More on Confidence Intervals for Partially Identified Parameters

Jörg Stoye*

New York University

– preliminary and incomplete, not for general circulation –

February 14, 2007

Abstract

This note extends Imbens and Manski's (2004) analysis of confidence regions for partially identified parameters. That paper's final result implicitly assumes superefficient estimation of a nuisance parameter. This aspect appears to have gone unnoticed before, and it implies that the assumptions imposed are not, in general, consistent. I establish the result using internally consistent assumptions that furthermore weaken the superefficiency condition. I furthermore propose a confidence region that does not rely on superefficiency and that embeds a specification test for nonemptiness of the identified set.

1 Introduction

Analysis of partial identification, that is, of models where only bounds on parameters are identified, has become an active field of econometrics.[1] Within this field, attention has only recently turned to general treatments of estimation and inference. An important contribution in this direction has been made by Imbens and Manski (2004, IM henceforth). Their major innovation is to point out that in constructing confidence regions for partially identified parameters, one might be interested in coverage probabilities for the parameter rather than its "identified set." The intuitively most obvious, and previously used, confidence regions have nominal coverage probabilities defined for the latter, which means that they

*I am indebted to Adam Rosen, whose insightful comments helped to much improve this paper. Of course, all errors are mine. Address: Department of Economics, New York University, 269 Mercer Street, New York, NY 10003, j.stoye@nyu.edu.

[1] See Manski (2003) for a survey and Haile and Tamer (2003) as well as Honoré and Tamer (2006) for recent contributions.

are conservative with respect to the former. IM go on to propose a number of confidence regions designed to cover real-valued parameters that can be asymptotically concluded to lie in an interval.

This paper refines and extends IM's technical analysis, specifically their last result, a confidence interval that exhibits uniform coverage of partially identified parameters if the length of the identified interval is a nuisance parameter. IM's proof of coverage for that confidence set relies on a somewhat hard to understad assumption that turns out to amount to superefficiency of a certain estimator, and also to be internally consistent only in a boundary case. I weaken the assumption in two different ways. IM's confidence interval remains valid under the first of these, which retains superefficiency. The other weakening requires a quite different analysis, and I propose a confidence region that also embeds a specification test for the underlying model.

2 Imbens and Manski Reconsidered

2.1 Background

In this section, I recapitulate a simplified version of IM's setup, sketch their main idea, and highlight a difficulty. The object of interest is the real-valued parameter $\theta_0(P)$ of a probability distribution $P(X)$; P must lie in a set $\mathcal{P}$ that is characterized by ex ante constraints (maintained assumptions). The random variable X is not completely observable, so that θ_0 may not be identified. Assume, however, that the observable aspects of $P(X)$ identify bounds $\theta_l(P)$ and $\theta_u(P)$ s.t. $\theta_0 \in [\theta_l, \theta_u]$ a.s. See the aforecited references for examples. The interval $\Theta_0 \equiv [\theta_l, \theta_u]$ will also be called *identified set*. Let $\Delta(P) \equiv \theta_u - \theta_l$ denote its length; obviously, Δ is identified as well. Assume that estimators $\widehat{\theta}_l$, $\widehat{\theta}_u$, and $\widehat{\Delta}$ exist and are connected by the identity $\widehat{\Delta} \equiv \widehat{\theta}_u - \widehat{\theta}_l$.

Confidence regions for identified sets of this type are conventionally formed as

$$CI_\alpha = \left[\widehat{\theta}_l - C_N \widehat{\sigma}_l N^{-1/2}, \widehat{\theta}_u + C_N \widehat{\sigma}_u N^{-1/2}\right],$$

where $\widehat{\sigma}_l$ respectively $\widehat{\sigma}_u$ are standard errors for $\widehat{\theta}_l$ respectively $\widehat{\theta}_u$, and where C_N is chosen s.t.

$$\left[\widehat{\theta}_u - C_N \widehat{\sigma}_u N^{-1/2}, \widehat{\theta}_u + C_N \widehat{\sigma}_u N^{-1/2}\right]$$

would be an equal-tailed level α confidence interval for θ_u; e.g. $C_N = \Phi^{-1}(0.975) \approx 1.96$ for a 95%-confidence interval. Under regularity conditions, $\Pr(\Theta_0 \subseteq CI_\alpha) \to \alpha$; see, for example, Horowitz and Manski (2000). IM's contribution is motivated by the observations that (i) one might be interested in coverage of θ_0 rather than Θ_0, (ii) whenever $\Delta > 0$, then $\Pr(\theta_0 \in CI_\alpha) \to 1 - \frac{1-\alpha}{2}$. In words, a 90% C.I. for Θ_0 is a 95% C.I. for θ_0. The reason is that asymptotically, Θ_0 is large relative to sampling errors, so that noncoverage risk is effectively one-sided at its boundary and vanishes otherwise. One would, therefore, be tempted to construct a level α C.I. for θ as $CI_{1-2(1-\alpha)}$.[2]

Unfortunately, this intuition works pointwise (see lemma 2 in IM) but not uniformly over interesting specifications of $\mathcal{P}$. Specifically, $\Pr(\theta_0 \in CI_\alpha) = \alpha$ if $\Delta = 0$ and also $\Pr(\theta_0 \in CI_\alpha) \to \alpha$ along any local parameter sequence that fulfils $\Delta_N = o(N^{-1/2})$, i.e. a sequence along which Δ becomes small relative to the sampling error. Whilst uniformity failures are standard in econometrics, this one is unpalatable because it concerns a very salient region of the parameter space; were it neglected, one would be led to construct confidence intervals that *shrink* as a parameter moves from point identification to slight underidentification.[3]

IM therefore conclude by proposing an intermediate confidence region that takes the uniformity problem into account. It is defined as

$$CI_\alpha^1 \equiv \left[\widehat{\theta}_l - C_N \cdot \widehat{\sigma}_l N^{-1/2}, \widehat{\theta}_u + C_N \cdot \widehat{\sigma}_u N^{-1/2}\right], \tag{1}$$

where C_N solves

$$\Phi\left(C_N + \sqrt{N}\frac{\widehat{\Delta}}{\max\{\widehat{\sigma}_l, \widehat{\sigma}_u\}}\right) - \Phi\left(-C_N\right) = \alpha. \tag{2}$$

This confidence region is still defined as the union of pointwise confidence sets, but the level of those sets is calibrated in a novel way by the choice of C_N. For a 95% confidence set, C_N will be $\Phi^{-1}(0.975) \approx 1.96$ if $\widehat{\Delta} = 0$, that is if point identification must be presumed, and will approach $\Phi^{-1}(0.95) \approx 1.64$ as $\widehat{\Delta}$ grows large relative to sampling error. IM first establish validity of CI_α^1 when Δ is known (lemma 3) and, in their final result (lemma 4), show uniform validity under the following assumption.

Assumption 1 ***(i)*** *There are estimators for the lower and upper bound $\widehat{\theta}_l$ and $\widehat{\theta}_u$ that satisfy:*

$$\sqrt{N}\begin{bmatrix} \widehat{\theta}_l - \theta_l \\ \widehat{\theta}_u - \theta_u \end{bmatrix} \xrightarrow{d} N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_l^2 & \rho\sigma_l\sigma_u \\ \rho\sigma_l\sigma_u & \sigma_u^2 \end{bmatrix}\right)$$

uniformly in $P \in \mathcal{P}$, and there are estimators $\left(\widehat{\sigma}_l^2, \widehat{\sigma}_l^2, \widehat{\rho}\right)$ that converge to their population values uniformly in $P \in \mathcal{P}$.

(ii) *For all $P \in \mathcal{P}$, $\underline{\sigma}^2 \leq \sigma_l^2, \sigma_u^2 \leq \overline{\sigma}^2$ for some positive and finite $\underline{\sigma}^2$ and $\overline{\sigma}^2$, and $\theta_u - \theta_l \leq \overline{\Delta} < \infty$.*

(iii) *For all $\epsilon > 0$, there are $v > 0$, K, and N_0 s.t. $N \geq N_0$ implies $Pr\left(\sqrt{N}\left|\widehat{\Delta} - \Delta\right| > K\Delta^v\right) < \epsilon$ uniformly in $P \in \mathcal{P}$.*

[2] To avoid uninstructive complications, I presume $\alpha \geq .5$ throughout.

[3] The problem would be avoided if $\mathcal{P}$ were restricted s.t. Δ is bounded away from 0. But such a restriction will frequently be inappropriate. For example, one cannot a priori bound from below the degree of item nonresponse in a survey or of attrition in a panel.

Even in cases where Δ is known a priori, e.g. interval data, the problem arguably disappears only in a superficial sense. Were it ignored, one would construct confidence intervals that work uniformly given any model but whose performance deteriorates across models as point identification is approached.

Whilst it is clear that uniformity can only obtain under restrictions on $\mathcal{P}$, it is important to note here that Δ is not bounded from below, thus the specific uniformity problem that arises near point identification is not assumed away. Having said that, conditions (i) and (ii) are fairly standard, but (iii) deserves some explanation. It implies that $\widehat{\Delta}$, which estimates the length of the identified interval (and hence the "degree of underidentification"), approaches its population counterpart Δ in a specific way. If $\Delta = 0$, then $\widehat{\Delta} = 0$ with probability approaching 1 in finite samples, i.e. if point identification obtains, then this will be learned exactly, and the limiting distribution of $\widehat{\Delta}$ must be degenerate. What's more, degenerate limiting distributions occur along any parameter sequence that converges to zero, as is formally stated in the following lemma.

Lemma 1 *Assumption 1(iii) implies that* $\sqrt{N}\left|\widehat{\Delta} - \Delta_N\right| \xrightarrow{p} 0$ *for any sequence of distributions* $\{P_N\} \subseteq \mathcal{P}$ *s.t.* $\Delta_N \equiv \Delta(P_N) \to 0$.

In words, assumption 1(iii) requires $\widehat{\Delta}$ to be superefficient at $\Delta = 0$. This feature appears to not have been previously recognized; it is certainly nonstandard and might even seem undesirable.[4] However, as will be shown below, it is fulfilled and can be put to good use in a leading application.

The difficulty is that assumptions 1(i)-(ii) and (iii) are inconsistent except if $\rho = 1$. To see this, note that if $\rho < 1$, then assumption 1(i)-(ii) implies that $\sqrt{N}\left(\widehat{\theta}_l - \theta_l, \widehat{\theta}_u - \theta_u\right)'$ uniformly converges to a nondegenerate, bivariate normal distribution whose variance is bounded from below, hence $\widehat{\Delta}$ uniformly converges to a normal distribution with accordingly bounded variance. In view of lemma 1, this is inconsistent with condition (iii).

2.2 Inference with Superefficiency

In this section, I maintain superefficiency but resolve the aforementioned difficulty. Specifically, I propose the following assumption.[5]

Assumption 2 ***(i)*** *There exists an estimator* $\widehat{\theta}_l$ *of* θ_l *that satisfies:*

$$\sqrt{N}\left[\widehat{\theta}_l - \theta_l\right] \xrightarrow{d} N\left(0, \sigma_l^2\right)$$

uniformly in $P \in \mathcal{P}$, *and there is an estimator* $\widehat{\sigma}_l^2$ *that converges to* σ_l^2 *uniformly in* $P \in \mathcal{P}$.

[4] When Hodgson originally defined a superefficient estimator, his intent was not, of course, to propose its use. For cautionary tales regarding the implicit, and sometimes inadvertent, use of superefficient estimators, see Leeb and Pötscher (2005).

[5] The following is not the only possible adjustment to lemma 4. For example, Guido Imbens pointed out that one could allow for ρ to be a restricted function of Δ. However, assumption 2 is weaker than other adjustments I am aware of.

(ii) *There exists an estimator* $\widehat{\Delta}$ *of* Δ *s.t.*

$$\sqrt{N}\left[\left(\widehat{\theta}_l+\widehat{\Delta}\right)-(\theta_l+\Delta)\right]\xrightarrow{d}N\left(0,\sigma_u^2\right)$$

uniformly in $P\in\mathcal{P}$, *and there is an estimator* $\widehat{\sigma}_u^2$ *that converges to* σ_u^2 *uniformly in* $P\in\mathcal{P}$.

(iii) *There exists* $x\in(0,1/2)$ *s.t.* $\sqrt{N}\left|\widehat{\Delta}-\Delta_N\right|\xrightarrow{p}0$ *for any sequence of distributions* $\{P_N\}\subseteq\mathcal{P}$ *with* $\Delta_N\leq N^{-x}$.

(iv) *For all* $P\in\mathcal{P}$, $\underline{\sigma}^2\leq\sigma_l^2,\sigma_u^2\leq\overline{\sigma}^2$ *for some positive and finite* $\underline{\sigma}^2$ *and* $\overline{\sigma}^2$.

Assumption 2 models a situation where θ_u is estimated only indirectly by $\widehat{\theta}_u\equiv\widehat{\theta}_l+\widehat{\Delta}$. (By symmetry, the case of estimating (θ_u,Δ) is covered as well.) Importantly, uniform joint asymptotic normality of $\left(\widehat{\theta}_l,\widehat{\theta}_l+\widehat{\Delta}\right)$ is not imposed. Furthermore, condition (iii) has been replaced with a requirement that is strictly weaker (by lemma 1, assumption 1(iii) implies that assumption 2(iii) holds for all $x>0$) and arguably more transparent about what is really being required.

Of course, assumption 2(iii) is again a superefficiency condition, but it faithfully models a leading application and IM's motivation, namely estimation of a mean with missing data. To see this, let $\theta_0=\mathbb{E}X$, where $X\in[0,1]$, and assume that one observes realizations of $(D,D\cdot X)$, where $D\in\{0,1\}$ indicates whether a data point is present ($D=1$) or missing ($D=0$). Then the identified set for θ_0 is

$$[\theta_l,\theta_u]=[(1-\Delta)\,\mathbb{E}(X|D=1),(1-\Delta)\,\mathbb{E}(X|D=1)+\Delta],$$

where $\Delta\equiv1-\mathbb{E}D$ (the definition as "one minus propensity score" insures consistency with previous use). The obvious estimator for Θ_0 is its sample analog

$$\widehat{\Theta}_0\equiv[\underbrace{\mathbb{E}_ND\mathbb{E}_N(X|D=1)}_{\widehat{\theta}_l},\underbrace{\mathbb{E}_ND\mathbb{E}_N(X|D=1)}_{\widehat{\theta}_l}+\underbrace{1-\mathbb{E}_ND}_{\widehat{\Delta}}],$$

where $\mathbb{E}_N$ denotes sample means. So in this application, indirect estimation of $\widehat{\theta}_u$ as $\widehat{\theta}_l+\widehat{\Delta}$ is natural. Uniform convergence of $\left(\widehat{\theta}_l,\widehat{\theta}_l+\widehat{\Delta}\right)$ to individually normal distributions follows from a uniform central limit theorem. Finally, it is interesting to note that $\widehat{\Delta}$ fulfils part (iii) of both assumptions 1 and 2, making it a natural example of a superefficient estimator. At the same time, this application is not covered by other results developed below, because the limiting distribution of $\left(\widehat{\theta}_l,\widehat{\theta}_l+\widehat{\Delta}\right)$ is not uniformly jointly normal.[6]

This section's main result is as follows.

[6] This observation adjusts an oversight in IM (pp. 1850-1851), who claim that their application fulfils assumption 1 but only verify individual asymptotic normality of $\widehat{\theta}_l$ and $\widehat{\theta}_u$. To see that $\widehat{\Delta}$ is not uniformly asymptotically normal, observe that $\Pr\left(\widehat{\Delta}<0\right)=0\neq\Phi(x)$ for any finite x, and the event that $\widehat{\Delta}<0$ does not "escape to infinity" along a local sequence $\Delta_N\to0$.

Proposition 1 *Let assumption 2 hold. Then*

$$\lim_{N\to\infty}\inf_{\theta\in\Theta}\inf_{P:\theta_0(P)=\theta}\Pr\left(\theta\in CI_\alpha^1\right)=\alpha.$$

A brief intuition for proposition 1 goes as follows. One could define an "oracle" threshold $\widetilde{C}_N$ by replacing all sample quantities in expressio (2) with population analogs. If CI_α^1 could be based on $\widetilde{C}_N$, its validity would follow relatively straightforwardly from normal approximations. Of course, $\widetilde{C}_N$ is not feasible. C_N as defined in (2) is essentially a plug-in estimator of $\widetilde{C}_N$, but without assumption 1(iii), it need not be consistent – $\left(\widehat{\Delta}-\Delta\right)$ is of order $O(N^{-1/2})$, hence $\sqrt{N}\left(\widehat{\Delta}-\Delta\right)$ does not vanish. To resolve the issue, distinguish between parameter sequences $\{\Delta_N\}$ s.t. Δ_N vanishes fast enough for condition (iii) to apply and sequences where this fails. In the former case, $\sqrt{N}\left(\widehat{\Delta}-\Delta\right)$ does vanish, so consistency of C_N is recovered. In the latter case, Δ_N grows uniformly large relative to sampling error, so that the uniformity problem does not arise to begin with. The "naive" $CI_{1-(1-\alpha)/2}$ is then a valid construction, CI_α^1 is asymptotically equivalent to it, and no harm is done by the fact that C_N is indeed inconsistent for C_N^*.

2.3 Inference with Joint Normality

Whilst assumption 2(iii) does have applications, it is of obvious interest to consider inference on θ_0 without superefficiency. For a potential application, imagine that $\widehat{\theta}_u$ and $\widehat{\theta}_l$ derive from separate inequality moment conditions (as in Pakes et al. 2006 and Rosen 2006). I therefore now turn to the following assumption:

Assumption 3 ***(i)*** *There are estimators for $\widehat{\theta}_l$ and $\widehat{\theta}_u$ that satisfy:*

$$\sqrt{N}\begin{bmatrix}\widehat{\theta}_l-\theta_l\\ \widehat{\theta}_u-\theta_u\end{bmatrix}\xrightarrow{d}N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\sigma_l^2 & \rho\sigma_l\sigma_u\\ \rho\sigma_l\sigma_u & \sigma_u^2\end{bmatrix}\right)$$

uniformly in $P\in\mathcal{P}$, and there are estimators $(\widehat{\sigma}_l,\widehat{\sigma}_u,\widehat{\rho})$ that converge to their population values uniformly in $P\in\mathcal{P}$.

(ii) *For all $P\in\mathcal{P}$, $\underline{\sigma}^2\le\sigma_l^2,\sigma_u^2\le\overline{\sigma}^2$ for some positive and finite $\underline{\sigma}^2$ and $\overline{\sigma}^2$, and $\Delta\le\overline{\Delta}<\infty$.*

Relative to assumption 1, assumption 3 simply removes superefficiency altogether. This leads to numerous difficulties that were previously avoided. At the core of these lies the fact that the joint distribution of $\sqrt{N}\left(\widehat{\theta}_l-\theta_l,\widehat{\theta}_u-\theta_u\right)'$ may be nondegenerate bivariate even when Δ is small. This complicates estimation of Δ, and it raises the possibility that $\widehat{\theta}_u<\widehat{\theta}_l$ in finite samples. Accounting for these issues is possible but necessitates numerous adjustments to CI_α^1.

Specifically, let $\widehat{\Delta} \equiv \widehat{\theta}_u - \widehat{\theta}_l$ and

$$\Delta^* \equiv \begin{cases} \widehat{\Delta}, & \widehat{\Delta} > N^{-1/3} \\ 0 & \text{otherwise.} \end{cases}.$$

Let $(C_l, C_u) \in \mathbb{R}_+^2$ minimize $(C_l + C_u)$ subject to the constraint that

$$\Pr\left(-\frac{C_l}{\widehat{\sigma}_l} \leq z_1, \widehat{\rho} z_1 \leq \frac{C_u + \sqrt{N}\Delta^*}{\widehat{\sigma}_u} + \sqrt{1-\widehat{\rho}^2} z_2\right) \geq \alpha \tag{3}$$

$$\Pr\left(-\frac{C_l + \sqrt{N}\Delta^*}{\widehat{\sigma}_l} + \sqrt{1-\widehat{\rho}^2} z_2 \leq \widehat{\rho} z_1, z_1 \leq \frac{C_u}{\widehat{\sigma}_u}\right) \geq \alpha, \tag{4}$$

where z_1 and z_2 are independent standard normal random variables. In typical cases, (C_l, C_u) will be uniquely characterized by the fact that both of (3,4) hold with equality, but it is conceivable that one of the conditions is slack at the solution. Finally, define

$$CI_\alpha^2 \equiv \begin{cases} \left[\widehat{\theta}_l - N^{-1/2} C_l, \widehat{\theta}_u + N^{-1/2} C_u\right], & \widehat{\theta}_l - N^{-1/2} C_l \leq \widehat{\theta}_u + N^{-1/2} C_u \\ \varnothing & \text{otherwise} \end{cases}. \tag{5}$$

If $\widehat{\theta}_u$ is too far below $\widehat{\theta}_l$, then CI_α^2 is empty, which should be interpreted as rejection of the maintained assumption that $\theta_u \geq \theta_l$.

Proposition 2 *Let assumption 2 hold. Then*

$$\lim_{N\to\infty} \inf_{\theta\in\Theta} \inf_{P:\theta_0(P)=\theta} \Pr\left(\theta \in CI_\alpha^2\right) = \alpha.$$

The preceding algebra may appear involved, but conceptually, CI_α^2 adapts CI_α^1 in a straightforward manner. To see this, note first that (2) can be equivalently written as

$$\Pr\left(-C_N \leq z \leq C_N + \sqrt{N}\frac{\widehat{\Delta}}{\max\{\widehat{\sigma}_l, \widehat{\sigma}_u\}}\right) = \alpha,$$

where z is standard normal. This expression reflects integration with respect to a univariate normal distribution because in the critical case, i.e. when $\Delta_N \to 0$, Δ_N is known and hence $\widehat{\theta}_u = \widehat{\theta}_l + \Delta_N$ asymptotically. The limiting distribution of $\left(\widehat{\theta}_l, \widehat{\theta}_u\right)$ is then degenerate bivariate and therefore captured by the above. Furthermore, this limiting case is characterized by $\sigma_l = \sigma_u$, leading to symmetry between θ_l and θ_u.

Under joint normality, the limiting distribution of $\sqrt{N}\left(\widehat{\theta}_l - \theta_l, \widehat{\theta}_u - \theta_u\right)'$ may be nondegenerate bivariate even when Δ is small. Hence, it would be with loss of generality to replace $\widehat{\sigma}_l$ respectively $\widehat{\sigma}_u$ with their mutual maximum, it would accordingly be with loss of generality to equate C_l and C_u, and calibration of C_l and C_u has to be with respect to a bivariate limiting distribution. The expressions

(3,4) capture these generalizations. Among other things, the loss of symmetry between C_l and C_u leads to two calibration equations which insure that nominal coverage at both θ_l and θ_u equals α.[7]

With Δ^* replaced by $\widehat{\Delta}$, (3,4) would therefore generalize (2). Indeed, one could use it to modify CI_α^1 without affecting its asymptotic validity; it is easy to show that given assumption 2, both versions of CI_α^1 are asymptotically equivalent. However, many parameter sequences that fulfil assumption 2 in the limit are characterized by $\sigma_l \neq \sigma_u$, as well as $\rho < 1$, for finite N. For such sequences, the modification might constitute a finite sample refinement, although one would have to take into account that it relies on an estimator of ρ.

There is an additional problem however: Absent superefficiency of $\widehat{\Delta}$, C_l and C_u need not converge to their "oracle" counterparts. This problem is here resolved by replacing $\widehat{\Delta}$ with a thresholding estimator that recovers superefficiency at 0. Of course, there is some price to be paid in terms of uniformity: CI_α^2 is uniformly valid and pointwise exact, but conservative along parameter sequences where $0 < \Delta_N \leq O(N^{-1/3})$.

The final new element of CI_α^2 is that it embeds a specification test. IM do not consider such a test, presumably for two reasons: It does not arise in their leading application, i.e. estimation of means with missing data, and it is trivial in their framework because the superefficiency assumption implies fast learning about Δ in the critical region where $\Delta \approx 0$. But the issue is substantively interesting in other applications, and is nontrivial when $\widehat{\Delta} < 0$ is a generic possibility. Of course, one could construct a version of CI_α^2 that is never empty; one example would be the convex hull of $\left\{\widehat{\theta}_l - N^{-1/2}C_l, \widehat{\theta}_u + N^{-1/2}C_u\right\}$. But in samples where $\widehat{\theta}_u$ is much below $\widehat{\theta}_l$, this would appear absurd; realistically, such samples would lead one to question one's model specification. This aspect motivates the specification test, which does not affect the interval's asymptotic validity. For the intended applications of proposition 2, e.g. moment inequalities, the specification test appears to be an attractive feature.

Some further remarks on CI_α^2 are in order.

- The construction of Δ^* can be refined in two ways. First, I defined a hard thresholding estimator for simplicity, but a smooth version would also insure validity and presumably improve performance for Δ close to the threshold. Second, the proof will go through as long as the threshold is of order $O(N^{-1/2+\varepsilon})$ for some $\varepsilon > 0$. Adjustment of ε is subject to the following trade-off: The smaller ε, the less conservative CI_α^2 is along local parameter sequences. On the other hand, as $\varepsilon \to 0$, the quality of the uniform approximation to $\lim_{N\to\infty} \inf_{\theta\in\Theta} \inf_{P:\theta_0(P)=\theta} \Pr\left(\theta \in CI_\alpha^2\right)$ deteriorates, and uniformity would break down if $\varepsilon = 0$. Fine-tuning this trade-off is a possible subject of further research.

[7] Nominal coverage at interior points will exceed α and is well understood to go to 1 pointwise (although not uniformly).

- The specification test is easily seen to be pointwise consistent. It also has some power against local alternatives of the type $\Delta_N = -hN^{-1/2}$, although this power will not converge to $1-\alpha$ as $h \to 0$. This reflects an impossibility result: It is easy to construct a stand-alone specification test with full asymptotic power against local alternatives, but this test would have power $(1-\alpha)$ if $\Delta = 0$. For CI_α^2 to have confidence level α at $\Delta = 0$, it would therefore have to be a 100% confidence interval conditional on the specification test being passed. Assuming that CI_α^2 is desired to be finite, the embedded specification test necessarily has inferior power.

3 Conclusion

This note extended Imbens and Manski's (2004) analysis of confidence regions for partially identified parameters. A brief summary of its contributions goes as follows. First, I notice that one assumption used for IM's final result boils down to superefficient estimation of a nuisance parameter. This nature of their assumption appears to have gone unnoticed before. Furthermore, it implies that the assumptions imposed for the result are not, in general, consistent. I establish the conclusion of IM's lemma 4 from an assumption that avoids this issue and furthermore weakens the superefficiency condition. I also show that superefficiency can be disposed of altogether, although this requires substantial adjustments to the confidence region. In particular, I propose a confidence region that is adapted to the bivariate nature of the problem and that embeds a specification test for nonemptiness of the identified set.

A Proofs

Lemma 1 The aim is to show that if $\Delta_N \to 0$, then

$$\forall \delta, \varepsilon > 0, \exists N^* : N \geq N^* \Longrightarrow \Pr\left(\sqrt{N}\left|\widehat{\Delta} - \Delta_N\right| > \delta\right) < \varepsilon.$$

Fix δ and ε. By assumption 1(iii), there exist N^{**}, $v > 0$, and k s.t.

$$N \geq N^{**} \Longrightarrow \Pr\left(\sqrt{N}\left|\widehat{\Delta} - \Delta\right| > K\Delta^v\right) < \varepsilon$$

uniformly over $\mathcal{P}$. Specifically, the preceding inequality will obtain if Δ is chosen in $(0, \delta^{1/v}K^{-1/v}]$, in which case $K\Delta^v \leq \delta$. Because $\Delta_N \to 0$, N^{***} can be chosen s.t. $N \geq N^{***} \Rightarrow \Delta_N \leq \delta^{1/v}K^{-1/v}$. Hence, the conclusion obtains by choosing $N^* = \max\{N^{**}, N^{***}\}$.

Proposition 1 I will prove that $\inf_{\{\theta_N\}\subseteq\Theta} \inf_{\{P_N\}:\theta_N\in\Theta_0(P_N)} \lim_{N\to\infty} \Pr\left(\theta_N \in CI_\alpha^1\right) \to \alpha$. This is a pointwise limit, but because I take it over sequences of distributions, proposition 1 is implied. In particular, the limit applies to a sequence $\{\theta_N, P_N\}$ s.t. $(\theta_N, P_N) \in \arg\min_{(\theta,P):\theta\in\Theta_0(P)} \Pr\left(\theta \in CI_\alpha^1|N\right)$.

For the remainder of this proof, I will identify sequences $\{P_N\}$ with the implied sequences $\{\Delta_N, \theta_N\} \equiv \{\Delta(P_N), \theta_0(P_N)\}$. For simplicity of notation, I also suppress the N subscript on $(\theta_l, \theta_u, \sigma_l, \sigma_u)$ as well as all estimators.

I present two arguments, one for the case that $\{\Delta_N\}$ is weakly dominated by N^{-x} and one for the case that $\{\Delta_N\}$ strictly dominates N^{-x}. Any sequence $\{P_N\}$ can be decomposed into two subsequences such that either subsequence is covered by one of the cases.

Let $\{\Delta_N\}$ be weakly dominated by N^{-x}, then $\sqrt{N}\left|\widehat{\Delta} - \Delta_N\right| \xrightarrow{p} 0$ by condition (iii), hence

$$\Phi\left(C_N + \sqrt{N}\frac{\widehat{\Delta}}{\max\{\widehat{\sigma}_l, \widehat{\sigma}_u\}}\right) \xrightarrow{p} \Phi\left(C_N + \sqrt{N}\frac{\Delta_N}{\max\{\widehat{\sigma}_l, \widehat{\sigma}_u\}}\right),$$

but it is also routinely shown that

$$\Phi\left(C_N + \sqrt{N}\frac{\Delta_N}{\max\{\widehat{\sigma}_l, \widehat{\sigma}_u\}}\right) \xrightarrow{p} \Phi\left(C_N + \sqrt{N}\frac{\Delta_N}{\max\{\sigma_l, \sigma_u\}}\right).$$

Now write

$$\begin{aligned}
& \Pr\left(\theta_l \in CI_\alpha^1\right) \\
= & \Pr\left(\widehat{\theta}_l - C_N\widehat{\sigma}_l N^{-1/2} \le \theta_l \le \widehat{\Delta} + C_N\widehat{\sigma}_u N^{-1/2}\right) \\
= & \Pr\left(-C_N\frac{\widehat{\sigma}_l}{\sigma_l} \le \frac{\sqrt{N}\left(\theta_l - \widehat{\theta}_l\right)}{\sigma_l} \le \frac{\sqrt{N}}{\sigma_l}\widehat{\Delta} + C_N\frac{\widehat{\sigma}_u}{\sigma_l}\right) \\
\to & \Phi\left(\frac{\sqrt{N}}{\sigma_l}\Delta_N + C_N\right) - \Phi\left(-C_N\right) \\
\ge & \Phi\left(\frac{\sqrt{N}}{\max\{\sigma_l, \sigma_u\}}\Delta_N + C_N\right) - \Phi\left(-C_N\right) \\
\to & \alpha,
\end{aligned}$$

where the last line follows from the preceding argument and the first convergence step uses numerous convergence assumptions including superefficiency of $\widehat{\Delta}$. the argument for θ_u is similar. Finally, if Δ can be considered known, then evaluation of second derivatives immediately shows that the uniform limit of $\Pr\left(a\theta_l + (1-a)\theta_u \in CI_\alpha^1\right)$ is concave in a over $[0,1]$, implying that $\min_{\theta_0\in\Theta_0}\Pr\left(\theta_0 \in CI_\alpha^1\right)$ is achieved on a subset of $\{\theta_l, \theta_u\}$.

Now let $\{\Delta_N\}$ strictly dominate N^{-x}. Then $\sqrt{N}\Delta_N$ dominates $N^{1/2-x}$ and hence diverges to infinity. Write

$$
\begin{aligned}
& \Pr\left(\theta_N \in CI_\alpha^1\right) \\
= \;& \Pr\left(\widehat{\theta}_l - C_N \widehat{\sigma}_l N^{-1/2} \leq \theta_N \leq \widehat{\theta}_l + \widehat{\Delta} + C_N \widehat{\sigma}_u N^{-1/2}\right) \\
= \;& \Pr\left(-C_N \widehat{\sigma}_l \leq \sqrt{N}\left(\theta_N - \theta_l\right) + \sqrt{N}\left(\theta_l - \widehat{\theta}_l\right) \leq \sqrt{N}\widehat{\Delta} + C_N \widehat{\sigma}_u\right) \\
= \;& \Pr\left(-C_N \widehat{\sigma}_l \leq \sqrt{N}\left(\theta_N - \theta_l\right) + \sqrt{N}\left(\theta_l - \widehat{\theta}_l\right)\right) \\
& - \Pr\left(\sqrt{N}\left(\theta_N - \theta_l\right) + \sqrt{N}\left(\theta_l - \widehat{\theta}_l\right) > \sqrt{N}\widehat{\Delta} + C_N \widehat{\sigma}_u\right).
\end{aligned}
$$

Fix any subsequence of $\{P_N\}$ s.t. $\limsup_{N\to\infty} \sqrt{N}\left(\theta_N - \theta_l\right) < \infty$ and consider

$$
\begin{aligned}
& \Pr\left(\sqrt{N}\left(\theta_N - \theta_l\right) + \sqrt{N}\left(\theta_l - \widehat{\theta}_l\right) > \sqrt{N}\widehat{\Delta} + C_N \widehat{\sigma}_u\right) \\
= \;& \Pr\left(\sqrt{N}\left(\theta_N - \theta_l\right) > \sqrt{N}\widehat{\Delta} + C_N \widehat{\sigma}_u - \sqrt{N}\left(\theta_l - \widehat{\theta}_l\right)\right) \\
\to \;& \Pr\left(\sqrt{N}\left(\theta_N - \theta_l\right) > \sqrt{N}\Delta_N + C_N \widehat{\sigma}_u - \sqrt{N}\left(\theta_l - \widehat{\theta}_l\right)\right) \\
\leq \;& \Pr\left(\sqrt{N}\left(\theta_l - \widehat{\theta}_l\right) > \sqrt{N}\Delta_N - \sqrt{N}\left(\theta_N - \theta_l\right)\right) \\
\to \;& 0,
\end{aligned}
$$

where the first convergence statement uses uniform consistency of $\widehat{\Delta}$, and the inequality uses that $C_N \widehat{\sigma}_u \geq 0$ by construction (and rearranges terms). The final step uses that $\sqrt{N}\Delta_N$ diverges and $\sqrt{N}\left(\theta_N - \theta_l\right)$ is bounded, hence $\sqrt{N}\Delta_N - \sqrt{N}\left(\theta_N - \theta_l\right)$ diverges, whereas $\sqrt{N}\left(\theta_l - \widehat{\theta}_l\right)$ converges to a random variable by assumption. It follows that

$$
\begin{aligned}
& \lim_{N\to\infty} \Pr\left(\theta_N \in CI_\alpha^1\right) \\
= \;& \lim_{N\to\infty} \Pr\left(-C_N \widehat{\sigma}_l \leq \sqrt{N}\left(\theta_N - \theta_l\right) + \sqrt{N}\left(\theta_l - \widehat{\theta}_l\right)\right) \\
\geq \;& \lim_{N\to\infty} \Pr\left(-C_N \widehat{\sigma}_l \leq \sqrt{N}\left(\theta_l - \widehat{\theta}_l\right)\right) \\
= \;& 1 - \Phi(C_N) \geq \alpha,
\end{aligned}
$$

where the first inequality uses that $\sqrt{N}\left(\theta_N - \theta_l\right) \geq 0$, and the second inequality uses the definition of C_N, as well as convergence of $\widehat{\sigma}_l$ and $\sqrt{N}\left(\theta_l - \widehat{\theta}_l\right)/\sigma_l$.

For any subsequence of $\{P_N\}$ s.t. $\sqrt{N}\left(\theta_N - \theta_l\right)$ diverges, one can apply a symmetric argument to show that

$$
\Pr\left(-C_N \widehat{\sigma}_l \leq \sqrt{N}\left(\theta_N - \theta_l\right) + \sqrt{N}\left(\theta_l - \widehat{\theta}_l\right)\right) \to 1,
$$

whereas

$$
\lim_{N\to\infty} \Pr\left(\sqrt{N}\left(\theta_N - \theta_l\right) + \sqrt{N}\left(\theta_l - \widehat{\theta}_l\right) > \sqrt{N}\widehat{\Delta} + C_N \widehat{\sigma}_u\right) \leq \alpha,
$$

which jointly implies the conclusion.

Proposition 2 As with proposition 1, the proof will establish a limit that is pointwise but applies to parameter sequences. Consider sequences s.t. $\Delta_N \geq N^{-5/12}$, then the proof entirely resembles the analog case of proposition 1. Thus, restrict attention to $\{\Delta_N\}$ s.t. $\Delta_N < N^{-5/12}$, hence $\Delta_N = o(N^{-1/2})$, hence $\Pr(\Delta^* = 0) \to 1$. Initially also restrict attention to $\Pr(\theta_l \in CI_\alpha^2)$ and $\Pr(\theta_u \in CI_\alpha^2)$. Let $(\widetilde{C}_l, \widetilde{C}_u)$ fulfil

$$\begin{aligned}
\Pr\left(-\frac{\widetilde{C}_l}{\sigma_l} \leq z_1, \rho z_1 \leq \frac{\widetilde{C}_u + \sqrt{N}\Delta}{\sigma_u} + \sqrt{1-\rho^2} z_2\right) &\geq \alpha \\
\Pr\left(-\frac{\widetilde{C}_l + \sqrt{N}\Delta}{\sigma_l} + \sqrt{1-\rho^2} z_2 \leq \rho z_1, z_1 \leq \frac{\widetilde{C}_u}{\sigma_u}\right) &\geq \alpha,
\end{aligned}$$

i.e. the oracle counterpart of (C_l, C_u). Observe that although suppressed in the notation, $(\widetilde{C}_l, \widetilde{C}_u)$ is a (nonrandom) function of N. Write

$$\begin{aligned}
& \Pr\left(\theta_l \in \left[\widehat{\theta}_l - N^{-1/2}\widetilde{C}_l, \widehat{\theta}_u + N^{-1/2}\widetilde{C}_u\right]\right) \\
= & \Pr\left(\widehat{\theta}_l - N^{-1/2}\widetilde{C}_l \leq \theta_l \leq \widehat{\theta}_u + N^{-1/2}\widetilde{C}_u\right) \\
= & \Pr\left(-\frac{\widetilde{C}_l}{\sigma_l} \leq \frac{\sqrt{N}}{\sigma_l}\left(\theta_l - \widehat{\theta}_l\right) \leq \frac{\sqrt{N}}{\sigma_l}\left(\widehat{\theta}_u - \theta_l\right) + \frac{\widetilde{C}_u}{\sigma_l}\right) \\
= & \Pr\left(-\frac{\widetilde{C}_l}{\sigma_l} \leq \frac{\sqrt{N}}{\sigma_l}\left(\theta_l - \widehat{\theta}_l\right) \leq \frac{\sqrt{N}}{\sigma_l}\left(\widehat{\theta}_u - \theta_u + \theta_u - \theta_l + \theta_l - \widehat{\theta}_l\right) + \frac{\widetilde{C}_u}{\sigma_l}\right) \\
\to & \Pr\left(-\frac{\widetilde{C}_l}{\sigma_l} \leq \frac{\sqrt{N}}{\sigma_l}\left(\theta_l - \widehat{\theta}_l\right) \leq \frac{\sqrt{N}}{\sigma_l}\Delta + \frac{\sqrt{N}}{\sigma_l}\left(1 - \rho\frac{\sigma_u}{\sigma_l}\right)\left(\theta_l - \widehat{\theta}_l\right) + \frac{\sigma_u}{\sigma_l}\sqrt{1-\rho^2} z_2 + \frac{\widetilde{C}_u}{\sigma_l}\right) \\
= & \Pr\left(-\frac{\widetilde{C}_l}{\sigma_l} \leq \frac{\sqrt{N}}{\sigma_l}\left(\theta_l - \widehat{\theta}_l\right), \rho\sigma_u\frac{\sqrt{N}}{\sigma_l}\left(\theta_l - \widehat{\theta}_l\right) \leq \frac{\widetilde{C}_u}{\sigma_l} + \frac{\sqrt{N}}{\sigma_l}\Delta + \frac{\sigma_u}{\sigma_l}\sqrt{1-\rho^2} z_2\right) \\
\to & \Pr\left(-\frac{\widetilde{C}_l}{\sigma_l} \leq z_1, \rho z_1 \leq \frac{\widetilde{C}_u}{\sigma_u} + \frac{\sqrt{N}}{\sigma_u}\Delta + \sqrt{1-\rho^2} z_2\right) \qquad (6) \\
\geq & \alpha.
\end{aligned}$$

Here, the first convergence statement uses that by assumption,

$$\left(\sqrt{N}\left(\widehat{\theta}_u - \theta_u\right) | \sqrt{N}\left(\theta_l - \widehat{\theta}_l\right)\right) \xrightarrow{d} N\left(-\rho\frac{\sigma_u}{\sigma_l}\sqrt{N}\left(\theta_l - \widehat{\theta}_l\right), \sigma_u^2(1-\rho^2)\right)$$

uniformly.

As pointed out in the text, (C_l, C_u) is not a uniformly consistent estimator of $(\widetilde{C}_l, \widetilde{C}_u)$. More specifically, it is routinely shown that $(\sigma_l, \sigma_u, \rho)$ can be replaced with their estimators in the above, but estimation error in $\sqrt{N}\Delta$ does not vanish.However, $\Delta^* = 0 \leq \Delta$ with probability appraching 1. Expression (6) is easily seen to increase in Δ for every (C_l, C_u) hence CI_α^2 is valid (if potentially conservative) at θ_l. The argument for θ_u is similar. Finally, (C_l, C_u) can minimize $(C_l + C_u)$ s.t. (3,4) only if at least one of (3,4) binds, implying that CI_α^2 is exact for at least one of $\{\theta_l, \theta_u\}$.

Now consider $\theta_0 \equiv a\theta_l + (1-a)\theta_u$, some $a \in [0,1]$. Observe that by assumption,

$$\sqrt{N}\left(a\widehat{\theta}_l + (1-a)\widehat{\theta}_u - \theta_0\right) \xrightarrow{d} N(0, \sigma_a),$$

where $\sigma_a^2 \equiv a^2\sigma_l^2 + (1-a)^2\sigma_u^2 - 2a(1-a)\rho\sigma_l\sigma_u$. Write

$$\begin{aligned}
& \Pr\left(\theta_0 \in \left[\widehat{\theta}_l - N^{-1/2}C_l, \widehat{\theta}_u + N^{-1/2}C_u\right]\right) \\
= \; & \Pr\left(\widehat{\theta}_l - N^{-1/2}C_l \leq \theta_0 \leq \widehat{\theta}_u + N^{-1/2}C_u\right) \\
= \; & \Pr\left(\frac{\sqrt{N}}{\sigma_a}(1-a)\left(\widehat{\theta}_l - \widehat{\theta}_u\right) - \frac{C_l}{\sigma_a} \leq \frac{\sqrt{N}}{\sigma_a}(\theta_0 - a\widehat{\theta}_l - (1-a)\widehat{\theta}_u) \leq \frac{\sqrt{N}}{\sigma_a}a\left(\widehat{\theta}_u - \widehat{\theta}_l\right) + \frac{C_u}{\sigma_a}\right) \\
= \; & \Pr\left(\frac{\sqrt{N}}{\sigma_a}(1-a)\left(\widehat{\Delta} - \Delta\right) + \frac{\sqrt{N}}{\sigma_a}(1-a)\Delta - \frac{C_l}{\sigma_a}\right. \\
& \left. \leq \frac{\sqrt{N}}{\sigma_a}(\theta_0 - a\widehat{\theta}_l - (1-a)\widehat{\theta}_u) \leq \frac{\sqrt{N}}{\sigma_a}a\left(\widehat{\Delta} - \Delta\right) + \frac{\sqrt{N}}{\sigma_a}a\Delta + \frac{C_u}{\sigma_a}\right).
\end{aligned}$$

Consider varying Δ whilst holding $(\theta_l, \sigma_l, \sigma_u, \rho)$ constant. C_l and C_u depend on Δ only through Δ^*, but recall that $\Pr(\Delta^* = 0) \to 1$. Also, $\frac{\sqrt{N}}{\sigma_a}\left(\widehat{\Delta} - \Delta\right)$ is asymptotically pivotal. Hence, the preceding probability's limit depends on Δ only through $\frac{\sqrt{N}}{\sigma_a}(1-a)\Delta$ and $\frac{\sqrt{N}}{\sigma_a}a\Delta$. As $\frac{\sqrt{N}}{\sigma_a} > 0$, the probability is minimized by setting $\Delta = 0$. In this case, however, $\theta_0 = \theta_l$, and coverage has already been established.

References

[1] Haile, P. and E.T. Tamer (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy* 111: 1-50.

[2] Honoré, B.E. and E.T. Tamer (2006): "Bounds on Parameters in Panel Dynamic Discrete Choice Models," *Econometrica* 74: 611–629.

[3] Horowitz, J. and C.F. Manski (2000): "Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data," *Journal of the American Statistical Association* 95: 77–84.

[4] Imbens, G. and C.F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica* 72: 1845-1857.

[5] Leeb, H. and B. Pötscher (2005): "Model Selection and Inference: Facts and Fiction," *Econometric Theory* 21: 21-59.

[6] Manski, C.F. (2003): *Partial Identification of Probability Distributions.* Springer Verlag.

[7] Pakes, A., J. Porter, K. Ho, and J. Ishii (2006): "Moment Inequalities and their Application," mimeo, Harvard University.

[8] Rosen, A.M. (2006): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," mimeo, UCL.