Calculating Confidence Intervals for Continuous and Discontinuous Functions of Parameters

Tiemen Woutersen and John C. Ham*
University of Arizona, and University of Maryland, IZA, IRP (Madison), and IFAU

February 27, 2013

Abstract. Applied researchers often need to estimate confidence intervals for functions of parameters, such as the effects of counterfactual policy changes. If the function is continuously differentiable and has non-zero and bounded derivatives, then they can use the delta method. However, if the function is nondifferentiable (as in the case of simulating functions with zero-one outcomes), has zero derivatives, or unbounded derivatives, then researchers usually use the nonparametric bootstrap or sample from the asymptotic distribution of the estimated parameter vector. Researchers also use these bootstrap approaches when the function is well-behaved but complicated. Indeed, these approaches are advocated by two very influential published articles. We first show that both of these bootstrap procedures can produce confidence intervals whose asymptotic coverage is less than advertised, i.e. confidence intervals that are too small. We then propose two procedures that provide correct coverage. In applications, we find that the bootstrap approaches mentioned above produce confidence intervals that are significantly smaller than their consistent counterparts, suggesting that previous empirical work is likely to have been overly optimistic in terms of the precision of estimated counterfactual effects.

Keywords: Confidence intervals, simulation, structural models, policy effects.
JEL codes: C12, C15.

*Comments are welcome, woutersen@email.arizona.edu and ham@econ.umd.edu. We thank Don Andrews, James Fisher, Bernd Fitzenberger, Soohyung Lee, Nicole Lott, Serkan Ozbeklik, Geert Ridder, John Rust, Huxin Su, Petra Todd, Mauricio Varela, Jonathan Wright and especially Kei Hirano and Xianghong Li for helpful comments and discussions. We also thank seminar participants at the University of Wisconsin, UC San Diego, and USC.

1. Introduction

We propose procedures to calculate confidence intervals for functions of parameters without restricting the derivatives of the functions and without requiring the functions to be continuous. These are the first procedures for these cases that have consistency proofs. The need for such procedures follows from applied work. Applied researchers often estimate confidence intervals for functions of estimated parameters, e.g. to carry out counterfactual policy analysis. If the function is differentiable and has non-zero and bounded derivatives, they can use the delta method,[1] although researchers are often reluctant to use it for complex, nonlinear functions whose derivatives satisfy these properties.

However, if the function has zero or unbounded derivatives, or is discontinuous, as in the case of simulating functions with zero-one outcomes, then the delta method is inappropriate. Krinsky and Robb (1986) propose the following approach as an alternative to the delta method to obtain a $(1-\alpha)$ confidence interval for a function evaluated at the parameter estimates: i) take a large number of draws from the asymptotic (normal) distribution of the parameter estimates; ii) calculate the function value for each draw; and iii) trim $(\alpha/2)$ from each tail of the resulting distribution of the function values. Their approach has been widely used in empirical work to obtain confidence intervals for complex, nonlinear, differentiable functions of the estimated parameters, such as consumer demand elasticities, the expected duration of unemployment and impulse functions,[2] as well as for nondifferentiable functions of the estimated parameters.[3] Finally, two prominent textbooks[4] also recommend this approach, and the '*wtp*' and '*wtpcikr*' commands in Stata (the leading software package used by applied economists) are based on Krinsky and Robb (1986). Although this procedure of sampling from the asymptotic distribution

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

[2] See Krinsky and Robb (1986) and Fitzenberger, Osikominu and Paul (2010) for applications to demand elasticities and unemployment duration respectively. Further see Inoue and Kilian (2011) for a recent overview of the impulse response function literature. A simple Google search lists forty-four published papers that refer to Krinsky and Robb (1986).

[3] A few (of many possible) examples are Gaure, Røed and Westlie (2010), Ham, Mountain and Chan (1997), Hitsch, Hortacsu and Ariely (2010), Merlo and Wolpin (2009) and Røed and Westlie (2011). Its use is advocated, but not implemented, by Eberwein, Ham and LaLonde (2002). 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.

[4] See Greene (2012, page 610) and Wooldridge (2010, page 441).

is sometimes called the parametric bootstrap, this term has more than one meaning in the literature, so instead we will refer to it as the Asymptotic Distribution bootstrap or AD-bootstrap.

Runkle (1987) recommends the following alternative to the Krinsky-Robb procedure to obtain a $(1-\alpha)$ confidence interval for a function evaluated at the parameter estimates: i) draw a bootstrap sample of the data, reestimate the model, and use the resulting parameter estimates to calculate the function; ii) repeat i) many times and trim $(\alpha/2)$ from each tail of the resulting distribution of function values. Runkle's article also has been very influential; in fact, it was included in the issue that commemorated the twentieth anniversary of the *Journal of Business and Economics Statistics* as one of the ten most influential papers in the history of the journal (Ghysels and Hall 2002, page 1). Moreover, Runkle's approach is endorsed by three prominent graduate econometrics textbooks.[5] We refer to this approach as the ADR-bootstrap. It is first order equivalent to the AD-bootstrap for cases where the version of the bootstrap that is used in Runkle's (1987) procedure estimates the asymptotic distribution of the parameters consistently.

We first give several important examples in which the widely used AD-bootstrap and ADR-bootstrap fail. We then provide a method of obtaining confidence intervals for these functions that is correct under relatively mild conditions that are likely to be satisfied in empirical work. We also provide a modification of our approach that offers potential efficiency gains in principle and in practice; this second method is asymptotically equivalent to the delta method when the latter is valid. Thus, our proposed procedure is valid under weaker conditions than the delta method but involves no efficiency loss. Therefore, our approach should be very useful in all of the cases where researchers have previously used the AD-bootstrap or the ADR-bootstrap, as well as in the case of differentiable functions where it is unclear whether the (generally numerical) derivatives actually are nonzero and bounded.

To implement our first procedure, the researcher obtains a $(1-\alpha)$ confidence interval

[5] Hamilton (1994, page 337), Cameron and Trivedi (2005, page 363) and Wooldridge (2010, page 439); see also Cameron and Trivedi (2010, page 434). Examples of the use of the Runkle (1987) method in applied work are Chaudhuri, Goldberg, and Jia (2006) and Ryan (2012), who use it to obtain confidence intervals for the effects of counterfactual policy changes, and Hoderlein and Mihaleva (2008), who use it to estimate confidence intervals for price elasticities.

for the function of interest by: i) sampling from the asymptotic distribution of the parameter estimator using the bootstrap or using a normal approximation; ii) keeping the draw only if it is in the $(1-\alpha)$ confidence interval for the estimated parameters; iii) calculating the function value for each draw; and iv) using all function values to construct the confidence interval for the function. This procedure differs from the AD-bootstrap and ADR-bootstrap in that they trim the extreme values of the function that come from both 'reasonable' values and 'unreasonable' (extreme) values of the parameter vector, while our approach deletes only function values that arise from 'unreasonable' values of the parameters.[6] We refer to our procedure as the confidence interval bootstrap or CI-bootstrap. We also provide a modification that offers potential efficiency gains over the CI-bootstrap, and refer to it as the weighted confidence interval bootstrap or the WCI-bootstrap. The substantial conditions that are necessary to apply our approach are: i) that one can sample from the asymptotic distribution of the estimators of the parameters and ii) that the set of points at which the function is discontinuous is small. For example, if the function is a scalar, then the second requirement is that the number of discontinuity points is finite.

We also apply our method to an example considered by Andrews (2000). Andrews showed that no version of the bootstrap can consistently estimate the distribution of his maximum likelihood estimator. Nevertheless, our method produces a confidence interval with correct coverage for his estimator.

We use the empirical work from two papers to obtain evidence demonstrating the difference between the procedures in practice. First, we consider work from Ham, Li and Shore-Sheppard (2011, hereafter HLSS), who construct both relatively simple differentiable functions, and relatively complicated nondifferentiable functions, of their parameter estimates describing the labor market dynamics of disadvantaged women in the U.S. Second, we consider confidence intervals for complex differentiable functions of estimated parameters from a rich model of dating and marriage that Lee and Ham (2012, hereafter LH) use to evaluate the efficacy of different matching mechanisms for online dating. We find first for HLSS' simple differentiable functions, the AD-bootstrap produces somewhat smaller confidence intervals than the (appropriate) delta method. Second, we find

[6] We formalize this notion of 'reasonable' values in Lemma 1.

that the AD-bootstrap produces much smaller confidence intervals than those from the (appropriate) CI-bootstrap for LH's complicated differentiable functions and HLSS' non-differentiable functions. Additionally, we find that the WCI-bootstrap offers substantial efficiency gains over the CI-bootstrap in the case of relatively simple differentiable functions. The upshot is that the size of many estimated confidence intervals in the literature may be substantially biased downwards.

We proceed as follows. In section 2 we show that in several important examples, the AD-bootstrap and ADR-bootstrap fail to provide a confidence interval with the correct coverage. In section 3 we show that the CI-bootstrap and the WCI-bootstrap provide consistent confidence intervals for both nondifferentiable and discontinuous functions. In section 4 we provide evidence on the difference between the CI, WCI and AD-bootstraps in practice, and we conclude in section 5.

2. Failures of the Delta Method and the AD-bootstrap when Calculating Confidence Intervals for Functions of Parameters

We examine the performance of the delta method, the AD-bootstrap, and the ADR-bootstrap in three important examples in this section. However, since the AD-bootstrap and ADR-bootstrap are equivalent asymptotically, for expositional ease in what follows we simply refer to the AD-bootstrap with the understanding that the ADR-bootstrap will behave in exactly the same way as the AD-bootstrap. In Example 1, both the delta method and the AD-bootstrap fail. In Example 2, the AD-bootstrap fails. In Example 3, the delta method is infeasible and no version of the bootstrap consistently estimates the asymptotic distribution of the function of the estimator; however, we show below that the CI-bootstrap can be used to construct a valid confidence interval.

Example 1: The Delta Method and the AD-bootstrap Fail

Suppose we observe a random sample, $X_1, ..., X_N$, from a normal distribution with mean μ and variance 1, and let $\hat{\mu} = \bar{X}_N = \frac{1}{N}\sum_i X_i$. Let $E(X) = \mu_0 = 0$ and consider $h(\mu) = \sqrt{|\mu|}$. The delta method yields the following symmetric 95% confidence interval with probability one,

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

The probability that the true value is inside this confidence interval is about 0.67 (in repeated samples) for any N. Fortunately, our method gives a confidence interval with the correct coverage probability of 95%.

In Example 1, the delta method fails because the derivative does not exist at one point. The failure is easy to spot here. However, such problems may be much harder to spot in more realistic applications such as the two empirical applications that we consider below, both of which involve estimating more than one hundred parameters. In example 1, the AD-bootstrap also fails. In particular, it yields a confidence interval with a coverage probability of 0% for any N. The reason for this is that the true value, $h(\mu = 0) = 0$, is the minimum of the function. Let $G(.)$ denote the distribution function of $h(\hat{\mu})$. Applying the AD-bootstrap and using the interval between the 2.5 and 97.5 percentiles of the function values, $[G(0.025), G(0.975)]$, yields a confidence interval that does *not* cover the true value, so the coverage probability is 0%.

Example 2: 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 the true values of the parameters are zero, i.e. $\beta_0 = \gamma_0 = 0$. Let the estimator $(\hat{\beta}, \hat{\gamma})$ 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 if $\rho = 1$. The AD-bootstrap samples from the normal distribution with mean $(\hat{\beta}, \hat{\gamma})$ 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 and using the interval between the 2.5 and 97.5 percentiles of the function values, $[G(0.025), G(0.975)]$, 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 $\rho = 0$.[7] Thus, the AD-bootstrap does not produce a confidence interval with the correct coverage.[8] We also note that the AD-bootstrap confidence interval coincides with the Bayesian credible interval (with flat priors) in this case, so the

[7] See the Appendix for details.

[8] Correct coverage of a confidence interval means that the coverage probability converges to a probability no smaller than its nominal probability, see Andrews and Cheng (2012a and 2012b).

Bayesian procedure also fails here. Note that the extreme failure of the AD-bootstrap for $\rho = 1$ occurs because $h(\beta_0 = 0, \gamma_0 = 0) = 0$ is the minimum value of the function.[9] Just as in example 1, this is the case even if the variance of the estimators would be arbitrarily small. We observe that the continuity of the coverage as a function of the true values of the parameters also causes the AD-bootstrap to fail for ρ close to one. Finally, we note that using the AD-bootstrap to calculate standard errors for average partial effects can also fail in this type of situation.

Example 3: A Problem from Andrews (2000)

Suppose we observe a random sample, $X_1, ..., X_N$, from a normal distribution with mean μ and variance 1 (denoted as $N(\mu, 1)$) and suppose that μ is restricted to be nonnegative. Andrews (2000) considers the maximum likelihood estimator $\hat{\mu} = \max(\bar{X}_N, 0)$ where $\bar{X}_N = \frac{1}{N}\sum_i X_i$. He shows that the nonparametric bootstrap fails to consistently estimate the distribution of $\hat{\mu}$ if $\mu = 0$, and that it is impossible to consistently estimate the distribution of $\hat{\mu}$ using any bootstrap method if $\mu = \frac{c}{\sqrt{N}}$ for any $c > 0$. We show that in this case the CI-bootstrap can consistently estimate the confidence interval for μ.

Of course, there are examples where the AD-bootstrap will have correct asymptotic coverage, but it is difficult to ascertain in general when this will be the case.[10]

3. Main result

In this section we provide a method of obtaining confidence intervals that is valid under reasonable assumptions that are likely to be satisfied in empirical work. We begin by discussing the CI-bootstrap and then consider the WCI-bootstrap. The latter has the advantage that it is asymptotically equivalent to the delta method when the delta method is valid. For both of these approaches, we consider using the asymptotic distribution and the bootstrap approximation.

[9] Thus, sampling from the asymptotic distribution yields values of the function that are larger than the minimum (and true value) with probability one, ensuring failure of the AD-bootstrap.

[10] For example, it is straightforward to show that if θ and $h(\theta)$ are scalars, then a sufficient condition for the AD-boostrap to work is for $h(\theta)$ to be monotonic. However, Example 2 shows that this does not generalize to the case where the parameter is of dimension two and the function is monotonic in its first and second argument.

Let the dimension of θ be equal to K and let $h(\theta)$ have dimension H. Note that allowing for $H > 1$ is important, since it allows one to obtain a joint confidence set for multiple counterfactual outcomes. For example, in a structural model with human capital accumulation (Eckstein and Wolpin 1999, Keane and Wolpin 2000) one can look at the effect of a policy change on both completed schooling and work experience (at any point in the life cycle). Alternatively, in a model of labor market dynamics (Eberwein, Ham and LaLonde 1997), it would be helpful to have a joint confidence set for the effect of participating in a training program on the expected duration of employment and the expected duration of unemployment.

Suppose that the estimator for θ, denoted by $\hat{\theta}$, is asymptotically normally distributed and consider the following confidence interval for the parameter θ.

$$CI^{\theta}_{1-\alpha} = \{\theta \in \Theta | N \cdot (\hat{\theta} - \theta)'(\hat{\Omega})^{-1}(\hat{\theta} - \theta) \leq \chi^2_{1-\alpha}(K)\}, \qquad (1)$$

where $\hat{\Omega}$ is the asymptotic variance-covariance matrix for $\hat{\theta}$ and $\chi^2_{1-\alpha}(K)$ is the $(1-\alpha)$ percentile of the χ^2 distribution with K degrees of freedom. Next, let $CI^{h(\theta)}_{1-\alpha}$ denote the set of values that we obtain if we apply the function $h(\theta)$ to every element of $CI^{\theta}_{1-\alpha}$. More precisely,

$$CI^{h(\theta)}_{1-\alpha} = \{\tau \in \mathbb{R}^H | \tau = h(\theta) \text{ for some } \theta \in CI^{\theta}_{1-\alpha}\}. \qquad (2)$$

Suppose that the researcher draws M^* times from the asymptotic distribution of $\hat{\theta}$. Let $\tilde{\theta}_1, ..., \tilde{\theta}_{M^*}$ denote these draws. The researcher then keeps the draws that satisfy $\tilde{\theta}_m \in CI^{\theta}_{1-\alpha}$, $m = 1, ..., M^*$.. Let $\tilde{\theta}_1, ..., \tilde{\theta}_M$ denote retained these draws. We now estimate the confidence set for the function of the parameters, $CI^{h(\theta)}_{1-\alpha}$, by applying the function $h(.)$ to the draws $\tilde{\theta}_1, ..., \tilde{\theta}_M$. In particular, let $\widehat{CI^{h(\theta)}_{1-\alpha}}$ be the set of all points in the image of $h(\theta)$, $\theta \in \Theta$, that are no farther than the Euclidian distance $\eta > 0$ away from $h(\tilde{\theta}_1), ..., h(\tilde{\theta}_{M-1})$, or $h(\tilde{\theta}_M)$.[11] The confidence set $\widehat{CI^{h(\theta)}_{1-\alpha}}$ is what arises from using what we described heuristically above as the CI-bootstrap.

We briefly note why the AD-bootstrap can fail for the case where $h(\theta)$ is a scalar. The AD-bootstrap samples from the entire asymptotic distribution of $\hat{\theta}$ and forms the

[11] Let Θ_h be the image of $h(\theta)$, $\theta \in \Theta$. If $\tilde{\theta}_s \in CI^{\theta}_{1-\alpha}$ is sampled, then any $h \in \Theta_h$ for which $||h(\tilde{\theta}_s) - h||^2 \leq \eta$ is included in the $(1-\alpha)$ confidence interval for $h(\theta)$.

confidence interval of $h(\theta)$ by trimming the extreme $(1-\alpha)/2$ values 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 from i) an extreme draw from the asymptotic distribution of θ or ii) a 'reasonable' draw for θ that results in an extreme value of $h(\theta)$.[12] 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 does not trim $h(\theta)$ for a 'reasonable' draw of θ. Moreover, note that constructing a confidence interval using the CI-bootstrap is no more difficult than constructing one using the AD-bootstrap.

A similar procedure can be used if the researcher draws J bootstrap samples to obtain the distribution of $\hat{\theta}$. Let $\tilde{\theta}_1, ..., \tilde{\theta}_J$ denote the bootstrap sample estimates and $\hat{\Omega}$ the variance-covariance matrix of the bootstrap samples. For each estimate, we calculate $B_j = (\hat{\theta} - \tilde{\theta}_j)'\hat{\Omega}^{-1}(\hat{\theta} - \tilde{\theta}_j)$, $j = 1, ..., J$. We then select the $(1-\alpha) \cdot J$ bootstrap estimates that have the smallest values of B_j, and call this set B. Let $\tilde{\theta}_1, ..., \tilde{\theta}_M$ denote these draws. As before, we estimate the confidence set by applying the function $h(.)$ to the draws $\tilde{\theta}_1, ..., \tilde{\theta}_M$. That is, $\widehat{CI_{1-\alpha}^{h(\theta)}}$ is the set of all points in the image of $h(\theta)$, $\theta \in \Theta$, that are no farther than the Euclidian distance $\eta > 0$ away from $h(\tilde{\theta}_1), ..., h(\tilde{\theta}_{M-1})$, or $h(\tilde{\theta}_M)$.

We next consider the case where the asymptotic distribution of $\hat{\theta}$ is unknown but a one can construct a confidence interval for it. For example $CI_{1-\alpha}^{\theta}$ is derived using bounds.[13] Note that one cannot calculate $h(\theta)$ for every $\theta \in CI_{1-\alpha}^{\theta}$. Therefore, we use a grid that has M points to approximate $CI_{1-\alpha}^{\theta}$. We then calculate $h(\theta)$ for each of these M grid points.[14] Next, let the confidence set $\widehat{CI_{1-\alpha}^{h(\theta)}}$ be the set of all points in the image of $h(\theta)$, $\theta \in \Theta$, that are no farther than the Euclidian distance $\eta > 0$ away from $h(\tilde{\theta}_1), ..., h(\tilde{\theta}_{M-1})$, or $h(\tilde{\theta}_M)$.

We state our first assumption in terms of properties of $CI_{1-\alpha}^{\theta}$. Later on, when we we discuss the weighted CI-bootstrap, we use properties of the asymptotic distribution as primitives in the assumptions, since the weighting may depend on this asymptotic

[12] In the lemma that follows, we formalize the notion that values that are closer to $\hat{\theta}$ are likely to be closer to the true value θ_0 as well (compared to values that are further away from $\hat{\theta}$).

[13] E.g. the confidence interval is derived using the techniques proposed by Imbens and Manski (2004), Chernozhukov, Hong, and Tamer (2007), or Stoye (2009).

[14] One can use equally spaced grids, Halton sequences, Halton (1964), or Sobol sequences, Sobol (1967). All these grids are dense in $CI_{1-\alpha}^{\theta}$ as M increases.

distribution. Let N denote the sample size. Also, let P be the data generating process and let $\mathcal{P}$ be a space of probability distributions. Our first assumption requires the true value of the parameter, $\theta_0(P)$, to be an element of $CI^{\theta}_{1-\alpha}$ with probability of at least $(1-\alpha)$, uniformly over $\mathcal{P}$.

Assumption 1

Let (i) $\theta_0(P) \in \Theta$, *which is compact; and (ii)*

$$\lim_{N\to\infty} \inf_{P\in\mathcal{P}} \Pr\{\theta_0(P) \in CI^{\theta}_{1-\alpha}\} \geq 1-\alpha$$

where $\alpha \in (0,1)$.

While the true parameter, $\theta_0(P)$, is of course a function of the data generating process, for expositional ease we often write it as θ_0. Note that Assumption 1 simply says that the confidence set for the parameter contains the true parameter value with probability $(1-\alpha)$ in the limit, uniformly over $\mathcal{P}$. This will certainly hold for any estimator that is uniformly asymptotically normally distributed, as well as for the subsampling and bootstrap confidence intervals for θ under appropriate regularity conditions (see Romano and Shaikh 2010, and Andrews and Guggenberger 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* $R < \infty$; *let* $\Theta_1, \Theta_2, ..., \Theta_{R-1}$ *and* Θ_R *have nonzero Lebesgue measure; and let* $h(\theta)$ *be uniformly continuous*[15] *for all* $\theta \in \Theta_r$, $r = 1, ..., R$.

The second assumption allows $h(\theta)$ to be discontinuous. For example, if θ is a scalar, then Assumption 2 requires that the number of discontinuities is finite. In general, the parameter space is partitioned into R subsets, and $h(\theta)$ is assumed to be uniformly continuous on each of these sets. The restriction is that R is finite. This condition is weaker than the conditions needed for the delta method.

Next, we propose a modified version of our procedure. This modified procedure uses weights and usually yields a smaller confidence interval than the CI-bootstrap. The idea

[15] The vector-function $h(\theta)$ is uniformly continuous on Θ_j if for any $\eta > 0$ there is an $\varepsilon > 0$ such that $||h(\theta_1) - h(\theta_2)|| < \eta$ for all θ_1, $\theta_2 \in \Theta_j$ with $||\theta_1 - \theta_2|| < \varepsilon$ where $||.||$ is the Euclidean norm.

is to use a weighted average of the elements of the parameter vector θ. These weights are comparable to the weights that are used in the general method of moments (GMM) procedure, in the sense that the reason to use them is to reduce variation or spread. For example, consider the function $h(\theta) = \Phi(\theta_1 + 2\theta_2)$; then the researcher could use a confidence interval for $\theta_1 + 2\theta_2$ rather than the confidence interval for (θ_1, θ_2). That is, the researcher could use a confidence interval for a weighted average. In general, let $\hat{\theta}$ be asymptotically normally distributed and let $\hat{\Omega}$ denote a consistent estimator for its asymptotic variance-covariance matrix. Define the vector $w = (w_1, w_2, ..., w_K)'$, where $w_1, w_2, ..., w_K$ are scalars if $h(\theta)$ is a scalar and column vectors with length H otherwise. Consider the following confidence interval for $w'\theta$,

$$WCI^{\theta}_{1-\alpha} = \{\theta \in \Theta | N \cdot (\hat{\theta} - \theta)' w (w'\hat{\Omega}w)^{-1} w'(\hat{\theta} - \theta) \leq \chi^2_{1-\alpha}(H)\}. \tag{3}$$

If $h(\theta)$ is a scalar, as in the applications reviewed in the introduction, then $H = 1$. Let $WCI^{h(\theta)}_{1-\alpha}$ denote the set of values that we obtain if we apply the function $h(\theta)$ to every element of $WCI^{\theta}_{1-\alpha}$. That is,

$$WCI^{h(\theta)}_{1-\alpha} = \{\tau \in \mathbb{R}^H | \tau = h(\theta) \text{ for some } \theta \in WCI^{\theta}_{1-\alpha}\}. \tag{4}$$

We estimate $WCI^{h(\theta)}_{1-\alpha}$ by drawing M^* times from the asymptotic distribution of $\hat{\theta}$, and keeping the draws that are elements of $WCI^{\theta}_{1-\alpha}$ (i.e. draws that satisfy. $N \cdot (\hat{\theta} - \theta)'w(w'\hat{\Omega}w)^{-1}w'(\hat{\theta} - \theta) \leq \chi^2_{1-\alpha}(H)$). Let $\tilde{\theta}_1, ..., \tilde{\theta}_M$ denote these retained draws. We now estimate this confidence set by applying the function $h(.)$ to the draws $\tilde{\theta}_1, ..., \tilde{\theta}_M$. In particular, let $\widehat{WCI}^{h(\theta)}_{1-\alpha}$ be the set of all points in the image of $h(\theta)$, $\theta \in \Theta$, that are no farther than the Euclidian distance $\eta > 0$ away from $h(\tilde{\theta}_1), ..., h(\tilde{\theta}_{M-1})$, or $h(\tilde{\theta}_M)$.

A similar procedure can be used if the researcher uses the bootstrap for the distribution of $\hat{\theta}$. Again, let $\tilde{\theta}_1, ..., \tilde{\theta}_J$ denote the bootstrap sample estimates and $\hat{\Omega}$ the variance-covariance matrix of the bootstrap samples. For each estimate, we calculate $\tilde{B}_j = (\hat{\theta} - \tilde{\theta}_j)'w(w'\hat{\Omega}w)^{-1}w'(\hat{\theta} - \tilde{\theta}_j)$, $j = 1, ..., J$. We then select the $(1 - \alpha) \cdot J$ bootstrap estimates that have the smallest values of $\tilde{B}_j$ and call this set $\tilde{B}$. Let $\tilde{\theta}_1, ..., \tilde{\theta}_M$ denote these draws. As before, we estimate the confidence set by applying the function $h(.)$ to the draws $\tilde{\theta}_1, ..., \tilde{\theta}_M$. That is, $\widehat{WCI}^{h(\theta)}_{1-\alpha}$ is the set of all points in the image of $h(\theta)$, $\theta \in \Theta$, that are no farther than the Euclidian distance $\eta > 0$ away from $h(\tilde{\theta}_1), ..., h(\tilde{\theta}_{M-1})$, or $h(\tilde{\theta}_M)$.

In applications, the weights w will often be estimated. One may estimate w by using numerical derivatives of $h(\theta)$ around the estimate $\hat{\theta}$. The numerical derivatives provide simple estimates for the weights, $\hat{w}$, and then one replaces w by $\hat{w}$ in forming $\widehat{WCI^{\theta}_{1-\alpha}}$ and $\tilde{B}$ to obtain confidence intervals for $h(\theta)$. Furthermore, we suggest to limit the ratio of the weights so that $\min_k(|\hat{w}_k|\})/\max_k(|\hat{w}_k|) \geq 1/100$. The WCI-bootstrap yields confidence intervals with the correct coverage for $h(\theta)$, even if some of the partial derivatives of $h(\theta)$ are infinite (as in Example 1) or zero, while of course this is not true for the delta method. Since the WCI-bootstrap is asymptotically equivalent to the delta method when the latter is valid (see the Appendix), the WCI-bootstrap is safer to use than the delta method but involves no loss of efficiency.

A somewhat more complicated procedure that avoids numerical differentiation is the following. First, consider the case where the researcher samples from the asymptotic distribution. In that case, we propose to let the initial confidence set be all values of $\theta \in \Theta$ for which $N \cdot (\hat{\theta} - \theta)'\hat{\Omega}^{-1}(\hat{\theta} - \theta) \leq \chi^2_{1-\alpha}(K)$. We then use the values of this initial confidence interval to estimate a linear approximation to the function $h(\theta)$. In particular, we use the asymptotic distribution to draw M values of the parameter that satisfy $N \cdot (\hat{\theta} - \theta)'\hat{\Omega}^{-1}(\hat{\theta} - \theta) \leq \chi^2_{1-\alpha}(K)$. Let $\tilde{\theta}_1, ..., \tilde{\theta}_M$ denote these points. We then calculate $h(\tilde{\theta}_1), ..., h(\tilde{\theta}_M)$ and regress $\{h(\tilde{\theta}_1), ..., h(\tilde{\theta}_M)\}$ on $\{\tilde{\theta}_1, ..., \tilde{\theta}_M\}$. Let $\hat{w}$ denote the least squares estimator and use the elements of $\hat{w}$ as weights. Note that $\hat{w}'\theta$ is just the best linear predictor and that again $h(\theta)$ is not required to be continuous. Next, construct a confidence set for $h(\theta)$ by again replacing w with $\hat{w}$.

A similar procedure can be used to estimate weights if the researcher uses the bootstrap. Once again, let $\tilde{\theta}_1, ..., \tilde{\theta}_J$ denote the bootstrap sample estimates and $\hat{\Omega}$ the variance-covariance matrix of the bootstrap samples. As in the case of the CI-bootstrap, we calculate $\tilde{A}_j = (\hat{\theta} - \tilde{\theta}_j)'\hat{\Omega}^{-1}(\hat{\theta} - \tilde{\theta}_j)$, $j = 1, ..., J$. We then select the $(1 - \alpha) \cdot J$ bootstrap estimates that have the smallest values of $\tilde{A}_j$ and call this set $\tilde{A}$. Next, we regress $h(\tilde{\theta}_j)$ on $\tilde{\theta}_j$ using all $j \in \tilde{A}$. This yields the weights $\hat{w}$. Next, construct a confidence set for $h(\theta)$ by again replacing w with $\hat{w}$.

Besides Assumption 2, we also need Assumption 3 for the WCI-bootstrap when we construct $\widehat{WCI_{1-\alpha}^{h(\theta)}}$ (i.e. sample from the asymptotic distribution of $\hat{\theta}$).

Assumption 3

Let (i) $\theta \in \Theta$, which is compact; (ii) for all k, $w_k \neq 0$, $\hat{w}_k \neq 0$, $\underset{P \in \mathcal{P}}{sup} |\hat{w}_k - w_k| = o_p(1)$; (iii)$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega)$ uniformly in $P \in \mathcal{P}$, and the estimator $\hat{\Omega}$ converges to Ω uniformly in $P \in \mathcal{P}$, where Ω has full rank; (iv) $\alpha \in (0, 1)$.

If the researcher uses the bootstrap to obtain the confidence set for θ, then we need an additional assumption for $\widehat{WCI_{1-\alpha}^{h(\theta)}}$. In particular, we require that the weighted average, $\hat{w}'\theta_0$, is in the confidence set $WCI_{1-\alpha}^{\theta}$ with a probability that is equal or larger than $(1-\alpha)$, uniformly in $P \in \mathcal{P}$. Romano and Shaikh (2010) give uniform convergence results for the bootstrap (and subsampling).

Assumption 4

If a version of the bootstrap is used, then

$$\lim_{N\to\infty} \inf_{P\in\mathcal{P}} \Pr(\hat{w}'\theta_0 \in WCI_{1-\alpha}^{\theta}) \geq 1 - \alpha.$$

Before stating our theorem, intuition for our result can be obtained by continuing our consideration of Example 3.

Example 3 (Continuation):

As noted above, Andrews (2000) considers $\hat{\mu} = \max(\bar{X}_N, 0)$ where $\bar{X}_N \sim N(\mu, \frac{1}{N})$ and $\mu \geq 0$. He first shows that the nonparametric bootstrap fails to consistently estimate the distribution of $\hat{\mu}$ if $\mu = 0$. Further, he demonstrates 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 any $c > 0$. However, the CI-bootstrap can be used to calculate a 95% confidence interval (or a $(1-\alpha)$ confidence interval) for $\hat{\mu}$ in spite of the absence of a consistent estimator of the asymptotic distribution. In particular, let $\gamma = E(X)$, $\hat{\gamma} = \bar{X}_N$, and $\mu = h(\gamma) = \max(\gamma, 0)$. The symmetric 95% confidence interval for is for μ 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 any $c > 0$.[16]

[16] This example also illustrates that one should perhaps not focus exclusively on the distribution of the bootstrap when the goal is to derive a confidence interval. Also, Hirano and Porter (2012) derive more impossibility results.

We now state our theorem.

Theorem

Let Assumptions 1-2 hold. Then the CI-bootstrap yields

$$\lim_{M\to\infty}\lim_{N\to\infty}\inf_{P\in\mathcal{P}}\Pr\left(h(\theta_0)\in\widehat{CI_{1-\alpha}^{h(\theta)}}\right)\geq 1-\alpha.$$

Let Assumptions 2-3 hold. Then sampling from the asymptotic distribution and using the WCI-bootstrap yields

$$\lim_{M\to\infty}\lim_{N\to\infty}\inf_{P\in\mathcal{P}}\Pr\left(h(\theta_0)\in\widehat{WCI_{1-\alpha}^{h(\theta)}}\right)\geq 1-\alpha.$$

Let Assumptions 2-4 hold. Then using a bootstrap procedure for $\hat{\theta}$ and the WCI-bootstrap yields

$$\lim_{M\to\infty}\lim_{N\to\infty}\inf_{P\in\mathcal{P}}\Pr\left(h(\theta_0)\in\widehat{WCI_{1-\alpha}^{h(\theta)}}\right)\geq 1-\alpha.$$

Proof: See appendix.

If one relaxes the uniformity requirements[17] in Assumption 1, 3, or 4, then the theorem holds without the uniformity property. Specifically, if we replace Assumption 1 by (i) $\theta_0\in\Theta$, which is compact; and (ii)

$$\lim_{N\to\infty}\Pr(\theta_0\in CI_{1-\alpha}^{\theta})\geq 1-\alpha,$$

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

$$\lim_{M\to\infty}\lim_{N\to\infty}\Pr\left(h(\theta_0)\in\widehat{CI_{1-\alpha}^{h(\theta)}}\right)\geq 1-\alpha.$$

Also, Assumption 3 puts only mild restrictions on the weights. In particular, one could use other estimators for the weights. For example, if the function $h(\theta)$ has a single index, then one also could calculate the weights using semiparametric least squares, as in Ichimura (1993), or one of the single index estimators reviewed by Horowitz (1998). In general, our weighting is analogous to the use of a weighting matrix when applying the method of moments estimator. In particular, using a weighting matrix that does *not* converge to the efficient weighting matrix does not, in general, cause the method of moments estimator

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

to be inconsistent, see Hansen (1982) and 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 using the WCI-bootstrap with nonzero weights rather than the CI-bootstrap. Using nonzero weights is analogous to the approach of Newey and West (1987) and Andrews (1991), who advocate using estimates of the variance-covariance matrix that are positive semi-definite.

The main difference between the CI and WCI-bootstrap on the one hand, and the AD-bootstrap on the other hand, is that the CI and WCI-bootstrap use values of θ that are close to $\hat{\theta}$, while the AD-bootstrap does not have this property. In particular, the AD-bootstrap trims extreme values of $h(\theta)$ rather than extreme values of θ. This explains why the AD-bootstrap yields an inconsistent confidence interval in Example 2. We formalize the notion that values of θ that are closer to $\hat{\theta}$ also are likely to be closer to the true value θ_0 in the following lemma.

Lemma

Let θ, v, and w be scalars. Let $(\hat{\theta} - \theta_0) \sim N(0, \sigma^2)$, $\sigma^2 > 0$, and $v^2 < w^2$. Then

$$P(|\hat{\theta} + v - \theta_0| \leq \varepsilon) > P(|\hat{\theta} + w - \theta_0| \leq \varepsilon) \text{ for any } \varepsilon > 0.$$

Proof: See appendix.

Note that most of the discussion of confidence intervals in the literature is about the coverage probability and about the length of the confidence interval. This lemma and our examples add another consideration to the discussion on confidence intervals in general. We now turn to investigating the differences between the AD-bootstrap and the CI-bootstrap within the context of two empirical studies.

4. A Comparison of the Confidence Intervals Produced by the AD-bootstrap and the CI-bootstrap in Two Empirical Studies

In this section we use the parameter estimates and data from two empirical studies to compare the length of the confidence intervals produced by the different methods discussed above. We first use results and data from Ham, Li and Shore-Sheppard (2011, hereafter HLSS). They 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 hazard functions for these women for i) nonemployment spells in progress at the start of the sample, i.e. left censored nonemployment spells; ii) employment spells in progress at the start of the sample, i.e. left censored employment spells; iii) nonemployment spells that begin after the start of the sample, i.e. fresh nonemployment spells and iv) employment spells that begin after the start of the sample, i.e. fresh employment spells.[18]

HLSS first consider the effect of a change in an independent variable on the expected duration of each type of spell. Since the expected duration is a relatively simple differentiable function of the estimated parameters, they use the delta method to calculate confidence intervals. In Table 1, we compare the confidence intervals produced by the delta method, the AD-bootstrap, the CI-bootstrap and the WCI-bootstrap for these expected durations; only the AD-bootstrap will produce incorrect confidence intervals. The first panel presents (for each type of spell) the confidence intervals for the sample average of the individual expected durations for each spell type. The remaining panels show the analogous confidence intervals (produced by each method for each type of spell) of the effect on the expected durations of i) having more schooling; ii) being African-American versus being white; iii) being Hispanic versus. being white and iv) having a child under 6 years versus. not having a child under 6 years. For ease of viewing, in each panel we also report the ratio of the confidence interval lengths produced by: i) the AD-bootstrap relative to the delta method; ii) the CI-bootstrap relative to the delta method; and iii) the WCI-bootstrap relative to the delta method. From Table 1 we conclude that: i) the inconsistent AD-bootstrap produces somewhat shorter confidence intervals than the delta method; ii) the CI-bootstrap produces substantially larger confidence intervals than the delta method; and iii) the WCI-bootstrap produces, on average, confidence intervals that are somewhat larger than those produced using the delta method but considerably smaller than those produced by the CI-bootstrap.

HLSS also consider the effect of the change in an independent variable on the estimated fraction of time a woman will spend in employment 3 years, 6 years, and 10 years after the

[18] They also estimate the joint distribution of the (correlated) unobserved heterogeneity components in each hazard function.

change, which depends on the parameters from all the hazard functions. This function is nondifferentiable so the delta method is no longer applicable and the CI-bootstrap or WCI-bootstrap should be used for estimating confidence intervals.[19] The first panel of Table 2 shows confidence intervals for the baseline fraction of time spent in employment at 3 years, 6 years, and 10 years after the start of the sample. In the remaining panels, we show the respective confidence intervals for the effects of changes in the demographic variables considered above on the fraction of time employed 3, 6, and 10 years after the change. In each case we also show the ratio of the confidence interval lengths produced by i) the AD-bootstrap relative to the CI-bootstrap and ii) the AD-bootstrap relative to the WCI-bootstrap. Table 2 shows that the CI and WCI-bootstrap confidence intervals are basically identical, while the AD-bootstrap produces substantially smaller confidence intervals than the consistent WCI-bootstrap and CI-bootstrap.

Finally, Lee and Ham (2012, here after LH) use data from an online dating service that proposes (opposite gender) matches to its individual members. The data indicate whether the man and woman agree to the date proposed by the company, and if not, whether the man, the woman, or both turned down the date. The data set also contains information on whether, conditional on a first date, the couple goes on a second date, and, if not, whether the man, the woman, or both turned down the second date. Finally, the data also indicate whether the couple marries. Denote the outcome that individual i of gender j ($j = M, F$) accepts (refuses) date d ($d = 1, 2$) as Y_d^j=1 (Y_d^j=0), and let the outcome where the couple marries (does not marry) be denoted by Y^3=1 (Y^3=0). LH estimate a fairly rich model of marriage and dating, and then simulate their estimated model to measure the relative efficiency of different possible matching algorithms that the dating company could use. Here we focus on the baseline probabilities of acceptance for the algorithm that the company actually uses. These probabilities are complicated differ-

[19] For example, if a woman starts the sample in nonemployment, they calculate her hazard function for month 1 of a left censored nonemployment spell, and draw a uniform random number from [0,1]. Suppose the random number is less than the hazard. Then, she moves to employment and a 1 is registered for this month of her simulated employment history. In the next month, we calculate her hazard for month 1 of a fresh employment spell, and again draw a random number. If this random number is less than the hazard, a 0 is registered for the second month of her simulated employment history as she moves back to unemployment; otherwise a 1 is registered for this month of her employment history as she stays in employment. This simulation is comparable to those used in structural modelling to estimate the effect of counterfactual policy changes.

entiable functions of the estimated parameters so it is sensible to use the CI-bootstrap to calculate confidence intervals the baseline probabilities. In Table 3, we contrast these confidence intervals with those produced by the AD-bootstrap. We find that the confidence intervals produced by the AD-bootstrap are about half of the length of those produced by the CI-bootstrap, but that the CI-bootstrap still produces quite narrow confidence intervals for the baseline probabilities.

Thus our results suggest that previous work is likely to have substantially overstated the precision of their counterfactual policy effects, and that there may well be a significant efficiency gain from moving from the CI-bootstrap to the WCI-bootstrap.

5. Conclusion

Applied researchers often need to estimate confidence intervals for functions of estimated parameters that are nondifferentiable, or have unbounded or zero derivatives. Currently, they use the (nonparametric) bootstrap or sample from the asymptotic distribution of the estimated parameters, since the delta method is not appropriate in these settings. Researchers also frequently use these procedures to obtain confidence intervals for well-behaved, but complicated, functions. Indeed, two heavily cited articles and four prominent graduate econometrics textbooks recommend one or both of these approaches. Further, one of these approaches can be implemented using pre-programmed commands in the widely used Stata software package.

We first show that both of these procedures produce confidence intervals that can be incorrect in the sense that the asymptotic coverage is less than intended, i.e. they produce confidence intervals that are too small. We then propose two procedures that have correct coverage under relatively weak conditions. In particular, our procedures are the first to give confidence intervals for functions of parameters without restricting the derivatives of the functions and without requiring the functions to be continuous. We use data and parameter estimates from two empirical studies to compare our approach to the traditional one, and find that the procedures currently used produce substantially downward biased confidence intervals.

Further, Andrews (2000) gives an example in which all versions of the bootstrap fail to consistently estimate the distribution of the maximum likelihood estimator. Our proposed procedures also work for this example, suggesting that it might be more fruitful to focus on the construction of confidence intervals, rather than on the distributions of various versions of the bootstrap.

Finally, one of our procedures (the WCI-bootstrap) produces asymptotically the same confidence interval as the delta method if the linear approximation holds, so in principle there is no efficiency loss in using the WCI-bootstrap in any application. Moreover, we find that in practice this procedure produces similar confidence intervals to the delta method in a situation where the latter is likely to be used.

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] Andrews, D. W. K. and Guggenberger, P. (2010): "Asymptotic Size and a Problem with Subsampling and the m out of n Bootstrap," *Econometric Theory*, 26: 426-468.

[5] Andrews, D. W. K. and X. Cheng (2012a); "Estimation and Inference with Weak, Semi-strong, and Strong Identification," *Econometrica,* 80: 2153-2211.

[6] Andrews, D. W. K. and X. Cheng (2012b); "Maximum Likelihood Estimation and Uniform Inference with Sporadic Identification Failure," Cowles Foundation Discussion Paper 1824.

[7] Cameron, A. C. and P. K. Trivedi (2005): *Microeconometrics, Methods and Applications*, Cambridge University Press, Cambridge.

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

[9] Chaudhuri, S., P. Goldberg, and P. Jia (2006): "Estimating the Effects of Global Patent Protection in Pharmaceuticals: A Case Study of Quinolones in India," *American Economic Review,* 96: 1477-1514.

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

[11] Eberwein, Z., C. J. Ham and R. LaLonde (1997): "The Impact of Being Offered and Receiving Classroom Training on the Employment Histories of Disadvantaged Women: Evidence From Experimental Data," *Review of Economic Studies*, 64: 655-682.

[12] Eckstein, Z. and K. Wolpin (1999): "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities and Abilities," *Econometrica*, 67: 1295–1339.

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

[14] 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.

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

[16] Greene, W. (2012): *Econometric Analyses*, 7th edition, Prentice Hall, New York.

[17] Ghysels, E. and A. Hall (2002): "Editors' Introduction to Twentieth Anniversary Commemorative Issue of the Journal of Business and Economic Statistics," *Journal of Business and Economic Statistics,* 20: 1-4.

[18] Halton, J. H. (1964), Algorithm 247: "Radical-inverse quasi-random point sequence," *Communications of the Association for Computing Machinery*, 7: 701-702.

[19] Ham, J., D. Mountain and L. Chan (1997): "Time-of-Use Prices and Electricity Demand: Allowing for Selection Bias in Experimental Data," *The Rand Journal of Economics,* 28: 113-141.

[20] 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.

[21] Hamilton, J. D. (1994): *Time Series Analysis,* Princeton University Press, Princeton NJ.

[22] Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica,* 50: 1029-1054.

[23] Hirano, K. and J. Porter (2012): "Impossibility Results for Nondifferentiable Functionals," *Econometrica,* 80: 1769–1790.

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

[25] Hoderlein, S. and S. Mihaleva (2008): "Increasing the Price Variation in a Repeated Cross Section," *Journal of Econometrics,* 147: 316-325.

[26] Horowitz, J. L. (1998): *Semiparametric Methods in Econometrics,* Springer-Verlag, New York.

[27] Ichimura, H. (1993). "Semiparametric Least Squares (SLS) and Weighted SLS Estimation of Single Index Models," *Journal of Econometrics,* 58: 71–120.

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

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

[30] Keane, M. and K. Wolpin (2000): "Eliminating Race Differences in School Attainment and Labor Market Success," *Journal of Labor Economics*,18: 614-652.

[31] Krinsky, I. and A. L. Robb (1986): "On Approximating the Statistical Properties of Elasticities," *The Review of Economics and Statistics,* 68: 715-719.

[32] Lee, Sooyhung and J. Ham (2012). "Dating, Marriage, and Matching Online," Mimeo, University of Maryland.

[33] 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.

[34] 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. North-Holland, Amsterdam.

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

[36] 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.*

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

[38] Ryan, S. (2012): "The Costs of Environmental Regulation in a Concentrated Industry," *Econometrica*, 80: 1019–1061.

[39] Runkle, D .E. (1987): "Vector Autoregressions and Reality," *Journal of Business and Economic Statistics,* 5: 437-442.

[40] Sobol, I. M. (1967): "Distribution of Points in a Cube and Approximate Evaluation of Integrals," *U.S.S.R. Computational Mathematics and Mathematical Physics*, 7: 86–112.

[41] Stoye, J. (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77: 1299-1315.

[42] Weisberg, S. (2005): *Applied Linear Regression*, 3rd edition, Wiley, New York, NY.

[43] Wooldridge, J. M. (2010) *Econometric Analysis of Cross Section and Panel Data, 2nd*, MIT Press, Cambridge MA.

6. Appendix

Example 1:

Note that the true value of μ is zero. Consider

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

Example 2

Consider $\rho = 1$. In that case $\hat{\beta} = \hat{\gamma}$, so that $h(\hat{\beta}, \hat{\gamma}) = h(\hat{\beta}, \hat{\beta})$. Therefore, we can define a new function that has just one scalar as its argument. In particular, define

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

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, $\underline{h}(\beta) > \underline{h}(0)$ 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. Hence, the coverage probability is also zero for the function $h(\beta, \gamma) = \frac{1}{2}\Phi(\beta) + \frac{1}{2}\Phi\left(-2\gamma - \sqrt{2\ln(2)}\right)$ if $\rho = 1$. Note that the coverage probability is continuous in ρ so that the coverage probability is also too low for some $\rho < 1$. In the simulations, based on 100,000 repetitions, the coverage probability was still too low for $\rho = 0.5$.

Example 3:

Note that

$$
\begin{aligned}
&P\left(\mu \in \left[\max\left(0, \bar{X}_N - \frac{1.96}{\sqrt{N}}\right), \max\left(0, \bar{X}_N + \frac{1.96}{\sqrt{N}}\right)\right]\right)\\
&= P\left(\max\left(0, \bar{X}_N - \frac{1.96}{\sqrt{N}}\right) \leq \mu \leq \max\left(0, \bar{X}_N + \frac{1.96}{\sqrt{N}}\right)\right)\\
&\geq P\left(\max\left(0, \bar{X}_N - \frac{1.96}{\sqrt{N}}\right) \leq \mu \leq \bar{X}_N + \frac{1.96}{\sqrt{N}}\right)\\
&= 1 - P\left(\max\left(0, \bar{X}_N - \frac{1.96}{\sqrt{N}}\right) > \mu\right) - P\left(\bar{X}_N + \frac{1.96}{\sqrt{N}} < \mu\right)\\
&= 1 - P\left(\max\left(-\bar{X}_N, -\frac{1.96}{\sqrt{N}}\right) > \mu - \bar{X}_N\right) - P\left(\frac{1.96}{\sqrt{N}} < \mu - \bar{X}_N\right).
\end{aligned}
$$

Notice that $P\left(-\bar{X}_N > \mu - \bar{X}_N\right) = 0$ since $\mu \geq 0$. Thus

$$
\begin{aligned}
&P\left(\mu \in \left[\max\left(0, \bar{X}_N - \frac{1.96}{\sqrt{N}}\right), \max\left(0, \bar{X}_N + \frac{1.96}{\sqrt{N}}\right)\right]\right)\\
&\geq 1 - P\left(-\frac{1.96}{\sqrt{N}} > \mu - \bar{X}_N\right) - P\left(\frac{1.96}{\sqrt{N}} < \mu - \bar{X}_N\right)\\
&= P\left(-1.96 \leq \sqrt{N}(\bar{X}_N - \mu) \leq 1.96\right)\\
&= 0.95
\end{aligned}
$$

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

Proof of Theorem:

Consider a uniformly continuous function $f(\theta)$ and let $f([0,1])$ denote the set of values of $f(\theta)$ where $\theta \in [0,1]$, i.e.

$$f([0,1]) = \{\tau \in \mathbb{R} | \tau = f(\theta) \text{ for some } \theta \in [0,1]\}.$$

Next, consider approximating this function on the interval $\theta \in [0,1]$ by evaluating the function at all the values of the M grid points, $G_M = \{\frac{1}{M}, ..., \frac{M-1}{M}, \frac{M}{M}\}$, and including all values that are no farther than $\eta > 0$ from $f(\frac{1}{M}), ..., f(\frac{M-1}{M})$, or $f(\frac{M}{M})$. The next lemma proves that this approximation contains the set $f([0,1])$.

Lemma A1:

Let $f : \mathbb{R} \to \mathbb{R}$ be uniformly continuous. Let $\theta \in [0,1]$, and $\eta > 0$. Then

$$f([0,1]) \subset \lim_{M\to\infty} \cup_{i=1}^{M} [f(i/M) - \eta, f(i/M) + \eta].$$

Proof: By construction, since η is fixed and $h(\theta)$ is uniformly continuous, there exists an $r > 0$ such that $y \in B_r(i/M)$ implies that $|f(y) - f(i/M)| < \eta$ where B_r denotes a ball with radius r. Thus, $f(B_r(i/M)) \subset [f(i/M) - \eta, [f(i/M) + \eta]$. Next, let $M > \frac{1}{r}$ so that $\cup_{i=1}^{M} B_r(i/M) = [0,1]$. Finally, $f([0,1]) \subset \cup_{i=1}^{M} f(B_r(i/M)) \subset \cup_{i=1}^{M}[f(i/M) - \eta, [f(i/M) + \eta]$.

Note that this lemma can easily be generalized to $\theta \in [0,1]^2$ as well as $\theta \in [0,1]^K$ or $\theta \in \Theta$, which is compact. Using this lemma, we now turn to the assumptions of the theorem.

We first consider the case where Assumptions 1-2 hold and we use the CI-bootstrap. The vector-function $h(\theta)$ is uniformly continuous on Θ_r, $r = 1, ..., R$, so that for any $\eta > 0$ there is an $\varepsilon > 0$ such that $||h(\theta_1) - h(\theta_2)|| < \eta$ for all θ_1, $\theta_2 \in \Theta_r$ with $||\theta_1 - \theta_2|| < \varepsilon$ where $||.||$ is the Euclidean norm. Therefore we can partition the confidence interval $CI^{\theta}_{1-\alpha}$ into 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 $||h(\theta_a)-h(\theta_b)|| < \eta$; 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. Also note that, without loss of generality, $CI^{\theta}_{1-\alpha}(1)$, $CI^{\theta}_{1-\alpha}(2), ..., CI^{\theta}_{1-\alpha}(Q)$ have a nonzero Lebesgue measure. Thus, for any $M \geq M_0$, where $M_0 < \infty$, we have that every set $CI^{\theta}_{1-\alpha}(1)$, $CI^{\theta}_{1-\alpha}(2), ..., CI^{\theta}_{1-\alpha}(Q)$ has one or more of the grid point as its elements since the grid is dense in $CI^{\theta}_{1-\alpha}$. Thus, calculating $h(\theta_1), ..., h(\theta_M)$ and including every point in the image of $h(\theta)$, $\theta \in \Theta$, that are no farther than $\eta > 0$ away from $h(\theta_1), ..., h(\theta_{M-1})$, or $h(\theta_M)$ gives $CI^{h(\theta)}_{1-\alpha} \subset \widehat{CI^{h(\theta)}_{1-\alpha}}$ for any $M \geq M_0$. Note that M_0 does not depend on N. Therefore, the requirement in Assumption 1,

$$\lim_{N\to\infty} \inf_{P\in\mathcal{P}} \Pr(\theta_0 \in CI^{\theta}_{1-\alpha}) \geq 1 - \alpha,$$

yields

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

Next, consider the case where Assumptions 2-3 hold and the researcher uses the WCI-bootstrap and samples from the asymptotic distribution of $\hat{\theta}$. Note that by Assumption 3 $w_k \neq 0$, $\hat{w}_k \neq 0$, $\underset{P \in \mathcal{P}}{sup} |\hat{w}_k - w_k| = o_p(1)$. Also note that

$$WCI^{\theta}_{1-\alpha} = \{\theta \in \Theta | N \cdot (\hat{\theta} - \theta)' \hat{w} (\hat{w}' \hat{\Omega} \hat{w})^{-1} \hat{w}' (\hat{\theta} - \theta) \leq \chi^2_{1-\alpha}(H)\}.$$

Just as we could partition $CI^{\theta}_{1-\alpha}$, we can also partition $WCI^{\theta}_{1-\alpha}$ (since Θ is compact). Thus, we partition the confidence interval $WCI^{\theta}_{1-\alpha}$ in Q sets, $WCI^{\theta}_{1-\alpha}(1)$, $WCI^{\theta}_{1-\alpha}(2), ..., WCI^{\theta}_{1-\alpha}(Q)$ such that (i) if $\theta_a \in WCI^{\theta}_{1-\alpha}(q)$ and $\theta_b \in WCI^{\theta}_{1-\alpha}(q)$ for some q, then $||h(\theta_a) - h(\theta_b)|| < \eta$; and (ii) $WCI^{\theta}_{1-\alpha}(1) \cup WCI^{\theta}_{1-\alpha}(2) ... \cup WCI^{\theta}_{1-\alpha}(Q) = WCI^{\theta}_{1-\alpha}$ where $Q < \infty$. Also note that, without loss of generality, $WCI^{\theta}_{1-\alpha}(1)$, $WCI^{\theta}_{1-\alpha}(2), ..., WCI^{\theta}_{1-\alpha}(Q)$ have a nonzero Lebesgue measure. Next, note that $\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega)$ uniformly in $P \in \mathcal{P}$ so that a value of each of the Q subsets is sampled with probability approaching one as $M \to \infty$. Therefore, calculating $h(\tilde{\theta}_1), ..., h(\tilde{\theta}_M)$ and including every point in the image of $h(\theta)$, $\theta \in \Theta$, that are no farther than $\eta > 0$ away from $h(\tilde{\theta}_1), ..., h(\tilde{\theta}_{M-1})$, or $h(\tilde{\theta}_M)$ gives $WCI^{h(\theta)}_{1-\alpha} \subset \widehat{WCI^{h(\theta)}_{1-\alpha}}$ with probability approaching one as $M \to \infty$. This yields the result for sampling from the asymptotic distribution.

Finally, consider the case where Assumptions 2-4 hold. In this case, one can use any version of the bootstrap as long as Assumption 4 is satisfied, i.e.

$$\lim_{N \to \infty} \inf_{P \in \mathcal{P}} \Pr(\hat{w}' \theta_0 \in WCI^{\theta}_{1-\alpha}) \geq 1 - \alpha.$$

Using the same reasoning as for sampling from the asymptotic distribution concludes the proof of the theorem.

Proof of Lemma:

Note that v is a constant. Thus, if θ, v, and w are scalars, then

$$P(|\hat{\theta} + v - \theta_0| \leq \varepsilon) = P(|Z\sigma + v| \leq \varepsilon),$$

where Z is a realization from a standard normal distribution. Note that this probability remains the same if v is replaced by $(-v)$. Similarly, $P(|\hat{\theta} + w - \theta_0| \leq \varepsilon)$ remains the same if w is replaced by $(-w)$. Thus, without loss of generality, we assume that $0 \leq v < w$. This gives

$$P(|\hat{\theta} + v - \theta_0| \leq \varepsilon) = P(-\varepsilon \leq Z\sigma + v \leq \varepsilon)$$
$$= P(-\frac{\varepsilon + v}{\sigma} \leq Z \leq \frac{\varepsilon - v}{\sigma}) = \int_{(-\varepsilon-v)/\sigma}^{(\varepsilon-v)/\sigma} \phi(z)dz.$$

Similarly,

$$P(|\hat{\theta} + w - \theta_0| \leq \varepsilon) = \int_{(-\varepsilon-w)/\sigma}^{(\varepsilon-w)/\sigma} \phi(z)dz.$$

This gives

$$P(|\hat{\theta} + v - \theta_0| \leq \varepsilon) - P(|\hat{\theta} + w - \theta_0| \leq \varepsilon) = \int_{(-\varepsilon-v)/\sigma}^{(\varepsilon-v)/\sigma} \phi(z)dz - \int_{(-\varepsilon-w)/\sigma}^{(\varepsilon-w)/\sigma} \phi(z)dz$$
$$= \int_{(\varepsilon-w)/\sigma}^{(\varepsilon-v)/\sigma} \phi(z)dz - \int_{(-\varepsilon-w)/\sigma}^{(-\varepsilon-v)/\sigma} \phi(z)dz$$

using $0 \leq v < w$. Note that the last equation holds, even if $\varepsilon - w < -\varepsilon - v$. Thus

$$P(|\hat{\theta} + v - \theta_0| \leq \varepsilon) - P(|\hat{\theta} + w - \theta_0| \leq \varepsilon) = \int_{-w/\sigma}^{-v/\sigma} \phi(z + \varepsilon)dz - \int_{-w/\sigma}^{-v/\sigma} \phi(z - \varepsilon)dz$$
$$= \int_{-w/\sigma}^{-v/\sigma} \{\phi(z + \varepsilon) - \phi(z - \varepsilon)\}dz.$$

Note that $0 \leq v < w$ so that $z \in [-w/\sigma, -v/\sigma)$ is negative. Also note that $\varepsilon > 0$ so that $\phi(z + \varepsilon) - \phi(z - \varepsilon) > 0$ for any $z \in [-w/\sigma, -v/\sigma]$. Therefore,
$P(|\hat{\theta} + v - \theta_0| \leq \varepsilon) - P(|\hat{\theta} + w - \theta_0| \leq \varepsilon) > 0$. This completes the proof.

WCI-bootstrap and the delta method

Here we show that the WCI-bootstrap and the delta method are asymptotically equivalent under the standard assumptions of the delta method. The standard assumptions[20] of the delta method are (i) $\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, \Omega)$, (ii) $\hat{\Omega} = \Omega + o_p(1)$, (iii) $h(\theta)$ is continuously differentiable in a neighborhood of θ_0; let $h_{Der}(\theta)$ denote this derivative and let $h_{Der} = h_{Der}(\theta_0)$. Let all elements of h_{Der} be nonzero. Let

$$CI_{1-\alpha}^{Delta} = \{h \in \mathbb{R}^H | h = h(\theta) \text{ for some } \theta \text{ for which}$$
$$N \cdot \{h(\hat{\theta}) - h\}'\{h_{Der}(\hat{\theta})'\hat{\Omega}h_{Der}(\hat{\theta})\}^{-1}\{h(\hat{\theta}) - h\} \leq \chi^2_{1-\alpha}(H)\}.$$

[20] See, for example, Greene (2012, page 1084).

The coverage of this confidence interval converges to $(1-\alpha)$ under the assumptions stated above. We now show that the WCI-bootstrap yields the same confidence interval asymptotically. Consider the confidence interval for the WCI-bootstrap as $M \rightarrow \infty$,

$$\widehat{WCI}_{1-\alpha}^{h(\theta)} = \{h \in \mathbb{R}^H | h = h(\theta) \text{ for some } \theta \text{ for which } N \cdot (\hat{\theta}-\theta)' \hat{w} (\hat{w}' \hat{\Omega} \hat{w})^{-1} \hat{w}' (\hat{\theta}-\theta) \leq \chi^2_{1-\alpha}(H)\}.$$

First, consider the case that $h(\theta)$ is a linear function of the parameters so that $h(\hat{\theta}) - h(\theta) = w'(\hat{\theta} - \theta)$, and $\hat{w} = h_{Der}(\hat{\theta}) = w$. This gives

$$CI_{1-\alpha}^{Delta} = \{h \in \mathbb{R}^H | h = w'\theta \text{ for some } \theta \text{ for which } N \cdot (\hat{\theta}-\theta)' w (w' \hat{\Omega} w)^{-1} w' (\hat{\theta}-\theta) \leq \chi^2_{1-\alpha}(H)\},$$

which is the same set as $\widehat{WCI}_{1-\alpha}^{h(\theta)}$.

Next, if $h(\theta)$ is continuously differentiable (and not necessarily linear), then $h(\hat{\theta}) - h(\theta) = h_{Der}(\bar{\theta})'(\hat{\theta} - \theta)$ where $\bar{\theta}$ is an intermediate value, $\bar{\theta} \in (\hat{\theta}, \theta)$. Note that $h_{Der}(\hat{\theta})$, $h_{Der}(\bar{\theta})$, and $\hat{w}$ (calculated using numerical differentiation or least squares) all converge in probability to $w = h_{Der} = h_{Der}(\theta_0)$. This gives

$$CI_{1-\alpha}^{Delta} = \{h \in \mathbb{R}^H | h = w'\theta \text{ for some } \theta \text{ for which } N \cdot (\hat{\theta}-\theta)' w (w' \hat{\Omega} w)^{-1} w' (\hat{\theta}-\theta) \leq \chi^2_{1-\alpha}(H) + o_p(1)\},$$

so that the confidence intervals of the delta method and WCI-bootstrap are first order equivalent.

Table 1: 95% Confidence Intervals for the Effects of Changes in the Demographic Variables (Separately) on the Expected Durations of Employment and Non-employment Spells

		Left-censored non-employment spells	Left-censored employment spells	Fresh non-employment spells	Fresh employment spells
Estimated Expected Duration (in months)	Estimate	39.305	42.248	11.821	11.929
	Delta Method	[37.872,40.738]	[41.055,43.441]	[10.811,12.832]	[10.969,12.900]
	AD-bootstrap	[38.009,40.491]	[40.957,43,327]	[10.897,12.884]	[11.031,12.965]
	CI-bootstrap	[36.623,41.566]	[39.956,44.252]	[10.207,14.037]	[10.366,14.163]
	WCI-bootstrap	[37.431,41.089]	[40.571,43.567]	[10.864,12.943]	[10.960,13.037]
Ratio of Lengths of Confidence Intervals	AD/Delta	0.866	0.993	0.983	1.002
	CI/Delta	1.725	1.801	0.528	1.966
	WCI/Delta	1.276	1.256	0.972	1.076
Effect on Expected Duration From Changes With Respect to:					
Age: (age=35) - (age=25)	Estimated Effect	7.471	5.095	-0.070	1.027
	Delta Method	[5.376,9.566]	[3.142,7.048]	[-1.113,0.972]	[0.184,1.870]
	AD-bootstrap	[5.399,9.429]	[3.209,7.057]	[-1.094,0.919]	[0.206,1.825]
	CI-bootstrap	[3.550,10.945]	[1.312,9.066]	[-2.527,2.066]	[-0.723,2.655]
	WCI-bootstrap	[5.167,9.623]	[2.992,7.397]	[-1.315,1.095]	[0.024,2.037]
Ratio of Lengths of Confidence Intervals	*AD/Delta*	0.962	0.985	0.965	0.960
	CI/Delta	1.765	1.985	2.203	2.004
	WCI/Delta	1.063	1.128	1.156	1.194
Schooling: (s = 12) - (s < 12)	Estimated Effect	-5.293	7.013	-1.940	2.970
	Delta Method	[-7.352,-3.234]	[4.775,9.251]	[-3.106,-0.773]	[1.983,3.958]
	AD-bootstrap	[-7.230,-3.237]	[4.795,9.211]	[-3.120,-0.766]	[2.013,3.940]
	CI-bootstrap	[-9.432,-1.204]	[2.372,11.364]	[-4.116,0.313]	[1.099,5.011]
	WCI-bootstrap	[-7.868,-2.615]	[4.462,9.735]	[-3.227,-0.600]	[1.823,4.057]
Ratio of Lengths of Confidence Intervals	AD/Delta	0.970	1.072	1.009	0.976
	CI/Delta	1.998	2.184	1.898	1.981
	WCI/Delta	1.276	1.280	1.126	1.131

Table 1 (Continued)

		Left-censored non-employment spells	Left-censored employment spells	Fresh non-employment spells	Fresh employment spells
Effect on Expected Duration From Changes With Respect to:					
Race: Black - White	Estimated Effect	2.524	-1.074	1.842	-0.440
	Delta Method	[-0.022,5.069]	[-3.390,1.242]	[0.434,3.249]	[-1.595,0.716]
	AD-bootstrap	[0.100,5.265]	[-3.420,1.235]	[0.424,3.150]	[-1.550,0.695]
	CI-bootstrap	[-1.702,7.678]	[-5.662,3.440]	[-0.752,4.612]	[-2.663,1.701]
	WCI-bootstrap	[-0.355,5.674]	[-3.908,1.648]	[0.329,3.331]	[-1.775,1.005]
Ratio of Lengths of Confidence Intervals	AD/Delta	1.015	0.910	0.968	0.971
	CI/Delta	1.842	1.788	1.665	1.888
	WCI/Delta	1.184	1.091	1.066	1.203
Race: Hispanic - White	Estimated Effect	2.708	-2.616	0.435	0.090
	Delta Method	[-0.051,5.467]	[-5.623,0.391]	[-1.076,1.947]	[-1.330,1.511]
	AD-bootstrap	[-0.0422,5.728]	[-5.216,0.093]	[-0.968,1.889]	[-1.234,1.477]
	CI-bootstrap	[-2.921,9.055]	[-7.613,2.476]	[-2.225,3.490]	[-2.702,3.003]
	WCI-bootstrap	[-0.522,6.455]	[-6.314,1.289]	[-1.105,2.103]	[-1.522,1.998]
Ratio of Lengths of Confidence Intervals	AD/Delta	1.046	0.883	0.945	0.954
	CI/Delta	2.170	1.678	1.891	1.887
	WCI/Delta	1.264	1.264	1.061	1.164
Number of children less than 6 years old (one - zero)	Estimated Effect	4.225	-1.965	1.151	0.165
	Delta Method	[2.450,6.000]	[-3.982,0.053]	[0.011,2.290]	[-0.721,1.051]
	AD-bootstrap	[2.335,6.042]	[-3.853,-0.080]	[0.012,2.295]	[-0.715,1.027]
	CI-bootstrap	[0.558,7.912]	[-5.525,1.700]	[-1.205,3.250]	[-1.903,1.653]
	WCI-bootstrap	[1.894,6.451]	[-4.359,0.653]	[-0.158,2.439]	[-0.875,1.164]
Ratio of Lengths of Confidence Intervals	AD/Delta	1.044	0.935	1.002	0.983
	CI/Delta	2.072	1.722	1.955	2.007
	WCI/Delta	1.284	1.242	1.140	1.151

Table 2: 95% Confidence Intervals For the Effect of Changing Demographic Variables on the Expected Fraction of Time Spent in Employment for Different Time Horizons

		3-year Period	6-year Period	10-year Period
Estimated Expected Fraction of Time in Employment	Estimate	0.431	0.439	0.449
	AD-bootstrap	[0.414,0.449]	[0.421,0.459]	[0.431,0.470]
	CI-bootstrap	[0.396,0.469]	[0.401,0.480]	[0.409,0.491]
	WCI-bootstrap	[0.396,0.469]	[0.401,0.480]	[0.409,0.489]
Ratio of Lengths of Confidence Intervals	AD/CI	0.479	0.481	0.476
	AD/WCI	0.479	0.481	0.488
Change on the Expected Fraction of Time Spent in Employment With Respect to:				
Schooling: (s = 12) - (s < 12)	Estimated Effect	0.09	0.097	0.100
	AD-bootstrap	[0.072,0.107]	[0.078,0.115]	[0.081,0.119]
	CI-bootstrap	[0.053,0.125]	[0.058,0.134]	[0.061,0.137]
	WCI-bootstrap	[0.053,0.126]	[0.058,0.136]	[0.061,0.140]
Ratio of Lengths of Confidence Intervals	AD/CI	0.486	0.487	0.500
	AD/WCI	0.479	0.474	0.481
Race: Black - White	Estimated Effect	-0.031	-0.034	-0.037
	AD-bootstrap	[-0.050,-0.009]	[-0.055,-0.011]	[-0.058,-0.012]
	CI-bootstrap	[-0.073,0.019]	[-0.080,0.019]	[-0.083,0.019]
	WCI-bootstrap	[-0.073,0.019]	[-0.079,0.019]	[-0.083,0.019]
Ratio of Lengths of Confidence Intervals	AD/CI	0.446	0.444	0.451
	AD/WCI	0.446	0.449	0.451

Table 2 (Continued)

		3-year Period	6-year Period	10-year Period
Change on the Expected Fraction of Time Spent in Employment With Respect to:				
Race: Hispanic - White	Estimated Effect	-0.026	-0.028	-0.029
	AD-bootstrap	[-0.046,-0.001]	[-0.050,-0.002]	[-0.052,-0.001]
	CI-bootstrap	[-0.068,0.023]	[-0.073,0.024]	[-0.077,0.026]
	WCI-bootstrap	[-0.070,0.025]	[-0.074,0.027]	[-0.077,0.029]
Ratio of Lengths of Confidence Intervals	AD/CI	0.495	0.495	0.495
	AD/WCI	0.474	0.475	0.481
Number of kids less than 6 years old: one - zero	Estimated Effect	-0.030	-0.033	-0.035
	AD-bootstrap	[-0.047,-0.014]	[-0.052,-0.016]	[-0.054,-0.017]
	CI-bootstrap	[-0.067,0.004]	[-0.072,0.003]	[-0.075,0.003]
	WCI-bootstrap	[-0.067,0.004]	[-0.072,0.003]	[-0.075,0.003]
Ratio of Lengths of Confidence Intervals	AD/CI	0.465	0.480	0.474
	AD/WCI	0.465	0.480	0.474

Table 3

A Comparison of the CI and AD Bootstrap 95% Confidence

Intervals for Predictions from the Lee-Ham Baseline Model

Probability of Outcome, Dating Company's Matching Algorithm						Prediction (1)	CI Bootstrap (2)	AD Bootstrap (3)	(2) length / (3) length (4)
Panel A									
Moments	Y_1^M	Y_1^W	Y_2^M	Y_2^W	Y_3				
[Omitted]	0	0	-	-	-	57.425	[56.576,57.841]	[56.793,57.579]	1.609
1	0	1	-	-	-	10.758	[10.422,11.210]	[10.545,10.966]	1.872
2	1	0	-	-	-	16.111	[15.481,16.869]	[15.878,16.488]	2.275
3	1	1	0	0	-	4.354	[4.198,4.672]	[4.301,4.494]	2.456
4	1	1	0	1	-	2.393	[2.304,2.539]	[2.329,2.450]	1.942
5	1	1	1	0	-	3.181	[3.083,3.396]	[3.133,3.300]	1.874
6	1	1	1	1	0	5.524	[5.144,5.970]	[5.295,5.809]	1.607
7	1	1	1	1	1	0.255	[0.231,0.300]	[0.238,0.289]	1.353