Chapter 18

Panel Data

18.1 Introduction

Modern econometrics is divided into two branches: microeconometrics and time series analysis. The latter is covered in chapter 19. The former has many elements, of which we have discussed several examples, such as qualitative dependent variables, duration models, count data, and limited dependent variables, all of which primarily involve different types of cross-sectional data. In light of this it would seem natural to call microeconometrics cross-sectional data analysis. We do not, however, because a major category of microeconometrics involves longitudinal or panel data in which a cross-section (of people, firms, countries, etc.) is observed over time. Thanks to the computer revolution, such data sets, in which we have observations on the same units in several different time periods, are more common and have become more amenable to analysis.

Two prominent examples of panel data are the PSID (Panel Study of Income Dynamics) data and the NLS (National Longitudinal Surveys of Labor Market Experience) data, both of which were obtained by interviewing several thousand people over and over again through time. These data sets were designed to enable examination of the causes and nature of poverty in the United States, by collecting information on such things as employment, earnings, mobility, housing, and consumption behavior. Indeed, thousands of variables were recorded. These data are typical of panel data in that they are short and wide, consisting of a very large number of cross-sectional units observed over a small number of time periods. Such data are expensive to obtain, involving tracking large numbers of people over extended time periods. Is this extra expense warranted?

Panel data have several attractive features that justify this extra cost, four of which are noted below.

1. Panel data can be used to deal with heterogeneity in the micro units. In any cross-section there is a myriad of unmeasured explanatory variables that affect

the behavior of the people (firms, countries, etc.) being analyzed. (Heterogeneity means that these micro units are all different from one another in fundamental unmeasured ways.) Omitting these variables causes bias in estimation. The same holds true for omitted time series variables that influence the behavior of the micro units uniformly, but differently in each time period. Panel data enable correction of this problem. Indeed, some would claim that the ability to deal with this omitted variable problem is the main attribute of panel data.

2. Panel data create more variability, through combining variation across micro units with variation over time, alleviating multicollinearity problems. With this more informative data, more efficient estimation is possible.
3. Panel data can be used to examine issues that cannot be studied using time series or cross-sectional data alone. As an example, consider the problem of separating economies of scale from technological change in the analysis of production functions. Cross-sectional data can be used to examine economies of scale, by comparing the costs of small and large firms, but because all the data come from one time period there is no way to estimate the effect of technological change. Things are worse with time series data on a single firm; we cannot separate the two effects because we cannot tell if a change in that firm's costs over time is due to technological change or due to a change in the size of the firm. As a second example, consider the distinction between temporary and long-term unemployment. Cross-sectional data tell us who is unemployed in a single year, and time series data tell us how the unemployment level changed from year to year. But neither can tell us if the same people are unemployed from year to year, implying a low turnover rate, or if different people are unemployed from year to year, implying a high turnover rate. Analysis using panel data can address the turnover question because these data track a common sample of people over several years.
4. Panel data allow better analysis of dynamic adjustment. Cross-sectional data can tell us nothing about dynamics. Time series data need to be very lengthy to provide good estimates of dynamic behavior, and then typically relate to aggregate dynamic behavior. Knowledge of individual dynamic reactions can be crucial to understanding economic phenomena. Panel data avoid the need for a lengthy time series by exploiting information on the dynamic reactions of each of several individuals.

18.2 Allowing for Different Intercepts

Suppose an individual's consumption y is determined linearly by his or her income x and we have observations on a thousand individuals ($N = 1000$) in each of four time periods ($T = 4$). A plot of all the data produces a scatter shown in simplified form (only a few observations are shown, not all 4000 observations!) in Figure 18.1. (Ignore the ellipses for the moment.) If we were to run ordinary least squares (OLS), we would produce a slope estimate shown by the line AA drawn through these data. But now suppose we identify these data by the cross-sectional unit (person, firm, or country, for example) to which they belong, in this case a person. This is shown in Figure 18.1 by drawing an ellipse for each person, surrounding all four time series observations

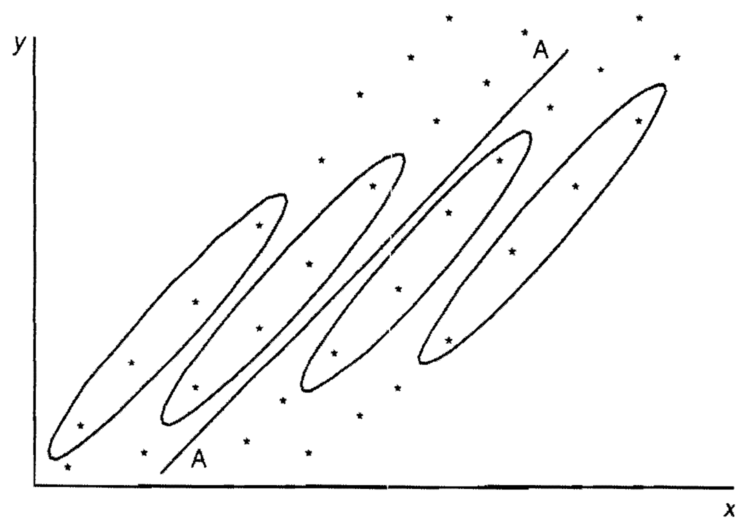

Figure 18.1 Panel data showing four observations on each of four individuals.

on that person. (There would be a thousand such ellipses in the actual data scatterplot, with roughly half above and half below *AA*; only four are drawn in Figure 18.1.) This way of viewing the data reveals that although each person in this example has the same slope, these people all have different intercepts. Most researchers would agree that this cross-sectional heterogeneity is the normal state of affairs – there are so many unmeasured variables that determine *y* that their influence gives rise to a different intercept for each individual. This phenomenon suggests that OLS is biased unless the influence of these omitted variables (embodied in different intercepts) is uncorrelated with the included explanatory variables. Two ways of improving estimation have been suggested, associated with two different ways of modeling the presence of a different intercept for each cross-sectional unit.

The first way is to put in a dummy for each individual (and omit the intercept). Doing this allows each individual to have a different intercept, and so OLS including all these dummies should guard against the bias discussed above. This "fixed effect" model gives rise to what is called the *fixed effects estimator* – OLS applied to the fixed effects model. At first glance this seems as though it would be difficult to estimate because (in our example above) we would require a thousand dummies. It turns out that a computational trick avoids this problem via an easy transformation of the data. This transformation consists of subtracting from each observation the average of the values within its ellipse – the observations for each individual have subtracted from them the averages of all the observations for that individual. OLS on these transformed data produces the desired slope estimate.

The fixed effects model has two major drawbacks:

1. By implicitly including a thousand dummy variables we lose 999 degrees of freedom (by dropping the intercept we save one degree of freedom). If we could find some way of avoiding this loss, we could produce a more efficient estimate of the common slope.

2. The transformation involved in this estimation process wipes out all explanatory variables that do not vary within an individual. This means that any explanatory variable that is time-invariant, such as gender, race, or religion, disappears, and so we are unable to estimate a slope coefficient for that variable. (This happens because within the ellipse in Figure 18.1, the values of these variables are all the same so that when we subtract their average they all become zero.)

The second way of allowing for different intercepts, the "random effects" model, is designed to overcome these two drawbacks of the fixed effects model. This model is similar to the fixed effects model in that it postulates a different intercept for each individual, but it interprets these differing intercepts in a novel way. This procedure views the different intercepts as having been drawn from a bowl of possible intercepts, so they may be interpreted as random (usually assumed to be normally distributed) and treated as though they were a part of the error term. As a result, we have a specification in which there is an overall intercept, a set of explanatory variables with coefficients of interest, and a composite error term. This composite error has two parts. For a particular individual, one part is the "random intercept" term, measuring the extent to which this individual's intercept differs from the overall intercept. The other part is just the traditional random error with which we are familiar, indicating a random deviation for that individual in that time period. For a particular individual the first part is the same in all time periods; the second part is different in each time period.

The trick to estimation using the random effects model is to recognize that the variance–covariance matrix of this composite error is nonspherical (i.e., not all off-diagonal elements are zero). In the example above, for all four observations on a specific individual, the random intercept component of the composite error is the same, so these composite errors will be correlated in a special way. Observations on different individuals are assumed to have zero correlation between their composite errors. This creates a variance–covariance matrix with a special pattern. The *random effects estimator* estimates this variance–covariance matrix and performs estimated generalized least squares (EGLS). The EGLS calculation is done by finding a transformation of the data that creates a spherical variance–covariance matrix and then performing OLS on the transformed data. In this respect it is similar to the fixed effects estimator except that it uses a different transformation.

18.3 Fixed Versus Random Effects

By saving on degrees of freedom, the random effects model produces a more efficient estimator of the slope coefficients than the fixed effects model. Furthermore, the transformation used for the random effects estimation procedure does not wipe out the explanatory variables that are time-invariant, allowing estimation of coefficients on variables such as gender, race, and religion. These results suggest that the random effects model is superior to the fixed effects model. So should we always use the

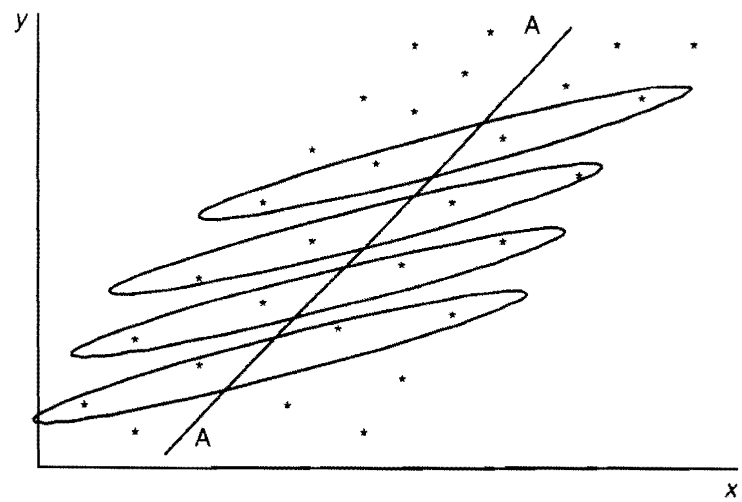

Figure 18.2 Panel data showing four observations on each of four individuals, with positive correlation between x and the intercept.

random effects model? Unfortunately, the random effects model has a major qualification that makes it applicable only in special circumstances.

This qualification is illustrated in Figure 18.2, where the data look exactly the same as in Figure 18.1, but the ellipses are drawn differently, to reflect a different allocation of observations to individuals. All persons have the same slope and different intercepts, just as before, but there is a big difference now – the common slope is not the same as the slope of the *AA* line, as it was in Figure 18.1. The main reason for this is that *the intercept for an individual is larger the larger is that individual's x value.* (Lines drawn through the observations in ellipses associated with higher x values cut the y axis at larger values.) This causes the OLS estimate using all the data to produce the *AA* line, clearly an overestimate of the common slope. This happens because as we move toward a higher x value, the y value increases for two reasons. First, it increases because the x value increases, and second, because there is likely to be a higher intercept. OLS estimation is biased upward because when x changes, OLS gives it credit for both of these y changes.

This bias does not characterize the fixed effects estimator because as described earlier the different intercepts are explicitly recognized by putting in dummies for them. But it is a problem for the random effects estimator because rather than being explicitly recognized, the intercepts are incorporated into the (composite) error term. As a consequence, the composite error term will tend to be bigger whenever the x value is bigger, creating correlation between x and the composite error term. Correlation between the error and an explanatory variable creates bias. As an example, suppose that wages are being regressed on schooling for a large set of individuals, and that a missing variable, ability, is thought to affect the intercept. Since schooling and ability are likely to be correlated, modeling this as a random effect will create correlation between the composite error and the regressor schooling, causing the random effects estimator to be biased. The bottom line here is that the random effects estimator should only be used whenever we are confident that its composite error is uncorrelated with the explanatory variables. A test for this, a variant of the *Hausman test* (discussed in the general notes), is based on seeing if the random effects estimate is insignificantly different from the unbiased fixed effects estimate.

Here is a summary of the discussion above. Estimation with panel data begins by testing the null that the intercepts are equal. If this null is accepted the data are pooled. If this null is rejected, a Hausman test is applied to test if the random effects estimator is unbiased. If this null is not rejected, the random effects estimator is used; if this null is rejected, the fixed effects estimator is used. For the example shown in Figure 18.1, OLS, fixed effects, and random effects estimators are all unbiased, but random effects is most efficient. For the example shown in Figure 18.2, OLS and random effects estimators are biased, but the fixed effects estimator is not.

There are two kinds of variation in the data pictured in Figures 18.1 and 18.2. One kind is variation from observation to observation within a single ellipse (i.e., variation *within* a single individual). The other kind is variation in observations from ellipse to ellipse (i.e., variation *between* individuals). The fixed effects estimator uses the first type of variation (in all the ellipses), ignoring the second type. Because this first type of variation is variation *within* each cross-sectional unit, the fixed effects estimator is sometimes called the "within" estimator. An alternative estimator can be produced by using the second type of variation, ignoring the first type. This is done by finding the average of the values within each ellipse and then running OLS on these average values. This is called the "between" estimator because it uses variation between individuals (ellipses). Remarkably, the OLS estimator on the pooled data is an unweighted average of the within and between estimators. The random effects estimator is a (matrix-) weighted average of these two estimators. Three implications of this are of note.

1. This is where the extra efficiency of the random effects estimator comes from – it uses information from both the within and the between estimators.
2. This is how the random effects estimator can produce estimates of coefficients of time-invariant explanatory variables – these variables vary between ellipses, but not within ellipses.
3. This is where the bias of the random effects estimator comes from when the explanatory variable is correlated with the composite error – the between estimator is biased. The between estimator is biased because a higher x value gives rise to a higher y value both because x is higher and because the composite error is higher (because the intercept is higher) – the estimating formula gives the change in x all the credit for the change in y.

18.4 Short Run Versus Long Run

Suppose that an individual's consumption (y) is determined in the long run by his or her level of income (x), producing a data plot such as that in Figure 18.2. But suppose that due to habit persistence, in the short run the individual adjusts consumption only partially when income changes. A consequence of this is that within an ellipse in Figure 18.2, as an individual experiences changes in income, changes in consumption are modest, compared to the long-run changes evidenced as we move from ellipse to ellipse (i.e., from one individual's approximate long-run income level to another individual's approximate long-run income level). If we had observations on only one cross-section we would have one observation (for the first time period, say) from each ellipse and an OLS regression would produce an estimate of the long-run relationship between consumption and income. If we had observations on only one cross-sectional unit over time (i.e., observations within a single ellipse) an OLS regression would produce an estimate of the short-run relationship between consumption and income. This explains why, contrary to many people's intuition, cross-sectional data are said to estimate long-run relationships whereas time series data estimate short-run relationships.

Because the fixed effects estimator is based on the time series component of the data, it estimates short-run effects. And because the random effects estimator uses both the cross-sectional and time series components of the data, it produces estimates that mix the short-run and long-run effects. A lesson here is that whenever we have reason to believe that there is a difference between short- and long-run reactions, we must build the appropriate dynamics into the model specification, such as by including a lagged value of the dependent variable as an explanatory variable.

One of the advantages of panel data is that they can be used to analyze dynamics with only a short time series. For a time series to reveal dynamic behavior it must be long enough to provide repeated reactions to changes – without such information the estimating procedure would be based on only a few reactions to change and so the resulting estimates could not be viewed with confidence. The power of panel data is that the required repeated reactions are found by looking at the reactions of the N different cross-sectional units, avoiding the need for a long time series.

Modeling dynamics typically involves including a lagged value of the dependent variable as an explanatory variable. Unfortunately, fixed and random effect estimators are biased in this case; to deal with this, special estimation procedures have been developed, as discussed in the general notes.

18.5 Long, Narrow Panels

The exposition above is appropriate for the context of a wide, short panel, in which N, the number of cross-sectional units, is large, and T, the number of time periods, is small. Whenever we have a long, narrow panel, analysis is typically undertaken in a different fashion. With a lot of time series observations on each of a small number of cross-sectional units, it is possible to estimate a separate equation for each cross-sectional unit. Consequently, the estimation task becomes one of finding some way to improve estimation of these equations by estimating them together. Suppose, for illustrative purposes, we have six firms each with observations over 30 years, and we are estimating an equation in which investment y is a linear function of expected profit x.

whenever we are confident that its composite error is uncorrelated with the explanatory variables. A test for this, a variant of the *Hausman test* (discussed in the general notes), is based on seeing if the random effects estimate is insignificantly different from the unbiased fixed effects estimate.

Here is a summary of the discussion above. Estimation with panel data begins by testing the null that the intercepts are equal. If this null is accepted the data are pooled. If this null is rejected, a Hausman test is applied to test if the random effects estimator is unbiased. If this null is not rejected, the random effects estimator is used; if this null is rejected, the fixed effects estimator is used. For the example shown in Figure 18.1, OLS, fixed effects, and random effects estimators are all unbiased, but random effects is most efficient. For the example shown in Figure 18.2, OLS and random effects estimators are biased, but the fixed effects estimator is not.

There are two kinds of variation in the data pictured in Figures 18.1 and 18.2. One kind is variation from observation to observation within a single ellipse (i.e., variation *within* a single individual). The other kind is variation in observations from ellipse to ellipse (i.e., variation *between* individuals). The fixed effects estimator uses the first type of variation (in all the ellipses), ignoring the second type. Because this first type of variation is variation *within* each cross-sectional unit, the fixed effects estimator is sometimes called the "within" estimator. An alternative estimator can be produced by using the second type of variation, ignoring the first type. This is done by finding the average of the values within each ellipse and then running OLS on these average values. This is called the "between" estimator because it uses variation between individuals (ellipses). Remarkably, the OLS estimator on the pooled data is an unweighted average of the within and between estimators. The random effects estimator is a (matrix-) weighted average of these two estimators. Three implications of this are of note.

1. This is where the extra efficiency of the random effects estimator comes from – it uses information from both the within and the between estimators.
2. This is how the random effects estimator can produce estimates of coefficients of time-invariant explanatory variables – these variables vary between ellipses, but not within ellipses.
3. This is where the bias of the random effects estimator comes from when the explanatory variable is correlated with the composite error – the between estimator is biased. The between estimator is biased because a higher x value gives rise to a higher y value both because x is higher and because the composite error is higher (because the intercept is higher) – the estimating formula gives the change in x all the credit for the change in y.

18.4 Short Run Versus Long Run

Suppose that an individual's consumption (y) is determined in the long run by his or her level of income (x), producing a data plot such as that in Figure 18.2. But suppose that due to habit persistence, in the short run the individual adjusts consumption only

There are several different ways in which the six equations (one for each firm) could be estimated together so as to improve efficiency.

1. We could assume that the intercept and slope coefficients are the same for each firm, in which case the data could be pooled and OLS used to estimate the single intercept and single slope.
2. More realistically, we could assume the six slopes to be the same but the intercepts to be different. By putting in dummies for the intercept differences, we could estimate a single equation by OLS, using all the data.
3. Even more realistically, in addition to assuming different intercepts (and equal slopes) we could assume that the variance of the error term is different for each equation. A single equation would be estimated by EGLS.
4. We could assume contemporaneous correlation among the cross-sectional errors. This would allow the error in the fourth equation, for example, to be correlated with the error in the fifth (and all other equations) *in the same time period*. Correlations between errors in different time periods are assumed to be zero. Estimation would be by EGLS following the SURE (seemingly unrelated estimation) procedure described in chapter 11.
5. We could allow the errors in each of the six equations to have different variances, and be autocorrelated within equations, but uncorrelated across equations.

Before choosing one of these estimation procedures we need to test the relevant assumptions to justify our choice. A variety of tests is available, as discussed in the general and technical notes to this section.

General Notes

18.1 Introduction

- Baltagi (2005) is an excellent source of information on panel data procedures, with extensive reference to its burgeoning literature. His introductory chapter (pp. 1–9) contains a description of the nature of prominent panel data sets, references to sources of panel data, examples of applications of these data, an exposition of the advantages of panel data, and discussion of limitations of panel data. Hsiao (2003a) is another well-known survey. Cameron and Trivedi (2005, pp. 58–9) has a concise description of several sources of microeconomic data, some of which are panel data. Pergamit *et al.* (2001) is a good description of the NLS data. Limitations of panel data include data collection problems, distortions caused by measurement errors that plague survey data, problems caused by the typically short time dimension, and sample selection problems due to self-selection, nonresponse, and attrition.
- Greene (2008, chapter 9) has a good textbook exposition of relationships among various estimators, computational considerations, and relevant test statistics.
- The second dimension of panel data need not be time. For example, we could have data on twins (or sisters), in which case the second "time period" for an individual is not an observation on that individual in a different time period but rather an observation on his or her twin (or one of her sisters). As another example we might have

data on N individuals writing a multiple-choice exam with T questions.

- Most panel data has a time dimension, so problems associated with time series analysis can become of concern. In particular, unit roots and cointegration may need to be tested for and accommodated in a panel data analysis. Some commentary on this dimension of panel data is provided in chapter 19 on time series.

18.2 Allowing for Different Intercepts

- The "fixed effects estimator" is actually the "OLS estimator applied when using the fixed effects model," and the "random effects estimator" is actually the "EGLS estimator applied when using the random effects model." This technical abuse of econometric terminology has become so common that it is understood by all as to what is meant and so should not cause confusion.
- The transformation used to produce the fixed effects estimator takes an individual's observation on an explanatory variable and subtracts from it the average of all of that individual's observations on that explanatory variable. In terms of Figures 18.1 and 18.2, each observation within an ellipse has subtracted from it its average value within that ellipse. This moves all the ellipses so that they are centered on the origin. The fixed effects estimate of the slope is produced by running OLS on all these observations, without an intercept.
- The fixed effects transformation is not the only transformation that removes the individual intercepts. An alternative transformation is first differencing – by subtracting the first period's observation on an individual from the second period's observation on that same individual, for example, the intercept for that individual is eliminated. Running OLS on the differenced data produces an alternative to the fixed effects estimator. If there are only two time periods, these two estimators are identical. When there are more than two time periods the choice between them rests on assumptions about the error term in the relationship being estimated. If the errors are serially uncorrelated, the fixed effects estimator is more efficient, whereas if the errors follow a random walk (discussed in chapter 19) the first-differencing estimator is more efficient. Wooldridge (2002, pp. 284–5) discusses this problem and the fact that these two estimators will both be biased, but in different ways, whenever the explanatory variables are not independent of the error term. In practice first differencing appears to be used mainly as a means of constructing estimators used when a lagged value of the dependent variable is a regressor, as discussed in the general notes to section 18.4.
- The random effects transformation requires estimates of the variance of each of the two components of the "composite" error – the variance of the "random intercepts" and the variance of the usual error term. Several different ways of producing these estimates exist. For example, fixed effects estimation could be performed, with the variance of the intercept estimates used to estimate the variance of the "random intercepts," and the variance of the residual used to estimate the variance of the usual error term. Armed with these estimates, random effects estimation can be performed. Monte Carlo studies suggest use of whatever estimates are computationally easiest.
- Both fixed and random effects estimators assume that the slopes are equal for all cross-sectional units. Robertson and Symons (1992) claim that this is hard to detect and that even small differences in slopes can create substantial bias, particularly in a dynamic context. On the other hand, Baltagi, Griffen, and Xiong (2000) claim that although some bias may be created, the efficiency gains from the pooling more than offset this. This view is supported by Attanasio, Picci, and Scorcu (2000).
- Whenever the number of time period observations for each cross-section is not the same we have an *unbalanced* panel. This requires modification to estimation, built into panel data estimation software. Extracting a balanced panel out of an unbalanced data set is not advised – doing so leads to a substantial loss of efficiency. As always, one must ask why the data are missing to be alert to selection bias problems; a check for selection bias here can take the form of comparing balanced and unbalanced estimates.

- The fixed and random effects estimators discussed in the body of this chapter were explained in the context of each individual having a different intercept. It is also possible for each time period to have a different intercept. In the second time period there may have been a big advertising campaign, for example, so everyone's consumption of the product in question may have risen during that period. In the fixed effects case, to deal with this dummies are added for the different time periods. In the random effects case, a time-period-specific error component is added. When there are intercept differences across both individuals and time periods, we speak of a two-way effects model, to distinguish it from the one-way effect model in which the intercepts differ only across individuals. Estimation is similar to the one-way effects case, but the transformations are more complicated. The one-way effect model is used far more often than the two-way effects model.

18.3 Fixed Versus Random Effects

- Another way of summarizing the difference between the fixed and random effects estimators is in terms of omitted variable bias. If the collective influence of the unmeasured omitted variables (that give rise to the different intercepts) is uncorrelated with the included explanatory variables, omitting them will not cause any bias in OLS estimation. In this case they can be bundled into the error term and efficient estimation undertaken via EGLS – the random effects estimator is appropriate. If, however, the collective influence of these omitted unmeasured variables is correlated with the included explanatory variables, omitting them causes OLS bias. In this case, they should be included to avoid this bias. The fixed effects estimator does this by including a dummy for each cross-sectional unit.
- There are two ways to test if the intercepts are different from one another. (If they do not differ from one another, OLS on the pooled data is the estimator of choice.) One way is to perform the fixed effects estimation and calculate the corresponding dummy variable coefficient (intercept) estimates. Do an *F* test in the usual way (a Chc test) to test if the coefficients on the dummy vai ables are identical. Another way is to perform tl random effects estimation and test if the varian of the intercept component of the composite err term is zero, using a Lagrange multiplier (LN test developed by Breusch and Pagan (198(as described for example in Greene (20C pp. 205–6). Be careful here – a common err among practitioners is to think that this LM te is testing for the appropriateness of the randc effects model, which it does not. To test f whether we should use the fixed or the rando effects estimator we need to test for whether t random effects estimator is unbiased, as explain below.
- The random effects estimator (sometimes call the *variance components* or *error componer* estimator) is recommended whenever it is u biased (i.e., whenever its composite error uncorrelated with the explanatory variables, explained earlier). This is an example of testi for independence between the error term and t explanatory variables, for which, as explain in chapter 9, the Hausman test is appropria Regardless of the truth of the null, the fix effects estimator is unbiased because it includ dummies for the different intercepts. But the ra dom effects estimator is unbiased only if the nı is true. Consequently, if the null is true the fix and random effects estimators should be appro: mately equal, and if the null is false they shou be different. The Hausman test tests the null testing if these two estimators are insignificant different from one another. Fortunately, there an easy way to conduct this test. Transform t data to compute the random effects estimator, t when regressing the transformed dependent va able on the transformed independent variabl add an extra set of independent variables, name the explanatory variables transformed for fix effects estimation. The Hausman test is calc lated as an *F* test for testing the coefficients these extra explanatory variables against zero.
- The fixed effects estimator is more robust selection bias problems than is the random effe estimator because if the intercepts incorpor

selection characteristics they are controlled for in the fixed effects estimation.

- One other consideration is sometimes used when deciding between fixed and random effects estimators. If the data exhaust the population (say, observations on all firms producing automobiles), then the fixed effects approach, which produces results conditional on the cross-section units in the data set, seems appropriate because these are the cross-sectional units under analysis. Inference is confined to these cross-sectional units, which is what is relevant. On the other hand, if the data are a drawing of observations from a large population (say, a thousand individuals in a city many times that size), and we wish to draw inferences regarding other members of that population, the random effects model seems more appropriate (so long as the random effects composite error is not correlated with the explanatory variables).
- The "between" estimator (OLS when each observation is the average of the data inside an ellipse) has some advantages as an estimator in its own right. Because it averages variable observations, it can reduce the bias caused by measurement error (by averaging out the measurement errors). In contrast, transformations that wipe out the individual intercept effect, such as that of the fixed effect estimator, may aggravate the measurement error bias (because all the variation used in estimation is variation within individuals, which is heavily contaminated by measurement error; in the PSID data, for example, it is thought that as much as 80% of wage *changes* is due to measurement error!). Similarly, averaging may alleviate the bias caused by correlation between the error and the explanatory variables.

18.4 Short Run Versus Long Run

- If a lagged value of the dependent variable appears as a regressor, both fixed and random effects estimators are biased. The fixed effects transformation subtracts each unit's average value from each observation. Consequently, each transformed value of the lagged dependent variable for that unit involves all the error terms associated with that unit, and so is contemporaneously correlated with the transformed error. Things are even worse for random effects because a unit's random intercept appears directly as an element of the composite error term and as a determinant of the lagged value of the dependent variable.

  One way of dealing with this problem is by using the first-differencing transformation to eliminate the individual effects (the heterogeneity), and then finding a suitable instrument to apply IV estimation. The first-differencing transformation is popular because for this transformation it is easier to find an instrument, in this case a variable that is correlated with the first-differenced lagged value of the dependent variable but uncorrelated with the first-differenced error. A common choice of instrument is y_{t-2} used as an instrumental variable for $(\Delta y)_{t-1}$, as suggested by Anderson and Hsiao (1981). This procedure does not make use of a large number of additional moment conditions, such as that higher lags of y are not correlated with $(\Delta y)_{t-1}$. This has led to the development of several GMM (generalized method of moments) estimators. Baltagi (2005, chapter 8) has a summary of all this, with references to the literature. One general conclusion, consistent with results reported earlier for GMM, is that researchers should avoid using a large number of IVs or moment conditions. See, for example, Harris and Mitayas (2004).
- How serious is the bias when using the lagged value of the dependent variable in a fixed effects panel data model? A Monte Carlo study by Judson and Owen (1999) finds that even with $T = 30$ this bias can be as large as 20%. They investigate four competing estimators and find that a "bias-corrected" estimator suggested by Kiviet (1995) is best. Computational difficulties with this estimator render it impractical in unbalanced panels, in which case they recommend the usual fixed effects estimator when T is greater than 30 and a GMM estimator (with a restricted number of moment conditions) for T less than 20, and note that the computationally simpler IV estimator of Anderson and Hsiao (1981) can be used when T is greater than 20. A general conclusion here, as underlined by Attanasio, Picci, and Scorcu (2000), is that for T greater than 30, the bias

created by using the fixed effects estimator is more than offset by its greater precision compared to IV and GMM estimators.

18.5 Long, Narrow Panels

- Greene (2008, chapter 10) has a good exposition of the several different ways in which estimation can be conducted in the context of long, narrow panels. A Chow test (as described in the general notes to section 15.4) can be used to test for equality of slopes across equations. Note, though, that if there is reason to believe that errors in different equations have different variances, or that there is contemporaneous correlation between the equations' errors, such testing should be undertaken by using the SURE estimator, not OLS; as explained in chapter 8, inference with OLS is unreliable if the variance–covariance matrix of the error is nonspherical. If one is not certain whether the coefficients are identical, Maddala (1991) recommends shrinking the separate estimates towards some common estimate. Testing for equality of variances across equations, and zero contemporaneous correlation among errors across equations, can be undertaken with a variety of LM, W, and LR tests, all described clearly by Greene.
- Estimating several equations together improves efficiency only if there is some connection among these equations. Correcting for different error variances across equations, for example, will yield no benefit if there is no constraint across the equations enabling the heteroskedasticity correction to improve efficiency. The main examples of such constraints are equality of coefficients across equations (they all have the same slope, for example), and contemporaneous correlation among errors as described in the general and technical notes of section 11.1 when discussing SURE. The qualifications to SURE introduced by Beck and Katz (1995, 1996), discussed in the technical notes to section 11.1, are worth reviewing for the context of long, narrow panels.
- Long, wide panels, such as the Penn World Tables widely used to study growth, are becoming more common. In this context the slope coefficients are often assumed to differ randomly and interest focuses on estimating the average effect of an explanatory variable. Four possible estimating procedures seem reasonable: estimate a separate regression for each unit and average the resulting coefficient estimates; estimate using fixed or random effects models assuming common slopes; average the data over units and estimate using these aggregated time series data; and average the data over time and use a cross-section regression on the unit averages. Although all four estimation procedures are unbiased when the regressors are exogenous, Pesaran and Smith (1995) show that when a lagged value of the dependent variable is present, only the first of these methods is asymptotically unbiased.
- A popular way of analyzing macroeconomic growth with large-N, large-T panel data is to use five- or ten-year averages of the data. The idea is that this will alleviate business-cycle effects and measurement error. Attanasio, Picci, and Scorcu (2000) argue that this is undesirable because it throws away too much information.

Technical Notes

18.2 Allowing for Different Intercepts

- The fixed effects estimator can be shown to be an instrumental variable estimator with the deviations from individual means as the instruments. This insight has been used to develop alternative instrumental variable estimators for this context. Verbeek (2000, pp. 321–2) is a textbook exposition.
- In addition to having different intercepts, each individual may have a different trend. First differencing the data will eliminate the different intercepts and convert the different trends into different intercepts for the first-differenced data.

18.3 Fixed versus Random Effects

- The transformation for fixed effects estimation is very simple to derive. Suppose the observation for the ith individual in the tth time period is written

$$y_{it} = \alpha_i + \beta x_{it} + \varepsilon_{it} \tag{18.1}$$

If we average the observations on the ith individual over the T time periods for which we have data on this individual we get

$$\bar{y}_i = \alpha_i + \beta\bar{x}_i + \bar{\varepsilon}_i \quad (18.2)$$

Subtracting equation (18.2) from equation (18.1) we get

$$y_{it} - \bar{y}_i = \beta(x_{it} - \bar{x}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

The intercept has been eliminated. OLS of $y^*_{it} = y_{it} - \bar{y}_i$ on $x^*_{it} = x_{it} - \bar{x}_i$ produces the fixed effects estimator. Computer software estimates the variance of the error term by dividing the sum of squared errors from this regression by $NT - K - N$ rather than by $NT - K$, in recognition of the N estimated means.

- For random effects estimation the estimating equation is written as

$$y_{it} = \mu + \beta x_{it} + (u_i + \varepsilon_{it})$$

where μ is the mean of the "random" intercepts $\alpha_i = \mu + u_i$, and the errors u_i and ε_{it} in the composite error term have variances σ_u^2 and σ_ε^2, respectively.

The transformation for random effects estimation can be shown to be

$$y^*_{it} = y_{it} - \theta\bar{y}_i \quad \text{and} \quad x^*_{it} = x_{it} - \theta\bar{x}_i$$
$$\text{where } \theta = 1 - \frac{\sigma_\varepsilon}{\sqrt{T\sigma_u^2 + \sigma_\varepsilon^2}}$$

This is derived by figuring out what transformation will make the transformed residuals such that they have a spherical variance–covariance matrix.

Notice that if all the individuals had the same intercept, so that $\sigma_u^2 = 0$, θ becomes 0 and the random effects estimator becomes OLS on all the raw data, as makes sense. A better way of looking at special cases is to draw on the result that the random effects estimator is a matrix-weighted average of the fixed effects (the "within") estimator and the "between" estimator (recall that the "between" estimator estimates the slope by running OLS on data averaged across time for each individual). For expository purposes this can be written as

$$\text{random effects} = \text{fixed effects} + \lambda \text{ between}$$

$$\text{where } \lambda = (1-\theta)^2 = \frac{\sigma_\varepsilon^2}{T\sigma_u^2 + \sigma_\varepsilon^2}$$

The fixed effects estimator ignores information provided by the "between" estimator, whereas the random effects estimator tries to use this information. The "between" estimator allows the differing intercepts to play a prominent role. This happens because the averaged data have attached to them averaged errors embodying a common intercept. Minimizing the sum of these squared errors allows the differing intercepts to have a heavy influence on the estimated slope. By eliminating the intercepts, fixed effects wipes out this influence. By not eliminating the intercepts, random effects allows this influence to play a role. The smaller the variance in the intercepts (and thus the weaker the justification for ignoring them via fixed effects), the greater λ and so the greater is the role for the "between" estimator in the random effects estimator.

To make all of this work we need estimates of the variances σ_u^2 and σ_ε^2, so that an estimate of θ can be produced. Typically σ_ε^2 is estimated as the estimated error variance from the within estimation, and σ_u^2 is estimated as the estimated error variance from the between estimation less $1/T$ times the σ_ε^2 estimate. Notice that if the number of cross-sectional units is small, the between estimator will not have many observations and so will likely produce a poor estimate of σ_u^2. This suggests that the random effects estimator should not be used whenever there is a small number of cross-sectional units.

- Asymptotic analysis in the context of panel data is complicated by the issue of what should be allowed to go to infinity, N or T? Asymptotic justification for random effects requires that the number of cross-sections N grow, to enable the variance of the distribution of the intercepts to be estimated using more and more observations and so be consistent. But for fixed effects it is the number of time periods T that must grow, to enable each of the N intercepts to be estimated using more and more observations and so be consistent.

- Suppose the dependent variable in a panel data set is qualitative. An obvious extension of the fixed effects method would be to estimate using logit or probit, allowing (via dummies) each individual to have a different intercept in the index function. Because of the nonlinearity of the logit/probit specification, there is no easy way to transform the data to eliminate the intercepts, as was done for the linear regression case. Consequently, estimation requires estimating the N intercepts (by including N dummies) along with the slope coefficients. Although these maximum likelihood estimates (MLEs) are consistent, because of the nonlinearity all estimates are biased in small samples. The dummy variable coefficient (the intercept) estimate for each individual is based on T observations; because in most applications T is small, this produces a bias that cannot be ignored. This is referred to as the "incidental parameters" problem: as N becomes larger and larger more and more parameters (the intercepts) need to be estimated, preventing the manifestation of consistency (unless T also grows). The bottom line here is that T needs to be sufficiently large (20 or more should be large enough) to allow this logit/probit fixed effects estimation procedure to be acceptable. Greene (2004) reports Monte Carlo results measuring the impact of using fixed effects in nonlinear models such as logit/probit, Tobit, and selection models.
- There is a caveat to the fixed effects logit/probit described above. If the dependent variable observation for an individual is one in all time periods, traditional likelihood maximization breaks down because any infinitely large intercept estimate for that individual creates a perfect fit for the observations on that individual – the intercept for that individual is not estimable. Estimation in this context requires throwing away observations on individuals with all one or all zero observations. An example of when this procedure should work well is when we have N students answering, say 50 multiple-choice exam questions, and nobody scored zero or 50 correct.
- But what if T is small, as is typically the case fo panel data? In this case, for logit (but not for pro bit), a clever way of eliminating the intercepts i possible by maximizing a likelihood conditiona on the sum of the dependent variable values fo each individual. Suppose $T = 3$ and for the *i*th individual the three observations on the dependen variable are (0, 1, 1), in that order. The sum o these observations is 2. Conditional on the sun of these observations equal to 2, the probability of (0, 1, 1) is calculated by the expression for the unconditional probability for (0, 1, 1), given by the usual logit formula for three observations divided by the sum of the unconditional probabili ties for all the different ways in which the sum o the dependent variable observations could be 2 namely (0, 1, 1), (1, 1, 0), and (1, 0, 1). In words if an individual has two ones in the three time peri ods, what is the probability that these two ones occurred during the second and third time peri ods rather than in some other way? Greene (2008 pp. 803–5) has an example showing how this pro cess eliminates the intercepts. This process (maxi mizing the conditional likelihood) is the usual way in which fixed effects estimation is undertaken fo qualitative dependent variable panel data models in econometrics. Larger values of T cause calcula tions to become burdensome, but software (such as LIMDEP) has overcome this problem.
- This technique of maximizing the conditiona likelihood cannot be used for probit, because fo probit the algebra described above does not elim inate the intercepts. Probit is used for randon effects estimation in this context, however. In this case the usual maximum likelihood approach is used, but it becomes computationally complicated because the likelihood cannot be written as the product of individual likelihoods (because some observations pertain to the same individual and so cannot be considered to have been generated independently). See Baltagi (2005, pp. 209–15 for discussion.
- Baltagi (2005, pp. 215–6) summarizes recen computational innovations (based on simulation

in estimating limited dependent variable models with panel data. Wooldridge (1995) suggests some simple tests for selection bias and ways to correct for such bias in linear fixed effects panel data models.

- Wooldridge (2002, pp. 262–3, 274–6) exposits estimation of robust variance–covariance matrices for random and fixed effects estimators.

18.5 Long, Narrow Panels

- When pooling data from different time periods or across different cross-sectional units, you may believe that some of the data are "more reliable" than others. For example, you may believe that more recent data should be given a heavier weight in the estimation procedure. Bartels (1996) proposes a convenient way of doing this.
- Tests for the nature of the variance–covariance matrix have good intuitive content. Consider the LR test for equality of the error variances across the N firms, given by

$$\mathrm{LR} = T\left(N\ln\hat{\sigma}^2 - \sum_{i=1}^{N}\ln\hat{\sigma}_i^2\right)$$

where $\hat{\sigma}^2$ is the estimate of the assumed-common error variance, and $\hat{\sigma}_j^2$ is the estimate of the ith firm's error variance. If the null of equal variances is true, the $\hat{\sigma}_i^2$ values should all be approximately the same as $\hat{\sigma}^2$ and so this statistic should be small, distributed as a chi-square with $N-1$ degrees of freedom.

The corresponding LM test is given by

$$\mathrm{LM} = \frac{T}{2}\sum_{i=1}^{N}\left[\frac{\hat{\sigma}_i^2}{\hat{\sigma}^2} - 1\right]^2$$

If the null is true the $\hat{\sigma}_i^2/\hat{\sigma}^2$ ratios should all be approximately unity and this statistic should be small.

As another example, consider the LR test for the $N(N-1)/2$ unique off-diagonal elements of the contemporaneous variance–covariance matrix (Σ) equal to zero, given by

$$\mathrm{LR} = T\left(\sum_{i=1}^{N}\ln\hat{\sigma}_i^2 - \ln\left|\hat{\Sigma}\right|\right)$$

If the null is true, the determinant of Σ is just the product of its diagonal elements, so $\Sigma\ln\hat{\sigma}_i^2$ should be approximately equal to $\ln|\hat{\Sigma}|$ and this statistic should be small, distributed as a chi-square with $N(N-1)/2$ degrees of freedom.

The corresponding LM test is given by

$$\mathrm{LM} = T\sum_{i=2}^{N}\sum_{j-1}^{i-1} r_{ij}^2$$

where r_{ij}^2 is the square of the correlation coefficient between the contemporaneous errors for the ith and jth firms. The double summation just adds up all the different contemporaneous correlations. If the null is true, all these correlation coefficients should be approximately zero and so this statistic should be small.