Interquantile Shrinkage in Regression Models

Liewen Jiang, Huixia Judy Wang, Howard D. Bondell

Abstract

Conventional analysis using quantile regression typically focuses on fitting the regression model at different quantiles separately. However, in situations where the quantile coefficients share some common feature, joint modeling of multiple quantiles to accommodate the commonality often leads to more efficient estimation. One example of common features is that a predictor may have a constant effect over one region of quantile levels but varying effects in other regions. To automatically perform estimation and detection of the interquantile commonality, we develop two penalization methods. When the quantile slope coefficients indeed do not change across quantile levels, the proposed methods will shrink the slopes towards constant and thus improve the estimation efficiency. We establish the oracle properties of the two proposed penalization methods. Through numerical investigations, we demonstrate that the proposed methods lead to estimations with competitive or higher efficiency than the standard quantile regression estimation in finite samples. Supplemental materials for the article are available online.

Key Words: Fused lasso; Non-crossing; Oracle; Quantile regression; Smoothing; Sup-norm.

1 Introduction

Quantile regression has attracted an increasing amount of attention after being introduced by Koenker and Bassett (1978). One major advantage of quantile regression over classical mean regression is its flexibility in assessing the effect of predictors on different locations of the response distribution. Regression at multiple quantiles provides more comprehensive statistical views than analysis at the mean or at a single quantile level. The standard approach of multiple-quantile regression is to fit the regression model at each quantile separately. However, in applications where the quantile coefficients enjoy some common features across quantile levels, joint modeling of multiple quantiles can lead to more efficient estimations. One special case is the regression model with independent and identically distributed (*i.i.d.*) errors so that the quantile slope coefficients are constant across all quantile levels. Under this assumption, Zou and Yuan (2008a) proposed the Composite Quantile Regression (CQR) method by combining the objective functions at multiple quantiles to estimate the common slopes. For models with *i.i.d.* errors, the composite estimator is more efficient than the conventional estimator obtained at a single quantile level. However, in practice, the *i.i.d.* error assumption is restrictive and needs to be verified. It is likely that the quantile slope coefficients may appear constant in certain quantile regions, but vary in others. For instance, the analysis of an international economic growth data shows that the impact of political instability on the GDP growth tends to be constant for $\tau \in [0, 0.5]$, but varies at upper quantiles; see Figure 3 in Section 5. In order to borrow information across quantiles to improve efficiency, the common structure of quantile slopes has to be determined.

One way to identify the commonality of quantile slopes at multiple quantile levels is through hypothesis testing. Koenker (2005) described the Wald-type test through direct estimation of the asymptotic covariance matrix of the quantile coefficient estimates at multiple quantiles. The Wald-type test can be used to test the equality of quantile slopes at a given set of quantile levels, and this was implemented in the function 'anova.rq' in the R package *quantreg*. The testing approach is feasible if we only test the equality of slopes at a few given

quantile levels. However, to identify the complete quantile regions with constant quantile slopes, at least $2^{p(K-1)}$ tests have to be conducted, where K is the total number of quantile levels and p is the number of predictors. This makes the testing procedure complicated, especially when K and p are large. To overcome this drawback, we propose penalization approaches to allow for simultaneous estimation and automatic shrinkage for interquantile differences of the slope coefficients.

Penalization methods are useful tools for variable selection. In conditional mean regression, various penalties have been introduced to produce sparse models. Tibshirani (1996) employed L_1-norm in Least Absolute Shrinkage and Selection Operator (lasso) for variable selection. Fan and Li (2001) proposed the Smoothly Clipped Absolute Deviation (SCAD) penalty, which is a nonconcave penalty and defined through its first order derivative. Zou (2006) extended the L_1-penalty by including adaptive weights, referred to as adaptive lasso penalty. Yuan and Lin (2006) introduced group lasso to identify significant factors represented as groups of predictors. In applications where the predictors have a natural ordering, Tibshirani, et al. (2005) introduced the Fused lasso, which penalizes the L_1-norm of both coefficients and their successive differences.

The penalization idea was also adopted for quantile regression models in various contexts. Koenker (2004) employed the lasso penalty to shrink random subject effects towards constant in studying conditional quantiles of longitudinal data. Li and Zhu (2008) studied L_1-norm quantile regression and computed the entire solution path. Further work on this problem, with improved algorithm, can be found in Kato (2010) and Osborne and Turlach (2011). Wu and Liu (2009a) further discussed SCAD and adaptive lasso methods in quantile regression and demonstrated their oracle properties. Zou and Yuan (2008b) adopted a group F_∞-norm penalty to eliminate covariates that have no impact on any quantile levels. Li and Zhu (2007) analyzed the quantiles of the Comparative Genomic Hybridization (CGH) data using fused quantile regression. Wang and Hu (2010) proposed fused adaptive lasso to accommodate the spatial dependence in studying multiple array CGH samples from two different groups.

In this work, we adopt the fusion idea to shrink the differences of quantile slopes at two

adjacent quantile levels towards zero. Therefore, the quantile regions with constant quantile slopes can be automatically identified and all the coefficients can be estimated simultaneously. We develop two types of fusion penalties in the multiple-quantile regression model: fused lasso and fused sup-norm, along with their adaptively weighted counterparts.

An additional benefit of our approach is that it is straightforward to include restrictions to enforce non-crossing of the quantile curves on any given domain of interest. The crossing issue refers to the fact that in typical quantile regression estimation, it is possible, and often the case that for a particular region of the covariate space, a more extreme quantile of the response is estimated to be smaller than one that is less extreme. Of course, this is not possible in reality for any valid distribution, but is a common issue plaguing quantile regression estimation. The quantile crossing is often seen in finite samples especially in multiple-quantile regression. The quantile crossing issue has been studied in He (1997), Tackeuchi, et al. (2006), Wu and Liu (2009b), and Bondell, Reich and Wang (2010).

The remainder of this paper is organized as follows. In section 2, we illustrate the proposed methods and discuss the computation issues. In section 3, we discuss the asymptotic properties of the proposed penalization estimators. A simulation study is conducted to assess the numerical performance of our proposed estimators in section 4. We apply the proposed methods to analyze international economic growth data in section 5. All technical details are provided in the appendix.

2 Proposed Method

2.1 Model Setup

Let Y be the response variable and $\mathbf{X} \in \boldsymbol{R}^p$ be the corresponding covariate vector. Suppose we are interested in regression at K quantile levels $0 < \tau_1 < \ldots < \tau_K < 1$, where K is a finite integer. Denote $Q_{\tau_k}(\mathbf{x})$ as the τ_k^{th} conditional quantile function of Y given $\mathbf{X} = \mathbf{x}$, that is, $P\{Y \leq Q_{\tau_k}(\mathbf{x}) | \mathbf{X} = \mathbf{x}\} = \tau_k$, for $k = 1, \ldots, K$. We consider the linear quantile regression

model

$$Q_{\tau_k}(\mathbf{x}) = \alpha_k + \mathbf{x}^T \boldsymbol{\beta}_k, \tag{1}$$

where $\alpha_k \in \boldsymbol{R}$ is the intercept and $\boldsymbol{\beta}_k \in \boldsymbol{R}^p$ is the slope vector at the quantile level τ_k. Let $\{y_i, \mathbf{x}_i\}, i = 1, \cdots, n$, be an observed sample. At a given quantile level τ_k, the conventional quantile regression method estimates $(\alpha_k, \boldsymbol{\beta}_k^T)^T$ by $(\tilde{\alpha}_k, \tilde{\boldsymbol{\beta}}_k^T)^T$, the minimizer of the quantile-specific objective function

$$\sum_{i=1}^{n} \rho_{\tau_k}(y_i - \alpha_k - \mathbf{x}_i^T \boldsymbol{\beta}_k),$$

where $\rho_\tau(r) = \tau r I(r > 0) + (\tau - 1) r I(r \leq 0)$ is the quantile check function and $I(\cdot)$ is the indicator function; see Koenker and Bassett (1978). Minimizing the quantile loss function at each quantile level separately is equivalent to minimizing the following combined loss function

$$\sum_{k=1}^{K} \sum_{i=1}^{n} \rho_{\tau_k}(y_i - \alpha_k - \mathbf{x}_i^T \boldsymbol{\beta_k}). \tag{2}$$

In some applications, however, it is likely that the quantile coefficients share some commonality across quantile levels. For example, the quantile slope may be constant in certain quantile regions for some predictors. Separate estimation at each quantile level will ignore such common features and thus lose efficiency. An alternative strategy is to model multiple quantiles jointly by borrowing information from neighboring quantiles. Zou and Yuan (2008a) proposed a composite quantile estimator by assuming that the quantile slope is constant across all quantiles. Such an assumption is restrictive, and it requires the model structure to be known beforehand, which is hard to determine in practice.

In the following sections, we denote $\beta_{k,l}$ as the slope corresponding to the l-th predictor at the quantile level τ_k, and $d_{k,l} = \beta_{k,l} - \beta_{k-1,l}$ as the slope difference at two neighboring quantiles τ_{k-1} and τ_k, with $k = 2, \ldots, K$ and $d_{1,l} = \beta_{1,l}$ for $l = 1, \ldots, p$. Let $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{d}_1^T, \ldots, \boldsymbol{d}_K^T)^T$ de-

note the collection of unknown parameters, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)^T$ and $\boldsymbol{d}_k = (d_{k,1}, \ldots, d_{k,p})^T$ for $k = 1, \ldots, K$. Therefore, the τ_k^{th} quantile coefficient vector can be written as

$$(\alpha_k, \boldsymbol{\beta}_k^T)^T = \boldsymbol{T}_k \boldsymbol{\theta},$$

where $\boldsymbol{T}_k = (\boldsymbol{D}_{k,0}, \boldsymbol{D}_{k,1}, \boldsymbol{D}_{k,2}) \in \boldsymbol{R}^{(p+1)\times(p+1)K}$, $\boldsymbol{D}_{k,0}$ is a $(p+1) \times K$ matrix with 1 in the first row and the k^{th} column, but zero elsewhere, $\boldsymbol{D}_{k,1} = \mathbf{1}_k^T \otimes (\mathbf{0}_p, \boldsymbol{I}_p)^T \in \boldsymbol{R}^{(p+1)\times pk}$ and $\boldsymbol{D}_{k,2}$ is a $(p+1) \times (K-k)p$ zero matrix. Here $\mathbf{0}_p$ is a $p \times 1$ zero vector, $\boldsymbol{I}_p$ is a $p \times p$ identity matrix and $\mathbf{1}_k$ is a $k \times 1$ vector with all 1's. Define $\mathbf{z}_{ik}^T = (1, \mathbf{x}_i^T)\boldsymbol{T}_k \in \boldsymbol{R}^{1\times(p+1)K}$. With these reparameterizations, the combined quantile objective function (2) can be rewritten as

$$\sum_{k=1}^{K} \sum_{i=1}^{n} \rho_{\tau_k}(y_i - \mathbf{z}_{ik}^T \boldsymbol{\theta}). \tag{3}$$

In order to capture the possible feature that some quantile slope coefficients are constant in some quantile regions, we propose to shrink the interquantile differences $\{d_{k,l}, k = 2, \ldots, K, l = 1, \ldots, p\}$ towards zero, thus inducing smoothing across quantiles.

2.2 Penalized Joint Quantile Estimators

We propose an adaptive fused penalization approach to shrink interquantile slope differences towards zero. The proposed adaptive fused (AF) penalization estimator is defined as

$$\hat{\boldsymbol{\theta}}_{AF} = \arg\min_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}), \text{ where } Q(\boldsymbol{\theta}) = \sum_{k=1}^{K} \sum_{i=1}^{n} \rho_{\tau_k}(y_i - \mathbf{z}_{ik}^T \boldsymbol{\theta}) + \lambda \sum_{l=1}^{p} \|Diag(\tilde{w}_{k,l})\boldsymbol{\theta}_{(l)}\|_\nu. \tag{4}$$

Here $\lambda \geq 0$ is a tuning parameter controlling the degree of penalization, $Diag(\tilde{w}_{k,l})$ is a diagonal matrix with elements $\tilde{w}_{2,l}, \ldots, \tilde{w}_{K,l}$ on the diagonal, $\tilde{w}_{k,l}$ is the adaptive weight for $d_{k,l}$, and $\boldsymbol{\theta}_{(l)} = (d_{2,l}, \ldots, d_{K,l})^T$ can be regarded as a group of parameters corresponding to the l^{th} predictor.

In this paper, we consider two choices of ν: $\nu = 1$ and $\nu = \infty$, corresponding to the Fused Adaptive Lasso (FAL) and Fused Adaptive Sup-norm (FAS) penalization approaches, respectively. Let $\tilde{d}_{k,l}$ be the initial estimator obtained from the conventional quantile regression. For fused adaptive lasso, we set the adaptive weights $\tilde{w}_{k,l} = |\tilde{d}_{k,l}|^{-1}$. For fused adaptive sup-norm, we let $\tilde{w}_{k,l} = 1/\max\{|\tilde{d}_{k,l}|, k = 2, \ldots, K\}$ be the group-wise weights. As a special case, when all the adaptive weights $\tilde{w}_{k,l} = 1$, fused adaptive lasso and fused adaptive sup-norm are reduced to Fused Lasso (FL) and Fused Sup-norm (FS), respectively. Notice that for fused adaptive lasso, the slope differences are penalized individually, leading to piecewise constant quantile slope coefficients. However, for fused adaptive sup-norm, the slope differences are penalized in a group manner, and consequently, either all elements in $\boldsymbol{\theta}_{(l)}$ will be shrunk to be 0, or none of them will be shrunk to 0. In this paper, we focus on the shrinkage of interquantile coefficient differences to achieve higher estimation efficiency. If one desires to remove any predictor from the model for variable selection purpose, additional penalization should be considered.

For regression at multiple quantiles, the estimated quantiles may cross, that is, a lower conditional quantile may be estimated larger than an upper quantile. To avoid quantile crossing, we propose adding additional constraints by adopting the idea in Bondell, Reich and Wang (2010). Without loss of generality, we assume that the desired domain for non-crossing is the region $[0, 1]^p$. We propose to minimize (4) subject to the following non-crossing constraint:

$$(d_\alpha)_k - \sum_{l=1}^{p} d_{k,l}^{-} \geq 0, \;\; k = 2, \ldots, K, \tag{5}$$

where $(d_\alpha)_k = \alpha_k - \alpha_{k-1}$ and $d_{k,l}^{-}$ is the negative part of $d_{k,l} = \beta_{k,l} - \beta_{k-1,l}$. As discussed in Bondell, Reich and Wang (2010), the non-crossing constraint (5) ensures the estimation of the conditional quantiles $Q_\tau(y|\mathbf{x})$ to be nondecreasing in τ for any $\mathbf{x} \in [0, 1]^p$. In addition, the asymptotic properties of the penalized estimators are not affected by adding the constraint (5).

2.3 Computation

For a given t, the minimization can be formulated as a linear programming problem with linear constraints, and thus can be solved by using any existing linear programming software. In our numerical studies, we use the R function "rq.fit.sfn" in the package *quantreg*. This function adopts the sparse Frisch-Newton interior-point algorithm so that the computational time is proportional to the number of nonzero entries in the design matrix.

Note that minimizing (4) with non-crossing constraint (5) is equivalent to solving

$$\min_{\boldsymbol{\theta}} \sum_{k=1}^{K}\sum_{i=1}^{n} \rho_{\tau_k}(y_i - \mathbf{z}_{ik}^T\boldsymbol{\theta}), \text{ s.t. } \sum_{l=1}^{p} \|Diag(\tilde{w}_{k,l})\boldsymbol{\theta}_{(l)}\|_\nu \le t, \text{ and } (d_\alpha)_k - \sum_{l=1}^{p} d_{k,l}^- \ge 0, \qquad (6)$$

where $t > 0$ is a tuning parameter that plays a similar role as λ. Consider the point where $d_{k,l} = 0, l = 1, \ldots, p, k = 2, \ldots, K$, that is, all slopes are constant across quantile levels. Note this point satisfies both inequality constraints in (6). Hence a solution to (6) always exists. Adopting this constraint formulation gives us a natural range of the tuning parameter $t \in [0, t_{\max}]$, where $t_{\max} = \sum_{l=1}^{p} \|Diag(\tilde{w}_{k,l})\tilde{\boldsymbol{\theta}}_{(l)}\|_\nu$ with $\tilde{\boldsymbol{\theta}}_{(l)}$ being the conventional RQ estimator.

The tuning parameter t can be selected by minimizing the Akaike Information Criterion (AIC) (Akaike, 1974):

$$\text{AIC}(t) = \sum_{k=1}^{K} \log\left[\sum_{i=1}^{n} \rho_{\tau_k}\left\{y_i - \mathbf{z}_{ik}^T\hat{\boldsymbol{\theta}}_k(t)\right\}\right] + \frac{1}{n} edf(t), \qquad (7)$$

where the first term measures the goodness of fit (see Bondell, Reich and Wang 2010 for a similar measure for joint quantile regression), $\hat{\boldsymbol{\theta}}_k(t)$ is the solution to (6) with the tuning parameter value t, and $edf(t)$ is the effective degree of freedom associated with the tuning parameter t. We set *edf* as the number of nonzero d's for fused adaptive lasso, and as the number of unique d's for fused adaptive sup-norm (Zhao, Rocha and Yu, 2009).

3 Asymptotic Properties

Define F_i as the conditional cumulative distribution function of Y given $\mathbf{X} = \mathbf{x}_i$. To establish the asymptotic properties of the proposed fused adaptive lasso and fused adaptive sup-norm estimators, we assume the following three regularity conditions.

(A1) For $k = 1, \ldots, K, i = 1, \ldots, n$, the conditional density function of Y given $\mathbf{X} = \mathbf{x}_i$, denoted as f_i, is continuous, and has a bounded first derivative, and $f_i\{Q_{\tau_k}(\mathbf{x}_i)\}$ is uniformly bounded away from zero and infinity.

(A2) $\max_{1 \le i \le n} \|\mathbf{x}_i\| = o(n^{1/2})$.

(A3) For $1 \le k \le K$, there exist some positive definite matrices $\mathbf{\Gamma}_k$ and $\mathbf{\Omega}_k$ such that $\lim_{n\to\infty} n^{-1} \sum_{i=1}^n \mathbf{z}_{ik}\mathbf{z}_{ik}^T = \mathbf{\Gamma}_k$ and $\lim_{n\to\infty} n^{-1} \sum_{i=1}^n f_i\{Q_{\tau_k}(\mathbf{x}_i)\}\mathbf{z}_{ik}\mathbf{z}_{ik}^T = \mathbf{\Omega}_k$.

In this paper, we focus on regression at multiple quantiles. In situations where the quantile slopes of a particular predictor are not constant across quantile levels, the boundedness in that predictor direction is needed to ensure the validity of the linear quantile regression model (1). In particular, if the support were unbounded in that direction, then the slopes must be constant across quantiles, otherwise, they would cross, and the model itself would be invalid. We note that partially (or fully) bounding the support is a special case of (A2).

3.1 Fused Adaptive Lasso Estimator

For ease of illustration, we consider $p = 1$ in this subsection. Notations can then be simplified by letting θ_j be the j^{th} element of $\boldsymbol{\theta}$ and $\tilde{w}_j = |\tilde{\theta}_{K+j}|^{-1}$ for $j = 2, \ldots, K$. Denote $\boldsymbol{\theta}_0 = (\theta_{j,0}, j = 1, \ldots, 2K)$ as the true value of $\boldsymbol{\theta}$. Let the index sets $\mathcal{A}_1 = \{1, \ldots, K\}$, $\mathcal{A}_2 = \{j : \theta_{j,0} \neq 0, j = K+1, \ldots, 2K\}$, and $\mathcal{A} = \mathcal{A}_1 \cup \mathcal{A}_2$. We write $\boldsymbol{\theta}_{\mathcal{A}} = (\theta_j : j \in \mathcal{A})^T$, and its truth as $\boldsymbol{\theta}_{\mathcal{A},0} = (\theta_{j,0} : j \in \mathcal{A})^T$.

Before discussing the asymptotic property of the fused adaptive lasso estimator, we first examine the oracle estimator in this setting with $p = 1$. Without loss of generality, we

assume that the quantile slopes β_k vary for the first $s < K$ quantiles, but remain constant for the remaining $(K - s)$ quantile levels. The properties of the oracle and fused adaptive lasso estimators for other more general cases follow the similar exposition, but with more complicated notations. Suppose the model structure is known, the oracle estimator $\hat{\boldsymbol{\theta}}_{\mathcal{A}} \in \boldsymbol{R}^{K+s}$ can be obtained by

$$\hat{\boldsymbol{\theta}}_{\mathcal{A}} = \arg\min_{\boldsymbol{\theta}_{\mathcal{A}}} \sum_{k=1}^{K} \sum_{i=1}^{n} \rho_{\tau_k}(y_i - \boldsymbol{z}_{ik,\mathcal{A}}^T \boldsymbol{\theta}_{\mathcal{A}}),$$

where $\boldsymbol{z}_{ik,\mathcal{A}} \in \boldsymbol{R}^{K+s}$ contains the first $K + s$ elements of $\boldsymbol{z}_{ik}$.

Proposition 1. *Under conditions (A1)-(A3), we have*

$$n^{1/2}(\hat{\boldsymbol{\theta}}_{\mathcal{A}} - \boldsymbol{\theta}_{\mathcal{A},0}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{\mathcal{A}}), \ \text{as } n \to \infty,$$

where $\boldsymbol{\Sigma}_{\mathcal{A}} = \left(\sum_{k=1}^{K} \boldsymbol{\Omega}_{k,\mathcal{A}}\right)^{-1} \left\{\sum_{k=1}^{K} \tau_k(1-\tau_k)\boldsymbol{\Gamma}_{k,\mathcal{A}}\right\} \left(\sum_{k=1}^{K} \boldsymbol{\Omega}_{k,\mathcal{A}}\right)^{-1}$, $\boldsymbol{\Omega}_{k,\mathcal{A}}$ *and* $\boldsymbol{\Gamma}_{k,\mathcal{A}}$ *are the top-left* $(K + s) \times (K + s)$ *submatrices of* $\boldsymbol{\Omega}_k$ *and* $\boldsymbol{\Gamma}_k$*, respectively.*

However, in practice, the true model structure is usually unknown beforehand. Hence, we take into account the full parameter vector $\boldsymbol{\theta}$ and estimate it as $\hat{\boldsymbol{\theta}}_{FAL} = \arg\min_{\boldsymbol{\theta}} Q(\boldsymbol{\theta})$, where $Q(\boldsymbol{\theta})$ is defined in (4) with $\nu = 1$. We show that $\hat{\boldsymbol{\theta}}_{FAL}$ has the following oracle property.

Theorem 1. *Suppose that conditions (A1)-(A3) hold. If* $n^{1/2}\lambda_n \to 0$ *and* $n\lambda_n \to \infty$ *as* $n \to \infty$*, we have*

1. sparsity: $Pr\Big(\{j : \hat{\theta}_{j,FAL} \neq 0, j = K + 1, \ldots, 2K\} = \mathcal{A}_2\Big) \to 1$*;*

2. asymptotic normality: $n^{1/2}(\hat{\boldsymbol{\theta}}_{\mathcal{A},FAL} - \boldsymbol{\theta}_{\mathcal{A},0}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{\mathcal{A}})$*, where* $\boldsymbol{\Sigma}_{\mathcal{A}}$ *is the covariance matrix of the oracle estimator given in Proposition 1.*

3.2 Fused Adaptive Sup-norm Estimator

To illustrate the asymptotic property of the fused adaptive sup-norm method, we consider the general case with p predictors. Let $\boldsymbol{\theta}_{(0)} = (\alpha_1, \ldots, \alpha_K)^T \in \boldsymbol{R}^K$ be the vector of intercepts, $\boldsymbol{\theta}_{(-1)} = \boldsymbol{d}_1 = (\beta_{1,1}, \ldots, \beta_{1,p})^T \in \boldsymbol{R}^p$ be the slope coefficient at τ_1, and $\boldsymbol{\theta}_{(l)} = (d_{2,l}, \ldots, d_{K,l})^T \in$

$\boldsymbol{R}^{K-1}$ be the interquantile slope differences associated with the l^{th} predictor. For notational convenience, we reorder $\boldsymbol{\theta}$ and define the new parameter vector $\boldsymbol{\theta} = (\boldsymbol{\theta}_{(-1)}^T, \boldsymbol{\theta}_{(0)}^T, \ldots, \boldsymbol{\theta}_{(p)}^T)^T$. The vector $\mathbf{z}_{ik}$ here is an updated vector with the order of elements corresponding to the new parameter vector $\boldsymbol{\theta}$. Define the index sets $\mathcal{B}_1 = \{-1, 0\}, \mathcal{B}_2 = \{l : \|\boldsymbol{\theta}_{(l)}\| \neq 0, l = 1, \ldots, p\}$, and $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$. Hence $\boldsymbol{\theta}_{\mathcal{B}} = (\boldsymbol{\theta}_{(l)}^T : l \in \mathcal{B})^T$ is the non-null subset of $\boldsymbol{\theta}$ and the true parameter vector $\boldsymbol{\theta}_{\mathcal{B},0} = (\boldsymbol{\theta}_{(l),0}^T : l \in \mathcal{B})^T$. Without loss of generality, we assume that the quantile slopes vary across quantiles for the first $g \geq 0$ predictors and remain constant for the others. That is, $\|\boldsymbol{\theta}_{(l)}\| = 0$ for $l = g+1, \ldots, p$.

Proposition 2. *Let $\hat{\boldsymbol{\theta}}_{\mathcal{B}}$ be the oracle estimator of $\boldsymbol{\theta}_{\mathcal{B},0}$ obtained by knowing the true structure. Assuming conditions (A1)-(A3) hold, we have*

$$n^{1/2}(\hat{\boldsymbol{\theta}}_{\mathcal{B}} - \boldsymbol{\theta}_{\mathcal{B},0}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}_{\mathcal{B}}), \text{ as } n \to \infty,$$

where $\boldsymbol{\Sigma}_{\mathcal{B}} = \left(\sum_{k=1}^K \boldsymbol{\Omega}_{k,\mathcal{B}}\right)^{-1} \left\{\sum_{k=1}^K \tau_k(1-\tau_k)\boldsymbol{\Gamma}_{k,\mathcal{B}}\right\} \left(\sum_{k=1}^K \boldsymbol{\Omega}_{k,\mathcal{B}}\right)^{-1}$, $\boldsymbol{\Omega}_{k,\mathcal{B}}$ and $\boldsymbol{\Gamma}_{k,\mathcal{B}}$ are the top-left $m \times m$ submatrices of $\boldsymbol{\Omega}_k$ and $\boldsymbol{\Gamma}_k$, respectively, where $m = K + p + g(k-1)$.

Theorem 2 shows that when the true model structure is unknown, the fused adaptive sup-norm penalized estimator of $\boldsymbol{\theta}$, denoted as $\hat{\boldsymbol{\theta}}_{FAS}$, has the following oracle property.

Theorem 2. *Suppose that conditions (A1)-(A3) hold. If $n^{1/2}\lambda_n \to 0$ and $n\lambda_n \to \infty$ as $n \to \infty$, we have*
1. sparsity: $Pr\left(\{l : \|\hat{\boldsymbol{\theta}}_{(l),FAS}\| \neq \mathbf{0}, l = 1, \ldots, p\} = \mathcal{B}_2\right) \to 1$;
2. asymptotic normality: $n^{1/2}(\hat{\boldsymbol{\theta}}_{\mathcal{B},FAS} - \boldsymbol{\theta}_{\mathcal{B}}) \to N(0, \boldsymbol{\Sigma}_{\mathcal{B}})$ in distribution, where $\boldsymbol{\Sigma}_{\mathcal{B}}$ is the covariance matrix of the oracle estimator given in Proposition 2.

4 Simulation Study

We consider three different examples to assess the finite sample performance of our proposed methods. In each example, the simulation is repeated 500 times with 9 quantile levels $\tau =$

$\{0.1, 0.2, \ldots, 0.9\}$ being considered. We compare the following approaches, the fused adaptive lasso (FAL) method, the fused lasso method without adaptive weights (FL), the fused adaptive sup-norm (FAS) method, the fused sup-norm method without adaptive weights (FS), and the conventional quantile regression method (RQ).

To evaluate various approaches, we examine three performance measurements. The first one is the mean of integrated squared error (MISE), defined as the average of ISE over 500 simulations, where

$$ISE = \frac{1}{n}\sum_{i=1}^{n}\{(\alpha_k + \mathbf{x}_i^T\boldsymbol{\beta}_k) - (\hat{\alpha}_k + \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_k)\}^2.$$

Here, $(\alpha_k + \mathbf{x}_i^T\boldsymbol{\beta}_k)$ and $(\hat{\alpha}_k + \mathbf{x}_i^T\hat{\boldsymbol{\beta}}_k)$ are the true and estimated τ_k^{th} conditional quantile of Y given $\mathbf{x}_i$. The MISE aims to assess the estimation efficiency. The second measurement is the overall oracle proportion (ORACLE), defined as the proportion of times where the true model is selected correctly, that is, all nonzero interquantile slope differences are estimated as nonzero and all zero ones are shrunk exactly to zero. The oracle proportion measures the structure identification ability. The third measurement is True Proportion (TP), defined as the proportion of times where the individual difference of quantile slopes at two adjacent quantile levels is estimated correctly as either zero or non-zero.

Example 1. This example corresponds to a model with a univariate predictor. The data is generated from

$$y_i = \alpha + \beta x_i + (1 + \gamma x_i)e_i, \quad i = 1, \ldots, 200, \tag{8}$$

where $x_i \overset{i.i.d}{\sim} U(0,1), e_i \overset{i.i.d}{\sim} N(0,1), \alpha = 1, \beta = 3$, and $\gamma \geq 0$ controls the degree of heteroscedasticity. Under model (8), the τ^{th} conditional quantile of Y given x is $Q_\tau(x) = \alpha(\tau) + \beta(\tau)x$, where $\alpha(\tau) = \alpha + \Phi^{-1}(\tau)$, $\beta(\tau) = \beta + \gamma\Phi^{-1}(\tau)$ and $\Phi^{-1}(\tau)$ is the τ^{th} quantile of $N(0,1)$. If $\gamma = 0$, (8) is a homoscedastic model with the constant quantile slope $\beta(\tau) = \beta$. However, if $\gamma \neq 0$, (8) becomes a heteroscedastic model with the quantile slope $\beta(\tau)$ varying in τ. For this example with $p = 1$, fused adaptive sup-norm is equivalent to fused sup-norm since the

penalty only involves one group-wise weight, which can be incorporated into the penalization parameter, λ.

Table 1: The MISE and ORACLE proportions of different methods in Example 1. The values in the parentheses are the standard errors of 10^3*MISE.

	10^3*MISE									ORACLE
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	-
					$\gamma = 2$					
FL	120.96	79.50	66.26	61.34	62.79	61.57	68.28	83.26	136.00	0.31
	(6.46)	(4.03)	(3.37)	(3.03)	(3.13)	(3.46)	(3.65)	(4.27)	(6.18)	-
FAL	121.36	84.06	70.76	63.78	66.75	65.93	74.43	88.69	141.40	0.018
	(6.53)	(4.30)	(3.62)	(3.15)	(3.24)	(3.82)	(3.95)	(4.53)	(6.45)	-
FS	118.50	75.53	61.15	56.02	55.27	56.04	61.63	77.05	131.88	1
	(6.26)	(3.77)	(3.07)	(2.79)	(2.77)	(3.04)	(3.33)	(3.85)	(5.97)	-
RQ	117.51	81.03	67.11	62.17	63.93	62.47	69.10	84.67	129.29	-
	(6.58)	(4.07)	(3.39)	(3.06)	(3.14)	(3.50)	(3.73)	(4.45)	(6.22)	-
					$\gamma = 0$					
FL	22.77	17.14	14.70	13.73	13.99	13.99	14.88	17.59	24.51	0.418
	(1.15)	(0.79)	(0.67)	(0.62)	(0.64)	(0.70)	(0.69)	(0.80)	(1.08)	–
FAL	24.24	17.50	15.00	13.94	14.16	14.14	15.07	17.90	25.74	0.298
	(1.22)	(0.79)	(0.67)	(0.63)	(0.65)	(0.71)	(0.70)	(0.84)	(1.13)	–
FS	22.57	16.86	14.92	14.03	13.81	13.95	14.98	17.44	23.74	0.364
	(1.14)	(0.76)	(0.69)	(0.63)	(0.64)	(0.70)	(0.68)	(0.79)	(1.03)	–
RQ	28.45	20.10	16.80	15.93	15.78	15.62	16.73	20.66	30.81	–
	(1.36)	(0.94)	(0.75)	(0.71)	(0.71)	(0.76)	(0.76)	(0.93)	(1.31)	–

MISE: Mean of Integrated Squared Errors; ORACLE: the proportion of times where the model is selected correctly among 500 simulations; FL: Fused Lasso; FAL: Fused Adaptive Lasso; FS: Fused Sup-norm; FAS: Fused Adaptive Sup-norm; RQ: Regular Quantile Regression.

Table 1 and Figure 1 contain the simulation results for example 1 with both $\gamma = 2$ and $\gamma = 0$. In the case with $\gamma = 2$, where the slope coefficients $\beta(\tau)$ vary across quantile levels, Figure 1 shows that the estimation efficiency of our proposed estimators is not significantly worse than RQ estimator, which is the oracle estimator under the model with no any sparsity. Table 1 further shows that the group-wise shrinkage method fused sup-norm performs the best with the smallest MISE and the highest ORACLE. When $\gamma = 0$, the slope coefficients remain invariant across quantile levels. Our proposed methods, by shrinking quantile slope coefficients to be constant, return significantly smaller mean squared error (MSE) and MISE

over τ compared to RQ (see Figure 1).

Asymptotically, fused adaptive lasso has the oracle property with adaptive weights, while fused lasso does not have such nice property except for special cases where the quantile slope coefficients are all constant under the condition $\sqrt{n}\lambda_n \to \infty$, or there are no common quantile slope coeffcients at all with $\sqrt{n}\lambda_n \to 0$. In example 1, the slope coefficients either remain invariant across quantile levels, or vary at all quantile levels. The ideal adaptive weighting scheme is assigning equal weights to all interquantile differences. The fused lasso method does not assign any weight and thus is equivalent to assigning equal component-wise weights. However, fused adaptive lasso adopts the weights calculated from the RQ estimates which contains variability, particularly in the tails. Consequently, fused lasso performs slightly better than fused adaptive lasso for both $\gamma = 0$ and 2.

Example 2. To further demonstrate the role of the adaptive component-wise weights, we consider a more complex univariate example. Let $x_i \sim U(0,1)$ for $i = 1, \ldots, 500$. Assume $Q_\tau(x_i) = \alpha(\tau) + \beta(\tau)x_i$ for any $0 < \tau < 1$, where $\alpha(\tau) = \alpha + \Phi^{-1}(\tau)$ and

$$\beta(\tau) = \begin{cases} \beta - \gamma\Phi^{-1}(0.49) + \gamma\Phi^{-1}(\tau) & \text{if } 0 < \tau < 0.49 \\ \beta & \text{if } 0.49 \leq \tau < 1, \end{cases}$$

with $\alpha = 0, \beta = 3$, and $\gamma = 15$. To generate data, we first generate a quantile level $u_i \sim U(0,1)$ and let $y_i = \alpha(u_i) + \beta(u_i)x_i$. Therefore, for the quantiles $\tau = \{0.1, 0.2, \ldots, 0.9\}$, $\beta(\tau)$ varies for $\tau = 0.1, \ldots, 0.4$, but remains as a constant for $\tau = 0.5, \ldots, 0.9$.

In this setting, $\beta(\tau)$ has different patterns across τ. Adaptive weights are expected to lead to more efficient shrinkage by controlling the shrinkage speeds of different coefficient differences $d_k = \beta(\tau_k) - \beta(\tau_{k-1}), k = 2, \ldots, 9$. By assigning larger weights to the upper quantiles at which $\beta(\tau)$ is a constant, fused adaptive lasso achieves much higher ORACLE and TP at the upper five quantiles than fused lasso (see Table 2). This suggests that if the interquantile slope differences are well distinguished, employing adaptive weights can effectively improve the structure identification accuracy.

Table 2: Percentage of correctly identifying $d_k = \beta(\tau_k) - \beta(\tau_{k-1})$ over 500 simulations in Examples 1-2.

Example 1, $\gamma = 2$, TP									
	$d_2 \neq 0$	$d_3 \neq 0$	$d_4 \neq 0$	$d_5 \neq 0$	$d_6 \neq 0$	$d_7 \neq 0$	$d_8 \neq 0$	$d_9 \neq 0$	ORACLE
FL	0.69	0.89	0.93	0.91	0.94	0.92	0.89	0.67	0.31
FAL	0.65	0.66	0.67	0.65	0.65	0.67	0.67	0.62	0.02
Example 1, $\gamma = 0$, TP									
	$d_2 = 0$	$d_3 = 0$	$d_4 = 0$	$d_5 = 0$	$d_6 = 0$	$d_7 = 0$	$d_8 = 0$	$d_9 = 0$	ORACLE
FL	0.91	0.82	0.81	0.78	0.79	0.77	0.82	0.88	0.42
FAL	0.84	0.81	0.84	0.85	0.87	0.81	0.83	0.81	0.30
Example 2, TP									
	$d_2 \neq 0$	$d_3 \neq 0$	$d_4 \neq 0$	$d_5 \neq 0$	$d_6 = 0$	$d_7 = 0$	$d_8 = 0$	$d_9 = 0$	ORACLE
FL	1.00	1.00	1.00	0.99	0.50	0.66	0.67	0.81	0.25
FAL	0.96	0.97	0.97	0.84	0.78	0.98	0.98	0.96	0.52

TP: percentage of correctly identifying each $d_k = \beta(\tau_k) - \beta(\tau_{k-1}), k = 2, \ldots, 9$ over 500 simulations. ORACLE: overall percentage of correctly shrinking all zero d_k to zero, and nonzero d_k as nonzero.

Example 3. In this example, we consider a bivariate case with $p = 2$. The data are generated from

$$y_i = \alpha + \beta_1 x_{i,1} + \beta_2 x_{i,2} + (1 + \gamma x_{i,2}) e_i, \ i = 1, \ldots, 200,$$

where $x_{i,1} \overset{i.i.d}{\sim} U(0,1)$, $x_{i,2} \overset{i.i.d}{\sim} U(0,1)$, $e_i \overset{i.i.d}{\sim} N(0,1)$, $\alpha = 1, \beta_1 = \beta_2 = 3$. Therefore, the τ^{th} conditional quantile of Y given x_1 and x_2 is

$$Q_\tau(x_1, x_2) = \alpha(\tau) + \beta_1(\tau) x_1 + \beta_2(\tau) x_2,$$

where $\alpha(\tau) = \alpha + \Phi^{-1}(\tau)$, $\beta_1(\tau) = \beta_1$ and $\beta_2(\tau) = \beta_2 + \gamma\Phi^{-1}(\tau)$. Unlike $\beta_1(\tau)$, which stays invariant for all τ, $\beta_2(\tau)$ is constant when $\gamma = 0$ but it varies across τ when $\gamma \neq 0$.

Figure 2 reveals that in the bivariate case with $\gamma = 2$, where the true $\beta_1(\tau)$ stay invariant across τ, our proposed approaches yield significantly smaller MSE of the estimated $\hat{\beta}_1(\tau)$ over τ than RQ method. On the other hand, the true $\beta_2(\tau)$ vary across quantiles, our proposed

fused lasso and fused adaptive lasso approaches result in comparable estimation efficiency with RQ, while fused sup-norm and fused adaptive sup-norm yield smaller MSE of the estimated $\hat{\beta}_2(\tau)$, especially in the middle quantiles. This is largely due to the nature of the group-wise shrinkage methods: fused sup-norm and fused adaptive sup-norm either shrink all interquantile slope differences $d_{k,l} = \beta_{k,l} - \beta_{k-1,l}, l = 1, 2, k = 2, \ldots, 9$ to be exactly zero, or none of them to be zero. It is very likely that fused sup-norm and fused adaptive sup-norm induce some smoothness on the quantile slope coefficients for the second predictor, but do not set any of them to be exactly equal. It is also shown in Table 3 that fused adaptive sup-norm has higher ORACLE than fused lasso and fused adaptive lasso. By imposing two distinguished group-wise weights on two groups of interquantile slope differences, fused adaptive sup-norm leads to better model selection results than fused sup-norm. We compare fused lasso and fused adaptive lasso in Table 4. When the true d's are indeed zero, fused adaptive lasso has higher TP than fused lasso (see $d_{2,1}, \ldots, d_{9,1}$ with $\gamma = 2$ in Table 4), but when the true d's are nonzero, fused adaptive lasso has lower TP's than fused lasso (see $d_{2,2} \ldots d_{9,2}$ with $\gamma = 2$ in Table 4). This implies that fused adaptive lasso may suffer from over-shrinkage problems in the varying regions.

Table 3 shows the simulation results when $\gamma = 0$. As we expect, all the proposed shrinkage methods yield significantly smaller MSE and MISE than RQ when the slope coefficients are constant for each predictor (see Figure 2). However, similar to Example 1, the quantile slope coefficients within each group are too close to be separated, which prevents the component-wise adaptive weights from playing effective roles. Thus, fused adaptive lasso has less estimation efficiency (higher MISE) and worse selection accuracy (lower ORACLE) than fused lasso, especially in the tails where the initial quantile estimates are more variable.

Table 3: The MISE and ORACLE proportion of different methods in Example 3 with $\gamma = 2$ and $\gamma = 0$, respectively.

					10^3*MISE					ORACLE
τ	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	–
					$\gamma = 2$					
FL	178.82	114.03	100.65	89.29	94.28	90.22	96.94	115.04	176.77	0.006
	(7.68)	(4.64)	(4.22)	(4.00)	(4.30)	(3.89)	(3.93)	(4.68)	(7.43)	–
FAL	175.74	119.45	104.15	92.23	99.77	95.69	101.74	122.40	175.00	0
	(7.65)	(5.10)	(4.52)	(4.20)	(4.61)	(4.13)	(4.31)	(5.05)	(7.63)	–
FS	156.21	107.46	91.09	80.49	79.45	81.17	88.78	106.70	157.90	0.264
	(6.68)	(4.41)	(3.88)	(3.54)	(3.45)	(3.63)	(3.57)	(4.26)	(6.19)	–
FAS	154.74	106.85	91.10	81.15	82.30	82.73	88.97	106.65	154.94	0.402
	(6.71)	(4.43)	(3.93)	(3.52)	(3.59)	(3.67)	(3.56)	(4.30)	(6.21)	–
RQ	181.88	122.62	106.27	94.27	98.19	97.62	103.84	122.58	173.38	–
	(7.34)	(4.70)	(4.39)	(4.04)	(4.37)	(4.17)	(4.08)	(4.87)	(7.09)	–
					$\gamma = 0$					
FL	31.00	25.03	22.58	20.53	20.84	21.04	22.55	25.68	30.99	0.276
	(1.19)	(0.94)	(0.83)	(0.79)	(0.80)	(0.78)	(0.82)	(0.98)	(1.22)	–
FAL	35.43	26.16	22.81	21.15	21.39	21.68	23.38	26.51	34.67	0.154
	(1.37)	(0.96)	(0.85)	(0.81)	(0.81)	(0.79)	(0.85)	(1.01)	(1.35)	(0.016)
FS	31.75	25.15	22.51	20.43	20.84	20.78	22.25	25.30	30.67	0.216
	(1.21)	(0.93)	(0.82)	(0.79)	(0.79)	(0.79)	(0.81)	(0.94)	(1.19)	–
FAS	31.81	25.13	22.56	20.57	20.74	20.65	22.12	25.35	31.09	0.228
	(1.22)	(0.93)	(0.82)	(0.81)	(0.80)	(0.78)	(0.80)	(0.94)	(1.21)	–
RQ	47.26	31.71	26.36	24.25	25.21	25.11	27.30	31.84	45.02	–
	(1.74)	(1.13)	(0.94)	(0.91)	(0.95)	(0.93)	(0.96)	(1.16)	(1.67)	–

Table 4: True Proportion of True Positive (TP) for each interquantile difference $d_{k,l}$ in Example 3 with $\gamma = 2$ and $\gamma = 0$, respectively, where $d_{k,l} = \beta_{k,l} - \beta_{k-1,l}, l = 1, 2, k = 2, \ldots, 9$. For $\gamma = 2$, the true coefficients $d_{k,1} = 0$, but $d_{k,2} \neq 0$ for all k. For $\gamma = 0, d_{k,l} = 0$ for all k and l.

	$d_{2,1}$	$d_{3,1}$	$d_{4,1}$	$d_{5,1}$	$d_{6,1}$	$d_{7,1}$	$d_{8,1}$	$d_{9,1}$	$d_{2,2}$	$d_{3,2}$	$d_{4,2}$	$d_{5,2}$	$d_{6,2}$	$d_{7,2}$	$d_{8,2}$	$d_{9,2}$
								$\gamma = 2$								
FL	0.79	0.69	0.62	0.58	0.59	0.62	0.79	0.61	0.85	0.92	0.92	0.92	0.92	0.89	0.87	0.58
FAL	0.79	0.80	0.81	0.77	0.82	0.79	0.80	0.79	0.63	0.67	0.67	0.66	0.69	0.65	0.69	0.58
								$\gamma = 0$								
FL	0.93	0.83	0.80	0.75	0.76	0.78	0.83	0.91	0.92	0.85	0.83	0.80	0.78	0.78	0.85	0.91
FAL	0.848	0.808	0.85	0.82	0.84	0.82	0.82	0.83	0.81	0.82	0.85	0.82	0.84	0.83	0.83	0.80

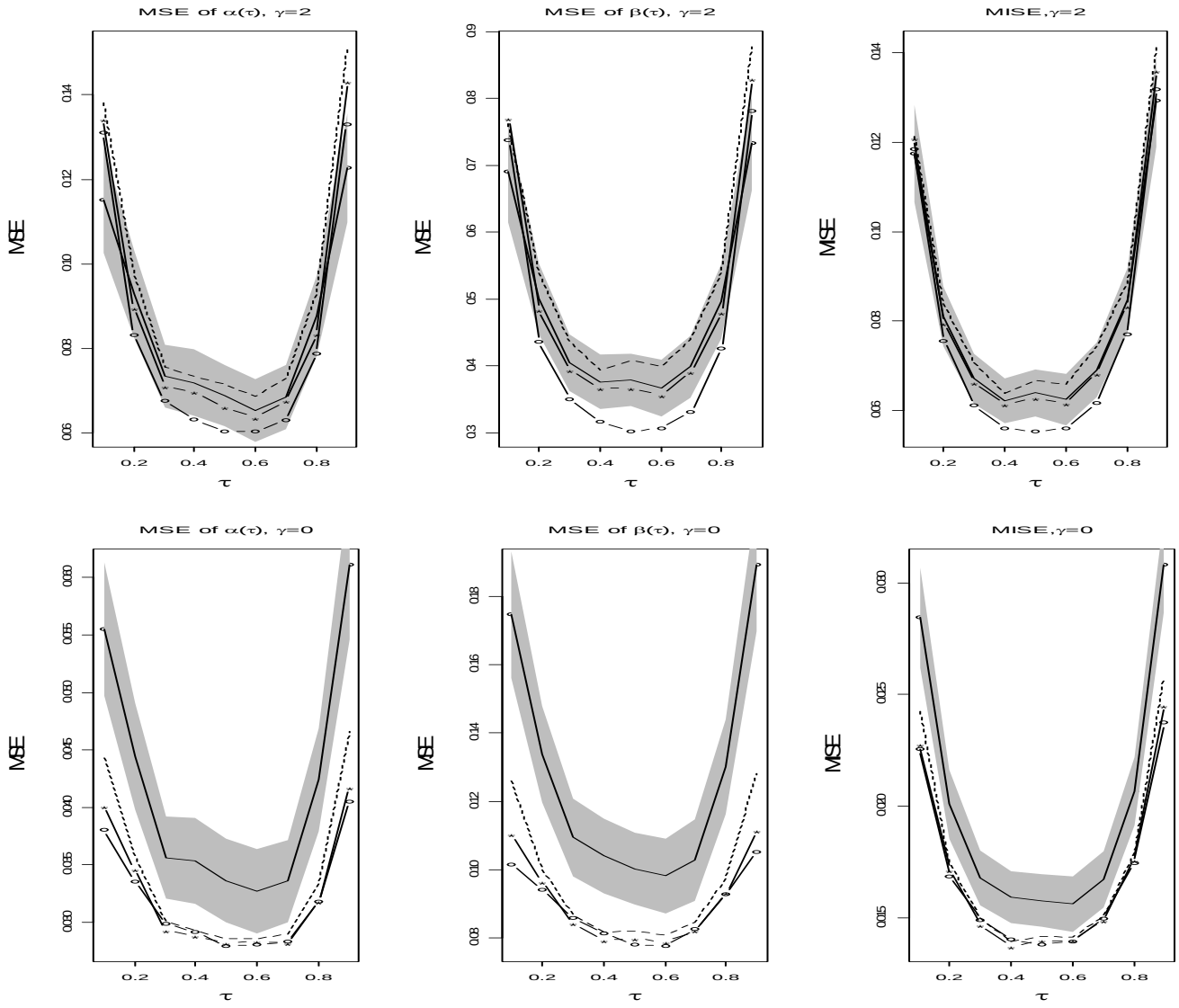

Figure 1: Mean squared error (MSE) of coefficient estimates and the mean integrated squared error (MISE) over τ for example 1. When $\gamma = 2$, slope coefficients vary across τ. When $\gamma = 0$, slope coefficients are constant over τ. The solid line is for RQ, the line with stars on it is for fused lasso, the dashed line is for fused adaptive lasso, and the dotted line is for fused sup-norm approach. The shaded area is the 90% confidence band calculated from RQ method.

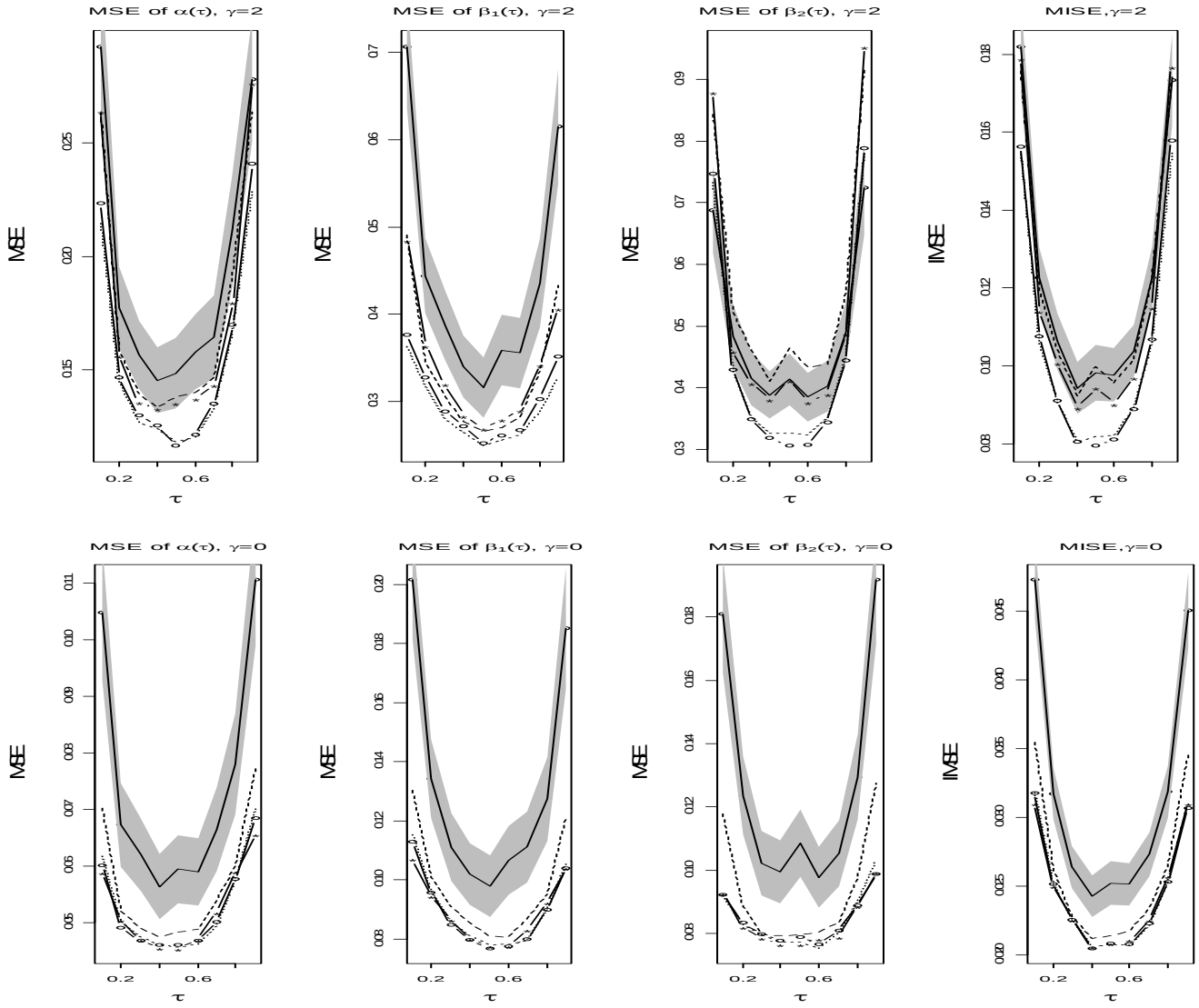

Figure 2: Mean squared error (MSE) of coefficient estimates and the mean integrated squared error (MISE) over τ for example 3. When $\gamma = 2$, slope coefficients are constant for the first predictor, but vary across quantiles for the second one. When $\gamma = 0$, slope coefficients are constant for both predictors. The solid line is for RQ, the line with stars on it is for fused lasso, the dashed line is for fused adaptive lasso, and the dotted line is for fused sup-norm approach. The shaded area is the 90% confidence band calculated from RQ method.

5 Analysis of Economic Growth Data

We apply the proposed methods to analyze the international economic growth data, which were originally from Barro and Lee (1994) and later studied by Koenker and Machado (1999).

The data set contains 161 observations. The first 71 observations correspond to 71 countries and their averaged annual growth percentages of per Capita Gross Domestic Product (GDP growth) from period 1965-1975 are recorded. The remaining 90 observations are for period 1975-1985. Some countries may appear in both periods. There are 13 covariates in total: the initial per capital GDP ($igdp$), male middle school education (mse), female middle school education (fse), female higher education (fhe), male higher education (mhe), life expectancy ($lexp$), human capital ($hcap$), the ratio of eduction and GDP growth (edu), the ratio of investment and GDP growth ($ivst$), the ratio of public consumption and GDP growth ($pcon$), black market premium ($blakp$), political instability (pol) and growth rate terms trade ($ttrad$). All covariates are standardized to lie in the interval $[0, 1]$ before analysis, and we focus on $\tau = \{0.1, 0.2, \ldots, 0.9\}$. Our purpose is to investigate the effects of covariates on multiple conditional quantiles of the GDP growth. In Figure 4, we list 3 predictors as examples.

Koenker and Machado (1999) studied the effects of covariates on certain conditional quantiles of the GDP growth by using the conventional quantile regression method (RQ). In this study, we consider simultaneous regression of multiple quantiles by employing the proposed adaptively weighted penalization methods fused adaptive lasso and fused adaptive sup-norm, and select the penalization parameter by minimizing the AIC value as described in Section 2.3. Figure 4 shows the quantile slope estimates from fused adaptive lasso, fused adaptive sup-norm and RQ, respectively, across the 9 quantile levels for 4 covariates out of the 13. In each plot, the shaded area is the 90% pointwise confidence band constructed from the inversion of rank score test described in Koenker (2005). The points connected by solid lines correspond to the estimated quantile coefficients from RQ, while the points connected by dashed lines are fused adaptive lasso estimates, and the stars with dashed lines are fused adaptive sup-norm estimates. The fused adaptive sup-norm shrinks the slope coefficients of mhe, $ivst$

and *blakp* to be constant over τ, but vary with τ for *pol*, while fused adaptive lasso tends to shrink neighboring quantile coefficients to be equal, resulting in piecewise constant quantile coefficients. In contrast, the solid lines (RQ) are more variable compared to the dashed lines (fused adaptive lasso and fused adaptive sup-norm), since RQ can not make any shrinkage.

To further verify the shrinkage results, we conduct hypothesis tests to check the constancy of slope coefficients by using the R function "anova.rqlist" in the *quantreg* package. This function is based on the Wald-type test described in Koenker (2005) and can be used to test if the slope coefficients are identical at some pre-specified quantiles. Take the covariate *ivst* for an example, the Wald-type test for equality of quantile coefficients at $\tau = 0.1, \ldots, 0.9$ returns a p-value of 0.9324, implying that the effect of *ivst* does not vary significantly across the nine quantile levels. This agrees with the results from fused adaptive lasso and fused adaptive sup-norm, where the nine quantile coefficients are shrunk to be a constant. On the other hand, for the covariate *pol*, the equality test on the quantile coefficients at $\tau = 0.1, \ldots, 0.9$ returns a p-value of 0.000996, showing that the effect of *pol* varies across the 9 quantile levels. The equality test at $\tau = 0.5, \ldots, 0.9$ also shows the significant difference. However, for the equality test at $\tau = 0.1, \ldots, 0.5$, the null hypothesis failed to be rejected. This agrees with the results from fused adaptive lasso, which shrinks the first 5 quantile coefficients to be a constant, but keeps the upper quantile coefficients vary.

To evaluate the prediction accuracy of the different methods, we performed a cross validation study. The data are randomly split into a testing set with 50 observations and a training set with the remaining 111 observations. For each method, we estimate the quantile coefficients based on the training set, denoted as $\hat{\boldsymbol{\beta}}(\tau_j), j = 1, \ldots, 9$ and predict the τ_j^{th} conditional quantile of the GDP growth on the testing set. Prediction error (PE), used to assess the prediction accuracy, is defined as $\text{PE} = \sum_{k=1}^{9} \sum_{j=1}^{50} \rho_{\tau_k}\{y_j - \mathbf{x}_j^T \hat{\boldsymbol{\beta}}(\tau_k)\}$, where $\{(y_j, \mathbf{x}_j), j = 1, \ldots, 50\}$ are in the test set. We repeat the cross validation 200 times and take the average of PE. For fused adaptive lasso, the mean PE is 252.96 (s.e=1.66), while it is 252.73 (s.e.=1.97) for fused adaptive sup-norm and 258.94 (s.e.=1.65) for RQ. It is clear that both proposed penalization approaches yield higher prediction accuracy than RQ.

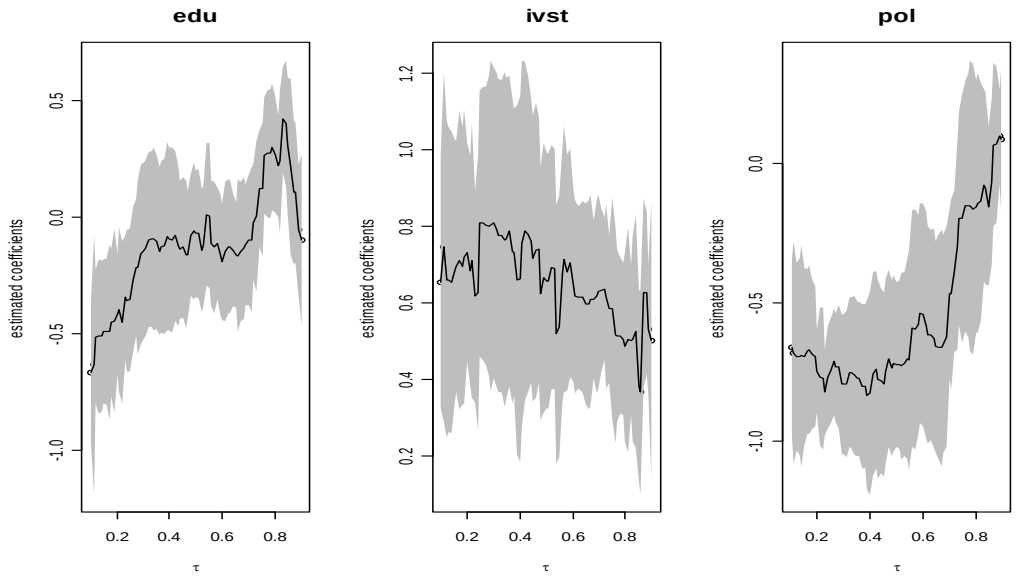

Figure 3: The effects of *edu* (the ratio of education and GDP growth), *ivst* (the ratio of investment and GDP growth) and *pol* (political instability) on different quantiles of the GDP growth.

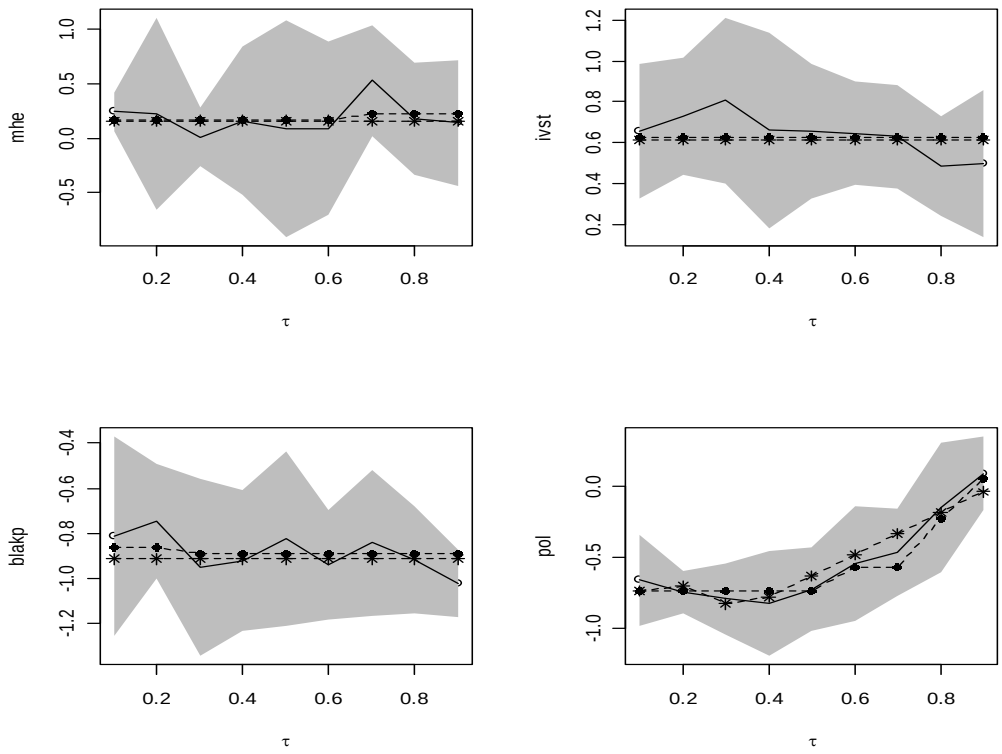

Figure 4: Estimated quantile coefficients for the four covariates *mhe* (male higher education), *ivst* (the ratio of investment and GDP growth), *blakp* (black market premium) and *pol* (political instability) from RQ (solid line), FAL (dashed line with dots), FAS (dashed line with stars) methods. Shaded areas are the 90% confidence bands from regular RQ method.

Supplemental Materials

R Code: R code to perform the method in this article along with a help file is supplied. (codeQR.zip)

Appendix: The supplemental files include the Appendix which gives all proofs. (appendixQR.pdf)

6 Acknowledgement

The authors are grateful to the editor, an associate editor, and two anonymous referees for their valuable comments. Wang's research was supported in part by NSF grant DMS-1007420. Bondell's research was supported in part by NSF grant DMS-1005612 and NIH grant P01-CA-142538.

References

Akaike, H. (1974), New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control* **19**, 716-723.

Barro, R. and Lee, J. (1994), Data Set for a Panel of 138 Countries. Cambridge, Mass.: National Bureau of Economic Research.

Bondell, H., Reich, B. and Wang, H. (2010), Non-crossing Quantile Regression Curve Estimation. *Biometrika* **97**, 825-838.

Fan, J. and Li, R. (2001), Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties. *Journal of the American Statistical Association* **96**, 1348-1360.

Geyer, C. (1994), On the Asymptotics of Constrained M-estimation. *Annals of Statistics* **22**, 1993-2010.

He, X. (1997), Quantile Curves Without Crossing. *The American Statistician* **51**, 186-192.

Kato, K. (2010), Solving l_1 regularization problems with piecewise linear losses. *Journal of Computational and Graphical Statistics* **19**, 1024-1040.

Koenker, R. and Bassett, G. (1978), Regression Quantiles. *Econometrica* **4**, 33-50.

Koenker, R. (2004), Quantile Regression for Longitudinal Data. *Journal of Multivariate Analysis* **91**, 74-89.

Koenker, R. (2005), Quantile Regression. Cambridge: Cambridge University Press.

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

Li, Y. and Zhu, J. (2007), Analysis of Array CGH Data for Cancer Studies Using Fused Quantile Regression. *Bioinformatics* **23**, 2470-2476.

Li, Y. and Zhu, J. (2008), L1-norm Quantile Regression. *Journal of Computational and Graphical Statistics* **17**, 163-185.

Osborne, M.R. and Turlach, B.A (2011), A homotopy algorithm for the quantile regression lasso and related piecewise linear problems. *Journal of Computational and Graphical Statistics* **20**, 972-987.

Tackeuchi, I., Le, Q., Sears, T. and Smola, A. (2006), Nonparametric Quantile Estimation. *Jounal of Machine Learning Research* **7**, 1231-1264.

Tibshirani, R. (1996), Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, series B* **58**, 267-288.

Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., and Knight, K. (2005), Sparsity and Smoothness via the Fused Lasso. *Journal of the Royal Statistical Society, series B* **67**, 91-108.

Wang, H. and Hu, J. (2010), Identification of differential aberrations in multiple-sample array CGH studies. *Biometrics* **67**, 353-362.

Wu, Y. and Liu, Y. (2009a), Variable selection in quantile regression. *Statistica Sinica* **19**, 801-817.

Wu, Y. and Liu, Y. (2009b), Stepwise Multiple Quantile Regression Estimation Using Non-crossing Constraints. *Statistica and Its Interface* **2**, 299-310.

Yuan, M. and Lin, Y. (2006), Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, series B* **68**, 49-67.

Zhao, P. and Rocha, G. and Yu, B. (2009), The composite absolute penalties family for grouped and hierarchical variable selection. *Annals of Statistics* **37**, 3468-3497.

Zou, H. (2006), The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* **101**, 1418-1429.

Zou, H. and Yuan, M. (2008a), Composite quantile regression and the oracle model selection theory. *The Annals of Statistics* **36**, 1108-1126.

Zou, H. and Yuan, M. (2008b), Regularized Simultaneous Model Selection in Multiple Quantiles Regression. *Computational Statistics and Data Analysis* **52**, 5296-5304.