cemmap
centre for microdata methods and practice

Uniform confidence bands for functions estimated nonparametrically with instrumental variables

Joel L. Horowitz
Sokbae Lee

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP19/10

UNIFORM CONFIDENCE BANDS FOR FUNCTIONS ESTIMATED NONPARAMETRICALLY WITH INSTRUMENTAL VARIABLES

JOEL L. HOROWITZ AND SOKBAE LEE

Abstract. This paper is concerned with developing uniform confidence bands for functions estimated nonparametrically with instrumental variables. We show that a sieve nonparametric instrumental variables estimator is pointwise asymptotically normally distributed. The asymptotic normality result holds in both mildly and severely ill-posed cases. We present an interpolation method to obtain a uniform confidence band and show that the bootstrap can be used to obtain the required critical values. Monte Carlo experiments illustrate the finite-sample performance of the uniform confidence band.

JEL Classification Codes: C13, C14.

Key words: Bootstrap, instrumental variables, sieve estimator, uniform confidence band.

1. Introduction

This paper is concerned with developing a uniform confidence band for the unknown function g in the model

$$Y = g(X) + U;\ E(U|W = w) = 0 \text{ for almost every } w, \tag{1.1}$$

where Y is a scalar dependent variable, $X \in \mathbb{R}^q$ is a continuously distributed explanatory variable that may be endogenous (that is, we allow the possibility that $E(U|X = x) \neq 0$), $W \in \mathbb{R}^q$ is a continuously distributed instrument for X, and U is an unobserved random variable. The unknown function g is nonparametric. It is assumed to satisfy mild regularity conditions but does not belong to a known,

---

Date: 21 June 2010.

The work of both authors was supported in part by the Economic and Social Research Council through its funding of the ESRC Centre for Microdata Methods and Practice (RES-589-28-0001). The research of Joel L. Horowitz was supported in part by NSF grant SES-0817552 and the research of Sokbae Lee was supported in part by ESRC research grant RES-000-22-2761 and by European Research Council (ERC) research grant ERC-2009-StG-240910-ROMETA. We would like to thank Xiaohong Chen and three anonymous referees for helpful comments.

finite-dimensional parametric family. The data are an independent random sample $\{(Y_i, X_i, W_i) : i = 1, \ldots, n\}$ from the distribution of (Y, X, W).

Nonparametric estimators of g in (1.1) have been developed by Newey and Powell (2003); Hall and Horowitz (2005); Darolles, Florens, and Renault (2006); and Blundell, Chen, and Kristensen (2007). Horowitz (2007) gave conditions for asymptotic normality of the kernel estimator of Hall and Horowitz (2005). Newey, Powell, and Vella (1999) presented a control function approach to estimating g in a model that is different from (1.1) but allows endogeneity of X and achieves identification through an instrument. The control function model is non-nested with (1.1) and is not discussed further in this paper. Chernozhukov, Imbens, and Newey (2007); Horowitz and Lee (2007); and Chernozhukov, Gagliardini, and Scaillet (2008) have developed methods for estimating a quantile-regression version of model (1.1). In the quantile regression, the condition $E(U|W = w) = 0$ is replaced by

$$P(U \leq 0|W = w) = \alpha \quad \text{for some } \alpha \in (0, 1). \tag{1.2}$$

Chen and Pouzo (2008, 2009) developed a method for estimating a large class of nonparametric and semiparametric conditional moment models with possibly non-smooth moments. This class includes (1.2).

This paper obtains asymptotic uniform confidence bands for g in (1.1) by using a modified version of the sieve estimator of Blundell, Chen, and Kristensen (2007). Sieve estimators of g are easier to compute than kernel-based estimators such as those of Darolles, Florens, and Renault (2006) and Hall and Horowitz (2005). Moreover, sieve estimators achieve the fastest possible rate of convergence under conditions that are weaker in important ways than those required by existing kernel-based estimators. The sieve estimator used in this paper was proposed by Horowitz (2009) in connection with a specification test for model (1.1). Here, we show that this estimator is pointwise asymptotically normal and that the bootstrap can be used to obtain simultaneous pointwise confidence intervals for $g(x_1), \ldots, g(x_L)$ on almost every finite grid of points $x_1, \ldots, x_L$. We obtain a uniform confidence band by using properties of g such as smoothness or monotonicity to interpolate between the grid points. Hall and Titterington (1988) used interpolation to obtain uniform confidence bands for nonparametrically estimated probability density and conditional mean functions.

A seemingly natural approach to constructing a uniform confidence band is to obtain the asymptotic distribution of a suitably scaled version of $\sup_x |\hat{g}(x) - g(x)|$, where $\hat{g}$ is the estimator of g. However, when $\hat{g}$ is a sieve estimator, this is a difficult problem that has been solved only for special cases in which g is a conditional mean function and certain restrictive conditions hold (Zhou, Shen, and Wolfe 1998; Wang and Yang 2009). Our interpolation approach avoids this problem. The resulting uniform confidence band is not asymptotically exact; its true and nominal coverage probabilities are not necessarily equal even asymptotically. But the confidence band can be made arbitrarily accurate (that is, the difference between the true and nominal asymptotic coverage probabilities can be made arbitrarily small) by making the grid $x_1, \ldots, x_L$ sufficiently fine. In practice, a confidence band can be computed at only finitely many points, so it makes little practical difference whether the confidence interval at each point is based on a finite-dimensional distribution or the distribution of a scaled version of $\sup_x |\hat{g}(x) - g(x)|$.

The remainder of the paper is organized as follows. Section 2 presents the sieve nonparametric IV estimator. Section 3 gives conditions under which the estimators of $g(x_1), \ldots, g(x_L)$ are asymptotically multivariate normally distributed when X and W are scalar random variables. Section 4 uses the results of Section 3 to obtain a uniform confidence band for g when X and W are scalars. Section 5 establishes consistency of the bootstrap for estimating the confidence band. Section 6 extends the results of Sections 3-5 to the case in which X and W are random vectors. Section 7 reports the results of a Monte Carlo investigation of the finite-sample coverage probabilities of the uniform confidence bands, and concluding comments are given in Section 8. The proofs of theorems are in the appendix.

2. The Sieve Nonparametric Estimator

This section describes Horowitz's (2009) sieve estimator of g when X and W are scalar random variables. Let f_W denote the probability density function of W, f_{XW} denote the probability density function of (X, W), and

$$m(w) := E(Y|W = w)f_W(w).$$

Assume, without loss of generality, that the supports of X and W are $[0, 1]$. This assumption can always be satisfied, if necessary, carrying out monotone transformations of X and W. Define the operator A by

$$(Av)(w) := \int_0^1 v(x)f_{XW}(x, w)dx.$$

Then g in (1.1) satisfies

$$Ag = m.$$

For a function $v : [0,1] \mapsto \mathbb{R}$ and integer $l \geq 0$, define

$$D_l v(x) := \frac{\partial^l v(x)}{\partial x^l}$$

whenever the derivative exists, with the convention $D_0 v(x) = v(x)$. Given an integer $s > 0$, define the Sobolev norm

$$\|v\|_s := \left\{ \sum_{l=0}^{s} \int_0^1 [D_l v(x)]^2 \, dx \right\}^{1/2}$$

and the function space

$$\mathcal{H}_s := \{ v : [0,1] \mapsto \mathbb{R} : \|v\|_s \leq C_g \},$$

where $C_g < \infty$ is a constant. Assume that $g \in \mathcal{H}_s$ for some $s > 0$ and that $\|g\|_s < C_g$.

The estimator of g is defined in terms of series expansions of g, m, and A. Let $\{\psi_j : j = 1, 2, \ldots\}$ be a complete, orthonormal basis for $L_2[0,1]$ (the space of square integrable functions on $[0,1]$). The expansions are

$$\begin{aligned} g(x) &= \sum_{j=1}^{\infty} b_j \psi_j(x), \\ m(w) &= \sum_{k=1}^{\infty} a_k \psi_k(w), \\ f_{XW}(x,w) &= \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} c_{jk} \psi_j(x) \psi_k(w), \end{aligned} \tag{2.1}$$

where

$$\begin{aligned} b_j &= \int_0^1 g(x) \psi_j(x) dx, \\ a_k &= \int_0^1 m(w) \psi_k(w) dw, \\ c_{jk} &= \int_{[0,1]^2} f_{XW}(x,w) \psi_j(x) \psi_k(w) dw dx. \end{aligned}$$

To estimate g, we need to estimate a_k, m, c_{jk}, and f_{XW}. The estimators are

$$\begin{aligned} \hat{a}_k &= n^{-1}\sum_{i=1}^{n} Y_i\psi_k(W_i), \\ \hat{m} &= \sum_{j=1}^{J_n} \hat{a}_j\psi_j, \qquad (2.2) \\ \hat{c}_{jk} &= n^{-1}\sum_{1=1}^{n} \psi_j(X_i)\psi_k(W_i), \end{aligned}$$

and

$$\hat{f}_{XW}(x,w) = \sum_{j=1}^{J_n}\sum_{k=1}^{J_n} \hat{c}_{jk}\psi_j(x)\psi_k(w),$$

respectively, where $J_n < \infty$ is the series truncation point. Define the operator $\hat{A}_n$ that estimates A by

$$(\hat{A}_n v)(w) := \int_0^1 v(x)\hat{f}_{XW}(x,w)dx. \qquad (2.3)$$

Define the subset of $\mathcal{H}_s$:

$$\mathcal{H}_{ns} := \left\{ v = \sum_{j=1}^{J_n} v_j\psi_j : \|v\|_s \le C_g \right\}.$$

The sieve estimator of g is defined as

$$\hat{g}_n := \arg\min_{v\in\mathcal{H}_{ns}} \left\| \hat{A}_n v - \hat{m} \right\|, \qquad (2.4)$$

where $\|\cdot\|$ is the norm on $L_2[0,1]$. Under the assumptions of Section 3, $\hat{A}_n^{-1}$ exists with probability approaching 1 as $n \to \infty$ and $P(\hat{A}_n\hat{g}_n = \hat{m}) \to 1$ as $n \to \infty$. Therefore,

$$\hat{g}_n = \hat{A}_n^{-1}\hat{m} \qquad (2.5)$$

with probability approaching 1 as $n \to \infty$.

When n is small, $\hat{g}_n$ in (2.5) may be numerically unstable. Blundell, Chen, and Kristensen (2007) propose stabilizing $\hat{g}_n$ by replacing (2.4) with the solution to a penalized least-squares problem. Blundell, Chen, and Kristensen (2007) provide an analytic, easily computed solution to this problem and present the results of numerical experiments on the penalization method's ability to stabilize $\hat{g}_n$ in small samples. We do not pursue this approach here, because it does not affect our theoretical results and we have not encountered numerical instability in our simulations.

3. Asymptotic Normality

This section gives conditions under which $\hat{g}_n(x)$ is asymptotically normally distributed. Proving asymptotic normality of an estimator usually requires assumptions that are stronger than those needed for consistency or convergence at the asymptotically optimal rate. The assumptions made here are stronger than those used by Blundell, Chen, and Kristensen (2007) and Horowitz (2009) to prove that their estimators are consistent with the optimal rate of convergence.

Define A^* to be the adjoint operator of A and

$$\rho_n := \sup_{h \in \mathcal{H}_{ns}: \|h\| \neq 0} \frac{\|h\|}{\|(A^* A)^{1/2} h\|}, \tag{3.1}$$

Blundell, Chen, and Kristensen (2007) call ρ_n the sieve measure of ill-posedness and discuss its relation to the eigenvalues of A^*A. Under suitable conditions, $\rho_n = O(J_n^r)$ if the eigenvalues, sorted in decreasing order, converge to zero at the rate J_n^{-2r} (mildly ill-posed case). If the eigenvalues converge exponentially fast (severely ill-posed case), then ρ_n is proportional to $\exp(cJ_n)$ for some finite $c > 0$.

Assumption 3.1. *(1) The supports of X and W are $[0,1]$. (2) $g \in \mathcal{H}_s$ and $\|g\|_s < C_g$ for some integer $s > 0$ and finite constant C_g. (3) The operator A is nonsingular. (4) (X, W) has a probability density function f_{XW} with respect to Lebesgue measure. In addition, f_{XW} has $r \geq s$ bounded derivatives with respect to any combination of its arguments. (5) The conditional density of X given W and the marginal density of W, denoted by $f_{X|W}(x|w)$ and $f_W(w)$, respectively, are bounded. (6) $\sup_{w \in [0,1]} E(Y^2|W = w) \leq C_Y$ for some $C_Y < \infty$.*

Assumption 3.2. *(1) The set of functions $\{\psi_j : j = 1, 2, \ldots\}$ is a complete, orthonormal basis for $L_2[0,1]$. (2) $\left\| g - \sum_{j=1}^J b_j \psi_j \right\| = O\left(J^{-s}\right)$. (3) $\|A_n - A\| = O(J_n^{-r})$ if $r < \infty$ and $\|A_n - A\| = O[\exp(-cJ_n)]$ for some $c > 0$ if $r = \infty$.*

Among other things, Assumptions 3.1 and 3.2 ensure that f_{XW} is at least as smooth as g. Moreover, A and A^* map $L_2[0,1]$ into $\mathcal{H}_s$. When Assumptions 3.1 (2) and 3.2 (4) hold, Assumptions 3.2 (2) and 3.2 (3) are satisfied by a variety of bases including trigonometric functions, orthogonal polynomials, and splines.

Let A_n be the operator on $L_2[0,1]$ whose kernel is

$$a_n(x,w) = \sum_{j=1}^{J_n}\sum_{k=1}^{J_n} c_{jk}\psi_j(x)\psi_k(w).$$

Let A_n^* denote the adjoint operator of A_n.

Assumption 3.3. *The ranges of A_n and A_n^* are contained in $\mathcal{H}_{ns}$ for all sufficiently large n. Moreover,*

$$\rho_n \sup_{h\in\mathcal{H}_{ns}:\|h\|\neq 0} \frac{\|(A_n - A)h\|}{\|h\|} = O(J_n^{-s}). \tag{3.2}$$

Assumption 3.3 ensures that A_n is a "sufficiently accurate" approximation to A on $\mathcal{H}_{ns}$. This assumption complements Assumption 3.2 (3), which specifies the accuracy of A_n as an approximation to A on the larger set $\mathcal{H}_s$. Condition (3.2) can be interpreted as a smoothness restriction on f_{XW} or as a restriction on the sizes of the values of c_{jk} for $j \neq k$. Condition (3.2) is satisfied automatically if $c_{jk} = c_{jj}\delta_{jk}$, where δ_{jk} is the Kronecker delta. Hall and Horowitz (2005) used a similar diagonality condition in their nonparametric instrumental variables estimator.

Assumption 3.4. *(1) $J_n^{-s} = o\left[\rho_n(J_n/n)^{1/2}\right]$. (2) $(\rho_n J_n)/n^{1/2} \to 0$.*

Assumption 3.4 (1) requires $\hat{g}_n$ to be undersmoothed. That is, as $n \to \infty$, J_n increases at a rate that is faster than the asymptotically optimal rate. As with other nonparametric estimators, undersmoothing ensures that the asymptotic bias of $\hat{g}_n$ is negligible. Assumption 3.4 (2) ensures that the asymptotic variance of $\hat{g}_n$ converges to zero.

Remark 1. (1) If $\rho_n = O(J_n^r)$ for some finite $r > 0$, then we can set $J_n \propto n^\eta$, where $\frac{1}{2r+2s+1} < \eta < \frac{1}{2r+2}$.

(2) If $\rho_n = \exp(cJ_n)$ for some finite $c > 0$, Assumption 3.4 is satisfied if

$$J_n = \frac{\log n}{2c} - \frac{2s\alpha_0 + 1}{2c}\log\log n$$

for some α_0 satisfying $0 < \alpha_0 < 1$. The rate of increase must be logarithmic, and the constant multiplying $\log n$ must be $1/(2c)$. If the constant is larger, the integrated variance of $\hat{g}_n - g$ does not converge to 0. If the constant is smaller, the bias dominates the variance. The higher order component of J_n is important. If it is 0 or too small, the integrated variance does not converge to 0. These requirements illustrate the delicacy of estimation in the severely ill-posed case.

Remark 2. Finding a theory-based method for choosing J_n is a difficult and important problem whose solution is beyond the scope of this paper. Pending a solution, we suggest using the following heuristic, which has worked well in Monte Carlo experimentation. The integrated variance of $\hat{g}_n$ is $E\left\|\hat{g}_n - E\hat{g}_n\right\|^2 = \sum_{j=1}^{J_n} \sigma_j^2$, where $\sigma_j^2 = \mathrm{Var}\langle \hat{g}_n, \psi_j\rangle$ and $\langle\cdot,\cdot\rangle$ denotes the inner product in $L_2[0,1]$. The variance components σ_j^2 can be estimated by using the standard formulae of GMM estimation. Chen and Pouzo (2008) and Horowitz (2009) have found through Monte Carlo experiments that as J_n increases from 1, $E\left\|\hat{g}_n - E\hat{g}_n\right\|^2$ changes little at first but increases by a factor of 10 or more when J_n crosses a "critical value." This suggests the following heuristic procedure for choosing J_n in applications: First, find the largest value of J_n that does not produce a very large increase in the estimated value of $E\left\|\hat{g}_n - E\hat{g}_n\right\|^2$. Call this value J_{n0}. Then achieve undersmoothng by choosing $J_n = J_{n0}^{\gamma}$ for some $\gamma > 1$.

Now define

$$\gamma_n(x, Y, X, W) := [Y - g(X)] \sum_{j=1}^{J_n} [(A_n^{-1}\psi_j)(x)]\psi_j(W). \tag{3.3}$$

Also, define

$$\sigma_n^2(x) := n^{-1}\mathrm{Var}\left[\gamma_n(x, Y, X, W)\right]. \tag{3.4}$$

Define $c_n \asymp d_n$ for any positive sequences of constants c_n and d_n to mean that c_n/d_n is bounded away from 0 and ∞.

Assumption 3.5. *For any $x \in [0,1]$, $\sigma_n(x) \asymp \|\sigma_n\|$ except, possibly, if x belongs to a set of Lebesgue measure 0.*

This condition is similar to Assumption 6 of Horowitz (2007). It rules out a form of superefficiency in which $\hat{g}_n(x) - g(x)$ converges to 0 more rapidly than $\|\hat{g}_n - g\|$.

Assumption 3.6. *(1) $\sup_{w\in[0,1]} E(|U|^{2+\delta}|W = w) \le C_U$ for some $C_U < \infty$ and for some $\delta > 0$. (2) $E|\psi_j(W)|^{2+\delta}$ is bounded uniformly over j. (3) $n^{-\delta/2} J_n^{1+\delta/2} \to 0$.*

Assumption 3.6 ensures that we can establish the asymptotic normality of the sieve estimator. Conditions (1) and (2) impose some moment restrictions on U and $\psi_j(W)$. Condition (3) holds for the mildly ill-posed case if $J_n \propto n^{\eta}$ as in Remark 1 (1), $\eta < \delta/(2+\delta)$ and

$$\frac{1}{2r + 2s + 1} < \frac{\delta}{2+\delta} < \frac{1}{2r+2}.$$

It holds in the severely ill-posed case for any $\delta > 0$ since in this case, J_n diverges at a logarithmic rate.

Let $\{x_1, \ldots, x_L\}$ denote a set of L points in $[0,1]$. The following theorem establishes the joint asymptotic normality of the sieve estimator of $\hat{g}_n(x_1), \ldots, \hat{g}_n(x_L)$.

Theorem 3.1. *Let Assumptions 3.1-3.6 hold. Then as $n \to \infty$,*

$$V_g(x_1, \ldots, x_L)^{-1/2} \left\{ \frac{[\hat{g}_n(x_1) - g(x_1)]}{\sigma_n(x_1)}, \ldots, \frac{[\hat{g}_n(x_L) - g(x_L)]}{\sigma_n(x_L)} \right\} \to_d \mathbf{N}(0, I),$$

except, possibly, if $x_1, ..., x_L$ belong to a set of Lebesgue measure 0 in $[0,1]^L$, where I is the L-dimensional identity matrix and $V_g(x_1, \ldots, x_L)$ is the $L \times L$ matrix whose (j,k) element is

$$V_{jk} := E\left[\frac{\gamma_n(x_j, Y, X, W)\gamma_n(x_k, Y, X, W)}{(Var[\gamma_n(x_j, Y, X, W)])^{1/2}(Var[\gamma_n(x_k, Y, X, W)])^{1/2}}\right].$$

3.1. **Estimation of $\sigma_n^2(x)$.** To make use of the asymptotic results obtained in Theorem 3.1, it is necessary to estimate $\sigma_n^2(x)$. To do this, let

$$\hat{\delta}_n(x, Y, X, W) := [Y - \hat{g}_n(X)] \sum_{k=1}^{J_n} \psi_k(W)(\hat{A}_n^{-1}\psi_k)(x). \tag{3.5}$$

Then $\sigma_n^2(x)$ can be estimated consistently by

$$s_n^2(x) := n^{-2} \sum_{i=1}^{n} \left\{ \hat{\delta}_n(x, Y_i, X_i, W_i) - \bar{\delta}_n(x) \right\}^2, \tag{3.6}$$

where

$$\bar{\delta}_n(x) := n^{-1} \sum_{i=1}^{n} \hat{\delta}_n(x, Y_i, X_i, W_i). \tag{3.7}$$

We now state the consistency of $s_n^2(x)$.

Theorem 3.2. *Let Assumptions 3.1-3.6 hold. Then as $n \to \infty$,*

$$\frac{s_n^2(x)}{\sigma_n^2(x)} \to_p 1.$$

4. Uniform Confidence Band

The results in Section 3 make it possible to form joint confidence intervals and, by interpolation, a uniform confidence band for g over $[a, b]$ for constants a and b such that $0 \le a < b \le 1$. To form joint confidence intervals, let $\{x_1, \ldots, x_L\}$ be points sampled from uniform distributions on the intervals $[a, a+(b-a)/L), [a+(b-a)/L, a+2(b-$

$a)/L), \ldots, [a+(L-1)(b-a)/L, b]$. Random sampling this way avoids exceptional sets of Lebesgue measure 0 in Theorem 3.1. Let z_α satisfy

$$P\left[\sup_{1\leq l\leq L} |\mathbb{Z}_l| > z_\alpha\right] = \alpha,$$

where $\mathbb{Z}_l$ is the l-th component of $\mathbb{Z} \sim \mathbf{N}[0, V_g(x_1, \ldots, x_L)]$. Then

$$\hat{g}(x_l) - z_\alpha s_n(x_l) \leq g(x_l) \leq \hat{g}(x_l) + z_\alpha s_n(x_l) \tag{4.1}$$

are joint asymptotic $100(1-\alpha)\%$ confidence intervals for $g(x_1), \ldots, g(x_L)$, $l = 1, \ldots, L$. We now describe two ways of obtaining a uniform confidence band for g by interpolating the joint confidence intervals. A method for estimating z_α is described in Section 5.

4.1. **A Uniform Confidence Band under Monotonicity.**

In this subsection, we develop a uniform confidence band when g is monotonic on $[a, b]$. The monotonicity assumption is common in economics. For example, market demand is a monotone decreasing function of price. Many functions of interest in economics (e.g., production functions, cost functions, among many others) are monotonic.

Let

$$\overline{x}_l := \operatorname{argmax}\{\hat{g}_n(x_l) + z_\alpha s_n(x_l), \hat{g}_n(x_{l+1}) + z_\alpha s_n(x_{l+1})\},$$

and

$$\underline{x}_l := \operatorname{argmin}\{\hat{g}_n(x_l) - z_\alpha s_n(x_l), \hat{g}_n(x_{l+1}) - z_\alpha s_n(x_{l+1})\}.$$

Then by the assumed monotonicity of g,

$$\hat{g}_n(\underline{x}_l) - z_\alpha s_n(\underline{x}_l) \leq g(x) \leq \hat{g}_n(\overline{x}_l) + z_\alpha s_n(\overline{x}_l)$$

uniformly over $x \in [x_l, x_{l+1}]$, $l = 1, \ldots, L-1$. Putting these intervals together gives a uniform confidence band for g over $[a, b]$. The asymptotic coverage probability is at least $1-\alpha$ and it can be made arbitrarily close to $1-\alpha$ by making L sufficiently large.

Remark 3. Our result does not require $\hat{g}_n$ to be monotonic. It is possible that an estimator of g that has been made monotonic through rearrangement would yield a confidence band that is narrower than ours is, at least in finite samples (Chernozhukov, Fernández-Val, and Galichon, 2009). Investigation of this issue is left to future research.

4.2. **A Uniform Confidence Band under Lipschitz Continuity.** In this subsection, we assume that g is Lipschitz continuous. That is,

$$|g(x) - g(y)| \leq C_L|x - y|$$

for some constant C_L and any $x, y \in [a, b]$.[1] For any $x \in [a + (b - a)/L, a + (L - 1)(b - a)/L]$, choose l such that $|x - x_l|$ is minimized. First note that (4.1) is equivalent to

$$\begin{aligned} &\hat{g}_n(x_l) - z_\alpha s_n(x_l) + [g(x) - g(x_l)] \\ &\leq g(x) \leq \hat{g}_n(x_l) + z_\alpha s_n(x_l) + [g(x) - g(x_l)]. \end{aligned} \tag{4.2}$$

Then (4.2) implies

$$\hat{g}_n(x_l) - z_\alpha s_n(x_l) - C_L|x - x_l| \leq g(x) \leq \hat{g}_n(x_l) + z_\alpha s_n(x_l) + C_L|x - x_l|,$$

so that

$$\hat{g}_n(x_l) - z_\alpha s_n(x_l) - \frac{C_L}{L} \leq g(x) \leq \hat{g}_n(x_l) + z_\alpha s_n(x_l) + \frac{C_L}{L} \tag{4.3}$$

uniformly over $x \in [x_l - 1/L, x_l + 1/L]$. Putting these intervals in (4.3) together gives a uniform confidence band for g over $[a, b]$. Again the asymptotic coverage probability exceeds $1 - \alpha$ but can be made arbitrarily close to $1 - \alpha$ by making L sufficiently large.

In applications, C_L is unknown. Replacing C_L in (4.3) with an estimator such as $\hat{C}_L = \max_{x\in[a,b]} |\hat{g}_n'(x)|$ is undesirable because $\hat{C}_L - C_L$ converges to 0 more slowly than $\hat{g}_n - g$ does. If one is willing to place an *a priori* upper bound on C_L, this upper bound can be used in place of C_L in (4.3). Putting an upper bound on C_L can be viewed as limiting the amount of wiggliness that is allowed for g. This is often reasonable in economic applications since many functions of interest such as Engel curves and earnings functions are unlikely to be very wiggly.

5. Bootstrap Estimation of z_α

This section shows that the bootstrap consistently estimates the joint asymptotic distribution of $[\hat{g}_n(x_1) - g(x_1)]/s_n(x_1)$,...,$[\hat{g}_n(x_L) - g(x_L)]/s_L(x_L)$. It follows that the bootstrap consistently estimates the critical value z_α in (4.1).

It is shown in the proof of theorem 3.1 that the leading term of the asymptotic expansion of $\hat{g}_n(x) - g(x)$ is

$$S_n(x) = n^{-1} \sum_{i=1}^{n} \delta_n(x, Y_i, X_i, W_i),$$

[1] Lipschitz continuity requires strengthening Assumption 3.1 (2) slightly to make the first derivative uniformly bounded, not just bounded in the Sobolev norm.

where

(5.1)

$$\delta_n(x,Y,X,W) := \sum_{k=1}^{J_n} \left\{ [Y\psi_k(W) - a_k] - \sum_{j=1}^{J_n} b_j \left[\psi_j(X)\psi_k(W) - c_{jk}\right] \right\} (A_n^{-1}\psi_k)(x).$$

Therefore, it suffices to show that the bootstrap consistently estimates the asymptotic distribution of $t_n(x_1), ..., t_n(x_L)$, where $t_n(x) := S_n(x)/s_n(x)$. Define $g_n(x) := \sum_{j=1}^{J_n} b_j\psi_j(x)$ for any $x \in [0,1]$. Define

$$\tilde{S}_n(x) := n^{-1} \sum_{i=1}^{n} \tilde{\delta}_n(x, Y_i, X_i, W_i), \tag{5.2}$$

where

$$\tilde{\delta}_n(x,Y,X,W) := [Y - g_n(X)] \sum_{k=1}^{J_n} \psi_k(W)(A_n^{-1}\psi_k)(x). \tag{5.3}$$

Then $S_n(x)$ can be rewritten as

$$S_n(x) = \tilde{S}_n(x) - E\tilde{S}_n(x).$$

Hence, $t_n(x) = [\tilde{S}_n(x) - E\tilde{S}_n(x)]/s_n(x)$. We now describe a bootstrap procedure that consistently estimates the asymptotic distribution of $t_n(x_1), \ldots, t_n(x_L)$.

Let $\{(Y_i^*, X_i^*, W_i^*) : i = 1, \ldots, n\}$ denote a bootstrap sample that is obtained by sampling the data $\{(Y_i, X_i, W_i) : i = 1, \ldots, n\}$ randomly with replacement. The bootstrap version of $\tilde{S}_n(x)$ is

$$\tilde{S}_n^*(x) := n^{-1} \sum_{i=1}^{n} \hat{\delta}_n(x, Y_i^*, X_i^*, W_i^*),$$

where $\hat{\delta}_n(x,Y,X,W)$ is defined in (3.5). A bootstrap version of $t_n(x)$ is

$$t_n^*(x) := \left[\tilde{S}_n^*(x) - \bar{\delta}_n(x)\right]/s_n(x), \tag{5.4}$$

where $\bar{\delta}_n(x)$ is defined in (3.7). The α-level bootstrap critical value, z_α^*, estimates z_α in (4.1) and can be obtained as the solution to

$$P^*\left[\sup_{1\le l\le L} |t_n^*(x_l)| > z_\alpha^*\right] = \alpha,$$

where P^* denotes the probability measure induced by bootstrap sampling conditional on the data $\{(Y_i, X_i, W_i) : i = 1, \ldots, n\}$. One nice feature of the bootstrap procedure is that it is unnecessary to estimate $V_g(x_1, \ldots, x_L)$.

An alternative bootstrap version of $t_n(x)$ is

$$t_n^{**}(x) := \left[\tilde{S}_n^*(x) - \bar{\delta}_n(x)\right] / s_n^*(x), \tag{5.5}$$

where $s_n^*(x)$ is the bootstrap analog of $s_n(x)$. Specifically,

$$s_n^*(x) := \left[n^{-2} \sum_{i=1}^{n} \left\{ \hat{\delta}_n^*(x, Y_i^*, X_i^*, W_i^*) - \bar{\delta}_n^*(x) \right\}^2 \right]^{1/2}, \tag{5.6}$$

where $\hat{A}_n^*$ and $\hat{g}_n^*$, respectively, are the same as $\hat{A}_n$ and $\hat{g}_n$ in (2.3) and (2.4), but with the bootstrap sample $\{(Y_i^*, X_i^*, W_i^*) : i = 1, \ldots, n\}$ in place of the estimation data,

$$\hat{\delta}_n^*(x, Y_i^*, X_i^*, W_i^*) := [Y_i^* - \hat{g}_n^*(X_i^*)] \sum_{k=1}^{J_n} \psi_k(W_i^*)[(\hat{A}_n^*)^{-1}\psi_k](x). \tag{5.7}$$

and

$$\bar{\delta}_n^*(x) := n^{-1} \sum_{i=1}^{n} \hat{\delta}_n^*(x, Y_i^*, X_i^*, W_i^*). \tag{5.8}$$

Let $\mathcal{L}^*(\ldots)$ denote the conditional distribution $L(\ldots|\{(Y_i, X_i, W_i) : i = 1, \ldots, n\})$ and let $d_\infty(H_1, H_2)$ denote the Kolmogorov distance, that is the sup norm of the difference between two distribution functions H_1 and H_2. The following theorem establishes the consistency of the bootstrap and implies that z_α^* is a consistent estimator of z_α.

Theorem 5.1. *Let Assumptions 3.1-3.6 hold. Then as $n \to \infty$,*

(5.9)
$$d_\infty\left(\mathcal{L}^*\{t_n^*(x_1), \ldots, t_n^*(x_L)\}, \mathbf{N}[0, V_g(x_1, \ldots, x_L)]\right) \to 0 \quad \textit{in probability,}$$

and

(5.10)
$$d_\infty\left(\mathcal{L}^*\{t_n^{**}(x_1), \ldots, t_n^{**}(x_L)\}, \mathbf{N}[0, V_g(x_1, \ldots, x_L)]\right) \to 0 \quad \textit{in probability.}$$

6. Multivariate Model

This section extends the results of Sections 2-5 to a multivariate model in which X and W are q-dimensional random vectors. Assume that the support of (X, W) is contained in $[0,1]^{2q}$. Let $\{\psi_j : j = 1, 2, \ldots\}$ be a complete, orthonormal basis for $L_2[0,1]^q$. Define the operator A by

$$(Av)(w) := \int_{[0,1]^q} v(x) f_{XW}(x, w) dx.$$

As in Section 2, the estimator of g is defined in terms of series expansions of g, m, and A. The expansions are like those in (2.1) with the following generalized Fourier coefficients:

$$b_j = \int_{[0,1]^q} g(x)\psi_j(x)dx,$$
$$a_k = \int_{[0,1]^q} m(w)\psi_j(w)dw,$$
$$c_{jk} = \int_{[0,1]^{2q}} f_{XW}(x,w)\psi_j(x)\psi_k(w)dwdx.$$

The estimators of a_k, m, c_{jk}, and f_{XW} are the same as in (2.2), but with the basis functions for $L_2[0,1]^q$. Also, define the operator $\hat{A}_n$ that estimates A by

$$(\hat{A}_n v)(w) := \int_{[0,1]^q} v(x)\hat{f}_{XW}(x,w)dx. \tag{6.1}$$

The sieve estimator of g is as in (2.4), where $\|\cdot\|$ is now the norm on $L_2[0,1]^q$. Then the asymptotic normality result of Section 3 can be extended to the multivariate model with minor modifications.[2]

As in Section 4, it is possible to form joint confidence set for g in the multivariate model. However, it is difficult to display joint confidence intervals or a uniform confidence set when X is multidimensional. Therefore, we consider a one-dimensional projection of a joint confidence set for g.

Assume without loss of generality that the first component of X is the direction of interest. Let $\{x_{11}, \ldots, x_{1L}\}$ be points sampled from uniform distributions on the intervals $[a, a+(b-a)/L)$, $[a+(b-a)/L, a+2(b-a)/L)$, $\ldots$, $[a+(L-1)(b-a)/L, b]$. Let $\sigma_n^2(x)$ denote a multivariate version of (3.4) and $s_n^2(x)$ denote a consistent estimator of $\sigma_n^2(x)$ as in (3.6). For a fixed value, say x_{-1}, of remaining components of X,

(6.2)
$$\hat{g}(x_{1l}, x_{-1}) - z_\alpha s_n(x_{1l}, x_{-1}) \leq g(x_{1l}, x_{-1}) \leq \hat{g}(x_{1l}, x_{-1}) + z_\alpha s_n(x_{1l}, x_{-1})$$

[2] If there are more instruments than covariates, we need to replace the basis $\{\psi_j(w)\}$ with a higher-dimensional basis. This complicates the notation but changes nothing of importance in the theory. In applications, a high-dimensional instrument increases the number of Fourier coefficients of f_{XW} that must be estimated, thereby creating a curse-of-dimensionality effect and reducing estimation precision.

are joint asymptotic $100(1-\alpha)\%$ confidence intervals for $\{g(x_{1l}) : l = 1, \ldots, L\}$ over $[a, b]$, where

$$P\left[\sup_{1\le l\le L} |\mathbb{Z}_l| > z_\alpha\right] = \alpha,$$

and $\mathbb{Z}_l$ is the l-th component of $\mathbb{Z}$. Here, $\mathbb{Z}$ is the L-dimensional mean-zero normal vector whose covariance matrix is the asymptotic covariance matrix of

$$\left\{\frac{[\hat{g}_n(x_{11}, x_{-1}) - g(x_{11}, x_{-1})]}{\sigma_n(x_{11}, x_{-1})}, \ldots, \frac{[\hat{g}_n(x_{1L}, x_{-1}) - g(x_{1L}, x_{-1})]}{\sigma_n(x_{1L}, x_{-1})}\right\}.$$

We can construct the uniform confidence band from (6.2) as in Section 4 by assuming monotonicity or Lipschitz continuity. As in Section 5, the critical value z_α can be obtained by the bootstrap.

7. Monte Carlo Experiments

This section reports the results of a Monte Carlo investigation of the coverage probabilities of the joint confidence intervals and uniform confidence bands using the bootstrap-based critical values of Section 5.

As in Horowitz (2007), realizations of (Y, X, W) were generated from the model

$$f_{XW}(x, w) = C_f \sum_{j=1}^{\infty} (-1)^{j+1} j^{-\alpha/2} \sin(j\pi x)\sin(j\pi w),$$
$$g(x) = 2.2x,$$
$$Y = E[g(X)|W] + V,$$

where C_f is a normalization constant chosen so that the integral of the joint density of (X, W) equals one and $V \sim \mathbf{N}(0, 0.01)$. Experiments were carried out with $\alpha = 1.2$ and $\alpha = 10$. The sample size is $n = 200$. There are 1000 Monte Carlo replications in each experiment.

The grid $(x_1, ..., x_L)$ used to form joint confidence intervals and uniform confidence bands consists of 100 points. The Monte Carlo results are not sensitive to variations in the value of L over the range 25 to 100. The basis functions are Legendre polynomials that have had their supports shifted and have been normalized to make them orthonormal on $[0, 1]$. The critical values are obtained by using the two bootstrap methods of Section 5 with 1000 bootstrap replications. The confidence bands were computed by using the piecewise monotonicity method of Section 4.1. The joint confidence intervals are for $(x_1, \ldots, x_L) \in [a, b]$ and the uniform confidence band is for any $x \in [a, b] = [0.2, 0.8]$, $[0.1, 0.9]$ or $[0.01, 0.99]$.

The results of the experiments are shown in Tables 1-2. In each table, columns 3-5 show the empirical coverage probabilities of the joint confidence intervals, and columns 6-8 show the empirical coverage probabilities of the uniform confidence bands. We show the results of experiments with $J_n = 3, 4, 5$, and 6. The results show that the differences between the nominal and empirical coverage probabilities are small when the critical value is based on $t_n^{**}(x)$ and $J_n = 3$ or 4.

8. Conclusions

This paper has given conditions under which a sieve nonparametric IV estimator is pointwise asymptotically normally distributed. The asymptotic normality result holds in both mildly and severely ill-posed cases. We have also shown that joint pointwise confidence intervals can be interpolated to obtain a uniform confidence band for the estimated function. The bootstrap can be used to estimate the critical values needed to form confidence intervals and bands. The results of Monte Carlo experiments show that the differences between nominal and empirical coverage probabilities are small when the critical values are obtained by using a suitable version of the bootstrap.

Appendix A. Proofs

Throughout the proofs, $\|\cdot\|$ is the L_2 norm if $(\cdot)$ is a function and the L_2 operator norm if $(\cdot)$ is an operator, e.g. $\|A\| = \sup_{\|h\|=1} \|Ah\|$.

We begin with the proof of Theorem 3.1. Because $\hat{g}_n = \hat{A}_n^{-1}\hat{m}$ with probability approaching 1, it suffices to establish the asymptotic distribution of $\hat{h} \equiv \hat{A}_n^{-1}\hat{m}$.

Define

$$m_n := \sum_{k=1}^{J_n} a_k \psi_k.$$

Then

$$A_n\hat{h} + (\hat{A}_n - A_n)\hat{h} = \hat{m},$$

so that

$$\begin{aligned} \hat{h} &= A_n^{-1}\hat{m} - A_n^{-1}(\hat{A}_n - A_n)\hat{h} \\ &= A_n^{-1}\hat{m} - A_n^{-1}(\hat{A}_n - A_n)g - A_n^{-1}(\hat{A}_n - A_n)(\hat{h} - g). \end{aligned} \tag{A.1}$$

Recall that $g_n = \sum_{j=1}^{J_n} b_j\psi_j$. Write

$$A_n^{-1}\hat{m} - g = A_n^{-1}(\hat{m} - m_n) + (A_n^{-1}m_n - g_n) + (g_n - g). \tag{A.2}$$

Combining (A.1) with (A.2) yields $\hat{h} - g = S_n + R_n$, where

$$S_n := A_n^{-1}(\hat{m} - m_n) - A_n^{-1}(\hat{A}_n - A_n)g \tag{A.3}$$

and $R_n := R_{n1} + R_{n2} + R_{n3}$ with

$$\begin{aligned} R_{n1} &= -A_n^{-1}(\hat{A}_n - A_n)(\hat{h} - g), \\ R_{n2} &= A_n^{-1}m_n - g_n, \\ R_{n3} &= g_n - g. \end{aligned}$$

Now using the series expansions, we have that

$$\begin{aligned} [A_n^{-1}(\hat{m} - m_n)](x) &= \sum_{k=1}^{J_n} (\hat{a}_k - a_k)(A_n^{-1}\psi_k)(x) \\ &= n^{-1} \sum_{i=1}^{n} \sum_{k=1}^{J_n} \left[Y_i \psi_k(W_i) - a_k \right] (A_n^{-1}\psi_k)(x) \end{aligned}$$

and

$$\begin{aligned} [A_n^{-1}(\hat{A}_n - A_n)g](x) &= \sum_{j=1}^{J_n} \sum_{k=1}^{J_n} b_j (\hat{c}_{jk} - c_{jk})(A_n^{-1}\psi_k)(x) \\ &= n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{J_n} \sum_{k=1}^{J_n} b_j \left[\psi_j(X_i)\psi_k(W_i) - c_{jk} \right] (A_n^{-1}\psi_k)(x). \end{aligned}$$

Therefore, it follows from (A.3) that

$$S_n(x) = n^{-1} \sum_{i=1}^{n} \delta_n(x, Y_i, X_i, W_i),$$

where $\delta_n(x, Y, X, W)$ is defined in (5.1). Recall that $\tilde{S}_n(x)$ and $\tilde{\delta}_n(x, Y, X, W)$ are defined in (5.2) and (5.3) in Section 5. Then

$$S_n(x) = \tilde{S}_n(x) - E\tilde{S}_n(x).$$

Observe that

$$(A_n^{-1}\psi_k)(x) = \sum_{j=1}^{J_n} c^{jk} \psi_j(x),$$

where c^{jk} is the (j,k) element of the inverse of the $J_n \times J_n$ matrix $[c_{jk}]$. Therefore,

$$\sum_{k=1}^{J_n} \psi_k(W)(A_n^{-1}\psi_k)(x) = \sum_{j=1}^{J_n}\sum_{k=1}^{J_n} c^{jk}\psi_k(W)\psi_j(x)$$
$$= \sum_{j=1}^{J_n} \psi_j(x)[(A_n^{-1})^*\psi_j](W),$$

where * denotes the adjoint operator. Recall that $U = Y - g(X)$. Then

$$\begin{aligned}\tilde{\delta}_n(x,Y,X,W) &= [Y - g_n(X)] \sum_{j=1}^{J_n} \psi_j(x)[(A_n^{-1})^*\psi_j](W) \\ &= U \sum_{j=1}^{J_n} \psi_j(x)[(A_n^{-1})^*\psi_j](W) \\ &\quad - [g_n(X) - g(X)] \sum_{j=1}^{J_n} \psi_j(x)[(A_n^{-1})^*\psi_j](W) \\ &\equiv \gamma_n(x,Y,X,W) + \tilde{\gamma}_n(x,Y,X,W).\end{aligned}$$

Therefore, since $E\gamma_n(x,Y,X,W) = 0$,

$$\begin{aligned}S_n(x) &= \tilde{S}_n(x) - E\tilde{S}_n(x) \\ &= n^{-1}\sum_{i=1}^{n} \gamma_n(x,Y_i,X_i,W_i) + n^{-1}\sum_{i=1}^{n} \{\tilde{\gamma}_n(x,Y_i,X_i,W_i) - E\tilde{\gamma}_n(x,Y,X,W)\} \\ &\equiv \Gamma_n(x) + \tilde{\Gamma}_n(x).\end{aligned}$$

We now prove five lemmas that are useful to prove Theorem 3.1.

Lemma A.1. *We have that*

$$\left\|A_n^{-1}\right\| = \rho_n[1 + O(J_n^{-s})].$$

Proof of Lemma A.1. First note that by Assumption 3.3, the eigenfunctions of $A_n^* A_n$ are in $\mathcal{H}_s$ for all sufficiently large n. Hence, since the dimension of $A_n^* A_n$ is J_n, we have that the eigenfunctions of $A_n^* A_n$ are in $\mathcal{H}_{ns}$ as well.

Now $\|A_n^{-1}\|^2$ is the largest eigenvalue of $(A_n^{-1})^* A_n^{-1} = (A_n A_n^*)^{-1}$, which is the inverse of the smallest eigenvalue of $A_n A_n^*$ or, equivalently, the inverse of the smallest eigenvalue of $A_n^* A_n$. Since the smallest eigenvalue of $A_n^* A_n$ minimizes $\|A_n\|^2$, it suffices to the find the inverse of

$$\inf_{h \in \mathcal{H}_{ns}: \|h\| \neq 0} \frac{\|A_n h\|}{\|h\|}.$$

But

$$\begin{aligned}
\rho_n^{-1} &= \inf_{h \in \mathcal{H}_{ns}: \|h\| \neq 0} \frac{\|Ah\|}{\|h\|} = \inf_{h \in \mathcal{H}_{ns}: \|h\| \neq 0} \frac{\|A_n h + (A - A_n)h\|}{\|h\|} \\
&\leq \inf_{h \in \mathcal{H}_{ns}: \|h\| \neq 0} \frac{\|A_n h\| + \|(A - A_n)h\|}{\|h\|} \\
&= \inf_{h \in \mathcal{H}_{ns}: \|h\| \neq 0} \frac{\|A_n h\|}{\|h\|} + O(\rho_n^{-1} J_n^{-s})
\end{aligned}$$

by (3.2). Therefore,

$$\inf_{h \in \mathcal{H}_{ns}: \|h\| \neq 0} \frac{\|A_n h\|}{\|h\|} \geq \rho_n^{-1} + O(\rho_n^{-1} J_n^{-s}) = \rho_n^{-1}[1 + O(J_n^{-s})],$$

which implies that

$$\left\|A_n^{-1}\right\| \leq \rho_n[1 + O(J_n^{-s})].$$

Also, note that since $\|A_n h + (A - A_n)h\| \geq \|A_n h\| - \|(A_n - A)h\|$,

$$\begin{aligned}
\rho_n^{-1} &= \inf_{h \in \mathcal{H}_{ns}: \|h\| \neq 0} \frac{\|Ah\|}{\|h\|} = \inf_{h \in \mathcal{H}_{ns}: \|h\| \neq 0} \frac{\|A_n h + (A - A_n)h\|}{\|h\|} \\
&\geq \inf_{h \in \mathcal{H}_{ns}: \|h\| \neq 0} \frac{\|A_n h\| - \|(A_n - A)h\|}{\|h\|} \\
&= \inf_{h \in \mathcal{H}_{ns}: \|h\| \neq 0} \frac{\|A_n h\|}{\|h\|} + O(\rho_n^{-1} J_n^{-s})
\end{aligned}$$

by (3.2). Therefore,

$$\inf_{h \in \mathcal{H}_{ns}: \|h\| \neq 0} \frac{\|A_n h\|}{\|h\|} \leq \rho_n^{-1} + O(\rho_n^{-1} J_n^{-s}) = \rho_n^{-1}[1 + O(J_n^{-s})],$$

which implies that

$$\left\|A_n^{-1}\right\| \geq \rho_n[1 + O(J_n^{-s})].$$

Therefore, we have proved the lemma. □

Lemma A.2. *We have that*

$$\|R_{n1}\| = O\left[\rho_n^2 (J_n/n)\right].$$

Proof of Lemma A.2. By Horowitz (2009),

$$\begin{aligned}
\left\|\hat{h} - g\right\| &= O_p\left[J_n^{-s} + \rho_n (J_n/n)^{1/2}\right] \\
&= O_p\left[\rho_n (J_n/n)^{1/2}\right],
\end{aligned}$$

where the last equality follows from undersmoothing (See Assumption 3.4 (1)). Horowitz (2009) proves that $\left\|\hat{A}_n - A_n\right\| = O_p[(J_n/n)^{1/2}]$. Therefore, by Lemma A.1,

$$\begin{aligned}\|R_{n1}\| &\leq \left\|A_n^{-1}\right\| \left\|(\hat{A}_n - A_n)(\hat{h} - g)\right\| \\ &\leq O(\rho_n) \left\|\hat{A}_n - A_n\right\| \left\|\hat{h} - g\right\| \\ &= O(\rho_n)\; O_p[(J_n/n)^{1/2}] \left\|\hat{h} - g\right\|,\end{aligned}$$

which proves the lemma. □

Lemma A.3. *We have that*

$$\|R_{n2}\| = O(J_n^{-s}).$$

Proof of Lemma A.3. Note that by Lemma A.1,

$$\|R_{n2}\| \leq \left\|A_n^{-1}\right\| \|m_n - A_n g_n\| \leq O(\rho_n) \|m_n - A_n g_n\|.$$

Also, note that

$$m = Ag = \sum_{j=1}^{\infty}\sum_{k=1}^{\infty} b_j c_{jk}\psi_k,$$

and

$$m_n = \sum_{j=1}^{\infty}\sum_{k=1}^{J_n} b_j c_{jk}\psi_k.$$

Moreover, $A_n g_n = \sum_{j=1}^{J_n}\sum_{k=1}^{J_n} b_j c_{jk}\psi_k$. Therefore,

$$m_n - A_n g_n = \sum_{j=J_n+1}^{\infty}\sum_{k=1}^{J_n} b_j c_{jk}\psi_k.$$

In addition,

$$(A - A_n)g = \sum_{j=J_n+1}^{\infty}\sum_{k=1}^{J_n} b_j c_{jk}\psi_k + \sum_{j=1}^{\infty}\sum_{k=J_n+1}^{\infty} b_j c_{jk}\psi_k.$$

Therefore,

$$\|(A - A_n)g\|^2 = \|m_n - A_n g_n\|^2 + \sum_{k=J_n+1}^{\infty}\left(\sum_{j=1}^{\infty} b_j c_{jk}\right)^2,$$

which implies that

$$\rho_n \|m_n - A_n g_n\| \leq \rho_n \|(A - A_n)g\|.$$

Now note that Assumption 3.3 implies that

$$\rho_n \sup_{h \in \mathcal{H}_{ns}} \|(A_n - A)h\| = O(J_n^{-s}). \tag{A.4}$$

Furthermore,

$$\rho_n \|(A - A_n)(g - g_n)\| \le \rho_n \|A - A_n\| \|g - g_n\| = O(J_n^{-s}) \tag{A.5}$$

by Assumptions 3.2 (2) and (3). Therefore, using (A.4) and (A.5), we have that

$$\begin{aligned} \rho_n \|(A - A_n)g\| &\le \rho_n \|(A - A_n)g_n\| + \rho_n \|(A - A_n)(g - g_n)\| \\ &= O(J_n^{-s}). \end{aligned}$$

Therefore, we have proved the lemma. □

Lemma A.4. *We have that*

$$\|\sigma_n\| = O\left[\rho_n (J_n/n)^{1/2}\right].$$

Proof. To show the lemma, write

$$\begin{aligned} \int_0^1 \sigma_n^2(x) dx &= n^{-1} E \int_0^1 [\gamma_n(x, Y, X, W)]^2 dx \\ &= E\left[\sigma_U^2(W) \sum_{j=1}^{J_n} \left\{[(A_n^{-1})^* \psi_j](W)\right\}^2\right], \end{aligned}$$

where $\sigma_U^2(w) := E[U^2|W = w]$ is bounded uniformly over w because of Assumption 3.6 (1).

Now $\|(A_n^{-1})^*\| = \|A_n^{-1}\|$. Therefore, by Lemma A.1 and the fact that f_W is bounded, $\|(A_n^{-1})^* \psi_j\|^2 = O(\rho_n^2)$ and $E\{[(A_n^{-1})^* \psi_j](W)\}^2 = O(\rho_n^2)$ uniformly over j. It follows that

$$n^{-1} E \|\gamma_n\|^2 = O\left(\rho_n^2 J_n / n\right). \tag{A.6}$$

Hence, we have proved the lemma. □

Lemma A.5. *We have that*

$$E\left\|\tilde{\Gamma}_n\right\|^2 = O(J_n^{-2s+1} \rho_n^2 / n).$$

Proof. Note that

$$\begin{aligned} E\left\|\tilde{\Gamma}_n\right\|^2 &= n^{-1} E \|\tilde{\gamma}_n - E\tilde{\gamma}_n\|^2 \\ &\le n^{-1} E \|\tilde{\gamma}_n\|^2 \\ &= n^{-1} E\left[E\{[g_n(X) - g(X)]^2 | W\} \sum_{j=1}^{J_n} \{[(A_n^{-1})^* \psi_j](W)\}^2\right]. \end{aligned}$$

Then by Assumption 3.1 (5) and 3.2 (2),

$$\begin{aligned} E\left[[g_n(X) - g(X)]^2 | W = w\right] &\leq \sup_{(x,w)} f_{X|W}(x|w) \left\| g_n - g \right\|^2 \\ &= O(J_n^{-2s}). \end{aligned}$$

Combining this with the fact that $E\{[(A_n^{-1})^* \psi_j](W)\}^2 = O(\rho_n^2)$ uniformly over j, we have that

$$\begin{aligned} n^{-1} E \left\| \tilde{\gamma}_n \right\|^2 &= O(J_n^{-2s}/n) E \left[\sum_{j=1}^{J_n} \{[(A_n^{-1})^* \psi_j](W)\}^2 \right] \\ &= O(J_n^{-2s+1} \rho_n^2 / n). \end{aligned} \tag{A.7}$$

Therefore, we have proved the lemma. □

Proof of Theorem 3.1. Note that by Assumption 3.2 (2), $\|R_{n3}\| = O(J_n^{-s})$. This is asymptotically negligible because of undersmoothing (Assumption 3.4 (1)). Therefore, by Lemmas A.2 and A.3 with the conditions on J_n in Assumption 3.4,

$$\|R_n\| = o_p\left[\rho_n (J_n/n)^{1/2}\right]. \tag{A.8}$$

Define

$$T_n(x) := n^{-1/2} \sum_{i=1}^{n} \frac{\gamma_n(x, Y_i, X_i, W_i)}{(\mathrm{Var}[\gamma_n(x, Y, X, W)])^{1/2}}.$$

Since

$$\begin{aligned} \gamma_n(x, Y, X, W) &= U \sum_{j=1}^{J_n} \psi_j(x) [(A_n^{-1})^* \psi_j](W) \\ &= U \sum_{j=1}^{J_n} [(A_n^{-1} \psi_j)(x)] \psi_j(W), \end{aligned}$$

the Lyapunov condition here is that

$$B_n \equiv n^{-\delta/2} (\mathrm{Var}[\gamma_n(x, Y, X, W)])^{-(2+\delta)/2} E \left| U \sum_{j=1}^{J_n} [(A_n^{-1} \psi_j)(x)] \psi_j(W) \right|^{2+\delta} \to 0$$

as $n \to \infty$.

Since $\|A_n^{-1}\| = \rho_n[1+O(J_n^{-s})]$ by Lemma A.1, we have that $(A_n^{-1}\psi_j)(x) = O(\rho_n)$ for almost every x. Hence, for almost every x,

$$\begin{aligned} E\left|U\sum_{j=1}^{J_n}[(A_n^{-1}\psi_j)(x)]\psi_j(W)\right|^{2+\delta} &\leq C_U E\left|\sum_{j=1}^{J_n}[(A_n^{-1}\psi_j)(x)]\psi_j(W)\right|^{2+\delta} \\ &\leq O(\rho_n^{2+\delta})E\left[\sum_{j=1}^{J_n}|\psi_j(W)|\right]^{2+\delta} \\ &\leq O(\rho_n^{2+\delta})\left[\sum_{j=1}^{J_n}\left\{E|\psi_j(W)|^{2+\delta}\right\}^{1/(2+\delta)}\right]^{2+\delta} \\ &= O\left[(\rho_n J_n)^{2+\delta}\right], \end{aligned}$$

where the first inequality follows from Assumption 3.6 (1), the second inequality is due to the observation that $(A_n^{-1}\psi_j)(x) = O(\rho_n)$ for almost every x, the third inequality is by the generalized Minkowski's inequality, and the fourth equality is from Lemma A.1 and Assumption 3.6 (2). Also, we have that $\sigma_n(x) \asymp \rho_n(J_n/n)^{1/2}$ and $\mathrm{Var}[\gamma_n(x,Y,X,W)] \asymp \rho_n^2 J_n$ by Lemma A.4 and Assumption 3.5. Then it follows that

$$B_n = O\left(n^{-\delta/2}J_n^{1+\delta/2}\right) = o(1),$$

where the last equality comes from Assumption 3.6 (3). Therefore, we have shown that the Lyapunov condition is satisfied.

Then a triangular-array version of the Lindeberg-Levy central theorem yields the result that

$$T_n(x) \to_d \mathbf{N}(0,1).$$

Now let $\{x_1,\ldots,x_L\}$ be a set of L points in $[0,1]$. Then, the Cramér-Wold device yields the result that

$$\left\{\frac{S_n(x_1)}{\sigma_n(x_l)},\ldots,\frac{S_n(x_L)}{\sigma_n(x_L)}\right\} \to_d \mathbf{N}[0,V_g(x_1,\ldots,x_L)].$$

Under the assumption that $\sigma_n(x) \asymp \|\sigma_n\|$, the theorem now follows from Lemmas A.4 and A.5. □

We will first prove Theorem 5.1 and then Theorem 3.2.

Proof of Theorem 5.1. Define

$$\Lambda_n(x,X,W) := -\left[\hat{g}_n(X) - g_n(X)\right]\sum_{k=1}^{J_n}\psi_k(W)(A_n^{-1}\psi_k)(x)$$

and

$$\bar{\Lambda}_n(x) := n^{-1} \sum_{i=1}^{n} \Lambda_n(x, X_i, W_i).$$

Also, define

$$\delta_n^*(x, Y, X, W) := [Y - \hat{g}_n(X)] \sum_{k=1}^{J_n} \psi_k(W)(A_n^{-1}\psi_k)(x). \tag{A.9}$$

Now write

$$\delta_n^*(x, Y, X, W) = \tilde{\delta}_n(x, Y, X, W) + \Lambda_n(x, X, W).$$

Define $\Delta_n := \hat{A}_n - A_n$. Then using the fact that

$$\hat{A}_n^{-1} - A_n^{-1} = \left[(I + A_n^{-1}\Delta_n)^{-1} - I\right] A_n^{-1}, \tag{A.10}$$

we have that

$$\tilde{S}_n^*(x) - \bar{\delta}_n(x) = \sum_{l=1}^{4} \tilde{S}_{nl}^*(x),$$

where

$$\tilde{S}_{n1}^*(x) = n^{-1} \sum_{i=1}^{n} \left[\tilde{\delta}_n(x, Y_i^*, X_i^*, W_i^*) - \bar{\delta}_n(x)\right],$$

$$\tilde{S}_{n2}^*(x) = n^{-1} \left[(I + A_n^{-1}\Delta_n)^{-1} - I\right] \sum_{i=1}^{n} \left[\tilde{\delta}_n(x, Y_i^*, X_i^*, W_i^*) - \bar{\delta}_n(x)\right],$$

$$\tilde{S}_{n3}^*(x) = n^{-1} \sum_{i=1}^{n} \left[\Lambda_n(x, X_i^*, W_i^*) - \bar{\Lambda}_n(x)\right],$$

and

$$\tilde{S}_{n4}^*(x) = n^{-1} \left[(I + A_n^{-1}\Delta_n)^{-1} - I\right] \sum_{i=1}^{n} \left[\Lambda_n(x, X_i^*, W_i^*) - \bar{\Lambda}_n(x)\right].$$

First, $\tilde{S}_{n1}^*(x)$ is a bootstrap analog of S_n, so consistency of the bootstrap distribution of $\tilde{S}_{n1}^*(x)/s_n(x)$ for that of $S_n/s_n(x)$ follows immediately from Theorem 1.1 of Mammen (1992). Similarly, the bootstrap distribution of $\sum_{l=1}^{L} \gamma_l \tilde{S}_{n1}^*(x_l)/s_n(x_l)$ is consistent for that of $\sum_{l=1}^{L} \gamma_l S_n(x_l)/s_n(x_l)$ for any real constants $\gamma_1, \ldots, \gamma_L$.

Now consider $\tilde{S}_{n2}^*$. Note that

$$\left\|A_n^{-1}\Delta_n\right\| \leq \left\|A_n^{-1}\right\| \left\|\Delta_n\right\| = O_p\left[\rho_n (J_n/n)^{1/2}\right] = o_p(1). \tag{A.11}$$

Therefore,

$$\left\|(I + A_n^{-1}\Delta_n)^{-1} - I\right\| = o_p(1). \tag{A.12}$$

Since

$$\tilde{S}^*_{n2}(x) = \left[(I + A_n^{-1}\Delta_n)^{-1} - I\right] \tilde{S}^*_{n1}(x),$$

(A.12) implies that

$$\left\|\tilde{S}^*_{n2}\right\| = o_p(1) \left\|\tilde{S}^*_{n1}\right\|.$$

Now consider $\tilde{S}^*_{n3}$ and $\tilde{S}^*_{n4}$. We have that

$$\tilde{S}^*_{n4}(x) = \left[(I + A_n^{-1}\Delta_n)^{-1} - I\right] \tilde{S}^*_{n3}(x).$$

Therefore, again (A.12) implies that

$$\left\|\tilde{S}^*_{n4}\right\| = o_p(1) \left\|\tilde{S}^*_{n3}\right\|.$$

It now suffices to show that $\tilde{S}^*_{n3}$ is asymptotically negligible. To do this, define

$$\nu_n(X) := \hat{g}_n(X) - g_n(X),$$

$$Z_n(W, x) := \sum_{k=1}^{J_n} \psi_k(W)(A_n^{-1}\psi_k)(x).$$

Then

$$\tilde{S}^*_{n3}(x) = n^{-1}\sum_{i=1}^{n} \nu_n(X_i^*)Z_n(W_i^*, x) - n^{-1}\sum_{i=1}^{n} \nu_n(X_i)Z_n(W_i, x).$$

Let V^* and E^*, respectively, denote the variance and expectation relative to the distribution induced by bootstrap sampling. Then $E^*\tilde{S}^*_{n3}(x) = 0$. Define $V_n^*(x) := V^*[\tilde{S}^*_{n3}(x)]$. Now note that

$$\begin{aligned} V_n^*(x) &\leq E^* n^{-2}\sum_{i=1}^{n} \nu_n^*(X_i^*)^2 Z_n^*(W_i^*, x)^2 \\ &= n^{-2}\sum_{i=1}^{n} \nu_n^*(X_i)^2 Z_n^*(W_i, x)^2. \end{aligned}$$

But, $\nu_n(X_i)^2 = O(\|\hat{g}_n - g_n\|^2) = O(\|\hat{g}_n - g\|^2)$ with probability 1. Therefore,

$$V_n^*(x) \leq n^{-2} O(\|\hat{g}_n - g\|^2) \sum_{i=1}^{n} Z_n^*(W_i, x)^2$$

with probability 1. Now,

$$n^{-2}\sum_{i=1}^{n} Z_n^*(W_i,x)^2 = n^{-2}\sum_{i=1}^{n}\left[\sum_{k=1}^{J_n}\psi_k(W_i)(A_n^{-1}\psi_k)(x)\right]^2$$
$$\equiv R_n(x).$$

But, $\|A_n^{-1}\psi_k\| = O(\rho_n)$, so $(A_n^{-1}\psi_k)(x) = O(\rho_n)$ for almost every x. Therefore,

$$R_n(x) \leq O(\rho_n^2)n^{-2}\sum_{i=1}^{n}\left[\sum_{k=1}^{J_n}|\psi_k(X_i)|\right]^2$$
$$= O_p\left(\frac{\rho_n^2 J_n^2}{n}\right)$$

by Markov's inequality for almost every x. Under Assumption 3.4 (2), $R_n(x) = o_p(1)$ for almost every x. It follows that for almost every x,

$$V_n^*(x) = o_p(\|\hat{g}_n - g\|^2).$$

This combined with the fact that $E^*\tilde{S}_{n3}^*(x) = 0$ implies that $\tilde{S}_{n3}^*(x)$ is asymptotically negligible for almost every x under sampling from the bootstrap distribution.

Now note that the estimator $s_n(x)$ is consistent for $\sigma_n(x)$ by Theorem 3.2. Therefore, the first conclusion (5.9) of Theorem 5.1 follows from consistency of the bootstrap distribution of the bootstrap distribution of $\sum_{l=1}^{L}\gamma_l\tilde{S}_{n1}^*(x_l)/s_n(x_l)$ for that of $\sum_{l=1}^{L}\gamma_l\tilde{S}_n(x_l)/s_n(x_l)$ and the Cramér-Wold device.

Similarly, the second conclusion (5.10) of Theorem 5.1 follows if we show that $s_n^*(x)$ is consistent for $\sigma_n(x)$, which is proved in Lemma A.6 below. □

Proof of Theorem 3.2. Note that we can write $s_n^2(x)$ as

$$s_n^2(x) = n^{-2}\sum_{i=1}^{n}\left\{\hat{\delta}_n(x,Y_i,X_i,W_i)\right\}^2 - n^{-1}\left[\bar{\delta}_n(x)\right]^2.$$

By the arguments used for $\tilde{S}_{n2}^*$ in the proof of Theorem 5.1, replacing $\hat{A}_n$ with A_n creates an asymptotically negligible error for almost every x, it suffices to prove the consistency of

$$n^{-2}\sum_{i=1}^{n}\{\delta_n^*(x,Y_i,X_i,W_i)\}^2 - n^{-1}\left[\bar{\delta}_n^*(x)\right]^2,$$

where δ_n^* is defined in (A.9) and

$$\bar{\delta}_n^*(x) := n^{-1} \sum_{i=1}^n \delta_n^*(x, Y_i, X_i, W_i).$$

Now

$$\delta_n^*(x, Y, X, W) = \tilde{\delta}_n(x, Y, X, W) + \Lambda_n(x, X, W). \tag{A.13}$$

Then the second term on the right-hand side of (A.13) is asymptotically negligible for almost every x by the arguments used with $\tilde{S}_{n3}^*$ in the proof of Theorem 5.1. Furthermore, recall that

$$\tilde{\delta}_n(x, Y, X, W) = \gamma_n(x, Y, X, W) + \tilde{\gamma}_n(x, Y, X, W). \tag{A.14}$$

Again, the second term on the right-hand side of (A.14) is asymptotically negligible for almost every x using the arguments used in the proof of Theorem 3.1. Therefore, it suffices to show that

(A.15)

$$\sigma_n^{-2}(x) \left\{ n^{-2} \sum_{i=1}^n [\gamma_n(x, Y_i, X_i, W_i)]^2 - n^{-1} \left[n^{-1} \sum_{i=1}^n \gamma_n(x, Y_i, X_i, W_i) \right]^2 \right\} \to_p 1.$$

We have shown that the Lyapunov condition is satisfied in the proof of Theorem 3.1. Therefore, (A.15) follows from a triangular-array version of the weak law of large numbers, e.g. Lemmas 11.4.2 and 11.4.3 of Lehmann and Romano (2005). □

Lemma A.6. *Let Assumptions 3.1-3.6 hold. Then as* $n \to \infty$,

$$\frac{[s_n^*(x)]^2}{\sigma_n^2(x)} \to_p 1,$$

conditional on the original observations $\{(Y_i, X_i, W_i) : i = 1, \ldots, n\}$.

Proof of Lemma A.6. The estimator $[s_n^*(x)]^2$ differs from $s_n^2(x)$ by replacing $\hat{g}_n$ with $\hat{g}_n^*$, $\hat{A}_n^{-1}$ with $(\hat{A}_n^*)^{-1}$, and $\{Y_i, X_i, W_i\}$ with $\{Y_i^*, X_i^*, W_i^*\}$.

Define $\Delta_n^* := \hat{A}_n^* - \hat{A}_n$. Then

$$(\hat{A}_n^*)^{-1} - \hat{A}_n^{-1} = \left[(I + \hat{A}_n^{-1}\Delta_n^*)^{-1} - I \right] \hat{A}_n^{-1}. \tag{A.16}$$

Now using (A.10), write

$$\begin{aligned} \hat{A}_n^{-1}\Delta_n^* &= A_n^{-1}\Delta_n^* + [\hat{A}_n^{-1} - A_n^{-1}]\Delta_n^* \\ &= A_n^{-1}\Delta_n^* + [(I + A_n^{-1}\Delta_n)^{-1} - I]A_n^{-1}\Delta_n^*. \end{aligned}$$

Thus, by (A.12),

$$\left\| \hat{A}_n^{-1}\Delta_n^* \right\| \le [1 + o_p(1)] \left\| A_n^{-1}\Delta_n^* \right\|. \tag{A.17}$$

Now as in (A.11),

$$\left\|A_n^{-1}\Delta_n^*\right\| \leq \left\|A_n^{-1}\right\| \left\|\Delta_n^*\right\| = O_{p^*}\left[\rho_n (J_n/n)^{1/2}\right] = o_{p^*}(1), \tag{A.18}$$

where p^* denotes bootstrap probability. It follows from (A.16)-(A.18) that

$$\left\|[(\hat{A}_n^*)^{-1} - \hat{A}_n^{-1}]h\right\| = o_{p^*}(1) \left\|\hat{A}_n^{-1}h\right\| \tag{A.19}$$

for any $h \in L_2[0,1]$.

Now define $\hat{m}^* = \sum_{k=1}^{J_n} a_k^* \psi_k$, where $a_k^* = n^{-1}\sum_{i=1}^n Y_i^* \psi_k(W_i^*)$. Set

$$\hat{g}_n^* = (\hat{A}_n^*)^{-1}\hat{m}^*.$$

Note that this is not the same as (2.4) with the bootstrap sample. Recall that $\hat{h} \equiv \hat{A}_n^{-1}\hat{m}$ is asymptotically equivalent to $\hat{g}_n$. Then

$$\begin{aligned}&\hat{g}_n^* - \hat{h}\\ &= [(A_n^*)^{-1} - \hat{A}_n^{-1}]\hat{m} + [(A_n^*)^{-1} - \hat{A}_n^{-1}](\hat{m}^* - \hat{m}) + \hat{A}_n^{-1}(\hat{m}^* - \hat{m}).\end{aligned}$$

Therefore, it follows from (A.19) and the fact that $\|\hat{m}^* - \hat{m}\| = O_{p^*}[(J_n/n)^{1/2}]$ that

$$\left\|\hat{g}_n^* - \hat{h}\right\| = O_{p^*}[\rho_n (J_n/n)^{1/2}]. \tag{A.20}$$

Consequently, it follows from that (A.19) and (A.20) that $s_n^*(x)^2$ is asymptotically equivalent to

$$n^{-2}\sum_{i=1}^n \left\{\hat{\delta}_n(x, Y_i^*, X_i^*, W_i^*) - \bar{\delta}_n(x)\right\}^2,$$

where $\hat{\delta}_n(x,Y,X,W)$ and $\bar{\delta}_n(x)$ are defined in (3.5) and (3.7), respectively. Then the lemma follows from the consistency of the bootstrap estimator of a sample average. □

TABLE 1. Results of Monte Carlo experiments with bootstrap critical values ($\alpha = 1.2$)

Range of x: $[a,b]$	J_n	Joint Confidence Intervals Nominal Probabilities 0.90	0.95	0.99	Uniform Confidence Band Nominal Probabilities 0.90	0.95	0.99
Bootstrap Critical Values I							
(0.2,0.8)	3	0.866	0.923	0.962	0.872	0.926	0.962
	4	0.913	0.953	0.986	0.920	0.957	0.986
	5	0.929	0.962	0.987	0.935	0.965	0.989
	6	0.933	0.966	0.989	0.938	0.970	0.990
(0.1,0.9)	3	0.851	0.893	0.944	0.859	0.904	0.948
	4	0.826	0.883	0.926	0.838	0.886	0.931
	5	0.874	0.914	0.963	0.883	0.921	0.964
	6	0.896	0.940	0.975	0.903	0.947	0.979
(0.01,0.99)	3	0.848	0.896	0.945	0.862	0.906	0.952
	4	0.808	0.864	0.921	0.830	0.870	0.929
	5	0.790	0.856	0.919	0.817	0.874	0.934
	6	0.788	0.849	0.916	0.825	0.873	0.937
Bootstrap Critical Values II							
(0.2,0.8)	3	0.911	0.951	0.981	0.914	0.951	0.981
	4	0.929	0.968	0.992	0.935	0.971	0.992
	5	0.948	0.981	0.997	0.953	0.984	0.997
	6	0.955	0.987	0.997	0.959	0.989	0.997
(0.1,0.9)	3	0.907	0.946	0.989	0.912	0.949	0.991
	4	0.904	0.938	0.986	0.907	0.940	0.988
	5	0.926	0.966	0.991	0.932	0.967	0.991
	6	0.949	0.980	0.997	0.956	0.982	0.997
(0.01,0.99)	3	0.905	0.946	0.989	0.911	0.955	0.993
	4	0.895	0.949	0.992	0.910	0.957	0.993
	5	0.922	0.964	0.995	0.931	0.973	0.997
	6	0.943	0.976	0.996	0.957	0.984	0.997

Note: This table shows coverage probabilities of the joint confidence intervals and uniform confidence band for $g(x)$. Two types of bootstrap critical values are considered: $t_n^*(x)$ in (5.4) (bootstrap critical value I) and $t_n^{**}(x)$ in (5.5) (bootstrap critical value II).

TABLE 2. Results of Monte Carlo experiments with bootstrap critical values ($\alpha = 10$)

Range of x: $[a,b]$	J_n	Joint Confidence Intervals Nominal Probabilities 0.90	0.95	0.99	Uniform Confidence Band Nominal Probabilities 0.90	0.95	0.99
		Bootstrap Critical Values I					
(0.2,0.8)	3	0.656	0.701	0.768	0.659	0.702	0.770
	4	0.727	0.770	0.846	0.738	0.778	0.848
	5	0.745	0.793	0.871	0.749	0.800	0.877
	6	0.776	0.821	0.890	0.789	0.831	0.897
(0.1,0.9)	3	0.652	0.699	0.765	0.660	0.702	0.768
	4	0.695	0.736	0.812	0.702	0.743	0.820
	5	0.699	0.755	0.829	0.710	0.765	0.843
	6	0.742	0.790	0.867	0.766	0.808	0.875
(0.01,0.99)	3	0.649	0.700	0.765	0.661	0.704	0.768
	4	0.692	0.732	0.811	0.708	0.745	0.819
	5	0.699	0.749	0.831	0.720	0.765	0.846
	6	0.745	0.793	0.865	0.773	0.820	0.882
		Bootstrap Critical Values II					
(0.2,0.8)	3	0.891	0.938	0.975	0.894	0.939	0.976
	4	0.915	0.948	0.983	0.915	0.950	0.983
	5	0.930	0.970	0.991	0.931	0.970	0.991
	6	0.960	0.977	0.995	0.961	0.980	0.995
(0.1,0.9)	3	0.892	0.940	0.979	0.893	0.940	0.979
	4	0.915	0.954	0.986	0.917	0.955	0.986
	5	0.936	0.970	0.991	0.937	0.971	0.991
	6	0.955	0.979	0.996	0.956	0.979	0.996
(0.01,0.99)	3	0.892	0.942	0.979	0.894	0.944	0.979
	4	0.917	0.959	0.986	0.923	0.960	0.986
	5	0.940	0.973	0.993	0.943	0.973	0.993
	6	0.962	0.984	1.000	0.965	0.984	1.000

Note: This table shows coverage probabilities of the joint confidence intervals and uniform confidence band for $g(x)$. Two types of bootstrap critical values are considered: $t_n^*(x)$ in (5.4) (bootstrap critical value I) and $t_n^{**}(x)$ in (5.5) (bootstrap critical value II).

References

Blundell, R., X. Chen, and D. Kristensen (2007): Semi-nonparametric IV estimation of shape-invariant Engel curves, *Econometrica*, 75, 1613–1669.

Chen, X. and D. Pouzo (2008): Estimation of nonparametric conditional moment models with possibly nonsmooth moments, Cemmap Working Papers, CWP12/08, available at: `http://cemmap.ifs.org.uk`.

Chen, X. and D. Pouzo (2009): Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals, *Journal of Econometrics*, 152, 46–60.

Chernozhukov, V., I. Fernández-Val, and A. Galichon (2009): Improving point and interval estimators of monotone functions by rearrangement, *Biometrika* 96, 559–575.

Chernozhukov, V, G. W. Imbens and W. K. Newey (2007): Instrumental variable estimation of nonseparable models, *Journal of Econometrics*, 139, 4–14.

Chernozhukov, V., P. Gagliardini, and O. Scaillet (2008): Nonparametric instrumental variable estimation of quantile structural effects, working paper, Department of Economics, Massachusetts Institute of Technology, Cambridge, MA.

Darolles, S., J. -P. Florens, and E. Renault (2006): Nonparametric instrumental regression, Working paper, GREMAQ, University of Social Science, Toulouse, France.

Hall, P. and J. L. Horowitz (2005): Nonparametric methods for inference in the presence of instrumental variables, *Annals of Statistics*, 33, 2904–2929.

Hall, P. and D. M. Titterington (1988): On confidence bands in nonparametric density estimation and regression, *Journal of Multivariate Analysis*, 27, 228–254.

Horowitz, J. L. (2007): Asymptotic Normality of a Nonparametric Instrumental Variables Estimator, *International Economic Review,* 48, 1329-1349.

Horowitz, J. L. (2009): Specification Testing in Nonparametric Instrumental Variables Estimation, *Journal of Econometrics*, forthcoming.

Horowitz, J. L. and S. Lee (2007): Nonparametric instrumental variables estimation of a quantile regression model, *Econometrica*, 75, 1191–1208.

Lehmann, E. L. and J. P. Romano (2005): *Testing Statistical Hypotheses* (Third Edition), New York: Springer-Verlag.

Mammen, E. (1992). *When Does Bootstrap Work? Asymptotic Results and Simulations*, New York: Springer-Verlag.

Newey, W. K., J. L. Powell, and F. Vella (1999): Nonparametric estimation of triangular simultaneous models, *Econometrica*, 67, 565-603.

Wang, J. and L. Yang (2009). Polynomial spline confidence bands for regression curves, *Statistica Sinica*, 19, 325-342.

Zhou, S., X. Shen, and D.A. Wolfe (1998). Local asymptotics of regression splines and confidence regions, *Annals of Statistics*, 26, 1760-1782.

Department of Economics, Northwestern University, 2001 Sheridan Road, Evanston, Illinois 60208, USA
E-mail address: joel-horowitz@northwestern.edu

Department of Economics, University College London, Gower Street, London WC1E 6BT, UK
E-mail address: l.simon@ucl.ac.uk