Maximum Likelihood for Cross-Lagged Panel Models with Fixed Effects

Paul D. Allison
University of Pennsylvania

Richard Williams
University of Notre Dame

Enrique Moral-Benito
Banco de España

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, xtdpdml, instrumental variable.

Maximum Likelihood for Cross-Lagged Panel Models with Fixed Effects

ABSTRACT

Panel data make it possible both to control for unobserved confounders and to 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 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.

1. 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 (Halaby 2004, Allison 2005a, Allison 2009, Firebaugh et al. 2013). 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., Markus 1979, Kessler and Greenberg 1981, Finkel 1995, Kenny

and Judd 1996, McArdle and Nesselroade 2014, Hamaker et al. 2015). 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 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 (Ahn and Schmidt 1995), have considerable small-sample bias, and often perform poorly 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 many 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 paper, 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 under most conditions.

This will be accomplished in several steps:

- Section 2 explores the relationship between the dynamic panel data models of econometrics and the cross-lagged panel models used in other social sciences.
- Section 3 reviews GMM estimation of dynamic panel data models and examines its limitations.
- Section 4 reviews the development of ML methods for dynamic panel data models.
- Section 5 shows how the ML method of Moral-Benito can be implemented in the SEM framework.
- Section 6 presents an empirical example.
- Section 7 presents results from a Monte Carlo study comparing the AB method and the ML-SEM method.
- Section 8 concludes with a discussion.

2. CROSS-LAGGED PANEL MODELS VS. 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,…,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 that

possibility is excluded. 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 vary over both individuals and time (possibly including lagged values), and another column vector of control variables z_i that vary over individuals but not over time.

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

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

where μ_t is an intercept that varies with time, β_1 and β_2 are scalar coefficients, δ_1 and γ_1 are row vectors of coefficients, ε_{it} is a random disturbance, and α_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:

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

where τ_t, is an intercept that varies with time, β_3, and β_4 are scalar coefficients, δ_2 and γ_2 are row vectors of coefficients, υ_{it} is a random disturbance, and η_i is a set of individual effects analogous to α_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: first, by the introduction of the unobserved individual effects, α_i and η_i and, second, by the

presumption that the coefficients for all variables are constant over time. The constancy assumption can certainly be relaxed, but will be maintained for now for simplicity.

The individual effects α_i and η_i may be specified either as random variables or as 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, ε_{it} and υ_{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 ε_{is} and υ_{is}. With respect to x and y, we cannot assume strict exogeneity because both variables appear as dependent variables. In fact, (1) and (2) together imply that ε_{it} and υ_{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), above, but with a few changes in meaning:

- x is typically a vector rather than a scalar
- x is usually not lagged
- α_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 α as a set of fixed constants ("fixed effects") is equivalent to allowing for unrestricted correlations between α and 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 ε_{is}. That is, the x variables are independent of all future values of ε but may be correlated with past values of ε.

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 non-linear equation. For example, if x is dichotomous, 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.

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 to a few dead ends.

3. 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 β 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 et al. 1988). To remove the fixed effects (α) from the equations, they reformulated the model in terms of first differences: $\Delta y_{it} = y_{it} - y_{it\text{-}1}$, $\Delta x_{it} = x_{it} - x_{it\text{-}1}$, and $\Delta w_{it} = w_{it} - w_{it\text{-}1}$. Note that first differencing not only removes α from the equation, but also z, the vector of time-invariant predictors. They then used lagged difference scores for y, x, and w as instrumental variables for Δy and Δx, and estimated the resulting system of equations 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 moments, 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 known 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 et al. 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 nonlinear 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.

4. 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 et al. (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".

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. dealt with the first issue by using the same device as Arellano and Bond—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. 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 points 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., 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 α_i and η_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 (α_i and η_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.

5. ML ESTIMATION VIA SEM

In this section, we show how Moral-Benito's 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 the 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

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

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 equations (1) and (2)? Although (1) and (2) can be estimated simultaneously, we follow the econometric tradition of focusing only on equation (1), while allowing (2) to determine certain constraints (or lack thereof) on $\mathbf{\Theta}$, the variance matrix for the exogenous variables.

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

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

and for $\mathbf{B}$,

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

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

- α with z
- α with all ε
- z with all ε
- all w with all ε
- all ε with each other
- x_{it} with ε_{is} whenever $s \geq t$

All other elements of $\mathbf{\Theta}$ are left unrestricted. Note that α is allowed to correlate with both w and x. And x is allowed to correlate with all *prior* realizations of ε, as a consequence of equation (2). The restriction that cov(α, 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.

6. EMPIRICAL EXAMPLE

As an example of how to implement ML-SEM for dynamic panel models, we re-analyze 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(wage) in each year

z = ED = years of education in 1976

The goal is to estimate equation (1), reproduced here

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

with x treated as predetermined, and w and z treated as strictly exogenous. To this point, we have assumed that x is a single variable, but the proposed method can easily handle multiple predetermined variables. For instance, we could make LWAGE predetermined rather than strictly exogenous.

By treating 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. It could, for example, be a logistic regression model, but it could also have any other functional form.

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 models. Like most SEM packages, CALIS requires that the data be in the "wide form" rather than the "long form".[1] 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, LWAGE1, LWAGE2,…, LWAGE7, etc. 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 1 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 α_i that is common to all equations. When CALIS encounters a variable name like ALPHA that is not on the input data

[1] Mplus can analyze "long form" data using its multilevel add-on, but the multilevel mode is not suitable for the dynamic panel models considered here.

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).

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 in order for ALPHA to truly behave as a set of fixed effects (Teachman et al. 2001, Allison and Bollen 1997, Bollen and Brand 2010). However, because ED does not vary over time, the correlation between ALPHA and ED is not identified, so it is constrained be 0 in Line 9. Lines 10-13 allow the error term ε 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.[2]

FIGURE 1 ABOUT HERE

By default, CALIS allows the intercept to differ for each equation, which is equivalent to allowing for an unrestricted effect of time itself. 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, in order to produce results that can be directly compared with AB

[2] 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.

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.

Table 1 displays the 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.

TABLE 1 ABOUT HERE

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 statistic comparing the fitted model with a "saturated" 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.

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.[3] 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 squared 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.[4] 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.

7. 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:

$$
\begin{aligned}
y_{i1} &= c\alpha_i + gu_{i1} \\
x_{i1} &= c\eta_i + gv_{i1} \\
y_{it} &= \beta_1 x_{it-1} + \beta_2 y_{it-1} + c\alpha_i + gu_{it} \\
x_{it} &= \beta_3 x_{it-1} + \beta_4 y_{it-1} + c\eta_i + gv_{it}
\end{aligned}
$$

for i=1,…, N and t=2,…,T. The time-invariant, unmeasured components α_i and η_i, were generated as bivariate standard normal variates with correlation ρ. The time-specific disturbances

[3] Stata, Mplus, and lavaan report somewhat lower values of these measures because they define the baseline model in a different way.
[4] The xtdpd model also included dummy variables for time in order to ensure comparability with the ML-SEM estimates.

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.

TABLE 2 ABOUT HERE

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, 1000 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 to estimate the models. The **sem** command[5] 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.[6] Program code for the Monte Carlo simulations is available in an online appendix.

We will focus on the estimates for β_1, the cross-lagged effect of *x* on *y*, and β_2, the autoregressive effect of *y* on *y*. For each of those parameters and for each condition, Appendix Table A.1 reports the mean and the standard deviation of the estimates for β_1 across the 1000 samples, as well as the "coverage"—the proportion of nominal 95 percent confidence intervals (calculated in each sample using the conventional normal approximation) that actually include the true values. If a method is performing well, the coverage should be close to .95.

[5] We actually used the user-written command, **xtdpdml**, which serves as a simplifying shell for the **sem** command. For details, see Williams et al. (2016)

[6] The ML models were less restrictive than the 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 model 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.

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% 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 the ML-SEM. To avoid that, we also excluded the same non-convergent 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 β_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% 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% 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 β_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 β_2 under all conditions. Every 95% 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% 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 β_2 is at or close to 1.0.

The bias in AB for β_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 estimated mean squared error for ML-SEM to the estimated mean squared error for AB. Mean squared 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 β_1 ranged from .83 (AB did 17% worse) to 1.12 (AB did 12% better), with a median of .96. So there was no clear winner for the cross-lagged effect. For β_2, however, the relative efficiency ranged from .34 (AB did 66% worse) to .87 (AB did 13% 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 β_1 and β_2 as a function of the number of time points:

TABLE 3 ABOUT HERE

For both β_1 and β_2, the relative efficiency of AB declines with the number of time points, although the decline is much more precipitous for β_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 achieved approximate stationarity by generating data from 1,000 time points. Then, for T=4, 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 β_1 regardless of the number of time points. For β_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.

TABLE 4 ABOUT HERE

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 for β_2 but not for β_1. When the standard deviation of α (and η) is held constant at 1.0, increases in the standard deviation of ε (and v) are associated with declines in the relative efficiency of AB for β_2 but not for β_1. The direction is reversed when the standard deviation of ε is held constant and the standard deviation of α is varied—higher standard deviations of α result in *higher* relative efficiency of AB for β_2.

TABLE 5 ABOUT HERE

Table 6 shows how relative efficiency is affected by the value of ρ, the correlation between the two fixed effects, α and η. For β_1 there is no apparent trend. For β_2, however, the relative efficiency of AB increases substantially as the correlation goes from 0 to .90.

TABLE 6 ABOUT HERE

Finally, Table 7 shows how relative efficiency is affected by the magnitude of β_1, the cross-lagged coefficient. As in the last two tables, the relative efficiencies of AB estimates of β_1 are virtually unaffected. For β_2, relative efficiency increases substantially as β_1 gets larger.

TABLE 7 ABOUT HERE

There were three other factors that had no apparent effect on relative efficiency: the sign of β_4 (the cross-lagged effect of y on x), sample size, and the magnitude of β_2 (the autoregressive

coefficient). The absence of a relationship with sample size and β_2 is somewhat surprising. We expected ML-SEM to do better in smaller samples, and we expected AB to perform more poorly when β_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 be less robust 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. 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 β_1 and .619 for β_2.

8. 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. As shown in this paper, cross-lagged panel models can be estimated within this framework.

However, 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 paper, 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. And the initial observations of the dependent variable are be treated just like any other exogenous variables. Reciprocal 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. Moreover, for the autoregressive parameter, the ML-SEM estimator was substantially more efficient than the AB estimator under all conditions. For this parameter, the efficiency of AB relative to ML-SEM declined markedly as the number of time points increased.

The ML-SEM method can be extended in several ways, as described in detail in Bollen and Brand (2010). 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. As pointed out by Bollen and Brand (2010), one can even allow the coefficient of α_i, the fixed effect, to vary with time instead of being

constrained to 1 for every time point. This option is very 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 full information maximum likelihood. Unlike AB, full information maximum likelihood can easily handle missing data on predictor variables.

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 9 time varying variables and T=11, then the covariance matrix will be 101 x 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 unbalanced data.

As can be seen from Figure 1 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. Different packages have different defaults 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. In fact, a user-written Stata command called **xtdpdml** is already available for ML-SEM (Williams et al. 2016). This command radically reduces the programming needed for ML-SEM, and is no more difficult to use than the built-in Stata commands for AB estimation (**xtabond**, **xtdpd**, or **xtdpdsys**).

References

Ahn, S. C. and Peter Schmidt (1995) "Efficient Estimation of Models for Dynamic Panel Data." *Journal of Econometrics* 68: 5-27.

Allison, Paul D. and Bollen, K.A (1997) "Change Score, Fixed Effects, and Random Component Models: A Structural Equation Approach," paper presented at the Annual Meeting of the American Sociological Association.

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) *Fixed Effects Regression Methods for Longitudinal Data Using SAS.* Cary, NC: The SAS Institute.

Allison, Paul D. (2005b) "Causal Inference with Panel Data." Paper presented at the Annual Meeting of the American Sociological Association, August.

Allison, Paul D. (2009) *Fixed Effects Regression Models*. Thousand Oaks, CA: Sage Publications.

Anderson, T.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, M. and S. Bond (1991) "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations." *The Review of Economic Studies* 58: 277-297.

Browne, Michael W. (1984) "Asymptotically distribution-free methods for the analysis of covariance structures." *British Journal of Mathematical and Statistical Psychology* 37: 62-83.

Baltagi, Badi H. (2013), *Econometric Analysis of Panel Data.* Fifth Edition. New York: John Wiley & Sons.

Bhargava, Alok and J. Dennis Sargan (1983) "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods." *Econometrica* 51: 1635-1659.

Bentler, Peter M., and David G. Weeks (1980) "Linear structural equations with latent variables." *Psychometrika* 45.3 289-308.

Blundell, R. and S. Bond (1998) "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics* 87: 115-144.

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.” In *Handbook of Econometrics*, eds. Z. Griliches and M. Intriligator. Amsterdam: North-Holland, pp. 1247–1318.

Cornwell, Christopher and Peter Rupert (1988) “Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators.” *Journal of Applied Econometrics* 3: 149-155.

Duncan, O.D. (1969). "Some linear models for two-wave, two-variable panel analysis." *Psychological Bulletin* 72: 177-182.

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.” In *Handbook of Causal Analysis for Social Research*, ed. S.L. Morgan. New York: Springer, pp. 113-131

Grassetti, Luca (2011) “A Note on Transformed Likelihood Approach in Linear Dynamic Panel Models.” *Statistical Methods & Applications* 20: 221-240.

Halaby, Charles N. (2004) “Panel Models in Sociological Research: Theory into Practice.” *Annual Review of Sociology* 30: 507-544.

Hamaker, Ellen L., Rebecca M. Kuiper, and Raoul PPP Grasman (2015) “A Critique of the Cross-Lagged Panel Model." *Psychological methods* 20: 102-116.

Han, Chirok, and Peter C.B. Phillips (2013) “First Difference Maximum Likelihood and Dynamic Panel Estimation.” *Journal of Econometrics* 175: 35-45.

Holtz-Eakin, D., W. Newey and H.S. Rosen (1988) “Estimating Vector Autoregressions with Panel Data.” *Econometrica* 56: 1371-1395.

Hsiao, Cheng (2014) *Analysis of Panel Data*. Third Edition. 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: 107-150.

Jöreskog, Karl G. (1978) “Structural analysis of covariance and correlation matrices.” *Psychometrika* 43: 443-477.

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 C. M. Judd (1996). A general procedure for the estimation of interdependence. *Psychological Bulletin, 119,* 138-148.

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: 1013–1038

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: 451-472.

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*: 1417-1426.

Roodman, D. (2009) “A Note on the Theme of Too Many Instruments.” *Oxford Bulletin of Economics and Statistics* 71: 135–158

Teachman, Jay (2014) “Latent Growth Curve Models with Random and Fixed Effects.” In *Emerging Methods in Family Research*, pp. 3-17. Springer International Publishing, 2014.

Teachman, Jay, Greg J. Duncan, W. Jean Yeung, and Dan Levy (2001) “Covariance Structure Models for Fixed and Random Effects.” *Sociological Methods & Research* 30: 271-288.

Williams, Richard, Paul D. Allison and Enrique Moral-Benito (2016) “xtdpdml: Linear Dynamic Panel-Data Estimation using Maximum Likelihood and Structural Equation Modeling.” (available at http://www3.nd.edu/~rwilliam/dynamic/SJPaper.pdf).

Wooldridge, Jeffrey M. (2010) *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press

```
 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;
14   run;
```

Figure 1. SAS Program for Estimating Dynamic Panel Model with Fixed Effects.

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

	ML-SEM			Arellano-Bond		
Predictor	Estimate	SE	z	Estimate	SE	z
wks(t-1)	0.188	0.020	9.59	0.163	0.039	4.18
lwage(t-1)	0.588	0.488	1.20	-1.276	.462	-2.76
union(t-1)	-1.206	0.522	-2.31	-1.175	.513	-2.29
ed	-0.107	0.056	-1.89			

Table 2. Parameter Values for Monte Carlo Simulation

N	50, 100, **400**, 1600
T	4, **5**, 7, 10
ρ	0, .25, .50, **.75**, .90
β_1	0, **.25**, .50, .75, 1.00
β_2	0, .25, .50, **.75**, 1.00, 1.25
β_4	**-.25**, 0, .25
g	.75, **1.00**, 1.50, 2.00
c	.50, **1.00**, 1.50 2.00

Table 3. How Number of Time Points Affects Relative Efficiency of the AB Method

	β_1	β_2
T=4	1.040	0.827
T=5	0.944	0.708
T=7	0.939	0.505
T=10	0.832	0.337

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

	β_1	β_2
T=4	0.982	0.562
T=5	0.989	0.527
T=7	1.038	0.371
T=10	1.029	0.305

Table 5. How Relative Efficiency Depends on the Variance of ε and α.

	β_1	β_2
SD(ε)=.25	0.924	0.716
SD(ε)=1.0	0.944	0.708
SD(ε)=1.5	0.974	0.640
SD(ε)=2.0	0.964	0.570
SD(α)=0.5	0.981	0.629
SD(α)=1.0	0.944	0.708
SD(α)=1.5	0.989	0.814
SD(α)=2.0	0.974	0.863

Table 6. How Relative Efficiency Depends on ρ.

	β_1	β_2
ρ=0	0.936	0.378
ρ =.25	1.044	0.467
ρ =.50	0.881	0.535
ρ =.75	0.944	0.708
ρ =.90	0.942	0.759

Table 7. How Relative Efficiency Depends on β_1.

	β_1	β_2
β_1=0	0.978	0.483
β_1=.25	0.944	0.708
β_1=.50	0.959	0.750
β_1=.75	0.875	0.814
β_1=1.00	0.938	0.779

Appendix A. Results from Monte Carlo Simulations

Table A.1 Performance of ML and AB Estimators of β_1, the lagged effect of x on y.

			Maximum Likelihood			Arellano-Bond			
Condition	True[a]	Non-Converge[e]	Mean[b]	SD[c]	Coverage[d]	Mean[b]	SD[c]	Coverage[d]	Relative Efficiency
N=50	0.25	19	0.256	0.0975	0.934	0.240	0.0918	0.939	1.121
N=100	0.25	23	0.250	0.0664	0.946	0.242	0.0664	0.943	0.985
N=400	0.25		0.250	0.0314	0.949	0.249	0.0323	0.952	0.944
N=1600	0.25		0.250	0.0161	0.947	0.250	0.0164	0.952	0.964
β_2=0	0.25		0.250	0.0325	0.950	0.252	0.0324	0.955	1.004
β_2=.25	0.25		0.251	0.0322	0.947	0.252	0.0310	0.954	1.076
β_2=.50	0.25		0.251	0.0313	0.950	0.249	0.0318	0.949	0.969
β_2=.75	0.25		0.250	0.0314	0.949	0.249	0.0323	0.952	0.944
β_2=.90	0.25		0.251	0.0324	0.953	0.248	0.0343	0.943	0.890
β_2=1.0	0.25		0.249	0.0328	0.949	0.249	0.0346	0.949	0.900
β_2=1.25	0.25		0.249	0.0335	0.957	0.248	0.0345	0.941	0.942
β_1=0	0.00	2	0.000	0.0331	0.947	-0.002	0.0334	0.947	0.978
β_1=.25	0.25		0.250	0.0314	0.949	0.249	0.0323	0.952	0.944
β_1=.50	0.50		0.500	0.0310	0.953	0.498	0.0316	0.958	0.959
β_1=.75	0.75		0.750	0.0304	0.954	0.749	0.0325	0.957	0.875
β_1=1.00	1.00		1.001	0.0300	0.948	1.000	0.0310	0.948	0.938
β_4=-.25	0.25		0.250	0.0314	0.949	0.249	0.0323	0.952	0.944
β_4=0	0.25	4	0.250	0.0337	0.948	0.250	0.0336	0.949	1.007
β_4=.25	0.25	23	0.250	0.0311	0.950	0.251	0.0309	0.949	1.014
ρ=0	0.25	4	0.251	0.0325	0.947	0.248	0.0336	0.941	0.936
ρ =.25	0.25		0.250	0.0337	0.939	0.248	0.0329	0.949	1.044
ρ =.50	0.25		0.250	0.0315	0.954	0.247	0.0334	0.951	0.881
ρ =.75	0.25		0.250	0.0314	0.949	0.249	0.0323	0.952	0.944
ρ =.90	0.25		0.252	0.0321	0.940	0.249	0.0331	0.953	0.942

SD(ε)=.25	0.25	15	0.251	0.0305	0.950	0.249	0.0317	0.944	0.924
SD(ε)=1.0	0.25		0.250	0.0314	0.949	0.249	0.0323	0.952	0.944
SD(ε)=1.5	0.25	1	0.250	0.0325	0.954	0.247	0.0328	0.951	0.974
SD(ε)=2	0.25	13	0.250	0.0318	0.949	0.247	0.0322	0.954	0.964
SD(α)=0.5	0.25		0.250	0.0336	0.938	0.247	0.0338	0.937	0.981
SD(α)=1.0	0.25	21	0.250	0.0314	0.949	0.249	0.0323	0.952	0.944
SD(α)=1.5	0.25		0.251	0.0315	0.949	0.250	0.0318	0.955	0.989
SD(α)=2.0	0.25	2	0.251	0.0317	0.945	0.250	0.0322	0.941	0.974
T=4	0.25		0.250	0.0402	0.947	0.249	0.0394	0.948	1.040
T=5	0.25		0.250	0.0314	0.949	0.249	0.0323	0.952	0.944
T=7	0.25	8	0.251	0.0247	0.952	0.248	0.0254	0.949	0.939
T=10	0.25		0.250	0.0188	0.949	0.246	0.0203	0.941	0.832
chi-square	0.25		0.252	0.0315	0.941	0.250	0.0331	0.947	0.908

[a]True value of the coefficient in the model producing the data.
[b]Mean of 1000 parameter estimates.
[c]Standard deviation of 1000 parameter estimates
[d]Percentage of nominal 95% confidence intervals that include the true value.
[e]Number of non-convergent samples for ML.

Table A.2. Performance of ML and AB Estimators of β_2, the lagged effect of *y* on itself.

		Maximum Likelihood			Arellano-Bond			
Condition	True[a]	Mean[b]	SD[c]	Coverage[d]	Mean[b]	SD[c]	Coverage[d]	Relative Efficiency
N=50	0.75	0.752	0.0917	0.953	0.680	0.1021	0.890	0.549
N=100	0.75	0.752	0.0675	0.951	0.716	0.0760	0.914	0.657
N=400	0.75	0.750	0.0325	0.953	0.742	0.0377	0.950	0.708
N=1600	0.75	0.750	0.0159	0.955	0.748	0.0195	0.946	0.656
β_2=0	0.00	0.001	0.0327	0.948	-0.005	0.0377	0.955	0.738
β_2=.25	0.25	0.248	0.0322	0.964	0.241	0.0456	0.951	0.554
β_2=.50	0.50	0.499	0.0378	0.952	0.488	0.0492	0.947	0.554
β_2=.75	0.75	0.750	0.0325	0.953	0.742	0.0377	0.950	0.708
β_2=.90	0.90	0.901	0.0257	0.961	0.895	0.0289	0.952	0.769
β_2=1.0	1.00	0.999	0.0219	0.956	0.996	0.0243	0.940	0.798
β_2=1.25	1.25	1.250	0.0150	0.947	1.249	0.0160	0.942	0.867
β_1=0	0.25	0.752	0.0425	0.960	0.730	0.0577	0.936	0.483
β_1=.25	0.25	0.750	0.0325	0.953	0.742	0.0377	0.950	0.708
β_1=.50	0.75	0.750	0.0249	0.953	0.747	0.0286	0.952	0.750
β_1=.75	0.75	0.750	0.0208	0.956	0.748	0.0229	0.946	0.814
β_1=1.00	0.75	0.751	0.0175	0.948	0.748	0.0198	0.933	0.779
β_4=-.25	0.75	0.750	0.0325	0.953	0.742	0.0377	0.950	0.708
β_4=0	0.75	0.751	0.0303	0.953	0.743	0.0344	0.942	0.749
β_4=.25	0.75	0.749	0.0294	0.938	0.743	0.0323	0.938	0.788
ρ=0	0.75	0.753	0.0418	0.948	0.727	0.0640	0.919	0.378
ρ =.25	0.75	0.752	0.0380	0.948	0.736	0.0540	0.929	0.467
ρ =.50	0.75	0.750	0.0352	0.948	0.737	0.0464	0.925	0.535
ρ =.75	0.75	0.750	0.0325	0.953	0.742	0.0377	0.950	0.708
ρ =.90	0.75	0.751	0.0305	0.950	0.742	0.0343	0.949	0.759
SD(ε)=.25	0.75	0.751	0.0254	0.942	0.745	0.0296	0.944	0.716
SD(ε)=1.0	0.75	0.750	0.0325	0.953	0.742	0.0377	0.950	0.708
SD(ε)=1.5	0.75	0.751	0.0455	0.951	0.731	0.0536	0.935	0.640
SD(ε)=2	0.75	0.750	0.0519	0.965	0.723	0.0632	0.934	0.570
SD(α)=0.5	0.75	0.752	0.0552	0.948	0.725	0.0650	0.919	0.629
SD(α)=1.0	0.75	0.750	0.0325	0.953	0.742	0.0377	0.950	0.708
SD(α)=1.5	0.75	0.751	0.0232	0.946	0.747	0.0256	0.943	0.814
SD(α)=2.0	0.75	0.751	0.0176	0.951	0.749	0.0189	0.961	0.863
T=4	0.75	0.752	0.0483	0.944	0.744	0.0529	0.946	0.827
T=5	0.75	0.750	0.0325	0.953	0.742	0.0377	0.950	0.708
T=7	0.75	0.750	0.0224	0.938	0.739	0.0295	0.928	0.505
T=10	0.75	0.750	0.0157	0.950	0.736	0.0231	0.898	0.337
chi-square	0.75	0.750	0.0317	0.959	0.744	0.0398	0.941	0.619

[a]True value of the coefficient in the model producing the data.
[b]Mean of 1000 parameter estimates.
[c]Standard deviation of 1000 parameter estimates
[d]Percentage of nominal 95% confidence intervals that include the true value.

Appendix 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) noxconditional
```

Explanation

- Line 1 reads the wide-form data into memory
- Lines 2-13 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.

This model can also be estimated in Stata with the user-written **xtdpdml** command, which acts as a simplifying shell for the **sem** command (Williams 2016). The code for this example is:

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

Mplus

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

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 much 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.

lavaan (R package)

```
 1   wage <- read.table("C:/wagenames.txt",header=T)
 2   wagemod <-'
 3   alpha =~ 1*wks2 + 1*wks3 + 1*wks4 + 1*wks5 + 1*wks6 + 1*wks7
 4   wks2 ~ a*wks1 + b*union1 + c*lwage1 + d*ed
 5   wks3 ~ a*wks2 + b*union2 + c*lwage2 + d*ed
 6   wks4 ~ a*wks3 + b*union3 + c*lwage3 + d*ed
 7   wks5 ~ a*wks4 + b*union4 + c*lwage4 + d*ed
 8   wks6 ~ a*wks5 + b*union5 + c*lwage5 + d*ed
 9   wks7 ~ a*wks6 + b*union6 + c*lwage6 + d*ed
10   wks2 ~~ union3 + union4 + union5 + union6
11   wks3 ~~ union4 + union5 + union6
12   wks4 ~~ union5 + union6
13   wks5 ~~ union6
14   alpha ~ wks1+lwage1+lwage2+lwage3+lwage4+lwage5+lwage6+
15      union1+union2+union3+union4+union5+union6
16   union6 ~ ed+wks1+lwage1+lwage2+lwage3+lwage4+lwage5+lwage6+
17      union1+union2+union3+union4+union5
18   union5 ~ ed+wks1+lwage1+lwage2+lwage3+lwage4+lwage5+lwage6+
19      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   wagefit <- sem(wagemod,data=wage)
25   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 =~ means "is measured by".
- Lines 4-9 specify the six regression equations. The symbol ~ 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 ~~ 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.
- Line 24 calls the **sem** function which actually fits the model, using the **wage** data and the **wagemod** model specification.
- Line 25 reports the estimates and associated statistics.