DEPARTMENT OF ECONOMICS

COLLEGE OF BUSINESS AND ECONOMICS

UNIVERSITY OF CANTERBURY

CHRISTCHURCH, NEW ZEALAND

A MONTE CARLO EVALUATION OF SOME COMMON PANEL DATA ESTIMATORS WHEN SERIAL CORRELATION AND CROSS-SECTIONAL DEPENDENCE ARE BOTH PRESENT

by W. Robert Reed and Haichun Ye

WORKING PAPER

No. 01/2007

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

WORKING PAPER No. 01/2007

A MONTE CARLO EVALUATION OF SOME COMMON PANEL DATA ESTIMATORS WHEN SERIAL CORRELATION AND CROSS-SECTIONAL DEPENDENCE ARE BOTH PRESENT

by W. Robert Reed and Haichun Ye*

Revised, April 30, 2007

ABSTRACT. This study employs Monte Carlo experiments to evaluate the performances of a number of common panel data estimators when serial correlation and cross-sectional dependence are both present. It focuses on fixed effects models with less than 100 cross-sectional units and between 10 and 25 time periods (such as are commonly employed in empirical growth studies). Estimator performance is compared on two dimensions: (i) root mean square error and (ii) accuracy of estimated confidence intervals. An innovation of our study is that our simulated panel data sets are designed to look like "real-world" panel data. We find large differences in the performances of the respective estimators. Further, estimators that perform well on efficiency grounds may perform poorly when estimating confidence intervals, and vice versa. Our experimental results form the basis for a set of estimator recommendations. These are applied to "out of sample" simulated panel data sets and found to perform well.

Keywords: Panel Data estimation, Monte Carlo analysis, FGLS, PCSE, Groupwise Heteroscedasticity, Serial Correlation, Cross-sectional Dependence, Stata, EViews

JEL Categories: C23, C15

Acknowledgments: This paper has benefited from comments by Bill Greene, Peter Kennedy, Peter Phillips, Gareth Thomas (from EViews), David Drukker (from Stata), and seminar participants at the universities of Auckland, Canterbury, Otago, and Waikato in New Zealand. Culpability for remaining errors is ours alone.

* W. Robert Reed (contact author), Department of Economics, University of Canterbury, Private Bag 4800, Christchurch 8020, New Zealand; email: bobreednz@yahoo.com. Haichun Ye, Department of Economics, University of Oklahoma, Norman, OK 73071, USA.

Acknowledgments: This paper has benefited from comments by Bill Greene, Peter Kennedy, Peter Phillips, Gareth Thomas (from EViews), David Drukker (from Stata), and seminar participants at the universities of Auckland, Canterbury, Otago, and Waikato in New Zealand. Culpability for remaining errors is ours alone.

> "My worry as an econometric theorist is not that there is tension between us (the theorists) and them (the applied economists). On the contrary, such tension can be healthy and inspiring. My worry is rather the lack of tension. There are two camps, a gap between them, and little communication."
>
> -- J. R. Magnus (1999)[1]

I. INTRODUCTION

Panel data can be characterized by complex error structures. The presence of nonspherical errors, if not properly addressed, can generate inefficiency in coefficient estimation and biasedness in the estimation of standard errors. Serial correlation has long been recognized as a potential problem for panel data. Cross-sectional dependence has recently received renewed attention (Driscoll and Kraay, 1998; De Hoyos and Sarafidis, 2006). It is likely that both of these are present in many empirical applications. This is a problem, because most common panel data estimators are unable to simultaneously handle both serial correlation and cross-sectional dependence.

One estimator that can is Parks' feasible GLS estimator (Parks, 1967). However, it can only be implemented when the number of time periods (T) is greater than or equal to the number of cross-sections (N). An additional problem is that Parks' FGLS estimator is known to underestimate standard errors in finite samples, often severely so. Beck and Katz (1995) report that a two-step, modified version of "inefficient" OLS -- known as "panel-corrected standard error" (PCSE) estimation -- performs substantially better than the asymptotically efficient FGLS(Parks) estimator in many circumstances.[2]

---

[1] This quote is taken from Peter Kennedy's A Guide to Econometrics, Fifth Edition (2003, page 405).

[2] In a similar vein, Grubb and Magee (1988) demonstrate that OLS can dominate FGLS even when the data generating process is characterized by serial correlation.

The poor performance of the Parks' FGLS estimator in finite samples is illustrative: It arises because the true error variance-covariance matrix is unknown. Substituting estimates for the elements of the population variance-covariance matrix impairs the performance of FGLS estimation. It opens up the door for asymptotically inefficient estimators to perform better in finite samples.[3]

All of this creates a confusing situation for researchers using panel data in which both serial correlation and cross-sectional dependence may be present. On the one hand, there is a plethora of panel data estimators available from statistical software packages like EViews, LIMDEP, RATS, SAS, Stata, TSP, and others. On the other hand, the finite sample performances of these estimators are not well known. At the end of the day, it is not clear which estimator one should use in a given research situation.

This study attempts to shed some light on this subject. It uses Monte Carlo analysis to evaluate the performances of a number of common panel data estimators when both serial correlation and cross-sectional dependence are present. Given the vast scope of this research area, our study is inevitably narrow in its focus: It works with data sets in which the number of cross-sectional units is less than 100 and the number of time periods range from 10 to 25 -- sizes typical for panel data studies of economic growth across countries and U.S. states. It studies fixed effects models, but not random effects. And it draws its set of panel data estimators from the menu of choices available in Stata and EViews. Estimator performance is compared on two dimensions: (i) root mean square error ("efficiency")[4] and (ii) accuracy of estimated confidence intervals ("coverage").

[3] This is consistent with the "shrinkage principle," well-known in the forecasting literature, that imposing incorrect restrictions on a model can improve forecast performance (Diebold, 2004, page 45).

[4] We follow other Monte Carlo studies in equating efficiency with MSE (cf. Beck and Katz, 1995), but recognize that FGLS is biased in small samples.

A criticism of the Monte Carlo approach is that the population parameter values used in experiments may not represent "real-world" data. This criticism is particularly valid for panel data, where the number of population parameters in the error variance-covariance matrix can be larger than the number of observations. A further complication is that the performance of the estimators can be a function of the distribution of the explanatory variables (see, for example, Peterson, 2007). An innovation of our study is that we attempt to address this concern by creating simulated data environments that look like "real-world" panel data.

We have three main findings. First, the choice of which estimator to use is an important one that can substantially impact one's research findings: There are large differences in the performances of the respective estimators. Second, we find that estimators that perform well on efficiency grounds may perform poorly when estimating confidence intervals, and vice versa. For example, in many settings FGLS(Parks) is the best overall estimator with respect to efficiency, but the worst when it comes to estimating confidence intervals. This means that researchers may have to use one estimator if they want the "best" coefficient estimates, and another if they desire reliable hypothesis testing.

Third, our experiments identify a number of data scenarios where one estimator performs better than the others with respect to (i) efficiency or (ii) coverage. This leads us to make a set of (tentative) recommendations concerning specific panel data estimator(s) one should use in given situations. As a check, we apply these recommendations to "out of sample" simulated panel data. Our recommendations perform well.

The paper proceeds as follows. Section II describes the Monte Carlo experiments, including a description of the specific estimators that we study and the performance measures we use to compare them. Section III characterizes the simulated panel data sets used in our experiments. Section IV analyzes the results of the Monte Carlo experiments and develops specific estimator recommendations. Section V applies these recommendations to new simulated data to see whether the recommendations are valid when applied elsewhere. Section VI concludes.

II. DESCRIPTION OF THE MONTE CARLO EXPERIMENTS

The data generating process. Our simulated data environments are designed to incorporate both serial correlation and cross-sectional dependence. We model the following fixed effects data generating process (DGP):

$$(1)\qquad \begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \\ \vdots \\ \boldsymbol{y_N} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{X_1} \\ \boldsymbol{X_2} \\ \vdots \\ \boldsymbol{X_N} \end{bmatrix} \beta_x + \begin{bmatrix} \boldsymbol{\varepsilon_1} \\ \boldsymbol{\varepsilon_2} \\ \vdots \\ \boldsymbol{\varepsilon_N} \end{bmatrix}, \text{ or } \quad \boldsymbol{y} = \boldsymbol{\beta_0} + \boldsymbol{X}\beta_x + \boldsymbol{\varepsilon},$$

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; $\boldsymbol{X_i}$ is a $T \times 1$ vector of observations of the exogenous explanatory variable; β_i, $i = 1,2,...N$, and β_x are scalars; and $\boldsymbol{\varepsilon}_i$ is a $T \times 1$ vector of error terms, where $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Omega_{NT}})$.

We want a structure for $\boldsymbol{\Omega}_{NT}$ that can simultaneously incorporate serial correlation and cross-sectional dependence in the error term. Accordingly, we adopt a version of Parks' (1967) well-known model. It assumes (i) groupwise heteroscedasticity; (ii) first-order serial correlation; and (iii) time-invariant cross-sectional dependence. We employ the following specification for $\boldsymbol{\Omega}_{NT}$:[5]

[5] In its most general form, the Parks model assumes groupwise, first-order serial correlation. In contrast, our experiments model the DGP with an AR(1) parameter, ρ, that is the same across groups. We do this for two reasons. First, Beck and Katz (1995) recommend that researchers should impose a common AR(1) parameter when

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

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 $\sigma_{\varepsilon,ij} = \dfrac{\sigma_{u,ij}}{1-\rho^2}$.

In order to generate panel data observations using this DGP, we must select values for the population parameters, including the distribution of the explanatory variable, $\boldsymbol{X}$. We would like these to be set at values that typify "real-world" panel data. The challenge in doing this is illustrated by the large number of elements in the error variance-covariance matrix. There are $\left[\dfrac{N(N+1)}{2}+1\right]$ unique parameters in $\boldsymbol{\Omega}_{NT}$. So, for example, when $N = 20$, we must set 211 population values for $\boldsymbol{\Omega}_{NT}$, each of which can take a wide range of values. Unfortunately, theory offers little guidance as to which of these parameters, or which relationships between parameters, are most significant for the performance of the estimators in finite samples.

Our solution to this problem is to estimate a large number of "real-world" panel data sets of various (N,T) sizes. The residuals from these regressions are used to estimate the elements of Parks-type, error variance-covariance matrices. These are then referenced to set the population parameters in the DGP. Population values for the β's and the distribution of $\boldsymbol{X}$ are set using a similar procedure. Details are provided in the Appendix.

---

estimating the Parks model, and we wanted our full Parks model estimators to be correctly specified. Second, having a single AR(1) parameter facilitates characterization and comparison of serial correlation within and across the simulated data sets.

The estimators. There are many estimators available to researchers working with panel data that may have serial correlation and cross-sectional dependence. We choose our estimators from the menu of panel data estimators available in Stata and EViews, though most of these can be found in other software packages. Within these two packages, there are a variety of commands and options available to the user, depending on the specific assumptions he/she makes about heteroscedasticity, serial correlation, and cross-sectional dependence. We focus on four categories of estimators: (i) OLS, (ii) FGLS, (iii) OLS/FGLS with "robust" standard errors, and (iv) the PCSE estimator, a two-step estimator that is neither OLS nor FGLS.[6]

OLS and FGLS estimators employ the following general formulae for $\hat{\boldsymbol{\beta}}$ and $Var\left(\hat{\boldsymbol{\beta}}\right)$:

(3) $$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{y}$$

(4) $$Var\left(\hat{\boldsymbol{\beta}}\right) = \left(\boldsymbol{X}'\hat{\boldsymbol{\Omega}}^{-1}\boldsymbol{X}\right)^{-1},$$

where $\hat{\boldsymbol{\Omega}}$ incorporates implicit assumptions about error heteroscedasticity, serial correlation, and cross-sectional dependence. OLS arises when $\boldsymbol{\Omega}_{NT}$ is assumed to be characterized by homoscedasticity, no serial correlation, and no cross-sectional dependence. At the other end of the spectrum, FGLS(Parks) assumes groupwise heteroscedasticity, first-order serial correlation, and time-invariant cross-sectional dependence.

FGLS(Parks) is not computable when $N > T$, because $\hat{\boldsymbol{\Sigma}}$ is not full rank (cf. Equation [2]). For this and other reasons, our study includes two FGLS estimators that do not incorporate cross-sectional dependence. These assume, respectively, that $\boldsymbol{\Omega}_{NT}$ is characterized by (i) groupwise heteroscedasticity, no serial correlation, and no cross-sectional dependence; and (ii)

[6] We do not consider dynamic panel data models (cf. Roodman, 2006) as these entail additional issues not primarily related to the structure of the error variance-covariance matrix.

groupwise heteroscedasticity and first-order serial correlation, but no cross-sectional dependence.

In contrast, OLS and (partial)[7] FGLS with "robust" standard errors employ the following general formulae:

(5) $$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{y}$$

(6) $$Var\left(\hat{\boldsymbol{\beta}}\right) = \left(\boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{W}^{-1}\hat{\boldsymbol{\Omega}}\boldsymbol{W}^{-1}\boldsymbol{X}\right)\left(\boldsymbol{X}'\boldsymbol{W}^{-1}\boldsymbol{X}\right)^{-1},$$

where $\boldsymbol{W}$ identifies the "weighting matrix" and $\hat{\boldsymbol{\Omega}}$ incorporates assumptions about the estimated error variance-covariance matrix. OLS coefficient estimates are produced when $\boldsymbol{W}$ is the identity matrix. Our study includes OLS estimators with "robust" standard errors where the robustness refers to (i) heteroscedasticity, (ii) heteroscedasticity + serial correlation, and (iii) heteroscedasticity + cross-sectional dependence.

There are many (partial) FGLS options that allow for robust standard errors. Greene (2003, pages 333f.) recommends weighting on groupwise heteroscedasticity. We follow-up Greene's recommendation by estimating three additional (partial) FGLS estimators.[8] Like the OLS estimators, these allow robustness with respect to (i) heteroscedasticity, (ii) heteroscedasticity + serial correlation, and (iii) heteroscedasticity + cross-sectional dependence.[9]

Beck and Katz's (1995) two-step, PCSE estimator uses the formulae,

(7) $$\hat{\boldsymbol{\beta}} = \left(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right)^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{y}}$$

(8) $$Var\left(\hat{\boldsymbol{\beta}}\right) = \left(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right)^{-1}\left(\tilde{\boldsymbol{X}}'\hat{\boldsymbol{\Sigma}}\tilde{\boldsymbol{X}}\right)\left(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right)^{-1},$$

[7] We thank Peter Phillips for recommending the use of "partial FGLS" to distinguish these estimators from conventional FGLS.
[8] Specifically, Greene (2003, page 333) recommends using a robust estimator that incorporates cross-sectional dependence.
[9] Estimators 9 through 11 can be thought of as the (partial) FGLS analogues for the OLS estimators 4, 3, and 2, respectively.

where $\tilde{\boldsymbol{X}}$ and $\tilde{\boldsymbol{y}}$ are the Prais-transformed vectors of the explanatory and dependent variables, and $\hat{\boldsymbol{\Sigma}}$ is the estimate of $\boldsymbol{\Sigma}$ in Equation (2).

TABLE 1 lists the eleven estimators included in this study. Many more could have been chosen from Stata and EViews, and from other software packages. These eleven represent our subjective judgment of the estimators most likely to be chosen by researchers working with panel data, in which serial correlation and cross-sectional dependence may be present.

Most of the estimators in TABLE 1 can be estimated by both Stata and EViews. However, there are slight differences in how the software packages calculate specific estimators.[10] We use Stata's "version" for the first 8 estimators in TABLE 1, and EViews' "version" for the remaining three. Additional details are supplied in the table.

Our analysis takes pains to exactly replicate the output one would obtain using the respective software packages. For example, Stata calculates the common AR(1) parameter, ρ, differently for the **xtgls** and **xtpcse** procedures. For the **xtgls** procedure with option **corr(ar1)**, Stata takes the average of the *N* group-specific, estimated AR(1) parameters, and then truncates the estimated value to lie between -1 and 1. In contrast, for the **xtpcse** procedure with option **corr(ar1)**, Stata first truncates each of the *N* group-specific, estimated AR(1) parameters to lie between -1 and 1, and then takes the average. Another example relates to the construction of confidence intervals. For most procedures, Stata uses *t* critical values to construct confidence intervals. However, both **xtgls** and **xtpcse** calculate confidence intervals using *Z* critical values.[11] Our procedures incorporate all of these details in calculating coefficient estimates and confidence intervals.

[10] For example, Stata uses linear methods for estimating ρ, while EViews employs a nonlinear procedure.

[11] Of particular note is the way that Stata calculates confidence intervals when **cluster( )** is chosen. Let $Var(\hat{\boldsymbol{\beta}})$ be the estimated coefficient covariance matrix unadjusted for degrees of freedom. The cluster option makes the

The experiments. As discussed above, our Monte Carlo experiments set the population parameters in the DGP to values estimated from "real-world" panel data sets. Since our goal is to produce experimental results that are generalizable to actual research situations, it is important that we reference a variety of "real-world" data, and that these data embody a wide range of heteroscedasticity, serial correlation, and cross-sectional dependence behaviors.

We use two sources of "real-world" data for this purpose: annual real per capita Personal Income data (PCPI) from U.S. states, and annual real per capita GDP data across countries. Further, we work with both the level and growth rate of these variables, and with two different residual-producing regression specifications.[12] This yields a total of eight families of "real-world" data. Within each family, we estimate data sets that vary in size from $T = 10$ to $T = 25$; and from $N = 5$ to either $N = 48$ (for the faux U.S. data) or $N = 77$ (for the faux international data). The characteristics of the corresponding simulated data sets are discussed below.

Our primary ("in-sample") Monte Carlo analysis is based on 144 experiments. A single experiment is defined by a unique DGP patterned after a specific-sized (N,T) panel data set from one of the eight families of "real-world" data. Within each experiment, we simulate a thousand panel data sets. For each simulated panel data set, we calculate estimates of β_x and $Var\left(\hat{\beta}_x\right)$ for each of the eleven estimators in TABLE 1. These estimates are then aggregated to produce the estimator-specific performance measures, (i) "*EFFICIENCY*" and (ii) "*COVERAGE*". Thus, one

---

following d.f. adjustment, $Var\left(\hat{\boldsymbol{\beta}}\right)_{CLUSTER} = q_c \, Var\left(\hat{\boldsymbol{\beta}}\right)$, where $q_c = \left(\frac{NT-1}{NT-K}\right)\left(\frac{C}{C-1}\right)$, and NT is the total number of observations, K is the number of estimated coefficients, and C is the number of "clusters" (either N or T in our notation). Further, in calculating confidence intervals (and p-values), Stata uses the t critical value with $C - 1$ degrees of freedom. Contrast this with the conventional approach of using $NT - K$ degrees of freedom. This difference can have a substantial impact on the width of the confidence interval. Note that EViews follows the latter when estimating its analog of **cluster( )** standard errors.

[12] The main difference between the two residual-producing regression specifications is that only one included time period fixed effects (both included group fixed effects). The inclusion of time period fixed effects substantially reduces (but does not eliminate) cross-sectional dependence (cf. Roodman, 2006).

EFFICIENCY value and one *COVERAGE* value is produced for each estimator in every experiment.

Measures of estimator performance. *EFFICIENCY* measures the ratio of the (square root) of mean square errors for the respective estimator and OLS. It is defined by

$$(9) \qquad EFFICIENCY = 100 \cdot \frac{\sqrt{\sum_{r=1}^{1000}\left(\hat{\beta}_{ESTIMATOR}^{(r)} - \beta_x\right)^2}}{\sqrt{\sum_{r=1}^{1000}\left(\hat{\beta}_{OLS}^{(r)} - \beta_x\right)^2}},$$

where β_x is the population value, and $\hat{\beta}_{ESTIMATOR}^{(r)}$ and $\hat{\beta}_{OLS}^{(r)}$ are the estimated values of β_x in a given replication using the respective estimator and OLS. A value less than *100* is interpreted to mean that the estimator is more efficient than OLS in that experiment.

COVERAGE calculates the percent of the one thousand, estimated 95% confidence intervals that include the true value of β_x (i.e., the coverage rate). A *COVERAGE* value less than *95* indicates that the estimated confidence intervals are too narrow on average. This would imply over-rejection of the null hypothesis. As stated above, the estimated confidence intervals for each of the eleven estimators are constructed to exactly match Stata and EViews output.

Each experiment produces one *EFFICIENCY* and one *COVERAGE* value for each of the eleven estimators. There are a total of *144* experiments. The performance results from these experiments are analyzed to identify relationships with observable data characteristics. Since the FGLS(Parks) estimator cannot be computed when $N > T$, we divide our experiments into two groups: $N \le T$ (*80* experiments) and $N > T$ (*64* experiments).

III. DESCRIPTION OF SIMULATED DATA USED IN OUR EXPERIMENTS

This section describes three statistics for measuring the degrees of groupwise heteroscedasticity, serial correlation, and cross-sectional dependence in the simulated panel data sets. There are two reasons for doing this. First, this is the first step towards linking observable data characteristics and estimator performance. Note that the population parameters of the underlying DGP are not very useful for this purpose, since they are not directly seen by the econometrician.

Second, the ultimate aim of this paper is to develop a set of recommendations regarding estimator selection as a function of observable data characteristics. If our simulated panel data sets display only a narrow range of error behaviours, we can have little confidence that our recommendations will generalize to other panel data sets. A necessary but not sufficient condition for us to have confidence in the wider applicability of our results is that our simulated data sets display a wide range of heteroscedasticity, serial correlation, and cross-sectional dependence behaviours.

"*HETCOEF*" is a measure of the degree of groupwise heteroscedasticity present in a given data set. It is computed by estimating group-specific standard deviations that are then sorted in ascending order. The "heteroscedasticity coefficient" is calculated as the ratio of the standard deviations associated with the 75th and 25th percentile ranking, respectively. For example, when $N=5$, the associated groupwise standard deviations, ranked in ascending order, are $\hat{\sigma}_1$, $\hat{\sigma}_2$, …, $\hat{\sigma}_5$; and *HETCOEF* is calculated by $\frac{\hat{\sigma}_4}{\hat{\sigma}_2}$. In many cases, linear interpolation is employed. For example, when $N=10$, *HETCOEF* is calculated by $\frac{\hat{\sigma}_7 + 0.75\cdot\left(\hat{\sigma}_8 - \hat{\sigma}_7\right)}{\hat{\sigma}_3 + 0.25\cdot\left(\hat{\sigma}_4 - \hat{\sigma}_3\right)}$. A value greater than or equal to one is guaranteed given the sorting of the group-specific standard deviations.

"*RHOHAT*" estimates the value of the common AR(1) parameter using observations from a given data set. It is calculated using the formula suggested by Greene (2003, page 326):

$$RHOHAT = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} e_{it}e_{i,t-1}}{\sum_{i=1}^{N}\sum_{t=2}^{T} e_{it}^{2}}$$, where the e_{it} are residuals from an OLS regression.

"*CSCORR*" is a measure of the degree of cross-sectional dependence present in a given panel data set. Define $\left| r_{ij} \right|$ as the absolute value of the sample correlation coefficient between the residuals from groups *i* and *j*. *CSCORR* calculates the average of this value across all (*i,j*) pairs.

Each of these measures describes a single panel data set. These are averaged over all one thousand data sets to produce a summary value for the given experiment. TABLE 2 reports these for the "in-sample" experiments in this study. As discussed above, FGLS(Parks) cannot be estimated when $N > T$. Accordingly, the table divides the experiments into those where $N \leq T$ (*80* experiments) and $N > T$ (*64* experiments). We first consider those experiments where $N \leq T$.

Collectively, the simulated data sets display a wide range of heteroscedasticity, serial correlation, and cross-sectional dependence behaviours. *HETCOEF*, the measure of groupwise heteroscedasticity, ranges from a minimum of *1.19* to a maximum of *2.31*.[13] The mean value of *HETCOEF* is *1.57*, which implies that the 75th percent-ranked, group-specific standard deviation is approximately 57% larger than its 25th percent-ranked counterpart. Overall, the null hypothesis of no groupwise heteroscedasticity is rejected in approximately three-fourths of the data sets.

[13] As these values are experiment-specific, the associated range on the level of individual data sets is even larger.

RHOHAT, the measure of serial correlation behaviour, ranges from a minimum of *-0.09* to a maximum value of *0.79*. The average *RHOHAT* value across all data sets where $N \leq T$ is *0.36*, and the null hypothesis of no serial correlation is rejected in approximately two thirds of these data sets.

We also find substantial cross-sectional dependence in our data sets. The minimum *CSCORR* value in the $N \leq T$ data sets is *0.19*, the maximum is *0.89*, and the mean is *0.41*. These numbers represent the average, absolute value of the correlations across all possible pairs of groups. The null hypothesis of no cross-sectional dependence is rejected in almost 90 percent of all data sets.

The characteristics of the $N > T$ data sets are generally comparable, aside from the obvious difference that they include a larger number of cross-sectional units. The difference in rejection rates is somewhat misleading, since the power of these tests is substantially influenced by the number of observations in the data set.

This statistical description provides evidence that our experimental data sets embody a wide range of heteroscedasticity, serial correlation, and cross-sectional dependence behaviours.

IV. ANALYSIS OF MONTE CARLO EXPERIMENTS

We first analyze estimator *EFFICIENCY* and then address *COVERAGE*. TABLE 3 enables a comparison of estimator efficiency. The top part of the table reports the results of averaging the *EFFICIENCY* values over the respective sets of experiments for each of the eleven estimators. Values less than *100* indicate that the respective estimator is, on average, more efficient than OLS.

Estimators 5, 9, 10, and 11 all use the same weighting matrix (based on groupwise heteroscedasticity) and hence all produce the same $\hat{\beta}_x$ values. As a result, they have identical

EFFICIENCY values and are grouped together in the table. Estimators 1-4 all produce identical, OLS coefficients. By construction, the corresponding *EFFICIENCY* values equal *100*, and thus are not reported in the table.

We first focus on those experiments where $N \leq T$ (cf. Column [1]). It is clear that Estimator 7, the FGLS(Parks) estimator, is substantially more efficient than the other eleven estimators. The latter provide efficiency gains that are, at best, only slightly better than OLS. In contrast, the FGLS(Parks) estimator is, on average, approximately 25% more efficient than OLS, as quantified by our *EFFICIENCY* measure.

These average figures can mask substantial variation across experiments. Accordingly, we also calculate the percent of experiments where the respective estimator does better than OLS. These results are reported in the bottom part of TABLE 3. All of the non-OLS estimators are more efficient than OLS in at least half of the experiments. However, only Estimator 7, the FGLS(Parks) estimator, consistently outperforms OLS: FGLS(Parks) is more efficient in over 95 percent of the experiments.

The next step in our analysis consists of relating observable characteristics of the panel data sets to the efficiency performance of FGLS(Parks). We hope that this will lead us to identify specific data situations where FGLS(Parks) can be expected to produce the greatest gains over OLS.

The first column of TABLE 4 reports the results of regressing the *EFFICIENCY* of the FGLS(Parks) estimator on the data characteristics *HETCOEF*, *RHOHAT*, *CSCORR*, *N*, and *T*. The observations are drawn from the 80 experiments where $N \leq T$. Negative coefficient estimates are interpreted as indicating greater efficiency. Heteroscedasticity robust standard errors are used to calculate t-statistics, and the respective p-values are reported below the estimated coefficients.

The following data characteristics are associated with significant efficiency gains for FGLS(Parks): lower (groupwise) heteroscedasticity, lower cross-sectional dependence, a smaller number of cross-sectional units, and a larger number of time periods. The contribution of serial correlation is insignificant.

The first two results, while consistent with previous research[14], may be surprising given that it is the presence of groupwise heteroscedasticity and cross-sectional dependence (along with serial correlation) that are ultimately responsible for the greater efficiency of FGLS. We suspect these results are finite sample phenomena driven by imprecise estimation of the error variance-covariance matrix: The larger the values of the underlying variance-covariance elements, the greater the potential for mis-estimation of these values to impair the efficiency of FGLS relative to OLS.

Evidence in favour of this interpretation is provided by the estimated coefficient for T. While increases in both N and T grow the number of observations in the data set, an increase in N expands the number of unique parameters in the error variance-covariance matrix, while an increase in T does not.[15] Thus, an increase in T should improve the precision of the variance-covariance estimates, yielding greater efficiency gains for FGLS(Parks). The estimated negative coefficient for T is consistent with this explanation.

The asymptotic efficiency result for the FGLS(Parks) estimator requires large T, fixed N. The intuition relates to the fact that as T/N gets larger, there is more data to estimate each

---

[14] Driscoll and Fray (1998) find that coverage rates for the SUR estimator decline as the degree of cross-sectional dependence increases (Table 1, page 554).

[15] When N increases by 1, the number of unique parameters in the Parks error variance-covariance matrix increases by $N+1$, while the number of observations increases by T, and recall that $T \geq N$.

parameter in the error variance-covariance matrix. This is supported by the empirical results of TABLE 4. When the variable (T/N) is substituted for N and T, the individual size variables become statistically superfluous.[16]

TABLE 4 makes clear that there are many data characteristics that contribute to the efficiency of the FGLS(Parks) estimator. In an ideal world, there would be one characteristic that could alert researchers when this estimator was likely to be most effective. After some experimentation, we found that (T/N) can serve this role.

FIGURE 1 plots the *EFFICIENCY* values for each of the respective estimators, where the individual experiments/observations are sorted in ascending order of (T/N). The vertical axis reports the estimator's *EFFICIENCY* value, and the horizontal axis indicates the ratio of time periods to cross-sectional units (T/N).

Due to the construction of our experiments, the latter values increase in discrete jumps, indicated by vertical grid lines in the graph. For example, all the observations before the first dashed line represent experiments where $T/N = 1.00$. The observations before the second dashed line and including the first dashed line represent experiments where $T/N = 1.25$; and so on.

The thick, solid line plots the *EFFICIENCY* values of the FGLS(Parks) estimator. While there is substantial variation, a clear negative trend is apparent, indicating greater relative efficiency in the FGLS(Parks) estimator as T/N increases. This plot indicates the following: When $T/N \geq 1.50$, there is separation between FGLS(Parks) and all other estimators, including OLS. It is consistently superior. When $1 \leq T/N < 1.50$, no estimator dominates. In this case,

[16] The *p*-value for the coefficient of N when it is included as an additional variable in a specification that already includes (T/N) is *0.344*. The *p*-value for the coefficient of T when it is added to a specification that already includes (T/N) is *0.192*.

FGLS(Parks) offers only marginal improvement on OLS, and is sometimes inferior. Other estimators tend to outperform both FGLS(Parks) and OLS in some instances, but do worse in others. This leads to the first recommendation:

> ***<u>Recommendation #1</u>: When the primary concern is efficiency and $T/N \geq 1.50$, use FGLS(Parks).***

Unfortunately, our experimental results are not able to identify a dominant estimator when $1 \leq T/N < 1.50$.

We continue our study of estimator *EFFICIENCY*, but now move to the case where $N > T$. With FGLS(Parks) no longer in the choice set, the question is whether another estimator can be found that is consistently more efficient than OLS for some identifiable data situations. The second column of TABLE 3 compares average *EFFICIENCY* (top part of table) and the percent of experiments where the respective estimator is more efficient than OLS (bottom part of table).

Estimator 5/9/10/11 and Estimator 6, which are (partial) FGLS estimators that weight on (i) groupwise heteroscedasticity and (ii) groupwise heteroscedasticity + serial correlation, perform quite similarly. Both offer some efficiency advantages relative to OLS, though the gains are not as substantial as in the previous case with FGLS(Parks).

As before, we attempt to relate estimator performance to observable data characteristics.[17] The second column of TABLE 4 shows that greater relative efficiency for FGLS(Groupwise Heteroscedasticity) is significantly associated with greater heteroscedasticity, greater cross-sectional dependence, and data sets with a larger number of cross-sections and a larger number of time periods. Serial correlation is insignificantly associated with the efficiency

[17] Asymptotic theory is not helpful in identifying key determinants in this case because the respective estimators incorrectly/incompletely model the error variance-covariance matrix.

of this estimator. With respect to the FGLS(Groupwise Heteroscedasticity + Serial Correlation) estimator, Column (3) shows its relative efficiency increases significantly with greater heteroscedasticity. The estimated contributions of the other data characteristics are statistically insignificant.

Once again we are faced with a situation where there are many panel data characteristics that contribute to estimator efficiency. Our hope is that one of these characteristics will either be sufficiently dominant, or sufficiently correlated with the other characteristics, that it can serve as a guide for selecting a "best" estimator.

FIGURE 2 reports the fruits of our experimentation. In this figure, the observations/experiments are sorted in ascending order of *HETCOEF*, so that groupwise heteroscedasticity increases from left to right. This follows up the observation from TABLE 4 that the relative efficiencies of both FGLS(Groupwise Heteroscedasticity) and FGLS(Groupwise Heteroscedasticity + Serial Correlation) increase as groupwise heteroscedasticity becomes more pronounced.

The solid vertical line in the figure splits the observations into two groups: those with *HETCOEF* values less than *1.67*, and those with *HETCOEF* values larger than *1.67*. This value appears to represent a threshold affecting the relative performance of these two estimators: For large values of groupwise heteroscedasticity (*HETCOEF > 1.67*), both FGLS(Groupwise Heteroscedasticity) and FGLS(Groupwise Heteroscedasticity + Serial Correlation) provide consistent efficiency gains relative to the other estimators. In contrast, no one estimator is distinctly preferred when *HETCOEF* falls below this value. This motivates the next recommendation:

***Recommendation #2*: When the primary concern is efficiency, $N > T$, and $HETCOEF > 1.67$, use either FGLS(Groupwise Heteroscedasticity) or FGLS(Groupwise Heteroscedasticity + Serial Correlation).**

Unfortunately, when $N > T$ and $HETCOEF < 1.67$, no estimator appears to consistently dominate the others.

Next up is a comparison of estimator performance with respect to accuracy in estimating confidence intervals. We continue to separate the cases where $N \le T$ and $N > T$. *COVERAGE* measures the percent of estimated 95% confidence intervals that contain the true value of β_x. Columns (1) and (3) of TABLE 5 report the average values for this measure for each estimator over the respective sets of experiments.

In a few of the individual experiments, *COVERAGE* is larger than 95 for some estimators. This can cause the average *COVERAGE* value to be misleading. For example, combining the values *90* and *100* produces an average value of *95*, suggesting that the respective estimator is highly accurate in its estimation of confidence intervals. To address this problem, we also calculate the absolute value of the difference between *95* and *COVERAGE* ($|95 - COVERAGE|$) for each experiment and estimator. The average value of this alternative measure is reported in Columns (2) and (4) of TABLE 5. In fact, the two measures are very similar, as can be confirmed by noting that the sum of the two average values is close to *95* for all estimators.

As a group, the eleven estimators do a poor job of estimating confidence intervals. Further, there are big differences between estimators. The most efficient estimator when $N \le T$ -- the FGLS(Parks) estimator -- is the worst estimator when it comes to estimating confidence intervals. On average, only *43.3* percent of the 95% confidence intervals estimated with this estimator contain the population value. This accords with similar findings reported by Beck and

Katz (1995). The estimator that comes closest to producing accurate confidence intervals is Estimator 8. This is, in fact, the PCSE estimator promoted by Beck and Katz (1995) as an alternative to FGLS(Parks).

We proceed with what is by now a familiar procedure: We regress the relevant performance measure (in this case, $|95 - COVERAGE|$) for the best estimator (Estimator 8 when $N \leq T$) on observable characteristics of the panel data sets. The results are reported in the first column of TABLE 6. A negative coefficient indicates that an increase in the respective characteristic is associated with more accurate confidence intervals for that estimator. It is apparent that several of the characteristics are significantly associated with this performance measure. As before, we find that plotting estimator performance against individual data characteristics identifies relationships that can lead to estimator recommendations.

FIGURE 3 plots observations of $|95 - COVERAGE|$ for each of the estimators, where the experiments are sorted in ascending order of *RHOHAT*. Panel (A) represents the case where $N \leq T$. All eleven estimators are reported in the figure, which results in an informative, albeit messy, graph. For reasons discussed below, we highlight Estimator 4 (OLS[Heteroscedasticity + Cross-sectional Dependence Robust]) and Estimator 8 (PCSE[Parks]).[18]

We observe that many of the estimators worsen in their estimation of confidence intervals once *RHOHAT* exceeds *0.30*. Interestingly, this same pattern is observed in the lower panel of FIGURE 3, which plots the experiments where $N > T$. Panels (A) and (B) make clear that none of the estimators, including the PCSE estimator, produce consistently accurate confidence intervals when $RHOHAT > 0.30$.[19]

[18] While it is not distinguished in the legend, FGLS(Parks) is identifiable by its wild swings and extreme values.
[19] As in the previous case, we find no evidence that this threshold value is significantly affected by either *N* or *T*.

In response, we concentrate our search for a "best" estimator on those cases where $RHOHAT < 0.30$. TABLE 7 repeats the analysis of TABLE 5 except that it only includes observations where the estimated ρ value is less than *0.30*. Several estimators now demonstrate reasonable performance: However, Estimator 4 (OLS[Heteroscedasticity and Cross-sectional Dependence Robust]) and Estimator 8 (the PCSE estimator) demonstrate the best overall performance, considered over both $N \le T$ and $N > T$. This leads to our final recommendation:

> ***<u>Recommendation #3</u>*: When the primary concern is constructing accurate confidence intervals and $RHOHAT < 0.30$, we recommend either Beck and Katz's (1995) PCSE estimator or the OLS(Heteroscedasticity + Cross-sectional Dependence Robust) estimator.**

When $RHOHAT > 0.30$, our results suggest that no estimator is consistently reliable, and estimated confidence intervals should not be trusted.

V. AN "OUT OF SAMPLE" CHECK

In an ideal world, we would be able to analytically derive the finite sample properties of our panel data estimators. Unfortunately, this is not the situation that we face. As a result, we have turned to Monte Carlo experimentation to establish performance patterns that could serve as the basis for estimator recommendations. The concern is that these recommendations will only be valid for the specific, simulated panel data sets on which they are based.

To address this concern, we simulate additional panel data sets patterned after an entirely different sort of "real world" data: tax burden data from U.S. states (both levels and growth rates), and government consumption share data across countries (levels only). The advantage of turning to dissimilar data is that it can introduce unforeseen data qualities that affect estimator performance. A total of *52* additional experiments are conducted: *30* involving panel data sets where $N \le T$, and *22* where $N > T$. If the results from our Monte Carlo experiments are generalizable, the previous recommendations should perform well in these additional data settings.

Recommendation #1 states that when $T/N \geq 1.50$ and efficiency is the primary concern, researchers should use FGLS(Parks). FIGURE 4 replicates the analysis of FIGURE 1, using observations from the additional, "out of sample" experiments. Despite using altogether different, simulated panel data sets, the two figures are quite similar. FGLS(Parks) consistently outperforms all other estimators when $T/N \geq 1.50$ for these "out of sample" data sets. The validity of Recommendation #1 is upheld in these additional experiments.

Recommendation #2 states that when the primary concern is efficiency, $N > T$, and the value of $HETCOEF > 1.67$, researchers should use either FGLS(Groupwise Heteroscedasticity) or FGLS(Groupwise Heteroscedasticity + Serial Correlation). FIGURE 5 puts this recommendation to the test by replicating the analysis of FIGURE 2 with the new data sets. Recommendation #2 is likewise confirmed as good advice: Both estimators consistently dominate the others when *HETCOEF* takes values larger than *1.67*.

The last recommendation addresses choice of estimator when the primary concern is accurate confidence intervals: FIGURE 6 applies the analysis of FIGURE 3 to the "out of sample" panel data sets. The superior performances of the OLS(Heteroscedasticity + Cross-sectional Dependence Robust) estimator and the PCSE estimator are evident in both panels when $RHOHAT < 0.30$. This provides confirmation that Recommendation #3 is also valid when applied to additional data sets.

These "out of sample" experiments provide some evidence that the recommendations based on the original set of Monte Carlo experiments can be generalized to other data sets. However, additional testing is warranted.

VI. CONCLUSION

A researcher is working with panel data that may contain both serial correlation and cross-sectional dependence. Statistical software packages such as EViews, LIMDEP, RATS, SAS, Stata, TSP, and others offer many estimator choices. Which one(s) should he/she use? At the present time, there is no definitive answer to this question. The finite sample properties of these estimators are analytically indeterminate. And while their asymptotic properties can be derived, studies such as Beck and Katz (1995) have demonstrated that these are unreliable predictors of actual estimator performance. Given this state of affairs, Monte Carlo studies offer the most promising way forward.

This study uses Monte Carlo experimentation to study the performance of a number of common panel data estimators. We focus on data scenarios where the number of cross-sectional units is less than 100 and the number of time periods range from 10 to 25 -- sizes typical for panel data studies of economic growth across countries and U.S. states. The experiments analyze a linear model with fixed effects and an error structure that allows both serial correlation and cross-sectional dependence. An innovation of our study is that it constructs simulated panel data sets that are patterned after "real world" data.

Our Monte Carlo experiments uncover large differences in how the various estimators perform -- differences that could substantially affect the results of empirical research. This highlights the importance of choosing a good estimator. Further, we find that estimators that perform well on efficiency grounds may perform poorly when estimating confidence intervals, and vice versa. A good example of this is provided by the FGLS(Parks) estimator, which is asymptotically efficient given our DGP. FGLS(Parks) was the best overall performer on efficiency grounds but the worst when it came to estimating confidence intervals. The lesson here is that researchers may have to use one estimator if they want the "best" coefficient estimates, and another if they desire reliable hypothesis testing.

Another interesting finding is that testing for the presence of nonspherical errors, such as heteroscedasticity, serial correlation, and cross-sectional dependence, is only of limited value in choosing the appropriate estimator. Even when these behaviours are present, it does not follow that the associated estimator is the best one to use. For example, OLS(Heteroscedasticity + Cross-sectional Dependence Robust), which ignores serial correlation, does a much better job of estimating confidence intervals than FGLS(Parks).

Our Monte Carlo experiments identify three panel data characteristics that are especially important for estimator performance. These are: (i) the ratio of the number of time periods to number of cross-sectional units (T/N), (ii) the degree of groupwise heteroscedasticity, as measured by a "heteroscedasticity coefficient" (*HETCOEF*); and (iii) the degree of serial correlation, as measured by an estimate of the AR(1) parameter (*RHOHAT*). The last two measures are constructed in a straightforward fashion from OLS residuals.

The results of our Monte Carlo experiments suggest the following three recommendations for researchers working with balanced panel data characterized by Parks-style heteroscedasticity, serial correlation, and cross-sectional dependence:

1. When the primary concern is efficiency and $T/N \geq 1.50$, use FGLS(Parks)

2. When the primary concern is efficiency, $N > T$, and $HETCOEF > 1.67$, use either FGLS(Groupwise Heteroscedasticity) or FGLS(Groupwise Heteroscedasticity + Serial Correlation)

3. When the primary concern is constructing accurate confidence intervals and $RHOHAT < 0.30$, use either Beck and Katz's (1995) PCSE estimator or OLS(Heteroscedasticity + Cross-sectional Dependence Robust).

These recommendations do not cover all possible data scenarios. Unfortunately, our Monte Carlo experiments were not able to identify dominant estimators in these other cases.

It is worth stating the obvious that these recommendations are not based on theory, but on empirical patterns observed from Monte Carlo experiments. As such, there are grounds for scepticism that these results apply to anything other than the artificial data from which they were derived. To address this concern, we apply these recommendations to additional simulated panel data sets, patterned after altogether different “real world” data. We find that the three recommendations continue to be valid when applied to these “out of sample” panel data sets.

In the end, only additional testing will determine whether these recommendations are robust for applications to other data. It is hoped that this study stimulates further research along these lines.

REFERENCES

Beck, Nathaniel and Jonathan N. Katz. "What To Do (and Not To Do) With Time Series Cross-Section Data." *American Political Science Review* Vol. 89, No. 3 (1995): 634-647.

De Hoyos, Rafael E. and Vasilis Sarafidid. "Testing for Cross-sectional Dependence in Panel Data Models." *Stata Journal* Vol. 6, No. 4 (2006): 482-496.

Diebold, Francis X. *Elements of Forecasting*, Third Edition. Mason, Ohio: South-Western, 2004.

Driscoll, John C. and Aart C. Kraay. "Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data." *The Review of Economics and Statistics* Vol. 80, No. 4 (1998): 549-560.

Greene, William H. *Econometric Analysis*, Fifth Edition. Upper Saddle River, NJ: Prentic Hall, 2003.

Grubb, D. and L. Magee. "A Variance Comparison of OLS and Feasible GLS Estimators." *Econometric Theory*, Vol. 4, No. 2 (1988): 329-335.

Kennedy, Peter. *A Guide to Econometrics*, Fifth Edition. Malden, MA: Blackwell Publishing Ltd., 2003.

Parks, Richard. "Efficient Estimation of a System of Regression Equations When Disturbances Are Both Serially and Contemporaneously Correlated." *Journal of the American Statistical Association* Vol. 62 (1967): 500-509.

Peterson, Mitchell A. "Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches." Mimeo, Northwestern University, January 2007.

Roodman, David. "How to Do xtabond2: An Introduction to 'Difference' and 'System' GMM in Stata." Mimeo, Center for Global Development, Working Paper Number 103, December 2006.

Wasylenko, Michael. "Taxation and Economic Development: The State of the Economic Literature." *New England Economic Review* (March/April 1997): 37-52.

Wooldridge, Jeffrey M. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press, 2002.

TABLE 1
List and Description of Panel Data Estimators To Be Studied

Estimator	*Name*	*Package*	*Command*
1	OLS	Stata	command = **xtreg**
2	OLS(Heteroscedasticity Robust)	Stata	command = **xtreg** options = **robust**
3	OLS(Heteroscedasticity + Serial Correlation Robust)	Stata	command = **xtreg** options = **cluster(**name of cross-sectional variable)
4	OLS(Heteroscedasticity + Cross-sectional Dependence Robust)	Stata	command = **xtreg** options = **cluster(**name of time period variable)
5	FGLS(Groupwise Heteroscedasticity)	Stata	command = **xtgls** options = **corr(independent) panels(heteroscedastic)**
6	FGLS(Groupwise Heteroscedasticity + Serial Correlation)	Stata	command = **xtgls** options = **corr(ar1) panels(heteroscedastic)**
7	FGLS(Parks)	Stata	command = **xtgls** options = **corr(ar1) panels(correlated)**
8	PCSE(Parks)	Stata	command = **xtpcse** options = **corr(ar1)**
9	FGLS(Weights =Groupwise Heteroscedasticity; Covariance = Heteroscedasticity + Cross-sectional Dependence Robust)	EViews	GLS Weights = **Cross-section weights** Coef covariance method = **White cross-section**
10	FGLS(Weights =Groupwise Heteroscedasticity; Covariance = Heteroscedasticity + Serial Correlation Robust)	EViews	GLS Weights = **Cross-section weights** Coef covariance method = **White period**
11	GLS(Weights =Groupwise Heteroscedasticity; Covariance = Heteroscedasticity Robust)	EViews	GLS Weights = **Cross-section weights** Coef covariance method = **White (diagonal)**

TABLE 2
Description of Simulated Data Sets Used in the Experiments

	Groupwise Heteroscedasticity (Measure = *HETCOEF*)	Serial Correlation (Measure = *RHOHAT*)	Cross-sectional Dependence (Measure = *CSCORR*)	Number of Cross-sections (N)	Number of Time Periods (T)
			(A) Experiments Where $N \leq T$		
Minimum	1.19	-0.09	0.19	5	10
Maximum	2.31	0.79	0.89	20	25
Mean	1.57	0.36	0.41	10	19
H_0:	No Groupwise Heteroscedasticity	No Serial Correlation	No Cross-sectional Dependence	----	----
Rejection Rate of H_0	72.9%	66.7%	87.7%	----	----
			(B) Experiments Where $N > T$		
Minimum	1.25	-0.05	0.22	20	10
Maximum	2.15	0.80	0.78	77	25
Mean	1.77	0.33	0.37	49	16
H_0:	No Groupwise Heteroscedasticity	No Serial Correlation	No Cross-sectional Dependence	----	----
Rejection Rate of H_0	96.6%	63.5%	99.3%	----	----

NOTES: The measures *HETCOEF*, *RHOHAT*, and *CSCORR* are described in Section III of the text. To test the null hypothesis of no groupwise heteroscedasticity, we apply the test described on the bottom of page 328 in Greene (2003). To test the null hypothesis of no serial correlation, we apply the test associated with Equation (7.76) in Wooldridge (2002, p. 176). To test the hypothesis of no cross-sectional dependence, we use Equation (13-68) on page 327 in Greene (2003).

TABLE 3
Comparison of Estimator *EFFICIENCY*

	DATA SETS	
	$N \leq T$	$N > T$
(A) Average *EFFICIENCY*		
Estimator 5/9/10/11	95.2	82.9
Estimator 6	95.1	83.1
Estimator 7	73.9	----
Estimator 8	100.8	101.0
(B) Percent of Experiments Where Estimator is More Efficient Than OLS		
Estimator 5/9/10/11	58.8	84.4
Estimator 6	71.3	79.7
Estimator 7	96.3	----
Estimator 8	62.5	51.6

NOTES: *EFFICIENCY* is defined in Equation (5) in the text. There are *80* experiments for which the respective panel data sets have sizes such that $N \leq T$, and *64* experiments where $N > T$.

TABLE 4
Determinants of *EFFICIENCY* for Estimator 7, and Estimators 5/9/10/11 and 6

VARIABLE	Data Sets (Estimator)		
	N* ≤ *T (Estimator 7)	***N* > *T*** (Estimator 5/9/10/11)	***N* > *T*** (Estimator 6)
Dependent Variable = *EFFICIENCY*			
Constant	20.746 *(0.232)*	247.72 *(0.000)*	169.52 *(0.000)*
HETCOEF	22.409 *(0.009)*	-68.906 *(0.000)*	-50.636 *(0.001)*
RHOHAT	-0.135 *(0.983)*	4.356 *(0.154)*	-9.687 *(0.112)*
CSCORR	76.929 *(0.000)*	-64.203 *(0.000)*	37.560 *(0.371)*
N	2.511 *(0.000)*	-0.117 *(0.017)*	-0.104 *(0.192)*
T	-2.080 *(0.000)*	-0.906 *(0.000)*	-0.148 *(0.782)*
R-squared	0.646	0.823	0.646
Mean of Dependent Variable	73.86	82.86	83.13
Number of Observations	80	64	64

NOTE: Coefficient estimates are derived from OLS regression with White standard errors. The coefficient *p*-values are reported in parenthesis below the respective estimates. Estimator 7 is the FGLS(Parks) estimator. Estimators 5/9/10/11 all produce the same coefficient estimates, equivalent to FGLS(Groupwise Heteroscedasticity), and hence are grouped together. Estimator 6 is the FGLS(Groupwise Heteroscedasticity + Serial Correlation) estimator.

TABLE 5
Comparison of Accuracy of Confidence Intervals Across Estimators

	$N \leq T$		$N > T$	
	COVERAGE	**Absolute Value of (*95-COVERAGE*) Over All Experiments**	***COVERAGE***	**Absolute Value of (*95-COVERAGE*) Over All Experiments**
Estimator 1	73.6	21.9	74.2	21.9
Estimator 2	73.7	21.8	77.9	18.8
Estimator 3	83.5	11.6	91.8	3.9
Estimator 4	72.7	22.5	74.0	21.3
Estimator 5	69.8	25.6	72.6	22.9
Estimator 6	86.4	9.3	88.8	7.2
Estimator 7	43.3	51.7	----	----
Estimator 8	87.8	7.2	88.1	6.9
Estimator 9	66.1	28.9	65.4	29.6
Estimator 10	68.1	26.9	80.1	14.9
Estimator 11	69.5	25.9	72.4	23.2

NOTES: *COVERAGE* is the percent of estimated 95% confidence intervals that contain the population value. Further detail is provided in Section II of the text.

TABLE 6
Determinants of Accuracy of Confidence Intervals for Estimator 8

VARIABLE	Data Sets (Estimator) $N \leq T$ (Estimator 8)	$N > T$ (Estimator 8)
Dependent Variable = $\lvert 95 - COVERAGE \rvert$		
Constant	6.5701 *(0.004)*	-11.712 *(0.081)*
HETCOEF	0.8197 *(0.392)*	4.5458 *(0.038)*
RHOHAT	12.796 *(0.000)*	11.7680 *(0.000)*
CSCORR	8.1917 *(0.000)*	28.358 *(0.000)*
N	-0.1456 *(0.002)*	-0.0067 *(0.706)*
T	-0.3818 *(0.000)*	-0.2158 *(0.026)*
R-squared	0.796	0.717
Mean of Dependent Variable	7.245	6.908
Number of Observations	80	64

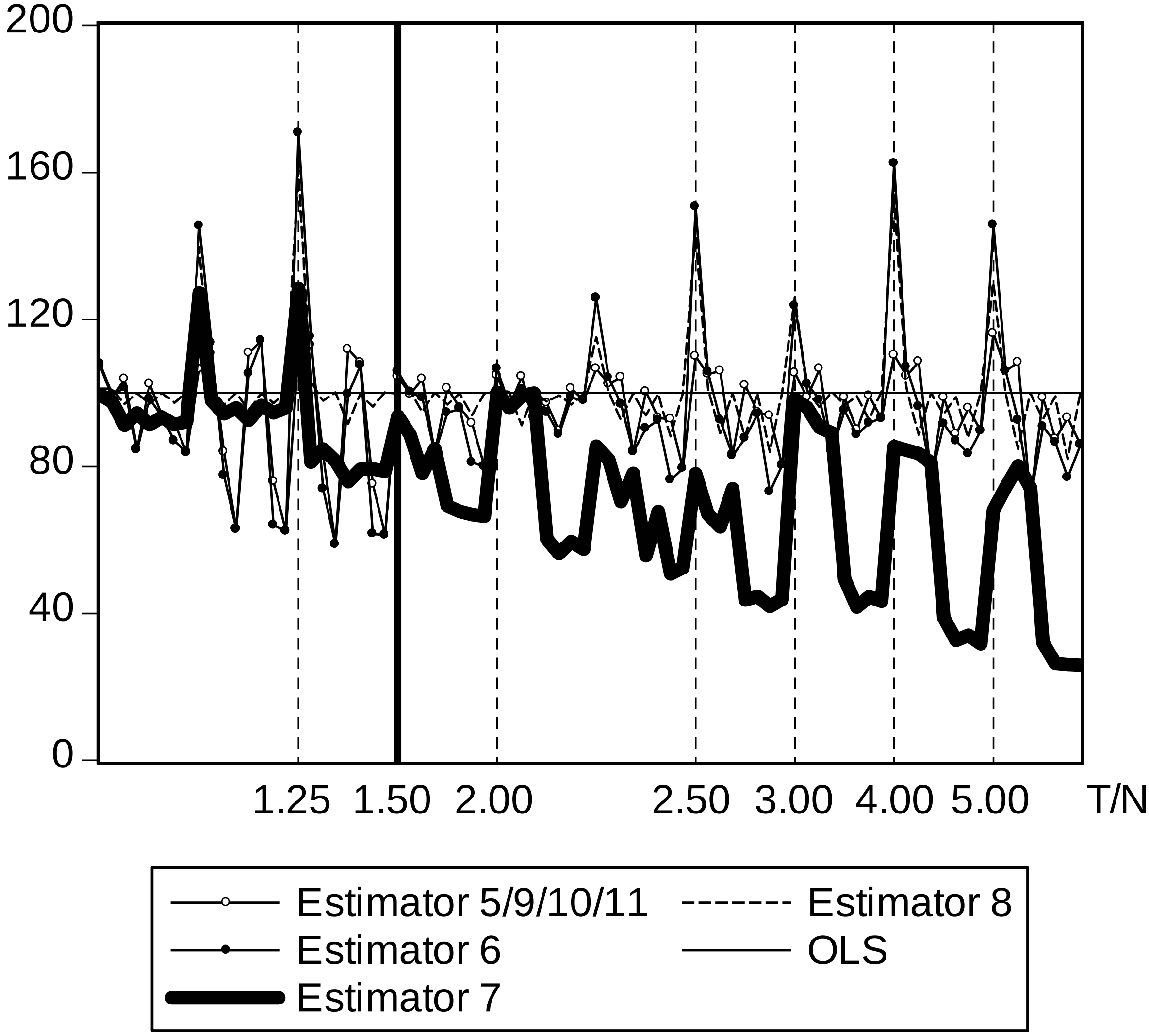
NOTES: Coefficient estimates are derived from OLS regression with White standard errors. The coefficient *p*-values are reported in parentheses below the respective estimates. Estimator 8 is the PCSE estimator.

TABLE 7
Comparison of Accuracy of Confidence Intervals Across Estimators When Data Set Is Characterized by *RHOHAT < 0.30*

	$N \leq T$	$N > T$
	Absolute Value of (*95-COVERAGE*) Over All Experiments	**Absolute Value of (*95-COVERAGE*) Over All Experiments**
Estimator 1	5.2	4.0
Estimator 2	4.5	1.8
Estimator 3	9.9	1.5
Estimator 4	3.7	1.4
Estimator 5	6.3	2.1
Estimator 6	4.9	2.0
Estimator 7	47.9	----
Estimator 8	3.1	2.4
Estimator 9	8.6	6.9
Estimator 10	19.9	6.4
Estimator 11	6.4	2.1

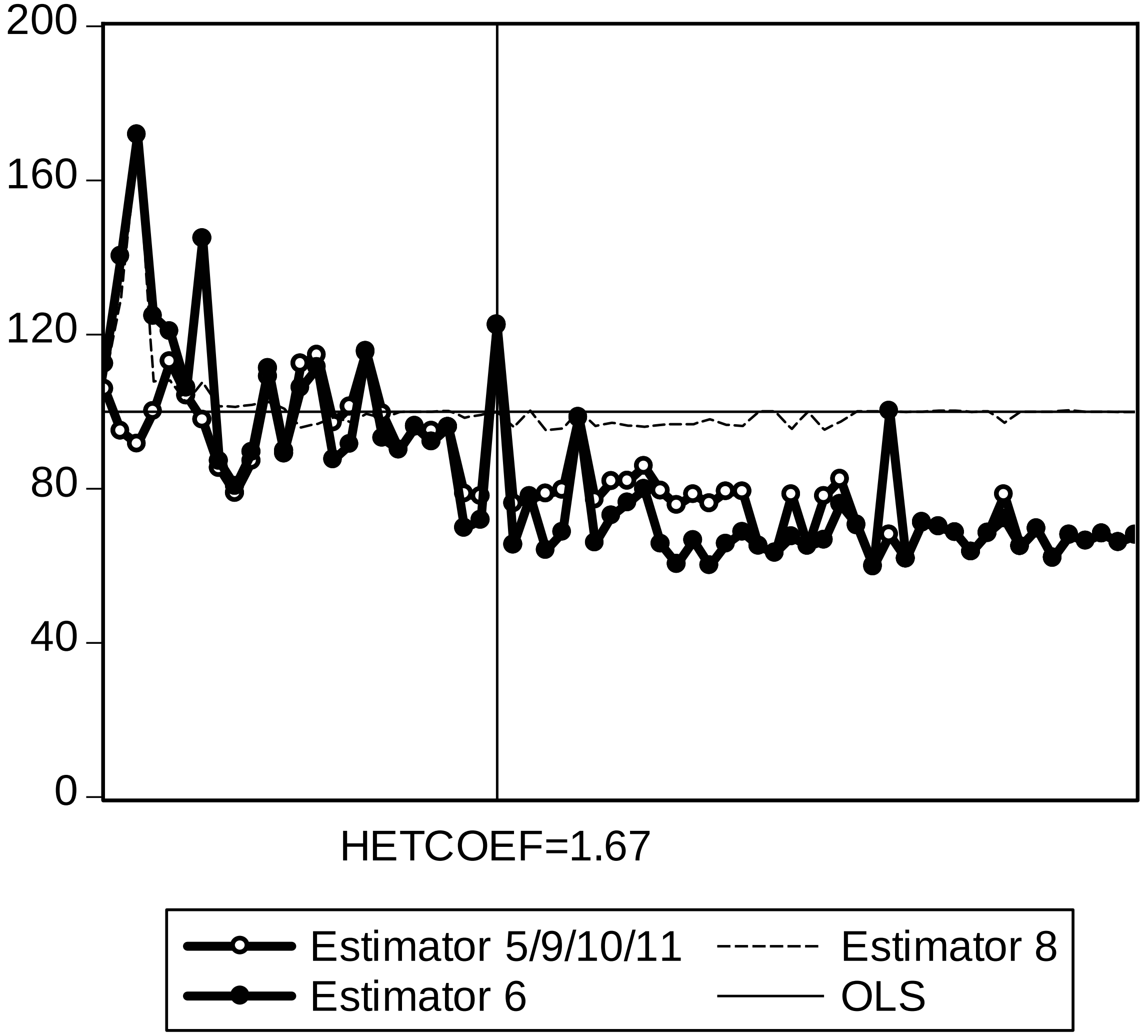
NOTES: There are *39* experiments where the respective data sets were sized with $N \leq T$, and *32* experiments where $N > T$.

FIGURE 1
Comparison of Estimator *EFFICIENCY*: $N \leq T$



NOTES: *EFFICIENCY* is represented by the vertical axis. The observations/experiments are sorted in ascending order of T/N. Observations to the left of the first vertical gridline have $T/N = 1.00$. Observations to the left of the second vertical gridline, and including the first gridline, have $T/N = 1.25$; and so on. Estimator 7 is the FGLS(Parks) estimator.

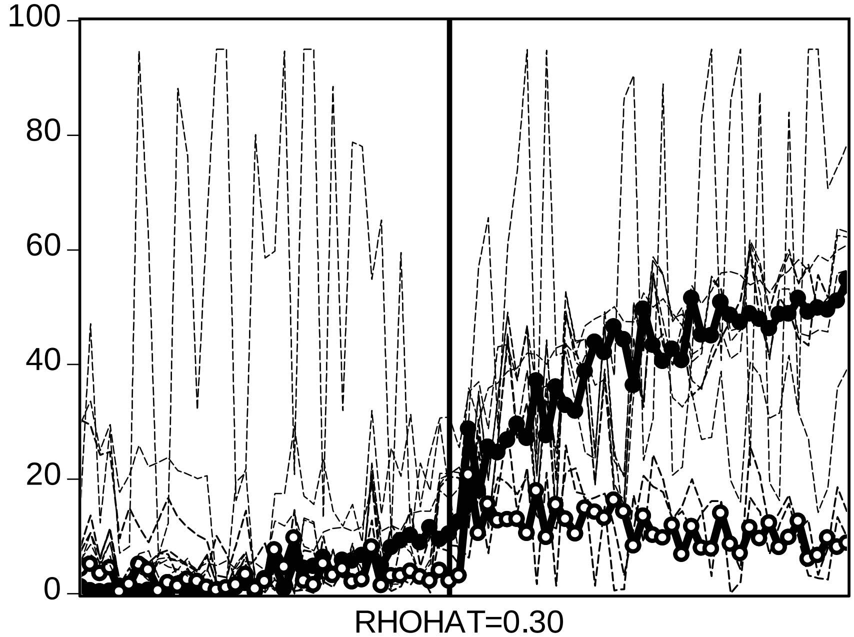
FIGURE 2
Comparison of Estimator *EFFICIENCY*: $N > T$



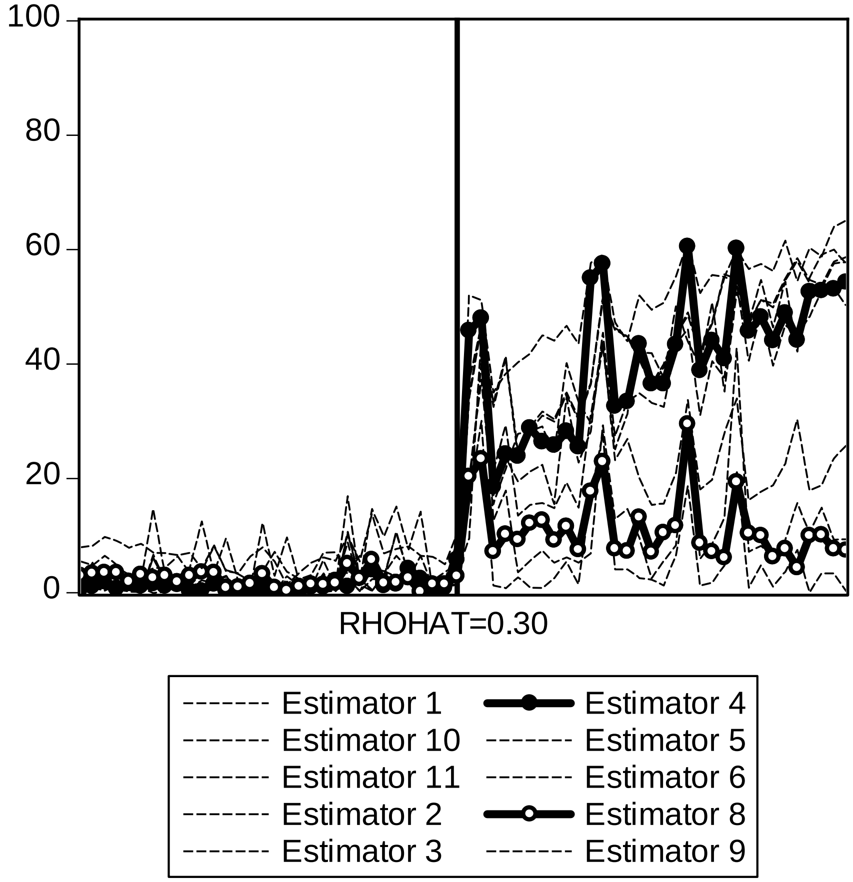
NOTES: *EFFICIENCY* is measured by the vertical axis. The observations/experiments are sorted in ascending order of *HETCOEF*, which is a measure of the degree of groupwise heteroscedasticity present in the data set. Further detail is in given in Section III in the text. Estimators 5/9/10/11 are equivalent to FGLS(Groupwise Heteroscedasticity). Estimator 6 is FGLS(Groupwise Heteroscedasticity + Serial Correlation).

FIGURE 3
Comparison of Accuracy of Confidence Intervals Across Estimators

(A) $N \leq T$

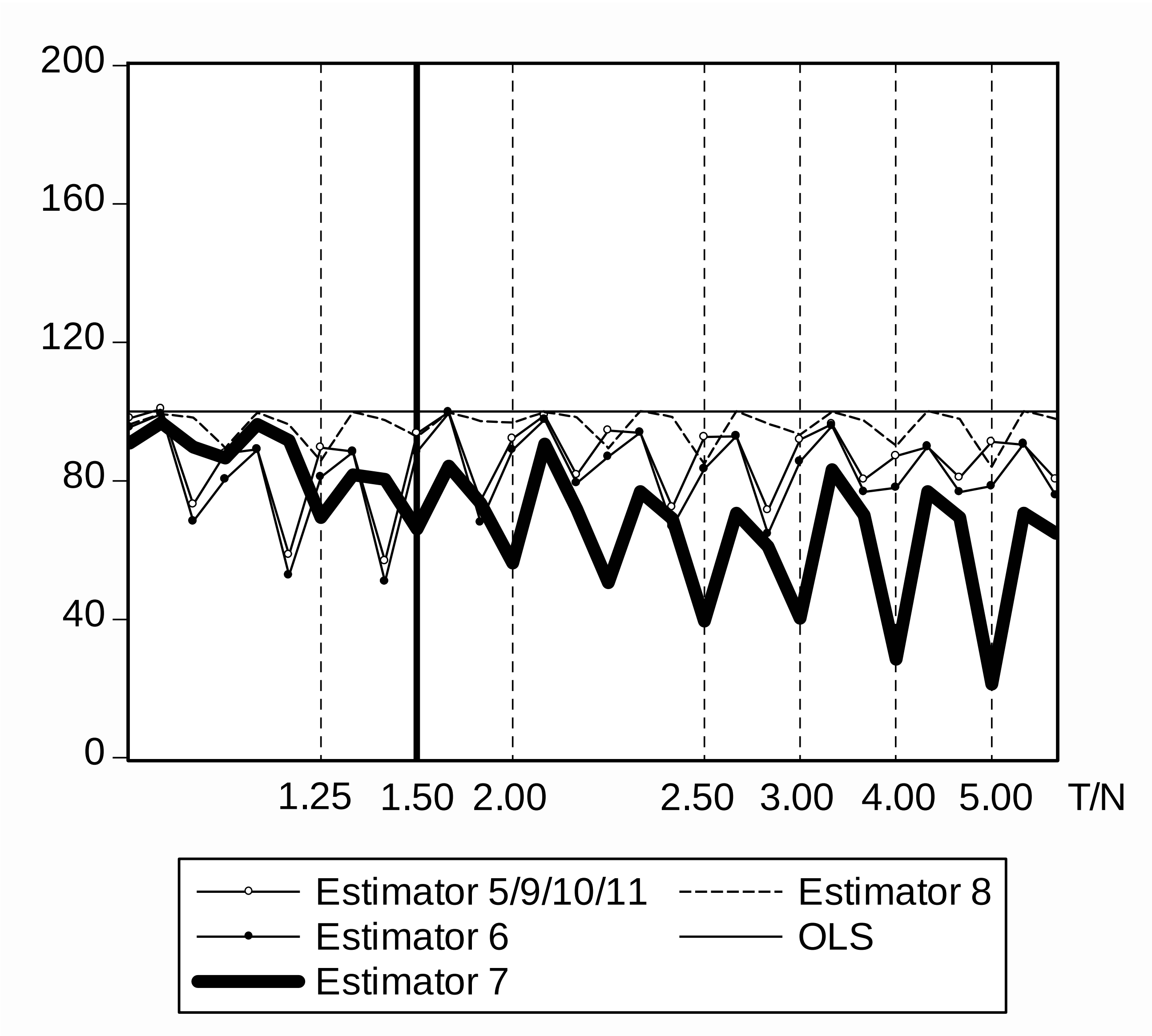


(B) $N > T$



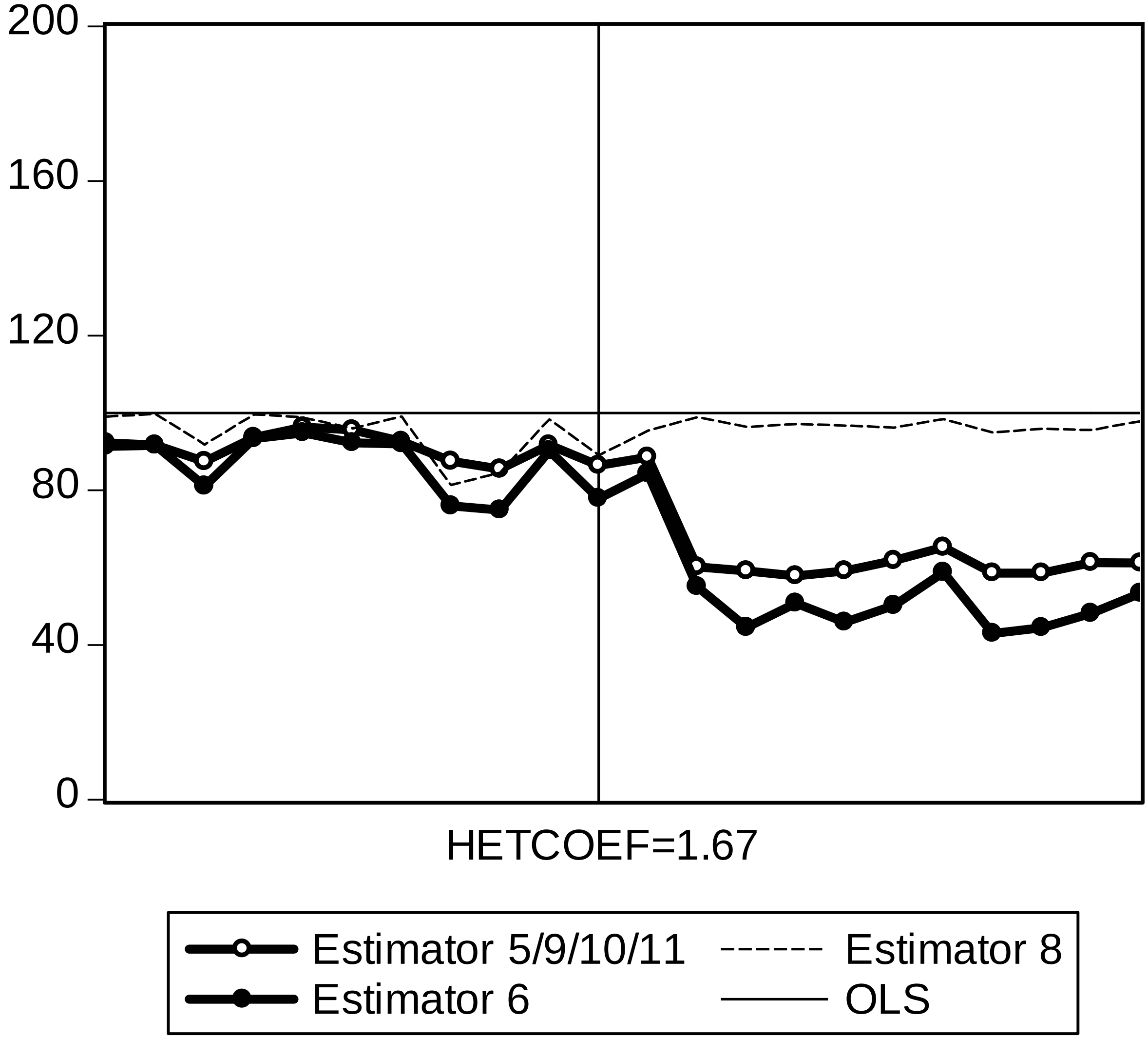
NOTES: The vertical axis reports accuracy of estimator confidence intervals using the measure $\left|95 - COVERAGE\right|$, which is described in Section IV of the text. The observations/experiments are sorted in ascending order of *RHOHAT*, which is a measure of the degree of serial correlation in the data set. This measure is defined in Section III in the text. Estimator 4 is OLS(Heteroscedasticity + Serial Correlation Robust). Estimator 8 is the PCSE estimator.

FIGURE 4
"Out of Sample" Comparison of Estimator *EFFICIENCY*: $N \leq T$



NOTES: This figure replicates the analysis of FIGURE 1 except that it uses the "out of sample" observations/experiments as described in Section V of the text.

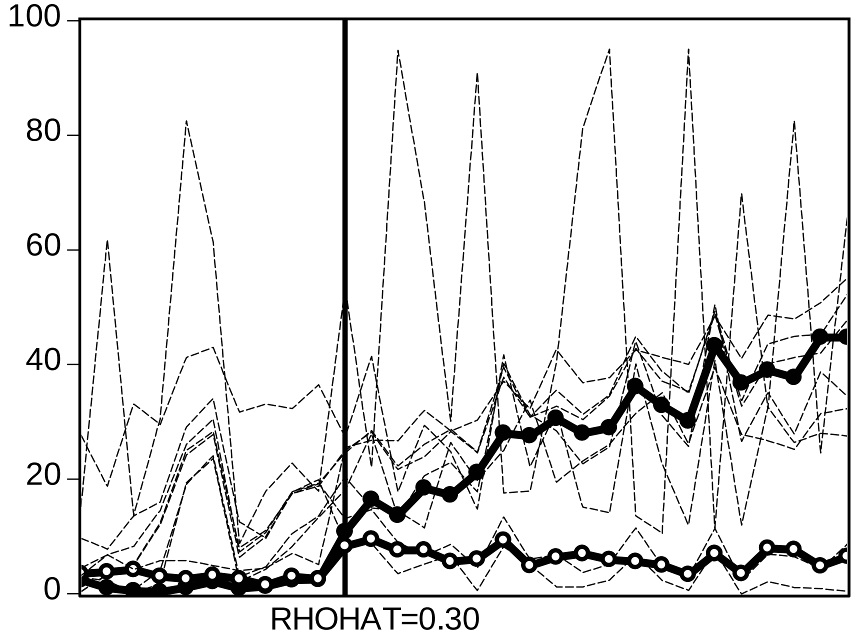
FIGURE 5
““Out of Sample” Comparison of Estimator *EFFICIENCY*: *N* > *T*



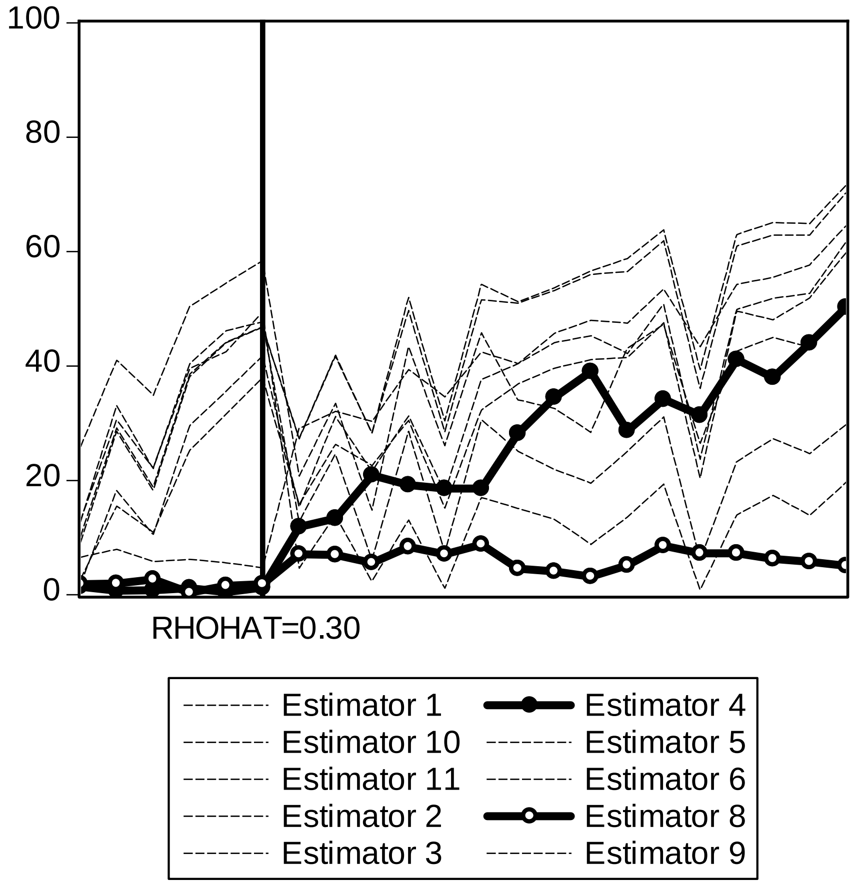
NOTES: This figure replicates the analysis of FIGURE 2 except that it uses the “out of sample” observations/experiments as described in Section V of the text.

FIGURE 6
"Out of Sample" Comparison of Accuracy of Confidence Intervals Across Estimators

(A) $N \leq T$



(B) $N > T$



NOTES: This figure replicates the analysis of FIGURE 3 except that it uses the “out of sample” observations/experiments as described in Section V of the text.

APPENDIX
Description of Procedure for Generating Simulated Panel Data Sets

Suppose we want to generate an artificial panel data set with N cross-sectional units and T time periods. Further, we want this data to "look like" the kind of data likely to be encountered in actual research. We represent the underlying DGP as a linear, fixed effects model with a Parks-style (Parks, 1967) error structure:

(A1) $$\begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \\ \vdots \\ \boldsymbol{y_N} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{X_1} \\ \boldsymbol{X_2} \\ \vdots \\ \boldsymbol{X_N} \end{bmatrix} \beta_x + \begin{bmatrix} \boldsymbol{\varepsilon_1} \\ \boldsymbol{\varepsilon_2} \\ \vdots \\ \boldsymbol{\varepsilon_N} \end{bmatrix}, \text{ or } \quad \boldsymbol{y} = \boldsymbol{\beta_0} + \boldsymbol{X}\beta_x + \boldsymbol{\varepsilon},$$

(A2) $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Omega}_{NT})$; and

(A3) $\boldsymbol{\Omega}_{NT} = \boldsymbol{\Sigma} \otimes \boldsymbol{\Pi}$,

$$\text{where } \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_{\varepsilon,11} & \sigma_{\varepsilon,12} & \cdots & \sigma_{\varepsilon,1N} \\ \sigma_{\varepsilon,21} & \sigma_{\varepsilon,22} & \cdots & \sigma_{\varepsilon,2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{\varepsilon,N1} & \sigma_{\varepsilon,N2} & \cdots & \sigma_{\varepsilon,NN} \end{bmatrix}, \ \boldsymbol{\Pi} = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{T-1} \\ \rho & 1 & \rho & \cdots & \rho^{T-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{T-1} & \rho^{T-2} & \rho^{T-3} & \cdots & 1 \end{bmatrix}, \text{ and } \sigma_{\varepsilon,ij} = \frac{\sigma_{u,ij}}{1-\rho^2}.$$

We use two different sets of $\boldsymbol{y}$ and $\boldsymbol{X}$ variables for the main Monte Carlo analyses: the log of real, annual U.S. state Per Capita Personal Income, and state tax burdens (for the state level analyses); and the log of real, annual per capita GDP, and government share of GDP (for the cross-country analyses). We illustrate our approach with the U.S. data.

We want to set values for the population parameters that are patterned after "real world" U.S. economic growth data. Towards that end, we start with forty years of PCPI and tax burden data on 48 states (omitting Alaska and Hawaii), covering the period 1960-1999. A long time series is crucial for our approach because we want to have multiple observations for each element of the error covariance matrix. Most studies use time series where T is between 10 and

25 years. By having a data series substantially longer than that, we can sample multiple T-year, panel data sets in order to construct a "representative" error structure for a T-year panel data set.

The first step consists of determining "representative" values for ρ and the $\sigma_{u,ij}$'s. We begin by creating a sample using the first N states in our data set.[20] Next, we choose the T-year period, 1960 to (1960+T-1). We then estimate a one-way fixed effects regression model for this sample, relating the dependent variable Y (= log of real U.S. state PCPI) to a set of state fixed effects (D^j) and the explanatory variable X (= tax burden).

$$\text{(A4)} \qquad Y_{it} = \sum_{j=1}^{N} \alpha_j D_{it}^j + \alpha_{N+1} X_{it} + error\ term_{it},$$

where $i=1,2, \ldots ,N;\ t=1960,1961,\ldots,1960+T\text{-}1;$ and D^j is a state dummy variable that takes the value 1 for state j. Equation (A4) is the "residual generating function." The residuals from this estimated equation are used to estimate ρ and the $\sigma_{u,ij}$'s in the usual manner, as if one were computing a conventional FGLS estimator. Denote the associated estimates from this sample as

$$\hat{\rho}_i \text{ and } \hat{\boldsymbol{\Phi}}_i = \begin{bmatrix} \hat{\sigma}_{u,11} & \hat{\sigma}_{u,12} & \cdots & \hat{\sigma}_{u,1N} \\ \hat{\sigma}_{u,21} & \hat{\sigma}_{u,22} & \cdots & \hat{\sigma}_{u,2N} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{u,N1} & \hat{\sigma}_{u,N2} & \cdots & \hat{\sigma}_{u,NN} \end{bmatrix}.$$

We repeat this process for every possible, T-contiguous year sample contained within the 40 years of data from 1960-1999 [i.e., 1960-(1960+T-1), 1961-(1961+T-1), 1962-(1962+T-1), …, (1999-T+1)-1999]. This produces a total of $40–T+1$ estimates of ρ and $\boldsymbol{\Phi}$, one for each T-contiguous year sample. We then average these to obtain "grand means" $\bar{\rho}$ and $\bar{\boldsymbol{\Phi}}$. Our "representative" $NT \times NT$ error structure, $\boldsymbol{\Omega}_{NT}$, is then constructed as follows:

[20] For example, since our data are organized alphabetically, the first fives states would be Alabama, Arizona, Arkansas, California, and Colorado.

(A5) $\boldsymbol{\Omega}_{NT} = \overline{\boldsymbol{\Sigma}} \otimes \overline{\boldsymbol{\Pi}}$,

where

(A6) $\overline{\boldsymbol{\Sigma}} = \frac{1}{\left(1-\overline{\rho}^2\right)}\overline{\boldsymbol{\Phi}}$,

and

$$\text{(A7)} \quad \overline{\boldsymbol{\Pi}} = \begin{bmatrix} 1 & \overline{\rho} & \overline{\rho}^2 & \cdots & \overline{\rho}^{T-1} \\ \overline{\rho} & 1 & \overline{\rho} & \cdots & \overline{\rho}^{T-2} \\ \overline{\rho}^2 & \overline{\rho} & 1 & \cdots & \overline{\rho}^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \overline{\rho}^{T-1} & \overline{\rho}^{T-2} & \overline{\rho}^{T-3} & \cdots & 1 \end{bmatrix}.$$

This becomes the population error covariance matrix used for the associated Monte Carlo experiment. Note that every element of $\boldsymbol{\Omega}_{NT}$ is based on error covariance matrices estimated from actual panel data. In this sense, $\boldsymbol{\Omega}_{NT}$ can be said to be "representative" of the kinds of error structures one encounters in "real world" data.

The next step consists of setting values for β_x and the distribution of $\boldsymbol{X}$. In the context of our example, β_x represents the impact on economic growth of an increase in taxes. We set $\beta_x = -0.01$ to be consistent with estimates from the literature (cf. Wasylenko, 1997).

The distribution of $\boldsymbol{X}$ is constructed similarly to how we constructed the "grand mean" values, $\overline{\rho}$ and $\overline{\boldsymbol{\Phi}}$: Define $\boldsymbol{X}_i$ as the $NT \times 1$ vector of observations of the explanatory variable for the first *N* states in the sample and the time period, 1960 to (1960+*T*-1). Repeat the process for all *T*-contiguous year samples contained within the 40 years of data from 1960-1999 [i.e., 1960-(1960+*T*-1), 1961-(1961+*T*-1), 1962-(1962+*T*-1), …, (1999-*T*+1)-1999]. This produces a total of $40–T+1$ distributions of $\boldsymbol{X}$, which are then averaged to obtain $\overline{\boldsymbol{X}}$. $\boldsymbol{X}$ is set equal to $\overline{\boldsymbol{X}}$.

The β_i , $i = 1,2,...N$ are nuisance parameters. In order to generate artificial $\boldsymbol{y}$ values that mimic "real world" $\boldsymbol{y}$ values, we exploit the fact that the OLS regression line must pass through the sample means. Define β_0 such that

(A8) $\boldsymbol{i}_{NT}\beta_0 = \bar{\boldsymbol{y}} - \bar{\boldsymbol{X}}\beta_x = \bar{\boldsymbol{y}} + 0.01\bar{\boldsymbol{X}}\beta$.

where $\bar{\boldsymbol{y}}$ is constructed in the same manner as $\bar{\boldsymbol{X}}$. Note that $\bar{\boldsymbol{y}}$ is used solely to generate a value for β_0 that can be used in the subsequent Monte Carlo analysis.

Given values for β_0, β_x, ρ, the $\sigma_{u,ij}$'s, and the distribution of $\boldsymbol{X}$, experimental observations are generated in the usual manner. Define $\boldsymbol{u}$ as an $NT \times 1$ vector of standard normal random variables. Define $\boldsymbol{Q}$ such that $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{\Omega}_{NT}$. Error terms are created by $\boldsymbol{\varepsilon} = \boldsymbol{Q}'\boldsymbol{u}$. These simulated errors are added to the deterministic component, $\beta_0 + \beta_x\, x_i$, to calculate stochastic observations of y_i, where $y_i = \beta_0 + \beta_x\, x_i + \varepsilon_i$, $i=1,2,...,NT.$ Given an experimental data set of NT observations of (y_i, x_i), we estimate β_x using the respective estimators from TABLE 1. We perform a thousand replications of this experiment, generating a thousand estimates of β_x for each estimator in each experiment.[21]

This same procedure can be modified in a straightforward manner to conduct Monte Carlo experiments for alternative N and T values. In turn, the same general procedure can be used to create artificial data patterned after other kinds of data, like international data on the log of real per capital GDP. We also use an alternative form of the "residual generating function" where we estimate a two-way fixed effects model, including dummy variables for time. This has the twin advantages of reducing cross-sectional dependence and increasing R^2. For example, for

[21] This is essentially the same experimental design employed by Beck and Katz(1995), except that we pattern our DGP population parameters after "real world" data.

U.S. state level income data with state fixed effects, typical R^2 values run between 60-70 percent. Adding time fixed effects typically raises this to over 90 percent.