Easy Bootstrap-Like Estimation of Asymptotic Variances[*]

Bo E. Honoré[†]   Luojia Hu[‡]

June 29, 2018

Abstract

The bootstrap is a convenient tool for calculating standard errors of the parameter estimates of complicated econometric models. Unfortunately, the bootstrap can be very time-consuming. In a recent paper, Honoré and Hu (2017), we propose a "Poor (Wo)man's Bootstrap" based on one-dimensional estimators. In this paper, we propose a modified, simpler method and illustrate its potential for estimating asymptotic variances.

KEYWORDS: standard error; bootstrap; inference; censored regression; two-step estimation.

JEL CODE: C10, C18, C15.

[†]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.

1 Introduction

Most standard estimators for cross-sectional econometric models have asymptotic distribution of the form

$$\sqrt{n}\left(\widehat{\theta} - \theta_0\right) \overset{d}{\longrightarrow} N\left(0, H^{-1}VH^{-1}\right) \tag{1}$$

where θ_0 is the k-dimensional parameter of interest, H and V are symmetric, positive definite matrices to be estimated. It is usually possible to get explicit expressions for H and V, but estimating them can be computationally difficult in complicated models. The bootstrap[1] provides a simple method for estimating $H^{-1}VH^{-1}$ directly.

One practical problem with the bootstrap is that it requires re-estimating the model a large number of times. This can be a limitation for complicated models where it is time-consuming to calculate the objective function that defines the estimator, or for estimators that are based on sample moments that are discontinuous in the parameter.

In Honoré and Hu (2017), we introduced a version of the bootstrap which is based on calculating one-dimensional estimators using a fixed set of directions in $\mathbb{R}^k$ for each bootstrap replication. The covariance of these one-dimensional estimators is then used to back out estimators of H and V via *nonlinear* least squares. The benefit of this approach is that it is often much easier to calculate one-dimensional than k-dimensional estimators.

In this note, we introduce a modified approach which permits using one-dimensional estimators in different directions in each bootstrap replication, and which makes it possible to back out estimators to H and V via *linear* regression. In order to highlight the idea behind the approach, we will be deliberately vague about the underlying regularity conditions.

Section 2 describes our basic idea in the context of an extremum estimator, but as mentioned, the approach applies equally well to GMM estimators. In Section 3, we illustrate the potential usefulness of the approach by considering Powell's (1984) the Censored Least Absolute Deviations. We choose this example because quantile regression estimators provide a classical example where the matrix H in (1) cannot be estimated by a simple sample analog. Section 4 demonstrates how the proposed approach can be used to estimate the

---

[1]The bootstrap can also be used to provide asymptotic refinements that can lead to more reliable inference in finite samples. That is not the topic of this note.

variance of two step estimators. Two step estimators also provide a classical example where it is cumbersome to estimate the variance of an estimator. Section 5 concludes.

2   Our Modified Approach

To fix ideas, consider an extremum estimator of the form

$$\widehat{\theta} = \arg\min_{t} \frac{1}{n} \sum_{i=1}^{n} q\left(z_i; t\right) \tag{2}$$

where z_i is the data for observation number i, n is the sample size, and $\theta_0 = \arg\min_t E\left[q\left(z_i; t\right)\right]$ is the true parameter value. Under random sampling and weak technical assumptions, (1) holds with $V = V\left[q'\left(z_i; \theta_0\right)\right]$ and $H = E\left[q''\left(z_i; \theta_0\right)\right]$, where the differentiation is with respect to the parameter. See for example Amemiya (1985). The insight in Honoré and Hu (2017) is to consider (infeasible) one-dimensional estimators of the form

$$\widehat{a}\left(\delta\right) = \arg\min_{a} \frac{1}{n} \sum_{i=1}^{n} q\left(z_i; \theta_0 + a\delta\right),$$

where δ is a fixed k-dimensional vector and a is a scalar. The joint asymptotic distribution of m such estimators, $\widehat{a}\left(\delta_1\right)$, ..., $\widehat{a}\left(\delta_m\right)$, is asymptotically normal with asymptotic variance

$$\Omega = \left(C'\left(I \otimes H\right)C\right)^{-1}\left(D'VD\right)\left(C'\left(I \otimes H\right)C\right)^{-1}, \tag{3}$$

where I is an $m \times m$ identity matrix,

$$\underset{(k\times m)}{D} = \begin{pmatrix} \delta_1 & \delta_2 & \cdots & \delta_m \end{pmatrix} \qquad \text{and} \qquad \underset{(km\times m)}{C} = \begin{pmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \delta_m \end{pmatrix}.$$

Equation (3) implies the relationship,

$$\left(C'\left(I \otimes H\right)C\right)\Omega\left(C'\left(I \otimes H\right)C\right) = \left(D'VD\right). \tag{4}$$

Honoré and Hu (2017) proved that for suitably chosen directions, δ_1, .., δ_m, equation (4) identifies[2] V and H from Ω, and proposed estimating V and H by nonlinear least squares

---

[2]Except for an innocuous scale normalization.

after estimating Ω with the bootstrap. Honoré and Hu (2017) also demonstrate that the same approach can be used for GMM estimators.

The argument leading to (1) is almost always based on the representation

$$\widehat{\theta} - \theta_0 \approx H^{-1} \frac{1}{n} \sum_{i=1}^{n} s_i \tag{5}$$

where $\approx$ means that the two sides differ by a magnitude which is asymptotically negligible relative to the right hand side, and s_i is a function of the data for individual i. For example, for the extremum estimator in (2), $s_i = q'(z_i; \theta_0)$ when q is smooth in the parameter. The same basic argument applies to the bootstrap (see Hahn (1996)). Specifically, consider a bootstrap sample $\{z_i^b\}$ of size[3] n, where the z_i^b's are drawn with replacement from the empirical distribution of $\{z_i\}$. Standard asymptotic theory implies that in each bootstrap replication, b, the estimator, $\widehat{\theta}_b = \arg\min_t \frac{1}{n} \sum_{i=1}^{n} q(z_i^b; t)$ has the linear representation

$$\widehat{\theta}_b - \widehat{\theta} \approx H^{-1} \frac{1}{n} \sum_{i=1}^{n} s_i^b \tag{6}$$

for the same H and in (5).

As in Honoré and Hu (2017), this paper considers (infeasible) estimators of the form

$$\widehat{a}(\delta) = \arg\min_a \frac{1}{n} \sum_{i=1}^{n} q(z_i; \theta_0 + a\delta),$$

where δ is a fixed k-dimensional vector. These estimator have the representation

$$\widehat{a}(\delta) \approx (\delta' H \delta)^{-1} \delta' \frac{1}{n} \sum_{i=1}^{n} s_i$$

and the corresponding (feasible) estimators in a bootstrap sample,

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

have the representation

$$\widehat{a}_b(\delta) \approx (\delta' H \delta)^{-1} \delta' \frac{1}{n} \sum_{i=1}^{n} s_i^b. \tag{7}$$

---

[3]In principle, the bootstrap sample size can differ from the actual sample size. We ignore this in order to keep the notation simpler.

Note that we can write (7) as

$$(\delta' H \delta) \, \widehat{a}_b (\delta) \approx \delta' s^b$$

where $s^b = \frac{1}{n} \sum\limits_{i=1}^{n} s_i^b$. Equivalently

$$\widehat{a}_b (\delta) \left(\delta' H \delta \right) - \delta' s^b \approx 0 \tag{8}$$

or

$$\sum_{j,\ell} \left(\widehat{a}_b (\delta) \, \delta_j \delta_\ell \right) h_{j\ell} - \sum_j \delta_j s_j^b \approx 0, \tag{9}$$

where s_j^b is the j'th element of s^b, δ_j is the j'th element of δ. Since $h_{j\ell} = h_{\ell j}$, equation (9) can be written as

$$\sum_{j} \left(\widehat{a}_b (\delta) \, \delta_j \delta_j \right) h_{jj} + \sum_{\ell < j} \left(2 \widehat{a}_b (\delta) \, \delta_j \delta_\ell \right) h_{j\ell} - \sum_j \delta_j s_j^b \approx 0. \tag{10}$$

As in Honoré and Hu (2017), the same idea applies to GMM estimators.

It is useful to think of (10) as a linear regression model where the parameters are the $h_{j\ell}$'s and the s_j^b's , the dependent variable is always 0 and (asymptotically) there is no error. Of course, for this to be useful, one needs to impose a scale normalization such as $h_{11} = 1$ or $\sum\limits_{j=1}^{k} h_{jj}^2 = 1$. See Appendix 1 for how to impose the restriction in practice.

In each bootstrap replication, each δ-vector gives an observation from (10). The s^b-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 s^b as a bootstrap-specific fixed effect that can be eliminated by a transformation similar to the "textbook" panel data deviations-from-means transformation. For details[4], see Appendix 2, where α_i plays the role of s^b. This provides an easy way to estimate the elements of H (up to scale).

Once H has been estimated, one could back out the s^b-vector for each bootstrap replication and use the sample variance of s^b to estimate V. Specifically, the s^b-vector for a bootstrap replication can be estimated by stacking the terms $\sum_j \left(\widehat{a}_b (\delta) \, \delta_j \delta_j \right) \widehat{h}_{jj} + \sum_{\ell < j} \left(2 \widehat{a}_b (\delta) \, \delta_j \delta_\ell \right) \widehat{h}_{j\ell}$

---

[4]If one wants to impose the normalization $\sum\limits_{j=1}^{k} h_{jj}^2 = 1$, then the method described in Appendix 1 can be applied to the regression (14) in Appendix 2.

for a given bootstrap replication and regressing them on the stacked δ''s (this is the D' from above).

One potential advantage of exploiting (10) to recover H and s^b is that it is straightforward to allow the directions δ to differ across replications. This is useful because it seems intuitive that in a given application, some choices of δ will be less informative for recovering H and s^b than others. For example, Honoré and Hu (2017) use all vectors of the from e_j, $e_j + e_\ell$ and $e_j - e_\ell$, where e_j denotes a vector that has 1 in its j'th element and zeros elsewhere. This treats all the elements of $\widehat{\theta}$ symmetrically. It would be more natural to treat all elements of $Avar\left(\widehat{\theta}\right)^{-1/2}\widehat{\theta}$ symmetrically. This would be scale and rotation invariant, and it amounts to taking the directions in Honoré and Hu (2017) (or any other set of symmetric directions) and pre-multiplying them by $Avar\left(\widehat{\theta}\right)^{1/2}$. Since $Avar\left(\widehat{\theta}\right)$ is not known, this is not feasible, but one could adjust the directions for a given bootstrap replication using preliminary estimates of H and V (and hence $Avar\left(\widehat{\theta}\right)$) based on the bootstrap replications so far.

3   Illustration: Censored Least Absolute Deviations

In this section we use the Censored Least Absolute Deviations (CLAD) estimator to illustrate our approach. Powell (1984) considered the model

$$y_i = \max\left\{0, x_i'\beta + \varepsilon_i\right\}$$

with $median\left(\varepsilon_i|\,x_i\right) = 0$ and proposed the Censored Least Absolute Deviations estimator,

$$\widehat{\beta} = \arg\min_b \sum |y_i - \max\left(0, x_i'b\right)|\,.$$

Under random sampling and weak regularity conditions,

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

with $V = E\left[1\left\{x_i'\beta > 0\right\}x_i x_i'\right]$ and $H = 2E\left[f_{\varepsilon_i|x_i}\left(0|\,x_i\right)1\left\{x_i'\beta > 0\right\}x_i x_i'\right]$. Note that in this case, the "Hessian" in asymptotic variance involves the conditional density of ε_i given x_i. The CLAD estimator, along with its uncensored predecessor, is one of the simplest and earliest asymptotically normal econometric estimators for which the asymptotic variance cannot be

estimated by a simple sample analog. See, for example, Buchinsky (1998) for a discussion of various approaches.

We consider 1,000 Monte Carlo replications from a random sample of size n from

$$y_i = \max\{0, x_i'\beta + \varepsilon_i\},$$

where we first generate $(\widetilde{x}_{i1}, \widetilde{x}_{i2}, \widetilde{x}_{i3}, \widetilde{x}_{i4})$ from a normal distribution with means 0, variances 1, and all covariances $\frac{1}{2}$. The explanatory variables are then $x_{ij} = 1\{\widetilde{x}_{ij} \geq 0\}$ for $j = 1 \cdots 3$, $x_{i4} = \widetilde{x}_{i4}$ and $x_{i5} = 1$. The error, ε_i, is $N\left(0, (1 + x_{i1})^2\right)$ and $\beta = \left(\frac{1}{5}, \frac{2}{5}, \frac{3}{5}, \frac{4}{5}, 1\right)$. This results in approximately 20% censoring. We choose n to be 10,000. This is unrealistically large given the number of explanatory variables. We choose a very large sample size because it allows us to focus on the marginal contribution of this paper to the estimation of asymptotic variances, without worrying about whether the asymptotic distribution is a good approximation to begin with, or whether the discrete nature of the empirical distribution of the data causes small sample issues for the bootstrap.

Since we know the data generating process, we can calculate the variance of the estimator implied by the asymptotic distribution. The corresponding standard errors are given in the first row of Table 1. The second row reports the standard deviation of the estimator across the Monte Carlo samples.

The next three rows of Table 1 report estimated standard errors based on the following bootstrap procedures with 1,000 bootstrap replications: (i) the regular multinomial bootstrap, (ii) the poor (wo)man's bootstrap from Honoré and Hu (2017), and (iii) the computationally easy bootstrap from Section 2. To simplify the comparison of (ii) and (iii), we use the directions, δ, proposed in Honoré and Hu (2017) for both.

The final row of Table 1 illustrates how our proposed procedure can sometimes be simplified by using the structure of the problem. For the CLAD estimator, the V-matrix is easy to estimate by a sample analog, but the H-matrix is more troublesome because it contains the conditional density of the errors. Also, in this case $s_i = 1\{x_i'\beta > 0\} sign(y_i - x_i'\beta) x_i$. In each bootstrap replication, we therefore use $\frac{1}{n}\sum_{i=1}^{n} 1\left\{x_i^{b\prime}\widehat{\beta} > 0\right\} sign\left(y_i^b - x_i^{b\prime}\widehat{\beta}\right) x_i^b$ to estimate s^b, and then we use (9) estimate the elements in H (by least absolute deviations). V is estimated by $\frac{1}{n}\sum_{i=1}^{n} 1\left\{x_i'\widehat{\beta} > 0\right\} x_i x_i'$.

7

Table 1: Average Estimated Standard Errors

	β_1	β_2	β_3	β_4	β_5
Square Root of Asymptotic Variance	0.044	0.038	0.038	0.024	0.028
Standard Deviation Across Replications	0.044	0.037	0.038	0.024	0.028
Average of Bootstrap Standard Errors	0.044	0.039	0.039	0.024	0.028
Standard Errors Based on Honoré and Hu (2017)	0.044	0.038	0.038	0.024	0.028
Standard Errors Based on (9) Section 2	0.045	0.039	0.039	0.024	0.028
S.E. Based on (9) using the Structure of the CLAD	0.046	0.040	0.040	0.025	0.030

The results presented in Table 1 suggest that the approach proposed here can be useful for estimating asymptotic variances. Somewhat surprisingly, the approach that used the structure of the asymptotic variance (in row six) performs slightly worse than the one based on (9) Section 2. On the other hand, the former is computationally simpler.

4 Our Approach with Two Step Estimators

The asymptotic variance of two step estimators does not have the representation in (9). However, they are still asymptotically linear (Newey (1984)), so the same basic idea applies. Specifically, suppose that we have a two step estimation problem

$$\widehat{\theta}_1 = \arg\min_t \frac{1}{n} \sum_{i=1}^{n} q(z_i; t) \qquad \text{and} \qquad \widehat{\theta}_2 = \arg\min_t \frac{1}{n} \sum r\left(z_i; \widehat{\theta}_1, t\right),$$

with first order conditions[5]

$$0 = \frac{1}{n} \sum_{i=1}^{n} q_1(z_i; \widehat{\theta}_1) \qquad \text{and} \qquad 0 = \frac{1}{n} \sum_{i=1}^{n} r_2(z_i; \widehat{\theta}_1, \widehat{\theta}_2). \tag{11}$$

In that case general GMM theory applies, and it follows that

$$\sqrt{n}\left(\begin{pmatrix} \widehat{\theta}_1 \\ \widehat{\theta}_2 \end{pmatrix} - \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}\right) \overset{d}{\longrightarrow}$$

$$N\left(0, \begin{pmatrix} Q_{11} & 0 \\ R_{21} & R_{22} \end{pmatrix}^{-1} \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix} \left(\begin{pmatrix} Q_{11} & 0 \\ R_{21} & R_{22} \end{pmatrix}^{-1}\right)'\right),$$

---

[5]Here, we implicitly assume that the objective functions are smooth, but similar expressions can often be obtained when they are not. See for example Huber (1967).

where $Q_{11} = E\left[\frac{\partial q_1(z_i;\theta_1)}{\partial \theta_1}\right]$, $R_{21} = E\left[\frac{\partial r_2(z_i;\theta_1,\theta_2)}{\partial \theta_1}\right]$, $R_{22} = E\left[\frac{\partial r_2(z_i;\theta_1,\theta_2)}{\partial \theta_2}\right]$, $V_{11} = V\left[q\left(z_i;\theta_1\right)\right]$, $V_{12} = \text{cov}\left[q\left(z_i;\theta_1\right), r\left(z_i;\theta_1,\theta_2\right)\right]$, and $V_{22} = V\left[r\left(z_i;\theta_1,\theta_2\right)\right]$.

Since (11) constitute a set of moment conditions and we pointed out in Section 2 that the approach discussed there applies to GMM estimators, it is tempting to conclude that two step estimators do not warrant a special treatment. However, the problem is that the one-dimensional estimation used in Section 2 will not preserve the simplicity of the two-step estimator. For example, Heckman's two-step estimator is based on two simple optimization problems (probit and OLS) which deliver two parameter vectors $\widehat{\alpha}$ and $\widehat{\beta}$ separately. In contrast, the procedure in Section 2 would lead to estimating linear combinations of the elements of α and β. See Section 5 of Honoré and Hu (2017). In this section, we therefore provide a different procedure that explicitly preserves the simplicity of the two-step estimator.

One way to see this in smooth cases is to do a Taylor series approximation to (11) around the true parameter values to get

$$
\begin{aligned}
0 &\approx \frac{1}{n}\sum_{i=1}^{n} q_1(z_i;\widehat{\theta}_1) \\
&\approx \frac{1}{n}\sum_{i=1}^{n} q_1(z_i;\theta_1) + \left(\frac{1}{n}\sum_{i=1}^{n} q_{11}(z_i;\theta_1)\right)\left(\widehat{\theta}_1 - \theta_1\right) \approx s_1 + Q_{11}\left(\widehat{\theta}_1 - \theta_1\right),
\end{aligned}
$$

where $s_1 = \frac{1}{n}\sum_{i=1}^{n} q_1(z_i;\theta_1)$, and

$$
\begin{aligned}
0 &\approx \frac{1}{n}\sum_{i=1}^{n} r_2(z_i;\widehat{\theta}_1,\widehat{\theta}_2) \\
&\approx \frac{1}{n}\sum_{i=1}^{n} r_2(z_i;\theta_1,\theta_2) + \left(\frac{1}{n}\sum_{i=1}^{n} r_{21}(z_i;\theta_1,\theta_2)\right)\left(\widehat{\theta}_1 - \theta_1\right) + \left(\frac{1}{n}\sum_{i=1}^{n} r_{22}(z_i;\theta_1,\theta_2)\right)\left(\widehat{\theta}_2 - \theta_2\right) \\
&\approx s_2 + R_{21}\left(\widehat{\theta}_1 - \theta_1\right) + R_{22}\left(\widehat{\theta}_2 - \theta_2\right),
\end{aligned}
$$

where $s_2 = \frac{1}{n}\sum_{i=1}^{n} r_2(z_i;\theta_1,\theta_2)$.

Now suppose that, as in Honoré and Hu (2017), we calculate (infeasible) directional estimators of the form

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

9

A Taylor series expansion of the first order conditions yields

$$
\begin{aligned}
0 &\approx \delta_1' Q_{11} \delta_1 \widehat{a}_1\left(\delta_1\right) + \delta_1' s_1 \\
0 &\approx \delta_2' R_{22} \delta_2 \widehat{a}_2\left(\delta_1, \delta_2\right) + \delta_2' s_2 + \delta_2' R_{21} \delta_1 \widehat{a}_1\left(\delta_1\right) \\
0 &\approx \delta_2' R_{22} \delta_2 \widehat{a}_3\left(\delta_2\right) + \delta_2' s_2.
\end{aligned}
$$

These are again linear in the elements of Q_{11}, R_{21} and R_{22} and we can use the same approach as in Section 2.

5   Conclusion

The bootstrap is a convenient tool for estimating asymptotic variances, but it can sometimes be quite time consuming. In Honoré and Hu (2017) we proposed a version of the bootstrap that is based on calculating one-dimensional estimators. This can lead to great computational gains in complicated models because search in one dimension is faster and more reliable than in higher dimensions.

This paper proposes a modification to the approach in Honoré and Hu (2017). The advantage of the approach is that while Honoré and Hu (2017) requires nonlinear least squares, the approach here can be implemented with linear regression. In also has the advantage that one can calculate one-dimensional estimators in different directions in different bootstrap replications. In Honoré and Hu (2017), the directions must be the same in each bootstrap replication.

The approach applies to extremum estimators as well as GMM estimators, including two-step estimators.

References

AMEMIYA, T. (1985): *Advanced Econometrics*. Harvard University Press.

BUCHINSKY, M. (1998): "Recent Advances in Quantile Regression Models," *Journal of Human Resources*, 33, 88–126.

HAHN, J. (1996): "A Note on Bootstrapping Generalized Method of Moments Estimators," *Econometric Theory*, 12(1), pp. 187–197.

HONORÉ, B. E., AND L. HU (2017): "Poor (Wo)man's Bootstrap," *Econometrica*, 85(4), 1277–1301.

HUBER, P. J. (1967): "The behavior of maximum likelihood estimates under nonstandard conditions," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, pp. 221–233, Berkeley, Calif. University of California Press.

NEWEY, W. K. (1984): "A method of moments interpretation of sequential estimators," *Economics Letters*, 14(23), 201 – 206.

POWELL, J. L. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–325.

Appendix 1

In order to estimate H and V from (9) we regress 0 on a set of explanatory variables subject to a scale normalization. In other words, we minimize a sum of squares of the form $\sum (x_i' b)^2$ subject to a normalization. If the normalization is that one of the elements of b is 1, then that can be done by a simple linear regression. If the normalization is of the form $\sum b_j^2 = 1$ where the sum is over some subset of the parameters (for example, the diagonal elements of H), we use the following.

Consider the minimization problem

$$\min b' \begin{pmatrix} A & B \\ B' & C \end{pmatrix} b \qquad \text{s.t.} \qquad b_1' b_1 = 1$$

where $b = (b_1', b_2')'$. The minimization problem

$$\min b_1' A b_1 + b_1' B b_2 + b_2' B' b_1 + b_2' C b_2 \qquad s.t. \qquad b_1' b_1 = 1$$

has Lagrangian

$$L = b_1' A b_1 + b_1' B b_2 + b_2' B' b_1 + b_2' C b_2 + \lambda \left(b_1' b_1 - 1 \right).$$

The first order condition with respect to b_2 is

$$2 B' b_1 + 2 C b_2 = 0$$

or

$$b_2 = -C^{-1} B' b_1 \tag{12}$$

while the first order condition with respect to b_1

$$2 A b_1 + 2 B b_2 + 2 \lambda b_1 = 0. \tag{13}$$

Substituting (12) into (13), we obtain

$$2 A b_1 - 2 B C^{-1} B' b_1 + 2 \lambda b_1 = 0$$

or

$$\left(A - B C^{-1} B' \right) b_1 + \lambda b_1 = 0$$

12

so the minimizing value of b_1 must be an eigen-vector of $(A - BC^{-1}B')$ and $-\lambda$ is the corresponding eigenvalue.

Returning to the original objective function and plugging in (12), we have

$$
\begin{aligned}
& b_1' A b_1 + b_1' B b_2 + b_2' B' b_1 + b_2' C b_2 \\
&= b_1' A b_1 - b_1' B C^{-1} B' b_1 - b_1' B C^{-1} B' b_1 + b_1' B C^{-1} C C^{-1} B' b_1 \\
&= b_1' \left(A - B C^{-1} B' \right) b_1 \\
&= b_1' \left(-\lambda b_1 \right) \\
&= -\lambda b_1' b_1 = -\lambda
\end{aligned}
$$

because b_1 is an eigenvector with eigenvalue $-\lambda$. So b_1 must be the eigenvector associated with the smallest (real) eigenvalue..

Finally, we will show that $(A - BC^{-1}B')$ is positive definite, so all its eigenvalues are real. This will establish that the minimizing value for b_1 is the eigenvector associated with the smallest eigenvalue. The solution for b_2 is then given by (12).

Note that $\begin{pmatrix} A & B \\ B' & C \end{pmatrix}$ is positive definite, hence its inverse is also positive definite. This inverse can be partitioned as

$$
\begin{pmatrix} (A - BC^{-1}B')^{-1} & ?? \\ ?? & ?? \end{pmatrix}
$$

Hence $(A - BC^{-1}B')^{-1}$ is positive definite, and then so is its inverse, $(A - BC^{-1}B')$.

Appendix 2

Consider the panel data regression model

$$
y_i = X_i \beta + Z_i \alpha_i + \varepsilon_i
$$

where y_i is $T_i \times 1$, X_1 is $T_i \times K$, Z_i is $T_i \times L$, ε_i is $T_i \times 1$ and $E\left[\varepsilon_i \mid X_i, Z_i\right] = 0$. Here, y_i, X_i and Z_i are observed data, β is the parameter of interest and α_i is a vector of individual specific "fixed" effects. Assume that $L < T_i$ and that $Z_i' Z_i$ has full rank and define $P_{Z_i} =$

$I - Z_i \left(Z_i' Z_i \right)^{-1} Z_i'$. Then

$$
\begin{aligned}
P_{Z_i} y_i &= P_{Z_i} X_i \beta + P_{Z_i} Z_i \alpha_i + P_{Z_i} \varepsilon_i \\
&= P_{Z_i} X_i \beta + P_{Z_i} \varepsilon_i,
\end{aligned}
$$

or

$$
\begin{pmatrix} P_{Z_1} y_1 \\ P_{Z_2} y_2 \\ \vdots \\ P_{Z_n} y_n \end{pmatrix} = \begin{pmatrix} P_{Z_1} X_1 \\ P_{Z_2} X_2 \\ \vdots \\ P_{Z_n} X_n \end{pmatrix} \beta + \begin{pmatrix} P_{Z_1} \varepsilon_1 \\ P_{Z_2} \varepsilon_2 \\ \vdots \\ P_{Z_n} \varepsilon_n \end{pmatrix}. \tag{14}
$$

As a result, β can be estimated by applying OLS to equation (14). When Z_i is a column vector of ones and α_i is one-dimensional, this is the usual deviations-from-means fixed effects estimator.