Enhancing quantile estimation via quantile combination under heteroscedasticity

Suin Kim[1] and Yoonsuh Jung[1*]

[1]Department of Statistics, Korea University, 145 Anam-ro, Seoul, 02841, South Korea.

*Corresponding author(s). E-mail(s): yoons77@korea.ac.kr;
Contributing authors: ksi2002@korea.ac.kr;

Abstract

Quantile regression is a robust methodology for estimating conditional quantiles of a response variable, particularly in datasets with heteroscedasticity. This study proposes an approach to enhance quantile regression by a weighted combination of multiple quantile estimates. While composite quantile regression is a popular approach for integrating multiple quantile losses, previous studies have focused on estimating central tendencies under homoscedasticity. In contrast, our method targets a specific quantile under heteroscedastic conditions. By selecting suitable local quantiles to be combined and estimating their optimal weights, our method can be more efficient than using only a single quantile. We establish some theoretical properties of our estimator under a linear location-scale model and extend our work to a nonlinear model. Results from simulation studies and real-world data analysis indicate that the proposed method yields more robust and efficient estimates compared to the original quantile regression. Moreover, our approach effectively reduces quantile crossing, a significant issue in quantile estimation.

Keywords: Quantile regression, M-estimation, heteroscedasticity, robust estimation

1 Introduction

Regression models range from simple linear models to advanced machine learning methods and deep neural networks, most of which focus on estimating the conditional mean of the target variable. However, point estimation alone often fails to capture the full data distribution, and black-box models can hinder the assessment of estimate reliability. Quantile regression (QR) addresses these limitations by estimating the conditional quantiles of the target variable. This approach is particularly valuable when covariate effects vary across quantiles, as in heteroscedastic or skewed data.

Previous studies have explored integrating multiple quantiles to enhance quantile regression performance. These methods fall into two categories: (i) combining loss functions across quantiles (Zou and Yuan, 2008; Bradic et al, 2011; Xu et al, 2017; Kai et al, 2010; Jiang et al, 2018) and (ii) combining estimators from different quantiles (Portnoy and Koenker, 1989; Zhao and Xiao, 2014). Both strategies are effective for improving the performance compared to using single quantile. However, previous research has primarily focused on estimating the central tendency, restricting the combined quantiles to evenly spaced intervals within (0, 1) (e.g., $0.05, \ldots, 0.95$ or $0.1, \ldots, 0.9$). Most of the existing methods, including composite quantile regression models, combine quantiles assuming homoscedasticity. As the demand for predicting quantiles increases (e.g., for uncertainty quantification (Yin et al, 2023; Pouplin et al, 2024)), we propose methods which integrate quantiles under heteroscedasticity.

We demonstrate that the proposed method can be useful to estimate specific quantiles beyond central tendency. We refer to this as combining quantile regression. To highlight the novel contribution of our approach, we present the key components of this paper.

- **Quantile range selection**: We introduce an empirical rule for selecting the appropriate range of quantiles to combine, tailored to the target quantile.

- **Optimal weight determination**: We develop a framework to calculate optimal weights for combining quantiles, thereby improving estimation efficiency when targeting a specific quantile.
- **Theoretical validation**: We establish the asymptotic properties of our estimator under a linear location-scale model.

Although the theoretical development of our work is focused on linear location-scale models, the proposed idea of combining quantile estimators can be applied to nonlinear models. In particular, we consider equal-weighted combinations in nonlinear settings, and demonstrate their empirical benefits through simulations.

The remainder of the paper is organized as follows. In Section 2, we review quantile regression under heteroscedasticity. Section 3 introduces our estimator, its theoretical properties, and the derivation of optimal weights and quantile selection. In Section 4, we validate our method through extensive simulations, highlighting its advantages over single-quantile estimation. Finally, Section 5 applies our approach to a real-world dataset to demonstrate its practical effectiveness.

2 Overview of quantile regression under heteroscedasticity

Consider a linear location-scale model in which the response variable y is related to the predictor variables $\mathbf{x} \in \mathbb{R}^p$ by

$$y = \mathbf{x}^\top \boldsymbol{\beta} + \mathbf{x}^\top \boldsymbol{\gamma} \epsilon, \tag{1}$$

where ϵ is an independent and identically distributed ($i.i.d.$) error with distribution function F and density f. We assume that $\mathbf{x}$ and ϵ are independent and that $\mathbb{E}(\epsilon) = 0$. Under this model, the true conditional quantile function of y is $Q_y(\tau \mid \mathbf{x}) = \mathbf{x}^\top \boldsymbol{\beta} +$

$\mathbf{x}^\top \boldsymbol{\gamma}\xi(\tau)$ for any $\tau \in (0, 1)$, where $\xi(\tau) \equiv Q_\epsilon(\tau)$ denotes the τ-th quantile of ϵ. The parameter of interest is the slope at the τ-th quantile, $\boldsymbol{\beta}_\tau = \boldsymbol{\beta} + \boldsymbol{\gamma}\,\xi(\tau)$.

Unlike mean regression, which often minimizes a squared loss, QR minimizes the check loss function $\rho_\tau(z) = z(\tau - \mathbb{I}(z < 0))$, where $\mathbb{I}(\cdot)$ is the indicator function. Given independent observations $(\mathbf{x}_i, y_i)$ for $i = 1, \ldots, n$, the standard QR estimator is defined by

$$\hat{\boldsymbol{\beta}}_\tau = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \rho_\tau(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}). \tag{2}$$

This optimization yields a consistent estimator for the true conditional quantiles of y (Koenker and Bassett Jr, 1978).

In the presence of heteroscedasticity, weighting observations can improve the efficiency of quantile regression. Following Koenker (2005) and Koenker and Zhao (1994), the recommended weights are $1/\sigma_i$, where $\sigma_i = \mathbf{x}_i^\top \boldsymbol{\gamma}$ for $i = 1, \ldots, n$. Since $\boldsymbol{\gamma}$ is unknown, a preliminary estimator $\hat{\boldsymbol{\gamma}}$ is used to compute $\hat{\sigma}_i = \mathbf{x}_i^\top \hat{\boldsymbol{\gamma}}$. One approach to obtain $\hat{\boldsymbol{\gamma}}$ is to fit an ordinary least squares regression of $|y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}|$ on $\mathbf{x}_i$, where $\hat{\boldsymbol{\beta}}$ is any linear and consistent estimator of $\boldsymbol{\beta}$. The weighted quantile regression estimator for the τ-th quantile is then

$$\hat{\boldsymbol{\beta}}_\tau(\hat{\boldsymbol{\gamma}}) = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{n} \frac{1}{\hat{\sigma}_i} \rho_\tau(y_i - \mathbf{x}_i^\top \boldsymbol{\beta}). \tag{3}$$

To analyze the joint asymptotic properties of the weighted estimator (3), consider a set of quantile levels $0 < \tau_1 < \cdots < \tau_K < 1$ with their corresponding estimators $\{\hat{\boldsymbol{\beta}}_{\tau_k}(\hat{\boldsymbol{\gamma}})\}_{k=1}^{K}$. We assume the following conditions.

(A1) The distribution function F is absolutely continuous, with a positive density f uniformly bounded away from 0 and ∞ at $\xi(\tau_k)$, for $k = 1, \ldots, K$, where $K < M < \infty$ for some constant M.

(A2) There exists a positive definite matrix $\mathbf{Q}_j = \lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \sigma_i^{-j} \mathbf{x}_i \mathbf{x}_i^\top$ for $j = 0, 1, 2$.

(A3) $\max_i \|\mathbf{x}_i/\sigma_i\| = O(n^{1/4})$.

(A4) $\sum_{i=1}^{n} \|\mathbf{x}_i/\sigma_i\|^3 = O(n)$.

(A5) $\hat{\boldsymbol{\gamma}} = \kappa\boldsymbol{\gamma} + O_p(n^{-1/2})$ for some positive constant κ (i.e., consistency up to scale).

Assumptions (A1) and (A2) are standard in quantile regression analysis (Koenker, 2005, Section 4.2); (A1) ensures that the error distribution is smooth near the quantiles of interest, and (A2) guarantees that the weighted design matrices are well-behaved in the limit. Assumptions (A3) and (A4) prevent any single predictor from dominating the estimation, thus ensuring stability and robustness (Koenker and Zhao, 1994). Finally, (A5) requires that the preliminary estimator $\hat{\boldsymbol{\gamma}}$ is consistent up to a scaling constant—a sufficient condition since the scale of $\boldsymbol{\gamma}$ does not affect the minimization in the weighted quantile regression estimator.

Under assumptions (A1)–(A5), Theorem 2.1 in Koenker and Zhao (1994) shows the asymptotic behavior of $\hat{\boldsymbol{\beta}}_\tau(\hat{\boldsymbol{\gamma}})$:

$$\sqrt{n}\left(\begin{pmatrix}\hat{\boldsymbol{\beta}}_{\tau_1}(\hat{\boldsymbol{\gamma}})\\ \vdots \\ \hat{\boldsymbol{\beta}}_{\tau_K}(\hat{\boldsymbol{\gamma}})\end{pmatrix} - \begin{pmatrix}\boldsymbol{\beta}_{\tau_1}\\ \vdots \\ \boldsymbol{\beta}_{\tau_K}\end{pmatrix}\right) \xrightarrow{d} N\left(\mathbf{0}, \mathbf{H}\otimes\mathbf{Q}_2^{-1}\right), \tag{4}$$

where $\mathbf{H}$ is a $K \times K$ symmetric matrix with elements $h_{kk'} = \frac{\tau_k(1-\tau_{k'})}{f(\xi(\tau_k))f(\xi(\tau_{k'}))}$ for $k \le k'$, and $\otimes$ denotes the Kronecker product. This asymptotic result forms the theoretical foundation for our combining quantile estimator.

3 Combining quantile regression targeting a specific quantile

3.1 Definition

To improve robustness and efficiency of quantile regression under heteroscedasticity, we propose a method that combines estimates from multiple quantiles to target a specific quantile of interest. We call this approach combining quantile regression (CQR). The CQR estimator for a target quantile $\tau \in (0,1)$ is defined as a weighted sum of the quantile regression estimators in (3) at $\tau_1, \dots, \tau_K$:

$$\hat{\boldsymbol{\beta}}_\tau^{CQR} = \sum_{k=1}^{K} w_k \, \hat{\boldsymbol{\beta}}_{\tau_k}(\hat{\boldsymbol{\gamma}}), \quad \text{subject to } \mathbf{w}^\top \mathbf{1} = 1 \quad \text{and} \quad \mathbf{w}^\top \mathbf{v} = \xi(\tau). \tag{5}$$

Here, $\mathbf{w} = (w_1, \dots, w_K)^\top$ is the weight vector, $\mathbf{1} = (1, \dots, 1)^\top$, and $\mathbf{v} = (\xi(\tau_1), \dots, \xi(\tau_K))^\top$. Although $\hat{\boldsymbol{\beta}}_\tau^{CQR}$ depends on $\mathbf{w}$, we omit this dependence for notational simplicity if it does not cause confusion. This form of the estimator has been considered in Zhao and Xiao (2014) and Bloznelis et al (2019), but their focus was primarily on central tendency, i.e., they combined the quantile estimator to estimate a common slope $\boldsymbol{\beta}$. Our approach extends these methods to the case of quantile-specific slope $\boldsymbol{\beta}_\tau$ whose another look is the heteroscedastic model.

The constraints in (5) allow our estimator to be consistent for the target quantile regression coefficient. The constraint $\mathbf{w}^\top \mathbf{1} = 1$ ensures that the weights sum to one, while $\mathbf{w}^\top \mathbf{v} = \xi(\tau)$ forces the weighted combination of $\xi(\tau_k)$s to be equal to $\xi(\tau)$. Together, these constraints guarantee that the target of $\hat{\boldsymbol{\beta}}_\tau^{CQR}$ is $\boldsymbol{\beta} + \boldsymbol{\gamma}\,(\mathbf{w}^\top \mathbf{v}) = \boldsymbol{\beta} + \boldsymbol{\gamma}\,\xi(\tau) = \boldsymbol{\beta}_\tau$ and $\hat{\boldsymbol{\beta}}_\tau^{CQR}$ is a consistent estimator of $\boldsymbol{\beta}_\tau$ under the conditions (A1)–(A5). We then state the following lemma regarding the asymptotic behavior of $\hat{\boldsymbol{\beta}}_\tau^{CQR}$.

Lemma 1. *Suppose (A1)–(A5) hold, and assume that* $\mathbf{w}$ *satisfies the constraints* $\mathbf{w}^\top \mathbf{1} = 1$ *and* $\mathbf{w}^\top \mathbf{v} = \xi(\tau)$*. Then, the estimator* $\hat{\boldsymbol{\beta}}_\tau^{CQR}$ *is asymptotically normal.*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_\tau^{CQR} - \boldsymbol{\beta}_\tau) \xrightarrow{d} N\left(0, (\mathbf{w}^\top \mathbf{H} \mathbf{w}) \cdot \mathbf{Q}_2^{-1}\right). \tag{6}$$

This result follows directly from the asymptotic normality in (4).

Because the asymptotic variance of $\hat{\boldsymbol{\beta}}_{\tau}^{CQR}$ depends on the weights, we seek the optimal weight vector $\mathbf{w}^*$ that minimizes this variance subject to the constraints. The corresponding optimization problem is

$$\mathbf{w}^* = \arg\min_{\mathbf{w}\in\Omega} \mathbf{w}^\top \mathbf{H}\mathbf{w}, \tag{7}$$

where $\Omega = \{\, \mathbf{w} \in \mathbb{R}^K : \mathbf{w}^\top \mathbf{1} = 1 \text{ and } \mathbf{w}^\top \mathbf{v} = \xi(\tau) \,\}$. We next present the explicit solution for the optimal weights, extending Theorem 3 in Xu and Zhao (2022) to accommodate cases where $\xi(\tau) \neq 0$.

Lemma 2. *Define* $\mathbf{B} = \begin{pmatrix} \mathbf{1} & \mathbf{v} \end{pmatrix}$ *and* $\boldsymbol{\delta} = (1, \xi(\tau))^\top$. *Then the optimal weights that minimize the asymptotic variance in* (6) *are given as*

$$\mathbf{w}^* = \mathbf{H}^{-1}\mathbf{B}\Big(\mathbf{B}^\top \mathbf{H}^{-1}\mathbf{B}\Big)^{-1}\boldsymbol{\delta}. \tag{8}$$

Proof. The optimization problem in (7) is a quadratic programming problem with equality constraint $\mathbf{B}^\top \mathbf{w} = \boldsymbol{\delta}$. Its Lagrangian expression is

$$\mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \mathbf{w}^\top \mathbf{H}\mathbf{w} - \boldsymbol{\lambda}^\top \Big(\mathbf{B}^\top \mathbf{w} - \boldsymbol{\delta}\Big),$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)^\top$ is the vector of Lagrange multipliers. Differentiating with respect to $\mathbf{w}$ and setting the derivative to zero gives

$$\nabla_{\mathbf{w}}\mathcal{L} = 2\mathbf{H}\mathbf{w} - \mathbf{B}\,\boldsymbol{\lambda} = \mathbf{0},$$

which in turn results in $\mathbf{w} = \frac{1}{2}\mathbf{H}^{-1}\mathbf{B}\,\boldsymbol{\lambda}$. Substituting this expression into the constraint and solve it for $\boldsymbol{\lambda}$ yields

$$\boldsymbol{\lambda} = 2\Big(\mathbf{B}^\top \mathbf{H}^{-1}\mathbf{B}\Big)^{-1}\boldsymbol{\delta}.$$

Finally, substituting back into the expression for $\mathbf{w}$ we obtain the optimal weights $\mathbf{w}^*$ in (8). $\square$

The minimum asymptotic variance by (8) is $\boldsymbol{\delta}^\top \left(\mathbf{B}^\top\mathbf{H}^{-1}\mathbf{B}\right)^{-1}\boldsymbol{\delta}\cdot\mathbf{Q}_2^{-1}$. In particular, when the target quantile level τ is included in the set $\{\tau_1, \ldots, \tau_K\}$, the single-quantile estimator $\hat{\boldsymbol{\beta}}_\tau$ corresponds to the weight vector $w_{(0)} = (0, 0, \ldots, 1, 0, \ldots, 0)^\top$, where the 1 is in the position corresponding to $\tau_k = \tau$. This weight vector satisfies both constraints $w_{(0)}^\top\mathbf{1} = 1$ and $w_{(0)}^\top\mathbf{v} = \xi(\tau)$, and thus $w_{(0)} \in \Omega$. Since the optimal weight w^* minimizes the convex objective function $w^\top\mathbf{H}w$ over Ω, it necessarily holds that $w^{*\top}\mathbf{H}w^* \leq w_{(0)}^\top\mathbf{H}w_{(0)} = h_{rr}$, where h_{rr} denotes the asymptotic variance of the standard quantile regression estimator at level τ. This inequality implies that the proposed combining estimator achieves asymptotic efficiency at least as good as the conventional single-quantile estimator, and strictly better unless the single-quantile solution is already optimal under the quadratic criterion. Therefore, the variance reduction is theoretically guaranteed whenever τ is included in the combining grid. The reduction in asymptotic variance highlights the advantage of leveraging information from multiple quantiles, particularly under heteroscedasticity.

3.2 Estimation of weights

In practice, the optimal weights $\mathbf{w}^*$ in (8) cannot be applied directly because they depend on unknown quantities of the conditional quantiles of the error term, $\xi(\tau_k)$, and the corresponding error densities, $f(\xi(\tau_k))$, for $k = 1, \ldots, K$. Therefore, we need to estimate them to implement the CQR estimator. We now describe the process of obtaining the estimated weights.

First, preliminary estimators $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ are obtained. These may be the same as those mentioned in Section 2 or derived using alternative methods. For example, as noted in Koenker (2005), one may use the interquantile range estimator $\hat{\boldsymbol{\gamma}} = \hat{\boldsymbol{\beta}}_{0.75} - \hat{\boldsymbol{\beta}}_{0.25}$, where $\hat{\boldsymbol{\beta}}_\tau$ denotes the quantile regression estimator at quantile τ. The chosen $\hat{\boldsymbol{\gamma}}$ must satisfy (A5) to ensure consistency up to scale. With these preliminary estimators, we calculate the residuals.

$$r_i = \frac{y_i - \mathbf{x}_i^\top \hat{\boldsymbol{\beta}}}{\mathbf{x}_i^\top \hat{\boldsymbol{\gamma}}}, \quad i = 1, \dots, n.$$

Let $\hat{\xi}(\tau_k)$ denote the sample τ_k-th quantile of the residuals $\{r_i\}_{i=1}^n$ for $k = 1, \dots, K$, which estimate the quantile of the error distribution $\xi(\tau_k)$. Next, the densities at these quantiles are estimated using kernel density estimation.

$$\hat{f}(\hat{\xi}(\tau_k)) = \frac{1}{nb_n} \sum_{i=1}^n K\left(\frac{\hat{\xi}(\tau_k) - r_i}{b_n}\right),$$

where $K(\cdot)$ is a kernel function with bounded support and bounded derivative, and b_n is the bandwidth parameter. With these estimates, we construct the estimated covariance matrix $\hat{\mathbf{H}}$ using the following element in kth row and k'th column.

$$\hat{h}_{kk'} = \frac{\tau_k(1 - \tau_{k'})}{\hat{f}(\hat{\xi}(\tau_k))\hat{f}(\hat{\xi}(\tau_{k'}))}, \quad k, k' = 1, \dots, K.$$

Replacing $\mathbf{B}$ and $\boldsymbol{\delta}$ in Lemma 2 with their plug-in estimators, denoted by $\hat{\mathbf{B}}$ and $\hat{\boldsymbol{\delta}}$ respectively, the estimated optimal weights $\hat{\mathbf{w}}^*$ are computed by

$$\hat{\mathbf{w}}^* = \hat{\mathbf{H}}^{-1}\hat{\mathbf{B}}\left(\hat{\mathbf{B}}^\top \hat{\mathbf{H}}^{-1}\hat{\mathbf{B}}\right)^{-1}\hat{\boldsymbol{\delta}}, \text{ where } \hat{\mathbf{B}} = \begin{pmatrix} \mathbf{1} & \hat{\mathbf{v}} \end{pmatrix} \text{ and } \hat{\boldsymbol{\delta}} = (1, \hat{\xi}(\tau))^\top. \tag{9}$$

This estimator is the explicit solution to the constrained optimization problem, and thus by construction, it satisfies the required constraints exactly, ensuring that $\sum_{k=1}^K \hat{w}_k^* = 1$. To ensure the consistency of the kernel density estimators, we impose the following additional assumption.

(A6) The kernel function $K(\cdot)$ has bounded support and a bounded derivative. The bandwidth b_n satisfies $b_n \to 0$ and $nb_n^4 \to \infty$ as $n \to \infty$.

Let $\hat{\boldsymbol{\beta}}_{\tau,\mathbf{w}^*}^{CQR}$ denote the combining quantile regression estimator with the true optimal weights in (8) and $\hat{\boldsymbol{\beta}}_{\tau,\hat{\mathbf{w}}^*}^{CQR}$ denote the estimator with the estimated weights in (9). Under assumptions (A1)–(A6), the estimated weights $\hat{\mathbf{w}}^*$ converge to the true optimal weights $\mathbf{w}^*$, and the estimator $\hat{\boldsymbol{\beta}}_{\tau,\hat{\mathbf{w}}^*}^{CQR}$ retains the same asymptotic properties as $\hat{\boldsymbol{\beta}}_{\tau,\mathbf{w}^*}^{CQR}$.

Theorem 3. *Suppose assumptions (A1)–(A6) hold. Then,*

(i) $\|\hat{\mathbf{w}}^* - \mathbf{w}^*\| = o_p(1)$.

(ii) $\hat{\boldsymbol{\beta}}_{\tau,\hat{\mathbf{w}}^*}^{CQR}$ *is asymptotically normal with the same distribution as* $\hat{\boldsymbol{\beta}}_{\tau}^{CQR}$ *in Lemma 1:*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{\tau,\hat{\mathbf{w}}^*}^{CQR} - \boldsymbol{\beta}_\tau) \xrightarrow{d} N\left(0, \mathbf{w}^{*T}\mathbf{H}\mathbf{w}^* \cdot \mathbf{Q}_2^{-1}\right).$$

Proof. (i) We first show that $\|\hat{\mathbf{w}}^* - \mathbf{w}^*\| = o_p(1)$. Note that $\mathbf{w}^*$ is invariant under the scaling of ϵ. Since $\mathbf{w}^*$ depends on $\xi(\tau_k)$ and $f(\xi(\tau_k))$, it is sufficient to show that their estimates converge to the true values up to scale.

We first establish a uniform bound for the term $\mathbf{x}_i^\top \hat{\boldsymbol{\gamma}}/\mathbf{x}_i^\top \boldsymbol{\gamma}$. By combining assumption (A5) on the convergence of $\hat{\boldsymbol{\gamma}}$ with assumption (A3) on the predictors, the Cauchy-Schwarz inequality yields:

$$\max_{1\le i\le n}\left|\frac{\mathbf{x}_i^\top \hat{\boldsymbol{\gamma}}}{\mathbf{x}_i^\top \boldsymbol{\gamma}} - \kappa\right| \le \left(\max_{1\le i\le n}\left\|\frac{\mathbf{x}_i}{\sigma_i}\right\|\right) \cdot \|\hat{\boldsymbol{\gamma}} - \kappa\boldsymbol{\gamma}\| = O(n^{1/4}) \cdot O_p(n^{-1/2}) = O_p(n^{-1/4}). \tag{10}$$

Now, to bound the reciprocal, let $\zeta_i = (\mathbf{x}_i^\top \hat{\boldsymbol{\gamma}}/\mathbf{x}_i^\top \boldsymbol{\gamma}) - \kappa$. A first-order Taylor expansion of $(\kappa + \zeta_i)^{-1}$ around $\zeta_i = 0$ gives:

$$\frac{\mathbf{x}_i^\top \boldsymbol{\gamma}}{\mathbf{x}_i^\top \hat{\boldsymbol{\gamma}}} - \frac{1}{\kappa} = \frac{1}{\kappa + \zeta_i} - \frac{1}{\kappa} = -\frac{\zeta_i}{\kappa^2} + O_p(n^{-1/2}).$$

Because the rate of the leading term is determined by $\max_i |\zeta_i| = O_p(n^{-1/4})$, this rate indicates the uniform convergence:

$$\max_{1\le i\le n} \left| \frac{\mathbf{x}_i^\top \boldsymbol{\gamma}}{\mathbf{x}_i^\top \hat{\boldsymbol{\gamma}}} - \frac{1}{\kappa} \right| = O_p(n^{-1/4}). \tag{11}$$

We now derive the uniform bound for the difference between the residual r_i and the scaled error ϵ_i/κ. This difference can be decomposed as:

$$\begin{aligned} r_i - \frac{\epsilon_i}{\kappa} &= \frac{\mathbf{x}_i^\top(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{\mathbf{x}_i^\top \hat{\boldsymbol{\gamma}}} + \left(\frac{\mathbf{x}_i^\top \boldsymbol{\gamma}}{\mathbf{x}_i^\top \hat{\boldsymbol{\gamma}}} - \frac{1}{\kappa} \right) \epsilon_i \\ &= \left(\frac{\mathbf{x}_i^\top(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{\sigma_i} \right) \left(\frac{\sigma_i}{\mathbf{x}_i^\top \hat{\boldsymbol{\gamma}}} \right) + \left(\frac{\mathbf{x}_i^\top \boldsymbol{\gamma}}{\mathbf{x}_i^\top \hat{\boldsymbol{\gamma}}} - \frac{1}{\kappa} \right) \epsilon_i. \end{aligned} \tag{12}$$

We bound the two terms on the right-hand side of (12) uniformly over i. For the first term, the $\sqrt{n}$-consistency of the preliminary estimator $\hat{\boldsymbol{\beta}}$ ensures that $\|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\| = O_p(n^{-1/2})$. Combining with assumption (A3), this yields:

$$\max_{1\le i\le n} \left| \frac{\mathbf{x}_i^\top(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{\sigma_i} \right| \le \left(\max_{1\le i\le n} \left\| \frac{\mathbf{x}_i}{\sigma_i} \right\| \right) \cdot \|\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\| = O(n^{1/4}) \cdot O_p(n^{-1/2}) = O_p(n^{-1/4}).$$

Since $\max_i |\sigma_i/(\mathbf{x}_i^\top \hat{\boldsymbol{\gamma}})| = O_p(1)$ by (11), the first term in (12) is $O_p(n^{-1/4})$ uniformly over i. The bound on the second term also follows directly from (11). Combining these results, we obtain the bound:

$$\left| r_i - \frac{\epsilon_i}{\kappa} \right| = O_p(n^{-1/4})(1 + |\epsilon_i|), \tag{13}$$

which is satisfied uniformly over i.

The uniform bound in (13) implies that for any $\eta > 0$, there exists a constant $M < \infty$ such that the inequality $|r_i - \epsilon_i/\kappa| \le M n^{-1/4}(1 + |\epsilon_i|)$ holds for all i simultaneously with probability at least $1-\eta$. Setting $\delta_n = Mn^{-1/4}$, we can bound r_i:

$$\frac{\epsilon_i}{\kappa} - \delta_n(1+|\epsilon_i|) \leq r_i \leq \frac{\epsilon_i}{\kappa} + \delta_n(1+|\epsilon_i|).$$

This inequality implies that the τ-th sample quantile of the residuals, $\hat{\xi}(\tau)$, is bounded by the sample quantiles of the lower and upper bounds. As $n \to \infty$, $\delta_n \to 0$, and the perturbation term $\delta_n(1+|\epsilon_i|)$ converges to zero in probability. Since the $\{\epsilon_i\}$ are i.i.d., the sample quantiles of both the lower and upper bounds converge in probability to the true τ-th quantile of the distribution of ϵ/κ, which is $\xi(\tau)/\kappa$. By the squeeze theorem, the sample quantile of the residuals must converge to the same limit: $\hat{\xi}(\tau) \xrightarrow{p} \xi(\tau)/\kappa$.

We now demonstrate the consistency of the kernel density estimator, $\hat{f}(x) = (nb_n)^{-1}\sum_i K\{(x-r_i)/b_n\} \xrightarrow{p} \kappa f(\kappa x)$, where $K(\cdot)$ is a predefined kernel function. The difference $\hat{f}(x) - \kappa f(\kappa x)$, scaled by nb_n, can be decomposed as

$$\sum_{i=1}^{n}\left[K\left\{\frac{x-r_i}{b_n}\right\} - K\left\{\frac{x-\epsilon_i/\kappa}{b_n}\right\}\right] + \left[\sum_{i=1}^{n} K\left\{\frac{x-\epsilon_i/\kappa}{b_n}\right\} - \kappa f(\kappa x)\right] =: A_n + B_n.$$

As shown by Silverman (1986), under *i.i.d.* errors, $|B_n| = o_p(nb_n)$. It remains to show that $|A_n| = o_p(nb_n)$.

Without loss of generality, assume the kernel $K(\cdot)$ has support bounded within $[-1,1]$. The summand in A_n is non-zero only for indices in the set $\mathcal{I} := \{1 \leq i \leq n : |r_i - x| \leq b_n \text{ or } |\epsilon_i/\kappa - x| \leq b_n\}$, allowing the summation to be restricted to $\mathcal{I}$.

For any $i \in \mathcal{I}$, we can bound $|\epsilon_i/\kappa - x|$ using the triangle inequality and equation (13):

$$\left|\frac{\epsilon_i}{\kappa} - x\right| \leq |r_i - x| + \left|r_i - \frac{\epsilon_i}{\kappa}\right| \leq b_n + O_p(n^{-1/4})(1+|\epsilon_i|).$$

This inequality implies that for $i \in \mathcal{I}$, $|\epsilon_i|$ must be uniformly bounded in probability, or $|\epsilon_i| = O_p(1)$. Thus, We can define a slightly larger set $\mathcal{I}_1 := \{1 \leq i \leq n : |\epsilon_i/\kappa - x| \leq$

$b_n + M_2 n^{-1/4}\}$ for some constant M_2, such that $\mathcal{I} \subset \mathcal{I}_1$ with probability approaching one.

Under assumption (A6), the derivative of $K(\cdot)$ is bounded. Therefore, the mean value theorem yields:

$$|A_n| \le \sum_{i\in\mathcal{I}} \frac{O(1)}{b_n} \left| r_i - \frac{\epsilon_i}{\kappa} \right| \le \frac{O_p(n^{-1/4})}{b_n} \sum_{i\in\mathcal{I}_1} (1 + |\epsilon_i|). \tag{14}$$

Since $|\epsilon_i|$ is bounded on this set, we can absorb $(1 + |\epsilon_i|)$ into the $O_p(\cdot)$ term, leaving $|A_n| \le O_p(1)/(n^{1/4} b_n)|\mathcal{I}_1|$. The size of the set $|\mathcal{I}_1|$ is a random variable whose expectation is:

$$\mathbb{E}[|\mathcal{I}_1|] = n \cdot \mathbb{P}\left(\left| \frac{\epsilon_i}{\kappa} - x \right| \le b_n + M_2 n^{-1/4} \right) \le n \cdot O(b_n + n^{-1/4}),$$

where the last step uses the boundedness of the density function $f(\cdot)$. It follows that $|\mathcal{I}_1| = O_p(n(b_n + n^{-1/4}))$. Substituting this into the bound for $|A_n|$ gives:

$$|A_n| = O_p\left(\frac{n(b_n + n^{-1/4})}{n^{1/4} b_n} \right) = O_p\left(n^{3/4} + \frac{n^{1/2}}{b_n} \right).$$

This rate is $o_p(nb_n)$ because the condition $nb_n^4 \to \infty$ from assumption (A6) ensures that $n^{3/4}/(nb_n) = 1/(n^{1/4}b_n) \to 0$ and $(n^{1/2}/b_n)/(nb_n) = 1/(\sqrt{n}b_n^2) \to 0$. Since both $|A_n|$ and $|B_n|$ are $o_p(nb_n)$, the proof is complete.

Without loss of generality, assume that the support of $K(\cdot)$ is bounded within $[-1, 1]$. Note that the summand in A_n is zero for indices $i \notin \mathcal{I}$, where $\mathcal{I} := \{1 \le i \le n : |r_i - x| \le b_n \text{ or } |\epsilon_i/\kappa - x| \le b_n\}$. Thus, the summation in A_n can be restricted to the set $\mathcal{I}$. Using equation (13), we have

$$\left| \frac{\epsilon_i}{\kappa} - x \right| \le |r_i - x| + \left| r_i - \frac{\epsilon_i}{\kappa} \right| \le b_n + O_p(n^{-1/4})(1 + |\epsilon_i|), \quad \text{for } i \in \mathcal{I}.$$

Hence, there exists a constant M_1 such that $|\epsilon_i/\kappa - x| \leq b_n + M_1 n^{-1/4}(1+|\epsilon_i|)$ for all $i \in \mathcal{I}$ with probability approaching one. Define the set $\mathcal{I}_1 := \{1 \leq i \leq n : |\epsilon_i/\kappa - x| \leq b_n + M_1 n^{-1/4}(1+|\epsilon_i|)\}$. Then, $\mathcal{I} \subset \mathcal{I}_1$. For $i \in \mathcal{I}_1$, $|\epsilon_i| \leq |x| + |\epsilon_i - x| \leq |x| + b_n + O_p(n^{-1/4})(1+|\epsilon_i|)$, which further implies that $\epsilon_i \leq (|x|+b_n+O_p(n^{-1/4}))/(1-O_p(n^{-1/4}) = O_p(1)$ uniformly over i. That is, if $i \in \mathcal{I}_1$, $|\epsilon_i$ is uniformly bounded in probability. Applying it to $\mathcal{I}_1$, we define the set $\mathcal{I}_2 := \{1 \leq i \leq n : |\epsilon_i| \leq M_2$ and $|\epsilon_i - x| \leq b_n + M_2 n^{-1/4}\}$. Then, $\mathcal{I}_1 \subset \mathcal{I}_2$.

Under condition (A6), which assumes the boundness of the derivative of K, the mean value theorem yields

$$|A_n| = \frac{O(1)}{b_n}\sum_{i\in\mathcal{I}}\left|r_i - \frac{\epsilon_i}{\kappa}\right| = \frac{O_p(1)}{n^{1/4}b_n}\sum_{i\in\mathcal{I}_2}(1+|\epsilon_i|) \leq \frac{O_p(1)}{n^{1/4}b_n}\sum_{i=1}^{n}\mathbb{I}\left[\left|\frac{\epsilon_i}{\kappa} - x\right| \leq b_n + M_2 n^{-1/4}\right], \tag{15}$$

with probability approaching one. The indicator fucntion in the second inequality restricts the summation to the elements in $\mathcal{I}_2$. Taking expectations in the inequality above and using the boundedness of the density $f(\cdot)$, we obtain

$$\mathbb{E}\left[\mathbb{I}\left\{\left|\frac{\epsilon_i}{\kappa} - x\right| \leq b_n + M_2 n^{-1/4}\right\}\right] \leq M_3\Big(b_n + M_2 n^{-1/4}\Big) \leq M_4(b_n + n^{-1/2}),$$

for some constants M_3 and M_4. Therefore, $|A_n| = O_p(n^{-1/2} + 1/b_n) = o_p(nb_n)$ under the condition $nb_n^2 \to \infty$ in (A6).

(ii) Next, we show that $\hat{\boldsymbol{\beta}}^{CQR}_{\tau,\hat{\mathbf{w}}^*}$ shares the same asymptotic distribution as $\hat{\boldsymbol{\beta}}^{CQR}_{\tau,\mathbf{w}^*}$. Consider the scaled difference between these two.

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{CQR}_{\tau,\hat{\mathbf{w}}^*} - \boldsymbol{\beta}_\tau) - \sqrt{n}(\hat{\boldsymbol{\beta}}^{CQR}_{\tau,\mathbf{w}^*} - \boldsymbol{\beta}_\tau) = \sqrt{n}\sum_{k=1}^{K}(\hat{w}_k^* - w_k^*)(\hat{\boldsymbol{\beta}}_{\tau_k} - \boldsymbol{\beta}_\tau). \tag{16}$$

Since $\hat{\boldsymbol{\beta}}_{\tau_k}$ are $\sqrt{n}$-consistent estimators of $\boldsymbol{\beta}_{\tau_k}$ and $\|\hat{\mathbf{w}}^* - \mathbf{w}^*\| = o_p(1)$, the righthand side of (16) is $o_p(1)$. Then,

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{CQR}_{\tau,\hat{\mathbf{w}}^*} - \boldsymbol{\beta}_\tau) = \sqrt{n}(\hat{\boldsymbol{\beta}}^{CQR}_{\tau,\mathbf{w}^*} - \boldsymbol{\beta}_\tau) + o_p(1).$$

Hence, $\hat{\boldsymbol{\beta}}^{CQR}_{\tau,\hat{\mathbf{w}}^*}$ shares the same asymptotic distribution as $\hat{\boldsymbol{\beta}}^{CQR}_{\tau,\mathbf{w}^*}$, thus completing the proof (as stated in Lemma 1). □

3.3 Selecting the range of combining quantiles

In this subsection, we discuss selecting combining quantiles for our method. Previous studies estimating central tendencies often use evenly spaced quantiles with $\tau_k = k/10$ for $k = 1, \ldots, 9$. However, when targeting a specific quantile τ, quantile selection can extend beyond this uniform spacing. We propose instead to symmetrically center the quantiles around τ. Given a fixed K, we define a symmetric interval $[\tau - c_\tau, \tau + c_\tau]$, where we call c_τ a *window width*. By evenly distributing quantiles within $[\tau - c_\tau, \tau + c_\tau]$, the selection task reduces to choosing the window width c_τ. For example, with $\tau = 0.3$ and $c_\tau = 0.2$, $K = 9$ combining quantiles become $(0.10, 0.15, \ldots, 0.50)$.

Under this definition of symmetric interval, to ensure all quantiles lie within $(0, 1)$, the window width is constrained by $c_\tau^{SM} = \min(\tau, 1 - \tau)$, referred to as the *symmetric maximum*. However, this may omit necessary information, particularly when targeting tail quantiles. We therefore generalize the interval to an asymmetric one, $[\max(0.05, \tau - c_\tau), \min(0.95, \tau + c_\tau))]$. Note that we restrict quantiles to $(0.05, 0.95)$ to avoid extreme values with insufficient sample density. The asymmetric window width can extend up to $c_\tau^{AM} = \max(0.95 - \tau, \tau - 0.05)$, which we denote as the *asymmetric maximum*.

It is important to note that the asymptotic variance of the combining quantile regression estimator depends on the window width parameter c_τ. With fixed K, the asymptotic variance in (6) can be expressed as a function of c_τ,

$\boldsymbol{\delta}^\top \Big(\mathbf{B}(c_\tau)^\top \mathbf{H}(c_\tau)^{-1}\mathbf{B}(c_\tau)\Big)^{-1} \boldsymbol{\delta} \cdot \mathbf{Q}_2^{-1}$. Here, $\mathbf{B}(c_\tau)$ and $\mathbf{H}(c_\tau)$ are analogous to $\mathbf{B}$ and $\mathbf{H}$ but now depend on c_τ.

The optimal window width could be defined as the minimizer of this asymptotic variance. However, due to the complex role of c_τ in determining the quantile levels τ_k and, consequently, the matrices $\mathbf{H}(c_\tau)$ and $\mathbf{B}(c_\tau)$, this minimization problem does not admit a closed-form solution. Even numerical optimization may be challenging because the objective function is a nontrivial function of c_τ that involves the inverse of $\mathbf{H}(c_\tau)$ and the interplay of the kernel density estimates of f at the corresponding quantiles. These difficulties motivate us to adopt a more practical approach.

Therefore, we propose an empirical guideline for the window width, based on comprehensive simulation results. A brief overview of a preliminary experiment is presented here, with further details provided in the subsequent section. We consider a univariate model: $y = \beta x + (\gamma_0 + \gamma x)\epsilon$, where $\beta = 1$, $\gamma_0 = 1$, $\gamma = 2$, and $\epsilon \sim N(0,1)$. For sample sizes $n = 100, 500, 1000$, we vary c_τ from 0.05 to c_τ^{AM} for target quantiles $\tau = 0.1, 0.2, 0.3$. We evaluate the performance of the estimated regression coefficients using both the theoretical optimal weights in (8) and the estimated weights in (9).

Denote the single quantile estimator at the b-th iteration as $\hat\beta_{\tau,b}$ and our estimator as $\hat\beta_{\tau,b}^{CQR}$. Then the relative MSE is computed as

$$\frac{\sum_{b=1}^{B}(\hat\beta_{\tau,b}^{CQR} - \beta_\tau)^2}{\sum_{b=1}^{B}(\hat\beta_{\tau,b} - \beta_\tau)^2}. \tag{17}$$

Relative MSE below 1 indicates the proposed method is preferred. Figure 1 shows that our method consistently yields lower MSE than single quantile regression (all relative MSEs < 1). For central quantiles, extending the window width beyond c_τ^{SM} does not improve performance. However, for tail quantiles such as $\tau = 0.1$, the full range $(0.05, 0.95)$ captures essential information. It is suggested to select the (possible) widest window width under the linear model in (1) because the relative MSE tends to

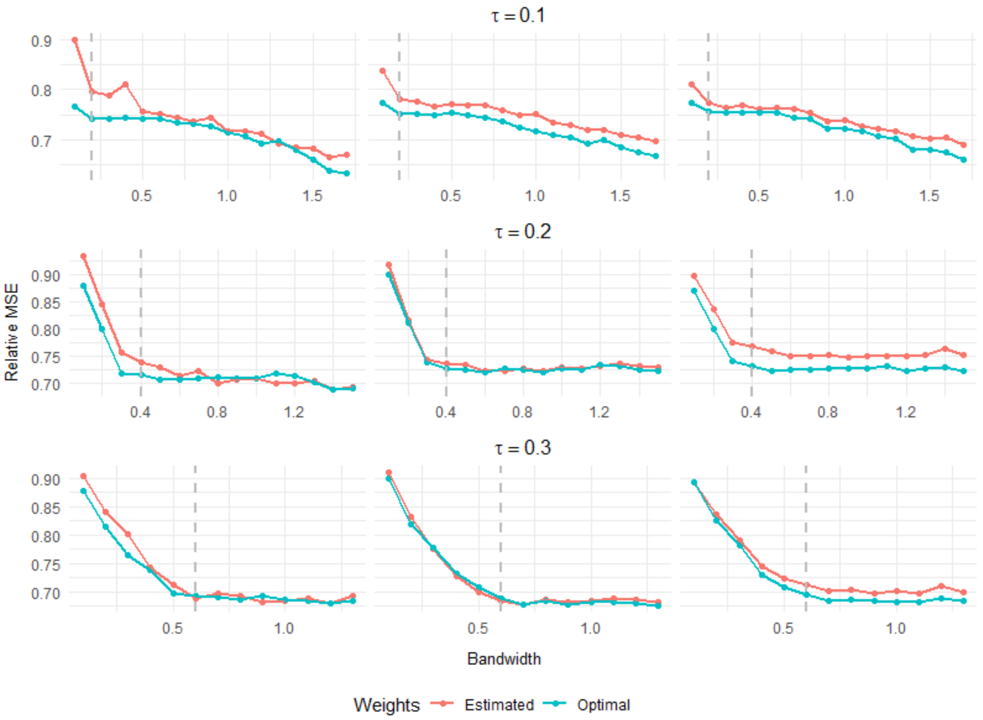


Fig. 1: Relative MSEs from 1000 Monte Carlo simulations, where vertical dotted lines indicate the value of the symmetric maximum window width c_τ^{SM} and panels show results for $n = 100, 500, 1000$ from left to right

decrease as the window width increases. We conjecture that this is due to the larger reduction in the variance of the proposed estimate with the wider window width. From these results, we derive the following empirical rule: use c_τ^{SM} for quantiles whose corresponding density is "sufficiently" high, and use c_τ^{AM} otherwise. What constitutes a "sufficient" density level will be discussed in detail with simulation results in the next section. Note that this rule may not be universally acceptable when extending our method to nonlinear data.

Extension to nonlinear data. While the previous discussion has focused on linear models, our framework extends naturally to nonlinear relationships. Consider the generalized model

$$y = g(\mathbf{x}) + h(\mathbf{x})\epsilon,$$

where g and h are unknown nonlinear functions. Suppose we have a nonlinear quantile regression model that estimates the conditional quantile function. Some examples are quantile smoothing splines (Koenker et al, 1994), support vector quantile regression (Takeuchi et al, 2006), quantile regression forests (Meinshausen and Ridgeway, 2006), and quantile neural networks (Cannon, 2011). Let $\hat{Q}_y(\tau|\mathbf{x})$ denote the conditional τ-th quantile estimate of y. Then, our combining quantile regression estimator becomes

$$\hat{Q}_y^{CQR}(\tau|\mathbf{x}) = \sum_{k=1}^{K} w_k \hat{Q}_y(\tau_k|\mathbf{x}).$$

Due to the complexity of these nonlinear models, the theoretical derivation of optimal weights is infeasible; hence, we consider equal weights, $\mathbf{w} = (1/K, \dots, 1/K)^\top$. This simplification introduces asymptotic bias because the original framework depends on optimal weights for bias correction. Consequently, the selection of window width becomes even more critical. Excessively dispersed quantile levels may exacerbate bias, whereas concentrating them around τ can mitigate bias due to the heightened correlation among the quantile estimates. Nonetheless, such concentration may also diminish the potential variance reduction attainable by combining estimators. Based on our empirical findings (presented in the next section), we propose the following heuristic window width

$$c_\tau = 0.625 \min(\tau, 1 - \tau). \tag{18}$$

Although this rule is not the best for all models and datasets, but it generally shows improved performance. Alternatively, the user may treat c_τ as a tuning parameter and select it through validation methods, though this approach can be computationally intensive when repeatedly fitting the quantile model for estimates at $\tau_1, \dots, \tau_K$.

4 Simulation study

4.1 Linear location-scale model

We conduct an extensive simulation study to assess the performance of our method within a linear location-scale framework. The primary objectives are: (i) to compare the estimation accuracy of CQR with traditional quantile regression under various conditions, and (ii) to examine the effect of window width and estimated weight on the performance of CQR.

Datasets are generated from the two-dimensional linear location-scale model,

$$y = \mathbf{x}^\top \boldsymbol{\beta} + (\gamma_0 + \mathbf{x}^\top \boldsymbol{\gamma})\epsilon,$$

where γ_0 is the intercept of scale. We consider two parameter settings:

- Setting 1: $\boldsymbol{\beta} = (1, 0.5)^\top$, $\gamma_0 = 0.5$, and $\boldsymbol{\gamma} = (4, 3)^\top$.
- Setting 2: $\boldsymbol{\beta} = (1, 2, 0, 0)^\top$, $\gamma_0 = 0.01$, and $\boldsymbol{\gamma} = (3, 5, 0.5, 0.5)^\top$.

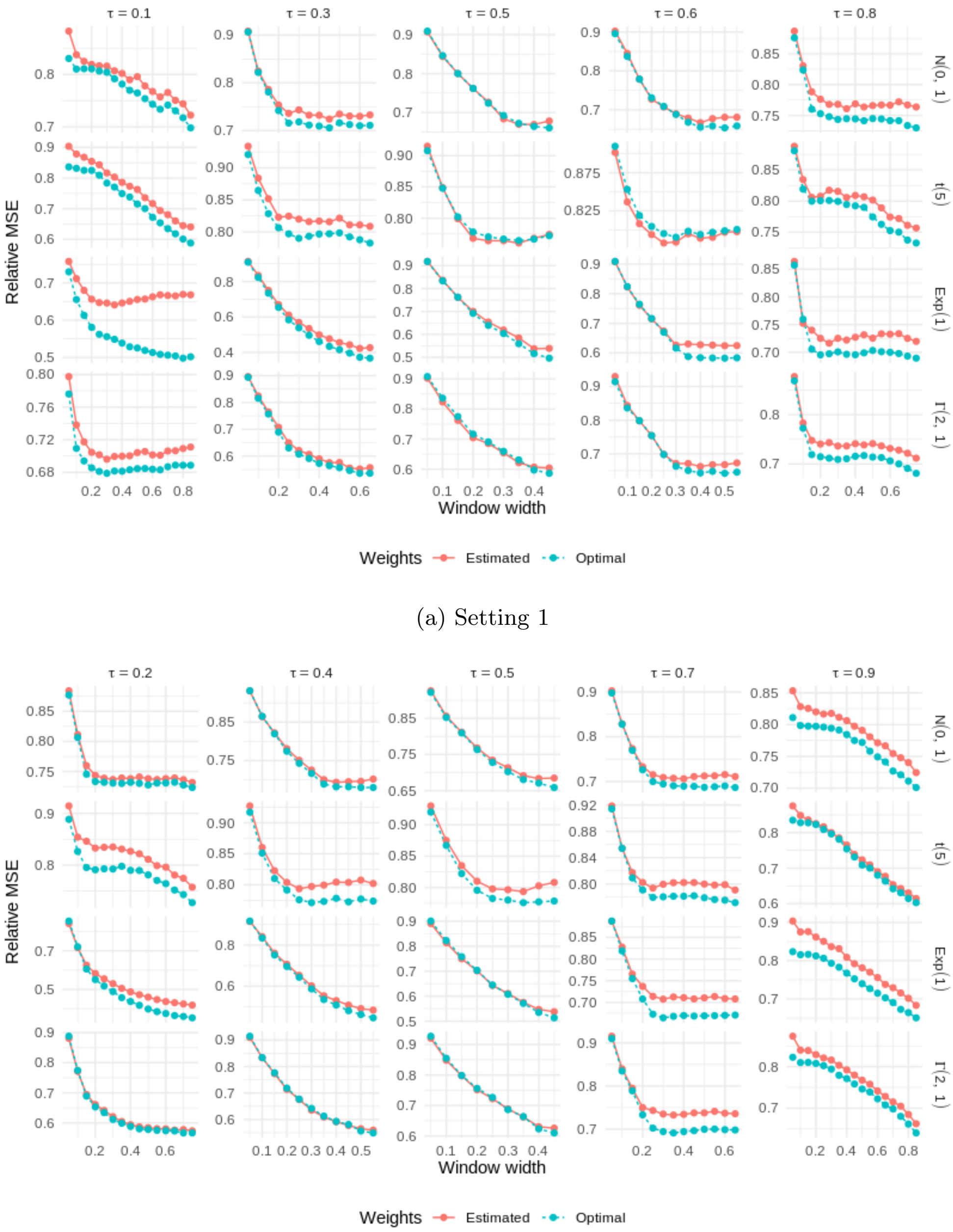
All covariates are independently drawn from $U(1, 5)$. In Setting 2, the last two covariates have zero coefficients in $\boldsymbol{\beta}$, contributing to heteroscedasticity only through non-zero coefficients in $\boldsymbol{\gamma}$. We examine three factors in our simulations. They are error distributions ($N(0, 1)$, $t(5)$, $\text{Exp}(1)$, and $\text{Gamma}(2, 1)$), sample sizes ($n = 100, 200, 500, 1000$), and target quantiles ($\tau = 0.1, 0.3, 0.5, 0.6, 0.8$ for Setting 1 and $\tau = 0.2, 0.4, 0.5, 0.7, 0.9$ for Setting 2).

We first investigate the effect of window width selection. The combining quantile estimator is computed using (i) the estimated weights following the procedure described in Section 3.2 and (ii) the theoretical optimal weights in (8). The preliminary estimators required for weight estimation are obtained via ordinary least squares. Window widths are varied from 0.05 to $c_\tau^{AM} = \max(0.95 - \tau, \tau - 0.05)$ in increments of 0.05. The number of combining quantiles is fixed at $K = 9$. Performance is evaluated using the relative MSE defined in (17) over $B = 300$ simulation iterations, where a relative MSE less than 1 indicates that our method shows higher estimation accuracy.

Figure 2 shows the relative MSE for $n = 500$. The results for other sample sizes are provided in the supplementary materials. In accordance with the experiment presented in Section 3.3, the 'best' window width frequently corresponds to the asymmetric maximum in scenarios where the error density is low (e.g., for tail quantiles with sparse samples). This phenomenon is observed, for example, under a standard normal error with $\tau = 0.1$ or 0.9, as well as in the case of an exponential error with $\tau = 0.9$. In other instances, performance ceases to improve once the window width reaches the symmetric maximum. Furthermore, when utilizing estimated weights, excessively wide window widths may results in a deterioration of performance due to estimation errors. Upon computing the density corresponding to the target quantile for each case, we derive the following heuristic: employ the asymmetric maximum when the density value is 0.25 or lower, and opt for the symmetric maximum in all other cases. Although the optimal window width is also influenced by the density of the quantile levels being combined, this guideline should provide a useful framework for practitioners seeking an efficient implementation.

Next, we investigate whether estimating weights provides an advantage over using equal weights, specifically focusing on the number of observations necessary to reliably estimate the weights. To this end, we compare the single quantile regression estimator with two variants of CQR:

1. Equally weighted (ECQR): Combines quantiles with equal weights ($w_k = 1/K$ for $k = 1, \dots, K$).
2. Estimated Weighted (WCQR): Combines quantiles using the estimated weights described in Section 3.2.



(a) Setting 1

(b) Setting 2

Fig. 2: Relative MSE of WCQR across different window widths for various target quantiles ($\epsilon \sim N(0, 1)$, sample size $n = 500$)

We fix number of combining quantiles $K = 9$. WCQR is assessed using both symmetric and asymmetric maximum window widths. Given that the asymmetric range is unsuitable for equal weighting, ECQR is evaluated solely using the symmetric maximum window width.

Table 1: Relative MSE of CQR with equal weights (ECQR) and estimated weights (WCQR) compared to single quantile regression from 1000 Monte Carlo simulation. The numbers below ECQR and WCQR indicate the window width

Setting 1

τ		0.1			0.3			0.5		0.6			0.8		
		ECQR	WCQR		ECQR	WCQR		ECQR	WCQR	ECQR	WCQR		ECQR	WCQR	
c_τ		0.05	0.05	0.85	0.25	0.25	0.65	0.45	0.45	0.35	0.35	0.55	0.15	0.15	0.75
ϵ	n														
N(0,1)	100	0.86	0.97	0.72	0.76	0.74	0.72	0.71	0.74	0.72	0.72	0.72	0.82	0.87	0.81
	200	0.87	0.92	0.74	0.77	0.72	0.71	0.64	0.66	0.72	0.70	0.69	0.82	0.79	0.76
	500	0.88	0.88	0.72	0.82	0.74	0.73	0.67	0.68	0.76	0.68	0.68	0.84	0.79	0.76
	1000	0.86	0.84	0.73	0.89	0.71	0.70	0.66	0.67	0.82	0.67	0.66	0.85	0.75	0.74
t(5)	100	0.90	1.14	0.68	0.91	0.85	0.83	0.96	0.85	0.93	0.85	0.87	0.92	0.90	0.82
	200	0.92	1.00	0.62	0.94	0.81	0.80	0.94	0.81	0.93	0.82	0.81	0.95	0.88	0.79
	500	0.90	0.90	0.64	1.02	0.82	0.81	0.93	0.77	1.11	0.79	0.80	1.00	0.81	0.76
	1000	0.91	0.84	0.63	1.31	0.81	0.78	0.88	0.74	1.11	0.74	0.73	1.05	0.81	0.75
Exp(1)	100	0.85	0.77	1.16	0.77	0.64	0.50	1.18	0.53	1.07	0.63	0.62	0.95	0.82	0.75
	200	0.81	0.74	0.74	0.70	0.58	0.42	1.55	0.52	1.26	0.60	0.60	0.93	0.76	0.72
	500	0.82	0.76	0.67	0.76	0.61	0.43	2.46	0.54	1.73	0.63	0.62	0.96	0.74	0.72
	1000	0.80	0.73	0.61	0.79	0.58	0.41	4.54	0.55	2.83	0.63	0.62	1.14	0.73	0.71
Γ(2,1)	100	0.85	0.88	0.93	0.66	0.62	0.55	1.01	0.64	0.98	0.67	0.68	0.92	0.82	0.76
	200	0.82	0.81	0.74	0.67	0.64	0.54	1.10	0.61	1.07	0.66	0.67	0.92	0.80	0.76
	500	0.81	0.80	0.71	0.68	0.65	0.56	1.49	0.60	1.36	0.67	0.67	0.92	0.75	0.71
	1000	0.80	0.78	0.71	0.68	0.64	0.52	2.21	0.59	1.91	0.66	0.65	1.01	0.71	0.69

Setting 2

τ		0.2			0.4			0.5		0.7			0.9		
		ECQR	WCQR		ECQR	WCQR		ECQR	WCQR	ECQR	WCQR		ECQR	WCQR	
c_τ		0.15	0.15	0.75	0.35	0.35	0.55	0.45	0.45	0.25	0.25	0.65	0.05	0.05	0.85
ϵ	n														
N(0,1)	100	0.82	0.84	0.79	0.76	0.75	0.77	0.72	0.76	0.74	0.78	0.79	0.88	0.99	0.78
	200	0.81	0.79	0.76	0.73	0.71	0.71	0.69	0.71	0.72	0.73	0.74	0.87	0.94	0.76
	500	0.81	0.76	0.73	0.74	0.70	0.70	0.67	0.68	0.71	0.71	0.71	0.86	0.85	0.72
	1000	0.81	0.76	0.74	0.74	0.68	0.68	0.67	0.69	0.72	0.72	0.72	0.87	0.84	0.71
t(5)	100	0.93	0.93	0.83	0.93	0.89	0.91	0.92	0.88	0.84	0.87	0.89	0.92	1.08	0.75
	200	0.93	0.89	0.83	0.89	0.83	0.86	0.92	0.85	0.85	0.85	0.87	0.92	0.99	0.67
	500	0.94	0.85	0.76	0.96	0.80	0.80	0.94	0.81	0.82	0.80	0.79	0.90	0.88	0.61
	1000	0.98	0.83	0.78	0.99	0.79	0.79	0.93	0.79	0.88	0.82	0.82	0.90	0.86	0.61
Exp(1)	100	0.79	0.81	0.67	0.77	0.61	0.56	1.14	0.63	0.89	0.80	0.78	0.92	1.07	0.74
	200	0.72	0.67	0.51	0.80	0.59	0.53	1.32	0.59	0.94	0.77	0.76	0.93	0.98	0.71
	500	0.72	0.63	0.42	0.83	0.55	0.48	1.64	0.54	0.91	0.71	0.71	0.90	0.90	0.68
	1000	0.72	0.61	0.39	0.92	0.54	0.47	2.50	0.55	1.00	0.69	0.69	0.91	0.86	0.66
Γ(2,1)	100	0.77	0.76	0.69	0.71	0.66	0.63	0.91	0.69	0.88	0.83	0.82	0.91	1.05	0.73
	200	0.74	0.73	0.63	0.69	0.63	0.58	0.88	0.61	0.83	0.76	0.76	0.88	0.94	0.68
	500	0.72	0.69	0.57	0.69	0.61	0.56	1.12	0.63	0.84	0.74	0.74	0.89	0.87	0.66
	1000	0.72	0.69	0.57	0.70	0.60	0.55	1.42	0.61	0.87	0.73	0.72	0.89	0.86	0.66

Table 1 demonstrates that WCQR, when using the asymmetric maximum window width, achieves the lowest relative MSE in most instances when $n \geq 200$. However, for smaller samples ($n = 100$), WCQR occasionally underperforms ECQR or QR primarily due to additional errors arising from weight and density estimation. This effect is particularly pronounced in Setting 2, where the inclusion of two additional covariates further complicates the estimation process. These findings suggest that weight estimation remains reliable only when a sufficiently large number of observations are available. In high-dimensional settings, weight estimation becomes more challenging, as it necessitates larger sample sizes, and the assumed location-scale model may no longer be adequate.

4.2 Nonlinearity and heteroscedasticity

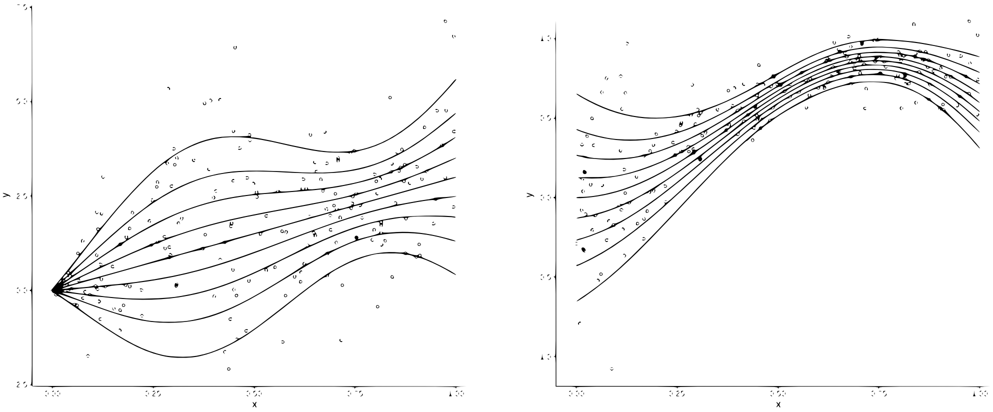
We extend the CQR method to synthetic datasets that exhibit nonlinearity and heteroscedasticity. Notably, determining the optimal weights under a nonlinear model presents a significant challenge due to its complexity and the lack of closed-form solutions. Consequently, we consider an equally weighted version of CQR. For clarity, we refer to CQR as the equally weighted version throughout this subsection.

Simulated datasets are generated from the model $y = g(x) + h(x)\epsilon$ with the following specifications:

- Setting 3: $g(x) = 3x$, $h(x) = 0.5 + 2x + sin(2\pi x - 0.5)$
- Setting 4: $g(x) = sin(0.25x) \times sin(1.5x)$, $h(x) = \sqrt{0.01 + 0.25(1 - sin(2.5x))^2}$.

In Setting 3, the conditional mean function $g(x)$ is linear, while the scale function $h(x)$ is nonlinear, introducing heteroscedasticity that varies with x. Setting 4 presents a more complex scenario where both $g(x)$ and $h(x)$ are nonlinear, resulting in a fully nonlinear conditional distribution of y. In both settings, the covariate x is drawn from a uniform distribution $U(0, 1)$, and the error term ϵ follows one of three distributions:

$N(0,1)$, $t(5)$, and $\text{Exp}(1)$. Figure 3 illustrates the true conditional quantile functions for Settings 3 and 4 with standard normal error.



(a) Setting 3

(b) Setting 4

Fig. 3: Examples of simulated datasets from Setting 3 (left panel) and Setting 4 (right panel) with $\epsilon \sim N(0,1)$, where the curves represent the true conditional quantile functions for $\tau = 0.1, 0.2, \ldots, 0.9$ from bottom to top

For estimation, we use quantile smoothing splines (QSS) and support vector quantile regression (SVQR) with a Gaussian kernel. QSS minimizes a penalized quantile loss, with smoothness controlled by the parameter λ. SVQR extends support vector regression to quantile estimation, using parameters σ (inverse kernel width) and C (regularization parameter). Again, the combining quantiles are equally spaced within $(\tau - c_\tau, \tau + c_\tau)$, where c_τ varies from 0.05 to c_τ^{SM}.

We vary sample sizes $n = 200, 400, 1000$, splitting each dataset evenly into training and validation sets. Model parameters are selected via grid search to minimize mean check loss on the validation data. The final model is fitted using the full samples, and performance is assessed on 5000 test samples that are separately generated. We compute the mean squared error (MSE) between the true and estimated conditional quantile functions on the test data:

$$\text{MSE} = \frac{1}{5000} \sum_{i=1}^{5000} \left(Q_y(\tau|x_i) - \hat{Q}_y(\tau|x_i) \right)^2 .$$

Following the approach in (17), we report the relative MSE of our model compared to the single quantile model over 100 Monte Carlo iterations.

Table 2 shows that the widest window width does not necessarily result in optimal performance in nonlinear settings. As previously discussed, our estimator exhibits asymptotic bias in the absence of appropriate weighting. Since the combining quantiles are centered around the target quantile, individual quantile functions can closely approximate the target quantile when they are sufficiently close. However, while a wider window width may reduce variance, it can also introduce greater bias.

This effect becomes more pronounced in the presence of high heteroscedasticity, which often results in greater variation among quantile functions, thereby amplifying the discrepancy between their average and the target quantile function. In such cases, the equally weighted CQR method exhibits limited effectiveness in mitigating bias. Consequently, narrower window widths are preferable, as they offer a more favorable trade-off between variance reduction and bias minimization.

In both settings, the optimal window widths are generally smaller than $c_\tau^{SM} = \min(\tau, 1-\tau)$. The error distribution significantly influences performance; distributions with heavier tails or skewness introduce additional challenges but do not negate the advantages of CQR when window widths are carefully selected. While accounting for data characteristics is crucial when determining window widths in nonlinear models, practical constraints, such as time and computational resources, may impose limitations on exhaustive tuning. In our experiments, using the empirical rule in (18) provides modest improvements across different settings, so we suggest it as a rule of thumb.

Table 2: Relative MSE of the combining quantile regression estimator compared to original quantile regression from 100 Monte Carlo simulation. We try the window width from 0.05 to $\min(\tau, 1-\tau)$ by 0.05. For simplicity, we only record the window width with the lowest relative MSE

Settting 3

		QSS										SVQR									
		$\tau = 0.1$		$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.6$		$\tau = 0.8$		$\tau = 0.1$		$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.6$		$\tau = 0.8$	
ϵ	n	c_τ	RE	c_τ	RE	c_τ	RE	c_τ	RE	c_τ	RE	c_τ	RE	c_τ	RE	c_τ	RE	c_τ	RE	c_τ	RE
N(0,1)	200	0.09	0.72	0.20	0.72	0.40	0.62	0.30	0.73	0.15	0.76	0.09	0.74	0.20	0.81	0.15	0.88	0.05	0.89	0.05	0.87
	400	0.07	0.72	0.20	0.70	0.35	0.66	0.25	0.75	0.09	0.77	0.09	0.68	0.20	0.69	0.20	0.80	0.15	0.84	0.10	0.89
	1000	0.07	0.87	0.15	0.73	0.40	0.65	0.20	0.79	0.10	0.83	0.09	0.66	0.20	0.64	0.20	0.86	0.15	0.78	0.07	0.79
t(5)	200	0.05	0.76	0.20	0.67	0.30	0.67	0.20	0.84	0.09	0.89	0.07	0.81	0.20	0.81	0.15	0.85	0.10	0.84	0.07	0.80
	400	0.05	0.76	0.15	0.81	0.35	0.72	0.20	0.83	0.10	0.81	0.07	0.80	0.20	0.67	0.25	0.82	0.07	0.98	0.07	0.88
	1000	0.05	0.84	0.15	0.82	0.35	0.74	0.20	0.75	0.09	0.88	0.07	0.70	0.20	0.64	0.20	0.71	0.15	0.85	0.09	0.82
Exp(1)	200	0.07	0.83	0.20	0.74	0.20	0.78	0.20	0.79	0.07	0.84	0.07	0.86	0.03	0.86	0.05	0.95	0.09	0.89	0.07	0.86
	400	0.09	0.73	0.20	0.79	0.15	0.85	0.15	0.81	0.05	0.85	0.09	0.67	0.10	0.82	0.09	0.88	0.05	0.90	0.07	0.86
	1000	0.09	0.72	0.20	0.80	0.15	0.81	0.10	0.79	0.07	0.87	0.09	0.63	0.20	0.84	0.10	0.84	0.05	0.89	0.07	0.80

Settting 4

		QSS										SVQR									
		$\tau = 0.1$		$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.6$		$\tau = 0.8$		$\tau = 0.1$		$\tau = 0.3$		$\tau = 0.5$		$\tau = 0.6$		$\tau = 0.8$	
ϵ	n	c_τ	RE	c_τ	RE	c_τ	RE	c_τ	RE	c_τ	RE	c_τ	RE	c_τ	RE	c_τ	RE	c_τ	RE	c_τ	RE
N(0,1)	200	0.07	0.72	0.20	0.63	0.30	0.70	0.25	0.68	0.15	0.67	0.07	0.83	0.20	0.74	0.40	0.62	0.30	0.82	0.15	0.79
	400	0.07	0.81	0.20	0.52	0.05	0.89	0.20	0.54	0.15	0.67	0.07	0.75	0.15	0.81	0.40	0.65	0.25	0.73	0.10	0.84
	1000	0.07	0.77	0.20	0.60	0.05	0.84	0.20	0.63	0.15	0.71	0.07	0.76	0.15	0.78	0.35	0.62	0.25	0.68	0.10	0.72
t(5)	200	0.05	0.81	0.20	0.65	0.15	0.72	0.15	0.77	0.10	0.81	0.07	0.57	0.20	0.80	0.25	0.78	0.15	0.81	0.10	0.81
	400	0.05	0.77	0.20	0.68	0.10	0.88	0.20	0.72	0.09	0.78	0.05	0.91	0.20	0.83	0.30	0.76	0.20	0.83	0.09	0.88
	1000	0.05	0.82	0.15	0.71	0.15	0.82	0.20	0.62	0.09	0.73	0.07	0.79	0.15	0.73	0.25	0.69	0.15	0.73	0.10	0.78
Exp(1)	200	0.09	0.60	0.20	0.76	0.20	0.71	0.15	0.81	0.09	0.77	0.03	0.95	0.20	0.84	0.15	0.82	0.10	0.88	0.07	0.86
	400	0.09	0.56	0.20	0.73	0.20	0.72	0.20	0.75	0.10	0.79	0.07	0.82	0.15	0.78	0.10	0.85	0.15	0.82	0.09	0.78
	1000	0.07	0.73	0.20	0.70	0.15	0.79	0.15	0.74	0.09	0.80	0.09	0.76	0.15	0.76	0.15	0.86	0.15	0.77	0.07	0.85

5 Real data analysis

In this section, we present the experimental results of our method applied to real-world datasets, which often pose challenges such as small sample sizes, outliers, and quantile crossing—where higher quantile functions fail to remain above lower ones, violating monotonicity. Our analysis focuses on the method's ability to reduce variance and enhance robustness against outliers. By integrating multiple quantile estimates, CQR yields more stable and reliable fits.

We analyze two datasets: the Barro-Lee educational attainment dataset and the Auto MPG dataset. To evaluate the effectiveness of CQR, we compare it with single quantile regression models, emphasizing improvements in robustness and reductions in quantile crossing. Model performance is assessed using two key metrics: difference-in-fit (DFFIT) and crossing loss.

- Originally, DFFITs measure the sensitivity of a fitted model to individual observations. In this study, we adapt this concept by computing the sum of absolute differences across all observations, interpreting it as an indicator of model stability—specifically, the degree to which the model's predictions are affected by the inclusion or exclusion of a single observation. It is defined as

$$\text{DFFITs} = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{Q}_{\tau}(y|\mathbf{x}_i) - \hat{Q}_{\tau}^{(-i)}(y|\mathbf{x}_i)\right|,$$

  where $\hat{Q}_{\tau}^{(-i)}(y|\mathbf{x}_i)$ denote the fitted quantile function of the i-th observation, estimated without including the i-th observation. Lower DFFITs values indicate a more robust model with reduced sensitivity to individual data points.

- Crossing Loss: Crossing loss quantifies violations of the monotonicity property in quantile functions—i.e., ensuring that estimated quantiles satisfy $\hat{Q}_{\tau_k}(y|\mathbf{x}) \leq \hat{Q}_{\tau_{k'}}(y|\mathbf{x})$ for all $\tau_k < \tau_{k'}$. A crossing violation occurs when a lower quantile estimate is greater than a higher quantile estimate. Following Sangnier et al (2016), we define the crossing loss as:

$$\text{Crossing Loss} = \sum_{\tau_k < \tau_{k'}}\sum_{i=1}^{n}\mathbb{I}\left(\hat{Q}_{\tau_k}(y|\mathbf{x}_i) > \hat{Q}_{\tau_{k'}}(y|\mathbf{x}_i)\right).$$

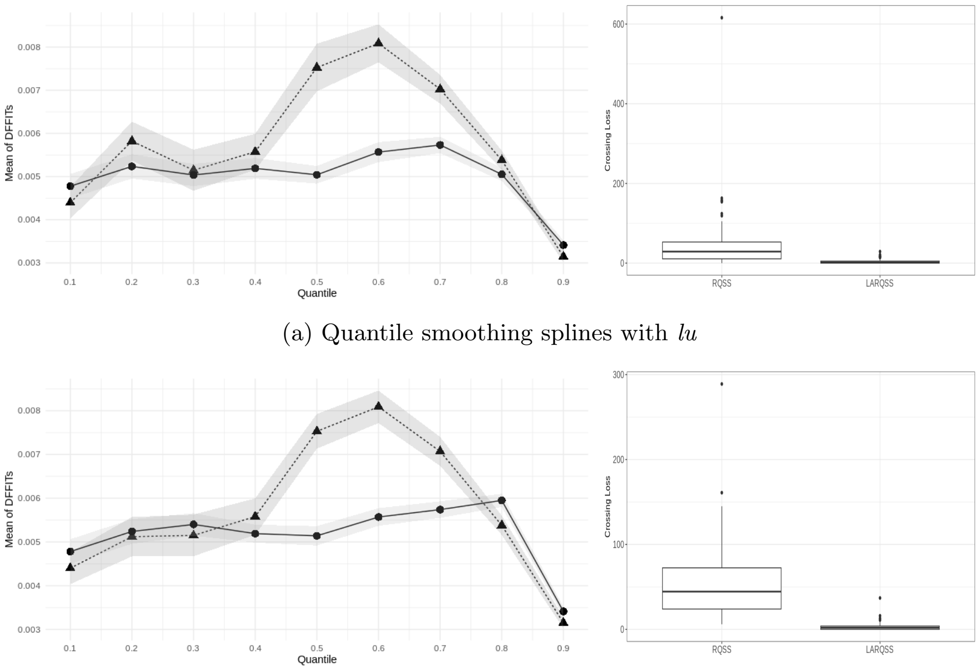
  A lower crossing loss value indicates fewer instances of quantile crossing, thereby improving model reliability.

We implement CQR using quantile smoothing splines (QSS) and support vector quantile regression (SVQR). The window width is set by the empirical rule mentioned in the earlier section, $c_\tau = 0.625\min(\tau, 1-\tau)$.

Barro-Lee. The Barro-Lee dataset comprises 1898 observations from 146 countries, detailing educational attainment levels from 1950 to 2010. The data is freely available at http://www.barrolee.com. The response variable, average years of schooling *yr_sch*, is modeled separately using two covariates:

- *lu*: Proportion of the population with no schooling.
- *lpc*: Proportion of the population that has completed primary schooling.

Quantile regression models are fitted using QSS and its CQR-implemented version (CQSS) for each covariate. The dataset is randomly split into training and test sets across 100 iterations.

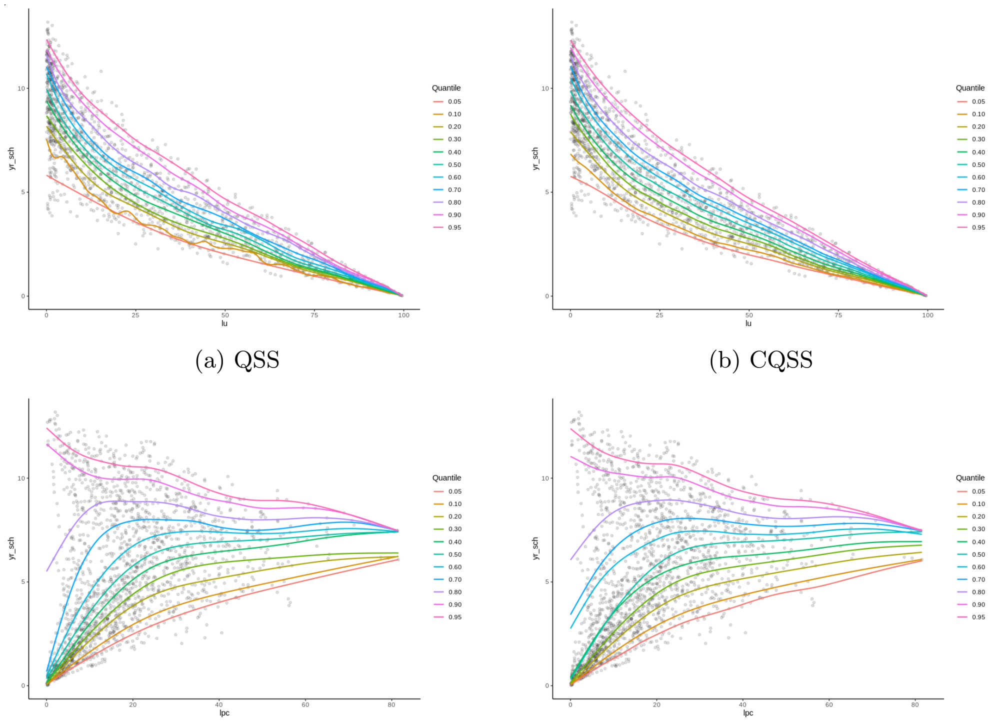


(a) Quantile smoothing splines with *lu*

(b) Quantile smoothing splines with *lpc*

Fig. 4: Mean DFFITs (left panel) for each target quantile, where the solid line represents QSS, the dashed line represents CQSS, and the shaded areas indicate one standard error, and a boxplot of crossing loss (right panel)

Figure 4 presents the mean DFFITs values and crossing loss across various target quantiles for QSS and CQSS. CQSS generally shows lower DFFITs values, particularly for central quantiles ($\tau = 0.4$ to $\tau = 0.7$), indicating reduced sensitivity to individual observations. Additionally, CQSS has smaller standard errors in DFFITs, suggesting improved model stability. Crossing loss results show that our method reduces quantile crossing compared to using only single quantile. This improvement is further illustrated in the fitted quantile curves in Figure 5. While QSS displays fluctuations and crossings in some regions, CQSS produces smoother, monotonic quantile fits. By combining multiple quantiles, our method can provide more robust quantile functions compared to the single quantile model.



(a) QSS

(b) CQSS

(c) QSS

(d) CQSS

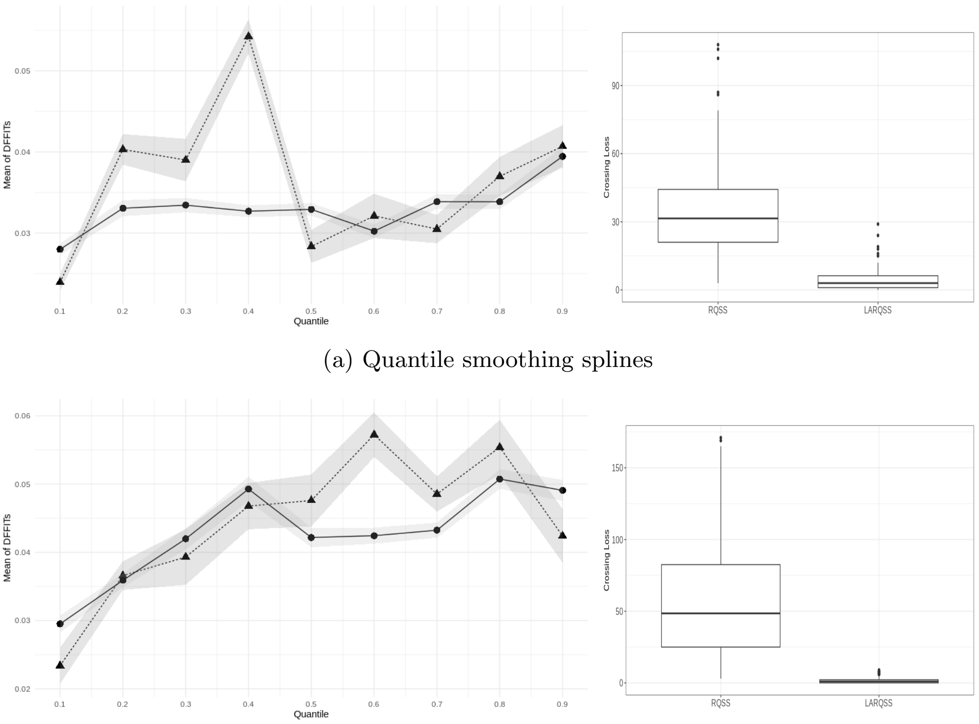
Fig. 5: Fitted quantile curves for $\tau = 0.05, 0.1, \ldots, 0.9, 0.95$ of *yr_sch* in the Barro-Lee dataset, with the top panels presenting QSS and CQSS fits for *lu* and the bottom panels displaying fits for *lpc*

Auto MPG. The Auto MPG dataset contains 398 observations on vehicles, including miles per gallon (mpg) and vehicle weight. This dataset is available in UCI machine learning repository. The objective is to predict mpg based on vehicle weight, following prior studies. Observations with missing values were excluded, and the dataset was randomly split into training and test sets over 100 iterations.

Figure 6 compares the mean DFFITs values for traditional and our models. CQSS and CSVQR (CQR-implemented SVQR) consistently achieve lower DFFITs values than their standalone counterparts, indicating enhanced robustness to individual data points. This improvement is particularly pronounced at specific quantiles, demonstrating CQR's ability to stabilize fits across the conditional distribution. Crossing loss metrics in Figure 6 further confirm that the combined models significantly reduce quantile crossing, improving monotonicity and reliability. Fitted quantile curves in Figure 7 further support these findings. Traditional models, such as QSS and SVQR, exhibit irregularities and crossings, particularly near the boundaries of the data. In contrast, the proposed models generate smoother and more consistent quantile estimates across the full range of vehicle weights. These results underscore the effectiveness of CQR in mitigating the risks of overfitting.

6 Conclusion

This study addresses the estimation of specific quantiles by combining estimates from multiple quantiles. Our method, combining quantile regression (CQR), reduces variance and mitigates issues such as quantile crossing. Key contributions include addressing practical considerations in the combined quantiles approach, such as weight estimation and selecting multiple quantiles. Simulation studies demonstrate that CQR consistently outperforms traditional quantile regression methods, particularly in the presence of heteroscedastic errors. Applications to real-world datasets further validate



(a) Quantile smoothing splines

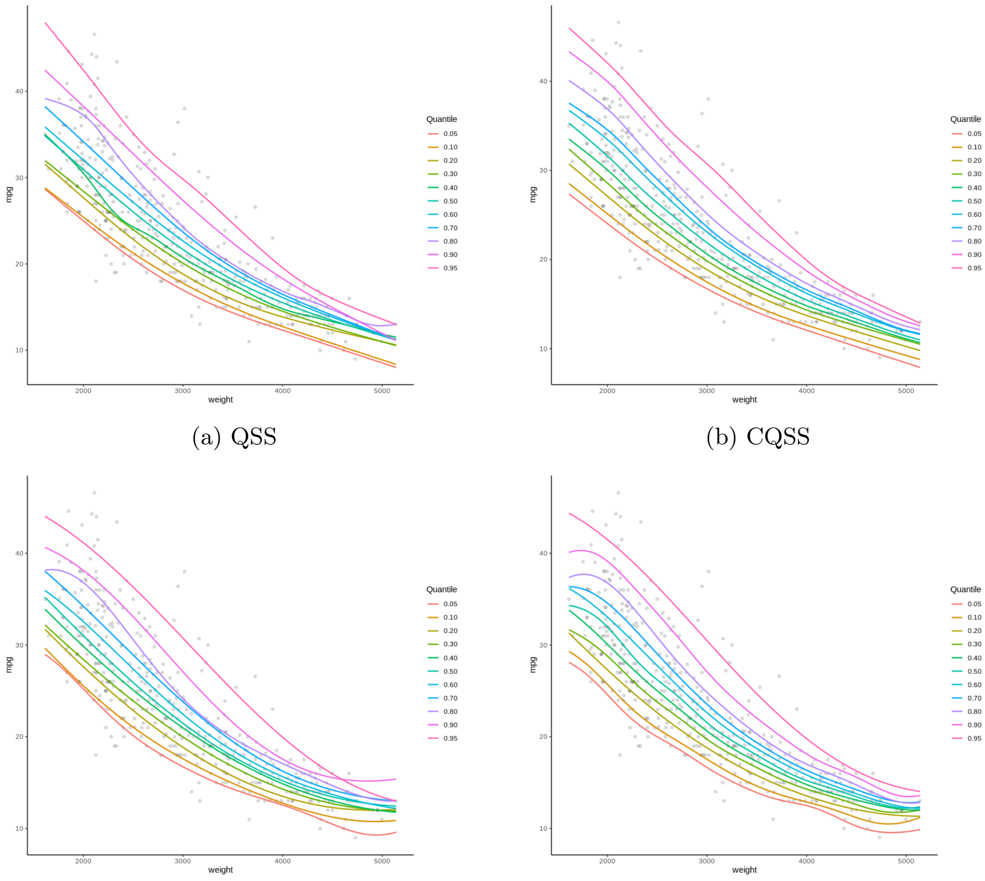
(b) Support vector quantile regression

Fig. 6: Mean DFFITs for each target quantile (left panel) and box plots of crossing loss (right panel) are shown, where in the DFFITs plots the solid line indicates QSS, the dashed line indicates CQSS, and shaded regions represent one standard error, and the box plots depict the distribution of crossing loss across iterations

its practical advantages, delivering more stable and accurate estimates across a wide range of quantiles.

Despite its advantages, the CQR method depends on certain assumptions, such as correct model specification and adequate data for accurate weight estimation. In nonlinear models, estimating optimal weights is more challenging due to the complexity of underlying asymptotic properties. Additionally, if not carefully calibrated, equally weighted CQR may introduce undesirable bias.

Future research could extend the optimal weight estimation process to nonlinear models, leveraging advanced asymptotic techniques or machine learning algorithms.



(a) QSS

(b) CQSS

(c) SVQR

(d) CSVQR

Fig. 7: Fitted curves for conditional quantiles ($\tau = 0.05, 0.1, \ldots, 0.9, 0.95$) of the mpg variable in the `mpg` dataset, where the top panels display the regression curves from quantile smoothing splines and the bottom panels display those from support vector quantile regression

Integrating CQR with machine learning or deep learning architectures could enhance adaptability and performance in high-dimensional or nonparametric settings. Additionally, exploring the application of CQR to time-series data with autocorrelation or spatial data with geographical dependencies could broaden its scope. Addressing challenges such as missing data or outliers using CQR also presents a promising avenue for further research.

7 Appeendix

In this appendix, we present detailed results from the simulation studies discussed in Section 4 of the main text.

7.1 Additional results for linear cases (Settings 1 and 2)

In the main text, we evaluated two settings under the linear location-scale model:

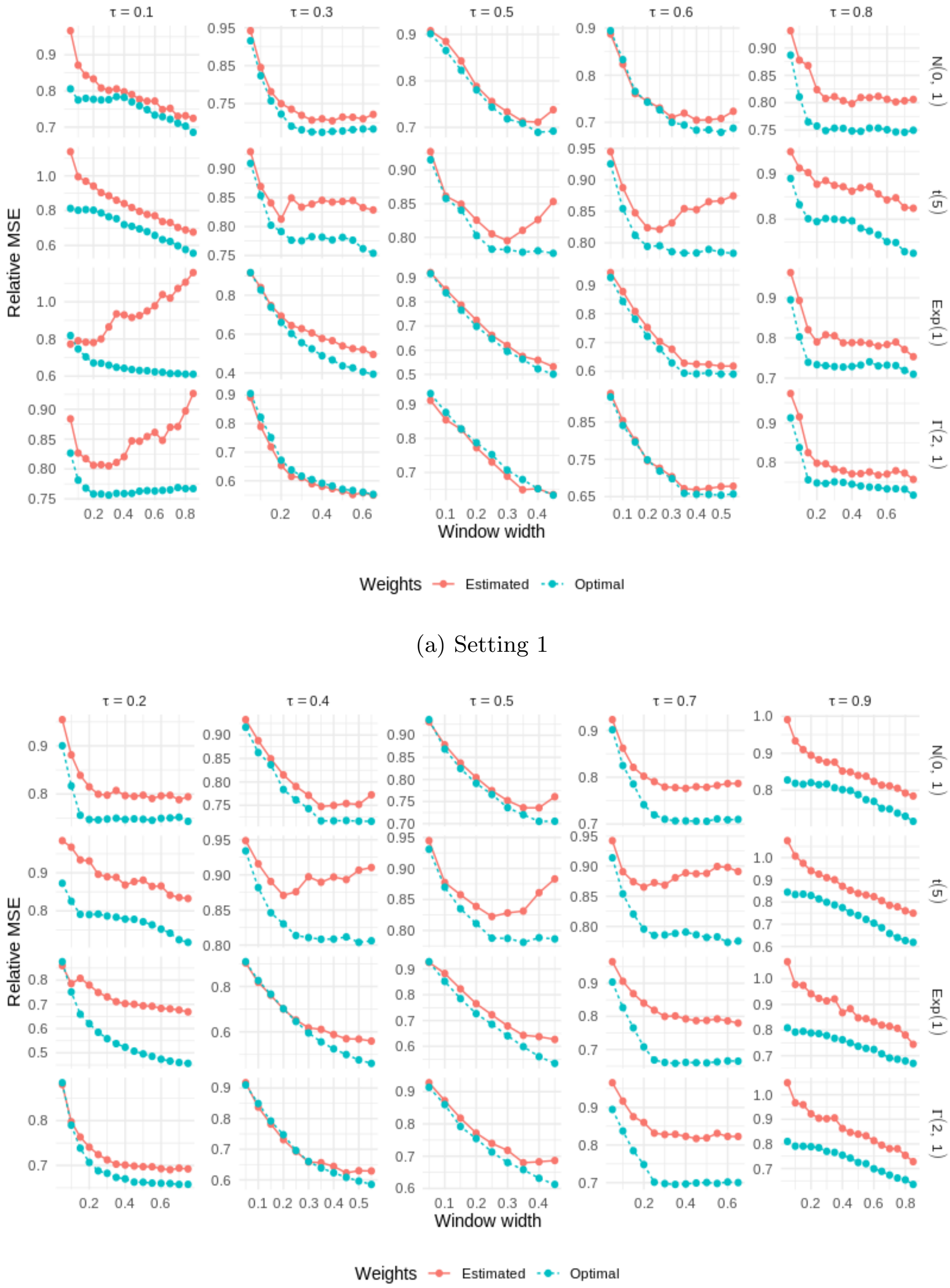
$$y = \mathbf{x}^\top \beta + (\gamma_0 + \mathbf{x}^\top \gamma)\epsilon,$$

with parameters

- Setting 1: $\beta = (1, 0.5)^\top$, $\gamma_0 = 0.5$, and $\gamma = (4, 3)^\top$.
- Setting 2: $\beta = (1, 2, 0, 0)^\top$, $\gamma_0 = 0.01$, and $\gamma = (3, 5, 0.5, 0.5)^\top$.

In both settings, the covariates $\mathbf{x}$ are independently drawn from a uniform distribution $U(1, 5)$. In the main text, we provided results with a sample size of $n = 500$; here in the appendix we cover additional sample sizes $n = 100, 200$, and 1000. All other simulation settings (e.g., error distributions, the number of combining quantiles, and the variation of window width) remain the same as in the main text.

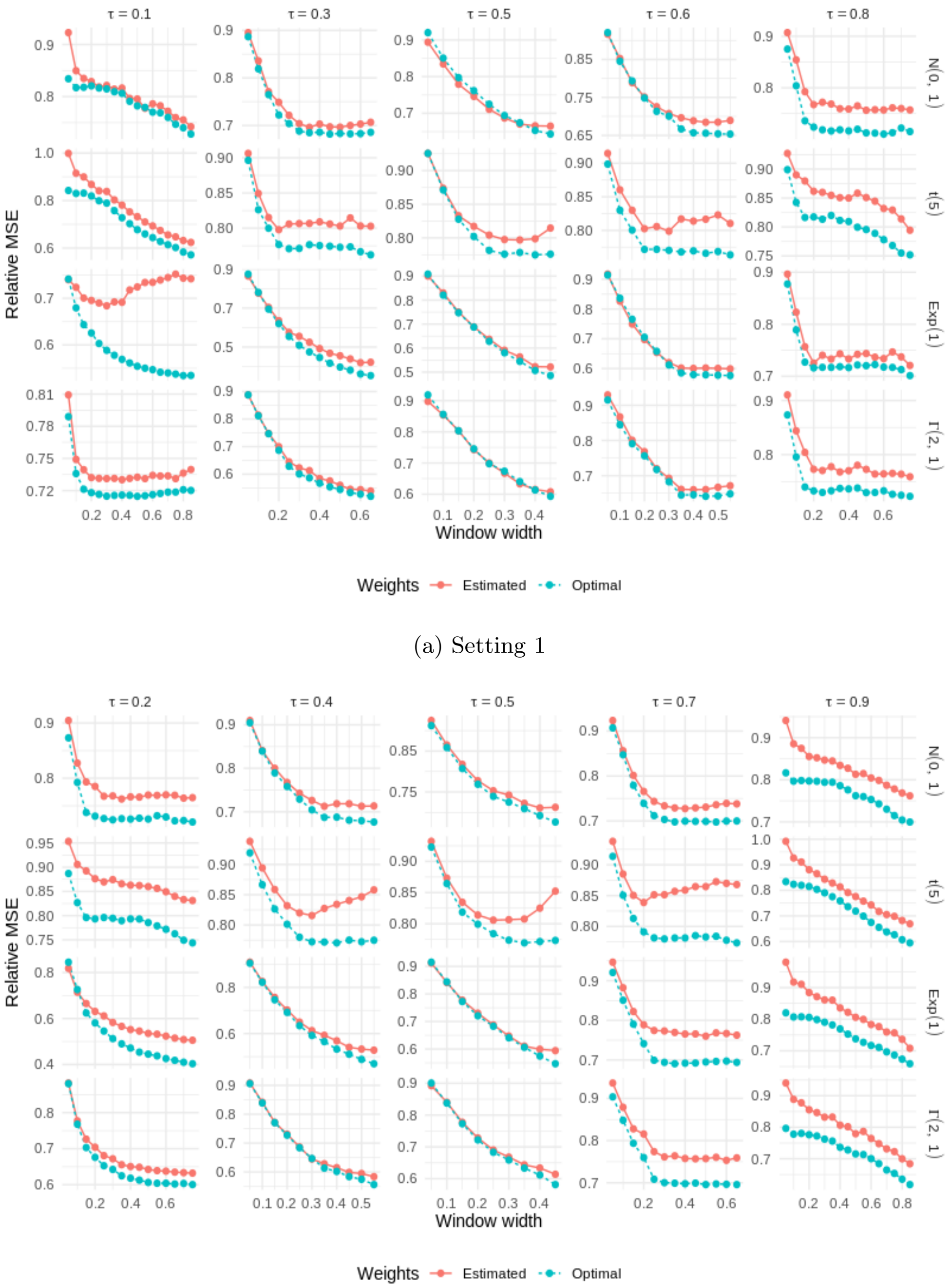
Figures 8, 9, and 10 show the relative mean squared errors (MSEs) across different window widths for $n = 100$, 200, and 1000, respectively. For $n = 200$ and $n = 1000$, the results closely align with those presented in the main text. In general, CQR with the asymmetric maximum window width outperforms single quantile regression. Our method performs well across various sample sizes once sufficient sample density is secured. For $n = 100$, however, CQR's performance with estimated weights sometimes declines as the window width increases. This drop is likely due to inaccuracies in weight estimation caused by unreliable density estimates in smaller datasets. In such cases, selecting a symmetric maximum, c_τ^{SM}, is more effective. Alternatively, it may



(a) Setting 1

(b) Setting 2

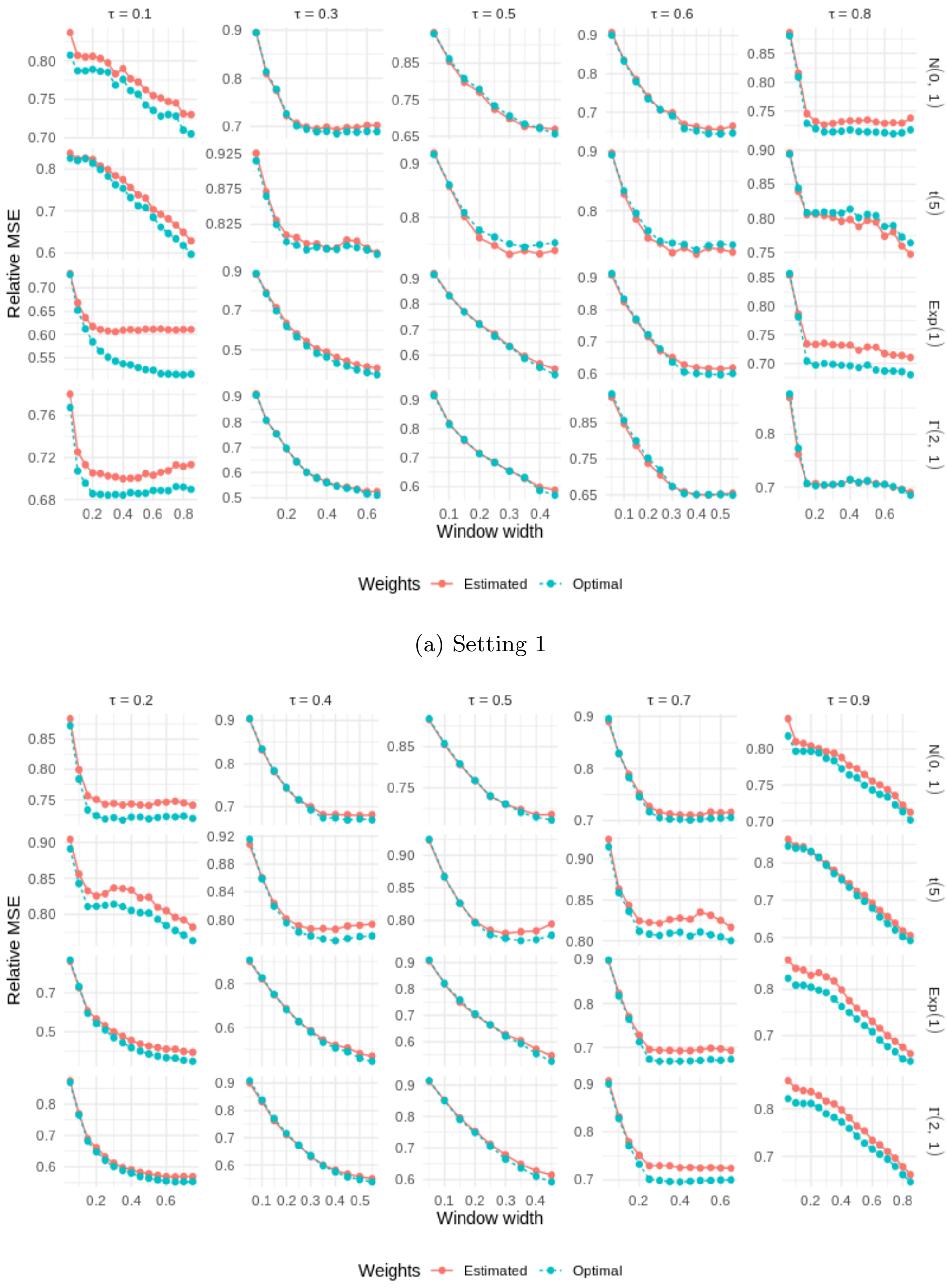
Fig. 8: The mean of relative MSEs from 300 Monte Carlo simulation datasets with sample size $n = 100$



(a) Setting 1

(b) Setting 2

Fig. 9: The mean of relative MSEs from 300 Monte Carlo simulation datasets with sample size $n = 200$



(a) Setting 1

(b) Setting 2

Fig. 10: The mean of relative MSEs from 300 Monte Carlo simulation datasets with sample size $n = 1000$

be preferable to use the equally weighted version of CQR with a smaller window width range, thereby avoiding the additional estimation error associated with weight estimation. These findings emphasize the need to consider sample size when choosing window widths and weights for CQR in linear models.

7.2 Additional results for nonlinear cases (Settings 3 and 4)

For the nonlinear models, we previously investigated the performance of equally weighted CQR due to the challenges of estimating optimal weights in nonlinear settings.

The models used are $y = g(x) + h(x)\epsilon$ with nonlinear functions g and f, specified as follows:

- Setting 3: $g(x) = 3x$ and $h(x) = 0.5 + 2x + \sin(2\pi x - 0.5)$.
- Setting 4: $g(x) = \sin(0.25x) \times \sin(1.5x)$ and $h(x) = \sqrt{0.01 + 0.25(1 - \sin(2.5x))^2}$.

The covariate x is drawn from $U(0, 1)$ in both settings, and the error term ϵ follows N(0,1), t(5), and Exp(1). For each target quantile, we vary the window width within the appropriate interval up to the symmetric maximum. In the main text, we reported only the window width that yielded the best performance for each case; here in the appendix we present the full results.

Tables 3 and 4 compare the MSEs across different window widths for Settings 3 and 4, respectively. The baseline refers to the single quantile model without combining. The results indicate that choosing an appropriate window width is crucial. A window width that is too wide often results in worse performance than the traditional model due to the bias introduced when averaging over quantiles with significantly different conditional quantile functions in nonlinear contexts. Optimal performance is generally achieved with narrower, carefully tuned window widths for each quantile.

Table 3: The mean of MSEs with 100 Monte Carlo simulations under Setting 3. The top half is the result of using a quantile smoothing spline, and the bottom half is the result of using a support vector quantile regression. Bold text indicates the lowest value. All values are multiplied by 10^3

ϵ	n			$\tau = 0.1$					$\tau = 0.3$		
		Baseline	0.06	0.10	0.14	0.18	Baseline	0.14	0.20	0.30	0.40
N(0,1)	200	5.63(0.56)	4.64(0.36)	4.36(0.35)	4.06(0.32)	**4.04(0.28)**	3.04(0.27)	2.61(0.19)	2.45(0.19)	2.31(0.17)	**2.19(0.17)**
	400	2.96(0.25)	2.42(0.20)	2.30(0.19)	**2.14(0.18)**	2.17(0.20)	1.97(0.15)	1.74(0.13)	1.57(0.11)	1.44(0.10)	**1.37(0.10)**
	1000	1.28(0.09)	1.17(0.08)	1.12(0.08)	**1.11(0.09)**	1.36(0.11)	0.82(0.06)	0.68(0.05)	0.63(0.04)	**0.60(0.04)**	0.64(0.04)
t(5)	200	7.12(0.64)	5.89(0.47)	**5.39(0.39)**	5.80(0.47)	8.07(0.79)	3.79(0.35)	3.21(0.29)	2.84(0.25)	2.65(0.23)	**2.53(0.21)**
	400	4.83(0.39)	3.83(0.23)	**3.69(0.21)**	3.73(0.19)	5.09(0.32)	2.00(0.17)	1.80(0.14)	1.71(0.14)	**1.63(0.13)**	1.74(0.13)
	1000	2.33(0.17)	2.07(0.16)	**1.95(0.14)**	2.14(0.14)	3.96(0.25)	0.92(0.06)	0.82(0.06)	0.77(0.06)	**0.75(0.05)**	0.85(0.06)
Exp(1)	200	0.36(0.03)	0.32(0.02)	0.30(0.02)	**0.30(0.02)**	0.32(0.02)	1.05(0.12)	0.96(0.12)	0.86(0.08)	0.81(0.07)	**0.77(0.06)**
	400	0.17(0.01)	0.15(0.01)	0.14(0.01)	0.13(0.01)	**0.13(0.01)**	0.51(0.03)	0.46(0.03)	0.43(0.03)	0.41(0.03)	**0.40(0.03)**
	1000	0.08(0.01)	0.07(0.01)	0.07(0.01)	0.06(0.01)	**0.06(0.00)**	0.24(0.03)	0.22(0.02)	0.21(0.02)	0.19(0.02)	**0.19(0.02)**
N(0,1)	200	4.83(0.39)	4.21(0.36)	4.04(0.35)	**4.00(0.36)**	4.16(0.37)	3.80(0.34)	3.34(0.31)	3.17(0.31)	2.96(0.28)	**2.79(0.26)**
	400	3.03(0.25)	2.73(0.24)	2.46(0.21)	**2.26(0.18)**	2.33(0.19)	1.74(0.14)	1.45(0.11)	1.43(0.11)	**1.41(0.12)**	1.42(0.14)
	1000	1.35(0.09)	1.14(0.08)	1.06(0.08)	**1.03(0.08)**	1.25(0.10)	0.80(0.06)	0.68(0.05)	0.65(0.05)	**0.62(0.05)**	0.65(0.05)
t(5)	200	14.22(1.12)	8.79(0.81)	8.25(0.74)	**8.11(0.71)**	9.34(0.85)	4.80(0.38)	4.41(0.33)	4.29(0.32)	4.00(0.29)	**3.83(0.29)**
	400	7.59(0.55)	7.01(0.52)	**6.94(0.51)**	7.01(0.51)	11.61(0.56)	2.12(0.16)	1.91(0.15)	1.86(0.15)	1.78(0.15)	**1.75(0.14)**
	1000	3.47(0.27)	2.91(0.21)	2.76(0.19)	**2.75(0.18)**	4.15(0.21)	1.15(0.09)	0.91(0.07)	0.86(0.07)	**0.84(0.06)**	0.87(0.06)
Exp(1)	200	0.69(0.07)	0.65(0.07)	**0.66(0.08)**	0.71(0.09)	1.99(0.10)	1.53(0.27)	1.35(0.22)	1.33(0.23)	**1.31(0.23)**	1.28(0.22)
	400	0.32(0.03)	0.27(0.03)	**0.26(0.03)**	0.26(0.03)	0.39(0.03)	0.63(0.05)	0.54(0.05)	0.51(0.05)	**0.49(0.04)**	0.51(0.05)
	1000	0.10(0.01)	0.09(0.01)	**0.08(0.01)**	0.08(0.01)	0.08(0.01)	0.27(0.02)	0.23(0.02)	0.22(0.02)	**0.21(0.02)**	0.21(0.02)
ϵ	n			$\tau = 0.6$					$\tau = 0.8$		
		Baseline	0.10	0.30	0.50	0.70	Baseline	0.10	0.18	0.20	0.30
N(0,1)	200	2.72(0.21)	2.40(0.18)	2.17(0.16)	**2.07(0.15)**	2.15(0.15)	3.50(0.34)	3.27(0.29)	3.15(0.27)	2.85(0.21)	**2.66(0.17)**
	400	1.66(0.14)	1.55(0.13)	1.32(0.10)	**1.24(0.09)**	1.47(0.11)	2.36(0.21)	2.08(0.19)	1.97(0.18)	**1.82(0.16)**	1.84(0.15)
	1000	0.70(0.05)	0.65(0.05)	0.59(0.04)	**0.56(0.04)**	0.71(0.04)	0.88(0.07)	0.83(0.06)	0.80(0.06)	**0.74(0.05)**	0.78(0.04)
t(5)	200	3.43(0.27)	3.11(0.24)	**2.93(0.22)**	2.94(0.21)	3.85(0.27)	5.64(0.42)	5.30(0.41)	5.21(0.39)	**5.08(0.37)**	5.51(0.40)
	400	1.86(0.16)	1.67(0.14)	1.58(0.12)	**1.55(0.12)**	2.23(0.14)	3.10(0.26)	2.85(0.25)	2.74(0.24)	**2.53(0.21)**	2.86(0.20)
	1000	0.90(0.07)	0.80(0.07)	**0.69(0.05)**	0.72(0.05)	1.35(0.08)	1.16(0.09)	1.06(0.08)	1.05(0.07)	**1.03(0.06)**	1.43(0.08)
Exp(1)	200	2.88(0.25)	2.61(0.22)	**2.32(0.18)**	2.56(0.19)	5.14(0.29)	5.80(0.41)	5.24(0.35)	5.04(0.34)	**4.94(0.32)**	5.70(0.36)
	400	1.62(0.14)	1.40(0.11)	**1.31(0.10)**	1.61(0.13)	4.32(0.23)	3.52(0.34)	3.15(0.29)	**2.97(0.27)**	3.04(0.26)	4.04(0.33)
	1000	0.73(0.06)	0.62(0.05)	**0.58(0.04)**	0.93(0.05)	3.33(0.11)	1.68(0.14)	1.52(0.13)	1.47(0.12)	**1.47(0.11)**	2.18(0.12)
N(0,1)	200	3.17(0.24)	3.00(0.23)	2.77(0.21)	**2.65(0.21)**	2.69(0.20)	4.23(0.33)	3.75(0.30)	3.62(0.29)	**3.44(0.27)**	3.34(0.27)
	400	1.90(0.19)	1.75(0.18)	1.58(0.16)	**1.38(0.13)**	1.48(0.11)	2.25(0.21)	2.08(0.21)	2.02(0.20)	**1.89(0.19)**	1.90(0.17)
	1000	0.74(0.06)	0.62(0.05)	0.53(0.04)	**0.50(0.04)**	0.69(0.05)	1.01(0.08)	0.88(0.07)	0.81(0.06)	**0.73(0.05)**	0.78(0.06)
t(5)	200	3.36(0.37)	3.04(0.33)	**2.72(0.29)**	2.93(0.31)	4.21(0.43)	5.16(0.54)	4.65(0.45)	4.44(0.43)	**4.16(0.41)**	4.96(0.56)
	400	2.10(0.19)	1.96(0.19)	**1.77(0.16)**	1.82(0.17)	2.44(0.23)	2.90(0.21)	2.72(0.23)	2.61(0.20)	**2.57(0.20)**	2.88(0.25)
	1000	0.95(0.09)	0.80(0.07)	**0.69(0.06)**	0.78(0.06)	1.51(0.10)	1.39(0.11)	1.17(0.10)	1.10(0.09)	**1.08(0.08)**	1.55(0.11)
Exp(1)	200	2.95(0.29)	2.68(0.29)	**2.61(0.30)**	3.18(0.34)	6.02(0.52)	5.77(0.37)	5.23(0.36)	**4.97(0.35)**	5.06(0.39)	6.01(0.51)
	400	1.76(0.17)	1.56(0.16)	**1.45(0.15)**	1.84(0.18)	4.25(0.31)	4.30(0.39)	3.75(0.34)	3.60(0.33)	**3.40(0.31)**	4.02(0.37)
	1000	0.80(0.06)	0.67(0.05)	**0.61(0.05)**	0.94(0.07)	3.34(0.15)	1.85(0.12)	1.65(0.12)	**1.59(0.11)**	1.60(0.10)	2.35(0.14)

Declarations

Conflict of interest The authors have no conflict of interest to declare that are relevant to the content of this article.

Funding Jung's work has been partially supported by National Research Foundation of Korea (NRF) grants funded by the Korean government (MIST) (No. 2022R1F1A1071126 and No. RS-2022-NR068754).

Table 4: The mean of MSEs and its standard error (in the parenthesis) with 100 Monte Carlo simulations under Setting 4. The top half is the result of using a quantile smoothing spline, and the bottom half is the result of using a support vector quantile regression. Bold text indicates the lowest value. All values are multiplied by 10.

ϵ	n			$\tau = 0.1$					$\tau = 0.3$		
		Baseline	0.06	0.10	0.14	0.18	Baseline	0.14	0.20	0.30	0.40
	200	2.30(0.13)	1.88(0.10)	1.71(0.09)	**1.66(0.09)**	1.85(0.10)	10.82(0.80)	8.73(0.59)	8.16(0.53)	7.45(0.47)	**6.81(0.41)**
N(0,1)	400	1.08(0.08)	0.95(0.07)	0.90(0.06)	**0.87(0.06)**	0.92(0.05)	6.32(0.45)	4.89(0.32)	4.50(0.29)	3.84(0.24)	**3.30(0.20)**
	1000	0.51(0.03)	0.44(0.03)	0.41(0.02)	**0.39(0.02)**	0.54(0.03)	2.87(0.22)	2.24(0.15)	2.04(0.13)	1.77(0.10)	**1.72(0.10)**
	200	3.37(0.23)	2.86(0.16)	**2.74(0.16)**	2.87(0.19)	3.93(0.32)	1.13(0.09)	0.95(0.07)	0.87(0.06)	0.76(0.05)	**0.73(0.05)**
t(5)	400	2.13(0.16)	1.73(0.12)	**1.65(0.11)**	1.74(0.11)	2.78(0.19)	0.75(0.05)	0.62(0.04)	0.56(0.03)	0.51(0.03)	**0.51(0.04)**
	1000	0.99(0.07)	0.85(0.07)	**0.81(0.07)**	0.91(0.07)	1.85(0.12)	0.33(0.02)	0.28(0.02)	0.26(0.02)	**0.23(0.01)**	0.28(0.02)
	200	0.07(0.01)	0.05(0.00)	0.05(0.00)	0.04(0.00)	**0.04(0.00)**	0.27(0.02)	0.25(0.02)	0.24(0.02)	0.22(0.02)	**0.21(0.02)**
Exp(1)	400	0.04(0.00)	0.03(0.00)	0.03(0.00)	0.02(0.00)	**0.02(0.00)**	0.15(0.01)	0.13(0.01)	0.13(0.01)	0.12(0.01)	**0.11(0.01)**
	1000	0.02(0.00)	0.02(0.00)	0.02(0.00)	**0.01(0.00)**	0.17(0.16)	0.07(0.00)	0.06(0.00)	0.05(0.00)	**0.05(0.00)**	**0.05(0.02)**
	200	1.60(0.14)	1.37(0.13)	**1.35(0.13)**	1.39(0.13)	1.45(0.12)	0.66(0.06)	**0.61(0.06)**	0.62(0.06)	0.63(0.05)	0.65(0.05)
N(0,1)	400	0.79(0.06)	**0.67(0.05)**	0.68(0.05)	0.70(0.05)	0.75(0.05)	0.36(0.03)	**0.33(0.02)**	0.35(0.02)	0.37(0.02)	0.39(0.02)
	1000	0.37(0.03)	0.33(0.03)	**0.33(0.02)**	0.34(0.02)	0.44(0.03)	0.19(0.02)	**0.17(0.02)**	0.18(0.01)	0.18(0.01)	0.21(0.02)
	200	2.53(0.21)	2.17(0.19)	**2.14(0.18)**	2.32(0.19)	2.91(0.25)	0.69(0.07)	**0.60(0.07)**	0.61(0.06)	0.63(0.07)	0.69(0.07)
t(5)	400	1.52(0.12)	1.31(0.11)	**1.30(0.10)**	1.41(0.11)	1.89(0.15)	0.41(0.04)	**0.37(0.03)**	0.38(0.03)	0.40(0.03)	0.45(0.03)
	1000	0.74(0.07)	**0.65(0.06)**	0.67(0.05)	0.78(0.06)	1.35(0.09)	0.24(0.02)	**0.22(0.02)**	0.22(0.02)	0.23(0.02)	0.29(0.02)
	200	0.04(0.00)	0.03(0.00)	0.03(0.00)	**0.03(0.00)**	0.03(0.00)	0.18(0.01)	0.15(0.01)	**0.15(0.01)**	0.15(0.01)	0.16(0.01)
Exp(1)	400	0.02(0.00)	0.02(0.00)	**0.02(0.00)**	0.02(0.00)	0.02(0.00)	0.12(0.01)	**0.10(0.10)**	0.10(0.01)	0.10(0.01)	0.11(0.01)
	1000	0.01(0.00)	0.01(0.00)	**0.01(0.00)**	0.01(0.00)	0.01(0.00)	0.06(0.01)	**0.05(0.00)**	0.05(0.00)	0.06(0.00)	0.06(0.00)

ϵ	n			$\tau = 0.6$					$\tau = 0.8$		
		Baseline	0.10	0.30	0.50	0.70	Baseline	0.10	0.18	0.20	0.30
	200	0.68(0.07)	0.55(0.05)	4.87(0.05)	**0.47(0.04)**	0.56(0.04)	1.30(0.10)	1.13(0.08)	1.01(0.07)	0.97(0.06)	**0.87(0.05)**
N(0,1)	400	0.49(0.05)	0.33(0.03)	**0.27(0.02)**	**0.29(0.02)**	0.38(0.03)	0.77(0.06)	0.62(0.04)	0.56(0.03)	0.55(0.03)	**0.51(0.03)**
	1000	0.18(0.01)	0.13(0.01)	**0.11(0.01)**	0.14(0.01)	0.22(0.01)	0.35(0.03)	0.30(0.02)	0.26(0.02)	0.25(0.02)	**0.25(0.02)**
	200	0.75(0.07)	0.61(0.06)	**0.58(0.05)**	0.72(0.06)	1.07(0.09)	2.01(0.15)	1.71(0.13)	1.63(0.13)	**1.63(0.12)**	1.79(0.14)
t(5)	400	0.42(0.04)	0.33(0.03)	**0.30(0.02)**	0.42(0.30)	0.68(0.04)	0.96(0.07)	0.81(0.05)	**0.75(0.05)**	0.75(0.05)	0.88(0.05)
	1000	0.22(0.02)	0.17(0.01)	**0.14(0.01)**	0.21(0.01)	0.46(0.03)	0.45(0.03)	0.38(0.02)	**0.32(0.02)**	0.33(0.02)	0.45(0.03)
	200	0.98(0.08)	0.84(0.07)	**0.79(0.06)**	1.21(0.08)	0.22(0.13)	2.66(0.21)	2.23(0.14)	**2.05(0.13)**	2.06(0.13)	2.43(0.16)
Exp(1)	400	0.59(0.04)	0.47(0.03)	**0.45(0.03)**	0.83(0.04)	1.70(0.08)	1.54(0.12)	1.27(0.09)	1.22(0.09)	**1.21(0.09)**	1.54(0.10)
	1000	0.27(0.02)	**0.21(0.01)**	0.23(0.01)	0.63(0.03)	1.48(0.05)	0.68(0.05)	0.58(0.04)	**0.54(0.03)**	0.55(0.03)	0.89(0.05)
	200	0.56(0.05)	0.48(0.04)	**0.46(0.04)**	0.49(0.04)	0.60(0.05)	1.12(0.09)	0.97(0.07)	**0.92(0.07)**	0.91(0.07)	0.94(0.07)
N(0,1)	400	0.33(0.03)	0.30(0.03)	0.29(0.02)	**0.28(0.02)**	0.35(0.03)	0.61(0.05)	0.55(0.04)	**0.54(0.04)**	0.53(0.04)	0.56(0.04)
	1000	0.17(0.01)	0.15(0.01)	0.14(0.01)	**0.12(0.01)**	0.20(0.02)	0.28(0.02)	**0.25(0.02)**	0.25(0.02)	0.25(0.02)	0.29(0.02)
	200	0.57(0.06)	**0.49(0.06)**	0.52(0.05)	0.57(0.06)	0.83(0.08)	1.58(0.13)	**1.37(0.11)**	1.39(0.11)	1.41(0.11)	1.55(0.12)
t(5)	400	0.30(0.02)	**0.28(0.02)**	0.28(0.02)	0.30(0.02)	0.53(0.04)	0.88(0.08)	0.80(0.07)	**0.79(0.06)**	0.80(0.06)	0.90(0.06)
	1000	0.20(0.02)	0.18(0.01)	0.17(0.01)	**0.16(0.01)**	0.39(0.03)	0.37(0.03)	**0.35(0.03)**	0.36(0.03)	0.36(0.03)	0.50(0.03)
	200	0.75(0.06)	0.66(0.05)	**0.65(0.05)**	0.75(0.05)	1.49(0.10)	2.08(0.15)	**1.81(0.14)**	2.05(0.13)	2.04(0.13)	2.42(0.15)
Exp(1)	400	0.52(0.04)	**0.47(0.04)**	0.49(0.04)	0.63(0.04)	1.47(0.09)	1.43(0.12)	**1.31(0.10)**	1.33(0.10)	1.35(0.10)	1.54(0.10)
	1000	0.20(0.01)	**0.19(0.01)**	0.20(0.01)	0.63(0.03)	1.24(0.05)	0.55(0.04)	**0.54(0.03)**	0.49(0.03)	0.51(0.03)	0.79(0.05)

References

Bloznelis D, Claeskens G, Zhou J (2019) Composite versus model-averaged quantile regression. Journal of Statistical Planning and Inference 200:32–46

Bradic J, Fan J, Wang W (2011) Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73(3):325–349

Cannon AJ (2011) Quantile regression neural networks: Implementation in r and application to precipitation downscaling. Computers & geosciences 37(9):1277–1284

Jiang R, Qian WM, Zhou ZG (2018) Weighted composite quantile regression for partially linear varying coefficient models. Communications in Statistics - Theory and Methods 47(16):3987–4005. https://doi.org/10.1080/03610926.2017.1366522, URL https://doi.org/10.1080/03610926.2017.1366522, https://doi.org/10.1080/03610926.2017.1366522

Kai B, Li R, Zou H (2010) Local composite quantile regression smoothing: an efficient and safe alternative to local polynomial regression. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72(1):49–69

Koenker R (2005) Quantile Regression. Econometric Society Monographs, Cambridge University Press

Koenker R, Bassett Jr G (1978) Regression quantiles. Econometrica: journal of the Econometric Society pp 33–50

Koenker R, Zhao Q (1994) L-estimatton for linear heteroscedastic models. Journaltitle of Nonparametric Statistics 3(3-4):223–235

Koenker R, Ng P, Portnoy S (1994) Quantile smoothing splines. Biometrika 81(4):673–680

Meinshausen N, Ridgeway G (2006) Quantile regression forests. Journal of machine learning research 7(6)

Portnoy S, Koenker R (1989) Adaptive l-estimation for linear models. The Annals of Statistics 17(1):362–381

Pouplin T, Jeffares A, Seedat N, et al (2024) Relaxed quantile regression: Prediction intervals for asymmetric noise. URL https://arxiv.org/abs/2406.03258, 2406.03258

Sangnier M, Fercoq O, d'Alché Buc F (2016) Joint quantile regression in vector-valued rkhss. Advances in Neural Information Processing Systems 29

Silverman BW (1986) Density estimation for statistics and data analysis. Chapman & Hall/CRC monographs on statistics and applied probability, Chapman & Hall, London, URL https://cds.cern.ch/record/1070306

Takeuchi I, Le QV, Sears TD, et al (2006) Nonparametric quantile estimation. Journal of Machine Learning Research 7(45):1231–1264

Xu Q, Deng K, Jiang C, et al (2017) Composite quantile regression neural network with applications. Expert Systems with Applications 76:129–139

Xu Z, Zhao Z (2022) Efficient estimation for models with nonlinear heteroscedasticity. Journal of Business & Economic Statistics 40(4):1498–1508

Yin X, Fallah-Shorshani M, McConnell R, et al (2023) Quantile extreme gradient boosting for uncertainty quantification. URL https://arxiv.org/abs/2304.11732, 2304.11732

Zhao Z, Xiao Z (2014) Efficient regressions via optimally combining quantile information. Econometric theory 30(6):1272–1314

Zou H, Yuan M (2008) Composite quantile regression and the oracle model selection theory. The Annals of Statistics 36(3):1108–1126