Random Effects Quantile Regression

Adam Maidman

December 8, 2014

1 Introduction

In a 2011 *Journal of the American Statistical Association* paper, Kim and Yang propose a semi-parametric approach to a random effects quantile regression model [2]. This quantile regression version of the well-known mean mixed effects model is an important contribution to statistics for obvious reasons which – for the sake of brevity – we won’t discuss here. Kim and Yang’s semi-parametric approach to the random effects quantile regression model allows for the modeling of the parametric random effects while avoiding taking into account any error distribution. The random effects are not required to be identically distributed or even Gaussian. Maximizing the semiparametric likelihood-like criterion function is very difficult and computationally expensive to calculate directly. Instead, a Bayesian approach is taken so that the estimator can be found using Markov chain Monte Carlo (MCMC) samplers.

Although Kim and Yang’s approach appears to yield good results, I critique the length of the MCMC chains and thus some of their estimations. The length of the authors’ chain was determined using diagnostics and suggestions from Cowles and Carlin’s 1996 paper [1]. The authors perhaps should have considered running the chain longer and choosing a stopping point based on the Monte Carlo Standard Error (mc-se) instead.

In the next two sections of this paper, we will discuss the random effects quantile regression model, the necessity for an MCMC approach, and compare numerical results in a simulation study.

2 Random Effects Quantile Regression Model

The random effects quantile regression model is useful when sampling from a population with several clusters. Suppose we have $n$ clusters with size $m_i$, where $i = 1, \ldots, n$. Let $\{y_{ij}, x_{ij}\}$, $j = 1, \ldots, m_i$ denote observations from the $i$th cluster drawn from the model

$$Q_{i,\tau}\{ y \mid x, b_i(\tau) \} = x^T b_i(\tau) \text{ for } \tau \in (0, 1).$$

We assume that $x \in \Xi_i \subseteq \mathbb{R}^q$ are random covariates including an intercept where $\Xi_i$ is the support of $x$ in the $i$th cluster with cluster specific density $g_i(b|\theta)$, and $Q_{i,\tau}\{y|x, b_i(\tau)\}$ is the $\tau$th conditional quantile of the response $y \in \mathbb{R}$ in the $i$th cluster. The $b_i(\tau)$’s are random $q$-dimensional variables with support on $\Upsilon_i$, common mean $\beta(\tau)$, and cluster specific covariance matrices $\Sigma_i(\tau)$. For ease
of notation, we omit the $\tau$ from future equations and expressions. To compare the quantile model with the well-known mean model, we write

$$y_{ij} = x_{ij}^T \beta + x_{ij}^T b_i^* + e_{ij}$$

where the $b_i^*$ are mean 0 random effects. This equation leads to the nice interpretation that if the random effects are positive, then the cluster has a higher response than the population average at that specific quantile.

From conditions stated in Kim and Yang’s paper which are standard for quantile regression models, we define the estimator for the $i^{th}$ cluster as

$$\tilde{b}_i = \arg\min_{b_i \in \mathbb{R}^q} \sum_{j=1}^{m_i} \rho_{\tau}(y_{ij} - x_{ij}^T b_i),$$

where $\rho_{\tau}(u) = u(\tau - I(u \leq 0))$ is the check function. Some technical details leads to the empirical likelihood for the cluster specific coefficient $b_i$,

$$L_{m_i}(b) = \max \left\{ \prod_{j=1}^{m_i} p_j \left| \sum_{j=1}^{m_i} p_j \varphi_{\tau}(y_{ij} - x_{ij}^T b) x_{ij} = 0, \sum_{j=1}^{m_i} p_j = 1, 0 \leq p_j \leq 1 \right. \right\},$$

where $\varphi_{\tau}$ is the derivative of the check function. We now define a semiparametric likelihood-like criterion function for $b_i$ and $\theta$, respectively, by

$$\tilde{L}_{m_i}(b_i|\theta) = L_{m_i}(b_i) g_i(b_i|\theta),$$

$$\prod_{i=1}^{n} \int_{b_i} \tilde{L}_{m_i}(b_i|\theta) db_i.$$

where $\theta$ is the parameter(s) necessary to define the density $g_i(b_i|\theta)$. Under some regularity conditions, the estimators for the $b_i$’s and $\theta$ are

$$\tilde{b}_{m_i} = \arg\max_{b_i \in \Upsilon_i} \tilde{L}_{m_i}(b_i|\theta_n),$$

$$\tilde{\theta}_n = \arg\max_{\theta \in \Theta} \prod_{i=1}^{n} \int_{b_i} \tilde{L}_{m_i}(b_i|\theta) db_i.$$

Computation of these estimators is daunting, so a Bayesian framework is proposed instead. The posterior mean will be the value of interest.

3 Markov Chain Monte Carlo

Suppose $\theta = (\beta, \Sigma)$ and $g_i(b_i|\theta)$ is the Gaussian density with mean $\beta$ and covariance $\Sigma$. Further suppose that the prior for $\theta$ is the improper prior $\pi(\beta, \Sigma) \propto |\Sigma|^{-(q+1)/2}$ which produces a proper posterior

$$\Sigma|b_1, \ldots, b_n \sim IW_{n-1}(S_n),$$

$$\beta|\Sigma, b_1, \ldots, b_n \sim N_q \left( n^{-1} \sum_{i=1}^{n} b_i, n^{-1} \Sigma \right),$$

where $S_n = \sum_{i=1}^{n} (b_i - \tilde{b}_i)(b_i - \tilde{b}_i)^T$. This yields the target density for the $b_i$

$$\tilde{\omega}_n(b_i|\theta) \propto L_{m_i}(b_i) g_i(b_i|\theta).$$
The authors suggest using an independent Metropolis-Hastings algorithm with a Gaussian proposal density to sample from $\tilde{\omega}_n(b_i|\theta)$. Acceptance probabilities should be in the range of $0.1 \sim 0.4$. I had difficulty finding a proposal density which yielded acceptance probabilities in this range, so I decided to run a symmetric Metropolis-Hastings algorithm with a Gaussian proposal density with a diagonal covariance matrix instead.

Following suggestions laid out by Cowles and Carlin’s MCMC diagnostics review paper, Kim and Yang ran their chain for only 4,000 iterations and discarded the first 1,000 iterations as burn-in, leaving a chain of length 3,000. I ran my chain for 35,000 iterations. I compared both chains under the following simulation

$$Q_{i,\tau}\{y \mid x, b_i(\tau)\} = x^T b_i(\tau), \quad b_i \in \mathbb{R}^2, \quad n = 4, \quad m_i = 20, \quad \tau = 0.5,$$

with independent and identically distributed Gaussian errors. Results for $b_{4,1}$, the first component of the 4\textsuperscript{th} cluster, are shown below.

The trace plot displays all 35,000 draws and highlights the 3,000 draws selected using Kim and Yang’s method. The autocorrelation plot shows the correlation with different lags for the 1,000\textsuperscript{th} through 4,000\textsuperscript{th} steps. The autocorrelation plot for the full 35,000 step chain looks similar and is not shown here. Although the trace plot for the 3,000 steps appears to be similar to the full 35,000 step chain, a 3,000 step chain is not comparable to the full chain and does not capture the mean as well as the full chain. Additionally, the autocorrelation plot suggests that “independent” draws are about 46 steps apart. This is unnerving for a chain of length 3,000.

<table>
<thead>
<tr>
<th>Steps</th>
<th>Estimates</th>
<th>MCSE</th>
<th>Estimates ± 2SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,000-4,000</td>
<td>1.326726</td>
<td>0.06241951</td>
<td>(1.202, 1.451)</td>
</tr>
<tr>
<td>1-35,000</td>
<td>1.247893</td>
<td>0.02210512</td>
<td>(1.204, 1.292)</td>
</tr>
</tbody>
</table>

Table 1: Estimates and Monte Carlo Standard Errors

Lastly, Table 1 shows that the estimates from the 3,000 step chain cannot be trusted because the interval in column 4 spans 1.2 - 1.4. Clearly, the longer chain is necessary as it estimates a reliable posterior mean of 1.2. Although Kim and Yang’s method can be used to find a posterior mean for this random effects quantile regression model, more thought regarding the number of steps for the Markov chain was needed.
References
