

Poor (Wo)man's Bootstrap*

Bo E. Honoré[†]

Luojia Hu[‡]

January 13, 2015.

Abstract

The bootstrap is a convenient tool for estimating 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 which is only based on estimation of one-dimensional parameters.

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

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 neighbours when the estimator is based on a non-smooth objective function or discontinuous moment conditions.

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

*This research was supported by the National Science Foundation and 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.

[†]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.

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 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 parameter of interest. We then demonstrate that this insight carries over to GMM estimators. We also point out 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 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 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 which 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 $\{z_i\}$,

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

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

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

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 expected value of the second derivative of q , but the setup also applies to many non-smooth objective functions such as Powell (1984).

While it is in principle possible to estimate V and H directly, many empirical researchers estimate $avar(\hat{\beta})$ 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 $avar(\hat{\beta})$.

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

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

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

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

where $\{z_i^b\}$ is the bootstrap sample. This is a one-dimensional minimization problem, so for complicated objective functions, it will be much easier to solve than the minimization problem that defines $\widehat{\beta}$ and its bootstrap equivalent.

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

$$\begin{aligned}\Omega_{\delta_1, \delta_2} &= \text{avar} \left(\begin{pmatrix} \widehat{a}(\delta_1) \\ \widehat{a}(\delta_2) \end{pmatrix} \right) \\ &= \begin{pmatrix} (\delta_1' H \delta_1)^{-1} \delta_1' V \delta_1 (\delta_1' H \delta_1)^{-1} & (\delta_1' H \delta_1)^{-1} \delta_1' V \delta_2 (\delta_2' H \delta_2)^{-1} \\ (\delta_1' H \delta_1)^{-1} \delta_1' V \delta_2 (\delta_2' H \delta_2)^{-1} & (\delta_2' H \delta_2)^{-1} \delta_2' V \delta_2 (\delta_2' H \delta_2)^{-1} \end{pmatrix}.\end{aligned}\tag{2}$$

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

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

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

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

The upper right hand corner of this is

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

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

$$k_2 h_{11} (h_{11} + 2h_{12} + h_{22}) = (1 + \rho k_1 h_{22}) \quad (3)$$

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

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

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

$$k_3 h_{11} (h_{11} - 2h_{12} + h_{22}) = (1 - \rho k_1 h_{22}) \quad (4)$$

Rewriting (3) and (4) in matrix form, we get

$$\begin{pmatrix} 2k_2 h_{11} & k_2 h_{11} - \rho k_1 \\ -2k_3 h_{11} & k_3 h_{11} + \rho k_1 \end{pmatrix} \begin{pmatrix} h_{12} \\ h_{22} \end{pmatrix} = \begin{pmatrix} 1 - k_2 h_{11}^2 \\ 1 - k_3 h_{11}^2 \end{pmatrix} \quad (5)$$

Appendix 1 shows that the determinant of the matrix on the left is positive definite. As a result, the two equations, (3) and (4), 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 $(\hat{a}(\delta_1), \dots, \hat{a}(\delta_m))$ as a solution to a set of nonlinear equations.

Specifically, define

$$D = \begin{pmatrix} \delta_1 & \delta_2 & \cdots & \delta_m \end{pmatrix} \quad \text{and} \quad C = \begin{pmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_m \end{pmatrix}. \quad (6)$$

The covariance matrix for the m estimators is then

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

which implies that

$$(C'(I \otimes H)C) \Omega (C'(I \otimes H)C) = (D'VD) \quad (7)$$

These need to be solved for the symmetric and positive definite matrices V and H . The calculation above shows that this has a unique solution¹ as long as D contains all vector of the form 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

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

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

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

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 has the form

$$\Sigma = (\Gamma'W_0\Gamma)^{-1} \Gamma'W_0SW_0\Gamma (\Gamma'W_0\Gamma)^{-1}$$

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

¹Except for scale.

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

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

by

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

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

$$\begin{aligned} \Omega_{\delta_1, \delta_2} &= \text{avar} \left(\begin{pmatrix} \hat{a}(\delta_1) \\ \hat{a}(\delta_2) \end{pmatrix} \right) \\ &= \begin{pmatrix} (\delta_1' \Gamma' W_0 \Gamma \delta_1)^{-1} \delta_1' \Gamma' W_0 S W_0 \Gamma \delta_1 (\delta_1' \Gamma' W_0 \Gamma \delta_1)^{-1} & (\delta_1' \Gamma' W_0 \Gamma \delta_1)^{-1} \delta_1' \Gamma' W_0 S W_0 \Gamma \delta_2 (\delta_2' \Gamma' W_0 \Gamma \delta_2)^{-1} \\ (\delta_2' \Gamma' W_0 \Gamma \delta_2)^{-1} \delta_2' \Gamma' W_0 S W_0 \Gamma \delta_1 (\delta_1' \Gamma' W_0 \Gamma \delta_1)^{-1} & (\delta_2' \Gamma' W_0 \Gamma \delta_2)^{-1} \delta_2' \Gamma' W_0 S W_0 \Gamma \delta_2 (\delta_2' \Gamma' W_0 \Gamma \delta_2)^{-1} \end{pmatrix} \end{aligned} \quad (8)$$

Of course (8) has exactly the same structure as (2) and we can therefore back out the matrices $\Gamma' W_0 \Gamma$ and $\Gamma' W_0 S W_0 \Gamma$ (up to scale) exactly the same way we backed out H and V above.

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

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

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 $\hat{\alpha}_{k\ell}$ that solves the k 'th moment with respect to the ℓ 'th element of the parameter,

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

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

$$\text{Acov}(\hat{\alpha}_{k\ell}, \hat{\alpha}_{jm}) = \frac{S_{kj}}{\gamma_{k\ell} \gamma_{jm}}$$

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

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

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

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

S_{kj} is identified as well.

Finally

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

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 GMM or method of moments estimators. As such their asymptotic variances have 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 which 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

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

with moment conditions (or limiting first order conditions),

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

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

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

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 $\theta = (\theta'_1, \theta'_2)'$ has been parameterized as $\hat{\theta} + a\delta$. 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 the sample selection model

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

where (ν_i, ε_i) has a bivariate normal distribution. α can be estimated by the probit maximum likelihood estimator, $\hat{\alpha}_{MLE}$, in a model with d_i as the outcome and z_i as the explanatory variables.

In a second step β is then estimated by the coefficients on x_i in the regression of y_i on x_i and $\lambda_i = \frac{\phi(z'_i \hat{\alpha}_{MLE})}{1 - \Phi(z'_i \hat{\alpha}_{MLE})}$ using only the sample for which $d_i = 1$. See Heckman (1979).

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 $E[q_1(z_i, \theta_1)]$ and $V[q(z_i, \theta_1)]$ can be estimated by applying the poor (wo)man's bootstrap to the first step in the estimation procedure alone. $E[r_2(z_i, \theta_1, \theta_2)]$ and $V[r(z_i, \theta_1, \theta_2)]$ 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 $E[r_1(z_i, \theta_1, \theta_2)]$ and $cov[q(z_i, \theta_1), r(z_i, \theta_1, \theta_2)]$, consider the three infeasible scalar estimators

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

for fixed δ_1 , δ_2 and δ_3 .

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

$$\begin{pmatrix} \delta'_1 E[q_1(z_i, \theta_1)] \delta_1 & 0 & 0 \\ \delta'_1 E[r_1(z_i, \theta_1, \theta_2)] \delta_2 & \delta'_2 E[r_2(z_i, \theta_1, \theta_2)] \delta_2 & 0 \\ 0 & 0 & \delta'_3 E[r_2(z_i, \theta_1, \theta_2)] \delta_3 \end{pmatrix}^{-1} \\ \begin{pmatrix} \delta'_1 V[q(z_i, \theta_1)] \delta_1 & \delta'_1 cov[q(z_i, \theta_1), r(z_i, \theta_1, \theta_2)] \delta_2 & \delta'_1 cov[q(z_i, \theta_1), r(z_i, \theta_1, \theta_2)] \delta_3 \\ \delta'_1 cov[q(z_i, \theta_1), r(z_i, \theta_1, \theta_2)] \delta_2 & \delta'_2 V[r(z_i, \theta_1, \theta_2)] \delta_2 & \delta'_2 V[r(z_i, \theta_1, \theta_2)] \delta_3 \\ \delta'_1 cov[q(z_i, \theta_1), r(z_i, \theta_1, \theta_2)] \delta_3 & \delta'_2 V[r(z_i, \theta_1, \theta_2)] \delta_3 & \delta'_3 V[r(z_i, \theta_1, \theta_2)] \delta_3 \end{pmatrix} \\ \begin{pmatrix} \delta'_1 E[q_1(z_i, \theta_1)] \delta_1 & 0 & 0 \\ \delta'_1 E[r_1(z_i, \theta_1, \theta_2)] \delta_2 & \delta'_2 E[r_2(z_i, \theta_1, \theta_2)] \delta_2 & 0 \\ 0 & 0 & \delta'_3 E[r_2(z_i, \theta_1, \theta_2)] \delta_3 \end{pmatrix}^{-1}.$$

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

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

which can be written as

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

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

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

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 $(\hat{a}_1, \hat{a}_2, \hat{a}_3)$ identifies $\delta_1' V [q(z_i, \theta_1), r(z_i, \theta_1, \theta_2)] \delta_2$ and $\delta_1' E [r_1(z_i, \theta_1, \theta_2)] \delta_2$. By choosing $\delta_1 = e_\ell$ and $\delta_2 = e_k$ this recovers all the elements of $\text{cov} [q(z_i, \theta_1), r(z_i, \theta_1, \theta_2)]$ and $E [r_1(z_i, \theta_1, \theta_2)]$.

3 Implementation

There are many ways to turn the identification strategy above into estimation of² H and V . One is to pick a set of δ -vectors and estimate the covariance matrix of the associated estimators. Denote

²Here we use the notation for extremum estimators. The same discussion applies to GMM estimators.

this estimator by $\widehat{\Omega}$. The matrices V and H can then be estimated by solving the nonlinear least squares problem

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

where D and C are defined in (6), $V_{11} = 1$, 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

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

where $\alpha_1 = 0$ (because $V_{11} = 1$) and $\alpha_k = \log(V_{kk})$. We can estimate the vector of α 's by regression $\log(y_{ij})$ 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 $(\widehat{a}(e_1), \dots, \widehat{a}(e_k))$.

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

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

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 (2)). We write this as

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

or

$$v_{ii} + v_{ij} - k_2 h_{ii}^2 - k_2 h_{ii} h_{jj} = 2k_2 h_{ii} h_{ij} \quad (10)$$

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

$$v_{ii} - v_{ij} - k_3 h_{ii}^2 - k_3 h_{ii} h_{jj} = -2k_3 h_{ii} h_{ij} \quad (11)$$

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

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

or

$$v_{jj} + v_{ij} - k_4 h_{jj}^2 - k_4 h_{ii} h_{jj} = 2k_4 h_{jj} h_{ij} \quad (12)$$

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

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

or

$$-v_{jj} + v_{ij} - k_5 h_{jj}^2 - k_5 h_{ii} h_{jj} = -2k_5 h_{jj} h_{ij} \quad (13)$$

Writing (10)–(13) in vector notation

$$\begin{pmatrix} v_{ii} + v_{ij} - k_2 h_{ii}^2 - k_2 h_{ii} h_{jj} \\ v_{ii} - v_{ij} - k_3 h_{ii}^2 - k_3 h_{ii} h_{jj} \\ v_{jj} + v_{ij} - k_4 h_{jj}^2 - k_4 h_{ii} h_{jj} \\ -v_{jj} + v_{ij} - k_5 h_{jj}^2 - k_5 h_{ii} h_{jj} \end{pmatrix} = \begin{pmatrix} 2k_2 h_{ii} \\ -2k_3 h_{ii} \\ 2k_4 h_{jj} \\ -2k_5 h_{jj} \end{pmatrix} h_{ij} \quad (14)$$

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 lower 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 (14) 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 $\hat{a}(\delta_p)$ and $\hat{a}(\delta_q)$, ω_{pq} :

$$\omega_{pq} = \text{cov}(\hat{a}(\delta_p), \hat{a}(\delta_q)) = (\delta_p' H \delta_p)^{-1} \delta_p' V \delta_q (\delta_q' H \delta_q)^{-1}$$

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 $\delta_q = e_k$, 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 estimation of standard error in a linear regression model. However, its familiarity makes it natural to use this model to illustrate the numerical properties of the approach.

We consider a linear regression model,

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

with 10 explanatory variables generated as follows. For each observation, we first generate a 9-dimensional normal, \tilde{x}_i with means equal to 0, variances equal to 1 and all covariances equal to $\frac{1}{2}$. x_{i1} to x_{i9} are then $x_{ij} = 1 \{\tilde{x}_{ij} \geq 0\}$ for $j = 1 \dots 3$, $x_{ij} = \tilde{x}_{ij} + 1$ for $j = 4$ to 6, $x_{i7} = \tilde{x}_{i7}$, $x_{i8} = \tilde{x}_{i8}/2$ and $x_{i9} = 10\tilde{x}_{i9}$. Finally $x_{i10} = 1$. ε_i is normally distributed conditional on x_i and with variance $(1 + x_{i1})^2$. We pick $\beta = (\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1, 0, 0, 0, 0, 0)$. This yields an R^2 of approximately? The scaling of x_{i8} and x_{i9} are 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 variance estimator based on estimating V and H from (9) by nonlinear least squares (N), and the variance estimator based on estimating V and H from (14) 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 \tilde{x}_i and with true coefficients equal to 0. For this design, we do not yet calculate the variance estimators based on (9) 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

$$y_i = g(f(x'_i\beta, \varepsilon_i))$$

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 the Kendall's rank correlation between y_i and x'_ib :

$$\hat{\beta} = \arg \max_b \sum_{i < j} (1 \{y_i > y_j\} - 1 \{y_i < y_j\}) (1 \{x'_ib > x'_jb\} - 1 \{x'_ib < x'_jb\})$$

The asymptotic distribution of this estimator was derived in Sherman (1993). Specifically, he showed that with³ $\beta' = (\theta', 1)'$, $\hat{\theta}$ will have a limiting normal distribution of the form considered in Section 2.1:

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

where

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

$$V = E \left[\tilde{S}(y, x'\beta)^2 g_0(x'\beta)^2 (x_0 - \bar{x}_0)(x_0 - \bar{x}_0)' \right]$$

with⁴ $\tilde{S}(y_0, t) = E[1 \{y_0 > y\} - 1 \{y_0 < y\} | x'\beta = t]$, $\tilde{S}_2(y_0, t) = \frac{\partial \tilde{S}(y_0, t)}{\partial t}$, and $\bar{x}_0 = E[x_0 | x'\beta]$, 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 (1987)'s estimator maximizes the Kendall's rank correlation between y_i and x'_ib . Cavanagh and Sherman (1998) proposed an alternative estimator of β based on by maximizing

$$\sum_{i=1}^n M(y_i) R_n(x'_ib)$$

³Since f is unspecified, it is clear that some kind of scale normalization is necessary.

⁴With the exception of V and H the notation here is chosen to make is as close as possible to that in Sherman (1993).

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

$$\sum_{i=1}^n \left(\sum_{k=1}^n 1\{y_i > y_k\} \right) \left(\sum_{j=1}^n 1\{x'_i b > x'_j b\} \right) = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n 1\{y_i > y_k\} 1\{x'_i b > x'_j b\} \quad (15)$$

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

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

where $\beta' = (\theta', 1)'$ and H_1 and V_1 having 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 $Avar(\hat{\theta})$ (or $Avar(\tilde{\theta})$) 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 \tilde{x}_{ij} with means equal to 0, variances equal to 1 and all covariances equal to $\frac{1}{2}$. We then define $x_{ij} = \tilde{x}_{ij}$ for $j = 1 \dots 2$, $x_{i3} = 1\{\tilde{x}_{i3} \geq 0\}$, and $x_{i4} = \tilde{x}_{i4} + 1$. 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 $\hat{\beta}$ 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 (15).

Table 4 compares the t-statistics based on the bootstrap estimator of the variance of $\hat{\theta}$, the variance estimator based on estimating V_1 and H_1 from (9) by nonlinear least squares (N), and the variance estimator based on estimating V_1 and H_1 from (14) 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 bigger difference between the t-statistics based on our approach and that 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 answer 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 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 approximating the asymptotic variance than does the usual bootstrap. We suspect that the reason is that our approach implicitly assumes that the asymptotics provides 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

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

and sector 2 otherwise.

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

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

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

We will now imagine a setting in which the econometrician has a data set with n observations from this model. x_{is} is composed of a constant and a normally distributed component which is independent across sectors and across individuals. In the data generating process these are $\beta_1 = (1, 1)'$, $\beta_2 = (\frac{1}{2}, 1)'$, $\gamma_1 = 0$ and $\gamma_2 = 1$. Finally, $\sigma_1^2 = 2$, $\sigma_2^2 = 3$, $\tau = 0$ and $\rho = 0.95$. In the estimation, we treat 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 woman's bootstrap standard errors.

The model is estimated by indirect inference matching the following parameters in the following regressions (all estimated by OLS; with the additional notation that $d_{i0} = 0$)

- The regression coefficients and residual variance in a regression of w_{it} on x_{i1} , x_{i2} , and $1\{d_{it-1} = 1\}$ using the subsample of observations in sector 1.
- The regression coefficients and residual variance in a regression of w_{it} on x_{i1} , x_{i2} , and $1\{d_{it-1} = 1\}$ using the subsample of observations in sector 2.
- The regression coefficients in a regression $1\{d_{it} = 1\}$ on x_{i1} and x_{i2} and $1\{d_{it-1} = 1\}$.

Let \hat{a} be the vector of those parameters based on the data and let $\widehat{V}[\hat{a}]$ 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 $\tilde{\alpha}(\theta)$ and estimates the model parameters by minimizing

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

over 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.

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 estimation of *one-dimensional* parameters. We believe that this method can be useful for applied researchers 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 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.
- NEWHEY, 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.
- 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 Matrix in Equation (5)

The determinant of the matrix on the left of (5) is

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

since H is positive definite.

Appendix 2: Calculating the Asymptotic Variance of the MRC Estimator

Following Cavanagh and Sherman (1998) let

$$f((y_1, x_1), (y_2, x_2), (y_3, x_3), b) = 1\{y_1 > y_3\} 1\{x'_1 b > x'_2 b\}$$

and let

$$\tau((y, x), b) = E[1\{y > Y_3\} 1\{x'b > X'_2 b\}] + E[1\{Y_1 > Y_3\} 1\{X'_1 b > x'b\}] + E[1\{Y_1 > y\} 1\{X'_1 b > X'_2 b\}]$$

The asymptotic variance of the estimator is

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

where

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

and

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

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

$$\begin{aligned} \hat{\tau}((y, x), b) &= \frac{1}{n^2} \sum_{j,k} 1\{y > y_j\} 1\{x'b > x'_k b\} + \frac{1}{n^2} \sum_{j,k} 1\{y_k > y\} 1\{x'_k b > x'b\} \\ &\quad + \frac{1}{n^2} \sum_{j,k} 1\{y_k > y\} 1\{x'_k b > x'_j b\} \\ &= \frac{1}{n^2} R_n(y) R_n(x'b) + \frac{1}{n^2} \sum_k R_n(y_k) 1\{x'_k b > x'b\} \\ &\quad + \frac{1}{n^2} \sum_k 1\{y_k > y\} R_n(x'_k b). \end{aligned}$$

We then numerically differentiate $\hat{\tau}((y, x), b)$ twice (using a step-size⁵ 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 $(X_1, X_2)'$ have a bivariate normal distribution with mean $(\mu_1, \mu_2)'$ and variance $\begin{pmatrix} \sigma_1^2 & \tau\sigma_1\sigma_2 \\ \tau\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$ and let $(Y_1, Y_2)' = (\exp(X_1), \exp(X_2))'$. We are interested in $E[\max\{Y_1, Y_2\}]$.

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

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

⁵Changing the step-size to 0.05 changes the asymptotic standard errors by approximately 2%.

Therefore

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

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

	$ T_E - T_B $	$ T_E - T_N $	$ T_E - T_L $	$ T_B - T_N $	$ T_B - T_L $	$ T_N - T_L $
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, $n = 2000$.

	$ T_E - T_B $	$ T_E - T_N $	$ T_E - T_L $	$ T_B - T_N $	$ T_B - T_L $	$ T_N - T_L $
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, $n = 2000$.

	$ T_E - T_B $	$ T_E - T_L $	$ T_B - T_L $
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

	$n = 200$			$n = 500$		
	$ T_B - T_N $	$ T_B - T_L $	$ T_N - T_L $	$ T_B - T_N $	$ T_B - T_L $	$ T_N - T_L $
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	0.023	0.018
Asymptotic	0.049	0.041	0.051	0.040	0.028	0.068	0.026	0.019
BS_N	0.054	0.042	0.053	0.041	0.031	0.070	0.027	0.018