Confidence Intervals for Partially Identified Parameters

Guido W. Imbens
Charles F. Manski

THE INSTITUTE FOR FISCAL STUDIES
DEPARTMENT OF ECONOMICS, UCL
cemmap working paper CWP09/03

Confidence Intervals for Partially Identified Parameters*

Guido W. Imbens
UC Berkeley, and NBER

Charles F. Manski
Northwestern University

This Draft: May 29, 2003

Abstract

In the last decade a growing body of research has studied inference on partially identified parameters (e.g., Manski, 1990, 2003). In many cases where the parameter of interest is real-valued, the identification region is an interval whose lower and upper bounds may be estimated from sample data. Confidence intervals may be constructed to take account of the sampling variation in estimates of these bounds. Horowitz and Manski (1998, 2000) proposed and applied interval estimates that asymptotically cover the entire identification region with fixed probability. Here we introduce conceptually different interval estimates that asymptotically cover each element in the identification region with fixed probability (but not necessarily every element simultaneously). We show that these two types of interval estimate are different in practice, the latter in general being shorter. The difference in length (in excess of the length of the identification set itself) can be substantial, and in large samples is comparable to the difference of one – and two–sided confidence intervals. A complication arises from the fact that the simplest version of the proposed interval is discontinuous in the limit case of point identification, leading to coverage rates that are not uniform in important subsets of the parameter space. We develop a modification depending on the width of the identification region that restores uniformity. We show that under some conditions, using the estimated width of the identification region instead of the true width maintains uniformity.

JEL Classification:
Keywords: *Bounds, Partial Identification, Identification Regions, Confidence Intervals*

*We are grateful to to Joerg Stoye for excellent research assistance, to Michael Jansson, Francesca Molinari and Tom Rothenberg for helpful comments, and to the NSF for financial support through grants SES 0136789 (Imbens) and SES 0001436 (Manski).

1 Introduction

In the last decade a growing body of research has studied inference in settings where parameters of interest are partially identified (e.g., Manski, 1990, 2003). Such methods have been applied and extended to a wide variety of settings, including the analysis of labor market programs (Blundell, et al., 2002), interval measurement (Manski and Tamer, 2002), auctions (Haile and Tamer, 2003), the effect of teenage pregnancies on labor market outcomes (Hotz, Mullins and Sanders. 1997, Mullins, 2003), measurement error (Horowitz and Manski, 1995; Bollinger, 1996, Molinari, 2002, Dominitz and Sherman, 2003), and selection problems (Manski, 1990; Lee, 2002). In many cases where the parameter is real-valued, the identification region is an interval whose lower and upper bounds may be estimated from sample data. Confidence intervals (CIs) may be constructed to take account of the sampling variation in estimates of these bounds. Early on, Manski, Sandefur, McLanahan and Powers (1992) computed separate confidence intervals for the lower and upper bounds of the identification regions of such parameters. Subsequently, Horowitz and Manski (1998, 2000) proposed and applied intervals that asymptotically cover the entire identification region with fixed probability. Chernozhukov, Hong, and Tamer (2003) extend this approach of constructing CIs that cover the entire identification region to problems with vector valued parameters and identification regions defined through minimization problems. They also develop a new implementation of such intervals through subsampling bootstrap methods.

Here, we introduce a conceptually different type of confidence interval. Rather than cover the entire identification region with fixed probability, we propose CIs that asymptotically cover the true value of the parameter with fixed probability. We show that, in general, coverage of a parameter is a less demanding objective than is coverage of the entire identification region. We prove that any specified confidence interval has a weakly larger coverage probability for the parameter than for its identification region. It follows that if a given interval achieves a specified coverage probability for the identification region, there exists a subset of this interval that achieves the same coverage probability for the parameter.

To illustrate the basic nature of our CIs for partially identified parameters, and to address some subtleties, we study in depth the construction of CIs for the mean of a bounded random variable when some data are missing and the distribution of missing data is unrestricted (beyond the bounds on their values). Initially we assume that the propensity score (i.e., the probability of observing an outcome) is known. We prove that, for any specified asymptotic coverage probability, CIs for the parameter are proper subsets of ones for the identification region, with the difference in width related to the difference in critical values for one– and two–sided tests. However, we find that the exact coverage probabilities of the simplest version of our new CIs do not converge to their nominal values uniformly across different values for the width of the identification region. Specifically, uniformity fails when the width of the region shrinks to zero; that is, as the parameter becomes point-identified. An unattractive consequence is that confidence intervals can be wider when the parameter is point–identified than when it is set-identified. To avoid this anomaly, we modify the proposed CI to ensure that its exact coverage probabilities do converge uniformly to their nominal values. We motivate the modified CI by showing that its exact and nominal coverage probabilities coincide when outcomes are normally distributed.

We then discuss implementation of the new CIs at a more general level, and provide conditions under which CIs with uniform asymptotic coverage can be constructed by substituting estimates for unknown nuisance parameters, including the width of the identification region. Finally we provide a brief empirical illustration.

2 Confidence Intervals for Parameters and for Their Identification Regions

Many problems of partial identification have the following abstract structure. Let $(\Omega, \mathcal{A}, P)$ be a specified probability space, and let $\mathcal{P}$ be a space of probability distributions on $(\Omega, \mathcal{A})$. The distribution P is not known, but a random sample of size N is available, with empirical distribution P_N. Let λ be a quantity which is known only to belong to a specified set Λ. Let $f(\cdot, \cdot) : \mathcal{P} \times \Lambda \to \mathbb{R}$ be a specified real-valued function. The object of interest is the real parameter $\theta = f(P, \lambda)$. Then the identification region for $f(P, \lambda)$ is the set $\{f(P, \lambda'), \lambda' \in \Lambda\}$. A natural estimate for the identification region is its sample analog $\{f(P_N, \lambda'), \lambda' \in \Lambda\}$.

Suppose that $\lambda_l(P) = \text{argmin}_{\lambda' \in \Lambda} f(P, \lambda')$ and $\lambda_u(P) = \text{argmax}_{\lambda' \in \Lambda} f(P, \lambda')$ exist for all $P \in \mathcal{P}$. Then the identification region necessarily is a subset of the closed interval $[f(P, \lambda_l(P)), f(P, \lambda_u(P))]$. We focus on the class of problems in which the identification region is this closed interval. Manski (2003) describes various problems in this class, including ones that arise when data are missing or contaminated. A particularly simple and important leading case will be examined in detail in Sections 3 and 4.

It is natural to estimate the identification region $[f(P, \lambda_l(P)), f(P, \lambda_u(P))]$ by its sample analog $[f(P_N, \lambda_l(P_N)), f(P_N, \lambda_u(P_N))]$, which is consistent under standard regularity conditions. It is also natural to construct confidence intervals for $[f(P, \lambda_l(P)), f(P, \lambda_u(P))]$ of the form $[f(P_N, \lambda_l(P_N)) - C_{N0}, f(P_N, \lambda_u(P_N)) + C_{N1}]$, where (C_{N0}, C_{N1}) are specified non-negative numbers that may depend on the sample data. In their study of nonparametric regression analysis with missing outcome or covariate data, Horowitz and Manski (2000) proposed CIs of this form and showed how (C_{N0}, C_{N1}) may be chosen to achieve a specified asymptotic probability of coverage of the identification region. Chernozhukov, Hong and Tamer (2003) study confidence sets with the same property in more general settings with vector valued parameters and identification regions defined through minimization of general objective functions.

In this paper, we study the use of intervals of the form $[f(P_N, \lambda_l(P_N)) - C_{N0}, f(P_N, \lambda_u(P_N)) + C_{N1}]$ as CIs for the partially identified parameter $f(P, \lambda)$. Our most basic finding is Lemma 2.1:

Lemma 2.1 *Let $C_{N0} \geq 0$, $C_{N1} \geq 0$, $\lambda \in \Lambda$, and $P \in \mathcal{P}$. The probability that the interval $[f(P_N, \lambda_l(P_N)) - C_{N0}, f(P_N, \lambda_u(P_N)) + C_{N1}]$ covers the parameter $f(P, \lambda)$ is at least as large as* *the probability that it covers the entire identification region* $[f(P,\lambda_l(P)), f(P,\lambda_u(P))]$.

Proof: The coverage probability for the parameter $f(P,\lambda)$ is

$$\alpha_N(P,\lambda) = \Pr\left(f(P,\lambda) \in [f(P_N,\lambda_l(P_N)) - C_{N0}, f(P_N,\lambda_u(P_N)) + C_{N1}]\right) \tag{2.1}$$

$$= Pr\left(f(P_N,\lambda_l(P_N)) \leq f(P,\lambda) + C_{N0} \cap f(P_N,\lambda_u(P_N)) \geq f(P,\lambda) - C_{N1}\right).$$

The coverage probability for the identification region $[f(P,\lambda_l(P)), f(P,\lambda_u(P))]$ is

$$a_N(P) = Pr\left([f(P,\lambda_l(P)), f(P,\lambda_u(P))] \subset [f(P_N,\lambda_l(P_N)) - C_{N0}, f(P_N,\lambda_u(P_N)) + C_{N1}]\right) \tag{2.2}$$

$$= Pr\left(f(P_N,\lambda_l(P_N)) \leq f(P,\lambda_l(P)) + C_{N0} \cap f(P_N,\lambda_u(P_N)) \geq f(P,\lambda_u(P)) - C_{N1}\right).$$

Then $\alpha_N(P,\lambda) \geq a_N(P)$ because

$$f(P_N,\lambda_l(P_N)) \leq f(P,\lambda_l(P)) + C_{N0} \cap f(P_N,\lambda_u(P_N)) \geq f(P,\lambda_u(P)) - C_{N1} \tag{2.3}$$

$$\Longrightarrow f(P_N,\lambda_l(P_N)) \leq f(P,\lambda) + C_{N1} \cap f(P_N,\lambda_u(P_N)) \geq f(P,\lambda) - C_{N1}.$$

□

Lemma 2.1 implies that the coverage probability for the parameter is at least as large as that for the identification region. The coverage probabilities $\alpha_N(P,\lambda)$ and $a_N(P)$ are functions of the unknown quantities (P,λ). The uniform coverage probabilities are

$$\alpha_N = \inf_{(P,\lambda)\in(\mathcal{P}\times\Lambda)} \alpha_N(P,\lambda)$$

$$a_N = \inf_{P\in\mathcal{P}} a_N(P)$$

The lemma implies that $\alpha_N \geq a_N$.

It is common in the construction of CIs to choose an interval that achieves at least a specified coverage probability, say α. Suppose that (C_{N0}, C_{N1}) is chosen to achieve at least coverage probability α for the identification region. Lemma 2.1 implies that there exists a subset of the interval $[f(P_N,\lambda_l) - C_{N0}, f(P_N,\lambda_u) + C_{N1}]$ that achieves at least coverage probability α for the parameter.

Finally, note that the inequalities proved in Lemma 2.1 are weak, not strict. They cannot be strict in general: if the parameter of interest is point-identified the two will be identical. Sections 3 and 4 shows that in settings where the parameter of interest is not point-identified the inequalities are strict given some conditions.

Given the potential differences between the two types of CIs, the researcher faces a substantive choice whether to consider intervals that cover the entire identification region or the true parameter value with some fixed probability. Although generally both intervals converge to the identification region as $N \to \infty$, their differences may be substantial in finite samples and the question cannot be avoided. In general, the answer depends on the application and the focus of the researcher.

3 Means with Missing Data and Known Propensity Score

In this section we construct CIs for the mean of a bounded random variable when some data are missing and the distribution of missing data is unrestricted. Let (Y, W) be a pair of random variables, where Y has compact support $\mathbb{Y}$ and W is binary with support $\{0, 1\}$; without loss of generality, let the smallest and largest elements of $\mathbb{Y}$ be 0 and 1 respectively. The parameter of interest is $\theta = \mathbb{E}[Y]$. The researcher has a random sample of $(W_i, Y_i \cdot W_i)$, $i = 1, \ldots, N$, so W_i is always observed and Y_i is only observed if $W_i = 1$. Let $\mu = \mathbb{E}[Y|W = 1]$ and $\lambda = \mathbb{E}[Y|W = 0]$ be the conditional means of Y in the two subpopulations, let $\sigma^2 = \mathbb{V}(Y|W = 1)$ be the conditional variance in the subpopulation with $W = 1$ and let $p = \mathbb{E}[W]$, with $0 < p \leq 1$, be the propensity score. In this section we assume, for purposes of exposition, that p is known. Later we will allow for unknown p, but assume that it is bounded away from zero by a positive number p_0.

Let $F(y)$ be the conditional distribution function of Y given $W = 1$. The distribution function $F(\cdot)$ is in the set of distribution functions $\mathcal{F}$ with variance $\underline{\sigma}^2 \leq \sigma^2 \leq \overline{\sigma}^2$ where $\underline{\sigma}^2$ and $\overline{\sigma}^2$ are known positive lower and upper bounds on the conditional variance of Y given $W = 1$. The conditional distribution of Y given $W = 0$ is unknown other than that $\lambda = \mathbb{E}[Y|W = 0]$ is in the interval $\Lambda = [0, 1]$. Given these definitions, the parameter of interest can be written as $\theta = \mu \cdot p + \lambda \cdot (1 - p)$. The identification region for θ is the closed interval

$$[\theta_l, \theta_u] = [\mu \cdot p, \mu \cdot p + 1 - p] .$$

With the probability p of observing Y known, the only unknown component of the interval boundaries is the conditional mean μ. This parameter can be estimated by its sample analog

$$\hat{\mu} = \frac{\sum_{i=1}^{N} W_i \cdot Y_i}{\sum_{i=1}^{N} W_i}.$$

Given this estimator for μ, the identification region $[\theta_l, \theta_u]$ can be estimated as

$$[\hat{\theta}_l, \hat{\theta}_u] = [\hat{\mu} \cdot p, \hat{\mu} \cdot p + 1 - p] .$$

This estimator is consistent for the identification region $[\theta_l, \theta_u]$.

3.1 Symmetric Confidence Intervals for the Parameter and its Identification Region

The first step towards constructing CIs is to consider inference for μ. Using standard large sample results we have

$$\sqrt{N}(\hat{\mu} - \mu) \xrightarrow{d} \mathcal{N}(0, \sigma^2/p).$$

A consistent estimator for σ^2 is

$$\hat{\sigma}^2 = \frac{1}{N_1 - 1} \sum_{i=1}^{N} W_i \cdot (Y_i - \hat{\mu})^2,$$

where $N_1 = \sum_{i=1}^{N} W_i$. Hence the standard $100 \cdot \alpha\%$ confidence interval for μ is

$$CI_\alpha^\mu = \left[\hat{\mu} - z_{(\alpha+1)/2} \cdot \frac{\hat{\sigma}}{\sqrt{p \cdot N}}, \hat{\mu} + z_{(\alpha+1)/2} \cdot \frac{\hat{\sigma}}{\sqrt{p \cdot N}}\right], \tag{3.4}$$

where z_τ is the τ quantile of the standard normal distribution, so that $\Phi(z_\tau) = \int_{-\infty}^{z_\tau} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-y^2/2} dy = \tau$.

Now consider symmetric CIs for the identification region $[\theta_l, \theta_u]$ and for the parameter θ. In each case, let the desired asymptotic coverage probability be α. Lemma 3.1 shows that the symmetric interval

$$CI_\alpha^{[\theta_l,\theta_u]} = \left[\left(\hat{\mu} - z_{(\alpha+1)/2} \cdot \frac{\hat{\sigma}}{\sqrt{p \cdot N}}\right) \cdot p, \left(\hat{\mu} + z_{(\alpha+1)/2} \cdot \frac{\hat{\sigma}}{\sqrt{p \cdot N}}\right) \cdot p + 1 - p\right] \tag{3.5}$$

asymptotically covers $[\theta_l, \theta_u]$ with probability α, and Lemma 3.2 shows that the interval

$$CI_\alpha^\theta = \left[\left(\hat{\mu} - z_\alpha \cdot \frac{\hat{\sigma}}{\sqrt{p \cdot N}}\right) \cdot p, \left(\hat{\mu} + z_\alpha \cdot \frac{\hat{\sigma}}{\sqrt{p \cdot N}}\right) \cdot p + 1 - p\right] \tag{3.6}$$

asymptotically covers θ with at least probability α. Thus, Lemma 3.1 shows that the entire interval $[\theta_l, \theta_u]$ will, in large samples, be in the confidence interval $CI_\alpha^{[\theta_l,\theta_u]}$ with probability α, and Lemma 3.2 shows that, whatever the value of the unidentified parameter λ may be, the parameter of interest $\theta = p \cdot \mu + (1-p) \cdot \lambda$ will be in the confidence interval CI_α^θ with at least probability α, as long as $p < 1$. The second interval differs from the first only in that its cutoff points are based on z_α rather than $z_{(\alpha+1)/2}$, making the second interval strictly shorter than the first.

Lemma 3.1 (Coverage properties of $CI_\alpha^{[\theta_l,\theta_u]}$)

$$\inf_{F \in \mathcal{F}, p_0 \leq p \leq 1} \lim_{N \to \infty} \Pr\left([\theta_l, \theta_u] \subset CI_\alpha^{[\theta_l,\theta_u]}\right) = \alpha. \tag{3.7}$$

Proof: See Appendix.

Lemma 3.2 (Coverage properties of CI_α^θ)

$$\inf_{F \in \mathcal{F}, \lambda \in \Lambda, p_0 \leq p < 1} \lim_{N \to \infty} \Pr\left(\theta \in CI_\alpha^\theta\right) \geq \alpha, \textit{ with equality when } \alpha \geq 0.5. \tag{3.8}$$

Proof: See Appendix.

3.2 A Uniform Confidence Interval for the Parameter

Although the confidence interval CI_α^θ has in large samples the appropriate confidence level for all values of p in the open interval $(0,1)$, it has an unattractive feature. In this subsection we first describe in details this feature, and then propose a modification to eliminate it. The issue is that for any N one can find a value of p such that the coverage is arbitrarily close to $100\cdot(2\alpha-1)\%$, rather than the nominal $100\cdot\alpha\%$. To see this, let us look at an example with $Y|W=1$ normal with mean μ and known variance σ^2. For ease of exposition we consider a slight modification of CI_α^θ with the estimated variance $\hat\sigma^2$ replaced by the true variance σ^2:

$$\overline{CI}_\alpha^\theta = \left[\left(\hat\mu - z_\alpha\cdot\frac{\sigma}{\sqrt{p\cdot N}}\right)\cdot p, \left(\hat\mu + z_\alpha\cdot\frac{\sigma}{\sqrt{p\cdot N}}\right)\cdot p + 1 - p\right]. \tag{3.9}$$

The coverage probability of $\overline{CI}_\alpha^\theta$ for $\theta=\mu\cdot p+\lambda\cdot(1-p)$ at $\lambda=0$ (so $\theta=\mu\cdot p$) is

$$\begin{aligned}\Pr\left(\theta\in\overline{CI}_\alpha^\theta\right) &= \Pr\left(\left(\hat\mu - z_\alpha\cdot\frac{\sigma}{\sqrt{pN}}\right)p\le\mu p\le\left(\hat\mu + z_\alpha\cdot\frac{\sigma}{\sqrt{pN}}\right)p+1-p\right)\\ &= 1-\Pr\left(\left(\hat\mu - z_\alpha\cdot\frac{\sigma}{\sqrt{pN}}\right)p>\mu p\right)-\Pr\left(\left(\hat\mu + z_\alpha\cdot\frac{\sigma}{\sqrt{pN}}\right)p+1-p<\mu p\right)\\ &= 1-\Pr\left(\frac{\hat\mu-\mu}{\sigma}\sqrt{pN}>z_\alpha\right)-\Pr\left(\frac{\hat\mu-\mu}{\sigma}\sqrt{pN}<-z_\alpha-\frac{(1-p)\sqrt{N}}{\sigma\sqrt{p}}\right)\\ &= 1-\Phi\left(-z_\alpha\right)-\Phi\left(-z_\alpha-\frac{(1-p)\sqrt{N}}{\sigma\sqrt{p}}\right)=\Phi\left(z_\alpha+\frac{(1-p)\sqrt{N}}{\sigma\sqrt{p}}\right)-\Phi\left(-z_\alpha\right).\end{aligned}$$

For any fixed $p\in(0,1)$, this coverage probability approaches α as $N\to\infty$. However, for any fixed $N<\infty$, the coverage probability approaches $2\alpha-1$ as $p\to1$. The finite-sample coverage of CI_α^θ is less than its asymptotic coverage because the asymptotic calculation sets to zero the probability that the lower bound of the identification region exceeds the estimate of the upper bound. This probability is generically positive in finite samples, and its magnitude increases as $p\to1$.

This example shows that the asymptotic coverage result in Lemma 3.2 is very delicate. The statement of the lemma supposes that $p<1$. At $p=1$ the parameter of interest $\theta=\mu p+\lambda(1-p)=\mu$ is point-identified, and the standard $100\alpha\%$ confidence interval for θ is CI_α^μ, given in (3.4). Interval CI_α^μ, which has width $2z_{(\alpha+1)/2}\cdot\hat\sigma/\sqrt{\cdot N}$, is not the limit of interval CI_α^θ as $p\to1$. The width of interval CI_α^θ is $2z_\alpha\cdot\hat\sigma\sqrt{p}/\sqrt{N}+1-p$. For any fixed value of N, the width of CI_α^θ converges to $2z_\alpha\cdot\hat\sigma/\sqrt{N}$ as $p\to1$. Since $z_\alpha<z_{(\alpha+1)/2}$, this is strictly less than the width of the standard interval for $p=1$. Thus, there is a discontinuity in the width of the confidence interval at $p=1$.

The discontinuity at $p=1$ is unsettling, especially its direction. It is counterintuitive that the CI for θ should be shorter when the parameter is partially identified than when it is point-identified. The anomaly arises because the coverage of CI_α^θ is not uniform in (F,λ,p), and in particular not uniform in p. Formally, Lemma 3.2 shows only that the coverage of interval CI_α^θ converges to

$100\alpha\%$ as $N \to \infty$, for every value of (F, λ, p). It does not show that the lowest coverage rate will converge to $100\alpha\%$. In other words, it may be true that

$$\lim_{N\to\infty} \inf_{F\in\mathcal{F},\lambda\in\Lambda,p_0\le p<1} \Pr\left(\theta \in CI_\alpha^\theta\right) < \alpha.$$

Uniformity of confidence intervals is not always feasible. For example, in instrumental variables settings uniformity of confidence intervals over parameter values where the structural parameters are not identified implies that the expected length of the intervals must be infinite (Gleser and Hwang, 1987; Dufour, 1997). In that case it has been argued that the parameter space should be restricted to regions where the structural parameters are identified, and uniformity should only be required to hold over the restricted parameter space. Here the issue is arguably different. The point-identified case with $p = 1$ is of great interest, and any reasonable parameter space would include it. We therefore think it desirable to construct confidence intervals that are uniform in p, at least for $p \in [p_0, 1]$, for some $p_0 > 0$.

We propose here a modification of CI_α^θ whose coverage probability does converge uniformly in p; indeed it converges uniformly in (F, λ, p). To motivate the modification, it is helpful to first consider the case where $Y|W = 1$ is normally distributed with unknown mean μ and known variance σ^2. In this case, we will be able to derive the exact (finite sample) coverage rate of confidence intervals. A pair of sufficient statistics for θ is $(\hat{\mu}, \hat{p})$, where $\hat{p} = \sum_i W_i/N$ is also ancillary. Note that $\hat{\mu}|\hat{p} \sim \mathcal{N}(\mu, \sigma^2/(N\hat{p}))$. Again we consider symmetric intervals of the form

$$\left[\hat{\theta}_l - D, \hat{\theta}_u + D\right].$$

The conditional coverage probability for such an interval, for a specific value of θ, is

$$\Pr\left(\hat{\theta}_l - D \le \theta \le \hat{\theta}_u + D \middle| \hat{p}\right) = \Pr\left(\hat{\mu}\cdot p - D \le \mu\cdot p + \lambda\cdot(1-p) \le \hat{\mu}\cdot p + 1 - p + D \middle| \hat{p}\right)$$

$$= \Pr\left(-D - \lambda\cdot(1-p) \le (\mu - \hat{\mu})\cdot p \le (1-\lambda)\cdot(1-p) + D \middle| \hat{p}\right)$$

$$= \Pr\left(-\sqrt{N\hat{p}}\cdot\frac{D + \lambda\cdot(1-p)}{\sigma p} \le \sqrt{N\hat{p}}\cdot\frac{\mu-\hat{\mu}}{\sigma} \le \sqrt{N\hat{p}}\cdot\frac{D + (1-\lambda)\cdot(1-p)}{\sigma p} \middle| \hat{p}\right)$$

$$= \Phi\left(\sqrt{N\hat{p}}\cdot\frac{D + (1-\lambda)\cdot(1-p)}{\sigma p}\right) - \Phi\left(-\sqrt{N\hat{p}}\cdot\frac{D + \lambda\cdot(1-p)}{\sigma p}\right).$$

This probability has local minima at the end points $\lambda = 0, 1$, with the probabilities identical at both end points, which are therefore the global minimum.[1] Thus, to get the coverage rate to be at least α for all values of λ, one needs to choose D to solve:

$$\Phi\left(\sqrt{N\hat{p}}\cdot\frac{D + (1-p)}{\sigma p}\right) - \Phi\left(-\sqrt{N\hat{p}}\cdot\frac{D}{\sigma p}\right) = \alpha.$$

[1] That the probabilities are identical at the endpoints is immediate. The endpoints give the local minima because the second derivative of the probability with respect to λ is negative for all values of λ. The first derivative with respect to λ is

$$\sqrt{N\hat{p}}\cdot\frac{1-p}{\sigma p}\cdot\left(\phi\left(-\sqrt{N\hat{p}}\cdot\frac{D + \lambda\cdot(1-p)}{\sigma p}\right) - \phi\left(\sqrt{N\hat{p}}\cdot\frac{D + (1-\lambda)\cdot(1-p)}{\sigma p}\right)\right).$$

This yields an exact CI conditional on $\hat{p}$. To facilitate the comparison with the previous interval let $C_N = D\sqrt{N\hat{p}}/(p\sigma)$. Then C_N is chosen to solve

$$\Phi\left(C_N + \sqrt{N\hat{p}} \cdot \frac{1-p}{\sigma p}\right) - \Phi\left(-C_N\right) = \alpha,$$

with the corresponding confidence interval

$$\left[\hat{\mu} \cdot p - C_N \frac{p\sigma}{\sqrt{N\hat{p}}}, \hat{\mu} \cdot p + (1-p) + C_N \frac{p\sigma}{\sqrt{N\hat{p}}}\right].$$

For any fixed $0 < p < 1$, $\lim_{N\to\infty} C_N = z_\alpha$, which would give us the interval CI_α^θ back. With p very close to one, however, there will be a substantial modification for finite N. With $p = 1$ the interval estimate is now identical to the standard one. For $0 < p < 1$ the confidence interval is strictly wider than the interval for $p = 1$.

For the general case with unknown distribution for $Y|W = 1$ we construct a confidence interval by replacing σ by $\hat{\sigma}$ and $\hat{p}$ by p as these modifications do not affect the asymptotic unconditional coverage rate:

$$\widetilde{CI}_\alpha^\theta = \left[\left(\hat{\mu} - C_N \cdot \hat{\sigma}/\sqrt{p \cdot N}\right) \cdot p, \left(\hat{\mu} + C_N \cdot \hat{\sigma}/\sqrt{p \cdot N}\right) \cdot p + 1 - p\right], \tag{3.10}$$

where C_N satisfies

$$\Phi\left(C_N + \sqrt{N} \cdot \frac{1-p}{\hat{\sigma}\sqrt{p}}\right) - \Phi\left(-C_N\right) = \alpha. \tag{3.11}$$

Lemma 3.3 shows that the new interval has a coverage rate that converges uniformly in (F, λ, p):

Lemma 3.3 (Coverage properties of $\widetilde{CI}_\alpha^\theta$)

$$\lim_{N\to\infty} \inf_{F\in\mathcal{F},\lambda\in\Lambda,p_0\le p\le 1} \Pr\left(\theta \in \widetilde{CI}_\alpha\right) \ge \alpha.$$

Proof: see Appendix.

It is useful to compare the three intervals, $CI_\alpha^{[\theta_l,\theta_u]}$, CI_α^θ, and $\widetilde{CI}_\alpha^\theta$, in terms of the constants that multiply $\hat{\sigma}/\sqrt{p \cdot N}$, the standard error of $\hat{\mu}$. Since the form of the intervals is the same for all three cases, and since the width of the intervals is strictly increasing in this constant we can compare the width by directly comparing these constants. For the first CI, $CI_\alpha^{[\theta_l,\theta_u]}$, the constant is $z_{(\alpha+1)/2}$, which solves

$$\Phi\left(C\right) - \Phi\left(-C\right) = \alpha. \tag{3.12}$$

The second derivative is

$$\left(\sqrt{N\hat{p}} \cdot \frac{1-p}{\sigma p}\right)^2 \cdot \left[-\phi\left(-\sqrt{N\hat{p}} \cdot \frac{D + \lambda \cdot (1-p)}{\sigma p}\right) \cdot \left(\sqrt{N\hat{p}} \cdot \frac{D + \lambda \cdot (1-p)}{\sigma p}\right)\right.$$
$$\left. -\phi\left(\sqrt{N\hat{p}} \cdot \frac{D + (1-\lambda) \cdot (1-p)}{\sigma p}\right) \cdot \left(\sqrt{N\hat{p}} \cdot \frac{D + (1-\lambda) \cdot (1-p)}{\sigma p}\right)\right] \le 0.$$

For the second interval CI^{θ}_{α}, the constant is z_α, which solves

$$\Phi\left(\infty\right)-\Phi\left(-C\right)=1-\Phi\left(-C\right)=\alpha, \tag{3.13}$$

and which is strictly smaller. For the third interval $\widetilde{CI}^{\theta}_{\alpha}$, the constant solves

$$\Phi\left(C+\sqrt{N}\cdot\frac{1-p}{\hat{\sigma}\sqrt{p}}\right)-\Phi\left(-C\right)=\alpha. \tag{3.14}$$

This is strictly between the first two constants, leading to the general result that $z_\alpha < C_N < z_{(\alpha+1)/2}$ and

$$CI^{\theta}_{\alpha}\subset\widetilde{CI}^{\theta}_{\alpha}\subset CI^{[\theta_l,\theta_u]}_{\alpha}.$$

Thus, the uniform confidence interval for the parameter is strictly narrower than the confidence interval for the identification region.

4 The General Case

In this section we develop a confidence interval that converges uniformly in more general settings, including ones in which the width of the identification region is a nuisance parameter that must be estimated. We then apply this to the case of missing data with unknown propensity score.

4.1 Confidence Intervals With Uniform Coverage Probabilities

We use the same structure as in Section 2. Let $(\Omega,\mathcal{A},P)$ be a specified probability space, and let $\mathcal{P}$ be a space of probability distributions on $(\Omega,\mathcal{A})$. The distribution P is not known, but a random sample of size N is available, with empirical distribution P_N. Let λ be a quantity which is known only to belong to a specified set Λ. Let $f(\cdot,\cdot):\mathcal{P}\times\Lambda\rightarrow\mathbb{R}$ be a specified real-valued function. The object of interest is the real-valued parameter $\theta=f(P,\lambda)$. Suppose that $\lambda_l(P)=\text{argmin}_{\lambda'\in\Lambda}f(P,\lambda')$ and $\lambda_u(P)=\text{argmax}_{\lambda'\in\Lambda}f(P,\lambda')$ exist. The first assumption requires that these functions do not depend on the probability distribution:

Assumption 4.1 (Invariance of Lower and Upper Bound)
$\lambda_l(P)=\lambda_l$ and $\lambda_u(P)=\lambda_u$ for all $P\in\mathcal{P}$.

Define $\theta_l=f(P,\lambda_l)$, $\theta_u=f(P,\lambda_u)$, with corresponding estimators $\hat{\theta}_l$ and $\hat{\theta}_u$. In many cases, although this is not necessary for the following argument, these estimators will be obtained as the sample-analogs, $\hat{\theta}_l=f(P_N,\lambda_l)$ and $\hat{\theta}_u=f(P_N,\lambda_u)$ Then the identification region $[f(P,\lambda_l),f(P,\lambda_u)]=[\theta_l,\theta_u]$ is naturally estimated by its sample analog $[\hat{\theta}_l,\hat{\theta}_u]$.

The next assumption ensures that well behaved estimators for the lower and upper bound exist:

Assumption 4.2 (Uniform Estimation of Bounds)
There are estimators for the lower and upper bound $\hat{\theta}_l$ *and* $\hat{\theta}_u$ *that satisfy:*

$$\sqrt{N}(\hat{\theta}_l - \theta_l) \xrightarrow{d} \mathcal{N}(0, \sigma_l^2), \qquad \sqrt{N}(\hat{\theta}_u - \theta_u) \xrightarrow{d} \mathcal{N}(0, \sigma_u^2),$$

uniformly in $P \in \mathcal{P}$, *and there are estimators for* σ_l^2 *and* σ_u^2 *that converge to the true values uniformly in* $P \in \mathcal{P}$.

Third, we impose some conditions on the set of probability distributions:

Assumption 4.3 (Set of Probability Distributions)
For all $P \in \mathcal{P}$,
(i) $\underline{\sigma}^2 \leq \sigma_l^2, \sigma_u^2 \leq \overline{\sigma}^2$ *for some positive and finite* $\underline{\sigma}^2$ *and* $\overline{\sigma}^2$,
(ii) $\theta_u - \theta_l \leq \overline{\Delta} < \infty$.

The fourth assumption ensures that the implied estimator for the width of the interval is well behaved. Specifically, when the true interval width $\Delta = \theta_u - \theta_l$ is close to zero (and the parameter is close to being point-identified), the estimated width $\hat{\Delta} = \hat{\theta}_u - \hat{\theta}_l$ cannot be allowed to be very large. This assumption is key to ensuring that the estimated width of the identification region can be used instead of the true width in the construction of the confidence interval. It allows one to avoid assuming a lower bound on the width of the identification region, which would rule out the point-identified case.

Assumption 4.4 (Convergence of Interval Width)
For all $\epsilon > 0$ *there are* $\nu > 0$, C, *and* N_0 *such that for* $N \geq N_0$

$$\Pr\left(\sqrt{N}|\hat{\Delta} - \Delta| > C \cdot \Delta^{\nu}\right) < \epsilon,$$

uniformly in $P \in \mathcal{P}$.

Given these assumptions we construct the confidence interval as:

$$\overline{CI}_{\alpha}^{\theta} = \left[\hat{\theta}_l - \overline{C}_N \cdot \hat{\sigma}_l/\sqrt{N}, \hat{\theta}_u + \overline{C}_N \cdot \hat{\sigma}_u/\sqrt{N}\right], \tag{4.15}$$

where $\overline{C}_N$ satisfies

$$\Phi\left(\overline{C}_N + \sqrt{N} \cdot \frac{\hat{\Delta}}{\max(\hat{\sigma}_l, \hat{\sigma}_u)}\right) - \Phi\left(-\overline{C}_N\right) = \alpha. \tag{4.16}$$

The following Lemma gives the general uniform coverage result.

Lemma 4.1 (Coverage properties of $\overline{CI}_{\alpha}^{\theta}$)
Suppose Assumptions 4.1-4.4 hold. Then

$$\lim_{N\to\infty} \inf_{P\in\mathcal{P},\lambda\in\Lambda} \Pr\left(\theta \in \overline{CI}_{\alpha}^{\theta}\right) \geq \alpha.$$

Proof: see Appendix.

4.2 The Missing Data Problem With Unknown Propensity Score

Here we return to the missing data problem of Section 3. We allow for an unknown propensity score (assuming p is bounded away from zero) and show that this problem fits the four assumptions sufficient for the application of Lemma 4.1. We continue to assume that the conditional variance of Y given $W = 1$ is bounded and bounded away from zero, guaranteeing that Assumption 4.3 is satisfied.

In this case $\Lambda = [0, 1]$, $\theta = \mu \cdot p + \lambda \cdot (1 - p)$, so that $\lambda_l(P) = 0$ and $\lambda_u(P) = 1$ for all P, and Assumption 4.1 is satisfied.

The lower bound can be estimated by

$$\hat{\theta}_l = \frac{1}{N} \sum_{i=1}^{N} W_i \cdot Y_i.$$

The upper bound can be estimated by

$$\hat{\theta}_u = \frac{1}{N} \sum_{i=1}^{N} \left(W_i \cdot Y_i + 1 - W_i \right).$$

Both estimators are asymptotically normal, with

$$\sqrt{N}(\hat{\theta}_l - \theta_l) \xrightarrow{d} \mathcal{N}\left(0, \sigma_l^2\right), \qquad \text{and } \sqrt{N}(\hat{\theta}_u - \theta_u) \xrightarrow{d} \mathcal{N}\left(0, \sigma_u^2\right),$$

where $\sigma_l^2 = \sigma^2 \cdot p + \mu^2 \cdot p \cdot (1 - p)$ and $\sigma_u^2 = \sigma^2 \cdot p + \mu^2 \cdot p \cdot (1 - p) + p \cdot (1 - p) - 2 \cdot \mu \cdot p \cdot (1 - p)$. Since the convergence is also uniform in P, Assumption 4.2 is satisfied.

Finally, consider Assumption 4.4. Let $\nu = 1/2$, and $N_0 = 1$. In the missing data case $\hat{\Delta} = 1 - \hat{p}$. The variance of $\hat{\Delta}$ is $\Delta(1 - \Delta)/N$. Hence, $\mathbb{E}[N \cdot (\hat{\Delta} - \Delta)^2] \leq \Delta$. Now apply Chebyshev's inequality, with $C = 1/\sqrt{\epsilon}$, so that $\Pr\left(\sqrt{N}|\hat{\Delta} - \Delta| > C \cdot \Delta^{\nu}\right) = \Pr\left(N(\hat{\Delta} - \Delta)^2 > C^2 \cdot \Delta^{2\nu}\right) < \mathbb{E}[N \cdot (\hat{\Delta} - \Delta)^2]/(C^2 \Delta^{2\nu}) \leq \Delta/(C^2 \Delta^{2\nu}) = 1/C^2 = \epsilon$. Hence Assumption 4.4 is satisfied, and Lemma 4.1 can be used to construct a CI by substituting $\hat{p}$ for p in $\widetilde{CI}_{\alpha}^{\theta}$ given in 3.10.

5 An Empirical Illustration

In this section we use real data to illustrate the confidence intervals proposed in this paper. The data were originally analyzed by Meyer, Viscusi, and Dubin (1995), who wanted to learn how an increase in the level of disability benefits affects the number of weeks a worker spent on disability; this variable is measured in whole weeks, and its distribution is highly skewed. The increase in benefits applies only to high-earning workers, not to low-earning ones. Meyer, Viscusi and Dubin estimated difference-in-difference models of the form

$$Y_i = \beta_0 + \beta_1 \cdot T_i + \beta_2 \cdot G_i + \beta_3 \cdot T_i \cdot G_i + \epsilon_i,$$

where Y_i is the outcome, the binary variable T_i indicates the post-change period, and the binary variable G_i indicates the high-earning group (the group affected by the change in benefits). The coefficient on the interaction, β_3, is the parameter of interest, expressing the effect of the change in benefits on disability durations. Meyer, Viscusi and Dubin reported results when the outcome is measured in weeks and in log-weeks.

Athey and Imbens (2002) suggest a generalization of the difference-in-difference model that they label the changes-in-changes model. Letting Y_i denote the observed outcome for individual i, and Y_i^N the outcome in the absence of the change in the benefits (equal to Y_i and observed unless $T_i = G_i = 1$), their model assumes that Y_i^N satisfies

$$Y_i^N = h(T_i, \epsilon_i),$$

with $\epsilon_i \perp T_i | G_i$. Group differences are expressed by differences in the conditional distribution of $\epsilon_i | G_i = g$ by g. Athey and Imbens take the parameter of interest to be

$$\tau = \mathbb{E}[Y|T = 1, G = 1] - \mathbb{E}[Y^N|T = 1, G = 1],$$

the difference between the expected outcome for the high earners in the second period, $\mathbb{E}[Y|T = 1, G = 1]$, and the expected outcome for the high earners in the second period in the absence of the change in benefits, $\mathbb{E}[Y^N|T = 1, G = 1]$. A key assumption is that $h(t, \epsilon)$ is weakly monotone in ϵ. Athey and Imbens show that with discrete data the parameter of interest is not point-identified and that sharp bounds can be constructed. These have the form

$$\tau \in \left[\mathbb{E}\left[Y \middle| T = 1, G = 1 \right] - \mathbb{E}\left[F_{Y,01}^{-1}(F_{Y,00}(Y)) \middle| T = 0, G = 1 \right], \right.$$
$$\left. \mathbb{E}\left[Y \middle| T = 1, G = 1 \right] - \mathbb{E}\left[F_{Y,01}^{-1}(\underline{F}_{Y,00}(Y)) \middle| T = 0, G = 1 \right] \right],$$

where $F_{Y,gt}(y) = Pr(Y \leq y | T = t, G = g)$, $\underline{F}_{Y,gt}(y) = Pr(Y < y | T = t, G = g)$, and $F_{Y,gt}^{-1}(q) = \inf\{y | F_{Y,gt}(y) \geq q\}$.

We estimate these bounds by substituting maximum likelihood estimates (with the outcome discrete this is straightforward). We then use the results in Athey and Imbens on asymptotic normality of these estimators and estimate the standard errors. These are used in Table 1 to construct three confidence intervals. First, we calculate the CI for the entire identification region, $CI_{0.95}^{[\theta_l, \theta_u]}$. Second, we calculate the CI for the parameter of interest, $CI_{0.95}^{\theta}$. Third, we calculate the CI for the parameter adjusted to ensure uniform convergence, $\overline{CI}_{0.95}^{\theta}$, using the estimated width of the identification region. We calculate this both for the outcome in levels and in logarithms.

We find that modifying the CI to have the appropriate coverage only for the parameter of interest, rather than for the entire identification region, makes a considerable difference. For the analysis in levels, ensuring that the convergence is uniform leads to an additional substantial change.

Table 1: Confidence Intervals for Effect of Benefit Change on Injury Durations in Discrete Changes-In-Changes Model (Meyer-Viscusi-Dubin Data)

Estimand	$\hat{\theta}_l$ (s.e.)	$\hat{\theta}_u$ (s.e.)	$\overline{C}_N$	$CI_{0.95}^{[\theta_l,\theta_u]}$	$CI_{0.95}^{\theta}$	$\overline{CI}_{0.95}^{\theta}$
Average Effect on Treated (levels)	0.15 (1.69)	1.14 (1.67)	1.753	[-3.17,4.42]	[-2.63,3.89]	[-2.82,4.08]
Average Effect on Treated (logs)	0.14 (0.13)	0.58 (0.17)	1.655	[-0.12,0.91]	[-0.08,0.86]	[-0.08,0.86]

Three 95% CIs (in square brackets) are reported for both parameters of interest. The first CI is based on the estimator of the lower bound minus 1.96, and the estimator of the upper bound plus 1.96 times their standard errors. The second CI is equal to the estimator of the lower bound minus 1.645, and the estimator of the upper bound plus 1.645 times their standard errors. The third CI is the adjusted interval for the parameter, given in (4.15)

6 Conclusion

In the last decade a growing body of research has studied inference in settings where parameters of interest are partially identified. Less attention has been focused on the construction of confidence intervals in such settings. When confidence intervals have been estimated, they have typically been constructed to provide coverage for the entire identification region with fixed probability. In this paper we introduce a conceptually different type of confidence interval that asymptotically covers the true value of the parameter with fixed probability. We show that, in general, coverage of a parameter is a less demanding objective than is coverage of the entire identification region. We show in a simple setting with missing data that CIs for the parameter are proper subsets of ones for the identification region, with the difference in width related to the difference in critical values for one– and two–sided tests. However, we find that the exact coverage probabilities of the simplest version of our new CIs do not converge to their nominal values uniformly across different values for the width of the identification region. Specifically, uniformity fails when the width of the region shrinks to zero; that is, as the parameter becomes point-identified. To avoid this anomaly, we modify the proposed CI to ensure that its exact coverage probabilities do converge uniformly to their nominal values. We motivate the modified CI by showing that its exact and nominal coverage probabilities coincide when outcomes are normally distributed. We also provide more general results on the implementation of the new CIs, and provide conditions under which CIs with uniform asymptotic coverage can be constructed by substituting estimates for unknown nuisance parameters, including the width of the identification region. Finally, in a brief empirical illustration we show that these results can lead to substantially different confidence intervals.

7 Appendix

Proof of Lemma 3.1: Fix F and p. Then

$$\begin{aligned}
&\Pr\left([\theta_l,\theta_u] \subset CI_\alpha^{[\theta_l,\theta_u]}\right) \\
&= Pr\left(\theta_l \geq \left(\hat{\mu} - z_{(\alpha+1)/2}\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right)\cdot p \text{ and } \theta_u \leq \left(\hat{\mu} + z_{(\alpha+1)/2}\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right)\cdot p + 1 - p\right) \\
&= 1 - Pr\left(\theta_l < \left(\hat{\mu} - z_{(\alpha+1)/2}\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right)\cdot p \text{ or } \theta_u > \left(\hat{\mu} + z_{(\alpha+1)/2}\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right)\cdot p + 1 - p\right) \\
&= 1 - Pr\left(\theta_l < \left(\hat{\mu} - z_{(\alpha+1)/2}\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right)\cdot p\right) - Pr\left(\theta_u > \left(\hat{\mu} + z_{(\alpha+1)/2}\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right)\cdot p + 1 - p\right) \\
&= 1 - Pr\left(\mu\cdot p < \left(\hat{\mu} - z_{(\alpha+1)/2}\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right)\cdot p\right) \\
&\qquad - Pr\left(\mu\cdot p + 1 - p > \left(\hat{\mu} + z_{(\alpha+1)/2}\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right)\cdot p + 1 - p\right) \\
&= 1 - Pr\left(\mu < \hat{\mu} - z_{(\alpha+1)/2}\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right) - Pr\left(\mu > \hat{\mu} + z_{(\alpha+1)/2}\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right) \\
&\longrightarrow 1 - (1-\alpha)/2 - (1-\alpha)/2 = \alpha
\end{aligned}$$

Note that the second step in the above derivation is an equality, rather than a weak inequality, because the two events whose union is taken are mutually exclusive. □

Proof of Lemma 3.2: Consider the three possibilities for λ: $\lambda = 0$, $\lambda = 1$, and $0 < \lambda < 1$. In the first case, with $\lambda = 0$, we have $\theta = \mu \cdot p$. Hence the coverage probability of CI_α^θ is

$$\begin{aligned}
&\Pr(\theta \in CI_\alpha^\theta) = \Pr\left(\left(\hat{\mu} - z_\alpha\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right)\cdot p \leq \mu\cdot p \leq \left(\hat{\mu} + z_\alpha\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right)\cdot p + 1 - p\right) \\
&\geq 1 - \Pr\left(\left(\hat{\mu} - z_\alpha\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right)\cdot p > \mu\cdot p\right) - \Pr\left(\mu\cdot p > \left(\hat{\mu} + z_\alpha\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right)\cdot p + 1 - p\right).
\end{aligned}$$

When $\alpha \geq 0.5$, which is the usual case of interest, the above weak inequality is an equality because $z_\alpha \geq 0$, which implies that the two events whose union is taken are mutually exclusive. Consider the two probabilities on the righthand side. The second probability equals

$$\Pr\left(\mu\cdot p > \left(\hat{\mu} + z_\alpha\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right)\cdot p + 1 - p\right) = \Pr\left(-z_\alpha - \sqrt{N}\cdot\frac{1-p}{\hat{\sigma}\sqrt{p}} > \sqrt{N}\cdot\frac{\hat{\mu}-\mu}{\hat{\sigma}/\sqrt{p}}\right),$$

which goes to zero as N goes to infinity since $\sqrt{N}(\hat{\mu}-\mu)\sqrt{p}/\hat{\sigma}$ converges to a standard normal distribution. The first probability satisfies

$$\Pr\left(\left(\hat{\mu} - z_\alpha\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right)\cdot p > \mu\cdot p\right) = \Pr\left(\sqrt{N}\cdot\frac{\hat{\mu}-\mu}{\hat{\sigma}/\sqrt{p}} > z_\alpha\right) \longrightarrow 1 - \alpha.$$

Hence for this value of λ the coverage of the interval converges to α. The same argument works for $\lambda = 1$. Now consider intermediate values for λ. In that case we have

$$\Pr(\theta \in CI_\alpha^\theta) = \Pr\left(\left(\hat{\mu} - z_\alpha\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\right)\cdot p \leq \mu\cdot p + \lambda\cdot(1-p) \leq \left(\hat{\mu} + z_\alpha\cdot\frac{\hat{\sigma}}{\sqrt{p\cdot N}}\cdot p + 1 - p\right).$$

$$\geq 1 - \Pr\left(\left(\hat{\mu} - z_\alpha \cdot \frac{\hat{\sigma}}{\sqrt{p \cdot N}}\right) \cdot p > \mu \cdot p + \lambda \cdot (1-p)\right)$$
$$-\Pr\left(\mu \cdot p + \lambda \cdot (1-p) > \left(\hat{\mu} + z_\alpha \cdot \frac{\hat{\sigma}}{\sqrt{p \cdot N}}\right) \cdot p + 1 - p\right)$$

For the first probability on the righthand side we have:

$$\Pr\left(\mu \cdot p + \lambda \cdot (1-p) < \hat{\mu} \cdot p - z_\alpha \cdot p \cdot \frac{\hat{\sigma}}{\sqrt{p \cdot N}}\right) = \Pr\left(\sqrt{N} \cdot \frac{\hat{\mu} - \mu}{\hat{\sigma}/\sqrt{p}} > z_\alpha + \sqrt{N} \cdot \lambda \cdot \frac{1-p}{\hat{\sigma}\sqrt{p}}\right),$$

which goes to zero as N goes to infinity, as long as $\lambda > 0$. Similarly, the second probability goes to zero with N. Hence for all intermediate values of λ the asymptotic coverage is 100%, irrespective of the nominal coverage rate. Thus, the lowest asymptotic coverage across all values of λ is $100\alpha\%$ when $\alpha \geq 0.5$ and is at least $100\alpha\%$ when $\alpha < 0.5$. $\square$

Before presenting a proof of Lemma 3.3 we present a number of preliminary results. First we state a result for uniform convergence to a central limit theorem.

Lemma 7.1 (Uniform Central Limit Theorem, Berry-Esseen) *Suppose $X_1, X_2, \ldots$ are independent and identically distributed random variables with distribution function $F \in \mathcal{F}$. Let $\mu(F) = \mathbb{E}_F[X]$, $\sigma^2(F) = \mathbb{E}_F[(X-\mu)^2]$, and let $0 < \underline{\sigma}^2 \leq \sigma^2(F) \leq \bar{\sigma}^2 < \infty$, and $\mathbb{E}_F[X^3] < \infty$ for all $F \in \mathcal{F}$. Then*

$$\sup_{-\infty < a < \infty, F \in \mathcal{F}} \left| \Pr\left(\sqrt{N}\left(\frac{\bar{X}_N - \mu}{\sigma}\right) < a\right) - \Phi(a)\right| \longrightarrow 0,$$

where $\bar{X}_N = \sum_{i=1}^N X_i/N$.

For a formal proof see, for example, Shorack and Wellner (1986). Next, we show that uniform convergence still holds if we use an estimated variance.

Lemma 7.2 (Uniform Central Limit Theorem with Estimated Variance)

$$\sup_{-\infty < a < \infty, F \in \mathcal{F}} \left| \Pr\left(\sqrt{N}\left(\frac{\bar{X}_N - \mu}{\hat{\sigma}}\right) < a\right) - \Phi(a)\right| \longrightarrow 0$$

Proof of Lemma 7.2: By the triangle inequality

$$\sup_{-\infty < a < \infty, F \in \mathcal{F}} \left| \Pr\left(\sqrt{N}\left(\frac{\bar{X}_N - \mu}{\hat{\sigma}}\right) < a\right) - \Phi(a)\right|$$
$$\leq \sup_{-\infty < a < \infty, F \in \mathcal{F}} \left| \Pr\left(\sqrt{N}\left(\frac{\bar{X}_N - \mu}{\sigma}\right) < a\right) - \Phi(a)\right|$$
$$+ \sup_{-\infty < a < \infty, F \in \mathcal{F}} \left| \Pr\left(\sqrt{N}\left(\frac{\bar{X}_N - \mu}{\hat{\sigma}}\right) < a\right) - \Pr\left(\sqrt{N}\left(\frac{\bar{X}_N - \mu}{\sigma}\right) < a\right)\right|$$

By lemma 7.1, the first term on the righthand side converges to zero, and all we need to prove is that

$$\sup_{-\infty < a < \infty, F \in \mathcal{F}} \left| \Pr\left(\sqrt{N}\left(\frac{\bar{X}_N - \mu}{\hat{\sigma}}\right) < a\right) - \Pr\left(\sqrt{N}\left(\frac{\bar{X}_N - \mu}{\sigma}\right) < a\right)\right| \longrightarrow 0$$

First, note that $\hat{\sigma}$ converges to σ uniformly in F because of the moment conditions on X. Since σ is bounded away from zero, this implies that $(\hat{\sigma} - \sigma)/\sigma$ converges to zero, also uniformly in F. So for any $\epsilon > 0$ and

$\eta > 0$, there is an N_0 such that, for $N > N_0$ and for all F, $\Pr((\hat{\sigma} - \sigma)/\sigma > \epsilon) < \eta$. By the Berry-Esseen theorem there is also for all $\delta > 0$ an N_1 such that for $N \geq N_1$

$$\left| \Pr\left(\sqrt{N} \left(\frac{\bar{X}_N - \mu}{\sigma} \right) < a \right) - \Phi(a) \right| < \delta,$$

uniformly in a. For $N \geq \max(N_0, N_1)$ we have

$$\sup_{-\infty < a < \infty, F \in \mathcal{F}} \left| \Pr\left(\sqrt{N} \left(\frac{\bar{X}_N - \mu}{\hat{\sigma}} \right) < a \right) - \Pr\left(\sqrt{N} \left(\frac{\bar{X}_N - \mu}{\sigma} \right) < a \right) \right|$$

$$= \sup_{-\infty < a < \infty, F \in \mathcal{F}} \left| \Pr\left(\sqrt{N} \left(\frac{\bar{X}_N - \mu}{\sigma} \right) < \frac{\hat{\sigma}}{\sigma} \cdot a \right) - \Pr\left(\sqrt{N} \left(\frac{\bar{X}_N - \mu}{\sigma} \right) < a \right) \right|$$

$$\leq \sup_{-\infty < a < \infty, F \in \mathcal{F}} \Pr\left(\left| \sqrt{N} \left(\frac{\bar{X}_N - \mu}{\sigma} \right) - a \right| < |a| \cdot \epsilon \right) + \eta$$

$$\leq \sup_{-\infty < a < \infty, F \in \mathcal{F}} \Phi(a + |a| \cdot \epsilon) - \Phi(a - |a| \cdot \epsilon) + \eta + 2\delta$$

$$\leq \sup_{-\infty < a < \infty, F \in \mathcal{F}, \omega \in [0,1]} 2 \cdot \phi(a + \omega \cdot |a| \cdot \epsilon) \cdot |a| \cdot \epsilon + \eta + 2\delta,$$

For $\epsilon < 1/2$ there is a bound on $\phi(a + \omega \cdot |a| \cdot \epsilon) \cdot |a|$ that does not depend on ϵ, so that by choosing ϵ, η and δ small enough we can make the entire expression arbitrarily small. $\square$

The previous two results, Lemmas 7.1-7.2, can be used to show that coverage for the standard confidence interval for the sample mean is uniform with respect to the underlying distribution.

Lemma 7.3 (Uniform Coverage for Confidence Interval for the Sample Mean)

$$\inf_{F \in \mathcal{F}} \Pr\left(\bar{X}_N - z_{(\alpha+1)/2} \cdot \frac{\hat{\sigma}}{\sqrt{N}} \leq \mu \leq \bar{X}_N + z_{(\alpha+1)/2} \cdot \frac{\hat{\sigma}}{\sqrt{N}} \right) \longrightarrow \alpha$$

Proof of Lemma 7.3:

$$\inf_{F \in \mathcal{F}} \Pr\left(\bar{X}_N - z_{(\alpha+1)/2} \cdot \frac{\hat{\sigma}}{\sqrt{N}} \leq \mu \leq \bar{X}_N + z_{(\alpha+1)/2} \cdot \frac{\hat{\sigma}}{\sqrt{N}} \right)$$

$$= \inf_{F \in \mathcal{F}} \Pr\left(-z_{(\alpha+1)/2} \leq \sqrt{N} \cdot \frac{\bar{X}_N - \mu}{\hat{\sigma}} \leq z_{(\alpha+1)/2} \right)$$

which goes to α. $\square$ **Proof of Lemma 3.3:** For fixed λ the coverage probability is

$$\Pr\left(\left(\hat{\mu} - C_N \cdot \hat{\sigma}/\sqrt{p \cdot N} \right) \cdot p \leq \mu \cdot p + \lambda \cdot (1-p) \leq \left(\hat{\mu} + C_N \cdot \hat{\sigma}/\sqrt{p \cdot N} \right) \cdot p + 1 - p \right)$$

$$= \Pr\left(-C_N \frac{\hat{\sigma}}{\sigma} - \sqrt{N} \cdot \frac{\lambda \cdot (1-p)}{\sigma \cdot \sqrt{p}} \leq \sqrt{N} \cdot \frac{\mu - \hat{\mu}}{\sigma/\sqrt{p}} \leq C_N \frac{\hat{\sigma}}{\sigma} + \sqrt{N} \cdot \frac{(1-\lambda) \cdot (1-p)}{\sigma \cdot \sqrt{p}} \right)$$

For any $\epsilon > 0$ there almost surely exists an N_0 such that for $N > N_0$, $|(\hat{\sigma} - \sigma)/\sigma| < \epsilon$, so that $\epsilon > 1 - \hat{\sigma}/\sigma$. Therefore for $N \geq N_0$,

$$\Pr\left(-C_N \frac{\hat{\sigma}}{\sigma} - \sqrt{N} \cdot \frac{\lambda \cdot (1-p)}{\sigma \cdot \sqrt{p}} \leq \sqrt{N} \cdot \frac{\mu - \hat{\mu}}{\sigma/\sqrt{p}} \leq C_N \frac{\hat{\sigma}}{\sigma} + \sqrt{N} \cdot \frac{(1-\lambda) \cdot (1-p)}{\sigma \cdot \sqrt{p}} \right)$$

$$\geq \Pr\left(-C_N(1-\epsilon) - \sqrt{N} \cdot \frac{\lambda \cdot (1-p)}{\sigma \cdot \sqrt{p}} \leq \sqrt{N} \cdot \frac{\mu - \hat{\mu}}{\sigma/\sqrt{p}} \leq C_N(1-\epsilon) + \sqrt{N} \cdot \frac{(1-\lambda) \cdot (1-p)}{\sigma \cdot \sqrt{p}} \right)$$

For N large enough this can be made arbitrarily close to

$$\Phi\left(C_N(1-\epsilon)+\sqrt{N}\cdot\frac{(1-\lambda)\cdot(1-p)}{\sigma\cdot\sqrt{p}}\right)-\Phi\left(-C_N(1-\epsilon)-\sqrt{N}\cdot\frac{\lambda\cdot(1-p)}{\sigma\cdot\sqrt{p}}\right)$$

$$=\Phi\left(C_N+\sqrt{N}\cdot\frac{(1-\lambda)\cdot(1-p)}{\sigma\cdot\sqrt{p}}\right)-\Phi\left(-C_N-\sqrt{N}\cdot\frac{\lambda\cdot(1-p)}{\sigma\cdot\sqrt{p}}\right)+2\epsilon C_N\phi(\omega),$$

for some ω. Because $C_N\leq z_{(\alpha+1)/2}$ (see the definition of C_N), and since $\phi(\cdot)$ is bounded, the last term can be made arbitrarily small by choosing ϵ small. The sum of the first two terms has a negative second derivative with respect to λ, and so it is minimized at $\lambda=0$ or $\lambda=1$. By the definition of C_N it follows that at those values for λ the value of the sum is α. Hence for any $\nu>0$, for N large enough we have

$$\Pr\left(\left(\hat{\mu}-C_N\cdot\hat{\sigma}/\sqrt{p\cdot N}\right)\cdot p\leq\mu\cdot p+\lambda\cdot(1-p)\leq\left(\hat{\mu}+C_N\cdot\hat{\sigma}/\sqrt{p\cdot N}\right)\cdot p+1-p\right)\geq\alpha-\nu.$$

□

Before proving Lemma 4.1 we establish a couple of preliminary results. Define $\breve{C}_N$ and $\ddot{C}_N$ by:

$$\Phi\left(\breve{C}_N+\sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)}\right)-\Phi\left(-\breve{C}_N\right)=\alpha,$$

$$\Phi\left(\ddot{C}_N+\sqrt{N}\cdot\frac{\hat{\Delta}}{\max(\sigma_l,\sigma_u)}\right)-\Phi\left(-\ddot{C}_N\right)=\alpha.$$

Note that $\overline{C}_N$ and $\ddot{C}_N$ are stochastic (as they depend on $\hat{\Delta}$), while $\breve{C}_N$ is a sequence of constants.

Lemma 7.4 (Difference Between $\overline{C}_N$ and $\ddot{C}_N$)
Suppose Assumptions 4.1-4.4 hold. Then

$$\left|\overline{C}_N-\ddot{C}_N\right|\longrightarrow 0,$$

uniformly in $P\in\mathcal{P}$.

Proof of Lemma 7.4:

By Assumption 4.2 $\hat{\sigma}_l$ and $\hat{\sigma}_u$ converge to their probability limits uniformly in $P\in\mathcal{P}$. Since both σ_l and σ_u are bounded away from zero on $\mathcal{P}$, this implies that $1/\max(\hat{\sigma}_u,\hat{\sigma}_l)$ converges to its probability limit uniformly. Define $\lambda=\hat{\Delta}/\max(\sigma_l,\sigma_u)$ and $\hat{\lambda}=\hat{\Delta}/\max(\hat{\sigma}_l,\hat{\sigma}_u)$. Then $\hat{\lambda}/\lambda$ converges to one uniformly.

By the definition of $\overline{C}_N$ and $\ddot{C}_N$,

$$\Phi\left(\ddot{C}_N+\sqrt{N}\cdot\lambda\right)-\Phi\left(-\ddot{C}_N\right)=\Phi\left(\overline{C}_N+\sqrt{N}\cdot\hat{\lambda}\right)-\Phi\left(-\overline{C}_N\right)=\alpha.$$

Hence

$$\left|\Phi\left(\ddot{C}_N+\sqrt{N}\cdot\lambda\right)-\Phi\left(-\ddot{C}_N\right)-\left(\Phi\left(\overline{C}_N\cdot\frac{\lambda}{\hat{\lambda}}+\sqrt{N}\cdot\lambda\right)-\Phi\left(-\overline{C}_N\cdot\frac{\lambda}{\hat{\lambda}}\right)\right)\right| \tag{7.17}$$

$$\leq\left|\Phi\left(\overline{C}_N+\sqrt{N}\cdot\hat{\lambda}\right)-\Phi\left(-\overline{C}_N\right)-\left(\Phi\left(\overline{C}_N\cdot\frac{\lambda}{\hat{\lambda}}+\sqrt{N}\cdot\lambda\right)-\Phi\left(-\overline{C}_N\cdot\frac{\lambda}{\hat{\lambda}}\right)\right)\right|$$

$$\leq\left|\Phi\left(\overline{C}_N+\sqrt{N}\cdot\hat{\lambda}\right)-\Phi\left(\overline{C}_N\cdot\frac{\lambda}{\hat{\lambda}}+\sqrt{N}\cdot\lambda\right)\right|+\left|\Phi\left(-\overline{C}_N\right)-\Phi\left(-\overline{C}_N\cdot\frac{\lambda}{\hat{\lambda}}\right)\right|$$

$$= \left| \Phi\left(\frac{\hat\lambda}{\lambda} \cdot \left(\overline{C}_N \cdot \frac{\lambda}{\hat\lambda} + \sqrt{N}\cdot\lambda \right) \right) - \Phi\left(\overline{C}_N \cdot \frac{\lambda}{\hat\lambda} + \sqrt{N}\cdot\lambda \right) \right| + \left| \Phi\left(-\frac{\hat\lambda}{\lambda} \cdot \overline{C}_N \cdot \frac{\lambda}{\hat\lambda} \right) - \Phi\left(-\overline{C}_N \cdot \frac{\lambda}{\hat\lambda} \right) \right| \tag{7.18}$$

By the mean value theorem, there exists a $\gamma \in [0,1]$ such that $\Phi(a\cdot b) - \Phi(a) = \phi(a\cdot(1+\gamma\cdot(b-1))\cdot a\cdot(b-1)$. Hence, with $|\hat\lambda/\lambda - 1| < \epsilon$, the first term of (7.18) can be bounded by

$$\left| \phi\left((1+\tilde\epsilon)\cdot \left(\overline{C}_N \cdot \frac{\lambda}{\hat\lambda} + \sqrt{N}\cdot\lambda \right) \right) \cdot \epsilon \cdot \left(\overline{C}_N \cdot \frac{\lambda}{\hat\lambda} + \sqrt{N}\cdot\lambda \right) \right|,$$

for some $|\tilde\epsilon| \le \epsilon$, and the second term by

$$\left| \phi\left((1+\bar\epsilon)\cdot \overline{C}_N \cdot \frac{\lambda}{\hat\lambda} \right) \cdot \epsilon \cdot \overline{C}_N \cdot \frac{\lambda}{\hat\lambda} \right|,$$

for some $|\bar\epsilon| \le \epsilon$. These expressions can be made arbitrarily small by choosing ϵ small enough, implying that (7.17) can be made arbitrarily small. Using a mean value theorem, equation (7.17) can be written, for some $\gamma \in [0,1]$, as

$$\left| \phi\left(\ddot{C}_N + \gamma\cdot\left(\overline{C}_N \cdot \frac{\lambda}{\hat\lambda} - \ddot{C}_N \right) + \sqrt{N}\cdot\lambda \right) \cdot \left(\ddot{C}_N - \overline{C}_N \cdot \frac{\lambda}{\hat\lambda} \right) \right.$$
$$\left. + \phi\left(\ddot{C}_N + \gamma\cdot\left(\overline{C}_N \cdot \frac{\lambda}{\hat\lambda} - \ddot{C}_N \right) \right) \cdot \left(\ddot{C}_N - \overline{C}_N \cdot \frac{\lambda}{\hat\lambda} \right) \right|$$
$$= \left| \phi\left(\ddot{C}_N + \gamma\cdot\left(\overline{C}_N \cdot \frac{\lambda}{\hat\lambda} - \ddot{C}_N \right) + \sqrt{N}\cdot\lambda \right) + \phi\left(\ddot{C}_N + \gamma\cdot\left(\overline{C}_N \cdot \frac{\lambda}{\hat\lambda} - \ddot{C}_N \right) \right) \right| \cdot \left| \ddot{C}_N - \overline{C}_N \cdot \frac{\lambda}{\hat\lambda} \right|$$

This can only be small if $|\ddot{C}_N - \overline{C}_N \cdot (\lambda/\hat\lambda)|$ is small, which, given uniform convergence of $\lambda/\hat\lambda$ to one requires $|\overline{C}_N - \ddot{C}_N| \to 0$. □

Lemma 7.5 *For all $\epsilon > 0$ there is an N_0 such that for $N \ge N_0$,*

$$\left| \Pr\left(\hat\theta_l - \overline{C}_N \cdot \hat\sigma_l/\sqrt{N} \le \theta_l \le \hat\theta_u + \overline{C}_N \cdot \hat\sigma_u/\sqrt{N} \right) - \Pr\left(\hat\theta_l - \ddot{C}_N \cdot \sigma_l/\sqrt{N} \le \theta_l \le \hat\theta_u + \ddot{C}_N \cdot \sigma_u/\sqrt{N} \right) \right| < \epsilon,$$

uniformly in $P \in \mathcal{P}$.

Proof of Lemma 7.5:

First, by uniform convergence of $\hat\sigma_l$ and $\hat\sigma_u$ to their probability limits, and by uniform convergence of $\overline{C}_N - \ddot{C}_N$ to zero, there is for all positive ϵ_1 and ϵ_2 an N_0 such that for $N > N_0$, $\Pr(|\overline{C}_N \hat\sigma_l/\sigma_l - \ddot{C}_N| > \epsilon_1) < \epsilon_2$. Also, there is an N_1 such that for $N > N_1$, $\sup_z |\Phi(z) - \Pr(\sqrt{N}(\hat\theta_l - \theta_l)/\sigma_l < z)| < \epsilon_3$. Next,

$$\left| \Pr\left(\hat\theta_l - \overline{C}_N \cdot \hat\sigma_l/\sqrt{N} \le \theta \le \hat\theta_u + \overline{C}_N \cdot \hat\sigma_u/\sqrt{N} \right) - \Pr\left(\hat\theta_l - \ddot{C}_N \cdot \sigma_l/\sqrt{N} \le \theta \le \hat\theta_u + \ddot{C}_N \cdot \sigma_u/\sqrt{N} \right) \right|$$
$$\le \left| \Pr\left(\hat\theta_l - \overline{C}_N \cdot \hat\sigma_l/\sqrt{N} > \theta \right) - \Pr\left(\hat\theta_l - \ddot{C}_N \cdot \sigma_l/\sqrt{N} > \theta \right) \right| \tag{7.19}$$
$$+ \left| \Pr\left(\theta > \hat\theta_u + \overline{C}_N \cdot \hat\sigma_u/\sqrt{N} \right) - \Pr\left(\theta > \hat\theta_u + \ddot{C}_N \cdot \sigma_u/\sqrt{N} \right) \right| \tag{7.20}$$

We will show that (7.19) can be made arbitrarily small. The same argument can be used to show that (7.20) can be made arbitrarily small. To show that (7.19) can be made arbitrarily small we write

$$\Pr\left(\hat\theta_l - \overline{C}_N \cdot \hat\sigma_l/\sqrt{N} > \theta \right) - \Pr\left(\hat\theta_l - \ddot{C}_N \cdot \sigma_l/\sqrt{N} > \theta \right)$$

$$= \Pr\left(\hat\theta_l - \overline{C}_N \cdot \hat\sigma_l/\sqrt{N} > \theta \middle| \left|\overline{C}_N \frac{\hat\sigma_l}{\sigma_l} - \ddot{C}_N\right| < \epsilon_1\right) \cdot \Pr\left(\left|\overline{C}_N \frac{\hat\sigma_l}{\sigma_l} - \ddot{C}_N\right| < \epsilon_1\right)$$

$$+\Pr\left(\hat\theta_l - \overline{C}_N \cdot \hat\sigma_l/\sqrt{N} > \theta \middle| \left|\overline{C}_N \frac{\hat\sigma_l}{\sigma_l} - \ddot{C}_N\right| \geq \epsilon_1\right) \cdot \Pr\left(\left|\overline{C}_N \frac{\hat\sigma_l}{\sigma_l} - \ddot{C}_N\right| \geq \epsilon_1\right) - \Pr\left(\hat\theta_l - \ddot{C}_N \cdot \sigma_l/\sqrt{N} > \theta\right)$$

$$\leq \Pr\left(\hat\theta_l - \overline{C}_N \cdot \hat\sigma_l/\sqrt{N} > \theta \middle| \left|\overline{C}_N \frac{\hat\sigma_l}{\sigma_l} - \ddot{C}_N\right| < \epsilon_1\right) \cdot \Pr\left(\left|\overline{C}_N \frac{\hat\sigma_l}{\sigma_l} - \ddot{C}_N\right| < \epsilon_1\right)$$

$$+\Pr\left(\left|\overline{C}_N \frac{\hat\sigma_l}{\sigma_l} - \ddot{C}_N\right| \geq \epsilon_1\right) - \Pr\left(\hat\theta_l - \ddot{C}_N \cdot \sigma_l/\sqrt{N} > \theta\right)$$

$$\leq \Pr\left(\hat\theta_l - \ddot{C}_N \cdot \sigma_l/\sqrt{N} + \epsilon_1\sigma_l/\sqrt{N} > \theta \middle| \left|\overline{C}_N \frac{\hat\sigma_l}{\sigma_l} - \ddot{C}_N\right| < \epsilon_1\right) \cdot \Pr\left(\left|\overline{C}_N \frac{\hat\sigma_l}{\sigma_l} - \ddot{C}_N\right| < \epsilon_1\right)$$

$$+\epsilon_2 - \Pr\left(\hat\theta_l - \ddot{C}_N \cdot \sigma_l/\sqrt{N} > \theta\right)$$

$$\leq \Pr\left(\hat\theta_l - \ddot{C}_N \cdot \sigma_l/\sqrt{N} + \epsilon_1\sigma_l/\sqrt{N} > \theta\right) + \epsilon_2 - \Pr\left(\hat\theta_l - \ddot{C}_N \cdot \sigma_l/\sqrt{N} \geq \theta\right)$$

$$= \Pr\left(\sqrt{N} \cdot \frac{\hat\theta_l - \theta_l}{\sigma_l} > \sqrt{N} \cdot \frac{\theta - \theta_l}{\sigma_l} + \ddot{C}_N - \frac{\epsilon_1}{\sigma_l}\right) + \epsilon_2 - \Pr\left(\sqrt{N} \cdot \frac{\hat\theta_l - \theta_l}{\sigma_l} > \sqrt{N} \cdot \frac{\theta - \theta_l}{\sigma_l} + \ddot{C}_N\right)$$

$$= \Pr\left(\sqrt{N} \cdot \frac{\theta - \theta_l}{\sigma_l} + \ddot{C}_N > \sqrt{N} \cdot \frac{\hat\theta_l - \theta_l}{\sigma_l} > \sqrt{N} \cdot \frac{\theta - \theta_l}{\sigma_l} + \ddot{C}_N - \frac{\epsilon_1}{\sigma_l}\right) + \epsilon_2$$

$$\leq \sup_z \Pr\left(z\, \sqrt{N} \cdot \frac{\hat\theta_l - \theta_l}{\sigma_l} > z - \frac{\epsilon_1}{\sigma_l}\right) + \epsilon_2$$

$$\leq \sup_z \left(\Phi(z) - \Phi\left(z - \frac{\epsilon_1}{\sigma_l}\right)\right) + 2\epsilon_3 + \epsilon_2$$

$$\leq \epsilon_1 \cdot \frac{\overline{\phi}}{\sigma_l} + \epsilon_2 + 2\epsilon_3,$$

where $\overline{\phi} = \sup_z \phi(z) = \phi(0) = 1/\sqrt{2\pi}$. Following the same logic one can show that

$$\Pr\left(\hat\theta_l - \overline{C}_N \cdot \hat\sigma_l/\sqrt{N} > \theta\right) - \Pr\left(\hat\theta_l - \ddot{C}_N \cdot \sigma_l/\sqrt{N} > \theta\right) \geq -\epsilon_1 \cdot \frac{\overline{\phi}}{\sigma_l} - \epsilon_2 - 2\epsilon_3.$$

Together the two imply that (7.19) can be made arbitrarily small. □

Lemma 7.6 *For any $\eta, \epsilon > 0$, there is an N_0 such that for $N \geq N_0$*

$$\Pr\left(\Phi\left(\ddot{C}_N + \sqrt{N} \cdot \frac{\Delta}{\max(\sigma_l, \sigma_u)}\right) - \Phi\left(-\ddot{C}_N\right) < \alpha - \eta\right) < \epsilon,$$

uniformly in $P \in \mathcal{P}$.

Proof of Lemma 7.6:
First, the statement in the Lemma is, because $\ddot{C}_N$ satisfies $\Phi(\ddot{C}_N + \sqrt{N}\lambda) - \Phi(-\ddot{C}_N) = \alpha$, equivalent to

$$\Pr\left(\left(\Phi\left(\ddot{C}_N + \sqrt{N} \cdot \lambda\right) - \Phi\left(-\ddot{C}_N\right)\right) - \left(\Phi\left(\ddot{C}_N + \sqrt{N} \cdot \frac{\Delta}{\max(\sigma_l, \sigma_u)}\right) - \Phi\left(-\ddot{C}_N\right)\right) > \eta\right) < \varepsilon,$$

where $\lambda = \hat\Delta / \max(\sigma_l, \sigma_u)$.

By Assumption 4.4 there are $\nu > 0$, $C > 0$, and N_0, such that with $\delta = \nu/5$, and $N \geq \max(N_0, C^{1/\delta})$, so that

$$\Pr\left(\sqrt{N}|\hat{\Delta} - \Delta| > N^\delta \Delta^\nu\right) \leq \Pr\left(\sqrt{N}|\hat{\Delta} - \Delta| > C\Delta^\nu\right) < \epsilon.$$

Then:

$$\begin{aligned}
&\left(\Phi\left(\ddot{C}_N + \sqrt{N}\cdot\lambda\right) - \Phi\left(-\ddot{C}_N\right)\right) - \left(\Phi\left(\ddot{C}_N + \sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)}\right) - \Phi\left(-\ddot{C}_N\right)\right)\\
&= \Phi\left(\ddot{C}_N + \sqrt{N}\cdot\lambda\right) - \Phi\left(\ddot{C}_N + \sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)}\right)\\
&= 1\{\hat{\Delta} \leq \Delta\}\cdot\left(\Phi\left(\ddot{C}_N + \sqrt{N}\cdot\lambda\right) - \Phi\left(\ddot{C}_N + \sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)}\right)\right)\\
&\quad +1\{\hat{\Delta} > \Delta, \sqrt{N}|\hat{\Delta} - \Delta| \leq N^\delta\Delta^\nu\}\cdot\left(\Phi\left(\ddot{C}_N + \sqrt{N}\cdot\lambda\right) - \Phi\left(\ddot{C}_N + \sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)}\right)\right)\\
&\quad +1\{\hat{\Delta} > \Delta, \sqrt{N}|\hat{\Delta} - \Delta| > N^\delta\Delta^\nu\}\cdot\left(\Phi\left(\ddot{C}_N + \sqrt{N}\cdot\lambda\right) - \Phi\left(\ddot{C}_N + \sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)}\right)\right)\\
&\leq 1\{\hat{\Delta} > \Delta, \sqrt{N}|\hat{\Delta} - \Delta| \leq N^\delta\Delta^\nu\}\cdot\left(\Phi\left(\ddot{C}_N + \sqrt{N}\cdot\lambda\right) - \Phi\left(\ddot{C}_N + \sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)}\right)\right)\\
&\quad +1\left\{\sqrt{N}|\hat{\Delta} - \Delta| > N^\delta\Delta^\nu\right\}.
\end{aligned}$$

The event in the second indicator function has probability less than ϵ. The first expression is, by a mean value theorem, equal to:

$$1\{\hat{\Delta} > \Delta, \sqrt{N}|\hat{\Delta} - \Delta| \leq N^\delta\Delta^\nu\}\cdot\phi\left(\ddot{C}_N + \sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)} + \gamma\cdot\sqrt{N}\cdot\frac{\hat{\Delta} - \Delta}{\max(\sigma_l,\sigma_u)}\right)\cdot\sqrt{N}\cdot\frac{\hat{\Delta} - \Delta}{\max(\sigma_l,\sigma_u)},$$

for some $0 \leq \gamma \leq 1$. Because the entire expression is zero unless $\hat{\Delta} > \Delta$, and $\ddot{C}_N$ and Δ are nonnegative, this can be bounded from above by its value at $\gamma = 0$ with $\ddot{C}_N$ dropped:

$$\begin{aligned}
&1\{\hat{\Delta} > \Delta, \sqrt{N}|\hat{\Delta} - \Delta| \leq N^\delta\Delta^\nu\}\cdot\phi\left(\sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)}\right)\cdot\sqrt{N}\cdot\frac{\hat{\Delta} - \Delta}{\max(\sigma_l,\sigma_u)}\\
&\quad\leq 1\{\hat{\Delta} > \Delta, \sqrt{N}|\hat{\Delta} - \Delta| \leq N^\delta\Delta^\nu\}\cdot\phi\left(\sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)}\right)\cdot\frac{N^\delta\Delta^\nu}{\max(\sigma_l,\sigma_u)}\\
&\quad\leq N^{-\delta}\cdot\phi\left(\sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)}\right)\cdot\frac{N^{2\delta}\Delta^\nu}{\max(\sigma_l,\sigma_u)}. \qquad (7.21)
\end{aligned}$$

Maximizing

$$\phi\left(\sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)}\right)\cdot\frac{N^{2\delta}\Delta^\nu}{\max(\sigma_l,\sigma_u)}, \qquad (7.22)$$

over Δ gives $\Delta = \max(\sigma_l, \sigma_u)\cdot\sqrt{\nu}\cdot N^{-1/2}$. Substituting this into (7.22) gives an expression that is decreasing in N if $\delta < 4\nu$, and which is therefore bounded, with the bound independent of the value of Δ and thus uniform over $P \in \mathcal{P}$. Hence (7.21) can be made smaller than η by choosing N large enough, uniformly in $P \in \mathcal{P}$, completing the proof. □

Lemma 7.7 (Difference Between $\ddot{C}_N$ and $\breve{C}_N$)
For any $\eta, \epsilon > 0$, there is an N_0 such that for $N \geq N_0$

$$\Pr\left(\ddot{C}_N < \breve{C}_N - \eta\right) < \epsilon,$$

uniformly in $P \in \mathcal{P}$.

Proof of Lemma 7.7:
Let $\underline{\phi} = \phi(z_{(\alpha+1)/2})$. Note that $\ddot{C}_N$ and $\breve{C}_N$ are positive and less than $z_{(\alpha+1)/2}$, and thus $\phi(\ddot{C}_N) \geq \underline{\phi}$ and $\phi(\breve{C}_N) \geq \underline{\phi}$ Using Lemma 7.6 there is an N_0 such that for $N \geq N_0$

$$\Pr\left(\Phi\left(\ddot{C}_N + \sqrt{N} \cdot \frac{\Delta}{\max(\sigma_l, \sigma_u)}\right) - \Phi\left(-\ddot{C}_N\right) < \alpha - \eta \cdot \underline{\phi}\right) < \epsilon,$$

uniformly in $P \in \mathcal{P}$. Conditional on

$$\Phi\left(\ddot{C}_N + \sqrt{N} \cdot \frac{\Delta}{\max(\sigma_l, \sigma_u)}\right) - \Phi\left(-\ddot{C}_N\right) > \alpha - \eta \cdot \underline{\phi},$$

and by the fact that $\breve{C}_N$ satisfies $\Phi(\breve{C}_N + \sqrt{N} \cdot \Delta / \max(\sigma_l, \sigma_u)) - \Phi(-\breve{C}_N) = \alpha$, we have

$$\Phi\left(\ddot{C}_N + \sqrt{N} \cdot \frac{\Delta}{\max(\sigma_l, \sigma_u)}\right) - \Phi\left(\breve{C}_N + \sqrt{N} \cdot \frac{\Delta}{\max(\sigma_l, \sigma_u)}\right) - \Phi\left(-\ddot{C}_N\right) + \Phi\left(-\breve{C}_N\right) > -\eta \cdot \underline{\phi}$$

By a mean value theorem this implies

$$\left(\phi\left(\ddot{C}_N + \sqrt{N} \cdot \frac{\Delta}{\max(\sigma_l, \sigma_u)} + \gamma \cdot (\breve{C}_N - \ddot{C}_N)\right) + \phi\left(\breve{C}_N + \gamma \cdot (\ddot{C}_N - \breve{C}_N)\right)\right) \cdot \left(\ddot{C}_N - \breve{C}_N\right) > -\eta \cdot \underline{\phi},$$

for some $\gamma \in [0, 1]$, and thus $\ddot{C_N} > \breve{C}_N - \eta$ with probability $1 - \epsilon$. $\square$

The combination of Lemmas 7.4 and 7.7 implies that for any $\epsilon > 0$ we can find an N_0 such that for $N \geq N_0$,

$$\Pr\left(\overline{C}_N < \breve{C}_N - \eta\right) < \epsilon,$$

uniformly in $P \in \mathcal{P}$. Note that Lemma 7.7 does not imply that $|\overline{C}_N - \breve{C}_N|$ converges to zero uniformly. This is not necessarily true unless we are willing to rule out values of Δ close to zero, which is exactly the point-identified area we are concerned with.

Proof of Lemma 4.1:
We will prove that for any positive ϵ, for N sufficiently large, the probability that

$$\hat{\theta}_l - \overline{C}_N \cdot \hat{\sigma}_l / \sqrt{N} \leq \theta \leq \hat{\theta}_u + \overline{C}_N \cdot \hat{\sigma}_u / \sqrt{N},$$

is at least $\alpha - \epsilon$, uniformly in $P \in \mathcal{P}$. We will prove this for $\theta = \theta_u$. The proof for $\theta = \theta_l$ is analogous, and similar to previous cases the coverage probability is minimized at the boundary of the identification region.

For arbitrary positive ϵ_1, ϵ_2, and ϵ_3, choose N large enough so that the following conditions are satisfied (i), $\sup_z |\Pr(\sqrt{N}(\hat{\theta}_l - \theta_l)/\sigma_l \leq z) - \Phi(z)| \leq \epsilon_1$, (ii), $\sup_z |\Pr(\sqrt{N}(\hat{\theta}_u - \theta_u)/\sigma_u \leq z) - \Phi(z)| \leq \epsilon_1$, and (iii),

$\Pr(\ddot{C}_N - \breve{C}_N < -\epsilon_2) < \epsilon_3$. Existence of such an N follows for conditions (i) and (ii) from Assumption 4.2, and for condition (iii) from Lemma 7.7.

Define the following events, E_1, E_2, E_3, E_4, and E_5:

$$E_1 \equiv \hat{\theta}_l - \overline{C}_N \cdot \hat{\sigma}_l/\sqrt{N} \leq \theta_u \leq \hat{\theta}_u + \overline{C}_N \cdot \hat{\sigma}_u/\sqrt{N},$$

$$E_2 \equiv \hat{\theta}_l - \ddot{C}_N \cdot \sigma_l/\sqrt{N} \leq \theta_u \leq \hat{\theta}_u + \ddot{C}_N \cdot \sigma_u/\sqrt{N},$$

$$E_3 \equiv \hat{\theta}_l - (\breve{C}_N - \epsilon_2) \cdot \sigma_l/\sqrt{N} \leq \theta_u \leq \hat{\theta}_u + (\breve{C}_N - \epsilon_2) \cdot \sigma_u/\sqrt{N},$$

$$E_4 \equiv \hat{\theta}_l - \breve{C}_N \cdot \sigma_l/\sqrt{N} \leq \theta_u \leq \hat{\theta}_u + \breve{C}_N \cdot \sigma_u/\sqrt{N},$$

$$E_5 \equiv \ddot{C}_N - \breve{C}_N > -\epsilon_2,$$

and let E_5^c be the complement of E_5. Note that $(E_5 \cap E_3) \Rightarrow E_2$ and thus $(E_5 \cap E_3) \Rightarrow (E_2 \cap E_3)$. Define also

$$P_3 \equiv \Phi(\breve{C}_N - \epsilon_2 + \sqrt{N} \cdot \Delta/\sigma_l) - \Phi(-\breve{C}_N + \epsilon_2).$$

$$P_4 \equiv \Phi(\breve{C}_N + \sqrt{N} \cdot \Delta/\sigma_l) - \Phi(-\breve{C}_N) = \alpha.$$

By conditions (i) and (ii), $|P_3 - \Pr(E_3)| \leq 2\epsilon_1$ and $|P_4 - \Pr(E_4)| \leq 2\epsilon_1$. Also, $|P_3 - P_4| \leq 2\epsilon_2\overline{\phi}$, and by (iii), $\Pr(E_6^c) < \epsilon_3$. By Lemma 7.5 it follows that for any $\epsilon_4 > 0$ we can choose N large enough so that $|\Pr(E_1) - \Pr(E_2)| < \epsilon_4$. Then, by elementary set theory:

$$\Pr(E_1) \geq \Pr(E_2) - \epsilon_4 \geq \Pr(E_2 \cap E_3) - \epsilon_4 \geq \Pr(E_5 \cap E_3) - \epsilon_4 \geq \Pr(E_3) - \Pr(E_5^c) - \epsilon_4$$

$$\geq P_3 - 2\epsilon_1 - \epsilon_3 - \epsilon_4 \geq P_4 - 2\epsilon_1 - \epsilon_3 - 2\epsilon_2\overline{\phi} - \epsilon_4 = \alpha - 2\epsilon_1 - \epsilon_3 - 2\epsilon_2\overline{\phi} - \epsilon_4.$$

Since $\epsilon_1, \ldots, \epsilon_4$ were chosen arbitrarily, one can make $\Pr(E_1) > \alpha - \epsilon$ for any $\epsilon > 0$. $\square$

REFERENCES

Athey, S., and G. Imbens (2002), "Identification and Inference in Nonlinear Difference-in-Differences Models," NBER technical working paper t0280.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir, (2002) "Changes in the Distribution of Male and Female Wages Accounting for Employment Compositional Change," Institute for Fiscal Studies.

Bollinger, C. (1996), "Bounding Mean Regressions when a Binary Regressor is Mismeasured," *Journal of Econometrics*, 73, 387-399.

Chernozhukov, V., H. Hong and E. Tamer, (2003), "Parameter Set Inference in a Class of Econometric Models," unpublished manuscript, Department of Economics, Princeton University.

Dominitz, J. and R. Sherman (2003), "Nonparametric Analysis of Mixture Models with Verification," California Institute of Technology.

Dufour, J.-M., (1997), "Some Impossibility Theorems in Econometrics With Applications to Structural and Dynamic Models," *Econometrica*, 65, 6: 1365-1387.

Gleser, L., and J. Hwang, (1987), “The Nonexistence of $100(1-\alpha)$ Confidence Sets of Finite Expected Diameter in Errors in Variables and Related Models,” *The Annals of Statistics*, 15, 1351-1362.

Haile, P. and E. Tamer (2003), "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111, 1-51.

Horowitz, J. and C. Manski (1995), "Identification and Robustness with Contaminated and Corrupted Data," *Econometrica*, 63, 281-302.

Horowitz, J. and C. Manski (1998), “Censoring of Outcomes and Regressors due to Survey Nonresponse: Identification and Estimation Using Weights and Imputations,” *Journal of Econometrics*, 84, 37-58.

Horowitz, J., and C. Manski, (2000): “Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data” *Journal of the American Statistical Association*, 95, 77-84.

Hotz, J., C. Mullins, and S. Sanders (1997), "Bounding Causal Effects Using Data from a Contaminated Natural Experiment: Analyzing the Effects of Teenage Childbearing," *Review of Economic Studies*, 64, 575-603.

Lee, D. (2002), “Trimming for Bounds on Treatment Effects with Missing Outcomes,” NBER technical working paper t0277.

Manski, C. (1990), “Nonparametric Bounds on Treatment Effects,” *American Economic Review Papers and Proceedings*, 80, 319-323.

Manski, C. (2003), *Partial Identification of Probability Distributions*, New York: Springer-Verlag.

Manski, C., G. Sandefur, S. McLanahan, and D. Powers (1992), “Alternative Estimates of the Effect of Family Structure During Adolescence on High School Graduation,” *Journal of the American Statistical Association*, 87, 25-37.

Manski, C. and E. Tamer (2002), "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519-546.

Meyer, B., K. Viscusi and D. Durbin (1995), “Workers’ Compensation and Injury Duration: Evidence from a Natural Experiment,” *American Economic Review*, 85, 322-340.

Molinari, F. (2002), "Identification of Probability Distributions with Misclassified Data," Department of Economics, Northwestern University.

Mullins, C. (2003), “Bounding Treatment Effects with Contaminated and Censored Data: Assessing the Impract of Early Childbearing on Children,” VanderBilt University.

Shorack, G., and J. Wellner, (1986), *Empirical Processes with Applications to Statistics*, Wiley, New York, NY.