Els UK Jobcode: HLR Ch29-P370481 16-7-2007 5:27p.m. Page:475 Trim:7.5in×9.25in Floats: Top/Bot TS: Integra, India

Chapter 29

Linear panel analysis

Steven E. Finkel

1 Introduction

Linear panel analysis refers to the statistical models and methods appropriate for the analysis of continuous or quantitative outcomes with data collected on multiple units (i.e., individuals, schools, or countries) at more than one point in time. This means that linear panel analysis is concerned with the various ways of analyzing change in continuous variables, in describing different patterns of change for different units, in modeling why some units change more than others and what variables are responsible for these differences. These methods may be distinguished from other kinds of longitudinal analyses, such as loglinear, transition or Markov models for analyzing over-time change in categorical outcomes, event history or duration models for analyzing temporal processes leading to the occurrence of specific events, and time-series methods for the analysis of change in continuous outcomes for a single unit over a relatively long period of time. Hence, what characterizes linear panel analysis is a focus on *continuous outcomes* for *multiple units* at *multiple points*.

It is conventional within this general rubric to make one further distinction. In some panel datasets, *time* is dominant, i.e., relatively few units have been observed for relatively long periods of time. In other data, "*N*" is dominant, i.e., many units have been observed for relatively few points in time. Although the two kinds of data have the same formal structure, time-dominant data, sometimes referred to as "time-series cross-sectional data," is typically analyzed with statistical methods rooted in the time-series econometric tradition (see Beck and Katz, 1995; Greene, 2003, see also Worrall, Chapter 15 in this volume). This chapter is concerned with the statistical methods used for the analysis of "N-dominant" panel data, typically with observations on hundreds or possibly thousands of units observed at two or more "waves," in time. Examples of "panel data" are the National Election Studies (NES) panels that track thousands of the same respondents across multiple presidential and congressional elections in 1956–58–60, 1972–74–76, and 2000–2002–2004, the multiwave US Panel Study of Income Dynamics (PSID), and the German Socio-Economic Panel (SOEP) covering some twenty-one waves of observation since 1986.[1]

There are several important motivations for analyzing panel versus cross-sectional data. Consider the hypothesis that economic performance contributes to the consolidation or stability of democratic regimes. This hypothesis

[1] Available at www.umich.edu/~nes/, psidonline.isr.umich.edu/, and www.diw.de/english/sop/ respectively.

Font: Melior Font size: 10/12pt Margins:Top:3pc Gutter:42pt T.Area:161mmx190mm 1 Color COP: open recto Depth: 45 lines

could be tested with cross-sectional data by predicting some sample of countries' level of democracy at a given point in time with relevant economic indicators, perhaps including additional variables related to the social or political characteristics of the countries as statistical controls. Statistically significant effects of the economic variables would then be taken as supporting the hypothesis; if the researcher was confident enough in the specification of the model or otherwise sufficiently bold, he or she might even claim that economic performance had a *causal* effect on countries' level of democracy.

Yet serious obstacles exist for successful causal inference in the cross-sectional context, many of which can be dealt with more easily through the analysis of panel data. First, cross-sectional data contains no direct measure of changes in Y nor changes in X—it is implicitly assumed in such designs that by comparing units with higher and lower values on the independent and dependent variables simulates "changes" in X (economic performance) and their impact on changes in Y (democracy). But what has been conducted is simply a fixed comparison of countries or units at a given point in time, which says little directly about what happens when individuals or units change. Far better from the point of view of testing theories of social, psychological, or political change by directly observing change over time, and of course that is what longitudinal or panel data provides at its very foundation.

But the problems of causal inference in cross-sectional designs go beyond the lack of direct observation of change. First, panel data offer decided advantages in dealing with the problem of *spurious* relationships between variables, such that some outside variable Z not considered by the researcher is actually responsible for the observed relationship between X and Y. Of course, theoretically-relevant variables that are observed in a given dataset should always be incorporated into any statistical model to attempt to avoid this kind of bias. But with panel data, the researcher has the ability under some conditions to control for *unmeasured* variables that may be confounding the observed relationship between X and Y. In the economic performance and democracy example, variables such as a country's political culture or the degree of "entreprenuerialism" in the population may cause both economic and democratic outcomes, and to the extent that these variables are unobserved in the typical cross-national dataset, the estimated effect of economic performance on democracy will suffer from omitted variable bias. Panel data are no panacea for this problem but they offer the researcher far greater options for incorporating "unobserved heterogeneity" between units into statistical models, and controlling their potentially damaging effects, than is possible in cross-sectional analyses.

A second obstacle in cross-sectional data is that because X and Y are measured at the same time, it is difficult to determine which one of the variables can be presumed to "cause" the other. Does economic performance lead to changes in a country's level of democracy, or does the country's level of democracy lead to higher levels of economic performance (or both)? With panel data, the researcher can track the impact of changes in X, or the level of X, at *earlier* points in time with *later* values of Y, and estimates of the effects of earlier values of Y on subsequent values of X can similarly be obtained. The ability to estimate dynamic models of reciprocal causality between X and Y by exploiting intertemporal change in both variables is one of the important advantages of longitudinal data analysis in general, and the methods used for estimating these kinds of models for continuous outcomes form a large part of the statistical toolkit for linear panel analysis.

Third, it is also the case that successful causal inference depends on the accurate measurement of the variables in any statistical model.

As is well-known, random measurement error in the independent variable of a bivariate model will attenuate the estimated effect of X on Y, with the direction of bias in multivariate models being completely indeterminate (Bollen, 1989; Wheaton, et al., 1977). While panel data are certainly not immune to these measurement problems, the longitudinal information contained in such data offer the researcher greater ability to model measurement error in the variables than is typically the case with cross-sectional data. As is the case with models of reciprocal causality, measurement error modeling can proceed with fewer potentially unrealistic and restrictive assumptions in panel analyses, thus strengthening confidence in the estimated causal linkages between variables.

Currently, there are several general methodological approaches or frameworks within which researchers conduct linear panel analysis. One stems from the econometric tradition, and focuses most explicitly on the problem of unobservables in the causal system (Baltagi, 2005; Frees, 2004; Hsiao, 2002; Woolridge, 2002). The analysis in this framework typically involves *pooling* or stacking the data across waves. This means that each row of data contains information on X and Y from a particular unit at only one of the panel waves, with information from unit (case) 1 at waves 1, 2, ... through time T, followed in the dataset with information from case 2 at waves 1, 2, ... T until the last row contains information on X and Y from the Nth case at the last wave (T) of observation. The pooling procedure thus yields N*T total observations for analysis. This setup allows the researcher considerable power in estimating a variety of panel models, including those where Y_t is predicted by X_t (or X_{t-1}) as well as by an additional factor U that represent *unobserved* variables or influences on Y for a particular unit i that remain stable over time. These factors give rise to what is referred to in the econometric literature as *unobserved heterogeneity*. In these models, a single effect of X_t or X_{t-1} on Y_t is typically produced that is purged of the potentially confounding influence of the unobservables, with the information contained in the panel data being used in alternative models to "sweep out" or take into account the unobserved heterogeneity in different ways, and to model possible dependence in the units' idiosyncratic error terms over time.

Another approach to panel analysis stems from the structural equation modeling (SEM) and path analysis traditions in sociology and psychology (Bollen, 1989; Duncan, 1975; Finkel, 1995; Kenny, 1979; Kessler and Greenberg, 1981). In this framework, a separate equation for each dependent (endogenous) variable at each panel wave of observation is specified with a set of independent variables, which may themselves be either exogenous, or unpredicted by other variables in the model, or endogenous variables that are caused elsewhere in the overall causal chain. Thus Y observed at time 2 of the panel may be predicted by Y at time 1 (the "lagged endogenous variable") and a series of time 1 Xs, Y at time 3 may be predicted by Y and the Xs at time 2, and so forth. The resulting series of equations are then, given appropriate assumptions about error processes and the distribution of observed variables, typically estimated simultaneously through maximum likelihood or related methods in software packages such as LISREL, EQS, MLWin, or SAS. The SEM approach is often extended by including additional equations to model random measurement error in observed indicators of the exogenous and endogenous variables, resulting in a set of measurement equations linking *latent* variables with one or more error-filled indicators, and a set of structural equations linking the latent variables together in the presumed causal system. Such models may also be extended to test alternative causal lag structures in the model, such that variables may be presumed to exert causal influence on endogenous variables either simultaneously (i.e., at the same wave of observation), or lagged by one or more time periods. The SEM approach is particularly useful for the panel analyst in estimating a variety of reciprocal causal effects models that guard against the possibility of biases induced by measurement error in the variables, and that allow the flexible testing of alternative lag structures.

Until recently, panel researchers have typically conducted their analyses within one framework or the other, with the choice depending to some extent on the leanings of the particular social science or other discipline, and to some extent on the nature of the substantive problem and the threats to causal inference that were presumed to be especially serious in a given research area (i.e., measurement error versus unobserved heterogeneity). But much work has been done in recent decades to bring the two traditions together, or perhaps more accurately, to bolster the toolkit of each framework so that it incorporates some of the major strengths of the other. Thus the SEM approach has recently been extended to incorporate models of unobserved heterogeneity, while the econometric approach has expanded to include more extensive models of measurement error, reciprocal causality, and dynamic processes than were typically the case in decades past. While it would not be accurate to state that there are no remaining differences between the SEM and econometric approaches, it is nevertheless the case that panel researchers now have the tools to at least attempt to overcome the common threats to successful causal inference using either analytical framework.

In what follows, I shall provide an overview of the basic econometric and SEM methods used for linear panel analysis. I shall begin with the problem of unobserved heterogeneity, outlining the "fixed" and "random" effects models that deal with this problem. The discussion will then turn to the SEM approach for estimating dynamic models of reciprocal causality and then models with measurement error in the observed indicators. Finally, I will outline briefly more advanced models within the econometric and SEM traditions that attempt to incorporate both unobserved heterogeneity and dynamic processes in order to strengthen the causal inference process.[2]

2 Unobserved heterogeneity models for linear panel analysis

Consider a simple, single equation model of the effects of some independent variables X_1, $X_2 \ldots X_j$ on a dependent variable Y, each measured at **T** points in time on a sample of **N** individuals. There are no reciprocal effects specified and perfect measurement is assumed for all variables. We can then write the basic panel model for the relationship between the *J* number of X variables and Y for the *i*th case at a given *t* point in time as:

$$Y_{it} = \alpha + \beta_1{}^* X_{1it} + \beta_2{}^* X_{2it} + \cdots \beta_j{}^* X_{Jit} + \varepsilon_{it} \quad (1)$$

This, of course, is the usual model analyzed in cross-sectional research (where T=1), using ordinary least squares (OLS) under the classical assumptions regarding the distribution of the idiosyncratic error terms ε_{it}. With panel data, though, it is clear that the classical assumptions regarding the error term are unlikely to be met, for the simple reason that the observations are not independent across time. That is, the error term for case *i* at time *t* is likely to be related to the error term for the same case at time $t+1$, $t+2$, etc. One strategy for handling this problem is to maintain the structure of equation (1) and estimate parameters with some variant of "robust" standard errors that

[2]This chapter will not discuss, except in passing, a third tradition for panel analysis, known variously as "hierarchical growth" or "mixed" models, or "latent growth" models. See Chapters 33–35 in this volume, Bollen and Curran (2005), and Singer and Willett (2002).

allow casewise dependence or other deviations from the standard assumptions.[3]

Most panel analysts, however, prefer to model the sources of the temporal dependence between observations of a given case more directly. One relatively simple setup is to assume that the dependence is produced by some stable, unobserved factor or factors **U** that are unique to a given unit and are also related to Y. This means that the "true" model (with U summarized as a single variable) is:

$$Y_{it} = \alpha + \beta_1{}^*X_{1it} + \beta_2{}^*X_{2it} + \cdots \beta_j{}^*X_{jit} + U_i + \varepsilon_{it} \quad (2)$$

For example, if Y is a country's extent of political repression and the Xs represent some measures of democracy and economic performance, the U term may encompass factors such as a country's "culture," history of violence, degree of ethnic homogeneity, and the like, all or some of which may not have been measured or have been otherwise available for inclusion in the analysis. If the model were at the individual level, such that Y represents, for example, a person's knowledge about politics, the U term may encompass intrinsic intelligence, motivation, or family socialization processes that produced individuals who are generally higher or lower on knowledge than would be expected from the values of the observed Xs in the model. The U_i are referred in the econometric literature as individual-specific, or unit effects.

In cross-sectional analyses, the unobserved stable U factor(s) are folded into the equation's unknown error term and the analyst can do little if anything about it. If the U are uncorrelated with the X_j observed variables, this would mean that the equation's explained variance is less than it might otherwise be, with inefficiency in the estimation of the standard errors for the β_j regression coefficients. But if (as seems likely), the U are related in some way to the observed $\mathbf{X_j}$, then the corresponding estimates of β_j will be biased. The potentially confounding effects of omitted variables is, of course, one of the most serious problems in nonexperimental research of any kind.

2.1 Fixed effects models

With panel data, however, at least some headway in attacking the problem can be made. Equation (2) may be rearranged to show that the presence of U implies that each unit has its own intercept $(\alpha + U_i)$, where U_i may be viewed as all the stable unit-level factors that lead that case to be larger or smaller than the overall average intercept (α) for the dependent variable in the sample. This suggests that one way of dealing with the problem with panel data would be to estimate equation (2) with OLS by including a dummy variable for N-1 units, with the coefficient on the dummies representing the individual-specific effects for each case (relative to an omitted baseline unit). This procedure is referred to as the LSDV (least squares dummy variables) method, and produces consistent estimates of the β_j coefficients for the X_j, controlling for stable unit effects that push the intercept for that case above or below the common (or baseline) intercept α. In this way we see that panel data can use the multiple observations on cases over time to begin to control for the effects of some kinds of variables that are not measured or observed in a given dataset.

The LSDV method is not usually applied, however, due to the need to include a potentially enormous number of dummy variables in large-N panel studies. A more common approach is to first express equation (2) in terms of the unit-level means of all observed variables, as in

[3]This is the motivating logic behind the class of panel estimators known as "generalized estimating equations (GEE)." See Hardin and Hilbe (2003) and Zorn (2001), and also Hardin and Hilbe, Chapter 31 in this volume.

$$\overline{Y}_i = \alpha + \beta_1 \overline{X}_{1i} + \beta_2 \overline{X}_{2i} + \cdots \beta_J \overline{X}_{Ji} + \overline{U}_i + \overline{\varepsilon}_i \quad (3)$$

This formulation is called the "between" equation, as all "within-unit" variation over time is averaged out, leaving a model that only considers variation between the individual units. Subtracting (3) from (2) yields the *fixed effects* (FE) model that sweeps the U terms out of the equation altogether:

$$Y_i - \overline{Y}_i = \beta_1(X_{1it} - \overline{X}_{1i}) + \beta_2(X_{2it} - \overline{X}_{2i}) + \cdots \beta_J(X_{Jit} - \overline{X}_{Ji}) + (\varepsilon_{it} - \overline{\varepsilon}_i) \quad (4)$$

Given that U_i is constant over time, $U_i = \overline{U}_i$ and hence drops out from equation (4). Estimating the resulting regression with OLS (with an appropriate adjustment for the standard errors due to the N number of unit means figuring in the estimation) yields consistent estimates of the β_j effects of the observed X_j variables.[4] The fixed effects estimator is also referred to as the "within" estimator, as it considers *only* variation in X and Y around their unit-level means, as all "between unit" variation (i.e., differences in $\overline{X}_i$ and $\overline{Y}_i$) are eliminated through the mean-differencing procedure.

A related approach to sweeping out the individual-level U_i is simply to lag equation (2) by one time period, as in:

$$Y_{it-1} = \alpha + \beta_1{}^* X_{1it-1} + \beta_2{}^* X_{2it-1} + \cdots \beta_j{}^* X_{jit-1} + U_{it-1} + \varepsilon_{it} \quad (5)$$

[4]That is, since N degrees of freedom are used in the calculation of the unit means, the appropriate df for the fixed effects model is NT-N-k-1. The constant term α in (2) may also be recovered by adding the grand mean for Y and each of the X_j to each of the deviation expressions in (4). Both of these adjustments are implemented automatically in STATA and other software packages for estimating econometric panel models.

Subtracting this equation from (2) yields the *first difference* (FD) equation that also produces consistent estimates of the β_j:

$$Y_i - Y_{it-1} = \beta_1(X_{1it} - X_{1it-1}) + \beta_2(X_{2it} - X_{2it-1}) + \beta_J(X_{Jit} - X_{Jit-1}) + (\varepsilon_{it} - \varepsilon_{it-1}) \quad (6)$$

While both FD and FE methods are consistent, the fixed effect procedure is more commonly applied in multiwave panel data, as it makes use of *all* of the over-time variation in X in its calculations, and provides a more parsimonious method of expressing X and Y at a given time period as deviations from an overall mean value. Subtracting X_{t-1} in the first difference approach is inefficient in the sense that only one of the lag values of X is used to difference out the unit effect, and the one that is used is to some extent arbitrary (why not subtract X_{t-2} or X_{t-3}?).

The FE and FD procedures are relatively simple, yet powerful methods for panel models controlling for unobserved heterogeneity between cases. As Allison (1994) and Halaby (2004) have shown, these methods are also equivalent to the popular "difference in difference estimator" (DID) for assessing treatment effects in quasi-experimental and policy research. If, as is plausible in almost all nonexperimental research situations, there are unobserved differences between members of the treatment and control groups, then it may be the case that such differences and not the treatment itself may produce the observed differences between the two groups. Constructing the difference between both the treatment group's and control group's pre-test score and post-test scores, or, in multiwave data, between the treatment group's and control groups's pre-test and post-test mean-deviated scores, essentially removes the unobserved (stable) differences between the two groups from consideration. The "difference in differences" between the treatment and control groups is then the pure effect of the treatment, controlling for these unobserved differences (and any observed covariates that are included in the model as additional controls).[5] Allison (1994) shows how this logic can be applied to estimating the effect of a variety of treatments with panel data, where the treatments may be given to individuals who were not randomly assigned to different conditions in quasi-experimental conditions, or where the "treatments" were simply the experience of some life event such as divorce, military service, unemployment, or job promotion between panel waves. Whenever stable, unobserved factors at the unit level may be related to the unit's likelihood of experiencing of these events and also related to the unit's value on some outcome variable, the FE or FD approach sweeps out the (stable) unobservables to produce consistent estimates of the effects of the event or treatment itself.[6]

Despite their clear strengths in eliminating the potentially confounding effects of unobserved hetereogeneity, the FE and FD models have certain features that render them problematic for some panel analyses. First, the differencing or mean-difference process eliminates not only the stable unobserved factors from consideration, but also all stable *observed* factors, such that the researcher can say nothing about the effects on the dependent variable of characteristics of units, countries, individuals, or other units that do not change over time. In many research situations, the effects of (nearly) stable individual-level factors such as education or sex, or, at the country level, stable political characteristics such as electoral systems or institutional arrangements are of prime theoretical interest, and FE/FD models are less attractive.[7] Second, as was noted above, the FE/FD method uses 1/T of the available degrees of freedom, which in short panels can be a relatively high cost. Third, it is also the case that the estimates of the unit-effect in short-term panel studies are based on only a few waves of observation, and hence may be unreliable to the extent that chance factors produce a few consistently high or low readings on the dependent variable for a given case over time. Since FE/FD methods take these unit-effects as given, they potentially overstate the "true" amount of temporal dependence produced by stable unobservables.[8]

[5]The logical equivalence of the DID estimator and the FE estimator is ensured only when a dummy variable for the wave of observation is also included in the FE model. This presents no special statistical problems and results in the "two-way" fixed effects model. This addition is necessary because the FE (and FD) models eliminate from consideration those observations that do not change on X over time; the inclusion of the time dummies thus captures the general time effect on Y that occurs for both treatment and control groups, with the additional impact of the treatment being given by its coefficient.

[6]It is also possible to test for the statistical significance of the unit effects as a whole through estimation of a nested F test that compares the R-squared of models with and without the inclusion of the N-1 unit effects.

[7]Interaction effects between time and stable unit-level factors may, however, enter FE models, so that the researcher may estimate how the impact of a stable variable changes across waves of the panel.

[8]It is also argued that FE methods are more appropriate in instances where the analyst wishes to make inferences conditioned on the observed units in the sample, as in analyses at the country level where, for example, Germany's unit effect is of interest and would be the same no matter how many different country samples are drawn. In instances where large numbers of individuals are sampled randomly from some population (and hence the specific individual effects are of less interest), the random effects approach to be discussed subsequently is arguably more appropriate. See Allison (1994) and Teachman et al. (2001) for counter-arguments on this point, claiming that FE and RE are simply alternative ways of dealing with the "nuisance" posed by the presence of the unit effects in (2) above.

2.2 Random effects models

An alternative approach to estimating equation panel models with unobserved heterogeneity, the *random effects* or *random intercepts* (RE) model overcomes some of these deficiencies, though not without some additional costs of its own. Consider the two sources of unobserved error in equation (2): the unobserved unit effect (Ui) and the unit-time-specific idiosyncratic error term ε_{it}. In the "random effects" approach, both of these sources of error are treated as the realization of random processes and assumed to be independent, normally-distributed variables (with variances denoted as σ_u^2 and σ_ε^2 respectively). Some units have a higher intercept on the dependent variable because of a large random U term, some units have a lower intercept because of a small random U term; added to this error is the idiosyncratic error ε_{it} which produces further deviations from the linear prediction of Y from the U and the B_jX_{jj} terms of the model. The only additional unknown in this setup compared to the pooled OLS panel model of equation (1) is σ_u^2, and thus one immediate benefit of the random effects model is that it saves a significant number of degrees of freedom compared to its FE or FD counterparts.

Estimation of the RE model proceeds under two assumptions. First, the two components of the composite error term, U_i and the ε_{it}, are assumed to be unrelated, otherwise no separate estimate of each would be possible. Second, and more important, both error terms must be assumed to be unrelated to the included X variables in the model, i.e., $E(X_{ji}U_i) = E(X_{ji}\varepsilon_{it}) = 0$. This is of course problematic, in that the setup assumes away the possible correlation between X and the unit-effect unobservables that prompts many panel analyses in the first place(!) The situation need not be so dire, however, as will be discussed in more detail below. Given these assumptions, the composite error term of the model, $U_i + \varepsilon_{it}$, has a fixed structure over time, with the variances (diagonal elements) being equal to $\sigma_u^2 + \sigma_\varepsilon^2$, and the covariances (off-diagonals) being equal to σ_u^2 for every time period. As such, estimation can proceed through feasible generalized least squares (FGLS) methods that weight the model by the inverse of the error variance–covariance matrix, in this case the weight θ calculated as:

$$\Theta = 1 - \frac{\sigma_\varepsilon^2}{\sqrt{T\sigma_u^2 + \sigma_\varepsilon^2}} \tag{7}$$

Once an estimate of θ is obtained through manipulation of the "within" and "between" regressions of equations (3) and (4) above[9], the model is then transformed as:

$$\begin{aligned} Y_i - \Theta\overline{Y}_i = \beta_1(X_{1it} - \Theta\overline{X}_{1i}) + \beta_2(X_{2it} - \Theta\overline{X}_{2i}) \\ + \cdots \beta_J(X_{Jit} - \Theta\overline{X}_{Ji}) + (U_{it} - \Theta\overline{U}_i) \\ + (\varepsilon_{it} - \Theta\overline{\varepsilon}_i) \end{aligned} \tag{8}$$

with this equation's error term (comprised of the two last terms in parentheses) now having constant variance and zero correlation across time units for each case.

Several features of the RE model are especially noteworthy. First, it can be seen that as θ approaches 1, it means that more and more of the composite error variance is made up of unit-level (between) variance σ_u^2. In the unlikely event that θ equals 1, all of the error variance is unit-level variance, and the RE estimator reduces to the FE mean-differenced estimator of the β. As θ gets closer to 0, more and more of the error variance is made up of the random σ_ε^2 component, with no unit-level variance to take into account, and the RE estimator thus reduces to the pooled OLS approach represented by equation (1). So the RE approach represents something of a middle ground between

[9]An estimate of σ_ε^2 is obtained from the "within" regression of (4), as all of the unit-level variation has been purged from this model. An estimate of σ_u^2 is obtained by manipulating the error term of the "between" regression of (3), which produces the error term $(\sigma_u^2 + \sigma_\varepsilon^2/T)$.

the FE and OLS models, weighted toward one or the other depending on how much of the error variance is comprised of unit-specific versus idiosyncratic components.[10] It may also be said that RE "adjusts" or "shrinks" the FE estimator back toward pooled OLS to the extent that the unit-level effects in general are either small, relative to overall error, or are unreliable due to a relatively small T (as can be seen from the denominator of equation (7)).[11]

Second, the RE model produces estimates of the effects of both changing *and* stable independent X variables, as its estimation equation (8) does not result in the elimination of any stable variable so long as θ is not 1 (or very close to 1, when estimates of stable variables will tend to be very imprecise). The ability to model Y as functions of both changing and unchanging X variables over time is one of the major advantages of the RE approach to unobserved heterogeneity; as noted, however, this (and other) advantages of RE may be enjoyed only to the extent that the assumptions of the model hold, i.e., that there is no correlation between the X_j and the U_i.

For this reason, there has been much debate in the panel literature over the applicability of FE versus RE. The "Hausman" test provides one way of adjudicating the dispute, by providing a test statistic to assess the significance of the difference between the FE and RE estimates. The logic behind the test is that, if the assumptions of the RE model hold, then FE and RE are two different ways to arrive at consistent estimates of the β, but RE is more efficient. If the assumptions of the RE model do not hold, then RE will be inconsistent, while FE will always be consistent. Thus, one should see similar estimates of the β_j if the RE assumptions hold, and different ones if they don't. Hausman (1978) showed that the statistic

$$\frac{\beta(FE) - \beta(RE)}{\mathrm{var}\beta(FE) - \mathrm{var}\beta(RE)}$$

is distributed as χ^2 with J degrees of freedom, with failure to reject the null hypothesis implying that the RE model is appropriate. If the Hausman null hypothesis is rejected, then it may be concluded that there is some violation of the RE assumptions and that a likely nonzero correlation between the X_j and the U_i exists which the analyst should take into account. One way to do so is through the FE or FD methods that eliminate the U_i from consideration in the estimation equation altogether. Another is by including in a random effects version of (2) the *unit-level mean* for each time-varying independent variable (i.e., $\overline{X}_j$) as additional predictors (Skrondal and Rabe-Hesketh, 2004, pp. 52–53). The RE "problem" may thus be viewed as an omitted variable issue, with the unit effects being potentially correlated (at the unit-level) with the included explanatory factors. Once the time-varying unit-level means are included, this correlation is essentially controlled for in the model, with the resultant composite error term satisfying the assumptions for FGLS estimation.[12]

2.3 Example: The effects of civic education on political knowledge in Kenya, 2002–2003

We illustrate these models with panel data collected on 401 individuals interviewed at three

[10]The proportion of total variance comprised of unit-specific variance is also referred to as the "intraclass correlation coefficient."

[11]It can also be seen from equation (7) that the RE and FE converge as $T \rightarrow \infty$, so the issue of fixed versus random effects is not as relevant in large T panel (or time-series cross-sectional) studies.

[12]Plümper and Troeger (forthcoming) provide a similar method for incorporating unit-level means into a fixed effect model. In both cases, it is still necessary to assume that the time-invariant X_J are unrelated to $U_{i.}$. See Hausman and Taylor (1981) for an approach to this problem involving different assumptions about correlations between particular time-invariant X variables with the unit effects.

waves between February 2002 and June 2003 in Kenya, as part of a study evaluating the effects of attending civic education and democracy training workshops on changes in democratic knowledge, attitudes and participation in the run-up to the Kenyan December 2002 presidential elections.[13] A total of 210 respondents were interviewed before they attended a civic education (CE) workshop between February and April 2002, with 191 individuals serving as the control group, as they were selected to match the "treatment" group on place of residence, gender, age, and educational status. There were two follow-up interviews for all respondents, one in November, some 7–9 months after the workshop, and another in April–May 2003, about one year after the initial workshop took place. Our concern here is whether the workshop exposure (the "treatment") led to significant changes in respondents' knowledge of politics, measured with four questions asking the name of various elected officials (the Vice President and the Provisional Commissioner) and various institutional provisions in the Kenyan political system (e.g., the length of the term of office of the President and the procedures for amending the Kenyan constitution).

Given that individuals selected themselves into the "treatment" group, in that attendance at local civic education workshops was purely voluntary, it is likely that the treatment group differs from the control group on some measured variables, such as interest in politics, and also on unmeasured factors that may influence political knowledge such as intrinsic intelligence, motivation, openness to political reform, personal discussion networks, and the like. To control for these potentially contaminating effects, a series of unobserved heterogeneity models were estimated with the three-wave data, and the results are shown in Table 29.1.

The model in the first column of the table shows the pooled OLS model of (1) above, i.e., one that contains no unobserved heterogeneity term whatsoever. In this model, individuals exposed to civic education workshops are on average .32 higher on the dependent variable after exposure than individuals in the control group, holding interest, education, sex, and age constant. The dummy variables for wave 2 and wave 3 show that, controlling for other independent variables, there is a .31 increase in knowledge in wave 2 compared to wave 1 (the baseline wave of the panel), and a .18 increase in knowledge in wave 3 compared to wave 1 for all individuals, including those in the control group. In the fixed effects model of column (2), the unit-specific effect specified in the theoretical model of (2) above is swept out through the mean-differencing process, leaving the "within" regression of individual deviations from their own means. In this model, the effect of civic education falls to .21, approximately two-thirds of its magnitude in the pooled model, though still significantly different from zero. Note that in the FE model, there are no estimates for the time-invariant factors of education, sex, and age, as they drop out of the model (along with U_i) through mean-differencing. Thus the FE model shows that exposure to civic education has a significant impact on later political knowledge, controlling for observed and unobserved stable factors at the individual level that may be correlated with both CE exposure and with knowledge.

An alternative random effects model is shown in column (3) of the table. As can be seen, the estimates are much closer to the pooled OLS model in column (1) than the fixed effects estimates, with CE now having a .31 effect on knowledge. Estimates of the stable observed control variables are also similar to their OLS values, as they should be, given the relatively

[13]For more information on the overall study, including details on the sampling and questionnaire design, see Finkel (2003), available at www.pitt.edu/~finkel.

Table 29.1 Unobserved heterogeneity models, Kenya three-wave civic education study

Variable	*(1)*	*(2)*	*(3)*	*(4)*
	Pooled OLS	*Fixed effects*	*Random effects*	*Random effects with unit-level means*
Civic education	.32	.21	.31	.21
	(.05)	(.10)	(.07)	(.10)
Political interest	.27	.13	.24	.12
	(.04)	(.05)	(.04)	(.05)
Education	.20		.20	.19
	(.02)	(dropped)	(.02)	(.06)
Male	.47		.48	.43
	(.05)	(dropped)	(.06)	(.06)
Age	−.005		−.005	−.005
	(.002)	(dropped)	(.002)	(.002)
Wave 2 dummy	.31	.41	.33	.41
	(.07)	(.08)	(.07)	(.08)
Wave 3 dummy	.18	.28	.20	.28
	(.07)	(.08)	(.07)	(.08)
Treatment mean				.13*
	----	----	----	(.13)
Political interest mean				.32
				(.09)
Constant	.78	2.01	.85	.33
	(.15)	(.13)	(.17)	(.21)
Adj. R-squared	.26	.10	.27	.28
Degrees of freedom	1193	796	1192	1190
Intraclass correlation	----	.46	.18	.18
Estimated θ	----	----	.22	.22

All coefficients statistically significant at $p<.05$ except *. Standard Errors in Parentheses. $N = 401$.

small estimated value for θ (.22). This model, however, assumes no correlation between the treatment or other observed variables and the random unit-level error term; a Hausman test shows that this assumption is likely to be violated, as the null hypothesis of no difference between the FE and RE models can be rejected ($\chi^2 = 13.88$, $p<.01$). Thus the choice here is between a FE specification, or a random effects specification with the unit-level means controlled (model 4). In both instances the CE effect is estimated to be .21; the difference between the models, aside from their assumption about the distribution of the U_i, is that the RE model also provides information on the effects of the time-invariant control variables that are differenced out of the FE specification.[14]

[14]It should be noted that the specification of the effect of CE in all of the models in Table 29.1 is that of an additive, permanent change, such that CE leads to an increase in knowledge in the immediate time period after exposure, with the effects persisting over time. Moreover, the effect of time itself is modeled with dummy variables for each time period. Allison (1994) shows other ways to model the impact of events, as well as alternatives models of time effects. Random effects models that incorporate randomly-varying time-trends are also referred to as "hierarchical longitudinal growth models" (see Raudenbush and Bryk, 2002, and also Luke, chapter 33 in this volume).

3 Dynamic panel analysis

To this point, it has been assumed that all of the temporal dependence in the model structure was rooted in the presence of the unit-specific effect (be it fixed or randomly distributed). That is, the only reason that a Y_i response at time t would be correlated with a Y_i response at time $t+1$ is the presence of the stable U_i term, which influences the response for the individual unit at all time periods. Nothing else in the models considered thus far suggests a dynamic process involving time-dependence in either the core of the model or in the idiosyncratic error terms. The longitudinal nature of the data, in other words, has been used primarily as a means to rid the model of the effects of the nuisance U_i term, and not to model any kind of dynamic processes *per se*. Controlling for unobserved stable unit effects is highly important for panel analysis, but it is often insufficient to take into account all of the temporal dependence in the data. Thus much of panel analysis is devoted to alternative means of modeling temporal dependence, either instead of, or in addition to, the heterogeneity models we have considered thus far.

3.1 Autocorrelated disturbances

One kind of additional temporal dependence is caused by correlations between the idiosyncratic error terms ε_i of successive panel waves. These *autocorrelated* disturbances could be the result of exogenous random shocks to the system that persist for several time periods, omitted variables that change over time, or correlated errors of measurement from one panel wave to the next. Such a model could be represented as in (2) above with the additional stipulation that the error terms are autocorrelated, as in:

$$Y_{it} = \alpha + \beta_1{}^*X_{1it} + \beta_2{}^*X_{2it} + \cdots \beta_j{}^*X_{jit} + U_i + \varepsilon_{it} \quad (9a)$$

$$\varepsilon_{it} = \rho\varepsilon_{it-1} + \nu_{it}, \text{ where } \nu_{it} \sim N(0, \sigma_v^2) \quad (9b)$$

This model is estimated in two stages. The first proceeds by estimating ρ through one of many available methods commonly used in time-series analyses (see Baltagi, 2005), and then weighting equation (9a) by the ρ estimate to produce a model with the well-behaved error term ν_{it}. In the second stage, either a fixed effects or random effects specification for the U_i is assumed, with estimation on the transformed equation (a) proceeding as in the models discussed earlier, either with mean-differencing (fixed effects) or estimation of the variance components of the error term and "θ-mean differencing" (random effects). One wave of observations is lost with the first-stage differencing procedure, so this model requires at least three waves of data.

3.2 Lagged endogenous variable models

An alternative model of temporal dependence in Y_i is perhaps more prevalent in the panel literature. In this model the Y_{it} response is determined by a series of X variables, either at time t and/or at a lag of $t-1$, along with the *lagged value of Y_i*, as in:

$$Y_{it} = \alpha + \beta_1{}^*Y_{it-1} + \beta_2{}^*X_{1it-1} + \beta_3{}^*X_{2it-1} + \cdots \beta_j{}^*X_{Jit-1} + \varepsilon_{it} \quad (10)$$

In this model, the lag value of Y (the "lagged endogenous variable") has a direct effect on the value of Y at the next time point, along with effects specified from prior values of X as well. We may include contemporaneous levels of X in the model as well, but for now we focus on the lagged effects for both the X_J and Y_{t-1}. The model captures the temporal dependence of adjacent responses on Y neither through their joint relationship with some stable unobserved U term, nor through the autocorrelation of adjacent unknown idiosyncratic errors ε_{it}, but rather because of the *direct influence* of Y at a given point in time on subsequent responses. With some simple algebra it can be shown that the model is equivalent to predicting the *change* in Y from its prior value, with the coefficient on lagged Y in the change-score version of the model being equal to $(\beta_1 - 1)$:

$$Y_{it} - Y_{it-1} = \alpha + (\beta_1 - 1)^* Y_{it-1} + \beta_2{}^* X_{1it-1} + \beta_3{}^* X_{2it-1} + \cdots \beta_j{}^* X_{Jit-1} + \varepsilon_{it} \quad (11)$$

This means that, so long as the lagged dependent (endogenous) variable is included on the right-hand side, there is no difference between the expression of the model in terms of "static" scores in Y or as a "dynamic" change-score model in ΔY. Further, the effects of the X_j variables are exactly the same, regardless of the specification of (10) or (11).

The model of equation (10) or (11) differs from those we have considered thus far in several important ways. One is the absence of the unit-specific error term U_i, meaning that it is assumed that *all* of the temporal dependence of responses over time is due to the causal mechanism linking the lagged endogenous variable and the lagged (or contemporaneous) X_j to Y at any given point in time. This assumption can be relaxed, however, and we will examine models that include both lagged Y and the unit-specific effects below. But the most fundamental difference is the presence of lagged Y as a predictor in equations (10) and (11) in the first place, and the inclusion of this term that has generated a good deal of controversy in the panel literature.

All agree that whenever there are strong substantive reasons for assuming that prior values of a variable have a direct *causal* effect on its subsequent value, the dynamic model is entirely appropriate. For example, in models of attitude formation and change, it is often assumed that there is some natural "state-dependence," such that attitudes are determined directly by their prior values unless disturbed by some exogenous shock. Economic models of wealth may also assume that an individual's store of accumulated wealth *causes* subsequent levels through different investment decisions, employment, and educational opportunities, and the like. These processes would stand in contrast to variables that need to be "created anew" in each time interval, as, for example, behavior such as voting or political participation where engaging in the activity at one point in time may not necessarily "cause" subsequent behaviour. In those cases, there may be little theoretical justification for estimating the dynamic model.

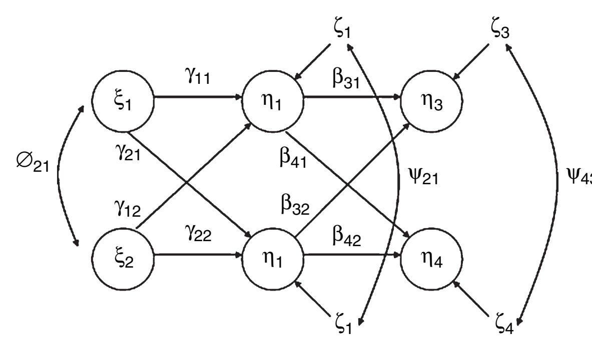
More controversial are the purely *statistical* reasons that have been advanced for including lagged Y in panel models. One is to serve as a proxy for unmeasured factors that lead to the response at both points in time, and many analysts argue that the heterogeneity models or autocorrelation models we have considered thus far deal more explicitly and more effectively with this problem (Allison, 1990; Liker, et al., 1985). Likewise, it has often been argued that including the lagged endogenous variable served as a statistical control for "regression to the mean" effects, whereby change in a variable is typically *negatively* related to its subsequent value (as seen in the $(\beta_1 - 1)$ coefficient for Y_1 in equation 11). This occurs because high (low) initial values were likely to be the product of some random forces that are not likely to be repeated in subsequent observations. Others claim, however, that once random measurement errors are taken into account (with models that we consider below), regression to the mean is usually not sufficient to justify the inclusion of Y_{t-1} in the model. These arguments are far from settled in the literature, but they do point to the need for researchers to consider carefully the "epistemological status" of the lagged endogenous variable (Arminger, 1987), and not enter it automatically in panel models.

4 Structural equation panel models

The dynamic specification of (10) is a common starting point for panel analysis in the *structural equation modeling* (SEM) tradition. Instead of pooling the data across waves and estimating a single coefficient for each of the independent variables over the N^*T units of observation, structural equation methods specify and estimate a system of equations, one for each dependent (endogenous) variable at each wave of observation. The overall model represents the interrelationships across all of the waves between the *exogenous* variables, i.e., those determined outside of the causal system and taken as "given," and the *endogenous* variables, which are determined by other exogenous or endogenous variables in the model. In principle, each of the equations—i.e., the equation for Y in wave 2, Y in wave 3, etc.—could be estimated separately, but the SEM approach uses the information provided by the variances and covariances between *all* the observed variables in the model, even those that are not directly related to one another in any of the model's equations, to allow the more efficient simultaneous estimation of all the model's parameters. In panel models, this cross-wave covariation, along with the intrinsic temporal ordering between variables at earlier and later waves, provides the researcher great flexibility in estimating complex models with reciprocal causal linkages between variables, and, under some conditions, models that allow random measurement error in the observed indicators over time. These are the principal ways that the structural equation approach is used by panel analysts, with the dynamic specification of (10) most often at the core of these models.

The SEM approach may be illustrated with an example of the relationship between an individual's partisan identification (i.e., the direction and strength of his or her attachment to the Republican or Democratic Party in the US), and presidential approval (i.e., whether he or she approves of the performance of the sitting President in office). In the political science literature, controversy rages over whether party identification is a stable characteristic which influences shorter-term perceptions such as candidate evaluations, presidential approval, or whether such short-term factors instead determine partisan change over time (e.g., Green and Palmquist, 1990). A structural equation model of these alternative processes is depicted in Figure 29.1, with the variables of partisan identification and presidential approval measured in three waves of observation in the National Election Panel Study of 2000–2002–2004. In order to facilitate the presentation through the chapter, the variables and coefficients are labeled with the LISREL nomenclature which is widely utilized within the SEM tradition (Jöreskog and Sörbom, 1994; Kaplan, 2000).

The model shows that party identification and presidential approval measured in 2000 are assumed to be exogenous variables, i.e., variables with causes outside the causal system, and are depicted as ξ_1 and ξ_2. Party identification and presidential approval in 2002 and 2004 are endogenous (η_1 through η_4), predicted by their own previous value and the previous value of the other variable.[15] That is,



ξ = Exogenous variable
η = Endogenous variable
β, γ = Regression coefficients
$\varnothing$ = Variance-covariances of the ξ exogenous variables
ψ = Variance-covariances of the ζ structural disturbances
ζ = Structural disturbance of the η

Figure 29.1 Three-wave, cross-lagged panel model

[15] All variables are coded in a "pro-Republican" direction so that high values on approval indicate either disapproval of a Democratic President (in 2000) or approval of a Republican President (in 2002 and 2004).

the model specifies a lagged effect from each variable on itself over time, and *cross-lagged* effects of party identification and presidential approval on each other. Each endogenous variable also has a random error term, depicted as ζ_1 through and ζ_4. The structural effect linking the exogenous variables ξ to endogenous variable η are labeled as γ coefficients, and the structural effects linking the endogenous variables to one another are labeled as β. Following common SEM practice, all variables are expressed as deviations from their mean, which eliminates consideration of the intercept in all of the structural equations.[16]

The four equations for the endogenous variables may therefore be written as:

$$\eta_1 = \gamma_{11}\xi_1 + \gamma_{12}\xi_2 + \zeta_1 \tag{12a}$$

$$\eta_2 = \gamma_{21}\xi_1 + \gamma_{22}\xi_2 + \zeta_2 \tag{12b}$$

$$\eta_3 = \beta_{31}\eta_1 + \beta_{32}\eta_2 + \zeta_3 \tag{12c}$$

$$\eta_4 = \beta_{41}\eta_1 + \beta_{42}\eta_2 + \zeta_4 \tag{12d}$$

or in matrix form as

$$\boldsymbol{\eta} = \mathbf{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{13}$$

where

$\boldsymbol{\eta}$ = a vector of m endogenous variables (in this case 4)

$\mathbf{B}$ = an m by m matrix of β coefficients linking the m endogenous variables (here 4×4)

$\boldsymbol{\Gamma}$ = an m by n matrix of γ coefficients linking the n exogenous to the m endogenous (4×2)

$\boldsymbol{\zeta}$ = a vector of m structural disturbances for the endogenous variables (here 4)

Two other matrices are relevant from the point of view of specification and estimation:

ϕ, the n by n matrix of the variances and covariances of the exogenous variables (here comprised of three distinct elements, the variances of ξ_1 and ξ_2 and their covariance, represented by the curved arrow labeled ϕ_{21}); and ψ, the m by m matrix of the variances and covariance of the structural disturbances ζ (here 4×4, with the diagonal elements representing each equation's error variance and additional covariances estimated between the structural disturbances for n_1 and n_3 at wave 2, and for n_2 and n_4 at wave 3). These error covariances represent the residual covariance between party identification and presidential approval at a given panel wave that cannot be explained through the stability and cross-lagged effects in the model.

Several features of this model are important to note. First, it can be seen that each of equations (12a) through (12d) is simply a version of the dynamic panel model of (10) with the lagged values of Y and X as independent variables. Thus the estimate for γ_{11} in (12a) and β_{31} in (12c) represent the "stability" or "autoregressive" effect of party ID in wave 1 or 2 on its own value in wave 2 or 3; alternatively, subtracting 1 from these estimates will result in the effect of party ID at wave 1 or 2 on the subsequent *change* in party ID over the next panel wave. Similarly, the estimates γ_{12} and β_{32} represent the lagged effect of presidential approval in wave 1 or 2 on subsequent values of party identification, or, equivalently, on the change in party identification over time. The corresponding coefficients for presidential approval and party identification in equations (12b) and (12d) have exactly the same interpretation in terms of the respective variables on presidential approval or the

[16]LISREL and other SEM analysis packages do allow for the estimation of intercepts, and these kinds of models are widely used in the analysis of group differences (Sörbom, 1982) and in multilevel structural equation models (Muthén, 1994).

change in presidential approval over time. Thus the SEM setup here is designed to provide information on the relative stability of the two variables, as well as on the relative magnitude of the two cross-lagged effects across the waves 1–2 and waves 2–3 periods.[17]

Second, the model as specified contains no contemporaneous reciprocal linkages between party identification and presidential approval at wave 2 or wave 3, i.e., there are assumed to be no effects between η_1 and η_2 or between η_3 and η_4. This means that the cross-lagged model in Figure 29.1 is *recursive*, in contrast to a *nonrecursive* model that would contain feedback effects between variables observed at the same panel wave. The assumption of no contemporaneous effects may not be justified, as in many panel analyses the length of time for X to affect Y may be significantly shorter than the time lag between waves of observations. Nevertheless, in the absence of strong theory the cross-lagged model is usually a satisfactory initial model.[18] Nonrecursive models present more difficulties than recursive models in identifying and estimating causal parameters, as will be seen below.

Structural equation panel models are estimated in the same way as all SEMs (see Bollen, 1989; Kaplan, 2000). The variances and covariances between observed variables are expressed in terms of the unknown parameters γ, β, ψ, and ϕ, given the model's assumptions. In this case we assume that all variables are expressed in mean deviation form, that the ζ disturbances are unrelated to both the η and ξ that appear as independent variables in their respective equations, and that there are no covariances between any of the ζ disturbances.

Second, it is determined whether the model as a whole, and individual equations within it, are *identified*, i.e., whether there is sufficient information in terms of the observed variances and covariances to produce unique estimates of each of the model's parameters. All recursive models are either identified or overidentified (with more known variances and covariances compared than unknown parameters), though this will not be the case in the nonrecursive simultaneous effects models we will consider shortly. In the model of Figure 29.1, there are 17 unknowns—3 φ variances and covariances of the exogenous variables, 4 γ linking exogenous to endogenous variables, 4 β linking endogenous variables, and 6 ψ variances and covariances of the endogenous variables—and 21 known variances and covariances. This means that the model is *overidentified* with 4 degrees of freedom. When models are overidentified, there will be more than one solution for at least some of the model's unknowns, and this additional information can be used to assess how well the model fits the data as a whole.

Third, under the assumption of multivariate normality of the observed variables, maximum likelihood methods are typically used to estimate the model parameters. Intuitively, the ML procedures find the estimates of the unknown parameters, which, taken together, minimize the difference between the implied and actual variance–covariance matrices (see, e.g., Kaplan, 2000, pp. 24–27 for more details).

[17] It is important to note that the "stability" represented by the lagged dependent variable is not absolute stability in the sense of "no change," but rather stability in the sense of relative rankings of cases over time. When the autoregressive coefficients are closer to 1, this indicates that units with higher values at time (t) tend also to have higher values at time (t+1), though significant absolute changes may have occurred, either due to the effects of the other variables in the model or an overall change that affects all units equally.

[18] It is also the case that the cross-lagged model can be derived from the "continuous time" panel model where both X and Y exert continual influence on one another over time, as opposed to having effects in discrete time intervals corresponding to the waves of measurement (Coleman, 1968; Tuma and Hannan, 1994).

Finally, the variance–covariance matrix that is implied from the estimated coefficients is compared to the observed variance–covariance matrix to make assessments of the fit of the model as a whole. If a model can reproduce the observed variances and covariance very well, it is "consistent" with the data. If a model cannot, it is "inconsistent" with the data.

A variety of measures are available for assessing the significance of particular models in terms of their overall explanatory power and in terms of making comparisons between alternative models with different unknown parameters. For example, the quantity $n^* \log L_0$, where L_0 is the likelihood function of the estimated model, is distributed as $\chi^{2,}$ with degrees of freedom equal to the number of estimated parameters. This "model χ^2", widely used as a measure of the overall fit of the model to the data with low values, relative to degrees of freedom, indicated better fit. Given the sensitivity of χ^2 to sample size, a variety of additional measures have been proposed to assess the explanatory power of a given model versus alternative baseline models; e.g., the Normed Fit Index (NFI) compares the model χ^2 to the χ^2 of a "complete independence" model with zero covariances among all variables in the population, while the Parsimony Normed Fit Index (PNFI) penalizes the NFI to the extent that the estimated model includes more and more parameters. See Hu and Bentler (1995) for a comprehensive overview of these issues. Finally, when models are "nested," such that one model can be defined by relaxing constraints on parameters in another model, the difference in χ^2 between the two models provides a test of the significance of the improvement in fit between the unconstrained versus constrained models.

The ML estimates of the cross-lagged model of Figure 29.1 for the NES panel data (N=738) are shown in Table 29.2. The model shows that there are significant cross-lagged effects in both directions between party identification and presidential approval, with the magnitude of the party-approval effect being approximately three to four times larger as the reverse. The model shows strong stabilities for the party and variable and weak stability for the approval of the President, especially between 2000 and 2002, which may be expected due to the change in Presidential administration (despite the coding changes described in footnote 12). Nevertheless, the model shows some support for the "revisionist" notion that short-term political evaluations influence the intensity of an individual's identification with particular political parties in the US. The estimates, however, are in the context of a poorly-fitting model, with a large and significant χ^2 and a relatively low Parsimony Normed Fit Index of .26.

Column (2) shows the flexibility of the SEM approach in terms of constraining particular parameters to be equal in order to test the relative explanatory power of alternative model specifications. In this model, the cross-lagged effects between party and approval from waves 1–2 and 2–3 are each constrained to be equal, thus gaining 2 degrees of freedom in the process. The results show nearly identical parameter estimates to model (1), and the difference in the two models' χ^2 is only .27, which, with 2 degrees of freedom difference, indicates that the unconstrained model does not fit the data significantly better than the constrained model. The PNFI shows a corresponding improvement to .39, reflecting the nearly equal explanatory power of this model with fewer estimated parameters. In practice, panel analysts will estimate a variety of alternative models, usually imposing equality constraints at the outset and relaxing them as necessary as indicated by χ^2 tests and other information about model fit. In this case, the overall indices show that neither of the models estimated provides a particularly good fit to the data, meaning that additional parameters, perhaps in the form of synchronous casual effects, may need to be included.

Table 29.2 Cross-lagged reciprocal effects models, party identification and presidential approval, American National Election Study, 2000–2002–2004

Stability effects	*(1)*	*(2)*	*(3)*
	No equality constraints	*Equality constraints*	*Measurement error in party identification*
	.80	.81	.97
Party identification, wave 1–2 γ_{11}	.79	.80	.96
	.88	.88	1.04
Party identification, wave 2–3 β_{31}	.83	.82	.96
	.15	.14	.09
Presidential approval, wave 1–2 γ_{22}	.16	.15	.10
	.56	.57	.52
Presidential approval, wave 2–3 β_{42}	.49	.49	.44
Cross-lagged effects			
	.31	.31[a]	.37[a]
Party to approval, wave 1–2 γ_{21}	.43	.43	.49
	.31	.31[a]	.37[a]
Party to approval, wave 2–3 β_{41}	.37	.37	.43
	.17	.16[b]	.03[b*]
Approval to party, wave 1–2 γ_{12}	.13	.12	.02
	.15	.16[b]	.03[b*]
Approval to party, wave 2–3 β_{32}	.10	.11	.02
Error covariances			
Wave 1 φ_{21}	2.25	2.25	2.26
	.62	.62	.66
Wave 2 ψ_{21}	.45	.45	.29
	.13	.13	.09
Wave 3 ψ_{43}	.44	.44	.25
	.10	.10	.06
Measurement error variance			.50
Party identification ε_1	----	----	
R-squared, party identification, wave 2	.77	.77	.95
R-squared, party identification, wave 3	.79	.79	.95
R-squared, presidential approval, wave 2	.29	.28	.31
R-squared, presidential approval, wave 3	.60	.60	.62
χ^2 (degrees of freedom)	133.33(4)	133.60(6)	10.91(5)
Normed fit index (NFI)	.97	.97	1.00
Parsimony normed fit index	.26	.39	.33

All variables statistically significant; standardized coefficients italicized.
Coefficients a, b constrained in models (2) and (3) to be equal. N = 738.

4.1 Alternative lag effects

Given enough waves of observation and enough knowns in the form of observed variances and covariances, the SEM approach allows considerable flexibility in specifying alternatives to the cross-lagged causal models of Figure 29.1. One possibility is a model with synchronous, or simultaneous impact of variables on one another at a given wave of observations. Such a model would be appropriate if the time lag for the causal influence of the independent variable is thought to be short, relative to the time period between observations. For example, in models of the impact of interpersonal networks, if the panel waves were separated by five years, a cross-lagged model may not capture network impact well, as its effect may either have dissipated in the intervening time period or the network itself may have changed since the initial panel wave, or both. In that case, a more accurate depiction of the causal process may be to include β effects between η_1 and η_2, and between η_3 and η_4 instead of the cross-lagged effects linking the variables from wave 1 to 2, and wave 2 to 3. It is also possible that both short- and longer-term causal lags could be present, in which case the additional β effects would be included along with the cross-lagged processes already specified.

In either case, the inclusion of synchronous causal effects yields a *nonrecursive* causal model whose estimation is significantly more complicated than in the recursive, cross-lagged only case. This is so for two main reasons. First, in synchronous effects models the assumption of no correlation between independent variables and the error term in their respective equations is untenable, meaning that the methods we have considered thus far would produce inconsistent estimates of the β. Imagine hypothetical causal arrows between η_1 and η_2 in Figure 29.1, with β_{21} representing the effect of party ID on approval in 2002, and β_{12} representing the reciprocal link between approval and party ID. It can immediately be seen that η_1 is related to ζ_2, since ζ_2 causes η_2 which causes η_1. Similar processes result in a nonzero correlation between η_2 and ζ_1, meaning that we cannot consistently estimate either β_{21} or β_{12}—or any of the other effects in their respective equations—with the information at hand.

Second, the inclusion of reciprocal synchronous effects raises the possibility that the model as a whole, or some of the individual equations, are not identified. For example, in a bivariate cross-section model with reciprocal causal effects, the model would not be identified because there would be four unknowns—the two β and the two structural disturbances—and only three known variances and covariances with which to estimate the unknown parameters.[19]

What is needed in the case of nonrecursive models in general, and synchronous effects panel models in particular, is more information in the form of additional observed variances and covariances. More information, of course, generates additional "knowns" so that the counting rule is more likely to be satisfied. But the additional variables must be related to the included variables in specific ways in order to be of use in solving the estimation problems posed in nonrecursive models. Specifically, what are needed are variables that can help satisfy the so-called *order condition* for identification, which states that if an equation involves m endogenous variables, then there must be at least $(m-1)$ *excluded exogenous variables* for the equation to be identified. That is, the equation must have at least one ξ that does *not* have an effect on the η endogenous dependent variable in question for every endogenous independent variable that does. For example,

[19] In that case, the model as a whole would not satisfy what is referred to as the *counting rule* for identification, whereby $(m+n)(m+n+1)/2 >= k$, with m and n representing the number of exogenous and endogenous variables, and k representing the number of free parameters.

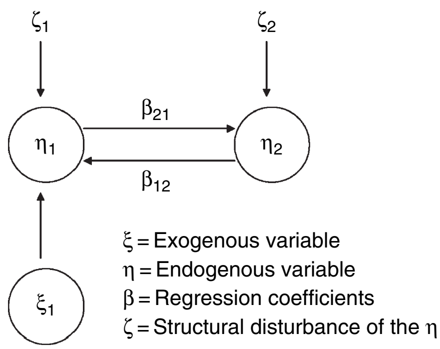


Figure 29.2 An instrument variable ($\S_1$) in a nonrecursive causal model

in Figure 29.2 we add a variable ξ_1 to the equation predicting η_1 in a reciprocally-related bivariate cross-sectional model. Following the order condition, ξ_1 identifies the η_2 equation, as there is now one excluded exogenous variable from the equation and one included endogenous variable (η_1). Given that no exogenous variable is excluded from the η_1 equation, it is still not identified. Thus identification depends on the availability of additional exogenous variables that affect one and only one of the two reciprocally-related variables in a synchronous effects model.

The excluded exogenous variable, also called an *instrumental variable*, is then used to estimate the model parameters consistently. In the simplest case, such as depicted in Figure 29.2, the assumptions regarding $\xi_{1,}$ namely, that it is uncorrelated with the ζ and it has no direct causal effect on $\eta_{2,}$ allow the expression of β_{21} as:

$$\text{Covariance}(\xi_1, \eta_2)/\text{Covariance}(\xi_1, \eta_1)$$

with the covariance between η_1 and the structural disturbance ζ_2 no longer contaminating the estimate of the β_{21} causal effect. In more complex cases, there may be more than one instrument available for inclusion in the model, in which case estimation proceeds through "two-stage least squares" (TSLS) methods or through the general maximum likelihood procedures discussed above. In the widely-utilized TSLS setup, the endogenous independent variable is first regressed on *all* exogenous variables, including all of the instruments, generating a "predicted η" which is uncorrelated with all of the model ζ. Then, in the second stage, the dependent variable is regressed on the exogenous independent variables and the "predicted η" from the first stage (with appropriate corrections to the standard errors in the second stage).

All of these procedures, however, depend on the satisfaction of the assumptions of instrumental variable analyses. That is, there must be variables included in the model that affect *one and only one* of the two reciprocally-related endogenous variables, and these variables must be exogenous in the sense of unrelated to the structural disturbances of the endogenous variables. These variables are difficult to find in many practical research situations. In panel designs, however, it may be reasonable under some conditions to assume that the lagged values of variables are exogenous and related to the endogenous variables in such ways as to facilitate identification and estimation. For example, consider a "pure" two-wave synchronous effects version of Figure 29.1 with no cross-lagged effect from either ξ_1 to η_2 or from ξ_2 to η_1. In that case ξ_1 would be used to identify the η_1 equation. Thus, if the assumptions of exogeneity can be justified, the longitudinal structure of panel data can provide additional information in the form of instrumental variables for purposes of identifying and estimating nonrecursive, reciprocal effects models.

In many instances, however, the situation will be complicated by the violation of the exogeneity assumptions. If there are autocorrelated disturbances present in the model, for example, then the lagged value of a variable will *not* be unrelated to the structural disturbance of the endogenous variables and hence will not be a suitable instrument. Further, in models with *both* cross-lagged and synchronous effects, the lagged value of a variable, even if unrelated to the structural disturbances, is assumed to have effects on both endogenous variables at subsequent waves of observation, and so cannot be used to identify the subsequent wave's equations. One possible solution would be to include additional exogenous variables as instruments, provided that they satisfy the exclusion restrictions discussed above. Another common possibility is to identify the model through the use of equality constraints, such that the cross-lagged effects may be assumed to be equal from waves 1 to 2 and 2 to 3, the synchronous effects may be assumed to be equal in waves 2 and 3, and perhaps the stability effects equal across waves as well. Given that some of these models will be nested within each other, comparison of chi-square goodness of fit measures can provide some insight into the models that are most consistent with the observed data; with that information, along with the parameter estimates found for the various models, the analyst may arrive at conclusions regarding the likely pattern of lag causal effects between variables over the course of the panel observations.

In the current example, we re-estimate model (2) in Table 29.2 by including both synchronous and cross-lagged effects between party and approval, and specifying equality constraints between the two sets of contemporaneous effects. The results show that the cross-lagged effects are still significant in both directions, while neither contemporaneous effect is significant, with estimated values of .001 from party to approval and −.05 from approval to party. The model χ^2 is 133.3 with 4 degrees of freedom, thus not a significant improvement from the cross-lagged only model in Table 29.2, given the loss of 2 degrees of freedom. We conclude that the cross-lagged model is superior to a model with both cross-lagged and synchronous effects, though neither model fits the data well.

4.2 Measurement error

The SEM approach to panel analysis is also often extended to include the estimation of causal effects, controlling for errors of measurement in the variables of interest. It is well-known that random measurement in independent variables causes estimation of regression coefficients to be biased, downward in the case of bivariate equations and in unknown directions in multivariate models (Bollen, 1989). In cross-sectional analyses, there is often not enough information to identify and estimate both the structural effects *and* the measurement error that may be present in the model, as there needs to be multiple observed indicators of the presumed "latent" (error-free) variable of interest in order to proceed. With panel data, though, the information that is provided from the same variables over time allows much more flexibility in estimating structural effects once measurement error is taken into account, and in estimating and assessing the measurement properties of particular indicators as well. In the LISREL and other applications of the SEM framework, all of these effects are estimated simultaneously, providing powerful additional tools to the panel analyst in strengthening the causal inference process.

Figure 29.3 shows a three-wave autoregressive panel model that includes a random measurement component. The model depicts each wave's indicator of y as a function of an unobserved latent "true" score η plus a random measurement component ε. In equation form, the model is written in two parts as:

$$y_{it} = \lambda_{kt}\, \eta_{jt} + \varepsilon_{it} \quad \text{“Measurement Model”} \quad (14a)$$

$$\eta_{it} = \beta_{t,t-1}\, \eta_{jt-1} + \zeta_{it} \quad \text{“Structural Model”} \quad (14b)$$

with the ε assumed to be normally distributed random variables, uncorrelated with the η and the structural disturbances ζ, and, in basic models, uncorrelated with each other over time. The

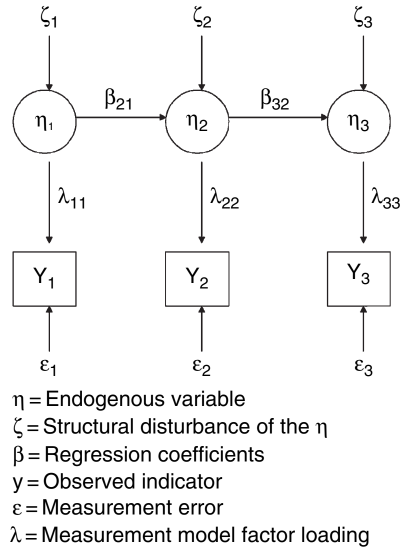


Figure 29.3 Three-wave autoregressive model with measurement error

model differs from (13) because of the presence of random measurement equation in the observed variables, which was assumed (unrealistically) to be zero in all of the models considered to this point. Expressing, for example, the structural portion of the model from wave 1 to wave 2 in terms of the fallible indicators shows the consequences of this omission, setting $\lambda = 1$ for simplicity:

$$y_{2i} = \beta_{21}\, y_{1i} - \beta_{21}\varepsilon_{1i} + \zeta_{2i} + \varepsilon_{2i} \tag{15}$$

As can be seen, the regression of y_2 on y_1, with both variables containing random error, will have a larger error variance than the true structural disturbance ζ_2, with a corresponding lower R-squared and inefficient estimates of the model coefficient's standard errors; even more consequential is that estimates of β_{21}, the autoregressive or "stability" effect in the model, will be inconsistent, as y_1 is intrinsically related to the error term in equation (15) due to the presence of ε_1.

This is an important result, showing that measurement error in the basic dynamic panel model yields an incorrect estimate of the stability parameter β, often lower than its true value. And to the extent that other independent variables are (positively) related to both y_1 and y_2, their effect in equation (15) is likely to be biased *upwards*, leading to potentially erroneous conclusions about their effects on y_2 or Δy. Thus, it is essential to take measurement error into account in panel models, and failure to do so is likely to negate many of the advantages of panel analysis in estimating dynamic causal processes.

Given the statistical nature of the problem, i.e., the correlation between the independent variable y_1 and the error term in equation (15), one solution would be to use instrumental variable analysis. If an exogenous variable ξ exists, such that it was uncorrelated with both the measurement error in y and the structural disturbance ζ, then estimation could proceed via the TSLS or the ML procedures considered thus far. In the first stage, y_1 would be regressed on all exogenous variables, including the instrumental variable(s), and in the second stage, y_2 would be regressed on all independent variables in its equation, along with the predicted value of y_1 from the first stage, which would be purged of the correlation with its error term. There are two drawbacks to this approach. One is practical, in that it is commonly difficult to find exogenous variables that satisfy the exclusion restriction discussed above, i.e., that are related to a fallible indicator in one wave of observation but not the next.

Second, the IV solution, even if successfully implemented, does not provide information about the measurement properties of indicators in the model, which may often be of considerable interest. In terms of the measurement model of (14a), we may wish to know the "reliability" of y, defined as the quantity

$$\text{Variance}(\eta)/\text{Variance}(y)$$

or the proportion of the observed variance that is comprised of "true score" variance. This information may be especially useful in multiple indicator models, i.e., those where more than one indicator for a given latent construct is available. In that case, assessing the reliabilities of the specific indicators, both absolutely and in relation to one another, provides additional information that can be used to determine the adequacy of the indicators or their suitability as measures of the underlying construct.

As the amount of information available increases, in terms of waves of observation and number of indicators of latent variables, panel models with measurement error may be identified and estimated with fewer and fewer restrictive assumptions. In the model of equations (14a) and (14b), for example, we have six observed variances and covariances, and 11 unknowns—the three λ, the three variances of the measurement error ε, two β, and three variances of the structural disturbances ζ. We may make some progress in setting all of the λ to equal 1; this has no substantive bearing on the model and simply puts the latent variable on the same measurement scale as the observed y indicators.[20] This leaves 8 unknowns and 6 knowns in the model. Wiley and Wiley (1970) propose that identification be achieved in this case by constraining the variance of the measurement error term ε to be equal across the three waves, gaining two degrees of freedom in the process as well. Under these assumptions, the model is just-identified and the relevant parameters can be solved through algebraic manipulation of the observed indicators' variances and covariances, as:

$$\mathrm{var}(\varepsilon) = \mathrm{var}(y_2) - [\mathrm{Covariance}(y_3, y_2)^{*}\mathrm{Covariance}(y_1, y_2)/\mathrm{Covariance}(y_1, y_3)] \tag{16a}$$

$$\beta_{21} = \mathrm{Covariance}(y1, y2)/[\mathrm{Variance}(y1) - \mathrm{Variance}(\varepsilon)] \tag{16b}$$

$$\beta_{32} = \mathrm{Covariance}(y1, y3)/\mathrm{Covariance}(y1, y2) \tag{16c}$$

Once these manipulations are accomplished, it is straightforward to solve the reliabilities of the indicators, given the observed variance $(y_t) = \mathrm{var}(\eta_t) + \mathrm{var}(\varepsilon)$.[21]

More generally, the maximum likelihood estimation of measurement error and structural coefficients may proceed simultaneously within the LISREL or other applications of the SEM approach. The measurement equations in (13a) would be summarized in matrix form as:

$$\mathbf{y} = \mathbf{\Lambda}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{17a}$$

and, if measurement error were assumed to be present in the exogenous ξ variables, then

$$\mathbf{x} = \mathbf{\Lambda}_{\mathbf{x}}\boldsymbol{\xi} + \boldsymbol{\delta} \tag{17b}$$

with the $\mathbf{\Lambda}$ matrices being $(q \times m)$ and $(p \times n)$ matrices of the λ linking the q indicators (y) of the m endogenous variables and p indicators (x) of the n exogenous variables, and $\boldsymbol{\varepsilon}$ and $\boldsymbol{\delta}$ being $(q \times 1)$ and $(p \times 1)$ vectors of the measurement errors in y and x, respectively. The implied variance-covariance matrix that expresses the knowns in terms of the unknown model parameters can then be expanded to include the

[20]In multiple indicator models, one indicator's λ is set to 1 for the same reason, while the other λ are free to vary.

[21]The single indicator, three-wave model may also be identified through the Heise (1969) procedure. In this model, the latent and observed variables are standardized, so that the unknowns are the two β stabilities and the three λ coefficients, which represent path coefficients linking the latent variables and observed indicators (and thus λ^2 represents the reliability coefficient). Under the assumption of equal reliabilities across waves, the model is just-identified.

unknowns in $\Lambda_y\Lambda_x$, and the variances of the ε and δ, denoted as $\boldsymbol{\theta}_y$ and $\boldsymbol{\theta}_x$, respectively (see Kaplan, 2000, p. 56). Maximum likelihood estimation then proceeds as before, assuming multivariate normality for all the observed variables (and assuming that appropriate constraints are imposed when necessary to achieve parameter identification).

There are several options for incorporating the information regarding measurement error into full-blown cross-lagged or synchronous structural equation models. In three-wave single-indicator models, one method is to fix the measurement error variances of the indicators at the values that are obtained through the Wiley-Wiley procedure, with the structural effects then estimated while correcting for the unreliability of the indicators. Another method is to allow the measurement error variances to be completely free parameters, with LISREL or alternative programs producing simultaneous estimates of measurement and structural coefficients in the model.

This procedure is illustrated in model (3) of Table 29.2, which shows a reanalysis of the cross-lagged panel model of Figure 29.1 while allowing for measurement error in the party identification equation. As can be seen, the stability of the party variable rises considerably compared to the previous estimation, and neither cross-lagged effect from approval to party identification now is statistically significant or substantively meaningful, with standardized values of .04 or less. By contrast, the cross-lagged effect from party to approval is somewhat larger than in the no-measurement-error model, and the overall model χ^2 is much improved (10.91 with 5 degrees of freedom, compared with 133.6 with 6 degrees of freedom in the previous model). Thus interpretation of the causal effects in panel models may be substantially altered once measurement error is taken in to account; in this case the measurement error model shows results that are much more in accord with traditional views of party identification as the "unmoved mover" of short-term political evaluations (Green and Palmquist, 1990). And in this case, the reliabilities of the party identification measures are calculated as approximately 90%, meaning that only 10% of the indicator variance was estimated to be "error;" when indicators exhibit less reliability, then differences between measurement error models and models assuming perfect measurement will be even greater.[22]

It should be emphasized, however, that the measurement models estimated thus far depend on their own set of assumptions that need to be justified. For the Wiley-Wiley procedure, for example, it must be assumed that the error variances are equal over time for the three-wave panel model to be identified. This may be unrealistic, and work conducted with longer-term panels suggests that error variances tend in fact to shrink over time (Jagodzinski, et al., 1987). In the present case, moreover, the Wiley-Wiley method failed to produce credible estimates in the case of presidential approval, indicating that the assumptions in the model were unlikely to be satisfied.

In such instances, the analyst may simulate results by plugging in different values of error variances, assuming perhaps some degree of constant shrinkage over time. But the more promising solution, as always in measurement error models, is whenever possible to add waves of observation and/or additional indicators for the latent constructs. With four waves of data, the assumption of equal error variances can be relaxed such that the measurement errors of the "inner" indicators y2 and y3 are identified without constraint; moreover, the stability effect from wave 2 to wave 3 (β_{32}) is overidentified, thus indicating that the fit of the model as

[22]For example, the observed variance of y_2, party identification in 2002, is 4.84. Given the estimated error variance of .5, this yields a reliability estimate of $(4.84-.50)/4.84=.90$.

a whole can be assessed. With longer-wave panels, more parameters will be identified without the restrictive constraints of the Wiley-Wiley procedure. And when multiple indicators are available, models that incorporate autocorrelation between the errors of measurement over time may also be estimated, with fewer constraints necessary as the waves of observation and number of indicators increase.

5 Dynamic panel models with unobserved heterogeneity

A natural extension of the models that we have been considering thus far is to incorporate both dynamic causal processes (and potentially measurement error) along with unobserved heterogeneity into a single model. Both the econometric and the SEM panel traditions provide the ability to incorporate and estimate these kinds of models.

The dynamic model with heterogeneity takes the following general form:

$$Y_{it} = \alpha + \beta_1{}^* Y_{it-1} + \beta_2{}^* X_{1it} + \beta_3{}^* X_{2it} + \cdots \beta_j{}^* X_{jit} + U_i + \varepsilon_{it} \tag{18}$$

with β_1 representing the effect of the lagged endogenous variable Y_{t-1} and U_i, as before, representing the unit-specific effect. It can be seen that the model combines equation (2), the basic model for unobserved heterogeneity, with equation (10), the basic dynamic model with contemporaneous effects of the Xs, though lagged values of X could also have been included in the model.[23] The combined model thus corresponds to a situation where, for reasons discussed earlier, the lagged endogenous variable is thought to exert direct causal influence on its subsequent values, *and* there are stable unobserved differences between the units that push individual cases higher or lower on the dependent variable, aside from the included Xs and aside from the dynamic processes represented by the effect of Y_{t-1}.

The inclusion of both kinds of effects in the same panel model results in some estimation difficulties, however. The problem stems from the fact that Y_{t-1}, the lagged endogenous variable, is intrinsically related to the composite error term of (18) due the presence of U_i. This can be seen by lagging equation (18) by one time period, as in:

$$Y_{it-1} = \alpha + \beta_1{}^* Y_{it-2} + \beta_2{}^* X_{1it-1} + \beta_3{}^* X_{2it-1} + \cdots \beta_j{}^* X_{jit-1} + U_i + \varepsilon_{it} \tag{19}$$

which shows the direct dependence of Y_{it-1} on U_i, and hence the bias produced by traditional methods of estimating (18). Moreover, one popular method for "sweeping out" the unit effect, the "fixed effect" transformation of equation (4), fails to correct the problem, as the solution for eliminating U_i from consideration produces a transformed error term of $(\varepsilon_{it} - \bar{\varepsilon}_i)$. Since $\bar{\varepsilon}_i$ contains some portion of ε_{it-1}, then the lagged endogenous variable (Y_{t-1}) is still related to the mean-differenced error term, and thus in relatively short panels the biases in estimating the dynamic effects in this model still exist.[24]

The solution to the problem lies in an adaptation of the first-difference (FD) model considered above in (6):

$$Y_{it} - Y_{it-1} = \alpha + \beta_1{}^*(Y_{it-1} - Y_{it-2}) + \beta_2{}^*(X_{1it} - X_{1it-1}) + \beta_3{}^*(X_{2it} - X_{2it-1}) + \cdots \beta_j{}^*(X_{jit} - X_{jit-1}) + (\varepsilon_{it} - \varepsilon_{it-1}) \tag{20}$$

[23] As will be seen, in the SEM version of the model, the lagged values are normally the values that are included in order to ensure model identification.

[24] As $T \to \infty, \bar{\varepsilon}_i \to 0$, and hence the error term in (18) would no longer be related to Y_{it-1} so long as no autocorrelation is present in ε_{it}. But with short panels, bias on the order of 1/T exists. See Nickell (1981).

In this model, the U_i have been differenced out of the equation, with the resulting error term the difference between the idiosyncratic error at times t and $t-1$. It can be seen, however, that the *differenced* lagged endogenous variable $(Y_{it-1} - Y_{it-2})$ will still be related to the differenced error term, given the presence of ε_{it-1} in the latter.

What are needed are instrumental variables that are related to the differenced lagged endogenous variable but unrelated to the differenced error term, and various candidates have been proposed in the literature. Anderson and Hsiao (1982) suggest several possible instruments, one being the *twice-differenced* lagged endogenous variable $(Y_{it-2} - Y_{it-3})$, and the other being the *level* of Y_{it-2}; both proposed instruments are unrelated to the error term in (20).[25] One drawback to the former solution, however, is that it requires at least four waves of data, and subsequent work also suggests that the twice-differenced lagged endogenous variable is often a poorly-performing instrument in that it is usually only weakly related to $(Y_{it-1} - Y_{it-2})$.

The Arellano-Bond solution rests on the fact that the panel structure of the data provides more and more potential instruments in equation (20) as the number of waves of observation increases (Arellano and Bond, 1991). For three wave data, Y_{it-2} may be used as an instrument for $(Y_{it-1} - Y_{it-2})$, for four waves of data Y_{it-2}, Y_{it-3} and $(Y_{it-2} - Y_{it-3})$ may be used, for five waves of data Y_{it-2}, Y_{it-3}, Y_{it-4} and all the respective changes scores may be used, and so on. So as one moves through the panel more and more instruments are included to arrive at more precise estimates of the dynamic and other effects in the model. In the Arellano-Bond formulation, the various lagged levels and differences of the exogenous X variables are also included as instruments for the same reason. One drawback to the method is its inapplicability when there is serial correlation in the original equation's (18) idiosyncratic error term, and the need for at least four waves of data to test this assumption.[26]

The strategy of using the panel structure of the data to find suitable instruments is also employed in econometric models that allow for reciprocal causality and measurement error. As has been discussed above, the statistical problem that results from either reciprocal causal effects specification, or from the presence of measurement error in variables, is an intrinsic correlation between the independent variable(s) and the structural disturbance term for that variable's equation. Hence some of the independent variables X in either (2), the original unobserved heterogeneity model, or the dynamic model (18) must be treated as endogenous. This leads naturally to the application of instrumental variables analyses in the context of first difference or fixed effects models, using lagged values of the independent variables as instruments under certain conditions (see Halaby, 2004, pp. 532–535; Woolridge, 2002, Chapter 12).

Dynamic models with unobserved heterogeneity may also be estimated within the SEM framework, though applications of this kind are less common in the literature (Dorman, 2001). An example of such a model is shown in Figure 29.4, with the basic features being the dynamic cross-lagged reciprocal effects specification considered earlier, along with an additional exogenous variable (ξ_3) that represents the individual-specific effect for each unit.

The unit-effect corresponds to the U_i term in equation (18), or, if no dynamic processes are specified, to the U_i term in the basic unobserved heterogeneity model of equation (2). It

[25] This strategy would not have worked with the fixed effect transformation, as Y_{it-2} would still be related to the portion of the error term $\bar{\varepsilon}_i$ which contains ε_{it-2}.

[26] See Wawro (2002) for discussion of alternative dynamic panel estimators, and application of these procedures to the controversy regarding the endogeneity of political party identification.

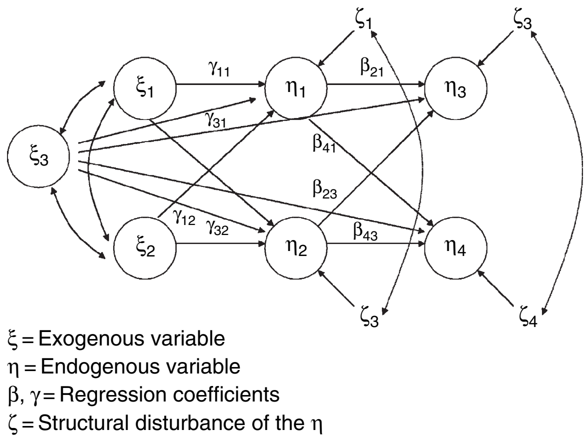


Figure 29.4 Three-wave cross-lagged model with stable unobserved variable (§$_3$)

is assumed to be stable over time (and thus can be represented by a single ξ), and it is assumed to be related to each of the other latent variables in the model as well. The unit-effect is specified within the SEM framework as an additional latent variable *with no observed indicator* and with variance set arbitrarily to 1—i.e., it is a "phantom variable" that is identified only if there is enough other excess information from the observed variances and covariances in the model. In the present case, the model is identified so long as some equality constraints are placed on the coefficients in the equations for the wave 2 and wave 3 variables, e.g., equal cross-lagged effects, equal stabilities, or equal error variances. Of course, alternative models that impose all of these constraints may also be estimated and compared. In the present example, an unobserved variable model linking party identification and presidential approval over time shows an excellent fit to the data (chi-square of 3.66 with 3 df), and the results indicate that there are no significant cross-lagged relationships between the two variables in *either* direction.

There are several attractive features of the SEM version of this model. First, the analyst may make use of the full range of SEM-related procedures for incorporating measurement error into the analysis, as the identification of the structural and measurement portions of the model are largely independent. Indeed, comprehensive models of this kind may often be necessary to estimate, as it otherwise may be unclear whether the unobserved variable represents the stable "unit-effect" or simply a latent variable that represents one or more of the observed constructs purged of measurement error. Second, alternative unobserved variable models may be tested and compared in terms of their ability to account for the observed data. This is especially useful in that a model with the unit-effect being related to only the endogenous variables over time will be nested within a model that has the unit-effect related to both endogenous and exogenous variables.[27] In this way the covariance structure analysis can provide a statistical test of the random effects versus fixed effects specification of the unobserved heterogeneity model (see Teachman, et al., 2001, for further development of this model). Finally, with enough waves of observation, more elaborate unobserved variables models may be specified, some that allow the unobserved variable to change over time (Dorman, 2001). In these ways the full power and flexibility of the SEM approach can be used to estimate models providing comprehensive attempts to overcome the most significant threats to successful causal inference in nonexperimental research.

6 Conclusion

This chapter has outlined two approaches to panel analysis, one focusing on the problem of unobserved heterogeneity, and the other focusing on dynamic causal processes

[27]That is, in one model the covariances between the unit effect and any other ξ would be fixed at zero, while in the other they would be estimated parameters.

and measurement error. Models incorporating unobserved heterogeneity were described mainly in the context of pooled econometric-type estimation, while the dynamic models with measurement error corrections were described mainly in the context of structural equation modeling procedures. As mentioned, however, recent work within the two traditions has resulted in a greater convergence of models and analytic strategies. This convergence is likely to continue, as newer developments in the field are even more explicitly synthetic in their approach.

For example, the recent work of Skrondal and Rabe-Hesketh (2004) incorporates all manner of latent variables, from unobserved heterogeneity to "true score variables" purged of measurement error in their indicators, to latent responses that represent missing values of partially-observed variables, into a single analytical framework. More generally, the realization that panel and "multilevel" data share the same logical structure (as the observations over time are nested within individual units) has led to the development of models that incorporate intratemporal growth processes at the "lower" level that may also vary randomly at the "higher" level. Such models incorporating randomly-varying intercepts and randomly-varying slopes may be estimated either through an extension of the random effects model discussed above, or with structural equation methods that treat the intercepts and slopes as "latent" variables that vary randomly across units (Bollen and Curran, 2005; Singer and Willett, 2002, Chapters 33–35 of this volume). Future developments in linear panel analysis, then, promise to build on the approaches presented in this chapter, to synthesize and to extend them in important new directions.

Acknowledgement

I thank George Krause for advice and encouragement, and Andrea Castagnola for exceptional research assistance.

Data and software

STATA 9.0 was used to estimate the pooled "econometric"-type models in the first section of this chapter. This software is available at: http://www.stata.com/. Many other software packages can be used for these models as well, including Mplus (http://www.statmodel.com/), SAS(http://www.sas.com/), MLwiN (http://www.cmm.bristol.ac.uk/), SPlus (http://www.insightful.com/products/splus/default.asp), and LIMDEP (http://www.limdep.com).

LISREL 8.72 was used to estimate the structural equation models in the second section of the chapter. This software is available at: http://www.ssicentral.com/lisrel/index.html. Other popular packages available for this kind of analysis are Mplus (http://www. statmodel.com/), SAS (PROC CALIS) (http://www.sas.com/), SPSS (AMOS) (http://www.spss.com/amos/), EQS (http://www.mvsoft.com/), MX (http://www.vcu.edu/mx/) and Smart Plus (http://www.smartpls.de/forum/).

The data used to estimate all models in this chapter can be found at www.pitt.edu/~finkel/data.htm.

References

Allison, P. (1990). Change-scores as dependent variables in regression analysis. *Sociological Methodology and Research*, 20: 93–114. AU1

Allison, P. (1994). Using panel data to estimate the effect of events. *Sociological Methodology and Research*, 23: 179–199. AU1

Anderson, T. W. and Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of Econo.*, 18: 47–82. AU2

Arellano, M. and Bond, S. (1991). Some tests of specification for panel Data: Monte Carlo evidence and an application to employment equations. *Review of Ecocomic Studies*, 58: 277–297. AU3

Arminger, G. (1987). Misspecification, asymptotic stability, and ordinal variables in the analysis of panel data. *Soc. Meth. and Res.*, 15: 336–348. AU4

Baltagi, B. H. (2005). *Econometric Analysis of Panel Data*. London: Wiley.

Beck, N. and Katz, J. (1995). What to do (and not to do) with time-series cross-section data. *American. Political Science Review*, 89: 634–647.

Bollen, K. and Curran, P. (2005). *Latent Curve Models*. New York: Wiley Interscience.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

Coleman, J. (1968). The mathematical study of change. In H. Blalock and A. Blalock (eds), *Methodology in Social Research*, pp. 428-478. New York: McGraw Hill.

Dorman, C. (2001). Modeling unmeasured third variables in longitudinal studies. *Struct. Eq. Mod.*, 8: 575–598.

Duncan, O. D. (1975). *An Introduction to Structural Equation Models*. New York: Academic Press.

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

Finkel, S. E. (2003). The Impact of the Kenya National Civic Education Programme on Democratic Attitudes, Knowledge, Values and Behavior. United States Agency for International Development Contract AEP-I-00-00-00018-00, Washington, DC: Management Systems International.

Frees, E. W. (2004). *Longitudinal and Panel Data: Analysis and Applications in the Social Sciences*. Cambridge, UK/New York: Cambridge University Press.

Green, D. and Palmquist, B. (1990). Of artifacts and partisan instability. *American Journal of Political Science*, 34: 872–902.

Greene, W. H. (2003). *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall.

Halaby, C. N. (2004). Panel models in sociological research: Theory into practice. *Annu. Rev. Sociol.*, 30: 507–544.

Hardin, J. and Hilbe, J. (2003). *Generalized Estimating Equations*. Boca Raton, Fl: Chapman & Hall.

Hausman, J. (1978). Specification tests in econometrics. *Econo.*, 46: 1251–1271.

Hausman, J. and Taylor, W. E. (1981). Panel data and unobservable individual effects. *Econo.*, 49: 1377–1398.

Heise, D. (1969). Separating reliability and stability in test-retest correlation. *American Sociological Review,* 334: 93–101.

Hsiao, C. (2002). *Analysis of Panel Data*. Cambridge UK: Cambridge University Press.

Hu, L. and Bentler, P. (1995). Evaluating model fit. In R. H. Hoyle (ed.), *Structural Equation Modeling: Concepts, Issues and Applications*. Thousand Oaks, CA: Sage.

Jagodzinski, W., Kuehnel, S. and Schmidt, P. (1987). Is there a "Socratic effect" in nonexperimental panel studies?, *Soc. Meth. and Res.*, 15: 259–303.

Jöreskog, K. and Sörbom, D. (1994). *LISREL 8 and PRELIS 8: Comprehensive Analysis of Linear Relationships in Multivariate Data*. Hillsdale, NJ: Erlbaum.

Kaplan, D. (2000). *Structural Equation Modeling: Foundations and Extensions*. Thousand Oaks, CA: Sage.

Kenny, D. A. (1979). *Correlation and Causality*. New York: Wiley.

Kessler, R. C. and Greenberg, D. F. (1981). *Linear Panel Analysis: Models of Quantitative Change*. New York: Academic Press.

Liker, J. K., Augustyniak, S. and Duncan, G. J. (1985). Panel data and models of change: A comparison of first-difference and conventional two-wave models. *Social Science Researh*, 14: 80–101.

Muthén, B. (1994). Multilevel covariance structure analysis. *Soc. Meth. and Res.*, 22: 376–398.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econo.*, 49: 1417–1426.

Plümper, T. and Troeger, V. (forthcoming). Efficient estimation of time-invariant and rarely changing variables in finite sample panel analyses with unit fixed effects. *Pol. Anal.*

Raudenbush, S. W. and Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Thousand Oaks, CA: Sage.

Singer, J. D. and Willett, J. B. (2002). *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. Oxford, UK/New York: Oxford University Press.

Skrondal, A. and Rabe-Hesketh, S. (2004). *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, Fl: Chapman & Hall/CRC.

Sörbom, D. (1982). Structural equation models with structured means. In K. Jöreskog and H. Wold (eds), *Systems under Indirect Observation, Part I*, pp. 183–195. Amsterdam: North Holland.

Teachman, J., Duncan, G., Yeung, W. J. and Levy, D. (2001). Covariance structure models for fixed and random effects. *Soc. Meth. and Res.*, 30: 271–288.

Tuma, N. B. and Hannan, M. T. (1994). *Social Dynamics*. New York: Academic Press.

Wawro, G. (2002). Estimating dynamic panel models in political science. *Pol. Anal.*, 10: 25–48.

Wheaton, B., Muthén, B., Alwin, D. and Summers, G. (1977). Assesing reliability and stability in panel models. In D. Heise (ed.), *Sociological Methodology*, pp. 84-136. San Francisco: Jossey-Bass.

Wiley, D. E. and Wiley, J. A. (1970). The estimation of measurement error in panel data. *American. Sociological Review*, 35: 112–117.

AU6

Woolridge, J. (2002). *Econometric Analysis of Cross-Sectional and Panel Data*. Cambridge, MA: MIT Press.

Zorn, C. J. W. (2001). Generalized estimating equation models for correlated data: A review with applications. *American. Journal of Political Science*, 45: 470–490.

Chapter No: 29

Query No	Contents
AU1:	Is this the correct title?
AU2:	Need the full journal title
AU3:	Need correct/full journal title
AU4:	Need the full journal title
AU5:	Need the full title of journal
AU6:	Check correct title of journal