DEPARTMENT OF ECONOMICS

COLLEGE OF BUSINESS AND ECONOMICS

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

A MONTE CARLO EVALUATION OF THE EFFICIENCY OF THE PCSE ESTIMATOR

by

Xiujian Chen
Department of Economics
California State University, Fullerton

Shu Lin
Department of Economics
Florida Atlantic University

W. Robert Reed*
Department of Economics
University of Canterbury

WORKING PAPER

No. 14/2006

Department of Economics, College of Business and Economics
University of Canterbury, Private Bag 4800, Christchurch
New Zealand

WORKING PAPER No. 14/2006

A MONTE CARLO EVALUATION OF THE EFFICIENCY OF THE PCSE ESTIMATOR

by

Xiujian Chen
Department of Economics
California State University, Fullerton

Shu Lin
Department of Economics
Florida Atlantic University

W. Robert Reed*
Department of Economics
University of Canterbury

November 3, 2006

Abstract

Panel data characterized by groupwise heteroscedasticity, cross-sectional correlation, and AR(1) serial correlation pose problems for econometric analyses. It is well known that the asymptotically efficient, FGLS estimator (Parks) sometimes performs poorly in finite samples. In a widely cited paper, Beck and Katz (1995) claim that their estimator (PCSE) is able to produce more accurate coefficient standard errors without any loss in efficiency in "practical research situations." This study disputes that claim. We find that the PCSE estimator is usually less efficient than Parks -- and substantially so -- except when the number of time periods is close to the number of cross-sections.

JEL Categories: C23, C15

Keywords: Panel data estimation, Monte Carlo analysis, FGLS, Parks, PCSE, finite sample

*The corresponding author is W. Robert Reed, Professor of Economics, University of Canterbury, Private Bag 4800, Christchurch, New Zealand. Email: bobreednz@yahoo.com. Phone: +64 3 364 2846. Fax: +64 3 364 2635.

Acknowledgments: An earlier draft of this paper was presented at the University of Oklahoma, and the 11th International Conference on Panel Data. We acknowledge helpful comments from Kevin Grier, Cynthia Rogers, and Aaron Smallwood.

1. INTRODUCTION

Panel data characterized by heteroscedasticity, serial correlation, and cross-sectional correlation raise serious issues for econometric analyses. An oft-employed procedure for data of this sort is the FGLS estimator proposed by Parks (1968). When the data generating process (DGP) is characterized by groupwise heteroscedasticity, time-invariant cross-sectional correlation, and first-order (AR[1]) serial correlation, the Parks estimator is asymptotically efficient.[1] It is well-known, however, that the Parks estimator can sometimes perform poorly in finite samples, particularly with respect to estimating coefficient standard errors.

A recent paper by Beck and Katz (1995) -- henceforth BK -- proposes an alternative, two-step estimator. In the first step, the data are transformed to eliminate serial correlation.[2] In the second step, OLS is applied to the transformed data, and the standard errors are corrected for cross-sectional correlation. Based on their Monte Carlo analyses, BK conclude that their "panel-corrected standard error" (PCSE) estimator produces more accurate standard error estimates, without any loss in efficiency. It is noteworthy that the PCSE procedure assumes the same error variance-covariance matrix (and estimates the same parameters) as the Parks estimator.[3]

If BK's results were generalizable to actual panel data, it would be a very useful finding. It promises an important benefit without any cost. Indeed, the paper and corresponding PCSE estimator have been highly influential. A recent count identifies over 500 citations of BK on Web of Science (e.g. Ferguson and Schularick, 2006;

[1] For example, in STATA, the "xtgls" options (i) "panels(hteroscedastic)", (ii) "panels(correlated)" and (iii) "corr(ar1/psar1)" correspond to these three types of nonspherical error variance-covariance behaviors.
[2] The "xtpcse" procedure in STATA uses a Prais-Winsten transformation in this first stage.
[3] As a result, the better performance of the PCSE estimator cannot be attributed to the "shrinkage principle" (Diebold, 2004, page 45).

Yermack, 2006; Dejuan and Luengo-Prado, 2006; and Lapré and Tsikriktsis, 2006). Further, the PCSE estimator is now a standard procedure in many statistical software packages, including STATA, GAUSS, RATS, and Shazam.

Unfortunately, our analysis is unable to confirm BK's efficiency results. Using a different set of Monte Carlo parameters patterned after actual panel data, we find that the PCSE estimator is almost always less efficient than Parks, often substantially so.

2. THE DATA GENERATING PROCESS AND A MEASURE FOR RELATIVE EFFICIENCY

Following BK, we assume that the DGP is given by:

$$(1) \qquad \begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{i} \\ \boldsymbol{i} \\ \vdots \\ \boldsymbol{i} \end{bmatrix} \beta_0 + \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix} \beta_x + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{bmatrix}, \text{ or } \quad \boldsymbol{y} = \boldsymbol{i}_{NT}\beta_0 + \boldsymbol{X}\beta_x + \boldsymbol{\varepsilon},$$

where N and T are the number of cross-sectional units and time periods, respectively; $\boldsymbol{y}_i$ is a $T \times 1$ vector of observations of the dependent variable for the i^{th} cross-sectional unit; $\boldsymbol{i}$ is a $T \times 1$ vector of ones; $\boldsymbol{X_i}$ is a $T \times 1$ vector of observations of the explanatory variable; β_0 and β_x are scalars; and $\boldsymbol{\varepsilon}_i$ is a $T \times 1$ vector of error terms, where $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Omega}_{NT})$.

The error structure, $\boldsymbol{\Omega}_{NT}$, is based on the Parks model (Parks, 1967). It assumes (i) groupwise heteroscedasticity, (ii) first-order serial correlation, and (iii) time-invariant cross-sectional correlation, imposing the following specification for $\boldsymbol{\Omega}_{NT}$:

$$(2) \qquad \boldsymbol{\Omega}_{NT} = \boldsymbol{\Sigma} \otimes \boldsymbol{\Pi},$$

where $\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{\varepsilon,11} & \sigma_{\varepsilon,12} & \cdots & \sigma_{\varepsilon,1N} \\ \sigma_{\varepsilon,21} & \sigma_{\varepsilon,22} & \cdots & \sigma_{\varepsilon,2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\varepsilon,N1} & \sigma_{\varepsilon,N2} & \cdots & \sigma_{\varepsilon,NN} \end{bmatrix}$,

$\boldsymbol{\Pi} = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix}$, $\varepsilon_{it} = \rho\varepsilon_{i,t-1} + u_{it}$, and $\sigma_{\varepsilon,ij} = \dfrac{\sigma_{u,ij}}{1-\rho^2}$.

There are a total of $\dfrac{N^2+N+2}{2}$ unique parameters in $\boldsymbol{\Omega}_{NT}$ (the $\sigma_{\varepsilon,ij}$'s and ρ). Each of these must be given a value in order to generate simulated data. BK emphasize that their Monte Carlo analyses were designed to provide guidance to researchers using panel data sets that are likely to be encountered in "practical research situations." However, even a moderately sized panel data set of 10 cross-sectional units requires setting more than 50 unique parameters. How is one to know whether the parameter values chosen by the researcher for the Monte Carlo simulations are those that typify practical research situations?

Our approach is to take estimated values of these parameters from previous research we have done. In particular, we use (i) real, per capita personal income data from U.S. states; and (ii) international real, per capita GDP data. Further, we work with studies in which the dependent variable is in (i) level and (ii) difference (growth) form. And, since these studies derived their estimates of the $\sigma_{\varepsilon,ij}$'s and ρ using residuals from regression equations, we reference two different regression specifications.[4]

[4] The main difference in the two regression specifications is that version 1 includes cross-sectional fixed effects, while version 2 includes both cross-sectional and time period fixed effects. To give an idea of the

This allows us to create eight different artificial data environments, each patterned after U.S. or international data (INT), income data that are either in level (L) or difference (D) form, and a particular type of residual-producing regression specification (1 or 2). We designate these US-L1, US-D1, US-L2, US-D2, INT-L1, INT-D1, INT-L2, and INT-D2. Our Monte Carlo experiments set values for the $\sigma_{\varepsilon,ij}$'s and ρ so that they "look like" estimated values from actual panel sets of a particular size (N,T) for each of the eight data environments.

Like BK, we generate 1000 simulated panel data sets for each (N,T) experiment. For every panel data set (replication), we estimate β_x in Equation (1) using both Parks and PCSE.[5] Define β_x^* as the true population value of β_x, and $\hat{\beta}_{Parks}^{(r)}$ and $\hat{\beta}_{PCSE}^{(r)}$ as the Parks and PCSE estimates of β_x for a given replication, *r*. We compare the efficiency performance of the two estimators using BK's measure of relative efficiency:

$$Relative\ Efficiency = 100 \cdot \frac{\sqrt{\sum_{r=1}^{1000}\left(\hat{\beta}_{Parks}^{(r)} - \beta_x^*\right)^2}}{\sqrt{\sum_{r=1}^{1000}\left(\hat{\beta}_{PCSE}^{(r)} - \beta_x^*\right)^2}}.$$

"Relative Efficiency" values less than 100 indicate that PCSE is less efficient than Parks.

3. RESULTS

Table 1 reports the results of our Monte Carlo simulations. The presentation is patterned after Table 5 (page 642) in BK. BK report that the PCSE estimator is either more efficient, or only slightly less efficient, than the Parks estimator except for "extreme

---

difference this makes for the residuals, the R^2 associated with the first specification usually ran around 0.60 for the US-Level data, compared to R^2 values that were typically over 0.90 using the second specification.

[5] All the programs used for this analysis were written in SAS/IML. The formulae for the Parks and PCSE estimators were constructed to exactly match the output from STATA's "xtgls" and "xtpcse" procedures, using the (i) "panels(correlated)" and "corr(ar1)", and (ii) "correlation(ar1)" options, respectively (we note that the default cross-sectional correlation option for the "xtpcse" option is groupwise heteroscedasticity and time-invariant cross-sectional correlation).

cases" (page 645) where the "average contemporaneous correlation is at least 0.50 and the time sample is quite long" (page 642). In contrast, we find that the PCSE estimator is almost always less efficient than Parks, sometimes substantially so. For panel data sets of size N=10 and T=20, we find that "Relative Efficiency" is less than 50% in four of the eight artificial data environments, and never higher than 74%.

As expected, we find that the efficiency advantage of Parks increases monotonically with T. As T increases, there are more observations available to estimate each cross-sectional covariance term. This increases the reliability of the FGLS estimates, enhancing the associated efficiency advantages. With a few exceptions, the efficiency of the PCSE estimator approaches that of Parks only when T is very close to N.

Why are our results different from those of BK? Table 2 makes it clear that it is not because our simulated data are driven by extreme values of either serial correlation or cross-sectional correlation. Each cell reports the (i) average correlation coefficient and (ii) average, absolute value of the cross-sectional correlation terms for the 1000 data sets corresponding to that experiment. There is a wide range of serial correlation behavior across the different artificial data environments. Further, most of the simulated data sets (cf. the last six columns) have average contemporaneous correlation values well below the 0.50 value that BK identify as problematic.

Most likely, the strong performance of the PCSE estimator reported by BK is an artifact of the particular parameter values they selected for their simulations. As discussed above, even small panel data sets with relatively few cross-sectional units have a large number of unique parameters in the error variance-covariance matrix. It is

difficult to know how one should set these parameters. For example, the average contemporaneous correlation may be less important for determining efficiency than the dispersion of the cross-sectional covariance values. Further, it is unclear which particular combinations represent "practical data situations."[6] For these reasons, we think that simulations using error variance-covariance parameter values that are patterned after real panel data sets provide a better means of evaluating likely estimator performance in "practical research situations."

4. CONCLUSION

This paper evaluates the efficiency of the PCSE estimator. Beck and Katz (1995) claim that the PCSE estimator is more efficient, or only slightly less efficient, than the Parks estimator except for extreme cases that researchers are unlikely to encounter in practice.

In contrast, this study finds that the PCSE estimator is usually less efficient than Parks -- and substantially so -- except when the number of time periods is close to the number of cross-sections. Our findings are consistent across a wide variety of Monte Carlo environments patterned after actual panel data. They indicate that researchers should be aware that use of the PCSE estimator may come at a considerable cost in efficiency.

[6] For example, the simulations underlying the Relative Efficiency results of BK's Table 5 assume no serial correlation.

REFERENCES

Beck, Nathaniel and Jonathan N. Katz, 1995, What to do (and not to do) with time-series cross-section data, American Political Science Review 89, 634-647.

Dejuan, Joseph P. and Maria Jose Luengo-Prado, 2006, Consumption and aggregate constraints: international evidence, Oxford Bulletin of Economics and Statistics 68, 81-99.

Diebold, Francis X, 2004, Elements of Forecasting, 3rd Edition. (South-Western, Mason: Ohio).

Ferguson, Niall and Moritz Schularick, 2006, The empire effect: the determinants of country risk in the first age of globalization: 1880-1913, Journal of Economic History 66, 283-312.

Lapré, Michael A. and Nikos Tsikriktsis, 2006, Organizational learning curves for customer dissatisfaction: heterogeneity across airlines, Management Science 52, 352-366.

Parks, Richard, 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.

Yermack, David, 2006, Flights of fancy: corporate jets, CEO perquisites, and inferior shareholder returns, Journal of Financial Economics 80, 211-242.

Table 1:
Relative Efficiency of PCSE Compared to Parks (%)

		Data Generating Process[c]							
N[a]	T[b]	US-L1	US-G1	US-L2	US-G2	INT-L1	INT-G1	INT-L2	INT-G2
5	10	99.4	95.7	61.8	56.7	108.9	99.9	61.5	57.6
	15	78.0	95.3	51.0	42.1	94.2	89.0	48.0	43.1
	20	55.5	83.3	41.2	33.1	94.4	80.9	38.8	31.7
	25	52.4	74.5	34.5	26.6	94.6	74.3	31.8	25.9
10	10	98.4	96.5	94.1	93.5	93.9	94.7	93.9	92.1
	15	94.0	88.0	71.5	67.9	82.3	84.8	71.2	66.6
	20	74.3	81.5	59.3	68.0	76.0	78.2	57.6	52.5
	25	54.4	66.1	49.6	44.7	71.3	74.0	49.9	43.9
20	20	91.1	96.1	97.1	96.8	97.5	95.9	97.3	95.9
	25	80.0	78.8	82.8	79.5	86.5	81.3	82.4	78.6

NOTE: $\textit{Relative Efficiency} = 100 \cdot \dfrac{\sqrt{\sum_{r=1}^{1000}\left(\hat{\beta}_{Parks}^{(r)} - \beta_x^*\right)^2}}{\sqrt{\sum_{r=1}^{1000}\left(\hat{\beta}_{PCSE}^{(r)} - \beta_x^*\right)^2}}$. Values less than 100% indicate that PCSE is less efficient than Parks.

[a] Number of cross-sectional units.
[b] Number of time periods.
[c] Indicates the type of actual panel data after which the simulated data are patterned. See Section 2 for category definitions.

Table 2:
Mean Serial Correlation and Mean Cross-Sectional Correlation of Simulated Data Sets

		Data Generating Process[c]							
N[a]	T[b]	US-L1	US-G1	US-L2	US-G2	INT-L1	INT-G1	INT-L2	INT-G2
5	10	0.35 / 0.65	0.09 / 0.59	0.42 / 0.36	0.19 / 0.34	0.41 / 0.36	-0.09 / 0.27	0.38 / 0.38	-0.08 / 0.34
	15	0.50 / 0.77	0.16 / 0.67	0.56 / 0.35	0.26 / 0.32	0.57 / 0.34	-0.06 / 0.23	0.56 / 0.36	-0.06 / 0.31
	20	0.59 / 0.86	0.20 / 0.70	0.66 / 0.34	0.29 / 0.31	0.66 / 0.35	-0.04 / 0.20	0.67 / 0.34	-0.02 / 0.30
	25	0.64 / 0.89	0.21 / 0.70	0.73 / 0.34	0.30 / 0.30	0.73 / 0.35	-0.04 / 0.19	0.74 / 0.33	-0.02 / 0.29
10	10	0.38 / 0.62	0.08 / 0.57	0.50 / 0.34	0.09 / 0.31	0.47 / 0.39	-0.03 / 0.30	0.49 / 0.37	-0.04 / 0.31
	15	0.53 / 0.71	0.17 / 0.65	0.63 / 0.31	0.16 / 0.28	0.62 / 0.37	0.01 / 0.27	0.66 / 0.35	0.00 / 0.28
	20	0.62 / 0.79	0.23 / 0.66	0.73 / 0.30	0.19 / 0.27	0.71 / 0.37	0.03 / 0.26	0.75 / 0.35	0.01 / 0.27
	25	0.69 / 0.81	0.24 / 0.66	0.79 / 0.28	0.20 / 0.26	0.76 / 0.37	0.03 / 0.24	0.79 / 0.35	0.02 / 0.25
20	20	0.63 / 0.77	0.19 / 0.65	0.71 / 0.32	0.12 / 0.29	0.72 / 0.35	0.04 / 0.25	0.72 / 0.30	0.02 / 0.25
	25	0.70 / 0.79	0.21 / 0.65	0.77 / 0.31	0.14 / 0.28	0.78 / 0.35	0.05 / 0.24	0.78 / 0.29	0.02 / 0.23

NOTE: Each cell summarizes serial and cross-sectional correlations of a 1000 simulated, panel data sets of respective size (N,T). For each simulated data set, we estimate a value for ρ and $\frac{N^2+N}{2}$ values for the respective $\sigma_{\varepsilon,ij}$'s. The first number in each cell reports the average value of $\hat{\rho}$ across the respective 10000 simulated data sets. The second number reports the average of the absolute value of the respective estimates of $\sigma_{\varepsilon,ij}$.

[a] Number of cross-sectional units.
[b] Number of time periods.
[c] Indicates the type of actual panel data after which the simulated data are patterned. See Section 2 for category definitions.