cemmap

centre for microdata methods and practice

A bootstrap method for constructing pointwise and uniform confidence bands for conditional quantile functions

Joel L. Horowitz
Anand Krishnamurthy

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP01/17

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

A BOOTSTRAP METHOD FOR CONSTRUCTING POINTWISE AND UNIFORM CONFIDENCE BANDS FOR CONDITIONAL QUANTILE FUNTIONS

by

Joel L. Horowitz
Department of Economics
Northwestern University
Evanston, IL 60201
joel-horowitz@northwestern.edu

Anand Krishnamurthy
Cornerstone Research
599 Lexington Ave.
New York, NY 10022
akrishnamurthy@cornerstone.com

January 2017

ABSTRACT

This paper is concerned with inference about the conditional quantile function in a nonparametric quantile regression model. Any method for constructing a confidence interval or band for this function must deal with the asymptotic bias of nonparametric estimators of the function. In estimation methods such as local polynomial estimation, this is usually done through undersmoothing or explicit bias correction. The latter usually requires oversmoothing. However, there are no satisfactory empirical methods for selecting bandwidths that under- or oversmooth. This paper extends the bootstrap method of Hall and Horowitz (2013) for conditional mean functions to conditional quantile functions. The paper also shows how the bootstrap method can be used to obtain uniform confidence bands. The bootstrap method uses only bandwidths that are selected by standard methods such as cross validation and plug-in. It does not use under- or oversmoothing. The results of Monte Carlo experiments illustrate the numerical performance of the bootstrap method.

Key words: Quantile regression; smoothing; confidence band; bootstrap

A BOOTSTRAP METHOD FOR CONSTRUCTING POINTWISE AND UNIFORM CONFIDENCE BANDS FOR CONDITIONAL QUANTILE FUNTIONS

1. INTRODUCTION

This paper is concerned with inference about the unknown function g in the nonparametric quantile regression model

(1.1) $Y = g(X) + \varepsilon; \quad P(\varepsilon \leq 0) = \tau$,

where X is an observed continuously distributed explanatory variable and ε is an unobserved continuously distributed random variable that is independent of X and whose τ quantile ($0 < \tau < 1$) is 0. Hall and Horowitz (2013) (hereinafter HH) describe a bootstrap method for constructing a pointwise confidence band for the unknown function $m(x) = E(Y \mid X = x)$ in a nonparametric mean regression. This paper extends the bootstrap method of HH to g in the quantile regression model (1.1). The paper also shows how the bootstrap method can be used to construct a uniform confidence band for g. The method for constructing a uniform confidence band for g can be used to construct a uniform confidence band for m, but this is not done here.

Any method for constructing a pointwise or uniform confidence band for g based on a nonparametric estimate must deal with the problem of asymptotic bias. For example, a local polynomial estimate of g with a bandwidth chosen by cross-validation or plug-in methods is asymptotically biased. Denote the estimate by $\hat{g}$. The expected value of $\hat{g}$ does not equal g, the asymptotic distribution of the scaled estimate is not centered at g, and the true coverage probability of an asymptotic confidence interval for g that is constructed from the normal distribution in the usual way is less than the nominal probability. This problem is usually overcome by undersmoothing or explicit bias reduction. Undersmoothing consists of making the bias asymptotically negligible by using a bandwidth whose rate of convergence is faster than the asymptotically optimal rate. In explicit bias reduction, an estimate of the asymptotic bias is used to construct an asymptotically unbiased estimate of g. Most explicit bias reduction methods involve some form of oversmoothing, that is using a bandwidth whose rate of convergence is slower than the asymptotically optimal rate. Undersmoothing and explicit bias correction methods are also available for the conditional mean function m.

Methods based on undersmoothing or oversmoothing require a bandwidth whose rate of convergence is faster or slower than the asymptotically optimal rate. As discussed by HH, there are no attractive, effective empirical ways to choose these bandwidths. In addition, undersmoothing can produce very wiggly confidence bands, even for smooth conditional quantile or conditional mean functions.

Explicit bias correction methods that rely on estimation of derivatives can also produce wiggly confidence bands.

The method presented in this paper, like the method of HH, uses bandwidths chosen by standard empirical methods such as cross validation or a plug-in rule. It does not under- or oversmooth and does not use auxiliary or other non-standard bandwidths. Instead, the method uses the bootstrap to estimate the bias of $\hat{g}$. The bootstrap estimate of the bias has stochastic noise that is comparable in size to the bias itself. However, combining a suitable quantile of the "distribution" of the bootstrap bias estimate with $\hat{g}$ enables us to obtain a pointwise confidence band with an asymptotic coverage probability that equals or exceeds $1-\alpha$ for any given $\alpha > 0$ at all but a user specified fraction of the possible values of x. The exceptional points are in regions where the function g has sharp peaks or troughs that cause the bias of $\hat{g}$ to be unusually large. These regions are typically visible in a plot of $\hat{g}$ and can also be found through a theoretical analysis. An asymptotic uniform confidence band that has no exceptional points is obtained by replacing the bootstrap bias estimate with an upper bound on the estimated bias.

Methods that use undersmoothing have been described by Bjerve, Doksum, and Yandell (1985); Hall (1992); Hall and Owen (1993); Neumann (1995); Chen (1996); Neumann and Polzehl (1998); Picard and Tribouley (2000); Chen, Härdle, and Li (2003); Claeskens and Van Keilegom (2003); Härdle, Huet, Mammen, and Sperlich (2004); and McMurry and Politis (2008). Methods based on oversmoothing have been described by Härdle and Bowman (1988); Härdle and Marron (1991); Hall (1992); Eubank and Speckman (1993); Sun and Loader (1994); Härdle, Huet, and Jolivet (1995); Xia (1998); and Schucany and Somers (1977). Calonico, Cattaneo, and Farrell (2016) describe an explicit bias correction method for conditional mean functions that does not require oversmoothing or an auxiliary bandwidth. It is not known whether this method can be extended to conditional quantile functions.

Section 2 of this paper presents an informal description of our method. The method is similar in some respects to that of HH for conditional mean functions, but the non-smoothness of quantile estimators presents problems that are different from those involved in estimating conditional mean functions. These require a separate treatment and modifications of parts of the method of HH. Section 2 also outlines the extension of our method to a heteroskedastic version of model (1.1). Section 3 presents formal theoretical results. Section 4 presents simulation results that illustrate the numerical performance of the method. Conclusions are presented in Section 5. The proofs of theorems are in the appendix, which is Section 6.

2. INFORMAL DESCRIPTION OF THE METHOD

Let $\{Y_i, X_i : i = 1,...,n\}$ denote an independent random sample of observations from the distribution of (Y, X) in model (1.1). Let $\hat{g}(x)$ denote a nonparametric estimator of $g(x)$. Denote the bias and variance of $\hat{g}(x)$ by $\beta(x) = E[\hat{g}(x)] - g(x)$ and $\sigma_{\hat{g}}^2(x) = Var[\hat{g}(x)]$, respectively. Assume that $\{\hat{g}(x) - E[\hat{g}(x)]\} / \sigma_{\hat{g}}(x) \to^d N(0,1)$ as $n \to \infty$. To minimize the complexity of the discussion in the remainder of this paper, we assume that X is a scalar random variable and $\hat{g}$ is a local linear quantile regression estimator. The main results of the paper continue to apply if X is a vector or $\hat{g}$ is a local polynomial estimator of odd degree different from 1. This paper does not treat series estimators. The local linear quantile estimation procedure is described in Step 1 in Section 2.1. To avoid boundary effects we restrict attention to a compact set $\mathcal{S}$ that is contained in an open subset of the support of X. Let h denote the bandwidth used in local polynomial estimation of g.

If $\beta(x)$ were known, an asymptotic $1-\alpha_0$ confidence interval for $g(x)$ would be

$$\hat{g}(x) - z_{1-\alpha/2}\sigma_{\hat{g}}(x) \le g(x) \le \hat{g}(x) + z_{1-\alpha/2}\sigma_{\hat{g}}(x) .$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution, $\alpha = \alpha(x, \alpha_0)$ satisfies

$$\Phi[z_{1-\alpha/2} - \beta(x)/\sigma_{\hat{g}}(x)] - \Phi[-z_{1-\alpha/2} - \beta(x)/\sigma_{\hat{g}}(x)] = 1 - \alpha_0 , \tag{2.1}$$

and Φ is the normal distribution function. In applications, $\beta(x)$ and $\sigma_{\hat{g}}(x)$ are unknown. Let $\hat{\sigma}_{\hat{g}}(x)$ be the estimate of $\sigma_{\hat{g}}(x)$ that is described in Section 2.1. Let $\hat{\lambda}(x)$ denote the bootstrap estimate of $\beta(x)/\sigma_{\hat{g}}(x)$ that is obtained in Step 5 of the procedure described in Section 2.1, and let $\hat{\alpha}(x, \alpha_0)$ denote the solution in α to

$$\Phi[z_{1-\alpha/2} - \hat{\lambda}(x)] - \Phi[-z_{1-\alpha/2} - \hat{\lambda}(x)] = 1 - \alpha_0 .$$

For $\xi \in [0,1]$, let $\hat{\alpha}_{\xi}(\alpha_0)$ be the ξ quantile of points in the set $\{\hat{\alpha}(x, \alpha_0) : x \in \mathcal{S}\}$. Define $\hat{z}(\alpha_0) = z_{1-\hat{\alpha}_{\xi}(\alpha_0)/2}$. Construct the pointwise confidence band

$$\mathcal{B}_n[\hat{\alpha}_{\xi}(\alpha_0)] = \left\{(x,y) : \hat{g}(x) - \hat{z}(\alpha_0)\hat{\sigma}_{\hat{g}}(x) \le y \le \hat{g}(x) + \hat{z}(\alpha_0)\hat{\sigma}_{\hat{g}}(x)\right\}.$$

It is shown in Section 3.3 that $\mathcal{B}_n$ has asymptotic coverage probability equal to or greater than $1-\alpha_0$ except for a proportion ξ of points $x \in \mathcal{S}$.

To construct a uniform confidence band for g, define

$$\hat{\lambda}_{\max} = \max_{x \in \mathcal{S}} \hat{\lambda}(x)$$

and

$$\hat{\lambda}_{\min} = \min_{x \in \mathcal{S}} \hat{\lambda}(x).$$

Let W_1 be the mean-zero Gaussian process defined in Section 3.1, and let $\hat{t}_U$ satisfy

$$P\left(-\hat{t}_U - \hat{\lambda}_{\min} \le W_1\left(\frac{x}{h}\right) \le \hat{t}_U - \hat{\lambda}_{\max} \quad \forall\, x \in \mathcal{S}\right) = 1 - \alpha_0,$$

where h is the bandwidth used for local linear quantile estimation of g. It is shown in Section 3.3 that

$$\mathcal{B}_U(\alpha_0) \equiv \{(y,x): \ \hat{g}(x) - \hat{t}_U \hat{\sigma}_{\hat{g}}(x) \le y \le \hat{g}(x) + \hat{t}_U \hat{\sigma}_{\hat{g}}(x); x \in \mathcal{S}\}$$

is an asymptotic uniform confidence band for g whose coverage probability equals or exceeds $1-\alpha_0$.

2.1 The Estimation Procedure

This section provides a step-by-step explanation of the method for constructing $\mathcal{B}_n$ and $\mathcal{B}_U$.

Step 1: *Local linear estimation of* g *and estimation of* $\sigma_{\hat{g}}^2$. Let K be a kernel function and h be a possibly random bandwidth. For any real v, define $K_h(v) = K(v/h)$. Define the check function

$$\rho_\tau(v) = v[\tau - I(v \le 0)],$$

where $0 < \tau < 1$ and I is the indicator function. The local linear estimator of $g(x)$ is $\hat{g}(x) = \hat{b}_0$, where

$$(\hat{b}_0, \hat{b}_1) = \arg\min_{b_0, b_1} \sum_{i=1}^{n} \rho_\tau[Y_i - b_0 - b_1(X_i - x)] K_h(X_i - x).$$

To obtain $\hat{\sigma}_{\hat{g}}(x)$, let $\hat{f}_X(x)$ be a consistent kernel nonparametric estimator of $f_X(x)$, the probability density function of X at x. Let $\tilde{\varepsilon}_i = Y_i - \hat{g}(X_i)$ be the residuals from estimating model (1.1), and let $\hat{f}_\varepsilon(0)$ be a consistent kernel nonparametric estimator of $f_\varepsilon(0)$, the probability density of ε at 0. Specifically,

$$\hat{f}_\varepsilon(0) = (nh_\varepsilon)^{-1} \sum_{i=1}^{n} K_{h_\varepsilon}(\tilde{\varepsilon}_i),$$

where h_ε is a bandwidth. Define

$$B_K = \int K^2(v)\,dv.$$

It is shown in Section 3.2 that the variance of the asymptotic distribution of $\hat{g}(x)$ is

$$\sigma_{\hat{g}}^2(x) = \frac{\tau(1-\tau)B_K}{(nh) f_X(x)[f_\varepsilon(0)]^2}$$

The variance of $\hat{g}(x)$ can be estimated by replacing $f_X(x)$ and $f_\varepsilon(0)$ with their consistent estimators to obtain

$$\hat{\sigma}_{\hat{g}}^2(x) = \frac{\tau(1-\tau)B_K}{(nh)\hat{f}_X(x)[\hat{f}_\varepsilon(0)]^2}.$$

Step 2: *Compute centered residuals*. Let q_n be the τ quantile of the residuals $\{\tilde{\varepsilon}_i\}$. That is

$$q_n = \inf\left[q:\ n^{-1}\sum_{i=1}^{n} I(\tilde{\varepsilon}_i \le q) \ge \tau\right].$$

The centered residuals are

$$\hat{\varepsilon}_i = \tilde{\varepsilon}_i - q_n.$$

The τ quantile of centered residuals is 0.

Step 3. *Construct the bootstrap resample*. The bootstrap resample is $\{Y_i^*, X_i:\ i = 1,...,n\}$, where

$$Y_i^* = \hat{g}(X_i) + \varepsilon_i^*$$

and the ε_i^*s are obtained by sampling the $\hat{\varepsilon}_i$s randomly with replacement. The X_is are not resampled.

Step 4: *Compute the bootstrap estimate of the asymptotic bias of* $\hat{g}(x)$. Let there be B bootstrap resamples that are indexed by $b = 1,...,B$. For resample b, define

$$T_{nb}^*(x) = n^{-1}\sum_{i=1}^{n}\left\{1 - \tau^{-1} I\left[Y_i^* \le \hat{b}_0 + \hat{b}_1(X_i - x)\right]\right\} K_h(X_i - x).$$

Let $E^*(T_{nb}^*)$ denote the bootstrap expectation of T_{nb}^* conditional on the data. Estimate $E^*(T_{nb}^*)$ by

$$\hat{T}_n(x) = B^{-1}\sum_{b=1}^{B} T_{nb}^*(x).$$

$\hat{T}_n(x)$ converges almost surely to $E^*[T_{nb}^*(x)]$ and can be made arbitrarily close to $E^*[T_{nb}^*(x)]$ by making B sufficiently large. The bootstrap estimate of $\beta(x) = E\hat{g}(x) - g(x)$ is

$$\hat{\beta}(x) = h^{-1}\left[\frac{\tau}{(1-\tau)B_K}\right]^{1/2} \frac{\hat{\sigma}_{\hat{g}}(x)}{[\hat{f}_X(x)]^{1/2}}\hat{T}_n(x).$$

In contrast to HH, we do not form a bootstrap estimate of $g(x)$. Instead, we form a bootstrap estimate of the asymptotic form of $E\hat{g}(x) - g(x)$ from the analytic expression for this asymptotic form. This expression is given by equation (3.4).

The estimate $\hat{\beta}(x)$ has random noise whose order of magnitude is the same as that of $E\hat{\beta}(x) - \beta(x)$. Therefore, $\hat{\beta}(x)$ is not consistent for $\beta(x)$. The method for finding pointwise and uniform confidence bands for g takes account of this inconsistency. See Steps 5 and 6 below.

Step 5: *Obtain the normalized estimate of the bias and effective significance level.* The normalized bias of $\hat{g}(x)$ is defined as $\lambda(x) = \beta(x)/\sigma_{\hat{g}}(x)$ and is estimated by $\hat{\lambda}(x) = \hat{\beta}(x)/\hat{\sigma}_{\hat{g}}(x)$. The effective significance level at point x, $\hat{\alpha}(x,\alpha_0)$, is defined as the solution in α to the equation

$$\Phi[z_{1-\alpha/2} - \hat{\lambda}(x)] - \Phi[-z_{1-\alpha/2} - \hat{\lambda}(x)] = 1 - \alpha_0 .$$

Step 6: *Construct a pointwise confidence band for* g. Let $\xi \in [0,1]$. Let $\hat{\alpha}_\xi(\alpha_0)$ be the ξ quantile of points in the set $\{\hat{\alpha}(x,\alpha_0): x \in \mathcal{S}\}$. Define $\hat{z}(\alpha_0) = z_{1-\hat{\alpha}_\xi(\alpha_0)/2}$. Construct the pointwise confidence band

$$\mathcal{B}_n[\hat{\alpha}_\xi(\alpha_0)] = \left\{(x,y): \hat{g}(x) - \hat{z}(\alpha_0)\hat{\sigma}_{\hat{g}}(x) \le y \le \hat{g}(x) + \hat{z}(\alpha_0)\hat{\sigma}_{\hat{g}}(x)\right\}.$$

It is shown in Section 3.3 that the pointwise band $\mathcal{B}_n[\hat{\alpha}_\xi(\alpha_0)]$ covers $g(x)$ with probability at least $1-\alpha_0$ except for a proportion ξ of points $x \in \mathcal{S}$. Specifically, for $\alpha(x,\alpha_0)$ as in (2.1), let $\alpha_\xi(\alpha_0)$ denote the ξ quantile of the points $\{\alpha(x,\alpha_0): x \in \mathcal{S}\}$. Define the set

(2.2) $$\mathcal{R}_\xi(\alpha_0) = \{x \in \mathcal{S}: \alpha(x,\alpha_0) > \alpha_\xi(\alpha_0)\}.$$

Then

$$\liminf_{n\to\infty} P\{[x, g(x)] \in \mathcal{B}_n[\hat{\alpha}_\xi(\alpha_0)\} \ge 1 - \alpha_0$$

for each $x \in \mathcal{R}_\xi(\alpha_0)$.

Step 7: *Construct a uniform confidence band for* g. Define

(2.3) $$\hat{\lambda}_{\max} = \max_{x\in\mathcal{S}} \hat{\lambda}(x)$$

and

(2.4) $$\hat{\lambda}_{\min} = \min_{x\in\mathcal{S}} \hat{\lambda}(x).$$

Let W_1 denote the mean-zero Gaussian process defined in Theorem 3.1 in Section 3.1. Define $\hat{t}_U$ as the solution in t to

(2.5) $$P\left[-t - \hat{\lambda}_{\min} \le W_1\left(\frac{x}{h}\right) \le t - \hat{\lambda}_{\max} \ \ \forall\, x \in \mathcal{S}\right] = 1 - \alpha_0 .$$

The asymptotic uniform confidence band is

$$\mathcal{B}_U(\alpha_0) \equiv \{(y,x): \hat{g}(x) - \hat{t}_U\hat{\sigma}_{\hat{g}}(x) \le y \le \hat{g}(x) + \hat{t}_U\hat{\sigma}_{\hat{g}}(x); x \in \mathcal{S}\}.$$

The quantities $\beta^*_{\max}$ and $\beta^*_{\min}$ can be computed by replacing $\mathcal{S}$ in (2.3) and (2.4) with a fine grid of equally spaced points. The critical value $\hat{t}_U$ can be computed by replacing $\mathcal{S}$ in (2.5) with the grid.

2.2. Heteroskedasticity

A heteroskedastic version of (1.1) is

(2.6) $Y = g(X) + \sigma(X)\varepsilon; \quad P(\varepsilon \le 0) = \tau$,

where $\sigma(\cdot)$ is a scale function and ε is independent of X. Identification of σ requires normalizing the scale of ε. This is done by setting the interquartile range (IQR) of ε equal to 1. Then

$$\sigma(x) = IQR(Y \mid X = x).$$

Let $\hat{g}(x)$ be the local linear quantile regression estimate of $g(x)$ and $\hat{\sigma}(x)$ be a consistent nonparametric estimate of $IQR(Y \mid X = x)$. The residuals of model (2.5) are

$$\breve{\varepsilon}_i = \frac{Y_i - \hat{g}(X_i)}{\hat{\sigma}(X_i)}.$$

The centered residuals of (2.6) are as in Step 2 after replacing $\tilde{\varepsilon}_i$ with $\breve{\varepsilon}_i$. The estimate of the variance of of the estimate of $g(x)$ in (2.6) is

$$\hat{\sigma}_{\hat{g}}^2(x) = \frac{\hat{\sigma}^2(x)\tau(1-\tau)B_K}{(nh)\hat{f}_X(x)[\hat{f}_\varepsilon(0)]^2},$$

where $\hat{f}_\varepsilon$ is now based on the $\breve{\varepsilon}_i$ s. Bootstrap sampling is done by setting

$$Y_i^* = \hat{g}(X_i) + \hat{\sigma}(X_i)\varepsilon_i^*,$$

where the ε_i^* s are sampled randomly with replacement from the centered $\breve{\varepsilon}_i$ s. Steps 4-6 for construction of pointwise and uniform confidence bands remain as in Section 2.1 but with the foregoing modifications of $\hat{\sigma}_{\hat{g}}^2(x)$ and the bootstrap sampling procedure.

3. **THEORETICAL RESULTS**

This section presents theorems giving conditions under which the pointwise and uniform confidence bands constructed in Steps 5-6 of Section 2.1 have the claimed coverage properties when $\hat{g}$ is a local linear quantile regression estimator. We make the following assumptions:

Assumption 1: (i) The data $\{Y_i, X_i : i = 1,...,n\}$ are an independent random sample from model (1.1); (ii) X in (1.1) has compact support; (iii) ε in (1.1) is independent of X and $P(\varepsilon \le 0) = \tau$ for some $\tau \in (0,1)$.

Assumption 2: (i) The distribution of X is absolutely continuous with respect to Lebesgue measure with probability density function f_X; (ii) f_X is bounded away from 0 on $\text{supp}(X)$ and twice continuously differentiable on the interior of $\text{supp}(X)$.

Assumption 3: The distribution of ε is absolutely continuous with respect to Lebesgue measure with probability density function f_ε; (ii) f_ε is twice continuously differentiable and $f_\varepsilon(0) > 0$.

Assumption 4: The function g in (1.1) is three times continuously differentiable on the interior of $\text{supp}(X)$; (ii) $\hat{g}$ is a local linear quantile regression estimator of g. (iii) There is a compact set $\mathcal{G} \in \mathbb{R}^2$ such that $[g(x), g'(x)] \in \mathcal{G}$ for each x.

Assumption 5: The kernel K is a probability density function with support $[-1,1]$, symmetrical around 0, and twice continuously differentiable on $(-1,1)$.

Assumption 6: The bandwidth h used to construct $\hat{g}$ satisfies:

(i) $h = \hat{d}n^{-1/5}$, where $\hat{d}$ is a function of the data $\{Y_i, X_i: i = 1,...,n\}$ and $\hat{d} \to^p d_0$ as $n \to \infty$ for some finite constant $d_0 > 0$.

(ii) There exists a finite constant $D_1 > 0$ such that

$$P\left(|\hat{d} - d_0| > n^{-D_1}\right) \to 0 \text{ as } n \to \infty.$$

(iii) There are constants D_2 and D_3 such that $0 < D_2 < D_3 < 1$ and

$$P\left(n^{-D_3} \le h \le n^{-D_2}\right) = 1 - O\left(n^{-C}\right)$$

as $n \to \infty$ for all finite $C > 0$.

Assumption 1 defines the data generation process. Assumptions 2-4 are smoothness assumptions. Assumption 5 specifies standard properties of K. Assumption 6 is satisfied by standard bandwidth choice methods such as cross-validation and plug-in methods. Under assumptions 1-6, local linear estimates of g obtained using the random bandwidth h and the deterministic bandwidth h_0 are asymptotically equivalent. See Lemma 6.1 in Section 6.

3.1 Asymptotic Approximations to $(nh)^{1/2}[\hat{g}(x) - g(x)]$ and the Bootstrap Bias Estimate

The asymptotic coverage probabilities of the pointwise and uniform confidence bands defined in Steps 5 and 6 of Section 2.1 depend on strong asymptotic approximations to $(nh)^{1/2}[\hat{g}(x) - g(x)]$ and the bootstrap estimate of $E\hat{g}(x) - g(x)$. These approximations are given by the following theorems.

Theorem 3.1: Let assumptions 1-6 hold. Define

$$\psi_0(x) = \frac{[\tau(1-\tau)B_K]^{1/2}}{f_X(x)^{1/2} f_\varepsilon(0)},$$

$$\kappa_2 = \int v^2 K(v)dv,$$

and $h_0 = d_0 n^{-1/5}$. There exists a Gaussian process $W_1(x)$ defined on the same probability space as the data such that $E[W_1(x)] = 0$ for all $x \in \mathcal{S}$, $E[W_1(x)]^2 = 1$ for all $x \in \mathcal{S}$, and for any $\eta > 0$

$$\lim_{n\to\infty} P\left\{ \sup_{x\in\mathcal{S}} \left| (nh)^{1/2}[\hat{g}(x) - g(x)] - \left[\frac{d_0^{5/2}\kappa_2}{2} g''(x) + \psi_0(x) W_1\left(\frac{x}{h_0}\right)\right]\right| > \eta \right\} = 0 . \blacksquare$$

It follows from Theorem 3.1 that for each $x \in \mathcal{S}$,

(3.1) $\quad (nh)^{1/2}[\hat{g}(x) - g(x)] \to^d N[\mu_g, V_g(x)]$,

where

(3.2) $\quad \mu_g = \dfrac{d_0^{5/2}\kappa_2}{2} g''(x)$

and

(3.3) $\quad V_g(x) = \psi_0^2(x)$.

Moreover, asymptotically,

(3.4) $\quad E\hat{g}(x) - g(x) = \dfrac{h_0^2 \kappa_2}{2} g''(x)$

and

(3.5) $\quad Var[\hat{g}(x)] = \sigma_{\hat{g}}^2(x) = \dfrac{\tau(1-\tau)B_K}{(nh_0) f_X(x)[f_\varepsilon(0)]^2}$.

Properties (3.1)-(3.5) were obtained previously by Fan, Hu, and Truong (1994) and Yu and Jones (1997).

$Var[\hat{g}(x)]$ can be estimated consistently by replacing $f_X(x)$ and $f_\varepsilon(0)$ on the right-hand sides of (3.3) and (3.5) by the consistent estimators $\hat{f}_X(x)$ and $\hat{f}_\varepsilon(0)$. Bootstrap estimation of $E\hat{g}(x) - g(x)$ relies on the strong approximation given by the following theorem.

Theorem 3.2: Let assumptions 1-6 hold. Let E^* denote the bootstrap expectation conditional on the data. Define

$$A_1(x) = \tau^{-1} d_0^{5/2} f_\varepsilon(0) f_X(x)$$

and

$$A_2(x) = \left[\frac{1-\tau}{\tau} B_K f_X(x)\right]^{1/2} .$$

Then

(i) For all $x \in \mathcal{S}$, and any $\eta > 0$

$$\lim_{n\to\infty} P\left\{\sup_{x\in\mathcal{S}}\left|\left(\frac{n}{h}\right)^{1/2} E^*[T^*_{nb}(x)] + \frac{A_1(x)\kappa_2}{2} g''(x) - A_2(x)W_1\left(\frac{x}{h_0}\right)\right| > \eta\right\} = 0 .$$

(ii) There exists a Gaussian process $\Delta(x)$ such that $E[\Delta(x)] = 0$ for all $x \in \mathcal{S}$, $E[\Delta(x)]^2 = 1$ for all $x \in [0,1]$, and for any $\eta > 0$

$$\lim_{n\to\infty} P\left\{\sup_{x\in\mathcal{S}}\left|\hat{\lambda}(x) - \left[\frac{\beta(x)}{\sigma_{\hat{g}}(x)} + \Delta(x)\right]\right| > \eta\right\} = 0 . \blacksquare$$

For any $\alpha \in (0,1)$ and $x \in \mathcal{S}$ define

$$\hat{\pi}(x,\alpha) = \Phi\left[z_{1-\alpha/2} - \hat{\lambda}(x)\right] - \Phi\left[-z_{1-\alpha/2} - \hat{\lambda}(x)\right].$$

The following corollary to Theorem 3.2 is used to establish the asymptotic coverage probabilities of the confidence bands constructed in Steps 5 and 6 of Section 2.1.

Corollary 3.3: Let assumptions 1-6 hold. Then for any $\eta > 0$ and $0 < \alpha < 1$

$$\lim_{n\to\infty} P\left\{\sup_{x\in\mathcal{S}}\left|\hat{\pi}(x,\alpha) - \left[\Phi\left(z_{1-\alpha/2} - \frac{\beta(x)}{\sigma_{\hat{g}}(x)} - \Delta(x)\right) - \Phi\left(-z_{1-\alpha/2} - \frac{\beta(x)}{\sigma_{\hat{g}}(x)} - \Delta(x)\right)\right]\right| > \eta\right\} = 0 . \blacksquare$$

3.3 Coverage Probabilities of Confidence Bands

This section shows that the pointwise and uniform confidence bands constructed in Steps 5 and 6 of Section 2.1 have asymptotic coverage probabilities of at least $1-\alpha_0$. We use the following notation. Let $\alpha_0 \in (0,1/2)$. Define $\hat{\alpha}(x,\alpha_0)$ as in Step 5. Define $T(x,\alpha_0)$ as the solution in T to

$$\Phi\left[T - \frac{\beta(x)}{\sigma_{\hat{g}}(x)} - \Delta(x)\right] - \Phi\left[-T - \frac{\beta(x)}{\sigma_{\hat{g}}(x)} - \Delta(x)\right] = 1 - \alpha_0 .$$

Define

$$A(x,\alpha_0) = 2\{1 - \Phi[T(x,\alpha_0)]\}$$

As in (2.1), define and $\alpha(x,\alpha_0)$ to be the solution in a to

$$\Phi\left[z_{1-a/2} - \frac{\beta(x)}{\sigma_{\hat{g}}(x)}\right] - \Phi\left[-z_{1-a/2} - \frac{\beta(x)}{\sigma_{\hat{g}}(x)}\right] = 1 - \alpha_0 .$$

Let $\alpha_\xi(\alpha_0)$ denote the ξ-level quantile of points in the set $\{\alpha(x,\alpha_0): x \in \mathcal{S}\}$. Define $\mathcal{R}_\xi(\alpha_0)$ as in (2.2). The following corollary gives the asymptotic coverage probability of the pointwise confidence band constructed in Step 5.

Theorem 3.4: Let assumptions 1-5 hold. For all $C > 0$ and $\eta > 0$,

(i) $\lim_{n\to\infty} P\left[\sup_{x\in\mathcal{S},\,|\Delta(x)|\le C} \left|\hat{\alpha}(x,\alpha_0) - A(x,\alpha_0)\right| > \eta\right] = 0$.

(ii) $\lim_{n\to\infty} P\left[\hat{\alpha}_\xi(\alpha_0) \le \alpha_\xi(\alpha_0)\right] = 1$

(iii) $\liminf_{n\to\infty} P\left\{[x, g(x)] \in \mathcal{B}_n[\hat{\alpha}_\xi(\alpha_0)]\right\} \ge 1 - \alpha_0$ for each $x \in \mathcal{R}_\xi(\alpha_0)$.

Theorem 3.4(iii) shows that the pointwise band $\mathcal{B}_n[\hat{\alpha}_\xi(\alpha_0)]$ covers $g(x)$ with probability at least $1-\alpha_0$ except for a proportion ξ of points $x \in \mathcal{S}$.

Now consider the uniform confidence band constructed in Step 6 of Section 2.1. It follows from Theorem 3.1 that up to asymptotically negligible terms

$$\frac{(nh)^{1/2}[\hat{g}(x) - g(x)]}{\sigma_{\hat{g}}(x)} = \frac{\beta(x)}{\sigma_{\hat{g}}(x)} + W_1\left(\frac{x}{h_0}\right)$$

uniformly over $x \in \mathcal{S}$. Recall that $\lambda(x) = \beta(x)/\sigma_{\hat{g}}(x)$. If $\lambda(x)$ were known, an asymptotic uniform $1-\alpha_0$ confidence band for g would be

$$-t_U \le \lambda(x) + W_1\left(\frac{x}{h_0}\right) \le t_U\,,$$

where t_U is the solution in t to

$$P\left[-t - \lambda(x) \le W_1\left(\frac{x}{h_0}\right) \le t - \lambda(x)\;\; \forall\, x \in \mathcal{S}\right] = 1 - \alpha_0\,.$$

Define

$$\lambda_{\max} = \max_{x\in\mathcal{S}} \lambda(x)$$

and

$$\lambda_{\min} = \min_{x\in\mathcal{S}} \lambda(x)\,.$$

Then

$$P\left[-t_U - \lambda_{\min} \le W_1\left(\frac{x}{h_0}\right) \le t_U - \lambda_{\max}\;\; \forall\, x \in \mathcal{S}\right] \le P\left[-t_U - \lambda(x) \le W_1\left(\frac{x}{h_0}\right) \le t_U - \lambda(x)\;\; \forall\, x \in \mathcal{S}\right].$$

Therefore, asymptotically

$$(3.6)\qquad P\left[-t_U \le \frac{(nh)^{1/2}[\hat{g}(x) - g(x)]}{\sigma_{\hat{g}}(x)} \le t_U\;\; \forall x \in \mathcal{S}\right] \ge 1 - \alpha_0$$

if t_U is chosen so that

$$P\left[-t_U-\lambda_{\min}\le W_1\left(\frac{x}{h_0}\right)\le t_U-\lambda_{\max}\quad \forall\, x\in\mathcal{S}\right]=1-\alpha_0 .$$

The quantities $\lambda_{\max}$ and $\lambda_{\min}$ are unknown in applications. A feasible confidence band can be obtained by replacing them with the bootstrap estimates

$$\hat{\lambda}_{\max}=\max_{x\in\mathcal{S}}\hat{\lambda}(x)$$

and

$$\hat{\lambda}_{\min}=\min_{x\in\mathcal{S}}\hat{\lambda}(x) .$$

Similarly, the unknown quantities $\sigma_{\hat{g}}(x)$ and h_0 can be replaced with $\hat{\sigma}_{\hat{g}}(x)$ and h, respectively. The critical value t_U is replaced by $\hat{t}_U$, which is the solution in t to

(3.7) $$P\left[-t-\hat{\lambda}_{\min}\le W_1\left(\frac{x}{h}\right)\le t-\hat{\lambda}_{\max}\quad \forall\, x\in\mathcal{S}\right]=1-\alpha_0 ,$$

where $\hat{\lambda}_{\max}$ and $\hat{\lambda}_{\min}$ are treated as non-stochastic constants, not random variables, when calculating the probability on the right-hand side of (3.7). The resulting uniform confidence band is

$$-\hat{t}_U\le\frac{(nh)^{1/2}[\hat{g}(x)-g(x)]}{\hat{\sigma}_{\hat{g}}(x)}\le\hat{t}_U;\quad \forall x\in\mathcal{S} .$$

The following theorem establishes the asymptotic coverage probability of this interval.

Theorem 3.5: Let assumptions 1-6 hold. Then

$$\liminf_{n\to\infty} P\left[-\hat{t}_U\le\frac{(nh)^{1/2}[\hat{g}(x)-g(x)]}{\hat{\sigma}_{\hat{g}}(x)}\le\hat{t}_U;\quad \forall x\in\mathcal{S}\right]\ge 1-\alpha_0 .\ \blacksquare$$

The covariance function of W_1 can be estimated consistently. See equation (6.13) in the appendix. The probability on the right-hand side of (3.7) can be computed by simulation with arbitrary accuracy by replacing the covariance function of W_1 with the consistent estimate and the continuum $\mathcal{S}$ with a grid of equally spaced points.

4. **NUMERICAL EXPERIMENTS**

This section reports the results of a set of Monte Carlo experiments that illustrate the finite-sample performance of the method described in Section 2.

4.1 Design

Data $\{Y_i, X_i:\ i=1,...,n\}$ were generated from the models

$$Y=g_j(X)+\varepsilon;\quad j=1,2,3,$$

where

$$g_1(x) = x + 5\phi(10x),$$

$$g_2(x) = \frac{\sin(1.5\pi x)}{1 + 18x^2[\operatorname{sgn}(x) + 1]},$$

$$g_3(x) = \frac{\sin(1.5\pi x)}{1 + 2x^2[\operatorname{sgn}(x) + 1]},$$

ϕ is the standard normal probability density function, and $X \sim U[-1,1]$. The distribution of ε is $N(\mu_\tau, 1)$ for $\tau = 0.25$, $\tau = 0.5$, and $\tau = 0.75$, respectively, where $\mu_{0.25} = 0.6745$, $\mu_{0.50} = 0$, and $\mu_{0.75} = -0.6745$. Thus, $P(\varepsilon \le 0) = \tau$ for each value of τ. The functions g_j were used in numerical experiments by HH and other authors. Graphs of these functions are shown in Figure 1. The function g_1 has a sharp peak and is the most challenging for our method. The function g_2 is less challenging than g_1. The function g_3 is the smoothest and least challenging.

The sample sizes in the experiments were $n =$ 100, 500, and 1000. The kernel function was

$$K(v) = 0.75(1 - v^2) I(|v| \le 1).$$

The bandwidth h for local linear estimation of the g_js was chosen using the plug-in method of Yu and Jones (1998). Bandwidths for estimating $f_X(x)$ and $f_\varepsilon(0)$ were chosen by Silverman's rule of thumb. To avoid boundary effects, the set $\mathcal{S}$ was chosen so that its boundaries were at least one bandwidth from the boundaries of $[-1,1]$. This resulted in $\mathcal{S} = [-0.9, 0.9]$ for experiments with g_1 and g_2 and $\mathcal{S} = [-0.85, 0.85]$ for experiments with g_3. $\mathcal{S}$ is narrower for the experiments with g_3 because that function is smoother than g_1 and g_2 and has a larger bandwidth. We set $\alpha_0 = 0.05$ and $1 - \xi = 0.95$. Pointwise confidence bands were computed using an equally spaced grid of points $x \in \mathcal{S}$ with a spacing of 0.05. The grid spacing was 0.02 for uniform confidence bands. There were 1000 Monte Carlo replications in each experiment.

We also computed pointwise confidence bands using undersmoothing and the explicit bias correction method of Schucany and Sommers (1977). There are no satisfactory empirical methods for choosing an undersmoothing bandwidth or the auxiliary bandwidth required for explicit bias correction. Therefore, for undersmoothing, we set the bandwidth equal to $\gamma_1 h$, where h is the bandwidth selected by the method of Yu and Jones (1998) and $\gamma_1 \le 1$ is a constant. For explicit bias correction, we set the auxiliary bandwidth equal to $\hat{d} n^{-\gamma_2/5}$, where $\gamma_2 \le 1$ is a constant and $\hat{d}$ is as in assumption 6. The values of γ_1 and γ_2 were chosen to achieve coverage probabilities of at least 0.95 for as large a proportion of values of x in the grid as possible. This approach cannot be used in applications and gives an advantage to undersmoothing and explicit bias correction. Nonetheless, it will be seen in Section 4.2 that the performance of these methods is poor compared to that of the method of Section 2.1.

4.2 Results of the Experiments

Tables 1-3 show properties of pointwise confidence bands for $\tau = 0.25$, 0.50 and 0.75, respectively. At all quantiles and sample sizes, the bootstrap method described in Section 2.1 has much higher proportions of values of x for which the probability of covering of $g(x)$ exceeds 0.95 than do the undersmoothing and explicit bias correction methods. When $n = 100$, the bootstrap method's proportions are exceed 0.70 for $j = 1$ and 2, and 0.95 for $j = 3$. When $n = 1000$, the bootstrap method's proportions exceed 0.92 for all values of j. By contrast, the proportion of values of x for which undersmoothing achieves a coverage probability of at least 0.95 is below 0.65 for all values of n and j. The absolute error in the coverage probability (column 5 of Tables 1-3) is the absolute value of the difference between the actual coverage probability and the nominal probability of 0.95. Thus, the absolute error increases when the actual coverage probability exceeds 0.95 as well as when the actual coverage probability is less than 0.95.

The proportion of values of x for which explicit bias correction achieves a coverage probability of at least 0.95 is below 0.20 for all values of n and j. Undersmoothing and explicit bias correction perform poorly despite choosing the bandwidth for undersmoothing and the auxiliary bandwidth for explicit bias correction to achieve optimal performance of these methods.

The relatively low proportions of points at which the coverage probability of the bootstrap method equals or exceeds 0.95 for g_1 are due to the sharp peak of this function in the vicinity of $x = 0$, which causes the bias of $\hat{g}_1$ to be especially large. HH provide a theoretical explanation for why the bootstrap method performs poorly in regions of unusually high bias. The phenomenon is illustrated in Table 4, which shows the proportion of points for which the bootstrap confidence band covers $g_1(x)$ with probability exceeding 0.95 when the interval $[-0.05, 0.05]$ containing the peak is excluded from $\mathcal{S}$. The proportions of points for which the coverage probability equals or exceeds 0.95 is at least 0.94 when $n = 500$ or $n = 1000$ except for $n = 500$ and $\tau = 0.25$, when the proportion is 0.91. The function g_2 has peaks and troughs at $x = 0.15$ and $x = -0.35$, but these are not as sharp as the peak of g_1. Consequently, they have little effect on the coverage probabilities for g_2 when $n \geq 500$.

Table 5 shows the coverage probabilities of uniform confidence bands obtained with the bootstrap method. The coverage probabilities for g_2 and g_3 at all quantiles equal or exceed 0.95 if

$n \geq 500$ and 0.93 if $n = 100$. The coverage probabilities for g_1 with $n \geq 500$ equal or exceed 0.94 except for $n = 500$ and $\tau = 0.25$, when the coverage probability is 0.92.

5. CONCLUSIONS

This paper has described a bootstrap method for constructing pointwise and uniform confidence bands for a conditional quantile function that is estimated nonparametrically. The method is based on local polynomial estimation and uses only a bandwidth that can be selected using standard methods such as cross validation or plug-in. In contrast to other methods for constructing confidence bands, the bootstrap method does not require bandwidths that under- or oversmooth the nonparametric function estimator. This is an important advantage of the bootstrap method, because there are no satisfactory empirical methods for selecting bandwidths that under- or oversmooth a nonparametric estimator. The bootstrap method presented here is an extension of the method of Hall and Horowitz (2013) for conditional mean functions to conditional quantile functions and uniform confidence bands. The results of Monte Carlo experiments have illustrated the good finite-sample performance of the bootstrap method and the poor performance of methods based on under- or oversmoothing.

6. MATHEMATICAL APPENDIX: PROOFS OF THEOREMS

Assumptions 1-6 hold throughout this appendix. We use linear functional notation. For any function f

$$\mathcal{P}f = \int f(x)dP(x); \qquad \mathcal{P}_n f = \int f(x)dP_n(x),$$

where $\mathcal{P}$ and $\mathcal{P}_n$, respectively, are the distribution function and empirical distribution function of (Y, X).

6.1 Replacing the random bandwidth h with the non-stochastic bandwidth h_0.

For any bandwidth s and any $x \in \mathcal{S}$, define

$$Q_n(x, b_0, b_1, s) = n^{-1}\sum_{i=1}^{n} \rho_\tau[Y_i - b_0 - b_1(X_i - x)]K_s(X_i - x).$$

Let d_0 be as in assumption 6 and $h_0 = d_0 n^{-1/5}$. The following lemma shows that replacing the random bandwidth h with the non-stochastic bandwidth h_0 has an asymptotically negligible effect on the local linear estimator of $g(x)$.

Lemma 6.1: Define

$$(\hat{b}_{0h}, \hat{b}_{1h}) = \arg \min_{(b_0,b_1)\in\mathcal{G}} Q_n(x, b_0, b_1, h)$$

and

$$(\hat{b}_{0h_0}, \hat{b}_{1h_0}) = \arg \min_{(b_0,b_1)\in\mathcal{G}} Q_n(x, b_0, b_1, h_0) .$$

For each $x \in \mathcal{S}$ and $j = 0$ or 1, $\hat{b}_{jh} - \hat{b}_{jh_0} = o_p(n^{-2/5})$

<u>Proof</u>: Let D_1 be as in assumption 6(ii). Define

$$a_n = n^{-D_4} \left(\frac{\log n}{nh_0} \right)^{1/2} ,$$

where $0 < D_4 < D_1$. Let $\mathcal{O}$ be an open neighborhood of $(\hat{b}_{0h_0}, \hat{b}_{1h_0})$ such that

$$a_n = \min_{(b_0,b_1)\in\mathcal{G}\backslash\mathcal{O}} Q_n(x, b_0, b_1, h_0) - Q_n(x, \hat{b}_{0h_0}, \hat{b}_{1h_0}, h_0) .$$

The proof takes place in three steps. Step 1 shows that if

(6.1) $$\sup_{(b_0,b_1)\in\mathcal{G}} | Q_n(x, b_0, b_1, h) - Q_n(x, b_0, b_1, h_0) | < a_n / 2 ,$$

then $\hat{\boldsymbol{b}}_h \equiv (\hat{b}_{0h}, \hat{b}_{1h}) \in \mathcal{O}$. Step 2 shows that (6.1) holds with probability approaching 1 as $n \to \infty$. Step 3 shows that $\hat{b}_{0h} - \hat{b}_{0h_0} = o_p(n^{-2/5})$ if $\hat{\boldsymbol{b}}_h \in \mathcal{O}$.

Step 1: Let A_n be the event

$$\sup_{(b_0,b_1)\in\mathcal{G}} | Q_n(x, b_0, b_1, h) - Q_n(x, b_0, b_1, h_0) | < a_n / 2 .$$

Then

(6.2) $$A_n \Rightarrow Q_n(x, \hat{b}_{0h}, \hat{b}_{1h}, h) > Q_n(x, \hat{b}_{0h}, \hat{b}_{1h}, h_0) - a_n / 2$$

and

(6.3) $$A_n \Rightarrow Q_n(x, \hat{b}_{0h_0}, \hat{b}_{1h_0}, h_0) > Q_n(x, \hat{b}_{0h_0}, \hat{b}_{1h_0}, h) - a_n / 2 .$$

But $Q_n(x, \hat{b}_{0h}, \hat{b}_{1h}, h) \le Q_n(x, \hat{b}_{0h_0}, \hat{b}_{1h_0}, h)$. Therefore, it follows from (6.3) that

(6.4) $$A_n \Rightarrow Q_n(x, \hat{b}_{0h_0}, \hat{b}_{1h_0}, h_0) > Q_n(x, \hat{b}_{0h}, \hat{b}_{1h}, h) - a_n / 2 .$$

Substituting (6.2) into (6.4) yields

(6.5) $$A_n \Rightarrow Q_n(x, \hat{b}_{0h_0}, \hat{b}_{1h_0}, h_0) > Q_n(x, \hat{b}_{0h}, \hat{b}_{1h}, h_0) - a_n ,$$

Equivalently,

(6.6) $$A_n \Rightarrow Q_n(x, \hat{b}_{0h}, \hat{b}_{1h}, h_0) - Q_n(x, \hat{b}_{0h_0}, \hat{b}_{1h_0}, h_0) < a_n .$$

Therefore, $A_n \Rightarrow (\hat{b}_{0h}, \hat{b}_{1h}) \in \mathcal{O}$ because $Q_n(x, \hat{b}_{0h}, \hat{b}_{1h}, h_0) - Q_n(x, \hat{b}_{0h_0}, \hat{b}_{1h_0}, h_0) \ge a_n$ if $(\hat{b}_{0h}, \hat{b}_{1h}) \notin \mathcal{O}$.

Step 2: A Taylor series expansion of $Q_n(x,b_0,b_1,h)$ about $\hat{h}=h_0$ yields

$$Q_n(x,b_0,b_1,h)-Q_n(x,b_0,b_1,h_0)=\frac{h_0}{\tilde{h}}\frac{1}{nh_0}\sum_{i=1}^{n}\rho_\tau[Y_i-b_0-b_1(X_i-x)]K'\left(\frac{X_i-x}{\tilde{h}}\right)\left(\frac{h-h_0}{\tilde{h}}\right)$$

$$=\frac{h_0}{\tilde{h}^2}\mathcal{P}\left\{\rho_\tau[Y_i-b_0-b_1(X_i-x)]K'\left(\frac{X_i-x}{\tilde{h}}\right)\left(\frac{h-h_0}{h_0}\right)\right\}$$

$$+\frac{h_0}{\tilde{h}^2}(\mathcal{P}_n-\mathcal{P})\left\{\rho_\tau[Y_i-b_0-b_1(X_i-x)]K'\left(\frac{X_i-x}{\tilde{h}}\right)\left(\frac{h-h_0}{h_0}\right)\right\},$$

where $\tilde{h}$ is between h_0 and h. By assumption 6 and Theorem 2.37 of Pollard (1984)

$$\frac{h_0}{\tilde{h}^2}(\mathcal{P}_n-\mathcal{P})\left\{\rho_\tau[Y_i-b_0-b_1(X_i-x)]K'\left(\frac{X_i-x}{\tilde{h}}\right)\left(\frac{h-h_0}{h_0}\right)\right\}=^{a.s.}O\left[n^{-D_1}\left(\frac{(\log n)^{1+\varepsilon}}{nh_0}\right)^{1/2}\right]$$

for any $\varepsilon>0$. Standard calculations for kernel estimators combined with assumption 3 yield the result that

$$\frac{h_0}{\tilde{h}^2}\mathcal{P}\left\{\rho_\tau[Y_i-b_0-b_1(X_i-x)]K'\left(\frac{X_i-x}{\tilde{h}}\right)\left(\frac{h-h_0}{h_0}\right)\right\}=O\left(n^{-D_1}h_0^2\right).$$

Therefore,

$$|Q_n(x,b_0,b_1,h)-Q_n(x,b_0,b_1,h_0)|<a_n/2$$

almost surely for all sufficiently large n.

Step 3: It follows from Theorem 2.37 of Pollard (1984) that

$$Q_n(x,b_0,b_1,h_0)-Q_n(x,\hat{b}_{0h_0},\hat{b}_{1h_0},h_0)=^{a.s.}\ell(x)(b_0-\hat{b}_{0h_0})+o(1),$$

where ℓ is a non-zero function that does not depend on n. Therefore, $\hat{\boldsymbol{b}}_h\in\mathcal{O}$ implies that $\hat{b}_{0h}-\hat{b}_{0h_0}=o_p(n^{-2/5})$. Q.E.D.

6.2 Proofs of Theorem 3.1, Theorem 3.2, and Corollary 3.3

Proof of Theorem 3.1: The constraint $(b_0,b_1)\in\mathcal{G}$ in Lemma 6.1 is non-binding with probability approaching 1 as $n\to\infty$. Therefore, it suffices to consider the local linear estimator of $g(x)$ obtained in Section 2.1 with the non-stochastic bandwidth h_0 in place of h. Denote this estimator by $\hat{g}_x$. Denote the estimator of $g'(x)$ by g'_x. Let $g''_x=d^2g(x)/dx^2$. Then

(6.7) $$(\hat{g}_x, \hat{g}'_x) = \arg\min_{b_0,b_1} n^{-1}\sum_{i=1}^{n} \rho_\tau[Y_i - b_0 - b_1(X_i - x)]K_{h_0}(X_i - x)$$

$$= \arg\min_{b_0,b_1} \mathcal{P}_n\left\{\rho_\tau[Y - b_0 - b_1(X - x)]K_{h_0}(X - x)\right\}.$$

For each $x \in \mathcal{S}$, define $\boldsymbol{b}_x = (b_{x0}, b_{x1})'$ to be an arbitrary 2×1 vector. An argument similar to that used to prove Lemma A.2 of Ruppert and Carroll (1980) shows that he first-order conditions for (6.7) are

(6.8) $$\mathcal{P}_n\left\{\tau - I[Y - b_{x0} - b_{x1}(X - x) \le 0]\right\}K_{h_0}(X - x) = O_p(h_0 / n).$$

(6.9) $$\mathcal{P}_n\left\{\tau - I[Y - b_{x0} - b_{x1}(X - x) \le 0]\right\}(X - x)K_{h_0}(X - x) = O_p(h_0 / n).$$

As is shown below, the asymptotic form of $\hat{g}_x - g_x$ depends only on (6.8). Therefore, only (6.8) is treated in the remainder of the proof. Define

$$T_{n1}(x) = \mathcal{P}\left\{\tau - I[Y - g_x - g'_x(X - x) \le 0]\right\}K_{h_0}(X - x),$$

$$T_{n2}(x) = (\mathcal{P}_n - \mathcal{P})\left\{\tau - I[Y - g_x - g'_x(X - x) \le 0]\right\}K_{h_0}(X - x),$$

$$T_{n3}(x, \boldsymbol{b}_x) = \mathcal{P}\left\{I[Y - b_{x0} - b_{x1}(X - x) \le 0] - I[Y - g_x - g'_x(X - x) \le 0]\right\}K_{h_0}(X - x),$$

and

$$T_{n4}(x, \boldsymbol{b}_x) = (\mathcal{P}_n - \mathcal{P})\left\{I[Y - b_{x0} - b_{x1}(X - x) \le 0] - I[Y - g_x - g'_x(X - x) \le 0]\right\}K_{h_0}(X - x).$$

In these definitions, $\boldsymbol{b}_x$ is an arbitrary, non-stochastic vector. Then

$$\mathcal{P}_n\left\{\tau - I[Y - b_{x0} - b_{x1}(X - x) \le 0]\right\}K_{h_0}(X - x) = T_{n1}(x) + T_{n2}(x) + T_{n3}(x, \boldsymbol{b}_x) + T_{n4}(x, \boldsymbol{b}_x).$$

We now derive the asymptotic forms of T_{n1}, T_{n2}, T_{n3}, and T_{n4}.

Analysis of $T_{n1}(x)$: Let F_ε denote the distribution function of ε in (1.1). Then a Taylor series expansion yields

$$T_{n1}(x) = \tau - \int F_\varepsilon[g(z) - g_x - g'_x(z - x)]K_{h_0}(z - x) f_X(z)dz$$

(6.10) $$= \frac{h_0^3\kappa_2}{2} f_X(x) f_\varepsilon(0) g''_x + O\left(h_0^4\right)$$

uniformly over $x \in \mathcal{S}$.

Analysis of $T_{n2}(x)$: Define

$$T_{n2a}(x) = \frac{\tau}{n}\sum_{i=1}^{n}\left\{K_{h_0}(X_i - x) - E[K_{h_0}(X_i - x)]\right\},$$

$$T_{n2b}(x) = n^{-1}\sum_{i=1}^{n}\{I(\varepsilon_i \le 0)K_{h_0}(X_i - x) - E[I(\varepsilon_i \le 0)K_{h_0}(X_i - x)]\},$$

and

$$T_{n2c}(x) = \frac{\tau}{n}\sum_{i=1}^{n}\Big(\{I[\varepsilon_i \le g_x - g_{X_i} + g'_x(X_i - x)] - I(\varepsilon_i \le 0)\}K_{h_0}(X_i - x)\}$$

$$-E\{I[\varepsilon_i \le g_x - g_{X_i} + g'_x(X_i - x)] - I(\varepsilon_i \le 0)\}K_{h_0}(X_i - x)\}\Big).$$

Then

$$T_{n2} = T_{n2a} + T_{n2b} + T_{n2c}.$$

Let F_X and F_{nX} and, respectively, denote the distribution and empirical distributions functions of X. Define the stochastic process $Z_{nx}(x) = n^{1/2}[F_{nX}(x) - F_X(x)]$. Define the limit process $Z_x^0(x)$ by $Z_{nx}(x) \rightsquigarrow Z_x^0(x)$ as $n \to \infty$. Then a change of variables and integration by parts yields

$$n^{1/2}T_{n2a}(x) = \tau\int K\left(\frac{v - x}{h_0}\right)dZ_{nx}(v)$$

$$= -\tau\int Z_{nx}(x + h_0\xi)K'(\xi)d\xi$$

$$= -\tau\int Z_x^0(x + h_0\xi)K'(\xi)d\xi - \tau\int[Z_{nx}(x + h_0\xi) - Z_x^0(x + h_0\xi)]K'(\xi)d\xi$$

$$\equiv n^{1/2}T_{n2a}^{(1)}(x, h_0) + n^{1/2}T_{n2a}^{(2)}(x, h_0).$$

It follows from Theorem 3 of Komlós, Major, and Tusnády (1975) that there are processes $\tilde{Z}_{nx}$ and $\tilde{Z}_x^0$ having the same distributions as Z_{nx} and Z_x^0 such that

$$P\left[\sup_{x\in\mathcal{S}}|\tilde{Z}_{nx}(x) - \tilde{Z}_x^0(x)| > C_1 n^{-1/2}\log n\right] < n^{-C_2},$$

where C_1 and C_2 are constants. Therefore, $(n/h_0)^{1/2}T_{n2a}(x)$ can be approximated by the mean-zero Gaussian process $(n/h_0)^{1/2}T_{n2a}^{(1)}(x, h_0)$ in the sense that

(6.10) $$P\left[\sup_{x\in\mathcal{S}}|(n/h_0)^{1/2}T_{n2a}(x) - (n/h_0)^{1/2}T_{n2a}^{(1)}(x, h_0)| > C_3(nh_0)^{-1/2}\log n\right] < n^{-C_2}.$$

Now consider T_{n2b}. Let $F_{\varepsilon x}$ denote the distribution function of (ε, X) and $F_{n\varepsilon x}$ denote the empirical distribution function. Define $Z_{n\varepsilon x}(\varepsilon, x) = n^{1/2}[F_{n\varepsilon x}(\varepsilon, x) - F_{\varepsilon x}(\varepsilon, x)]$, and let $Z_{\varepsilon x}^0(\varepsilon, x)$ denote the limiting Gaussian process of $Z_{n\varepsilon x}(\varepsilon, x)$. Integration by parts and a change of variables yields

$$n^{1/2}T_{n2b}(x) = \int I(s \le 0) K\left(\frac{v-x}{h_0}\right) dZ_{n\varepsilon x}(s,v)$$

$$= -\int Z_{\varepsilon x}^0(0, x+h_0\xi)K'(\xi)d\xi - \int [Z_{n\varepsilon x}(0, x+h_0\xi) - Z_{\varepsilon x}^0(0, x+h_0\xi)]K'(\xi)d\xi$$

$$\equiv n^{1/2}T_{n2b}^{(1)}(x,h_0) + n^{1/2}T_{n2b}^{(2)}(x,h_0).$$

To bound $n^{1/2}T_{n2b}^{(2)}$, let $\hat{n}_0$ denote the number of observations for which $\varepsilon_i \le 0$. Assume without loss of generality that these are the first $\hat{n}_0$ observations. The corresponding X_i 's are a random sample of X, because X and ε are independent. We have

$$F_{n\varepsilon x}(0,x) = n^{-1}\sum_{i=1}^{n} I(\varepsilon_i \le 0) I(X_i \le x)$$

$$= \frac{\hat{n}_0}{n}\hat{n}_0^{-1}\sum_{i=1}^{\hat{n}_0} I(X_i \le x)$$

$$\equiv \frac{\hat{n}_0}{n}F_{\hat{n}_0 x}(x).$$

Moreover, because ε and X are independent,

$$F_{n\varepsilon x}(0,x) - F_{\varepsilon x}(0,x) = \frac{\hat{n}_0}{n}F_{\hat{n}_0 x}(x) - F_\varepsilon(0)F_x(x)$$

$$= F_{n\varepsilon}(0)[F_{\hat{n}_0 x}(x) - F_x(x)] + \left[\frac{\hat{n}_0}{n} - F_\varepsilon(0)\right]F_{\hat{n}_0 x}(x)$$

$$= F_\varepsilon(0)\hat{n}_0^{-1/2}Z_x^0(x) + n^{-1/2}\xi F_{\hat{n}_0}(x) + F_\varepsilon(0)\hat{n}_0^{-1/2}r_{n1}(x) + F_{\hat{n}_0 x}(0)n^{-1/2}r_{n2},$$

where $\xi \sim N(0, V_\xi)$, $V_\xi = F_\varepsilon(0)[1 - F_\varepsilon(0)]$,

$$n^{1/2}\left[\frac{\hat{n}_0}{n} - F_\varepsilon(0)\right] \to^d \xi ,$$

$$r_{n1}(x) = \hat{n}_0^{1/2}[F_{\hat{n}_0 x}(x) - F_x(x)] - Z_x^0(x) ,$$

and

$$r_{n2} = n^{1/2}\left[\frac{\hat{n}_0}{n} - F_U(0)\right] - \xi .$$

By Theorem 3 of Komlós, Major, and Tusnády (1975), there are a version $\tilde{r}_{n1}(x)$ of $r_{n1}(x)$ and constants C_4 and C_5 such that

$$P\left(\sup_{x\in\mathcal{S}} |\tilde{r}_{n1}(x)| > C_4 n^{-1/2}\log n\right) < n^{-C_5}.$$

A similar result applies to $r_{n2}(x)$. Therefore, $(n/h_0)^{1/2} T_{n2b}(x)$ can be approximated by the mean-zero Gaussian process $(n/h_0)^{1/2} T_{n2b}^{(1)}(x,h_0)$ in the sense that there are finite constants C_6 and C_7 such that

$$(6.11)\quad P\left[\sup_{x\in\mathcal{S}} |(n/h_0)^{1/2} T_{n2b}(x) - (n/h_0)^{1/2} T_{n2b}^{(1)}(x,h_0)| > C_6 (nh_0)^{-1/2}\log n\right] < n^{-C_7}.$$

It follows from Theorem 2.37 of Pollard (1984) that

$$(6.12)\quad \left(\frac{n}{h_0}\right)^{1/2} T_{n2c}(x) =^{a.s.} o[h_0^{3/2}(\log n)^{1/2+\delta}]$$

uniformly over $x\in\mathcal{S}$ for any $\delta > 0$. Combining (6.10)-(6.12) yields the result that $(n/h_0)^{1/2} T_{n2}(x)$ can be approximated by the mean-zero Gaussian process $(n/h_0)^{1/2}[T_{n2a}^{(1)}(x,h_0) + T_{n2b}^{(1)}(x,h_0)]$. The sample paths of this process are uniformly continuous in h_0 (Dudley 1967). A straightforward but lengthy calculation shows that the covariance function of this process converges to

$$(6.13)\quad C(x_1,x_2) = \left(\frac{1-\tau}{\tau}\right) f_X(x_1)\int K(\zeta)K(\zeta+\delta)d\zeta,$$

where $\delta = (x_1 - x_2)/h_0$. Let $W_1(\cdot)$ denote the mean-zero Gaussian process whose covariance function is $C(x_1,x_2)/C(x_1,x_1)$. Then it follows from Theorem 5.8 of Boucheron, Lugosi, and Massart (2013) and criterion B of Loève (1978, p. 268) that for any $\eta > 0$

$$(6.14)\quad \lim_{n\to\infty} P\left\{\sup_{x\in\mathcal{S}}\left|\left(\frac{n}{h_0}\right)^{1/2} T_{n2}(x) - \left[f_X(x)\left(\frac{1-\tau}{\tau}\right)B_K\right]^{1/2} W_1\left(\frac{x}{h_0}\right)\right| > \eta\right\} = 0.$$

Analysis of T_{n3}. We have

$$T_{n3}(x,\boldsymbol{b}_x) = \mathcal{P}\left\{I[Y - b_{x0} - b_{x1}(X-x) \le 0] - I[Y - g_x - g_x'(X-x) \le 0]\right\} K_{h_0}(X-x)$$

$$(6.14)\quad = \int \{F_\varepsilon[b_{x0} - g_z + b_{x1}(z-x)] - F_\varepsilon[g_x - g_z + g_x'(z-x)]\} K_{h_0}(z-x) f_X(z)dz.$$

Suppose there is a constant $C_1 < \infty$ such that

$$(6.15)\quad \sup_{x\in\mathcal{S}} |b_{x0} - g_x| \le C\left(\frac{\log n}{nh_0}\right)^{1/2}$$

and

(6.16) $\sup_{x\in\mathcal{S}} | b_{x1} - g'_x | \le C_2 \left(\frac{\log n}{nh_0^3} \right)^{1/2}$.

Define

$\mathcal{V} = \{ \boldsymbol{b}_x :$ (6.15) and (6.16) hold for all $x \in \mathcal{S} \}$.

Then the change of variables $\xi = (z - x) / h_0$ and Taylor series expansions about zero of the F_ε terms in the integral on the right-hand side of (6.15) yield

(6.17) $\sup_{x\in\mathcal{S};\, \boldsymbol{b}_x \in\mathcal{V}} | T_{n3}(x, \boldsymbol{b}_x) - h_0 f_\varepsilon(0) f_X(x)(b_0 - g_x) | \le C_2 \left(\frac{\log n}{n} \right)$

for some constant $C_2 < \infty$ and all sufficiently large n. It follows from Proposition 2 of Guerre and Sabbah (2012) that

(6.18) $\sup_{x\in\mathcal{S}} | \hat{g}_x - g_x | = O_p \left[\left(\frac{\log n}{nh_0} \right)^{1/2} \right]$

and

(6.19) $\sup_{x\in\mathcal{S}} | \hat{g}'_x - g'_x | = O_p \left[\left(\frac{\log n}{nh_0^3} \right)^{1/2} \right]$.

Let $\hat{\boldsymbol{b}}_x = (\hat{g}_x, \hat{g}'_x)$. Then (6.17)-(6.19) imply that

(6.20) $\sup_{x\in\mathcal{S}} | T_{n3}(x, \hat{\boldsymbol{b}}_x) - h_0 f_\varepsilon(0) f_X(x)(\hat{b}_0 - g_x) | = O_p \left[\left(\frac{\log n}{n} \right) \right]$.

Analysis of T_{n4}. We have

$$\left\{ I[Y - b_{x0} - b_{x1}(X - x) \le 0] - I[Y - g_x - g'_x(X - x) \le 0] \right\} K_{h_0}(X - x)$$

$$= \{ I[g_x - g_X + g'_x(X - x) < \varepsilon \le b_{x0} - g_X - b_{x1}(X - x)]$$

$$- I[b_{x0} - g_X - b_{x1}(X - x) < \varepsilon \le g_x - g_X + g'_x(X - x)]\} K_{h_0}(X - x).$$

Let (6.15) and (6.16) hold. Then,

$$\mathcal{P}\left(\left\{ I[Y - b_{x0} - b_{x1}(X - x) \le 0] - I[Y - g_x - g'_x(X - x) \le 0] \right\}^2 K_{h_0}(X - x)^2 \right) \le C \left(\frac{h_0 \log n}{n} \right)$$

for some $C < \infty$ and all sufficiently large n. It follows from Theorem (2.37) of Pollard (1984) that

$$\sup_{x\in\mathcal{S},\, \boldsymbol{b}_x\in\mathcal{V}} | T_{n4}(x, \boldsymbol{b}_x) | \ll h_0^{1/4} \left(\frac{\log n}{n} \right)^{3/4}$$

almost surely. Therefore, it follows from (6.18) and (6.19) that

(6.21) $$T_{n4}(x,\hat{\boldsymbol{b}}_x) = O_p\left[h_0^{1/4}\left(\frac{\log n}{n}\right)^{3/4}\right].$$

Now combine (6.10), (6.14), (6.20), and (6.21) to obtain

$$\mathcal{P}_n\left\{\tau - I[Y - \hat{b}_{x0} - \hat{b}_{x1}(X-x) \le 0]\right\}K_{h_0}(X-x) = \frac{h_0^3\kappa_2}{2} f_X(x) f_\varepsilon(0) g_x''$$

$$+\left(\frac{h_0}{n}\right)^{1/2}\left[f_X(x)\left(\frac{1-\tau}{\tau}\right)B_K\right]^{1/2} W_1\left(\frac{x}{h_0}\right) + h_0 f_\varepsilon(0) f_X(x)(\hat{g}_x - g_x)| + O\left(h_0^4\right)$$

$$+O_p\left[\left(\frac{\log n}{n}\right)\right] + o_p\left[\left(\frac{h_0}{n}\right)^{1/2}\right] + O_p\left[h_0^{1/4}\left(\frac{\log n}{n}\right)^{3/4}\right]$$

(6.22) $$= \frac{h_0^3\kappa_2}{2} f_X(x) f_\varepsilon(0) g_x'') + \left(\frac{h_0}{n}\right)^{1/2}\left[f_X(x)\left(\frac{1-\tau}{\tau}\right)B_K\right]^{1/2} W_1\left(\frac{x}{h_0}\right) + h_0 f_\varepsilon(0) f_X(x)(\hat{g}_x - g_x)| + o_p\left[\left(\frac{h_0}{n}\right)^{1/2}\right]$$

uniformly over $x \in \mathcal{S}$. The theorem follows from combining (6.8) and (6.22). Q.E.D.

Proof of Theorem 3.2: Define $\tilde{T}_n(x) = E^* T_{nb}^*(x)$. By definition,

$$\hat{\varepsilon}_j = \tilde{\varepsilon}_j - q_n = \varepsilon_i + g_{X_j} - \hat{g}_{X_j} - q_n .$$

Then conditional on the original data,

$$\tilde{T}_n(x) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}[1-\tau^{-1}I(\varepsilon_j \le \hat{g}_{X_j} - g_{X_j} + \hat{g}_x - \hat{g}_X + \hat{g}_x'(X_i - x) + q_n)]K_{h_0}(X_i - x) .$$

By construction,

$$n^{-1}\sum_{j=1}^{n}[1-\tau^{-1}I(\hat{\varepsilon}_j \le 0)] = O_p(n^{-1}) ,$$

so

$$\tilde{T}_n(x) = -\frac{1}{\tau n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{I[\varepsilon_j \le \hat{g}_{X_j} - g_{X_j} + \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n)]$$

$$- I(\varepsilon_j \le \hat{g}_{X_j} - g_{X_j} + q_n)\}K_{h_0}(X_i - x) + O_p(n^{-1})$$

$$= -\frac{1}{\tau n}\sum_{i=1}^{n}\mathcal{P}_{n(\varepsilon,Z)}\{-I(\varepsilon \le \hat{g}_Z - g_Z + \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n)]$$

$$- I(\varepsilon \le \hat{g}_Z - g_Z + q_n)\}K_{h_0}(X_i - x) + O_p(n^{-1}),$$

where $\mathcal{P}_{n(\varepsilon,Z)}$ is the empirical measure of (ε, X) and this notation is used instead of $\mathcal{P}_{n(\varepsilon,X)}$ to avoid confusion with the data $\{X_j: j = 1,...,n\}$. Define

$$v = \tilde{\varepsilon} - (\hat{g}_Z - g_Z),$$

and let $\mathcal{P}_v$ and $\mathcal{P}_{nv}$, respectively, denote the population and empirical measures of v. Then,

$$\tilde{T}_n(x) = -\frac{1}{\tau n}\sum_{i=1}^{n}\mathcal{P}_{nv}\{-I(v \le \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n)] - I(v \le q_n)\}K_{h_0}(X_i - x) + O_p(n^{-1})$$

(6.23) $$= -\frac{1}{\tau n}\sum_{i=1}^{n}\mathcal{P}_v\{-I(v \le \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n)] - I(v \le q_n)\}K_{h_0}(X_i - x)$$

$$+ \frac{1}{\tau n}\sum_{i=1}^{n}(\mathcal{P}_{nv} - \mathcal{P}_v)\{-I(v \le \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n)] - I(v \le q_n)\}K_{h_0}(X_i - x) + O_p(n^{-1}).$$

The summands on the right-hand side of (6.23) are non-zero only if $|X_i - x| \le h_n$, $g(x)$ is continuous, and $\hat{g}(x) - g(x) \to^p 0$ uniformly over $x \in \mathcal{S}$ by Proposition 2 of Guerre and Sabbah (2012). In addition, the empirical process $\varphi_n(t) = (\mathcal{P}_{nv} - \mathcal{P}_v)I(v \le t)$ is stochastically equicontinuous. Therefore, the second term on the right-hand side of (6.23) is $O_p(h_0 n^{-1/2})$, and

$$\tilde{T}_n(x) = -\frac{1}{\tau n}\sum_{i=1}^{n}\mathcal{P}_v\{-I(v \le \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n)] - I(v \le q_n)\}K_{h_0}(X_i - x) + O_p(h_0 n^{-1/2}).$$

Because ε and X are independent,

(6.24) $$\tilde{T}_n(x) = -\frac{1}{\tau n}\sum_{i=1}^{n}\mathcal{P}_Z\mathcal{P}_\varepsilon\{-I(\varepsilon \le \hat{g}_Z - g_Z + \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n)]$$

$$- I(\varepsilon \le \hat{g}_Z - g_Z + q_n)\}K_{h_0}(X_i - x) + O_p(h_0 n^{-1/2}).$$

Define

$$\hat{A}_1(x, X_i, Z) = \hat{g}_Z - g_Z + \hat{g}_x - \hat{g}_{X_i} + \hat{g}'_x(X_i - x) + q_n$$

and

$$\hat{A}_2(Z) = \hat{g}_Z - g_Z + q_n .$$

We have $|\hat{g}(x) - g(x)| = O_p[(nh_0)^{-1/2}(\log n)^{1/2}]$ and $|\hat{g}'(x) - g'(x)| = O_p[(nh_0^3)^{-1/2}(\log n)^{1/2}]$ uniformly over $x \in \mathcal{S}$. Moreover, in the summand on the right-hand side of (6.24), only terms for which $|X_i - x| \le h_0$ are non-zero. Therefore, arguments like those used to obtain (6.24) show that

$$\tilde{T}_n(x) = -\frac{f_\varepsilon(0)}{\tau n}\sum_{i=1}^{n}\mathcal{P}_Z[\hat{A}_1(x, X_i, Z) - \hat{A}_2(Z)]K_{h_0}(X_i - x) + O_p(h_0 n^{-1/2})$$

$$\equiv \tilde{T}_{na}(x) + \tilde{T}_{nb}(x) + O_p(h_0 n^{-1/2}),$$

where

$$\tilde{T}_{na}(x) = -\frac{f_\varepsilon(0)}{\tau n}\sum_{i=1}^{n}[g_x - g_{X_i} + g'_x(X_i - x)]K_{h_0}(X_i - x)$$

and

$$\tilde{T}_{nb}(x) = -\frac{f_\varepsilon(0)}{\tau n}\sum_{i=1}^{n}[(\hat{g}_x - g_x) - (\hat{g}_{X_i} - g_{X_i}) + (\hat{g}'_x - g'_x)(X_i - x)]K_{h_0}(X_i - x) .$$

Standard calculations for kernel estimators show that

$$\tilde{T}_{na}(x) = -\frac{\kappa_2 h_0^3 f_\varepsilon(0) g''(x) f_X(x)}{2\tau} + O(h_0^4) ,$$

uniformly over $x \in \mathcal{S}$ and

$$\left(\frac{n}{h_0}\right)^{1/2}\tilde{T}_{na} = -\frac{\kappa_2 d_0^{5/2} f_\varepsilon(0) g''(x) f_X(x)}{2\tau} + O(h_0) \tag{6.25}$$

uniformly over $x \in \mathcal{S}$.

Now consider $\tilde{T}_{nb}(x)$. It follows from Theorem 3.1 that

$$(\hat{g}_x - g_x) - (\hat{g}_{X_i} - g_{X_i}) = \frac{d_0^{5/2}\kappa_2}{2}(nh_0)^{-1/2}[g''(x) - g''(X_i)] + (nh_0)^{-1/2}\psi_0(x)W_1\left(\frac{x}{h_0}\right)$$

$$- (nh_0)^{-1/2}\psi_0(X_i)W_1\left(\frac{X_i}{h_0}\right) + o_p[(nh_0)^{-1/2}].$$

Combining this result with assumption 4 yields

$$-\frac{f_\varepsilon(0)}{\tau n}\sum_{i=1}^{n}[(\hat{g}_x-g_x)-(\hat{g}_{X_i}-g_{X_i})]K_{h_0}(X_i-x)=-\frac{f_\varepsilon(0)}{\tau n}(nh_0)^{-1/2}\psi_0(x)W_1\left(\frac{x}{h_0}\right)\sum_{i=1}^{n}K_{h_0}(X_i-x)$$

$$+\frac{f_\varepsilon(0)}{\tau n}(nh_0)^{-1/2}\sum_{i=1}^{n}\psi_0(X_i)W_1\left(\frac{X_i}{h_0}\right)K_{h_0}(X_i-x)+o_p\left[\left(\frac{h_0}{n}\right)^{1/2}\right]$$

$$=-\left(\frac{h_0}{n}\right)^{1/2}\frac{f_\varepsilon(0)f_X(x)\psi_0(x)}{\tau}W_1\left(\frac{x}{h_0}\right)+o_p\left[\left(\frac{h_0}{n}\right)^{1/2}\right]$$

uniformly over $x\in\mathcal{S}$. Therefore,

$$\tilde{T}_{nb}(x)=-\left(\frac{h_0}{n}\right)^{1/2}\frac{f_\varepsilon(0)f_X(x)\psi_0(x)}{\tau}W_1\left(\frac{x}{h_0}\right)$$

$$-\frac{f_\varepsilon(0)}{\tau n}(\hat{g}'_x-g'_x)\sum_{i=1}^{n}(X_i-x)K_{h_0}(X_i-x)+o_p\left[\left(\frac{h_0}{n}\right)^{1/2}\right]$$

uniformly over $x\in\mathcal{S}$. In addition, $|\hat{g}'(x)-g'(x)|=O_p[(nh_0^3)^{-1/2}(\log n)^{1/2}]$ uniformly over $x\in\mathcal{S}$, and

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-x)K_{h_0}(X_i-x)\ll h_0^{3/2}\left(\frac{\log n}{n}\right)^{1/2}=o\left[\left(\frac{h_0}{n}\right)^{1/2}\right]$$

almost surely uniformly over $x\in\mathcal{S}$ by Theorem 2.37 of Pollard (1984). Therefore,

$$\tilde{T}_{nb}(x)=-\left(\frac{h_0}{n}\right)^{1/2}\frac{f_\varepsilon(0)f_X(x)\psi_0(x)}{\tau}W_1\left(\frac{x}{h_0}\right)+o_p\left[\left(\frac{h_0}{n}\right)^{1/2}\right] \tag{6.26}$$

uniformly over $x\in\mathcal{S}$. Combining (6.25) and (6.26) yields

$$\left(\frac{n}{h_0}\right)^{1/2}\tilde{T}_n=\frac{\kappa_2 d_0^{5/2}f_\varepsilon(0)f_X(x)}{2\tau}g''(x)-\frac{f_\varepsilon(0)f_X(x)\psi_0(x)}{\tau}W_1\left(\frac{x}{h_0}\right)+o_p(1)$$

$$=\frac{A_1(x)\kappa_2}{2}g''(x)-A_2(x)W_1\left(\frac{x}{h_0}\right)+o_p(1)$$

uniformly over $x\in\mathcal{S}$. This proves part (i) of the theorem. Part (ii) is an immediate consequence of part (i), uniform consistency of $\hat{f}_X(x)$, and consistency of $\hat{f}_\varepsilon(0)$. Q.E.D.

Proof of Corollary 3.3: A Taylor series expansion of $\hat{\pi}(x,\alpha)$ about $\hat{\lambda}(x)=\lambda(x)$ yields

$$\hat{\pi}(x,\alpha)=\Phi[z_{1-\alpha/2}-\lambda(x)-\Delta(x)]-\Phi[-z_{1-\alpha/2}-\lambda(x)-\Delta(x)]+r_n(x)[\hat{\lambda}(x)-\lambda(x)],$$

where

$$r_n(x) = [\{\phi[z_{1-\alpha/2} - \tilde{\lambda}_1(x) + \Delta(x)] - \phi[-z_{1-\alpha/2} + \tilde{\lambda}_2(x) - \Delta(x)]\}$$

and $\tilde{\lambda}_1(x)$ and $\tilde{\lambda}_2(x)$ are between $\hat{\lambda}(x)$ and $\lambda(x)$. The corollary now follows from Theorem 3.2(ii) and boundedness of $r_n(x)$. Q.E.D.

6.3 Proofs of Theorems 3.4 and 3.5

Proof of Theorem 3.4: Part (i) follows from Corollary 3.3. The process $\Delta(\cdot)$ is a non-stochastic multiple of W_1 and has uniformly continuous sample paths (Dudley 1967). Parts (ii) and (iii) of the theorem follow from arguments identical to those used to prove results (4.12) and (4.13) of HH. Q.E.D.

Proof of Theorem 3.5: It suffices to show that asymptotically, $\hat{\lambda}_{\max} \geq \max_{x\in\mathcal{S}} \lambda(x)$ and $\hat{\lambda}_{\min} \leq \min_{x\in\mathcal{S}} \lambda(x)$. We prove that $\hat{\lambda}_{\max} \geq \max_{x\in\mathcal{S}} \lambda(x)$ asymptotically. The proof for $\hat{\lambda}_{\min}$ is similar.

To show that $\hat{\lambda}_{\max} \geq \max_{x\in\mathcal{S}} \lambda(x)$ asymptotically, observe that $\lambda(x)$ is a continuous function on the compact interval $\mathcal{S}$. Therefore, there is a point $x^* \in \mathcal{S}$ such that $\max_{x\in\mathcal{S}} \lambda(x) = \lambda(x^*)$. Assume that x^* is unique. The proof for a unique x^* holds with minor modifications if x^* is not unique. Given any $\varepsilon > 0$, choose $\delta > 0$ so that

$$|\lambda(x) - \lambda(x^*)| < \varepsilon$$

whenever

$$|x - x^*| \leq \delta .$$

Because

$$\sup_{x\in\mathcal{S}} \hat{\lambda}(x) \geq \sup_{|x-x^*|\leq\delta} \hat{\lambda}(x) ,$$

it suffices to show that

$$\sup_{|x-x^*|\leq\delta} \hat{\lambda}(x) \geq \lambda(x^*) .$$

By Theorem 3.2(ii)

$$\hat{\lambda}(x) = \frac{\beta(x)}{\sigma_{\hat{g}}(x)} + \Delta(x) .$$

If $x \in [x^* - \delta, x^* + \delta]$, then

$$\Delta(x) > \varepsilon \Rightarrow \hat{\lambda}(x) = \frac{\beta(x)}{\sigma_{\hat{g}}(x)} + \Delta(x) > \frac{\beta(x)}{\sigma_{\hat{g}}(x)} + \varepsilon > \frac{\beta(x^*)}{\sigma_{\hat{g}}(x^*)} = \lambda(x^*) .$$

Therefore, $\Delta(x) > \varepsilon$ for some $x \in [x^* - \delta, x^* + \delta]$ implies that $\hat{\lambda}_{\max} \geq \lambda(x^*)$. To prove that $\Delta(x) > \varepsilon$ for some $x \in [x^* - \delta, x^* + \delta]$, let $x_0, ..., x_{J_n}$ be a set of points such that

$$x^* - \delta = x_0 < x_1 < ... < x_{J_n} = x^* + \delta .$$

Let $x_j - x_{j-1} > 2h_0$ for each $j = 1,..., J_n$ and $J_n \to \infty$ as $n \to \infty$. This is possible because $h_0 \to 0$ and δ remains fixed as $n \to \infty$. Then $\Delta(x_0),...,\Delta(x_{J_n})$ are independent random variables that are normally distributed with means of 0 and variances that are bounded away from 0 as $n \to \infty$. Let $Var[\Delta(x)] \geq \sigma_{\min}^2 > 0$, Then as $n \to \infty$,

$$P\left\{ \bigcap_{j=0}^{J_n} [\Delta(x_j) \leq \varepsilon] \right\} \leq [\Phi(\varepsilon / \sigma_{\min})]^{J_n+1} \to 0 ,$$

and

$$P[\hat{\lambda}_{\max} \geq \lambda(x^*)] \to 1 .$$

TABLE 1: SIMULATION RESULTS FOR $\tau = 0.25$

Method	n	j	Prop. with Cov. Prob. ≥ 0.95	Av. Abs. Error of Cov. Prob.	Av. Width
Bootstrap method of Sec. 2.1	100	1	0.76	0.034	1.76
		2	0.73	0.022	1.57
		3	0.97	0.024	1.26
	500	1	0.89	0.049	1.01
		2	1.0	0.033	0.83
		3	1.0	0.028	0.59
	1000	1	0.92	0.051	0.80
		2	0.95	0.032	0.62
		3	0.94	0.026	0.44
Undersmooth	100	1	0	0.097	1.24
		2	0	0.07	1.18
		3	0.03	0.04	1.16
	500	1	0.30	0.065	0.68
		2	0.22	0.016	0.71
		3	0.35	0.01	0.66
	1000	1	0.32	0.062	0.52
		2	0.41	0.014	0.57
		3	0.40	0.009	0.56
Bias Corr.	100	1	0	0.18	1.38
		2	0	0.19	1.37
		3	0	0.15	1.31
	500	1	0	0.11	1.07
		2	0.08	0.052	0.91
		3	0.06	0.036	0.078
	1000	1	0	0.068	0.88
		2	0.08	0.14	0.81
		3	0.09	0.018	0.059

TABLE 2: SIMULATION RESULTS FOR $\tau = 0.50$

Method	n	j	Prop. with Cov. Prob. ≥ 0.95	Av. Abs. Error of Cov. Prob.	Av. Width
Bootstrap method of Sec. 2.1	100	1	0.73	0.034	1.64
		2	0.76	0.022	1.48
		3	0.97	0.024	1.17
	500	1	0.89	0.048	0.95
		2	1	0.035	0.79
		3	0.95	0.027	0.56
	1000	1	0.92	0.056	0.74
		2	0.97	0.035	0.58
		3	0.94	0.029	0.41
Undersmooth	100	1	0	0.075	1.20
		2	0	0.053	1.19
		3	0.06	0.026	1.17
	500	1	0.27	0.021	0.78
		2	0.41	0.016	0.62
		3	0.51	0.094	0.57
	1000	1	0.49	0.025	0.56
		2	0.51	0.011	0.51
		3	0.63	0.099	0.43
Bias Corr.	100	1	0	0.14	1.38
		2	0	0.12	1.35
		3	0	0.092	1.26
	500	1	0	0.069	1.01
		2	0.11	0.028	0.86
		3	0.17	0.019	0.69
	1000	1	0.05	0.038	0.80
		2	0.11	0.025	0.63
		3	0.20	0.013	0.51

TABLE 3: SIMULATION RESULTS FOR $\tau = 0.75$

Method	n	j	Prop. with Cov. Prob. ≥ 0.95	Av. Abs. Error of Cov. Prob.	Av. Width
Bootstrap method of Sec. 2.1	100	1	0.76	0.032	1.74
		2	0.89	0.023	1.58
		3	0.94	0.020	1.20
	500	1	0.86	0.045	0.98
		2	1	0.033	0.78
		3	0.91	0.022	0.57
	1000	1	0.86	0.046	0.76
		2	0.97	0.035	0.62
		3	0.94	0.026	0.42
Undersmooth	100	1	0	0.092	1.20
		2	0	0.071	1.10
		3	0	0.045	0.90
	500	1	0.27	0.056	0.68
		2	0.27	0.025	0.62
		3	0	0.033	0.45
	1000	1	0.35	0.055	0.52
		2	0.32	0.024	0.46
		3	0.11	0.028	0.34
Bias Corr.	100	1	0	0.011	1.07
		2	0	0.019	1.37
		3	0	0.015	1.31
	500	1	0	0.0.068	0.89
		2	0.08	0.05	0.91
		3	0.06	0.036	0.78
	1000	1	0	0.068	0.89
		2	0.08	0.048	0.81
		3	0.09	0.018	0.59

TABLE 4: THE BOOTSTRAP METHOD'S COVERAGE PROBABILITIES FOR g_1 WHEN THE INTERVAL [-0.05,0.05] IS REMOVED FROM S

n	τ	Prop. with Cov. Prob. ≥ 0.95	Av. Abs. Error of Cov. Prob.	Av. Width
100	0.25	0.76	0.020	1.71
	0.50	0.79	0.022	1.62
	0.75	0.85	0.021	1.75
500	0.25	0.91	0.033	0.92
	0.50	1.0	0.039	0.88
	0.75	0.94	0.035	0.94
1000	0.25	0.94	0.034	0.70
	0.50	1.0	0.041	0.67
	0.75	0.94	0.035	0.71

TABLE 5: COVERAGE PROBABILITIES OF UNIFORM CONFIDENCE BANDS OBTAINED BY THE BOOTSTRAP METHOD

τ	n	j	Cov. Prob.
0.25	100	1	0.84
		2	0.94
		3	0.95
	500	1	0.92
		2	0.99
		3	0.97
	1000	1	0.94
		2	1
		3	0.98
0.50	100	1	0.85
		2	0.95
		3	0.93
	500	1	0.95
		2	1
		3	0.97
	1000	1	0.95
		2	1
		3	0.98
0.75	100	1	0.86
		2	0.96
		3	0.96
	500	1	0.94
		2	1
		3	0.97
	1000	1	0.95
		2	1.0
		3	0.97

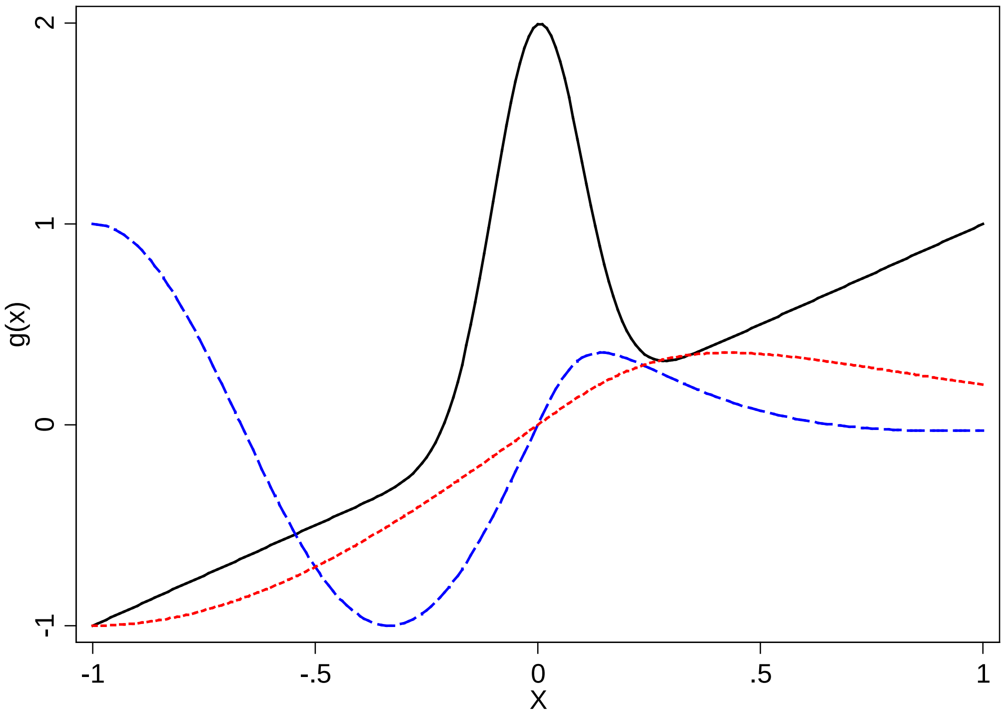

Figure 1: Conditional quantile functions. Solid line is $g_1(x)$. Long dashes are $g_2(x)$. Short dashes are $g_3(x)$.

REFERENCES

Bjerve, S., K.A. Doksum, and B.S. Yandell (1985). Uniform confidence bounds for regression based on a simple moving average. *Scandinavian Journal of Statistics*, 12, 159-169.

Boucheron, S., G. Lugosi, and P. Massart (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford, U.K.: Oxford University Press.

Boyd, S. and L. Vandenberghe (2004). *Convex Optimization*. Cambridge University Press: Cambridge, U.K.

Calonico, S., M.D. Cattaneo, and M.H. Farrell (2016). On the effect of bias estimation on coverage accuracy in nonparametric inference. Working paper, Department of Economics, University of Michigan.

Chen, S.X. (1996). Empirical likelihood confidence intervals for nonparametric density estimation. *Biometrika*, 83, 329-341.

Chen, S.X., W. Härdle, and M. Li (2003). An empirical likelihood goodness-of-fit test for time series. *Journal of the Royal Statistical Society Series B*, 65, 663-678.

Claeskens, G. and I. Van Keilegom (2003). Bootstrap confidence bands for regression curves and their derivatives. *Annals of Statistics*, 31, 1852-1884.

Dudley, R.M. (1967). The sizes of compact subsets of Hilbert space and continuity of Gaussian processes. *Journal of Functional Analysis*, 1, 290-330.

Eubank, R.L. and P.L. Speckman (1993). Confidence regions in nonparametric regression. *Journal of the American Statistical Association*, 88, 1287-1301.

Fan, J., T.-C. Hu, and Y.K. Truong (1994). Robust nonparametric function estimation. *Scandinavian Journal of Statistics*, 21, 433-446.

Guerre, E. and C. Sabbah (2012). Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory*, 28, 87-129.

Hall, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Annals of Statistics*, 20, 675-694.

Hall, P. and J.L. Horowitz (2013). A simple bootstrap method for constructing nonparametric confidence bands for functions. *Annals of Statistics*, 41, 1892-1921.

Hall, P. and A.B. Owen (1993). Empirical likelihood confidence bands in density estimation. *Journal of Computational and Graphical Statistics*, 2, 273-289.

Härdle, W. and A.W. Bowman (1988). Bootstrapping in nonparametric regression: Local adaptive smoothing and confidence bands. *Journal of the American Statistical Association*, 83, 102-110.

Härdle, W., S. Huet, and E. Jolivet (1995). Better bootstrap confidence intervals for regression curve estimation. *Statistics*, 26, 287-306.

Härdle, W., S. Huet, E. Mammen, and S. Sperlich (2004). Bootstrap inference in semiparametric generalized additive models. *Econometric Theory*, 20, 265-300.

Härdle, W. and J.S. Marron (1991). Bootstrap simultaneous error bars for nonparametric regression. *Annals of Statistics*, 19, 778-796.

Komlós, J., P. Major, and G. Tusnády (1975). An approximation of partial sums of independent RV'-s and the sample DF. I. *Zeitschrift fur Warscheinlichkeitstheorie und Verwand Gebiete*, 32, 111-131.

Loève, M. (1978). *Probability Theory II*. New York: Springer-Verlag.

McMurry, T.L. and D.N. Politis. (2008). Bootstrap confidence intervals in nonparametric regression with built-in bias correction. *Statistics and Probability Letters*, 78, 2463-2469.

Neumann, M.H. (1995). Automatic bandwidth choice and confidence intervals in nonparametric regression. *Annals of Statistics*, 1937-1959.

Neumann, M.H. and J. Polzehl (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *Journal of Nonparametric Statistics*, 9, 307-333.

Picard, D. and K. Tribouley (2000). Adaptive confidence interval for pointwise curve estimation. *Annals of Statistics*, 28, 298-335.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer.

Schucany, W.R. and J.P. Sommers (1977). Improvement of kernel type density estimators. *Journal of the American Statistical Association*, 72, 420-423.

Sun, J. and C.R. Loader (1994). Simultaneous confidence bands for linear regression and smoothing. *Annals of Statistics*, 22, 1328-1345.

Xia, Y. (1998). Bias-corrected confidence bands in nonparametric regression. *Journal of the Royal Statistical Society Series B*, 60, 797-811.

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