Annu. Rev. Polit. Sci. 2001. 4:271–93

TIME-SERIES–CROSS-SECTION DATA: What Have We Learned in the Past Few Years?

Nathaniel Beck
Department of Political Science, University of California, San Diego, La Jolla, California 92093-0521; e-mail: beck@ucsd.edu

Key Words robust standard errors, feasible generalized least squares, random coefficients, spatial econometrics, error correction

■ **Abstract** This article treats the analysis of "time-series–cross-section" (TSCS) data, which has become popular in the empirical analysis of comparative politics and international relations (IR). Such data consist of repeated observations on a series of fixed (nonsampled) units, where the units are of interest in themselves. An example of TSCS data is the post–World War II annual observations on the political economy of OECD nations. TSCS data are also becoming more common in IR studies that use the "dyad-year" design; such data are often complicated by a binary dependent variable (the presence or absence of dyadic conflict). Among the issues considered here are estimation and specification. I argue that treating TSCS issues as an estimation nuisance is old-fashioned; those wishing to pursue this approach should use ordinary least squares with panel correct standard errors rather than generalized least squares. A modern approach models dynamics via a lagged dependent variable or a single equation error correction model. Other modern issues involve the modeling of spatial impacts (geography) and heterogeneity. The binary dependent variable common in IR can be handled by treating the TSCS data as event history data.

1. INTRODUCTION

Time-series–cross-section (TSCS) data are commonly analyzed in political science and related disciplines. TSCS data are characterized by repeated observations (often annual) on the same fixed political units (usually states or countries). For convenience, I refer to units and countries interchangeably, and similarly refer to time period and years. Although I refer only to political science applications, there are also many purely economic uses of TSCS data, particularly in the study of economic growth (e.g. Grier & Tullock 1989).

Although there are myriad applications, the prototypical application is the study of political economy, and in particular the impact of political arrangements on economic performance in advanced industrial societies. Here I use an example

1094-2939/01/0623-0271$14.00

drawn from Garrett (1998), who examines the political economy of government economic policy and performance in 14 OECD nations from 1966 to 1990. In particular, he is interested in whether labor organization and political partisanship affect economic policy and/or performance, and whether the impacts of those variables have changed over time (whether globalization has limited the impact of domestic political arrangements).

Other applications rely on more or fewer repeated observations. Although there is no strict lower limit to the number of repeated observations, there must be enough for some averaging operations to make sense. Thus, for example, our[1] simulations use a minimum of 15 repeated observations. There is no reason in principle that the observations need be annual, but they typically are. Although quarterly or monthly data would increase the number of repeated observations, such data are often not meaningful in the political economy context. There is no upper limit to the number of repeated observations we can study, and more repeated observations simply improve the performance of TSCS estimators.

Although the 14 units studied by Garrett is a typical number (for other studies, see Hall & Franzese 1998, Hicks & Kenworthy 1998, Iversen 1999, Radcliffe & Davis 2000), other researchers have studied the 50 American states (e.g. Fiorina 1994, Fording 1997, Smith 1997, Su et al 1993) or the 100 or so nations on which good data exist (e.g. Blanton 2000, Burkhart & Lewis-Beck 1994, Gasiorowski 2000, Poe et al 1999), or even the 654 parishes of Louisiana (Giles & Hertz 1994). The critical issue, as Section 2 shows, is that the units be fixed and not sampled, and that inference be conditional on the observed units.

TSCS data has also become of interest in international relations (IR). Many quantitative IR researchers use a "dyad-year" design (Maoz & Russett 1993), in which pairs of nations are observed annually for long periods of time (ranging from 40 to over 100 years). The dependent variable of interest in these studies is often the binary indicator of whether a dyad was in conflict in a given year. Binary dependent variables cause special problems. (Although binary dependent variables are most common in the study of international conflict, they also arise in studies of the diffusion of innovation, such as Berry & Berry 1990. The same logic holds for such studies.)

I discuss the characteristics of TSCS data in Section 2 and use that characterization to discuss estimation issues in Sections 3 and 4. Section 3 concerns old-fashioned estimation issues, which center on generalized least squares. Generalized least squares methods treat the interesting properties of TSCS data as nuisances that cause estimation difficulties. These methods have poor statistical performance. Section 4 provides the current solution to the old-fashioned estimation nuisance problem. Although the solution is also old-fashioned, it at least has good statistical properties.

[1]This paper reports work done jointly with Jonathan N Katz, and the use of "our" or "we" always denotes Katz and myself.

The next sections consider a more modern approach, which treats the interesting properties of TSCS data as something to be modeled. I begin with dynamics in Section 5. I then turn to the modeling of spatial or geographic factors (Section 6), followed by the modeling of heterogeneity (Section 7). Section 8 discusses the binary dependent variable case of interest in IR.

2. CHARACTERIZING TSCS DATA

TSCS uses the following notation (assuming a "rectangular" structure for notational convenience but with no loss of generality):

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

where $\mathbf{x}_{i,t}$ is a K vector of exogenous variables and observations are indexed by both unit (i) and time (t). Assume Ω to be the $NT \times NT$ covariance matrix of the errors with typical element $E(\epsilon_{i,t}\epsilon_{j,s})$. I assume, until Section 8, that the dependent variable, y, is continuous (at least in the sense of social science, where seven-point scales and the like are treated as continuous). Given the nature of typical TSCS data, I often refer to the units as countries and the time periods as years, but the discussion generalizes to any data set that is TSCS.

Equation 1 hides as much as it reveals. In particular, it does not distinguish "panel" data from TSCS data. Panel data are repeated cross-section data, but the units are sampled (usually they are survey respondents obtained in some random sampling scheme), and they are typically observed only a few times. TSCS units are fixed; there is no sampling scheme for the units, and any "resampling" experiments must keep the units fixed and only resample complete units (Freedman & Peters 1984). In panel data, the people observed are of no interest; all inferences of interest concern the underlying population that was sampled, rather than being conditional on the observed sample. TSCS data are exactly the opposite; all inferences of interest are conditional on the observed units. For TSCS data, we cannot even contemplate a thought experiment of resampling a new "Germany," although we can contemplate observing a new draw of German data for some year.

The difference between TSCS and panel data has both theoretical and practical consequences, which go hand in hand. Theoretically, all asymptotics for TSCS data are in T; the number of units is fixed and even an asymptotic argument must be based on the N observed units. We can, however, contemplate what might happen as $T \rightarrow \infty$, and methods can be theoretically justified based on their large-T behavior.

Panel data have the opposite characteristic. However many waves a panel has, that number is fixed by the design, and there can be no justification of methods by an appeal to asymptotics in T. There are, however, reasonable asymptotics in N, as sample sizes can be thought of as getting larger and larger.

Many common panel methods are justified by asymptotics in N. In particular, the currently popular "general estimating equation" approach of Liang & Zeger

(1986) is known to have good properties only as N becomes large. Thus, although it might be a very useful method for panel data, there is no reason to believe that the general estimating equation is a good approach for TSCS data.

The methods Katz and I propose require that T be large enough that averages over the T time periods for each unit make sense. We also use standard time-series methods to model the dynamics of TSCS data; this is possible only when T is not tiny. Panel data methods, conversely, are constructed to deal with small Ts; one would not attempt to use a lagged dependent variable when one has only three repeated observations per unit!

Thus, TSCS methods are justified by asymptotics in T and typically require a reasonably large T to be useful. Again, there is no hard and fast minimum T for TSCS methods to work, but one ought to be suspicious of TSCS methods used for, say, $T < 10$. On the other hand, TSCS methods do not require a large N, although a large N is typically not harmful. Thus, estimation on 14 OECD nations does not violate any assumption that justifies a (correct) TSCS method, but estimation on thousands of fixed units via TSCS methods is perfectly acceptable (though it might be numerically difficult). In contrast, panel methods are designed for and work well with very small Ts (three, or perhaps even two) but require a large N for the theoretical properties of the estimators to have any practical consequences. Panel estimators are also designed to avoid practical issues that arise from the large (and asymptotically infinitely large) N that characterizes panel data. Because much of the econometric literature conflates the analysis of panel data with the analysis of TSCS data, it is critical to keep in mind the distinction between the two types of data.

3. ESTIMATION ISSUES

Equation 1 can be estimated by ordinary least squares (OLS). OLS is optimal if the error process for Equation 1 meets the Gauss-Markov assumptions. Several of the Gauss-Markov assumptions are often suspect for TSCS data. An old fashioned approach is to treat these violations as a nuisance and correct for them using feasible generalized least squares (FGLS). I use the term old-fashioned because this perspective views violations of the Gauss-Markov assumptions as an estimation nuisance rather than something to be modeled. The modern perspective, at least in time series, is to regard these “violations” as interesting features to be modeled and not swept under the rug (Hendry & Mizon 1978). The situation for TSCS data differs from that for single time series in that the FGLS approach can do considerable harm with TSCS data; this is because it is possible to estimate some error properties of TSCS data that cannot be estimated with either a single time series or cross section. I therefore consider the old-fashioned methods primarily as a warning of what can go wrong and for historical reasons. (Lest I be accused of beating a dead horse, I note that the horse was quite alive only three or four years ago, and that I can claim some credit for helping to kill it!)

The Gauss-Markov assumption is that each of the $\epsilon_{i,t}$ is independent and identically distributed; that is,

$$E(\epsilon_{i,t}\epsilon_{j,s}) = \begin{cases} \sigma^2 & \text{if } i = j \text{ and } s = t \\ 0 & \text{otherwise.} \end{cases} \qquad 2.$$

It is well known that OLS will be inefficient, and its reported standard errors may be incorrect, if the error process does not look like Equation 2. In particular, the errors may show (*a*) panel heteroskedasticity, i.e. each country may have its own error variance (Equation 4); (*b*) contemporaneous correlation of the errors, i.e. the error for one country may be correlated with the errors for other countries in the same year (Equation 3); or (*c*) serially correlated errors, i.e. the errors for a given country are correlated with previous errors for that country (Equation 6). We would expect the errors from TSCS models to often "violate" the Gauss-Markov assumptions. If nations vary so that the error variance varies from nation to nation, we expect to observe panel heteroskedasticity; alternatively, we may observe panel heteroskedasticity because one or two units do not fit the basic specification well. Panel heteroskedasticity is one type of interunit heterogeneity. As we shall see in Section 7, there are many more interesting forms of unit heterogeneity to assess and model.

We will observe contemporaneously correlated errors if unobserved features of some countries are related to unobserved features in other countries. (Although we use the term errors, these are only errors of the observer, i.e. omitted variables.) Thus, if the Dutch and German economies are linked, then we would expect omitted variables for each country also to be linked.

Finally, since TSCS data consist of a group of time series, we would expect the data to show the usual features of time-series data, that is, temporally dependent observations. In the old-fashioned approach, these are modeled as serially correlated errors.

These are seen as nuisances in the old-fashioned approach. After discussing the old-fashioned approach, I return to the modeling of each of these interesting features of TSCS data. In this section and the next, however, I treat these problems as nuisances that cause problems in estimation. For expository purposes, I assume in this section and the next that observations are temporally independent, returning to that issue in Section 5. As we shall see, the modern approach to dynamics fits easily with our recommended estimation fix.

Feasible Generalized Least Squares

TSCS models with contemporaneously correlated and panel-heteroskedastic errors have Ω as an $NT \times NT$ matrix block diagonal matrix with an $N \times N$ matrix of contemporaneous covariances, Σ [having typical element $E(\epsilon_{i,t}\epsilon_{j,t})$], along the block diagonal. This follows from the assumption that the error process can be characterized by

$$E(\epsilon_{i,t}\epsilon_{j,s}) = \begin{cases} \sigma_i^2 & \text{if } i = j \text{ and } s = t \\ \sigma_{i,j} & \text{if } i \neq j \text{ and } s = t \\ 0 & \text{otherwise.} \end{cases} \qquad 3.$$

Note that the data provide T sets of residuals to estimate Σ. Thus, FGLS could be used to estimate Equation 1 with the panel heteroskedastic and contemporaneously correlated error matrix. Such a procedure first does OLS, uses the OLS residuals to estimate Σ, and uses the standard FGLS formulae to estimate model parameters and standard errors. This procedure was first described by Parks (1967) and was popularized by Kmenta (1986), so it is usually known as Parks or Parks-Kmenta.

Unlike many common FGLS applications, this procedure requires estimating an enormous number of parameters for the error covariances. Note that FGLS assumes that the parameters of Σ are known, not estimated. We have shown (Beck & Katz 1995) that the properties of the Parks estimator for typical TSCS Ts are very bad, and that, in particular, the estimated standard errors could be underestimating variability by from 50% to 200%, depending on T. The FGLS standard errors underestimate sampling variability because FGLS assumes that Σ is known, not estimated. Our conclusion is that the Parks-Kmenta estimator simply should not be used.

Table 1 shows why many researchers liked Parks-Kmenta; it gives the nice t-ratios that are so prized by journal editors. The Parks-Kmenta estimator of the basic Garrett model is in the third set of columns of the table. Note that with a T of 25, standard errors are anywhere between 50% and 100% smaller than corresponding OLS standard errors. Although this may make one's results easier to publish, the simple fact is that the Parks-Kmenta standard errors are wrong, and perhaps worse, they are wrong in the direction of being wildly optimistic.

Some researchers, noting this problem, avoided Parks-Kmenta but still used FGLS to correct for panel heteroskedasticity. In this model, all error covariances between different units are assumed to be zero, but each unit has its own error variance, σ_i^2. Panel heteroskedastic errors thus yield

$$E(\epsilon_{i,t}\epsilon_{j,s}) = \begin{cases} \sigma_i^2 & \text{if } i = j \text{ and } s = t \\ 0 & \text{otherwise.} \end{cases} \qquad 4.$$

Panel heteroskedasticity differs from simple heteroskedasticity in that error variances are constant within a unit.

This model appears to avoid the craziness of Parks-Kmenta, since only N error parameters need be estimated (and N is typically not enormous for TSCS studies). The FGLS correction for panel studies proceeds, as usual, by a first round of OLS and a second round of weighted OLS, with weights being inversely proportional to the estimated σ_i for each unit (and these σ_i estimated in the obvious manner). This procedure can be called panel-weighted least squares (PWLS).

Although our simulations do not show that PWLS has horrible properties (at least for reasonable N and T), we do feel that it is very problematic. This

TABLE 1 Comparison of ordinary least squares (OLS) and feasible generalized least squares (FGLS) estimates of the Garrett model of economic growth in 14 advanced industrial democracies, 1966–1990[a]

	OLS/PCSE[b]			Panel. Het.[c]		Cont. Corr.[d]	
Variable	$\hat{\beta}$	**SE**	**PCSE**	$\hat{\beta}$	**SE**	$\hat{\beta}$	**SE**
GDP Lagged	0.13	0.05	0.07	0.10	0.05	0.08	0.05
OIL	−6.62	6.26	6.42	−5.92	5.79	−4.71	3.12
DEMAND	.064	0.11	0.16	0.60	0.10	0.72	0.07
LABOR	−0.13	0.62	0.56	−0.44	0.52	−0.03	0.29
LEFT	−0.68	0.42	0.31	−0.57	0.27	−0.65	0.13
LEFTxLABOR	0.23	0.16	0.12	0.23	0.14	0.23	0.06
PER6673	1.41	0.55	0.74	1.56	0.49	1.85	0.40
PER7479	0.04	0.56	0.77	0.29	0.49	0.52	0.40
PER8084	−0.54	0.58	0.80	−0.51	0.51	−0.18	0.42
PER8690	−0.14	0.55	0.76	−0.2	0.49	0.13	0.40
CONSTANT	2.39	1.36	1.36	2.90	1.19	1.84	0.79

[a]All models estimated with fixed effects.

[b]OLS with OLS standard errors and panel correct standard errors (PCSE).

[c]FGLS estimates correcting for panel heteroskedasticity.

[d]FGLS correcting for both contemporaneous correlation of the errors and panel heteroskedasticity (Parks-Kmenta).

is because the weights used in the procedure are simply how well the observations for a unit fit the original OLS regression plane. The second round of FGLS simply downweights the observations for a country if that country does not fit the OLS regression plane well. Thus, on the second round, fit will be good! In other, non-TSCS applications, the correction for heteroskedasticity is theoretical and does not simply downweight poorly fitting observations. The closest analog to PWLS in the cross-sectional world would be to run a cross-sectional regression and then weight each observation by the inverse of its residual. That would yield nice R^2's and t's, but it would be an odd procedure. We reanalyzed (Beck & Katz 1996) a PWLS study of Burkhart & Lewis-Beck (1994) that analyzed economic growth in more than 100 countries and found that three quarters of the weight in the second-round regression came from only 20 advanced industrial societies that fit the first-round regression well. PWLS does less harm in the Garrett example (the middle columns of Table 1), but it does seem wrong to use a procedure that weights observations by how well they fit a prior regression.

This is not to say that we should ignore heterogeneity across units; far from it. As Section 7 demonstrates, the modern approach is to model this heterogeneity directly, or at least to inquire whether all units are governed by the same regression equation. But given the choice between simply ignoring panel heteroskedasticity (and hence using OLS) or using PWLS, the former seems less mischievous. It is also easy, as we will see in the next section, to modify OLS to avoid incorrect standard errors in the presence of panel heteroskedasticity; this modification comes at almost no cost.

4. PANEL CORRECT STANDARD ERRORS

The results of the previous section are negative. Parks-Kmenta has very poor properties and the FGLS correction for panel heteroskedasticity is, in my view, inherently flawed. This does not mean, however, that OLS is a good estimator for TSCS data; the errors are likely, after all, to show both panel heteroskedasticity and contemporaneous correlation of the errors. Under these conditions, OLS is still consistent, though inefficient, and the OLS standard errors may be wrong. Although inefficiency may be an important issue, it is easy to at least compute panel correct standard errors (PCSEs), which correctly measure the sampling variability of the OLS estimates, $\hat{\beta}$.

The usual OLS formula for the standard errors may be misleading for TSCS data. The correct formula is given by the square roots of the diagonal terms of

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

The OLS standard errors are produced by assuming that the Ω matrix is just a constant times an identity matrix; this assumption will often be incorrect for TSCS data.

Fortunately, it is easy, given the repeated time structure of TSCS data, to estimate Ω by $\hat{\Omega}$ and then use $\hat{\Omega}$ in place of Ω in Equation 5 to produce PCSEs. The PCSEs are simply the square roots of the diagonal terms of Equation 5, using $\hat{\Omega}$. Although the details are a bit complicated (see Beck & Katz 1995), the basic idea is quite simple. Ω is a block diagonal matrix, where each block is identical. We thus have T replications of the error that can be used to estimate this block; for large T, this estimate is quite good. Simulations reported by Beck & Katz (1995) indicate that PCSEs are very accurate (to within a few percent) for $T > 15$. This is true even when the errors meet the Gauss-Markov assumptions. Thus, there is no cost, and some potential gain, to using PCSEs in place of the usual OLS standard errors. They are, in addition, easy to compute and are implemented in standard software (such as Stata and LIMDEP). We therefore recommend that researchers who are worried about using OLS because of the complicated errors found in TSCS data use the OLS estimates of β combined with PCSEs. This approach, though still old-fashioned, works much better than the FGLS "fix" of the previous section.

The third column of Table 1 shows the PCSEs for the Garrett model. The reported PCSEs differ from their OLS counterparts by about one third. Thus, for example, the OLS standard errors understate our uncertainty about the critical political interaction term; with correct standard errors, we can clearly reject the null hypothesis that this interaction has no effect on economic growth. Since the

PCSEs are always as good as (and usually better than) their OLS counterparts, this leads to the correct inference that nations with powerful left parties and centralized bargaining grow faster than nations with only one of those attributes.

5. MODELS WITH TEMPORALLY DEPENDENT OBSERVATIONS

So far I have dealt only with cross-sectional issues. I now turn to problems caused by dynamics. Obviously TSCS data will often show dynamics. The old-fashioned treatment is to think of these dynamics as a nuisance, that is, to model them as serially correlated errors so that

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

where the νs are independent and identically distributed. Models with such an error process must be estimated by FGLS (Prais-Winsten or the like).[2] This correction requires the estimation of only one serial correlation parameter, so it does not have bad statistical properties. But it is not consistent with modern time-series analysis.

Modern time-series analysts model the dynamics directly as part of the specification. The simplest form of this is the use of lagged dependent variables, but depending on the data, other forms, such as single equation error correction (Davidson et al 1978), may be tried. Whatever we can do for time series we can do for TSCS data. Because typical TSCS data are annual, it is often the case that a single lagged dependent variable is all that the data warrant; it is easy to test for more complicated dynamics via standard Lagrange multiplier tests described below.

Following the identical argument for time series, we can usually replace serially correlated error models with models involving a lagged dependent variable. This simplifies other estimation issues so long as the error process, conditional on the lagged dependent variable being in the specification, is temporally independent. It is easy to test for this independence via a standard Lagrange multiplier test.[3] The

[2]FGLS here consists of running OLS, then estimating the autoregressive parameter in a regression of the residuals on their lags, and then transforming the data by subtracting the estimated autoregressive parameter times the prior observation from the current observation. This is identical to the single time-series correction of serially correlated errors; see Beck & Katz (1996) for more details.

[3]This test regresses the OLS residuals on their lags, as well as all other independent variables, including the lagged dependent variable. The advantage of this test over others is that all estimation is done under the null hypothesis of no serial correlation of the errors, so all estimation can be done using OLS. A similar regression, but with higher-order lagged residuals, tests whether the data indicate the need for a more complex dynamic structure, involving higher order lags. Tests on the Garrett model indicate that the single lag of the dependent variable is adequate, and that there is no remaining serial correlation of the errors. As in the single time-series case, the use of a lagged dependent variable with temporally dependent errors makes OLS inconsistent. As in that case, this is seldom a problem in practice.

modeling of dynamics via a lagged dependent variable allows researchers to estimate their specification using the same methodology recommended in Section 4, OLS with PCSEs.

The Garrett data appear to be well modeled by stationary time-series methods, so no attempt was made to investigate error correction models of the growth of gross domestic product. But this need not have been the case. Had Garrett modeled the level of GDP, rather than its rate of growth, we might well have observed nonstationary data ("unit roots"). Issues of nonstationary data have been extensively studied for single time series; they are also analyzed for large-N panel studies. But we know little about nonstationary TSCS data. One solution that should work well is the TSCS variant of the single equation error correction model. As in the single time-series case, this models the change in a dependent variable of interest as a function of changes in the relevant independent variables and the amount by which the variables are out of equilibrium. We thus have

$$\Delta y_{i,t} = \Delta \mathbf{x}_{i,t}\beta + \rho(y_{i,t-1} - \mathbf{x}_{i,t-1}\gamma) + \epsilon_{i,t};$$
$$i = 1, \ldots, N; \quad t = 2, \ldots, T, \qquad 7.$$

which could then be estimated by OLS with PCSEs. Both Iversen (1999) and Franzese (2001) have used error correction models for studying OECD political economy with some success.

6. SPATIAL MODELING

The Parks FGLS method was designed to correct for OLS problems caused by relationships between the various units. Although the correction has bad properties, we would expect the observations of the various countries in a TSCS study to be interrelated. It is better, however, to try to model this relationship than to leave it as an unspecified nuisance. Such modeling is standard among economic geographers and spatial econometricians (Anselin 1988). These ideas are hard to implement in simple cross-sectional models, but they are easy to implement in TSCS models if we are willing to assume that spatial effects operate with a temporal lag.

The Parks approach simply allowed for an unspecified contemporaneous correlation of the errors that varied by unit but not by time. The inordinate number of parameters in this model is the cause of its terrible performance. Spatial econometricians, on the other hand, assume that there are a few parameters that describe the relationship among units. Geographers typically assume that units are related in proportion to their closeness. Although closeness is often defined in a purely geographic manner, there is no reason it cannot be defined by the degree of economic relationship between nations.

Spatial econometricians allow for two types of spatial dependence. The errors of nearby units may be correlated (spatial autocorrelation), with the degree of correlation inversely related to the closeness of the two units. Alternatively, a "spatial lag," that is, the weighted sum of the dependent variable of all other units, may be added to the specification. The weights in this sum are again proportional to the closeness of the units. If one has only cross-sectional data, spatial econometrics presents formidable technical challenges. But with TSCS data, one can assume that neighbors have an effect only with a temporal lag. Then, so long as the remaining errors are temporally independent, one can use spatial econometric methods quite easily.

Although it is easy to add spatially autocorrelated errors to the model using FGLS methods, it seems more natural to use spatial lags. Remember that the "errors" are errors of the observer, that is, omitted variables that the researcher either could not or did not choose to include in the model. Thus, if we assume that only the errors show autocorrelation, then we are assuming that only the unmeasured variables are spatially related and that there is no relationship between measured variables. This is, however, far from the ideal way to think about spatial independence, at least for models of open economies (Alt 1985, Franzese 2001, Garrett 1998).[4] We would expect that the economies of trading partners are linked, so that if the economies of a nation's trading partners are doing well, that nation's economy also should be doing well. If we assume that this spatial effect also operates with a temporal lag (so that it takes some time, say a year, for an improving German economy to be reflected in an improving Dutch economy), then we simply add to Equation 1 a term $S_{i,t-1}$, which is a weighted average of the lagged dependent variable for all units except unit i.

This average is a weighted average, with prespecified (not estimated!) weights. Geographers often use a weighting scheme in which abutting units are given a weight of one and all others are weighted zero. This scheme might make sense in some cases, and it is used in Berry & Berry's (1990) study of the adoption of state lotteries. But for political economy models, the right weighting is almost certainly the importance of trade between unit i and the other units; the German economy, lagged one year, will have a strong effect on the Dutch economy in proportion to the importance of Dutch-German trade in the Dutch economy (that is, as a proportion of Dutch GDP). Thus, for each nation and each year, a researcher must compute the weighted average of the economic performance of all trading partners, weighted by the importance of that trade.

Many researchers use a spatial lag of the dependent variable in their models without explicitly noting the geographic interconnection of their data. Thus, for example, the lagged spatial lag has already been included in the Garrett specification I have been estimating. His *DEMAND* variable is a trade weighted average (for each country) of the growth of GDP in all other OECD countries. Its inclusion in the model reflects two obvious ideas: (*a*) When the world's economy improves, nation *X*'s economy will probably improve, and (*b*) how much better nation *X* is

[4]I focus here on political economy models, but the same logic holds for models of the adoption of innovations by states, such as those of Berry & Berry (1990). In those models, it is the spatial lag of the innovations of other states that has an effect; states are more likely to adopt a lottery, for example, if geographically nearby states have done so.

doing reflects how much better other nations are doing, weighted by the importance of those others nations to X's economy. As the tables show, the performance of one's trading partners has an important impact on one's own economic performance. Omitting this variable will yield either inefficient or biased estimates of the impact of domestic variables on economic performance.

As long as the performance of one's partners operates with a temporal lag, it is easy to estimate a model that contains a spatial lag by OLS. One must ensure that the errors show no temporal correlation, but the lagged dependent variable will usually do so. Any remaining contemporaneous correlation of the errors, or panel heteroskedasticity, can be handled by the use of PCSEs. For this methodology to work, care must be taken that no independent variable is correlated with its own error term. Since the temporal lag of the spatial lag is clearly correlated with the temporal lag of all other unit error terms, it is critical that the temporally lagged dependent variable sop up all, or almost all, of the dynamics so there is no (or almost no) remaining temporal serial correlation of the errors. This is easy to test for by using the Lagrange multiplier test described in the previous section. Although my own experience indicates that the lagged dependent variable eliminates serial correlation of the errors, clearly this will not always be the case, and so a strict adherence to testing is required.

7. HETEROGENEITY

Equation 1 assumes that the countries are completely homogeneous, differing only in the levels of their explanatory variables. The FGLS approach was designed to guard against one type of heterogeneity, unequal unit error variances. A more modern approach would presume that heterogeneity is an interesting feature of TSCS data and attempt to model that heterogeneity. That approach would allow for heterogeneity in the parameters of interest, the β in Equation 1.

TSCS researchers never assume a completely homogeneous universe. Instead, they typically assert that some subset of countries is homogeneous and include those countries in the analysis, excluding all others. It would be interesting to allow for a more nuanced view of heterogeneity. Before discussing the modeling of heterogeneity, it is first relevant to discuss tests for homogeneity.

Assessing Heterogeneity via Cross Validation

The standard test for homogeneity is a normal F-test. We estimate the two models

$$H_0 : y_{i,t} = \mathbf{x}_{i,t}\beta + \epsilon_{i,t} \qquad 8.$$

$$H_1 : y_{i,t} = \mathbf{x}_{i,t}\beta_i + \epsilon_{i,t} \qquad 9.$$

and test the null hypotheses $H_0 : \beta_i = \beta$. This is done by the usual F-statistic, which compares the difference in sums of squares residuals from estimating Equations 8 and 9, divided by the appropriate number of degrees of freedom and then divided by the mean square error of Equation 9.

We might reject the null of pooling either because of slight variations in all the β_i or because one particular country is not well fit by Equation 8. We also might fail to notice an "outlying" country because the F-test averages variation over all units. Finally, we might reject the null of pooling when there is little parameter variation because of the large sample sizes common in TSCS data sets. In any event, great care must be taken in interpreting the result of the F-test.[5]

We can assess whether the F-test rejected homogeneity because one or two countries were outliers by using cross validation (Stone 1974). For the Garrett data, this procedure also assesses whether the small F-statistic resulted from averaging many homogeneous units with one outlying unit. The simplest form of cross validation is to leave out one observation, fit a regression with all the others, and "predict" the left-out observation. For all but very small data sets, this procedure is useless. But for TSCS data, it makes perfect sense, except that instead of leaving out one observation at a time, we leave out one unit at a time. We can then compare specifications by seeing how they perform in terms of mean absolute (or square) "prediction" error, or we can see if any units are predicted less well than the others. Table 2 demonstrates such an exercise.

In that table, we see that typical mean absolute forecast errors range from 1.2% to 2% (the unit is percent growth in GDP), except for Japan, which has a forecast error of 3.2%. Thus, Japan fits the basic specification much less well than any other OECD nation. A researcher might be well advised to drop the Japanese data from the Garrett specification.[6] Dropping Japan from the analysis increases the impact of the critical *LEFTxLABOR* interaction term by about 20%.

Fixed Effects Models

Having checked for unit heterogeneity, how do we proceed? The old-fashioned approach was to eliminate the estimation problems caused by the nuisance of heterogeneity. The modern approach is to model this heterogeneity. The simplest way of doing so is to assume that each unit has its own intercept, thus changing Equation 1 to

$$y_{i,t} = \mathbf{x}_{i,t}\beta + f_i + \epsilon_{i,t}, \qquad 10.$$

[5]A test for pooling in the Garrett model, taking fixed effects as given, yields an F-statistic of 1.29 with 130 and 196 degrees of freedom. This just misses being statistically significant at the conventional level ($P < 0.06$). But the statistic is very small and is only nearly significant because of the huge number of degrees of freedom in the numerator. Thus, given a choice between a fully pooled and a completely unpooled model, I would simply estimate the pooled model, with fixed effects, for these data. To pursue the issue further, one would do an F-test for pooling on interesting subsets of the coefficients.

[6]This makes sense particularly because the Japanese political economy differs in many ways from other OECD political economies. I would be less sanguine if a country such as the Netherlands had the worst forecast errors. It is important to make sure one is not achieving good fits simply by dropping units that fit poorly. But in this case, both cross validation and knowledge of the OECD nations leads me to argue that Japan should not be included in the analysis.

TABLE 2 Out of sample forecast errors (ordinary least squares) by country for Garrett model of economic growth in 14 advanced industrial democracies, 1966–1990[a]

Country	Mean Absolute Error (%)
United States	1.9
Canada	1.7
United Kingdom	1.7
Netherlands	1.6
Belgium	1.6
France	1.2
Germany	1.4
Austria	1.3
Italy	1.7
Finland	2.0
Sweden	1.2
Norway	1.5
Denmark	1.7
Japan	3.2

[a]No effects in model.

where f_i is a dummy variable marking unit i. Note that this model introduces very little heterogeneity. The effect of the independent variables does not vary by country; each country simply has its own intercept.

The panel literature contains great debates about how to model the effects, and in particular, whether we should treat them as fixed or random. That argument, however, is for panel data; for TSCS data, it is clear that fixed effects are appropriate (Hsaio 1986: 41–43). Hsaio shows that fixed effects are appropriate if one wants to make inferences to the observed units, whereas the random effects model (which assumes that the effects are drawn from some distribution) is appropriate if one thinks of the observed units as a sample from a larger population and if one wants to make inferences about the larger population. In TSCS data, the units (countries) are fixed and we are not interested in extending inference to a larger, hypothetical, population of similar countries. Furthermore, with a large T, fixed and random effects converge. As is well known (see e.g. Greene 1999: 568–70), fixed effects and random effects differ by $\frac{\sigma_\epsilon 2}{\sigma_\epsilon^2 + T\sigma_\alpha^2}$. As T gets large, this term goes to zero, so that the random-effects and fixed-effects estimators become identical.

It is easy to test whether the fixed effects are required via a simple F-test comparing the sums of squared errors from Equations 1 and 10. Obviously, fixed effects are not theoretical variables; to say that Germany grew faster because it was

Germany is hardly a satisfying explanation. But if we cannot otherwise explain unit variations in growth, and the *F*-test indicates we need fixed effects, then estimating Equation 1 without fixed effects means that we are estimating a misspecified model (Green et al 2001).[7]

We should note that the use of fixed effects comes with its own costs. Fixed effects are clearly collinear with any independent variables that are unchanging attributes of the units, so they force us to drop such unchanging variables from the specification. These variables (perhaps characteristics such as democracy) might be of interest. And although we can estimate Equation 10 with slowly changing independent variables, the fixed effects will soak up most of the explanatory power of those slowly changing variables. Thus, if a variable such as type of bargaining system changes over time, but slowly, the fixed effects will make it hard for such variables to appear either substantively or statistically significant (Beck & Katz 2001).

TSCS analysts should test to see whether fixed effects are needed in the specification. If not, then there is no problem. If an *F*-test indicates that fixed effects are required, then researchers should make sure they are not losing the explanatory power of slowly changing or stable variables of interest. (Sometimes we use stable variables of no substantive interest, and it is not harmful to eliminate these from the specification.) If variables of interest are being lost because of the inclusion of fixed effects, the researcher must weigh the gains from including fixed effects against their costs. If the gains, in terms of decreased sum of squared errors, are slight, albeit statistically significant, then it might be better to omit the fixed effects and suffer slight omitted-variable bias. Like most interesting issues, this is a matter of judgment, not slavish adherence to some 0.05 test level (see Beck & Katz 2001 for a fuller discussion).

The Random Coefficients Model

The random coefficients model (RCM) is an interesting compromise between assuming complete homogeneity and assuming complete heterogeneity. This model is the same as the Bayesian hierarchical model. Western (1998) provides a full discussion of this model in the context of TSCS data.

The RCM is a compromise between estimating the fully pooled Equation 1 and a fully unpooled estimate, that is, a separate OLS for each unit. There are not enough data for the latter (that is, separate OLS estimations will have huge standard errors), but the former requires the very strong assumption of complete pooling. The RCM uses the idea of "borrowing strength" (Gelman et al 1995). This Bayesian notion shrinks each of the individual unit OLS estimates back to the overall (pooled) estimate. RCMs simply generalize random effects from the intercept to all parameters of interest.

[7]The *F*-statistic on the null hypothesis that fixed effects are not needed in the Garrett specification is 4.88 with 13 and 326 degrees of freedom. We can therefore reject the null hypothesis that fixed effects are not needed in the specification with a *P* value of zero to many decimal places.

The RCM is

$$y_{i,t} = \mathbf{x}_{i,t}\beta_i + \mathbf{z}_{i,t}\gamma\epsilon_{i,t}, \tag{11.}$$

where the $\beta_i \sim N(\beta, \sigma_\beta^2)$ and the γ represent fixed coefficients. Note that if we allow only the intercept to be random, the RCM reduces to the usual random effects model. This is a complex model, but it can be estimated by either classical maximum likelihood (Pinheiro & Bates 2000) or Bayesian methods, typically with diffuse priors (Western 1998).[8]

The RCM can be made more useful, as Western showed (1998), by allowing the β_is to be functions of other unit variables, z_i. This step allows for modeling differential effects as a function of differing institutions. (The z_is are time invariant, so they only measure properties of units.) This is particularly important in comparative politics, where we might expect that the effect of some x on the dependent variable is contingent on structural features that vary from country to country. For example, the Garrett model asserts that the effect of having a left government is contingent on the type of labor bargaining in each country. We can then write:

$$\beta_i = \mathbf{z}_i\gamma + \beta + \alpha_i. \tag{12.}$$

Substituting Equation 12 into Equation 11, we see that this model is simply an interactive model with random coefficients on the linear terms only.

This model is difficult to estimate. In its most general form, with many random coefficients, it requires the estimation of a huge number of parameters in the variance covariance of the random terms. Researchers can ask less of the data by allowing as many as possible of the coefficients to be fixed. It also makes sense to assume, as Western does, that the random coefficients are independently distributed, so no covariance terms need be estimated. This simplifies the estimation problem dramatically.

The RCM is seldom used in TSCS applications. One reason is that the RCM estimates of the β will be similar to the OLS estimates, in that OLS is still consistent, albeit inefficient if the coefficients really are random. Given the large sample sizes of typical TSCS data, this inefficiency may not be important. The OLS standard errors will of course be wrong. But, as Western (1998) shows, the RCM is of greater interest if we care about the estimated unit coefficients, the β_is. These are of great interest in the study of comparative politics. In what countries does politics have the greatest, or least, impact on economic performance?

All RCMs work by estimating the individual OLS estimates of the β_i, estimated one country at a time, and then shrinking them back to the overall pooled estimate of β. The degree of shrinkage is proportional to both how homogeneous

[8]In econometrics, the RCM dates back to Swamy (1971), who proposed a two-step estimator. Madalla & Hu (1996) report the inferior performance of the two-step estimator. Katz and I also have unpublished simulations showing that the two-step estimator performs poorly, even for very simple models. A variety of technical reasons for this poor performance are detailed by Beck (2000).

the units appear to be and how confident we are in our unit-by-unit estimates. This confidence is determined by *T*. In panel data, with small *T*, the RCM (known in that context as the hierarchical model) has been of great use. The issue is much less clear for TSCS data, especially as *T* gets large. It is also the case that we will observe the least shrinkage (that is, the least "borrowing of strength") when we most need it (i.e. when the units are heterogeneous) and the greatest shrinkage when we least need it (i.e. when the units are relatively homogeneous).[9] Thus, although RCMs appear to be an interesting way of modeling heterogeneity, the verdict on their utility for comparative politics is not yet clear.

8. BINARY DEPENDENT VARIABLES

So far we have assumed a continuous dependent variable, but what if we have binary TSCS (BTSCS) data?[10] Such data are particularly common in IR, where the dependent variable is whether a pair of nations was involved in a dispute in a given year and the independent variables are characteristics of the dyad (such as how democratic each nation is, or the level of dyadic trade). IR BTSCS analysts typically use the dyad-year design, where the variables are observed on pairs of nations annually over a long period (Maoz & Russett 1993). Observations on pairs of nations can produce a large *N* (4000 or so in the largest studies); the *T* in these studies ranges from ∼50 to 100. These data sets present interesting problems that were usually ignored until very recently. My own current efforts deal with some BTSCS issues using an old-fashioned "treat the nuisance" approach, but some recent work attempts to model the interesting features of BTSCS data (see Beck & Tucker 1997 for further discussion).

Binary-dependent-variable panel studies are of great interest in biostatistics and econometrics. Many medical trials involve observing a large number of subjects a few times, with each observation recording a binary dependent variable (e.g. whether the subject showed a particular symptom). Although these studies are interesting, their relevance to BTSCS data is unclear; all of them are justified by asymptotics in *N*, and all were designed to work for small *T*s. To give but one example, it is hard to allow for fixed effects in binary panel data, but fixed effects are as good (or as bad!) in BTSCS data as in TSCS data. Other popular BTSCS panel

[9]See Madalla & Hu (1996) for a computation of the shrinkage as a function of sample size and heterogeneity. Using his formula, I find that the RCM would shrink the unit OLS coefficients by only about 25%, back to the overall pooled estimates for the Garrett model we have been using. We have already seen that the pooled estimates for this model are quite acceptable. Thus, the RCM results for the Garrett model I have been using are not very interesting, and I do not report them here. This is not to say that RCM estimates of the β_i might not be more interesting in other data sets, such as that used by Western (1998).

[10]We could complicate matters further by allowing the data to be polychotomous or censored, but I restrict myself to the simple binary dependent variable case. Little is known about the more complex cases.

approaches, such as Liang & Zeger's (1986) "general estimating equation," may possibly work with BTSCS data, but all that has been proved about this method is that it works well for panel data.

Until 1999 or so, most analysts working with BTSCS data used ordinary logit (all results hold for probit also). This approach ignores both temporal and spatial dependencies in the data, which leads to inefficient estimation and incorrect standard errors. A modern approach attempts to model these features of the data. (See Gleditsch & Ward 2000 for a discussion of spatial issues.)

Dynamics: An Old-Fashioned Fix

Although it may seem odd to admit that an article I published two years ago has an old-fashioned solution, the fix for dynamic issues recommended in Beck et al (1998) treats dynamic issues as a nuisance that impedes estimation. The best solution would obviously be to directly model the dynamics; unfortunately this is very difficult. But failing to deal with the dynamics, either by old-fashioned or modern methods, can cause serious problems.[11]

Our old-fashioned solution for BTSCS data in IR is to think of them as event history data. Thus, the conflict data sets really contain information on the length of time between conflicts. This approach works if conflict is rare, so that most dyads manifest a long period of peace followed by an interlude of conflict.[12]

The advantage of thinking of BTSCS data as event history data is that event historians always allow for the possibility that the observations are temporally linked. For an event historian, the probability that a dyad will be in conflict, given that it is currently at peace, is a function of the past history of the dyad and not simply a function of current conditions.

To be more specific, it is easy to view BTSCS dyad-year data as grouped duration data suitable for a grouped Cox (1972) proportional hazards semiparametric estimation. The discrete version of the grouped duration Cox model is a complementary log-log model with a series of time dummy variables added to the specification. It is hard to see much difference in practice between the complementary log-log specification and the more common logit or probit. We thus recommend that researchers with data similar to the IR dyad-year conflict data do logit or probit but add the temporal dummies to their specifications.

Table 3 compares a discrete grouped time logit analysis (from Beck et al 1998) with an ordinary logit analysis (from Oneal & Russett 1997). The major consequence of including the temporal variables is that the pacific effects of international trade disappear.

---

[11]The problems are most severe when the errors are correlated and the independent variables trend. We have presented (Beck & Katz 1997) simulation results showing that, in the presence of severe trending and autocorrelation, standard errors can be off by 50% or more.

[12]The event history approach also sensitizes us to many other features of IR BTSCS data [see Beck et al (1998) and Beck & Katz (2001) for other insights we can reach by thinking like event historians].

TABLE 3 Comparison of ordinary logit and grouped duration analyses ($N = 20{,}990$)

	Ordinary	Logit	Grouped Duration	Logit[a]
Variable	$\hat{\beta}$	**SE**	$\hat{\beta}$	**SE**
Democracy	−0.50	0.07	−0.55	0.08
Economic growth	−2.23	0.85	−1.15	0.92
Alliance	−0.82	0.08	−0.47	0.09
Contiguous	1.31	0.08	0.70	0.09
Capability ratio	−0.31	0.04	−0.30	0.04
Trade	−66.13	13.44	−12.67	10.50
Constant	−3.29	0.07	−0.94	0.09
Log likelihood	−3477.6		−2554.7	
Degrees of freedom	20983		20036[b]	

[a]34 temporal dummy variables in specification not shown.

[b]Three dummy variables and 916 observations dropped.

Dynamics: A Modern Approach

A modern approach would, as we saw in Section 5, explicitly model the dynamics. Although this has not been done, I can sketch a reasonable approach. It starts with the latent variable setup for a logit model:

$$\begin{aligned} y^*_{i,t} &= \mathbf{x}_{i,t}\beta + \rho y^*_{i,t-1} + \epsilon_{i,t} \\ y_{i,t} &= 1 \text{ if } y^*_{i,t} > 0 \\ y_{i,t} &= 0 \text{ if } y^*_{i,t} \leq 0. \end{aligned} \qquad 13.$$

This model looks like the standard dynamic setup, with the underlying latent variable for conflict, the propensity to be in conflict, being a function of its lagged value. This model is difficult to estimate by standard methods, but is estimable using Markov Chain Monte Carlo methods (see Jackman 2000a,b for political science applications). This appears to be a promising approach for future research.

The explicit model in Equation 13 also helps us to think about alternative dynamic specification. Some have suggested using the realization of the lagged dependent variable, $y_{i,t-1}$, in place of the latent lag in Equation 13. Looking at the specification, we see this is equivalent to asserting that a current conflict increases the propensity for a future conflict, but a current pacific year decreases that future propensity. Such a model is very easy to estimate (simply add $y_{i,t-1}$ to the logit), but is it sensible to do this? If a dyad has a high propensity for conflict but fails to engage in conflict during a given year, does this increase or decrease the likelihood of future conflict? A more sensible model, which might be called "strain relief," replaces the lagged latent variable in Equation 13 with the strain in the dyad, i.e. the lagged difference between the propensity for conflict and whether a conflict actually occurred ($y^*_{i,t-1} - \gamma y_{i,t-1}$). This strain relief model is akin to the continuous dependent variable error correction model that is so appealing. This model can also be estimated by Markov Chain Monte Carlo methods. Although this has not yet been done, the approach is promising.

Heterogeneity

BTSCS researchers have recognized the issue of heterogeneity. Thus, many analysts (Maoz & Russett 1993) limit their analysis to the so-called politically relevant dyads (involving major powers or contiguous states), assuming that other dyads are unlikely to be in conflict. Assuming two types of homogeneous dyads, those never in conflict and those that are politically relevant, is of course a very strong assumption. Beck et al (2000) argued that a more flexible way to handle this type of heterogeneity is to allow for massive interactions in a neural network model. These massive interactions allow for an independent variable of interest, say democracy, to have almost no effect in most dyads but a large effect in a few "conflict prone" dyads. More work is required before we can be confident about the utility of this approach.

A much simpler approach has been suggested by Green et al (2001). They suggest simply adding fixed effects (basically a dyadic-specific dummy variable) to the logit. Although fixed effects cannot be used in panel models, they may perhaps be useful for BTSCS data. However, the use of fixed effects in the IR conflict data causes us to believe that dyads that are never in conflict can give us no information about the impact of variables such as democracy on conflict. Because 90% of dyads are never in conflict, and since these dyads are more likely to be democratic, this is a very serious problem. For this reason, fixed effects are almost never a good idea for BTSCS data (Beck & Katz 2001). This is not to say that modeling heterogeneity is not important; far from it. But at present we do not have a good method for modeling heterogeneity in BTSCS data.

9. CONCLUSION

Katz and I became interested in TSCS data in about 1993. TSCS models were becoming of great interest in political science because they appeared to allow students of comparative politics (broadly defined) to use powerful statistical methods that had been the province of students of American politics (typically studying voting behavior via large-*N* surveys). Most studies either ignored TSCS issues or treated those issues as a nuisance, using an FGLS estimation method. In the early 1990s, researchers were using some procedures that, to my mind, had poor statistical properties or seemed otherwise dangerous. One reason that TSCS data is of interest is that the richness of the data allows us to do many things; but many of those things should not be done.

By now, most political science articles appear to use our recommended methodology of OLS estimates of β coupled with PCSEs, with dynamics modeled via a lagged dependent variable. Of course this seems to me like a reasonable way to estimate the fully pooled TSCS model. My hope is that researchers will take the error correction dynamic model seriously, although in practice stationary models have appeared to perform adequately. The use of (single equation) error correction models does not require researchers to leave the simple OLS world. This is not to imply that error correction models are easy to estimate or that there are no issues in using them. But a full discussion of these issues would take this essay too far afield.

Now that estimation issues have been dealt with, interest should focus on specification. Presumably political scientists have a comparative advantage in specification, not estimation! One current area of research is modeling TSCS data with a binary dependent variable; this is a particularly hot topic in the study of international conflict. Although this can be thought of as an arena of high technique, the important issues have to do with specification—that is, do the BTSCS data look like event history data, or should they be modeled with a lagged latent variable in a dynamic probit setup? Two other pressing issues are the incorporation of spatial effects and the modeling of heterogeneity. The former seems straightforward, and many researchers might take pleasure in noting that they have been estimating spatial models all along. Heterogeneity appears to be a more difficult problem, although "high-tech" approaches such as Bayesian RCMs should not detract from the use of simpler ideas, such as cross validation. In the end, all these issues are primarily issues of specification, not (difficult) estimation. Thus, we can begin to return the modeling of phenomena in comparative politics and IR back to political scientists, rather than leaving it in the hands of econometricians or biometricians.

ACKNOWLEDGMENTS

This paper reports work done jointly with Jonathan N Katz, and the use of "our" or "we" denotes Katz and myself. Thanks to Geoffrey Garrett, John Oneal, and Bruce Russett for supplying the data used here, to Simon Jackman for helpful discussions, and to Chad Rector and Laurie Rice for comments on a previous draft.

LITERATURE CITED

Alt J. 1985. Political parties, world demand and unemployment: domestic and international sources of economic activity. *Am. Polit. Sci. Rev.* 79:1016–40

Anselin L. 1988. *Spatial Econometrics: Methods and Models*. Boston: Kluwer Acad. 284 pp.

Beck N. 2000. Issues in the analysis of time-series–cross-section data in the year 2000. In *Social Science Methodology in the New Millennium. Proc. Int. Conf. Logic Methodol., 5th, Cologne, Oct. 3–6*, ed. E de Leeuw. Cologne: Int. Sociol. Assoc.

Beck N, Katz JN. 1995. What to do (and not to do) with time-series cross-section data. *Am. Polit. Sci. Rev.* 89:634–47

Beck N, Katz JN. 1996. Nuisance vs. substance: specifying and estimating time-series–cross-section models. *Polit. Anal.* 6:1–36

Beck N, Katz JN. 1997. *The analysis of binary time-series–cross-section data and/or the democratic peace*. Presented at Annu. Meet. Soc. Polit. Methodol., July, Columbus, OH

Beck N, Katz JN. 2001. Throwing out the baby with the bathwater: a comment on Green, Kim and Yoon. *Int. Organ.* In press

Beck N, Katz JN, Tucker R. 1998. Taking time seriously: time-series–cross-section analysis with a binary dependent variable. *Am. J. Polit. Sci.* 42:1260–88

Beck N, King G, Zeng L. 2000. Improving quantitative studies of international conflict: a conjecture. *Am. Polit. Sci. Rev.* 94:21–36

Beck N, Tucker R. 1997. *Conflict in time and space*. Cent. Int. Aff. Work. Pap. 97–8, Harvard Univ., https://wwwc.cc. columbia.edu/sec/dlc/ciao/wps/tur01

Berry FS, Berry WD. 1990. State lottery adoptions as policy innovations: an event history analysis. *Am. Polit. Sci. Rev.* 84:395–415

Blanton SL. 2000. Promoting human rights and democracy in the developing world: US rhetoric versus US arms exports. *Am. J. Polit. Sci.* 44:123–31

Burkhart R, Lewis-Beck M. 1994. Comparative democracy: the economic development thesis. *Am. Polit. Sci. Rev.* 88:903–10

Cox DR. 1972. Regression models and life tables. *J. R. Stat. Soc. Ser. B* 34:187–220

Davidson J, Hendry D, Srba F, Yeo S. 1978. Econometric modelling of the aggregate time-series relationship between consumers' expenditures and income in the United Kingdom. *Econ. J.* 88:661–92

Fiorina M. 1994. Divided government in the American states—a byproduct of legislative professionalism. *Am. Polit. Sci. Rev.* 88:304–16

Fording R. 1997. The conditional effect of violence as a political tactic: mass insurgency, welfare generosity, and electoral context in the American states. *Am. J. Polit. Sci.* 41:1–29

Franzese RJ. 2001. *The Political Economy of Macroeconomic Policy in Developed Democracies*. New York: Cambridge Univ. Press. In press

Freedman D, Peters S. 1984. Bootstrapping a regression equation: some empirical results. *J. Am. Stat. Assoc.* 79:97–106

Garrett G. 1998. *Partisan Politics in the Global Economy*. New York: Cambridge Univ. Press. 208 pp.

Gasiorowski MJ. 2000. Democracy and macroeconomic performance in underdeveloped countries: an empirical analysis. *Comp. Polit. Stud.* 33:319–49

Gelman A, Carlin JB, Stern HS, Rubin DB. 1995. *Bayesian Data Analysis*. London: Chapman & Hall. 526 pp.

Giles M, Hertz K. 1994. Racial threat and partisan identification. *Am. Polit. Sci. Rev.* 88:317–26

Gleditsch KS, Ward MD. 2000. Peace and war in time and space: the role of democratization. *Int. Stud. Q.* 44:1–29

Green D, Kim SY, Yoon D. 2001. Dirty pool. *Int. Organ.* In press

Greene W. 1999. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall. 1004 pp. 4th ed.

Grier K, Tullock G. 1989. An empirical analysis of cross-national growth, 1951–80. *J. Monetary Econ.* 24:259–76

Hall P, Franzese R. 1998. Mixed signals: central bank independence, coordinated wage bargaining, and European monetary union. *Int. Organ.* 52:505–35

Hendry D, Mizon M. 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. *Econ. J.* 88:549–63

Hicks A, Kenworthy L. 1998. Cooperation and political economic performance in affluent democratic capitalism. *Am. J. Soc.* 103:1631–72

Hsaio C. 1986. *Analysis of Panel Data*. New York: Cambridge Univ. Press. 246 pp.

Iversen T. 1999. *Contested Economic Institutions: The Politics of Macroeconomics and Wage Bargaining in Advanced Democracies*. New York: Cambridge Univ. Press. 221 pp.

Jackman S. 2000a. Estimation and inference via Bayesian simulation: an introduction to Markov Chain Monte Carlo. *Am. J. Polit. Sci.* 44:375–404

Jackman S. 2000b. Estimation and inference are "missing data" problems: unifying social-science statistics via Bayesian simulation. *Polit. Anal.* 8:307–32

Kmenta J. 1986. *Elements of Econometrics*. New York: Macmillan. 655 pp. 2nd ed.

Liang K-Y, Zeger SL. 1986. Longitudinal data analysis using generalized linear models. *Biometrika* 73:13–22

Madalla GS, Hu H. 1996. The pooling problem. In *The Econometrics of Panel Data*, ed. L Mátyás, P Sevestre, pp. 307–22. Dordrecht: Kluwer Acad. 2nd ed.

Maoz Z, Russett BM. 1993. Normative and structural causes of democratic peace, 1946–1986. *Am. Polit. Sci. Rev.* 87:639–56

Oneal JR, Russett BM. 1997. The classical liberals were right: democracy, interdependence, and conflict, 1950–1985. *Int. Stud. Q.* 41:267–94

Parks R. 1967. Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated. *J. Am. Stat. Assoc.* 62:500–9

Pinheiro JC, Bates DM. 2000. *Mixed Effects Models in S and S-Plus*. New York: Springer. 528 pp.

Poe SC, Tate CN, Keith LC. 1999. Repression of the human rights to personal integrity revisited: a global cross-national study covering the years 1976–1993. *Int. Stud. Q.* 43:291–313

Radcliffe B, Davis P. 2000. Labor organization and electoral participation in industrial democracies. *Am. J. Polit. Sci.* 44:132–41

Smith K. 1997. Explaining variation in state-level homicide rates: Does crime policy pay? *J. Polit.* 59:350–67

Stone M. 1974. Crossvalidatory choice and assessment of statistical prediction. *J. R. Stat. Soc. Ser. B* 36:111–33

Su T-T, Kamlet M, Mowery D. 1993. Modeling United States budgetary and fiscal policy outcomes—a disaggregated, systemwide perspective. *Am. J. Polit. Sci.* 37:213–45

Swamy PAVB. 1971. *Statistical Inference in Random Coefficient Models*. New York: Springer-Verlag. 209 pp.

Western B. 1998. Causal heterogeneity in comparative research: a Bayesian hierarchical modelling approach. *Am. J. Polit. Sci* 42:1233–59

Annual Review of Political Science
Volume 4, 2001

CONTENTS

TRANSNATIONAL POLITICS: Contention and Institutions in International Politics, *Sidney Tarrow* 1
MECHANISMS IN POLITICAL PROCESSES, *Charles Tilly* 21
DEMOCRATIZATION AND ECONOMIC REFORM, *V. Bunce* 43
PSYCHOLOGY AND INTERNATIONAL RELATIONS THEORY, *J. M. Goldgeier, P. E. Tetlock* 67
POLITICAL TRADITIONS AND POLITICAL CHANGE: The Significance of Postwar Japanese Politics for Political Science, *Bradley Richardson, Dennis Patterson* 93
RELIGION AND COMPARATIVE POLITICS, *Anthony Gill* 117
TOWARD A FOURTH GENERATION OF REVOLUTIONARY THEORY, *Jack A. Goldstone* 139
POLITICAL CONSEQUENCES OF MINORITY GROUP FORMATION, *M. Hechter, D. Okamoto* 189
POLITICAL KNOWLEDGE, POLITICAL ENGAGEMENT, AND CIVIC EDUCATION, *William A. Galston* 217
THEORIES OF DELEGATION, *J. Bendor, A. Glazer, T. Hammond* 235
TIME-SERIES–CROSS-SECTION DATA: What Have We Learned in the Past Few Years?, *Nathaniel Beck* 271
THE ORIGINS AND WAR PRONENESS OF INTERSTATE RIVALRIES, *John Vasquez, Christopher S. Leskiw* 295
THE POLITICAL ECONOMY OF INTERNATIONAL MONETARY RELATIONS, *J. Lawrence Broz, Jeffry A. Frieden* 317
BIOLOGY AND POLITICS: Linking Nature and Nurture, *R.D. Masters* 345
WOMEN'S MOVEMENTS AT CENTURY'S END: Excavation and Advances in Political Science, *Karen Beckwith* 371
TAKING STOCK: The Constructivist Research Program in International Relations and Comparative Politics, *Martha Finnemore, Kathryn Sikkink* 391
VIACRATIC AMERICA: Plessy on Foot v. Brown on Wheels, *Douglas W. Rae* 417