Honoré, Bo E.; Hu, Luojia

Working Paper

Poor (wo)man's bootstrap

Working Paper, No. 2015-01

Provided in Cooperation with:
Federal Reserve Bank of Chicago

Suggested Citation: Honoré, Bo E.; Hu, Luojia (2015) : Poor (wo)man's bootstrap, Working Paper, No. 2015-01, Federal Reserve Bank of Chicago, Chicago, IL

Federal Reserve Bank of Chicago

Poor (Wo)man's Bootstrap

Bo E. Honoré and Luojia Hu

March 2015

WP 2015-01

Poor (Wo)man's Bootstrap*

Bo E. Honoré† Luojia Hu‡

March 4, 2015

Abstract

The bootstrap is a convenient tool for calculating standard errors of the parameters 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.

The paper contains no new difficult math. But we believe that it can be useful.

KEYWORDS: standard error; bootstrap; inference; structural models; parametric estimation.

JEL CODE: C10, C18.

1 Introduction

The bootstrap is often used for estimating standard errors in applied work even when analytical expression exists for a consistent estimator. The bootstrap is convenient from a programming point of view, because it relies on the same estimation procedure that delivers the point estimates. Moreover, it does not explicitly force the researcher to make choices regarding bandwidths or number of nearest neighbors when the estimator is based on a non–smooth objective function or discontinuous moment conditions.

*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. Jan De Loecker and Aureo de Paula provided constructive suggestions as did seminar participants at the Federal Reserve Bank of Chicago, University of Aarhus, University of Copenhagen and Central European University.

†Mailing Address: Department of Economics, Princeton University, Princeton, NJ 08544-1021. Email: honore@Princeton.edu.

‡Mailing Address: Economic Research Department, Federal Reserve Bank of Chicago, 230 S. La Salle Street, Chicago, IL 60604. Email: lhu@frbchi.org.

Unfortunately, the bootstrap can be computationally burdensome if the estimator is complex. For example, in many structural econometric models it can take hours or days to get a single bootstrap draw of the estimator. 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 variance of the estimator of interest. The need for faster alternatives to the standard bootstrap also motivated the papers by Heagerty and Lumley (2000) and Hong and Scaillet (2006). Unfortunately, their approach assumes 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.

We emphasize that the contribution is the convenience of the approach, and 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 asymptotically normally distributed extremum estimator. We introduce a set of simple infeasible one-dimensional estimators related to the estimator of interest, and we show how their asymptotic variances can be used to back out the asymptotic variance of the estimator of the parameter of interest. In Appendix 4, we extend the results in Hahn (1996) to demonstrate that the bootstrap can be used to estimate the variances of the simple infeasible one-dimensional estimators. This suggests a computationally simple method for estimating the variance of the estimator of the parameter-vector of interest. We also demonstrate that this insight carries over to GMM estimators and that an alternative, and even simpler approach can be applied to method of moments estimators. It turns out that our procedure is not necessarily convenient for two-step estimators. In Section 2.5, we therefore propose a modified version specifically tailored for this scenario.

In Section 3, we discuss how the asymptotic variances of the simpler infeasible estimators can be estimated using the bootstrap and we propose a practical procedure for mapping them into the asymptotic variance of interest.

We illustrate our approach in Section 4. We first focus on the OLS estimator. The advantage of this 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 the maximum rank correlation estimator and to an indirect inference estimator of a structural econometric model. The former is chosen because it is an estimator that can be time–consuming to estimate and whose variance depends on unknown densities and conditional expectations. The latter provides an example of the kind of model where we think the approach will be useful in current empirical research.

2 Basic Idea

2.1 M–estimators

Consider an extremum estimator of a parameter β based on a random sample $\{z_i\}$,

$$\widehat{\beta} = \arg\min_b Q_n(b) = \arg\min_b \sum_{i=1}^{n} q(z_i, b).$$

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

$$avar\left(\widehat{\beta}\right) = H^{-1}VH^{-1}$$

where V and H are both symmetric and positive definite. When q is a smooth function of b, V is the variance of the derivative of q with respect to b 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 $avar\left(\widehat{\beta}\right)$ 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 $avar\left(\widehat{\beta}\right)$.

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

$$\widehat{a}(\delta) = \arg\min_{a} Q_n(\beta + \delta a) \tag{1}$$

where δ is a fixed vector. The bootstrap equivalent of this is

$$\arg\min_{a} \sum_{i=1}^{n} q\left(z_i^b, \widehat{\beta} + \delta a\right) \tag{2}$$

where $\left\{z_i^b\right\}$ 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 $\widehat{\beta}$ and its bootstrap equivalent. The validity of the bootstrap distribution of $\widehat{a}(\delta)$ follows from Appendix 4.

It is easiest to illustrate why this works by considering a case where β is two–dimensional. For this case, consider two vectors δ_1 and δ_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 $\widehat{\beta}$, the infeasible estimators $\widehat{a}(\delta_1)$ and $\widehat{a}(\delta_2)$ will be jointly asymptotically normal with variance

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

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

$$\Omega_{(1,0),(0,1)} = \begin{pmatrix} h_{11}^{-2} v_{11} & h_{11}^{-1} v_{12} h_{22}^{-1} \\ h_{11}^{-1} v_{12} h_{22}^{-1} & h_{22}^{-2} v_2 \end{pmatrix}$$

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

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

where we have already noted that ρ 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 δ_1 and δ_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

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

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},

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

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},

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

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

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

Appendix 1 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 β, β_i and β_j, 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 $\frac{v_{jj}}{v_{ii}}$, $\frac{v_{ij}}{v_{ii}}$, and all the elements scaled by $\sqrt{\frac{v_{jj}}{v_{ii}}}$. 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 $(\widehat{a}(\delta_1), \cdots, \widehat{a}(\delta_m))$ as a solution to a set of nonlinear equations.

Specifically, define

$$D = \begin{pmatrix} \delta_1 & \delta_2 & \cdots & \delta_m \end{pmatrix} \qquad \text{and} \qquad C = \begin{pmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_m \end{pmatrix}. \tag{7}$$

The covariance matrix for the m estimators is then

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

which implies that

$$\left(C'(I \otimes H) C\right) \Omega \left(C'(I \otimes H) C\right) = \left(D'VD\right) \tag{8}$$

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[1] as long as D contains all vectors of the from e_j, $e_j + e_k$ and $e_j - e_k$.

2.2 GMM

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

$$E\left[f(x_i, \theta_0)\right] = 0$$

where x_i is "data for observation i" and it is assumed that this defines a unique θ_0. The GMM estimator for θ_0 is

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

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

[1] Except for scale.

has the form

$$\Sigma = \left(\Gamma' W_0 \Gamma\right)^{-1} \Gamma' W_0 S W_0 \Gamma \left(\Gamma' W_0 \Gamma\right)^{-1}$$

where W_0 is the probability limit of W_n, $S = V\left[f\left(x_i, \theta_0\right)\right]$ and $\Gamma = \frac{\partial}{\partial \theta'} E\left[f\left(x_i, \theta_0\right)\right]$. Hahn (1996) showed that the limiting distribution of the GMM estimator can be estimated by the bootstrap.

Now let δ be some fixed vector and consider the problem of estimating a scalar parameter, α, from

$$E\left[f\left(x_i, \theta_0 + \alpha\delta\right)\right] = 0$$

by

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

The asymptotic variance of two such estimators corresponding to different δ would be

$$\Omega_{\delta_1,\delta_2} = avar\left(\left(\begin{array}{c} \widehat{a}\left(\delta_1\right) \\ \widehat{a}\left(\delta_2\right) \end{array}\right)\right) \tag{9}$$

$$= \left(\begin{array}{cc} \left(\delta_1' \Gamma' W_0 \Gamma \delta_1\right)^{-1} \delta_1' \Gamma' W_0 S W_0 \Gamma \delta_1 \left(\delta_1' \Gamma' W_0 \Gamma \delta_1\right)^{-1} & \left(\delta_1' \Gamma' W_0 \Gamma \delta_1\right)^{-1} \delta_1' \Gamma' W_0 S W_0 \Gamma \delta_2 \left(\delta_2' \Gamma' W_0 \Gamma \delta_2\right)^{-1} \\ \left(\delta_1' \Gamma' W_0 \Gamma \delta_1\right)^{-1} \delta_1' \Gamma' W_0 S W_0 \Gamma \delta_2 \left(\delta_2' \Gamma' W_0 \Gamma \delta_2\right)^{-1} & \left(\delta_2' \Gamma' W_0 \Gamma \delta_2\right)^{-1} \delta_2' \Gamma' W_0 S W_0 \Gamma \delta_2 \left(\delta_2' \Gamma' W_0 \Gamma \delta_2\right)^{-1} \end{array}\right)$$

Of course (9) has exactly the same structure as (3) and we can therefore back out the matrices $\Gamma' W_0 \Gamma$ and $\Gamma' W_0 S W_0 \Gamma$ (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 $\widehat{a}\left(\delta\right)$ in this GMM setting is proved in Appendix 4.

2.3 Method of Moments

We next consider the just identified case where the number of parameters equals the number of moments. In this case, the weighting matrix plays no role for the asymptotic distribution of the estimator. Specifically, the asymptotic variance is

$$\Sigma = \left(\Gamma^{-1}\right) S \left(\Gamma^{-1}\right)'$$

This is very similar to the expression for the asymptotic variance of the extremum estimator. The difference is that the Γ matrix is typically only symmetric if the moment condition corresponds to the first-order condition for an optimization problem.

We first note that there is no loss of generality in normalizing the diagonal elements of S to 1. Now consider the $\widehat{\alpha}_{k\ell}$ that solves the k'th moment with respect to the ℓ'th element of the parameter,

$$\frac{1}{n}\sum_{i=1}^{n} f_k\left(x_i, \theta_0 + \widehat{\alpha}_{k\ell} e_\ell\right) \approx 0$$

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

$$Acov\left(\widehat{\alpha}_{k\ell}, \widehat{\alpha}_{jm}\right) = \frac{S_{kj}}{\gamma_{k\ell}\gamma_{jm}}$$

where S_{kj} and γ_{jk} denote the elements in S and Γ. In particular

$$Avar\left(\widehat{\alpha}_{kk}\right) = \frac{S_{kk}}{\gamma_{kk}^2} = \frac{1}{\gamma_{kk}^2}$$

Since the moment conditions are invariant to sign–changes, there is no loss in generality in assuming $\gamma_{kk} > 0$. Hence γ_{kk} is identified. Since

$$Acov\left(\widehat{\alpha}_{kk}, \widehat{\alpha}_{jj}\right) = \frac{S_{kj}}{\gamma_{kk}\gamma_{jj}},$$

S_{kj} is identified as well.

Finally

$$Acov\left(\widehat{\alpha}_{kk}, \widehat{\alpha}_{jm}\right) = \frac{S_{kj}}{\gamma_{kk}\gamma_{jm}}$$

so γ_{jm} is also identified.

2.4 Indirect Inference

Simulation-based inference has become increasingly popular as a way to estimate complicated structural econometric models. See Smith (2008) for an introduction and Gourieroux and Monfort (2007) for a textbook treatment. These models often result in simulation moments that are discontinuous functions of the parameters. In this case, a given bootstrap replication should use the same draws of the unobservables for the calculation of all δ.

2.5 Two-Step Estimators

Finite dimensional two–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 above can therefore in principle be applied. However, the one-dimensional estimation used in the bootstrap does not preserve the simplicity of the two-step structure. In this

section we therefore propose a version of the poor (wo)man's bootstrap that is suitable for two-step estimators.

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

$$\begin{aligned}\widehat{\theta}_1 &= \arg\min_{t_1} \frac{1}{n}\sum Q\left(z_i, t_1\right) \\ \widehat{\theta}_2 &= \arg\min_{t_2} \frac{1}{n}\sum R\left(z_i, \widehat{\theta}_1, t_2\right)\end{aligned}$$

with moment conditions (or limiting first-order conditions),

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

where θ_1 and θ_2 are k_1 and k_2-dimensional parameters of interest and q and r are smooth functions. Although our exposition requires this, the results also apply when one or both steps involve GMM estimation with possibly non-smooth functions.

The estimator $\widehat{\theta} = \left(\widehat{\theta}_1', \widehat{\theta}_2'\right)'$ will have a limiting normal distribution with asymptotic variance

$$\begin{pmatrix} E\left[q_1\left(z_i, \theta_1\right)\right] & 0 \\ E\left[r_1\left(z_i, \theta_1, \theta_2\right)\right] & E\left[r_2\left(z_i, \theta_1, \theta_2\right)\right] \end{pmatrix}^{-1} V \begin{bmatrix} q\left(z_i, \theta_1\right) \\ r\left(z_i, \theta_1, \theta_2\right) \end{bmatrix} \left(\begin{pmatrix} E\left[q_1\left(z_i, \theta_1\right)\right] & 0 \\ E\left[r_1\left(z_i, \theta_1, \theta_2\right)\right] & E\left[r_2\left(z_i, \theta_1, \theta_2\right)\right] \end{pmatrix}^{-1} \right)'.$$

This 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. Unfortunately, this is not necessarily convenient because the poor (wo)man's bootstrap would use the bootstrap sample to estimate scalar a where $\theta = \left(\theta_1', \theta_2'\right)'$ has been parameterized as $\widehat{\theta} + a\delta$. When δ places weight on elements from both θ_1 and θ_2, the estimation of a no longer benefits from the simplicity of the two-step setup.

Example 1 *Consider the standard sample selection model*

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

where (ν_i, ε_i) *has a bivariate normal distribution.* α *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* β *is then estimated by the coefficients on* x_i *in the regression of* y_i *on* x_i *and* $\lambda_i = \frac{\phi\left(z_i'\widehat{\alpha}_{MLE}\right)}{1-\Phi\left(z_i'\widehat{\alpha}_{MLE}\right)}$ *using only the sample for which* $d_i = 1$*. See Heckman (1979).*

We now demonstrate 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.

We first note that the elements of $E\left[q_1\left(z_i, \theta_1\right)\right]$ and $V\left[q\left(z_i, \theta_1\right)\right]$ can be estimated by applying the poor (wo)man's bootstrap to the first step in the estimation procedure alone. $E\left[r_2\left(z_i, \theta_1, \theta_2\right)\right]$ and $V\left[r\left(z_i, \theta_1, \theta_2\right)\right]$ can be estimated by applying the poor (wo)man's bootstrap to the second step of the estimation procedure holding $\widehat{\theta}_1$ fixed.

To estimate the elements of $E\left[r_1\left(z_i, \theta_1, \theta_2\right)\right]$ and $cov\left[q\left(z_i, \theta_1\right), r\left(z_i, \theta_1, \theta_2\right)\right]$, consider the three infeasible scalar estimators

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

for fixed δ_1, δ_2 and δ_3.

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

$$\begin{pmatrix}
\delta_1' E\left[q_1\left(z_i, \theta_1\right)\right]\delta_1 & 0 & 0 \\
\delta_1' E\left[r_1\left(z_i, \theta_1, \theta_2\right)\right]\delta_2 & \delta_2' E\left[r_2\left(z_i, \theta_1, \theta_2\right)\right]\delta_2 & 0 \\
0 & 0 & \delta_3' E\left[r_2\left(z_i, \theta_1, \theta_2\right)\right]\delta_3
\end{pmatrix}^{-1}$$

$$\begin{pmatrix}
\delta_1' V\left[q\left(z_i, \theta_1\right)\right]\delta_1 & \delta_1' cov\left[q\left(z_i, \theta_1\right), r\left(z_i, \theta_1, \theta_2\right)\right]\delta_2 & \delta_1' cov\left[q\left(z_i, \theta_1\right), r\left(z_i, \theta_1, \theta_2\right)\right]\delta_3 \\
\delta_1' cov\left[q\left(z_i, \theta_1\right), r\left(z_i, \theta_1, \theta_2\right)\right]\delta_2 & \delta_2' V\left[r\left(z_i, \theta_1, \theta_2\right)\right]\delta_2 & \delta_2' V\left[r\left(z_i, \theta_1, \theta_2\right)\right]\delta_3 \\
\delta_1' cov\left[q\left(z_i, \theta_1\right), r\left(z_i, \theta_1, \theta_2\right)\right]\delta_3 & \delta_2' V\left[r\left(z_i, \theta_1, \theta_2\right)\right]\delta_3 & \delta_3' V\left[r\left(z_i, \theta_1, \theta_2\right)\right]\delta_3
\end{pmatrix}$$

$$\begin{pmatrix}
\delta_1' E\left[q_1\left(z_i, \theta_1\right)\right]\delta_1 & 0 & 0 \\
\delta_1' E\left[r_1\left(z_i, \theta_1, \theta_2\right)\right]\delta_2 & \delta_2' E\left[r_2\left(z_i, \theta_1, \theta_2\right)\right]\delta_2 & 0 \\
0 & 0 & \delta_3' E\left[r_2\left(z_i, \theta_1, \theta_2\right)\right]\delta_3
\end{pmatrix}^{-1}.$$

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

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

which can be written as

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

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 elements of this matrix by $\omega_{\ell k}$ we have

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

There is no loss in generality in normalizing

$$r_2 = 1$$

so now we know r_1 and ρ. We also know v from ω_{33}.

This implies that the asymptotic variance of $(\widehat{a}_1, \widehat{a}_2, \widehat{a}_3)$ identifies $\delta_1' V\left[q\left(z_i, \theta_1\right), r\left(z_i, \theta_1, \theta_2\right)\right]\delta_2$ and $\delta_1' E\left[r_1\left(z_i, \theta_1, \theta_2\right)\right]\delta_2$. Choosing $\delta_1 = e_\ell$ and $\delta_2 = e_k$ recovers all the elements of $cov\left[q\left(z_i, \theta_1\right), r\left(z_i, \theta_1, \theta_2\right)\right]$ and $E\left[r_1\left(z_i, \theta_1, \theta_2\right)\right]$.

2.6 Alternative Approach

In this subsection we present an alternative insight for estimating a variance matrix by bootstrapping one-dimensional components of the parameter vector.

For simplicity, consider the extremum estimator discussed in Section 2.1. Standard asymptotic theory implies that in each boostrap replication, b, the estimator of the full parameter vector has the linear representation

$$\widehat{\theta}_b - \widehat{\theta} \doteq H^{-1} s_b \tag{10}$$

where H is fixed and s_b is approximately normal and i.i.d. across bootstrap samples. The directional estimator defined in (2) has the corresponding representation

$$\widehat{a}_b(\delta) \doteq \left(\delta' H \delta\right)^{-1} \delta' s_b. \tag{11}$$

We write (11) as

$$\left(\delta' H \delta\right) \widehat{a}_b(\delta) \doteq \delta' s_b$$

or equivalently

$$\widehat{a}_b(\delta) \left(\delta' H \delta\right) - \delta' s_b \doteq 0$$

or

$$\sum_{k,\ell} \left(\widehat{a}_b(\delta) \delta(k) \delta(\ell)\right) h_{k\ell} - \sum_k \delta(k) s_b(k) \doteq 0 \tag{12}$$

where $s_b(k)$ is the k'th element of s_b, $\delta(k)$ is the k'th element of δ and $h_{k\ell} = h_{\ell k}$.

It is useful to think of (12) as a linear regression model where the dependent variable is always 0 and (asymptotically) there is no error. In each bootstrap replication, each δ-vector gives an observation from (12). The s-vector differs across bootstrap replications, but the elements of H are the same. In other words, if we focus on H, we can think of the s's as bootstrap-specific fixed effects, which could be eliminated by a transformation similar to the "textbook" panel data deviations-from-means transformation. This gives one an easy way to estimate the elements of H (up to scale).

Once H has been estimated, one could back out the s for each bootstrap replication and use the sample variance for s as an estimate of V. Alternatively, we can use the estimate of H obtained from (12) and then exploit

$$\left(C'(I \otimes H) C\right) \Omega \left(C'(I \otimes H) C\right) = \left(D' V D\right) \tag{13}$$

to back out V.

This approach relies on the linear representations in (10) and (11). As such, it applies to the GMM setting and the two-step estimators as well as to the extremum estimators.[2]

[2]For extremum estimators, V is essentially the variance of q'. There are many examples in which it is easy to

One potential advantage of exploiting (12) to recover H and s_b is that it is straightforward to allow the directions δ to differ across replications. We plan to explore this in future work.

3 Implementation

There are many ways to turn the identification strategy above into estimation of[3] H and V. One is to pick a set of δ–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

$$\min_{V,H} \sum_{ij} \left(\left\{ \left(C'\left(I \otimes H\right) C\right) \widehat{\Omega} \left(C'\left(I \otimes H\right) C\right) - \left(D'VD\right) \right\}_{ij} \right)^2 \tag{14}$$

where D and C are defined in (7), $V_{11} = 1$, and V and H are positive definite matrices.

From a computational point of view, it can be time–consuming to recover the estimates of V and H by a nonlinear minimization problem. We therefore illustrate the usefulness of our approach by estimating V and H along the lines of the identification proof.

For all i, j, we estimate $y_{ij} = V_{jj}/V_{ii}$ exactly as prescribed by the identification. Taking logs, this gives a set of equations of the form

$$\log\left(y_{ij}\right) = \sum_k \alpha_k 1\left(k = j\right) - \alpha_k 1\left(k = i\right)$$

where $\alpha_1 = 0$ (because $V_{11} = 1$) and $\alpha_k = \log\left(V_{kk}\right)$. We can estimate the vector of α's by regression $\log\left(y_{ij}\right)$ on a set of dummy variables. This gives estimates of the diagonal elements of V. The correlation structure in V is the same as the correlation structure in the variance of $\left(\widehat{a}\left(e_1\right), \cdots, \widehat{a}\left(e_k\right)\right)$.

To estimate H we first use that $Avar\left(\widehat{a}\left(e_j\right)\right) = \frac{V_{jj}}{h_{jj}^2}$. Since H is positive definite, we therefore estimate h_{jj} by $\sqrt{\widehat{V}_{jj} \Big/ Avar\left(\widehat{a}\left(e_j\right)\right)}$.

To estimate the off-diagonal elements, h_{ij}, we use the estimated covariances between $\widehat{a}\left(e_i\right)$ and $\widehat{a}\left(e_i + e_j\right)$, between $\widehat{a}\left(e_i\right)$ and $\widehat{a}\left(e_i - e_j\right)$, between $\widehat{a}\left(e_j\right)$ and $\widehat{a}\left(e_i + e_j\right)$, and between $\widehat{a}\left(e_j\right)$ and $\widehat{a}\left(e_i - e_j\right)$.

---

estimate V directly, but difficult to estimate H. The least absolute deviations estimator is a classic example of this. In those cases, (12) can be used to estimate H.

[3] Here we use the notation for extremum estimators. The same discussion applies to GMM estimators.

Specifically, the asymptotic covariance between $\widehat{a}(e_i)$ and $\widehat{a}(e_i + e_j)$ is

$$k_2 = h_{ii}^{-1}(v_{ii} + v_{ij})(h_{ii} + 2h_{ij} + h_{jj})^{-1}$$

(see equation (3). We write this as

$$k_2 h_{ii}(h_{ii} + 2h_{ij} + h_{jj}) = (v_{ii} + v_{ij})$$

or

$$v_{ii} + v_{ij} - k_2 h_{ii}^2 - k_2 h_{ii} h_{jj} = 2k_2 h_{ii} h_{ij} \tag{15}$$

Now consider the asymptotic covariance between $\widehat{a}(e_i)$ and $\widehat{a}(e_i - e_j)$:

$$k_3 = h_{ii}^{-1}(v_{ii} - v_{ij})(h_{ii} - 2h_{ij} + h_{jj})^{-1}$$

or

$$v_{ii} - v_{ij} - k_3 h_{ii}^2 - k_3 h_{ii} h_{jj} = -2k_3 h_{ii} h_{ij} \tag{16}$$

Next consider the asymptotic covariance between $\widehat{a}(e_j)$ and $\widehat{a}(e_i + e_j)$:

$$k_4 = h_{jj}^{-1}(v_{jj} + v_{ij})(h_{ii} + 2h_{ij} + h_{jj})^{-1}$$

or

$$v_{jj} + v_{ij} - k_4 h_{jj}^2 - k_4 h_{ii} h_{jj} = 2k_4 h_{jj} h_{ij} \tag{17}$$

Finally consider the asymptotic covariance between $\widehat{a}(e_j)$ and $\widehat{a}(e_i - e_j)$:

$$k_5 = h_{jj}^{-1}(-v_{jj} + v_{ij})(h_{ii} - 2h_{ij} + h_{jj})^{-1}$$

or

$$-v_{jj} + v_{ij} - k_5 h_{jj}^2 - k_5 h_{ii} h_{jj} = -2k_5 h_{jj} h_{ij} \tag{18}$$

Writing (15)–(18) in vector notation

$$\begin{pmatrix} v_{ii} + v_{ij} - k_2 h_{ii}^2 - k_2 h_{ii} h_{jj} \\ v_{ii} - v_{ij} - k_3 h_{ii}^2 - k_3 h_{ii} h_{jj} \\ v_{jj} + v_{ij} - k_4 h_{jj}^2 - k_4 h_{ii} h_{jj} \\ -v_{jj} + v_{ij} - k_5 h_{jj}^2 - k_5 h_{ii} h_{jj} \end{pmatrix} = \begin{pmatrix} 2k_2 h_{ii} \\ -2k_3 h_{ii} \\ 2k_4 h_{jj} \\ -2k_5 h_{jj} \end{pmatrix} h_{ij} \tag{19}$$

The off-diagonal element h_{ij} could then be estimated by regressing the vector on the left-hand side (y) on the vector on the right-hand side (x). To lessen the influence of any one of the four equations, we use weighted regression where the weight is $\frac{1}{\sqrt{|x_\ell|}}$.

It is worth noting that (19) does not contain all the "linear" information about the off-diagonal elements, h_{ij}. Consider, for example, any two vectors δ_p and δ_q and their associated $\widehat{a}(\delta_p)$ and $\widehat{a}(\delta_q)$, ω_{pq}:

$$\omega_{pq} = acov\left(\widehat{a}(\delta_p), \widehat{a}(\delta_q)\right) = \left(\delta_p' H \delta_p\right)^{-1} \delta_p' V \delta_q \left(\delta_q' H \delta_q\right)^{-1}$$

or

$$\begin{aligned} \delta_p' V \delta_q &= \left(\sum_{ij} \delta_{pi}\delta_{pj} h_{ij}\right) \omega_{pq} \left(\sum_{k\ell} \delta_{qk}\delta_{q\ell} h_{k\ell}\right) \\ &= \sum_{ijk\ell} \delta_{pi}\delta_{pj}\delta_{qk}\delta_{q\ell}\omega_{pq} h_{ij} h_{k\ell} \end{aligned}$$

This gives a quadratic system. However, by restricting attention to $\delta_q = e_k$, we get

$$\delta_p' V \delta_q - \sum_i \delta_{pi}\delta_{pi}\delta_{qk}\delta_{q\ell}\omega_{pq} h_{ii} h_{kk} = \sum_{i \neq j} \delta_{pi}\delta_{pj}\delta_{qk}\delta_{q\ell}\omega_{pq} h_{ij} h_{kk}$$

This is linear in the h_{ij}'s.

4 Illustrations

4.1 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, $\widetilde{x}_i$ with means equal to 0, variances equal to 1 and all covariances equal to $\frac{1}{2}$. x_{i1} to x_{i9} are then $x_{ij} = 1\{\widetilde{x}_{ij} \geq 0\}$ for $j = 1 \cdots 3$, $x_{ij} = \widetilde{x}_{ij} + 1$ for $j = 4$ to 6, $x_{i7} = \widetilde{x}_{i7}$, $x_{i8} = \widetilde{x}_{i8}/2$ and $x_{i9} = 10\widetilde{x}_{i9}$. Finally $x_{i10} = 1$. ε_i is normally distributed conditional on x_i and with variance $(1 + x_{i1})^2$. We pick $\beta = \left(\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1, 0, 0, 0, 0, 0\right)$. 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 (14) by nonlinear least squares (N), and the variance estimator based on estimating V and H from (19) by OLS (L). All 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.

To explore the sensitivity of the approach to the dimensionality of the parameter, we also consider a design with 10 additional regressors, all generated like $\widetilde{x}_i$ and with true coefficients equal to 0. For this design, we do not yet calculate the variance estimators based on (14) by nonlinear least squares (N). The results are in Table 3.

Tables 1-3 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 (testing the true parameter values) based on our approach and on the regular bootstrap is smaller than the difference between the t-statistics based on the bootstrap and the Eicker-Huber-White variance estimator.

4.2 Maximum Rank Correlation Estimator

Han (1987) and Cavanagh and Sherman (1998) defined maximum rank correlation estimators for β in the model

$$y_i = g\left(f\left(x_i'\beta, \varepsilon_i\right)\right)$$

where β is a k–dimensional parameter of interest, f is strictly increasing in each of its arguments and g is increasing. This model includes many single-equation econometric models as special cases.

The estimator proposed by Han (1987) maximizes Kendall's rank correlation between y_i and $x_i'b$:

$$\widehat{\beta} = \arg\max_b \sum_{i<j} \left(1\left\{y_i > y_j\right\} - 1\left\{y_i < y_j\right\}\right)\left(1\left\{x_i'b > x_j'b\right\} - 1\left\{x_i'b < x_j'b\right\}\right)$$

The asymptotic distribution of this estimator was derived in Sherman (1993). Specifically, he showed that with[4] $\beta' = \left(\theta', 1\right)'$, $\widehat{\theta}$ will have a limiting normal distribution of the form considered in Section 2.1:

$$\sqrt{n}\left(\widehat{\theta} - \theta\right) \xrightarrow{d} N\left(0, H^{-1}VH^{-1}\right)$$

[4]Since f is unspecified, it is clear that some kind of scale normalization is necessary.

where

$$H = \frac{1}{2}E\left[\widetilde{S}_2\left(y, x'\beta\right) g_0\left(x'\beta\right)\left(x_0 - \overline{x}_0\right)\left(x_0 - \overline{x}_0\right)'\right],$$
$$V = E\left[\widetilde{S}\left(y, x'\beta\right)^2 g_0\left(x'\beta\right)^2\left(x_0 - \overline{x}_0\right)\left(x_0 - \overline{x}_0\right)'\right]$$

with[5] $\widetilde{S}(y_0, t) = E\left[1\left\{y_0 > y\right\} - 1\left\{y_0 < y\right\} | x'\beta = t\right]$, $\widetilde{S}_2(y_0, t) = \frac{\partial \widetilde{S}(y_0, t)}{\partial t}$, and $\overline{x}_0 = E\left[x_0 | x'\beta\right]$, where x_0 is the first $k-1$ elements of x (i.e., the elements associated with θ) and g_0 is the marginal density of $x'\beta$.

As mentioned above, Han's (1987) estimator maximizes Kendall's rank correlation between y_i and $x_i'b$. Cavanagh and Sherman (1998) proposed an alternative estimator of β based on maximizing

$$\sum_{i=1}^{n} M\left(y_i\right) R_n\left(x_i'b\right)$$

where $M\left(\cdot\right)$ is an increasing function and $R_n\left(x_i'b\right) = \sum_{j=1}^{n} 1\left\{x_i'b > x_j'b\right\}$ is the rank of $x_i'b$ in the set $\left\{x_j'b : j = 1, ..., n\right\}$. When $M\left(\cdot\right) = R_n\left(\cdot\right)$, the objective function is a linear function of Spearman's rank correlation. In that case the objective function is

$$\sum_{i=1}^{n}\left(\sum_{k=1}^{n} 1\left\{y_i > y_k\right\}\right)\left(\sum_{j=1}^{n} 1\left\{x_i'b > x_j'b\right\}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} 1\left\{y_i > y_k\right\} 1\left\{x_i'b > x_j'b\right\} \tag{20}$$

The estimator proposed by Cavanagh and Sherman (1998) is also asymptotically normal,

$$\sqrt{n}\left(\widehat{\theta} - \theta\right) \xrightarrow{d} N\left(0, H_1^{-1} V_1 H_1^{-1}\right)$$

where $\beta' = \left(\theta', 1\right)'$ and H_1 and V_1 have a structure similar to H and V. See Appendix 2.

Direct estimation of H and V (or H_1 and V_1) requires nonparametric estimation. It is therefore tempting to instead estimate $Avar\left(\widehat{\theta}\right)$ (or $Avar\left(\widetilde{\theta}\right)$) by the bootstrap. On the other hand, the maximum rank correlation estimators are cumbersome to calculate in higher dimensions, which can make this approach problematic in practice. The approach suggested in this paper is therefore potentially useful.

To investigate this, we consider a relatively simple data-generating process with

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

and only four explanatory variables generated along the lines of the explanatory variables in section 4.1: For each observation, i, we first generate $\widetilde{x}_{ij}$ with means equal to 0, variances equal to 1 and all

[5] With the exception of V and H the notation here is chosen to make it as close as possible to that in Sherman (1993).

covariances equal to $\frac{1}{2}$. We then define $x_{ij} = \widetilde{x}_{ij}$ for $j = 1 \cdots 2$, $x_{i3} = 1\{\widetilde{x}_{i3} \geq 0\}$, and $x_{i4} = \widetilde{x}_{i4} + 1$. The error, ε_i, is normal with mean 0 and variance 1.5^2. A normalization is needed since the maximum rank correlation estimator only estimates β up to scale. Two natural normalizations are $\|\beta\| = 1$ and $\beta_1 = 1$. One might fear that the quality of the normal approximation suggested by the asymptotic distribution will depend on which normalization one applies. Since this issue is unrelated to the contribution of this paper, we use $\beta = (1, 0, 0, 0)'$ and estimate with the normalization that $\beta_1 = 1$. The low dimension of β makes it possible to estimate the variance of $\widehat{\beta}$ by the usual bootstrap and compare the results to the ones obtained by the approach proposed here. For now, we only consider the estimator defined by minimizing (20).

Table 4 compares the t–statistics based on the bootstrap estimator of the variance of $\widehat{\theta}$, the variance estimator based on estimating V_1 and H_1 from (14) by nonlinear least squares (N), and the variance estimator based on estimating V_1 and H_1 from (19) by OLS (L). We use sample sizes of 200 and 500 and the results presented here are based on 400 Monte Carlo replications, each using 400 bootstrap samples to calculate the standard errors. Compared to the linear regression model, there is a bigger difference between the t–statistics based on our approach and those based on the usual bootstrap. However, the differences are small enough that they are unlikely to be of serious consequence in empirical applications.

While an applied researcher would primarily be interested in the effect of the various bootstrap methods on the resulting t-statistics, it is also interesting to investigate how precisely they estimate the asymptotic standard errors of the estimators. To address this, we calculate the standard error of the estimator suggested by the asymptotics using the expression provided in Cavanagh and Sherman (1998). See Appendix 2. We then compare this to the standard deviation of the estimator as well as to the average standard errors based on the three bootstrap methods. The results are presented in Table 5. Interestingly, it seems that our approach does a better job of approximating the asymptotic variance than does the usual bootstrap. We suspect that the reason is that our approach implicitly assumes that the asymptotics provide a good approximation for *one*–dimensional estimation problems.

4.3 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 (2014).

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

$$\log(w_{si2}) = x_{si}'\beta_s + 1\{d_{i1} = s\}\gamma_s + \varepsilon_{si2}$$

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

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

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

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

This gives closed-form solutions for $w_{1i1} + \rho E\left[\max\{w_{1i2}, w_{2i2}\} | x_{1i}, x_{2i}, d_{i1} = 1\right]$ and $w_{2i1} + \rho E\left[\max\{w_{1i2}, w_{2i2}\} | x_{1i}, x_{2i}, d_{i2} = 1\right]$.

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 $\beta_1 = (1, 1)'$, $\beta_2 = \left(\frac{1}{2}, 1\right)'$, $\gamma_1 = 0$ and $\gamma_2 = 1$. Finally, $\sigma_1^2 = 2$, $\sigma_2^2 = 3$, $\tau = 0$ and $\rho = 0.95$. In the estimation, we treat ρ and τ 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 400 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 $\widehat{a}$ be the vector of those parameters based on the data and let $\widehat{V}[\widehat{\alpha}]$ be the associated estimated variance. For a candidate vector of structural parameters, θ, the researcher simulates the model R times (holding the draws of the errors constant across different values of θ), calculates the associated $\widetilde{\alpha}(\theta)$ and estimates the model parameters by minimizing

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

over θ.

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. Since the results for maximum rank correlation suggest that the nonlinear version outperforms the linear version, we do not consider the latter here. Table 6 presents 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. The computation time makes it infeasible to perform a Monte Carlo study that includes the usual bootstrap method.

5 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, and in which the regular bootstrap mothod is computationally infeasible.

References

Altonji, J. G., A. A. Smith, and I. Vidangos (2013): "Modeling Earnings Dynamics," *Econometrica*, 81(4), 1395–1454.

Cavanagh, C. L., and R. P. Sherman (1998): "Rank Estimators of Monotonic Index Models," *Journal of Econometrics*, 84, 351–381.

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.

Han, A. (1987): "Nonparametric Analysis of a Generalized Regression Model," *Journal of Econometrics*, 35, 303–316.

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.

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 (2014): "Identification in a Dynamic Roy Model," *In Preparation*.

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.

Newey, 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.

Sherman, R. T. (1993): "The Limiting Distribution of the Maximum Rank Correlation Estimator," *Econometrica*, 61, 123–137.

Smith, A. A. (2008): "Indirect Inference," in *The New Palgrave Dictionary of Economics*, ed. by S. Durlauf, and L. Blume. Macmillan, 2 edn.

Appendix 1: Non-Singularity of the Matrix in Equation (6)

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

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

since H is positive definite.

Appendix 2: Calculating the Asymptotic Variance of the MRC Estimator

Following Cavanagh and Sherman (1998), let

$$f\left(\left(y_1,x_1\right),\left(y_2,x_2\right),\left(y_3,x_3\right),b\right)=1\left\{y_1>y_3\right\}1\left\{x_1'b>x_2'b\right\}$$

and let

$$\tau\left(\left(y,x\right),b\right)=E\left[1\left\{y>Y_3\right\}1\left\{x'b>X_2'b\right\}\right]+E\left[1\left\{Y_1>Y_3\right\}1\left\{X_1'b>x'b\right\}\right]+E\left[1\left\{Y_1>y\right\}1\left\{X_1'b>X_2'b\right\}\right]$$

The asymptotic variance of the estimator is

$$9V^{-1}\Delta V^{-1}$$

where

$$\Delta = V\left[\frac{\partial}{\partial b}\tau\left(\left(Y,X\right),\beta\right)\right]$$

and

$$V = E\left[\frac{\partial^2}{\partial b\partial b'}\tau\left(\left(Y,X\right),\beta\right)\right]$$

To calculate the analytical standard errors in Table 5, we draw a sample of size 100,000. For each observation we estimate $\tau\left(\left(y,x\right),b\right)$ by

$$\begin{aligned}\widehat{\tau}\left(\left(y,x\right),b\right) &= \frac{1}{n^2}\sum_{j,k}1\left\{y>y_j\right\}1\left\{x'b>x_k'b\right\}+\frac{1}{n^2}\sum_{j,k}1\left\{y_k>y_j\right\}1\left\{x_k'b>x'b\right\}\\ &\quad+\frac{1}{n^2}\sum_{j,k}1\left\{y_k>y\right\}1\left\{x_k'b>x_j'b\right\}\\ &\doteq \frac{1}{n^2}R_n\left(y\right)R_n\left(x'b\right)+\frac{1}{n^2}\sum_{k}R_n\left(y_k\right)1\left\{x_k'b>x'b\right\}\\ &\quad+\frac{1}{n^2}\sum_{k}1\left\{y_k>y\right\}R_n\left(x_k'b\right).\end{aligned}$$

We then numerically differentiate $\widehat{\tau}\left(\left(y,x\right),b\right)$ twice (using a step-size[6] of 0.01). This yields estimates of Δ and V.

Appendix 3: Maximum of Two Lognormals

The following is taken from Kotz, Balakrishnan, and Johnson (2000).

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

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

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

[6] Changing the step-size to 0.05 changes the asymptotic standard errors by approximately 2%.

Therefore

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

Appendix 4: 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). 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 θ_0 is the unique solution to $G\left(t\right) = 0$ where $G\left(t\right) \equiv E\left[g\left(Z_i, t\right)\right]$, Z_i is the vector of data for observation i and g is a known function. The parameter space is Θ.

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

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

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 τ_n but from a bootstrap sample $\left\{\widehat{Z}_{n1},\ldots,\widehat{Z}_{nn}\right\}$. Specifically

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

where $\widehat{G}_n(t) \equiv \frac{1}{n}\sum_{i=1}^{n} g\left(\widehat{Z}_{ni},t\right)$. $\widehat{A}_n$ is constructed from $\left\{\widehat{Z}_{ni}\right\}_{i=1}^{n}$ 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) θ_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) $\sup_t|G_n(t)-G(t)|\longrightarrow 0$ *as* $n\longrightarrow\infty$ *a.s.;*

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

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

(vii) $|A_nG_n(\tau_n)|\leq o_p(1)+\inf_t|A_nG_n(t)|$ *and* $\left|\widehat{A}_n\widehat{G}_n(\widehat{\tau}_n)\right|\leq o_B(1)+\inf_t\left|\widehat{A}_n\widehat{G}_n(t)\right|$

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

Theorem 1 (Hahn Theorem 1) *Assume that*

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

(ii) $|A_nG_n(\tau_n)|\leq o_p\left(n^{-1/2}\right)+\inf_t|A_nG_n(t)|$ *and* $\left|\widehat{A}_n\widehat{G}_n(\widehat{\tau}_n)\right|\leq o_B\left(n^{-1/2}\right)+\inf_t\left|\widehat{A}_n\widehat{G}_n(t)\right|$*;*

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

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

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

(v) $G(t)$ *is differentiable at* θ_0*, an interior point of the parameter space,* Θ*, with derivative* Γ *with full rank; and*

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

Then

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

and

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

where

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

and

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

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 θ_0, which we partition as $\theta_0' = \left(\theta_0^1, \theta_0^2\right)$.

The infeasible estimator of θ_0, which assumes that θ_0^2 is known, is

$$\gamma_n = \arg\min_t \left| A_n G_n \left(\begin{pmatrix} t \\ \theta_0^2 \end{pmatrix} \right) \right| \tag{21}$$

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 θ_0^1 and θ_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

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

be the usual GMM estimator of θ_0. We consider the bootstrap estimator

$$\widehat{\gamma}_n = \arg\min_t \left| \widehat{A}_n \widehat{G}_n \begin{pmatrix} t \\ \widehat{\theta}^2 \end{pmatrix} \right| \tag{22}$$

where $\widehat{G}_n(t) \equiv \frac{1}{n}\sum_{i=1}^n g\left(\widehat{Z}_{ni}, t\right)$. $\widehat{A}_n$ is constructed from $\left\{\widehat{Z}_{ni}\right\}_{i=1}^n$ 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 $\widehat{\gamma}_n$ can be used to approximate the distribution of γ_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 $\widehat{\gamma}_n$. 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 conditionsa 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 $\widehat{\gamma}_n = \theta_0^1 + o_B(1)$.*

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

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

This follows from

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

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

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

and

$$n^{1/2}\left(\widehat{\gamma}_n - \gamma_n\right) \xRightarrow{p} N(0, \Omega)$$

where

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

and

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

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

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

Combining this with the triangular inequality we have

$$\begin{aligned}
\left|AG\left(\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}\right) - AG\left(\theta_0\right)\right| \leq\ & \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}\right) - AG\left(\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}\right) - \widehat{A}_n\widehat{G}_n\left(\theta_0\right) + AG\left(\theta_0\right)\right| \\
&+ \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}\right) - \widehat{A}_n\widehat{G}_n\left(\theta_0\right)\right| \\
\leq\ & o_B\left(n^{-1/2}\right) + o_B\left(1\right)\left|G\left(\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}\right) - G\left(\theta_0\right)\right| \\
&+ \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}\right) - \widehat{A}_n\widehat{G}_n\left(\theta_0\right)\right| \qquad (24)
\end{aligned}$$

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 (24) and collecting the $G\left(\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}\right) - G\left(\theta_0\right)$ terms yield

$$\begin{aligned}
&\left(C_1 - o_B\left(1\right)\right)\left|G\left(\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}\right) - G\left(\theta_0\right)\right| \qquad (25) \\
\leq\ & o_B\left(n^{-1/2}\right) + \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}\right) - \widehat{A}_n\widehat{G}_n\left(\theta_0\right)\right| \\
\leq\ & o_B\left(n^{-1/2}\right) + \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}\right)\right| + \left|\widehat{A}_n\widehat{G}_n\left(\theta_0\right)\right| \\
\leq\ & o_B\left(n^{-1/2}\right) + \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix} \theta_0^1 \\ \widehat{\theta}^2 \end{pmatrix}\right)\right| + \left|\widehat{A}_n\widehat{G}_n\left(\theta_0\right)\right| \qquad (26)
\end{aligned}$$

Stochastic equicontinuity implies that

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

or

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

so (26) implies

$$\begin{aligned}
&\left(C_1-o_B\left(1\right)\right)\left|G\left(\begin{pmatrix}\widehat{\gamma}_n\\ \widehat{\theta}^2\end{pmatrix}\right)-G\left(\theta_0\right)\right|\\
\leq\;& o_B\left(n^{-1/2}\right)+\left|\widehat{A}_n\left(G\left(\begin{pmatrix}\theta_0^1\\ \widehat{\theta}^2\end{pmatrix}\right)-G\left(\theta_0\right)\right)\right|+2\left|\widehat{A}_n\right|\left|\widehat{G}_n\left(\theta_0\right)\right|+\left|\widehat{A}_n\right|o_B\left(n^{-1/2}\right)\\
\leq\;& o_B\left(n^{-1/2}\right)+\left|\widehat{A}_n\right|\left|\left(G\left(\begin{pmatrix}\theta_0^1\\ \widehat{\theta}^2\end{pmatrix}\right)-G\left(\theta_0\right)\right)\right|+2\left|\widehat{A}_n\right|\left|\widehat{G}_n\left(\theta_0\right)-G_n\left(\theta_0\right)\right|\\
&+2\left|\widehat{A}_n\right|\left|G_n\left(\theta_0\right)\right|+\left|\widehat{A}_n\right|o_B\left(n^{-1/2}\right)\\
=\;& o_B\left(n^{-1/2}\right)+O_B\left(1\right)O_p\left(n^{-1/2}\right)+O_B\left(1\right)O_B\left(n^{-1/2}\right)\\
&+O_B\left(1\right)O_p\left(n^{-1/2}\right)+O_B\left(1\right)o_B\left(n^{-1/2}\right)
\end{aligned}\tag{27}$$

Note that

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

As above, the nonsingularity of Γ implies nonsingularity of ΓE_1, and hence, there exists a constant $C_2>0$ such that $\left|\Gamma E_1 x\right|\geq C_2\left|x\right|$ for all x. Applying this to the equation above and collecting terms give

$$C_2\left|\widehat{\gamma}_n-\theta_0^1\right|\leq\left|\Gamma E_1\left(\widehat{\gamma}_n-\theta_0^1\right)\right|=\left|G\left(\begin{pmatrix}\widehat{\gamma}_n\\ \theta_0^2\end{pmatrix}\right)-G\left(\theta_0\right)\right|+o_B\left(1\right)\left|\widehat{\gamma}_n-\theta_0^1\right|\tag{28}$$

Combining (28) with (27) yields

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

or

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

Part 2: Asymptotic Normality. Let

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

Define

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

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

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

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

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

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

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

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

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

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

$$\left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix} \widehat{\gamma}_n \\ \widehat{\theta}^2 \end{pmatrix}\right) - \widetilde{L}_n(\widehat{\gamma}_n)\right| = o_B\left(n^{-1/2}\right) \tag{29}$$

It follows from Hahn

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

We thus have

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

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

Next, we will show that

$$\left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}\right)-\widetilde{L}_n\left(\widehat{\sigma}_n\right)\right|=o_B\left(n^{-1/2}\right) \tag{30}$$

We have

$$\begin{aligned}\left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}\right)-\widetilde{L}_n\left(\widehat{\sigma}_n\right)\right| &= \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}\right)-A\Gamma\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}-\theta_0\right)-\widehat{A}_n\widehat{G}_n\left(\theta_0\right)\right| \\ &\leq \left|\widehat{A}_n\widehat{G}_n\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}\right)-AG\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}\right)-\widehat{A}_n\widehat{G}_n\left(\theta_0\right)+AG\left(\theta_0\right)\right| \\ &\quad+\left|AG\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}\right)-AG\left(\theta_0\right)-A\Gamma\left(\begin{pmatrix}\widehat{\sigma}_1\\ \widehat{\theta}^2\end{pmatrix}-\theta_0\right)\right| \\ &= o_B\left(n^{-1/2}\right)+o\left(\begin{pmatrix}\widehat{\sigma}_1\\ \widehat{\theta}^2\end{pmatrix}-\theta_0\right) \\ &= o_B\left(n^{-1/2}\right)\end{aligned}$$

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

Combining (29) and (30) with the definitions of $\widehat{\gamma}_n$ and $\widehat{\sigma}_n$ we get

$$\left|\widetilde{L}_n\left(\widehat{\gamma}_n\right)\right|=\left|\widetilde{L}_n\left(\widehat{\sigma}_n\right)\right|+o_B\left(n^{-1/2}\right) \tag{31}$$

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

$$\begin{aligned}
\left|\widetilde{L}_n\left(\widehat{\sigma}_n\right)\right| &\leq \left|A\Gamma\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}-\theta_0\right)\right|+\left|\widehat{A}_n\widehat{G}_n\left(\theta_0\right)\right| \\
&\leq \left|A\Gamma\left(\begin{pmatrix}\widehat{\sigma}_n\\ \widehat{\theta}^2\end{pmatrix}-\begin{pmatrix}\gamma_n\\ \widehat{\theta}^2\end{pmatrix}\right)\right|+\left|\widehat{A}_n\widehat{G}_n\left(\theta_0\right)-\widehat{A}_nG_n\left(\theta_0\right)\right| \\
&\quad +\left|A\Gamma\left(\begin{pmatrix}\gamma_n\\ \widehat{\theta}^2\end{pmatrix}-\theta_0\right)\right|+\left|\widehat{A}_nG_n\left(\theta_0\right)\right| \\
&= O_B\left(n^{-1/2}\right)+O_B\left(1\right)O_B\left(n^{-1/2}\right)+O_p\left(n^{-1/2}\right)+O_B\left(1\right)O_p\left(n^{-1/2}\right) \qquad (32)
\end{aligned}$$

Squaring both sides of (31) we have

$$\left|\widetilde{L}_n\left(\widehat{\gamma}_n\right)\right|^2=\left|\widetilde{L}_n\left(\widehat{\sigma}_n\right)\right|^2+o_B\left(n^{-1}\right) \qquad (33)$$

because (32) implies that the cross-product term can be absorbed in the $o_B\left(n^{-1}\right)$. On the other hand, for any t

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

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

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

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

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

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

Compare this to (33) to conclude that

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

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

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

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

and

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

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

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

which has the form of (21).

Table 1: Ordinary Least Squares, $n = 200$.

	$\lvert T_E - T_B \rvert$	$\lvert T_E - T_N \rvert$	$\lvert T_E - T_L \rvert$	$\lvert T_B - T_N \rvert$	$\lvert T_B - T_L \rvert$	$\lvert T_N - T_L \rvert$
β_1	0.031	0.027	0.025	0.017	0.026	0.014
β_2	0.029	0.023	0.022	0.017	0.025	0.012
β_3	0.031	0.027	0.025	0.018	0.023	0.011
β_4	0.032	0.027	0.026	0.020	0.023	0.011
β_5	0.033	0.026	0.023	0.020	0.021	0.011
β_6	0.032	0.029	0.027	0.022	0.025	0.012
β_7	0.031	0.025	0.024	0.020	0.024	0.010
β_8	0.033	0.027	0.027	0.020	0.018	0.011
β_9	0.034	0.026	0.025	0.021	0.023	0.006
β_{10}	0.033	0.034	0.026	0.018	0.022	0.023

Table 2: Ordinary Least Squares, $n = 2000$.

	$\lvert T_E - T_B \rvert$	$\lvert T_E - T_N \rvert$	$\lvert T_E - T_L \rvert$	$\lvert T_B - T_N \rvert$	$\lvert T_B - T_L \rvert$	$\lvert T_N - T_L \rvert$
β_1	0.025	0.025	0.025	0.004	0.003	0.002
β_2	0.021	0.021	0.021	0.003	0.003	0.002
β_3	0.024	0.024	0.024	0.004	0.003	0.002
β_4	0.023	0.022	0.022	0.004	0.004	0.003
β_5	0.025	0.025	0.025	0.004	0.004	0.003
β_6	0.025	0.025	0.025	0.004	0.004	0.003
β_7	0.026	0.025	0.026	0.004	0.003	0.003
β_8	0.024	0.023	0.023	0.004	0.004	0.003
β_9	0.022	0.023	0.023	0.003	0.003	0.001
β_{10}	0.023	0.023	0.023	0.006	0.005	0.005

Table 3: OLS, $n = 2000$.

	$\lvert T_E - T_B \rvert$	$\lvert T_E - T_L \rvert$	$\lvert T_B - T_L \rvert$
β_1	0.024	0.024	0.005
β_2	0.022	0.023	0.006
β_3	0.022	0.022	0.005
β_4	0.022	0.022	0.005
β_5	0.024	0.024	0.006
β_6	0.022	0.022	0.005
β_7	0.024	0.023	0.005
β_8	0.022	0.022	0.005
β_9	0.024	0.025	0.005
β_{10}	0.022	0.022	0.005
β_{11}	0.021	0.020	0.005
β_{12}	0.024	0.024	0.005
β_{13}	0.021	0.021	0.005
β_{14}	0.021	0.020	0.005
β_{15}	0.024	0.024	0.005
β_{16}	0.023	0.023	0.005
β_{17}	0.023	0.023	0.005
β_{18}	0.022	0.022	0.005
β_{19}	0.022	0.022	0.005
β_{20}	0.021	0.021	0.006

Table 4: Maximum Rank Correlation

	$n = 200$			$n = 500$		
	$\lvert T_B - T_N \rvert$	$\lvert T_B - T_L \rvert$	$\lvert T_N - T_L \rvert$	$\lvert T_B - T_N \rvert$	$\lvert T_B - T_L \rvert$	$\lvert T_N - T_L \rvert$
β_2	0.138	0.091	0.104	0.109	0.059	0.075
β_3	0.129	0.080	0.138	0.109	0.041	0.093
β_4	0.128	0.095	0.073	0.103	0.059	0.058

Table 5: Maximum Rank Correlation

	$n = 200$					$n = 500$				
	Actual	Asymp	Boots	BS_N	BS_L	Actual	Asymp	Boots	BS_N	BS_L
β_2	0.167	0.145	0.189	0.161	0.228	0.099	0.092	0.108	0.094	0.103
β_3	0.317	0.258	0.321	0.294	0.397	0.181	0.163	0.191	0.173	0.189
β_4	0.181	0.142	0.191	0.166	0.180	0.098	0.090	0.107	0.095	0.102

Table 6: Structural Model

	β_{11}	β_{12}	β_{21}	β_{22}	γ_1	γ_2	$\log(\sigma_1)$	$\log(\sigma_2)$
Actual	0.042	0.041	0.049	0.039	0.026	0.063	0.022	0.018
Asymptotic	0.049	0.041	0.051	0.040	0.028	0.068	0.026	0.019
BS_N	0.055	0.042	0.054	0.041	0.032	0.070	0.028	0.019

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.

Working Paper Series *(continued)*

Repos, Fire Sales, and Bankruptcy Policy **WP-12-15**
Gaetano Antinolfi, Francesca Carapella, Charles Kahn, Antoine Martin, David Mills, and Ed Nosal

Speculative Runs on Interest Rate Pegs
The Frictionless Case **WP-12-16**
Marco Bassetto and Christopher Phelan

Institutions, the Cost of Capital, and Long-Run Economic Growth:
Evidence from the 19th Century Capital Market **WP-12-17**
Ron Alquist and Ben Chabot

Emerging Economies, Trade Policy, and Macroeconomic Shocks **WP-12-18**
Chad P. Bown and Meredith A. Crowley

The Urban Density Premium across Establishments **WP-13-01**
R. Jason Faberman and Matthew Freedman

Why Do Borrowers Make Mortgage Refinancing Mistakes? **WP-13-02**
Sumit Agarwal, Richard J. Rosen, and Vincent Yao

Bank Panics, Government Guarantees, and the Long-Run Size of the Financial Sector:
Evidence from Free-Banking America **WP-13-03**
Benjamin Chabot and Charles C. Moul

Fiscal Consequences of Paying Interest on Reserves **WP-13-04**
Marco Bassetto and Todd Messer

Properties of the Vacancy Statistic in the Discrete Circle Covering Problem **WP-13-05**
Gadi Barlevy and H. N. Nagaraja

Credit Crunches and Credit Allocation in a Model of Entrepreneurship **WP-13-06**
Marco Bassetto, Marco Cagetti, and Mariacristina De Nardi

Financial Incentives and Educational Investment:
The Impact of Performance-Based Scholarships on Student Time Use **WP-13-07**
Lisa Barrow and Cecilia Elena Rouse

The Global Welfare Impact of China: Trade Integration and Technological Change **WP-13-08**
Julian di Giovanni, Andrei A. Levchenko, and Jing Zhang

Structural Change in an Open Economy **WP-13-09**
Timothy Uy, Kei-Mu Yi, and Jing Zhang

The Global Labor Market Impact of Emerging Giants: a Quantitative Assessment **WP-13-10**
Andrei A. Levchenko and Jing Zhang

Working Paper Series *(continued)*

Size-Dependent Regulations, Firm Size Distribution, and Reallocation **WP-13-11**
François Gourio and Nicolas Roys

Modeling the Evolution of Expectations and Uncertainty in General Equilibrium **WP-13-12**
Francesco Bianchi and Leonardo Melosi

Rushing into American Dream? House Prices, Timing of Homeownership,
and Adjustment of Consumer Credit **WP-13-13**
Sumit Agarwal, Luojia Hu, and Xing Huang

The Earned Income Tax Credit and Food Consumption Patterns **WP-13-14**
Leslie McGranahan and Diane W. Schanzenbach

Agglomeration in the European automobile supplier industry **WP-13-15**
Thomas Klier and Dan McMillen

Human Capital and Long-Run Labor Income Risk **WP-13-16**
Luca Benzoni and Olena Chyruk

The Effects of the Saving and Banking Glut on the U.S. Economy **WP-13-17**
Alejandro Justiniano, Giorgio E. Primiceri, and Andrea Tambalotti

A Portfolio-Balance Approach to the Nominal Term Structure **WP-13-18**
Thomas B. King

Gross Migration, Housing and Urban Population Dynamics **WP-13-19**
Morris A. Davis, Jonas D.M. Fisher, and Marcelo Veracierto

Very Simple Markov-Perfect Industry Dynamics **WP-13-20**
Jaap H. Abbring, Jeffrey R. Campbell, Jan Tilly, and Nan Yang

Bubbles and Leverage: A Simple and Unified Approach **WP-13-21**
Robert Barsky and Theodore Bogusz

The scarcity value of Treasury collateral:
Repo market effects of security-specific supply and demand factors **WP-13-22**
Stefania D'Amico, Roger Fan, and Yuriy Kitsul

Gambling for Dollars: Strategic Hedge Fund Manager Investment **WP-13-23**
Dan Bernhardt and Ed Nosal

Cash-in-the-Market Pricing in a Model with Money and
Over-the-Counter Financial Markets **WP-13-24**
Fabrizio Mattesini and Ed Nosal

An Interview with Neil Wallace **WP-13-25**
David Altig and Ed Nosal

Working Paper Series *(continued)*

Firm Dynamics and the Minimum Wage: A Putty-Clay Approach **WP-13-26**
Daniel Aaronson, Eric French, and Isaac Sorkin

Policy Intervention in Debt Renegotiation:
Evidence from the Home Affordable Modification Program **WP-13-27**
Sumit Agarwal, Gene Amromin, Itzhak Ben-David, Souphala Chomsisengphet, Tomasz Piskorski, and Amit Seru

The Effects of the Massachusetts Health Reform on Financial Distress **WP-14-01**
Bhashkar Mazumder and Sarah Miller

Can Intangible Capital Explain Cyclical Movements in the Labor Wedge? **WP-14-02**
François Gourio and Leena Rudanko

Early Public Banks **WP-14-03**
William Roberds and François R. Velde

Mandatory Disclosure and Financial Contagion **WP-14-04**
Fernando Alvarez and Gadi Barlevy

The Stock of External Sovereign Debt: Can We Take the Data at 'Face Value'? **WP-14-05**
Daniel A. Dias, Christine Richmond, and Mark L. J. Wright

Interpreting the *Pari Passu* Clause in Sovereign Bond Contracts:
It's All Hebrew (and Aramaic) to Me **WP-14-06**
Mark L. J. Wright

AIG in Hindsight **WP-14-07**
Robert McDonald and Anna Paulson

On the Structural Interpretation of the Smets-Wouters "Risk Premium" Shock **WP-14-08**
Jonas D.M. Fisher

Human Capital Risk, Contract Enforcement, and the Macroeconomy **WP-14-09**
Tom Krebs, Moritz Kuhn, and Mark L. J. Wright

Adverse Selection, Risk Sharing and Business Cycles **WP-14-10**
Marcelo Veracierto

Core and 'Crust': Consumer Prices and the Term Structure of Interest Rates **WP-14-11**
Andrea Ajello, Luca Benzoni, and Olena Chyruk

The Evolution of Comparative Advantage: Measurement and Implications **WP-14-12**
Andrei A. Levchenko and Jing Zhang

Working Paper Series *(continued)*

Saving Europe?: The Unpleasant Arithmetic of Fiscal Austerity in Integrated Economies **WP-14-13**
Enrique G. Mendoza, Linda L. Tesar, and Jing Zhang

Liquidity Traps and Monetary Policy: Managing a Credit Crunch **WP-14-14**
Francisco Buera and Juan Pablo Nicolini

Quantitative Easing in Joseph's Egypt with Keynesian Producers **WP-14-15**
Jeffrey R. Campbell

Constrained Discretion and Central Bank Transparency **WP-14-16**
Francesco Bianchi and Leonardo Melosi

Escaping the Great Recession **WP-14-17**
Francesco Bianchi and Leonardo Melosi

More on Middlemen: Equilibrium Entry and Efficiency in Intermediated Markets **WP-14-18**
Ed Nosal, Yuet-Yee Wong, and Randall Wright

Preventing Bank Runs **WP-14-19**
David Andolfatto, Ed Nosal, and Bruno Sultanum

The Impact of Chicago's Small High School Initiative **WP-14-20**
Lisa Barrow, Diane Whitmore Schanzenbach, and Amy Claessens

Credit Supply and the Housing Boom **WP-14-21**
Alejandro Justiniano, Giorgio E. Primiceri, and Andrea Tambalotti

The Effect of Vehicle Fuel Economy Standards on Technology Adoption **WP-14-22**
Thomas Klier and Joshua Linn

What Drives Bank Funding Spreads? **WP-14-23**
Thomas B. King and Kurt F. Lewis

Inflation Uncertainty and Disagreement in Bond Risk Premia **WP-14-24**
Stefania D'Amico and Athanasios Orphanides

Access to Refinancing and Mortgage Interest Rates: **WP-14-25**
HARPing on the Importance of Competition
Gene Amromin and Caitlin Kearns

Private Takings **WP-14-26**
Alessandro Marchesiani and Ed Nosal

Momentum Trading, Return Chasing, and Predictable Crashes **WP-14-27**
Benjamin Chabot, Eric Ghysels, and Ravi Jagannathan

Early Life Environment and Racial Inequality in Education and Earnings
in the United States **WP-14-28**
Kenneth Y. Chay, Jonathan Guryan, and Bhashkar Mazumder

Working Paper Series *(continued)*

Poor (Wo)man's Bootstrap **WP-15-01**
Bo E. Honoré and Luojia Hu