Bayesian Quantile Regression over Semiparametric Ordinal Models using MCMC

Under: Prof. M.A. Rahman

Anshul Goel
150110

Department of Economic Sciences
Indian Institute of Technology, Kanpur

April 20, 2018

Motivation

- Robustness is very profound term in the domain of Statistics
- Surprisingly, the estimators in common use like least squares are not very robust and perform very poorly in case of even modest deviations
- Many robust estimates to sample mean like median and trimmed means were in use in astronomy as old as in 18th century. Some robust estimates to least square errors were also introduced like leat absolute error estimates and later on quantile regressions
- The quantile regression estimates discussed in the work offer comparable efficiency to least square estimates for Gaussian error error distributions and higher for other wide class of error distributions

Robust Statistics

A statistic is defined to be robust if it gives the following desired features:

- Reasonably good (optimal) efficiency at assumed means where efficiency is defined as low error in point estimate and narrow intervals for interval estimate
- Insensitive to small deviations in model assumptions
- Large deviations dont break down the estimate

Thus, robust statistic is one that gives good estimate for the processes in the neighborhood of chosen model.

Regression Quantiles: Definition

- Sample Quantiles: Let $\{y_t : t = 1, ..., T\}$ be a random sample on a random variable Y having a distribution F. Then, the θth sample quantile, $0 < \theta < 1$, may be defined as any solution to the minimisation problem:

$$min_{b \in \mathbb{R}} \left[\sum_{t \in \{t : y_t \geq b\}} \theta |y_t - b| + \sum_{t \in \{t : y_t \leq b\}} (1 - \theta) |y_t - b| \right] \qquad (1)$$

- Generalising the concept of sample quantile to regression quantile by bypassing the use ordered samples and using the above helps to solve the issue of identifying outliers in regression and thus regression quantile is defined to solution of problem:

$$min_{b \in \mathbb{R}} \left[\sum_{t \in \{t : y_t \geq x_t' \beta\}} \theta |y_t - x_t' \beta| + \sum_{t \in \{t : y_t \leq x_t' \beta\}} (1 - \theta) |y_t - x_t' \beta| \right] \qquad (2)$$

Regression Quantiles: Remarks

Resemblance in Asymptotic behaviour:

- The analogy between sample quantiles and regression quantiles is strengthened by resemblance in asymptotic behaviour. Thus, *the regression median is more efficient than the least squares estimator in linear model for any distribution for which sample median is more efficient than the mean*
- A straight forward extension of theory of "Systematic Statistics"- estimators which are linear combination of sample quantiles to estimators which are combination of regression quantiles

Regression Quantile: Bayesian Modelling Approach

- Yu and Moyeed (2001) proposed that minimising (2) is equivalent to maximising a likelihood function under the asymmmetric Laplace error distribution but the resulting posterior for β_p (p denoting the p^{th} quantile) came out to be non-tractable. The density of error term, ϵ was defined as:

$$f_p(\epsilon) = p(1-p)exp\{-\rho_p(\epsilon)\} \tag{3}$$

where $\rho_p(u) = u\{p - I(u < 0)\}$, $I(.)$ is indicator function. The mean and variances are: $E(\epsilon) = \frac{1-2p}{p(1-p)}$ and $Var(\epsilon) = \frac{1-2p+2p^2}{p^2(1-p)^2}$

- Later, the location-scale mixture representation of asymmetric Laplace distribution was suggested which gives fully tractable densities as follows:

$$\epsilon = \theta z + \tau\sqrt{z}u \tag{4}$$

where z is a standard exponential distribution and u is a standard normal and $\theta = \frac{1-2p}{p(1-p)}$ and $\tau^2 = \frac{2}{p(1-p)}$

Semiparametric Regression: Introduction

$$y = x_0'\beta_0 + f(x_1, ..., x_{K_1}) + u \tag{5}$$

where K_0 columns of x_0' are covariates assumed to enter linearly in mean function, $f(.)$ is unspecified function of K_1 covariates $x_1, ..., x_{K_1}$ and $u \sim \mathcal{N}(0, \sigma^2)$. Assumptions:

- Additivity: $f(x_1, ..., x_{K_1}) = \sum_k^{K_1} f_k(x_k)$
- $f_k(x_k)$ are smooth functions called normal cubic splines which implies two conditions:
  - $f_k'(x_k), f_k''(x_k)$ are continuous between smallest and largest values of x_k
  - "natural" implies that $f_k''(x_k) = 0$ at smallest and largest value of x_k

The motive now is to reduce the unknown function f to $S\beta$ where S is the basis for cubic specification

Semiparametric Regression: Procedure I

The procedure is explained for x_k and subscript k is dropped

- Step 1: Set M knots $min(x) = \tau_1, \tau_2, \tau_3, ..., \tau_M = max(x)$, where $M < n$
- Step 2: Choosing a basis for representing $f(.)$. The LS (Lancaster and Salkauskas) basis are considered as the coefficient β come out to be the function ordinates at the knots. For any point x_i, the representation of $f(x)$ as cubic spline is:

$$f(x_i) = \sum_{m=1}^{M} (\Phi_m(x_i) b_m + \Psi_m(x_i) s_m), i = 1, ..., n \tag{6}$$

where

$$\Phi_m = \begin{cases} 0 & x_i < \tau_{m-1} \\ \frac{-2}{h_m^3}(x_i - \tau_{m-1})^2 (x_i - \tau_m - 0.5h_m) & \tau \leq x_i < \tau_m \\ \frac{2}{h_{m+1}^3}(x_i - \tau_{m+1})^2 (x_i - \tau_m + 0.5h_{m+1}) & \tau_m leq x_i < \tau_{m+1} \\ 0 & x_i \geq \tau_{m+1} \end{cases} \tag{7}$$

Semiparametric Regression: Procedure II

$$\Psi_m = \begin{cases} 0 & x_i < \tau_{m-1} \\ \frac{1}{h_m^2}(x_i - \tau_{m-1})^2(x_i - \tau_m) & \tau \leq x_i < \tau_m \\ \frac{1}{h_{m+1}^2}(x_i - \tau_{m+1})^2(x_i - \tau_m) & \tau_m leq x_i < \tau_{m+1} \\ 0 & x_i \geq \tau_{m+1} \end{cases} \tag{8}$$

where $h_m = \tau_m - \tau_{m-1}$ and $f(\tau_m) = b_m, f'(\tau_m) = s_m$ Following the approach of M. Lancaster and Salkauskas (1986), $f(x_i)$ can be reduced to:
$f(x_i) = z_i' b$
where $z_i' = \Phi'(x_i) + \Psi'(x_i)A^{-1}C$
$\Phi'(x_i) = \big(\Phi_m(X_i), ..., \Phi_M(X_i)\big)$
$\Psi'(x_i) = \big(\Psi_m(X_i), ..., \Psi_M(X_i)\big)$

Semiparametric Regression: Identification

- From the expression of z_i before, the sum of each row of $Z = (z_1, z_2, ..., z_n)$ is 1 which is collinear with intercept in x_0
- To get identification of b, set $\sum_i^M b_m = 0$ which would lead to $z_i' b = b_2(z_i 2 - z_i 1) + ... + b_M(z_i m - z_i 1)$
- Define a $n \times M - 1$ basis matrix S:

$$S = \begin{bmatrix} z_{12} - z_{11} & z_{13} - z_{11} & \dots & z_{1M} - z_{11} \\ \vdots & \vdots & \vdots & \vdots \\ z_{n2} - zn1 & z_{n3} - z_{n1} & \dots & z_{nM} - zn1 \end{bmatrix} \tag{9}$$

$\beta = (b_2, ..., b_M)$ and the natural cubic spline approximation to $f(x)$ is $S\beta$

Semiparametric Regression: Estimation I

- Considering equation (5), the chosen priors are: $\beta_0 \in \mathcal{N}_{K_0}(b_{00}, B_{00})$ and $\sigma^2 \in IG(\alpha_0/2, \delta_0/2)$
- For β of $S\beta$, the priors are chosen on difference in ordinates at end notes and second difference of ordinates at interior notes: $\frac{b_2 - b_1}{h_2} = \frac{b_2 - (-1)(b_2 + b_3 + \ldots + b_M)}{h_2} = \frac{2b_2 + b_3 + \ldots + b_M}{h_2} \in \mathcal{N}(0, \sigma_e^2)$. Similarly, $\frac{b_M - b_{M-1}}{h_M} \in \mathcal{N}(0, \sigma_e^2)$, conditioned on variance σ_e^2.
- The prior for interior knots conditioned on variance σ_d^2 is: $\frac{b_{m+1} - b_m}{h_{m+1}} - \frac{b_m - b_{m-1}}{h_m} \in \mathcal{N}(0, \sigma_d^2)$, for $m = 3, \ldots, M-1$
- Thus, $\Delta\beta | \sigma_e^2, \sigma_d^2 \in \mathcal{N}_{M-1}(0, T) \implies \beta | \sigma_e^2, \sigma_d^2 \in \mathcal{N}_{M-1}(0, \Delta^{-1} T (\Delta^{-1})')$

Semiparametric Regression: Estimation II

- where:

$$\Delta = \begin{bmatrix} \frac{2}{h_2} & \frac{1}{h_2} & \frac{2}{h_2} & \frac{2}{h_2} & \frac{2}{h_2} & \dots & \frac{2}{h_2} & \frac{2}{h_2} \\ \frac{1}{h_2} & -(\frac{1}{h_2}+\frac{1}{h_3}) & \frac{1}{h_3} & 0 & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{h_3} & -(\frac{1}{h_3}+\frac{1}{h_4}) & \frac{1}{h_4} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & & & \\ 0 & 0 & 0 & 0 & 0 & \dots & -\frac{1}{h_M} & \frac{1}{h_M} \end{bmatrix}$$

$$T = \begin{bmatrix} \sigma_e^2 & 0 & 0 \\ 0 & \sigma_d^2 \mathcal{I}_{M-3} & 0 \\ 0 & 0 & \sigma_e^2 \end{bmatrix}$$

- $\sigma_e^2 \in IG(\alpha_{e0}/2, \delta_{e0}/2)$ and $\sigma_d^2 \in IG(\alpha_{d0}/2, \delta_{d0}/2)$
- Genralising for K_1 functions $f_k(x_k)$ and allow different x_k to take different number of knots M_k, regression coeffiecients β_k and smoothing parameters $(\sigma_{ek}^2, \sigma_{dk}^2)$. Thus, y is: $y = X_0\beta_0 + S_1\beta_1 + ... + S_{K_1}\beta_{K_1} + u$

Semiparametric Regression: Estimation III

- Thus, $y = X_0\beta_0 + S\beta_1 + u$ where $S = (S_1, ..., S_{K_1}), \beta_{(1)} = (\beta_1', ..., \beta_{K_1}')$
- Combining all the covariates and parameters into linear regression:

$$y = W\beta + u \tag{10}$$

where $W = (X_0, S)$ and $\beta = (\beta_0', \beta_{(1)}')$

- This leads to the prior distribution for $\beta|\{\sigma_e^2, \sigma_d^2\} \in \mathcal{N}_K(b_0, B_0)$ where

$$B_0 = \begin{bmatrix} B_{00} & 0 & 0 & \dots & 0 \\ 0 & \Delta_1^{-1}T_1(\Delta_1^{-1})' & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \Delta_{K_1}^{-1}T_{K_1}(\Delta_{K_1}^{-1})' \end{bmatrix}$$

$b_0 = (b_{00}', 0_{K-K_0}') where K = K_0 + \sum_k^{K_1} M_k - K_1$

Ordinal Models: Introduction

- Ordinal Model arise when the dependent variable is discrete and are ordered with only an ordinal meaning and no cardinal interpretation.
- The model is represented using continuous latent variable z_i:

$$z_i = x_i' \beta_p + \epsilon_i, \qquad \forall i = 1, ..., n \tag{11}$$

- The observed variable can have J outcomes which are related to z_i through the cut point vector γ_p:

$$\gamma_{p,j-1} < z_i \leq \gamma_{p,j} \implies y_i = j, \qquad \forall i = 1, ..., n; j = 1, ..., J \tag{12}$$

where $\gamma_{p,0} = -\infty$ and $\gamma_{p,j} = \infty$, location restriction needs that $\gamma_{1,p} = 0$ and scale restriction implies that variance $V(\epsilon) = \frac{1-2p+2p^2}{p^2(1-p)^2}$ is fixed which is true for a given p

- Ordering of cut points causes complication in sampling as it is difficult to maintain order while sampling, which is solved by a monotone transformation, a logarithmic transformation:

$$\delta_{p,j} = \ln(\gamma_{p,j} - \gamma_{p,j-1}), \qquad 2 \leq j \leq J-1 \tag{13}$$

Ordinal Models: Estimation

- The normal-exponential mixture representation of asymmetric Laplace distribution is used for estimating quantile ordinal model, hence z_i:

$$z_i = x_i'\beta_p + \theta w_i + \tau\sqrt{w_i}u_i, \qquad \forall i = 1, ..., n \tag{14}$$

Thus, $z_i|\beta_p, w_i \in \mathcal{N}(x_i'\beta_p + \theta w_i, \tau^2 w_i)$

- The priors for β_p, δ_p and w_i are:

$$\beta_p \in \mathcal{N}(\beta_{p0}, B_{p0}) \tag{15}$$

$$\delta_p \in \mathcal{N}(\delta_{p0}, D_{p0}) \tag{16}$$

$$w_i \in \varepsilon(1) \tag{17}$$

- Thus, the joint posterior density can be written as:

$$\pi(z, \beta_p, \delta_p, w|y) \propto \left\{ \prod_{i=1}^{n}\prod_{j=1}^{J} 1\{\gamma_{p,j-1} < z_i \le \gamma_{p,j}\}\mathcal{N}(z_i|x_i'\beta_p + \theta w_i, \tau^2 w_i \right.$$

$$\left. (w_i|1) \right\} \times \mathcal{N}(\beta_{p0}, B_{p0})\mathcal{N}(\delta_{p0}, D_{p0} \tag{18}$$

Future Work

- The semiparametric ordinal model, in terms of latent variable y_i^* (just to avoid confusion with z_i, used in Semiparametric study), is given as:

$$y_i^* = x_{0i}'\beta_p + f(x_{1i}, ..., x_{K_1 i}) + \theta w_i + \tau\sqrt{w_i}u_i, \qquad \forall i = 1, ..., n \quad (19)$$

  where the priors for parameters as same as discussed in ordinal models and semiparametric regression. The future work would be focussed on developing Sampling Algorithm for the above model
- Thus, the model in (18) would be written as: $Y^* = W\beta + u$ as done in Semiparametric
- Next step would involve introducing the cut-point γ_p^* which would relate Y^* to the observed variable Y on the similar lines of discussion of ordinal models

References I

- Koenker, R. and Bassett, G. (1978). Regression Quantiles. Econometrica, 46(1): 3350. MR0474644. doi: http://dx.doi.org/10.2307/1913643
- Kozumi, H. and Kobayashi, G. (2011). Gibbs Sampling Methods for Bayesian Quantile Regression. Journal of Statistical Computation and Simulation, 81(11): 1565 1578. MR2851270. doi:http://dx.doi.org/10.1080/00949655.2010.496117.
- Rahman, M. A. (2016). Bayesian Quantile Regression for Ordinal Models. doi:10.1214/15-BA939. https://projecteuclid.org/euclid.ba/1423083637
- Jeliazkov, I., Graves, J., and Kutzbach, M. (2008). Fitting and Comparison of Models for Multivariate Ordinal Outcomes. Advances in Econometrics: Bayesian Econometrics, 23: 115156. doi: http://dx.doi.org/10.1016/S0731-9053(08)23004-5.

References II

- Teh, Y W. (2010). Dirichlet Process. Encyclopedia of Machine Learning, Springer.
- Greenberg, Edward. Introduction to Bayesian Econometrics. Second Edition. Cambridge University Press.
- Lin X. and Carroll R. J. (2001). Semiparametric Regression for Clustered Data Using Generalized Estimating Equations, Journal of the American Statistical Association, 96:455, 1045-1056, doi: `https://doi.org/10.1198/016214501753208708`
- Willink R. (2008). What is robustness in data analysis? Metrologia, Volume 45, Number 4, doi: `https://doi.org/10.1088/0026-1394/45/4/010`
- Chib S. and Greenberg E. (1995) "Understanding the Metropolis-Hastings Algorithm." The American Statistician 49, no. 4: 327-35. doi: `https://doi.org/10.2307/2684568`.