
Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models
Author(s): Nathaniel Beck and Jonathan N. Katz
Source: *Political Analysis*, Vol. 6 (1996), pp. 1-36
Published by: Oxford University Press on behalf of the Society for Political Methodology
Stable URL: https://www.jstor.org/stable/25791568
Accessed: 21-08-2018 15:45 UTC

Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models

Nathaniel Beck and Jonathan N. Katz

Abstract

In a previous article we showed that ordinary least squares with panel corrected standard errors is superior to the Parks generalized least squares approach to the estimation of time-series-cross-section models. In this article we compare our proposed method with another leading technique, Kmenta's "cross-sectionally heteroskedastic and timewise autocorrelated" model. This estimator uses generalized least squares to correct for both panel heteroskedasticity and temporally correlated errors. We argue that it is best to model dynamics via a lagged dependent variable rather than via serially correlated errors. The lagged dependent variable approach makes it easier for researchers to examine dynamics and allows for natural generalizations in a manner that the serially correlated errors approach does not. We also show that the generalized least squares correction for panel heteroskedasticity is, in general, no improvement over ordinary least squares and is, in the presence of parameter heterogeneity, inferior to it. In the conclusion we present a unified method for analyzing time-series-cross-section data.

Introduction

The analysis of time-series-cross-section (TSCS) data is becoming more common in the social sciences. Such data are used in cross-national polit-

Thanks to Ross Burkhart and Michael Lewis-Beck for providing data and to Elizabeth Gerber, William Greene, William Heller, Gary King, Andrew Levin, Brian Loynd, James Stimson, Glenn Sueyoshi, and Diana Weinhold for helpful comments. Katz thanks the National Science Foundation for a Graduate Fellowship funding his work on this project while at UCSD. Part of this article was delivered at the 1993 Annual Meeting of the Midwest Political Science Association, Chicago, Illinois.

ical economy studies, comparisons of policy across the American states, and the analysis of disaggregated budgets. TSCS data are characterized by repeated observations on fixed units such as states or nations. The number of units typically ranges from about 10 to 100. These units are observed repeatedly, with typical data containing twenty or more annual observations on each unit.[1]

Both the temporal and spatial properties of TSCS data make ordinary least squares (OLS) problematic. In particular, models for TSCS data often allow for temporally and spatially correlated errors as well as for panel heteroskedasticity.[2] There are several ways to "correct" for these complicated error processes, all of them based on generalized least squares (GLS). Elsewhere we showed that the GLS correction for spatially correlated errors (Parks 1967) leads to very bad estimates of standard errors (Beck and Katz 1995). In that article we also proposed a new estimator for the variability of OLS estimators, "panel corrected standard errors" (PCSEs). Monte Carlo analysis shows that PCSEs are excellent estimators of the variability of OLS estimates. We therefore proposed that analysts of TSCS data use OLS with PCSEs to correct for cross-sectional complications of the error process.

There are alternatives to our proposed method. The leading alternative is Kmenta's (1986, 618-22) "cross-sectionally heteroskedastic and timewise autocorrelated" (CHTA) model.[3] This procedure uses GLS to correct for panel heteroskedasticity and serially correlated errors. CHTA has been widely used by TSCS analysts.[4] It does not suffer from the same problems as the Parks method. In this article we assess the relative performance of CHTA and OLS in the context of data observed in common research situations.[5]

---

[1]TSCS data are distinguished from cross-sectionally dominated "panel data," which have a few repeated observations on a large number of sampled units. We assume the reader is familiar with standard TSCS analysis (Hsaio 1986; Stimson 1985).

[2]Panel heteroskedasticity, as distinct from ordinary heteroskedasticity, allows the error variances to vary from unit to unit while requiring that they be constant within each unit.

[3]This method was popularized for political scientists in Stimson's (1985) influential review of TSCS methods. His version of CHTA is called "GLS-ARMA." CHTA and GLS-ARMA are equivalent. Other authors using CHTA cite Greene (1993, 444-59). For convenience we refer to Kmenta's CHTA in this article.

[4]Recent articles in major journals that used CHTA include Blais, Blake, and Dion (1993), Burkhart, and Lewis-Beck (1994), Clingermayer, and Wood (1995), Giles and Hertz (1994), Levobic (1994) and Pollins (1989). Note that, while Giles and Hertz wrote that they used the Parks method, a personal communication indicates that they used CHTA.

[5]Since we have shown that PCSEs are never worse than, and are often superior to, OLS standard errors for TSCS data, whenever we refer to OLS we assume that standard errors are PCSEs.

We argue that CHTA is not more efficient than OLS, in general, nor does it provide assessments of variability as accurately as do PCSEs. Since we are interested in the performance of the various estimators in finite samples, we cannot obtain analytic assessments of how well the estimators perform. We therefore compare CHTA and OLS/PCSE via Monte Carlo experiment.[6] We also argue that CHTA can lead investigators to ignore important features of the data, while our proposed OLS/PCSE methodology leads investigators to consider those important features.

Since CHTA proceeds by first eliminating serial correlation of the errors and then by eliminating panel heteroskedasticity, we can consider the two components separately. We first contrast the CHTA correction for serially correlated errors with estimation of models that include lagged dependent variables. We argue in the third section that there are significant advantages to modeling the dynamics with a lagged dependent variable. This latter argument relies on the modern approach to the analysis of single time series. It is buttressed by Monte Carlo simulations.

CHTA corrects for panel heteroskedasticity using panel weighted least squares. In the fourth section we use Monte Carlo experiments both to assess the relative efficiency of this method as compared with OLS and to assess the accuracy of reported standard errors. We also consider the performance of the two estimators in the presence of parameter heterogeneity. We argue that the empirical weights used by panel weighted least squares can mislead investigators and that there is typically little or no gain from weighting.

In the next section we lay out the details of the TSCS model and discuss why OLS may not be optimal. We also briefly lay out the CHTA solution and our proposed solution, OLS with PCSEs. After comparing the properties and performance of these two solutions, in the fifth section we reconsider the findings of the CHTA analysis of Burkhart, and Lewis-Beck (1994). The conclusion lays out a simple, unified method for analyzing TSCS data. The appendix treats some technical issues concerning PCSEs.

The Estimation of Time-Series-Cross-Section Models

The generic TSCS model has the form

$$y_{i,t} = \mathbf{x}_{i,t}\beta + \epsilon_{i,t}; \qquad i = 1, \ldots, N; \qquad t = 1, \ldots, T, \tag{1}$$

[6]All simulations were written using GAUSS 3.1 and are available by ftp from weber.ucsd.edu.

where $\mathbf{x}_{i,t}$ is a K vector of exogenous variables and observations are indexed by both unit (i) and time (t). We denote the $NT \times NT$ covariance matrix of the errors as $\mathbf{\Omega}$, with its typical element being $E(\epsilon_{i,t}\epsilon_{j,s})$.

The vector of observations on the dependent variable is denoted as $\mathbf{Y}$, with the matrix of independent variables being denoted $\mathbf{X}$. All vectors and matrices are stacked by unit. We work with fixed effects models, so the exogenous variables may include a vector of unit-specific dummy variables. We do not allow for random effects since they are not relevant in the TSCS case.[7]

Feasible Generalized Least Squares Estimates of TSCS Models

OLS is an optimal estimator[8] of equation 1 if the errors follow a simple, spherical form, where

$$\mathbf{\Omega} = \sigma^2\mathbf{I}. \tag{2}$$

TSCS analysts allow for more complicated error structures. The errors may show panel heteroskedasticity:

$$\mathrm{Var}(\epsilon_{i,t}^2) = \sigma_i^2. \tag{3}$$

Errors may also be contemporaneously correlated:

$$E(\epsilon_{i,t}\epsilon_{j,t}) = \sigma_{ij} \tag{4}$$

$$E(\epsilon_{i,t}\epsilon_{j,t'}) = 0 \qquad \text{for } t \neq t'. \tag{5}$$

[7] Hsaio (1986, 43) notes that if "inferences are going to be confined to the effects in the model, the effects are more appropriately considered fixed." In the typical cases of comparative politics research, the units are nations or states, with inference being confined to the set of nations or states being studied. For example, in the analysis of advanced industrial societies, inferences are conditional on the set of advanced industrial societies that are observed.

In any event, for the typical time samples used in TSCS data, there is little if any difference between fixed and random effects estimators. To see this, note that the random effects estimator differs from the fixed effects estimator insofar as $\sigma_\epsilon^2/(\sigma_\epsilon^2 + T\sigma_\mu^2)$ differs from zero (σ_ϵ^2 is the error variance, and σ_μ^2 is the variance of the random effects) (Greene 1993, 473). This shows that as T gets large the random effects estimator approaches the fixed effects estimator. In typical TSCS applications, T will be at least 20, so the differences between the two estimators will usually be small.

[8] There are clearly many other issues in TSCS modeling that go beyond simple parameter estimation. In particular, many controversies hinge on choice of units and whether some units are "outliers" that should be excluded from analysis. This article does not examine these issues directly. But clearly any further analysis must be improved by better parameter estimates and improved standard errors. Thus, for example, analysis of whether parameters vary across subgroups will depend strongly on first getting correct estimates of standard errors. We return to this issue in our reanalysis of the Burkhart and Lewis-Beck data.

Finally, the errors may manifest serial correlation:

$$\epsilon_{i,t} = \rho\epsilon_{i,t-1} + \nu_{i,t}, \tag{6}$$

where $\nu_{i,t}$ are incoming "shocks," that is, independent, identically distributed ("iid"), zero-mean random variables.[9]

The assumption of serially correlated errors is one way to model the dynamics inherent in TSCS data. We could, alternatively, model the dynamics via a lagged dependent variable:

$$y_{i,t} = \phi y_{i,t-1} + \mathbf{x}_{i,t}\beta + \epsilon_{i,t}. \tag{7}$$

The errors in this model may also show panel heteroskedasticity, contemporaneous correlation or serial correlation. As we show in the third section the errors in equation 7 will usually be serially independent. We refer to this equation with serially independent errors as the "dynamic model." Since the dynamics inherent in equation 1 with serially correlated errors are considered to be merely a nuisance that impedes estimation, we refer to this setup as the "static model."

TSCS analysts are usually unwilling to assume that errors are spherical and hence do not consider OLS to be optimal for their data. They have therefore turned to "feasible generalized least squares" (FGLS) to estimate the static model. This requires using OLS to estimate equation 1, with the OLS residuals from this estimation used to estimate $\mathbf{\Omega}$. This estimate of $\mathbf{\Omega}$ is used to transform the data, with the new, transformed model having a spherical error process. This transformed model can be estimated by OLS. CHTA uses two such transformations: one to eliminate serial correlation of the errors and one to eliminate panel heteroskedasticity.

Ordinary Least Squares with Panel Corrected Standard Errors

In Beck and Katz (1995) we proposed that analysts deal with the complications of TSCS error processes by using OLS but with panel corrected standard errors. While OLS is not efficient in the presence of nonspherical errors, it does yield consistent estimates. The simulations we reported showed that the efficiency loss of OLS would not be large in practical research situations.

OLS standard errors will be inaccurate in the presence of nonspherical errors, however, in that they do not provide good estimates of the sampling variability of the OLS parameters estimates. Our simulation

[9] Thus, we assume that the errors follow a common first-order autoregressive process. Some analysts allow for unit specific ρ_i.

showed that PCSEs are accurate in the presence of either contemporaneously correlated or panel heteroskedastic errors. If, as we argue will usually be the case, the errors in equation 7 are serially independent, OLS with PCSEs should provide good estimates of equation 7. PCSEs are calculated using the OLS residuals from equation 7. Since we are assuming that the errors in equation 7 are temporally independent, the variance-covariance matrix of the errors takes a simple form

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma} \otimes \mathbf{I}_T, \tag{8}$$

where $\boldsymbol{\Sigma}$ is the $N \times N$ matrix of error variances and contemporaneous covariances (with σ_i^2 from eq. 3 along the diagonal and σ_{ij} from equation 4 off the diagonal) and $\otimes$ denotes the Kronecker product.[10]

Let $\mathbf{E}$ denote the $T \times N$ matrix of the OLS residuals. $\mathbf{E}'\mathbf{E}/T$ provides a consistent estimate of $\boldsymbol{\Sigma}$.[11] PCSEs are thus estimated by the square root of the diagonal of

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \left(\frac{\mathbf{E}'\mathbf{E}}{T} \otimes \mathbf{I}_T \right) \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{9}$$

Our interest here is in comparing OLS/PCSE with CHTA. We wish to compare the performance of these two methods for the types of data seen in research situations. Both methods treat the dynamic properties of TSCS data as well as attempt to remedy problems caused by cross-sectional complications. Both methods treat these issues independently. OLS/PCSE treats dynamics via a lagged dependent variable and handles cross-sectional complications via the PCSEs. CHTA first transforms the data to eliminate serially correlated errors and then transforms the transformed data to eliminate panel heteroskedasticity. We can therefore examine how well each method handles dynamics and how well each

[10] A more complete derivation of PCSEs is presented in the appendix.

[11] While dividing by T yields a consistent estimate, we could just as well have divided by $T - K$. There is no theory as to which is better in small samples (it doesn't matter in large samples). While conventionally we divide in similar situations by $T - K$, the theory behind this is not well established. While dividing by $T - K$ yields an unbiased estimate, dividing by T yields an estimate with lower mean squared error. One argument for dividing by $T - K$ is that it is more conservative in that it yields higher estimates of variability. But this is not necessarily an improvement if dividing by T yields the correct estimate of variability. Our simulation had $K = 2$ and so shed little light on this issue. At present all we can say is that PCSEs computed by dividing by T are superior to OLS standard errors and that they perform well in our simulations. We have no evidence about whether we could improve matters even more by dividing by $T - K$. For reasonable values of N and K (say more than 20 and less than five), the difference between the two methods of computing PCSEs will be under 15 percent.

handles cross-sectional issues, separately. We do this in the next two sections. Since both methods treat dynamics first, we begin with that issue.

The Dynamics of Time-Series-Cross-Section Models

Equation 1, with first order autoregressive errors (eq. 6), can be estimated by FGLS; we refer to this method as "AR1." This method proceeds by first estimating equation 1 by OLS and then using the OLS residuals to estimate ρ in equation 6. The observations are then transformed by the well-known Prais-Winsten transformation (see, e.g., Greene 1993, 432) to produce serially independent errors.[12] This transformation is:

$$y^*_{i,t} = y_{i,t} - \hat{\rho} y_{i,t-1}, \qquad t \geq 2 \tag{10}$$

$$y^*_{i,1} = \frac{y_{i,1}}{\sqrt{1-\hat{\rho}^2}} \tag{11}$$

with $\mathbf{X}$ transformed similarly.

Implicit in this notation is the idea that the errors for all units follow the same autoregressive process with a common autoregressive parameter, ρ. Kmenta's CHTA allows for unit-specific ρ_i, and many CHTA analyses follow Kmenta. In Beck and Katz (1995) we argued that it is better to assume a common autoregressive process. The essence of the argument is that TSCS analysts start with the assumption that the parameters of interest, β, do not vary by unit; this "pooling" assumption is at the heart of TSCS analysis. Why then should the "nuisance" serial correlation parameters vary by unit? We then showed, by a series of Monte Carlo experiments, that the assumption of a common serial correlation process leads to superior estimates of β even when the data are generated with diverse, unit-specific ρ_i.

The inferiority of the unit-specific serial correlation estimator is a consequence of each of the ρ_i being estimated using only a small number (T) of observations. It is well known that autoregressions estimated using 20 or 30 observations will lead to erratic results. Among other things, autoregressive parameters estimated in small samples are biased downward (Hurwicz 1950; Nickell 1981). The easiest way to see why this bias exists is to note that the dependent variable is centered prior to its being used in the standard regression formula. In an autoregression this centering induces a correlation between the centered dependent

[12] This transformation is required so as not to lose the information contained in the first observation for each unit. Monte Carlo studies of single time series have found this method to be superior to simply discarding the first observation.

variable and the independent variables. The latter are just the lags of the dependent variable and hence must be correlated with the average of the dependent variable, the term that is subtracted for centering. The smaller the sample size, the larger that correlation. Hurwicz showed that the degree of downward bias is approximately $1 + 3\rho/(T-1)$. Thus, for example, for $T = 20$ and $\rho = .5$, the estimate of each unit-specific serial correlation is downwardly biased by about 12 percent; when ρ rises to .8, this bias increases to about 18 percent. Thus, the Kmenta procedure corrects for serial correlation using unstable downwardly biased estimates. The estimate of a common ρ suffers neither from instability nor downward bias, since it is estimated using $N \times T$ observations.

We can, alternatively, add a lagged dependent variable to equation 1 to produce equation 7, which can then be estimated by OLS; we refer to this method as LDV. The LDV model of dynamics makes it clear why the unit-specific correction for serial correlation seems odd. The LDV analogue of unit-specific serial correlation would be the model

$$y_{i,t} = \phi_i y_{i,t-1} + \mathbf{x}_{i,t}\beta + \epsilon_{i,t}. \tag{12}$$

This model is never seen in practice. Why should ϕ vary by unit when β does not? If anything, ϕ, which measures speed of adjustment, is more likely to be homogeneous across units than is β. For the remainder of this article we assume that the dynamics are modeled with a common ρ (or ϕ).

There is an alternative to both AR1 and LDV that encompasses each of them.[13] Equation 1 with serially correlated errors can be rewritten as

$$y_{i,t} = \rho y_{i,t-1} + \mathbf{x}_{i,t}\beta - \mathbf{x}_{i,t-1}(\rho\beta) + \nu_{i,t}, \tag{13}$$

where the $\nu_{i,t}$ are iid, zero mean errors.[14] The only difference between equation 13 and the dynamic model (eq. 7) is that the former contains a lagged $\mathbf{x}$ term, although the coefficient on this term is constrained. In this sense, we can see LDV and AR1 as differing in the constraint each imposes on the coefficient of $\mathbf{x}_{i,t-1}$; the dynamic model assumes this coefficient is zero while the static model with serially correlated errors assumes it is the negative product of the other model coefficients. Both

[13] The argument here is identical to that for a single time series. While we believe that the approach we propose is consistent with modern "London School of Economics" econometrics (Gilbert 1990), this is not the place to write an article on such practice. We rely here on Harvey's (1990) standard text.

[14] This is derived by writing equation 1 for time period $t-1$, solving for $\epsilon_{i,t-1}$ and then substituting this in equation 1 for time period t.

of these models can be seen as special cases of the more general model:

$$y_{i,t} = \phi y_{t-1} + \mathbf{x}_{i,t}\beta + \mathbf{x}_{i,t-1}\gamma + \nu_{i,t}. \tag{14}$$

Thus, the two standard ways of treating the dynamics of TSCS data are both simplifications of a more complex model. While it is unlikely that either simplification is exactly correct, it will quite often be the case that either are good enough. While there are some theoretical justifications for the lagged dependent variable model (in terms of partial adjustment), in practice there is seldom a theoretical reason to prefer the dynamic model or the static model with serially correlated errors.[15]

We can, of course, test equation 7 to see if it appears adequate. In particular, researchers should test to see whether its residuals show serial correlation. The simplest way to do this is via a Lagrange multiplier test (Engle 1984). To test the null hypothesis that the errors from equation 7 are serially independent, we regress the residuals from an OLS estimation of equation 7 on the first lag of those residuals as well as all the independent variables used in the OLS estimation. The estimated coefficient on the lagged residual term yields an estimate of the remaining serial correlation of the errors. A test of the null hypothesis that this coefficient is zero is a test of the null hypothesis that the remaining errors are serially independent. If we reject that null hypothesis, OLS no longer would be the appropriate way to estimate equation 7. Fortunately, it is unlikely that there will be much, if any, remaining serial correlation of the errors after including a lagged dependent variable.[16]

---

[15] We limit our discussion to first-order dynamics. Most TSCS data consists of well under 50 annual observations, so it is unlikely that the data will clearly indicate that a more complicated dynamic structure is necessary. Most TSCS analysts implicitly assume that all dynamics are first-order. More complicated dynamics may become important if we move to models with quarterly or monthly data. Our proposed methodology generalizes straightforwardly to more complicated dynamics; the AR1 methodology does not. Much of the data of interest to TSCS analysts are naturally measured annually, and so, in general, simple dynamics will be appropriate. Nothing in this article argues against testing for higher-order dynamics; the methodology for doing this is a simple generalization of the methodology we discuss.

[16] If the errors show serial correlation in the presence of a lagged dependent variable, the standard estimation strategy is instrumental variables. While this has fine asymptotic properties, it may perform very poorly in practical research situations. Problems with instrumental variables arise because it is difficult to find instruments that are well correlated with the variables they are instruments for while themselves being uncorrelated with the errors. Thus, it may well be the case that it is better to estimate equation 7 with OLS, even in the presence of a small, but statistically significant, level of residual serial correlation of the errors. One advantage of the Lagrange multiplier test is that it provides an immediate estimate of the level of residual correlation. TSCS researchers can then examine that level and not simply assume that errors show nontrivial residual serial correlation based on a test of sig-

We can also test whether the dynamic model is an adequate simplification of the more general model, equation 14.[17] If the lagged dependent variable causes the remaining errors to be serially independent, this test can be done via the usual comparisons of sums of squared errors. TSCS researchers have, in our experience, not considered the more general model (eq. 14). It is surely worthwhile to permit this more general model to be chosen as the appropriate specification; if the more general model is appropriate, we need simply include a lagged $\mathbf{x}$ term in equation 7. This causes no new estimation problems, and so, when we refer to equation 7, it may contain lagged $\mathbf{x}$ terms.

It is also possible to design tests to compare equation 1 with serially correlated errors against the more general alternative of equation 14 through a common factors (COMFAC) test (Harvey 1990, 283–87). In practice, however, researchers who model the dynamics via serially correlated errors do not consider the more general alternative.[18] Thus, while equation 7 leads naturally to considering more general dynamic models, equation 1, with serially correlated errors, does not.

Note that equation 7 makes the dynamics of the model explicit. Equation 13 makes the dynamics explicit as well, but researchers tend in practice simply to transform away the serial correlation and then estimate equation 1. In this approach, the dynamics are simply a nuisance that leads to estimation difficulties; once those difficulties are dealt with, these analysts concentrate on the parameters of interest, namely, the β in equation 1. We think this ignores an important part of the model.[19]

Making the dynamics explicit has another important advantage; it allows us to explicitly consider issues of unit root TSCS data. Just as for a single time series, TSCS models have a unit root if the estimated value of ϕ in equation 7 is one. Little is known about unit roots in the

nificance. This is particularly important for relatively large values of N and T where small levels of residual serial correlation may be statistically significant.

[17] Given the relatively short time series of TSCS models, we would probably not need to consider lag lengths longer than one. If relevant, say, with quarterly or monthly data, this obviously could be done. The preferred procedure is to start with the most general lag structure and test simplifications.

[18] We have never seen a COMFAC analysis used in any political science TSCS analysis. This type of analysis appears rarely, if ever, in the analyses of single time series in political science.

[19] This point is well known in the single time series world (Hendry and Mizon 1978; Beck 1991) but is often ignored in the TSCS world. Thus, many researchers, modeling dynamics with serial correlation, do not even report their estimated serial correlation coefficients. Four of the six articles referenced in the introduction (Blais, Blake, and Dion 1993; Giles and Hertz 1994; Levobic 1994; Pollins 1989) did not report anything concerning estimation of the dynamics.

TSCS context.[20]

The underlying logic of unit roots and the corresponding ideas of modeling short- versus long-run phenomena, as well as error correction, however, may have an enormous payoff in the TSCS arena.[21] Many researchers estimate a model with serially correlated errors with little regard for the size of this serial correlation; in many data sets that we have seen, this correlation exceeds 0.9.[22] The correction for serially correlated errors in this case is effectively taking first differences. This may eliminate serially correlated errors, but it also leads to researchers estimating

$$\Delta y_{i,t} = \Delta \mathbf{x}_{i,t}\beta + \nu_{i,t}, \tag{15}$$

where $\nu_{i,t}$ is a zero mean, iid error process.

Equation 15 drops any attempt to model the long-run relationship between series and instead concentrates only on the short run. It is possible, just as for a single time series, to combine short- and long-run phenomena for TSCS models. The TSCS analogue of the error correction model is

$$\Delta y_{i,t} = \alpha + \Delta \mathbf{x}_{i,t}\beta + \phi(y_{i,t-1} - \mathbf{x}_{i,t-1}\gamma) + \epsilon_{i,t}. \tag{16}$$

If a panel model can be represented in error correction form, then β represents the short-term impact of $\mathbf{x}_{i,t}$, on $y_{i,t}$, and ϕ represents the long-term impact, that is, the rate at which $y_{i,t}$ and $\mathbf{x}_{i,t}$ return to their long term equilibrium relationship. When the residuals from

[20] The only paper we have seen is an unpublished work by Levin and Lin (1993). It should be noted that there are many different approaches to modeling "stochastic" trends in the single time series literature (Stock and Watson 1988) and each of these have an analogue in TSCS data. TSCS data contains relatively short time series; this may make it difficult to distinguish the various different approaches to modeling long-run behavior. It is also possible that the cross-sectional richness of TSCS data may offset the temporal poverty of those data. In any event, we do not consider these issues in this article but note that our proposed methodology both makes these issues more apparent and allows for their future consideration.

[21] See the articles in Engle, and Granger (1991) for a discussion of these issues in the single time series context.

[22] For example, the reader of Blais, Blake, and Dion's (1993) study of the political determinants of government spending would never know that they corrected for a level of serial correlation that *probably* exceeded 0.9. Thus, in effect, while the reader of Blais, Blake, and Dion sees equations relating spending to the left-right composition of the government, the near first differencing reduces the analysis to one of only short-run effects, that is, the effect of changes in government on changes in spending. This short-run model may be of interest, but it is not the model explicitly formulated by them. We would have liked to redo their analysis to see the effects of modeling both long- and short-run effects, but, unfortunately, we were not able to obtain these data.

estimating equation 1 show a level of serial correlation near one, or when the estimate of ϕ in equation 7 is near one, it is critical to separate the short- and long-run impacts of the independent variables on $y_{i,t}$ in this manner. This situation appears to arise commonly in TSCS data, especially in the comparative political economy arena.

In short, both the dynamic model and the static model with serially correlated errors will often be consistent with the data. The dynamic model causes researchers to think about the dynamics of their model, while the static model with serially correlated errors leads to researchers thinking of dynamics as a nuisance that causes estimation problems. The former is clearly preferred. This preference is based on the general advantages of fully dynamic models, although the data, of course, may better fit the static model with serially correlated errors. To investigate whether either specification performs better in estimating parameters, we turn to Monte Carlo experiments.

Monte Carlo Experiments

The setup of these experiments is similar to those reported in Beck and Katz 1995. The first step in the simulation was to generate the data. We ran experiments with various combinations of N and T chosen to reflect values typically found in TSCS studies. For a given N and T, then, observations on a single independent variable, $x_{i,t}$ ($i = 1, \ldots, N$; $t = 1, \ldots, T$) were generated and fixed over the 1,000 replications used in the experiment. Since the performance of time-series estimators varies with the level of trend in the data, we could not simply generate the $x_{i,t}$ as independent draws from a normal. Instead, we used

$$x_{i,t} = \delta x_{i,t-1} + \mu_{i,t}, \tag{17}$$

where each $\mu_{i,t}$ is drawn from an iid standard normal. By varying δ, we could adjust the amount of trend in $x_{i,t}$. One problem with simulating data according to equation 17 is that we require knowledge of $x_{i,0}$. We dealt with this problem by generating $T + 50$ observations on each time series (with $x_{i,-50} = 0$), retaining only the T final observations for use in analysis.

Once the independent variable was created, we needed to create NT observations of the dependent variable for each of the one thousand replications of the experiment. We chose to use equation 14 for generating the $y_{i,t}$, since it encompasses both models of interest as special cases.

Hence,

$$y_{i,t}^{(l)} = \alpha + \phi y_{t-1} + \beta x_{i,t} + \gamma x_{i,t-1} + \nu_{i,t}^{(l)}; \quad \begin{array}{l} i = 1, \ldots, N; \\ t = 1, \ldots, T; \\ l = 1, \ldots, 1{,}000, \end{array} \tag{18}$$

where both α and β were fixed at 10 over all experiments. Since the statistical evaluations of AR1 and LDV are independent of any cross-sectional complications in the data, for the current experiments we assumed that the data were both spatially independent and homoskedastic. More formally, the $\nu_{i,t}^{(l)}$ were generated by independent draws from a normal distribution with mean zero and a variance set such that the parameters were about twice their standard errors. This simplification leads to no loss of generality.

By varying the coefficient γ, we could generate data to appear as though they came from equation 7 ($\gamma = 0$) or equation 1 with serially correlated errors ($\gamma = -\phi\beta$). Obviously LDV will outperform AR1 when the data are generated to follow equation 7, with the reverse occurring when the data are generated statistically with serially correlated errors. We therefore generated data that fell between these two models. More specifically, we drew γ on each iteration of the experiment from a normal distribution with mean $\gamma = -\rho\beta/2$ and experimentally manipulated variance.

The LDV estimate of β for replication l is referred to as $\beta_{\text{LDV}}^{(l)}$; the AR1 estimate for that replication is $\beta_{\text{AR1}}^{(l)}$. 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 1,000 $\beta_{\text{AR1}}^{(l)}$ or $\beta_{\text{LDV}}^{(l)}$. The quality of the LDV or AR1 estimates of variability can then be assessed by comparing the root mean square average of the 1,000 estimated standard errors with the corresponding standard deviation of the 1,000 estimates. The measure of accuracy we focus on, "overconfidence," is the percentage by which, say, AR1 understates variability; that is,

$$\text{Overconfidence} = 100 \frac{\sqrt{\sum_{l=1}^{1{,}000} (\beta_{\text{AR1}}^{(l)} - \overline{\beta_{\text{AR1}}})^2}}{\sqrt{\sum_{l=l}^{1{,}000} [\text{SE}(\beta_{\text{AR1}}^{(l)})]^2}}. \tag{19}$$

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

We were also interested in the relative efficiency of AR1 and LDV. Since the true value of β is known, the root mean square error of the LDV and AR1 estimates of β can be calculated. The relative efficiency of LDV, as compared with AR1, is given by

$$\text{Efficiency} = 100 \frac{\sqrt{\sum_{l=1}^{1,000}(\beta_{\text{AR1}}^{(l)} - \beta)^2}}{\sqrt{\sum_{l=1}^{1,000}(\beta_{\text{LDV}}^{(l)} - \beta)^2}}. \tag{20}$$

Efficiency greater than 100 percent indicates that LDV is superior, in mean square error terms, to AR1.

Monte Carlo Results

We ran many experiments of this type. Table 1 shows results for experiments with $N = 15$ and $T = 20$ or 40. In this table we report two sets of experiments, with the variance of γ set first to one-half and then to one. We also experimentally varied ϕ and δ. Since our interest is in estimating β rather than the dynamics, we report only statistics that address the quality of this estimate. (Both LDV and AR1 provided good estimates of the dynamic parameter, ϕ.)[23]

With the smaller T of 20, LDV appears clearly superior to AR1. In the worst case, LDV standard errors are overconfident by 25 percent. This degree of overconfidence occurred only for strongly trending data ($\delta = .75$). For less strongly trending data, LDV estimates of variability were accurate to within 10 percent or better. AR1 estimates of variability were relatively accurate with little trend in the data ($\delta = .25$), though never more accurate than LDV estimates of variability. When the data showed stronger trend however, AR1 estimates of variability were highly overconfident; with very strong trend ($\delta = .75$) this overconfidence would make AR1 estimates of variability useless. These problems persist with the simulations based on the larger T of 40. Interestingly, estimated standard errors do not become more accurate with the increase in the size of the time sample. It should be remembered that our time samples are small, and so no asymptotic results apply to them.

Turning to efficiency comparisons, LDV is generally more efficient than AR1. In the smaller time sample, LDV is sometimes more than four times as efficient as AR1, while it is never worse than a third less efficient than AR1. This pattern persists for the larger time sample,

[23] Unlike our other Monte Carlo experiments, the results varied with choice of parameter. While the qualitative results are robust, quantitative results were more variable than we would have liked. Nothing in the unreported results would lead us or anyone else to prefer AR1 to LDV, however.

although there are cases in which AR1 is dramatically more efficient than LDV. (With strongly trending data and a large value of ϕ, AR1 is about three times more efficient than LDV.) But for every case in which AR1 is much more efficient than LDV, there is another case in which this relationship is reversed.

It should be remembered that these are comparisons of two estimators using data generated when both should be at least somewhat wrong. We did not use our simulations to see what would have happened if we had estimated the correct model, equation 14, since obviously such a model would have estimated the parameters quite well. Fortunately, in most (but not all) of the experiments in which AR1 was more efficient than LDV, the Lagrange multiplier test indicated that LDV was not the appropriate estimation method.[24] There is no common analogue to this test for AR1 estimation. Thus, even when LDV performed less well than AR1, it often would have been possible to see that LDV should have been discarded in favor of estimating the correct model, equation 14.

At worst, these results indicate no clear statistical preference for AR1 over LDV and, with some exceptions, a statistical preference for LDV over AR1. Given the other advantages of LDV, we therefore suggest that TSCS analysts should generally begin with equation 7. This should be estimated by OLS and then tested to ensure that the residuals are serially independent. In practice we find that this is usually the case. If not, it is easy to include additional lags of the independent variables in the specification, to see if the more complicated model, equation 14, is necessary. If the coefficient on the lagged dependent variable is near one, we propose that TSCS analysts consider the TSCS equivalent of error correction, equation 16.

Correcting for Heteroskedasticity: Panel Weighted Least Squares

The CHTA correction for panel heteroskedasticity is "panel weighted least squares" (PWLS).[25] PWLS can be used on either equation 1 or 7. Since we are not concerned with dynamics in this section, we work with estimates of the parameters of equation 1 with serially independent, but panel heteroskedastic, and possibly contemporaneously correlated, errors. If $e_{i,t}$ is the vector of OLS residuals from estimating equation 1,

[24] This is indicated by the proportion of times this test rejected the null hypothesis of no serial correlation.

[25] PWLS is seldom, if ever, used by itself. It is, however, almost invariably used as part of CHTA. The separability of the CHTA temporal and cross-sectional corrections allows us to analyze the two corrections separately. We thus use the acronym PWLS for convenience; in actual use it is simply one component of CHTA.

TABLE 1. Comparison of Estimating Dynamic vs Static Model with Serially Correlated Errors (N=15)

T	δ[a]	ϕ[b]	VAR(γ)[c]	Optimism[d] LDV	Optimism[d] AR1	Efficiency[e]	LM[f]
20	0.25	0.25	0.50	100	104	114	3
20	0.25	0.50	0.50	99	100	192	5
20	0.25	0.75	0.50	93	113	406	18
20	0.50	0.25	0.50	103	175	267	3
20	0.50	0.50	0.50	104	117	186	5
20	0.50	0.75	0.50	100	79	93	19
20	0.75	0.25	0.50	108	345	446	3
20	0.75	0.50	0.50	110	222	452	5
20	0.75	0.75	0.50	112	81	76	22
20	0.25	0.25	1.00	102	106	116	3
20	0.25	0.50	1.00	101	106	192	5
20	0.25	0.75	1.00	94	140	410	18
20	0.50	0.25	1.00	108	222	284	3
20	0.50	0.50	1.00	108	128	190	5
20	0.50	0.75	1.00	104	84	95	18
20	0.75	0.25	1.00	118	470	459	4
20	0.75	0.50	1.00	117	288	462	4
20	0.75	0.75	1.00	125	87	79	22
40	0.25	0.25	0.50	103	101	104	3
40	0.25	0.50	0.50	102	117	241	6
40	0.25	0.75	0.50	95	158	327	34
40	0.50	0.25	0.50	107	176	269	3
40	0.50	0.50	0.50	107	91	67	5
40	0.50	0.75	0.50	104	108	131	32
40	0.75	0.25	0.50	113	390	496	3
40	0.75	0.50	0.50	114	205	346	5
40	0.75	0.75	0.50	123	70	30	28
40	0.25	0.25	1.00	106	101	104	3
40	0.25	0.50	1.00	105	138	250	6
40	0.25	0.75	1.00	99	209	337	33
40	0.50	0.25	1.00	114	224	284	3
40	0.50	0.50	1.00	113	90	67	5
40	0.50	0.75	1.00	115	133	135	32
40	0.75	0.25	1.00	127	530	509	3
40	0.75	0.50	1.00	126	270	358	5
40	0.75	0.75	1.00	146	71	30	28

[a] From equation 17.

[b] From equation 14.

[c] Variance of generated γ.

[d] From equation 19.

[e] From equation 20.

[f] Proportion of replications where LM test for serial correlation rejects null hypothesis.

PWLS estimates the error variances by

$$\widehat{\sigma_i^2} = \frac{\sum_{t=1}^{T} e_{i,t}^2}{T}. \tag{21}$$

PWLS then weights each observation by the inverse of the $\hat{\sigma}_i$ and performs a second round of OLS.[26]

We would not expect PWLS to be subject to the pathologies we found in the much more complicated Parks method. While PWLS standard errors do not take into account the extra variability resulting from using estimated variances, these variances are reasonably well estimated for TSCS data with long time samples. There are no analytic results on what T is large enough to allow PWLS standard errors to be acceptably accurate. Our Monte Carlo experiments shed some light on this issue.

TSCS researchers who opt for PWLS should realize that PWLS is different from weighted least squares (WLS) for purely cross-sectional data. Cross-sectional WLS uses a theoretical relationship between some variables and the unknown error variance to weight the regression function; PWLS uses the estimated unit error variances for the same purpose. PWLS is therefore much closer to being the panel analogue of robust estimators, such as iteratively reweighted least squares, than it is of WLS. While empirical weighting techniques are quite popular in the statistics literature, they are almost unused in political science. Why should we use PWLS with TSCS data while ignoring its counterpart for cross-sectional data?

PWLS weights each unit by how well it fits equation 1. This may be problematic if there is any unit-to-unit variation in the model parameters. In this case PWLS will give more weight to the unit with the lowest estimated error variance. If the variation from unit to unit is random, this will lead to less efficient estimation. Thus, suppose the data are generated according to:

$$y_{i,t} = x_{i,t}\beta_i + \epsilon_{i,t}, \tag{22}$$

where the β_i are random draws from some distribution with mean β. Such an assumption seems not implausible.[27] If our interest is in estimating the mean β, PWLS may be less efficient than OLS.

[26] The choice of divisor here is irrelevant, since PWLS depends only on relative weights. These relative weights are completely determined by the numerator. In particular, none of the disadvantages shown by PWLS in the Monte Carlo experiments is in any manner a consequence of our choice of divisor in equation 21.

[27] Random coefficient models like this can be estimated in the TSCS context (Hsaio 1986, 131-38). Here we are simply interested in the consequence of random parameter variation for OLS and PWLS.

We seek to compare the performance of OLS/PCSE and PWLS in actual research situations. To that end we again turn to Monte Carlo experiments. We conducted two different Monte Carlo studies: the first compares the performance of the two estimators with homogeneous parameters, while the second assesses their relative performance in the face of parameter heterogeneity.

Monte Carlo Experiments

We were interested in both the performance of PWLS standard errors and the relative efficiency of PWLS as compared with OLS. The basic setup of the experiments is similar to that of the dynamic experiments: an independent variable was generated, 1,000 replications of an error process were generated, and these were used to generate 1,000 replications of the dependent variable. Since we are interested here in the performance of estimators in relation to cross-sectional issues, we generated the independent variable and errors with no dynamics. Thus, our experiments correspond to either the second stage of CHTA or the estimation of the dynamic model with serially independent errors.

While these experiments were temporally simple, they were quite complicated spatially. This is because OLS (and PWLS) standard errors will be accurate if the heteroskedasticity or contemporaneous correlation of the errors is unrelated to the structure of the independent variables.[28] Thus, we had to generate the independent variable so that the $x_{i,t}$ showed correlation between units as well as unequal variances among those units.

To accomplish this, for every time period, t, we drew $x_{i,t}$ from an N-variate normal. The first half of the marginal normals was assigned a variance of one, while the second half was assigned some other value, which was experimentally manipulated. For ease of exposition, we forced all unit correlations to be equal. We thus assumed that the correlation matrix of the N-variate normal was:

$$\mathbf{R} = \begin{pmatrix} 1 & r & \dots\dots & r \\ r & 1 & r \;\; . & r \\ \dots & \dots & \dots\dots & \dots \\ r & \dots\dots & r & 1 \end{pmatrix}. \tag{23}$$

These simplifications make it much easier to present our results without loss of generality.

For each of the 1,000 replications, errors $\epsilon_{i,t}$ were drawn from an N-variate normal distribution that was proportional to that which was

[28]This can be seen either by examining equation 9 or by extrapolating White's discussion (1980) to the TSCS case.

used to generate the $x_{i,t}$. (The constant of proportionality was chosen so that coefficients were between two and three times their estimated standard errors.) The errors thus showed both panel heteroskedasticity and contemporaneous correlation of the errors.

The value of r in equation 23 measures the level of contemporaneous correlation of the errors. We have seen no textbook treatment of the degree of panel heteroskedasticity. We measure it by the standard deviation of the normalized weights that would be used in PWLS.[29]

For the homogeneous parameter case, we generated the 1,000 replications of the $y_{i,t}$ according to

$$y_{i,t}^{(l)} = \alpha + \beta x_{i,t} + \epsilon_{i,t}^{(l)}; \qquad \begin{aligned} i &= 1, \dots, N; \\ t &= 1, \dots, T; \\ l &= 1, \dots, 1,000, \end{aligned} \tag{24}$$

where both α and β were set to 10.

We proceeded similarly for the experiments to assess the performance of PWLS with heterogeneous parameters, except that the 1,000 replications of $y_{i,t}$ were generated according to

$$y_{i,t}^{(l)} = \alpha + \beta_i x_{i,t} + \epsilon_{i,t}^{(l)}; \qquad \begin{aligned} i &= 1, \dots, N; \\ t &= 1, \dots, T; \\ l &= 1, \dots, 1,000, \end{aligned} \tag{25}$$

where α was fixed at 10 and the β_i were independent draws from a normal distribution with a mean of 10 and variance experimentally manipulated.[30]

We assessed the overconfidence of estimated standard errors by means of equation 19, substituting PWLS for AR1 and OLS/PCSE for LDV. Efficiency for homogeneous parameters is also computed as in the first experiment, using equation 20. This was slightly modified for the heterogeneous parameter experiment, since for any given replication the average realized β_i might differ from 10. We therefore computed effi-

[29] The weight for the unit i is $w_i = 1/\sigma_i$. We define "standardized heteroskedasticity" as the standard deviation of the $w_i/\bar{w}$. This measure varies from zero for homoskedasticity to just under one.

[30] We chose the normal for simplicity. There is no constraint on the β_i, so heterogeneity distributions like the gamma are not appropriate. The shape of the normal seems plausible for these experiments. For the purposes of these experiments we need not worry about whether normal heterogeneity yields a tractable mixing distribution.

ciency around this average realized $\bar{\beta}_i$, that is,

$$\text{Efficiency} = 100 \frac{\sqrt{\sum_{l=1}^{1,000}(\beta_{\text{PWLS}}^{(l)} - \bar{\beta}_i)^2}}{\sqrt{\sum_{l=1}^{1,000}(\beta_{\text{OLS}}^{(l)} - \bar{\beta}_i)^2}}. \tag{26}$$

As before, efficiency greater than 100 percent indicates that OLS is superior, in mean square error terms, to PWLS.

Monte Carlo Results

The first experiments assessed PWLS and OLS for the homogeneous parameter case. Results of these experiments for $N = 15$ are shown in table 2.[31] PCSEs perform well. For experiments with $T \geq 10$ the PCSEs are within 10 percent of true variability. Even when T is as small as five, the PCSEs are never more than 25 percent off, even under conditions of extreme heteroskedasticity and contemporaneous correlation of the errors.[32]

Beginning with the case in which the errors were generated to be heteroskedastic but uncorrelated, PWLS standard errors are no better and often worse, than PCSEs; for $T = 5$ PCSEs were about 40 percent more accurate; this advantage declined to 20 percent for $T = 10$ and then essentially disappeared. But even for large T, large heteroskedasticity and no contemporaneous correlation of the errors, PCSEs were slightly more accurate than PWLS standard errors. These experiments show that, while PWLS should not be used with extremely short time samples, it appears to be acceptably accurate for time samples of 20 years or more *if the errors show only panel heteroskedasticity.*

When we induced contemporaneous correlation of the errors as well as heteroskedasticity, the advantage of PCSEs became marked. With moderate (0.25) contemporaneous correlation of the errors, PWLS standard errors were about 50 percent overconfident. PWLS standard errors were over 200 percent overconfident with unrealistically high contemporaneous correlations of 0.50. PWLS was not designed to correct for contemporaneously correlated errors, and its estimated standard errors are clearly inaccurate in the presence of such correlation.

It should be noted that these experiments probably overstate the degree of overconfidence of PWLS standard errors that would be observed

[31] Results for other experiments, not reported here, were very similar to those reported in table 2.

[32] We have examined many TSCS data sets. Heteroskedasticity of 0.3 is large, and heteroskedasticity over 0.5 is rare; similarly, we have not seen average contemporaneous correlation of the errors approach 0.5.

in actual data. This is because we simulated our errors so that they would show a covariance structure strongly related to the covariance structure of the independent variables. If there is less of a relationship between the structure of the errors and the structure of the independent variables in the actual data, then PWLS standard errors will be less overconfident than they are in our simulations. We have not seen PWLS analysts report any statistics that would allow them or their readers to assess how well their standard errors might perform.

In any event, we would expect that the structure of the independent variables and the errors will be related in actual data. Consider the case of cross-national political economy data. Why should the errors of, say, Belgium and the Netherlands covary differently from the independent variables for those two countries? This, combined with high overconfidence of PWLS standard errors in our simulated data, suggests that PWLS standard errors cannot be taken as reliable guides to the sampling variability of parameter estimates in the presence of contemporaneously correlated TSCS data.

We also examined the performance of the traditional White (1980) heteroskedasticity consistent standard errors (which do not take into account the panel nature of the heteroskedasticity). These standard errors are what would be computed by the "robust" option in TSCS modules such as SHAZAM. Their overconfidence is reported in the column labeled "White." They performed almost identically to PCSEs in the presence of only heteroskedasticity, but were markedly inferior to PCSEs (although superior to PWLS standard errors) in the presence of contemporaneously correlated errors. These results clearly show that PCSEs are preferred to White's heteroskedasticity consistent standard errors for TSCS data.

Given the cost of using PWLS in terms of inaccurate standard errors, is there a compensating efficiency gain in using PWLS over OLS? Table 2 shows that, in the absence of contemporaneously correlated errors, OLS is just about as efficient as PWLS in the presence of low heteroskedasticity. As heteroskedasticity increases, PWLS does become more efficient than OLS, being about 10 percent more efficient when standardized heteroskedasticity is 0.30 and about 20 percent more efficient when heteroskedasticity rises to 0.50, a figure higher than we have seen.

This advantage disappears when heteroskedasticity is joined with contemporaneous correlation of the errors. With very high heteroskedasticity, and very high contemporaneous correlation of the errors, OLS and PWLS are equally efficient. More importantly, PWLS is only a few percentage points more efficient than OLS when the errors show even

TABLE 2. Overconfidence and Efficiency of OLS and PWLS: N = 15

T	Het.	Correlation	Overconfidence[a]			Efficiency[c]
			White[b]	PCSE	PWLS	
5	0.00	0.00	103	104	138	110
5	0.00	0.25	135	113	172	106
5	0.00	0.50	221	124	265	105
5	0.15	0.00	102	103	138	107
5	0.15	0.25	121	112	171	105
5	0.15	0.25	210	123	264	105
5	0.30	0.00	103	104	139	99
5	0.30	0.25	126	110	172	98
5	0.30	0.50	192	122	262	102
5	0.50	0.00	103	104	141	91
5	0.50	0.25	123	109	173	91
5	0.50	0.50	174	119	257	98
10	0.00	0.00	103	102	118	106
10	0.00	0.25	137	105	156	105
10	0.00	0.25	223	108	246	105
10	0.15	0.00	102	102	119	102
10	0.15	0.25	119	105	155	104
10	0.15	0.50	214	108	245	105
10	0.30	0.00	102	102	119	94
10	0.30	0.25	126	105	155	98
10	0.30	0.50	195	108	243	102
10	0.50	0.00	101	102	120	86
10	0.50	0.25	121	105	154	90
10	0.50	0.50	174	107	240	98
20	0.00	0.00	96	96	103	103
20	0.00	0.25	146	101	155	102
20	0.00	0.50	231	104	242	102
20	0.15	0.00	95	95	103	99
20	0.15	0.25	140	100	154	102
20	0.15	0.50	219	103	241	103

[a] Percentage from equation 19.
[b] Heteroskedasticity Consistent Standard Errors.
[c] Percentage from equation 20.

TABLE 2. – Continued

T	Het.	Correlation	Overconfidence[a]			Efficiency[c]
			White[b]	PCSE	PWLS	
20	0.30	0.00	96	96	103	90
20	0.30	0.25	132	99	154	98
20	0.30	0.50	199	103	240	102
20	0.50	0.00	103	103	103	80
20	0.50	0.25	125	98	153	93
20	0.50	0.50	180	102	239	101
30	0.00	0.00	102	101	106	102
30	0.00	0.25	149	106	155	101
30	0.00	0.50	230	107	237	101
30	0.15	0.00	102	102	106	97
30	0.15	0.25	144	105	154	100
30	0.15	0.50	220	107	237	102
30	0.30	0.00	102	102	107	89
30	0.30	0.25	136	104	154	97
30	0.30	0.50	202	106	237	101
30	0.50	0.00	103	106	107	80
30	0.50	0.25	129	103	154	92
30	0.50	0.50	183	106	236	99
40	0.00	0.00	105	104	109	102
40	0.00	0.25	143	104	149	102
40	0.00	0.50	222	105	228	101
40	0.15	0.00	104	103	109	99
40	0.15	0.25	139	103	149	101
40	0.15	0.50	213	104	228	102
40	0.30	0.00	102	102	109	91
40	0.30	0.25	132	102	149	97
40	0.30	0.50	194	104	227	101
40	0.50	0.00	101	101	109	82
40	0.50	0.25	125	101	149	91
40	0.50	0.50	176	103	227	99

[a] Percentage from equation 19.

[b] Heteroskedasticity Consistent Standard Errors.

[c] Percentage from equation 20.

moderate contemporaneous correlation. In short, PWLS is only more efficient than OLS in the presence of high heteroskedasticity and low contemporaneous correlation of the errors. In other cases, OLS is either as efficient or almost as efficient as PWLS without suffering from the overoptimistic errors of PWLS.

Fortunately, researchers do not have to guess about whether PWLS might be superior for their own research. The OLS residuals from equation 1 or 7 can be used to estimate both the heteroskedasticity and contemporaneous correlation of the errors.[33] At that point researchers can consult table 2 to decide whether the efficiency advantage of PWLS is sufficiently great to offset its more inaccurate standard errors.

We argued that PWLS might be vulnerable to parameter heterogeneity. Table 3 reports the relative efficiency of PWLS and OLS in the presence of heterogeneous parameters. These experiments induced only small amounts of heterogeneity in β. While we do not know how heterogeneous parameters are in actual data, we do not believe that these experiments represent unrealistic cases.[34]

To make matters as favorable as possible for PWLS, we considered data in which the errors were heteroskedastic but contemporaneously independent. The previous experiment tells us that PWLS will be more efficient than OLS given high heteroskedasticity and parameter homogeneity. This advantage holds us in the presence of small amounts of heterogeneity. But when the parameters are drawn from a normal with a standard deviation of as much as 0.5 (i.e., most β_i were within about 10 percent of their mean), OLS becomes as efficient as PWLS, with its advantage increasing somewhat with increasing heterogeneity. Thus, the advantage of PWLS over OLS, even when the errors there are highly heteroskedastic and contemporaneously independent, is eliminated by a small amount of parameter heterogeneity. While we cannot know if actual research situations will show this amount of parameter heterogeneity, it does not seem counterintuitive that unit β_i will vary from the overall average β by 10 to 20 percent.

We can sum up all the evidence presented to conclude that, while PWLS standard errors may well be overconfident, OLS/PCSEs perform well. Further, viewed in the most favorable light, the efficiency advantage of PWLS over OLS is, at best, slight. Only in situations where the errors indicate extremely high heteroskedasticity and contemporaneously uncorrelated errors should PWLS even be considered. In such

[33] The RATS program we wrote to compute PCSEs does these computations.

[34] In the highest variance experiment, almost all the unit β_i are within 20 percent of the mean β; for the other two experiments almost all the β_i differ from the mean β by less than 15 or 10 percent, respectively.

TABLE 3. Efficiency of OLS and PWLS: Heterogeneous Parameters ($N = 15$, no contemporaneous correlation of errors)

T	Heteroskedasticity	Parameter Heterogeneity[a]	Efficiency[b]
20	0.00	0	103
20	0.15	0	100
20	0.30	0	91
20	0.50	0	81
20	0.00	0.25	103
20	0.15	0.25	101
20	0.30	0.25	95
20	0.50	0.25	88
20	0.00	0.50	104
20	0.15	0.50	104
20	0.30	0.50	101
20	0.50	0.50	98
20	0.00	1	106
20	0.15	1	110
20	0.30	1	111
20	0.50	1	107
40	0.00	0	102
40	0.15	0	97
40	0.30	0	89
40	0.50	0	81
40	0.00	0.25	102
40	0.15	0.25	99
40	0.30	0.25	95
40	0.50	0.25	94
40	0.00	0.50	103
40	0.15	0.50	105
40	0.30	0.50	106
40	0.50	0.50	106
40	0.00	1	106
40	0.15	1	115
40	0.30	1	118
40	0.50	1	112

[a] Standard deviation of distribution from which β_i are drawn.
[b] From equation 26.

cases it would probably be wise to attempt to theoretically reduce the heteroskedasticity rather than doing so empirically via PWLS.[35]

Our recommendation is that TSCS researchers use OLS in preference to PWLS unless there are strong grounds for choosing PWLS. While this recommendation is supported by our experiments, it is also supported by a preference for using simple, well-understood methods like OLS. As soon as we start using empirical weights, we have less understanding of what drives the final estimates. Complicated estimation schemes are sometimes necessary, but they should only be used when necessary and when the gains from complication exceed the costs of moving away from well-understood methods. This will seldom, if ever, be the case with PWLS.

Finally, even if researchers choose to use PWLS, they should at least report the estimated weights. None of the six analyses cited in the introduction made such a report. As we shall see in the next section, the weighted sample may look rather different from the original one. If weighting is important, the weights are interesting model parameters and should be reported like any other estimate.

Reanalysis

We use our proposed methodology to reanalyze the findings of Burkhart and Lewis-Beck 1994 (hereafter BLB) about the relationship of economic development and democracy. They analyzed annual data from 1972 through 1989 on 131 countries. Their measure of democracy is an index compiled by Raymond Gastil, which runs from two to 14; their measure of economic development is the common log of energy consumption per capita. They initially estimated the dynamic form, which we prefer. Included in this equation were, in addition to energy consumption, two interaction terms between energy consumption and dummy variables measuring whether a nation was on the periphery or the semiperiphery of the world system.[36] The relationship between economic development and democracy is one of the oldest empirical questions in comparative politics. BLB report very strong evidence for such a relationship.

BLB appropriately tested to see whether the residuals from the dynamic estimation were serially correlated. Based on Durbin's h-test they concluded that there was serial correlation of the errors. They therefore

[35] Stimson (1985) argues for including unit dummy variables to reduce heteroskedasticity. It is also sometimes possible to change dependent variables, say, from the level of spending to spending as a proportion of GDP. Reducing heteroskedasticity by respecification clearly deals with estimation problems in a manner that is both theoretically and econometrically justified.

[36] We used exactly the data provided by BLB and had no difficulty reproducing their basic regression results. All details on measurement may be found in BLB.

TABLE 4. Estimates of Burkhart and Lewis-Beck Model
Dependent variable: Gastil Democracy Index, 131 countries, 1972–89

	CHTA[a]		OLS			PWLS	
	b	SE	*b*	SE	PCSE	*b*	SE
Democracy_1	.09	.02	.96	.006	.012	.995	.002
Energy	2.49	.22	.14	.033	.045	.016	.007
Energy × SP[b]	−1.33	.19	−.034	.021	.022	.006	.005
Energy × P[c]	−1.54	.18	−.042	.022	.019	−.008	.007
Constant	.35	.06	−.005	.085	.099	.004	.019
ρ	.90						
Error Corr.[d]			−.042			.023	
$\hat{\sigma}$			.90			.92	

[a] From Burkhart, and Lewis-Beck (1994, 906).
[b] Semiperiphery.
[c] Periphery.
[d] From Lagrange multiplier test.

used an instrumental variables estimation to estimate the level of serial correlation, transformed the data based on this estimate, and then used OLS on the transformed data.[37] We report their results in the first two columns of table 4.

We reestimated their model. There are a variety of asymptotically equivalent Lagrange multiplier tests of the null hypothesis of serially independent errors. Of these tests, the one that most strongly rejects this null has a TR^2 of 4.013, which, with one degree of freedom, has a P value of .045. We note that our test statistics are rather different from the Durbin's h reported by BLB (7.87), which was based on estimation of unit-specific serial correlations.[38] We have argued that the estimates of unit-specific serial correlation will be misleading. We also have argued that allowing for unit-specific serial correlation makes little theoretical sense. We can see this very clearly in the BLB model. While they use unit-specific ρ in their testing procedure, their substantive model assumes both a common ρ and a common ϕ.

But, regardless of any test, the estimated (*common*) serial corre-

[37] While BLB indicate that they corrected their estimates for panel heteroskedasticity, our reanalysis indicates that they did not do so. Burkhart, and Lewis-Beck (1994, 905) state that "heteroskedasticity was corrected ... with the 'force homoskedastic model' option in Microcrunch . . . ." This confusingly named option does not produce weighted least squares. The Microcrunch *User's Guide* states: "The normal specification for the GLS-ARMA model is heteroskedastic error (i.e., the estimator includes a weighted least squares analogue). . . . Users may override that default by specifying 'Y' to a Homoskedastic Error prompt. . . ." (Atunes and Stimson 1988, 47).

[38] This information was provided in a personal communication by Ross Burkhart.

lation of the residuals, as shown in the Lagrange multiplier auxiliary regression, is –0.04. This low level of serial correlation only approaches statistical significance because of BLB's huge sample size. They chose to use instrumental variables to eliminate the small problems caused by a trivial amount of serial correlation of the errors. Instrumental variables is problematic insofar as the instruments and the original variables are not well correlated. The instruments used by BLB explain only half the variance of the original variables (Burkhart, and Lewis-Beck 1994, 909). Thus, the cure of instrumental variables for this data set is almost surely worse than the mild illness of possibly slightly serially correlated errors.

Given the small level of serial correlation of the OLS errors and the inefficiencies introduced by instrumental variables, we chose to use the OLS parameter estimates. It should also be noted that OLS allows us to retain one additional year's worth of observations, giving us a sample size about 6 percent larger than that of BLB. Our OLS estimates are in the middle columns of table 4. Since the errors were likely to show both heteroskedasticity and contemporaneous correlation, we report PCSEs as well as traditional OLS standard errors.[39]

Like BLB, we find a significant relationship between economic development and democracy, although the standard error on the energy coefficient in our analysis is relatively larger than that found by BLB. For the energy variable, the CHTA t-ratio is 11.4, while our corresponding ratio (using PCSEs) is 3.1. The larger standard errors on the periphery and semiperiphery interactions with energy consumption lead us, unlike BLB, to not reject the null hypothesis that there is no interaction between economic development, democracy, and position in the world system.[40]

Our estimated coefficient for the energy variable is much smaller than that reported by BLB. But note that the two coefficients have very different substantive interpretations. BLB are essentially working with a model in first differences (they transform by subtracting 90 percent of the previous observation from the current observation) where the lagged first difference has almost no effect. Thus, almost all the effect of a permanent change in economic development occurs instantaneously. Using our estimates, the initial rise in the democracy rating would be about one-twentieth as large as the BLB estimates, with subsequent

[39]The PCSEs are not all that different from the OLS standard errors, although the PCSE for energy consumption is about a third larger than the corresponding OLS standard error. It is impossible to know a priori when PCSEs and OLS standard errors will differ.

[40]The F-statistic for the test of the hypothesis that neither interaction belongs in the specification is 1.92, which yields a P value of .15 with 2 and 2,222 degrees of freedom.

years showing similar increases until a new equilibrium is reached many years later. The new equilibrium level of democracy would be similar to the first-year increase in democracy as given in the BLB model. Examination of the Gastil democracy index shows that our picture of extremely sluggish movement in that index is an accurate portrayal of the data.[41]

BLB did not actually use PWLS. Since they argued that it was important to correct for panel heteroskedasticity, it is interesting to see what would have happened had they done so. These estimates are in the last two columns of table 4. They show an integrated process with an estimate of the lagged democracy coefficient of .995. There is a huge amount of heteroskedasticity in these data (standardized value of 3.7). Are the PWLS estimates superior to the OLS estimates?

There are 12 nations with perfect Gastil scores of 14 each year. These, not surprisingly, are all OECD nations. These 12 nations have very small OLS residuals and so are weighted extremely heavily in the PWLS estimations. If we compute the weights of each nation in the PWLS estimates, we find that 75 percent of that weight belongs to these 12 nations. Thus, the PWLS results largely reflect the performance of 12 nations whose democracy index remained constant. This clearly accounts for the estimate of the lagged democracy coefficient being one. While we might be interested in the cause of advanced industrial democracy, the BLB article is about the role of economic development in worldwide democracy. The PWLS estimates are useless for this assessment. While OLS may not be as efficient as PWLS in the presence of large amounts of heteroskedasticity, the OLS estimates are not subject to this extreme weighting problem. OLS standard errors are inconsistent in the presence of panel heteroskedasticity, but our PCSEs correct this problem.

The most important lesson of this reanalysis is that it is important to look at the substantive import of both heteroskedasticity and serial correlation of the errors rather than treating them as nuisances that impede estimation. Using a complicated instrumental variables technique to correct for a small amount of serial correlation of the errors led Burkhart and Lewis-Beck astray. Had they actually corrected for heteroskedasticity, as they claimed to do, they would have been equally led astray. Our proposed simpler methodology does not lead investigators in the wrong direction.

[41] Our estimates are also internally consistent. In the BLB model, all determinants of democracy other than economic development have a slow, geometric impact while economic development has only an instantaneous effect on democracy. All the other determinants of democracy are contained in the BLB "error" term.

Conclusion

Kmenta's "cross-sectionally heteroskedastic timewise autocorrelated" method first corrects for serially correlated errors via transformation and then uses PWLS to correct for heteroskedasticity. Thus CHTA treats the dynamics as a nuisance rather than as an intrinsic part of the model. The PWLS portion of CHTA is usually no more efficient than OLS, and estimated standard errors from CHTA may be incorrect in the presence of contemporaneous correlation of the errors. In situations of high heteroskedasticity it may be the case that CHTA is more efficient than OLS; it is easy to see whether this is the case by examining the estimated unit variances. We have not seen such situations arise often in practice. But, even if such situations were to occur, parameter heterogeneity would eliminate whatever advantage CHTA may possess. CHTA does not produce dramatically incorrect estimates or standard errors. But each of its two components leads away from the data, allows researchers to ignore dynamic issues, and uses empirically derived weights for observations, possibly causing estimates to change in a manner that it is not easy for analysts to understand.

Our proposed method, ordinary least squares with panel corrected standard errors, is in general as efficient as CHTA and provides more accurate standard errors. But we prefer it not only because of its superior statistical properties but, rather, because it forces us to think about the dynamics of our models and it does not engage in empirical weighting of the data. We therefore believe it should be the method of choice for TSCS data.

Our proposed method is easy to use. The dynamics of TSCS data can usually be treated by including a lagged dependent variable in the specification. The resulting specification, equation 7, can then be estimated by OLS. The computed standard errors may well be incorrect in the presence of either panel heteroskedasticity or contemporaneous correlation of the errors. It is easy, however, to remedy this problem by computing PCSEs.

Researchers can test the adequacy of this setup in a variety of ways. Lagrange multiplier tests can be used to test for any remaining serial correlation of the errors as well as for whether a more general dynamic model, involving lagged independent variables, is required. Both tests involve regressions based on OLS residuals of the dynamic model and so are easy to compute. Researchers should not use instrumental variables to estimate the dynamic model unless the residuals show at least moderate serial correlation; researchers should estimate the dynamic model using OLS even if the residuals show statistically significant, but substantively small, amounts of serial correlation. The former situation should be rare.

Similarly, researchers can examine whether the errors show enough heteroskedasticity to make it worthwhile to consider PWLS. Rejection of the null hypothesis of homoskedasticity should not lead researchers to automatically use PWLS. Only when the level of heteroskedasticity is sufficiently high that the gains from PWLS might offset its costs should the weighting procedure even be considered. Even in such a case, the disadvantages of PWLS in the presence of contemporaneously correlated errors or parameter heterogeneity might lead researchers to avoid weighting. If weighting is used, researchers should carefully scrutinize and report the weights used.

Are we doing more than telling TSCS researchers to return to tried and true methods? Many TSCS analysts have modeled dynamics via a lagged dependent variable, and surely the most common method for analyzing TSCS data is OLS. Our proposed PCSEs are new, and our simple Lagrange multiplier test for serially correlated errors in the presence of lagged dependent variables, while an easy extension of its single time-series counterpart, has not been used before. But, while we extoll the statistical virtues of our new PCSEs, we are also happy to propose a very simple method for analyzing TSCS data.

Political methodology is moving to more and more sophisticated methods. In many cases this is entirely appropriate. But even when appropriate, this move often leads to the use of techniques that produce standard errors about which we have little, if any, intuition. Even where the complications are appropriate, the use of complicated methods makes it difficult for authors to communicate with readers. As methods become more and more complicated, and as software becomes more and more sophisticated, we may find ourselves running "options" that we really don't understand. But, perhaps most importantly, these complicated methods often move us away from looking at and thinking about the data.[42] We recommend that TSCS analysts begin with OLS estima-

[42] King (1990, 11) states this well: "Along the same lines, I propose a new statistical criterion that we should consider as important as any of the more usual ones. We should ask of every new estimator: 'What did it do to the data?' Statistical criteria such as consistency, unbiasedness, minimum mean square error, admissibility, etc., are all very important. . . . However, in the end, statistical analyses never involve more than taking a lot of numbers and summarizing them with a few numbers. Knowing that one's procedures meet some desirable statistical criterion is comforting but insufficient. We must also fully understand (and communicate) just what was done to the data to produce the statistics we report. In part, this is just another call for full reporting of statistical procedures, but it is also a suggestion that we hold off using even those statistical procedures that meet the usual statistical criteria until

tion of the dynamic specification using panel corrected standard errors. This makes it easy for researchers, and readers, to see "what was done to the data."

APPENDIX: TECHNICAL DETAILS ON PCSE'S

In this appendix we will first prove the consistency of PCSEs in the presence of contemporaneously correlated and/or panel heteroskedastic errors. We then turn to the question of why PCSEs produce accurate finite sample estimators of variability and why the Parks standard errors do not.

PCSEs Are Consistent

In order to prove consistency we need to make several (standard) assumptions.[43]

1. $y_{i,t} = \mathbf{x}_{i,t}\beta + \epsilon_{i,t}$.
2. $\mathbf{X}'\mathbf{X}/NT \xrightarrow{a.s.} M$, a nonsingular and finite matrix.[44]
3. $E[\mathbf{X}'\epsilon|\mathbf{X}] = 0$.
4. ϵ is distributed with zero mean and variance-covariance matrix $\mathbf{\Omega}$, where $\mathbf{\Omega}$ is finite and positive definite. In the case of panel data, $\mathbf{\Omega} = \mathbf{\Sigma} \otimes \mathbf{I}_T$. If $\mathbf{\Omega}$ is to be finite and positive definite, then, by standard properties of Kronecker products, $\mathbf{\Sigma}$ must also be finite and positive definite.

The first two assumptions guarantee the existence of the OLS estimate of β,

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

The third assumption guarantees unbiasedness of this estimate regardless of the covariance structure of the errors (White 1984, chaps. 2, 3).

We are, however, interested in the estimated standard errors of the parameters. The proof of consistency of our PCSEs is in two parts. We

we can show precisely and intuitively how the data are summarized. Developing estimators that are robust, adaptive, nonparametric, semiparametric, distribution free, heteroskedasticity-consistent, or otherwise unrestrictive is important, but until we clarify just what estimators like these do to our data, they are not worth using."

[43] These assumptions are much stronger than we need. We use them because they are more common in political science applications. For a more general set of conditions, see White (1984, theorem 6.3).

[44] It is more common to assume that the independent variable are fixed. Fixed regressors, however, is a very strong and implausible assumption given our usual models and data. $\mathbf{X}$ refers to the stacked matrix of independent variables.

first show that the covariance matrix of $\hat{\beta}$ is of the form in equation 9. We then show that the our estimate of $\mathbf{\Omega}$ is consistent and hence that the overall estimate of the covariance matrix of the parameters must be consistent.

PROPOSITION 1.

$$Var\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

Proof. Given (1) and the definition of $\hat{\beta}$,

$$\hat{\beta} = \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon.$$

Since β is nonstochastic, it does not contribute anything to the variance of $\hat{\beta}$. We also note that the second term has zero mean because $(\mathbf{X}'\mathbf{X})$ converges to a finite matrix by (2) and $\mathbf{X}'\epsilon$ has zero mean by (3). Using (4), we get:

$$\begin{aligned}\mathrm{Var}(\hat{\beta}) &= \mathrm{Var}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\epsilon\epsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E(\epsilon\epsilon')\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}.\end{aligned}$$

The problem is that we do not know $\mathbf{\Omega}$. The panel structure of the data yields $\mathbf{\Omega} = \mathbf{\Sigma} \otimes \mathbf{I}_T$. Given this, to estimate $\mathbf{\Omega}$ we only need to estimate $\mathbf{\Sigma}$, the $N \times N$ contemporaneous covariance matrix of the errors. Let ϵ_t be the N vector of unit errors at time period t. Thus,

$$\mathbf{\Sigma} = E(\epsilon_t \epsilon_t').$$

ϵ_t is not observed, but we do observe the residuals $\mathbf{e}_t = \mathbf{y}_t - \mathbf{X}\hat{\beta}$. Thus, we can estimate $\mathbf{\Sigma}$ by

$$\hat{\mathbf{\Sigma}} = T^{-1}\sum_{t=1}^{T} \mathbf{e}_t \mathbf{e}_t' = T^{-1}\mathbf{E}'\mathbf{E},$$

where $\mathbf{E}$ is the $T \times N$ matrix of residuals. □

PROPOSITION 2. *Given assumptions (1) to (4), then* $\hat{\mathbf{\Sigma}} \xrightarrow{a.s.} \mathbf{\Sigma}$. *Hence,* $\hat{\mathbf{\Omega}} \xrightarrow{a.s.} \mathbf{\Omega}$.

Proof. See White 1984, proposition 7.2.[45]

[45] The proof entails repeated use of one of the Laws of Large Numbers and some bounding conditions.

PROPOSITION 3. *Given assumptions (1) to (4), PCSEs are consistent estimates of the standard errors of $\hat{\beta}$.*

Proof. Follows directly from propositions 1 and 2.

PCSEs Have Good Finite Sample Properties

PCSEs are consistent. But why do they perform well in finite samples? While $\hat{\Omega}$ is a consistent estimator of Ω, it clearly is not an accurate small sample estimator. Unless T is much larger than N, $\hat{\Omega}$ estimates each of the elements of Ω with very few degrees of freedom.

The key to understanding why PCSEs perform well is to note that our interest is in estimating $\mathbf{X}'\Omega\mathbf{X}$ not Ω. To see why $\mathbf{X}'\hat{\Omega}\mathbf{X}$ provides good finite sample estimates of $\mathbf{X}'\Omega\mathbf{X}$, consider an even more extreme case, that of White's (1980) heteroskedasticity-consistent standard errors.

For the simple heteroskedasticity (not panel heteroskedasticity) case, Ω is a diagonal matrix, with entries being the error variance for each observation. White uses $\mathbf{X}'\hat{\Omega}\mathbf{X}$ to estimate $\mathbf{X}'\Omega\mathbf{X}$, where $\hat{\Omega}$ is a diagonal matrix consisting of the squares of the OLS residuals. Clearly the squares of the OLS residuals are not very good estimates of the corresponding error variances. But even so, White's estimate of $\mathbf{X}'\Omega\mathbf{X}$ performs well in finite samples.[46] His estimate of Ω is a weighted average of all the squares and cross products of the independent variables, with the weights provided by the estimated residuals. As a weighted average, it can be expected to be an accurate finite sample estimator for sample sizes seen in practical research situations.

The same argument holds for PCSEs. In fact, PCSEs will show even better finite sample performance than do White's heteroskedasticity-consistent standard errors, since PCSEs take into account the panel structure of the data. Thus, while each element of White's $\hat{\Omega}$ is computed using only one residual, each element of the PCSE $\hat{\Omega}$ is computed using T observations.[47] As a consequence, the estimate of Ω that enters into the PCSEs is superior to that used by White. Therefore, PCSEs should have good finite sample performance. This expectation is borne out in our Monte Carlo studies.

[46] MacKinnon and White (1985) examined the finite sample performance of the heteroskedasticity-consistent standard errors. While they showed that jackknifed standard errors outperformed heteroskedasticity-consistent standard errors, the advantage of the former was not great and the latter performed quite well.

[47] Thus, we showed that $\hat{\Omega}$ is a consistent estimate of Ω. White had to rely on $X'\hat{\Omega}X$ being a consistent estimate of $X'\Omega X$.

Parks Standard Errors Do Not Perform Well

PCSEs and Parks (1967) standard errors are built from the same estimate of Ω.[48] Why, then, do PCSEs perform well in finite samples while the Parks standard errors do not? The simple answer is that the Parks standard errors are not estimating the variability of the Parks FGLS estimator and hence there is no reason for them to be a good estimate of that variability. The Parks standard errors are not inaccurate because they are computed using $\hat{\Omega}$ but rather because they are estimating the wrong thing (Freedman and Peters 1984).

The problem with the Parks standard errors is that they are estimating the variability of a generalized least squares, not a feasible generalized least squares, estimator. Thus, they estimate only one portion of the overall variability of the Parks estimates, ignoring that portion of variability due to the use of an estimate of Ω rather than its known value. In finite samples this omitted variability can be considerable unless T is much greater than N. This expectation is, again, borne out by our Monte Carlo studies.

REFERENCES

Atunes, George, and James Stimson. 1988. *User's Guide to Microcrunch*. Houston: SofTex Micro Systems.

Beck, Nathaniel. 1991. "Comparing Dynamic Specifications: The Case of Presidential Approval." *Political Analysis* 3:51–87.

Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to Do (and Not to Do) with Times-Series-Cross-Section Data." *American Political Science Review* 89:634–47.

Blais, André, Donald Blake, and Stéphane Dion. 1993. "Do Parties Make a Difference? Parties and the Size of Government in Liberal Democracies." *American Journal of Political Science* 37:40–62.

Burkhart, Ross, and Michael Lewis-Beck. 1994. "Comparative Democracy: The Economic Development Thesis." *American Political Science Review* 88:903–10.

Clingermayer, James C., and B. Dan Wood. 1995. "Disentangling Patterns of State Debt Financing." *American Political Science Review* 89:108-20.

Engle, Robert. 1984. Wald, Likelihood Ratio and Lagrange Multiplier Tests in Econometrics. In *Handbook of Econometrics*, edited by Z. Griliches and M. Intriligator. Vol. 2. Amsterdam: North-Holland, pp. 775–826.

Engle, Robert, and C. W. J. Granger, eds. 1991. *Long Run Relationships: Readings in Cointegration*. New York: Oxford University Press.

Freedman, David, and Stephen Peters. 1984. "Bootstrapping a Regression Equation: Some Empirical Results." *Journal of the American Statistical Association* 79:97–106.

[48] The computation of the Parks standard errors is laid out in Beck and Katz (1995).

Gilbert, Christopher. 1990. Professor Hendry's Econometric Methodology. In *Modeling Economic Series*, edited by C. W. J. Granger. New York: Oxford University Press, pp. 279–303.

Giles, Michael, and Kaenan Hertz. 1994. "Racial Threat and Partisan Identification." *American Political Science Review* 88:317–26.

Greene, William. 1993. *Econometric Analysis.* 2d ed. New York: Macmillan.

Harvey, Andrew. 1990. *The Econometric Analysis of Time Series.* 2d ed. Cambridge: MIT Press.

Hendry, David, and Graham Mizon. 1978. "Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on a Study of the Demand for Money by the Bank of England." *Economic Journal* 88:549–63.

Hsaio, 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*, edited by T. Koopmans. New York: Wiley, pp. 365–83.

King, Gary. 1990. "On Political Methodogy." *Political Analysis* 2:1–29.

Kmenta, Jan. 1986. *Elements of Econometrics.* 2d ed. New York: Macmillan.

Levin, Andrew, and C.-F. Lin. 1993. "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties." Discussion Paper 92-23 (revised), Department of Economics, University of California, San Diego.

Levobic, James. 1994. "Riding Waves or Making Waves? The Services and the U.S. Defense Budget, 1981–1993." *American Political Science Review* 88:839–52.

MacKinnon, James, and Halbert White. 1985. "Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics* 29:305–25.

Nickell, S. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49:1417–26.

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.

Stimson, James. 1985. "Regression in Space and Time: A Statistical Essay." *American Journal of Political Science* 29:914–47.

Stock, James, and Mark Watson. 1988. "Variable Trends in Economic Time Series." *Journal of Economic Perspectives* 2(3): 147–74.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix and a Direct Test for Heteroskedasticity." *Econometrica* 48:817–38.

White, Halbert. 1984. *Asymptotic Theory for Econometricians.* Orlando: Academic Press.