

# 贝叶斯空间分位数回归

丁玉璐

武汉东湖学院基础课部, 湖北 武汉

收稿日期: 2026年3月23日; 录用日期: 2026年4月27日; 发布日期: 2026年5月26日

## 摘要

自Koenker和Bassett在1978年提出分位数回归模型以来, 分位数回归已经取得了长足的发展, 尤其是随着信息技术的深入普及, 分位数回归的研究和应用成为统计学和现代计量经济学的研究热点之一。但考虑到传统的分位数回归方法在空气污染问题研究中的不足, 本文尝试在传统分位数模型中加入空间影响因素, 并结合贝叶斯理论, 建立贝叶斯空间分位数模型, 对影响我国城市空气质量的主要因素进行分析。

## 关键词

分位数回归, 贝叶斯方法, 贝叶斯空间分位数回归

# Bayesian Spatial Quantile Regression

Yulu Ding

Basic Courses Department, Wuhan Donghu University, Wuhan Hubei

Received: March 23, 2026; accepted: April 27, 2026; published: May 26, 2026

## Abstract

Since Koenker and Bassett proposed the quantile regression model in 1978, quantile regression has made significant progress. Especially with the widespread adoption of information technology, the research and application of quantile regression has become a research hotspot in statistics and modern econometrics. However, considering the shortcomings of traditional quantile regression methods in air pollution research, this paper attempts to incorporate spatial influencing factors into the traditional quantile model and, combined with Bayesian theory, establish a Bayesian spatial quantile model to analyze the main factors affecting urban air quality in my country.

## Keywords

Quantile Regression, Bayesian Method, Bayesian Spatial Quantile Regression

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

在大量研究分析及其建模的应用过程中, 回归模型已被广大研究者所青睐。回归模型主要研究自变量和因变量之间的关系[1], 它是一种预测性建模技术。但传统回归模型只是描述了自变量和因变量之间条件期望的关系[2], 不能够全面的了解随机变量[3]。针对这一问题, 著名经济学家和统计学家 Koenker R 与 Bassett G 在 1978 年提出了“分位数回归”思想, 这一思想主要是表达了自变量与因变量之间的条件分位数的关系。自从分位数回归这一思想的提出到现今, 已经过去了约 30 年, 经过学者们的不断探索, 在各个不同的领域都用到了分位数回归的思想, 特别是最近的几年, 随着科学技术的不断发展, 计算机也得到了迅猛的发展, 分位数回归在互联网领域也得到了应用。

然而, 随着分位数回归应用范围的不断扩展[4], 逐渐地发现其中存在的一些问题[5]: 例如针对研究的小样本模型[6]无法保证其大样本时统计量性质的优良性, 又如要非常复杂才能得到假设目标函数不可导[7]等等。于是研究者们就对分位数回归做了一些改进, 即利用贝叶斯估计的不确定性推理[8], 将贝叶斯估计方法与分位数回归相结合[9], 发现解决了分位数回归所面临的一些问题。同时, 结合马尔科夫链蒙特卡洛方法(Markov Chain Monte Carlo, MCMC) [10], 统计学家提出了贝叶斯分位数回归的思想。

## 2. 分位数回归

### 2.1. 分位数回归的提出

回归分析是近几十年来在应用研究领域中最重要统计方法之一。标准的线性回归模型为:

$$Y = X\beta + \varepsilon \quad (2.1)$$

其中  $Y$  是包括了观测值  $Y_1, \dots, Y_n$  的列向量,  $X$  为观测数据矩阵,  $\beta$  表示参数,  $\varepsilon$  表示随机误差项。

传统回归模型又称均值回归, 它研究了一个变量对另一个变量的具体依赖关系, 其是通过已知的后者或设定值, 来预测或估计前者的均值; 分位数回归首先由 Koenker Roger 和 Bassett Gilbert Jr 于 1978 年提出: 利用解释变量和被解释变量之间的条件分位数进行建模, 要揭示解释变量对被解释变量本来分布的位置、刻度和形状的影响。

### 2.2. 分位数回归参数估计方法

#### (一) 分位数回归原理

一般的分位数回归模型为:

$$Y_i = \alpha + X_i\beta_\tau + \varepsilon_{ii}, i=1, 2, \dots, N$$

其中,  $i$  是个体维度,  $\tau$  表示分位数,  $Y_{ii}$  是被解释变量,  $X_{ii}$  是解释变量,  $\alpha$  是常数项,  $\varepsilon_{ii}$  表示随机误差项。

#### (二) 分位数回归估计方法

##### (1) 参数估计

参数估计也即通过样本的综合指标来估计总体的指标。用样本均值对总体均值进行估计也即用样本率估计总体率。参数估计由点估计和区间估计两部分。

点估计通常表示总体的某个特征值，是根据样本来估计总体分布过程中所出现的未知参数函数，常用的构造点估计的方法有四种：

① 矩估计法。用样本矩估计总体矩。

② 最大似然估计法。由 1912 年英国统计学家费希尔提出，它是用样本分布密度来构造似然函数，进而得出参数的最大似然估计值。

③ 最小二乘法。

④ 贝叶斯估计法。它是基于贝叶斯学派的观点所提出的估计方法。由于可用来估计未知参数估计量的方法有很多，那么选择一个优良的估计量成为了一个问题。先要对其优良性定出相应准则，但这一准则并不是唯一的，它可根据理论研究及实际应用的方便对其进行选择。

(2) 分位数回归的估计方法

假设条件分布  $y|x$  的总体  $q$  分位数  $y_q(x)$  是  $x$  的线性函数：

$$y_q(x_i) = x_i' \beta_q \tag{2.2}$$

$\beta_q$  称为“ $q$  分位数回归系数”，其中估计量  $\hat{\beta}_q$  由下面最小化问题定义：

$$\min_{\beta_q} \sum_{i: y_i \geq x_i' \beta_q} q |y_i - x_i' \beta_q| + \sum_{i: y_i < x_i' \beta_q} (1-q) |y_i - x_i' \beta_q|$$

如果  $q=1/2$ ，则为“中位数回归”。

### 2.3. 分位数回归的假设检验

分位数回归的假设检验由两个部分组成：首先是与均值回归比较类似[11]，即拟合优度检验、Wald 检验等等；然后有斜率对称性检验和斜率相等检验等等。

(一) 拟合优度检验

拟合优度检验统计量[12]为：

$$R_{(\tau)}^1 = 1 - \frac{\hat{V}_{(\tau)}}{\tilde{V}_{(\tau)}}$$

其中  $\hat{V}_{(\tau)} \leq \tilde{V}_{(\tau)}$ ，即  $R_{(\tau)}^1$  的值在 0 和 1 之间。 $\hat{V}_{(\tau)}$  小于  $\tilde{V}_{(\tau)}$  表明解释变量具有比较强的作用，那么  $R_{(\tau)}^1$  的值就会越接近于 1，相反则  $R_{(\tau)}^1$  越接近于 0。

此统计量的值越大，则说明拟合效果越好，最小化  $\tau$  的分位数回归目标函数可表示如下：

$$\hat{V}_{(\tau)} = \min_{\beta_{(\tau)}} \sum_i \rho_{\tau}(Y_i - X_i' \beta_{(\tau)})$$

(二) Wald 检验

Wald 检验即为给定一个分位数回归参数估计量的渐进协方差矩阵，就可以构造出 Wald 形式的统计量，然后进行各种约束形式的参数检验。Wald 统计量的表达式为：

$$w = \frac{n(e_*' e_* - e' e)}{e' e} \sim \chi^2(q) \tag{2.3}$$

其中， $e_*' e_*$  为有约束模型的残差平方和， $e' e$  表示为无约束模型的残差平方和。

原假设被设定为： $H_0 : g(\beta) = C$

如果约束为真，则  $g(\beta^{MLE}) - C \rightarrow 0$  不显著，其中  $\beta^{MLE}$  表示为无约束极大似然的估计值。当  $g(\beta^{MLE}) - C$  显著时，那么约束条件表示为无效，从而拒绝原假设。检验统计量为：

$$w = (g(\hat{\beta}) - C)' [Var(g(\hat{\beta}) - C)]^{-1} (g(\hat{\beta}) - C) \sim \chi^2(q)$$

### (三) 斜率相等检验

斜率相等检验即为，针对不同的分位点，检验估计得到的结构参数是否是相等的[13]。

原假设被设定为：

$$H_0: \beta_{i(\tau_1)} = \beta_{i(\tau_2)} = \dots = \beta_{i(\tau_p)} \quad i = 1, \dots, k$$

如果接受原假设，那么对于不同的分位点，每个斜率都是不变的，这时应用普通最小二乘估计；若拒绝了原假设，那么就需要应用分位数回归进行估计。

### (四) 斜率对称性检验

对于给定的  $X$  和  $Y$ ，检验其分布是否是对称的。假设要检验的分位数回归模型共有  $m$  (奇数) 个，它的中间值  $T_{(m+1)/2}$  是 0.5，那么其他的  $T$  都是关于 0.5 对称的，也即  $T_j = 1 - T_{m-j+1}, j = 1, \dots, (m-1)/2$ 。参数估计量按  $T_k$  的大小顺序。则对称性检验的零假设为：

$$H_0: \frac{\beta_{(\tau_j)} + \beta_{(\tau_{m-j+1})}}{2} = \beta_{(0.5)}, \quad j = 1, \dots, (m-1)/2$$

其中， $m$  表示分位数回归的个数，若接受了斜率相等假设，那么不用进行斜率对称性检验就可以了。如果拒绝了斜率相等性假设，那么可以进行斜率对称性检验。若是接受原假设，就认为斜率具有对称性，否则，可认为斜率不具对称性。

## 2.4. 分位数回归与经典最小二乘的比较

当数据出现异方差显著或尖峰厚尾分布等的情况时，OLS 估计的优良性将不再稳定，且稳健性也比较差。反而分位数回归估计比 OLS 估计更稳健，以下表 1 为最小二乘与分位数回归进行的对比。

**Table 1.** Comparison between OLS and QR

**表 1.** 普通最小二乘与分位数回归对比

	普通最小二乘	分位数回归
基本思想	使所构建的方程和样本间的差距最短	同普通最小二乘估计方法
目的	借助于数学模型对客观世界存在事物的不确定性关系进行的数量化描写	同普通最小二乘估计方法
原理	以平均数为基准，求解最短距离	以不同分位数为基准，进而求解最短距离
算法	最小二乘法	加权最小一乘法
前提假设	独立，正态，同方差	独立
假设要求	强假设	弱假设
检验类型	参数检验	非参数检验
承载信息	描述平均的总体信息	充分表示了整个分布上的各部分的信息
极端值	无法考虑极端值的影响	可以充分考虑极端值的影响
异方差	影响大	影响小
拟合曲线	只能拟合一条曲线	可以拟合一簇曲线
计算方法	求偏导解的行列式，算法具有完备	自助法进行标准误差的估计，多种算法可求目标函数

### 3. 基于贝叶斯方法的分位数回归

#### 3.1. 贝叶斯估计

贝叶斯是指托马斯·贝叶斯(1702~1761),他证明了一个特例(皮埃尔-西蒙·拉普拉斯(1749~1827)推导了此定理的一般版本,天体力学、医疗统计学、可靠度及法学上都有应用)。早期的贝叶斯推断是用拉普拉斯的不充分原则进而得到的均匀先验,称作逆向机率[14]。自1920年以后,逆向机率很大程度的被另一群即称为频率论统计的方式取代。

二十世纪开始,拉普拉斯分为两派,出现了客观 bayes 方法和主观 bayes 方法。客观 bayes 方法中,统计分析只会依照假设的模型,及分析的资料及给定先验分布的方式。主观贝叶斯方法(即是提供信息的贝叶斯方法)中,先验的规格依信念(也即分析希望要呈现的主张)而定,信念可以由专家整理资讯后总结产生,也可以根据以往的研究等产生。

自1980年发现了马尔科夫蒙特卡罗方法以来,贝叶斯方法的研究及应用有了大幅的发展[15],解决了许多运算上的问题,也有越来越多的人愿意参与到这类复杂问题的讨论中。bayes 方法也被许多领域接受认可,比如在机器学习、医药、环境等领域中得到了极广泛的应用。

贝叶斯估计主要利用的是贝斯定理,以及结合先验概率和新的证据从而得到新的概率的一种估计方法。从另一种角度看 bayes 估计提出了计算假设概率的一种方法。

bayes 估计是将后验概率推导为两个前、先验概率及似然函数的结果。bayes 推断是根据 bayes 定理从而计算后验概率:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} \quad (3.1)$$

其中,|表示某事件成立的条件; $H$ 为假说,它的概率可能会受到实验数据的影响; $E$ 表示证据; $P(H)$ 表示先验概率,也即观察到数据 $E$ 之前,假说 $H$ 的概率; $P(H|E)$ 后验概率; $P(E|H)$ 是假定成立时,观察到 $E$ 的机率。在 $H$ 不变时,这是 $E$ 的函数,也是似然函数; $P(E)$ 有时会称为边缘似然率;针对不同的 $H$ 数值,只有 $P(H)$ 和 $P(E|H)$ (都在分子)会影响 $P(H|E)$ 的数值。

贝叶斯定理也可以写成下式:

$$P(H|E) = \frac{P(E|H)}{P(E)} \cdot P(H)$$

其中系数 $\frac{P(E|H)}{P(E)}$ 可以解释成 $E$ 对 $H$ 机率的影响。

##### (一) 先验分布与后验分布

先验分布又称为验前分布、事前分布,反映在进行试验统计之前可以根据相关参数的其他知识从而得到相应分布,它与试验结果和随机抽样无关。bayes 学派认为,在对数据做观察获得样本之前,对于 $\theta$ 人们通常会提前有一些认知,因为是实验之前的认知,所以称为先验知识。因此,贝叶斯学派们认为, $\theta$ 应该被称作为随机变量。 $\theta$ 的密度函数为 $h(\theta)$ , $\theta$ 的分布函数称 $H(\theta)$ ,分别称它们为先验密度函数和先验分布函数,它们两者一起称先验分布。参数空间 $(\Theta, B_{\Theta})$ 上任何一个概率分布 $\pi$ 称为参数 $\theta$ 先验分布。它反映人们对参数经验的认识。

##### (二) 蒙特卡罗方法

蒙特卡罗方法首先是一种计算方法,还是一种随机模拟的方法,它是以概率统计的理论方法为基础,可通过利用随机数解决许多的问题。如将所求问题与之对应的概率理论模型相联系,运用计算机进行实

证数据的模拟分析,就可得到所求问题的最优解。

根据概率论的定义可以知道,要想获得某个事件的发生的概率,有一种方法就是经过大量的试验发生频率,用这些频率来估算事物的发生概率,那么当样本容量足够大的时候,就可认为这一频率就是其相对应的概率。所以,有时就可以先对确定的随机变量做大量的试验也即随机抽样,然后再把所得到的结果值分别带入与其相对应的函数式中,以此来确定结构是否合理,最后可求出结构的相应失效概率。

假设具有统计独立的随机变量为  $X_i = 1, 2, 3, \dots, k$ , 那么相应的概率密度函数分别表示为  $f_{x_1}, f_{x_2}, \dots, f_{x_k}$ , 其中功能函数为  $Z = g(x_1, x_2, \dots, x_k)$ 。

首先根据各随机变量的相应分布,产生  $N$  组观测随机数  $x_1, x_2, \dots, x_k$ , 计算功能函数值  $Z_i = g(x_1, x_2, \dots, x_k) (i = 1, 2, \dots, N)$ 。假如其中存在有  $L$  组随机数,那么其对应的功能函数值  $Z_i \leq 0$ 。

### 3.2. 非对称拉普拉斯分布

ALD 是 bayes 分位数回归估计方法的重点内容之一,它也是 bayes 分位数回归的假定误差项分布。近几十年来,国内外也有许多学者在讨论和研究 ALD, ALD 也得到了广泛的推广[16]。下面对 ALD 做一个简单的介绍:

ALD 是在 LD 的理论基础上进行扩展得来的,在了解 ALD 之前,首先对拉普拉斯分布有一个简单的了解: LD 的随机变量概率密度分布为:

$$f(x | \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right) = \frac{1}{2b} \begin{cases} \exp\left(-\frac{\mu - x}{b}\right) & \text{if } x < \mu \\ \exp\left(-\frac{\mu - x}{b}\right) & \text{if } x \geq \mu \end{cases} \quad (3.2)$$

其中,  $\mu$  为位置参数,  $b > 0$  为尺度参数。若  $\mu = 0$ , 则正半部分恰好是尺度为 1/2 的指数分布。如下图 1。

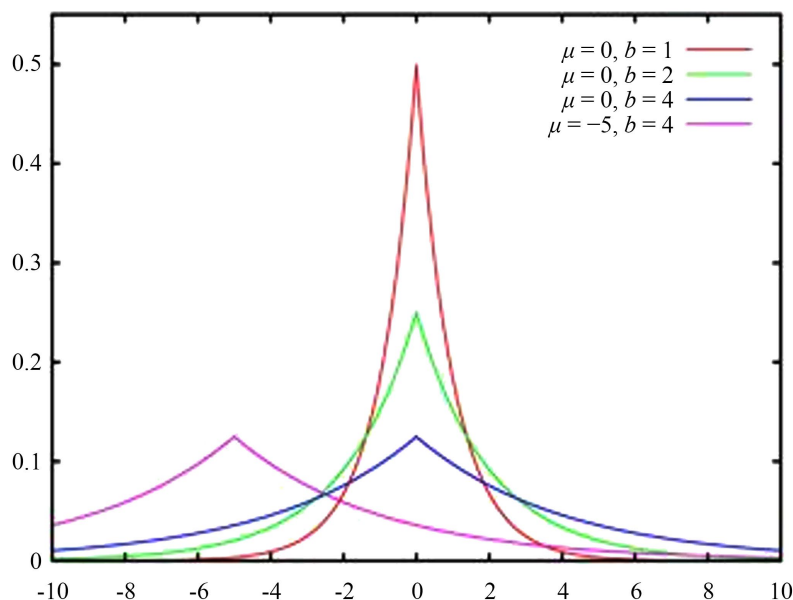


Figure 1. Probability density distribution of LD with varying parameters

图 1. 不同的参数下 LD 的概率密度分布

ALD 是在 LD 的基础上引入的非对称性。即若一个随机变量  $x \sim \text{ALD}(\mu, \sigma, \tau)$ ，那么其对应的密度函数为：

$$f(x; \mu, \sigma, \tau) = \frac{\tau(1-\tau)}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}(\tau - I(x \leq \mu))\right), x \in (-\infty, +\infty) \quad (3.3)$$

其中， $-\infty < \mu < +\infty$  表示为位置参数， $\sigma > 0$  表示尺度参数， $0 < \tau < 1$  表示偏度参数。则对应的分布函数可表示为：

$$F(x; \mu, \sigma, \tau) = \begin{cases} \tau \exp\left(\frac{1-\tau}{\sigma}(x-\mu)\right), & x \leq \mu \\ 1 - (1-\tau) \exp\left(-\frac{\tau}{\sigma}(x-\mu)\right), & x > \mu \end{cases} \quad (3.4)$$

在上式中，当  $x = \mu$  时， $F(x; \mu, \sigma, \tau) = \tau$ ，也即随机变量  $x$  在概率  $\tau$  处的分位数为位置参数  $\mu$ ，所以在分位数回归模型中令误差项服从 ALD 是合适的。这也是把分位数回归模型的随机误差项设置为服从 ALD 的原因。

假定线性回归模型为：

$$y_i = Q_\tau(y_i | x_i; \beta) + \varepsilon \quad (3.5)$$

令  $\varepsilon \sim \text{ALD}(0, \sigma, \tau)$ ，则  $y_i \sim \text{ALD}(Q_\tau(y_i | x_i; \beta), \sigma, \tau)$ 。其中， $\sigma$  为尺度参数， $\tau$  表示偏度参数。则(2.9)的密度函数可写为：

$$\begin{aligned} f(y_i; Q_\tau(y_i | x_i; \beta), \sigma, \tau) &= \frac{\tau(1-\tau)}{\sigma} \exp\left(-\frac{y_i - Q_\tau(y_i | x_i; \beta)}{\sigma}(\tau - I(y_i \leq Q_\tau(y_i | x_i; \beta)))\right) \\ &= \frac{\tau(1-\tau)}{\sigma} \exp\left(-\frac{1}{\sigma} \rho_\tau(y_i - Q_\tau(y_i | x_i; \beta))\right) \end{aligned}$$

根据以上式子，似然函数可以表示为：

$$L(Q_\tau(Y | X; \beta), \sigma, \tau | Y) = \frac{\tau^n (1-\tau)^n}{\sigma^n} \exp\left(-\frac{1}{\sigma} \sum_{i=1}^n \rho_\tau(y_i - Q_\tau(y_i | x_i; \beta))\right)$$

由此可知道，在给定了分位数水平情况下，极小化的分位数损失函数与其极大化似然函数是等价的。由于 ALD 满足了两个重要的性质：(1) ALD 分布满足线性运算；(2) 随机变量的  $\tau$  分位数刚好为位置参数。所以可借助 ALD 对分位数回归模型中的参数进行估计。

#### 4. 贝叶斯空间分位数回归

空间分位数回归是对空间回归和分位数回归两个发展比较成熟且各自相较独立想法的结合。本文参考 Yu、Moyeed(2001)提出的 bayes 分位数回归方法，将分位数回归方法和空间方法相结合，并结合空间的相关性建立如下模型：

$$q(\tau | X_i, s_i) = X_i' \beta(\tau, s_i) \quad (4.1)$$

其中，条件密度  $y_i$  为空间和时间位置上大气污染的综合质量指标， $t_i$  和  $s_i$  分别表示日期和空间的地理位置。主要关注的是估计的条件密度  $y_i$ 、函数  $s_i$  和协变量  $X_i = (X_{i1}, \dots, X_{ip})'$ ，其中  $X_{i1} = 1$  表示截距。特别地，想要研究导致极端日空气质量变化的主要原因，极端事件通常用回归水平来表述，水平值  $c_n$  表示  $n$  天的返回值以及  $P(y_i > c_n) = 1/n$ 。考虑到产生极端事件的因素，因此模型  $y_i$  的条件密度通过分位函数

$q(\tau | X_i, s_i)$  来表示, 定义  $P\{y_i < q(\tau | X_i, s_i)\} = \tau \in [0, 1]$ 。  $\beta(\tau, s_i) = (\beta_1(\tau, s_i), \dots, \beta_p(\tau, s_i))'$  是  $\tau^{\text{th}}$  分位数水平上的空间变化系数。直接对分位数函数进行建模, 可以明确每个协变量对极值的影响程度。

作为模型(4.1)的特例, 还有一些常见的模型。例如: 对所有的  $\tau$  和  $s$ , 令  $\beta_j(\tau, s) \equiv \beta_j$ , 则可得到通用的位置偏移线性回归模型  $\sum_{j=2}^p X_{ij} \beta_j$  和由  $\beta_1(\tau, s)$  决定的剩余密度, 且对所有的  $\tau, j > 1$ , 都使  $\beta_j(\tau, s) \equiv \beta_j(s)$ , 这意味着在不同的空间可以通过空间过程  $\beta_j(s)$  给出空间变系数(Gelfand *et al.*, 2003)  $X_j$  的模型。其中允许  $\beta_j(\tau, s)$  随  $s$  和  $\tau$  假设的放松, 使得协变量的值仅是影响到了平均响应变量, 并且给出了一个密度回归模型, 其中协变量可以影响响应分布的形状。特别是, 对中心( $\tau = 0.5$ )和尾部( $\tau \approx 0$  和  $\tau \approx 1$ )的密度协变量会造成不同的影响。

#### 4.1. 模型的分位数过程

通过忽略空间位置并且假设只有截取的模型  $X_i = 1$  对分位数函数进行建模。在这种情况下, 分位数函数(2.9)缩减到  $q(\tau) = \beta(\tau)$ , 必须构造过程  $\beta(\tau)$  以使得  $q(\tau)$  随着  $\tau$  的变化是不减少的。令

$$\beta(\tau) = \sum_{m=1}^M B_m(\tau) \alpha_m$$

其中  $M$  表示基函数的个数,  $B_m(\tau)$  是一个已知  $\tau$  的基函数,  $\alpha_m$  是未知系数, 确定分位数函数的形状。我们用伯恩斯坦基多项式:

$$B_m(\tau) = \binom{M}{m} \tau^m (1-\tau)^{M-m}$$

以上具有一个明显的特质, 即如果  $\alpha_m \geq \alpha_{m-1}, m > 1$ , 则  $\beta(\tau)$  和  $q(\tau)$  是关于  $\tau$  的增函数。这使得复杂的单调性约束的序列减少了许多, 只有简单的约束:  $\delta_m = \alpha_m - \alpha_{m-1} \geq 0, m = 2, \dots, M$ 。只有这一约束是足够的, 但是为保证其函数的递增性却是不必要的条件。与典型的半参数模型的情况一样, 对于有限的  $M$  模型, 这个模型不能跨越所有的连续单调函数[2]。然而, 随着数量增加, 这些约束在伯恩斯坦多项式基础上引发了未有的密集的状态, 从  $[0, 1] \rightarrow M$  支持连续空间的单调函数(Chang *et al.*, 2007)。

自限制  $\alpha = (\alpha_1, \dots, \alpha_M)$  表达的是相邻项之间的差别, 我们重设置参数  $\delta_1 = \alpha_1$  和  $\delta_m = \alpha_m - \alpha_{m-1}, m = 2, \dots, M$ , 则原始基函数系数是  $\alpha_m = \sum_{i=1}^m \delta_i$ , 参考 Following Cai and Dunson (2008)。为确保分位数限制, 则引入一个潜在的无约束变量  $\delta_m^*$  和  $\delta_1 = \delta_1^*$ , 使得:

$$\delta_m = \begin{cases} \delta_m^*, & \delta_m^* \geq 0 \\ 0, & \delta_m^* < 0 \end{cases}$$

其中,  $m > 1$ 。

$\delta_m^*$  具有独立的正常先验  $\delta_m^* \sim N(\bar{\delta}_m(\Theta), \sigma^2)$ ,  $\Theta$  为未知的超参数。我们选择  $\bar{\delta}_m(\Theta)$  使得分位数过程是以参数分布  $f_0(y | \Theta)$  为中心的。比如, 一个  $N(\mu_0, \sigma_0^2)$  随机变量和  $\Theta = (\mu_0, \sigma_0)$ , 使得  $q_0(\tau | \Theta)$  的分位数函数为  $f_0(y | \Theta)$ , 接着选择适当的  $\bar{\delta}_m(\Theta)$  使:

$$q_0(\tau | \Theta) \approx \sum_{m=1}^M B_m(\tau) \bar{\alpha}_m(\Theta) \quad (4.2)$$

其中,  $\bar{\alpha}_m(\Theta) = \sum_{j=1}^m \bar{\delta}_j(\Theta)$ 。  $\bar{\delta}_m(\Theta)$  对应下列的岭回归估计量:

$$(\bar{\delta}_1(\Theta), \dots, \bar{\delta}_M(\Theta))' = \arg \min_d \sum_{k=1}^K \left( q_0(\tau_k | \Theta) - \sum_{m=1}^M B_m(\tau_k) \left[ \sum_{i=1}^m d_i \right] \right)^2 + \lambda \sum_{m=1}^M d_m^2$$

其中,  $d_m \geq 0, m > 1$ ,  $\{\tau_1, \dots, \tau_k\}$  是一个  $(0, 1)$  密集网格。发现简单的参数分位数曲线通常可以用小于  $M$  的

项来近似表示, 因此几个不同的  $d$  在本质上是相同的, 包括一些为负值的元素  $\bar{\delta}$ 。为确保数值的稳定性, 加上岭惩罚  $\lambda \sum_{m=1}^p d_m^2$ 。调整常数项, 使  $\lambda \rightarrow 0$  适用于未处罚参数, 对所有的  $\bar{\delta} = 0$ , 使  $\lambda \rightarrow \infty$ 。令  $\lambda = 1$  这使得参数分位数曲线趋于近似, 即  $\bar{\delta}$  的值是一个从短期到长期的顺利的一个变化过程。对于  $\sigma \rightarrow 0$  的分位数函数也即类似的对于参数分位数函数  $q_0(\tau | \Theta)$  增加了收缩, 与  $f_0(y | \Theta)$  是类似的。

### 4.2. 具有协变量的空间分位数模型

加上协变量, 则条件分位数函数变成:

$$q(\tau | X_i) = X_i' \beta(\tau) = \sum_{j=1}^p X_{ij} \beta_j(\tau) \tag{4.3}$$

在 4.1 节中, 分位数曲线使用了 Bernstein 基多项式建模:

$$\beta_j(\tau) = \sum_{m=1}^M B_m(\tau) \alpha_{jm}$$

其中  $\alpha_{jm}$  是未知系数。过程  $\beta_j(\tau)$  必须被构造, 以使得对于所有的  $X_i$ , 在  $\tau$  中  $q(\tau | X_i)$  的值是不减少的。收集具有共同基函数的项

$$X_i' \beta(\tau) = \sum_{m=1}^M B_m(\tau) \theta_m(X_i) \tag{4.4}$$

其中,  $\theta_m(X_i) = \sum_{j=1}^p X_{ij} \alpha_{jm}$ 。因此, 如果  $\theta_m(X_i) \geq \theta_{m-1}(X_i)$ ,  $m > 1$ , 那么  $X_i' \beta(\tau)$  和  $q(\tau | X_i)$  是关于  $\tau$  的增函数。

指定  $\alpha_{jm}$  之前需确保其单调性, 假设  $X_{i1} = 1$ , 那么对于截距和其余的协变量可以适当的缩放, 使得  $X_{ij} \in [0, 1], j > 1$ 。由于约束是用相邻项之间的差来表示的, 重设置参数使得,  $\delta_{j1} = \alpha_{j1}$  和  $\delta_{jm} = \alpha_{jm} - \alpha_{j,m-1}$ ,  $m = 2, \dots, M$ 。通过引入潜在的无约束变量  $\delta_{jm}^* \sim N(\bar{\delta}_{jm}(\Theta), \sigma_j^2)$  来确保分位数限制并使

$$\delta_{jm} = \begin{cases} \delta_{jm}^*, & \delta_{jm}^* + \sum_{j=2}^p I(\delta_{jm}^* < 0) \delta_{jm}^* \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

对于所有的  $j = 1, \dots, p$  和  $m = 1, \dots, M$ , 使  $X_{i1} = 1$ ,  $X_{ij} \in [0, 1], j = 2, \dots, p$ 。因此,  $X_{ij} \delta_{jm} \geq X_{ij} I(\delta_{jm} < 0) \delta_{jm} \geq I(\delta_{jm} < 0) \delta_{jm}, j > 1$

$$\begin{aligned} \theta_m(X_i) - \theta_{m-1}(X_i) &= \sum_{j=1}^p X_{ij} \delta_{jm} \geq \delta_{1m} + \sum_{j=2}^p X_{ij} I(\delta_{jm} < 0) \delta_{jm} \\ &\geq \delta_{1m} + \sum_{j=2}^p I(\delta_{jm} < 0) \delta_{jm} \geq 0 \end{aligned}$$

对所有的  $X_i$  给出一个有效的分位数过程。在 4.1 节中心拦截曲线参数分位数函数  $q_0(\Theta)$ 。其余的系数  $\bar{\delta}_{jm}(\Theta) = 0, j > 1$ 。虽然这个模型比较灵活, 但我们仍然假设分位数过程是协变量的线性函数。在某些应用中, 线性分位数关系可能是过度限制性的, 在这种情况下, 可以添加交互项或基函数等进行转换, 以使得模型更加灵活。若将二次项或高阶项添加到模型中, 式(3.5)可能就会不适用, 因式(3.5)不必要地限制了协变量组合的分位数函数, 这些协变量的组合是不可能出现的, 例如, 线性项为零, 二次项为 1。此外, 预测器和响应之间的线性关系对于响应的转换是不变的。为了减弱变换的敏感性, 可以设置一个非线性模型  $q(\tau | X_j)$ ,  $q(\tau | X_j)$  和  $T(q(\tau | X_j))$  为跨越同一类函数的变换  $T$ 。

对于空间数据, 允许分位数过程在每个空间位置是不同的:

$$\beta_j(\tau, s) = \sum_{m=1}^M B_m(\tau) \alpha_{jm}(s) \quad (4.5)$$

$\alpha_{jm}(s)$  是空间变化的基函数系数。在每个空间位置执行单调性约束引入潜在的高斯参数  $\delta_{jm}^*(s)$ ，潜参数与基函数系数的关系为  $\alpha_{jm}(s) = \sum_{j=1}^m \delta_{ji}^*(s)$  和

$$\delta_{jm}^*(s) = \begin{cases} \delta_{jm}^*(s), & \delta_{1m}^*(s) + \sum_{j=2}^P I(\delta_{jm}^*(s) < 0) \delta_{jm}^*(s) \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

对于所有的  $j = 1, \dots, p$  和  $m = 1, \dots, M$  均成立。

在不同的空间对于有条件的密度函数  $\delta_{jm}^*(s)$ ， $\delta_{jm}^*(s)$  是独立的高斯空间过程。即  $E(\delta_{jm}^*(s)) = \bar{\delta}_{jm}(\Theta)$  和指数空间协方差  $Cov(\delta_{jm}^*(s), \delta_{jm}^*(s')) = \sigma_j^2 \exp(-\|s - s'\|/\rho_j)$ ，其中  $\sigma_j^2$  表示的是  $\delta_{jm}^*(s)$  的方差， $\rho_j$  决定了空间相关函数的范围。

## 5. 小结

本文主要介绍了分位数回归的基本思想，及分位数回归的假设检验、与经典最小二乘法的比较，接着第二节介绍了基于 bayes 方法的分位数回归，包括：bayes 估计原理、马尔科夫链、非对称拉普拉斯分布等，最后介绍了 bayes 空间分位数回归的思想及理论模型并假设随机误差项服从 ALD。

## 参考文献

- [1] Dunson, D.B. and Taylor, J.A. (2005) Approximate Bayesian Inference for Quantiles. *Journal of Nonparametric Statistics*, **17**, 385-400. <https://doi.org/10.1080/10485250500039049>
- [2] Taddy, M.A. and Kottas, A. (2010) A Bayesian Nonparametric Approach to Inference for Quantile Regression. *Journal of Business & Economic Statistics*, **28**, 357-369. <https://doi.org/10.1198/jbes.2009.07331>
- [3] Machado, J.A.F. and Mata, J. (2005) Counterfactual Decomposition of Changes in Wage Distributions Using Quantile Regression. *Journal of Applied Econometrics*, **20**, 445-465. <https://doi.org/10.1002/jae.788>
- [4] Melly, B. (2005) Decomposition of Differences in Distribution Using Quantile Regression. *Labour Economics*, **12**, 577-590. <https://doi.org/10.1016/j.labeco.2005.05.006>
- [5] Koenker, R. and Hallock, K.F. (2001) Quantile Regression. *Journal of Economic Perspectives*, **15**, 143-156. <https://doi.org/10.1257/jep.15.4.143>
- [6] Gelfand, A.E., Kim, H., Sirmans, C.F. and Banerjee, S. (2003) Spatial Modeling with Spatially Varying Coefficient Processes. *Journal of the American Statistical Association*, **98**, 387-396. <https://doi.org/10.1198/016214503000170>
- [7] Yu, K. and Zhang, J. (2005) A Three-Parameter Asymmetric Laplace Distribution and Its Extension. *Communications in Statistics—Theory and Methods*, **34**, 1867-1879. <https://doi.org/10.1080/03610920500199018>
- [8] Yu, K. and Moyeed, R.A. (2001) Bayesian Quantile Regression. *Statistics & Probability Letters*, **54**, 437-447. [https://doi.org/10.1016/s0167-7152\(01\)00124-9](https://doi.org/10.1016/s0167-7152(01)00124-9)
- [9] Lancaster, T. and Jae Jun, S. (2010) Bayesian Quantile Regression Methods. *Journal of Applied Econometrics*, **25**, 287-307. <https://doi.org/10.1002/jae.1069>
- [10] Kozumi, H. and Kobayashi, G. (2011) Gibbs Sampling Methods for Bayesian Quantile Regression. *Journal of Statistical Computation and Simulation*, **81**, 1565-1578. <https://doi.org/10.1080/00949655.2010.496117>
- [11] Ghany, M.G., Nelson, D.R., Strader, D.B., Thomas, D.L. and Seeff, L.B. (2011) An Update on Treatment of Genotype 1 Chronic Hepatitis C Virus Infection: 2011 Practice Guideline by the American Association for the Study of Liver Diseases. *Hepatology*, **54**, 1433-1444. <https://doi.org/10.1002/hep.24641>
- [12] Reich, B.J. and Fuentes, M. (2007) A Multivariate Semiparametric Bayesian Spatial Modeling Framework for Hurricane Surface Wind Fields. *The Annals of Applied Statistics*, **1**, 249-264. <https://doi.org/10.1214/07-aos108>
- [13] Gustafsson, B. and Li, S. (2000) Economic Transformation and the Gender Earnings Gap in Urban China. *Journal of Population Economics*, **13**, 305-329. <https://doi.org/10.1007/s001480050140>
- [14] Buchinsky, M. and Hahn, J. (1998) An Alternative Estimator for the Censored Quantile Regression Model. *Econometrica*, **66**, 653-671. <https://doi.org/10.2307/2998578>

- [15] Hjort, N.L. and Walker, S.G. (2009) Quantile Pyramids for Bayesian Nonparametrics. *The Annals of Statistics*, **37**, 105-131. <https://doi.org/10.1214/07-aos553>
- [16] Tokdar, S.T. and Kadane, J.B. (2012) Simultaneous Linear Quantile Regression: A Semiparametric Bayesian Approach. *Bayesian Analysis*, **7**, 51-72. <https://doi.org/10.1214/12-ba702>