ANOTHER LOOK AT WHAT TO DO WITH TIME-SERIES CROSS-SECTION DATA

by

Xiujian Chen, Shu Lin, and W. Robert Reed*

Abstract

Our study revisits Beck and Katz' (1995) comparison of the Parks and PCSE estimators using time-series, cross-sectional data (TSCS). Our innovation is that we construct simulated statistical environments that are designed to closely match "real-world," TSCS data. We pattern our statistical environments after income and tax data on U.S. states from 1960-1999. While PCSE generally does a better job than Parks in estimating standard errors, it too can be unreliable, sometimes producing standard errors that are substantially off the mark. Further, we find that the benefits of PCSE can come at a substantial cost in estimator efficiency. Based on our study, we would give the following advice to researchers using TSCS data: Given a choice between Parks and PCSE, we recommend that researchers use PCSE for hypothesis testing, and Parks if their primary interest is accurate coefficient estimates.

Revised, April 13, 2005

*The contact author is W. Robert Reed, email: breed@ou.edu, fax: (405) 325-5842, phone: (405) 325-2358.

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.

I. INTRODUCTION

Empirical studies frequently employ data consisting of repeated time-series observations on fixed, cross-sectional units. While providing a rich amount of information, time-series cross-sectional (TSCS) data are likely to be characterized by complex error structures. The application of OLS to data with nonspherical errors produces inefficient coefficient estimates, and the corresponding standard error estimates are biased. In contrast, GLS produces coefficient and standard error estimates that are efficient and unbiased, respectively, given certain assumptions. Two such assumptions are (i) the error covariance structure is correctly specified, and (ii) the elements of the error covariance matrix are known. Feasible GLS (FGLS) is used when the structure of the error covariance matrix is known, but its elements are not. The finite sample properties of FGLS are analytically indeterminate.

Beck and Katz (1995) (henceforth, BK) use Monte Carlo methods to study the performance of FGLS in a statistical environment characterized by (i) groupwise heteroscedasticity, (ii) first-order serial correlation, and (iii) contemporaneous cross-sectional correlation. They dub the corresponding FGLS estimator "Parks" (after Parks [1967]). BK report three major findings:

1. Parks produces dramatically inaccurate standard errors.
2. An alternative method, based on OLS but using "panel-corrected standard errors," (henceforth, PCSE) produces accurate standard errors.
3. The efficiency advantage of Parks over PCSE is at best slight, except in extreme cases of cross-sectional correlation, and then only when the number of time periods (T) is at least twice the number of cross-section units (N).

Consequently, BK prescribe that researchers use the PCSE procedure when working with TSCS data.[1]

BK has been very influential. A recent count identified over 350 citations, primarily in the political science literature (cf. Web of Science, www.isinet.com/products/citation/wos). Their PCSE estimator has been applied in studies using both U.S. and international data. It was recently added as an estimation procedure within the statistical software package STATA (StataCorp, 2001).[2]

Our paper constructs a statistical environment modeled after real-world TSCS data and revisits BK's analysis of the Parks and PCSE estimators. We first construct a "Parks-type" statistical environment, and then attempt to replicate BK's findings using similar Monte Carlo techniques. We confirm BK's result that Parks consistently underestimates coefficient standard errors. However, we find that PCSE can also substantially underestimate coefficient standard errors. Further, we find that PCSE is much less efficient than reported by BK.

We next construct a completely general statistical environment, and repeat our analysis. We once again obtain the result that PCSE generally does a better job than FGLS when estimating standard errors. However, the standard error benefits of PCSE over Parks are less, and the costs in terms of diminished efficiency are greater.

Our results suggest that PCSE is superior to Parks when the researcher's main focus is hypothesis testing. However, even PCSE estimates of standard errors can be misleading. Further, Parks is superior to PCSE when the main concern is obtaining

[1] A further advantage of PCSE is that it is able to incorporate cross-sectional correlation when the number of time series observations (T) is less than the number of cross-sectional observations (N), whereas standard FGLS cannot.

[2] The corresponding command is "xtpcse".

accurate coefficient estimates. We conclude that researchers should use both procedures, relying on the PCSE estimates for hypothesis testing, and Parks for coefficient estimates.

Our paper proceeds as follows. Section II re-evaluates BK's Monte Carlo analysis within a "Parks-type" statistical environment. We set the values of the elements of the population covariance matrix equal to their respective values in real-world TSCS data. Section III repeats this analysis, generalizing the statistical environment so that it more closely approximates real-world TSCS data. Section IV concludes.

II. RE-EVALUATING BK WITHIN A "PARKS-TYPE" STATISTICAL ENVIRONMENT

IIA. Methodology for producing a "Parks-type" statistical environment patterned on actual TSCS data

BK build their Monte Carlo analysis around the following TSCS model:

$$\begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \\ \vdots \\ \boldsymbol{y_N} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X_1} \\ \boldsymbol{X_2} \\ \vdots \\ \boldsymbol{X_N} \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\varepsilon_1} \\ \boldsymbol{\varepsilon_2} \\ \vdots \\ \boldsymbol{\varepsilon_N} \end{bmatrix}, \text{ or } \quad \boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}\,;$$

where $\boldsymbol{y_i}$ is a $T \times 1$ vector of observations on the dependent variable in the i^{th} group, $i = 1,2,...,N$; $\boldsymbol{X_i}$ is a $T \times K$ matrix of exogenous variables; $\boldsymbol{\beta}$ is a $K \times 1$ vector of coefficients; $\boldsymbol{\varepsilon}_i$ is a $T \times 1$ vector of error terms; and $\boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \boldsymbol{\Omega}_{NT})$.

Following Parks (1967), they allow $\boldsymbol{\Omega}_{NT}$ to consist of (i) groupwise heteroscedasticity; (ii) groupwise, first-order serial correlation; and (iii) cross-sectional (spatial) correlation. Specifically,

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

where $\sigma_{\varepsilon,ij} = \dfrac{\sigma_{u,ij}}{1-\rho_i\rho_j}$ and $\boldsymbol{\Sigma}_{ij} = \begin{bmatrix} 1 & \rho_j & \rho_j^2 & \cdots & \rho_j^{T-1} \\ \rho_i & 1 & \rho_j & \cdots & \rho_j^{T-2} \\ \rho_i^2 & \rho_i & 1 & \cdots & \rho_j^{T-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_i^{T-1} & \rho_i^{T-2} & \rho_i^{T-3} & \cdots & 1 \end{bmatrix}$.

They proceed by selecting various combinations of N and T, specifying the elements of the respective error covariance matrices (the $\boldsymbol{\Omega}_{NT}$'s) by positing values for the population parameters $\sigma_{u,ij}$, ρ_i, and ρ_j, $i, j = 1,2,..., N$.

Given $\boldsymbol{\Omega}_{NT}$, experimental observations are generated in the usual manner. Define $\boldsymbol{u}$ as a 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 a deterministic component, $\beta_0 + \beta_x x_i$, to calculate stochastic observations of y_i, where $y_i = \beta_0 + \beta_x x_i + \varepsilon_i$, $i=1,2,...,NT$. BK create the x_i's from a zero-mean normal distribution (fixed in all replications), and set β_0 and β_x equal to 10 in all experiments. They perform 1000 replications for each experiment.

BK compare the (i) Parks and (ii) PCSE estimates of β_x. They employ several performance measures, including "Level" and "Efficiency." "Level" calculates the percent of estimated 95% confidence intervals that include the true value of β_x. "Efficiency" measures the relative efficiency of PCSE to Parks and is defined by

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

An Efficiency value less than (greater than) 100 indicates that PCSE is less efficient (more efficient) than Parks.

IIB. Constructing a "Parks-type" statistical environment based on actual TSCS data

Our experiments follow BK's methodology with one major exception: We pattern our simulated statistical environment on actual data according to the following two-stage procedure: In the first stage, we estimate the parameters $\sigma_{u,ij}$, ρ_i, and ρ_j, $i, j = 1,2,..., N$ from actual TSCS data. In the second stage, we use these estimated values as population values in the subsequent Monte Carlo experiments. This ensures that our simulated data look like real TSCS data.[3]

For our "real-world" TSCS data, we use two data sets. The first data set consists of annual, state-level observations on income (specifically, the log of real Per Capita Personal Income). The second data set consists of annual, state-level observations on taxes (specifically, Tax Burden, defined as the ratio of total state and local taxes over Personal Income).

We select these data for several reasons. First, many of the studies that employ PCSE use state-level data (Nicholson-Crotty, 2004; Barrilleaux and Berkman, 2003; Kousser, 2002; Boehmke, 2002; Crowley and Skocpol, 2001; and Fording, 2001).

[3] BK recommend that empirical estimation of PCSE's restrict the autocorrelation parameters to be the same across groups (i.e., $\rho_i = \rho_j = \rho$ for all $i, j = 1,2,..., N$). Accordingly, we directly impose this on the simulated statistical environment and then look to the TSCS data to provide a "realistic" value for ρ.

Second, both state incomes and state taxes have been the subject of much previous research, and continue to be actively researched.[4]

Third, there exists a long time series for both sets of data. We employ 40 years of data, on 48 states (omitting Alaska and Hawaii), covering the period 1960-1999. A long time series is crucial for our approach. Most studies use time series where *T* is between 10 and 25 years (cf. Table 1 in BK). By having a data series substantially longer that *T*, we can sample multiple *T*-year, TSCS data sets in order to construct a "representative" error structure for a *T*-year (cross-sectional) time series. We then use this representative error structure to generate experimental observations through the standard Monte Carlo procedure.

Our approach works like this: Suppose we want to construct a Parks-type error covariance structure ($\boldsymbol{\Omega}_{NT}$, $N=5$, $T=10$) for a regression model with either state income or state taxes as the dependent variable. We begin by choosing the first 5 states in our data set.[5] Next, we choose the 10-year period, 1960-1969. We then estimate a fixed effects model relating the respective dependent variable (Y) to a set of state fixed effects (D^j), and an explanatory variable X (more on X below):

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

where $i=1,2, \ldots ,N$; $t=1,2,\ldots,T$; $N=5$; $T=10$; and D^j is a state dummy variable that takes the value 1 for state j. We refer to this equation as the "residual generating function."

[4] For literature on state incomes, see Dye and Feiock (1995), Jones (1990), Brace (1991), Brierly and Feiock (1993), Hendrick and Garand (1991), Dye (1980), and Brace et al. (1989). For literature related to state taxes/revenues, see Reed (2005), McAtee, Yackee, and Lowery (2003) and Alt and Lowery (1994, 2000, 2003).

[5] The first 5 states in our data sets are Alabama, Arkansas, Arizona, California, and Colorado.

The residuals from this estimated equation are used to estimate the "Parks-method" error covariance matrix, $\hat{\boldsymbol{\Omega}}_{5,10}$, in the standard manner. Our innovation is that we do this for every possible 10-contiguous year period in our 40 year sample (i.e., 1960-1969, 1961-1970, …, 1990-1999 – a total of $40–T+1$ subsamples). We then average these error covariance matrices to obtain a "representative" covariance matrix, $\overline{\boldsymbol{\Omega}}_{5,10}$. This becomes our "population" covariance matrix for the Monte Carlo experiments.

We proceed by generating experimental observations of y_i, where $y_i = \beta_0 + \beta_x x_i + \varepsilon_i$, $i=1,2,...,NT$, $NT=50$, and the errors are simulated from the population error covariance matrix, $\overline{\boldsymbol{\Omega}}_{5,10}$.[6] We set the values of β_0 and the x_i's to be representative of their respective data sets, and fix the value of β_x consistent with the empirical literature on income/taxes.[7]

Given an experimental data set of $NT=50$ observations of $(y_i x_i)$, we estimate β_x using the Parks and PCSE estimators, respectively. We perform 1000 replications of this experiment, generating 1000 estimates of β_x for both the Parks and PCSE estimators. These 1000 estimates are then analyzed to compare the performance of the two

---

[6] Note that the ε_i are orthogonal to the x_i by construction. Further, the influence of fixed effects is "filtered out" via the residual generating function employed in the first stage of the data-generating procedure. Hence there is no need to include fixed effects in the simulated data generating equation.

[7] For the income equations, we use Tax Burden as the explanatory variable and set $\beta_x = -0.01$ (see, for example, Helms [1985] and Wasylenko [1997]). For the tax equations, we use the log of real Per Capita Personal Income and set $\beta_x = -1.0$ (see, for example, Reed [2005]). The fact that each of the variables appears in the other residual generating function as an explanatory variable may raise concerns. With respect to the literature, these are common specifications. As a practical matter, the inclusion/exclusion of these explanatory variables in the residual-generating functions has a negligible effect on the results. Our only motivation for including them is to address potential concerns that the resulting error structure be independent of the explanatory variable in the simulated data.

estimators. This same procedure can be modified in a straightforward manner to conduct Monte Carlo experiments for alternative *N* and *T* values.

At this point it bears revisiting the claim that our procedure approximates the error structure that the researcher is likely to encounter in real-life research problems. Admittedly, the residual-generating function specified above represents a stripped down version of the specifications usually employed by researchers. Other variables typically would be included in the specification.

Unfortunately, there is no single specification that dominates the empirical literature on state incomes/taxes. As a result, we experimented with alternative residual generating functions that added a lagged dependent variable and/or time fixed effects. We found that our main results were qualitatively unaffected by these more elaborate specifications. Accordingly, we only report results based on the residual-generating function with state fixed effects.

Our study conducts experiments for a wide range of "sizes" of TSCS data sets: We set values for *N* equal to *5, 10*, *20*, and *48*; and values for *T* equal to *10, 15, 20,* and *25* -- a total of sixteen *N* and *T* combinations. This range encompasses most of the data sets reported in BK's Table 1.

The first column of TABLE 1 summarizes salient characteristics of the data for the "Parks-type" statistical environment. The top part of TABLE 1 reports on the income data, the bottom part on the tax data. "*Mean* R^2" refers to the average R^2 for the respective residual generating functions in the first stage of the data-generating process. In other words, a typical fixed-effects regression equation "explained" approximately

73% of the variation in the (real) income data, and *70%* of the variation in the (real) tax data.

The subsequent rows characterize the serial correlation, heteroscedasticity, and cross-sectional correlation behavior of the *simulated* data produced in the second-stage of the data-generating process, employing the "Parks-type" population covariance matrix, $\overline{\boldsymbol{\Omega}}$. These data comprised the actual observations used to estimate β_x with the Parks and PCSE procedures, respectively.

Both the simulated income and tax data evidenced substantial degrees of serial correlation. The average of the estimated $\hat{\rho}$ values, "*Mean* $\hat{\rho}$" (averaged over all replications and experiments) was *0.61* for the income data, and *0.58* for the tax data.

As a measure of groupwise heteroscedasticity, we estimated group-specific standard errors ($\hat{\sigma}_i$, $i=1,\ldots,N$) for each replication and rank-ordered them from smallest to largest. We then calculated a "heteroscedasticity coefficient" (h), defined as the ratio of the upper quartile value of $\hat{\sigma}_i$ over its lower quartile value, again averaged over all replications and experiments. The "heteroscedasticity coefficient" value for the income data was *1.24*, and the corresponding value for the tax data was *1.59*.

Finally, both the simulated income and simulated tax data were characterized by substantial cross-sectional correlation. "*Mean* r_{ij}" is defined as the mean (absolute) value of the contemporaneous correlation between errors from groups i and j, averaged over all possible cross-sectional correlations, and over all replications and experiments. "*Mean* r_{ij}" for the income data was *0.74*, and *0.36* for the tax data. Note that the income data displayed a much greater degree of cross-sectional correlation than the tax data.

In summary, our simulated data were characterized by precisely the kinds of statistical problems (i.e., serial correlation, groupwise heteroscedasticity, and cross-sectional correlation) that the Parks and PCSE procedures are designed to handle.

IIC. Monte Carlo experiments assuming a "Parks-type" statistical environment

TABLE 2 reports the results of our Monte Carlo experiments assuming a "Parks-type" statistical environment. Each experiment consisted of a 1000 replications of simulated TSCS data of size *NT*. Separate experiments were conducted for all sixteen *NT* combinations. We note that the Parks method is not applicable when $N > T$, which is probably why BK do not report Monte Carlo results for these cases. However, as BK's Table 1 shows, $N > T$ for many TSCS data sets, and thus we think researchers will be interested to know how PCSE fares (in an absolute sense) in these environments.

The left hand side of TABLE 2 reports the performance of Parks and PCSE with respect to "Level." The top panel (Panel A) of TABLE 2 reports the results using the simulated income data. Centering our attention first on the Parks results, we find -- consistent with BK -- that Parks substantially, in some cases dramatically, underestimates coefficient standard errors, resulting in confidence intervals that are too narrow (i.e., "overconfident"). The Parks "Level" values range from a high of *67.2* percent for $N=5$, $T=25$; to an abysmally low *8.8* percent for $N=20$, $T=20$. In other words, when $N=20$ and $T=20$, less than 10 percent of the 95% confidence intervals include the true value of β_x, causing the null hypothesis to be rejected much too frequently. While the biasedness of FGLS was known before BK, their work was important in establishing the degree to which Parks underestimates standard errors. Our research confirms this finding of theirs.

Turning now to the PCSE "Level" values, we come across our first surprising finding: While PCSE always does a better job than Parks when estimating confidence intervals, it also underestimates standard errors. The PCSE "Level" values range from a high of *88.0* percent (*N=48, T=25)*, to a low of *72.0* percent (*N=5, T=10)*. Across all sixteen *NT* experiments, the mean "Level" value for PCSE is *79.3*, substantially less than its "expected" value of 95.

The source of our surprise comes from the fact that there is no mention of this possibility in BK. Upon reflection, however, this result should have been anticipated. The analytic expressions for the PCSE standard errors, like those for the FGLS standard errors, assume that the elements of the population covariance matrix are known. In reality, they are unknown and must be estimated. Estimation of these parameters introduces an additional degree of uncertainty that is not incorporated in the standard error formulae.[8] Thus, the standard error formulae are biased downwards.

The right hand side of TABLE 2 reports the efficiency of PCSE relative to Parks. Values less than 100 indicate that PCSE is less efficient than Parks. The right hand side of TABLE 2 makes clear that the improvement of PCSE with respect to standard errors comes at a cost of lower efficiency. Actually, BK would have predicted these "Efficiency" results. They write:

---

[8] The formulae for the Parks and PCSE estimates of the coefficient covariance matrix are $Cov(\hat{\boldsymbol{\beta}})_{Parks-FGLS} = (\boldsymbol{X'\hat{\Omega}^{-1}X})^{-1}$ and $Cov(\hat{\boldsymbol{\beta}})_{PCSE} = (\boldsymbol{X'X})^{-1}(\boldsymbol{X'\hat{\Omega}X})(\boldsymbol{X'X})^{-1}$, respectively. The reason given by BK for the poor performance of $Cov(\hat{\boldsymbol{\beta}})_{Parks-FGLS}$ is that there are relatively few observations to estimate the large number of parameters in $\hat{\boldsymbol{\Omega}}$. The crux of the problem is this: "Each element of the matrix of contemporaneous covariances is estimated using, on average, *2T/N* observations. Many…panel studies have ratios of *T* to *N* very close to 1, so covariances are being estimated with only slightly more than two observations per estimate!" (page 637). It should be clear that this problem also affects $Cov(\hat{\boldsymbol{\beta}})_{PCSE}$, since the same $\hat{\boldsymbol{\Omega}}$ appears in both (Parks and PCSE) covariance expressions.

> [PCSE] is, as expected, more efficient than Parks when the errors are uncorrelated (spherical). But even when the average correlation of the errors rises to .25, [PCSE] remains slightly more efficient than Parks. Parks becomes more efficient than [PCSE] when average contemporaneous correlations rise to .50, but this advantage is noticeable only when the number of time points is at least double the number of units. Even here, the efficiency advantage of Parks over [PCSE] is under 20%. Only when the average contemporaneous correlation of the errors rises to .75 is the advantage of Parks marked, and then only when *T* is twice *N* (page 642).

Referring back to TABLE 1 we see that the simulated, income data sets are indeed characterized by a high degree of contemporaneous correlation (the average contemporaneous correlation across all the data sets used in Panel (A) of TABLE 2 is *0.74*). Large efficiency costs occur only when *T* is more than twice the size of *N*. Therefore, we next turn to the tax data, where the average correlation of the errors is *0.36* (cf. TABLE 1) and, according to BK, cross-sectional correlation should not be much of a problem.

In fact, the results for the simulated tax data are very similar to those for the income data. Specifically, we once again find that:

1. Parks substantially underestimates coefficient standard errors, resulting in confidence intervals that are much too narrow.

2. PCSE produces more reliable standard error estimates than Parks. However, PCSE also underestimates coefficient standard errors, producing overly narrow confidence intervals.

3. The improvement in standard error estimates provided by PCSE comes at the cost of decreased efficiency.

Whether this tradeoff in improved standard error estimation is worth the cost in diminished efficiency is, of course, a subjective evaluation that each researcher must make for themselves. However, two things are noteworthy here. First, we find substantial efficiency costs even when the degree of cross-sectional correlation would be in the "acceptable" range according to BK.[9] The "*Mean* r_{ij}" value for the tax data is only 0.36 (cf. TABLE 1), and was never higher than 0.41 for any of the individual experiments. Yet the average "Efficiency" value across all experiments was only *81.9*. And second, when faced with similar efficiency losses (cf. the last column in their Table 5, page 642) BK counsel that "researchers should consider alternatives to [PCSE]" (page 642).

We conclude this section by reporting that we obtained these same results using several different residual generating functions, all within the "Parks-type" statistical environment studied by BK. Of course, in real life, there is no guarantee that the statistical environment falls within the "Parks-type" category. How do Parks and PCSE compare in a statistical environment that more closely matches the kind of TSCS data that researchers are likely to encounter in real life? That is the subject of our next section.

III. EXTENDING BK'S ANALYSIS TO A MORE GENERAL STATISTICAL ENVIRONMENT

While a "Parks-type" statistical environment is generally viewed as being quite general, it should be noted that it imposes substantial limitations on $\boldsymbol{\Omega}$. Given that $\boldsymbol{\varepsilon}$ is $NT \times 1$, there are $\frac{NT(NT+1)}{2}$ unique parameters in the unrestricted version of $\boldsymbol{\Omega}$. In contrast, there are $\frac{N^2+3N}{2}$ unique parameters in the Parks specification of $\boldsymbol{\Omega}$ (counting the group-specific AR[1] parameters). In other words, the Parks model scales down the

[9] In BK's Table 5 (page 642), "Efficiency" is never below *100* when the average contemporaneous correlation is 0.25. For experiments where $10 \le N \le 20$, $10 \le T \le 30$ and the average contemporaneous correlation is 0.50, "Efficiency" is never below *88*. Compare this to the results in Panel (B) of TABLE 2: Here, "Efficiency" is always less than *88* except when T=N.

number of unique parameters in $\boldsymbol{\Omega}$ by approximately $\frac{1}{T^2}$. As it is common in empirical studies using TSCS data for T to range between 10 and 25 years of data (or more), this constitutes a substantial restriction on $\boldsymbol{\Omega}$. Since both FGLS (Parks) and BK's PCSE procedure are not designed to be applied outside the "Parks-type" statistical environment, it is unclear how they will behave, both absolutely and relatively, in a more "realistic" statistical environment.

This section addresses the following questions: Suppose one uses a "real-world data set" and assumes (incorrectly) that it fits the Parks-type statistical model. Will PCSE still underestimate coefficient standard errors? Will PCSE still do a better job than Park of estimating standard errors? And will PCSE still be less efficient than Parks? While BK never compare Parks and PCSE outside a "Parks-type" statistical environment, we think that researchers will find our results of interest given the widespread popularity of the PCSE methodology.

To construct our more general statistical environment, we repeat the process described in Section (IIB) up to the point where the residuals from the "residual generating function" are used to construct the sample covariance matrix. Rather than constructing a "Park-type" error covariance matrix, we construct the unrestricted error covariance matrix, $\hat{\boldsymbol{\Omega}}_{NT} = \boldsymbol{ee}'$ (similar to how "robust" covariance matrices are calculated). As before, these sample covariance matrices are then averaged to obtain the "representative" error covariance matrix, $\overline{\boldsymbol{\Omega}}_{NT}$. $\overline{\boldsymbol{\Omega}}_{NT}$ becomes the "population" covariance matrix for the subsequent NT Monte Carlo experiment.

The right-hand side column of TABLE 1 reports the salient characteristics of the data for this generalized statistical environment. Of course, "*Mean* R^2" is the same as in the left-hand side column, since the first-stage of the data-generating process – which produces the residuals used to construct the sample covariance matrices – is identical (same original data, same residual generating functions). While the specific values differ, it is clear that the simulated data in this "generalized statistical environment" are likewise characterized by substantial degrees of serial correlation, groupwise heteroscedasticity, and cross-sectional correlation.

TABLE 3 reports the results from the Monte Carlo experiments. These results are somewhat different from those of TABLE 2. For example, it is no longer true that Parks and PCSE always underestimate coefficient standard errors. When these procedures are applied in a "generalized statistical environment," they can either under- or over-estimate coefficient standard errors. For example, for the income data, the "Level' values for Parks range from a low of *39.9* (*N=20, T=20*) to a high of *100* (several experiments). For the tax data, the corresponding range is *48.6* to *100*. The same is true for the PCSE estimates: For both income and tax data, the corresponding "Level" values lie on both sides of 95.

Nor is it necessarily the case that PCSE always produces more accurate hypothesis tests than Parks. For example, for the income data when *N=10* and *T=20*, Parks produces a marginally more accurate "Level" result than PCSE (*96.1* versus *97.6*). Similar examples can be found for the tax data. Indeed, were it not for a couple of egregious exceptions (*N=10/T=10* and *N=20/T=20*), one might even be led to conclude that Parks was as good, if not slightly better, than PCSE for hypothesis testing with these latter data. That being said, PCSE overall appears to estimate coefficient standard errors more accurately than Parks. However, PCSE can also be grossly inaccurate. For example, there are cases where PCSE estimated coefficient standard errors that are twice, or more, their true size.[10]

Turning now to "Efficiency," we see that it is still true that there are efficiency costs in using PCSE rather than Parks to estimate β_x . If anything, the efficiency costs are greater in the "generalized statistical environment." The average value of "Efficiency" over all sixteen experiments was *51.8* for the income data, and *76.3* for the tax data. The latter value would have been considerably lower were it not for one outlier case where PCSE was substantially more efficient than Parks ($N=5$, $T=25$). Both values are lower than their counterparts in TABLE 2. Further, it is no longer true that PCSE compares well with Parks on efficiency grounds when N and T are approximately equal. This is evidenced by both income and tax data (cf. $N=10/T=10$ and $N=20/T=20$).

The following summarizes our main findings from this analysis of the Parks and PCSE estimators within a "generalized" statistical environment:

1. In a "generalized" statistical environment, both Parks and PCSE can either under- or overestimate coefficient standard errors, so that we cannot sign the direction of the bias associated with using these techniques for hypothesis testing.

2. PCSE usually, but not always, produces more reliable standard error estimates than Parks. However, PCSE estimates can sometimes be highly unreliable.

3. Whenever PCSE provides a benefit in the form of more accurate standard error estimates, it comes at a cost of reduced efficiency.

---

[10] This was true for the income data when $N=5$ and $T=10$; and true for the tax data in the following cases: $N=5/T=15$; $N=10/T=20$; $N=10/T=25$; and $N=20/T=25$. While not reported in the text, we calculated a "Standard Error Ratio" consisting of the ratio of the average estimated standard error based on the associated covariance formula, over the sample standard deviation calculated from the 1000 values of $\hat{\beta}_x$. This is essentially the inverse of BK's "Overconfidence" measure.

We note that these findings remained valid when alternative, more fully specified residual-generating functions were used to construct our statistical environments.

IV. CONCLUSION

Time-series, cross-sectional (TSCS) data are extremely useful to researchers and have been widely employed in published research. However, the complex nature of the associated error structure can cause inaccurate estimates of coefficients and their standard errors. Beck and Katz (1995) study the properties of FGLS (Parks) and "OLS with Panel-Corrected Standard Errors" (PCSE) within a simulated statistical environment characterized by serial correlation, groupwise heteroscedasticity, and cross-sectional correlation. They find that Parks produces estimates of coefficient standard errors that are too small, and that the extent of this bias can be substantial. In contrast, PCSE produces accurate estimates of standard errors, at little to no cost in efficiency, except in extreme cases. Consequently, BK prescribe that researchers use the PCSE procedure when working with TSCS data

Our study revisits BK's comparison of the Parks and PCSE estimators. Our innovation is that we construct simulated statistical environments that are designed to closely match "real-world," TSCS data. We pattern our statistical environments after income and tax data on U.S. states from 1960-1999. For these data, we find that the benefits of PCSE are smaller, and the costs greater, than a reading of BK would suggest: While PCSE generally does a better job than Parks in estimating standard errors, it too can be unreliable, sometimes producing standard errors that are substantially off the mark. Further, we find that the benefits of PCSE can come at a substantial cost in estimator efficiency.

Based on our study, we would give the following advice to researchers using TSCS data: Given a choice between Parks and PCSE, we recommend that researchers use PCSE for hypothesis testing, and Parks if their primary interest is accurate coefficient estimates. We caution that our advice is predicated on the assumption that researchers' TSCS data resemble our simulated income and tax data. It would be valuable to supplement our findings with results from other simulated statistical environments patterned after actual TSCS data. That is a topic for future research.

REFERENCES

Alt, J. and R. Lowry. "Divided Government, Fiscal Institutions, and Budget Deficits: Evidence from the States." *American Political Science Review* Vol. 88, No. 4 (1994): 811-828.

______________ . "A Dynamic Model off State Budget Outcomes under Divided Partisan Government." *Journal of Politics* Vol. 62, No. 4 (2000): 1035-1069.

______________ . "Party Differences in State Budget Outcomes Are There After All: Response to 'Reexamining the Dynamic Model of Divided Partisan Government'." *Journal of Politics* Vol. 65, No. 2 (2003): 491-497.

Barrilleaux C., and M. Berkman. "Do Governors Matter? Budgeting Rules and the Politics of State Policymaking." *Political Research Quarterly* Vol. 56, No. 4 (2003): 409-417.

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.

Boehmke, F.J. "The Effect of Direct Democracy on the Size and Diversity of State Interest Group Populations." *Journal of Politics* Vol. 64, No. 3 (2002): 827-844.

Brace, P. "The Changing Context of State Political Economy." *Journal of Politics* Vol. 53, No. 2 (1991): 297-317.

Brace, P., Cohen, Y., Gray, V., and D. Lowry. "How Much Do Interest Groups Influence State Economic Growth?" *American Political Science Review* Vol. 83, No. 4 (1989): 1297-1301.

Brierly, A.B. and R. C. Feick. "Accounting for State Economic Growth: A Production Function Approach." *Political Research Quarterly* Vol. 46, No. 3 (1993): 657-669.

Crowley, J.E. and T. Skocpol. "The Rush to Organize: Explaining Associational Formation in the United States, 1860s-1920s." *American Journal of Political Science* Vol. 45, No. 4 (2001): 813-829.

Dye, T.R. "Taxing, Spending, and Economic Growth in the American States." *Journal of Politics* Vol. 42, No. 4 (1980): 1085-1107.

Dye, T.R. and R.C. Feiock. "State Income-Tax Adoption and Economic-Growth." *Social Science Quarterly* Vol. 76, No. 3 (1995): 648-654.

Fording, R.C. “The Political Response to Black Insurgency: A Critical Test of Competing Theories of the State.” *American Political Science Review* Vol. 95, No. 1 (2001): 115-130.

Helms, L. Jay. “The Effect of State and Local Taxes on Economic Growth: A Time Series-Cross Section Approach.” *The Review of Economics and Statistics* (1985): 574-82.

Hendrick, R.M., and J.C. Garand. “Variation in State Economic Growth: Decomposing State, Regional, and National Effects.” *Journal of Politics* Vol. 53, No. 4 (1991): 1093-1110.

Jones, B.D. “Public Policies and Economic-Growth in the American States.” *Journal of Politics* Vol. 52, No. 1 (1990): 219-233.

Kousser, T. “The Politics of Discretionary Medicaid Spending, 1980-1993.” *Journal of Health, Politics, Policy, and Law* Vol. 27, No. 4 (2002): 639-671.

McAtee, A., Yackee, S., and D. Lowery. “Reexamining the Dynamic Model of Divided Partisan Government.” *Journal of Politics* Vol. 65, No. 2 (2003): 477-490.

Nicholson-Crotty, S. “Goal Conflict and Fund Diversion in Federal Grants to the States.” *American Journal of Political Science* Vol. 48, No. 1 (2004): 110-122.

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

Reed, W. R. “Democrats, Republicans, and Taxes: Evidence that Political Parties Matter.” *Journal of Public Economics* (2005): Forthcoming.

StataCorp. *Stata Statistical Software: Release 7.0.* College Station, TX: Stata Corporation, 2001.

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

TABLE I
Summary of Diagnostics

	"PARKS-TYPE" STATISTICAL ENVIRONMENT	*GENERALIZED STATISTICAL ENVIRONMENT*
	Income Data	
Mean R^2	0.728	0.728
Mean $\hat{\rho}$	0.61	0.82
Mean $h = \left(\frac{\hat{\sigma}_{75th\ percentile}}{\hat{\sigma}_{25th\ percentile}} \right)$	1.24	1.31
Mean $r_{ij} = \left(\frac{\hat{\sigma}_{\varepsilon,ij}}{\sqrt{\hat{\sigma}_{\varepsilon,ii}}\sqrt{\hat{\sigma}_{\varepsilon,jj}}} \right)$	0.74	0.64
	Tax Data	
Mean R^2	0.701	0.701
Mean $\hat{\rho}$	0.58	0.64
Mean $h = \left(\frac{\hat{\sigma}_{75th\ percentile}}{\hat{\sigma}_{25th\ percentile}} \right)$	1.59	1.55
Mean $r_{ij} = \left(\frac{\hat{\sigma}_{\varepsilon,ij}}{\sqrt{\hat{\sigma}_{\varepsilon,ii}}\sqrt{\hat{\sigma}_{\varepsilon,jj}}} \right)$	0.36	0.41

NOTE: Means are calculated over all replications (1000 replications per experiment) and experiments (a total of 16 experiments based on 16 possible *N* and *T* combinations).

TABLE II
Performance of Parks and PCSE Estimators in a "Parks-type" Statistical Environment

		A. Income Data							
		Level				*Efficiency*			
		T=10	*T=15*	*T=20*	*T=25*	*T=10*	*T=15*	*T=20*	*T=25*
N=5	*Parks* *PCSE*	62.1 72.0	65.2 72.6	67.1 77.5	67.2 76.9	101.4	98.9	79.6	58.8
N=10	*Parks* *PCSE*	29.2 79.6	49.3 75.3	51.1 75.3	49.4 81.6	98.2	94.1	84.8	61.6
N=20	*Parks* *PCSE*	---- 82.5	---- 79.3	8.8 78.8	11.2 86.9	----	----	98.0	83.6
N=48	*Parks* *PCSE*	---- 82.1	---- 78.7	---- 81.9	---- 88.0	----	----	----	----
MEAN	*Parks* *PCSE*	46.1 79.3				85.9			

TABLE II: Continued
Performance of Parks and PCSE Estimators in a "Parks-type" Statistical Environment

		B. Tax Data							
		Level				*Efficiency*			
		T=10	*T=15*	*T=20*	*T=25*	*T=10*	*T=15*	*T=20*	*T=25*
N=5	***Parks***	66.1	70.2	76.5	77.6	84.9	77.8	72.2	71.7
	PCSE	83.3	82.0	84.6	86.1				
N=10	***Parks***	23.4	52.3	61.0	65.3	96.8	84.0	76.0	72.8
	PCSE	86.1	85.3	87.1	86.0				
N=20	***Parks***	----	----	8.4	23.2	----	----	97.4	85.7
	PCSE	87.6	89.9	87.4	86.6				
N=48	***Parks***	----	----	----	----	----	----	----	----
	PCSE	86.5	88.0	90.0	89.8				
MEAN	***Parks***	52.4				81.9			
	PCSE	86.6							

NOTE: "Level" and "Efficiency" are defined in the text.(cf. Section IIA). "Mean" refers to the average value over all replications (1000 replications per experiment) and all experiments. For the PCSE "Level" estimates, there are a total of 16 experiments. For the Parks "Level" estimates and the "Efficiency" estimates, there are only 10 experiments, because Parks cannot be calculated when $N < T$.

TABLE III
Performance of Parks and PCSE Estimators in a Generalized Statistical Environment

		A. Income Data							
		Level				***Efficiency***			
		T=10	***T=15***	***T=20***	***T=25***	***T=10***	***T=15***	***T=20***	***T=25***
N=5	***Parks***	100	100	99.9	99.8	95.2	48.5	30.6	21.6
	PCSE	100	100	83.7	98.8				
N=10	***Parks***	81.7	100	96.1	93.0	49.8	28.6	49.0	52.9
	PCSE	100	99.4	97.6	97.5				
N=20	***Parks***	----	----	39.9	59.7	----	----	77.6	64.7
	PCSE	96.0	94.2	98.9	96.2				
N=48	***Parks***	----	----	----	----	----	----	----	----
	PCSE	94.2	91.0	99.7	87.5				
MEAN	***Parks***	87.0				51.8			
	PCSE	95.9							

TABLE III: Continued
Performance of FGLS and PCSE Estimators in a Generalized Statistical Environment

		B. Tax Data							
		Level				***Efficiency***			
		T=10	***T=15***	***T=20***	***T=25***	***T=10***	***T=15***	***T=20***	***T=25***
N=5	***Parks*** ***PCSE***	100 100	100 100	99.6 100	99.7 100	38.5	74.2	100.3	176.4
N=10	***Parks*** ***PCSE***	72.9 100	99.4 99.9	99.8 100	99.7 100	60.4	37.5	65.9	81.2
N=20	***Parks*** ***PCSE***	---- 98.9	---- 99.8	48.6 100	90.9 100	----	----	72.3	56.4
N=48	***Parks*** ***PCSE***	---- 98.8	---- 100	---- 100	---- 100	----	----	----	----
MEAN	***Parks*** ***PCSE***	91.1 99.8				76.3			

NOTE: "Level" and "Efficiency" are defined in the text.(cf. Section IIA). "Mean" refers to the average value over all replications (1000 replications per experiment) and all experiments. For the PCSE "Level" estimates, there are a total of 16 experiments. For the Parks "Level" estimates and the "Efficiency" estimates, there are only 10 experiments, because Parks cannot be calculated when $N < T$.