ECON 5360 Class Notes
Panel Data

1 Introduction

Panel (a.k.a., longitudinal or time series-cross sectional) data are observed both across sections and over time. The advantages of panel data are that it

1. increases the number of observations,
2. increases the precision of parameter estimates and
3. allows one to sort out effects that may be impossible with only cross sectional or only time series data (e.g., technological progress versus economies of scale).

Three famous U.S. panel data sets are the

- Panel Study of Income Dynamics (PSID),
- National Longitudinal Survey (NLS) and
- Current Population Survey (CPS).

For example, the PSID follows 6,000 families and 15,000 individuals since 1968, asking questions related to income, job changes, marital status, other socioeconomic and demographic characteristics, etc.

Although there are clear advantages of panel data, there are also some complications. Below, I present an introduction to estimation with panel data. Begin with the following model

$$y_{it} = \alpha_i + x_{it}'\beta + \epsilon_{it} \tag{1}$$

where $i = 1, ..., n$ and $t = 1, ..., T$. Of course, when $\alpha_i = \alpha$ for $i = 1, ..., n$, then the data can simply be pooled together, the model written in standard form $Y = X\beta + \epsilon$, and estimated with standard linear techniques.

2 Fixed-Effects (FE) Model

In the fixed-effects model, we treat α_i as a group-specific constant term to be estimated with the other parameters. Now write the model as

$$Y = D\alpha + X\beta + \epsilon \tag{2}$$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{nT \times 1} \qquad D = \begin{bmatrix} i_T & 0 & \cdots & 0 \\ 0 & i_T & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & i_T \end{bmatrix}_{nT \times n} \qquad X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & & x_{2k} \\ \vdots & & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}_{nT \times k}$$

and

$$\alpha = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix}_{n \times 1} \qquad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{k \times 1} .$$

This commonly referred to as the Least Squares Dummy Variable (LSDV) model. Theoretically, there are no problems in estimating (2) – assuming the standard assumptions hold, the Gauss-Markov theorem applies and we obtain unbiased and efficient estimates. There may be computational problems, however. Notice that the stacked coefficient vector is of length $(n+k)$. Therefore, the standard least squares formula requires inverting a matrix of size $(n+k) \times (n+k)$. Since many panel data sets have $n > 1000$, this can lead to numerical errors.

2.1 Partitioned Regression

A partitioned regression provides a simple solution to the above problem. Recall that

$$b = (X' M_D X)^{-1} (X' M_D Y)$$

where

$$M_D = I - D(D'D)^{-1}D' = \begin{bmatrix} I_T - \frac{1}{T} ii' & 0 & \cdots & 0 \\ 0 & I_T - \frac{1}{T} ii' & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & I_T - \frac{1}{T} ii' \end{bmatrix}$$

is a symmetric, idempotent "residual-maker" matrix for the regression on the dummy variables D. Since M_D is symmetric, idempotent and premultiplication produces the average over $t = 1, ..., T$ for each i, the

partitioned regression is equivalent to regressing

$$Y_* = M_D Y = Y - \bar{Y} = \begin{bmatrix} y_1 - \bar{y}_1 \\ y_2 - \bar{y}_2 \\ \vdots \\ y_n - \bar{y}_n \end{bmatrix}_{nT \times 1}$$

on

$$X_* = M_D X = X - \bar{X} = \begin{bmatrix} x_{11} - \bar{x}_{11} & x_{12} - \bar{x}_{12} & \cdots & x_{1k} - \bar{x}_{1k} \\ x_{21} - \bar{x}_{21} & x_{22} - \bar{x}_{22} & & x_{2k} - \bar{x}_{2k} \\ \vdots & & \ddots & \vdots \\ x_{n1} - \bar{x}_{n1} & x_{n2} - \bar{x}_{n2} & \cdots & x_{nk} - \bar{x}_{nk} \end{bmatrix}_{nT \times k}.$$

This only requires inversion of a $(k \times k)$ matrix. The group-specific constant terms can then be recovered according to

$$a_i = \bar{y}_i - b' \bar{x}_i$$

for $i = 1, ..., n$. The partitioned regression approach is basically a two-stage estimation procedure:

- Step #1. Transform the data by subtracting group means.
- Step #2. Run OLS on the transformed data.

2.2 Variance Estimation

The variance estimator for b is as expected

$$\widehat{var}(b) = s^2 (X' M_D X)^{-1}$$

where the appropriate estimator for σ_ϵ^2 is

$$s^2 = \frac{e'e}{nT - n - k} = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} (y_{it} - a_i - x'_{it} b)^2}{nT - n - k}.$$

Keep in mind that if you use the partitioned-matrix approach, standard econometric programs may incorrectly use the degrees of freedom correction $nT - k$ when $nT - n - k$ is appropriate. Finally, the variance estimator for the a_i is

$$\widehat{var}(a_i) = \frac{s^2}{T} + \bar{x}'_i \{ s^2 (X' M_D X)^{-1} \} \bar{x}_i.$$

2.3 Testing for Group-Specific Effects

The standard F test can be used to test whether the pooled or fixed-effects model is more appropriate. The null hypothesis is

$$H_0\text{: } \alpha_1 = \alpha_2 = \cdots = \alpha_n.$$

The F statistic is

$$F = \frac{(R^2_{LSDV} - R^2_{pooled})/(n-1)}{(1 - R^2_{LSDV})/(nT - n - k)} \sim F(n-1, nT-n-k)$$

which could alternatively be written in sum-of-squared-errors form.

3 Random-Effects (RE) Model

An alternative approach is to treat α_i in equation (1) as a random draw from a distribution rather than being nonstochastic. The advantages are

- The RE model has fewer parameters to estimate.
- The RE model allows for additional explanatory variables that have equal value for all observations within a group (e.g., education level of parents, number of siblings, etc.).

The disadvantage of the RE approach is that

- if the unobserved group-specific effects are correlated with the explanatory variables, then the estimates will be biased and inconsistent.
- the estimator is a bit more complicated.

3.1 Basic Framework

From equation (1), we will let $\alpha_i = \alpha + \mu_i$ so that the model is

$$y_{it} = \alpha + x'_{it}\beta + (\mu_i + \epsilon_{it})$$

where the following assumptions are made

- $E(\epsilon_{it}) = 0$, $\forall i, t$
- $E(\mu_i) = 0$, $\forall i$
- $E(\epsilon^2_{it}) = \sigma^2_\epsilon$, $\forall i, t$
- $E(\mu^2_i) = \sigma^2_\mu$, $\forall i$

- $E(\epsilon_{it}\mu_j) = 0$, $\forall i, t, j$
- $E(\epsilon_{it}\epsilon_{js}) = 0$, $\forall s \neq t$ or $i \neq j$
- $E(\mu_i\mu_j) = 0$, $\forall i \neq j$.

The RE model can be written in matrix form as

$$Y = X\beta + \eta$$

where $\eta \sim N(0, \Omega)$. Given the assumptions above, the $(nT \times nT)$ variance-covariance matrix Ω has the following structure

$$\Omega = (I_n \otimes \Sigma_T) = \begin{bmatrix} \Sigma_T & 0 & \cdots & 0 \\ 0 & \Sigma_T & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \Sigma_T \end{bmatrix}$$

where

$$\Sigma_T = \begin{bmatrix} \sigma_\epsilon^2 + \sigma_\mu^2 & \sigma_\mu^2 & \cdots & \sigma_\mu^2 \\ \sigma_\mu^2 & \sigma_\epsilon^2 + \sigma_\mu^2 & \cdots & \sigma_\mu^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\mu^2 & \sigma_\mu^2 & \cdots & \sigma_\epsilon^2 + \sigma_\mu^2 \end{bmatrix}_{T \times T}.$$

3.2 Estimation

Although OLS will produce consistent estimates, because Ω is not diagonal, the OLS estimates will be inefficient. Therefore, GLS is the efficient estimator. Recall that the GLS estimator is

$$\hat{\beta}_{GLS} = (X_*'X_*)^{-1}(X_*'Y_*) = ((PX)'(PX))^{-1}((PX)'(PY)) = (X'\Omega^{-1}X)^{-1}(X'\Omega^{-1}Y). \tag{3}$$

For the RE model,

$$P = \Omega^{-1/2} = [I_n \otimes \Sigma_T^{-1/2}]$$

where

$$\Sigma_T^{-1/2} = \frac{1}{\sigma_\epsilon}[I_T - \frac{\theta}{T} i_T i_T']$$

and

$$\theta = 1 - \frac{\sigma_\epsilon}{\sqrt{\sigma_\epsilon^2 + T\sigma_\mu^2}}.$$

Therefore, the GLS estimator can be calculated by running a regression of the pseudo-deviations

$$Y_* = \begin{bmatrix} y_1 - \theta\bar{y}_1 \\ y_2 - \theta\bar{y}_2 \\ \vdots \\ y_n - \theta\bar{y}_n \end{bmatrix}$$

on the similarly transformed X_*.

3.3 Feasible Estimation

To make (3) operational, all that is left is to estimate σ_ϵ^2 and σ_μ^2. We will do this sequentially – first we estimate σ_ϵ^2 and then use that to estimate σ_μ^2.

3.3.1 Estimation of σ_ϵ^2.

We begin by using the "within-groups" information given by the difference between

$$y_{it} = \alpha + \beta' x_{it} + (\mu_i + \epsilon_{it}) \tag{4}$$

and

$$\bar{y}_i = \alpha + \beta' \bar{x}_i + (\mu_i + \bar{\epsilon}_i). \tag{5}$$

This produces

$$y_{it} - \bar{y}_i = \beta'(x_{it} - \bar{x}_i) + (\epsilon_{it} - \bar{\epsilon}_i). \tag{6}$$

Now that the unobserved group-specific random effects are gone, we estimate (6) using the LSDV estimator and use the residuals to get the following estimate of σ_ϵ^2:

$$\hat{\sigma}_\epsilon^2 = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(e_{it} - \bar{e}_i)^2}{nT - n - k}.$$

3.3.2 Estimation of σ_μ^2.

Now we use the "between-groups" information to estimate σ_μ^2. Consider (5) again

$$\bar{\epsilon}_i + \mu_i = \bar{y}_i - \alpha - \beta' \bar{x}_i.$$

The variance of (5) is

$$\sigma_{**}^2 = \frac{\sigma_\epsilon^2}{T} + \sigma_\mu^2.$$

Therefore, we can estimate σ_μ^2 using

$$\hat{\sigma}_\mu^2 = \frac{e'_{**}e_{**}}{n-k} - \frac{\hat{\sigma}_\epsilon^2}{T}.$$

Finally, insert the estimates $\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_\mu^2$ into P and calculate $\hat{\beta}_{GLS}$.

4 Choosing Between Fixed and Random Effects Models

A frequent question with panel data is which model to use – fixed or random effects. The answer boils down to whether the unobserved group-specific effects are correlated with the explanatory variables or not. If they are, then the RE model will produce inconsistent estimates. If they are not, then the RE model may be preferable. There are two methods for choosing between RE and FE models.

4.1 Think Through the Problem

Consider the following two problems where the RE model is almost certainly inappropriate.

1. Returns to Schooling. Labor economists use panel datasets to explain individual wages as a function of years of schooling, as well as other socioeconomic and demographic characteristics. Individuals almost certainly have unobserved innate abilities that are likely to be correlated with observable explanatory variables such as years of schooling, marital status, type of employment, etc.

2. Economic Growth and R&D Spending. Consider a regression of GDP per capita on a number of different country-specific variables such as research and development (R&D) spending, saving rates, population growth rates, schooling, and capital-labor ratios. There are likely to be unobserved, country-specific cultural differences that influence economic growth and, at the same time, are correlated with the explanatory variables such as saving rates, population growth rates, etc.

4.2 Hausman Specification Test

The motivation behind the Hausman test is that under the null hypothesis of no correlation (i.e., H_0: $corr(X_{it}, \mu_i) = 0$), then both the FE and RE estimators are consistent but only the RE estimator is efficient. Under the alternative, while the FE estimator is consistent, the RE estimator is not. The statistic is

$$W = (b_{LSDV} - \hat{\beta}_{GLS})'[var(b) - var(\hat{\beta}_{GLS})]^{-1}(b_{LSDV} - \hat{\beta}_{GLS}) \overset{asy}{\sim} \chi^2(k-1).$$

5 Heteroscedasticity and Autocorrelation

In general, there are two ways to handle nonspherical disturbances – robust estimation of the asymptotic variance-covariance matrix (e.g., White's estimator or the Newey-West estimator) or respecification of the

error structure and application of generalized least squares. I will present only the latter (see Greene section 13.7 for a discussion of robust estimation). Note that LIMDEP has canned routines for handling heteroscedasticity and autocorrelation in panel-data models.

5.1 Heteroscedasticity in the FE Model

The most straightforward way to handle heteroscedasticity in the FE model is to begin by calculating an estimate of $\sigma^2_{\epsilon,i}$ using the LSDV residuals

$$\hat{\sigma}^2_{\epsilon,i} = \frac{1}{T}\sum\nolimits_{t=1}^{T} e^2_{i,t}. \tag{7}$$

Feasible GLS estimates are then calculated by

$$\hat{\beta}_{FGLS} = (X'\hat{\Omega}^{-1}X)^{-1}(X'\hat{\Omega}^{-1}Y)$$

where

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}^2_{\epsilon,1}I_T & 0 & \cdots & 0 \\ 0 & \hat{\sigma}^2_{\epsilon,2}I_T & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\sigma}^2_{\epsilon,n}I_T \end{bmatrix}_{nT\times nT}.$$

5.2 Heteroscedasticity in the RE Model

Begin by considering the composite error term $\mu_i + \epsilon_{it}$. Although it makes sense to allow $E(\mu_i^2) = \sigma^2_{\mu,i}$, we will only have one observation for each i on μ_i. Therefore, estimation of $\sigma^2_{\mu,i}$ would have to be $\hat{\mu}_i^2$, which is probably not desirable. Therefore, if we let the $E(\epsilon^2_{it}) = \sigma^2_{\epsilon,i}$, then all we have to do is adjust our transformation parameter as follows:

$$\theta_i = 1 - \frac{\sigma_{\epsilon,i}}{\sqrt{\sigma^2_{\epsilon,i} + T\sigma^2_{\mu}}}.$$

A similar result holds for unbalanced panels where heteroscedasticity is introduced by varying group sizes. The only remaining task is to estimate $\sigma^2_{\epsilon,i}$ and σ^2_{μ}. Greene suggests using the LSDV residuals to estimate the $\sigma^2_{\epsilon,i}$ since the model has been purged of the μ_i. This estimate is given in equation (7). The group-specific variance σ^2_{μ} can then be estimated by

$$\hat{\sigma}^2_{\mu} = \frac{1}{n}\sum\nolimits_{i=1}^{n}(\hat{\sigma}^2_{OLS,i} - \hat{\sigma}^2_{\epsilon,i})$$

where $\hat{\sigma}^2_{OLS,i}$ are comparable to $\hat{\sigma}^2_{\epsilon,i}$, but estimated using the pooled OLS residuals. The RE estimator then proceeds as described earlier.

5.3 Autocorrelation in the FE Model

Autocorrelation in the FE is fairly simple (although care needs to be made that the data are stacked in the correct manner). Begin with AR(1) errors

$$\epsilon_{it} = \rho_i \epsilon_{i,t-1} + \nu_{it} \tag{8}$$

where some researchers choose to set $\rho_i = \rho$ for all $i = 1, ..., n$. Either way, the $\rho_i s$ can be estimated with the LSDV residuals, the data pseudo-differenced, and feasible GLS applied.

5.4 Autocorrelation in the RE Model

Autocorrelation in the RE model is only slightly more complex. Of course, in the RE model there is always going to be autocorrelation in the composite error term $\mu_i + \epsilon_{it}$ because μ_i does not vary over time. Therefore, it only makes sense to specify autocorrelation in ϵ_{it}, as is done in equation (8). The LSDV residuals can be used to get an estimate of ρ (or ρ_i if desired) and Cochrane-Orcutt can then be applied. The transformed model will take the form

$$y_{it} - \rho_i y_{i,t-1} = \alpha(1 - \rho_i) + \beta'(x_{it} - \rho_i x_{i,t-1}) + \mu_i(1 - \rho_i) + \nu_{it}$$

for $t = 2, ..., T$.

6 Other Types of Panel-Data Models

6.1 Unbalanced Panels

Up to this point, we have implicitly assumed that each cross section is observed for T periods. However, because of sample attrition and new entry, it is common to have not observe each cross section for all periods. In this case, the total number of observations will not be nT, but rather $\sum_{i=1}^{n} T_i$. Below, I describe how to modify the FE and RE models for an unbalanced panel, which is already programmed into most econometric software packages.

6.1.1 FE Model

The FE model works with an unbalanced panel with no real change, provided the dummy variables are updated to no longer have the same number of ones in each column.

6.1.2 RE Model

The RE model is only slightly more complicated. With an unbalanced panel, now define

$$\theta_i = 1 - \frac{\sigma_\epsilon}{\sqrt{\sigma_\epsilon^2 + T_i\sigma_\mu^2}}$$

where $i = 1, ..., n$. The data are then transformed according to

$$Y_* = \begin{bmatrix} y_1 - \theta_1\bar{y}_1 \\ y_2 - \theta_2\bar{y}_2 \\ \vdots \\ y_n - \theta_n\bar{y}_n \end{bmatrix}$$

and similarly for X. Of course, if $T_i = T$, then this collapses back to the standard RE estimator.

6.2 Time-Specific Effects

It is possible, using either the FE or RE approach, to incorporate time-specific effects:

$$y_{it} = \alpha_i + \beta' x_{it} + \gamma_t + \epsilon_{it}.$$

Using the FE approach, γ_t can be estimated by incorporating dummy variables for $T-1$ of the time periods. The RE approach, which treats γ_t as a random variable, is more complicated.

6.3 Dynamic Models

Sometimes it is desirable to allow for dynamic effects in the model

$$y_{it} = \alpha_i + \gamma y_{i,t-1} + x_{it}'\beta + \epsilon_{it}.$$

The problem with estimating such a model, either of the FE or RE nature, is that after transforming the model to get rid of the group-specific effect (e.g., first-differencing the model), the right-hand side variables are correlated with the error term. The solution to this problem involves using values of $y_{i,t-s}$ for $s > 1$ as instrumental variables.

7 Gauss Application

Consider the following panel data model taken from Woolridge (2002)

$$Murders_{it} = \alpha_i + \beta_1 Executions_{it} + \beta_2 Unemp_{it} + \epsilon_{it}$$

where $Murders_{it}$ is the number of murders in state i in year t per 10,000 people; $Executions_{it}$ is the total number of executions for the current and prior two years; $Unemp_{it}$ is the current unemployment rate; $i = 1, ..., 50$; and $t = 1987, 1990, 1993$. See Gauss example #7 for further details.