More on Confidence Intervals for Partially Identified Parameters

Jörg Stoye*

New York University

December 17, 2008

Abstract

This paper extends Imbens and Manski's (2004) analysis of confidence intervals for interval identified parameters. The extension is motivated by the discovery that for their final result, Imbens and Manski implicitly assume locally superefficient estimation of a nuisance parameter.

I re-analyze the problem both with assumptions that merely weaken this superefficiency condition and with assumptions that remove it altogether. Imbens and Manski's confidence region is valid under weaker assumptions than theirs, yet superefficiency is required. I also provide a confidence interval that is valid under superefficiency but can be adapted to the general case. A methodological contribution is to observe that the difficulty of inference comes from a pre-estimation problem regarding a nuisance parameter, clarifying the connection to other work on partial identification.

Keywords: Bounds, identification regions, confidence intervals, uniform convergence, superefficiency.

JEL classification codes: C10, C14.

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 recently turned to general treatments of estimation and inference. An important contribution in this direction is due to Imbens and Manski (2004, IM henceforth), who point out that one might be interested in confidence regions for partially identified parameters themselves rather than "identified sets." The intuitively most obvious, and previously used, confidence regions are of the latter type, meaning that they are conservative for the parameters. IM propose a number of confidence regions for real-valued parameters that can be asymptotically concluded to lie in an interval.

This paper refines and extends IM's main technical result, a confidence interval that exhibits uniform coverage of partially identified parameters if the length of the identified interval is a nuisance parameter. IM rely on a high-level assumption that turns out to imply locally superefficient estimation of this nuisance parameter and that will fail in many applications. I take this discovery as point of departure for a new analysis of the problem, providing different confidence intervals that are valid both with and without superefficiency.

A brief summary and overview of results goes as follows. In section 2, I set up the inference problem, briefly summarize IM's contribution, and explain the aforementioned issue. Section 3 contains the re-analysis. It reconstructs IM's results from weaker and, as will be shown, more generic assumptions, but also proposes a different confidence region that is easily adapted to the case of no superefficiency. Section 4 concludes, and the appendix contains all proofs.

2   Background

Following Woutersen (2006), I consider a generalization of IM's setup that removes some nuisance parameters. 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. 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. 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

$$CI_\alpha = \left[\widehat{\theta}_l - \frac{c_\alpha \widehat{\sigma}_l}{\sqrt{N}}, \widehat{\theta}_u + \frac{c_\alpha \widehat{\sigma}_u}{\sqrt{N}} \right],$$

---

[1] See Manski (2003) for a survey and many references.

where $\widehat{\sigma}_l \, [\widehat{\sigma}_u]$ is a standard error for $\widehat{\theta}_l \, [\widehat{\theta}_u]$, and where c_α is chosen s.t.

$$\Phi(c_\alpha) - \Phi(-c_\alpha) = 1 - \alpha. \tag{1}$$

For example, $c_\alpha = \Phi^{-1}(0.975) \approx 1.96$ for a 95%-confidence interval. Under regularity conditions, a simple Bonferroni argument establishes that $\Pr(\Theta_0 \subseteq CI_\alpha) \to 1 - \alpha$. 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 $\Pr(\theta_0 \in CI_\alpha) \to 1 - \alpha/2$. In words, a 90% C.I. for Θ_0 is a 95% C.I. for θ_0. The reason is that asymptotically, Δ is large relative to sampling error, so that noncoverage risk is effectively one-sided at $\{\theta_l, \theta_u\}$ and vanishes otherwise. One would, therefore, be tempted to construct a level α C.I. for θ as $CI_{2\alpha}$.[2]

Unfortunately, this intuition works pointwise but not uniformly over interesting specifications of $\mathcal{P}$. Specifically, $\Pr(\theta_0 \in CI_\alpha) = 1 - \alpha$ if $\Delta = 0$ and also $\Pr(\theta_0 \in CI_\alpha) \to 1 - \alpha$ along any local parameter sequence where $\Delta_N \leq O(N^{-1/2})$, i.e. whenever Δ fails to diverge relative to sampling error. While 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 propose the following confidence region:

$$CI_\alpha^1 \equiv \left[\widehat{\theta}_l - \frac{c_\alpha^1 \widehat{\sigma}_l}{\sqrt{N}}, \widehat{\theta}_u + \frac{c_\alpha^1 \widehat{\sigma}_u}{\sqrt{N}} \right],$$

where c_α^1 solves

$$\Phi\left(c_\alpha^1 + \frac{\sqrt{N}\widehat{\Delta}}{\max\{\widehat{\sigma}_l, \widehat{\sigma}_u\}} \right) - \Phi\left(-c_\alpha^1 \right) = 1 - \alpha. \tag{2}$$

Comparison of (2) with (1) reveals that (2) takes into account the estimated length of the identified set. For a 95% confidence set, c_α^1 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 show uniform validity of CI_α^1 under the following assumption.

---

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

[3] In fact, CI_α is not uniformly valid (at level $1 - \alpha$) for the identified set either. Application of the union/intersection method yields an interval that is pointwise but not uniformly equivalent to CI_α, with problems again occurring if Δ_N is local to zero. This is not the focus of this paper however.

All of these problem could be avoided by bounding Δ 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. Also, even when Δ is known a priori, e.g. with 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.

Assumption 1 *(i) There exist estimators $\widehat{\theta}_l$ and $\widehat{\theta}_u$ that satisfy:*

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

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 $P \in \mathcal{P}$.

While it is clear that uniformity can obtain only under restrictions on $\mathcal{P}$, it is important to note 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 $\widehat{\Delta}$ 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 local parameter sequence that converges to zero, as is formally stated in the following lemma.[4]

Lemma 1 *Assumption 1(iii) implies that $\sqrt{N} \left| \widehat{\Delta} - \Delta_N \right| \xrightarrow{p} 0$ for all sequences 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 zero. It seems that this was not previously recognized, and it is certainly a nonstandard restriction. To be sure, there are relevant cases where superefficiency obtains naturally. One important example is that sample probabilities are superefficient estimators of population probabilities near zero; this is why Imbens and Manski (2004) are able to verify their assumptions for the mean with missing data. A new and less obvious example will be presented in this paper. Also, superefficiency has a somewhat dubious reputation due to pitfalls encountered when it is induced artificially. Such constructions incur large local risk that is, furthermore, easily overlooked in inappropriate asymptotic frameworks; see Leeb and Pötscher (2005) for some cautionary tales. These problems need not arise when superefficiency is assumed, and they do not affect CI_α^1. Having said that, imposing superefficieny by assumption raises questions about generality. It fails when $\widehat{\theta}_l$ and $\widehat{\theta}_u$ come from individual moment conditions (as in the ATM example of Pakes et al. 2007) or when worst-case bounds are tightened by means of testable assumptions (i.e.

---

[4]This paper makes heavy use of local parameters, and to minimize confusion, I reserve the subscript $(\cdot)_N$ for deterministic functions of N, including local parameters; hence the use of c_α where IM used C_N. Estimators are denoted by $\widehat{(\cdot)}$ throughout.

the refined bounds may overlap). More generally, one might wonder whether CI_α^1 applies in cases that interestingly generalize the mean with missing data.

I now turn to answering these questions. Regarding the range of applications of CI_α^1, the supereficiency condition can be weakened and is then implied whenever $\widehat{\theta}_u \geq \widehat{\theta}_l$ by construction. Regarding generality, CI_α^1 is not valid without superefficiency, but one can construct an alternative interval that is easily adapted to settings with and without superefficiency.

3   Re-analysis of the Inference Problem

It is instructive to begin by assuming that Δ_N is known.

Assumption 2 (i) *There exists an estimator $\widehat{\theta}_l$ 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 $P \in \mathcal{P}$.

(ii) $\Delta_N \geq 0$ *is known.*

(iii) *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 $\overline{\sigma}^2$.*

Applications of this scenario include inference about the mean from interval data, where the length of intervals (e.g., income brackets) does not vary on the support of θ, as well as worst-case bounds on the average treatment effect when potential outcomes are supported on $[0,1]$. Define

$$\widetilde{CI}_\alpha = \left[\widehat{\theta}_l - \frac{\widetilde{c}_\alpha \widehat{\sigma}_l}{\sqrt{N}}, \widehat{\theta}_u + \frac{\widetilde{c}_\alpha \widehat{\sigma}_l}{\sqrt{N}}\right],$$

where

$$\Phi\left(\widetilde{c}_\alpha + \frac{\sqrt{N}\Delta_N}{\widehat{\sigma}_l}\right) - \Phi\left(-\widetilde{c}_\alpha\right) = 1 - \alpha. \tag{3}$$

Lemma 2 establishes that this confidence interval is uniformly valid.

Lemma 2 *Let assumption 2 hold. Then*

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

This result generalizes IM's lemma 3 from means with missing data to the present setting It has the same proof idea: The normal approximation to $\Pr\left(\theta_0 \in \widetilde{CI}_\alpha\right)$ is concave in θ_0 and equals $(1 - \alpha)$ if $\theta_0 \in \{\theta_l, \theta_u\}$. The lemma's main purpose is as a backdrop for the case with unknown Δ_N, when $\widetilde{CI}_\alpha$ is not feasible; pre-estimation of Δ_N will turn out to be the root cause of most problems. I turn to this case now.

Thus, impose assumption 1(i)-(ii), drop assumption 1(iii), but consider the following.

Assumption 3 *There exists a sequence $\{a_N\}$ s.t. $a_N \to 0$, $a_N\sqrt{N} \to \infty$, and $\sqrt{N}\left|\widehat{\Delta} - \Delta_N\right| \xrightarrow{p} 0$ for all sequences of distributions $\{P_N\} \subseteq \mathcal{P}$ with $\Delta_N \leq a_N$.*

Assumption 3 is again a superefficiency condition, although more transparent than assumption 1(iii). More importantly, it is weaker: By lemma 1, assumption 1(iii) implies that the above holds with the quantifier "for all sequences $\{a_N\}$...". One consequence of this weakening is the availability of a simple sufficient condition for superefficiency.[5]

Lemma 3 *Let assumption 1(i)-(ii) hold and assume that $\Pr(\widehat{\theta}_u \geq \widehat{\theta}_l) = 1$ for all $\mathcal{P}$. Then assumption 3 is implied.*

Thus, assumption 3 obtains whenever $(\widehat{\theta}_l, \widehat{\theta}_u)$ are jointly asymptotically normal and are almost surely ordered, in particular when they are ordered by construction. This immediately verifies superefficiency for a good number of applications including worst-case bounds on smooth functions of population moments or (under regularity conditions) quantiles, where in either case, partial identification could be due either to missing data or to interval observations with (a priori) unknown length of intervals. Assumption 1(iii) is not similarly implied and accordingly harder to verify (although ultimately fulfilled in the examples just given). Nonetheless, assumption 3 suffices for CI_α^1 to be valid.

Proposition 1 *Let assumptions 1(i)-(ii) and 3 hold. Then*

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

Proposition 1 strengthens IM's main result; furthermore, a clear understanding of its reliance on superefficiency informs a concise and intuitive proof. Think of CI_α^1 as feasible version of $\widetilde{CI}_\alpha$, with c_α^1 being an estimator of $\widetilde{c}_\alpha$. Validity of CI_α^1 would easily follow from consistency of c_α^1, but such consistency seems to fail: $\left(\widehat{\Delta} - \Delta\right)$ is usually of order $O(N^{-1/2})$, so that $\left(\sqrt{N}\widehat{\Delta} - \sqrt{N}\Delta\right)$ does not vanish. This is where superefficiency comes into play. Think in terms of local parameters Δ_N, and distinguish between sequences where Δ_N vanishes fast enough for assumption 3 to apply and sequences where this fails. In the former case, $\left(\sqrt{N}\widehat{\Delta} - \sqrt{N}\Delta\right)$ does vanish, so the asymptotics are as if Δ were known. In the latter case, Δ_N grows large relative to sampling error, so that the uniformity problem does not arise to begin with.

The intuition highlights one channel through which superefficiency simplifies the analysis, namely the vanishing of sampling variation in Δ_N. There is a second, more subtle channel: Asymptotically, $\widehat{\theta}_u = \widehat{\theta}_l + \Delta_N$, meaning that the limiting sampling distribution is univariate. This simplification allows IM to calibrate c_α through a single equation (2), even though the estimation problem is generally

---

[5] I thank Thierry Magnac for suggesting this result.

bivariate. One can avoid reliance on this simplification as follows: Let (c_l^2, c_u^2) minimize $(\widehat{\sigma}_l c_l + \widehat{\sigma}_u c_u)$ subject to the constraint that

$$\Pr\left(-c_l \le z_1, \widehat{\rho} z_1 \le c_u + \frac{\sqrt{N}\widehat{\Delta}}{\widehat{\sigma}_u} + \sqrt{1-\widehat{\rho}^2} z_2\right) \ge 1 - \alpha \tag{4}$$

$$\Pr\left(-c_l - \frac{\sqrt{N}\widehat{\Delta}}{\widehat{\sigma}_l} - \sqrt{1-\widehat{\rho}^2} z_2 \le \widehat{\rho} z_1, z_1 \le c_u\right) \ge 1 - \alpha, \tag{5}$$

where z_1 and z_2 are independent standard normal random variables.[6] In typical cases, (c_l^2, c_u^2) will be uniquely characterized by the fact that both of (4,5) hold with equality, but it is conceivable that one condition is slack at the solution. Let

$$CI_\alpha^2 \equiv \left[\widehat{\theta}_l - \frac{\widehat{\sigma}_l c_l^2}{\sqrt{N}}, \widehat{\theta}_u + \frac{\widehat{\sigma}_u c_u^2}{\sqrt{N}}\right].$$

Then the following holds.

Proposition 2 *Let assumptions 1(i)-(ii) and 3 hold. Then*

$$\lim_{N\to\infty} \inf_{\theta\in\Theta} \inf_{P:\theta_0(P)=\theta} \Pr\left(\theta_0 \in CI_\alpha^2\right) = 1 - \alpha.$$

Calibration of (c_l^2, c_u^2) takes into account that the underlying estimation problem is bivariate. Expression (4) ensures that the nominal size of CI_α^2 equals at least $(1-\alpha)$ if $\theta_0 = \theta_l$, equation (5) does the same if $\theta_0 = \theta_u$, where equality obtains if the according condition binds.[7] In contrast, expression (2) enforces $c_l = c_u$ and properly calibrates both only if $\sigma_l = \sigma_u$ and $\rho = 1$. In general, the nominal size of CI_α^1 for any finite N exceeds $(1-\alpha)$ at one end and falls short of it at the other one, so that the interval appears invalid. This does not affect first-order asymptotics because for large Δ_N, the testing problem is asymptotically one-sided, whereas for small Δ_N, superefficiency ensures that $\sigma_l = \sigma_u$ and $\rho = 1$ in the limit. Nonetheless, the nominal size of CI_α^1 really converges to $(1-\alpha)$ rather than equalling it, and CI_α^1 corresponds to a biased hypothesis test, i.e. Θ_0 is not an upper contour set of nominal coverage probability. The construction of CI_α^2 avoids these issues; indeed, CI_α^2 is literally defined as the shortest confidence interval with correct nominal size.

The improvement may be mostly conceptual if superefficiency obtains, because CI_α^1 and CI_α^2 are then asymptotically equivalent. It becomes more serious when superefficiency fails. The reason is that validity of CI_α^2 depends on assumption 3 only through the first channel mentioned above. This dependency can be eliminated by a straightforward modification: Consider a shrinkage estimator

$$\Delta^* \equiv \begin{cases} \widehat{\Delta}, & \widehat{\Delta} > b_N \\ 0 & \text{otherwise} \end{cases},$$

---

[6] Appendix B exhibits closed-from expressions for (4,5), illustrating that they can be evaluated without simulation.

[7] By the nominal size of CI at θ_l, say, I mean $\int_{CI} \frac{1}{\widehat{\sigma}_l}\phi((x - \hat{\theta}_l)/\hat{\sigma}_l)dx$, i.e. its size at θ_l as predicted from sample data. This size would be exact in finite samples if normal approximations were perfect and variances known. Confidence regions are typically constructed by setting it equal to $1 - \alpha$.

where b_N is some pre-assigned sequence s.t. $b_N \to 0$ and $b_N\sqrt{N} \to \infty$. By replacing $\widehat{\Delta}$ with Δ^* in the calibration of (c_l^2, c_u^2), one can artificially restore superefficiency where it is needed. Of course, this comes at the aforementioned cost of inducing superefficiency, more on which below.

Let (c_l^3, c_u^3) minimize $(\widehat{\sigma}_l c_l + \widehat{\sigma}_u c_u)$ subject to the constraint that

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

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

where z_1 and z_2 are as before, and define

$$CI_\alpha^3 \equiv \begin{cases} \left[\widehat{\theta}_l - \frac{\widehat{\sigma}_l c_l^3}{\sqrt{N}}, \widehat{\theta}_u + \frac{\widehat{\sigma}_u c_u^3}{\sqrt{N}}\right], & \widehat{\theta}_l - \frac{\widehat{\sigma}_l c_l^3}{\sqrt{N}} \leq \widehat{\theta}_u + \frac{\widehat{\sigma}_u c_u^3}{\sqrt{N}} \\ \varnothing & \text{otherwise} \end{cases}.$$

Before discussing CI_α^3 further, I state this paper's final result.

Proposition 3 *Let assumption 1(i)-(ii) hold. Then*

$$\lim_{N\to\infty} \inf_{\theta\in\Theta} \inf_{P:\theta_0(P)=\theta} \Pr\left(\theta_0 \in CI_\alpha^3\right) = 1-\alpha.$$

The definition of CI_α^3 reveals an additional modification: If $\widehat{\theta}_u$ is too far below $\widehat{\theta}_l$, the interval is empty, which can be interpreted as rejection of the maintained assumption that $\theta_u \geq \theta_l$. In other words, CI_α^3 embeds a specification test. IM do not consider such a test, presumably for two reasons: It does not arise in their leading application, and it is trivial under their assumptions or whenever $\widehat{\theta}_u$ and $\widehat{\theta}_l$ are ordered by construction. But it is interesting in applications such as moment inequalities, where $\widehat{\Delta} < 0$ is a generic possibility and samples with $\widehat{\theta}_u$ much below $\widehat{\theta}_l$ might raise doubts whether $\theta_u \geq \theta_l$. In other cases, the possibility of CI_α^3 being empty may be unattractive. It can be avoided by arbitrarily replacing $\varnothing$ in the above definition. One plausible alternative would be a Wald confidence region based on imposing $\theta_u = \theta_l$, i.e. to estimate both θ_l and θ_u by a variance-weighted average $\widehat{\theta}$ of $\widehat{\theta}_l$ and $\widehat{\theta}_u$ and to construct a confidence region by adding and subtracting $\Phi^{-1}(1-\alpha)$ standard errors of $\widehat{\theta}$. Any such enlargement of CI_α^3 will make it potentially conservative if $\Delta = 0$; adjusting (c_l^3, c_u^3) to achieve exact size would cause complications.

Whichever way one resolves this issue, an intriguing aspect of CI_α^3 is that it is analogous to CI_α^2 except for the use of Δ^*. (The event that $\widehat{\theta}_l - \widehat{\sigma}_l c_l^3/\sqrt{N} > \widehat{\theta}_u + \widehat{\sigma}_u c_u^3/\sqrt{N}$ uniformly vanishes under superefficiency and is impossible if $\widehat{\theta}_u \geq \widehat{\theta}_l$.) Together, CI_α^2 and CI_α^3 therefore provide a unified approach to inference for interval identified parameters. In contrast, CI_α^1 does not become valid upon replacing $\widehat{\Delta}$ with Δ^* in (3).

Some further remarks on these results are in order. First, inspection of proofs reveals that propositions 1 and 2 do not really require joint (as opposed to marginal) normality of estimators. Hence IM's

result can be strengthened even further, although the additional gain in generality may be marginal. Furthermore, joint normality cannot be dropped from lemma 3, so it is still needed if one wishes to invoke the simple sufficient condition. Joint normality is also required to claim that CI_α^2 has exact nominal size and corresponds to a nominally unbiased test.

Second, two ways in which Δ^* can be modified are as follows. I defined a soft thresholding estimator for simplicity, but making Δ^* a smooth function of $\widehat{\Delta}$ would also ensure validity and might improve performance for Δ close to b_N. Also, the sequence b_N is left to adjustment by the user. While the law of the iterated logarithm makes $b_N = (\log \log N)^{1/2}$ a salient choice, such adjustment is generally subject to the following trade-off: The slower b_N vanishes, the less distortion is caused by shrinkage, but the quality of uniform normal approximations deteriorates, and they break down for $b_N = O(N^{-1/2})$. I do not expand on these possibilities to avoid redundancy with independent work by Andrews and Soares (2007).

A modification of CI_α^3 that is not recommended is to let $c_l^3 = \Phi^{-1}(1 - 2\alpha)$ and $c_u^3 = \Phi^{-1}(1 - 2\alpha)$, implying that $CI_\alpha^3 = CI_{2\alpha}$, whenever $\Delta^* > 0$. This would make CI_α^3 shorter without affecting first-order asymptotics, because shrinking $\widehat{\Delta}$ suffices to ensure uniformity. But the new interval ignores the two-sided nature of noncoverage risk and hence has nominal size below $1 - \alpha$ (although approaching $1 - \alpha$ as N grows large). The improvement in length is, therefore, spurious.

Third, $\Delta^* = 0$ can be interpreted as failure of a pre-test to reject $H_0 : \theta_u = \theta_l$, where the size of the pre-test approaches 1 as $N \to \infty$. Thus, shrinking $\widehat{\Delta}$ resembles the "conservative pre-test" solution to the parameter-on-the-boundary problem given by Andrews (2000, section 4). Nonetheless, one should not interpret CI_α^3 as being based on model selection. If $\theta_u = \theta_l$ were known, the aforementioned Wald confidence region would be efficient, and a post-model selection confidence region would use it – but it would be invalid if $\Delta_N = O(N^{-1/2})$. CI_α^3 avoids the trap by employing a shrinkage estimator of Δ in the pre-test but not in the subsequent construction of the interval. Having said that, there is a tight connection between problems encountered here and known issues with post-model selection estimators (Leeb and Pötscher 2005), the underlying problem being discontinuity of pointwise limit distributions. In particular, now that superefficiency of Δ^* is induced rather than assumed, it does follow that Δ^* estimates Δ_N with large local risk near zero. Uniform validity of CI_α^3 accordingly comes at the price of asymptotic dissimilarity, i.e. the interval is conservative along certain local parameter sequences. This feature cannot be avoided by any of the modifications mentioned above (or in Andrews 2000 or Andrews and Soares 2007).

Finally, this paper's assumptions improve on preceding work but still entail substantial loss of generality. Uniform validity of normal approximations could be replaced by uniform validity of the bootstrap after minor adjustments, and $\sqrt{N}$-consistency plays no special role in and of itself. But these modifications will not help if upper and lower bounds are characterized as minima respectively

9

maxima over a number of moment conditions (as in the hospitals example of Pakes et al. 2007; but, see subsequent work by Chernozhukov et al. 2008). Also, it is not obvious how to generalize CI_α^3 to a multivariate moment conditions (but, see subsequent work by Fan and Park 2007). Other, independently developed methods (Andrews and Guggenberger 2007, Andrews and Soares 2007) will work more generally. When CI_α^3 applies, however, it is attractive for numerous reasons: It is trivial to compute, it is by construction the shortest interval whose nominal size is exact at both ends, and it is is easily adjusted to the presence or absence of superefficiency.

4 Conclusion

This paper extended Imbens and Manski's (2004) analysis of confidence regions for partially identified parameters. Core findings are as follows. I establish that one assumption used for IM's final result boils down to locally superefficient estimation of a nuisance parameter. This fact appears to have gone unnoticed before. The inference problem is then re-analyzed with and without superefficiency. IM's confidence region is found to be valid under weaker conditions than theirs, a sufficient condition being that estimators of bounds are jointly asymptotically normal and ordered by construction. Furthermore, valid inference can be achieved by a confidence region that corresponds to a nominally unbiased hypothesis test, is easily adapted to the case without superefficiency, and embeds a specification test.

A conceptual contribution is to recognize that much of the inference problem stems from pre-estimation of Δ. This insight allows for brief and transparent proofs and clarifies the connection to related work. For example, once the boundary problem is recognized, analogy to Andrews (2000) suggests that a straightforward normal approximation, as well as the bootstrap, will fail, whereas subsampling might work. Indeed, carefully specified subsampling techniques are known to yield valid inference for parameters identified by moment inequalities, of which the present scenario is a special case (Andrews and Guggenberger 2007, Chernozhukov et al. 2007, Romano and Shaikh 2008). The bootstrap, on the other hand, does not work in the same setting, unless it is modified in several ways, one of which is analogous to shrinking $\widehat{\Delta}$ (Bugni 2007). Against the backdrop of these (subsequent to IM) results, validity of simple normal approximations in IM appears as a puzzle that is now resolved. At the same time, the updated version of these normal approximations has practical value because it provides closed-form and otherwise attractive inference for important, if relatively simple, applications.

A Proofs

Preliminaries Most proofs consider sequences $\{P_N\}$ that will be identified with the implied sequences $\{\Delta_N, \theta_N\} \equiv \{\Delta(P_N), \theta_0(P_N)\}$. For ease of notation, I suppress the N subscript on $(\theta_l, \sigma_l, \sigma_u)$

and on estimators. Some algebraic steps treat $(\theta_l, \sigma_l, \sigma_u)$ as constant; this is w.l.o.g. because by compactness implied in assumption 1(ii), any sequence $\{P_N\}$ induces a sequence of values $(\theta_l, \sigma_l, \sigma_u)$ with finitely many accumulation points, and the argument can be conducted separately for the according subsequences.

Proofs of lemma 3 and the propositions establish that $\inf_{\{\theta_N\} \subseteq \Theta} \inf_{\{P_N\}: \theta_N \in \Theta_0(P_N)} \Pr\left(\theta_N \in CI_\alpha^i\right) \to$ $1 - \alpha, i = 1, 2, 3$. These are pointwise limits, but they imply the claims because they are taken over sequences, in particular they apply along least favorable sequences. Proofs present two arguments, one for the case that $\{\Delta_N\}$ is "small" and one for the case that it is "large" in a sense that will be delimited. Any sequence $\{P_N\}$ can be decomposed into one large and one small subsequence.

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_0, $v > 0$, and K s.t.

$$N \geq N_0 \Longrightarrow \Pr\left(\sqrt{N}\left|\widehat{\Delta} - \Delta_N\right| > K\Delta_N^v\right) < \varepsilon$$

uniformly over $\mathcal{P}$ (and hence Δ_N). As $\Delta_N \to 0$, one can choose N_1 s.t. $N \geq N_1 \Rightarrow \Delta_N \leq \delta^{1/v} K^{-1/v} \Rightarrow$ $K\Delta_N^v \leq \delta$. Combining these and choosing $N^* \equiv \max\{N_0, N_1\}$ yields

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

as required.

Lemma 2 Parameterize θ_0 as $\theta_0 = \theta_l + a\Delta_N$ for some $a \in [0,1]$. Then algebra reveals that

$$
\begin{aligned}
&\Pr\left(\theta_0 \in \widetilde{CI}_\alpha\right) \\
=\ &\Pr\left(-\tilde{c}_\alpha \frac{\widehat{\sigma}_l}{\sigma_l} - \frac{\sqrt{N}}{\sigma_l} a\Delta_N \leq \frac{\sqrt{N}\left(\theta_l - \widehat{\theta}_l\right)}{\sigma_l} \leq \frac{\sqrt{N}}{\sigma_l}(1-a)\Delta_N + \tilde{c}_\alpha \frac{\widehat{\sigma}_l}{\sigma_l}\right) \\
\to\ &\Phi\left(\tilde{c}_\alpha + \frac{\sqrt{N}}{\sigma_l}(1-a)\Delta_N\right) - \Phi\left(-\tilde{c}_\alpha - \frac{\sqrt{N}}{\sigma_l} a\Delta_N\right)
\end{aligned}
$$

uniformly over $\mathcal{P}$. Besides uniform asymptotic normality of $\widehat{\theta}_l$, this convergence statement uses that by uniform consistency of $\widehat{\sigma}_l$ in conjunction with the lower bound on σ_l, $\widehat{\sigma}_l/\sigma_l \to 1$ uniformly, and also that the derivative of the standard normal c.d.f. is uniformly bounded.

Evaluation of derivatives establishes that the last expression in the preceding display is strictly concave in a, hence it is minimized at $a \in \{0, 1\}$. But in those cases, the algebra simplifies to

$$\Pr\left(\theta_l \in \widetilde{CI}_\alpha\right) \to \Phi\left(\frac{\sqrt{N}}{\sigma_l}\Delta_N + \tilde{c}_\alpha\right) - \Phi\left(-\tilde{c}_\alpha\right) = 1 - \alpha$$

and similarly for θ_u.

Lemma 3 By assumption 3, $\sqrt{N}(\widehat{\Delta}-\Delta_N) \to N(0,\sigma_\Delta^2)$ uniformly in $\mathcal{P}$, where $\sigma_\Delta^2 \equiv \sigma_l^2+\sigma_u^2-2\rho\sigma_l\sigma_u$. Hence, one can fix a sequence $\varepsilon_N \to 0$ s.t. $\sup_{P\in\mathcal{P},d\in\mathbb{R}} \left|\Pr(\sqrt{N}(\widehat{\Delta}-\Delta_N) \le d) - \Phi(d/\sigma_\Delta^2)\right| \le \varepsilon_N$ for all N. (The lemma restates lemma 2 if $\sigma_\Delta^2 = 0$, so this case can be ignored.) Fix a nonpositive sequence $\delta_N \to -\infty$ s.t. $\Phi(\gamma\delta_N) > O(\varepsilon_N)$ for any fixed $\gamma \ge 0$. This is possible because of well known uniform bounds on the standard normal c.d.f., e.g. $\Phi(\gamma\delta_N) > -(\gamma\delta_N)^{-1}(2\pi)^{-1/2}\exp(-(\gamma\delta_N)^2/2)$ as $\delta_N \to -\infty$; using this bound, one can verify that $\delta_N = -(\log(-\log\varepsilon_N))^{1/2}$ will do.

Fix any sequence P_N s.t. $\Delta_N \le a_N \equiv -\delta_N N^{-1/2}$. I will show that $\sigma_\Delta^2 \to 0$, implying assumption 3. Assume this fails, then σ_Δ^2 must have an accumulation point $\sigma_{\Delta\infty}^2 > 0$. Along any subsequence converging to $\sigma_{\Delta\infty}^2$, one would have

$$\Pr\left(\widehat{\theta}_u \le \widehat{\theta}_l\right) = \Pr\left(\widehat{\Delta} \le 0\right) = \Pr\left(\sqrt{N}(\widehat{\Delta}-\Delta_N) \le -\sqrt{N}\Delta_N\right)$$
$$\ge \Phi\left(-\sqrt{N}\Delta_N/\sigma_{\Delta\infty}\right) - \varepsilon_N \ge \Phi(\delta_N/\sigma_{\Delta\infty}) - \varepsilon_N > 0$$

for N large enough, a contradiction. Note how the conclusion of lemma 2 (and hence, assumption 1(iii)) is not implied because for $\Delta_N > a_N$, the second inequality above might fail.

Proposition 1 Let $\Delta_N \le a_N$, then $\sqrt{N}\left|\widehat{\Delta}-\Delta_N\right| \overset{p}{\to} 0$ by assumption 3, thus

$$\sqrt{N}\left(\widehat{\theta}_u - \theta_u\right) = \sqrt{N}\left(\widehat{\theta}_l + \widehat{\Delta} - \theta_l - \Delta_N\right) \overset{p}{\to} \sqrt{N}\left(\widehat{\theta}_l - \theta_l\right).$$

In conjunction with conditions (i)-(ii), this implies $\sigma_u = \sigma_l$, hence $\widehat{\sigma}_u - \widehat{\sigma}_l \overset{p}{\to} 0$. Also using assumption 3 again, it follows that

$$\Phi\left(c_\alpha^1 + \sqrt{N}\frac{\widehat{\Delta}}{\max\{\widehat{\sigma}_l,\widehat{\sigma}_u\}}\right) \overset{p}{\to} \Phi\left(c_\alpha^1 + \sqrt{N}\frac{\Delta_N}{\widehat{\sigma}_l}\right)$$

and the argument can be completed as in lemma 2.

Let $\Delta_N > a_N$, then $\sqrt{N}\Delta_N \to \infty$, hence $\limsup_{N\to\infty} \sqrt{N}(\theta_N-\theta_l) = \infty$ or $\limsup_{N\to\infty} \sqrt{N}(\theta_u-\theta_N) = \infty$ or both. Some algebra reveals that

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

Assume $\limsup_{N\to\infty} \sqrt{N}(\theta_N-\theta_l) < \infty$. By consistency of $\widehat{\Delta}$, divergence of $\sqrt{N}\Delta_N$ implies divergence in probability of $\sqrt{N}\widehat{\Delta}$. Thus

$$\Pr\left(\sqrt{N}(\theta_N-\theta_l) + \sqrt{N}\left(\theta_l-\widehat{\theta}_l\right) > \sqrt{N}\widehat{\Delta} + c_\alpha^1\widehat{\sigma}_u\right)$$
$$\le \Pr\left(\sqrt{N}\left(\theta_l-\widehat{\theta}_l\right) > \sqrt{N}\widehat{\Delta} - \sqrt{N}(\theta_N-\theta_l)\right) \to 0,$$

where I used that $c_\alpha^1 \widehat{\sigma}_u \geq 0$ by construction and that $\sqrt{N}\left(\theta_l - \widehat{\theta}_l\right)$ converges to a random variable by assumption. It follows that

$$\lim_{N\to\infty} \Pr\left(\theta_N \in CI_\alpha^1\right) = \lim_{N\to\infty} \Pr\left(-c_\alpha^1 \widehat{\sigma}_l \leq \sqrt{N}\left(\theta_N - \theta_l\right) + \sqrt{N}\left(\theta_l - \widehat{\theta}_l\right)\right)$$

$$\geq \lim_{N\to\infty} \Pr\left(-c_\alpha^1 \widehat{\sigma}_l \leq \sqrt{N}\left(\theta_l - \widehat{\theta}_l\right)\right) = 1 - \Phi(c_\alpha^1) \geq 1 - \alpha,$$

where the first inequality uses $\sqrt{N}\left(\theta_N - \theta_l\right) \geq 0$, and the second inequality uses the definition of c_α^1 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. $\sqrt{N}\left(\theta_N - \theta_u\right)$ fails to diverge, the argument is symmetric. If both diverge, coverage probability converges to 1. To see that a coverage probability of $1 - \alpha$ can be attained, consider the case of $\Delta = 0$.

Proposition 2   For a short proof that does not use joint normality, inspection of (4-5) reveals that CI_α^2 is asymptotically equivalent to CI_α^1 given local superefficiency. The longer argument below shows why CI_α^2 generally has exact nominal size and will also be needed for proposition 3. Let $(\widetilde{c}_l, \widetilde{c}_u)$ fulfil

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

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

and write

$$\lim_{N\to\infty} \Pr\left(\theta_l \in \left[\widehat{\theta}_l - \frac{\widetilde{c}_l \widehat{\sigma}_l}{\sqrt{N}}, \widehat{\theta}_u + \frac{\widetilde{c}_u \widehat{\sigma}_u}{\sqrt{N}}\right]\right)$$

$$= \lim_{N\to\infty} \Pr\left(-\widetilde{c}_l \frac{\widehat{\sigma}_l}{\sigma_l} \leq \frac{\sqrt{N}}{\sigma_l}\left(\theta_l - \widehat{\theta}_l\right), \frac{\sqrt{N}}{\sigma_l}\left(\theta_u - \widehat{\theta}_u\right) \leq \frac{\sqrt{N}}{\sigma_l}\Delta + \widetilde{c}_u \frac{\widehat{\sigma}_u}{\sigma_l}\right)$$

$$= \lim_{N\to\infty} \Pr\left(-\widetilde{c}_l \leq \frac{\sqrt{N}}{\sigma_l}\left(\theta_l - \widehat{\theta}_l\right), \frac{1}{\sigma_l}\left(\rho\frac{\sigma_u}{\sigma_l}\sqrt{N}\left(\theta_l - \widehat{\theta}_l\right) - \sigma_u\sqrt{1-\rho^2}z_2\right) \leq \frac{\sqrt{N}}{\sigma_l}\Delta + \widetilde{c}_u \frac{\sigma_u}{\sigma_l}\right)$$

$$= \lim_{N\to\infty} \Pr\left(-\widetilde{c}_l \leq z_1, \rho z_1 \leq \widetilde{c}_u + \frac{\sqrt{N}}{\sigma_u}\Delta + \sqrt{1-\rho^2}z_2\right) \tag{6}$$

$$\geq 1-\alpha \tag{7}$$

uniformly. Here, the first step can be verified algebraically; the second step uses that by assumption,

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

uniformly, that $(\widehat{\sigma}_l, \widehat{\sigma}_u)$ are uniformly consistent, and that $\sigma_l, \sigma_u \geq \underline{\sigma}$, so that neither can vanish; and the third step uses convergence of $\frac{\sqrt{N}}{\sigma_l}\left(\theta_l - \widehat{\theta}_l\right)$. The argument for θ_u is similar; note that in contrast to the very first step of proposition 1, assumption 3 was not used.

13

As before, for any sequence $\{\Delta_N\}$ s.t. $\Delta_N < a_N$, superefficiency implies that (c_l^2, c_u^2) is consistent for $(\widetilde{c}_l, \widetilde{c}_u)$ and CI_α^2 therefore valid at $\{\theta_l, \theta_u\}$. Convexity of coverage probability over $[\theta_l, \theta_u]$ follows as before. For $\Delta_N \geq a_N$, the argument entirely resembles proposition 1. Finally, (7) will bind if (4) binds, implying that CI_α^2 will then have exact nominal size at θ_l. A similar argument applies for θ_u. But (c_l^2, c_u^2) can minimize $(c_l\widehat{\sigma}_l + c_u\widehat{\sigma}_u)$ subject to (4-5) only if at least one of (4-5) binds, hence CI_α^2 is nominally exact.

Proposition 3 Let $c_N \equiv \left(N^{-1/2} b_N\right)^{1/2}$, thus $N^{1/2} c_N = \left(N^{1/2} b_N\right)^{1/2} \to \infty$, and for parameter sequences s.t. $\Delta_N > c_N$, the proof is again as before. For the other case, consider $(\widetilde{c}_l, \widetilde{c}_u)$ as defined in the previous proof. By uniform convergence of estimators and uniform bounds on (σ_l, σ_u), $\Pr(\widehat{\Delta} \leq b_N)$ is uniformly asymptotically bounded below by

$$\Phi\left(\sqrt{N}\left(b_N - c_N\right)/2\overline{\sigma}\right) = \Phi\left(\left(N^{1/2} b_N - \left(N^{1/2} b_N\right)^{1/2}\right)/2\overline{\sigma}\right) \to 1.$$

Hence, $\Delta^* = 0 \leq \Delta$ with probability approaching 1. Expression (6) is easily seen to increase in Δ for every $(\widetilde{c}_l, \widetilde{c}_u)$, hence CI_α^3 is valid (if potentially conservative) at θ_l. The argument for θ_u is similar.

Now parameterize the true parameter value as $\theta_N \equiv a\theta_l + (1-a)\theta_u$ for some $a \in [0, 1]$, then some algebra yields

$$\Pr\left(\theta_N \in \left[\widehat{\theta}_l - \frac{c_l^3 \widehat{\sigma}_l}{\sqrt{N}}, \widehat{\theta}_u + \frac{c_u^3 \widehat{\sigma}_u}{\sqrt{N}}\right]\right)$$
$$= \Pr\left(\sqrt{N}(1-a)\left(\Delta - \widehat{\Delta}\right) - \sqrt{N}(1-a)\Delta - c_l^3 \widehat{\sigma}_l\right.$$
$$\left. \leq a\sqrt{N}\left(\theta_l - \widehat{\theta}_l\right) + (1-a)\sqrt{N}\left(\theta_u - \widehat{\theta}_u\right) \leq \sqrt{N} a\left(\widehat{\Delta} - \Delta\right) + \sqrt{N} a\Delta + c_u^3 \widehat{\sigma}_u\right).$$

Consider varying Δ, holding $(\theta_l, \sigma_l, \sigma_u, \rho, a)$ constant. The cutoff values c_l^3 and c_u^3 depend on Δ only through Δ^*, but recall that $\Pr(\Delta^* = 0) \to 1$. The estimators $(\widehat{\sigma}_l, \widehat{\sigma}_u)$ are uniformly consistent, and the joint limiting distribution of all other random variables depends only on $(\sigma_l, \sigma_u, \rho, a)$. Hence, the preceding probability's limit is minimized at $\Delta = 0$, in which case $\theta_N = \theta_l$ and coverage was already shown.

B   Closed-Form Expressions for (c_l, c_u)

This appendix provides a closed-form equivalent of (4,5). These expressions can be written as

$$\int_{-\infty}^{c_l} \Phi\left(\frac{\widehat{\rho}}{\sqrt{1 - \widehat{\rho}^2}} z + \frac{c_u}{\sqrt{1 - \widehat{\rho}^2}} + \frac{\sqrt{N}\widehat{\Delta}}{\widehat{\sigma}_u \sqrt{1 - \widehat{\rho}^2}}\right) d\Phi(z) \geq 1 - \alpha$$

$$\int_{-\infty}^{c_u} \Phi\left(\frac{\widehat{\rho}}{\sqrt{1 - \widehat{\rho}^2}} z + \frac{c_l}{\sqrt{1 - \widehat{\rho}^2}} + \frac{\sqrt{N}\widehat{\Delta}}{\widehat{\sigma}_l \sqrt{1 - \widehat{\rho}^2}}\right) d\Phi(z) \geq 1 - \alpha$$

if $-1 < \widehat{\rho} < 1$,

$$\Phi(c_l) - \Phi\left(-c_u - \frac{\sqrt{N}\widehat{\Delta}}{\widehat{\sigma}_u}\right) \geq 1 - \alpha$$

$$\Phi(c_u) - \Phi\left(-c_l - \frac{\sqrt{N}\widehat{\Delta}}{\widehat{\sigma}_l}\right) \geq 1 - \alpha$$

if $\widehat{\rho} = 1$ (compare these expressions to (2)), and

$$\Phi\left(\min\left\{c_l, c_u + \frac{\sqrt{N}\widehat{\Delta}}{\widehat{\sigma}_u}\right\}\right) \geq 1 - \alpha$$

$$\Phi\left(\min\left\{c_u, c_l + \frac{\sqrt{N}\widehat{\Delta}}{\widehat{\sigma}_l}\right\}\right) \geq 1 - \alpha,$$

implying that $\Phi(c_l) = \Phi(c_u) = 1 - \alpha$ at the minimization problem's solution, if $\widehat{\rho} = -1$. There is no discontinuity at the limit because $\Phi\left(\frac{\widehat{\rho}}{\sqrt{1-\widehat{\rho}^2}}z + \frac{c_u}{\sqrt{1-\widehat{\rho}^2}} + \frac{\sqrt{N}\widehat{\Delta}}{\widehat{\sigma}_u\sqrt{1-\widehat{\rho}^2}}\right) \to \mathbb{I}\left\{z \geq [\leq]c_u + \frac{\sqrt{N}\widehat{\Delta}}{\widehat{\sigma}_u}\right\}$ as $\widehat{\rho} \to 1[-1]$.

References

[1] Andrews, D.W.K. (2000): Inconsistency of the Bootstrap when a Parameter is on the Boundary of the Parameter Space, Econometrica, 68, 399-405.

[2] Andrews, D.W.K., and P. Guggenberger (2007): Validity of Subsampling and 'Plug-in Asymptotic' Inference for Parameters Defined by Moment Inequalities, Cowles Foundation Discussion Paper 1620.

[3] Andrews, D.W.K., and G. Soares (2007): Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection, Cowles Foundation Discussion Paper 1631.

[4] Bugni, F.A. (2007): Bootstrap Inference in Partially Identified Models, Unpublished Manuscript, Northwestern University.

[5] Chernozhukov, V., H. Hong, and E.T. Tamer (2007): Parameter Set Inference in a Class of Econometric Models, Econometrica, 75, 1243-1284.

[6] Chernozhukov, V., S. Lee, and A. Rosen (2008): Intersection Bounds: Estimation and Inference, Unpublished Mansucript, University College London.

[7] Fan, Y., and S.S. Park (2007): Confidence Intervals for Some Partially Identified Parameters, Unpublished Mansucript, Vanderbilt University.

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

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

[10] Manski, C.F. (2003): Partial Identification of Probability Distributions. New York: Springer Verlag.

[11] Pakes, A., J. Porter, K. Ho, and J. Ishii (2006): Moment Inequalities and their Application, Unpublished Mansucript, Harvard University.

[12] Romano, J.P., and A.M. Shaikh (2008): Inference for Identifiable Parameters in Partially Identified Econometric Models, Journal of Statistical Planning and Inference, 138, 2786-2807.

[13] Woutersen, T. (2006): A Simple Way to Calculate Confidence Intervals for Partially Identified Parameters, Unpublished Mansucript, Johns Hopkins University.