Modeling Dynamics in Time-Series–Cross-Section Political Economy Data

ANNUAL REVIEWS **Further**

Click here for quick links to Annual Reviews content online, including:

- Other articles in this volume
- Top cited articles
- Top downloaded articles
- Our comprehensive search

Nathaniel Beck[1] and Jonathan N. Katz[2]

[1]Department of Politics, New York University, New York, New York 10003; email: nathaniel.beck@nyu.edu

[2]Division of the Humanities and Social Sciences, California Institute of Technology, Pasadena, California 91125; email: jkatz@caltech.edu

Annu. Rev. Polit. Sci. 2011. 14:331–52

First published online as a Review in Advance on March 17, 2011

The *Annual Review of Political Science* is online at polisci.annualreviews.org

This article's doi: 10.1146/annurev-polisci-071510-103222

1094-2939/11/0615-0331$20.00

Keywords

lagged dependent variables, correlated errors, error correction model, non-stationarity, model specification

Abstract

This article deals with a variety of dynamic issues in the analysis of time-series–cross-section (TSCS) data. Although the issues raised are general, we focus on applications to comparative political economy, which frequently uses TSCS data. We begin with a discussion of specification and lay out the theoretical differences implied by the various types of dynamic models that can be estimated. It is shown that there is nothing pernicious in using a lagged dependent variable and that all dynamic models either implicitly or explicitly have such a variable; the differences between the models relate to assumptions about the speeds of adjustment of measured and unmeasured variables. When adjustment is quick, it is hard to differentiate between the various models; with slower speeds of adjustment, the various models make sufficiently different predictions that they can be tested against each other. As the speed of adjustment gets slower and slower, specification (and estimation) gets more and more tricky. We then turn to a discussion of estimation. It is noted that models with both a lagged dependent variable and serially correlated errors can easily be estimated; it is only ordinary least squares that is inconsistent in this situation. There is a brief discussion of lagged dependent variables combined with fixed effects and issues related to non-stationarity. We then show how our favored method of modeling dynamics combines nicely with methods for dealing with other TSCS issues, such as parameter heterogeneity and spatial dependence. We conclude with two examples.

1. INTRODUCTION

iid: independent identically distributed

Time-series–cross-section (TSCS) data are perhaps the most commonly used data in comparative politics (broadly defined) and particularly in comparative political economy.[1] There are various issues related to TSCS data; here we focus on some important ones related to the dynamic (time-series) properties of the data. Obviously many of these issues are similar to those for a single time series, but the context of comparative political economy and the relatively short lengths of the TSCS time periods make for some interesting special issues. We assume that the reader is familiar with the basic statistical issues of time-series data. Because various specification issues are covered for political scientists elsewhere (Beck 1985, 1991; Keele & Kelly 2006; De Boef & Keele 2008), we go fairly quickly over the basic issues, spending more time on issues relating to the dynamic modeling and interpretation of those models in political economy.

1.1. Notation and Nomenclature

Our interest is in modeling the dynamics of TSCS models. By dynamics we mean any process where some variable has an impact that is distributed over time. Let $y_{i,t}$ be an observation for unit i at time t where $i = 1, \ldots N$ and $t = 1, \ldots T$. We assume that y is measured as a continuous variable, or at least can be taken as continuous. In what follows, we typically do not care if we have one or more than one independent variable or variables, so let $x_{i,t}$ be either an observation on a single independent variable or a vector of such variables; if the latter, it is assumed that the dynamics apply similarly to all the components of that vector.

Where we need to differentiate dynamics, we use a second variable (or vector of variables), $z_{i,t}$. Because the constant term is typically irrelevant for what we do, we omit it from our notation (or include it in the vector of other independent variables). We assume that these independent variables are exogenous, which is clearly both a strong and critical assumption. We postpone discussing specifications for $y_{i,t}$ until the next section.

TSCS data are most commonly used in comparative political economy, and so we refer to time periods as years and units as countries. Given the yearly frequency, we focus on models where explicit lags are only of one or two years; we would not expect to see the higher-order dynamic processes common in standard time-series analysis of monthly or quarterly data. Although we focus on low-order processes, it is trivial to extend the interpretation, tests, and estimation to higher-order dynamics.

T must be large enough so that averaging over time makes sense. In our prior work (Beck & Katz 1995, 1996), we did not examine situations where $T < 15$. Political economy data often span a sample period of 30 or more years and so there are no problems. This article does not discuss standard survey panels that typically have five or fewer waves, and there is no reason to believe that the methods discussed here apply to such survey panels. We make no assumptions about N, although in comparative political economy it is seldom less than 20 (advanced industrial nations); it can be quite large (>100,000 for some studies in international relations using dyadic data).

We distinguish two types of error terms: $\nu_{i,t}$ refers to an independent identically distributed (iid) error process, whereas $\varepsilon_{i,t}$ refers to a generic error process that may or may not be iid. Unless specifically stated, we restrict non-iid processes to simple first-order autoregressions. This simplifies notation with no loss to our argument; it is simple to extend our argument to more complicated error processes. Because coefficients are only interpretable in

[1]Comparative politics refers to any comparison of political units, so it encompasses almost all of international relations, which has countries or country pairs as the unit of analysis, and any study that compares units (regions, states, or counties) within a single country. Our language comes from comparative political economy, but everything applies to all other types of TSCS studies (both within and beyond political science) so long as the data are observed over a long enough period. Adolph et al. (2005) report that by the early 2000s ~5% of all political science articles in JSTOR journals used TSCS data.

the context of a specific model, we superscript coefficients to indicate the specification they refer to whenever confusion might arise. We use the terms specification and model interchangeably, and refer to both as equations when referring to a specific equation in the text.

When relevant, we use L as the lag operator, so that

$$Ly_{i,t} = y_{i,t-1} \text{ if } t > 1 \qquad 1.$$

and $Ly_{i,1}$ is missing. The first difference operator is then $\Delta = 1 - L$.

1.2. Stationarity

We initially, and for most of the article, assume that the data are drawn from a stationary process. A univariate process is stationary if its various moments and cross-moments do not vary with the time period. In particular, the initial sections assume that the data are drawn from a "covariance stationary" process, that is,

$$\mathrm{E}(y_{i,t}) = \mu \qquad 2a.$$

$$\mathrm{Var}(y_{i,t}) = \sigma^2 \qquad 2b.$$

$$\mathrm{E}(y_{i,t}y_{i,t-k}) = \sigma_k \qquad 2c.$$

(and similarly for any other random variables).

Stationary processes are mean reverting, and the best long-run forecast for a stationary process is that mean. Thus, we can think of the mean as the "equilibrium" of a stationary process. Alternatively, we can think of the statistical properties of the data as not varying simply as a function of time (so, for example, there are no trends in the data). We briefly discuss nonstationary data in Section 4.

1.3. Missing Data

To simplify notation, we assume that the dataset is rectangular, that is, each country is observed for the same time period (which is called the "sample period" even though it is not a sample of anything). It is easy to extend everything to a world in which some countries are not observed for a few years at either the beginning or the end of the period under study, and the only cost of so doing would be an additional subscript in the notation.

Missing data in the interior of the period under study is not benign. At a minimum, their absence causes all the standard problems associated with missing data (Little & Rubin 1987). The default solution, list-wise deletion, is well known to be an inferior solution to the missing-data problem for cross-sectional data. But the problem is more severe for TSCS data because the specification invariably includes temporal lags of the data; even if the model has only first-order lags, each observation with missing data leads to the deletion of two data points. Thus, even more than with cross-sectional data, multiple imputation techniques are *de rigueur* for dynamic models that have more than a trivial amount of "missingness." Obviously the amount of missingness will vary as we move from studies of advanced industrial societies to studies of poorer nations, and so the attention paid to missingness can vary.

Although it is easy to say that analysts should use Little & Rubin's multiple imputations, the standard methods for cross-sectional imputations (hot decking or assuming that both missing and non-missing observations are essentially multivariate normal) are not appropriate for TSCS data. This is because we know a lot about TSCS data. Thus, for example, we would expect that missing economic variables are likely to be highly related to observed values in nearby countries with similar economies, or that observations on trending time series can be imputed with interpolated values. Honaker & King's (2010) Amelia II allows users this kind of flexibility. But there is no mechanistic solution to the missing-data problem in political economy. To give but one example, missingness is often related to civil wars; if we simply use some complicated averaging method to impute missing economic data during a civil war, our imputations are likely to be overly optimistic. Analysts using TSCS datasets with significant missing data can only be warned that they must take extreme care.

LDV: lagged dependent variable

FDL: finite distributed lag model

AR1: first-order autoregressive error process

SC: first-order serially correlated error model

1.4. Roadmap

The next section, on the interpretation of alternative dynamic specification, is the heart of the article. There we deal only with stationary data. The third section briefly examines combining dynamics with cross-sectional issues, in particular accounting for heterogeneity across units. The fourth section extends the argument to slowly moving and non-stationary data. Two examples of dynamic modeling with political economy TSCS data are discussed in the fifth section, and we offer some general conclusions in the final section.

2. DYNAMIC MODELS: STATIONARY DATA

There are various specifications for any time-series model; for reviews considering applications to political science, see Beck (1991) and De Boef & Keele (2008). All time-series specifications have identical counterparts in TSCS models. These specifications appear in any standard text, so we discuss general specification issues without either citation or claim of originality.

In our prior work (Beck & Katz 1996) we argued that a lagged dependent variable (LDV) specification is often adequate; because that has sparked some discussion (Achen 2000, Keele & Kelly 2006), we spend some time on this issue. After discussing a variety of specifications, we discuss issues of interpretation and estimation.[2]

2.1. Dynamic Specifications

The generic static (non-dynamic) specification is

$$y_{i,t} = \beta^s x_{i,t} + \nu_{i,t}. \qquad 3.$$

This specification is static because any changes in x or the errors are felt instantaneously and their effect dissipates instantaneously; there are no delayed effects. (It may be that $x_{i,t}$ is measured with a lag, so the effect could be felt with a lag, but the model is still inflexible in that the effect is completely and only felt at the one specified year.)

There are several ways to add dynamics to the static specification. The simplest is the finite distributed lag (FDL) model, which assumes that the impact of x sets in over two (or a few) periods but then dissipates completely. This specification has

$$y_{i,t} = \beta^{f1} x_{i,t} + \beta^{f2} x_{i,t-1} + \nu_{i,t} \qquad 4.$$

with the obvious generalization for higher-ordered lags. Equation 3 is nested inside Equation 4 (Equation 3 is a special case of Equation 4), so testing between the two is simple in principle (although the correlation of x and its lags makes for a number of practical issues).

Another commonly used dynamic specification is to assume that the errors follow a first-order autoregressive (AR1) process rather than the iid process of Equation 3. If we assume that the errors follow an AR1 process, we have a serially correlated (SC) error model:

$$y_{i,t} = \beta^{SC} x_{i,t} + \nu_{i,t} + \rho\varepsilon_{i,t-1} \qquad 5a.$$

$$= \beta^{SC} x_{i,t} + \frac{\nu_{i,t}}{1-\rho L} \qquad 5b.$$

$$= \beta^{SC} x_{i,t} + \rho y_{i,t-1} - \beta^{SC}\rho x_{i,t-1} + \nu_{i,t}. \qquad 5c.$$

The formulation in Equation 5c makes the dynamics implied by the model clearer and also makes it easier to compare various models.

Another alternative is the LDV model (with iid errors):

$$y_{i,t} = \beta^{LDV} x_{i,t} + \phi y_{i,t-1} + \nu_{i,t} \qquad 6a.$$

$$= \beta^{LDV} \frac{x_{i,t}}{1-\phi L} + \frac{\nu_{i,t}}{1-\phi L}. \qquad 6b.$$

As Equation 6b makes clear, the LDV model simply assumes that the effect of x decays geometrically (and for a vector of independent variables, all decay geometrically at the same rate). Note also that the compound error term is an infinite geometric sum (with the same decay parameter as for x); this error term is equivalent to

[2] We often refer to Achen's (2000) critiques of the use of LDVs. Although it is odd to spend time critiquing a decade-old unpublished paper, this paper has been influential (more than 300 Google Scholar citations as of this writing). We only deal with the portions of Achen's paper relevant to issues raised here.

a first-order moving average (MA1) error process, again with its decay parameter constrained to equal ϕ, the rate at which the effect of x on y decays.

Both the SC and LDV specifications are special cases of the autoregressive distributed lag (ADL) model,

$$y_{i,t} = \beta^{ADL} x_{i,t} + \theta y_{i,t-1} + \gamma x_{i,t-1} + \nu_{i,t}, \qquad 7.$$

where Equation 5c imposes the constraint that $\gamma = -\beta^{ADL}\theta$ and Equation 6a assumes that $\gamma = 0$. The nesting of both the LDV and SC specifications within the ADL specification allows for testing between the various models.

For interpretative purposes, it can be helpful to rewrite the ADL model in error correction (EC) form (Davidson et al. 1978). To do this, subtract $y_{i,t-1}$ from both sides of the ADL model to get a first difference in y on the left-hand side, and add and subtract $\beta^{ADL} x_{i,t-1}$ from the right-hand side to get a first difference of x in the specification. This leads to

$$\Delta y_{i,t} = \beta^{EC} \Delta x_{i,t} - \lambda(y_{i,t-1} - \kappa x_{i,t-1}) + \nu_{i,t}, \qquad 8.$$

which allows for the nice interpretation that short-run changes in y are a function of both short-run changes in x and how much x and y were out of equilibrium last year, where the equilibrium y and x are given by $y_{i,t} = \kappa x_{i,t}$ and the speed of equilibration (per year) is λ. The coefficients of the EC model can be easily derived from the corresponding ADL model: $\beta^{EC} = \beta^{ADL}$, $\lambda = \theta - 1$ and $\kappa = \frac{\gamma + \beta^{ADL}}{\theta - 1}$. For comparison with other models the ADL model works better, but for direct substantive interpretation of the coefficients the EC model is easier to work with (since one can directly read off the short-run impact of a change in x as well as various long-run impacts). Because the two are identical, either one can be estimated. We return to the EC model when we deal with nonstationary data in Section 4.

2.2. Interpretation

To see how the various specifications differ, we turn to unit and impulse response functions. Since x itself is stochastic, assume the process has run long enough for y to be at its equilibrium value (stationarity implies the existence of such an equilibrium). We can then think of a one-time shock in x (or ν) of one unit, with a subsequent return to equilibrium (zero for the error) the next year; if we then plot y against this, we get an impulse response function (IRF). Alternatively, we can shock x by one unit and let it stay at the new value; the plot of y against x is a unit response function (URF).

The static specification assumes that all variables have an instantaneous and only an instantaneous impact. Thus, the IRF for either x or ν is a spike, associated with an instantaneous change in y, and if x or ν then returns to its previous value in the next period, y immediately also returns to its previous value. The URF is simply a step function, with the height of the single step being β^s.

The FDL model generalizes this. The URF has two steps, of height β^{f1} and $\beta^{f1} + \beta^{f2}$, and the interval between the steps is one year. Thus, unlike in the simple static model, if x changes, it takes two years for the full effect of the change to be felt, but the effect is fully felt in those two years. For example, it may take one year for a party to have an impact on unemployment, but it may be the case that after that year the new party in office has done all it can and will do in terms of changing unemployment. Similarly, an institutional change may not have all of its impact immediately, but the full impact may occur within the space of a year.

We could add more lags to Equation 4, allowing for more complicated dynamics. But time series within a country are often temporally correlated, so multicolinearity makes it difficult to get good estimates of the coefficients of the FDL specification with many lags of x. Given the annual frequency of much of the data seen in comparative political economy, the problem of having to add too many lags to Equation 4 (say more than one additional lag) may not, in practice, be a problem. It may be unlikely that interesting institutional changes have only an immediate impact, but the FDL model *might* be appropriate. It surely should be borne in mind in thinking about appropriate

MA1: first-order moving average process

ADL: autoregressive distributed lag model

EC: error correction model

IRF: impulse response function

URF: unit response function

specifications, and, as we shall see, it combines nicely with some other specifications.

The SC model has a different IRF for x and the error. The IRF for x is a spike, identical to that of the static model; the IRF for the error has a declining geometric form with rate of decay ρ. It seems odd that all the omitted variables have a declining geometric IRF but the x we are modeling has only an instantaneous impact. Maybe that is correct, but this is not the first specification that would occur to us. The SC specification can be generalized by adding more lags of x, but we would still have very different dynamics for x and the unobserved variables in the "error" term. One should clearly have a reason to believe that dynamics are of this form before using the SC specification.

The LDV model has an IRF for both x and the error that has a declining geometric form; the initial response is β^{LDV} (or 1 for the error); this declines to zero geometrically at a rate ϕ. Although the effect never completely dissipates, it becomes tiny fairly quickly unless ϕ is almost one. The URF starts with an effect β^{LDV} immediately, increasing to $\frac{\beta^{LDV}}{1-\phi}$. If ϕ is close to one, the long-run impact of x can be 10 or more times the immediate impact.

Although the ADL specification appears to be much more flexible, it actually has an IRF similar to that of the LDV specification, other than in the first year (and is identical for a shock to the error process). Initially, y changes by β^{ADL} units; in the next period the change is $\beta^{ADL}\theta + \gamma$, which then dies out geometrically at a rate θ. Thus, the ADL specification is only a bit more general than the LDV specification. It does allow for the maximal impact of x to occur a year later, rather than instantaneously (or, more generally, the effect of x after one period is not constrained to be the immediate impact with one year's decay). This may be important in some applications. A comparison of the various IRFs and URFs is in **Figure 1**, which clearly shows that the difference between the specifications has simply to do with the timing of the adjustment to y after a change in x.

Before we get to slightly more complicated models, this analysis tells us several things. The various models differ in the assumptions they impose on the dynamics that govern how x and the errors impact y. None of the dynamic specifications can be more or less right *a priori*. Later, we discuss some estimation (see Section 2.5) and testing (see Section 2.6) issues, but for now we can say that various theories would suggest various specifications. The most important issue is whether we think a change in some variable is felt only immediately or whether its impact is distributed over time; in the latter case, do we think that a simple structure, such as a declining geometric form, is adequate? How would we expect an institutional change to affect some y of interest in terms of the timing of that effect? If only immediately or completely in one or two years, the SC or FDL model seems right; if we expect some initial effect that increases to some limit over time, the LDV or ADL model would be used. But there is nothing atheoretical about the use of a lagged dependent variable, and there is nothing that should lead anyone to think the use of a lagged dependent variable causes incorrect harm. It may cause "correct" harm, in that it may keep us from incorrectly concluding that x has a big effect when it does not, but that cannot be a bad thing. As has been well known, and as Hibbs (1974) showed three decades ago for political science, the correct modeling and estimation of time-series models often undoes seemingly obvious findings.

A related way to say the same thing is that each of the models (for stationary data) implies some long-run equilibrium and a speed with which that equilibrium is reached after some shock to the system. It is easy to solve for equilibria (if they exist) by noting that in equilibrium both x and y are stable. Let y_E and x_E refer to equilibrium x and y. Then for the ADL model it must be true that

$$y_E = \beta^{ADL} x_E + \theta y_E + \gamma x_E, \qquad 9.$$

yielding $y_E = \frac{(\beta^{ADL}+\gamma x_E}{1-\theta}$ ($|\theta| < 1$ by stationarity). This is easier to see in the EC form, where $y_{i,t-1} = \kappa x_{i,t-1}$ in equilibrium and λ is the rate (per year) at which y returns to this equilibrium. All the models for stationary data imply both a long-run equilibrium and a speed of

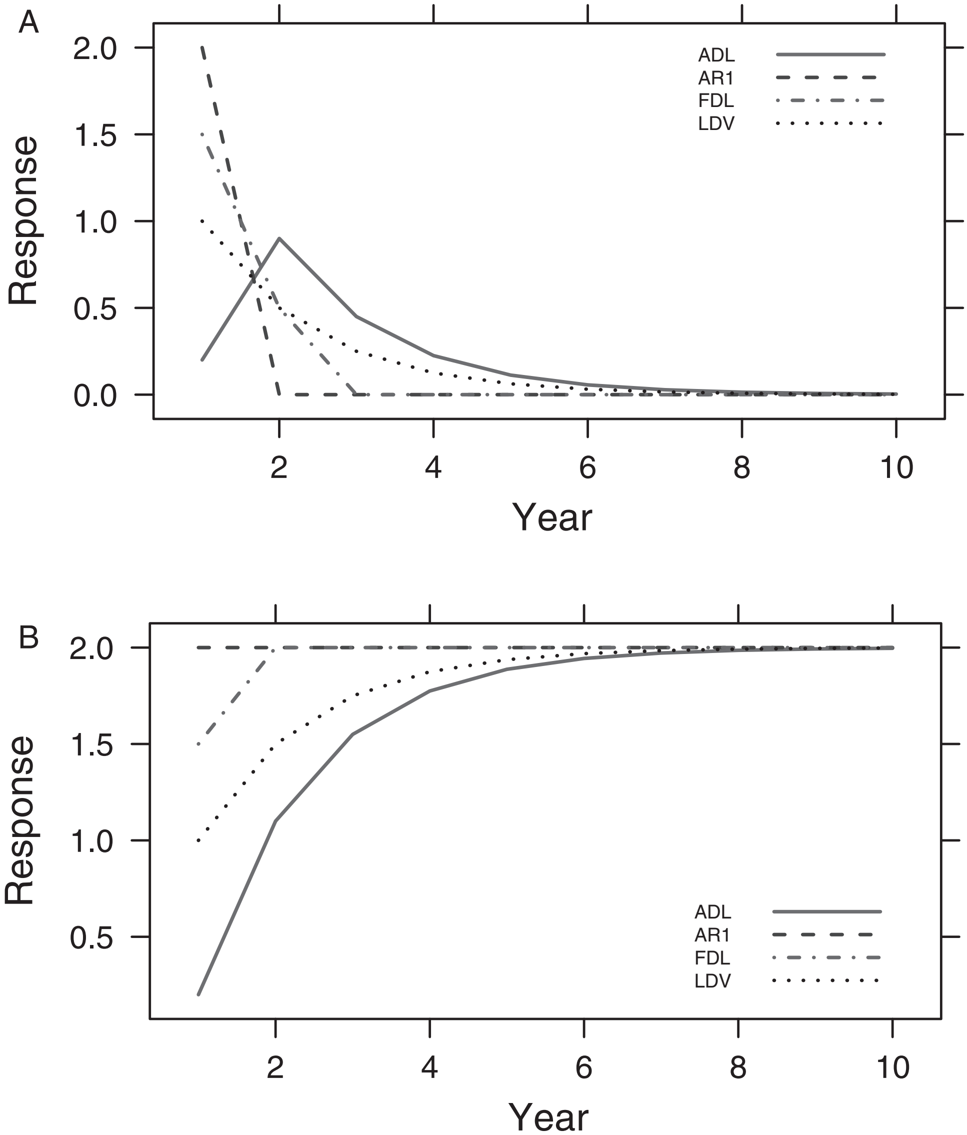

Figure 1

Comparison of impulse and unit response functions for four specifications: autoregressive distributed lag (ADL, $y_{i,t} = 0.2x_{i,t} + 0.5y_{i,t-1} + 0.8x_{i,t-1}$), finite distributed lag (FDL, $y_{i,t} = 1.5x_{i,t} + 0.5x_{i,t-1}$), lagged dependent variable (LDV, $y_{i,t} = 1.0x_{i,t} + 0.5y_{i,t-1}$), and autoregressive errors (Static, $y_{i,t} = 2x_{i,t}$).

equilibration, with the different parameter constraints determining these long-run features. Each of these models implies different short- and long-run reactions of y to x, and standard econometric methods (see Section 2.6) can be used to discriminate between them.

2.3. Higher-Order Processes and Other Complications

We can generalize any of these models to allow for non-iid errors and higher-order dynamics. However, because our applications typically use annual data, it is often the case that first-order error processes suffice, and it would be unusual to have more than second-order processes. As we shall see, it is easy to test for higher-order error processes, so there is no reason to simply assume that errors are iid or only follow a first-order process. For notational simplicity, we restrict ourselves to second-order processes, but the generalization is obvious.

Consider the LDV model with AR1 errors, in which

$$y_{i,t} = \beta^{LDV} x_{i,t} + \phi y_{i,t-1} + \frac{\nu_{i,t}}{1-\omega L}. \qquad 10.$$

After multiplying through by $(1-\omega L)$, we get a model with two lags of y, x and lagged x and some constraints on the parameters; if we generalize the ADL model similarly, we get a model with two lags of both y and x and more constraints. The interpretation of this model is similar to the model with iid errors.

We have already seen that the LDV model with iid errors is equivalent to a model where the effect of all the independent variables and the error decline at the same geometric rate. But if we assume that the "errors," that is, omitted or unmeasured variables, follow an MA1 process with the same decay rate, ϕ, as for the measured variables (which may or may not be a good assumption), then we have

$$y_{i,t} = \beta^{LDV} x_{i,t} + \phi y_{i,t-1} + (1-\phi L)\nu_{i,t}, \qquad 11a.$$

which simplifies to

$$y_{i,t} = \beta^{LDV} \frac{x_{i,t}}{1-\phi L} + \nu_{i,t}. \qquad 11b.$$

That is, we have a model that combines a geometrically declining impact of x on y with iid errors. It is surely more likely that the "errors" are correlated than that they are independent. Of course, the most likely case is that the errors are neither iid nor MA1 with the same dynamics as x, so we should entertain a more general specification, where the effects of both measured and unmeasured variables have a declining geometric impact with different rates of decline. The simplest such specification is Equation 10. We return to this more general specification in Section 2.5.

2.4. More Complicated Dynamics—Multiple Independent Variables

We typically have more than one independent variable. How much dynamic generality can or should be allowed for? One easy generalization is to allow for two independent (or sets of independent) variables, x and z. Allowing also for a separate speed of adjustment for the errors yields

$$y_{i,t} = \beta \frac{x_{i,t}}{1-\phi_x L} + \gamma \frac{z_{i,t}}{1-\phi_z L} + \frac{\nu_{i,t}}{1-\phi_\nu L}. \qquad 12.$$

Obviously each new variable now requires us to estimate two additional parameters. Also, on multiplying out the lag structures, we see that with three separate speeds of adjustment we have a third-order lag polynomial multiplying y, which means that we will have the first three lags of y on the right-hand side of the specification (and two lags of both x and z) and a second-order moving average error process. Although there are many constraints on the parameters of this model, the need for three lags of y costs us three years' worth of observations (assuming the original dataset contained as many observations as were available). With k independent variables, we would lose $k+1$ years of data; for a typical problem, where T is perhaps 30 and k is perhaps 5, this is nontrivial. Thus, we are unlikely to ever be able to (or want to) estimate a model where each variable has its own speed of adjustment.

But we might get some leverage by allowing for two kinds of independent variables: those where adjustment (speed of return to equilibrium) is relatively fast (x) and those where the system returns to equilibrium much more slowly. To simplify, assume the error process shows the same slower adjustment speed as z; we can obviously build more complex models, but they bring nothing additional to this discussion. We then would have

$$y_{i,t} = \beta_x x_{i,t} + \beta_z \frac{z_{i,t}}{1-\phi L} + \frac{\nu_{i,t}}{1-\phi L} \qquad 13a.$$

$$= \beta_x x_{i,t} - \phi\beta_x x_{i,t-1} + \beta_z z_{i,t} + \phi y_{i,t-1} + \nu_{i,t}. \qquad 13b.$$

Thus, at the cost of one extra parameter, we can allow some variables to have only an immediate or very quick effect, while others have a slower effect, with that effect setting in geometrically. With enough years we could estimate more complex models, allowing for multiple dynamic processes, but such an opportunity is unlikely to present itself in studies of comparative political economy. We could also generalize the model by allowing for the lags of x and z to enter without constraint. It is possible to test for whether these various complications are supported by the data, or whether they simply ask too much of the data. As always, it is easy enough to test and then make a decision.

2.5. Estimation Issues

As is well known, a specification with no lagged dependent variable but serially correlated errors is easy to estimate using any of several variants of feasible generalized least squares, with the Cochrane-Orcutt iterated procedure being the most well known. It is also easy to estimate such a model via maximum likelihood, breaking up the full likelihood into a product of conditional likelihoods.

The LDV model with iid errors is optimally estimated by ordinary least squares regression (OLS). However, it is also well known that OLS yields inconsistent estimates of the LDV model if the error process is serially correlated. Perhaps less well known is that Cochrane-Orcutt or maximum likelihood provides consistent estimates of the LDV model with serially correlated errors by accounting for that serial correlation (Hamilton 1994, p. 226). Thus, it is easy to correctly estimate the LDV model while allowing for serially correlated errors if analysts wish to do so. But we hope that analysts will not wish to do so.

It is often the case that the inclusion of a lagged dependent variable eliminates almost all serial correlation of the errors. To see this, start with the SC equation:

$$y_{i,t} = \beta^{SC} x_{i,t} + \varepsilon_{i,t} \qquad \text{14a.}$$

$$\varepsilon_{i,t} = \nu_{i,t} + \rho\varepsilon_{i,t-1}. \qquad \text{14b.}$$

Remember that the error term is simply all the omitted variables, that is, everything that determines y that is not explained by x. If we adjoin $y_{i,t-1}$ to the specification, the error in that new specification is $\varepsilon_{i,t} - \phi y_{i,t-1}$, where $\varepsilon_{i,t}$ is the original error in Equation 14a, not some generic error term. Since the $\varepsilon_{i,t}$ are serially correlated because they contain a common omitted variable, and $y_{i,t-1}$ contains the omitted variables at time $t-1$, including $y_{i,t-1}$ will almost certainly lower the degree of serial correlation, and often will eliminate it. But there is no reason to simply hope for this; we can estimate (using OLS) the LDV model assuming iid errors, and then test the null hypothesis that the errors are independent using a Lagrange multiplier test. (This only requires that OLS be consistent under the null of iid errors, which it is.) The test is trivial to implement, by simply regressing the residuals from the OLS regression on the lag(s) of this residual and all the independent variables including the lagged dependent variable with NTR^2 from this regression, being asymptotically distributed χ^2 with degrees of freedom equal to the number of lags tested. If, as often happens, we do not reject the null that the remaining errors are iid, we can continue with the OLS estimates. If we do reject that null, we should estimate a more complicated model.

Obviously, failing to reject the null of no serial correlation of the errors is not the same thing as knowing there is no serial correlation of the errors. Is this incorrect logic in interpreting a failure to reject the null likely to cause problems? There are two reasons to be sanguine here. First, the large amount of data in typical TSCS studies gives the Lagrange multiplier test good power. In our first example (Section 5.1), with ~300 total observations, the Lagrange multiplier test detected a serial correlation of the errors of ~0.10. It is also the case that ignoring a small amount of serial correlation (that is, estimating the LDV model with OLS as if there were no serial correlation) leads to only small amounts of bias. As Achen (2000, p. 13) elegantly shows, the estimation bias in incorrectly using OLS to estimate the LDV model with serially correlated errors is directly

ADLLDV2 specification: the autoregressive distributed lag specification with a second lag of the dependent variable

proportional to that serial correlation. Applied researchers make many assumptions to simplify analysis, assumptions that are never exactly correct. Ignoring a small serial correlation of the errors is surely one of the more benign mistakes.

As we shall see in Section 3, a number of fruitful avenues of investigation are open if the errors are either uncorrelated or sufficiently uncorrelated that we can ignore their small correlation. But what if a researcher is not so sanguine? As we have seen in Section 2.5, one can easily estimate, using methods other than OLS, the ADL model with serially correlated errors. But a more fruitful approach, as shown in Section 2.3, is to include second-order lags of the variables in the ADL specification; this "ADL2" specification can be appropriately estimated by OLS, once again allowing the researcher to more easily examine other interesting features of the data. Of course the same Lagrange multiplier testing procedure should first be used to test for remaining serial correlation, but with annual data we can be hopeful that we will not need highly complicated lag structures in the preferred specification.

Obviously more parsimonious specifications are easier to interpret (and convey to the reader), and so more complicated specifications with higher-order lags come at a cost. Thus, we might want to consider models intermediate between the ADL and ADL2 models. One obvious choice is to simply append a second lag of the dependent variable to the ADL specification; this is analogous to moving from the static to the LDV specifications, as discussed above. This simpler specification, ADLLDV2, should be tested to see if the errors are iid. The ADLLDV2 specification may be a good compromise between parsimony and fidelity to important features of the data; in our first example this is our preferred model. In other cases even simpler models may provide a better tradeoff between the various goals.

2.6. Discriminating Between Models

We can use the fact that the ADL model nests the LDV and SC models to test which specification better fits the data. The LDV model assumes $\gamma = 0$ (in Equation 7), whereas the SC model assumes $\gamma = -\theta\beta^{ADL}$. Thus, we can estimate the full ADL model and test whether $\gamma = 0$ or $\gamma = -\theta\beta^{ADL}$.[3] If both simplifications are rejected, we can retain the more complicated ADL model.[4] Even in the absence of a precise test, the ADL estimates will often indicate which simplification is not too costly to impose.

For fast dynamics (where θ is close to zero), it will be hard to distinguish between the LDV and SC specifications—or, alternatively, it does not make much difference which specification we use. To see this, note that if the SC model is correct, but we estimate the LDV model, we are incorrectly omitting the lagged x variable although it should be in the specification, but with a constrained coefficient $\theta\beta$. As θ goes to zero, the bias from failing to include this term goes to zero. Similarly, if we incorrectly estimate the SC model when the LDV model is correct, we have incorrectly included in the specification the lagged x variable, with coefficient $-\theta\beta$. Again, as θ goes to zero, this goes to zero. Thus, we might find ourselves not rejecting either the LDV or SC specifications in favor of the more general specification, but for small θ it matters little. As θ grows larger the two models diverge, and so we have a better chance of discriminating between the specifications.

This view is different from the conventional wisdom on omitted-variable bias. It is normally thought to be worse to incorrectly exclude than to incorrectly include a variable. This

[3]The first test is an ordinary t-test. The second is easiest via a linear approximation to the nonlinear constraint using a Taylor series (Greene 2008, pp. 96–98); this test is implemented in some common statistical packages such as Stata.

[4]Starting with the ADL model and then testing whether simplifications are consistent with the data is part of the idea of general-to-simple testing (also called the encompassing approach) espoused by Hendry and his colleagues (Hendry & Mizon 1978, Mizon 1984). This approach could start with a more complicated model with higher-order specifications, but given annual data, the ADL model with no more than two lags is often the most complicated specification that need be considered.

difference is because both models constrain the coefficient of the lagged x, and so the SC model "forces" the lagged x to be in the specification. But if we start with the ADL model and then test for whether simplifications are consistent with the data, we will not be misled. This testing of simplifications is easy to extend to more complicated models, such as Equation 13b.

3. COMBINING DYNAMIC AND CROSS-SECTIONAL ISSUES

Modeling dynamics with TSCS data is only half the job; clearly analysts also need to model the cross-sectional properties of the data. We have discussed various cross-sectional issues for TSCS data elsewhere (Beck & Katz 1995, Beck 2001). Here we discuss some issues that relate to the interaction of modeling dynamics and cross-sectional issues. For reasons of space, we omit discussion of dynamics with discrete dependent variables (see Beck et al. 1998). Dynamics are no less important in models with discrete dependent variables, but the recommended modeling is different for that situation.

3.1. Independent Errors Simplify Cross-Sectional Modeling

We have advocated modeling dynamics by including appropriate current and lagged values of the x's and lagged values of the dependent variable so that the resulting errors appear to be serially independent, enabling easy interpretation and estimation. This approach makes it much simpler to model cross-sectional situations. Most standard programs that allow for modeling complicated cross-sectional situations do not allow for temporally correlated errors. Although this is a practical rather than a theoretical issue, some estimation methods are sufficiently complex that one really wants to use a "canned" program (see Sidebar "Why Not Just Correct the Standard Errors?").

In particular, realistic political economy models often should allow for spatial effects, that is, they should recognize that variables in one country impact other countries. Models of the political causes of economic performance, for example, must take into account that the economic performance of any country is a function of the economic performance of its trading partners. These issues have been discussed in the context of TSCS data elsewhere (Beck et al. 2006, Franzese & Hayes 2007), and here we simply point out that our preferred approach to dynamics makes it easy for analysts to deal with this critical cross-sectional issue.

Another cross-sectional feature that should be considered (see Beck & Katz 2007 and the references cited there) is that parameters may vary randomly by country, and possibly as a function of country-level covariates. It is easy to allow for this by using the "random coefficients model" (which is equivalent to a "mixed" or "hierarchical" or "multilevel" model) if the error process is iid. Note that one of the randomly varying parameters can be that of the lagged dependent variable, the parameter that controls the speed of adjustment in the model. Perhaps countries differ in that speed of adjustment. As we see in Section 5.1, this issue is easy to examine when errors are iid.

WHY NOT JUST CORRECT THE STANDARD ERRORS?

Most standard programs that allow for modeling complicated cross-sectional situations do not allow for temporally correlated errors. Some researchers try to solve this problem by using simple models and then correcting the standard errors using some variant of Huber's (1967) "robust" standard errors. This is the reasoning behind our recommendation to use PCSEs (panel corrected standard errors) to deal with some difficult cross-sectional complications of the error process. There are similar autocorrelation consistent standard errors (Newey & West 1987). We do not recommend these because failing to account for serially correlated errors often leads to substantial inefficiencies in estimation as well as incorrect standard errors; failing to account for cross-sectional problems in the data is usually less serious. In any event, users of our preferred methods have no need to resort to autocorrelation consistent standard errors.

3.2. Fixed Effects and Lagged Dependent Variables

Perhaps the most common cross-sectional issue is heterogeneity of the intercepts. In the TSCS context, this is usually dealt with by adding "fixed effects" (country-specific intercepts) to the specification. We would adjoin these country-specific intercepts to the preferred ADL specification. But here we get into potential trouble, since it is well known that autoregressive models with fixed effects lead to biased parameter estimates (Nickell 1981). This bias is induced because centering all variables by country, which eliminates the heterogeneity of the constant term, induces a correlation between the centered lagged dependent variable and the centered error term.

It is also well known that this bias is of order $\frac{1}{T}$, and almost all of the work on this problem has been in the context of small-T "panels." When T is 2 or 3, the bias is indeed severe (50% or so). But when T is 20 or more, the bias becomes small.

Various corrections for this bias are well known. Most of them involve the use of instrumental variables, building on the work of Anderson & Hsiao (1982). As is often the case, it is hard to find good instruments, and so the instrumental variable corrections often obtain consistency at the price of rather poor finite sample properties. Other estimators (Kiviet 1995) are hard to combine with other methods and hard to generalize to even non-rectangular data sets.

We ran Monte Carlo experiments to compare OLS estimation of a simple LDV model with fixed effects to the Kiviet and Anderson-Hsiao estimators (Beck & Katz 2009). For the T's seen typically in TSCS analysis (20 or more), OLS performs about as well as Kiviet and much better than Anderson-Hsiao. Given the advantages of the OLS method discussed in the previous subsection, we do not hesitate to recommend OLS when country-specific intercepts must be adjoined to the specification of a TSCS model. Judson & Owen (1999) give similar advice following a similar discussion of this issue.

4. NON-STATIONARITY IN POLITICAL ECONOMY TSCS DATA

Before we look at some examples, one topic remains: what to do with non-stationary data. During the past two decades, with the pioneering work of Engle & Granger (1987), time-series econometrics has been dominated by the study of non-stationary series. There are many ways to violate the assumptions of stationarity presented in Equation 2, but most of the work has focused on the issue of unit roots or integrated series in which shocks to the series accumulate forever. These series are long-memoried; even distant shocks persist to the present. The key question is how to estimate models where the data are integrated (we restrict ourselves to integration of order one with no loss of generality). Such data, denoted I(1), are not stationary but their first difference is stationary. The simplest example of such an I(1) process is a random walk, where

$$y_{i,t} = y_{i,t-1} + \nu_{i,t}, \qquad 15.$$

with $\nu_{i,t}$ being stationary by definition. Integrated data look very different from data generated by a stationary process. Most importantly, they do not have equilibria (because there is no mean reversion), and the best prediction of an integrated series many periods ahead is the current value of that series.

There is a huge literature on estimating models with integrated data. Such methods must take into account that standard asymptotic theory does not apply, and also that

$$\lim_{t\to\infty} \mathrm{Var}(y_{i,t}) = \infty. \qquad 16.$$

Thus, if we wait long enough, any integrated series will wander "infinitely" far from its mean. Much work on both diagnosing and estimating models with integrated series builds on both these issues. Our interest is not in the estimation of single time series but rather TSCS political economy data.[5]

Political economy data is typically observed annually for relatively short periods of time (often 20–40 years). Of most relevance, during that time period we often observe very few cycles. Thus, although the series may be very persistent, we have no idea if a longer time period would show the series to be stationary (though with a slow speed of adjustment) or non-stationary. These annual observations on, for example, GDP or left political control of the economy are very different from the daily observations we may have on financial rates. So although it may appear from an autoregression that some political economy series have unit roots, is this the right characterization of these series? For example, using Huber & Stephens' (2001) data, an autoregression of social security on its lag yields a point estimate of the autoregressive coefficient of 1.003 with a standard error of 0.009; a similar autoregression of Christian Democratic party cabinet participation yields 1.03 with a standard error of 0.001. It does not take heavy-duty statistical testing to see we cannot reject the null (that the autoregressive coefficient is one) in favor of the alternative (that it is less than one). But does this mean that we think the series might be I(1)?

If these series had unit roots, they would tend to wander far from their means, and the variance of the observations would grow larger and larger over time (a similar argument is made by Alvarez & Katz 2000). But by definition both the proportion of the budget spent on social security and Christian Democratic cabinet participation are bounded between 0% and 100%, which then bounds how large their variances can become. Further, if either series were I(1), then we would be equally likely to see an increase or decrease in either variable regardless of its present value. Do we really believe that there is no tendency for social security spending to be more likely to rise when it is low and to fall when high, or for Christian Democratic cabinet strength to exhibit a similar tendency? In the Huber & Stephens data, social security spending ranges only between 3% and 33% of the budget, and Christian Democratic cabinet strength ranges between 0% and 34%. Even though these series are very persistent, they simply cannot be I(1). The impressive apparatus built over the past two decades to estimate models with I(1) series does not provide the tools needed for many, if not most, political economy TSCS datasets.

One possibility is to induce stationarity by first differencing all slowly changing variables, leading to a model that explains changes in y by changes in x. In practice, first-difference models often perform poorly (at least from the perspective of the researcher, for whom changes in x appear unrelated to changes in y). Modeling first differences also throws out any long-run information about y and x, so the effect of a change in x is the same regardless of whether y is high or low by historical standards.

Fortunately, the modeling issue is not really about the univariate properties of any series but the properties of the stochastic process that generated the y's conditional on the observed covariates. Even with data similar to Huber & Stephens', the errors may appear stationary and so the methods of the previous section can be used. In particular, whether the series are integrated or stationary but slowly moving, they may be well modeled by the EC specification (Equation 8), which, as we have seen, is just an alternative parameterization of the ADL model. The EC form is nice because it combines the short-run first-differences model with the long-run tendency for series to be in equilibrium. If the estimated λ in the EC specification is zero, that indicates that y and x have no long-run equilibrium relationship. We have already seen that if x and y are stationary, they always have a long-run relationship, so this is only a problem if the series are integrated. In other words, if the series are stationary but adjust very slowly, the EC (or equivalent ADL) model is a good place to start, and if the series are integrated, either the EC model will work

[5]There is a literature on panel unit roots (Im et al. 2003, Levin et al. 2002), but at this point the literature is still largely about testing for unit roots.

(the series are said to be co-integrated) or the residuals will appear highly correlated. Because our preferred methodology chooses specifications with almost uncorrelated residuals, it should never lead to choosing an incorrect EC (or ADL) specification.

Why do we propose ignoring much of what has dominated econometric argument for two decades? First, economists study many series (such as interest or exchange rates) that inherently are in levels, and so are likely to be integrated; variables in political economy are often expressed as a proportion of GDP or the government budget and hence are much less likely to be integrated. Other political variables, such as party control of government, may be persistent, but cannot possibly be integrated (they take on values of zero and one only, and so have neither infinite variance nor no tendency to revert back toward the mean). A second difference is that economists have no theory about whether one short-run exchange rate adjusts to a second rate, or the second rate adjusts to the first, or both; this leads to complicated estimation issues. In many political economy models, it is clear that y adjusts to x but not vice versa. We think that left governments increase spending but we do not think that low spending leads directly to a right-wing government (Beck 1992). Thus, even with highly persistent data, the EC (or ADL) model, estimated by OLS, will quite often work well, and, when it fails, simple tests and a rigorous methodology will indicate that failure.[6]

5. EXAMPLES

In this section we consider two examples to explore the practical issues in estimating dynamics in political economy TSCS datasets. The first example, presented in some detail, looks at the political determinants of capital taxation rates where adjustment speeds are fairly slow. The second example, presented more cursorily, looks at the impact of political variables on the growth of GDP. In the GDP example, where the dynamics are quite fast, the specification choice has fewer consequences. All computations were done using Stata 11.1, with data kindly provided by Geoff Garrett. Although our analysis is different from those of Garrett & Mitchell, we began by easily replicating their results.

5.1. Capital Taxation Rates

Our first example models capital taxation rates in 16 OECD nations from 1961 through 1993 using the data and specification of Garrett & Mitchell (2001).[7] Obviously tax rates move relatively slowly over time; the autoregressive coefficient of tax rates is 0.77. Thus, although tax rates are clearly stationary, it will take some number of years for the system to get close to fully adjusting; it takes about 2.65 years for any shock to dissipate.

Before estimation, one should examine the data to see whether there is sufficient within-country heterogeneity to make TSCS analysis meaningful, to see whether there appears to be very much inter-country heterogeneity that might need to be modeled, and to see whether there are any temporal pattens, such as trends, that need to be modeled. For the first two issues a standard country-specific box plot of tax rates, shown in **Figure 2**, is appropriate; for the third question time-series plots by country, shown in **Figure 3**, are more useful.

[6] There is a slight technical problem in that the distribution of the estimated λ is not normal if the series are not co-integrated. Instead, they have a Dickey-Fuller type distribution, which has fatter tails. Thus, there may be some cases where a standard test of the null hypothesis that $\lambda = 0$ yields incorrect conclusions. But given the large n and T of TSCS data, in many cases it is clear that the EC model is adequate or not, and if we incorrectly assume stationarity, consistent application of appropriate standard methods will indicate the problem.

[7] The data set is not rectangular; some countries only report tax rates for a portion of the period under study. In total there were 322 observations after one drops missing data at the beginning of a period (and omits observations with missing lagged data so that all results pertain to the same data). The extra lag in the ADLLDV2 model leads to the loss of the first observation for each country, yielding 306 observations for that estimation. This loss of data points is an additional reason why analysts may not prefer this specification.

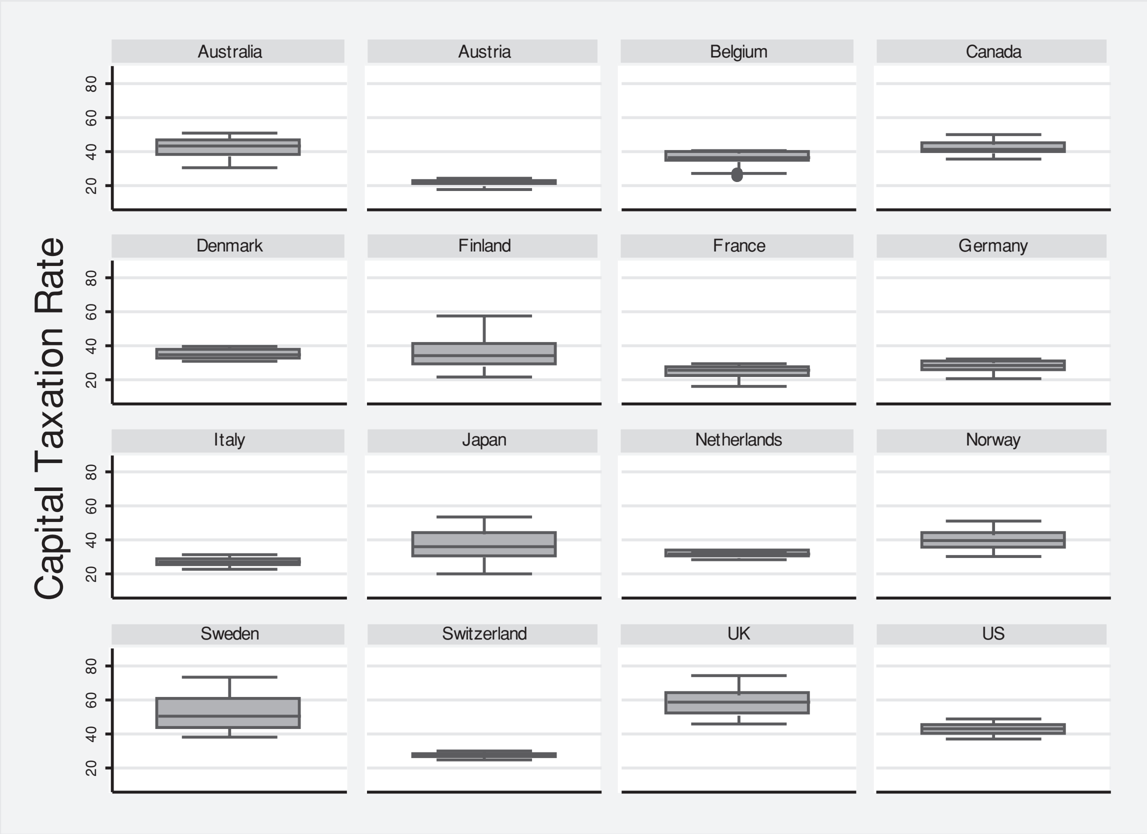

Figure 2

Box plots of capital taxation rates by country, 1967–1992.

Whereas some countries (Austria, Denmark, the Netherlands, and New Zealand) show little if any variation in their tax rates, the other countries show enough variation over time to make a TSCS analysis of interest. There is also some inter-country heterogeneity, with France, Sweden, and the United Kingdom having generally higher rates. **Figure 3** shows that taxation rates in some countries are strongly trending whereas others show little trend; this figure also clearly shows the beginning of period missingness pattern in the data. A regression of tax rates on time shows a trend of ~0.33% (with a small standard error) per annum in those rates. Thus, it appears as though a TSCS analysis of these data is sensible, and it may be the case that there will be unexplained temporal and cross-sectional heterogeneity. Following Garrett & Mitchell, we mean-centered all observations by country and year, which is equivalent to allowing for year- and country-specific intercepts.[8]

[8]Of course one can only decide whether these year- and country-specific intercepts are needed after a specification is chosen, and because these intercepts are atheoretical, one should attempt to find specifications where they are not necessary. Alas, this is often impossible. Here the intercepts were significant in all specifications. We might have preferred a model with a time trend instead of year-specific intercepts, but the difference between the two specifications was negligible, and we preferred to stay consistent with Garrett & Mitchell. Obviously in actual research such decisions should be made with care, and researchers should not simply do what others have done before.

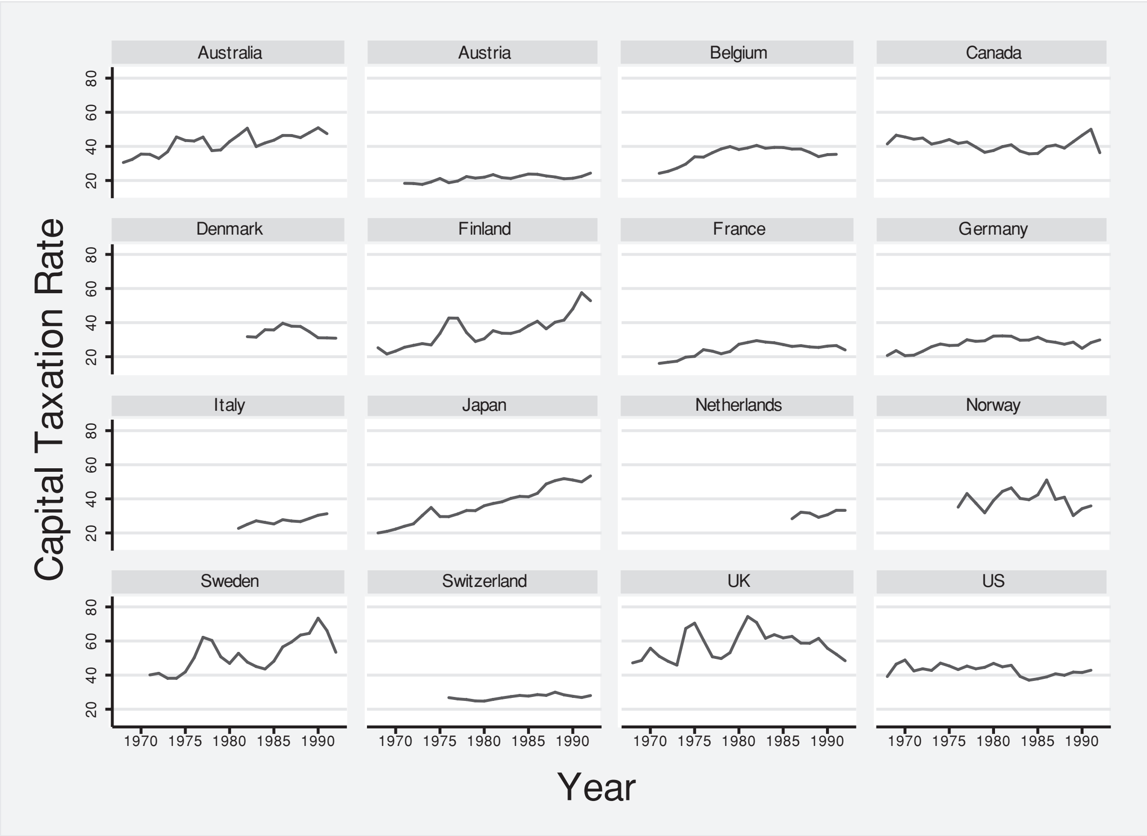

Figure 3

Time-series plots of capital taxation rates by country.

Garrett & Mitchell (2001) wish to explain capital taxation rates (this is only one of their analyses) by variables related to the economy, the demand for services, and political factors. We work more or less with the Garrett & Mitchell specification, dropping a few variables that were not substantively interesting nor statistically significant in any specification. We thus regress the capital tax rate (*CAPTAX*) on unemployment (*UNEM*), economic growth (*GDPPC*), the proportion of the population that is elderly (*AGED*), vulnerability of the workforce as measured by low wage imports (*VULN*), foreign direct investment (*FDI*), and two political variables: the proportion of the cabinet portfolios held by the left (*LEFT*) and the proportion held by Christian Democrats (*CDEM*). Because we mean-centered all variables, there are no intercepts in the model. **Table 1** reports the results of the various dynamic estimations. All standard errors are our recommended panel-corrected standard errors (Beck & Katz 1995), which are easy to compute with our recommended methodology.

The static model (not shown) is clearly inadequate; a Lagrange multiplier test for serial correlation of the errors strongly rejects the null hypothesis of serially independent errors. Because the static model is nested inside both the LDV and SC models, standard Wald tests (a t-test of either H_0:$\rho = 0$ or H_0: $\beta_{TAX_L} = 0$) clearly show that the static model can be rejected in favor of either of these two models.

Table 1 Comparison of SC, LDV, ADL, and ADLLDV2 estimates of Garrett & Mitchell's (2001) model of capital taxation in 16 OECD nations, 1967–1992 (country and year centered)

	SC		LDV		ADL		ADLLDV2	
Variable	$\hat{\beta}$	PCSE	$\hat{\beta}$	PCSE	$\hat{\beta}$	PCSE	$\hat{\beta}$	PCSE
VULN	−0.22	0.12	−0.10	0.07	−0.28	0.13	−0.33	0.14
FDI	0.51	0.26	0.37	0.21	0.59	0.26	0.48	0.28
UNEM	−0.18	0.22	−0.34	0.14	−0.68	0.27	−0.68	0.30
AGED	1.42	0.51	0.35	0.24	0.26	0.71	−0.27	0.87
GDPPC	−0.69	0.11	−0.62	0.12	−0.80	0.13	−0.81	0.14
LEFT	0.004	0.012	0.006	0.009	0.003	0.013	0.002	0.014
CDEM	0.018	0.022	0.015	0.012	0.015	0.025	0.031	0.024
TAX_L			0.70	0.06	0.76	0.07	0.93	0.10
$VULN_L$					0.21	0.14	0.24	0.15
FDI_L					−0.55	0.29	−0.56	0.31
$UNEM_L$					0.48	0.26	0.62	0.28
$AGED_L$					0.24	0.76	0.98	0.94
$GDPPC_L$					0.29	0.12	0.36	0.14
$LEFT_L$					0.005	0.013	0.004	0.014
$CDEM_L$					0.005	0.024	−0.010	0.025
TAX_{L2}							−0.26	0.09
ρ	0.66							

Abbreviations: ADL, autoregressive distributed lag; ADLLDV2, the ADL specification with a second lag of the dependent variable; LDV, lagged dependent variable; PCSE, panel corrected standard errors; SC, first-order serially correlated error model. See text for definitions of variables in the left-hand column.

But we must compare both the LDV and SC specifications to the more general ADL specification. Again, since both these specifications are nested inside the ADL specification we can use standard Wald tests (in this case an *F*-test of the null hypothesis that the coefficients on all the lagged *x*'s are zero); that null is decisively rejected, so the more general ADL specification is preferred.

The ADL specification still shows serial correlation of the errors; a Lagrange multiplier test of the null hypothesis of independent errors shows we can reject that null of iid errors. (A regression of the residuals from the ADL specification on the lagged residuals and all the other independent variables has an R^2 of 0.046, which, multiplied by the number of observations in that regression, yields a statistic of 14; since this statistic is distributed χ^2_1, the null hypothesis of independent errors is clearly rejected.) As discussed in Section 2.5, we added a second lag of capital taxation to the specification; results of estimating this specification are in the ADLLDV2 columns. We cannot reject the null hypothesis of independent errors for this regression ($\chi^2_1 = 1.1$). The ADLLDV2 specification is both statistically superior to the simpler specifications and shows iid errors. There are, of course, many other specifications that a substantive article would test (multiple speeds of adjustment, for example), but we do not pursue these here.

All the models show that a one-time unit shock to the error process dies out exponentially (or nearly exponentially for the ADLLDV2 model) with similar decay rates ranging from 24% to 34% per annum for the first three models; for the ADLLDV2 model, the initial decay rate is only 7% in the first year but increases to 33% (one minus the sum of the coefficients on the lagged dependent variable) after the first year. Given the standard errors on these coefficients, the decay rates are quite similar. Thus, for example, a classical confidence interval for the decay rate in the ADL model is (11%, 38%), and in the ADLLDV2 model after the first year it is (17%, 49%).

Turning to the estimated effect of the various independent variables (omitting the two political variables, which show almost no effect but huge standard errors), recall that the SC specification assumes that the effect of the variables is only instantaneous, the LDV model assumes the effect decays geometrically, and the ADL and ADLLDV2 models allow us to test those assumptions. In those latter specifications, the coefficients on the current and lagged values of *VULN* and *FDI* are close in absolute value and of opposite sign. Thus, the impact of those variables on capital taxation rates is more or less only instantaneous, and the ADL coefficient estimates of this instantaneous effect are similar to the SC estimates but different from the LDV estimates. Of course, the ADL specifications allow us to study the speed of adjustment, whereas the SC specification just assumes instantaneous adjustment.

The coefficients on *UNEM* and *GDPPC* and their lags are of opposite sign but do not quite offset each other. Here the ADL estimates are, as we would expect, much closer to the LDV estimates than to the SC estimates. But again, we need not make the assumptions about decay rates that the LDV specification imposes; instead, we can examine what the decay process looks like. Interestingly, and contrary to Achen's notion of the lagged dependent variable "dominating" a regression, the coefficients of all four of these substantive coefficients are as large as or larger than the similar coefficients in the SC specification.

The variable *AGED* determines tax rates in the SC specification but fails to show any impact in any of the other specifications. Intuitively, although *AGED* perhaps "ought" to affect tax rates, its coefficient in the SC specification "seems" a bit large; would a one-point increase in the aged population be expected to lead to a more-than-one-point increase in capital taxation rates? Perhaps it is not so simple to discuss which results "make sense," and making sense is hardly a statistical criterion. Note also that *AGED* is itself highly trending (its autoregression has a coefficient of 0.93 with a standard error of 0.01). Although we can reject the null that *AGED* has a unit root, it, like the capital tax rate, changes very slowly. Thus, we might suspect that the simple contemporaneous relationship between the two variables is spurious (in the sense of Granger & Newbold 1974). Of course we cannot know the "truth" here, but it is not obvious that the ADL (or LDV) results on the impact of *AGED* are somehow foolish or wrong.

The moral so far is that researchers should estimate a model flexible enough to account for various types of dynamics; they should also try hard to make sure that the error process is close to iid. The ADLLDV2 model performs very well here, both in terms of its passing various tests and its interpretability (with the simpler ADL model being easier to interpret but not quite passing the statistical tests). While no specification will be good in all situations, it is clear that researchers should not consider more general specifications before accepting highly constrained ones such as either the SC or LDV model.

Our focus is on dynamics, but no TSCS analysis is complete without a final assessment of heterogeneity over countries. Remember that our analysis uses country-centered data, so there can be no heterogeneity in the various centered means. But we can see if the model fails to work for some subset of countries by cross-validation (Stone 1974), leaving out one country at a time. Thus we reran the ADLLDV2 specification, leaving out one country at a time and then using the estimated values to predict capital tax rates in the omitted country. The mean absolute prediction error was then computed for each country. For all observations, the absolute forecast error was about 2.3. Four countries—Japan, Norway, Sweden, and the United Kingdom—had mean absolute forecast errors above 3.5, indicating at least some lack of homogeneity. We do not pursue this issue further here, but clearly this issue would be pursued in a more complete analysis. (We also do not present other post-estimation analyses that should be standard, such as residual plots by countries.)

We also assessed heterogeneity by testing for parameter heterogeneity (by country). Here, since we focus on dynamics, we fit the ADL specification allowing for the coefficient of the lagged dependent variable (θ) for each country to be a random draw from a normal distribution with zero mean. This allows us to see whether the general speed of adjustment varies by country. Results of this estimation reveal no statistically (or substantively) significant parameter heterogeneity (on the lagged dependent variable); the estimated standard deviation on the normal from which the coefficients were drawn was only 0.09.

The standard error of the estimated standard deviation was 0.07. A test of the null hypothesis that θ does not vary by country yields a statistic of 0.70; this statistic is χ^2_1, so far from the critical value for rejection of the null. We can look at the individual country estimates of θ. Most are within 0.01 of the overall estimate of θ, with only the coefficient for the United Kingdom really differing; the estimated θ for the United Kingdom is 0.11 under the overall estimate for θ, though with a standard error of ~0.07. Given all this, we prefer not to pursue whether further investigation of the speed of adjustment in tax rates in the United Kingdom is needed, but clearly this type of analysis in other situations might prove extremely useful.

5.2. The Growth of Gross Domestic Product

Our second example relates to political economy explanations of the growth of GDP in 14 OECD nations observed from 1966 through 1990 (yielding 336 observations), using data from Garrett (1998). Our treatment is cursory for reasons of space. We use one of his models, taking the growth in GDP as a linear additive function of political factors and economic controls. The political variables are the proportion of cabinet posts occupied by left parties (*LEFT*), the degree of centralized labor bargaining as a measure of corporatism (*CORP*), and the product of the latter two variables (*LEFT* x *CORP*). The economic and control variables are a dummy marking the relatively prosperous period through 1973 (*PER73*), overall OECD GDP growth (*DEMAND*), trade openness (*TRADE*), capital mobility (*CAPMOB*), and a measure of oil imports (*OILD*). All variables, following Garrett's use of country fixed effects, were mean centered by country. As before, all standard errors are panel corrected.

GDP growth appears stationary, with an autoregressive coefficient of 0.32. Thus, all specifications are expected to show relatively fast dynamics, with quick returns to equilibrium. Turning to models with explanatory variables, results of estimating various specifications are in **Table 2**.

The static model showed modest serial correlation of the errors; a Lagrange multiplier test showed we could clearly reject the null of serially independent errors ($\chi^2_1 = 8.6$). Substantively, the serial correlation of the errors is small (0.12), so the OLS results are similar to the slightly more correct results in the two dynamic specifications.

Given the rapid speed of adjustment (the coefficient on the LDV is 0.16), it is not surprising that all three specifications show similar estimates. Very few coefficients are significant in any of the specifications, but the two variables that show a strong impact in the static specification continue to show a strong impact in the two dynamic specifications.

The similarity of the SC and LDV estimates is not surprising; because of the fast dynamics, the two models are not really very different. After one period, the various independent variables in the LDV specification have only 3% of their original impact; the long-run effects in the LDV specification are only 18% larger than the immediate impacts. Thus, the two specifications are saying more or less the same things, and the estimated coefficients are quite similar. Substantively, it appears as though GDP growth in a country is largely determined by

Table 2 Comparison of SC and LDV estimates of Garrett's (1998) model of economic growth in 14 OECD nations, 1966–1990 (country centered)

	Static		SC		LDV	
Variable	$\hat{\beta}$	PCSE	$\hat{\beta}$	PCSE	$\hat{\beta}$	PCSE
DEMAND	0.007	0.0012	0.007	0.002	0.007	0.001
TRADE	−0.018	0.019	−0.021	0.021	−0.019	0.019
CAPMOB	−0.20	0.21	−0.25	0.23	−0.24	0.21
OILD	−7.86	7.34	−6.69	7.89	−5.85	7.08
PER73	1.76	0.42	1.76	0.45	1.45	0.43
CORP	0.54	0.56	0.43	0.61	0.30	0.56
LEFT	−0.075	0.17	−0.076	0.18	−0.062	0.17
LEFTxCORP	0.10	0.53	0.10	0.56	0.17	0.52
GDP_L					0.16	0.07
ρ			0.12			

Abbreviations: LDV, lagged dependent variable; PCSE, panel corrected standard errors; SC, first-order serially correlated error model. See text for definitions of variables in the left-hand column.

GDP growth in its trading partners, and politics appears to play little if any role.

Both specifications were tested against the full ADL specification that contained all the one-year lags of the independent variables. Standard hypothesis tests do not come close to rejection of the simpler SC or LDV models in favor of the ADL model. Since the LDV and AIC specifications are not nested, discriminating between them is not so simple. Because both specifications have the same number of parameters, discrimination using standard information criteria (AIC or BIC) simplifies to comparisons of goodness of fit, on which criterion both specifications perform almost equally well. This will often be the case, since both the SC and LDV specifications imply very quick adjustments to equilibrium when the dynamic parameters are near zero.

In summary, the data are consistent with very short-run impacts, and it does not much matter how we exactly specify those dynamics. In terms of the Achen critique, there are two predictors of GDP that are strong in the SC model; they remain about equally strong in the LDV model. As we have argued, there is nothing about lagged dependent variables that makes them "dominate a regression" or makes "real" effects disappear. Given the nature of dynamics, this will always be the case when variables adjust quickly.

6. CONCLUSION

There is no cookbook for modeling the dynamics of TSCS models. Instead, careful examination of the specifications, and what they entail substantively, can allow TSCS analysts to think about how to model these dynamics. Well-known econometric tests help in this process, and standard methods make it easy to estimate the appropriate dynamic model. Modeling decisions are less critical where variables equilibrate quickly; as the adjustment process slows, the various specifications imply more and more different characteristics of the data. Analysts should take advantage of this to choose the appropriate model, namely, one that implies dynamics consistent with theoretical concerns. The specification chosen should of course be flexible enough to allow for testing against alternative dynamic specifications.

Being more specific, we have provided evidence that, contra Achen, there is nothing pernicious in the use of a model with a lagged dependent variable. Obviously attention to issues of testing and specification are as important here as anywhere, but there is nothing about lagged dependent variables that makes them generically harmful. As we have seen, there is a variety of generic dynamic specifications, and researchers should choose among them by using the same general methodology they use in other cases. The ADL model (or its ADL2 complication) is a good place to start; at that point, various specializations of the model can be tested against this general specification. Analysts, should, of course, interpret the dynamic results in substantive terms, focusing on long- as well as short-run effects.

Instead of pushing dynamics into a complicated error process that then must be "fixed up" to allow for estimation, it is much better to model the dynamics directly—that is, in terms of observable variables. There are both theoretical and practical advantages to this. The theoretical advantage is that dynamic issues become much more than nuisances for estimation. The practical advantage is that it is easy to estimate models with (approximately) independent error processes via OLS, and easy then to estimate these models with additional complicating cross-sectional features.

There are many important features of TSCS data, both in the temporal and spatial realms. Both sets of features are interesting, and neither should be swept under the rug. Fortunately, the econometrics involved with good TSCS modeling are not difficult, and a clear eye on specification and testing allows researchers to find substantively interesting features of the data.

DISCLOSURE STATEMENT

The authors are not aware of any affiliations, memberships, funding, or financial holdings that might be perceived as affecting the objectivity of this review.

ACKNOWLEDGMENTS

For research support we thank the National Science Foundation. An earlier version was presented at the Annual Meeting of the Society for Political Methodology, Stanford University, Stanford, California, July 29–31, 2004. We thank Geoff Garrett, Evelyne Huber, and John Stephens for providing data. We also thank the many colleagues who have discussed TSCS issues with us and allowed us to present in various forums, among whom Chris Achen and Simon Jackman (for efforts well beyond the call of duty) must be singled out.

LITERATURE CITED

Achen C. 2000. *Why lagged dependent variables can suppress the explanatory power of other independent variables*. Presented at Annu. Meet. Soc. Polit. Methodol., Jul. 20–22, Univ. Calif. Los Angeles

Adolph C, Butler DM, Wilson SE. 2005. *Which time-series cross-section estimator should I use now? Guidance from Monte Carlo experiments*. Presented at Annu. Meet. Am. Polit. Sci. Assoc., Sep. 1–4, Washington, DC

Alvarez RM, Katz JN. 2000. *Aggregation and dynamics of survey responses: the case of presidential approval*. Soc. Sci. Work. Pap. 1103, Div. Hum. Soc. Sci., Calif. Inst. Technol.

Anderson TW, Hsiao C. 1982. Formulation and estimation of dynamic models using panel data. *J. Econ.* 18:47–82

Beck N. 1985. Estimating dynamic models is not merely a matter of technique. *Polit. Methodol.* 11:71–90

Beck N. 1991. Comparing dynamic specifications: the case of presidential approval. *Polit. Anal.* 3:51–87

Beck N. 1992. The methodology of cointegration. *Polit. Anal.* 4:237–47

Beck N. 2001. Time-series–cross-section data: What have we learned in the past few years? *Annu. Rev. Polit. Sci.* 4:271–93

Beck N, Gleditsch KS, Beardsley K. 2006. Space is more than geography: using spatial econometrics in the study of political economy. *Int. Stud. Q.* 50:27–44

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 versus substance: specifying and estimating time-series–cross-section models. *Polit. Anal.* 6:1–36

Beck N, Katz JN. 2007. Random coefficient models for time-series–cross-section data: Monte Carlo experiments. *Polit. Anal.* 15:182–95

Beck N, Katz JN. 2009. *Modeling dynamics in time-series–cross-section political economy data*. Soc. Sci. Work. Pap. 1304, Div. Hum. Soc. Sci., Calif. Inst. Technol.

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

Davidson JH, Hendry DF, 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

De Boef S, Keele L. 2008. Taking time seriously: dynamic regression. *Am. J. Polit. Sci.* 52:184–200

Engle R, Granger CW. 1987. Co-integration and error correction: representation, estimation and testing. *Econometrica* 55:251–76

Franzese RJ, Hayes JC. 2007. Spatial econometric models of cross-sectional interdependence in political science panel and time-series-cross-section data. *Polit. Anal.* 15:140–64

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

Garrett G, Mitchell D. 2001. Globalization, government spending and taxation in the OECD. *Eur. J. Polit. Res.* 39:144–77

Granger CW, Newbold P. 1974. Spurious regressions in econometrics. *J. Econometrics* 2:111–20

Greene W. 2008. *Econometric Analysis*. Upper Saddle River, NJ: Pearson Prentice Hall. 6th ed.

Hamilton JD. 1994. *Time Series Analysis*. Princeton, NJ: Princeton Univ. Press

Hendry D, 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. *Econ. J.* 88:549–63

Hibbs D. 1974. Problems of statistical estimation and causal inference in time-series regression models. In *Sociological Methodology 1973–1974*, ed. H Costner, pp. 252–308. San Francisco: Jossey-Bass

Honaker J, King G. 2010. What to do about missing values in time series cross-section data. *Am. J. Polit. Sci.* 54:561–81

Huber E, Stephens JD. 2001. *Development and Crisis of the Welfare State*. Chicago: Univ. Chicago Press

Huber PJ. 1967. The behavior of maximum likelihood estimates under non-standard conditions. In *Proc. Fifth Annu. Berkeley Symp. Math. Stat. Probab.*, ed. LM LeCam, J Neyman, I:221–33. Berkeley: Univ. Calif. Press

Im KS, Pesaran MH, Shin Y. 2003. Testing for unit roots in heterogeneous panels. *J. Econometrics* 115:53–74

Judson KA, Owen AL. 1999. Estimating dynamic panel data models: a guide for macroeconomists. *Econ. Lett.* 65:9–15

Keele L, Kelly NJ. 2006. Dynamic models for dynamic theories: the ins and outs of lagged dependent variables. *Polit. Anal.* 14:186–205

Kiviet JF. 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel models. *J. Econometrics* 68:53–78

Levin A, Lin C-F, Chu C-SJ. 2002. Unit root tests in panel data: asymptotic and finite-sample properties. *J. Econometrics* 108:1–24

Little RJA, Rubin DB. 1987. *Statistical Analysis with Missing Data*. New York: Wiley

Mizon G. 1984. The encompassing approach in econometrics. In *Econometrics and Quantitative Economics*, ed. D Hendry, K Wallis, pp. 135–72. Oxford, UK: Blackwell

Newey WK, West KD. 1987. A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 55:703–8

Nickell S. 1981. Biases in dynamic models with fixed effects. *Econometrica* 49:1417–26

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

Annual Review of Political Science

Volume 14, 2011

Contents

Indexes

Errata

An online log of corrections to *Annual Review of Political Science* articles may be found at http://polisci.annualreviews.org/