Time-Series Cross-Section Methods

Nathaniel Beck

The Oxford Handbook of Political Methodology
Edited by Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier

Print Publication Date: Aug 2008
Subject: Political Science, Political Methodology, Comparative Politics
Online Publication Date: Sep 2009 DOI: 10.1093/oxfordhb/9780199286546.003.0020

Abstract and Keywords

This article outlines the literature on time-series cross-sectional (TSCS) methods. First, it addresses time-series properties including issues of nonstationarity. It moves to cross-sectional issues including heteroskedasticity and spatial autocorrelation. The ways that TSCS methods deal with heterogeneous units through fixed effects and random coefficient models are shown. In addition, a discussion of binary variables and their relationship to event history models is provided. The best way to think about modeling single time series is to think about modeling the time-series component of TSCS data. On the cross-sectional side, the best approach is one based on thinking about cross-sectional issues like a spatial econometrician. In general, the critical insight is that TSCS and binary TSCS data present a series of interesting issues that must be carefully considered, and not a standard set of nuisances that can be dealt with by a command in some statistical package.

Keywords: time-series cross-sectional methods, nonstationarity, heteroskedasticity, spatial autocorrelation, binary variables, fixed effects, random coefficient models, heterogeneous units

TIME-SERIES cross-section (TSCS) data consist of comparable time-series data observed on a variety of units. The paradigmatic applications are to the study of comparative political economy, where the units are countries (often the advanced industrial democracies) and where for each country we observe annual data on a variety of political and economic variables. A standard question for such studies relates to the political determinants of economic outcomes and policies. There have now been hundreds of such studies published.[1]

TSCS data resemble "panel" data, where a large number of units, who are almost invariably survey respondents, are observed over a small number of "waves" (interviews). Any procedure that works well as the number of units gets large should work well for panel data; however, any procedure which depends on a large number of time points will not necessarily work well for panel data. TSCS data typically have the opposite structure to panel data: a relatively small number of units observed for some reasonable length of time. Thus methods that are appropriate for the analysis of panel data are not necessarily appropriate for TSCS data and vice versa.[2]

(p. 476) All of these types of data are a particular form of “multilevel” (or “hierarchical”) data. Multilevel data are organized by grouping lower-level observations in some meaningful way. Thus in the study of education, we may observe students, but students can be thought of as members of a class, and classes can be thought of as members of a school. It is often helpful to think of TSCS data as multilevel data, but with much additional structure.

Political scientists have been familiar with panel data in the guise of election studies for well over half a century. TSCS data has become popular only more recently. Adolph, Butler, and Wilson (2005) examined political science articles in journals indexed in JSTOR. They found relatively few uses of TSCS terminology before 1975, with an explosion of articles using terms related to TSCS analysis starting in the late 1980s. They report that 5 percent of all political science articles from 1996 to 2000 made reference to TSCS or panel terminology and found approximately 200 empirical analyses using TSCS data during that period. Beck and Katz's (1995) methodological discussion of TSCS data is the most heavily cited *American Political Science Review* article published since 1985.

Why the explosive growth in TSCS analyses starting in the mid-1980s? Credit must be given to Stimson's (1985) pioneering essay discussing the importance of both panel and TSCS data in political science. This was the first political science article to discuss general methodological issues related to TSCS data. But it should also be noted that students of comparative politics were ready to hear Stimson's message in 1985.

Most quantitative comparative political science before 1985 consisted of cross- sectional regressions. Researchers interested in the political economy of advanced industrial democracies (Cameron 1978; Lange and Garrett 1985) found themselves running regression on fifteen observations, and hence involved with controversies that relied heavily on whether one particular influential observation (often Norway) completely determined the findings. Since it was not possible to either create more advanced industrial democracies, nor to add nonadvanced industrial democracies to the data-set, the ability to add many additional observations in the temporal domain was clearly attractive (Alvarez, Garrett, and Lange, 1991). TSCS analyses are now the standard in studies of comparative political economy. This, combined with the dyad- year studies of conflict related to the “Democratic Peace” (for example, Maoz and Russett 1993, which is the second-most cited *American Political Science Review* article since 1990), account for much of the popularity of TSCS analyses.

1 Exploratory Analysis

TSCS analysts, like all analysts, should initially examine their data to discern important properties. TSCS data should be examined in the usual ways for whether the data seem to have long tails or skewness, whether there are extreme outliers, and other (p. 477) such issues that must be examined before any regression-type analysis is undertaken. In addition, analysts should examine whether the various cross-sectional units appear similar, whether there is substantial variation in each unit over time, and whether the data show interesting time-series properties. This is most relevant for the dependent variable or variables.

One can look at the time-series properties of the data by usual time-series methods. Thus one can plot the data against time to examine for trends, one can look at correlograms (constructed so as to respect the grouping of the data, which requires a program that is aware of the time-series and cross-sectional properties of the data), and one can look at the autoregression to examine whether the variables have unit roots. While testing for unit roots in such data is a new and active area (Im, Pesaran, and Shin 2003), it surely is easy enough to run an autoregression on TSCS data and see whether the coefficient of the lagged variable is near one.

For exploratory analysis of cross-sectional issues, the best method is the box plot, plotting variables by unit. Since the units are typically both meaningful (countries) and not too numerous (about twenty), much information can be gleaned from these plots. In particular, one can discern if the center and spread of the variables differs by unit, or whether one or a few units are considerably different from the bulk of the units. At that point investigators could consider whether one unit should be dropped from the analysis, or whether this unit to unit heterogeneity (in mean and/or variance) can be modeled.

2 Notation

We will work with a single equation setup. While it is possible to extend this to a simultaneous equations framework, such extensions are extremely rare. We use the subscript i to denote unit, and t for time. t always denotes calendar time (typically a year), so that the same t subscript in different units stands for the same time point. The dependent variable is $y_{i,t}$ with the k-vector of independent variables being $x_{i,t}$. A very simple (pooled) model is

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

This assumes that the data structure is rectangular; that is, each of the N units are observed for the same T time periods. It is easy to handle units that start and end at slightly different time points; data missing in the interior of the time frame can be handled by standard multiple imputation methods (King et al. 2001).

Of course the important issues relate to what independent variables we include to explain y. While one more good predictor of y is worth more than all the various econometric manipulations, we focus on those manipulations, assuming that analysts (p. 478) have done an appropriate job of choosing explanatory variables (and related issues such as functional form).

For the econometric issues, we denote the covariance of the errors by Ω (with variance terms on the diagonal). This is an $NT \times NT$ matrix, with typical element $E(\epsilon_{i,t}\epsilon_{j,s})$. If Ω meets the Gauss-Markov assumptions (that is, all off-diagonal elements are zero and all diagonal elements are identical) then OLS is appropriate and estimation is simple. If the Gauss-Markov assumptions are not met then OLS will still be consistent, but it will not be fully efficient and the reported standard errors will be inaccurate. Much attention has been paid to how the Gauss-Markov assumptions might be violated for TSCS data, and how to improve estimation given those violations. We begin with issues related to the time-series properties of the data, and then consider cross-sectional complications.

3 Dynamic Issues

3.1 Stationary Data

The easiest way to understand the implications of the time-series properties of TSCS data is to remember that for each unit we simply have time-series data, and hence all that we know about time-series data holds (*mutatis mutandis*) for TSCS data. Thus all the standard tests for whether the data are serially independent continue to apply, all the estimation methods for dealing with serially dependent data still work, and all the issues relevant to modeling dynamic processes continue to be relevant.[3] Since TSCS data are often annual, we restrict ourselves here to first-order processes; the generalization to higher-order processes is straightforward.

As always, we can view the dynamics issue as an estimation nuisance or we can treat dynamics as a serious modeling challenge. The latter approach is favored by modern analysts (Hendry, Pagan, and Sargan 1984; Hendry and Mizon 1978). But of course we still need to make sure that the estimation method chosen is appropriate. We begin by noting that dynamic specifications often include a lagged dependent variable. Whether or not the specification contains a lagged dependent variable, OLS is appropriate (at least with respect to time-series issues) so long as the errors are serially independent. It is always appropriate to test for serially correlated errors using a Lagrange multiplier test. This test is easy to implement. One first runs OLS, then regresses the OLS residuals on the lagged values of the residuals (one lag to test for first order serial correlation) and all the other independent variables in the specification, and then compares the statistic NTR^2 to a x^2 distribution with one (p. 479) degree of freedom. If one fails to reject the null of no serial correlation of the errors, then OLS is appropriate.

If the errors are serially correlated then one can estimate the model by either maximum likelihood or feasible generalized least squares. While iterative methods such as Cochrane—Orcutt may find a local maximum if there is a lagged dependent variable in the specification, this is not usually a problem in practice, and if one is worried, starting the iterations at a few different sets of parameter variables will invariably find the global maximum (Hamilton 1994, 226). But in typical situations the inclusion of a lagged dependent variable in the specification will eliminate almost all serial correlation of the errors (since the lagged dependent variable implicitly includes lagged error terms into the specification).

To see that the use of lagged dependent variables is not fundamentally different from a static model with serially correlated errors, note that both models are a specialized form of the more general "auto-regressive distributed lag" model

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

where the errors are serially independent. The lagged dependent variable model assumes $\gamma = 0$ whereas the model with serially correlated (first-order autoregressive) errors assumes $\gamma = -\beta_\phi$. One can always estimate the more general autoregressive distributed lag model and test whether the data support either of the two restrictions on γ This is easy to do via OLS and surely worth doing in almost all cases.

3.2 Nonstationary Data

There is one additional issue that has important consequences for modeling dynamics: unit roots. In equation (2) ϕ may be very near one. While the exact test statistic for determining whether we can reject that there is a unit root is controversial (Im, Pesaran, and Shin 2003), the presence of unit roots (or nonstationarity) has critical consequences. Estimating a model with either a lagged dependent variable or serially correlated errors in the presence of a unit root can lead to dramatically misleading results (spurious regressions). Thus after estimating a dynamic model one should certainly see if the residuals appear stationary (that is, whether an autoregression of the residuals on their lags shows a coefficient on the lagged residual term near one), as well as examine whether the coefficients on any lagged dependent variable terms are near one. A precise test of the null that the data have no unit roots is much less important than seeing whether certain coefficients are very near one (especially given the large sample sizes, and hence small standard errors, typically seen in TSCS data).

What should analysts do if they find that the data appear to be nonstationary? Clearly they should not ignore the problem. They could simply model the short run; that is, take first differences, giving up on any attempt to model the long run. Alternatively, they could use the various error-correction models associated with the work of Engle and Granger (1987) and Davidson et al. (1978), though almost (p. 480) all the work associated with this approach is for single time-series data rather than TSCS data. Currently this must be marked as an area for further study, though the consequences for ignoring this issue maybe grave.

4 Cross-sectional Issues

If, after suitable modeling, we find serially independent errors, there may remain violations of the Gauss-Markov assumptions due to cross-sectional (spatial) complications in the data. In particular, the errors for the different units may have differing variances (panel heteroskedasticity) or may be correlated across units. If there is correlation of errors across the units, it is almost invariably assumed both to not vary over time and to only occur for units observed at the same time point; thus the typical violation allowed for is contemporaneous correlation of the errors. We would expect contemporaneous correlation of the errors to be likely in studies of political economy in open economies; shocks that affect one nation can also be expected to affect its trading partners (either as a common shock or through the unexpected impact on trade).

4.1 Traditional Approaches

Researchers concerned about such spatial violations turned to “feasible generalized least squares” (FGLS). This method uses OLS to estimate the model, then takes the residuals from OLS to estimate the covariance matrix of the errors, Ω, and then, based on that estimate, transforms the data so the resulting transformed observations satisfy the Gauss-Markov assumptions. The FGLS procedure for contemporaneously correlated and panel heteroskedastic errors was derived by Parks (1967).

Beck and Katz (1995) showed that this procedure has extremely poor statistical properties unless $T >> N$, which is rare. Thus this method is seldom used any more. Researchers worried about correcting standard errors due to contemporaneously correlated and panel heteroskedastic errors can use "panel corrected standard errors" (PCSEs) in place of the OLS standard errors.

This consists of using the usual formula for the OLS standard errors when the Gauss-Markov assumption that $\Omega = \sigma^2 I$ is violated. The covariance of the OLS estimate of β in this case is

$$(\mathbf{X}'\mathbf{X})^{-1}\left(\mathbf{X}'\Omega\,\mathbf{X}\right)(\mathbf{X}'\mathbf{X})^{-1}. \tag{3}$$

Under the assumptions that the errors at different time points are independent, Ω is a block diagonal matrix, with the blocks down the diagonal being $N \times N$ matrices of the contemporaneous covariances of the errors (and diagonal terms being the unit (p. 481) specific variances). Let V be this matrix. V can be estimated using the T replicates of the OLS residuals, e_{it} (since the covariance matrix of the errors is assumed to be stable over time). Variances and covariances are estimated in the obvious way (so $V_{i,j}$ is estimated by

$$\frac{\sum_{t=1}^{T} e_{i,t} e_{j,t}}{T}).)$$

. PCSEs are then the square roots of the diagonal terms of

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\hat{\mathbf{V}} \otimes \mathbf{I}_T\right)\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \tag{4}$$

where ⊗ is the Kronecker product. Since the estimate of the V matrix is based on an average of T replicates, it has good properties for typical TSCS data.

It might appear that this procedure is similar to White's (1982) heteroskedasticity consistent standard errors. But White's procedure is only robust to heteroskedasticity, and does not assume that variances are constant within a unit; it also does not allow for the errors to be contemporaneously correlated. White's procedure also depends heavily on asymptotics, since there is one variance term per observation. With PCSEs there are T observations per estimate, and so as long as T is large the procedure performs well. (The performance of PCSEs is purely a function of T, not N.) Monte Carlo experiments showed that PCSEs are very close to OLS standard errors when the Gauss-Markov assumptions hold, and can be considerably better than OLS standard errors when those assumptions are violated so long as $T > 15$ (Beck and Katz 1995).

4.2 Spatial Insights

PCSEs simply correct the OLS standard errors because the simplifying assumptions of OLS are likely invalid for TSCS data; it clearly would be better to model the process and use that model to improve estimation. Steps along this line can be taken using the ideas of spatial econometrics (Anselin 1988). The FGLS procedure allows for an arbitrary contemporaneous correlation of the errors; by not imposing any structure on these errors an inordinate number of extra parameters must be estimated. Spatial methods allow for a simple parameterization of the correlation of the errors, allowing for better estimation of TSCS data with such errors. This parameterization is known as the "spatially lagged error" model. After examining this, we shall see that an alternative, the "spatial autoregressive" model, may be even more useful.

Spatial methods assume an a priori specified weighting matrix which ties together the errors. Thus it differs from FGLS methods which allow for an arbitrary pattern of correlation of the errors. Of course the spatial methods are only superior if the specified weights are close to being correct. Geographers often assume that geographically nearby units are similar, but any method of weighting can be appropriate. Beck, Gleditsch, and Beardsley (2006) argued that for studies of political economy it is often better to assume that nations are tied together by their level of trade (in proportion to GDP).

Let W denote the (pre-specified) spatial weighting matrix, which is $NT \times NT$. Letting $w_{i,t,j,s}$ be a typical element of this matrix, the assumption that all error correlations are contemporaneous implies the $w_{i,t,j,s} = 0$ if $t \neq s$. For contemporaneous (p. 482) observations we would have nonzero weights (and by definition $w_{i,t,i,t} = 0$). The weights are then typically normalized to each row sums to one.

The spatially lagged errors model then has

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta} + \epsilon_{i,t} + \lambda \mathbf{W}_{i,t}\boldsymbol{\epsilon}_t \qquad (5)$$

where $\mathrm{w}_{i,t}$ corresponds to the row of W corresponding to unit i and ϵ is the vector of errors for the other units, both at time t.
Thus there is only one extra parameter, λ, to estimate. Such a model is easy to estimate using standard spatial econometrics packages and usually produces estimates with good properties. If $\lambda = 0$, this reduces to the standard nonspatial linear regression model.

It should be stressed that this setup assumes that the errors are serially independent (perhaps after including some lagged variables in the specification). It also does not allow for panel heteroskedasticity, though this is often not a problem in practice if the dependent variables are suitably measured (say as a proportion of GDP).

The spatial lagged error model assumes that the errors across units are related, but otherwise the observations are independent. Another plausible model is that the dependent variables are linked; if the dependent variable is, say, unemployment, a country will have more unemployment if its trading partners have more unemployment (with weighting matrix as before). If we are willing to assume that these spatial effects occur with a temporal lag, and if there is no remaining serial correlation of the errors, the spatial autoregressive model (also temporally lagged) has

$$y_{i,t} = \mathbf{x}_{i,t}\beta + \gamma \mathrm{W}_{\mathrm{i,t}}\mathbf{y}_{t-1} + \epsilon_{i,t} \qquad (6)$$

where y_{t-1} is a vector of the lagged value of y for all the other units. While this spatial autoregressive lag has often been included in specifications based on informal reasoning, it is best to think about this model in the spatial econometrics context. In particular, this model is only easy to estimate if the errors are serially independent (perhaps after inclusion of the temporal lagged $y_{i,\,t-1}$ in the specification). If the errors are serially independent the model can be estimated by OLS; one simply adjoins to the model the lagged weighted average of the dependent variable for all other units. This simplicity comes from assuming that the spatial lag enters also with a temporal lag. Whether or not this assumption is compelling varies by substantive problem.

5 Heterogeneous Units

So far we have assumed a fully pooled model; that is, all units obey the same specification with the same parameter values. There are many ways to allow for unit to unit heterogeneity in equation (1) (or its dynamic counterparts). In keeping with the spirit of TSCS data, we will assume that there is intra-unit homogeneity.

(p. 483) 5.1 Fixed Effects

The simplest way to allow for unit heterogeneity is to allow the intercepts to vary by unit, the "fixed effects" model. This has

$$y_{i,t} = \mathbf{x}_{i,t}\beta + f_i + \epsilon_{i,t}. \qquad (7)$$

(In this notation there can be no constant term.) This model is easy to estimate by OLS. If the fixed effects are omitted there will be omitted variable bias if the fixed effects both explain y and are correlated with $\mathbf{x}$ (Green, Kim, and Yoon 2001). Researchers should clearly check for the possibility that there are fixed effects by estimating equation (7) and then testing the null hypothesis that all the f_i are equal. This is most easily done via a standard F-test on a model with a constant term and one particular effect dropped, with the null being tested that all the other effects are zero.

The fixed effects model is equivalent to unit centering all observations, so that the only question at issue is whether temporal variation in **x** is associated with temporal variation in *y*; all cross-sectional effects are eliminated by the unit centering. This makes it impossible to estimate the impacts of any variables that do not vary over time; it also becomes almost impossible to estimate the effects of variables that seldom change. This is a serious problem in political economy, where we often care about the impact of institutions, which almost by definition change slowly if at all.

Fortunately it is often the case that well-specified models do not require fixed effects. Ideally one would like to explain the effects by substantive variables, and not simply conclude that Germany grew faster because it was Germany. It will also often be the case that misspecified dynamics may lead to the unnecessary use of fixed effects (Beck and Katz 2001). And it may be better to decide not to include fixed effects if the test of the null of no fixed effects only marginally rejects that null. But where fixed effects are needed in the model, failure to include them can lead to omitted variable bias.

It might appear that random effects would solve the problems caused by fixed effects. Random effects models are similar to equation (7) except the f_i 's are seen as draws from a normal distribution. Unfortunately the random effects model assumes that the effects are orthogonal to the independent variables, thus assuming away all the interesting issues which lead to omitted variable bias. While random effect models may be very useful for panel data, they do not deal with the important issues of heterogeneity in TSCS data.

5.2 Assessing Heterogeneity by Cross-validation

A useful tool to examine heterogeneity is cross-validation (leaving out one unit at a time). Cross-validation is a well-known technique for evaluating models by seeing how well they “predict” data that have not already been used to estimate the model (p. 484) (Stone 1974). Cross-validation can easily be used with TSCS data both to assess models and to see whether some units appear not to pool.

For TSCS data, cross-validation proceeds by estimating the model leaving out one unit, and then using those estimates to “predict” the values of y for the omitted unit. This is done for each unit. Models can then be compared according to the mean squared error of prediction. The idea behind cross-validation is that by examining predictions for data that were not used to estimate the model we can ensure against overfitting. (This becomes more critical as we move to more complicated models with more independent variables or nonlinear functional forms, where it is easier to overfit the data.)

Cross-validation (leaving out a unit at a time) can not only be used to compare models, but can also be used to see if one (or a few) units seem to follow a different pattern than do the bulk of other units. While it is clearly bad practice to simply exclude from data-sets units that fit the model less well, it does make sense to assess whether all units belong in the model. Thus it might make sense to exclude a unit from the analysis if it has *both* large cross-validation prediction errors and might be thought different from other units on some other grounds. Thus, for example, even if Turkey were included in a European Union data-set (at some time in the future), we might suspect that it should not be included in a model of the political economy of advanced democracies. It will often be the case in studies of comparative politics that we know a lot more about the units than just the values of the variables.

Cross-validation is useful for finding whether one or a small number of units should be excluded from the regression. We may also suspect that different units follow different regimes. Thus the older West European countries may not pool with the newer East European members of the European Union. This hunch could be tested by creating an East European dummy variable and then interacting it with all the independent variables with a standard *F*-test used to assess whether the interaction terms substantially improve goodness of fit. It also might be the case that an East European dummy variable would do as well as the various country fixed effects. The fact that some data-collection organization collects data on a set of units does not mean that those units are sufficiently homogeneous to allow for estimation of the fully pooled model.

5.3 Random Coefficient Models

It is, of course, possible to estimate a model for each unit separately, with

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\beta}_i + \epsilon_{i,t} \qquad (8)$$

(with the appropriate additions for dynamics). If T is large enough it is not ridiculous to estimate N separate time series (and time-series analysis on single countries surely has a long tradition). But T will typically not be large enough for unit time-series analyses to be sensible. (Beck and Katz 2007 found with simulated data that the fully (p. 485) pooled model gives better estimates of the unit β_i even when there is heterogeneity when $T < 30$.) And even for larger T's, separate time-series analyses on each country make it difficult to claim that one is doing comparative politics.

A very nice compromise is the “random coefficients model” (RCM). This allows for unit heterogeneity, but also assumes that the various unit-level coefficients are draws from a common (normal) distribution. Thus the RCM adjoins to equation (8)

$$\boldsymbol{\beta}_i \sim N(\boldsymbol{\beta}, \Gamma) \qquad (9)$$

where Γ is a matrix of variance and covariance terms to be estimated. Γ indicates the degree of heterogeneity of the unit parameters (Γ = 0 indicates perfect homogeneity). An important (and restrictive) assumption is that the stochastic process which generates the β_i is independent of the error process, and is also uncorrelated with the vector of independent variables, although of course this assumption is less restrictive than the assumption of complete homogeneity. The RCM has a venerable heritage, going back at least to Smith (1973); Western (1998) first discussed this model in the context of comparative politics.

While this model is often estimated by Bayesian methods, it is also feasible to estimate it via standard maximum likelihood (Pinheiro and Bates 2000). While until recently it was hard for researchers to estimate the RCM, innovations in commonly used software packages no longer allow for this excuse. Beck and Katz (2007) found that the RCM gave superior estimates of the overall β, whether or not there was significant unit heterogeneity, and also provided good estimates of the unit β_i. These estimates of the β_i are an average of the fully pooled and the separate unit by unit estimates, where the amount of averaging is a function of both how disparate the unit by unit estimates are and the confidence we have in those estimates (their standard errors). Thus the RCM estimates of the β_i "shrink" the unit by unit estimates back toward the pooled estimate, with the degree of shrinkage chosen empirically; alternatively we can talk of improving the low-precision unit by unit estimates by allowing them to "borrow strength" from the other estimates.

Since we would expect there to be some amount of unit heterogeneity, analysts should routinely estimate the RCM. They may then decide there is enough homogeneity to go back to the fully pooled model. Since the RCM provides good estimates of the unit β_i, the method can also be used to see whether the parameters for some unit or units are sufficiently different from the other units that we should not pool the two types of observations.

The RCM is a special case of the multilevel model; yearly data are nested inside countries. Unlike the more general data, the RCM assumes that the lower- level data are connected by being yearly observations (rather than, say, being survey respondents nested by country). Thus one can use all the various dynamic specifications we have discussed as well as the various multilevel methods in analyzing the RCM.

One especially nice feature of the RCM (inherited from the multilevel model) is that we can model the variation of the unit coefficients as a function of unit-level (p. 486) variables. Thus the marginal effect of some $x_{i,t}$ on $y_{i,t}$ can be made dependent on some unit level z_i. In particular, we can make equation (9) more general by

$$\beta_i = \alpha + \mathbf{z}_i \kappa + \mu_i \tag{10}$$

where the μ are drawn from k-variate normal distribution. The ability to model the relationship between independent and dependent variables as conditional on unit- level covariates makes the

RCM an extremely powerful tool of comparative analysis. It allows us to move from saying that the effect of some variable is different in France and Germany to this impact differs because of some institutional difference between the two nations.

6 Binary Dependent Variables

So far we have restricted ourselves to models with a continuous dependent variable. This works well in the study of comparative political economy, where the dependent variable is usually an economic policy or outcome. But TSCS data are also common in the study of international relations (Maoz and Russett 1993); here the dependent variable is often dichotomous, with the most common example being whether or not a pair of nations are in conflict in any given year.[4]

The most common data setup in the study of international relations is the “dyad- year” design. This is a form of TSCS data, but the units are pairs of actors (usually pairs of nations) observed over time (usually annually). The dependent variable may be continuous (as in the study of international trade) or binary (as in the study of conflict). The dyads may either be directed or undirected dyads; with directed dyads the dependent variable is what is sent to a receiver (so AB and BA are different dyads) whereas with undirected dyads there is no distinction between sender and receiver (so the dyad AB is the same as the dyad BA).

Dyad-year data present an additional complication over and above standard TSCS data: The dyads AB and AC are not likely to be fully independent. Beck, Gleditsch, and Beardsley (2006) discussed how to use spatial econometric methods to deal with this issue in the context of trade. One simply allows for a spatially lagged error, where (p. 487) dyads containing an overlapping member are considered to be adjacent and others are considered to be nonadjacent. The problem is much more complicated for a binary dependent variable.

Binary TSCS (BTSCS) data present other problems which are likely to be more important. In particular, the lack of a simple residual makes it harder to model either the time-series or the cross-sectional property of the error process. We can think of BTSCS data as generated by realizations of ones and zeros from underlying latent variable, so

$$y^*_{i,t} = \mathbf{x}_{i,t}\beta + \epsilon_{i,t} \tag{11}$$

$$y_{i,t} = \begin{cases} 1 \text{ if } y^*_{i,t} > 0 \\ 0 \text{ otherwise.} \end{cases} \tag{12}$$

While all would be easy if we observed the latent *y**, we do not. Thus we cannot use the “prediction errors” from a logit[5] analysis of the realized y to make inferences about Ω, the covariance matrix of the *latent* errors.

Poirier and Ruud (1988) showed that “ordinary” probit (that is, standard probit analysis assuming serially independent errors in the latent variable) is consistent in the presence of serially correlated errors; ordinary probit is, however, inefficient and produces incorrect standard errors; the same is true for ordinary logit. Poirier and Ruud showed that the standard errors could be corrected by using a robust covariance matrix. Poirier and Ruud's results are for a single time series, but they also hold for the time-series component of BTSCS data. Beck, Katz, and Tucker (1998) showed that using Huber's (1967) robust standard errors assuming that all observations for a unit are correlated provides reasonably correct assessments of the variability of the ordinary logit standard errors, even though the Huber procedure does not take into account that the error structure in a unit should be that of serial correlation. But we can do much better than simply fix the standard errors.

6.1 Event History Approaches

BTSCS data can be seen as sequences of zeros terminated by a one and sequences of ones terminated by a zero. Thus we either can see the data as yearly binary data or as event history data giving the lengths of sequences of ones and zeros, with time- varying covariates.[6] The latter approach is more informative. In event history analysis our interest is in explaining the lengths of spells; that is, the lengths of sequences of zeros terminated by a one. If we remove from the data-set all the sequences of ones after the first one, we have standard discrete time-duration data. This can be analyzed by standard discrete time methods, with the most common being the discrete time version of the Cox (1975) proportional hazards model.

(p. 488) Beck, Katz, and Tucker (1998) showed that logit analysis can be seen as estimating the discrete time yearly hazard of transitioning from a zero to a one (peace to conflict) *if* all years with a conflict after the first are removed from the data. The hazard rate is the rate of transitioning from a zero to a one conditional on not having transitioned previously; it is the elimination of the subsequent one observations that enforces this conditioning. Beck, Katz, and Tucker showed that a logit analysis which added a series of dummy variables marking the year since the last event (that is, the number of preceding zeros) is completely analogous to the discrete time Cox model. The years since the last event dummies correspond to the integral over a year of the Cox baseline hazard; since the baseline hazards are unspecified, no information is lost in treating their yearly integrals as a dummy variable.[7] These temporal dummies account, in the logit analysis, for what is known in the event history world as duration dependence; that is, the chance of a spell terminating varies with the length of that spell.

The tie between event history analysis and the yearly logits is seen by noting that the Cox proportion hazards model assumes that the instantaneous hazard rate (the rate at which transitions from zero to one occur in a small time interval given no prior transition) has the form

$$h_i(t) = h_0(t)e^{\mathbf{x}_{i,t}\beta} \qquad (13)$$

where *i* refers to units, t refers to continuous time, and $h_0(t)$ is an unspecified baseline hazard function.
We only observe whether or not a transition occurred between time *t* - 1 and *t* (assuming annual data) and are interested in the probability of a transition in year *t*, $P(y_{i,t} = 1)$. Assuming no prior transitions, and standard event history formulae, we get

$$P(y_{i,t} = 1) = 1 - \exp(-\int_{t-1}^{t} h_i(\tau)d\tau) \qquad (14a)$$

$$= 1 - \exp(-\int_{t-1}^{t} e^{\mathbf{x}_{i,t}\beta}h_0(\tau)d\tau) \qquad (14b)$$

$$= 1 - \exp(-e^{\mathbf{x}_{i,t}\beta}\int_{t-1}^{t} h_0(\tau)d\tau) \qquad (14c)$$

Since the baseline hazard is unspecified, we can just treat the integral of the baseline hazard as an unknown constant. Defining

$$\alpha_t = \int_{t-1}^{t} h_0(\tau)d\tau \text{ and} \qquad (15)$$

$$\kappa_t = \log(\alpha_t) \qquad (16)$$

(p. 489) we then have

$$P(y_{i,t} = 1) = 1 - \exp\left(-e^{\mathbf{x}_{i,t}\beta}\alpha_t\right) \qquad (17a)$$

$$= 1 - \exp\left(-e^{\mathbf{x}_{i,t}\beta+\kappa_t}\right). \qquad (17b)$$

This is exactly a binary dependent variable model with a cloglog “link.” Thus the Cox proportional hazard in continuous time is *exactly* a binary dependent variable with dummy variables marking the length of the spell and a cloglog link. Since there is seldom reason to

prefer one binary link function to another, logit (or probit) should be an adequate substitute for the cloglog model for researchers who are more familiar with logit and probit.
This method allows analysts to assess the impact of the independent variables on either the length of spells of zeros or on the probability of transitioning from a zero to a one given that a unit had not previously transitioned. One might then estimate a completely different model for the lengths of spells of ones or the probability of transitioning from a one to a zero given that a unit had not previously transitioned in this manner. Thus we have one model that determines why dyads remain at peace and another for lengths of conflict; there is no reason those models should be the same. Note that ordinary logit assumes that the same model determines the transition from zero to one and from one to one since in ordinary logit the probability of observing a one is not conditional on whether a unit was previously a one or a zero.

Thinking of BTSCS data as event history data has other advantages. It forces us to think about both left and right censoring. Left censoring is the problem caused by not knowing that the first year of a data-set was the first year of a spell of zeros. There is no easy solution here but at least the event history approach makes the problem clear. Right censoring occurs when a unit ends with a spell of zeros. This causes no problems since a right censored observation is just marked with a string of zeros that do not terminate with a one.

The event history approach also helps deal with dyads that transition several times from zero to one and then back to zero again. In event history analysis this is called repeated spells. The one thing we are sure of is that second spells are usually different from first spells, and so on. Ordinary logit assumes that all spells are identical. The event history approach would lead, at a minimum, to using as an explanatory variable the number of previous transitions from zero to one that a unit has experienced.

Since BTSCS and grouped event history data are identical, it is also possible to use event history methods rather than binary data methods when the former allow for simpler solutions to hard problems. Thus, for example, Jones and Branton (2005) use the Cox proportional hazard model rather than a logit-based method to deal with issues of repeated events and “competing risks” (where a spell can end in one of several ways) in a study of policy innovation in the American states. Note that since competing risk data look exactly like unordered multichotomous data, event history methods can enable researchers to move beyond dichotomous to multichotomous dependent variables.

(p. 490) 6.2 Markov Transition Models

The event history approach is very similar to the Markov transition model (Amemiya 1985; Ware, Lipsitz, and Speizer 1988); this approach is best known in political science through the work of Przeworski et al. (2000). Beck et al. (2001) show the similarity of the two approaches. The Markov transition model assumes that the probability of observing a one in the current year is a function of covariates and whether or not a one was observed in the prior year; as in the event history approach there is a different model for transitioning from zero to one and for one remaining one (which is just one minus the probability of a transition from one to zero).

The Markov transition model can be written as:

$$P(y_{i,t} = 1) = \text{logit}(\mathbf{x}_{i,t}\boldsymbol{\beta} + \mathbf{x}_{i,t}y_{i,t-1}\boldsymbol{\gamma}). \tag{18}$$

β is the impact of the independent variables when the prior state of y was zero, while β + Γ is the similar impact when the prior state of y was one. If γ = 0 then the probability of being a one currently does not depend on the prior state of y. In that case one can model $P(y_{i,t} = 1)$ without reference to the lagged $y_{i,t-1}$ and so ordinary logit would be appropriate. Note that this is a testable hypothesis, not a necessary assumption.

As with the event history approach, there is no assumption that transitions to one from zero are modeled the same way as transitions from one to zero. Unlike the event history approach, the transition model assumes duration independence and so the transition model is a special case of the event history model. The hypothesis that there is no duration dependence can be tested by an F-test on the null hypothesis that all the spell time dummy variables are zero. But by allowing the transitions from zero to one to differ from the reverse transitions makes the Markov transition model a much more plausible alternative than ordinary logit.

7 Conclusion

TSCS (and BTSCS) data present many interesting complications. As always we can treat these complications as an estimation nuisance or as interesting substantive issues to be modeled. The latter approach is usually preferable.

The best way to think about modeling single time series is the best way to think about modeling the time-series component of TSCS data. For the last decade or so a good methodology for time-series data has been well known (based on the autoregressive distributed lag model) and there is no reason not to use that methodology for stationary TSCS data. The data may appear to have unit roots, and at that point analysts should surely consider using the methods associated with analyzing nonstationary data.

(p. 491) On the cross-sectional side the best approach is one based on thinking about cross- sectional issues like a spatial econometrician. Economies are linked, innovations diffuse. These are things to be modeled, not econometric nuisances to be eliminated. There are better and worse ways to eliminate nuisances, but those methods should not be the first choice.

Allowing for parameter heterogeneity is now easy to do and makes much sense. The random coefficient model seems to perform very well and the idea behind it is very attractive. Explaining difference in the effect of the independent variables from unit to unit by unit-level variables is also a very attractive feature of this model. There is no reason not to start with the random coefficient model, and then test for whether there is nontrivial heterogeneity.

Turning to BTSCS data, the key insight is to think about such data as event history data. At that point either the Markov transition model or the logit analysis with the dummy variables counting spell length seem attractive. Thinking about BTSCS data as event history data also leads to a variety of insights that come less easily to those who simply think of this as binary dependent variable data. What is most critical is to realize that understanding the causes of the lengths of sequences of zeros is very different from understanding the length of sequences of ones.

In all cases the critical insight is that TSCS and BTSCS data present a series of interesting issues that must be carefully considered, and not a standard set of nuisances that can be dealt with by a command in some statistical package. There is never a statistical panacea, and there is no such panacea for TSCS or BTSCS data.

References

ADOLPH, C. BUTLER, D. M. and WILSON, S. E. 2005. Which time-series cross-section estimator should I use now? Guidance from Monte Carlo experiments. Presented at the Annual Meeting of the American Political Science Association, Washington, DC.

ALVAREZ, R. M. GARRETT, G. and LANGE, P. 1991. Government partisanship, labor organization and macroeconomic performance. *American Political Science Review*, 85: 539–56.

AMEMIYA, T. 1985. *Advanced Econometrics*. Cambridge, Mass.: Harvard University Press.

ANSELIN, L. 1988. *Spatial Econometrics: Methods and Models*. Boston: Kluwer Academic.

BECK, N. EPSTEIN, D. JACKMAN, S. and O'HALLORAN, S. 2001. Alternative models of dynamics in binary time—series-cross-section models: the example of state failure. Presented at the Annual Meeting of the Society for Political Methodology, Emory University.

—— **GLEDITSCH**, K. S. and **BEARDSLEY**, K. 2006. Space is more than geography: using spatial econometrics in the study of political economy. *International Studies Quarterly*, 50: 27–44.

—— and **KATZ**, J. N. 1995. What to do (and not to do) with time-series cross-section data. *American Political Science Review*, 89: 634–47.

—— ——2001. Throwing out the baby with the bath water: a comment on Green, Kim and Yoon. *International Organization*, 55: 487–95.

—— ——2004. Time series cross section issues: dynamics Presented at the Annual Meeting of the Society for Political Methodology, Stanford University.

(p. 492) **BECK**, N. and **KATZ**, J. N. 2007. Random coefficient models for time-series—cross-section data: Monte Carlo experiments. *Political Analysis*, 15: 182–95.

—— ——and **TUCKER**, R. 1998. Taking time seriously: time-series—cross-section analysis with a binary dependent variable. *American Journal of Political Science*, 42: 1260–88.

BOX-STEFFENSMEIER, J. M. and **JONES**, B. S. 2004. *Event History Modeling: A Guide for Political Scientists*. New York: Cambridge University Press.

CAMERON, A. C. and **TRIVEDI**, P. K. 1998. *Regression Analysis of Count Data*. New York: Cambridge University Press.

CAMERON, D. 1978. The expansion of the public economy: a comparative analysis. *American Political Science Review*, 72: 1243–61.

COX, D. R. 1975. Partial likelihood. *Biometrika*, 62: 269–76.

DAVIDSON, J. **HENDRY**, D. **SRBA**, F. and **YEO**, S. 1978. Econometric modelling of the aggregate time-series relationship between consumers' expenditures and income in the United Kingdom. *Economic Journal*, 88: 661–92.

ENGLE, R. and **GRANGER**, C. W. J. 1987. Co-integration and error correction: representation, estimation and testing. *Econometrica*, 55: 251–76.

GARRETT, G. 1998. *Partisan Politics in the Global Economy*. New York: Cambridge University Press.

GREEN, D. **KIM**, S. Y. and **YOON**, D. 2001. Dirty pool. *International Organization*, 55: 441–68.

HAMILTON, J. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

HENDRY, D. and MIZON, G. 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.

—— PAGAN, A. and SARGAN, J. D. 1984. Dynamic specification. **Ch. 18** in *Handbook of Econometrics*, vol. ii, ed. Z. Griliches and M. Intriligator. Amsterdam: North-Holland.

HSIAO, C. 2003. *Analysis of Panel Data*, 2nd edn. New York: Cambridge University Press.

HUBER, P. J. 1967. The behavior of maximum likelihood estimates under non-standard conditions. Pp. 221–33 in *Proceedings of the Fifth Annual Berkeley Symposium on Mathematical Statistics and Probability*, vol. i, ed. L. M. LeCam and J. Neyman. Berkeley: University of California Press.

IM, K. S. PESARAN, M. H. and SHIN, Y. 2003. Testing for unit roots in heterogeneous panels. *Journal of Econometrics*, 115: 53–74.

JONES, B. S. and BRANTON, R. P. 2005. Beyond logit and probit: Cox duration models of single, repeating and competing events for state policy adoption. *State Politics and Policy Quarterly*, 5: 420–43.

KING, G. HONAKER, J. JOSEPH, A. and SCHEVE, K. 2001. Analyzing incomplete political science data: an alternative algorithm for multiple imputation. *American Political Science Review*, 95: 49–69.

LANGE, P. and GARRETT, G. 1985. The politics of growth: strategic interaction and economic performance in the advanced industrial democracies, 1974–1980. *Journal of Politics*, 47: 792–827.

MAOZ, Z. and RUSSETT, B. M. 1993. Normative and structural causes of democratic peace, 1946–1986. *American Political Science Review*, 87: 639–56.

PARKS, R. 1967. Efficient estimation of a system of regression equations when disturbances are both serially and contemporaneously correlated. *Journal of the American Statistical Association*, 62: 500–9.

PINHEIRO, J. C. and BATES, D. M. 2000. *Mixed Effects Models in S and S-Plus*. New York: Springer.

(p. 493) POIRIER, D. J. and RUUD, P. A. 1988. Probit with dependent observations. *Review of Economic Studies*, 55: 593–614.

PRZEWORSKI, A. ALVAREZ, M. CHEIBUB, J. A. and LIMONGI, F. 2000. *Democracy and Development: Political Regimes and Economic Well-being in the World, 1950-1990*. Cambridge: Cambridge University Press.

SMITH, A. F. M. 1973. A general Bayesian linear model. *Journal of the Royal Statistical Society, Series B*, 35: 67–75.

STIMSON, J. 1985. Regression in space and time: a statistical essay. *American Journal of Political Science*, 29: 914–47.

STONE, M. 1974. Crossvalidatory choice and assessment of statistical prediction. *Journal of the Royal Statistical Society, Series B*, 36: 111–33.

WARE, J. H. LIPSITZ, S. and SPEIZER, F. E. 1988. Issues in the analysis of repeated categorical outcomes. *Statistics in Medicine*, 7: 95–107.

WESTERN, B. 1998. Causal heterogeneity in comparative research: a Bayesian hierarchical modelling approach. *American Journal of Political Science*, 42: 1233–59.

WHITE, H. 1982. Maximum likelihood estimation of misspecified models. *Econometrica*, 50: 1–25.

WOOLDRIDGE, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Mass.: MIT Press.

Notes:

(1) A typical study is Garrett's (1998) analysis of the role of political and labor market variables in determining economic policies and outcomes in fourteen OECD advanced industrial democracies observed annually from 1966 through 1990. Adolph, Butler, and Wilson (2005) provide an inclusive list along with some properties of each data-set analyzed.

(2) Hsiao (2003) is still the best general text for reading about these issues.

(3) For a longer discussion of some of these issues in the context of TSCS data, see Beck and Katz (2004).

(4) One could also have a dependent variable that is ordered or multichotomous. Very little is known about how to estimate such a model, although we shall see below that an event history approach allows researchers to study multichotomous dependent variables. While limited dependent variable panel models are a very active research topic

(Wooldridge 2002), there is little corresponding research for TSCS data. Researchers with limited dependent variable TSCS data should surely use Huber's (1967) robust (“sandwich”) standard errors, but that is only a band-aid. Beck et al. (2001) discuss some very complicated latent variable approaches, but these are new and unproven. I do not deal further with these complications here. It is also possible to analyze models where the dependent variable is an event count (Cameron and Trivedi 1998), and recent software innovations make estimation of TSCS models with a count dependent variable easier.

(5) Everything said here also holds for probit or other binary variable “link” functions.

(6) See Box-Steffensmeier and Jones (2004) for a general introduction to event history analysis. The rest of this section assumes some familiarity with event history methods and terminology.

(7) Beck, Katz, and Tucker (1998) consider some refinements of this procedure to enforce smoothness, but all the basics are captured with the dummy variable approach. This section is a condensation of that article which should be consulted for more details.

Nathaniel Beck

Nathaniel Beck is Professor and Chair of the Wilf Family Department of Political Science, New York University.