Nonparametric Estimation of an Additive Quantile Regression Model

Joel L. Horowitz
Department of Economics
Northwestern University
Evanston, IL 60208, USA
joel-horowitz@northwestern.edu

and

Sokbae Lee *
Centre for Microdata Methods and Practice
Institute for Fiscal Studies
and
Department of Economics
University College London
London, WC1E 6BT, UK
l.simon@ucl.ac.uk

April 2004

Abstract

This paper is concerned with estimating the additive components of a nonparametric additive quantile regression model. We develop an estimator that is asymptotically normally distributed with a rate of convergence in probability of $n^{-r/(2r+1)}$ when the additive components are r-times continuously differentiable for some $r \geq 2$. This result holds regardless of the dimension of the covariates and, therefore, the new estimator has no curse of dimensionality. In addition, the estimator has an oracle property and is easily extended to a generalized additive quantile regression model with a link function. The numerical performance and usefulness of the estimator are illustrated by Monte Carlo experiments and an empirical example.

*We thank Andrew Chesher, Hidehiko Ichimura, and Roger Koenker for helpful comments and suggestions. Special thanks go to Andrew Chesher for encouraging us to work on this project. A preliminary version of the paper was presented at the Cemmap Workshop on Quantile Regression Methods and Applications in April, 2003. The research of Joel L. Horowitz was supported in part by NSF Grant SES-9910925 and this work was partially supported by ESRC Grant RES-000-22-0704.

1 Introduction

This paper is concerned with nonparametric estimation of the functions $m_{1,\alpha}, \ldots, m_{d,\alpha}$ in the quantile-regression model

$$Y = \mu_\alpha + m_{1,\alpha}(X^1) + \cdots + m_{d,\alpha}(X^d) + U_\alpha, \tag{1.1}$$

where Y is a real-valued dependent variable, X^j $(j = 1, \ldots, d)$ is the j'th component of random vector $X \in \mathbf{R}^d$ for some finite $d \geq 2$, μ_α is an unknown constant, $m_{1,\alpha}, \ldots, m_{d,\alpha}$ are unknown functions, and U_α is an unobserved random variable whose α-th quantile conditional on $X = x$ is zero for almost every x. Estimation is based on an *iid* random sample $\{(Y_i, X_i) : i = 1, \ldots, n\}$ of (Y, X). We describe an estimator of the additive components $m_{1,\alpha}, \ldots, m_{d,\alpha}$ that converges in probability pointwise at the rate $n^{-r/(2r+1)}$ when the $m_{j,\alpha}$ are r times continuously differentiable. In contrast to previous estimators, the rate of convergence of our estimators does not depend on the dimension of X, so asymptotically there is no curse of dimensionality. Moreover, our estimators have an oracle property. Specifically, the centered, scaled estimator of each additive component is asymptotically normally distributed with the same mean and variance that it would have if the other components were known. Finally, it is straightforward to extend our estimator to the generalized additive model

$$G(Y) = \mu_\alpha + m_{1,\alpha}(X^1) + \cdots + m_{d,\alpha}(X^d) + U_\alpha,$$

where G is a known, strictly increasing function.

To the best of our knowledge, there are three existing methods for estimating the model (1.1): spline, marginal integration, and backfitting estimators. Doksum and Koo (2000) consider a spline estimator, but they do not provide pointwise rates of convergence or an asymptotic distribution. It is not known whether the spline estimator can achieve the optimal pointwise rate. Huang (2003) discusses the difficulty of obtaining pointwise asymptotic normality with spline estimators in the context of additive nonparametric mean regression models. De Gooijer and Zerom (2003) develop a marginal integration estimator of (1.1), but their first step is an unrestricted, d-dimensional, nonparametric quantile regression. Consequently, their method suffers from the curse of dimensionality (see Remark 6 of De Gooijer and Zerom (2003)). Fan and Gijbels (1996, pp.296-297) propose a backfitting estimator of the model (1.1). Backfitting may avoid the curse of dimensionality because it uses only one-dimensional smoothers. However, the rate of convergence and other asymptotic distributional properties of the backfitting estimator are unknown. De Gooijer and Zerom (2003) carry out a simulation study to compare the finite sample performances of their marginal integration estimator and the backfitting method.

This paper presents an estimator of (1.1) that avoids the curse of dimensionality. It builds on Horowitz and Mammen (2002) (hereinafter HM), who developed an estimator of the additive components of a nonparametric additive mean regression model with a link function. The estimator of HM converges in probability pointwise at the rate $n^{-2/5}$ when the additive components are twice continuously differentiable, regardless of the dimension of X. Thus, the estimator has no curse of dimensionality. This paper extends the HM approach to additive quantile regression models. As in HM, we use a two-stage estimation procedure that does not require full-dimensional, unrestricted nonparametric estimation. In the first stage, the additive components are estimated by a series quantile-regression estimator that imposes the additive structure of (1.1). In the second stage, the estimator of each additive component is obtained by a one-dimensional local polynomial quantile regression in which the other components are replaced by their first-stage series estimates. Although the estimation method proposed here is similar in concept to that of HM, mean and quantile regressions are sufficiently different to make the extension of HM to quantile regressions non-trivial and to require a separate treatment.

The key to the ability of our estimator to avoid the curse of dimensionality is that by imposing additivity at the outset, the first-stage series estimator achieves a faster-converging bias than does a full-dimensional nonparametric estimator. Although the variance of the series estimator converges relatively slowly, the second estimation step creates an averaging effect that reduces variance, thereby achieving the optimal rate of convergence. The approach used here differs from typical two-stage estimation, which aims at estimating a single parameter by updating an initial consistent estimator. Here, there are several unknown functions, but we update the estimator of only one. We show that asymptotically, the estimation error of the other functions does not appear in the updated estimator of the function of interest.

The remainder of this paper is organized as follows. Section 2 provides an informal description of the two-stage estimator. Asymptotic properties of the estimator are given in Section 3. Section 4 reports the results of a Monte Carlo investigation of the finite sample properties of the estimator, and Section 5 applies the estimator to an empirical example. Concluding comments are in Section 6. We use the following notation. We let subscripts index observations of random variables and superscripts denote components of vectors. Thus, if X is a random vector, X_i is the i'th observation of X, X^j is the j'th component of X, and X_i^j is the i'th observation of the j'th component. We suppress the subscript α in the notation whenever this can be done without causing confusion.

2 Informal Description of the Estimator

This section describes a two-stage procedure for estimating $m_j(\cdot)$. For any $x \in \mathbf{R}^d$, define $m(x) = m_1(x^1) + \ldots + m_d(x^d)$, where x^j is the j-th component of x. We assume that the support of X is $\mathcal{X} \equiv [-1,1]^d$, and normalize $m_1, \ldots, m_d$ so that

$$\int_{-1}^{1} m_j(v)dv = 0$$

for $j = 1, \ldots, d$.

To describe the first-stage series estimator, let $\{p_k : k = 1, 2, \cdots\}$ denote a basis for smooth functions on $[-1,1]$. Conditions that the basis functions must satisfy are given in Section 3. For any positive integer κ, define

$$P_\kappa(x) = [1, p_1(x^1), \ldots, p_\kappa(x^1), \ldots, p_\kappa(x^2), \ldots, p_1(x^d), \ldots, p_\kappa(x^d)]'.$$

Then for $\theta_\kappa \in \mathbf{R}^{\kappa d+1}$, $P_\kappa(x)'\theta_\kappa$ is a series approximation to $\mu + m(x)$. To obtain asymptotic results, κ must satisfy certain conditions as $n \to \infty$. Upper and lower bounds on the number of terms κ are given in Section 3. For a random sample $\{(Y_i, X_i) : i = 1, \ldots, n\}$, let $\hat{\theta}_{n\kappa}$ be a solution to

$$\min_\theta S_{n\kappa}(\theta) \equiv n^{-1} \sum_{i=1}^{n} \rho_\alpha[Y_i - P_\kappa(X_i)'\theta], \tag{2.1}$$

where $\rho_\alpha(\cdot)$ is the check function such that $\rho_\alpha(u) = |u| + (2\alpha - 1)u$ for $0 < \alpha < 1$. The first-stage series estimator of $\mu + m(x)$ is defined as

$$\tilde{\mu} + \tilde{m}(x) = P_\kappa(x)'\hat{\theta}_{n\kappa},$$

where $\tilde{\mu}$ is the first component of $\hat{\theta}_{n\kappa}$. For any $j = 1, \ldots, d$ and any $x^j \in [-1,1]$, the series estimator $\tilde{m}_j(x^j)$ of $m_j(x^j)$ is the product of $[p_1(x^j), \cdots, p_\kappa(x^j)]$ with the appropriate components of $\hat{\theta}_{n\kappa}$. It can be seen that the same basis functions $\{p_1, \ldots, p_k\}$ are used to approximate $m_j(\cdot)$ and that no cross products are needed because of the additive form of (1.1).

To describe the second-stage estimator of (say) $m_1(x^1)$, define

$$m_{-1}(\tilde{X}_i) = m_2(\tilde{X}_i^2) + \cdots + m_d(\tilde{X}_i^d) \quad \text{and} \quad \tilde{m}_{-1}(\tilde{X}_i) = \tilde{m}_2(\tilde{X}_i^2) + \cdots + \tilde{m}_d(\tilde{X}_i^d),$$

where $\tilde{X}_i = (\tilde{X}_i^2, \ldots, \tilde{X}_i^d)$. Assume that m_1 is r-times continuously differentiable on $[-1, 1]$. Then the second-stage estimator of $m_1(x^1)$ is a $(r-1)$-th local polynomial estimator with $m_{-1}(\tilde{X}_i)$ replaced by the first-stage estimates $\tilde{m}_{-1}(\tilde{X}_i)$. Specifically, the estimator $\hat{m}_1(x^1)$ is defined as $\hat{m}_1(x^1) = \hat{b}_0$, where $\hat{b}_n = (\hat{b}_0, \hat{b}_1, \ldots, \hat{b}_{r-1})$ minimizes

(2.2)

$$S_n(b) \equiv (n\delta_n)^{-1} \sum_{i=1}^{n} \rho_\alpha \Big[Y_i - \tilde{\mu} - b_0 - \sum_{k=1}^{r-1} b_k \left[\delta_n^{-1}(X_i^1 - x^1) \right]^k - \tilde{m}_{-1}(\tilde{X}_i) \Big] K\left(\frac{x^1 - X_i^1}{\delta_n} \right),$$

K (kernel function) is a probability density function on $[-1, 1]$, and δ_n (bandwidth) is a sequence of real numbers converging to zero as $n \to \infty$. The regularity conditions for K and δ_n are given in Section 3. The second-stage estimators of $m_2(x^2), \ldots, m_d(x^d)$ are obtained similarly. Then the estimator of the regression surface is obtained by $\tilde{\mu} + \hat{m}_1(x^1) + \ldots + \hat{m}_d(x^d)$.

Since quantile regression is equivariant to monotone transformations of Y, it is straightforward to extend the estimator of (1.1) to a generalized additive model that has the form

$$G(Y) = \mu + m_1(X^1) + \cdots + m_d(X^d) + U, \tag{2.3}$$

where G is a known, strictly increasing function, and the α-th quantile of U conditional $X = x$ is zero for almost every x. The estimator of the α-th quantile of Y conditional $X = x$ can be easily obtained by $G^{-1}[\tilde{\mu} + \hat{m}_1(x^1) + \cdots + \hat{m}_d(x^d)]$, where $\tilde{\mu} + \hat{m}_1(x^1) + \cdots + \hat{m}_d(x^d)$ is obtained by the estimation procedure described above with $G(Y_i)$ being substituted for Y_i.

We end this section by mentioning computational aspects of the estimation procedure. Both the first-stage and second-stage minimization problems, (2.1) and (2.2) are linear programming problems and therefore can be solved easily by using computation methods developed for linear quantile regression methods. Moreover, the new estimator does not require iterations (backfitting approach) or n first-stage estimates (marginal integration method).

3 Asymptotic Results

This section gives asymptotic results for the estimator described in Section 2. We need some additional notation. For any matrix A, let $\|A\| = [\text{trace}(A'A)]^{1/2}$ be the Euclidean norm. Let $d(\kappa) = \kappa d + 1$ and $b_{\kappa 0}(x) = \mu + m(x) - P_\kappa(x)'\theta_{\kappa 0}$. Define $\zeta_\kappa = \sup_{x \in \mathcal{X}} \|P_\kappa(x)\|$ and $\Phi_\kappa = E[f(0|X)P_\kappa(X)P_\kappa(X)']$.

To establish asymptotic results, we need the following conditions.

Assumption 3.1. *The data $\{(Y_i, X_i) : i = 1, \ldots, n\}$ are i.i.d. and the α-th quantile of Y conditional on $X = x$ is $\mu + m(x)$ for almost every $x \in \mathcal{X}$.*

Assumption 3.2. *The support of X is $\mathcal{X} \equiv [-1, 1]^d$. The distribution of X is absolutely continuous with respect to Lebesgue measure. The probability density function of X (denoted by $f_X(x)$) is bounded, bounded away from zero, and is twice continuously differentiable on $\mathcal{X}$.*

Assumption 3.3. *Let $F(u|x)$ be the distribution function of U_α conditional on $X = x$. Assume that $F(0|x) = \alpha$ for almost every $x \in \mathcal{X}$ and that $F(\cdot|x)$ has a probability density function $f(\cdot|x)$. There is a constant $L_f < \infty$ such that $|f(u_1|x) - f(u_2|x)| \leq L_f|u_1 - u_2|$ for all u_1 and u_2 in a neighborhood of zero and for all $x \in \mathcal{X}$. Also, there are constants $c_f > 0$ and $C_f < \infty$ such that $c_f \leq f(u|x) \leq C_f$ for all u in a neighborhood of zero and for all $x \in \mathcal{X}$.*

Assumption 3.4. *For each j, $m_j(\cdot)$ is r-times continuously differentiable on $[-1, 1]$ for some $r \geq 2$.*

Assumption 3.5. *The smallest eigenvalue of Φ_κ is bounded away from zero for all κ, and the largest eigenvalue of Φ_κ is bounded for all κ.*

Assumption 3.6. *The basis functions $\{p_k : k = 1, 2, \ldots\}$ satisfy the following conditions:*
(a) each p_k is continuous,
(b) $\int_{-1}^{1} p_k(v)dv = 0$,

$$(c) \quad \int_{-1}^{1} p_j(v)p_k(v)dv = \begin{cases} 1 & \text{if } j = k \\ 0 & \text{otherwise,} \end{cases}$$

(d) as $\kappa \to \infty$,

$$\zeta_\kappa = O(\kappa^{1/2}), \tag{3.1}$$

(e) there are vectors $\theta_{\kappa 0} \in \mathbf{R}^{d(\kappa)}$ such that

$$\sup_{x \in \mathcal{X}} |\mu + m(x) - P_\kappa(x)'\theta_{\kappa 0}| = O(\kappa^{-r}) \tag{3.2}$$

as $\kappa \to \infty$.

Assumption 3.7. $(\kappa^4/n)(\log n)^2 \to 0$ *and κ^{1+2r}/n is bounded.*

It is necessary to make some comments on regularity conditions. Assumption 3.1 defines a data generating process. If necessary, the bounded support condition in Assumption 3.2 can be satisfied by carrying out a monotone transformation of X. Among other things, Assumption 3.3 requires that $f(\cdot|x)$ be bounded away from zero in a neighborhood of zero uniformly over x. This is a convenient condition to establish asymptotic results. Without this condition, the rate of convergence and limiting distribution would not be the same. See, for example, Knight (1998) for asymptotic results for linear median regression estimators under more general assumptions on $f(u|x)$. As in Newey (1997) and HM, Assumption 3.5 insures the non-singularity of the covariance matrix of the asymptotic form of the first-stage estimator. Assumption 3.6 imposes restrictions on the basis functions. When f_X is bounded away from zero and m_j is continuously r-times differentiable, the conditions in Assumption 3.6 are satisfied by B-splines. Due to the additive structure of (1.1), the uniform approximation error in (3.2) is of order $O(\kappa^{-r})$ regardless of the dimension of X. This helps the second-stage estimator avoid the curse of dimensionality.

The following theorem gives a uniform convergence result for the first-stage series estimator.

Theorem 3.1. *Let Assumptions 3.1-3.7 hold. Then as $n \to \infty$,*

$$(a) \left\| \hat{\theta}_{n\kappa} - \theta_{\kappa 0} \right\| = O_p\left[(\kappa/n)^{1/2} + \kappa^{-r} \right],$$

and

$$(b) \sup_{x \in \mathcal{X}} |\tilde{m}(x) - m(x)| = O_p\left[\kappa/(n^{1/2}) + \kappa^{1/2-r} \right].$$

The uniform convergence rate for the first-stage estimator is not optimal. However, it appears (to our best knowledge) that this is a first uniform convergence result for series estimators of additive quantile regression models with $d \geq 2$. He and Shi (1994, 1996) obtained L_2 rates of convergence for B-spline estimators of univariate and bivariate quantile regression models and Portnoy (1997) and He and Portnoy (2000) derived local asymptotic properties of smoothing splines for $d \leq 2$.

The following theorem establishes a Bahadur-type expansion of the first-stage estimator.

Theorem 3.2. *Let Assumptions 3.1-3.7 hold. Then as $n \to \infty$,*

$$\hat{\theta}_{n\kappa} - \theta_{\kappa 0} = n^{-1}\Phi_\kappa^{-1}\sum_{i=1}^{n} P_\kappa(X_i)\Big\{\alpha - 1\big[U_i \leq 0\big]\Big\} + n^{-1}\Phi_\kappa^{-1}\sum_{i=1}^{n} f(0|X_i)P_\kappa(X_i)b_{\kappa 0}(X_i) + R_n,$$

where the remainder term R_n satisfies

$$\|R_n\| = O_p\Big[(\kappa^2/n)^{3/4}(\log n)^{1/2} + \kappa^{3/2}/n\Big] + o_p(n^{-1/2}).$$

To state asymptotic results for the second-stage estimator, we need additional assumptions.

Assumption 3.8. *(a) $\kappa = C_\kappa n^\nu$ for some constant C_κ satisfying $0 < C_\kappa < \infty$ and some ν satisfying*

$$\frac{1}{2r+1} < \nu < \frac{2r+3}{12r+6}.$$

(b) $\delta_n = C_\delta n^{-1/(2r+1)}$ for some constant C_δ satisfying $0 < C_\delta < \infty$.

Assumption 3.9. *The function K is a bounded, continuous probability density function on $[-1, 1]$ and is symmetrical about 0.*

Assumption 3.8 (a) requires that $r \geq 2$. Assumption 3.8 (b) and Assumption 3.9 are standard in the nonparametric estimation literature.

Define

$$\bar{P}_\kappa(\tilde{x}) = [1, 0, \ldots, 0, p_1(x^2), \ldots, p_\kappa(x^2), \ldots, p_1(x^d), \ldots, p_\kappa(x^d)]'.$$

The following condition is used to establish the limiting distribution of the two-stage estimator.

Assumption 3.10. *The largest eigenvalue of $E[\bar{P}_\kappa(\tilde{X})\bar{P}_\kappa(\tilde{X})'|X^1 = x^1]$ is bounded for all κ and each component of $E[\bar{P}_\kappa(\tilde{X})\bar{P}_\kappa(\tilde{X})'|X^1 = x^1]$ is twice continuously differentiable with respect to x^1.*

Let $\mu_j = \int_{-1}^{1} u^j K(u)\, du$ denote the moments of K and let $S(K)$ be the $(r \times r)$ matrix, whose (i,j) component is μ_{i+j-2}. Also, let $e_1 = (1, 0, \ldots, 0)$ be the unit column vector. As in Ruppert and Wand (1994) and Fan and Gijbels (1996, pp. 63-66), let $K_*(u) = e_1' S(K)^{-1}(1, u, \ldots, u^{r-1})' K(u)$ be the equivalent kernel. K_* is a kernel of order r if r is even. Let $f_{X^1}(x^1)$ denote the probability density function of X^1, $f_1(u|x^1)$ the probability density function of U_α conditional on $X^1 = x^1$, and $D^k m_j(x^j)$ the k-th order derivative of m_j. The main result of the paper is as follows:

Theorem 3.3. *Let Assumptions 3.1-3.10 hold. Also, assume that r is even, where r is defined in Assumption 3.4. Then as $n \to \infty$, for any x^1 satisfying $|x^1| \leq 1 - \delta_n$,*

(a) $|\hat{m}_1(x^1) - m_1(x^1)| = O_p\left[n^{-r/(2r+1)}\right]$.

(b) $n^{r/(2r+1)}[\hat{m}_1(x^1) - m_1(x^1)] \to_d \mathbf{N}[B(x^1), V(x^1)]$,

where

$$B(x^1) = \left\{\int_{-1}^{1} u^r K_*(u)\, du\right\} \{r!\}^{-1} C_\delta^r D^r m_1(x^1) \quad \textit{and}$$

$$V(x^1) = \left\{\int_{-1}^{1} [K_*(u)]^2\, du\right\} C_\delta^{-1} \alpha(1-\alpha)/\{f_{X^1}(x^1)[f_1(0|x^1)^2]\}.$$

(c) If $j \neq 1$, then $n^{r/(2r+1)}[\hat{m}_1(x^1) - m_1(x^1)]$ and $n^{r/(2r+1)}[\hat{m}_j(x^j) - m_j(x^j)]$ are asymptotically independently normally distributed for any x^j satisfying $|x^j| \leq 1 - \delta_n$.

The theorem implies that the second-stage estimator achieves the optimal rate and has the same asymptotic distribution that it would have if $m_2, \ldots, m_d$ were known. Because of Assumption 3.8 (a), it is required that m_j be at least twice continuously differentiable. This required differentiability is independent of the dimension of X and, therefore, our estimator avoids the curse of dimensionality.

If r were odd, that is even $(r-1)$ polynomial fits were used in (2.2), then the asymptotic bias would depend on $D^r f_{X^1}(x^1)$ and therefore would not be design adaptive. See, for example, Ruppert and Wand (1994) and Fan and Gijbels (1996, pp. 61-63) for this issue in detail.

4 Monte Carlo Experiments

This section reports the results of a small set of Monte Carlo experiments that compare the numerical performance of the two-stage estimator with the marginal integration method.

Experiments were carried out with $d = 2$ and $d = 5$. The experiments with $d = 2$ were carried out with the design identical to that of De Gooijer and Zerom (2003, Section 4). Specifically, the experiments consist of estimating m_1 and m_2 in

$$Y = m_1(X^1) + m_2(X^2) + 0.25\varepsilon, \tag{4.1}$$

where $m_1(x^1) = 0.75x^1$ and $m_2(x^2) = 1.5\sin(0.5\pi x^2)$. The covariates X^1 and X^2 are bivariate normal with mean zero, unit variance, and correlation ρ. We consider $\alpha = 0.5$ and sample sizes $n = 100$ and 200. As in De Gooijer and Zerom (2003), experiments were carried out with $\rho = 0.2$ (low correlation between covariates) and $\rho = 0.8$ (high correlation).

B-splines were used for the first-stage of the two-stage estimator with $\kappa_n = 4$ and local linear fitting was used for the second-stage. Also, the kernel K is taken to be the normal density function. The bandwidth $\delta_{1n} = 3\hat{\sigma}_{X^1}n^{-1/5}$ was chosen for estimation of m_1 and $\delta_{2n} = \hat{\sigma}_{X^2}n^{-1/5}$ was for estimation of m_2, where $\hat{\sigma}_{X^j}$ is the sample standard deviation of X^j for $j = 1, 2$. The normal density function does not satisfy the finite support condition in Assumption 3.9, but these kernel and bandwidths were chosen to be identical to those in De Gooijer and Zerom (2003) in order to compare the finite-sample performance of the two-stage estimator vis-à-vis those of the marginal integration method and the backfitting approach reported in De Gooijer and Zerom (2003).

To see whether the two-stage estimator avoids the curse of dimensionality in finite-samples, three additional covariates were added to (4.1). More specifically, the experiments with $d = 5$ consist of estimating m_1 and m_2 in

$$Y = m_1(X^1) + m_2(X^2) + X^3 + X^4 + X^5 + 0.25\varepsilon,$$

where X^j are independently distributed as $U[-1, 1]$ for $j = 3, 4, 5$. Since the local linear fitting is used, the second stage estimator has the rate of convergence $n^{-2/5}$ regardless of d.

All the experiments were carried out in R using libraries 'splines' (to generate B-spline basis) and 'quantreg' (to solve (2.1) and (2.2)). The R language is available as free software at http://www.r-project.org. There were 100 Monte Carlo replications per experiment and the absolute deviation error (ADE) was computed for each estimated function on the interval $[-2, 2]$. An average of the ADE's (AADE) was the criterion used in De Gooijer and Zerom (2003).

Table 1 shows the AADE values for the marginal integration and two-stage estimators for combinations of d, ρ and n. We computed the pilot estimator of the marginal integration estimator directly through the check function method, whereas De Gooijer and Zerom

(2003) obtained the pilot estimator by inverting the conditional distribution function. As in De Gooijer and Zerom (2003), local linear approximation is adopted in the direction of interest and local constant approximation is used in the nuisance directions. The asymptotic distribution of the marginal integration estimator is identical regardless of the choice between two alternative pilot estimators. As was reported by De Gooijer and Zerom (2003), the marginal integration estimator performs poorly when there is high correlation among covariates. When $d = 2$ and $\rho = 0.8$, the finite-sample performance of the two-stage estimator is considerably better than that of the mariginal integration estimator. Also, the performance of the two-stage estimator is comparable between $\rho = 0.2$ and $\rho = 0.8$ when $d = 2$. That is also the case between $d = 2$ and $d = 5$ for both $\rho = 0.2$ and $\rho = 0.8$. These are consistent with the asymptotic results established in Section 3 because the limiting distribution of the two-stage estimator does not depend on d or ρ. However, the marginal integration estimator performs very poorly when $d = 5$ and $\rho = 0.8$. In that case, the AADE's for the marginal integration estimator are more than twice as large as those for the two-stage estimator.

In summary, the results of experiments suggest that the two-stage estimator outperforms the marginal integration estimator when there is high correlation among covariates and/or the dimension of covariates is relatively large. Furthermore, the results indicate that the two-stage estimator performs well with high-dimensional covariates.

5 An Empirical Example

Yafeh and Yosha (2003) used a sample of Japanese firms in the chemical industry to examine whether 'concentrated shareholding is associated with lower expenditure on activities with scope for managerial private benefits'. In this section, we concentrate on only one of regressions considered by Yafeh and Yosha (2003). The dependent variable Y is general sales and administrative expenses deflated by sales (denoted by MH5 in Yafeh and Yosha (2003)). This measure is one of five measures of expenditures on activities with scope for managerial private benefits considered by Yafeh and Yosha (2003). The covariates include a measure of ownership concentration (denoted by TOPTEN, cumulative shareholding by the largest ten shareholders), and firm characteristics: the log of assets, firm age, and leverage (the ratio of debt to debt plus equity). The regression model is

$$\text{MH5} = \beta_0 + \beta_1 \text{TOPTEN} + \beta_2 \log(\text{Assets}) + \beta_3 \text{Age} + \beta_4 \text{Leverage} + U. \tag{5.1}$$

The sample size is 185. This dataset is available at the Economic Journal web site at http://www.res.org.uk.

We estimated the additive conditional median function using the two-stage estimator. Estimation results for other conditional quantile functions are available on request. Before estimation begins, the covariates are standardized to have mean zero and variance 1. B-splines were used for the first-stage with $\kappa_n = 3$ and local linear fitting was used for the second-stage with the bandwidth $h_n = 1.25$. The kernel K is taken to be the normal density function. Varying h_n from 1 to 1.5 did not change estimation results significantly. Figure 1 summarizes estimation results. Each panel shows the estimated function of interest and 90% symmetric pointwise confidence interval (without bias correction). The confidence interval was obtained using asymptotic approximation based on Theorem 3.3. The unknown components of the asymptotic variance in Theorem 3.3 were estimated by kernel density estimators.

It can be seen that the effects of ownership concentration (TOPTEN) are nonlinear. This suggests that the relationship between MH5 and TOPTEN conditional on firm characteristics cannot be well described by a linear location-shift model. The effects of firm size (log(Assets)) are also highly nonlinear. This may be due to the fact that MH5 includes expenditures that are not related with managerial private benefits. The effects of firm age are mostly negligible, compared to effects of other covariates. The effects of leverage are linear. The estimation results suggest that the linear model is misspecified. To verify this, we used the test of linearity for median regression models in Horowitz and Spokoiny (2002). The test gives a test statistic of 2.192 with a 5% critical value of 1.999. Thus, the test rejects the linear median regression model at the 5% level.

In summary, our estimation results confirm the qualitative conclusion of Yafeh and Yosha (2003) and indicate that a model that is a more flexible than linear median regression models is needed to study the relationship between concentrated shareholding and expenditures for managerial private benefits.

6 Conclusions

This paper has developed an estimator of the additive components of a nonparametric additive quantile regression model. It is shown that the estimator converges in probability of $n^{-2/(2r+1)}$ when the unknown functions are r-times continuously differentiable for some $r \geq 2$. This result holds regardless of the dimension of the covariates. In addition, the estimator has an oracle property. Specifically, the estimator of each additive component has the same asymptotic distribution that it would have if the other components were known. Finally, the estimator described here is easily extended to a generalized additive quantile regression model with a known link function.

A Appendix: Proofs

Throughout the Appendix, let C denote a generic constant that may be different in different uses. Let $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote minimum and maximum eigenvalues of a symmetric matrix A.

A.1 Proofs of Theorems 3.1 and 3.2

It is useful to introduce some additional notation that is used in Koenker and Bassett (1978) and Chaudhuri (1991). Let $\mathcal{N} = \{1, \dots, n\}$ and $\mathcal{H}_\kappa$ denote the collection of all $d(\kappa)$-element subsets of $\mathcal{N}$. Also, let $B(h)$ denote the submatrix (subvector) of a matrix (vector) B with rows (components) that are indexed by the elements of $h \in \mathcal{H}_\kappa$. In particular, let $\mathbf{P}_\kappa(h)$ denote a $d(\kappa) \times d(\kappa)$ matrix, whose rows are the vectors $P_\kappa(X_i)'$ such that $i \in h$, and let $\mathbf{Y}_\kappa(h)$ denote a $d(\kappa) \times 1$ vector, whose elements are Y_i such that $i \in h$. Let $\mathbf{P}_\kappa$ denote a $n \times d(\kappa)$ matrix, whose rows are the vectors $P_\kappa(X_i)'$ for $i = 1, \dots, n$.

The following lemmas are useful to prove Theorems 3.1 and 3.2. The first lemma is from Koenker and Bassett (1978, Theorem 3.1).

Lemma A.1. *Suppose that $\mathbf{P}_\kappa$ has rank $= d(\kappa)$. Then there is a subset $h_\kappa \in \mathcal{H}_\kappa$ such that the problem* (2.1) *has at least one solution of the form $\hat{\theta}_{n\kappa} = \mathbf{P}_\kappa(h_\kappa)^{-1}\mathbf{Y}_\kappa(h_\kappa)$.*

Define

$$\overline{H}_{n1\kappa}(\hat{\theta}_{n\kappa}) = \sum_{i=1, i \in h_\kappa^c}^{n} \Big\{\alpha - 1\big[Y_i \le P_\kappa(X_i)'\hat{\theta}_{n\kappa}\big]\Big\} P_\kappa(X_i)'\mathbf{P}_\kappa(h_\kappa)^{-1}.$$

Lemma A.2. *$\hat{\theta}_{n\kappa} = \mathbf{P}_\kappa(h_\kappa)^{-1}\mathbf{Y}_\kappa(h_\kappa)$ is a unique solution to* (2.1) *almost surely for all sufficiently large n.*

Proof. The matrix $\mathbf{P}_\kappa$ has rank $= d(\kappa)$ almost surely for all sufficiently large n. By Lemma A.1, there exists a $h_\kappa \in \mathcal{H}_\kappa$ such that the problem (2.1) has at least one solution of the form $\hat{\theta}_{n\kappa} = \mathbf{P}_\kappa(h_\kappa)^{-1}\mathbf{Y}_\kappa(h_\kappa)$.

As in Theorem 3.3 in Koenker and Bassett (1978) [see also Fact 6.4 in Chaudhuri (1991)], $\hat{\theta}_{n\kappa} = \mathbf{P}_\kappa(h_\kappa)^{-1}\mathbf{Y}_\kappa(h_\kappa)$ is a unique solution to (2.1) if and only if each component in $\overline{H}_{n1\kappa}(\hat{\theta}_{n\kappa})$ is strictly between $\alpha - 1$ and α, i.e. $\overline{H}_{n1\kappa}(\hat{\theta}_{n\kappa}) \in (\alpha-1, \alpha)^{d(\kappa)}$. Also, if $\hat{\theta}_{n\kappa} = \mathbf{P}_\kappa(h_\kappa)^{-1}\mathbf{Y}_\kappa(h_\kappa)$ is a solution to (2.1), then $\overline{H}_{n1\kappa}(\hat{\theta}_{n\kappa}) \in [\alpha-1, \alpha]^{d(\kappa)}$.

Since the distribution of $P_\kappa(X_i)$ is absolutely continuous with respect to Lebesgue measure (except for the first component), the probability that $\overline{H}_{n1\kappa}(\hat{\theta}_{n\kappa})$ lies on the boundary of the cube $[\alpha-1, \alpha]^{d(\kappa)}$ is zero for all sufficiently large n. Therefore, $\hat{\theta}_{n\kappa} = \mathbf{P}_\kappa(h_\kappa)^{-1}\mathbf{Y}_\kappa(h_\kappa)$ is a unique solution almost surely for all sufficiently large n. □

Let $\Phi_{n\kappa} = n^{-1}\sum_{i=1}^n f(0|X_i)P_\kappa(X_i)P_\kappa(X_i)'$. Let 1_n be the indicator function such that $1_n = 1\{\lambda_{\min}(\Phi_{n\kappa}) \geq \lambda_{\min}(\Phi_\kappa)/2\}$. As in the proofs of Theorem 1 of Newey (1997) and Lemma 4 of HM, one can show that $\|\Phi_{n\kappa} - \Phi_\kappa\|^2 = O_p(\kappa^2/n) = o_p(1)$. Thus, $\Pr(1_n = 1) \to 1$ as $n \to \infty$.

Define

$$
\begin{aligned}
G_{n\kappa}(\theta) &= n^{-1}\Phi_{n\kappa}^{-1}\sum_{i=1}^n \Big\{\alpha - 1\big[U_i \leq P_\kappa(X_i)'(\theta - \theta_{\kappa 0}) - b_{\kappa 0}(X_i)\big]\Big\}P_\kappa(X_i), \\
G^*_{n\kappa}(\theta) &= n^{-1}\Phi_{n\kappa}^{-1}\sum_{i=1}^n \Big\{\alpha - F\big[P_\kappa(X_i)'(\theta - \theta_{\kappa 0}) - b_{\kappa 0}(X_i)\big|X_i\big]\Big\}P_\kappa(X_i),
\end{aligned}
$$

and $\tilde{G}_{n\kappa}(\theta) = G_{n\kappa}(\theta) - G^*_{n\kappa}(\theta)$.

Lemma A.3. *As $n \to \infty$,*

$$
1_n \left\|\tilde{G}_{n\kappa}(\theta_{\kappa 0})\right\| = O_p\left[(\kappa/n)^{1/2}\right].
$$

Proof. Notice that since the data are i.i.d., $f(\cdot|x)$ is bounded away from zero in a neighborhood of zero for all x (in particular, $f(0|x) \geq c_f$ for all x), and the smallest eigenvalue of $\Phi_{n\kappa}$ is bounded away from zero (when $1_n = 1$),

$$
\begin{aligned}
&E\left[1_n \left\|\tilde{G}_{n\kappa}(\theta_{\kappa 0})\right\|^2 \Big| X_1, \ldots, X_n\right] \\
&\leq 1_n n^{-2}\sum_{i=1}^n \left[E\Big[\Big\{F\big[-b_{\kappa 0}(X_i)\big|X_i\big] - 1\big[U \leq -b_{\kappa 0}(X_i)\big]\Big\}^2\Big|X_i\Big]P_\kappa(X_i)'\Phi_{n\kappa}^{-2}P_\kappa(X_i)\right] \\
&\leq C1_n n^{-2}\sum_{i=1}^n \mathrm{trace}\Big[P_\kappa(X_i)'\Phi_{n\kappa}^{-2}P_\kappa(X_i)\Big] \\
&\leq C1_n n^{-2}\sum_{i=1}^n c_f^{-1}f(0|X_i)\mathrm{trace}\Big[P_\kappa(X_i)'\Phi_{n\kappa}^{-2}P_\kappa(X_i)\Big]
\end{aligned}
$$

$$
\begin{aligned}
&\leq C 1_n n^{-1} \mathrm{trace}\Big\{ \Phi_{n\kappa}^{-2} \Big\{ n^{-1} \sum_{i=1}^{n} f(0|X_i) P_\kappa(X_i) P_\kappa(X_i)' \Big] \Big\} \\
&= C 1_n n^{-1} \mathrm{trace}\Big(\Phi_{n\kappa}^{-1} \Big) \\
&\leq C n^{-1} d(\kappa).
\end{aligned}
$$

Therefore, the lemma follows from Markov's inequality. □

Lemma A.4. *As $n \to \infty$,*

$$
1_n \left\| G_{n\kappa}(\hat{\theta}_{n\kappa}) \right\| \leq C \zeta_\kappa \kappa / n \quad \textit{almost surely.}
$$

Proof. By Lemma A.2, there is a unique index set $h_\kappa \in \mathcal{H}_\kappa$ such that $\hat{\theta}_{n\kappa} = \mathbf{P}_\kappa(h_\kappa)^{-1} \mathbf{Y}_\kappa(h_\kappa)$ almost surely for all sufficiently large n. Now write $G_{n\kappa}(\hat{\theta}_{n\kappa}) = G_{n\kappa 1}(\hat{\theta}_{n\kappa}) + G_{n\kappa 2}(\hat{\theta}_{n\kappa})$, where

$$
G_{n\kappa 1}(\hat{\theta}_{n\kappa}) = n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1, i \in h_\kappa}^{n} \Big\{ \alpha - 1\big[U_i \leq P_\kappa(X_i)'(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) - b_{\kappa 0}(X_i) \big] \Big\} P_\kappa(X_i),
$$

and

$$
G_{n\kappa 2}(\hat{\theta}_{n\kappa}) = n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1, i \in h_\kappa^c}^{n} \Big\{ \alpha - 1\big[U_i \leq P_\kappa(X_i)'(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) - b_{\kappa 0}(X_i) \big] \Big\} P_\kappa(X_i).
$$

Notice that $\max_{1 \leq i \leq n} 1_n \left\| \Phi_{n\kappa}^{-1} P_\kappa(X_i) \right\| \leq C \max_{1 \leq i \leq n} \| P_\kappa(X_i) \| = C \zeta_\kappa$ for some constant $C < \infty$ since the smallest eigenvalue of $\Phi_{n\kappa}$ is bounded away from zero (when $1_n = 1$). Thus, we have $1_n \left\| G_{n\kappa 1}(\hat{\theta}_{n\kappa}) \right\| \leq C \zeta_\kappa d(\kappa)/n$.

Now consider $G_{n\kappa 2}(\hat{\theta}_{n\kappa})$. Notice that $G_{n\kappa 2}(\hat{\theta}_{n\kappa})' = n^{-1} \overline{H}_{n1\kappa}(\hat{\theta}_{n\kappa}) \mathbf{P}_\kappa(h_\kappa) \Phi_{n\kappa}^{-1}$. As was explained in the proof of Lemma A.2, each component in $\overline{H}_{n1\kappa}(\hat{\theta}_{n\kappa})$ is between $\alpha - 1$ and α. Thus, $\left\| \overline{H}_{n1\kappa}(\hat{\theta}_{n\kappa}) \right\| \leq d(\kappa)^{1/2}$. Since the smallest eigenvalue of $\Phi_{n\kappa}$ is bounded away from zero (when $1_n = 1$), we can find a constant $C < \infty$ (independent of κ) such that

$$
1_n \left\| \mathbf{P}_\kappa(h_\kappa) \Phi_{n\kappa}^{-1} \right\| \leq C \left\| \mathbf{P}_\kappa(h_\kappa) \right\|.
$$

Also notice that

$$
\begin{aligned}
\left\| \mathbf{P}_\kappa(h_\kappa) \right\|^2 &= \mathrm{trace}[\mathbf{P}_\kappa(h_\kappa)' \mathbf{P}_\kappa(h_\kappa)] = \mathrm{trace}[\mathbf{P}_\kappa(h_\kappa) \mathbf{P}_\kappa(h_\kappa)'] \\
&= \sum_{i \in h_\kappa} \left\| P_\kappa(X_i) \right\|^2 \leq \zeta_\kappa^2 d(\kappa).
\end{aligned}
$$

Hence, $\|\mathbf{P}_\kappa(h_\kappa)\| \leq \zeta_\kappa d(\kappa)^{1/2}$. Therefore,

$$1_n \left\| G_{n\kappa 2}(\hat{\theta}_{n\kappa}) \right\| \leq n^{-1} \left\| \overline{H}_{n1\kappa}(\hat{\theta}_{n\kappa}) \right\| 1_n \left\| \mathbf{P}_\kappa(h_\kappa) \Phi_{n\kappa}^{-1} \right\| \leq C \zeta_\kappa d(\kappa)/n.$$

Since arguments used in this proof hold uniformly over h_κ, the lemma follows immediately.

$\square$

The next lemma is based on the elegant argument of Welsh (1989).

Lemma A.5. *As $n \to \infty$,*

$$\sup_{\|\theta - \theta_{\kappa 0}\| \leq C(\kappa/n)^{1/2}} 1_n \left\| \tilde{G}_{n\kappa}(\theta) - \tilde{G}_{n\kappa}(\theta_{\kappa 0}) \right\| = O_p \left[d(\kappa)^{1/2} \zeta_\kappa^{1/2} (d(\kappa)/n)^{3/4} (\log n)^{1/2} \right].$$

Proof. Let $B_n = \{\theta : \|\theta - \theta_{\kappa 0}\| \leq C(d(\kappa)/n)^{1/2}\}$. As in the proof of Theorem 3.1 of Welsh (1989), cover the ball B_n with cubes $\mathcal{C} = \{\mathcal{C}(\theta_l)\}$, where $\mathcal{C}(\theta_l)$ is a cube containing $(\theta_l - \theta_{\kappa 0})$ with sides of $C(d(\kappa)/n^5)^{1/2}$ such that $\theta_l \in B_n$. Then the number of the cubes covering the ball B_n is $L = (2n^2)^{d(\kappa)}$. Also, we have that $\|(\theta - \theta_{\kappa 0}) - (\theta_l - \theta_{\kappa 0})\| \leq C(d(\kappa)/n^{5/2})$ for any $(\theta - \theta_{\kappa 0}) \in \mathcal{C}(\theta_l)$, where $l = 1, \cdots, L$.

First note that

$$\begin{aligned} &\sup_{\theta \in B_n} 1_n \left\| \tilde{G}_{n\kappa}(\theta) - \tilde{G}_{n\kappa}(\theta_{\kappa 0}) \right\| \\ &\leq \max_{1 \leq l \leq L} \sup_{(\theta - \theta_{\kappa 0}) \in \mathcal{C}(\theta_l)} 1_n \left\| \tilde{G}_{n\kappa}(\theta) - \tilde{G}_{n\kappa}(\theta_l) \right\| + \max_{1 \leq l \leq L} 1_n \left\| \tilde{G}_{n\kappa}(\theta_l) - \tilde{G}_{n\kappa}(\theta_{\kappa 0}) \right\|. \end{aligned} \tag{A.1}$$

Define $\gamma_n = C(d(\kappa)/n^{5/2})$. Now using the fact that $1[u \leq \cdot]$ and $F[\cdot|x]$ are monotone increasing functions, we have

$$\begin{aligned} &\sup_{(\theta - \theta_{\kappa 0}) \in \mathcal{C}(\theta_l)} 1_n \left\| \tilde{G}_{n\kappa}(\theta) - \tilde{G}_{n\kappa}(\theta_l) \right\| \\ &\leq \sup_{(\theta - \theta_{\kappa 0}) \in \mathcal{C}(\theta_l)} n^{-1} \sum_{i=1}^n 1_n \left\| \Phi_{n\kappa}^{-1} P_\kappa(X_i) \right\| \\ &\times \Big| \Big\{ 1\big[U_i \leq P_\kappa(X_i)'(\theta - \theta_{\kappa 0}) - b_{\kappa 0}(X_i)\big] - F\big[P_\kappa(X_i)'(\theta - \theta_{\kappa 0}) - b_{\kappa 0}(X_i) \big| X_i\big] \Big\} \\ &- \Big\{ 1\big[U_i \leq P_\kappa(X_i)'(\theta_l - \theta_{\kappa 0}) - b_{\kappa 0}(X_i)\big] - F\big[P_\kappa(X_i)'(\theta_l - \theta_{\kappa 0}) - b_{\kappa 0}(X_i) \big| X_i\big] \Big\} \Big| \\ &\leq n^{-1} \sum_{i=1}^n 1_n \left\| \Phi_{n\kappa}^{-1} P_\kappa(X_i) \right\| \Big[\Big\{ 1\big[U_i \leq P_\kappa(X_i)'(\theta_l - \theta_{\kappa 0}) - b_{\kappa 0}(X_i) + \|P_\kappa(X_i)\| \gamma_n\big] \\ &- F\big[P_\kappa(X_i)'(\theta_l - \theta_{\kappa 0}) - b_{\kappa 0}(X_i) - \|P_\kappa(X_i)\| \gamma_n \big| X_i\big] \Big\} \end{aligned}$$

$$
\begin{aligned}
&- \Big\{ 1\big[U_i \le P_\kappa(X_i)'(\theta_l - \theta_{\kappa 0}) - b_{\kappa 0}(X_i)\big] - F\big[P_\kappa(X_i)'(\theta_l - \theta_{\kappa 0}) - b_{\kappa 0}(X_i)\big|X_i\big]\Big\}\Big] \\
&\le \Big| n^{-1}\sum_{i=1}^{n} 1_n \left\|\Phi_{n\kappa}^{-1}P_\kappa(X_i)\right\| \Big[\Big\{1\big[U_i \le P_\kappa(X_i)'(\theta_l - \theta_{\kappa 0}) - b_{\kappa 0}(X_i) + \|P_\kappa(X_i)\|\,\gamma_n\big] \\
&- F\big[P_\kappa(X_i)'(\theta_l - \theta_{\kappa 0}) - b_{\kappa 0}(X_i) + \|P_\kappa(X_i)\|\,\gamma_n\big|X_i\big]\Big\} \\
&- \Big\{ 1\big[U_i \le P_\kappa(X_i)'(\theta_l - \theta_{\kappa 0}) - b_{\kappa 0}(X_i)\big] - F\big[P_\kappa(X_i)'(\theta_l - \theta_{\kappa 0}) - b_{\kappa 0}(X_i)\big|X_i\big]\Big\}\Big]\Big| \\
&+ n^{-1}\sum_{i=1}^{n} 1_n \left\|\Phi_{n\kappa}^{-1}P_\kappa(X_i)\right\| \Big\{F\big[P_\kappa(X_i)'(\theta_l - \theta_{\kappa 0}) - b_{\kappa 0}(X_i) + \|P_\kappa(X_i)\|\,\gamma_n\big|X_i\big] \\
&- F\big[P_\kappa(X_i)'(\theta_l - \theta_{\kappa 0}) - b_{\kappa 0}(X_i) - \|P_\kappa(X_i)\|\,\gamma_n\big|X_i\big]\Big\}.
\end{aligned}
\tag{A.2}
$$

Consider the second term in (A.2). By Assumption 3.3,

$$
\begin{aligned}
\max_{1\le l\le L} n^{-1}\sum_{i=1}^{n} & 1_n \left\|\Phi_{n\kappa}^{-1}P_\kappa(X_i)\right\| \Big\{F\big[P_\kappa(X_i)'(\theta_l - \theta_{\kappa 0}) - b_{\kappa 0}(X_i) + \|P_\kappa(X_i)\|\,\gamma_n\big|X_i\big] \\
&- F\big[P_\kappa(X_i)'(\theta_l - \theta_{\kappa 0}) - b_{\kappa 0}(X_i) - \|P_\kappa(X_i)\|\,\gamma_n\big|X_i\big]\Big\} \\
&\le C\gamma_n \max_{1\le i\le n} 1_n \left\|\Phi_{n\kappa}^{-1}P_\kappa(X_i)\right\| \|P_\kappa(X_i)\| \\
&\le C(d(\kappa)/n^{5/2})\zeta_\kappa^2.
\end{aligned}
\tag{A.3}
$$

Now consider the second term in (A.1), that is $\max_{1\le l\le L} 1_n \left\|\tilde{G}_{n\kappa}(\theta_l) - \tilde{G}_{n\kappa}(\theta_{\kappa 0})\right\|$. Let $\Delta^{(j)}_{\tilde{G}_{n\kappa}}(\theta_l)$ denote the j-th element of $[\tilde{G}_{n\kappa}(\theta_l) - \tilde{G}_{n\kappa}(\theta_{\kappa 0})]$. Then we have

$$
1_n\Delta^{(j)}_{\tilde{G}_{n\kappa}}(\theta_l) = 1_n e_{(j)}'[\tilde{G}_{n\kappa}(\theta_l) - \tilde{G}_{n\kappa}(\theta_{\kappa 0})],
$$

where $e_{(j)}$ is a unit vector whose components are all zero except for the j-th component being one. Notice that conditional on $\{X_1,\ldots,X_n\}$, the summands in $1_n\Delta^{(j)}_{\tilde{G}_{n\kappa}}(\theta_l)$ are independently distributed with mean 0 and that the summands in $1_n\Delta^{(j)}_{\tilde{G}_{n\kappa}}(\theta_l)$ are bounded uniformly (over j and l) by $n^{-1}C\zeta_\kappa$ for all sufficiently large n. Furthermore, the variance of $1_n\Delta^{(j)}_{\tilde{G}_{n\kappa}}(\theta_l)$ conditional on $\{X_1,\ldots,X_n\}$ is bounded by $Cn^{-2}\sum_{i=1}^{n} 1_n|e_{(j)}'\Phi_{n\kappa}^{-1}P_\kappa(X_i)|^2|P_\kappa(X_i)'(\theta_l - \theta_{\kappa 0})|$. Notice that using the fact that $f(0|x)$ is bounded away from zero (that is, $f(0|x)\ge c_f$ for all x) and that the smallest eigenvalue of $\Phi_{n\kappa}^{-1}$ is bounded away from zero (when $1_n = 1$)

for all κ,

$$
\begin{aligned}
&n^{-1}\sum_{i=1}^{n}\left|e_{(j)}'\Phi_{n\kappa}^{-1}P_\kappa(X_i)\right|^2 |P_\kappa(X_i)'(\theta_l-\theta_{\kappa 0})| \\
&\leq n^{-1}\sum_{i=1}^{n} c_f^{-1} f(0|X_i)\left|e_{(j)}'\Phi_{n\kappa}^{-1}P_\kappa(X_i)\right|^2 |P_\kappa(X_i)'(\theta_l-\theta_{\kappa 0})| \\
&\leq C\max_{1\leq i\leq n}\left|P_\kappa(X_i)'(\theta_l-\theta_{\kappa 0})\right| e_{(j)}'\Phi_{n\kappa}^{-1}\left[n^{-1}\sum_{i=1}^{n} f(0|X_i)P_\kappa(X_i)P_\kappa(X_i)'\right]\Phi_{n\kappa}^{-1}e_{(j)} \\
&\leq C\zeta_\kappa (d(\kappa)/n)^{1/2}\lambda_{\max}(\Phi_{n\kappa}^{-1}) \\
&\leq C\zeta_\kappa (d(\kappa)/n)^{1/2}
\end{aligned}
$$

uniformly (over j and l) for all sufficiently large n. Therefore, the conditional variance of $1_n\Delta^{(j)}_{\tilde{G}_{n\kappa}}(\theta_l)$ is bounded uniformly (over j and l) by $n^{-1}C\zeta_\kappa(d(\kappa)/n)^{1/2}$ for all sufficiently large n. Let $\varepsilon_n = d(\kappa)^{1/2}\zeta_\kappa^{1/2}(d(\kappa)/n)^{3/4}(\log n)^{1/2}$. An application of Bernstein's inequality (see, for example, van der Vaart and Wellner (1996, p.102)) to the sum $\Delta^{(j)}_{\tilde{G}_{n\kappa}}(\theta_l)$ gives

$$
\begin{aligned}
&\Pr\Big(\max_{1\leq l\leq L} 1_n\left\|\tilde{G}_{n\kappa}(\theta_l)-\tilde{G}_{n\kappa}(\theta_{\kappa 0})\right\| > C\varepsilon_n \Big| X_1,\ldots,X_n\Big) \\
&\leq \sum_{l=1}^{L}\Pr\Big(1_n\left\|\tilde{G}_{n\kappa}(\theta_l)-\tilde{G}_{n\kappa}(\theta_{\kappa 0})\right\| > C\varepsilon_n \Big| X_1,\ldots,X_n\Big) \\
&\leq \sum_{l=1}^{L}\sum_{j=1}^{d(\kappa)}\Pr\Big(1_n\left|\Delta^{(j)}_{\tilde{G}_{n\kappa}}(\theta_l)\right| > C\varepsilon_n d(\kappa)^{-1/2}\Big| X_1,\ldots,X_n\Big) \\
&\leq 2\,(2n^2)^{d(\kappa)}d(\kappa)\exp\left[-C\frac{\varepsilon_n^2 d(\kappa)^{-1}}{n^{-1}\zeta_\kappa(d(\kappa)/n)^{1/2}+n^{-1}\zeta_\kappa\varepsilon_n d(\kappa)^{-1/2}}\right] \\
&\leq C\exp\left[2d(\kappa)\log n + \log d(\kappa) - Cd(\kappa)\log n\right] \\
&\leq C\exp\left[-Cd(\kappa)\log n\right] \qquad \text{(A.4)}
\end{aligned}
$$

for all sufficiently large n. In particular, it is required here that $\zeta_\kappa = O(\kappa^{1/2})$ and $(\kappa^3/n)(\log n)^2 \to 0$.

Now consider the first term in (A.2). Let $\tilde{T}_{n\kappa}(\theta_l)$ denote the expression inside $|\cdot|$ in the first term in (A.2). Notice that conditional on $\{X_1,\ldots,X_n\}$, the summands in $\tilde{T}_n(\theta_l)$ are independently distributed with mean 0 and with range bounded by $n^{-1}C\zeta_\kappa$ and that the variance of the summands in $\tilde{T}_{n\kappa}(\theta_l)$ conditional on $\{X_1,\ldots,X_n\}$ is bounded by $n^{-1}C\zeta_\kappa^3\gamma_n = n^{-1}C(d(\kappa)/n^{5/2})\zeta_\kappa^3$ uniformly over l for all sufficiently large n. Another application of Bernstein's inequality to $\tilde{T}_n(\theta_l)$ gives

$$
\begin{aligned}
\Pr(\max_{1\le l\le L} |\tilde{T}_{n\kappa}(\theta_l)| > C\varepsilon_n | X_1,\ldots,X_n) &\le \sum_{l=1}^{L} \Pr(|\tilde{T}_{n\kappa}(\theta_l)| > C\varepsilon_n | X_1,\ldots,X_n) \\
&\le 2\,(2n^2)^{d(\kappa)} \exp\left[-C\frac{\varepsilon_n^2}{n^{-1}(d(\kappa)/n^{5/2})\zeta_\kappa^3 + n^{-1}\zeta_\kappa\varepsilon_n}\right] \\
&\le 2\,(2n^2)^{d(\kappa)} \exp\left[-Cn\varepsilon_n/\zeta_\kappa\right] \\
&\le C\exp\left[2d(\kappa)\log n + \log d(\kappa) - Cd(\kappa)n^{1/4}(\log n)^{1/2}\right] \qquad \text{(A.5)}
\end{aligned}
$$

for all sufficiently large n. Now the lemma follows by combining (A.3), (A.4), and (A.5). □

Lemma A.6. *As $n \to \infty$,*

$$1_n G^*_{n\kappa}(\theta) = -1_n(\theta - \theta_{\kappa 0}) + 1_n n^{-1}\Phi_{n\kappa}^{-1}\sum_{i=1}^{n} f(0|X_i)P_\kappa(X_i)b_{\kappa 0}(X_i) + R^*_{n\kappa},$$

where $\|R^*_{n\kappa}\| = O\left[\zeta_\kappa \|\theta - \theta_{\kappa 0}\|^2 + \zeta_\kappa \kappa^{-2r}\right]$.

Proof. Define

$$1_n \tilde{G}^*_{n\kappa}(\theta) = -1_n(\theta - \theta_{\kappa 0}) + 1_n n^{-1}\Phi_{n\kappa}^{-1}\sum_{i=1}^{n} f(0|X_i)P_\kappa(X_i)b_{\kappa 0}(X_i).$$

Using a first-order Taylor series expansion, Assumptions 3.3 and 3.5, and equation 3.2, we have

$$
\begin{aligned}
1_n \left\|G^*_{n\kappa}(\theta) - \tilde{G}^*_{n\kappa}(\theta)\right\| &\le C \max_{1\le i\le n} 1_n \left\|\Phi_{n\kappa}^{-1}P_\kappa(X_i)\right\| \left[n^{-1}\sum_{i=1}^{n}\left\{P_\kappa(X_i)'(\theta - \theta_{\kappa 0}) - b_{\kappa 0}(X_i)\right\}^2\right] \\
&\le C\zeta_\kappa\left\{(\theta-\theta_{\kappa 0})'\Phi_{n\kappa}(\theta-\theta_{\kappa 0}) + \max_{1\le i\le n} b_{\kappa 0}(X_i)^2\right\} \\
&\le C\zeta_\kappa\lambda_{\max}(\Phi_{n\kappa})(\theta-\theta_{\kappa 0})'(\theta-\theta_{\kappa 0}) + C\zeta_\kappa \max_{1\le i\le n} b_{\kappa 0}(X_i)^2 \\
&\le O\left[\zeta_\kappa \|\theta - \theta_{\kappa 0}\|^2\right] + O(\zeta_\kappa\kappa^{-2r}) \qquad \text{(A.6)}
\end{aligned}
$$

for all sufficiently large n, which proves the lemma. □

Lemma A.7. *As $\kappa \to \infty$,*

$$1_n \left\| n^{-1}\Phi_{n\kappa}^{-1}\sum_{i=1}^{n} f(0|X_i)P_\kappa(X_i)b_{\kappa 0}(X_i)\right\| = O(\kappa^{-r}).$$

Proof. Let $\bar{B}_\kappa$ be a $(n \times 1)$ vector whose elements are $f(0|X_i)^{1/2} b_{\kappa 0}(X_i)$ and $\bar{\mathbf{P}}_\kappa$ be a $(n \times d(\kappa))$ matrix whose rows are $f(0|X_i)^{1/2} P_\kappa(X_i)'$. Then $\Phi_{n\kappa} = \bar{\mathbf{P}}'_\kappa \bar{\mathbf{P}}_\kappa / n$ and $\sum_{i=1}^n f(0|X_i) P_\kappa(X_i) b_{\kappa 0}(X_i) = \bar{\mathbf{P}}'_\kappa \bar{B}_\kappa$. Therefore, using the fact that $\bar{\mathbf{P}}_\kappa (\bar{\mathbf{P}}'_\kappa \bar{\mathbf{P}}_\kappa)^{-1} \bar{\mathbf{P}}'_\kappa$ is idempotent (so that its largest eigenvalue is just one),

$$
\begin{aligned}
& 1_n \left\| n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1}^n f(0|X_i) P_\kappa(X_i) b_{\kappa 0}(X_i) \right\|^2 \\
&= 1_n \left\| \Phi_{n\kappa}^{-1} \bar{\mathbf{P}}'_\kappa \bar{B}_\kappa / n \right\|^2 \\
&= 1_n n^{-2} \bar{B}'_\kappa \bar{\mathbf{P}}_\kappa \Phi_{n\kappa}^{-2} \bar{\mathbf{P}}'_\kappa \bar{B}_\kappa \\
&\leq 1_n n^{-2} \lambda_{\max}(\Phi_{n\kappa}^{-1}) \bar{B}'_\kappa \bar{\mathbf{P}}_\kappa (\bar{\mathbf{P}}'_\kappa \bar{\mathbf{P}}_\kappa / n)^{-1} \bar{\mathbf{P}}'_\kappa \bar{B}_\kappa \\
&\leq 1_n n^{-1} \lambda_{\max}(\Phi_{n\kappa}^{-1}) \lambda_{\max}[\bar{\mathbf{P}}_\kappa (\bar{\mathbf{P}}'_\kappa \bar{\mathbf{P}}_\kappa)^{-1} \bar{\mathbf{P}}'_\kappa] \left\| \bar{B}_\kappa \right\|^2 \\
&\leq C \max_{1 \leq i \leq n} b_{\kappa 0}(X_i)^2
\end{aligned}
$$

for all sufficiently large n. The lemma now follows from equation (3.2). □

Proof of Theorem 3.1. Write

$$
1_n G_{n\kappa}(\hat{\theta}_{n\kappa}) = 1_n \tilde{G}_{n\kappa}(\theta_{\kappa 0}) + 1_n[\tilde{G}_{n\kappa}(\hat{\theta}_{n\kappa}) - \tilde{G}_{n\kappa}(\theta_{\kappa 0})] + 1_n G^*_{n\kappa}(\hat{\theta}_{n\kappa}). \tag{A.7}
$$

By Lemma A.6, (A.7) can be rewritten as

$$
\begin{aligned}
1_n(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) &= -1_n G_{n\kappa}(\hat{\theta}_{n\kappa}) + 1_n \tilde{G}_{n\kappa}(\theta_{\kappa 0}) + 1_n[\tilde{G}_{n\kappa}(\hat{\theta}_{n\kappa}) - \tilde{G}_{n\kappa}(\theta_{\kappa 0})] \\
&\quad + 1_n n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1}^n f(0|X_i) P_\kappa(X_i) b_{\kappa 0}(X_i) + R^*_{n\kappa}.
\end{aligned} \tag{A.8}
$$

To prove part (a), suppose that $\left\| \hat{\theta}_{n\kappa} - \theta_{\kappa 0} \right\| \leq C[(\kappa/n)^{1/2} + \kappa^{-r}]$ for any constant $C > 0$. Then, by applying Lemmas A.3 - A.7 to equation (A.8), we have

$$
\begin{aligned}
1_n \left\| \hat{\theta}_{n\kappa} - \theta_{\kappa 0} \right\| &\leq 1_n \left\| G_{n\kappa}(\hat{\theta}_{n\kappa}) \right\| + 1_n \left\| \tilde{G}_{n\kappa}(\theta_{\kappa 0}) \right\| + 1_n \left\| \tilde{G}_{n\kappa}(\hat{\theta}_{n\kappa}) - \tilde{G}_{n\kappa}(\theta_{\kappa 0}) \right\| \\
&\quad + 1_n \left\| n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1}^n f(0|X_i) P_\kappa(X_i) b_{\kappa 0}(X_i) \right\| + \| R^*_{n\kappa} \| \\
&\leq O_p\left[(\kappa/n)^{1/2} \right] + O_p\left(\kappa^{3/2}/n \right) + O_p\left[(\kappa^2/n)^{3/4} (\log n)^{1/2} \right] + O_p(\kappa^{-r}).
\end{aligned} \tag{A.9}
$$

Therefore, the right-hand side of (A.9) is less than $C[(\kappa/n)^{1/2} + \kappa^{-r}]$ with probability approaching one (w.p.a.1) provided that $(\kappa^4/n)(\log n)^2 \to 0$ and κ^{1+2r}/n is bounded. This implies that w.p.a.1,

$$
1_n \left\| \hat{\theta}_{n\kappa} - \theta_{\kappa 0} \right\| \leq C[(\kappa/n)^{1/2} + \kappa^{-r}],
$$

which in turn implies that w.p.a.1,

$$\left\|\hat{\theta}_{n\kappa} - \theta_{\kappa 0}\right\| \leq C[(\kappa/n)^{1/2} + \kappa^{-r}]$$

since $\Pr(1_n = 1) \to 1$ as $n \to \infty$. Therefore, part (a) of Theorem 3.1 is proved. Part (b) follows by combining part (a) with $\zeta_\kappa = O(\kappa^{1/2})$. □

Proof of Theorem 3.2. By applying Lemmas A.4 - A.6, and Theorem 3.1 (a) to (A.8), we have

$$1_n(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) = 1_n \tilde{G}_{n\kappa}(\theta_{\kappa 0}) + 1_n n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1}^{n} f(0|X_i) P_\kappa(X_i) b_{\kappa 0}(X_i) + R_n,$$

where the remainder term R_n satisfies

$$\|R_n\| = O_p\Big[(\kappa^2/n)^{3/4}(\log n)^{1/2} + \kappa^{3/2}/n\Big].$$

Define

$$\bar{G}_{n\kappa}(\theta_{\kappa 0}) = n^{-1} \Phi_{n\kappa}^{-1} \sum_{i=1}^{n} \Big\{\alpha - 1\big[U_i \leq 0\big]\Big\} P_\kappa(X_i).$$

By using arguments similar to those used in the proof of Lemma A.3, we have

$$E\left[1_n \left\|\tilde{G}_{n\kappa}(\theta_{\kappa 0}) - \bar{G}_{n\kappa}(\theta_{\kappa 0})\right\|^2 \Big| X_1, \ldots, X_n\right] \leq C n^{-1} d(\kappa) \sup_{x \in \mathcal{X}} |b_{\kappa 0}(x)|.$$

Hence,

$$1_n \left\|\tilde{G}_{n\kappa}(\theta_{\kappa 0}) - \bar{G}_{n\kappa}(\theta_{\kappa 0})\right\| = o_p(n^{-1/2})$$

by Markov's inequality. The theorem now follows from the fact that $\|\Phi_{n\kappa} - \Phi_\kappa\| = O_p(\kappa^2/n) = o_p(1)$ and $\Pr(1_n = 1) \to 1$ as $n \to \infty$. □

A.2 Proof of Theorem 3.3

We need additional notation to prove Theorem 3.3. Recall that $D^k m_j(x^j)$ denotes the k-th order derivative of m_j. Define

$$\begin{aligned}
\beta_n(x^1) &= \Big[m_1(x^1), \delta_n D^1 m_1(x^1), \ldots, \delta_n^{r-1} D^{(r-1)} m_1(x^1)\{(r-1)!\}^{-1}\Big]', \\
Z_{ni}(x^1) &= \Big[1, \delta_n^{-1}(X_i^1 - x^1), \ldots, \{\delta_n^{-1}(X_i^1 - x^1)\}^{(r-1)}\Big]', \\
K_{ni}(x^1) &= K\left(\frac{x^1 - X_i^1}{\delta_n}\right), \quad \text{and} \\
B_i(x^1) &= m_1(X_i^1) - m_1(x^1) - \sum_{k=1}^{(r-1)} \{k!\}^{-1} D^k m_1(x^1)(X_i^1 - x^1).
\end{aligned}$$

To simplify the notation, dependence on x^1 of $\beta_n(x^1), Z_{ni}(x^1), K_{ni}(x^1)$, and $B_i(x^1)$ will be suppressed throughout the proof (when there is no confusion). For example, $B_i = m_1(X_i^1) - Z_{ni}'\beta_n$. Also, define

$$\bar{b}_{\kappa 0}(\tilde{x}) = \mu + m_{-1}(\tilde{x}) - \bar{P}_\kappa(\tilde{x})'\theta_{\kappa 0}.$$

where $\bar{P}_\kappa(\tilde{x})$ is defined in the main text. Recall that

$$\bar{P}_\kappa(\tilde{x}) = [1, 0, \ldots, 0, p_1(x^2), \ldots, p_\kappa(x^2), \ldots, p_1(x^d), \ldots, p_\kappa(x^d)]'.$$

Then $\tilde{\mu} + \tilde{m}_{-1}(\tilde{X}_i) = \bar{P}_\kappa(\tilde{X}_i)'\hat{\theta}_{n\kappa}$ and $[\tilde{\mu} - \mu] + [\tilde{m}_{-1}(\tilde{X}_i) - m_{-1}(\tilde{X}_i)] = \bar{P}_\kappa(\tilde{X}_i)'(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) - \bar{b}_{\kappa 0}(\tilde{X}_i)$. Finally, define

$$\begin{aligned}
G_n(b, x^1) &= (n\delta_n)^{-1} \sum_{i=1}^n \Big\{\alpha - 1\big[U_i \le Z_{ni}'(b - \beta_n) - B_i\big]\Big\} Z_{ni} K_{ni}, \\
\tilde{G}_n(b, x^1) &= (n\delta_n)^{-1} \sum_{i=1}^n \Big\{\alpha - 1\big[U_i \le Z_{ni}'(b - \beta_n) - B_i + \bar{P}_\kappa(\tilde{X}_i)'(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) - \bar{b}_{\kappa 0}(\tilde{X}_i)\big]\Big\} Z_{ni} K_{ni}, \\
G_n^*(b, x^1) &= (n\delta_n)^{-1} \sum_{i=1}^n \Big\{\alpha - F\big[Z_{ni}'(b - \beta_n) - B_i \big| X_i\big]\Big\} Z_{ni} K_{ni}, \\
\tilde{G}_n^*(b, x^1) &= (n\delta_n)^{-1} \sum_{i=1}^n \Big\{\alpha - F\big[Z_{ni}'(b - \beta_n) - B_i + \bar{P}_\kappa(\tilde{X}_i)'(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) - \bar{b}_{\kappa 0}(\tilde{X}_i) \big| X_i\big]\Big\} Z_{ni} K_{ni},
\end{aligned}$$

$\Delta_{G_n}(b, x^1) = G_n(b, x^1) - G_n^*(b, x^1)$, and $\Delta_{\tilde{G}_n}(b, x^1) = \tilde{G}_n(b, x^1) - \tilde{G}_n^*(b, x^1)$.

The following lemmas are useful to prove Theorem 3.3.

Lemma A.8. *As $n \to \infty$, for any x^1 such that $|x^1| \le 1 - \delta_n$,*

$$\left\|\tilde{G}_n(\hat{b}_n, x^1)\right\| = O\left[(n\delta_n)^{-1}\right] \quad \textit{almost surely.}$$

Proof. Notice that the minimization problem (2.2) is just a kernel-weighted linear quantile regression problem and therefore, it has a linear programming representation. Also, notice that each component of Z_{ni} is bounded by one whenever K_{ni} is nonzero. Then the lemma can be proved by using arguments identical to those used in the proof of Lemma A.4. □

Lemma A.9. *As $n \to \infty$, for any x^1 such that $|x^1| \le 1 - \delta_n$,*

$$\left\|\Delta_{G_n}(\beta_n, x^1)\right\| = O_p\left[(n\delta_n)^{-1/2}\right].$$

Proof. Notice that the mean of $\Delta_{G_n}(\beta_n, x^1)$ is zero. Then the lemma follows by calculating $E[\|\Delta_{G_n}(\beta_n, x^1)\|^2]$ and then applying Markov's inequality. □

Lemma A.10. *As $n \to \infty$, for any x^1 such that $|x^1| \le 1 - \delta_n$,*

$$\sup_{\|b-\beta_n\| \le C(n\delta_n)^{-1/2}} \left\|\Delta_{G_n}(b, x^1) - \Delta_{G_n}(\beta_n, x^1)\right\| = O_p\left[(n\delta_n)^{-3/4}(\log n)^{1/2}\right].$$

Proof. The proof of Lemma A.10 is analogous to that of Lemma A.5. Let $\tilde{B}_n = \{b : \|b - \beta_n\| \le C(n\delta_n)^{-1/2}\}$. As in the proofs of Lemma A.5 and Theorem 3.1 of Welsh (1989), cover the ball $\tilde{B}_n$ with cubes $\tilde{\mathcal{C}} = \{\mathcal{C}(b_l)\}$, where $\mathcal{C}(b_l)$ is a cube containing $(b_l - \beta_n)$ with sides of $C(n^5\delta_n)^{-1/2}$ such that $b_l \in \tilde{B}_n$. Then the number of the cubes covering the ball $\tilde{B}_n$ is $\tilde{L} = 4n^4$. Also, we have that $\|(b - \beta_n) - (b_l - \beta_n)\| \le \sqrt{2}C(n^5\delta_n)^{-1/2} \equiv \tilde{\gamma}_n$ for any $(b - \beta_n) \in \mathcal{C}(b_l)$, where $l = 1, \cdots, \tilde{L}$.

As in the proof of Lemma A.5 (in particular, equations (A.1) and (A.2)),

$$\begin{aligned}
&\sup_{b \in \tilde{B}_n} \left\|\Delta_{G_n}(b, x^1) - \Delta_{G_n}(\beta_n, x^1)\right\| \\
&\le \max_{1 \le l \le \tilde{L}} \sup_{(b-\beta_n) \in \mathcal{C}(b_l)} \left\|\Delta_{G_n}(b, x^1) - \Delta_{G_n}(b_l, x^1)\right\| + \max_{1 \le l \le \tilde{L}} \left\|\Delta_{G_n}(b_l, x^1) - \Delta_{G_n}(\beta_n, x^1)\right\| \\
&\le \max_{1 \le l \le \tilde{L}} \left\|\Delta_{G_n}(b_l, x^1) - \Delta_{G_n}(\beta_n, x^1)\right\| \\
&+ \max_{1 \le l \le \tilde{L}} \Bigg|(n\delta_n)^{-1} \sum_{i=1}^{n} \|Z_{ni}K_{ni}\| \\
&\quad \times \Bigg[\Big\{1\big[U_i \le Z'_{ni}(b_l - \beta_n) - B_i + \|Z_{ni}\|\,\tilde{\gamma}_n\big] - F\big[Z'_{ni}(b_l - \beta_n) - B_i + \|Z_{ni}\|\,\tilde{\gamma}_n \big| X_i\big]\Big\} \\
&\quad - \Big\{1\big[U_i \le Z'_{ni}(b_l - \beta_n) - B_i\big] - F\big[Z'_{ni}(b_l - \beta_n) - B_i \big| X_i\big]\Big\}\Bigg]\Bigg| \\
&+ \max_{1 \le l \le \tilde{L}} (n\delta_n)^{-1} \sum_{i=1}^{n} \|Z_{ni}K_{ni}\| \Big\{F\big[Z'_{ni}(b_l - \beta_n) - B_i + \|Z_{ni}\|\,\tilde{\gamma}_n \big| X_i\big] \\
&\quad - F\big[Z'_{ni}(b_l - \beta_n) - B_i - \|Z_{ni}\|\,\tilde{\gamma}_n \big| X_i\big]\Big\}.
\end{aligned} \tag{A.10}$$

Now with some modifications, arguments similar to those in the proof of Lemma A.5 yield the desired result. □

Lemma A.11. *As $n \to \infty$, for any x^1 such that $|x^1| \le 1 - \delta_n$,*

$$\left\|\Delta_{\tilde{G}_n}(b, x^1) - \Delta_{G_n}(b, x^1)\right\| = O_p\left[(n\delta_n)^{-1/2}[\kappa^2/n]^{1/4}(\log n)^{1/2}\right]$$

for any b satisfying $\|b - \beta_n\| \le C(n\delta_n)^{-1/2}$.

Proof. To prove the lemma, define

$$\tilde{H}_n(b, x^1, \theta) = (n\delta_n)^{-1} \sum_{i=1}^{n} \Big\{ \alpha - 1\big[U_i \le Z_{ni}'(b - \beta_n) - B_i + \bar{P}_\kappa(\tilde{X}_i)'(\theta - \theta_{\kappa 0}) - \bar{b}_{\kappa 0}(\tilde{X}_i)\big] \Big\} Z_{ni} K_{ni},$$

$$\tilde{H}_n^*(b, x^1, \theta) = (n\delta_n)^{-1} \sum_{i=1}^{n} \Big\{ \alpha - F\big[Z_{ni}'(b - \beta_n) - B_i + \bar{P}_\kappa(\tilde{X}_i)'(\theta - \theta_{\kappa 0}) - \bar{b}_{\kappa 0}(\tilde{X}_i) \big| X_i\big] \Big\} Z_{ni} K_{ni},$$

and

$$\Delta_{\tilde{H}_n}(b, x^1, \theta) = \tilde{H}_n(b, x^1, \theta) - \tilde{H}_n^*(b, x^1, \theta).$$

Then $\tilde{G}_n(b, x^1) = \tilde{H}_n(b, x^1, \hat{\theta}_{n\kappa})$, $\tilde{G}_n^*(b, x^1) = \tilde{H}_n(b, x^1, \hat{\theta}_{n\kappa})$, and $\Delta_{\tilde{G}_n}(b, x^1) = \Delta_{\tilde{H}_n}(b, x^1, \hat{\theta}_{n\kappa})$. The lemma follows if one can show that

$$\sup_{\|\theta - \theta_{\kappa 0}\| \le C[\kappa/n]^{1/2}} \Big\| \Delta_{\tilde{H}_n}(b, x^1, \theta) - \Delta_{G_n}(b, x^1) \Big\| = O_p\left[(n\delta_n)^{-1/2} [\kappa^2/n]^{1/4} (\log n)^{1/2} \right]$$

for any b satisfying $\|b - \beta_n\| \le C(n\delta_n)^{-1/2}$. This can be proved by using virtually the same arguments as in the proofs of Lemmas A.5 and A.11. □

Define

$$Q_n = (n\delta_n)^{-1} \sum_{i=1}^{n} f(0|X_i) Z_{ni} Z_{ni}' K_{ni}.$$

Lemma A.12. *As $n \to \infty$, for any x^1 such that $|x^1| \le 1 - \delta_n$,*

$$\tilde{G}_n^*(\hat{b}_n, x^1) = -Q_n(\hat{b}_n - \beta_n) + O_p(\delta_n^{\,r}) + O_p\Big[(\hat{b}_n - \beta_n)^2\Big] + o_p\Big[(n\delta_n)^{-1/2}\Big].$$

Proof. Let $\tilde{\Delta}_i(\hat{b}_n, x^1) = Z_{ni}'(\hat{b}_n - \beta_n) - B_i + \bar{P}_\kappa(\tilde{X}_i)'(\hat{\theta}_{n\kappa} - \theta_{\kappa 0}) - \bar{b}_{\kappa 0}(\tilde{X}_i)$. A first-order Taylor expansion of $F\big[\tilde{\Delta}_i(\hat{b}_n, x^1)\big|X_i\big]$ gives

$$\begin{aligned}
\tilde{G}_n^*(\hat{b}_n, x^1) = &-(n\delta_n)^{-1} \sum_{i=1}^{n} \tilde{\Delta}_i(\hat{b}_n, x^1) f(0|X_i) Z_{ni} K_{ni} \\
&- (n\delta_n)^{-1} \sum_{i=1}^{n} \tilde{\Delta}_i(\hat{b}_n, x^1) \big[f(\tilde{\Delta}_i^*(\hat{b}_n, x^1)|X_i) - f(0|X_i) \big] Z_{ni} K_{ni} \\
&\equiv \tilde{G}_{n1}^*(\hat{b}_n) + \tilde{G}_{n2}^*(\hat{b}_n),
\end{aligned}$$

where $\tilde{\Delta}_i^*(\hat{b}_n, x^1)$ is between 0 and $\tilde{\Delta}_i(\hat{b}_n, x^1)$.

Write $\tilde{G}_{n1}^*(\hat{b}_n)$ further as

$$\tilde{G}_{n1}^*(\hat{b}_n) = \tilde{G}_{n11}^* + \tilde{G}_{n12}^* + \tilde{G}_{n13}^* + \tilde{G}_{n14}^*, \tag{A.11}$$

where

$$\tilde{G}^*_{n11} = -(n\delta_n)^{-1}\sum_{i=1}^n Z'_{ni}(\hat{b}_n - \beta_n)f(0|X_i)Z_{ni}K_{ni},$$

$$\tilde{G}^*_{n12} = (n\delta_n)^{-1}\sum_{i=1}^n B_i f(0|X_i)Z_{ni}K_{ni},$$

$$\tilde{G}^*_{n13} = -(n\delta_n)^{-1}\sum_{i=1}^n \bar{P}_\kappa(\tilde{X}_i)'(\hat{\theta}_{n\kappa} - \theta_{\kappa 0})f(0|X_i)Z_{ni}K_{ni},$$

and

$$\tilde{G}^*_{n14} = (n\delta_n)^{-1}\sum_{i=1}^n \bar{b}_{\kappa 0}(\tilde{X}_i)f(0|X_i)Z_{ni}K_{ni}.$$

The first term is $\tilde{G}^*_{n11} = -Q_n(\hat{b}_n - \beta_n)$. Next consider the second term $\tilde{G}^*_{n12}$. Notice that $\max_{1\leq i\leq n}|B_i| \leq C\delta_n^r$ since m_1 is continuously r-times differentiable. Also, it is easy to see that

$$\left\|(n\delta_n)^{-1}\sum_{i=1}^n f(0|X_i)Z_{ni}K_{ni}\right\| = O_p(1).$$

Therefore,

$$\left\|\tilde{G}^*_{n12}\right\| = O_p\left(\max_{1\leq i\leq n}|B_i|\right) = O_p(\delta_n^r).$$

Now consider the third term $\tilde{G}^*_{n13}$ in (A.11). Using Theorem 3.2, we have

$$\begin{aligned}
\tilde{G}^*_{n13} = &-(n^2\delta_n)^{-1}\sum_{i=1}^n \bar{P}_\kappa(\tilde{X}_i)'\Phi_\kappa^{-1}P_\kappa(X_j)\Big\{\alpha - 1\big[U_j \leq 0\big]\Big\}f(0|X_i)Z_{ni}K_{ni} \\
&- (n\delta_n)^{-1}\sum_{i=1}^n \bar{P}_\kappa(\tilde{X}_i)'\bar{B}_{n\kappa}f(0|X_i)Z_{ni}K_{ni} \\
&- (n\delta_n)^{-1}\sum_{i=1}^n \bar{P}_\kappa(\tilde{X}_i)'R_n f(0|X_i)Z_{ni}K_{ni} \\
&\equiv \tilde{G}^*_{n131} + \tilde{G}^*_{n132} + \tilde{G}^*_{n133},
\end{aligned}$$

where the remainder term R_n is defined in Theorem 3.2 and

$$\bar{B}_{n\kappa} = n^{-1}\sum_{j=1}^n \Phi_\kappa^{-1}f(0|X_j)P_\kappa(X_j)b_{\kappa 0}(X_j).$$

First, consider $\tilde{G}^*_{n131}$. To show that

$$\left\|\tilde{G}^*_{n131}\right\| = o_p\left[(n\delta_n)^{-1/2}\right], \tag{A.12}$$

define, for $k = 1, \ldots, r$,

$$g_n^{(k)} = \sum_{j=1}^{n} a_j^{(k)} \{\alpha - 1[U_j \leq 0]\},$$

where

$$\begin{aligned} a_j^{(k)} &= -n^{-3/2}\delta_n^{-1/2} \sum_{i=1}^{n} \bar{P}_\kappa(\tilde{X}_i)' \Phi_\kappa^{-1} P_\kappa(X_j) f(0|X_i) Z_{ni}^{(k)} K_{ni} \\ &\equiv -n^{-3/2}\delta_n^{-1/2} \sum_{i=1}^{n} A_{ij}^{(k)} K_{ni} \end{aligned}$$

and $Z_{ni}^{(k)}$ is the k-th component of Z_{ni}. Then the k-th component of $\tilde{G}^*_{n131}$ is $(n\delta_n)^{-1/2} g_n^{(k)}$ for $k = 1, \ldots, r$. Therefore, to prove (A.12), it suffices to show that $g_n^{(k)} = o_p(1)$ for $k = 1, \ldots, r$. Notice that $E\left[g_n^{(k)} \middle| X_1, \ldots, X_n\right] = 0$ and $\text{Var}\left[g_n^{(k)} \middle| X_1, \ldots, X_n\right] \leq C \sum_{j=1}^{n} \left[a_j^{(k)}\right]^2$. Hence,

$$\max_{1 \leq j \leq n} \left|a_j^{(k)}\right| = o_p(n^{-1/2}) \tag{A.13}$$

implies that $g_n^{(k)} = o_p(1)$. This equation (A.13) can be proved using arguments similar to those used in the proof of Lemma 7 of HM. The proof of (A.13) will be given at the end of the appendix.

Next consider $\tilde{G}^*_{n132}$. By $K_{ni} = 1(|x^1 - X_i^1| \leq \delta_n) K_{ni}$ and the Schwarz inequality,

$$\begin{aligned} \left\|\tilde{G}^*_{n132}\right\| &\leq \left[(n\delta_n)^{-1} \sum_{i=1}^{n} \left\{\bar{P}_\kappa(\tilde{X}_i)' \bar{B}_{n\kappa} 1(|x^1 - X_i^1| \leq \delta_n)\right\}^2\right]^{1/2} \\ &\quad \times \left[(n\delta_n)^{-1} \sum_{i=1}^{n} \|f(0|X_i) Z_{ni} K_{ni}\|^2\right]^{1/2}. \end{aligned} \tag{A.14}$$

By the standard methods for bounding kernel estimators,

$$(n\delta_n)^{-1} \sum_{i=1}^{n} \|f(0|X_i) Z_{ni} K_{ni}\|^2 = O_p(1). \tag{A.15}$$

Also,

$$\begin{aligned} &\left\|(n\delta_n f_{X^1}(x^1))^{-1} \sum_{i=1}^{n} \bar{P}_\kappa(\tilde{X}_i) \bar{P}_\kappa(\tilde{X}_i)' 1(|x^1 - X_i^1| \leq \delta_n) - E\left[\bar{P}_\kappa(\tilde{X}) \bar{P}_\kappa(\tilde{X})' | X^1 = x^1\right]\right\|^2 \\ &\quad = O_p\left(\kappa^2/n + \kappa^2\delta_n^4\right) = o_p(1), \end{aligned} \tag{A.16}$$

where $f_{X^1}(x^1)$ is the density of X^1. Then by (A.16) and Assumption 3.10, the largest eigenvalue of the first term inside $\|\cdot\|$ in (A.16) is bounded for all sufficiently large n. It follows that

$$\begin{aligned}
&(n\delta_n)^{-1}\sum_{i=1}^{n}\left\{\bar{P}_\kappa(\tilde{X}_i)'\bar{B}_{n\kappa}1(|x^1-X_i^1|\leq\delta_n)\right\}^2 \\
&\leq C\bar{B}_{n\kappa}'\left[(n\delta_n f_{X^1}(x^1))^{-1}\sum_{i=1}^{n}\bar{P}_\kappa(\tilde{X}_i)\bar{P}_\kappa(\tilde{X}_i)'1(|x^1-X_i^1|\leq\delta_n)\right]\bar{B}_{n\kappa} \\
&\leq C\left\|\bar{B}_{n\kappa}\right\|^2
\end{aligned}$$

for all sufficiently large n. In view of Lemma A.7, it can be shown that $\left\|\bar{B}_{n\kappa}\right\| = O(\kappa^{-r})$. Therefore, by (A.14),

$$\left\|\tilde{G}_{n132}^*\right\| \leq O_p(\kappa^{-r})O_p(1) = o_p\left[(n\delta_n)^{-1/2}\right] \tag{A.17}$$

provided that $\delta_n \propto n^{-1/(2r+1)}$ and $n^{r/(2r+1)}\kappa^{-r} \to 0$. In particular, if $\kappa \propto n^\nu$, then

$$\nu > \frac{1}{2r+1}. \tag{A.18}$$

Now consider $\tilde{G}_{n133}^*$. Arguments identical to those used to prove (A.17) gives

$$\left\|\tilde{G}_{n133}^*\right\| \leq O_p(\|R_n\|)O_p(1) = O_p\left[(\kappa^2/n)^{3/4}(\log n)^{1/2}+\kappa^{3/2}/n\right] = o_p\left[(n\delta_n)^{-1/2}\right]$$

provided that $\delta_n \propto n^{-1/(2r+1)}$ and $\kappa n^{-(2r+3)/(12r+6)}(\log n)^{1/3} \to 0$. In particular, if $\kappa \propto n^\nu$, then

$$\nu < \frac{2r+3}{12r+6}. \tag{A.19}$$

Combining (A.19) with (A.18) gives

$$\frac{1}{2r+1} < \nu < \frac{2r+3}{12r+6}, \tag{A.20}$$

which requires that r must be larger than or equal to 2. Combining the results for $\tilde{G}_{n13k}^*$ for $k = 1, 2, 3$ gives

$$\tilde{G}_{n13}^* = o_p\left[(n\delta_n)^{-1/2}\right].$$

Next consider the fourth term $\tilde{G}_{n14}^*$ in (A.11). Notice that

$$\begin{aligned}
\left\|\tilde{G}_{n14}^*\right\| &\leq \left[(n\delta_n)^{-1}\sum_{i=1}^{n}\|f(0|X_i)Z_{ni}K_{ni}\|\right]O(\kappa^{-r}) \\
&= O_p(1)O(\kappa^{-r}) \\
&= o_p\left[(n\delta_n)^{-1/2}\right].
\end{aligned}$$

Therefore, combining the results for $\tilde{G}^*_{n1k}$ for $k = 1, 2, 3, 4$ gives

$$\tilde{G}^*_{n1}(\hat{b}_n) = -Q_n(\hat{b}_n - \beta_n) + O_p(\delta_n^r) + o_p\Big[(n\delta_n)^{-1/2}\Big].$$

Now consider $\tilde{G}^*_{n2}(\hat{b}_n)$. It follows from Assumption 3.3 and Theorem 3.1 (b) that

$$\begin{aligned}
\tilde{G}^*_{n2}(\hat{b}_n) &= O_p\left[\{\tilde{\Delta}_i(x^1, \hat{b}_n)\}^2\right] \\
&= O_p\left[(\hat{b}_n - \beta_n)^2 + \delta_n^{2r} + k^2/n + \kappa^{-2r}\right] \\
&= O_p\left[(\hat{b}_n - \beta_n)^2\right] + o_p\Big[(n\delta_n)^{-1/2}\Big].
\end{aligned}$$

Then the lemma follows from combining the results for $\tilde{G}^*_{nk}(\hat{b}_n)$ for $k = 1, 2$. □

Proof of Theorem 3.3. Write

$$\begin{aligned}
\tilde{G}_n(\hat{b}_n, x^1) &= \Delta_{G_n}(\beta_n, x^1) + [\Delta_{G_n}(\hat{b}_n, x^1) - \Delta_{G_n}(\beta_n, x^1)] \\
&\quad + [\Delta_{\tilde{G}_n}(\hat{b}_n, x^1) - \Delta_{G_n}(\hat{b}_n, x^1)] + \tilde{G}^*_n(\hat{b}_n, x^1).
\end{aligned} \tag{A.21}$$

To prove part (a), suppose that $\left\|\hat{b}_n - \beta_n\right\| \leq C(n\delta_n)^{-1/2}$ for any constant $C > 0$. Notice that Q_n is invertible and $\left\|Q_n^{-1}\right\| = O_p(1)$ for all sufficiently large n. By applying Lemmas A.8 - A.12 to equation (A.21), we have

$$\begin{aligned}
\left\|\hat{b}_n - \beta_n\right\| &\leq \left\|Q_n^{-1}\right\| \left\|Q_n(\hat{b}_n - \beta_n)\right\| \\
&\leq O_p\Big[(n\delta_n)^{-1/2}\Big] + O_p(\delta_n^r) + o_p\Big[(n\delta_n)^{-1/2}\Big]
\end{aligned} \tag{A.22}$$

for all sufficiently large n. Therefore, the right-hand side of (A.22) is less than $C(n\delta_n)^{-1/2}$ (w.p.a.1), provided that $(n\delta_n)^{1/2}\delta_n^r$ is bounded. This implies that w.p.a.1,

$$\left\|\hat{b}_n - \beta_n\right\| \leq C(n\delta_n)^{-1/2}.$$

which in turn implies that

$$\left\|\hat{b}_n - \beta_n\right\| = O_p\Big[(n\delta_n)^{-1/2}\Big] = O_p\Big[n^{-r/(2r+1)}\Big].$$

Therefore, part (a) of Theorem 3.3 is proved.

Combining Lemmas A.8, A.10 - A.12, and (A.21) with part (a) of the theorem gives

$$Q_n(\hat{b}_n - \beta_n) = G_n(\beta_n, x^1) - G^*_n(\beta_n, x^1) + \tilde{G}^*_{n12} + r_{n1},$$

where $\tilde{G}^*_{n12}$ was defined in the proof of Lemma A.12, and the remainder term r_{n1} satisfies $\|r_{n1}\| = o_p\Big[(n\delta_n)^{-1/2}\Big]$. By a first-order Taylor expansion and Assumption (3.3), it is easy to show that

$$\left\|\tilde{G}^*_{n12} - G^*_n(\beta_n, x^1)\right\| = O_p\left(\left\{\max_{1\le i\le n}|B_i|\right\}^2\right) = O_p(\delta_n^{2r}) = o_p\Big[(n\delta_n)^{-1/2}\Big].$$

Therefore, it follows that

$$\left\|\hat{b}_n - \beta_n - Q_n^{-1}G_n(\beta_n)\right\| = o_p\Big[(n\delta_n)^{-1/2}\Big] \tag{A.23}$$

for all sufficiently large n. Furthermore, methods similar to those used to establish asymptotic properties of kernel estimators give $\|Q^*_n - Q^*\| = o_p(1)$, where

$$\begin{aligned} Q_* &= \int_{\tilde{x}\in[-1,1]^{d-1}} f(0|x^1,\tilde{x}) f_X(x^1,\tilde{x})d\tilde{x}\int_{-1}^{1}\begin{pmatrix} 1 & u & \dots & u^{r-1} \\ & & \dots & \\ u^{r-1} & u^r & \dots & u^{2(r-1)}\end{pmatrix} K(u)\,du \\ &= [f_1(0|x^1) f_{X^1}(x^1)]\,S(K). \end{aligned}$$

Then it follows that

$$\hat{m}_1(x^1) - m_1(x^1) = e_1'(\hat{b}_n - \beta_n) = e_1' Q_*^{-1} G_n(\beta_n) + r_{n2},$$

where the remainder term r_{n2} satisfies $\|r_{n2}\| = o_p\Big[(n\delta_n)^{-1/2}\Big]$.

Recall that $e_1' S(K)^{-1}(1, u, \dots, u^{r-1})' K(u)$ is a kernel of order r. Then parts (b) and (c) can be proved by using arguments identical to those used to establish asymptotic normality of local polynomial estimators. □

Proof of (A.13). To show (A.13), first notice that

$$\begin{aligned} a_j^{(k)} &= -n^{-3/2}\delta_n^{-1/2} A_{jj}^{(k)} K_{nj} - n^{-3/2}\delta_n^{-1/2}\sum_{i=1, i\neq j}^{n} A_{ij}^{(k)} K_{ni} \\ &= -n^{-3/2}\delta_n^{-1/2}\sum_{i=1, i\neq j}^{n} A_{ij}^{(k)} K_{ni} + O_p\left(n^{-3/2}\delta_n^{-1/2}\zeta_\kappa^2\right) \\ &= -n^{-3/2}\delta_n^{-1/2}\sum_{i=1, i\neq j}^{n} A_{ij}^{(k)} K_{ni} + o_p\left(n^{-1/2}\right) \end{aligned}$$

uniformly over j. Write

$$
\begin{aligned}
& n^{-3/2}\delta_n^{-1/2}\sum_{i=1,i\neq j}^{n} A_{ij}^{(k)}K_{ni} \\
&= n^{-3/2}\delta_n^{-1/2}\sum_{i=1,i\neq j}^{n} E[A_{ij}^{(k)}|X_i]K_{ni} + n^{-3/2}\delta_n^{-1/2}\sum_{i=1,i\neq j}^{n}\left\{A_{ij}^{(k)} - E[A_{ij}^{(k)}|X_i]\right\}K_{ni} \\
&\equiv a_{j1}^{(k)} + a_{j2}^{(k)}.
\end{aligned}
$$

By $E[P_\kappa(X_j)|X_i] = E[P_\kappa(X_j)]$ for $j \neq i$, $K_{ni} = 1(|x^1 - X_i^1| \leq \delta_n)K_{ni}$, and the Schwarz inequality,

$$
\begin{aligned}
\left|a_{j1}^{(k)}\right| &\leq n^{-3/2}\delta_n^{-1/2}\sum_{i=1,i\neq j}^{n}\left|\bar{P}_\kappa(\tilde{X}_i)'\Phi_\kappa^{-1}E[P_\kappa(X_j)]f(0|X_i)Z_{ni}^{(k)}K_{ni}\right| \\
&\leq n^{-1/2}\delta_n^{1/2}\left((n\delta_n)^{-1}\sum_{i=1,i\neq j}^{n}\left\{\bar{P}_\kappa(\tilde{X}_i)'\Phi_\kappa^{-1}E[P_\kappa(X_j)]1(|x^1 - X_i^1| \leq \delta_n)\right\}^2\right)^{1/2} \\
&\quad\times\left((n\delta_n)^{-1}\sum_{i=1,i\neq j}^{n}\left\{f(0|X_i)Z_{ni}^{(k)}K_{ni}\right\}^2\right)^{1/2}. \qquad \text{(A.24)}
\end{aligned}
$$

As in (A.15) and (A.16),

$$
(n\delta_n)^{-1}\sum_{i=1,i\neq j}^{n}\left\{f(0|X_i)Z_{ni}^{(k)}K_{ni}\right\}^2 = O_p(1) \qquad \text{(A.25)}
$$

and

$$
\begin{aligned}
&\left\|(n\delta_n f_{X^1}(x^1))^{-1}\sum_{i=1,i\neq j}^{n}\bar{P}_\kappa(\tilde{X}_i)\bar{P}_\kappa(\tilde{X}_i)'1(|x^1 - X_i^1| \leq \delta_n) - E\left[\bar{P}_\kappa(\tilde{X})\bar{P}_\kappa(\tilde{X})'|X^1 = x^1\right]\right\|^2 \\
&\qquad = O_p\left(\kappa^2/n + \kappa^2\delta_n^4\right) = o_p(1). \qquad \text{(A.26)}
\end{aligned}
$$

Then by (A.26) and Assumption 3.10, the largest eigenvalue of the first term inside $\|\cdot\|$ in (A.26) is bounded for all sufficiently large n. Furthermore, in view of Assumption 3.6 (c), elements of $E[P_\kappa(X_j)]$ are the Fourier coefficients of the density of X. Since the density of

X is bounded, $E[P_\kappa(X_j)]'E[P_\kappa(X_j)]$ converges as $\kappa \to \infty$. Combining all these gives

$$
\begin{aligned}
&(n\delta_n)^{-1} \sum_{i=1, i\neq j}^{n} \left\{ \bar{P}_\kappa(\tilde{X}_i)'\Phi_\kappa^{-1} E[P_\kappa(X_j)] 1(|x^1 - X_i^1| \leq \delta_n) \right\}^2 \\
&\leq C E[P_\kappa(X_j)]'\Phi_\kappa^{-1} \left[(n\delta_n f_{X^1}(x^1))^{-1} \sum_{i=1, i\neq j}^{n} \bar{P}_\kappa(\tilde{X}_i)\bar{P}_\kappa(\tilde{X}_i)' 1(|x^1 - X_i^1| \leq \delta_n) \right] \Phi_\kappa^{-1} E[P_\kappa(X_j)] \\
&\leq C E[P_\kappa(X_j)]'E[P_\kappa(X_j)] \\
&\leq C
\end{aligned}
$$

for all sufficiently large n, where the constant C can be chosen uniformly over j. Combining this and (A.25) with (A.24) proves that $\max_{1\leq j\leq n} \left| a_{j1}^{(k)} \right| = o_p(n^{-1/2})$.

Now, it remains to prove that $\max_{1\leq j\leq n} \left| a_{j2}^{(k)} \right| = o_p(n^{-1/2})$. To do so, notice that

$$
\begin{aligned}
E\left[\left| a_{j2}^{(k)} \right|^2 \middle| \{X_i\}_{i=1, i\neq j}^{n} \right] &= n^{-3}\delta_n^{-1} \sum_{i=1, i\neq j}^{n} \sum_{l=1, l\neq j}^{n} \bar{P}_\kappa(\tilde{X}_i)'\Phi_\kappa^{-1} \\
&\quad \times E\left[\{P_\kappa(X_j) - E[P_\kappa(X_j)]\}\{P_\kappa(X_j) - E[P_\kappa(X_j)]\}' \middle| \{X_i\}_{i=1, i\neq j}^{n} \right] \\
&\quad \times \Phi_\kappa^{-1}\bar{P}_\kappa(\tilde{X}_l) f(0|X_i) Z_{ni}^{(k)} K_{ni} f(0|X_l) Z_{nl}^{(k)} K_{nl} \\
&\leq C n^{-3}\delta_n^{-1} \sum_{i=1, i\neq j}^{n} \sum_{l=1, l\neq j}^{n} \bar{P}_\kappa(\tilde{X}_i)'\Phi_\kappa^{-1}\bar{P}_\kappa(\tilde{X}_l) f(0|X_i) Z_{ni}^{(k)} K_{ni} f(0|X_l) Z_{nl}^{(k)} K_{nl} \\
&\leq C n^{-1}\delta_n \left[(n\delta_n f_{X^1}(x^1))^{-1} \sum_{i=1, i\neq j}^{n} K_{ni}\bar{P}_\kappa(\tilde{X}_i)' \right] \Phi_\kappa^{-1} \\
&\quad \times \left[(n\delta_n f_{X^1}(x^1))^{-1} \sum_{l=1, l\neq j}^{n} \bar{P}_\kappa(\tilde{X}_l) K_{nl} \right] \\
&\leq C n^{-1}\delta_n \, E\left[\bar{P}_\kappa(\tilde{X}) \middle| X^1 = x^1 \right]' E\left[\bar{P}_\kappa(\tilde{X}) \middle| X^1 = x^1 \right]
\end{aligned}
$$

for all sufficiently large n. Elements of $E[\bar{P}_\kappa(\tilde{X})|X^1 = x^1]$ are the Fourier coefficients of the conditional density of $\tilde{X}$ given $X^1 = x^1$, which is bounded. Hence, $E[\bar{P}_\kappa(\tilde{X})|X^1 = x^1]'E[\bar{P}_\kappa(\tilde{X})|X^1 = x^1]$ converges as $\kappa \to \infty$, implying that

$$
E\left[\left| a_{j2}^{(k)} \right|^2 \middle| \{X_i\}_{i=1, i\neq j}^{n} \right] = o_p(n^{-1})
$$

uniformly over j. This in turn implies that $\max_{1\leq j\leq n} \left| a_{j2}^{(k)} \right| = o_p(n^{-1/2})$ by Markov inequality. $\square$

References

Chaudhuri, P. (1991): Nonparametric estimates of regression quantiles and their local bahadur representation, *Annals of Statistics* 19, 760-777.

De Gooijer, J.G. and D. Zerom (2003): On additive conditional quantiles with high-dimensional covariates, *Journal of the American Statistical Association*, 98, 135-146.

Doksum, K. and J.-Y. Koo (2000): On spline estimators and prediction intervals in nonparametric regression, *Computational Statistics & Data Analysis*, 35, 76-82.

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

He, X. and P. Shi (1994): Convergence rate of B-spline estimators of nonparametric donditional quantile functions, *Journal of Nonparametric Statistics* 3, 299–308.

He, X. and P. Shi (1996): Bivariate tensor-product B-splines in a partly linear model, *Journal of Multivariate Analysis* 58, 162–181.

He, X. and S. Portnoy (2000): Some asymptotic results on bivariate quantile splines, *Journal of Statistical Planning and Inference*, 91, 341-349.

Horowitz, J.L. and E. Mammen (2002): Nonparametric estimation of an additive model with a link function, Cemmap Working Paper CWP19/02, http://cemmap.ifs.org.uk.

Horowitz, J.L. and V.G. Spokoiny (2002): An adaptive, rate-optimal test of linearity for median regression models, *Journal of the American Statistical Association*, 97, 822-835.

Huang, J.Z. (2003): Local asymptotics for polynomial spline regression, *Annals of Statistics*, 31, 1600-1635.

Knight, K. (1998): Limiting distributions for L_1 regression estimators under general conditions, *Annals of Statistics* 26, 755-770.

Koenker, R. and G. Bassett (1978): Regression quantiles, *Econometrica*, 46, 33-50.

Newey, W. K. (1997): Convergence rates and asymptotic normality for series estimators, *Journal of Econometrics*, 79, 147-168.

Portnoy, S. (1997): Local asymptotics for quantile smoothing splines, *Annals of Statistics*, 25, 414-434.

Ruppert, D. and M.P. Wand (1994): Multivariate locally weighted least squares estimation, *Annals of Statistics*, 22, 1346-1370.

van der Vaart, A. W. and J. A. Wellner (1996), *Weak Convergence and Empirical Processes*, New-York: Springer-Verlag.

Welsh, A. H. (1989): On M-processes and M-estimation, *Annals of Statistics*, 17, 337-361.

Yafeh, Y. and O. Yosha (2003): Large shareholders and Banks: Who monitors and how? *Economic Journal*, 113, 128-146.

Table 1. Results of Monte Carlo Experiments

ρ	n	Results for m_1		Results for m_2	
		MI	2S	MI	2S
			$d = 2$		
0.2	100	0.0834	0.0783	0.1430	0.1497
	200	0.0560	0.0519	0.0964	0.1125
0.8	100	0.1638	0.0920	0.2957	0.1620
	200	0.1331	0.0621	0.2602	0.1146
			$d = 5$		
0.2	100	0.1268	0.0688	0.1914	0.1466
	200	0.0917	0.0534	0.1293	0.1176
0.8	100	0.1810	0.0893	0.4060	0.1618
	200	0.1650	0.0638	0.3578	0.1208

Note: The values shown in Table 1 are the average absolute deviation errors (AADE's) for the marginal integration (MI) and two-stage (2S) estimators.

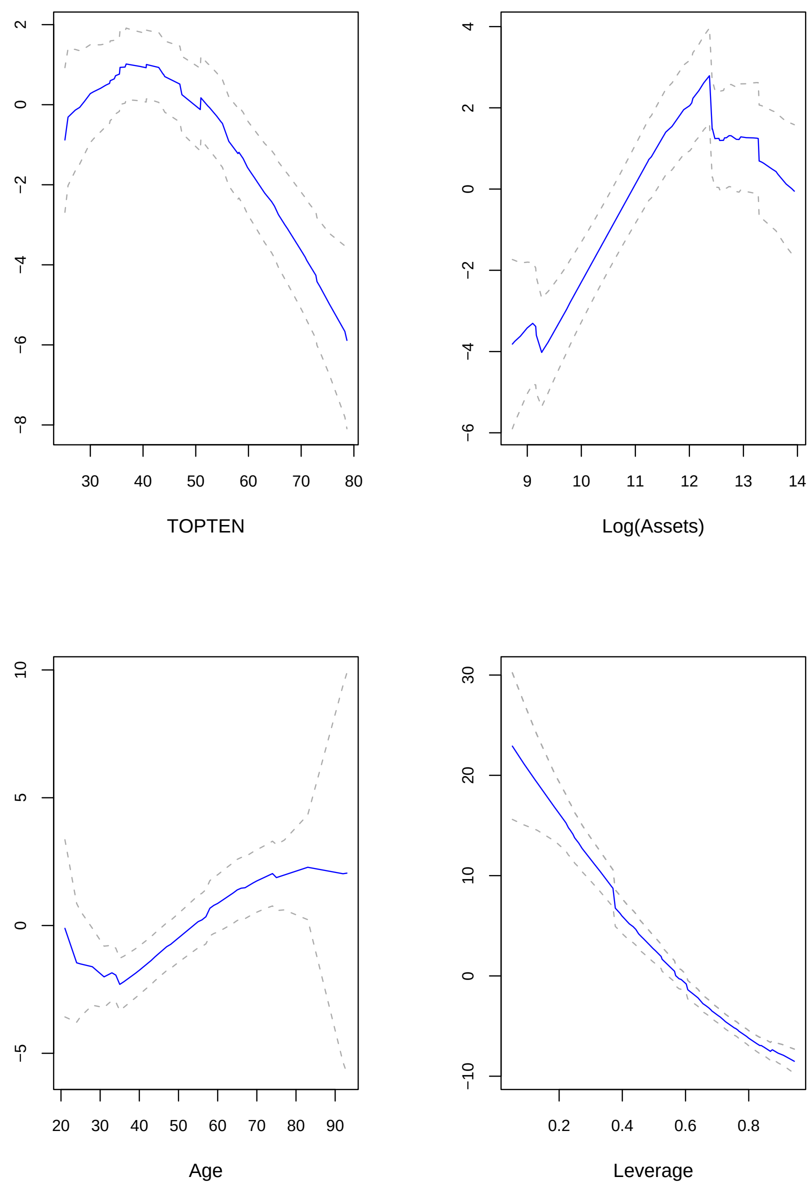

Figure 1. Estimation Results

Note: The estimated additive components of a nonparametric additive median regression model are shown along with their 90% pointwise confidence intervals.