分位水平即可, 因为过度的增加分位水平带来的估计精度的提升远不及它带来的计算压力。

图 5 和图 6 分别绘制了在混合正态误差分布下和对数正态误差分布下 BPBCQR3、BPBQR 以及 PIRLS 估计量对本文设置的马蹄形区域系数内函数的估计结果。可以看出, 在混合正态分布误差下, BPBCQR3 和 BPBQR 估计方法得到的结果相差不大, 能较好的接近于真实函数。而 PIRLS 由于异常值的影响, 几乎失效。在对数正态分布下, 由等高线可以看出 BPBCQR3 估计量消除了误差分布带来的偏倚, 估计结果更值得信赖。而分位回归和均值回归都没有做到这一点。

表 6 $MN(0, 0.9, 0.1, 0.5^2, 10^2)$ 模拟结果

n	方法	$\beta_0(s)$		$\beta_1(s)$		Time
		MISE	MASE	MISE	MASE	
n_1	PIRLS	7.467 (10.338)	3.608 (1.461)	6.761 (10.927)	3.324 (1.551)	3.831
	BPBQR	0.416(0.120)	1.277(0.191)	0.345(0.109)	1.180(0.198)	1.928
	BPBCQR3	0.380(0.119)	1.215(0.199)	0.316(0.098)	1.257(0.181)	4.709
	BPBCQR6	0.381(0.115)	1.216(0.198)	0.314(0.096)	1.124(0.178)	7.456
	BPBCQR9	0.381(0.115)	1.217(0.197)	0.313(0.096)	1.122(0.178)	11.158
n_2	PIRLS	3.073(3.296)	2.455(0.756)	3.182(4.261)	2.321(0.948)	7.159
	BPBQR	0.232(0.049)	0.931(0.120)	0.199(0.048)	0.880(0.125)	3.396
	BPBCQR3	0.207(0.052)	0.874(0.122)	0.179(0.043)	0.827(0.113)	27.397
	BPBCQR6	0.208(0.050)	0.877(0.119)	0.178(0.042)	0.826(0.111)	38.191
	BPBCQR9	0.209(0.049)	0.877(0.117)	0.178(0.041)	0.826(0.111)	52.801

表 7 $LN(0, 0.5^2)$ 模拟结果

n	方法	$\beta_0(s)$		$\beta_1(s)$		Time
		MISE	MASE	MISE	MASE	
n_1	PIRLS	6.692(0.875)	6.348(0.443)	6.583(0.890)	6.345(0.453)	2.735
	BPBQR	5.672(0.777)	5.850(0.414)	5.630(0.782)	5.860(0.427)	1.966
	BPBCQR3	0.321(0.077)	1.114(0.147)	0.269(0.078)	1.035(0.163)	4.497
	BPBCQR6	0.324(0.076)	1.117(0.147)	0.271(0.078)	1.041(0.164)	7.724
	BPBCQR9	0.328(0.078)	1.123(0.148)	0.272(0.077)	1.043(0.162)	9.908
n_2	PIRLS	6.379(0.553)	6.301(0.284)	6.420(0.584)	6.348(0.295)	7.227
	BPBQR	5.416(0.470)	5.807(0.257)	5.417(0.470)	5.828(0.262)	3.479
	BPBCQR3	0.170(0.039)	0.791(0.104)	0.151(0.035)	0.756(0.104)	29.742
	BPBCQR6	0.171(0.039)	0.794(0.107)	0.151(0.036)	0.757(0.105)	39.599
	BPBCQR9	0.171(0.038)	0.795(0.103)	0.151(0.035)	0.757(0.104)	51.005

表 8 $t(3)$ 模拟结果

n	方法	$\beta_0(s)$		$\beta_1(s)$		Time
		MISE	MASE	MISE	MASE	
n_1	PIRLS	1.762(1.287)	2.661(0.614)	1.687(3.594)	2.346(0.803)	2.877
	BPBQR	1.122(0.419)	2.139(0.408)	1.010(0.413)	2.015(0.408)	2.305
	BPBCQR3	1.080(0.421)	2.088(0.405)	1.022(0.461)	2.026(0.480)	4.811
	BPBCQR6	1.061(0.404)	2.072(0.393)	1.000(0.444)	2.009(0.475)	7.669
	BPBCQR9	1.060(0.407)	2.071(0.395)	0.994(0.439)	2.004(0.471)	10.948
n_2	PIRLS	0.844(0.341)	1.851(0.363)	0.688(0.414)	1.663(0.379)	7.460
	BPBQR	0.533(0.154)	1.464(0.172)	0.488(0.172)	1.391(0.255)	3.777
	BPBCQR3	0.506(0.159)	1.433(0.251)	0.487(0.178)	1.390(0.276)	19.779
	BPBCQR6	0.505(0.159)	1.431(0.252)	0.482(0.177)	1.382(0.274)	53.458
	BPBCQR9	0.503(0.160)	1.429(0.253)	0.479(0.178)	1.377(0.277)	60.137

表 9 $Laplace(0, 1, 1)$ 模拟结果

n	方法	$\beta_0(s)$		$\beta_1(s)$		Time
		MISE	MASE	MISE	MASE	
n_1	PIRLS	1.319(0.463)	2.338(0.447)	0.969(0.426)	2.003(0.429)	2.731
	BPBQR	0.803(0.314)	1.787(0.348)	0.761(0.404)	1.729(0.403)	2.003
	BPBCQR3	0.961(0.395)	1.962(0.400)	0.887(0.417)	1.860(0.431)	4.684
	BPBCQR6	0.953(0.394)	1.956(0.403)	0.872(0.414)	1.844(0.427)	7.844
	BPBCQR9	0.950(0.389)	1.955(0.401)	0.871(0.413)	1.841(0.426)	13.210
n_2	PIRLS	0.706(0.222)	1.680(0.295)	0.528(0.171)	1.466(0.240)	7.558
	BPBQR	0.419(0.127)	1.269(0.201)	0.363(0.125)	1.201(0.213)	4.352
	BPBCQR3	0.486(0.139)	1.392(0.216)	0.458(0.151)	1.359(0.233)	21.541
	BPBCQR6	0.485(0.140)	1.389(0.218)	0.454(0.152)	1.352(0.234)	38.956
	BPBCQR9	0.484(0.141)	1.388(0.217)	0.452(0.152)	1.347(0.236)	66.571

表 10 $Cauchy(0, 0.5^2)$ 模拟结果

n	方法	$\beta_0(s)$		$\beta_1(s)$		Time
		MISE	MASE	MISE	MASE	
n_1	PIRLS	225961.6 (2178550.1)	49.679 (217.693)	244919.1 (220139.2)	50.908 (221.743)	2.839
	BPBQR	0.546(0.194)	1.463(0.261)	0.489(0.206)	1.399(0.274)	1.978
	BPBCQR3	0.627(0.257)	1.580(0.320)	0.592(0.261)	1.539(0.327)	5.261
	BPBCQR6	0.626(0.252)	1.579(0.318)	0.585(0.256)	1.532(0.326)	8.002
	BPBCQR9	0.627(0.250)	1.581(0.317)	0.582(0.255)	1.530(0.326)	11.602
n_2	PIRLS	32480.16 (208625.77)	46.286 (99.942)	41194.85 (347800.99)	55.894 (163.299)	7.133
	BPBQR	0.287(0.069)	1.042(0.139)	0.249(0.090)	0.979(0.146)	3.719
	BPBCQR3	0.327(0.094)	1.122(0.176)	0.290(0.086)	1.073(0.167)	48.862
	BPBCQR6	0.328(0.094)	1.123(0.173)	0.289(0.086)	1.070(0.169)	48.162
	BPBCQR9	0.329(0.094)	1.124(0.174)	0.288(0.085)	1.069(0.167)	75.832

注：柯西误差下，在计算 BPBQR 估计结果的统计量的均值和标准差时，剔除了 2 次异常情况；PIRLS 统计量按原 100 次模拟结果计算。

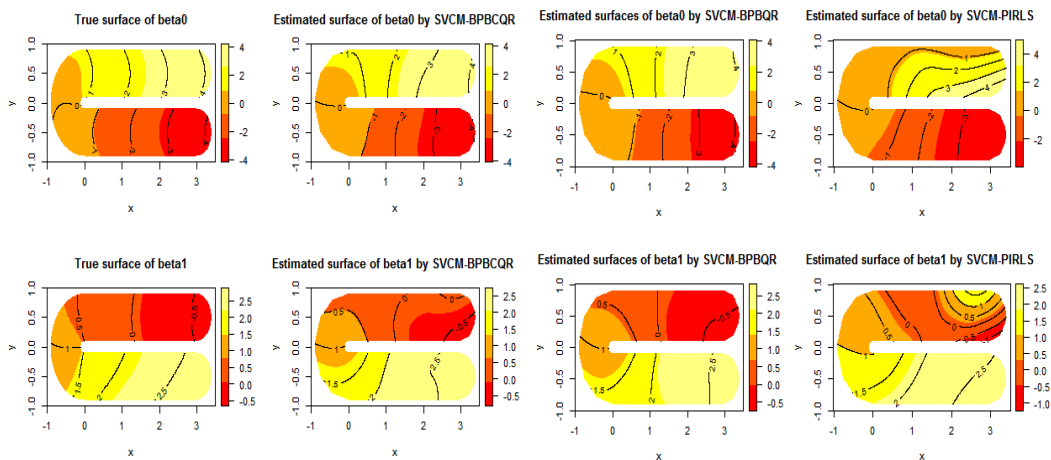


图 5 $MN(0, 0.9, 0.1, 0.5^2, 10^2)$ 模拟结果

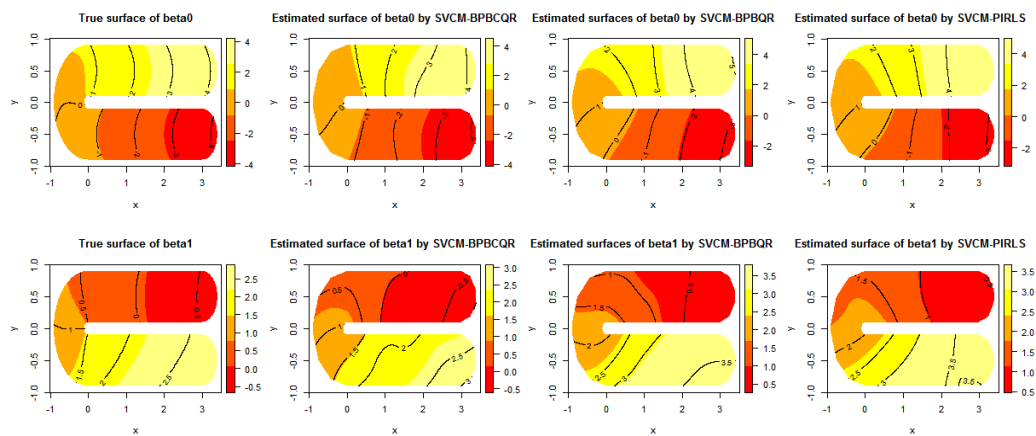


图 6 $LN(0, 0.5^2)$ 模拟结果

5 空间部分线性变系数分位回归模型估计

5.1 空间部分线性变系数模型

本节考虑将部分线性变系数模型推广至空间部分变系数模型(SPLVCMs)。令 $\{(Y_i, \mathbf{X}_i, \mathbf{Z}_i, \mathbf{S}_i)\}_{i=1}^n = \{Y_i, X_{i0}, X_{i1}, \dots, X_{ip}, Z_{i0}, Z_{i1}, \dots, Z_{iq}, S_{i1}, S_{i2}\}_{i=1}^n$ 是来自总体 $(Y, \mathbf{X}, \mathbf{S})$ 联合分布的 n 个独立的样本。其中, $Y \in \mathbb{R}^1$ 是响应变量, $\mathbf{X} = (X_1, X_2, \dots, X_p)^\top \in \mathbb{R}^p$ 为 p 维的解释变量, 设定其对响应变量 Y 是线性影响关系, 这意味着这些协变量对响应变量的影响具有空间平稳性。 $\mathbf{Z} = (Z_0, Z_1, \dots, Z_q)^\top \in \mathbb{R}^{q+1}$ 对 Y 的影响具有空间依赖性(即空间非平稳的), $Z_{i0} \equiv 1$ 是常数项, $\mathbf{S} = (S_1, S_2)^\top \in \mathbb{R}^2$ 总体 $(Y, \mathbf{X}, \mathbf{Z})$ 在任意形状有界区域 $\Omega \subset \mathbb{R}^2$ 中产生数据的观测位置。给定上述记号, 空间部分线性变系数模型可以表示为

$$\begin{aligned} Y_i &= \mathbf{X}_i^\top \boldsymbol{\alpha} + \mathbf{Z}_i^\top \boldsymbol{\beta}(\mathbf{S}_i) + \varepsilon_i \\ &= \mathbf{X}_i^\top \boldsymbol{\alpha} + \mathbf{Z}_i^\top \boldsymbol{\beta}(S_{i1}, S_{i2}) + \varepsilon_i, \end{aligned} \quad (5.1)$$

模型 (5.1) 中, $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_p)^\top$ 表示线性部分的 p 维系数向量, $\boldsymbol{\beta}(\cdot) = \boldsymbol{\beta}(\cdot, \cdot) = (\beta_0(\cdot, \cdot), \beta_1(\cdot, \cdot), \dots, \beta_q(\cdot, \cdot))^\top$ 表示变系数部分二维系数函数向量, ε_i 随机误差项, 假设其服从于 $F(\cdot)$ 。对于模型(5.1)参数估计的理论研究, 目前还没有较为系统的讨论。

5.2 偏二元惩罚样条 QR 估计

经典的均值回归估计的是响应变量 Y 的条件均值, 进而刻画解释变量带来的平均影响趋势。不同于均值回归, 分位回归需要估计 Y 的 τ ($0 < \tau < 1$) 阶条件分位数。首先从理论上推导出分位回归估计思想。一般地, 对于一个随机向量 $(\mathbf{X}, \mathbf{Z}, \mathbf{S}, Y)$, 在给定协变量 $\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, \mathbf{S} = \mathbf{s}$ 的情况下, Y 的条件累计分布函数记为 $F_{Y|\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}, \mathbf{S}=\mathbf{s}}(y|\mathbf{x}, \mathbf{z}, \mathbf{s})$, 该条件随机变量 $Y|\mathbf{X}, \mathbf{Z}, \mathbf{S}$ 的 τ 阶分位数定义为

$$Q_\tau(Y|\mathbf{X} = \mathbf{x}, \mathbf{S} = \mathbf{s}) = \arg \inf \{y \in \mathbb{R} : F_{Y|\mathbf{X}=\mathbf{x}, \mathbf{Z}=\mathbf{z}, \mathbf{S}=\mathbf{s}}(y|\mathbf{x}, \mathbf{z}, \mathbf{s}) \geq \tau\}. \quad (5.2)$$

令 $\rho_\tau(u) = u[\tau - I(u < 0)]$ 为检验函数, 即分位数在 τ 的损失函数。则利用下面式子就可得到 τ 阶条件分位数的估计

$$Q_\tau(\mathbf{x}, \mathbf{z}, \mathbf{s}) = \arg \min_{a \in \mathbb{R}} E \{ \rho_\tau(Y - a) | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}, \mathbf{S} = \mathbf{s} \}. \quad (5.3)$$

现在需要重新考虑本文的空间部分线性变系数模型(5.1), 与之对应 $Q_\tau(\mathbf{X}, \mathbf{Z}, \mathbf{S}) = c_\tau + \sum_{l=1}^p \mathbf{X}_l^\top \alpha_l + \sum_{j=0}^q \mathbf{Z}_j^\top \beta_{\tau,j}(\mathbf{S})$, 其中 $c_\tau = F^{-1}(\tau)$ 为 ε 的 $100\tau\%$ 分位数, 这里假定为 $Quantile_\tau(\varepsilon) = 0$. 若已知样本 $\{(Y_i, \mathbf{X}_i, \mathbf{Z}_i, \mathbf{S}_i)\}_{i=1}^n$, 并得到参数估计, 需要对下面式子关于 $\alpha_\tau, \beta_\tau(\cdot)$ 最小化

$$\sum_{i=1}^n \rho_\tau \{ Y_i - \mathbf{X}_i^\top \alpha_\tau - \mathbf{Z}_i^\top \beta_\tau(\mathbf{S}_i) \}, \quad (5.4)$$

其中 $\alpha_\tau = (\alpha_{0\tau}, \alpha_{1\tau}, \dots, \alpha_{p\tau})^\top$, $\beta_\tau(\cdot) = (\beta_{0,\tau}(\cdot), \beta_{1,\tau}(\cdot), \dots, \beta_{q,\tau}(\cdot))^\top$, 为了简单表示, 下文中系数函数的分位数标记 τ 省略. 对于变系数部分的非参数函数 $\beta_j(\mathbf{s}), j = 0, 1, \dots, q$, 常见的处理方式就是将其参数化, 然后利用参数的估计办法即可, 基于前文中提到的, 利用基于三角剖分的局部二元样条逼近方法将其近似为

$$\beta_j(\mathbf{s}) \approx \boldsymbol{\xi}_j^\top \mathbf{B}_j(\mathbf{s}), \quad j = 0, 1, \dots, q, \quad (5.5)$$

则式(5.4)转化为关于 $\alpha, \boldsymbol{\xi}_0, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_p$ 的最小化问题

$$\sum_{i=1}^n \rho_\tau \left\{ Y_i - \sum_{l=0}^p X_{il} \alpha_l - \sum_{j=0}^q Z_{ij} \boldsymbol{\xi}_j^\top \mathbf{B}_j(\mathbf{S}_i) \right\}, \quad (5.6)$$

式(5.6)称为空间部分线性变系数分位回归模型, 也视为 SPLVCMs 二元样条分位回归估计, 记为 SPLVCM-BBQR 估计. 但是对于式(5.6), 如果直接进行分位回归估计, 可能会使得变系数部分的二元系数函数不具备良好的光滑性能, 而且很有可能会和实际情况不吻合. 为了考虑变系数部分的光滑约束和空间区域观测数据的稀疏性和不规则特点. 这里, 本文考虑对式(5.6)添加一个粗糙惩罚, 即惩罚样条逼近法. 具体如下:

$$\sum_{i=1}^n \rho_\tau \left\{ Y_i - \sum_{l=0}^p X_{il} \alpha_l - \sum_{j=0}^q Z_{ij} \boldsymbol{\xi}_j^\top \mathbf{B}_j(\mathbf{S}_i) \right\} + \frac{1}{2} \sum_{j=0}^q \lambda_j \mathcal{J}(\beta_j), \quad (5.7)$$

并且对于任意的二元函数 $g(\mathbf{s}) = g(s_1, s_2)$ 有

$$\mathcal{J}(g) = \int \int_{\Omega} \{ (\nabla_{s_1}^2 g)^2 + (\nabla_{s_2}^2 g)^2 + 2(\nabla_{s_1}^1 \nabla_{s_2}^1 g) \} ds_1 ds_2,$$

其中, $\nabla_{s_k}^\zeta g(\mathbf{s})$ 为 $g(\mathbf{s})$ 在区域 Ω 上任一点 (s_1, s_2) 方向 $s_k, k = 1, 2$ 上的 ζ 阶导数, $\lambda_j, j = 0, 1, \dots, q$ 为相应的非负光滑参数, 也称调谐参数. 对于式(5.7)中的惩罚项也可以通过二元样条逼近方法近似, 即

$$\mathcal{J}(\beta_j) \approx \mathcal{J}(\boldsymbol{\xi}_j^\top \mathbf{B}_j(\mathbf{S}_i)) = \boldsymbol{\xi}_j^\top \mathbf{P}_j \boldsymbol{\xi}_j, \quad (5.8)$$

其中, \mathbf{P}_j 对应第 β_j 的分块对角惩罚矩阵。考虑到光滑约束条件, 结合式(5.8), 式(5.7)又可以改写为下面式子

$$\begin{aligned} \min_{\boldsymbol{\alpha}, \boldsymbol{\xi}_0, \boldsymbol{\xi}_1, \dots, \boldsymbol{\xi}_p} \sum_{i=1}^n \rho_\tau \left\{ Y_i - \sum_{l=0}^p X_{il} \alpha_l - \sum_{j=0}^q Z_{ij} \boldsymbol{\xi}_j^\top \mathbf{B}_j(\mathbf{S}_i) \right\} + \frac{1}{2} \sum_{j=0}^q \lambda_j \boldsymbol{\xi}_j^\top \mathbf{P}_j \boldsymbol{\xi}_j, \\ \text{s.t. } \boldsymbol{\Psi}_j \boldsymbol{\xi}_j = \mathbf{0}. \end{aligned} \quad (5.9)$$

原则上对于不同的系数函数, 亦可以采用不同的三角剖分、Bernstein 基多项式和约束矩阵以及对角惩罚矩阵等。为了计算方便, 这里对所有系数函数采用相同的处理。所以假设 $T_j \equiv T$, 且有 N 个三角形, $\mathbf{B}_j(\mathbf{S}_i) \equiv \mathbf{B}(\mathbf{S}_i) = \{B_m, m \in \mathcal{M}\}^\top$, $|\mathcal{M}| = N(d+1)(d+2)/2$, $\boldsymbol{\Psi}_j \equiv \boldsymbol{\Psi}$, $\mathbf{P}_j \equiv \mathbf{P}$ 。除此之外, 为了式(5.9)变为无约束问题, 采用一种基于 QR 分解的技术处理光滑约束矩阵 $\boldsymbol{\Psi}$ (假设其秩为 ϱ), 详见 Kim & Wang(2021)。使得某些 $\boldsymbol{\xi}_j = \mathbf{Q}^* \boldsymbol{\xi}_j^*$ 能够保证光滑约束 $\boldsymbol{\Psi} \boldsymbol{\xi}_j = \mathbf{0}$, 其中 \mathbf{Q}^* 来自

$$\boldsymbol{\Psi}^\top = \mathbf{Q}\mathbf{R} = \begin{pmatrix} \mathbf{Q}^\varrho & \mathbf{Q}^* \end{pmatrix} \begin{pmatrix} \mathbf{R}_1 \\ \mathbf{R}_2 \end{pmatrix}, \quad (5.10)$$

其中 \mathbf{Q} 为正交矩阵, \mathbf{R} 为上三角阵, \mathbf{Q}^ϱ 是 \mathbf{Q} 的前 ϱ 列, $\mathbf{R}_2 = \mathbf{0}$ 。经过处理, 最终约束最优化问题转化为无约束惩罚问题

$$\sum_{i=1}^n \rho_\tau \left\{ Y_i - \sum_{l=0}^p X_{il} \alpha_l - \sum_{j=0}^q Z_{ij} (\mathbf{Q}^* \boldsymbol{\xi}_j^*)^\top \mathbf{B}(\mathbf{S}_i) \right\} + \frac{1}{2} \sum_{j=0}^q \lambda_j (\mathbf{Q}^* \boldsymbol{\xi}_j^*)^\top \mathbf{P} \mathbf{Q}^* \boldsymbol{\xi}_j^*, \quad (5.11)$$

式(5.11)关于 $\boldsymbol{\alpha}, \boldsymbol{\xi}_0^*, \boldsymbol{\xi}_1^*, \dots, \boldsymbol{\xi}_q^*$ 的最小化解称为空间部分线性变系数分位回归模型的偏二元惩罚样条估计, 记为 SPLVCM-BPBQR 估计量。记 $\hat{\boldsymbol{\alpha}} = (\hat{\alpha}_0, \hat{\alpha}_1, \dots, \hat{\alpha}_p)^\top$ 为线性部分估计量。令 $\boldsymbol{\xi}^* = (\boldsymbol{\xi}_0^*, \boldsymbol{\xi}_1^*, \dots, \boldsymbol{\xi}_q^*)^\top$, 且 $\hat{\boldsymbol{\xi}}^* = (\hat{\boldsymbol{\xi}}_0^*, \hat{\boldsymbol{\xi}}_1^*, \dots, \hat{\boldsymbol{\xi}}_q^*)^\top$ 为式(5.11)变系数部分的解, 则对于 $\hat{\boldsymbol{\xi}}_j = \{\hat{\xi}_{j,m}, m \in \mathcal{M}\}^\top = \mathbf{Q}^* \hat{\boldsymbol{\xi}}_j^*$, 得到 $\beta_j(s)$ 的 SPLVCM-BPBQR 估计量为

$$\hat{\beta}_j(s) = \hat{\boldsymbol{\xi}}_j^\top \mathbf{B}(s) = \sum_{m \in \mathcal{M}} \hat{\xi}_{j,m} B_m(s), \quad j = 0, 1, \dots, q. \quad (5.12)$$

5.3 估计量的理论性质

本节研究所提出估计量 SPLVCM-BPBQR 的渐近性质。若假设二元系数函数 $\beta_j(\cdot), j = 0, 1, \dots, q$ 均来自一般的 Sobolev 空间, 则需要考虑逼近偏倚, 并且当三角剖分中三角形个数满足 $N \asymp O(n^{1/(d+2)})$ 以及一些适当的正则性条件下, 可以有,

$\|\hat{\beta}_j(\cdot, \cdot) - \beta_j(\cdot, \cdot)\|_{L_2(\Omega)}^2 = O_p \{n^{-(d+1)/(d+2)}\}$, 并且这是 Stone(1982)中给出的最优收敛速率。详情可参见 Mu 等(2019)。类似于 Andriyana 等(2014), 下面考虑假设二元系数函数 $\beta_j(\cdot), j = 0, 1, \dots, q$ 是 Benstein 基多项式 $\{B_m, m \in \mathcal{M}\}^\top$ 张成的空间中的元素。则式(2.22)的逼近变成等式不存在逼近偏倚。首先引入记号, 令 $\mathbf{Z}^* = (\mathbf{Q}^* \mathbf{Z}_0^*, \mathbf{Q}^* \mathbf{Z}_1^*, \dots, \mathbf{Q}^* \mathbf{Z}_q^*), \mathbf{Z}_j^* = (Z_{1j} \mathbf{B}(\mathbf{S}_1), Z_{2j} \mathbf{B}(\mathbf{S}_2), \dots, Z_{nj} \mathbf{B}(\mathbf{S}_n))^\top$, $\boldsymbol{\xi}^* = (\boldsymbol{\xi}_0^*, \boldsymbol{\xi}_1^*, \dots, \boldsymbol{\xi}_q^*)^\top$ 。

这里引入一些假设, 使其保证理论性质能够成立:

(C1) 令 $f_{\mathbf{S}}(\mathbf{s})$ 为二维随机向量 $\mathbf{S} = (S_1, S_2)^\top$ 的联合密度函数, 并且 $f_{\mathbf{S}}(\mathbf{s})$ 连续、有上界, 有 $f_{\mathbf{S}}(\mathbf{s}) > 0$ 成立。

(C2) 对于任意的 i , 随机向量 $\mathbf{X}_i, \mathbf{Z}_i$ 的每个分量依概率有界。

(C3) ε 有累计分布函数 $F_\varepsilon(\cdot)$ 和密度函数 $f_\varepsilon(\cdot)$ 。对于每一个 p 维向量 \mathbf{u} ,

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \int_0^{u_0 + \boldsymbol{\Pi}_i \mathbf{u}} \sqrt{n} [F_\varepsilon(a + t/\sqrt{n}) - F_\varepsilon(a)] dt \\ = \frac{1}{2} f_\varepsilon(a) (u_0, \mathbf{u}^\top) \begin{pmatrix} 1 & \mathbf{0} \\ \mathbf{0} & \mathcal{D} \end{pmatrix} (u_0, \mathbf{u}^\top)^\top. \end{aligned}$$

(C4) 二元系数函数 $\beta_j(\cdot), j = 0, 1, \dots, q$ 均来自 Benstein 基多项式 $\{B_m, m \in \mathcal{M}\}^\top$ 张成的空间中的元素。

(C5) 三角剖分 T 是 π 拟均匀的, 即存在一个正常量 π 使得 $|T|/r_T \leq \pi$ 。

(C6) 三角剖分的大小满足 $|T| \rightarrow 0, |T|^{-2} \log n/n \rightarrow 0$ 。

(C7) 对于任意 $g \in \mathbb{S}_d^y$, $\Delta \in T$, 存在一个独立于 g 和 Δ 的正常数 \mathbb{F}_1 和在三角形 $\Delta \in T$ 中最大观测数量 \mathbb{F}_2 , 使得对于所有的 $\Delta \in T$, 有如下式子成立

$$\mathbb{F}_1 \|g\|_{\infty, \Delta} \leq \{\sum_{\mathbf{s}_i \in \Delta, i=1, \dots, n} g(\mathbf{s}_i)^2\}^{1/2} \leq \mathbb{F}_2 \|g\|_{\infty, \Delta},$$

其中, $\|g\|_{\infty, \Delta}$ 为 g 在三角形 Δ 上的无穷范数, 常量 \mathbb{F}_1 和 \mathbb{F}_2 满足 $\mathbb{F}_1/\mathbb{F}_2 = O(1)$ 。

条件(C1)和(C2)是非参数回归常用的假设, 如 Guo 等(2012), Mu 等(2018)。条件(C3)是分位回归估计中对误差项的有关假设(Koenker, 2005)。条件(C4)是非参数估计中对系数函数的限定和假设。条件(C5)意味着三角剖分后的各个三角形尽可能均匀, 这可以通过前文中描述的三角剖分选择方法实现。条件(C7)保证了离散样条的存在性, 并且限制了某一个三角形上有过多的观测值, (C5)和(C7)均来自 Mu 等(2018)。

令 Y_1, Y_2, \dots, Y_n 是独立的随机变量, α_0, ξ_0^* 表示未知待估参数的真实值. 假定条件 τ 分位函数为 $q_\tau(Y_i | \mathbf{X}_i, \mathbf{Z}_i, \mathbf{S}_i) = q_\tau(Y_i | \mathbf{X}_i, \mathbf{Z}_i^*) = \mathbf{X}_i^\top \alpha_\tau + \mathbf{Z}_i^{*\top} \xi_\tau^* = \mathbf{\Pi}_i^\top \boldsymbol{\theta}_\tau$, 其中 $\mathbf{\Pi}_i = (\mathbf{X}_i^\top, \mathbf{Z}_i^{*\top})^\top$, $\boldsymbol{\theta}_\tau = (\alpha_\tau, \xi_\tau^*)^\top$. 记 $\boldsymbol{\zeta} = (\boldsymbol{\eta}, \boldsymbol{\varphi})^\top$, 其中 $\boldsymbol{\eta} = \sqrt{n}(\alpha - \alpha_0)$, $\boldsymbol{\varphi} = \sqrt{n}(\boldsymbol{\xi}^* - \boldsymbol{\xi}_0^*)$, 即 $\boldsymbol{\varphi} = (\varphi_0^\top, \varphi_1^\top, \dots, \varphi_q^\top) \in \mathbb{R}^{|\mathcal{M}| \times (q+1)}$, $\varphi_j = (\varphi_{j1}, \varphi_{j2}, \dots, \varphi_{j|\mathcal{M}|}) \in \mathbb{R}^{|\mathcal{M}|}$, $\varphi_{jm} = \sqrt{n}(\xi_{jm}^* - \xi_{0;jm}^*)$.

引入 Koenker(2005)假设:

(D1) 存在正定矩阵 \mathcal{D} 和 \mathcal{D}_1 , 使得

$$(I) \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \mathbf{\Pi}_i^\top \mathbf{\Pi}_i = \mathcal{D}$$

$$(II) \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n f_{\epsilon, i}(\epsilon_i(\tau)) \mathbf{\Pi}_i^\top \mathbf{\Pi}_i = \mathcal{D}_1(\tau)$$

(D2) $\max_{i=1, 2, \dots, n} \|\mathbf{\Pi}_i\|_2 / \sqrt{n} \rightarrow 0$

定理 3 如果假设条件(A1-A7)、(B1-B2)成立. 若存在 $\bar{\lambda} > 0$, 使得当 $n \rightarrow \infty$ 时, $\lambda_{\max} / \sqrt{n} \rightarrow \bar{\lambda}$, 则

$$\sqrt{n} \begin{pmatrix} \hat{\alpha} - \alpha \\ \hat{\xi}^* - \xi^* \end{pmatrix} \xrightarrow{d} \mathcal{N} \left(\begin{pmatrix} \mathbf{0}_p \\ -\mathcal{D}_1^{-1} \nabla_{\mathcal{Q}_3}^1 \end{pmatrix}, \tau(1-\tau) \mathcal{D}_1^{-1} \mathcal{D} (\mathcal{D}_1^{-1})^\top \right),$$

其中 $\mathbf{0}_p$ 为 p 维的列零向量, $\mathcal{Q}_3 = 1/2\bar{\lambda} \mathbf{K} \sum_{j=0}^p \sum_{m \in \mathcal{M}} \varphi_{jm} \text{sgn}(\xi_{jm}^*) |\xi_{jm}^*|$ 且 $\mathbf{K} = \mathbf{Q}^{*\top} \mathbf{P} \mathbf{Q}^*$, $\nabla_{\mathcal{Q}_3}^1$ 为 \mathcal{Q}_3 关于 $\boldsymbol{\varphi} = \sqrt{n}(\boldsymbol{\xi}^* - \boldsymbol{\xi}_0^*)$ 的梯度向量.

证明 令

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) = \mathcal{L}(\alpha, \boldsymbol{\xi}^*) &= \sum_{i=1}^n \rho_\tau \left\{ Y_i - \sum_{l=1}^p X_{il} \alpha_l - \sum_{j=0}^q Z_{ij} (\mathbf{Q}^* \boldsymbol{\xi}_j^*)^\top \mathbf{B}(\mathbf{S}_i) \right\} \\ &\quad + \frac{1}{2} \sum_{j=0}^q \lambda_j (\mathbf{Q}^* \boldsymbol{\xi}_j^*)^\top \mathbf{P} \mathbf{Q}^* \boldsymbol{\xi}_j^*, \end{aligned}$$

并考虑目标函数 $\mathcal{Q}_n(\boldsymbol{\zeta}) = \mathcal{L} \left(\boldsymbol{\theta} + \frac{\boldsymbol{\zeta}}{\sqrt{n}} \right) - \mathcal{L}(\boldsymbol{\theta})$, 因为 $\hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta})$ 及凸性, 故 $\hat{\boldsymbol{\zeta}} = (\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\varphi}})^\top = \arg \min_{\boldsymbol{\zeta}} \mathcal{Q}_n(\boldsymbol{\zeta})$. 即 $\mathcal{Q}_n(\boldsymbol{\zeta})$ 的极限行为决定了 $\boldsymbol{\zeta}$ 的渐近分布. 为了简化记号令 $\mathbf{K} = \mathbf{Q}^{*\top} \mathbf{P} \mathbf{Q}^*$, 进一步有

$$\begin{aligned} \mathcal{Q}_n(\boldsymbol{\zeta}) &= \sum_{i=1}^n \left\{ \rho_\tau \left(Y_i - \mathbf{\Pi}_i^\top \boldsymbol{\theta} - \mathbf{\Pi}_i^\top \frac{\boldsymbol{\zeta}}{\sqrt{n}} \right) - \rho_\tau \left(Y_i - \mathbf{\Pi}_i^\top \boldsymbol{\theta} \right) \right\} \\ &\quad + \frac{1}{2} \sum_{j=0}^p \lambda_j \mathbf{K} \left\{ \left\| \boldsymbol{\xi}_j^* + \frac{\boldsymbol{\varphi}_j}{\sqrt{n}} \right\|_2^2 - \|\boldsymbol{\xi}_j^*\|_2^2 \right\} \\ &= \mathcal{Q}_{1n}(\boldsymbol{\theta}) + \mathcal{Q}_{2n}(\boldsymbol{\theta}) + \mathcal{Q}_{3n}(\boldsymbol{\theta}). \end{aligned}$$

类似于定理 1 的证明, 利用 Knight(1998)不等式, 根据 Cramer-Wald 定理和 Lindeberg-Feller 中心极限定理, 利用假设, 有 $Q_{1n}(\zeta) \xrightarrow{d} -\zeta W$, 其中 $W \xrightarrow{d} \mathcal{N}(\mathbf{0}, \tau(1-\tau)\mathbf{D})$ 。

对于 $Q_{2n}(\zeta)$, 令

$$Q_{2ni}(\zeta) = \int_0^{\mathbf{\Pi}_i^\top \zeta / \sqrt{n}} \{I(Y_i - \mathbf{\Pi}_i^\top \boldsymbol{\theta} \leq s) - I(Y_i - \mathbf{\Pi}_i^\top \boldsymbol{\theta} \leq 0)\} ds,$$

则 $Q_{2n}(\zeta)$ 可以分解为 $Q_{2n}(\zeta) = \sum_{i=1}^n EQ_{2ni}(\zeta) + \sum_{i=1}^n \{Q_{2ni}(\zeta) - EQ_{2ni}(\zeta)\}$, 其中 $\sum_{i=1}^n EQ_{2ni}(\zeta) \xrightarrow{p} \frac{1}{2} \zeta^\top \mathbf{D}_1 \zeta$, 考虑方差满足 $Var\{Q_{2n}(\zeta)\} = o_p(1)$ 。

对于 $Q_{3n}(\zeta)$, 类似于 Andriyana 等(2014), 根据 $\lambda_{\max} = \{\lambda_j, j = 0, 1, \dots, q\}$, 如果当 $n \rightarrow \infty$ 时有 $\lambda_{\max}/\sqrt{n} \rightarrow \bar{\lambda} \geq 0$, 则

$$Q_{3n}(\zeta) \rightarrow \frac{1}{2} \bar{\lambda} \mathbf{K} \sum_{j=0}^q \sum_{m \in \mathcal{M}} \varphi_{j,m} \text{sgn}(\xi_{j,m}^*) |\xi_{j,m}^*| = Q_3,$$

进一步令 $\nabla_{Q_3}^1$ 为 Q_3 关于 $\boldsymbol{\varphi} = \sqrt{n}(\boldsymbol{\xi}^* - \boldsymbol{\xi}_0^*)$ 的梯度向量, 且 $\nabla_{Q_3}^1$ 独立于 $\boldsymbol{\varphi}$ 。

综上, 可得结论

$$\sqrt{n} \begin{pmatrix} \hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha} \\ \hat{\boldsymbol{\xi}}^* - \boldsymbol{\xi}^* \end{pmatrix} \xrightarrow{d} \mathcal{N} \left(\begin{pmatrix} \mathbf{0}_p \\ -\mathbf{D}_1^{-1} \nabla_{Q_3}^1 \end{pmatrix}, \tau(1-\tau) \mathbf{D}_1^{-1} \mathbf{D} (\mathbf{D}_1^{-1})^\top \right).$$

证毕。

5.4 算法与实现

5.4.1 SPLVCMQR-ADMM 算法

这里给出基于 ADMM 算法的 SPLVCM-BPBQR 估计实现, 记为 SPLVCMQR-ADMM 算法。

令 $\epsilon_i = \epsilon_i(\boldsymbol{\alpha}, \boldsymbol{\xi}^*) = Y_i - \sum_{l=0}^p X_{il} \alpha_l - \sum_{j=0}^q Z_{ij} (\mathbf{Q}^* \boldsymbol{\xi}_j^*)^\top \mathbf{B}(\mathbf{S}_i)$, $i = 1, 2, \dots, n$ 为部分线性空间变系数模型的分位回归估计的残差项, 并且为了方便记号, 将惩罚项改写为 $\mathcal{J}_\lambda(\boldsymbol{\xi}^*) = 1/2 \sum_{j=0}^q \lambda_j (\mathbf{Q}^* \boldsymbol{\xi}_j^*)^\top \mathbf{P} \mathbf{Q}^* \boldsymbol{\xi}_j^* = 1/2 \boldsymbol{\lambda} \otimes \boldsymbol{\xi}^{*\top} \mathbf{Q}^{*\top} \mathbf{P} \mathbf{Q}^* \boldsymbol{\xi}^*$, 其中光滑参数 $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \dots, \lambda_q)^\top$, 类似地, 这里也应用 ADMM 算法将正则分位回归(5.11)重新表述为凸优化问题(5.13):

$$\begin{aligned} \min_{\alpha \in \mathbb{R}^{p+1}, \xi^* \in \mathbb{R}^{|\mathcal{M}| \times (p+1)}} & \sum_{i=1}^n \rho_{\tau}(\epsilon_i) + \mathcal{J}_{\lambda}(\xi^*) \\ \text{subject to} & \mathbf{X}^{\top} \alpha + \mathbf{Z}^{*\top} \xi^* + \epsilon = \mathbf{Y} \end{aligned} \quad (5.13)$$

其中 $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)^{\top}$, $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^{\top}$, $\mathbf{Z}^* = (\mathbf{Q}^* \mathbf{Z}_0^*, \mathbf{Q}^* \mathbf{Z}_1^*, \dots, \mathbf{Q}^* \mathbf{Z}_p^*)^{\top}$, $\mathbf{Z}_j^* = (Z_{1j} \mathbf{B}(\mathbf{S}_1), Z_{2j} \mathbf{B}(\mathbf{S}_2), \dots, Z_{nj} \mathbf{B}(\mathbf{S}_n))$, $\xi^* = (\xi_0^*, \xi_1^*, \dots, \xi_p^*)^{\top}$.

下面是惩罚分位回归估计 SPLVCM-BPBQR 的 ADMM 算法。假设迭代到第 t 步, 则第 $t+1$ 步迭代机制为

$$\begin{aligned} \epsilon^{(t+1)} &= \arg \min_{\epsilon \in \mathbb{R}^n} \sum_{i=1}^n \rho_{\tau}(\epsilon_i) + \frac{\hbar}{2} \left\| \mathbf{Y} - \epsilon - \mathbf{X}^{\top} \alpha^{(t)} - \mathbf{Z}^{*\top} \xi^{*(t)} + \frac{\mathbf{u}^{(t)}}{\hbar} \right\|_2^2 \\ (\alpha^{(t+1)\top}, \xi^{*(t+1)}) &= \arg \min_{\alpha \in \mathbb{R}^{p+1}, \xi^* \in \mathbb{R}^{|\mathcal{M}| \times (q+1)}} \\ & \frac{\hbar}{2} \left\| \mathbf{Y} - \epsilon^{(t+1)} - \mathbf{X}^{\top} \alpha - \mathbf{Z}^{*\top} \xi^* + \frac{\mathbf{u}^{(t)}}{\hbar} \right\|_2^2 + \mathcal{J}_{\lambda}(\xi^*) \\ \mathbf{u}^{(t+1)} &= \mathbf{u}^{(t)} + \hbar(\mathbf{Y} - \epsilon^{(t+1)} - \mathbf{X}^{\top} \alpha^{(t+1)} - \mathbf{Z}^{*\top} \xi^{*(t+1)}), \end{aligned} \quad (5.14)$$

其中, μ 缩放后的为拉格朗日乘数, \hbar 是一个惩罚参数。Boyd 等(2011)建议取值 $\hbar = 1.2$ 。下面引进软门限算子来简化(5.14)式中 ϵ 的更新。即

$$\epsilon^{(t+1)} = \mathcal{S}_{\frac{1}{\hbar}} \left(\mathbf{Y} - \mathbf{X}^{\top} \alpha^{(t)} - \mathbf{Z}^{*\top} \xi^{*(t)} + \frac{\mathbf{u}^{(t)}}{\hbar} - (2\tau_{n \times 1} - \mathbf{1}_{n \times 1})/\hbar \right), \quad (5.15)$$

其中, 软门限算子 $\mathcal{S}_{\delta} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ 定义为 $(\mathcal{S}_{\delta}(\mathbf{v}))_i = (v_i - a)_+ - (-v_i - a)_+$ 。

同样地, 对 (α^{\top}, ξ^*) 的更新虽然没有类似的形式, 但可以看作是一个具有 L_2 惩罚的最小二乘优化问题, 即岭回归问题, 可以利用现有的数值方法来解决这个问题, 并更新 ξ^*

$$(\alpha^{(t+1)}, \xi^{*(t+1)}) = \left(\mathbf{\Pi}^{\top} \mathbf{\Pi} + \begin{pmatrix} \mathbf{0}_{p \times p} & \mathbf{0} \\ \mathbf{0} & \lambda \otimes \mathbf{Q}^{*\top} \mathbf{P} \mathbf{Q}^* \end{pmatrix} / \hbar \right)^{-1} \mathbf{\Pi}^{\top} \left(\mathbf{Y} - \epsilon^{(t+1)} + \frac{\mathbf{u}^{(t)}}{\hbar} \right), \quad (5.16)$$

其中, $\mathbf{\Pi} = (\mathbf{X}^{\top}, \mathbf{Z}^{*\top})^{\top}$ 注意到 $\mathbf{\Pi}^{\top} \mathbf{\Pi} + \lambda \otimes \mathbf{Q}^{*\top} \mathbf{P} \mathbf{Q}^* / \hbar$ 的逆只需要计算一次, 所以 ADMM 迭代过程会非常快速且有效(Kong, 2015)。

类似地, 用原残差 ϵ_{pri} 和对偶残差 ϵ_{dual} (Boyd, 2011) 定义 ADMM 算法的终止条件。在第 k 次迭代原残差 $\epsilon_{pri}^{(k)}$ 和对偶残差 $\epsilon_{dual}^{(k)}$ 的计算公式为

$$\begin{aligned} \epsilon_{pri}^{(k)} &= \mathbf{Y} - \mathbf{X}^{\top} \alpha^{(k)} - \mathbf{Z}^{*\top} \xi^{*(k)} - \epsilon^{(k)} \\ \epsilon_{dual}^{(k)} &= \hbar \mathbf{\Pi}^{\top} (\Theta^{(k)} - \Theta^{(k-1)}), \end{aligned} \quad (5.17)$$

相应地, 算法终止条件为

$$\begin{aligned} \|\boldsymbol{\epsilon}_{pri}^{(k)}\|_2 &\leq e_{pri}, \quad \|\boldsymbol{\epsilon}_{dual}^{(k)}\|_2 \leq e_{dual}, \\ e_{pri} &= \sqrt{n}e_{abs} + e_{rel} \max \left\{ \|\boldsymbol{\Pi}^\top \boldsymbol{\Theta}^{(k)}\|_2^2, \|\boldsymbol{\epsilon}^{(k)}\|_2^2, \|\mathbf{Y}\|_2^2 \right\} \\ e_{dual} &= \sqrt{|\mathcal{M}|} \times pe_{abs} + e_{rel} \|\boldsymbol{\Pi}^\top \mathbf{u}^{(k)}\|_2^2, \end{aligned}$$

其中, $\boldsymbol{\Theta}^{(k)} = (\boldsymbol{\alpha}^{(k)\top}, \boldsymbol{\xi}^{*(k)\top})$, e_{abs} 和 e_{rel} 为绝对误差限和相对误差限, 在实际计算中, 可以指定任意小的数。这里取 $e_{abs} = 10^{-4}$, $e_{rel} = 10^{-2}$ (Boyd 等, 2011)。

5.4.2 光滑参数的选取

光滑参数向量 $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \dots, \lambda_p)^\top$ 的选择关系到数据拟合优度和系数函数 $\beta_j(\mathbf{s})$ 有效性之间的权衡, 一般可以利用交叉核实(CV)和广义交叉核实(GCV)来选取。考虑到 CV 会有数据欠光滑的风险, 本文选取 GCV 方法, 即在一个网格上最小化如下 GCV 准则:

$$GCV(\boldsymbol{\lambda}) = \frac{n^{-1} \|\mathbf{Y} - \mathbf{L}(\boldsymbol{\lambda})\mathbf{Y}\|^2}{\{n^{-1} \text{tr}(\mathbf{I} - \mathbf{L}(\boldsymbol{\lambda}))\}^2}, \quad (5.18)$$

其中 $\mathbf{L}(\boldsymbol{\lambda}) = \boldsymbol{\Pi} \left(\boldsymbol{\Pi}^\top \boldsymbol{\Pi} + \begin{pmatrix} \mathbf{0}_{p \times p} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\lambda} \otimes \mathbf{Q}^{*\top} \mathbf{P} \mathbf{Q}^* \end{pmatrix} \right)^{-1} \boldsymbol{\Pi}^\top$, $\text{tr}(\cdot)$ 表示矩阵的迹。

5.5 数值模拟

本小节同通过数值模拟来验证本章所提出方法的优良性能。在数值模拟中, 依然考虑 Wood 等(2008)构造的一种在矩形 $[-1, 3.5] \times [-1, 1]$ 内的修正马蹄形区域, 它被均匀的分为 $N_s = 401 \times 901$ 个网格点, 函数只在马蹄形区域内取值。这里假设数据生成机制来自以下模型:

$$Y_i = -0.5X_1 + 1.5X_2 + \beta_0(\mathbf{S}_i)Z_0 + \beta_1(\mathbf{S}_i)Z_1 + \varepsilon_i, i = 1, 2, \dots, n,$$

其中, 二元系数函数的设定也与前文一致。 $\beta_0(s_1, s_2)$ 选取 Wood 等(2008)中的函数, $\beta_1(s_1, s_2)$ 选取 Kim & Wang(2021)中的函数。协变量 $X_1 \sim N(0, 1)$, $X_2 \sim N(0, 2^2)$, $Z_j \sim U(0, 1), j = 0, 1$ 。为了说明 SPLVCM-BPBQR 估计方法的稳健性, 需要模拟误差分布为尖峰、厚尾或者非对称的情形。这里考虑 3 种误差情形: 混合正态分布 $0.9N(0, 0.5^2) + 0.1N(0, 10^2)$ 、学生 t 分布 $t(3)$ 、拉普拉斯分布 $Laplace(0, 1, 1)$ 。其中, 拉普拉斯分布误差利用 R 包 GeneralizedHyperbolic 中的函数 $rskewlap()$ 产生。需要说明, 常见于其他文献的误差分布正态分布和柯西分布,

由于篇幅限制不再展示模拟结果。在柯西误差条件下，均值回归往往会失效。而在正态误差条件下，最小二乘结果要略微好于分位回归估计结果。为了考虑所提出的方法在不同样本量情形下的表现，设置三种样本量： $n_1 = 800, n_2 = 2000, n_3 = 4000$ 。对于每一种情形，模拟 100 次。

此外，这里采用的三角剖分 T 也与前文一致(有 77 个三角形，65 个顶点)。为了考虑惩罚样条展开阶数对分位回归估计的影响，在非参数部分模拟中考虑设定阶数 $d = 2$ 和 3 。其他设定为：光滑度 $\nu = 1$ ，惩罚参数 (λ_0, λ_1) 通过 GCV 选择。为了与其他估计方法比较，考虑普通最小二乘估计 (SPLVCM-OLS) 和本文方法 SPLVCM-BPBQR 的比较。其中，最小二乘估计由最小化如下式子可得

$$\sum_{i=1}^n \left\{ Y_i - \sum_{l=0}^p X_{il} \alpha_l - \sum_{j=0}^q Z_{ij} (\mathbf{Q}^* \boldsymbol{\xi}_j^*)^\top \mathbf{B}(\mathbf{S}_i) \right\}^2 + \frac{1}{2} \sum_{j=0}^q \lambda_j (\mathbf{Q}^* \boldsymbol{\xi}_j^*)^\top \mathbf{P} \mathbf{Q}^* \boldsymbol{\xi}_j^*.$$

为了便于比较各估计量的估计精度，这里计算关于参数的一些统计量。对于线性部分，计算线性系数估计量的均方误差 (SE) 和绝对偏差 (AE)。针对变系数部分，计算系数函数估计量的平均积分均方误差 MISE 和平均积分绝对误差 MASE (计算公式详见 3.5 节)。

表 11 $0.9N(0, 0.5^2) + 0.1N(0, 10^2)$ 误差线性部分模拟结果

n	方法	SE		AE	
		α_1	α_2	α_1	α_2
n_1	OLS	6.623E-03	1.791E-03	0.0659	0.0328
		(8.966E-03)	(2.854E-03)	(0.0480)	(0.0269)
n_1	BPBQR	5.562E-04	1.654E-04	0.0193	0.0101
		(7.059E-04)	(2.488E-04)	(0.0136)	(0.0080)
n_2	OLS	2.593E-03	7.168E-04	0.0412	0.0193
		(3.441E-03)	(1.421E-03)	(0.0301)	(0.0186)
n_2	BPBQR	2.648E-04	5.548E-05	0.0130	0.0060
		(3.761E-04)	(7.375E-05)	(0.0098)	(0.0044)
n_3	OLS	1.143E-03	3.615E-04	0.0274	0.0149
		(1.405E-03)	(5.605E-04)	(0.0198)	(0.0119)
n_3	BPBQR	1.016E-04	2.335E-05	0.0082	0.0038
		(1.414E-04)	(3.572E-05)	(0.0058)	(0.0029)

表 11-表 13 展示了三种假设误差分布下，线性部分系数估计的模拟结果，其中括号里是相应的统计量的标准差(下同)。其中均方误差 SE 采用科学计数法(保留四位有效数字)，绝对误差 AE 保留四位小数。可以看出，随着样本量的增加，

无论是普通最小二乘还是本文提出的方法,都趋向于更好的结果。在样本量一定的条件下,BPBQR估计结果要远远优于OLS所估计的结果,各统计量的标准差的对比结果说明所提出的方法不仅更加精确,而且要比最小二乘稳健的多。从三种不同的误差分布看,在混合正态分布误差下,即存在大量离群点的下,BPBQR估计量的优越性最为明显。这再一次说明了本文方法对异常值不敏感,是一种相对稳健的估计方法。

表 12 $t(3)$ 误差线性部分模拟结果

n	方法	SE		AE	
		α_1	α_2	α_1	α_2
n_1	OLS	3.764E-03 (5.636E-03)	1.156E-03 (1.634E-03)	0.0485 (0.0378)	0.0274 (0.0203)
	BPBQR	2.363E-03 (3.249E-03)	6.963E-04 (9.505E-04)	0.0397 (0.0281)	0.0214 (0.0156)
n_2	OLS	1.665E-03 (2.308E-03)	3.514E-04 (3.794E-04)	0.0319 (0.0255)	0.0159 (0.0099)
	BPBQR	1.280E-03 (1.750E-03)	2.589E-04 (3.343E-04)	0.0285 (0.0217)	0.0130 (0.0095)
n_3	OLS	7.809E-04 (1.084E-03)	1.931E-04 (2.450E-04)	0.0217 (0.0177)	0.0114 (0.0080)
	BPBQR	5.020E-04 (7.178E-04)	1.217E-04 (1.441E-04)	0.0176 (0.0139)	0.0091 (0.0062)

表 13 $Laplace(0, 1, 1)$ 误差线性部分模拟结果

n	方法	SE		AE	
		α_1	α_2	α_1	α_2
n_1	OLS	3.067E-03 (3.639E-03)	6.617E-04 (8.180E-04)	0.0459 (0.0312)	0.0210 (0.0150)
	BPBQR	1.924E-03 (2.706E-03)	4.380E-04 (5.447E-04)	0.0359 (0.0254)	0.0168 (0.0125)
n_2	OLS	1.049E-03 (1.561E-03)	2.636E-04 (3.318E-04)	0.0252 (0.0204)	0.0132 (0.0095)
	BPBQR	6.515E-04 (1.049E-03)	1.753E-04 (1.995E-04)	0.0198 (0.0162)	0.0111 (0.0073)
n_3	OLS	5.858E-04 (9.350E-04)	1.131E-04 (1.578E-04)	0.0187 (0.0154)	0.0086 (0.0063)
	BPBQR	3.864E-04 (7.519E-04)	6.446E-05 (1.117E-04)	0.0145 (0.0133)	0.0059 (0.0054)

表 14-表 16 展示了非参数部分数值模拟的结果。可以看出样本量越大，估计精度越高。此外，展开阶数，也对估计量的精度有较大的影响。整体而言，阶数越高，拟合精度越好。在最小二乘估计中，由于误差项为非正态分布，所以阶数的影响在部分模拟下与上述结论不一致（如混合正态误差，样本量为 800 时，3 阶逼近结果要比 2 阶差）。在系数 $\beta_2(\mathbf{s})$ 的估计结果中，部分估计量的标准差在 3 阶情况下要略大于 2 阶（如 Laplace 误差下，MISE 和 MASE 估计量的标准差）。

表 14 $0.9N(0, 0.5^2) + 0.1N(0, 10^2)$ 误差非参部分模拟结果

n	方法	d	$\beta_0(\mathbf{s})$		$\beta_1(\mathbf{s})$	
			MISE	MASE	MISE	MASE
n_1	OLS	2	6.1212(7.3734)	3.5296(1.3453)	5.5224(7.1358)	3.0645(1.4025)
		3	18.4962(28.3041)	4.3850(2.0929)	13.9149(4.0413)	4.0413(2.1959)
	BPBQR	2	0.4459(0.1707)	1.3082(0.2403)	0.3826(0.1338)	1.2306(0.2282)
		3	0.3776(0.1679)	1.2005(0.2497)	0.3357(0.1331)	1.1460(0.2384)
n_2	OLS	2	2.7996(3.3441)	2.4392(0.7940)	2.8974(4.1575)	2.2805(0.8631)
		3	6.3840(6.9142)	2.7708(0.9443)	6.4902(7.1543)	4.6432(0.9916)
	BPBQR	2	0.2361(0.0519)	0.9364(0.1282)	0.2049(0.0563)	0.8862(0.1329)
		3	0.1894(0.0514)	0.8451(0.1314)	0.1721(0.0521)	0.8213(0.1234)
n_3	OLS	2	1.4051(1.5211)	1.7816(0.4940)	1.2330(1.1998)	1.6616(0.4682)
		3	3.3894(4.7478)	2.0050(0.6725)	2.8460(3.0618)	1.8919(0.6136)
	BPBQR	2	0.1569(0.0306)	0.7442(0.0907)	0.1306(0.0308)	0.6904(0.0983)
		3	0.1209(0.0097)	0.6653(0.0897)	0.1082(0.0296)	0.6379(0.0963)

表 15 $t(3)$ 误差非参部分模拟结果

n	方法	d	$\beta_0(\mathbf{s})$		$\beta_1(\mathbf{s})$	
			MISE	MASE	MISE	MASE
n_1	OLS	2	1.9312(1.5251)	2.7356(0.6020)	1.3118(0.7606)	2.3104(0.5213)
		3	3.0427(10.0443)	2.7402(1.4288)	5.7025(43.3518)	2.4044(1.6447)
	BPBQR	2	1.0916(0.4144)	2.0953(0.4364)	0.9785(0.3935)	1.9952(0.4020)
		3	1.0148(0.4117)	2.0054(0.4403)	0.9575(0.4013)	1.9646(0.4178)
n_2	OLS	2	0.9167(0.2824)	1.9283(0.3262)	0.6970(0.2404)	1.7007(0.3192)
		3	0.7932(0.2557)	1.8011(0.3056)	0.6400(0.2388)	1.6251(0.3206)
	BPBQR	2	0.5770(0.1794)	1.5108(0.2557)	0.5004(0.1935)	1.4107(0.2744)
		3	0.5246(0.1800)	1.4458(0.2623)	0.4836(0.1941)	1.3876(0.2742)
n_3	OLS	2	0.5850(0.1996)	1.5207(0.2726)	0.4650(0.1416)	1.3901(0.2385)
		3	0.5032(0.1801)	1.4118(0.2473)	0.4145(0.1231)	1.3146(0.2203)
	BPBQR	2	0.3537(0.0878)	1.1780(0.1682)	0.3132(0.0911)	1.1135(0.1758)
		3	0.3121(0.0856)	1.1100(0.1733)	0.2905(0.0905)	1.0754(0.1801)

表 16 $Laplace(0, 1, 1)$ 误差非参部分模拟结果

n	方法	d	$\beta_0(s)$		$\beta_1(s)$	
			MISE	MASE	MISE	MASE
n_1	OLS	2	1.3461(0.4427)	2.3671(0.4188)	1.0151(0.4316)	2.0491(0.4269)
		3	1.1877(0.4595)	2.2210(0.4399)	0.9266(0.4362)	1.9535(0.4519)
	BPBQR	2	0.8781(0.2990)	1.8779(0.3387)	0.7199(0.3231)	1.6845(0.3642)
		3	0.7891(0.2900)	1.7739(0.3472)	0.6955(0.3334)	1.6440(0.3728)
n_2	OLS	2	0.6603(0.2312)	1.6246(0.3128)	0.5462(0.1858)	1.4995(0.2664)
		3	0.5827(0.2134)	1.5337(0.3100)	0.4985(0.1810)	1.4338(0.2678)
	BPBQR	2	0.4270(0.1435)	1.2860(0.2556)	0.3857(0.1372)	1.2258(0.2289)
		3	0.3771(0.1384)	1.2122(0.2233)	0.3616(0.1344)	1.1817(0.2239)
n_3	OLS	2	0.4020(0.1283)	1.2421(0.2174)	0.3391(0.1039)	1.1668(0.1916)
		3	0.3549(0.1432)	1.1725(0.2513)	0.3069(0.1074)	1.1135(0.2044)
	BPBQR	2	0.2719(0.0725)	1.0188(0.1531)	0.2329(0.0730)	0.9521(0.1482)
		3	0.2379(0.0698)	0.9632(0.1525)	0.2163(0.0742)	0.9171(0.1514)

在模拟中，还记录了不同估计量在不同条件设置下 100 次模拟的平均运行时间。如表 17 所示，这里以混合正态分布误差下的运行时间为例进行说明。从结果可以看出，所提出的估计算法平均运行时间皆比普通的 OLS 算法的少得多。随着样本量的增加和展开阶数的增加，这种优势愈发明显。尤其当样本量较大阶数较高时($n=2000, 4000, d=3$)，OLS 平均估计时间是 BPBQR 估计的 3-5 倍甚至更多。

表 17 $0.9N(0, 0.5^2) + 0.1N(0, 10^2)$ 误差下的运算时间

n	d	OLS	BPBQR
n_1	2	2.5878	1.9366
	3	16.7989	8.0694
n_2	2	7.0823	3.4824
	3	53.6311	9.2962
n_3	2	21.9970	7.3675
	3	132.0833	31.7689

如图 7 所示，在模拟中也展示了非参数系数函数的可视化估计结果。整体而言，无论是最小二乘还是分位回归估计，两种方法 β_0 的估计结果要优于 β_1 。进一步可以明显看出，分位回归估计得到的系数函数曲面更接近于真实曲面，BPBQR 估计要明显优于 OLS，而 OLS 估计几乎已经无法刻画研究的“真相”，接近失效。

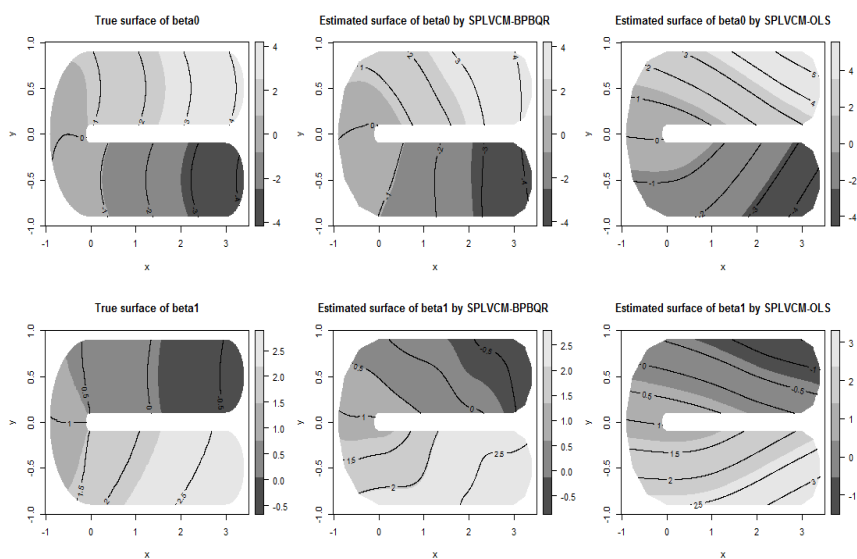


图 7 非参数函数 $\beta_0(\cdot), \beta_1(\cdot)$ 的真实曲面(左)、BPBQR 估计(中)、OLS 估计(右)

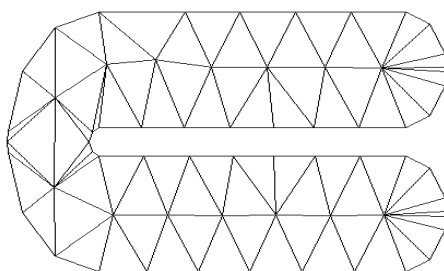


图 8 马蹄形区域三角剖分

注:图 8 数据来源于 Kim & Wang(2021)附件: <https://www.tandfonline.com/doi/suppl/10.1080/10618600.2020.1754225>

6 实际数据应用

6.1 失业率数据分析

失业率问题的研究一直是相关科学领域的热点话题，它反映着一个地区的经济问题，而且失业率过高，会造成严重的社会问题。研究区域人口基本特征与反映失业情况的失业率之间是否有空间非平稳性的影响关系具有一定的现实意义，如各地区相关部门可以根据区域性非平稳关系、根据不同的人口特征，制定相应的就业鼓励政策。

本小节利用佛罗里达州 2014 年的一个实际数据集进行分析以说明所提出的估计量的应用价值。该数据集的另一种形式，包含交通数据交通事故发生频率和车辆里程数，被 Kim & Wang(2021)用来说明其方法的应用，并且其结果已经说明空间变系数模型比其他方法更有效。该数据集包括 11249 个人口普查区块组和 5 个变量，这些数据均来自美国人口普查局。对于每一个人口普查区组，协变量有总人口(AP)的对数值，男性比例(MR)，拉美裔人口比例(HR)，65 岁及以上人口比例(OR)和失业率(UR)。

6.1.1 模型与评价

现在主要研究失业率(UR)和协变量 4 个协变量有总人口(AP)的对数值，男性比例(MR)，拉美裔人口比例(HR)及 65 岁及以上人口比例(OR)的影响关系。由于数据的非负性质，建模之前首先男性比例(MR)，拉美裔人口比例(HR)，65 岁及以上人口比例(OR)和失业率(UR)对数化。协变量 X 对数化具体变换为 $X_i^* = \log(X_i + 0.001)$ 。

对于本文所关注的问题，设定如下模型：

(1)空间变系数模型(SVCM)，即模型(6.1)。分别考虑所提出的分位回归估计(SVCMQR)和复合分位回归估计(SVCMCQR)。

$$UR_i^* = \beta_0(\mathbf{S}_i) + \beta_1(\mathbf{S}_i)AP_i + \beta_2(\mathbf{S}_i)MR_i^* + \beta_3(\mathbf{S}_i)HR_i^* + \beta_4(\mathbf{S}_i)OR_i^* + \epsilon_i, \quad (6.1)$$

(2)全局线性回归模型(LM)，即模型(6.2)。分别考虑普通最小二乘估计(LMOLS)和分位回归估计(LMQR)。

$$UR_i^* = \beta_0 + \beta_1 AP_i + \beta_2 MR_i^* + \beta_3 HR_i^* + \beta_4 OR_i^* + \epsilon_i, \quad (6.2)$$

其中 $\beta_0(\cdot)$ 为待估的二元截距函数，可称为基准曲面， $\beta_1(\cdot), \dots, \beta_4(\cdot)$ 分别为协变量 AP, MR, HR, OR 相应的待估二元函数。

在具体应用中，考虑与 Kim & Wang(2021)一致的佛罗里达州区域三角剖分(见图 9)。其他设定为：惩罚样条展开阶数 $d = 3$ ，惩罚参数 λ 由 GCV 准则得到。

在模型(6.1)中，可以将协变量 OR 定义为老龄化程度。在分位回归估计中考虑三个分位水平：0.1, 0.5, 0.9 来比较协变量不同分位水平对失业率的影响关系变化情况，相应的估计量记为 $SVCMQR_{0.1}$, $SVCMQR_{0.5}$, $SVCMQR_{0.9}$ 。对于复合分位回归估计量，这里根据模拟结果考虑三个分位水平，估计量记为 $SVCMCQR_3$ 。对于模型(6.2)的拟合，将其普通最小二乘估计量和 0.5 分位水平的分位回归估计量记为 LMOLS 和 LMQR0.5。

表 11 汇总了各模型均方误差 MSE 和绝对偏差 MAE，结果如下

表 11 模型误差比较

方法	$SVCMQR_{0.1}$	$SVCMQR_{0.5}$	$SVCMQR_{0.9}$	$SVCMCQR_3$	LMOLS	LMQR0.5
MSE	2.1659	2.0622	2.2028	13.1389	2.1045	2.3967
MAE	0.9205	0.8950	0.9221	3.3447	1.0090	0.9575

可以看出，在该数据应用下，空间变系数 0.5 分位回归估计 $SVCMQR_{0.5}$ 估计的拟合精度优于 BPBCQR 估计量，所以下面就分位回归估计展开分析。

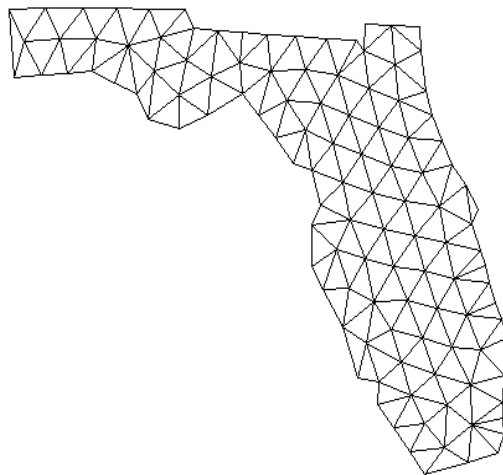


图 9 佛罗里达州三角剖分

注:图 11 数据来源于 Kim & Wang(2021)附件: <https://www.tandfonline.com/doi/suppl/10.1080/10618600.2020.1754225>

6.1.2 拟合优度检验

根据 6.2.1 节中三种模型的精度评估, 不难发现空间变系数分位回归模型在该数据集下表现良好。所以, 这里进一步通过假设检验进行验证。下面展示对模型(6.2)进行了拟合优度检验结果。

利用实际数据集的全部样本, 产生 100 个 Bootstrap 样本, 设定检验显著性水平为 0.05。建立原假设

$$H_0: \beta_j(\mathbf{s}) = \beta_j(\mathbf{s}; \boldsymbol{\vartheta}) \text{ v.s. } H_1: \beta_j(\mathbf{s}) \neq \beta_j(\mathbf{s}; \boldsymbol{\vartheta}), 0 \leq j \leq 1.$$

经检验, 在 0.05 的显著性水平下, 观测统计量的值为 0.03821, Bootstrap 样本经验累积分布如图 10 所示, 检验 p 值显著为 0.000。故有充分的理由认为模型(6.1)的设定是合理的。

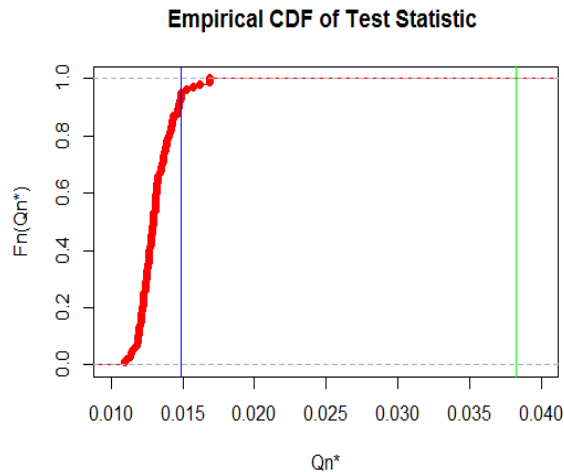


图 10 检验统计量的 ECDF: 图中蓝色标记上 0.05 分位数; 绿色标记观测统计量的值

下面, 采用模型(6.1)的拟合结果, 分析所考虑的因素对失业率的影响关系以及非平稳效应等。

6.1.3 结果分析

利用分位回归模型和估计方法自身的优势, 本文选取三个分位水平: 即 0.1(低分位区), 0.5(中分位区), 0.9(高分位区)。

图 11-13 分别展示了 $\beta_0(\cdot), \dots, \beta_4(\cdot)$ 在 0.1, 0.5, 0.9 分位水平下的估计情况。可以看出各系数函数均表现出了不同的空间非平稳性, 其次, 在不同分位水平下,

各系数的正负影响关系保持了基本一致性，而区别在于不同分位水平下，空间非平稳效应的振幅和大小发生了一定的改变。故以下分析以 0.5 分位水平为例。

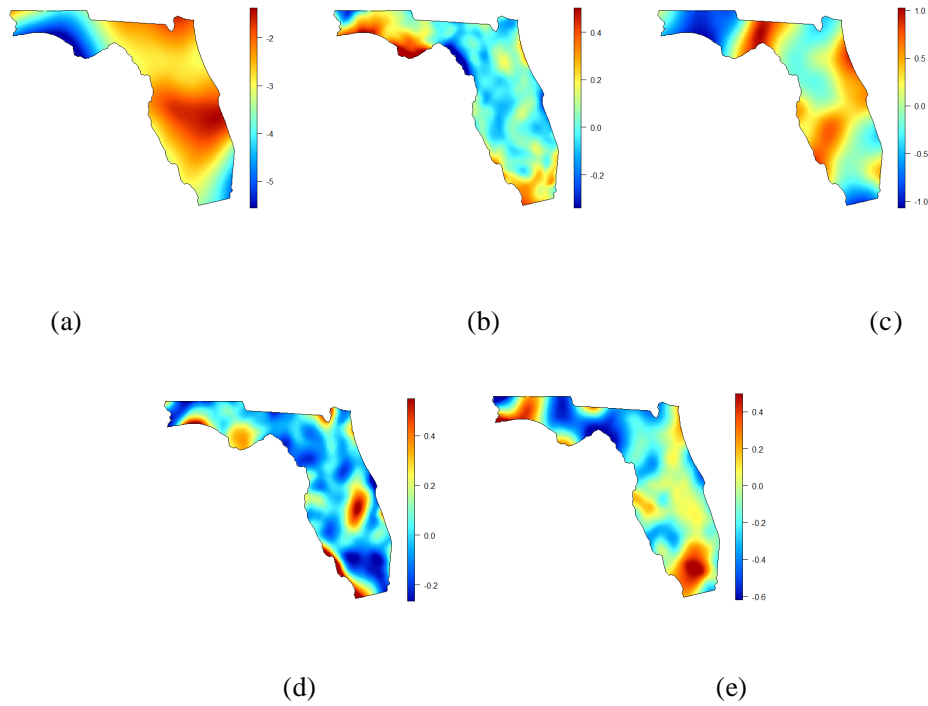


图 11 0.1 分位水平各系数函数估计图：(a-e)分别代表常数和协变量 AP, MR*, HR*, OR* 的二元系数函数 $\beta_0(\cdot), \dots, \beta_4(\cdot)$ 的估计量(下同)

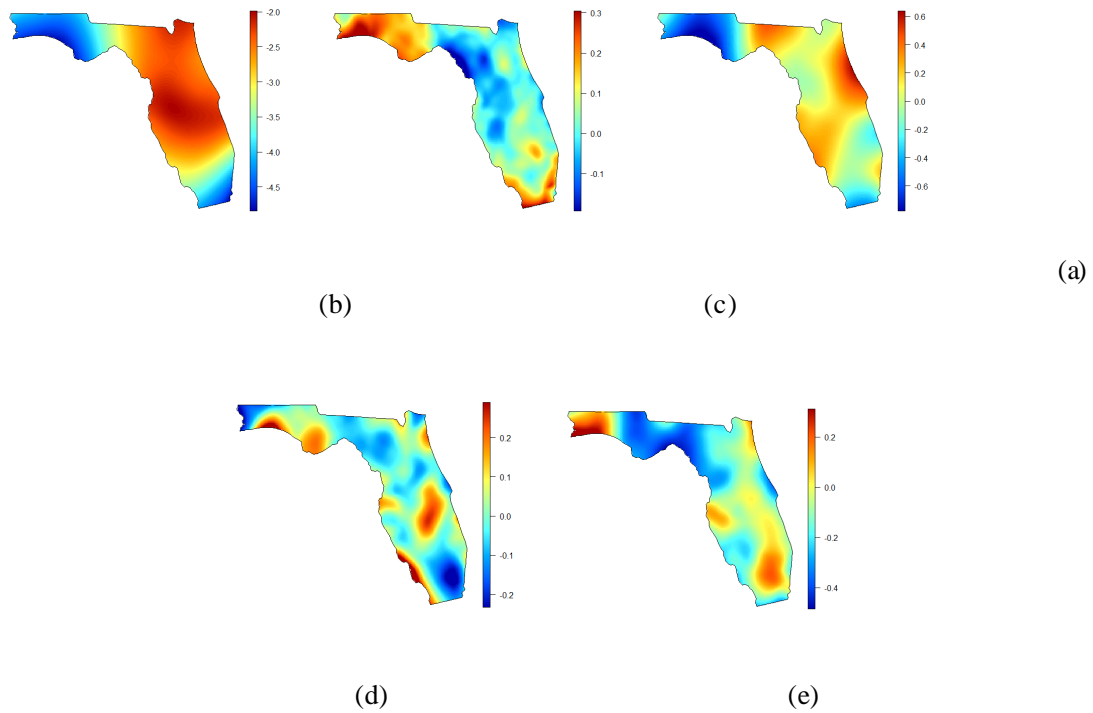


图 12 0.5 分位水平各系数函数估计图

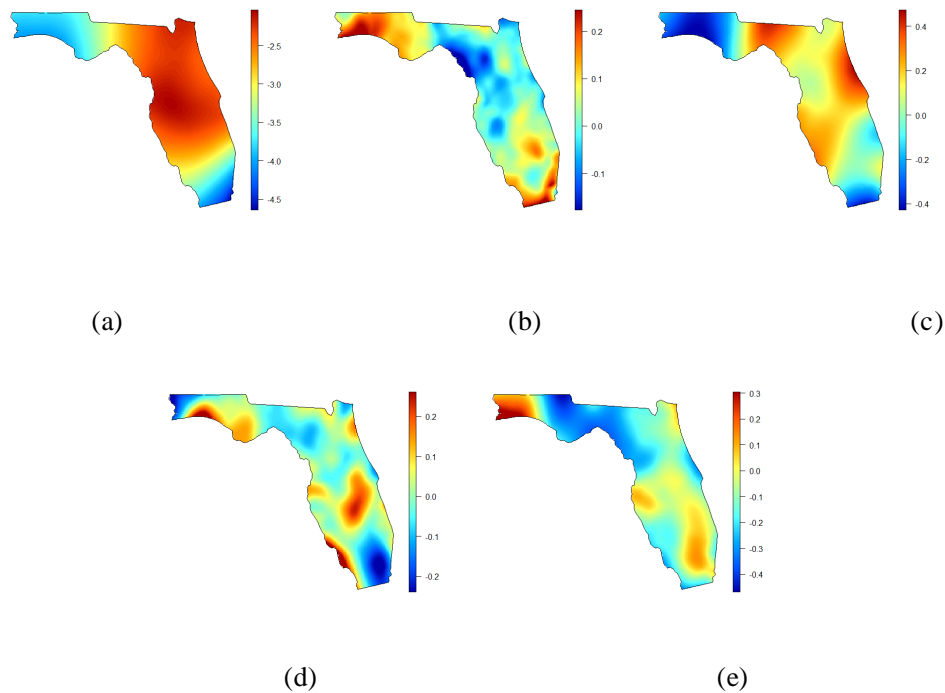


图 13 0.9 分位水平各系数函数估计图

图 12 显示了中位水平的协变量对失业率的影响变化情况。从图 12(a)可以看出，在佛罗里达州的彭萨克拉地区和塔拉哈西以西地区以及迈阿密周边地区失业率较低。中部地区如杰克逊维尔、奥兰多、开普科勒尔等失业率相对较高。图 12(b) 是总人口的对数值系数表现出的空间效应。彭萨克拉地区及其沿海一带、塔拉哈西等地区、迈阿密南部表现为正向作用。图 12(c) 为男性比率对失业率的影响。塔拉哈西以东、奥兰多及杰克逊维尔沿海一带、开普科勒尔西北地区显示为正向作用。得到的结论是高失业区，男性比例对失业率有一定的正向影响作用；低失业区，男性比例对失业率有反向影响作用。从图 12(d) 可以看出拉美裔人口比例对失业率的正负影响没有明显的地区效应。图 12(e) 显示 65 岁及以上人口比例的空间非平稳性。

可以看出无论哪个水平，老龄化程度对失业率的影响有很大的空间非平稳性。在 0.5 分位水平下，彭萨克拉周边地区表现为正向影响。塔拉哈西及周边地区呈现较弱的正向或者负向效应。在彭萨克拉西部地区，西棕榈滩周边地区，迈阿密部分地区，老龄化程度越高，失业率越高。

综合而言，总人口、男性比例以及 65 岁及以上人口比例对失业率的影响作用较为明显，并且存在明显的空间非平稳效应。

6.2 空气质量数据分析

空气质量问题一直以来备受关注,研究空气质量问题中的空间非平稳性具有一定的现实意义。空气污染物表现出的空间非平稳性等区域特点可以为空气污染防治提供重要的理论支持,如区域联防联控、源头治理等。这里主要关注所提出的方法在空气质量数据分析中的应用,以显示其应用价值。所用数据来自 2019 年我国 13 个省级行政区(13 个省级行政区分别是青海、甘肃、宁夏、山西、陕西、河南、安徽、湖北、江西、湖南、贵州、重庆、四川) 151 个地市州的空气质量数据,括空气质量等级(Qtlv)、空气悬浮颗粒物($\leq 2.5\mu m$)浓度(PM2.5)、空气悬浮颗粒物($\leq 10\mu m$)浓度(PM10)、二氧化硫浓度(SO2)、二氧化氮浓度(NO2)、一氧化氮浓度(NO)、一氧化碳(CO)、和臭氧浓度(O3)考虑对 PM2.5 浓度数据展开研究。为了解决空气质量数据的缺失问题,这里按季度汇总了空气质量数据,并且由于第一季度污染相对严重,下面以第一季度为例展开相关讨论。

6.2.1 模型设定

设定 PM2.5 浓度为响应变量,二氧化硫浓度(SO2)、二氧化氮浓度(NO2)、一氧化碳(CO)、和臭氧浓度(O3)为解释变量。图 14(左)展示了 13 个省级行政区组成的空间区域边界以及数据样本的空间位置。不难看出西北地区,如青海省样本点较为稀疏。在本例中,利用 Lai & Wang(2019)提供的 R 包 *Triangulation* 中的函数 *TriMesh()* 实现研究中 13 个省级行政区组成的空间区域的三角剖分。最终划分结果如图 14(右)所示,由 150 个顶点和 183 个三角形构成。

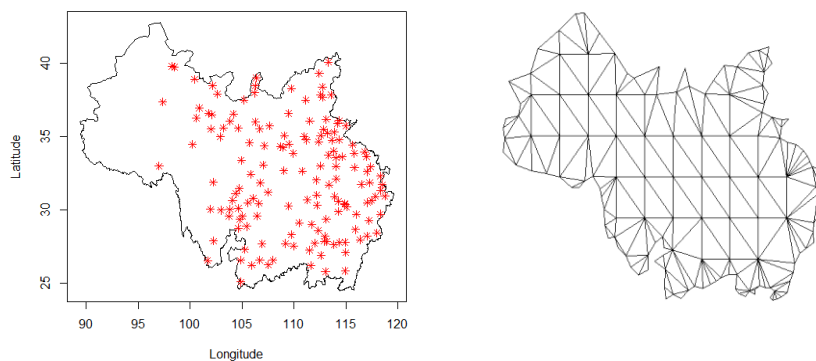


图 14 空间数据样本点(左)与三角剖分(右)

针对空气质量问题建立如下模型，考虑如下估计方法：

(1)空间变系数模型(SVCM)：考虑估计方法有 0.5 分位水平的分位回归估计 SVCMQR0.5，复合分位回归估计 SVCMCQR3；

$$PM2.5 = \beta_0(\mathbf{u}) + \beta_1(\mathbf{u})SO2 + \beta_2(\mathbf{u})NO2 + \beta_3(\mathbf{u})CO + \beta_4(\mathbf{u})O3 + \epsilon, \quad (6.3)$$

(2)线性模型(LM)：考虑估计方法有最小二乘估计 LMOLS 和分位回归估计 LMQR。

$$PM2.5 = \beta_0 + \beta_1SO2 + \beta_2NO2 + \beta_3CO + \beta_4O3 + \epsilon, \quad (6.4)$$

表 12 汇总了各模型均方误差 MSE 和绝对偏差 MAE，结果如下

方法	SVCMQR0.5	SVCMCQR3	SPLVCM0.5	LMOLS	LMQR0.5
MSE	0.3163	0.3334	0.3362	0.3472	0.3767
MAE	0.4121	0.4722	0.4888	0.4415	0.4628

可以看出，本文提出的估计方法 SVCMQR0.5，SVCMCQR3 的拟合结果较优，下面以 SVCMQR0.5 展开数据分析。

6.2.2 拟合优度检验

为了验证 SVCMQR0.5 估计下模型(6.3)拟合优良情况，这里进一步通过假设检验进行验证。利用空气质量数据集 151 个全部样本，产生 100 个 Bootstrap 样本，设定检验显著性水平为 0.05。建立原假设

$$H_0: \beta_j(\mathbf{s}) = \beta_j(\mathbf{s}; \vartheta) \text{ v.s } H_1: \beta_j(\mathbf{s}) \neq \beta_j(\mathbf{s}; \vartheta), \quad 0 \leq j \leq 1.$$

经检验，检验 p 值为 0.1，大于显著性水平 0.05。故根据现有样本无法判断模型(6.3)的合理性。

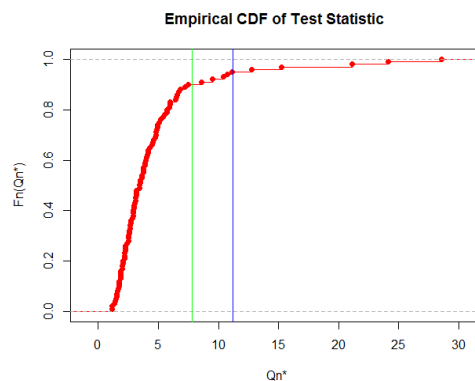


图 15 检验统计量的 ECDF：图中蓝色标记上 0.05 分位数；绿色标记观测统计量的值

6.2.3 结果分析

由于拟合优度检验没有通过,故本文依据估计结果,考虑空间部分线性变系数分位回归模型进行拟合,记为 SPLVCM0.5。根据 SVCMQR0.5 结果,将 CO 设为线性部分协变量,其他未变系数部分协变量(相比其他协变量,CO 协变量空间平稳性较好),最终拟合结果如表 12 所示,可以看到 SPLVCM0.5 的结果依然略差于 SVCMQR0.5, SVCMCQR3。故下面依然采用 SVCMQR0.5 的估计结果。

各协变量系数函数估计结果如图 16 所示。

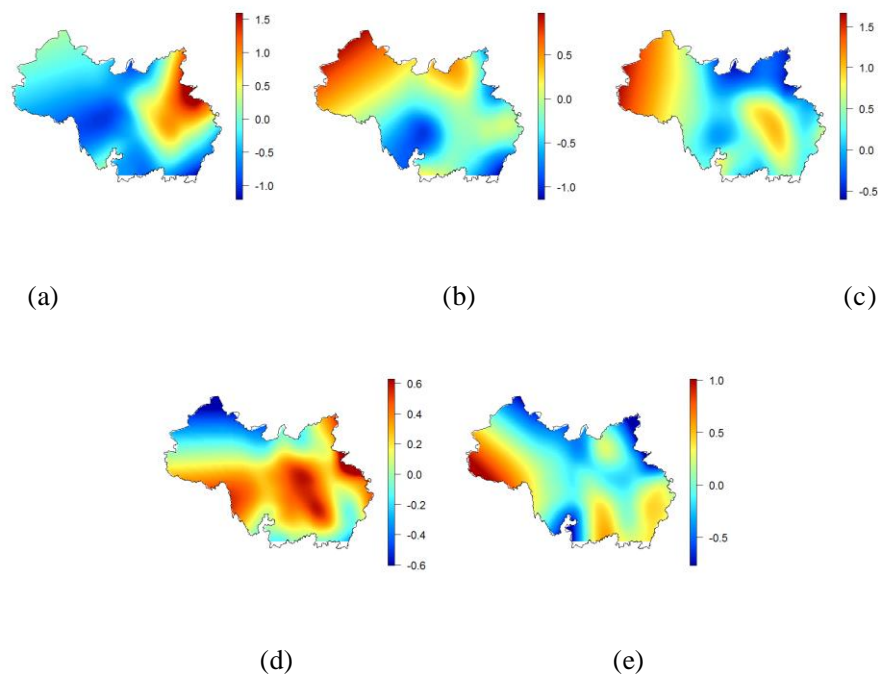


图 16 0.5 分位水平各系数函数估计图: (a-e)分别代表常数和协变量 SO₂, NO₂, CO, O₃ 的二元系数函数 $\beta_0(\cdot), \dots, \beta_4(\cdot)$ 的估计量

图 16(a)是常数项的系数函数估计结果,可称其为基准面。图 16(b)是二氧化硫浓度的空间非平稳效应,在甘肃西北、青海等地表现为正效应。图 16(c)显示,二氧化氮浓度在青海西北、湖北周边地区存在正效应。图 16(d)显示,一氧化碳浓度的正效应遍及大部分地区,如四川、陕西、重庆、湖南及湖北地区,甘肃中部以北显示为负效应。图 16(e)是臭氧浓度的系数函数估计情况,可以看出,臭氧浓度在大部分地区表现为负效应,明显的正效应仅存在于青海西南地区和江西及周边地区。

7 总结与展望

7.1 全文总结

本文讨论了一种研究空间结构数据非平稳性的非参数回归模型,即空间变系数模型。该模型可以有效的研究空间非平稳性问题,探索数据中的空间动态特性。针对二元系数函数提出了该模型的二元惩罚样条分位回归估计方法和复合分位回归方法,参数估计数值模拟结果表明的这里的估计方法更加稳健、有效。在非正态误差分布下,所提出的估计方法要一致优于惩罚迭代重加权最小二乘方法,即使在正态误差下,也能够基本达到最小二乘估计的估计性能,并且从分位回归模拟结果可以发现所提出的估计方法有更快的运算速度。此外,本文还提出了对于空间变系数模型的拟合优度检验方法,在实际应用中,该拟合优度检验可以帮助检验空间变系数模型中假设系数随着空间地理位置而发生变化是否合理。

对于同一个研究问题,考虑到空间平稳性的存在,即平稳性和非平稳性的共存。本文提出了一种新的半参数模型,即空间部分线性变系数模型,该模型是线性模型和空间变系数模型的结合,保留了线性模型的可解释性和非参数模型的灵活性等优点。该模型能够同时研究空间平稳数据和非平稳数据,具有重要的现实意义。针对该模型,提出了一种基于惩罚样条逼近变系数部分系数函数的分位回归估计方法,该方法能够有效的应用于非正态误差数据分析中,得到比最小二乘方法更为稳健的结果。

同时,对于空间变系数模型的分位回归估计量和空间部分线性变系数模型的分位回归估计量,在一定的正则性条件下,讨论了它们的大样本性质,并给出了部分证明,进而从理论上保证了本文提出的估计量的有效性。针对估计量计算实现,采用交替方向乘子算法,该算法相较于其他算法,比较灵活,并且还具备较快的收敛速度,能更好的应用于惩罚目标函数下的参数估计问题。

7.2 研究展望

在今后的研究中,本文所研究的内容还可以从以下几个角度展开深入讨论。其一,可以进一步讨论 B 型-样条展开阶数选择问题,阶数选择要权衡参数

估计精度和计算时间，可以通过构建一个准则函数来寻找两者之间的最优点。

其二，由于数据的复杂性，可以考虑与可加模型、函数型数据模型相结合，从统计的角度讨论异源异构数据的融合分析，以及相应融合模型的参数估计问题。

其三，针对高维空间统计问题，亦可利用 LASSO 方法研究线性部分的变量选择，甚至也可以利用 Group LASSO 方法讨论变系数部分的变量选择问题，甚至，针对空间统计模型的超高维问题，也是值得研究的课题。

其四，随着数据产生背景的复杂化，时空数据的研究被更多的研究者提及，所以可以将本文的空间截面变系数分位回归模型推广至时空混合变系数模型，研究数据的时空分布和相应参数的分位回归估计。

最后，空间高维数据模型甚至时空高维数据模型的降维问题也是一个值得讨论和研究的课题。

参考文献

- [1]Ahmad I, Li L Q. Efficient Estimation of a Semiparametric Partially Linear Varying Coefficient Model[J]. The Annals of Statistics, 2005, 33(1):258-283.
- [2]Andriyana Y, Gijbels I, Verhasselt A. P-splines quantile regression estimation in varying coefficient models[J]. TEST, 2014, 23(1):153-194.
- [3]Azzimonti L, Sangalli L M, Secchi P, Domanin M, Nobile, F. BloodFlow Velocity Field Estimation Via Spatial Regression With PDE Penalization[J].Journal of the American Statistical Association, 2015,110, 1057–1071.
- [4]Boyd S, Parikh N, E Chu, et al. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers[J]. Foundations & Trends in Machine Learning, 2010, 3(1):1-122.
- [5]Brunsdon C, Fotheringham A S, Charlton E M. Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity[J]. Geographical Analysis,1996, 28, 281–298.
- [6]Cai Z, Yao F Q. Functional-Coefficient Regression Models for Nonlinear Time Series[J]. Journal of the American Statistical Association, 2000, 95(451):941-956.
- [7]Chen J, Chen Z. Extended Bayesian information criteria for model selection with large model space. Biometrika,2008, 95, 759–771.
- [8]Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. Journal of the American Statistical Association,2001,96, 1348–1360.
- [9]Fan J, Zhang W. Statistical methods with varying coefficient models[J]. Statistics & Its Interface, 2008, 1(1):179-195.
- [10]Fan J, Huang T. Profile likelihood inferences on semiparametric varying coefficient partially linear models[J]. Bernoulli,2005, 11:1031–1057.
- [11]Fotheringham A S, Brunsdon C, Charlton E M. Geographically Weighted Regression: The Analysis of Spatially Varying Relationships[M]. Chichester:Wiley,2002.
- [12]Gabay D, Mercier B. A dual algorithm for the solution of nonlinear variational problems via finite element approximation[J]. Computers & Mathematics with

- Applications, 1976, 2(1):17-40.
- [13]Guo J, Tang M, Tian M, et al. Variable selection in high-dimensional partially linear additive models for composite quantile regression[J]. Computational Statistics & Data Analysis, 2013, 65(Complete):56-67.
- [14]Guo J, Tian M Z, Zhu K. New efficient and robust estimation in varying-coefficient models with heteroscedasticity[J]. Statistica Sinica, 2012, 22(3):1075-1101.
- [15]Hastie T J, Tibshirani R J. Varying-coefficient Models[J]. Journal of the Royal Statistical Society.1993, 55, 757-796.
- [16]Hestenes M R. Multiplier and gradient methods[J]. Journal of Optimization Theory & Applications, 1969, 4(5):303-320.
- [17]Huang J Z, Zhou W L. Varying-coefficient models and basis function approximations for the analysis of repeated measurements[J]. Biometrika, 2002, 89(1):111-128.
- [18]Kim M, Wang L. Generalized Spatially Varying Coefficient Models[J]. Journal of Computational and Graphical Statistics, 2021,30(1), 1-10.
- [19]Knight K. Limiting distributions for ll regression estimators under general conditions. The Annals of Statistics, 1998, 26: 755–770.
- [20]Koenker R, Bassett G W. Regression quantiles[J]. Econometrica, 1978, 46(1):211-244.
- [21]Koenker R, Park B J. An interior point algorithm for nonlinear quantile regression[J]. Journal of Econometrics, 1996, 71(1-2):265-283.
- [22]Koenker R. Quantile regression[M]. Cambridge University Press,2005.
- [23]Kong L, Shu H, Heo G, et al. Estimation for bivariate quantile varying coefficient model[J]. Statistics, 2015.
- [24]Lai M J, Schumaker L L. Spline Functions on Triangulations[M]. Cambridge University Press, 2007.
- [25]Lai M J, Wang L. Bivariate Penalized Splines for Regression[J]. Statistica Sinica, 2013, 23, 1399–1417.
- [26]Lai M J, Wang L. R Package ‘Triangulation’: Triangulation in 2D domain ,2019,

<https://github.com/FIRST-Data-Lab/Triangulation>.

- [27]Mu J, Wang G, Wang L. Estimation and inference in spatially varying coefficient models[J]. *Environmetrics*, 2018, 29.
- [28]Pietrosanu M, Gao J, Kong L, et al. Advanced algorithms for penalized quantile and composite quantile regression[J]. *Computational Statistics*, 2020:1-14.
- [29]Ramsay T. Spline smoothing over difficult regions[J]. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 2002, 64(2):307-319.
- [30]Shen Y, Liang H Y. Quantile regression for partially linear varying-coefficient model with censoring indicators missing at random[J]. *Computational Statistics & Data Analysis*, 2018,117:1-18.
- [31]Stone, Charles J. Optimal Global Rates of Convergence for Nonparametric Regression[J]. *The Annals of Statistics*, 1982, 10(4):1040-1053.
- [32]Sun Y, Yan H, Zhang W, et al. A semiparametric spatial dynamic model[J]. *The Annals of Statistics*, 2014, 42(2):700-727.
- [33]Tang Y, Song X, Wang H J, et al. Variable selection in high-dimensional quantile varying coefficient models[J]. *Journal of Multivariate Analysis*, 2013, 122(Complete):115-132.
- [34]Wang H J, Zhou Z J. Quantile regression in partially linear varying coefficient models [J]. *The Annals of Statistics*, 2009, 37(6B):3841-3866.
- [35]Wang H, Ranalli M G. Low-Rank Smoothing Splines on Complicated Domains[J]. *Biometrics*,2007, 63, 209–217.
- [36] Wang H, Leng C. Unified LASSO Estimation Via Least Square Approximation[J]. *Journal of the American Statistical Association*,2007, 102, 1039–1048.
- [37]Wang N, Mei C L, Yan X D. Local linear estimation of spatially varying coefficient models: an improvement on the geographically weighted regression technique[J]. *Environment & Planning A*, 2008, 40(4):986-1005.
- [38]Wilhelm M, Sangalli L M. Generalized Spatial Regression with Differential Regularization[J]. *Journal of Statistical Computation and Simulation*, 2016,86,2497–2518.
- [39]Wood S N, Bravington M V, Hedley S L. Soap Film Smoothing[J]. *Journal of the*

- Royal Statistical Society, Series B,2008, 70, 931–955.
- [40]Zhou L, Pan H. Smoothing noisy data for irregular regions using penalized bivariate splines on triangulations[J]. Computational Statistics, 2014, 29(1-2):263-281.
- [41]Zou H, Yuan M. Composite quantile regression and the oracle Model Selection Theory[J]. The Annals of Statistics, 2008, 36(3):1108-1126.
- [42]李家勋. 空间变系数分位数回归模型的稳健模型识别问题研究[D].湖南师范大学,2020.
- [43]刘艳霞, 芮荣祥, 田茂再.部分线性变系数模型的新复合分位数回归估计[J].应用数学学报,2021,44(02):159-174.
- [44]张波, 郭海兵. 部分线性变系数模型的一种新的轮廓(Profile)最小二乘估计[J].数理统计与管理, 2015(02):89-97.
- [45]张园园, 邓文礼, 田茂再.基于变系数模型的自适应分位回归方法[J]. 数学年刊, 2012, 33A(5), 539-556.

作者攻读硕士期间研究成果目录

发表或完成的论文目录:

- [1]梁永玉,田茂再.基于分层贝叶斯时空 Poission 模型的流行病建模研究[J].系统科学与数学,2022,42(02): 462-472.
- [2]梁永玉,田茂再.空间部分线性变系数模型的分位回归估计[J].统计与决策,2022, 38(09):36-41.
- [3]梁永玉,任潇潇,曹苏周,牛成英.基于贝叶斯网络的祁连山自然保护区水资源安全影响因素诊断[J].水文.2021,41(5):65-71+12.
- [4]梁永玉,田茂再.函数型数据视角下基于伪分位数聚类的空气质量治理区域划分.数理统计与管理(已录用),2022.
- [5]梁永玉,田茂再.空间变系数模型的二元惩罚样条分位回归估计.数理统计与管理(返修),2022.
- [6]梁永玉,田茂再.广义函数型部分空间变系数模型的估计与应用.应用数学学报(审查中),2022.

参与课题目录:

- [1]全国统计科学研究项目(2017LZ41),2020年09月已结项;
- [2]甘肃人文社科项目(19ZZ17),2021年09月已结项。

后 记

至此，草文已成，硕士生涯，遂将毕矣。韶光易逝，如白驹过隙。忆往昔，余入金城求学，已历七载。吾发肤受之父母，德才学业，受之师长，外加同窗之助，得以成长。育教恩，同窗谊，虽无以为报，但岂能忘焉！故倾情著文，聊表感激。

吾乃河州人也，能至今日，不离父母之心。父母皆谷人，以农耕为生，知读书之重。余幼时，体弱多病，不好学，常假借不适之名辍学家中。至五六，既不识书，亦恶习之。父母见之，大怒。守学于吾身左右，亲教算术英文，督察甚久，才得以矫正。自幼学起，以读大学为盼，以期不受农耕之苦。现双亲岁至天命，发渐白，犹耕耘劳作，为余操劳。每念之，深感自责内疚，惟有孝顺，方能解忧一二。

余自入学，得拜于田门，受教于恩师田茂再先生。先生博学鸿儒，治学严谨，言传身教，常为研究、授业，倾以余力，至忘我之境，同门皆为此崇拜至甚。先生与我隔千里，指点教导却未减丝毫。每逢先生至金城，便留两三日，携大家而游，近则黄河五泉，远至甘、肃两州，赏月牙清泉，论莫高文化。黄河之滨，白塔山亭，绿皮车里，皆为同门研讨之地。学乐相融，堪比天伦。三年以来，先生常予吾以谆谆教诲、循循善诱。吾今日之寸进，全赖先生之心血。师恩似海，无以为报，而请所授之，铭于心，施于行。

时至今，于兰财七年之久。得统院众师长教导，受益良多，倍感荣幸。吾每遇瓶颈，便向诸师，而今学业顺遂，亦是诸师之功。为此，甚是感恩，谨表谢意。

路漫漫，遇良友者不胜枚举，甚幸。同门之情，乃肺腑也，挚友有三，皆属莫逆也。朝夕间，谈笑风生而品金城佳肴，切磋琢磨亦学彼此之长，可谓亦师亦友。

自入学，恰又逢瘟。何其有幸，生于华夏，得以免于其难。神州医护、华夏儿女皆为其所动，不辞劳苦，在此亦表敬意。

言虽有尽而感激无穷也。故谨借诗文一二，或用诸位之娱遣，或表余之心愿：

《登蘭山有感》

千里栈道五泉悬，碧水石畔玉珠连。几遇垂髫笑开颜，恰如亲临桃花源。

《无题》

晚风萧瑟秋意凉，学苑书香衣衫长。要问吾辈志何方，不负韶华更自赏。

《庚子春寄情》

白雪如衾暖乾坤，小筑几家唱南村。惟愿阴云尽散去，从此九州无一瘟。

永玉于辛丑年冬月