Econometrica, Vol. 81, No. 2 (March, 2013), 667–737

INTERSECTION BOUNDS: ESTIMATION AND INFERENCE

BY VICTOR CHERNOZHUKOV, SOKBAE LEE, AND ADAM M. ROSEN[1]

We develop a practical and novel method for inference on intersection bounds, namely bounds defined by either the infimum or supremum of a parametric or non-parametric function, or, equivalently, the value of a linear programming problem with a potentially infinite constraint set. We show that many bounds characterizations in econometrics, for instance bounds on parameters under conditional moment inequalities, can be formulated as intersection bounds. Our approach is especially convenient for models comprised of a continuum of inequalities that are separable in parameters, and also applies to models with inequalities that are nonseparable in parameters. Since analog estimators for intersection bounds can be severely biased in finite samples, routinely underestimating the size of the identified set, we also offer a median-bias-corrected estimator of such bounds as a by-product of our inferential procedures. We develop theory for large sample inference based on the strong approximation of a sequence of series or kernel-based empirical processes by a sequence of "penultimate" Gaussian processes. These penultimate processes are generally not weakly convergent, and thus are non-Donsker. Our theoretical results establish that we can nonetheless perform asymptotically valid inference based on these processes. Our construction also provides new adaptive inequality/moment selection methods. We provide conditions for the use of nonparametric kernel and series estimators, including a novel result that establishes strong approximation for any general series estimator admitting linearization, which may be of independent interest.

KEYWORDS: Bound analysis, conditional moments, partial identification, strong approximation, infinite-dimensional constraints, linear programming, concentration inequalities, anti-concentration inequalities, non-Donsker empirical process methods, moderate deviations, adaptive moment selection.

1. INTRODUCTION

THIS PAPER DEVELOPS a practical and novel method for estimation and inference on intersection bounds. Such bounds arise in settings where the parameter of interest, denoted θ^*, is known to lie within the bounds $[\theta^l(v), \theta^u(v)]$ for

[1]We are especially grateful to D. Chetverikov, K. Kato, Y. Luo, A. Santos, five anonymous referees, and a co-editor for making several extremely useful suggestions that have led to substantial improvements. We thank T. Armstrong, R. Blundell, A. Chesher, F. Molinari, W. Newey, C. Redmond, N. Roys, S. Stouli, and J. Stoye for detailed discussion and suggestions, and participants at numerous seminars and conferences for their comments. This paper is a revised version of "Inference on Intersection Bounds," which initially was presented and circulated at the University of Virginia and the Harvard/MIT econometrics seminars in December 2007, and presented at the March 2008 CEMMAP/Northwestern Conference on Inference in Partially Identified Models With Applications. We gratefully acknowledge financial support from the National Science Foundation, the Economic and Social Research Council (RES-589-28-0001, RES-000-22-2761), and the European Research Council (ERC-2009-StG-240910-ROMETA).

 DOI: 10.3982/ECTA8718

each v in some set $\mathcal{V} \subseteq \mathbb{R}^d$, which may be uncountably infinite. The identification region for θ^* is then

$$\Theta_I = \bigcap_{v \in \mathcal{V}} \left[\theta^l(v), \theta^u(v)\right] = \left[\sup_{v \in \mathcal{V}} \theta^l(v), \inf_{v \in \mathcal{V}} \theta^u(v)\right]. \tag{1.1}$$

Intersection bounds stem naturally from exclusion restrictions (Manski (2003)) and appear in numerous applied and theoretical examples.[2] A leading case is that where the bounding functions are conditional expectations with continuous conditioning variables, yielding conditional moment inequalities. More generally, the methods of this paper apply to any estimator for the value of a linear programming problem with an infinite-dimensional constraint set.

This paper covers both parametric and nonparametric estimators of bounding functions $\theta^l(\cdot)$ and $\theta^u(\cdot)$. We provide formal justification for parametric, series, and kernel-type estimators via asymptotic theory based on the strong approximation of a sequence of empirical processes by a sequence of Gaussian processes. This includes an important new result on strong approximation for series estimators that applies to any estimator that admits a linear approximation, essentially providing a functional central limit theorem for series estimators for the first time in the literature. In addition, we generalize existing results on the strong approximation of kernel-type estimators to regression models with multivariate outcomes, and we provide a novel multiplier method to approximate the distribution of such estimators. For each of these estimation methods, the paper provides the following information:

(i) Confidence regions that achieve a desired asymptotic level.

(ii) Novel adaptive inequality selection (AIS) needed to construct sharp critical values, which in some cases result in confidence regions with exact asymptotic size.[3]

[2]Examples include average treatment effect bounds from instrumental variable restrictions (Manski (1990)), bounds on the distribution of treatment effects in a randomized experiment (Heckman, Smith, and Clements (1997)), treatment effect bounds from nonparametric selection equations with exclusion restrictions (Heckman and Vytlacil (1999)), monotone instrumental variables and the returns to schooling (Manski and Pepper (2000)), English auctions (Haile and Tamer (2003)), the returns to language skills (Gonzalez (2005)), changes in the distribution of wages (Blundell, Gosling, Ichimura, and Meghir (2007)), the study of disability and employment (Kreider and Pepper (2007)), unemployment compensation reform (Lee and Wilke (2009)), set identification with Tobin regressors (Chernozhukov, Rigobon, and Stoker (2010)), endogeneity with discrete outcomes (Chesher (2010)), estimation of income poverty measures (Nicoletti, Foliano, and Peracchi (2011)), bounds on average structural functions and treatment effects in triangular systems (Shaikh and Vytlacil (2011)), and set identification with imperfect instruments (Nevo and Rosen (2012)).

[3]The previous literature (e.g., Chernozhukov, Hong, and Tamer (2007)) and contemporaneous papers (such as Andrews and Shi (2013)) use “nonadaptive" cutoffs such as $C\sqrt{\log n}$. Ideally C should depend on the problem at hand and so careful calibration might be required in practice. Our new AIS procedure provides data-driven, adaptive cutoffs, which do not require calibration.

(iii) Convergence rates for the boundary points of these regions.

(iv) A characterization of local alternatives against which the associated tests have nontrivial power.

(v) Half-median-unbiased estimators of the intersection bounds.

Moreover, our paper also extends inferential theory based on empirical processes in Donsker settings to non-Donsker cases. The empirical processes arising in our problems do not converge weakly to a Gaussian process, but can be strongly approximated by a sequence of "penultimate" Gaussian processes, which we use directly for inference without resorting to further approximations, such as the extreme value approximations in Bickel and Rosenblatt (1973). These new methods may be of independent interest for a variety of other problems.

Our results also apply to settings where a parameter of interest, say μ, is characterized by intersection bounds of the form (1.1) on an auxiliary function $\theta(\mu)$. Then the bounding functions have the representation

$$\theta^l(v) := \theta^l(v;\mu) \quad \text{and} \quad \theta^u(v) := \theta^u(v;\mu), \tag{1.2}$$

and thus inference statements for $\theta^* := \theta(\mu)$ bounded by $\theta^l(\cdot)$ and $\theta^u(\cdot)$ can be translated to inference statements for the parameter μ. This includes cases where the bounding functions are a collection of conditional moment functions indexed by μ. When the auxiliary function is additively separable in μ, the relation between the two is simply a location shift. When the auxiliary function is nonseparable in μ, inference statements on θ^* still translate to inference statements on μ, though the functional relation between the two is more complex.

This paper overcomes significant complications for estimation of and inference on intersection bounds. First, because the bound estimates are suprema and infima of parametric or nonparametric estimators, closed-form characterization of their asymptotic distributions is typically unavailable or difficult to establish. As a consequence, researchers have often used the canonical bootstrap for inference, yet the recent literature indicates that the canonical bootstrap is not generally consistent in such settings; see, for example, Andrews and Han (2009), Bugni (2010), and Canay (2010).[4] Second, since sample analogs of the bounds of Θ_I are the suprema and infima of estimated bounding functions, they have substantial finite sample bias, and estimated bounds tend to be much tighter than the population bounds. This was noted by Manski and Pepper (2000, 2009), and some heuristic bias adjustments have been proposed by Haile and Tamer (2003) and Kreider and Pepper (2007).

Note that our AIS procedure could be iterated via stepdown, for example, as in Chetverikov (2012). We omit the details for brevity.

[4]Recent papers by Andrews and Shi (2013) and Kim (2009) provide justification for subsampling procedures for the statistics they employ for inference with conditional moment inequalities. We discuss these papers further in our literature review below.

We solve the problem of estimation and inference for intersection bounds by proposing bias-corrected estimators of the upper and lower bounds, as well as confidence intervals. Specifically, our approach employs a precision correction to the estimated bounding functions $v \mapsto \hat{\theta}^l(v)$ and $v \mapsto \hat{\theta}^u(v)$ before applying the supremum and infimum operators. We adjust the estimated bounding functions for their precision by adding to each of them an appropriate critical value times their pointwise standard error. Then, depending on the choice of the critical value, the intersection of these precision-adjusted bounds provides (i) confidence sets for either the identified set Θ_I or the true parameter value θ^*, or (ii) bias-corrected estimators for the lower and upper bounds. Our bias-corrected estimators are half-median-unbiased in the sense that the upper bound estimator $\hat{\theta}^u$ exceeds θ^u and the lower bound estimator $\hat{\theta}^l$ falls below θ^l each with probability at least $\frac{1}{2}$ asymptotically. Due to the presence of the inf and sup operators in the definitions of θ^u and θ^l, achieving unbiasedness is impossible in general, as shown by Hirano and Porter (2012), and this motivates our half-unbiasedness property. Bound estimators with this property are also proposed by Andrews and Shi (2013; henceforth AS). An attractive feature of our approach is that the only difference in the construction of our estimators and confidence intervals is the choice of a critical value. Thus, practitioners need not implement two entirely different methods to construct estimators and confidence bands with desirable properties.

This paper contributes to a growing literature on inference on set-identified parameters bounded by inequality restrictions. The prior literature focused primarily on models with a finite number of unconditional inequality restrictions. Some examples include Andrews and Jia (2012), Andrews and Guggenberger (2009), Andrews and Soares (2010), Beresteanu and Molinari (2008), Bugni (2010), Canay (2010), Chernozhukov, Hong, and Tamer (2007), Galichon and Henry (2009), Romano and Shaikh (2008, 2010), and Rosen (2008), among others. We contribute to this literature by considering inference with a continuum of inequalities. Contemporaneous and independently written research on conditional moment inequalities includes AS, Kim (2009), and Menzel (2009). Our approach differs from all of these. Whereas we treat the problem with fundamentally nonparametric methods, AS provided inferential statistics that transform the model's conditional restrictions to unconditional ones through the use of instrument functions.[5] In this sense our approach is similar in spirit to that of Härdle and Mammen (1993) (although they used the L^2 norm and we use a sup test), while the approach of AS parallels that of Bierens (1982) for testing a parametric specification against a nonparametric alternative. As such,

[5]Thus, the two approaches also require different assumptions. We rely on the strong approximation of a studentized version of parametric or nonparametric bounding function estimators (e.g., conditional moment functions in the context of conditional moment inequalities), while AS required that a functional central limit theorem hold for the *transformed* unconditional moment functions, which involve instrument functions not present in this paper.

these approaches are complementary, each with their relative merits, as we describe in further detail below. Kim (2009) proposed an inferential method related to that of AS, but where data-dependent indicator functions play the role of instrument functions. Menzel (2009) considered problems where the number of moment inequalities defining the identified set is large relative to the sample size. He provided results on the use of a subset of such restrictions in any finite sample, where the number of restrictions employed grows with the sample size, and examined the sensitivity of estimation and inference methods to the rate with which the number of moments used grows with the sample size.

The classes of models to which our approach and others in the recent literature apply have considerable overlap, most notably in models comprised of conditional moment inequalities and, equivalently, models whose bounding functions are conditional moment functions. Relative to other approaches, our approach is especially convenient for inference in parametric and nonparametric models with a continuum of inequalities that are separable in parameters, that is, those admitting representations of the form

$$\sup_{v\in\mathcal{V}} \theta^l(v) \leq \theta^* \leq \inf_{v\in\mathcal{V}} \theta^u(v),$$

as in (1.1). Our explicit use of nonparametric estimation of bounding functions renders our method applicable in settings where the bounding functions depend on covariates in addition to the variable V, that is, where the function $\theta(x)$ *at a point* x is the object of interest, with

$$\sup_{v\in\mathcal{V}} \theta^l(x, v) \leq \theta(x) \leq \inf_{v\in\mathcal{V}} \theta^u(x, v).$$

When the functions $\theta^l(x, v)$ and $\theta^u(x, v)$ are nonparametrically specified, these can be estimated by either the series or kernel-type estimators we study in Section 4. At present most other approaches do not appear to immediately apply when we are interested in $\theta(x)$ at a point x, when covariates X are continuously distributed, with the exception of the recent work by Fan and Park (2011) in the context of instrumental variable (IV) and monotone instrumental variable (MIV) bounds, and that of Andrews and Shi (2011), which extends methods developed in AS to this case.[6]

To better understand the comparison between our point and interval estimators and those of AS when both are applicable, consider as a simple example the case where $\theta^* \leq E[Y|V]$ almost surely, with $E[Y|V = v]$ continuous in v. Then the upper bound on θ^* is $\theta_0 = \inf_{v\in\mathcal{V}} E[Y|V = v]$ over some region $\mathcal{V}$. θ_0 is a nonparametric functional and can, in general, only be estimated at a

[6]The complication is that inclusion of additional covariates in a nonparametric framework requires a method for localization of the bounding function around the point x. With some nontrivial work and under appropriate conditions, the other approaches can likely be adapted to this context.

nonparametric rate: that is, one cannot construct point or interval estimators that converge to θ_0 at superefficient rates, that is, rates that exceed the optimal nonparametric rate for estimating $\theta(v) := \theta^u(v) = E[Y|V = v]$.[7] Our procedure delivers point and interval estimators that can converge to θ_0 at this rate, up to an undersmoothing factor. However, there exist point and interval estimators that can achieve faster (superefficient) convergence rates at *some* values of $\theta(\cdot)$. In particular, if the bounding function $\theta(\cdot)$ happens to be flat on the argmin set $V_0 = \{v \in \mathcal{V} : \theta(v) = \theta_0\}$, meaning that V_0 is a set of positive Lebesgue measure, then the point and interval estimator of AS can achieve the convergence rate of $n^{-1/2}$. As a consequence, their procedure for testing $\theta_{na} \leq \theta_0$ against $\theta_{na} > \theta_0$, where $\theta_{na} = \theta_0 + C/\sqrt{n}$ for $C > 0$, has nontrivial asymptotic power, while our procedure does not. If, however, $\theta(\cdot)$ is not flat on V_0, then the testing procedure of AS no longer has power against the aforementioned $n^{-1/2}$ alternatives, and results in point and interval estimators that converge to θ_0 at a suboptimal rate.[8] In contrast, our procedure delivers point and interval estimators that can converge at nearly the optimal rate, and hence can provide better power in these cases. In applications, both flat and nonflat cases are important, and we therefore believe that both testing procedures are useful.[9] For further comparisons, we refer the reader to our Monte Carlo section (Section 7) and to the Supplemental Material (Chernozhukov, Lee, and Rosen (2013)) Appendices K and L, which confirm these points both analytically and numerically.[10]

There are some more recent additions to the literature on conditional moment inequalities. Lee, Song, and Whang (2013) developed a test for condi-

[7]Suppose, for example, that $V_0 = \arg\inf_{v\in\mathcal{V}} \theta(v)$ is a singleton, with $\theta_0 = \theta(v)$ for some $v \in \mathcal{V}$. Then θ_0 is a nonparametric function evaluated at a single point, which cannot be estimated faster than the optimal nonparametric rate. Lower bounds on the rates of convergence in nonparametric models are characterized, for example, by Stone (1982) and Tsybakov (2009). Having a uniformly superefficient procedure would contradict these lower bounds.

[8]With regard to confidence intervals/interval estimators, we mean here that the upper bound of the confidence interval does not converge at this rate.

[9]Note also that nonflat cases can be justified as generic if, for example, one takes $\theta(\cdot)$ as a random draw from the Sobolev ball equipped with the Gaussian (Wiener) measure.

[10]See Supplemental Material Appendix K for specific examples, and see Armstrong (2011b) for a comprehensive analysis of the power properties of the procedure of Andrews and Shi (2013). We also note that this qualitative comparison of local asymptotic power properties conforms with related results regarding tests of parametric models versus nonparametric (PvNP) alternatives, which involve moment *equalities*. Recall that our test relies on nonparametric estimation of bound-generating functions, which often take the form of conditional moment inequalities, and is similar in spirit to the approach of, for example, Härdle and Mammen (1993) in the PvNP testing literature. On the other hand, the statistics employed by AS rely on a transformation of conditional restrictions to unconditional ones in similar spirit to Bierens (1982). Tests of the latter type have been found to have power against some $n^{-1/2}$ alternatives, while the former do not. However, tests of the first type typically have nontrivial power against a larger class of alternatives, and so achieve higher power against some classes of alternatives. For further details, see, for example, Horowitz and Spokoiny (2001) and the references therein.

tional moment inequalities using a one-sided version of L^p-type functionals of kernel estimators. Their approach is based on a least favorable configuration that permits valid but possibly conservative inference using standard normal critical values. Armstrong (2011b) and Chetverikov (2011) both proposed interesting and important approaches to estimation and inference based on conditional moment inequalities, which can be seen as introducing full studentization in the procedure of AS, fundamentally changing its behavior. The resulting procedures use a collection of fully studentized nonparametric estimators for inference, bringing them much closer to the approach of this paper. Their implicit nonparametric estimators are locally constant with an adaptively chosen bandwidth. Our approach is specifically geared to smooth cases, where $\theta^u(\cdot)$ and $\theta^l(\cdot)$ are continuously differentiable of order $s \geq 2$, resulting in more precise estimates and hence higher power in these cases.[11] On the other hand, in less smooth cases, the procedures of Armstrong (2011b) and Chetverikov (2011) automatically adapt to deliver optimal estimation and testing procedures, and so can perform somewhat better than our approach in these cases. Armstrong (2011a) derived the convergence rate and asymptotic distribution for a test statistic related to that in Armstrong (2011b) when evaluated at parameter values on the boundary of the identified set, drawing a connection to the literature on nonstandard M-estimation. Ponomareva (2010) studied bootstrap procedures for inference using kernel-based estimators, including one that can achieve asymptotically exact inference when the bounding function is uniquely maximized at a single point and locally quadratic. Our simulation-based approach does not rely on these conditions for its validity, but will automatically achieve asymptotic exactness with appropriately chosen smoothing parameters in a sufficiently regular subset of such cases.

Plan of the Paper

We organize the paper as follows. In Section 2, we motivate the analysis with examples and provide an informal overview of our results. In Section 3, we provide a formal treatment of our method under high-level conditions. In Section 4, we provide conditions and theorems for validity for parametric and nonparametric series and kernel-type estimators. We provide several examples that demonstrate the use of primitive conditions to verify the conditions of Section 3. This includes sufficient conditions for the application of these estimators to models comprised of conditional moment inequalities. In Section 5, we provide a theorem that establishes strong approximation for series estimators admitting an asymptotic linear representation and that covers the examples of Section 4. Likewise, we provide a theorem that establishes strong approximation for kernel-type estimators in Appendix G.2 of the Supplemental Material.

[11]Note that to harness power gains higher order kernels or series estimators should be used; our analysis allows for either.

In Section 6, we provide step-by-step implementation guides for parametric and nonparametric series and kernel-type estimators. In Section 7, we illustrate the performance of our method using both series and kernel-type estimators in Monte Carlo experiments, which we compare to that of AS in terms of coverage frequency and power. Our method performs well in these experiments, and we find that our approach and that of AS perform favorably in different models, depending on the shape of the bounding function. Section 8 concludes. In Appendices A–D, we recall the definition of strong approximation and provide proofs, including the proof of the strong approximation result for series estimators. The Supplemental Material contains further appendices. The first of these, Appendix E, provides proofs omitted from the main text.[12] Appendices F–H concern kernel-type estimators, providing primitive conditions for their application to conditional moment inequalities, strong approximation results, and the multiplier method, enabling inference via simulation, and proofs. Appendix I provides additional details on the use of primitive conditions to verify an asymptotic linear expansion needed for strong approximation of series estimators and Appendix J gives some detailed arguments for local polynomial estimation of conditional moment inequalities. Appendix K provides local asymptotic power analysis that supports the findings of our Monte Carlo experiments. Appendix L provides further Monte Carlo evidence.

Notation

For any two reals a and b, $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. $Q_p(X)$ denotes the pth quantile of random variable X. We use wp $\to 1$ as shorthand for "with probability approaching 1 as $n \to \infty$." We write $\mathcal{N}_k =_d N(0, I_k)$ to denote that the k-variate random vector $\mathcal{N}_k$ is distributed multivariate normal with mean zero and variance the $k \times k$ identity matrix. To denote probability statements conditional on observed data, we write statements conditional on $\mathcal{D}_n$. $\mathbb{E}_n$ and $\mathbb{P}_n$ denote the sample mean and empirical measure, respectively. That is, given independent and identically distributed (i.i.d.) random vectors $X_1, \dots, X_n$, we have $\mathbb{E}_n f = \int f \, d\mathbb{P}_n = n^{-1} \sum_{i=1}^n f(X_i)$. In addition, let $\mathbb{G}_n f = \sqrt{n}(\mathbb{E}_n - E) f = n^{-1/2} \sum_{i=1}^n [f(X_i) - Ef(X)]$. The notation $a_n \lesssim b_n$ means that $a_n \leq C b_n$ for all n; $X_n \lesssim_{\mathrm{P}_n} c_n$ abbreviates $X_n = O_{\mathrm{P}_n}(c_n)$. $X_n \to_{\mathrm{P}_n} \infty$ means that for any constant $C > 0$, $\mathrm{P}_n(X_n < C) \to 0$. We use V to denote a generic compact subset of $\mathcal{V}$, and we write $\operatorname{diam}(\mathsf{V})$ to denote the diameter of V in the Euclidean metric. $\| \cdot \|$ denotes the Euclidean norm, and for any two sets A, B in Euclidean space, $d_H(A, B)$ denotes the Hausdorff pseudo-distance between A and B with respect to the Euclidean norm. C stands for a generic positive constant, which may be different in different places, unless stated otherwise. For a set V and an element v in Euclidean space, let $d(v, \mathsf{V}) := \inf_{v' \in \mathsf{V}} \|v - v'\|$. For a function $p(v)$, let $\operatorname{lip}(p)$ denote the Lipschitz

[12]Specifically, Appendix E contains the proofs of Lemmas 2 and 4.

coefficient, that is, $\mathrm{lip}(p) := L$ such that $\|p(v_1) - p(v_2)\| \le L\|v_1 - v_2\|$ for all v_1 and v_2 in the domain of $p(v)$.

2. MOTIVATING EXAMPLES AND INFORMAL OVERVIEW OF RESULTS

In this section, we briefly describe three examples of intersection bounds from the literature and provide an informal overview of our results.

EXAMPLE A—Treatment Effects and Instrumental Variables: In the analysis of treatment response, the ability to uniquely identify the distribution of potential outcomes is typically lacking without either experimental data or strong assumptions. This owes to the fact that for each individual unit of observation, only the outcome from the received treatment is observed; the counterfactual outcome that would have occurred given a different treatment is not known. Although we focus here on treatment effects, similar issues are present in other areas of economics. In the analysis of markets, for example, observed equilibrium outcomes reveal quantity demanded at the observed price, but do not reveal what demand would have been at other prices.

Suppose only that the support of the outcome space is known, $Y \in [0, 1]$, but no other assumptions are made regarding the distribution of counterfactual outcomes. Manski (1989, 1990) provided worst-case bounds on mean treatment outcomes for any treatment t conditional on observables $(X, V) = (x, v)$,

$$\theta^l(x, v) \le E\big[Y(t)|X = x, V = v\big] \le \theta^u(x, v),$$

where the bounds are

$$\theta^l(x, v) := E\big[Y \cdot 1\{Z = t\}|X = x, V = v\big],$$
$$\theta^u(x, v) := E\big[Y \cdot 1\{Z = t\} + 1\{Z \ne t\}|X = x, V = v\big],$$

where Z is the observed treatment. If V is an instrument satisfying $E[Y(t)|X, V] = E[Y(t)|X]$, then for any fixed x, bounds on $\theta^* := \theta^*(x) := E[Y(t)|X = x]$ are given by

$$\sup_{v \in \mathcal{V}} \theta^l(x, v) \le \theta^*(x) \le \inf_{v \in \mathcal{V}} \theta^u(x, v)$$

for any $\mathcal{V} \subseteq \mathrm{Supp}(V|X = x)$, where the subset $\mathcal{V}$ will be taken as known for estimation purposes. Similarly, bounds implied by restrictions such as monotone treatment response, monotone treatment selection, and monotone instrumental variables, as in Manski (1997) and Manski and Pepper (2000), also take the form of intersection bounds.

EXAMPLE B—Bounding Distributions to Account for Selection: Similar analysis applies to inference on distributions whose observations are censored due to selection. This approach was used by Blundell et al. (2007) to study changes in male and female wages. The starting point of their analysis is that the cumulative distribution $F(w|x,v)$ of wages W at any point w, conditional on observables $(X,V)=(x,v)$, must satisfy the worst-case bounds

$$\theta^l(x,v) \le F(w|x,v) \le \theta^u(x,v), \tag{2.1}$$

where D is an indicator of employment, and hence observability of W, so that

$$\theta^l(x,v) := E\big[D\cdot 1\{W\le w\}|X=x, V=v\big],$$
$$\theta^u(x,v) := E\big[D\cdot 1\{W\le w\}+(1-D)|X=x, V=v\big].$$

This relation is used to bound quantiles of conditional wage distributions. Additional restrictions motivated by economic theory are then used to tighten the bounds.

One such restriction is an exclusion restriction of the continuous variable out-of-work income, V. They considered the use of V as either an excluded or monotone instrument. The former restriction implies bounds on the parameter $\theta^* := F(w|x)$,

$$\sup_{v\in\mathcal{V}} \theta^l(x,v) \le F(w|x) \le \inf_{v\in\mathcal{V}} \theta^u(x,v) \tag{2.2}$$

for any $\mathcal{V}\subseteq \text{support}(V|X=x)$, while the weaker monotonicity restriction, namely that $F(w|x,v)$ is weakly increasing in v, implies the bounds on $\theta^* := F(w|x,v_0)$ for any v_0 in support$(V|X=x)$,

$$\sup_{v\in\mathcal{V}_l} \theta^l(x,v) \le F(w|x,v_0) \le \inf_{v\in\mathcal{V}_u} \theta^u(x,v), \tag{2.3}$$

where $\mathcal{V}_l=\{v\in\mathcal{V}: v\le v_0\}$ and $\mathcal{V}_u=\{v\in\mathcal{V}: v\ge v_0\}$.

EXAMPLE C—(Conditional) Conditional Moment Inequalities: Our inferential method can also be used for pointwise inference on parameters restricted by (possibly conditional) conditional moment inequalities. Such restrictions arise naturally in empirical work in industrial organization; see, for example, Pakes, Porter, Ho, and Ishii (2005) and Berry and Tamer (2007).

To illustrate, consider the restriction

$$E\big[m_j(X,\mu_0)|Z=z\big]\ge 0 \quad \text{for all } j=1,\dots,J \text{ and } z\in\mathcal{Z}_j, \tag{2.4}$$

where each $m_j(\cdot,\cdot)$, $j=1,\dots,J$, is a real-valued function, (X,Z) are observables, and μ_0 is the parameter of interest. Note that this parameter can depend on a particular covariate value. Suppose, for instance, that $Z=(Z_1,Z_2)$ and interest lies in the subgroup of the population with $Z_1 = z_1$, so that the researcher wishes to condition on $Z_1 = z_1$. In this case, $\mu_0 = \mu_0(z_1)$ depends on z_1. Conditioning on this value, we have from (2.4) that

$$E\big[m_j(X, \mu_0)|Z_1 = z_1, Z_2 = z_2\big] \geq 0$$
$$\text{for all } j = 1, \ldots, J \text{ and } z_2 \in \operatorname{Supp}(Z_2|Z_1 = z_1),$$

which is equivalent to (2.4) with $\mathcal{Z}_j = \operatorname{Supp}(Z|Z_1 = z_1)$. Note also that regions $\mathcal{Z}_j$ can depend on the inequality j as in (2.3) of the previous example and that the previous two examples can, in fact, be cast as special cases of this one.

Suppose that we would like to test (2.4) at level α for the conjectured parameter value $\mu_0 = \mu$ against an unrestricted alternative. To see how this can be done, define

$$v = (z, j), \quad \mathcal{V} := \big\{(z, j) : z \in \mathcal{Z}_j, j \in \{1, \ldots, J\}\big\},$$
$$\theta(\mu, v) := E\big[m_j(X, \mu)|Z = z\big],$$

and $\hat{\theta}(\mu, v)$ a consistent estimator. Under some continuity conditions, this is equivalent to a test of $\theta_0(\mu) := \inf_{v \in \mathcal{V}} \theta(\mu, v) \geq 0$ against $\inf_{v \in \mathcal{V}} \theta(\mu, v) < 0$. Our method for inference delivers a statistic

$$\hat{\theta}_\alpha(\mu) = \inf_{v \in \mathcal{V}} \big[\hat{\theta}(\mu, v) + \hat{k} \cdot s(\mu, v)\big]$$

such that $\limsup_{n \to \infty} P(\theta_0(\mu) \geq \hat{\theta}_\alpha(\mu)) \leq \alpha$ under the null hypothesis. Here, $s(\mu, v)$ is the standard error of $\hat{\theta}(\mu, v)$ and $\hat{k}$ is an estimated critical value, as we describe below. If $\hat{\theta}_\alpha(\mu) < 0$, we reject the null hypothesis, while if $\hat{\theta}_\alpha(\mu) \geq 0$, we do not.

Informal Overview of Results

We now provide an informal description of our method for estimation and inference. Consider an upper bound θ_0 on θ^* of the form

$$\theta^* \leq \theta_0 := \inf_{v \in \mathcal{V}} \theta(v), \tag{2.5}$$

where $v \mapsto \theta(v)$ is a bounding function and $\mathcal{V}$ is the set over which the infimum is taken. We focus on describing our method for the upper bound (2.5), as the lower bound is entirely symmetric. In fact, any combination of upper and lower bounds can be combined into upper bounds on an auxiliary function of θ^* of the form (2.5), and this can be used for inference on θ^*, as we describe in Section 6.[13]

[13]Alternatively, one can combine one-sided intervals for lower and upper bounds for inference on the identified set Θ_I using Bonferroni's inequality, or for inference on θ^* using the method

What are good *estimators* and *confidence regions* for the bound θ_0? A natural idea is to base estimation and inference on the sample analog: $\inf_{v\in\mathcal{V}} \hat{\theta}(v)$. However, this estimator does not perform well in practice. First, the analog estimator tends to be downward biased in finite samples. As discussed in the Introduction, this will typically result in bound estimates that are much narrower than those in the population; see, for example, Manski and Pepper (2000, 2009) for more on this point. Second, inference must appropriately take into account sampling error of the estimator $\hat{\theta}(v)$ across all values of v. Indeed, different levels of precision of $\hat{\theta}(v)$ at different points can severely distort the perception of the minimum of the bounding function $\theta(v)$. Figure 1 illustrates these problems geometrically. The solid curve is the true bounding function $v \mapsto \theta(v)$ and the dash-dotted thick curve is its estimate $v \mapsto \hat{\theta}(v)$. The remaining dashed curves represent eight additional potential realizations of the estimator, illustrating its precision. In particular, we see that the precision of the estimator is much lower on the right side than on the left. A naïve sample analog estimate for θ_0 is provided by the minimum of the dash-dotted curve, but this estimate can, in fact, be quite far away from θ_0. This large deviation from the true value arises from both the lower precision of the estimated curve on the right side of the figure and from the downward bias created by taking the minimum of the estimated curve.

To overcome these problems, we propose a *precision-corrected* estimate of θ_0,

$$\hat{\theta}_0(p) := \inf_{v\in\mathcal{V}}\big[\hat{\theta}(v) + k(p)\cdot s(v)\big], \tag{2.6}$$

where $s(v)$ is the standard error of $\hat{\theta}(v)$, and $k(p)$ is a critical value, the selection of which is described below. That is, our estimator $\hat{\theta}_0(p)$ minimizes the *precision-corrected curve* given by $\hat{\theta}(v)$ plus critical value $k(p)$ times the pointwise standard error $s(v)$. Figure 2 shows a precision-corrected curve as a dashed curve with a particular choice of critical value k. In this figure, we see that the minimizer of the precision-corrected curve can indeed be much closer to θ_0 than the sample analog $\inf_{v\in\mathcal{V}} \hat{\theta}(v)$.

These issues are important both in theory and in practice, as can be seen in the application in our working paper version (Chernozhukov, Lee, and Rosen (2009)). There we used the data from the National Longitudinal Survey of Youth of 1979 (NLSY79), as in Carneiro and Lee (2009), to estimate bounds on expected log wages Y_i as a function of years of schooling t. We used Armed Forces Qualifying Test score (AFQT) normalized to have mean zero as a monotone instrumental variable, and estimated the MIV-MTR (mono-

described in Chernozhukov, Lee, and Rosen (2009, Section 3.7), which is a slight generalization of methods previously developed by Imbens and Manski (2004) and Stoye (2009).

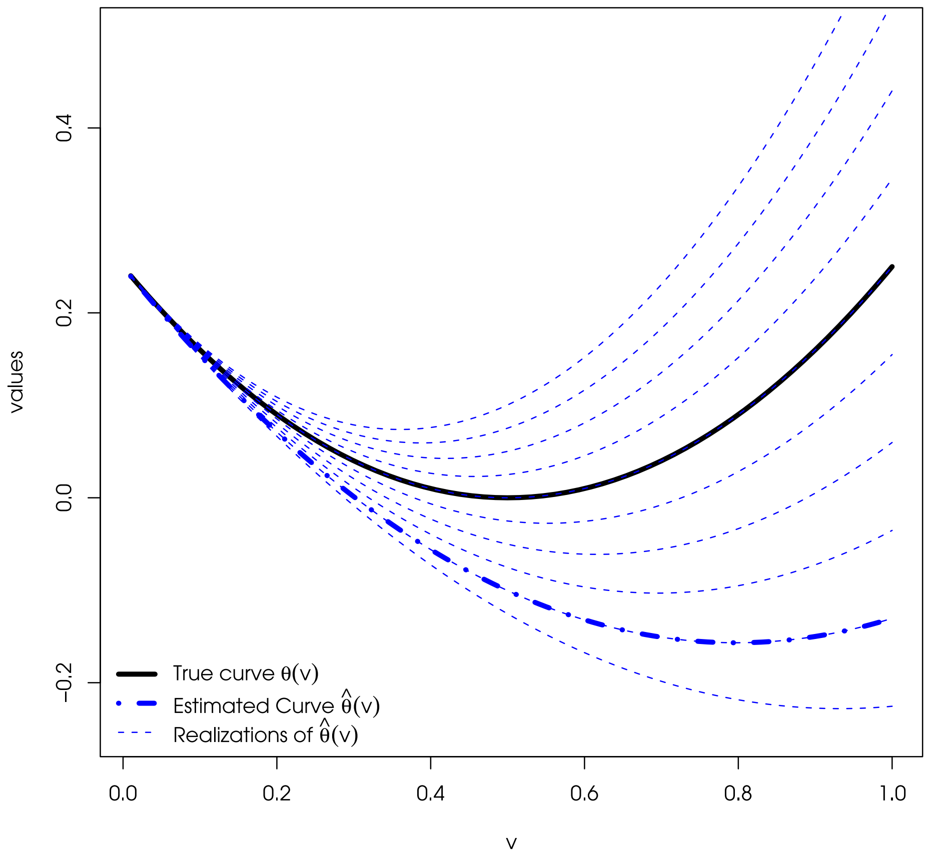

FIGURE 1.—Illustration of how variation in the precision of the analog estimator at different points may impede inference. The solid curve is the true bounding function $\theta(v)$, while the dash–dotted curve is a single realization of its estimator, $\hat{\theta}(v)$. The lighter dashed curves depict eight additional representative realizations of the estimator, illustrating its precision at different values of v. The minimum of the estimator $\hat{\theta}(v)$ is indeed quite far from the minimum of $\theta(v)$, making the empirical upper bound unduly tight.

tone instrument variable–monotone treatment response) bounds of Manski and Pepper (2000). Figures 3 and 4 highlight the same issues as the schematic Figures 1 and 2 with the NLSY data and the MIV-MTR upper bound for the parameter $\theta^* = \mathrm{P}[Y_i(t) > y|V_i = v]$, at $y = \log(24)$ ($\sim$90th percentile of hourly wages) and $v = 0$ for college graduates ($t = 16$).[14]

[14]The parameter θ^* used for this illustration differs from the conditional expectations bounded in Chernozhukov, Lee, and Rosen (2009). For further details regarding the application and the data, we refer to that version, which is available at http://cemmap.ifs.org.uk/wps/cwp1909.pdf.

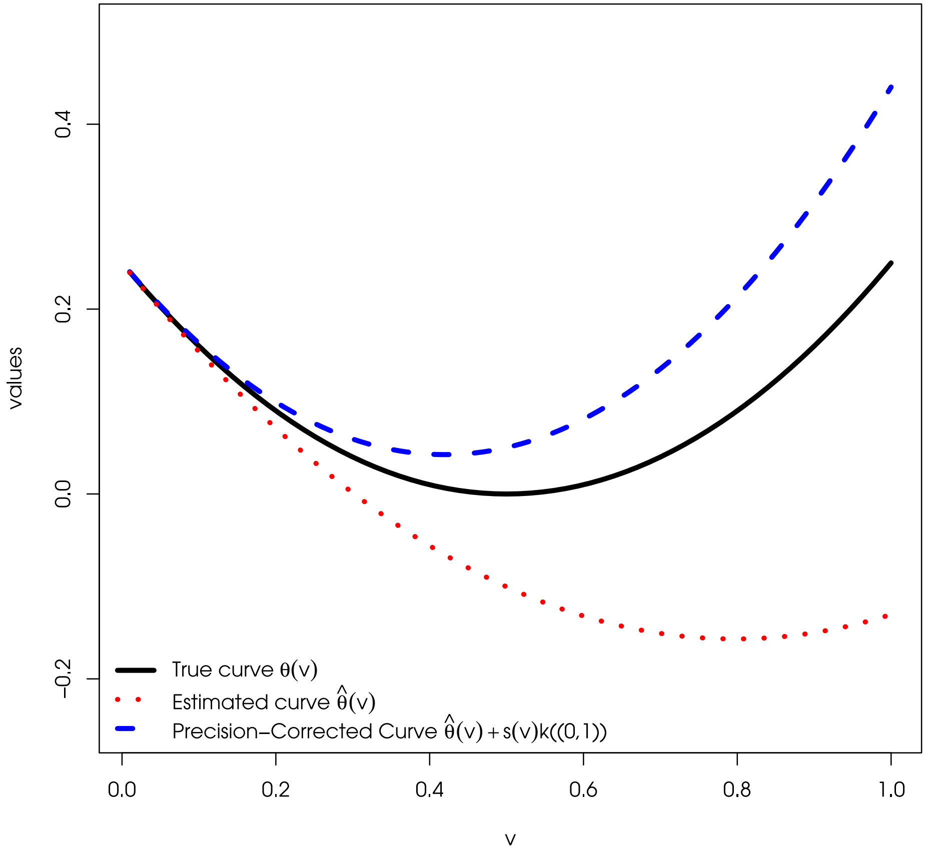

FIGURE 2.—Depiction of a precision-corrected curve (dashed curve) that adjusts the boundary estimate $\hat{\theta}(v)$ (dotted curve) by an amount proportional to its pointwise standard error. The minimum of the precision-corrected curve is closer to the minimum of the true curve (solid) than the minimum of $\hat{\theta}(v)$, removing the downward bias.

Figure 3 shows the nonparametric series estimate of the bounding function using B-splines as described in Section 7.2 (solid curve) and 20 bootstrap estimates (dashed curves). The precision of the estimate is worst when the AFQT is near 2, as demonstrated by the bootstrap estimates. At the same time, the bounding function has a decreasing shape with the minimum at AFQT $= 2$. Figure 4 shows a precision-corrected curve (solid curve) that adjusts the bound estimate $\hat{\theta}(v)$ (dashed curve) by an amount proportional to its pointwise standard error; the horizontal dashed line shows the end point of a 95% one-sided confidence interval. As in Figure 2, the minimizer of the precision-corrected curve is quite far from that of the uncorrected estimate of the bounding function.

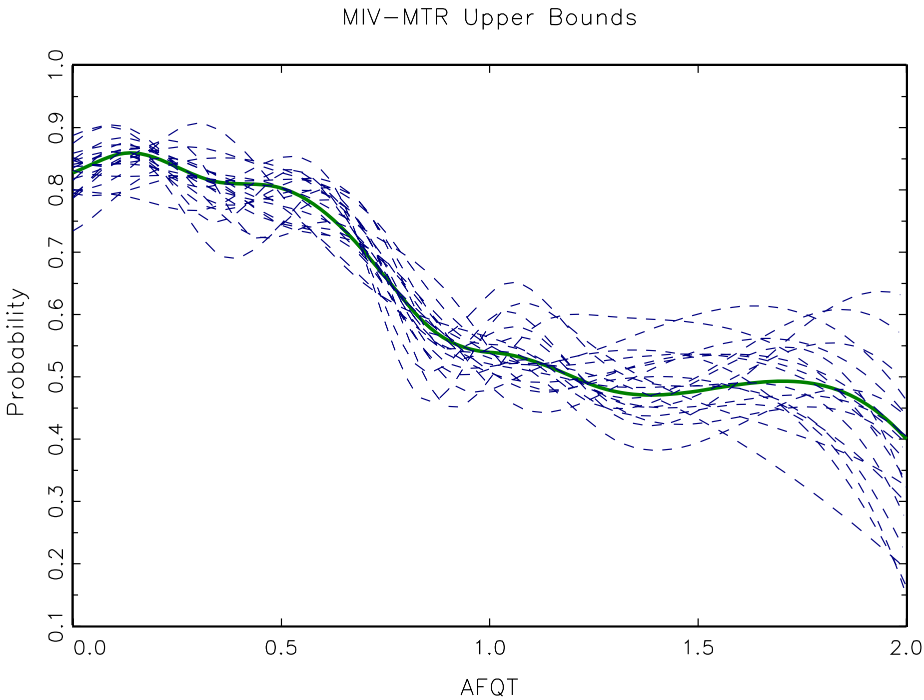

FIGURE 3.—Depiction based on NLSY data and the application in Section 5 of Chernozhukov, Lee, and Rosen (2009) of an estimate of the bounding function (solid curve) and 20 bootstrapped estimates (dashed curves).

The degree of precision correction, both in these figures and in general, is driven by the critical value $k(p)$. The main input in the selection of $k(p)$ for the estimator $\hat{\theta}_0(p)$ in (2.6) is the standardized process

$$Z_n(v) = \frac{\theta(v) - \hat{\theta}(v)}{\sigma(v)},$$

where $\sigma(v)/s(v) \to 1$ in probability uniformly in v. Generally, the finite sample distribution of the process Z_n is unknown, but we can approximate it uniformly by a sequence of Gaussian processes Z_n^* such that for an appropriate sequence of constants $\bar{a}_n$,

$$\bar{a}_n \sup_{v \in \mathcal{V}} \left| Z_n(v) - Z_n^*(v) \right| = o_p(1). \tag{2.7}$$

For any compact set V, used throughout to denote a generic compact subset of $\mathcal{V}$, we then approximate the quantiles of $\sup_{v \in \mathsf{V}} Z_n^*(v)$ either by analytical methods based on asymptotic approximations or by simulation. We then use

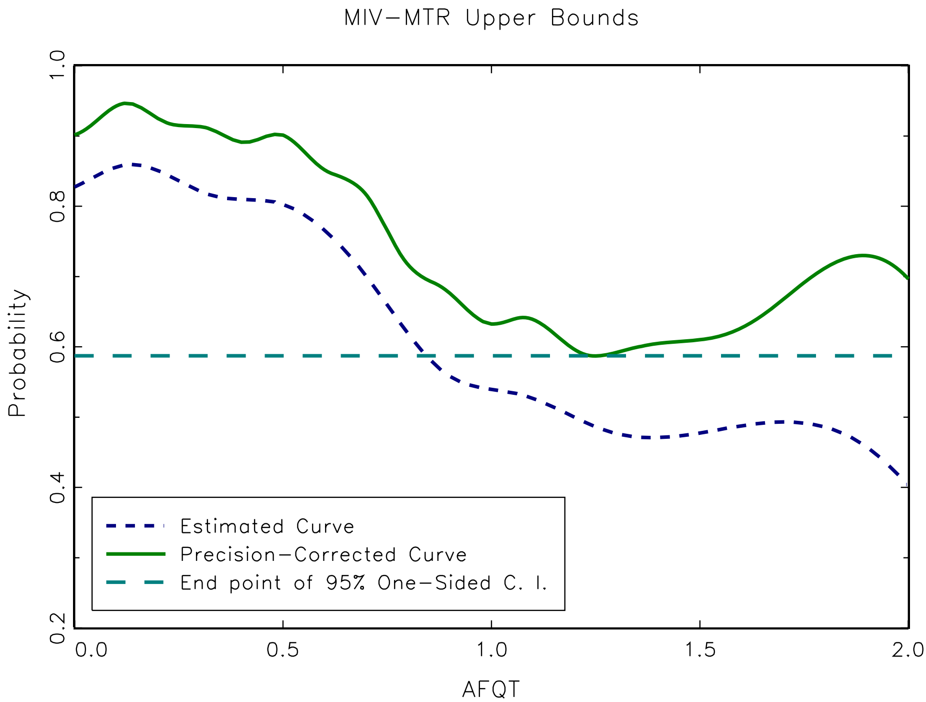

FIGURE 4.—Depiction based on NLSY data and the application in Section 5 of Chernozhukov, Lee, and Rosen (2009) of a precision-corrected curve (solid curve) that adjusts the boundary estimate $\hat{\theta}(v)$ (dashed curve) by an amount proportional to its pointwise standard error. The horizontal dashed line shows the end point of the 95% one-sided confidence interval.

the p-quantile of this statistic, $k_{n,\mathsf{V}}(p)$, in place of $k(p)$ in (2.6). We show that, in general, simulated critical values provide sharper inference and, therefore, we advocate their use.

For the estimator in (2.6) to exceed θ_0 with probability no less than p asymptotically, we require that wp $\to 1$ the set V contains the argmin set

$$V_0 := \arg\inf_{v \in \mathcal{V}} \theta(v).$$

A simple way to achieve this is to use $\mathsf{V} = \mathcal{V}$, which leads to asymptotically valid but conservative inference. For construction of the critical value $k_{n,\mathsf{V}}(p)$ above, we thus propose the use of a preliminary set estimator $\hat{V}_n$ for V_0 using a novel adaptive inequality selection procedure. Because the critical value $k_{n,\mathsf{V}}(p)$ is nondecreasing in V for n large enough, this yields an asymptotic critical value no larger than those based on $\mathsf{V} = \mathcal{V}$. The set estimator $\hat{V}_n$ is shown to be sandwiched between two nonstochastic sequences of sets: a lower envelope $\underline{V}_n$ and an upper envelope $\overline{V}_n$ with probability going to 1. We show in Lemma 1 that our inferential procedure using $\hat{V}_n$ concentrates on the lower envelope V_n, which is a neighborhood of the argmin set V_0. This validates our use of the set estimator $\hat{V}_n$. The upper envelope $\overline{V}_n$, a larger—but nonetheless shrinking—neighborhood of the argmin set V_0, plays an important role in the derivation of estimation rates and local power properties of our procedure. Specifically, because this set contains V_0 wp $\to 1$, the tail behavior of $\sup_{v \in \overline{V}_n} Z_n^*(v)$ can be used to bound the estimation error of $\hat{\theta}_0(p)$ relative to θ_0.

Moreover, we show that in some cases, inference based on simulated critical values using $\hat{V}_n$ in fact "concentrates" on V_0 rather than just V_n. These cases require the scaled penultimate process $\bar{a}_n Z_n^*$ to behave sufficiently well (i.e., to be stochastically equicontinuous) within r_n neighborhoods of V_0, where r_n denotes the rate of convergence of the set estimator $\hat{V}_n$ to the argmin set V_0. When this holds, the tail behavior of $\sup_{v \in V_0} Z_n^*(v)$ rather than $\sup_{v \in \overline{V}_n} Z_n^*(v)$ bounds the estimation error of our estimator. This typically leads to small improvements in the convergence rate of our estimator and the local power properties of our approach. The conditions for this to occur include the important special case where V_0 is singleton and where the bounding function is locally quadratic, although it can hold more generally. The formal conditions are given in Section 3.5, where we provide conditions for consistency and rates of convergence of $\hat{V}_n$ for V_0, and in Section 3.6, where we provide the aforementioned equicontinuity condition and a formal statement of the result regarding when inference concentrates on V_0.

At an abstract level, our method does not distinguish parametric estimators of $\theta(v)$ from nonparametric estimators; however, details of the analysis and regularity conditions are quite distinct. Our theory for nonparametric estimation relies on undersmoothing, although for locally constant or sign-preserving estimation of bounding functions, this does not appear essential, since the approximation bias is conservatively signed. In such cases, our inference algorithm still applies to nonparametric estimates of bounding functions without undersmoothing, although our theory would require some minor modifications to handle this case. We do not formally pursue this here, but we provide some simulation results for kernel estimation without undersmoothing as part of the additional Monte Carlo experiments reported in Supplemental Material Appendix L.

For all estimators, parametric and nonparametric, we employ strong approximation analysis to approximate the quantiles of $\sup_{v \in \mathsf{V}} Z_n(v)$, and we verify our conditions separately for each case. The formal definition of strong approximation is provided in Appendix A.

3. ESTIMATION AND INFERENCE THEORY UNDER GENERAL CONDITIONS

3.1. *Basic Framework*

In this and subsequent sections, we allow the model and the probability measure to depend on n. Formally, we work with a probability space $(A, \mathcal{A}, \mathrm{P}_n)$ throughout. This approach is conventionally used in asymptotic statistics to ensure robustness of statistical conclusions with respect to perturbations in P_n. It guarantees the validity of our inference procedure under any sequence of probability laws P_n that obey our conditions, including the case with fixed P. We thus generalize our notation in this section to allow model parameters to depend on n.

The basic setting is as follows:

CONDITION C.1—Setting: There is a nonempty compact set $\mathcal{V} \subset \mathcal{K} \subset \mathbb{R}^d$, where $\mathcal{V}$ can depend on n and $\mathcal{K}$ is a bounded fixed set, independent of n. There is a continuous real-valued function $v \mapsto \theta_n(v)$. There is an estimator $v \mapsto \hat{\theta}_n(v)$ of this function, which is an almost surely (a.s.) continuous stochastic process. There is a continuous function $v \mapsto \sigma_n(v)$ that represents non-stochastic normalizing factors bounded by $\bar{\sigma}_n := \sup_{v\in\mathcal{V}} \sigma_n(v)$, and there is an estimator $v \mapsto s_n(v)$ of these factors, which is an a.s. continuous stochastic process, bounded above by $\bar{s}_n := \sup_{v\in\mathcal{V}} s_n(v)$.

We are interested in constructing point estimators and one-sided interval estimators for

$$\theta_{n0} = \inf_{v\in\mathcal{V}} \theta_n(v).$$

The main input in this construction is the standardized process

$$Z_n(v) = \frac{\theta_n(v) - \hat{\theta}_n(v)}{\sigma_n(v)}.$$

In the following discussion, we require that this process can be approximated by a standardized Gaussian process in the metric space $\ell^\infty(\mathcal{V})$ of bounded functions that map $\mathcal{V}$ to $\mathbb{R}$, which can be simulated for inference.

CONDITION C.2—Strong Approximation: (a) Z_n is strongly approximated by a sequence of penultimate Gaussian processes Z_n^* that have zero mean and a.s. continuous sample paths,

$$\sup_{v\in\mathcal{V}} \left| Z_n(v) - Z_n^*(v) \right| = o_{\mathrm{P}_n}(\delta_n),$$

where $E_{\mathrm{P}_n}[(Z_n^*(v))^2] = 1$ for each $v \in \mathcal{V}$, and $\delta_n = o(\bar{a}_n^{-1})$ for the sequence of constants $\bar{a}_n$ defined in Condition C.3 below. (b) Moreover, for simulation purposes, there is a process $Z_n^\star$, whose distribution is zero-mean Gaussian conditional on the data $\mathcal{D}_n$ and such that $E_{\mathrm{P}_n}[(Z_n^\star(v))^2|\mathcal{D}_n] = 1$ for each $v \in \mathcal{V}$, that can approximate an identical copy $\bar{Z}_n^*$ of Z_n^*, where $\bar{Z}_n^*$ is independent of $\mathcal{D}_n$, namely there is an $o(\delta_n)$ term such that

$$\mathrm{P}_n\Big[\sup_{v\in\mathcal{V}}\big|\bar{Z}_n^*(v) - Z_n^\star(v)\big| > o(\delta_n)\big|\mathcal{D}_n\Big] = o_{\mathrm{P}_n}(1/\ell_n)$$

for some $\ell_n \to \infty$ chosen below.

For convenience we refer to Appendix A, where the definition of strong approximation is recalled. The penultimate process Z_n^* is often called a coupling, and we construct such couplings for parametric and nonparametric estimators under both high-level and primitive conditions. It is convenient to work with Z_n^*, since we can rely on the fine properties of Gaussian processes. Note that Z_n^* depends on n and generally does not converge weakly to a fixed Gaussian process, and, therefore, is not asymptotically Donsker. Nonetheless, we can perform either analytical or simulation-based inference based on these processes.

Our next condition captures the so-called *concentration* properties of Gaussian processes.

CONDITION C.3—Concentration: For all n sufficiently large and for any compact, non-empty $\mathsf{V} \subseteq \mathcal{V}$, there is a normalizing factor $a_n(\mathsf{V})$ that satisfies

$$1 \leq a_n(\mathsf{V}) \leq a_n(\mathcal{V}) =: \bar{a}_n, \quad a_n(\mathsf{V}) \text{ is weakly increasing in } \mathsf{V},$$

such that

$$\mathcal{E}_n(\mathsf{V}) := a_n(\mathsf{V})\Big(\sup_{v\in\mathsf{V}} Z_n^*(v) - a_n(\mathsf{V})\Big)$$

obeys

$$\mathrm{P}_n\big[\mathcal{E}_n(\mathsf{V}) \geq x\big] \leq \mathrm{P}[\mathcal{E} \geq x], \tag{3.1}$$

where $\mathcal{E}$ is a random variable with continuous distribution function such that for some $\eta > 0$,

$$\mathrm{P}(\mathcal{E} > x) \leq \exp(-x/\eta).$$

The *concentration* condition will be verified in our applications by appealing to the Talagrand–Samorodnitsky inequality for the concentration of the suprema of Gaussian processes, which is sharper than the classical concentration inequalities.[15] These concentration properties play a key role in our analysis, as they determine the uniform speed of convergence $\bar{a}_n\bar{\sigma}_n$ of the estimator

[15]For details, see Lemma 12 in Appendix C.1.

$\hat{\theta}_{n0}(p)$ to θ_{n0}, where the estimator is defined later. In particular, this property implies that for any compact $V_n \subseteq \mathcal{V}$, $E_{\mathrm{P}_n}[\sup_{v\in V_n} Z_n^*(v)] \lesssim \bar{a}_n$. As there is concentration, there is an opposite force, called *anti-concentration*, which implies that under Conditions C.2(a) and C.3, for any $\delta_n = o(1/\bar{a}_n)$, we have

$$\sup_{x\in\mathbb{R}} \mathrm{P}_n\left(\left|\sup_{v\in V_n} Z_n^*(v) - x\right| \le \delta_n\right) \to 0. \tag{3.2}$$

This follows from a generic anti-concentration inequality derived in Chernozhukov, Chetverikov, and Kato (2011), which is quoted in Appendix B for convenience. Anti-concentration simplifies the construction of our confidence intervals. Finally, the exponential tail property of $\mathcal{E}$ plays an important role in the construction of our adaptive inequality selector, introduced below, since it allows us to bound moderate deviations of the one-sided estimation noise, namely $\sup_{v\in\mathsf{V}} Z_n^*(v)$.

Our next assumption requires uniform consistency as well as suitable estimates of σ_n.

CONDITION C.4—Uniform Consistency: We have that

$$\text{(a)}\quad \bar{a}_n\bar{\sigma}_n = o(1) \quad \text{and} \quad \text{(b)}\quad \sup_{v\in\mathcal{V}}\left|\frac{s_n(v)}{\sigma_n(v)} - 1\right| = o_{\mathrm{P}_n}\left(\frac{\delta_n}{\bar{a}_n + \ell\ell_n}\right),$$

where $\ell\ell_n \nearrow \infty$ is a sequence of constants defined below.

In what follows, we let

$$\ell_n := \log n \quad \text{and} \quad \ell\ell_n := \log \ell_n,$$

but it should be noted that ℓ_n can be replaced by other slowly increasing sequences.

3.2. *The Inference and Estimation Strategy*

For any compact subset $\mathsf{V} \subseteq \mathcal{V}$ and $\gamma \in (0, 1)$, define

$$\kappa_{n,\mathsf{V}}(\gamma) := Q_\gamma\left(\sup_{v\in\mathsf{V}} Z_n^*(v)\right).$$

The following result is useful for establishing validity of our inference procedure.

LEMMA 1—Inference Concentrates on a Neighborhood V_n of V_0: *Under Conditions* C.1–C.4,

$$\mathrm{P}_n\left(\sup_{v\in\mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} \le x\right) \ge \mathrm{P}_n\left(\sup_{v\in V_n} Z_n^*(v) \le x\right) - o(1)$$

uniformly in $x \in [0, \infty)$, *where*

$$V_n := \{v \in \mathcal{V} : \theta_n(v) \leq \theta_{n0} + \kappa_n \sigma_n(v)\} \quad \textit{for} \quad \kappa_n := \kappa_{n,\mathcal{V}}(\gamma_n'), \tag{3.3}$$

where γ_n' *is any sequence such that* $\gamma_n' \nearrow 1$ *with* $\kappa_n \leq \bar{a}_n + \eta(\ell\ell_n + C')/\bar{a}_n$ *for some constant* $C' > 0$.

Thus, with probability converging to 1, the inferential process "concentrates" on a neighborhood of V_0 given by V_n. The "size" of the neighborhood is determined by κ_n, a high quantile of $\sup_{v\in\mathcal{V}} Z_n^*(v)$, which summarizes the maximal one-sided estimation error over $\mathcal{V}$. We use this to construct half-median-unbiased estimators for θ_{n0} as well as one-sided interval estimators for θ_{n0} with correct asymptotic level, based on analytical and simulation methods for obtaining critical values proposed below.

REMARK 1—Sharp Concentration of Inference: In general, it is not possible for the inferential processes to concentrate on subsets smaller than V_n. However, as shown, in Section 3.6, in some special cases (e.g., when V_0 is a well identified singleton), the inference process will, in fact, concentrate on V_0. In this case, our simulation-based construction will automatically adapt to deliver median-unbiased estimators for θ_{n0} as well as one-sided interval estimators for θ_{n0} with exact asymptotic size. Indeed, in the special but extremely important case of V_0 being singleton, we can achieve

$$\mathrm{P}_n\Big(\sup_{v\in V_n} Z_n^*(v) > x\Big) = \mathrm{P}\big(N(0,1) > x\big) - o(1)$$

under some regularity conditions. In this case, our simulation-based procedure will automatically produce a critical value that approaches the pth quantile of the standard normal, delivering asymptotically exact inference.

Our construction relies first on an *auxiliary critical value* $k_{n,\mathcal{V}}(\gamma_n)$, chosen so that wp $\to 1$,

$$k_{n,\mathcal{V}}(\gamma_n) \geq \kappa_{n,\mathcal{V}}(\gamma_n'), \tag{3.4}$$

where we set $\gamma_n := 1 - 0.1/\ell_n \nearrow 1$ and $\gamma_n \geq \gamma_n' = \gamma_n - o(1)$. This critical value is used to obtain a preliminary set estimator

$$\hat{V}_n = \Big\{v \in \mathcal{V} : \hat{\theta}_n(v) \leq \inf_{\tilde{v}\in\mathcal{V}}\big(\hat{\theta}_n(\tilde{v}) + k_{n,\mathcal{V}}(\gamma_n)s_n(\tilde{v})\big) + 2k_{n,\mathcal{V}}(\gamma_n)s_n(v)\Big\}. \tag{3.5}$$

The set estimator $\hat{V}_n$ is then used in the construction of the *principal critical value* $k_{n,\hat{V}_n}(p)$, $p \geq 1/2$, where we require that wp $\to 1$,

$$k_{n,\hat{V}_n}(p) \geq \kappa_{n,V_n}\big(p - o(1)\big). \tag{3.6}$$

The principal critical value is fundamental to our construction of confidence regions and estimators, which we now define.

DEFINITION 1—Generic Interval and Point Estimators: Let $p \geq 1/2$. Then our interval estimator takes the form

$$\hat{\theta}_{n0}(p) = \inf_{v \in \mathcal{V}} \big[\hat{\theta}_n(v) + k_{n,\hat{V}_n}(p) s_n(v)\big], \tag{3.7}$$

where the half-median-unbiased estimator corresponds to $p = 1/2$.

The principal and auxiliary critical values are constructed below so as to satisfy (3.4) and (3.6) using either analytical or simulation methods. As a consequence, we show in Theorems 1 and 2 that

$$\mathrm{P}_n\big\{\theta_{n0} \leq \hat{\theta}_{n0}(p)\big\} \geq p - o(1) \tag{3.8}$$

for any fixed $1/2 \leq p < 1$. The construction relies on the new set estimator $\hat{V}_n$, which we call an *adaptive inequality selector* (AIS), since it uses the problem-dependent cutoff $k_{n,\mathcal{V}}(\gamma_n)$, which is a bound on a high quantile of $\sup_{v \in \mathcal{V}} Z_n^*(v)$. The analysis therefore must take into account the moderate deviations (tail behavior) of the latter.

Before proceeding to the details of its construction, we note that the argument for establishing the coverage results and analyzing power properties of the procedure depends crucially on the result (proven in Lemma 2 below)

$$\mathrm{P}_n\{V_n \subseteq \hat{V}_n \subseteq \overline{V}_n\} \to 1,$$

where

$$\overline{V}_n := \big\{v \in \mathcal{V} : \theta_n(v) \leq \theta_{n0} + \bar{\kappa}_n \bar{\sigma}_n\big\} \quad \text{for} \quad \bar{\kappa}_n := 7(\bar{a}_n + \eta \ell\ell_n / \bar{a}_n), \tag{3.9}$$

where $\eta > 0$ is defined by Condition C.3. Thus, the preliminary set estimator $\hat{V}_n$ is sandwiched between two deterministic sequences of sets, facilitating the analysis of its impact on the convergence of $\hat{\theta}_{n0}(p)$ to θ_{n0}.

3.3. *Analytical Method and Its Theory*

Our first construction is quite simple and demonstrates the main—though not the finest—points. This construction uses the majorizing variable $\mathcal{E}$ that appears in Condition C.3.

DEFINITION 2—Analytical Method for Critical Values: For any compact set V and any $p \in (0, 1)$, we set

$$k_{n,\mathsf{V}}(p) = a_n(\mathsf{V}) + c(p)/a_n(\mathsf{V}), \tag{3.10}$$

where $c(p) = Q_p(\mathcal{E})$ is the pth quantile of the majorizing variable $\mathcal{E}$ defined in Condition C.3, where for any fixed $p \in (0,1)$, we require that $\mathsf{V} \mapsto k_{n,\mathsf{V}}(p)$ is weakly monotone increasing in V for sufficiently large n.

The first main result is as follows.

THEOREM 1—Analytical Inference, Estimation, and Power Under Conditions C.1–C.4: *Suppose Conditions C.1–C.4 hold. Consider the interval estimator given in Definition 1 with critical value function given in Definition 2. Then, for a given $p \in [1/2, 1)$, the following statements hold*:

(i) *The interval estimator has asymptotic level p*:

$$\mathrm{P}_n\{\theta_{n0} \le \hat{\theta}_{n0}(p)\} \ge p - o(1).$$

(ii) *The estimation risk is bounded by*, *wp* $\to 1$ *under* P_n,

$$\left|\hat{\theta}_{n0}(p) - \theta_{n0}\right| \le 4\bar{\sigma}_n\left(a_n(\overline{V}_n) + \frac{O_{\mathrm{P}_n}(1)}{a_n(\overline{V}_n)}\right) \lesssim_{\mathrm{P}_n} \bar{\sigma}_n \bar{a}_n.$$

(iii) *Hence*, *any alternative $\theta_{na} > \theta_{n0}$ such that*

$$\theta_{na} \ge \theta_{n0} + 4\bar{\sigma}_n\left(a_n(\overline{V}_n) + \frac{\mu_n}{a_n(\overline{V}_n)}\right), \quad \mu_n \to_{\mathrm{P}_n} \infty,$$

is rejected with probability converging to 1 *under* P_n.

Thus, $(-\infty, \hat{\theta}_{n0}(p)]$ is a valid one-sided interval estimator for θ_{n0}. Moreover, $\hat{\theta}_{n0}(1/2)$ is a half-median-unbiased estimator for θ_{n0} in the sense that

$$\liminf_{n\to\infty} \mathrm{P}_n\left[\theta_{n0} \le \hat{\theta}_{n0}(1/2)\right] \ge 1/2.$$

The rate of convergence of $\hat{\theta}_{n0}(p)$ to θ_{n0} is bounded above by the uniform rate $\bar{\sigma}_n\bar{a}_n$ for estimation of the bounding function $v \mapsto \theta_n(v)$. This implies that the test of $H_0: \theta_{n0} = \theta_{na}$ that rejects if $\theta_{na} > \hat{\theta}_{n0}(p)$ asymptotically rejects all local alternatives that are more distant[16] than $\bar{\sigma}_n\bar{a}_n$, including fixed alternatives as a special case. In Section 4 below, we show that in parametric cases, this results in power against $n^{-1/2}$ local alternatives. For series estimators, $\bar{a}_n\bar{\sigma}_n$ is proportional to $(\log n)^c\sqrt{K/n}$, where c is some positive constant and $K \to \infty$ is the number of series terms. For kernel-type estimators of bounding functions, the rate $\bar{a}_n\bar{\sigma}_n$ is proportional to $(\log n)^c/\sqrt{nh^d}$, where c is some positive constant and h is the bandwidth, assuming some undersmoothing is done. For example, if the bounding function is s-times differentiable, $\bar{\sigma}_n$ can be made close

[16]Here and below we ignore various constants appearing in front of terms like $\bar{\sigma}_n\bar{a}_n$.

to $(\log n/n)^{s/(2s+d)}$ apart from some undersmoothing factor by considering a local polynomial estimator; see Stone (1982). For both series and kernel-type estimators, we show below that $\bar{a}_n$ can be bounded by $\sqrt{\log n}$.

3.4. *Simulation-Based Construction and Its Theory*

Our main and preferred approach is based on the simple idea of simulating quantiles of relevant statistics.

DEFINITION 3—Simulation Method for Critical Values: For any compact set $\mathsf{V} \subseteq \mathcal{V}$, we set

$$k_{n,\mathsf{V}}(p) = Q_p\Big(\sup_{v\in\mathsf{V}} Z_n^\star(v)\big|\mathcal{D}_n\Big). \tag{3.11}$$

We have the following result for simulation inference, analogous to that obtained for analytical inference.

THEOREM 2—Simulation Inference, Estimation, and Power Under Conditions C.1–C.4: *Suppose Conditions* C.1–C.4 *hold. Consider the interval estimator given in Definition* 1 *with the critical value function specified in Definition* 3. *Then, for a given* $p \in [1/2, 1)$, *the following statements hold*:

(i) *The interval estimator has asymptotic level* p:

$$\mathrm{P}_n\big\{\theta_{n0} \le \hat{\theta}_{n0}(p)\big\} \ge p - o(1).$$

(ii) *The estimation risk is bounded by*, *wp* $\to 1$ *under* P_n,

$$\big|\hat{\theta}_{n0}(p) - \theta_{n0}\big| \le 4\bar{\sigma}_n\left(a_n(\overline{V}_n) + \frac{O_{\mathrm{P}_n}(1)}{a_n(\overline{V}_n)}\right) \lesssim_{\mathrm{P}_n} \bar{\sigma}_n\bar{a}_n.$$

(iii) *Any alternative* $\theta_{na} > \theta_{n0}$ *such that*

$$\theta_{na} \ge \theta_{n0} + 4\bar{\sigma}_n\left(a_n(\overline{V}_n) + \frac{\mu_n}{a_n(\overline{V}_n)}\right), \quad \mu_n \to_{\mathrm{P}_n} \infty,$$

is rejected with probability converging to 1 *under* P_n.

3.5. *Properties of the Set Estimator* $\hat{V}_n$

In this section we establish some containment properties for the estimator $\hat{V}_n$. Moreover, these containment properties imply a useful rate result under the following condition:

CONDITION V—Degree of Identifiability for V_0: There exist constants $\rho_n > 0$ and $c_n > 0$, possibly dependent on n, and a positive constant δ, independent of n, such that

$$\theta_n(v) - \theta_{n0} \geq \bigl(c_n d(v, V_0)\bigr)^{\rho_n} \wedge \delta, \quad \forall v \in \mathcal{V}. \tag{3.12}$$

We say $(c_n, 1/\rho_n)$ characterize the degree of identifiability of V_0, as these parameters determine the rate at which V_0 can be consistently estimated. Note that if $V_0 = \mathcal{V}$, then this condition holds with $c_n = \infty$ and $\rho_n = 1$, where we adopt the convention that $0 \cdot \infty = 0$.

We have the following result, whose first part we use in the proof of Theorems 1 and 2 above, and whose second part we use below in the proof of Theorem 3.

LEMMA 2—Estimation of V_n and V_0: *Suppose Conditions C.1–C.4 hold.*

(i) *Containment*: *Then wp* $\to 1$, *for either analytical or simulation methods*,

$$V_n \subseteq \hat{V}_n \subseteq \overline{V}_n$$

for V_n *defined in* (3.3) *with* $\gamma_n' = \gamma_n - o(1)$ *and* $\overline{V}_n$ *defined in* (3.9).

(ii) *Rate*: *If also Condition* V *holds and* $\bar{\kappa}_n \bar{\sigma}_n \to 0$, *then wp* $\to 1$,

$$\begin{aligned} d_H(\hat{V}_n, V_0) &\leq d_H(\hat{V}_n, V_n) + d_H(V_n, V_0) \\ &\leq d_H(\overline{V}_n, V_n) + d_H(V_n, V_0) \leq r_n := 2(\bar{\kappa}_n \bar{\sigma}_n)^{1/\rho_n}/c_n. \end{aligned}$$

3.6. *Automatic Sharpness of Simulation Construction*

When the penultimate process Z_n^* does not lose equicontinuity too fast and when V_0 is sufficiently well identified, our simulation-based inference procedure becomes sharp in the sense of not only achieving the right level, but in fact automatically achieving the right size. In such cases, we typically have some small improvements in the rates of convergence of the estimators. The most important case covered is that where V_0 is singleton (or a finite collection of points) and θ_n is locally quadratic, that is, $\rho_n \geq 2$ and $c_n \geq c > 0$ for all n. These sharp situations occur when the inferential process concentrates on V_0 and not just on the neighborhood V_n, in the sense described below. For this to happen, we impose the following condition.

CONDITION S—Equicontinuity Radii Are not Smaller Than r_n: When Condition V holds, the scaled penultimate process $\bar{a}_n Z_n^*$ has an equicontinuity radius φ_n that is no smaller than $r_n := 2(\bar{\kappa}_n \bar{\sigma}_n)^{1/\rho_n}/c_n$, namely

$$\sup_{\|v - v'\| \leq \varphi_n} \bar{a}_n \bigl|Z_n^*(v) - Z_n^*(v')\bigr| = o_{\mathrm{P}_n}(1), \quad r_n \leq \varphi_n.$$

When Z_n^* is Donsker, that is, asymptotically equicontinuous, this condition holds automatically, since, in this case, $\bar{a}_n \propto 1$ and for any $o(1)$ term, equicontinuity radii obey $\varphi_n = o(1)$, so that consistency $r_n = o(1)$ is sufficient. When Z_n^* is not Donsker, its finite-sample equicontinuity properties decay as $n \to \infty$, with radii φ_n characterizing the decay. However, as long as φ_n is not smaller than r_n, we have just enough finite-sample equicontinuity left to achieve the following result.

LEMMA 3—Inference Sometimes Concentrates on V_0: *Suppose Conditions* C.1–C.4, S, *and* V *hold*. *Then*,

$$\mathrm{P}_n\left(\sup_{v\in\mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} \leq x\right) = \mathrm{P}_n\left(\sup_{v\in V_0} Z_n^*(v) \leq x\right) + o(1).$$

Under the stated conditions, our inference and estimation procedures *automatically* become sharp in terms of size and rates.

THEOREM 3—Sharpness of Simulation Inference: *Suppose Conditions* C.1–C.4, S, *and* V *hold*. *Consider the interval estimator given in Definition* 1 *with the critical value function specified in Definition* 3. *Then*, *for a given* $p \in [1/2, 1)$, *the following statements hold*:

(i) *The interval estimator has asymptotic size* p:

$$\mathrm{P}_n\big\{\theta_{n0} \leq \hat{\theta}_{n0}(p)\big\} = p + o(1).$$

(ii) *Its estimation risk is bounded by*, $wp \to 1$ *under* P_n,

$$\big|\hat{\theta}_{n0}(p) - \theta_{n0}\big| \leq 4\bar{\sigma}_n\left(a_n(V_0) + \frac{O_{\mathrm{P}_n}(1)}{a_n(V_0)}\right) \lesssim_{\mathrm{P}_n} \bar{\sigma}_n a_n(V_0).$$

(iii) *Any alternative* $\theta_{na} > \theta_{n0}$ *such that*

$$\theta_{na} \geq \theta_{n0} + 4\bar{\sigma}_n\left(a_n(V_0) + \frac{\mu_n}{a_n(V_0)}\right), \quad \mu_n \to_{\mathrm{P}_n} \infty,$$

is rejected with probability converging to 1 *under* P_n.

4. INFERENCE ON INTERSECTION BOUNDS IN LEADING CASES

4.1. *Parametric Estimation of Bounding Function*

We now show that the above conditions apply to various parametric estimation methods for $v \mapsto \theta_n(v)$. This is an important practical, and indeed tractable, case. The required conditions cover standard parametric estimators of bounding functions such as least squares, quantile regression, and other estimators.

CONDITION P—Finite-Dimensional Bounding Function: We have that (i) $\theta_n(v) := \theta_n(v, \beta_n)$, where $\mathcal{V} \times \mathcal{B} \mapsto \theta_n(v, \beta)$ is a known function parameterized by finite-dimensional vector $\beta \in \mathcal{B}$, where $\mathcal{V}$ is a compact subset of $\mathbb{R}^d$

and $\mathcal{B}$ is a subset of $\mathbb{R}^k$, where the sets do not depend on n. (ii) The function $(v,\beta)\mapsto p_n(v,\beta):=\partial\theta_n(v,\beta)/\partial\beta$ is uniformly Lipschitz with Lipschitz coefficient $L_n\le L$, where L is a finite constant that does not depend on n. (iii) An estimator $\hat\beta_n$ is available such that

$$\Omega_n^{-1/2}\sqrt{n}(\hat\beta_n-\beta_n)=\mathcal{N}_k+o_{\mathrm{P}_n}(1),\quad \mathcal{N}_k=_d N(0,I_k),$$

that is, $\mathcal{N}_k$ is a random k-vector with the multivariate standard normal distribution. (iv) $\|p_n(v,\beta_n)\|$ is bounded away from zero, uniformly in v and n. The eigenvalues of Ω_n are bounded from above and away from zero, uniformly in n. (v) There is also a consistent estimator $\hat\Omega_n$ such that $\|\hat\Omega_n-\Omega_n\|=O_{\mathrm{P}_n}(n^{-b})$ for some constant $b>0$, independent of n.

EXAMPLE 1—A Saturated Model: As a simple, but relevant example we consider the following model. Suppose that v takes on a finite set of values, denoted $1,\dots,k$, so that $\theta_n(v,\beta)=\sum_{j=1}^k\beta_j 1(v=j)$. Suppose first that $\mathrm{P}_n=\mathrm{P}$ is fixed, so that $\beta_n=\beta_0$, a fixed value. Condition P(ii) and the boundedness requirement of P(iv) follow from $\partial\theta_n(v,\beta)/\partial\beta_j=1(v=j)$ for each $j=1,\dots,k$. Condition P(v) applies to many estimators. Then if the estimator $\hat\beta$ satisfies $\Omega^{-1/2}\sqrt{n}(\hat\beta-\beta_0)\to_d N(0,I_k)$, where Ω is positive definite, the strong approximation in Condition P(iii) follows from Skorohod's theorem and Lemma 9.[17] Suppose next that P_n and the true value $\beta_n=(\beta_{n1},\dots,\beta_{nk})'$ change with n. Then if

$$\Omega_n^{-1/2}\sqrt{n}(\hat\beta_n-\beta_n)=\frac{1}{\sqrt{n}}\sum_{i=1}^n u_{i,n}+o_{\mathrm{P}_n}(1),$$

with $\{u_{i,n}\}$ i.i.d. vectors with mean zero and variance matrix I_k for each n, and $E\|u_{i,n}\|^{2+\delta}$ bounded uniformly in n for some $\delta>0$, then $\Omega_n^{-1/2}\sqrt{n}(\hat\beta_n-\beta_n)\to_d N(0,I_k)$, and again Condition P(iii) follows from Skorohod's theorem and Lemma 9.

LEMMA 4—Conditions P and V imply Conditions C.1–C.4 and S: *Condition* P *implies Conditions* C.1–C.4, *where*, *for* $p_n(v,\beta):=\frac{\partial\theta_n(v,\beta)}{\partial\beta}$,

$$Z_n(v)=\frac{\theta_n(v)-\hat\theta_n(v)}{\sigma_n(v)},\quad Z_n^*(v)=\frac{p_n(v,\beta_n)'\Omega_n^{1/2}}{\|p_n(v,\beta_n)'\Omega_n^{1/2}\|}\mathcal{N}_k,$$

$$Z_n^\star(v)=\frac{p_n(v,\hat\beta_n)'\hat\Omega_n^{1/2}}{\|p_n(v,\hat\beta_n)'\hat\Omega_n^{1/2}\|}\mathcal{N}_k,\quad \sigma_n(v)=\left\|n^{-1/2}p_n(v,\beta_n)'\Omega_n^{1/2}\right\|,$$

[17]See Theorem 1.10.3 of van der Vaart and Wellner (1996, p. 58) and the subsequent historical discussion attributing the earliest such results to Skorohod (1956), later generalized by Wichura and Dudley.

$$s_n(v) = \|n^{-1/2} p_n(v, \hat{\beta}_n)' \hat{\Omega}_n^{1/2}\|, \quad \delta_n = o(1), \quad \bar{a}_n \lesssim 1, \quad \bar{\sigma}_n \lesssim \sqrt{1/n},$$

$$a_n(\mathsf{V}) = \left(2\sqrt{\log\left\{C\left(1 + C' L_n \operatorname{diam}(\mathsf{V})\right)^d\right\}}\right) \vee (1 + \sqrt{d}),$$

for some positive constants C and C', and $P[\mathcal{E} > x] = \exp(-x/2)$. Furthermore, if also Condition V *holds and $c_n^{-1}(\ell\ell_n/\sqrt{n})^{1/\rho_n} = o(1)$, then Condition* S *holds.*

The following theorem is an immediate consequence of Lemma 4 and Theorems 1, 2, and 3.

THEOREM 4—Estimation and Inference With Parametrically Estimated Bounding Functions: *Suppose Condition* P *holds and consider the interval estimator $\hat{\theta}_{n0}(p)$ given in Definition* 1 *with simulation-based critical values specified in Definition* 3 *for the simulation process $Z_n^\star$ specified above.* (a) *Then* (i) $\mathrm{P}_n[\theta_{n0} \leq \hat{\theta}_{n0}(p)] \geq p - o(1)$, (ii) $|\theta_{n0} - \hat{\theta}_{n0}(p)| = O_{\mathrm{P}_n}(\sqrt{1/n})$, *and* (iii) $\mathrm{P}_n(\theta_{n0} + \mu_n\sqrt{1/n} \geq \hat{\theta}_{n0}(p)) \to 1$ *for any $\mu_n \to_{\mathrm{P}_n} \infty$.* (b) *If Condition* V *holds with $c_n \geq c > 0$ and $\rho_n \leq \rho < \infty$, then* $\mathrm{P}_n[\theta_{n0} \leq \hat{\theta}_{n0}(p)] = p + o(1)$.

We next provide two examples that generalize the simple, but well used, saturated example of Example 1 to more substantive cases. Aside from being practically relevant due to the common use of parametric restriction in applications, these examples offer a natural means of transition to the next section, which deals with series estimation and which can be viewed as parametric estimation with parameters of increasing dimension and vanishing approximation errors.

EXAMPLE 2—Linear Bounding Function: Suppose that $\theta_n(v, \beta_n) = p_n(v)'\beta_n$, where $p_n(v)'\beta : \mathcal{V} \times \mathcal{B} \mapsto \mathbb{R}$. Suppose that (a) $v \mapsto p_n(v)$ is Lipschitz with Lipschitz coefficient $L_n \leq L$ for all n, with the first component equal to 1, (b) there is an estimator available that is asymptotically linear,

$$\Omega_n^{-1/2}\sqrt{n}(\hat{\beta}_n - \beta_n) = \frac{1}{\sqrt{n}}\sum_{i=1}^n u_{i,n} + o_{\mathrm{P}_n}(1),$$

with $\{u_{i,n}\}$ i.i.d. vectors with mean zero and variance matrix I_k for each n and $E\|u_{i,n}\|^{2+\delta}$ bounded uniformly in n for some $\delta > 0$, and (c) Ω_n has eigenvalues bounded away from zero and from above. These conditions imply Condition P(i)–(iv). Indeed, Conditions P(i), (ii), and (iv) hold immediately, while Condition P(iii) follows from the Lindeberg–Feller central limit theorem (CLT), which implies that under P_n,

$$\Omega_n^{-1/2}\sqrt{n}(\hat{\beta}_n - \beta_n) \to_d N(0, I_k),$$

and the strong approximation follows by the Skorohod representation and Lemma 9 by suitably enriching the probability space if needed. Note that if

$\theta_n(v, \beta_n)$ is the conditional expectation of Y_i given $V_i = v$, then $\hat{\beta}_n$ can be obtained by the mean regression of Y_i on $p_n(V_i)$, $i = 1, \dots, n$; if $\theta_n(v, \beta_n)$ is the conditional u-quantile of Y_i given $V_i = v$, then $\hat{\beta}_n$ can be obtained by the u-quantile regression of Y_i on $p_n(V_i)$, $i = 1, \dots, n$. Regularity conditions that imply the conditions stated above can be found in, for example, White (1984) and Koenker (2005). Finally estimators of Ω_n depend on the estimator of β_n: for mean regression, the standard estimator is the Eicker–Huber–White estimator; for quantile regression, the standard estimator is Powell's (1984) estimator. For brevity, we do not restate sufficient conditions for Condition P(v), but these are readily available for common estimators.

EXAMPLE 3—Conditional Moment Inequalities: This is a generalization of the previous example where now the bounding function is the minimum of J conditional mean functions. Referring to the conditional moment inequality setting specified in Section 2, suppose we have an i.i.d. sample of (X_i, Z_i), $i = 1, \dots, n$, with $\text{support}(Z_i) = \mathcal{Z} \subseteq [0, 1]^d$. Let $v = (z, j)$, where j denotes the enumeration index for the conditional moment inequality, $j \in \{1, \dots, J\}$, and suppose $\mathcal{V} \subseteq \mathcal{Z} \times \{1, \dots, J\}$. The parameters J and d do not depend on n. Hence

$$\theta_{n0} = \min_{v \in \mathcal{V}} \theta_n(v) = \min_{(z,j) \in \mathcal{V}} \theta_n(z, j).$$

Suppose that $\theta_n(v) = E_{\mathrm{P}_n}[m(X, \mu, j)|Z = z] = b(z)'\chi_n(j)$ for $b : \mathcal{Z} \mapsto \mathbb{R}^m$, denoting some transformation of z, with m independent of n and where $\chi_n(j)$ are the population regression coefficients in the regression of $Y(j) := m(X, \mu, j)$ on $b(Z)$, $j = 1, \dots, J$, respectively, under P_n. Suppose that the first $J_0/2$ pairs correspond to moment inequalities generated from moment equalities so that $\theta_n(j) = -\theta_n(j-1)$, $j = 2, 4, \dots, J_0$, and so these functions are replicas of each other up to sign; also note that $\chi_n(j) = -\chi_n(j-1)$, $j = 2, 4, \dots, J_0$. Then we can rewrite

$$\theta_n(v) = E_{\mathrm{P}_n}\big[m(X, \mu, j)|Z = z\big] = b(z)'\chi_n(j) := p_n(v)'\beta_n,$$

$$\beta_n = \big(\chi_n(j)', j \in \mathcal{J}\big)', \quad \mathcal{J} := \{2, 4, \dots, J_0, J_0 + 1, J_0 + 2, \dots, J\}',$$

where β_n is a K-vector of regression coefficients, and $p_n(v)$ is a K-vector such that $p_n(z, j) = [0'_m, \dots, 0'_m, (-1)^{j+1} b'_m(z), 0'_m, \dots, 0'_m]'$ with $b'_m(z)$ appearing in the $\lceil j/2 \rceil$th block for $1 \le j \le J_0$ and $p_n(z, j) = [0'_m, \dots, 0'_m, b'_m(z), 0'_m, \dots, 0'_m]'$ with $b(z)$ appearing in the jth block for $J_0 + 1 \le j \le J$, where 0_m is an m-dimensional vector of zeroes.[18]

We impose the following conditions:

(a) $b(z)$ includes constant 1;

[18]Note the absence of $\chi_n(j)$ for odd j up to J_0 in the definition of the coefficient vector β_n. This is required to enable nonsingularity of $E_{\mathrm{P}_n}[\epsilon_i \epsilon_i' | Z_i = z]$. Imposing nonsingularity simplifies the proofs, and is not needed for practical implementation.

(b) $z \mapsto b(z)$ has Lipschitz coefficient bounded above by L;

(c) for $Y_i = (Y_i(j), j \in \mathcal{J})'$ and for $\epsilon_i := Y_i - E_{\mathrm{P}_n}[Y_i|Z_i]$, the eigenvalues of $E_{\mathrm{P}_n}[\epsilon_i \epsilon_i' | Z_i = z]$ are bounded away from zero and from above, uniformly in $z \in \mathcal{Z}$ and n;

(d) $Q = E_{\mathrm{P}_n}[b(Z_i)b(Z_i)']$ has eigenvalues bounded away from zero and from above, uniformly in n;

(e) $E_{\mathrm{P}_n}\|b(Z_i)\|^4$ and $E_{\mathrm{P}_n}\|\epsilon_i\|^4$ are bounded from above uniformly in n.

Then it follows from, for example, White (1984) that for $\hat{\chi}_n(j)$ denoting the ordinary least squares estimator obtained by regressing $Y_i(j), i = 1, \ldots, n$, on $b(Z_i), i = 1, \ldots, n$,

$$\sqrt{n}\big(\hat{\chi}_n(j) - \chi_n(j)\big) = Q^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b(Z_i)\epsilon_i(j) + o_{\mathrm{P}_n}(1), \quad j \in \mathcal{J},$$

so that

$$\sqrt{n}(\hat{\beta}_n - \beta_n) = (I_{|\mathcal{J}|} \otimes Q)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \underbrace{\big(I_{|\mathcal{J}|} \otimes b(Z_i)\big)\epsilon_i}_{u_i} + o_{\mathrm{P}_n}(1).$$

By conditions (c) and (d), $E_{\mathrm{P}_n}[u_i u_i']$ and Q have eigenvalues bounded away from zero and from above, so the same is true of $\Omega_n = (I_{|\mathcal{J}|} \otimes Q)^{-1} E_{\mathrm{P}_n}[u_i u_i'] \times (I_{|\mathcal{J}|} \otimes Q)^{-1}$. These conditions verify Condition P(i), (ii), and (iv). Application of the Lindeberg–Feller CLT, Skorohod's theorem, and Lemma 9 verifies Condition P(iii). By the argument given in Chapter VI of White (1984), Condition P(v) holds for the standard analog estimator for Ω_n,

$$\hat{\Omega}_n = (I_{|\mathcal{J}|} \otimes \hat{Q})^{-1} \mathbb{E}_n\big[\hat{u}_i \hat{u}_i'\big](I_{|\mathcal{J}|} \otimes \hat{Q})^{-1},$$

where $\hat{Q} = \mathbb{E}_n[b(Z_i)b(Z_i)']$ and $\hat{u}_i = (I_{|\mathcal{J}|} \otimes b(Z_i))\hat{\epsilon}_i$, with $\hat{\epsilon}_i(j) = Y_i(j) - b(Z_i)'\hat{\chi}_n(j)$ and $\hat{\epsilon}_i = (\hat{\epsilon}_i(j), j \in \mathcal{J})'$.

4.2. *Nonparametric Estimation of $\theta_n(v)$ via Series*

Series estimation is effectively like parametric estimation, but the dimension of the estimated parameter tends to infinity and bias arises due to approximation based on a finite number of basis functions. If we select the number of terms in the series expansion so that the estimation error is of larger magnitude than the approximation error, that is, if we undersmooth, then the analysis closely mimics the parametric case.

CONDITION NS: The function $v \mapsto \theta_n(v)$ is continuous in v. The series estimator $\hat{\theta}_n(v)$ has the form $\hat{\theta}(v) = p_n(v)'\hat{\beta}_n$, where $p_n(v) := (p_{n,1}(v), \ldots, p_{n,K_n}(v))'$ is a collection of K_n continuous series functions mapping $\mathcal{V} \subset \mathcal{K} \subset \mathbb{R}^d$ to $\mathbb{R}^{K_n}$, and $\hat{\beta}_n$ is a K_n-vector of coefficient estimates and $\mathcal{K}$ is a fixed compact set. Furthermore, the following statements hold:

(i)(a) The estimator satisfies the linearization and strong approximation condition

$$\frac{\hat{\theta}_n(v) - \theta_n(v)}{\| p_n(v)'\Omega_n^{1/2}\|/\sqrt{n}} = \frac{p_n(v)'\Omega_n^{1/2}}{\| p_n(v)'\Omega_n^{1/2}\|}\mathcal{N}_n + R_n(v),$$

where

$$\mathcal{N}_n =_d N(0, I_{K_n}), \quad \sup_{v\in\mathcal{V}}\big|R_n(v)\big| = o_{\mathrm{P}_n}(1/\log n).$$

(b) The matrices Ω_n are positive definite, with eigenvalues bounded from above and away from zero, uniformly in n. Moreover, there are sequences of constants ζ_n and ζ_n' such that $1 \le \zeta_n' \lesssim \| p_n(v)\| \le \zeta_n$ uniformly for all $v \in \mathcal{V}$ and $\sqrt{\zeta_n^2 \log n/n} \to 0$, and $\| p_n(v) - p_n(v')\|/\zeta_n' \le L_n\|v - v'\|$ for all $v, v' \in \mathcal{V}$, where $\log L_n \lesssim \log n$ uniformly in n.

(ii) There exists $\hat{\Omega}_n$ such that $\|\hat{\Omega}_n - \Omega_n\| = O_{\mathrm{P}_n}(n^{-b})$, where $b > 0$ is a constant.

Condition NS is not primitive, but reflects the functionwise large sample normality of series estimators. It requires that the studentized nonparametric process is approximated by a sequence of Gaussian processes, which take a very simple intuitive form, rather than by a fixed single Gaussian process. Indeed, the latter would be impossible in nonparametric settings, since the sequence of Gaussian processes is not asymptotically tight. Note also that the condition implicitly requires that some undersmoothing takes place so that the approximation error is negligible relative to the sampling error. We provide primitive conditions that imply Condition NS(i) in three examples presented below. In particular, we show that the asymptotic linearization for $\hat{\beta}_n - \beta_n$, which is available from the literature on series regression (e.g., from Andrews (1991) and Newey (1997)), and the use of Yurinskii's (1977) coupling imply Condition NS(i). This result could be of independent interest, although we only provide sufficient conditions for the strong approximation to hold.

Note that under Condition NS, the uniform rate of convergence of $\hat{\theta}_n(v)$ to $\theta_n(v)$ is given by $\sqrt{\zeta_n^2/n}\sqrt{\log n} \to 0$, where $\zeta_n \propto \sqrt{K_n}$ for standard series terms such as B-splines or trigonometric series.

LEMMA 5—Condition NS Implies Conditions C.1–C.4: *Condition NS implies Conditions C.1–C.4 with*

$$Z_n(v) = \frac{\theta_n(v) - \hat{\theta}_n(v)}{\sigma_n(v)}, \quad Z_n^*(v) = \frac{p_n(v)'\Omega_n^{1/2}}{\| p_n(v)'\Omega_n^{1/2}\|}\mathcal{N}_n,$$

$$Z_n^\star(v) = \frac{p_n(v)'\hat{\Omega}_n^{1/2}}{\| p_n(v)'\hat{\Omega}_n^{1/2}\|}\mathcal{N}_n, \quad \sigma_n(v) = \big\|n^{-1/2}p_n(v)'\Omega_n^{1/2}\big\|,$$

$$s_n(v) = \|n^{-1/2} p_n(v)'\hat{\Omega}_n^{1/2}\|, \quad \delta_n = 1/\log n,$$

$$\bar{a}_n \lesssim \sqrt{\log n}, \quad \bar{\sigma}_n \lesssim \sqrt{\zeta_n^2/n},$$

$$a_n(\mathsf{V}) = \left(2\sqrt{\log\{C(1 + C' L_n \operatorname{diam}(\mathsf{V}))^d\}}\right) \vee (1 + \sqrt{d})$$

for some constants C *and* C', *where* $\operatorname{diam}(\mathsf{V})$ *denotes the diameter of the set* V *and* $P[\mathcal{E} > x] = \exp(-x/2)$.

REMARK 2: Lemma 5 verifies the main conditions, Conditions C.1–C.4. These conditions enable construction of simulated or analytical critical values. For the latter, the pth quantile of $\mathcal{E}$ is given by $c(p) = -2\log(1 - p)$, so we can set

$$k_{n,\mathsf{V}}(p) = a_n(\mathsf{V}) - 2\log(1 - p)/a_n(\mathsf{V}), \tag{4.1}$$

where

$$a_n(\mathsf{V}) = \left(2\sqrt{\log\{\ell_n(1 + \ell_n L_n \operatorname{diam}(\mathsf{V}))^d\}}\right) \tag{4.2}$$

is a feasible scaling factor that bounds the scaling factor in the statement of Lemma 5, at least for all large n. Here, all unknown constants have been replaced by slowly growing numbers ℓ_n such that $\ell_n > C \vee C'$ for all large n. Note also that $\mathsf{V} \mapsto k_{n,\mathsf{V}}(p)$ is monotone in V for all sufficiently large n, as required in the analytical construction given in Definition 2. A sharper analytical approach can be based on Hotelling's tube method; for details, refer to Chernozhukov, Lee, and Rosen (2009). That approach is tractable for the case of $d = 1$, but does not immediately extend to $d > 1$. Note that the simulation-based approach is effectively a numeric version of the exact version of the tube formula and is less conservative than using simplified tube formulas.

LEMMA 6—Condition NS Implies Condition S in Some Cases: *Suppose Condition NS holds. Then* (i) *the radius* φ_n *of equicontinuity of* Z_n^* *obeys*

$$\varphi_n \leq o(1) \cdot \left(\frac{1}{L_n\sqrt{\log n}}\right)$$

for any $o(1)$ *term*. (ii) *If Condition V holds and*

$$\left(\sqrt{\frac{\zeta_n^2}{n}\log n}\right)^{1/\rho_n} c_n^{-1} = o\left(\frac{1}{L_n\sqrt{\log n}}\right), \tag{4.3}$$

then Condition S holds. (iii) *If* V_0 *is singleton and* (4.3) *holds*, $\rho_n \leq 2$ *and* $c_n \geq c > 0$ *for all* n, *then* $a_n(V_0) \propto 1$ *and* (4.3) *reduces to*

$$L_n^4 K_n \log^3 n/n \to 0.$$

The following theorem is an immediate consequence of Lemmas 5 and 6 and Theorems 1, 2, and 3.

THEOREM 5—Estimation and Inference With Series-Estimated Bounding Functions: *Suppose Condition* NS *holds and consider the interval estimator* $\hat{\theta}_{n0}(p)$ *given in Definition* 1 *with either analytical critical value* $c(p) = -2\log(1-p)$ *or simulation-based critical values from Definition* 3 *for the simulation process* $Z_n^\star$ *above*. (a) *Then* (i) $\mathrm{P}_n[\theta_{n0} \leq \hat{\theta}_{n0}(p)] \geq p - o(1)$, (ii) $|\theta_{n0} - \hat{\theta}_{n0}(p)| = O_{\mathrm{P}_n}(\sqrt{\log n}\sqrt{\zeta_n^2/n})$, *and* (iii) $\mathrm{P}_n(\theta_{n0} + \mu_n\sqrt{\log n}\sqrt{\zeta_n^2/n} \geq \hat{\theta}_{n0}(p)) \to 1$ *for any* $\mu_n \to_{\mathrm{P}_n} \infty$. (b) *Moreover, for the simulation-based critical values, if Condition* V *and relation* (4.3) *hold, then* (i) $\mathrm{P}_n[\theta_{n0} \leq \hat{\theta}_{n0}(p)] = p - o(1)$, (ii) $|\theta_{n0} - \hat{\theta}_{n0}(p)| = O_{\mathrm{P}_n}(\sqrt{\zeta_n^2/n})$, *and* (iii) $\mathrm{P}_n(\theta_{n0} + \mu_n\sqrt{\zeta_n^2/n} \geq \hat{\theta}_{n0}(p)) \to 1$ *for any* $\mu_n \to_{\mathrm{P}_n} \infty$.

We next present some examples with primitive conditions that imply Condition NS.

EXAMPLE 4—Bounding Function is Conditional Quantile: Suppose that $\theta_n(v) := Q_{Y_i|V_i}[\tau|v]$ is the τth conditional quantile of Y_i given V_i under P_n, assumed to be a continuous function in v. Suppose we estimate $\theta_n(v)$ with a series estimator. There is an i.i.d. sample (Y_i, V_i), $i = 1, \dots, n$, with support$(V_i) \subseteq [0, 1]^d$ for each n, that is defined on a probability space equipped with probability measure P_n. Suppose that the intersection region of interest is $\mathcal{V} \subseteq$ support(V_i). Here the index d does not depend on n, but all other parameters, unless stated otherwise, can depend on n. Then $\theta_n(v) = p_n(v)'\beta_n + A_n(v)$, where $p_n : [0, 1]^d \mapsto \mathbb{R}^{K_n}$ are the series functions, β_n is the quantile regression coefficient in the population, $A_n(v)$ is the approximation error, and K_n is the number of series terms that depend on n. Let C be a positive constant.

We impose the following technical conditions to verify Conditions NS(i) and NS(ii):

Uniformly in n:

(i) p_n are either B-splines of a fixed order or trigonometric series terms or any other terms $p_n = (p_{n1}, \dots, p_{nK_n})'$ such that $\|p_n(v)\| \lesssim \zeta_n = \sqrt{K_n}$ for all $v \in$ support(V_i), $\|p_n(v)\| \gtrsim \zeta_n' \geq 1$ for all $v \in \mathcal{V}$, and $\log \mathrm{lip}(p_n) \lesssim \log K_n$;

(ii) the mapping $v \mapsto \theta_n(v)$ is sufficiently smooth, namely $\sup_{v\in\mathcal{V}} |A_n(v)| \lesssim K_n^{-s}$ for some $s > 0$;

(iii) $\lim_{n\to\infty} (\log n)^c K_n^{-s+1} = 0$ and $\lim_{n\to\infty} (\log n)^c \sqrt{n} K_n^{-s}/\zeta_n' = 0$ for each $c > 0$;

(iv) eigenvalues of $\Lambda_n = E_{\mathrm{P}_n}[p_n(V_i)p_n(V_i)']$ are bounded away from zero and from above;

(v) $f_{Y_i|V_i}(\theta_n(v)|v)$ is bounded uniformly over $v \in \mathcal{V}$ away from zero and from above;

(vi) $\lim_{n\to\infty} K_n^5(\log n)^c/n = 0$ for each $c > 0$;

(vii) the restriction on the bandwidth sequence in Powell's estimator $\hat{Q}_n$ of $Q_n = E_{\mathrm{P}_n}[f_{Y_i|V_i}(\theta_n(V_i)|V_i)p_n(V_i)p_n(V_i)']$ specified in Belloni, Chernozhukov, and Fernandez-Val (2011) holds.

Suppose that we use the standard quantile regression estimator

$$\hat{\beta}_n = \arg\min_{b\in\mathbb{R}^{K_n}} \mathbb{E}_n\big[\rho_\tau\big(Y_i - p_n(V_i)'b\big)\big],$$

so that $\hat{\theta}_n(v) = p_n(v)'\hat{\beta}$ for $\rho_\tau(u) = (\tau - 1(u<0))u$. Then by Belloni, Chernozhukov, and Fernandez-Val (2011), under conditions (i)–(vi), the asymptotically linear representation

$$\sqrt{n}(\hat{\beta}_n - \beta_n) = Q_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n \underbrace{p_n(V_i)\epsilon_i}_{u_i} + o_{\mathrm{P}_n}\left(\frac{1}{\log n}\right),$$

holds for $\epsilon_i = (\tau - 1(w_i \le \tau))$, where w_i, $i = 1,\dots,n$, are i.i.d., uniform, and independent of V_i, $i = 1,\dots,n$. Note that by conditions (iv) and (v), $S_n := E_{\mathrm{P}_n}[u_i u_i'] = \tau(1-\tau)\Lambda_n$ and Q_n have eigenvalues bounded away from zero and from above uniformly in n, and so the same is also true of $\Omega_n = Q_n^{-1}S_nQ_n^{-1}$. Given other restrictions imposed in condition (i), Condition NS(i)(b) is verified. Next using condition (iv) and others, the strong approximation required in Condition NS(i)(a) follows by invoking Theorem 7 and Corollary 1 in Section 5, which is based on Yurinskii's coupling. To verify Condition NS(ii), consider the plug-in estimator $\hat{\Omega}_n = \hat{Q}_n^{-1}\hat{S}_n\hat{Q}_n^{-1}$, where $\hat{Q}_n$ is Powell's estimator for Q_n and $\hat{S}_n = \tau(1-\tau)\cdot\mathbb{E}_n[p_n(V_i)p_n(V_i)]$. Then under condition (vii), it follows from the proof of Theorem 7 in Belloni, Chernozhukov, and Fernandez-Val (2011) that $\|\hat{\Omega}_n - \Omega_n\| = O_{\mathrm{P}_n}(n^{-b})$ for some $b > 0$.

EXAMPLE 5—Bounding Function is Conditional Mean: Now suppose that $\theta_n(v) = E_{\mathrm{P}_n}[Y_i|V_i = v]$, assumed to be a continuous function with respect to $v \in \operatorname{support}(V_i)$, and that the intersection region is $\mathcal{V} \subseteq \operatorname{support}(V_i)$. Suppose we are using the series approach to approximating and estimating $\theta_n(v)$. There is an i.i.d. sample (Y_i, V_i), $i = 1,\dots,n$, with $\operatorname{support}(V_i) \subseteq [0,1]^d$ for each n. Here d does not depend on n, but all other parameters, unless stated otherwise, can depend on n. Then we have $\theta_n(v) = p_n(v)'\beta_n + A_n(v)$ for $p_n : [0,1]^d \mapsto \mathbb{R}^{K_n}$ representing the series functions, β_n is the coefficient of the best least squares approximation to $\theta_n(v)$ in the population, and $A_n(v)$ is the approximation error. The number of series terms K_n depends on n.

We impose the following technical conditions:

Uniformly in n:

(i) p_n are either B-splines of a fixed order or trigonometric series terms or any other terms $p_n = (p_{n1},\dots,p_{nK_n})'$ such that $\|p_n(v)\| \lesssim \zeta_n = \sqrt{K_n}$ for all $v \in \operatorname{support}(V_i)$, $\|p_n(v)\| \gtrsim \zeta_n' \ge 1$ for all $v \in \mathcal{V}$, and $\log\operatorname{lip}(p_n) \lesssim \log K_n$;

(ii) the mapping $v \mapsto \theta_n(v)$ is sufficiently smooth, namely $\sup_{v\in\mathcal{V}} |A_n(v)| \lesssim K_n^{-s}$ for some $s > 0$;
(iii) $\lim_{n\to\infty}(\log n)^c\sqrt{n}K_n^{-s} = 0$ for each $c > 0$[19];
(iv) for $\epsilon_i = Y_i - E_{\mathrm{P}_n}[Y_i|V_i]$, $E_{\mathrm{P}_n}[\epsilon_i^2|V_i = v]$ is bounded away from zero uniformly in $v \in \mathrm{support}(V_i)$;
(v) eigenvalues of $Q_n = E_{\mathrm{P}_n}[p_n(V_i)p_n(V_i)']$ are bounded away from zero and from above;
(vi) $E_{\mathrm{P}_n}[|\epsilon_i|^4|V_i = v]$ is bounded from above uniformly in $v \in \mathrm{support}(V_i)$;
(vii) $\lim_{n\to\infty}(\log n)^c K_n^5/n = 0$ for each $c > 0$.
We use the standard least squares estimator

$$\hat{\beta}_n = \mathbb{E}_n\big[p_n(V_i)p_n(V_i)'\big]^{-1}\mathbb{E}_n\big[p_n(V_i)Y_i\big],$$

so that $\hat{\theta}_n(v) = p_n(v)'\beta_n$. Then by Newey (1997), under conditions (i)–(vii), we have the asymptotically linear representation

$$\sqrt{n}(\hat{\beta}_n - \beta_n) = Q_n^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\underbrace{p_n(V_i)\epsilon_i}_{u_i} + o_{\mathrm{P}_n}(1/\log n).$$

For details, see Supplemental Material Appendix I. Note that $E_{\mathrm{P}_n}(u_i u_i')$ and Q_n have eigenvalues bounded away from zero and from above uniformly in n, and so the same is also true of $\Omega_n = Q_n^{-1}E_{\mathrm{P}_n}(u_i u_i')Q_n^{-1}$. Thus, under condition (i), Condition NS(i)(a) is verified. The strong approximation Condition NS(i)(a) now follows by invoking Theorem 7 in Section 5. Finally, Newey (1997) verified that Condition NS(ii) holds for the standard analog estimator $\hat{\Omega}_n = \hat{Q}_n^{-1}\mathbb{E}_n(\hat{u}_i\hat{u}_i')\hat{Q}_n^{-1}$ for $\hat{u}_i = p_n(V_i)(Y_i - \hat{\theta}_n(V_i))$ and $\hat{Q}_n = \mathbb{E}_n[p_n(V_i)p_n(V_i)']$ under conditions that are implied by those above.

Finally, note that if we had $\epsilon_i \sim N(0, \sigma^2(V_i))$, conditional on V_i, we could establish Condition NS(i) with a much weaker growth restriction than (vii). Thus, while our use of Yurinskii's coupling provides concrete sufficient conditions for strong approximation, the functionwise large sample normality is likely to hold under weaker conditions in many situations.

EXAMPLE 6—Bounding Function From Conditional Moment Inequalities: Consider now Example C in Section 2, where now the bounding function is the minimum of J conditional mean functions. Suppose we have an i.i.d. sample of (X_i, Z_i), $i = 1, \dots, n$, with $\mathrm{support}(Z_i) = \mathcal{Z} \subseteq [0, 1]^d$, defined on a probability space equipped with probability measure P_n. Let $v = (z, j)$, where j denotes

[19]This condition, which is based on Newey (1997), can be relaxed to $(\log n)^c K_n^{-s+1} \to 0$ and $(\log n)^c\sqrt{n}K_n^{-s}/\zeta_n' \to 0$ using the recent results of Belloni, Chernozhukov, and Kato (2010) for least squares series estimators.

the enumeration index for the conditional moment inequality, $j \in \{1, \dots, J\}$, and $\mathcal{V} \subseteq \mathcal{Z} \times \{1, \dots, J\}$. The parameters J and d do not depend on n. Hence

$$\theta_{n0} = \min_{v \in \mathcal{V}} \theta_n(v)$$

for $\theta_n(v) = E_{\mathrm{P}_n}[m(X_i, \mu, j)|Z_i = z]$, assumed to be a continuous function with respect to $z \in \mathcal{Z}$. Suppose we use the series approach to approximate and estimate $\theta_n(z, j)$ for each j. Then $E_{\mathrm{P}_n}[m(X, \mu, j)|z] = b_n(z)'\chi_n(j) + A_n(z, j)$ for $b_n : [0, 1]^d \mapsto \mathbb{R}^{m_n}$ denoting an m_n-vector of series functions, $\chi_n(j)$ is the coefficient of the best least squares approximation to $E_{\mathrm{P}_n}[m(x, \mu, j)|z]$ in the population, and $A_n(z, j)$ is the approximation error. Let $\mathcal{J}$ be a subset of $\{1, \dots, J\}$ defined as in the parametric Example 3 (to handle inequalities associated with equalities).

We impose the following conditions:

Uniformly in n:

(i) $b_n(z)$ are either B-splines of a fixed order or trigonometric series terms or any other terms $b_n(z) = (b_{n1}(z), \dots, b_{nm_n}(z))'$ such that $\|b_n(z)\| \lesssim \zeta_n = \sqrt{m_n}$ for all $z \in \mathcal{Z}$, $\|b_n(z)\| \gtrsim \zeta_n' \geq 1$ for all $z \in \mathcal{Z}$, and $\log \operatorname{lip}(b_n(z)) \lesssim \log m_n$;

(ii) the mapping $z \mapsto \theta_n(z, j)$ is sufficiently smooth, namely $\sup_{z \in \mathcal{Z}} |A_n(z, j)| \lesssim m_n^{-s}$ for some $s > 0$, for all $j \in \mathcal{J}$;

(iii) $\lim_{n\to\infty} (\log n)^c \sqrt{n} m_n^{-s} = 0$ for each $c > 0$[20];

(iv) for $Y(j) := m(X, \mu, j)$, $Y_i := (Y_i(j), j \in \mathcal{J})'$, and $\boldsymbol{\epsilon}_i := Y_i - E_{\mathrm{P}_n}[Y_i|Z_i]$, the eigenvalues of $E_{\mathrm{P}_n}[\boldsymbol{\epsilon}_i \boldsymbol{\epsilon}_i'|Z_i = z]$ are bounded away from zero, uniformly in $z \in \mathcal{Z}$;

(v) eigenvalues of $Q_n = E_{\mathrm{P}_n}[b_n(Z_i)b_n(Z_i)']$ are bounded away from zero and from above;

(vi) $E_{\mathrm{P}_n}[\|\boldsymbol{\epsilon}_i\|^4|Z_i = z]$ is bounded above, uniformly in $z \in \mathcal{Z}$;

(vii) $\lim_{n\to\infty} m_n^5 (\log n)^c / n = 0$ for each $c > 0$.

The above construction implies $\theta_n(v) = b_n(z)'\chi_n(j) + A_n(z, j) =: p_n(v)'\beta_n + A_n(v)$ for $\beta_n = (\chi_n'(j), j \in \mathcal{J})'$, where $p_n(v)$ and β_n are vectors of dimension $K_n := m_n \times |\mathcal{J}|$, defined as in parametric Example 3. Consider the standard least squares estimator $\hat{\beta}_n = (\hat{\chi}_n'(j), j \in \mathcal{J})'$ consisting of $|\mathcal{J}|$ least square estimators, where $\hat{\chi}_n(j) = \mathbb{E}_n[b_n(Z_i)b_n(Z_i)']^{-1}\mathbb{E}_n[b_n(Z_i)Y_i(j)]$. Then it follows from Newey (1997) that for $Q_n = E_{\mathrm{P}_n}[b_n(Z_i)b_n(Z_i)']^{-1}$,

$$\begin{aligned}\sqrt{n}\big(\hat{\chi}_n(j) - \chi_n(j)\big) = {} & \frac{1}{\sqrt{n}} \sum_{i=1}^n Q_n^{-1} b_n(Z_i)\boldsymbol{\epsilon}_i(j) \\ & + o_{\mathrm{P}_n}(1/\log n), \quad j \in \mathcal{J},\end{aligned}$$

[20]See the previous footnote on a possible relaxation of this condition.

so that

$$\sqrt{n}(\hat{\beta}_n - \beta_n) = (I_{|\mathcal{J}|} \otimes Q_n)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \underbrace{\big(I_{|\mathcal{J}|} \otimes b_n(Z_i)\big)\epsilon_i}_{u_i} + o_{\mathrm{P}_n}(1/\log n).$$

By conditions (iv), (v), and (vi), $E_{\mathrm{P}_n}[u_i u_i']$ and Q_n have eigenvalues bounded away from zero and from above, so the same is true of $\Omega_n = (I_{|\mathcal{J}|} \otimes Q_n)^{-1} E_{\mathrm{P}_n}[u_i u_i'](I_{|\mathcal{J}|} \otimes Q_n)^{-1}$. This and condition (i) imply that Condition NS(i)(b) holds. Application of Theorem 7, based on Yurinskii's coupling, verifies Condition NS(i)(a). Finally, Condition NS(ii) holds for the standard plug-in estimator for Ω_n, by the same argument as given in the proof of Theorem 2 of Newey (1997).

4.3. *Nonparametric Estimation of $\theta_n(v)$ via Kernel Methods*

In this section, we provide conditions under which kernel-type estimators satisfy Conditions C.1–C.4. These conditions cover both standard kernel estimators as well as local polynomial estimators.

CONDITION NK: Let $v = (z, j)$ and $\mathcal{V} \subseteq \mathcal{Z} \times \{1, \dots, J\}$, where $\mathcal{Z}$ is a compact convex set that does not depend on n. The estimator $v \mapsto \hat{\theta}_n(v)$ and the function $v \mapsto \theta_n(v)$ are continuous in v. In what follows, let e_j denote the J-vector with jth element 1 and all other elements 0. Suppose that (U, Z) is a $(J + d)$-dimensional random vector, where U is a generalized residual such that $E[U|Z] = 0$ a.s. and Z is a covariate, the density f_n of Z is continuous and bounded away from zero and from above on $\mathcal{Z}$, uniformly in n, and the support of U is bounded uniformly in n. $\mathbf{K}$ is a twice continuously differentiable, possibly higher order, product kernel function with support on $[-1, 1]^d$, $\int \mathbf{K}(u)\,du = 1$, and h_n is a sequence of bandwidths such that $h_n \to 0$ and $nh_n^d \to \infty$ at a polynomial rate in n.

(i) We have that uniformly in $v \in \mathcal{V}$,

$$\big(nh_n^d\big)^{1/2}\big(\hat{\theta}_n(v) - \theta_n(v)\big) = \mathbb{B}_n(g_v) + o_{\mathbf{P}_n}(\delta_n),$$
$$g_v(U, Z) := \frac{e_j' U}{(h_n^d)^{1/2} f_n(z)} \mathbf{K}\left(\frac{z - Z}{h_n}\right),$$

where $\mathbb{B}_n$ is a P_n-Brownian bridge such that $v \mapsto \mathbb{B}_n(g_v)$ has continuous sample paths over $\mathcal{V}$. Moreover, the latter process can be approximated via the Gaussian multiplier method, namely there exist sequences $o(\delta_n)$ and $o(1/\ell_n)$ such that

$$\mathrm{P}_n\Big(\sup_{v \in \mathcal{V}} \big|\mathbb{G}_n^o(g_v) - \bar{\mathbb{B}}_n(g_v)\big| > o(\delta_n) \big| \mathcal{D}_n\Big) = o_{\mathrm{P}_n}(1/\ell_n)$$

for some independent (from data) copy $v \mapsto \bar{\mathbb{B}}_n(g_v)$ of the process $v \mapsto \mathbb{B}_n(g_v)$. Here, $\mathbb{G}_n^o(g_v) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i g_v(U_i, Z_i)$, where η_i are i.i.d. $N(0,1)$, independent of the data $\mathcal{D}_n$ and of $\{(U_i, Z_i)\}_{i=1}^n$, which are i.i.d. copies of (U, Z). Covariates $\{Z_i\}_{i=1}^n$ are part of the data, and $\{U_i\}_{i=1}^n$ are a measurable transformation of data.

(ii) There exists an estimator $z \mapsto \hat{f}_n(z)$, having continuous sample paths, such that $\sup_{z\in\mathcal{Z}} |\hat{f}_n(z) - f_n(z)| = O_{\mathrm{P}_n}(n^{-b})$, and there are estimators $\hat{U}_i$ of generalized residuals such that $\max_{1\le i\le n} \|\hat{U}_i - U_i\| = O_{\mathrm{P}_n}(n^{-\tilde{b}})$ for some constants $b > 0$ and $\tilde{b} > 0$.

Condition NK(i) is a high-level condition that captures the large sample Gaussianity of the entire estimated function where estimation is done via a kernel or local method. Under some mild regularity conditions, specifically those stated in Appendix G, Condition NK(i) follows from the Rio–Massart coupling (Rio (1994) and Massart (1989)) and from the Bahadur expansion holding uniformly in $v \in \mathcal{V}$:

$$\left(nh_n^d\right)^{1/2}\left(\hat{\theta}_n(v) - \theta_n(v)\right) = \mathbb{G}_n(g_v) + o_{\mathbf{P}_n}(\delta_n).$$

Uniform Bahadur expansions have been established for a variety of local estimators; see, for example, Masry (1996) and Kong, Linton, and Xia (2010), including higher order kernel and local polynomial estimators. It is possible to use more primitive sufficient conditions stated in Appendix G based on the Rio–Massart coupling, but these conditions are merely sufficient and other primitive conditions may also be adequate. Our general argument, however, relies only on validity of Condition NK(i).

For simulation purposes, we define

$$\mathbb{G}_n^o(\hat{g}_v) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_i \hat{g}_v(U_i, Z_i),$$

$$\eta_i \text{ i.i.d. } N(0,1), \text{ independent of the data } \mathcal{D}_n,$$

$$\hat{g}_v(U_i, Z_i) = \frac{e_j' \hat{U}_i}{(h_n^d)^{1/2} \hat{f}_n(z)} \mathbf{K}\left(\frac{z - Z_i}{h_n}\right).$$

LEMMA 7—Condition NK Implies Conditions C.1–C.4: *Condition NK Implies Conditions C.1–C.4 with $v = (z, j) \in \mathcal{V} \subseteq \mathcal{Z} \times \{1, \dots, J\}$,*

$$Z_n(v) = \frac{\theta_n(v) - \hat{\theta}_n(v)}{\sigma_n(v)}, \quad Z_n^*(v) = \frac{\mathbb{B}_n(g_v)}{\sqrt{E_{\mathrm{P}_n}[g_v^2]}}, \quad Z_n^\star(v) = \frac{\mathbb{G}_n^o(\hat{g}_v)}{\sqrt{\mathbb{E}_n[\hat{g}_v^2]}},$$

$$\sigma_n^2(v) = E_{\mathrm{P}_n}\left[g_v^2\right]/\left(nh_n^d\right), \quad s_n^2(v) = \mathbb{E}_n\left[\hat{g}_v^2\right]/\left(nh_n^d\right), \quad \delta_n = 1/\log n,$$

$$\bar{a}_n \lesssim \sqrt{\log n}, \quad \bar{\sigma}_n \lesssim \sqrt{1/(nh^d)},$$

$$a_n(\mathsf{V}) = \Big(2\sqrt{\log\big\{C\big(1 + C'\big(1 + h_n^{-1}\big)\operatorname{diam}(\mathsf{V})\big)^d\big\}}\Big) \vee (1 + \sqrt{d}),$$

for some constants C *and* C', *where* $\operatorname{diam}(\mathsf{V})$ *denotes the diameter of the set* V. *Moreover*, $\mathrm{P}[\mathcal{E} > x] = \exp(-x/2)$.

REMARK 3: Lemma 7 verifies the main conditions, Conditions C.1–C.4. These conditions enable construction of either simulated or analytical critical values. For the latter, the pth quantile of $\mathcal{E}$ is given by $c(p) = -2\log(1-p)$, so we can set

$$k_{n,\mathsf{V}}(p) = a_n(\mathsf{V}) - 2\log(1-p)/a_n(\mathsf{V}), \tag{4.4}$$

where

$$a_n(\mathsf{V}) = \Big(2\sqrt{\log\big\{\ell_n\big(1 + \ell_n\big(1 + h_n^{-1}\big)\operatorname{diam}(\mathsf{V})\big)^d\big\}}\Big) \tag{4.5}$$

is a feasible version of the scaling factor in which unknown constants have been replaced by the slowly growing sequence ℓ_n. Note that $\mathsf{V} \mapsto k_{n,\mathsf{V}}(p)$ is monotone in V for large n, as required in the analytical construction given in Definition 2. A sharper analytical approach can be based on Hotelling's tube method or on the use of extreme value theory. For details of the extreme value approach, we refer the reader to Chernozhukov, Lee, and Rosen (2009). Note that the simulation-based approach is effectively a numeric version of the exact version of the tube formula and is less conservative than using simplified tube formulas. In Chernozhukov, Lee, and Rosen (2009), we established that inference based on extreme value theory is valid, but the asymptotic approximation is accurate only when sets V are "large" and does not seem to provide an accurate approximation when V is small. Moreover, it often requires a very large sample size for accuracy even when V is large.

LEMMA 8—Condition NK Implies Condition S in Some Cases: *Suppose Condition* NK *holds*. *Then* (i) *the radius* φ_n *of equicontinuity of* Z_n^* *obeys*

$$\varphi_n \le o(1) \cdot \left(\frac{h_n}{\sqrt{\log n}}\right)$$

for any $o(1)$ *term*. (ii) *If Condition* V *holds and*

$$\left(\sqrt{\frac{\log n}{nh^d}}\log n\right)^{1/\rho_n} c_n^{-1} = o\left(\frac{h_n}{\sqrt{\log n}}\right), \tag{4.6}$$

then Condition S *holds*.

The following theorem is an immediate consequence of Lemmas 7 and 8 and Theorems 1, 2, and 3.

THEOREM 6—Estimation and Inference for Bounding Functions Using Local Methods: *Suppose Condition* NK *holds and consider the interval estimator* $\hat{\theta}_{n0}(p)$ *given in Definition* 1 *with either analytical critical values specified in Remark* 3 *or simulation-based critical values given in Definition* 3 *for the simulation process* $Z_n^\star$ *specified above*. (a) *Then* (i) $\mathrm{P}_n[\theta_{n0} \le \hat{\theta}_{n0}(p)] \ge p - o(1)$, (ii) $|\theta_{n0} - \hat{\theta}_{n0}(p)| = O_{\mathrm{P}_n}(\sqrt{\log n/(nh_n^d)})$, *and* (iii) $\mathrm{P}_n(\theta_{n0} + \mu_n\sqrt{\log n/(nh_n^d)} \ge \hat{\theta}_{n0}(p)) \to 1$ *for any* $\mu_n \to_{\mathrm{P}_n} \infty$. (b) *Moreover*, *for simulation-based critical values*, *if Condition* V *and* (4.6) *hold*, *then* (i) $\mathrm{P}_n[\theta_{n0} \le \hat{\theta}_{n0}(p)] = p - o(1)$, (ii) $|\theta_{n0} - \hat{\theta}_{n0}(p)| = O_{\mathrm{P}_n}(\sqrt{1/(nh_n^d)})$, *and* (iii) $\mathrm{P}_n(\theta_{n0} + \mu_n\sqrt{1/(nh_n^d)} \ge \hat{\theta}_{n0}(p)) \to 1$ *for any* $\mu_n \to_{\mathrm{P}_n} \infty$.

In Appendix F in the Supplemental Material, we provide an example where the bounding function is obtained from conditional moment inequalities and where Condition NK holds under primitive conditions. We provide only one example for brevity, but more examples can be covered as for series estimation in Section 4.2. In Appendix G in the Supplemental Material, we provide conditions under which the required strong approximation in Condition NK(i) holds.

5. STRONG APPROXIMATION FOR ASYMPTOTICALLY LINEAR SERIES ESTIMATORS

In the following theorem, we establish strong approximation for series estimators appearing in the previous section as part of Condition NS(i). In Appendix I of the Supplemental Material, we demonstrate as a leading example how the required asymptotically linear representation can be achieved from primitive conditions for the case of estimation of a conditional mean function.

THEOREM 7—Strong Approximation for Asymptotically Linear Series Estimators: *Let* $(A, \mathcal{A}, \mathrm{P}_n)$ *be the probability space for each* n *and let* $n \to \infty$. *Let* $\delta_n \to 0$ *be a sequence of constants converging to* 0 *at no faster than a polynomial rate in* n. (a) *Assume the series estimator has the form* $\hat{\theta}_n(v) = p_n(v)'\hat{\beta}_n$, *where* $p_n(v) := (p_{n,1}(v), \ldots, p_{n,K_n}(v))'$ *is a collection of* K_n*-dimensional approximating functions such that* $K_n \to \infty$ *and* $\hat{\beta}_n$ *is a* K_n*-vector of estimates*. (b) *Assume the estimator* $\hat{\beta}_n$ *satisfies an asymptotically linear representation around some* K_n*-dimensional vector* β_n,

$$\Omega_n^{-1/2}\sqrt{n}(\hat{\beta}_n - \beta_n) = n^{-1/2}\sum_{i=1}^n u_{i,n} + r_n, \quad \|r_n\| = o_{\mathrm{P}_n}(\delta_n), \tag{5.1}$$

$$(5.2)\qquad u_{i,n},\quad i=1,\dots,n\quad \textit{are independent with}$$
$$E_{\mathrm{P}_n}[u_{i,n}]=0,\quad E_{\mathrm{P}_n}\big[u_{i,n}u_{i,n}'\big]=I_{K_n},$$

$$(5.3)\qquad \Delta_n=\sum_{i=1}^n E\|u_{i,n}\|^3/n^{3/2}\quad \textit{such that } K_n\Delta_n/\delta_n^3\to 0,$$

where Ω_n is a sequence of $K_n\times K_n$ invertible matrices. (c) *Assume the function $\theta_n(v)$ admits the approximation $\theta_n(v)=p_n(v)'\beta_n+A_n(v)$, where the approximation error $A_n(v)$ satisfies $\sup_{v\in\mathcal{V}}\sqrt{n}|A_n(v)|/\|g_n(v)\|=o(\delta_n)$ for $g_n(v):=p_n(v)'\Omega_n^{1/2}$. Then we can find a random normal vector $\mathcal{N}_n=_d\mathcal{N}(0,I_{K_n})$ such that $\|\Omega_n^{-1/2}\sqrt{n}(\hat\beta_n-\beta_n)-\mathcal{N}_n\|=o_{\mathrm{P}_n}(\delta_n)$ and*

$$\sup_{v\in\mathcal{V}}\left|\frac{\sqrt{n}(\hat\theta_n(v)-\theta_n(v))}{\|g_n(v)\|}-\frac{g_n(v)}{\|g_n(v)\|}\mathcal{N}_n\right|=o_{\mathrm{P}_n}(\delta_n).$$

The following corollary covers the cases considered in the examples in the previous section.

COROLLARY 1—A Leading Case of Influence Function: *Suppose the conditions of Theorem 7 hold with $u_{i,n}:=\Omega_n^{-1/2}Q_n^{-1}p_n(V_i)\epsilon_i$, where (V_i,ϵ_i) are i.i.d. with $E_{\mathrm{P}_n}[\epsilon_i p_n(V_i)]=0$, $S_n:=E_{\mathrm{P}_n}[\epsilon_i^2 p_n(V_i)p_n(V_i)']$, and $\Omega_n:=Q_n^{-1}S_n(Q_n^{-1})'$, where Q_n^{-1} is a nonrandom invertible matrix, and $\|\Omega_n^{-1/2}Q_n^{-1}\|\le\tau_n$, $E_{\mathrm{P}_n}[|\epsilon_i|^3|V_i=v]$ is bounded above uniformly in $v\in\mathrm{support}(V_i)$, and $E_{\mathrm{P}_n}[\|p_n(V_i)\|^3]\le C_nK_n^{3/2}$. Then the key growth restriction on the number of series terms $K_n\Delta_n/\delta_n^3\to 0$ holds if $\tau_n^6C_n^2{K_n}^5/(n\delta_n^6)\to 0$.*

REMARK 4—Applicability: In this paper, $\delta_n=1/\log n$. Sufficient conditions for linear approximation Theorem 7(b) follow from results in the literature on series estimation, for example, Andrews (1991), Newey (1995, 1997), and Belloni, Chernozhukov, and Fernandez-Val (2011). See also Chen (2007) and references therein for a general overview of sieve estimation and recent developments. The main text provides several examples, including mean and quantile regression, with primitive conditions that provide sufficient conditions for the linear approximation.

6. IMPLEMENTATION

In Section 6.1, we lay out steps for implementation of parametric and series estimation of bounding functions, while in Section 6.2, we provide implementation steps for kernel-type estimation. The end goal in each case is to obtain estimators $\hat\theta_{n0}(p)$ that provide bias-corrected estimates or the end points of confidence intervals, depending on the chosen value of p (e.g., $p=1/2$ or $p=1-\alpha$). As before, we focus here on the upper bound. If instead $\hat\theta_{n0}(p)$

were the lower bound for θ^*, given by the supremum of a bounding function, the same algorithm could be applied to perform inference on $-\theta^*$, bounded above by the infimum of the negative of the original bounding function, and then any inference statements for $-\theta^*$ could trivially be transformed to inference statements for θ^*. Indeed, any set of lower and upper bounds can be similarly transformed to a collection of upper bounds, and the above algorithm can be applied to perform inference on θ^*, for example, according to the methods laid out for inference on parameters bounded by conditional moment inequalities in Section 3.[21] Alternatively, if one wishes to perform inference on the identified set in such circumstances, one can use the intersection of upper and lower one-sided intervals each based on $\tilde{p} = (1+p)/2$ as an asymptotic level-p confidence set for Θ_I, which is valid by Bonferroni's inequality.[22]

6.1. *Parametric and Series Estimators*

Let β_n denote the bounding function parameter vector if parametric estimation is used, while β_n denotes the coefficients of the series terms if series estimation is used, as in Section 4.2. K denotes the dimension of β_n and I_K denotes the K-dimensional identity matrix. As in the main text, let $p_n(v) = \partial\theta_n(v, \hat{\beta}_n)/\partial\beta_n$, which are simply the series terms in the case of series estimation.

ALGORITHM 1—Implementation for Parametric and Series Estimation:

Step 1. Set $\tilde{\gamma}_n \equiv 1 - 0.1/\log n$. Simulate a large number R of draws denoted $Z_1, \ldots, Z_R$ from the K-variate standard normal distribution $\mathcal{N}(0, I_K)$.

Step 2. Compute $\hat{\Omega}_n$, a consistent estimator for the large sample variance of $\sqrt{n}(\hat{\beta}_n - \beta_n)$.

Step 3. For each $v \in \mathcal{V}$, compute $\hat{g}(v) = p_n(v)'\hat{\Omega}_n^{1/2}$ and set $s_n(v) = \|\hat{g}(v)\|/\sqrt{n}$.

Step 4. Compute

$$k_{n,\mathcal{V}}(\tilde{\gamma}_n) = \tilde{\gamma}_n\text{-quantile of } \Big\{\sup_{v\in\mathcal{V}}\big(\hat{g}(v)'Z_r/\|\hat{g}(v)\|\big), r = 1, \ldots, R\Big\},$$

and

$$\hat{V}_n = \Big\{v \in \mathcal{V} : \hat{\theta}_n(v) \le \min_{v\in\mathcal{V}}\big(\hat{\theta}_n(v) + k_{n,\mathcal{V}}(\tilde{\gamma}_n)s_n(v)\big) + 2k_{n,\mathcal{V}}(\tilde{\gamma}_n)s_n(v)\Big\}.$$

[21]For example, if we have $\theta_n^l(z) \le \theta_n^* \le \theta_n^u(z)$ for all $z \in \mathcal{Z}$, then we can equivalently write $\min_{z\in\mathcal{Z}} \min_{j=1,2} g_n(\theta_n^*, z, j) \ge 0$, where $g_n(\theta_n^*, z, 1) = \theta_n^u(z) - \theta_n^*$ and $g_n(\theta_n^*, z, 2) = \theta_n^* - \theta_n^l(z)$. Then we can apply our method through use of the auxiliary function $g_n(\theta_n, z, j)$, in similar fashion as in Example C with multiple conditional moment inequalities.

[22]In an earlier version of this paper (Chernozhukov, Lee, and Rosen (2009)), we provided a different method for inference on a parameter with both lower and upper bounding functions, which can also be used for valid inference on θ^*.

Step 5. Compute

$$k_{n,\hat{V}_n}(p) = p\text{-quantile of } \Big\{\sup_{v\in\hat{V}_n}\big(\hat{g}(v)'Z_r/\|\hat{g}(v)\|\big), r=1,\ldots,R\Big\},$$

and set

$$\hat{\theta}_{n0}(p) = \inf_{v\in\mathcal{V}}\big[\hat{\theta}_n(v) + k_{n,\hat{V}_n}(p)\|\hat{g}(v)\|/\sqrt{n}\big].$$

An important special case of the parametric setup is that where the support of v is finite, as in Example 1 of Section 4.1, so that $\mathcal{V} = \{1,\ldots,J\}$. In this case, the algorithm applies with $\theta_n(v,\beta_n) = \sum_{j=1}^{J} 1[v=j]\beta_{nj}$, that is, where for each j, $\theta_n(j,\beta_n) = \beta_{nj}$ and $\hat{g}(v) = (1[v=1],\ldots,1[v=J])\cdot\hat{\Omega}_n^{1/2}$. Note that this covers the case where the bounding function is a conditional mean or quantile with discrete conditioning variable, such as conditional mean estimation with discrete regressors, in which case $\beta_{nj} = E[Y|V=j]$ can be estimated by a sample mean.

REMARK 5: In the case of series estimation, if desired, one can bypass simulation of the stochastic process by instead employing the analytical critical value in Step 4, $k_{n,\mathsf{V}}(p) = a_n(\mathsf{V}) - 2\log(1-p)/a_n(\mathsf{V})$ from Remark 2 in Section 4.2. This is convenient because it does not involve simulation, though it requires computation of $a_n(\hat{V}_n) = 2\sqrt{\log\{\ell_n(1+\ell_n L_n \operatorname{diam}(\hat{V}_n))^d\}}$. Moreover, it could be too conservative in some applications. Thus, we recommend using simulation, unless the computational cost is too high.

6.2. *Kernel-Type Estimators*

In this section, we describe the steps for implementation of kernel-type estimators.

ALGORITHM 2—Implementation for Kernel Case:

Step 1. Set $\gamma_n \equiv 1 - 0.1/\log n$. Simulate $R\times n$ independent draws from $N(0,1)$, denoted by $\{\eta_{ir} : i=1,\ldots,n, r=1,\ldots,R\}$, where n is the sample size and R is the number of simulation repetitions.

Step 2. For each $v\in\mathcal{V}$ and $r=1,\ldots,R$, compute $\mathbb{G}_n^o(\hat{g}_v; r) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta_{ir}\times \hat{g}_v(U_i,Z_i)$, where $\hat{g}_v(U_i,Z_i)$ is defined in Section 4.3, that is,

$$\hat{g}_v(U_i,Z_i) = \frac{e_j'\hat{U}_i}{(h_n^d)^{1/2}\hat{f}_n(z)}\mathbf{K}\left(\frac{z-Z_i}{h_n}\right).$$

Let $s_n^2(v) = \mathbb{E}_n[\hat{g}_v^2]/(nh_n^d)$ and $\mathbb{E}_n[\hat{g}_v^2] = n^{-1}\sum_{i=1}^{n}\hat{g}_v^2(U_i,Z_i)$. Here, $\hat{U}_i$ is the kernel-type regression residual and $\hat{f}_n(z)$ is the kernel density estimator of density of Z_i.

Step 3. Compute $k_{n,\mathcal{V}}(\gamma_n) = \gamma_n$-quantile of $\{\sup_{v\in\mathcal{V}} \mathbb{G}_n^o(\hat{g}_v; r)/\sqrt{\mathbb{E}_n[\hat{g}_v^2]}, r = 1, \dots, R\}$ and $\hat{V}_n = \{v \in \mathcal{V} : \hat{\theta}_n(v) \leq \min_{v\in\mathcal{V}}(\hat{\theta}_n(v) + k_{n,\mathcal{V}}(\gamma_n)s_n(v)) + 2k_{n,\mathcal{V}}(\gamma_n) \times s_n(v)\}$.

Step 4. Compute $k_{n,\hat{V}_n}(p) = p$-quantile of $\{\sup_{v\in\hat{V}_n} \mathbb{G}_n^o(\hat{g}_v; r)/\sqrt{\mathbb{E}_n[\hat{g}_v^2]}, r = 1, \dots, R\}$ and set $\hat{\theta}_{n0}(p) = \inf_{v\in\mathcal{V}}[\hat{\theta}(v) + k_{n,\hat{V}_n}(p)s_n(v)]$.

REMARK 6: (i) The researcher also has the option of employing an analytical approximation in place of simulation if desired. This can be done by using $k_{n,\mathsf{V}}(p) = a_n(\mathsf{V}) - 2\log(1-p)/a_n(\mathsf{V})$ from Remark 3, but requires computation of

$$a_n(\hat{V}_n) = 2\sqrt{\log\left\{\ell_n\left(1 + \ell_n\left(1 + h_n^{-1}\right)\operatorname{diam}(\hat{V}_n)\right)^d\right\}}.$$

This approximation could be too conservative in some applications, and thus we recommend using simulation, unless the computational cost is too high. (ii) In the case where the bounding function is nonseparable in a parameter of interest, a confidence interval for this parameter can be constructed as described in Section 6.1, where Step 1 is carried out once and Steps 2–4 are executed iteratively on a set of parameter values approximating the parameter space. However, the bandwidth, $\hat{f}_n(z)$, and $\mathbf{K}(\frac{z-Z_i}{h_n})$, do not vary across iterations and thus only need to be computed once.

7. MONTE CARLO EXPERIMENTS

In this section, we present results of Monte Carlo experiments to illustrate the finite-sample performance of our method. We consider a Monte Carlo design with bounding function

$$\theta(v) := L\phi(v), \tag{7.1}$$

where L is a constant and $\phi(\cdot)$ is the standard normal density function. Throughout the Monte Carlo experiments, the parameter of interest is $\theta_0 = \sup_{v\in\mathcal{V}} \theta(v)$.[23]

7.1. *Data-Generating Processes*

We consider four Monte Carlo designs (data-generating processes (DGP)) for the sake of illustration.[24] In the first Monte Carlo design, labeled DGP1, the bounding function is completely flat so that $V_0 = \mathcal{V}$. In the second design,

[23]Previous sections focused on $\theta_0 = \inf_{v\in\mathcal{V}} \theta(v)$ rather than $\theta_0 = \sup_{v\in\mathcal{V}} \theta(v)$. This is not a substantive difference, as for any function $\theta(\cdot)$, $\sup_{v\in\mathcal{V}} \theta(v) = -\inf_{v\in\mathcal{V}}(-\theta(v))$.

[24]We consider some additional Monte Carlo designs in Appendix L in the Supplemental Material.

DGP2, the bounding function is nonflat, but smooth in a neighborhood of its maximizer, which is unique so that V_0 is a singleton. In DGP3 and DGP4, the bounding function is also nonflat and smooth in a neighborhood of its (unique) maximizer, though relatively peaked. Illustrations of these bounding functions are provided in Figures S.1 and S.2 in the Supplemental Material. In practice, the shape of the bounding function is unknown, and the inference and estimation methods we consider do not make use of this information. As we describe in more detail below, we evaluate the finite-sample performance of our approach in terms of coverage probability for the true point θ_0 and coverage for a false parameter value θ that is close to but below θ_0. We compare the performance of our approach to that of the Cramer–von Mises statistic proposed by AS. DGP1 and DGP2 in particular serve to effectively illustrate the relative advantages of both procedures as we describe below. Neither approach dominates.

For all DGPs, we generated 1000 independent samples from the model

$$V_i \sim \text{Unif}[-2, 2], \quad U_i = \min\bigl\{\max\{-3, \sigma\tilde{U}_i\}, 3\bigr\}, \quad \text{and}$$
$$Y_i = L\phi(V_i) + U_i,$$

where $\tilde{U}_i \sim N(0, 1)$, and L and σ are constants. We set these constants in the manner

$$\text{DGP1:} \quad L = 0 \text{ and } \sigma = 0.1; \qquad \text{DGP2:} \quad L = 1 \text{ and } \sigma = 0.1;$$
$$\text{DGP3:} \quad L = 5 \text{ and } \sigma = 0.1; \qquad \text{DGP4:} \quad L = 5 \text{ and } \sigma = 0.01.$$

We considered sample sizes $n = 500$ and $n = 1000$, and we implemented both series and kernel-type estimators to estimate the bounding function $\theta(v)$ in (7.1). We set $\mathcal{V}$ to be an interval between the 0.05 and 0.95 sample quantiles of V_i's so as to avoid undue influence of outliers at the boundary of the support of V_i. For both types of estimators, we computed critical values via simulation as described in Section 6, and we implemented our method with both the conservative but simple, nonstochastic choice $\hat{V} = \mathcal{V}$ and the set estimate $\hat{V} = \hat{V}_n$ described in Section 3.2.

7.2. *Series Estimation*

For basis functions, we use polynomials and cubic B-splines with knots equally spaced over the sample quantiles of V_i. The number $K = K_n$ of approximating functions was obtained by the simple rule-of-thumb

$$K = \underline{\hat{K}}, \quad \hat{K} := \hat{K}_{\text{cv}} \times n^{-1/5} \times n^{2/7}, \tag{7.2}$$

where $\underline{a}$ is defined as the largest integer that is smaller than or equal to a, and $\hat{K}_{\text{cv}}$ is the minimizer of the leave-one-out least squares cross-validation score.

If $\theta(v)$ is twice continuously differentiable, then a cross-validated K has the form $K \propto n^{1/5}$ asymptotically. Hence, the multiplicative factor $n^{-1/5} \times n^{2/7}$ in (7.2) ensures that the bias is asymptotically negligible from undersmoothing.[25]

7.3. *Kernel-Type Estimation*[26]

We use local linear smoothing since it is known to behave better at the boundaries of the support than the standard kernel method. We used the kernel function $K(s) = \frac{15}{16}(1-s^2)^2 1(|s| \leq 1)$ and the rule-of-thumb bandwidth

$$h = \hat{h}_{\mathrm{ROT}} \times \hat{s}_v \times n^{1/5} \times n^{-2/7}, \tag{7.3}$$

where $\hat{s}_v$ is the square root of the sample variance of the V_i and $\hat{h}_{\mathrm{ROT}}$, is the rule-of-thumb bandwidth for estimation of $\theta(v)$ with studentized V, as prescribed in Section 4.2 of Fan and Gijbels (1996). The exact form of $\hat{h}_{\mathrm{ROT}}$ is

$$\hat{h}_{\mathrm{ROT}} = 2.036 \left[\frac{\tilde{\sigma}^2 \int w_0(v)\, dv}{n^{-1} \sum_{i=1}^{n} \{\tilde{\theta}^{(2)}(\tilde{V}_i)\}^2 w_0(\tilde{V}_i)} \right]^{1/5} n^{-1/5},$$

where $\tilde{V}_i$'s are studentized V_i's, $\tilde{\theta}^{(2)}(\cdot)$ is the second-order derivative of the global quartic parametric fit of $\theta(v)$ with studentized V_i, $\tilde{\sigma}^2$ is the simple average of squared residuals from the parametric fit, $w_0(\cdot)$ is a uniform weight function that has value 1 for any $\tilde{V}_i$ that is between the 0.10 and 0.90 sample quantiles of $\tilde{V}_i$. Again, the factor $n^{1/5} \times n^{-2/7}$ is multiplied in (7.3) to ensure that the bias is asymptotically negligible due to undersmoothing.

7.4. *Simulation Results*

To evaluate the relative performance of our inference method, we also implemented one of the inference methods proposed by AS, specifically their

[25] For B-splines, the optimal $\hat{K}_{\mathrm{cv}}$ was first selected from the first $5 \times n^{1/5}$ values starting from 5, with $n^{1/5}$ rounded up to the nearest integer. If the upper bound was selected, the cross-validation (CV) score of $\hat{K}_{\mathrm{cv}}$ was compared to that of $\hat{K}_{\mathrm{cv}} + 1$ iteratively, such that $\hat{K}_{\mathrm{cv}}$ was increased until further increments resulted in no improvement. This allows $\hat{K}_{\mathrm{cv}} \propto n^{1/5}$ and provides a crude check against the upper bound binding in the CV search, though in these DGPs, results differed little from those searching over $\{5, 6, 7, 8, 9\}$, reported in Chernozhukov, Lee, and Rosen (2009). For polynomials, the CV search was limited to the set $\{3, 4, 5, 6\}$ due to multicollinearity issues that arose when too many terms were used.

[26] Appendices G and H in the Supplemental Material provide strong approximation results and proofs for kernel-type estimators, including the local linear estimator used here.

TABLE I

RESULTS FOR MONTE CARLO EXPERIMENTS (CLR WITH SERIES ESTIMATION USING B-SPLINES)

DGP	Sample Size	Critical Value Estimating V_n?	Ave. Smoothing Parameter	Cov. Prob.	False Cov. Prob.	Ave. Argmax Set Min.	Ave. Argmax Set Max.
1	500	No	9.610	0.944	0.149	−1.800	1.792
1	500	Yes	9.610	0.944	0.149	−1.800	1.792
1	1000	No	10.490	0.947	0.013	−1.801	1.797
1	1000	Yes	10.490	0.947	0.013	−1.801	1.797
2	500	No	9.680	0.992	0.778	−1.800	1.792
2	500	Yes	9.680	0.982	0.661	−0.762	0.759
2	1000	No	10.578	0.997	0.619	−1.801	1.797
2	1000	Yes	10.578	0.982	0.470	−0.669	0.670
3	500	No	11.584	0.995	0.903	−1.800	1.792
3	500	Yes	11.584	0.984	0.765	−0.342	0.344
3	1000	No	13.378	0.994	0.703	−1.801	1.797
3	1000	Yes	13.378	0.971	0.483	−0.290	0.290
4	500	No	18.802	0.996	0.000	−1.800	1.792
4	500	Yes	18.802	0.974	0.000	−0.114	0.114
4	1000	No	20.572	1.000	0.000	−1.801	1.797
4	1000	Yes	20.572	0.977	0.000	−0.098	0.091

Cramér–von Mises-type (CvM) statistic with both plug-in asymptotic (PA/Asy) and asymptotic generalized moment selection (GMS/Asy) critical values. For instrument functions, we used countable hypercubes and the S-function of AS Section 3.2.[27] We set the weight function and tuning parameters for the CvM statistic exactly as in AS (see AS Section 9). These values performed well in their simulations, but our Monte Carlo design differs from theirs, and alternative choices of tuning parameters could perform more or less favorably in our design. We did not examine sensitivity to the choice of tuning parameters for the CvM statistic.

The coverage probability (CP) of confidence intervals with nominal level 95% is evaluated for the true lower bound θ_0, and false coverage probability (FCP) is reported at $\theta = \theta_0 - 0.02$. There were 1000 replications for each experiment. Tables I, II, and III summarize the results. The acronyms CLR and AS refer to our inference method and that of AS, respectively.

We first consider the performance of our method for DGP1. In terms of coverage for θ_0, both series estimators and the local linear estimator perform reasonably well, with the series estimators performing best. The polynomial series and local linear estimators perform somewhat better in terms of false coverage probabilities, which decrease with the sample size for all estimators.

[27]All three S-functions in AS Section 3.2 are equivalent in our design, since there is a single conditional moment inequality.

TABLE II

RESULTS FOR MONTE CARLO EXPERIMENTS (AS WITH CVM-TYPE STATISTIC)

DGP	Sample Size	Critical Value	Cov. Prob.	False Cov. Prob.
1	500	PA/Asy	0.959	0.007
1	500	GMS/Asy	0.955	0.007
1	1000	PA/Asy	0.958	0.000
1	1000	GMS/Asy	0.954	0.000
2	500	PA/Asy	1.000	1.000
2	500	GMS/Asy	1.000	0.977
2	1000	PA/Asy	1.000	1.000
2	1000	GMS/Asy	1.000	0.933
3	500	PA/Asy	1.000	1.000
3	500	GMS/Asy	1.000	1.000
3	1000	PA/Asy	1.000	1.000
3	1000	GMS/Asy	1.000	1.000
4	500	PA/Asy	1.000	1.000
4	500	GMS/Asy	1.000	1.000
4	1000	PA/Asy	1.000	1.000
4	1000	GMS/Asy	1.000	1.000

The argmax set V_0 is the entire set $\mathcal{V}$, and our set estimator $\hat{V}_n$ detects this. Turning to DGP2, we see that coverage for θ_0 is in all cases roughly 0.98–0.99. There is nontrivial power against the false parameter θ in all cases, with the series estimators giving the lowest false coverage probabilities. For DGP3, the bounding function is relatively peaked compared to the smooth but nonflat bounding function of DGP2. Consequently, the average end points of the preliminary set estimator $\hat{V}_n$ become more concentrated around 0, the maximizer of the bounding function. Performance in terms of coverage probabilities improves in nearly all cases, with the series estimators performing significantly better when $n = 1000$ and $\hat{V}_n$ is used. With DGP4, the bounding function remains as in DGP3, but now with the variance of Y_i decreased by a factor of 100. The result is that the bounding function is more accurately estimated at every point. Moreover, the set estimator $\hat{V}_n$ is now a much smaller interval around 0. Coverage frequencies for θ_0 do not change much relative to DGP3, but false coverage probabilities drop to 0. Note that in DGP2–DGP4, our method performs better when V_n is estimated in that it makes the coverage probability more accurate and the false coverage probability smaller. DGP3 and DGP4 serve to illustrate the convergence of our set estimator $\hat{V}_n$ when the bounding function is peaked and precisely estimated, respectively.

In Table II, we report the results of using the CvM statistic of AS to perform inference. For DGP1 with a flat bounding function, the CvM statistic with both the PA/Asy and GMS/Asy performs well. Coverage frequencies for θ_0 were close to the nominal level, closer than our method using polynomial

TABLE III

RESULTS FOR MONTE CARLO EXPERIMENTS (OTHER ESTIMATION METHODS)

DGP	Sample Size	Critical Value Estimating V_n?	Ave. Smoothing Parameter	Cov. Prob.	False Cov. Prob.	Ave. Argmax Set Min.	Ave. Argmax Set Max.
CLR With Series Estimation Using Polynomials							
1	500	No	5.524	0.954	0.086	−1.800	1.792
1	500	Yes	5.524	0.954	0.086	−1.800	1.792
1	1000	No	5.646	0.937	0.003	−1.801	1.797
1	1000	Yes	5.646	0.937	0.003	−1.801	1.797
2	500	No	8.340	0.995	0.744	−1.800	1.792
2	500	Yes	8.340	0.989	0.602	−0.724	0.724
2	1000	No	9.161	0.996	0.527	−1.801	1.797
2	1000	Yes	9.161	0.977	0.378	−0.619	0.620
3	500	No	8.350	0.998	0.809	−1.800	1.792
3	500	Yes	8.350	0.989	0.612	−0.300	0.301
3	1000	No	9.155	0.996	0.560	−1.801	1.797
3	1000	Yes	9.155	0.959	0.299	−0.253	0.252
4	500	No	8.254	1.000	0.000	−1.800	1.792
4	500	Yes	8.254	0.999	0.000	−0.081	0.081
4	1000	No	9.167	0.998	0.000	−1.801	1.797
4	1000	Yes	9.167	0.981	0.000	−0.069	0.069
CLR With Local Linear Estimation							
1	500	No	0.606	0.923	0.064	−1.799	1.792
1	500	Yes	0.606	0.923	0.064	−1.799	1.792
1	1000	No	0.576	0.936	0.003	−1.801	1.796
1	1000	Yes	0.576	0.936	0.003	−1.801	1.796
2	500	No	0.264	0.995	0.871	−1.799	1.792
2	500	Yes	0.264	0.989	0.808	−0.890	0.892
2	1000	No	0.218	0.996	0.779	−1.801	1.796
2	1000	Yes	0.218	0.990	0.675	−0.776	0.776
3	500	No	0.140	0.995	0.943	−1.799	1.792
3	500	Yes	0.140	0.986	0.876	−0.426	0.424
3	1000	No	0.116	0.992	0.907	−1.801	1.796
3	1000	Yes	0.116	0.986	0.816	−0.380	0.377
4	500	No	0.078	0.991	0.000	−1.799	1.792
4	500	Yes	0.078	0.981	0.000	−0.142	0.142
4	1000	No	0.064	0.997	0.000	−1.801	1.796
4	1000	Yes	0.064	0.991	0.000	−0.127	0.127

series or local linear regression. The CvM statistic has a lower false coverage probability than the CLR confidence intervals in this case, although at a sample size of 1000, the difference is not large. For DGP2, the bounding function is nonflat but smooth in a neighborhood of V_0 and the situation is much different. For both PA/Asy and GMS/Asy critical values with the CvM statistic, coverage frequencies for θ_0 were 1. Our confidence intervals also overcovered in this case, with coverage frequencies of roughly 0.98–0.99. Moreover, the

TABLE IV
COMPUTATION TIMES OF MONTE CARLO EXPERIMENTS

	AS	Series (B-splines)	Series (Polynomials)	Local Linear
Total minutes for simulations	8.67	57.99	19.44	83.75
Average seconds for each test	0.03	0.22	0.07	0.31
Ratio relative to AS	1.00	6.69	2.24	9.66

CvM statistic has low power against the false parameter θ, with coverage 1 with PA/Asy and coverage 0.977 and 0.933 with sample size 500 and 1000, respectively, using GMS/Asy critical values. For DGP3 and DGP4, both critical values for the CvM statistic gave coverage for θ_0 and the false parameter θ equal to 1. Thus under DGP2–DGP4, our confidence intervals perform better by both measures. In summary, overall neither approach dominates.

Thus, in our Monte Carlo experiments, the CvM statistic exhibits better power when the bounding function is flat, while our confidence intervals exhibit better power when the bounding function is nonflat. AS established that the CvM statistic has power against some $n^{-1/2}$ local alternatives under conditions that are satisfied under DGP1, but that do not hold when the bounding function has a unique minimum.[28] We have established local asymptotic power for nonparametric estimators of polynomial order less distant than $n^{-1/2}$ that apply whether the bounding function is flat or nonflat. Our Monte Carlo results accord with these findings.[29] In the Supplemental Material, we present further supporting Monte Carlo evidence and local asymptotic power analysis to show why our method performs better than the AS method in nonflat cases.

In Table IV, we report computation times for our Monte Carlo experiments.[30] The fastest performance in terms of total simulation time was achieved with the CvM statistic of AS, which took roughly 9 minutes to execute a total of 16,000 replications. Simulations using our approach with B-spline series, polynomial series, and local linear polynomials took roughly 58, 19, and 84 minutes, respectively. Based on these times, the table shows for each statistic

[28]Specifically, Assumptions LA3 and LA3′ of AS Theorem 4 do not hold when the sequence of models has a fixed bounding function with a unique minimum. As they discuss after the statement of Assumptions LA3 and LA3′, in such cases GMS and plug-in asymptotic tests have trivial power against $n^{-1/2}$ local alternatives.

[29]We did not do CP correction in our reported results. Our conclusion will remain valid even with CP correction as in AS, since our method performs better in DGP2–DGP4, where we have overcoverage.

[30]These computation times were obtained on a 2011 iMac desktop with a 2.7 GHz processor and 8 GB RAM using our implementation. Generally speaking, performance time for both methods will depend on the efficiency of one's code, so more efficient implementation times for both methods may be possible.

the average time for a single test and the relative performance of each method to that obtained using the CvM statistic.[31]

In practice, one will not perform Monte Carlo experiments, but will rather be interested in computing a single confidence region for the parameter of interest. When the bounding function is separable, our approach offers the advantage that the critical value does not vary with the parameter value being tested. As a result, we can compute a confidence region in the same amount of time it takes to compute a single test. On the other hand, to construct a confidence region based on the CvM statistic, one must compute the statistic and its associated critical value at a large number of points in the parameter space, where the number of points required will depend on the size of the parameter space and the degree of precision desired. If, however, the bounding function is not separable in the parameter of interest, then both approaches use parameter-dependent critical values.

8. CONCLUSION

In this paper, we provided a novel method for inference on intersection bounds. Bounds of this form are common in the recent literature, but two issues have posed difficulties for valid asymptotic inference and bias-corrected estimation. First, the application of the supremum and infimum operators to boundary estimates results in finite-sample bias. Second, unequal sampling error of estimated bounding functions complicates inference. We overcame these difficulties by applying a precision correction to the estimated bounding functions before taking their intersection. We employed strong approximation to justify the magnitude of the correction so as to achieve the correct asymptotic size. As a by-product, we proposed a bias-corrected estimator for intersection bounds based on an asymptotic median adjustment. We provided formal conditions that justified our approach in both parametric and nonparametric settings, the latter using either kernel or series estimators.

At least two of our results may be of independent interest beyond the scope of inference on intersection bounds. First, our result on the strong approximation of series estimators is new. This essentially provides a functional central limit theorem for any series estimator that admits a linear asymptotic expansion, and is applicable quite generally. Second, our method for inference applies to any value that can be defined as a linear programming problem with

[31]The difference in computation time between polynomial and B-spline series implementation was almost entirely due to the simpler cross-validation search from polynomials. With a search over only four values of $\hat{K}_{\text{cv}}$, simulation time for B-splines took roughly 20 minutes. Cross-validation is also in part accountable for slower performance relative to AS. We followed Section 9 of AS in choosing tuning parameters for the CvM statistic, which does not involve cross-validation. Using B-splines with a deterministic bandwidth resulted in a computation time of 12 minutes, roughly 1.4 times the total computation time for the CvM statistic. Nonetheless, we prefer cross-validation for our method in practice.

either a finite- or an infinite-dimensional constraint set. Estimators of this form can arise in a variety of contexts, including, but not limited to, intersection bounds. We therefore anticipate that although our motivation lay in inference on intersection bounds, our results may have further application.

APPENDIX A: DEFINITION OF STRONG APPROXIMATION

The following definitions are used extensively.

DEFINITION 4—Strong Approximation: Suppose that for each n, there are random variables Z_n and Z_n' defined on a probability space $(A, \mathcal{A}, \mathrm{P}_n)$ and taking values in the separable metric space (S, d_S). We say that $Z_n =_d Z_n' + o_{\mathrm{P}_n}(\delta_n)$ for $\delta_n \to 0$ if there are identically distributed copies of Z_n and Z_n', denoted $\bar{Z}_n$ and $\bar{Z}_n'$, defined on $(A, \mathcal{A}, \mathrm{P}_n)$ (suitably enriched if needed) such that

$$d_S(\bar{Z}_n, \bar{Z}_n') = o_{\mathrm{P}_n}(\delta_n).$$

Note that copies $\bar{Z}_n$ and $\bar{Z}_n'$ can always be defined on $(A, \mathcal{A}, \mathrm{P}_n)$ by suitably enriching this space by taking product probability spaces. It turns out that for the Polish spaces, this definition implies the following stronger and much more convenient form.

LEMMA 9—A Convenient Implication for Polish Spaces via Dudley and Philipp: *Suppose that (S, d_S) is Polish, that is, complete, separable metric space and that $(A, \mathcal{A}, \mathrm{P}_n)$ has been suitably enriched. Suppose that Definition 4 holds. Then there is also an identical copy Z_n^* of Z_n' such that $Z_n = Z_n^* + o_{\mathrm{P}_n}(\delta_n)$, that is,*

$$d_S(Z_n, Z_n^*) = o_{\mathrm{P}_n}(\delta_n)$$

PROOF: We start with the original probability space $(A', \mathcal{A}', \mathrm{P}_n')$ that can carry Z_n and $(\bar{Z}_n, \bar{Z}_n')$. To apply Lemma 2.11 of Dudley and Philipp (1983), we need to carry a standard uniform random variable $U \sim U(0,1)$ that is independent of Z_n. To guarantee this, we can always consider $U \sim U(0,1)$ on the standard space $([0,1], \mathcal{F}, \lambda)$, where $\mathcal{F}$ is the Borel sigma algebra on $[0,1]$ and λ is the usual Lebesgue measure, and then enrich the original space $(A', \mathcal{A}', \mathrm{P}_n')$ by formally creating a new space $(A, \mathcal{A}, \mathrm{P}_n)$ as the product of $(A', \mathcal{A}', \mathrm{P}_n')$ and $([0,1], \mathcal{F}, \lambda)$. Then using the Polishness of (S, d_S), given the joint law of $(\bar{Z}_n, \bar{Z}_n')$, we can apply Lemma 2.11 of Dudley and Philipp (1983) to construct Z_n^* such that (Z_n, Z_n^*) has the same law as $(\bar{Z}_n, \bar{Z}_n')$, so that $d_S(\bar{Z}_n, \bar{Z}_n') = o_{\mathrm{P}_n}(\delta_n)$ implies $d_S(Z_n, Z_n^*) = o_{P_n}(\delta_n)$. *Q.E.D.*

Since in all of our cases, the relevant metric spaces are either the space of continuous functions defined on a compact set equipped with the uniform metric or finite-dimensional Euclidean spaces, which are all Polish spaces, we can use Lemma 9 throughout the paper. Using this implication of strong approximation makes our proofs slightly simpler.

APPENDIX B: PROOFS FOR SECTION 3

B.1. *Some Useful Facts and Lemmas*

A useful result in our case is the anti-concentration inequality derived in Chernozhukov, Chetverikov, and Kato (2011).

LEMMA 10—Anti-Concentration Inequality (Chernozhukov, Chetverikov, and Kato (2011)): *Let $X = (X_t)_{t\in T}$ be a separable Gaussian process indexed by a semimetric space T such that $E_P[X_t] = 0$ and $E_P[X_t^2] = 1$ for all $t \in T$. Then*

$$\sup_{x\in\mathbb{R}} P\Big(\Big|\sup_{t\in T} X_t - x\Big| \le \epsilon\Big) \le C\epsilon\Big(E_P\Big[\sup_{t\in T} X_t\Big] \vee 1\Big), \quad \forall \epsilon > 0, \tag{B.1}$$

where C is an absolute constant.

An immediate consequence of this lemma is the following result.

COROLLARY 2—Anti-concentration for $\sup_{v\in V_n} Z_n^*(v)$: *Let V_n be any sequence of compact nonempty subsets in $\mathcal{V}$. Then under Conditions C.2 and C.3, we have that for $\delta_n \to 0$ such that $\delta_n = o(1/\bar{a}_n)$,*

$$\sup_{x\in\mathbb{R}} \mathrm{P}_n\Big(\Big|\sup_{v\in V_n} Z_n^*(v) - x\Big| \le \delta_n\Big) = o(1).$$

PROOF: Continuity in Condition C.2 implies separability of Z_n^*. Condition C.3 implies that $E_{\mathrm{P}_n}[\sup_{v\in V_n} Z_n^*(v)] \le E_{\mathrm{P}_n}[\sup_{v\in\mathcal{V}} Z_n^*(v)] \le K\bar{a}_n$ for some constant K that depends only on η, so that

$$\sup_{x\in\mathbb{R}} \mathrm{P}_n\Big(\Big|\sup_{v\in V_n} Z_n^*(v) - x\Big| \le \delta_n\Big) \le C\delta_n[K\bar{a}_n \vee 1] = o(1). \qquad \text{Q.E.D.}$$

LEMMA 11—Closeness in Conditional Probability Implies Closeness of Conditional Quantiles Unconditionally: *Let X_n and Y_n be random variables and let $\mathcal{D}_n$ be a random vector. Let $F_{X_n}(x|\mathcal{D}_n)$ and $F_{Y_n}(y|\mathcal{D}_n)$ denote the conditional distribution functions, and let $F_{X_n}^{-1}(p|\mathcal{D}_n)$ and $F_{Y_n}^{-1}(p|\mathcal{D}_n)$ denote the corresponding conditional quantile functions. If $\mathrm{P}_n(|X_n - Y_n| > \xi_n|\mathcal{D}_n) = o_{\mathrm{P}_n}(\tau_n)$ for some sequence $\tau_n \searrow 0$, then with unconditional probability P_n converging to 1, for some $\varepsilon_n = o(\tau_n)$,*

$$F_{X_n}^{-1}(p|\mathcal{D}_n) \le F_{Y_n}^{-1}(p + \varepsilon_n|\mathcal{D}_n) + \xi_n$$

and

$$F_{Y_n}^{-1}(p|\mathcal{D}_n) \le F_{X_n}^{-1}(p + \varepsilon_n|\mathcal{D}_n) + \xi_n, \quad \forall p \in (0, 1 - \varepsilon_n).$$

PROOF: We have that for some $\varepsilon_n = o(\tau_n)$, $\mathrm{P}_n[\mathrm{P}_n\{|X_n - Y_n| > \xi_n|\mathcal{D}_n\} \leq \varepsilon_n] \to 1$, that is, there is a set Ω_n such that $\mathrm{P}_n(\Omega_n) \to 1$ such that $\mathrm{P}_n\{|X_n - Y_n| > \xi_n|\mathcal{D}_n\} \leq \varepsilon_n$ for all $\mathcal{D}_n \in \Omega_n$. So, for all $\mathcal{D}_n \in \Omega_n$,

$$
\begin{aligned}
&F_{X_n}(x|\mathcal{D}_n) + \varepsilon_n \geq F_{Y_n+\xi_n}(x|\mathcal{D}_n) \quad \text{and} \\
&F_{Y_n}(x|\mathcal{D}_n) + \varepsilon_n \geq F_{X_n+\xi_n}(x|\mathcal{D}_n), \quad \forall x \in \mathbb{R},
\end{aligned}
$$

which implies the inequality stated in the lemma, by definition of the conditional quantile function and equivariance of quantiles to location shifts. *Q.E.D.*

B.2. *Proof of Lemma 1—Concentration of Inference on V_n*

Step 1. Letting

$$
\begin{aligned}
&A_n := \sup_{v\in V_n} Z_n(v), \quad B_n := \sup_{v\in\mathcal{V}} Z_n(v), \\
&R_n := \Bigl(\sup_{v\in\mathcal{V}}\bigl|Z_n(v)\bigr| + \kappa_n\Bigr) \sup_{v\in\mathcal{V}}\left|\frac{\sigma_n(v)}{s_n(v)} - 1\right|, \\
&A_n^* := \sup_{v\in V_n} Z_n^*(v), \quad B_n^* := \sup_{v\in\mathcal{V}} Z_n^*(v), \\
&R_n^* := \Bigl(\sup_{v\in\mathcal{V}}\bigl|Z_n^*(v)\bigr| + \kappa_n\Bigr) \sup_{v\in\mathcal{V}}\left|\frac{\sigma_n(v)}{s_n(v)} - 1\right|,
\end{aligned}
$$

we obtain

$$
\begin{aligned}
\sup_{v\in\mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} &= \sup_{v\in\mathcal{V}}\left\{\frac{\theta_{n0} - \theta_n(v)}{s_n(v)} + Z_n(v)\frac{\sigma_n(v)}{s_n(v)}\right\} \\
&= \sup_{v\in V_n}\left\{\frac{(\theta_{n0} - \theta_n(v))}{s_n(v)} + Z_n(v)\frac{\sigma_n(v)}{s_n(v)}\right\} \\
&\quad \vee \sup_{v\notin V_n}\left\{\frac{(\theta_{n0} - \theta_n(v))}{s_n(v)} + Z_n(v)\frac{\sigma_n(v)}{s_n(v)}\right\} \\
&\leq_{(1)} \sup_{v\in V_n}\left\{Z_n(v)\frac{\sigma_n(v)}{s_n(v)}\right\} \\
&\quad \vee \sup_{v\notin V_n}\left\{\frac{-\kappa_n\sigma_n(v)}{s_n(v)} + Z_n(v)\frac{\sigma_n(v)}{s_n(v)}\right\} \\
&\leq A_n \vee (B_n - \kappa_n) + 2R_n \\
&\leq_{(2)} A_n^* \vee \bigl(B_n^* - \kappa_n\bigr) + 2R_n^* + o_{\mathrm{P}_n}(\delta_n),
\end{aligned}
$$

where in inequality (1) we used that $\theta_n(v) \geq \theta_{n0}$ and $\theta_{n0} - \theta_n(v) \leq -\kappa_n \sigma_n(v)$ outside V_n, and in inequality (2) we used Condition C.2. Next, since we assumed in the statement of the lemma that $\kappa_n \lesssim \bar{a}_n + \ell\ell_n$, by Condition C.4, $R_n^* = O_{\mathrm{P}_n}(\bar{a}_n + \bar{a}_n + \ell\ell_n) o_{\mathrm{P}_n}(\delta_n/(\bar{a}_n + \ell\ell_n)) = o_{\mathrm{P}_n}(\delta_n)$. Therefore, there is a deterministic term $o(\delta_n)$ such that $\mathrm{P}_n(2R_n^* + o_{\mathrm{P}_n}(\delta_n) > o(\delta_n)) = o(1)$.[32]

Hence uniformly in $x \in [0, \infty)$,

$$\begin{aligned}
&\mathrm{P}_n\left(\sup_{v \in \mathcal{V}} \frac{(\theta_{n0} - \hat{\theta}_n(v))}{s_n(v)} > x\right) \\
&\quad \leq \mathrm{P}_n\big(A_n^* + o(\delta_n) > x\big) + \mathrm{P}_n\big(B_n^* - \kappa_n + o(\delta_n) > 0\big) + o(1) \\
&\quad \leq \mathrm{P}_n\big(A_n^* > x\big) + \mathrm{P}_n\big(B_n^* - \kappa_n > 0\big) + o(1) \\
&\quad \leq \mathrm{P}_n\big(A_n^* > x\big) + \big(1 - \gamma_n'\big) + o(1),
\end{aligned}$$

where the last two inequalities follow by Corollary 2 and by $\kappa_n = Q_{\gamma_n'}(B_n^*)$.

Step 2. To complete the proof, we must show that there is $\gamma_n' \nearrow 1$ that obeys the stated condition. Let $1 - \gamma_n' \searrow 0$ such that $1 - \gamma_n' \geq C/\ell_n$. It suffices to show that

$$\text{(B.2)} \qquad \kappa_n \leq \left(\bar{a}_n + \frac{c(\gamma_n')}{\bar{a}_n}\right) \leq \left(\bar{a}_n + \frac{\eta\ell\ell_n + \eta \log C^{-1}}{\bar{a}_n}\right),$$

where $c(\gamma_n') = Q_{\gamma_n'}(\mathcal{E})$. To show the first inequality in (B.2), note that

$$\begin{aligned}
\mathrm{P}_n\left(\sup_{v \in \mathcal{V}} Z_n^*(v) \leq \big(\bar{a}_n + c\big(\gamma_n'\big)/\bar{a}_n\big)\right) &=_{(1)} \mathrm{P}_n\big(\mathcal{E}_n(\mathcal{V}) \leq c\big(\gamma_n'\big)\big) \\
&\geq_{(2)} \mathrm{P}_n\big(\mathcal{E} \leq c\big(\gamma_n'\big)\big) = \gamma_n',
\end{aligned}$$

where equality (1) holds by definition of $\mathcal{E}_n(\mathcal{V})$ and inequality (2) holds by Condition C.3. To show the second inequality in (B.2), note that by Condition C.3, $\mathrm{P}(\mathcal{E} > t) \leq \exp(-t\eta^{-1})$ for some constant $\eta > 0$, so that $c(\gamma_n') \leq -\eta \log(1 - \gamma_n') \leq \eta\ell\ell_n + \eta \log C^{-1}$. *Q.E.D.*

B.3. *Proof of Theorem 1—Analytical Construction*

Part (i)—Level. Observe that

$$\begin{aligned}
\mathrm{P}_n\big(\theta_{n0} \leq \hat{\theta}_{n0}(p)\big) &= \mathrm{P}_n\left(\sup_{v \in \mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} \leq k_{n,\hat{V}_n}(p)\right) \\
&\geq_{(1)} \mathrm{P}_n\left(\sup_{v \in \mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} \leq k_{n,V_n}(p)\right) - \mathrm{P}_n(V_n \not\subseteq \hat{V}_n)
\end{aligned}$$

[32]Throughout the paper, we use the elementary fact that if $X_n = o_{\mathrm{P}_n}(\Delta_n)$ for some $\Delta_n \searrow 0$, then there is an $o(\Delta_n)$ term such that $\mathrm{P}_n\{|X_n| > o(\Delta_n)\} \to 0$.

$$\geq_{(2)} \mathrm{P}_n\Big(\sup_{v\in V_n} Z_n^*(v) \leq k_{n,V_n}(p)\Big) - o(1)$$
$$= \mathrm{P}_n\big(\mathcal{E}_n(V_n) \leq c(p)\big) - o(1)$$
$$\geq_{(3)} \mathrm{P}_n\big(\mathcal{E} \leq c(p)\big) - o(1) =_{(4)} p - o(1),$$

where inequality (1) follows by monotonicity of $\mathsf{V} \mapsto k_{n,\mathsf{V}}(p) = a_n(\mathsf{V}) + c(p)/a_n(\mathsf{V})$ for large n holding by construction, inequality (2) holds by Lemma 1, by $\mathrm{P}_n(V_n \not\subseteq \hat{V}_n) = o(1)$ holding by Lemma 2, and also by the fact that the critical value $k_{n,V_n}(p) \geq 0$ is nonstochastic; and inequality (3) and equality (4) hold by Condition C.3.

Part (ii)—Estimation Risk. We have that under P_n,

$$\begin{aligned}
&\big|\hat{\theta}_{n0}(p) - \theta_{n0}\big| \\
&\quad = \Big|\inf_{v\in\mathcal{V}}\big[\hat{\theta}_n(v) + k_{n,\hat{V}_n}(p)s_n(v)\big] - \theta_{n0}\Big| \\
&\quad = \left|\sup_{v\in\mathcal{V}}\left(\left[\frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} - k_{n,\hat{V}_n}(p)\right]\sigma_n(v)\frac{s_n(v)}{\sigma_n(v)}\right)\right| \\
&\quad \leq_{(1)} \left(\left|\sup_{v\in\mathcal{V}}\frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)}\right| + k_{n,\hat{V}_n}(p)\right)\bar{\sigma}_n\left(1 + o_{\mathrm{P}_n}\left(\frac{\delta_n}{\bar{a}_n + \ell\ell_n}\right)\right) \\
&\quad \leq_{(2)} \left(\left|\sup_{v\in\mathcal{V}}\frac{\theta_{n0} - \hat{\theta}_n(v)}{\sigma_n(v)}\right| + k_{n,\hat{V}_n}(p)\right)\bar{\sigma}_n\left(1 + o_{\mathrm{P}_n}\left(\frac{\delta_n}{\bar{a}_n + \ell\ell_n}\right)\right)^2 \\
&\quad \leq_{(3)} \Big(\sup_{v\in V_n}\big|Z_n^*(v)\big| + o_{\mathrm{P}_n}(\delta_n) + k_{n,\hat{V}_n}(p)\Big) \\
&\qquad \times \bar{\sigma}_n\left(1 + o_{\mathrm{P}_n}\left(\frac{\delta_n}{\bar{a}_n + \ell\ell_n}\right)\right)^2 \quad \mathrm{wp} \to 1 \\
&\quad \leq_{(4)} \Big(\sup_{v\in V_n}\big|Z_n^*(v)\big| + o_{\mathrm{P}_n}(\delta_n) + k_{n,\overline{V}_n}(p)\Big) \\
&\qquad \times \bar{\sigma}_n\left(1 + o_{\mathrm{P}_n}\left(\frac{\delta_n}{\bar{a}_n + \ell\ell_n}\right)\right)^2 \quad \mathrm{wp} \to 1 \\
&\quad \leq_{(5)} 3\left|a_n(\overline{V}_n) + \frac{O_{\mathrm{P}_n}(1)}{a_n(\overline{V}_n)} + o_{\mathrm{P}_n}(\delta_n)\right| \\
&\qquad \times \bar{\sigma}_n\left(1 + o_{\mathrm{P}_n}\left(\frac{\delta_n}{\bar{a}_n + \ell\ell_n}\right)\right)^2 \quad \mathrm{wp} \to 1 \\
&\quad \leq_{(6)} 4\left|a_n(\overline{V}_n) + \frac{O_{\mathrm{P}_n}(1)}{a_n(\overline{V}_n)}\right|\bar{\sigma}_n \quad \mathrm{wp} \to 1,
\end{aligned}$$

where inequality (1) holds by Condition C.4 and the triangle inequality; inequality (2) holds by Condition C.4; inequality (3) follows because wp $\to 1$, for some $o(\delta_n)$,

$$\sup_{v\in V_0} Z_n^*(v) - o(\delta_n) \leq_{(a)} \sup_{v\in V_0} Z_n(v) \leq_{(b)} \sup_{v\in\mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{\sigma_n(v)}$$
$$\leq_{(c)} \Big(\sup_{v\in V_n} Z_n^*(v)\Big) \vee 0 + o(\delta_n),$$

where inequality (a) is by Condition C.2, inequality (b) is by definition of Z_n, and inequality (c) is by the proof of Lemma 1, so that wp $\to 1$,

$$\left|\sup_{v\in\mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{\sigma_n(v)}\right| \leq \sup_{v\in V_n}\big|Z_n^*(v)\big| + o_{\mathrm{P}_n}(\delta_n).$$

Inequality (4) follows by Lemma 2, which implies $V_n \subseteq \hat{V}_n \subseteq \overline{V}_n$ wp $\to 1$, so that

$$k_{n,\hat{V}_n}(p) \leq k_{n,\overline{V}_n}(p) = a_n(\overline{V}_n) + \frac{c(p)}{a_n(\overline{V}_n)}.$$

Condition C.3 gives inequality (5). Inequality (6) follows because $a_n(\overline{V}_n) \geq 1$, $\bar{a}_n \geq 1$, and $\delta_n = o(1)$; this inequality is the claim that we needed to prove.

Part (iii). We have that

$$\theta_{na} - \theta_{n0} \geq 4\bar{\sigma}_n\left(a_n(\overline{V}_n) + \frac{\mu_n}{a_n(\overline{V}_n)}\right) > \hat{\theta}_{n0}(p) - \theta_{n0} \quad \text{wp } \to 1,$$

with the last inequality occurring by Part (ii) since $\mu_n \to_{\mathrm{P}_n} \infty$. *Q.E.D.*

B.4. *Proof of Theorem 2—Simulation Construction*

Part (i)—Level Consistency. Let us compare critical values

$$k_{n,V_n}(p) = Q_p\Big(\sup_{v\in V_n} Z_n^\star(v)\big|\mathcal{D}_n\Big) \quad \text{and} \quad \kappa_{n,V_n}(p) = Q_p\Big(\sup_{v\in V_n} \bar{Z}_n^*(v)\Big).$$

The former is data dependent, while the latter is deterministic. Note that $k_{n,V_n}(p) \geq 0$ by Condition C.2(b) for $p \geq 1/2$. By Condition C.2, for some deterministic term $o(\delta_n)$,

$$\mathrm{P}_n\Big(\Big|\sup_{v\in V_n} Z_n^\star(v) - \sup_{v\in V_n} \bar{Z}_n^*(v)\Big| > o(\delta_n)\big|\mathcal{D}_n\Big) = o_{\mathrm{P}_n}(1),$$

which implies by Lemma 11 that for some $\varepsilon_n \searrow 0$, wp $\to 1$,

$$k_{n,V_n}(p) \geq \big(\kappa_{n,V_n}(p - \varepsilon_n) - o(\delta_n)\big)_+ \quad \text{for all } p \in [1/2, 1 - \varepsilon_n). \tag{B.3}$$

The result follows analogously to the proof in Part (i) of Theorem 1, namely

$$
\begin{aligned}
&\mathrm{P}_n\big(\theta_{n0} \le \hat{\theta}_{n0}(p)\big) \\
&\quad = \mathrm{P}_n\bigg(\sup_{v\in\mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} \le k_{n,\hat{V}_n}(p)\bigg) \\
&\quad \ge_{(1)} \mathrm{P}_n\bigg(\sup_{v\in\mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} \le k_{n,V_n}(p)\bigg) - o(1) \\
&\quad \ge_{(2)} \mathrm{P}_n\bigg(\sup_{v\in\mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} \le \big(\kappa_{n,V_n}(p-\varepsilon_n) - o(\delta_n)\big)_+\bigg) - o(1) \\
&\quad \ge_{(3)} \mathrm{P}_n\Big(\sup_{v\in V_n} Z_n^*(v) \le \big(\kappa_{n,V_n}(p-\varepsilon_n) - o(\delta_n)\big)_+\Big) - o(1) \\
&\quad \ge \mathrm{P}_n\Big(\sup_{v\in V_n} Z_n^*(v) \le \kappa_{n,V_n}(p-\varepsilon_n) - o(\delta_n)\Big) - o(1) \\
&\quad \ge_{(4)} p - \varepsilon_n - o(1) = p - o(1),
\end{aligned}
$$

where inequality (1) follows by monotonicity of $\mathsf{V} \mapsto k_{n,\mathsf{V}}(p)$ and by $\mathrm{P}_n(V_n \nsubseteq \hat{V}_n) = o(1)$ shown in Lemma 2; inequality (2) holds by the comparison of quantiles in equation (B.3); inequality (3) holds by Lemma 1; inequality (4) holds by anti-concentration Corollary 2.

Parts (ii) and (iii)—Estimation Risk and Power. By Lemma 2, wp $\to 1$, $\hat{V}_n \subseteq \overline{V}_n$, so that $k_{n,\hat{V}_n}(p) \le k_{n,\overline{V}_n}(p)$. By Condition C.2 for some deterministic term $o(\delta_n)$,

$$
\text{(B.4)} \qquad \mathrm{P}_n\Big(\Big|\sup_{v\in\overline{V}_n} Z_n^\star(v) - \sup_{v\in\overline{V}_n} \bar{Z}_n^*(v)\Big| > o(\delta_n)|\mathcal{D}_n\Big) = o_{\mathrm{P}_n}(1/\ell_n),
$$

which implies by Lemma 11 that for some $\varepsilon_n \searrow 0$, wp $\to 1$, for all $p \in (\varepsilon_n, 1-\varepsilon_n)$,

$$
\text{(B.5)} \qquad k_{n,\overline{V}_n}(p) \le \kappa_{n,\overline{V}_n}(p+\varepsilon_n) + o(\delta_n),
$$

where the terms $o(\delta_n)$ are different in different places. By Condition C.3, for any fixed $p \in (0,1)$,

$$
\kappa_{\overline{V}_n}(p+\varepsilon_n) \le a_n(\overline{V}_n) + c(p+\varepsilon_n)/a_n(\overline{V}_n) = a_n(\overline{V}_n) + O(1)/a_n(\overline{V}_n).
$$

Thus, combining the inequalities above and $o(\delta_n) = o(\bar{a}_n^{-1}) = o(a_n^{-1}(\overline{V}_n))$ by Condition C.2, wp $\to 1$,

$$
k_{n,\hat{V}_n}(p) \le a_n(\overline{V}_n) + O(1)/a_n(\overline{V}_n).
$$

Now Parts (ii) and (iii) follow as in the proof of Parts (ii) and (iii) of Theorem 1, using this bound on the simulated critical value instead of the bound on the analytical critical value. *Q.E.D.*

B.5. *Proof of Lemma 3—Concentration on V_0*

By Conditions S and V, wp $\to$ 1,

$$\left|\sup_{v\in V_n} Z_n^*(v) - \sup_{v\in V_0} Z_n^*(v)\right| \le \sup_{\|v-v'\|\le r_n} \left|Z_n^*(v) - Z_n^*(v')\right| = o_{\mathrm{P}_n}\left(\bar{a}_n^{-1}\right). \tag{B.6}$$

Conclude similarly to the proof of Lemma 1, using anti-concentration Corollary 2, that

$$\begin{aligned}
\mathrm{P}_n\left(\sup_{v\in\mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} \le x\right) &\ge \mathrm{P}_n\left(\sup_{v\in V_0} Z_n^*(v) + o(\bar{a}_n^{-1}) \le x\right) - o(1) \\
&\ge \mathrm{P}_n\left(\sup_{v\in V_0} Z_n^*(v) \le x\right) - o(1).
\end{aligned}$$

This gives a lower bound. Similarly, using Conditions C.3 and C.4 and anti-concentration Corollary 2,

$$\begin{aligned}
\mathrm{P}_n\left(\sup_{v\in\mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} \le x\right) &\le \mathrm{P}_n\left(\sup_{v\in V_0} Z_n(v)\frac{\sigma_n(v)}{s_n(v)} \le x\right) \\
&\le \mathrm{P}_n\left(\sup_{v\in V_0} Z_n^*(v) - o(\delta_n) \le x\right) + o(1) \\
&\le \mathrm{P}_n\left(\sup_{v\in V_0} Z_n^*(v) \le x\right) + o(1),
\end{aligned}$$

where the $o(\cdot)$ terms above are different in different places and the first inequality follows from

$$\sup_{v\in\mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} \ge \sup_{v\in V_0} \frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} = \sup_{v\in V_0} Z_n(v)\frac{\sigma_n(v)}{s_n(v)}.$$

This gives the upper bound. *Q.E.D.*

B.6. *Proof of Theorem 3—When Simulation Inference Becomes Sharp*

Part (i)—Size. By Lemma 2, wp $\to$ 1, $\hat{V}_n \subseteq \overline{V}_n$, so that $k_{n,\hat{V}_n}(p) \le k_{n,\overline{V}_n}(p)$ wp $\to$ 1. So let us compare critical values

$$k_{n,\overline{V}_n}(p) = Q_p\left(\sup_{v\in\overline{V}_n} Z_n^\star(v)\big|\mathcal{D}_n\right) \quad \text{and} \quad \kappa_{n,V_0}(p) = Q_p\left(\sup_{v\in V_0} \bar{Z}_n^*(v)\right).$$

The former is data dependent, while the latter is deterministic. Recall that by Condition C.2, wp $\to 1$, we have (B.4). By Condition V, $d_H(\overline{V}_n, V_0) \leq r_n$, and so by Condition S, we have for some $o(\bar{a}_n^{-1})$,

$$\mathrm{P}_n\Big(\Big|\sup_{v\in\overline{V}_n} \bar{Z}_n^*(v) - \sup_{v\in V_0} \bar{Z}_n^*(v)\Big| > o\big(\bar{a}_n^{-1}\big)\big|\mathcal{D}_n\Big) = o_{\mathrm{P}_n}(1).$$

Combining (B.4) and this relation, we obtain that for some $o(\bar{a}_n^{-1})$ term,

$$\mathrm{P}_n\Big(\Big|\sup_{v\in\overline{V}_n} Z_n^\star(v) - \sup_{v\in V_0} \bar{Z}_n^*(v)\Big| > o\big(\bar{a}_n^{-1}\big)\big|\mathcal{D}_n\Big) = o_{\mathrm{P}_n}(1).$$

This implies by Lemma 11 that for some $\varepsilon_n \searrow 0$, and any $p \in (\varepsilon_n, 1-\varepsilon_n)$, wp $\to 1$,

$$k_{n,\hat{V}_n}(p) \leq k_{n,\overline{V}_n}(p) \leq \kappa_{n,V_0}(p+\varepsilon_n) + o\big(\bar{a}_n^{-1}\big). \tag{B.7}$$

Hence, for any fixed p,

$$\begin{aligned}
&\mathrm{P}_n\big(\theta_{n0} \leq \hat{\theta}_{n0}(p)\big)\\
&\quad = \mathrm{P}_n\bigg(\sup_{v\in\mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} \leq k_{n,\hat{V}_n}(p)\bigg)\\
&\quad \leq_{(1)} \mathrm{P}_n\bigg(\sup_{v\in\mathcal{V}} \frac{\theta_{n0} - \hat{\theta}_n(v)}{s_n(v)} \leq \kappa_{n,V_0}(p+\varepsilon_n) + o\big(\bar{a}_n^{-1}\big)\bigg) + o(1)\\
&\quad \leq_{(2)} \mathrm{P}_n\Big(\sup_{v\in V_0} Z_n^*(v) \leq \kappa_{n,V_0}(p+\varepsilon_n) + o\big(\bar{a}_n^{-1}\big)\Big) + o(1)\\
&\quad \leq_{(3)} p + \varepsilon_n + o(1) = p + o(1),
\end{aligned}$$

where inequality (1) is by the quantile comparison (B.7), inequality (2) is by Lemma 3, and inequality (3) is by anti-concentration Corollary 2. Combining this with the lower bound of Theorem 2, we have the result.

Parts (ii) and (iii)—Estimation Risk and Power. We have that by Condition C.3,

$$\kappa_{n,V_0}(p+\varepsilon_n) \leq a_n(V_0) + c(p+\varepsilon_n)/a_n(V_0) = a_n(V_0) + O(1)/a_n(V_0).$$

Hence combining this with equation (B.7), we have, wp $\to 1$,

$$k_{n,\hat{V}_n}(p) \leq a_n(V_0) + O(1)/a_n(V_0) + o\big(\bar{a}_n^{-1}\big) = a_n(V_0) + O(1)/a_n(V_0).$$

Then Parts (ii) and (iii) follow identically to the proof of Parts (ii) and (iii) of Theorem 1, using this bound on the simulated critical value instead of the bound on the analytical critical value. *Q.E.D.*

APPENDIX C: PROOFS FOR SECTION 4

C.1. *Tools and Auxiliary Lemmas*

We shall heavily rely on the Talagrand–Samorodnitsky (TS) inequality, which was obtained by Talagrand and sharpens earlier results by Samorodnitsky. Here it is restated from van der Vaart and Wellner (1996, Proposition A.2.7, p. 442).

Talagrand–Samorodnitsky Inequality

Let X be a separable zero-mean Gaussian process indexed by a set T. Suppose that for some $\Gamma > \sigma(X) = \sup_{t\in T} \sigma(X_t)$, $0 < \epsilon_0 \leq \sigma(X)$,

$$N(\varepsilon, T, \rho) \leq \left(\frac{\Gamma}{\varepsilon}\right)^{\nu}, \quad \text{for} \quad 0 < \varepsilon < \epsilon_0,$$

where $N(\varepsilon, T, \rho)$ is the covering number of T by ε-balls with respect to (w.r.t.) the standard deviation metric $\rho(t, t') = \sigma(X_t - X_{t'})$. Then there exists a universal constant D such that for every $\lambda \geq \sigma^2(X)(1 + \sqrt{\nu})/\epsilon_0$, we have

$$\text{(C.1)} \qquad \mathrm{P}\Big(\sup_{t\in T} X_t > \lambda\Big) \leq \left(\frac{D\Gamma\lambda}{\sqrt{\nu}\sigma^2(X)}\right)^{v} \big(1 - \Phi\big(\lambda/\sigma(X)\big)\big),$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function.

The following lemma is an application of this inequality that we use.

LEMMA 12—Concentration Inequality via Talagrand–Samorodnitsky: *Let Z_n be a separable zero-mean Gaussian process indexed by a set V such that $\sup_{v\in \mathsf{V}} \sigma(Z_n(v)) = 1$. Suppose that for some $\Gamma_n(\mathsf{V}) > 1$ and $d \geq 1$,*

$$N(\varepsilon, \mathsf{V}, \rho) \leq \left(\frac{\Gamma_n(\mathsf{V})}{\varepsilon}\right)^{d} \quad \textit{for} \quad 0 < \varepsilon < 1,$$

where $N(\varepsilon, \mathsf{V}, \rho)$ is the covering number of V by ε-balls w.r.t. the standard deviation metric $\rho(v, v') = \sigma(Z_n(v) - Z_n(v'))$. Then for

$$a_n(\mathsf{V}) = \big(2\sqrt{\log L_n(\mathsf{V})}\big) \vee (1 + \sqrt{d}), \quad L_n(\mathsf{V}) := C_n'\left(\frac{\Gamma_n(\mathsf{V})}{\sqrt{d}}\right)^{d},$$

where for D denoting Talagrand's constant in (C.1) *and C_n' such that*

$$C_n' \geq D^d C_d \frac{1}{\sqrt{2\pi}}, \quad C_d := \max_{\lambda\geq 0} \lambda^{d-1} e^{-\lambda^2/4},$$

we have for $z \geq 0$,

$$\mathrm{P}\Big(a_n(\mathsf{V})\Big(\sup_{v\in\mathsf{V}} Z_n(v) - a_n(\mathsf{V})\Big) > z\Big)$$
$$\leq \exp\left(-\frac{z}{2} - \frac{z^2}{4a_n^2(\mathsf{V})}\right) \leq \exp\left(-\frac{z}{2}\right).$$

PROOF: We apply the TS inequality by setting $t = v$, $X = Z$, $\sigma(X) = 1$, $\epsilon_0 = 1$, $\nu = d$, with $\lambda \geq (1 + \sqrt{d})$, so that

$$\mathrm{P}\Big(\sup_{v\in\mathsf{V}} Z_n(v) > \lambda\Big) \leq \left(\frac{D\Gamma_n(\mathsf{V})\lambda}{\sqrt{d}}\right)^d \big(1 - \Phi(\lambda)\big)$$
$$\leq \left(\frac{D\Gamma_n(\mathsf{V})\lambda}{\sqrt{d}}\right)^d \frac{1}{\sqrt{2\pi}}\frac{1}{\lambda} e^{-\lambda^2/2} \leq L_n(\mathsf{V})e^{-\lambda^2/4}.$$

Setting, for $z \geq 0$, $\lambda = \frac{z}{a_n(\mathsf{V})} + a_n(\mathsf{V}) \geq (1 + \sqrt{d})$, we obtain

$$L_n(\mathsf{V}) \exp\left(-\frac{\lambda^2}{4}\right) \leq \exp\left(-\frac{z}{2} - \frac{z^2}{4a_n^2(\mathsf{V})}\right). \qquad Q.E.D.$$

The following lemma is an immediate consequence of Corollary 2.2.8 of van der Vaart and Wellner (1996).

LEMMA 13—Maximal Inequality for a Gaussian Process: *Let X be a separable zero-mean Gaussian process indexed by a set T. Then for every $\delta > 0$,*

$$E \sup_{\rho(s,t)\leq\delta} |X_s - X_t| \lesssim \int_0^\delta \sqrt{\log N(\varepsilon, T, \rho)}\, d\varepsilon,$$
$$E \sup_{t\in T} |X_t| \lesssim \sigma(X) + \int_0^{2\sigma(X)} \sqrt{\log N(\varepsilon, T, \rho)}\, d\varepsilon,$$

where $\sigma(X) = \sup_{t\in T}\sigma(X_t)$ and $N(\varepsilon, T, \rho)$ is the covering number of T with respect to the semimetric $\rho(s, t) = \sigma(X_s - X_t)$.

PROOF: The first conclusion follows from Corollary 2.2.8 of van der Vaart and Wellner (1996), since covering and packing numbers are related by $N(\varepsilon, T, \rho) \leq D(\varepsilon, T, \rho) \leq N(\varepsilon/2, T, \rho)$. The second conclusion follows from the special case of the first conclusion: for any $t_0 \in T$, $E\sup_{t\in T}|X_t| \lesssim E|X_{t_0}| + \int_0^{\mathrm{diam}(T)} \sqrt{\log N(\varepsilon, T, \rho)}\, d\varepsilon \leq \sigma(X) + \int_0^{2\sigma(X)} \sqrt{\log N(\varepsilon, T, \rho)}\, d\varepsilon$. Q.E.D.

C.2. *Proof of Lemma 5*

Step 1—Verification of Condition C.1. This condition holds by inspection, in view of the continuity of $v \mapsto p_n(v)$ and by Ω_n and $\hat{\Omega}_n$ being positive definite.

Step 2—Verification of Condition C.2. Set $\delta_n = 1/\log n$. Condition NS(i) directly assumes Condition C.2(a).

To show Condition C.2(b), we employ the maximal inequality stated in Lemma 13. Set $X_t = Z_n^*(v) - Z_n^\star(v)$, $t = v$, and $T = \mathcal{V}$, and note that for some absolute constant C, conditional on $\mathcal{D}_n$,

$$N(\varepsilon, T, \rho) \le \left(\frac{1 + C Y_n \operatorname{diam}(T)}{\varepsilon} \right)^d, \quad 0 < \varepsilon < 1,$$

since $\sigma(X_t - X_{t'}) \lesssim Y_n \|t - t'\|$, $T \subset \mathbb{R}^d$, where Y_n is an upper bound on the Lipschitz constant of the function

$$v \mapsto \frac{p_n(v)'\Omega_n^{1/2}}{\|p_n(v)'\Omega_n^{1/2}\|} - \frac{p_n(v)'\hat{\Omega}_n^{1/2}}{\|p_n(v)'\hat{\Omega}_n^{1/2}\|},$$

where $\operatorname{diam}(T)$ is the diameter of set T under the Euclidean metric. Using inequality (E.6) in the Supplemental Appendix, we can bound

$$Y_n \le 2L_n \frac{\lambda_{\max}(\Omega_n^{1/2})}{\lambda_{\min}(\Omega_n^{1/2})} + 2L_n \frac{\lambda_{\max}(\hat{\Omega}_n^{1/2})}{\lambda_{\min}(\hat{\Omega}_n^{1/2})} = O_{\mathrm{P}_n}(L_n),$$

where L_n is the constant defined in Condition NS(i) and by assumption $\log L_n \lesssim \log n$. Here we use the fact the eigenvalues of Ω_n and $\hat{\Omega}_n$ are bounded away from zero and from above by Condition NS(i) and Condition NS(ii). Therefore, $\log N(\varepsilon, T, \rho) \lesssim \log n + \log(1/\varepsilon)$.

Using (E.6) again gives

$$\begin{aligned}
\sigma(X) &\lesssim \sup_{v \in \mathcal{V}} \left\| \frac{p_n(v)'\Omega_n^{1/2}}{\|p_n(v)'\Omega_n^{1/2}\|} - \frac{p_n(v)'\hat{\Omega}_n^{1/2}}{\|p_n(v)'\hat{\Omega}_n^{1/2}\|} \right\| \\
&\le \sup_{v \in \mathcal{V}} 2 \frac{\|p_n(v)'(\hat{\Omega}_n^{1/2} - \Omega_n^{1/2})\|}{\|p_n(v)'\Omega_n^{1/2}\|} \\
&\le \sup_{v \in \mathcal{V}} 2 \frac{\|p_n(v)'\Omega_n^{1/2}(\Omega_n^{-1/2}\hat{\Omega}_n^{1/2} - I)\|}{\|p_n(v)'\Omega_n^{1/2}\|} \\
&\le \|\Omega_n^{-1/2}\hat{\Omega}_n^{1/2} - I\| \le \|\Omega_n^{-1/2}\| \|\hat{\Omega}_n^{1/2} - \Omega_n^{1/2}\| = O_{\mathrm{P}_n}\big(n^{-b}\big)
\end{aligned}$$

for some constant $b > 0$, where we have used that the eigenvalues of Ω_n and $\hat{\Omega}_n$ are bounded away from zero and from above under Conditions NS(i) and NS(ii), and the assumption $\|\hat{\Omega}_n - \Omega_n\| = O_{\mathrm{P}_n}(n^{-b})$. Hence

$$E\Big(\sup_{t\in T}|X_t||\mathcal{D}_n\Big) \lesssim \sigma(X) + \int_0^{2\sigma(X)} \sqrt{\log(n/\varepsilon)}\, d\varepsilon = O_{\mathrm{P}_n}\big(n^{-b}\sqrt{\log n}\big).$$

Hence for each $C > 0$,

$$\begin{aligned}
&\mathrm{P}_n\Big(\sup_{v\in\mathcal{V}}\big|Z_n^*(v) - Z_n^\star(v)\big| > C\delta_n\big|\mathcal{D}_n\Big) \\
&\quad\lesssim \frac{1}{C\delta_n} O_{\mathrm{P}_n}\big(n^{-b}\sqrt{\log n}\big) = o_{\mathrm{P}_n}(1/\ell_n),
\end{aligned}$$

which verifies Condition C.2(b).

Step 3—Verification of Condition C.3. We shall employ Lemma 12, which has the required notation in place. We only need to compute an upper bound on the covering numbers $N(\varepsilon, \mathsf{V}, \rho)$ for the process Z_n^*. We have that

$$\begin{aligned}
\sigma\big(Z_n^*(v) - Z_n^*(\tilde{v})\big) &\le \left\| \frac{p_n(v)'\Omega_n^{1/2}}{\|p_n(v)'\Omega_n^{1/2}\|} - \frac{p_n(\tilde{v})'\Omega_n^{1/2}}{\|p_n(\tilde{v})'\Omega_n^{1/2}\|} \right\| \\
&\le 2\left\| \frac{(p_n(v) - p_n(\tilde{v}))'\Omega_n^{1/2}}{\|p_n(v)'\Omega_n^{1/2}\|} \right\| \\
&\le 2L_n \frac{\lambda_{\max}(\Omega_n^{1/2})}{\lambda_{\min}(\Omega_n^{1/2})}\|v - \tilde{v}\| \le CL_n\|v - \tilde{v}\|,
\end{aligned}$$

where C is some constant that does not depend on n, by the eigenvalues of Ω_n bounded away from zero and from above. Hence it follows that

$$N(\varepsilon, \mathsf{V}, \rho) \le \left(\frac{1 + CL_n \operatorname{diam}(\mathsf{V})}{\varepsilon}\right)^d, \quad 0 < \varepsilon < 1,$$

where the diameter of V is measured by the Euclidean metric. Condition C.3 now follows by Lemma 12, with $a_n(\mathsf{V}) = (2\sqrt{\log L_n(\mathsf{V})}) \vee (1 + \sqrt{d})$ and $L_n(\mathsf{V}) = C'(1 + CL_n \operatorname{diam}(\mathsf{V}))^d$, where C' is some positive constant.

Step 4—Verification of Condition C.4. Under Condition NS, we have that

$$a_n(\mathsf{V}) \le \bar{a}_n := a_n(\mathcal{V}) \lesssim \sqrt{\log \ell_n + \log n} \lesssim \sqrt{\log n},$$

so that Condition C.4(a) follows if $\sqrt{\log n}\sqrt{\zeta_n^2/n} \to 0$.

To verify Condition C.4(b), note that uniformly in $v \in \mathcal{V}$,

$$
\begin{aligned}
&\left| \frac{\| p_n(v)'\hat{\Omega}_n^{1/2}\|}{\| p_n(v)'\Omega_n^{1/2}\|} - 1 \right| \\
&\quad \le \left| \frac{\| p_n(v)'\hat{\Omega}_n^{1/2}\| - \| p_n(v)'\Omega_n^{1/2}\|}{\| p_n(v)'\Omega_n^{1/2}\|} \right| \\
&\quad \le \frac{\| p_n(v)'(\hat{\Omega}_n^{1/2} - \Omega_n^{1/2})\|}{\| p_n(v)'\Omega_n^{1/2}\|} \le \frac{\| p_n(v)'\Omega^{1/2}(\Omega_n^{-1/2}\hat{\Omega}_n^{1/2} - I)\|}{\| p_n(v)'\Omega_n^{1/2}\|} \\
&\quad \le \|\Omega_n^{-1/2}\hat{\Omega}_n^{1/2} - I\| \le \|\Omega_n^{-1/2}\| \|\hat{\Omega}_n^{1/2} - \Omega_n^{1/2}\| = o_{\mathrm{P}_n}(\delta_n/\bar{a}_n)
\end{aligned}
$$

by $\|\hat{\Omega}^{1/2} - \Omega_n^{1/2}\| = O_{\mathrm{P}_n}(n^{-b})$ and $\|\Omega_n^{-1/2}\|$ bounded, both implied by the assumptions. *Q.E.D.*

C.3. *Proof of Lemma 6*

To show claim (i), we need to establish that for $\varphi_n = o(1) \cdot (\frac{1}{L_n\sqrt{\log n}})$, with any $o(1)$ term, we have that $\sup_{\|v-\tilde{v}\|\le\varphi_n} |Z_n^*(v) - Z_n^*(\tilde{v})| = o_{\mathrm{P}_n}(1)$.

Consider the stochastic process $X = \{Z_n^*(v), v \in \mathcal{V}\}$. We shall use the standard maximal inequality stated in Lemma 13. From the proof of Lemma 5, we have $\sigma(Z_n^*(v) - Z_n^*(\tilde{v})) \le CL_n\|v - \tilde{v}\|$, where C is some constant that does not depend on n, and $\log N(\varepsilon, \mathsf{V}, \rho) \lesssim \log n + \log(1/\varepsilon)$. Since $\|v - \tilde{v}\| \le \varphi_n \Longrightarrow \sigma(Z_n^*(v) - Z_n^*(\tilde{v})) \le C\frac{o(1)}{\sqrt{\log n}}$, we have

$$
\begin{aligned}
E \sup_{\|v-\tilde{v}\|\le\varphi_n} |X_v - X_{\tilde{v}}| &\lesssim \int_0^{Co(1)/\sqrt{\log n}} \sqrt{\log(n/\varepsilon)}\, d\varepsilon \\
&\lesssim \frac{o(1)}{\sqrt{\log n}}\sqrt{\log n} = o(1).
\end{aligned}
$$

Hence the conclusion follows from Markov's inequality.

Under Condition V, by Lemma 2, $r_n \lesssim (\sqrt{\log n \frac{\zeta_n^2}{n}})^{1/\rho_n} c_n^{-1}$, so $r_n = o(\varphi_n)$ if

$$
\left(\sqrt{\log n \frac{\zeta_n^2}{n}}\right)^{1/\rho_n} c_n^{-1} = o\left(\frac{1}{L_n\sqrt{\log n}}\right). \tag{C.2}
$$

Thus, Condition S holds. The remainder of the lemma follows by direct calculation. *Q.E.D.*

APPENDIX D: PROOFS FOR SECTION 5

D.1. *Proof of Theorem 7 and Corollary 1*

The first step of our proof uses Yurinskii's (1977) coupling. For completeness we now state the formal result from Pollard (2002, p. 244).

Yurinskii's Coupling

Consider a sufficiently rich probability space $(A, \mathcal{A}, \mathrm{P})$. Let $\xi_1, \dots, \xi_n$ be independent K_n-vectors with $E\xi_i = 0$ for each i and let $\Delta := \sum_i E\|\xi_i\|^3$ be finite. Let $S = \xi_1 + \cdots + \xi_n$. For each $\delta > 0$, there exists a random vector T with $N(0, \mathrm{var}(S))$ distribution such that

$$\mathrm{P}\{\|S - T\| > 3\delta\} \leq C_0 B\left(1 + \frac{|\log(1/B)|}{K_n}\right), \quad \text{where} \quad B := \Delta K_n \delta^{-3}$$

for some universal constant C_0.

The proof has two steps: in the first, we couple the estimator $\sqrt{n}(\hat{\beta}_n - \beta_n)$ with the normal vector; in the second, we establish the strong approximation.

Step 1. To apply the coupling, consider

$$\sum_{i=1}^{n} \xi_i, \qquad \xi_i = u_{i,n}/\sqrt{n} \sim (0, I_{K_n}/n).$$

Then we have that $\sum_{i=1}^{n} E\|\xi_i\|^3 = \Delta_n$. Therefore, by Yurinskii's coupling,

$$\mathrm{P}_n\left\{\left\|\sum_{i=1}^{n} \xi_i - \mathcal{N}_n\right\| \geq 3\delta_n\right\} \to 0 \quad \text{if} \quad K_n \Delta_n/\delta_n^3 \to 0.$$

Combining this with the assumption on the linearization error r_n, we obtain

$$\begin{aligned}
&\left\|\Omega_n^{-1/2}\sqrt{n}(\hat{\beta}_n - \beta_n) - \mathcal{N}_n\right\| \\
&\quad \leq \left\|\sum_{i=1}^{n} \xi_i - \mathcal{N}_n\right\| + \left\|\Omega_n^{-1/2}\sqrt{n}(\hat{\beta}_n - \beta_n) - \sum_{i=1}^{n} \xi_i\right\| \\
&\quad = o_{\mathrm{P}_n}(\delta_n) + r_n = o_{\mathrm{P}_n}(\delta_n).
\end{aligned}$$

Step 2. Using the result of Step 1 and that

$$\frac{\sqrt{n}\, p_n(v)'(\hat{\beta}_n - \beta_n)}{\|g_n(v)\|} = \frac{\sqrt{n} g_n(v)' \Omega_n^{-1/2}(\hat{\beta}_n - \beta_n)}{\|g_n(v)\|},$$

we conclude that

$$\text{(D.1)} \qquad \left|S_n(v)\right| := \left|\frac{\sqrt{n}g_n(v)'\Omega_n^{-1/2}(\hat{\beta}_n - \beta_n)}{\|g_n(v)\|} - \frac{g_n(v)'\mathcal{N}_n}{\|g_n(v)\|}\right|$$
$$\leq \left\|\sqrt{n}\Omega_n^{-1/2}(\hat{\beta}_n - \beta_n) - \mathcal{N}_n\right\| = o_{\mathrm{P}_n}(\delta_n)$$

uniformly in $v \in \mathcal{V}$. Finally,

$$\sup_{v\in\mathcal{V}}\left|\frac{\sqrt{n}(\hat{\theta}_n(v) - \theta_n(v))}{\|g_n(v)\|} - \frac{g_n(v)'\mathcal{N}_n}{\|g_n(v)\|}\right|$$
$$\leq \sup_{v\in\mathcal{V}}\left|\frac{\sqrt{n}(\hat{\theta}_n(v) - \theta_n(v))}{\|g_n(v)\|} - \frac{\sqrt{n}g_n(v)'\Omega_n^{-1/2}(\hat{\beta}_n - \beta_n)}{\|g_n(v)\|}\right|$$
$$+ \sup_{v\in\mathcal{V}}\left|\frac{\sqrt{n}g_n(v)'\Omega_n^{-1/2}(\hat{\beta}_n - \beta_n)}{\|g_n(v)\|} - \frac{g_n(v)'\mathcal{N}_n}{\|g_n(v)\|}\right|$$
$$= \sup_{v\in\mathcal{V}}\left|\sqrt{n}A_n(v)/\left\|g_n(v)\right\|\right| + \sup_{v\in\mathcal{V}}\left|S_n(v)\right| = o(\delta_n) + o_{\mathrm{P}_n}(\delta_n),$$

using the assumption on the approximation error $A_n(v) = \theta(v) - p_n(v)'\beta_n$ and (D.1). This proves the theorem.

Step 3. To show the corollary note that

$$E_{\mathrm{P}_n}\|u_{i,n}\|^3 \leq \|\Omega_n^{-1/2}Q_n^{-1}\|^3 \cdot E_{\mathrm{P}_n}\left\|p_n(V_i)\epsilon_i\right\|^3 \lesssim \tau_n^3 K_n^{3/2} C_n,$$

using the boundedness assumptions stated in the corollary.

REFERENCES

ANDREWS, D. W. K. (1991): "Asymptotic Normality of Series Estimators for Nonparametric and Semiparametric Regression Models," *Econometrica*, 59 (2), 307–345. [697,707]

ANDREWS, D. W. K., AND P. GUGGENBERGER (2009): "Validity of Subsampling and Plug-In Asymptotic Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669–709. [670]

ANDREWS, D. W. K., AND S. HAN (2009): "Invalidity of the Bootstrap and m Out of n Bootstrap for Interval Endpoints Defined by Moment Inequalities," *Econometrics Journal*, 12 (s1), S172–S199. [669]

ANDREWS, D. W. K., AND P. JIA (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80 (6), 2805–2826. [670]

ANDREWS, D. W. K., AND X. SHI (2011): "Nonparametric Inference Based on Conditional Moment Inequalities," Discussion Paper 1840, Cowles Foundation. [671]

——— (2013): "Inference for Parameters Defined by Conditional Moment Inequalities," *Econometrica* (forthcoming). [668-670,672]

ANDREWS, D. W. K., AND G. SOARES (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78 (1), 119–157. [670]

ARMSTRONG, T. B. (2011a): "Asymptotically Exact Inference in Conditional Moment Inequality Models," Working Paper, Stanford University. [673]

——— (2011b): "Weighted KS Statistics for Inference on Conditional Moment Inequalities," Working Paper, Stanford University. [672,673]

BELLONI, A., V. CHERNOZHUKOV, AND I. FERNANDEZ-VAL (2011): "Conditional Quantile Processes Under Increasing Dimensions," Working Paper CWP19/11, CeMMAP. [700,707]

BELLONI, A., V. CHERNOZHUKOV, AND K. KATO (2010): "(Some) New Asymptotic Theory for Linear Series Estimators," Working Paper, MIT. [701]

BERESTEANU, A., AND F. MOLINARI (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76 (4), 763–814. [670]

BERRY, S. T., AND E. TAMER (2007): "Identification in Models of Oligopoly Entry," in *Advances in Econometrics, Ninth World Congress*, Vol. 2, ed. by R. Blundell, W. Newey, and T. Persson. Cambridge: Cambridge University Press, 46–85. [676]

BICKEL, P. J., AND M. ROSENBLATT (1973): "On Some Global Measures of the Deviations of Density Function Estimates," *The Annals of Statistics*, 1 (6), 1071–1095. [669]

BIERENS, H. (1982): "Consistent Model Specification Tests," *Journal of Econometrics*, 20, 105–134. [670,672]

BLUNDELL, R., A. GOSLING, H. ICHIMURA, AND C. MEGHIR (2007): "Changes in the Distribution of Male and Female Wages Accounting for Unemployment Composition Using Bounds," *Econometrica*, 75 (2), 323–363. [668,676]

BUGNI, F. (2010): "Bootstrap Inference for Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78 (2), 735–753. [669,670]

CANAY, I. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156 (2), 408–425. [669,670]

CARNEIRO, P., AND S. LEE (2009): "Estimating Distributions of Potential Outcomes Using Local Instrumental Variables With an Application to Changes in College Enrollment and Wage Inequality," *Journal of Econometrics*, 149, 191–208. [678]

CHEN, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models (Chapter 76)," in *Handbook of Econometrics*,Vol. 6. Amsterdam: Elsevier, 5549–5632. [707]

CHERNOZHUKOV, V., CHETVERIKOV, D., AND K. KATO (2011): "Anti-Concentration and Honest Adaptive Confidence Bands," Working Paper, Hiroshima University. [686,719]

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75 (5), 1243–1284. [668,670]

CHERNOZHUKOV, V., S. LEE, AND A. ROSEN (2009): "Intersection Bounds: Estimation and Inference," Working Paper CWP19/09, CeMMAP. [678,679,698,705,708,712]

——— (2013): "Supplement to 'Intersection Bounds: Estimation and Inference'," *Econometrica Supplementary Material*, 81, http://www.econometricsociety.org/ecta/Supmat/8718_proofs.pdf; http://www.econometricsociety.org/ecta/Supmat/8718_data_and_programs.zip. [672]

CHERNOZHUKOV, V., R. RIGOBON, AND T. STOKER (2010): "Set Identification and Sensitivity Analysis With Tobin Regressors," *Quantitative Economics*, 1 (2), 255–277. [668]

CHESHER, A. D. (2010): "Instrumental Variable Models for Discrete Outcomes," *Econometrica*, 78 (2), 575–601. [668]

CHETVERIKOV, D. (2011): "Adaptive Rate Optimal Test of Conditional Moment Inequalities," Working Paper, MIT. [673]

——— (2012): "Testing Regression Monotonicity in Econometric Models," Working Paper, MIT. [669]

DUDLEY, R., AND W. PHILIPP (1983): "Invariance Principles for Sums of Banach Space Valued Random Elements and Empirical Processes," *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 62, 509–552. [718]

FAN, J., AND I. GIJBELS (1996): *Local Polynomial Modelling and Its Applications*. London, U.K.: Chapman & Hall. [712]

FAN, Y., AND S. PARK (2011): "Bias-Correction and Confidence Sets Under IV and MIV Assumptions," Working Paper, Vanderbilt University. [671]

GALICHON, A., AND M. HENRY (2009): "A Test of Non-Identifying Restrictions and Confidence Regions for Partially Identified Parameters," *Journal of Econometrics*, 159 (2), 186–196. [670]

GONZALEZ, L. (2005): "Nonparametric Bounds on the Returns to Language Skills," *Journal of Applied Econometrics*, 20, 771–795. [668]

HAILE, P. A., AND E. TAMER (2003): "Inference With an Incomplete Model of English Auctions," *Journal of Political Economy*, 111 (1), 1–51. [668,669]

HÄRDLE, W., AND E. MAMMEN (1993): "Comparing Nonparametric versus Parametric Regression Fits," *The Annals of Statistics*, 21 (4), 1926–1947. [670,672]

HECKMAN, J. J., AND E. J. VYTLACIL (1999): "Local Instrumental Variables and Latent Variable Models for Identifying and Bounding Treatment Effects," *Proceedings of the National Academy of Sciences*, 96, 4730–4734. [668]

HECKMAN, J. J., J. SMITH, AND N. CLEMENTS (1997): "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64, 487–537. [668]

HIRANO, K., AND J. R. PORTER (2012): "Impossibility Results for Nondifferentiable Functionals," *Econometrica*, 80 (4), 1769–1790. [670]

HOROWITZ, J., AND V. G. SPOKOINY (2001): "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model Against a Nonparametric Alternative," *Econometrica*, 69 (3), 599–631. [672]

IMBENS, G., AND C. F. MANSKI (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72 (6), 1845–1857. [678]

KIM, K. I. (2009): "Set Estimation and Inference With Models Characterized by Conditional Moment Inequalities," Working Paper, University of Minnesota. [669-671]

KOENKER, R. (2005): *Quantile Regression*. Cambridge: Cambridge University Press. [695]

KONG, E., O. LINTON, AND Y. XIA (2010): "Uniform Bahadur Representation for Local Polynomial Estimates of m-Regression and Its Application to the Additive Model," *Econometric Theory*, 26, 1529–1564. [704]

KREIDER, B., AND J. PEPPER (2007): "Disability and Employment: Reevaluating the Evidence in Light of Reporting Errors," *Journal of the American Statistical Association*, 102 (478), 432–441. [668,669]

LEE, S., AND R. WILKE (2009): "Reform of Unemployment Compensation in Germany: A Nonparametric Bounds Analysis Using Register Data," *Journal of Business & Economic Statistics*, 27 (2), 193–205. [668]

LEE, S., K. SONG, AND Y.-J. WHANG (2013): "Testing Functional Inequalities," *Journal of Econometrics*, 172, 14–32. [672]

MANSKI, C. F. (1989): "Anatomy of the Selection Problem," *The Journal of Human Resources*, 24 (3), 343–360. [675]

——— (1990): "Nonparametric Bounds on Treatment Effects," *American Economic Review*, 80 (2), 319–323. [668]

——— (1997): "Monotone Treatment Response," *Econometrica*, 65 (6), 1311–1334. [675]

——— (2003): *Partial Identification of Probability Distributions*. New York: Springer-Verlag. [668]

MANSKI, C. F., AND J. V. PEPPER (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68 (4), 997–1010. [668,675,678,679]

——— (2009): "More on Monotone Instrumental Variables," *Econometrics Journal*, 12 (s1), S200–S216. [678]

MASRY, E. (1996): "Multivariate Local Polynomial Regression for Time Series: Uniform Strong Consistency and Rates," *Journal of Time Series Analysis*, 17, 571–599. [704]

MASSART, P. (1989): "Strong Approximation for Multivariate Empirical and Related Processes, via KMT Constructions," *The Annals of Probability*, 17 (1), 266–291. [704]

MENZEL, K. (2009): "Estimation and Inference With Many Weak Moment Inequalities," Working Paper, MIT. [670,671]

NEVO, A., AND A. M. ROSEN (2012): "Identification With Imperfect Instruments," *Review of Economics and Statistics*, 94 (3), 659–671. [668]

NEWEY, W. K. (1995): "Convergence Rates for Series Estimators," in *Statistical Methods of Economics and Quantitative Economics: Essays in Honor of C.R. Rao*, ed. by G. Maddalla, P. Phillips, and T. Srinavasan. Cambridge, MA: Blackwell, 254–275. [707]

——— (1997): "Convergence Rates and Asymptotic Normality for Series Estimators," *Journal of Econometrics*, 79, 147–168. [697,701-703,707]

NICOLETTI, C., F. FOLIANO, AND F. PERACCHI (2011): "Estimating Income Poverty in the Presence of Missing Data and Measurement Error Problems," *Journal of Business & Economic Statistics*, 29 (1), 61–72. [668]

PAKES, A., J. PORTER, K. HO, AND J. ISHII (2005): "The Method of Moments With Inequality Constraints," Working Paper, Harvard University. [676]

POLLARD, D. (2002): *A User's Guide to Measure Theoretic Probability*. Cambridge: Cambridge University Press. [732]

PONOMAREVA, M. (2010): "Inference in Models Defined by Conditional Moment Inequalities With Continuous Covariates," Working Paper, University of Western Ontario. [673]

POWELL, J. L. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25 (3), 303–325. [695]

RIO, E. (1994): "Local Invariance Principles and Their Application to Density Estimation," *Probability Theory and Related Fields*, 98, 21–45. [704]

ROMANO, J. P., AND A. M. SHAIKH (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 138 (9), 2786–2807. [670]

——— (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78 (1), 169–211. [670]

ROSEN, A. M. (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146, 107–117. [670]

SHAIKH, A. M., AND E. J. VYTLACIL (2011): "Partial Identification in Triangular Systems of Equations With Binary Dependent Variables," *Econometrica*, 79 (3), 949–955. [668]

SKOROHOD, A. (1956): "Limit Theorems for Stochastic Processes," *Theory of Probability and Its Applications*, 1, 261–290. [693]

STONE, C. J. (1982): "Optimal Global Rates of Convergence for Nonparametric Regression," *The Annals of Statistics*, 10 (4), 1040–1053. [672,690]

STOYE, J. (2009): "More on Confidence Regions for Partially Identified Parameters," *Econometrica*, 77 (4), 1299–1315. [678]

TSYBAKOV, A. B. (2009): *Introduction to Nonparametric Estimation*. New York: Springer. [672]

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. New York: Springer-Verlag. [693,727,728]

WHITE, H. (1984): *Asymptotic Theory for Econometricians*. New York: Academic Press. [695,696]

YURINSKII, V. (1977): "On the Error of the Gaussian Approximation for Convolutions," *Theory of Probability and Its Applications*, 22, 236–247. [697]

Dept. of Economics, Massachusetts Institute of Technology, Cambridge, MA 02139, U.S.A.; vchern@mit.edu; http://www.mit.edu/~vchern/,

Dept. of Economics, Seoul National University, Seoul 151-742, Korea; sokbae@gmail.com; https://sites.google.com/site/sokbae/,

and

Centre for Microdata Methods and Practice, Institute for Fiscal Studies, Dept. of Economics, University College London, Gower Street, London WC1E 6BT, U.K.; adam.rosen@ucl.ac.uk; http://www.homepages.ucl.ac.uk/~uctparo/.

Manuscript received July, 2009; final revision received September, 2012.