

# Maximum Likelihood for Cross-lagged Panel Models with Fixed Effects

Socius: Sociological Research for a Dynamic World  
 Volume 3: 1–17  
 © The Author(s) 2017  
 Reprints and permissions:  
[sagepub.com/journalsPermissions.nav](http://sagepub.com/journalsPermissions.nav)  
 DOI: 10.1177/2378023117710578  
[srd.sagepub.com](http://srd.sagepub.com)



Paul D. Allison<sup>1</sup>, Richard Williams<sup>2</sup>, and Enrique Moral-Benito<sup>3</sup>

## Abstract

Panel data make it possible both to control for unobserved confounders and allow for lagged, reciprocal causation. Trying to do both at the same time, however, leads to serious estimation difficulties. In the econometric literature, these problems have been solved by using lagged instrumental variables together with the generalized method of moments (GMM). Here we show that the same problems can be solved by maximum likelihood (ML) estimation implemented with standard software packages for structural equation modeling (SEM). Monte Carlo simulations show that the ML-SEM method is less biased and more efficient than the GMM method under a wide range of conditions. ML-SEM also makes it possible to test and relax many of the constraints that are typically embodied in dynamic panel models.

## Keywords

panel data, dynamic panel model, fixed effects, cross-lagged model, generalized method of moments, GMM, Arellano-Bond, FIML, SEM, structural equation model, maximum likelihood, predetermined variable, sequentially exogenous variable, xtdpml, instrumental variable

## Introduction

Panel data have two big attractions for making causal inferences with non-experimental data:

- the ability to control for unobserved, time-invariant confounders
- the ability to model the direction of causal relationships.

Controlling for unobservables can be accomplished with fixed effects methods that are now well known and widely used (Allison 2005b, 2009; Firebaugh, Warner, and Massoglia 2013; Halaby 2004). For examining causal direction, the most popular approach has long been the cross-lagged panel model, originating with the two-wave, two-variable model proposed by Duncan (1969) and elaborated by many others (e.g., Finkel 1995; Hamaker, Kuiper, and Grasman 2015; Kenny and Judd 1996; Kessler and Greenberg 1981; Markus 1979; McArdle and Nesselroade 2014). In these models,  $x$  and  $y$  at time  $t$  affect both  $x$  and  $y$  at time  $t + 1$ .

Unfortunately, attempting to combine fixed effects models with cross-lagged panel models leads to serious estimation problems that are well known in the econometric literature. Economists typically refer to such models as

*dynamic* panel models because of the lagged effect of the dependent variable on itself. The estimation difficulties include error terms that are correlated with predictors, the so-called incidental parameters problem, and uncertainties about the treatment of initial conditions. For reviews of the extensive literature on dynamic panel data models, see Wooldridge (2010), Baltagi (2013), or Hsiao (2014).

The most popular econometric method for estimating dynamic panel models has long been the generalized method of moments (GMM) that relies on lagged variables as instruments. This method has been incorporated into several widely available software packages, including SAS, Stata, LIMDEP, RATS, and plm (an R package), usually under the name of Arellano-Bond (AB) estimators. While the AB approach provides consistent estimators of the coefficients, there is evidence that the estimators are not fully efficient, have considerable small-sample bias, and often perform poorly

<sup>1</sup>University of Pennsylvania, Philadelphia, PA, USA

<sup>2</sup>University of Notre Dame, Notre Dame, IN, USA

<sup>3</sup>Banco de España, Madrid, Madrid, Spain

## Corresponding Author:

Paul D. Allison, Department of Sociology, University of Pennsylvania, 3718 Locust Walk, Philadelphia, PA 19104-6299, USA.  
 Email: [allison@soc.upenn.edu](mailto:allison@soc.upenn.edu)



when the autoregressive parameter (the effect of a variable on itself at a later point in time) is near 1.0.

In recent years, econometricians have explored maximum likelihood (ML) estimation as a way to overcome some of the limitations of the GMM methodology. These efforts have culminated in the work of Moral-Benito (2013), who developed an ML method that effectively addresses the key problems of dynamic panel data models. Unfortunately, there is currently little software available to implement his method. In this article, we show that the model and method of Moral-Benito falls within the framework of linear structural equation models (SEM) and that it can therefore be estimated in a straightforward way with widely available SEM packages. Using simulated data, we show that the ML-SEM method outperforms the AB method with respect to bias and efficiency under most conditions. ML-SEM also has several other advantages over the AB method:

- Error variances can easily be allowed to vary with time.
- The unobserved, time-invariant factor can have different effects at different times.
- Missing values on predictors can easily be handled by full information maximum likelihood (FIML).
- Many goodness-of-fit measures are available to assess the over-identifying restrictions of the model.
- There is no need to choose among many possible instrumental variables.
- Latent variables with multiple indicators can be incorporated into the model.
- Time-invariant variables can be included in the model.

ML-SEM does have a few downsides, however:

- Programming can be complex and tedious (although this problem has essentially been solved with the new `xtdpml` command for Stata).
- Convergence failures sometimes occur.
- Computation may be noticeably slower than with AB, especially when using FIML to handle unbalanced or other forms of missing data.

In this article, we:

- explore the relationship between the dynamic panel data models of econometrics and the cross-lagged panel models used in other social sciences,
- review GMM estimation of dynamic panel data models and examine its limitations,
- review the development of ML methods for dynamic panel data models,
- show how the ML method of Moral-Benito can be implemented in the SEM framework,
- present an empirical example,

- present results from a Monte Carlo study comparing the AB method and the ML-SEM method,
- conclude with a discussion.

## Cross-lagged Panel Models versus Dynamic Panel Data Models

### Cross-lagged Panel Models

We begin with a cross-lagged panel model that is specified in a way that facilitates comparisons with the dynamic panel models of econometrics. The data consist of a sample of  $N$  individuals, each of whom is observed at  $T$  points in time ( $t = 1, \dots, T$ ). Thus, the data set is “balanced,” having the same number of observations for each individual. Although the methods to be considered can be extended to unbalanced data, the initial development is simpler if we exclude that possibility. We also presume that the number of time points is substantially smaller than the number of individuals.

At each time point, we observe two quantitative variables,  $x_{it}$  and  $y_{it}$ , and we want to allow for the possibility that they have a lagged, reciprocal causal relationship. We may also observe a column vector of control variables  $w_{it}$  that varies over both individuals and time (possibly including lagged values) and another column vector of control variables  $z_i$  that varies over individuals but not over time.

Consider the following equation for  $y$ , with  $i = 1, \dots, N$  and  $t = 2, \dots, T$ :

$$y_{it} = \mu_t + \beta_1 x_{it-1} + \beta_2 y_{it-1} + \delta_1 w_{it} + \gamma_1 z_i + \alpha_i + \varepsilon_{it}, \quad (1)$$

where  $\mu_t$  is an intercept that varies with time,  $\beta_1$  and  $\beta_2$  are scalar coefficients,  $\delta_1$  and  $\gamma_1$  are row vectors of coefficients,  $\varepsilon_{it}$  is a random disturbance, and  $\alpha_i$  represents the combined effects on  $y$  of all unmeasured variables that are both constant over time and have constant effects. The lags for  $x$  and  $y$  are shown here as lags of one time unit, but the lags could be greater and could be different for each variable.

We also specify an analogous equation for  $x$ :

$$x_{it} = \tau_t + \beta_3 x_{it-1} + \beta_4 y_{it-1} + \delta_2 w_{it} + \gamma_2 z_i + \eta_i + \upsilon_{it} \quad (2)$$

where  $\tau_t$  is an intercept that varies with time,  $\beta_3$ , and  $\beta_4$  are scalar coefficients,  $\delta_2$  and  $\gamma_2$  are row vectors of coefficients,  $\upsilon_{it}$  is a random disturbance, and  $\eta_i$  is a set of individual effects analogous to  $\alpha_i$  in Equation 1. Equations 1 and 2 do *not* allow for simultaneous causation, which would require problematic assumptions in order to estimate and interpret the causal effects.

These two equations differ from the classic cross-lagged panel model in two ways: (1) by the introduction of the unobserved individual effects,  $\alpha_i$  and  $\eta_i$  and (2) by the presumption that the coefficients for all variables are constant over time. The constancy assumption can easily be relaxed, but we retain it now for simplicity.

The individual effects  $\alpha_i$  and  $\eta_i$  may be specified either as random variables or sets of fixed parameters. Outside of economics, they are usually treated as random variables that are independent of all other exogenous variables (e.g., Hamaker et al. 2015).

More needs to be said about the random disturbance terms,  $\varepsilon_{it}$  and  $\nu_{it}$ . We assume that they are independent of each other (both within and between time points) and normally distributed with means of 0 and constant variance (at least across individuals, although we will allow for variances that change over time). We also assume that  $w_{it}$  and  $z_i$  are strictly exogenous, meaning that for any  $t$  and any  $s$ ,  $w_{it}$  and  $z_i$  are independent of  $\varepsilon_{is}$  and  $\nu_{is}$ . With respect to  $x$  and  $y$ , we cannot assume strict exogeneity because both variables appear as dependent variables. In fact, Equations 1 and 2 together imply that  $\varepsilon_{it}$  and  $\nu_{it}$  are correlated with all *future* values of  $x$  and  $y$ .

### Dynamic panel data models

The basic dynamic panel data model found in the econometric literature is essentially the same as Equation 1 but with a few changes in meaning:

- $x$  is typically a vector rather than a scalar,
- $x$  is usually not lagged,
- $\alpha_i$  is treated as a set of fixed constants rather than as a set of random variables.

The first two differences are relatively unimportant, but the third is crucial. Treating  $\alpha$  as a set of fixed constants (“fixed effects”) is equivalent to allowing for unrestricted correlations between  $\alpha$  and all the time-varying predictors, both  $x$  and  $w$ . Allowing these correlations supports a claim that these models control for all time-invariant confounders, either observed or unobserved.

As in the cross-lagged panel model,  $w_{it}$  and  $z_i$  are assumed to be strictly exogenous. But  $x_{it}$  is assumed to be *predetermined* (Arellano 2003) or equivalently, *sequentially exogenous* (Wooldridge 2010). This means that for all  $s > t$ ,  $x_{it}$  is independent of  $\varepsilon_{is}$ . That is, the  $x$  variables are independent of all future values of  $\varepsilon$  but may be correlated with past values of  $\varepsilon$ .

The assumption that  $x_{it}$  is predetermined allows for the *existence* of Equation 2, but it also allows for a much wider range of possibilities. In particular, Equation 2 could be modified to have multiple lags of  $y$ , or it could be a nonlinear equation. For example, if  $x$  is dichotomous (as in the example in we present later), a logistic regression equation could substitute for Equation 2.

It should now be fairly clear that the cross-lagged panel model can be regarded as a special case of the dynamic panel data model. You can get from the latter to the former by (a) lagging  $x$  and reducing it from a vector to a scalar, (b) converting fixed effects into random effects, and (c)

imposing the structure of Equation 2 on the dependence of  $x$  on prior  $y$ s.

We agree with economists that the less restricted model is a better way to go. The ability to control for unmeasured confounders is a huge advantage in making claims of causality. And not having to specify the functional form of the dependence of  $x$  on  $y$  both simplifies the estimation problem and reduces the danger of misspecification. If you are interested in the dependence of  $x$  on  $y$ , you can always specify a second dynamic panel data model for  $y$  and estimate that separately.

On the other hand, we believe that those working in the cross-lagged panel tradition have chosen the better approach to estimation. Except in the simple case of two-wave data, most cross-lagged models are formulated as structural equation models and estimated by maximum likelihood using standard SEM packages. Economists have taken a rather different path, one that has led to possibly inferior estimators and a few dead ends.

### GMM Estimation

Estimation of the dynamic panel data model represented by Equation 1 is not straightforward for reasons that are well known in the econometric literature. First, the presence of the lagged dependent variable as a predictor implies that conventional fixed effects methods will yield biased estimates of the  $\beta$  coefficients (Arellano 2003). Second, even if the lagged dependent variable is excluded, the fact that the  $x$ s are merely predetermined, not strictly exogenous, implies that conventional fixed effects methods will yield biased estimates of the coefficients whenever  $T > 3$  (Wooldridge 2010).

Until recently, econometricians focused almost exclusively on instrumental variable methods. The dominant method is usually attributed to Arellano and Bond (1991), although there were important earlier precedents (Anderson and Hsiao 1981; Holtz-Eaken, Newey, and Rosen 1988). To remove the fixed effects ( $\alpha$ ) from the equations, the model is reformulated in terms of first differences:  $\Delta y_{it} = y_{it} - y_{it-1}$ ,  $\Delta x_{it} = x_{it} - x_{it-1}$  and  $\Delta w_{it} = w_{it} - w_{it-1}$ . Note that first differencing not only removes  $\alpha$  from the equation but also  $z$ , the vector of time-invariant predictors. Lagged difference scores for  $y$ ,  $x$ , and  $w$  are then used as instrumental variables for  $\Delta y$  and  $\Delta x$ , and the resulting system of equations is estimated by the generalized method of moments (GMM).

Models with instrumental variables imply multiple restrictions on the moments in the data, specifically, that covariances between instruments and certain error terms are 0. GMM chooses parameter estimates that minimize the corresponding observed moments. Since there are multiple moment restrictions, the method requires a weight matrix that optimally combines the observed moments into a unidimensional criterion. In many settings, GMM requires iteration to minimize that criterion. But for the moments used in the AB method, minimization is accomplished by solving a linear equation that requires no iteration.

AB estimators come in two forms, the one-step method (the usual default in software) and the two-step method. The latter uses results from the first step to reconstruct the weight matrix, but there is little evidence that its performance is any better than the one-step method (Judson and Owen 1999). Another extension is the GMM system estimator of Blundell and Bond (1998), which uses both levels and first differences of the lagged variables as instruments. This method produces more efficient estimators but at the cost of making the rather unrealistic assumption that the initial observations reflect stationarity of the process generating the data.

AB estimators are believed to suffer from three problems:

1. *Small sample bias.* AB estimators are consistent, that is, they converge in probability to the true values as sample size increases. However, simulation evidence indicates that they are prone to bias in small samples, especially when the autoregressive parameter (the effect of  $y_{t-1}$  on  $y_t$ ) is near 1.0 (Blundell and Bond 1998; Kiviet, Pleus, and Poldermans 2014).
2. *Inefficiency.* AB estimators do not make use of all the moment restrictions implied by the model. As a consequence, they are not fully efficient. Ahn and Schmidt (1995) proposed an efficient GMM estimator that *does* make use of all restrictions, but its non-linear form makes it more difficult to implement. In any case, their method is not generally available in commercial software packages.
3. *Uncertainty about the choice of instruments.* Anyone who has attempted to use the AB method knows that there are many choices to be made regarding what variables to use as instruments and whether they are to be entered as levels or first differences. In principle, it would make sense to use all possible instruments that are consistent with the model. But available evidence suggests that “too many” instruments can be just as bad as too few, leading to additional small-sample bias (Roodman 2009). This problem is especially acute when  $T$  is large, in which case the number of potential instruments is also large.

## ML Estimation of Dynamic Panel Models

In an effort to solve some of these problems, there has been quite a bit of work in the econometric literature on ML estimation of dynamic panel models. However, that work has yet to have a significant impact on empirical applications. Bhargava and Sargan (1983) considered ML estimation of dynamic panel models, but they assumed that the time-varying predictors were uncorrelated with the fixed effects, which is precisely what we do *not* want do. The seminal paper of Hsiao, Pesaran, and Tahmiscioglu (2002) proposed a ML

estimator that *does* allow the predictors in each equation to be correlated with the fixed effects. In their view, accomplishing this was difficult for two reasons:

There are two issues involved in the estimation of the fixed effects dynamic panel data model when the time-series dimension is short. One is the introduction of individual-specific effects that increase with the number of observations in the cross-section dimension. The other is the initial value problem. Both lead to the violation of the conventional regularity conditions for the MLE of the structural parameters to be consistent because of the presence of “incidental parameters.” (P. 139)

The issue of incidental parameters is a well-known problem in maximum likelihood estimation. It’s what happens when the number of parameters increases with the sample size, thereby invalidating the usual asymptotic arguments for consistency and efficiency of ML estimators (Nickell 1981).

Hsiao et al. (2002) dealt with the incidental parameters problem by using the same device as Arellano and Bond (1991)—taking first differences of the time-varying variables, thereby eliminating the individual-specific fixed effects. The likelihood was then formulated in terms of the difference scores. To deal with the initial value problem, they introduced assumptions of stationarity for the generation of the initial values from some prior, unobserved process, assumptions that they admitted may be “controversial.” They also presented simulation evidence indicating that the performance of their ML estimator was somewhat better than that of several different GMM estimators.

Although the use of first differences solves the incidental parameters problem for the fixed effects, it greatly complicates the subsequent development of the method. Moreover, Han and Phillips (2013) argued that the first-difference likelihood is not a true likelihood function and that consequently, it may behave in pathological ways, especially when the autoregressive coefficients have values near 1.0.

For many years, there was no readily available software to implement the ML method of Hsiao et al. (2002). However, Grassetti (2011) showed that implementation is possible with conventional *random effects* software by working with variables that are differences from initial values instead of differences between adjacent time points. Recently, Kripfganz (2016) introduced a Stata command, `xtdpdqml`, that implements both the method of Hsiao et al. and the “random effects” model of Bhargava and Sargan (1983). However, Kripfganz also pointed out that that the method of Hsiao et al. does *not* yield consistent estimators for models with predetermined variables.

In contrast to Hsiao et al. (2002), Moral-Benito (2013) showed that parameters in Equations 1 or 2 can be directly estimated by maximum likelihood without first differencing and without any assumptions about initial conditions. The key insight is that  $\alpha_i$  and  $\eta_i$  do not have to be treated as fixed parameters. As pointed out long ago by Mundlak (1978) and



further elaborated by Chamberlain (1982, 1984), the fixed effects model is equivalent to a random effects model that allows for unrestricted correlations between the individual-specific effects ( $\alpha_i$  and  $\eta_i$ ) and the time-varying predictors. Once that approach is adopted, there is no longer any need to impose arbitrary assumptions on the initial observations,  $y_1$  and  $x_1$ . They can be treated as strictly exogenous, which is entirely appropriate given the lack of knowledge about what precedes those observations.

## ML Estimation Via SEM

In this section, we show how Moral-Benito's (2013) method can be implemented with SEM software. The essential features of the ML-SEM method for cross-lagged panel models with fixed effects were previously described by Allison (2000, 2005a, 2005b, 2009), but his approach was largely pragmatic and computational. Moral-Benito provided a rigorous theoretical foundation for this method.

The justification for using SEM software rests on the fact that Equations 1 and 2 are a special case of the linear structural equation model proposed by Jöreskog (1978) and generalized by Bentler and Weeks (1980). In its most general form, the model may be compactly specified as

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{B}\mathbf{y} + \boldsymbol{\Gamma}\mathbf{x}, \quad (3)$$

where  $\mathbf{y}$  is a  $p \times 1$  vector of endogenous variables that may be either observed or latent;  $\mathbf{x}$  is a  $k \times 1$  vector of exogenous variables that, again, may be either observed or latent (including any disturbance terms in the model);  $\boldsymbol{\mu}$  is a vector of intercepts, and  $\mathbf{B}$  and  $\boldsymbol{\Gamma}$  are matrices of coefficients. The endogenous vector  $\mathbf{y}$  and any latent variables in  $\mathbf{x}$  are assumed to have a multivariate normal distribution conditional on the observed exogenous variables. The  $\mathbf{B}$  matrix has zeros on the main diagonal, and both  $\mathbf{B}$  and  $\boldsymbol{\Gamma}$  may have many additional restrictions. Most commonly, these restrictions take the form of setting certain parameters equal to 0, but there may also be equality restrictions. The remaining parameter  $\boldsymbol{\Theta}$  is the variance matrix for  $\mathbf{x}$ , which usually has many elements set to 0.

There are several widely available software packages that will estimate any special case of this model via maximum likelihood. These include LISREL, EQS, Amos, Mplus, PROC CALIS (in SAS), sem (in Stata), lavaan (for R), and OpenMx (for R). Remarkably, the earliest version of LISREL, introduced in 1973, could probably have estimated the dynamic panel models considered here, albeit with considerably more programming effort than with contemporary packages.

How does this model relate to the cross-lagged Equations 1 and 2? Although Equations 1 and 2 can be estimated simultaneously, we follow the econometric tradition of focusing only on Equation 1 while allowing Equation 2 to determine certain constraints (or lack thereof) on  $\boldsymbol{\Theta}$ , the variance matrix for the exogenous variables.

Equation 1 is a special case of Equation 3 in the following sense. Without loss of generality, we treat  $w_{it}$  and  $z_i$  as scalars rather than vectors. We then have,  $\mathbf{y}' = (y_{i2}, \dots, y_{iT})$ ,  $\mathbf{x}' = (\alpha_i, z_i, y_{i1}, x_{i1}, \dots, x_{i(T-1)}, w_{i2}, \dots, w_{iT}, \varepsilon_{i2}, \dots, \varepsilon_{iT})$  and  $\boldsymbol{\mu}' = (\mu_2, \dots, \mu_T)$ . For  $\boldsymbol{\Gamma}$ , we have

$$\boldsymbol{\Gamma} = \begin{bmatrix} 1 & \gamma_1 & \beta_2 & \beta_1 & 0 & \dots & \delta_1 & 0 & \dots & 1 & 0 & \dots \\ 1 & \gamma_1 & 0 & 0 & \beta_1 & \dots & 0 & \delta_1 & \dots & 0 & 1 & \dots \\ 1 & \gamma_1 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots \\ 1 & \gamma_1 & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 & 0 & \dots \end{bmatrix},$$

and for  $\mathbf{B}$ ,

$$\mathbf{B} = \begin{bmatrix} 0 & 0 & 0 & \dots \\ \beta_2 & 0 & 0 & \dots \\ 0 & \beta_2 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

For  $\boldsymbol{\Theta}$ , the following covariances are set to 0:

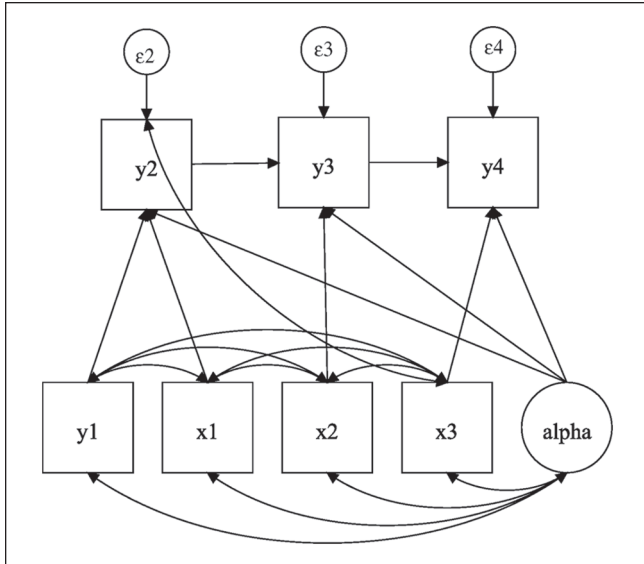
- $\alpha$  with  $z$
- $\alpha$  with all  $\varepsilon$
- $z$  with all  $\varepsilon$
- all  $w$  with all  $\varepsilon$
- all  $\varepsilon$  with each other
- $x_{it}$  with  $\varepsilon_{is}$  whenever  $s \geq t$ .

All other elements of  $\boldsymbol{\Theta}$  are left unrestricted. Note that  $\alpha$  is allowed to correlate with both  $w$  and  $x$ . And  $x$  is allowed to correlate with all *prior* realizations of  $\varepsilon$ , as a consequence of Equation 2. The restriction that  $\text{cov}(\alpha, z) = 0$ , while perhaps undesirable, is essential for identification. That is, we must assume that the fixed effects are uncorrelated with any time-invariant variables.<sup>1</sup>

Figure 1 displays a path diagram of this model for the case in which  $T = 4$ , with no  $w$  variables and no  $z$  variables.<sup>2</sup> That is, we only have the endogenous  $y$  variables and the predetermined  $x$  variables. Notice that all the  $x$  variables are allowed to freely correlate with each other, as well as with  $y_1$ , which is treated like any other exogenous variable. Similarly, the latent variable alpha (enclosed in a circle) is allowed to correlate with all the exogenous variables, including  $y_1$ . Alpha affects each  $y$  variable (with a coefficient of 1, not shown). The coefficients for the effects of the  $x$ s on the  $y$ s are constrained to be the same at all three time points, but this constraint can be easily relaxed.

<sup>1</sup>For an alternative parameterization and a derivation of the likelihood function, see Moral-Benito, Allison, and Williams (2017).

<sup>2</sup>The path diagrams in Figures 1 and 3 were produced by Mplus, version 7.4.



**Figure 1.** Path diagram for 4-period dynamic panel model.

What makes  $x$  predetermined in this diagram is the correlation between  $x_3$  and  $\varepsilon_2$ . If this correlation were omitted,  $x$  would be strictly exogenous rather than predetermined. Again, the rule is that for any predetermined variable,  $x$  at time  $t$  is allowed to correlate with the error term for  $y$  at any prior time point.

How do the assumptions of ML-SEM differ from those of AB? ML-SEM makes stronger assumptions in three respects. First, and most importantly, ML-SEM assumes multivariate normality for all endogenous variables while AB makes no distributional assumptions. However, ML-SEM produces consistent estimators even when the normality assumption is violated (Moral-Benito 2013). And if there is concern about normality, robust standard errors can be used for constructing confidence intervals and hypothesis tests. Second, to identify the effects of time-invariant variables, we introduced the assumption that  $\text{cov}(\varepsilon, z) = 0$ . But if you have any reason to doubt that assumption, you can just exclude time-invariant variables from the model. They will still be controlled as part of the  $\alpha$  term. Lastly, ML-SEM makes use of the moment restrictions implied by the assumption that there is no serial correlation in the error terms in Equation 1. Although the use of these restrictions was recommended by Ahn and Schmidt (1995) to improve efficiency, they have generally not been incorporated into AB estimation because they imply nonlinear estimating equations.

On the other hand, ML-SEM makes it possible to relax many assumptions that are built into AB. Most notably, the default in ML-SEM is to allow for an unrestricted effect of time itself and for different error variances at each time point. It is also possible to allow  $\alpha$ , the latent variable for the individual effects, to have different coefficients at different time points.

## Empirical Example

As an example of how to implement ML-SEM for dynamic panel models, we reanalyze data described by Cornwell and Rupert (1988) for 595 household heads who reported a non-zero wage in each of 7 years from 1976 to 1982. For purposes of illustration, we use only the following variables from that data set:

$y$  = WKS = number of weeks employed in each year

$x$  = UNION = 1 if wage set by union contract, else 0, in each year

$w$  = LWAGE =  $\ln(\text{wage})$  in each year

$z$  = ED = years of education in 1976.

Let's suppose that our principal goal is to estimate the effect of UNION at time  $t$  on WKS at time  $t + 1$ . However, we also want to allow for the possibility that UNION is itself affected by earlier values of WKS. That is, we want to treat UNION as a predetermined variable. We also want to control for LWAGE, a time-varying variable, and we want to estimate the effect of ED, a time-invariant variable, on WKS. If we suspected that LWAGE was affected by earlier values of WKS, we could also treat it as predetermined. But for simplicity we will assume that it is strictly exogenous.

Our goal, then, is to estimate Equation 1, reproduced here

$$y_{it} = \mu_t + \beta_1 x_{i,t-1} + \beta_2 y_{i,t-1} + \delta_1 w_{it} + \gamma_1 z_i + \alpha_i + \varepsilon_{it}, \quad (1)$$

with  $x$  treated as predetermined and  $w$  and  $z$  treated as strictly exogenous. By specifying UNION as predetermined, we allow for the possibility that number of weeks worked at time  $t$  could affect union status at time  $t + 1$  or, indeed, at any future time. However, we don't have to specify the functional form of that relationship. Since UNION is dichotomous, a logistic regression model might be a natural first choice for this dependence. But we don't have to make that choice because the model is agnostic regarding the dependence of  $x$  on  $y$ .

Note that although we have lagged the effect of  $x$  in Equation 1 to be consistent with the cross-lagged panel approach, the ML-SEM method does not require  $x$  to be lagged. We could have written  $x_t$  rather than  $x_{t-1}$ , and the estimation would have been equally straightforward. On the other hand, if we choose not to lag  $x$ , it would then be problematic to estimate another model in which  $x$  depends on unlagged  $y$ . That would imply a model with simultaneous, reciprocal causation, which would not be identified without additional instrumental variables.

We use PROC CALIS in SAS to illustrate the estimation of Equation 1 because CALIS has a syntax and default settings that are particularly well suited to dynamic panel

```

1  proc calis data=my.wagewide;
2  path
3  wks2 <- wks1 union1 lwage1 ed alpha = a b c d 1,
4  wks3 <- wks2 union2 lwage2 ed alpha = a b c d 1,
5  wks4 <- wks3 union3 lwage3 ed alpha = a b c d 1,
6  wks5 <- wks4 union4 lwage4 ed alpha = a b c d 1,
7  wks6 <- wks5 union5 lwage5 ed alpha = a b c d 1,
8  wks7 <- wks6 union6 lwage6 ed alpha = a b c d 1,
9  alpha <-> ed = 0,
10 wks2 <-> union3 union4 union5 union6,
11 wks3 <-> union4 union5 union6,
12 wks4 <-> union5 union6,
13 wks5 <-> union6,
14 <-> wks2 wks3 wks4 wks5 wks6 wks7 = v v v v v v;
15 run;

```

**Figure 2.** SAS program for estimating dynamic panel model with fixed effects.

models. Like most SEM packages, CALIS requires that the data be in the “wide form” rather than the “long form.”<sup>3</sup> For our example, the wide-form data set has one record per person, with seven variables corresponding to each conceptual variable at the seven time points. Thus, we have WKS1, WKS2, . . . , WKS7, and LWAGE1, LWAGE2, . . . , LWAGE7, and so on. Of course, there is only one variable for ED, which did not vary over time. In contrast, most software packages for the analysis of panel data (including those for the AB method) expect the data to be in the long form, with separate records for each individual at each point in time.

Because the setup for this kind of model will be unfamiliar to most readers, it’s worth examining it in some detail. Figure 2 shows the CALIS program for estimating the model. (Equivalent code for Stata, Mplus, and lavaan can be found in Appendix B). Line 1 invokes the CALIS procedure for the data set MY.WAGEWIDE. Line 2 begins the PATH statement, which continues until the end of Line 13. Lines 3 through 8 specify an equation for each of the six time points. Note that there is no equation specified for WKS1 because at Time 1 we do not observe the lagged values of the predictor variables.

The variable ALPHA refers to the “fixed effects” variable  $\alpha_i$  that is common to all equations. When CALIS encounters a variable name like ALPHA that is not on the input data set, it presumes that the name refers to a latent variable. After the equals sign on each line, there is a list of coefficient names or values corresponding to the predictor variables in the equation. Because the coefficient names are the same for each equation, the corresponding coefficients are also constrained to be the same. The coefficient of ALPHA is constrained to have a value of 1.0 in each equation, consistent with Equation 1.

<sup>3</sup>Mplus can analyze “long form” data using a multilevel add-on, but its multilevel mode is currently not suitable for the dynamic panel models considered here. However, the forthcoming version 8 of Mplus may allow for dynamic panel analysis in long form.

By default in PROC CALIS, latent exogenous variables like ALPHA are allowed to covary with all observed exogenous variables, including WKS1 and all the LWAGE and UNION variables. This is exactly what we want to achieve for ALPHA to truly behave as a set of fixed effects (Allison and Bollen 1997; Bollen and Brand 2010; Teachman et al. 2001). However, because ED does not vary over time, the correlation between ALPHA and ED is not identified, so it is constrained to be 0 in Line 9. Lines 10 through 13 allow the error term  $\varepsilon$  in each equation to be correlated with *future* values of UNION. This is the key device that allows UNION to be a predetermined (sequentially exogenous) variable.<sup>4</sup>

By default, CALIS allows the intercept to differ for each equation, which is equivalent to treating time as a set of dummy variables. It is easy to constrain the intercepts to be the same if desired. With a little more difficulty, one can impose constraints that correspond to a linear or quadratic effect of time.

Also by default, the error variances are allowed to differ across equations, which is not the case for most AB software. Line 14 constrains the error variances to be the same for each equation to produce results that can be directly compared with AB estimates. This line assigns names to the error variances for each variable. Because they are given the same name (v), the corresponding parameter estimates are constrained to be the same. In most applications, however, it is probably better to leave the variances unconstrained.

Figure 3 is a path diagram for the model. Here are a few key points to note about this diagram:

- There is no correlation between ALPHA and ED.
- All other observed exogenous variables are freely correlated with one another.
- All error terms are correlated with future values for UNION. This implicitly allows for the possibility that  $y$  affects future  $x$ , with no restrictions on that relationship.

Table 1 displays the numerical results in the first four columns. Not surprisingly, there is a highly significant effect of  $WKS(t-1)$  on  $WKS(t)$ , although the magnitude of the effect is not large. There is also a significant negative effect (at the .05 level) of  $UNION(t-1)$  on  $WKS(t)$  and a not quite significant negative effect of ED on  $WKS(t)$ . By constraint, these coefficient estimates are the same for all six equations.

This is only a small portion of the output from PROC CALIS. The full output also contains estimates of the variances and covariances for all the exogenous variables, including ALPHA and the error terms for each equation. As with all SEM software, there is also a likelihood ratio chi-square

<sup>4</sup>An equivalent method is to specify five additional regressions for UNION2 through UNION6 as dependent variables, with predictor variables that include all values of LWAGE, prior values of WKS, prior values of UNION, and ALPHA.

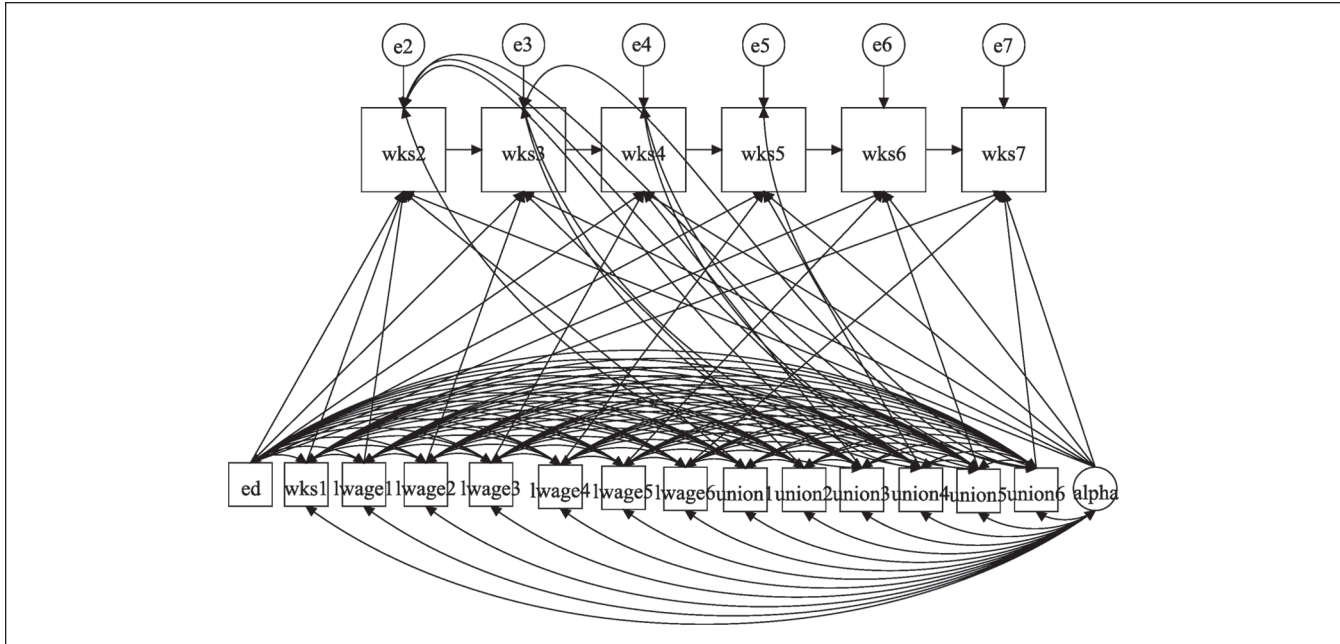


Figure 3. Path diagram for empirical example.

Table 1. Alternative Estimates for Dynamic Model with Fixed and Lagged Effects.

Predictor	Maximum Likelihood Structural Equation Modeling			Arellano-Bond		
	Estimate	SE	z	Estimate	SE	z
wks( $t - 1$ )	.188	.020	9.59	.163	.039	4.18
lwage( $t - 1$ )	.588	.488	1.20	-1.276	.462	-2.76
union( $t - 1$ )	-1.206	.522	-2.31	-1.175	.513	-2.29
ed	-.107	.056	-1.89			

statistic comparing the fitted model with a “saturated” or “just-identified” model that perfectly reproduces the covariance matrix for all the variables. For this example, it has a value of 138.48 with 76 degrees of freedom, yielding a  $p$  value less than .0001. The 76 degrees of freedom correspond to 76 over-identifying restrictions on the covariance matrix of the observed variables that are implied by the model.

Because this is a goodness-of-fit statistic, higher  $p$  values indicate a better fitting model. So by conventional significance standards, the model does not fit the data. However, the consensus in the SEM literature is that with large sample sizes, it may be hard to find any reasonably parsimonious model that yields a  $p$  value greater than .05. There are numerous alternative measures of fit that are relatively insensitive to sample size, and many of these are reported by PROC CALIS. For example, Bentler’s Comparative Fit Index is .995 while Bentler and Bonnet’s Non-Normed

Index (also known as the Tucker-Lewis Index) is .987.<sup>5</sup> Values near 1.0 are desirable, so these measures suggest a very good fit to the data. One of the most popular measures of fit is the root mean square error of approximation (RMSEA). For this example, it has a value of .037. Anything less than .05 is considered to be a good fit.

For comparison, we also estimated the same model using the standard AB method, as implemented with the Stata command `xtdpd`. The last three columns of Table 1 display the results. Note that this method—because it is based on difference scores—cannot produce any estimates for the effect of ED, which does not change over time.<sup>6</sup> The lagged effects of UNION and WKS (on itself) are similar to the estimates produced by PROC CALIS. However, the coefficient for the lagged effect of LWAGE is dramatically different from the ML estimate. This naturally raises the question of which method performs better, in general.

### Monte Carlo Study

To evaluate the performance of the ML-SEM method and compare it with the AB method, we generated observations from the following cross-lagged panel model:

<sup>5</sup>Stata, Mplus, and lavaan report somewhat lower values of these measures because they define the baseline model in a different way.

<sup>6</sup>The `xtdpd` model also included dummy variables for time to ensure comparability with the maximum likelihood structural equation modeling (ML-SEM) estimates.



**Table 2.** Parameter Values for Monte Carlo Simulation.

N	50, 100, <b>400</b> , 1,600
T	4, 5, 7, 10
$\rho$	0, .25, .50, <b>.75</b> , .90
$\beta_1$	0, <b>.25</b> , .50, .75, 1.00
$\beta_2$	0, .25, .50, <b>.75</b> , 1.00, 1.25
$\beta_4$	<b>-.25</b> , 0, .25
g	.75, <b>1.00</b> , 1.50, 2.00
c	.50, <b>1.00</b> , 1.50, 2.00

Note: The numbers in bold are the values for the baseline model.

$$y_{it} = c\alpha_i + gu_{it}$$

$$x_{it} = c\eta_i + gv_{it}$$

$$y_{it} = \beta_1 x_{it-1} + \beta_2 y_{it-1} + c\alpha_i + gu_{it}$$

$$x_{it} = \beta_3 x_{it-1} + \beta_4 y_{it-1} + c\eta_i + gv_{it}$$

for  $i = 1, \dots, N$  and  $t = 2, \dots, T$ . The time-invariant “fixed effects,”  $\alpha_i$  and  $\eta_i$ , were generated as bivariate standard normal variates with correlation  $\rho$ . The time-specific disturbances  $u_{it}$  and  $v_{it}$  were each standard normal and independent of all other exogenous variables. Parameters and data structures were varied as shown in Table 2.

The numbers in bold are the values for the *baseline* model. Each parameter was varied in turn while keeping all others at their baseline values. For each condition, 1,000 samples were generated. There were a total of 30 different conditions.

For each sample, we estimated the parameters in the equation for  $y$  as the dependent variable using both ML-SEM and AB. We used Stata both to generate the data and estimate the models. The `sem` command<sup>7</sup> was used for ML-SEM, and the `xtdpd` command was used for AB. For the latter, we used the default one-step method with all available instruments.<sup>8</sup> Program code for the Monte Carlo simulations is available in an online appendix.

We will focus on the estimates for  $\beta_1$ , the cross-lagged effect of  $x$  on  $y$ , and  $\beta_2$ , the autoregressive effect of  $y$  on itself. For each of those parameters and for each condition, Appendix Table A.1 reports the mean and the standard deviation of the estimates for  $\beta_1$  across the 1,000 samples as well

<sup>7</sup>We actually used the user-written command, `xtdpdml`, which serves as a simplifying shell for the `sem` command. For details, see Williams, Allison, and Moral-Benito (2016)

<sup>8</sup>The ML models were less restrictive than the Arellano-Bond (AB) models. Specifically, the ML-SEM models allowed for a different intercept and a different error variance at each point in time, while the AB models constrained those estimates to be the same for all time points. Since the data-generating process embodied those constraints, both AB and ML-SEM should produce consistent estimates. However, the fact that AB estimated fewer parameters may have given it some advantage in assessing the relative efficiency of the two estimators.

as the “coverage”—the proportion of nominal 95 percent confidence intervals (calculated in each sample using the conventional normal approximation) that actually included the true values. If a method is performing well, the coverage should be close to .95.

Unlike AB, ML-SEM requires an iterative algorithm, and that algorithm sometimes fails to converge, especially with small samples and more extreme parameter values. Of the 30 conditions in the simulation, 12 had convergence failures for at least one of the 1,000 samples. For each condition, Appendix Table A.1 gives the number of convergence failures. Of those conditions that had convergence failures, the number of failures ranged from 1 to 23 (out of 1,000 samples). Thus, in the worst case, only about 2 percent of the samples suffered convergence failures.

Samples with convergence failures are likely to be more extreme or unusual than those without such failures, and the exclusion of those samples could give an unfair advantage to ML-SEM. To avoid that possibility, we also excluded the same nonconvergent samples from the AB estimation.

In Appendix Table A.1, both ML-SEM and AB estimators appear at first glance to produce approximately unbiased point estimates of  $\beta_1$  (the cross-lagged effect of  $x$  on  $y$ ) under all 30 conditions. However, the AB estimator actually shows a small downward bias under many conditions. Specifically, for AB, 11 of the thirty 95 percent Monte Carlo confidence intervals (not shown) do *not* include the true value because the upper confidence limit is less than the true value. For ML-SEM, on the other hand, every 95 percent Monte Carlo confidence interval includes the true value. Despite the downward bias in AB, the two estimators did about equally well for interval estimation. For both ML-SEM and AB, the median coverage over the 30 conditions was .949. ML-SEM coverage ranged from .934 to .957. For AB, the coverage ranged from .937 to .958.

For  $\beta_2$  (the lagged effect of  $y$  on itself), AB does substantially worse than ML-SEM, as shown in Table A.2. Again, ML-SEM produces approximately unbiased estimates of  $\beta_2$  under all conditions. Every 95 percent Monte Carlo confidence interval included the true value. On the other hand, AB estimates are persistently smaller than the true values, and only *one* of the 95 percent Monte Carlo confidence intervals included the true value. This downward bias is generally small, however, except for the smaller sample sizes of  $N = 50$  and  $N = 100$ , where it’s quite apparent. Somewhat surprisingly, given earlier literature, the bias is small even when  $\beta_2$  is at or close to 1.0.

The bias in AB for  $\beta_2$  translates into slightly worse coverage for interval estimates. For ML-SEM, the median coverage over the 30 conditions was .951 with a range from .937 to .965. For AB, the median coverage was .941, ranging from .890 (for  $N = 50$ ) to .961.

Next, we examine the relative efficiency of ML-SEM and AB. We calculate relative efficiency as the ratio of the

**Table 3.** How Number of Time Points Affects Relative Efficiency of the Arellano-Bond Method.

	$\beta_1$	$\beta_2$
$T = 4$	1.040	.827
$T = 5$	.944	.708
$T = 7$	.939	.505
$T = 10$	.832	.337

estimated mean square error for ML-SEM to the estimated mean square error for AB. Mean square error is the sampling variance plus the square of the bias. Across 30 different conditions, the relative efficiency of AB compared with ML-SEM for estimating  $\beta_1$  ranged from .83 (AB did 17 percent worse) to 1.12 (AB did 12 percent better), with a median of .96. So there was no clear winner for the cross-lagged effect. For  $\beta_2$ , however, the relative efficiency ranged from .34 (AB did 66 percent worse) to .87 (AB did 13 percent worse) with a median of .68. To put this in perspective, if the relative efficiency is .50, then using AB rather than ML would be equivalent to discarding half of the sample.

What affects relative efficiency? Although the relative efficiencies for all conditions are shown in the last column of Tables A.1 and A.2, we also present some of them here to highlight key results. Table 3 gives the relative efficiencies of the estimators for  $\beta_1$  and  $\beta_2$  as a function of the number of time points.

For both  $\beta_1$  and  $\beta_2$ , the relative efficiency of AB declines with the number of time points, although the decline is much more precipitous for  $\beta_2$ , the autoregressive coefficient. These declines are consistent with the literature suggesting that when there are many time points—and therefore many instruments—AB is vulnerable to overfitting.

Unfortunately, there is a potential problem with the results in Table 3. Because our data-generating model was not constrained to be stationary, the variances and covariances at later time points may have differed from those at earlier time points. So the declines in efficiency observed in Table 3 could reflect not the *number* of time points but rather changes over time in the pattern of variances and covariances. To avoid this possible confounding, we first produced approximate stationarity by generating data from 1,000 time points. Then, for the  $T = 4$  condition, we used the data from the next four time points to estimate the model. The same strategy was used for  $T = 5, 7,$  and  $10$ . Results in Table 4 show that the ML and AB estimators do about equally well for  $\beta_1$  regardless of the number of time points. For  $\beta_2$ , however, the efficiency of AB is quite low at all time points and declines noticeably as the number of time points grows larger. In this case, the inferiority of AB stems both from downward bias in the coefficients and from larger standard errors.

Previous work (Kiviet 2005) has suggested that the ratio of the fixed effects variance to the error variance may be an important factor in the efficiency of AB. Table 5 confirms this

**Table 4.** Number of Time Points and Relative Efficiency under Approximate Stationarity.

	$\beta_1$	$\beta_2$
$T = 4$	.982	.562
$T = 5$	.989	.527
$T = 7$	1.038	.371
$T = 10$	1.029	.305

**Table 5.** How Relative Efficiency Depends on the Variance of  $\varepsilon$  and  $\alpha$ .

	$\beta_1$	$\beta_2$
$SD(\varepsilon) = .25$	.924	.716
$SD(\varepsilon) = 1.0$	.944	.708
$SD(\varepsilon) = 1.5$	.974	.640
$SD(\varepsilon) = 2.0$	.964	.570
$SD(\alpha) = .5$	.981	.629
$SD(\alpha) = 1.0$	.944	.708
$SD(\alpha) = 1.5$	.989	.814
$SD(\alpha) = 2.0$	.974	.863

for  $\beta_2$  but not for  $\beta_1$ . When the standard deviations of  $\alpha$  and  $\eta$  are held constant at 1.0, increases in the standard deviations of  $\varepsilon$  and  $\nu$  are associated with declines in the relative efficiency of AB for  $\beta_2$  but not for  $\beta_1$ . The direction is reversed when the standard deviation of  $\varepsilon$  is held constant and the standard deviation of  $\alpha$  is varied—higher standard deviations of  $\alpha$  result in *higher* relative efficiency of AB for  $\beta_2$ .

Table 6 shows how relative efficiency is affected by the value of  $\rho$ , the correlation between the two fixed effects,  $\alpha$  and  $\eta$ . For  $\beta_1$ , there is no apparent trend. For  $\beta_2$ , however, the relative efficiency of AB increases substantially as the correlation goes from 0 to .90.

Finally, Table 7 shows how relative efficiency is affected by the magnitude of  $\beta_1$ , the cross-lagged coefficient. As in the last two tables, the relative efficiencies of AB estimates for  $\beta_1$  are virtually unaffected. For  $\beta_2$ , relative efficiency increases substantially as  $\beta_1$  gets larger.

There were three other factors that had no apparent effect on relative efficiency: the sign of  $\beta_4$  (the cross-lagged effect of  $y$  on  $x$ ), sample size, and the magnitude of  $\beta_2$  (the autoregressive coefficient). The absence of a relationship with sample size and  $\beta_2$  is somewhat surprising. We expected ML-SEM to do better in smaller samples, and we expected AB to perform more poorly when  $\beta_2$  was close or equal to 1.0. In fact, AB did quite well when  $\beta_2 = 1$ , both in absolute terms and relative to ML-SEM.

Because ML-SEM is based on the assumption of multivariate normality, it has been suggested that it may do worse than AB when distributions are not normal. To check this out, for the baseline set of parameter values, all the random draws were made from a chi-square distribution with two degrees of freedom, a distribution that is highly skewed to the right.

**Table 6.** How Relative Efficiency Depends on  $\rho$ .

	$\beta_1$	$\beta_2$
$\rho = 0$	.936	.378
$\rho = .25$	1.044	.467
$\rho = .50$	.881	.535
$\rho = .75$	.944	.708
$\rho = .90$	.942	.759

**Table 7.** How Relative Efficiency Depends on  $\beta_1$ .

	$\beta_1$	$\beta_2$
$\beta_1 = 0$	.978	.483
$\beta_1 = .25$	.944	.708
$\beta_1 = .50$	.959	.750
$\beta_1 = .75$	.875	.814
$\beta_1 = 1.00$	.938	.779

The last rows of Tables A.1 and A.2 show that both estimators did well under this condition, but ML-SEM did better. The relative efficiency of AB was .908 for  $\beta_1$  and .619 for  $\beta_2$ .

## Discussion and Conclusion

Panel data have a lot of potential for improving our ability to make causal inferences from non-experimental data. But appropriate methods are needed take advantage of such data. The linear dynamic panel model of econometrics protects against two major threats to valid causal inference: unmeasured confounders and reverse causation. The Arellano-Bond method can produce approximately unbiased estimates of the parameters of that model under a wide range of conditions. One of our goals was to show that cross-lagged panel models can be estimated within this econometric framework by estimating each side of the cross-lagged model separately. Compared with simultaneous estimation, this approach has the advantage producing estimates that are robust to misspecification of the lag structure or the functional form of the other side.

Unfortunately, the AB method is reputed to be problematic when the autoregressive parameter is near 1.0, and its efficiency has also been questioned. Maximum likelihood methods based on first differences have been offered as an alternative, but they rely on questionable assumptions about the initial conditions.

In this article, we have shown that the linear dynamic panel model with predetermined regressors is a special case of the well-known linear structural equation model. Instead of relying on difference scores to eliminate the fixed effects, maximum likelihood estimation of this model is accomplished by allowing the fixed effects to have unrestricted correlations with the time-varying predictors. The initial observations of the dependent variable are treated just like any other exogenous variables.

Cross-lagged causation is accommodated by allowing the error term in each equation to correlate with future values of the time-dependent predictors. Many different statistical packages, both freeware and commercial, can implement the ML-SEM method.

Monte Carlo simulations showed that ML-SEM produced approximately unbiased estimates under all the conditions studied. Confidence interval coverage was also excellent. The AB estimator also did very well for the cross-lagged parameter, although with some downward bias. For the autoregressive parameter, however, the downward bias in AB was much more substantial, and its efficiency (relative to ML-SEM) was poor under most conditions. For this parameter, the efficiency of AB relative to ML-SEM also declined markedly as the number of time points increased.

So one key conclusion from the simulations is that if you are primarily interested in the cross-lagged coefficients, AB and ML-SEM have statistical properties that are about equally good. Nevertheless, there are still plenty of reasons why one might prefer ML-SEM. As detailed by Bollen and Brand (2010), the ML-SEM method can be extended in several ways. Although maximum likelihood is the default estimator for all SEM packages, most packages offer alternative methods, including the asymptotic distribution free method of Browne (1984). Many packages also have options for robust standard errors. Many of the constraints that are implied by the linear dynamic panel model can be easily relaxed in the SEM setting. We already showed how the error variances can be allowed to vary with time. One could also allow the coefficients to vary with time.

It's even possible to allow the coefficient of  $\alpha$ , the fixed effect, to vary with time instead of being constrained to 1 for every time point. This option is attractive because it removes one of the principal limitations of the classic fixed effects estimator: that it does not control for unmeasured time-invariant variables when their *effects* change over time. It is also possible to allow for individual-specific trends that are correlated with the time-varying predictors (Teachman 2014).

With regard to unbalanced samples and missing data more generally, most SEM packages have the option of handling missing data by FIML. Unlike AB, FIML can easily handle missing data on predictor variables.<sup>9</sup>

Although we have not considered models with simultaneous reciprocal effects, such effects can certainly be built into SEM models if appropriate instruments are available. Finally, some SEM packages (like Mplus or the **gsem** command in Stata) can estimate similar models for categorical dependent variables.

Are there any downsides to ML-SEM? As noted earlier, ML-SEM is not suitable when  $T$  is large relative to  $N$ . This is easily seen from the fact that ML-SEM operates on the full covariance matrix for all the variables at all points in time. For example, if the predictors in the model consist of nine

<sup>9</sup>For examples of the use of FIML to handle missing data, see Williams et al. (2016) and Moral-Benito et al. (2017).

time-varying variables and  $T = 11$ , then the covariance matrix will be  $101 \times 101$ . And unless  $N > 101$ , that matrix will not have full rank, causing the maximization algorithm to break down. As noted earlier, ML-SEM will sometimes fail to converge. And even if it converges, computation time can be considerably greater than for AB, especially when using FIML to handle missing data.

As can be seen from Figure 2 and Appendix B, program code for ML-SEM can be more complex than for the AB method. Not only are ML-SEM programs typically longer, but it can also be challenging for the analyst to figure out exactly how to specify the model so that the correct covariances are either set to 0 or left unconstrained. This

challenge is exacerbated by the fact that different packages have different defaults for the covariances and different ways of overriding those defaults.

Of course, this is just a programming issue, and it would certainly be feasible to write a Stata command, a SAS macro, or an R function that would automatically set up the correct model with minimal input from the user. As noted earlier, we have already developed a Stata command called **xtdpdml** that does exactly that (Williams, Allison, and Moral-Benito 2016). This command radically reduces the programming needed for ML-SEM and is actually simpler to use than the built-in Stata commands for AB estimation (**xtabond**, **xtdpd**, or **xtdpdsys**).

## Appendix

### A. Results from Monte Carlo Simulations

**Table A.1.** Performance of Maximum Likelihood and Arellano-Bond Estimators of  $\beta_1$ , the Lagged Effect of  $x$  on  $y$ .

Condition	True <sup>a</sup>	Nonconverge <sup>b</sup>	Maximum Likelihood			Arellano-Bond			Relative Efficiency
			Mean <sup>c</sup>	SD <sup>d</sup>	Coverage <sup>e</sup>	Mean <sup>c</sup>	SD <sup>d</sup>	Coverage <sup>e</sup>	
N = 50	.25	19	.256	.0975	.934	.240	.0918	.939	1.121
N = 100	.25	23	.250	.0664	.946	.242	.0664	.943	.985
N = 400	.25		.250	.0314	.949	.249	.0323	.952	.944
N = 1,600	.25		.250	.0161	.947	.250	.0164	.952	.964
$\beta_2 = 0$	.25		.250	.0325	.950	.252	.0324	.955	1.004
$\beta_2 = .25$	.25		.251	.0322	.947	.252	.0310	.954	1.076
$\beta_2 = .50$	.25		.251	.0313	.950	.249	.0318	.949	.969
$\beta_2 = .75$	.25		.250	.0314	.949	.249	.0323	.952	.944
$\beta_2 = .90$	.25		.251	.0324	.953	.248	.0343	.943	.890
$\beta_2 = 1.0$	.25		.249	.0328	.949	.249	.0346	.949	.900
$\beta_2 = 1.25$	.25		.249	.0335	.957	.248	.0345	.941	.942
$\beta_1 = 0$	.00	2	.000	.0331	.947	-.002	.0334	.947	.978
$\beta_1 = .25$	.25		.250	.0314	.949	.249	.0323	.952	.944
$\beta_1 = .50$	.50		.500	.0310	.953	.498	.0316	.958	.959
$\beta_1 = .75$	.75		.750	.0304	.954	.749	.0325	.957	.875
$\beta_1 = 1.00$	1.00		1.001	.0300	.948	1.000	.0310	.948	.938
$\beta_1 = -.25$	.25		.250	.0314	.949	.249	.0323	.952	.944
$\beta_1 = 0$	.25	4	.250	.0337	.948	.250	.0336	.949	1.007
$\beta_1 = .25$	.25	23	.250	.0311	.950	.251	.0309	.949	1.014
$\rho = 0$	.25	4	.251	.0325	.947	.248	.0336	.941	.936
$\rho = .25$	.25		.250	.0337	.939	.248	.0329	.949	1.044
$\rho = .50$	.25		.250	.0315	.954	.247	.0334	.951	.881
$\rho = .75$	.25		.250	.0314	.949	.249	.0323	.952	.944
$\rho = .90$	.25		.252	.0321	.940	.249	.0331	.953	.942
$SD(\varepsilon) = .25$	.25	15	.251	.0305	.950	.249	.0317	.944	.924
$SD(\varepsilon) = 1.0$	.25		.250	.0314	.949	.249	.0323	.952	.944
$SD(\varepsilon) = 1.5$	.25	1	.250	.0325	.954	.247	.0328	.951	.974
$SD(\varepsilon) = 2$	.25	13	.250	.0318	.949	.247	.0322	.954	.964

(continued)



Table A1. (continued)

Condition	True <sup>a</sup>	Nonconverge <sup>b</sup>	Maximum Likelihood			Arellano-Bond			Relative Efficiency
			Mean <sup>c</sup>	SD <sup>d</sup>	Coverage <sup>e</sup>	Mean <sup>c</sup>	SD <sup>d</sup>	Coverage <sup>e</sup>	
SD( $\alpha$ ) = .5	.25		.250	.0336	.938	.247	.0338	.937	.981
SD( $\alpha$ ) = 1.0	.25	21	.250	.0314	.949	.249	.0323	.952	.944
SD( $\alpha$ ) = 1.5	.25		.251	.0315	.949	.250	.0318	.955	.989
SD( $\alpha$ ) = 2.0	.25	2	.251	.0317	.945	.250	.0322	.941	.974
T = 4	.25		.250	.0402	.947	.249	.0394	.948	1.040
T = 5	.25		.250	.0314	.949	.249	.0323	.952	.944
T = 7	.25	8	.251	.0247	.952	.248	.0254	.949	.939
T = 10	.25		.250	.0188	.949	.246	.0203	.941	.832
Chi-square	.25		.252	.0315	.941	.250	.0331	.947	.908

<sup>a</sup>True value of the coefficient in the model producing the data.

<sup>b</sup>Number of nonconvergent samples for maximum likelihood.

<sup>c</sup>Mean of 1,000 parameter estimates.

<sup>d</sup>Standard deviation of 1,000 parameter estimates.

<sup>e</sup>Proportion of nominal 95 percent confidence intervals that include the true value.

Table A2. Performance of Maximum Likelihood and Arellano-Bond Estimators of  $\beta_2$ , the Lagged Effect of  $y$  on Itself.

Condition	True <sup>a</sup>	Maximum Likelihood			Arellano-Bond			Relative Efficiency
		Mean <sup>b</sup>	SD <sup>c</sup>	Coverage <sup>d</sup>	Mean <sup>b</sup>	SD <sup>c</sup>	Coverage <sup>d</sup>	
N = 50	.75	.752	.0917	.953	.680	.1021	.890	.549
N = 100	.75	.752	.0675	.951	.716	.0760	.914	.657
N = 400	.75	.750	.0325	.953	.742	.0377	.950	.708
N = 1,600	.75	.750	.0159	.955	.748	.0195	.946	.656
$\beta_2 = 0$	.00	.001	.0327	.948	-.005	.0377	.955	.738
$\beta_2 = .25$	.25	.248	.0322	.964	.241	.0456	.951	.554
$\beta_2 = .50$	.50	.499	.0378	.952	.488	.0492	.947	.554
$\beta_2 = .75$	.75	.750	.0325	.953	.742	.0377	.950	.708
$\beta_2 = .90$	.90	.901	.0257	.961	.895	.0289	.952	.769
$\beta_2 = 1.0$	1.00	.999	.0219	.956	.996	.0243	.940	.798
$\beta_2 = 1.25$	1.25	1.250	.0150	.947	1.249	.0160	.942	.867
$\beta_1 = 0$	.25	.752	.0425	.960	.730	.0577	.936	.483
$\beta_1 = .25$	.25	.750	.0325	.953	.742	.0377	.950	.708
$\beta_1 = .50$	.75	.750	.0249	.953	.747	.0286	.952	.750
$\beta_1 = .75$	.75	.750	.0208	.956	.748	.0229	.946	.814
$\beta_1 = 1.00$	.75	.751	.0175	.948	.748	.0198	.933	.779
$\beta_4 = -.25$	.75	.750	.0325	.953	.742	.0377	.950	.708
$\beta_4 = 0$	.75	.751	.0303	.953	.743	.0344	.942	.749
$\beta_4 = .25$	.75	.749	.0294	.938	.743	.0323	.938	.788
$\rho = 0$	.75	.753	.0418	.948	.727	.0640	.919	.378
$\rho = .25$	.75	.752	.0380	.948	.736	.0540	.929	.467
$\rho = .50$	.75	.750	.0352	.948	.737	.0464	.925	.535
$\rho = .75$	.75	.750	.0325	.953	.742	.0377	.950	.708
$\rho = .90$	.75	.751	.0305	.950	.742	.0343	.949	.759
SD( $\varepsilon$ ) = .25	.75	.751	.0254	.942	.745	.0296	.944	.716
SD( $\varepsilon$ ) = 1.0	.75	.750	.0325	.953	.742	.0377	.950	.708
SD( $\varepsilon$ ) = 1.5	.75	.751	.0455	.951	.731	.0536	.935	.640
SD( $\varepsilon$ ) = 2	.75	.750	.0519	.965	.723	.0632	.934	.570
SD( $\alpha$ ) = .5	.75	.752	.0552	.948	.725	.0650	.919	.629
SD( $\alpha$ ) = 1.0	.75	.750	.0325	.953	.742	.0377	.950	.708
SD( $\alpha$ ) = 1.5	.75	.751	.0232	.946	.747	.0256	.943	.814

(continued)

Table A2. (continued)

Condition	True <sup>a</sup>	Maximum Likelihood			Arellano-Bond			Relative Efficiency
		Mean <sup>b</sup>	SD <sup>c</sup>	Coverage <sup>d</sup>	Mean <sup>b</sup>	SD <sup>c</sup>	Coverage <sup>d</sup>	
SD( $\alpha$ ) = 2.0	.75	.751	.0176	.951	.749	.0189	.961	.863
T = 4	.75	.752	.0483	.944	.744	.0529	.946	.827
T = 5	.75	.750	.0325	.953	.742	.0377	.950	.708
T = 7	.75	.750	.0224	.938	.739	.0295	.928	.505
T = 10	.75	.750	.0157	.950	.736	.0231	.898	.337
Chi-square	.75	.750	.0317	.959	.744	.0398	.941	.619

<sup>a</sup>True value of the coefficient in the model producing the data.

<sup>b</sup>Mean of 1,000 parameter estimates.

<sup>c</sup>Standard deviation of 1,000 parameter estimates.

<sup>d</sup>Proportion of nominal 95 percent confidence intervals that include the true value.

## B. Example Program Code for Other Software Packages

### Stata

```

1 use c:\wagewide.dta, clear
2 sem (wks2 <- lwage1@a wks1@b union1@c ed@d
   Alpha@1 E2@1) ///
3   (wks3 <- lwage2@a wks2@b union2@c ed@d
   Alpha@1 E3@1) ///
4   (wks4 <- lwage3@a wks3@b union3@c ed@d
   Alpha@1 E4@1) ///
5   (wks5 <- lwage4@a wks4@b union4@c ed@d
   Alpha@1 E5@1) ///
6   (wks6 <- lwage5@a wks5@b union5@c ed@d
   Alpha@1 ) ///
7   (wks7 <- lwage6@a wks6@b union6@c ed@d
   Alpha@1 ), ///
8 var(e.wks2@0 e.wks3@0 e.wks4@0 e.wks5@0) ///
9 cov(Alpha*(ed E*)@0) cov(_OEx*(E2 E3 E4
   E5)@0) ///
10 cov(E2*(E3 E4 E5)@0) cov(E3*(E4 E5)@0)
   cov(E4*(E5)@0) ///
11 cov(E2*(union3 union4 union5 union6)) ///
12 cov(E3*(union4 union5 union6)) ///
13 cov(E4*(union5 union6)) cov(E5*union6) noxcondi-
   tional ///
14 var(E2-E5@f e.wks6@f e.wks7@f)

```

### Explanation

- Line 1 reads the wide-form data into memory.
- Lines 2–14 are all one single **sem** command. The `///` at the end of each line allows the command to be spread across multiple lines in a DO file.
- Lines 2–7 specify the linear equations for years 2 through 7.
- The rule is that variable names (like **lwage1**) that begin with lower-case letters are observed variables, while variable names that begin with upper-case letters (like **Alpha**) are latent.

- `@a` assigns the name `a` to the coefficient for `lwage1`. Giving parameters the same name constrains them to be equal. `Alpha@1` constrains the coefficient of `Alpha` to be 1.0.
- An unfortunate limitation of the **sem** command is that it does not allow the error term in an equation to be correlated with observed, exogenous variables. But that's exactly what we need to do for the dynamic panel model: allow error terms to be correlated with future values of the time-dependent predictors, in this case **union**. The workaround is to suppress the original error terms (by setting their variances equal to 0 in line 8) and introducing new latent error terms `E2–E5` in lines 2–5. There is no need to do that at times 6 and 7 because there are no future values of **union** in the model.
- In line 9, the first **cov** option sets to 0 the covariance between **Alpha** and the time-invariant predictor **ed**, as well as the covariances between **Alpha** and the new error terms. The second **cov** option sets to 0 the covariances between the new error terms and all of the observed, exogenous variables (`_OEx`).
- Line 10 constrains the all the new error terms to be uncorrelated with each other.
- Lines 11–13 allow the new error terms to be correlated with future values of the predetermined predictor, **union**.
- The **noxconditional** option on line 13 requests that the means, variances, and covariances of the observed exogenous variables be included in the parameters. For unknown reasons, the model will not run correctly without this option.
- Line 14 constrains the error variance to be the same at all time points. This is generally not advisable but was done here to ensure comparability with the AB method.

This model can also be estimated in Stata with the user-written **xtpdml** command, which acts as a simplifying shell for

the `sem` command (Williams Allison, and Moral-Benito 2016). The code for this example is:

```
use c:\wages.dta, clear
xtset id t
xtpdml wks L.lwage, pre(L.union) inv(ed) errorinv
```

*Mplus*

```
1 data: file = 'C:\wagewide.csv';
2 variable: names =
3     id ed fem blk wks1 wks2 wks3 wks4 wks5
4     wks6 wks7
5     lwage1 lwage2 lwage3 lwage4 lwage5
6     lwage6 lwage7
7     union1 union2 union3 union4 union5 union6
8     union7
9     ms1 ms2 ms3 ms4 ms5 ms6 ms7;
10 usevar=ed wks1 wks2 wks3 wks4 wks5 wks6 wks7
11     lwage1 lwage2 lwage3 lwage4 lwage5
12     lwage6
13     union1 union2 union3 union4 union5 union6;
14 model:
15     alpha by wks2@1 wks3@1 wks4@1 wks5@1
16     wks6@1 wks7@1;
17         wks2 on lwage1 (1)
18             wks1 (2)
19             union1 (3)
20             ed (4);
21         wks3 on lwage2 (1)
22             wks2 (2)
23             union2 (3)
24             ed (4);
25         wks4 on lwage3 (1)
26             wks3 (2)
27             union3 (3)
28             ed (4);
29         wks5 on lwage4 (1)
30             wks4 (2)
31             union4 (3)
32             ed (4);
33         wks6 on lwage5 (1)
34             wks5 (2)
35             union5 (3)
36             ed (4);
37         wks7 on lwage6 (1)
38             wks6 (2)
39             union6 (3)
40             ed (4);
41     alpha with wks1 lwage1-lwage6 union1-union6;
42     wks2 with union3-union6;
43     wks3 with union4-union6;
44     wks4 with union5 union6;
45     wks5 with union6;
46     ed on wks1 lwage1-lwage6 union1-union6;
47     wks2-wks7 (5);
```

*Explanation*

- Line 1 specifies the location of the data file. It must be a text file in free format: one record per person, no variable names, and spaces between values.
- Lines 2–6 assign names to the variables in the order in which they appear on the data file.
- Lines 7–9 restrict the variables to those that actually appear in the model.
- Line 10 begins the model specification.
- Line 11 *defines* the latent variable **alpha** by specifying its indicators, each of which has a “factor loading” constrained to be 1.0.
- Lines 12–15 specify the first regression equation. In order to constrain coefficients to be the same across equations, the predictors must be on different lines, with a number in parentheses at the end of each line.
- Lines 16–35 specify the regressions for the remaining time points. Coefficients for variables followed by the same numbers are constrained to be the same.
- Line 36 allows the latent variable **alpha** to be correlated with the predictor variables, except for **ed**.
- Lines 37–40 allow the error term in each equation to be correlated with future values of the predetermined variable **union**.
- Line 41 allows **ed** to be correlated with all the other exogenous variables. It accomplishes this by specifying a regression with **ed** as the dependent variable.
- Line 42 constrains the error variances to be the same at all time points. This is generally not advisable but was done here to ensure comparability with the AB method.

*lavaan (R package)*

```
1 wage <- read.table("C:/wagenames.txt",header=T)
2 wagemod <- '
3 alpha =~ 1*wks2 + 1*wks3 + 1*wks4 + 1*wks5 +
4 1*wks6 + 1*wks7
5 wks2 ~ a*wks1 + b*union1 + c*lwage1 + d*ed
6 wks3 ~ a*wks2 + b*union2 + c*lwage2 + d*ed
7 wks4 ~ a*wks3 + b*union3 + c*lwage3 + d*ed
8 wks5 ~ a*wks4 + b*union4 + c*lwage4 + d*ed
9 wks6 ~ a*wks5 + b*union5 + c*lwage5 + d*ed
10 wks7 ~ a*wks6 + b*union6 + c*lwage6 + d*ed
11 wks2 ~~ union3 + union4 + union5 + union6
12 wks3 ~~ union4 + union5 + union6
13 wks4 ~~ union5 + union6
14 wks5 ~~ union6
15 alpha ~ wks1+lwage1+lwage2+lwage3+lwage4
16 +lwage5+lwage6+
17 union1+union2+union3+union4+union5+union6
18 union6 ~ ed+wks1+lwage1+lwage2+lwage3+lwage4
19 +lwage5+lwage6+
20 union1+union2+union3+union4
```

```

20 union4 ~ ed+wks1+lwage1+lwage2+lwage3+lwage4
  +lwage5+lwage6+
21   union1+union2+union3
22 union3 ~ ed+wks1+lwage1+lwage2+lwage3+lwage4
  +lwage5+lwage6+
23   union1+union2
24 wks2 ~ f*wks2
25 wks3 ~ f*wks3
26 wks4 ~ f*wks4
27 wks5 ~ f*wks5
28 wks6 ~ f*wks6
29 wks7 ~ f*wks7 ‘
30 wagefit <- sem(wagemod,data=wage)
31 summary(wagefit)

```

### Explanation

- Line 1 reads the data from a text file in free format with variable names in the first row. This data set is assigned the name **wage**.
- Lines 2–23 specify the model, which is stored in the object **wagemod**. The model specification is a “literal” that is demarcated by single quotes.
- Line 3 defines the latent variable **alpha** by naming its indicators, each with a “factor loading” of 1.0. The symbol  $\sim$  means “is measured by”.
- Lines 4–9 specify the six regression equations. The symbol  $\sim$  means “is regressed on.” The letters preceding each variable are the names of the coefficients. Coefficients with the same names are constrained to be equal.
- Lines 10–13 allow the error terms in each of the equations to be correlated with future values of the predetermined variable **union**. The symbol  $\sim\sim$  means “is correlated with.”
- Lines 14–15 allow the latent variable **alpha** to be correlated with other exogenous variables, except for **ed**.
- Lines 16–23 allow the **union** variables to be correlated with other exogenous variables. This is necessary because lines 10–13 caused the **union** variables to be treated as endogenous, and the default in lavaan is to presume that endogenous and exogenous variables are uncorrelated.
- Lines 24–29 impose the constraint that error variances are the same at all time points. This is not generally advisable but was done here to ensure comparability with the AB method.
- Line 30 calls the **sem** function, which actually fits the model, using the **wage** data and the **wagemod** model specification.
- Line 31 reports the estimates and associated statistics.

### Authors’ Note

These data are widely available on the Web. We downloaded them from <http://people.stern.nyu.edu/wgreene/Econometrics/PanelDataSets.htm>.

### References

- Ahn, Seung C., and Peter Schmidt. 1995. “Efficient Estimation of Models for Dynamic Panel Data.” *Journal of Econometrics* 68:5–27.
- Allison, Paul D. 2000. “Inferring Causal Order from Panel Data.” Paper presented at the Ninth International Conference on Panel Data, June 22, Geneva, Switzerland.
- Allison, Paul D. 2005a. “Causal Inference with Panel Data.” Paper presented at the Annual Meeting of the American Sociological Association, August.
- Allison, Paul D. 2005b. *Fixed Effects Regression Methods for Longitudinal Data Using SAS*. Cary, NC: The SAS Institute.
- Allison, Paul D. 2009. *Fixed Effects Regression Models*. Thousand Oaks, CA: Sage Publications.
- Allison, Paul D., and Kenneth A. Bollen. 1997. “Change Score, Fixed Effects, and Random Component Models: A Structural Equation Approach.” Paper presented at the Annual Meeting of the American Sociological Association.
- Anderson, Theodore W., and Cheng Hsiao. 1982. “Formulation and Estimation of Dynamic Models Using Panel Data.” *Journal of Econometrics* 18:67–82.
- Arellano, Manuel. 2003. *Panel Data Econometrics*. Oxford: Oxford University Press.
- Arellano, Manuel, and Stephen Bond. 1991. “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations.” *The Review of Economic Studies* 58:277–97.
- Baltagi, Badi H. 2013. *Econometric Analysis of Panel Data*. 5th ed. New York: John Wiley & Sons.
- Bentler, Peter M., and David G. Weeks. 1980. “Linear Structural Equations with Latent Variables.” *Psychometrika* 45:289–308.
- Bhargava, Alok, and J. Dennis Sargan. 1983. “Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods.” *Econometrica* 51:1635–59.
- Blundell, Richard, and Stephen Bond. 1998. “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models.” *Journal of Econometrics* 87:115–44.
- Bollen, Kenneth A., and Jennie E. Brand. 2010. “A General Panel Model with Random and Fixed Effects: A Structural Equations Approach.” *Social Forces* 89:1–34.
- Browne, Michael W. 1984. “Asymptotically Distribution-free Methods for the Analysis of Covariance Structures.” *British Journal of Mathematical and Statistical Psychology* 37:62–83.
- Chamberlain, Gary. 1982. “Multivariate Regression Models for Panel Data.” *Journal of Econometrics* 18:5–46.
- Chamberlain, Gary 1984. “Panel Data.” Pp. 1247–318 in *Handbook of Econometrics*, edited by Z. Griliches, and M. Intriligator. Amsterdam: North-Holland.
- Cornwell, Christopher, and Peter Rupert. 1988. “Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators.” *Journal of Applied Econometrics* 3:149–55.
- Duncan, Otis D. 1969. “Some Linear Models for Two-wave, Two-variable Panel Analysis.” *Psychological Bulletin* 72:177–82.
- Finkel, Steven E. 1995. *Causal Analysis with Panel Data*. Thousand Oaks, CA: Sage Publications.
- Firebaugh, Glenn, Cody Warner, and Michael Massoglia. 2013. “Fixed Effects, Random Effects, and Hybrid Models for Causal Analysis.” Pp. 113–31 in *Handbook of Causal Analysis for Social Research*, edited by S. L. Morgan. New York: Springer.



- Grassetti, Luca. 2011. "A Note on Transformed Likelihood Approach in Linear Dynamic Panel Models." *Statistical Methods & Applications* 20:221–40.
- Halaby, Charles N. 2004. "Panel Models in Sociological Research: Theory into Practice." *Annual Review of Sociology* 30:507–44.
- Hamaker, Ellen L., Rebecca M. Kuiper, and Raoul P. P. P. Grasman. 2015. "A Critique of the Cross-lagged Panel Model." *Psychological Methods* 20:102–16.
- Han, Chirok, and Peter C. B. Phillips. 2013. "First Difference Maximum Likelihood and Dynamic Panel Estimation." *Journal of Econometrics* 175:35–45.
- Holtz-Eakin, Douglas, Whitney Newey, and Harvey S. Rosen. 1988. "Estimating Vector Autoregressions with Panel Data." *Econometrica* 56:1371–95.
- Hsiao, Cheng. 2014. *Analysis of Panel Data*. 3rd ed. London: Cambridge University Press.
- Hsiao, Cheng, M. Hashem Pesaran, and A. Kamil Tahmiscioglu. 2002. "Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods." *Journal of Econometrics* 109:107–50.
- Jöreskog, Karl G. 1978. "Structural Analysis of Covariance and Correlation Matrices." *Psychometrika* 43:443–77.
- Judson, Ruth A., and Ann L. Owen. 1999. "Estimating Dynamic Panel Data Models: A Guide for Macroeconomists." *Economics Letters* 65:9–15.
- Kenny, David A., and Charles M. Judd. 1996. "A General Procedure for the Estimation of Interdependence." *Psychological Bulletin* 119:138–48.
- Kessler, Ronald C., and David F. Greenberg. 1981. *Linear Panel Analysis: Models of Quantitative Change*. New York: Academic Press.
- Kiviet, Jan F. 2005. "Judging Contending Estimators by Simulation: Tournaments in Dynamic Panel Data Models." Tinbergen Institute Discussion Paper, No. 05-112/4.
- Kiviet, Jan F., Milan Pleus, and Rutger Poldermans. 2014. "Accuracy and Efficiency of Various GMM Inference Techniques in Dynamic Micro Panel Data Models." Amsterdam School of Economics Discussion Paper, No. 2014/9.
- Kripfganz, Sebastian. 2016. "Quasi-maximum Likelihood Estimation of Linear Dynamic Short-T Panel Data Models." *The Stata Journal* 16:1013–38.
- Markus, Gregory B. 1979. *Analyzing Panel Data*. Thousand Oaks, CA: Sage Publications.
- McArdle, John J., and John R. Nesselroade. 2014. *Longitudinal Data Analysis Using Structural Equation Modeling*. Washington, DC: American Psychological Association.
- Moral-Benito, Enrique. 2013. "Likelihood-based Estimation of Dynamic Panels with Predetermined Regressors." *Journal of Business & Economic Statistics* 31:451–72.
- Moral-Benito, Enrique, Paul D. Allison, and Richard Williams. 2017. "Dynamic Panel Data Modeling Using Maximum Likelihood: An Alternative to Arellano-Bond." Retrieved May 5, 2017 ([https://www3.nd.edu/~rwilliam/dynamic/Benito\\_Allison\\_Williams.pdf](https://www3.nd.edu/~rwilliam/dynamic/Benito_Allison_Williams.pdf)).
- Mundlak, Yair. 1978. "On the Pooling of Time Series and Cross Section Data." *Econometrica* 46:69–85.
- Nickell, Stephen. 1981. "Biases in Dynamic Models with Fixed Effects." *Econometrica: Journal of the Econometric Society* 49:1417–26.
- Roodman, David. 2009. "A Note on the Theme of Too Many Instruments." *Oxford Bulletin of Economics and Statistics* 71:135–58.
- Teachman, Jay. 2014. "Latent Growth Curve Models with Random and Fixed Effects." Pp. 3–17 in *Emerging Methods in Family Research*, edited by S. M. McHale, S. M. Amato, and A. Booth. New York: Springer International Publishing.
- Teachman, Jay, Greg J. Duncan, W. Jean Yeung, and Dan Levy. 2001. "Covariance Structure Models for Fixed and Random Effects." *Sociological Methods & Research* 30:271–88.
- Williams, Richard, Paul D. Allison, and Enrique Moral-Benito. 2016. "xtdpml: Linear Dynamic Panel-data Estimation Using Maximum Likelihood and Structural Equation Modeling." Retrieved May 5, 2017 (<http://www3.nd.edu/~rwilliam/dynamic/SJPaper.pdf>).
- Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

### Author Biographies

**Paul D. Allison** is professor of sociology at the University of Pennsylvania and president of Statistical Horizons LLC. After completing his doctorate in sociology at the University of Wisconsin, he did postdoctoral study in statistics at the University of Chicago and the University of Pennsylvania. He has published eight books and more than 70 articles on topics that include linear regression, logistic regression, structural equation models, inequality measures, missing data, survival analysis, and scientific careers. At present, his principal methodological research is focused on the analysis of longitudinal data, especially with determining the causes and consequences of events, and on methods for handling missing data. A former Guggenheim Fellow, Allison received the 2001 Lazarsfeld Award for distinguished contributions to sociological methodology. In 2010, he was named a Fellow of the American Statistical Association.

**Richard Williams** is associate professor and a former chairman of the Department of Sociology at the University of Notre Dame. His work has appeared in the *American Sociological Review*, *Social Forces*, *Social Problems*, *Demography*, *Cityscape*, and *Sociological Methods and Research*. Recent work by Williams has looked at issues involving the analysis of categorical data. He notes that assumptions of heterogeneity and proportionality are often violated in commonly used logit and ordered logit models. He discusses how heterogeneous choice models and generalized ordered logit models provide potential solutions to these problems. He is the author of the Stata programs `gologit2` and `oglm` that make the estimation of these models possible. In his work, he also shows how the use of adjusted predictions and marginal effects can be major aids in interpreting the results from categorical models. He won the 2015 Stata Journal Editors' Prize for his work in these areas.

**Enrique Moral-Benito** is an economist at Banco de España (DG Economics, Statistics and Research) and adjunct professor of quantitative methods at CEMFI. He received a PhD in economics from UIMP-CEMFI in 2010. His main research interests are aggregate productivity, economic growth, and panel data econometrics. His work has appeared in journals such as the *Review of Economics and Statistics*, *Journal of Applied Econometrics*, *Journal of Business and Economic Statistics*, and *European Economic Review*.