Efficient combination of conditional quantile information for the single-index model

Thomas Parker∗

September 3, 2013

Abstract

A local-linear estimator — called a quantile averaging estimator — is proposed for estimation of a single-index model, where an important restriction is placed on the distribution of the error term: that its distribution is a location-scale model. The estimator could be called a semiparametric $L$-estimator, mixing together information provided by separate single-index quantile regressions in a weighted average for both the parametric and nonparametric parts of the model. This method is related to local-linear composite quantile regression, although information provided from the conditional quantiles is used in a different way. The two methods are compared in simulation studies and empirical application of the method is illustrated an investigation of the Boston housing data.

1 Introduction

Suppose that the data may be represented using the single-index model

$$Y = g_0(X^\top \beta_0) + \sigma_0(X^\top \beta_0)\epsilon,$$  \hfill (1)

where it is assumed that the observed variables $\{y_i, X_i\}_{i=1}^n$ are iid but $g_0$ and $\sigma_0$ are unknown functions. This semiparametric model is useful for dimension reduction and has been well-studied; see for example Horowitz (2009) for a thorough introduction. Typically estimators of this model (abstracting away from the scale function $\sigma_0$) minimize the $L_2$ distance between observations and predictions:

$$\{\hat{\beta}, \hat{g}\} = \arg\min_{\beta, g} \mathbb{E} \left[ (Y - g(X^\top \beta))^2 \right].$$  \hfill (2)

While popular, this may have drawbacks in some settings. Generally speaking, this loss function results in the best possible estimators when the distribution of the error term $\epsilon$ is Gaussian; however, in

∗The author would like to thank participants at the 2013 Asian Meeting of the Econometric Society for their attention and Jungmo Yoon for his computer code. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET:www.sharcnet.ca) and Compute/Calcul Canada.
other settings it need not retain its optimality. Efficiency might be one reason: were the density of \( \epsilon \) known, then it would be more efficient to estimate the parameters of the model by optimizing a loss function based on the loglikelihood (put another way, squared-error loss corresponds to optimizing a loglikelihood based the Gaussian density). Robustness may also be an issue: for example, when \( \epsilon \) follows a distributed that has heavier tails than the normal distribution, the squared-error loss function can result in suboptimal estimates compared to the absolute-error loss function. In this paper we propose estimators that may be used to deal with either one of these issues.

Here we maintain a further important assumption, which is that the distribution of \( \epsilon \) follow a distribution that is a member of a location-scale model: \( \epsilon \sim F \in \mathcal{F} \), where

\[
\mathcal{F} = \left\{ F : F(\epsilon) = F_{\mu}(\frac{\epsilon - \mu}{\sigma}) \right\},
\]

and where \( F_{\mu} \) is a “standardized” distribution function. This assumption, while somewhat restrictive, generalizes from the typical assumptions of normality that underly least-squares-style estimation of model (1) but provides structure that allows one to estimate the model in a manner that is sensitive to the shape of \( F_{\mu} \). In particular in this paper we focus on estimators that combine information from the conditional quantiles of \( y \) to improve efficiency over least-squares-style estimation. The estimators examined here are designed for efficiency, while robustness to outliers is but a side-effect. However, these estimators could also easily be redirected towards use as robust regression estimators. The scale function \( \sigma_0 \) is treated as an infinite-dimensional nuisance parameter, although we note that the model could be extended to multi-index models of location and scale (cf. Keilegom and Wang (2010)). We do not consider it a drawback that the scale is not a fully nonparametric function — it seems reasonable to model scale semiparametrically, if one is willing to model the location in this way.

A number of recent papers have dealt with nonparametric composite quantile regression when errors can be assumed to follow a location-scale model; these include Kai et al. (2010), Sun et al. (2013) and Zhao and Xiao (2010) (also the paper by Linton and Xiao (2007) is similar in its goal, although the methodology differs). However, there is relatively little literature extending these estimators to semiparametric models. Composite quantile regression for the single-index model has been proposed in Fan et al. (2013) for the purpose of estimating \( \beta \), with an emphasis on variable selection for high-dimensional data. The focus of this paper is somewhat different. We focus on a related estimator, which we call quantile averaging estimator. The idea behind this estimator is straightforward in principle: simply estimate several single-index quantile regressions and combine them using [weighted] averages. This is inspired by the nonparametric regression proposal of Zhao and Xiao (2010).

2 The estimators

We assume that the data may be well described by a model consisting of (1) and (3). Equation (1) implies that information about the unknown functions \( g_0 \) and \( \sigma_0 \) is only available from the response \( Y \) and the index variable \( U := X^T \beta_0 \). This restriction gives the model an advantage over purely nonpara-
metric models in that it achieves dimension reduction and allows $\beta_0$ to be estimated at a parametric rate. However, the assumption made in (3) provides more structure that can be used to design an estimator, because it implies information about all the conditional quantiles of $Y$ given $X$ — or equivalently, given $U$. Equation (3) implies that the conditional quantile function of $Y$ given $U$ is linked to the quantile function of $\epsilon$ in the following way:

$$F_{Y|U}^{-1}(\tau|U) = g_0(U) + \sigma_0(U)F_\epsilon^{-1}(\tau).$$

(4)

We define $Q_0(u, \tau) := F_{Y|U}^{-1}(\tau|u)$ for an argument $u$ in the interior of the support of $U$, and because $\tau$ will almost always be fixed, we denote derivatives of $Q_0$ with respect to $u$ as $Q_0'(u, \tau)$ and $Q_0''(u, \tau)$.

Here we provide an analysis for two different but related estimators, a weighted composite quantile regression estimator and a weighted quantile averaging regression estimator. The difference between these two estimators lies in the way that information from different conditional quantiles is used.

2.1 Composite quantile regression

The (uniformly weighted) composite quantile regression estimator was introduced by Fan et al. (2013) for the single index model. We review it here because it will be used as a comparison method in the empirical investigations below. If $g_0$ were known, we would estimate $\beta_0$ by solving the sample analog of

$$\hat{\beta}_{CQR} = \min_{\beta} E\left[\rho_w\left(Y - g_0(X^T \beta)\right)\right],$$

(5)

where $\rho_w(u) = \sum_{k=1}^{m} w_k \rho_{\tau_k}(u)$, $w$ is a vector of weights and $\rho_\tau(u) = u(\tau - I(u < 0))$ is the typical quantile regression loss function. $\hat{\beta}_{W_{CQR}}$ is very nearly a classical $M$-estimator, minimizing the expected value of a special loss function, one that is made up of the contributions of $m$ different individual quantile regression loss functions. Specifically, one estimates (taking $\rho_w$ apart into its constituent parts)

$$(\hat{g}_{CQR}, \hat{g}'_{CQR}, \hat{\beta}_{CQR}) = \arg\min_{\{a_{jk}, b_j\}, \beta} \sum_{k=1}^{m} \sum_{j=1}^{n} \sum_{i=1}^{n} \rho_{\tau_k}(y_i - a_{jk} - b_j(X_i - X_j)^T \beta) \omega_{ij}$$

(6)

where $\omega_{ij} = K((U_i - U_j)/h) = \sum_{\ell} K((U_{\ell} - U_j)/h)$. Note that for each $j$ (corresponding to an estimate of $g_0(U_j)$ and its derivative) there are $m$ estimates of the function (the $\{a_{jk}\}_{k}$) and only one estimate of its derivative ($\hat{b}_j$). The characteristic that makes this estimator differ from an $M$ estimator is the fact that these $m$ different function values are estimated and subsequently averaged together to form an estimate of $g_0$.

One could also apply weights to the component quantile regression loss functions in the above minimization problem, for the reasons discussed in the introduction. Only a uniformly weighted average was considered by Fan et al. (2013), and we call this the CQR estimator. In contrast we call a composite quantile regression estimator with nonuniform weights a weighted composite quantile regression (WCQR) estimator.

Composite quantile regression is attractive because it is the solution to a single optimization prob-
lem (even if it may not strictly be an M-estimator). It also has the attractive feature that the estimator forces there to be a single estimate of $g'$ — problems associated with several estimates of the derivative function will become apparent in Section 3. On the other hand, the single problem that must be solved is quite large and can be difficult to compute. Zhao and Xiao (2010) defined optimal weights for a linear (i.e., parametric) WCQR estimator; here we investigate their claims empirically, albeit indirectly, since the estimator is tailored to a single-index model.

2.2 Quantile averaging regression

The quantile averaging estimator (QAE\(^1\)) uses several single-index quantile regressions as building blocks. The (weighted) quantile averaging estimator of $g_0$ and $\beta_0$ generally takes the form

$$
\left( \hat{g}_{QAE}(u), \hat{\beta}_{QAE} \right) = \left( \int [0,1] \varphi(\tau) \hat{Q}(u, \tau) d\tau, \int [0,1] \eta(\tau) \hat{\beta}(\tau) d\tau \right),
$$

(7)

where the estimates $\{\hat{Q}(u, \tau), \hat{\beta}(\tau)\}$ come from single-index quantile regressions and $\varphi$ and $\eta$ are (possibly different) weight functions. In most applications these integrals are weighted averages over $m$ estimates, where $m$ is a relatively small, often arbitrarily chosen number (e.g. $m = 9$). This estimator is an $L$-statistic, an extension of the estimator of Koenker and Portnoy (1987); Portnoy and Koenker (1989) or Koenker and Zhao (1994) from the linear model to the semiparametric single-index model. As such, it is not the solution to an optimization problem. However the constituent parts used to compute it are simple to obtain, are themselves $M$-estimators, and estimation is easily scalable (Mosteller, 1946).

In a single-index quantile regression one estimates $\beta(\tau)$ by solving the sample analog (once again assuming known $Q_0(\cdot, \tau)$) of

$$
\hat{\beta}(\tau) = \min_{\beta} \mathbb{E} \left[ \rho_{\tau} \left( Y - Q_0(X^T \beta, \tau_k) \right) \right].
$$

(8)

However, the single-index restriction (1) implies that $\beta(\tau) \equiv \beta_0$ and so each $\hat{\beta}(\tau_k)$ is an estimator of the same vector $\beta_0$. This means that in theory we should be able to run separate quantile regressions and average their estimated index coefficients to arrive at a more efficient estimate of $\beta_0$. An estimator of this sort was proposed by Koenker and Portnoy (1987) for the linear model and revisited in Portnoy and Koenker (1989), Koenker and Zhao (1994) and Zhao and Xiao (2010). In this paper we extend this line of research to the semiparametric single-index model.

In model (1), of course, $g_0$ and $\sigma_0$ are unknown functions. That the index coefficients across regressions should be identical is just one of two major conclusions that can be drawn from the combination of (1) and (4). The second conclusion is with regard to $g_0$. When estimating $\beta(\tau)$ in (8), we use an estimate of $Q_0(u, \tau) = g_0(u) + \sigma_0(u)F^{-1}_\epsilon(\tau)$. Consider a set of $m$ (true) conditional quantile functions

\(^1\)We chose $E$ for quantile averaging estimator, so as not to cause confusion for those used to using QAR to refer to quantile autoregression.
and a vector of weights $\varphi$ summing to one. Then under the conditions of the model,

$$\sum_{k=1}^{m} \varphi_k Q_0(u, \tau_k) = \sum_k \varphi_k g_0(u) + \sum_k \varphi_k \sigma_0(u) F_{\epsilon}^{-1}(\tau_k)$$

$$= g_0(u) + \sigma_0(u) \sum_k \varphi_k F_{\epsilon}^{-1}(\tau_k).$$

(9)

If we restrict $\varphi$ further so that it is orthogonal to the vector of quantiles of $\epsilon$, then the weighted average of the conditional quantile curves becomes an estimate of $g_0(\cdot)$. This restriction will be used in the weighted quantile averaging method proposed below. Although one goal of this paper is to investigate the effect of heteroskedasticity on the quantile averaging estimator, no special scale estimator is proposed. Another $L$-estimate for the scale function could however easily be made with the proposed quantile averaging estimator.

It is interesting to note the differences between the quantile averaging and composite quantile estimators. The major theoretical difference is that the derivative $g'_0$ is left unrestricted in the component quantile estimates of the quantile averaging estimator, while it is forced to be equal across quantiles in the composite quantile regression. This restriction seems to have two major benefits: first, the estimates are not affected by the estimation of several unnecessary coefficients that would otherwise need to be averaged as in the QA estimator, and second, it has the (unintended) effect of enforcing monotonicity of the estimated functions across quantiles. These advantages come at a modest cost. Forcing the derivative estimates to be identical during estimation while averaging across level estimates seems to be motivated from practical considerations. However, estimation can be quite costly\(^2\), making such practical considerations irrelevant. The reason behind these difficulties with the CQR estimator is computational but the fact remains that this algorithm remains somewhat impractical for applied researchers is still needed.

2.3 Estimation details

The basic quantile averaging regression uses $m$ single-index quantile regressions. These estimates are labeled $\{\hat{\beta}(\tau_k)\}$ and $\hat{Q}(\cdot, \tau_k)$ for $k = 1, 2, \ldots m$; they were estimated with the algorithm proposed in Wu et al. (2008), although the estimator of Kong and Xia (2012) could also have been used. Assuming one already has $m$ single-index quantile regressions to work with, the algorithm for the weighted quantile averaging estimator is below:

1. Make an initial estimate of $\beta_0$ by averaging the $\hat{\beta}(\tau_k)$ estimates:

$$\tilde{\beta} = \sum_{k=1}^{m} \frac{1}{m} \hat{\beta}(\tau_k).$$

(11)

2. Construct index variables $\tilde{U}_j = X_j^T \tilde{\beta}$.

\(^2\)For example, in the simulation experiments conducted below, which had sample size 200, approximately 12 gigabytes of memory were required to conduct composite quantile regression estimation.
3. Make an initial estimate of $g_0$ by averaging the $m$ preliminary estimates:

$$
\hat{g}(\tilde{U}_j) = \sum_{k=1}^{m} \frac{1}{m} \hat{Q}(\tilde{U}_j, \tau_k)
$$

(12)

4. From these estimates, estimate the scale function $\sigma_0$ using a local-linear median regression of the absolute values of the residuals $\tilde{\epsilon} = y - \hat{g}(\tilde{U})$ on the index:

$$
\hat{\sigma}(\tilde{U}_j) = \arg\min_{\sigma} \sum_{j=1}^{n} \sum_{i=1}^{n} \rho_{1/2}(\left| \tilde{\epsilon}_i \right| - \sigma - \gamma(\tilde{U}_i - \tilde{U}_j)) \frac{1}{h_{\sigma}} K \left( \frac{\tilde{U}_i - \tilde{U}_j}{h_{\sigma}} \right).
$$

(13)

5. Use $\hat{g}$ and $\hat{\sigma}$ to estimate normalized residuals, and estimate the density function $\hat{f}_\epsilon(u)$.

6. Use $\hat{f}_\epsilon$ to estimate weights $\hat{\varphi}$, and repeat steps 1 to 3 once but replacing uniform weights with $\hat{\varphi}$.

Local linear estimators were used for the constituent quantile regressions. This style of estimation appears quite popular in the literature, but we note that spline estimators could be used for the nonparametric parts. The algorithm used for individual single-index quantile regressions is given in Appendix B for reference. Also, as mentioned above, these estimators could be used in conjunction with scale estimators that could themselves be estimated using a different index than that used for the location function; see Keilegom and Wang (2010) for more on this topic. Currently an MM (majorization-minimization) method (Hunter and Lange, 2000) is used for the composite quantile regression estimator, although this seems better-suited to the sparse methods described in Koenker (2005, §6.8).

3  Asymptotic results

We make a number of assumptions on the data in order to state the asymptotic results below. It is assumed that $X \in \mathcal{X} \subseteq \mathbb{R}^p$ and $\tau \in \mathcal{T} = [\epsilon, 1 - \epsilon]$ for some $\epsilon > 0$. We first state identification conditions, which are standard (Horowitz, 2009, p.14):

A1  $g_0$ and $\sigma_0$ are twice differentiable, nonperiodic and nonconstant on supp$(U)$.

A2  $X$ has at least one continuously distributed component with a nonzero coefficient, and supp$(X)$ is not contained in a proper linear subspace of $\mathbb{R}^p$.

A3  $U := X^T \beta_0$ has density $f_U$ which is continuously differentiable and $0 < f_U(u) < \infty$ for all $u \in$ supp$(U)$, which is assumed to be a connected set.

A4  $\beta_0 = 0$ (there is no intercept term) and $\beta_1 = 1$.

Next we list conditions that ensure enough regularity that the asymptotic results below hold. These are very similar to the conditions given in Qu and Yoon (2011).
\( B1 \) \( f_\varepsilon(F_\varepsilon^{-1}(\cdot)) \) is Lipschitz continuous over \( \mathcal{T} \), and there exist \( c, \bar{c} \) and \( \delta > 0 \) such that
\[
\zeta < f_\varepsilon(F_\varepsilon^{-1}(\tau) + \eta) < \bar{c}
\]
for all \( |\eta| < \delta \) and \( \tau \in \mathcal{T} \).

\( B2 \) Letting \( \xi(\tau) = f_\varepsilon(F_\varepsilon^{-1}(\tau)) \),
\[
\lim_{\tau \to 0} \frac{\xi^2(\tau) + \xi^2(1 - \tau)}{\tau} = 0 \quad \text{and} \quad \lim_{\tau \to 0} \tau^2 \int_{-\tau}^{1-\tau} \left| \xi''(t) \right|^2 dt = 0
\]
\( B3 \) \( F_\varepsilon^{-1} \) is once differentiable with respect to \( \tau \), and \( F_\varepsilon^{-1}(\tau) \) and \( dF_\varepsilon^{-1}(\tau)/d\tau \) are finite and Lipschitz continuous over \( \mathcal{T} \).

The kernel function \( K \) is nonnegative, differentiable on the interior of its support and compactly supported. Furthermore, \( \int K(z)dz = 1, \int zK(z)dz = 0, \int z^2K(z)dz = \mu_2 \) and \( \int K^2(z)dz = \nu_0 < \infty \).

\( B4 \) The bandwidth sequence \( \{h_n\} \) is assumed to satisfy \( h_{n,\tau} = c(\tau)h_n \), where \( h_n = O(n^{-1/5}) \) and \( nh_n \to \infty \), and \( c(\tau) \) is Lipschitz continuous and \( 0 < c(\tau) < \infty \) for all \( \tau \in \mathcal{T} \).

The above identification conditions — the \( A \) assumptions — are just as in Horowitz (2009), and the \( B \) assumptions are regularity conditions similar to those in Zhao and Xiao (2010) and Qu and Yoon (2011) but tailored to the present model. The final assumption on the bandwidth is similar to other results in the literature but restricts bandwidths at neighboring quantiles to be close to one another. This is needed to obtain uniformity results for the estimator of the nonparametric part.

### 3.1 Nonparametric part

The nonparametric estimator of the \( \tau \)-th conditional quantile function is a local-linear estimator; that is, given a value \( \tilde{\beta} \) (and thus \( \tilde{U} \)), we estimate \( Q_0 \) (the \( \tau \)-th conditional quantile function) and its derivative in the regression
\[
(\hat{Q}(U_j, \tau), \hat{Q}'(U_j, \tau)) = \arg\min_{a_j, b_j} \sum_{i=1}^{n} \rho_\tau \left( y_i - a_j - b_j(U_i - U_j) \right) \omega_{ij},
\]
Thus the “intercept” estimate \( \hat{a}_j \) is an estimate of \( Q_0(U_j, \tau) \), and \( \hat{b}_j \) of \( Q'_0(U_j, \tau) = g'_0(U_j) + \sigma'_0(U_j)F_\varepsilon^{-1}(\tau) \). This is different from the slope estimate of the composite quantile regression because there it is assumed that a single “slope” parameter exists across all quantiles. The following Lemma gives the asymptotic distribution of \( \hat{Q}(u, \tau) \) and is slightly different than Wu et al. (2008). The author is currently busy using the results of Qu and Yoon (2011) to show that this result holds uniformly in \( \tau \).
Lemma 1. Let $x$ be such that $x^\top \beta_0$ is in the interior of $\text{supp}(U)$ and suppose $\hat{\beta}$ is some consistent estimate of $\beta_0$. The single-index quantile regression in the location-scale case has asymptotic distribution

$$\sqrt{n} (\hat{Q}(x^\top \hat{\beta}, \tau) - Q_0(x^\top \beta_0, \tau)) \sim N \left(0, \frac{(1-\tau)\sigma_0^2(x^\top \beta_0)\nu_0}{f_U(x^\top \beta_0)f_\varepsilon^2(F_\varepsilon^{-1}(\tau))} \right).$$

(17)

For the next few results we define a few relevant terms. First we define the vectors

$$Q_m := [F_\varepsilon^{-1}(\tau_1), \ldots, F_\varepsilon^{-1}(\tau_m)]^\top$$

and

$$q_m := [f_\varepsilon(F_\varepsilon^{-1}(\tau_1)), \ldots, f_\varepsilon(F_\varepsilon^{-1}(\tau_m))]^\top$$

and the matrix

$$H = \left\{ \frac{\tau_\ell \wedge \tau_k - \tau_\ell \tau_k}{f_\varepsilon(F_\varepsilon^{-1}(\tau_\ell))f_\varepsilon(F_\varepsilon^{-1}(\tau_k))} \right\}_{1 \leq k, \ell \leq m}.$$  

(20)

The terms of $H$ occur quite regularly in the asymptotic theory of quantile regressions; in the present case they are combined together in the asymptotic variance matrix that appears in Theorem 1 and in finding optimal weights. Next, define $\Lambda_m := I(F_\varepsilon^{-1}H^{-1}11$. Under the maintained assumptions (specifically, B2) Theorem 1 of Zhao and Xiao (2010) implies that

$$\lim_{m \to \infty} \Lambda_m = I(F_\varepsilon),$$

(21)

where $I(F_\varepsilon)$ is the Fisher information for location of the distribution of $\varepsilon$.

Finally define the scaled bias term

$$\alpha_\varphi := \frac{\mu_2 h^2}{2} \sum_{k=1}^q \varphi_k Q''(x^\top \beta_0, \tau_k) = \frac{\mu_2 h^2}{2} \left( g_\varepsilon''(x^\top \beta_0) + \sigma_0''(x^\top \beta_0) \varphi^\top Q_m \right) = O_p \left(h^2\right),$$

(22)

where the second equality comes from the constraint that the weight vector adds to one. However, a second, more troublesome bias term appears below in the statement of Theorem 1, which must be dealt with differently because it does not vanish.

Theorem 1. Given weight vector $\varphi$, the weighted quantile averaging estimator $\hat{g}_{QAE}$ of $g_0$ has asymptotic distribution

$$\sqrt{n} (\hat{g}_{QAE}(x^\top \beta) - g_0(x^\top \beta_0) - \sigma_0(x^\top \beta_0)\varphi^\top Q_m - \alpha_\varphi) \sim N \left(0, \frac{\nu_0 \sigma_0^2(x^\top \beta_0)}{f_U(x^\top \beta_0)\nu_0 \varphi^\top H \varphi} \right).$$

(23)

Calculating the optimal weight vector is straightforward given the result of Theorem 1 — minimizing the variance by adjusting $\varphi$ is a matter of solving a quadratic program because the weights must satisfy the constraint that they sum to 1. If the distribution of $\varepsilon$ has median zero and is symmetric (i.e., $F_\varepsilon^{-1}(\tau) = -F_\varepsilon^{-1}(1-\tau)$), this weighting scheme minimizes the asymptotic variance of $\hat{g}_{QAE}$ and is consistent. However, if the distribution of $\varepsilon$ cannot be assumed symmetric, then $\sigma_0(x^\top \beta_0)\varphi^\top Q_m \neq 0$
and the estimator is biased. We propose to add the constraint that the weight vector is orthogonal to the quantile vector $Q_m$.

**Corollary 1.** When $\varepsilon$ is symmetrically distributed, the weights that minimize the variance of $\hat{g}$ as $m, n \to \infty$ are

$$\varphi^* = (1^\top H^{-1} 1)^{-1} H^{-1} 1.$$  

(24)

This weighting scheme results in an estimator of $g$ with asymptotic distribution

$$\sqrt{n h} \left( \tilde{g}_{QAE}(x^\top \hat{\beta}) - g_0(x^\top \beta_0) - \frac{1}{2} g''_0(x^\top \beta_0) \mu_2 h^2 \right) \sim \mathcal{N} \left( 0, \frac{\sigma_0^2(x^\top \beta_0) \nu_0}{f_U(x^\top \beta_0)} \Lambda_m^{-1} \right).$$  

(25)

When the distribution of $\varepsilon$ is not symmetric, the weights that minimize estimation variance while maintaining a vanishing bias term as $m, n \to \infty$ are

$$\varphi^* = \frac{(q_m^\top H^{-1} q_m) H^{-1} 1}{1^\top H^{-1} 1} - \frac{(1^\top H^{-1} q_m) H^{-1} q_m}{1^\top H^{-1} 1}.$$  

(26)

This weighting scheme results in an estimator with asymptotic distribution

$$\sqrt{n h} \left( \tilde{g}_{QAE}(x^\top \hat{\beta}) - g_0(x^\top \beta_0) - \frac{1}{2} g''_0(x^\top \beta_0) \mu_2 h^2 \right) \sim \mathcal{N} \left( 0, \frac{\sigma_0^2(x^\top \beta_0) \nu_0}{f_U(x^\top \beta_0)} \right) \frac{(q_m^\top H^{-1} q_m)}{1^\top H^{-1} 1 (q_m^\top H^{-1} q_m) - (q_m^\top H^{-1} 1)^2}.$$  

(27)

Note that when the distribution of $\varepsilon$ cannot be considered symmetric the estimate is less precise:

$$\Lambda_m^{-1} = \frac{1}{1^\top H^{-1} 1} \leq \frac{1}{1^\top H^{-1} 1 - \frac{(1^\top H^{-1} q_m)^2}{q_m^\top H^{-1} q_m}} = \frac{(q_m^\top H^{-1} q_m)}{1^\top H^{-1} 1 (q_m^\top H^{-1} q_m) - (q_m^\top H^{-1} 1)^2}.$$  

(28)

However, it is hoped that this estimator might also be considered despite the fact that its variance may not have the smallest variance that is theoretically possible. The advantages of this methodology are twofold: first, the stability of the single-index quantile regression estimators is passed on to the averaging estimator, and second, the method is flexible enough that one may easily implement custom weighting depending on a particular situation.

### 3.2 Parametric part

The proposed $\hat{\beta}_{QAE}$ is also a weighted average of estimates from several quantile regressions. To achieve full semiparametric efficiency, weighting should be done using the conditional variance of $y|X$, which is straightforward since a location-scale structure has been assumed. Rather than focusing on weights that depend on observations $i$, we focus the analysis on weights over quantiles — that is, how best to combine the several conditional quantile estimates $\{\hat{\beta}(\tau_k)\}_k$ into a single estimate $\hat{\beta}$.

Asymptotic covariance matrices will depend on the relationship between data $X$ and the index $U$, as
well as on the variance of \( y \) conditional on \( X \) (which has been assumed to be equivalent to dependence on \( U \)). Define

\[
C = E \left[ \frac{1}{\sigma_0(U)} \left( g_0(U) \right)^2 (X - E[X|U])(X - E[X|U])^\top \right] \tag{29}
\]

This matrix is common in the single-index literature, specialized (using Assumption (3)) so that \( f_\epsilon \) is not in the definition. Define the associated matrix

\[
D(\tau) = E \left[ \frac{1}{\sigma_0(U)} \left( Q_0(U, \tau) \right)^2 (X - E[X|U])(X - E[X|U])^\top \right] \tag{30}
\]

which depends on \( \tau \) through \( Q_0 = g_0 + \sigma_0 F_\epsilon^{-1}(\tau) \). For the below results also define the closely related matrix

\[
D_0(\tau) = E \left[ \left( Q_0(U, \tau) \right)^2 (X - E[X|U])(X - E[X|U])^\top \right] \tag{31}
\]

which differs from \( D \) only in that it is lacking a term related to \( \sigma_0 \).

The next Lemma gives the Bahadur representation for the index estimate of a single-index quantile regression estimate when the location-scale structure is assumed to hold for the distribution of \( \epsilon \). This Lemma differs from Theorem 3 of Wu et al. (2008) because the location and scale functions must be accounted for here. Because of these functions the result looks somewhat more complex, but under simplifying restrictions their asymptotic distribution results are the same.

**Lemma 2.** Under the assumptions of the model and the \( A \) and \( B \) assumptions, the estimate \( \hat{\beta}(\tau) \) in the single-index quantile regression has Bahadur representation

\[
\sqrt{n} \left( \hat{\beta}(\tau) - \beta_0 \right) = \frac{1}{\hat{f}_\epsilon(\hat{F}_\epsilon^{-1}(\tau))} D^{-1}(\tau) \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_\tau \left( \epsilon_i - \hat{F}_\epsilon^{-1}(\tau) \right) Q_0(U_i, \tau) (X_i - E[X|U_i]) (1 + o_p(1)).
\]

The asymptotic distribution of \( \hat{\beta}(\tau) \) for the location-scale model is

\[
\sqrt{n} \left( \hat{\beta}(\tau) - \beta_0 \right) \sim N \left( 0, D^{-1}(\tau)D_0(\tau)D^{-1}(\tau) \right). \tag{32}
\]

This asymptotic distribution does not suffer from any bias term as the nonparametric estimator did. However, the dependence of \( D \) and \( D_0 \) on \( \tau \) (through \( Q_0 \)) leads to complicated variance formulas for the quantile averaging estimator when the data are heteroskedastic. Under homoskedasticity, the sandwich matrix in (33) reduces to \( \sigma^2 C^{-1} \). Note that, were all derivative estimates constrained to be equal, as they are for the composite quantile regression estimator, some of these complications are removed. In order to present the following result in a somewhat tractable way, define the matrix

\[
D_0(\tau_k, \tau_\ell) = E \left[ Q_0'(U, \tau_k)Q_0'(U, \tau_\ell)(X - E[X|U])(X - E[X|U])^\top \right] \tag{34}
\]

**Theorem 2.** Under the above assumptions, the asymptotic distribution of \( \hat{\beta}_{QAE} \), for any vector of weights \( \eta \), is

\[
\sqrt{n} \left( \hat{\beta}_{QAE} - \beta_0 \right) \sim N(0, V),
\]

where
where
\[
V = \sum_{k=1}^{q} \sum_{\ell=1}^{q} \frac{\varphi_k \varphi_\ell (\tau_k \wedge \tau_\ell - \tau_k \tau_\ell)}{f_\epsilon(f_\epsilon^{-1}(\tau_k))f_\epsilon(f_\epsilon^{-1}(\tau_\ell))} D^{-1}(\tau_k) D_0(\tau_k, \tau_\ell) D^{-1}(\tau_\ell).
\] (36)

Note that when \( \epsilon \) is homoskedastic, the above formula reduces to
\[
\sqrt{n} \left( \hat{\beta}_{QAE} - \beta_0 \right) \sim \mathcal{N} \left( 0, \eta^\top H \eta \sigma^2 C^{-1} \right). \] (37)

In this case then, it is very clear how weights should be chosen to minimize the asymptotic variance of the estimator. Because of the difficulty of expression (36), we suggest using the weight vector that was proposed for the symmetric case of the nonparametric proposal, that is, to let \( \eta^* = (1^\top H^{-1} 1) H^{-1} 1 \), where \( C \) was defined in (29).

Next we derive the asymptotic distribution of the composite quantile regression estimator. This estimator has a simpler covariance matrix to write down; simulations will tell whether it performs better than the quantile averaging estimator.

**Theorem 3.** Let \( \hat{\beta}_{CQR} \) be the composite quantile regression estimator. Then
\[
\sqrt{n} \left( \hat{\beta}_{CQR} - \beta_0 \right) \sim \mathcal{N} \left( 0, J_\eta C^{-1} C_0 C^{-1} \right)
\] (38)

where
\[
J_\eta = \frac{\sum_k \sum_\ell \eta_k \eta_\ell (\tau_k \wedge \tau_\ell - \tau_k \tau_\ell)}{\sum_k \sum_\ell \eta_k \eta_\ell f_\epsilon(f_\epsilon^{-1}(\tau_k))f_\epsilon(f_\epsilon^{-1}(\tau_\ell))}
\] (39)

and
\[
C_0 = E \left[ \left( g_0'(U) \right)^2 (X - E[X|U])(X - E[X|U])^\top \right]. \] (40)

Note that under homoskedasticity, the asymptotic distribution of this estimator is
\[
\sqrt{n} \left( \hat{\beta}_{CQR} - \beta_0 \right) \sim \mathcal{N} \left( 0, J_\varphi \sigma^2 C^{-1} \right),
\] (41)

which is similar to expression (37); in either case (of heteroskedasticity or homoskedasticity) it is possible to minimize variance of the estimator by adjusting \( \eta \) in the \( J_\eta \) term. Minimizing \( J_\eta \) leads to the following Corollary (as noted in Koenker (2005, §5.5) for the linear model).

**Corollary 2.** The optimal weight vector for the weighted composite quantile regression estimator \( \hat{\beta}_{WCQR} \) is
\[
\eta^* = \Gamma^{-1}_m q_m
\] (42)

where \( \Gamma_m \) is the \( m \times m \) matrix with \((k, \ell)\) entry \( \tau_k \wedge \tau_\ell - \tau_k \tau_\ell \). The resulting weighted estimator has asymptotic distribution
\[
\sqrt{n} \left( \hat{\beta}_{CQR} - \beta_0 \right) \sim \mathcal{N} \left( 0, \Lambda^{-1}_m C^{-1} C_0 C^{-1} \right)
\] (43)

However, we note that for weighted CQR estimation, one must also maintain the constraint that \( \eta_k \geq 0 \) for all \( k \), because otherwise the objective function is no longer convex. This slight inconvenience
is discussed briefly in Zhao and Xiao (2010) in the context of the parametric linear WCQR estimator, where it is noted that a sufficient condition for these constraints to be slack as \(n, m \to \infty\) is that \(f_\epsilon\) is log-concave.

### 4 Simulation experiment

The results of a small simulation experiment are presented below; in this preliminary draft, only limited results are shown.

Data was generated using the bottom of a quadratic function: specifically, three uniformly distributed regressors were generated, and the response was generated as

\[
Y = g(X_1 + X_2 + X_3) + 0.25 \epsilon, \tag{44}
\]

where \(\epsilon \sim \mathcal{N}(0, 1)\) and \(g(x) = (x - 3/2)^2\). The function estimates were evaluated using two measures: relative error, defined as

\[
RE(\hat{g}) = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\hat{g}(\hat{U}_i) - g_0(U_i)}{g_0(U_i)} \right| \tag{45}
\]

and average squared error, defined as

\[
ASE(\hat{g}) = \frac{1}{n} \sum_{i=1}^{n} (\hat{g}(\hat{U}_i) - g_0(U_i))^2 \tag{46}
\]

where the \(U_i\) come from the true (not estimated) values in the sample. We report nonparametric measures first: Table 1 shows these results from 290 simulation runs.

<table>
<thead>
<tr>
<th>RE (avg)</th>
<th>RE (s.e.)</th>
<th>ASE (avg)</th>
<th>ASE (s.e.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.7221</td>
<td>0.0304</td>
<td>0.7129</td>
</tr>
<tr>
<td>Median</td>
<td>0.7180</td>
<td>0.0371</td>
<td>0.7037</td>
</tr>
<tr>
<td>QA</td>
<td>0.7122</td>
<td>0.0337</td>
<td>0.6858</td>
</tr>
</tbody>
</table>

Table 1: Performance measures for the nonparametric parts of single-index estimators using mean-regression, median-regression and quantile averaging methods.

As can be seen, the quantile averaging regression performs quite well in terms of both measures.

Index coefficient estimates were evaluated using empirical bias, (simulated) standard error and root mean squared error, as well as an \(L_2\) estimate and a measure of the angle between the estimates and true coefficient values (which are all 1). Specifically the \(L_2\) measure is the standardized \(L_2\) norm of the coefficients,

\[
L_2 := \left\| \hat{\beta} - \beta_0 \right\| / \left\| \beta_0 \right\| \tag{47}
\]
so that smaller values of $L_2$ correspond to better estimates of $\beta_0$. The angle statistic is defined as

$$\text{angle} := \frac{\langle \hat{\beta}, \beta_0 \rangle}{\| \hat{\beta} \| \| \beta_0 \|}$$

so that larger values of $\text{angle}$ (closer to 1) are better.

On the other hand, the finite-dimensional parts are more difficult to work with. First Table 2 shows individual single-index quantile regression index coefficient estimates along with the weighted quantile averaging estimate.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>-0.3561</td>
<td>-0.1984</td>
<td>1.0360</td>
<td>0.7804</td>
<td>1.0954</td>
<td>0.8053</td>
<td>0.4724</td>
<td>0.8370</td>
<td>0.7125</td>
<td>0.6164</td>
</tr>
<tr>
<td>q1</td>
<td>-0.4100</td>
<td>-0.4787</td>
<td>1.1379</td>
<td>1.1637</td>
<td>1.2095</td>
<td>1.2583</td>
<td>0.7741</td>
<td>0.9605</td>
<td>0.5558</td>
<td>0.6942</td>
</tr>
<tr>
<td>q2</td>
<td>-0.3991</td>
<td>-0.3435</td>
<td>0.8941</td>
<td>0.9187</td>
<td>0.9791</td>
<td>0.9808</td>
<td>0.5821</td>
<td>0.7879</td>
<td>0.6049</td>
<td>0.6849</td>
</tr>
<tr>
<td>q3</td>
<td>-0.3339</td>
<td>-0.3167</td>
<td>0.9601</td>
<td>0.9756</td>
<td>1.0165</td>
<td>1.0257</td>
<td>0.5710</td>
<td>0.8461</td>
<td>0.6525</td>
<td>0.6640</td>
</tr>
<tr>
<td>q4</td>
<td>-0.3984</td>
<td>-0.3940</td>
<td>1.0057</td>
<td>0.9285</td>
<td>1.0818</td>
<td>1.0086</td>
<td>0.6205</td>
<td>0.8414</td>
<td>0.5510</td>
<td>0.7378</td>
</tr>
<tr>
<td>q5</td>
<td>-0.5890</td>
<td>-0.6140</td>
<td>1.1042</td>
<td>1.0672</td>
<td>1.2515</td>
<td>1.2312</td>
<td>0.8453</td>
<td>0.9084</td>
<td>0.3805</td>
<td>0.7771</td>
</tr>
</tbody>
</table>

Table 2: Performance measures for the index coefficients of individual conditional quantiles and the mean regression estimator.

As can be seen in Table 2, estimates of each conditional quantile may not perform as well as the mean-regression estimate; this is to be expected because the data are normally distributed, and the mean estimate should be optimal here. Finally we present simulation results for the weighted quantile averaging estimator, to see whether it might be possible to make up this lack of efficiency by combining conditional quantile estimates.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>mn</td>
<td>-0.3560</td>
<td>-0.1984</td>
<td>1.0360</td>
<td>0.7804</td>
<td>1.0954</td>
<td>0.8053</td>
<td>0.4724</td>
<td>0.8370</td>
<td>0.7125</td>
<td>0.6164</td>
</tr>
<tr>
<td>med</td>
<td>-0.3339</td>
<td>-0.3167</td>
<td>0.9601</td>
<td>0.9756</td>
<td>1.0165</td>
<td>1.0257</td>
<td>0.5710</td>
<td>0.8461</td>
<td>0.6525</td>
<td>0.6640</td>
</tr>
<tr>
<td>wqa</td>
<td>-0.4892</td>
<td>-0.4771</td>
<td>0.9417</td>
<td>0.9397</td>
<td>1.0612</td>
<td>1.0539</td>
<td>0.7553</td>
<td>0.7393</td>
<td>0.5960</td>
<td>0.5609</td>
</tr>
</tbody>
</table>

Table 3: Performance measures for the index estimates for mean-regression, median-regression and quantile averaging methods.

As can be seen in Table 3, the weighted quantile averaging regression estimates are close to the mean and median regression estimates, but they lag behind the other two in terms of bias. There are a few reasons for this. First of all, it may happen that an individual quantile estimate does not converge to its proper limit, pulling the average estimate away from the target. This is a feature that would be easy to correct when analyzing an individual sample, but is difficult to work with in a simulation experiment with many automatically-generated datasets of large size. Secondly, the weighting scheme may be too rough for the weighted averages; currently smoothed versions of the weighting function are being investigated, as well as non-crossing methods for the nonparametric part of the estimation strategy (which should affect the parametric part as well).
5 Empirical illustration

We used the Boston Housing data set as an illustration because it is well-known and has been used in this literature previously. The data consist of 506 observations in Boston that were collected from the 1970 US Census (each observation corresponds to one census tract). The model investigated is fairly simple:

\[ \text{medv}_i = g(\beta_{\text{room}} \text{room}_i + \beta_{\text{ltax}} \ln \text{ltax}_i + \beta_{\text{ptratio}} \text{ptratio}_i + \beta_{\text{llstat}} \ln \text{llstat}_i) + \epsilon_i \]  

(49)

where

1. medv is the median value of a home (in thousands of 1970 US dollars) in the census tract tract
2. room is the average number of rooms per house
3. ltax is the logarithm of the property tax rate
4. ptratio is the pupil-to-teacher ratio at schools in the tract
5. llstat is the percentage of the population in the tract that is considered lower status.

We illustrate the index from an estimated linear model against the response variable in Figure 1. It can be seen that there are several outliers, which are misclassified and/or topcoded values that could affect the estimation of the model (we note that these data have been corrected in Gilley and Pace (1996)). Furthermore, it can be seen that the data is scattered in a nonlinear pattern around the estimated index line: a plot of residuals against the estimated index is provided in the right panel reveals that the linear model suffers from some form of misspecification.

A single-index model might alleviate the misspecification apparent in the linear model. As an illustration of the quantile averaging method, Figure 3 (in the appendix) shows several individual single-index quantile regressions, and then one panel shows the averaging regression. The final plot in Figure 3 is reproduced in Figure 2, compared with the single-index model estimated using mean-regression methods.

The mean-regression estimator looks more smooth because it is the solution to a single optimization problem; the quantile averaging estimator could be smoothed if one desired a smoother appearance in plots. Although these function estimates look very similar when plotted on their respective index values, the coefficient estimates are different. These are reported in Table 4, as are the coefficient estimates from the linear model. The coefficients in the linear model are altered to be comparable to the others — the intercept estimate is omitted and \( \hat{\beta}_1 \) is set to 1. Therefore all coefficient estimates should be interpreted as relative to the effect of the average number of bedrooms per house on the
median house value in a census tract. Presumably as the average number of bedrooms increases, the median home value increases, so the signs of these coefficients can be read as they are. The negative values associated with the tax rate show that higher taxes are associated with lower home values, but the difference between the mean and quantile average shows that the relationship between the two is more complex, which may be expected — the fact that the mean estimate is closer to zero implies that in some tracts higher taxes must be associated with no median drop in the price or with a median price increase. Also one should keep in mind that the coefficient estimates from the linear models are only included as comparisons — because of the nonlinearity in the residuals, the linear models are technically misspecified. The difference between the coefficient estimates could be considered some rough indication of the degree to which the linear models are misspecified.

Although the single-index model is difficult to interpret as a structural model, it is good for prediction purposes. In Table 5 we include here also one table showing the sum of squared residuals and sum of absolute residuals from each of the four models used in Table 4. From Table 5 it can be seen that

\[ \hat{\beta}_{tax} \quad \hat{\beta}_{ptratio} \quad \hat{\beta}_{stat} \]

<table>
<thead>
<tr>
<th>Model</th>
<th>$\hat{\beta}_{tax}$</th>
<th>$\hat{\beta}_{ptratio}$</th>
<th>$\hat{\beta}_{stat}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear model, mean</td>
<td>-0.2825</td>
<td>-0.2148</td>
<td>-2.5537</td>
</tr>
<tr>
<td>Linear model, median</td>
<td>-0.6419</td>
<td>-0.1701</td>
<td>-1.2755</td>
</tr>
<tr>
<td>SI model, mean</td>
<td>-0.7140</td>
<td>-0.1900</td>
<td>-2.3333</td>
</tr>
<tr>
<td>SI model, quantile average</td>
<td>-0.9167</td>
<td>-0.1519</td>
<td>-1.4249</td>
</tr>
</tbody>
</table>

3Indeed, the data is available in the R packages MASS and m1bench.
Figure 2: The composite quantile regression estimator (left) formed by averaging the 9 [interior] conditional deciles. This is compared to a least-squares style single-index model. Although the functions may look similar in this plot, the coefficient estimates (reported in the text) are still somewhat different.

|                  | $\sum_i r_i^2$ | $\sum_i |r_i|$ |
|------------------|-----------------|-----------------|
| Linear model, mean | 11343.28        | 1717.855        |
| Linear model, median | 12401.88        | 1638.304        |
| SI model, mean    | 9717.17         | 1487.996        |
| SI model, quantile average | 9886.172 | 1418.803        |

Table 5: Measures of predictive power from the four different estimators shown in Table 4. Included are sums of squared or absolute residuals (labeled $\{r_i\}$ above) from estimation of the model(s).

the semiparametric models have smaller predictive errors, whether measured using squared residuals or absolute values.

6 Conclusion

Quantile averaging regression is a flexible semiparametric method that, in part, adapts to the form of the error distribution when the errors may be assumed to follow a location-scale model. A small empirical experiment reveals that the model performs in a manner quite similar to the more-familiar mean-regression style method of estimation; however, in theory the quantile averaging estimator can be tailored to be more robust to outliers or to be more efficient than the least-squares methodology under nonnormal error distributions.
A Mathematical proofs

Proof of Lemma 1. The proof is similar to the proof of Theorem 2 of Wu et al. (2008) and the proofs of Theorems 1 and 5 of Kai et al. (2010). We outline the steps, highlighting the differences between a straightforward application of Kai et al. (2010). The proof of Theorem 5 of Kai et al. (2010) is used twice; once it is used conditioning on \( U = X^\top \beta_0 \), and once it is used conditioning on \( \hat{U} = X^\top \hat{\beta} \).

Here we calculate the asymptotic distribution when conditioning on \( \hat{U} \). First make some definitions: let \( u \) be some value in the interior of the domain of \( U \). Let \( Q_0(u, \tau) := g_0(u) + \sigma_0(u)F^{-1}_\mu(\tau) \) and because \( \tau \) remains fixed, denote the first derivative of \( Q_0 \) with respect to \( u \) as \( Q_0'(u, \tau) := g_0'(u) + \sigma_0'(u)F^{-1}_\mu(\tau) \) (and analogously for the second derivative). Define

\[
\hat{\theta} = \left[ \frac{\sqrt{n\tilde{h}}(\tilde{a} - Q_0(u, \tau))}{h\sqrt{n\tilde{h}}(\tilde{b} - Q_0'(u, \tau))} \right]
\]

\[
\hat{Z}_i = \left[ \frac{1}{\hat{U}_i - u} \right], \quad \hat{K}_i = K \left( \frac{\hat{U}_i - u}{h} \right), \quad \hat{\Delta}_i = \frac{\hat{\theta}^\top \hat{Z}_i}{\sqrt{n\tilde{h}}}
\]

\[
\hat{d}_i = Q_0(\hat{U}_i, \tau) - Q_0(u, \tau) - (\hat{U}_i - u)Q_0'(u, \tau)
\]

\[
\hat{W}_n = -\frac{1}{\sqrt{n\tilde{h}}} \sum_{i=1}^n \hat{Z}_i \psi_\tau \left( \hat{\epsilon}_i - F^{-1}_\mu(\tau) + \frac{\hat{d}_i}{\sigma_0(\hat{U}_i)} \right) \hat{K}_i,
\]

where \( \psi_\tau(u) = \tau - I(u \leq 0) \).

First, it can be verified that minimizing

\[
\sum_{i=1}^n \rho_\tau \left( y_i - \tilde{a} - \tilde{b}(\hat{U}_i - u) \right) \hat{K}_i
\]

with respect to \( \tilde{a} \) and \( \tilde{b} \) is equivalent to minimizing the objective function

\[
\sum_{i=1}^n \left( \rho_\tau \left( \left( \hat{\epsilon}_i - F^{-1}_\mu(\tau) \right) \sigma_0(\hat{U}_i) + \hat{d}_i - \hat{\Delta}_i \right) - \rho_\tau \left( \left( \hat{\epsilon}_i - F^{-1}_\mu(\tau) \right) \sigma_0(\hat{U}_i) + \hat{d}_i \right) \right) \hat{K}_i
\]

with respect to \( \hat{\theta} \) (via \( \{\hat{\Delta}_i\}_i \)). Knight’s identity implies that minimizing this latter objective function is in turn equivalent to minimizing

\[
L_n(\hat{\theta}) = -\hat{\theta}^\top \hat{W}_n + B_n(\hat{\theta}),
\]

with respect to \( \hat{\theta} \), where

\[
B_n(\hat{\theta}) = \sum_{i=1}^n \hat{K}_i \int_0^{\hat{\Delta}_i} \left( I \left( \hat{\epsilon}_i \leq F^{-1}_\mu(\tau) - \frac{\hat{d}_i}{\sigma_0(\hat{U}_i)} + \frac{\hat{\Delta}_i}{\sigma_0(\hat{U}_i)} \right) - I \left( \hat{\epsilon}_i \leq F^{-1}_\mu(\tau) - \frac{\hat{d}_i}{\sigma_0(\hat{U}_i)} \right) \right) \, dz.
\]
Let
\[ S = \frac{f_\epsilon(F_\epsilon^{-1}(\tau))f_U(u)}{\sigma_0(u)} \begin{bmatrix} 1 & 0 \\ 0 & \mu_2 \end{bmatrix}. \] (59)

Next we show that \( B_n(\hat{\theta}) = \frac{1}{2} \hat{\theta}^\top S \hat{\theta} + o_p(1) \) uniformly in \( \hat{\theta} \). The expected value of \( B_n \) is

\[
E[B_n(\hat{\theta})|\hat{U}] = \sum_{i=1}^n \tilde{K}_i \int_0^{\Delta_i} F_\epsilon\left(F_\epsilon^{-1}(\tau) - \frac{\tilde{d}_i}{\sigma_0(\hat{U}_i)} + \frac{z}{\sigma_0(\hat{U}_i)}\right) - F_\epsilon\left(F_\epsilon^{-1}(\tau) - \frac{\tilde{d}_i}{\sigma_0(\hat{U}_i)}\right) d\tau \] (60)

\[
= \sum_{i=1}^n \frac{\tilde{K}_i \tilde{\Delta}_i^2}{2\sigma_0(\hat{U}_i)} f_\epsilon(F_\epsilon^{-1}(\tau)) + o_p(1) \] (61)

\[
= \frac{1}{2} \frac{f_\epsilon(F_\epsilon^{-1}(\tau))f_U(u)}{\sigma_0(u)} \begin{bmatrix} 1 & 0 \\ 0 & \mu_2 \end{bmatrix} \hat{\theta} + o_p(1) \] (64)

\[
= \frac{1}{2} \hat{\theta}^\top S \hat{\theta} + o_p(1). \] (65)

It can be similarly shown that \( \text{Var}(B_n(\hat{\theta})|\hat{U}) = o_p(1) \). The above steps rely on the differentiability of \( g_0, \sigma_0 \) and \( F_\epsilon \), the consistency of \( \hat{\beta} \), and that \( K \) is a second-order kernel function. Then the convexity lemma of Pollard (1991) implies that

\[
\hat{\theta} = \text{argmin}_\theta L_n(\theta) = \frac{1}{2} \theta^\top S \theta - \theta^\top \hat{W}_n + o_p(1) \] (66)

asymptotically has the Bahadur representation

\[
\hat{\theta} = S^{-1} \hat{W}_n + o_p(1) \] (67)

uniformly on any compact set in \( \Theta \).

These steps can be repeated, conditioning on \( U = X^\top \beta_0 \), and the resulting estimator is denoted
with a bar below. The result is that, defining

\[
\hat{\theta} = \left[ \frac{\sqrt{n}h(\bar{a} - Q_0(u, \tau))}{h\sqrt{n}h(b - Q_0(u, \tau))} \right] \tag{68}
\]

\[
Z_i = \left[ \frac{1}{U_i - u} \right], \quad K_i = K \left( \frac{U_i - u}{h} \right) \tag{69}
\]

\[
d_i = Q_0(U_i, \tau) - Q_0(u, \tau) - (U_i - u)Q'_0(u, \tau) \tag{70}
\]

\[
\hat{W}_n = -\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} Z_i \psi_{\tau} \left( \epsilon_i - F_{\tau}^{-1}(\tau) + \frac{d_i}{\sigma_0(U_i)} \right) K_i, \tag{71}
\]

there is an analogous Bahadur representation,

\[
\hat{\theta} = S^{-1}\hat{W}_n + o_P(1). \tag{72}
\]

Kai et al. (2010, Theorem 5) show that that this converges to a normal limit. Below we calculate the expectation and variance of \( \hat{\theta} \). It can be verified that

\[
E \left[ S^{-1}\hat{W}_n | U \right] = S^{-1} \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} E \left[ Z_i \psi_{\tau} \left( \epsilon_i - F_{\tau}^{-1}(\tau) + \frac{d_i}{\sigma_0(U_i)} \right) K_i | U \right] \tag{73}
\]

\[
= S^{-1} \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} E \left[ f_\epsilon(F_{\tau}^{-1}(\tau)) \left( \frac{U_i - u}{2\sigma_0(U_i)} Q''_0(u, \tau)Z_i | U \right) \right] (1 + o_P(1)) \tag{74}
\]

\[
= S^{-1} f_\epsilon(F_{\tau}^{-1}(\tau))Q''_0(u, \tau)f_U(u) \frac{1}{2\sigma_0(u)} \begin{bmatrix} \mu_2 \\ \mu_3 \end{bmatrix} (1 + o(1)) \tag{75}
\]

now utilizing the definition of \( S \),

\[
= \frac{\sigma_0(u)}{f_\epsilon(F_{\tau}^{-1}(\tau))f_U(u)} \begin{bmatrix} 1 & 0 \\ 0 & 1/\mu_2 \end{bmatrix} \frac{f_\epsilon(F_{\tau}^{-1}(\tau))Q''_0(u, \tau)f_U(u)}{2\sigma_0(u)} \frac{1}{h^2} \begin{bmatrix} \mu_2 \\ \mu_3/\mu_2 \end{bmatrix} (1 + o(1)) \tag{76}
\]

\[
= \frac{1}{2} Q''_0(u, \tau)h^2 \begin{bmatrix} \mu_2 \\ \mu_3/\mu_2 \end{bmatrix} (1 + o(1)) \tag{77}
\]

The variance of \( \hat{\theta} \) can be shown equal to (following Kai et al. (2010, p.65))

\[
\text{Var} \left( S^{-1} \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \psi_{\tau} \left( \epsilon_i - F_{\tau}^{-1}(\tau) \right) Z_i K_i | U \right) = \frac{\sigma}{\tau} - \tau f_U(u)S^{-1} \begin{bmatrix} v_0 \\ v_1 \end{bmatrix} S^{-1}(1 + o(1)) \tag{78}
\]

\[
: = \frac{\tau}{\tau} - \tau S^{-1}SS^{-1}(1 + o(1)). \tag{79}
\]
In other words,

$$\text{Var} \left( S^{-1} \hat{W}_n | U \right) = \frac{\tau(1-\tau)\sigma^2_0(u)}{f_0(u)f_0'(F^{-1}_0(\tau))} \begin{bmatrix} v_0 & v_{1/\mu_2} \\ v_{1/\mu_2} & v_{2/\mu_2} \end{bmatrix} (1 + o(1)).$$  \hfill (80)$$

The asymptotic distribution of $\hat{\theta}$ and $\bar{\theta}$ is the same: note that

$$\hat{\theta} - \bar{\theta} = S^{-1} \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \left( \psi_{\tau} \left( y_i - \bar{\hat{a}}(U_i - u) \right) \hat{Z}_i \hat{\kappa}_i - \psi_{\tau} \left( y_i - \bar{\hat{a}}(U_i - u) \right) \hat{Z}_i \kappa_i \right)$$  \hfill (81)$$

$$= S^{-1} \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \psi_{\tau} \left( y_i - \bar{\hat{a}}(U_i - u) \right) (\hat{Z}_i \kappa_i - Z_i \kappa_i)$$  \hfill (82)$$

because the consistency of $\hat{\beta}$ implies $\psi_{\tau} \left( y_i - \bar{\hat{a}}(U_i - u) \right) = \psi_{\tau} \left( y_i - \bar{\hat{a}}(U_i - u) \right)$ almost surely as $n \to \infty$. Then for some $C > 0$,

$$E \left[ (\hat{\theta} - \bar{\theta}) (\hat{\theta} - \bar{\theta})^\top \right] < CS^{-1} \frac{1}{h} E \left[ \psi_{\tau}^2 \left( y_i - \bar{\hat{a}}(U_i - u) \right) (\hat{Z}_i \kappa_i - Z_i \kappa_i) (\hat{Z}_i \kappa_i - Z_i \kappa_i)^\top \right] S^{-1} = O(o(1)) = o(1).$$  \hfill (84)$$

This implies (taking the first element of the vector) that

$$\sqrt{nh} \left( Q_0(u, \tau) - Q_0(u, \tau) - \frac{1}{2} Q_0''(u, \tau) h^2 \mu_2 \right) \sim \mathcal{N} \left( 0, \frac{\tau(1-\tau)\sigma^2_0(u)\nu_0}{f_0(u)f_0'(F^{-1}_0(\tau))} \right).$$  \hfill (85)$$

Showing that $u$ may be replaced with $x^\top \hat{\beta}$ is accomplished as follows:

$$\sqrt{nh} \left( \hat{Q}(x^\top \hat{\beta}, \tau) - Q_0(x^\top \beta_0, \tau) \right)$$  \hfill (86)$$

$$= \sqrt{nh} \left( \hat{Q}(x^\top \hat{\beta}, \tau) - \hat{Q}(x^\top \hat{\beta}_0, \tau) + \hat{Q}(x^\top \hat{\beta}_0, \tau) - Q_0(x^\top \beta_0, \tau) \right)$$  \hfill (87)$$

$$= \sqrt{nh} \hat{Q}'(x^\top \beta_0, \tau) X^\top (\beta - \beta_0)(1 + o(1)) + \sqrt{nh} \left( \hat{Q}(x^\top \beta_0, \tau) - Q_0(x^\top \beta_0, \tau) \right)$$  \hfill (88)$$

and because $\sqrt{nh}(\hat{\beta} - \beta_0) = O_p(1)$ and $h = o_p(1)$,

$$= \sqrt{nh} \left( Q(x^\top \beta_0, \tau) - Q_0(x^\top \beta_0, \tau) \right) + o_p(1).$$  \hfill (89)$$

\textbf{Proof of Theorem 1.} Lemma 1 implies that

$$\hat{Q}(x^\top \hat{\beta}, \tau) - Q_0(x^\top \beta_0, \tau_k) = \frac{1}{2} Q_0''(x^\top \beta_0, \tau_k) \mu_2 h^2 (1 + o_p(1))$$

$$= \frac{\sigma_0(x^\top \beta_0)}{f_U(x^\top \beta_0 f_\epsilon(F_\epsilon^{-1}(\tau_k)))} \frac{1}{nh} \sum_{i=1}^{n} \psi_{\tau_k} \left( e_i - F_\epsilon^{-1}(\tau_k) \right) K_i + o_p \left( 1/\sqrt{nh} \right).$$  \hfill (90)$$
Recall that $Q_0(u, \tau) := g_0(u) + \sigma_0(u)F^{-1}_e(\tau)$. Because the weights are constrained to sum to one, when combining the conditional quantile estimates one finds

\[
\hat{g}_{QAE}(x^\top \hat{\beta}) - g_0(x^\top \beta_0) - \sigma_0(x^\top \beta_0) \sum_{k=1}^q \varphi_k F_0^{-1}(\tau_k) - \frac{\mu_2 h^2}{2} \sum_{k=1}^q \varphi_k Q_0'(x^\top \beta_0, \tau_k)(1 + o_p(1))
\]

\[
= \frac{\sigma_0(x^\top \beta_0)}{f_0(x^\top \beta_0)} \frac{1}{nh} \sum_{i=1}^n \sum_{k=1}^q \frac{\varphi_k}{f_\delta(F_0^{-1}(\tau_k))} \psi_{\tau_k} (e_i - F^{-1}_e(\tau_k)) K_i + o_p\left(\frac{1}{\sqrt{nh}}\right). \tag{91}
\]

Calculating the variance implies the result.

**Proof of Corollary 1.** The optimal weights in the symmetric case can be solved using Lagrange multipliers. The optimal weights solve

\[
\min_{\varphi} \varphi^\top H \varphi \quad \text{s.t.} \quad \varphi^\top 1 = 1 \tag{92}
\]

and with the additional constraint that $\varphi^\top q_m = 0$ if no symmetry assumption is made. These can both be solved in the same way (there are either one or two linear constraints of $\varphi$), resulting in

\[
\varphi^*_s = (1^\top H^{-1} 1)^{-1} H^{-1} 1 \tag{93}
\]

in the symmetric case, and

\[
\varphi^* = \frac{(q_m^\top H^{-1} q_m) \cdot H^{-1} 1 - (1^\top H^{-1} q_m) \cdot H^{-1} q_m}{(1^\top H^{-1} 1) \cdot (q_m^\top H^{-1} q_m) - (1^\top H^{-1} q_m)^2} \tag{94}
\]

in the general case. To find the asymptotic variances for the second result these optimal weights can be inserted into the variance given in Theorem 1.

**Proof of Lemma 2.** Taking estimates of the nonparametric part $\{\hat{a}_j, \hat{b}_j\}$ as given, $\hat{\beta} (\tau)$ is the minimizer of

\[
\Xi(\beta) = \sum_{j=1}^n \sum_{i=1}^n \rho_{\tau_k} \left( y_i - \hat{a}_j - \hat{b}_j (X_i - X_j)^\top \beta \right) \omega_{ij}. \tag{95}
\]

Using the notation $X_{ij} = X_i - X_j$, $\hat{\beta} = \sqrt{n}(\hat{\beta} - \beta_0)$ (suppressing dependence on $\tau$) and $y_{ij} = y_i - \hat{a}_j - \hat{b}_j X_{ij}^\top \beta_0$, minimization of the above objective function is equivalent to minimizing the related objective function

\[
\Xi'(\hat{\beta}) = \sum_{j=1}^n \sum_{i=1}^n \left( \rho_{\tau_k} \left( y_{ij} - \frac{1}{\sqrt{n}} \hat{b}_j X_{ij}^\top \hat{\beta} \right) - \rho_{\tau_k} (y_{ij}) \right) \omega_{ij}. \tag{96}
\]
Once again using Knight’s identity, this is equal to

\[-\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sum_{i=1}^{n} b_j X_{ij}^\top \beta \psi_{\tau_i}(y_{ij}) \omega_{ij} + \sum_{j=1}^{n} \sum_{i=1}^{n} \omega_{ij} \int_{0}^{\frac{1}{\sqrt{n}} b_j X_{ij}^\top \hat{\beta}} I(y_{ij} \leq z) - I(y_{ij} \leq 0) \, dz. \quad (97)\]

Lemma 3 shows that the first term above converges to a normal random variable with mean zero and covariance matrix \(D_0(\tau)\). Lemma 4 shows that the second term converges in probability to the matrix \(\frac{1}{2} f_\epsilon(F^{-1}_\epsilon(\tau)) \hat{\beta}^\top D(\tau) \hat{\beta}\). Then the convexity lemma and Slutsky’s theorem imply the results (the Bahadur representation and asymptotic normality respectively).

**Lemma 3.** Under the conditions of Lemma 2,

\[-\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sum_{i=1}^{n} b_j X_{ij} \psi_{\tau_i}(y_{ij}) \omega_{ij} \sim \mathcal{N}(0, \tau(1 - \tau) D_0(\tau)) \quad (98)\]

**Proof of Lemma 3.** First rearrange the sums and use the approximation \(\omega_{ij} = (nhf(U_j))^{-1} K\left(\frac{U_j - U}{h}\right) (1 + o_p(1))\) (cf. Fan et al. (2013, p. 15)).

\[-\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \sum_{i=1}^{n} b_j X_{ij} \psi_{\tau_i}(y_{ij}) \omega_{ij} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{b}_j X_{ij} \psi_{\tau_i}(y_{ij}) \frac{K\left(\frac{U_i - U_j}{h}\right)}{nhf(U_j)} (1 + o_p(1)). \quad (99)\]

Then because

\[y_{ij} = y_i - a_j - \hat{b}_j(U_i - U_j) = g_0(U_i) + \sigma_0(U_i) \epsilon_i - \hat{a}_j - \hat{b}_j(U_i - U_j) = (\epsilon_i - F^{-1}_\epsilon(\tau)) \sigma_0(U_i) + o_p(1), \quad (100)\]

the above sum can be rewritten

\[-\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\tau_i} (\epsilon_i - F^{-1}_\epsilon(\tau)) \sum_{j=1}^{n} \hat{b}_j X_{ij} \frac{K\left(\frac{U_i - U_j}{h}\right)}{nhf(U_j)} (1 + o_p(1)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\tau_i} (\epsilon_i - F^{-1}_\epsilon(\tau)) Q_0(U_i, \tau) \left(X_i - E[X|U_j]\right) (1 + o_p(1)). \quad (102)\]

This converges to a normal random variable with mean zero and covariance matrix as in the statement of the Lemma.

**Lemma 4.** Under the conditions of Lemma 2,

\[\sum_{j=1}^{n} \sum_{i=1}^{n} \omega_{ij} \int_{0}^{\frac{1}{\sqrt{n}} b_j X_{ij}^\top \hat{\beta}} I(y_{ij} \leq z) - I(y_{ij} \leq 0) \, dz = \frac{1}{2} f_\epsilon(F^{-1}_\epsilon(\tau)) \hat{\beta}^\top D(\tau) \hat{\beta} (1 + o_p(1)) \quad (103)\]
Proof of Lemma 4. Note that
\[ y_{ij} = y_i - a_j - b_j(U_i - U_j) = g_0(U_i) + \sigma_0(U_i)\epsilon_i - a_j - b_j(U_i - U_j) = (\epsilon_i - F^{-1}_\epsilon(\tau))\sigma_0(U_i) + o_p(1). \tag{104} \]

Therefore
\[ \int_0^{1/nh} I(y_{ij} \leq z) - I(y_{ij} \leq 0)dz = \int_0^{1/nh} I(\epsilon_i \leq F^{-1}_\epsilon(\tau) + z/\sigma_0(U_i)) - I(\epsilon_i \leq F^{-1}_\epsilon(\tau)) dz(1 + o_p(1)). \tag{106} \]

To find the limit in probability of this term, first take expectations of \( y \) conditional on \( U \). Then
\[ \int_0^{1/nh} b_j X_{ij}^\top \hat{\beta} \int_0^{1/nh} F_\epsilon \left( F^{-1}_\epsilon(\tau) + z/\sigma_0(U_i) \right) - F_\epsilon \left( F^{-1}_\epsilon(\tau) \right) dz = \frac{1}{2nh} \int_0^{1/nh} f_\epsilon \left( F^{-1}_\epsilon(\tau) \right) \frac{1}{\sigma_0(U_i)} \hat{\beta}^\top X_{ij} X_{ij}^\top \beta K \left( \frac{U_i - U_j}{h} \right) (1 + o_p(1)). \tag{108} \]

Put this back into the double summation, rewriting \( \omega_{ij} = (nhf_\epsilon(U_j))^{-1}K \left( \frac{U_i - U_j}{h} \right) (1 + o_p(1)). \) This is
\[ \frac{1}{2n^2 h} \sum_{j=1}^n b_j^2 f_\epsilon \left( F^{-1}_\epsilon(\tau) \right) \sum_{i=1}^n \frac{1}{\sigma_0(U_i)} \hat{\beta}^\top X_{ij} X_{ij}^\top \beta K \left( \frac{U_i - U_j}{h} \right) (1 + o_p(1)). \tag{109} \]

Consider the inside sum: fixing \( j \),
\[ \frac{1}{nh} \sum_{i=1}^n \frac{1}{\sigma_0(U_i)} X_{ij} X_{ij}^\top K \left( \frac{U_i - U_j}{h} \right) = E \left[ \frac{1}{nh} \sum_{i=1}^n \frac{1}{\sigma_0(U_i)} X_{ij} X_{ij}^\top K \left( \frac{U_i - U_j}{h} \right) \right] (1 + o_p(1)) \tag{110} \]
\[ = \frac{1}{\sigma_0(U_j)} \left( E \left[ X|U_j \right] - X_j \right) \left( E \left[ X|U_j \right] - X_j \right)^\top f_\epsilon(U_j)(1 + o_p(1)). \tag{111} \]

Therefore (109) is equal to
\[ \frac{1}{2n} \sum_{j=1}^n f_\epsilon \left( F^{-1}_\epsilon(\tau) \right) b_j^2 \hat{\beta}^\top E \left[ \frac{1}{\sigma_0(U_j)} \right] \left( Q_0'(U_j, \tau) \right)^2 (X - E[X|U]) (X - E[X|U])^\top \beta (1 + o_p(1)) \]
\[ = \frac{1}{2} f_\epsilon \left( F^{-1}_\epsilon(\tau) \right) \hat{\beta}^\top E \left[ \frac{1}{\sigma_0(U)} \left( Q_0'(U, \tau) \right)^2 (X - E[X|U]) (X - E[X|U])^\top \right] \beta (1 + o_p(1)). \tag{112} \]

To show that this converges in probability it is sufficient to show that the variance of this quantity...
converges to zero. Similarly to Kai et al. (2010, p. 64-65), we note

\[
\sum_{j=1}^{n} \sum_{i=1}^{n} \text{Var} \left( \omega_{ij} \int_{0}^{\frac{1}{\sqrt{n}} \hat{b}_{ij} X_{ij}^{\top} \hat{\beta}} I \left( \epsilon_{i} \leq F_{\epsilon}^{-1}(\tau) + \frac{z}{\sigma_{0}(U_{i})} \right) - I \left( \epsilon_{i} \leq F_{\epsilon}^{-1}(\tau) \right) \, dz \right)
\]

(113)

\[
\leq \sum_{j=1}^{n} \sum_{i=1}^{n} \mathbb{E} \left[ \left( \omega_{ij} \int_{0}^{\frac{1}{\sqrt{n}} \hat{b}_{ij} X_{ij}^{\top} \hat{\beta}} I \left( \epsilon_{i} \leq F_{\epsilon}^{-1}(\tau) + \frac{z}{\sigma_{0}(U_{i})} \right) - I \left( \epsilon_{i} \leq F_{\epsilon}^{-1}(\tau) \right) \, dz \right)^{2} \right]
\]

(114)

\[
\leq \sum_{j=1}^{n} \sum_{i=1}^{n} \omega_{ij}^{2} \int_{0}^{\frac{1}{\sqrt{n}} \hat{b}_{ij} X_{ij}^{\top} \hat{\beta}} \int_{0}^{\frac{1}{\sqrt{n}} \hat{b}_{ij} X_{ij}^{\top} \hat{\beta}} F_{\epsilon} \left( F_{\epsilon}^{-1}(\tau) + \frac{1}{\sqrt{n}} \left| \hat{b}_{ij} X_{ij}^{\top} \hat{\beta} \right| \right) - F_{\epsilon} \left( F_{\epsilon}^{-1}(\tau) \right) \, dz_{1} \, dz_{2}
\]

(115)

\[
= o \left( \sum_{j=1}^{n} \sum_{i=1}^{n} \omega_{ij}^{2} \frac{1}{n} \left( \hat{b}_{ij} X_{ij}^{\top} \hat{\beta} \right)^{2} \right) = o_{p}(1).
\]

(116)

\[\text{Proof of Theorem 2.}\] First define

\[
B_{n}(\tau) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{\tau} \left( \epsilon_{i} - F_{\epsilon}^{-1}(\tau) \right) Q_{0}^\prime(U_{i}, \tau) \left( X_{i} - \mathbb{E}\left[ X|U_{i} \right] \right).
\]

(117)

Using Lemma 2 and the above definition it can be verified that \( \hat{\beta} \) has the following Bahadur representation:

\[
\sqrt{n} \left( \hat{\beta} - \beta_{0} \right) = \sum_{k=1}^{q} \frac{\eta_{k}}{f_{\epsilon}(F_{\epsilon}^{-1}(\tau_{k}))} D_{\epsilon}^{-1}(\tau_{k}) B_{n}(\tau_{k})(1 + o_{p}(1)).
\]

(118)

This has mean zero and, noting that

\[
\text{Cov} \left( B_{n}(\tau_{k}), B_{n}(\tau_{l}) \right) = \left( \tau_{k} \wedge \tau_{l} - \tau_{k} \wedge \tau_{l} \right) D_{0}(\tau_{k}, \tau_{l})
\]

(119)

it can be verified that the asymptotic variance is as in the assertion of the Theorem. \[\square\]

\[\text{Proof of Theorem 3.}\] The proof of this Theorem is similar to the proof of Theorem 2, so only the details are sketched. Given \( \{a_{jk}, b_{j}\}_{j,k} \), the composite quantile regression estimator is

\[
\hat{\beta}_{CQR} = \arg\min_{\beta} \sum_{k=1}^{m} \sum_{j=1}^{n} \sum_{i=1}^{n} \eta_{k} \rho_{\tau_{k}} \left( y_{ijk} - \hat{a}_{jk} - \hat{b}_{ij} X_{ij}^{\top} \hat{\beta} \right) \omega_{ij}.
\]

(120)

This is equivalent to minimizing (after making the definitions \( y_{ijk} = y - \hat{a}_{jk} - \hat{b}_{ij} X_{ij}^{\top} \beta_{0} \) and \( \hat{\beta} = \sqrt{n}(\hat{\beta} - \beta_{0}) \)) the objective function

\[
L(\hat{\beta}) = \sum_{k=1}^{m} \sum_{j=1}^{n} \sum_{i=1}^{n} \eta_{k} \omega_{ij} \left( \rho_{\tau_{k}} \left( y_{ijk} - \frac{1}{\sqrt{n}} \hat{b}_{ij} X_{ij}^{\top} \hat{\beta} \right) - \rho_{\tau_{k}}(y_{ijk}) \right)
\]

(121)

\[24\]
with respect to $\beta$. Knight’s identity implies that

\[
L(\beta) = -\frac{1}{\sqrt{n}} \sum_{k=1}^{m} \sum_{j=1}^{n} \sum_{i=1}^{n} \eta_k \hat{b}_j X_{ij}^\top \hat{B} \psi_{\tau_k}(y_{ijk}) \omega_{ij} + \sum_{k=1}^{m} \sum_{j=1}^{n} \sum_{i=1}^{n} \eta_k \omega_{ij} \int_{0}^{1} \frac{\hat{b}_j X_{ij}^\top \hat{B}}{\hat{b}_j X_{ij}^\top \hat{B}} I(y_{ijk} \leq z) - I(y_{ijk} \leq 0) dz,
\]

which we write as \( L(\beta) = L_1(\beta) + L_2(\beta) \). Lemmas 5 and 6 show that \( L_1 \) and \( L_2 \) converge to a random vector and a matrix respectively. More specifically,

\[
L(\beta) = L_1(\beta) + L_2(\beta) = W_n \beta + \frac{1}{2} \sum_{k=1}^{m} \eta_k f_r(F_r^{-1}(\tau_k)) \beta^\top C \beta + o_p(1)
\]

and the convexity lemma of Pollard can be used to show that uniformly in $\beta$, the minimizer of the above quadratic equation behaves like a normal random variable with mean and variance $V$ defined in the statement of the Theorem.

**Lemma 5.** Under the conditions of Theorem 3,

\[
-\frac{1}{\sqrt{n}} \sum_{k=1}^{m} \sum_{i=1}^{n} \sum_{j=1}^{n} \hat{b}_j \psi_{\tau_k}(y_{ijk}) X_{ij}^\top \omega_{ij} \sim \mathcal{N}(0, \Xi),
\]

where

\[
\Xi = \sum_{k=1}^{m} \sum_{i=1}^{n} \eta_k \eta_i (\tau_k \land \tau_i - \tau_k \tau_i) E \left[ (g'(U))^2 (X - E[X|U]) (X - E[X|U])^\top \right]
\]

**Proof of Lemma 5.** The proof is similar in its details to the proof of Lemma 3, so the general ideas are sketched. It can be verified that

\[
-\frac{1}{\sqrt{n}} \sum_{k=1}^{m} \eta_k \sum_{j=1}^{n} \sum_{i=1}^{n} \hat{b}_j X_{ij} \psi_{\tau_k}(y_{ijk}) \omega_{ij} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{nh} \sum_{j=1}^{n} \hat{b}_j X_{ij} \frac{K(U_i - U_j)}{f(U)} \sum_{k=1}^{m} \eta_k \psi_{\tau_k}(\epsilon_i - F_r^{-1}(\tau_k))(1 + o_p(1))
\]

\[
= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} g'(U_i) (X_i - E[X|U_i]) \sum_{k=1}^{m} \eta_k \psi_{\tau_k}(\epsilon_i - F_r^{-1}(\tau_k))(1 + o_p(1)).
\]

This converges to a mean-zero normal random vector with covariance matrix as described in the statement of the Lemma. Note that for the composite quantile regression estimator, the $\hat{b}_j$ terms estimate $g'(U_j)$ (not $Q_0'(U_j, \tau)$, as is the case for quantile averaging).
Lemma 6. Under the conditions of Theorem 3,

\[
\sum_{k=1}^{m} \eta_k \sum_{j=1}^{n} \sum_{i=1}^{n} \omega_{ij} \int_{0}^{1} \rho_{\omega_{ij}} b_{ij} X_{ij} \hat{\beta} d\hat{z}
\]

\[I(y_{ijk} \leq z) - I(y_{ijk} \leq 0)dz\]

\[= \frac{1}{2} \tilde{\beta}^T C \tilde{\beta} \sum_{k=1}^{m} \eta_k f_{\epsilon}(F_{\epsilon}^{-1}(\tau_k))(1 + o_P(1)) \quad (130)\]

Proof of Lemma 6. This Lemma could be considered a corollary of Lemma 4. Indeed, its proof is nearly the same as that of Lemma 4 — the only differences are the definition of \(y_{ijk}\) and the corresponding weighted sum of \(f_{\epsilon}(F_{\epsilon}^{-1}(\tau_k))\) terms, and that \(\hat{b}_j\) has a different estimand, resulting in a different matrix. \(\blacksquare\)

B Estimation algorithms

B.1 Weighted composite quantile regression

The following basic algorithm was used in estimation of the basic CQR estimator. This is nearly the same algorithm as was proposed in Fan et al. (2013). We assume that a vector of weights, \(\varphi\) is given beforehand.

1. Start with an initial estimate of \(\beta_0\), for example a direct median estimate following the proposal of Chaudhuri et al. (1997).

2. Given estimate \(\hat{\beta}\) (recall that \(X_j^T \hat{\beta} := \hat{U}_j\)), estimate the function \(g_0\) and its derivative \(g_0'\) by running \(n\) composite quantile regressions for \(j = 1, \ldots n\):

\[
\{\hat{g}(\hat{U}_j), \hat{g}'(\hat{U}_j)\}_{j=1}^{n} = \arg\min_{\{a_{jk}, b_j\}} \sum_{j=1}^{n} \sum_{i=1}^{n} \rho_w \left( y_i - a_{jk} - b_j (U_i - \hat{U}_j) \right) \omega_{ij} \quad (131)
\]

where \(\omega_{ij} = K \left( \hat{U}_i - \hat{U}_j / h \right) / \sum_{t=1}^{n} K \left( \hat{U}_i - \hat{U}_j / h \right)\).

3. Given estimates \(\{a_{jk}, \hat{b}\}\) and \(\{\omega_{ij}\}_{ij}\) used in the previous step, estimate \(\beta_0\) in the composite quantile regression

\[
\hat{\beta} = \arg\min_{\beta} \sum_{j=1}^{n} \sum_{i=1}^{n} \rho_w \left( y_i - a_j - \hat{b}(X_i - X_j)^T \beta \right) \omega_{ij} \quad (132)
\]

4. Repeat step 2 and 3 to convergence.

5. The final estimate of \(g_0\) is obtained as

\[
\hat{g}(\hat{U}_j) = \sum_{k=1}^{m} \varphi_k a_{jk}, \quad (133)
\]
interpolating as necessary to make predictions away from the set \{U_j\}_j.

To estimate the \textit{weighted} composite quantile regression, use the above algorithm twice in the following larger estimation scheme:

1. Estimate the model using uniform weights in the definition of \( \rho_w \) — that is, for \( m \) quantiles, set \( \varphi_k = 1/m \).

2. Given final estimates \( \{\hat{g}, \hat{\beta}\} \) from part 1, construct index \( \hat{U} = X^\top \hat{\beta} \) and unweighted residuals \( \hat{\epsilon} - y - \hat{g}(\hat{U}) \), and estimate the scale of \( \epsilon \) in \( n \) median regressions of the absolute value of the estimated residuals on the estimated index:

\[
\hat{\sigma}(U_j) = \arg\min_{\sigma_j} \sum_{i=1}^n \rho_{1/2} \left( |\hat{\epsilon}_i| - \sigma_j \hat{U}_i \right) \omega_{ij} \tag{134}
\]

3. Construct weights \( \hat{\varphi} \) using the rule specified in the corollary to Theorem 3 in the text.

4. Re-estimate the model using the CQR algorithm, this time with \( \hat{\varphi} \) as weights.

\textbf{B.2 Single-index regression for a single conditional quantile}

The algorithm proposed by Wu et al. (2008) is used to estimate a single-index quantile regression for any \( \tau \):

1. Start with an initial estimate of \( \beta_0 \), for example the direct estimator proposed by Chaudhuri et al. (1997).

2. Given an estimate \( \hat{\beta}(\tau) \) (and the resulting estimated index \( \hat{U} := X^\top \hat{\beta}(\tau) \)), estimate \( g_0 \) and \( g'_0 \) in \( n \) separate quantile regressions:

\[
\{\{\hat{Q}(\hat{U}_j, \tau)\}, \{\hat{Q}'(\hat{U}_j, \tau)\}\} = \arg\min_{\{a_j\}, \{b_j\}} \sum_{i=1}^n \sum_{i=1}^n \rho_{\tau_k} \left( y_i - a_j - b_j (\hat{U}_i - \hat{U}_j) \right) \omega_{ij} \tag{135}
\]

3. Given estimates \( \{\hat{a}_j, \hat{b}_j\} \) and \( \{\omega_{ij}\}_{i,j} \) used in the previous step, estimate \( \beta_0 \) in the quantile regression

\[
\hat{\beta} = \arg\min_\beta \sum_{j=1}^n \sum_{i=1}^n \rho_{\tau_k} \left( y_i - \hat{a}_j - \hat{b}_j (X_i - X_j)^\top \beta \right) \omega_{ij} \tag{136}
\]

4. Repeat steps 2 and 3 to convergence.
C Illustrative figure

Figure 3: A figure illustrating the basic methodology of the quantile averaging estimator: the first $m = 9$ panels show individual single-index quantile regressions, while the final panel shows the estimator that results from averaging conditional quantile curves and index coefficients.
References


