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

\[ Y = q_\alpha(X) + Z'\beta_\alpha + U_\alpha, \]  

(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, \( \beta_\alpha \) is a \( d_z \times 1 \) vector of unknown parameters, and \( U_\alpha \) is an unobserved random variable that satisfies \( \text{Prob}(U_\alpha \leq 0|X = x, Z = z) = \alpha \) for all \( x \) and \( z \), where \( \alpha \) indexes the quantile of interest. If \( \alpha = 0.5 \), the model reduces to a partially linear median regression model.


The main purpose of this paper is to develop an asymptotically efficient estimator of \( \beta_\alpha \). The paper first develops a simple, two-stage estimator of \( \beta_\alpha \) that is an average of nonparametric estimators. In the first stage, \( \beta_\alpha \) is estimated locally at each data point by some nonparametric method. These estimates of \( \beta_\alpha \) 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 \( \beta_\alpha \) is calculated based on a projection formula, and then two types of efficient estimators of \( \beta_\alpha \) are constructed, depending on the assumption about \( U_\alpha \). If \( U_\alpha \) is homoskedastic, then an optimally weighted version of the AQR estimator can attain the efficiency bound. When \( U_\alpha \) 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 \( \beta_\alpha \) and \( q_\alpha \) are identified. Section 3 describes the AQR estimator of \( \beta_\alpha \). 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 \( \beta_\alpha \) is derived. In Section 6, it is shown that the bound is attainable by constructing efficient estimators of \( \beta_\alpha \). Section 7 presents...
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 $\beta_\alpha$ AND $q_\alpha$

Before we consider estimation of $\beta_\alpha$ and $q_\alpha$ in (1), we have to find conditions under which the partially linear quantile regression model can be identified. The model is identified if $\beta_\alpha$ and $q_\alpha$ 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 $U_\alpha$ is positive at zero for all $x$ and $z$, and
2. $\text{var}(Z|X = x)$ is nonsingular at every $x$,

then $q_\alpha$ and $\beta_\alpha$ 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 $\beta_\alpha$ and $q_\alpha$ 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 $\beta_\alpha$. The estimation procedure involves two stages: in the first stage, $\beta_\alpha$ 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 $\beta_\alpha$.

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 $\delta_n$ 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
all $d_\alpha$-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 $z \in R^{d_\alpha}$ with $u \in A_k$, let $z^u = \prod_{i=1}^{d_\alpha} z_i^{u_i}$. In addition, given $X_1, X_2 \in R^{d_\alpha}$, define

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

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 $\beta_\alpha$ is defined by $\hat{b}_\alpha(X_i) = \tilde{b}$, where $(\tilde{c}', \tilde{b}')'$ is the solution to the following minimization problem:

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

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 $\beta_\alpha$ at a nonparametric rate. A more efficient estimator for $\beta_\alpha$ 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 $\beta_\alpha$. 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 $\beta_\alpha$ is now obtained by

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

where $\tau_\alpha(x)$ is a trimming function such that $\tau_\alpha(x) = 1(x \in \mathcal{X})$ with a compact subset $\mathcal{X}$ of $R^{d_\alpha}$. The trimming function is introduced to estimate the parameters $\beta_\alpha$ 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 $\beta_\alpha$ 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-
Partially linear quantile regression

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 $\beta_\alpha$ 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 $F_\alpha(\cdot \mid x, z)$ and $f_\alpha(\cdot \mid x, z)$, respectively, denote the cumulative distribution function and the density function of $U_\alpha$ conditional on $(X, Z) = (x, z)$. Moreover, let $g(x)$ denote the density function of $X$ and let $g_{x,z}(x, z)$ 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 $\mathcal{W}_z = \mathcal{W}_x^c \times \mathcal{W}_z^d$ 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 $R^{d_z}$.

Assumption 2.

(a) $F_\alpha(0 \mid x, z) = \alpha$ for all $(x, z)$ in $\mathcal{W}_x \times \mathcal{W}_z$,
(b) $f_\alpha(u \mid x, z)$ 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 $D^{\alpha}m(x) \equiv \partial^{[\alpha]}m(x)/\partial x^{\alpha}_1 \ldots \partial x^{\alpha}_d$ exist and are continuous for all $x \in \mathcal{W}$ and $[\alpha] \leq l$ and (ii) there exists a constant $M > 0$ such that

$$\|D^{\alpha}m(x_1) - D^{\alpha}m(x_2)\| \leq M\|x_1 - x_2\|^\gamma$$

for all $x_1, x_2 \in \mathcal{W}$ and $[\alpha] = l$.
In other words, a function $m$ is $l$-times continuously differentiable and the $l$th derivative is Hölder continuous with exponent $\gamma$. 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 $p_\alpha > 3d_x/2$ on $\mathcal{W}_x$.

Assumption 4. There exists a $\gamma$, $0 < \gamma \leq 1$ 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 $\kappa$ is a positive real number satisfying $1/(2p_\alpha) < \kappa < 1/(3d_x)$.

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

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

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

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 $p_\alpha$ of $q_\alpha$ 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 $\delta_n$ converge to zero faster than $1/(2p_\alpha + d_x)$, which is the asymptotically optimal rate for a nonparametric estimator of $q_\alpha$. 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 $\delta_n$ in Assumption 6 is used to make the estimator have no asymptotic bias, whereas the right inequality is necessary to make the
remainder terms of the Bahadur-type expansion negligible. For the trimming function $\tau_x$, 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. 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 $\text{var}(Z|X = x)$ is Hölder continuous.

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

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

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

where

$$V = \alpha(1 - \alpha)E\left[\left(\tau^*_x(X)\right)^2 e_\xi' \Sigma(X)^{-1} \Omega \Sigma(X)^{-1} e_\xi\right]$$

with

$$\Omega = \begin{pmatrix} 1 & Z' \\ Z & ZZ' \end{pmatrix}$$

and

$$\tau^*_x(x) = \frac{1(x \in \mathcal{X})}{\Pr(X \in \mathcal{X})}.$$

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 $U_\alpha$ 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 $n \to \infty$,

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

where

$$\bar{V} = \frac{\alpha(1 - \alpha)}{f^2_\alpha(0)} E\left[\left(\tau^*_x(X)\right)^2 E[Z' | X] - E[Z | X] E[Z | X]' \right]^{-1}.$$
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 $U_a$ is independent of $(X, Z)$. He and Liang (2000) develop estimators for linear and partially linear errors-in-variables 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 $\beta_a$ 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 $U_a$. An asymptotically efficient estimator will be constructed in Section 6.

We end this section by considering estimation of the nonparametric component $q_a(\cdot)$ of the model (1). Because the parametric component $\beta_a$ 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 $q_a(\cdot)$ as asymptotically efficiently as if $\beta_a$ were known. The function $q_a(\cdot)$ can be estimated by carrying out a local polynomial quantile regression of $Y - Z'\beta_a$ 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 $\beta_a$ 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 $\beta_a$ is the inverse of the expectation of the outer product of the efficient score for $\beta_a$, namely, $V_B = \{E[S_aS'_a]\}^{-1}$, where the efficient score $S_a$ is defined by the projection of the score function for $\beta_a$ 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 $q_a(\cdot)$ of the regression function:

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

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-
ric submodel for the distribution of \((U_a, X, Z)\). Instead, we use the existing results of Newey and Powell (1990) to derive the efficient scores \((S_{\beta_a}, S_{\eta})\) for \((\beta'_a, \eta')\). Then the efficient score \(S_{\eta}\) for \(\beta_a\) will be calculated by finding the projection of \(S_{\beta_a}\) onto the orthogonal complement of the tangent space for \(\eta\).

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

\[
S_{\beta_a} = k(U_a, X, Z)Z \quad \text{and} \quad S_{\eta} = k(U_a, X, Z)h(X),
\]

where

\[
k(U_a, X, Z) = \frac{1}{\alpha(1 - \alpha)} f_a(0|X, Z)[\alpha - 1(U_a \leq 0)].
\]

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

\[
k(U_a, X, Z)h^*(X) = k(U_a, X, Z)\{E[k(U_a, X, Z)^2|X]\}^{-1}E[S_{\beta_a}k(U_a, X, Z)|X]
\]

\[
= k(U_a, X, Z)T(X),
\]

where

\[
T(X) = \frac{E[f^2_a(0|X, Z)Z|X]}{E[f^2_a(0|X, Z)|X]}.
\]

Thus, the efficient score for \(\beta_a\) is

\[
S_a(Y, X, Z, q_a, \beta_a) = S_{\beta_a} - k(U_a, X, Z)h^*(X)
\]

\[
= \frac{f_a(0|X, Z)}{\alpha(1 - \alpha)} \{\alpha - 1\{Y - q_a(X) - Z'\beta_a \leq 0\}\}[Z - T(X)].
\]

This yields the efficiency bound \(V_B\)

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

provided that \(E[S_aS_a']\) is nonsingular. This result implies that the AQR estimator is not efficient in general. If the distribution of \(U_a\) 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.
6. EFFICIENT ESTIMATION OF $\beta_\alpha$

This section constructs efficient estimators of $\beta_\alpha$. When $U_\alpha$ 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 $U_\alpha$ permits general heteroskedasticity, then a one-step estimator will be constructed.

6.1. Homoskedastic $U_\alpha$

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 $\beta_\alpha$ can be obtained by choosing a proper weighting function when $U_\alpha$ is homoskedastic. To show this, let $w(x)$ be a $d_x \times d_x$ matrix-valued, weighting function such that $E[t_x X w(X)]$ is nonsingular and $w(x)$ is in $H_\gamma(\mathcal{V}_\alpha)$ for some $\gamma$, $0 < \gamma \leq 1$. A weighted AQR estimator $\hat{\beta}_\alpha$ can be defined as

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

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

$$\sqrt{n}(\hat{\beta}_\alpha - \beta_\alpha) \rightarrow_d N(0,\tilde{V}_w),$$

where

$$\tilde{V}_w = \frac{\alpha(1-\alpha)}{f_\alpha^2(0)} E[t_x X w(X)]^{-1} E[t_x X w(X) \text{var}(Z|X)^{-1} w(X)]$$

$$\times E[t_x X w(X)]^{-1}.$$ 

Clearly the optimal choice of the weight function is to set $w(x) = \text{var}(Z|X = x)$. In practical applications, it is likely that $\text{var}(Z|X = x)$ 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 $\hat{\beta}_\alpha^*$ denote the weighted AQR estimator with weight function $\hat{w}(x)$. In the following theorems, $\cdot | \cdot$ and $o_p(1)$ will be understood componentwise.
THEOREM 4. Assume that the conditions in Corollary 3 hold. Moreover, suppose that

$$\sup_{x_i \in X} |\hat{\beta}_\alpha(X_i) - \text{var}(Z|X_i)| = o_p(1).$$

Then as $n \to \infty$,

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

where

$$\bar{V}_\alpha = \frac{\alpha(1-\alpha)}{f_\alpha^2(0)} E[\tau_\alpha(X) \text{var}(Z|X)]^{-1}.$$

Sufficient conditions for uniform consistency of $\hat{\beta}$ 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 $\bar{V}_\alpha$ 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 $\tau_x$ 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 $U_\alpha$ permits restricted heteroskedasticity, in other words, $U_\alpha$ 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 $\bar{V}_\alpha$ for the weighted AQR estimator $\hat{\beta}_\alpha$ has the following form:

$$\bar{V}_\alpha = \alpha(1-\alpha)E[\tau_\alpha(X)w(X)]^{-1}E[\tau_\alpha(X)w(X)f_\alpha^2(0|X)\text{var}(Z|X)]^{-1}w(X) \times E[\tau_\alpha(X)w(X)]^{-1}.$$

This reveals that the optimal weighting function is $w(x) = f_\alpha^2(0|x)\text{var}(Z|X = x)$. The efficiency bound, therefore, can be attained by using a consistent estimator of $f_\alpha^2(0|x)\text{var}(Z|X = x)$ as the weighting function. The conditional density $f_\alpha(0|x)$ can be consistently estimated using estimated $U_\alpha$’s. On the other hand, if $U_\alpha$ 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.
6.2. Heteroskedastic $U_{\alpha}$

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

If $S_{\alpha}$ were known except for $\beta_{\alpha}$, 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 $S_{\alpha}$ 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 $\beta_{\alpha}$. 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 $\beta_{\alpha}$.

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 $S_{\alpha}$ 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 $R^{d_z}$, respectively. The trimming function $\tau(x, z)$ is introduced for the same reason as before. For a given $b$, a smoothed feasible score function $\hat{S}_{\alpha}(b)$ is then defined as

$$
\hat{S}_{\alpha}(b) = \tau(X_i, Z_i) \frac{\hat{f}_{\alpha}(0|X_i, Z_i)}{\alpha(1 - \alpha)} \left[ \alpha - 1 + J \left( \frac{Y_i - \hat{\hat{q}}_{\alpha}(X_i) - Z_i b}{j_n} \right) \right]
$$

$$
\times [Z_i - \hat{T}(X_i)],
$$
To complete the description of the one-step efficient estimator, we need to specify the nonparametric estimators of \( q_a(X_i), f_a(0|X_i, Z_i), \) and \( T(X_i) \). First of all, \( q_a(X_i) \) can be estimated by the first element of \( \hat{c} \) of (2). For \( f_a(0|X, Z) \), a standard kernel density estimator may be used. Observe that \( f_a(0|X, Z) \) can be written as \( f_a(0|X, Z) = f_1(0, x, z)/f_2(x, z) \), where \( f_1 \) and \( f_2 \) are joint densities of \((U_a, X, Z)\) and \((X, Z)\), respectively. This suggests that the conditional density of \( U_a \) 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 \( \hat{f}_a(0|X_i, Z_i) \) is defined as

\[
\hat{f}_a(0|X_i, Z_i) = \frac{(n\nu_1^{d_x+1}+d_z+1)^{-1} \sum_{j=1}^{n} \tau(X_j, Z_j) K_{\alpha x}(\hat{U}_{\alpha j}, \frac{X_j - X_i}{\nu_1}, \frac{Z_j - Z_i}{\nu_1})}{(n\nu_2^{d_x+1}+d_z+1)^{-1} \sum_{j=1}^{n} \tau(X_j, Z_j) K_{\alpha z}(\frac{X_j - X_i}{\nu_2}, \frac{Z_j - Z_i}{\nu_2})},
\]

where \( \hat{U}_{\alpha j} = Y_i - q_a(X_i) - Z_i \hat{\beta}_a \), \( K_{\alpha x} \) is a \((d_x + d_z + 1)\)-dimensional kernel function with a bandwidth \( \nu_1 \), and \( K_{\alpha z} \) is a \((d_x + d_z)\)-dimensional kernel function with a bandwidth \( \nu_2 \). Finally, \( T(X_i) \) can be estimated by

\[
\hat{T}(X_i) = \frac{\sum_{j=1}^{n} \tau(X_j, Z_j) \hat{f}_a^2(0|X_j, Z_j) Z_j K_x(\frac{X_j - X_i}{\gamma_n})}{\sum_{j=1}^{n} \tau(X_j, Z_j) \hat{f}_a^2(0|X_j, Z_j) K_x(\frac{X_j - X_i}{\gamma_n})},
\]

where \( K_x \) is a \( d_x \)-dimensional kernel function with a bandwidth \( \gamma_n \).

The following additional regularity conditions are useful to derive the asymptotic properties of the one-step estimator of \( \beta_a \).

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 \( \mathcal{X} \times \mathcal{Z} \subset \mathcal{W}_x \times \mathcal{W}_z \).

Assumption 10. The conditional density \( f_a(u|x, z) \) 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 \( J^{(i)}(v) = d^i J(v)/dv^i \).
(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) $\int_{-1}^1 J^{(1)}(v) \, dv = 1$, $\int_{-1}^1 v J^{(1)}(v) \, dv = 0$, and $\int_{-1}^1 v^2 J^{(1)}(v) \, dv > 0$.

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

Assumption 13.

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

where

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

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

$$\sqrt{n}(\hat{\beta}_n^* - \beta_\alpha) \to_d N(0, V_n),$$

where

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

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 $\tau$. Just like the case of homoskedastic $U_\alpha$, the effect of the trimming function is likely to be eliminated by letting the support of $\tau$ grow.

For statistical inference, it is necessary to obtain a consistent estimator of the variance $V_n$. There may be several ways to estimate this variance matrix, but a
simple consistent estimator of $V_*$ that is a by-product of the estimation procedure is

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

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

$$
\hat{V}_* = \alpha(1 - \alpha) \left\{ n^{-1} \sum_{i=1}^{n} \left[ \tau_i \hat{f}^2(0|X_i,Z_i) Z_i Z_i' - \frac{\hat{E}[\tau f^2_0(0|X,Z)|X_i]}{\hat{E}[\tau f^2_0(0|X,Z)|X_i]} \right] \right\}^{-1} ,
$$

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

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

where

$$
\Omega_i = \begin{pmatrix} 1 & Z_i' \\ Z_i & Z_i Z_i' \end{pmatrix} ,
\quad \hat{\tau}_x^2(x) = \frac{1(x \in \mathcal{X})}{n^{-1} \sum_{i=1}^{n} 1(X_i \in \mathcal{X})} ,
$$

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

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

where $X_i$ and $Z_i$ were drawn from a bivariate standard normal distribution with correlation 0.5 and $\varepsilon_i$ was drawn from the standard normal distribution that is
independent of $X$ and $Z$. Three different functions for $q$ and two different functions for $\sigma$ were simulated:

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

and

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

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 $\beta$ was set to be 1. The trimming function used in the experiments was $\tau_3(x) = 1(|x| \leq 2)$.10

Computing the AQR estimates requires choosing the order of polynomial $k$ and the bandwidth $\delta_n$ in (2). In the experiments $k = 3$. The asymptotic results of Section 3 only provide the range of $\delta_n$ in terms of the asymptotic order. A higher order asymptotic theory is required to obtain an asymptotically optimal rule-of-thumb bandwidth for the estimation of the nonparametric component $q_\alpha$. Let $h_n$ be the rule-of-thumb bandwidth for the estimation of $q$ when $k = 3$, suggested by Fan and Gijbels (1996, p. 202). Specifically, $h_n$ is of the form

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

where $w_0(\cdot)$ is a weight function, $\hat{q}_\alpha(\cdot)$ 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 $\alpha$th sample quantile of the residuals. Also, $\hat{q}_\alpha^{(4)}(\cdot)$ denotes the fourth derivative of $\hat{q}_\alpha(\cdot)$. 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 $h_n$ converges at rate $n^{-1/9}$, which is optimal for the estimation of $q$. Under-smoothing is required for the estimation of $\beta$. A simple bandwidth such as $\hat{\delta}_n = h_n \times n^{1/5} \times n^{-1/5}$ converges at rate $n^{-1/5}$ and satisfies Assumption 6 when $p_\alpha \geq 3$. 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 $\beta$. The asymptotic RMSEs were calculated (ignoring the effect of trimming) by the formula of the
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
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 \( f_a(0|X_i, Z_i) \). 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

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

The AQR estimates of \( \beta \) 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 \( \beta \). 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 \( \beta_\alpha \) can be
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 $\beta_\alpha$ do not depend on bandwidths, a higher order theory is required to choose opti-

Figure 2. Results of the Monte Carlo experiments for the one-step estimates of $\beta$. 

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 $\beta_\alpha$ do not depend on bandwidths, a higher order theory is required to choose opti-
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 \( \gamma \) in Assumption 4(a) and (b) and Assumption 8 do not have to be the same. For brevity, we assume that \( \gamma \) denotes the minimum of the three \( \gamma \)'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 \( \beta_n \). For example, \( \beta_n \) can be identified and estimated only using observations for which \( \Sigma(x) \) is nonsingular as long as \( \text{Prob}\{\Sigma(x) \text{ is nonsingular}\} \) is positive. Assumption 8 is adopted here to minimize the complexity of the proof.

8. In fact, any \( n^{-1/2} \)-consistent estimator of \( \beta_n \) 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, \( \hat{q}_n \) 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 \( |X_j - X_t| < \delta_n \) 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


PARTIALLY LINEAR QUANTILE REGRESSION 21


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


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 $U_{ai} = Y_i - q_\alpha(X_i) - Z_i\beta_\alpha$.

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

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

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

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

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

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

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 $\beta_\alpha$ 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

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

$$
= \frac{1}{n \Pr(X \in \mathcal{X})} \sum_{i=1}^n \tau_x(X_i)\{\hat{b}_\alpha(X_i) - \beta_\alpha\} + R_{\tau_x,n}
$$

$$
= T_n + R_{\tau_x,n}.
$$

First notice that $R_{\tau_x,n}$ results from the replacement of $n_{\tau_x}$ 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,
we need additional notation. Let \( b(\delta, X_j - X_i, Z_j) \) denote the \((s(A_k) + d_z)\)-dimensional vector

\[
\begin{align*}
  b(\delta, X_j - X_i, Z_j) &= (\{\delta^\alpha(X_j - X_i)\}^u, \{u \leq k\}', Z_j)' \\
  &\quad \text{(A.1)}
\end{align*}
\]

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

\[
G_n(X_i) = \int_{[-1,1]^d} \int_{\mathbb{W}_z} f_\alpha(0|X_i + \delta_x t, z) b(1, t, z) b(1, t, z)' dP_Z(z|X_i + \delta_x t) \\
\times g_{\delta_x}(t, X_i) \, dt,
\]

where

\[
g_{\delta_x}(t, X_i) = \frac{g(X_i + \delta_x t)}{\int_{[-1,1]^d} g(X_i + \delta_x t) \, dt}.
\]

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 \( e' = (0, I_{d_z}) \), where \( 0 \) 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 \( |X_j - X_i| \leq \delta_n \) for \( j \neq i \), \( j = 1, \ldots, n \), and let \( q_n^*(X_j, X_i) \) 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

\[
\hat{b}_n(X_i) - \beta_\alpha = e' \{N_n(X_i) G_n(X_i) \}^{-1}
\]

\[
= \sum_{j=1, j \neq i}^n b(\delta, X_j - X_i, Z_j) [\alpha - 1\{Y_j \leq q_n^*(X_j, X_i) + Z_j^\prime \beta_\alpha\}] \\
\times 1\{|X_j - X_i| \leq \delta_n\} + R_n(X_i),
\]

where

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

provided that \( 1/(2p_q + d) \leq \kappa < 1/(3d_z) \). It is worth noting that \( b(\delta, X_j - X_i, Z_j) \) 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 \( \hat{b}_n(X_i) - \beta_\alpha \) 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 \( 1/(2p_q) < \kappa < 1/(3d_z) \)), we have

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

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

\[
U_n = \sum_{1 \leq i < j \leq n} \xi_n(\xi_i, \xi_j),
\]
with $\xi_i = (Y_i, X_i, Z_i)$, $\xi_n(\xi_i, \xi_j) = \eta_n(\xi_i, \xi_j) + \eta_n(X_i, X_j, Z_i)$, $p_n(X) = \delta_n^d \int_{[-1,1]^d} g(X + \delta_n t) \, dt$, and

$$\eta_n(\xi_i, \xi_j) = \frac{1}{n} \tau_n(X_i) e^\prime\{np_n(X_i) G_n(X_i)\}^{-1} b(\delta_n, X_j - X_i, Z_i) \{ \alpha - 1(U_{ij} \leq 0) \} \times 1\{|X_j - X_i| \leq \delta_n\}.$$ 

Define $P_n$ to be the projection of $U_n$, so that

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

where

$$m_n(\xi_i) = E[\xi_n(\xi_i, \xi_j) | \xi_i] = \frac{1}{n^2} \{ \alpha - 1(U_{ii} \leq 0) \} \times E[\tau_n(X_i) e^\prime\{np_n(X_i) G_n(X_i)\}^{-1} b(\delta_n, X_i - X_j, Z_j) 1\{|X_i - X_j| \leq \delta_n\} | X_i, Z_i].$$

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

$$E(U_n - P_n)^2 = \frac{n(n - 1)}{2} (E\xi_n^2(\xi_1, \xi_2) - 2Em_n^2(\xi_1)) \leq \frac{n(n - 1)}{2} E\xi_n^2(\xi_1, \xi_2) \leq 2n(n - 1)E\eta_n^2(\xi_1, \xi_2).$$

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

$$E\eta_n^2(\xi_1, \xi_2) = O(1/n^4 \delta_n^{2d}).$$

This implies that

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

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

$$E(P_n - \bar{P}_n)^2 = o_p(n^{-1}), \quad (A.3)$$
where
\[ \hat{P}_n = \frac{n - 1}{n^2} \sum_{i=1}^{n} \{ \alpha - 1(U_{ai} \leq 0) \} \tau^*_a(X_i) \nu(X_i, Z_i) \]

with
\[ \nu(x, z) = e_z^* E \left[ f_a(0 \mid X, Z) \begin{pmatrix} 1 & Z' \\ Z & ZZ' \end{pmatrix} \mid X = x \right]^{-1} \begin{pmatrix} 1 \\ z \end{pmatrix}. \]

Conditioning on \((X_i, Z_i)\), we have
\[ E(P_n - \hat{P}_n)^2 = \frac{(n - 1)^2 \alpha(1 - \alpha)}{n^3} E \left[ \tau^*_a(X_i) e' \{ p_n(X_j) G_n(X_j) \}^{-1} b(\delta_n, X_i - X_j, Z_i) \right. \]
\[ \times \left. 1 \{ |X_j - X_i| \leq \delta_n \} \right] \nu(X_i, Z_i) \nu^2(X_i, Z_i) \right] \quad \text{(A.4)} \]

As in Lemma 4.2(b) of Chaudhuri et al. (1997), it can be shown that
\[ \delta_n^d \{ p_n(X_j) G_n(X_j) \}^{-1} \]
\[ = \left\{ g(X_i) \int_{[-1,1]^{d_1}} E[b(1, t, Z)b(1, t, Z)'f_a(0 \mid X, Z) \mid X = X_j] dt \right\}^{-1} + O_{L^2}(\delta_n^y), \]

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
\[ \int_{[-1,1]^{d_1}} \tau^*_a(X_i - \delta_n u)e' \left\{ \int_{[-1,1]^{d_1}} E[b(1, t, Z)b(1, t, Z)'f_a(0 \mid X, Z) \mid X = X_i - \delta_n u] dt \right\}^{-1} \]
\[ \times b(1, u, Z_i) du + O_{L^2}(\delta_n^y). \]

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
\[ P_n = \hat{P}_n + o_p(n^{-1/2}). \]

Therefore,
\[ \sqrt{n}(\hat{\beta}_a - \beta_a) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ \alpha - 1(U_{ai} \leq 0) \} \tau^*_a(X_i) \nu(X_i, Z_i) + o_p(1), \]

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

Proof of Theorem 5. Let \( I_u = V_{\nu}^{-1} \). Write
\[ \sqrt{n}(\hat{\beta}_a - \beta_a) = \sqrt{n}(\hat{\beta}_a - \beta_a) + \left[ -n^{-1} \sum_{i=1}^{n} \partial \hat{S}_{ai}(\hat{\beta}_a) / \partial b \right]^{-1} n^{-1/2} \sum_{i=1}^{n} \hat{S}_{ai}(\hat{\beta}_a). \]
Let

\[ S_{\alpha}(b) = \tau(X_i, Z_i) \frac{f_\alpha(0|X_i, Z_i)}{\alpha(1 - \alpha)} [\alpha - 1\{Y_i - q_\alpha(X_i) - Z_i'b \leq 0\}] [Z - T_s(X)]. \]

It will be shown subsequently that

\[ -n^{-1} \sum_{i=1}^{n} \frac{\partial \hat{S}_{\alpha}(\hat{\beta}_\alpha)}{\partial b} = I_s + o_p(1), \quad \text{and} \]

\[ n^{-1/2} \sum_{i=1}^{n} \hat{S}_{\alpha}(\hat{\beta}_\alpha) = n^{-1/2} \sum_{i=1}^{n} S_{\alpha}(\beta_\alpha) - I_s \sqrt{n} (\hat{\beta}_\alpha - \beta_\alpha) + o_p(1). \]

It then follows that

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

which gives the desired result immediately.

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

\[ \sup_{X_i \in X} |\hat{q}(X_i) - q_\alpha(X_i)| = o_p(n^{-1/3}). \]

In fact, we require that \( 1/(2p_n + d_\alpha) < \kappa < 1/(3d_\alpha) \) 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

\[ n^{-1} \sum_{i=1}^{n} \tau_s(X_i)[\hat{q}(X_i) - q_\alpha(X_i)] = O_p(n^{-1/2}). \]

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

\[ -n^{-1} \sum_{i=1}^{n} \frac{\partial \hat{S}_{\alpha}(\hat{\beta}_\alpha)}{\partial b} = \frac{1}{nj} \sum_{i=1}^{n} \frac{\tau_i \hat{f}_i}{\alpha(1 - \alpha)} \left( \frac{\hat{U}_{ai}}{j_n} \right) [Z_i - \hat{T}(X_i)] Z_i' \]

\[ = \frac{1}{nj} \sum_{i=1}^{n} \frac{\tau_i f_i}{\alpha(1 - \alpha)} \left( \frac{\hat{U}_{ai}}{j_n} \right) [Z_i - T_s(X_i)] Z_i' + o_p(1) \]

\[ = \frac{1}{nj} \sum_{i=1}^{n} \frac{\tau_i f_i}{\alpha(1 - \alpha)} \left( \frac{U_{ai}}{j_n} \right) [Z_i - T_s(X_i)] Z_i' + o_p(1) \]

\[ = I_s + o_p(1), \]

where the first equality comes from differentiation, the second equality follows in view of the uniform consistency of \( \hat{f}_a(0|X_i, Z_i) \) 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,
\[ \frac{1}{nj_n} \sum_{i=1}^{n} f^{(1)} \left( \frac{\hat{U}_{ai}}{j_n} \right) = \frac{1}{nj_n} \sum_{i=1}^{n} f^{(1)} \left( \frac{U_{ai}}{j_n} \right) + \frac{1}{nj_n} \sum_{i=1}^{n} f^{(2)} \left( \frac{U_{ai}}{j_n} \right)(\hat{U}_{ai} - U_{ai}) + R_{n,J}, \]  

where the remainder term \( R_{n,J} \) is 

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

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

\[ \frac{1}{nj_n^2} \sum_{i=1}^{n} f^{(2)} \left( \frac{U_{ai}}{j_n} \right) = O_p(1) \]

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 \( f^{(2)} \) is Lipschitz continuous to obtain 

\[ |R_{n,J}| \leq M_J \left[ \frac{1}{nj_n^3} \sum_{i=1}^{n} \tau_i(\hat{U}_{ai} - U_{ai}) \right] \sup_i \tau_i|\hat{U}_{ai} - U_{ai}| \]

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

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

\[ n^{-1/2} \sum_{i=1}^{n} \delta_{ai}(\hat{\beta}_a) = n^{-1/2} \sum_{i=1}^{n} \tau_i f_t \alpha(1 - \alpha) \left[ \alpha - 1 + J \left( \frac{Y_i - \hat{q}_a(X_i) - Z_i \hat{\beta}_a}{j_n} \right) \right] \times [Z_i - T_*(X_i)] [1 + o_p(1)] \]

\[ = n^{-1/2} \sum_{i=1}^{n} \delta_{ai}(\hat{\beta}_a) [1 + o_p(1)]. \]

Using a Taylor series expansion, 

\[ n^{-1/2} \sum_{i=1}^{n} \delta_{ai}(\hat{q}_a, \hat{\beta}_a) = n^{-1/2} \sum_{i=1}^{n} \delta_{ai}(q_a, \beta_a) + \left[ n^{-1} \sum_{i=1}^{n} \frac{\partial \delta_{ai}(q_a, \beta_a)}{\partial b} \right] \sqrt{n} (\hat{\beta}_a - \beta_a) \]

\[ + n^{-1/2} \sum_{i=1}^{n} \frac{\partial \delta_{ai}(q_a, \beta_a)}{\partial q} [\hat{q}_a(X_i) - q_a(X_i)] + R_{n,s}, \]

where \( R_{n,s} \) is a Taylor series remainder term. As before, the expression in the brackets of the second term is \( -I_s + o_p(1) \). 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,
\[ n^{-1/2} \sum_{i=1}^{n} \left[ f\left( \frac{U_{ai}}{\tilde{j}_n} \right) - 1(U_{ai} > 0) \right] = O_p(n^{1/2}j_n^2 + j_n^{1/2}). \]

This implies that using the restriction on \( j_n \),

\[ n^{-1/2} \sum_{i=1}^{n} \bar{S}_{ai}(q_{ai}, \beta_{ai}) = n^{-1/2} \sum_{i=1}^{n} S_{ai}(\beta_{ai}) + o_p(1). \]

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,

\[ n^{-1/2} \sum_{i=1}^{n} \frac{\delta \bar{S}_{ai}(q_{ai}, \beta_{ai})}{\delta q} = -n^{-1/2} \sum_{i=1}^{n} \frac{\tau_i f_i}{\alpha(1-\alpha)j_n} J_{11}^{-1} \left( \frac{U_{ai}}{j_n} \right) [Z_i - T_v(X_i)] \]

\[ = n^{1/2} O_p(j_n^2 + (nj_n)^{-1/2}) \]

by Chebyshev inequality. Hence,

\[ \left| n^{-1/2} \sum_{i=1}^{n} \frac{\delta \bar{S}_{ai}(q_{ai}, \beta_{ai})}{\delta q} [\hat{q}_{ai}(X_i) - q_{ai}(X_i)] \right| \]

\[ \leq n^{-1/2} \sum_{i=1}^{n} \frac{\delta \bar{S}_{ai}(q_{ai}, \beta_{ai})}{\delta q} \sup_{X_i \in \mathcal{X}} |\hat{q}_{ai}(X_i) - q_{ai}(X_i)| \]

\[ = o_p(1), \]

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

**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

\[ q_{ai}^n(x + \delta_n t, x) = \sum_{u \in A_k} c_{n,u}(x) t^u, \quad (A.10) \]

with the coefficients \( c_{n,u}(x) = (u!)^{-1} D^n u(x) \delta_n^{[u]} \), where \( u! = u_1! \ldots u_d! \). Let \( c_{ai}(x) = (c_{n,u}(x))_{u \in A_k} \) denote the \( s(A_k) \)-dimensional vector of Taylor coefficients in (A.10). In addition, let \( \hat{a}_{ai}(X_i) = (\hat{c}_{ai}(X_i), \hat{b}_{ai}(X_i))' \) denote the \( (s(A_k) + d_z)' \)-dimensional vector that minimizes (2) and let \( a_{ai}(X_i) = (c_{ai}(X_i), \beta_{ai})' \). 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.
LEMMA 1. Let Assumptions 1–5, 7, and 8 hold. Then we have

\[
\hat{a}_n(X_i) - a_n(X_i) = \{N_n(X_i)G_n(X_i)\}^{-1} \times \sum_{j=1, j \neq i}^{n} b(\delta_n, X_j - X_i, Z_j) [\alpha - 1\{Y_j \leq q_n^\alpha(X_j, X_i) + Z_j \beta_n]\]
\[
\times 1\{|X_j - X_i| \leq \delta_n\} + R_n(X_i),
\]

where the remainder term \(R_n(X_i)\) satisfies

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

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

**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

\[
\Pr(\lim \inf E_n) = 1,
\]

where

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

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

\[
\Pr(\lim \inf F_n) = 1,
\]

where

\[
F_n = \{|\hat{a}_n(X_i) - a_n(X_i)| \leq K_1 n^{-(1-\kappa d_x)/2} (\log n)^{1/2} \text{ for all } X_i \in \mathcal{X}\},
\]

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

\[
\tilde{H}_n(\delta_n, X_i) = \int_{[-1,1]^d} \int_{W_i} F_a[q^*_\alpha(X_i + \delta_n t, X_i) - q_\alpha(X_i + \delta_n t)|X_i + \delta_n t, z] \\
\times b(1, t, z) dP_Z(z|X_i + \delta_n t) g_{\delta_n}(t, X_i) dt,
\]

and

\[
R_n^{(1)}(X_i) = \tilde{H}_n(\delta_n, X_i) - \tilde{H}_n(\delta_n, X_i) - G_n(X_i)\{\hat{a}_\alpha(X_i) - a_\alpha(X_i)\}. \tag{A.13}
\]

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

\[
R_n^{(1)}(X_i) = (n^{-1 - \tau_d}) \log n
\]
a almost surely as \( n \to \infty \) uniformly in \( X_i \in W_i \).

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

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

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

\[
X_n(X_i) = (n^{1 - \tau d_e} / 4 \log n)^{3/4} \tag{A.14}
\]
a almost surely as \( n \to \infty \) uniformly in \( X_i \in W_i \). To exploit Bernstein’s inequality, it is assumed here that \( Z \) is bounded.

Step 5. Write

\[
[N_n(X_i)]^{-1} \sum_{j=1, j \neq i}^{n} b(\delta_n, X_j - X_i, Z_j)[\alpha - 1\{Y_j \leq q^*_\alpha(X_j, X_i) + Z_j \beta_\alpha\}1\{|X_j - X_i| \leq \delta_n\} \\
= [N_n(X_i)]^{-1} X_n(X_i) + \tilde{H}_n(\delta_n, X_i) - \tilde{H}_n(\delta_n, X_i) + R_n^{(2)}(X_i), \tag{A.15}
\]
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

$$R_n^{(2)}(X_i) \leq Mn^{-1-a_X}$$

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 $n \to \infty$. 

}$