Binary Quantile Regression and Variable Selection: A new Approach

Katerina Aristodemou

Brunel University, UK

Keming Yu

Brunel University, UK

Abstract.

In this paper we propose a new estimation method for binary quantile regression and variable selection which can be implemented by an iteratively re-weighted least squares approach. In contrast to existing estimators, this method is computationally simple, guaranteed to converge to a unique solution and implemented with standard software packages. We demonstrate our methods using Monte-Carlo experiments and then apply the method to the widely used work-trip mode choice data analysis. The results indicate that the proposed estimators work well in finite samples.

Keywords: Adaptive lasso, binary regression, iteratively re-weighted least squares, quantile regression, smoothed maximum score estimator, work-trip mode choice, variable selection,

1 Introduction

Applications of regression models for binary response are very common and models such as logistic regression and probit regression, are widely used in many fields. However, these conventional binary regression models, focus on the estimation of the conditional mean function, which is not always the prime interest for a researcher. Also, they assume that the errors are independent of the regressors, which is rarely the case in practice. Quantile regression (Koenker (2005)) extends the mean regression model to conditional quantiles of the response variable and can provide estimation for a family of quantile functions that describe the entire underlining distribution of the response variable. Furthermore, quantile regression parameter estimates are not biased by a location-scale shift of the conditional distribution of the dependent variable. Quantile regression has been used by many researchers in different fields and has also been extended to the analysis of censored data, count data and proportions.

The potential benefits of binary quantile regression have been recognised by several authors (e.g. Manski (1975), Horowitz (1992), Kordas (2006) and Benoit and Van den Poel (2012)) who developed different estimation techniques for the binary quantile regression model.

The general binary regression model is defined as:

$$\begin{aligned} y^* &= \boldsymbol{x}'\boldsymbol{\beta} + \epsilon_i, \\ y &= I\{y^* \geq 0\}, \end{aligned} \tag{1}$$

where, y_i^* is a continuous, scalar latent variable, y is the observed binary outcome of this latent variable, $I(\cdot)$ is the indicator function, $\boldsymbol{x}$ is a $p \times 1$ vector of explanatory variables, $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters and ϵ is a scalar random error term. If the distribution of ϵ conditional on x is known up to a finite set of parameters, $\boldsymbol{\beta}$ can be estimated by different techniques, including maximum likelihood. If it is assumed that ϵ has a Normal distribution then the binary probit model arises, whereas, if a logistic distribution is assumed then the model (1) becomes the binary logit model.

Specifying the distribution of ϵ a priori, will yield inconsistent estimators if the distribution of ϵ is misspecified. A more flexible model is obtained by imposing only one assumption on ϵ, the quantile restriction $Q_\tau(\epsilon_i|x_i) = 0$.

Let $Q_\tau\left(y^*|\boldsymbol{x}\right)$ denote the conditional quantile of the latent variable y^* given $\boldsymbol{x}$, defined as:

$$Q_\tau\left(y^*|\boldsymbol{x}\right) \equiv F_{y^*}^{-1}(\tau|\boldsymbol{x}) \equiv \boldsymbol{x}'\boldsymbol{\beta}(\tau),$$

where $F(\cdot)$ is the distribution function of the latent variable y^* and $\tau \in [0,1]$.

By the equivalence property to monotone transformations of the conditional quantile function (Powell (1986)), the τ^{th} conditional quantile function of the observed variable y_i in the model (1) can be expressed as:

$$Q_\tau\left(y|x\right) = I\{\boldsymbol{x}'\boldsymbol{\beta}(\tau) \geq 0\}. \tag{2}$$

Binary quantile regression was first introduced by Manski (1975, 1985). In these papers he introduced the Maximum Score Estimator (MSE), which requires very weak assumptions on the relation of errors to regression variables and can accommodate for heteroscedasticity of unknown form. Estimates of the regression parameters in model (1) can be obtained by:

$$\widehat{\boldsymbol{\beta}}(\tau) = \arg \max_{\{\boldsymbol{\beta}:||\boldsymbol{\beta}||=1\}} \sum_{i=1}^{n} [y_i - (1-\tau)] I\{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau) \geq 0\}, \tag{3}$$

where, $(x_i, y_i, i = 1, ..., n)$ is a random sample of observation and $0 < \tau < 1$ is the τ^{th} regression quantile. Identification of $\boldsymbol{\beta}$ is only possible up to a scale, thus to make estimation possible a scale normalisation is necessary. Manski (1975, 1985) used the normalisation $||\boldsymbol{\beta}|| = 1$, where $||\cdot||$ denotes the Euclidean norm.

Manski (1985) provided the conditions under which the maximum score and binary quantile regression estimators are consistent. However, this work faces important technical drawbacks in both optimising the objective function and inferring the regression parameters. The rate of convergence of $\widehat{\boldsymbol{\beta}}(\tau)$ and its asymptotic distribution were derived by Cavanagh (1987). Kim and Pollard (1990) showed that it is not asymptotically normal, but the estimator converges in distribution to the maximum of a complicated multidimensional stochastic process. Furthermore, the model is nonlinear in parameters thus its estimation is computationally more demanding than conventional linear quantile regression models. Delgado et al. (2001) attempted to solve the problem by using sub-sampling methods to form confidence intervals. They provided simulation evidence that suggests inconsistency of the bootstrap, a result that was later proved by Abrevaya and Huang (2005).

The maximum score estimator has a slow rate of convergence and a complicated asymptotic distribution because it is obtained by maximising a step function. To remedy some of these shortcomings Horowitz (1992) developed a smoothed maximum score estimator (SMSE) under a linear median regression specification for the latent variable in the binary model, which can be computed using standard optimisation routines. Kordas (2006) extended this estimator to a family of conditional quantile functions giving the opportunity for a complete understanding of the conditional distribution of the latent response variable given covariates:

$$\widehat{\boldsymbol{\beta}}_{smse}(\tau) = \arg \max_{\{\boldsymbol{\beta}:|\beta_1|=1\}} \sum_{i=1}^{n} [y_i - (1-\tau)] K\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau)}{h_n}\right) \tag{4}$$

where K is a smooth continuous function and h_n is a sequence of real positive constants converging to zero as the sample size increases. Identification of $\boldsymbol{\beta}$ up to scale requires that $\boldsymbol{x}$ has at least one component whose probability distribution conditional on the remaining components is absolutely continuous with respect to the Lebesgue measure (Manski (1985)). To make estimation possible Horowitz (1992) imposes the normalisation, $|\beta_1| = 1$. This requires to arrange the components of $\boldsymbol{x}$ appropriately, so that x_1, satisfies this condition and accordingly, to re-arrange the components of $\boldsymbol{\beta}$ so that β_1 is the coefficient corresponding to x_1. Kordas (2006) discusses two possible normalisation methods $||\boldsymbol{\beta}|| = 1$ or $|\beta_p| = 1$. In this work the latter normalisation method was chosen.

Horowitz's approach is computationally simpler than the maximum score estimator. Also, under stronger conditions than in Manski (1975, 1985), Horowitz's estimator converges at a faster rate and is asymptotically normally distributed.

Benoit and Van den Poel (2012) provided numerical evidence for the usefulness of Bayesian quantile regression for binary response models based on the Asymmetric Laplace distribution.

Although both the maximum score and smoothed maximum score estimators have desirable asymptotic properties, they are difficult to implement in practice, and most importantly, they do not necessarily guarantee convergence and a unique solution. Specifically, the objective function in the maximum score estimator is discontinuous (step-function) therefore it cannot be solved using a gradient-based optimisation method, whereas, the objective function of the smoothed maximum score estimator can have several local maxima, therefore stochastic search algorithms are necessary to identify the global maximum (e.g. the simulated annealing algorithm suggested by Horowitz (1992)). Even though algorithms for solving both the MSE and the SMSE are readily available these are not included in standard software packages. Furthermore, the non-standard structure of their objective functions cannot always guarantee global convergence. These practical limitations motivate the development of the estimator described in this chapter. An alternative estimation approach is proposed, based on a nonlinear asymmetrical weighted loss function, which can be implemented by an iteratively reweighted least square algorithm (IRLS). The IRLS algorithm is computationally simple and guarantees convergence to a unique solution (Kokic et al. (1997)).

The remainder of the paper is organised as follows. Section 2 introduces the Binary quantile regression, provides the asymptotic properties of the estimator and describes the proposed estimation approach and the corresponding algorithm for binary quantile regression. Section 3 introduces the method of variable selection via the modern adaptive lasso technique and describes how this method can be implemented in the framework of the binary quantile regression. An estimation approach

5

and the algorithm for variable selection using a penalised binary quantile regression objective function are provided. Section 4 illustrates the proposed methods through a Monte Carlo study and a real example. Concluding remarks are provided in Section 5. Technical proofs can be found in the Appendix.

2 Binary quantile regression

The estimator in equation (3) can be viewed as a $\tau - quantile$ version of the general linear binary quantile regression problem (Koenker and Bassett (1978)), which is obtained by solving:

$$\widehat{\boldsymbol{\beta}}(\tau) = \arg \min_{\{\boldsymbol{\beta}:|\beta_1|=1\}} \mathcal{R}_u(x) \tag{5}$$

where,

$$\mathcal{R}_u(x) = \sum_{i=1}^{n} w_i(\tau)|y_i - I\{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau) \geq 0\}|$$

and

$$w_i(\tau) = \begin{cases} \tau & \text{if } y_i - I\{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau) \geq 0\} \geq 0; \\ (1-\tau) & \text{if } y_i - I\{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau) \geq 0\} < 0. \end{cases}$$

A smoothed version of the model (5) can be contracted by replacing the indicator function with a smooth cumulative distribution function (cdf), $K(\cdot)$ (Horowitz (1992)), such as:

$$\widehat{\boldsymbol{\beta}}_{smse}(\tau) = \arg \min_{\{\boldsymbol{\beta}:|\beta_1|=1\}} \sum_{i=1}^{n} w_i(\tau) \left| y_i - K\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau)}{h_n}\right)\right| \tag{6}$$

where,

$$w_i(\tau) = \begin{cases} \tau & \text{if } y_i - K\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau)}{h_n}\right) \geq 0; \\ (1-\tau) & \text{if } y_i - K\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau)}{h_n}\right) < 0. \end{cases}$$

6

and $K(\cdot)$ satisfies the following properties,

$$\begin{aligned} &K1 : |K(v) < M| \text{ for some finite M and v } \in (-\infty, \infty) \\ &K2 : \lim_{v\to-\infty} K(v) = 0 \text{ and } \lim_{v\to\infty} K(v) = 1. \end{aligned} \tag{7}$$

2.1 Estimation of the Smoothed Binary Quantile Regression Model

In this sub-section an alternative estimation approach for estimating binary quantile regression models is developed, which is simple, is guaranteed to converge to a unique solution and can be implemented with standard software packages.

In a recent paper, Blevins and Khan (2013) demonstrated that for binary data the maximum score objective function in equation (5) is equivalent to the quadratic loss objective function under the median restriction, i.e for $\mathbf{w} = 0.5$. Since quantile regression can be viewed as a generalisation of median regression, in this chapter this work is extended to the estimation of binary regression quantiles using a non-linear least asymmetric weighted squares (LAWS) approach. For any given quantile the estimator in model (5) is mathematically equivalent to the nonlinear LAWS estimator. Hence, the binary quantile regression objective function in equation (5), under Kordas (2006) normalisation can be written as:

$$\widehat{\boldsymbol{\beta}}_{laws}(\tau) = \arg \min_{\{\boldsymbol{\beta}:|\beta_p|=1\}} \sum_{i=1}^{n} w_i(\tau) \left(y_i - I\{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau) \geq 0\}\right)^2 \tag{8}$$

where, $\widehat{\boldsymbol{\beta}}_{laws}(\tau) = (\widehat{\beta}', 1)'$ and

$$w_i(\tau) = \frac{\mathcal{R}_u\left(y_i - I\{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau) \geq 0\}\right)}{\left(y_i - I\{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau) \geq 0\}\right)^2} \tag{9}$$

In the case of binary data it can be shown that equation (9) is equal to

$$w_i(\tau) = \begin{cases} \tau & \text{if } y_i - I\{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau) \geq 0\} \geq 0; \\ (1-\tau) & \text{if } y_i - I\{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau) \geq 0\} < 0. \end{cases} \tag{10}$$

The concept of LAWS was first introduced by Newey and Powell (1987), who used the so-called regression expectiles to investigate the underlying conditional distribution. Recently LAWS re-gained interest in the context of semiparametric or geoadditive regression (see for example Schnabel and Eilers (2009) and Sobotka and Kneib (2012)). Breckling and Chambers (1988) proposed a M-quantile regression based on an asymmetric loss function and Jones (1994) showed that expectiles are quantiles of a transformation of the original distribution. Nonparametric estimation of regression expectiles was considered by Yao and Tong (1996) who used a kernel method based on a locally linear fit. Compared to quantile regression, the LAWS is reasonably efficient under normality conditions (Efron (1991)). Confidence intervals for expectiles based on an asymptotic Normal distribution were introduced by Sobotka et al. (2013).

2.2 Estimation Algorithm

The algorithm to estimate the model (8) is a nonlinear weighted least squares algorithm. However, since the weights are determined by the residuals that vary from iteration to iteration, a nonlinear IRLS approach is implemented.

To enable estimation, following Horowitz (1992), the standard Normal distribution, with cdf $\Phi(\cdot)$ is taken as the Kernel density and a customary normalisation $\boldsymbol{\beta}_n = 1$ is imposed. Then, the nonlinear binary regression estimator is obtained by minimising the nonlinear smoothed LAWS function (slaws):

$$\widehat{\boldsymbol{\beta}}_{slaws}(\tau) = \arg \min_{\{\boldsymbol{\beta}:|\beta_p|=1\}} \sum_{i=1}^{n} w_i(\tau) \left(y_i - \Phi\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau)}{h_n} \right) \right)^2 \tag{11}$$

where, $\widehat{\boldsymbol{\beta}}_{slaws}(\tau) = (\widehat{\beta}', 1)'$ and

$$w_i(\tau) = \begin{cases} \tau & \text{if } y_i - \Phi\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau)}{h_n}\right) \geq 0; \\ (1-\tau) & \text{if } y_i - \Phi\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau)}{h_n}\right) < 0. \end{cases} \tag{12}$$

The steps of the algorithm for fitting the binary quantile regression model are described in Algorithm 1. These steps can be easily implemented using standard software packages such as R or Stata.

8

Algorithm 1 Binary quantile regression via nonlinear LAWS

1: Obtain an initial estimate of β by running standard nonlinear OLS regression.
2: Obtain an initial estimate of the residuals $\epsilon_i^0 = y_i - \Phi\left(\frac{\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}(\tau)}{h_n}\right)$.
3: Construct the weights, $w_i^0(\tau)$ using equation (12) and estimate equation (11) via nonlinear WLS regression.
4: Obtain new estimates of the residuals, $\epsilon_i^1 = y_i - \Phi\left(\frac{\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_{slaws}(\tau)}{h_n}\right)$.
5: Update the weights to obtain $w_i^1(\tau)$ using equation (12).
6: Estimate equation (11) by nonlinear WLS regression.
7: Repeat steps 4 to 6 until convergence.

2.3 Asymptotic Properties

Regarding the asymptotic properties of the estimator, it can be shown that, under the following assumptions, Theorem 1 can be established.

Assumption 1. The vectors (x_i', ϵ_i') are identically and independently distributed random variables.

Assumption 2. $F_{\epsilon_i}(\cdot)$ is a distribution function with $F(0) = \tau$ and $Q_\tau(\epsilon_i|x_i) = 0$ for $\tau \in (0, 1)$.

Assumption 3. $\beta_n \in \mathbb{B}$, the closure of an open convex set of $\Re^{p-1}$.

Assumption 4. The support of x_i is not contained in any proper linear subspace of $\Re^p$.

Assumption 5. The density function, $f_{\epsilon_i|x_i}(\cdot)$ is positive in a neighborhood of 0.

Assumption 6. The weights $w_i(\tau)$ are independent of the regression parameters.

Assumption 7. The n vectors $x_j, j = 1...p-1$ are independently distributed with the first component of $x_{i1} \equiv 1$ for all i almost surely.

Assumption 8. $0 < P(y_i = 1|x_i) < 1$ for almost every x_i.

Theorem 1. *(proof is provided in Appendix)*
If $h_n \to 0$, *then* $\widehat{\beta}(\tau) - \beta_0(\tau) \xrightarrow{p} 0$.

Furthermore, under regularity conditions identical to the ones in Horowitz (1992), the estimator enjoys asymptotic properties similar to those of the maximum score estimator Manski (1975, 1985). In particular, the rate of convergence can be as fast as the $O(n^{-1/3})$ and it has a non-Gaussian limiting distribution.

The slower rate of convergence relative to the smoothed maximum score estimator in Horowitz (1992) is due to a bias condition, where the bias of the estimator converges at the rate of h_n. This is in contrast to the rate of h_n^2 for the smoothed maximum score estimator. However, according to Blevins and Khan (2013) this bias condition can be easily corrected, e.g. by using a different kernel function to the Normal cdf, or via other bias-reducing mechanisms, such as jackknifing.

3 Variable Selection via Penalised Binary Quantile Regression

Variable selection plays an important role in the model-building process. A common problem when constructing a predictive model is the large number of candidate predictor variables. Identifying the smallest set of relevant variables has many advantages: (i) the process is cost-effective, usually simpler, and potentially faster, (ii) it improves the prediction performance of the predictors (iii) knowledge about the relevant variables can enhance the understanding of the underlying problem. Furthermore, multicollinearity and overfitting are areas of concern when a large number of independent variables are incorporated in a regression model.

The problem of overfitting also arises in quantile regression models. First, Koenker (2004) developed a L1-regularisation quantile regression method to shrink individual effects in longitudinal data towards a common value and Li and Zhu (2008) considered the L1-norm (LASSO) regularised quantile regression. The lasso is a regularised technique for simultaneous estimation and variable selection (Sobotka et al. (2013)). Even though the lasso is generally able to provide consistent variable selection and optimal prediction, scenarios exist in which the lasso selection cannot be consistent.

To solve this problem Zou (2006) developed a new version of the lasso, the adaptive lasso. This is a weighted L1 penalty which allows different penalisation parameters for different regression coefficients. The weights are determined by an initial estimator, $\widehat{\boldsymbol{\beta}}(\tau)$, e.g. the classical quantile regression estimator, and are used to construct weights based on the importance of each predictor. The most important advantage of the adaptive lasso is its oracle property, which estimators based on the classical lasso do not enjoy. The oracle property requires that as the sample size increases the coefficient of non-relevant terms approaches zero and the probability of selecting the correct model goes to 1. Also, it requires that consistent model selection does not come at the expense of efficiency: the asymptotic distribution of the non-zero components of $\widehat{\boldsymbol{\beta}}$ must be the same as the "oracle model", when $\boldsymbol{y}$ is regressed only on the relevant variables. Wu and Liu (2009) considered variable selection through penalised quantile regression with adaptive lasso penalties in the framework of a linear model.

It should be noted that in Bayesian terms, the lasso procedure can be interpreted as a posterior mode estimate under independent Laplace priors for the regression coefficients (Tibshirani (1996), Park and Casella (2008)). Based on this principle Li et al. (2010) proposed a Bayesian regularized quantile regression model by assuming that the model residuals come from the skewed Laplace distribution. The Laplace distribution has the attractive property that it can be represented as a scale mixture of normals with an exponential mixing density which leads to the development of a hierarchical Bayesian interpretation of the Lasso, which can be easily estimate by a Gibbs sampling algorithm. Benoit et al. (2013) extended this work to bayesian lasso binary quantile regression.

In this section the modern adaptive lasso variable selection technique is extended to Binary quantile regression, in the framework of the nonlinear LAWS approach. Suppose that $\widehat{\boldsymbol{\beta}}(\tau)$ is a consistent estimator of $\boldsymbol{\beta}(\tau)$, the binary quantile regression estimator in equation (5). Then the $\tau-quantile$ version of the adaptive lasso binary quantile regression estimator, $\widehat{\boldsymbol{\beta}}^*$, is given by:

$$\widehat{\boldsymbol{\beta}}^*(\tau) = \arg \min_{\{\boldsymbol{\beta}:|\beta_1|=1\}} \sum_{i=1}^{n} w_i(\tau) \left|y_i - I\{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau) \geq 0\}\right| + \lambda_n \sum_{j=1}^{p} w_j^{lasso} \left|\boldsymbol{\beta}_j\right| \tag{13}$$

where, $w_i(\tau)$ is defined in equation (10), $\mathbf{w}^{lasso} = \frac{1}{|\widehat{\boldsymbol{\beta}}(\tau)|}$ is a known weights vector (Zou (2006)) and λ is a nonnegative regularisation parameter which controls the level of penalisation, with greater values implying more aggressive model selection. The second term in equation (13) is the adaptive lasso binary quantile regression penalty function, that is crucial for the success of the lasso.

3.1 Estimation Algorithm

In this sub-section the estimation approach to obtain the penalised binary quantile regression estimator in equation (13) is presented. The approach is simple and has the advantage of being implementable in standard software packages such as R or Stata.

Like the estimator for non-penalised binary quantile regression, developed in section 2, the estimator of the adaptive lasso binary quantile regression in equation (13) is mathematically equivalent to the penalised nonlinear LAWS estimator given:

$$\widehat{\boldsymbol{\beta}}^*_{adapt.lasso_{laws}}(\tau) = \arg \min_{\{\boldsymbol{\beta}:|\beta_p|=1\}} \sum_{i=1}^{n} w_i(\tau) \left(y_i - I\{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau) \geq 0\}\right)^2 + \lambda_n \sum_{j=1}^{p} w_j^{lasso} \left|\boldsymbol{\beta}_j\right| \tag{14}$$

where, $\widehat{\boldsymbol{\beta}}_{laws}(\tau)$ is a consistent estimator of $\boldsymbol{\beta}(\tau)$ in equation (8), $w_i(\tau)$ is defined as before, $\mathbf{w}^{lasso} = \frac{1}{|\widehat{\boldsymbol{\beta}}_{laws}(\tau)|}$ and λ is a nonnegative regularisation parameter.

Again, as in the non-penalised binary quantile regression estimator, to enable estimation the Indicator function is replaced by the standard Normal kernel density, $\Phi(\cdot)$. Then, the nonlinear adaptive lasso smoothed binary quantile regression estimator is defined as:

12

$$\widehat{\boldsymbol{\beta}}^{*}_{adapt.lasso_{slaws}}(\tau) = \arg \min_{\{\boldsymbol{\beta}:\|\beta_p\|=1\}} \sum_{i=1}^{n} w_i(\tau) \left(y_i - \Phi\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}(\tau)}{h_n}\right)\right)^2 + \lambda_n \sum_{j=1}^{p} w_j^{lasso} \left|\boldsymbol{\beta}_j\right| \tag{15}$$

where, $w_i(\tau)$ is defined in equation (12), $\widehat{\boldsymbol{\beta}}_{slaws}(\tau)$, is a consistent estimator of the binary quantile regression estimator in equation (11), $\mathbf{w}^{lasso} = \frac{1}{|\widehat{\boldsymbol{\beta}}_{slaws}(\tau)|}$ and λ is a nonnegative regularisation parameter.

The estimator can be obtained by an iteratively re-weighted least square algorithm (IRLS). The steps of the algorithm for fitting the adaptive lasso binary quantile regression model are described in Algorithm 2.

Algorithm 2 Variable Selection via Penalised Binary quantile regression

1: Obtain an initial estimate for non-penalised binary quantile regression, $\widehat{\boldsymbol{\beta}}_{slaws}(\tau)$, via Algorithm 1.

2: Calculate $\mathbf{w}^{lasso} = \frac{1}{|\widehat{\boldsymbol{\beta}}_{slaws}(\tau)|}$.

3: Use the initial estimates $\widehat{\boldsymbol{\beta}}_{slaws}(\tau)$ to obtain an initial estimate of the residuals $\epsilon_i^0 = y_i - \Phi\left(\frac{\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_{slaws}(\tau)}{h_n}\right)$.

4: Construct the initial weights, $w_i^0(\tau)$ using equation (12).

5: Use $\mathbf{w}^{lasso}$ and $w_i^0(\tau)$ to optimise the objective function in equation (15) via direct numerical optimisation.

6: Obtain new estimates of the residuals, $\epsilon_i^1 = y_i - \Phi\left(\frac{\boldsymbol{x}_i'\widehat{\boldsymbol{\beta}}_{salaslaws}(\tau)}{h_n}\right)$.

7: Update the weights to obtain $w_i^1(\tau)$ using equation (12).

8: Re-estimate equation (15) via direct numerical optimisation.

9: Repeat steps 6 to 8 until convergence.

Choice of λ

The selection of the tuning parameters λ should be based on a data-driven approach to allow for increasing flexibility with the sample size. The most common way for its selection is the method of K-fold cross-validation. This is a measure of the out-of-sample estimation error under different configurations for tuning parameters, without collecting additional data.

13

The first step of the approach involves selecting a grid of candidate values for λ and dividing the data into K roughly equal folds. For each candidate value of λ the model is fitted K-1 times, each time leaving out one of the folds and the model prediction error of computed using the Kth fold by:

$$E_k(\lambda) = \sum_{i \in K^{th} fold} \left(y_i - \widehat{y}_{(-i)}(\lambda)\right)^2, \tag{16}$$

where, $\widehat{y}_{(-i)}(\lambda)$ is the fitted value from the model that excludes the fold containing i.

This gives the cross-validation error

$$CV(\lambda) = \frac{1}{K} \sum_{k=1}^{K} E_k(\lambda) \tag{17}$$

The selected tuning parameter is the one that minimises the cross-validation error.

3.2 Oracle properties

In this section we show that with the proper choice of $\lambda \equiv \lambda_n$ above, the adaptive lasso in (15) enjoys the oracle properties under the following technical conditions:
(i) Error assumption (cf Pollard (1991)): The regression errors $\{\epsilon_i\}$ in equation (1) are independent and identically distributed, with τth quantile zero and a continuous, positive density $f(.)$ in a neighborhood of zero.
(ii) Let $\phi'(.)$ be the first derivative of the standard normal density or the second derivative of standard normality cumulative function $\Phi(.)$. Let h_n be the bandwidth which exists a constant $C > 0$ and $\nu > 0$, $h_n = Cn^{-1/(2\nu+1)}$. The design $\boldsymbol{x}_i$, $i = 1, ..., n$ satisfy $\lim_{n\to\infty}(\sum_{i=1}^{n} \phi'(\boldsymbol{x}_i)'\boldsymbol{x}_i\boldsymbol{x}_i')\phi'(\boldsymbol{x}_i)/n = \Sigma$, where Σ exists and is a positive definite matrix. Denote the top-left q-by-q submatrix of Σ by Σ_A and the right-bottom (p - q)-by- (p - q) submatrix of Σ by Σ_{A^c}.

Theorem 2. *(proof is provided in Appendix)*
Let $\mathcal{A} = \{j : \beta_j \neq 0\}$ and assume that $|\mathcal{A}| = q < p$, then the true regression model

depends only on a subset of $\boldsymbol{x}$*. Suppose that* $\lambda_n = o(\sqrt{n})$ *and* $\lambda_n n^{(\nu-1)/2} \to \infty$*, then*

(i) $\widehat{\boldsymbol{\beta}}(adapt_{lasso})$ *can identify the right subset model* $\mathcal{A}$.

(ii) $\widehat{\boldsymbol{\beta}}(adapt_{lasso})$ *has the optimal estimation rate,*

$$\sqrt{n}\left(\widehat{\boldsymbol{\beta}}(adapt_{lasso}) - \boldsymbol{\beta}(adapt_{lasso})\right) \to N(0,\ \tau(1-\tau)\Sigma_A^{-1}/f(0)^2).$$

4 Numerical Experiments

In this section the proposed approach for binary quantile regression and variable selection is demonstrated through two simulated and one real examples. The first simulation example is carried out to examine the performance of the proposed binary quantile regression estimator, using a nonlinear least asymmetric weighted squares (LAWS) approach. The second simulation example demonstrates the proposed approach for variable selection in binary quantile regression models. The real example is based on the widely studied transport-choice dataset described in Horowitz (1993). All programs were written and executed in the free statistical package R.

4.1 Simulation Example 1 - Binary Quantile Regression

In the first simulation experiment the following model was considered for simulating data:

$$y_i^* = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \epsilon_i, \tag{18}$$

where $x_{pi} \sim N(0,1)$, $i = 1, ..., n$ and $n = 500$ and $\boldsymbol{\beta} = (-0.1, -1, 1)$.

For the model error ϵ_i the following three specifications were considered:

- a homoscedastic symmetric error specification: $\epsilon_i \sim N(0,1)$.
- a homoscedastic asymmetric error distribution: $\epsilon_i \sim \chi^2(1)$, minus its median.
- a heteroscedastic error distribution: $\epsilon_i \sim (2 + x_{1i})N(0,1)$.

[Table 1 about here.]

The model parameters were estimated using the proposed binary quantile regression approach. For each case 150 Monte Carlo simulations were run. Table 1 summarises the estimated parameters and the standard errors for β_0 and β_1 under all three error specifications[1]. The results of the analysis indicate that even in a relatively small sample size the estimator works relatively well, especially in the homoscedastic cases. Therefore, it can be concluded that the proposed binary quantile regression estimator is a viable alternative to the smoothed maximum score estimator given that its implementation simplicity does not come at the expense of finite sample performance.

4.2 Simulation Example 2 - Variable Selection

In this sub-section the performance of the proposed penalised binary quantile regression approach is investigated through a simulated example.

In this example data was simulated from the following regression model:

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta}(\tau) + \epsilon_i, \tag{19}$$

where $x_i \sim N(0,1)$, $i = 1, ..., n$, $n = 200$ and

$$\boldsymbol{\beta} = (0.5, 1.5, 0, 0, 2, 0, -1, 1)$$

.

20 validation and 20 training and 200 testing observations were simulated from the model and three homoscedastic and one heteroscedastic specifications for the model error ϵ_i were considered,

- a homoscedastic symmetric error specification: $\epsilon_i \sim N(0,1)$
- a Laplace distribution: $\epsilon_i \sim Laplace(0,1)$

[1] The value of β_2 has been normalised to 1.

- a mixture of two Normal distributions: $\epsilon_i \sim 0.1N(0,1) + 0.9N(0,9)$
- a heteroscedastic error distribution: $\epsilon_i \sim (2 + x_{1i})N(0,1)$

The model was fitted using the generated data set. The experiment was repeated 100 times. All the penalised quantile regression estimates were obtained via direct numerical optimisation using the R function `optim`. The penalty parameter in lasso λ was chosen using the a cross-validation method.

[Table 2 about here.]

In the analysis the estimated parameters were compared to the true parameter values. For every data generating process the bias was calculated, which was averaged over the 100 generated datasets from each scenario.

The results of the simulations are summarised in Table 2. It can be observed that, in general, the proposed method performs well when comparing the estimates $\widehat{\beta_j}$ with the true values β_j as the majority of the estimated biases are around or smaller than $|0.1|$.

4.3 Work-trip Mode-Choice Data Example

In order to assess the practical applicability of the proposed approach the method was tested on a previously published maximum score dataset (Horowitz (1993)). Mode choice modelling and prediction relate closely to transportation policies and can be useful for estimating travel demand and for mitigating traffic congestion. The dataset contains 842 observations sampled randomly from the Washington, D.C. area transportation study for each of the following four dependent variables: (i) the number of cars owned by traveller households, CARS, measured in car units; (ii) the transit out-of-vehicle travel time minus automobile out-of-vehicle travel time, DOVTT, measured in minutes; (iii) the transit in-vehicle travel time minus automobile in-vehicle travel time, DIVTT, also measured in minutes; and (iv) the transit

17

fare minus automobile travel cost, DCOST, measured in US dollars. The dependent variable of the resulting binary choice model was CHOOSE, which equals to 1 if the car is used and 0 otherwise, representing the latent variable “willingness to use a car”. All continuous variables were standardised to have zero mean and unit standard deviation for better comparison with results in the literature. Scale normalisation is achieved by setting the coefficient of DCOST equal to 1, as in Horowitz (1993), to enable the comparison of the obtained results to previous research.

Table 3 provides estimates of the model parameters for the median case ($\tau = 0.5$) as well as a comparison with the results obtained by three different estimation approaches, namely the smoothed maximum score estimator (Horowitz (1993)), a mixed integer optimisation (MIP) method (Florios and Skouras (2008)) and a Bayesian binary quantile regression (BBQR) approach based on the asymmetric Laplace distribution (Benoit and Van den Poel (2012)).

[Table 3 about here.]

The analysis suggests that the results obtained by Horowitz (1993) are quite different from the ones obtained by Florios and Skouras (2008), and Benoit and Van den Poel (2012). According to Horowitz (1993), DCOST and CARS are the most important variables influencing the work-trip mode choice, with DCOST being by far the most important variable. In contrast, the results obtained by the other two methods, which are very similar between them, show that the variable CARS is by far the most important variable with the other variables having a small impact. The difficulty in computing maximum score estimates, discussed in Section 1, has been identified by many authors. In the context of computing estimators such algorithms are problematic because the statistical properties of such procedures can differ from those of exact estimates, e.g. as the ones provided by (Florios and Skouras (2008)).

The proposed LAWS approach delivers very similar estimates to the ones obtained both under MIP and BBQR. Furthermore, the technique is able to provide a more in-depth view of the relationship of the dependent variable and the covariates, as it allows to estimate the relationships at different parts of the distribution of the response variable. Figure 1 illustrates the effect of covariates on the response variable at 0.10, 0.25, 0.50 , 0.75 and 0.90 quantile levels. The solid line represents the point estimates of the regression coefficients for the different quantiles and the dotted lines represent the upper and lower levels of a 95% confidence interval.

[Figure 1 about here.]

These results indicate that the effect of CARS and DOVTT on the unobserved willingness to take the car become stronger for higher conditional quantiles. This means that the effect of these variables is not constant across various quantiles of the latent variable. Specifically, commuters who have a low willingness to use the car are less affected by the number of cars whereas commuters with high willingness to use a car are more affected by the number of cars. Furthermore, commuters with increasing willingness to use a car are more affected by increasing out-of vehicle transportation time. In addition the results indicate that CARS is the most important variable as it has three times higher effect than the second variable, followed by the variable DCOST. The effect of DOVTT on the unobserved willingness to take the car is much lower than both CARS and DCOST, whereas, the respective effect of DIVTT is very small as compared to all the other variables.

5 Conclusions

In this paper an alternative estimation approach to binary quantile regression and variable selection is proposed. The approach is based on a nonlinear asymmetrical weighted loss function which can be implemented by an iteratively reweighted least square algorithm (IRLS). Existing algorithms for fitting quantile regression models are not computational straight forward, hence they do not necessarily guarantee convergence and a unique solution. Also, due to their non-standard objective functions they cannot be computed using standard software packages. The main advantage of the proposed approach is that the IRLS algorithm converge to a unique

19

solution, whereas its computational simplicity makes it an attractive alternative to conventional methods. The results of the simulation study indicate that the ease of implementation does not come at the expense of finite sample performance.

Address for correspondence

Katerina Aristodemou, Department of Mathematical Sciences, Brunel University, Kingston Lane, Uxbridge, Middlesex UB8 3PH, UK.

E-mail: katerina.aristodemou@brunel.ac.uk

Appendix A

Proof of theorem 1

Proof. To establish consistency we use the results of Blevins and Khan (2013), who applied the standard consistency theorem of Newey and McFadden (1994) (Theorem 2.1). The proof is similar to those in Manski (1985) and Horowitz (1992).

Let $S_\tau(\boldsymbol{\beta}(\tau)) = [(2Pr(y=1|x_i)-1)-(1-2\tau)]\, I(\boldsymbol{x}_i'\boldsymbol{\beta}(\tau) \geq 0)$ be the population score function. Under Assumptions 4 and 5, for any $0 < \tau < 1$, $S_\tau(\boldsymbol{\beta}(\tau)) \leq S_\tau(\boldsymbol{\beta}_0(\tau))$ with equality only if $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta}_0(\tau)$ (Manski (1985)'s Lemma 3 and Corollary 2).

As in Blevins and Khan (2013) the observations are iid by Assumption 1, compactness of $\mathbb{B}$ is established by Assumption 3 and the objective function is a sample average of bounded functions that are continuous in the parameters. Continuity of the objective function follows from Assumption 5.

To establish consistency it is necessary to show that as $n \to \infty$ the stochastic objective function $S_\tau(\boldsymbol{\beta}(\tau))$ converges in probability to a limit function $S_\tau(\boldsymbol{\beta}_0(\tau))$. Since $\widehat{\boldsymbol{\beta}}(\tau)$ maximises $S_\tau(\boldsymbol{\beta}(\tau))$ by definition it follows that $\widehat{\beta}(\tau) - \beta_0(\tau) \xrightarrow{p} 0$ (Amemiya (1985), Theorem 4.2.1).

Blevins and Khan (2013) proved that under the above assumptions $S_\tau(\boldsymbol{\beta}(\tau)) \xrightarrow{p}$

$S_\tau(\boldsymbol{\beta}_0(\tau))$ by showing that, under the assumption $h_n \to 0$ the component of the limiting objective function that depends on $\beta(\tau)$ is

$$E\left[[1-2(Pr(y=1|x_i))](I\{x_i'\boldsymbol{\beta}(\tau)\geq 0\} - I\{x_i'\boldsymbol{\beta}_0(\tau)\geq 0\})\right],$$

which is clearly 0 for $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta}_0(\tau)$.

In a similar manner, under Assumption 6, the component of the limiting objective function that depends on $\beta(\tau)$ in this case is

$$E\left[[1-2(Pr(y=1|x_i)) - (1-2\tau)](I\{x_i'\boldsymbol{\beta}(\tau)\geq 0\} - I\{x_i'\boldsymbol{\beta}_0(\tau)\geq 0\})\right],$$

which is also clearly 0 for $\boldsymbol{\beta}(\tau) = \boldsymbol{\beta}_0(\tau)$. By the strict monotonicity of $K(\cdot)$ and Assumptions 2, 4 and 5, it follows that this component is also strictly positive if $\boldsymbol{\beta}(\tau) \neq \boldsymbol{\beta}_0(\tau)$ for all $0 < \tau < 1$. Therefore it is also minimised at $\boldsymbol{\beta}_0(\tau)$. Moreover, let $S^*_{n,\tau}$ denote the objective function in (8). Under Assumptions 3 and 7 by Lemma 4 of Horowitz (1992) $|S_{n,\tau} - S^*_{n,\tau}| \xrightarrow{p} 0$ a.s. uniformly. Thus, consistency is established.

□

Proof of theorem 2

Proof. Let $\boldsymbol{\beta} = bbeta^* + \frac{\boldsymbol{u}}{\sqrt{n}}$. For a fixed $0 < \tau < 1$, based on (15) (15) (exactly same inference based on equation (14))

consider

$$\Gamma_n(\boldsymbol{u}) = \sum_{i=1}^{n} w_i(\tau)\left(y_i - \Phi\left(\frac{\boldsymbol{x}_i'(\boldsymbol{\beta}^* + \frac{\boldsymbol{u}}{\sqrt{n}})}{h_n}\right)\right)^2 +$$
$$+\lambda_n \sum_{j=1}^{p} w_j^{lasso}\left|\boldsymbol{\beta}_j^* + \frac{u_j}{\sqrt{n}}\right|.$$

Let $\hat{\boldsymbol{u}}^{(n)} = argmin_{\boldsymbol{u}}\Gamma_n(\boldsymbol{u})$, then $\hat{\boldsymbol{u}}^{(n)} = \sqrt{n}\left(\widehat{\boldsymbol{\beta}}^*_{adapt.lasso_{slaws}}(\tau) - \boldsymbol{\beta}^*\right)$. Using the Taylor expansion and let

$$H^{(n)}(\boldsymbol{u}) = \Gamma_n(\boldsymbol{u}) - \Gamma_n(0),$$

then

$$H^{(n)}(\boldsymbol{u}) = A_1^{(n)} + A_2^{(n)} + A_3^{(n)} + A_4^{(n)},$$

with

$$\begin{aligned}
A_1^{(n)} &= \sum_{i=1}^{n} w_i(\tau)\Bigg(y_i - \phi\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}^*}{h_n}\right)\frac{\boldsymbol{x}_i'\boldsymbol{u}}{\sqrt{n}}\boldsymbol{u} \\
A_2^{(n)} &= \frac{1}{2}\sum_{i=1}^{n} w_i(\tau)\phi'\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}^*}{h_n}\right)\boldsymbol{u}'\frac{\boldsymbol{x}_i\boldsymbol{x}_i'}{n}\boldsymbol{u} \\
A_3^{(n)} &= \frac{\lambda_n}{\sqrt{n}}\sum_{j=1}^{p}\hat{w}_j\sqrt{n}\Bigg(|\boldsymbol{\beta}^* + \frac{u_j}{\sqrt{n}}| - |\boldsymbol{\beta}_j^*|\Bigg) \\
A_4^{(n)} &= n^{-3/2}\sum_{i=1}^{n} w_i(\tau)\frac{1}{6}\phi''\left(\frac{\boldsymbol{x}_i'\tilde{\boldsymbol{\beta}^*}}{h_n}\right)(\boldsymbol{x}_i'\boldsymbol{u})^3.
\end{aligned}$$

where $\phi()$ is the derivative function of $\Phi(.)$, $\tilde{\boldsymbol{\beta}^*}$ is between $\boldsymbol{\beta}^*$ and $\boldsymbol{\beta}^* + \frac{\boldsymbol{u}}{\sqrt{n}}$. Now we show the asymptotic limit of each term.

First, note that $w_i(\tau)(y_i - \phi\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}^*}{h_n}\right) = \rho_\tau(y_i\phi\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}^*}{h_n}\right))$, where $\rho_\tau(.)$ is the 'check function' in quantile regression, so that the asymptotic limit of $\rho_\tau(y_i\phi\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}^*}{h_n}\right))$ can be derived along the same line as that in linear and nonlinear quantile regression (Koenker, 2005; Oberhofer and Haupt, 2015). Then combining the central limit theory and asymptotic normality of quantile regression, we have $A_1^{(n)} \to_d \boldsymbol{u}' N\Bigg(0,\, \tau(1-\tau)\Sigma_A^{-1}/f(0)^2\Bigg)$.

Second, for the term $A_2^{(n)}$, note that $w_i(\tau)\phi'\left(\frac{\boldsymbol{x}_i'\boldsymbol{\beta}^*}{h_n}\right)\frac{\boldsymbol{x}_i\boldsymbol{x}_i'}{n} \to_p \tau(1-\tau)\Sigma^{-1}/f(0)^2$, so $A_2^{(n)} \to_p \frac{1}{2}\tau(1-\tau)\boldsymbol{u}'\Sigma_A^{-1}\boldsymbol{u}/f(0)^2$.

The limit property of $A_3^{(n)}$ follows standard discussion of adaptive lasso (the proof of Theorem 2 of Zou (2006)):

$$\frac{\lambda_n}{\sqrt{n}}\hat{w}_j\sqrt{n}\Bigg(|\beta_j^* + \frac{u_j}{\sqrt{n}}| - |\beta_j^*|\Bigg) \to_p \begin{cases} 0 & \text{if } \beta_j \neq 0 \\ 0 & \text{if } \beta_j = 0 \text{if } u_j = 0 \\ \infty & \text{if } \beta_j = 0 \text{if } u_j \neq 0. \end{cases}$$

The $A_4^{(n)}$ satisfies $6\sqrt{n}A_4^{(n)}$ is bounded due to the exponential form of normalily density and its derivatives.

Therefore, by Slutsky's theorem, we see that $H^{(n)}(\boldsymbol{u}) \to_d H(\boldsymbol{u})$ for every $\boldsymbol{u}$, where

$$H(\boldsymbol{u}) = \tau(1-\tau)\boldsymbol{u}' A \Sigma_A \boldsymbol{u}_A - 2\boldsymbol{u}'_A W_A$$

if $u_j = 0$ for $j \notin A$, and $W = N\Big(0,\ \tau(1-\tau)\Sigma\Big)$. $H^{(n)}$ is convex and the unique minimum of H

□

23

References

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

Amemiya, T. (1985). *Advanced econometrics*. Harvard university press.

Benoit, D. F., Alhamzawi, R., and Yu, K. (2013). Bayesian lasso binary quantile regression. *Computational Statistics*, 28(6):2861–2873.

Benoit, D. F. and Van den Poel, D. (2012). Binary quantile regression: a bayesian approach based on the asymmetric laplace distribution. *Journal of Applied Econometrics*, 27(7):1174–1188.

Blevins, J. R. and Khan, S. (2013). Local nlls estimation of semi-parametric binary choice models. *The Econometrics Journal*, 16(2):135–160.

Breckling, J. and Chambers, R. (1988). M-quantiles. *Biometrika*, 75(4):761–771.

Cavanagh, C. (1987). Limiting behavior of estimators defined by optimization. *unpublished manuscript (Department of Economics, Harvard University)*.

Delgado, M., Rodrıguez-Poo, J., and Wolf, M. (2001). Subsampling inference in cube root asymptotics with an application to manskis maximum score estimator. *Economics Letters*, 73(2):241–250.

Efron, B. (1991). Regression percentiles using asymmetric squared error loss. *Statistica Sinica*, 1:93–125.

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

Horowitz, J. (1992). A smoothed maximum score estimator for the binary response model. *Econometrica: Journal of the Econometric Society*, 60(3):505–531.

24

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

Jones, M. (1994). Expectiles and m-quantiles are quantiles. *Statistics & Probability Letters*, 20(2):149–153.

Kim, J. and Pollard, D. (1990). Cube root asymptotics. *The Annals of Statistics*, 18(1):191–219.

Koenker, R. (2004). Quantile regression for longitudinal data. *Journal of Multivariate Analysis*, 91(1):74–89.

Koenker, R. (2005). *Quantile regression*, volume 38. Cambridge university press.

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 46(1):33–50.

Kokic, P., Chambers, R., Breckling, J., and Beare, S. (1997). A measure of production performance. *Journal of Business & Economic Statistics*, 15(4):445–451.

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

Li, Q., Xi, R., Lin, N., et al. (2010). Bayesian regularized quantile regression. *Bayesian Analysis*, 5(3):533–556.

Li, Y. and Zhu, J. (2008). L1-norm quantile regression. *Journal of Computational and Graphical Statistics*, 17(1):163–185.

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

Manski, C. (1985). Semiparametric analysis of discrete response:: Asymptotic properties of the maximum score estimator. *Journal of Econometrics*, 27(3):313–333.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. *Handbook of econometrics*, 4:2111–2245.

Newey, W. K. and Powell, J. L. (1987). Asymmetric least squares estimation and testing. *Econometrica: Journal of the Econometric Society*, 55(4):819–847.

Park, T. and Casella, G. (2008). The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686.

Powell, J. (1986). Censored regression quantiles. *Journal of econometrics*, 32(1):143–155.

Schnabel, S. K. and Eilers, P. H. (2009). Optimal expectile smoothing. *Computational Statistics & Data Analysis*, 53(12):4168–4177.

Sobotka, F., Kauermann, G., Schulze Waltrup, L., and Kneib, T. (2013). On confidence intervals for semiparametric expectile regression. *Statistics and Computing*, 23(2):135–148.

Sobotka, F. and Kneib, T. (2012). Geoadditive expectile regression. *Computational Statistics & Data Analysis*, 56(4):755–767.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288.

Wu, Y. and Liu, Y. (2009). Variable selection in quantile regression. *Statistica Sinica*, 19(2):801.

Yao, Q. and Tong, H. (1996). Asymmetric least squares regression estimation: A nonparametric approach. *Journal of nonparametric statistics*, 6(2-3):273–292.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American statistical association*, 101(476):1418–1429.

26

Table 1: Simulation Example 1 - Estimated Parameters and (Standard Deviations)

	Normal		Heteroscedastic		Asymmetric	
τ	β_0	β_1	β_0	β_1	β_0	β_1
0.10	-1.21	-0.97	-2.09	-1.90	-0.52	-1.01
	(0.05)	(0.05)	(0.11)	(0.12)	(0.03)	(0.04)
0.25	-0.66	-0.91	-1.1	-1.36	-0.33	-0.99
	(0.04)	(0.05)	(0.06)	(0.09)	(0.03)	(0.04)
0.50	-0.09	-0.89	0.01	-0.83	-0.02	-0.94
	(0.03)	(0.04)	(0.04)	(0.05)	(0.03)	(0.04)
0.75	0.48	-0.90	0.96	-0.49	0.61	-0.86
	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.05)
0.90	1.01	-0.94	1.87	-0.27	1.54	-0.87
	(0.05)	(0.05)	(0.08)	(0.07)	(0.07)	(0.06)

Table 2: Simulation Example 2 - Estimated Bias for Model Parameters

τ	$\widehat{\beta_0}$	$\widehat{\beta_1}$	$\widehat{\beta_2}$	$\widehat{\beta_3}$	$\widehat{\beta_4}$	$\widehat{\beta_5}$	$\widehat{\beta_6}$
			Normal (0,1)				
0.10	0.30,	0.09	0.03	-0.004	0.05	-0.02	-0.03
0.25	0.08	0.03	-0.0009	-0.02	0.04	-0.01	-0.04
0.5	-0.05	0.01	-0.02	-0.03	0.008	-0.02	0.009
0.75	-0.09	-0.02	-0.01	0.007	0.07	-0.03	-0.04
0.90	-0.26	0.099	0.008	0.003	0.11	-0.007	-0.06
			$Laplace(0,1)$				
0.10	0.08	0.08	0.04	0.05	0.09	-0.004	-0.04
0.25	-0.01	-0.08	-0.07	-0.01	-0.08	0.009	0.001
0.5	-0.003	-0.03	-0.04	-0.04	-0.02	-0.02	-0.03
0.75	0.06	0.02	-0.07	-0.06	0.04	-0.1	-0.11
0.90	-0.07	-0.12	-0.08	-0.11	-0.09	-0.03	-0. 13
			Normal mixture				
0.10	0.34	0.09	-0.01	-0.04	0.20	-0.009	-0.09
0.25	0.18	0.06	-0.01	0.02	0.09	-0.004	-0.06
0.5	-0.04	0.0008	-0.04	-0.01	0.02	-0.03	-0.04
0.75	-0.18	0.04	-0.03	-0.01	0.04	-0.03	-0.08
0.90	-0.35	0.02	-0.04	-0.04	0.09	-0.02	-0.06
			Heteroscedastic model				
0.10	0.05	0.40	0.06	-0.08	0.09	-0.05	-0.06
0.25	0.12	0.05	0.02	0.005	-0.22	-0.01	0.08
0.50	-0.29	-0.22	-0.03	-0.06	-0.17	-0.04	-0.02
0.75	0.03	0.03	-0.05	-0.07	0.01	-0.09	0.10
0.90	-0.10	-0.03	-0.09	0.16	0.13	-0.0002	-0.17

28

Table 3: Mode-Choice Data: Model Parameters Estimates

AUTHOR	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
Benoit and Van den Poel(2012)	4.825	3.375	1.018	0.282	1	BBQR
Current study	-1.493	3.545	0.455	0.274	1	LAWS

Figure 1: Mode-choice Dataset: Quantile Curves for Model Parameters

30