DISCUSSION PAPER SERIES

IZA DP No. 5816

Calculating Confidence Intervals for Continuous and Discontinuous Functions of Estimated Parameters

John Ham
Tiemen Woutersen

June 2011

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Calculating Confidence Intervals for Continuous and Discontinuous Functions of Estimated Parameters

John Ham
University of Maryland,
IRP (Madison), IFAU and IZA

Tiemen Woutersen
Johns Hopkins University

Discussion Paper No. 5816
June 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 5816
June 2011

ABSTRACT

Calculating Confidence Intervals for Continuous and Discontinuous Functions of Estimated Parameters*

The delta method is commonly used to calculate confidence intervals of functions of estimated parameters that are differentiable with non-zero, bounded derivatives. When the delta method is inappropriate, researchers usually first use a bootstrap procedure where they i) repeatedly take a draw from the asymptotic distribution of the parameter values and ii) calculate the function value for this draw. They then trim the bottom and top of the distribution of function values to obtain their confidence interval. This note first provides several examples where this procedure and/or delta method fail to provide an appropriate confidence interval. It next presents a method that is appropriate for constructing confidence intervals for functions that are discontinuous or are continuous but have zero or unbounded derivatives. In particular the coverage probabilities for our method converge uniformly to their nominal values, which is not necessarily true for the other methods discussed above.

JEL Classification: C12, C15

Keywords: confidence intervals, simulation, structural models, policy effects

Corresponding author:

John C. Ham
Department of Economics
University of Maryland
3105 Tydings Hall
College Park, MD 20742
USA
E-mail: ham@econ.umd.edu

* Comments are welcome. We thank Nicole Lott, Geert Ridder, John Rust, Jonathan Wright and especially Xianghong Li for helpful comments and discussions.

1. Introduction

Applied researchers often would like to estimate confidence intervals for functions of estimated parameters. If the function is differentiable and has non-zero and bounded derivatives they can use the delta method.[1] However, if the function is nondifferentiable, or has zero or unbounded derivatives, as in the case of simulating functions with zero-one outcomes, researchers usually follow the following procedure to obtain $(1-\alpha)*100\%$ confidence intervals for the function evaluated at the parameter estimates: sample from the asymptotic (normal) distribution of the parameter estimates, calculate the function value for each draw, and then trim $\alpha/2$ from each tail of the distribution of the function values.[2] Further, authors sometimes will use this approach for relatively complex differentiable functions (with bounded nonzero derivatives) such as expected durations of unemployment (Fitzenberger, Osikominu and Paul 2010) or impulse functions.[3] Although this procedure of sampling from the asymptotic distribution is sometimes called the parametric bootstrap, this term has more than one meaning, and thus instead we describe it as the AD-bootstrap. This note provides several important examples in which the AD-bootstrap fails. Further, one cannot avoid problems with the AD-bootstrap using the approach advocated by Cameron and Trivedi (2009) in their influential book on implementing sophisticated econometric estimators in the widely used Stata program; they suggest to take a bootstrap sample, reestimate the model and the function of interest, repeat this many times, trim the bootstrap distribution of the function values, and use this trimmed distribution to calculate the relevant confidence interval.[4] The issue here is that the AD-bootstrap is first order equivalent to the Cameron-Trivedi procedure for cases that the latter estimates the asymptotic distribution of the function consistently (which will not always be the case).

The purpose of this note is to provide a method of obtaining confidence intervals for

[1] See, e.g., Weisberg (2005) for a description of the delta method.

[2] This approach is used by: Fitzenberger, Osikominu and Paul (2010), Gaure, Røed and Westlie (2010), Hitsch, Hortacsu and Ariely (2010), Merlo and Wolpin (2009) and Røed and Westlie (2010). Its use is advocated, but not implemented, by Eberwein, Ham and LaLonde (2001). A review of the literature indicates that many studies either i) do not give a confidence interval for the simulated results or ii) give a confidence interval for the simulated results but do not state how they construct it.

[3] See Inoue and Kilian (2011) for a recent overview of the impulse response function literature.

[4] This assumes that the Cameron and Trivedi procedure produces a consistent estimate of the asymptotic distribution of the estimator.

these functions under relatively mild conditions that are likely to be satisfied in empirical work. Further, our approach has the interesting feature that it can be used to calculate a confidence interval for a parameter for which asymptotic distribution cannot be estimated using any known version of the bootstrap method. More precisely, Romano and Shaikh (2010) give conditions under which subsampling and the ordinary bootstrap yield uniform asymptotically valid results for obtaining a confidence interval of a parameter. Suppose these conditions hold, so that it is possible to construct a confidence interval for the estimated parameters for which the coverage probabilities converge uniformly to their nominal level. Below we show that one can construct functions of the subsample estimates where the resulting confidence interval for the function will have a coverage probability that is incorrect for any sample size. We provide a method of constructing such confidence intervals that works with all of these examples, and then show that our method provides confidence intervals that are substantially different from the (incorrect) confidence intervals produced by the AD-bootstrap in a serious empirical project.

Our note proceeds as follows. In Section 2 we show that in several important examples, the AD-bootstrap fails to provide a confidence interval with the correct coverage. Section 3 proposes a simple procedure to estimate consistent confidence intervals for both differentiable and nondifferentiable functions which we refer to as the CI-bootstrap. This procedure samples from the $1-\alpha$ confidence interval of the estimated parameter, calculates the function value for each of the draws, and then uses all of the function values to obtain $1-\alpha$ confidence interval for the value of the function at the estimated parameter. In Section 4 we draw on empirical work by Ham, Li and Shore-Sheppard (2011), who simulate a discontinuous function to estimate the effect of changes in demographic on the expected time spent in employment at three, six and ten years after the change, to compare our approach with the AD-bootstrap. We find that the two methods produce substantially different confidence intervals and our results suggest that previous empirical work is likely to have been overly optimistic in terms of finding confidence intervals that were too small. Thus differences between the AD-bootstrap and the CI-bootstrap are likely to be important in practice, as well as in principle.

2. Failures of the delta method and AD-bootstrap when calculating confidence intervals for functions of estimated parameters

In example one the AD-bootstrap (and any other bootstrap that estimates the distribution of the estimator of the parameter consistently) fails. In the second example the delta method is infeasible while no version of the bootstrap consistently estimates the asymptotic distribution of the function of the estimator. In the third example, the delta method fails again.

Example 1: Simulating a Probit Model

Suppose that one is interested in the function $h(\beta, \gamma) = \frac{1}{2}\Phi(\beta) + \frac{1}{2}\Phi\left(-2\gamma - \sqrt{2\ln(2)}\right)$ where $\beta = \gamma = 0$. Let the estimator $\binom{\hat{\beta}}{\hat{\gamma}}$ have a normal distribution with mean zero and a known variance-covariance matrix $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. The delta method cannot be used since the function has a zero derivative at the true value of the parameters. The AD-bootstrap samples from the normal distribution with mean $\binom{\hat{\beta}}{\hat{\gamma}}$ and covariance matrix $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ in order to construct the distribution of $h(\hat{\beta}, \hat{\gamma})$. Let $G(.)$ denote this distribution function of $h(\hat{\beta}, \hat{\gamma})$. Applying the AD-bootstrap (or any other bootstrap that consistently estimates the distribution of $\hat{\beta}, \hat{\gamma})$ and using the interval between the 2.5 and 97.5 percentile, $[G(0.025), G(0.975)]$, does *not* yield a confidence interval with 95% coverage for many values of ρ. For example, the coverage is 0% for the AD-bootstrap if $\rho = 1$, 90% for $\rho = 0.5$, and 93% for $\rho = 0$.[5] Thus the AD-bootstrap (and any other bootstrap that consistently estimates the distribution of $h(\hat{\beta}, \hat{\gamma})$) does not produce a confidence interval with the correct coverage[6]. We would also note that the AD-bootstrap confidence interval coincides with the Bayesian credible interval (with flat priors) in this case, so the Bayesian procedure also fails here.

Example 2: (Andrews, 2000)

Suppose we observe a random sample, $X_1, ..., X_N$, from a normal distribution with mean μ and variance 1 (denoted $N(\mu, 1)$) and suppose that μ is restricted to be nonnegative. Andrews (2000) considers the maximum likelihood estimator $\hat{\mu} = \max(\bar{X}_N, 0)$ where

[5] See the Appendix for more detail on this and the following examples.

[6] Correct coverage of a confidence interval means that the coverage probability converges to a probability no smaller than its nominal probability.

$\bar{X}_N = \frac{1}{N}\sum X_i$. He shows that the regular bootstrap fails to consistently estimate the distribution of $\hat{\mu}$ if $\mu = 0$ and that it is impossible to consistently estimate the distribution of $\hat{\mu}$ if $\mu = \frac{c}{\sqrt{N}}$ for some $c > 0$. This note does not attempt to estimate the distribution of $\hat{\mu}$ but it shows below that the CI-bootstrap can consistently estimates the confidence interval for μ.

Example 3: A Continuous Function

Suppose we observe a random sample, $X_1, ..., X_N$, from a normal distribution with mean μ and variance 1 and let $\hat{\mu} = \bar{X}_N = \frac{1}{N}\sum X_i$. Let $\mu_0 = E(X) = 0$ and consider

$$h(\mu) = \begin{cases} \sqrt{|\mu|} \text{ if } \mu \geq 0 \\ -\sqrt{|\mu|} \text{ if } \mu < 0 \end{cases}$$

The delta method yields the following symmetric 95% confidence interval:

$$\left[\text{sign}(\bar{X}_N)\sqrt{|\bar{X}_N|} - \frac{1.96}{2\sqrt{|\bar{X}_N|}}, \text{sign}(\bar{X}_N)\sqrt{|\bar{X}_N|} + \frac{1.96}{2\sqrt{|\bar{X}_N|}}\right].$$

The probability that the true value is inside this confidence interval is about 0.67 while our method gives a confidence interval with the correct size, 0.95.

Of course, there are examples where the delta method and/or the AD-bootstrap will have correct coverage.

3. Main result and applications

In this section we provide a method of obtaining confidence intervals that is valid under reasonable assumptions which are likely to be satisfied in empirical work. Let the dimension of θ be equal to K and let $h(\theta)$ have dimension H. Let $CI^{\theta}_{1-\alpha}$ denote the $(1-\alpha)$ confidence interval of θ. Suppose we sample M times from $CI^{\theta}_{1-\alpha}$ and include every point in $CI^{\theta}_{1-\alpha}$ that is no farther than the Euclidian distance $\eta > 0$ from a sampled point.[7] For each sampled point, we calculate $h(\theta)$. Let $\widehat{CI^{h(\theta)}_{1-\alpha}}$ denote the resulting confidence interval. More precisely, let

$$CI^{h(\theta)}_{1-\alpha} = \{\omega \in \mathbb{R}^H | \omega = h(\theta) \text{ for some } \theta \in CI^{\theta}_{1-\alpha}\}.$$

[7] If θ_S is sampled, then any $\theta \in \Theta$ for which $||\theta_S - \theta||_2 \leq \eta$ is included in the CI for some parameter space Θ.

In fact $\widehat{CI_{1-\alpha}^{h(\theta)}}$ is what we described heuristically above as the CI-bootstrap; we here note intuitively why the AD-bootstrap can fail. The AD-bootstrap samples from the entire asymptotic distribution of $\hat{\theta}$ and forms the confidence interval of $h(\theta)$ by trimming the extreme $(1-\alpha)/2$ from the upper and lower tails of the resulting distribution for $h(\theta)$. Note that the extreme values of $h(\theta)$ that the AD-bootstrap trims can arise either i) they are based on an extreme draw from the asymptotic distribution of θ or ii) a **"**reasonable**"** draw for θ results in an extreme value of $h(\theta)$.The CI-bootstrap instead samples from the $1-\alpha$ confidence interval of θ and includes all of the resulting values of $h(\theta)$ in its $1-\alpha$ confidence interval, and thus appropriately only trims case i) above.[8] Moreover, note that constructing a confidence interval using the CI-bootstrap is no more difficult than constructing one using the AD-bootstrap. We now state the assumptions necessary for the CI-bootstrap to hold.

Assumption 1

Let (i) $\theta \in \Theta$, *which is compact; and (ii)*

$$\lim_{N\to\infty} \inf_{P\in\mathcal{P}} \Pr(\theta \in CI_{1-\alpha}^{\theta}) \geq 1-\alpha.$$

Note that Assumption 1 simply says that the confidence interval for the parameter contains the true parameter value with probability $1-\alpha$ in the limit. This will certainly hold for any estimator that is uniformly asymptotically normally distributed, as well as for the subsampling and bootstrap of θ under appropriate regularity conditions (see Romano and Shaikh (2010)).

Assumption 2

Let $h(\theta)$ *be bounded for all* $\theta \in \Theta$. *Let there exist a partitioning of the parameter space such that* $\Theta_1 \cup \Theta_2 ... \cup \Theta_R = \Theta$, *where* $R < \infty$, *and let* $h(\theta)$ *be uniformly continuous*[9] *for all* $\theta \in \Theta_r$, $r = 1, ..., R$.

[8] It is perhaps worth noting that in results available from the authors, one can show that a sufficient condition for the AD-bootstrap to work in the univariate case is that $h()$ be a monotonic function. However, as example 1 shows, when we move beyond a univariate function, this no longer holds, since the function here is monotonic in both parameters.

[9] The vector-function $h(\theta)$ is uniformly continuous on Θ_j if for each $\varepsilon > 0$ there is an $\eta > 0$ such that $||h(\theta_1) - h(\theta_2)|| < \varepsilon$ for all $\theta_1, \theta_2 \in \Theta_j$ with $||\theta_1 - \theta_2|| < \eta$ where $||.||$ is the Euclidean norm.

The second assumption allows $h(\theta)$ to be discontinuous. In general, the parameter space is partitioned into R subspaces so that the regions of discontinuity can have dimension $R-1$, with the only restriction that R is finite. This again is likely to be satisfied in any empirical application.

Before proving the theorem, intuition for our result can be obtained if we continue our consideration of *Example 2.* As noted above,

Andrews (2000) considers the maximum likelihood estimator $\hat{\mu} = \max(\bar{X}_N, 0)$ where $\bar{X}_N = \frac{1}{N}\sum X_i$ is normally distributed with mean μ, $\mu \geq 0$, and variance $\frac{1}{N}$. He shows that the regular bootstrap fails to consistently estimate the distribution of $\hat{\mu}$ if $\mu = 0$, and that it is impossible to consistently estimate (using any version of the bootstrap) the distribution of $\hat{\mu}$ if $\mu = \frac{c}{\sqrt{N}}$ for some $c > 0$. However, the CI-bootstrap can be used to calculate a 95% confidence interval (or a κ% confidence interval for any κ) for $\hat{\mu}$ in spite of the absence of a consistent estimator of the asymptotic distribution. In particular, the symmetric 95% confidence is $\left[\max\left(0, \bar{X}_N - \frac{1.96}{\sqrt{N}}\right), \max\left(0, \bar{X}_N + \frac{1.96}{\sqrt{N}}\right)\right]$, which contains μ with probability 0.95, including the case where $\mu = \frac{c}{\sqrt{N}}$ for some $c > 0$.[10]

At this point it is appropriate to state our theorem:

Theorem

Let Assumptions 1-2 hold. Then

$$\lim_{M\to\infty} \lim_{N\to\infty} \inf_{P\in\mathcal{P}} \Pr\left(h(\theta) \in \widehat{CI_{1-\alpha}^{h(\theta)}}\right) \geq 1-\alpha.$$

Proof: See appendix.

If one relaxes the uniformity requirement[11] in Assumption 1, i.e. one assumes that

$$\lim_{M\to\infty} \lim_{N\to\infty} \Pr(\theta \in CI_{1-\alpha}^{\theta}) \geq 1-\alpha,$$

then the theorem holds without the uniformity result, i.e.

$$\lim_{M\to\infty} \lim_{N\to\infty} \Pr(h(\theta) \in CI_{1-\alpha}^{h(\theta)}) \geq 1-\alpha.$$

[10] This example also illustrates that one should perhaps not focus too much on the distribution of the bootstrap when the goal simply is to derive a confidence interval. Also, Hirano and Porter (2010) derive more impossibility results in the spirit of Andrews (2000).

[11] Andrews' (1987) emphasizes of the importance of uniform convergence.

Remark 1: Shrinking the Size of the Confidence Interval produced by the CI-bootstrap.

As is well known, the delta method uses a linear approximation of $h(\theta)$. One can use a similar linear approximation here which in many cases will the confidence regions smaller for the CI-bootstrap. To begin, suppose that $h(\theta)$ is a scalar and the parameter vector θ has length K. Moreover, suppose one samples from the asymptotic distribution and calculates the quantity $\lambda^{[m]} = \sum_{k=1}^{K}(\theta_{k,j}^{[m]} - \hat{\theta}_k)\hat{w}_k$ where m denotes the draw, $m = 1, ..., M$, and $\hat{w}_k, k = 1, ..., K$, is some nonzero weight. If $h(\theta)$ is continuously differentiable, as in example 1, then one can choose $\hat{w}_k$ to be the partial derivative, $\hat{w}_k = \left\{ \left(\frac{\partial h(\theta)}{\partial \theta_k} \right) \Big|_{\theta=\hat{\theta}} \right\}$ for all k. If $h(\theta)$ is not continuously differentiable, then one can choose the weights to be numerical derivatives. That is, $\hat{w}_k = \{h(\theta + i_k\varepsilon_k) - h(\theta - i_k\varepsilon_k)\}/(2\varepsilon_k)$ for all k where i_k is a row vector whose kth element is one and all other elements are zero. In what follows we refer to this approach as the weighted CI-bootstrap. Note that the delta method fails in general when the linear approximation fails to hold asymptotically (i.e. the function is nondifferentiable asymptotically), but the method presented here remains valid if the linear approximation fails. In using this approximation, we found that in simulations we had the best results by choosing nonstochastic, nonzero weights or asymptotically nonstochastic, nonzero weights i.e. $\hat{w}_k = w_k + o_p(1)$, where $w_k \neq 0$ for all k. After choosing the weights, one can the derive a valid confidence interval for λ. Let $[\underline{c}, \overline{c}]$ denote this confidence interval[12]. One can then select the parameter draws for which $\underline{c} \leq \lambda_m \leq \overline{c}$, calculate the value of $h(\theta_m)$ for each draw (within the confidence interval) $m = 1, ..., M$, and calculate $\widehat{CI_{1-\alpha}^{h(\theta)}}$ as above.

For linear functions of the form $h(\theta) = c_0 + \sum_{k=1}^{K} c_k\theta_k$, one obviously has $\frac{\partial h(\theta)}{\partial \theta_k} = c_k$. If the function $h(\theta)$ is a single index function such that $h(\theta) = g\left(\theta_1 + \sum_{k=2}^{K} \hat{w}_k\theta_k\right)$ for some $g(.)$ and $\theta_1 \neq 0$, then $\frac{\partial g(\theta)}{\partial \theta_k} = \frac{\partial g(\theta)}{\partial \theta_1}\hat{w}_k$, $\hat{w}_1 = 1$, and $\frac{\partial g(\theta)}{\partial \theta_k} / \frac{\partial g(\theta)}{\partial \theta_1} = \hat{w}_k$ for $k \geq 2$. Thus, the class of functions that can be approximated is much larger than just linear functions. More importantly, the difference from the delta method is that a failure to approximate

[12] For example, $\underline{c}$ is the 2.5 percentile, $\overline{c}$ is the 97.5 percentile and $[\underline{c}, \overline{c}]$ is a 95% confidence interval.

$h(\theta)$ well does not cause this modification of the CI-bootstrap to be inconsistent. Also note that if $h(\theta)$ is a linear function of the parameters, then the delta method and the method proposed in this note, using $\lambda^{[m]} = \sum_{k=1}^{K} \left(\theta_{k,j}^{[m]} - \hat{\theta}_k\right) \left\{ \left(\frac{\partial h(\theta)}{\partial \theta_k}\right)\Big|_{\theta=\hat{\theta}} \right\}$, coincide; see the appendix for details. If $h(\theta)$ is a vector valued function, then the weighting is done in the same way as for the delta method, with $\hat{w}_k, k = 1, ..., K$, replacing the partial derivatives. Analogous to the case above where $h(\theta)$ is a single index function, this approximation gives $CI_{1-\alpha}^{\theta}$, which is then used to calculate $CI_{1-\alpha}^{h(\theta)}$. This weighting is similar to using a weighting matrix when applying the method of moment estimator. In particular, using a weighting matrix that does *not* converge to the efficient weighting matrix does not, in general, cause the method of moment estimator to be inconsistent, see Newey and McFadden (1994). The same is true here for the choice of weights, $\hat{w}_k$, $k = 1, ..., K$. Choosing an efficient weighting matrix is, in general, a good idea and here we suggest to use, if possible, the partial derivatives or approximations of them (if the partial derivatives do not exist). Moreover, we suggest to use nonzero weights, just like Newey and West (1987) and Andrews (1991) suggest to use estimates of the (efficient) weighting matrix that are positive semi-definite. We provide an algorithm for implementing the weighted CI-bootstrap in the appendix.

Remark 2 : Consider the parameter estimate $\hat{\mu}$ from example 2. As noted above, Andrews (2000) shows that the regular bootstrap fails to consistently estimate the distribution of $\hat{\mu}$ if $\mu = 0$ and that it is impossible to consistently estimate the distribution of $\hat{\mu}$ if $\mu = \frac{c}{\sqrt{N}}$ for some $c > 0$.Then the CI-bootstrap can be used directly to calculate a confidence interval for a function of $\hat{\mu}$, $h(\hat{\mu})$.

We now turn to investigating the differences between the AD-bootstrap and the CI-bootstrap within the context of a serious empirical study and defer studying the weighted CI-bootstrap for a later draft of the note.

4. A Comparison of the confidence intervals produced by the AD- and CI-bootstraps for the fraction of time a disadvantaged woman spends in employment.

Ham, Li and Shore-Sheppard (2011) estimate a model of the employment dynamics of disadvantaged mothers (i.e. single mothers with a high school degree or less) for the U.S. Specifically, they estimate four hazard functions for these women: i) a nonemployment spell in progress at the start of the sample, i.e. a left censored nonemployment spell; ii) for those in an employment spell in progress at the start of the sample, i.e. a left censored employment spell; iii) a nonemployment spell that begins after the start of the sample, i.e. a fresh nonemployment spell and iv) for an employment spell that begins after the start of the sample, i.e. a fresh employment spell.[13] Next they use simulation to interpret the values of their parameter estimates. Specifically, they first simulate the model for each member of the sample at the actual individual values of the explanatory variables to obtain baseline values for the fraction of the sample in employment at 3 years, 6 years and 10 years after the start of the simulation;[14] The results of this simulation are in Part A of Table 1. Considering column 1 of row 1 in Part A, we see that the expected time spent in employment 3 years from the beginning of the sample is 0.431. Column 1 of the next two rows indicates that the confidence intervals for the AD-bootstrap and the CI-bootstrap at the 95% confidence level are $[0.414, 0.449]$ and $[0.396, 0.469]$ respectively. Thus both methods provide relatively tight confidence intervals for this proportion, but the interval produced by the CI-bootstrap are larger than those for the AD-bootstrap. They obtain similar results for the expected proportion of time spent in employment at 6 and 10 years.

Part B of the table shows the effects of changing some of the explanatory variables. For example, Part B first shows the difference in expected portion of time employed by those with a high school degree and those with less than a high school degree at 3, 6 and 10 years. To obtain this difference they first simulate the model when the dummy variable for 12 years of schooling is set equal to 1. They next simulate the model when the dummy variable for 12 years of schooling is set equal to 0. In both cases they leave unchanged the

[13] They allow the unobserved heterogeneity components in the hazard function to be correlated and this correlation is taken into account in the simulations.

[14] All results in Table 1 are based on 20,000 draws.

values of the other explanatory variables. The entries in the first row of Part B show the differences in the fraction employed between the two groups at 3, 6 and 10 years. The first entry in this row indicates that, as one would expect, high school graduates spend 9 percentage points more time in employment at 3 years than do high school dropouts. From the next two lines we see that the 95% confidence intervals for the AD-bootstrap and CI-bootstrap are $[0.072, 0.107]$ and $[0.053, 0.126]$ respectively. Thus both confidence intervals allow one to reject the null hypothesis that this effect is zero, but again the confidence interval for the CI-bootstrap is larger than the confidence interval for the AD-bootstrap. The results for the effects of changing the level of education on the fraction employed at 6 and 10 years out are quite similar.

The next line of column 1 shows that the difference in the expected fraction of time spent in employment between Whites and African Americans at 3 years is -0.031; again this result is intuitively plausible. Now 95% the confidence intervals for the AD-bootstrap and CI-bootstrap are $[-0.050, -0.009]$ and $[-0.073, 0.019]$ respectively. Thus not only is the 95% confidence interval for the CI-bootstrap larger than the confidence interval for the AD-bootstrap, but now only the latter allows one to reject the null hypothesis that the effect is significantly different from zero. The corresponding results in columns 2 and 3 for the effect at 6 and 10 years out respectively tell the same story. Finally column 1 of the next line indicates that they estimate the effect of having at least one child less than 6 years of age versus not having a child in that age range is to reduce the fraction of time spend in employment by 3 percentage points at three years out; the next two lines indicate that only the AD-bootstrap allows one to reject the null hypothesis that this effect is zero. The results in columns 2 and 3 for the effect 6 and 10 years out are qualitatively similar.

The results in Table 1 suggest that the CI-bootstrap will produce wider confidence intervals than that for the AD-bootstrap. This result is intuitively plausible. Extreme values of the function can come from extreme draws of the parameter distribution and nonextreme draws from the parameter distribution, and the AD-bootstrap inappropriately trims extreme function values from both sources, while the CD-bootstrap only trims those arising from extreme values of the parameters. Since many previous empirical studies use the AD-bootstrap, our results indicate that it is likely that they have inadvertently provided confidence levels that are too small, and correspondingly have overstated the statistical significance of their simulation results.

5. Conclusion

The AD-bootstrap is commonly used to calculate confidence intervals for functions of estimated parameters that are nondifferentiable or have non-zero but unbounded derivatives. However, this approach produces inappropriate confidence intervals in the sense that they have less than complete coverage; moreover our empirical results suggest that these confidence intervals will often be too small in a meaningful way. This note presents an alternative that provides appropriate confidence intervals for these type of functions, as well as differentiable functions with bounded derivatives under conditions that are likely to be satisfied in empirical work.

Further, the coverage probabilities for our method converge uniformly to their correct values, which is not necessarily true of other methods. Further, Andrews (2000) gives an example in which all versions of the bootstrap fail to consistently estimate the distribution of the maximum likelihood estimator. The CI-bootstrap works in this example and also in other examples where the delta method or the bootstrap fails. An interesting property of the weighted CI-bootstrap is that it produces the same confidence interval as the delta method if the linear approximation holds, so there is no efficiency loss in that sense.

References

[1] Andrews, D.W.K. (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica,* 55: 1465–1472.

[2] Andrews, D.W.K. (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59: 817–858.

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

[4] Cameron, A. C. and P. K. Trivedi (2010): *Microeconometrics Using Stata (* Revised Edition), College Station, Texas: Stata Press.

[5] Eberwein, C., J. Ham, and R. LaLonde (2002): "Alternative Methods of Estimating Program Effects in Event History Models," *Labour Economics*, 9: 249-278

[6] Fitzenberger, B., A. Osikominu and M. Paul (2010): "The Heterogeneous Effects of Training Incidence and Duration on Labor Market Transitions," Mimeo, Albert–Ludwigs–University Freiburg.

[7] Gaure, S., K. Røed,K. and L. Westlie (2010): "Job Search Institutions and Job Match Quality," Mimeo, The Ragnar Frisch Centre for Economic Research.

[8] Ham, J., X. Li and L. Shore-Sheppard (2011): "Correcting for Seam Bias when Estimating Discrete Variable Models, with an Application to Analyzing the Employment Dynamics of Disadvantaged Women in the SIPP," Mimeo, University of Maryland.

[9] Hirano, K. and J. Porter (2010): "Impossibility Results for Nondifferentiable Functionals," University of Arizona working paper.

[10] Hitsch,G., A. Hortacsu and D. Ariely (2010): "Matching and Sorting in Online Dating," *American Economic Review* 100:130-163.

[11] Inoue, A., and L. Kilian (2011): "Inference on Impulse Response Functions in Structural VAR Models," Mimeo, North Carolina State University.

[12] Merlo, A. and K. Wolpin (2009): "The Transition from School to Jail: Youth Crime and High School Completion Among Black Males," Mimeo, University of Pennsylvania.

[13] Newey, W. K., and D. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in the *Handbook of Econometrics,* Vol. 4, ed. by R. F. Engle and D. MacFadden. Amsterdam: North-Holland.

[14] Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55: 703–708.

[15] Røed, K. and L. Westlie (2011) “Unemployment Insurance in Welfare States: Soft Constraints and Mild Sanctions," Forthcoming in the *Journal of the European Economic Association.*

[16] Romano, J. P. and A. M. Shaikh (2010). “On the Uniform Asymptotic Validity of Subsampling and the Bootstrap,” Mimeo, University of Chicago.

[17] Weisberg, S. (2005). *Applied Linear Regression*, third edition. New York:Wiley.

6. Appendix

Example 1

Consider $\rho = 1$ and let

$$\begin{aligned} \underline{h}(\beta) &= \frac{1}{2}\Phi(\beta) + \frac{1}{2}\Phi\left(-2\beta - \sqrt{2\ln(2)}\right) \\ \underline{h}'(\beta) &= \frac{1}{2}\phi(\beta) - \phi\left(-2\beta - \sqrt{2\ln(2)}\right). \end{aligned}$$

Note that

$$\begin{aligned} \underline{h}'(\beta = 0) &= \frac{1}{2}\phi(0) - \phi\left(-\sqrt{2\ln(2)}\right) \\ &= \frac{1}{2}\frac{1}{\sqrt{2\pi}} - \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left\{-\sqrt{2\ln(2)}\right\}^2\right] \\ &= \frac{1}{2}\frac{1}{\sqrt{2\pi}} - \frac{1}{\sqrt{2\pi}}\exp\{-\ln(2)\} = 0. \end{aligned}$$

Checking the second order conditions and the limits yields that $\underline{h}(0)$ is the minimum. Thus, $\underline{h}(0) > \underline{h}(\beta)$ for any $\beta \neq 0$. Therefore, the true value $\underline{h}(0) = \frac{1}{4} + \frac{1}{2}\Phi\left(-\sqrt{2\ln(2)}\right)$ is outside any two-sided AD-confidence interval of $\underline{h}(\beta)$. Thus, the coverage probability is zero in this case so that the coverage probability is also zero for the function in example one, $h(\beta,\gamma) = \frac{1}{2}\Phi(\beta) + \frac{1}{2}\Phi\left(-2\gamma - \sqrt{2\ln(2)}\right)$ if $\rho = 1$. Note that the coverage probability is continuous in ρ so that the coverage probability is also too low for some $\rho < 1$, as shown by the examples in the text.

Example 2:

Using $\mu \geq 0$ yields $P\left(\mu \in \left[\bar{X}_N - \frac{1.96}{\sqrt{N}}, 0\right)\right) = 0$ if $\bar{X}_N - \frac{1.96}{\sqrt{N}} < 0$, so

$$\begin{aligned} &P\left(\mu \in \left[\max\left(0, \bar{X}_N - \frac{1.96}{\sqrt{N}}\right), \bar{X}_N + \frac{1.96}{\sqrt{N}}\right]\right) \\ &= P\left(\mu \in \left[\bar{X}_N - \frac{1.96}{\sqrt{N}}, \bar{X}_N + \frac{1.96}{\sqrt{N}}\right]\right) \\ &= P\left(0 \in \left[\bar{X}_N - \mu - \frac{1.96}{\sqrt{N}}, \bar{X}_N - \mu + \frac{1.96}{\sqrt{N}}\right]\right) \\ &= P\left(0 \in \left[\sqrt{N}(\bar{X}_N - \mu) - 1.96, \sqrt{N}(\bar{X}_N - \mu) + 1.96\right]\right) \\ &= P(|\sqrt{N}(\bar{X}_N - \mu)| \leq 1.96) \\ &= 0.95 \end{aligned}$$

since $\sqrt{N}(\bar{X}_N - \mu)$ has a standard normal distribution. This holds for any $\mu \geq 0$, including $\mu = \frac{c}{\sqrt{N}}$.

Example 3:

Note that the true value of μ is zero.

$$P\left(0 \in \left[\text{sign}(\bar{X}_N)\sqrt{|\bar{X}_N|} - \frac{1.96}{2\sqrt{|\bar{X}_N|}}\frac{1}{\sqrt{N}}, \text{sign}(\bar{X}_N)\sqrt{|\bar{X}_N|} + \frac{1.96}{2\sqrt{|\bar{X}_N|}}\frac{1}{\sqrt{N}}\right]\right)$$
$$= P\left(\text{sign}(\bar{X}_N)\sqrt{|\bar{X}_N|} - \frac{1.96}{2\sqrt{|\bar{X}_N|}}\frac{1}{\sqrt{N}} \leq 0 \leq \text{sign}(\bar{X}_N)\sqrt{|\bar{X}_N|} + \frac{1.96}{2\sqrt{|\bar{X}_N|}}\frac{1}{\sqrt{N}}\right)$$
$$= P\left(-\frac{1.96}{2\sqrt{|\bar{X}_N|}}\frac{1}{\sqrt{N}} \leq -\text{sign}(\bar{X}_N)\sqrt{|\bar{X}_N|} \leq \frac{1.96}{2\sqrt{|\bar{X}_N|}}\frac{1}{\sqrt{N}}\right)$$
$$= P\left(-\frac{1.96}{2}\frac{1}{\sqrt{N}} \leq -\text{sign}(\bar{X}_N)|\bar{X}_N| \leq \frac{1.96}{2}\frac{1}{\sqrt{N}}\right)$$
$$= P\left(-\frac{1.96}{2}\frac{1}{\sqrt{N}} \leq \bar{X}_N \leq \frac{1.96}{2}\frac{1}{\sqrt{N}}\right)$$
$$= \Phi\left(\frac{1.96}{2}\right) - \Phi\left(-\frac{1.96}{2}\right) \approx 0.67$$

Proof of Theorem

The vector-function $h(\theta)$ is uniformly continuous on Θ_r , $r = 1, ..., R$, so that for each $\varepsilon > 0$ there is an $\eta > 0$ such that $||h(\theta_1) - h(\theta_2)|| < \varepsilon$ for all $\theta_1, \theta_2 \in \Theta_r$ with $||\theta_1 - \theta_2|| < \eta$ where $||.||$ is the Euclidean norm. Therefore, we partition the confidence interval $CI^{\theta}_{1-\alpha}$ in Q sets, $CI^{\theta}_{1-\alpha}(1)$, $CI^{\theta}_{1-\alpha}(2)$, ..., $CI^{\theta}_{1-\alpha}(Q)$ such that (i) if $\theta_a \in CI^{\theta}_{1-\alpha}(q)$ and $\theta_b \in CI^{\theta}_{1-\alpha}(q)$ for some q, then $||h(\theta_1) - h(\theta_2)|| < \varepsilon$; and (ii) $CI^{\theta}_{1-\alpha}(1) \cup CI^{\theta}_{1-\alpha}(2) ... \cup CI^{\theta}_{1-\alpha}(Q) = CI^{\theta}_{1-\alpha}$ where $Q < \infty$. Note that such a partition is possible since Θ is compact. Next, suppose that M, the number of samples from $CI^{\theta}_{1-\alpha}$, is sufficiently large that so that we have a realization from each set. Let these samples be denoted by $\tilde{\theta}_1, \tilde{\theta}_2, .., \tilde{\theta}_M$. Note that for every point in $CI^{\theta}_{1-\alpha}$ we have that $\min_{j=1,...,M} ||h(\theta) - h(\tilde{\theta}_j)|| < \varepsilon$. Therefore, as $M \to \infty$, $CI^{h(\theta)}_{1-\alpha} \in \widehat{CI^{h(\theta)}_{1-\alpha}}$ so that the requirement in Assumption 1,

$$\lim_{N\to\infty} \inf_{P\in\mathcal{P}} \Pr(\theta \in CI^{\theta}_{1-\alpha}) \geq 1 - \alpha,$$

yields

$$\lim_{M\to\infty} \lim_{N\to\infty} \inf_{P\in\mathcal{P}} \Pr(h(\theta) \in \widehat{CI^{h(\theta)}_{1-\alpha}}) \geq 1 - \alpha.$$

Q.E.D.

Delta method:

$$h(\hat{\theta}) = h(\theta_0) + H(\bar{\theta})(\hat{\theta} - \theta_0)$$

$$
\begin{aligned}
h(\hat{\theta}) &= h(\theta_0) + H(\bar{\theta})(\hat{\theta} - \theta_0) \\
&= h(\theta_0) + H(\theta_0)(\hat{\theta} - \theta_0) + o(||\hat{\theta} - \theta_0||)
\end{aligned}
$$

7. Appendix algorithm for the Weighted CI-bootstrap

In the first step, we approximate the derivatives. This is only done to make the confidence interval smaller and **so that** numerical derivatives or approximate derivatives (i.e. good guesses) can be used. The resulting confidence interval will still be valid if nonzero, random numbers are used, but it will be wider.

1. Calculate $\left.\frac{\partial h(\theta)}{\partial \theta_k}\right|_{\theta=\hat{\theta}}$ for $k = 1, ..., K$.

2. Sample M times from the distribution of $\hat{\theta}$ (using the bootstrap or just by sampling from the asymptotic distribution of $\hat{\theta}$). Denote these samples by $\{\hat{\theta}^{[1]}, \hat{\theta}^{[2]}, ..., \hat{\theta}^{[M]}\}$.

3. Calculate $W^{[m]} = \sum_{k=1}^{K} (\theta_k^{[m]} - \hat{\theta}_k) \left\{ \left.\left(\frac{\partial h(\theta)}{\partial \theta_k}\right)\right|_{\theta=\hat{\theta}} \right\}$ and $h(\hat{\theta}^{[m]})$ for $m = 1, ..., M$. Make a matrix S and put the realizations of W in the first column and the realizations of h in the second column, i.e. $S_{m1} = W^{[m]}$ and $S_{m2} = h(\hat{\theta}^{[m]})$.

4. Sort the matrix S using column one. This yields S^{sorted}.

Thus, $S_{11}^{\text{sorted}} = max\{\hat{\theta}^{[1]}, \hat{\theta}^{[2]}, ..., \hat{\theta}^{[M]}\}$ and $S_{M1}^{\text{sorted}} = min\{\hat{\theta}^{[1]}, \hat{\theta}^{[2]}, ..., \hat{\theta}^{[M]}\}$.

The 95% confidence interval is then $[S_{2,0.025 \cdot M}^{\text{sorted}}, S_{2,0.975 \cdot M}^{\text{sorted}}]$.

Table 1: 95% Confidence Intervals for the Fraction of Time Spent in Employment for Different Time Horizons and the Effect of Changes in Demographic Variables on these Fractions

	3-year Period	6-year Period	10-year Period
Panel A: Average Expected Fraction of Time in Employment	0.431	0.439	0.449
CI-bootstrap, 95% level	[0.396, 0.469]	[0.401, 0.480]	[0.409, 0.491]
AD-bootstrap, 95% level	[0.414, 0.449]	[0.421, 0.459]	[0.431, 0.470]
Panel B: Changes with respect to:			
Years of Schooling: (s = 12) vs (s < 12)	0.09	0.097	0.1
CI-bootstrap, 95% level	[0.053, 0.125]	[0.058, 0.134]	[0.061, 0.137]
AD-bootstrap, 95% level	[0.072, 0.107]	[0.078, 0.115]	[0.081, 0.119]
Race: African American vs White	-0.031	-0.034	-0.037
CI-bootstrap, 95% level	[-0.073, 0.019]	[-0.080, 0.019]	[-0.083, 0.019]
AD-bootstrap, 95% level	[-0.050, -0.009]	[-0.055, -0.011]	[-0.058, -0.012]
Number of Children less than 6 Years: One Child vs No Children	-0.030	-0.033	-0.035
CI-bootstrap, 95% level	[-0.067, 0.004]	[-0.072, 0.003]	[-0.075, 0.003]
AD-bootstrap, 95% level	[-0.047, -0.014]	[-0.052, -0.016]	[-0.054, -0.017]

Notes:
1. Based on data and parameter estimates from Ham et al (2011).
2. Based on 20,000 draws each.