Estimating Dynamic Panel Data Models in Political Science

Gregory Wawro
Department of Political Science, Columbia University,
New York, NY 10027
e-mail: gjw10@columbia.edu

Panel data are a very valuable resource for finding empirical solutions to political science puzzles. Yet numerous published studies in political science that use panel data to estimate models with dynamics have failed to take into account important estimation issues, which calls into question the inferences we can make from these analyses. The failure to account explicitly for unobserved individual effects in dynamic panel data induces bias and inconsistency in cross-sectional estimators. The purpose of this paper is to review dynamic panel data estimators that eliminate these problems. I first show how the problems with cross-sectional estimators arise in dynamic models for panel data. I then show how to correct for these problems using generalized method of moments estimators. Finally, I demonstrate the usefulness of these methods with replications of analyses in the debate over the dynamics of party identification.

1 Introduction

Many aspects of politics are inherently dynamic, and the literature in political science has no shortage of works that attempt to model the dynamics of political phenomena. At their simplest, these models include lags of dependent variables. Yet even these simple models can present substantial problems when it comes to estimating them. This is especially the case when estimating dynamic models using panel data. Panel data are a very valuable resource for finding empirical solutions to social science puzzles. These data enable researchers to analyze causal relationships more thoroughly because they provide opportunities to examine variation both within and across cross-sectional units, as well as allowing us to account for unobserved effects.[1]

Numerous studies in political science have used panel data to estimate models with dynamics. These studies include dynamic models of party identification (Jackson 1975;

Author's note: A previous version of this article was presented at the Political Methodology Summer 2000 Conference. The author would like to thank Don Green and David Yoon for sharing their data and thoughts on this project and Pedro Albarran, Manuel Arellano, Larry Bartels, Charlie Himmelberg, and Nolan McCarty for helpful comments. This research was supported by a grant from the Institute for Social and Economic Research and Policy at Columbia University and a John M. Olin Faculty Fellowship. The author alone is responsible for the contents of this article. All of the data and code necessary to replicate the analysis in this article are available at the *Political Analysis* Web site.

[1]See Baltagi (1995, pp. 1–6) and Hsiao (1986, pp. 1–5) for discussions of the advantages that panel data offer.

Fiorina 1981; Markus 1982; Franklin and Jackson 1983; Green and Palmquist 1990; Kabashima and Ishio 1998; Green and Yoon 2002), campaign finance (Krasno et al. 1994; Box-Steffensmeier and Lin 1996), and protest activity (Finkel and Mueller 1998). However, with the exception of Green and Yoon, these studies have not taken into account important estimation issues, which calls into question the inferences we can make from these analyses. First, these studies estimate models on a period-by-period basis (i.e., they estimate a separate, cross-sectional model for each time period), which is inefficient because this approach does not take advantage of the panel structure of the data and the information it provides. Second, these studies do not adequately account for individual-specific effects, which is one of the main motivations for doing panel analysis in the first place. With dynamic models, the failure to account explicitly for unobserved individual effects can lead to biased and inconsistent estimates of parameters of interest. Cross-sectional estimators, such as ordinary least squares (OLS), lose the desirable properties of unbiasedness and consistency because at least one of the explanatory variables on the right-hand side of the regression equation will be correlated with the disturbance term unless individual-specific effects are adequately accounted for. Standard instrumental variables fixups are unlikely to work in the dynamic context. Even panel data estimators such as least-squares dummy variables (LSDV) produce biased and inconsistent estimates for dynamic panel models.[2]

The purpose of this article is to discuss, in a political science context, dynamic panel data estimators that surmount the problems mentioned above. While much of what is in this article is available from reviews in the economics literature [e.g., Baltagi (1995, pp. 125–148) and Arellano and Honoré, 2001], this article discusses these methods in a way that is more relevant to political scientists and gives more intuition behind the methods than do existing discussions.

The article proceeds as follows. Sections 2 and 3 discuss issues related to modeling dynamics and show how the problems of bias and inconsistency arise in dynamic panel data models. Section 4 discusses the Anderson–Hsiao estimator, a very simple estimator for dynamic panel data. Section 5 reviews generalized method of moments (GMM) estimation and Section 6 discusses GMM estimators for dynamic panel data. I then demonstrate the usefulness of these methods in Section 7 with replications of analyses in the longstanding debate over the dynamics of party identification. Section 8 concludes.

2 Modeling Persistence

A useful way to build intuition about dynamic panel data models is to think about various approaches to modeling persistence in a data generating process (DGP). In a typical time-series analysis, lags of the dependent variable are included as regressors to model persistence. There are various reasons for including lags. The inclusion of lags accounts for partial adjustment of behavior over time. Individuals might partially adjust their behavior over time, for example, to reach a long-run equilibrium. Another motivation for including lags would be to account for particular factors, including exogenous shocks, that we believe to have continual effects over time. The coefficients on lagged dependent variables indicate whether these factors have a greater impact over time or whether their impact decays and the

[2]A literature search revealed no articles published in major political science journals that used appropriate methods for dynamic panel data. The monographs on panel data by Markus (1979a) and Finkel (1995) focus mainly on using panel data to correct for measurement error and largely ignore issues of modeling unit-specific heterogeneity—a key issue in the models discussed in this article.

rate at which it decays. Including lags of dependent variables as regressors is a parsimonious way of accounting for the persistent effects of explanatory variables in the past and can also help eliminate serial correlation in the disturbance term (cf. Beck and Katz 1996).

An alternative way to model persistence in the data is to assume that there are individual-specific effects that do not vary over time. The modeling of individual-specific effects is at the heart of panel data analysis.[3] Methods for panel data are concerned primarily with accounting for factors that have persistent effects within cross-sectional units over time but vary across cross-sectional units. One of the key advantages of panel data is that, with the appropriate methods, they enable researchers to account for the effects of variables that are too costly or difficult to observe. Accounting for this unobserved individual heterogeneity enables researchers to obtain better estimates of the effects of variables of interest that they are able to observe.

Some models include both of these approaches to modeling persistence. *Dynamic panel models* include as part of their specification both lagged dependent variables and unobserved individual-specific effects. These models are very powerful tools that allow for empirical modeling of dynamics while accounting for individual-level heterogeneity. They enable us to parse out whether past behavior directly affects current behavior or whether individuals are simply predisposed to behave one way or another. An individual's behavior in the past might have a direct impact on her behavior in current and future periods, for example, if she learns that a particular pattern of behavior brings about a preferred outcome. But it could also be the case that we observe a high correlation among past, current, and future behavior simply because an individual has a predilection to behave in a particular way. Because dynamic panel models explicitly include variables to account for past behavior and time-invariant individual-specific effects, they enable us to understand better what factors drive behavior over time, differentiating between "true" dynamics and factors that vary across, but not within, individuals over time, even though such factors are unobservable.

The study by Green and Yoon (2002) nicely demonstrates the usefulness of models that combine individual specific effects with dynamics. Green and Yoon are concerned with heterogeneity in the process that generates party identification at the individual level. This process has been traditionally modeled as a dynamic one, and a key concern of theirs is the degree of persistence of shocks to party identification, which is modeled by including a lag of the dependent variable. Yet they are also concerned with individual-level heterogeneity, which they model in part with individual-specific effects that do not vary over time. By including individual-specific effects in their model, they are able to get around the unattractive assumption of other models that each cross-sectional unit has the same intercept, which implies that every individual returns to the same equilibrium level of partisanship. By using dynamic panel methods to estimate this model, they are able to test whether the individual-level heterogeneity that is posited by macro-level models of partisanship does in fact exist. Their analysis constitutes a major improvement over analyses that use simple cross-section estimators. However, as I demonstrate in a replication of their analysis, we must be careful when choosing from among the various dynamic panel estimators that are available. But

[3]This article focuses mainly on methods for panel data as opposed to time-series cross-section (TSCS) data. The traditional distinction made in the literature between these two types of data is that the former has many cross-sectional units but relatively few time periods, while the latter has few cross-sectional units but relatively many time periods. The desirable properties of the estimators discussed in this article depend on the number of cross-sectional units being large (i.e., the asymptotics are in N, instead of T) and so may perform well on data with a large number of both time periods and cross-sectional units, but I do not address this issue in this article.

first, I discuss why special estimators are necessary to estimate the kind of model examined by Green and Yoon.

3 Dynamic Panel Data and Cross-Sectional Estimators

OLS, which is used by most of the political science studies mentioned above, is both biased and inconsistent when used to estimate dynamic models with panel data. It is straightforward to see how these problems arise in the dynamic case. Consider the following representative regression model for panel data:

$$y_{i,t} = \gamma y_{i,t-1} + \mathbf{x}_{i,t}\boldsymbol{\beta} + \alpha_i + u_{i,t} \tag{1}$$

where i denotes the cross-sectional units ($i = 1, \ldots, N$), t denotes the time period ($t = 1, \ldots, T$), $\mathbf{x}_{i,t}$ is a vector of exogenous explanatory variables, γ and $\boldsymbol{\beta}$ are parameters to be estimated, and $u_{i,t}$ is a random disturbance term.[4] α_i is an unobservable individual-specific effect which is constant across time within individuals.[5] For the disturbance term, we assume that

$$E[u_{i,t} \mid y_{i,t-1}, \ldots, y_{i,1}, \mathbf{x}_{i,t}, \mathbf{x}_{i,t-1}, \ldots, \mathbf{x}_{i,1}] = 0 \tag{2}$$

(that is, we assume that the disturbance term has mean zero conditional on all past values of the endogenous variable and all past and present values of the exogenous variables). Throughout most of the exposition in this article, I assume that the $u_{i,t}$ are serially uncorrelated and homoskedastic. However, in the discussion that follows, I will point out when violations of the assumption of spherical disturbances affect the properties of the estimators. I will also discuss tests of these assumptions and corrections when they are violated.

Most of the political science studies cited above estimate some variation of Eq. (1) ignoring the individual-specific effect, α_i. That is, they estimate

$$y_{i,t} = \gamma y_{i,t-1} + \mathbf{x}_{i,t}\boldsymbol{\beta} + u^*_{i,t} \tag{3}$$

where $u^*_{i,t} = \alpha_i + u_{i,t}$. The individual effect here is relegated to the disturbance term. This would be appropriate if we believe that the explanatory variables in the equation adequately account for all individual heterogeneity. But to believe this, especially given the data that social scientists typically have, is to be overly optimistic. Indeed, part of the attractiveness of panel data is that they enable us to estimate models where we can account for individual-specific variables that are theoretically relevant but too difficult or too costly to measure.

[4] Equation (1) can be generalized to include additional lags of $y_{i,t}$ and lags of $\mathbf{x}_{i,t}$, as well as period-specific effects. Although the estimators that are discussed below can handle more complex models of this type, I employ simpler models such as Eq. (1) for ease of exposition. Unless noted otherwise, I assume that panels are balanced or, if they are unbalanced, that observations are missing at random. Nonrandom attrition from dynamic panel data is a very difficult problem and I do not address it in this article.

[5] Although whether we treat the individual-specific effects as fixed or random is often an important issue, it is trivial for the dynamic panel data estimators discussed in this paper because these estimators remove these effects, for example, by differencing.

To see why ignoring individual-specific effects is problematic, consider what happens if we lag Eq. (1) by one period:

$$y_{i,t-1} = \gamma y_{i,t-2} + \mathbf{x}_{i,t-1}\boldsymbol{\beta} + \alpha_i + u_{i,t-1}. \tag{4}$$

Since $y_{i,t-1}$ is correlated with α_i by construction, $y_{i,t-1}$ is correlated with the disturbance $u^*_{i,t}$ in Eq. (3), which means that the OLS estimates of γ and β are biased and inconsistent.[6] We should be leery, therefore, about drawing inferences based on parameter estimates produced by cross-sectional regression methods.

Unfortunately, even standard panel data estimators are not appropriate for estimating models like Eq. (1). For example, the standard "within-group" or LSDV transformation to remove the individual effects produces biased and inconsistent estimates because correlation remains between the transformed lagged dependent variable and the transformed disturbance (Baltagi 1995, pp. 125–126; Kiviet 1995). The bias is of order $1/T$ and so remains a problem in typical panel data sets where T is small (Nickell 1981).

To obtain consistent estimates, the individual effects have to be dealt with first. The most common approach is to transform the equation to remove the individual-specific effects. This eliminates the problem of correlation between the lagged dependent variable and the individual-specific component in the error term. Since the individual specific effects are removed, it does not matter whether we conceive of them as fixed or random in the original model. However, these transformations create a different kind of correlation between the lagged endogenous variable and the disturbance in the transformed equation. Instrumental variables are then employed to eliminate this problem in the transformed equation. In the next section, I discuss the first instrumental variable estimator developed for dynamic panel data.

4 The Anderson–Hsiao Estimator

The first-difference estimators developed by Anderson and Hsiao (1981, 1982) are a good place to start the discussion of specific dynamic panel data estimators since it is relatively easy to see how these estimators correct for the problems associated with cross-sectional estimators. Anderson and Hsiao first pointed out that first-differencing Eq. (22) (below) eliminates the problem of correlation between the lagged endogenous variable and the individual-specific effect. First-differencing Eq. (1) gives

$$y_{i,t} - y_{i,t-1} = \gamma(y_{i,t-1} - y_{i,t-2}) + (\mathbf{x}_{i,t} - \mathbf{x}_{i,t-1})\boldsymbol{\beta} + u_{i,t} - u_{i,t-1} \tag{5}$$

which can be rewritten

$$\Delta y_{i,t} = \gamma \Delta y_{i,t-1} + \Delta \mathbf{x}_{i,t}\boldsymbol{\beta} + \Delta u_{i,t} \tag{6}$$

where Δ denotes the difference operator. Although the individual-specific effects have been differenced out, there is still correlation between right- hand- side variables and the disturbance term because $y_{i,t-1}$ in $\Delta y_{i,t-1}$ is, by construction, correlated with $u_{i,t-1}$ in $\Delta u_{i,t}$. Instrumental variables can be used to purge this correlation, however, and a set of instruments is conveniently supplied by the panel structure of the data. Anderson and Hsiao

[6]It is also possible that the $\mathbf{x}_{i,t}$ are correlated with α_i, which exacerbates the problem.

note that $y_{i,t-2} - y_{i,t-3}$ and $y_{i,t-2}$ are correlated with $y_{i,t-1} - y_{i,t-2}$ but not $u_{i,t} - u_{i,t-1}$, assuming that there is no serial correlation. The same is true for $\mathbf{x}_{i,t-2} - \mathbf{x}_{i,t-3}$ and $\mathbf{x}_{i,t-2}$. This makes these variables valid as instruments for estimating the parameters in Eq. (6). Anderson and Hsiao proposed two instrumental variable estimators that are consistent: the first uses $\Delta y_{i,t-2}$ as an instrument and the second simply uses $y_{i,t-2}$.

While the Anderson–Hsiao estimators solve the problems that cross-sectional estimators have when used with dynamic panel data, they are not without problems of their own. Arellano (1989) shows that the estimator that uses differences as instruments suffers from singularities as well as large variances over a range of values for γ. Thus the estimator that uses instruments in levels is preferred to the one that uses instruments in differences. Monte Carlo work, however, has shown that the levels estimator can have large biases and large standard errors, particularly when γ is close to 1.[7] The Anderson–Hsiao levels estimator can work adequately for lower values of γ and with larger samples, but subsequently developed estimators are likely to serve researchers better and are recommended.

The subsequent improvements on Anderson and Hsiao's estimators have built on their innovation of using instrumental variables made available by the panel structure of the data. These studies have adopted the GMM framework to derive estimators that surmount the problems of Anderson–Hsiao. The GMM estimators that have been developed employ *orthogonality restrictions* or *moment conditions* to derive valid instruments. The key intuition behind these methods is that the panel structure of the data provides a large number of instrumental variables in the form of lagged endogenous and exogenous variables. These estimators are generally more efficient than the Anderson–Hsiao estimators because they use additional instrumental variables that the Anderson–Hsiao estimators neglect. Since GMM estimators are not commonly used (at least self-consciously) in political science, in the next section I give a brief review of the concepts behind GMM estimation before turning to specific estimators for dynamic panel data. Readers who are familiar with GMM can skip the next section and still find the remaining sections of the paper accessible.

5 A Brief Review of GMM Estimation

Although the explicit use of GMM estimators is rare in political science, the key working parts of these estimators should be familiar to political scientists. GMM provides a unifying framework under which most estimators encountered in political science can be derived.[8] Readers should recognize similarities between GMM estimators and the more familiar techniques of generalized least squares (GLS) and two-stage least squares (2SLS).

The main idea behind GMM is that, from a set of basic assumptions about a DGP, we can establish population moment conditions and then use sample analogs of these moment conditions to compute parameter estimates. The population moment conditions typically involve expectations of functions of the disturbance term and explanatory variables, while the sample analogues of the population moment conditions take the form of sample moments (e.g., sample means).

[7] At the extreme, Arellano and Bover (1995, p. 46) concluded from a Monte Carlo study that the Anderson–Hsiao levels estimator is "useless" when $N = 100$, $T = 3$, and the coefficient on the lagged endogenous variable is .8.

[8] This section draws from the discussions of GMM estimators by Greene (1997) and Wooldridge (no date). Hansen (1982) shows that GMM estimators are consistent and asymptotically normally distributed. Thus, if an estimator can be shown to be a GMM estimator (i.e., can be derived using the GMM framework discussed in this section), then the "goodness" properties of consistency and asymptotic efficiency automatically follow.

To fix ideas about how GMM works, consider the cross-sectional regression

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + u_i \tag{7}$$

where we adopt the key identifying assumption

$$E[\mathbf{x}_i' u_i] = \mathbf{0} \tag{8}$$

(here $\mathbf{x}_i$ is a $1 \times k$ matrix of explanatory variables, $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters to be estimated, and u_i is the disturbance). This basic assumption defines a set of moment conditions and is a weaker variant of the assumption in Eq. (2) discussed above. Again, this assumption simply means that the explanatory variables and disturbance term are uncorrelated. Substituting in for u_i, we can rewrite Eq. (8) as

$$E[\mathbf{x}_i'(y_i - \mathbf{x}_i\boldsymbol{\beta})] = \mathbf{0} \tag{9}$$

to get the moment conditions in terms of observables and parameters. Population moments are estimated consistently with sample moments, so the next step is to write down the sample analog of Eq. (9):

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i'(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}) = \mathbf{0} \tag{10}$$

where $\hat{\boldsymbol{\beta}}$ is our estimator. Multiplying this out and solving for $\hat{\boldsymbol{\beta}}$ gives

$$\hat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \mathbf{x}_i'\mathbf{x}_i\right)^{-1}\left(\sum_{i=1}^{N} \mathbf{x}_i' y_i\right) \tag{11}$$

which is identical to the equation for the OLS estimator of $\boldsymbol{\beta}$. We can rewrite Eq. (11) as $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ by stacking the $\mathbf{x}_i$ and y_i for observations $i = 1, \ldots, N$ into an $N \times K$ matrix $\mathbf{X}$ and an $N \times 1$ vector $\mathbf{y}$, respectively.

Now suppose that the assumption given by Eq. (8) does not hold, for example, because $x_{i,k}$ in $\mathbf{x}_i$ is correlated with u_i. If there are some variables $\mathbf{z}_i$ available for which

$$E[\mathbf{z}_i' u_i] = \mathbf{0} \tag{12}$$

and if the elements of $\mathbf{z}_i$ are partially correlated with $x_{i,k}$, then the $\mathbf{z}_i$ can serve as instrumental variables. The population moment conditions for the GMM estimator of $\boldsymbol{\beta}$ are

$$E[\mathbf{z}_i'(y_i - \mathbf{x}_i\boldsymbol{\beta})] = \mathbf{0} \tag{13}$$

which have the sample analog

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{z}_i'(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}) = \mathbf{0}. \tag{14}$$

If the number of columns in $\mathbf{z}_i$ (i.e., the number of moment conditions) is greater than the number of parameters to be estimated (which is typically the case), then our equation is overidentified and there is not a closed-form solution as with Eq. (10) (which was just identified).[9] To get around this problem we choose $\hat{\boldsymbol{\beta}}$ so that it minimizes the quadratic

$$\left(\sum_{i=1}^{N} \mathbf{z}_i'(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}})\right)' \mathbf{W} \left(\sum_{i=1}^{N} \mathbf{z}_i'(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}})\right) \tag{15}$$

where $\mathbf{W}$ is a positive semidefinite weighting matrix. The solution to this minimization problem does have a closed form, and with a little manipulation, we obtain

$$\hat{\beta} = \left(\mathbf{X}'\mathbf{ZWZ}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{ZWZ}'\mathbf{y}\right). \tag{16}$$

Readers should note the similarities between this GMM estimator and expressions for 2SLS estimators (note the $\mathbf{Z}$'s) and GLS estimators (note the $\mathbf{W}$'s). It can be shown that the asymptotic variance of $\sqrt{N}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is

$$\boldsymbol{\Omega} = (E[\mathbf{X}_i'\mathbf{Z}_i]\mathbf{W}E[\mathbf{Z}_i'\mathbf{X}_i])^{-1}E[\mathbf{X}_i'\mathbf{Z}_i]\mathbf{WVW}E[\mathbf{Z}_i'\mathbf{X}_i](E[\mathbf{X}_i'\mathbf{Z}_i]\mathbf{W}E[\mathbf{Z}_i'\mathbf{X}_i])^{-1} \tag{17}$$

where

$$\mathbf{V} = \text{Var}[\mathbf{Z}_i'u_i] = E[\mathbf{Z}_i'u_iu_i'\mathbf{Z}_i].$$

To make the GMM estimator efficient, we choose $\mathbf{W} = \mathbf{V}^{-1}$, which makes $\boldsymbol{\Omega}$ as small as possible (Hansen 1982). Substituting $\mathbf{V}^{-1}$ for $\mathbf{W}$ in Eq. (17) gives

$$\boldsymbol{\Omega} = (\mathbf{X}_i'\mathbf{Z}_i\mathbf{V}^{-1}\mathbf{Z}_i'\mathbf{X}_i)^{-1}. \tag{18}$$

A consistent estimator for $\mathbf{V}^{-1}$, which produces standard errors that are robust to nonspherical disturbances, is

$$\hat{\mathbf{W}} = \hat{\mathbf{V}}^{-1} = \left\{\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\hat{u}_i\hat{u}_i'\mathbf{Z}_i\right\}^{-1} \tag{19}$$

where the $\hat{u}_i$ are estimated residuals produced using a first-stage, consistent estimator of $\boldsymbol{\beta}$. Plugging in $\hat{\mathbf{W}}$ and $\hat{\mathbf{V}}^{-1}$ in Eqs. (16) and (18) produces the asymptotically optimal GMM estimator.

This review of GMM estimation has been stated in terms of cross-sectional models. In the next sections, I discuss in detail several GMM estimators that have been derived for panel data models with lagged dependent variables. The basic form of these estimators is the same as the form for cross-sectional models, but a key feature of GMM estimators for dynamic panel models is that they exploit the panel structure of the data to construct instruments that satisfy moment conditions like Eq. (13).

[9]Note that if the number of moment conditions is less than the number of coefficients to be estimated, then the model is not identified.

6 GMM Estimators for Dynamic Panel Data

Researchers analyzing dynamic panel data have many GMM estimators from which to choose. In this section, I discuss some of the key estimators that have been developed, pointing out what assumptions underlie the estimators, how well the estimators perform relative to each other, and how to test the validity of the assumptions behind the estimators to choose the most appropriate estimator for the data in question.

6.1 *A First-Difference Estimator*

Arellano and Bond (1991) note that the Anderson–Hsiao estimator is inefficient because it does not use all available instruments and can be improved upon by placing it in a GMM framework—a significant contribution which led to a growth industry of developing GMM estimators for dynamic panel data.[10] If we assume that $E(u_{i,t}) = 0$ and $E(u_{i,t}u_{i,s}) = 0$ (i.e., the disturbance has mean zero and is not serially correlated), then the transformed residuals in Eq. (5) have zero covariance between all $y_{i,t}$ and $\mathbf{x}_{i,t}$ dated $t-2$ and earlier. This means that we can go back through the panel from period $t-2$ to obtain appropriate instrumental variables for purging the correlation between $\Delta y_{i,t-1}$ and $\Delta u_{i,t}$. The transformed residuals satisfy a large number of moment conditions of the form

$$E[\mathbf{z}'_{i,t}\Delta u_{i,t}] = \mathbf{0}, \quad t = 2, \ldots, T \tag{20}$$

where $\mathbf{z}_{i,t} = (y_{i,t-2}, \mathbf{x}_{i,t-2}, y_{i,t-3}, \mathbf{x}_{i,t-3}, \ldots, y_{i,1}, \mathbf{x}_{i,1})'$ denotes the instrument set at period t.[11]

For notational efficiency, we can stack the time periods to write down a system of T equations for each individual:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\theta} + \mathbf{u}_i \tag{21}$$

where

$$\mathbf{y}_i = \begin{bmatrix} \Delta y_{i,3} \\ \Delta y_{i,4} \\ \vdots \\ \Delta y_{i,T} \end{bmatrix}, \quad \mathbf{X}_i = \begin{bmatrix} \Delta y_{i,2} & \Delta \mathbf{x}_{i,3} \\ \Delta y_{i,3} & \Delta \mathbf{x}_{i,4} \\ \vdots & \vdots \\ \Delta y_{i,T-1} & \Delta \mathbf{x}_{i,T} \end{bmatrix}, \quad \text{and} \quad \mathbf{u}_i = \begin{bmatrix} \Delta u_{i,3} \\ \Delta u_{i,4} \\ \vdots \\ \Delta u_{i,T} \end{bmatrix}.$$

The set of instruments is given by the block diagonal matrix

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{z}_{i,3} & & & 0 \\ & \mathbf{z}_{i,4} & & \\ & & \ddots & \\ 0 & & & \mathbf{z}_{i,T} \end{bmatrix}.$$

[10]Holtz-Eakin et al. (1988), who work in the context of vector autoregressions, also noted the availability of additional instruments.

[11]If we assume strict exogeneity of all $\mathbf{x}_{i,t}$ [i.e., $E(\mathbf{x}'_{i,t}u_{i,s}) = 0$ for all t and s], then $\mathbf{x}_{i,t}$ at periods later than $t-2$ can also be used as instruments.

Note that this means that the number of instruments increases as we move through the panel. For example, suppose that Eq. (1) includes $y_{i,t-1}$ as the only explanatory variable:

$$y_{i,t} = \gamma y_{i,t-1} + \alpha_i + u_{i,t}. \tag{22}$$

Then the instrument matrix becomes

$$\mathbf{Z}_i^* = \begin{bmatrix} y_{i,1} & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & y_{i,1} & y_{i,2} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & y_{i,1} & y_{i,2} & \cdots & y_{i,t-2} \end{bmatrix}.$$

Hence, if $T = 5$, then we would have

$$\mathbf{Z}_i^* = \begin{bmatrix} y_{i,1} & 0 & 0 & 0 & 0 \\ 0 & y_{i,1} & y_{i,2} & 0 & 0 \\ 0 & 0 & y_{i,1} & y_{i,2} & y_{i,3} \end{bmatrix}$$

for our matrix of instruments. Using the notation for the stacked equations, we can write the vector of population moment conditions as

$$E[\mathbf{Z}_i'\mathbf{u}_i] = \mathbf{0}. \tag{23}$$

The sample analogue of Eq. (23) that we use to construct an optimal GMM estimator for $\boldsymbol{\theta} = (\gamma, \boldsymbol{\beta})$ is

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{u}_i = \mathbf{0}. \tag{24}$$

Let

$$\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{bmatrix}, \quad \text{and} \quad \mathbf{Z} = \begin{bmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_N \end{bmatrix}$$

(here we are just stacking the observations for all of the cross-sectional units for all periods). Then we can reexpress Eq. (24) as

$$\frac{1}{N}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) = \mathbf{0}.$$

The optimal GMM estimator is then given by

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}'\mathbf{Z}\hat{\mathbf{V}}^{-1}\mathbf{Z}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}\hat{\mathbf{V}}^{-1}\mathbf{Z}'\mathbf{y} \tag{25}$$

where $\hat{\mathbf{V}}$ is a consistent estimate of $\mathbf{V}$, the limiting variance of the sample moments, $E[\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i]$. In general, the optimal choice for $\hat{\mathbf{V}}$ is

$$\hat{\mathbf{V}}_r = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i\mathbf{Z}_i$$

where $\hat{\mathbf{u}}_i$ is an estimate of the vector of residuals, $u_{i,t}$, obtained from an initial consistent estimator. Arellano and Bond (1991) suggest using $\hat{\mathbf{V}}_c = \frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{H}\mathbf{Z}_i$ to produce the initial consistent estimator, where

$$\mathbf{H} = \begin{bmatrix} 2 & -1 & 0 & 0 & \cdots & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 \\ 0 & -1 & 2 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 2 \end{bmatrix}.$$

By the properties of GMM estimators, with T fixed and $N \to \infty$, $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically distributed as $N(\boldsymbol{\theta}, \boldsymbol{\Sigma})$ (Hansen 1982). The asymptotic variance $\boldsymbol{\Sigma}$ is equal to

$$\left\{E(\mathbf{X}_i'\mathbf{Z}_i)E[\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i]^{-1}E(\mathbf{Z}_i'\mathbf{X}_i)\right\}^{-1}.$$

A consistent estimator of the asymptotic variance is

$$\hat{\boldsymbol{\Sigma}} = \left(\mathbf{X}'\mathbf{Z}\hat{\mathbf{V}}_r^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}.$$

Standard errors for the first-difference estimates are obtained by taking the square root of the diagonal of $\hat{\boldsymbol{\Sigma}}$. If the disturbances are heteroskedastic, then the two-step estimator is more efficient. In practice, however, the asymptotic standard errors for the one-step estimator appear to be more reliable for making inferences in small samples (Arellano and Bond 1991; Blundell and Bond 1998).[12]

A discussion of the basic first-difference GMM estimator is useful for pedagogical purposes because it is relatively easy to see how it corrects for the problems associated with cross-sectional estimators. However, this estimator can perform very poorly in certain estimation situations—similar to those in which the Anderson–Hsiao estimator does poorly. Blundell and Bond (1998) performed an extensive Monte Carlo study and found that the first-difference estimator displays large downward biases and a serious lack of precision in estimating the autoregressive parameter when it is greater than .8. For the intuition behind this result, suppose that $T = 3$. Then Eq. (22) implies that

$$\Delta y_{i,2} = (\gamma - 1)y_{i,1} + \alpha_i + u_{i,2}. \tag{26}$$

Thus, $y_{i,1}$, the instrument for $\Delta y_{i,2}$ in the first-differenced equation, is only weakly correlated with $\Delta y_{i,2}$ for values of γ close to 1, which leads to downward bias in the autoregressive parameter.[13] If we expect a high degree of persistence in the dependent variable, then we should consider some alternative estimators that can overcome the problems associated with first-differencing. Even if we do not expect a high degree of persistence in the data, several alternative estimators are more attractive because they can give substantial efficiency

[12] Arellano and Bond's first-difference estimator can be implemented using the **xtabond** routine in Version 7 of Stata.

[13] Blundell and Bond show that downward bias also occurs if the variance of the individual effect is small relative to the variance of the disturbance term.

gains over the first-difference estimator. In the next section, I discuss GMM estimators that attempt to improve on the basic first-difference estimator by exploiting additional moment conditions.

6.2 *(Possibly Superior) Alternatives to First-Differencing*

One of the key ways in which researchers have sought to improve on the basic first-difference estimators is to exploit additional moment conditions that arise under alternative assumptions of the DGP. Arellano and Bover (1995) and Blundell and Bond (1998) note that under certain assumptions differences can be used as instruments to estimate equations in levels, in addition to the instruments that are available after first-differencing. Again, suppose that we have the simple dynamic model given by Eq. (22), and assume that

$$E[\alpha_i y_{i,t}] = E[\alpha_i y_{i,s}] \tag{27}$$

for all t and s (i.e., the lagged dependent variable has a constant correlation with the individual-specific effects), and $E[y_{i,t-1}u_{i,s}] = 0$, where $t - 1 < s$ (i.e., the lagged dependent variable is uncorrelated with present and past values of the disturbance). We can treat this model as part of a system of two equations—one in differences and one in levels:

$$\Delta y_{i,t} = \gamma \Delta y_{i,t-1} + \Delta u_{i,t} \tag{28}$$

$$y_{i,t} = \gamma y_{i,t-1} + u^*_{i,t} \tag{29}$$

where the second equation denotes the DGP in levels (recall from Section 3 that $u^*_{i,t} = \alpha_i + u_{i,t}$). This setup gives us moment conditions in levels in addition to the moment conditions we have with first-differencing. For example, for $T = 3$ we get the levels moment condition

$$E[u^*_{i,3}\Delta y_{i,2}] = 0. \tag{30}$$

This follows since

$$E[u^*_{i,3}\Delta y_{i,2}] = E[(\alpha_i + u_{i,3})(y_{i,2} - y_{i,1})] \tag{31}$$

$$= E[\alpha_i y_{i,2}] - E[\alpha_i y_{i,1}] + E[u_{i,3}y_{i,2}] - E[u_{i,3}y_{i,1}]. \tag{32}$$

The last two expectations are 0 because of our assumptions in Eq. (2), while the difference between the first two expectations is 0 because of Eq. (27). The additional moment condition leads to the matrix of instruments

$$\mathbf{Z}_i^+ = \begin{bmatrix} \mathbf{Z}_i^* & \mathbf{0} \\ \mathbf{0} & \Delta y_{i,2} \end{bmatrix}.$$

In their Monte Carlo studies, Arellano and Bover (1995) and Blundell and Bond (1998) show that estimators that exploit these kinds of additional moment conditions by using instrument matrices such as $\mathbf{Z}_i^+$ perform much better than the basic first-difference estimator, especially as γ approaches 1.[14]

[14] Ahn and Schmidt (1995) point out other moment conditions that can be exploited under alternative assumptions. If we assume that the $u_{i,t}$ are uncorrelated with each other and uncorrelated with α_i and $y_{i,0}$ (the initial value of

6.3 *The Forward Orthogonal Deviations Estimator*

Arellano and Bover (1995) propose an alternative transformation to first-differencing called forward orthogonal deviations. This transformation involves subtracting the mean of all future observations in the sample for each individual. With this transformation, the disturbance in Eq. (1) becomes

$$\tilde{u}_{it} = w_{it}\left(u_{i,t} - \frac{u_{i,t+1} + \cdots + u_{i,T}}{T-t}\right) \quad \text{for} \quad t = 1, \ldots, T-1$$

where $w_{i,t} = \sqrt{(T-t)/(T-t+1)}$ is a weight that equalizes the variance of the transformed errors. To implement this transformation we premultiply the stacked levels regression (i.e., not differenced)[15] by the forward orthogonal deviations operator

$$\boldsymbol{A} = diag[(T-1)/T, \ldots, 1/2]^{1/2}\boldsymbol{A}^{+} \tag{33}$$

where

$$\boldsymbol{A}^{+} = \begin{bmatrix} 1 & -(T-1)^{-1} & -(T-1)^{-1} & \cdots & -(T-1)^{-1} & -(T-1)^{-1} & -(T-1)^{-1} \\ 0 & 1 & -(T-2)^{-1} & \cdots & -(T-2)^{-1} & -(T-2)^{-1} & -(T-2)^{-1} \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1/2 & -1/2 \\ 0 & 0 & 0 & \cdots & 0 & 1 & -1 \end{bmatrix}.$$

For the intuition on how this transformation works, note that the elements in the rows of this matrix sum to 0, which means that it will remove the individual-specific effects, and the upper triangularity of $\boldsymbol{A}^{+}$ ensures the validity of the lagged endogenous variables as instruments. The transformation thus works in a manner similar to the first-difference transformation. As Arellano and Honoré (2001) point out, a key difference between the two transformations is that orthogonal deviations does not introduce a moving average process in the disturbance, assuming that the disturbance is not initially autocorrelated and has constant variance. Thus orthogonal deviations can intuitively "be regarded as doing first differences to eliminate the effects plus a GLS transformation to remove the serial correlation induced by differencing" (Arellano and Bover 1995, p. 42).[16]

These alternative estimators exploit additional moment conditions that give us more instruments than the basic first-difference estimator.[17] Since valid instruments are typically

---

$y_{i,t}$), then we get the following nonlinear moment conditions: $E[u_{i,T}\Delta u_{i,t}] = 0$ for $t = 2, \ldots, T-1$. Adding the homoskedasticity assumption that $E(u_{i,t}^2)$ is the same for all T gives still more moment conditions that can be exploited. Ahn and Schmidt go on to show that employing these additional moment conditions leads to substantial gains in asymptotic efficiency, especially if γ is close to unity.

[15] By the "stacked levels regression" I mean an equation that looks like Eq. (21) but is in levels instead of differences, such that $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,T})'$, $\mathbf{X}_i = (y_{i,2} \quad \mathbf{x}_{i,1}, \ldots, y_{i,T-1} \quad \mathbf{x}_{i,T})'$, and $\mathbf{u}_i = (u_{i,1}, \ldots, u_{i,T})'$.

[16] Keane and Runkle (1992) derive a "forward filter" estimator that is similar to the forward orthogonal deviations estimator. Their estimator gives efficiency gains when the disturbances are nonspherical within cross-sectional units and is especially useful when the disturbances are serially correlated. The key similarity is the use of an upper-triangular matrix to remove individual specific effects while preserving the validity of instruments.

[17] The DPD98 software written in Gauss by Arellano and Bond can compute the orthogonal deviations estimator and the first-difference estimator, as well as the versions of these estimators that use differences as instruments for the levels equation. The results I report for the replications below were obtained using the DPD98 code,

difficult to come by in political science data, it is tempting to be ecstatic about the plethora of valid instruments that the panel structure of the data provides for dynamic panel models. Yet we may not want to use all of the instruments that are available. In finite samples, there is a bias/efficiency trade-off that starts to bite as T increases in size. As the number of periods increases, more and more instruments become available extending back to the first period. Yet instruments from the earlier periods in the panel become weaker the farther we progress through the panel. Using all of the instruments in the panel, while efficient, can cause severe downward bias in GMM estimators when our sample is finite (Ziliak 1997). The bias occurs because we are overfitting the model by including instruments that are too weakly correlated with the explanatory variables. There is no general rule as to when the bias/efficiency trade-off becomes problematic. But if T is large relative to N, then we may want to try using fewer instruments for the later periods to see if it makes a difference. One approach is to use a "stacked" estimator that reduces the number of instruments by choosing linear combinations of the moment conditions rather than treating them as separate (e.g., see Himmelberg and Wawro 1998).

While using any of the dynamic panel data estimators will theoretically be an improvement over cross-sectional estimators, not all of the dynamic panel data estimators will perform equally well. If one is particularly concerned with efficiency, the systems of equations estimators that exploit moment conditions in levels are more attractive. The systems estimators are certainly preferred if one expects the autoregressive parameter to be close to 1. In addition to these rules of thumb, there are specification checks that can help one choose between estimators that are more and estimators that are less demanding in terms of the assumptions necessary to generate additional moment restrictions. I discuss these checks in the next subsection.

6.4 *Specification Tests*

It is important to note that the consistency of the estimators discussed in the previous section depends crucially on the assumption that the $u_{i,t}$ in Eq. (1) are serially uncorrelated. If serial correlation exists, then some of our instruments will be invalid and the moment conditions used to identify parameters will not hold. To judge the reliability of our estimates, it is advisable to conduct tests for serial correlation. Arellano and Bond (1991) note that if there is no serial correlation in the $u_{i,t}$ in Eq. (1), then the first-differenced residuals should display negative first-order serial correlation but not second-order serial correlation. If the $u_{i,t}$ are not serially correlated, then first-differencing produces the first-order moving average, or MA(1), process $u_{i,t} - u_{i,t-1}$. But suppose that our disturbances for the levels equation are $u_{i,t} - \rho u_{i,t-1}$. Then differencing produces the disturbance term $u_{i,t} - u_{i,t-1} - \rho(u_{i,t-1} - u_{i,t-2})$, which displays second-order serial correlation. With this type of serial correlation, $y_{i,t-2}$ is not valid as an instrument since it will be correlated with $u_{i,t-2}$ in the differenced disturbance term, although lagged y's at period $t - 3$ and earlier remain valid instruments.

Arellano and Bond (1991) give tests of first- and second-order serial correlation based on the residuals from the two-step estimator of the first-differenced equation. A drawback of this test is that the complete test (i.e., for both first- and second-order serial correlation) can be performed only for samples where $T \geq 5$, although the test statistic for first-order serial correlation is defined for samples where $T \geq 4$.

which is freely available at http://www.cemfi.es/~arellano/.

Arellano and Bond also derive an omnibus test of overidentifying restrictions (or moment conditions), which helps to determine whether our assumptions about serial correlation are correct. It tests whether the moment conditions over and above those needed to identify the parameters are valid. One advantage of this test is that it is defined for samples where $T \geq 3$. Its disadvantages are that it can reject the restrictions due to forms of misspecification other than serial correlation and its asymptotic distribution is known only when the disturbance term is homoskedastic.

The Sargan test (cf. Sargan 1958; Hansen 1982) that they develop is performed by computing

$$s = \hat{\mathbf{u}}'\mathbf{Z}\left(\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i\right)^{-1}\mathbf{Z}'\hat{\mathbf{u}} \tag{34}$$

where $\hat{\mathbf{u}} = (\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_N)'$, which represents the stacked vectors of estimated first-differenced residuals for all i and T. The test statistic s is distributed asymptotically as χ^2 with degrees of freedom equal to the number of columns in $\mathbf{Z}$ minus the number of explanatory variables. A significant χ^2 value for this test indicates that the overidentifying restrictions are invalid. The intuition behind this test is that if the moment conditions given by Eq. (23) hold, then the sample moments given by Eq. (24) when evaluated at the parameter estimates should be close to 0, and hence the value of the quadratic function in Eq. (34) should be small. The farther away from 0 the sample moments are, the larger the s statistic will be. Rejection of the overidentifying restrictions should lead one to reconsider the specification of the model, possibly reducing the number of instruments employed or including more lags to eliminate serial correlation.

Differences between Sargan test statistics can be used to test the validity of additional moment conditions. For example, we could test the validity of adding moment conditions in levels to an estimator that uses moment conditions in differences. The difference between the Sargan statistics is distributed as χ^2 with degrees of freedom equal to the number of new moment conditions that are used. A significant χ^2 value would indicate that the additional moment conditions are not valid and should not be used. Even though there may be gains to using the additional moment conditions, we should use them only if the specification tests do not reject the assumptions necessary to produce the moment conditions. In the applications of dynamic panel estimators in Section 7, I show the importance of using specification tests to determine which of the various estimators available are most appropriate for the data in question.

7 Applying Dynamic Panel Methods: Stability in Party Identification

Some of the most popular panel data sets in political science are those that have been produced by the American National Election Studies (ANES). The ANES panel studies have been used in a contentious debate about the long-standing notion that party identification is extremely stable and changes only at a glacial pace (Jackson 1975; Markus 1979b, 1982; Franklin and Jackson 1983; Jennings and Markus 1984; Green and Palmquist 1990). Several studies have estimated dynamic models of party identification to assess the stability of party identification. In this section, I demonstrate the usefulness of dynamic panel estimators by replicating some of this work. An important lesson to be learned from the replications is that it may not be enough simply to use a dynamic panel estimator instead of a cross-sectional estimator—researchers must be careful when choosing among available dynamic panel estimators. Dynamic panel estimators are generally more reliable than cross-sectional estimators, yet some types of dynamic panel estimators will be more reliable than others. As this section demonstrates, different dynamic panel estimators can give different results, but the judicious use of specification checks can help choose the most appropriate estimator from which to draw inferences.

7.1 *Reanalysis of Green and Palmquist: Party Identification and "Short-Term Forces"*

In an important article in this literature, Green and Palmquist (1990) model party identification and short-term forces (STF) as part of the following system of equations:

$$P_{i,t} = \beta_{12} STF_{i,t} + \gamma_{11} P_{i,t-1} + u_{1i} \quad (35)$$

$$STF_{i,t} = \beta_{21} P_{i,t} + \gamma_{22} STF_{i,t-1} + u_{2i}. \quad (36)$$

However, they estimate only the parameters in the first structural equation in the system, using $STF_{i,t-1}$ as an instrument for $STF_{i,t}$. Thus, they have not corrected for the problem of correlation between $P_{i,t-1}$ and individual-specific effects that would be contained in u_{1i}.[18]

The primary motivation of Green and Palmquist's analysis is the problem of errors in measuring party identification and how such errors can lead to incorrect inferences about the effects of STF on party identification. After applying the Wiley–Wiley method to correct for measurement error, they find that the relationships between STF and party identification vanish. However, Green and Yoon (2002) argue that the technique that Green and Palmquist use is invalid when the model assumes that intercepts vary over individuals. Since this is one of the motivating assumptions for employing dynamic panel data techniques, it is not appropriate to use the Wiley–Wiley method. Green and Yoon instead estimate a dynamic panel model of party identification using the Anderson–Hsiao estimator. Following their lead, I choose to deal with the problems presented by individual specific effects and lagged dependent variables and leave issues of measurement error to the side. However, if we assumed that the measurement error was constant across periods, then the individual-specific effect can help account for measurement error in the dependent variable. It is important to note that the dynamic panel methods discussed above perform the same endogeneity correction that Green and Palmquist apply because the instrument set will include lags of the STF variable, although like Green and Palmquist, I do not estimate the full system.

To test their model, Green and Palmquist examined the relationship between party identification and a battery of evaluations of the presidential candidates in the 1980 elections. I report only four of the nine regressions that Green and Palmquist performed to maintain a focused discussion of the differences among estimators. I focus on general assessments of the candidates and approval related to economic conditions, since the economy was a major issue in the 1980 election. The results I report here are the equations that include the feeling thermometer difference scores, the assessment of the candidates' ability to solve economic problems, the general approval rating, and the assessment of Carter's handling of

[18]Note that in the model, Green and Palmquist specify the error terms as being constant over time within individuals, implying that the endogenous variables are not affected by shocks that vary by period. It makes more sense to specify the error terms as having (at least) two components: one that is constant within individuals over time and one that varies both across individuals and across periods.

Table 1 Estimates of Green and Palmquist's dynamic party identification equations

	Pooled IV		*OD*			*ODL*		
STF variable	$P_{t-1}{}^a$	$STF_t{}^a$	$P_{t-1}{}^a$	$STF_t{}^a$	s^b	$P_{t-1}{}^a$	$STF_t{}^a$	s^b
Feeling thermometer	.814	.006	−.128	−.006	1.221	.220	.014	8.866
difference scores	(.019)	(.001)	(.171)	(.007)	(.317)	(.104)	(.004)	(.031)
Solve economic	.832	.201	−.099	.004	.003	.166	.092	.664
problems	(.027)	(.094)	(.158)	(.097)	(.959)	(.104)	(.081)	(.882)
Approval rating	.858	.106	.026	.032	5.746	.203	.148	1.079
	(.019)	(.037)	(.184)	(.087)	(.017)	(.101)	(.072)	(.782)
Carter's handling of	.863	.080	−.081	.028	3.287	.230	.142	4.228
inflation	(.016)	(.039)	(.147)	(.058)	(.070)	(.092)	(.054)	(.238)

[a]Standard errors in parentheses.
[b]Sargan test statistic, with p values in parentheses.

inflation.[19] Green and Palmquist found statistically significant effects for these STF variables when estimating the equations period by period, but these effects disappeared with the (likely incorrect) correction for measurement error mentioned above. The results given by dynamic panel estimators are not directly comparable to those produced by estimating separate equations for each time period since the former will give us one set of parameter estimates for all periods. For purposes of comparison, I computed pooled instrumental variables estimates of the parameters pooling the data for all of the periods. I computed both the simple first-difference and the orthogonal deviations (OD) estimators. Since previous analyses have found that the parameter on lagged party identification is close to 1, I also computed the versions of these estimators that use additional moment conditions in levels because these estimators tend to perform better than the simpler estimators under these conditions. The results for the first-difference and OD estimators were nearly identical, so I report only the latter.

The results are reported in Table 1. I report the results in the same manner as Green and Palmquist: each row in the Table 1 gives the estimates of the coefficients from Eq. (35) for a particular STF variable. The pooled IV results are very similar to what they found for the equations where they did not correct for measurement error. The coefficients on the lagged dependent variables and the STF variables are approximately the same as those estimated by Green and Palmquist. From these results, we would conclude that there is a high degree of persistence in an individual's partisanship over time ($\gamma > .8$ in every equation), but short-term assessments of the candidates can move party identification. Yet readers should keep in mind that the inferences from this pooled IV estimator are not to be trusted.

Nor are all of the dynamic panel estimates to be trusted. In addition to the coefficients and standard errors for the lagged party identification and STF variables, I also report in Table 1 the χ^2 statistics and p values from the Sargan specification test.[20] The Sargan test rejects the

[19]The results for the replication of the remaining five regressions are available at the *Political Analysis* Web site. The dependent variable ranges from 0 for strong Democrats to 5 for strong Republicans. The thermometer differences range from +100 for "warmest" to −100 for "coldest" feelings toward Reagan vis-á-vis Carter. The assessment of the candidates' ability to solve economic problems ranges from −3 to +3, with +3 being the most favorable toward Reagan. The approval variables range from values of 1 for strong approval to values of 5 for strong disapproval.

[20]Even though the one-step estimates can be more reliable for making inferences in small samples, I report the

overidentifying restrictions for several items. For the simple OD estimator, the Sargan test rejects for the equations examining the general approval rating and the president's handling of inflation (at the .05 level for the former and at the .1 level for the latter). For the orthogonal deviations plus levels (ODL) estimator, the Sargan test rejects for the feeling thermometer difference scores equation. The properties of the OD and ODL estimators depend crucially on the validity of the sample moment conditions used to identify the model's parameters. If these conditions do not hold, then it is not appropriate to draw inferences based on these estimators.

It is appropriate, however, to make inferences on the remaining equations. The OD estimates for the feeling thermometer scores indicate that neither the lag of party identification nor the STF variables have effects that are statistically distinguishable from 0. Similar null results are given by both the OD and the ODL estimates for the equation that contains the assessment of the candidates' ability to solve economic problems. These estimates comport well with Green and Palmquist's original conclusion that party identification is extremely stable. Party identification according to these results appears to be simply a function of a time-invariant, individual-specific component and random noise.

However, we would draw somewhat different conclusions based on the estimates produced by the ODL estimator for the general approval rating and inflation items. The lagged party identification variable has coefficients that are bounded away from 0 for both of these items, but they are only about one-fourth the size of the pooled IV estimates of γ. The ODL estimates of the standard errors for γ are on average 30% smaller than those for the OD estimator, which is what we would expect from a more efficient estimator. The coefficients on the STF variables are bounded away from zero, although we do not see the same reduction in the standard errors as we saw with γ. The appropriate inference to draw from these results is that the more an individual disapproved of Carter's performance, the more strongly she identified with the Republican party.

Although the results are mixed, these replications show the importance of using specification checks to choose among available dynamic panel estimators. While the systems estimators are generally preferred, we should not use them unless they pass specification checks. The benefit of increased efficiency that these estimators provide comes at a cost of additional assumptions about the DGP that may not be satisfied. If those assumptions are not met, as the Sargan test indicates for the equation that includes the feeling thermometer item, then we should rely on other, possibly less efficient estimators. But we should not make inferences based on simpler estimators if specification checks reject them, as with the OD estimator for the general approval rating and the inflation items.

The substantive implication of this reanalysis is that party identification is less stable than Green and Palmquist's results indicate. They found no effects for STF in any of their equations. With a more appropriate estimator that passes specification checks, however, we find that some assessments of the president's performance do affect individuals' party identification, which provides support for those who argue that party identification is *not* an "unmoved mover."

7.2 *Reanalysis of Green and Yoon's Dynamic Party Identification Model*

In one of the first applications of dynamic panel models in political science, Green and Yoon (2002) employ these methods to try to understand the relationship between changes in party

two-step estimates because the differences between the one-step and the two-step estimates were miniscule.

identification at the macro level and changes at the micro level. They contend that using these methods enables us to determine whether or not the patterns observed in aggregate movements of partisanship over time are due to the individual-level heterogeneity posited by Box-Steffensmeier and Smith's (1996) fractionally integrated time-series models of macropartisanship. They employ the Anderson–Hsiao first-difference estimator to estimate the dynamic model

$$P_{i,t} = \gamma P_{i,t-1} + \alpha_i + u_{i,t} \tag{37}$$

in four separate panel surveys. They find that γ is not statistically distinguishable from 0 in any of the surveys and conclude that party identification at the micro level has no memory, in contrast to the model proposed by Box-Steffensmeier and Smith. That is, shocks to party identification die out immediately because they cannot affect identification in subsequent periods through lagged party identification. An individual's party identification at a given time is simply a function of her individual-specific mean (α_i) and a contemporaneous shock ($u_{i,t}$).

While their application of dynamic panel estimators to the analysis of party identification is innovative, Green and Yoon are perhaps too quick to dismiss the role that past shocks to party identification play in determining current party identification. As discussed in Section 4, the Anderson–Hsiao estimator has been found to be downwardly biased and can be very imprecise. Indeed, in their Monte Carlo simulations where they show that treating party identification as a linear continuous variable does not seriously impair the Anderson–Hsiao estimator, Green and Yoon find the downward bias with this estimator that previous studies have found. The Anderson–Hsiao estimator is also inefficient compared with other, more recently developed estimators. Both of these problems can lead to incorrect inferences about the effect of lagged party identification.

I reestimated the simple dynamic model in Eq. (37) for the 1992–1996 ANES panel using the more efficient estimators developed by Arellano and Bover (1995) discussed in Section 6.2. Following Green and Yoon's lead, I estimated the models both with and without time dummies. The results are reported in Table 2. The first thing to note is that the Sargan test rejects the overidentifying restrictions for three of the four models estimated. The equation with time dummies estimated by OD is the only one to pass both the Sargan test and the serial-correlation test. The results produced by the simple OD estimator lead to the same inferences as made by Green and Yoon.[21] That is, since the coefficient on the lag of party identification is not bounded away from 0, past shocks to party identification do not appear to have effects in current periods.

Readers may question the usefulness of dynamic panel methods, since in many instances, we reject the assumptions about the DGP that underlie the models we estimate. But dismissing these methods on such grounds is tantamount to killing a messenger who bears bad news. A more appropriate response is to consider the possible inadequacies of the empirical models we are trying to estimate. If the hypothesized model is appropriate for the DGP, then specification tests should confirm this. If they do not, then we should rethink our model specification.

[21] I also estimated the models using the first-difference estimators, which produced results that were very close to the orthogonal deviations estimator. Again, I report the two-step estimates of standard errors instead of the one-step estimates because the difference between them is so small that it does not affect our inferences.

Table 2 Estimates of a simple dynamic party identification equation for the 1992–1996 panel survey—Orthogonal deviations and orthogonal deviations plus levels estimators ($N = 500$)

	γ[a]	*1st-order serial correlation test*[b]	*Sargan test*[b]
	Orthogonal deviations		
No time dummies	−.228	−1.640	16.289
	(.173)	(.101)	(.000)
Time dummies	−.099	−3.046	3.945
	(.110)	(.002)	(.139)
	Orthogonal deviations plus levels		
No time dummies	.439	−6.17	25.55
	(.057)	(.000)	(.000)
Time dummies	.389	−6.78	25.75
	(.058)	(.000)	(.000)

[a]Standard errors in parentheses.
[b]p values in parentheses.

One possible adjustment is to use fewer moment conditions.[22] Some moment conditions do not hold in the presence of serial correlation and/or heteroskedasticity, which would lead the Sargan test to reject. Not using these moment conditions can give more reliable estimates. Results from estimating the specification using instruments dated $t - 3$ instead of $t - 2$ and earlier are reported in Table 3. The results for the OD estimates of the equation for time dummies are not robust. The OD results for both equations are dubious because Arellano and Bond's serial correlation test indicates that there is not first-order serial correlation in the first-differenced residuals even though they should follow an MA(1) process if they were serially uncorrelated to begin with. Unfortunately, the Sargan test is not available for the OD estimates because the model is just identified, so we need to rely on the serial correlation test, which indicates that the OD estimates are not reliable. For the ODL estimates, the model with period-specific effects passes both the serial correlation and the Sargan tests, while the model without these effects fails the Sargan tests. ODL with time dummies produces a coefficient on the lag of party identification of approximately .4 and gives standard errors that are almost 90% smaller than the standard errors for the comparable OD estimates. From this result, we would conclude that past shocks do affect current party identification through the lagged dependent variable.

To reach a more definitive conclusion about the nature of persistence in individual-level partisanship, I added the 1997 wave of the ANES panel. The 1997 wave includes about a quarter of the respondents from the 1992–1996 waves, which means that we have an unbalanced panel. Fortunately, respondents for the 1997 wave were randomly selected from the earlier waves, so no selection bias should be introduced.

Adding this wave to the analysis has several advantages. First, it gives us additional moment conditions, which should increase the precision of our estimates. Second, it enables us to perform a more thorough battery of tests for the moment conditions. Increasing the

[22]Another response, suggested by the replication of Green and Palmquist's analysis, is to include STF variables in the specification. Recall that the Sargan test did not reject dynamic models of party identification that included variables measuring presidential performance evaluations. I estimated a model that included the general approval rating, but the Sargan test still rejected the specification.

Table 3 Estimates of a simple dynamic party identification equation for the 1992–1996 panel survey—Orthogonal deviations and orthogonal deviations plus levels estimators with fewer instruments ($N = 500$)

	γ[a]	*1st-order serial correlation test*[b]	*Sargan test*[b]
	Orthogonal deviations		
No time dummies	−1.702	.349	*N.A.*
	(.632)	(.727)	
Time dummies	−.826	.240	*N.A.*
	(.626)	(.811)	
	Orthogonal deviations plus levels		
No time dummies	.471	−7.47	15.33
	(.071)	(.000)	(.000)
Time dummies	.431	−9.35	3.78
	(.077)	(.000)	(.151)

[a]Standard errors in parentheses.
[b]p values in parentheses.

number of periods to five enables us to do the complete test for both first- and second-order serial correlation discussed in Section 6.4. We can also perform a Sargan test for the OD estimator using instruments dated $t - 3$ and earlier because adding another wave provides more instruments and produces an overidentified model (recall that with $t = 4$ the model was just identified for the OD estimator). Finally, it is useful to demonstrate that the methods discussed in this article work on an unbalanced panel.

The results are reported in Table 4. The specification checks again show the OD estimates to be unreliable. The models with and without time dummies both fail the first-order serial correlation test, and the model without time dummies displays evidence of negative second-order serial correlation when it should display none. The ODL estimator for the equation with time dummies passes all of the specification tests to which it is subjected. It

Table 4 Estimates of a simple dynamic party identification equation for the 1992–1997 panel survey—Orthogonal deviations and orthogonal deviations plus levels estimators

	γ[a]	*1st-order serial correlation test*[b]	*2nd order serial correlation test*[b]	*Sargan test*[b]
		Orthogonal deviations		
No time dummies	−1.031	.840	−2.597	3.281
	(.356)	(.401)	(.009)	(.194)
Time dummies	−.287	−.426	−1.030	2.479
	(.422)	(.670)	(.303)	(.290)
		Orthogonal deviations plus levels		
No time dummies	.470	−8.166	.522	13.762
	(.071)	(.000)	(.602)	(.017)
Time dummies	.441	−9.638	.542	3.219
	(.069)	(.000)	(.588)	(.666)

Note. $N = 497$ for 1992 through 1996 waves. $N = 130$ for 1997 wave.
[a]Standard errors in parentheses.
[b]p values in parentheses.

shows evidence of serial correlation in the first-differenced errors, as it should, but does not display second-order serial correlation, as it should not. Furthermore, the Sargan test does not reject the overidentifying restrictions for this model. The ODL estimator again estimates the autoregressive parameter to be approximately .4 for the model with time dummies and gives standard errors that are much smaller than the simple OD estimators. This estimate of γ is bounded away from 0, indicating that past party identification affects current party identification.

The results produced by the ODL estimator for the dynamic party identification model with period-specific effects are the most robust of the various models and estimators examined. The estimated coefficients are approximately the same for the 1992–1996 panel and the 1992–1997 panel and pass the most specification tests. The ODL estimator also gives dramatic increases in efficiency, producing standard errors that are between 50 and 90% smaller than those produced by the OD estimator. Based on the ODL results, we would reach a conclusion about change in partisanship at the micro level substantively different from that reached by Green and Yoon. Past shocks to party identification appear to have persistent effects across periods in a way that is consistent with the Box-Steffensmeier and Smith model of macropartisanship. Party identification at the individual level appears to have memory, and past shocks can lead individuals to drift away from their mean level of partisanship over time, leading to the kinds of shifts in partisanship we see at the macro level.

8 Discussion

This article has discussed how to estimate dynamic models properly using panel data. Dynamic panel data can be a powerful resource for answering questions of interest in political science, but researchers need to employ methods that are more sophisticated than standard regression techniques if they are to realize the gains these data offer. Researchers in political science have largely failed to keep pace with the innovations in the literature on panel methods in the past two decades, despite Stimson's (1985) early warning about the complexities of analyzing this kind of data.

This article demonstrates, both theoretically and substantively, the costs of using inferior cross-sectional methods. In the replications of work on the dynamics of party identification, I found that dynamic panel estimators led to conclusions that were substantively different from those produced by cross-sectional methods. I also showed how differences can arise between different dynamic panel estimators and that there can be substantial gains to moving from simpler to more involved dynamic panel estimators that minimize bias and maximize efficiency. Estimators that employ a systems of equations framework proved to be more robust than simpler estimators and indicated that party identification is less stable than simpler estimators would lead us to believe. Although the computation of the systems estimators is somewhat more involved, they are highly recommended. But using the simpler dynamic panel estimators is certainly better than using cross-sectional methods. Whatever estimators are employed, it is crucial that researchers systematically apply specification checks to make sure that assumptions underlying the estimators hold before drawing inferences.

While this article has argued that political scientists need to worry more about the methods they employ when it comes to analyzing panel data, another implication of this article is that we need to be more concerned about how we collect panel data in the first place. The most reliable estimators for dynamic panel models require at least three time periods' worth of observations on cross-sectional units. Panel surveys that consist of only two waves, such as that analyzed by Finkel and Mueller (1998), are of severely limited usefulness for estimating dynamic models and can lead to very inaccurate inferences. Samples with at least four or five periods are necessary to perform thorough specification checks for some estimators. The potential gains, in terms of the reliability of estimators, from adding another wave to a two-wave panel survey justify the additional costs. Although the properties of these estimators are asymptotic in N (i.e., the number of cross-sectional units), we should consider collecting fewer cross-sectional units in favor of collecting additional waves. Yet since the costs of adding additional waves to a survey increase nonlinearly, the gains from adding more waves beyond three drop off dramatically, and it is not likely that we will recoup the costs of collecting more than four or five waves. While increasing the number of waves beyond two is more costly, it is worthwhile to incur the additional costs since the returns on what would already be a substantial investment will be scant if we cannot use appropriate methods to analyze the data.

The additional costs of data collection should force us to think harder about model specification, particularly in terms of including lagged dependent variables to model persistence. In many analyses, the inclusion of lags seems more an afterthought than a reasoned model specification decision firmly grounded in theory. Given the increased costs in terms of collecting data, researchers should be especially careful that the benefits of esimating a panel model with lagged dependent variables outweigh the costs.

References

Ahn, Seung C., and Peter Schmidt. 1995. “Efficient Estimation of Models for Dynamic Panel Data.” *Journal of Econometrics* 68:5–27.

Anderson, T. W., and Cheng Hsiao. 1981. “Estimation of Dynamic Models with Error Components.” *Journal of the American Statistical Association* 76:598–606.

Anderson, T. W., and Cheng Hsiao. 1982. “Formulation and Estimation of Dynamic Models Using Panel Data.” *Journal of Econometrics* 18:47–82.

Arellano, Manuel. 1989. “A Note on the Anderson-Hsiao Estimator for Panel Data.” *Economics Letters* 31:337–341.

Arellano, Manuel, and Stephen Bond. 1991. “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations.” *Review of Economic Studies* 58:277–297.

Arellano, Manuel, and Olympia Bover. 1995. “Another Look at the Instrumental Variable Estimation of Error Component Models.” *Journal of Econometrics* 68:29–51.

Arellano, Manuel, and Bo Honoré. 2001. “Panel Data Models: Some Recent Developments.” In *Handbook of Econometrics,* eds. J. J. Heckman and E. E. Leamer. Amsterdam: North-Holland (in press). (Available at http://www.cemfi.es/∼arellano/.)

Baltagi, Badi H. 1995. *Econometric Analysis of Panel Data*. New York: Wiley.

Beck, Nathaniel, and Jonathan N. Katz. 1996. “Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models.” *Political Analysis* 6:1–36.

Blundell, Richard, and Stephen Bond. 1998. “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models.” *Journal of Econometrics* 87:115–143.

Box-Steffensmeier, Janet, and Tse-min Lin. 1996. “A Dynamic Model of Campaign Spending in Congressional Elections.” *Political Analysis* 6:37–66.

Box-Steffensmeier, Janet, and Renée Smith. 1996. “The Dynamics of Aggregate Partisanship.” *American Political Science Review* 90:567–580.

Finkel, Steven E. 1995. *Causal Analysis with Panel Data.* Thousand Oaks, CA: Sage.

Finkel, Steven E., and Edward N. Mueller. 1998. “Rational Choice and the Dynamics of Political Action: Evaluating Alternative Models with Panel Data.” *American Political Science Review* 92:37–51.

Fiorina, Morris P. 1981. *Retrospective Voting in American National Elections.* New Haven, CT: Yale University Press.

Franklin, Charles H., and John E. Jackson. 1983. “The Dynamics of Party Identification.” *American Political Science Review* 77:957–973.

Green, Donald P., and Bradley Palmquist. 1990. “Of Artifacts and Partisan Instability.” *American Journal of Political Science* 34:872–902.

Green, Donald P., and David H. Yoon. 2002. "Reconciling Individual and Aggregate Evidence Concerning Partisan Stability: Applying Time-Series Models to Panel Survey Data" *Political Analysis* 10:1–24.

Greene, William H. 1997. *Econometric Analysis*, 3rd ed. New York: Macmillan.

Hansen, L. P. 1982. "Large Sample Properties of Generalized Method of Moment Estimators." *Econometrica* 50:1029–1054.

Himmelberg, Charles P., and Gregory Wawro. 1998. "A Dynamic Panel Analysis of Campaign Contributions in Elections for the U.S. House of Representatives." Paper presented at the Political Methodology Summer Conference and the Annual Meeting of the American Political Science Association.

Holtz-Eakin, Douglas, Whitney Newey, and Harvey S. Rosen. 1988. "Estimating Vector Autoregressions with Panel Data." *Econometrica* 56:1371–1395.

Hsiao, Cheng. 1986. *Analysis of Panel Data*. Cambridge: Cambridge University Press.

Jackson, John E. 1975. "Issues and Party Alignment." In *The Future of Political Parties,* eds. Louis Maisel and Paul M. Sacks. Beverly Hills, CA: Sage, pp. 101–123.

Jennings, M. Kent, and Gregory B. Markus. 1984. "Partisan Orientations over the Long Haul: Results From a Three–Wave Political Socialization Panel Study." *American Political Science Review* 78:1000–1018.

Kabashima, I., and Y. Ishio. 1998. "The Instability of Party Identification Among Eligible Japanese Voters: A Seven–Wave Panel Study, 1993–6." *Party Politics* 4:151–176.

Keane, Michael P., and David E. Runkle. 1992. "On the Estimation of Panel Data Models with Serial Correlation When Instruments Are Not Strictly Exogenous." *Journal of Business and Economic Statistics* 10:1–9.

Kiviet, Jan F. 1995. "On Bias, Inconsistency, and Efficiency of Various Estimators in Dynamic Panel Data Models." *Journal of Econometrics* 68:53–78.

Krasno, Jonathan S., Donald P. Green, and Jonathan A. Cowden. 1994. "The Dynamics of Campaign Fundraising in House Elections." *Journal of Politics* 56:459–474.

Markus, Gregory B. 1979a. *Analyzing Panel Data*. Beverly Hills, CA: Sage.

Markus, Gregory B. 1979b. "The Political Environment and the Dynamics of Public Attitudes: A Panel Study." *American Journal of Political Science* 23:338–359.

Markus, Gregory B. 1982. "Political Attitudes During an Election Year: A Report on the 1980 NES Panel Study." *American Political Science Review* 76:538–560.

Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica* 49:1417–1426.

Sargan, J. D. 1958. "The Estimation of Economic Relationships Using Instrumental Variables." *Econometrica* 26:393–415.

Stimson, James A. 1985. "Regression in Space and Time: A Statistical Essay." *American Journal of Political Science* 29:914–947.

Wooldridge, Jeffrey M. No date. "*Econometric Analysis of Cross Section and Panel Data*." Unpublished manuscript.

Ziliak, James P. 1997. "Efficient Estimation with Panel Data When Instruments Are Predetermined: An Empirical Comparison of Moment-Condition Estimators." *Journal of Business and Economic Statistics* 15:419–431.