

## POOR (WO)MAN'S BOOTSTRAP

BO E. HONORÉ

Department of Economics, Princeton University

LUOJIA HU

Economic Research Department, Federal Reserve Bank of Chicago

The bootstrap is a convenient tool for calculating standard errors of the parameter estimates of complicated econometric models. Unfortunately, the fact that these models are complicated often makes the bootstrap extremely slow or even practically infeasible. This paper proposes an alternative to the bootstrap that relies only on the estimation of *one-dimensional* parameters. We introduce the idea in the context of M and GMM estimators. A modification of the approach can be used to estimate the variance of two-step estimators.

KEYWORDS: Standard error, bootstrap, inference, structural models, two-step estimation.

### 1. INTRODUCTION

THE BOOTSTRAP is often used for estimating standard errors in applied work. This is true even when an analytical expression exists for a consistent estimator of the asymptotic variance. The bootstrap is convenient from a programming point of view because it relies on the same estimation procedure that delivers the point estimates. Moreover, for estimators that are based on non-smooth objective functions or on discontinuous moment conditions, direct estimation of the matrices that enter the asymptotic variance typically forces the researcher to make choices regarding tuning parameters such as bandwidths or the number of nearest neighbors. The bootstrap avoids this. Likewise, estimation of the asymptotic variance of two-step estimators requires calculation of the derivative of the estimating equation in the second step with respect to the first-step parameters. This calculation can also be avoided by the bootstrap.

Unfortunately, the bootstrap can be computationally burdensome if the estimator is complex. For example, in many structural econometric models, it can take hours to get a single bootstrap draw of the estimator. This is especially problematic because the calculations in [Andrews and Buchinsky \(2001\)](#) suggest that the number of bootstrap replications used in many empirical economics papers is too small for accurate inference. This paper will demonstrate that in many cases it is possible to use the bootstrap distribution of much simpler alternative estimators to back out a bootstrap-like estimator of the asymptotic variance of the estimator of interest. The need for faster alternatives to the standard bootstrap also motivated the papers by, for example, [Davidson and MacKinnon \(1999\)](#),

---

Bo E. Honoré: [honore@Princeton.edu](mailto:honore@Princeton.edu)

Luojia Hu: [luhu@frbchi.org](mailto:luhu@frbchi.org)

This research was supported by the Gregory C. Chow Econometric Research Program at Princeton University. The opinions expressed here are those of the authors and not necessarily those of the Federal Reserve Bank of Chicago or the Federal Reserve System. We are very grateful to the editor and the referees for comments and constructive suggestions. We also thank Joe Altonji, Jan De Loecker, and Aureo de Paula as well as seminar participants at the Federal Reserve Bank of Chicago, University of Aarhus, University of Copenhagen, Central European University, Science Po, Brandeis University, Simon Fraser University, Yale University, and the University of Montreal.

Andrews (2002), Heagerty and Lumley (2000), Hong and Scaillet (2006), Kline and Santos (2012), and Armstrong, Bertanha, and Hong (2014). Unfortunately, their approaches assume that one can easily estimate the “Hessian” in the sandwich form of the asymptotic variance of the estimator. In practice, this can be difficult for estimators defined by optimization of non-smooth objective functions or by discontinuous moment conditions. It can also be cumbersome to derive explicit expressions for the “Hessian” in smooth problems. The main motivation for this paper is the difficulty of obtaining an estimator of the “Hessian.” Part of the contribution of Chernozhukov and Hong (2003) was also to provide an alternative way to do inference without estimating asymptotic variances from their analytical expressions. However, Kormiltsina and Nekipelov (2012) pointed out that the method proposed by Chernozhukov and Hong (2003) can be problematic in practice.

In this paper, we propose a method for estimating the asymptotic variance of a  $k$ -dimensional estimator by a bootstrap method that requires estimation of  $k^2$  one-dimensional parameters in each bootstrap replication. For estimators that are based on non-smooth or discontinuous objective functions, this will lead to substantial reductions in computing times as well as in the probability of locating local extrema of the objective function. The contribution of the paper is the convenience of the approach. We do not claim that any of the superior higher order asymptotic properties of the bootstrap or of the  $k$ -step bootstrap carries over to our proposed approach. However, these properties are not usually the main motivation for the bootstrap in applied economics.

We first introduce our approach in the context of an extremum estimator (Section 2.1). We consider a set of simple infeasible one-dimensional estimators related to the estimator of interest, and we show how their asymptotic covariance matrix can be used to back out the asymptotic variance of the estimator of the parameter of interest. Mimicking Hahn (1996), we show that the bootstrap can be used to estimate the joint asymptotic distribution of those one-dimensional estimators. This suggests a computationally simple method for estimating the variance of the estimator of the parameter vector of interest. We then demonstrate in Section 2.2 that this insight carries over to GMM estimators.

Section 3 shows that an alternative, and even simpler, approach can be applied to method of moments estimators. In Section 4, we discuss why, in general, the number of directional estimators must be of order  $O(k^2)$ , and we discuss how this can be significantly reduced when the estimation problem has a particular structure.

It turns out that our procedure is not necessarily convenient for two-step estimators. In Section 5, we therefore propose a modified version specifically tailored for this scenario. While our method can be used to estimate the full joint asymptotic variance of the estimators in the two steps, we focus on estimation of the correction to the variance of the second-step estimator which is needed to account for the estimation error in the first step. We also discuss how our procedure simplifies when the first-step or the second-step estimator is computationally simple.

We illustrate our approach by Monte Carlo studies in Section 6. The basic ideas introduced in Section 2 are illustrated in a linear regression model estimated by OLS and in a dynamic Roy model estimated by indirect inference. The motivation for the OLS example is that it is well understood and that its simplicity implies that the asymptotics often provide a good approximation in small samples. This allows us to focus on the marginal contribution of this paper rather than on issues about whether the asymptotic approximation is useful in the first place. Of course, the linear regression model does not provide an example of a case in which one would actually need to use our version of the bootstrap. We therefore also consider indirect inference estimation of a structural econometric model (a dynamic Roy model). This provides an example of the kind of model where we think

the approach will be useful in current empirical research. Finally, we illustrate the extensions discussed in Section 5 by applying our approach to a two-step estimator of a sample selection model inspired by [Helpman, Melitz, and Rubinstein \(2008\)](#) (see Section 6.3).

We emphasize that the contribution of this paper is the computational convenience of the approach. We are not advocating the approach in situations in which it is easy to use the bootstrap. That is why we use the term “poor (wo)man’s bootstrap.” We are also not implying that higher order refinements are undesirable when they are practical.

## 2. BASIC IDEA

### 2.1. *M* Estimators

We first consider an extremum estimator of a  $k$ -dimensional parameter  $\theta$  based on a random sample  $\{z_i\}$ ,

$$\hat{\theta} = \arg \min_{\tau} Q_n(\tau) = \arg \min_{\tau} \sum_{i=1}^n q(z_i, \tau).$$

Subject to the usual regularity conditions, this will have asymptotic variance of the form

$$A \text{var}(\hat{\theta}) = H^{-1} V H^{-1},$$

where  $V$  and  $H$  are both symmetric and positive definite. When  $q$  is a smooth function of  $\tau$ ,  $V$  is the variance of the derivative of  $q$  with respect to  $\tau$  and  $H$  is the expected value of the second derivative of  $q$ , but the setup also applies to many non-smooth objective functions such as in [Powell \(1984\)](#).

While it is in principle possible to estimate  $V$  and  $H$  directly, many empirical researchers estimate  $A \text{var}(\hat{\theta})$  by the bootstrap. That is especially true if the model is complicated, but unfortunately, that is also the situation in which the bootstrap can be time-consuming or even infeasible. The point of this paper is to demonstrate that one can use the bootstrap variance of much simpler estimators to estimate  $A \text{var}(\hat{\theta})$ .

The basic idea pursued here is to back out the elements of  $H$  and  $V$  from the covariance matrix of a number of infeasible one-dimensional estimators of the type

$$\hat{a}(\delta) = \arg \min_a Q_n(\theta + \delta a), \tag{1}$$

where  $\delta$  is a fixed  $k$ -dimensional vector.

The (nonparametric) bootstrap equivalent of (1) is

$$\arg \min_a \sum_{i=1}^n q(z_i^b, \hat{\theta} + \delta a), \tag{2}$$

where  $\{z_i^b\}$  is the bootstrap sample. This is a one-dimensional minimization problem, so for complicated objective functions, it will be much easier to solve than the minimization problem that defines  $\hat{\theta}$  and its bootstrap equivalent. Our approach will therefore be to estimate the joint asymptotic variance of  $\hat{a}(\delta)$  for a number of directions,  $\delta$ , and then use that asymptotic variance estimate to back out estimates of  $H$  and  $V$  (except for a scale normalization). In [Appendix A](#), we mimic the arguments in [Hahn \(1996\)](#) and note that the joint bootstrap distribution of the estimators  $\hat{a}(\delta)$  for different directions,  $\delta$ , can

be used to estimate the joint asymptotic distribution of  $\widehat{a}(\delta)$ . Although convergence in distribution does not guarantee convergence of moments, this can be used to estimate the variance of the asymptotic distribution of  $\widehat{a}(\delta)$  (by using robust covariance estimators). Since the mapping (discussed below) from this variance to  $H$  and  $V$  is continuous, this implies the consistency of our proposed method.

It is easiest to illustrate why our approach works by considering a case where  $\theta$  is two-dimensional. We first note that the estimation problem remains unchanged if  $q$  is scaled by a positive constant  $c$ , but in that case  $H$  would be scaled by  $c$  and  $V$  by  $c^2$ . There is therefore no loss of generality in assuming  $v_{11} = 1$ . In other words, the symmetric matrices  $H$  and  $V$  depend on five unknown quantities. Now consider two vectors  $\delta_1$  and  $\delta_2$  and the associated estimators  $\widehat{a}(\delta_1)$  and  $\widehat{a}(\delta_2)$ . Under the conditions that yield asymptotic normality of the original estimator  $\widehat{\theta}$ , the infeasible estimators  $\widehat{a}(\delta_1)$  and  $\widehat{a}(\delta_2)$  will be jointly asymptotically normal with variance

$$\begin{aligned} \Omega_{\delta_1, \delta_2} &= A \operatorname{var} \left( \begin{pmatrix} \widehat{a}(\delta_1) \\ \widehat{a}(\delta_2) \end{pmatrix} \right) \\ &= \begin{pmatrix} (\delta_1' H \delta_1)^{-1} \delta_1' V \delta_1 (\delta_1' H \delta_1)^{-1} & (\delta_1' H \delta_1)^{-1} \delta_1' V \delta_2 (\delta_2' H \delta_2)^{-1} \\ (\delta_1' H \delta_1)^{-1} \delta_1' V \delta_2 (\delta_2' H \delta_2)^{-1} & (\delta_2' H \delta_2)^{-1} \delta_2' V \delta_2 (\delta_2' H \delta_2)^{-1} \end{pmatrix}. \end{aligned} \tag{3}$$

It will be useful to explicitly write the  $(j, \ell)$ th elements of  $H$  and  $V$  as  $h_{j\ell}$  and  $v_{j\ell}$ , respectively. In the following, we use  $e_j$  to denote a vector that has 1 in its  $j$ th element and zeros elsewhere. With  $\delta_1 = e_1$  and  $\delta_2 = e_2$ , we have

$$\Omega_{e_1, e_2} = \begin{pmatrix} h_{11}^{-2} & h_{11}^{-1} v_{12} h_{22}^{-1} \\ h_{11}^{-1} v_{12} h_{22}^{-1} & h_{22}^{-2} v_{22} \end{pmatrix}.$$

The matrix  $\Omega_{e_1, e_2}$  is clearly informative about some of the elements of  $(h_{11}, h_{12}, h_{22}, v_{12}, v_{22})$ , but since it is a symmetric 2-by-2 matrix, it cannot provide enough information to identify all five elements. On the other hand, it turns out that the joint covariance that considers the estimators in two additional directions does identify all five elements. This is a special case of the following theorem:

**THEOREM 1:** *Let  $\delta_1, \delta_2, \delta_3,$  and  $\delta_4$  be nonproportional 2-by-1 vectors, and let  $H$  and  $V$  be symmetric 2-by-2 matrices. Assume that  $H$  is positive definite, and that  $v_{11} = 1$ . Then knowledge of  $(\delta_j' H \delta_j)^{-1} \delta_j' V \delta_\ell (\delta_\ell' H \delta_\ell)^{-1}$  for all combinations of  $\delta_j$  and  $\delta_\ell$  identifies  $H$  and  $V$ .*

**PROOF:** See Appendix B. Q.E.D.

Theorem 1 leaves many degrees of freedom with regard to the choice of directions,  $\delta$ . In order to treat all coordinates symmetrically, we focus on directions of the form  $e_j, e_j + e_\ell,$  and  $e_j - e_\ell$ . We then have the following:

**COROLLARY 2:** *Let  $H$  and  $V$  be symmetric  $k$ -by- $k$  matrices. Assume that  $H$  is positive definite, that  $v_{11} = 1$  and that  $v_{jj} > 0$  for  $j > 1$ . Then knowledge of  $(\delta_j' H \delta_j)^{-1} \delta_j' V \delta_\ell (\delta_\ell' H \delta_\ell)^{-1}$  for all combinations of  $\delta_j$  and  $\delta_\ell$  of the form  $e_j, e_j + e_\ell$  ( $l < j$ ), or  $e_j - e_\ell$  ( $l < j$ ) identifies  $H$  and  $V$ .*

**PROOF:** For each  $j$  and  $\ell$ , Theorem 1 identifies  $\frac{v_{\ell\ell}}{v_{jj}}, \frac{v_{j\ell}}{v_j}$ , and all the elements of  $H$  scaled by  $\sqrt{\frac{v_{\ell\ell}}{v_{jj}}}$ . These can then be linked together by the fact that  $v_{11}$  is normalized to 1. Q.E.D.

One can characterize the information about  $V$  and  $H$  contained in the covariance matrix of the estimators  $(\widehat{a}(\delta_1), \dots, \widehat{a}(\delta_m))$  as a solution to a set of nonlinear equations.

Specifically, define

$$D = (\delta_1 \quad \delta_2 \quad \dots \quad \delta_m) \quad \text{and} \quad C = \begin{pmatrix} \delta_1 & 0 & \dots & 0 \\ 0 & \delta_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \delta_m \end{pmatrix}. \tag{4}$$

The covariance matrix for the  $m$  one-dimensional estimators is then

$$\Omega = (C'(I \otimes H)C)^{-1}(D'VD)(C'(I \otimes H)C)^{-1},$$

which implies that

$$(C'(I \otimes H)C)\Omega(C'(I \otimes H)C) = (D'VD).$$

These need to be solved for the symmetric and positive definite matrices  $V$  and  $H$ . Corollary 2 above shows that this has a unique solution (except for scale) as long as  $D$  contains all vectors of the form  $e_j$ ,  $e_j + e_\ell$ , and  $e_j - e_\ell$ .

In practice, one would first estimate the parameter  $\theta$ . Using  $B$  bootstrap samples,  $\{z_i^b\}_{i=1}^n$ , one would then obtain  $B$  draws of the vectors  $(\widehat{a}(\delta_1), \dots, \widehat{a}(\delta_m))$ . Let  $\widehat{\Omega}$  denote  $n$  times a robust estimate of their variance matrix. There are then many ways to turn the identification strategy above into estimation of  $H$  and  $V$ . One is to pick a set of  $\delta$ -vectors and estimate the covariance matrix of the associated estimators. Denote this estimator by  $\widehat{\Omega}$ . The matrices  $V$  and  $H$  can then be estimated by solving the nonlinear least squares problem

$$\min_{V,H} \sum_{j\ell} \{ (C'(I \otimes H)C)\widehat{\Omega}(C'(I \otimes H)C) - (D'VD) \}_{j\ell}^2, \tag{5}$$

where  $D$  and  $C$  are defined in (4),  $v_{11} = 1$ , and  $V$  and  $H$  are positive definite matrices.

### 2.2. GMM

We now consider variance estimation for GMM estimators. The starting point is a set of moment conditions

$$E[f(z_i, \theta)] = 0,$$

where  $z_i$  is “data for observation  $i$ ” and it is assumed that this defines a unique  $\theta$ . The GMM estimator for  $\theta$  is

$$\widehat{\theta} = \arg \min_{\tau} \left( \frac{1}{n} \sum_{i=1}^n f(z_i, \tau) \right)' W_n \left( \frac{1}{n} \sum_{i=1}^n f(z_i, \tau) \right),$$

where  $W_n$  is a symmetric, positive definite matrix. Subject to weak regularity conditions (see Hansen (1982) or Newey and McFadden (1994)), the asymptotic variance of the GMM estimator has the form

$$\Sigma = (\Gamma'W\Gamma)^{-1}\Gamma'WSW\Gamma(\Gamma'W\Gamma)^{-1}, \tag{6}$$

where  $W$  is the probability limit of  $W_n$ ,  $S = V[f(z_i, \theta)]$ , and  $\Gamma = \frac{\partial}{\partial \theta'} E[f(z_i, \theta)]$ . [Hahn \(1996\)](#) showed that the limiting distribution of the GMM estimator can be estimated by the bootstrap.

Now let  $\delta$  be some fixed vector and consider the problem of estimating a scalar parameter,  $\alpha$ , from

$$E[f(z_i, \theta + \alpha\delta)] = 0$$

by

$$\hat{a}(\delta) = \arg \min_a \left( \frac{1}{n} \sum_{i=1}^n f(z_i, \theta + a\delta) \right)' W_n \left( \frac{1}{n} \sum_{i=1}^n f(z_i, \theta + a\delta) \right).$$

The asymptotic variance of two such estimators corresponding to different  $\delta$  would be

$$\begin{aligned} \Omega_{\delta_1, \delta_2} &= A \text{var} \left( \begin{pmatrix} \hat{a}(\delta_1) \\ \hat{a}(\delta_2) \end{pmatrix} \right) \\ &= \begin{pmatrix} (\delta_1' \Gamma' W \Gamma \delta_1)^{-1} \delta_1' \Gamma' W S W \Gamma \delta_1 (\delta_1' \Gamma' W \Gamma \delta_1)^{-1} \\ (\delta_1' \Gamma' W \Gamma \delta_1)^{-1} \delta_1' \Gamma' W S W \Gamma \delta_2 (\delta_2' \Gamma' W \Gamma \delta_2)^{-1} \\ (\delta_1' \Gamma' W \Gamma \delta_1)^{-1} \delta_1' \Gamma' W S W \Gamma \delta_2 (\delta_2' \Gamma' W \Gamma \delta_2)^{-1} \\ (\delta_2' \Gamma' W \Gamma \delta_2)^{-1} \delta_2' \Gamma' W S W \Gamma \delta_2 (\delta_2' \Gamma' W \Gamma \delta_2)^{-1} \end{pmatrix}. \end{aligned} \tag{7}$$

Of course, (7) has exactly the same structure as (3) and we can therefore back out the matrices  $\Gamma'W\Gamma$  and  $\Gamma'WSW\Gamma$  (up to scale) in exactly the same way that we backed out  $H$  and  $V$  above.

The validity of the bootstrap as a way to approximate the distribution of  $\hat{a}(\delta)$  in this GMM setting is discussed in [Appendix A](#). The result stated there is a minor modification of the result in [Hahn \(1996\)](#).

### 3. METHOD OF MOMENTS

A key advantage of the approach developed in [Section 2](#) is that the proposed bootstrap procedure is based on a minimization problem that uses the same objective function as the original estimator. In this section, we discuss modifications of the proposed bootstrap procedure to just identified method of moments estimators. It is, of course, possible to think of this case as a special case of generalized method of moments. Since the GMM weighting matrix plays no role for the asymptotic distribution in the just identified case, (6) becomes  $\Sigma = (\Gamma' \Gamma)^{-1} \Gamma' S \Gamma (\Gamma' \Gamma)^{-1}$  and the approach in [Section 2](#) can be used to recover  $\Gamma' \Gamma$  and  $\Gamma' S \Gamma$ . Here we will introduce an alternative bootstrap approach which can be used to estimate  $\Gamma$  and  $S$  directly. In doing this, we implicitly assume that all elements of  $\Gamma$  are nonzero.

The just identified method of moments estimator is defined by<sup>1</sup>

$$\frac{1}{n} \sum_{i=1}^n f(z_i, \hat{\theta}) \approx 0$$

<sup>1</sup>The “ $\approx$ ” notation is used as a reminder that the sample moments can be discontinuous and that it can therefore be impossible to make them exactly zero.

and, using the notation from Section 2.2, the asymptotic variance is  $\Sigma = (\Gamma^{-1})S(\Gamma^{-1})'$ . This is very similar to the expression for the asymptotic variance of the extremum estimator in Section 2.1. The difference is that the  $\Gamma$  matrix is typically only symmetric if the moment condition corresponds to the first-order condition for an optimization problem.

We start by noting that there is no loss of generality in normalizing the diagonal elements of  $\Gamma$ ,  $\gamma_{pp}$ , to 1 since the scale of  $f$  does not matter (at least asymptotically). Now consider the infeasible one-dimensional estimator,  $\widehat{a}_{p\ell}$ , that solves the  $p$ th moment with respect to the  $\ell$ th element of the parameter, holding the other elements of  $\theta$  fixed at the true value:

$$\frac{1}{n} \sum_{i=1}^n f_p(z_i, \theta_0 + \widehat{a}_{p\ell} e_\ell) \approx 0.$$

It is straightforward to show that the asymptotic covariance between two such estimators is

$$A \text{cov}(\widehat{a}_{p\ell}, \widehat{a}_{jm}) = \frac{s_{pj}}{\gamma_{p\ell} \gamma_{jm}},$$

where  $s_{pj}$  and  $\gamma_{jp}$  denote the elements in  $S$  and  $\Gamma$ , respectively. In particular,  $A \text{var}(\widehat{a}_{pp}) = s_{pp}/\gamma_{pp}^2 = s_{pp}$ . Hence  $s_{pp}$  is identified. Since  $A \text{cov}(\widehat{a}_{pp}, \widehat{a}_{jj}) = s_{pj}/(\gamma_{pp} \gamma_{jj}) = s_{pj}$ ,  $s_{pj}$  is identified as well. In other words,  $S$  is identified. Having already identified  $s_{pj}$  and  $\gamma_{jj}$ , the remaining elements of  $\Gamma$  are identified from  $A \text{cov}(\widehat{a}_{pp}, \widehat{a}_{jm}) = s_{pj}/\gamma_{pp} \gamma_{jm} = s_{pj}/\gamma_{jm}$ .

In practice, one would first generate  $B$  bootstrap samples,  $\{z_i^b\}_{i=1}^n$ . For each sample, the estimators,  $\widehat{a}_{p\ell}$ , are calculated from

$$\frac{1}{n} \sum_{i=1}^n f_p(z_i^b, \widehat{\theta} + \widehat{a}_{p\ell} e_\ell) \approx 0.$$

The matrix  $S$  can then be estimated by  $\widehat{\text{cov}}(\widehat{a}_{11}, \widehat{a}_{22}, \dots, \widehat{a}_{kk})$ . The elements of  $\Gamma$ ,  $\gamma_{jm}$ , can be estimated by  $\frac{\widehat{s}_{pj}}{\widehat{\text{cov}}(\widehat{a}_{pp}, \widehat{a}_{jm})}$  for arbitrary  $p$  or by  $\sum_{\ell=1}^k w_\ell \frac{\widehat{s}_{p\ell}}{\widehat{\text{cov}}(\widehat{a}_{\ell\ell}, \widehat{a}_{jm})}$  where the weights add up to 1,  $\sum_{\ell=1}^k w_\ell = 1$ . The weights could be chosen on the basis of an estimate of the variance of  $(\frac{\widehat{s}_{p1}}{\widehat{\text{cov}}(\widehat{a}_{11}, \widehat{a}_{jm})}, \dots, \frac{\widehat{s}_{pk}}{\widehat{\text{cov}}(\widehat{a}_{kk}, \widehat{a}_{jm})})$ .

The elements for  $\Gamma$  and  $S$  can also be estimated by minimizing

$$\sum_{p,\ell,j,m} \left( \widehat{\text{cov}}(\widehat{a}_{p\ell}, \widehat{a}_{jm}) - \frac{s_{pj}}{\gamma_{p\ell} \gamma_{jm}} \right)^2$$

with the normalizations,  $\gamma_{jj} = 1$ ,  $s_{pj} = s_{jp}$ , and  $s_{jj} > 0$  for all  $j$ . Alternatively, it is also possible to minimize

$$\sum_{p,\ell,j,m} \left( \widehat{\text{cov}}(\widehat{a}_{p\ell}, \widehat{a}_{jm}) \gamma_{p\ell} \gamma_{jm} - s_{pj} \right)^2.$$

To impose the restriction that  $S$  is positive semidefinite, it is convenient to normalize the diagonal of  $\Gamma$  to be 1 and parameterize  $S$  as  $TT'$ , where  $T$  is a lower triangular matrix.

#### 4. REDUCING THE NUMBER OF DIRECTIONAL ESTIMATORS

Needless to say, choosing  $D$  to contain all vectors of the form  $e_j$ ,  $e_j + e_\ell$ , and  $e_j - e_\ell$  will lead to a system that is wildly overidentified. Specifically, if the dimension of the

parameter vector is  $k$ , then we will be calculating  $k^2$  one-dimensional estimators. This will lead to a covariance matrix with  $(k^4 + k^2)/2$  unique elements. On the other hand,  $H$  and  $V$  are both symmetric  $k$ -by- $k$  matrices. In that sense, we have  $k^4 + k^2$  equations with  $k^2 + k - 1$  unknowns.<sup>2</sup>

Unfortunately, it turns out that the bulk of this overidentification is in  $V$ . To see this, suppose that  $V$  is known and that one has bootstrapped the joint distribution of  $m - 1$  one-dimensional estimators in directions  $\delta_\ell$  ( $\ell = 1, \dots, m - 1$ ). The variance of each of those one-dimensional estimators is  $(\delta'_\ell H \delta_\ell)^{-1} \delta'_\ell V \delta_\ell (\delta'_\ell H \delta_\ell)^{-1}$ . As a result, we can consider  $(\delta'_\ell H \delta_\ell)$  known.

Now imagine that we add one more one-dimensional estimator in the direction  $\delta_m$ . The additional information from this will be the variance of the estimator,  $(\delta'_m H \delta_m)^{-1} \delta'_m V \delta_m (\delta'_m H \delta_m)^{-1}$ , and its covariance with each of the first  $m - 1$  one-dimensional estimators,  $(\delta'_\ell H \delta_\ell)^{-1} \delta'_\ell V \delta_m (\delta'_m H \delta_m)^{-1}$ . Since  $V$  is known, and we already know  $(\delta'_\ell H \delta_\ell)$ , the only new information from the  $m$ th estimator is  $(\delta'_m H \delta_m)$ . In other words, each estimator gives one scalar piece of information about  $H$ . Since  $H$  has  $k(k + 1)/2$  elements, we need at least that many one-directional estimators.

Of course, the analysis in the previous section requires one to consider  $k^2$  directions, while the discussion above suggests that with known  $V$ , calculation of  $H$  requires only  $k(k + 1)/2$  one-dimensional estimators. In this sense, the approach in the previous section is wasteful, because it calculates approximately twice as many one-dimensional estimators as necessary (if  $V$  is known). We now demonstrate one way to reduce the number of one-dimensional estimators by (essentially) a factor of 2 without sacrificing identification (including identification of  $V$ ). In the previous section, we considered estimators in the directions  $e_j$  ( $j = 1, \dots, k$ ),  $e_j + e_\ell$  ( $\ell < j$ ), and  $e_j - e_\ell$  ( $\ell < j$ ). Here we consider only estimators in the directions  $e_j$  ( $j = 1, \dots, k$ ),  $e_j + e_1$  ( $j > 1$ ) and  $e_j - e_1$  ( $j > 1$ ).

We start by considering the one-dimensional estimators in the directions  $e_j$  ( $j = 1, \dots, k$ ),  $e_j + e_1$  ( $j = 2, \dots, k$ ), and  $e_j - e_1$  ( $j = 2, \dots, k$ ). There are  $3k - 2$  such estimators. By the argument above, their asymptotic variance identifies all elements of the  $H$  and  $V$  matrices of the form  $h_{11}$ ,  $h_{1j}$ ,  $h_{jj}$ ,  $v_{11}$ ,  $v_{1j}$ , and  $v_{jj}$  (after we have normalized  $v_{11} = 1$ ). This gives the diagonal elements of  $H$  and  $V$  as well as their first rows (and columns). The asymptotic correlation between  $\widehat{a}(e_j)$  and  $\widehat{a}(e_\ell)$  is  $v_{j\ell} / \sqrt{v_{jj} v_{\ell\ell}}$ . This gives the remaining elements of  $V$ . There are  $(k - 1)(k - 2)/2$  remaining elements of  $H$ ,  $h_{j\ell}$  with  $j > \ell > 1$ . To recover  $h_{j\ell}$ , consider the asymptotic covariance between  $\widehat{a}(e_1)$  and  $\widehat{a}(e_j + e_\ell)$ :

$$h_{11}^{-1}(v_{1j} + v_{1\ell})(h_{jj} + h_{\ell\ell} + 2h_{j\ell})^{-1},$$

which yields  $h_{j\ell}$ .

It is therefore possible to identify all of  $V$  and all of  $H$  with a total of  $(3k - 2) + (k - 1)(k - 2)/2 = (k^2 + 3k - 2)/2$  one-dimensional estimators. One disadvantage of this approach is that it treats the first element of the parameter vector differently from the others. We will therefore not pursue it further.

As mentioned above, the bulk of the overidentification is in  $V$ . This implies that we can recover  $V$  with much less information if  $H$  is known (or easily estimated). Specifically, the  $k$  one-dimensional estimators in the directions  $e_j$  ( $j = 1, \dots, k$ ) will have covariance matrix  $\text{diag}((h_{11}^{-1}, \dots, h_{kk}^{-1}))V \text{diag}((h_{11}^{-1}, \dots, h_{kk}^{-1}))$  from which  $V$  can be recovered.<sup>3</sup>

<sup>2</sup> $H$  and  $V$  both have  $(k^2 + k)/2$  unique elements and we impose one normalization.

<sup>3</sup>This insight can potentially be used to reduce the computational burden in (5).



### 4.1. Simplification When Information Equality Holds

Efficient generalized method of moments estimation in Section 2.2 implies that  $(\Gamma'W\Gamma) = \Gamma'WSW\Gamma$  and maximum likelihood estimation in Section 2.1 implies that  $H = V$ . Either way, the asymptotic variance of the estimator reduces to<sup>4</sup>  $H^{-1}$ , while the asymptotic variance of the  $k$  one-dimensional estimators in the directions  $e_1, \dots, e_k$ ,  $\widehat{a}(e_1), \dots, \widehat{a}(e_k)$ , is

$$\text{diag}(H)^{-1}H \text{diag}(H)^{-1}$$

(see equations (3) and (7)). The asymptotic variance of  $\widehat{a}(e_j)$  is therefore  $h_{jj}^{-1}$ . In other words,  $\text{diag}(H)^{-1} = \text{diag}(V(\widehat{a}(e_1), \dots, \widehat{a}(e_k)))$  and hence

$$H = \text{diag}(V(\widehat{a}(e_1), \dots, \widehat{a}(e_k)))^{-1}V(\widehat{a}(e_1), \dots, \widehat{a}(e_k)) \text{diag}(V(\widehat{a}(e_1), \dots, \widehat{a}(e_k)))^{-1}.$$

Therefore, it is possible to estimate the variance of the parameter of interest by bootstrapping only  $k$  one-dimensional estimators.

### 4.2. Exploiting Specific Model Structures

It is sometimes possible to reduce the computational burden by exploiting specific properties of the estimator of interest. For example, consider the case where a subvector can be easily estimated if one holds the remaining parts of the parameter vector fixed. Regression models of the type  $y_i = \beta_0 + x_{i1}^{\alpha_1}\beta_1 + x_{i2}^{\alpha_2}\beta_2 + \varepsilon_i$  are a textbook example of this; for fixed  $\alpha_1$  and  $\alpha_2$ , the  $\beta$ 's can be estimated by OLS. The same applies to regression models with Box–Cox transformations. The model estimated in Section 6.3 is yet another example where some parameters are easy to estimate for given values of the remaining parameters.

To explore the benefits of this situation, write  $\theta = (\alpha', \beta)'$ , where  $\beta$  can be easily estimated for fixed  $\alpha$ . In the following, we split  $H$  and  $V$  as

$$H = \begin{pmatrix} H_{\alpha\alpha} & H_{\alpha\beta} \\ H_{\beta\alpha} & H_{\beta\beta} \end{pmatrix} \quad \text{and} \quad V = \begin{pmatrix} V_{\alpha\alpha} & V_{\alpha\beta} \\ V_{\beta\alpha} & V_{\beta\beta} \end{pmatrix}.$$

Furthermore, we denote the  $j$ th columns of  $H_{\alpha\beta}$  and  $V_{\alpha\beta}$  by  $H_{\alpha\beta_j}$  and  $V_{\alpha\beta_j}$ , respectively. Similarly,  $H_{\beta_j\beta_\ell}$  and  $V_{\beta_j\beta_\ell}$  will denote the  $(j, \ell)$  elements of  $H_{\beta\beta}$  and  $V_{\beta\beta}$ .

Let  $\tilde{\theta}_j = (\alpha', \beta_j)'$ . The approach from Section 2.1 can be used to back out  $V_{\alpha\alpha}, V_{\alpha\beta_j}, H_{\alpha\alpha}$ , and  $H_{\alpha\beta_j}$ . In other words, we know all of  $H$  and  $V$  except for the off-diagonals of  $H_{\beta\beta}$  and  $V_{\beta\beta}$ . If the dimension of  $\alpha$  is 1, this will require  $3k - 2$  one-dimensional estimators:  $k$  in the directions  $e_j$  ( $j = 1, \dots, k$ ),  $k - 1$  in the directions  $e_j + e_1$  ( $j > 1$ ), and  $k - 1$  in the directions  $e_j - e_1$ .

In the process of applying the identification approach from Section 2.1, one also recovers the correlation of  $\widehat{\beta}(\delta_j)$  and  $\widehat{\beta}(\delta_\ell)$ . As noted above, this correlation is  $V_{\beta_j\beta_\ell} / \sqrt{V_{\beta_j\beta_j}V_{\beta_\ell\beta_\ell}}$ . As a result, we can also recover all of  $V_{\beta\beta}$ .

Now let  $\widehat{\beta}$  be the estimator of  $\beta$  that fixes  $\alpha$ . Its variance is  $H_{\beta\beta}^{-1}V_{\beta\beta}H_{\beta\beta}^{-1}$ . So to identify  $H_{\beta\beta}$ , we need to solve an equation of the form  $A = XV_{\beta\beta}X$ . Equations of this form (when  $A$  and  $V_{\beta\beta}$  are known) are called Riccati equations; see also Honoré and Hu (2015). When  $A$  and  $V_{\beta\beta}$  are symmetric, positive definite matrices, they have a unique symmetric

<sup>4</sup>In the case of a GMM estimator, define  $H$  to equal  $(\Gamma'W\Gamma)$ .

positive definite solution for  $X$ . In other words, one can back out all of  $H_{\beta\beta}$ . Of course, when  $\hat{\beta}$  is easy to calculate for fixed value of  $\alpha$ , it is also often easy to estimate  $H_{\beta\beta}$  and  $V_{\beta\beta}$  directly without using the bootstrap. This would further reduce the computational burden.

5. TWO-STEP ESTIMATORS

Many empirical applications involve a multi-step estimation procedure where each step is computationally simple and uses the estimates from the previous steps. Heckman’s two-step estimator is a textbook example of this. Let

$$d_i = 1\{z_i'\alpha + \nu_i \geq 0\},$$

$$y_i = d_i \cdot (x_i'\beta + \varepsilon_i),$$

where  $(\nu_i, \varepsilon_i)$  has a bivariate normal distribution. The parameter,  $\alpha$ , can be estimated by the probit maximum likelihood estimator,  $\hat{\alpha}_{MLE}$ , in a model with  $d_i$  as the outcome and  $z_i$  as the explanatory variables. In a second step,  $\beta$  is then estimated by the coefficients on  $x_i$  in the regression of  $y_i$  on  $x_i$  and  $\lambda_i = \frac{\phi(z_i'\hat{\alpha}_{MLE})}{1-\Phi(z_i'\hat{\alpha}_{MLE})}$  using only the sample for which  $d_i = 1$ . See Heckman (1979).

Finite-dimensional multi-step estimators can be thought of as GMM or method of moments estimators. As such, their asymptotic variances have a sandwich structure and the poor (wo)man’s bootstrap approach discussed in Sections 2.2 or 3 can therefore in principle be applied. However, the one-dimensional estimation used there does not preserve the simplicity of the multi-step structure. For example, Heckman’s two-step estimator is based on two simple optimization problems (probit and OLS) which deliver  $\hat{\alpha}$  and  $\hat{\beta}$  separately, whereas the procedure in Section 2.2 uses a more complicated estimation problem that involves minimization with respect to linear combinations of elements of both  $\alpha$  and  $\beta$ . Likewise, the approach in Section 3 would involve solving the OLS moment equations with respect to elements of  $\alpha$ . The simplicity of the multi-step procedure is lost either way. In this section, we therefore propose a version of the poor (wo)man’s bootstrap that is suitable for multi-step estimation procedures.

To simplify the exposition, we consider a two-step estimation procedure where the estimator in each step is defined by minimization problems

$$\hat{\theta}_1 = \arg \min_{\tau_1} \frac{1}{n} \sum Q(z_i, \tau_1),$$

$$\hat{\theta}_2 = \arg \min_{\tau_2} \frac{1}{n} \sum R(z_i, \hat{\theta}_1, \tau_2),$$
(8)

with limiting first-order conditions

$$E[q(z_i, \theta_1)] = 0,$$

$$E[r(z_i, \theta_1, \theta_2)] = 0,$$

where  $\theta_1$  and  $\theta_2$  are  $k_1$ - and  $k_2$ -dimensional parameters of interest and  $q(\cdot, \cdot)$  and  $r(\cdot, \cdot, \cdot)$  are smooth functions. Although our exposition requires this, the results also apply when one or both steps involve an extremum estimator with possibly non-smooth objective function or GMM with possibly discontinuous moment function.

Under random sampling,  $\hat{\theta} = (\hat{\theta}_1', \hat{\theta}_2')$  will have a limiting normal distribution with asymptotic variance

$$\begin{pmatrix} Q_1 & 0 \\ R_1 & R_2 \end{pmatrix}^{-1} \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \left( \begin{pmatrix} Q_1 & 0 \\ R_1 & R_2 \end{pmatrix}^{-1} \right)', \tag{9}$$

where  $\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} = V[\begin{pmatrix} q(z_i, \theta_1) \\ r(z_i, \theta_1, \theta_2) \end{pmatrix}]$ ,  $Q_1 = E[\frac{\partial q(z_i, \theta_1)}{\partial \theta_1}]$ ,  $R_1 = E[\frac{\partial r(z_i, \theta_1, \theta_2)}{\partial \theta_1}]$ , and  $R_2 = E[\frac{\partial r(z_i, \theta_1, \theta_2)}{\partial \theta_2}]$ . Getting  $R_1$  and  $V_{12}$  is usually the difficult part. It is often easy to estimate  $V_{11}$ ,  $V_{22}$ ,  $Q_1$ , and  $R_2$  directly, and when that is not possible, they can be estimated using the poor (wo)man's bootstrap procedure above. It follows from (9) that the asymptotic variance of  $\hat{\theta}_2$  is

$$R_2^{-1}R_1Q_1^{-1}V_{11}Q_1^{-1}R_1'R_2^{-1} - R_2^{-1}V_{21}Q_1^{-1}R_1'R_2^{-1} - R_2^{-1}R_1Q_1^{-1}V_{12}R_2^{-1} + R_2^{-1}V_{22}R_2^{-1}, \tag{10}$$

where the first three terms represent the correction for the fact that  $\hat{\theta}_2$  is based on an estimator of  $\theta_1$ .

The matrix in (9) has the usual sandwich structure, and the poor (wo)man's bootstrap can therefore be used to back out all the elements of the two matrices involved. However, this is not necessarily convenient because the poor (wo)man's bootstrap would use the bootstrap sample to estimate scalar  $a$  where  $\theta = (\theta_1', \theta_2')$  has been parameterized as  $\hat{\theta} + a\delta$ . When  $\delta$  places weight on elements from both  $\theta_1$  and  $\theta_2$ , the estimation of  $a$  no longer benefits from the simplicity of the two-step setup.

As noted above, the elements of  $Q_1$  and  $V_{11}$  can often be estimated directly and, if not, they can be estimated by applying the poor (wo)man's bootstrap to the first step in the estimation procedure alone. The matrices  $R_2$  and  $V_{22}$  are also often easily obtained or can be estimated by applying the poor (wo)man's bootstrap to the second step of the estimation procedure holding  $\hat{\theta}_1$  fixed. For example, for Heckman's two-step estimator,  $Q_1$  and  $V_{11}$  can be estimated by the scaled Hessian and score-variance for probit maximum likelihood estimation;  $R_2$  and  $V_{22}$  can be estimated by the (scaled) " $X'X$ " and " $X'e'eX$ " where  $X$  is the design matrix in the regression and  $e$  is the vector of residuals.

To estimate the elements of  $R_1$  and  $V_{12}$ , consider the three infeasible scalar estimators

$$\begin{aligned} \hat{a}_1(\delta_1) &= \arg \min_{a_1} \frac{1}{n} \sum Q(z_i, \theta_1 + a_1\delta_1), \\ \hat{a}_2(\delta_1, \delta_2) &= \arg \min_{a_2} \frac{1}{n} \sum R(z_i, \theta_1 + \hat{a}_1\delta_1, \theta_2 + a_2\delta_2), \\ \hat{a}_3(\delta_3) &= \arg \min_{a_3} \frac{1}{n} \sum R(z_i, \theta_1, \theta_2 + a_3\delta_3), \end{aligned}$$

for fixed  $\delta_1$ ,  $\delta_2$ , and  $\delta_3$ . In Section S-1 of the Supplemental Material (Honoré and Hu (2017)), we show that choosing  $\delta_1 = e_j$  and  $\delta_2 = \delta_3 = e_m$  (for  $j = 1, \dots, k_1$  and  $m = 1, \dots, k_2$ ) identifies all the elements of  $V_{12}$  and  $R_1$ . This requires calculation of  $k_1 + k_1k_2 + k_2$  one-dimensional estimators.

While this identification argument relies on three infeasible estimators, the strategy can be used to estimate  $V_{12}$  and  $R_1$  via the bootstrap. In practice, one would first estimate the parameters  $\theta_1$  and  $\theta_2$ . Using  $B$  bootstrap samples,  $\{z_i^b\}_{i=1}^n$ , one would then obtain  $B$

draws of the vector  $(\widehat{a}_1(e_j), \widehat{a}_2(e_j, e_m), \widehat{a}_3(e_m))$  for  $j = 1, \dots, k_1$  and  $m = 1, \dots, k_2$ , obtained from

$$\begin{aligned} \widehat{a}_1(e_j) &= \arg \min_{a_1} \frac{1}{n} \sum Q(z_i^b, \widehat{\theta}_1 + a_1 e_j), \\ \widehat{a}_2(e_j, e_m) &= \arg \min_{a_2} \frac{1}{n} \sum R(z_i^b, \widehat{\theta}_1 + \widehat{a}_1 e_j, \widehat{\theta}_2 + a_2 e_m), \\ \widehat{a}_3(e_m) &= \arg \min_{a_3} \frac{1}{n} \sum R(z_i^b, \widehat{\theta}_1, \theta_2 + a_3 e_m). \end{aligned}$$

These  $B$  draws can be used to estimate the variance-covariance matrix of  $(\widehat{a}_1(e_j), \widehat{a}_2(e_j, e_m), \widehat{a}_3(e_m))$  and one can then mimic the logic in Section 2.1 to estimate  $V_{12}$  and  $R_1$ .

Many two-step estimation problems have the feature that one of the steps is relatively easier than the other. For example, the second step in Heckman’s (1979) two-step estimator is a linear regression, while the first is maximum likelihood. Similarly, the second step in Powell’s (1987) estimator of the same model also involves a linear regression, while the first-step estimator is an estimator of a semiparametric discrete choice model such as Klein and Spady (1993). On the other hand, the first step in the estimation procedure used in Helpman, Melitz, and Rubinstein (2008) is probit maximum likelihood estimation, which is computationally easy relative to the nonlinear least squares used in the second step. In these situations, it may be natural to apply the one-dimensional bootstrap procedure proposed here to the more challenging step in the estimation procedure, while re-estimating the entire parameter vector in the easier step in each bootstrap sample. We next develop this idea for the case where the first-step estimation is easy. In Section S-2 of the Supplemental Material, we consider the case where the second step is relatively easy. In both cases, it turns out that the correction to the variance for  $\widehat{\theta}_2$  (the first three terms in (10)) can be calculated from the covariances between a first-step estimator and two second-step estimators: one that uses the estimated first-step parameter and one that uses the true value of the first parameter.

Consider again three estimators of the form

$$\begin{aligned} \widehat{a}_1 &= \arg \min_{a_1} \frac{1}{n} \sum Q(z_i, \theta_1 + a_1), \\ \widehat{a}_2(\delta) &= \arg \min_{a_2} \frac{1}{n} \sum R(z_i, \theta_1 + \widehat{a}_1, \theta_2 + a_2 \delta), \\ \widehat{a}_3(\delta) &= \arg \min_{a_3} \frac{1}{n} \sum R(z_i, \theta_1, \theta_2 + a_3 \delta), \end{aligned}$$

but now note that  $\widehat{a}_1$  is a vector of the same dimension as  $\theta_1$ . The asymptotic variance of  $(\widehat{a}_1, \widehat{a}_2(\delta), \widehat{a}_3(\delta))$  is

$$\begin{pmatrix} Q_1 & 0 & 0 \\ \delta'R_1 & \delta'R_2\delta & 0 \\ 0 & 0 & \delta'R_2\delta \end{pmatrix}^{-1} \begin{pmatrix} V_{11} & V_{12}\delta & V_{12}\delta \\ \delta'V'_{12} & \delta'V_{22}\delta & \delta'V_{22}\delta \\ \delta'V'_{12} & \delta'V_{22}\delta & \delta'V_{22}\delta \end{pmatrix} \begin{pmatrix} Q_1 & \delta'R_1 & 0 \\ 0 & \delta'R_2\delta & 0 \\ 0 & 0 & \delta'R_2\delta \end{pmatrix}^{-1}. \quad (11)$$

Multiplying (11) yields a matrix with nine blocks. The upper-middle block is  $-(\delta'R_2\delta)^{-1}Q_1^{-1}V_{11}Q_1^{-1}R_1'\delta + Q_1^{-1}V_{12}\delta(\delta'R_2\delta)^{-1}$ , while the upper-right block is  $Q_1^{-1}V_{12} \times \delta(\delta'R_2\delta)^{-1}$ . With  $R_2$ ,  $V_{11}$ , and  $Q_1$  known and  $\delta = e_j$ , the latter identifies  $V_{12}\delta$  which is the

$j$ th column of  $V_{12}$ . The difference between the upper-middle block and the upper-right block is  $-(\delta'R_2\delta)^{-1}Q_1^{-1}V_{11}Q_1^{-1}R_1'\delta$ . This identifies  $R_1'\delta$  which is the  $j$ th column of  $R_1'$ .

This approach requires calculation of only  $2k_2$  one-dimensional estimators using the more difficult second-step objective function. Moreover, the approach gives closed-form estimates of  $V_{12}$  and  $R_1$ .

6. ILLUSTRATIONS

6.1. Linear Regression

There are no reasons to apply our approach to the estimation of standard errors in a linear regression model. However, its familiarity makes it natural to use this model to illustrate the numerical properties of the approach.

We consider a linear regression model,

$$y_i = x_i'\beta + \varepsilon_i,$$

with 10 explanatory variables generated as follows. For each observation, we first generate a nine-dimensional normal,  $\tilde{x}_i$ , with means equal to 0, variances equal to 1, and all covariances equal to  $\frac{1}{2}$ .  $x_{i1}$  to  $x_{i9}$  are then  $x_{ij} = 1\{\tilde{x}_{ij} \geq 0\}$  for  $j = 1, \dots, 3$ ,  $x_{ij} = \tilde{x}_{ij} + 1$  for  $j = 4$  to 6,  $x_{i7} = \tilde{x}_{i7}$ ,  $x_{i8} = \tilde{x}_{i8}/2$ , and  $x_{i9} = 10\tilde{x}_{i9}$ . Finally,  $x_{i10} = 1$ .  $\varepsilon_i$  is normally distributed conditional on  $x_i$  and with variance  $(1 + x_{i1})^2$ . We pick  $\beta = (\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1, 0, 0, 0, 0, 0)$ . This yields an  $R^2$  of approximately 0.58. The scaling of  $x_{i8}$  and  $x_{i9}$  is meant to make the design a little more challenging for our approach.

We perform 400 Monte Carlo replications, and in each replication, we calculate the OLS estimator, the Eicker–Huber–White variance estimator (E), the bootstrap variance estimator (B), and the variance estimator based on estimating  $V$  and  $H$  from (5) by non-linear least squares (N). The bootstraps are based on 400 bootstrap replications. Based on these, we calculate  $t$ -statistics for testing whether the coefficients are equal to the true values for each of the parameters. Tables I and II report the mean absolute differences in these test statistics for sample sizes of 200 and 2000, respectively.

Tables I and II suggest that our approach works very well when the distribution of the estimator of interest is well approximated by its limiting distribution. Specifically, the difference between the  $t$ -statistics based on our approach and on the regular bootstrap (column 3) is smaller than the difference between the  $t$ -statistics based on the bootstrap and the Eicker–Huber–White variance estimator (column 1).

TABLE I  
ORDINARY LEAST SQUARES,  $n = 200$ . MEAN ABSOLUTE DIFFERENCE IN  $t$ -STATISTICS

	$ T_E - T_B $	$ T_E - T_N $	$ T_B - T_N $
$\beta_1$	0.031	0.029	0.013
$\beta_2$	0.029	0.022	0.022
$\beta_3$	0.031	0.026	0.022
$\beta_4$	0.032	0.026	0.024
$\beta_5$	0.033	0.024	0.023
$\beta_6$	0.032	0.027	0.026
$\beta_7$	0.031	0.024	0.025
$\beta_8$	0.033	0.026	0.023
$\beta_9$	0.034	0.025	0.023
$\beta_{10}$	0.033	0.027	0.021

TABLE II  
ORDINARY LEAST SQUARES,  $n = 2000$ . MEAN ABSOLUTE DIFFERENCE IN  $t$ -STATISTICS

	$ T_E - T_B $	$ T_E - T_N $	$ T_B - T_N $
$\beta_1$	0.025	0.025	0.004
$\beta_2$	0.021	0.021	0.003
$\beta_3$	0.024	0.024	0.004
$\beta_4$	0.023	0.022	0.004
$\beta_5$	0.025	0.025	0.004
$\beta_6$	0.025	0.025	0.004
$\beta_7$	0.026	0.025	0.004
$\beta_8$	0.024	0.023	0.004
$\beta_9$	0.022	0.023	0.003
$\beta_{10}$	0.023	0.023	0.006

6.2. Structural Model

The method proposed here should be especially useful when estimating nonlinear structural models such as those of Lee and Wolpin (2006), Altonji, Smith, and Vidangos (2013), and Dix-Carneiro (2014). To illustrate the usefulness of the poor (wo)man’s bootstrap in such a situation, we consider a very simple two-period Roy model. There are two sectors, labeled 1 and 2. A worker is endowed with a vector of sector-specific human capital,  $x_{si}$ , and sector-specific income in period 1 is  $\log(w_{si1}) = x'_{si}\beta_s + \varepsilon_{si1}$ , and sector-specific income in period 2 is  $\log(w_{si2}) = x'_{si}\beta_s + 1\{d_{i1} = s\}\gamma_s + \varepsilon_{si2}$ , where  $d_{i1}$  is the sector chosen in period 1. We parameterize  $(\varepsilon_{1it}, \varepsilon_{2it})$  to be bivariate normally distributed and i.i.d. over time.

Workers maximize discounted income. First consider time period 2. Here  $d_{i2} = 1$  and  $w_{i2} = w_{1i2}$  if  $w_{1i2} > w_{2i2}$ , that is, if

$$x'_{1i}\beta_1 + 1\{d_{i1} = 1\}\gamma_1 + \varepsilon_{1i2} > x'_{2i}\beta_2 + 1\{d_{i1} = 2\}\gamma_2 + \varepsilon_{2i2}$$

and  $d_{i2} = 2$  and  $w_{i2} = w_{2i2}$  otherwise. In time period 1, workers choose sector 1 ( $d_{i1} = 1$ ) if

$$w_{1i1} + \rho E[\max\{w_{1i2}, w_{2i2}\}|x_{1i}, x_{2i}, d_{i1} = 1] > w_{2i1} + \rho E[\max\{w_{1i2}, w_{2i2}\}|x_{1i}, x_{2i}, d_{i1} = 2]$$

and sector 2 otherwise. In Section S-3 of the Supplemental Material, we demonstrate that the expected value of the maximum of two dependent lognormally distributed random variables is

$$\begin{aligned} & \exp(\mu_1 + \sigma_1^2/2) \left( 1 - \Phi \left( \frac{\mu_2 - \mu_1 - (\sigma_1^2 - \tau\sigma_1\sigma_2)}{\sqrt{\sigma_2^2 - 2\tau\sigma_1\sigma_2 + \sigma_1^2}} \right) \right) \\ & + \exp(\mu_2 + \sigma_2^2/2) \left( 1 - \Phi \left( \frac{\mu_1 - \mu_2 - (\sigma_2^2 - \tau\sigma_1\sigma_2)}{\sqrt{\sigma_2^2 - 2\tau\sigma_1\sigma_2 + \sigma_1^2}} \right) \right), \end{aligned}$$

where the underlying normal random variables have means  $\mu_1$  and  $\mu_2$ , variances  $\sigma_1^2$  and  $\sigma_2^2$ , and correlation  $\tau$ . This gives closed-form solutions for  $w_{1i1} + \rho E[\max\{w_{1i2}, w_{2i2}\}|x_{1i}, x_{2i}, d_{i1} = 1]$  and  $w_{2i1} + \rho E[\max\{w_{1i2}, w_{2i2}\}|x_{1i}, x_{2i}, d_{i2} = 1]$ .

We will now imagine a setting in which the econometrician has a data set with  $n$  observations from this model.  $x_{is}$  is composed of a constant and a normally distributed component

that is independent across sectors and across individuals. In the data-generating process,  $\beta_1 = (1, 1)'$ ,  $\beta_2 = (\frac{1}{2}, 1)'$ ,  $\gamma_1 = 0$ ,  $\gamma_2 = 1$ ,  $\sigma_1^2 = 2$ ,  $\sigma_2^2 = 3$ ,  $\tau = 0$ , and  $\rho = 0.95$ . In the estimation, we treat  $\rho$  and  $\tau$  as known, and we estimate the remaining parameters. Fixing the discount rate parameter is standard and we assume independent errors for computational convenience. The sample size is  $n = 2000$  and the results presented here are based on 400 Monte Carlo replications, each using 1000 bootstrap samples to calculate the poor (wo)man's bootstrap standard errors.

The model is estimated by indirect inference matching the following parameters in the regressions (all estimated by OLS, with the additional notation that  $d_{i0} = 0$ ): (i) the regression coefficients and residual variance in a regression of  $w_{it}$  on  $x_{i1}$ ,  $x_{i2}$ , and  $1\{d_{it-1} = 1\}$  using the subsample of observations in sector 1; (ii) the regression coefficients and residual variance in a regression of  $w_{it}$  on  $x_{i1}$ ,  $x_{i2}$ , and  $1\{d_{it-1} = 1\}$  using the subsample of observations in sector 2; and (iii) the regression coefficients in a regression  $1\{d_{it} = 1\}$  on  $x_{i1}$  and  $x_{i2}$  and  $1\{d_{it-1} = 1\}$ .

Let  $\hat{a}$  be the vector of those parameters based on the data and let  $\widehat{V}[\hat{\alpha}]$  be the associated estimated variance. For a candidate vector of structural parameters,  $\theta$ , the researcher simulates the model two times (holding the draws of the errors constant across different values of  $\theta$ ), calculates the associated  $\tilde{\alpha}(\theta)$ , and estimates the model parameters by minimizing

$$(\hat{a} - \tilde{\alpha}(\theta))' \widehat{V}[\hat{\alpha}]^{-1} (\hat{a} - \tilde{\alpha}(\theta))$$

over  $\theta$ . Note that  $\tilde{\alpha}(\theta)$  is discontinuous in the parameter because there will be some values of  $\theta$  for which the individual is indifferent between the sectors.

This example is deliberately chosen in such a way that we can calculate the asymptotic standard errors. See [Gourieroux and Monfort \(2007\)](#). We use these as a benchmark when evaluating our approach. [Tables III and IV](#) present the results. With the possible exception of the intercept in sector 1, both the standard errors suggested by the asymptotic distribution and the standard errors suggested by the poor (wo)man's bootstrap approximate the standard deviation of the estimator well ([Table III](#)). The computation times make it infeasible to perform a Monte Carlo study that includes the usual bootstrap method. For example, estimating the model with 2000 observations once took approximately 900 seconds. By comparison, calculating all the one-dimensional parameters (once) took less than 5 seconds on the same computer. In addition, the computing cost of minimizing equation (5) was approximately 90 seconds. With 1000 bootstrap replications, this suggests that it would take more than 10 days to do the regular bootstrap in one sample, while our approach would take approximately one and a half hours. [Table IV](#) illuminates

TABLE III  
STRUCTURAL MODEL. ASYMPTOTIC AND ESTIMATED STANDARD ERRORS

	Actual	Asymptotic	Mean BS	Median BS
$\beta_{11}$	0.044	0.049	0.053	0.052
$\beta_{12}$	0.040	0.041	0.042	0.042
$\beta_{21}$	0.050	0.051	0.052	0.052
$\beta_{22}$	0.039	0.040	0.041	0.041
$\gamma_1$	0.027	0.029	0.031	0.031
$\gamma_2$	0.064	0.068	0.068	0.068
$\log(\sigma_1)$	0.023	0.026	0.026	0.026
$\log(\sigma_2)$	0.018	0.019	0.018	0.018

TABLE IV  
STRUCTURAL MODEL. REJECTION PROBABILITIES (20% LEVEL OF SIGNIFICANCE)

	Asymptotic s.e.	Poor (Wo)man's BS s.e.
$\beta_{11}$	16%	13%
$\beta_{12}$	16%	16%
$\beta_{21}$	22%	19%
$\beta_{22}$	19%	18%
$\gamma_1$	19%	16%
$\gamma_2$	17%	17%
$\log(\sigma_1)$	16%	15%
$\log(\sigma_2)$	18%	19%

the performance of the proposed bootstrap procedure for doing inference by comparing the rejection probabilities based on our standard errors to the rejection probabilities based on the true asymptotic standard errors.

### 6.3. Two-Step Estimation

In this section, we illustrate the use of the poor (wo)man's bootstrap applied to two-step estimators using a modification of the empirical model in [Helpman, Melitz, and Rubinstein \(2008\)](#). We first estimate the model and then use the estimated model and the explanatory variables as the basis for a Monte Carlo study.

The econometric model has the feature that the first step can be estimated by a standard probit. We therefore use it to illustrate the situation where the first estimation step is easy as discussed in Section 5. The model also has the feature that in the second step, some of the parameters can be estimated by ordinary least squares for fixed values of the remaining parameters. The example will therefore also illustrate simplification described in Section 4.2. Section S-4 of the Supplemental Material gives the mathematics for combining the insights in Sections 5 and 4.2. Finally, we have deliberately chosen the example to be simple enough that we can compare our approach to the regular bootstrap in the Monte Carlo study.

#### 6.3.1. Model Specification

In one of their specifications, [Helpman, Melitz, and Rubinstein \(2008\)](#) used a parametric two-step sample selection estimation procedure that assumes joint normality to estimate a model for trade flows from an exporting country to an importing country. The estimation involves a probit model for positive trade flow from one country to another in the first step, followed by nonlinear least squares in the second step using only observations that have the dependent variable equal to 1 in the probit. It is a two-step estimation problem, because some of the explanatory variables in the second step are based on the index estimated in the first step. In this specification, the expected value of the logarithm of trade flows in the second equation is of the form

$$x' \beta_1 + \lambda(-z' \hat{\gamma}) \beta_2 + \log(\exp(\beta_3(\lambda(-z' \hat{\gamma}) + z' \hat{\gamma})) - 1), \tag{12}$$

where  $\hat{\gamma}$  is the first-step probit estimator and  $\lambda(\cdot) = \frac{\varphi(\cdot)}{1 - \Phi(\cdot)}$ . Since probit maximum likelihood estimation is based on maximizing a concave objective function, it this is an example whether the first-step estimation of  $\gamma$  is computationally relatively easy. Moreover,



the second step has the feature discussed in Section 4.2, namely, that it is easy to estimate some parameters (here  $\beta_1$  and  $\beta_2$ ) conditional on the rest (here  $\beta_3$ ). One of the key explanatory variables in  $x$  and in  $z$  is logarithm of the distance between countries.

As pointed out in Santos Silva and Tenreyro (2015), this econometrics specification is problematic, both because of the presence of the sample selection correction term inside a nonlinear function and because it is difficult to separately identify  $\beta_2$  and  $\beta_3$ . To illustrate our approach, we therefore consider a modified reduced-form specification that has some of the same features as the model estimated in Helpman, Melitz, and Rubinstein (2008). Specifically, we estimate a sample selection model for trade in which the selection equation (i.e., the model for positive trade flows) is the same as in Helpman, Melitz, and Rubinstein (2008), but in which the outcome (i.e., the logarithm of trade flows) is linear using the same explanatory variables as Helpman, Melitz, and Rubinstein (2008) except that we allow distance to enter through a Box-Cox transformation rather than through its logarithm. Following Helpman, Melitz, and Rubinstein (2008), we estimate this model by a two-step procedure, but in our case the second step involves nonlinear least squares estimation of the equation

$$y_i = \beta_0 \frac{x_0^\lambda - 1}{\lambda} + x_1' \beta_1 + \lambda(-z' \hat{\gamma}) \beta_2 + \text{error}_i,$$

where  $x_0$  is the distance between the exporting country and the importing country. When  $x_1$  contains a constant or a saturated set of dummies, this model can be written as

$$y_i = \tilde{\beta}_0 x_0^\lambda + x_1' \tilde{\beta}_1 + \lambda(-z' \hat{\gamma}) \beta_2 + \text{error}_i. \quad (13)$$

Like (12), equation (13) has one parameter that enters nonlinearly. As a result, the second step again has the feature discussed in Section 4.2.

Helpman, Melitz, and Rubinstein (2008) used a panel from 1980 to 1989 of the trade flows (exports) from each of 158 countries to the remaining 157 countries.<sup>5</sup> In the specification that we mimic, the explanatory variables in the selection equation are (1) DISTANCE (the logarithm of the geographical distance between the capitals), (2) BORDER (a binary variable indicating if both countries share a common border), (3) ISLAND (a binary variable indicating if both countries in the pair are islands), (4) LANDLOCK (indicating if both countries in the pair are landlocked), (5) COLONY (indicating if one of the countries in the pair colonized the other one), (6) LEGAL (indicating if both countries in the pair have the same legal system), (7) LANGUAGE (indicating if both countries in the pair speak the same language), (8) RELIGION (a variable measuring the similarity in the shares of Protestants, Catholics, and Muslims in the countries in the pair; a higher number indicates a bigger similarity), (9) CU (indicating whether two countries have the same currency or have a 1:1 exchange rate), (10) FTA (indicating if both countries are part of a free trade agreement), (11) WTONONE and (12) WTOBOTH (binary variables indicating if neither or both countries are members of the WTO, respectively). They also include a full set of year dummies as well as import country and export country fixed effects (which are estimated as parameters). The explanatory variables in the second equation are the same variables except for RELIGION.

In our Monte Carlo study, we use the same explanatory variables as Helpman, Melitz, and Rubinstein (2008) except that we replace the country fixed effects by continent fixed

<sup>5</sup>See <http://scholar.harvard.edu/melitz/publications>.

effects. The reason is that when we simulated from the estimated model, we frequently generated data from which it was impossible to estimate all the probit parameters.<sup>6</sup>

To illustrate that our method can be used in “less than ideal” situations, we generate data from the full model, but estimate the selection equation (the probit) using only data from 1980. This is because some papers estimate the first step and the second step using different samples. Using only data from one year in the selection necessitates replacing the year dummies in the selection equation with a constant. In the second estimation step, we use data from all the years and include a full set of year dummies.

### 6.3.2. Monte Carlo Results

We first estimate the model using the actual data. This gives the values of  $\gamma$ ,  $\lambda$ ,  $\tilde{\beta}_0$ ,  $\tilde{\beta}_1$ , and  $\beta_2$  to be used in the data-generating process. We then set the correlation between the errors in the selection and the outcome equations to 0.5 and we calibrate the variance of the error in the second equation to 3<sup>2</sup>. This roughly matches the variance of the residuals in the second equation in the data-generating process to the same variance in the data.

The Monte Carlo study uses 400 replications. These replications use the same explanatory variables as in the actual data, and they only differ in the draws of the errors. In each replication, we estimate the parameters and calculate the standard errors using (1) the asymptotic variance that corrects for the two-step estimation, (2) the poor (wo)man’s bootstrap, and (3) the regular bootstrap.<sup>7</sup> In each Monte Carlo replication, we use the same 1000 samples to calculate the two versions of the bootstraps’ standard errors. The results are reported Tables V–VII.

Table V reports the standard deviations of the parameter estimates across the 400 Monte Carlo replications in column 1. Column 2 reports the means of the estimated

TABLE V  
SELECTION MODEL. MEANS OF ESTIMATED STANDARD ERRORS

	Actual	With Correction	Poor (Wo)man’s BS	Regular BS
$\tilde{\beta}_0$	0.290	0.278	0.280	0.283
BORDER	0.077	0.072	0.080	0.080
ISLAND	0.052	0.054	0.059	0.059
LANDLOCKED	0.098	0.093	0.100	0.100
LEGAL	0.024	0.023	0.025	0.025
LANGUAGE	0.026	0.025	0.027	0.027
COLONIAL	0.074	0.068	0.074	0.074
CU	0.127	0.119	0.129	0.128
FTA	0.106	0.094	0.104	0.105
WTONONE	0.045	0.040	0.044	0.043
WTOBOTH	0.025	0.023	0.026	0.026
MILLS	0.052	0.047	0.051	0.051
$\lambda$	0.060	0.054	0.057	0.058

<sup>6</sup>Even when we replaced the country dummies with continent dummies, we sometimes generated data sets from which we could not estimate the probit parameters. When that happened, we re-drew the data.

<sup>7</sup>By concentrating out the coefficients that enter linearly in the second step, it is trivial to do a full bootstrap in this example. We deliberately set it up like this in order to compare the results of our approach to the results from a regular bootstrap.

TABLE VI  
SELECTION MODEL. MEDIANS OF ESTIMATED STANDARD ERRORS

	Actual	With Correction	Poor (Wo)man's BS	Regular BS
$\tilde{\beta}_0$	0.290	0.277	0.278	0.280
BORDER	0.077	0.072	0.080	0.079
ISLAND	0.052	0.054	0.059	0.059
LANDLOCKED	0.098	0.093	0.100	0.100
LEGAL	0.024	0.023	0.025	0.025
LANGUAGE	0.026	0.025	0.027	0.027
COLONIAL	0.074	0.068	0.074	0.073
CU	0.127	0.119	0.129	0.128
FTA	0.106	0.094	0.103	0.105
WTONONE	0.045	0.040	0.043	0.043
WTOBOTH	0.025	0.023	0.026	0.026
MILLS	0.052	0.046	0.050	0.051
$\lambda$	0.060	0.054	0.057	0.058

standard errors using the asymptotic expressions with correction for the two-step estimation. Columns 3 and 4 report the means of the standard errors estimated using the poor (wo)man's bootstrap and the regular bootstrap, respectively. The results for the year dummies and continent fixed effects are omitted. The bootstrap and the poor (wo)man's bootstrap are almost identical in all cases. Moreover, in almost all cases, they are closer to the actual than the standard errors based on the asymptotic distribution. Table VI reports almost identical results for the medians of the estimated standard errors.

Table VII presents the size of the *t*-statistics that test that the parameters equal their true values using different estimates of the standard errors. The results based on the bootstrap and the poor (wo)man's bootstrap are again almost identical in all cases. They are also close to those based on the asymptotic distribution with correction for the two-step estimation.

TABLE VII  
SELECTION MODEL. REJECTION PROBABILITIES (20% LEVEL OF SIGNIFICANCE)

	With Correction	Poor (Wo)man's BS	Regular BS
$\tilde{\beta}_0$	24%	23%	23%
BORDER	26%	20%	20%
ISLAND	20%	16%	16%
LANDLOCKED	20%	17%	18%
LEGAL	25%	20%	20%
LANGUAGE	21%	18%	18%
COLONIAL	25%	21%	22%
CU	21%	19%	19%
FTA	24%	20%	20%
WTONONE	25%	22%	22%
WTOBOTH	25%	19%	19%
MILLS	27%	24%	23%
$\lambda$	25%	23%	21%

7. CONCLUSION

This paper has demonstrated that it is possible to estimate the asymptotic variance for broad classes of estimators using a version of the bootstrap that only relies on the estimation of *one-dimensional* parameters. We believe that this method can be useful for applied researchers who are estimating complicated models in which it is difficult to derive or estimate the asymptotic variance of the estimator of the parameters of interest. The contribution relative to the bootstrap is to provide an approach that can be used when researchers find it time-consuming to reliably recalculate the estimator of the whole parameter vector in each bootstrap replication. This will often be the case when the estimator requires solving an optimization problem to which one cannot apply gradient-based optimization techniques. In those cases, one-dimensional search will be not only faster, but also more reliable.

We have discussed the method in the context of the regular (nonparametric) bootstrap applied to extremum estimators, generalized method of moments estimators, and two-step estimators. However, the same idea can be used without modification for other bootstrap methods such as the weighted bootstrap or the block bootstrap.

APPENDIX A: VALIDITY OF BOOTSTRAP

Hahn (1996) established that under random sampling, the bootstrap distribution of the standard GMM estimator converges weakly to the limiting distribution of the estimator in probability. Here, we state the same result under the same regularity conditions for estimators that treat part of the parameter vector as known. Whenever possible, we use the same notation and the same wording as Hahn (1996). In particular,  $o_p^\omega(\cdot)$ ,  $O_p^\omega(\cdot)$ ,  $o_B(\cdot)$ , and  $O_B(\cdot)$  are defined on page 190 of that paper. A number of papers have proved the validity of the bootstrap in different situations. We choose to tailor our derivation after Hahn (1996) because it so closely mimics the classic proof of asymptotic normality of GMM estimators presented in Pakes and Pollard (1989).

We first review Hahn’s (1996) setup. The parameter of interest  $\theta_0$  is the unique solution to  $G(t) = 0$  where  $G(t) \equiv E[g(Z_i, t)]$ ,  $Z_i$  is the vector of data for observation  $i$ , and  $g$  is a known function. The parameter space is  $\Theta$ .

Let  $G_n(t) \equiv \frac{1}{n} \sum_{i=1}^n g(Z_i, t)$ . The GMM estimator is defined by

$$\tau_n \equiv \arg \min_t |A_n G_n(t)|,$$

where  $A_n$  is a sequence of random matrices (constructed from  $\{Z_i\}$ ) that converges to a nonrandom and nonsingular matrix  $A$ .

The bootstrap estimator is the GMM estimator defined in the same way as  $\tau_n$  but from a bootstrap sample  $\{\widehat{Z}_{n1}, \dots, \widehat{Z}_{nn}\}$ . Specifically,

$$\widehat{\tau}_n \equiv \arg \min_t |\widehat{A}_n \widehat{G}_n(t)|,$$

where  $\widehat{G}_n(t) \equiv \frac{1}{n} \sum_{i=1}^n g(\widehat{Z}_{ni}, t)$ .  $\widehat{A}_n$  is constructed from  $\{\widehat{Z}_{ni}\}_{i=1}^n$  in the same way that  $A_n$  was constructed from  $\{Z_i\}_{i=1}^n$ .

Our result is based on the same GMM setting as in Hahn (1996). The difference is that we are primarily interested in an infeasible estimator that assumes that one part of the parameter vector is known. We will denote the true parameter vector by  $\theta_0$ , which we partition as  $\theta_0^1 = (\theta_0^1, \theta_0^2)$ .

The infeasible estimator of  $\theta_0$ , which assumes that  $\theta_0^2$  is known, is

$$\gamma_n = \arg \min_t \left| A_n G_n \left( \begin{pmatrix} t \\ \theta_0^2 \end{pmatrix} \right) \right| \tag{14}$$

or

$$\gamma_n = \arg \min_t G_n \left( \begin{pmatrix} t \\ \theta_0^2 \end{pmatrix} \right)' A_n' A_n G_n \left( \begin{pmatrix} t \\ \theta_0^2 \end{pmatrix} \right).$$

Let the dimensions of  $\theta_0^1$  and  $\theta_0^2$  be  $k_1$  and  $k_2$ , respectively. It is convenient to define  $E_1 = (I_{k_1 \times k_1} : 0_{k_1 \times k_2})'$  and  $E_2 = (0_{k_2 \times k_1} : I_{k_2 \times k_2})'$ . Post-multiplying a matrix by  $E_1$  or  $E_2$  will extract the first  $k_1$  or the last  $k_2$  columns of the matrix, respectively.

Let

$$(\hat{\theta}^1, \hat{\theta}^2)' = \arg \min_{(t^1, t^2)} G_n \left( \begin{pmatrix} t^1 \\ t^2 \end{pmatrix} \right)' A_n' A_n G_n \left( \begin{pmatrix} t^1 \\ t^2 \end{pmatrix} \right)$$

be the usual GMM estimator of  $\theta_0$ . We consider the bootstrap estimator

$$\hat{\gamma}_n = \arg \min_t \left| \hat{A}_n \hat{G}_n \left( \begin{pmatrix} t \\ \hat{\theta}^2 \end{pmatrix} \right) \right|, \tag{15}$$

where  $\hat{G}_n(t) \equiv \frac{1}{n} \sum_{i=1}^n g(\hat{Z}_{ni}, t)$ .  $\hat{A}_n$  is constructed from  $\{\hat{Z}_{ni}\}_{i=1}^n$  in the same way that  $A_n$  was constructed from  $\{Z_i\}_{i=1}^n$ . Below, we adapt the derivations in Hahn (1996) to show that the distribution of  $\hat{\gamma}_n$  can be used to approximate the distribution of  $\gamma_n$ . We use exactly the same regularity conditions as Hahn (1996). The only exception is that we need an additional assumption to guarantee the consistency of  $\hat{\gamma}_n$ . For this, it is sufficient that the moment function,  $G$ , is continuously differentiable and that the parameter space is compact. This additional stronger assumption would make it possible to state the conditions in Proposition 1 more elegantly. We do not restate those conditions because that would make it more difficult to make the connection to Hahn's (1996) result.

PROPOSITION 1—Adaption of Hahn's (1996) Proposition 1: *Suppose that the conditions in Proposition 1 of Hahn (1996) are satisfied. In addition, suppose that  $G$  is continuously differentiable and that the parameter space is compact. Then  $\gamma_n = \theta_0^1 + o_p(1)$  and  $\hat{\gamma}_n = \theta_0^1 + o_B(1)$ .*

THEOREM 3—Adaption of Hahn's (1996) Theorem 1: *Assume that the conditions in Proposition 1 and Theorem 1 of Hahn (1996) are satisfied. Then*

$$n^{1/2}(\gamma_n - \theta_0^1) \implies N(0, \Omega)$$

and

$$n^{1/2}(\hat{\gamma}_n - \gamma_n) \xrightarrow{p} N(0, \Omega),$$

where

$$\Omega = (E_1' \Gamma' A' A \Gamma E_1)^{-1} E_1' \Gamma' A' A S A' A \Gamma E_1 (E_1' \Gamma' A' A \Gamma E_1)^{-1}$$

and

$$V = E[g(Z_i, \theta_0)g(Z_i, \theta_0)'].$$

The proofs of Proposition 1 and Theorem 3 are provided in Section S-5 of the Supplemental Material.

Theorem 3 is stated for GMM estimators. This covers extremum estimators and the two-step estimators as special cases. Theorem 3 also covers the case where one is interested in different infeasible lower-dimensional estimators as in Section 4.2. To see this, consider two estimators of the form

$$\widehat{a}(\delta_1) = \arg \min_a \left( \frac{1}{n} \sum_{i=1}^n f(x_i, \theta_0 + a\delta_1) \right)' W_n \left( \frac{1}{n} \sum_{i=1}^n f(x_i, \theta_0 + a\delta_1) \right)$$

and

$$\widehat{a}(\delta_2) = \arg \min_a \left( \frac{1}{n} \sum_{i=1}^n f(x_i, \theta_0 + a\delta_2) \right)' W_n \left( \frac{1}{n} \sum_{i=1}^n f(x_i, \theta_0 + a\delta_2) \right),$$

and let  $A_n$  denote the matrix-square root of  $W_n$ . We can then write

$$(\widehat{a}(\delta_1), \widehat{a}(\delta_2)) = \arg \min \left| \begin{pmatrix} A_n & 0 \\ 0 & A_n \end{pmatrix} \frac{1}{n} \sum_{i=1}^n \begin{pmatrix} f(x_i, \theta_0 + a\delta_1) \\ f(x_i, \theta_0 + a\delta_2) \end{pmatrix} \right|,$$

which has the form of (14).

### APPENDIX B: PROOF OF THEOREM 1

Let  $\omega_{\delta_\ell, \delta_k} = (\delta'_\ell H \delta_\ell)^{-1} \delta'_\ell V \delta_j (\delta'_j H \delta_j)^{-1}$ , let  $h_{j\ell}$  denote the elements of  $H$ , and write  $V = \begin{pmatrix} 1 & \rho v \\ \rho v & v^2 \end{pmatrix}$  with  $v > 0$ . We use  $e_j$  to denote a vector that has 1 in its  $j$ th element and zeros elsewhere.

First, note that we can always rotate the coordinate system so that two of the directions are  $e_1$  and  $e_2$ . Let  $(a_1, a_2)'$  and  $(b_1, b_2)'$  be the other two directions (after the rotation).

Second, note that

$$\begin{aligned} \sqrt{\frac{1}{\omega_{e_1, e_1}}} &= h_{11}, \\ \frac{\omega_{e_1, e_2}}{\sqrt{\omega_{e_1, e_1} \omega_{e_2, e_2}}} &= \frac{h_{11}^{-1} \rho v h_{22}^{-1}}{\sqrt{h_{11}^2 h_{22}^2 v^2}} = \rho, \\ \sqrt{\omega_{e_2, e_2}} &= \frac{v}{h_{22}} \equiv k_1, \\ \omega_{e_1, (a_1, a_2)'} &= h_{11}^{-1} (a_1 + a_2 \rho v) (a_1^2 h_{11} + 2a_1 a_2 h_{12} + a_2^2 h_{22})^{-1} \equiv k_2, \\ \omega_{e_1, (b_1, b_2)'} &= h_{11}^{-1} (b_1 + b_2 \rho v) (b_1^2 h_{11} + 2b_1 b_2 h_{12} + b_2^2 h_{22})^{-1} \equiv k_3. \end{aligned}$$

So  $\rho$  and  $h_{11}$  are identified. Using the third equation, the last two equations can be written as

$$\begin{aligned} k_2 h_{11} (a_1^2 h_{11} + 2a_1 a_2 h_{12} + a_2^2 h_{22}) &= (a_1 + a_2 \rho k_1 h_{22}), \\ k_3 h_{11} (b_1^2 h_{11} + 2b_1 b_2 h_{12} + b_2^2 h_{22}) &= (b_1 + b_2 \rho k_1 h_{22}), \end{aligned}$$

or

$$\begin{pmatrix} 2k_2h_{11}a_1a_2 & k_2h_{11}a_2^2 - a_2\rho k_1 \\ 2k_3h_{11}b_1b_2 & k_3h_{11}b_2^2 - b_2\rho k_1 \end{pmatrix} \begin{pmatrix} h_{12} \\ h_{22} \end{pmatrix} = \begin{pmatrix} a_1 - a_1^2k_2h_{11}^2 \\ b_1 - b_1^2k_3h_{11}^2 \end{pmatrix}. \tag{16}$$

Below, we show that the determinant of the matrix on the left cannot be zero. As a result, (16) has a unique solution for  $h_{12}$  and  $h_{22}$ . Once we have  $h_{22}$ , we then get the remaining unknown,  $v$ , from  $v = k_1h_{22}$ . This will complete the proof.

The determinant of the matrix on the left of (16) is

$$\begin{aligned} & (2k_2h_{11}a_1a_2)(k_3h_{11}b_2^2 - b_2\rho k_1) - (k_2h_{11}a_2^2 - a_2\rho k_1)(2k_3h_{11}b_1b_2) \\ &= (2h_{11}^{-1}(a_1 + a_2\rho v)(a_1^2h_{11} + 2a_1a_2h_{12} + a_2^2h_{22})^{-1}h_{11}a_1a_2) \\ & \quad \times \left( h_{11}^{-1}(b_1 + b_2\rho v)(b_1^2h_{11} + 2b_1b_2h_{12} + b_2^2h_{22})^{-1}h_{11}b_2^2 - b_2\rho \frac{v}{h_{22}} \right) \\ & \quad - \left( h_{11}^{-1}(a_1 + a_2\rho v)(a_1^2h_{11} + 2a_1a_2h_{12} + a_2^2h_{22})^{-1}h_{11}a_2^2 - a_2\rho \frac{v}{h_{22}} \right) \\ & \quad \times (2h_{11}^{-1}(b_1 + b_2\rho v)(b_1^2h_{11} + 2b_1b_2h_{12} + b_2^2h_{22})^{-1}h_{11}b_1b_2) \\ &= \frac{(2(a_1 + a_2\rho v)a_1a_2)}{(a_1^2h_{11} + 2a_1a_2h_{12} + a_2^2h_{22})} \\ & \quad \times \left( \frac{(b_1 + b_2\rho v)b_2^2h_{22}}{(b_1^2h_{11} + 2b_1b_2h_{12} + b_2^2h_{22})h_{22}} - \frac{b_2\rho v(b_1^2h_{11} + 2b_1b_2h_{12} + b_2^2h_{22})}{(b_1^2h_{11} + 2b_1b_2h_{12} + b_2^2h_{22})h_{22}} \right) \\ & \quad - \frac{2(b_1 + b_2\rho v)b_1b_2}{(b_1^2h_{11} + 2b_1b_2h_{12} + b_2^2h_{22})} \\ & \quad \times \left( \frac{(a_1 + a_2\rho v)a_2^2h_{22}}{(a_1^2h_{11} + 2a_1a_2h_{12} + a_2^2h_{22})h_{22}} - \frac{a_2\rho v(a_1^2h_{11} + 2a_1a_2h_{12} + a_2^2h_{22})}{(a_1^2h_{11} + 2a_1a_2h_{12} + a_2^2h_{22})h_{22}} \right) \\ &= \frac{(2(a_1 + a_2\rho v)a_1a_2)}{(a_1^2h_{11} + 2a_1a_2h_{12} + a_2^2h_{22})} \left( \frac{(b_1b_2^2h_{22} - b_2\rho v(b_1^2h_{11} + 2b_1b_2h_{12}))}{(b_1^2h_{11} + 2b_1b_2h_{12} + b_2^2h_{22})h_{22}} \right) \\ & \quad - \frac{2(b_1 + b_2\rho v)b_1b_2}{(b_1^2h_{11} + 2b_1b_2h_{12} + b_2^2h_{22})} \left( \frac{(a_1a_2^2h_{22} - a_2\rho v(a_1^2h_{11} + 2a_1a_2h_{12}))}{(a_1^2h_{11} + 2a_1a_2h_{12} + a_2^2h_{22})h_{22}} \right). \end{aligned}$$

Multiplying out and canceling terms, we get

$$\begin{aligned} & \frac{2a_1a_2b_1b_2(a_1b_2 - a_2b_1)(h_{11}v^2\rho^2 - 2h_{12}v\rho + h_{22})}{(b_1^2h_{11} + 2b_1b_2h_{12} + b_2^2h_{22})(a_1^2h_{11} + 2a_1a_2h_{12} + a_2^2h_{22})h_{22}} \\ &= 2a_1a_2b_1b_2(a_1b_2 - a_2b_1) \frac{[(-\rho v, 1)H(-\rho v, 1)']}{[(b_1, b_2)H(b_1, b_2)'][(a_1, a_2)H(a_1, a_2)'] [e_2'He_2]}. \end{aligned}$$

Each of the four terms in brackets is positive because  $H$  is positive definite. Moreover, since none of four directions is proportional to another  $a_1a_2b_1b_2(a_1b_2 - a_2b_1)$  cannot be zero.

APPENDIX C: EXPLOITING THE STRUCTURE IN HELPMAN ET AL.

In the specification used by [Helpman, Melitz, and Rubinstein \(2008\)](#) and in the modification in Section 6.3.1, it is relatively easy to re-estimate the first-step parameter in each bootstrap replication. In the second step, it is easy to estimate  $\beta_1$  and  $\beta_2$  for given value of  $\beta_3$ , since this is a linear regression. We therefore consider estimators of the form

$$\begin{aligned} \hat{a}_1 &= \arg \min_{a_1} \frac{1}{n} \sum Q(z_i, \theta_1 + a_1), \\ \hat{a}_2(\Delta) &= \arg \min_{a_2} \frac{1}{n} \sum R(z_i, \theta_1 + \hat{a}_1, \theta_2 + \Delta a_2), \\ \hat{a}_3(\Delta) &= \arg \min_{a_3} \frac{1}{n} \sum R(z_i, \theta_1, \theta_2 + \Delta a_3), \end{aligned}$$

where  $\hat{a}_2(\Delta)$  and  $\hat{a}_3(\Delta)$  are now the vectors of dimension  $l < k_2$  and  $\Delta$  is  $k_2$ -by- $l$ .  $\Delta$  either picks out the vector  $(\beta'_1, \beta'_2)'$  or the scalar  $\beta_3$ .

Using the notation from Section 5, the asymptotic variance of  $(\hat{a}_1, \hat{a}_2(\Delta), \hat{a}_3(\Delta))$  is

$$\begin{pmatrix} Q_1 & 0 & 0 \\ \Delta'R_1 & \Delta'R_2\Delta & 0 \\ 0 & 0 & \Delta'R_2\Delta \end{pmatrix}^{-1} \begin{pmatrix} V_{11} & V_{12}\Delta & V_{12}\Delta \\ \Delta'V'_{12} & \Delta'V_{22}\Delta & \Delta'V_{22}\Delta \\ \Delta'V'_{12} & \Delta'V_{22}\Delta & \Delta'V_{22}\Delta \end{pmatrix} \begin{pmatrix} Q_1 & R'_1\Delta & 0 \\ 0 & \Delta'R_2\Delta & 0 \\ 0 & 0 & \Delta'R_2\Delta \end{pmatrix}^{-1}.$$

Using the expression for partitioned inverse and multiplying out gives a matrix with nine blocks. The second and third blocks in the first row of blocks are  $-Q_1^{-1}V_{11}Q_1^{-1} \times R'_1\Delta(\Delta'R_2\Delta)^{-1} + Q_1^{-1}V_{12}\Delta(\Delta'R_2\Delta)^{-1}$  and  $Q_1^{-1}V_{12}\Delta(\Delta'R_2\Delta)^{-1}$ , respectively. With  $R_2$  and  $Q_1$  known and  $\Delta = (I_{l \times l} : 0_{l \times k_2})'$ , the block  $Q_1^{-1}V_{12}\Delta(\Delta'R_2\Delta)^{-1}$  identifies  $\Delta V'_{12}$  which is the first  $l$  row of  $V'_{12}$  (and hence the first  $l$  columns of  $V_{12}$ ). The difference between the last two blocks in the top row of blocks is  $-Q_1^{-1}V_{11}Q_1^{-1}R'_1\Delta(\Delta'R_2\Delta)^{-1}$ . This identifies  $R'_1\Delta$ , which is the first  $l$  columns of  $R'_1$ .

REFERENCES

ALTONJI, J. G., A. A. SMITH, AND I. VIDANGOS (2013): "Modeling Earnings Dynamics," *Econometrica*, 81 (4), 1395–1454. [1290]

ANDREWS, D. W. K. (2002): "Higher-Order Improvements of a Computationally Attractive k-Step Bootstrap for Extremum Estimators," *Econometrica*, 70 (1), 119–162. [1278]

ANDREWS, D. W. K., AND M. BUCHINSKY (2001): "Evaluation of a Three-Step Method for Choosing the Number of Bootstrap Repetitions," *Journal of Econometrics*, 103 (1–2), 345–386. [1277]

ARMSTRONG, T. B., M. BERTANHA, AND H. HONG (2014): "A Fast Resample Method for Parametric and Semiparametric Models," *Journal of Econometrics*, 179 (2), 128–133. [1278]

CHERNOZHUKOV, V., AND H. HONG (2003): "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 115 (2), 293–346. [1278]

DAVIDSON, R., AND J. G. MACKINNON (1999): "Bootstrap Testing in Nonlinear Models," *International Economic Review*, 40 (2), 487–508. [1277]

DIX-CARNEIRO, R. (2014): "Trade Liberalization and Labor Market Dynamics," *Econometrica*, 82 (3), 825–885. [1290]

GOURIEROUX, C., AND A. MONFORT (2007): *Simulation-Based Econometric Methods*. Oxford Scholarship Online Monographs. [1291]

HAHN, J. (1996): "A Note on Bootstrapping Generalized Method of Moments Estimators," *Econometric Theory*, 12 (1), 187–197. [1278,1279,1282,1296,1297]

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50 (4), 1029–1054. [1281]



- HEAGERTY, P. J., AND T. LUMLEY (2000): "Window Subsampling of Estimating Functions With Application to Regression Models," *Journal of the American Statistical Association*, 95 (449), 197–211. [1278]
- HECKMAN, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47 (1), 153–161. [1286, 1288]
- HELPMAN, E., M. MELITZ, AND Y. RUBINSTEIN (2008): "Estimating Trade Flows: Trading Partners and Trading Volumes," *Quarterly Journal of Economics*, 123, 441–487. [1279,1288,1292,1293,1300]
- HONG, H., AND O. SCAILLET (2006): "A Fast Subsampling Method for Nonlinear Dynamic Models," *Journal of Econometrics*, 133 (2), 557–578. [1278]
- HONORÉ, B. E., AND L. HU (2015): "Simpler Bootstrap Estimation of the Asymptotic Variance of U-Statistic Based Estimators," Working Paper Series WP-2015-7, Federal Reserve Bank of Chicago. [1285]
- (2017): "Supplement to 'Poor (Wo)man's Bootstrap'," *Econometrica Supplemental Material*, 85, <http://dx.doi.org/10.3982/ECTA13465>. [1287]
- KLEIN, R. W., AND R. H. SPADY (1993): "An Efficient Semiparametric Estimator for Binary Response Models," *Econometrica*, 61 (2), 387–421. [1288]
- KLINE, P., AND A. SANTOS (2012): "A Score Based Approach to Wild Bootstrap Inference," *Journal of Econometric Methods*, 1 (1), 23–41. [1278]
- KORMILTSINA, A., AND D. NEKIPELOV (2012): "Approximation Properties of Laplace-Type Estimators," in *DSGE Models in Macroeconomics: Estimation, Evaluation, and New Developments*. Advances in Econometrics, Vol. 28, ed. by N. Balke, F. Canova, F. Milani, and M. Wynne. Emerald Group Publishing. [1278]
- LEE, D., AND K. I. WOLPIN (2006): "Intersectoral Labor Mobility and the Growth of the Service Sector," *Econometrica*, 74 (1), 1–46. [1290]
- NEWWEY, W. K., AND D. MCFADDEN (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*. Handbooks in Economics, ed. by R. F. Engle and D. L. McFadden. Amsterdam: Elsevier, 2111–2245. [1281]
- PAKES, A., AND D. POLLARD (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–1057. [1296]
- POWELL, J. L. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–325. [1279]
- (1987): "Semiparametric Estimation of Bivariate Latent Models," Working Paper 8704, Social Systems Research Institute, University of Wisconsin–Madison. [1288]
- SANTOS SILVA, J. M. C., AND S. TENREYRO (2015): "Trading Partners and Trading Volumes: Implementing the Helpman–Melitz–Rubinstein Model Empirically," *Oxford Bulletin of Economics and Statistics*, 77 (1), 93–105. [1293]

---

Co-editor Elie Tamer handled this manuscript.

*Manuscript received 15 May, 2015; final version accepted 17 January, 2017; available online 13 April, 2017.*