Econometric Theory, **19**, 2003, 1–31. Printed in the United States of America.
DOI: 10.1017/S0266466603191013

EFFICIENT SEMIPARAMETRIC ESTIMATION OF A PARTIALLY LINEAR QUANTILE REGRESSION MODEL

SOKBAE LEE
Institute for Fiscal Studies and
University College London

This paper is concerned with estimating a conditional quantile function that is assumed to be partially linear. The paper develops a simple estimator of the parametric component of the conditional quantile. The semiparametric efficiency bound for the parametric component is derived, and two types of efficient estimators are considered. Asymptotic properties of the proposed estimators are established under regularity conditions. Some Monte Carlo experiments indicate that the proposed estimators perform well in small samples.

1. INTRODUCTION

Many econometrics problems are concerned with estimating a conditional location function such as conditional mean, conditional median, or conditional quantile. Since the seminal work of Koenker and Bassett (1978), there have been many theoretical and applied papers that are related to the estimation of conditional quantiles, including the conditional median as a special case. Most of these papers are based on a priori assumptions about the functional form of the conditional quantile function. The estimation results can be misleading, however, if the model is misspecified. On the other hand, a fully nonparametric method such as the local polynomial estimator in Chaudhuri (1991a, 1991b) could reduce the possibility of misspecification, whereas the curse of dimensionality occurs as the dimension of independent variables increases. Semiparametric methods that can obtain dimension reduction, therefore, are useful because they can avoid the loss of precision due to the curse of dimensionality and they make weaker assumptions about the functional form of the regression model.

This paper is a part of my Ph.D. dissertation submitted to the University of Iowa. I am grateful to my adviser, Joel Horowitz, for his insightful comments, suggestions, guidance, and support. I also thank John Geweke, Gene Savin, two anonymous referees, the co-editor Oliver Linton, and participants at the 2001 Midwest Econometrics Group Annual Meeting in Kansas City for many helpful comments and suggestions. Of course, the responsibility for any errors is mine. Address correspondence to: Sokbae Lee, Department of Economics, University College London, Gower Street, London WC1E 6BT, U.K.; e-mail: l.simon@ucl.ac.uk.

© 2003 Cambridge University Press 0266-4666/03 $9.50

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

In particular, this paper develops an estimation method for a partially linear quantile regression model. The model has the form

$$Y = q_\alpha(X) + Z'\beta_\alpha + U_\alpha, \tag{1}$$

where Y is a scalar dependent variable, X is a $d_x \times 1$ vector of continuous random variables, Z is a $d_z \times 1$ vector of continuous or discrete random variables, $q_\alpha(\cdot)$ is an unknown real-valued function, β_α is a $d_z \times 1$ vector of unknown parameters, and U_α is an unobserved random variable that satisfies $\text{Prob}(U_\alpha \leq 0 | X = x, Z = z) = \alpha$ for all x and z, where α indexes the quantile of interest. If $\alpha = 0.5$, the model reduces to a partially linear median regression model.

There is a growing literature on estimating semi- and nonparametric quantile regression models.[1] For example, see Chaudhuri (1991a, 1991b), Fan, Hu, and Truong (1994), and Welsh (1996) for local polynomial quantile regression; see Chaudhuri, Doksum, and Samarov (1997), Chen and Khan (2000, 2001), Khan (2001), and Khan and Powell (2001) for semiparametric estimators based on local polynomial approximations; see Koenker, Ng, and Portnoy (1994) and He, Ng, and Portnoy (1998) for smoothing splines; see He and Shi (1994, 1996) for *B*-spline approximations; see He and Liang (2000) for errors-in-variable models. Among these papers, three are especially concerned with partially linear quantile regression models. He and Shi (1996) consider M-type regression splines using bivariate tensor-product *B*-splines. He and Liang (2000) develop estimators for linear and partially linear errors-in-variables models. Chen and Khan (2001) propose an estimation method for a partially linear censored quantile regression. The aforementioned estimators are not asymptotically efficient under the conditional heteroskedasticity of U_α in (1).

The main purpose of this paper is to develop an asymptotically efficient estimator of β_α.[2] The paper first develops a simple, two-stage estimator of β_α that is an average of nonparametric estimators. In the first stage, β_α is estimated locally at each data point by some nonparametric method. These estimates of β_α are averaged in the second stage to obtain the parametric rate of convergence. This estimator, which will be called the average quantile regression (AQR) estimator, is $n^{-1/2}$-consistent and asymptotically normal, but it is not asymptotically efficient. The semiparametric efficiency bound for β_α is calculated based on a projection formula, and then two types of efficient estimators of β_α are constructed, depending on the assumption about U_α. If U_α is homoskedastic, then an optimally weighted version of the AQR estimator can attain the efficiency bound. When U_α is possibly heteroskedastic, a one-step asymptotically efficient estimator is used to attain the efficiency bound.

The rest of the paper is organized as follows. Section 2 gives sufficient conditions under which β_α and q_α are identified. Section 3 describes the AQR estimator of β_α. In Section 4, asymptotic properties of the AQR estimator are given under a certain set of regularity conditions. In Section 5, the semiparametric efficiency bound for β_α is derived. In Section 6, it is shown that the bound is attainable by constructing efficient estimators of β_α. Section 7 presents

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

some Monte Carlo experiments that illustrate the finite sample performance of the proposed estimators. Concluding remarks are given in Section 8. The proofs of theorems are in the Appendix.

2. IDENTIFICATION OF β_α AND q_α

Before we consider estimation of β_α and q_α in (1), we have to find conditions under which the partially linear quantile regression model can be identified. The model is identified if β_α and q_α are uniquely determined by the population distribution of (Y, X, Z). The following result gives sufficient conditions for the identification of the model.

THEOREM 1. *Suppose that* $\text{Prob}(Y \leq q_\alpha(X) + Z'\beta_\alpha | X = x, Z = z) = \alpha$ *for all x and z. If*

(1) the conditional density of U_α *is positive at zero for all x and z, and*
(2) $\text{var}(Z|X = x)$ *is nonsingular at every x,*

then q_α *and* β_α *are identified.*

The sufficient conditions can be relaxed but are stated in the form used to derive asymptotic properties of the proposed estimators. Condition (1) is standard but can be relaxed. For example, Knight (1998) derives the asymptotic distributions for linear median regression estimators under a more general assumption than condition (1). See also Smirnov (1952) for limiting distributions of sample quantiles under general assumptions. Condition (2) excludes a constant variable for Z. Furthermore, it requires that no components of Z be perfectly predictable by components of X. A similar but less stringent exclusion restriction is also needed for partially linear mean regression. For example, it is assumed in Robinson (1988) that $E[\text{var}(Z|X)]$ is positive definite. See Robinson (1988) for detailed discussion. It is assumed throughout the remainder of this paper that β_α and q_α are identified.

3. DESCRIPTION OF THE AVERAGE QUANTILE REGRESSION ESTIMATOR

This section presents the main idea behind the estimation method and describes the AQR estimator of β_α. The estimation procedure involves two stages: in the first stage, β_α is estimated locally at each data point; in the second stage, these local estimates are averaged to obtain an $n^{-1/2}$-consistent estimator of β_α.

To describe the estimation procedure, we need some notation. Let $\{(Y_i, X_i, Z_i); 1 \leq i \leq n\}$ be a random sample of (Y, X, Z) in (1) with size n. Let $C_n(X_i)$ be a cube in R^{d_x} centered at X_i with side length $2\delta_n$, where δ_n is a sequence of positive real numbers such that $\delta_n \to 0$. For $u = (u_1, \ldots, u_{d_x})$, a d_x-dimensional vector of nonnegative integers, let $[u] = u_1 + \cdots + u_{d_x}$. Let A_k be the set of

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

all d_x-dimensional vectors u such that $[u] \leq k$ for some integer $k \geq 0$ and let $s(A_k)$ denote the number of elements in A_k. For $z \in R^{d_x}$ with $u \in A_k$, let $z^u = \prod_{i=1}^{d_x} z_i^{u_i}$. In addition, given $X_1, X_2 \in R^{d_x}$, define

$$P_n(c, X_1, X_2) = \sum_{u \in A_k} c_u \left[\frac{X_1 - X_2}{\delta_n} \right]^u,$$

where $c = (c_u)_{u \in A_k}$ is a vector of dimension $s(A_k)$. Define $\rho_\alpha(t) = |t| + (2\alpha - 1)t$. For each X_i, the first-stage estimator $\hat{b}_\alpha(X_i)$ of β_α is defined by $\hat{b}_\alpha(X_i) \equiv \hat{b}$, where $(\hat{c}', \hat{b}')'$ is the solution to the following minimization problem:

$$\min_{c,b} \sum_{j=1, j \neq i}^{n} \rho_\alpha\{Y_j - P_n(c, X_j, X_i) - b' Z_j\} 1(X_j \in C_n(X_i)), \tag{2}$$

where $1(\cdot)$ is a standard indicator function. Notice that the preliminary estimator defined here is a leave-one-out estimator and a multivariate uniform kernel is used in (2).

It is important to note that only local data points around X_i are used in (2). As a consequence, $\hat{b}_\alpha(X_i)$ converges in probability to β_α at a nonparametric rate. A more efficient estimator for β_α requires the use of all data points. Because $\hat{b}_\alpha(X_i)$ can be obtained at each data point X_i, the proposed estimation strategy in this paper is to average all $\hat{b}_\alpha$'s to estimate β_α. This averaging method ensures that all data points are used and thus leads to a faster rate of convergence.

In the second stage, the AQR estimator $\hat{\beta}_\alpha$ of β_α is now obtained by

$$\hat{\beta}_\alpha = \frac{\sum_{i=1}^{n} \tau_x(X_i) \hat{b}_\alpha(X_i)}{\sum_{i=1}^{n} \tau_x(X_i)},$$

where $\tau_x(x)$ is a trimming function such that $\tau_x(x) = 1(x \in \mathcal{X})$ with a compact subset $\mathcal{X}$ of R^{d_x}. The trimming function is introduced to estimate the parameters β_α without being overly influenced by the tail behavior of the distribution of X. It will turn out in Section 4 that this AQR estimator is $n^{-1/2}$-consistent for β_α and asymptotically normal.

The first-stage estimation procedure is a simplified version of the local polynomial estimation procedure developed in Chaudhuri (1991a, 1991b) and Chaudhuri et al. (1997). The estimation method considered here is different from that used in Chaudhuri (1991a, 1991b) and Chaudhuri et al. (1997) in that only a linear term with respect to Z is adopted, no cross products between X and Z are included in (2), and local weighting is carried out in terms of only X, not both X and Z, because the partially linear form of the regression model is assumed in this paper. A similar type of preliminary estimation procedure is also considered in Chen and Khan (2001). The main advantage of using this method is that the Bahadur-type expansion of the estimator is al-

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

ready well established and can be easily specialized for the purpose of this paper.

We conclude this section by mentioning the numerical algorithm of (2). Because the uniform kernel is independent of c and b, the problem (2) can be easily shown to have a linear programming representation. From the perspective of linear quantile regression, problem (2) can be understood as just weighted quantile regression with weight equal to the uniform kernel. All computational algorithms developed for linear quantile regression, therefore, can be used to solve problem (2) (see, e.g., Buchinsky, 1998, and references therein).

4. ASYMPTOTIC PROPERTIES OF THE AVERAGE QUANTILE REGRESSION ESTIMATOR

This section gives regularity conditions under which the AQR estimator of β_α is $n^{-1/2}$-consistent and asymptotically normal. Let $\|\cdot\|$ denote the Euclidean norm.

Assumption 1. $\{(Y_i, X_i, Z_i); 1 \leq i \leq n\}$ is a random sample of (Y, X, Z) in (1).

Let $F_\alpha(\cdot|x,z)$ and $f_\alpha(\cdot|x,z)$, respectively, denote the cumulative distribution function and the density function of U_α conditional on $(X,Z) = (x,z)$. Moreover, let $g(x)$ denote the density function of X and let $g_{xz}(x,z)$ denote the joint density of X and Z with respect to an appropriate measure. Suppose that Z can be divided into $Z = (Z^{(c)}, Z^{(d)})$, where $Z^{(c)}$ denote the continuous components of Z and $Z^{(d)}$ denote the remaining discrete components. Assume that $Z^{(d)}$ has finitely many mass points. Let $\mathcal{W}_x$ and $\mathcal{W}_z = \mathcal{W}_z^c \times \mathcal{W}_z^d$ be supports of X and $Z = (Z^{(c)}, Z^{(d)})$ such that $\mathcal{W}_x$ and $\mathcal{W}_z^c$ are nonempty convex sets in R^{d_x} and $R^{d_z^c}$.

Assumption 2.

(a) $F_\alpha(0|x,z) = \alpha$ for all (x,z) in $\mathcal{W}_x \times \mathcal{W}_z$,
(b) $f_\alpha(u|x,z)$ is bounded away from zero and continuously differentiable with respect to u in a neighborhood of zero for all (x,z) in $\mathcal{W}_x \times \mathcal{W}_z$, and
(c) g is positive on $\mathcal{W}_x$ except on the boundary.

Following the nonparametric estimation literature, a function m: $R^d \to R$ will be said to have the order of smoothness p on a convex set $\mathcal{W}$ in R^d with $p = l + \gamma$, where $l \geq 0$ is an integer and $0 < \gamma \leq 1$, and will be written as $m \in H_p(\mathcal{W})$, if (i) partial derivatives $D^u m(x) \equiv \partial^{[u]} m(x)/\partial x_1^{u_1} \ldots \partial x_d^{u_d}$ exist and are continuous for all $x \in \mathcal{W}$ and $[u] \leq l$ and (ii) there exists a constant $M > 0$ such that

$$\|D^u m(x_1) - D^u m(x_2)\| \leq M\|x_1 - x_2\|^\gamma \quad \text{for all } x_1, x_2 \in \mathcal{W} \quad \text{and} \quad [u] = l.$$

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

In other words, a function m is l-times continuously differentiable and the lth derivative is Hölder continuous with exponent γ. The order of smoothness, if it applies to a vector or matrix-valued function, will be understood componentwise.

Assumption 3. The function $q_\alpha(\cdot)$ has the order of smoothness $p_q > 3d_x/2$ on $\mathcal{W}_x$.

Assumption 4. There exists a γ, $0 < \gamma \leq 1$ such that

(a) $f_\alpha(u|x,z)$ and $g_{xz}(x,z)$, as functions of x, belong to $H_\gamma(\mathcal{W}_x)$ for all u in a neighborhood of zero and every z in $\mathcal{W}_z$, and
(b) $g \in H_\gamma(\mathcal{W}_x)$.

Assumption 5. The distribution of Z has bounded support.

Assumption 6. $\delta_n \propto n^{-\kappa}$, where κ is a positive real number satisfying $1/(2p_q) < \kappa < 1/(3d_x)$.

Assumption 7. The trimming function $\tau_x(x)$ has compact support $\mathcal{X}$, where $\mathcal{X}$ has a nonempty interior and $\mathcal{X} \subset \mathcal{W}_x$.

Assumption 8. For all $x \in \mathcal{W}_x$, the matrix $\Sigma(x)$ is nonsingular, where

$$\Sigma(x) = E\left[f_\alpha(0|X,Z) \begin{pmatrix} 1 & Z' \\ Z & ZZ' \end{pmatrix} \middle| X = x \right].$$

Moreover, $\Sigma(x)$, considered as a function of x, is in $H_\gamma(\mathcal{W}_x)$ with some γ, $0 < \gamma \leq 1$.[3]

It is necessary to make some comments regarding regularity conditions. Condition (a) of Assumption 2 imposes the conditional quantile restriction, and condition (b) is important for identification. Condition (c) ensures that there will be sufficiently many X_j's near X_i asymptotically as $n \to \infty$. This condition with Assumption 4(b) guarantees that the marginal density of X is bounded away from zero and infinity on $\mathcal{X}$. As done in Chen and Khan (2001), it is possible to include discrete random variables for X. This, however, is not explicitly done here for the sake of simplicity.

Assumption 3 requires that the order of smoothness p_q of q_α grow as the dimension of X increases.[4] Assumptions 4 and 5 are needed to derive a Bahadur-type expansion similar to that developed in Chaudhuri et al. (1997).[5]

Assumption 6 restricts the range of the bandwidth.[6] As is common in the semiparametric estimation literature, undersmoothing is required. That is, Assumption 6 requires that δ_n converge to zero faster than $1/(2p_q + d_x)$, which is the asymptotically optimal rate for a nonparametric estimator of q_α. This is not surprising because averaging the first-stage estimators makes the variance of the AQR estimator become smaller than those of the first-stage estimators. The left inequality for δ_n in Assumption 6 is used to make the estimator have no asymptotic bias, whereas the right inequality is necessary to make the

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

remainder terms of the Bahadur-type expansion negligible. For the trimming function τ_x, a $\mathcal{X}$ that is too small can induce the loss of efficiency, whereas a $\mathcal{X}$ that is too large allows the estimator to be unduly influenced by the tail behavior of $g(x)$.

Finally, Assumption 8 ensures that the variance of the asymptotic distribution of the estimator is well defined.[7] In a homoskedastic case where $f_\alpha(0|x,z)$ is independent of x and z, Assumption 8 is satisfied if identification conditions in Theorem 1 hold and $\mathrm{var}(Z|X=x)$ is Hölder continuous.

The next theorem establishes the $n^{-1/2}$-consistency and asymptotic normality of the AQR estimator of β_α. Let e_z' be the $d_z \times (d_z+1)$ matrix such that $e_z' = (\mathbf{0}, I_{d_z})$, where $\mathbf{0}$ denotes the d_z-dimensional zero vector and I_{d_z} an identity matrix.

THEOREM 2. *Suppose that the order of the polynomial in (2) is* $k = [p_q]$. *Let* $\hat{\beta}_\alpha$ *denote the* AQR *estimator of* β_α. *Let Assumptions 1–8 hold. Then as* $n \to \infty$,

$$\sqrt{n}(\hat{\beta}_\alpha - \beta_\alpha) \to_d \mathbf{N}(0, V),$$

where

$$V = \alpha(1-\alpha)E[\{\tau_x^*(X)\}^2 e_z' \Sigma(X)^{-1} \Omega \Sigma(X)^{-1} e_z]$$

with

$$\Omega = \begin{pmatrix} 1 & Z' \\ Z & ZZ' \end{pmatrix} \quad \textit{and} \quad \tau_x^*(x) = \frac{1(x \in \mathcal{X})}{\Pr(X \in \mathcal{X})}.$$

Although the $n^{-1/2}$-consistency and asymptotic normality of the AQR estimator are established, the variance V in Theorem 2 is somewhat complicated. It will be shown in Section 5 that in general this variance is different from the efficient variance bound. The variance V in Theorem 2 can be simplified under a stronger condition than in Assumption 2. The following corollary restates Theorem 2 under the assumption of homoskedasticity.

COROLLARY 3. *Assume that the conditions in Theorem 2 hold. Furthermore, suppose that the conditional density of* U_α *given* x *and* z, *evaluated at zero, is independent of* (x,z), *namely,* $f_\alpha(0|x,z) = f_\alpha(0)$ *for all* (x,z) *in* $\mathcal{W}_x \times \mathcal{W}_z$. *Then as* $n \to \infty$,

$$\sqrt{n}(\hat{\beta}_\alpha - \beta_\alpha) \to_d \mathbf{N}(0, \widetilde{V}),$$

where

$$\widetilde{V} = \frac{\alpha(1-\alpha)}{f_\alpha^2(0)} E[\{\tau_x^*(X)\}^2 \{E[ZZ'|X] - E[Z|X]E[Z|X]'\}^{-1}].$$

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

The AQR estimator may be compared with other existing estimators in the literature. He and Shi (1996) consider M-type regression splines using bivariate tensor-product B-splines. They establish the asymptotic results under the assumption that U_α is independent of (X, Z). He and Liang (2000) develop estimators for linear and partially linear errors-in-varables models. Their asymptotic results are established under a stringent assumption that $E(Z|X = x) = 0$ for all x. Chen and Khan (2001) propose an estimation method for a partially linear censored quantile regression. Their estimator uses a two-stage estimation procedure. In the first stage, the conditional quantile function is nonparametrically estimated by the local polynomial method, which is also the case for the AQR estimator. In the second stage, they estimate β_α by a least-squares-type estimator using differenced values of the estimated conditional quantiles as dependent variables. The implementation of their estimator requires a kind of tuning parameter that they call a "selection function." None of the existing estimators in the literature are asymptotically efficient under the conditional heteroskedasticity of U_α. An asymptotically efficient estimator will be constructed in Section 6.

We end this section by considering estimation of the nonparametric component $q_\alpha(\cdot)$ of the model (1). Because the parametric component β_α can be estimated with an $n^{-1/2}$ rate, which is faster than the fastest possible rate of convergence for the nonparametric component, it is possible to estimate $q_\alpha(\cdot)$ as asymptotically efficiently as if β_α were known. The function $q_\alpha(\cdot)$ can be estimated by carrying out a local polynomial quantile regression of $Y - Z'\hat{\beta}_\alpha$ on X. See Fan and Gijbels (1996, p. 202) and Yu and Jones (1998) for rule-of-thumb bandwidths.

5. THE SEMIPARAMETRIC EFFICIENCY BOUND

In this section, the semiparametric efficiency bound for β_α will be derived by adopting the method used in Newey and Powell (1993). The semiparametric efficiency bound may be viewed as the supremum of the Cramér-Rao-type bounds for regular parametric submodels. This bound can be calculated rigorously by a projection formula. More specifically, the efficiency bound V_B for β_α is the inverse of the expectation of the outer product of the efficient score for β_α, namely, $V_B = \{E[S_\alpha S_\alpha']\}^{-1}$, where the efficient score S_α is defined by the projection of the score function for β_α onto the orthogonal complement of the tangent space in the nonparametric direction. See, for example, Newey (1990) and Bickel, Klaassen, Ritov, and Wellner (1993) for further discussion.

We consider the following parametric submodel for the nonparametric component $q_\alpha(\cdot)$ of the regression function:

$$q_{\alpha,\eta}(\cdot) = q_\alpha(\cdot) + \eta' h(\cdot),$$

where h is an arbitrary function of X that satisfies $E\|h\|^2 < \infty$. As in Newey and Powell (1993), this paper does not attempt to specify an explicit paramet-

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

ric submodel for the distribution of (U_α, X, Z). Instead, we use the existing results of Newey and Powell (1990) to derive the efficient scores $(S_{\beta_\alpha}, S_\eta)$ for $(\beta_\alpha', \eta')'$. Then the efficient score S_α for β_α will be calculated by finding the projection of S_{β_α} onto the orthogonal complement of the tangent space for η.

To begin, it is important to notice that the parametric submodel can be written as a linear quantile regression model with parameters $(\beta_\alpha', \eta')'$. It then follows from Newey and Powell (1990) that the efficient scores for β_α and η have the form

$$S_{\beta_\alpha} = k(U_\alpha, X, Z)Z \quad \text{and} \quad S_\eta = k(U_\alpha, X, Z)h(X),$$

where

$$k(U_\alpha, X, Z) = \frac{1}{\alpha(1-\alpha)} f_\alpha(0|X, Z)[\alpha - 1(U_\alpha \le 0)].$$

By Proposition A.3.5 of Bickel et al. (1993, p. 433), the projection of S_{β_α} onto the tangent space for η can be calculated by

$$\begin{aligned} k(U_\alpha, X, Z)h^*(X) &= k(U_\alpha, X, Z)\{E[k(U_\alpha, X, Z)^2|X]\}^{-1}E[S_{\beta_\alpha}k(U_\alpha, X, Z)|X] \\ &= k(U_\alpha, X, Z)T(X), \end{aligned}$$

where

$$T(X) = \frac{E[\, f_\alpha^2(0|X, Z)Z|X]}{E[\, f_\alpha^2(0|X, Z)|X]}.$$

Thus, the efficient score for β_α is

$$\begin{aligned} S_\alpha(Y, X, Z, q_\alpha, \beta_\alpha) &= S_{\beta_\alpha} - k(U_\alpha, X, Z)h^*(X) \\ &= \frac{f_\alpha(0|X, Z)}{\alpha(1-\alpha)}[\alpha - 1\{Y - q_\alpha(X) - Z'\beta_\alpha \le 0\}][Z - T(X)]. \end{aligned}$$

This yields the efficiency bound V_B

$$V_B = \alpha(1-\alpha)\left\{E[\, f_\alpha^2(0|X, Z)ZZ'] - E\left[\frac{E[\, f_\alpha^2(0|X, Z)Z|X]E[\, f_\alpha^2(0|X, Z)Z'|X]}{E[\, f_\alpha^2(0|X, Z)|X]}\right]\right\}^{-1}$$

provided that $E[S_\alpha S_\alpha']$ is nonsingular. This result implies that the AQR estimator is not efficient in general. If the distribution of U_α is homoskedastic and X and Z are jointly normal, then the variance of the AQR estimator is the same as the efficiency bound, ignoring the effect of trimming. This does not necessarily mean that the AQR estimator is efficient in that special case because imposition of the additional restriction could change the efficiency bound.

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

6. EFFICIENT ESTIMATION OF β_α

This section constructs efficient estimators of β_α. When U_α is homoskedastic with respect to Z, more precisely if $f_\alpha(0|x,z)$ is independent of z, an optimally weighted AQR estimator will deliver asymptotic efficiency; if U_α permits general heteroskedasticity, then a one-step estimator will be constructed.

6.1. Homoskedastic U_α

In this section, we will show that the efficiency bound can be attained by considering a weighted version of the AQR estimator. Because the AQR estimator is just a simple average of the nonparametric estimators, it is plausible to conjecture that a weighted average can improve asymptotic efficiency. Indeed, an efficient estimator of β_α can be obtained by choosing a proper weighting function when U_α is homoskedastic. To show this, let $w(x)$ be a $d_z \times d_z$ matrix-valued, weighting function such that $E[\tau_x(X)w(X)]$ is nonsingular and $w(x)$ is in $H_\gamma(\mathcal{W}_x)$ for some γ, $0 < \gamma \leq 1$. A weighted AQR estimator $\tilde{\beta}_\alpha$ can be defined as

$$\tilde{\beta}_\alpha = \left[\frac{1}{n}\sum_{i=1}^{n} \tau_x(X_i)w(X_i)\right]^{-1}\left[\frac{1}{n}\sum_{i=1}^{n} \tau_x(X_i)w(X_i)\hat{b}_\alpha(X_i)\right].$$

It is straightforward to show that when $f_\alpha(0|x,z) = f_\alpha(0)$ for all x and z,

$$\sqrt{n}(\tilde{\beta}_\alpha - \beta_\alpha) \to_d \mathbf{N}(0, \widetilde{V}_w),$$

where

$$\widetilde{V}_w = \frac{\alpha(1-\alpha)}{f_\alpha^2(0)} E[\tau_x(X)w(X)]^{-1}E[\tau_x(X)w(X)\mathrm{var}(Z|X)^{-1}w(X)]$$
$$\times E[\tau_x(X)w(X)]^{-1}.$$

Clearly the optimal choice of the weight function is to set $w(x) = \mathrm{var}(Z|X = x)$. In practical applications, it is likely that $\mathrm{var}(Z|X = x)$ is unknown; however, it can be replaced by its uniformly consistent estimator. For instance, we can use the following weighting function:

$$\hat{w}(x) = \hat{E}[ZZ'|X = x] - \hat{E}[Z|X = x]\hat{E}[Z|X = x]',$$

where $\hat{E}[\cdot|X = x]$ denotes a kernel estimator of the corresponding expectation. Let $\tilde{\beta}_\alpha^*$ denote the weighted AQR estimator with weight function $\hat{w}(x)$. In the following theorems, $|\cdot|$ and $o_p(1)$ will be understood componentwise.

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

THEOREM 4. *Assume that the conditions in Corollary 3 hold. Moreover, suppose that*

$$\sup_{X_i \in \mathcal{X}} |\hat{w}(X_i) - \mathrm{var}(Z|X_i)| = o_p(1).$$

Then as $n \to \infty$,

$$\sqrt{n}(\tilde{\beta}^*_\alpha - \beta_\alpha) \to_d \mathbf{N}(0, \widetilde{V}_*),$$

where

$$\widetilde{V}_* = \frac{\alpha(1-\alpha)}{f^2_\alpha(0)} E[\tau_x(X)\mathrm{var}(Z|X)]^{-1}.$$

Sufficient conditions for uniform consistency of $\hat{w}$ on a compact set can be easily obtained using the results of the literature, for example, Bierens (1983, 1987) and Andrews (1995). It is easy to see that when $f_\alpha(0|x,z) = f_\alpha(0)$ for all (x,z), the variance $\widetilde{V}_*$ is the same as the efficiency bound except for the existence of the trimming function. The estimator constructed in this paper is not efficient in a strict sense because it does not use all observations. It is expected that the loss of efficiency due to the existence of the trimming function could be eliminated by letting the support of τ_x grow very slowly as the sample size increases. For example, Robinson (1988) considers the trimming function $1(|\hat{f}_i| > b)$ (in his notation), where $\hat{f}_i$ is a kernel estimator of the probability density function of X_i and b is a positive constant. The effect of trimming is eliminated in Robinson (1988) by letting b converge to zero very slowly. In addition, Klein and Spady (1993) use elaborate trimming procedures to obtain an efficient semiparametric estimator for binary response models. Details are not worked out here, however.

If U_α permits restricted heteroskedasticity, in other words, U_α is homoskedastic with respect to Z, it is also possible to construct an efficient estimator via optimal weighting. More specifically, if $f_\alpha(0|x,z) = f_\alpha(0|x)$ for all z, then the asymptotic variance $\widetilde{V}_w$ for the weighted AQR estimator $\tilde{\beta}_\alpha$ has the following form:

$$\widetilde{V}_w = \alpha(1-\alpha)E[\tau_x(X)w(X)]^{-1}E[\tau_x(X)w(X)\{f^2_\alpha(0|X)\mathrm{var}(Z|X)\}^{-1}w(X)] \times E[\tau_x(X)w(X)]^{-1}.$$

This reveals that the optimal weighting function is $w(x) = f^2_\alpha(0|x)\mathrm{var}(Z|X = x)$. The efficiency bound, therefore, can be attained by using a consistent estimator of $f^2_\alpha(0|x)\mathrm{var}(Z|X = x)$ as the weighting function. The conditional density $f_\alpha(0|x)$ can be consistently estimated using estimated U_α's. On the other hand, if U_α permits general heteroskedasticity, then no weighted AQR estimator can deliver asymptotic efficiency. In the next section, an efficient estimator will be obtained by the one-step method.

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

6.2. Heteroskedastic U_α

In this section, a one-step asymptotically efficient estimator of β_α is constructed by taking one step from the AQR estimator of β_α.[8] Let $\hat{\beta}_\alpha$ denote the AQR estimator defined in Section 3.

If S_α were known except for β_α, a one-step asymptotically efficient estimator $\hat{\beta}_\alpha^*$ would be obtained by

$$\hat{\beta}_\alpha^* = \hat{\beta}_\alpha + \left[\sum_{i=1}^n S_\alpha(Y_i, X_i, Z_i, q_\alpha, \hat{\beta}_\alpha) S_\alpha(Y_i, X_i, Z_i, q_\alpha, \hat{\beta}_\alpha)' \right]^{-1} \times \sum_{i=1}^n S_\alpha(Y_i, X_i, Z_i, q_\alpha, \hat{\beta}_\alpha).$$

Of course, this estimator is not feasible because S_α contains unknown population quantities such as $q_\alpha(X)$, $f_\alpha(0|X,Z)$, and $T(X)$. Moreover, as pointed out by Newey and Powell (1990), the score function is not continuous in parameters β_α. As a result of this discontinuity, Newey and Powell (1990) make use of a sample splitting method for the efficient estimation of a (censored) linear quantile regression model. The sample splitting method adopted in Newey and Powell (1990) consists of using each half of the observations to estimate the efficient score for the other half. As a result, the ordering of the data may matter for the estimation of β_α.

Instead of using the technique of Newey and Powell (1990), this paper smooths the score function. Differentiability of the score function enables us to use standard Taylor series methods to obtain the asymptotic properties of the one-step estimator. The smoothing method requires the introduction of an additional tuning parameter, but we feel that this is acceptable because it is very hard to find any reasonable rule to choose the ordering of the data in practical applications. In addition, it will be shown in Section 7 that the simulation results are somewhat insensitive to the choice of the tuning parameter.[9]

Horowitz (1998a) uses a smoothed least-absolute-deviations estimator for a linear median-regression model to obtain asymptotic refinements of bootstrap tests. Following his idea, we replace the indicator function in S_α with a smooth function. Specifically, let J be a bounded, differentiable function satisfying $J(t) = 0$ if $t \leq -1$ and $J(t) = 1$ if $t \geq 1$. The function J can be regarded as the integral of a kernel function. Let $\{j_n\}$ be a sequence of positive real numbers that converges to zero. In addition, let $\tau(x,z) = 1(x \in \mathcal{X}, z^{(c)} \in \mathcal{Z})$, where $\mathcal{X}$ and $\mathcal{Z}$ are some compact subsets of R^{d_x} and $R^{d_z^c}$, respectively. The trimming function $\tau(x,z)$ is introduced for the same reason as before. For a given b, a smoothed feasible score function $\hat{S}_{\alpha i}(b)$ is then defined as

$$\hat{S}_{\alpha i}(b) = \tau(X_i, Z_i) \frac{\hat{f}_\alpha(0|X_i, Z_i)}{\alpha(1-\alpha)} \left[\alpha - 1 + J\left(\frac{Y_i - \hat{q}_\alpha(X_i) - Z_i' b}{j_n} \right) \right] \times [Z_i - \hat{T}(X_i)],$$

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

where $\hat{q}_\alpha(x)$, $\hat{f}(0|x,z)$, and $\hat{T}(x)$ denote consistent estimators of $q_\alpha(x)$, $f_\alpha(0|x,z)$, and $T(x)$, respectively. Notice that $1 - J(\cdot/j_n)$ can be arbitrarily close to $1(\cdot \leq 0)$ for sufficiently large n. An actual one-step estimator proposed here is

$$\hat{\beta}_\alpha^* = \hat{\beta}_\alpha + \left[-\sum_{i=1}^n \partial \hat{S}_{\alpha i}(\hat{\beta}_\alpha)/\partial b \right]^{-1} \sum_{i=1}^n \hat{S}_{\alpha i}(\hat{\beta}_\alpha). \tag{3}$$

To complete the description of the one-step efficient estimator, we need to specify the nonparametric estimators of $q_\alpha(X_i)$, $f_\alpha(0|X_i,Z_i)$, and $T(X_i)$. First of all, $q_\alpha(X_i)$ can be estimated by the first element of $\hat{c}$ of (2). For $f_\alpha(0|x,z)$, a standard kernel density estimator may be used. Observe that $f_\alpha(0|x,z)$ can be written as $f_\alpha(0|x,z) = f_1(0,x,z)/f_2(x,z)$, where f_1 and f_2 are joint densities of (U_α, X, Z) and (X,Z), respectively. This suggests that the conditional density of U_α at zero can be estimated consistently by obtaining the ratio of the kernel estimator of f_1 to the kernel estimator of f_2. More specifically, the kernel estimator $\hat{f}_\alpha(0|X_i,Z_i)$ is defined as

$$\hat{f}_\alpha(0|X_i,Z_i) = \frac{(n\nu_{1n}^{d_x+d_z+1})^{-1} \sum_{j=1}^n \tau(X_j,Z_j) K_{uxz}\left(\frac{\hat{U}_{\alpha j}}{\nu_{1n}}, \frac{X_j - X_i}{\nu_{1n}}, \frac{Z_j - Z_i}{\nu_{1n}} \right)}{(n\nu_{2n}^{d_x+d_z})^{-1} \sum_{j=1}^n \tau(X_j,Z_j) K_{xz}\left(\frac{X_j - X_i}{\nu_{2n}}, \frac{Z_j - Z_i}{\nu_{2n}} \right)},$$

where $\hat{U}_{\alpha i} = Y_i - \hat{q}_\alpha(X_i) - Z_i'\hat{\beta}_\alpha$, K_{uxz} is a $(d_x + d_z + 1)$-dimensional kernel function with a bandwidth ν_{1n}, and K_{xz} is a $(d_x + d_z)$-dimensional kernel function with a bandwidth ν_{2n}. Finally, $T(X_i)$ can be estimated by

$$\hat{T}(X_i) = \frac{\sum_{j=1}^n \tau(X_j,Z_j) \hat{f}_\alpha^2(0|X_j,Z_j) Z_j K_x\left(\frac{X_j - X_i}{\gamma_n} \right)}{\sum_{j=1}^n \tau(X_j,Z_j) \hat{f}_\alpha^2(0|X_j,Z_j) K_x\left(\frac{X_j - X_i}{\gamma_n} \right)},$$

where K_x is a d_x-dimensional kernel function with a bandwidth γ_n.

The following additional regularity conditions are useful to derive the asymptotic properties of the one-step estimator of β_α.

Assumption 9. The trimming function $\tau(x,z)$ has compact support $\mathcal{X} \times \mathcal{Z}$, where $\mathcal{X} \times \mathcal{Z}$ has a nonempty interior and $\mathcal{X} \times \mathcal{Z} \subset \mathcal{W}_x \times \mathcal{W}_z^c$.

Assumption 10. The conditional density $f_\alpha(u|x,z)$ is continuously twice differentiable with respect to u in a neighborhood of zero for all (x,z) in $\mathcal{W}_x \times \mathcal{W}_z$.

Assumption 11. Let $J^{(i)}(v) = d^i J(v)/dv^i$.

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

(a) $J(\cdot)$ is bounded, $J(v) = 0$ if $v \leq -1$, and $J(v) = 1$ if $v \geq 1$.
(b) J is twice differentiable, $J^{(1)}(v)$ is symmetrical about $v = 0$, $J^{(2)}$ is Lipschitz continuous, and $J^{(i)}(v)$ is bounded for $i = 1,2$.
(c) $\int_{-1}^{1} J^{(1)}(v)\,dv = 1$, $\int_{-1}^{1} vJ^{(1)}(v)\,dv = 0$, and $\int_{-1}^{1} v^2 J^{(1)}(v)\,dv > 0$.

Assumption 12. $j_n \propto n^{-\eta}$, where η is a positive real number satisfying $\frac{1}{4} < \eta < \frac{5}{18}$.

Assumption 13.

(a) $\sup_{(X_i, Z_i) \in \mathcal{X} \times \mathcal{Z}} |\hat{f}_\alpha(0|X_i, Z_i) - f_\alpha(0|X_i, Z_i)| = o_p(1)$.

(b) $\sup_{X_i \in \mathcal{X}} |\hat{T}(X_i) - T_*(X_i)| = o_p(1)$,

where

$$T_*(x) = \frac{E[\tau(X,Z) f_\alpha^2(0|X,Z) Z | X = x]}{E[\tau(X,Z) f_\alpha^2(0|X,Z) | X = x]}.$$

Assumptions 10–12 are necessary to make smoothing have no effect on the asymptotic distribution of the one-step estimator. Just like the assumption for $\hat{w}$ in Theorem 4, Assumption 13 is a high-level assumption that requires that kernel estimators be uniformly consistent on the compact sets. It is easy to obtain sufficient conditions for Assumption 13 using the results of Bierens (1983, 1987) and Andrews (1995). The main result of this paper is as follows.

THEOREM 5. *Let $\hat{\beta}_\alpha^*$ denote the one-step estimator defined in (3). Let Assumptions 1–13 hold. Then as $n \to \infty$ (assuming that V_* is well defined),*

$$\sqrt{n}(\hat{\beta}_\alpha^* - \beta_\alpha) \to_d \mathbf{N}(0, V_*),$$

where

$$V_* = \alpha(1-\alpha)\left\{ E[\tau f_\alpha^2(0|X,Z) ZZ'] - E\left[\frac{E[\tau f_\alpha^2(0|X,Z) Z | X] E[\tau f_\alpha^2(0|X,Z) Z' | X]}{E[\tau f_\alpha^2(0|X,Z) | X]} \right] \right\}^{-1}$$

with $\tau = \tau(X,Z)$.

The variance of Theorem 5 is the same as the efficiency bound except for the effect of the trimming function τ. Just like the case of homoskedastic U_α, the effect of the trimming function is likely to be eliminated by letting the support of τ grow.

For statistical inference, it is necessary to obtain a consistent estimator of the variance V_*. There may be several ways to estimate this variance matrix, but a

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

simple consistent estimator of V_* that is a by-product of the estimation procedure is

$$\hat{V}_* = \left[-n^{-1} \sum_{i=1}^{n} \partial \hat{S}_{\alpha i}(\hat{\beta}_\alpha)/\partial b \right]^{-1}.$$

It is shown in the proof of Theorem 5 in the Appendix that $\hat{V}_*$ converges to V_* in probability. An alternative way to estimate V_* is to replace the components of V_* with its empirical counterparts. A consistent estimator can be given by

$$\hat{V}_* = \alpha(1-\alpha) \left\{ n^{-1} \sum_{i=1}^{n} \left[\tau_i \hat{f}^2(0|X_i, Z_i) Z_i Z_i' - \frac{\hat{E}[\tau f_\alpha^2(0|X,Z)Z|X_i]\hat{E}[\tau f_\alpha^2(0|X,Z)Z'|X_i]}{\hat{E}[\tau f_\alpha^2(0|X,Z)|X_i]} \right] \right\}^{-1},$$

where $\hat{f}(0|x,z)$ and $\hat{E}[\cdot|x]$ denote consistent nonparametric estimators of $f_\alpha(0|x,z)$ and $E[\cdot|x]$.

Adopting the same idea, one can obtain a consistent estimator of the asymptotic variance V of the AQR estimator. Specifically, the consistent estimator $\hat{V}$ has the form

$$\hat{V} = \alpha(1-\alpha) \left[n^{-1} \sum_{i=1}^{n} \{\hat{\tau}_x^*(X_i)\}^2 e_z' \hat{\Sigma}(X_i)^{-1} \Omega_i \hat{\Sigma}(X_i)^{-1} e_z \right],$$

where

$$\Omega_i = \begin{pmatrix} 1 & Z_i' \\ Z_i & Z_i Z_i' \end{pmatrix}, \qquad \hat{\tau}_x^*(x) = \frac{1(x \in \mathcal{X})}{n^{-1} \sum_{i=1}^{n} 1(X_i \in \mathcal{X})},$$

and $\hat{\Sigma}(x)$ is a nonparametric estimator of $\Sigma(x)$ using the estimated $f_\alpha(0|X,Z)$. It is also straightforward to obtain consistent estimators of the variances of the AQR estimator and the weighted AQR estimator when U_α is homoskedastic.

7. MONTE CARLO EXPERIMENTS

This section presents the results of a Monte Carlo investigation of the finite sample performance of the proposed estimators in the previous sections. In all experiments $\alpha = 0.5$ and $n = 100$. Following Robinson (1988) and Chen and Khan (2001), we considered the following model:

$$Y_i = q(X) + Z'\beta + \sigma(X_i, Z_i)\varepsilon_i, \qquad i = 1, \ldots, n,$$

where X_i and Z_i were drawn from a bivariate standard normal distribution with correlation 0.5 and ε_i was drawn from the standard normal distribution that is

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

independent of X and Z. Three different functions for q and two different functions for σ were simulated:

$$q_1(x) = 1 + x, \qquad q_2(x) = x + 4\exp(-2x^2)/\sqrt{2\pi}, \qquad q_3(x) = \sin(\pi x),$$

and

$$\sigma_1(x,z) = \tfrac{1}{3}, \qquad \sigma_2(x,z) = C\exp[0.25(x+z)],$$

where C is a constant that was chosen to make $\sigma(X_i, Z_i)$ have standard deviation $\frac{1}{3}$. The function q_2, which is taken from Härdle (1990, p. 122), has a bell-shaped hump around zero. The parameter β was set to be 1. The trimming function used in the experiments was $\tau_x(x) = 1(|x| \leq 2)$.[10]

Computing the AQR estimates requires choosing the order of polynomial k and the bandwidth δ_n in (2). In the experiments $k = 3$. The asymptotic results of Section 3 only provide the range of δ_n in terms of the asymptotic order. A higher order asymptotic theory is required to obtain an asymptotically optimal δ_n. However, there is a simple, informal selection rule based on the rule-of-thumb bandwidth for the estimation of the nonparametric component q_α. Let $\tilde{h}_n$ be the rule-of-thumb bandwidth for the estimation of q when $k = 3$, suggested by Fan and Gijbels (1996, p. 202). Specifically, $\tilde{h}_n$ is of the form

$$\tilde{h}_n = 2.813\left[\frac{\alpha(1-\alpha)[\,\hat{f}\{\hat{F}^{-1}(\alpha)\}]^{-2}\int w_0(x)\,dx}{\sum_{i=1}^{n}[(\hat{q}_\alpha^{(4)}(X_i))]^2 w_0(X_i)}\right]^{1/9},$$

where $w_0(\cdot)$ is a weight function, $\hat{q}_\alpha(x)$ is obtained from a global polynomial fit, $\hat{f}(\cdot)$ is a kernel density estimate of the residuals of the global polynomial fit, and $\hat{F}^{-1}(\alpha)$ is the αth sample quantile of the residuals. Also, $\hat{q}_\alpha^{(4)}(x)$ denotes the fourth derivative of $\hat{q}_\alpha(x)$. The weight function was set to be $w_0(x) = 1(|x| \leq 2)$. The global polynomial fit was obtained by carrying out the median regression of Y on the constant term, $X, X^2, \ldots, X^5$, and Z. The bandwidth $\tilde{h}_n$ converges at rate $n^{-1/9}$, which is optimal for the estimation of q. Undersmoothing is required for the estimation of β. A simple bandwidth such as $\hat{\delta}_n = \tilde{h}_n \times n^{1/9} \times n^{-1/5}$ converges at rate $n^{-1/5}$ and satisfies Assumption 6 when $p_{q_\alpha} \geq 3$. By some preliminary simulations, the averages of ad hoc bandwidths $\hat{\delta}_n$ ranged between 0.7 and 1.1 across the designs considered in the experiments. In the experiments, $\delta_n \in \{0.5, 0.6, \ldots, 1.5\}$, which includes the range of the averages of $\hat{\delta}_n$. There were 1,000 replications in each experiment. The computations were carried out in GAUSS with GAUSS pseudo-random number generators.

Figure 1 shows the asymptotic (dashed lines) and empirical (solid lines) root mean squared errors (RMSEs) of the AQR estimates of β. The asymptotic RMSEs were calculated (ignoring the effect of trimming) by the formula of the

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

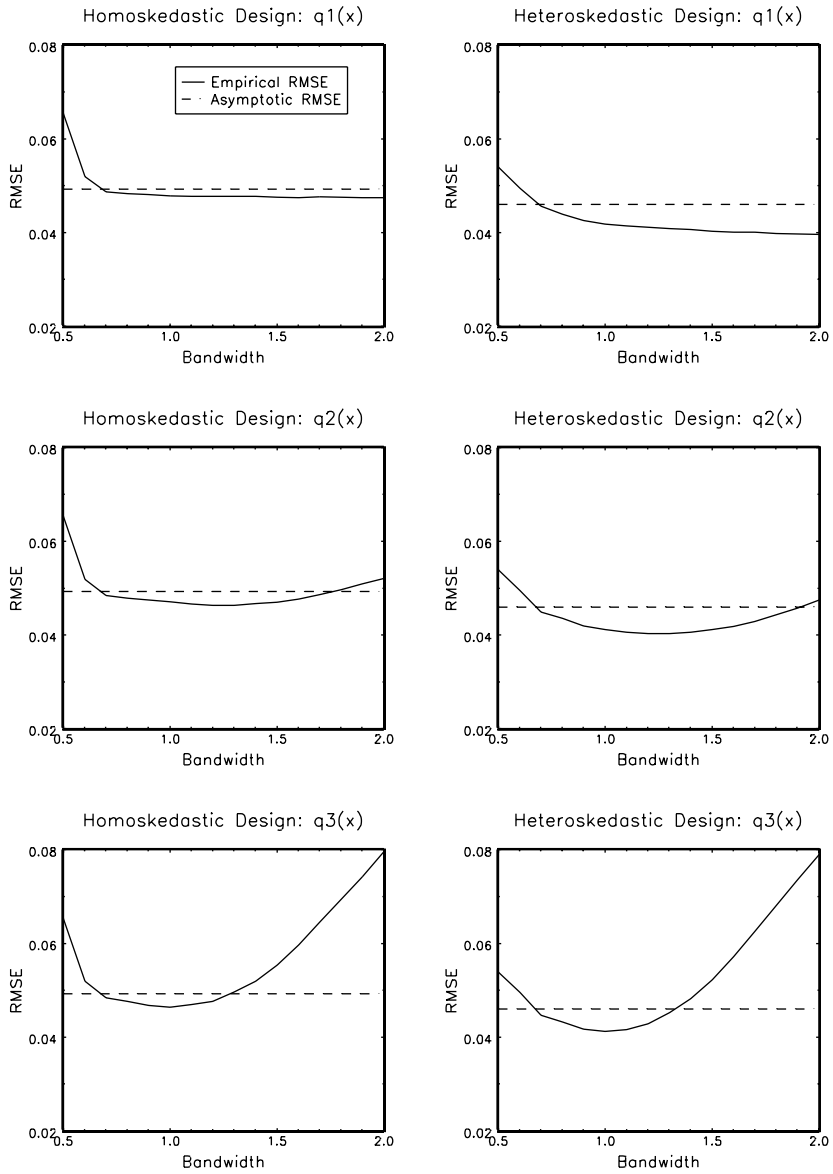


FIGURE 1. Results of the Monte Carlo experiments for the AQR estimates of β.

asymptotic variance given in Section 4. For all designs, the empirical RMSEs are quite close to the asymptotic RMSEs over a wide range of bandwidths including the range of the average $\hat{\delta}_n$. It can be seen that the results for q_3 are more sensitive to the bandwidth than those for q_1 and q_2, but the results for all

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

designs are quite insensitive to the bandwidth in the range of the rule of thumb. The empirical biases of the AQR estimates were also computed and were negligible relative to the empirical standard deviations, so they are not reported here.

More tuning parameters are required to compute the one-step efficient estimates. The trimming function was set to be $\tau(x,z) = 1(|x| \leq 2, |z| \leq 2)$. Gaussian kernels were used to estimate $f_\alpha(0|X_i, Z_i)$. Bandwidths were $\nu_{1n} = n^{-1/7}$ and $\nu_{2n} = n^{-1/6}$ with respect to standardized designs. For the estimation of $T(X_i)$, $\gamma_n = s_x n^{-1/5}$, where s_x is the sample standard deviation of X. The smoothing function J is the integral of the quartic kernel such that

$$J(v) = \begin{cases} 0 & \text{if } v < -1 \\ 0.5 + \dfrac{15}{16}\left(v - \dfrac{2}{3}v^3 + \dfrac{1}{5}v^5\right) & \text{if } |v| \leq 1 \\ 1 & \text{if } v > 1. \end{cases}$$

The AQR estimates of β and the estimates of $q(\cdot)$ were computed using bandwidth $\delta_n = 0.8$ Finally, the bandwidth j_n has to be chosen, but the asymptotic theory in Section 5 provides only qualitative restriction for the choice of j_n. The experiments focused on the sensitivity to the choice of j_n. In the experiments, $j_n = \{0.1, 0.2, \ldots, 2.0\}$.

Figure 2 shows the asymptotic (dashed lines) and empirical (solid lines) RMSEs of the one-step estimates of β. In addition, it also shows the empirical RMSEs (dotted lines) of the AQR estimates. Asymptotic results given in previous sections indicate that the AQR estimator is as efficient as the one-step estimator for homoskedastic designs. Furthermore, the optimally weighted AQR estimator is basically the same as the AQR estimator because $\text{var}(Z|X)$ is a constant in the experiments. On the other hand, it can be checked by some calculation that the asymptotic RMSE of the AQR estimator exceeds that of the one-step estimator by a factor of 1.4 for heteroskedastic designs. One noteworthy result is that the empirical RMSEs of the one-step estimates are somewhat larger than asymptotic counterparts for the heteroskedastic designs. This is not too surprising because the one-step estimates use several nonparametric estimates, which can be inaccurate for small sample size such as $n = 100$. The one-step estimates perform better than the AQR estimates for most of the values of bandwidth j_n. It also appears that the results are somewhat insensitive to the choice of the bandwidth j_n as long as j_n is not too small. In summary, the results of Monte Carlo experiments indicate that our proposed estimators work reasonably well in the finite samples.

8. CONCLUSIONS

This paper has developed a new estimation method for the partially linear quantile regression model. It is shown that the parametric component β_α can be

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

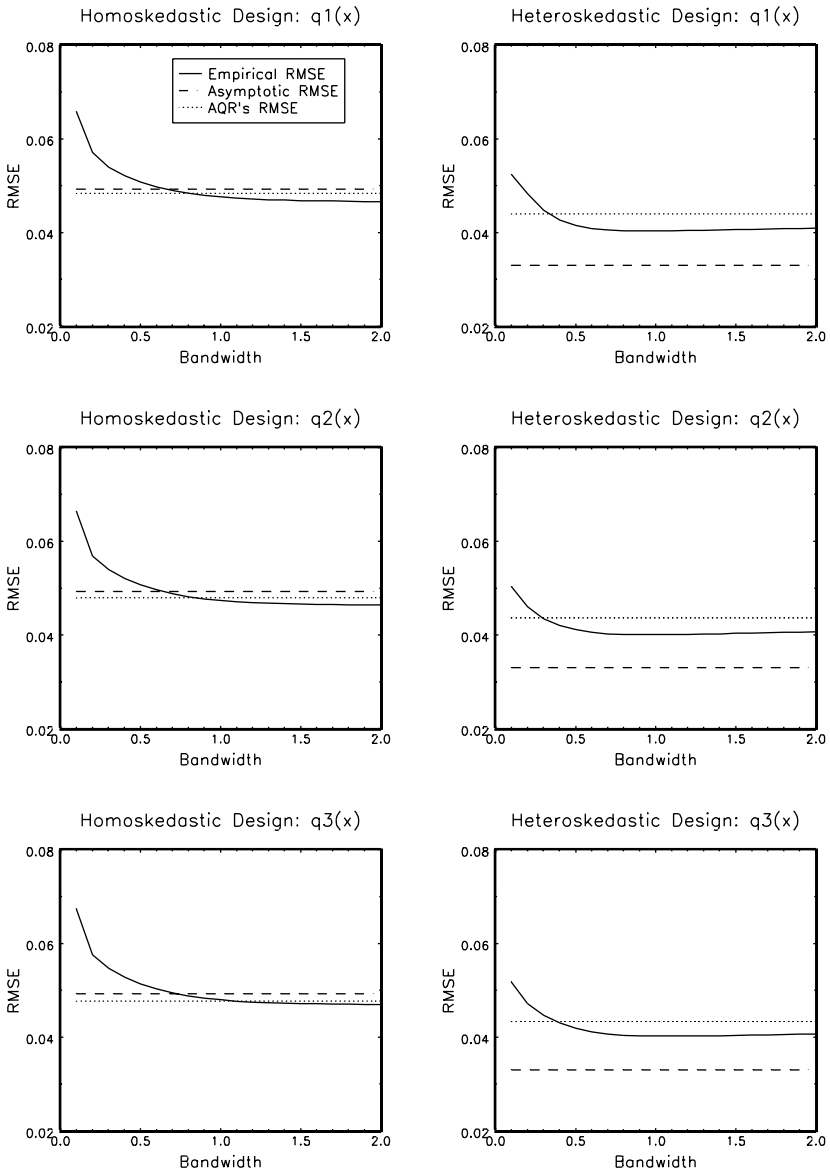


FIGURE 2. Results of the Monte Carlo experiments for the one-step estimates of β.

efficiently estimated. This paper does not investigate methods for optimally choosing the tuning parameters that are required to implement the estimation method. Because the asymptotic distributions of the proposed estimators of β_α do not depend on bandwidths, a higher order theory is required to choose opti-

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

mal bandwidths. There is no theoretical work (that we are aware of) regarding a higher order approximation for semiparametric quantile regression. This is a topic of future research. Another problem that needs to be studied is how to conduct a specification test of the regression model. Although partially linear regression is quite flexible, it still has a possibility of misspecification. It would be also an interesting problem to test a particular parametric quantile regression model against a partially linear alternative.

NOTES

1. For semi- and nonparametric mean regression models, see Härdle (1990) and Horowitz (1998b) among many others.

2. See Newey and Powell (1990) and Zhao (2001) for asymptotically efficient estimation of linear quantile regression models.

3. The exponents γ in Assumption 4(a) and (b) and Assumption 8 do not have to be same. For brevity, we assume that γ denotes the minimum of the three γ's.

4. This is common among semiparametric regression estimators. Higher order kernels are often used when the first-stage estimation is based on kernel-type estimators. Assumption 3 is not needed to derive the efficiency bound in Section 5, however.

5. In particular, Assumption 5 is made to exploit Bernstein's inequality. This assumption can be satisfied by dropping observations with very large values of Z, if necessary.

6. Assumption 6 allows for only deterministic bandwidth sequences. It is necessary to use data-based bandwidths in applications; however, it is beyond the scope of this paper to investigate the asymptotic properties of the estimator with data-dependent bandwidth sequences. In simple cases such as nonparametric density and regression estimation, the usual kinds of data-based bandwidth selection do not affect the first-order asymptotics of the estimators (see, e.g., Andrews, 1995).

7. The condition that $\Sigma(x)$ is nonsingular at every x is stronger than needed to estimate β_α. For example, β_α can be identified and estimated only using observations for which $\Sigma(x)$ is nonsingular as long as Prob$\{\Sigma(x)$ is nonsingular$\}$ is positive. Assumption 8 is adopted here to minimize the complexity of the proof.

8. In fact, any $n^{-1/2}$-consistent estimator of β_α can be used as an initial estimator.

9. Another alternative could be to invoke stochastic equicontinuity arguments in Andrews (1994). Unlike the mean regression case, $\hat{q}_\alpha$ is a step function and therefore is not smooth enough to apply the existing results in the literature.

10. If the number of observations that satisfy $|X_j - X_i| < \delta_n$ in (2) is less than 5, which is the number of regressors in the local cubic fitting, then the estimation procedure will break down. Hence, those points were additionally excluded in the experiments.

REFERENCES

Andrews, D.W.K. (1994) Empirical process methods in econometrics. In R.F. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. IV, pp. 2247–2294. New York: North-Holland.

Andrews, D.W.K. (1995) Nonparametric kernel estimation for semiparametric models. *Econometric Theory* 11, 560–596.

Bickel, P.J., C.A.J. Klaassen, Y. Ritov, & J.A. Wellner (1993) *Efficient and Adaptive Estimation for Semiparametric Models.* Baltimore: Johns Hopkins University Press.

Bierens, H.J. (1983) Uniform consistency of kernel estimators of a regression function under generalized conditions. *Journal of the American Statistical Association* 78, 699–707.

Bierens, H.J. (1987) Kernel estimators of regression functions. In T.F. Bewley (ed.), *Advances in Econometrics: Fifth World Congress*, vol. I, pp. 99–144. New York: Cambridge University Press.

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

Buchinsky, M. (1998) Recent advances in quantile regression models. *Journal of Human Resources* 33, 88–126.
Chaudhuri, P. (1991a) Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis* 39, 246–269.
Chaudhuri, P. (1991b) Nonparametric estimates of regression quantiles and their local Bahadur representation. *Annals of Statistics* 19, 760–777.
Chaudhuri, P., K. Doksum, & A. Samarov (1997) On average derivative quantile regression. *Annals of Statistics* 25, 715–744.
Chen, S. & S. Khan (2000) Estimating censored regression models in the presence of nonparametric multiplicative heteroskedasticity. *Journal of Econometrics* 98, 283–316.
Chen, S. & S. Khan (2001) Semiparametric estimation of a partially linear censored regression model. *Econometric Theory* 17, 567–590.
Fan, J. & I. Gijbels (1996) *Local Polynomial Modelling and Its Applications.* London: Chapman and Hall.
Fan, J., T.-C. Hu, & Y.K. Truong (1994) Robust nonparametric function estimation. *Scandinavian Journal of Statistics* 21, 433–446.
Härdle, W. (1990) *Applied Nonparametric Regression.* New York: Cambridge University Press.
He, X. & H. Liang (2000) Quantile regression estimates for a class of linear and partially linear errors-in-variables models. *Statistica Sinica* 10, 129–140.
He, X., P. Ng, & S. Portnoy (1998) Bivariate quantile smoothing splines. *Journal of the Royal Statistical Society, Series B* 60, 537–550.
He, X. & P. Shi (1994) Convergence rate of B-spline estimators of nonparametric conditional quantile functions. *Journal of Nonparametric Statistics* 3, 299–308.
He, X. & P. Shi (1996) Bivariate tensor-product B-splines in a partly linear model. *Journal of Multivariate Analysis* 58, 162–181.
Horowitz, J.L. (1998a) Bootstrap methods for median regression models. *Econometrica* 66, 1327–1351.
Horowitz, J.L. (1998b) *Semiparametric Methods in Econometrics.* New York: Springer-Verlag.
Khan, S. (2001) Two stage rank estimation of quantile index models. *Journal of Econometrics* 100, 319–355.
Khan, S. & J.L. Powell (2001) Two-step estimation of semiparametric censored regression models. *Journal of Econometrics* 103, 73–110.
Klein, R.W. & R.H. Spady (1993) An efficient semiparametric estimator for binary response models. *Econometrica* 61, 387–421.
Knight, K. (1998) Limiting distributions for L_1 regression estimators under general conditions. *Annals of Statistics* 26, 755–770.
Koenker, R. & G. Bassett (1978) Regression quantiles. *Econometrica* 50, 43–61.
Koenker, R., P. Ng, & S. Portnoy (1994) Quantile smoothing splines *Biometrika* 81, 673–680.
Newey, W.K. (1990) Semiparametric efficiency bounds. *Journal of Applied Econometrics* 5, 99–135.
Newey, W.K. & J.L. Powell (1990) Efficient estimation of linear and type I censored regression models under conditional quantile restrictions. *Econometric Theory* 6, 295–317.
Newey, W.K. & J.L. Powell (1993) Efficiency bounds for some semiparametric selection models. *Journal of Econometrics* 58, 169–184.
Robinson, P.M. (1988) Root-*n*-consistent semiparametric regression. *Econometrica* 56, 931–954.
Smirnov, N.V. (1952) Limit distributions for the terms of a variational series. *American Mathematical Society Translation*, no. 67, 64 pp.
Welsh, A.H. (1996) Robust estimation of smooth regression and spread functions and their derivatives. *Statistica Sinica* 6, 347–366.
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.

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

APPENDIX

A.1. The Proofs of Theorems. This part of the Appendix provides the proofs of theorems. The proofs of Corollary 3 and Theorem 4 are omitted because they are straightforward modifications of the proof of Theorem 2. The stochastic order symbols such as $o_p(1)$ or $O_p(1)$ will be understood componentwise. In addition, $|\cdot|$ is also considered componentwise. Let $U_{\alpha i} = Y_i - q_\alpha(X_i) - Z_i'\beta_\alpha$.

Proof of Theorem 1. Let $F_y(\cdot|x,z)$ denote the conditional distribution of Y given $X = x$ and $Z = z$. Then we have $F_y^{-1}(\alpha|x,z) = q_\alpha(x) + z'\beta_\alpha$. Notice that F_y^{-1} is unique given x and z by condition (1). Rewrite this as

$$(1 \quad z')\begin{pmatrix} q_\alpha(x) \\ \beta_\alpha \end{pmatrix} = F_y^{-1}(\alpha|x,z).$$

Premultiplying both sides by $(1 \quad z')'$, we have

$$\begin{pmatrix} 1 & z' \\ z & zz' \end{pmatrix}\begin{pmatrix} q_\alpha(x) \\ \beta_\alpha \end{pmatrix} = \begin{pmatrix} F_y^{-1}(\alpha|x,z) \\ zF_y^{-1}(\alpha|x,z) \end{pmatrix}$$

for all x and z. Taking expectations on both sides given x, we have

$$\begin{pmatrix} 1 & E[Z|X=x]' \\ E[Z|X=x] & E[ZZ'|X=x] \end{pmatrix}\begin{pmatrix} q_\alpha(x) \\ \beta_\alpha \end{pmatrix} = \begin{pmatrix} E[F_y^{-1}(\alpha|X,Z)|X=x] \\ E[ZF_y^{-1}(\alpha|X,Z)|X=x] \end{pmatrix}.$$

Notice that $q_\alpha(x)$ is just a point given x. If condition (2) is satisfied, then the matrix on the left-hand side is invertible. This implies that $q_\alpha(x)$ and β_α are uniquely determined. Because the choice of x is arbitrary, this completes the proof. ■

Proof of Theorem 2. Let $n_{\tau_x} = \sum_{i=1}^n \tau_x(X_i)$. Write

$$\begin{aligned}\hat{\beta}_\alpha - \beta_\alpha &= \frac{1}{n_{\tau_x}}\sum_{i=1}^n \tau_x(X_i)\{\hat{b}_\alpha(X_i) - \beta_\alpha\} \\ &= \frac{1}{n\Pr(X \in \mathcal{X})}\sum_{i=1}^n \tau_x(X_i)\{\hat{b}_\alpha(X_i) - \beta_\alpha\} + R_{\tau_x,n} \\ &\equiv T_n + R_{\tau_x,n}.\end{aligned}$$

First notice that $R_{\tau_x,n}$ results from the replacement of n_{τ_x} with $n\Pr(X \in \mathcal{X})$. It is easy to see that $R_{\tau_x,n} = o_p(n^{-1/2})$ because $\Pr(X \in \mathcal{X})$ is the expectation of $\tau_x(X)$.

To obtain the asymptotic distribution of $\hat{\beta}_\alpha - \beta_\alpha$, one needs to plug in a Bahadur-type expansion of $\hat{b}_\alpha(X_i) - \beta_\alpha$. Fortunately this type of expansion has been already well established in Chaudhuri (1991a, 1991b) and Chaudhuri et al. (1997). A modified version of the Bahadur-type linear representation for partially linear quantile regression is given in the second section of this Appendix. To describe the linear representation,

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

we need additional notation. Let $b(\delta_n, X_j - X_i, Z_j)$ denote the $(s(A_k) + d_z)$-dimensional vector

$$b(\delta_n, X_j - X_i, Z_j) = (\{\delta_n^{-[u]}(X_j - X_i)^u, [u] \le k\}', Z_j')' \tag{A.1}$$

and let $G_n(X_i)$ denote the $(s(A_k) + d_z) \times (s(A_k) + d_z)$ matrix

$$G_n(X_i) = \int_{[-1,1]^{d_x}} \int_{\mathcal{W}_z} f_\alpha(0|X_i + \delta_n t, z) b(1,t,z) b(1,t,z)' \, dP_Z(z|X_i + \delta_n t) \times g_{\delta_n}(t, X_i)\, dt, \tag{A.2}$$

where

$$g_{\delta_n}(t, X_i) = \frac{g(X_i + \delta_n t)}{\displaystyle\int_{[-1,1]^{d_x}} g(X_i + \delta_n t)\, dt}.$$

Also, $P_Z(z|x)$ denotes a probability measure with respect to Z given $X = x$. Let e' denote the $d_z \times (s(A_k) + d_z)$ matrix such that $e' = (\mathbf{0}, I_{d_z})$, where $\mathbf{0}$ denotes the $d_z \times s(A_k)$-dimensional zero matrix and I_{d_z} an identity matrix. In addition, let $N_n(X_i)$ denote the number of all j's satisfying $|X_j - X_i| \le \delta_n$ for $j \ne i$, $j = 1, \ldots, n$, and let $q_\alpha^*(X_j, X_i)$ be the k-order Taylor polynomial defined in (A.10), which follows.

It follows from Lemma 1 in the second section of this Appendix that

$$\begin{aligned}\hat{b}_\alpha(X_i) - \beta_\alpha &= e'\{N_n(X_i) G_n(X_i)\}^{-1} \\ &= \sum_{j=1, j\ne i}^{n} b(\delta_n, X_j - X_i, Z_j)[\alpha - 1\{Y_j \le q_\alpha^*(X_j, X_i) + Z_j'\beta_\alpha\}] \\ &\quad \times 1\{|X_j - X_i| \le \delta_n\} + R_n(X_i),\end{aligned}$$

where

$$\max_{X_i \in \mathcal{X}} |R_n(X_i)| = o(n^{-1/2}) \quad \text{almost surely as } n \to \infty,$$

provided that $1/(2p_q + d) < \kappa < 1/(3d_x)$. It is worth noting that $b(\delta_n, X_j - X_i, Z_j)$ and G_n are different from those of Chaudhuri et al. (1997) because a partially linear regression model is considered here.

Now substituting the Bahadur-type expansion of $\hat{b}_\alpha(X_i) - \beta_\alpha$ into T_n and following arguments similar to those in the proof of Theorem 2.1 of Chaudhuri et al. (1997) (in particular, we require $1/(2p_q) < \kappa < 1/(3d_x)$), we have

$$T_n = U_n + o_p(n^{-1/2}),$$

where U_n is a U-statistic with the kernel dependent on n:

$$U_n = \sum_{1 \le i < j \le n} \xi_n(\zeta_i, \zeta_j),$$

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

with $\zeta_i = (Y_i, X_i, Z_i)$, $\xi_n(\zeta_i, \zeta_j) = \eta_n(\zeta_i, \zeta_j) + \eta_n(\zeta_j, \zeta_i)$, $p_n(X) = \delta_n^{d_x} \int_{[-1,1]^{d_x}} g(X + \delta_n t)\, dt$, and

$$\eta_n(\zeta_i, \zeta_j) = \frac{1}{n} \tau_x^*(X_i) e' \{ n p_n(X_i) G_n(X_i) \}^{-1} b(\delta_n, X_j - X_i, Z_j) \{ \alpha - 1(U_{\alpha j} \le 0) \} \times 1\{ |X_j - X_i| \le \delta_n \}.$$

Define P_n to be the projection of U_n, so that

$$P_n = (n-1) \sum_{i=1}^{n} m_n(\zeta_i),$$

where

$$\begin{aligned} m_n(\zeta_i) &= E[\xi_n(\zeta_i, \zeta_j) | \zeta_i] \\ &= \frac{1}{n^2} \{ \alpha - 1(U_{\alpha i} \le 0) \} \\ &\quad \times E[\tau_x^*(X_j) e' \{ p_n(X_j) G_n(X_j) \}^{-1} b(\delta_n, X_i - X_j, Z_i) 1\{ |X_i - X_j| \le \delta_n \} | X_i, Z_i]. \end{aligned}$$

As in the proof of Theorem 2.1 of Chaudhuri et al. (1997), an application of the standard Hoeffding decomposition of U_n yields

$$\begin{aligned} E(U_n - P_n)^2 &= \frac{n(n-1)}{2} (E\xi_n^2(\zeta_1, \zeta_2) - 2Em_n^2(\zeta_1)) \\ &\le \frac{n(n-1)}{2} E\xi_n^2(\zeta_1, \zeta_2) \\ &\le 2n(n-1) E\eta_n^2(\zeta_1, \zeta_2). \end{aligned}$$

Using the facts that (a) in view of Assumption 8, $\| G_n^{-1}(X_i) \|$ is uniformly bounded for $X_i \in \mathcal{X}$ as $n \to \infty$, (b) $p_n(X_i) = \delta_n^{d_x} \int_{[-1,1]^{d_x}} f_\alpha(X_i + \delta_n t)\, dt$, and (c) each component of $b(\delta_n, X_2 - X_1, Z_2) 1\{ |X_2 - X_1| \le \delta_n \}$ is bounded, we have

$$E\eta_n^2(\zeta_1, \zeta_2) = O(1/n^4 \delta_n^{2d_x}).$$

This implies that

$$E(U_n - P_n)^2 = O\left(\frac{1}{n^2 \delta_n^{2d_x}} \right)$$

and, hence, using Assumption 6,

$$U_n = P_n + o_p(n^{-1/2}).$$

The next step is to evaluate the limit of the projection P_n. We will show subsequently that

$$E(P_n - \tilde{P}_n)^2 = o_p(n^{-1}), \qquad \textbf{(A.3)}$$

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

where

$$\tilde{P}_n = \frac{n-1}{n^2} \sum_{i=1}^{n} \{\alpha - 1(U_{\alpha i} \le 0)\} \tau_x^*(X_i) \nu(X_i, Z_i)$$

with

$$\nu(x,z) = e_z' E\left[f_\alpha(0|X,Z) \begin{pmatrix} 1 & Z' \\ Z & ZZ' \end{pmatrix} \middle| X = x \right]^{-1} \begin{pmatrix} 1 \\ z \end{pmatrix}.$$

Conditioning on (X_i, Z_i), we have

$$E(P_n - \tilde{P}_n)^2 = \frac{(n-1)^2 \alpha(1-\alpha)}{n^3} E\{E[\tau_x^*(X_j) e' \{p_n(X_j) G_n(X_j)\}^{-1} b(\delta_n, X_i - X_j, Z_i) \times 1\{|X_i - X_j| \le \delta_n\} | X_i, Z_i] - \tau_x^*(X_i) \nu(X_i, Z_i)\}^2. \tag{A.4}$$

As in Lemma 4.2(b) of Chaudhuri et al. (1997), it can be shown that

$$\delta_n^{d_x} \{p_n(X_j) G_n(X_j)\}^{-1} = \left\{ g(X_j) \int_{[-1,1]^{d_x}} E[b(1,t,Z) b(1,t,Z)' f_\alpha(0|X,Z) | X = X_j] \, dt \right\}^{-1} + O_{L^2}(\delta_n^\gamma),$$

where $O_{L^2}(\cdot)$ denotes a remainder term that is bounded in the L_2 norm. By a change of variables, the inner expectation in (A.4) becomes

$$\int_{[-1,1]^{d_x}} \tau_x^*(X_i - \delta_n u) e' \left\{ \int_{[-1,1]^{d_x}} E[b(1,t,Z) b(1,t,Z)' f_\alpha(0|X,Z) | X = X_i - \delta_n u] \, dt \right\}^{-1} \times b(1,u,Z_i) \, du + O_{L^2}(\delta_n^\gamma).$$

Using this and a Taylor series expansion, we can show that the outer expectation in (A.4) converges to zero, which proves (A.3). By Chebyshev inequality, (A.3) implies

$$P_n = \tilde{P}_n + o_p(n^{-1/2}).$$

Therefore,

$$\sqrt{n}(\hat{\beta}_\alpha - \beta_\alpha) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{\alpha - 1(U_{\alpha i} \le 0)\} \tau_x^*(X_i) \nu(X_i, Z_i) + o_p(1),$$

from which the desired result follows immediately using the multivariate central theorem. ■

Proof of Theorem 5. Let $I_* = V_*^{-1}$. Write

$$\sqrt{n}(\hat{\beta}_\alpha^* - \beta_\alpha) = \sqrt{n}(\hat{\beta}_\alpha - \beta_\alpha) + \left[-n^{-1} \sum_{i=1}^{n} \partial \hat{S}_{\alpha i}(\hat{\beta}_\alpha)/\partial b \right]^{-1} n^{-1/2} \sum_{i=1}^{n} \hat{S}_{\alpha i}(\hat{\beta}_\alpha).$$

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

Let

$$S_{\alpha i}(b) = \tau(X_i, Z_i)\,\frac{f_\alpha(0|X_i, Z_i)}{\alpha(1-\alpha)}\,[\alpha - 1\{Y_i - q_\alpha(X_i) - Z_i' b \le 0\}][Z - T_*(X)].$$

It will be shown subsequently that

$$-n^{-1}\sum_{i=1}^{n}\partial\hat{S}_{\alpha i}(\hat{\beta}_\alpha)/\partial b = I_* + o_p(1), \quad \text{and} \tag{A.5}$$

$$n^{-1/2}\sum_{i=1}^{n}\hat{S}_{\alpha i}(\hat{\beta}_\alpha) = n^{-1/2}\sum_{i=1}^{n}S_{\alpha i}(\beta_\alpha) - I_*\sqrt{n}(\hat{\beta}_\alpha - \beta_\alpha) + o_p(1). \tag{A.6}$$

It then follows that

$$\sqrt{n}(\hat{\beta}_\alpha^* - \beta_\alpha) = I_*^{-1} n^{-1/2}\sum_{i=1}^{n}S_{\alpha i}(\beta_\alpha) + o_p(1),$$

which gives the desired result immediately.

As in Lemma 4.3(a) of Chaudhuri et al. (1997), it can be shown that

$$\sup_{X_i \in \mathcal{X}} |\hat{q}(X_i) - q_\alpha(X_i)| = o_p(n^{-1/3}). \tag{A.7}$$

In fact, we require that $1/(2p_q + d_x) < \kappa < 1/(3d_x)$ in view of Lemma 1 (in particular, see step 3 in the proof). Furthermore, by similar arguments as in the proof of Theorem 2, it can be proved that

$$n^{-1}\sum_{i=1}^{n}\tau_x(X_i)[\hat{q}(X_i) - q_\alpha(X_i)] = O_p(n^{-1/2}). \tag{A.8}$$

Let $\tau_i = \tau(X_i, Z_i)$, $f_i = f_\alpha(0|X_i, Z_i)$, and $\hat{f}_i = \hat{f}_\alpha(0|X_i, Z_i)$ for shorthand notation. To show (A.5), note that

$$\begin{aligned}
-n^{-1}\sum_{i=1}^{n}\partial\hat{S}_{\alpha i}(\hat{\beta}_\alpha)/\partial b &= \frac{1}{nj_n}\sum_{i=1}^{n}\frac{\tau_i\hat{f}_i}{\alpha(1-\alpha)}J^{(1)}\left(\frac{\hat{U}_{\alpha i}}{j_n}\right)[Z_i - \hat{T}(X_i)]Z_i' \\
&= \frac{1}{nj_n}\sum_{i=1}^{n}\frac{\tau_i f_i}{\alpha(1-\alpha)}J^{(1)}\left(\frac{\hat{U}_{\alpha i}}{j_n}\right)[Z_i - T_*(X_i)]Z_i' + o_p(1) \\
&= \frac{1}{nj_n}\sum_{i=1}^{n}\frac{\tau_i f_i}{\alpha(1-\alpha)}J^{(1)}\left(\frac{U_{\alpha i}}{j_n}\right)[Z_i - T_*(X_i)]Z_i' + o_p(1) \\
&= I_* + o_p(1),
\end{aligned}$$

where the first equality comes from differentiation, the second equality follows in view of the uniform consistency of $\hat{f}_\alpha(0|X_i, Z_i)$ and $\hat{T}(X_i)$, and the fourth equality can be proved using methods similar to those used to calculate the asymptotic mean of kernel density estimators. To prove the third equality, note that by a Taylor series expansion,

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

$$\frac{1}{nj_n}\sum_{i=1}^{n} J^{(1)}\left(\frac{\hat{U}_{\alpha i}}{j_n}\right) = \frac{1}{nj_n}\sum_{i=1}^{n} J^{(1)}\left(\frac{U_{\alpha i}}{j_n}\right) + \frac{1}{nj_n^2}\sum_{i=1}^{n} J^{(2)}\left(\frac{U_{\alpha i}}{j_n}\right)(\hat{U}_{\alpha i} - U_{\alpha i}) + R_{n,J}, \tag{A.9}$$

where the remainder term $R_{n,J}$ is

$$R_{n,J} = \frac{1}{nj_n^2}\sum_{i=1}^{n}\left[J^{(2)}\left(\frac{\tilde{U}_{\alpha i}}{j_n}\right) - J^{(2)}\left(\frac{U_{\alpha i}}{j_n}\right)\right](\hat{U}_{\alpha i} - U_{\alpha i})$$

and $\tilde{U}_{\alpha i}$ is between $\hat{U}_{\alpha i}$ and $U_{\alpha i}$. Observe that the second term in (A.9) is of order $o_p(1)$, because $\sup_i \tau_i|\hat{U}_{\alpha i} - U_{\alpha i}| = o_p(1)$ by (A.7) and

$$\frac{1}{nj_n^2}\sum_{i=1}^{n} J^{(2)}\left(\frac{U_{\alpha i}}{j_n}\right) = O_p(1)$$

with the restriction that $nj_n^3 \to \infty$. Now we need to show that $R_{n,J} = o_p(1)$. To see this, use the fact that $J^{(2)}$ is Lipschitz continuous to obtain

$$|R_{n,J}| \le M_J\left|\frac{1}{nj_n^3}\sum_{i=1}^{n}\tau_i(\hat{U}_{\alpha i} - U_{\alpha i})\right| \sup_i \tau_i|\tilde{U}_{\alpha i} - U_{\alpha i}|$$

for some constant M_J. It follows from (A.7), (A.8), and the restriction on j_n that $R_{n,J} = o_p(j_n^{-3}n^{-5/6}) = o_p(1)$.

For (A.6), first note that once again by the uniform consistency of $\hat{f}_\alpha(0|X_i, Z_i)$ and $\hat{T}(X_i)$,

$$\begin{aligned} n^{-1/2}&\sum_{i=1}^{n}\hat{S}_{\alpha i}(\hat{\beta}_\alpha) \\ &= n^{-1/2}\sum_{i=1}^{n}\frac{\tau_i f_i}{\alpha(1-\alpha)}\left[\alpha - 1 + J\left(\frac{Y_i - \hat{q}_\alpha(X_i) - Z_i'\hat{\beta}_\alpha}{j_n}\right)\right] \\ &\quad \times [Z_i - T_*(X_i)][1 + o_p(1)] \\ &\equiv n^{-1/2}\sum_{i=1}^{n}\tilde{S}_{\alpha i}(\hat{q}_\alpha, \hat{\beta}_\alpha)[1 + o_p(1)]. \end{aligned}$$

Using a Taylor series expansion,

$$\begin{aligned} n^{-1/2}\sum_{i=1}^{n}\tilde{S}_{\alpha i}(\hat{q}_\alpha, \hat{\beta}_\alpha) &= n^{-1/2}\sum_{i=1}^{n}\tilde{S}_{\alpha i}(q_\alpha, \beta_\alpha) + \left[n^{-1}\sum_{i=1}^{n}\frac{\partial\tilde{S}_{\alpha i}(q_\alpha, \beta_\alpha)}{\partial b}\right]\sqrt{n}(\hat{\beta}_\alpha - \beta_\alpha) \\ &\quad + n^{-1/2}\sum_{i=1}^{n}\frac{\partial\tilde{S}_{\alpha i}(q_\alpha, \beta_\alpha)}{\partial q}[\hat{q}_\alpha(X_i) - q_\alpha(X_i)] + R_{n,S}, \end{aligned}$$

where $R_{n,S}$ is a Taylor series remainder term. As before, the expression in the brackets of the second term is $-I_* + o_p(1)$. Using the same arguments as in (A.9), one can show that the remainder term $R_{n,S}$ is $o_p(1)$. As in Lemma 3(k) of Horowitz (1998a), it is not difficult to show that using Chebyshev inequality,

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

$$n^{-1/2}\sum_{i=1}^{n}\left[J\left(\frac{U_{\alpha i}}{j_n}\right)-1(U_{\alpha i}>0)\right]=O_p(n^{1/2}j_n^2+j_n^{1/2}).$$

This implies that using the restriction on j_n,

$$n^{-1/2}\sum_{i=1}^{n}\tilde{S}_{\alpha i}(q_\alpha,\beta_\alpha)=n^{-1/2}\sum_{i=1}^{n}S_{\alpha i}(\beta_\alpha)+o_p(1).$$

To prove (A.6), it now suffices to show that the third term in the Taylor series expansion is $o_p(1)$. To see this, note that under Assumption 10,

$$\begin{aligned}n^{-1/2}\sum_{i=1}^{n}\frac{\partial\tilde{S}_{\alpha i}(q_\alpha,\beta_\alpha)}{\partial q}&=-n^{-1/2}\sum_{i=1}^{n}\frac{\tau_i f_i}{\alpha(1-\alpha)j_n}J^{(1)}\left(\frac{U_{\alpha i}}{j_n}\right)[Z_i-T_*(X_i)]\\&=n^{1/2}O_p(j_n^2+(nj_n)^{-1/2})\end{aligned}$$

by Chebyshev inequality. Hence,

$$\begin{aligned}&\left|n^{-1/2}\sum_{i=1}^{n}\frac{\partial\tilde{S}_{\alpha i}(q_\alpha,\beta_\alpha)}{\partial q}[\hat{q}_\alpha(X_i)-q_\alpha(X_i)]\right|\\&\quad\le\left|n^{-1/2}\sum_{i=1}^{n}\frac{\partial\tilde{S}_{\alpha i}(q_\alpha,\beta_\alpha)}{\partial q}\right|\sup_{X_i\in\mathcal{X}}|\hat{q}_\alpha(X_i)-q_\alpha(X_i)|\\&\quad=o_p(1),\end{aligned}$$

where the last equality follows from (A.7) and the restriction on j_n. ■

A.2. Linear Representation for the First-Step Estimators. The result presented here is an extension of the Bahadur-type representation for the local polynomial conditional quantile estimators given in Lemma 4.1 in Chaudhuri et al. (1997). The following lemma assumes partially linear quantile regression and gives a uniform result in the conditional variables X_i.

Define the k-order Taylor polynomial

$$q_\alpha^*(x+\delta_n t,x)=\sum_{u\in A_k}c_{n,u}(x)t^u, \tag{A.10}$$

with the coefficients $c_{n,u}(x)=(u!)^{-1}D^u(x)\delta_n^{[u]}$, where $u!=u_1!\ldots u_d!$. Let $c_\alpha(x)=(c_{n,u}(x))_{u\in A_k}$ denote the $s(A_k)$-dimensional vector of Taylor coefficients in (A.10). In addition, let $\hat{a}_\alpha(X_i)=(\hat{c}_\alpha(X_i)',\hat{b}_\alpha(X_i)')'$ denote the $(s(A_k)+d_z)$-dimensional vector that minimizes (2) and let $a_\alpha(X_i)=(c_\alpha(X_i)',\beta_\alpha')'$. Moreover, let $b(\delta_n,X_j-X_i,Z_j)$ and $G_n(X_i)$ be defined in (A.1) and (A.2), respectively.

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

LEMMA 1. *Let Assumptions 1–5, 7, and 8 hold. Then we have*

$$\hat{a}_\alpha(X_i) - a_\alpha(X_i) = \{N_n(X_i)G_n(X_i)\}^{-1} \times \sum_{j=1,\, j\neq i}^{n} b(\delta_n, X_j - X_i, Z_j)[\alpha - 1\{Y_j \le q_\alpha^*(X_j, X_i) + Z_j'\beta_\alpha\}] \times 1\{|X_j - X_i| \le \delta_n\} + R_n(X_i), \tag{A.11}$$

where the remainder term $R_n(X_i)$ *satisfies*

$$\max_{x_i \in \mathcal{X}} |R_n(X_i)| = o(n^{-3(1-\kappa d_x)/4}[\log n]^{3/4}) \quad \textit{almost surely as } n \to \infty,$$

provided that $\delta_n \propto n^{-\kappa}$ *with* $1/(2p_q + d) < \kappa < 1/d_x$.

Proof of Lemma 1. We will provide the main ideas and indicate the differences that result from partially linear quantile regression. We will split the proof into several steps.

Step 1. Step 1 of the proof of Lemma 4.1 in Chaudhuri et al. (1997) can be repeated without any modification for our purpose. Specifically, we can choose a pair of positive constants c_1 and c_2 such that

$$\Pr(\liminf E_n) = 1,$$

where

$$E_n = \{c_1 n^{1-\kappa d_x} \le N_n(X_i) \le c_2 n^{1-\kappa d_x} \quad \text{for all } X_i \in \mathcal{X}\},$$

provided that $0 < \kappa < 1/d_x$.

Step 2. The rate of the uniform convergence can be derived by using arguments similar to those in Chaudhuri et al. (1997). It can be shown that there exists a positive constant K_1 such that

$$\Pr(\liminf F_n) = 1, \tag{A.12}$$

where

$$F_n = \{|\hat{a}_\alpha(X_i) - a_\alpha(X_i)| \le K_1 n^{-(1-\kappa d_x)/2}(\log n)^{1/2} \quad \text{for all } X_i \in \mathcal{X}\},$$

provided that $\kappa > 1/(2p_q + d_x)$. Under Assumptions 1–5, 7, and 8, Fact 6.5 in Chaudhuri (1991b) and Fact 5.2 in Chaudhuri (1991a) can be easily restated for our purpose. Then (A.12) follows from simple modifications of the arguments used in the proofs of Theorem 3.2 in Chaudhuri (1991b) and Theorems 3.2 and 3.3 in Chaudhuri (1991a).

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

Step 3. Let $\hat{q}_\alpha(x + \delta_n t, x)$ denote the Taylor polynomial with estimated coefficients, which is (A.10) with $\hat{c}_\alpha(x)$. Define

$$\tilde{H}_n(\delta_n, X_i) = \int_{[-1,1]^{d_x}} \int_{\mathcal{W}_z} F_\alpha[q^*_\alpha(X_i + \delta_n t, X_i) - q_\alpha(X_i + \delta_n t)|X_i + \delta_n t, z] \times b(1, t, z)\, dP_Z(z|X_i + \delta_n t)\, g_{\delta_n}(t, X_i)\, dt,$$

$$\hat{H}_n(\delta_n, X_i) = \int_{[-1,1]^{d_x}} \int_{\mathcal{W}_z} F_\alpha[\hat{q}_\alpha(X_i + \delta_n t, X_i) - q_\alpha(X_i + \delta_n t) + z'(\hat{b}_\alpha(X_i) - \beta_\alpha)|X_i + \delta_n t, z] \times b(1, t, z)\, dP_Z(z|X_i + \delta_n t)\, g_{\delta_n}(t, X_i)\, dt,$$

and

$$R^{(1)}_n(X_i) = \hat{H}_n(\delta_n, X_i) - \tilde{H}_n(\delta_n, X_i) - G_n(X_i)\{\hat{a}_\alpha(X_i) - a_\alpha(X_i)\}. \quad \textbf{(A.13)}$$

Using the result in step 2 and the continuous differentiability of $f_\alpha(u|x, z)$ with respect to u in a neighborhood of zero, we can show that

$$R^{(1)}_n(X_i) = (n^{-(1-\tau d_x)} \log n)$$

almost surely as $n \to \infty$ uniformly in $X_i \in \mathcal{W}_x$.

Step 4. For each X_i, define an $(s(A_k) + d_z)$-dimensional random vector $\chi_n(X_i)$ as

$$\chi_n(X_i) = \sum_{j=1, j\neq i}^{n} \{b(\delta_n, X_j - X_i, Z_j)1\{Y_j \le \hat{q}_\alpha(X_j, X_i) + Z_j'\hat{b}_\alpha(X_i)\} - \hat{H}_n(\delta_n, X_i)\}1\{|X_j - X_i| \le \delta_n\} - \sum_{j=1, j\neq i}^{n} \{b(\delta_n, X_j - X_i, Z_j)1\{Y_j \le q^*_\alpha(X_j, X_i) + Z_j'\beta_\alpha\} - \tilde{H}_n(\delta_n, X_i)\}1\{|X_j - X_i| \le \delta_n\}.$$

It follows from arguments similar to those used in the proof of Theorem 3.3 in Chaudhuri (1991b) that

$$\chi_n(X_i) = (n^{(1-\tau d_x)/4}[\log n]^{3/4}) \quad \textbf{(A.14)}$$

almost surely as $n \to \infty$ uniformly in $X_i \in \mathcal{W}_x$. To exploit Bernstein's inequality, it is assumed here that Z is bounded.

Step 5. Write

$$[N_n(X_i)]^{-1} \sum_{j=1, j\neq i}^{n} b(\delta_n, X_j - X_i, Z_j)[\alpha - 1\{Y_j \le q^*_\alpha(X_j, X_i) + Z_j'\beta_\alpha\}]1\{|X_j - X_i| \le \delta_n\} = [N_n(X_i)]^{-1}\chi_n(X_i) + \hat{H}_n(\delta_n, X_i) - \tilde{H}_n(\delta_n, X_i) + R^{(2)}_n(X_i), \quad \textbf{(A.15)}$$

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press

where on the event E_n. As in step 3 of the proof of Theorem 3.3 in Chaudhuri (1991b), it can be shown that

$$R_n^{(2)}(X_i) \leq Mn^{-(1-\kappa d_x)} \tag{A.16}$$

for some constant M, which is uniform in $X_i \in \mathcal{X}$. Finally, the desired result of the lemma follows by combining (A.13)–(A.16) and the fact that $G_n(X_i)$ is invertible as $n \to \infty$. ■

https://doi.org/10.1017/S0266466603191013 Published online by Cambridge University Press