



What to do (and not to do) with Time-Series Cross-Section Data

Nathaniel Beck, Jonathan N. Katz

The American Political Science Review, Volume 89, Issue 3 (Sep., 1995), 634-647.

Your use of the JSTOR database indicates your acceptance of JSTOR's Terms and Conditions of Use. A copy of JSTOR's Terms and Conditions of Use is available at <http://www.jstor.org/about/terms.html>, by contacting JSTOR at jstor-info@umich.edu, or by calling JSTOR at (888)388-3574, (734)998-9101 or (FAX) (734)998-9113. No part of a JSTOR transmission may be copied, downloaded, stored, further transmitted, transferred, distributed, altered, or otherwise used, in any form or by any means, except: (1) one stored electronic and one paper copy of any article solely for your personal, non-commercial use, or (2) with prior written permission of JSTOR and the publisher of the article or other text.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

The American Political Science Review is published by American Political Science Association. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/apsa.html>.

The American Political Science Review
©1995 American Political Science Association

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact jstor-info@umich.edu.

©2001 JSTOR

WHAT TO DO (AND NOT TO DO) WITH TIME-SERIES CROSS-SECTION DATA

NATHANIEL BECK *University of California, San Diego*
 JONATHAN N. KATZ *California Institute of Technology*

We examine some issues in the estimation of time-series cross-section models, calling into question the conclusions of many published studies, particularly in the field of comparative political economy. We show that the generalized least squares approach of Parks produces standard errors that lead to extreme overconfidence, often underestimating variability by 50% or more. We also provide an alternative estimator of the standard errors that is correct when the error structures show complications found in this type of model. Monte Carlo analysis shows that these "panel-corrected standard errors" perform well. The utility of our approach is demonstrated via a reanalysis of one "social democratic corporatist" model.

We shall show that a commonly used technique for the analysis of time-series cross-section (TSCS) data produces incorrect results. Our result either invalidates or calls into question the findings of at least five articles published in the *American Political Science Review*, as well as a like number in other leading journals in political science and sociology. Table 1 provides an incomplete list of relevant articles whose conclusions are based on the use of this problematic technique. All of these articles use an application of the generalized least squares (GLS) method first described by Parks (1967), a method designed to deal with some common problems that occur in TSCS data. We show that the Parks method produces dramatically inaccurate standard errors when used for the type of data commonly analyzed by students of comparative politics. We then offer a new method that is both easier to implement and produces accurate standard errors.

Time-series cross-section data are characterized by having repeated observations on fixed units, such as states or nations. The number of units analyzed would typically range from about 10 to 100, with each unit observed over a relatively long time period (often 20 to 50 years). Both the temporal and spatial properties of TSCS data make the use of ordinary least squares (OLS) problematic. In particular, models for TSCS data often allow for temporally and spatially correlated errors, as well as for heteroscedasticity. Parks proposed a method for dealing with these problems based on GLS.¹ The use of this method can lead to dramatic underestimates of parameter variability in common research situations.

Why the severe problems with the Parks method? Is it not just an application of well-known GLS? While GLS has optimal properties for TSCS data, it assumes that we have knowledge about the error process that, in practice, we never have. Thus analysts use not GLS, but feasible generalized least squares (FGLS). It is "feasible" because it uses an estimate of the error process, avoiding the GLS assumption that the error process is known. The FGLS formula for standard

errors, however, assumes that the error process is known, not estimated. In many applications this is not a problem because the error process has few enough parameters that they can be well estimated. Such is not the case for TSCS models, where the error process has a large number of parameters. This oversight causes estimates of the standard errors of the estimated coefficients to understate their true variability. We provide a measure of how much the Parks standard errors understate true sampling variability, that is, how much the Parks method falsely inflates confidence in the findings of TSCS studies.

Unfortunately, it is not possible to provide analytic formulae for the degree of overconfidence introduced by the Parks method. Instead, we provide evidence from Monte Carlo experiments using simulated data to assess the performance of the various estimators. This evidence clearly shows the overconfidence induced by the Parks method. The Parks estimator may understate variability by between 50% and 300% in practical research situations. It is this extreme overconfidence that leads us either to overturn or to cast doubt on the findings of many analyses based on the Parks method.

Having demonstrated the problems of the Parks method, we instead advocate a simpler method for estimating TSCS models. It is well known that even though OLS estimates of TSCS model parameters may not be optimal, they often perform well in practical research situations. It is also well known that the OLS estimates of standard errors may be highly inaccurate in such situations. We therefore propose to retain OLS parameter estimates but replace the OLS standard errors with *panel-corrected standard errors*. Monte Carlo analysis shows that these new estimates of sampling variability are very accurate, even in the presence of complicated panel error structures.

We shall detail the problems of the Parks method, laying out the structure of TSCS models and showing why OLS is problematic. In order to understand Parks' solution and why it is problematic, it is neces-

TABLE 1**Articles Using the Parks Method**

CITATION	N ^a	T ^b	MODEL
Number of Units Is Less Than Number of Time Points			
Hicks & Swank 1992	18	23	Political determinants of welfare spending in advanced industrial societies
Hicks 1994b ^c	17	18	Political and union determinants of economic growth in advanced industrial societies
Pampel 1993 ^c	18	36	Political and social determinants of fertility rates in advanced industrial society
Schneider & Ingraham 1984	7, 18	59	Political determinants of social program expansion in advanced industrial societies
Su, Kamlet & Mowery 1993 ^d	5, 7, 10, 11	26	Political determinants of U.S. budget by category
Swank 1992	16	20	Political determinants of tax policy in advanced industrial societies
Number of Units Exceeds Number of Time Points ^e			
Alvarez, Garrett & Lange 1991	16	15	Political and union determinants of economic growth in advanced industrial societies
Friedland & Sanders 1985 ^c	12	6	Political determinants of economic growth in advanced industrial societies
Giles and Hertz 1994 ^{c,f}	64	15	Party registration and race in Louisiana parishes
Lin 1994 ^c	20, 42, 62	26	Influence of government spending on cross-national economic growth
Pampel & Williamson 1988 ^c	18	7	Political determinants of welfare spending in advanced industrial societies
Scholz, Twombly & Headrick 1991	30	10	Political determinants of enforcement of regulatory standards in New York counties
Scholz & Wei 1986	50	8	Regulatory enforcement in the states
Wood 1992	50	9	Policy implementation in the states

^aNumber of cross-sectional units.^bNumber of time periods.^cUsed Parks in addition to other methods; estimates using other methods not discussed in this article.^dEstimated with seemingly unrelated regressions; same as Parks correction for contemporaneously correlated errors.^eParks method cannot produce results.^fArticle indicates use of Park. Private communication with the authors indicates that Park was not in fact used.

sary first to consider some properties of FGLS. The Parks method is then laid out. This method provides an estimation strategy that takes into account both cross-sectional and temporal complications of the data. These two components of the method are presented and assessed separately. These assessments show that the Parks standard errors are likely to lead to extreme overconfidence for typical TSCS data, with the correction for cross-sectional complications being much more problematic than the correction for temporal complications.

We shall then present our proposed method for estimating TSCS models. We argue that this method, which combines ordinary least squares parameter estimates with panel-corrected standard errors, *should* perform well. As with the Parks method, a full assessment of this method for the types of data encountered in research situations is only possible via Monte Carlo analysis.

We then present this Monte Carlo evidence. Our results demonstrate the extreme overconfidence induced by the Parks standard errors. Our simulations show that while the Parks correction for cross-sectional complication causes much of the problem, the correction for temporal complications is also problem-

atic. In addition, the Monte Carlo evidence shows that panel-corrected standard errors perform extremely well, even in the presence of complicated panel error structures. The Monte Carlo evidence also shows that OLS parameter estimates are themselves, at worst, not much inferior to the Parks parameter estimates. Thus the costs of the inaccurate Parks standard errors are in no sense paid for by the superiority of the Parks estimator of the model parameters.

Finally, we use our proposed method to reanalyze Hicks and Swank's (1992) results obtained with the Parks method. We show that the strengths of many of their findings about the political causes of government spending in advanced industrial societies are an artifact of their use of the Parks method. We also briefly reconsider the findings of other studies that used that method. The conclusion presents a unified method for analyzing TSCS data.²

THE PARKS METHOD AND ITS FLAWS

Our analysis is limited to what Stimson (1985) called temporally dominated TSCS models, where a limited

number of units are observed for a relatively long period of time.³ The critical assumption of TSCS models is that of "pooling"; that is, all units are characterized by the same regression equation at all points in time. Given this assumption, we can write the generic TSCS model as

$$y_{i,t} = x_{i,t}\beta + \epsilon_{i,t}; i = 1, \dots, N; t = 1, \dots, T \quad (1)$$

where $x_{i,t}$ is a vector of one or more (k) exogenous variables and observations are indexed by both unit (i) and time (t).⁴ We shall denote the matrix of independent variables for all observations as X and the vector of observations on the dependent variable as Y . We assume that the data are stacked by unit.⁵ We denote the $NT \times NT$ covariance matrix of the errors with typical element $E(\epsilon_{i,t}\epsilon_{j,s})$ by Ω . TSCS models can be difficult to estimate because the error process of such models may be more complicated than is typical of either time-series or cross-sectional models. Different assumptions about this error process lead to different preferred methods of estimation.

Ordinary Least Squares Is Problematic for Time-Series Cross-Section Data

Ordinary least squares is optimal (best linear unbiased) for TSCS models if the errors are assumed to be generated in an uncomplicated ("spherical") manner. In particular, for OLS to be optimal it is necessary to assume that all the error processes have the same variance (homoscedasticity) and that all of the error processes are independent of each other. The latter assumption can be broken down into the assumption that errors for a particular unit at one time are unrelated to errors for that unit at all other times (no serial correlation) and that errors for one unit are unrelated to the errors for every other unit (no spatial correlation). Under these assumptions TSCS models should be estimated by OLS and OLS standard errors are correct. Most analysts, however, are not willing to accept the assumption of spherical errors for TSCS models.

Ordinary least squares is not optimal in the presence of nonspherical errors, in the sense that there will be other estimators that make more efficient use of the data. More seriously, if the errors are not spherical, there is no guarantee that the OLS standard errors will be correct. We use the term *correct standard errors* to indicate that we have accurate estimates of the variability of parameter estimates. Correct standard errors allow for the correct computation of confidence intervals and statistical tests. Incorrect standard errors will lead us to be either too confident or insufficiently confident about whether our findings might merely be statistical artifacts.

It is, of course, always possible that the errors of any regression model may be nonspherical. The problem is, however, much more acute for TSCS models. In particular, we might expect TSCS errors to be contemporaneously correlated in that large errors for unit i at time t will often be associated with large

errors for unit j at time t . This is likely in the cross-national context, where the economies of, say, the Netherlands and Belgium are linked. It is also a likely problem in other TSCS contexts, such as the study of disaggregated budgets, where large errors in one budget category may be associated with large errors in other categories in the same year. These contemporaneous correlations may differ by unit. For example, the errors in the Scandinavian economies may be linked together but remain independent of errors in the southern European countries.

We might also expect the errors in TSCS models to show "panel heteroscedasticity," where the variances of the error process differ from unit to unit. The errors of a cross-national panel study, for example, may show panel heteroscedasticity because the scale of the dependent variable, such as the level of government spending, may differ between countries. The assumption of panel heteroscedasticity is more stringent than just cross-sectional heteroscedasticity because we continue to assume that the error variances within each unit do not differ over time; this assumed structure allows for certain estimation strategies not available in the nonpanel case.

Finally, it is possible that the errors may show temporal dependence. The most typical assumption is that the errors show first-order serial correlation. Some analysts assume that the degree of serial correlation differs from unit to unit, while others assume it is constant across units.

Time-series cross-section analysts do put some structure on the assumed error process. In particular, they assume that for any given unit, the error variance is constant, so that the only source of heteroscedasticity is differing error variances across units. Analysts also assume that all spatial correlation both is contemporary and does not vary with time. The temporal dependence exhibited by the errors is also assumed to be time-invariant and may also be invariant across units. All of these assumptions allow analysts to attempt to improve on OLS for TSCS data. Since these assumptions are all based on the panel nature of the data, we call them the *panel error assumptions*.⁶

Feasible Generalized Least Squares

Equation 1 can be estimated by generalized least squares regardless of any complexities of the error process, so long as the covariance matrix of those errors, Ω , is known (up to a scale factor). Given that assumption, GLS is fully efficient and yields consistent estimates of the standard errors (see, e.g., Kmenta 1986, 609–16). Generalized least squares works by transforming equation 1 with a general error covariance matrix to another linear equation where the error covariance matrix is suitable for OLS estimation (spherical). The GLS estimates of β are given by

$$(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y \quad (2)$$

with estimated covariance matrix

$$(\mathbf{X}'\hat{\mathbf{\Omega}}^{-1}\mathbf{X})^{-1}. \quad (3)$$

The problem is that the covariance matrix of the errors, $\mathbf{\Omega}$, is never known in practice (even up to a scale factor). Thus an estimate of $\mathbf{\Omega}$, $\hat{\mathbf{\Omega}}$, is used in expressions 2 and 3. This procedure, FGLS, provides consistent estimates of β if $\hat{\mathbf{\Omega}}$ is estimated by residuals computed from consistent estimates of β . Ordinary least squares provides such consistent estimates. We denote the FGLS estimates of β by $\hat{\beta}$. The application of FGLS to TSCS models with panel errors was first described by Parks (1967).

Feasible generalized least squares performs well in large samples. In the limit, it is equivalent to full maximum likelihood, and so has all the optimal asymptotic properties of maximum likelihood (Cramer 1986, 79–128). We know little about the finite sample properties of FGLS other than that it yields unbiased estimators under very general conditions that are usually met in practice (Kakwani 1967). The better the estimate of $\mathbf{\Omega}$, of course, the better the FGLS estimator; obviously, estimation of $\mathbf{\Omega}$ will improve as NT increases in relationship to the number of parameters in $\mathbf{\Omega}$ that must be estimated.

It is difficult to assess the performance of FGLS in finite samples. There are by now many Monte Carlo studies showing that FGLS may be less efficient than its OLS counterpart, especially in very small samples. Our interest is in how well FGLS estimates of variability (expression 3) perform in finite samples. It is known that FGLS standard errors underestimate true variability, at least for normal errors (Freedman and Peters 1984). There are, unfortunately, no analytic results that indicate whether this underestimate is of importance to applied researchers, nor, in particular, are there any analytic results about the performance of the Parks estimates of variability for TSCS models. While we will assess this variability with Monte Carlo experiments, we can get some hint about this variability by closer examination of the Park method.

The Parks Method

The Parks method is FGLS for TSCS models where the errors show panel heteroscedasticity, contemporaneous correlation, and unit specific serial correlation. The correction for contemporaneous correlation of the errors automatically corrects for any panel heteroscedasticity, so we need only consider the corrections for contemporaneous and serial correlation of the errors here.⁷

The Parks method consists of two sequential FGLS transformations, first eliminating serial correlation of the errors then eliminating contemporaneous correlation of the errors. This is done by initially estimating equation 1 by OLS. The residuals from this estimation are used to estimate the unit-specific serial correlation of the errors, which are then used to transform the model into one with serially independent errors.⁸ Residuals from this estimation are then used to estimate the contemporaneous correlation of

the errors, and the data is once again transformed to allow for OLS estimation with now spherical errors.⁹ We can therefore consider the consequences of the two corrections separately.

Correcting for Contemporaneous Correlation of the Errors. We first consider the Parks correction for contemporaneously correlated errors. The TSCS model with contemporaneously correlated errors is then exactly expression 1 with the variance covariance matrix of the errors having zeros for all noncontemporaneous observations and free parameters allowing for contemporaneous pairwise correlation of the errors and panel heteroscedasticity. We can write this compactly as $\mathbf{\Omega} = \mathbf{\Sigma} \otimes \mathbf{I}_T$, where $\mathbf{\Sigma}$ is the $N \times N$ matrix of contemporaneous covariances, with typical element $E_t(\epsilon_{i,t}\epsilon_{j,t})$.¹⁰ While these parameters differ among pairs of units, they do not vary by time. Feasible generalized least squares, therefore, requires estimating all the pairwise contemporaneous covariances. The matrix of all these estimates is denoted $\hat{\mathbf{\Sigma}}$. There are $N \times (N + 1)/2$ contemporaneous covariances; each of these is estimated using NT observations.

The Parks correction for contemporaneously correlated errors cannot be used unless T is at least as big as N (Beck et al. 1993).¹¹ But even when T is greater than N , so that FGLS can be used, estimation of standard errors is problematic unless T is considerably larger than N . Each element of the matrix of contemporaneous covariances of the errors is estimated using, on average, $2T/N$ observations. Many cross-national 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! Studies on the political economy of advanced industrial nations seldom have T to N ratios that exceed 3; thus the elements of the covariance matrix of the errors are estimated with, on average, six observations. Theory does not tell us how inaccurate the Parks method is in these cases, but we should be prepared to see highly overconfident Parks standard errors in the typical cross-national panel case. We shall provide Monte Carlo evidence of this.

Correcting for Serially Correlated Errors. The Parks correction for serially correlated errors assumes the errors follow a unit-specific first-order autoregressive (AR1) process

$$\epsilon_{i,t} = \rho_i \epsilon_{i,t-1} + \nu_{i,t} \quad (4)$$

where the $\nu_{i,t}$ s are (mean zero) variables independently distributed across time. Some analysts impose the additional assumption that the ρ_i are homogeneous across units, that is, $\rho_i = \rho$. Ordinary least squares residuals are used to estimate either the common ρ or the ρ_i ; this estimate is then used to transform the data, using the well-known Prais-Winsten transformation (see, e.g., Kmenta 1986, 304).

The FGLS correction for a single ρ requires estimating one extra, unaccounted-for parameter. This is unlikely to cause FGLS standard errors to estimate variability inaccurately in the typical cross-national

panel situation. The FGLS correction for unit-specific serially correlated errors, used by Parks, is likely to cause more serious underestimates of variability. The essence of the problem is that each ρ_i is estimated using an autoregression based on only T observations. It is well known that such estimates are biased downward (Hurwicz 1950). As a consequence, the Parks estimates, which correct based on these inaccurate autoregressions, may be inferior to OLS estimates. The underestimates of the ρ_i , when combined with trending data, can cause the Parks estimates of standard errors to misestimate variability substantially.

The assumption of unit-specific serial correlations also seems odd at a theoretical level. Time-series cross-section analysts assume that the "interesting" parameters of the model, β , do not vary across units; this assumption of pooling is at the heart of TSCS analysis. Why should we expect the "nuisance" ρ to not show similar pooling? ρ can be interpreted as how long it takes for prior shocks to be removed from the system. Why should this "memory" be the only model parameter that varies from unit to unit?

The choice whether to correct for serially correlated errors assuming either heterogeneous or homogeneous ρ depends on the small sample properties of the two types of estimators. While we would expect the unit-specific serial correlation correction to lead to more inaccurate estimates of variability, it is also possible that allowing for variation among the ρ_i might improve overall estimation. We can only assess the small sample performance of the two corrections for serial correlation through Monte Carlo experimentation. Before looking at the results of those experiments, we first consider a new method for estimating the variability of OLS estimators. We can then compare the performance of the Parks estimator with our new method.

ORDINARY LEAST SQUARES WITH PANEL-CORRECTED STANDARD ERRORS

If the errors in equation 1 meet one or more of the panel error assumptions, then OLS estimates of β will be consistent but inefficient; the degree of inefficiency depends on the data and the exact form of the error process. The OLS standard errors will also be inaccurate,¹² but they can be corrected so that they provide accurate estimates of the variability of the OLS estimates of β . This correction takes into account the contemporaneous correlation of the errors (and perforce heteroscedasticity). Any serial correlation of the errors must be eliminated before the panel-corrected standard errors are calculated. The correction for contemporaneous correlation of the errors is only possible because we have repeated information on the contemporaneous correlation of the errors; our proposed method does not work outside the TSCS context.

The correct formula for the sampling variability of

the OLS estimates is given by the square roots of the diagonal terms of

$$\text{Cov}(\hat{\beta}) = (X'X)^{-1}\{X'\Omega X\}(X'X)^{-1}. \quad (5)$$

If the errors obey the spherical assumption, this simplifies to the usual OLS formula, where the OLS standard errors are the square roots of the diagonal terms of $\hat{\sigma}^2 (X'X)^{-1}$, where $\hat{\sigma}^2$ is the usual OLS estimator of the common error variance, σ^2 . If the errors obey the panel structure, then this formula provides incorrect standard errors. Expression 5, however, can still be used, in combination with that panel structure of the errors, to provide accurate, *panel-corrected standard errors* (PCSEs).¹³

For panel models with contemporaneously correlated and panel heteroscedastic errors, Ω is an $NT \times NT$ block diagonal matrix with an $N \times N$ matrix of contemporaneous covariances, Σ , along the diagonal. To estimate equation 5, we need an estimate of Σ . Since the OLS estimates of expression 1 are consistent, we can use the OLS residuals from that estimation to provide a consistent estimate of Σ . Let $e_{i,t}$ be the OLS residual for unit i at time t . We can estimate a typical element of Σ by

$$\hat{\Sigma}_{i,j} = \frac{\sum_{t=1}^T e_{i,t}e_{j,t}}{T},$$

with the estimate $\hat{\Sigma}$ being comprised of all these elements.¹⁴ We then use this to form the estimator $\hat{\Omega}$ by creating a block diagonal matrix with the $\hat{\Sigma}$ matrices along the diagonal.¹⁵ As the number of time points increases, $\hat{\Sigma}$ becomes an increasingly better estimator of Σ . We cannot, however, assess the finite sample performance of PCSEs by analytic methods. Thus we shall have to evaluate them with the same Monte Carlo experiments we use to evaluate Parks.

MONTE CARLO ANALYSIS

We have argued that the Parks method may not perform well in correcting for a variety of TSCS complications and, in particular, may lead to substantial underestimates of variability in finite samples. We have also argued that using OLS with PCSEs is a reasonable estimation strategy. We designed a series of Monte Carlo experiments to assess Parks, PCSEs, and OLS in the TSCS context. As argued, we can consider the two components of the Parks GLS "correction" separately.

Design of the Experiments

All experiments used simulated data that were generated to mimic some property of TSCS data. The setup of each simulation was similar. For a given N and T , observations on an independent variable, $x_{i,t}$, ($i = 1, \dots, N$; $t = 1, \dots, T$), were generated as random draws from a zero-mean normal distribution. Experiments were run with various combination of N

and T chosen to reflect values typically found in cross-national panel models (see Table 1).

For each experiment, we generated, using GAUSS, one thousand replicates of the NT error terms ($\epsilon_{i,t}^{(l)}$, $i = 1, \dots, N$; $t = 1, \dots, T$; $l = 1, \dots, 1000$) according to the model being studied. (A superscript (l) denotes a specific replicate.) The errors were always generated as zero-mean NT -variate normals, with standard deviations chosen so that estimated coefficients were roughly twice their standard errors. We chose the covariance structure of the errors to mimic the property of TSCS models being examined in a given set of experiments; this structure is discussed in the context of each set of experiments.

The $x_{i,t}$ were fixed over these one thousand replications. The one thousand replicates of $\epsilon_{i,t}$ —and the fixed $x_{i,t}$ —were used to generate one thousand replicates of $y_{i,t}$, using

$$y_{i,t}^{(l)} = \beta_0 + \beta_1 x_{i,t} + \epsilon_{i,t}^{(l)}; i = 1, \dots, N; \\ t = 1, \dots, T; l = 1, \dots, 1000$$

where both β_0 and β_1 were fixed at 10. We report only statistics concerning the estimation of β_1 .

The OLS estimate of β_1 for replication l is referred to as $\hat{\beta}^{(l)}$; the Parks estimate for that replication is $\tilde{\beta}^{(l)}$. We are concerned with the performance of the estimated standard errors. An accurate measure of the sampling variability of each estimator is the standard deviation of the one thousand $\tilde{\beta}^{(l)}$ s or $\hat{\beta}^{(l)}$ s. The quality of the OLS or Parks estimates of variability can then be assessed by comparing the root mean square average of the one thousand estimated standard errors with the corresponding standard deviation of the one thousand estimates. The measure of accuracy we focus on, *overconfidence*, is the percentage by which, say, FGLS understates variability; that is

$$\text{Overconfidence} = 100 \frac{\sqrt{\sum_{l=1}^{1,000} (\tilde{\beta}^{(l)} - \bar{\beta})^2}}{\sqrt{\sum_{l=1}^{1,000} (\text{s.e.}(\tilde{\beta}^{(l)}))^2}}.$$

Overconfidence of 200%, for example, indicates that the true sampling variability of an estimator is, on average, twice the reported estimate of that variability.

An alternate measure of the accuracy of standard errors is the true “level” of reported “95% confidence intervals.” The “95% confidence interval” for each of the one thousand replications was computed using the estimated standard errors. We can then see how often these intervals contained the true value of β . Insofar as this proportion is under 95%, estimated variability leads to overconfidence.

We were also interested in the relative efficiency of Parks and OLS. Since the true value of β_1 is known, the root mean square error of the OLS and Parks estimates of β_1 can be calculated. The relative efficiency of OLS as compared to Parks is given by

$$\text{Efficiency} = 100 \frac{\sqrt{\sum_{l=1}^{1,000} (\tilde{\beta}^{(l)} - \beta_1)^2}}{\sqrt{\sum_{l=1}^{1,000} (\hat{\beta}^{(l)} - \beta_1)^2}}.$$

Efficiency greater than 100% indicates that OLS is superior, in mean square error terms, to Parks.

Overconfidence of the Parks Correction

Correcting for Contemporaneously Correlated Errors. The first set of experiments was designed to evaluate the overconfidence of standard errors produced by the Parks FGLS transformation to eliminate contemporaneous correlation of the errors. This overconfidence is simply a function of N and T and does not depend on the actual contemporaneous correlation of the errors. We therefore generated the data as simply as possible, with the errors drawn as independent normals. The errors, in other words, were generated so as to appear spherical.¹⁶

The striking results of this experiment are reported in Table 2. In the worst cases, where T is exactly N , Parks is overconfident by 400% for $N = T = 10$ and by 600% for $N = T = 20$. Reported “95% confidence intervals” contain the true value of β only 30% of the time when $N = T = 10$; this falls to 20% when $N = T = 20$.

The Parks-induced overconfidence remains a serious problem as T grows in relation to N . Even in the most favorable case for Parks ($N = 10$, $T = 40$), its standard errors are 30% overconfident and reported 95% confidence intervals contain the true β only 87% of the time. In what is perhaps the most typical configuration, with $N = 15$ and $T = 30$, FGLS errors are more than 50% overconfident and reported 95% confidence intervals contain the true value of β only about 80% of the time. Readers confronted with analysis using the Parks correction for contemporaneous correlation of the errors in such a situation should substitute 2.6 for the conventional critical t -ratio of 1.68. These results show that TSCS researchers should simply avoid the Parks correction for contemporaneously correlated errors unless the ratio of the T to N is well above three, a situation not normally seen.

Correcting for Serially Correlated Errors. The Parks method first eliminates serial correlation of the errors by transforming the data, allowing for a separate transformation for each unit. We have argued that this may induce substantial overconfidence, a problem that could be eliminated by assuming that the serial correlation of the errors followed an identical process for all units. In order to test this claim of overconfidence, as well as to compare the relative efficiency of unit-specific versus common error processes, we needed to design a more complex set of experiments. The added complexity is necessary because the performance of the estimators varies with the degree of trend in the data.

The experiments generated data where the simulated errors were generated as AR1 processes with

TABLE 2

Overconfidence of Parks Correction for Contemporaneously Correlated Errors

N ^a	T ^b	OVERCONFIDENCE	LEVEL ^d
		(%) ^c	
10	10	408	31
	15	186	70
	20	152	78
	30	131	87
	40	130	87
15	15	529	24
	20	213	63
	30	156	79
	40	138	84
20	20	631	21
	30	187	70
	40	153	81

^aNumber of cross-sectional units.^bNumber of time points.

$$^c\text{Overconfidence} = 100 \frac{\sqrt{\sum_{i=1}^{1,000} (\hat{\rho}_i^{(l)} - \bar{\rho})^2}}{\sqrt{\sum_{i=1}^{1,000} (\text{s.e.}(\hat{\rho}_i^{(l)}))^2}}$$

^dPercentage of 95% confidence intervals containing β .

unit-specific ρ_i . More precisely, the errors were generated according to expression 4, where the ρ_i were set to ρ_1 for the first half of the units and ρ_2 for the second half; the separation between these two values was varied experimentally. The degree of trend in the independent variable was also experimentally manipulated. Thus the independent variables were generated according to $x_{i,t} = \delta x_{i,t-1} + \mu_{i,t}$, where the $\mu_{i,t}$ are zero-mean, independently distributed standard normal variates. Values of δ near 1 indicate strongly trending data. The experiments then proceeded as before.

Results of the experiments are presented in Table 3. Experiments with different values of N showed that the relative performance of the two estimators was sensitive only to the value of T ; we therefore only show the results for $N = 15$. In addition, since the results so strongly favor estimation with a common ρ , we present only the results least favorable to that method.

The experiments show that standard errors assuming a common ρ were never far off. Even in the worst cases, where half the units had $\rho = .9$ and half had $\rho = .3$, common ρ standard errors were never more than 20% too low; more typically, the common ρ estimator of variability was within 5% of the true value.¹⁷

The situation was very different for the unit-specific ρ_i estimator. The overconfidence of this estimator can be substantial; this overconfidence varies with T , δ and ρ . The most important determinant of overconfidence is δ , that is, the degree of trend in the independent variable. When $\delta = .9$, the Parks stan-

dard errors understate variability by almost 100% for T less than 30 and by about 30% for T above 30. Overconfidence caused by correcting for unit-specific serial correlation disappears only when the data trend mildly and T is at least 30.

Efficiency considerations clinch the case for estimation with a common ρ . When T is under 20, estimation assuming a common ρ is always more efficient than estimation assuming a varying ρ , even when the errors were generated with two very different ρ_i . Even when T gets as large as 40, estimation assuming a common ρ remains more efficient than estimation assuming a varying ρ , unless the errors were generated with two extremely different ρ_i . Even in this extreme case, estimation assuming varying ρ_i is less than 20% more efficient than estimation assuming a common ρ . Such extreme cases are unlikely to arise in practice. These experiments clearly show that TSCS analysts should correct for serially correlated errors assuming a common ρ .¹⁸

Thus, while the Parks (unit-specific) correction for serial correlation causes fewer problems in terms of overconfidence than does the Parks correction for contemporaneously correlated errors, it does induce substantial overconfidence. Where the data trend strongly, as they do in many political economy studies, this overconfidence can be as much as 100%.

Ordinary Least Squares with Panel-corrected Standard Errors

The overconfidence induced by the Parks standard errors makes it unusable except in the rarest of research situations. We proposed a simpler estimator, OLS with PCSEs. How does the proposed estimator perform in the Monte Carlo experiments? We first assess the accuracy of PCSEs and then compare the efficiency of OLS estimators with those produced by the Parks method.¹⁹

The Accuracy of Panel-corrected Standard Errors. OLS standard errors are accurate in the presence of either panel heteroscedasticity or contemporaneous correlation of the errors if the terms in the error covariance matrix, Ω , are not related to the squares and cross products of the independent variables. Since we wished to study the performance of PCSEs when OLS standard errors were incorrect, we designed experiments with contemporaneously correlated and/or panel-heteroscedastic error structures that were related to the panel structure of the independent variables.

For each value of t , the N -vector $x_{i,t}$ ($i = 1, \dots, N$) was generated as a draw from a zero-mean N -variate normal distribution. Varying degrees of heteroscedasticity were simulated by setting the variance of the first half of the units to 1 while the variance of the second half of the units was experimentally manipulated. The covariance matrix of this multivariate distribution was constructed so that all pairs of units were equally correlated, with the degree of correla-

TABLE 3

Comparison of Feasible Generalized Least Squares Corrections for Common and Varying ρ

T ^a	δ^b	ρ_1^c	ρ_2^d	OVERCONFIDENCE (%) ^e		EFFICIENCY (%) ^f
				SINGLE ρ	VARYING ρ	
10	.9	.9	.3	95	186	147
10	.5	.9	.3	98	124	136
10	.3	.9	.3	100	108	129
20	.9	.9	.3	117	199	129
20	.5	.9	.3	104	125	118
20	.3	.9	.3	104	111	111
30	.9	.9	.5	103	133	102
30	.9	.9	.3	100	136	93
30	.5	.9	.5	100	102	100
30	.5	.9	.3	96	101	96
30	.3	.9	.5	100	98	100
30	.3	.9	.3	97	96	96
40	.9	.9	.5	102	132	102
40	.9	.9	.3	104	125	83
40	.5	.9	.5	100	104	101
40	.5	.9	.3	99	102	94
40	.3	.9	.5	100	101	101
40	.3	.9	.3	99	98	94

^aNumber of time points; number of units fixed at 15.^b $x_{i,t} = \delta x_{i,t-1} + \mu_{i,t}$.^cSerial correlation of errors of first half of units.^dSerial correlation of errors of second half of units.

$$^e \text{Overconfidence} = 100 \frac{\sqrt{\sum_{l=1}^{1,000} (\hat{\beta}^{(l)} - \bar{\beta})^2}}{\sqrt{\sum_{l=1}^{1,000} (\text{s.e.}(\hat{\beta}^{(l)}))^2}}.$$

$$^f \text{Efficiency} = 100 \frac{\sqrt{\sum_{l=1}^{1,000} (\beta_{\text{Varying } \rho^{(l)}} - \beta_1)^2}}{\sqrt{\sum_{l=1}^{1,000} (\beta_{\text{Common } \rho^{(l)}} - \beta_1)^2}}.$$

Over 100% indicates superiority of common ρ estimator.

tion also experimentally manipulated.²⁰ Errors were then generated so that the variances and covariances of the errors were proportional to the corresponding variances and covariances of the independent variable. The errors could therefore show panel heteroscedasticity and contemporaneous correlation, either alone or in combination.

Table 4 shows a few key results from these experiments; a more complete table is in our companion article. These experiments set $N = 15$, vary T , and allow for various combinations of heteroscedasticity²¹ and contemporaneous correlation of the errors.

Panel-corrected standard errors performed excellently in these experiments. They were always within 10% of true variability, even under conditions of extremely high heteroscedasticity and contemporaneous correlation of the errors.²² In a typical research situation, we would expect PCSEs to be off by only a very few percentage points.

Of equal importance, in the case of homoscedasticity and contemporaneously independent errors, where OLS standard errors are accurate, PCSEs performed exactly as well as the OLS standard errors. But (as expected) as the errors became less spherical, the performance of the OLS standard errors declined.²³ Thus PCSEs dominate OLS standard errors; when PCSEs are not necessary, they perform as well

as the OLS standard errors, and when OLS standard errors perform poorly, PCSEs still perform well. Since PCSEs are not difficult to compute, they should replace OLS standard errors for TSCS data.²⁴

The Relative Efficiency of OLS and Parks. Panel-corrected standard errors perform well and are more accurate than Parks standard errors. Parks, however, was designed to take account of the panel error structure and hence be more efficient than OLS. The combination of OLS and PCSEs can only be clearly recommended if the OLS estimates of the parameters of equation 1 are, at worst, not much less efficient than the Parks estimates. We therefore designed experiments to compare the relative efficiency of OLS and Parks.

These experiments generated the errors completely independently of the independent variable.²⁵ The experiments generated the errors to show contemporaneous correlation of the errors. As before, for ease of exposition we generated the errors so that all units showed the same level of contemporaneous correlation of those errors.

Results of these experiments are in Table 5. Each entry in the table represents the relative efficiency of OLS as compared to Parks, with, for example, the first entry of 102 indicating that OLS is 2% more

TABLE 4

Ordinary Least Squares and Panel-corrected Standard Errors

T ^a	HETEROSCE- DASTICITY ^b	CONTEMPO- RANEOUS CORRELATION	OVER- CONFIDENCE (%) ^c	
			OLS	PCSE
10	0	0	102	102
10	0	.25	135	105
10	.3	0	119	102
10	.3	.25	144	105
20	0	0	95	96
20	.3	0	113	96
20	.3	.5	231	103
30	0	0	101	101
30	0	.5	229	107
30	.3	.5	234	106
40	0	0	104	104
40	0	.5	220	105
40	.3	0	120	102
40	.3	.5	225	104

^aNumber of time points; number of units fixed at 15.^bStandard deviation of $1/\sigma_i$, normalized.^cOverconfidence = $100 \frac{\sqrt{\sum_{i=1}^{1,000} (\hat{\beta}^{(i)} - \bar{\beta})^2}}{\sqrt{\sum_{i=1}^{1,000} (\text{s.e.}(\hat{\beta}^{(i)}))^2}}$.

accurate (in terms of the square root of mean squared error) in estimating β_1 than is Parks.

Ordinary least squares 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, OLS remains slightly more efficient than Parks. Parks becomes more efficient than OLS 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 OLS 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 .

Researchers can use the OLS residuals to compute the average contemporaneous correlation of the residuals.²⁶ Researchers should find OLS acceptable unless the average contemporaneous correlation is at least .50 and the time sample is quite long. We have done this calculation for a variety of TSCS data sets that were sent to us, and none of them met this condition. It is, of course, possible that some TSCS data might show extremely high contemporaneous correlation of the errors. For such data, researchers should consider alternatives to OLS, although the inaccurate standard errors of the Parks method would not make it the alternative of choice. A better strategy would be to model the cause of the high unit correlations directly, allowing whatever is causing unit errors to covary to be a variable in equation 1. But the need to do so should occur very rarely.

REANALYSIS

Hicks and Swank

We now use our methodology to reanalyze the study of Hicks and Swank (1992) and to draw some conclusions about a variety of other analyses. The underlying model assessed by Hicks and Swank, the *social democratic corporatist* model, has played an important role in the recent study of comparative political economy.²⁷ The Hicks and Swank study was chosen both because it exemplifies the issues we have been discussing and because the authors were kind enough to make their data available. Hicks and Swank used the Parks procedure as implemented in the computer package SHAZAM; reanalysis was done using RATS. We had no difficulty replicating the original Hicks and Swank results using RATS; all of the differences between our findings and theirs are due to changes in methodology.

Hicks and Swank are interested in the political and economic determinants of welfare spending in 18 advanced industrial societies for the 23-year period 1960–82. Here we reanalyze their “short model” containing only variables that pass a “jackknife” test (Hicks and Swank 1992, 667).²⁸ Their dependent variable is welfare spending as a proportion of gross domestic product. They use a variety of political, institutional, and economic independent variables. The political variables are electoral turnout and nine measures reflecting the strength of various parties: the strength of the Left, Center, and Right in the

TABLE 5

Relative Efficiency of Ordinary Least Squares Compared to Parks (%)

N ^b	T ^c	CONTEMPORANEOUS CORRELATION OF THE ERRORS			
		0	.25	.5	.75
10	10	102	100	99	97
	20	109	101	88	72
	30	112	105	90	68
	40	109	101	87	66
15	15	101	100	99	98
	20	108	102	93	84
	30	111	101	88	72
	40	111	100	83	64
20	20	102	101	100	99
	25	107	102	97	90
	30	107	100	91	80
	40	112	104	92	76

Note: Efficiency = $100 \frac{\sqrt{\sum_{i=1}^{1,000} (\hat{\beta}^{(i)} - \bar{\beta}_1)^2}}{\sqrt{\sum_{i=1}^{1,000} (\hat{\beta}^{(i)} - \beta_1)^2}}$.

Over 100% indicates superiority of OLS.

^aNumber of cross-sectional units.^bNumber of time points.

TABLE 6

Estimates of the Hicks and Swank Model of Social Security Spending in 18 Advanced Industrial Societies, 1960–82^a

	PARKS-AR1 ^b			OLS-AR1 ^c			
	b	se	t	b	OLS se	PCSE	t ^d
Turnout	5.7	.75	7.6	4.1	2.6	2.5	1.6
Government							
Left	.3	.72	.5	-1.11	1.8	1.7	-.6
Center	3.0	.65	4.6	1.01	1.8	1.6	.6
Right	1.4	.64	2.1	-.12	1.9	1.7	-.1
Opposition							
Left	-6.8	1.6	-4.1	-9.8	4.1	3.8	-2.6
Center	7.1	1.2	6.0	4.5	3.3	2.8	1.6
Right	3.0	1.2	2.5	.44	4.0	3.6	.1
Left corporatism	1.8	.33	5.3	1.8	.76	1.1	1.6
State centralization	1.8	.19	9.3	1.5	.70	.54	2.9
Bureaucratic traditionalism	.95	.25	3.8	1.6	.67	.73	2.2
Center-Left	-5.4	1.21	-4.5	-6.1	3.6	3.5	-1.7
Left-Center	2.9	.90	3.3	5.6	2.8	2.6	2.2
Left-Right	4.9	.93	5.3	5.1	3.2	2.9	1.7
lnGDP	1.0	.25	4.1	.61	.99	.87	.7
Price level (×100)	5.4	.37	14.4	4.7	1.3	1.1	4.3
Aged share of population	72.0	6.1	11.7	73.0	22.0	19.0	3.9
Post-OPEC	1.5	.08	18.1	1.6	.28	.23	7.0
Constant	-2.3	1.1	-2.0	.22	.13	.11	2.0
ρ^e	.90	—	—	.94	—	—	—

Note: The dependent variable is SOCIAL SECURITY SPENDING AS % OF GDP.

^aHicks and Swank estimates multiplied by 100.

^bFrom Hicks and Swank 1992, table 3, col. 4, correcting for both contemporaneously correlated and unit-specific AR1 errors.

^cCorrecting for AR1 errors (common ρ).

^dComputed using PCSEs.

^eAverage of unit-specific ρ_i for Parks-AR1, single ρ for OLS-AR1.

government; three similar measures for opposition party strength; and three interaction terms between government and opposition (center governments and left opposition, left governments and center opposition, and left governments and right opposition). A factor analysis of institutional variables yields three additional explanatory factor scores: left corporatism, state centralization, and bureaucratic patrimonialism. Finally, the model includes a variety of economic and social controls: the natural log of gross domestic product, the rate of inflation, the proportion of the population that is elderly, and a post-1973 OPEC oil-shock dummy. The model is linear in parameters and variables are measured so that all effects in the model are contemporaneous; dynamics are captured by (unit-specific) serially correlated errors. The model does not contain country dummy variables.

The estimates reported by Hicks and Swank, computing with the Parks method allowing for country-specific ρ_i are in Table 6, columns 1–3, labeled “Parks-AR1.” Their *t*-ratios are impressive, with 13 of 17 coefficients having *t*-ratios over four. However, our analysis shows that the Hicks and Swank standard errors may understate variability by a factor of three.

We reestimated the Hicks and Swank equation using OLS after correcting for serial correlation of the errors, assuming a common ρ . Results of this estima-

tion are in Table 6, columns 4–7. Both the usual standard errors (OLS se) and PCSEs are reported.²⁹ The *t*-ratios reported were computed using PCSEs.

The Hicks and Swank data, transformed to eliminate serially correlated errors, showed both heteroscedasticity (standardized measure of .37) and contemporaneous correlation of the errors (average of .25). Under these conditions, PCSEs and OLS standard errors should differ. They typically did differ but only by about 10%. While we base our conclusions on the PCSEs, we would have made similar findings using OLS standard errors.³⁰

Panel-corrected standard errors are, as predicted, roughly three times the standard errors obtained by Hicks and Swank.³¹ The Hicks and Swank data, when estimated with corrected standard errors, are not consistent with many of their conclusions. Hicks and Swank find that their “results support expectations, strongly rejecting most null hypotheses” (1992, 665). Our reanalysis finds that of the thirteen political and institutional variables in the short model, only four show *t*-ratios exceeding 2.0. Hicks and Swank find that the “evidence for positive voter turnout effects is pervasive and robust” (1992, 668). Our reanalysis finds this effect to be marginally statistically insignificant. Hicks and Swank find that “significant positive estimates for social democratic cor-

poratism, state administrative/political capabilities and traditional political legacies [are] everywhere strong and robust" (1992, 668). We find the social democratic corporatist effect to be marginally statistically insignificant, the state administrative/political capabilities effect to be moderately statistically significant and only the traditional bureaucratic legacies effect to be strongly statistically significant. Hicks and Swank find that "overall, strong evidence emerges for strategic interactions among parties" (1992, 668). We find that of the three interaction effects that pass the jackknife test, two are marginally significant and one is moderately so. Like Hicks and Swank, we do find the economic and social controls to have a generally strong impact on welfare spending, but those are not the variables of interest. While it is possible to argue about the credibility of findings that are at the margin of statistical significance, our reanalysis clearly casts doubt on the strength of the Hicks and Swank conclusions.

Other Studies

It is impossible to use the Parks method if the length of the time frame, T , is smaller than the number of units, N . Several published studies, therefore, present results that are either logically impossible to obtain or are completely a function of numerical inaccuracies. We have shown elsewhere that Alvarez, Garrett, and Lange's (1991) original results on the interaction of the strength of labor, party control, and economic outcome were simply artifacts of numerical inaccuracies in an old, unsupported, SAS procedure. Reanalysis using the methods recommended here supported their principal finding about economic growth but not their findings on unemployment or inflation (Beck et al. 1993). Several other articles listed in the lower half of Table 1 report having used the Parks method with $N > T$.³³ It is possible that these impossible results were obtained using the problematic routine. But while we cannot know how the results reported in these articles were achieved, we can be sure that they cannot have been achieved by appropriate use of the Parks method.

Turning to studies where the Parks method can in theory be useful, we reanalyzed Swank's (1992) study of the impact of politics on tax policy in 16 Organization for Economic Cooperation and Development nations observed over 20 years. With a sample this size, our simulations indicate that Swank's estimated standard errors should be overconfident by a factor of at least three; our reanalysis bore this out. The reanalysis, using OLS and PCSEs, showed that few of his coefficients, and none of his political coefficients, were significantly different from zero.

We did not reanalyze data sets used in the other Parks analyses (upper half of Table 1) and so can only use the Monte Carlo results to assess the inaccuracy of their standard errors. The findings of Hicks (1994b) about the short-run political causes of economic growth and unemployment, which are based on Parks analyses, are very problematic.³⁴ With obser-

vations on 17 countries over only 18 years, reported standard errors for these analyses should be overconfident by over 600%. Thus, while he reports findings with impressive t -ratios that are often near 10, correct computation of standard errors would lead to few if any rejections of the null hypothesis that none of the social democratic corporatist variables affect economic performance. Hicks concludes that "in the short run of year-to-year economic fluctuations, the social democratic corporatist (SDC) theory of economic performance is distinctly upheld for the case of economic growth" (p. 208). This is almost certainly an artifact of computing overconfident standard errors.

Su, Kamlet, and Mowery (1993) estimated models for disaggregated middle-class and defense budget categories with contemporaneously correlated errors.³⁵ Their most surprising finding was that party had a statistically significant effect on middle-class spending programs. This finding is based on 10 subprograms observed over 26 years. Our simulations indicate that their "significant" t -ratio of 2.5 is on the cusp of being significant at the .05 level.

Pampel's (1993) model relating fertility to cohort size and other socioeconomic variables worked with a larger time frame (36 years). His estimated standard errors are about 40% overconfident due to the correction for contemporaneously correlated errors, plus an additional factor for using unit-specific ρ_i to correct for serial correlation. The size of the latter factor is not known, because it is a function of the unknown trend in his data. Thus he should have used a critical value for his t -tests of between 2.5 and 3. Even with this higher critical value, it appears that many of his findings would remain statistically significant, although some of his findings of significant interactions between cohort size and socioeconomic variables would be overturned by the use of a more realistic critical value.

Finally, Schneider and Ingraham (1984) have an even longer time frame (59 years). With such a time frame the Parks standard errors are only slightly overconfident; none of the findings of that paper would be overturned by using more accurate standard errors. It is only with time frames as long as this study that the Parks method might prove useful. Such long time frames are exceedingly rare in the political science literature.

CONCLUSIONS

In his recent discussion of the analysis of TSCS data in comparative political economy, Hicks notes that "we should, pending information on the small sample properties of standard errors and t statistics in the Parks-Kmenta model, be wary of downward bias in standard errors and upward bias in t -statistics to the extent that $N(N - 1)/2$ approaches NT " (1994a, 186). The present article answers Hicks' request and shows clearly that the downward bias in standard errors makes the Parks technique unusable unless there are substantially more time points (T) than there are

cross-sectional units (N). In particular, the Parks technique is extremely misleading for the types of TSCS data typically analyzed by political scientists. As a consequence, the substantive conclusions of many articles that use the Parks method to estimate TSCS models are, at best, open to doubt. We counterbalance this negative conclusion by providing a simple methodology for analyzing TSCS data.

Time-series cross-section analysts should proceed by first examining the temporal properties of their data. This can be done, as we argue elsewhere, with lagged dependent variables or, as is typically done in the cross-national panel literature and as we do here, by transforming the data to eliminate serial correlation of the errors (Beck and Katz n.d.). If researchers choose the latter route, there is no doubt that they should transform based on an estimate of the common serial correlation; researchers correcting for serial correlation only should follow a similar strategy.

Once the dynamics are accounted for or transformed away, TSCS analysts can estimate model parameters by OLS. Our Monte Carlo evidence shows that this will seldom lead to a substantial loss of efficiency for the types of TSCS data typically analyzed by political scientists. Standard errors should be calculated using PCSEs. Our Monte Carlo evidence shows that there can be no harm from using PCSEs, while, in some circumstances, they may be considerably more accurate than the usual OLS standard errors. The combination of OLS with PCSEs allows for accurate estimation of variability in the presence of panel error structures without inducing the severe problems caused by the Parks method.

Researchers who worry that their data may fall into one of the extreme cases of heteroscedasticity or contemporaneous correlation of the errors can check for these problems by examining the structure of the OLS residuals. Only if these problems are severe, and only if sample sizes are large enough, should researchers contemplate a more complicated FGLS estimation strategy. Those contemplating such a strategy must trade-off the potential advantages of FGLS against the disadvantages of inaccurate standard errors. We have not seen a TSCS data set that makes it necessary even to consider this trade-off.

Many will take comfort in our finding that the workhorse of political methodology, OLS, is superior to the more complicated GLS approach to the analysis of TSCS data. We do not argue that complicated methods will always have problems, or that OLS is in general superior to complicated methods. It is, however, critical that we learn to assess the properties of complicated estimation strategies, and in particular that we study these properties for the types of data actually analyzed, rather than for large samples observed only in "asymptopia." This is particularly true for assessments of variability, about which we often have little or no intuition. Ordinary least squares, with corrected standard errors, will not always prove to be superior to more complicated techniques, but it clearly is superior to the Parks method for analyzing

the types of time-series cross-section data that are used by students of comparative politics.

Notes

We would like to thank Michael Alvarez, Geoffrey Garrett, Peter Lange, Alexander Hicks, and Duane Swank for generously providing their data for replication purposes. William Greene, Gary King, and Glenn Sueyoshi deserve more than the usual thanks for helping us to figure out both what we were doing and how to communicate it. We also thank Michael Alvarez, Charles Franklin, Ronald Gallant, Elizabeth Gerber, Sung Hahm, William Heller, Mark Kamlet, Brian Loynd, Glenn Mitchell, Chris Mooney, Jimmy Sanders, Renee Smith, James Stimson, and Michael Thies for helpful comments and conversations. We are grateful to Peter Williams for providing new LATEX styles. Katz thanks the National Science Foundation for a graduate fellowship that funded his work on this project while he was at the University of California, San Diego. Earlier versions were delivered at the 1993 annual meetings of the American Political Science Association in Washington, the Political Methodology Group in Tallahassee, and the Midwest Political Science Association in Chicago. All computer codes and data related to this article may be obtained via ftp to weber.ucsd.edu.

1. This method was popularized in Kmenta's text, so it is sometimes referred to as Parks-Kmenta and sometimes attributed only to Kmenta (1986, 622–25). We call it "Parks" throughout this article.

2. We have written a companion article, Beck and Katz n.d., that treats the dynamic issues of TSCS estimation in detail; it also examines estimators that correct for panel heteroscedasticity only.

3. We assume that the reader is familiar with the basics of TSCS models, as laid out in Stimson 1985 and more fully in Hsiao 1986. Stimson distinguished models that are cross-sectionally dominated from ones that are not. The former (e.g., the National Election Studies Panels) have observations on thousands of units observed a few times. While TSCS models are formally equivalent to such panel designs, the problems faced by TSCS modelers are very different from those faced by electoral panel modelers. The present article does not consider issues that arise in cross-sectionally dominated designs. While we typically call the designs we study TSCS models, in a few places it is more convenient to refer to our designs as *panels*; the two terms are used interchangeably here. Earlier versions of this article used the term *cross-national panel*. While this term correctly connotes the size of cross-section we consider, we do not wish in any way to limit the range of applications of our method to cross-national data.

4. The exogenous variables may contain unit-specific dummy variables, allowing intercepts to vary by unit. Such a model is called a *fixed effects* model. Fixed effects present no special problems for TSCS models, because the number of unit-specific dummy variables required is not large. We do not consider random effects models, which are heavily used in cross-sectionally dominated panel models but are not important for TSCS work. We show in Beck and Katz n.d. that it is easy to include lagged dependent variables in expression 1.

5. That is, the data are ordered so that the second observation is the observation on unit 1 for the second time period and, in general, the observation following unit i for time period t is the observation for unit i for time period $t + 1$ (or, following the last observation on unit i , it is the first observation on unit $i + 1$).

6. We can state the various "panel error" assumptions symbolically as:

Panel Heteroscedasticity. $E(\epsilon_{i,t}^2) \neq E(\epsilon_{j,t}^2)$, but $E(\epsilon_{i,t}^2) = E(\epsilon_{i,t'}^2)$, so we can write $E(\epsilon_{i,t}^2) = \sigma_i^2$.

Contemporaneously Correlated Errors. $E(\epsilon_{i,t}\epsilon_{j,t}) = E(\epsilon_{i,t'}\epsilon_{j,t'}) \neq 0$, but $E(\epsilon_{i,t}\epsilon_{j,t'}) = 0$, so we can write $E(\epsilon_{i,t}\epsilon_{j,t}) = \sigma_{ij}$, with all other covariances being zero.

Unit-specific Serially Correlated Errors. $\epsilon_{i,t} = \rho_i \epsilon_{i,t-1} + \nu_{i,t}$, where the ν are incoming "shocks" that are temporally independent, identically distributed, zero-mean random variables.

Common Serially Correlated Errors. $\epsilon_{i,t} = \rho \epsilon_{i,t-1} + \nu_{i,t}$, where the $\nu_{i,t}$ are incoming "shocks."

7. The FGLS method to correct for only panel heteroscedasticity is panel-weighted least squares, which is incorporated in Stimson's (1985) GLS-ARMA technique. The FGLS correction for panel heteroscedasticity is not subject to the extreme problems of the correction for contemporaneously correlated errors so we do not further consider it here. Those contemplating using this technique should see Beck and Katz n.d.

8. The same method may be used with a common serial correlation process, but the Parks method allows for unit-specific serial correlations.

9. Analysts modeling temporal dynamics with a lagged dependent variable need only use the second transformation. Other analysts use only the first stage of the Parks method, correcting only for unit-specific serial correlation. As we shall see, the use of either transformation alone leads to serious underestimates of variability.

10. \otimes is the Kronecker product. For example, if $N = 2$ and $T = 3$, the variance covariance matrix of the errors is

$$\Omega = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \sigma_{12} & 0 & 0 \\ 0 & \sigma_1^2 & 0 & 0 & \sigma_{12} & 0 \\ 0 & 0 & \sigma_1^2 & 0 & 0 & \sigma_{12} \\ \sigma_{12} & 0 & 0 & \sigma_2^2 & 0 & 0 \\ 0 & \sigma_{12} & 0 & 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_{12} & 0 & 0 & \sigma_2^2 \end{pmatrix}.$$

11. Feasible generalized least squares requires that $\hat{\Sigma}$ be invertible. It is an $N \times N$ matrix, but its rank is, at most, the lesser of T and N . Thus it is not invertible unless T is at least as large as N .

12. The OLS standard errors will not be consistent. The degree of inaccuracy is a complicated function of the relationship between the $X'X$ matrix and the variances and covariances of the error process. If these are only slightly related then the OLS standard errors will only be slightly incorrect. This is shown clearly for the cross-sectional case by White (1980). One consequence of this is that it is not possible to infer the overconfidence of OLS standard errors as a simple function of sample size.

13. We use this term to distinguish our proposed estimate of variability from White's (1980) heteroscedasticity-consistent standard errors. Many standard packages, such as RATS or SHAZAM compute robust standard errors based on White's method; none of these take into account the panel structure of the errors, even within a panel estimation module. Our Monte Carlo results, presented in Beck and Katz n.d., show that researchers should not use robust standard errors calculated by typical computer packages to approximate PCSEs. It should also be stressed that our use of the term *panel-corrected* indicates that the standard errors computed will not be inaccurate due to the panel structure of the data. Other pathologies, such as a time-varying error structure, may indeed cause PCSEs to be inaccurate. In this regard, PCSEs are no better, but no worse, than any other estimator proposed for TSCS data.

14. We are using the maximum likelihood, rather than the unbiased, estimate of Σ . The unbiased estimate would be obtained by dividing by $T - k$. It is not clear which method is superior. In our simulations, we use only one independent variable, so the difference between the two estimators is small.

15. In symbols, if E denotes the $T \times N$ matrix of the OLS residuals, we can estimate Σ by

$$\hat{\Sigma} = \frac{(E'E)}{T}$$

and hence estimate Ω by

$$\hat{\Omega} = \frac{(E'E)}{T} \otimes I_T,$$

where \otimes is the Kronecker product. Panel-corrected standard errors are thus computed by taking the square root of the diagonal elements of

$$(X'X)^{-1}X' \left(\frac{E'E}{T} \otimes I_T \right) X(X'X)^{-1}.$$

16. We examined the overconfidence of the Parks standard errors in many more complicated contexts. Our findings about the overconfidence of the Parks standard errors in those experiments are identical, to within the small sampling error of the experiments, to what is reported in Table 2.

17. As in all our experiments, we are assessing the estimation of β_1 and its variability.

18. Some TSCS analysts do not use the Parks method but do correct for serially correlated errors using unit-specific ρ_i (e.g., Levobich 1994; Pollins 1989; Rosh 1988). The simulation results apply equally to this method. Correcting for unit-specific serial correlation leads to substantial overconfidence, whether used alone or as part of the Parks method.

19. We deal only with the performance of these estimators assuming that serial correlation has already been eliminated.

20. The contemporaneous correlation of the errors matrix was assumed to take this simple form for ease of exposition. None of the experimental results are artifacts of this simple form, since the various estimation techniques allow for freely varying contemporaneous correlations.

21. We have seen no textbook measure of the degree of panel heteroscedasticity. It is measured for Table 4 by the standard deviation of the normalized weights that would be used in panel-weighted least squares. The variance of the i^{th} unit is σ_i^2 . Let $w_i = 1/\sigma_i$. We define *standardized heteroscedasticity* as the standard deviation of the w_i/\bar{w} .

22. This result is based on both Table 4 and results reported in Beck and Katz n.d.

23. Again, we designed the experiments where OLS performance would be degraded. In practical situations the OLS standard errors may perform well.

24. The code to compute PCSEs in RATS is available via ftp from weber.ucsd.edu.

25. The inefficiency of OLS in the presence of nonspherical errors does not depend on the relationship of the error process and the independent variables.

26. This statistic is computed in our RATS program.

27. This importance is best seen in the variety of papers in Janoski and Hicks 1994.

28. For definition and operationalization of these variables, see Hicks and Swank 1992, 663–64. We received data only on variables contained in the short model. Results from this model are similar to other results reported by Hicks and Swank.

29. Panel-corrected standard errors were computed after the data were transformed to eliminate serial correlation.

30. This is not an argument for using OLS standard errors. It is difficult to know, in advance, whether OLS standard errors are accurate, while we are confident about the accuracy of PCSEs. The only way we know that OLS standard errors are accurate for this reanalysis is that they are close to the estimated PCSEs. Since there is no reason for TSCS researchers to compute possibly misleading OLS standard errors, we focus on the PCSEs here.

31. We also estimated the Hicks and Swank model using our methods but allowing for unit-specific ρ_i . These estimates (not reported here) had, as predicted, PCSEs that were between 25% and 50% smaller than the PCSEs computed under the assumption of a common ρ . If we compare PCSEs to Parks standard errors when both correct for serially correlated errors similarly, we find that the PCSEs are, as predicted, about two-and-a-half times larger than the corresponding Parks standard errors.

32. Hicks and Swank appear to equate a strong effect with

one having a large t -ratio. We therefore assess their findings using this standard.

33. As noted in that table, some of these studies used Parks in addition to other estimation methods. We have no findings on the conclusions based on other methods.

34. His analysis of the long run does not use Parks and so is not discussed here.

35. This study uses a "seemingly unrelated regressions" methodology which is identical to the FGLS correction for contemporaneously correlated errors.

References

- Alvarez, R. Michael, Geoffrey Garrett, and Peter Lange. 1991. "Government Partisanship, Labor Organization, and Macroeconomic Performance." *American Political Science Review* 85:539–56.
- Beck, Nathaniel, and Jonathan N. Katz. N.d. "Nuisance or Substance: Specifying and Estimating Times-Series-Cross-Section Models." *Political Analysis*. Forthcoming.
- Beck, Nathaniel, Jonathan N. Katz, R. Michael Alvarez, Geoffrey Garrett, and Peter Lange. 1993. "Government Partisanship, Labor Organization, and Macroeconomic Performance: A Corrigendum." *American Political Science Review* 87:945–48.
- Cramer, J. 1986. *Econometric Applications of Maximum Likelihood Methods*. New York: Cambridge University Press.
- Freedman, David, and Stephen Peters. 1984. "Bootstrapping a Regression Equation: Some Empirical Results." *Journal of the American Statistical Association* 79:97–106.
- Friedland, Roger, and Jimmy Sanders. 1985. "The Public Economy and Economic Growth in Western Market Economies." *American Sociological Review* 50:421–37.
- Giles, Michael, and Kaenan Hertz. 1994. "Racial Threat and Partisan Identification." *American Political Science Review* 88:317–26.
- Hicks, Alexander. 1994a. "Introduction to Pooling." In *The Comparative Political Economy of the Welfare State*, ed. Thomas Janoski and Alexander Hicks. New York: Cambridge University Press.
- Hicks, Alexander. 1994b. "The Social Democratic Corporatist Model of Economic Performance in Short- and Medium-Run Perspective." In *The Comparative Political Economy of the Welfare State*, ed. Thomas Janoski and Alexander Hicks. New York: Cambridge University Press.
- Hicks, Alexander, and Duane Swank. 1992. "Politics, Institutions and Welfare Spending in Industrialized Democracies, 1960–1982." *American Political Science Review* 86:658–74.
- Hsiao, Cheng. 1986. *Analysis of Panel Data*. New York: Cambridge University Press.
- Hurwicz, L. 1950. "Least-Squares Bias in Time Series." In *Statistical Inference in Dynamic Economic Models*, ed. T. Koopmans. New York: Wiley.
- Janoski, Thomas, and Alexander Hicks. 1994. *The Comparative Political Economy of the Welfare State*. New York: Cambridge University Press.
- Kakwani, N. 1967. "The Unbiasedness of Zellner's Seemingly Unrelated Regression Equation Estimators." *Journal of the American Statistical Association* 62:141–42.
- Kmenta, Jan. 1986. *Elements of Econometrics*. 2d ed. New York: Macmillan.
- Levobich, James. 1994. "Riding Waves or Making Waves? The Services and the U.S. Defense Budget, 1981–1993." *American Political Science Review* 88:839–52.
- Lin, Stephen. 1994. "Government Spending and Economic Growth." *Applied Economics* 26:83–94.
- Pampel, Fred. 1993. "Relative Cohort Size and Fertility: The Socio-political Context of the Easterlin Effect." *American Sociological Review* 58:496–514.
- Pampel, F., and J. Williamson. 1988. "Welfare Spending in Advanced Industrial Democracies, 1950–1980." *American Journal of Sociology* 93:1424–56.
- Parks, Richard. 1967. "Efficient Estimation of a System of Regression Equations When Disturbances Are Both Serially and Contemporaneously Correlated." *Journal of the American Statistical Association* 62:500–509.
- Pollins, Brian. 1989. "Does Trade Still Follow the Flag?" *American Political Science Review* 83:465–80.
- Rosh, Robert M. 1988. "Third World Militarization." *Journal of Conflict Resolution* 32:771–98.
- Schneider, Sandra K., and Patricia Ingraham. 1984. "The Impact of Political Participation on Social Policy Adoption and Expansion." *Comparative Politics* 17:107–21.
- Scholz, John T., Jim Twombly, and Barbara Headrick. 1991. "Street-level Political Controls over Federal Bureaucracy." *American Political Science Review* 85:829–50.
- Scholz, J., and F. H. Wei. 1986. "Regulatory Enforcement in a Federalist System." *American Political Science Review* 80:1249–70.
- Stimson, James. 1985. "Regression in Space and Time: A Statistical Essay." *American Journal of Political Science* 29:914–47.
- Su, Tsai-tsu, Mark Kamlet, and David Mowery. 1993. "Modeling United States Budgetary and Fiscal Policy Outcomes—a Disaggregated, Systemwide Perspective." *American Journal of Political Science* 37:213–45.
- Swank, Duane. 1992. "Politics and the Structural Dependence of the State in Democratic Capitalist Nations." *American Political Science Review* 86:38–54.
- White, Halbert. 1980. "A Heteroscedasticity-consistent Covariance Matrix and a Direct Test for Heteroscedasticity." *Econometrica* 48:817–38.
- Wood, B. Dan. 1992. "Modeling Federal Implementation as a System: The Clean Air Case." *American Journal of Political Science* 36:40–67.

Nathaniel Beck is Professor of Political Science, University of California at San Diego, La Jolla, CA 92093.

Jonathan N. Katz is Assistant Professor of Political Science, California Institute of Technology, Pasadena, CA 91125.