



Federal Reserve Bank of Chicago

**Poor (Wo)man's Bootstrap**

*Bo E. Honoré and Luojia Hu*

REVISED  
April 2016

WP 2015-01

# Poor (Wo)man's Bootstrap\*

Bo E. Honoré<sup>†</sup>

Luojia Hu<sup>‡</sup>

April 2016

## Abstract

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.

JEL CODE: C10, C18, C15.

---

\*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. The most recent version of this paper can be found at <http://www.princeton.edu/~honore/papers/PoorWomansBootstrap.pdf>.

<sup>†</sup>Mailing Address: Department of Economics, Princeton University, Princeton, NJ 08544-1021. Email: [honore@Princeton.edu](mailto:honore@Princeton.edu).

<sup>‡</sup>Mailing Address: Economic Research Department, Federal Reserve Bank of Chicago, 230 S. La Salle Street, Chicago, IL 60604. Email: [luh@frbchi.org](mailto:luh@frbchi.org).

# 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 Davidson and MacKinnon (1999), Heagerty and Lumley (2000), Hong and Scaillet (2006) and Kline and Santos (2012). Unfortunately, their approaches assume that one can easily estimate the “Hessian” in the sandwich form of the asymptotic variance of the estimator. It is the difficulty of doing this that is the main motivation for this paper. Part of the contribution of Chernozhukov and Hong (2003) is also to provide an alternative way to do inference without estimating asymptotic variances from their analytical expressions. However, Kormiltsina and Nekipelov (2012) point 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 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.

Following the discussion of M-estimators and GMM-estimators, we illustrate our approach 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 perform a small Monte Carlo of the approach applied to an indirect inference estimator 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.

Section 4 shows that an alternative, and even simpler approach can be applied to method of moments estimators. In Section 5, 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 6, 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 this in Section 7 by applying our approach to a two-step estimator of a sample selection model inspired by Helpman, Melitz, and Rubinstein (2008).

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 feasible.

## 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

$$avar(\hat{\theta}) = H^{-1}VH^{-1}$$

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  $\text{avar}(\widehat{\theta})$  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  $\text{avar}(\widehat{\theta})$ .

It will be useful to explicitly write

$$H = \begin{pmatrix} h_{11} & h_{12} & \cdots & h_{1k} \\ h_{12} & h_{22} & \cdots & h_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ h_{1k} & h_{2k} & \cdots & h_{kk} \end{pmatrix} \quad \text{and} \quad V = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1k} \\ v_{12} & v_{22} & \cdots & v_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ v_{1k} & v_{2k} & \cdots & v_{kk} \end{pmatrix}.$$

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

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

where  $\delta$  is a fixed vector.

The (nonparametric) bootstrap equivalent of (1) is

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

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  $\widehat{\theta}$  and its bootstrap equivalent. Our approach will therefore be to estimate the joint asymptotic variance of  $\widehat{a}(\delta)$  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 1, we mimic the arguments in Hahn (1996) and prove that the joint bootstrap distribution of the estimators  $\widehat{a}(\delta)$  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).

It is easiest to illustrate why this works by considering a case where  $\theta$  is two-dimensional. For this case, 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

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

With  $\delta_1 = (1, 0)$  and  $\delta_2 = (0, 1)$ , we have

$$\Omega_{(1,0),(0,1)} = \begin{pmatrix} h_{11}^{-2} v_{11} & h_{11}^{-1} v_{12} h_{22}^{-1} \\ h_{11}^{-1} v_{12} h_{22}^{-1} & h_{22}^{-2} v_{22} \end{pmatrix}.$$

So the correlation in  $\Omega_{(1,0),(0,1)}$  gives the correlation in  $V$ . We also 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  $v_{11} = 1$ .

This gives

$$V = \begin{pmatrix} 1 & \rho v \\ \rho v & v^2 \end{pmatrix}, \quad v > 0$$

where we have already noted that  $\rho$  is identified from the correlation between  $\widehat{a}(\delta_1)$  and  $\widehat{a}(\delta_2)$ . We now argue that one can also identify  $v$ ,  $h_{11}$ ,  $h_{12}$  and  $h_{22}$ .

In the following,  $k_j$  will be used to denote objects that are identified from  $\Omega_{\delta_1, \delta_2}$  for various choices of  $\delta_1$  and  $\delta_2$ . We use  $e_j$  to denote a vector that has 1 in its  $j$ 'th element and zeros elsewhere.

We first consider  $\delta_1 = e_1$  and  $\delta_2 = e_2$  and we then have

$$\Omega_{(1,0),(0,1)} = \begin{pmatrix} h_{11}^{-2} & \rho v h_{22}^{-1} h_{11}^{-1} \\ \rho v h_{22}^{-1} h_{11}^{-1} & h_{22}^{-2} v^2 \end{pmatrix}$$

so we know  $k_1 = \frac{v}{h_{22}}$ . We also know  $h_{11}$ .

Now also consider a third estimator based on  $\delta_3 = e_1 + e_2$ . We have

$$\Omega_{(1,0),(1,1)} = \begin{pmatrix} h_{11}^{-2} & h_{11}^{-1} (1 + \rho v) (h_{11} + 2h_{12} + h_{22})^{-1} \\ h_{11}^{-1} (1 + \rho v) (h_{11} + 2h_{12} + h_{22})^{-1} & (1 + 2\rho v + v^2) (h_{11} + 2h_{12} + h_{22})^{-2} \end{pmatrix}.$$

The upper right-hand corner of this is

$$k_2 = h_{11}^{-1} (1 + \rho v) (h_{11} + 2h_{12} + h_{22})^{-1}.$$

Using  $v = k_1 h_{22}$  yields a linear equation in the unknowns,  $h_{12}$  and  $h_{22}$ ,

$$k_2 h_{11} (h_{11} + 2h_{12} + h_{22}) = (1 + \rho k_1 h_{22}). \quad (4)$$

Now consider the covariance between the estimators based on  $e_1$  and a fourth estimator based on  $e_1 - e_2$ ; in other words, consider the upper right-hand corner of  $\Omega_{(1,0),(1,-1)}$ :

$$k_3 = h_{11}^{-1} (1 - \rho v) (h_{11} - 2h_{12} + h_{22})^{-1}.$$

We rewrite this as a linear equation in  $h_{12}$  and  $h_{22}$ ,

$$k_3 h_{11} (h_{11} - 2h_{12} + h_{22}) = (1 - \rho k_1 h_{22}). \quad (5)$$

Rewriting (4) and (5) in matrix form, we get

$$\begin{pmatrix} 2k_2 h_{11} & k_2 h_{11} - \rho k_1 \\ -2k_3 h_{11} & k_3 h_{11} + \rho k_1 \end{pmatrix} \begin{pmatrix} h_{12} \\ h_{22} \end{pmatrix} = \begin{pmatrix} 1 - k_2 h_{11}^2 \\ 1 - k_3 h_{11}^2 \end{pmatrix}. \quad (6)$$

Appendix 2 shows that the determinant of the matrix on the left is positive definite. As a result, the two equations, (4) and (5), always have a unique solution for  $h_{12}$  and  $h_{22}$ . Once we have  $h_{22}$ , we then get the remaining unknown,  $v$ , from  $v = k_1 h_{22}$ .

The identification result for the two-dimensional case carries over to the general case in a straightforward manner. For each pair of elements of  $\theta$ ,  $\theta_j$  and  $\theta_\ell$ , the corresponding elements of  $H$  and  $V$  can be identified as above, subject to the normalization that one of the diagonal elements of  $V$  is 1. This yields  $\frac{v_{\ell\ell}}{v_{jj}}$ ,  $\frac{v_{j\ell}}{v_j}$ , and all the elements 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.

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

Specifically, define

$$D = \begin{pmatrix} \delta_1 & \delta_2 & \dots & \delta_m \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} \delta_1 & 0 & \dots & 0 \\ 0 & \delta_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \delta_m \end{pmatrix}. \quad (7)$$



The covariance matrix for the  $m$  estimators is then

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

which implies that

$$(C' (I \otimes H) C) \Omega (C' (I \otimes H) C) = (D'VD). \quad (8)$$

These need to be solved for the symmetric and positive definite matrices  $V$  and  $H$ . The calculation 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  $e_j - e_\ell$ .

There are 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

$$\min_{V,H} \sum_{j\ell} \left( \left\{ (C' (I \otimes H) C) \widehat{\Omega} (C' (I \otimes H) C) - (D'VD) \right\}_{j\ell} \right)^2 \quad (9)$$

where  $D$  and  $C$  are defined in (7),  $V_{11} = 1$ , 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

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

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

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

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

$$\Sigma = (\Gamma'W\Gamma)^{-1} \Gamma'W S W \Gamma (\Gamma'W\Gamma)^{-1} \quad (10)$$

where  $W$  is the probability limit of  $W_n$ ,  $S = V[f(x_i, \theta)]$  and  $\Gamma = \frac{\partial}{\partial \theta'} E[f(x_i, \theta)]$ . 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

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

by

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

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

$$\Omega_{\delta_1, \delta_2} = \text{avar} \left( \begin{pmatrix} \hat{a}(\delta_1) \\ \hat{a}(\delta_2) \end{pmatrix} \right) = \begin{pmatrix} (\delta_1' \Gamma' W \Gamma \delta_1)^{-1} \delta_1' \Gamma' W S W \Gamma \delta_1 (\delta_1' \Gamma' W \Gamma \delta_1)^{-1} & (\delta_1' \Gamma' W \Gamma \delta_1)^{-1} \delta_1' \Gamma' W S W \Gamma \delta_2 (\delta_2' \Gamma' W \Gamma \delta_2)^{-1} \\ (\delta_1' \Gamma' W \Gamma \delta_1)^{-1} \delta_1' \Gamma' W S W \Gamma \delta_2 (\delta_2' \Gamma' W \Gamma \delta_2)^{-1} & (\delta_2' \Gamma' W \Gamma \delta_2)^{-1} \delta_2' \Gamma' W S W \Gamma \delta_2 (\delta_2' \Gamma' W \Gamma \delta_2)^{-1} \end{pmatrix}. \quad (11)$$

Of course (11) has exactly the same structure as (3) and we can therefore back out the matrices  $\Gamma' W \Gamma$  and  $\Gamma' W S W \Gamma$  (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 proved in Appendix 1. The result stated there is a minor modification of the result in Hahn (1996).

### 3 Illustrations of Main Idea

#### 3.1 Linear Regression

There are few reasons why one would want 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 9-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  $j = 1 \cdots 3$ ,  $x_{ij} = \tilde{x}_{ij} + 1$  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 (9) by nonlinear 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 1 and 2 report the mean absolute differences in these test statistics for sample sizes of 200 and 2,000, respectively.

Tables 1 and 2 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).

## 3.2 Structural Model

The method proposed here should be especially useful when estimating nonlinear structural models such as Lee and Wolpin (2006), Altonji, Smith, and Vidangos (2013) and Dix-Carneiro (2014). To illustrate its usefulness in such a situation, we consider a very simple two-period Roy model like the one studied in Honoré and de Paula (2015).

There are two sectors, labeled one and two. A worker is endowed with a vector of sector-specific human capital,  $x_{si}$ , and sector-specific income in period one is

$$\log(w_{si1}) = x'_{si}\beta_s + \varepsilon_{si1}$$

and sector-specific income in period two is

$$\log(w_{s_{i2}}) = x'_{s_i} \beta_s + 1 \{d_{i1} = s\} \gamma_s + \varepsilon_{s_{i2}}$$

where  $d_{i1}$  is the sector chosen in period one. 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  $w_{1i2} > w_{2i2}$ , i.e., if

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

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

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

and sector 2 otherwise.

In Appendix 3, we demonstrate that the expected value of the maximum of two dependent lognormally distributed random variables with means  $(\mu_1, \mu_2)'$  and variance  $\begin{pmatrix} \sigma_1^2 & \tau\sigma_1\sigma_2 \\ \tau\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$  is

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

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 these are  $\beta_1 = (1, 1)'$ ,  $\beta_2 = (\frac{1}{2}, 1)'$ ,  $\gamma_1 = 0$  and  $\gamma_2 = 1$ . Finally,  $\sigma_1^2 = 2$ ,  $\sigma_2^2 = 3$ ,  $\tau = 0$  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  $d_{i0} = 0$ )

- 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.
- 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.
- 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  $\hat{V}[\hat{\alpha}]$  be the associated estimated variance. For a candidate vector of structural parameters,  $\theta$ , the researcher simulates the model  $R$  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

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

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 3 and 4 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 woman’s bootstrap approximate the standard deviation of the estimator well (Table 3). 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 (9) 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 4 illuminates 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 the true asymptotic standard errors.

## 4 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 methods 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 play no role for the asymptotic distribution in the just identified case, (10) 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.

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

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

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

---

<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.

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

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

where  $s_{pj}$  and  $\gamma_{jp}$  denote the elements in  $S$  and  $\Gamma$ , respectively. In particular

$$Avar(\widehat{a}_{pp}) = \frac{s_{pp}}{\gamma_{pp}^2} = s_{pp}.$$

Hence  $s_{pp}$  is identified. Since

$$Acov(\widehat{a}_{pp}, \widehat{a}_{jj}) = \frac{s_{pj}}{\gamma_{pp}\gamma_{jj}} = s_{pj},$$

$s_{pj}$  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

$$Acov(\widehat{a}_{pp}, \widehat{a}_{jm}) = \frac{s_{pj}}{\gamma_{pp}\gamma_{jm}} = \frac{s_{pj}}{\gamma_{jm}}.$$

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

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

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

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

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

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

## 5 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  $k^4 + k^2$  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,

---

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



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 two without sacrificing identification (including identification of  $V$ ). In the previous section, we considered estimators in the directions  $e_j$  ( $j = 1, \dots, k$ ),  $e_j + e_\ell$  ( $\ell \neq j$ ) and  $e_j - e_\ell$  ( $\ell \neq j$ ). Here we consider only estimators in the directions  $e_j$  ( $j = 1, \dots, k$ ),  $e_j + e_\ell$  ( $\ell \neq j$ ) and  $e_j - e_1$ .

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  $\hat{a}(e_j)$  and  $\hat{a}(e_\ell)$  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  $\hat{a}(e_1)$  and  $\hat{a}(e_j + e_\ell)$

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

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.

## 5.1 Simplification When Information Equality Holds

Efficient Generalized Method of Moments estimation in Section 2.2 implies that  $(\Gamma'WT) = \Gamma'WSWT$  and maximum likelihood estimation in Section 2.1 implies that  $H = V$ . Either way, the asymptotic variance of the estimator reduces to<sup>3</sup>  $H^{-1}$  while the asymptotic variance of the  $k$  one-dimensional estimators in the directions  $e_1, \dots, e_k, \hat{a}(e_1), \dots, \hat{a}(e_k)$ , is

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

---

<sup>3</sup>In the case of a GMM estimator, define  $H$  to equal  $(\Gamma'WT)$ .

(see equations (3) and (11)). 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

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

As a result, it is possible to estimate the variance of the parameter of interest by bootstraping only  $k$  one-dimensional estimators.

## 5.2 Exploiting Specific Model Structures

It is also sometimes possible to reduce the computational burden by exploiting specific properties of the estimator of interest. For example, it is sometimes the case that a subvector can be easily estimated if one holds the remaining parts of the parameter vector fixed. Regression models of the type  $y_i = \beta_0 + x_{i1}^{\alpha_1} \beta_1 + x_{i2}^{\alpha_2} \beta_2 + \varepsilon_i$  is 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 7 is yet another example where some parameters are easy to estimate for given values of the remaining parameters.

To explore the benefits of such 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  $\widetilde{\theta}_j = (\alpha', \beta_j)'$ . 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 one, 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 positive definite solution for  $X$ . In other words, one can back out all of  $H_{\beta\beta}$ . Of course, when  $\widehat{\beta}$  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.

## 6 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

$$\begin{aligned} d_i &= 1 \{z_i'\alpha + \nu_i \geq 0\} \\ y_i &= d_i \cdot (x_i'\beta + \varepsilon_i) \end{aligned}$$

where  $(\nu_i, \varepsilon_i)$  has a bivariate normal distribution.  $\alpha$  can be estimated by the probit maximum likelihood estimator,  $\widehat{\alpha}_{MLE}$ , 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  $\lambda_i = \frac{\phi(z_i'\widehat{\alpha}_{MLE})}{1 - \Phi(z_i'\widehat{\alpha}_{MLE})}$  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 4 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 4 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

$$\begin{aligned}\hat{\theta}_1 &= \arg \min_{\tau_1} \frac{1}{n} \sum Q(z_i, \tau_1) \\ \hat{\theta}_2 &= \arg \min_{\tau_2} \frac{1}{n} \sum R(z_i, \hat{\theta}_1, \tau_2)\end{aligned}\tag{12}$$

with limiting first-order conditions

$$\begin{aligned}E[q(z_i, \theta_1)] &= 0 \\ E[r(z_i, \theta_1, \theta_2)] &= 0\end{aligned}$$

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

$$\begin{pmatrix} Q_1 & 0 \\ R_1 & R_2 \end{pmatrix}^{-1} \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \left( \begin{pmatrix} Q_1 & 0 \\ R_1 & R_2 \end{pmatrix}^{-1} \right)'\tag{13}$$

where  $\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} = \text{var} \left( \begin{pmatrix} q(z_i, \theta_1) \\ r(z_i, \theta_1, \theta_2) \end{pmatrix} \right)$ ,  $Q_1 = E \left[ \frac{\partial q(z_i, \theta_1)}{\partial \theta_1} \right]$ ,  $R_1 = E \left[ \frac{\partial r(z_i, \theta_1, \theta_2)}{\partial \theta_1} \right]$  and  $R_2 = E \left[ \frac{\partial r(z_i, \theta_1, \theta_2)}{\partial \theta_2} \right]$ . 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 woman's bootstrap procedure above.

The asymptotic variance of  $\widehat{\theta}_2$  is

$$R_2^{-1}R_1Q_1^{-1}V_{11}Q_1^{-1}R_1'R_2^{-1} - R_2^{-1}V_{21}Q_1^{-1}R_1'R_2^{-1} - R_2^{-1}R_1Q_1^{-1}V_{12}R_2^{-1} + R_2^{-1}V_{22}R_2^{-1} \quad (14)$$

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

As mentioned, (13) 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  $\widehat{\theta} + a\delta$ . 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.

In Appendix 4, we show that it is possible to modify the poor (wo)man's bootstrap so it can be applied to two-step estimators using only one-dimensional estimators that are defined by only one of the two original objective functions.

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  $\widehat{\theta}_1$  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

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

for fixed  $\delta_1$ ,  $\delta_2$  and  $\delta_3$ . In Appendix 4, 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  $(\hat{a}_1(e_j), \hat{a}_2(e_j, e_m), \hat{a}_3(e_m))$  for  $j = 1, \dots, k_1$  and  $m = 1, \dots, k_2$ , obtained from

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

These  $B$  draws can be used to estimate the variance-covariance matrix of  $(\hat{a}_1(e_j), \hat{a}_2(e_j, e_m), \hat{a}_3(e_m))$  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 (1979)'s two-step estimator is a linear regression, while the first is maximum likelihood. Similarly, the second step in Powell (1987)'s 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 Roger W. Klein (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 the entire parameter vector in the easier step in each bootstrap sample. The next two subsections develop this idea. In both cases, it turns out that the correction to the variance for  $\hat{\theta}_2$  (the first three terms in (14)) 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.

## 6.1 Bootstrapping with Easy First-Step Estimator

We first consider a method which requires the first step estimator to be recalculated in each bootstrap replication, but which only requires calculation of  $2k_2$  one-dimensional estimators

in the second step. As above, we present the ideas in terms of the variance-covariance matrix of infeasible estimators that require knowledge of the true parameter vectors. The approach can then be made feasible by replacing these true values with the original sample estimates when performing the bootstrap.

Consider again three estimators of the form

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

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

$$\begin{pmatrix} Q_1 & 0 & 0 \\ \delta' R_1 & \delta' R_2 \delta & 0 \\ 0 & 0 & \delta' R_2 \delta \end{pmatrix}^{-1} \begin{pmatrix} V_{11} & V_{12} \delta & V_{12} \delta \\ \delta' V'_{12} & \delta' V_{22} \delta & \delta' V_{22} \delta \\ \delta' V'_{12} & \delta' V_{22} \delta & \delta' V_{22} \delta \end{pmatrix} \begin{pmatrix} Q_1 & \delta' R_1 & 0 \\ 0 & \delta' R_2 \delta & 0 \\ 0 & 0 & \delta' R_2 \delta \end{pmatrix}^{-1} \quad (15)$$

where

$$\begin{pmatrix} Q_1 & 0 & 0 \\ \delta' R_1 & \delta' R_2 \delta & 0 \\ 0 & 0 & \delta' R_2 \delta \end{pmatrix}^{-1} = \begin{pmatrix} Q_1^{-1} & 0 & 0 \\ -(\delta' R_2 \delta)^{-1} \delta' R_1 Q_1^{-1} & (\delta' R_2 \delta)^{-1} & 0 \\ 0 & 0 & (\delta' R_2 \delta)^{-1} \end{pmatrix}.$$

Multiplying (15) 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  $Q_1^{-1} V_{12} \delta (\delta' R_2 \delta)^{-1}$ . 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 columns 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  $R_1$ .

## 6.2 Bootstrapping with Easy Second-Step Estimator

We next consider the case where the first step estimator is computationally challenging while it is feasible to recalculate the second step estimator in each bootstrap sample. We again consider estimators of the form

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

but now  $\widehat{a}_2$  is a vector of the same dimension as  $\theta_2$ . The asymptotic variance of  $(\widehat{a}_1(\delta), \widehat{a}_2(\delta), \widehat{a}_3)$  is

$$\begin{pmatrix} \delta' Q_1 \delta & 0 & 0 \\ R_1 \delta & R_2 & 0 \\ 0 & 0 & R_2 \end{pmatrix}^{-1} \begin{pmatrix} \delta' V_{11} \delta & \delta' V_{12} & \delta' V_{12} \\ V'_{12} \delta & V_{22} & V_{22} \\ V'_{12} \delta & V_{22} & V_{22} \end{pmatrix} \begin{pmatrix} \delta' Q_1 \delta & \delta' R'_1 & 0 \\ 0 & R_2 & 0 \\ 0 & 0 & R_2 \end{pmatrix}^{-1} \quad (16)$$

where

$$\begin{pmatrix} \delta' Q_1 \delta & 0 & 0 \\ R_1 \delta & R_2 & 0 \\ 0 & 0 & R_2 \end{pmatrix}^{-1} = \begin{pmatrix} (\delta' Q_1 \delta)^{-1} & 0 & 0 \\ -R_2^{-1} R_1 \delta (\delta' Q_1 \delta) & R_2^{-1} & 0 \\ 0 & 0 & R_2^{-1} \end{pmatrix}.$$

Multiplying (16) yields a matrix with nine blocks. The upper-middle block is  $-(\delta' Q_1 \delta)^{-1} (\delta' V_{11} \delta) (\delta' Q_1 \delta) \delta' R'_1 R_2^{-1} + (\delta' Q_1 \delta)^{-1} \delta' V_{12} R_2^{-1}$  while the upper-right block is  $(\delta' Q_1 \delta)^{-1} \delta' V_{12} R_2^{-1}$ . The latter identifies  $\delta' V_{12}$ . When  $\delta = e_j$  this is the  $j$ 'th row of  $V_{12}$ . The difference between the upper-middle block and upper-right block gives  $-(\delta' Q_1 \delta)^{-1} (\delta' V_{11} \delta)^{-1} (\delta' Q_1 \delta) \delta' R'_1 R_2^{-1}$  which in turn gives  $\delta' R'_1$  or  $R'_1 \delta$ . When  $\delta$  equals  $e_j$  this is the  $j$ 'th column of  $R_1$ .

This approach requires calculation of only  $2k_1$  one-dimensional estimators using the more difficult first step objective function. Moreover, as above, the approach gives closed form estimates of  $V_{12}$  and  $R_1$ .

## 6.3 Implementation

The two previous subsections identified  $R_1$  and  $V_{12}$  in closed forms using a subset of the information contained in the asymptotic variance of  $(\widehat{a}_1, \widehat{a}_2, \widehat{a}_3)$ . Here we present one way to



use all the components of this variance to estimate  $R_1$  and  $V_{12}$ . For simplicity, we consider the case where one recalculates the entire first step estimator in each bootstrap sample. This is the case considered in Section 6.1.

Consider estimating the second step parameter in  $J$  different directions in each bootstrap replication,

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

The asymptotic variance of  $(\hat{a}_1, \{\hat{a}_2(\delta_j)\}_{j=1}^J, \{\hat{a}_3(\delta_j)\}_{j=1}^J)$  is of the form  $\Omega = A^{-1}B(A')^{-1}$  where

$$A = \begin{pmatrix} Q_1 & 0 & 0 \\ D'R_1 & C'(I \otimes R_2)C & 0 \\ 0 & 0 & C'(I \otimes R_2)C \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} V_{11} & V_{12}D & V_{12}D \\ D'V'_{12} & D'V_{22}D & D'V_{22}D \\ D'V'_{12} & D'V_{22}D & D'V_{22}D \end{pmatrix}.$$

This gives

$$\begin{aligned}& \begin{pmatrix} V_{11} & V_{12}D & V_{12}D \\ D'V'_{12} & D'V_{22}D & D'V_{22}D \\ D'V'_{12} & D'V_{22}D & D'V_{22}D \end{pmatrix} & (17) \\ & = \begin{pmatrix} Q_1 & 0 & 0 \\ D'R_1 & C'(I \otimes R_2)C & 0 \\ 0 & 0 & C'(I \otimes R_2)C \end{pmatrix} \Omega \begin{pmatrix} Q_1 & R'_1D & 0 \\ 0 & C'(I \otimes R_2)C & 0 \\ 0 & 0 & C'(I \otimes R_2)C \end{pmatrix}.\end{aligned}$$

This suggests estimating  $V_{12}$  and  $R_1$  by minimizing

$$\sum_{i,j} \left( \left( \begin{array}{ccc} \widehat{V}_{11} & V_{12}D & V_{12}D \\ D'V'_{12} & D'\widehat{V}_{22}D & D'\widehat{V}_{22}D \\ D'V'_{12} & D'\widehat{V}_{22}D & D'\widehat{V}_{22}D \end{array} \right) - \left( \begin{array}{ccc} Q_1 & 0 & 0 \\ D'R_1 & C'(I \otimes \widehat{R}_2)C & 0 \\ 0 & 0 & C'(I \otimes \widehat{R}_2)C \end{array} \right) \widehat{\Omega} \left( \begin{array}{ccc} Q_1 & R'_1D & 0 \\ 0 & C'(I \otimes \widehat{R}_2)C & 0 \\ 0 & 0 & C'(I \otimes \widehat{R}_2)C \end{array} \right) \right)_{ij}^2$$

over  $V_{12}$  and  $R_1$ .

When  $\delta_j = e_j$ ,  $D = I$  and  $C'(I \otimes R_2)C = \text{diag}(R_2) \stackrel{\text{def}}{=} M$ . Using this and multiplying out the right hand side of (17) gives

$$\begin{pmatrix} V_{11} & V_{12} & V_{12} \\ V'_{12} & V_{22} & V_{22} \\ V'_{12} & V_{22} & V_{22} \end{pmatrix} = \begin{pmatrix} Q_1 & 0 & 0 \\ R_1 & M & 0 \\ 0 & 0 & M \end{pmatrix} \begin{pmatrix} \Omega_{11} & \Omega_{12} & \Omega_{13} \\ \Omega_{21} & \Omega_{22} & \Omega_{23} \\ \Omega_{31} & \Omega_{32} & \Omega_{33} \end{pmatrix} \begin{pmatrix} Q_1 & R'_1 & 0 \\ 0 & M & 0 \\ 0 & 0 & M \end{pmatrix} =$$

$$\begin{pmatrix} Q_1\Omega_{11}Q_1 & Q_1\Omega_{11}R'_1 + Q_1\Omega_{12}M & Q_1\Omega_{13}M \\ R_1\Omega_{11}Q_1 + M\Omega_{21}Q_1 & R_1\Omega_{11}R'_1 + M\Omega_{21}R'_1 + R_1\Omega_{12}M + M\Omega_{22}M & R_1\Omega_{13}M + M\Omega_{23}M \\ M\Omega_{13}Q_1 & M\Omega_{31}R'_1 + M\Omega_{32}M & M\Omega_{33}M \end{pmatrix}.$$

The approach in Section 6.1 uses the last two parts of the first row to identify  $V_{12}$  and  $R_1$ . The upper left and lower right hand corners are not informative about  $V_{12}$  or  $R_1$ . Moreover, the matrix is symmetric. All the remaining information is therefore contained in the last two parts of the second row.  $R_1$  enters the middle block nonlinearly, which leaves three blocks of equations that are linear in  $V_{12}$  and  $R_1$ :

$$V_{12} = Q_1\Omega_{11}R'_1 + Q_1\Omega_{12}M$$

$$V_{12} = Q_1\Omega_{13}M$$

$$V_{23} = R_1\Omega_{13}M + M\Omega_{23}M.$$

These overidentify  $V_{12}$  and  $R_1$ , but they could be combined through least squares.

## 7 Illustration of 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 in Section 6.1. 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 5.2. Appendix 5 gives the mathematical details for combining the insights in Sections 6.1 and 5.2. 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.

### 7.1 Model Specification

In one of their specifications, Helpman, Melitz, and Rubinstein (2008) use 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 one 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

$$x'\beta_1 + \lambda(-z'\hat{\gamma})\beta_2 + \log(\exp(\beta_3(\lambda(-z'\hat{\gamma}) + z'\hat{\gamma})) - 1) \quad (18)$$

where  $\hat{\gamma}$  is the first step probit estimator and  $\lambda(\cdot) = \frac{\varphi(\cdot)}{1-\Phi(\cdot)}$ . Since probit maximum likelihood estimator is based on maximizing a concave objective function, this is an example where the first step estimation of  $\gamma$  is computationally relatively easy as in Section 6.1. Moreover, the

second step has the feature discussed in Section 5.2, 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

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

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

$$y_i = \tilde{\beta}_0 x_0^\lambda + x_1' \tilde{\beta}_1 + \lambda (-z' \hat{\gamma}) \beta_2. \quad (19)$$

Like (18), equation (19) has one parameter that enters nonlinearly. As a result, the second step again has the feature discussed in Section 5.2.

Helpman, Melitz, and Rubinstein (2008) use a panel from 1980 to 1989 of the trade flows (exports) from each of 158 countries to the remaining 157 countries<sup>4</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

---

<sup>4</sup>See [HTTP://SCHOLAR.HARVARD.EDU/MELITZ/PUBLICATIONS](http://scholar.harvard.edu/melitz/publications).

pair are landlocked), (5) COLONY (indicating if one of the countries in the pair colonialized 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) WTO<sub>NONE</sub> and (12) WTO<sub>BOTH</sub> (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 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>5</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.

## 7.2 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  $3^2$ . This roughly matches the variance of the residuals in the second equation in the data generating process to the same variance in the data.

---

<sup>5</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.

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 does not correct for the two step estimation, (2) the asymptotic variance that does make that correction, (3) the poor (wo)man’s bootstrap and (4) the regular bootstrap<sup>6</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 5–7.

Table 5 reports the standard deviations of the parameters estimates across the 400 Monte Carlo replications in column 1. Columns 2 and 3 report the means of the estimated standard errors using the asymptotic expressions without and with correction for the two-step estimation. Columns 4 and 5 report the means of the standard errors estimated using the poor (wo)mans bootstrap and the regular bootstrap, respectively. The results for the year dummies and continent fixed effects are omitted. The bootstrap and the poor (wo)mans 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. The standard errors that do not correct for the two-step estimation are the worst in almost all cases. Table 6 report almost identical results for the medians of the estimated standard errors.

Table 7 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. All three are much closer to the nominal size of 20% than the test based on the asymptotic distribution without correction.

---

<sup>6</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.

## 8 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 re-calculate 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 not only be 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.

## References

- ALTONJI, J. G., A. A. SMITH, AND I. VIDANGOS (2013): “Modeling Earnings Dynamics,” *Econometrica*, 81(4), 1395–1454.
- ANDREWS, D. W., AND M. BUCHINSKY (2001): “Evaluation of a Three-Step Method for Choosing the Number of Bootstrap Repetitions,” *Journal of Econometrics*, 103(12), 345 – 386.
- CHERNOZHUKOV, V., AND H. HONG (2003): “An MCMC approach to classical estimation,” *Journal of Econometrics*, 115(2), 293–346.
- DAVIDSON, R., AND J. G. MACKINNON (1999): “Bootstrap Testing in Nonlinear Models,” *International Economic Review*, 40(2), 487–508.

- DIX-CARNEIRO, R. (2014): “Trade Liberalization and Labor Market Dynamics,” *Econometrica*, 82(3), 825–885.
- GOURIEROUX, C., AND A. MONFORT (2007): *Simulation-based Econometric Methods*. Oxford Scholarship Online Monographs.
- HAHN, J. (1996): “A Note on Bootstrapping Generalized Method of Moments Estimators,” *Econometric Theory*, 12(1), pp. 187–197.
- HANSEN, L. P. (1982): “Large Sample Properties of Generalized Method of Moments Estimators,” *Econometrica*, 50(4), pp. 1029–1054.
- 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), pp. 197–211.
- HECKMAN, J. J. (1979): “Sample Selection Bias as a Specification Error,” *Econometrica*, 47(1), 153–61.
- HELPMAN, E., M. MELITZ, AND Y. RUBINSTEIN (2008): “Estimating Trade Flows: Trading Partners and Trading Volumes,” *Quarterly Journal of Economics*, 123, 441–487.
- HONG, H., AND O. SCAILLET (2006): “A fast subsampling method for nonlinear dynamic models,” *Journal of Econometrics*, 133(2), 557 – 578.
- HONORÉ, B., AND A. DE PAULA (2015): “Identification in a Dynamic Roy Model,” *In Preparation*.
- 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.
- KLINE, P., AND A. SANTOS (2012): “A Score Based Approach to Wild Bootstrap Inference,” *Journal of Econometric Methods*, 1(1), 23–41.



- KORMILTSINA, A., AND D. NEKIPELOV (2012): “Approximation Properties of Laplace-Type Estimators,” in *DSGE Models in Macroeconomics: Estimation, Evaluation, and New Developments*, ed. by N. Balke, F. Canova, F. Milani, and M. Wynne, vol. 28 of *Advances in Econometrics*. Emerald Group Publishing.
- KOTZ, S., N. BALAKRISHNAN, AND N. JOHNSON (2000): *Continuous Multivariate Distributions, Models and Applications*, Continuous Multivariate Distributions. Wiley, second edn.
- LEE, D., AND K. I. WOLPIN (2006): “Intersectoral Labor Mobility and the Growth of the Service Sector,” *Econometrica*, 74(1), pp. 1–46.
- NEWBY, W. K., AND D. MCFADDEN (1994): “Large Sample Estimation and Hypothesis Testing,” in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, no. 4 in *Handbooks in Economics*, pp. 2111–2245. Elsevier, North-Holland, Amsterdam, London and New York.
- PAKES, A., AND D. POLLARD (1989): “Simulation and the Asymptotics of Optimization Estimators,” *Econometrica*, 57, 1027–1057.
- POWELL, J. L. (1984): “Least Absolute Deviations Estimation for the Censored Regression Model,” *Journal of Econometrics*, 25, 303–325.
- (1987): “Semiparametric Estimation of Bivariate Latent Models,” Working Paper no. 8704, Social Systems Research Institute, University of Wisconsin–Madison.
- ROGER W. KLEIN, R. H. S. (1993): “An Efficient Semiparametric Estimator for Binary Response Models,” *Econometrica*, 61(2), 387–421.
- 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.

## Appendix 1: 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. In this appendix, we establish 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) results. 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

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

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  $\{Z_i\}_{i=1}^n$ .

Hahn (1996) proved the following results.

**Proposition 1 (Hahn Proposition 1)** *Assume that*

- (i)  $\theta_0$  is the unique solution to  $G(t) = 0$ ;
- (ii)  $\{Z_i\}$  is an i.i.d. sequence of random vectors;

(iii)  $\inf_{|t-\theta_0|\geq\delta} |G(t)| > 0$  for all  $\delta > 0$

(iv)  $\sup_t |G_n(t) - G(t)| \rightarrow 0$  as  $n \rightarrow \infty$  a.s.;

(v)  $E[\sup_t |g(Z_i, t)|] < \infty$ ;

(vi)  $A_n = A + o_p(1)$  and  $\hat{A}_n = A + o_B(1)$  for some nonsingular and nonrandom matrix  $A$ ;  
and

(vii)  $|A_n G_n(\tau_n)| \leq o_p(1) + \inf_t |A_n G_n(t)|$  and  $|\hat{A}_n \hat{G}_n(\hat{\tau}_n)| \leq o_B(1) + \inf_t |\hat{A}_n \hat{G}_n(t)|$

Then  $\tau_n = \theta_0 + o_p(1)$  and  $\hat{\tau}_n = \theta_0 + o_B(1)$ .

**Theorem 1 (Hahn Theorem 1)** Assume that

(i) Conditions (i)-(vi) in Proposition 1 are satisfied;

(ii)  $|A_n G_n(\tau_n)| \leq o_p(n^{-1/2}) + \inf_t |A_n G_n(t)|$  and  $|\hat{A}_n \hat{G}_n(\hat{\tau}_n)| \leq o_B(n^{-1/2}) + \inf_t |\hat{A}_n \hat{G}_n(t)|$ ;

(iii)  $\lim_{t \rightarrow \theta_0} e(t, \theta_0) = 0$  where  $e(t, t') \equiv E[(g(Z_i, t) - g(Z_i, t'))^2]^{1/2}$ ;

(iv) for all  $\varepsilon > 0$ ,

$$\lim_{\delta \rightarrow 0} \limsup_{n \rightarrow \infty} P \left( \sup_{e(t, t') \leq \delta} |G_n(t) - G(t) - G_n(t') + G(t')| \geq n^{-1/2} \varepsilon \right) = 0;$$

(v)  $G(t)$  is differentiable at  $\theta_0$ , an interior point of the parameter space,  $\Theta$ , with derivative  $\Gamma$  with full rank; and

(vi)  $\{g(\cdot, t) : t \in \Theta\} \subset L_2(P)$  and  $\Theta$  is totally bounded under  $e(\cdot, \cdot)$ .

Then

$$n^{1/2}(\tau_n - \theta_0) = -n^{-1/2}(\Gamma' A' A \Gamma)^{-1} \Gamma' A' A_n G_n(\theta_0) + o_p(1) \implies N(0, \Omega)$$

and

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

where

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

and

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

Our paper 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  $\theta_0' = (\theta_0^1, \theta_0^2)$ .

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

$$\gamma_n = \arg \min_t \left| A_n G_n \left( \begin{pmatrix} t \\ \theta_0^2 \end{pmatrix} \right) \right| \quad (20)$$

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

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

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

$$\hat{\gamma}_n = \arg \min_t \left| \hat{A}_n \hat{G}_n \left( \begin{pmatrix} t \\ \hat{\theta}^2 \end{pmatrix} \right) \right| \quad (21)$$

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 2 (Adaption of Hahn's (1996) Proposition 1)** *Suppose that the conditions in Proposition 1 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  $\hat{\gamma}_n = \theta_0^1 + o_B(1)$ .*

**Proof.** As in Hahn (1996), the proof follows from standard arguments. The only difference is that we need

$$\sup_t \left| \hat{G}_n \left( \begin{pmatrix} t \\ \hat{\theta}^2 \end{pmatrix} \right) - G \left( \begin{pmatrix} t \\ \theta_0^2 \end{pmatrix} \right) \right| = o_p^\omega(1)$$

This follows from

$$\begin{aligned} & \left| \hat{G}_n \left( \begin{pmatrix} t \\ \hat{\theta}^2 \end{pmatrix} \right) - G \left( \begin{pmatrix} t \\ \theta_0^2 \end{pmatrix} \right) \right| \\ &= \left| \hat{G}_n \left( \begin{pmatrix} t \\ \hat{\theta}^2 \end{pmatrix} \right) - G \left( \begin{pmatrix} t \\ \hat{\theta}^2 \end{pmatrix} \right) + G \left( \begin{pmatrix} t \\ \hat{\theta}^2 \end{pmatrix} \right) - G \left( \begin{pmatrix} t \\ \theta_0^2 \end{pmatrix} \right) \right| \\ &\leq \left| \hat{G}_n \left( \begin{pmatrix} t \\ \hat{\theta}^2 \end{pmatrix} \right) - G \left( \begin{pmatrix} t \\ \hat{\theta}^2 \end{pmatrix} \right) \right| + \left| G \left( \begin{pmatrix} t \\ \hat{\theta}^2 \end{pmatrix} \right) - G \left( \begin{pmatrix} t \\ \theta_0^2 \end{pmatrix} \right) \right| \end{aligned}$$

As in Hahn (1996), the first part is  $o_p^\omega(1)$  by bootstrap uniform convergence. The second part is bounded by  $\sup \left| \frac{\partial G(t_1, t_2)}{\partial t_2} \right| \left| \hat{\theta}^2 - \theta_0^2 \right|$ . This is  $O_p(\hat{\theta}^2 - \theta_0^2) = O_p(n^{-1/2})$  by the assumptions that  $G$  is continuously differentiable and that the parameter space is compact.

■

**Theorem 2 (Adaption of Hahn's (1996) Theorem 1)** *Assume that the conditions in Proposition 2 and Theorem 1 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

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

and

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

**Proof.** We start by showing that  $\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}$  is  $\sqrt{n}$ -consistent, and then move on to show asymptotic normality.

**Part 1.  $\sqrt{n}$ -consistency.** For  $\widehat{\theta}^2$  root- $n$  consistency follows from Pakes and Pollard (1989). Following Hahn (1996), we start with the observation that

$$\begin{aligned}
& \left| \widehat{A}_n \widehat{G}_n \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) - AG \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) - \widehat{A}_n \widehat{G}_n(\theta_0) + AG(\theta_0) \right| \\
& \leq \left| \widehat{A}_n \right| \left| \widehat{G}_n \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) - G \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) - \widehat{G}_n(\theta_0) + G(\theta_0) \right| + \left| \widehat{A}_n - A \right| \left| G \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) - G(\theta_0) \right| \\
& \leq o_B(n^{-1/2}) + o_B(1) \left| G \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) - G(\theta_0) \right| \tag{22}
\end{aligned}$$

Combining this with the triangular inequality we have

$$\begin{aligned}
\left| AG \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) - AG(\theta_0) \right| & \leq \left| \widehat{A}_n \widehat{G}_n \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) - AG \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) - \widehat{A}_n \widehat{G}_n(\theta_0) + AG(\theta_0) \right| \\
& \quad + \left| \widehat{A}_n \widehat{G}_n \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) - \widehat{A}_n \widehat{G}_n(\theta_0) \right| \\
& \leq o_B(n^{-1/2}) + o_B(1) \left| G \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) - G(\theta_0) \right| \\
& \quad + \left| \widehat{A}_n \widehat{G}_n \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) - \widehat{A}_n \widehat{G}_n(\theta_0) \right| \tag{23}
\end{aligned}$$

The nonsingularity of  $A$  implies the existence of a constant  $C_1 > 0$  such that  $|Ax| \geq C_1|x|$  for all  $x$ . Applying this fact to the left-hand side of (23) and collecting the  $G \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) - G(\theta_0)$  terms yield

$$(C_1 - o_B(1)) \left| G \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) - G(\theta_0) \right| \tag{24}$$

$$\begin{aligned}
& \leq o_B(n^{-1/2}) + \left| \widehat{A}_n \widehat{G}_n \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) - \widehat{A}_n \widehat{G}_n(\theta_0) \right| \\
& \leq o_B(n^{-1/2}) + \left| \widehat{A}_n \widehat{G}_n \left( \begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix} \right) \right| + \left| \widehat{A}_n \widehat{G}_n(\theta_0) \right| \\
& \leq o_B(n^{-1/2}) + \left| \widehat{A}_n \widehat{G}_n \left( \begin{pmatrix} \theta_0^1 \\ \widehat{\theta}^2 \end{pmatrix} \right) \right| + \left| \widehat{A}_n \widehat{G}_n(\theta_0) \right| \tag{25}
\end{aligned}$$

Stochastic equicontinuity implies that

$$\widehat{A}_n \widehat{G}_n \left( \left( \begin{array}{c} \theta_0^1 \\ \widehat{\theta}^2 \end{array} \right) \right) = \widehat{A}_n \left( G \left( \left( \begin{array}{c} \theta_0^1 \\ \widehat{\theta}^2 \end{array} \right) \right) - G(\theta_0) \right) + \widehat{A}_n \widehat{G}_n(\theta_0) + \widehat{A}_n o_B(n^{-1/2})$$

or

$$\left| \widehat{A}_n \widehat{G}_n \left( \left( \begin{array}{c} \theta_0^1 \\ \widehat{\theta}^2 \end{array} \right) \right) \right| \leq \left| \widehat{A}_n \left( G \left( \left( \begin{array}{c} \theta_0^1 \\ \widehat{\theta}^2 \end{array} \right) \right) - G(\theta_0) \right) \right| + \left| \widehat{A}_n \widehat{G}_n(\theta_0) \right| + \left| \widehat{A}_n \right| o_B(n^{-1/2})$$

so (25) implies

$$\begin{aligned} & (C_1 - o_B(1)) \left| G \left( \left( \begin{array}{c} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{array} \right) \right) - G(\theta_0) \right| \\ & \leq o_B(n^{-1/2}) + \left| \widehat{A}_n \left( G \left( \left( \begin{array}{c} \theta_0^1 \\ \widehat{\theta}^2 \end{array} \right) \right) - G(\theta_0) \right) \right| + 2 \left| \widehat{A}_n \right| \left| \widehat{G}_n(\theta_0) \right| + \left| \widehat{A}_n \right| o_B(n^{-1/2}) \\ & \leq o_B(n^{-1/2}) + \left| \widehat{A}_n \right| \left| G \left( \left( \begin{array}{c} \theta_0^1 \\ \widehat{\theta}^2 \end{array} \right) \right) - G(\theta_0) \right| + 2 \left| \widehat{A}_n \right| \left| \widehat{G}_n(\theta_0) - G_n(\theta_0) \right| \\ & \quad + 2 \left| \widehat{A}_n \right| \left| G_n(\theta_0) \right| + \left| \widehat{A}_n \right| o_B(n^{-1/2}) \\ & = o_B(n^{-1/2}) + O_B(1) O_p(n^{-1/2}) + O_B(1) O_B(n^{-1/2}) \\ & \quad + O_B(1) O_p(n^{-1/2}) + O_B(1) o_B(n^{-1/2}) \end{aligned} \tag{26}$$

Note that

$$G \left( \left( \begin{array}{c} \widehat{\gamma}_n \\ \theta_0^2 \end{array} \right) \right) = \Gamma E_1(\widehat{\gamma}_n - \theta_0^1) + o_B(1) |\widehat{\gamma}_n - \theta_0^1|$$

As above, the nonsingularity of  $\Gamma$  implies nonsingularity of  $\Gamma E_1$ , and hence, there exists a constant  $C_2 > 0$  such that  $|\Gamma E_1 x| \geq C_2 |x|$  for all  $x$ . Applying this to the equation above and collecting terms give

$$C_2 |\widehat{\gamma}_n - \theta_0^1| \leq |\Gamma E_1(\widehat{\gamma}_n - \theta_0^1)| = \left| G \left( \left( \begin{array}{c} \widehat{\gamma}_n \\ \theta_0^2 \end{array} \right) \right) - G(\theta_0) \right| + o_B(1) |\widehat{\gamma}_n - \theta_0^1| \tag{27}$$

Combining (27) with (26) yields

$$\begin{aligned} & (C_1 - o_B(1)) (C_2 - o_B(1)) |\widehat{\gamma}_n - \theta_0^1| \\ & \leq (C_1 - o_B(1)) \left| G \left( \left( \begin{array}{c} \widehat{\gamma}_n \\ \theta_0^2 \end{array} \right) \right) - G(\theta_0) \right| \\ & \leq o_B(n^{-1/2}) + O_B(1) O_p(n^{-1/2}) + O_B(1) O_B(n^{-1/2}) + O_B(1) O_p(n^{-1/2}) + O_B(1) o_B(n^{-1/2}) \end{aligned}$$

or

$$|\hat{\gamma}_n - \theta_0^1| \leq O_B(1) (O_p(n^{-1/2}) + O_B(n^{-1/2}))$$

**Part 2: Asymptotic Normality.** Let

$$\tilde{L}_n(t) = A\Gamma \left( \begin{pmatrix} t \\ \hat{\theta}^2 \end{pmatrix} - \begin{pmatrix} \theta_0^1 \\ \theta_0^2 \end{pmatrix} \right) + \hat{A}_n \hat{G}_n(\theta_0)$$

Define

$$\begin{aligned} \hat{\sigma}_n &= \arg \min_t |\tilde{L}_n(t)| = \\ & \arg \min_t \left( A\Gamma \left( \begin{pmatrix} t \\ \hat{\theta}^2 \end{pmatrix} - \begin{pmatrix} \theta_0^1 \\ \theta_0^2 \end{pmatrix} \right) + \hat{A}_n \hat{G}_n(\theta_0) \right)' \\ & \left( A\Gamma \left( \begin{pmatrix} t \\ \hat{\theta}^2 \end{pmatrix} - \begin{pmatrix} \theta_0^1 \\ \theta_0^2 \end{pmatrix} \right) + \hat{A}_n \hat{G}_n(\theta_0) \right) \end{aligned}$$

Solving for  $\hat{\sigma}_n$  gives

$$\begin{aligned} \hat{\sigma}_n &= \theta_0^1 - B_{11}^{-1} (B_{21}'x + C_1') \\ &= \theta_0^1 - ((\Gamma E_1)' A' A \Gamma E_1)^{-1} \\ & \quad \left( (\Gamma E_1)' A' A \Gamma E_2 (\hat{\theta}^2 - \theta_0^2) + (\Gamma E_1)' A' \hat{A}_n \hat{G}_n(\theta_0) \right) \\ &= \theta_0^1 - ((\Gamma E_1)' A' A \Gamma E_1)^{-1} (\Gamma E_1)' A' \\ & \quad \left( A \Gamma E_2 (\hat{\theta}^2 - \theta_0^2) + \hat{A}_n \hat{G}_n(\theta_0) \right) \end{aligned}$$

Mimicking the calculation on the top of page 195 of Hahn (1996),

$$\begin{aligned} (\hat{\sigma}_n - \gamma_n) &= -((\Gamma E_1)' A' A \Gamma E_1)^{-1} (\Gamma E_1)' A' \left( A \Gamma E_2 (\hat{\theta}^2 - \theta_0^2) + \hat{A}_n \hat{G}_n(\theta_0) \right) \\ & \quad + (E_1' \Gamma' A' A \Gamma E_1)^{-1} E_1' \Gamma' A' A G_n(\theta_0) \\ &= -((\Gamma E_1)' A' A \Gamma E_1)^{-1} (\Gamma E_1)' A' \\ & \quad \left( A \Gamma E_2 (\hat{\theta}^2 - \theta_0^2) + \hat{A}_n \hat{G}_n(\theta_0) - A G_n(\theta_0) \right) \\ &= -\Delta \left( \rho_n + \hat{A}_n \hat{G}_n(\theta_0) - A G_n(\theta_0) \right) \end{aligned}$$

where  $\Delta = ((\Gamma E_1)' A' A \Gamma E_1)^{-1} (\Gamma E_1)' A'$  and  $\rho_n = A \Gamma E_2 (\hat{\theta}^2 - \theta_0^2)$ . Or

$$(\hat{\sigma}_n - \gamma_n + \Delta \rho_n) = -\Delta \left( \hat{A}_n \hat{G}_n(\theta_0) - A G_n(\theta_0) \right)$$



From this it follows that  $\hat{\sigma}_n - \gamma_n = O_B(n^{-1/2})$ .

Next we want to argue that  $\sqrt{n}(\hat{\sigma}_n - \hat{\gamma}_n) = o_B(1)$ .

We next proceed as in Hahn (1996) (page 194). First we show that

$$\left| \hat{A}_n \hat{G}_n \left( \begin{pmatrix} \hat{\gamma}_n \\ \hat{\theta}^2 \end{pmatrix} \right) - \tilde{L}_n(\hat{\gamma}_n) \right| = o_B(n^{-1/2}) \quad (28)$$

It follows from Hahn

$$\left| \hat{A}_n \hat{G}_n \left( \begin{pmatrix} \hat{\gamma}_n \\ \hat{\theta}^2 \end{pmatrix} \right) - AG \left( \begin{pmatrix} \hat{\gamma}_n \\ \hat{\theta}^2 \end{pmatrix} \right) - \hat{A}_n \hat{G}_n(\theta_0) + AG(\theta_0) \right| = o_B(n^{-1/2})$$

We thus have

$$\begin{aligned} \left| \hat{A}_n \hat{G}_n \left( \begin{pmatrix} \hat{\gamma}_n \\ \hat{\theta}^2 \end{pmatrix} \right) - \tilde{L}_n(\hat{\gamma}_n) \right| &= \left| \hat{A}_n \hat{G}_n \left( \begin{pmatrix} \hat{\gamma}_n \\ \hat{\theta}^2 \end{pmatrix} \right) - A\Gamma \left( \begin{pmatrix} \hat{\gamma}_n \\ \hat{\theta}^2 \end{pmatrix} - \begin{pmatrix} \theta_0^1 \\ \theta_0^2 \end{pmatrix} \right) - \hat{A}_n \hat{G}_n(\theta_0) \right| \\ &\leq \left| \hat{A}_n \hat{G}_n \left( \begin{pmatrix} \hat{\gamma}_n \\ \hat{\theta}^2 \end{pmatrix} \right) - AG \left( \begin{pmatrix} \hat{\gamma}_n \\ \hat{\theta}^2 \end{pmatrix} \right) - \hat{A}_n \hat{G}_n(\theta_0) + AG(\theta_0) \right| \\ &\quad + \left| AG \left( \begin{pmatrix} \hat{\gamma}_n \\ \hat{\theta}^2 \end{pmatrix} \right) - AG(\theta_0) - A\Gamma \left( \begin{pmatrix} \hat{\gamma}_n \\ \hat{\theta}^2 \end{pmatrix} - \theta_0 \right) \right| \\ &= o_B(n^{-1/2}) + o \left( \begin{pmatrix} \hat{\gamma}_n \\ \hat{\theta}^2 \end{pmatrix} - \theta_0 \right) \\ &= o_B(n^{-1/2}) \end{aligned}$$

This uses the fact that  $\begin{pmatrix} \hat{\gamma}_n \\ \hat{\theta}^2 \end{pmatrix}$  is  $\sqrt{n}$ -consistent.

Next, we will show that

$$\left| \hat{A}_n \hat{G}_n \left( \begin{pmatrix} \hat{\sigma}_n \\ \hat{\theta}^2 \end{pmatrix} \right) - \tilde{L}_n(\hat{\sigma}_n) \right| = o_B(n^{-1/2}) \quad (29)$$

We have

$$\begin{aligned} \left| \hat{A}_n \hat{G}_n \left( \begin{pmatrix} \hat{\sigma}_n \\ \hat{\theta}^2 \end{pmatrix} \right) - \tilde{L}_n(\hat{\sigma}_n) \right| &= \left| \hat{A}_n \hat{G}_n \left( \begin{pmatrix} \hat{\sigma}_n \\ \hat{\theta}^2 \end{pmatrix} \right) - A\Gamma \left( \begin{pmatrix} \hat{\sigma}_n \\ \hat{\theta}^2 \end{pmatrix} - \theta_0 \right) - \hat{A}_n \hat{G}_n(\theta_0) \right| \\ &\leq \left| \hat{A}_n \hat{G}_n \left( \begin{pmatrix} \hat{\sigma}_n \\ \hat{\theta}^2 \end{pmatrix} \right) - AG \left( \begin{pmatrix} \hat{\sigma}_n \\ \hat{\theta}^2 \end{pmatrix} \right) - \hat{A}_n \hat{G}_n(\theta_0) + AG(\theta_0) \right| \\ &\quad + \left| AG \left( \begin{pmatrix} \hat{\sigma}_n \\ \hat{\theta}^2 \end{pmatrix} \right) - AG(\theta_0) - A\Gamma \left( \begin{pmatrix} \hat{\sigma}_n \\ \hat{\theta}^2 \end{pmatrix} - \theta_0 \right) \right| \\ &= o_B(n^{-1/2}) + o \left( \begin{pmatrix} \hat{\sigma}_n \\ \hat{\theta}^2 \end{pmatrix} - \theta_0 \right) \\ &= o_B(n^{-1/2}) \end{aligned}$$

For the last step we use  $\hat{\sigma}_n - \theta_0^1 = (\hat{\sigma}_n - \gamma_n) + (\gamma_n - \theta_0^1) = O_B(n^{-1/2}) + O_p(n^{-1/2})$ .

Combining (28) and (29) with the definitions of  $\hat{\gamma}_n$  and  $\hat{\sigma}_n$  we get

$$\left| \tilde{L}_n(\hat{\gamma}_n) \right| = \left| \tilde{L}_n(\hat{\sigma}_n) \right| + o_B(n^{-1/2}) \quad (30)$$

Exactly as in Hahn (1996) and Pakes and Pollard (1989), we start with

$$\begin{aligned} \left| \tilde{L}_n(\hat{\sigma}_n) \right| &\leq \left| A\Gamma \left( \begin{pmatrix} \hat{\sigma}_n \\ \hat{\theta}^2 \end{pmatrix} - \theta_0 \right) \right| + \left| \hat{A}_n \hat{G}_n(\theta_0) \right| \\ &\leq \left| A\Gamma \left( \begin{pmatrix} \hat{\sigma}_n \\ \hat{\theta}^2 \end{pmatrix} - \begin{pmatrix} \gamma_n \\ \hat{\theta}^2 \end{pmatrix} \right) \right| + \left| \hat{A}_n \hat{G}_n(\theta_0) - \hat{A}_n G_n(\theta_0) \right| \\ &\quad + \left| A\Gamma \left( \begin{pmatrix} \gamma_n \\ \hat{\theta}^2 \end{pmatrix} - \theta_0 \right) \right| + \left| \hat{A}_n G_n(\theta_0) \right| \\ &= O_B(n^{-1/2}) + O_B(1) O_B(n^{-1/2}) + O_p(n^{-1/2}) + O_B(1) O_p(n^{-1/2}) \quad (31) \end{aligned}$$

Squaring both sides of (30) we have

$$\left| \tilde{L}_n(\hat{\gamma}_n) \right|^2 = \left| \tilde{L}_n(\hat{\sigma}_n) \right|^2 + o_B(n^{-1}) \quad (32)$$

because (31) implies that the cross-product term can be absorbed in the  $o_B(n^{-1})$ . On the other hand, for any  $t$

$$\tilde{L}_n(t) = A\Gamma \left( \begin{pmatrix} t \\ \hat{\theta}^2 \end{pmatrix} - \begin{pmatrix} \theta_0^1 \\ \theta_0^2 \end{pmatrix} \right) + \hat{A}_n \hat{G}_n(\theta_0)$$

has the form  $\tilde{L}_n(t) = y - Xt$  where  $X = -A\Gamma E_1$  and  $y = -A\Gamma E_1 \theta_0^1 + A\Gamma E_2 (\hat{\theta}^2 - \theta_0^2) + \hat{A}_n \hat{G}_n(\theta_0)$

$\hat{\sigma}_n$  solves a least squares problem with first-order condition  $X' \tilde{L}_n(\hat{\sigma}_n) = 0$ . Also

$$\begin{aligned} \left| \tilde{L}_n(t) \right|^2 &= (y - Xt)'(y - Xt) \\ &= ((y - X\hat{\sigma}_n) - X(t - \hat{\sigma}_n))'((y - X\hat{\sigma}_n) - X(t - \hat{\sigma}_n)) \\ &= (y - X\hat{\sigma}_n)'(y - X\hat{\sigma}_n) + (t - \hat{\sigma}_n)' X' X (t - \hat{\sigma}_n) \\ &\quad - 2(t - \hat{\sigma}_n)' X' (y - X\hat{\sigma}_n) \\ &= \left| \tilde{L}_n(\hat{\sigma}_n) \right|^2 + |X(t - \hat{\sigma}_n)|^2 - 2(t - \hat{\sigma}_n)' X' \tilde{L}_n(\hat{\sigma}_n) \\ &= \left| \tilde{L}_n(\hat{\sigma}_n) \right|^2 + |(A\Gamma E_1)(t - \hat{\sigma}_n)|^2 \end{aligned}$$

Plugging in  $t = \hat{\gamma}_n$  we have

$$\left| \tilde{L}_n(\hat{\gamma}_n) \right|^2 = \left| \tilde{L}_n(\hat{\sigma}_n) \right|^2 + |(A\Gamma E_1)(\hat{\gamma}_n - \hat{\sigma}_n)|^2$$

Compare this to (32) to conclude that

$$(A\Gamma E_1)(\hat{\gamma}_n - \hat{\sigma}_n) = o_B(n^{-1/2})$$

$A\Gamma E_1$  has full rank by assumption so  $(\hat{\gamma}_n - \hat{\sigma}_n) = o_B(n^{-1/2})$  and  $n^{1/2}(\hat{\gamma}_n - \gamma_n) = n^{1/2}(\hat{\sigma}_n - \gamma_n) + o_B(n^{-1/2})$  and since  $n^{1/2}(\hat{\sigma}_n - \gamma_n) \xrightarrow{p} N(0, \Omega)$ , we obtain  $n^{1/2}(\hat{\gamma}_n - \gamma_n) \xrightarrow{p} N(0, \Omega)$ . ■

Theorem 2 is stated for GMM estimators. This covers extremum estimators and the two-step estimators as special cases. Theorem 2 also covers the case where one is interested in different infeasible lower-dimensional estimators as in Section 2. To see this, consider two estimators of the form

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

and

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

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

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

which has the form of (20).

## Appendix 2: Non-Singularity of the Matrix in Equation (6)

The determinant of the matrix on the left of (6) is

$$\begin{aligned}
& 2k_2h_{11}(k_3h_{11} + \rho k_1) + 2k_3h_{11}(k_2h_{11} - \rho k_1) \\
&= 2h_{11} \left[ h_{11}^{-2} \frac{(1 + \rho v)(1 - \rho v)}{(h_{11} + 2h_{12} + h_{22})(h_{11} - 2h_{12} + h_{22})} h_{11} \right] \\
&\quad + 2h_{11} \left[ \rho \frac{v}{h_{22}} h_{11}^{-1} \frac{(1 + \rho v)(h_{11} - 2h_{12} + h_{22}) - (1 - \rho v)(h_{11} + 2h_{12} + h_{22})}{(h_{11} + 2h_{12} + h_{22})(h_{11} - 2h_{12} + h_{22})} \right] \\
&= \frac{2(1 - \rho^2 v^2)}{(h_{11} + 2h_{12} + h_{22})(h_{11} - 2h_{12} + h_{22})} + 2\rho \frac{v}{h_{22}} \frac{2v\rho h_{11} - 4h_{12} + 2v\rho h_{22}}{(h_{11} + 2h_{12} + h_{22})(h_{11} - 2h_{12} + h_{22})} \\
&= \frac{4(1 - \rho^2 v^2) + 2\rho \frac{v}{h_{22}} (2v\rho h_{11} - 4h_{12} + 2v\rho h_{22})}{(h_{11} + 2h_{12} + h_{22})(h_{11} - 2h_{12} + h_{22})} \\
&= \frac{-8v\rho \frac{h_{12}}{h_{22}} + 4v^2 \rho^2 \frac{h_{11}}{h_{22}} + 4}{(h_{11} + 2h_{12} + h_{22})(h_{11} - 2h_{12} + h_{22})} \\
&= \frac{4 \begin{pmatrix} -\rho V \\ 1 \end{pmatrix}' H \begin{pmatrix} -\rho V \\ 1 \end{pmatrix}}{\begin{pmatrix} 0 \\ 1 \end{pmatrix}' H \begin{pmatrix} 0 \\ 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix}' H \begin{pmatrix} 1 \\ 1 \end{pmatrix} \begin{pmatrix} -1 \\ 1 \end{pmatrix}' H \begin{pmatrix} -1 \\ 1 \end{pmatrix}} > 0
\end{aligned}$$

since  $H$  is positive definite.

## Appendix 3: Maximum of Two Lognormals

Let  $(X_1, X_2)'$  have a bivariate normal distribution with mean  $(\mu_1, \mu_2)'$  and variance  $\begin{pmatrix} \sigma_1^2 & \tau\sigma_1\sigma_2 \\ \tau\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$  and let  $(Y_1, Y_2)' = (\exp(X_1), \exp(X_2))'$ . We are interested in  $E[\max\{Y_1, Y_2\}]$ .

Kotz, Balakrishnan, and Johnson (2000) present the moment-generating function for  $\min \{X_1, X_2\}$  is

$$\begin{aligned} M(t) &= E[\exp(\min \{X_1, X_2\} t)] = \exp(t\mu_1 + t^2\sigma_1^2/2) \Phi\left(\frac{\mu_2 - \mu_1 - t(\sigma_1^2 - \tau\sigma_1\sigma_2)}{\sqrt{\sigma_2^2 - 2\tau\sigma_1\sigma_2 + \sigma_1^2}}\right) \\ &\quad + \exp(t\mu_2 + t^2\sigma_2^2/2) \Phi\left(\frac{\mu_1 - \mu_2 - t(\sigma_2^2 - \tau\sigma_1\sigma_2)}{\sqrt{\sigma_2^2 - 2\tau\sigma_1\sigma_2 + \sigma_1^2}}\right) \end{aligned}$$

Therefore

$$\begin{aligned} E[\max \{Y_1, Y_2\}] &= E[Y_1] + E[Y_2] - E[\min \{Y_1, Y_2\}] \\ &= E[\exp(X_1)] + E[\exp(X_2)] - E[\min \{\exp(X_1), \exp(X_2)\}] \\ &= \exp(\mu_1 + \sigma_1^2/2) + \exp(\mu_2 + \sigma_2^2/2) - E[\exp(\min \{X_1, X_2\})] \\ &= \exp(\mu_1 + \sigma_1^2/2) + \exp(\mu_2 + \sigma_2^2/2) \\ &\quad - \exp(\mu_1 + \sigma_1^2/2) \Phi\left(\frac{\mu_2 - \mu_1 - (\sigma_1^2 - \tau\sigma_1\sigma_2)}{\sqrt{\sigma_2^2 - 2\tau\sigma_1\sigma_2 + \sigma_1^2}}\right) \\ &\quad - \exp(\mu_2 + \sigma_2^2/2) \Phi\left(\frac{\mu_1 - \mu_2 - (\sigma_2^2 - \tau\sigma_1\sigma_2)}{\sqrt{\sigma_2^2 - 2\tau\sigma_1\sigma_2 + \sigma_1^2}}\right) \\ &= \exp(\mu_1 + \sigma_1^2/2) \left(1 - \Phi\left(\frac{\mu_2 - \mu_1 - (\sigma_1^2 - \tau\sigma_1\sigma_2)}{\sqrt{\sigma_2^2 - 2\tau\sigma_1\sigma_2 + \sigma_1^2}}\right)\right) \\ &\quad + \exp(\mu_2 + \sigma_2^2/2) \left(1 - \Phi\left(\frac{\mu_1 - \mu_2 - (\sigma_2^2 - \tau\sigma_1\sigma_2)}{\sqrt{\sigma_2^2 - 2\tau\sigma_1\sigma_2 + \sigma_1^2}}\right)\right) \end{aligned}$$

## Appendix 4: Identification of the Variance of Two-Step Estimators

Consider the two-step estimation problem in equation (12) in Section 6. As mentioned, the asymptotic variance of  $\hat{\theta}_2$  is

$$R_2^{-1} R_1 Q_1^{-1} V_{11} Q_1^{-1} R_1' R_2^{-1} - R_2^{-1} V_{21} Q_1^{-1} R_1' R_2^{-1} - R_2^{-1} R_1 Q_1^{-1} V_{12} R_2^{-1} + R_2^{-1} V_{22} R_2^{-1}$$

where  $\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} = \text{var}\left(\begin{pmatrix} q(z_i, \theta_1) \\ r(z_i, \theta_1, \theta_2) \end{pmatrix}\right)$ ,  $Q_1 = E\left[\frac{\partial q(z_i, \theta_1)}{\partial \theta_1}\right]$ ,  $R_1 = E\left[\frac{\partial r(z_i, \theta_1, \theta_2)}{\partial \theta_1}\right]$  and  $R_2 = E\left[\frac{\partial r(z_i, \theta_1, \theta_2)}{\partial \theta_2}\right]$ . It is often easy to estimate  $V_{11}$ ,  $V_{22}$ ,  $Q_1$  and  $R_2$  directly. When it is not,

they can be estimated using the poor woman's bootstrap procedure above. We therefore focus on  $V_{12}$  and  $R_1$ .

Consider one-dimensional estimators of the form

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

The asymptotic variance of  $(\widehat{a}_1(\delta_1), \widehat{a}_2(\delta_1, \delta_2), \widehat{a}_3(\delta_3))$  is

$$\begin{pmatrix} \delta'_1 Q_1 \delta_1 & 0 & 0 \\ \delta'_1 R_1 \delta_2 & \delta'_2 R_2 \delta_2 & 0 \\ 0 & 0 & \delta'_3 R_2 \delta_3 \end{pmatrix}^{-1} \begin{pmatrix} \delta'_1 V_{11} \delta_1 & \delta'_1 V_{12} \delta_2 & \delta'_1 V_{12} \delta_3 \\ \delta'_1 V'_{12} \delta_2 & \delta'_2 V_{22} \delta_2 & \delta'_2 V_{22} \delta_3 \\ \delta'_1 V'_{12} \delta_3 & \delta'_2 V_{22} \delta_3 & \delta'_3 V_{22} \delta_3 \end{pmatrix} \begin{pmatrix} \delta'_1 Q_1 \delta_1 & \delta'_2 R_1 \delta_1 & 0 \\ 0 & \delta'_2 R_2 \delta_2 & 0 \\ 0 & 0 & \delta'_3 R_2 \delta_3 \end{pmatrix}^{-1}.$$

When  $\delta_2 = \delta_3$ , this has the form

$$\begin{pmatrix} q_1 & 0 & 0 \\ r_1 & r_2 & 0 \\ 0 & 0 & r_2 \end{pmatrix}^{-1} \begin{pmatrix} V_q & V_{qr} & V_{qr} \\ V_{qr} & V_r & V_r \\ V_{qr} & V_r & V_r \end{pmatrix} \begin{pmatrix} q_1 & r_1 & 0 \\ 0 & r_2 & 0 \\ 0 & 0 & r_2 \end{pmatrix}^{-1}$$

where  $q_1 = \delta'_1 Q_1 \delta_1$ ,  $r_1 = \delta'_1 R_1 \delta_2$ ,  $r_2 = \delta'_2 R_2 \delta_2$ ,  $V_q = \delta'_1 V_{11} \delta_1$ ,  $V_{qr} = \delta'_1 V_{12} \delta_2$  and  $V_r = \delta'_2 V_{22} \delta_2$ .

This can be written as

$$\begin{pmatrix} \frac{V_q}{q_1^2} & \frac{1}{q_1 r_2} V_{qr} - \frac{V_q r_1}{q_1^2 r_2} & \frac{1}{q_1 r_2} V_{qr} \\ \frac{1}{q_1} \left( \frac{1}{r_2} V_{qr} - \frac{V_q r_1}{q_1 r_2} \right) & \frac{1}{r_2} \left( \frac{V_r}{r_2} - \frac{1}{q_1} \frac{r_1}{r_2} V_{qr} \right) - \frac{1}{q_1} \frac{r_1}{r_2} \left( \frac{1}{r_2} V_{qr} - \frac{V_q r_1}{q_1 r_2} \right) & \frac{1}{r_2} \left( \frac{V_r}{r_2} - \frac{1}{q_1} \frac{r_1}{r_2} V_{qr} \right) \\ \frac{1}{q_1 r_2} V_{qr} & \frac{V_r}{r_2^2} - \frac{1}{q_1} \frac{r_1}{r_2^2} V_{qr} & \frac{V_r}{r_2^2} \end{pmatrix}$$

Normalize so  $V_q = 1$ , and parameterize  $V_r = v^2$  and  $V_{qr} = \rho \sqrt{V_q V_r} = \rho v$  gives the matrix

$$\begin{pmatrix} \frac{1}{q_1^2} & \frac{1}{q_1 r_2} \rho v - \frac{1}{q_1^2} \frac{r_1}{r_2} & \frac{1}{q_1 r_2} \rho v \\ \frac{1}{q_1} \left( \frac{1}{r_2} \rho v - \frac{1}{q_1} \frac{r_1}{r_2} \right) & \frac{1}{r_2} \left( \frac{v^2}{r_2} - \frac{1}{q_1} \frac{r_1}{r_2} \rho v \right) - \frac{1}{q_1} \frac{r_1}{r_2} \left( \frac{1}{r_2} \rho v - \frac{1}{q_1} \frac{r_1}{r_2} \right) & \frac{1}{r_2} \left( \frac{v^2}{r_2} - \frac{1}{q_1} \frac{r_1}{r_2} \rho v \right) \\ \frac{1}{q_1 r_2} \rho v & \frac{v^2}{r_2^2} - \frac{1}{q_1} \frac{r_1}{r_2^2} \rho v & \frac{v^2}{r_2^2} \end{pmatrix}$$

Denoting the  $(\ell, m)$ 'th element of this matrix by  $\omega_{\ell m}$  we have

$$\begin{aligned}
\omega_{33} - \omega_{32} &= \frac{1}{q_1} \frac{r_1}{r_2^2} \rho v = \frac{r_1}{r_2} \omega_{31} \\
\frac{\omega_{33} - \omega_{32}}{\omega_{31}} &= \frac{r_1}{r_2} \\
\rho &= \frac{\omega_{31}}{\sqrt{\omega_{11}\omega_{33}}}
\end{aligned}$$

Since  $r_2$  is known, this gives  $r_1$  and  $\rho$ . We also know  $v$  from  $\omega_{33}$ .

This implies that the asymptotic variance of  $(\hat{a}_1(\delta_1), \hat{a}_2(\delta_1, \delta_2), \hat{a}_3(\delta_3))$  identifies  $\delta_1' V_{12} \delta_2$  and  $\delta_1' R_1 \delta_2$ . Choosing  $\delta_1 = e_j$  and  $\delta_2 = e_m$  (for  $j = 1, \dots, k_1$  and  $m = 1, \dots, k_2$ ) recovers all the elements of  $V_{12}$  and  $R_1$ .

## Appendix 5. Exploiting the Structure in Helpman et al

In the specification used by Helpman, Melitz, and Rubinstein (2008) and in the modification in Section 7.1, 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- $l$ . In the application,  $\Delta$  is either pick out the vector  $(\beta_1', \beta_2)'$  or the scalar  $\beta_3$ .

Using the notation from Section 6, 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  $-Q_1^{-1} V_{11} Q_1^{-1} R_1' \Delta (\Delta' R_2 \Delta)^{-1} + Q_1^{-1} V_{12} \Delta (\Delta' R_2 \Delta)^{-1}$  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  $R'_1$ .



Table 1: 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	0.027	0.017
$\beta_2$	0.029	0.023	0.017
$\beta_3$	0.031	0.027	0.018
$\beta_4$	0.032	0.027	0.020
$\beta_5$	0.033	0.026	0.020
$\beta_6$	0.032	0.029	0.022
$\beta_7$	0.031	0.025	0.020
$\beta_8$	0.033	0.027	0.020
$\beta_9$	0.034	0.026	0.021
$\beta_{10}$	0.033	0.034	0.018

Table 2: 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

Table 3: 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	0.028	0.031	0.031
$\gamma_2$	0.064	0.068	0.069	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 4: Structural Model

Rejection Probabilities (20% level of significance)

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

Table 5: Selection Model  
Means of Estimated Standard Errors

	Actual	No Correction	With Correction	Poor Woman's BS	Regular BS
$\tilde{\beta}_0$	0.290	0.276	0.278	0.280	0.283
BORDER	0.077	0.065	0.072	0.080	0.080
ISLAND	0.052	0.045	0.054	0.059	0.059
LANDLOCKED	0.098	0.076	0.093	0.100	0.100
LEGAL	0.024	0.019	0.023	0.025	0.025
LANGUAGE	0.026	0.021	0.025	0.027	0.027
COLONIAL	0.074	0.066	0.068	0.074	0.074
CU	0.127	0.099	0.119	0.129	0.128
FTA	0.106	0.091	0.094	0.104	0.105
WTONONE	0.045	0.034	0.040	0.044	0.043
WTOBOTH	0.025	0.020	0.023	0.026	0.026
MILLS	0.052	0.041	0.047	0.051	0.051
$\lambda$	0.060	0.051	0.054	0.057	0.058

Table 6: Selection Model  
Medians of Estimated Standard Errors

	Actual	No Correction	With Correction	Poor Woman's BS	Regular BS
$\tilde{\beta}_0$	0.290	0.276	0.277	0.278	0.280
BORDER	0.077	0.065	0.072	0.080	0.079
ISLAND	0.052	0.045	0.054	0.059	0.059
LANDLOCKED	0.098	0.076	0.093	0.100	0.100
LEGAL	0.024	0.019	0.023	0.025	0.025
LANGUAGE	0.026	0.021	0.025	0.027	0.027
COLONIAL	0.074	0.066	0.068	0.074	0.073
CU	0.127	0.099	0.119	0.129	0.128
FTA	0.106	0.091	0.094	0.103	0.105
WTONONE	0.045	0.034	0.040	0.043	0.043
WTOBOTH	0.025	0.020	0.023	0.026	0.026
MILLS	0.052	0.041	0.046	0.050	0.051
$\lambda$	0.060	0.051	0.054	0.057	0.058

Table 7: Selection Model

Rejection Probabilities (20% level of significance)

	No Correction	With Correction	Poor Woman's BS	Regular BS
$\tilde{\beta}_0$	24%	24%	23%	23%
BORDER	29%	26%	20%	20%
ISLAND	30%	20%	16%	16%
LANDLOCKED	28%	20%	17%	18%
LEGAL	30%	25%	20%	20%
LANGUAGE	31%	21%	18%	18%
COLONIAL	25%	25%	21%	22%
CU	31%	21%	19%	19%
FTA	24%	24%	20%	20%
WTONONE	37%	25%	22%	22%
WTOBOTH	31%	25%	19%	19%
MILLS	32%	27%	24%	23%
$\lambda$	26%	25%	23%	21%

## Working Paper Series

A series of research studies on regional economic issues relating to the Seventh Federal Reserve District, and on financial and economic topics.

Examining Macroeconomic Models through the Lens of Asset Pricing <i>Jaroslav Borovička and Lars Peter Hansen</i>	WP-12-01
The Chicago Fed DSGE Model <i>Scott A. Brave, Jeffrey R. Campbell, Jonas D.M. Fisher, and Alejandro Justiniano</i>	WP-12-02
Macroeconomic Effects of Federal Reserve Forward Guidance <i>Jeffrey R. Campbell, Charles L. Evans, Jonas D.M. Fisher, and Alejandro Justiniano</i>	WP-12-03
Modeling Credit Contagion via the Updating of Fragile Beliefs <i>Luca Benzoni, Pierre Collin-Dufresne, Robert S. Goldstein, and Jean Helwege</i>	WP-12-04
Signaling Effects of Monetary Policy <i>Leonardo Melosi</i>	WP-12-05
Empirical Research on Sovereign Debt and Default <i>Michael Tomz and Mark L. J. Wright</i>	WP-12-06
Credit Risk and Disaster Risk <i>François Gourio</i>	WP-12-07
From the Horse's Mouth: How do Investor Expectations of Risk and Return Vary with Economic Conditions? <i>Gene Amromin and Steven A. Sharpe</i>	WP-12-08
Using Vehicle Taxes To Reduce Carbon Dioxide Emissions Rates of New Passenger Vehicles: Evidence from France, Germany, and Sweden <i>Thomas Klier and Joshua Linn</i>	WP-12-09
Spending Responses to State Sales Tax Holidays <i>Sumit Agarwal and Leslie McGranahan</i>	WP-12-10
Micro Data and Macro Technology <i>Ezra Oberfield and Devesh Raval</i>	WP-12-11
The Effect of Disability Insurance Receipt on Labor Supply: A Dynamic Analysis <i>Eric French and Jae Song</i>	WP-12-12
Medicaid Insurance in Old Age <i>Mariacristina De Nardi, Eric French, and John Bailey Jones</i>	WP-12-13
Fetal Origins and Parental Responses <i>Douglas Almond and Bhashkar Mazumder</i>	WP-12-14

## Working Paper Series *(continued)*

Repos, Fire Sales, and Bankruptcy Policy <i>Gaetano Antinolfi, Francesca Carapella, Charles Kahn, Antoine Martin, David Mills, and Ed Nosal</i>	WP-12-15
Speculative Runs on Interest Rate Pegs The Frictionless Case <i>Marco Bassetto and Christopher Phelan</i>	WP-12-16
Institutions, the Cost of Capital, and Long-Run Economic Growth: Evidence from the 19th Century Capital Market <i>Ron Alquist and Ben Chabot</i>	WP-12-17
Emerging Economies, Trade Policy, and Macroeconomic Shocks <i>Chad P. Bown and Meredith A. Crowley</i>	WP-12-18
The Urban Density Premium across Establishments <i>R. Jason Faberman and Matthew Freedman</i>	WP-13-01
Why Do Borrowers Make Mortgage Refinancing Mistakes? <i>Sumit Agarwal, Richard J. Rosen, and Vincent Yao</i>	WP-13-02
Bank Panics, Government Guarantees, and the Long-Run Size of the Financial Sector: Evidence from Free-Banking America <i>Benjamin Chabot and Charles C. Moul</i>	WP-13-03
Fiscal Consequences of Paying Interest on Reserves <i>Marco Bassetto and Todd Messer</i>	WP-13-04
Properties of the Vacancy Statistic in the Discrete Circle Covering Problem <i>Gadi Barlevy and H. N. Nagaraja</i>	WP-13-05
Credit Crunches and Credit Allocation in a Model of Entrepreneurship <i>Marco Bassetto, Marco Cagetti, and Mariacristina De Nardi</i>	WP-13-06
Financial Incentives and Educational Investment: The Impact of Performance-Based Scholarships on Student Time Use <i>Lisa Barrow and Cecilia Elena Rouse</i>	WP-13-07
The Global Welfare Impact of China: Trade Integration and Technological Change <i>Julian di Giovanni, Andrei A. Levchenko, and Jing Zhang</i>	WP-13-08
Structural Change in an Open Economy <i>Timothy Uy, Kei-Mu Yi, and Jing Zhang</i>	WP-13-09
The Global Labor Market Impact of Emerging Giants: a Quantitative Assessment <i>Andrei A. Levchenko and Jing Zhang</i>	WP-13-10

## Working Paper Series *(continued)*

Size-Dependent Regulations, Firm Size Distribution, and Reallocation <i>François Gourio and Nicolas Roys</i>	WP-13-11
Modeling the Evolution of Expectations and Uncertainty in General Equilibrium <i>Francesco Bianchi and Leonardo Melosi</i>	WP-13-12
Rushing into American Dream? House Prices, Timing of Homeownership, and Adjustment of Consumer Credit <i>Sumit Agarwal, LuoJia Hu, and Xing Huang</i>	WP-13-13
The Earned Income Tax Credit and Food Consumption Patterns <i>Leslie McGranahan and Diane W. Schanzenbach</i>	WP-13-14
Agglomeration in the European automobile supplier industry <i>Thomas Klier and Dan McMillen</i>	WP-13-15
Human Capital and Long-Run Labor Income Risk <i>Luca Benzoni and Olena Chyruk</i>	WP-13-16
The Effects of the Saving and Banking Glut on the U.S. Economy <i>Alejandro Justiniano, Giorgio E. Primiceri, and Andrea Tambalotti</i>	WP-13-17
A Portfolio-Balance Approach to the Nominal Term Structure <i>Thomas B. King</i>	WP-13-18
Gross Migration, Housing and Urban Population Dynamics <i>Morris A. Davis, Jonas D.M. Fisher, and Marcelo Veracierto</i>	WP-13-19
Very Simple Markov-Perfect Industry Dynamics <i>Jaap H. Abbring, Jeffrey R. Campbell, Jan Tilly, and Nan Yang</i>	WP-13-20
Bubbles and Leverage: A Simple and Unified Approach <i>Robert Barsky and Theodore Bogusz</i>	WP-13-21
The scarcity value of Treasury collateral: Repo market effects of security-specific supply and demand factors <i>Stefania D'Amico, Roger Fan, and Yuriy Kitsul</i>	WP-13-22
Gambling for Dollars: Strategic Hedge Fund Manager Investment <i>Dan Bernhardt and Ed Nosal</i>	WP-13-23
Cash-in-the-Market Pricing in a Model with Money and Over-the-Counter Financial Markets <i>Fabrizio Mattesini and Ed Nosal</i>	WP-13-24
An Interview with Neil Wallace <i>David Altig and Ed Nosal</i>	WP-13-25



## Working Paper Series *(continued)*

Firm Dynamics and the Minimum Wage: A Putty-Clay Approach <i>Daniel Aaronson, Eric French, and Isaac Sorkin</i>	WP-13-26
Policy Intervention in Debt Renegotiation: Evidence from the Home Affordable Modification Program <i>Sumit Agarwal, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, Tomasz Piskorski, and Amit Seru</i>	WP-13-27
The Effects of the Massachusetts Health Reform on Financial Distress <i>Bhashkar Mazumder and Sarah Miller</i>	WP-14-01
Can Intangible Capital Explain Cyclical Movements in the Labor Wedge? <i>François Gourio and Leena Rudanko</i>	WP-14-02
Early Public Banks <i>William Roberds and François R. Velde</i>	WP-14-03
Mandatory Disclosure and Financial Contagion <i>Fernando Alvarez and Gadi Barlevy</i>	WP-14-04
The Stock of External Sovereign Debt: Can We Take the Data at ‘Face Value’? <i>Daniel A. Dias, Christine Richmond, and Mark L. J. Wright</i>	WP-14-05
Interpreting the <i>Pari Passu</i> Clause in Sovereign Bond Contracts: It’s All Hebrew (and Aramaic) to Me <i>Mark L. J. Wright</i>	WP-14-06
AIG in Hindsight <i>Robert McDonald and Anna Paulson</i>	WP-14-07
On the Structural Interpretation of the Smets-Wouters “Risk Premium” Shock <i>Jonas D.M. Fisher</i>	WP-14-08
Human Capital Risk, Contract Enforcement, and the Macroeconomy <i>Tom Krebs, Moritz Kuhn, and Mark L. J. Wright</i>	WP-14-09
Adverse Selection, Risk Sharing and Business Cycles <i>Marcelo Veracierto</i>	WP-14-10
Core and ‘Crust’: Consumer Prices and the Term Structure of Interest Rates <i>Andrea Ajello, Luca Benzoni, and Olena Chyruk</i>	WP-14-11
The Evolution of Comparative Advantage: Measurement and Implications <i>Andrei A. Levchenko and Jing Zhang</i>	WP-14-12

## Working Paper Series *(continued)*

Saving Europe?: The Unpleasant Arithmetic of Fiscal Austerity in Integrated Economies <i>Enrique G. Mendoza, Linda L. Tesar, and Jing Zhang</i>	WP-14-13
Liquidity Traps and Monetary Policy: Managing a Credit Crunch <i>Francisco Buera and Juan Pablo Nicolini</i>	WP-14-14
Quantitative Easing in Joseph's Egypt with Keynesian Producers <i>Jeffrey R. Campbell</i>	WP-14-15
Constrained Discretion and Central Bank Transparency <i>Francesco Bianchi and Leonardo Melosi</i>	WP-14-16
Escaping the Great Recession <i>Francesco Bianchi and Leonardo Melosi</i>	WP-14-17
More on Middlemen: Equilibrium Entry and Efficiency in Intermediated Markets <i>Ed Nosal, Yuet-Yee Wong, and Randall Wright</i>	WP-14-18
Preventing Bank Runs <i>David Andolfatto, Ed Nosal, and Bruno Sultanum</i>	WP-14-19
The Impact of Chicago's Small High School Initiative <i>Lisa Barrow, Diane Whitmore Schanzenbach, and Amy Claessens</i>	WP-14-20
Credit Supply and the Housing Boom <i>Alejandro Justiniano, Giorgio E. Primiceri, and Andrea Tambalotti</i>	WP-14-21
The Effect of Vehicle Fuel Economy Standards on Technology Adoption <i>Thomas Klier and Joshua Linn</i>	WP-14-22
What Drives Bank Funding Spreads? <i>Thomas B. King and Kurt F. Lewis</i>	WP-14-23
Inflation Uncertainty and Disagreement in Bond Risk Premia <i>Stefania D'Amico and Athanasios Orphanides</i>	WP-14-24
Access to Refinancing and Mortgage Interest Rates: HARPing on the Importance of Competition <i>Gene Amromin and Caitlin Kearns</i>	WP-14-25
Private Takings <i>Alessandro Marchesiani and Ed Nosal</i>	WP-14-26
Momentum Trading, Return Chasing, and Predictable Crashes <i>Benjamin Chabot, Eric Ghysels, and Ravi Jagannathan</i>	WP-14-27
Early Life Environment and Racial Inequality in Education and Earnings in the United States <i>Kenneth Y. Chay, Jonathan Guryan, and Bhashkar Mazumder</i>	WP-14-28

**Working Paper Series** *(continued)*

Poor (Wo)man's Bootstrap  
*Bo E. Honoré and Luojia Hu*

**WP-15-01**