Chapter 2

Exploiting panel datasets

2.1The uses of panel data

Much econometric modelling is based on 'one-dimensional' data sets: time-series or cross-sectional analysis. Both these methods have their problems. Time-series models excel in their treatment of dynamic effects, but suffer from the multicollinearity of series. Cross-sectional analysis makes use of a wide variety of functional forms, but is necessarily limited in its treatment of dynamic effects.

Pooling data on individuals over time into one dataset allows the econometrician to deal with a range of relationships between units of information within a single coherent structure. Panel data can be seen as cross-sections observed and linked over time, or multiple structural time-series. It is argued that by combining the best of both worlds better estimators result[1]. The use of all information available within the same model makes for inherently more efficient estimators, while the larger amount of data increases the degrees of freedom for hypothesis testing. This latter effect produces increased flexibility in model design by allowing more scope for the use of instrumental variables, simultaneous equation specifications, lagged variables and other techniques needing many degrees of freedom.

More importantly, a panel dataset enables the researcher to discriminate between competing hypotheses indistinguishable under simpler models. Consider estimating the success rates over time of a training program. The hypothesised relationship may be

$$y_{it} = f(x_{it}, \beta, \alpha_i) \quad \textbf{(2.1)}$$

[1] In the context of this discussion "better" merely refers to some arbitrary criterion such as mean-squared error or efficiency used to evaluate different models.

where α_i is some element specific to the individual i ("individual heterogeneity" or an "individual specific effect") which does not vary over time; for example, an element of "motivation". Initial results find that 40% of the class pass each exam, but do the same 40% pass every exam, or does everyone have a 40% chance of passing? In other words, is this unobserved effect significant in determining the outcome? If the term α_i was identified and found to be a significant factor in the probability of passing in any particular period, this would imply that someone passing one exam is more likely to pass or have passed other exams. The hypothesis that the same 40% pass each time appears more likely.

This effect could not have been identified by treating the data as purely cross-sectional (that is, with no connection between observations in different periods). Treating each period equation separately means that the individual-specific effect is not identified and must be subsumed into the constant term. Pooling all the data would appear to show serial correlation in the errors. However, a panel model could determine the relative importance of the unobserved heterogeneity; and distinguish it from the apparent "serial correlation" in the results.

Panel models provide the opportunity to test and control for a much wider range of measurement errors and unmeasurable effects then the simple example above. By using the panel to its full extent, both intercepts and slope coefficients which vary over time and/or individuals can be estimated from structural or reduced forms. This gives great scope for flexibility in the model without having to identify all the relevant variables: the ability to "group" observations by period or individual is all that is needed in many cases.

Hsiao (1986, pp5-7) provides some examples, reproduced in part in Figure 2.1, where the apparent cross-sectional relationship is belied by the panel estimates. The bold lines represent pooled estimates (that is, ignoring any panel structure), while the others represent the "true" structure which could be revealed by the appropriate panel estimators. In Figure 2.1(a), the panel estimates have common positive slope coefficients, as does the pooled estimate.

However, as each individual has a different intercept, the pooling estimate is clearly inefficient. In Figure 2.1(b) the effect of ignoring different intercepts means that the pooled estimate no longer even has the correct sign for the slope. Thus, although only intercepts vary over individuals, this suffices for the pooled results to give an entirely erroneous view of the relationship[2]. Moreover, there is no a priori indication of how the pooled slope is biased from true. Identical slope coefficients but different unmeasured effects have led to very different pooled estimates in (a) and (b).

In Figures 2.1(c) and (d) the slope coefficients also vary. Clearly a pooled regression on the individuals in Figure 2.1(c) would indicate little or no relationship between the variables on the x and y axes, while in Figure 2.1(d) the pooled regression appears to produce a nonlinear relationship. A properly specified panel model would be able to determine the true structure of the relationship.

The ability of panel techniques to combine information on individuals and time is their strongest asset. Unfortunately, this ability can also cause significant problems. The rationale for panel models is that interrelationships over time and between individuals are constant and so can be factored out. If the assumed interrelationship is wrong, then the error may affect all elements of the regression. A misspecified cross-section in period t (for example) should not affect estimation of the relationship in t+1 which uses different data; but if an individual specific effect is misspecified, it may corrupt the results from all periods. This is especially relevant in non-linear models; see section 2.4.

The most obvious, and important, source of misspecification is selection bias. Panels, by

[2] For example, some cross-section studies carried out by the author appeared to reveal a positive relationship between the proportion of manufacturing in GDP and energy consumption in developed countries. A simple panel study using the same data showed a negative relationship, a reverse of case (b) above. The implication was that the cross-section results were spurious, arising from significant national differences, and the original model was too crude to pick this up.

their very nature, are more susceptible than other data sets to missing observations, a problem which increases as the panel grows over time. For example, consider a two-period panel composed of welfare recipients. Those who remain in the panel for the second period may be less "employable", if those who have found jobs leave the panel. Whether this attrition is random or correlated with the dependent variables is crucial for the results of any estimation. At best it reduces efficiency; at worst, it can distort results significantly.

This issue is extremely complex for panel models, and currently unsolved in the general case. There is some current research on this issue (see Ridder (1990) and Ritchie (1994) for a theoretical treatment; Bell and Ritchie (1993b, 1994) for a study of selection bias in the NES); but even simple static models with multinormal spherical errors present formidable computational difficulties. The applied work for this thesis presents a practical but rather ad hoc approach to selection bias.

A second (and much less frequently discussed) source of error peculiar to panels is an overdependence on the ability to account for unmeasured variables. Consider the training example taken over two years, when the underlying "motivation" changes significantly and that the change is reflected across all individuals. Four regression models may be considered:

$$\begin{aligned}
&(a)\ y_{it} = \mu + x_{it}\beta + u_{it}\\
&(b)\ y_{it} = \mu_t + x_{it}\beta_t + u_{it}\\
&(c)\ y_{it} = \mu + x_{it}\beta + \alpha_i + u_{it}\\
&(d)\ y_{it} = \mu_t + x_{it}\beta_t + \alpha_i + u_{it}
\end{aligned} \qquad \textbf{(2.2)}$$

for t=1,2, i=1..N. (2.2a) and (2.2b) are cross-sectional models; (2.2c) and (2.2d) are panel models. However, in (2.2a) and (2.2c) the coefficients are assumed to be constant over time, and so only (2.2b) and (2.2d) can identify the structural shift between years one and two. Clearly (2.2a) is the most restricted model and (2.2d) the least, and the performance of the estimators will reflect this; but it is difficult to say whether the flexible cross-section (2.2b) or the poorly-specified panel model (2.2c) will perform better. Chapter nine returns to this in an applied context.

The problem becomes more important when variation in the slope coefficients is allowed. If performance in a training program is improving as recruitment, teaching and testing methods improve, it may be desirable to allow the slope coefficients to vary over time. Separate cross-sections, such as (2.2b), allow for this. So may a panel model, and one such as(2.2d) is at least as efficient as the cross-sections. However, the overwhelming majority of models used in applied work are of the form of (2.2c) with a time-varying intercept. If the slope coefficients vary significantly the cross-section may give better results. Relying on the panel attributes of parsimonious models instead of using more general specifications can lead to poor outcomes.

This question of misspecification is not limited to panels, and the particular problems caused by the use of panel models do not raise any significant new issues. A general discussion of the misspecification of panel models is beyond the scope of this thesis, and so is only considered in relation to the particular matter at hand. Unless explicitly stated otherwise, it is assumed that the models are correctly specified.

2.2Fixed and random effects

The unmeasured effects in a panel model structure may be "fixed" or "random" with respect to the model. This has practical and theoretical consequences, and leads to different models, estimation procedures, and possibilities for inference. Consider a simple linear panel model:

$$y_{it} = x_{it}\beta + \mu + u_{it} \qquad u_{it} = \alpha_i + \varepsilon_{it} \qquad \textbf{(2.3)}$$

In this model, an individual-specific term is assumed to capture all the omitted information. μ is the mean intercept for all participants; α_i is the individual variation around μ. This term is sometimes called an "incidental parameter", as the focus of interest is the value of β. The concern about the α_i term arises from its effect on the other variables; identifying its value is often not required.

The individual-specific terms may be deemed "fixed" (that is, parameters of the model). They then amount to a set of coefficients on individual-specific dummy variables which may be estimated by OLS (or any method appropriate to the assumed error structure of ε_{it}). All results are then conditional on these parameters. Note that OLS estimates of the dummy coefficients are inconsistent while T remains small (see section 2.3), but this does not affect the consistency of the estimates of β and μ.

However, if α_i is considered "random" (that is, a random component of the variance of the dependent variable), then autocorrelated residuals from OLS estimates will reflect the distribution of both ε_{it} and α_i. Note that the error term in (2.3) in vector form becomes,

$$u_i = J_T \alpha_i + \varepsilon_i$$
$$E(u_i) = 0 \quad E(u_i u_{i'}) = J_T J_{T'} \sigma_\alpha^2 + I_T \sigma_\varepsilon^2 \qquad \textbf{(2.4)}$$

where J_T is a T-vector of ones. OLS is unbiased and consistent, but inefficient unless σ_α^2 is zero (and assuming zero correlation between x_{it} and α_i). GLS solution methods may be used to identify the two distributions and estimate (2.3) efficiently.

Section 2.3 concentrates on the practical effect of different assumptions; for the moment, consider the theoretical implications. The choice is largely a subjective one, and most texts on panel data consider how this choice may be made. Hsiao(1992) suggests that the key theoretical issues come down to (a) what is the purpose of the study? and (b) what is the context of the data?

The argument is usually based around sample versus population study. If interest lies in the characteristics of participants in the sample, or if the participation list is exhaustive, then a fixed effects model may be most appropriate. If the aim is to determine population parameters from a sample, then the random-effects specification may be more useful.

For example, in a study of training programmes in different industries over time, a significant industry-specific effect may be the result of institutionally-based practices. The size of the effect may be useful information in itself, enabling predictions for each industry and facilitating inter-industry comparisons. In such a case it appears reasonable to treat these industry-specific factors as fixed, and allowing a different constant term for each industry captures that effect. Interest lies in using the heterogeneity as a predictive and explanatory tool for the behaviour of individual industries in the sample. If all industries of interest are included in the survey, then being unable to predict the size of this effect for other industries - a consequence of the fixed-effect assumption - is irrelevant.

Alternatively, consider the effect of training programmes on individuals over time. It is plausible to assume that each employee responds to the programme in a unique manner which persists over time. However, the interest is less in these individual differences but in the overall effect of the programme. Making general predictions for the programme requires the distribution of these individual effects over the workforce. Accordingly, the trainees in question are assumed to be random drawings from the population of workers with correspondingly random unobserved traits. Then the performance of a new participant on the program can be predicted with more confidence than by extrapolation from the specific (fixed) heterogeneity of current trainees.

As a third example, the applied work in this thesis centres around Mincer-type wage equations where the worker is the observation unit. While acknowledging that the NES remains a sample drawing, with hundreds of thousands of individuals appearing in the dataset there is some justification for approaching it as a population. The random-effect and fixed-effect specifications can both be justified on the population/sample argument. On the other hand, the unobserved characteristics of each person are of less interest than how these characteristics manifest themselves over the populace as a whole. The aim is to be able to make predictions about the population, not to identify any one particular individual's wage.

Thus a random-effects model may be deemed appropriate[3].

With heterogeneity over both individuals and time, a combination of fixed and random effects may be employed. Consider extending (2.2) to include an effect specific to period t, λ_t:

$$y_{it} = x_{it}\beta + \mu + u_{it} \qquad u_{it} = \alpha_i + \lambda_t + \varepsilon_{it} \tag{2.5}$$

A common specification in applied work is to have one effect fixed and one random; models with both effects fixed or random are less common. Often this is for practical reasons: introducing dummy variables for the "large" dimension may be cumbersome and inefficient, whilst random-effects estimation of the "small" dimension tends to be complex and inefficient. As panels tend to be "short and fat" (that is, with T relatively small and N large), a common solution is to estimate random individual effects and to introduce time dummies for the time effects.

To some extent this also reflects research interests. Much panel work is done on micro-data, with population inferences being drawn. There is little interest in the performance of individuals as opposed to the whole, and the number of parameters in a fixed-effects specification may be very large. A better solution is to look for any overall distribution of such individual effects.

For time effects the opposite holds: calculating a "distribution" of time intercepts is likely to be fairly meaningless, but by treating intertemporal differences as parameters of the model there is more scope for comparing directly different periods. Fixed effects are particularly useful when slope coefficients are allowed to vary: the evolution of coefficients through time may be very enlightening.

---

3 In fact, the estimation method to be discussed uses the fixed-effects method, for practical reasons outlined in the next section, although knowledge of the individual effects has little practical use.

For example, estimation on the full NES panel has little of interest to say about the "motivation" of some hundred thousand individuals; but an observable shift in the intercept over time is of interest. thus a random individual effect and a fixed time effect appears sensible.

All this assumes that the underlying structure of the model is known. If the true structure is unknown, then the feasibility, efficiency and consistency of the specifications must be considered.

The fixed-effects model is distribution free; it is conditioned on the extant values for α_i without the need to describe the source or distribution of this effect. It is robust to alternative specifications of the individual heterogeneity because the distribution function is irrelevant to the estimation method.

Estimation of the random effect requires further assumptions about the error terms. This is not necessarily a significant drawback. Although ML estimation of the specification in (2.3) requires a specific functional form for the panel effects, GLS estimation is feasible and practical with merely a consistent estimate of the covariance matrix, given by fixed-effects estimates.

Much more serious is the necessary assumption of independence of the explanatory variables and the random effect. Mundlak (1978) argued that the fixed effects model is conditional on the explanatory variables, and that the random-effects model is a misspecification that fails to take account of this conditioning. An appropriate model should use $E(\alpha_i \mid x_i)$. Replacing α_i in (2.3) by a linear approximation

$$E(\alpha_i \mid x_i) = \sum_t x_{i't} a_t + \omega_i \quad \omega \_ N(0, \sigma^2_\omega) \qquad \textbf{(2.6)}$$

where a_t is a vector of constants to be estimated, allows for correlation between the unmeasurable and explanatory variables. Mundlak then suggested restricting the model to a function of the mean value of the explanatory variables

$$E(\alpha_i \mid x_i) = \bar{x}_{i'} a + \omega_i \qquad \omega \_ N(0, \sigma^2_\omega) \tag{2.7}$$

and it can be shown (Hsiao (1986) pp 44-45) that the GLS estimator of β collapses to the fixed-effects estimator; the difference between the "true" and "false" GLS estimates of β is the GLS estimate of a. Therefore, there are not two models: the apparent difference is a specification error.

In a more general approach, Chamberlain (1984) notes that, if α_i is correlated with $(x_{i1}..x_{iT})$, y_{it} is potentially a function of all the explanatory variables and any estimator should take account of all lead and lag values of x_{it}. This gives a multivariate regression of all the ys (Tx1) on all the xs (TxKT) with an arbitrary error structure:

$$\begin{gathered} y_{it} = Z_{i'} \zeta_t + u_{it} \\ E(u_{it}) = 0 \qquad E(u_{it} u_{is}) = \sigma^2_{ts} \end{gathered} \tag{2.8}$$

where $Z_i = [x_{i1} \ldots x_{iT}]$. This is the basis for Chamberlain's "minimum-distance" panel estimator, described in more detail in section 2.6.2.

Mundlak's result depends on the very restrictive assumption of the source of the heterogeneity, while Chamberlain's is much more general, but both raise an important point: the random effects estimator is unbiased and consistent only if the explanatory variables and the heterogeneity are independent of each other. The fixed-effects model, giving parameter estimates conditioned on the explanatory variables, is unaffected. This idea of the fixed effects model as a conditioning estimator (as opposed to the marginal estimator of the random-effects model) acknowledges the fact that the former is valid for both fixed- and random-effects specifications[4].

[4] Note that, if the assumption of zero correlation between the explanatory variables and the individual heterogeneity is violated, then cross-section estimates, which ignore the heterogeneity completely, will also be biased and/or inconsistent.

The fixed-effects estimator is therefore flexible and robust; however, generally it is inefficient. This is because the fixed-effects approach seeks to isolate individuals or periods, and so restricts itself to smaller samples relative to the number of parameters to be identified; the random-effects model looks for common characteristics and makes holistic assessments of the data. For example, as the number of periods shrinks, making efficient use of information across individuals becomes increasingly important; the large number of parameters to be estimated in the fixed-effects model becomes a growing burden. As N large and T small is a common structure for panel datasets, then a consistent random-effects model is generally more efficient. Taylor (1980) estimated that, if the assumptions of the random-effects model hold, then this construction is more efficient for the cases (T>2, N-K>8) and (T>1, N-K>9).

2.3Estimation of static linear panel data models

This section is only intended as a brief introduction to some aspects of panel estimation methods, so the mathematics are kept to a minimum and areas covered are selective[5]. This section concentrates on a few basic models, with N individuals and T periods, as the qualitative aspects of these carry over in a straightforward manner to more complex specifications. In chapter 5, the fixed-effects specifications implemented in the analysis software will be discussed in more detail.

As focus of the concern here is in illuminating certain aspects of panel models (and not developing the econometric methodology), this section begins with the simplest models. The general linear model is

$$y_{it} = x_{it}\beta_{it} + u_{it} \qquad \textbf{(2.9)}$$

[5] Full discussion is provided in Hsiao (1986) or Matyas and Sevestre (1992). In the following discussion a balanced panel is assumed; that is, $T_i=T_j=T$ for all i, j. This does not change the results materially.

with i=1..N, t=1..T, and x_{it} being a (1xK) vector of explanatory variables. This allows any parameter to vary over time and individuals, giving (NTxK) parameters in NT equations; the model is unidentified without some restrictions on the parameters. The type of restriction imposed can lead to very different models; so that, if one is removing or adding variables before re-estimating, the choice of initial model may influence the path taken. For example, a pooled model on the data in Figure 2.1(d) may indicate that a quadratic form is needed, whereas an initial panel specification could indicate that this is unnecessary. Thus, while panel estimates may be fairly robust in many cases, this does not obviate the need for general tests on the specification of the model.

The simplest parameterisation is the pooled model: in system form,

$$y = X\beta + u \qquad \textbf{(2.10)}$$

where y, X and u are stacked to give NTx1, NTxK, and NTx1 matrices. There is assumed to be no significant consistent variation in the coefficients. Separate time-series or cross-section estimates give the same result as the pooled model, but the greater combined number of observations in the panel lead to smaller standard errors. Estimates are consistent whether N, T or both tend to infinity. The estimator is efficient under the assumption that the residual errors have no time- or individual-specific element. If the disturbance is non-spherical for other forms of heteroscedasticity or autocorrelation, any of the usual transformation or estimation methods for the particular form of the error terms is appropriate.

2.3.1The covariance approach

A first extension to (2.10) is to let an intercept vary across individuals[6], as in (2.3) above. For true panel specifications, a number of solution methods become appropriate. This section demonstrates a common approach, using techniques from variance analysis.

6 Time-specific effects are dealt with in a qualitatively identical manner. The relevant equations are found by swapping T and N and the t-i subscripts.

If the individual variation α_i is treated as a fixed effect, it becomes a parameter to be estimated. A simple solution is to employ N dummy variables, such that

$$y = X\beta + I_N \otimes J_T A + u \qquad \textbf{(2.11)}$$

where A = $[\alpha_1\ \alpha_2\ ..\ \alpha_N]'$, an Nx1 vector of the individual fixed effects, I_N is the N-element identity matrix, and J_T is a T-vector of ones. Equation (2.11) is known as the least-squares dummy-variable (LSDV) estimator. The N+K parameters in this equation can be estimated by OLS[7]. This solution has a significant drawback. Consider the normal equations:

$$\begin{bmatrix} \beta \\ A \end{bmatrix} = \begin{pmatrix} X'X & X' I_N \otimes J_T \\ I_N \otimes J_{T'} X & I_N \otimes J_{T'} J_T \end{pmatrix}^{-1} \begin{pmatrix} X'y \\ I_N \otimes J_{T'} y \end{pmatrix} \qquad \textbf{(2.12)}$$

This requires the potentially formidable task of inverting an (N+K)x(N+K) matrix. A more practical alternative is to take deviations from individual means. For an individual equation, let

$$\tilde{y}_{it} \equiv y_{it} - \overline{y}_i = (x_{it} - \overline{x}_i)\beta + u_{it} - \overline{u}_i \equiv \tilde{x}_{it}\beta + \tilde{u}_{it} \qquad \textbf{(2.13)}$$

The individual effect, constant over time, drops out of the regression. The system equations become

$$\tilde{y} = \tilde{X}\beta + \tilde{u} \qquad \textbf{(2.14)}$$

and the normal equations are now

$$\hat{\beta}_w = \left(\tilde{X}'\tilde{X}\right)^{-1}\left(\tilde{X}'\tilde{y}\right) \qquad \textbf{(2.15)}$$

which only involves a KxK inversion. The panel effects can be found from the individual means:

$$\hat{\alpha}_i = \overline{y}_i - \overline{x}_i \hat{\beta}_w \qquad \textbf{(2.16)}$$

(2.15) is the "within" or "covariance" estimator (from the analysis-of-covariance technique

[7] Note that estimates of the individual-specific effect only become consistent for large T not large N. The estimates of the slope coefficients are consistent for N and/or T large.

used). The estimated slopes in (2.15) are BLUE and consistent if N and/or T is large; however, the estimates of the individual intercepts, while unbiased, are not consistent unless T is large. Expanding (2.16):

$$\begin{aligned}\hat{\alpha}_i &= (\bar{x}_i \beta + \bar{u}_i + \alpha_i) - \bar{x}_i \hat{\beta}_w \\ &= \bar{x}_i(\beta - \hat{\beta}_w) + \alpha_i + \frac{1}{T}\sum_t u_{it}\end{aligned} \tag{2.17}$$

As $E(u_{it}) = 0$, the estimate of α_i is unbiased, but remains inconsistent unless $T \rightarrow \infty$ as

$$\underset{N\rightarrow\infty}{\hat{\alpha}_i} \rightarrow \alpha_i + \frac{1}{T}\sum_t u_{it} \tag{2.18}$$

This is because the number of parameters to estimate increases with N and there remains insufficient variation amongst individuals to uncover heterogeneity whilst T is small.

Now consider random effects, distributed over individuals according to some function. The random individual effects and the residual errors have to be estimated together. Define

$$v_{it} \equiv \alpha_i + u_{it} \quad \alpha_i \_ N(0,\sigma_\alpha^2) \quad u_{it} \_ N(0,\sigma_u^2) \tag{2.19}$$

Other covariances are zero. The structure of the covariance matrix for an individual is

$$E(v_i' v_i) = J_T J_{T'} \sigma_\alpha^2 + I_T \sigma_u^2 \equiv \Omega_i \tag{2.20}$$

This is called the "variance components" or "error components" model, for obvious reasons. OLS is inefficient, as the error terms v_{it},v_{is} are serially correlated. However, GLS is both feasible and practical. Under the assumptions of (2.19) the covariance matrix for the regression is block-diagonal, as is its inverse:

$$\begin{aligned}\Omega &= diag(\Omega_1, \Omega_2, \ldots \Omega_N) \\ \Omega^{-1} &= diag(\Omega_1^{-1}, \Omega_2^{-1}, \ldots \Omega_N^{-1})\end{aligned} \tag{2.21}$$

with

$$\Omega_i^{-1} = \frac{1}{\sigma_u^2}\left[\left(I_T - \frac{1}{T}J_T J_{T'}\right) + \psi\frac{1}{T}J_T J_{T'}\right] \quad \psi \equiv \frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2} \tag{2.22}$$

The first part in the inverted variance term calculates deviations from the mean of any matrix to which it is applied; it is the same matrix used to generate the within estimator, β_w. The second part calculates the mean of a matrix and multiplies it by a constant which reflects the relative variances of the two error components. The normal equations for the GLS solution are

$$\begin{bmatrix}\hat{\beta}\\ \hat{\mu}\end{bmatrix}_{gls} = \left(\sum_i X_{i'}\Omega_i^{-1}X_i\right)^{-1}\left(\sum_i X_{i'}\Omega_i^{-1}y_i\right) \tag{2.23}$$

The covariance matrix, and thus the solution for β, contains two additive terms. The structure in (2.22) indicates one term should reflect "within-group" variation, and the other the differences between group means - "between-group" variation. It can be shown (see, for example, Hsiao (1986) pp34-41) that the GLS solution breaks down into

$$\hat{\beta}_{gls} = \Delta\hat{\beta}_b + (I_K - \Delta)\hat{\beta}_w \tag{2.24}$$

where the "between" estimator

$$\hat{\beta}_b = \left(\sum_i(\overline{X}_i - \overline{X})'(\overline{X}_i - \overline{X})\right)^{-1}\left(\sum_i(\overline{X}_i - \overline{X})'(\overline{y}_i - \overline{y})\right) \tag{2.25}$$

is calculated by taking the deviations of inter-group means from the whole mean of the regression. The GLS estimate is a weighted average of the within and between estimators, with the weights given by

$$\Delta = \left(\sum_i\sum_t(X_{it} - \overline{X}_i)'(X_{it} - \overline{X}_i) + \psi T\sum_i(\overline{X}_i - \overline{X})'(\overline{X}_i - \overline{X})\right)^{-1}\left(\sum_i(\overline{X}_i - \overline{X})'(\overline{X}_i - \overline{X})\right)$$

(2.26)

The between estimator, showing the inter-individual variation, is effectively the OLS estimate on the data, ignoring all panel aspects. The weighting is provided by the relative importance of the error components and the size of T. If T is one or if σ_u^2 is very large relative to σ_α^2, then the weights are allocated evenly between the two estimators, and the pooled OLS estimate results. In this case either variation among the observations for an individual is small or the number of observations for an individual is small; random variation between individuals is the dominant force. Alternatively, if σ_α^2 or T is large, the within estimator dominates. If σ_α^2 is large, then individuals are sufficiently different to make separate estimation on each individual a sensible strategy. If T is large, between-group variation becomes irrelevant: there are enough observations for each individual to be treated as a separate model. Each α_i can be thought of as drawn once (randomly) and then fixed for N sample drawings which are large enough to be estimated separately.

Feasible estimation of (2.23) requires known or consistently estimated error components. Consistent estimates are given by the residuals from separate covariance and pooled OLS regressions. The error components model can also be estimated by ML. When T is small and N large, this method is wholly consistent; however, if N is small and T large, the estimate of σ_α is inconsistent. When T is large, the model becomes a series of N separate regressions; the ML estimate collapses to the covariance estimator (see Hsiao (1986) for details).

The covariance estimator can be used on the random effects model; obviously, the transformation removes the individual effect whatever its nature. For large T the two estimators coincide, but by ignoring the information on within-group variance rather than using it, the covariance estimator of a random-effects model is less efficient than an error components model when T is small. However, the random-effects model requires zero correlation between the error-component and the explanatory variables for consistency and unbiasedness. This is not an issue for the fixed-effects model as the estimates are conditioned on the parameters. Thus covariance estimation of a random-effects model may be consistent when GLS estimation is not.

(2.11) is the simplest of panel models and can be expanded in many ways. Consider adding time- or individual-specific variables; that is, the coefficients are common to all, but the variables are not:

$$y_{it} = x_{it}\beta + z_i\delta + \mu + \alpha_i + u_{it} \quad \textbf{(2.27)}$$

z_i is a vector of individual data which does not change over time. In this model, a fixed α_i cannot be separated from the coefficient z_i if the covariance transformation is used. Random effects can still be identified. The minimum-distance estimator of Chamberlain (1984) uses invariant individual-specific vectors for its initial estimates of the parameters.

Another simple extension is to have both time- and individual-specific effects:

$$y_{it} = x_{it}\beta + \mu + \alpha_i + \lambda_t + u_{it} \quad \textbf{(2.28)}$$

This presents no new qualitative aspects. If both effects are fixed, then some restrictions are needed to prevent exact collinearity between the dummies. If both effects are random, then a variance component for time has to be calculated. This requires a third, "between-periods" estimator, which is the equivalent of the between-groups estimator used to find the individual variance.

However, the quantitative effects of incorporating more fixed or random variables are quite different, especially if the panel is unbalanced (that is, individuals have different numbers of observations). Adding time dummies is a straightforward matter and the balance of the panel has little practical effect. However, calculating a new error-component is complex, and if the panel is unbalanced the problem is significantly harder. Generalising (2.22) for unbalanced panels merely requires the substitution of T_i for T, but the three-component equivalent of (2.22) has six additive terms; with unbalanced panels the inverse is exceedingly complex[8].

In addition to the theoretical considerations introduced in section 2.2, two additional elements

[8] See Wansbeek and Kapteyn (1989) for the solution to the unbalanced three-component inverse.

have been introduced. First, the covariance estimator is consistent (if not necessarily efficient) under a wider range of assumptions. Secondly, the covariance estimator is easier to estimate than the random effects one. The practical differences increase if the panel is unbalanced. If the efficiency loss is small, or if collinearity between the individual heterogeneity and the explanatory variables is suspected, then the error-components model loses its appeal. The implication for the choice of fixed or random effects estimators is clear: the fixed-effects specification, although inefficient, is both tractable and robust.

Linear panel models more complex than those discussed above add little to the methodological issues raised, and are not considered here. However, three general features of linear models can be identified. Firstly, the dichotomy between fixed and random effects remains strong as models become more complex. Although some solution methods, such as analysis of covariance, may be applicable to both, the results for one or other assumption are general sub-optimal in some way. This is because the choice of random or fixed effects leads to qualitatively different assessments of intergroup or interperiod influences; applying one model to both hypotheses implies either irrelevant or relevant but unused information.

Secondly, the general structure of the solutions outlined persists. The fixed-effects specification may be seen as a question of effective use of dummy variables or covariance transformations. The random-effects model has a complex error structure requiring GLS or ML estimation. Choice of fixed or random effect determines the whole approach to estimation.

Thirdly, the more complex the models, the bigger the practical advantage of the fixed-effects approach. The difficulties of estimating fixed-effects models increase arithmetically with the complexity of the models; for random -effects, the relationship is more closely exponential. While the fixed-effects model does involve lots of 'incidental' parameters, generally the model can be transformed to remove these effects[9]. Although the number of parameters in the

---

[9] Often a combination of dummy variables and transformations is used. In the models to be described in

random-effects model rises much more slowly, isolating the distribution of these effects becomes increasingly problematic.

2.3.2The differencing approach

Where there is only one random effect, differencing of the equations is a practical alternative to the covariance approach. In both (2.11) and (2.19), differencing removes the individual-specific effect:

$$\tilde{y}_{it} \equiv y_{it} - y_{it-1} = (x_{it} - x_{it-1})\beta + u_{it} - u_{it-1} \equiv \tilde{x}_{it}\beta + \tilde{u}_{it} \qquad \textbf{(2.29)}$$

with $E(u_{it})=0$. Estimation can then proceed as usual. When T=2, the differencing approach and the covariance estimator coincide.

Differencing has significant advantages over the covariance estimator in dynamic linear models, to be detailed below. It also shares the distributional advantage of the covariance estimator: namely, that no distribution for the individual effect needs to be specified and that any correlation between the individual-specific effect and the other explanatory variables will not lead to inconsistent or biased estimators. However, for static models it is less appealing.

Firstly, the differencing estimator is less efficient than the covariance estimator - compare (5.90) and (5.124) for the expected variance of the regression. This is because, although both models involve N restrictions on the model, y_{i1} and y_{iT} each only contribute once to the information set of the differencing estimator, whereas all observations contribute equally to the covariance estimator.

Secondly, differencing is clearly not appropriate for a random-effects specification. The implication of the differencing approach is that the individual effects are fixed nuisance

---

chapter 5, individual effects are transformed out while dummies are used for time effects.

parameters to be removed before estimation of the main coefficients proceeds. Thus, for a random-effects specification, differencing will be less efficient than GLS estimation for the same reason that the fixed-effect covariance estimator is inefficient.

Thirdly, the estimates of the differencing approach are less amenable to interpretation if the coefficients are allowed to vary over time. Consider

$$y_{it} = x_{it}\beta_t + \alpha_i + \lambda_t + u_{it} \qquad \textbf{(2.30)}$$

Estimation of this equation by differencing leads to the model

$$\tilde{y}_{it} = \tilde{x}_{it}\beta_t + x_{it-1}\tilde{\beta}_t + \tilde{\lambda}_t + \tilde{u}_{it} \qquad \textbf{(2.31)}$$

where $\tilde{y}_{it} \equiv y_{it} - y_{it-1}$ and the other variables are similarly defined (see equation (5.129) for details). For the slope coefficients both levels and differences in the coefficients are being estimated, and with a little manipulation all the slope coefficients are identifiable. However, only the change in the time intercept is being estimated, and so only the relative values of the constant (relative to either λ_0 or λ_T) can be identified from the model. The levels of the intercepts can of course be recovered from the equations by calculating the predicted means of the regression for one period but the fact that the returned coefficients are changes rather than levels seems to be ignored by most authors[10]. Section 5.4 discusses this issue and a restricted approach which falls between the models of (2.29) and (2.30).

2.4 Non-linear models

Allowing for individual heterogeneity has serious disadvantages when a non-linear functional form is specified. Because of the additive nature of the linear equation, estimation of the main coefficient vector and the panel effects could be separated, for example by covariance or GLS estimation. Under a non-linear specification, this may no longer be possible.

[10] For example, the GAUSS program DPD returns "intercepts"; the fact that only T-1 are returned indicates that these are actually changes in the intercept.

Take a more general form of (2.5):

$$y_{it} = F(x_{it}, \beta, \mu, \alpha_i, \lambda_t) \qquad \textbf{(2.32)}$$

where F(...) is some function non-linear in the parameters. As mentioned earlier, if the panel terms to be fixed coefficients, these will only be estimated consistently if the right "dimension" of the panel is large; that is, a consistent estimate of a time-specific effect requires large N, and a consistent estimate of an individual-specific effect needs large T.

Assuming a typical panel structure with T small and N large, consistent estimates of λ_t are possible but not α_i[11]. In a linear regression with additive terms this does not affect the consistency of the main parameter estimates. However, (2.32) involves maximising the joint probability of all the parameters, and so inconsistent estimates of the incidental parameters will lead to the main parameters being inconsistently estimated (see Hsiao (1986) pp159-161 for an example).

This problem was considered in some detail by Neyman and Scott (1948), who suggested finding alternative functions for the main parameters which are independent of the incidental parameters[12]. Hsiao (1986) illustrates some simple cases, but notes that in general such functions are difficult to find. Perhaps most importantly, there does not seem to be one for the probit model; that is, there is no consistent estimator for small T for a fixed individual-effect probit model.

An alternative is to use a random-effects model, leading to a multivariate non-linear regression model with the likelihood function of the main parameters augmented by the marginal

[11] This is for the same reason as in the linear case; namely, that there are insufficient observations on each individual, but there are a large number of observations for each period.

[12] An alternative under investigation by the author is to use linear approximations to the non-linear functions; for example, replacing a probit or logit form with a linear probability model.

distribution of the random effects. The random-effects specification does at least provide consistent estimators, and simplifying assumptions can be made to reduce some of the complexity of the models (see Hsiao (1986) pp164-167). It has the obvious disadvantage of needing a specific distribution for the random effect. There is also the potential for correlation between the random effects and the explanatory variables, which is less easy to resolve than in the linear case: using the linear specifications of Mundlak (1978) or Chamberlain (1984) may impose unwarranted restrictions on a non-linear model.

Thus, fixed-effects models are relatively simple but may not provide consistent estimates of any parameters. In contrast, random-effects estimates are consistent, but only as long as assumptions on the distribution and independence of the effects is justified.

2.5Dynamic panel estimators

Dynamic panel models can be extremely informative. Apparent dynamic effects could result from heterogeneity, serial correlation, or (in the case of non-linear models) state dependence. Consider finding that a candidate who has passed one exam is more likely to pass another. This could be because candidates differ in their abilities (heterogeneity); because those who pass have acquired (randomly) some extra knowledge which is useful in both exams (serial correlation); or because those who pass avoid resits and so have more time to study for remaining exams ("true" state dependence). If a dynamic model can be constructed, then testing for combinations of these various forms of true and spurious state dependence is a possibility.

2.5.1Linear models

Dynamic models for panels are more complex than static ones. The reason is the initial condition of the dependent variable. Let

$$y_{it} = x_{it}\beta + y_{it-1}\gamma + \mu + u_{it} \qquad u_{it} = \alpha_i + \lambda_t + \varepsilon_{it} \qquad \textbf{(2.33)}$$

y_{it} and y_{it-1} depend on α_i, which implies that the initial condition y_{i0} does too - and, potentially, so does the entire pre-sample history. The problem is to remove or control for the individual-specific effect without introducing correlation between the dependent variable and the error term. For example, taking first-differences of (2.33) leads to

$$(y_{it} - y_{it-1}) = (x_{it} - x_{it-1})\beta + (y_{it-1} - y_{it-2})\gamma + (\lambda_t - \lambda_{t-1}) + (\varepsilon_{it} - \varepsilon_{it-1}) \qquad \textbf{(2.34)}$$

where the individual effect has been removed, but the error term is now correlated with the explanatory variables as $E(y_{it-1}\varepsilon_{it-1}) \neq 0$.

Consistent ML and GLS estimates may be available for a range of assumptions about the initial conditions. However, a relatively simple and popular solution to this problem is to difference the model, as in (2.34), and then use instrumental variables estimation. The panel structure itself provides instruments in the form of lagged or lagged-differenced dependent variables, an instrument set which grows over time as more lagged variables become available (Arellano and Bond (1991)). Unbiased and consistent estimators therefore exist for both the fixed and random effects.

The dynamic specification has received some criticism. The use of lagged dependent variables as instruments has been criticised on the grounds of a poor correlation over the long lags necessary to instrument differenced models. More fundamentally, it has been suggested that variable-coefficient models which take account of cross-period correlation a more general way, such as (2.8) or (2.30), offer both flexibility and consistency without the need for instrumentation (Chamberlain(1984)).

This thesis does not intend to discuss the relative merits of these two approaches. Both IV estimates of (2.34) and more general variable-coefficient models are available under the software to be described later. However, most of the applied work on the NES by the author and others has utilised cross-sectional analysis or the variable-coefficients approach.

2.5.2 Non-linear models

Estimation of the linear dynamic model is relatively straightforward. This is not the case for non-linear specifications. Consider the dynamic non-linear form:

$$y_{it} = F(x_{it}, y_{it-1}, \beta, \mu, \alpha_i, \lambda_t) \qquad \textbf{(2.35)}$$

where F(...) is once again some arbitrary non-linear function. In this case the panel model runs into severe difficulties. The problem, as for the linear dynamic model, is the determination of initial conditions, but it has been complicated by the non-separability of the terms in the model as discussed in section 2.4. If the parameter estimates are jointly calculated, then estimates of the initial conditions also need to be jointly calculated. Unlike the models in section 2.5.1, the incidental parameters can no longer be divorced from the ones of interest - and therefore estimation of the initial conditions (and possibly pre-sample history) must be included in the maximisation procedure.

On the assumption that the panel effects are fixed, the earlier conclusion that the ML estimates are inconsistent as long as T is small still holds. Moreover, Monte Carlo tests by Heckman (1981b) suggest that this inconsistency seems much more significant than for the static case discussed above.

The random-effects assumption needs information in y_{i0}, just as for the linear case. Unless y_{i0} is assumed to be independent of α_i, the marginal distribution of α_i needs to be integrated over y_{i0} as well; therefore information on y_{i0} is needed. And if y_{i0} is to be calculated, this may require information on y_{i-1}, y_{i-2}, etcetera if they too depend on α_i. Solutions to this issue have been suggested, but they are all unsatisfactory for one reason or another[13]. At the time of

[13] Suggestions including assuming that the initial conditions are truly exogenous and independent of α_i, or assuming that the pre-sample process is in equilibrium. Both these assumptions are very strong and hard to

writing there appears to be no general consistent estimator for dynamic non-linear models.

2.6General estimation techniques

In recent years there has been some interest in more generalised estimation methods. Two of these techniques are briefly considered here.

2.6.1Generalised method of moments (GMM)

GMM developed in the early eighties as a unifying approach to a wide range of problems[14]. Its name comes from estimators derived by minimising a set of moments conditions in a quadratic form. As such it shares characteristics with 'traditional' methods such as OLS, but the minimisation criterion is specified in such a way that wide range of problems may be treated within the same overall framework. Thus OLS, linear and non-linear IV, GLS, etcetera can be seen as restricted versions of the same basic estimator. A properly-specified GMM estimator is consistent and efficient.

GMM has found a strong foothold in the estimation of time series models. In panel models its main application has been in the area of linear dynamic models, particularly in the methodology of Arellano and Bond (1991). Their estimation program, DPD, uses the GMM nomenclature in allowing for a range of dynamic specifications and has had a notable impact in some areas of econometrics.

Because GMM is largely a unifying terminology until the appropriate functional forms are specified, GMM as an estimation method per se will not be pursued. The techniques in this

---

justify in applied work. Heckman (1981b) drops the pretence of consistency altogether and instead offers relatively practical approximating solution.

[14] Hall (1993) and Ogaki (1993) provide comprehensive surveys of GMM techniques and applications.

thesis can be seen as GMM estimators subject to known restrictions (for example, the linear IV model and Sargan's test are expounded using the GMM terminology), but this is a methodological issue. All of the methods used have been chosen for their practicality without reference to an overall framework.

2.6.2 Minimum-distance estimation (MDE)

Chamberlain's (1984) minimum-distance estimator is a general estimator in that it allows for a variety of specifications. While GMM and MDE both emphasise flexibility in the functional forms, MDE aims to provide robust and efficient estimation of unknown specifications rather than an efficient technique for known problems.

The starting point is the recognition that a random individual-specific effect could be correlated with any or all of the explanatory variables, and therefore a proper specification for a random-effects model should be

$$y_{it} = x_{it}\beta + Z_{i'}\zeta + u_{it} \qquad \textbf{(2.36)}$$

where Z_i is a 1xKT vector of all the x variables, as in (2.8). Moreover, the variance of u_{it} should be generously specified given the potential for heteroscedasticity and (particularly) autocorrelation in a panel dataset. Given these requirements, Chamberlain recommends estimating T separate equations of the form

$$y_{it} = Z_{i'}\pi_t + u_{it} \qquad \textbf{(2.37)}$$

Define $\Pi' = [\pi_1' \ldots \pi_T']$. Then, for a general specification of $Var(u_{it})$, the separate estimates of π_t are asymptotically normally distributed with mean Π and variance given by the covariance matrix of these separate estimates. This information is then used in a second stage regression to impose restrictions on the structure of Π in a manner analogous to that of simultaneous equation systems or GMM methods.

The slope coefficients in (2.36) may be allowed to vary over time, which makes the second-round estimates more complex but otherwise has little effect on the technique. This approach can also be applied to the non-linear counterpart of (2.36), with appropriate adjustments. However, the linear specification of the panel effect does imply a restriction on non-linear models, which may or may not be warranted.

MDE is very flexible as it places no restrictions on the error terms; it therefore has relatively low information requirements and is robust to the specification of u_{it}. However, the price paid is that MDE is only efficient within a class of estimators imposing no restrictions on the error term. Any prior knowledge about the true distribution of the error term implies a more efficient estimator exists.

A second difficulty with MDE is the large number of parameters to be investigated as T grows. This is a particular problem if N is relatively small or K large. Clearly, MDE is also unlikely to be attractive for random time-effects which would involve a matrix inversion of order NK. However, the variable-coefficient fixed-effects estimator to be described in Chapter 4 shares many features with MDE (including the data requirement), and it seems likely that MDE is a practical option[15].

2.7 Summary

As discussed in section 2.1, the panel structure allows for a much richer range of models. However, it is apparent that the usefulness of the more complex panel models is restricted by the feasibility of the estimator. This is especially true of non-linear models. The rest of this thesis is concerned with static, linear models, as estimators for these have been implemented in the analysis program to be outlined in chapters 5 and 6. The rationale for this decision is

[15] The possibility of implementing this estimator in the NES analysis software is currently being investigated.

examined in chapter 4.