Econometric Theory, **30**, 2014, 1272–1314.
doi:10.1017/S0266466614000176

EFFICIENT REGRESSIONS VIA OPTIMALLY COMBINING QUANTILE INFORMATION

ZHIBIAO ZHAO
Penn State University

ZHIJIE XIAO
Boston College

We develop a generally applicable framework for constructing efficient estimators of regression models via quantile regressions. The proposed method is based on optimally combining information over multiple quantiles and can be applied to a broad range of parametric and nonparametric settings. When combining information over a fixed number of quantiles, we derive an upper bound on the distance between the efficiency of the proposed estimator and the Fisher information. As the number of quantiles increases, this upper bound decreases and the asymptotic variance of the proposed estimator approaches the Cramér–Rao lower bound under appropriate conditions. In the case of nonregular statistical estimation, the proposed estimator leads to super-efficient estimation. We illustrate the proposed method for several widely used regression models. Both asymptotic theory and Monte Carlo experiments show the superior performance over existing methods.

1. INTRODUCTION

For regression estimations, the most widely used approach is the least squares (LS) method (for finite-dimensional models) or local LS method (in infinite-dimensional settings). If the data are normally distributed, the LS estimator has the likelihood interpretation and is the most efficient estimator. In the absence of Gaussianity, the LS estimation is usually less efficient than methods that exploit the distributional information, although may still be consistent under appropriate regularity conditions. Without these regularity conditions, the LS estimator may not even be consistent, for example, when the data have a heavy-tailed distribution such as the Cauchy distribution. Monte Carlo evidence indicates that the LS estimator can be quite sensitive to certain outliers. In empirical analysis, many applications (such as finance and economics) involve data with heavy-tailed or skewed

We thank Roger Koenker, Steve Portnoy, Victor Chernozhukov, two referees, and seminar attendants at University of Illinois for their very helpful comments. Zhao's research was partially supported by a NIDA grant P50-DA10075-15. Xiao thanks Boston College for research support. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDA or the NIH. Address correspondence to Zhijie Xiao, Department of Economics, Boston College, Chestnut Hill, MA 02467, USA and Shandong University, China; e-mail: zhijie.xiao@bc.edu.

© Cambridge University Press 2014

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

distributions, and the LS estimators may have poor performance in these cases. It is therefore important to develop robust and efficient estimation procedures for general innovation distributions.

If the underlying distribution were known, the Maximum Likelihood Estimator (MLE) could be constructed. Under regularity conditions, the MLE is asymptotic normal and asymptotically efficient in the sense that the limiting covariance matrix attains the Cramér–Rao lower bound. In practice, the true density function is generally unknown and so the MLE is not feasible. Nevertheless, the MLE and the Cramer-Rao bound serve as a standard against which we should measure our estimator.

For this reason, statisticians have devoted a great deal of research effort to the construction of estimation procedures that can extract distributional information from the data, and thus deliver more efficient estimators than the conventional LS method. For the location model $Y = \alpha + \varepsilon$, where ε has a symmetric density, the adaptive likelihood or score function based estimators of α were constructed in Beran (1974) and Stone (1975). Bickel (1982) further extended the idea to slope estimation of classical linear models. For nonlinear models, adaptive likelihood based estimations are usually technically challenging.

We believe that the quantile regression technique (Koenker and Bassett, 1978; Koenker, 2005) can provide a useful method in efficient statistical estimation. Intuitively, an estimation method that exploits the distributional information can potentially provide more efficient estimators. Since quantile regression provides a way of estimating the whole conditional distribution, appropriately using quantile regressions may improve estimation efficiency. Under regularity assumptions, the least-absolute-deviation (LAD) regression (i.e., quantile regression at median) can provide better estimators than the LS regression in the presence of heavy-tailed distributions. In addition, for certain distributions, a quantile regression at a nonmedian quantile may deliver a more efficient estimator than the LAD method. More importantly, additional efficiency gain can be achieved by combining information over multiple quantiles.

Although combining quantile regression over multiple quantiles can potentially improve estimation efficiency, this is often much easier to say than it is to do in a satisfactory way. To combine information from quantile regression, one may consider combining information over different quantiles via the criterion or loss function. For example, Zou and Yuan (2008) and Bradic, Fan, and Wang (2011) proposed the composite quantile regression (CQR) for parameter estimation and variable selection in the classical linear regression models. For nonparametric regression models, Kai, Li, and Zou (2010) proposed a local CQR estimation procedure, which is asymptotically equivalent to the local LS estimator as the number of quantiles increases. Alternatively, one may combine information based on estimators at different quantiles. Along this direction, Portnoy and Koenker (1989) studied asymptotically efficient estimation for the simple linear regression model. Although the proposed estimator is efficient asymptotically, it is not the best estimator with fixed quantiles. Also see Chamberlain (1994),

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

Xiao and Koenker (2009), and Chen, Linton, and Jacho-Chavez (2011) for related work on combination of estimators.

In this paper we consider regression estimation by combining information across k quantiles $\tau_j = j/(k+1)$, $j = 1, \ldots, k$. We show that for a wide range of parametric and nonparametric regression models, more efficient estimators can be constructed via optimally combining quantile regressions. We argue that it is essential to combine quantile information appropriately to achieve efficiency gain. In particular, simple averaging multiple quantile regression estimators is asymptotically equivalent to the LS method. We show that, by optimally combining information across quantiles$\tau_1, \ldots, \tau_k$, the efficiency of the proposed optimal weighted quantile average estimator is at most Φ_k away from the Fisher information, where Φ_k is defined as (43). As the number of quantiles $k \to \infty$, under appropriate regularity conditions, we have $\Phi_k \to 0$ and the estimator is asymptotically efficient. Interestingly, in the case of nonregular statistical estimation when these regularity conditions do not hold, the proposed estimators may lead to superefficient estimation.

The proposed methodology provides a generally applicable framework for constructing more efficient estimators under a broad variety of settings. For finite-dimensional parametric estimations, the method can be applied to construct efficient estimators for parameters in both linear and nonlinear regression models with homoscedastic errors and parameters in location-scale models with conditional heteroscedasticity. We show that, in the presence of conditional heteroscedasticity, some appropriate preliminary quantile regression is needed to improve the efficiency and to facilitate the quantile combination. Different restrictions (and thus optimal weights) are needed for estimation of the location parameters and scalar parameters. For nonparametric function estimations, the asymptotic bias of the proposed estimator is the same as that of the conventional nonparametric estimators (such as the local LS and the local LAD estimators) and meanwhile the inverse of the asymptotic variance is at most Φ_k away from the optimal Fisher information. Our extensive simulation studies show that the proposed method significantly outperforms the widely used LS, LAD, and the CQR method (Zou and Yuan, 2008; Kai, Li, and Zou, 2010) for both parametric and nonparametric models.

The rest of this paper is organized as follows: we provide a general discussion on the framework and assumptions for constructing efficient estimators based on quantile regressions in Section 2. Three leading cases of regressions are then investigated in Sections 3–5. In particular, we study the parametric regression models with homoscedasticity in Section 3. In Section 4, we study the parametric models with heteroscedasticity. Nonparametric models are investigated in Section 5. We focus on methodology and our discussions in Sections 3–5 consider the case with finite k. The case with k increasing to infinity is discussed in Section 6. Simulation studies are contained in Section 8, and an application to financial data is given in Section 9 to highlight the proposed method. Proofs are given in the Appendix.

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

2. MODEL SETUP, ASSUMPTIONS, AND THE WEIGHTED QUANTILE AVERAGE ESTIMATOR

We consider regression models of the following form:

$$Y = m(X) + \sigma(X)\varepsilon, \tag{1}$$

where (X, Y, ε) is the triplet of covariate, response, and noise, with ε independent of X. Here $m(\cdot)$ and $\sigma(\cdot)$ are two functions that depend on unknown parameters θ, where θ may be of finite dimension (parametric case) or infinite dimension (nonparametric case).

Denote by $Q_\varepsilon(\tau)$ the τ-th quantile of ε for $\tau \in (0, 1)$. Then the τ-th conditional quantile of Y given X, denoted by $Q_Y(\tau|X)$, is given by

$$Q_Y(\tau|X) = m(X) + \sigma(X)Q_\varepsilon(\tau). \tag{2}$$

As the inverse of the conditional distribution function, $Q_Y(\tau|X)$ fully captures the distributional relationship between Y and X. Intuitively, different distributional information may be obtained from different quantiles, and an appropriate combination of multiple quantiles may be more informative about the distribution than the conditional mean in the LS methods.

Throughout this paper we consider combining information over k equally spaced quantiles $\tau_j = j/(k+1)$, $j = 1, \ldots, k$. The discussion of this paper focuses on the case where k is assumed to be a given finite number. We consider the case that $k \to \infty$ increases with n in Section 6.3.

We briefly introduce the idea of our proposed estimator. Let θ be a parameter of interest. From the conditional quantile in (2), we can usually identify some perturbed version of θ, denoted by $\theta(\tau)$. Suppose there exists a class $\mathcal{W}$ of weights such that

$$\sum_{j=1}^{k} \omega_j \theta(\tau_j) = \theta, \qquad \omega = [\omega_1, \ldots, \omega_k]^T \in \mathcal{W}. \tag{3}$$

Given data on (X, Y), we can use a quantile regression based on (2) to obtain some consistent estimate, denoted by $\widehat{\theta}(\tau)$, of $\theta(\tau)$. In light of (3), we propose the following estimate of θ:

$$\widehat{\theta}_{\mathrm{WQAE}}(\omega) = \sum_{j=1}^{k} \omega_j \widehat{\theta}(\tau), \quad \omega \in \mathcal{W}. \tag{4}$$

We term $\widehat{\theta}_{\mathrm{WQAE}}(\omega)$ by the *weighted quantile average estimator* (WQAE) of θ. Since $\widehat{\theta}_{\mathrm{WQAE}}(\omega)$ is a consistent estimate of θ for each $\omega \in \mathcal{W}$, we propose in this paper estimation of θ based on $\omega \in \mathcal{W}$ that minimizes the limiting variance of $\widehat{\theta}_{\mathrm{WQAE}}(\omega)$ in parametric settings, or the asymptotic mean squared error in nonparametric settings. If ω is chosen in such a way, we call the corresponding estimator $\widehat{\theta}_{\mathrm{WQAE}}(\omega^*)$ the optimal WQAE (OWQAE).

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

The proposed estimation can be applied to a wide range of regression models. In this paper, we focus on the following three leading cases of regression (1):

Case 1: Parametric regression models with homoscedastic errors.

Case 2: Location-scale models with conditional heteroscedasticity in $\sigma(\cdot)$.

Case 3: Nonparametric regressions.

We study each of these cases in the following three sections 3–5.

Suppose we have samples $\{(X_t, Y_t)\}_{t=1}^n$ from (1) with corresponding noises $\{\varepsilon_t\}_{t=1}^n$. To facilitate the study of asymptotic theory, we impose the following assumptions.

Assumption 2.1. (i) $\{(X_t, \varepsilon_t)\}_{t\in\mathbb{Z}}$ is strictly stationary; for each t, ε_t is independent of $\{X_t, X_{t-1}, \ldots; \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$. (ii) $\{X_t\}_{t\in\mathbb{Z}}$ is an ergodic process.

Assumption 2.2. Denote by $f_\varepsilon(\cdot)$ and $F_\varepsilon(\cdot)$ the density and distribution functions of ε. f_ε is positive, twice differentiable, and bounded on $\{u : 0 < F_\varepsilon(u) < 1\}$.

Assumption 2.1 provides a convenient framework for studying asymptotic theory. First, strong mixing condition implies the ergodicity, and thus Assumption 2.1(ii) is weaker than the widely used strong mixing conditions in time series analysis. Next, we illustrate two useful properties below.

(P1) Martingale structure. Let $\mathcal{F}_t$ be the σ-algebra generated by $\{X_{t+1}, X_t, \ldots; \varepsilon_t, \varepsilon_{t-1}, \ldots\}$. By Assumption 2.1(i), ε_t is independent of $\mathcal{F}_{t-1}$. For any functions $h_1(\cdot)$ and $h_2(\cdot)$ such that $\mathbb{E}[|h_1(X_t)h_2(\varepsilon_t)|] < \infty$ and $\mathbb{E}[h_2(\varepsilon_t)] = 0$, we have $\mathbb{E}[h_1(X_t)h_2(\varepsilon_t)|\mathcal{F}_{t-1}] = h_1(X_t)\mathbb{E}[h_2(\varepsilon_t)] = 0$. Therefore,$\{h_1(X_t)h_2(\varepsilon_t)\}_{t\in\mathbb{Z}}$ are martingale differences with respect to the filtration $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$, and consequently $\sum_{t=1}^n h_1(X_t)h_2(\varepsilon_t)$ is a martingale.

(P2) Law of large numbers. By ergodic theorem (Thm. 3.5.7, Stout, 1974), for any function $\ell(\cdot)$ such that $\mathbb{E}[|\ell(X_t)|] < \infty$, the ergodicity in Assumption 2.1(ii) implies

$$\lim_{n\to\infty} \frac{1}{n}\sum_{t=1}^{n} \ell(X_t) = \mathbb{E}[\ell(X_0)], \quad \text{in probability.} \tag{5}$$

In subsequent sections we adopt the following notation. For a random vector Z, we use $\|Z\|$ to signify the Euclidean norm of Z, and write $Z \in \mathcal{L}^q$, $q > 0$, if $\mathbb{E}(\|Z\|^q) < \infty$.

3. HOMOSCEDASTIC PARAMETRIC REGRESSION MODELS

In this section we study the parametric regression model (case 1 in Section 2). Corresponding to the general representation (1), let $\sigma(X) \equiv 1$ and $m(X) = \alpha + m(X, \beta)$, where $\beta \in \mathbb{R}^p$ is the vector of unknown parameters, and the intercept α is added to ensure the identifiability of β, we have

$$Y = \alpha + m(X, \beta) + \varepsilon. \tag{6}$$

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

We are interested in the estimation of β.

By (2), we have the conditional quantile representation

$$Q_Y(\tau|X) = \alpha(\tau) + m(X,\beta), \quad \text{where} \quad \alpha(\tau) = \alpha + Q_\varepsilon(\tau). \tag{7}$$

Given samples $\{(X_t, Y_t)\}_{t=1}^n$ from (6), let $(\widehat{\alpha}(\tau), \widehat{\beta}(\tau))$ be an estimator of $(\alpha(\tau), \beta)$ from a quantile regression based on (7):

$$\left(\widehat{\alpha}(\tau), \widehat{\beta}(\tau)\right) = \underset{(a,b)}{\operatorname{argmin}} \sum_{t=1}^{n} \rho_\tau\{Y_t - a - m(X_t, b)\}, \tag{8}$$

where $\rho_\tau(z) = z(\tau - \mathbf{1}_{z\le 0})$ and $\mathbf{1}$ is the indicator function. Denote by $\dot{m}(x,\beta)$ is the partial derivative vector of $m(x,\beta)$ with respect to β. Define

$$D_n = \begin{bmatrix} 1\ \dot{m}(X_1,\beta)^T \\ \vdots \quad \vdots \\ 1\ \dot{m}(X_n,\beta)^T \end{bmatrix} \in \mathbb{R}^{n\times(p+1)} \quad \text{and} \quad Z_t = \begin{bmatrix} 1 \\ \dot{m}(X_t,\beta) \end{bmatrix} \in \mathbb{R}^{p+1}.$$

Similar to the linear regression, D_n serves as the design matrix and Z_t is the equivalent covariate corresponding to observation t for the quantile regression (8). A leading example is the classical linear regression model (see, e.g., Koenker, 1984), corresponding to $m(X,\beta) = X^T\beta$. In this case, $Q_Y(\tau|X) = [\alpha + Q_\varepsilon(\tau)] + X^T\beta$ and $\dot{m}(X,\beta) = X$.

Assumption 3.1. The quantile regression estimator has the Bahadur representation

$$\begin{bmatrix} \widehat{\alpha}(\tau) \\ \widehat{\beta}(\tau) \end{bmatrix} = \begin{bmatrix} \alpha(\tau) \\ \beta \end{bmatrix} + \frac{\left(D_n^T D_n\right)^{-1}}{f_\varepsilon(Q_\varepsilon(\tau))} \sum_{t=1}^{n} Z_t\left[\tau - \mathbf{1}_{\varepsilon_t < Q_\varepsilon(\tau)}\right] + o_p\left(n^{-1/2}\right), \tag{9}$$

uniformly in $\tau \in \mathcal{T} := [\delta, 1-\delta]$ with some small constant $\delta > 0$.

Assumption 3.1 is an asymptotic representation of the quantile regression estimator. Under regularity conditions on the regression function $m(\cdot)$, error density, and the parameter space, a Bahadur representation can be obtained over τ on a subset of [0,1]. See, e.g., Portnoy and Koenker (1989), Jurečková and Procházka (1994), and He and Shao (1996) for related study. Also see Section 4 for discussions on the conditional heteroscedastic parametric models.

Since β does not depend on τ, we can use $\widehat{\beta}(\tau)$ to estimate β with any choice of τ. By Theorem 3.1 (also, see the definition of Σ_β there),

$$\sqrt{n}\left[\widehat{\beta}(\tau) - \beta\right] \Rightarrow N\left(0, \Sigma_\beta^{-1}\frac{\tau(1-\tau)}{f_\varepsilon^2(Q_\varepsilon(\tau))}\right). \tag{10}$$

For example, the case with $\tau = 0.5$ corresponds to the median quantile regression or LAD estimation of β.

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

As discussed in Section 2, we want to combine information over the k quantiles $\tau_j = j/(k+1)$, $j = 1, \ldots, k$, where k is assumed to be a given finite number such that $\tau_j \in \mathcal{T}$. Since $\widehat{\beta}(\tau)$ is a consistent estimate of β, from (3)–(4), we consider the WQAE of β:

$$\widehat{\beta}_{\mathrm{WQAE}}(\omega) = \sum_{j=1}^{k} \omega_j \widehat{\beta}(\tau_j), \qquad \text{where} \qquad \sum_{j=1}^{k} \omega_j = 1. \tag{11}$$

THEOREM 3.1. *Suppose Assumptions 2.1–3.1 hold and* $\dot{m}(X, \beta) \in \mathcal{L}^2$ *with* $X \overset{d}{=} X_t$. *Then*

$$\sqrt{n}\left[\widehat{\beta}_{\mathrm{WQAE}}(\omega) - \beta\right] \Rightarrow N\left(0, \Sigma_\beta^{-1} S(\omega)\right), \tag{12}$$

with $\Sigma_\beta = \mathbb{E}\left[\dot{m}(X, \beta)\dot{m}(X, \beta)^T\right] - \mathbb{E}[\dot{m}(X, \beta)]\mathbb{E}\left[\dot{m}(X, \beta)^T\right]$ *assumed to be nonsingular, and*

$$S(\omega) = \omega^T H \omega \quad \text{with} \quad H = \left\{ \frac{\min(\tau_j, \tau_{j'}) - \tau_j \tau_{j'}}{f_\varepsilon(Q_\varepsilon(\tau_j)) f_\varepsilon(Q_\varepsilon(\tau_{j'}))} \right\}_{1 \le j, j' \le k} \in \mathbb{R}^{k \times k}. \tag{13}$$

The proposed estimator, the OWQAE, of β is obtained by choosing ω to minimize the asymptotic variance of $\widehat{\beta}_{\mathrm{WQAE}}(\omega)$.

THEOREM 3.2. *Under the assumptions of Theorem 3.1, the optimal weight is*

$$\omega^* = \underset{\omega_1 + \cdots + \omega_k = 1}{\mathit{argmin}} \; S(\omega) = \frac{H^{-1} e_k}{e_k^T H^{-1} e_k}, \qquad \text{where} \qquad e_k = (1, \ldots, 1)^T. \tag{14}$$

With ω^* *in (14), the OWQAE of* β *has the following limiting distribution:*

$$\sqrt{n}[\widehat{\beta}_{\mathrm{WQAE}}(\omega^*) - \beta] \Rightarrow N(0, \Sigma_\beta^{-1} \Omega_k^{-1}), \quad \text{where} \quad \Omega_k = e_k^T H^{-1} e_k. \tag{15}$$

Remark 3.1. A quick way of combining quantile information is to take a simple average of the quantile regression estimators. This is easy to implement and has been used in the literature (see, e.g., Kai, Li, and Zou, 2010 for nonparametric estimation) as a method of combining quantile information. If we use $\omega = [1/k, \ldots, 1/k]^T$ in (11), the resulting unweighted estimator has the asymptotic normality in Theorem3.1 with $S(\omega)$ replaced by

$$R_k := \frac{1}{k^2} \sum_{j=1}^{k} \sum_{j'=1}^{k} \frac{\min(\tau_j, \tau_{j'}) - \tau_j \tau_{j'}}{f_\varepsilon(Q_\varepsilon(\tau_j)) f_\varepsilon(Q_\varepsilon(\tau_{j'}))}. \tag{16}$$

Clearly, $R_k \ge S(\omega^*)$. See Section 6.2 for more discussions on the property of R_k.

We compute ω^* for some examples below using $k = 9$ quantiles $0.1, 0.2, \ldots, 0.9$.

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

Example 1. Let ε be Student-t distributed. For t_1 (Cauchy distribution), the optimal weight $\omega^* = \{-0.03,-0.04,0.08,0.29,0.40,0.29,0.08,-0.04,-0.03\}$, quantiles $\tau = 0.4, 0.5, 0.6$ contribute almost all information whereas quantiles $\tau = 0.1, 0.2, 0.8, 0.9$ have negative weights, so the unweighted quantile average estimator would not perform well. However, for N(0,1), $\omega^* = \{0.13,0.11,0.11,0.10,0.10,0.10,0.11,0.11,0.13\}$ are close to the uniform weights, and thus the OWQAE, the unweighted quantile average estimator, and the LS estimator have comparable performance.

Example 2. Let ε have normal mixture distributions. For Mixture 1: 0.5N(0,1)+0.5N(0,0.5^6) (different variances), $\omega^* = \{-0.002,-0.102, 0.183,0.277,0.287,0.277,0.183,-0.102,-0.002\}$, quantiles $0.3, \ldots, 0.7$ contain substantial information whereas quantiles 0.2 and 0.8 have negative weight. For Mixture 2: 0.5N(−2,1)+0.5N(2,1) (different means), $\omega^* = \{0.185,0.156,0.153,0.078,-0.144,0.078,0.153,0.156,0.185\}$, quantiles $\tau = 0.1, 0.2, 0.3, 0.7, 0.8, 0.9$ are comparable while the median performs the worst.

Example 3. Let ε be Gamma random variable with parameter $d > 0$. For $d = 1$ (exponential distribution), $\omega_1^* = 1.152, \omega_2^* = -0.124$, and $\omega_i^* \approx 0$ for $i = 3, \ldots, 9$. Quantiles 0.1 and 0.2 contain almost all information.

As shown in Examples 1–3, different quantiles may carry substantially different amount of information, and inappropriately utilizing such information may result in a significant loss of efficiency. The latter phenomenon provides strong evidence in favor of our proposed optimally weighted quantile based estimators.

In practice, the optimal weight ω^* in (14), which depends on the sparsity or quantile-density function $f_\varepsilon(Q_\varepsilon(\tau))$, needs to be estimated. We make the following assumption on the estimate, denoted by $\widehat{f_\varepsilon}(\widehat{Q}_\varepsilon(\tau))$, of $f_\varepsilon(Q_\varepsilon(\tau))$.

Assumption 3.2. $\sup_{\tau\in\mathcal{T}} |\widehat{f_\varepsilon}(\widehat{Q}_\varepsilon(\tau)) - f_\varepsilon(Q_\varepsilon(\tau))| = o_p(1)$ for $\mathcal{T}$ in Assumption 3.1.

Plugging the consistent estimate $\widehat{f_\varepsilon}(\widehat{Q}_\varepsilon(\tau))$ of $f_\varepsilon(Q_\varepsilon(\tau))$ into the matrix H in (14), we can obtain the following consistent estimate of the optimal weight ω^*:

$$\widehat{\omega}^* = \frac{\widehat{H}^{-1}e_k}{e_k^T\widehat{H}^{-1}e_k}, \qquad \text{where} \qquad \widehat{H} = \left\{\frac{\min(\tau_j,\tau_{j'}) - \tau_j\tau_{j'}}{\widehat{f_\varepsilon}(\widehat{Q}_\varepsilon(\tau_j))\widehat{f_\varepsilon}(\widehat{Q}_\varepsilon(\tau_{j'}))}\right\}_{1\le j,j'\le k}. \quad (17)$$

Theorem 3.3 asserts that $\widehat{\beta}_{\mathrm{WQAE}}(\widehat{\omega}^*)$ with the estimated weight $\widehat{\omega}^*$ achieves the same efficiency of $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$.

THEOREM 3.3. *Under the assumptions of Theorem 3.1 and Assumption 3.2, we have*

$$\sqrt{n}\left[\widehat{\beta}_{\mathrm{WQAE}}(\widehat{\omega}^*) - \beta\right] \Rightarrow N\left(0, \Sigma_\beta^{-1}\Omega_k^{-1}\right). \quad (18)$$

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

4. THE LOCATION-SCALE MODELS

Another class of widely used regression models is the location-scale models (case 2 in Section 2) that allow for conditional heteroscedasticity. There is a large literature in econometrics and statistics studying the location-scale models. Koenker and Zhao (1994) studied L estimation of a location-scale model in the following form:

$$Y_t = X_t^T \beta + \sigma_t \varepsilon_t, \quad \text{where} \quad \sigma_t = X_t^T \gamma, \tag{19}$$

under the condition $X_t^T \gamma > 0$. Zhao (2001) studied asymptotically efficient median regression using the k-nearest neighbors method. In this section we study the location-scale models via optimal quantile combination.

In the model (19), the positive constraint $X_t^T \gamma > 0$ is somewhat restrictive to allow for flexible applications. For example, it is violated for normally distributed covariates X. For this reason, many researchers consider an alternative form of σ_t which can be expressed as a linear function of absolute values of the regressors and other variables:

$$Y_t = X_t^T \beta + \sigma_t \varepsilon_t, \quad \text{where} \quad \sigma_t = U_t^T \gamma := \sigma(U_t), \tag{20}$$

where U_t is a vector of absolute values of the regressors and other covariates (see, e.g., Koenker and Zhao, 1996 for studies on related models). For example, let $X_t^T = (x_{t1}, \cdots, x_{tp})$, one may consider a location-scale model with $\sigma_t = \sigma(U_t) = U_t^T \gamma = \gamma_0 + \sum_{j=1}^p \gamma_j |x_{tj}|$, where $U_t^T = (1, |x_{t1}|, \cdots, |x_{tp}|)$, $\gamma = (\gamma_0, \gamma_1, \cdots, \gamma_p)$, $\gamma_0 > 0$, $\gamma_1 \geq 0, \cdots, \gamma_p \geq 0$.

In this section we consider the location-scale regression model (20).[1] We are interested in the estimation of β and γ. By (2), we have the conditional quantile representation

$$Q_Y(\tau|X) = X^T \beta + U^T \gamma(\tau) \quad \text{with} \quad \gamma(\tau) = \gamma\, Q_\varepsilon(\tau). \tag{21}$$

Given a sample of size n, we may estimate $(\beta, \gamma(\tau))$ using a quantile regression similar to (8). However, in the presence of conditional heteroscedasticity, it is more efficient to use a weighted quantile regression with the weights reflecting the conditional heteroscedasticity. In addition, the weighted quantile regression estimates have nice properties that help combining quantile information.

Thus, following the idea of Koenker and Zhao (1994), we consider the weighted quantile regression:

$$(\widehat{\beta}(\tau), \widehat{\gamma}(\tau)) = \underset{(b,r)}{\operatorname{argmin}} \sum_{t=1}^n \frac{1}{\widetilde{\sigma}_t} \rho_\tau \{Y_t - X_t^T b - U_t^T r\}, \tag{22}$$

where $\widetilde{\sigma}_t = U_t^T \widetilde{\gamma}$ and $\widetilde{\gamma}$ is a consistent estimate of γ.

Assumption 4.1. (i) $\{(X_t, U_t, \varepsilon_t)\}_{t\in\mathbb{Z}}$ is strictly stationary; for each t, ε_t is independent of $\{(X_t, U_t), (X_{t-1}, U_{t-1}), \ldots; \varepsilon_{t-1}, \varepsilon_{t-2}, \ldots\}$. (ii) $\{(X_t, U_t)\}_{t\in\mathbb{Z}}$ is an ergodic process.

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

Assumption 4.2. (i) $\|X_t\| + \|U_t\| \le c_1$ and $U_t^T \gamma \ge c_2$ for some constants $c_1, c_2 > 0$. (ii) $\widetilde{\gamma} - \gamma = o_p(n^{-1/4})$. (iii) Let (X, U) be distributed as (X_t, U_t). Define the matrices

$$M_1 = \mathbb{E}\left[\frac{XX^T}{\sigma(U)^2}\right], \quad M_2 = \mathbb{E}\left[\frac{XU^T}{\sigma(U)^2}\right], \quad M_3 = \mathbb{E}\left[\frac{UU^T}{\sigma(U)^2}\right],$$

and

$$M_\beta = M_1 - M_2 M_3^{-1} M_2^T, \qquad M_\gamma = M_3 - M_2^T M_1^{-1} M_2.$$

The matrices M_1, M_3, M_β, and M_γ are nonsingular.

Assumption 4.1 is a modification of Assumption 2.1 by allowing for more covariates (X_t, U_t). In Assumption 4.2, (i) is imposed simply for technical convenience and can be replaced by some finite moment conditions, (ii) requires that $\widetilde{\gamma}$ must be reasonably close to γ, and (iii) is used to avoid some singular design matrix.

THEOREM 4.1. *Suppose Assumptions 2.2, 4.1, and 4.2 hold. Then we have*

$$\begin{bmatrix} \widehat{\beta}(\tau) \\ \widehat{\gamma}(\tau) \end{bmatrix} = \begin{bmatrix} \beta \\ \gamma(\tau) \end{bmatrix} + \frac{1}{n f_\varepsilon(Q_\varepsilon(\tau))} \sum_{t=1}^{n} Z_t \left[\tau - \mathbf{1}_{\varepsilon_t < Q_\varepsilon(\tau)}\right] + o_p\left(n^{-1/2}\right), \tag{23}$$

with

$$Z_t = \begin{bmatrix} M_\beta^{-1}(X_t - M_2 M_3^{-1} U_t)/\sigma(U_t) \\ M_\gamma^{-1}(U_t - M_2^T M_1^{-1} X_t)/\sigma(U_t) \end{bmatrix}.$$

We now construct estimators of β and γ by optimally combining information over quantiles $\tau_1, \ldots, \tau_k$.

First, we consider estimation of β. As in Section 3, we consider the WQAE $\widehat{\beta}_{\mathrm{WQAE}}(\omega)$ given by (11). Using the Bahadur representation (23) from Theorem 4.1, the same argument in Theorem 3.1 yields

$$\sqrt{n}\left[\widehat{\beta}_{\mathrm{WQAE}}(\omega) - \beta\right] \Rightarrow N\left(0, M_\beta^{-1} S(\omega)\right),$$

with $S(\omega)$ given in Theorem 3.1. Therefore, the optimal weight can be constructed in a similar way as described by Theorem 3.2, and ω^* is given by (14). The OWQAE has the following limiting distribution

$$\sqrt{n}\left[\widehat{\beta}_{\mathrm{WQAE}}(\omega^*) - \beta\right] \Rightarrow N\left(0, M_\beta^{-1} \Omega_k^{-1}\right),$$

with Ω_k given in (15). If we use the estimated optimal weight $\widehat{\omega}^*$ in (17), under the additional Assumption 3.2, the conclusion in Theorem 3.3 also holds here.

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

Next, we consider estimation of the scale parameter γ via quantile combination. As will be clear from later analysis, the construction of WQAE and choice of optimal weights related to the scale parameter will be different from those of β. For this reason, we denote the weights used in γ estimation by $\pi = [\pi_1, \ldots, \pi_k]^T$. From (21)–(22), $\widehat{\gamma}(\tau)$ is an estimation of $\gamma(\tau) = \gamma Q_\varepsilon(\tau)$. Then, for any π satisfying $\sum_{j=1}^k \pi_j Q_\varepsilon(\tau_j) = 1$, $\sum_{j=1}^k \pi_j \gamma(\tau_j) = \gamma$. Therefore, we propose the following WQAE of γ:

$$\widehat{\gamma}_{\mathrm{WQAE}}(\pi) = \sum_{j=1}^k \pi_j \widehat{\gamma}(\tau_j), \qquad \text{where} \qquad \sum_{j=1}^k \pi_j Q_\varepsilon(\tau_j) = 1. \tag{24}$$

THEOREM 4.2. *Under the assumptions in Theorem 4.1, we have the asymptotic normality:*

$$\sqrt{n}\left[\widehat{\gamma}_{\mathrm{WQAE}}(\pi) - \gamma\right] \Rightarrow N\left(0, M_\gamma^{-1} S(\pi)\right),$$

where $S(\pi) = \pi^T H \pi$ *with* H *defined in (13). Furthermore, the optimal weight is*

$$\begin{aligned}\pi^* &= \underset{\pi_1 Q_\varepsilon(\tau_1) + \cdots + \pi_k Q_\varepsilon(\tau_k) = 1}{\mathit{argmin}} S(\pi) \\ &= \frac{H^{-1} q}{q^T H^{-1} q}, \quad \mathit{where} \quad q = [Q_\varepsilon(\tau_1), \ldots, Q_\varepsilon(\tau_k)]^T. \end{aligned} \tag{25}$$

With π^* *in (25), the OWQAE of* γ *has the following limiting distribution:*

$$\sqrt{n}\left[\widehat{\gamma}_{\mathrm{WQAE}}(\pi^*) - \gamma\right] \Rightarrow N\left(0, M_\gamma^{-1} \Lambda_k^{-1}\right), \quad \mathit{where} \quad \Lambda_k = q^T H^{-1} q. \tag{26}$$

Therefore, the optimal weights for the OWQAE of β and γ are different, and their corresponding OWQAEs have different efficiency. This is due to the structure of the conditional quantile representation (21): β does not depend on the quantile τ whereas γ relies on τ through the coefficient $Q_\varepsilon(\tau)$.

Similar to the case of β, the conclusion in Theorem 3.3 also holds for $\widehat{\gamma}_{\mathrm{WQAE}}(\widehat{\pi}^*)$ when we use estimated optimal weight $\widehat{\pi}^*$ by plugging in consistent estimates of (q, H) into (25).

To implement the weighted quantile regression (22), we need to find a consistent estimate $\widetilde{\gamma}$ of γ. We propose the following procedure:

(i) For each quantile $\tau = \tau_1, \ldots, \tau_k$, fit the unweighted quantile regression

$$\left(\widetilde{\beta}(\tau), \widetilde{\gamma}(\tau)\right) = \underset{(b,r)}{\mathrm{argmin}} \sum_{t=1}^n \rho_\tau \left\{Y_t - X_t^T b - U_t^T r\right\}. \tag{27}$$

By the same argument in Theorem 4.1, $\left(\widetilde{\beta}(\tau), \widetilde{\gamma}(\tau)\right) = (\beta, \gamma(\tau)) + O_p\left(n^{-1/2}\right)$.

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

(ii) Let

$$\widetilde{\gamma} = \sum_{j=1}^{k} |\widetilde{\gamma}(\tau_j)|. \tag{28}$$

Then $\widetilde{\gamma} = \gamma \sum_{j=1}^{k} |Q_\varepsilon(\tau_j)| + O_p\left(n^{-1/2}\right)$. Note that, in (22), it suffices for $\widetilde{\gamma}$ to estimate γ up to a multiplication factor. Thus, $\widetilde{\gamma}$ in (28) satisfies Assumption 4.2(ii).

Finally, we point out an identifiability issue of the optimal weight π^* in (25). Since $Q_\varepsilon(\tau)$ is identifiable up to a scale factor, if we multiply $Q_\varepsilon(\tau)$ by a constant c, π^* and hence $\widehat{\gamma}_{\mathrm{WQAE}}(\pi^*)$ in (24) will be multiplied by a factor $1/c$. This is due to the nonidentifiability of the parameter γ in (20). To ensure identifiability, we may impose some constraint on ε; see Section 7 for more discussions on estimating π^*.

5. NONPARAMETRIC REGRESSIONS

In this section we study the nonparametric regression (case 3 in Section 2). We assume that both $m(\cdot)$ and $\sigma(\cdot)$ in (1) are nonparametric functions, and we are interested in the estimation of $m(\cdot)$. Although our theory is also applicable for multivariate case, to avoid the issue of "curse of dimensionality", we consider the univariate case $X \in \mathbb{R}$.

Recall the conditional quantile $Q_Y(\tau|X)$ in (2). Without further assumptions, we cannot identify $m(X)$ from $Q_Y(\tau|X)$ at a single quantile. To ensure identifiability, we assume that ε has a symmetric density, which is satisfied for many commonly used distributions, such as normal distribution, Student-t distribution, Cauchy distribution, uniform distribution on a symmetric interval, Laplace distribution, symmetric stable distribution, many normal mixture distributions, and their truncated versions on symmetric intervals.

Consider weights $\omega_1, \ldots, \omega_k$ satisfying the constraints

$$\sum_{j=1}^{k} \omega_j = 1 \quad \text{and} \quad \omega_j = \omega_{k+1-j},\ j = 1, \ldots, k. \tag{29}$$

Under the symmetric density assumption above, $Q_\varepsilon(\tau) + Q_\varepsilon(1-\tau) = 0$. Therefore, with quantiles $\tau_j = j/(k+1)$ and using (2) and (29), we have

$$\sum_{j=1}^{k} \omega_j Q_Y(\tau_j|X) = \sum_{j=1}^{k} \omega_j \left[m(X) + \sigma(X) Q_\varepsilon(\tau_j)\right] = m(X). \tag{30}$$

This identity suggests estimation of $m(\cdot)$ by plugging in consistent estimation of $Q_Y(\tau_j|X)$.

Given samples $\{(X_t, Y_t)\}_{t=1}^{n}$, we can estimate $Q_Y(\tau|x)$ by the local linear quantile regression (Yu and Jones, 1998):

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

$$\left(\widehat{Q}_Y(\tau|x), \widehat{b}\right) = \underset{(a,b)}{\operatorname{argmin}} \sum_{t=1}^{n} \rho_\tau\{Y_t - a - b(X_t - x)\} K\left(\frac{X_t - x}{h}\right), \tag{31}$$

for a kernel function $K(\cdot)$ and bandwidth h. From (30), we propose the WQAE of $m(x)$:

$$\widehat{m}_{\mathrm{WQAE}}(x|\omega) = \sum_{j=1}^{k} \omega_j \widehat{Q}_Y(\tau_j|x). \tag{32}$$

Assumption 5.1. (i) f_ε is symmetric, positive, and twice continuously differentiable on its support; the density function $p_X(\cdot) > 0$ of X is differentiable, $m(\cdot)$ is three times differentiable, and $\sigma(\cdot) > 0$ is differentiable, in the neighborhood of x. (ii) $nh \to \infty$ and $nh^9 \to 0$. (iii) $K(\cdot)$ integrates to one, is symmetric, and has bounded support. Write

$$\mu_K = \int_{\mathbb{R}} u^2 K(u) du, \qquad \varphi_K = \int_{\mathbb{R}} K^2(u) du,$$

THEOREM 5.1. *Suppose Assumptions 2.1 and 5.1 hold. Let $S(\omega)$ be defined in (13). Then*

$$\sqrt{nh}\left\{\widehat{m}_{\mathrm{WQAE}}(x|\omega) - m(x) - \frac{1}{2}m''(x)\mu_K h^2\right\} \Rightarrow N\left(0, \frac{\varphi_K \sigma^2(x)}{p_X(x)} S(\omega)\right). \tag{33}$$

Furthermore, ω^* in (14) minimizes $S(\omega)$ subject to the constraints (29), and

$$\sqrt{nh}\left\{\widehat{m}_{\mathrm{WQAE}}\left(x|\omega^*\right) - m(x) - \frac{1}{2}m''(x)\mu_K h^2\right\} \Rightarrow N\left(0, \frac{\varphi_K \sigma^2(x)}{p_X(x)} \Omega_k^{-1}\right), \tag{34}$$

where Ω_k is defined in (15).

For comparison, we briefly review some alternative nonparametric estimation methods. The widely used local linear LS regression estimator, denoted by $\widehat{m}_{\mathrm{LS}}(x)$, is obtained by replacing the quantile loss $\rho_\tau(\cdot)$ in (31) with the square loss. If $\mathbb{E}(\varepsilon_t) = 0$ and $\mathrm{var}(\varepsilon_t) < \infty$, under some regularity conditions (Fan and Gijbels, 1996),

$$\sqrt{nh}\left\{\widehat{m}_{\mathrm{LS}}(x) - m(x) - \frac{1}{2}m''(x)\mu_K h^2\right\} \Rightarrow N\left(0, \frac{\varphi_K \sigma^2(x)}{p_X(x)} \mathrm{var}(\varepsilon)\right). \tag{35}$$

When ε_t's are Gaussian, the local LS estimation corresponds to the local likelihood criterion. In the absence of Gaussianity, asymptotic results of $\widehat{m}_{\mathrm{LS}}(x)$ generally still hold but this estimator is less efficient in terms of mean-squared error than estimators that exploit the distributional information. For heavy-tailed data,

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

local quantile regression is a robust estimation method; see, e.g., Yu and Jones (1998). The local median regression estimator, denoted by $\widehat{m}_{\mathrm{LAD}}(x)$, corresponds to $\tau = 0.5$ in (31). By Theorem5.1,

$$\sqrt{nh}\left\{\widehat{m}_{\mathrm{LAD}}(x) - m(x) - \frac{1}{2}m''(x)\mu_K h^2\right\} \Rightarrow N\left(0, \frac{\varphi_K \sigma^2(x)}{p_X(x)} \frac{1}{4f_\varepsilon^2(0)}\right). \tag{36}$$

Recently, Kai, Li, and Zou (2010) proposed a local composite quantile regression (CQR) estimator which takes a simple average of multiple quantile estimations. The local CQR estimator, denoted by $\widehat{m}_{\mathrm{CQR}}(x)$, has the asymptotic normality

$$\sqrt{nh}\left\{\widehat{m}_{\mathrm{CQR}}(x) - m(x) - \frac{1}{2}m''(x)\mu_K h^2\right\} \Rightarrow N\left(0, \frac{\varphi_K \sigma^2(x)}{p_X(x)} R_k\right), \tag{37}$$

where R_k is defined in (16). Intuitively, $\widehat{m}_{\mathrm{LS}}(x)$ uses information from the local sample average, $\widehat{m}_{\mathrm{LAD}}(x)$ uses information from the local sample median, $\widehat{m}_{\mathrm{CQR}}(x)$ uses information from multiple quantiles with uniform weight, and the proposed OWQAE $\widehat{m}_{\mathrm{WQAE}}(x|\omega^*)$ combines information from multiple quantiles optimally.

If the error density f_ε were known, we could replace the quantile loss $\rho_\tau(\cdot)$ in (31) by the log likelihoodlog $f_\varepsilon(Y_t - a - b(X_t - x))$ and obtain a likelihood-based estimator, denoted by $\widehat{m}_{\mathrm{MLE}}(x)$, see, e.g., Fan, Farman, and Gijbels (1998). Under appropriate conditions,

$$\sqrt{nh}\left\{\widehat{m}_{\mathrm{MLE}}(x) - m(x) - \frac{1}{2}m''(x)\mu_K h^2\right\} \Rightarrow N\left(0, \frac{\varphi_K \sigma^2(x)}{p_X(x)} \mathcal{I}(f_\varepsilon)^{-1}\right), \tag{38}$$

where $\mathcal{I}(f_\varepsilon)$ is the Fisher information of f_ε. Under some regularity conditions, the local likelihood estimator is the most efficient estimator. In practice, f_ε is unknown and $\widehat{m}_{\mathrm{MLE}}(x)$ is infeasible. In Section 6.2, it is shown that $\Omega_k \to \mathcal{I}(f_\varepsilon)$, and therefore the optimal WQAE $\widehat{m}_{\mathrm{WQAE}}(x|\omega^*)$ achieves the same asymptotic efficiency of the infeasible estimator $\widehat{m}_{\mathrm{MLE}}(x)$.

We now compare the efficiency of $\widehat{m}_{\mathrm{WQAE}}(x|\omega^*)$ to $\widehat{m}_{\mathrm{LS}}(x)$, $\widehat{m}_{\mathrm{LAD}}(x)$, and $\widehat{m}_{\mathrm{CQR}}(x)$. From (34)–(37), all four estimators have the asymptotic normality with different s^2:

$$\sqrt{nh}\left\{\widehat{m}(x) - m(x) - \frac{1}{2}m''(x)\mu_K h^2\right\} \Rightarrow N\left(0, \frac{\varphi_K \sigma^2(x)}{p_X(x)} s^2\right).$$

Define the asymptotic mean-squared error (AMSE) as $\mathrm{AMSE}\{\widehat{m}(x)|h\} = \left[m''(x)\mu_K h^2/2\right]^2 + \varphi_K \sigma^2(x)s^2/[nhp_X(x)]$. Minimizing the AMSE, we obtain the optimal bandwidth:

$$\begin{aligned} h^* &= \operatorname*{argmin}_h \mathrm{AMSE}\{\widehat{m}(x)|h\} \\ &= \{\mu_K m''(x)\}^{-2/5}\left\{\frac{\varphi_K \sigma^2(x)}{np_X(x)}\right\}^{1/5}\left(s^2\right)^{1/5} \propto \left(s^2\right)^{1/5}, \end{aligned} \tag{39}$$

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

and the associated optimal AMSE evaluated at the optimal bandwidth h^*

$$\text{AMSE}\{\widehat{m}(x)|h^*\} = \frac{5}{4}\left\{\mu_K m''(x)\right\}^{2/5}\left\{\frac{\varphi_K \sigma^2(x)}{np_X(x)}\right\}^{4/5}\left(s^2\right)^{4/5} \propto \left(s^2\right)^{4/5}. \quad \textbf{(40)}$$

In Section 6.4, we tabulate s^2 for different distributions.

Theorem 5.2 studies $\widehat{m}_{\text{WQAE}}(x|\omega^*)$ when we use the estimated optimal weight $\widehat{\omega}^*$ in (17).

THEOREM 5.2. *Under the assumptions of Theorem 3.1 and Assumption 3.2, when we use the estimated weight $\widehat{\omega}^*$ in (17), $\widehat{m}_{\text{WQAE}}(x|\widehat{\omega}^*)$ has the same asymptotic normality as $\widehat{m}_{\text{WQAE}}(x|\omega^*)$.*

The discussion of the selection of the bandwidth h is deferred to Section 8.4.

6. EFFICIENCY COMPARISON

6.1. The k-Quantile Optimal Efficiency Ω_k and Λ_k

The parameters in Sections 3–5 can be classified into two types: location-type and scale-type parameters. For β in (6) and (20) and the nonparametric function $m(\cdot)$ in Section 5, these parameters do not directly interact with the error ε, and we call them location-type parameters. For γ in (20), it is directly related to ε, and we call it a scale-type parameter.

Our discussion in the previous sections considers combination of information over a fixed number of quantiles. From the results in Sections 3–5, for the location-type parameters mentioned above, their OWQAE has the asymptotic variance proportional to Ω_k^{-1} with Ω_k defined in (15); for the scale-type parameter γ in (20), the OWQAE has the asymptotic variance proportional to Λ_k^{-1} with Λ_k defined in (26). Since the efficiency of an estimator is inversely proportional to its variance, we call Ω_k and Λ_k the *k-quantile optimal efficiency* of the location-type and the scale-type parameters, respectively. The larger Ω_k and Λ_k, the better performance of the corresponding estimators.

It is well-known that, under appropriate conditions, the variance of any unbiased parameter estimator has the Cramér–Rao lower bound: the inverse of the Fisher information of the underlying distribution. To illustrate the Fisher information for the location-type and scale-type parameters, consider the simple location-scale model $Y = \beta + \gamma\,\varepsilon$ with location parameter β and scale parameter γ. Note that Y has the density $f_Y(y;\beta,\gamma) = f_\varepsilon((y-\beta)/\gamma)/\gamma$. Under the specification $(\beta,\gamma) = (0,1)$, we can show that the Fisher information for the location parameter β is

$$\mathcal{I}(f_\varepsilon) = \int_{\mathbb{R}} \frac{\left[f_\varepsilon'(u)\right]^2}{f_\varepsilon(u)}du = \int_0^1 \left\{\frac{\partial f_\varepsilon(Q_\varepsilon(\tau))}{\partial \tau}\right\}^2 d\tau, \quad \textbf{(41)}$$

and the Fisher information for the scale parameter γ is

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

$$\mathcal{J}(f_\varepsilon)=\int_{\mathbb{R}}\frac{\left[f_\varepsilon(u)+uf_\varepsilon'(u)\right]^2}{f_\varepsilon(u)}du=\int_0^1\left\{\frac{\partial[Q_\varepsilon(\tau)f_\varepsilon(Q_\varepsilon(\tau))]}{\partial\tau}\right\}^2d\tau. \tag{42}$$

We assume $\mathcal{I}(f_\varepsilon)<\infty$ and $\mathcal{J}(f_\varepsilon)<\infty$. The Fisher information $\mathcal{I}(f_\varepsilon)$ and $\mathcal{J}(f_\varepsilon)$ serve as a natural standard when we measure the efficiency of our estimators in the case of regular estimation.

THEOREM 6.1. *Suppose Assumption 2.2 holds. Let* $\Delta=1/(k+1)$.

(i) *For* Ω_k *in (15), we have* $|\Omega_k-\mathcal{I}(f_\varepsilon)|\le\Phi_k$, *where* $g(t)=f_\varepsilon(Q_\varepsilon(t))$, *and*

$$\Phi_k=\frac{g^2(\Delta)+g^2(1-\Delta)}{\Delta}+\frac{\Delta^2}{2}\int_\Delta^{1-\Delta}\left[g''(t)\right]^2dt+\int_0^\Delta\left\{\left[g'(t)\right]^2+\left[g'(1-t)\right]^2\right\}dt. \tag{43}$$

(ii) *For* Λ_k *in (26), we have* $|\Lambda_k-\mathcal{J}(f_\varepsilon)|\le\Psi_k$, *where* $h(t)=Q_\varepsilon(t)f_\varepsilon(Q_\varepsilon(t))$, *and*

$$\Psi_k=\frac{h^2(\Delta)+h^2(1-\Delta)}{\Delta}+\frac{\Delta^2}{2}\int_\Delta^{1-\Delta}\left[h''(t)\right]^2dt+\int_0^\Delta\left\{\left[h'(t)\right]^2+\left[h'(1-t)\right]^2\right\}dt. \tag{44}$$

Theorem 6.1 indicates that, by optimally combining k quantiles $\tau_1,\dots,\tau_k$, the k-quantile optimal efficiency Ω_k (respectively Λ_k) for the OWQAE of the location-type (respectively scale-type) parameters is at most Φ_k (respectively Ψ_k) away from the corresponding Fisher information $\mathcal{I}(f_\varepsilon)$ (resp. $\mathcal{J}(f_\varepsilon)$). This result holds for any fixed k.

6.2. Asymptotic Behavior of Ω_k and Λ_k

In all previous sections, k is assumed to be a given finite number. In the following few sections, we discuss the behavior of the proposed estimators as k increases with n. In this section, we consider the asymptotic behavior of Ω_k and Λ_k as $k\to\infty$. For regular estimation, it is shown that Ω_k and Λ_k approach the corresponding Cramér–Rao efficiency bound. In Section 6.3, we discuss the OWQAE as $k\to\infty$ and when we use the true optimal weight or the estimated optimal weight. In Section 6.4, we discuss the asymptotic relative efficiency of OWQAE compared to some existing methods. Finally, Section 6.5 briefly considers some nonregular estimation.

The Cramér–Rao efficiency analysis is based on the basic assumption of finite Fisher information $\mathcal{I}(f_\varepsilon)<\infty$ and $\mathcal{J}(f_\varepsilon)<\infty$. From (41) and (42), this implies that $\int_0^\tau\{[g'(t)]^2+[g'(1-t)]^2\}dt\to0$ and $\int_0^\tau\{[h'(t)]^2+[h'(1-t)]^2\}dt\to0$ as $\tau\to0$, where $g(\tau)$ and $h(\cdot)$ are defined in Theorem 6.1. Thus, by Theorem 6.1, we have the following result.

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

THEOREM 6.2. *Suppose Assumption 2.2 holds. Let $g(\cdot)$ and $h(\cdot)$ be defined in Theorem 6.1.*

(i) *If*

$$\lim_{\tau\to 0}\frac{g^2(\tau)+g^2(1-\tau)}{\tau}=0 \quad \textit{and} \quad \lim_{\tau\to 0}\tau^2\int_{\tau}^{1-\tau}\left[g''(t)\right]^2dt=0, \tag{45}$$

then, for Φ_k in (43), $\lim_{k\to\infty}\Phi_k=0$, and

$$\lim_{k\to\infty}\Omega_k=\mathcal{I}(f_\varepsilon).$$

(ii) *If (45) holds with $g(\cdot)$ replaced by $h(\cdot)$, then, for Ψ_k in (44), $\lim_{k\to\infty}\Psi_k=0$, and*

$$\lim_{k\to\infty}\Lambda_k=\mathcal{J}(f_\varepsilon).$$

The condition (45) is conventionally imposed in the study of efficient estimations. Basically, it requires that the error density decay sufficiently fast at the boundary (the corresponding estimation is sometimes called as regular estimation), otherwise one may estimate the parameters at a faster rate; see, e.g., Akahira and Takeuchi (1995) for a discussion of this issue, also see Section 6.5 for discussions on related issues. By Theorem 6.2, as $k\to\infty$, the k-quantile optimal efficiency Ω_k and Λ_k attain the corresponding Fisher information.

From Theorem 6.2, Ω_k and Λ_k have different limit as $k\to\infty$. As discussed in Section 4, this is due to the extra dependence of $\gamma\, Q_\varepsilon(\tau)$ on $Q_\varepsilon(\tau)$ in the scale-type parameter.

Proposition 6.1 presents an alternative sufficient condition for (45).

PROPOSITION 6.1. *Suppose f_ε has support on $\mathbb{R}$ and Assumption 2.2 holds. Then (45) holds if*

$$\lim_{u\to\pm\infty}\left\{\frac{f_\varepsilon^2(u)}{\min\{1-F_\varepsilon(u),F_\varepsilon(u)\}}+\left|\frac{\partial^2\left[\log f_\varepsilon(u)\right]}{\partial u^2}\right|\frac{\min\{1-F_\varepsilon(u),F_\varepsilon(u)\}^{3/2}}{f_\varepsilon(u)}\right\}=0. \tag{46}$$

Write $x\propto y$ if x/y is bounded away from 0 and ∞. If $f_\varepsilon(u)\propto|u|^{-a}$ as $|u|\to\infty$ for some $a>1$, then $1-F_\varepsilon(u)\propto|u|^{1-a}$ as $u\to\infty$ and $F_\varepsilon(u)\propto|u|^{1-a}$ as $u\to-\infty$. Thus, by Proposition 6.1, we have the following result.

COROLLARY 6.1. *Suppose that there exist $a>1$ and $b>0$ such that $f_\varepsilon(u)\propto|u|^{-a}$ and $\partial^2[\log f_\varepsilon(u)]/\partial u^2\propto|u|^{-b}$ as $|u|\to\infty$. Then (45) is satisfied if $b+3(a-1)/2>1$.*

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

Many commonly used distributions with support on $\mathbb{R}$ satisfy (45). (i) For standard normal density f_ε, $\partial^2 \log f_\varepsilon(u)/\partial u^2 = -1$ and $1 - F_\varepsilon(u) = [1 + o(1)] f_\varepsilon(u)/u$ as $u \to \infty$, we can verify (46). (ii) For Laplace distribution with density $f_\varepsilon(u) = 0.5 \exp(-|u|)$, $u \in \mathbb{R}$, $\partial^2 [\log f_\varepsilon(u)]/\partial u^2 = 0$ for $u \neq 0$, so (46) can be easily verified. (iii) For logistic distribution, $g(\tau) = c(\tau - \tau^2)$ for some constant $c > 0$, so (45) holds. (iv) For Student-t distribution with $d > 0$ degrees of freedom, Corollary 6.1 holds with $a = d + 1$ and $b = 2$. (v) For normal mixture $\theta N(\mu_1, \sigma_1^2) + (1 - \theta) N(\mu_2, \sigma_2^2)$, $\mu_1, \mu_2 \in \mathbb{R}$, $\sigma_1^2, \sigma_2^2 > 0$, $\theta \in [0, 1]$, we can verify (46).

Recall that the unweighted quantile average estimators have asymptotic variance proportional to R_k defined in (16). The following results show that such a simple averaging estimator is asymptotically equivalent to the LS estimator as $k \to \infty$. This result indicates that if we use a simple average over quantiles, even as we use more and more quantiles, there is no efficiency gain of combining quantile information. Thus, proper weighting over different quantiles is crucial.

THEOREM 6.3. (i) $R_k \geq \Omega_k^{-1}$*; as* $k \to \infty$*, the equality holds if and only if* ε *is normally distributed.* (ii) *If* $\text{var}(\varepsilon) < \infty$*, then* $\lim_{k \to \infty} R_k = \text{var}(\varepsilon)$.

6.3. Behavior of the OWQAE as $k \to \infty$

In Sections 3–5, the asymptotic normalities of the OWQAE are established for k quantiles with a fixed k. In this section we consider the case that k increases with n. To keep the length, we consider only $\widehat{\beta}_{\text{WQAE}}(\omega^*)$ for the parametric regression case in Section 3.

Since the uniform Bahadur representation holds on a subinterval of [0,1], we modify Assumption 3.1 so that the Bahadur representation holds uniformly over expanding subintervals of [0,1] when the the number of quantiles increases with n.

Assumption 6.1. The asymptotic Bahadur representation (9) holds over $\tau \in \mathcal{T}_n = [\delta_n, 1 - \delta_n]$ with $\delta_n = (\log n)^{-\epsilon}$ for some $\epsilon > 0$. Let the number of quantiles $k = k_n = \lfloor \delta_n^{-1} \rfloor - 1$.

First, we consider the OWQAE with the theoretical optimal weight ω^* in (14).

COROLLARY 6.2. *Consider* $\widehat{\beta}_{\text{WQAE}}(\omega)$ *in (11) and* ω^* *in (14). Suppose Assumptions 2.1, 2.2, and 6.1 and (45) hold. Further assume* $\dot{m}(X_t, \beta) \in \mathcal{L}^q$ *for some* $q > 2$. *Then*
$n^{1/2} [\widehat{\beta}_{\text{WQAE}}(\omega^*) - \beta] \Rightarrow N\left(0, \Sigma_\beta^{-1} \mathcal{I}(f_\varepsilon)^{-1}\right)$.

Thus, if we use more and more quantiles as $n \to \infty$, the efficiency of the OWQAE with the theoretical optimal weight ω^* approaches the Fisher information. The same conclusion also holds for the estimators in Sections 4–5 provided that, as in Assumption 6.1, appropriate Bahadur representations hold uniformly on $\mathcal{T}_n$.

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

We next briefly discuss limiting behavior of the OWQAE with estimated weight when k_n is chosen as in Assumption 6.1. Again, we discuss the behavior of the parametric model in Section 3. As $k_n \to \infty$, the asymptotic analysis of the proposed estimator is complicated and depends on the behavior of quantile regression estimators and quantile-density estimators at the extreme.

Let $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*) = \sum_{j=1}^{k_n} \omega_j^* \widehat{\beta}(\tau_j)$ be the OWQAE with the optimal weight ω^* in (14), and let $\widehat{\beta}_{\mathrm{WQAE}}(\widehat{\omega}^*) = \sum_{i=1}^{k_n} \widehat{\omega}_j^* \widehat{\beta}(\tau_j)$ be the OWQAE with estimated weight $\widehat{\omega}^*$ [see, e.g., (17)]. Then, using $\sum_{j=1}^{k_n} \omega_j^* = 1$ and $\sum_{j=1}^{k_n} \widehat{\omega}_j^* = 1$, we have

$$\sqrt{n}\left[\widehat{\beta}_{\mathrm{WQAE}}(\widehat{\omega}^*) - \beta\right] = \sum_{j=1}^{k_n} \left(\widehat{\omega}_j^* - \omega_j^*\right) \sqrt{n}\left[\widehat{\beta}(\tau_j) - \beta\right] + \sqrt{n}\left[\widehat{\beta}_{\mathrm{WQAE}}(\omega^*) - \beta\right]. \tag{47}$$

From Corollary 6.2, in order to prove

$$\sqrt{n}\left[\widehat{\beta}_{\mathrm{WQAE}}(\widehat{\omega}^*) - \beta\right] \Rightarrow N\left(0, \Sigma_\beta^{-1} \mathcal{I}(f_\varepsilon)^{-1}\right), \tag{48}$$

it suffices to prove

$$\sum_{j=1}^{k} \left(\widehat{\omega}_j^* - \omega_j^*\right) \sqrt{n}\left[\widehat{\beta}(\tau_j) - \beta\right] = o_p(1). \tag{49}$$

We need additional regularity conditions regarding the behavior of density $f_\varepsilon(Q_\varepsilon(\tau))$ when τ approaches the boundary, and conditions on the density estimators.

Assumption 6.2. Let k_n and $\mathcal{T}_n$ be chosen as in Assumption 6.1. There exists some constant $\eta > 0$ such that $\inf_{\tau \in \mathcal{T}_n} f_\varepsilon(Q_\varepsilon(\tau)) \geq c(\log n)^{-\eta}$ and $\sum_{j=1}^{k_n} |\widehat{\omega}_j^* - \omega_j^*| = o_p[(\log n)^{-(\eta+\epsilon/2)}]$, where ϵ is the constant in Assumption 6.1.

Under Assumption 6.1 and the condition $\sum_{j=1}^{k_n} |\widehat{\omega}_j^* - \omega_j^*| = o_p\left[(\log n)^{-(\eta+\epsilon/2)}\right]$ in Assumption 6.2, by (71) in the proof of Theorem 3.1, we have

$$\sum_{j=1}^{k_n} \left(\widehat{\omega}_j^* - \omega_j^*\right) \sqrt{n}\left[\widehat{\beta}(\tau_j) - \beta\right] = \left[\Sigma_\beta^{-1} + o_p(1)\right] \sum_{j=1}^{k_n} \frac{\widehat{\omega}_j^* - \omega_j^*}{f_\varepsilon(Q_\varepsilon(\tau_j))} N_j + o_p(1), \tag{50}$$

where

$$N_j = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left\{\dot{m}(X_t, \beta) - \mathbb{E}[\dot{m}(X, \beta)]\right\} \left[\tau_j - \mathbf{1}_{\varepsilon_t < Q_\varepsilon(\tau_j)}\right].$$

Assume without loss of generality that $\dot{m}(X_t, \beta)$ is scalar-valued. By property (P1) in Section 2, the summands of N_j are martingale differences. By the

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

condition $\dot{m}(X_t, \beta) \in \mathcal{L}^2$ and the orthogonality of martingale differences, $\mathbb{E}(N_j^2) = O(1)$ uniformly in j. Thus,

$$\mathbb{E}\left(\max_{1\le j\le k_n} N_j^2\right) \le \sum_{j=1}^{k_n} \mathbb{E}(N_j^2) = O(k_n),$$

and thus $\max_{1\le j\le k_n} |N_j| = O_p(\sqrt{k_n})$. Recall k_n in Assumption 6.1. Under Assumption 6.2,

$$\left|\sum_{j=1}^{k_n} \frac{\widehat{\omega}_j^* - \omega_j^*}{f_\varepsilon(Q_\varepsilon(\tau_j))} N_j\right| \le \frac{\max_{1\le j\le k_n} |N_j|}{\inf_{\tau\in\mathcal{T}_n} f_\varepsilon(Q_\varepsilon(\tau))} \sum_{j=1}^{k_n} |\widehat{\omega}_j^* - \omega_j^*| = o_p(1).$$

Thus, (49) follows from (50), and we conclude that (48) holds.

6.4. Comparison of Asymptotic Relative Efficiency

We now compare the efficiency of the proposed OWQAE to some existing methods.

First, we consider the parametric case in Section 3. Theorem 3.2 gives $\sqrt{n}\big[\widehat{\beta}_{\mathrm{WQAE}}(\omega^*) - \beta\big] \Rightarrow N\big(0, \Sigma_\beta^{-1} S(\omega^*)\big)$. For parameter estimations, the most widely used method is the ordinary LS estimator, denoted by $\widehat{\beta}_{\mathrm{LS}}$, which minimizes the squared errors. Assuming $\mathrm{var}(\varepsilon) < \infty$ and other appropriate conditions, we have the asymptotic normality $\sqrt{n}\big(\widehat{\beta}_{\mathrm{LS}} - \beta\big) \Rightarrow N\big(0, \Sigma_\beta^{-1}\mathrm{var}(\varepsilon)\big)$. For the quantile regression based estimator $\widehat{\beta}(\tau)$ with a single quantile τ, the asymptotic normality in (10) holds. All the three estimators $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$, $\widehat{\beta}_{\mathrm{LS}}$, and $\widehat{\beta}(\tau)$ have asymptotic normality of the form: $\sqrt{n}\big(\widehat{\beta} - \beta\big) \Rightarrow N\big(0, \Sigma_\beta^{-1} s^2\big)$, where $s^2 = \mathrm{var}(\varepsilon)$ (assuming finite) for $\widehat{\beta}_{\mathrm{LS}}$, $s^2 = \tau(1-\tau)/f_\varepsilon^2(Q_\varepsilon(\tau))$ for $\widehat{\beta}(\tau)$, and $s^2 = S(\omega^*)$ for $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$. For comparison, we use $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$ as the benchmark and define its asymptotic relative efficiency (ARE) to $\widehat{\beta}_{\mathrm{LS}}$ and $\widehat{\beta}(\tau)$ as

$$\mathrm{ARE}\big(\widehat{\beta}_{\mathrm{LS}}\big) = \frac{\mathrm{var}(\varepsilon)}{S(\omega^*)} \qquad \text{and} \qquad \mathrm{ARE}\big(\widehat{\beta}(\tau)\big) = \frac{\tau(1-\tau)}{f_\varepsilon^2(Q_\varepsilon(\tau))S(\omega^*)}. \tag{51}$$

A value of $\mathrm{ARE} \ge 1$ indicates better performance of $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$. Clearly, $\mathrm{ARE}\big(\widehat{\beta}(\tau_j)\big) \ge 1, j = 1, \ldots, k$. Under the conditions in Theorem 6.2, $\lim_{k\to\infty} S(\omega^*) = 1/\mathcal{I}(f_\varepsilon) \le \mathrm{var}(\varepsilon)$ (the Cramér–Rao inequality) so that $\lim_{k\to\infty} \mathrm{ARE}\big(\widehat{\beta}_{\mathrm{LS}}\big) \ge 1$. Intuitively, both $\widehat{\beta}_{\mathrm{LS}}$ and $\widehat{\beta}(\tau)$ use only partial information: sample average and sample τ-th quantile, respectively. By contrast, $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$ combines strength across quantiles and thus can be more efficient.

Using $k = 9$ quantiles, Table 1 tabulates $\mathrm{ARE}\big(\widehat{\beta}_{\mathrm{LS}}\big)$ and $\mathrm{ARE}\big(\widehat{\beta}(\tau)\big)$, $\tau = 0.1, \ldots, 0.9$, for some commonly used distributions. For all nonnormal distributions considered, $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$ significantly outperforms $\widehat{\beta}_{\mathrm{LS}}$ and $\widehat{\beta}(\tau)$. For N(0,1), $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$ and $\widehat{\beta}_{\mathrm{LS}}$ are comparable, and both are about 50% more

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

TABLE 1. Theoretical ARE [see (51)] of $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$ compared to $\widehat{\beta}_{\mathrm{LS}}$, Zou and Yuan (2008)'s CQR estimator $\widehat{\beta}_{\mathrm{CQR}}$, and $\widehat{\beta}(\tau)$, using 9 quantiles $\tau_j = j/10$, $j = 1, \ldots, 9$. Mixture 1: 0.5N(0,1)+0.5N(0,0.5^6); Mixture 2: 0.5N(−2,1) + 0.5N(2,1). For Student-t_1, t_2, LS is not applicable [numbers ≥ 1 indicate better performance of $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$]

ε			$\widehat{\beta}(\tau)$ with $\tau =$								
distribution	$\widehat{\beta}_{\mathrm{LS}}$	$\widehat{\beta}_{\mathrm{CQR}}$	τ_1	τ_2	τ_3	τ_4	τ_5	τ_6	τ_7	τ_8	τ_9
Student-t_1	NA	1.58	47.12	6.40	2.34	1.40	1.19	1.40	2.34	6.40	47.12
Student-t_2	NA	1.12	9.01	2.85	1.65	1.27	1.17	1.27	1.65	2.85	9.01
N(0,1)	0.96	1.03	2.80	1.96	1.67	1.54	1.51	1.54	1.67	1.96	2.80
Mixture 1	10.17	2.43	91.95	32.64	3.26	1.80	1.55	1.80	3.26	32.64	91.95
Mixture 2	3.28	2.80	3.01	2.81	3.68	7.84	56.24	7.84	3.68	2.81	3.01
Laplace	2.00	1.32	9.00	4.00	2.33	1.50	1.00	1.50	2.33	4.00	9.00
Gamma(1)	9.00	3.67	1.00	2.25	3.86	6.00	9.00	13.50	21.00	36.00	81.00
Gamma(2)	2.44	1.73	1.12	1.49	1.91	2.42	3.10	4.08	5.65	8.68	17.34
Gamma(3)	1.71	1.41	1.26	1.42	1.64	1.94	2.35	2.94	3.88	5.68	10.75
Beta(1,1)	1.67	2.04	1.80	3.20	4.20	4.80	5.00	4.80	4.20	3.20	1.80
Beta(1,2)	2.35	2.33	1.05	2.11	3.16	4.22	5.27	6.33	7.38	8.44	9.49
Beta(1,3)	3.31	2.71	1.02	2.12	3.32	4.66	6.19	8.00	10.27	13.33	19.04

efficient than $\widehat{\beta}(0.5)$. For Student-t with one (t_1) or two (t_2) degrees of freedom, LS is not applicable due to infinite variance; $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$ is about 20% more efficient than $\widehat{\beta}(0.5)$ and even substantially more efficient than $\widehat{\beta}(\tau)$ for other choices of τ. Thus, potentially much improved efficiency and robustness can be achieved by using the proposed estimator $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$. For linear models, Zou and Yuan (2008) studied composite quantile regression (CQR) method, and we include the efficiency of their method for comparison purpose. Clearly, the OWQAE is significantly more efficient than the CQR.

We briefly mention efficiency comparison of the nonparametric estimator $\widehat{m}_{\mathrm{WQAE}}(x|\omega^*)$ relative to the local LS, local LAD, and Kai, Li, and Zou (2010)'s local CQR estimators in Section 5. By (40),

$$\text{The nonparametric relative efficiency} \;=\; (\text{The parametric relative efficiency})^{4/5}.$$

Thus, the same efficiency comparison result (up to an exponent 4/5) in Table 1 also holds for the nonparametric estimator $\widehat{m}_{\mathrm{WQAE}}(x|\omega^*)$ in Section 5.

6.5. Asymptotic Superefficiency

By Corollary 6.2, under (45), $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$ is an asymptotically efficient estimator of β, with limiting covariance matrix approaching the Fisher information bound. The corresponding conditions (45) are not mathematical trivialities but are real restricting conditions to obtain the efficiency results. In the case when those

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

"usual" regularity conditions do not hold, the previously discussed efficiency result may not hold, and we may obtain different results from the likelihood-based estimation. For example, we may have a different rate of convergence, and in general, asymptotically efficient estimators do not exist. These "unusual" cases are sometimes called "nonregular" statistical estimation. In this section, we briefly discuss the case of nonregular estimation. In this case, Theorem 6.4 shows that, by using quantile regression with optimal weighting, superefficient estimators may be obtained in the sense that the efficiency is larger than the Fisher information $\mathcal{I}(f_\varepsilon)$.

THEOREM 6.4. *Recall Ω_k in (15). Let $g(\tau)$ be defined in Theorem 6.1. Assume*

$$\lim_{\tau\to 0}\frac{g^2(\tau)+g^2(1-\tau)}{\tau}=c. \tag{52}$$

(i) *If $0<c<\infty$ and $\lim_{\tau\to 0}\tau^2\int_\tau^{1-\tau}[g''(t)]^2dt=0$, then $\lim_{k\to\infty}\Omega_k=c+\mathcal{I}(f_\varepsilon)$.*

(ii) *If $c=\infty$, then $\lim_{k\to\infty}\Omega_k=\infty$.*

Condition (45) covers the regular case $c=0$ in (52). Theorem 6.4 indicates that, for the nonregular case $c>0$ in (52), under appropriate conditions, for large k, the variance of the (standardized) optimally weighted quantile regression based estimator $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$ is smaller than the Cramér–Rao bound. In particular, if $c=\infty$, as the number of quantiles k increases, the asymptotic variance approaches zero. In this sense, the estimator $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$ is asymptotically superefficient.

Corollary 6.3 concerns a special case of superefficiency when the density f_ε is positive at the boundary.

COROLLARY 6.3. *Denote the support of f_ε by $\mathcal{D}$, then $\lim_{k\to\infty}\Omega_k=\infty$ in any of the following three cases:* (i) *$\mathcal{D}=[D_1,D_2]$ with $f_\varepsilon(D_1)+f_\varepsilon(D_2)>0$;* (ii) *$\mathcal{D}=[D_1,\infty)$ with $f_\varepsilon(D_1)>0$; or* (iii) *$\mathcal{D}=(-\infty,D_2]$ with $f_\varepsilon(D_2)>0$.*

For the truncated version of the distributions in Section 6.2, we have $\lim_{k\to\infty}\Omega_k=\infty$. For example, for the truncated normal on $[-1,1]$, Corollary 6.3(i) applies. For uniform distribution on $[0,1]$, we can show $\Omega_k=2k+2\to\infty$.

Similar results can also be established for Λ_k. We omit the details.

7. ESTIMATION OF THE OPTIMAL WEIGHT

To construct the proposed OWQAE $\widehat{\beta}_{\mathrm{WQAE}}(\omega^*)$ in Sections 3–5, we need to obtain estimations of the optimal weight ω^* in (14) and π^* in (25). It suffices to estimate $Q_\varepsilon(\tau)$ and $f_\varepsilon(Q_\varepsilon(\tau))$. We can accomplish this through a two-step procedure: first, use a preliminary estimator to obtain residuals; second, estimate

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

$Q_\varepsilon(\tau)$ and $f_\varepsilon(Q_\varepsilon(\tau))$ based on the residuals. Here we illustrate the idea using the models in Sections 3–5.

(**Case 1: Parametric model in Section 3**.) Since $f_\varepsilon(Q_\varepsilon(\tau))$ remains the same if we change ε to $c+\varepsilon$ for any c, α in (6) can be absorbed into ε. We propose the procedure:

(i) Use the uniform weight $\omega = [1/k, \ldots, 1/k]^T$ to obtain the preliminary estimator $\widehat{\beta}$, and compute the "residuals" (a combination of both α and ε) as $\widehat{\varepsilon}_t = Y_t - m(X_t, \widehat{\beta})$.

(ii) To estimate $f_\varepsilon(u)$, use the nonparametric density estimate $\widehat{f_\varepsilon}(u) = (nb)^{-1}\sum_{t=1}^n K\{(u-\widehat{\varepsilon}_t)/b\}$, where we follow Silverman (1986) to choose the rule-of-thumb bandwidth b:

$$b = 0.9n^{-1/5} * \min\left\{\mathrm{sd}(\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_n), \frac{\mathrm{IQR}(\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_n)}{1.34}\right\}.$$

Here, "sd" and "IQR" are the sample standard deviation and sample interquartile.

(iii) Estimate $f_\varepsilon(Q_\varepsilon(\tau))$ by $\widehat{f_\varepsilon}(\widehat{Q}_\varepsilon(\tau))$, where $\widehat{Q}_\varepsilon(\tau)$ is the sample τ-th quantile of $\widehat{\varepsilon}_1, \ldots, \widehat{\varepsilon}_n$.

(iv) Plug $\widehat{f_\varepsilon}(\widehat{Q}_\varepsilon(\tau))$ into (14) to obtain the estimated optimal weight $\widehat{\omega}^*$.

(**Case 2: Location-scale model in Section 4**.) To ensure identifiability, we assume without loss of generality that $\sum_{j=1}^k |Q_\varepsilon(\tau_j)| = 1$, otherwise we can consider the reparametrized model $Y = X^T\beta + (U^T\gamma^*)\varepsilon^*$ with $\gamma^* = c\gamma, \varepsilon^* = \varepsilon/c$, and $c = \sum_{j=1}^k |Q_\varepsilon(\tau_j)|$. Note that this assumption bears no effect on the optimal weight ω^* since $f_\varepsilon(Q_\varepsilon(\tau))$ is invariant under the transformation $c\varepsilon$ for any $c > 0$. For each quantile $\tau = \tau_1, \ldots, \tau_k$, we fit the quantile regression (27) to obtain $(\widetilde{\beta}(\tau), \widetilde{\gamma}(\tau))$. Define the preliminary estimator

$$\widetilde{\beta} = \frac{1}{k}\sum_{j=1}^k \widetilde{\beta}(\tau_j) \quad \text{and} \quad \widetilde{\gamma} = \sum_{j=1}^k |\widetilde{\gamma}(\tau_j)|. \tag{53}$$

Then $(\widetilde{\beta}, \widetilde{\gamma})$ consistently estimates (β, γ). We use the procedure to compute ω^* and π^*:

(i) Use $\widetilde{\beta}$ and $\widetilde{\gamma}$ in (53) to compute the errors $\widetilde{\varepsilon}_t = (Y_t - X_t^T\widetilde{\beta})/(U_t^T\widetilde{\gamma})$, $t = 1, \ldots, n$. To better mimic the constraint $\sum_{j=1}^k |Q_\varepsilon(\tau_j)| = 1$, consider the transformed errors

$$\widehat{\varepsilon}_t = \frac{\widetilde{\varepsilon}_t}{\sum_{j=1}^k |\widetilde{Q}_\varepsilon(\tau_j)|}, \quad t = 1, \ldots, n,$$

where $\widetilde{Q}_\varepsilon(\tau)$ is the sample τ-quantile of $\widetilde{\varepsilon}_1, \ldots, \widetilde{\varepsilon}_n$.

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

(ii) Use the same steps (ii)–(iii) in case 1 to obtain estimates $\widehat{f}_\varepsilon(\widehat{Q}_\varepsilon(\tau))$ and $\widehat{Q}_\varepsilon(\tau)$.

(iii) Use (14) to compute $\widehat{\omega}^*$ and use (25) to compute $\widehat{\pi}^*$.

(**Case 3: Nonparametric regression model in Section 5**.) As in case 2, $\omega^*$is invariant under the transformation $c\varepsilon, c > 0$. Assume without loss of generality that $|\varepsilon|$ has median one. Then the conditional median of $|Y - m(X)|$ given X is $\sigma(X)$, and we can apply local median quantile regression to estimate $\sigma(\cdot)$. We propose the procedure:

(i) Use (32) with the uniform weight to obtain the preliminary estimator $\widehat{m}(\cdot)$.

(ii) Compute $Y_t - \widehat{m}(X_t)$ and estimate $\sigma(\cdot)$ by local linear median quantile regression:

$$(\widehat{\sigma}(x), \widehat{b}) = \underset{\sigma,b}{\operatorname{argmin}} \sum_{t=1}^{n} \rho_{0.5}\big\{\ |Y_t - \widehat{m}(X_t)| - \sigma - b(X_t - x)\big\} K\left(\frac{X_t - x}{\ell}\right). \tag{54}$$

For the bandwidth ℓ, following Yu and Jones (1998), we use $\ell = \ell_{\mathrm{LS}}(\pi/2)^{1/5}$, where ℓ_{LS} is the plug-in bandwidth (Ruppert, Sheather, and Wand, 1995) for local linear LS regression based on the data $(X_i, |Y_i - \widehat{m}(X_i)|^2), i = 1, \ldots, n$.

(iii) Compute the errors $\widehat{\varepsilon}_t = \big[Y_t - \widehat{m}(X_t)\big]/\widehat{\sigma}(X_t)$ and obtain the estimator $\widehat{f}_\varepsilon(\widehat{Q}_\varepsilon(\tau))$ as in the parametric regression case 1.

(iv) Use (14) to obtain $\widehat{\omega}_1, \ldots, \widehat{\omega}_k$ and symmetrize them: $\widehat{\omega}_j^* = \big(\widehat{\omega}_j + \widehat{\omega}_{k+1-j}\big)/2, j = 1, \ldots, k$.

8. MONTE CARLO STUDIES

We conduct Monte Carlo studies to investigate the sampling performance of the proposed procedures in a variety of regression models. In all settings below, we use 1000 realizations to evaluate the performance of various methods.

8.1. Linear Models with Homoscedastic Errors

For linear models, we compare six estimation methods. OLS: ordinary LS estimator; LAD: the median quantile estimator with $\tau = 0.5$ in (8); QAU, QAO, QAE: the WQAE in (11) with the uniform weights, theoretical optimal weight ω^* in (14), and estimated optimal weight $\widehat{\omega}^*$ (cf. Section 7), respectively; CQR: Zou and Yuan (2008)'s CQR estimator. For QAU, QAO, QAE, and CQR, we use $k = 9$ quantiles $0.1, 0.2, \ldots, 0.9$. With 1000 realizations, we use QAE as the benchmark to which the other five methods are compared based on the empirical relative efficiency:

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

$$\text{RE(Method)} = \frac{\text{MSE(Method)}}{\text{MSE(QAE)}} \quad \text{and} \quad \text{MSE} = \frac{1}{1000}\sum_{j=1}^{1000}\left[\widehat{\beta}(j) - \beta\right]^2, \tag{55}$$

where "Method" stands for OLS, LAD, QAU, QAO, CQR, and $\widehat{\beta}(j)$ is the estimator of β in the j-th realization. A value of RE ≥ 1 indicates better performance of QAE.

We consider both independent data and time series data:

$$\text{Model 1}: Y_t = \alpha + X_t\beta + \varepsilon_t, \quad X_t \sim N(0,1), \quad (\alpha, \beta) = (0,1), \tag{56}$$

$$\text{Model 2}: Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 |Y_{t-2}| + 0.2\varepsilon_t, \quad (\alpha, \beta_1, \beta_2) = (0, 0.3, 0.5). \tag{57}$$

Model 2 is a variant of the threshold autoregressive model with a linear component Y_{t-1}. For the innovation ε_t, we consider 12 distributions: Normal distribution N(0,1), Student-t distribution with one (t_1) and two (t_2) degrees of freedom, the two normal mixture distributions in Example 2, Laplace distribution, Beta distributions Beta(1,1), Beta(1,2), Beta(1,3), and Gamma distributions Gamma(1), Gamma(2), Gamma(3).

The results are summarized in Table 2(a) for Model 1 with sample sizes $n = 100, 300$, and in Table 2(b) for Model 2 with $n = 300$. For N(0,1), Student-t_2, and Laplace distributions, QAE and CQR are comparable; for all other distributions, QAE significantly outperforms CQR. Also, QAE outperforms OLS for all nonnormal distributions whereas they are comparable for N(0,1). For $n = 300$, the superior performance of QAE is even more remarkable, which agrees with our asymptotic theory. For almost all cases considered, QAE substantially outperforms the LAD estimator and the relative efficiency can be as high as almost 2000%. It is worth pointing out that, for Beta and Gamma distributions, the relative efficiencies are much higher than the other distributions considered, owing to the superefficiency phenomenon in Section 6.5. We also note that QAE with estimated optimal weight has comparable performance to QAO with theoretical optimal weight. We conclude that the proposed OWQAE offers a more efficient alternative to existing methods.

8.2. Nonlinear Models with Homoscedastic Errors

We consider two nonlinear models (one independent data and the other time series data):

$$\text{Model 3}: Y_t = \alpha + \exp(\beta X_t) + 0.5\varepsilon_t, \ X_t \sim N(0,1), \quad (\alpha, \beta) = (0, 0.6), \tag{58}$$

$$\text{Model 4}: Y_t = \alpha + \sqrt{0.5 + \beta_1 Y_{t-1}^2 + \beta_2 Y_{t-2}^2} + 0.2\varepsilon_t, \quad (\alpha, \beta_1, \beta_2) = (0, 0.3, 0.5). \tag{59}$$

Model 4 is Engle (1982)'s ARCH model. Again, we consider the 12 distributions in Section 8.1 for ε_t. Table 3 summarizes the empirical relative efficiency

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

TABLE 2. (Linear regression model) Empirical relative efficiency of OWQAE with estimated optimal weight compared to five methods: OLS, LAD, QAU, QAO, and CQR. Mixture 1: $0.5\mathrm{N}(0,1)+0.5\mathrm{N}(0,0.5^6)$; Mixture 2: $0.5\mathrm{N}(-2,1)+0.5\mathrm{N}(2,1)$. For Student-$t_1, t_2$, due to infinite variance, the OLS is not stable and varies significantly in simulations [numbers ≥ 1 indicate better performance of OWQAE]

(a) Model 1 in (56)

ε_t	$n=100$					$n=300$				
Distribution	OLS	LAD	QAU	QAO	CQR	OLS	LAD	QAU	QAO	CQR
Student-t_1	NA	0.86	2.55	0.74	1.11	NA	1.04	3.16	0.89	1.37
Student-t_2	NA	0.95	1.14	0.81	0.89	NA	1.06	1.30	0.99	1.02
N(0,1)	0.86	1.33	0.90	0.90	0.91	0.89	1.43	0.95	0.95	0.97
Mixture 1	5.10	0.89	3.70	0.77	1.45	7.79	1.27	6.12	0.95	2.00
Mixture 2	2.22	10.85	2.31	1.18	2.03	2.65	19.88	2.84	1.09	2.30
Laplace	1.24	0.85	1.06	0.85	0.89	1.48	0.87	1.23	0.87	1.01
Gamma(1)	5.69	5.46	4.81	0.67	2.51	7.09	6.98	6.12	0.76	2.91
Gamma(2)	2.08	2.60	1.96	0.88	1.48	2.23	2.85	2.14	0.94	1.61
Gamma(3)	1.49	2.02	1.48	0.92	1.25	1.64	2.21	1.61	0.99	1.35
Beta(1,1)	1.49	4.20	1.73	0.91	1.77	1.58	4.58	1.88	0.95	1.90
Beta(1,2)	1.71	3.65	1.93	1.09	1.71	2.02	4.47	2.34	1.97	1.99
Beta(1,3)	2.39	4.30	2.59	1.36	2.01	2.78	5.12	3.06	3.03	2.25

(b) Model 2 in (57), $n=300$

ε_t	β_1					β_2				
Distribution	OLS	LAD	QAU	QAO	CQR	OLS	LAD	QAU	QAO	CQR
Student-t_1	NA	1.00	1.85	0.89	1.28	NA	0.85	2.74	0.80	1.07
Student-t_2	NA	1.10	1.19	1.04	0.99	NA	1.04	1.23	1.13	1.00
N(0,1)	0.91	1.45	0.94	0.94	0.98	0.90	1.36	0.94	0.94	0.95
Mixture 1	6.85	1.14	4.84	0.92	1.78	6.82	1.16	5.20	0.88	1.77
Mixture 2	2.70	19.79	2.86	1.11	2.35	2.80	19.51	2.99	1.18	2.43
Laplace	1.36	0.89	1.17	0.89	0.99	1.37	0.91	1.21	0.91	1.00
Gamma(1)	6.70	6.23	5.38	0.79	2.83	6.24	6.22	5.35	0.75	2.80
Gamma(2)	2.39	2.94	2.22	0.95	1.70	2.42	2.80	2.23	0.93	1.68
Gamma(3)	1.53	1.95	1.50	0.97	1.27	1.68	2.07	1.59	0.91	1.36
Beta(1,1)	1.50	4.30	1.79	0.93	1.80	1.50	4.15	1.77	0.94	1.81
Beta(1,2)	1.92	4.24	2.21	0.86	1.95	2.23	4.89	2.52	0.85	2.21
Beta(1,3)	2.78	4.71	3.04	0.82	2.29	2.77	4.60	3.01	0.80	2.28

[cf. (55)] of the proposed OWQAE compared to the other methods OLS, LAD, QAU, and QAO (see Section 8.1). The proposed OWQAE is significantly superior to the OLS, LAD, and QAU, and comparable to the QAO with theoretical optimal weight.

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

TABLE 3. (Nonlinear regression model) Empirical relative efficiency of OWQAE with estimated weights compared to four methods: OLS, LAD, QAU, QAO. Mixture 1: 0.5N(0,1)+0.5N(0,0.5^6); Mixture 2: 0.5N(−2,1)+0.5N(2,1). For Student-t_1, t_2, due to infinite variance, the OLS is not stable and varies significantly in simulations [numbers ≥ 1 indicate better performance of OWQAE]

ε_t	Model 3 in (58) $\beta = 0.6$				Model 4 in (59) $\beta_1 = 0.3$				$\beta_2 = 0.5$			
Distribution	OLS	LAD	QAU	QAO	OLS	LAD	QAU	QAO	OLS	LAD	QAU	QAO
Student-t_1	NA	0.48	1.61	0.55	NA	0.75	2.06	0.74	NA	0.73	2.38	0.75
Student-t_2	NA	1.00	1.37	1.55	NA	1.02	1.23	0.91	NA	0.99	1.21	1.06
N(0,1)	0.79	1.81	0.96	0.95	0.97	1.45	0.99	0.99	0.92	1.38	0.97	0.97
Mixture 1	3.75	1.01	3.16	0.86	6.45	1.12	4.62	0.89	6.99	1.09	5.08	0.91
Mixture 2	1.81	9.54	1.90	1.02	2.56	16.83	2.72	1.15	2.75	18.08	2.85	1.15
Laplace	1.36	1.19	1.15	1.19	1.52	0.93	1.28	0.93	1.42	0.89	1.21	0.89
Gamma(1)	4.21	6.58	3.94	0.75	6.74	6.42	5.78	0.77	6.67	6.44	5.56	0.73
Gamma(2)	1.71	2.78	1.69	0.97	2.23	2.59	2.07	0.97	2.39	2.90	2.21	0.92
Gamma(3)	1.31	1.89	1.28	1.14	1.59	2.04	1.53	0.96	1.64	2.05	1.58	0.92
Beta(1,1)	1.15	4.61	1.98	0.94	1.67	4.48	1.95	0.93	1.57	4.02	1.82	0.94
Beta(1,2)	1.08	3.72	1.97	2.50	1.98	3.98	2.24	0.89	2.06	4.21	2.30	0.85
Beta(1,3)	1.58	4.38	2.93	5.27	2.84	4.85	3.02	0.81	2.96	4.94	3.09	0.89

8.3. Location-Scale Models with Conditional Heteroscedasticity

Consider two location-scale models (one independent data and the other time series data):

$$\text{Model 5}: Y_t = \beta X_t + (\gamma_0 + \gamma_1 |X_t|)\varepsilon_t,\ X_t \sim N(0,1),$$
$$(\beta, \gamma_0, \gamma_1) = (0.6, 0.5, 1.0), \tag{60}$$
$$\text{Model 6}: Y_t = \beta Y_{t-1} + (\gamma_0 + \gamma_1 |Y_{t-1}|)\varepsilon_t,\ \ (\beta, \gamma_0, \gamma_1) = (0.4, 0.5, 0.5). \tag{61}$$

Model 6 is the ARCH model in Koenker and Zhao (1996), and it is different from Engle's ARCH model where the conditional heteroscedasticity takes the form in (59). Due to the conditional heteroscedasticity, it is slightly more difficult to estimate the parameters. We use Model 5 to illustrate five estimation methods.

(i): (LS method) If ε_t has zero mean and unit variance, the Gaussian-likelihood based estimation method is to minimize the loss function

$$\sum_{t=1}^{n} \left\{ \frac{(Y_t - bX_t)^2}{(r_0 + r_1|X_t|)^2} + \log\left[(r_0 + r_1|X_t|)^2\right] \right\}. \tag{62}$$

This is essentially an LS type estimation and the Gaussianity is not necessary for the consistency. In general, if ε_t has variance σ^2, then this LS method produces consistent estimators of β and $\sigma(\gamma_0, \gamma_1)$.

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

(ii): (LAD method) First, apply (27) with $\tau = 0.5$ to obtain the LAD estimator $\widehat{\beta}_{\mathrm{LAD}}$ of β. Second, apply median quantile regression based on absolute residuals:

$$\sum_{t=1}^{n} \rho_{0.5}\Big(|Y_t - \widehat{\beta}_{\mathrm{LAD}} X_t| - r_0 - r_1 |X_t|\Big). \tag{63}$$

This LAD regression produces estimators of $Q_{|\varepsilon|}(0.5)(\gamma_0, \gamma_1)$, where $Q_{|\varepsilon|}(0.5)$ is the median quantile of $|\varepsilon_t|$.

(iii): (OWQAE with theoretical optimal weights). As in Section 8.1, we denote this method by QAO.

(iv): (OWQAE with estimated optimal weights). As in Section 8.1, we denote this method by QAE. As discussed in Section 7, under the constraint $\sum_{j=1}^{k} |Q_\varepsilon(\tau_j)| = 1$, QAE produces consistent estimators of $(\beta, \gamma_0, \gamma_1)$. Without the latter constraint, QAE produces consistent estimators of β and $\sum_{j=1}^{k} |Q_\varepsilon(\tau_j)|(\gamma_0, \gamma_1)$.

(v): (OWQAE based on the unweighted quantile regression (27)). This method works the same as the OWQAE above and the only difference is to use $\tilde{\beta}(\tau)$ and $\tilde{\gamma}(\tau)$ in (27) to form the OWQAE. Denote this method by QAEU. Again, QAEU produces estimators of β and $\sum_{j=1}^{k} |Q_\varepsilon(\tau_j)|(\gamma_0, \gamma_1)$. We include this method to evaluate the performance of the OWQAE based on the unweighted quantile regression (27).

As discussed above, the five estimation methods produce consistent estimators of β and $\lambda(\gamma_0, \gamma_1)$ for some constant λ depending on the distribution of ε_t. To make sensible comparison, we divide the corresponding estimators by λ to obtain consistent estimators of (γ_0, γ_1). Furthermore, to ensure the consistency of the LS method, we consider the properly centered ε_t for the 12 distributions in Section 8.1. That is, if ε_t has finite mean, then we center it so that $\mathbb{E}(\varepsilon_t) = 0$.

The results are summarized in Table 4(a) for Model 5 and in Table 4(b) for Model 6. In both models, sample size $n = 300$. We make three observations. First, the OWQAE delivers much superior overall performance than OLS and LAD. Second, in all cases considered, the OWQAE using heteroscedasticity-weighted quantile regression (22) clearly outperforms the OWQAE using unweighted quantile regression (27). Third, the OWQAE with estimated weights is comparable to the QAO with theoretical optimal weights.

8.4. Nonparametric Regression Models

In our data analysis, we use the standard Gaussian kernel for $K(\cdot)$. We now address the bandwidth selection issue. By (39), the optimal bandwidth h^* is proportional to $(s^2)^{1/5}$. Denote by h^*_{LS} and h^*_{OWQAE} the bandwidth for the

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

TABLE 4. (Location-scale model) Empirical relative efficiency of OWQAE with estimated weights compared to four methods: OLS, LAD, QAO, OWQAE using unweighted quantile regression in (27). Mixture 1: 0.5N(0,1)+0.5N(0,0.5^6); Mixture 2: 0.5N(−2,1)+0.5N(2,1). For Student-t_1, t_2, due to infinite variance, the OLS is not stable and varies significantly in simulations [numbers ≥ 1 indicate better performance of OWQAE]

(a) Model 5 in (60)

	Location parameter				Scale parameter							
ε_t	$\beta = 0.6$				$\gamma_0 = 0.5$				$\gamma_1 = 1.0$			
Distribution	OLS	LAD	QAO	QAEU	OLS	LAD	QAO	QAEU	OLS	LAD	QAO	QAEU
Student-t_1	NA	0.96	0.85	1.17	NA	1.17	0.88	1.20	NA	1.22	0.92	1.25
Student-t_2	NA	1.05	0.91	1.13	NA	1.66	0.94	1.28	NA	1.70	0.96	1.27
N(0,1)	0.89	1.43	0.93	1.12	0.68	2.29	0.92	1.28	0.68	2.37	0.94	1.26
Mixture 1	7.05	1.21	0.91	1.18	0.55	1.88	0.85	1.23	0.55	2.05	0.89	1.27
Mixture 2	2.57	23.03	1.03	1.23	0.66	2.07	0.86	1.28	0.69	2.58	0.86	1.20
Laplace	1.53	0.80	0.80	1.11	0.93	1.89	0.96	1.25	0.96	1.81	0.96	1.21
Gamma(1)	7.38	6.40	0.79	1.17	7.31	3.72	0.50	1.13	6.38	3.18	0.44	1.14
Gamma(2)	2.32	2.89	0.92	1.13	2.70	2.61	0.79	1.23	2.60	2.76	0.81	1.21
Gamma(3)	1.67	2.37	0.92	1.13	1.79	2.63	0.83	1.15	1.76	2.84	0.88	1.20
Beta(1,1)	1.48	4.39	0.95	1.09	0.66	3.87	0.83	1.19	0.65	3.56	0.85	1.17
Beta(1,2)	1.95	4.48	0.87	1.12	1.04	3.41	0.71	1.10	1.07	3.75	0.73	1.21
Beta(1,3)	2.75	4.97	0.84	1.11	1.71	3.64	0.69	1.14	1.63	3.30	0.65	1.10

(b) Model 6 in (61)

	Location parameter				Scale parameter							
ε_t	$\beta = 0.4$				$\gamma_0 = 0.5$				$\gamma_1 = 0.5$			
Distribution	OLS	LAD	QAO	QAEU	OLS	LAD	QAO	QAEU	OLS	LAD	QAO	QAEU
Student-t_1	NA	0.65	0.55	NA	NA	16.84	0.37	NA	NA	3.85	0.30	NA
Student-t_2	NA	1.03	0.89	6.76	NA	3.69	0.88	3.38	NA	4.01	0.96	6.02
N(0,1)	0.89	1.45	0.97	1.12	0.62	1.94	0.92	1.21	0.67	2.10	0.97	1.19
Mixture 1	6.47	1.12	0.85	1.22	0.51	1.76	0.77	1.07	0.60	1.92	0.99	1.10
Mixture 2	2.40	22.77	1.00	6.59	0.70	20.72	0.91	34.50	0.64	9.51	0.84	5.82
Laplace	1.48	0.80	0.81	1.53	0.92	2.31	0.92	1.50	0.91	2.35	0.97	1.44
Gamma(1)	5.71	6.16	0.78	1.10	5.13	3.65	0.40	1.12	7.79	5.61	0.62	1.21
Gamma(2)	2.13	2.86	0.93	1.22	2.63	4.37	0.69	1.58	2.73	4.64	0.80	1.53
Gamma(3)	1.44	2.04	0.95	1.57	1.89	4.55	0.84	2.12	1.77	3.85	0.84	1.76
Beta(1,1)	1.59	4.30	0.92	1.01	0.67	3.14	0.71	1.00	0.83	3.79	0.98	1.01
Beta(1,2)	1.73	4.27	0.86	0.98	1.00	2.44	0.65	0.99	1.15	3.21	0.89	0.99
Beta(1,3)	2.48	5.17	0.81	1.01	1.50	2.45	0.54	0.98	1.48	3.79	0.87	1.00

least squares estimator and the proposed OWQAE. Then $h^*_{\mathrm{OWQAE}} = h^*_{\mathrm{LS}}[S(\omega^*)/\mathrm{var}(\varepsilon)]^{1/5}$, where $S(\omega^*)$ is defined in (13). In practice, to select h^*_{LS}, we can use the plug-in bandwidth selector in Ruppert, Sheather, and Wand (1995),

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

implemented using the command `dpill` in the `R` package `KernSmooth`. We then select h^*_{OWQAE} by plugging in estimates of $S(\omega^*)$ and $\text{var}(\varepsilon)$ using the two-step procedure in Section 7, but for the purpose of comparison we shall use their true values in our simulation studies. Similarly, we can choose the optimal bandwidths for the other two estimators. Kai, Li, and Zou (2010) adopted the same strategy. For the preliminary estimator $\widehat{m}(\cdot)$ in step (i) of Section 7, we use the plug-in bandwidth selector h^*_{LS}.

We compare the empirical performance of the four methods (LS, LAD, CQR, OWQAE) in Section 5. With 1000 realizations, we use the least squares estimator $\widehat{m}_{\text{LS}}$ as the benchmark to which the other three methods are compared based on the relative efficiency:

$$\text{RE}(\widehat{m}) = \frac{\text{MISE}(\widehat{m}_{\text{LS}})}{\text{MISE}(\widehat{m})} \quad \text{and} \quad \text{MISE}(\widehat{m}) = \frac{1}{1000}\sum_{j=1}^{1000}\int_{\ell_1}^{\ell_2}[\widehat{m}_j(x) - m(x)]^2 dx,$$

where $\widehat{m}_j$ is the estimator in the j-th realization, and $[\ell_1, \ell_2]$ is the interval over which m is estimated. A value of RE ≥ 1 indicates that $\widehat{m}$ outperforms the LS estimator. To facilitate computation, the integral is approximated using 20 grid points.

Consider $n = 200$ samples from the model

$$\text{Model 7}: \quad Y = \sin(2X) + 2\exp(-16X^2) + 0.5\varepsilon, \quad X \sim Unif[-1.6, 1.6]. \quad \textbf{(64)}$$

The same model was also used in Kai, Li, and Zou (2010) with the normal design $X \sim$ N(0,1). Here we use the uniform design to avoid some computational issues. For ε, we consider nine symmetric distributions: N(0,1), truncated normal on $[-1, 1]$, truncated Cauchy on$[-10, 10]$, truncated Cauchy on $[-1, 1]$, Student-t with 3 (t_3) degrees of freedom, Standard Laplace distribution, uniform distribution on $[-0.5, 0.5]$, and two normal mixture distributions: 0.5N(2,1)+0.5N(−2,1) and 0.95N(0,1)+0.05N(0,9). The first normal mixture can be used to model a two-cluster population, whereas the second normal mixture can be viewed as a noise contamination model. Let $[\ell_1, \ell_2] = [-1.5, 1.5]$.

The relative efficiencies of the four methods, with $\widehat{m}_{\text{LS}}(x)$ as the benchmark, are summarized in Table 5(a). Overall, OWQAE either significantly outperforms or is comparable to the other three methods. For example, for N(0,1) and 0.95N(0,1)+0.05N(0,9), OWQAE is comparable to LS; for other distributions, OWQAE has about 20% efficiency gain over LS for most distributions and more than 60% efficiency gain for 0.5N(2,1)+0.5N(−2,1). When compared with CQR, OWQAE outperforms CQR for all but the four distributions: N(0,1), Student-t_3, Laplace distribution, and 0.95N(0,1)+0.05N(0,9), for which they are comparable. While OWQAE underperforms LAD for truncated Cauchy on $[-10, 10]$, it has substantial efficiency gains for N(0,1), truncated N(0,1) on $[-1, 1]$, truncated Cauchy on $[-1, 1]$, uniform on $[-0.5, 0.5]$, and 0.5N(2,1)+0.5 N(−2,1).

The empirical performance of the proposed method in Table 5(a) is not as impressive as its theoretical performance in (40). For example, for truncated

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

TABLE 5. (Nonparametric regression model) Empirical relative efficiency of the local least-absolute-deviation estimator (LAD), Kai, Li and Zou (2010)'s local CQR estimator, and the proposed OWQAE, relative to the benchmark local LS estimator. For CQR and OWQAE: $\tau_j = j/(k+1)$, $j = 1, \ldots, k$, $k = 9, 19, 29$

(a) Model 7 in (64)

			CQR, $k =$			OWQAE, $k =$		
Distribution of ε	LS	LAD	9	19	29	9	19	29
N(0,1)	1	0.73	0.99	1.00	0.99	0.96	0.95	0.95
Truncated N(0,1) on $[-1,1]$	1	0.54	0.93	0.98	0.99	1.14	1.25	1.26
Truncated Cauchy on $[-10,10]$	1	1.67	1.16	1.07	1.05	1.40	1.27	1.21
Truncated Cauchy on $[-1,1]$	1	0.53	0.93	0.98	0.99	1.07	1.20	1.20
Student-t with 3 d.f.'s	1	1.44	1.39	1.21	1.16	1.42	1.28	1.19
Standard Laplace	1	1.26	1.11	1.06	1.05	1.16	1.10	1.06
Uniform on $[-0.5,0.5]$	1	0.51	0.94	0.98	0.99	1.23	1.24	1.21
0.5N(−2,1)+0.5N(2,1)	1	0.31	0.92	0.97	0.99	1.57	1.60	1.62
0.95N(0,1)+0.05N(0,9)	1	0.88	1.13	1.08	1.06	1.05	0.98	0.95

(b) Model 8 in (65)

			CQR, $k =$			OWQAE, $k =$		
Distribution of ε	LS	LAD	9	19	29	9	19	29
N(0,1)	1	0.66	0.96	0.98	0.98	0.94	0.95	0.95
Truncated N(0,1) on $[-1,1]$	1	0.43	0.84	0.91	0.93	1.14	1.56	1.80
Truncated Cauchy on $[-10,10]$	1	2.39	1.35	1.17	1.12	1.97	1.78	1.61
Truncated Cauchy on $[-1,1]$	1	0.46	0.84	0.91	0.94	1.02	1.35	1.55
Student-t with 3 d.f.'s	1	1.73	1.75	1.52	1.41	1.72	1.51	1.44
Standard Laplace	1	1.48	1.18	1.11	1.09	1.31	1.24	1.16
Uniform on $[-0.5,0.5]$	1	0.36	0.84	0.91	0.94	1.46	2.21	2.70
0.5N(−2,1)+0.5N(2,1)	1	0.20	0.88	0.95	0.97	2.12	2.18	2.17
0.95N(0,1)+0.05N(0,9)	1	0.86	1.18	1.15	1.12	1.10	1.02	0.95

N(0,1) on $[-1,1]$, the theoretical AREs according to (40) are 1, 0.48, 0.86, 0.93, 0.95, 1.13, 1.69, 2.17, compared to 1, 0.54, 0.93, 0.98, 0.99, 1.14, 1.25, 1.26 in Table 5(a). To explain this phenomenon, the plot (not included here) of the function$m(x) = \sin(2x) + 2\exp(-16x^2)$ exhibits large curvature and sharp changes on$[-0.5, 0.5]$, and thus a large estimation bias could easily offset the asymptotic efficiency improvements, especially for a moderate sample size. To appreciate this, use the same X and ε in (64), and consider model

$$\text{Model 8}: \quad Y = 1.8X + 0.5\varepsilon. \tag{65}$$

Then the bias term $h^2\mu_K m''(x) = 0$ vanishes and the variance plays a dominating role. For all four estimation methods, we use the same bandwidth:

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

the plug-in bandwidth selector for local linear regression. We summarize the relative efficiencies in Table 5(b). The overall pattern of the empirical relative efficiencies is consistent with that of the theoretical ones in (40), and the proposed OWQAE significantly outperforms other methods for almost all distributions considered. Also, using more quantiles ($k = 29$) significantly improves the performance of OWQAE for truncated N(0,1) on $[-1, 1]$, truncated Cauchy on $[-1, 1]$, and uniform on $[-0.5, 0.5]$. The latter property is not shared by the CQR method.

In summary, for most nonnormal distributions considered, the proposed method can have substantial efficiency improvements over other methods, and the empirical performance is consistent with our asymptotic theory.

9. AN EMPIRICAL APPLICATION

To highlight the proposed approach, we consider a simple application of this method to the widely studied cross-section of stock returns. The Capital Asset Pricing Model (CAPM, see Sharpe, 1964; Black, 1972) has long served as the backbone of both theoretical and empirical finance. It is generally agreed that investors demand a higher expected return for investment in riskier securities. Over the past three decades a number of studies have empirically examined the performance of the CAPM in the cross-section of returns, and it is also well documented that the rate of return to holding common stocks is to some extend predictable over time. A large number of papers have studied the appropriateness of CAPM model in explaining how investors assess the risk and how they determine what risk premium to demand, and several alternative models have also been proposed in the literature. However, empirical evidence is ambiguous. The support for other asset-pricing models is no better. In addition, the theory behind the CAPM has an intuitive appeal that other models lack. For these (and other) reasons, in spite of the controversy in empirical studies, the CAPM is still widely used in financial applications and still the preferred model used in MBA and other managerial finance courses.

The focus of this section is not on the choice of alternative models. In this section, we consider applications of the methods that we discussed in the previous sections on the traditional widely used CAPM cross-sectional regression (similar to those of Fama-MacBeth which can be used to study the predictability of returns). The cross-sectional regression equation at time t is

$$R_{i,t} = \lambda_0 + \lambda_1 \beta_{im,t-1} + \varepsilon_{i,t}, \tag{66}$$

where λ_0 is the intercept term, λ_1 is the slope coefficient, and $\beta_{im,t-1}$ is the conditional beta of the excess return for asset i in month t. The dating convention indicates that the conditional beta is formed using only information available at time $t-1$. This regression model provides a decomposition of each excess return over each period into two components: the first component, $\lambda_1 \beta_{im,t-1}$, represents the part of return of asset i that is related to the cross-sectional structure of risk,

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

as measured by the betas. The remaining component of the return is uncorrelated to the measures of risk. Thus, the asset pricing model implies that the predictability of returns should be related to the risk.

We consider a population of stocks traded on the New York Stock Exchange (NYSE) from January 2009 to December 2010. We study monthly stock returns. These data are available from CRSP (the Center for Research in Security Prices) as well as many other data resources. Following the literature of many empirical studies, the stocks are considered if their returns in the current month and also the previous 60 month are available, and we exclude firms with negative book-to-market equity (using information from Compustat). The cross-sectional regression model (66) is usually estimated by the least squares method in practice. On the other hand, cumulated empirical evidence in finance indicates that stock returns are not normally distributed. In fact, it is well-known that the distributions of returns are heavy-tailed. Therefore, it is important to consider estimation procedures which have good properties in the absence of Gaussianity.

We estimate the cross-sectional regression model (66) using four methods: the traditional OLS regression, the LAD estimation, a simple equally weighted quantile averaging estimation (denoted by QAU), and the optimally weighted quantile averaging estimation (denoted by OWQAE). We use $k = 9$ quantiles $0.1, 0.2, \ldots, 0.9$, for quantile combination. For the purpose of comparison, we evaluate the performance of these estimators based on their out-of-sample prediction. In particular, we estimate the cross-sectional regression model (66) based on cross-sectional data at each month of 2009, and then use the estimated coefficients $\widehat{\lambda}_0$ and $\widehat{\lambda}_1$ to construct forecast of return at the corresponding month of 2010. We compare both the mean squared prediction error (MSE) and the mean absolute deviation (MAD) of the predictions. In particular, we calculate the mean squared prediction error and the mean absolute prediction error by

$$\text{MSE} = \sum_i \left(R_{i,t+1} - \widehat{R}_{i,t+1}\right)^2, \ \text{MAD} = \sum_i \left|R_{i,t+1} - \widehat{R}_{i,t+1}\right|,$$

for each month, and then average these mean squared prediction errors and mean absolute prediction errors respectively over all months. Table A1 summarizes the results.

Table A1. Prediction Errors Based on (66)

	OLS	LAD	QAU	OWQAE
MAD	51.39	46.94	48.35	44.78
MSE	10.50	8.89	9.45	7.95

Numbers in this table are multiplied by 500 for convenience.

Model (66) is the basic regression model that characterizes the risk premiums. We next consider an extension of model (66) which adds conditional

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

heteroscedastic effect of capitalization (the "size" effect). We consider an analog of (20),

$$R_{i,t} = \lambda_0 + \lambda_1 \beta_{im,t-1} + \sigma_{i,t}\varepsilon_{i,t}, \tag{67}$$

where $\sigma_{i,t} = \gamma\, Cap_{i,t}$ and $Cap_{i,t}$ is the market capitalization. Again, we estimate the cross-sectional regression model (67) using the four estimation methods mentioned before. More specifically, the two-stage Weighted Least Squares (WLS), the two-stage Weighted Least Absolute Deviation (WLAD), the QAU and OWQAE based on quantiles $0.1, 0.2, \ldots, 0.9$. Table A2 reports the mean squared prediction errors (MSE) and the mean absolute prediction errors (MAD) that are calculated in a similar way as Table A1.

Table A2. Prediction Errors Based on (67)

	WLS	WLAD	QAU	OWQAE
MAD	51.01	46.80	48.30	44.09
MSE	10.40	8.86	9.41	7.90

Numbers in this table are multiplied by 500 for convenience.

The empirical results from Tables A1–A2 indicate that least squares method-based estimation is less efficient than other methods. In particular, the proposed OWQAE estimator performs relatively better than other methods.

10. FURTHER DISCUSSIONS

We propose a general method of combining quantile regression information to improve efficiency of regression estimators. The proposed method is simple and more efficient regression estimators can be constructed based on a relatively small number of quantiles.

The proposed method has a wide range of applicability and can be potentially applied to many other models. We briefly discuss a few directions of interesting applications of our approach, without giving full details.

The first direction is efficient estimation for the varying-coefficient model:

$$Y = \alpha(U) + X^T\beta(U) + \varepsilon, \tag{68}$$

where $\alpha(\cdot)$ is the functional intercept and $\beta(\cdot)$ is the p-dimensional column vector of functional coefficients. Then the conditional τ-th quantile of Y given (X, U) is

$$Q_Y(\tau|X,U) = \alpha_\tau(U) + X^T\beta(U) \quad \text{with} \quad \alpha_\tau(U) = \alpha(U) + Q_\varepsilon(\tau).$$

A useful application is the varying-coefficient longitudinal model when we have longitudinal measurements from multiple subjects. Wang, Zhu, and Zhou (2009)

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

studied quantile regression for a partially linear varying-coefficient longitudinal model. In their work, the coefficients depend on the quantile, and they estimated the coefficients for each quantile without combining information across quantiles. We will explore further in a future paper.

A second direction is volatility estimation in time series. In financial econometrics, volatility plays an important role in asset pricing and risk management. Here we briefly discuss volatility estimation for both parametric and nonparametric ARCH models.

(Nonparametric volatility) Consider nonparametric ARCH(1): $X_t = \sigma(X_{t-1})\varepsilon_t$. Let $Q_{\varepsilon^2}(\tau)$ be the τ-th quantile of ε_t^2, and $Q_{X_t^2|X_{t-1}}(\tau)$ the conditional τ-th quantile of X_t^2 given X_{t-1}. Then $Q_{X_t^2|X_{t-1}=x}(\tau) = \sigma^2(x)Q_{\varepsilon^2}(\tau)$ and $Q_{X_t^2|X_{t-1}=x}(\tau)/Q_{X_t^2|X_{t-1}=0}(\tau) = \sigma^2(x)/\sigma^2(0)$ for all τ. Given estimates $\widehat{Q}_{X_t^2|X_{t-1}=x}(\tau)$ of $Q_{X_t^2|X_{t-1}=x}(\tau)$, we can construct efficient estimators of $\sigma^2(x)/\sigma^2(0)$ by combining $\widehat{Q}_{X_t^2|X_{t-1}=x}(\tau_j)/\widehat{Q}_{X_t^2|X_{t-1}=0}(\tau_j)$, $j = 1, \ldots, k$.

(Parametric volatility) Consider parametric ARCH(p) model:

$$X_t = \sigma_t \varepsilon_t \quad \text{and} \quad \sigma_t^2 = \beta_0 + \beta_1 X_{t-1}^2 + \cdots + \beta_p X_{t-p}^2, \quad \beta_0 > 0, \beta_1, \ldots, \beta_p \geq 0.$$

Let $\mathcal{I}_{t-1}$ be the information up to time $t-1$. Denote by $Q_{\varepsilon^2}(\tau)$ the τ-th quantile of ε_t^2, and by $Q_{X_t^2|\mathcal{I}_{t-1}}(\tau)$ the conditional τ-th quantile of X_t^2 given $\mathcal{I}_{t-1}$. Then

$$Q_{X_t^2|\mathcal{I}_{t-1}}(\tau) = \beta_0(\tau) + \sum_{j=1}^{p} \beta_j(\tau) X_{t-j}^2 \quad \text{and} \quad \beta_j(\tau) = \beta_j Q_{\varepsilon^2}(\tau), \ j = 0, \ldots, p.$$

Therefore, we can apply quantile regression with quantile τ to obtain consistent estimates $\widehat{\beta}_j(\tau)$ of $\beta_j(\tau)$. Note that $\beta_j(\tau)/\beta_0(\tau) = \beta_j/\beta_0$ for all τ. Therefore,

$$\frac{\widehat{\beta}_0(\tau) + \sum_{j=1}^{p} \widehat{\beta}_j(\tau) X_{t-j}^2}{\widehat{\beta}_0(\tau)} \approx \frac{\beta_0 + \sum_{j=1}^{p} \beta_j X_{t-j}^2}{\beta_0} = \sigma_t^2/\beta_0, \quad \text{for all } \tau \in (0,1).$$

We can construct efficient estimators of σ_t^2/β_0 by combining quantiles $\tau_1, \ldots, \tau_k$.

Similar ideas also apply to generalized ARCH models. Since substantial work is needed here, we will explore further in a separate project.

NOTE

1. Note that different forms of the location-scale model can be studied similarly to our analysis in this section, and optimally weighted quantile averaging estimators can be constructed (but the construction of optimal weights will be different, depending on the specific form of the regression model).

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

REFERENCES

Akahira, M. & K. Takeuchi (1995) *Non-Regular Statistical Estimation*. Springer.

Beran, R. (1974) Asymptotically efficient adaptive rank estimates in location models. *Annals of Statistics* 2, 248–266.

Bickel, P. (1982) On adaptive estimation. *Annals of Statistics* 10, 647–671.

Black, F. (1972) Capital market equilibrium with restricted borrowing. *Journal of Business* 45, 444–454.

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

Chamberlain, G. (1994) Quantile regression, censoring and the structure of wages. *Advances in Econometrics, Sixth World Congress*, vol. 1, pp. 171–210.

Chen, X., O. Linton, & D. Jacho-Chavez (2011) Averaging of Moment Condition Estimators. Working paper.

Engle, R.F. (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of U.K. inflation. *Econometrica* 50, 987–1007.

Fan, J., M. Farmen, & I. Gijbels (1998) Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society, Series B* 60, 591–608.

Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and Its Applications*. Chapman and Hall.

He, X. & Q.M. Shao (1996) A general Bahadur representation of M-estimators and its application to linear regression with nonstochastic designs. *Annals of Statistics* 24, 2608–2630.

Jurečková, J. & B. Procházka (1994) Regression quantiles and trimmed least squares estimator in nonlinear regression model. *Journal of Nonparametric Statistics* 3, 201–222.

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

Knight, K. (1989) Limit Theory for Autoregressive-parameter estimates in an infinite-variance random walk. *The Canadian Journal of Statistics* 17, 261–278.

Koenker, R. (1984) A note on L-estimates for linear models. *Statistics & Probability Letters* 2, 323–325.

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

Koenker, R. & G. Bassett (1978) Regression quantiles. *Econometrica* 46, 33–49.

Koenker, R. & Q. Zhao (1994) L-estimation for linear heteroscedastic models. *Journal of Nonparametric Statistics* 3, 223–235.

Koenker, R. & Q. Zhao (1996) Conditional quantile estimation and inference for ARCH models. *Econometric Theory* 12, 793–813.

Pollard, D. (1991) Asymptotics for least absolute deviation regression estimators. *Econometric Theory* 7, 186–199.

Portnoy, S. & R. Koenker (1989) Adaptive *L*-estimation of linear models. *Annals of Statistics* 17, 362–381.

Ruppert, D., S.J. Sheather, & M.P. Wand (1995) An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association* 90, 1257–1270.

Sharpe, W.F. (1964) Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance* 19, 425–442.

Silverman, B.W. (1986) *Density Estimation*. Chapman and Hall.

Stone, C. (1975) Adaptive maximum likelihood estimation of a location parameter. *Annals of Statistics* 3, 267–284.

Stout, W.F. (1974) *Almost Sure Convergence*. Academic Press.

Wang, H., Z. Zhu, & J. Zhou (2009) Quantile regression in partially linear varying coefficient models. *Annals of Statistics* 37, 3841–3866.

Xiao, Z. & R. Koenker (2009) Conditional quantile estimation and inference for GARCH models. *Journal of the American Statistical Association* 104, 1696–1712.

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

Yu, K. & M.C. Jones (1998) Local linear quantile regression. *Journal of the American Statistical Association* 93, 228–237.

Zhao, Q. (2001) Asymptotically efficient median regression in the presence of heteroskedasticity of unknown form. *Econometric Theory* 17, 765–784.

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

APPENDIX: Proofs

Proof of Theorem 3.1. By the ergodicity in Assumption 2.1(ii) and (5),

$$\frac{D_n^T D_n}{n} \to \begin{bmatrix} 1 & \mathbb{E}\left[\dot{m}(X,\beta)^T\right] \\ \mathbb{E}[\dot{m}(X,\beta)]\, \mathbb{E}\left[\dot{m}(X,\beta)\dot{m}(X,\beta)^T\right] \end{bmatrix} := \Sigma, \quad \text{in probability.} \tag{69}$$

Recall the definition of Σ_β in Theorem 3.1. Then we can easily verified that

$$\Sigma^{-1} = \begin{bmatrix} 1+\mathbb{E}\left[\dot{m}(X,\beta)^T\right]\Sigma_\beta^{-1}\mathbb{E}[\dot{m}(X,\beta)] & -\mathbb{E}\left[\dot{m}(X,\beta)^T\right]\Sigma_\beta^{-1} \\ -\Sigma_\beta^{-1}\mathbb{E}[\dot{m}(X,\beta)] & \Sigma_\beta^{-1} \end{bmatrix}. \tag{70}$$

By Assumption 3.1 and (69)–(70),

$$\begin{aligned} \widehat{\beta}(\tau) = \beta + \frac{\Sigma_\beta^{-1}+o_p(1)}{nf_\varepsilon(Q_\varepsilon(\tau))}\sum_{t=1}^n \{\dot{m}(X_t,\beta)-\mathbb{E}[\dot{m}(X,\beta)]\}\,[\tau-\mathbf{1}_{\varepsilon_t<Q_\varepsilon(\tau)}] \\ +o_p\left(n^{-1/2}\right). \end{aligned} \tag{71}$$

Therefore,

$$\begin{aligned} \sqrt{n}\left[\sum_{j=1}^k \omega_j \widehat{\beta}(\tau_j)-\beta\right] = \frac{\Sigma_\beta^{-1}+o_p(1)}{\sqrt{n}} \\ \times \sum_{t=1}^n \{\dot{m}(X_t,\beta)-\mathbb{E}[\dot{m}(X,\beta)]\}d_t + o_p(1), \end{aligned} \tag{72}$$

where

$$d_t = \sum_{j=1}^k \frac{\omega_j}{f_\varepsilon(Q_\varepsilon(\tau_j))}\left[\tau_j - \mathbf{1}_{\varepsilon_t<Q_\varepsilon(\tau_j)}\right]. \tag{73}$$

By the Cramér–Wold device, it suffices to consider the case that $\dot{m}(X_t,\beta)$ is scalar-valued. Let $\mathcal{F}_t$ be the σ-algebra generated by $\{X_{t+1}, X_t, \ldots; \varepsilon_t, \varepsilon_{t-1}, \ldots\}$. By property (P1) in Section 2, $\{\{\dot{m}(X_t,\beta)-\mathbb{E}[\dot{m}(X,\beta)]\}d_t\}_{t\in\mathbb{Z}}$ form martingale differences with respect to $\{\mathcal{F}_t\}_{t\in\mathbb{Z}}$. Using $\operatorname{cov}\{\tau-\mathbf{1}_{\varepsilon_t\le Q_\varepsilon(\tau)}, \tau'-\mathbf{1}_{\varepsilon_t\le Q_\varepsilon(\tau')}\} = \min(\tau,\tau')-\tau\tau'$, we have

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

$\mathbb{E}(d_t^2) = S(\omega)$ with $S(\omega)$ defined in (13). Since ε_t is independent of $\mathcal{F}_{t-1}$, by (5),

$$\frac{1}{n}\sum_{t=1}^{n}\mathbb{E}\left[\left(\{\dot{m}(X_t,\beta)-\mathbb{E}[\dot{m}(X,\beta)]\}d_t\right)^2\Big|\mathcal{F}_{t-1}\right]$$
$$=\frac{S(\omega)}{n}\sum_{t=1}^{n}\{\dot{m}(X_t,\beta)-\mathbb{E}[\dot{m}(X,\beta)]\}^2\to\Sigma_\beta S(\omega). \tag{74}$$

Since k is fixed, by Assumption 2.2, d_t is bounded. Thus, the assumption $\dot{m}(X_t,\beta)\in\mathcal{L}^2$ ensures the Lindeberg condition. The result then follows from the martingale CLT. ■

Proof of Theorem 3.2. The optimal weight follows from the Lagrange multiplier method. The asymptotic normality follows from Theorem 3.1 and $S(\omega^*) = 1/\left(e_k^T H^{-1} e_k\right)$. ■

Proof of Theorem 3.3. Recall d_t in (72)–(73). Since k is fixed, by Slutsky's theorem, it is easy to see that $\sum_{t=1}^{n}\{\dot{m}(X_t,\beta)-\mathbb{E}[\dot{m}(X,\beta)]\}d_t$ has the same asymptotic distribution if we replace ω_j therein by any $\widetilde{\omega}_j$ such that $\widetilde{\omega}_j = \omega_j + o_p(1)$. Thus, the result follows. ■

Proof of Theorem 4.1. Define the vectors

$$V_t=\begin{bmatrix}X_t\\U_t\end{bmatrix},\ \theta=\begin{bmatrix}b\\r\end{bmatrix},\ \theta(\tau)=\begin{bmatrix}\beta\\\gamma(\tau)\end{bmatrix},\ \widehat{\theta}(\tau)=\begin{bmatrix}\widehat{\beta}(\tau)\\\widehat{\gamma}(\tau)\end{bmatrix},\delta=\sqrt{n}[\theta-\theta(\tau)].$$

Then

$$Y_t - X_t^T b - U_t^T r = U_t^T\gamma[\varepsilon_t - Q_\varepsilon(\tau)] - V_t^T\delta/\sqrt{n}.$$

Since $\widehat{\theta}(\tau)$ minimizes the criterion function in (22), the reparametrized parameter $\widehat{\delta} = \sqrt{n}\left[\widehat{\theta}(\tau)-\theta(\tau)\right]$ minimizes the loss function

$$L(\delta)=\sum_{t=1}^{n}\frac{1}{U_t^T\widetilde{\gamma}}\left[\rho_\tau\left\{U_t^T\gamma[\varepsilon_t-Q_\varepsilon(\tau)]-V_t^T\delta/\sqrt{n}\right\}-\rho_\tau\left\{U_t^T\gamma[\varepsilon_t-Q_\varepsilon(\tau)]\right\}\right].$$

Suppose we can establish the quadratic approximation

$$L(\delta) = L^*(\delta) + o_p(1), \tag{75}$$

where

$$L^*(\delta) = -\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\tau-\mathbf{1}_{\varepsilon_t<Q_\varepsilon(\tau)}\right]\frac{V_t^T\delta}{U_t^T\gamma}+\frac{f_\varepsilon(Q_\varepsilon(\tau))}{2}\delta^T\mathbb{E}\left[\frac{V_0V_0^T}{\left(U_0^T\gamma\right)^2}\right]\delta. \tag{76}$$

Then the convexity lemma in Pollard (1991) gives $\widehat{\delta} = \widehat{\delta}^* + o_p(1)$, where

$$\widehat{\delta}^* = \operatorname*{argmin}_\delta L^*(\delta) = \frac{1}{f_\varepsilon(Q_\varepsilon(\tau))}\left\{\mathbb{E}\left[\frac{V_0V_0^T}{(U_0^T\gamma)^2}\right]\right\}^{-1}$$
$$\times\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\frac{V_t}{U_t^T\gamma}\left[\tau-\mathbf{1}_{\varepsilon_t<Q_\varepsilon(\tau)}\right].$$

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

The desired result then follows by using block matrix inverse of $\{\mathbb{E}[V_0V_0^T/(U_0^T\gamma)^2]\}^{-1}$.

It remains to prove (75). In view of $L(\delta)$, define

$$\widetilde{L}(\delta)=\sum_{t=1}^{n}\frac{1}{U_t^T\gamma}\left[\rho_\tau\left\{U_t^T\gamma\,[\varepsilon_t-Q_\varepsilon(\tau)]-V_t^T\delta/\sqrt{n}\right\}-\rho_\tau\left\{U_t^T\gamma\,[\varepsilon_t-Q_\varepsilon(\tau)]\right\}\right].$$

It suffices to prove $\widetilde{L}(\delta)=L^*(\delta)+o_p(1)$ and $L(\delta)=\widetilde{L}(\delta)+o_p(1)$.

First, we prove $\widetilde{L}(\delta)=L^*(\delta)+o_p(1)$. Using $\rho_\tau(cz)=c\rho_\tau(z)$ for $c>0$, we can rewrite

$$\widetilde{L}(\delta)=\sum_{t=1}^{n}\left[\rho_\tau\left\{\varepsilon_t-Q_\varepsilon(\tau)-\frac{V_t^T\delta}{\sqrt{n}U_t^T\gamma}\right\}-\rho_\tau\left\{\varepsilon_t-Q_\varepsilon(\tau)\right\}\right].$$

Applying Knight (1989)'s identity

$$\rho_\tau(u-v)-\rho_\tau(u)=-v(\tau-\mathbf{1}_{u<0})+\int_0^v\left(\mathbf{1}_{u\le s}-\mathbf{1}_{u\le 0}\right)ds, \tag{77}$$

we can obtain

$$\begin{aligned}\widetilde{L}(\delta)=&-\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left[\tau-\mathbf{1}_{\varepsilon_t<Q_\varepsilon(\tau)}\right]\frac{V_t^T\delta}{U_t^T\gamma}\\&+\sum_{t=1}^{n}\mathbb{E}(\xi_t|\mathcal{G}_{t-1})+\sum_{t=1}^{n}\left[\xi_t-\mathbb{E}(\xi_t|\mathcal{G}_{t-1})\right],\end{aligned} \tag{78}$$

where $\mathcal{G}_t$ is the σ-algebra generated by $\{(X_{t+1},U_{t+1}),(X_t,U_t),\ldots;\varepsilon_t,\varepsilon_{t-1},\ldots\}$, and

$$\xi_t=\int_0^{\frac{V_t^T\delta}{\sqrt{n}U_t^T\gamma}}\left[\mathbf{1}_{\varepsilon_t\le Q_\varepsilon(\tau)+s}-\mathbf{1}_{\varepsilon_t\le Q_\varepsilon(\tau)}\right]ds.$$

By Assumption 4.1, ε_t is independent of $\mathcal{G}_{t-1}$. Thus,

$$\mathbb{E}(\xi_t|\mathcal{G}_{t-1})=\int_0^{\frac{V_t^T\delta}{\sqrt{n}U_t^T\gamma}}[F_\varepsilon(s+Q_\varepsilon(\tau))-F_\varepsilon(Q_\varepsilon(\tau))]ds. \tag{79}$$

By Assumption 4.2(i), there exists some constant c such that

$$\frac{|V_t^T\delta|}{U_t^T\gamma}\le c. \tag{80}$$

Thus, from (79) and Taylor's expansion $F_\varepsilon(s+Q_\varepsilon(\tau))-F_\varepsilon(Q_\varepsilon(\tau))=sf_\varepsilon(Q_\varepsilon(\tau))+o(s)$,

$$\begin{aligned}\sum_{t=1}^{n}\mathbb{E}(\xi_t|\mathcal{G}_{t-1})&=\frac{f_\varepsilon(Q_\varepsilon(\tau))}{2}\delta^T\left[\frac{1}{n}\sum_{t=1}^{n}\frac{V_tV_t^T}{(U_t^T\gamma)^2}\right]\delta+o(1)\\&\to\frac{f_\varepsilon(Q_\varepsilon(\tau))}{2}\delta^T\mathbb{E}\left[\frac{V_0V_0^T}{(U_0^T\gamma)^2}\right]\delta,\quad\text{in probability,}\end{aligned} \tag{81}$$

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

where the convergence follows from the ergodicity and (5). Since $\{\xi_t - \mathbb{E}(\xi_t|\mathcal{G}_{t-1})\}_{t\in\mathbb{Z}}$ are martingale differences with respect to $\{\mathcal{G}_{t-1}\}_{t\in\mathbb{Z}}$, by their orthogonality,

$$\mathbb{E}\left(\left\{\sum_{t=1}^{n}\left[\xi_t - \mathbb{E}(\xi_t|\mathcal{G}_{t-1})\right]\right\}^2\right) = \sum_{t=1}^{n}\mathbb{E}\left\{\left[\xi_t - \mathbb{E}(\xi_t|\mathcal{G}_{t-1})\right]^2\right\} \le \mathbb{E}\left[\left(\sqrt{n}\xi_0\right)^2\right]. \quad \textbf{(82)}$$

From (80), we have $|\sqrt{n}\xi_0| \le c\mathbf{1}_{|\varepsilon_0 - Q_\varepsilon(\tau)|\le c/\sqrt{n}}$, which combined with (82) gives $\sum_{t=1}^{n}[\xi_t - \mathbb{E}(\xi_t|\mathcal{G}_{t-1})] = o_p(1)$. Thus, by (78) and (81), we have $\widetilde{L}(\delta) = L^*(\delta) + o_p(1)$.

Next, we prove the approximation $L(\delta) = \widetilde{L}(\delta) + o_p(1)$. Let

$$\eta_t = \rho_\tau\left\{U_t^T\gamma\left[\varepsilon_t - Q_\varepsilon(\tau)\right] - V_t^T\delta/\sqrt{n}\right\} - \rho_\tau\left\{U_t^T\gamma\left[\varepsilon_t - Q_\varepsilon(\tau)\right]\right\}.$$

Then it is easy to see that

$$L(\delta) - \widetilde{L}(\delta) = \sum_{t=1}^{n}\frac{\eta_t U_t^T(\gamma - \widetilde{\gamma})}{(U_t^T\gamma)^2} + \sum_{t=1}^{n}\frac{\eta_t(U_t^T\gamma - U_t^T\widetilde{\gamma})^2}{(U_t^T\gamma)^2 U_t^T\widetilde{\gamma}} := N_1 + N_2. \quad \textbf{(83)}$$

By the same argument leading to the quadratic approximation $\widetilde{L}(\delta) = L^*(\delta) + o_p(1)$ above, we can show that each element of $\sum_{t=1}^{n}\eta_t U_t^T/(U_t^T\gamma)^2$ has a quadratic approximation of the order $O_p(1)$. Thus, $N_1 = O_p(\|\widetilde{\gamma} - \gamma\|) = o_p(1)$. For N_2, by Assumption 4.2(i)–(ii), $U_t^T\gamma - U_t^T\widetilde{\gamma} = o_p(n^{-1/4})$ and $|\eta_t| \le |V_t^T\delta|/\sqrt{n} = O(n^{-1/2})$, which give $N_2 = o_p(1)$. Thus, we conclude that $L(\delta) = \widetilde{L}(\delta) + o_p(1)$, completing the proof. ■

Proof of Theorem 4.2. The asymptotic normally follows from the Bahadur representation in Theorem 4.1 and the same martingale CLT argument in Theorem 3.1. The optimal weight follows from the Lagrange multiplier method. ■

Proof of Theorem 5.1. Write $K_t = K\{(X_t - x)/h\}$. For IID data, Kai, Li, and Zou (2010) have shown the following asymptotic representation:

$$\begin{aligned}\widehat{Q}_Y(\tau|x) - Q_Y(\tau|x) = {} & \frac{1}{2}m''(x)\mu_K h^2 \\ & + \frac{\sigma(x)}{p_X(x)f_\varepsilon(Q_\varepsilon(\tau))}\frac{1}{nh}\sum_{t=1}^{n}\left[\tau - \mathbf{1}_{\varepsilon_t < Q_\varepsilon(\tau)}\right]K_t \\ & + o_p\left(\frac{1}{\sqrt{nh}}\right).\end{aligned}$$

Examining their argument and using properties (P1) and (P2) in Section 2, we see that the asymptotic representation also holds under Assumption 2.1. Therefore, by (30),

$$\widehat{m}_{\mathrm{WQAE}}(x|\omega) = m(x) + \frac{1}{2}m''(x)\mu_K h^2 + \frac{\sigma(x)}{nhp_X(x)}\sum_{t=1}^{n}d_t K_t + o_p\left(\frac{1}{\sqrt{nh}}\right),$$

where d_t is defined in (73). The desired asymptotic normality then follows from the same martingale CLT argument in Theorem 3.1.

It is easy to see that, under the symmetric density assumption, the optimal weight ω^* in (14) automatically satisfies the symmetric weight constraint in (29). ■

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

Proof of Theorem 5.2. This follows from the same argument of Theorem 3.3. ■

Proof of Theorem 6.1. Recall $\tau_j = j/(k+1)$. Define $k \times k$ matrices Γ and P:

$$\Gamma = \left[\min(\tau_j, \tau_{j'}) - \tau_j \tau_{j'}\right]_{1 \le j, j' \le k},$$
$$P = \operatorname{diag}\Big\{ f_\varepsilon(Q_\varepsilon(\tau_1)), \ldots, f_\varepsilon(Q_\varepsilon(\tau_k)) \Big\}. \tag{84}$$

Here "diag" stands for the diagonal matrix. By direct matrix multiplications, we can verify

$$\Gamma^{-1} = (k+1) \begin{bmatrix} 2 & -1 & 0 & 0 & \ldots \\ -1 & 2 & -1 & 0 & \ldots \\ \vdots & & & & \\ \ldots & 0 & -1 & 2 & -1 \\ 0 & 0 & \ldots & -1 & 2 \end{bmatrix}, \tag{85}$$

with $2(k+1)$ on the diagonal, $-(k+1)$ on the super-/subdiagonals, and 0 elsewhere.

(i) Recall $g(\tau) = f_\varepsilon(Q_\varepsilon(\tau))$ and $\Delta = 1/(k+1)$. By $H_k = P^{-1}\Gamma P^{-1}$ and (85),

$$\Omega_k = e_k^T P \Gamma^{-1} P e_k = (k+1)\Big\{ g^2(\tau_1) + g^2(\tau_k) + \sum_{j=2} \left[g(\tau_j) - g(\tau_{j-1})\right]^2 \Big\}$$
$$= (k+1)\left[g^2(\tau_1) + g^2(\tau_k)\right] + W_k + \int_\Delta^{1-\Delta} \left[g'(t)\right]^2 dt, \tag{86}$$

where

$$W_k = (k+1) \sum_{j=2}^{k} \left[g(\tau_j) - g(\tau_{j-1})\right]^2 - \int_\Delta^{1-\Delta} \left[g'(t)\right]^2 dt.$$

We can rewrite W_k as

$$W_k = (k+1) \sum_{j=2}^{k} \left\{ \left[\int_{\tau_{j-1}}^{\tau_j} g'(t) dt \right]^2 - (\tau_j - \tau_{j-1}) \int_{\tau_{j-1}}^{\tau_j} \left[g'(t)\right]^2 dt \right\}$$
$$= -\frac{k+1}{2} \sum_{j=2}^{k} \int_{\tau_{j-1}}^{\tau_j} \int_{\tau_{j-1}}^{\tau_j} \left[g'(t) - g'(s)\right]^2 dt ds.$$

For $t, s \in [\tau_{j-1}, \tau_j]$, we have $|g'(t) - g'(s)| = |\int_s^t g''(v) dv| \le \int_{\tau_{j-1}}^{\tau_j} |g''(v)| dv$, uniformly. Thus, by the Cauchy–Schwarz inequality,

$$\max_{t,s \in [\tau_{j-1}, \tau_j]} |g'(t) - g'(s)|^2 \le \left[\int_{\tau_{j-1}}^{\tau_j} |g''(v)| dv \right]^2 \le \Delta \int_{\tau_{j-1}}^{\tau_j} \left[g''(v)\right]^2 dv.$$

Applying the above inequality, we can obtain

$$|W_k| \le \frac{k+1}{2} \sum_{j=2}^{k} (\tau_j - \tau_{j-1})^2 \max_{t,s \in [\tau_{j-1}, \tau_j]} |g'(t) - g'(s)|^2 \le \frac{\Delta^2}{2} \int_\Delta^{1-\Delta} |g''(t)|^2 dt.$$

The result then follows from (86) and the identity $\int_0^1 [g'(\tau)]^2 d\tau = \mathcal{I}(f_\varepsilon)$ in (41).

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

(ii) Using $H = P^{-1}\Gamma P^{-1}$, we can write $\Lambda_k = q^T H^{-1} q = (Pq)^T \Gamma^{-1}(Pq)$. Note that $Pq = [h(\tau_1), \ldots, h(\tau_k)]^T$ with $h(t) = Q_\varepsilon(t) f_\varepsilon(Q_\varepsilon(t))$. Using (85) and the argument in (i) above, we can easily obtain the desired result. ■

Proof of Proposition 6.1. Let $u = Q_\varepsilon(\tau)$ so that $\tau = F_\varepsilon(u)$. Since f_ε has support $\mathbb{R}$, $u \to -\infty$ as $\tau \to 0$. Recall $g(\tau) = f_\varepsilon(Q_\varepsilon(\tau))$. By the chain rule, we can show $g''(\tau) = \left[\partial^2 \log f_\varepsilon(u)/\partial u^2\right]/f_\varepsilon(u)$. Then one can easily show that (46) is equivalent to

$$\lim_{\tau \to 0} \left\{ \frac{g^2(\tau) + g^2(1-\tau)}{\tau} + \left[|g''(\tau)| + |g''(1-\tau)|\right] \tau^{3/2} \right\} = 0. \tag{87}$$

For example, $\lim_{\tau \to 0} g^2(\tau)/\tau = 0$ if and only if $\lim_{u \to -\infty} f_\varepsilon^2(u)/F_\varepsilon(u) = 0$, and $\lim_{\tau \to 0} g^2(1-\tau)/\tau = 0$ if and only if $\lim_{u \to \infty} f_\varepsilon^2(u)/[1 - F_\varepsilon(u)] = 0$. It remains to show (87) implies (45).

Let $\epsilon > 0$ be any given number. By (87), there exists $0 < \tau_0^* < 1/2$ such that $|g''(t)| < \sqrt{\epsilon} t^{-3/2}$ and $|g''(1-t)| < \sqrt{\epsilon} t^{-3/2}$ for all $t \in (0, \tau_0^*)$. Fix τ_0^*. By Assumption 2.2, there exists $c < \infty$ such that $|g''(\tau)| < c$ for $\tau \in [\tau_0^*, 1 - \tau_0^*]$. Let $\tau^* = \min\left\{\tau_0^*, \sqrt{\epsilon}\,[c(1 - 2\tau_0^*)]^{-1/2}\right\}$. Then $\tau^{*2} c(1 - 2\tau_0^*) < \epsilon$. For $\tau < \tau^*$, applying $\int_\tau^{1-\tau} = \int_\tau^{\tau_0^*} + \int_{\tau_0^*}^{1-\tau_0^*} + \int_{1-\tau_0^*}^{1-\tau}$, we have

$$\begin{aligned} \tau^2 \int_\tau^{1-\tau} |g''(t)|^2 dt &\le \tau^2 \left(\int_\tau^{\tau_0^*} \epsilon t^{-3} dt + \int_{\tau_0^*}^{1-\tau_0^*} c\, dt + \int_\tau^{\tau_0^*} \epsilon t^{-3} dt \right) \\ &\le \tau^2 \left[\frac{\epsilon}{2\tau^2} + c(1 - 2\tau_0^*) + \frac{\epsilon}{2\tau^2} \right] \le 2\epsilon, \end{aligned}$$

completing the proof. ■

Proof of Theorem 6.3.

(i) For $S(\omega)$ in (13), $R_k = S\left((1/k, \ldots, 1/k)^T\right)$. By the uniqueness of the minimizer ω^* of $S(\omega)$ [see (14)], $R_k \ge \Omega_k^{-1}$ with equality if and only if $\omega^* = (1/k, \ldots, 1/k)^T$. Let $g(\tau) = f_\varepsilon(Q_\varepsilon(\tau))$, and for convenience write $g(\tau_0) = g(\tau_{k+1}) = 0$. For $\omega^* = (\omega_1^*, \ldots, \omega_k^*)^T$ in (14), by $H^{-1} = P\Gamma^{-1}P$ and (85), we can show

$$\omega_j^* = \frac{(k+1)[2g(\tau_j) - g(\tau_{j-1}) - g(\tau_{j+1})]g(\tau_j)}{\Omega_k}, \tag{88}$$

where Ω_k is defined in (15). Note that, for $j = \lfloor (k+1)\tau \rfloor$ with $\tau \in (0,1)$, $\lim_{k \to \infty} (k+1)^2 \left[2g(\tau_j) - g(\tau_{j-1}) - g(\tau_{j+1})\right] g(\tau_j) = -g''(\tau) g(\tau)$. Thus, as $k \to \infty$, $\omega_j^* = 1/k$ for all j implies $g''(\tau)g(\tau) = -c$, $\tau \in (0,1)$, for some $c > 0$. Define the transformation $u = Q_\varepsilon(\tau)$. By the chain rule, we can show that $g''(\tau)g(\tau) = -c$ is equivalent to $\left[f_\varepsilon''(u) f_\varepsilon(u) - f_\varepsilon'(u)^2\right]/f_\varepsilon^2(u) = -c$ or $\{\log f_\varepsilon(u)\}'' = -c$. Thus, $f_\varepsilon(u)$ must be a normal density.

(ii) See the proof in Kai, Li, and Zou (2010). ■

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press

Proof of Corollary 6.2. As in the proof of Theorem 3.1, we use martingale CLT and consider scalar-valued $\dot{m}(X_t, \beta)$. Similar to d_t in (73), with the optimal weight ω^*, define

$$d_t^* = \sum_{j=1}^{k_n} \frac{\omega_j^*}{f_\varepsilon(Q_\varepsilon(\tau_j))} \left[\tau_j - \mathbf{1}_{\varepsilon_t < Q_\varepsilon(\tau_j)}\right].$$

By Theorem 6.2, $S(\omega^*) = 1/\Omega_{k_n} \to 1/\mathcal{I}(f_\varepsilon)$. Thus, the convergence of the conditional variances follows from the argument in (74). It remains to verify the Lindeberg condition. By Assumption 2.2, $g(\tau) = f_\varepsilon(Q_\varepsilon(\tau))$ is bounded on $\tau \in (0, 1)$. Thus, from (88),

$$|d_t^*| \le \frac{k_n + 1}{\Omega_{k_n}} \sum_{j=1}^{k_n} |2g(\tau_j) - g(\tau_{j-1}) - g(\tau_{j+1})| \le c_1 k_n^2,$$

for some constant c_1. For any given $c_2 > 0$,

$$\begin{aligned} &\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}\left[\left(\{\dot{m}(X_t, \beta) - \mathbb{E}[\dot{m}(X, \beta)]\} d_t^*\right)^2 \mathbf{1}_{|\{\dot{m}(X_t,\beta) - \mathbb{E}[\dot{m}(X,\beta)]\} d_t^*| \ge c_2\sqrt{n}}\right] \\ &\quad \le c_1^2 k_n^4 \mathbb{E}\left(\{\dot{m}(X, \beta) - \mathbb{E}[\dot{m}(X, \beta)]\}^2 \mathbf{1}_{|\dot{m}(X,\beta) - \mathbb{E}[\dot{m}(X,\beta)]| \ge c_2\sqrt{n}/(c_1 k_n^2)}\right). \end{aligned} \tag{89}$$

Note that, for any random variable $U \in \mathcal{L}^q, q > 2$, and constant $c > 0$,

$$\mathbb{E}(U^2 \mathbf{1}_{|U| \ge c}) \le \mathbb{E}\left(\frac{|U|^q}{c^{q-2}} \mathbf{1}_{|U| \ge c}\right) \le \frac{\mathbb{E}(|U|^q)}{c^{q-2}}. \tag{90}$$

Applying (90) to (89) and using $k_n = O[(\log n)^\epsilon]$, we have the Lindeberg condition. ■

Proof of Theorem 6.4. (i) It follows from (86) and the proof of Theorem 6.1. (ii) From $\lim_{\tau \to 0}[g^2(\tau) + g^2(1 - \tau)]/\tau = \infty$ and (86), $1/S(\omega^*) \ge (k + 1)[g^2(\tau_1) + g^2(\tau_k)] \to \infty$. ■

https://doi.org/10.1017/S0266466614000176 Published online by Cambridge University Press