

A Fast Resample Method for Parametric and Semiparametric Models *

Tim Armstrong^a Marinho Bertanha^b Han Hong^c

May 31, 2013

Abstract

We propose a fast resample method for two step nonlinear parametric and semiparametric models, which does not require recomputation of the second stage estimator during each resample iteration. The fast resample method directly exploits the score function representations computed on each bootstrap sample, thereby reducing computational time considerably. This method is used to approximate the limit distribution of parametric and semiparametric estimators, possibly simulation based, that admit an asymptotic linear representation. Monte Carlo experiments demonstrate the desirable performance and vast improvement in the numerical speed of the fast bootstrap method.

Key words: Score function, Bootstrap, Subsampling, Nonlinear models.

JEL Classification: C12, C15, C22, C52.

*We thank Xiaohong Chen particularly for discussions, and acknowledge generous support by the National Science Foundation (SES 1024504). Armstrong acknowledges generous support from a fellowship from the endowment in memory of B.F. Haley and E.S. Shaw through the Stanford Institute for Economic Policy Research. We are also grateful to the editor, the associate editor and three anonymous referees for suggestions and comments.

^a The corresponding author is Tim Armstrong, Department of Economics, Yale University, Email tim.b.armstrong@gmail.com.

^{b,c} Department of Economics, Stanford University.

1 Introduction

Resampling methods are attractive because they do not require analytic derivation of the limiting distribution and a consistent estimator for it. However, bootstrap and other resampling methods can be computationally intensive in many nonlinear econometric models where estimation of the parameter requires numerical optimization. See for example Gonçalves and White (2004).

The estimating equation, or score function bootstrap approach, considered by Hu and Zidek (1995), Hu and Kalbfleisch (2000), Davidson and MacKinnon (1999), Andrews (2002) and Kline and Santos (2011), cleverly avoids this computational complexity by exploiting the asymptotic influence function representation of nonlinear estimators. The *fast* resampling procedure proposed in this paper extends this literature to multistage semiparametric estimators, retaining the computational advantage while avoiding the difficulty of analytically characterizing the effect of initial stage estimation on the asymptotic distribution of the final stage parameter. The key idea in our setting is to achieve the latter goal by incorporating resampling of first stage estimators of possibly nonparametric functions in the second stage influence function representation.

In addition to nonlinearity and the difficulty of numerical optimization, many semiparametric estimators in economics also depend on a first step nonparametric estimator of an infinite dimensional function. While an extensive theory is available for demonstrating parametric rate of convergence to a limiting normal distribution (see for example the general results of Chen, Linton, and Van Keilegom (2003), the time series generalization by Chen, Hahn, and Liao (2011), and the earlier contributions by Newey (1994) and Andrews (1994)) practical inference for these models remains difficult.

Several methods have been developed for these models. If the asymptotic variance can be characterized analytically, using the pathwise derivative calculation in Newey (1994) and making use of Chen, Linton, and Van Keilegom (2003) to allow for general nonsmooth moment conditions, a consistent estimator using sample analogues is often available. Another approach, described and validated by Akerberg, Chen, and Hahn (2011) and available when the first stage is estimated using sieves, is to estimate the first and

second stage together, treating the first stage sieve estimate as parametric. In addition, a slightly modified version of the canonical bootstrap has been shown to work in these settings by Chen, Linton, and Van Keilegom (2003), which we will hereafter refer to as CLK.

Each of these approaches has limitations. While the first approach is useful in many settings, there are many cases where the asymptotic distribution is difficult to characterize analytically. Consistent estimation of the asymptotic distribution based on sample analogs is also subject to implementation errors. The second approach, while useful when the first stage is estimated using sieves, is harder to apply when the first stage is estimated using kernels and when analytic asymptotic distribution is difficult. The third approach is typically easy to implement in programming, but can be computationally burdensome because of difficulties with nonlinear optimization. The methods proposed in the present paper provide an useful alternative to the existing approaches, which are particularly useful when the researcher does not use parametric sieve methods in the first stage, and the effect of the first stage nonparametric estimate on the asymptotic variance is difficult to characterize analytically. Examples from the literature include Hotz, Miller, Sanders, and Smith (1994), Srisuma and Linton (2012), Hong and Shum (2010) and Bajari, Hong, and Khwaja (2006). From a computational standpoint, our methods are tailored to the case where the second stage is difficult to compute, but the first stage requires relatively little time for computation: the only quantity that must be computed repeatedly for our method is the first stage nonparametric estimate.

The paper is organized as follows. In Section 2 we first outline our framework in the context of two stage semiparametric estimators in which the first stage is possibly nonparametric and the second stage is parametric. Then we describe how fast resampling works and how it can be used in conducting inference on the estimator. We allow for a broad category of direct estimators and simulation based indirect estimators admitting an asymptotically linear representation. The estimator can be either one stage or multi-stage, and the first stage can be parametric or nonparametric. Section 3 provides the formal results of the consistency of the fast resampling procedure. A brief discussion follows

on how to extend the fast resampling procedure to multi-step estimation problems, and some primitive conditions that imply stochastic equicontinuity when data is not iid. In Section 4 we demonstrate through a Monte Carlo experiment that fast bootstrap achieves desirable performance and improves computational speed when compared with standard bootstrap.

2 Framework and fast resampling methods

Consider an estimator $\hat{\theta}$ of a parameter $\theta_0 \in \Theta \in \mathbb{R}^d$ formed from a sample $\underline{X} = (X_1, \dots, X_n)$ (iid or dependent data). The estimator $\hat{\theta}$ can potentially depend on an initial estimate \hat{h} of a nuisance parameter $h_0 \in \mathcal{H}$ which can be either finite dimensional or infinite dimensional (our procedure can be generalized in a straightforward way to multistage estimators with more than two stages in which each stage depends on an estimate in the previous stage). Often $\hat{\theta}$ is obtained by equating a set of moment conditions of dimension k to approximately zero

$$\hat{g}_n(\theta, \hat{h}) \equiv \frac{1}{n} \sum_{i=1}^n g(X_i, \theta, \hat{h})$$

such that

$$g(\theta, h_0) \equiv Eg(X_i, \theta, h_0) = 0 \quad \text{if and only if} \quad \theta = \theta_0.$$

In the rest of the paper we will focus on this GMM setup. It can be modified with minor changes to the M-estimator framework. When the number of moment conditions ‘ k ’ in $g(X_i, \theta, h)$ is greater than the number of parameters ‘ d ’, the GMM estimator $\hat{\theta}$ is often defined as the minimizer of a quadratic objective function

$$\hat{\theta} = \arg \min_{\theta \in \Theta} g_n(\theta, \hat{h})' W g_n(\theta, \hat{h}) = \arg \min_{\theta \in \Theta} \|g_n(\theta, \hat{h})\|. \quad (1)$$

where W is a $k \times k$ positive definite weighting matrix.

Under mild conditions, $\hat{\theta}$ usually has the following influence function representation that depends on \hat{h} :

$$\sqrt{n}(\hat{\theta} - \theta_0) = -(\Gamma_1' W \Gamma_1)^{-1} \Gamma_1' W \sqrt{n} g_n(\theta_0, \hat{h}) + o_p(1) \quad (2)$$

Here, $\Gamma_1 = \frac{\partial}{\partial \theta} E g(X, \theta, h_0) \Big|_{\theta=\theta_0}$. Hence asymptotic normality of $\sqrt{n}(\hat{\theta} - \theta_0)$ depends on the validity of the following condition:

$$\sqrt{n}g_n(\theta_0, \hat{h}) \xrightarrow{d} N(0, V_1).$$

Under suitable conditions, one can separate the dependence of $g(\cdot)$ on the initial estimate \hat{h} and obtain an asymptotic linear influence function representation of the following form:

$$\sqrt{n}g_n(\theta_0, \hat{h}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n [g(X_i, \theta_0, h_0) + \psi(X_i)] + o_p(1) \quad (3)$$

for some function $\psi(X_i)$ that represents the impact of replacing h_0 with \hat{h} on the second stage estimator. Under appropriate mixing conditions and conditions on the moments of the sample mean in equation 3, the normalized estimator $\sqrt{n}(\hat{\theta} - \theta_0)$ will converge to a normal distribution with variance matrix $(\Gamma_1' W \Gamma_1)^{-1} \Gamma_1' W V_1 W \Gamma_1 (\Gamma_1' W \Gamma_1)^{-1}$ where

$$V_1 = \lim_{n \rightarrow \infty} \text{Var} \left\{ \frac{1}{n^{1/2}} \sum_{i=1}^n [g(X_i, \theta_0, h_0) + \psi(X_i)] \right\}, \quad (4)$$

which reduces to $V_1 = E [g(X, \theta_0, h_0) + \psi(X)] [g(X, \theta_0, h_0) + \psi(X)]'$ in the iid case.

One approach to inference is to estimate this asymptotic distribution using \hat{W} and estimates $\hat{\Gamma}$ and \hat{V}_1 of Γ and V_1 . Although it is usually possible to obtain consistent estimates of these matrices, in particular $\hat{\Gamma}$ and \hat{W} , estimating V_1 can be burdensome and often times requires substantial and difficult analytic calculations.

One alternative is to bootstrap. Using the original data \underline{X} , we draw repeated bootstrap samples $\underline{X}^* = (X_1^*, \dots, X_n^*)$ and use the distribution of the statistic or a modified version of it over the bootstrap samples to estimate the unknown sampling distribution. The nonparametric multinomial bootstrap is designed for iid data and draws (X_1^*, \dots, X_n^*) as iid samples from the empirical distribution $F_n(x) = \frac{1}{n} \sum_{i=1}^n I[X_i \leq x]$.

For stationary dependent data, the bootstrap sample \underline{X}^* can be drawn using the moving block bootstrap (MBB), defined as follows. Let $X_{n+i} \equiv X_i$ and the t -th block of the data with $b < n$ elements to be $B_{t,b} = \{X_t, \dots, X_{t+b-1}\}$ for $t = 1, \dots, n$. Let $k = \lfloor \frac{n}{b} \rfloor$, and I_1, \dots, I_k be a sequence of iid Uniform $\{1, \dots, n\}$. The MBB sample \underline{X}^* will be all

the X'_l s in $\{B_{I_1,b}, B_{I_2,b}, \dots, B_{I_k,b}\}$. That is, $\underline{X}^* = (X_1^*, \dots, X_l^*)$, where $X_1^* = X_{I_1}; \dots; X_b^* = X_{I_1+b-1}; X_{b+1}^* = X_{I_2}; \dots; X_l^* = X_{I_k+b-1}$ and $l = kb \leq n$. In other words, the k random numbers I_j choose k out of n blocks of size b .

CLK provide conditions under which the distribution of $\hat{\theta}^*$ estimated using the multinomial bootstrap procedure provides a consistent estimate of the asymptotic variance of $\hat{\theta}$ in the iid case, which are generalized to allow for dependent data in Chen, Hahn, and Liao (2011) and in Lemma 4.2 in Chen (2007) under stationary beta-mixing conditions. In other words, after generating \underline{X}^* from \underline{X} as described above, $\hat{\theta}^*$ is obtained by minimizing the expression

$$\left[\sum_{i=1}^n g(X_i^*, \theta, \hat{h}^*) - g(X_i, \hat{\theta}, \hat{h}) \right]' \hat{W} \left[\sum_{i=1}^n g(X_i^*, \theta, \hat{h}^*) - g(X_i, \hat{\theta}, \hat{h}) \right] \quad (5)$$

over θ , where \hat{h}^* is an estimate of h_0 formed using the bootstrapped sample \underline{X}^* . The bootstrap estimate of the distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$ is the distribution of $\sqrt{n}(\hat{\theta}^* - \hat{\theta})$ conditional on the data \underline{X} .

Drawing repeated simulations from the bootstrap distribution requires repeatedly solving a minimization problem. This may be difficult computationally. We propose a bootstrap procedure that avoids this minimization problem, but also does not require computing an estimate of V_1 . Rather than minimizing the expression in (5), we base our bootstrap procedure on the influence function representation in (2).

Using either multinomial or MBB bootstrap draws \underline{X}^* , and an estimate \hat{h}^* of h_0 based on these draws, our bootstrap procedure estimates the distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$ using the distribution of

$$\hat{\eta}^* = - \left(\hat{\Gamma}'_1 \hat{W} \hat{\Gamma}_1 \right)^{-1} \hat{\Gamma}'_1 \hat{W} \sqrt{n} \left(g_n^* \left(\hat{\theta}, \hat{h}^* \right) - g_n \left(\hat{\theta}, \hat{h} \right) \right) \quad (6)$$

conditional on \underline{X} , where we have defined

$$g_n^*(\theta, h) = \frac{1}{n} \sum_{i=1}^n g(X_i^*, \theta, h).$$

and $\hat{\Gamma}_1$ and \hat{W} are consistent estimates of Γ_1 and W respectively. Intuitively, (6) is the bootstrap analog of the linear influence function representation of (2). Implementation

wise, the empirical distribution of $\hat{\eta}^*$ across a large number of bootstrap simulations serves as an approximation for the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta_0)$. The number of bootstrap simulations is only limited by the amount of computing power.

Each of the resampling methods described above requires a consistent estimate $\hat{\Gamma}_1$ of Γ_1 . For example, when the moment function is differentiable with respect to θ , one can use

$$\hat{\Gamma}_1 = \frac{\partial}{\partial \theta} \hat{g}_n(\hat{\theta}, \hat{h}).$$

When $\hat{g}_n(\hat{\theta}, \hat{h})$ is not differentiable in θ , numerical derivative can be used to compute $\hat{\Gamma}_1$ and is consistent under weak conditions on the step size parameters (e.g. Hong, Mahajan, and Nekipelov (2010)).

3 Consistency of the fast resample algorithm

In this section, we show that our modified bootstrap procedure is consistent under essentially the same conditions used to show consistency of the procedure based on the centered bootstrap objective function (5) in CLK (the only additional condition is the availability of consistent estimates $\hat{\Gamma}_1$ and \hat{W}). The proof follows from an application of the results in CLK to a certain artificially defined parameter.¹ Since the regularity conditions are the same as in CLK, we refer the reader to this paper for the conditions, and methods for verifying them (in their notation, $M_n^*(\theta, h)$, $M_n(\theta, h)$ and $M(\theta, h)$ play the role of our $g_n^*(\theta, h)$, $g_n(\theta, h)$ and $g(\theta, h)$ respectively). We use ' $\overset{p}{\rightsquigarrow}$ ' to denote conditional weak convergence in probability, that is, conditional on the observed sample \underline{X} , the bootstrapped statistic converges in distribution (see, e.g., Kosorok, 2008, for a formal definition).

Theorem 1. *Suppose that $\hat{\Gamma}_1 \overset{p}{\rightarrow} \Gamma$ and $\hat{W} \overset{p}{\rightarrow} W$, and that the conditions of Theorems 2 and B in CLK hold. Then*

$$\sqrt{n}(\hat{\theta} - \theta_0) \overset{d}{\rightarrow} N(0, \Omega) \quad \text{and} \quad \hat{\eta}^* \overset{p}{\rightsquigarrow} N(0, \Omega)$$

where $\Omega = (\Gamma_1' W \Gamma_1)^{-1} \Gamma_1' W V_1 W \Gamma_1 (\Gamma_1' W \Gamma_1)^{-1}$ and Γ_1 and V_1 are defined previously.

¹We thank an anonymous referee for suggesting this simplified proof.

Proof. The asymptotic normality of $\hat{\theta}$ is a restatement of Theorem 2 of CLK. For the conditional convergence in distribution of the bootstrap estimator $\hat{\eta}^*$, first apply Theorem B of CLK to the moment condition $E[g(X_i, \theta_0, h) - \beta] = 0$ at (β_0, h_0) , where β is taken to be the parameter of interest (and is known to be zero). The estimator $\hat{\beta}$ is simply $g_n(\theta_0, \hat{h})$, and the bootstrap estimate is $\hat{\beta}^* = g_n^*(\theta_0, \hat{h}^*)$, so, by Theorem B of CLK,

$$\sqrt{n} \left[g_n^*(\theta_0, \hat{h}^*) - g_n(\theta_0, \hat{h}) \right] = \sqrt{n}(\hat{\beta}^* - \hat{\beta}) \overset{p}{\rightsquigarrow} N(0, V_1)$$

for V_1 given in (4). By this along with the consistency assumption and bootstrap stochastic equicontinuity assumption (2.5'B) in CLK, it follows θ_0 can be replaced by $\hat{\theta}$ in the display above, giving

$$\sqrt{n} \left[g_n^*(\hat{\theta}, \hat{h}^*) - g_n(\hat{\theta}, \hat{h}) \right] \overset{p}{\rightsquigarrow} N(0, V_1).$$

The result follows by this and consistency of $\hat{\Gamma}_1$ and \hat{W} . □

The first part of Theorem 1 restates Theorem 2 in CLK, and is restated only to show that the distribution approximated by the bootstrap, given in the second half of the display, is in fact the correct one (both are normal with the same Ω). It follows from this that confidence intervals based on the fast bootstrap have the correct coverage asymptotically. In the context of dependent data and moving block bootstrap, Radulović (1996) provides conditions on the block size selection, the beta mixing coefficient, and the envelope function of the moment condition for the stochastic equicontinuity condition and its bootstrap version required in Theorem 1.

CLK also provide examples in which primitive conditions that satisfy their Theorems 2 and B can be verified, including a partial linear endogeneity quantile regression model. The same arguments in their paper can also be used to verify other models, for example the quantile treatment effect model in Firpo (2007). In addition, there are empirical models in which the moment conditions are smoothly differentiable in the underlying parameters but are nevertheless difficult to compute numerically. Examples include semiparametric estimators for dynamic discrete choice models (Rust (1987) and Hotz, Miller, Sanders, and Smith (1994)). Differentiability of value functions in dynamic discrete choice models is shown in the recent work by Norets (2010).

4 Monte Carlo Simulations

4.1 Model

To study the finite sample performance of the fast bootstrap method, we conduct a small monte carlo simulation of a two step model. We use the dynamic discrete choice model of Rust (1987) with the estimation procedure proposed in Hotz and Miller (1993). Dynamic discrete models similar to this have driven large interest in the industrial organization literature; see, for example, Bajari, Benkard, and Levin (2007), Ryan (2012), Aguirregabiria and Mira (2007) and Arcidiacono and Miller (2011). The first stage estimation of this model impacts the asymptotic variance of the second stage in a non-trivial way, which makes the analytical derivation of a variance estimator difficult. It becomes appealing to use resampling methods to draw inference on the parameters of interest which are estimated in the second stage. Moreover, restrictions on the cost function will produce a non-linear moment function, which makes the second stage difficult to optimize. Hence, the fast bootstrap will be computationally less expensive than the conventional bootstrap.

The model is as follows. A manager has to decide at each period t whether to keep or replace a machine of age s . The cost of keeping the machine is $c(s, \theta) + v_0$, where θ is a finite dimensional parameter, and v_0 is a random shock. The replacement cost is $R + v_1$, where $R \geq 0$ is a parameter and v_1 a random shock. Although this is an infinite horizon problem, we limit the state variable s to $\{1, 2, \dots, M\}$ by assuming that the cost of keeping the machine does not change after $M = 30$. The shocks $v_t = (v_{0t}, v_{1t})$ are observed by the agent at time t , and assumed to be iid standard bivariate normal with zero covariance. The manager chooses $a_t \in \{0, 1\}$ in each period t , where 1 means he decides to replace the machine; 0 otherwise. He discounts the future by a factor of $\beta \in (0, 1)$.

The cost function will be a 3th-order polynomial in s , $c(s; \theta) = \sum_{k=0}^3 \theta_k s^k$. The cost of keeping a machine is always positive. We will assume the marginal cost is always increasing, that is, the cost function is convex. For any parameter vector θ , we can

generate such a cost function if we specify $\theta_0 = f_0(\theta_1, \theta_3)$ and $\theta_2 = f_2(\theta_3)$, where

$$f_2(\theta_3) = -\frac{1}{2} \min_{s \in [0, M]} 6\theta_3 s$$

$$f_0(\theta_1, \theta_3) = -\min_{s \in [0, M]} \theta_1 s + f_2(\theta_3) s^2 + \theta_3 s^3$$

We define the value function $V(s)$ and choice specific value functions $V_0(s)$ and V_1 (which does not depend on s):

$$V(s) = \max_{\{a_t\}_t} -\mathbb{E} \left\{ \sum_{t=0}^{\infty} \beta^t [a_t(R + v_{1t}) + (1 - a_t)(c(s_t, \theta) + v_{0t})] \middle| s_0 = s \right\}$$

$$V_0(s) = -c(s, \theta) + \beta V(\min\{s + 1; M\})$$

$$V_1 = -R + \beta V(1)$$

The manager's policy function will be $\sigma(s, v) = \mathbb{I}\{V_1 - v_1 \geq V_0(s) - v_0\}$, and $p_{(s)} = P[a = 1|s] = \Phi([V_1 - V_0(s)]/\sqrt{2})$. Therefore, we can express the policy function in terms of $p = (p_{(1)}, \dots, p_{(30)})$: $\sigma(s, v, p) = \mathbb{I}\{\sqrt{2}\Phi^{-1}(p_{(s)}) \geq v_1 - v_0\}$. Next, we want to write the probability of choice a conditional on state s as a function of p and structural parameters $\lambda = [\theta_1 \ \theta_3 \ R]'$.

Using the expression for the policy function and joint normality of v_t , we can define the expected optimal payoff at a given period given s to be:

$$U(s; p, \lambda) = -\mathbb{E} [\sigma(s, v)(R + v_1) + (1 - \sigma(s, v))(c(s, \theta) + v_0) | s]$$

$$= -[(1 - p_{(s)})s^0 \dots (1 - p_{(s)})s^3 \ p_{(s)}] \gamma(\lambda) + \sqrt{2}\phi(\Phi^{-1}(p_{(s)}))$$

$$\equiv \psi_1(s; p)' \gamma(\lambda) + \psi_2(s; p) \tag{7}$$

where $\gamma(\lambda) = [f_0(\theta_1, \theta_3) \ \theta_1 \ f_2(\theta_3) \ \theta_3 \ R]'$.

Next, we need to consider all future paths for s_{t+k} , $k \geq 1$ given we are at s_t today. The one period transition probability can be written in terms of the parameter p : $P(s_{t+1} = s' | s_t = s; p) = \mathbb{I}\{s' = 1\}p_{(s)} + \mathbb{I}\{s' = \min\{s + 1, M\}\}(1 - p_{(s)})$. This allows us to retrieve $P(s_{t+k} = s' | s_t = s; p)$ for any $k \geq 1$. Using this and equation (7), we can write the choice specific value functions in terms of the parameters p, λ .

$$\begin{aligned}
V_0(s; p, \lambda) &= -c(s, \theta) + \beta \sum_{t=0}^{\infty} \beta^t \sum_{u=1}^M P(s_t = u | s_0 = \min\{s+1; M\}; p) U(u; p, \lambda) \\
&= \psi_3(s; p)' \gamma(\lambda) + \psi_4(s; p) \\
V_1(p, \lambda) &= -R + \beta \sum_{t=0}^{\infty} \beta^t \sum_{u=1}^M P(s_t = u | s_0 = 1; p) U(u; p, \lambda) \\
&= \psi_5(p)' \gamma(\lambda) + \psi_6(p)
\end{aligned}$$

where $\psi_3(s; p)$, $\psi_4(s; p)$, $\psi_5(p)$, $\psi_6(p)$ are implicitly defined through equation (4.1) and $\psi_1(s; p)$ and $\psi_2(s; p)$.

This leads to an expression for the conditional probabilities of machine replacement:

$$P[a_t = 1 | s_t; p, \lambda] = \Phi \left(\frac{(\psi_5(p) - \psi_3(s_t; p))' \gamma(\lambda) + \psi_6(p) - \psi_4(s_t; p)}{\sqrt{2}} \right) \quad (8)$$

4.2 Estimation

The econometrician observes n machines that are independently operated. At one given period t , the sample contains $\{a_i, s_i\}_{i=1}^n$, where we suppress the time subscript t for simplicity. We assume β to be known and are interested in estimating λ using moment condition (8). In the first stage, we compute $\hat{p}(s)$ by averaging out a_i for each value of s .

Let $X_{j,i}(p)$ denote the j -th element of $-\psi_3(s_i; p)$, $j = 1, \dots, 5$. Define $c(p) \equiv X_{1,i}(p) + X_{5,i}(p)$. Given that $c(p)$ and $\psi_5(p)$ do not vary with s_i , we transform the left-hand side of (8) to:

$$\begin{aligned}
&\Phi \left(\frac{1}{\sqrt{2}} \left(\psi_5(p)' \gamma(\lambda) + c(p) \gamma_1(\lambda) + \sum_{j=2}^4 X_{j,i}(p) \gamma_j(\lambda) + X_{5,i}(p) (\gamma_5(\lambda) - \gamma_1(\lambda)) + \psi_6(p) - \psi_4(s_i; p) \right) \right) \\
&= \Phi \left(\tilde{X}_i(p)' \delta(p, \lambda) + Z_i(p) \right)
\end{aligned}$$

where $\tilde{X}_i(p) = [1 \ X_{2,i}(p) \ \dots \ X_{5,i}(p)]' / \sqrt{2}$, $Z_i(p) = (\psi_6(p) - \psi_4(s_i; p)) / \sqrt{2}$, and $\delta(p, \lambda)$ the corresponding parameters².

²This transformation helps circumvent numerical issues when inverting $\tilde{X}' \tilde{X}$.

The moment condition g in the second stage is a function of the observed data a_i, s_i , first stage parameter p , and second stage structural parameters λ .

$$g(a_i, s_i; p, \lambda) = \tilde{X}_i(p) \left[a_i - \Phi \left(\tilde{X}_i(p)' \delta(p, \lambda) + Z_i(p) \right) \right]$$

Using the estimated \hat{p} from the 1st stage, we want to compute $\hat{\lambda} = [\hat{\theta}_1 \ \hat{\theta}_3 \ \hat{R}]$ by minimizing $g_n(\hat{p}, \lambda)' W g_n(\hat{p}, \lambda)$ for some P.D. matrix W , where $g_n(p, \lambda) = (1/n) \sum_{i=1}^n g(a_i, s_i; p, \lambda)$. The fast bootstrap statistic is $\hat{\eta}^* = -(\hat{\Gamma}'_1 W \hat{\Gamma}_1)^{-1} \hat{\Gamma}'_1 W \sqrt{n} \left[g_n^*(\hat{p}^*, \hat{\lambda}) - g_n(\hat{p}, \hat{\lambda}) \right]$, and Γ_1 is estimated by the derivative of g_n wrt λ evaluated at $(\hat{p}, \hat{\lambda})$.

4.3 Monte Carlo

The structural parameters are set to $\theta_1 = -5$, $\theta_3 = 2$ and $R = 1.3$. Fixing $\beta = 0.8$, we numerically solve for the value function $V(s)$ and the vector p . We consider sample sizes of 100, 400, 800 and 1600. We simulate 400 different samples of the largest size by drawing s_i uniformly over $\{1, \dots, M\}$ and using the true p to draw actions a_i .

For each sample size and simulation, we compute \hat{p} , $\hat{\lambda}$ and $\hat{\Gamma}_1$. Then, we bootstrap this sample 400 times, and compute \hat{p}^* and $\hat{\eta}^*$. We also compute $\hat{\lambda}^*$ to compare the conventional bootstrap to the fast one.

4.4 Results

The results from the Monte Carlo simulations are reported in tables 1 to 5 below. Table 1 reports the empirical coverage probabilities across simulations of the confidence intervals for λ_0 at three confidence levels, generated according to the fast bootstrap and the conventional bootstrap. The coverage rate produced by the fast bootstrap is somewhat more conservative than the specified level. On the other hand, the conventional bootstrap has lower coverage rates for some sample sizes.

Tables 2 to 4 report false coverage comparisons for the 90%-confidence interval produced by both conventional and fast-bootstrap. In both cases the coverage rate goes to zero as we move away from the true parameter, although this convergence is slightly slower for the fast bootstrap.

Finally, table 5 shows the median time in minutes required to run one simulation of this Monte Carlo exercise, for each sample size and both resampling methods. We can see that, even for this simple example, the conventional bootstrap method can be quite time consuming. The fast-bootstrap was about 3 to 30 times faster depending on the sample size.

[Table 1 about here.]

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

[Table 5 about here.]

5 Conclusion

We have proposed a fast resampling method. The method directly exploits estimating (score) functions computed on each resampled draw and avoids recomputing the estimators for each of them. While this paper focuses on the fast bootstrap method, an earlier version of the paper also shows the validity of a similar fast subsampling method based on the score function representation. We have not emphasized the fast subsampling method because it is not more computationally or inferentially advantageous than fast bootstrap. Fast resampling is easy to perform, and achieves satisfactory performance while improving considerably numerical speed. These advantages should be of interest for applied researchers using nonlinear and dynamic models to conduct effective inference.

Our analysis also suggests that while analytical or numerical variance formulas, resampling and MCMC can each be used to obtain valid asymptotic inference, using them in combination instead of in isolation can offer more powerful tools for computing standard errors and constructing confidence intervals and test statistics.

The main advantage of the score function based bootstrap method for two step estimators is to provide a computational feasible inference method for the first order asymptotic

distribution. It is likely to be less accurate than a studentized bootstrap which either requires analytic knowledge of the asymptotic variance or is computationally more extensive. In one step parametric models, Andrews (2002) studies the higher order accuracy properties of a studentized version of the bootstrap method. It is beyond the scope of this paper, but a potentially important direction of future effort is to extend the results of Andrews (2002) to two step semiparametric models.

References

- ACKERBERG, D., X. CHEN, AND J. HAHN (2011): “A Practical Asymptotic Variance Estimator for Two-Step Semiparametric Estimators,” Cowles Foundation Discussion Paper No. 1803.
- AGUIRREGABIRIA, V., AND P. MIRA (2007): “Sequential estimation of dynamic discrete games,” *Econometrica*, 75(1), 1–53.
- ANDREWS, D. (1994): “Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity,” *Econometrica*, 62, 43–72.
- (2002): “Higher-Order Improvements of a Computationally Attractive k-Step Bootstrap for Extremum Estimators,” *Econometrica*, 70(1), 119–162.
- ARCIDIACONO, P., AND R. A. MILLER (2011): “Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity,” *Econometrica*, 79(6), 1823–1867.
- BAJARI, P., C. L. BENKARD, AND J. LEVIN (2007): “Estimating dynamic models of imperfect competition,” *Econometrica*, 75(5), 1331–1370.
- BAJARI, P., H. HONG, AND A. KHWAJA (2006): “Moral hazard, adverse selection and health expenditures: A semiparametric analysis,” Discussion paper, National Bureau of Economic Research.

- CHEN, X. (2007): “Large Sample Sieve Estimation of Semi-nonparametric Models,” in *Handbook of Econometrics, Vol 6*. Elsevier Science.
- CHEN, X., J. HAHN, AND Z. LIAO (2011): “Semiparametric Two-Step GMM with Weakly Dependent Data,” Working paper, Yale University and UCLA.
- CHEN, X., O. LINTON, AND I. VAN KEILEGOM (2003): “Estimation of Semiparametric Models when the Criterion Function Is Not Smooth,” *Econometrica*, 71, 1591–1608.
- DAVIDSON, R., AND J. MACKINNON (1999): “Bootstrap Testing in Nonlinear Models,” *International Economic Review*, 40, 487–508.
- FIRPO, S. (2007): “Efficient semiparametric estimation of quantile treatment effects,” *Econometrica*, 75(1), 259–276.
- GONÇALVES, S., AND H. WHITE (2004): “Maximum Likelihood and the Bootstrap for Nonlinear Dynamic Models,” *Journal of Econometrics*, 119, 199–219.
- HONG, H., A. MAHAJAN, AND D. NEKIPELOV (2010): “Extremum Estimation and Numerical Derivatives,” working paper, UC Berkeley and Stanford University.
- HONG, H., AND M. SHUM (2010): “Pairwise-difference estimation of a dynamic optimization model,” *The Review of Economic Studies*, 77(1), 273–304.
- HOTZ, V. J., AND R. A. MILLER (1993): “Conditional choice probabilities and the estimation of dynamic models,” *The Review of Economic Studies*, 60(3), 497–529.
- HOTZ, V. J., R. A. MILLER, S. SANDERS, AND J. SMITH (1994): “A simulation estimator for dynamic models of discrete choice,” *The Review of Economic Studies*, 61(2), 265–289.
- HU, F., AND J. D. KALBFLEISCH (2000): “The estimating function bootstrap,” *Canadian Journal of Statistics*, 28(3), 449–481.
- HU, F., AND J. V. ZIDEK (1995): “A bootstrap based on the estimating equations of the linear model,” *Biometrika*, 82(2), 263–275.

- KLINE, P., AND A. SANTOS (2011): “A score based approach to wild bootstrap inference,” *The Journal of Econometric Methods*, 1.
- KOSOROK, M. (2008): *Introduction to Empirical Processes and Semiparametric Inference*. Springer Verlag.
- NEWBY, W. (1994): “The Asymptotic Variance of Semiparametric Estimators,” *Econometrica*, 62, 1349–82.
- NORETS, A. (2010): “Continuity and differentiability of expected value functions in dynamic discrete choice models,” *Quantitative economics*, 1(2), 305–322.
- RADULOVIĆ, D. (1996): “The Bootstrap for Empirical Processes Based on Stationary Observations,” *Stochastic Processes and their Applications*, 65(2), 259–279.
- RUST, J. (1987): “Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher,” *Econometrica: Journal of the Econometric Society*, pp. 999–1033.
- RYAN, S. P. (2012): “The costs of environmental regulation in a concentrated industry,” *Econometrica*, 80(3), 1019–1061.
- SRISUMA, S., AND O. LINTON (2012): “Semiparametric estimation of Markov decision processes with continuous state space,” *Journal of Econometrics*, 166(2), 320–341.

Table 1: Coverage of Confidence Intervals

	Confidence Level	Sample Size			
		100	400	800	1600
<i>a. Fast Bootstrap</i>					
θ_1	90%	0.9875	0.9725	0.9600	0.9475
	95%	0.9950	1.0000	0.9875	0.9800
	99%	0.9975	1.0000	1.0000	0.9975
θ_3	90%	0.9550	0.9800	0.9675	0.9375
	95%	0.9775	0.9925	0.9850	0.9750
	99%	0.9850	1.0000	0.9975	0.9950
R	90%	0.9600	0.9700	0.9375	0.9325
	95%	0.9750	0.9750	0.9600	0.9625
	99%	0.9975	0.9900	0.9900	0.9825
<i>b. Conventional Bootstrap</i>					
θ_1	90%	0.8875	0.9625	0.9375	0.9050
	95%	0.9225	1.0000	0.9750	0.9475
	99%	0.9875	1.0000	0.9975	0.9925
θ_3	90%	0.8025	0.9500	0.9300	0.8950
	95%	0.8050	0.9850	0.9675	0.9475
	99%	0.8075	1.0000	0.9925	0.9850
R	90%	0.9200	0.9775	0.9425	0.9175
	95%	0.9375	0.9850	0.9700	0.9625
	99%	0.9550	0.9975	0.9925	0.9850

Table 2: False Coverage for θ_1

Parameter Value	Sample Size			
	100	400	800	1600
<i>a. Fast Bootstrap</i>				
-8.0	0.9325	0.2325	0.0050	0.0000
-7.0	0.9500	0.5675	0.0975	0.0075
-6.0	0.9800	0.8825	0.7950	0.2650
-4.0	0.9950	0.9175	0.6075	0.1625
-3.0	1.0000	0.4225	0.0375	0.0000
-2.0	0.9900	0.1275	0.0025	0.0000
<i>b. Conventional Bootstrap</i>				
-8.0	0.8175	0.0750	0.0000	0.0000
-7.0	0.8475	0.3575	0.0500	0.0025
-6.0	0.8650	0.8275	0.5425	0.1550
-4.0	0.9000	0.8350	0.4275	0.1425
-3.0	0.8975	0.3225	0.0275	0.0000
-2.0	0.8225	0.0575	0.0025	0.0000

Probability that 90%-confidence interval built in table 1 contains alternative parameter values. True $\theta_1 = -5.0$.

Table 3: False Coverage for θ_3

Parameter Value	Sample Size			
	100	400	800	1600
<i>a. Fast Bootstrap</i>				
0.0	0.4625	0.0000	0.0000	0.0000
1.0	0.9800	0.0400	0.0000	0.0000
1.5	0.9850	0.5875	0.0950	0.0125
2.5	0.9275	0.6575	0.2000	0.0350
3.0	0.8625	0.1550	0.0000	0.0000
4.0	0.6325	0.0150	0.0000	0.0000
<i>b. Conventional Bootstrap</i>				
0.0	0.2425	0.0000	0.0000	0.0000
1.0	0.7425	0.0100	0.0000	0.0000
1.5	0.8625	0.4125	0.0725	0.0025
2.5	0.6950	0.4775	0.0825	0.0125
3.0	0.5525	0.0425	0.0000	0.0000
4.0	0.2300	0.0000	0.0000	0.0000

Probability that 90%-confidence interval built in table 1 contains alternative parameter values. True $\theta_3=2.0$.

Table 4: False Coverage for R

Parameter Value	Sample Size			
	100	400	800	1600
<i>a. Fast Bootstrap</i>				
0.7	0.9100	0.0400	0.0000	0.0000
1.0	0.9900	0.6550	0.0950	0.0025
1.2	0.9825	0.9600	0.9000	0.7275
1.4	0.9425	0.8875	0.8425	0.7150
1.6	0.8600	0.4200	0.1175	0.0125
1.9	0.6900	0.0775	0.0000	0.0000
<i>b. Conventional Bootstrap</i>				
0.7	0.9650	0.0250	0.0000	0.0000
1.0	0.9475	0.6100	0.0675	0.0025
1.2	0.9325	0.9800	0.8600	0.6125
1.4	0.9075	0.8925	0.8275	0.6600
1.6	0.8250	0.4100	0.1075	0.0100
1.9	0.6475	0.0575	0.0000	0.0000

Probability that 90%-confidence interval built in table 1 contains alternative parameter values. True $R=1.3$.

Table 5: Median computation time per simulation in minutes

Bootstrap Method	Sample Size			
	100	400	800	1600
Fast	0.7163	2.7759	5.5835	10.3878
Conventional	21.2070	21.8642	30.8541	37.5082

The simulations were performed in MATLAB(R) using a Unix server running 8 parallel processes with 8-core CPUs