



分位数回归的思想

- How to “go further” ?
- 分位数回归的思想最早是由**Koenker and Bassett (1978)**提出的。它是对古典条件均值模型为基础的最小二乘的拓展。
- 普通最小二乘法是利用因变量的条件均值来建模，通过使残差平方和达到最小来获得回归参数的估计。
- 分位数回归则利用因变量的条件分位数来建模，通过最小化加权的残差绝对值之和来估计回归参数。它可以称之为“加权的最小一乘回归法”。



分位数的概念

- 定义：设随机变量 Y 的分布函数为
$$F(y) = P(Y \leq y)$$

则 Y 的第 τ 分位数为

$$F^{-1}(\tau) = \inf\{y : F(y) \geq \tau\}$$


其中中位数可以表示为 $F^{-1}(1/2)$ 。



分位数回归思想的数学公式化

- 对于 Y 的一组随机样本 $\{y_1, y_2, \dots, y_n\}$, 样本均值是 $\min \sum_{i=1}^n (y_i - \xi)^2$ 的最优解。
- 样本中位数是最小化残差绝对值和的解, 即

$$F^{-1}(1/2) = \arg \min_{\xi \in R} \sum_{i=1}^n |y_i - \xi|$$


- 
- 对于其他的第 τ 分位数,我们可以求解下式:

$$\min_{\xi \in \mathbb{R}} \left[\sum_{i \in \{i: y_i \geq \xi\}} \tau |y_i - \xi| + \sum_{i \in \{i: y_i < \xi\}} (1 - \tau) |y_i - \xi| \right]$$

- 等价的表示为:

$$\min_{\xi \in \mathbb{R}} \sum_{i=1}^n \rho_{\tau}(y_i - \xi)$$

其中 $\rho_{\tau}(z) = \tau z I_{[0, \infty)}(z) - (1 - \tau) z I_{(-\infty, 0)}(z)$, $I(\cdot)$ 为示性函数。

- 
- 对于一般线性条件均值函数 $E(Y|X=x)=x'\beta$, 通过求解

$$\hat{\beta} = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \sum_{i=1}^n (y_i - x_i' \beta)^2$$

得到参数估计值。而一般线性条件分位数函数为 $Q(\tau|X=x)=x'\beta(\tau)$, 通过求解 得到参数估计值

$$\hat{\beta}(\tau) = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_{\tau}(y_i - x_i' \beta)$$

- 对于任意的 $\tau \in (0,1)$, 估计 $\hat{\beta}(\tau)$ 称为第 τ 分位数下的回归系数估计。




分位数回归参数的估计方法（点估计）

- 求解 $\hat{\beta}(\tau) = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_{\tau}(y_i - x_i' \beta)$ 等价于求解以下线性规划问题：

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

其中 \mathbf{e} 为 n 维单位向量。目前对上式的算法主要有如下几种：

- 
- 1.单纯形算法(Simplex Method): 该算法估计出来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的降低 (见Koenker and Orey, 1993)。
 - 2.内点算法(Interior Point Method): 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 (见Portnoy and Koenker, 1997)。
 - 3.平滑算法 (Smoothing Method): 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 (见Chen, 2004) 。
 - 其他方法: 如adaptive method 等。



分位数回归参数的估计方法（区间估计）

依据目前的文献，区间估计方法也可分为三种：

1. 直接估计法(Direct Estimation Method)，见Koenker和Bassett (1982)以及Koenker和Machado (1999)。该方法依据估计出来的回归分位系数的渐进正态性来计算置信区间。比较有代表性的是Sparsity算法，它是一种最直接且运算速度也最快的算法，但该算法得到的估计值对于随机项为独立同分布这一假设十分敏感。
2. 秩得分法(Rank Score Method)，见Koenker(1994)。秩得分法算法比较简单，但是对于大型数据处理效率较慢。
3. 重复抽样法(Resampling method)，见He和Hu (2002)。该方法使用了MCMB(Markov Chain Marginal Bootstrap)算法，这种算法能够进行高效率的运算，大大节省了运算时间。重复抽样法能够克服直接法和秩得分法的缺陷，但是对于小样本时计算出的参数估计值不够稳定。



分位数回归参数的显著性检验方法

- 在分位数回归模型中，设 $\beta(\tau) = (\beta_1(\tau), \beta_2(\tau))$ Koenker 与 Machado (1999) 提出了检验假设 $H_0: \beta_2(\tau) = 0$
- (其中 $\beta_2(\tau) \in R^q$) 的两个统计量：

$T_w(\tau)$ 和 $T_{LR}(\tau)$ 在原假设下都服从 χ_q^2

从而，它们都可能用来检验回归系数的显著性。



分位数回归模型的拟合优度

- Koenker与Machado（1999）依据最小二乘回归中拟合优度 R^2 的计算思想，提出了分位数回归中拟合优度的计算方法，定义为 $R^1(\tau)$ ，且 $0 \leq R^1(\tau) \leq 1$ 。
- 最小二乘回归中的 R^2 依据残差平方和度量了回归平方和占总离差平方和的比重，而 $R^1(\tau)$ 则按照残差绝对值的加权和，度量了在某个分位数下分位数回归的拟合效果。因此不像 R^2 反映的是整个分布的拟合情况，描述的是在某个分位数下的局部拟合效果。



与普通线性最小二乘回归方法的比较

- 1.在模型假设方面：OLS法要求满足经典假设的几个条件；QR法只要求扰动项 $e_i \perp F_i$ 的条件下 $F_i^{-1}(\tau)=0$ 。
- 2.在计算方面：OLS法求解简单；QR法复杂，但由于计算机技术的发展，其不难完成。
- 3.在估计的优良性方面：两者都有各自的优良性。由于QR法在模型的假设方面要求较少，较容易得到满足。特别是其估计方法（加权最小一乘估计方法）决定了其估计具有较强的稳健性。



分位数回归的应用

- 在环境科学研究方面:Chock, Winkler和Chen(2000)使用非参数分位数回归法研究了匹兹堡这座城市中日死亡率和空气污染集中度的相互关系; Hilary和Andrzej(2002)运用分位数回归对天气数据进行了分析。
- 在生态学研究方面: Koenker and Schorfheide(1994)分析了上世纪全球气候改变的过程; Dunham, Cade和Terrell(2002)研究了不同的河流对鲑鱼密度的影响。
- 在生存分析研究方面: Koenker和Hallock(2001)研究了诸多因数对于新生儿出体重的影响; Cole和Green(1992)以及Royston和Altman(1994)还讨论了分位数回归在医学上的应用。



分位数回归的应用

- 在劳动力市场研究方面：Buchinsky(1996)研究了美国的劳动力市场；Schultz和Mwabu(1998)研究了南非的劳动力市场；Montenegro(1998)分析了智利的情况；Fithzenberger、Hujer、Macurdy 和 Schnabe(2001)讨论了德国劳动力市场的情况；值得一提的是Machado与Mata (2001)扩展了Oaxaca的分解法，并提出了一种依据分位数回归过程来模拟边际分布的方法来研究葡萄牙的劳动力市场；Yu, Philippe 和Zhang (2005)使用贝叶斯分位数回归研究了英国1991-2001年间的工资结构的分布情况；Angrist, Chernozhukov 和Fernandez-val (2006)使用分位数回归方法研究了美国的工资结构的分布情况；Papapetrou (2006) 则研究了希腊公私企业中的工资差距情况。



分位数回归的应用

- 在需求分析研究方面：**Deaton(1997)**利用分位数回归分析了巴基斯坦的**Engel**曲线；**Hendricks**与**Koenker(1992)**使用分位数回归研究了电力消费需求情况；**Manning, Blumberg and Moulton(1995)**讨论了酒精使用的需求情况；**Taylor (2007)**运用指数加权分位数回归来预测超级市场的日销售情况。



分位数回归的应用

- 在收入不平等问题研究方面：**Gosling, Machin and Meghir (1996)** 研究了英国家庭的收入和财富的分布状况；**Conley and Galenson (1998)** 探讨了美国的几个城市在19世纪中期财富的累积情况；**Trede (1998) and Morillo (2000)** 比较了美国与德国的收入机动性。



分位数回归的应用

- 在金融资产组合研究方面：Bassett and Chen（2000,2001）运用分位数回归来评估共同基金的投资类型；Barnes和Hughes(2002)研究了美国资本市场的CAMP模型；Ma和Pohlman(2005)运用分位数回归讨论了资产收益预测与最优资产组合的构造。



分位数回归的应用

- 在金融风险研究方面：在金融市场的风险管理中，VaR已经成为其标志性指标。VaR能够将金融机构面临的所有种类各异、形式多样的风险加总成为一个简单的数字。Engle and Manganell（2004）首次在VaR的基础上提出了条件VaR模型（CAViaR），并使用分位数回归的方法来估计参数，通过对数据的模拟得出了这一模型对于厚尾数据的表现为最优的结论；Taylor（1999）使用分位数回归的方法来估计多期收益的风险值VaR；Chernozhukov and Umantsev（2001）使用分位数回归深入研究了CVaR的模型与估计方面的问题；Chen和Chen(2003)分别用分位数回归方法和方差-协方差法计算日经225指数的VaR值，实证结果为前者很大程度上优于后者；Georgios and Leonidas（2005）使用CAViaR 模型估计了美国和希腊证券市场中的市场风险值。