Supplementary Materials for Advanced Algorithms for Penalized Quantile and Composite Quantile Regression

Matthew Pietrosanu · Jueyu Gao · Linglong Kong · Bei Jiang · Di Niu · Hengshuai Yao · Jui Shangling

A Quantile and Composite Quantile Regression Without Adaptive Lasso Regularization

This supplementary appendix is structured as follows. Section A.1 presents details of our approach to solving the standard quantile regression problem without regularization via alternating direction method of multipliers (ADMM), majorize-minimization (MM), and coordinate descent (CD) algorithms. For the sake of comparison, we also introduce a basic interior point (IP) approach. Section A.2 gives details on the generalization from quantile to composite quantile regression, again without regularization.

A.1 Non-Regularized Quantile Regression

The following Subsections A.1.1 through A.1.3 detail our approach to non-regularized quantile regression using the ADMM, MM, and CD algorithms. We place particular emphasis on

Drs. Linglong Kong, Bei Jiang, and Di Niu are supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC).

M. Pietrosanu, J. Gao, L. Kong, and B. Jiang
Department of Mathematical & Statistical Sciences, University of Alberta, Edmonton, AB, Canada
Tel.: +780-492-3396
Fax: +780-492-6826
E-mail: pietrosa@ualberta.ca, jueyu@ualberta.ca, lkong@ualberta.ca, bei1@ualberta.ca

D. Niu
Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada
Tel.: + 780-492-3332
E-mail: dniu@ualberta.ca

H. Yao
Huawei Hisilicon, Edmonton, AB, Canada
E-mail: hengshuai.yao@huawei.com

J. Shangling
Huawei Hisilicon, Shanghai, China
E-mail: jui.shangling@huawei.com

the ADMM approach and first review its general setup. Subsection A.1.4 introduces a basic IP method and a reformulation of the quantile regression problem accessible to the `Rmosek` optimization package for R (Friberg, 2013). We use the notation presented in the main text throughout.

A.1.1 Alternating Direction Method of Multipliers Algorithm

Before proceeding with an application to quantile regression, we review the general ADMM algorithm, which decomposes a given additively separable convex optimization problem into a number of sub-convex optimization problems. The general formulation of the ADMM problem is

$$\begin{aligned} \min_{\boldsymbol{x},\boldsymbol{z}} \quad & f(\boldsymbol{x}) + g(\boldsymbol{z}) \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} = \boldsymbol{c}, \end{aligned}$$

where f and g are convex, real-valued functions of $\boldsymbol{x}$ and $\boldsymbol{z}$, $\boldsymbol{A}$ and $\boldsymbol{B}$ are matrices, and $\boldsymbol{c}$ is a constant vector. The augmented Lagrangian (Powell, 1967) of the above problem is written as

$$L_\rho(\boldsymbol{x},\boldsymbol{z},\boldsymbol{y}) = f(\boldsymbol{x}) + g(\boldsymbol{z}) + \boldsymbol{y}^T(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} - \boldsymbol{c}) + \frac{\rho}{2}||\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} - \boldsymbol{c}||_2^2,$$

where ρ is a tuning parameter. Setting $u = \frac{1}{\rho}y$ and $u_k = \frac{1}{\rho}y_k$, we can obtain the (more convenient) scaled augmented Lagrangian

$$L_\rho^{\mathrm{s}}(\boldsymbol{x},\boldsymbol{z},\boldsymbol{y}) = f(\boldsymbol{x}) + g(\boldsymbol{z}) + \frac{\rho}{2}||\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z} - \boldsymbol{c} + \boldsymbol{u}||_2^2 - \frac{\rho}{2}\boldsymbol{u}_2.$$

The ADMM method optimizes the scaled augmented Lagrangian using the iterative scheme

$$\begin{aligned} \boldsymbol{x}^{(t+1)} &= \arg\min_{\boldsymbol{x}}\Big[f(\boldsymbol{x}) + \frac{\rho}{2}||\boldsymbol{A}\boldsymbol{x} + \boldsymbol{B}\boldsymbol{z}^{(t)} - \boldsymbol{c} + \boldsymbol{u}^{(t)}||_2^2\Big], \\ \boldsymbol{z}^{(t+1)} &= \arg\min_{\boldsymbol{z}}\Big[g(\boldsymbol{z}) + \frac{\rho}{2}||\boldsymbol{A}\boldsymbol{x}^{(t+1)} + \boldsymbol{B}\boldsymbol{z} - \boldsymbol{c} + \boldsymbol{u}^{(t)}||_2^2\Big], \\ \boldsymbol{u}^{(t+1)} &= \boldsymbol{u}^{(t)} + \boldsymbol{A}\boldsymbol{x}^{(t+1)} + \boldsymbol{B}\boldsymbol{z}^{(t+1)} - \boldsymbol{c}. \end{aligned}$$

A generic stopping condition for the algorithm can be defined in terms of the primal and dual residuals, given by $\boldsymbol{r}_{\mathrm{primal}}^{(t+1)} = \boldsymbol{A}\boldsymbol{x}^{(t+1)} + \boldsymbol{B}\boldsymbol{z}^{(t+1)} - c$ and $\boldsymbol{r}_{\mathrm{dual}}^{(t+1)} = \rho\boldsymbol{A}^T\boldsymbol{B}(\boldsymbol{z}^{(t+1)} - \boldsymbol{z}^{(t)})$. The program can be made to terminate if both

$$\begin{aligned} ||\boldsymbol{r}_{\mathrm{primal}}^{(t)}||_2 &\le \varepsilon_{\mathrm{primal}} = \sqrt{p}\varepsilon_{\mathrm{abs}} + \varepsilon_{\mathrm{rel}}\max\{||\boldsymbol{A}\boldsymbol{x}^{(t)}||_2, ||\boldsymbol{B}\boldsymbol{z}^{(t)}||_2, ||\boldsymbol{c}||_2\} \\ ||\boldsymbol{r}_{\mathrm{dual}}^{(t)}||_2 &\le \varepsilon_{\mathrm{dual}} = \sqrt{n}\varepsilon_{\mathrm{abs}} + \varepsilon_{\mathrm{rel}}||\boldsymbol{A}^T\boldsymbol{y}^{(t)}||_2, \end{aligned}$$

where p and n are the length of c and $\boldsymbol{A}^T\boldsymbol{y}^{(t)}$, respectively. In our applications, we set $\varepsilon_{\mathrm{abs}} = 10^{-2}$ and $\varepsilon_{\mathrm{rel}} = 10^{-4}$.

We apply the ADMM algorithm (Boyd et al., 2011) by reformulating quantile regression as the convex optimization problem

$$\begin{aligned} \min_{\boldsymbol{\beta}\in\mathbb{R}^{p+1}} \quad & \sum_{i=1}^n \rho_\tau(r_i) \\ \text{subject to} \quad & \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{r} = \boldsymbol{Y}, \end{aligned}$$

where $\boldsymbol{r}$ is a vector of residuals. The intercept term is accounted for in both $\boldsymbol{\beta}$ and $\boldsymbol{X}$. Using the general procedure of ADMM, taking $f = 0$ and g as a function of $\boldsymbol{r}$ to be the entire objective function, we obtain the iterative scheme

$$\begin{aligned}
\boldsymbol{\beta}^{(t+1)} &= \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+1}}{\arg\min}\ \frac{\rho}{2}||\boldsymbol{Y}-\boldsymbol{r}^{(t)}-\boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{u}^{(t)}/\rho||_2^2,\\
\boldsymbol{r}^{(t+1)} &= \underset{\boldsymbol{r}\in\mathbb{R}^{n}}{\arg\min}\ \sum_{i=1}^{n}\rho_\tau(r_i)+\frac{\rho}{2}||\boldsymbol{Y}-\boldsymbol{r}-\boldsymbol{X}\boldsymbol{\beta}^{(t+1)}+\boldsymbol{u}^{(t)}/\rho||_2^2,\\
\boldsymbol{u}^{(t+1)} &= \boldsymbol{u}^{(t)}+\rho(\boldsymbol{Y}-\boldsymbol{r}^{(t+1)}-\boldsymbol{X}\boldsymbol{\beta}^{(t+1)}),
\end{aligned}$$

where $\boldsymbol{u}$ is the rescaled Lagrange multiplier and $\rho > 0$ is a penalty parameter. The update for $\boldsymbol{r}$ can be written in a closed form as $S_{1/\rho}\big(\boldsymbol{c}-(2\boldsymbol{\tau}_{n\times 1}-\mathbf{1}_{n\times 1})/\rho\big)$, where $\boldsymbol{c}=\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}^{(t)}+\boldsymbol{u}^{(t)}/\rho$ and, for real a, the function $S_a:\mathbb{R}^m\to\mathbb{R}^m$ is defined component-wise via $(S_a(\boldsymbol{v}))_i=(v_i-a)_+-(-v_i-a)_+$. The closed form for the update of $\boldsymbol{\beta}$ is given by $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{Y}-\boldsymbol{r}^{(t)}+\boldsymbol{u}^{(t)}/\rho)$. For reference, ρ is chosen to be 1.2 by Boyd et al. (2011). In the quantile regression setting, we have that

$$\begin{aligned}
\boldsymbol{r}_{\text{primal}}^{(t+1)} &= \boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}^{(t+1)}-\boldsymbol{r}^{(t+1)},\\
\boldsymbol{r}_{\text{dual}}^{(t+1)} &= \rho\boldsymbol{X}^T(\boldsymbol{r}^{(t+1)}-\boldsymbol{r}^{(t)}),\\
\varepsilon_{\text{primal}} &= \sqrt{n}\varepsilon_{\text{abs}}+\varepsilon_{\text{rel}}\max\{||\boldsymbol{X}\beta^{(t+1)}||_2^2,||\boldsymbol{r}^{(t+1)}||_2^2,||\boldsymbol{Y}||_2^2\},\\
\varepsilon_{\text{dual}} &= \sqrt{p}\varepsilon_{\text{abs}}+\varepsilon_{\text{rel}}||\boldsymbol{X}^T\boldsymbol{u}^{(t+1)}||_2^2.
\end{aligned}$$

A.1.2 Majorize-Minimization Algorithm

We use the MM algorithm developed by Hunter and Lange (2000) and Hunter and Li (2005) to solve the quantile regression problem without regularization. Our approach is exactly the same as in the main text, but we instead ignore the majorization of the penalty term in the quantile regression objective function. Construct a function $\rho_\tau^\varepsilon(r)$ based on some perturbation parameter $\varepsilon > 0$ that will be used to approximate the quantile regression objective function $L(\boldsymbol{\beta})$. For any residual r, define $\rho_\tau^\varepsilon(r)=\rho_\tau(r)-\frac{\varepsilon}{2}\ln(\varepsilon+|r|)$, and the subsequent approximation of $L(\boldsymbol{\beta})$ by $L^\varepsilon(\boldsymbol{\beta})=\sum_{i=1}^n\rho_\tau^\varepsilon(r_i)$. At the t-th iteration of the algorithm, for each current residual value $r_i^{(t)}=r_i^{(t)}(\boldsymbol{\beta}^{(t)})$, $\rho_\tau^\varepsilon(r)$ is majorized by the quadratic function

$$\xi_\tau^\varepsilon(r|r_i^{(t)})=\frac{1}{4}\left[\frac{r^2}{\varepsilon+|r_i^{(t)}|}+(4\tau-2)r+c\right],$$

for some solvable constant c that satisfies the equation $\xi(r_i^{(t)}|r^{(t)})=\rho_\tau^\varepsilon(r^{(t)})$. The MM algorithm minimizes the majorizer of $L^\varepsilon(\boldsymbol{\beta})$, namely,

$$Q^\varepsilon(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)})=\sum_{i=1}^n\xi_\tau^\varepsilon(r_i|r_i^{(t)}),$$

with the argument minimum taken as the updated value $\boldsymbol{\beta}^{(t+1)}$ of $\boldsymbol{\beta}$. For the t-th iteration of the algorithm, given an updated value $\boldsymbol{\beta}^{(t)}$ for $\boldsymbol{\beta}$, we generate and minimize a new majorized quadratic function $Q^\varepsilon(\cdot|\boldsymbol{\beta}^{(t)})$ and implement a Newton-Raphson iterative method to obtain an updated value $\boldsymbol{\beta}^{(t+1)}$ for $\boldsymbol{\beta}$.

A.1.3 Coordinate Descent Algorithm

To implement quantile regression, we use an extended version of the greedy CD method put forward by Edgeworth and, more recently, further developed by Wu and Lange (2008). In each iteration, for fixed $\boldsymbol{\beta}\in\mathbb{R}^p$, replace b_0 by the τ-th sample quantile of the residuals $y_i-\boldsymbol{X}_i^T\boldsymbol{\beta}$ for $i=1,\ldots,n$: this will necessarily decrease the value of the objective function.

Define $\Theta_i = \rho_\tau(r_i)$ for $i = 1, \ldots, n$. For each element β_m for $m = 1, \ldots, p$ of $\boldsymbol{\beta}$, rewrite the loss function as

$$L(b_0, \boldsymbol{\beta}) = L_m(b_0, \boldsymbol{\beta}) = \sum_{i=1}^{n} |x_{im}| \left| \frac{y_i - b_0 - \sum_{j=1,\, j\neq m}^{p} x_{ij}\beta_j}{x_{im}} - \beta_m \right| \cdot \Theta_i,$$

so that the CD algorithm applies. For each fixed m, sort the values of

$$z_i = \frac{y_i - b_0 - \sum_{j=1,\, j\neq m}^{p} x_{ij}\beta_j}{x_{im}}$$

for $i = 1, \ldots, n$ and update β_m to be the i^*-th order statistic $z_{(i^*)}$ satisfying both

$$\sum_{j=1}^{i^*-1} w_{(j)} < \frac{1}{2}\sum_{j=1}^{n} w_{(j)} \qquad \text{and} \qquad \sum_{j=1}^{i^*} w_{(j)} \geq \frac{1}{2}\sum_{j=1}^{n} w_{(j)},$$

where $w_i = |x_{im}| \cdot \Theta_i$. In other words, using the weights w_i, the selected $z_{(i^*)}$ is the weighted median of all z_i (for the fixed value of m). At the end of each iteration, we check for the convergence of $\boldsymbol{\beta}$ and stop the algorithm using an absolute value difference threshold of 10^{-3}.

A.1.4 Interior Point Algorithm

Interior point (IP) methods generally reach an optimal solution by travelling within rather than on the boundary of the feasible set. Though studied as early as the 1950s and 1960s, IP methods arguably first gained widespread interest with the landmark paper by Karmarkar (1984), who proposed an efficient, polynomial time IP algorithm for linear programs with performance rivalling the existing simplex method. Nesterov and Nemirovskii (1994) later extended these results to a range of convex optimization problems while maintaining polynomial time. In the present day, advanced IP methods and code for both linear and non-linear programs are widely available and well-studied in the literature (Roos et al., 2006). IP algorithms have also received considerable attention and success in applications to non-linear, non-convex optimization problems (Byrd et al., 1999).

We can implement quantile regression using an IP algorithm by reformulating the optimization problem as a linear program and making use of existing optimization packages such as `Rmosek` (Friberg, 2013). `Rmosek` can implement an IP algorithm to solve problems of the form

$$\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n} \quad & \boldsymbol{c}^T\boldsymbol{x} + c_0 \\ \text{subject to} \quad & \boldsymbol{l}^c \leq \boldsymbol{A}\boldsymbol{x} \leq \boldsymbol{u}^c \\ & \boldsymbol{l}^x \leq x \leq \boldsymbol{u}^x, \end{aligned}$$

where $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ is a constraint matrix; $\boldsymbol{c} \in \mathbb{R}^n$ and $c_0 \in \mathbb{R}$ the objective function coefficients and constant; $\boldsymbol{l}^c, \boldsymbol{u}^c \in \mathbb{R}^m$ the lower and upper constraint bounds; and $\boldsymbol{l}^x, \boldsymbol{u}^x \in \mathbb{R}^n$ the lower and upper variable bounds. For notational simplicity, $\leq$ is taken to mean component-wise comparison of vectors. Alternatively, other R packages such as `quantreg` exist specifically for quantile regression and make use of IP methods. The IP approach for quantile regression in `quantreg` is based on the method of Portnoy and Koenker (1997), with recent modifications including the prediction-correction algorithm of Mehrotra (1992). Lasso penalized quantile regression in `quantreg` uses a Frisch-Newton method.

Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n_{\geq 0}$ be a vector of the positive and negative parts, respectively, of the residuals $\boldsymbol{r} = (r_1, \ldots, r_n)$, and $\boldsymbol{\beta} \in \mathbb{R}^{p+1}$ a vector of parameters including the intercept. The quantile regression problem without regularization can be formulated for use in existing IP optimization routines such as `Rmosek` via

$$\begin{aligned}
\min_{\boldsymbol{\beta}\in\mathbb{R}^{p+1},\ \boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^n} \quad & \tau\mathbf{1}_{n\times1}^T\boldsymbol{u} + (1-\tau)\mathbf{1}_{n\times1}^T\boldsymbol{v} \\
\text{subject to} \quad & \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u} - \boldsymbol{v} \\
& \mathbf{0}_{n\times1} \le \boldsymbol{u} \le \boldsymbol{\infty}_{n\times1} \\
& \mathbf{0}_{n\times1} \le \boldsymbol{v} \le \boldsymbol{\infty}_{n\times1}.
\end{aligned}$$

As an aside, to incorporate an adaptive lasso penalty into the problem, we can rewrite the problem as a linear program accessible to existing IP routines via

$$\begin{aligned}
\min_{\boldsymbol{\beta}\in\mathbb{R}^{p+1},\ \boldsymbol{u},\boldsymbol{v}\in\mathbb{R}^n} \quad & \tau\mathbf{1}_{n\times1}^T\boldsymbol{u} + (1-\tau)\mathbf{1}_{n\times1}^T\boldsymbol{v} + p_\lambda(|\boldsymbol{\beta}|) \\
\text{subject to} \quad & \boldsymbol{\beta} \le \boldsymbol{\beta}^* \\
& -\boldsymbol{\beta} \le \boldsymbol{\beta}^* \\
& \mathbf{0}_{n\times1} \le \boldsymbol{\beta}^* \le \boldsymbol{\infty}_{n\times1} \\
& \mathbf{0}_{n\times1} \le \boldsymbol{u} \le \boldsymbol{\infty}_{n\times1} \\
& \mathbf{0}_{n\times1} \le \boldsymbol{v} \le \boldsymbol{\infty}_{n\times1}.
\end{aligned}$$

A.2 Composite Quantile Regression

This section shows details of the extension from quantile to composite quantile regression without regularization. Subsections A.2.1, A.2.2, and A.2.3 extend the above non-regularized quantile regression procedures using ADMM, MM, and CD algorithms, respectively. Subsection A.2.4 formulates the problem for use in `Rmosek` (Friberg, 2013) or other IP methods for linear programs. We use the notation presented in the main text throughout.

A.2.1 Alternating Direction Method of Multipliers Algorithm

Written in the ADMM form, the composite quantile regression problem can be expressed as

$$\begin{aligned}
\min_{\beta\in\mathbb{R}^{p+1}} \quad & \sum_{k=1}^{K}\sum_{i=1}^{n}\rho_{\tau_k}(r_{ik}) \\
\text{subject to} \quad & \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{r} = \boldsymbol{Y}^*,
\end{aligned}$$

where we assume that the intercept term is accounted for in both $\boldsymbol{\beta}$ and $\boldsymbol{X}$. The ADMM approach is applied in exactly the same way as in Subsection A.1.1, yielding the iterative update scheme

$$\begin{aligned}
\boldsymbol{\beta}^{(t+1)} &= \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^{p+K}} \ \frac{\rho}{2}||\boldsymbol{Y}^* - \boldsymbol{r}^{(t)} - \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{u}^{(t)}/\rho||_2^2, \\
\boldsymbol{r}^{(t+1)} &= \arg\min_{\boldsymbol{r}\in\mathbb{R}^{nK}} \ \sum_{k=1}^{K}\sum_{i=1}^{n}\rho_{\tau_k}(r_{ik}) + \frac{\rho}{2}||\boldsymbol{Y}^* - \boldsymbol{r} - \boldsymbol{X}^*\boldsymbol{\beta}^{(t+1)} + \boldsymbol{u}^{(t)}/\rho||_2^2, \\
\boldsymbol{u}^{(t+1)} &= \boldsymbol{u}^{(t)} + \rho(\boldsymbol{Y}^* - \boldsymbol{r}^{(t+1)} - \boldsymbol{X}^*\boldsymbol{\beta}^{(t+1)}),
\end{aligned}$$

where $\boldsymbol{c} = \boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta}^{(t)} + \boldsymbol{u}^{(t)}/\rho$; and residuals

$$\begin{aligned}
\boldsymbol{r}_{\text{primal}}^{(t+1)} &= \boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta}^{(t+1)} - \boldsymbol{r}^{(t+1)}, \\
\boldsymbol{r}_{\text{dual}}^{(t+1)} &= \rho\boldsymbol{X}^{*T}(\boldsymbol{r}^{(t+1)} - \boldsymbol{r}^{(t)}), \\
\varepsilon_{\text{primal}} &= \sqrt{n}\varepsilon_{\text{abs}} + \varepsilon_{\text{rel}}\max\{||\boldsymbol{X}^*\boldsymbol{\beta}^{(t+1)}||_2^2, ||\boldsymbol{r}^{(t+1)}||_2^2, ||\boldsymbol{Y}^*||_2^2\}, \\
\varepsilon_{\text{dual}} &= \sqrt{p}\varepsilon_{\text{abs}} + \varepsilon_{\text{rel}}||\boldsymbol{X}^{*T}\boldsymbol{u}^{(t+1)}||_2^2.
\end{aligned}$$

A generic stopping condition requiring $||\boldsymbol{r}^{(t)}_{\text{primal}}|| \leq \varepsilon_{\text{primal}}$ and $||\boldsymbol{r}^{(t)}_{\text{dual}}|| \leq \varepsilon_{\text{dual}}$ for termination can be imposed. We again take $\boldsymbol{u}$ as the rescaled Lagrange multiplier and $\rho > 0$ as a penalty parameter. Generalizing from quantile regression, the update for $\boldsymbol{r}$ can be written in a closed form as $S_{1/\rho}\big(\boldsymbol{c} - (2\boldsymbol{\tau}^* - \boldsymbol{1}_{n\times 1})/\rho\big)$, with S_a as defined previously for real a. The closed form update for $\boldsymbol{\beta}$ is given by $(\boldsymbol{X}^{*T}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*T}(\boldsymbol{Y}^* - \boldsymbol{r}^{(t)} + \boldsymbol{u}^{(t)}/\rho)$.

A.2.2 Majorize-Minimization Algorithm

An extension of the MM algorithm from quantile to composite quantile regression simply involves the incorporation of additional quantile levels. We use the same function $\rho^\varepsilon_\tau(r) = \rho_\tau(r) - \frac{\varepsilon}{2}\ln(\varepsilon + |r|)$ to approximate the composite quantile regression objective function via $L^\varepsilon(\boldsymbol{\beta}) = \sum_{k=1}^K \sum_{i=1}^n \rho^\varepsilon_{\tau_k}(r_{ik})$. We also use the same function ξ as defined in Subsection A.1.2 to majorize ρ^ε_τ. At the t-th iteration of the algorithm, for each current residual value $r^{(t)}_{ik} = r^{(t)}_{ik}(\boldsymbol{\beta}^{(t)})$, we have that $\rho^\varepsilon_{\tau_k}(r)$ is majorized by the quadratic function

$$\xi^\varepsilon_{\tau_k}(r|r^{(t)}_{ik}) = \frac{1}{4}\left[\frac{r^2}{\varepsilon + |r^{(t)}_{ik}|} + (4\tau_k - 2)r + c\right],$$

for some solvable constant c that satisfies the equation $\xi(r^{(t)}_{ik}|r^{(t)}_{ik}) = \rho^\varepsilon_{\tau_k}(r^{(t)}_{ik})$. The MM algorithm minimizes the majorizer of $L^\varepsilon(\boldsymbol{\beta})$, namely,

$$Q^\varepsilon(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}) = \sum_{k=1}^K \sum_{i=1}^n \xi^\varepsilon_{\tau_k}(r_{ik}|r^{(t)}_{ik}),$$

with the argument minimum taken as the updated value $\boldsymbol{\beta}^{(t+1)}$ of $\boldsymbol{\beta}$. In practice, for the t-th iteration of the algorithm, given an updated value $\boldsymbol{\beta}^{(t)}$ for $\boldsymbol{\beta}$, we generate and minimize a new majorized quadratic function $Q^\varepsilon(\cdot|\boldsymbol{\beta}^{(t)})$ using a Newton-Raphson iterative method. The argument minimum is taken as the updated value $\boldsymbol{\beta}^{(t+1)}$ for $\boldsymbol{\beta}$.

A.2.3 Coordinate Descent Algorithm

To apply the CD method to composite quantile regression, we rewrite the composite quantile regression objective function in the required CD form. For any $m = 1, \ldots, p$, we have

$$L_m(b_1, \ldots, b_k, \boldsymbol{\beta}) = \sum_{k=1}^K \sum_{i=1}^n |x_{im}| \left| \frac{y_i - b_k - \sum_{j=1,\, j\neq m}^p x_{ij}\beta_j}{x_{im}} - \beta_m \right| \cdot \Theta_{ik},$$

with $\Theta_{ik} = \rho_{\tau_k}(r_{ik})$ for $i = 1, \ldots, n$ and $k = 1, \ldots, K$. In each iteration, and for fixed $\boldsymbol{\beta}$, replace b_k, for $k = 1, \ldots, K$, with the τ-th sample quantile of the residuals $y_i - \boldsymbol{X}_i^T\boldsymbol{\beta}$ for $i = 1, \ldots, n$. To update β_m for $m = 1, \ldots, p$, sort the numbers

$$z_{ik} = \frac{y_i - b_k - \sum_{j=1,\, j\neq m}^p x_{ij}\beta_j}{x_{im}},$$

for $i = 1, \ldots, n$ and $k = 1, \ldots, K$. Update β_m with the value of the i^*-th order statistic $z_{(i^*)}$ satisfying both

$$\sum_{j=1}^{i^*-1} w_{(j)} < \frac{1}{2}\sum_{j=1}^{nK} w_{(j)} \qquad \text{and} \qquad \sum_{j=1}^{i^*} w_{(j)} \geq \frac{1}{2}\sum_{j=1}^{nK} w_{(j)},$$

where $w_{ik} = |x_{im}| \cdot \Theta_{ik}$. At the end of each iteration, we check for the convergence of $\boldsymbol{\beta}$ and stop the algorithm using an absolute value difference threshold of 10^{-3}.

A.2.4 Interior Point Algorithm

The extension of the previous IP method from quantile to composite quantile regression simply requires us to account for the extra quantile levels in the objective function and the resulting extra residuals. The problem can be formulated as a linear program via

$$
\begin{aligned}
\min_{\boldsymbol{\beta}\in\mathbb{R}^{p+K},\ \boldsymbol{u}_k,\boldsymbol{v}_k\in\mathbb{R}^n} \quad & \sum_{k=1}^{K} \tau_k \mathbf{1}_{n\times 1}^T \boldsymbol{u}_k + (1-\tau_k)\mathbf{1}_{n\times 1}^T \boldsymbol{v}_k \\
\text{subject to} \quad & \boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{u}_k - \boldsymbol{v}_k \\
& \mathbf{0} \le \boldsymbol{u}_k \le \boldsymbol{\infty} \\
& \mathbf{0} \le \boldsymbol{v}_k \le \boldsymbol{\infty},
\end{aligned}
$$

where each constraint is to hold for all $k = 1, \ldots, K$.

B Composite Quantile Regression with Adaptive Lasso Regularization

Here we give explicit details regarding the ADMM, MM, and CD methods for composite quantile regression with adaptive lasso regularization. An IP approach is also given for comparison.

B.1 Alternating Direction Method of Multipliers Algorithm

Applying ADMM in the composite quantile setting with adaptive lasso regularization, we obtain the iterative update scheme

$$
\begin{aligned}
\boldsymbol{r}^{(t+1)} &= \underset{\boldsymbol{r}\in\mathbb{R}^{nK}}{\arg\min} \ \sum_{k=1}^{K}\sum_{i=1}^{n} \rho_{\tau_k}(r_{ik}) + \frac{\rho}{2}||\boldsymbol{Y}^* - \boldsymbol{r} - \boldsymbol{X}^*\boldsymbol{\beta}^{(t)} + \boldsymbol{u}^{(t)}/\rho||_2^2, \\
\boldsymbol{\beta}^{(t+1)} &= \underset{\boldsymbol{\beta}\in\mathbb{R}^{p+K}}{\arg\min} \ \frac{\rho}{2}||\boldsymbol{Y}^* - \boldsymbol{r}^{(t+1)} - \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{u}^{(t)}/\rho||_2^2 + \lambda \sum_{j=1}^{p} {}^{|\beta_j|}/_{|\beta_j^{\text{CQR}}|^2}, \\
\boldsymbol{u}^{(t+1)} &= \boldsymbol{u}^{(t)} + \rho(\boldsymbol{Y}^* - \boldsymbol{r}^{(t+1)} - \boldsymbol{X}^*\boldsymbol{\beta}^{(t+1)}),
\end{aligned}
$$

where $\boldsymbol{c} = \boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta}^{(t)} + \boldsymbol{u}^{(t)}/\rho$; and residuals

$$
\begin{aligned}
\boldsymbol{r}_{\text{primal}}^{(t+1)} &= \boldsymbol{Y}^* - \boldsymbol{X}^*\boldsymbol{\beta}^{(t+1)} - \boldsymbol{r}^{(t+1)}, \\
\boldsymbol{r}_{\text{dual}}^{(t+1)} &= \rho \boldsymbol{X}_*^{*T}(\boldsymbol{r}^{(t+1)} - \boldsymbol{r}^{(t)}), \\
\varepsilon_{\text{primal}} &= \sqrt{n}\varepsilon_{\text{abs}} + \varepsilon_{\text{rel}} \max\{||\boldsymbol{X}_*^*\boldsymbol{\beta}_*^{(t+1)}||_2^2, ||\boldsymbol{r}^{(t+1)}||_2^2, ||\boldsymbol{b}^* - \boldsymbol{Y}||_2^2\}, \\
\varepsilon_{\text{dual}} &= \sqrt{p}\varepsilon_{\text{abs}} + \varepsilon_{\text{rel}}||\boldsymbol{X}^{*T}\boldsymbol{u}^{(t+1)}||_2^2.
\end{aligned}
$$

We again take $\boldsymbol{u}$ as the rescaled Lagrange multiplier and $\rho > 0$ as a penalty parameter. As before, the update for $\boldsymbol{r}$ can be written in a closed form as $S_{1/\rho}\big(\boldsymbol{c} - (2\boldsymbol{\tau}^* - \mathbf{1}_{n\times 1})/\rho\big)$, with S_a as defined previously for real a. With adaptive lasso regularization, the update for $\boldsymbol{\beta}$ does not have a closed form but can be viewed as a least squares optimization problem with adaptive lasso penalty. We implement existing numerical methods to solve this problem and update $\boldsymbol{\beta}$.

B.2 Majorize-Minimization Algorithm

An extension of the MM method for adaptive lasso regularized quantile regression to regularized composite quantile regression involves a minor change to incorporate multiple quantile levels into the majorized objective function. Using the same function $\rho_\tau^\varepsilon(r) = \rho_\tau(r) - \frac{\varepsilon}{2}\ln(\varepsilon + |r|)$ as before with perturbation parameter $\varepsilon > 0$ to approximate $\rho_\tau(r)$, we can approximate the regularized quantile regression objective function via

$$\sum_{k=1}^{K}\sum_{i=1}^{n}\rho_{\tau_k}^{\varepsilon}(r_{ik}) + +\lambda\sum_{j=1}^{p}\frac{1}{|\beta_j^{\mathrm{CQR}}|^2}\left[|\beta_j^{(t)}| + \frac{\big(\beta_j^2 - (\beta_j^{(t)})^2\big)\mathrm{sgn}(\beta_j^{(t)})}{2|\beta_j^{(t)} + \varepsilon|}\right].$$

Define, as before,

$$\xi_{\tau_k}^{\varepsilon}(r|r_{ik}^{(t)}) = \frac{1}{4}\left[\frac{r^2}{\varepsilon + |r_{ik}^{(t)}|} + (4\tau_k - 2)r + c\right].$$

In the t-th iteration of the MM algorithm, the (approximated) objective function is majorized by

$$Q^{\varepsilon}(\boldsymbol{\beta}|\boldsymbol{\beta}^{(t)}) = \sum_{k=1}^{K}\sum_{i=1}^{n}\xi_{\tau_k}^{\varepsilon}(r_{ik}|r_{ik}^{(t)}) + \lambda\sum_{j=1}^{p}\frac{1}{|\beta_j^{\mathrm{CQR}}|^2}\left[|\beta_j^{(t)}| + \frac{\big(\beta_j^2 - (\beta_j^{(t)})^2\big)\mathrm{sgn}(\beta_j^{(t)})}{2|\beta_j^{(t)} + \varepsilon|}\right].$$

Given an updated value $\boldsymbol{\beta}^{(t)}$ for $\boldsymbol{\beta}$, we generate a new majorizing function $Q^{\varepsilon}(\cdot|\boldsymbol{\beta}^{(t)})$ and implement a Gauss-Newton iterative method to estimate and update the value of $\boldsymbol{\beta}$.

B.3 Coordinate Descent Algorithm

As discussed in the main text, the CD method for regularized composite quantile regression simply adjusts the objective function to account for the extra quantile levels as

$$L_m(b_1, \ldots, b_k, \boldsymbol{\beta}) = \sum_{k=1}^{K}\sum_{i=1}^{n}|x_{im}|\left|\frac{y_i - b_k - \sum_{j=1,\, j\neq m}^{p} x_{ij}\beta_j}{x_{im}} - \beta_m\right| \cdot \Theta_{ik} + p_\lambda(|\boldsymbol{\beta}|).$$

In each iteration, for $k = 1, \ldots, K$, replace each b_k with the τ-th sample quantile of the residuals $y_i - \boldsymbol{X}_i^T\boldsymbol{\beta}$ for $i = 1, \ldots, n$. Define $z_{ik} = \frac{1}{x_{im}}\big(y_i - b_k - \sum_{j=1,\, j\neq m}^{p} x_{ij}\beta_j\big)$ if $r_{ik} \geq 0$ and $z_{ik} = 0$ if $r_{ik} < 0$. Update β_m to the value of the i^*-th order statistic $z_{(i^*)}$ satisfying both

$$\sum_{j=1}^{i^*-1} w_{(j)} < \frac{1}{2}\sum_{j=1}^{nK} w_{(j)} \qquad \text{and} \qquad \sum_{j=1}^{i^*} w_{(j)} \geq \frac{1}{2}\sum_{j=1}^{nK} w_{(j)},$$

where $w_{ik} = |x_{im}| \cdot \Theta_{ik}$ if $r_{ik} \geq 0$ and $w_{ik} = \lambda/|\beta_m^{\mathrm{CQR}}|^2$ if $r_{ik} < 0$. At the end of each iteration, check for the convergence of $\boldsymbol{\beta}$ and stop the algorithm using an absolute value difference threshold of 10^{-3}.

B.4 Interior Point Algorithm

Adaptive lasso regularized composite quantile regression is formulated by incorporating an appropriate penalty term into the linear program of Subsection A.2.4. This form is appropriate for the IP implementation in the `Rmosek` package (Friberg, 2013) and is given by

$$
\begin{aligned}
\min_{\boldsymbol{\beta}\in\mathbb{R}^{p+K},\ \boldsymbol{u}_k,\boldsymbol{v}_k\in\mathbb{R}^n} \quad & \sum_{k=1}^{K} \tau_k \mathbf{1}_{n\times 1}^T \boldsymbol{u}_k + (1-\tau_k)\mathbf{1}_{n\times 1}^T \boldsymbol{v}_k + p_\lambda(|\boldsymbol{\beta}|) \\
\text{subject to} \quad & \boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}_k - \boldsymbol{v}_k \\
& \boldsymbol{\beta} \le \boldsymbol{\beta}^* \\
& -\boldsymbol{\beta} \le \boldsymbol{\beta}^* \\
& \mathbf{0} \le \boldsymbol{u}_k \le \boldsymbol{\infty} \\
& \mathbf{0} \le \boldsymbol{v}_k \le \boldsymbol{\infty},
\end{aligned}
$$

where constraints are to hold for all $k = 1, \ldots, K$.

Acknowledgements Jueyu Gao acknowledges the supervision of Drs. Linglong Kong and Edit Gombay during his graduate studies. The authors have no declarations of interest to declare.

References

Boyd S, Parikh N, Chu E, Eckstein J (2011) Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning 3(1):1–122

Byrd R, Hribar M, Nocedal J (1999) An interior point algorithm for large-scale nonlinear programming. SIAM Journal on Optimization 9(4):877–900

Friberg H (2013) Rmosek: the R-to-Mosek optimization interface. URL `https://cran.r-project.org/package=Rmosek`, R package version 1.2.5.1

Hunter D, Lange K (2000) Quantile regression via an MM algorithm. Journal of Computational and Graphical Statistics 9(1):60–77

Hunter D, Li R (2005) Variable selection using MM algorithms. Annals of Statistics 33(4):1617–1642

Karmarkar N (1984) A new polynomial-time algorithm for linear programming. Combinatorica 4(4):373–395

Mehrotra S (1992) On the implementation of a primal-dual interior point method. SIAM Journal on Optimization 2(4):575–601

Nesterov Y, Nemirovskii A (1994) Interior-point polynomial algorithms in convex programming. Society for Industrial and Applied Mathematics, Philadelphia, USA

Portnoy S, Koenker R (1997) The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. Statistical Science 12(4):279–300

Powell M (1967) A method for non-linear constraints in minimization problems. United Kingdom Atomic Energy Authority

Roos C, Terlaky T, Vial JP (2006) Interior point methods for linear optimization. Springer, New York, USA

Wu T, Lange K (2008) Coordinate descent algorithms for lasso penalized regression. Annals of Applied Statistics 2(1):224–244