分位数回归理论前沿及应用

邱俊鹏◎著

THE FRONTIER AND APPLICATION OF
QUANTILE REGRESSION

经济管理出版社
ECONOMY & MANAGEMENT PUBLISHING HOUSE

国家自然科学基金项目成果（专著）
项目批准号：71601120

分位数回归理论前沿及应用

邱俊鹏◎著

THE FRONTIER AND APPLICATION OF QUANTILE REGRESSION

经济管理出版社
ECONOMY & MANAGEMENT PUBLISHING HOUSE

图书在版编目（CIP）数据

分位数回归理论前沿及应用/邱俊鹏著. —北京：经济管理出版社，2018.9
ISBN 978-7-5096-6002-7

Ⅰ.①分… Ⅱ.①邱… Ⅲ.①自回归模型—研究 Ⅳ.①O212.1

中国版本图书馆 CIP 数据核字（2018）第 206287 号

组稿编辑：张永美
责任编辑：魏晨红　曹　魏
责任印制：黄章平
责任校对：王纪慧

出版发行：经济管理出版社
（北京市海淀区北蜂窝 8 号中雅大厦 A 座 11 层　100038）
网　　址：www. E-mp. com. cn
电　　话：（010）51915602
印　　刷：北京虎彩文化传播有限公司
经　　销：新华书店
开　　本：720mm×1000mm/16
印　　张：14. 75
字　　数：198 千字
版　　次：2018 年 9 月第 1 版　　2018 年 9 月第 1 次印刷
书　　号：ISBN 978-7-5096-6002-7
定　　价：49. 00 元

前　言

分位数回归作为一种不同于均值回归、能提供更丰富信息的方法，早在1978年就被Koenker提出，至今关于它的理论和应用研究仍方兴未艾。在国外，分位数回归无论在理论上还是应用上都已经得到了长足的发展，而在国内还相对滞后，大多以应用为主。本书在分位数回归的两大前沿领域：贝叶斯估计方法和量化政策评价方法做探索性研究，并结合实证案例和蒙特卡罗模拟使读者对其加深理解和掌握。

分位数回归的估计一般采用频率方法，但其估计的“渐近有效性”，使得在小样本容量条件下假设检验的相关结论值得怀疑。Keming. Yu在2001年将贝叶斯方法应用于分位数回归模型，并证实了该估计方法的有效性。从此分位数回归的贝叶斯估计方法及应用受到越来越多研究者的关注。贝叶斯估计方法的一大优势是能实现小样本容量条件下参数后验分布的准确模拟，同时基于参数后验分布的最高后验密度(Highest Posterior Density，HPD）区间实现相关假设检验的准确推断。另外，贝叶斯方法还可以根据研究者的历史经验，引入参数的先验信息，从而后验分布更具灵活性。贝叶斯方法在一定程度上弥补了传统分位数回归估计方法的不足，然而深入比较这两种估计方法的优劣还有待进一步研究。比如如果先验分布不同，这种优势还仍然存在吗？或者在什么情况下，这种优势会更显著？假设检验的准确性如何？其次，传统的分位数回归估计方法也在改进，如可以采用自举方法得到标准差，那么在这种情况下，贝叶斯方法较传统方法还具有优势吗？

分位数回归在量化政策评价中具有独特的优势，可以考察政策对结果变量（Outcomes）的异质性影响。Frolich 和 Melly（2010）指出，约95%的应用计量经济研究是关于变量之间的平均影响，而对于变量分布的影响常常被忽略。均值回归很重要，但它有时不能充分揭示变量之间的关系。譬如，在评价最低工资标准政策产生的经济效益时，我们可以通过均值回归研究政策对工资的平均影响，但对于研究者更关心的政策对低收入人群的影响却无能为力，此时如果采用分位数回归则可以看到政策对整个收入分布的影响（当然包括低收入者）。均值回归中在相关关系到因果关系的确认之间存在"一段距离"，同样，在分位数回归中也一样。解决分位数回归中因果关系的方法我们称之为分位数处理效应。目前分位数回归的双差分方法、断点回归方法已相继被提出。需要指出的是，大多数分位数处理效应方法属于有条件的（Conditional），而不是无条件分位数处理效应（Unconditional Quantile Treatment Effect），两者的政策含义不同。仍以最低工资标准的政策评价为例，采用前者得到的是政策对具有某些相同特征（如年龄、性别、受教育水平、家庭收入等）群体的收入分布的影响，后者是政策对整个收入分布的影响，这才是政策制定者和研究者所真正关心的。

本书将对分位数回归的前沿方法，包括贝叶斯分位数估计方法、分位数双差分方法、分位数断点设计方法、无条件分位数回归方法的理论，及其在政策评价中的应用进行探索性研究。

全书共分六章。第一章作为理论准备和铺垫，介绍分位数回归的基本原理、传统估计方法及有限样本性质。第二章对分位数的贝叶斯估计方法进行深入研究。包括参数化规模参数，研究先验分布的设定对特定抽样算法条件下参数估计量统计性质的影响，不同抽样算法对参数估计量有效性的影响，比较研究贝叶斯分位数回归和传统分位数回归估计量的有效性和假设检验的准确性。第三章为第二章的拓展，将因变量离散化，考察二元离散选择模型的分位数贝叶斯估计方法。第四、第五、第

六章主要围绕分位数处理效应方法。第四章是关于条件分位数处理效应，介绍分位数双差分、分位数断点回归模型，并分别以我国的最低工资标准和入学年龄政策为例，说明这些方法在量化政策评价中的应用。第五章介绍基于倾向得分权重的非参数无条件分位数处理效应估计方法、基于再中心化影响函数方法和基于结构分位数方程的一般分位数回归三种前沿方法。第六章通过模拟和实例对前述处理效应方法进行比较研究，并提出未来的研究方向。

目　录

第一章　分位数回归基础 …… 1

第一节　分位数回归的基本原理 …… 1

第二节　分位数回归的一般估计方法 …… 3

一、单纯形法（Simplex Algorithm） …… 3

二、内点法（Interior Point Method） …… 5

第三节　分位数回归的有限样本分布和渐近分布 …… 6

一、分位数回归的有限样本分布 …… 6

二、分位数回归系数估计量的渐近分布 …… 8

第二章　分位数回归的贝叶斯估计 …… 10

第一节　非对称拉普拉斯分布（ALD） …… 12

第二节　贝叶斯估计方法 …… 14

一、先验分布的设定 …… 16

二、后验分布与 MCMC …… 18

第三节　基于 ALD 的贝叶斯分位数回归估计方法 …… 24

第四节　贝叶斯分位数回归估计方法的软件实现 …… 27

一、用泊松分布指定非对称拉普拉斯分布 …… 28

二、用二项分布指定非对称拉普拉斯分布 …… 31

第五节 尺度参数是否参数化对贝叶斯分位数回归估计结果的影响 …… 32
一、实验设计与抽样结果 …… 33
二、结论 …… 33
第六节 不同先验设定下分位数回归估计量性质的比较 …… 36
一、实验设计与抽样结果 …… 38
二、结论 …… 46
第七节 分位数回归估计量有效性和假设检验准确性的比较 … 47
一、分位数回归估计量有效性的比较 …… 47
二、分位数回归估计量假设检验准确性的比较 …… 49
第八节 应用研究：一种研究资本市场极端风险来源的新方法 …… 52
一、资本市场风险传染的文献综述 …… 54
二、模型设定 …… 57
三、实证分析 …… 61
四、总结与启示 …… 71

第三章 离散选择分位数回归的贝叶斯估计 …… 73

第一节 二元选择分位数回归的贝叶斯估计方法 …… 73
一、二元选择分位数回归 …… 73
二、基于 ALD 的贝叶斯二元选择分位数回归估计方法 …… 75
第二节 不同先验设定下二元选择分位数回归估计量性质的比较 …… 81
第三节 不同抽样方法对二元选择分位数回归估计量的影响 …… 82
第四节 不同估计方法对二元选择分位数回归估计量的影响 …… 87
第五节 删失模型分位数回归的贝叶斯估计方法 …… 89
一、删失模型分位数回归的贝叶斯估计方法 …… 89

二、删失模型分位数回归估计量有效性和假设检验准确性的比较 …… 91
第六节 应用研究：黄金可以对冲股票市场和通货膨胀风险吗 …… 97
一、相关文献综述 …… 99
二、模型设定及相关说明 …… 100
三、实证分析 …… 104
四、结论 …… 111

第四章 分位数处理效应模型及其应用 …… 112

第一节 分位数双差分模型 …… 113
一、线性双差分模型 …… 114
二、非参数双差分模型 …… 116
三、非线性双差分模型 …… 117
四、分位数双差分 …… 119
第二节 应用研究：最低工资标准提升的收入效益研究 …… 119
一、相关文献综述 …… 120
二、数据处理与统计描述 …… 123
三、最低工资标准提升对收入影响的实证分析 …… 126
四、结论与启示 …… 136
第三节 分位数断点回归模型 …… 139
第四节 应用研究：赢在“起跑线”？出生季与未来表现 …… 140
一、出生季与未来表现：从数据出发 …… 142
二、计量方法与识别策略 …… 147
三、实证结果及解释 …… 151
四、结论与启示 …… 156
本节附录 …… 158

第五章　无条件分位数回归方法及其应用 …………………… 160

第一节　无条件分位数回归方法 …………………………………… 160

一、基于倾向得分权重的非参数 UQTE 估计方法 ………… 162

二、基于再中心化影响函数 RIF 的方法 ………………… 164

三、无条件分位数处理效应的一般框架 ………………… 165

第二节　应用研究：高等教育对消费倾向和消费结构的异质性影响 ………………………………………………… 167

一、相关文献综述 ……………………………………………… 169

二、估计方法说明和数据统计特征描述 ………………… 170

三、高等教育对消费倾向和消费结构影响的实证分析 …… 174

四、结论和建议 ………………………………………………… 180

第六章　几种分位数处理效应方法的比较研究及拓展 …… 181

第一节　模拟比较 ………………………………………………… 181

第二节　实例比较 ………………………………………………… 183

一、数据说明 …………………………………………………… 183

二、估计方法及实证结果 ……………………………………… 186

三、各种处理效应的估计结果 ………………………………… 187

四、结论 ………………………………………………………… 190

第三节　无条件分位数处理效应方面的研究展望 …………… 191

附　录 ………………………………………………………… 195

参考文献 ……………………………………………………… 201

后　记 ………………………………………………………… 223

第一章　分位数回归基础

传统的最小二乘回归模型考察的是随机变量均值与解释变量的关系。分位数回归不仅能够度量回归变量对因变量分布中心的影响，而且能度量回归变量对因变量整个分布的影响，因此比经典的最小二乘回归法更具有优势。在经济、金融等社会研究领域，分位数回归的应用也越来越广泛。人们越来越关注随机变量在任意概率水平分位点与解释变量的关系，尤其是在随机变量分布的上尾处和下尾处，如劳动经济学中对收入分布和工作参与意愿的研究，以及后文关于金融风险和对冲的研究等。本章主要介绍分位数回归的理论，包括基本原理、传统估计方法和渐近分布，为后文的估计方法拓展、应用以及分位数的最新进展提供理论基础。

第一节　分位数回归的基本原理

设 Y 为实值随机变量，其分布函数为 $F(y)=P(Y\leqslant y)$，则对于任意的 $0<\tau<1$，有：

$$F^{-1}(\tau)=\inf\{x:\ F(x)\geqslant\tau\} \tag{1.1}$$

称式（1.1）为 Y 的 τ 分位数，当 $\tau=1/2$ 时，即为中位数。在分位数回归模型中，定义损失函数为分段线性函数：

$\rho_\tau(u)=u(\tau-I(u<0))$

其中，I 为指示函数。在决策理论中，损失函数不同，决策内容亦不同。损失函数期望的最优化问题可以表示为：

$$\min E[\rho_\tau(Y-\hat{\xi})] = (\tau-1)\int_{-\infty}^{\hat{\xi}}(y-\hat{\xi})\mathrm{d}F(y) + \tau\int_{\hat{\xi}}^{+\infty}(y-\hat{\xi})\mathrm{d}F(y)$$

对 $\hat{\xi}$ 求一阶导数，得：

$$(1-\tau)\int_{-\infty}^{\hat{\xi}}\mathrm{d}F(y) - \tau\int_{\hat{\xi}}^{+\infty}\mathrm{d}F(y) = F(\hat{\xi}) - \tau = 0$$

由于 F 为单调函数，因此满足 $\{\xi: F(\xi)=\tau\}$ 的任一元素都可以使期望损失最小。当解唯一时，$\hat{\xi}=F^{-1}(\tau)$；当有多个解时，取解区间的左端点，显然引入损失函数可以求得分位数的点估计。考虑一个样本 $\{y_i\}_{i=1}^{n}$，求它的 τ 概率水平的分位数问题可以转换成下面的最优化问题：

$$\min\sum_{i=1}^{n}\rho_\tau(y_i-\hat{\xi})$$

给定信息集 x，y 的条件分位数函数可以表示为 $q_y(\tau \mid x)=x'\beta$，系数向量 β 的估计由：

$$\min R(\beta) = \sum_{i=1}^{n}\rho_\tau(y_i - x'_i\beta) \tag{1.2}$$

得到。可以进一步写为：

$$\min(\sum\tau\,|y_i - x'_i\beta| + \sum(1-\tau)\,|y_i - x'_i\beta|) \tag{1.3}$$

可以看出，系数向量 β 的估计随 τ 的变化而不同。分位数回归的本质是通过 τ 在 0~1 取值，调节回归线（或者回归平面）的位置和方向，以优化目标函数。

第二节　分位数回归的一般估计方法

分位数回归是以古典条件均值模型为基础的最小二乘法的延伸，它用几个分位函数来估计整体模型。分位数回归法的特殊情况就是中位数回归（最小一乘回归），用对称权重解决残差绝对值之和最小化问题，而其他条件分位数回归则需要用非对称权重解决残差绝对值之和的最小化。

分位数回归的优点主要体现在以下几个方面：首先，它对模型中的随机扰动项不需做任何分布的假定，且分位数回归估计量具有在大样本理论下的渐近优良性质。其次，分位数回归由于是对所有分位数进行回归，因此对于数据中出现的异常点具有稳健性。最后，不同于普通的最小二乘回归，分位数回归对于因变量具有单调变换性。

目前，常用的分位数回归方法有单纯形法和内点法。

一、单纯形法（Simplex Algorithm）

Koenker 和 Orey（1993）把分两步解决最优化问题的单纯形算法（Barrodale 和 Roberts，1973）扩展到所有回归分位数中。

分位数回归的最优化模型式（1.3）可以重新表述为：

$$\min\{\tau e'_n u+(1-\tau)e'_n v \mid y-x'_i b=u-v,\ b\in\mathbb{R}^p,\ (u,\ v)\in\mathbb{R}^{2n}_+\}$$

其中，p 表示 b 中元素的个数。令 H 是集合 $N=\{1,\ 2,\ \cdots,\ n\}$ 中 p 个元素构成的子集。设 $h\in H$ 表示第 h 个 p 元素子集，$X(h)=\{x_i:\ i\in h\}$，$y(h)=\{y_i:\ i\in h\}$。h 在集合 N 中的补集记为 $\bar{h}$，$X(\bar{h})$ 和 $y(\bar{h})$ 的定义同上。于是，通过点集 $\{(x_i,\ y_i),\ i\in h\}$ 的线性规划任意基本解可以表达为：

$$b(h)=X(h)^{-1}y(h)$$
$$u(h)=v(h)=0$$
$$u(\bar{h})=(y-Xb(h))^{+}$$
$$v(\bar{h})=(y-Xb(h))^{-}$$

最优解在各个方向上的导数必须是非负的，考察目标函数 $R(b)$ 在 $b(h)$ 处的方向导数：

$$\nabla R(b(h),\ w)$$
$$=\frac{\mathrm{d}}{\mathrm{d}t}R(b(h)\ +tw)\big|_{t=0}$$
$$=\frac{\mathrm{d}}{\mathrm{d}t}\sum_{i=1}^{n}(y_i-x'_ib(h)-Xtw)(\tau-I((y_i-x'_ib(h)-x'_itw)<0))\big|_{t=0}$$
$$=-\sum_{i=1}^{n}\psi_{\tau}^{*}(y_i-x'_ib(h),-x'_iw)x'_iw$$

其中，$\psi_{\tau}^{*}(u,\ v)=\begin{cases}\tau-I(u<0),\ u\neq 0\\ \tau-I(v<0),\ u=0\end{cases}$。

引入参数 $v=X(h)w$，当且仅当满足下列条件时达到最优：

$$\forall v\in\mathbb{R}^{p},\ 0\leqslant-\sum_{i=1}^{n}\psi_{\tau}^{*}(y_i-x'_ib(h),-x'_iX(h)^{-1}v)x'_iX(h)^{-1}v$$

根据定义可知，$\mathbb{R}^{p}$ 空间上的单位基向量 $e'_i=x'_iX(h)^{-1}$，所以上式可改写为：

$$\forall v\in\mathbb{R}^{p},\ 0\leqslant-\sum_{i\in h}\psi_{\tau}^{*}(0,\ -v_i)v_i-\xi'v=-\sum_{i\in h}\psi_{\tau}^{*}(\tau-I(-v_i<0))v_i-\xi'v$$

其中，$\xi'=\sum_{i\in h}\psi_{\tau}^{*}(y_i-x'_ib(h),\ -x'_iX(h)^{-1}v)x'_iX(h)^{-1}$，方向 v 在空间 $\mathbb{R}^{p}$ 上的正则性由 $v=\pm e_k(k=1,2,\cdots,p)$ 来保证，所以最优梯度条件成立的充分条件是 $2p$ 个正则单位方向 $\{\pm e_i:i=1,2,\cdots,p\}$ 成立。

该算法估计得到的参数具有很好的稳定性，但是在处理大型数据时运算的速度会显著降低。

二、内点法（Interior Point Method）

单纯形算法在处理大型数据时效率低下，Portnoy 和 Koenker（1997）尝试把内点算法使用在分位数回归中，得出在处理大型数据时内点算法的运算速度远快于单纯形算法。

解式（1.3）可以重新表述为下面的线性规划问题：

$\min\{\tau e'u+(1-\tau)e'v \mid y=Xb+u-v,\ (u,\ v)\in\mathbb{R}_{+}^{2n}\}$

并且有对偶表达式：

$\max\{y'd \mid X'd=0,\ d\in[\tau-1,\ \tau]^{n}\}$

或者写为：

$\max\{y'a \mid X'a=(1-\tau)X'e,\ a\in[0,\ 1]^{n}\}$

其中，$a=d+1-\tau$。分位数回归的对偶表达式非常适合采用解决有界变量线性规划问题的内节点方法。本节所采用的算法是由 Portnoy 和 Koenker（1997）提出的 Frisch-Newton 方法。该算法也会在后文传统分位数回归方法中采用，可以在 R 软件中调用 quantreg 包执行。

加入松弛变量 s 和约束 $a+s=e$，可以得到障碍函数（Barrier Function）：

$$B(a,\ s,\ \mu)=y'a+\mu\sum_{i=1}^{n}(\log a_i+\log s_i)$$

在约束条件 $X'a=(1-\tau)X'e$ 和 $a+s=e$ 下，求目标函数的极大值。Newton 步长为 δ_a，求解：

$$\max y'\delta_a+\mu\delta'_a(A^{-1}+S^{-1})e-\frac{1}{2}\mu\delta'_a(A^{-2}+S^{-2})\delta_a$$

以 $X'\delta_a=0$ 为约束，满足：

$y+\mu(A^{-1}+S^{-1})e-\mu(A^{-2}+S^{-2})\delta_a=Xb$

两边同乘以 $X'(A^{-2}+S^{-2})^{-1}$，并结合约束条件，可以求得：

$b=(X'WX)^{-1}X'W(y+\mu(A^{-1}+S^{-1})e)$

其中，$W=(A^{-2}+S^{-2})^{-1}$。这是最初的障碍算法，在每一步中设 $\mu=0$

将产生这个算法的仿射变换，而且在每一步迭代中基本的线性代数是没有必要改变的，只是对角权重矩阵 W 发生了变化。

迭代一直进行到对偶图小于设定的 ε 为止。另外，由于此对偶图在解的位置上为零，因此这个判断收敛的准则比一般迭代算法更直接。

以上两种方法各有利弊。单纯形算法在处理大型数据时效率很低，但是估计出来的参数具有很好的稳定性。而且当数据中存在大量异常值时，单纯形算法可以计算出参数解，而这时内点算法可能失败。内点算法对于那些具有大量观察值和少量变量的数据集运算效率很高。但是内点算法也有自身的缺陷：它在每一步计算时都要进行因数分解，当自变量比较多的时候效率比较低；如果要达到和单纯形算法一样的精确度，就必须进行舍入步骤的计算，这也会降低算法的运行效率。此外，平滑算法（Smoothing Method）也是一种可供选择的估计方法。平滑算法用一个平滑函数来逼近，因此，可以反复使用牛顿—拉尔夫方法，这样经过有限步骤以后就能获得参数解，它兼顾了运算效率以及运算速度。Chen 等（2000）把这种算法扩展到计算回归分位数中。平滑算法在理论上比较简单，它适合处理具有大量观察值以及很多变量的数据集。后来，Chen 等（2000）结合这三种算法提出了一种自适应算法。

第三节　分位数回归的有限样本分布和渐近分布

一、分位数回归的有限样本分布

假定 Y_1，Y_2，…，Y_n 为一组独立同分布随机变量，累积分布函数为 F，对于一定的概率水平 τ，假定 F 在 $\xi_\tau = F^{-1}(\tau)$ 的领域内具有连续

的密度函数f，且$f(\xi_\tau)>0$，则样本τ分位数的目标函数为：

$$\hat{\xi}_\tau \equiv \inf\{\arg\min n^{-1}\sum_{i=1}^{n}\rho_\tau(Y_i-\xi)\}$$

目标函数$\sum_{i=1}^{n}\rho_\tau(Y_i-\xi)$是凸函数之和，因此也是凸函数。对$\xi$求一阶导数得：

$$g_n(\xi)=n^{-1}\sum_{i=1}^{n}(I(Y_i<\xi)-\tau)$$

根据$g_n(\xi)$的单调性，当且仅当$g_n(\xi)<0$时，$\hat{\xi}_\tau>\xi$成立。由此可知：

$$\begin{aligned}P\{\sqrt{n}(\hat{\xi}_\tau-\xi_\tau>\delta)\}&=P\{g_n(\xi_\tau+\delta/\sqrt{n})<0\}\\&=P\{n^{-1}\sum_{i=1}^{n}(I(Y_i<\xi_\tau+\delta/\sqrt{n})-\tau)<0\}\end{aligned}$$

$g_n(\xi_\tau+\delta/\sqrt{n})$的期望为：

$$\begin{aligned}E(g_n(\xi_\tau+\delta/\sqrt{n}))&=n^{-1}\sum_{i=1}^{n}E(I(Y_i<\xi_\tau+\delta/\sqrt{n})-\tau)\\&=(1-\tau)F(\xi_\tau+\delta/\sqrt{n})-\tau(1-F(\xi_\tau+\delta/\sqrt{n}))\\&=F(\xi_\tau+\delta/\sqrt{n})-\tau\\&=\frac{F(\xi_\tau+\delta/\sqrt{n})-F(\xi_\tau)}{\delta/\sqrt{n}}\delta/\sqrt{n}\to f(\xi_\tau)\delta/\sqrt{n}\end{aligned}$$

$g_n(\xi_\tau+\delta/\sqrt{n})$的方差为：

$$\begin{aligned}V(g_n(\xi_\tau+\delta/\sqrt{n}))&=V(n^{-1}\sum_{i=1}^{n}(I(Y_i<\xi_\tau+\delta/\sqrt{n})-\tau))\\&=n^{-2}\sum_{i=1}^{n}V(I(Y_i<\xi_\tau+\delta/\sqrt{n})-\tau)\\&=n^{-2}\sum_{i=1}^{n}F(\xi_\tau+\delta/\sqrt{n})(1-F(\xi_\tau+\delta/\sqrt{n}))\\&=F(\xi_\tau+\delta/\sqrt{n})(1-F(\xi_\tau+\delta/\sqrt{n}))/n\to\tau(1-\tau)/n\end{aligned}$$

$$P(\sqrt{n}(\hat{\xi}_\tau-\xi_\tau)>\delta)=P(g_n(\xi_\tau+\delta/\sqrt{n})<0)$$