FACULTEIT ECONOMIE
EN BEDRIJFSKUNDE

TWEEKERKENSTRAAT 2
B-9000 GENT
Tel. : 32 - (0)9 – 264.34.61
Fax. : 32 - (0)9 – 264.35.92

WORKING PAPER

Binary quantile regression: A Bayesian approach based on the asymmetric Laplace density

Dries F. Benoit[1]

Dirk Van den Poel[2]

August 2010

2010/662

---

[1] PhD Candidate, Ghent University
[2] Corresponding author: Prof. Dr. Dirk Van den Poel, Professor of Marketing Modeling/analytical Customer Relationship Management, Faculty of Economics and Business Administration, dirk.vandenpoel@ugent.be; more papers about customer relationship management can be obtained from the website: www.crm.UGent.be

D/2009/7012/33

Binary quantile regression: A Bayesian approach based on the asymmetric Laplace density

Dries F. Benoit and Dirk Van den Poel

Ghent University, Faculty of Economics and Business Administration, Department of Marketing, Tweekerkenstraat 2, B-9000 Ghent, Belgium.

Abstract

In this article, we develop a Bayesian method for quantile regression in the case of dichotomous response data. The frequentist approach to this type of regression has proven problematic in both optimizing the objective function and making inference on the regression parameters. By accepting additional distributional assumptions on the error terms, the Bayesian method proposed sets the problem in a parametric framework in which these problems are avoided, i.e. it is relatively straightforward to calculate point predictions of the estimators with their corresponding credible intervals. To test the applicability of the method, we ran two Monte-Carlo experiments and applied it to Horowitz' (1993) often studied work-trip mode choice dataset. Compared to previous estimates for the latter dataset, the method proposed interestingly leads to a different economic interpretation.

Keywords: quantile regression, binary regression, maximum score, asymmetric Laplace distribution, Bayesian inference, work-trip mode choice.

Corresponding author: Dirk Van den Poel (Dirk.VandenPoel@UGent.be); Tel. +32 9 264 89 80; Fax. + 32 9 264 42 79; website about CRM teaching: http://www.mma.UGent.be website about CRM research: http://www.crm.UGent.be .

1. Introduction

The classical theory of linear models focuses on the conditional mean function, i.e. the function that describes how the mean of *y* changes with the vector of covariates *x*. However, the mean may not be of prime interest for the researcher or additional information might be required about the whole conditional distribution of the response variable. For example, Whittaker et al. (2005) show that bank account use requires different predictors for different quantiles of the response distribution. Moreover, least squares methods, which focus on the conditional mean function, assume that the error has exactly the same distribution whatever values of *x* are taken. The components of the vector of *x* are expected to affect only the location of the conditional distribution of *y*, not to affect its scale or any other aspect of its distributional shape. In practice, however, these assumptions are often hard to maintain.

Quantile regression (Koenker and Basset, 1978; Koenker, 2005) extends the mean regression model to conditional quantiles of the response variable. Note that quantile regression comprises median regression (or equivalently L_1-regression), as the median is the most central quantile which separates the higher half of a sample from the lower half. The technique provides a more nuanced view of the relationship of the dependent variable and the covariates, since it allows the user to examine the relationship between a set of covariates and the different parts of the distribution of the response variable. An additional advantage is that parameter estimates of the quantile regression approach are not biased by a location-scale shift of the conditional distribution of the dependent variable. These two distinct advantages were not only acknowledged by theoretical statisticians, but have also encouraged researchers from varying disciplines to apply quantile regression in their research. The applications range from the field of ecology (e.g. Brown and Peet, 2003) over cancer research (e.g. Li and Zhu, 2007) to economics (e.g. Buchinski, 1994; 1998). See Yu et al. (2003) for a more complete overview of the different fields of quantile regression applications.

Furthermore, quantile regression has been extended to model dependent variables other than ratio/scale variables. These extensions include, *inter alia*, models for left-censored data (Powell, 1986; Yu and Stander, 2006), count data (Machado and Santos Silva, 2005) or proportions (Hewson and Yu, 2008).

In the case of a binary response variable, adopting quantile regression is not an obvious choice. The dependent variable takes on only two values and hence does not yield continuous quantiles that can be modelled via regression. However, several authors have recognized the potential benefits of binary quantile regression. Manski (1975; 1985) defined the general semi-parametric binary quantile regression estimator. For unclear reasons, subsequent research has focused exclusively on the median case (Koenker and Hallock, 2001). Kordas (2006) has recently explored the consequences of estimating other quantiles than the median for binary regression models and has shown that also in the dichotomous case the approach leads to a much richer view of how covariates influence the response variable.

The frequentist approach to binary quantile regression, emerging from Manski's work, faces some major technical drawbacks. First, the method has difficulty optimizing the regression parameters. Moreover, building confidence intervals around the estimates has proven problematic. In this paper, we adopt a Bayesian approach to binary quantile regression, which is quite different from previous approaches in this context. We show, both theoretically and in three applications, how our Bayesian approach to binary quantile regression can avoid the difficulties outlined above by imposing stronger assumptions on the error terms.

The remainder of the paper is organised as follows: Section 2 extends the ideas of median (L_1) regression to the quantile regression approach. For better understanding, we consider both the frequentist approach and the more recent Bayesian approach based on the asymmetric Laplace distribution (ALD). Section 3 treats the frequentist approach to binary quantile regression and discusses its limitations in optimizing and inference. Section 3 further shows how the Bayesian approach proposed avoids these difficulties by putting the problem in a parametric framework. In Section 4, this new strategy is then applied to two Monte-Carlo experiments and to one real-life application (i.e. Horowitz' (1993) often studied work-trip mode choice dataset). Finally, Section 5 presents the main findings, some limitations of the method and directions for further research.

2. From median regression to quantile regression

Consider the standard model where *y* and *x* are both continuous variables:

$$y_i = \mu(x_i) + \varepsilon_i \, . \tag{1}$$

If we assume that $Med(\varepsilon \mid x) = 0$, then $\mu(x_i)$ is a conditional median function. Since we assume that the relation between y and x is linear, we obtain a linear conditional median (L_1) model:

$$Med(y_i \mid \mathrm{x_i}) = x_{\mathrm{i}}^{'} \beta \, . \tag{2}$$

In this model, we find the regression coefficients by solving:

$$\arg\min_{\beta \in \Re} \sum_{i=1}^{n} | y_i - x_i^{'} \beta | \ . \tag{3}$$

Quantile regression proceeds by extending the median case to all other quantiles of interest. Contrary to the commonly used quadratic loss function for mean regression, the quantile regression links to a special class of loss functions which has robust properties (Huber, 1981).

$$\beta_{hat}(\tau) = \arg\min_{\beta \in \Re} \sum_{i=1}^{n} \rho_{\tau}(y_i - x_i^{'} \beta) \ . \tag{4}$$

In this equation, $\tau \in (0,1)$ is any quantile of interest, $\rho_{\tau}(z) = z(\tau - \mathrm{I}(z < 0))$ and $\mathrm{I}(\cdot)$ denotes the indicator function. The quantile $\beta_{hat}(\tau)$ is called the τ^{th} regression quantile. Note that the case where τ equals 0.5, which minimizes the sum of absolute residuals, corresponds to median (L_1) regression. Frequentist approaches to quantile regression then construct confidence intervals for $\beta_{hat}(\tau)$ by asymptotic theory or by bootstrapping.

Koenker and Machado (1999) were the first to show that likelihood based inference using independently distributed asymmetric Laplace densities is directly related to the minimization problem in Equation 4. This finding was picked up by Yu and Moyeed (2001) and was the start for the development of a Bayesian approach to quantile regression. It should be noted however, that some other Bayesian approaches have emerged mostly for median, rather than full quantile regression. Tsionas (2003) proposed an approach based on a scale mixture of normals, which itself leads to an ALD. Other methods are based on Dirichlet Process Priors (Kottas and Gelfand, 2001; Kottas and Krnjajic, 2009) or substitution likelihoods (Dunson and Taylor, 2005). However, the above semiparametric methods for quantile regression require complex choices of prior distributions and prior (hyper-) parameters. This is avoided in the methodology proposed here.

We concur with Hewson and Yu (2008) and Yu and Stander (2007) that this is one of the distinct advantages of Bayesian quantile regression based on the ALD.

In contrast to other parameterizations of the ALD (see Kotz et al., 2001), Yu and Zhang (2005) propose a three-parameter ALD with a skewness parameter that can be used directly to model the quantile of interest:

$$f_p(y \mid \mu, \sigma, \tau) = \frac{\tau(1\text{-}\tau)}{\sigma}\exp\left\{-\rho_\tau\left(\frac{y-\mu}{\sigma}\right)\right\}, \tag{5}$$

where

$$\rho_\tau(y) = y(\tau - I(y < 0)). \tag{6}$$

Equation 6 is identical to the loss function in the optimization problem in Equation 4. Thus, minimizing Equation 4 is equivalent to maximizing a regression likelihood using ALD errors with $\mu = x_i'\beta$ (Yu and Moyeed, 2001). Bayesian implementation of quantile regression begins by forming a likelihood comprised of independent asymmetric Laplace densities with $\mu = x_i'\beta$, specifying the quantile of interest, τ, and placing priors on the model parameters β and σ. Inference about model parameters then follows conventional Bayesian procedures which lead to exact inference about $\beta_{hat}(\tau)$ as opposed to the frequentist asymptotic inference which has shown to be unreliable in this context (Bilias et al., 2000).

[INSERT FIGURE 1 ABOUT HERE]

Figure 1 shows how the skewness of the ALD changes with altering values for τ. For example where $\tau = 0.1$ almost al the mass of the ALD is situated in the right tail. In the case where $\tau = 0.5$ both tails of the ALD have equal mass and the distribution then equals the more common double exponential distribution. In contrast to the normal distribution with a quadratic term in the exponent, the ALD is linear in the exponent. This results in a more peaked mode for the ALD together with thicker tails. On the other hand, the normal distribution has heavier shoulders compared to the ALD.

3. Binary quantile regression

3.1 Standard binary regression

The most frequently used form of the binary response model is:

$$\begin{aligned} y_i^* &= x_i^{'}\beta + u_i\,, \\ y_i &= 1 \text{ if } y_i^* \geq 0\,, \\ y_i &= 0 \text{ otherwise}\,. \end{aligned} \tag{7}$$

Where y_i is the indicator of the i^{th} individual's response determined by the underlying latent variable y_i^*, x_i is a 1 x k vector of explanatory variables, β is a k x 1 vector of parameters, u_i is a random error term and $i = 1, \ldots, n$.

Let $F(u|x)$ denote the cumulative distribution function of u conditional on the event $x_i = x$.

$$P(y_i = 1 \mid x_i, \beta) = 1 - F(-x_i^{'}\beta_i)\,. \tag{8}$$

Often it is assumed that $F(u|x)$ is either the cumulative normal or the cumulative logistic distribution, independent of x. In the former case, the model described in Equation 7 results in the binary probit model. In the latter case, the model becomes the binary logit model. The symmetry of the logistic or normal distribution makes it possible to further simplify the right hand side of Equation 8 to $F(x_i^{'}\beta_i)$.

3.2 Binary quantile regression: frequentist approaches

The first step towards a binary quantile regression model in econometrics was set by Manski's Maximum Score Estimator (Manski, 1975, 1985). This estimator imposes extremely weak assumptions on the distribution of the error term. The only condition is that the median of *u* conditional on *x* is zero (see Equation 7). Thus, the maximum score estimator does not require the researcher to know the functional form of the relationship between *x* and the distribution of *u*. Furthermore, it accommodates for heteroskedasticity of unknown form. Manski (1975) mainly focused on the median case, but later he acknowledged extending the estimation to the more general quantiles (Manski, 1985). This leads to the following form of the maximum score estimator:

$$\beta_{hat}(\tau) = \arg\max_{\beta \in \Re} n^{-1} \sum_{i=1}^{n} \rho_\tau (2y_i - 1)(2y_i - 1)\,\mathrm{sgn}(x_i^{'}\beta)\,. \tag{9}$$

for any quantile $\tau \in (0,1)$. With, sgn(·) is the signum function and again $\rho_\tau(\cdot)$ is the loss function as in Equation 4 and 6.

Scale normalization is needed because the parameter β is identified only up to a scale. Note that $I(\sigma\, x_i^{'}\beta > 0) = I(x_i^{'}\beta > 0)$ for all $\sigma > 0$. Two types of scale normalization are frequently used. For the first type, the condition $\|\beta\| = 1$ could be imposed or, for the second type, one coordinate of β could be set to unity (either +1 or -1). In the former case $\|\beta\|$ denotes the Euclidian norm of vector β. The latter case presumes some kind of a priori information on the sign of the fixed element of β. In many situations reasonable guesses about the sign of this coordinate might prove difficult to make.

Kim and Pollard (1990) showed that the maximum score estimator has a slow rate of convergence and a complicated asymptotic distribution. The complexity of the limiting distribution of the maximum score estimator limits its usefulness for statistical inference. This complexity follows directly from the absence of an asymptotic first-order condition of the objective function in Equation 9.

Delgado et al. (2001) have aimed to solve the problem by using subsampling. They theoretically justify subsampling for the maximum score estimator and provide simulation evidence that suggests inconsistency of the bootstrap. The latter was eventually proved by Abrevaya and Huang (2005). One major drawback of this method is the great computational expense and consequently it is only applicable in low dimensional problems with small sample sizes (Delgado et al., 2001). An alternative solution to the intractable limiting distribution of the maximum score estimator was proposed by Horowitz (1992). By using ideas related to the kernel method in nonparametric density estimation, he smoothed Manski's maximum score function so that it becomes continuous and differentiable. This method should lead to an asymptotically normal distribution for the smoothed median estimator. Recently, Kordas (2006) has extended this method to the quantiles other than the median. A drawback of this smoothed estimator is that it requires stronger restrictions on the smoothness of the error distribution than Manski's original estimator (Horowitz, 1992). Furthermore, simulation studies indicate that even for sample sizes of n = 1000 the normal approximation is inaccurate and even with bootstrapping it is difficult to estimate standard errors for this estimator (Abrevaya and Huang, 2005). Finally, Koenker (2005) points out that the, rather arbitrary, selection of smoothing parameters highly influences wether smoothing actually improves inference in applications. Kotlyarova and Zinde-Walsh (2006), however, provide a method for selecting the optimal bandwidth that shows good performance on large datasets.

Recently, Skouras (2003) and Florios and Skouras (2008) focussed on the problematic optimization of the objective function in Equation 9. Florios and Skouras (2008) give an overview of all empirical applications of the maximum score estimator, including Horowitz (1993) and conclude that none of the algorithms used guarantee a global optimal solution. They propose reformulating the problem as mixed integer programs (MIP) and show the superior performance on some real and simulated datasets. Nonetheless, the authors explicitly avoid inference for their estimator: *"we have not attempted to evaluate statistical significance because […] standard errors for maximum score estimators are difficult to estimate, even with bootstrapping (Abrevaya and Huang, 2005)"* (Florios and Skouras, 2008, p.88).

The elements above make clear that in the frequentist approach to binary quantile regression, both calculating a consistent estimator and making inferences about this estimator is problematic. Several solutions are proposed for this problem, but these all have specific drawbacks. In the next section, our Bayesian approach to the problem is developed.

3.3 Binary quantile regression: a Bayesian approach

Consider again the binary response model in Equation 7. The method proposed for binary quantile regression makes use of data augmentation (Tanner and Wong, 1987). The idea of data augmentation has shown to be effective in the Bayesian approach to many regression methods for binary or multinomial dependent variables (e.g. Albert and Chib, 1993; Groenewald and Mokgatlhe, 2004; Holmes and Held, 2006) and we hold that this is also the case for the current approach to binary quantile regression. Therefore, n latent variables $y_1^*,\ldots,y_n^*$ are introduced into the problem. These latent variables are asymmetric Laplace distributed as described in Equation 5.

$$y^* \sim ALD(\mu = x_i^{'}\beta\,,\ \sigma = 1,\ \tau). \tag{10}$$

The parameter σ is set to unity for identification reasons, similar to the reasons outlined in Section 3.2. The parameter τ should be specified at the quantile of interest. For example, $\tau = 0.5$ in the case of binary median regression. Further, define $y_i = 1$ if $y_i^* > 0$ and $y_i = 0$ if $y_i^* < 0$. Then,

$$P(y_i = 1 \mid x_i, \beta) = 1 - F_{y^*}(-x_i^{'}\beta)\,, \tag{11}$$

where $F_{y^*}(\cdot)$ is the cumulative distribution function of the asymmetric Laplace variable y^*.

The joint posterior density of the unobservables β and y^* given the data $y = (y_1, y_2, \ldots, y_n)$ and the quantile of interest, τ, is then given by:

$$\pi(\beta, y^* \mid y, \tau) \propto \quad \pi(\beta)\prod_{i=1}^{n}\{I(y_i^* > 0)I(y_i = 1) + I(y_i^* \le 0)I(y_i = 0)\}\, F_{y^*}(y_i^*; x_i^{'}\beta, 1, \tau)\,. \tag{12}$$

where $\pi(\beta)$ is the prior on the regression coefficients and $I(\cdot)$ is the indicator function. This joint posterior distribution does not fit any known class of distributions. Therefore, it is not possible to sample from this posterior directly. However, thanks to the development of Markov Chain Monte Carlo (MCMC) algorithms, computing this kind of posterior becomes fairly straightforward. Splitting up the complicated posterior in the posterior distribution of β conditional on y^*, and in the posterior distribution of y^* conditional on β, often facilitates sampling from the joint posterior. In the current case, one of the two fully conditional distributions is of a known form. This suggests a *Metropolis-Hastings within Gibbs* algorithm as an appropriate sampling scheme for the current setting.

From Equation 12, we find that the fully conditional distribution of y^* is given by:

$$\pi(y^* \mid \beta, y, \tau) \sim ALD(x_i^{'}\beta, 1, \tau) \text{ truncated at the left by 0, if } y_i = 1,$$

$$\pi(y^* \mid \beta, y, \tau) \sim ALD(x_i^{'}\beta, 1, \tau) \text{ truncated at the right by 0, if } y_i = 0. \qquad (13)$$

This is a distribution of a known form and consequently, direct sampling is possible. Sampling from the three-parameter ALD is straightforward as the ALD occurs as a simple linear combination of two independent exponential variates (Yu and Zhang, 2005). If ξ and η are independent and identical standard exponential distributions, then $\frac{\xi}{p} - \frac{\eta}{(1-p)} \sim$ ALD(0,1,p). Analogous with the normal distribution, any ALD can be derived from the standard ALD. That is, if X ~ ALD(0,1,p), then Y ~ $\mu + \sigma$X ~ ALD(μ,σ,p).

Next, from Equation 12 we can derive that the posterior density of β given y^*, τ and data is given by:

$$\pi(\beta \mid y^*, y, \tau) \propto \pi(\beta) \prod_{i=1}^{n} F_{y^*}(y_i^*; x_i^{'}\beta, 1, \tau). \qquad (14)$$

This fully conditional posterior density is in fact the posterior density for the regression parameter in the Bayesian quantile regression as discussed in Section 2. In contrast to the fully conditional posterior density for the latent data, this posterior is of an unknown form. A standard conjugate prior distribution is not available for the quantile regression formulation (Yu and

Moyeed, 2001), so MCMC methods may be used for extracting this posterior distribution, e.g. Metropolis-Hastings. This allows for the use of virtually any prior distribution on the regression parameters and even an improper uniform prior, $\pi(\beta) \sim U(-\infty,+\infty)$, will result in a proper posterior distribution, as proven in Yu and Moyeed (2001). Note that the posterior in Equation 14 is determined by assuming ALD distributed errors. This illustrates how the assumption of ALD errors influences estimation and inference of the model parameters. However, the applications in Section 4 indicate that the proposed method seems quite robust against departures from this assumption.

The *Metropolis-Hastings within Gibbs* sampler is now straightforward to implement. Given the data, the prior and the quantile of interest, the joint posterior distribution in Equation 12 can then be sampled from by sequentially drawing values from the distributions given in Equation 13 and 14. For every step, one should condition on the most recently drawn value of the conditioning arguments. Any value can be taken as a starting value, but good choices of starting values can strongly reduce the burn-in period of the algorithm proposed. For example, in the case where τ = 0.5, a good starting value could be the maximum likelihood estimate of the model under a probit or logit link. When a sufficiently large set of values is drawn from the joint posterior distribution, it becomes straightforward to compute point predictions for the model parameters, credible intervals or any other quantity of interest. This is a major advantage compared to the classical approaches to binary quantile regression where optimization and inference are awkward.

4. Applications

We apply our Bayesian approach to binary quantile regression to two Monte Carlo simulations and one real life example. In these examples, vague normal priors for β (with mean = 0 and standard deviation = 10) are chosen to minimize their influence on the posterior distributions. However, any prior distribution or parameter setting could be selected without eroding the methodology proposed here. Simulating realizations from the posterior distribution is done by the *Metropolis-Hastings within Gibbs* algorithm described in Section 3.3. For the Metropolis step, the update is performed using a random-walk Metropolis-Hastings algorithm with a Gaussian proposal density centred at the current state of the chain. The scale parameter of the proposal density was chosen so that an acceptance rate between 30% and 40% was achieved. Convergence of the MCMC chains was checked using the time-series plots of the draws of the different marginal distributions. All programs were written and executed in the free statistical package R.

4.1 Monte Carlo experiment 1

The purpose of this example is to illustrate how the proposed methodology is able to capture effects for discrete choice applications where heterogeneity of covariates is an issue. This situation occurs frequently and binary quantile regression is then a very appealing empirical strategy (Buchinski, 1998; Koenker and Hallock, 2001; Yu, Lu and Stander, 2003; Yu and Stander, 2007). Therefore, we generated n = 200 observations from the following heteroscedastic regression model:

$$y_i^* = 1.5x_i + \varepsilon_i, \qquad \text{with } \varepsilon_i \sim \mathrm{N}(0,\, 2x_i),$$

$$\text{and } x_i \sim \mathrm{U}(0, 10)\,. \qquad (15)$$

Figure 2 shows a graphical representation of the data together with some quantile regression lines for a number of quantiles of interest. These fits were obtained by using the Bayesian approach to quantile regression proposed by Yu and Moyeed (2001). Since the error variance of y^* is positively correlated with *x*, slope coefficients differ across quantiles.

INSERT FIGURE 2 ABOUT HERE

Figure 2 shows that the OLS estimate and the median regression estimate are quite similar. The regression lines are almost plotted on top of each other. Based on the mean and median fit, we might conclude that higher values on *x* will result in a higher value for *y*. However, the regression lines for quantiles other than the median give a more detailed insight into the effects. For the lower quantiles, the effect of the covariate exerts a negative effect on the dependent variable, while for the middle to higher quantiles the effect is positive. The effect becomes more pronounced for more extreme quantiles.

The binary regression model can be presented as the model in Equation 15, but with an extreme form of censoring of the dependent variable, y^*, from both above and from below. By defining $y_i = 0$, if $y_i^* \leq 0$ and $y_i = 1$, if $y_i^* > 0$, we now have simulated data from the binary response model as in Equation 7. The proposed method for binary quantile regression was applied to this data.

INSERT FIGURE 3 ABOUT HERE

Figure 3 shows the progress of the *Metropolis-Hastings within Gibbs* estimation of the regression slope for the different quantiles of interest. From the figure we can see that the chain mixes quite well. The influence of the initial values wears off fast and afterwards the chain concentrates and stabilizes on regions in the parameter space with higher probability.

INSERT TABLE 1 ABOUT HERE

Comparing the quantile regression lines in Figure 2 with the results of the Bayesian binary quantile regressions in Table 1, we find that the method proposed was able to recover the general effects that are present when the underlying latent variable is known. The results of the median model and the conventional logistic regression model are very similar. Looking at the beta parameters for these two models, we can conclude that the variable x exerts a positive effect on the dependent variable, y. For the other quantiles, as in Figure 2, the regression parameters suggest that the effect is negative for the lowest quantiles and is positive for middle and high quantiles. This example shows that, although only binary outcomes were observed and the errors were not exact ALD, the binary quantile regression approach proposed is able to expose the main effects of the covariate with the unobserved continuous latent variable.

4.2 Monte Carlo experiment 2

In this experiment, an extensive comparison of the proposed methods with the two main frequentist approaches ones is conducted. In their analysis of the Tobit quantile regression model, Yu and Stander (2007) used a model, which also appeared in Buchinsky and Hahn (1998), to compare their Bayesian estimator to existing fequentist estimators. The present experiment uses a very similar model, adjusted to the binary dependent variable situation:

$$y_i^* = -1 + 0.5x_{1i} + 1x_{2i} + \varepsilon_i \ . \qquad (16)$$

where the regressors were each drawn from a standard normal distribution and the error term has multiplicative heteroskedasticity obtained by taking $\varepsilon_i \sim \xi\nu(x_i)$ with $\xi \sim N(0,1)$ and $\nu(x_i) = 1 + 0.025(x_{1i} + x_{1i}^2 + x_{2i} + x_{2i}^2)$. Again, we define $y_i = 0$, if $y_i^* \leq 0$ and $y_i = 1$, if $y_i^* > 0$ to obtain the binary response model as in Equation 7.

A total of 1,000 Monte Carlo repetitions were conducted. For every repetition, we generated three datasets containing 200, 400 and 600 observations and three different approaches to binary quantile regression were fitted to the data (with $\tau = 0.5$): binary regression quantiles along with subsampling standard errors (Manski, 1985; Manski and Thompson, 1986), smoothed binary regression quantiles with asymptotic standard errors (Horowitz, 1992; Kordas, 2006) and finally the current Bayesian approach to binary quantile regression using the asymmetric Laplace density.

INSERT TABLE 2 ABOUT HERE

Table 2 summaries the biases, root mean square errors (RMSE) and 95% credible intervals for β_0 and β_1. The value of β_2 was set to unity to achieve scale normalization and is consequently not included in the table of parameters. As expected, results for all three methods show decreasing biases, RMSE and credible intervals for increasing sample sizes. The more data becomes available, the more the point estimate tends to the true value of the parameter and the more uncertainty about the estimate decreases. However, the figures clearly show considerably lower biases, lower mean square errors and more precise credible intervals for the method proposed compared to the other two methods, even without having exact ALD distributed errors. Note that the differences between the two frequentist methods on the one hand and the Bayesian method proposed on the other hand, decrease when more data becomes available. This finding corroborates the finding that BRQ's and sBRQ's require large amounts of data for reliable inference (Abrevaya and Huang, 2005; Kottas and Krnjajic, 2009). This indicates that when only a small amount of data is available, the Bayesian method proposed might be the better approach to binary quantile regression.

4.3 Real data: Work-trip mode-choice

Finally, the Bayesian procedure for the computation of the quantile regression estimates was tested on the widely studied transport-choice dataset described in Horowitz (1993) (also appearing in: McDonald, 1996; Gozalo and Linton, 2000; Horowitz, 2004 and Florios and Skouras, 2008). This dataset contains 842 observations randomly sampled from the Washington D.C. area transportation study. The data were obtained from home interviews and each record includes information for a single work trip: mode of transportation (DEPENDENT = 1, car), number of cars owned by the traveller's household (CARS), transit out-of-vehicle travel minus automobile out-of-vehicle travel time in minutes (DOVTT), transit in-vehicle travel time minus automobile in-vehicle travel time in minutes (DIVTT) and transit fare minus automobile travel cost in 1968 cents (DCOST). All continuous variables were standardized to have mean zero and unit standard deviation for both better comparison of the size of the effects and numerical stability of the method proposed. As in Horowitz (1993), the parameter DCOST is set to unity. This is not necessary for the Bayesian method where scale normalization is normally achieved by setting the variance of the error distribution equal to one, but it makes it easier to compare the results of the current study to those obtained in previous research.

INSERT TABLE 2 ABOUT HERE

Table 2 gives the estimates for the different parameters in the model for $\tau = 0.5$ (i.e. the median case). Also the computed estimates from Horowitz (1993) and Florios and Skouras (2008) are provided. Following the resulting estimates from Horowitz (1993), we could conclude that DCOST and CARS are the most important determinants of work-trip mode choice. Moreover, DCOST is by far the most important variable since its parameter is twenty times larger than the second largest parameter.

In contrast, using mixed integer programming (MIP) for optimizing the objective function in Equation 9, Florios and Skouras (2008) came up with very different results (see Table 2). It is remarkable that using a totally different methodology, the parameter estimates of the current Bayesian methodology are almost identical to the exact estimates obtained by the MIP method. Both MIP and the Bayesian estimates indicate that CARS is by far the most important variable. Note that if CARS is one or larger, the other variables must take on extremely negative values for the model to predict that no car is used for the work-trip. According to Florios and Skouras (2008), this finding suggests that the estimates for the other variables should be treated as zero. And indeed, they showed that keeping only CARS as covariate leads to a model predicting *travel by car* for car owners and *travel by transit* for non-car owners. This simple model, with only CARS as predictor, reduces the score by only 11 hits out of a total of 842. However, since Florios and Skouras (2008) do not provide standard errors for the estimates, concluding that a number of parameters in the model should be zero is somewhat guesswork. The Bayesian approach avoids this kind of speculations because the methodology provides exact and full inferences conditional on the data observed. Table 2 also contains the Bayesian credible intervals for the different parameters and they show that only the interval for DIVTT includes zero. This means we can be quite confident that the other variables do exert a positive effect on the work-trip transportation choice, unlike Florios and Skouras (2008) assumption. Also note that the estimates produced by the MIP method all fall in the 95% credible intervals provided by the Bayesian approach.

To obtain a more complete picture of the effects, a series of binary quantile regression models over the grid $\tau = \{0.05, 0.10, \ldots, 0.95\}$ is estimated. Figure 4 gives a graphical summary of this analysis. The point estimates plotted and the credible intervals are, respectively, the expectation, $Q_{.025}$ percentile and $Q_{.975}$ percentile obtained from the marginal posterior distribution of the different parameters. The solid line with filled dots represents the point estimates of the regression coefficients for the different quantiles. The shaded area depicting a 95% pointwise credible band is obtained from the marginal posterior distribution of the different parameters.

INSERT FIGURE 4 ABOUT HERE

For the interpretation of this kind of plots, it is recommended to imagine the underlying, unobserved continuous variable. In this application this is the willingness to take the car to go to work. From Figure 4, we can see that the effects of most variables become stronger for the higher conditional quantiles of the unobserved *willingness to take the car* distribution. This means that these variables exert heterogeneous effects across various quantiles of the latent variable. This is clearly the case for the variable CARS and, yet to a lesser extent, for the variable DOVTT. This means that commuters with a high preference to take the car to go to work are more affected by the number of cars they own than others. The same is true for the out of vehicle transportation time. This suggests that improving the density of the public transportation net will affect commuters with a high preference for taking the car much more than commuters who are rather positive toward public transportation. The difference in cost between taking public transport and using a car (DCOST) clearly has a positive effect on the willingness to take the car. This contrasts with the suggestion of Florios and Skouras (2008) that this parameter should be treated as zero. From the mean point estimates it could be concluded that commuters who prefer cars are more price sensitive, but the relatively large credible intervals prompt caution for this kind of interpretation. The effect of the variable DIVTT turns out to be not significantly different from zero for the total quantile process.

5. Discussion

This paper proposes a Bayesian methodology for modeling binary regression quantiles. The general body of literature on binary regression quantiles consists of elaborations of Manski's maximum score estimator. As shown in this study, the main portion of this literature focuses on the difficult optimization of the maximum score estimator and the problems in constructing appropriate confidence intervals for the estimator. These difficulties may account why so few applications are found in the field.

By assuming the asymmetric Laplace density for the underlying latent variable in combination with the data augmentation method, the Bayesian machinery makes it possible to model binary regression quantiles in a straightforward way. Parameter point estimates and credible intervals can easily be extracted from the posterior densities computed. The benefits and possibilities of the proposed approach to binary regression quantiles have been outlined in practice in two simulation studies and one real-life application.

Recently, Kottas and Krnjajic (2009) discussed limitations of quantile regression estimators for continuous dependent variables based on the ALD. They explore generalizations of the asymmetric Laplace density for quantile regression using a Dirichlet process mixture model. Monte Carlo experiments showed that their Bayesian semiparametric method is more robust than quantile regression estimation based on the ALD. A similar criticism can be addressed to the current methodology. Relaxing the assumption of ALD distributed errors by using a Dirichlet process as a prior could be an interesting path for further research. By doing so, the shape of the error density could adapt to the data and thus provide better fit compared to parametric error distributions and reduces the risk of model misspecification. Such an estimator would also resemble the Maximum Score Estimator more and would make comparisons between such an estimator and the MS estimator less awkward. However, Richardson (1999) noted that popular forms of priors are those which have parameters that can be set straightforwardly and which lead to posteriors with a relatively immediate form. In this respect, Bayesian quantile regression based on the ALD is preferable to Bayesian semiparametric quantile regression (Hewson and Yu, 2008; Yu and Stander, 2007).

In conclusion, we showed that the Bayesian ALD-based method for binary quantile regression is a viable strategy when the researcher is explicitly interested in modeling conditional quantiles, when heteroskedasticity is an issue or when only small sample sizes are available. As with quantile regression for a continuous dependent variable, we believe applications in a broad range of research domains, not limited to econometrics, can benefit from the method proposed in this study.

References

Abrevaya J, Huang J. 2005. On the bootstrap of the maximum score estimator. *Econometrica* **73**(4): 1175-1204.

Albert JH, Chib S. 1993. Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association* **88**: 669-679.

Bilias Y, Chen S, Ying Z. 2000. Simple resampling methods for censored regression quantiles. *Journal of Econometrics* **99**(2): 373-386.

Brown RL, Peet RK. 2003. Diversity and invisibility of southern Appalachian plant communities. *Ecology* **84**(1): 32-39.

Buchinski M. 1994. Changes in the U.S. wage structure 1963-1987: Application of quantile regression. *Econometrica* **62**(2): 405-458.

Buchinski M. 1998. The dynamics of changes in the female wage distribution in the USA: a quantile regression approach. *Journal of Applied Econometrics* **13**(1): 1-30.

Buchinski M, Hahn J. 1998. An alternative estimator for the censored quantile regression model. *Econometrica* **66**(3): 653-671.

Delgado MA, Rodriguez-Poo JM, Wolf M. 2001. Subsampling inferences in cube root asymptotics with an application to Manski's maximum score estimator. *Economic Letters* **73**(2): 241-250.

Dunson DB, Taylor JA. 2005. Approximate Bayesian inference for quantiles. *Journal of Nonparametric Statistics* **17**(3): 385-400.

Florios K, Skouras S. 2008. Exact computation of max weighted score estimators. *Journal of Econometrics* **146**(1): 86-91.

Gozalo P, Linton O. 2000. Local nonlinear least squares: Using parametric information in nonparametric regression. *Journal of Econometrics* **99**(1): 63-106.

Groenewald PC, Mokgatlhe L. 2004. Bayesian computation for logistic regression. *Computational Statistics & Data Analysis* **48**(4): 857-868.

Hewson P, Yu K. 2008. Quantile regression for binary performance indicators. *Applied Stochastic Models in Business and Industry* **24**(5): 401-418.

Holmes CC, Held L. 2006. Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Analysis* **1**(1): 145-168.

Horowitz JL. 1992. A smoothed maximum score estimator for the binary response model. *Econometrica* **60**(3): 505-531.

Horowitz JL. 1993. Semiparametric estimation of a work-trip mode choice model. *Journal of Econometrics* **58**(1-2): 49-70.

Horowitz JL. 2004. Semiparametric models. In *Handbook of Computational Statistics*, Gentle JE, Hardle W, Mori Y (eds). Springer: New York.

Huber PJ. 1981. *Robust Statistics*. Wiley: New York.

Kim J, Pollard D. 1990. Cube Root Asymptotics. *Annals of Statistics* **18**(1): 191-219.

Koenker R. 2005. *Quantile Regression*. Cambridge University Press: Cambridge.

Koenker R, Basset G. 1978. Regression Quantiles. *Econometrica* **46**(1): 33-50.

Koenker R, Hallock F. 2001. Quantile Regression. *Journal of Economic Perspectives* **15**(4): 143-156.

Koenker R, Machado JAF. 1999. Goodness of Fit and Related Inference Processes for Quantile Regression. *Journal of the American Statistical Association* **94**(448): 1296-1310.

Kordas G. 2006. Smoothed binary regression quantiles. *Journal of Applied Econometrics* **21**(3): 387-407.

Kotlyarova Y, Zinde-Walsh V. 2006. Non- and semi-parametric estimation in models with unknown smoothness. *Economics Letters* **93**(3): 379-386.

Kottas A, Gelfand AE. 2001. Bayesian semiparametric median regression modelling. *Journal of the American Statistical Association* **97**(460): 1020-1033.

Kottas A, Krnjajic M. 2009. Bayesian Semiparametric Modelling in Quantile Regression. *Scandinavian Journal of Statistics* **36**(2): 297-319.

Kotz S, Kozubowski TJ, Podgórski K. 2001. *The Laplace distribution and generalizations: A Revisit with Applications to Comunications, Economics, Engineering and Finance*. Birkhäuser: Boston.

Li YJ, Zhu J. 2007. Analysis of array CGH data for cancer studies using fused quantile regression. *Bioinformatics* **22**(18): 2470-2476.

Machado J, Santos Silva J. 2005. Quantiles for counts. *Journal of the American Statistical Association* **100**(Dec): 1226-1237.

Manski CF. 1975. Maximum score estimation of the stochastic utility model of choice. *Journal of Econometrics* **3**(3): 205-228.

Manski CF. 1985. Semiparametric analysis of discrete response: asymptotic properties of the maximum score estimator. *Journal of Econometrics* **27**(3): 313-333.

Manski CF, Thompson TS. 1986. Operational Characteristics of Maximum Score Estimation. *Journal of Econometrics* **32**(1): 85-108.

McDonald JB. 1996. An application and comparison of some flexible parametric and semi-parametric qualitative response models. *Economics Letters* **53**(2): 145-152.

Powell J. 1986. Censored regression quantiles. *Journal of Econometrics* **32**(1): 134-155.

Richardson S. 1999. Contribution to the discussion of Walker et al., Bayesian nonparametric inference for random distribution and related functions. *Journal of the Royal Statistical Society Series B – Statistical Methodology* **61**(3): 485-527.

Skouras S. 2003. An algorithm for computing estimators that optimize step functions. *Computational Statistics & Data Analysis* **42**(3): 349-361.

Tanner MA, Wong WH. 1987. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* **82**: 528-540.

Tsionas EG. 2003. Bayesian quantile inference. *Journal of Statistical Computation and Simulation* **73**(9): 659-674.

Whittaker J, Whitehead C, Somers M. 2005. The neglog transformation and quantile regression for the analysis of a large credit scoring database. *Journal of the Royal Statistical Society Series C - Applied Statistics* **54**(5): 863-878.

Yu K, Lu Z, Stander J. 2003. Quantile regression: applications and current research areas. *Journal of the Royal Statistical Society Series D – The Statistician* **52**(3): 331-350.

Yu K, Moyeed RA. 2001. Bayesian Quantile Regression. *Statistics and Probability Letters* **54**(4): 437-447.

Yu K, Stander J. 2007. Bayesian analysis of a tobit quantile regression model. *Journal of Econometrics* **137**(1): 260-276.

Yu K, Zhang J. 2005. A three-parameter asymmetric Laplace distribution and its extention. *Cummunications in Statistics: Theory & Methods* **34**(9-10): 1636-1640.

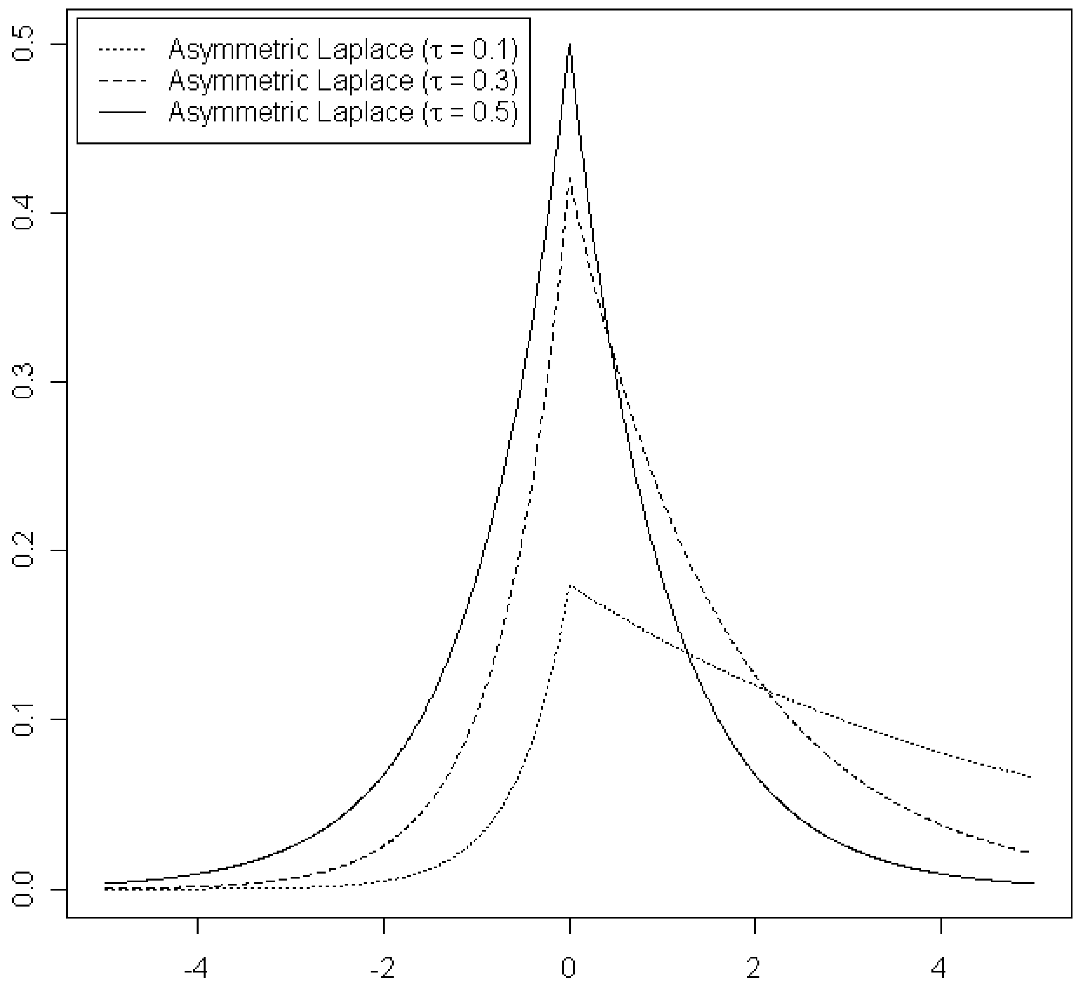

Figure 1: Standard Asymmetric Laplace Density (ALD)

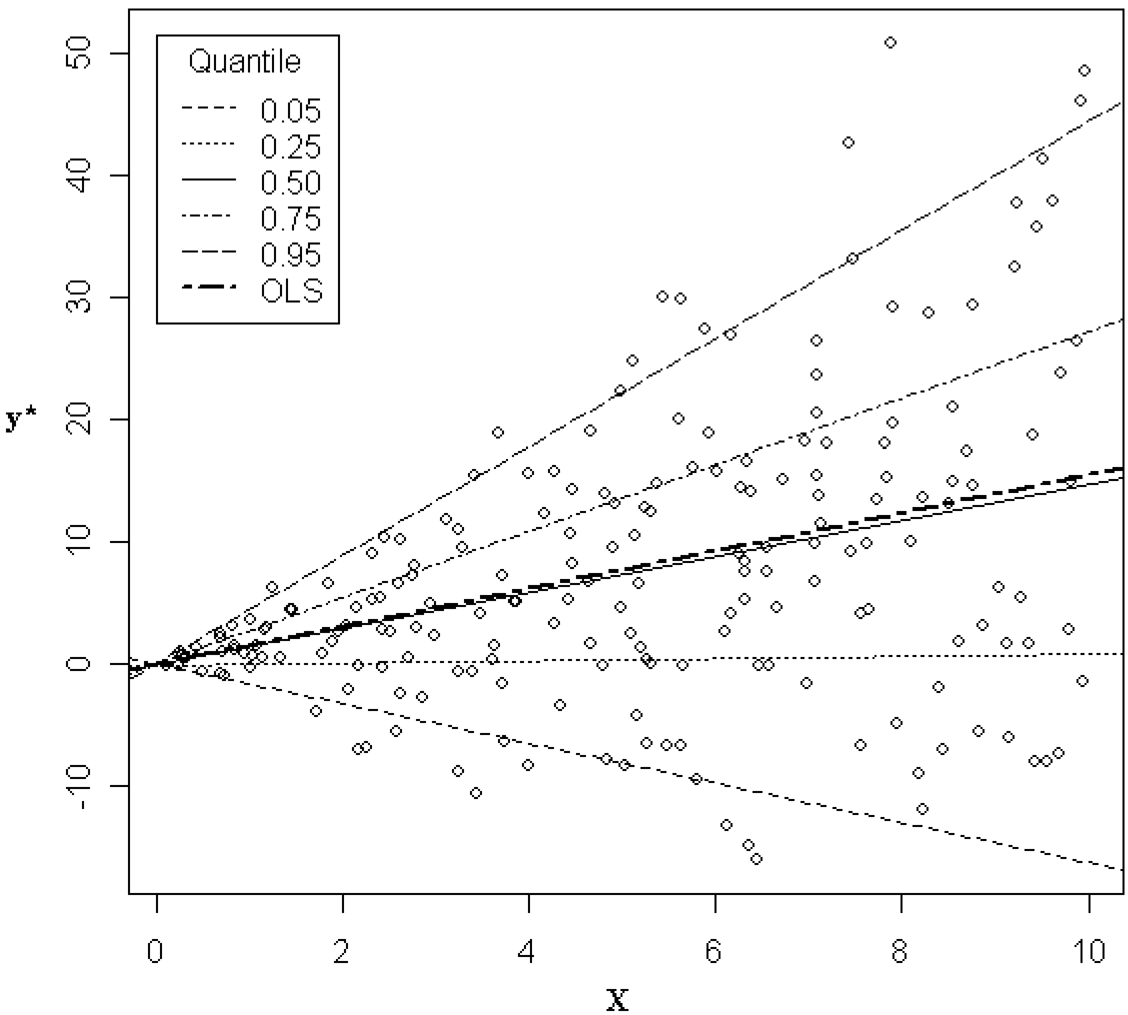

Figure 2: Quantile Regression

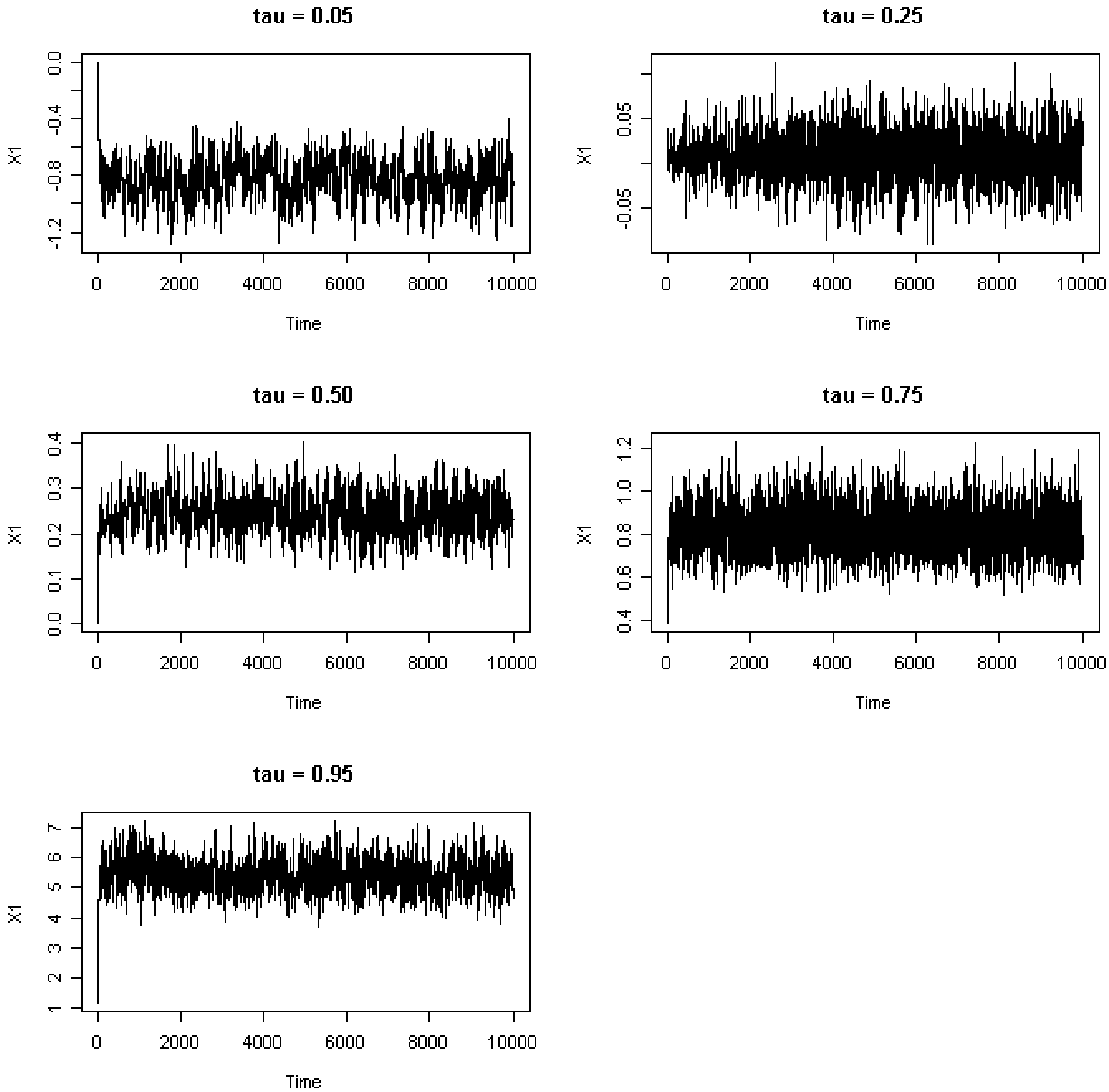

Figure 3: Time-series plots for the binary quantile regression parameter estimates.

Tau	Beta	95% Credible Interval	
0.05	-0.8391	-1.1527	-0.5553
0.25	0.0084	-0.0445	0.0616
0.50	0.2456	0.1621	0.3395
0.75	0.8103	0.6338	1.0021
0.95	5.4273	4.5450	6.2357
Model	**Beta**	**95% Confidence Interval**	
logit	0.1789	0.1223	0.2404

Table 1: Binary regression: quantile regression estimates and logit estimates

N		β_0			β_1		
		BRQ	sBRQ	Bayes(ALD)	BRQ	sBRQ	Bayes(ALD)
200	Bias	-1.32	-1.46	-1.09	0.17	0.19	0.09
	RMSE	1.81	1.88	1.22	0.83	0.83	0.31
	2.5%	-3.98	-3.24	-2.96	-0.07	0.19	0.21
	97.5%	-1.45	-1.69	-1.58	1.87	1.19	1.09
400	Bias	-1.18	-1.28	-0.98	0.10	0.09	0.04
	RMSE	1.45	1.46	1.01	0.54	0.46	0.16
	2.5%	-3.79	-2.91	-2.46	0.01	0.17	0.29
	97.5%	-1.53	-1.66	-1.65	1.62	1.01	0.84
600	Bias	-1.12	-1.17	-0.95	0.08	0.05	0.03
	RMSE	1.34	1.28	0.97	0.45	0.34	0.12
	2.5%	-3.45	-2.68	-2.31	0.08	0.19	0.33
	97.5%	-1.57	-1.67	-1.69	1.42	0.90	0.76

Table 2: Bias, root mean square errors (RMSE) and 95% credible intervals for the parameters β_0 and β_1 (with $\tau = 0.5$).

	INTERCEPT	CARS	DOVTT	DIVTT	DCOST	Method
Horowitz (1993)	-0.276	0.052	0.011	0.005	1	MSCORE
Florios and Skouras (2008)	5.122	3.916	0.962	0.401	1	MIP
Current study	4.825	3.375	1.018	0.282	1	Bayes(ALD)
95% credible interval (lower)	*3.331*	*2.287*	*0.328*	*-0.230*	-	*Bayes(ALD)*
95% credible interval (upper)	*7.621*	*5.378*	*2.183*	*0.847*	-	*Bayes(ALD)*

Table 3: Estimates for the work-trip mode choice model (standardized)

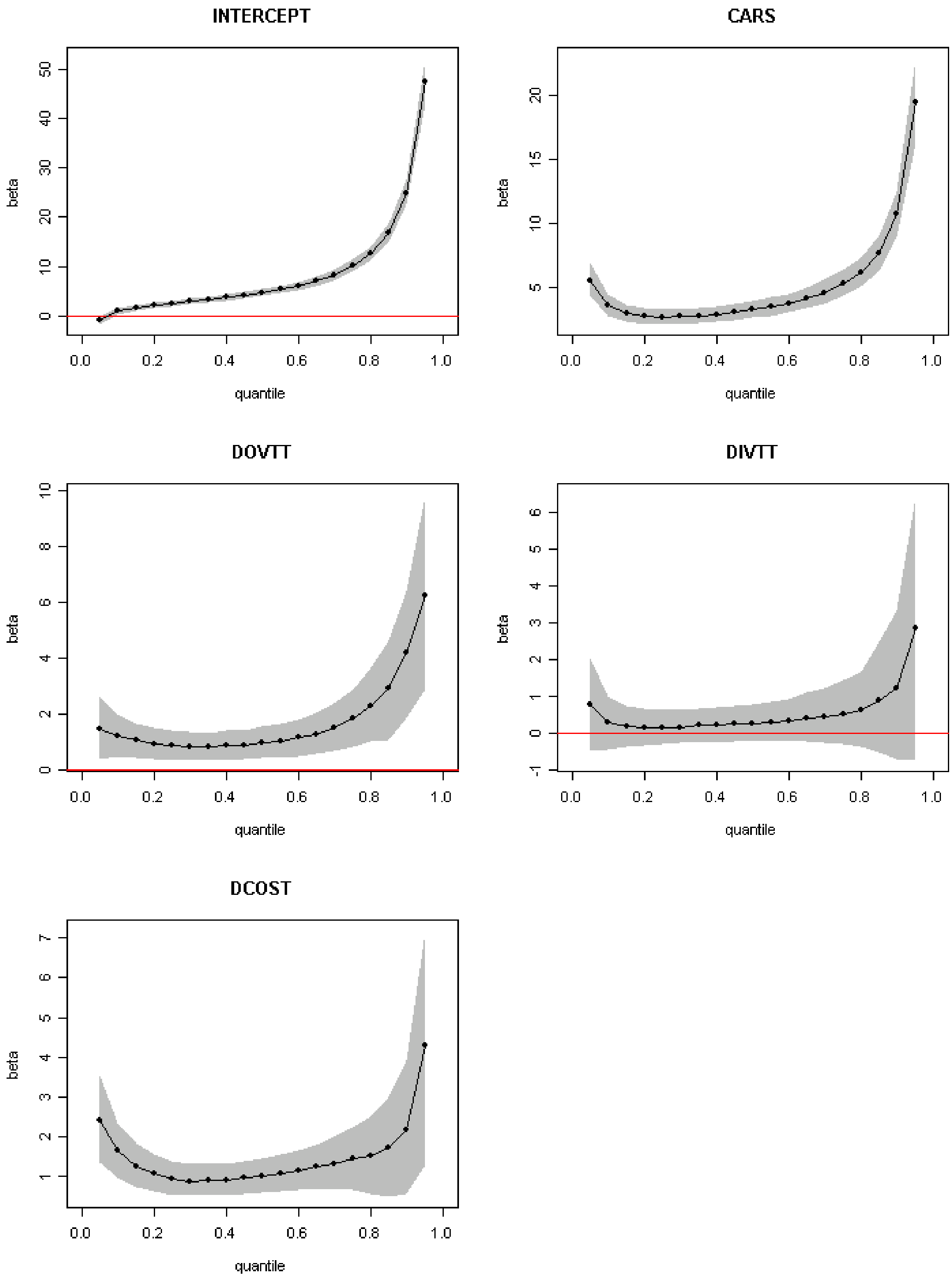

Figure 4: Quantile process