

**Discussion Paper**

No. 2019-39 | June 19, 2019 | <http://www.economics-ejournal.org/economics/discussionpapers/2019-39>

Please cite the corresponding Journal Article at  
<http://www.economics-ejournal.org/economics/journalarticles/2020-4>

## Bootstrap methods for inference in the Parks model

*Mantobaye Moundigbaye, Clarisse Messemer, Richard W. Parks,  
W. Robert Reed*

### Abstract

This paper shows how to bootstrap hypothesis tests in the context of the Parks (*Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated* 1967) estimator. It then demonstrates that the bootstrap outperforms Parks's top competitor. The Parks estimator has been a workhorse for the analysis of panel data and seemingly unrelated regression equation systems because it allows the incorporation of cross-sectional correlation together with heteroskedasticity and serial correlation. Unfortunately, the associated, asymptotic standard error estimates are biased downward, often severely. To address this problem, Beck and Katz (*What to do (and not to do) with time series cross-section data* 1995) developed an approach that uses the Prais-Winsten estimator together with "panel corrected standard errors" (PCSE). While PCSE produces standard error estimates that are less biased than Parks, it forces the user to sacrifice efficiency for accuracy in hypothesis testing. The PCSE approach has been, and continues to be, widely used. This paper develops an alternative: a nonparametric bootstrapping procedure to be used in conjunction with the Parks estimator. We demonstrate its effectiveness using an innovative experimental approach that creates artificial panel datasets modelled after actual panel datasets. Our approach provides a Pareto-improving option by allowing researchers to retain the efficiency of the Parks estimator while producing more accurate hypothesis test results than the PCSE.

**JEL** C13 C15 C23 C33

**Keywords** Parks model, PCSE, SUR, panel data, cross-sectional correlation, bootstrap, Monte Carlo, simulation

### Authors

Mantobaye Moundigbaye, Department of Economics and Finance, University of Canterbury, New Zealand

Clarisse Messemer, Bonneville Power Administration, Portland, Oregon, USA

Richard W. Parks, Department of Economics, University of Washington, Washington, USA

W. Robert Reed, Department of Economics and Finance, University of Canterbury, New Zealand, bob.reed@canterbury.ac.nz

**Citation** Mantobaye Moundigbaye, Clarisse Messemer, Richard W. Parks, W. Robert Reed (2019). Bootstrap methods for inference in the Parks model. *Economics Discussion Papers*, No 2019-39, Kiel Institute for the World Economy.

<http://www.economics-ejournal.org/economics/discussionpapers/2019-39>

Received May 27, 2019 Accepted as Economics Discussion Paper June 7, 2019 Published June 19, 2019

© Author(s) 2019. Licensed under the [Creative Commons License - Attribution 4.0 International \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

## I. INTRODUCTION

The Parks (1967) estimator was designed as an efficient estimator for systems of equations with both serially and contemporaneously correlated disturbances. Such models include the SUR model and associated restricted forms, such as time-series, cross-section/panel data models. Its superior efficiency is well established (Kmenta and Gilbert, 1970; Guilkey and Schmidt, 1973; Messemer, 2003; Chen et al., 2010; Moundigbaye et al., 2018). It has been widely used and is available in many econometric software packages including RATS, SHAZAM, SAS, and Stata.<sup>1</sup>

Kmenta and Gilbert (1970) were the first to note that Parks's estimated standard errors, while consistent, can be substantially biased in finite samples. More recently, Beck and Katz (1995) documented that the estimated standard errors for the Parks estimator have severe downward bias when the time dimension is small relative to the number of cross-sections. To address this deficiency, they recommend using a Prais-Winsten estimator with corresponding "panel-corrected standard errors" (PCSE).

This approach has been widely adopted, as evidenced by more than 2,400 Web of Science citations. It remains a popular estimation choice. Despite being more than twenty years old, the number of annual citations received by Beck and Katz (1995) continues to increase over time. While the PCSE approach generally involves less size distortion than Parks with asymptotic standard errors, it does not eliminate it (cf. Reed and Webb, 2010). This paper demonstrates that the combined use of the Parks estimator with bootstrapping constitutes an approach that is superior to the PCSE approach in both estimator efficiency and inference accuracy.

The remainder of the paper is organized as follows. Section 2 briefly introduces the

---

<sup>1</sup> The Parks estimator provides the framework for Stata's *xtgls* procedure.

SUR model with autoregressive errors and associated Parks estimator. Section 3 presents a nonparametric bootstrap procedure for the Parks estimator. Section 4 demonstrates the superior performance of the bootstrap procedure. It employs an innovative experimental approach where testing is performed on synthetic datasets designed to resemble actual datasets. Section 5 concludes.

## II. THE SUR MODEL AND THE PARKS ESTIMATOR

The Parks estimator was constructed as an efficient estimator for the SUR model with autocorrelated disturbances,

$$(1) \quad \mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \dots, N,$$

where  $\mathbf{y}_i$  and  $\boldsymbol{\varepsilon}_i$  are  $T \times 1$  vectors,  $\mathbf{X}_i$  is  $T \times k_i$ , and  $\boldsymbol{\beta}_i$  is  $k_i \times 1$ . The  $N$  equations can be stacked and represented in compact form as,

$$(2) \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where  $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}$ ,  $\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_N \end{bmatrix}$ ,  $\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_N \end{bmatrix}$ , and  $\boldsymbol{\varepsilon} = \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{bmatrix}$ , with  $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ , and  $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Omega}$ .

$\boldsymbol{\Omega}$ .

It is assumed that the disturbance vector,  $\boldsymbol{\varepsilon}_{(t)} = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt})'$ , is generated by a stationary, first-order autoregressive process,

$$(3) \quad \boldsymbol{\varepsilon}_{(t)} = \begin{bmatrix} \varepsilon_{1t} \\ \vdots \\ \varepsilon_{Nt} \end{bmatrix} = \begin{bmatrix} \rho_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \rho_N \end{bmatrix} \begin{bmatrix} \varepsilon_{1t-1} \\ \vdots \\ \varepsilon_{Nt-1} \end{bmatrix} + \begin{bmatrix} \nu_{1t} \\ \vdots \\ \nu_{Nt} \end{bmatrix} = \boldsymbol{\Pi} \boldsymbol{\varepsilon}_{(t-1)} + \boldsymbol{\nu}_{(t)}, \quad t = 2, 3, \dots, T,^2$$

where  $\boldsymbol{\Pi}$  is an  $N \times N$ , diagonal matrix consisting of scalars having absolute value less than 1.

Consistent with stationarity, the disturbances for the first observation,  $\boldsymbol{\varepsilon}_{(1)}$ , are assumed to be

---

<sup>2</sup>To clarify notation, notice that the vector  $\boldsymbol{\varepsilon}_i$  contains the  $T$  disturbances for the  $i$ th equation whereas the vector  $\boldsymbol{\varepsilon}_{(t)}$  contains the  $N$  disturbances for the different equations or cross-sectional elements at time  $t$ .

generated by  $\boldsymbol{\varepsilon}_{(1)} = \mathbf{A}^{-1}\mathbf{v}_{(1)}$ .<sup>3</sup> The innovations,  $\mathbf{v}_{(t)}$ ,  $t = 1, 2, \dots, T$  are independent and identically distributed random variables with  $E(\mathbf{v}_{(t)}) = \mathbf{0}$  and covariance matrix

$$(4) \quad E(\mathbf{v}_{(t)}\mathbf{v}_{(t)}') = \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{N1} & \cdots & \sigma_{NN} \end{bmatrix}.$$

In summary, the covariance model above assumes a diagonal  $\boldsymbol{\Pi}$  matrix, with  $N$  parameters specifying equation-specific, first-order serial correlation; together with a symmetric  $\boldsymbol{\Sigma}$  matrix with  $\frac{N(N+1)}{2}$  parameters, specifying the contemporaneous covariances.

The Parks estimator, which is a feasible generalized least squares procedure associated with the covariance model above, is represented by

$$(5) \quad \hat{\boldsymbol{\beta}}_P = (\mathbf{X}'\hat{\mathbf{P}}'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}_T)\hat{\mathbf{P}}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{P}}'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}_T)\hat{\mathbf{P}}\mathbf{y},$$

and

$$(6) \quad \hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_P) = (\mathbf{X}'\hat{\mathbf{P}}'(\hat{\boldsymbol{\Sigma}}^{-1} \otimes \mathbf{I}_T)\hat{\mathbf{P}}\mathbf{X})^{-1},$$

where  $\hat{\mathbf{P}}$  is the Prais-Winsten transformation matrix and  $\hat{\boldsymbol{\Sigma}}$  is the estimate of  $\boldsymbol{\Sigma}$  in Equation (4). Estimation of individual parameters is described in Judge et al. (1985, pages 485-490). Further details are given in the Appendix.

As a point of comparison, Beck and Katz's (1995) PCSE estimator is given by,

$$(7) \quad \hat{\boldsymbol{\beta}}_{PCSE} = (\mathbf{X}'\hat{\mathbf{P}}'\hat{\mathbf{P}}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{P}}'\hat{\mathbf{P}}\mathbf{y}$$

$$(8) \quad \hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_{PCSE}) = (\mathbf{X}'\hat{\mathbf{P}}'\hat{\mathbf{P}}\mathbf{X})^{-1}(\mathbf{X}'\hat{\mathbf{P}}'(\hat{\boldsymbol{\Sigma}} \otimes \mathbf{I}_T)\hat{\mathbf{P}}\mathbf{X})(\mathbf{X}'\hat{\mathbf{P}}'\hat{\mathbf{P}}\mathbf{X})^{-1}.$$

---

<sup>3</sup> Appendix 1 of Guilkey and Schmidt (1973) and Judge et al. (1985, p. 485-487) show how to obtain the elements of the matrix  $\mathbf{A}$ . If we let  $\mathbf{V}_0 = E(\boldsymbol{\varepsilon}_{(t)}\boldsymbol{\varepsilon}_{(t)}')$  for  $t = 1, 2, \dots, T$ , then from (3) and stationarity,  $\mathbf{V}_0 = \boldsymbol{\Pi}\mathbf{V}_0\boldsymbol{\Pi}' + \boldsymbol{\Sigma}$ . Guilkey and Schmidt show that this equation can be solved for the elements of  $\mathbf{V}_0$  in terms of the elements of  $\boldsymbol{\Pi}$  and  $\boldsymbol{\Sigma}$ . From  $\boldsymbol{\varepsilon}_{(1)} = \mathbf{A}^{-1}\mathbf{v}_{(1)}$  we have  $\mathbf{V}_0 = \mathbf{A}^{-1}\boldsymbol{\Sigma}(\mathbf{A}^{-1})'$  or  $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{V}_0\mathbf{A}'$ . If  $\mathbf{H}$  and  $\mathbf{B}$  are the Cholesky factors of  $\boldsymbol{\Sigma}$  and  $\mathbf{V}_0$ , respectively, then  $\mathbf{H}\mathbf{H}' = \mathbf{A}\mathbf{B}\mathbf{B}'\mathbf{A}'$  and  $\mathbf{A} = \mathbf{H}\mathbf{B}^{-1}$ .

### III. HYPOTHESIS TESTING: THE BOOTSTRAP VERSUS ASYMPTOTIC-BASED TESTS

Asymptotic-based tests. A common approach for testing linear hypotheses of the form

$H_0: \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$  involves the Wald statistic,

$$(9) \quad g = (\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r})'(\mathbf{R}\widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}})\mathbf{R}')^{-1}(\mathbf{R}\widehat{\boldsymbol{\beta}} - \mathbf{r}),$$

where the restriction matrix  $\mathbf{R}$  has  $q$  rows (the number of restrictions) and  $K$  columns (where  $K = \sum_{i=1}^M k_i$ ).<sup>4</sup> The test statistic,  $g$ , is asymptotically distributed as  $\chi_q^2$ . As noted above, there is ample evidence that asymptotic-based tests for the Parks model do not provide accurate inference. Rejection probabilities tend to be substantially in excess of their nominal levels and confidence intervals are too small.

Bootstrap methods can improve upon asymptotic-based tests. Horowitz (1997) and others have provided extensive surveys of the bootstrap literature. Horowitz (1997, p. 201) gives a succinct statement of the key bootstrap results:

“The bootstrap provides a higher-order asymptotic approximation to critical values for tests based on “smooth” asymptotically pivotal statistics. When a bootstrap-based critical value is used for such a test, the difference between the test’s true and nominal levels decreases more rapidly with increasing sample size than it does when the critical value is obtained from first-order asymptotic theory. Given a sufficiently large sample, the nominal level of the test will be closer to the true level when a bootstrap critical value is used than when a critical value based on first-order asymptotic theory is used.”

Since the Wald statistic,  $g$ , is asymptotically pivotal,<sup>5</sup> bootstrap approaches will provide improved accuracy compared with tests that rely on the asymptotic distribution.

---

<sup>4</sup> The bootstrapping approach that we propose is not limited to linear hypotheses. Suppose that we want to test the set of non-linear hypotheses  $\mathbf{R}(\boldsymbol{\beta}) = 0$  where the non-linear function has dimension  $q$ . The corresponding Wald statistic has the form  $g_{NL} = \mathbf{R}(\widehat{\boldsymbol{\beta}}_P)' \left[ \frac{\partial \mathbf{R}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}'} \Big|_{\widehat{\boldsymbol{\beta}}_P} \widehat{\mathbf{V}}(\widehat{\boldsymbol{\beta}}_P) \frac{\partial \mathbf{R}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \Big|_{\widehat{\boldsymbol{\beta}}_P} \right]^{-1} \mathbf{R}(\widehat{\boldsymbol{\beta}}_P)$ . This Wald statistic has the same asymptotic distribution as its linear analog, (9), and bootstrap testing would follow the same steps as those outlined below.

<sup>5</sup> With regard to the Wald statistic being asymptotically pivotal, Horowitz (1997) writes, “The arguments in Section 2a show that the bootstrap provides higher-order asymptotic approximations to the distributions and critical values of ‘smooth’ asymptotically pivotal statistics. These include test statistics whose asymptotic distributions are standard normal or chi-square.” Note that Wald statistics converge in distribution to chi-square, whose only parameter is the degrees of freedom. Hence the Wald statistic is asymptotically pivotal because it does not depend on parameters of the model’s data generating process.

Bootstrap Procedure. Below we give the steps for implementing a bootstrap test of the null hypothesis,  $H_0$ , in the context of a SUR model with AR(1) disturbances (i.e., the Parks model). Although the results that we show are based on a nonparametric bootstrap, it is useful for explanatory purposes to show how the nonparametric method differs from the simpler parametric method.

STEP 1: Estimate  $\beta_P$  from the unrestricted model and compute the test statistic,  $g$ , from (9). Call this test statistic  $\hat{g}$ .

STEP 2: Re-estimate the model under the restrictions imposed by the null hypothesis,  $R\beta = r$ , to obtain  $\tilde{\beta}$ ,  $\tilde{\Pi}$ ,  $\tilde{\Sigma}$ , and  $\tilde{A}$ . For the nonparametric bootstrap, we also need  $\tilde{E}$ , the  $N \times T$  matrix of residuals based on the constrained estimates.

STEP 3: Use the restricted estimates  $\tilde{\beta}$ ,  $\tilde{\Pi}$ , and  $\tilde{\Sigma}$  as the parameters of the data generating process described by Equations (1) through (4) above, together with the restriction,  $R\beta = r$ , to generate a bootstrap sample satisfying the hypothesis to be tested.

For a parametric bootstrap, one might start with the following:

STEP 3a: Draw an  $N \times T$  matrix  $U = \{u_{(t)}\}$  of standard normal random variables,  $u_{(t)} \sim N(\mathbf{0}, I_N)$ .

But for the nonparametric bootstrap we rely on information contained in the collection of constrained residuals,  $\tilde{E}$ . Using the restricted parameter estimates from STEP 2, make the following set of transformations to obtain a set of residuals,  $\{\tilde{u}_{(t)}\}$ , that can be treated as the starting point of the data generating process (DGP). Let  $\tilde{v}_{(t)} = \tilde{\epsilon}_{(t)} - \tilde{\Pi}\tilde{\epsilon}_{(t-1)}$  for  $t = 2, 3, \dots, T$ , and  $\tilde{v}_{(1)} = \tilde{A}\tilde{\epsilon}_{(1)}$ . Then let  $\tilde{u}_{(t)} = \tilde{H}^{-1}\tilde{v}_{(t)}$ , where  $\tilde{H}$  is the Cholesky factor of  $\tilde{\Sigma}$ .

For the bootstrap samples to satisfy the null hypothesis, the  $\{\tilde{u}_{(t)}\}$  should have row means of zero, and they should be uncorrelated. If those conditions are not met, one can whiten the residuals by first subtracting row averages from each of the corresponding elements of  $\{\tilde{u}_{(t)}\}$  to get the centered  $U_c$ . Correlation among the  $U_c$  rows can be eliminated by premultiplying  $U_c$  by the transposed inverse of the Cholesky factor of  $U_c U_c'$ . Hereafter, references to the set of residuals,  $\tilde{U} = \{\tilde{u}_{(t)}\}$  assumes that they have been whitened.

STEP 3a\*: For a nonparametric bootstrap, we draw a sample of  $T$   $\{\tilde{\mathbf{u}}_{(t)}\}$  vectors with replacement from this empirical distribution to form the columns of  $\mathbf{U}$ . It is the nonparametric counterpart to  $\mathbf{U}$  in Step 3a above, where the sampling was from a known distribution.

From this point onward, the remaining steps for both parametric and nonparametric bootstraps are the same.

STEP 3b: Transform the columns of  $\mathbf{U}$  to have covariance  $\tilde{\Sigma}$ ; i.e., construct  $\tilde{\mathbf{V}} = \{\tilde{\mathbf{H}}\mathbf{u}_{(t)}\} = \tilde{\mathbf{H}}\mathbf{U}$  where  $\tilde{\mathbf{H}}$  is the lower triangular Cholesky factor of  $\tilde{\Sigma}$ , such that  $\tilde{\mathbf{H}}\tilde{\mathbf{H}}' = \tilde{\Sigma}$ .

STEP 3c: Construct the disturbance vector  $\tilde{\boldsymbol{\varepsilon}}_{(t)}$ , where  $\tilde{\boldsymbol{\varepsilon}}_{(1)} = \tilde{\mathbf{A}}^{-1}\mathbf{v}_{(1)}$  and  $\tilde{\boldsymbol{\varepsilon}}_{(t)} = \tilde{\Pi}\tilde{\boldsymbol{\varepsilon}}_{(t-1)} + \tilde{\mathbf{v}}_{(t)}$  for  $t = 2, 3, \dots, T$ . Use these disturbances, model (2), and the restrictions implied by the hypothesis to be tested to generate the first bootstrap sample.

STEP 4: Estimate parameters for the unconstrained model from the first bootstrap sample ( $b=1$ ), compute the test statistic,  $g_1$ , and store it.

STEP 5: Repeat the process of generating a bootstrap sample, estimating the model, and computing the test statistic until one has  $B$  bootstrap samples and test statistics  $g_1, g_2, \dots, g_B$ . Davidson and MacKinnon (2004) recommend choosing  $B$  such that when  $\alpha$  is the level of significance of the test, the product  $\alpha(B + 1)$  is an integer, e.g.  $B = 999$ . Estimate the  $\alpha$ -level critical value for the test,  $g_{c\alpha}$ , as the  $(1 - \alpha)$ th quantile from the empirical distribution of the  $g_b$ 's.

STEP 6: Reject  $H_0$  at nominal  $\alpha$  level if the test statistic computed from the original sample,  $\hat{g}$  satisfies  $\hat{g} > g_{c\alpha}$ . Alternatively compute a  $p$ -value from STEP 5 as the fraction of the bootstrap samples with  $g_b > \hat{g}$ .

The above procedure can be suitably modified at STEPS 1 and 3 to deal with test statistics that depend on estimates of the restricted model (Lagrange multiplier tests) or on estimates of both restricted and unrestricted models (Likelihood ratio tests).

#### IV. RESULTS FROM MONTE CARLO EXPERIMENTS

Description of experiments. In this section we perform a series of Monte Carlo experiments to assess the performance of the bootstrap procedure described above. To do that, we construct synthetic panel datasets that are made to “look like” real datasets. The reason we do that is because the Parks model has  $\left(\frac{N^2+3N}{2}\right)$  unique parameters in the error variance-covariance

matrix. There is little guidance in how best to assign values to these parameters for the purpose of designing meaningful experiments.

Beck and Katz's simulation experiments are based on a substantial simplification of the error variance-covariance matrix. For example, when  $N=15$ , they reduce the number of unique parameters from 135 to 3 by assigning half of the panel units one variance value, the other half another, and setting all cross-sectional correlations the same. Further, their experiments do not allow for the interaction of cross-sectional and serial correlation.<sup>6</sup> The problem with this approach is that it greatly reduces the realism of the error variance-covariance matrix. This raises concerns about the external validity of the associated simulation results. In contrast, we adopt an innovative approach that assigns a unique value to every element in the error variance-covariance matrix, along with the values of the independent variable(s) in the experimental DGPs. We derive these values from actual panel datasets. The details of how we do this are given below.

The first dataset we work with is Grunfeld's (1958) investment data, one of the most widely used panel datasets in applied econometrics (Kleiber and Zeileis, 2010)<sup>7</sup>. The dataset consists of annual observations of three variables for 10 U.S. firms over a 20-year period (1935-1954).<sup>8</sup> The dependent variable is firm gross investment in plant and equipment ( $I$ ). The two explanatory variables are the market value of the firm at the end of the previous year ( $F$ ) and a capital stock measure ( $C$ ).

---

<sup>6</sup> Beck and Katz (1985, pages 640f.): "Varying degrees of heteroscedasticity were simulated by setting the variance of the first half of the units to 1 while the variance of the second half of the units was experimentally manipulated. The covariance matrix of this multivariate distribution was constructed so that all pairs of units were equally correlated, with the degree of correlation also experimentally manipulated. Errors were then generated so that the variances and covariances of the errors were proportional to the corresponding variances and covariances of the independent variable. The errors could therefore show panel heteroscedasticity and contemporaneous correlation, either alone or in combination."

<sup>7</sup> Greene (2012, page 342) writes, "The Grunfeld investment data...are a classic data set that have been used for decades to develop and demonstrate estimators for seemingly unrelated regressions."

<sup>8</sup> We use the version of the dataset that appears in Hill, Griffiths, and Lin (2008).



Our procedure for creating the synthetic datasets is best illustrated by example. Suppose we want to generate a synthetic dataset that “looks like” the Grunfeld data, except that it has dimensions  $N=2$  and  $T=20$ . We begin by extracting the data for the first two firms in the Grunfeld dataset. In this case, the data matrix  $\mathbf{X}$  consists of a constant term and the variables  $F$  and  $C$ . The twenty, time series values of the independent variables ( $F$  and  $C$ ) are set equal to their actual values in the Grunfeld data. To generate artificial values for the dependent variable  $I$ , we multiply  $\mathbf{X}$  by a coefficient vector  $\boldsymbol{\beta}$  whose elements are all set to zero, and then add simulated error terms.

The DGP for the simulated error terms is constructed so that the errors have the same nonspherical properties as residuals from a regression of the Grunfeld data. Specifically, we estimate Equation (2) using SUR and collect the residuals. These residuals are used to estimate the elements of  $\boldsymbol{\Pi}$  and  $\boldsymbol{\Sigma}$ . The estimated values are then set as the population values for the error variance-covariance matrix in the DGP that produces the simulated error terms. The simulated error terms are added to  $\mathbf{X}\boldsymbol{\beta}$  to produce simulated values of the dependent variable  $I$ .

By generating a new set of error terms, multiple synthetic datasets having dimensions  $N=2$  and  $T=20$  can be produced, each of which is constructed to have characteristics similar to the Grunfeld data. We then use these synthetic panel datasets to run experiments testing three linear restrictions, each having the form  $\mathbf{R}\boldsymbol{\beta} = \mathbf{0}$ :

$$(10.a) \quad \mathbf{R}_1 = [0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0],$$

$$(10.b) \quad \mathbf{R}_2 = [0 \quad 1 \quad 0 \quad 0 \quad -1 \quad 0], \text{ and}$$

$$(10.c) \quad \mathbf{R}_3 = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \end{bmatrix}.$$

Restriction (10.a) tests the statistical significance of a single parameter estimate: the coefficient on  $F$  in the equation for the first firm. (10.b) tests a linear combination of parameter estimates: whether the coefficient on  $F$  has the same value in the equations for the first and

second firms. (10.c) tests multiple linear combinations of parameter estimates: whether (i) the constant terms are equal in the equations for the first and second firms, and (ii) the coefficient on  $F$  is the same in the equations for the first and second firms. Our experiments are designed so that the respective null hypotheses are always true. We chose these three restrictions because they each represent a common type of hypothesis test found in empirical research.

(10.a)-(10.c) are easily modified to allow for different numbers of firms,  $N$ , in the synthetic, Grunfeld-type panel datasets. For these datasets, restriction matrices will have  $3N$  columns. The analogues to (10.a)–(10.c) for an alternative  $N$  value are identical in the first six columns, with zeros in the remaining  $3N - 6$  columns.

Results from synthetic panel datasets modelled after the Grunfeld data. The first three rows of the top panel of Table 1 ( $T=20$ ) report results of Monte Carlo experiments based on the Grunfeld data with  $N=5$ , testing each of the restrictions in (10.a)-(10.c). Each experiment consists of 1000 replications. The first column reports 5%, asymptotic critical values for the  $\chi^2$  distribution with 1 degree of freedom (Restrictions 1 and 2) and 2 degrees of freedom (Restriction 3). The next column reports the average critical values determined by the nonparametric bootstrap procedure described in Section III above. Note that these are *average* critical values because a critical value is produced for each replication, and the table summarizes the results from 1000 replications.

For example, in testing the significance of the coefficient for  $F$  in the equation for the first firm (Restriction 1), the asymptotic critical value is 3.841. This compares to an average critical value of 8.619 for the bootstrap procedure. While this is only one experiment, these results are qualitatively what we would expect: estimated standard errors for the Parks model are well-known for being biased downwards in finite samples, implying that that the asymptotic critical values will be smaller than they should be. A similar pattern emerges when testing Restrictions 2 and 3.

The next set of three rows repeats the experiments, except now the full set of 10 firms is used in creating the synthetic Grunfeld panel datasets. Note that the addition of data for the extra firms does more than just increase the sample size. It introduces a new set of variances and covariances, increasing the number of unique elements in the error variance-covariance matrix from 20 to 65. This exacerbates the bias in the Parks standard errors. While the asymptotic critical values are unchanged, the bootstrapped critical values increase to reflect the greater imprecision caused by having to estimate additional parameters.

The lower panel of Table 1 ( $T=11$ ) repeats the previous experiments, but this time only uses the first 11 years of the Grunfeld data. The reason for doing this is that the finite sample bias in the estimated Parks's asymptotic standard errors is known to increase as a function of  $N/T$  (Moundigbaye et al., 2018). We investigate this by decreasing  $T$  to where it is just larger than the number of firms ( $T=11, N=10$ ), noting that the Parks estimator cannot be estimated when  $T < N$ .<sup>9</sup> The six rows of the lower panel report the results for  $N=5$  and  $N=10$ . The results again correspond to expectations. Compared to their values in the top panel, the smaller  $N/T$  values are associated with larger bootstrapped critical values. Note that the asymptotic critical values are unchanged.

Table 2 calculates the Type I error rates associated with the critical values in Table 1. These should ideally equal 0.05, though some deviation is expected due to sampling error. Column (1) reports Type I error rates associated with the critical values taken from the  $\chi^2$  distribution. Column (2) does the same when the critical values are determined from the bootstrapping procedure described in STEP 6 above. The last column reports rejection rates associated with the PCSE estimator. In all cases, hypotheses are rejected whenever the sample

---

<sup>9</sup> In principle, the Parks estimator can be calculated when  $T=N$ . We found, however, that we sometimes encountered problems in our simulations in this case, so we set the lower bound of  $T=N+1$ .

statistic is greater than the critical value for a given replication. The values in the table report rejection rates for the 1000 replications for each experiment.

There are a number of noteworthy results here. First, the rejection rates using the asymptotic critical values range from 0.166 to 0.624. In other words, if we were to use a 5% significance level, we would reject the true null hypothesis anywhere from 17% to 62% of the time. Further, as foreshadowed above, performance deteriorates markedly as  $N$  increases holding  $T$  constant (within each panel in the table), and as  $T$  decreases holding  $N$  constant (from top panel to bottom panel). In contrast, the bootstrap procedure does much better. Rejection rates for the bootstrap range from 0.018 to 0.060, close to the 0.05 benchmark.

The last column provides a comparison with Beck and Katz's PCSE estimator. As noted in the introduction, the PCSE procedure has been promoted as producing standard errors, and associated test results, that are superior to the Parks estimator -- though at some cost in efficiency. Indeed, the improved performance of the PCSE over the Parks estimator with asymptotic standard errors is evident by comparing rejection rates in Column (3) and Column (1), respectively. It is also evident, however, that it performs substantially worse than the bootstrap procedure. A comparison of Columns (3) with Column (2) shows that rejection rates for the PCSE are further from the 0.05 benchmark than those for the bootstrap in 11 of 12 experiments, and of equal distance in one ( $T=11, N=10, Restriction=1$ ).

The last row in each panel reports mean column values across the different experiments for  $T=20$  and  $T=11$ , respectively. It provides a crude measure of overall performance, with values closer to 0.05 indicating better overall performance. The bootstrap procedure demonstrates superior performance over both the PCSE estimator and the Parks model with asymptotic critical values. In the  $T=20$  experiments, the mean rejection rates are 0.044 versus 0.154 and 0.276, respectively. In the  $T=11$  experiments, they are 0.039 versus 0.182 and 0.405.

Results from synthetic panel datasets modelled after two additional datasets. In this next section, we perform further performance tests. The goal is to investigate whether our bootstrap procedure continues to perform well when tested on synthetic datasets very different from those based on the Grunfeld data. Whereas the Grunfeld data related a firm’s investment to its market value and capital stock, the next two datasets we work with relate (i) tourism and crime in Italian provinces from 1985-2003, and (ii) foreign aid and real per capita GDP growth for a set of least developed countries (LDCs) from 1960-2000. We chose these studies because, in addition to offering a sharp contrast to Grunfeld, the associated data are strongly balanced, allow a relatively large number of  $N$  and  $T$  combinations, and are publicly available.

The first of these studies was published by Bruckner in 2013 in the *Journal of Applied Econometrics*. It estimates the effect of real per capita GDP growth on the growth in development aid for 44 countries over 25 years. Accordingly, our simulated datasets are comprised of these two variables.<sup>10,11</sup> We use the maximum number of time periods ( $T=25$ ) while allowing  $N$  to take on the values 5, 10, 15, 20, and 24 across the different experiments.

The second study was published by Biagi, Brandano, and Detotto in *Economics E-Journal* in 2012. It studies the effect of tourism on crime in 95 Italian provinces over a period of 18 years. In addition to the dependent variable measuring crime and the key explanatory variable measuring number of tourists, it includes control variables for economic growth, the level of income, the unemployment rate, population density, a measure of educational attainment, and a measure of criminal “deterrence.” Accordingly, our corresponding, simulated

---

<sup>10</sup> The regressions underlying our hypothesis tests are modelled after the regression reported in Table I, Column 3 on page 131 of Bruckner (2013).

<sup>11</sup> The corresponding restriction matrices for  $N=2$  are given by:

$$(10.a') \mathbf{R}_1 = \begin{bmatrix} 0 & 1 & 0 & 0 \end{bmatrix},$$

$$(10.b') \mathbf{R}_2 = \begin{bmatrix} 0 & 1 & 0 & -1 \end{bmatrix}, \text{ and}$$

$$(10.c') \mathbf{R}_3 = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}.$$

For larger  $N$ , the  $\mathbf{R}_i$  are identical in the first four columns, with zeros in the remaining  $2N-4$  columns.

datasets consist of eight variables.<sup>12,13</sup> We again used the maximum number of time periods ( $T=18$ ) while allowing  $N$  to take values equal to 5, 10, 15, and 17 across the different experiments. We followed the same procedure as previously in constructing synthetic panel datasets to resemble these two additional datasets.

Table 3 repeats the analysis of Table 2, focussing on the Type I error rates associated with testing restrictions  $\mathbf{R}_1$ ,  $\mathbf{R}_2$ , and  $\mathbf{R}_3$ . The top panel reports the experiments using the synthetic datasets derived from the Bruckner data. As was the case with the Grunfeld datasets, hypothesis tests which rely on asymptotic critical values (Column 1) generally perform poorly, with rejection rates ranging from 0.114 to 0.768. The latter value is not exceptional for the Parks estimator (see, for example, Table 2 in Beck and Katz, 1995; and Figures 5 and 6 in Moundigbaye et al., 2018).

The bootstrap (Column 2) again does much better, producing type I error rates that have a mean rate of 0.031 and stay within 0.034 of the nominal 0.05 test size. The PCSE approach (Column 3) does slightly better than the bootstrap with this data set. Rejection rates have a mean value across all experiments of 0.045 and stay within 0.022 of the nominal size. The fact that the PCSE can, in some circumstances, do very well, is not surprising (see, for example, Table 5, Columns 3 and 4 in Moundigbaye et al., 2018).

The bottom panel of Table 3 repeats the comparison, this time using synthetic panel datasets derived from the Biagi et al. (2012) data on Italian crime rates. The results for the Parks estimator with asymptotic critical values are similar to previous results, with a mean

---

<sup>12</sup> The regressions underlying our hypothesis tests are modelled after the regression reported in Table I, Column 1 on page 13 of Biagi et al. (2012).

<sup>13</sup> The corresponding restriction matrices for  $N=2$  are given by:

$$(10.a'') \mathbf{R}_1 = [0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0],$$

$$(10.b'') \mathbf{R}_2 = [0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ -1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0], \text{ and}$$

$$(10.c'') \mathbf{R}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

For larger  $N$ , the  $\mathbf{R}_i$  are identical in the first sixteen columns, with zeros in the remaining  $8N-16$  columns.

Type I error rate of 0.510. The corresponding Type I error rates for the PCSE approach are unacceptably large, as they were in Table 2, with a mean rejection rate of 0.234. The bootstrap procedure performs substantially better, with a mean rejection rate of 0.038.

## V. CONCLUSION

Although the Parks estimator has desirable efficiency properties both asymptotically and in finite samples; its poorly estimated standard errors limit its usefulness for inference. As a result, Beck and Katz's (1995) PCSE approach has become widely adopted as an alternative, as evidenced by its more than 2,400 Web of Science citations. Despite being more than 20 years old, the paper's annual citations continue to increase over time. Unfortunately, the PCSE approach, while reducing size distortion, does not entirely eliminate it; and it is based on a Prais-Winsten estimator that is less efficient than the Parks estimator.

This paper develops a bootstrap procedure for testing hypotheses using the more efficient Parks estimator. We illustrate the bootstrap's use in a number of experiments where we simulate panel datasets to "look like" real datasets. We show that the bootstrap procedure performs well with these data. While the PCSE approach can sometimes also perform well, the bootstrap usually performs better, often substantially better.

Up to this point, researchers working with panel datasets where the number of time periods is larger than the number of cross-sections have had to give up the efficiency of the Parks estimator to obtain greater accuracy in hypothesis testing using PCSE. The bootstrapping procedure presented here allows researchers to retain the efficient Parks estimator and to have test results that are generally more accurate than those offered by the PCSE approach.

## REFERENCES

- Beck, N. and Katz, J.N. (1995). What to do (and not to do) with time series cross-section data. *American Political Science Review*, 89, 634-647.
- Biagi, B., Brandano, M.G. & Detotto, C. (2012). The effect of tourism on crime in Italy: A dynamic panel approach. *Economics: The Open-Access, Open-Assessment E-Journal*, 6, 1–24.  
<http://dx.doi.org/10.5018/economics-ejournal.ja.2012-25>.
- Bruckner, M. (2013). On the simultaneity problem in the aid and growth debate. *Journal of Applied Econometrics*, 28, 126-150.
- Chen, X., Lin, S., and Reed, W.R. (2010). A Monte Carlo evaluation of the efficiency of the PCSE estimator. *Applied Economics Letters*, 17, 7-10.
- Davidson, R. and MacKinnon, J.G. (2004). *Econometric Theory and Methods*, New York, Oxford University Press.
- Greene, W.H. (2012). *Econometric Analysis*, 7th edition. Upper Saddle River, N.J: Prentice Hall.
- Grunfeld, Y. (1958). *The Determinants of Corporate Investment*, Unpublished Ph.D. Thesis, Department of Economics, University of Chicago.
- Guilkey, D.K. and Schmidt, P. (1973). Estimation of Seemingly Unrelated Regressions with vector autoregressive errors. *Journal of the American Statistical Association*, 68, 642-647.
- Hill, R.C., Griffiths, W.E. and Lim, G.C. (2008). *Principles of Econometrics*, 3rd edition, New York, John Wiley and Sons.
- Horowitz, J.L. (1997) Bootstrap methods in econometrics: Theory and numerical performance, in M. Kreps and K. F. Wallis, *Advances in Economics and Econometrics: Theory and Applications*, Cambridge, Cambridge University Press, Vol. III, 188-222.
- Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepohl, H., and Lee, T-C. (1985). *The Theory and Practice of Econometrics*, 2nd edition, New York, John Wiley and Sons.
- Kleiber, C. and Zeileis, A. (2010). The Grunfeld data at 50. *German Economic Review*, 11(4), 404-417.
- Kmenta, J. and Gilbert, R.F. (1970). Estimation of seemingly unrelated regressions with autoregressive disturbances. *Journal of the American Statistical Association*, 65, 186-197.



- Messemer, C. (2003) “Kmenta-Gilbert Revisited: Small Sample Properties of the Parks Estimator.” Essays in Multivariate Time Series Analysis with Applications to the Movement of Real Average Earnings in the US Economy, Unpublished Ph.D. dissertation, Department of Economics, University of Washington.
- Moundigbaye, M., Rea, W.S., and Reed, W.R. (2018). Which panel data estimator should I use?: A corrigendum and extension. *Economics: The Open-Access, Open-Assessment E-Journal*, 12 (2018-4): 1–31.  
<http://dx.doi.org/10.5018/economics-ejournal.ja.2018-4>.
- Parks, R.W. (1967). Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated. *Journal of the American Statistical Association*, 62, 500-509.
- Reed, W.R. and Webb, R. (2010). The PCSE Estimator is good – just not as good as you think. *Journal of Time Series Econometrics*, 2(1), Article 8.
- Rilstone, P. and Veall, M. (1996). Using bootstrapped confidence intervals for improved inferences with Seemingly Unrelated Regression equations. *Econometric Theory*, 12(3), 569-580.

**TABLE 1**  
**Comparison of Bootstrapped with Asymptotic Critical Values**

<i>N</i>	<i>Restriction</i>	<u>5% Critical Values</u>	
		<i>Asymptotic</i> (1)	<i>Bootstrapped</i> (2)
<b><i>T = 20</i></b>			
<b>5</b>	<b>1</b>	3.841	8.619
<b>5</b>	<b>2</b>	3.841	9.803
<b>5</b>	<b>3</b>	5.991	13.786
<b>10</b>	<b>1</b>	3.841	14.318
<b>10</b>	<b>2</b>	3.841	24.129
<b>10</b>	<b>3</b>	5.991	23.831
<b><i>T = 11</i></b>			
<b>5</b>	<b>1</b>	3.841	17.611
<b>5</b>	<b>2</b>	3.841	22.199
<b>5</b>	<b>3</b>	5.991	36.485
<b>10</b>	<b>1</b>	3.841	63.838
<b>10</b>	<b>2</b>	3.841	45.079
<b>10</b>	<b>3</b>	5.991	67.832

NOTE: *N* and *T* correspond to the number of cross-sectional units and time periods, respectively. Restrictions 1 through 3 are described in the text related to discussion of Equations (10.a) through (10.c). Asymptotic critical values are for the  $\chi^2$  distribution with 1 degree of freedom (Restrictions 1 and 2) and 2 degrees of freedom (Restriction 3). The bootstrapped critical values are averages over 1000 replications for each experiment. The procedure for calculating bootstrapped critical values is discussed in Section III in the text.

**TABLE 2**  
**Comparison of Type I Error Rates Using Asymptotic and Bootstrapped Critical Values:**  
**Grunfeld Data**

<i>N</i>	<i>Restriction</i>	<i>Type I Error Rates</i>		
		<i>Asymptotic</i> <i>(1)</i>	<i>Bootstrapped</i> <i>(2)</i>	<i>PCSE</i> <i>(3)</i>
<i>T = 20</i>				
<i>5</i>	<i>1</i>	0.166	0.050	0.126
<i>5</i>	<i>2</i>	0.174	0.050	0.112
<i>5</i>	<i>3</i>	0.244	0.040	0.220
<i>10</i>	<i>1</i>	0.304	0.060	0.112
<i>10</i>	<i>2</i>	0.326	0.018	0.092
<i>10</i>	<i>3</i>	0.444	0.044	0.262
<i>Mean</i>	<i>---</i>	<i>0.276</i>	<i>0.044</i>	<i>0.154</i>
<i>T = 11</i>				
<i>5</i>	<i>1</i>	0.240	0.028	0.094
<i>5</i>	<i>2</i>	0.288	0.038	0.136
<i>5</i>	<i>3</i>	0.426	0.052	0.300
<i>10</i>	<i>1</i>	0.392	0.030	0.070
<i>10</i>	<i>2</i>	0.462	0.032	0.134
<i>10</i>	<i>3</i>	0.624	0.052	0.360
<i>Mean</i>	<i>---</i>	<i>0.405</i>	<i>0.039</i>	<i>0.182</i>

NOTE: *N* and *T* correspond to the number of cross-sectional units and time periods, respectively. Restrictions 1 through 3 are described in the text related to discussion of Equations (10.a) through (10.c). Type I error rates report the percent of 1000 Monte Carlo experiments where the null hypothesis was rejected.

**TABLE 3**  
**Robustness Check: Further Analysis with Two Additional Datasets**

<i>N</i>	<i>Restriction</i>	<i>Type I Error Rates</i>		
		<i>Asymptotic</i> (1)	<i>Bootstrapped</i> (2)	<i>PCSE</i> (3)
<b>Aid and Growth Data: <i>T</i> = 25 (Bruckner, 2013)</b>				
<i>5</i>	<i>1</i>	0.162	0.050	0.042
<i>5</i>	<i>2</i>	0.114	0.038	0.048
<i>5</i>	<i>3</i>	0.122	0.042	0.042
<i>10</i>	<i>1</i>	0.226	0.042	0.046
<i>10</i>	<i>2</i>	0.216	0.024	0.040
<i>10</i>	<i>3</i>	0.212	0.030	0.028
<i>15</i>	<i>1</i>	0.338	0.030	0.048
<i>15</i>	<i>2</i>	0.350	0.026	0.072
<i>15</i>	<i>3</i>	0.362	0.040	0.034
<i>20</i>	<i>1</i>	0.564	0.020	0.046
<i>20</i>	<i>2</i>	0.598	0.016	0.042
<i>20</i>	<i>3</i>	0.618	0.032	0.038
<i>24</i>	<i>1</i>	0.682	0.022	0.042
<i>24</i>	<i>2</i>	0.768	0.034	0.066
<i>24</i>	<i>3</i>	0.768	0.026	0.042
<i>Mean</i>		<i>0.407</i>	<i>0.031</i>	<i>0.045</i>
<b>Tourism and Crime Data: <i>T</i> = 18 (Biagi et al., 2012)</b>				
<i>5</i>	<i>1</i>	0.254	0.050	0.19
<i>5</i>	<i>2</i>	0.296	0.070	0.19
<i>5</i>	<i>3</i>	0.436	0.062	0.34
<i>10</i>	<i>1</i>	0.454	0.050	0.19
<i>10</i>	<i>2</i>	0.424	0.038	0.15
<i>10</i>	<i>3</i>	0.596	0.066	0.37
<i>15</i>	<i>1</i>	0.570	0.014	0.17
<i>15</i>	<i>2</i>	0.584	0.028	0.20
<i>15</i>	<i>3</i>	0.820	0.012	0.34
<i>17</i>	<i>1</i>	0.494	0.018	0.19
<i>17</i>	<i>2</i>	0.480	0.022	0.16
<i>17</i>	<i>3</i>	0.708	0.020	0.32
<i>Mean</i>		<i>0.510</i>	<i>0.038</i>	<i>0.234</i>

NOTE:  $N$  and  $T$  correspond to the number of cross-sectional units and time periods, respectively. Restrictions 1 through 3 are described in the text related to the discussion of Equations (10.a) through (10.c) for the Bruckner (2013) and Biagi et al. (2012) datasets. Type I error rates report the percent of 1000 Monte Carlo experiments where the null hypothesis was rejected.

**APPENDIX**  
**Step-by- Step Procedure for Computing the Parks Estimator**

Assume one has to estimate the panel data equation:

$$y = X\beta + e. \quad (\text{A.0})$$

The Parks FGLS estimator with SUR residuals is implemented in 5 steps as below:

i. Implement SUR with contemporaneously correlated disturbances.

- Estimate equation (A.0) using ordinary least squares and collect the residuals to compute the contemporaneous variance-covariance matrix  $\hat{\Sigma}_{OLS}$ .

$$\hat{\Sigma}_{OLS} = \frac{1}{T} \Xi'_{OLS} \Xi_{OLS}, \quad (\text{A.1})$$

where  $\Xi_{OLS}$  is a  $T \times N$  matrix whose columns are made of individual OLS errors.

- Construct the OLS residuals full variance-covariance matrix as a block diagonal matrix.

$$\hat{\Omega}_{OLS} = \hat{\Sigma}_{OLS} \otimes I_T. \quad (\text{A.2})$$

- The SUR estimate with contemporaneously correlated errors is given by:

$$\tilde{\beta}_{SUR} = [X'(\hat{\Omega}_{OLS}^{-1} \otimes I_T)X]^{-1} [X'(\hat{\Omega}_{OLS}^{-1} \otimes I_T)y]. \quad (\text{A.3})$$

ii. Use errors from the SUR model with contemporaneously correlated errors to estimate the first-order serial correlation coefficient for each equation in system (A.0).

$$\tilde{\epsilon} = y - X\tilde{\beta}_{SUR}; \quad (\text{A.4})$$

and

$$\hat{\rho}_i = \frac{\sum_{t=2}^T \tilde{\epsilon}_{it} \tilde{\epsilon}_{i,t-1}}{\sum_{t=1}^{T-1} \tilde{\epsilon}_{it}^2}. \quad (\text{A.5})$$

- For  $t = 2, 3, \dots, T$ , and for each  $i$ , transform  $y_i$  and  $X_i$  series using  $\hat{\rho}_i$ .

$$y_i^* = \tilde{P}_{0i} y_i \quad \text{and} \quad X_i^* = \hat{P}_{0i} X_i, \quad (\text{A.6})$$

where

$$\hat{P}_{0i} = \begin{bmatrix} -\hat{\rho}_i & 1 & 0 & \dots & 0 \\ 0 & -\hat{\rho}_i & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -\hat{\rho}_i & 1 \end{bmatrix}. \quad (\text{A.7})$$

- Regress  $y_i^\bullet$  on  $X_i^\bullet$  using OLS and collect the residuals to compute the estimate  $\widehat{\Sigma}$  of the contemporaneous variance-covariance matrix  $\Sigma$ .

$$\widehat{\Sigma} = \frac{1}{T-1} \mathbf{E}' \mathbf{E}, \quad (\text{A.8})$$

where  $\mathbf{E}$  is a  $(T-1) \times N$  matrix whose columns are OLS residuals from the regression of  $y_i^\bullet$  on  $X_i^\bullet$ .

iii. Construct the full transformation matrix  $\widehat{\mathbf{P}}$ , such that:

$$\widehat{\mathbf{P}} \widehat{\mathbf{\Omega}} \widehat{\mathbf{P}}' = \widehat{\Sigma} \otimes \mathbf{I}_T. \quad (\text{A.9})$$

$\widehat{\mathbf{P}}$  has the form:

$$\widehat{\mathbf{P}} = \begin{bmatrix} \widehat{\mathbf{P}}_{11} & 0 & \dots & 0 \\ \widehat{\mathbf{P}}_{21} & \widehat{\mathbf{P}}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \widehat{\mathbf{P}}_{N1} & \dots & \widehat{\mathbf{P}}_{N,N-1} & \widehat{\mathbf{P}}_{NN} \end{bmatrix}, \quad (\text{A.10})$$

where

$$\widehat{\mathbf{P}}_{ii} = \begin{bmatrix} \widehat{\alpha}_{ii} & 0 & 0 & \dots & 0 \\ -\widehat{\rho}_i & 1 & 0 & \ddots & \vdots \\ 0 & -\widehat{\rho}_i & 1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & -\widehat{\rho}_i & 1 \end{bmatrix}; \quad (\text{A.11})$$

$$\widehat{\mathbf{P}}_{ij} = \begin{bmatrix} \widehat{\alpha}_{ii} & 0 & \dots & 0 \\ 0 & 0 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 \end{bmatrix}; \quad (\text{A.12})$$

$$\mathbf{A} = \begin{bmatrix} \widehat{\alpha}_{11} & 0 & \dots & 0 \\ \widehat{\alpha}_{21} & \widehat{\alpha}_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \widehat{\alpha}_{N1} & \widehat{\alpha}_{N2} & \dots & \widehat{\alpha}_{NN} \end{bmatrix}. \quad (\text{A.13})$$

where  $\mathbf{A}$  is defined as the product  $H'(B)^{-1}$ ,  $B$  and  $H$  are upper triangular matrices satisfying  $H'H = \widehat{\Sigma}$  and  $B'B = V_0$ , and  $V_0$  is the contemporaneous variance-covariance matrix of the error term  $\tilde{\mathbf{E}}$  and is defined as below. By construction, the matrix  $\mathbf{A}$  contains initial parameters ensuring the stationarity of the error term in equation (A.0).

$$V_0 = \begin{bmatrix} \frac{\hat{\sigma}_{11}}{1-\hat{\rho}_1^2} & \frac{\hat{\sigma}_{12}}{1-\hat{\rho}_1\hat{\rho}_2} & \cdots & \frac{\hat{\sigma}_{1N}}{1-\hat{\rho}_1\hat{\rho}_N} \\ \frac{\hat{\sigma}_{21}}{1-\hat{\rho}_2\hat{\rho}_1} & \frac{\hat{\sigma}_{22}}{1-\hat{\rho}_2^2} & \cdots & \frac{\hat{\sigma}_{2N}}{1-\hat{\rho}_2\hat{\rho}_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\hat{\sigma}_{N1}}{1-\hat{\rho}_{N1}\hat{\rho}_N} & \frac{\hat{\sigma}_{N2}}{1-\hat{\rho}_1\hat{\rho}_{N2}} & \cdots & \frac{\hat{\sigma}_{NN}}{1-\hat{\rho}_N^2} \end{bmatrix}; \quad (\text{A.14})$$

$$\hat{\sigma}_{11} = \hat{\Sigma}_{ii}. \quad (\text{A.15})$$

It follows that:

$$\hat{\Omega}^{-1} = \hat{P}'(\hat{\Sigma}^{-1} \otimes \mathbf{I}_T)\hat{P}. \quad (\text{A.16})$$

- iv. Apply  $\hat{P}$  to transform  $y$  and  $X$  and use OLS on the transformed data to get the Parks estimator of the SUR model with contemporaneously and first order serially correlated errors.

$$y^* = \hat{P}y. \quad (\text{A.17})$$

$$X^* = \hat{P}X. \quad (\text{A.18})$$

$$\hat{\beta}_{Parks} = [X^{*'}(\hat{\Sigma}^{-1} \otimes \mathbf{I}_T)X^*]^{-1}[X^{*'}(\hat{\Sigma}^{-1} \otimes \mathbf{I}_T)y^*]. \quad (\text{A.19})$$

A consistent estimator of the covariance matrix of  $\hat{\beta}_{Parks}$  is defined as:

$$V(\hat{\beta}_{Parks}) = [X^{*'}(\hat{\Sigma}^{-1} \otimes \mathbf{I}_T)X^*]^{-1}. \quad (\text{A.20})$$