Econometrica, Vol. 72, No. 6 (November, 2004), 1845–1857

NOTES AND COMMENTS

CONFIDENCE INTERVALS FOR PARTIALLY IDENTIFIED PARAMETERS

BY GUIDO W. IMBENS AND CHARLES F. MANSKI[1]

Recently a growing body of research has studied inference in settings where parameters of interest are partially identified. In many cases the parameter is real-valued and the identification region is an interval whose lower and upper bounds may be estimated from sample data. For this case confidence intervals (CIs) have been proposed that cover the entire identification region with fixed probability. 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 this probability. However, 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. To avoid the problems associated with this, we modify the proposed CI to ensure that its exact coverage probabilities do converge uniformly to their nominal values. We motivate this modified CI through exact results for the Gaussian case.

KEYWORDS: Bounds, identification regions, confidence intervals, uniform convergence.

1. INTRODUCTION

IN THE LAST DECADE a growing body of research has studied inference in settings where parameters of interest are partially identified (see Manski (2003) for an overview of this literature). 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 these estimates. Early on, Manski, Sandefur, McLanahan, and Powers (1992) computed separate confidence intervals for the lower and upper bounds. Subsequently, Horowitz and Manski (2000) developed CIs that asymptotically cover the entire identification region with fixed probability. Recently Chernozhukov, Hong, and Tamer (2003) extended this approach to settings with vector-valued parameters defined through minimization problems.

Here, we introduce a conceptually different type of confidence interval. Rather than cover the entire identification region with fixed probability α, we propose CIs that asymptotically cover the true value of the parameter with this probability. The key insight is that when the identification region has positive width, the true parameter can be close to at most one of the region's boundaries. Suppose that the true value is close to the upper bound of the identification region. Then, asymptotically the probability that the estimate for the lower bound exceeds the true value can be ignored when making inference on the true parameter. This allows the researcher to allocate the entire probability

[1]We thank Joerg Stoye for excellent research assistance, Michael Jansson, Francesca Molinari, Tom Rothenberg, a coeditor, and two anonymous referees for comments, and the NSF for support through Grants SES 0136789 (Imbens) and SES 0314312 (Manski). Imbens also acknowledges financial support from the Giannini Foundation and the Agricultural Experimental Station at UC Berkeley.

of making an error, $1-\alpha$, to values above the upper-bound point estimate. We do not know whether the true parameter is close to the upper or lower bound, so one-sided intervals with confidence level α are constructed around both bounds.

To illustrate the nature of our CIs for partially identified parameters, we construct CIs for the mean of a bounded random variable when some data are missing and the distribution of missing data is unrestricted. We show that our CIs for the parameter are proper subsets of the corresponding CIs for the identification region, with the difference in width related to the difference in critical values for one- and two-sided tests. However, 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. A 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 this modified CI through exact results for the Gaussian case.

2. CONFIDENCE INTERVALS FOR PARAMETERS AND 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 that 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 $\{f(P,\lambda'), \lambda'\in\Lambda\}$. 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 $P\in\mathcal{P}$. We focus on the class of problems in which the identification region is the closed interval $[f(P,\lambda_l(P)), f(P,\lambda_u(P))]$. Manski (2003) describes various problems in this class.

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 nonnegative numbers that may depend on the sample 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. In this paper, we study the use of these intervals as CIs for the partially identified parameter $f(P,\lambda)$. Our most basic finding is Lemma 1:

LEMMA 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 $[f(P,\lambda_l(P)), f(P,\lambda_u(P))]$.*

All proofs are given in the Appendix.

An implication of the lemma is that researchers face a substantive choice whether to report intervals that cover the entire identification region or intervals that cover the true parameter value with some fixed probability. Although both intervals generally converge to the identification region as $N \to \infty$, their difference typically is of the order $O_p(N^{-1/2})$. Which CI is of interest depends on the application.

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 takes values in the bounded set $\mathbb{Y}$ and W is binary with values $\{0, 1\}$; without loss of generality, let the smallest and largest elements of $\mathbb{Y}$ be 0 and 1, respectively. The researcher has a random sample of $(W_i, Y_i \cdot W_i)$, $i = 1, \dots, N$, so W_i is always observed and Y_i is only observed if $W_i = 1$. Define $\mu = \mathbb{E}[Y|W = 1]$, $\lambda = \mathbb{E}[Y|W = 0]$, $\sigma^2 = \mathbb{V}(Y|W = 1)$, and $p = \mathbb{E}[W]$, with $0 < p \leq 1$. We assume initially that p is known. This will be relaxed in Section 4. Let $F(y)$ be the conditional distribution function of Y given $W = 1$, an element of the set of distribution functions $\mathcal{F}$ with variance $\underline{\sigma}^2 \leq \sigma^2 \leq \overline{\sigma}^2$, for some positive and finite $\underline{\sigma}^2$ and $\overline{\sigma}^2$. The distribution of Y given $W = 0$ is unknown; hence, $\lambda \in [0, 1]$. The parameter of interest is $\theta = \mathbb{E}[Y] = \mu \cdot p + \lambda \cdot (1 - p)$. The identification region for θ is the closed interval $[\theta_l, \theta_u] = [\mu \cdot p, \mu \cdot p + 1 - p]$.

With p known, the only unknown determinant of the interval boundaries is the conditional mean μ. This parameter can be estimated by its sample analog $\hat{\mu} = \sum_{i=1}^N W_i \cdot Y_i / N_1$ (where $N_1 = \sum_{i=1}^N W_i$), and the identification region can be consistently estimated by $[\hat{\theta}_l, \hat{\theta}_u] = [\hat{\mu} \cdot p, \hat{\mu} \cdot p + 1 - p]$. 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 $\hat{\sigma}^2 = \sum_{i=1}^N W_i \cdot (Y_i - \hat{\mu})^2/(N_1 - 1)$. Hence, the standard $100 \cdot \alpha\%$ confidence interval for μ is

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

where z_τ is the τ quantile of the standard normal distribution: $\Phi(z_\tau) = \int_{-\infty}^{z_\tau} (1/\sqrt{2\pi}) \times e^{-y^2/2}\, dy = \tau$. In the point identified case with $p = 1$ we have $\theta = \theta_l = \theta_u = \mu$, and thus in that case CI_α^μ is also the appropriate CI for θ and $[\theta_l, \theta_u]$.

Now consider symmetric CIs for the identification region $[\theta_l, \theta_u]$ and the parameter θ. The CI for $[\theta_l, \theta_u]$ substitutes the lower (upper) confidence bound for μ into the lower (upper) bound for the identification region:

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

Note that as $p \to 1$, $CI_\alpha^{[\theta_l, \theta_u]} \to CI_\alpha^\mu$. The CI for θ adjusts the critical values to obtain the appropriate coverage for θ:

$$CI_\alpha^\theta = \left[\left(\hat{\mu} - z_\alpha \cdot \frac{\hat{\sigma}}{\sqrt{p \cdot N}}\right) \cdot p, \left(\hat{\mu} + z_\alpha \cdot \frac{\hat{\sigma}}{\sqrt{p \cdot N}}\right) \cdot p + 1 - p\right]. \tag{3}$$

Note that this is a proper interval only if $2z_\alpha \hat{\sigma}/\sqrt{pN} > -(1 - p)/p$, which is always true if $\alpha \geq .5$ and will be true with probability arbitrarily close to one for N large if

$\alpha < .5$. One can modify the interval if this condition is not satisfied without affecting the asymptotic properties. The following lemma describes the large sample properties of these intervals.

LEMMA 2: *For any* $p_0 > 0$,
(i) $\inf_{F\in\mathcal{F},p_0\le p\le 1}\lim_{N\to\infty}\Pr([\theta_l,\theta_u]\subset CI_\alpha^{[\theta_l,\theta_u]})=\alpha$;
(ii) $\inf_{F\in\mathcal{F},\lambda\in\Lambda,p_0\le p<1}\lim_{N\to\infty}\Pr(\theta\in CI_\alpha^{\theta})=\alpha$.

Although the confidence interval CI_α^θ has in large samples the appropriate confidence level for all values of p, λ, and F, it has an unattractive feature. The issue is that for any N one can find a value of p and λ such that the coverage is arbitrarily close to $100\cdot(2\alpha-1)\%$, rather than the nominal $100\cdot\alpha\%$. To see this, we consider an example with $Y|W=1$ normal with mean μ and known variance σ^2. Let $\hat{p}=\sum_i W_i/N$. The exact coverage probability of CI_α^θ for θ, conditional on $\hat{p}$, at $\lambda=0$ (so $\theta=\mu\cdot p$) is

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

For any fixed $p\in(0,1)$, this coverage probability approaches α with probability one as $N\to\infty$. However, for any fixed $N<\infty$, the coverage probability approaches $2\alpha-1$ with probability one as $p\to 1$. This example shows that the asymptotic coverage result in Lemma 2 is very delicate. One can also see this by considering the width of interval CI_α^θ equal to $2z_\alpha\cdot\hat{\sigma}\sqrt{p}/\sqrt{N}+1-p$. As $p\to 1$, for fixed N, this width converges to $2z_\alpha\cdot\hat{\sigma}/\sqrt{N}$. This is strictly less than the width of CI_α^μ, which is the standard interval for θ for the point-identified case with $p=1$. 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 does not converge uniformly in (F,λ,p) and, in particular, not uniformly in p.

We propose here a modification of CI_α^θ whose coverage probability does converge uniformly in (F,λ,p). To motivate the modification, again consider the case where $Y|W=1\sim\mathcal{N}(0,\sigma^2)$ with known σ^2. The conditional coverage rate for symmetric intervals of the form $[\hat{\theta}_l-D,\hat{\theta}_u+D]$ is

$$\begin{aligned}&\Pr(\hat{\theta}_l-D\le\theta\le\hat{\theta}_u+D|\hat{p})\\&\quad=\Phi\left(\sqrt{N\hat{p}}\cdot\frac{D+(1-\lambda)\cdot(1-p)}{\sigma p}\right)-\Phi\left(-\sqrt{N\hat{p}}\cdot\frac{D+\lambda\cdot(1-p)}{\sigma p}\right).\end{aligned}$$

To get the coverage rate to be at least α for all values of λ, one needs to choose D to solve:

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

To facilitate comparison with the previous CI, let $C_N=D\sqrt{N\hat{p}}/(p\sigma)$ so that C_N solves

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

with the corresponding confidence interval

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

For any fixed $0 < p < 1$, $\lim_{N\to\infty} C_N = z_\alpha$, which would give us the interval CI^{θ}_{α} back. For fixed N, as $p \to 1$, the interval estimate now converges to CI^{μ}_{α} with no discontinuity at $p = 1$. 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 CI by replacing σ by $\hat{\sigma}$ and $\hat{p}$ by p:

$$\widetilde{CI}^{\theta}_{\alpha} = \left[\left(\hat{\mu} - C_N\cdot\hat{\sigma}/\sqrt{p\cdot N}\right)\cdot p, \left(\hat{\mu} + C_N\cdot\hat{\sigma}/\sqrt{p\cdot N}\right)\cdot p + 1 - p\right], \tag{4}$$

where C_N satisfies

$$\Phi\left(C_N + \sqrt{N}\cdot\frac{1-p}{\hat{\sigma}\sqrt{p}}\right) - \Phi(-C_N) = \alpha. \tag{5}$$

Lemma 3 shows that the new interval has a coverage rate that converges uniformly in (F, λ, p):

LEMMA 3: *For any* $p_0 > 0$,

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

It is interesting to compare the three intervals $CI^{[\theta_l,\theta_u]}_{\alpha}$, CI^{θ}_{α}, and $\widetilde{CI}^{\theta}_{\alpha}$ in terms of the constants that multiply $\hat{\sigma}/\sqrt{p\cdot N}$, the standard error of $\hat{\mu}$. The form of the intervals is the same for all three cases and the width of the intervals is strictly increasing in this constant, so we can compare the widths by directly comparing these constants. For $CI^{[\theta_l,\theta_u]}_{\alpha}$, the constant is $z_{(\alpha+1)/2}$, which solves $\Phi(C) - \Phi(-C) = \alpha$. For CI^{θ}_{α}, the constant is z_α, which solves $\Phi(\infty) - \Phi(-C) = \alpha$, and which is strictly smaller. For $\widetilde{CI}^{\theta}_{\alpha}$, the constant is C_N which solves

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

Unless $p = 1$, this is strictly between the first two constants so $CI^{\theta}_{\alpha} \subset \widetilde{CI}^{\theta}_{\alpha} \subset CI^{[\theta_l,\theta_u]}_{\alpha}$. If the parameter is point identified ($p = 1$), then $C_N = z_{(\alpha+1)/2}$ and $CI^{\theta}_{\alpha} \subset \widetilde{CI}^{\theta}_{\alpha} = CI^{[\theta_l,\theta_u]}_{\alpha} = CI^{\mu}_{\alpha}$.

4. THE GENERAL CASE

Here 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 use the same structure and notation as in Section 2. Define $\theta_l = f(P, \lambda_l(P))$, $\theta_u = f(P, \lambda_u(P))$, and $\Delta = \theta_u - \theta_l$, and let $\hat{\theta}_l$, $\hat{\theta}_u$, and $\hat{\Delta} = \hat{\theta}_u - \hat{\theta}_l$ be estimators for θ_l, θ_u, and Δ. Then the identification region $[\theta_l, \theta_u]$ is naturally estimated by its sample analog $[\hat{\theta}_l, \hat{\theta}_u]$.

We consider the following set of assumptions:

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

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

uniformly in $P \in \mathcal{P}$, *and there are estimators* $\hat{\sigma}_l^2$, $\hat{\sigma}_u^2$, *and* $\hat{\rho}$ *for* σ_l^2, σ_u^2, *and* ρ *that converge to their population values uniformly in* $P \in \mathcal{P}$. (ρ *may be equal to one in absolute value, as in the case where the width of the identification region is known*.)

(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* $\nu > 0$, K, *and* N_0 *such that* $N \geq N_0$ *implies* $\Pr(\sqrt{N}|\hat{\Delta} - \Delta| > K\Delta^{\nu}) < \epsilon$, *uniformly in* $P \in \mathcal{P}$.

Given Assumption 1 we construct the confidence interval as:

$$\overline{CI}_\alpha^\theta = \left[\hat{\theta}_l - \overline{C}_N \cdot \hat{\sigma}_l/\sqrt{N},\ \hat{\theta}_u + \overline{C}_N \cdot \hat{\sigma}_u/\sqrt{N}\right], \tag{6}$$

where $\overline{C}_N$ satisfies

$$\Phi\left(\overline{C}_N + \sqrt{N} \cdot \frac{\hat{\Delta}}{\max(\hat{\sigma}_l, \hat{\sigma}_u)}\right) - \Phi(-\overline{C}_N) = \alpha. \tag{7}$$

The following lemma gives the general uniform coverage result.

LEMMA 4: *Suppose Assumption* 1 *holds. Then*

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

Next we return to the missing data problem of Section 3. We allow for an unknown p (assuming p is bounded away from zero) and show that this problem fits the assumption sufficient for the application of Lemma 4. Because the conditional variance of Y given $W = 1$ is bounded and bounded away from zero, Assumption 1(ii) is satisfied. The lower bound can be estimated by $\hat{\theta}_l = (1/N)\sum_{i=1}^N W_i \cdot Y_i$. The upper bound can be estimated by $\hat{\theta}_u = (1/N)\sum_{i=1}^N (W_i \cdot Y_i + 1 - W_i)$. Both estimators are asymptotically normal, with $\sqrt{N}(\hat{\theta}_l - \theta_l) \xrightarrow{d} \mathcal{N}(0, \sigma_l^2)$ and $\sqrt{N}(\hat{\theta}_u - \theta_u) \xrightarrow{d} \mathcal{N}(0, \sigma_u^2)$, 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 1(i) is satisfied. Finally, consider Assumption 1(iii). 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 $K = 1/\sqrt{\epsilon}$, so that

$$\begin{aligned}\Pr\big(\sqrt{N}|\hat{\Delta} - \Delta| > K \cdot \Delta^\nu\big) &= \Pr\big(N(\hat{\Delta} - \Delta)^2 > K^2 \cdot \Delta^{2\nu}\big) \\ &< \mathbb{E}[N \cdot (\hat{\Delta} - \Delta)^2]/(K^2 \Delta^{2\nu}) \\ &\leq \Delta/(K^2 \Delta^{2\nu}) = 1/K^2 = \epsilon.\end{aligned}$$

Hence Assumption 1 is satisfied, and Lemma 4 can be used to construct a CI which is equivalent to that obtained by substituting $\hat{p}$ for p in $\widetilde{CI}_\alpha^\theta$ given in (4).

Dept. of Economics, and Dept. of Agricultural and Resource Economics, 661 Evans Hall, University of California at Berkeley, Berkeley, CA 94720-3880, U.S.A., and NBER; imbens@econ.berkeley.edu; http://elsa.berkeley.edu/users/imbens/

and

Dept. of Economics and Institute for Policy Research, Northwestern University, 2001 Sheridan Rd., Evanston, IL 60208-2600, U.S.A.; cfmanski@northwestern.edu; http://www.faculty.econ.northwestern.edu/faculty/manski/.

Manuscript received May, 2003; final revision received March, 2004.

APPENDIX

PROOF OF LEMMA 1: Define the following two events:

$$\begin{aligned}A_1 &= f(P, \lambda) \in \big[f(P_N, \lambda_l(P_N)) - C_{N0}, f(P_N, \lambda_u(P_N)) + C_{N1}\big], \\ A_2 &= \big[f(P, \lambda_l(P)), f(P, \lambda_u(P))\big] \subset \big[f(P_N, \lambda_l(P_N)) - C_{N0}, f(P_N, \lambda_u(P_N)) + C_{N1}\big].\end{aligned}$$

Because $f(P, \lambda) \in [f(P, \lambda_l(P)), f(P, \lambda_u(P))]$, it follows that A_2 implies A_1 and that the coverage probability for the set (equal to the probability of the event A_2) is less than the coverage probability for the parameter (equal to the probability of the set A_1). *Q.E.D.*

PROOF OF LEMMA 2: For the first part, fix F and p. Then

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

$$= 1 - \Pr\left(\mu \cdot p < \left(\hat{\mu} - z_{(\alpha+1)/2} \cdot \frac{\hat{\sigma}}{\sqrt{p \cdot N}}\right) \cdot p\right)$$

$$- \Pr\left(\mu \cdot p + 1 - p > \left(\hat{\mu} + z_{(\alpha+1)/2} \cdot \frac{\hat{\sigma}}{\sqrt{p \cdot N}}\right) \cdot p + 1 - p\right)$$

$$= 1 - \Pr\left(\mu < \hat{\mu} - z_{(\alpha+1)/2} \cdot \frac{\hat{\sigma}}{\sqrt{p \cdot N}}\right) - \Pr\left(\mu > \hat{\mu} + z_{(\alpha+1)/2} \cdot \frac{\hat{\sigma}}{\sqrt{p \cdot N}}\right),$$

which converges to $1 - (1 - \alpha)/2 - (1 - \alpha)/2 = \alpha$ as N gets large. For the second part consider the three possibilities for λ: $\lambda = 0$, $\lambda = 1$, and $0 < \lambda < 1$. If $\lambda = 0$, we have $\theta = \mu \cdot p$. Hence, the coverage probability of CI_α^θ is, for N large enough so that $2z_\alpha \hat{\sigma}/\sqrt{pN} > -(1 - p)/p$,

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

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

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

The second term converges to $1 - \alpha$. The third term converges to zero, which implies the coverage rate is α. A similar argument applies when $\lambda = 1$. When $\lambda \in (0, 1)$ the coverage probability converges to one. *Q.E.D.*

Before presenting a proof of Lemma 3 we present a number of preliminary results.

LEMMA 5 (Uniform Central Limit Theorem, Berry–Esseen): *Suppose $X_1, X_2, \ldots$ are independent and identically distributed random variables with c.d.f. $F \in \mathcal{F}$. Let $\bar{X}_N = \sum_{i=1}^N X_i/N$, $\mu(F) = \mathbb{E}_F[X]$, $\sigma^2(F) = \mathbb{E}_F[(X - \mu)^2]$, and let $0 < \underline{\sigma}^2 \le \sigma^2(F) \le \bar{\sigma}^2 < \infty$, and $\mathbb{E}_F[|X^3|] < \infty$ for all $F \in \mathcal{F}$. Then*

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

See, e.g., Shorack (2000). Next, we show that we can use this to construct confidence intervals for sample means with asymptotically uniform convergence even with estimated variances.

LEMMA 6: *Under the same conditions as in Lemma 5,*

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

PROOF OF LEMMA 6: First,

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

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

Hence it will suffice to show that

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

By the triangle inequality:

$$\begin{aligned}&\left|\Pr\left(\sqrt{N}\left(\frac{\bar{X}_N-\mu}{\hat{\sigma}}\right)<a\right)-\Phi(a)\right|\\&\quad\le\left|\Pr\left(\sqrt{N}\left(\frac{\bar{X}_N-\mu}{\sigma}\right)<a\frac{\hat{\sigma}}{\sigma}\right)-\Phi\left(a\frac{\hat{\sigma}}{\sigma}\right)\right|+\left|\Phi\left(a\frac{\hat{\sigma}}{\sigma}\right)-\Phi(a)\right|.\end{aligned}$$

By Lemma 5 the first term converges to zero, and by uniform convergence of $\hat{\sigma}$ to σ the second one converges to zero. *Q.E.D.*

PROOF OF LEMMA 3: First we prove that the asymptotic coverage probability is greater than or equal to α. For fixed λ the coverage probability is

$$\begin{aligned}&\Pr\left(\left(\hat{\mu}-C_N\cdot\hat{\sigma}/\sqrt{p\cdot N}\right)\cdot p\le\mu\cdot p+\lambda\cdot(1-p)\le\left(\hat{\mu}+C_N\cdot\hat{\sigma}/\sqrt{p\cdot N}\right)\cdot p+1-p\right)\\&\quad=\Pr\left(-C_N\frac{\hat{\sigma}}{\sigma}-\sqrt{N}\cdot\frac{\lambda\cdot(1-p)}{\sigma\cdot\sqrt{p}}\right.\\&\qquad\left.\le\sqrt{N}\cdot\frac{\mu-\hat{\mu}}{\sigma/\sqrt{p}}\le C_N\frac{\hat{\sigma}}{\sigma}+\sqrt{N}\cdot\frac{(1-\lambda)\cdot(1-p)}{\sigma\cdot\sqrt{p}}\right).\end{aligned}$$

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 $N\ge N_0$,

$$\begin{aligned}&\Pr\left(-C_N\frac{\hat{\sigma}}{\sigma}-\sqrt{N}\cdot\frac{\lambda\cdot(1-p)}{\sigma\cdot\sqrt{p}}\le\sqrt{N}\cdot\frac{\mu-\hat{\mu}}{\sigma/\sqrt{p}}\le C_N\frac{\hat{\sigma}}{\sigma}+\sqrt{N}\cdot\frac{(1-\lambda)\cdot(1-p)}{\sigma\cdot\sqrt{p}}\right)\\&\quad\ge\Pr\left(-C_N(1-\epsilon)-\sqrt{N}\cdot\frac{\lambda\cdot(1-p)}{\sigma\cdot\sqrt{p}}\right.\\&\qquad\left.\le\sqrt{N}\cdot\frac{\mu-\hat{\mu}}{\sigma/\sqrt{p}}\le C_N(1-\epsilon)+\sqrt{N}\cdot\frac{(1-\lambda)\cdot(1-p)}{\sigma\cdot\sqrt{p}}\right).\end{aligned}$$

For N large enough this can be made arbitrarily close to

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

for some ω. Because $C_N\le z_{(\alpha+1)/2}$ (see 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

$$\begin{aligned}&\Pr\left(\left(\hat{\mu}-C_N\cdot\hat{\sigma}/\sqrt{p\cdot N}\right)\cdot p\le\mu\cdot p+\lambda\cdot(1-p)\le\left(\hat{\mu}+C_N\cdot\hat{\sigma}/\sqrt{p\cdot N}\right)\cdot p+1-p\right)\\&\quad\ge\alpha-\nu.\end{aligned}$$

To prove equality, note that at $p=1$ the CI is identical to CI_α^μ, so in that case the asymptotic coverage rate is equal to α. *Q.E.D.*

Before proving Lemma 4 we establish a couple of preliminary results. Define $\breve{C}_N$ and $\ddot{C}_N$ by

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

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

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.

Next we give two results without proof that show that one can ignore estimation error in σ_l and σ_u.

LEMMA 7: *Suppose Assumption* 1 *holds. Then, uniformly in* $P \in \mathcal{P}$,

$$|\overline{C}_N - \ddot{C}_N| \longrightarrow 0.$$

LEMMA 8: *For all* $\epsilon > 0$, *there is an* N_0 *such that for* $N \geq N_0$, *uniformly in* $P \in \mathcal{P}$ *and* $\lambda \in \Lambda$,

$$\begin{aligned}&\big|\Pr\big(\hat{\theta}_l - \overline{C}_N\cdot\hat{\sigma}_l/\sqrt{N} \leq \theta \leq \hat{\theta}_u + \overline{C}_N\cdot\hat{\sigma}_u/\sqrt{N}\big)\\ &\quad - \Pr\big(\hat{\theta}_l - \ddot{C}_N\cdot\sigma_l/\sqrt{N} \leq \theta \leq \hat{\theta}_u + \ddot{C}_N\cdot\sigma_u/\sqrt{N}\big)\big| < \epsilon.\end{aligned}$$

The next two lemmas account for the effects of estimation error in Δ.

LEMMA 9: *For any* $\eta, \epsilon > 0$, *there is an* N_0 *such that for* $N \geq N_0$, *uniformly in* $P \in \mathcal{P}$,

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

PROOF OF LEMMA 9: Because $\ddot{C}_N$ satisfies $\Phi(\ddot{C}_N + \sqrt{N}\hat{\Delta}/\max(\sigma_l,\sigma_u)) - \Phi(-\ddot{C}_N) = \alpha$, we only need to prove that

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

By Assumption 1(iii) there are ν, K, and N_0 such that with $\delta = \nu/5$ and $N \geq \max(N_0, K^{1/\delta})$,

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

Then:

$$\begin{aligned}&\Phi\left(\ddot{C}_N + \sqrt{N}\cdot\frac{\hat{\Delta}}{\max(\sigma_l,\sigma_u)}\right) - \Phi\left(\ddot{C}_N + \sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)}\right)\\ &\quad = \mathbb{1}\{\hat{\Delta} \leq \Delta\}\cdot\left(\Phi\left(\ddot{C}_N + \sqrt{N}\cdot\frac{\hat{\Delta}}{\max(\sigma_l,\sigma_u)}\right) - \Phi\left(\ddot{C}_N + \sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)}\right)\right)\\ &\qquad + \mathbb{1}\{\hat{\Delta} > \Delta, \sqrt{N}|\hat{\Delta} - \Delta| \leq N^{\delta}\Delta^{\nu}\}\\ &\qquad \times\left(\Phi\left(\ddot{C}_N + \sqrt{N}\cdot\frac{\hat{\Delta}}{\max(\sigma_l,\sigma_u)}\right) - \Phi\left(\ddot{C}_N + \sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)}\right)\right)\\ &\qquad + \mathbb{1}\{\hat{\Delta} > \Delta, \sqrt{N}|\hat{\Delta} - \Delta| > N^{\delta}\Delta^{\nu}\}\\ &\qquad \times\left(\Phi\left(\ddot{C}_N + \sqrt{N}\cdot\frac{\hat{\Delta}}{\max(\sigma_l,\sigma_u)}\right) - \Phi\left(\ddot{C}_N + \sqrt{N}\cdot\frac{\Delta}{\max(\sigma_l,\sigma_u)}\right)\right)\end{aligned} \tag{8}$$

$$\leq \mathbb{1}\{\hat{\Delta} > \Delta, \sqrt{N}|\hat{\Delta} - \Delta| \leq N^{\delta}\Delta^{\nu}\} \tag{9}$$

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

$$+ \mathbb{1}\{\sqrt{N}|\hat{\Delta} - \Delta| > N^{\delta}\Delta^{\nu}\}, \tag{10}$$

using the fact that (8) is nonpositive. The expectation of (10) is less than ϵ. By a mean value theorem (9) is, for some $\gamma \in [0, 1]$, equal to

$$\mathbb{1}\{\hat{\Delta} > \Delta, \sqrt{N}|\hat{\Delta} - \Delta| \leq N^{\delta}\Delta^{\nu}\}$$

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

Because the product is zero unless $\hat{\Delta} > \Delta$, and $\ddot{C}_N, \Delta \geq 0$, this can be bounded from above by

$$\mathbb{1}\{\hat{\Delta} > \Delta, \sqrt{N}|\hat{\Delta} - \Delta| \leq N^{\delta}\Delta^{\nu}\} \cdot \phi\left(\sqrt{N} \cdot \frac{\Delta}{\max(\sigma_l, \sigma_u)}\right) \cdot \sqrt{N} \cdot \frac{\hat{\Delta} - \Delta}{\max(\sigma_l, \sigma_u)} \tag{11}$$

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

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

Maximizing this over Δ gives

$$N^{\delta-\nu}\exp(-\nu/2)\nu^{\nu/2}\max(\sigma_l, \sigma_u)^{\nu-1}/\sqrt{2\pi}.$$

Given that $\delta < \nu$, this can be bounded arbitrarily close to zero uniformly in $P \in \mathcal{P}$. *Q.E.D.*

LEMMA 10: *For any $\eta, \epsilon > 0$, there is an N_0 such that for $N \geq N_0$, uniformly in $P \in \mathcal{P}$,*

$$\Pr(\ddot{C}_N < \breve{C}_N - \eta) < \epsilon.$$

PROOF: 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 9 there is an N_0 such that for $N \geq N_0$

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

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

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

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

$$- \Phi(-\ddot{C}_N) + \Phi(-\breve{C}_N) > -\eta \cdot \underline{\phi}$$

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

$$\left. + \phi(\breve{C}_N + \gamma \cdot (\ddot{C}_N - \breve{C}_N))\right) \cdot (\ddot{C}_N - \breve{C}_N) > -\eta \cdot \underline{\phi}$$

$$\Longrightarrow \underline{\phi} \cdot (\ddot{C}_N - \check{C}_N) > -\eta \cdot \underline{\phi}$$

$$\Longrightarrow \ddot{C}_N - \check{C}_N > -\eta,$$

for some $\gamma \in [0,1]$ by the mean value theorem, and thus with probability $\ddot{C}_N - \check{C}_N > -\eta$ with probability at least $1-\epsilon$. *Q.E.D.*

Note that Lemma 10 does not imply that $|\overline{C}_N - \check{C}_N|$ 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 with which we are concerned.

PROOF OF LEMMA 4: We will prove that for any positive ϵ, for N sufficiently large,

$$\Pr\bigl(\hat{\theta}_l - \overline{C}_N \cdot \hat{\sigma}_l/\sqrt{N} \le \theta \le \hat{\theta}_u + \overline{C}_N \cdot \hat{\sigma}_u/\sqrt{N}\bigr) \ge \alpha - \epsilon,$$

uniformly in $P \in \mathcal{P}$. We will prove this for $\theta = \theta_u$. The proof for $\theta = \theta_l$ is analogous, and by joint normality of the estimators for the upper and lower boundary 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), $\sup_z |\Pr(\sqrt{N}(\hat{\theta}_l - \theta_l)/\sigma_l \le z) - \Phi(z)| \le \epsilon_1$, (ii), $\sup_z |\Pr(\sqrt{N}(\hat{\theta}_u - \theta_u)/\sigma_u \le z) - \Phi(z)| \le \epsilon_1$, and (iii), $\Pr(\ddot{C}_N - \check{C}_N < -\epsilon_2) < \epsilon_3$. Existence of such an N follows for conditions (i) and (ii) from Assumption 4.2, and for condition (iii) from Lemma 10. Define the following events:

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

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

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

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

$$E_5 \equiv \ddot{C}_N - \check{C}_N > -\epsilon_2, \quad \text{and} \quad E_5^c \equiv \ddot{C}_N - \check{C}_N \le -\epsilon_2.$$

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

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

and

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

By conditions (i) and (ii), $|P_3 - \Pr(E_3)| \le 2\epsilon_1$ and $|P_4 - \Pr(E_4)| \le 2\epsilon_1$. Also, $|P_3 - P_4| \le 2\epsilon_2 \bar{\phi}$, and by (iii), $\Pr(E_6^c) < \epsilon_3$. By Lemma 8 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

$$\begin{aligned}\Pr(E_1) &\ge \Pr(E_2) - \epsilon_4 \ge \Pr(E_2 \cap E_3) - \epsilon_4 \ge \Pr(E_5 \cap E_3) - \epsilon_4 \ge \Pr(E_3) - \Pr(E_5^c) - \epsilon_4 \\ &\ge P_3 - 2\epsilon_1 - \epsilon_3 - \epsilon_4 \ge P_4 - 2\epsilon_1 - \epsilon_3 - 2\epsilon_2\bar{\phi} - \epsilon_4 = \alpha - 2\epsilon_1 - \epsilon_3 - 2\epsilon_2\bar{\phi} - \epsilon_4.\end{aligned}$$

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

REFERENCES

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2003): "Parameter Set Inference in a Class of Econometric Models," Unpublished Manuscript, Department of Economics, Princeton University.

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.

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.

SHORACK, G. (2000): *Probability for Statisticians*. New York: Springer-Verlag.