

Economics 508  
**Lecture 17**  
**Panel Data**

Consider a model of the form

$$y_{it} = x_{it}\beta + z_i\gamma + \alpha_i + u_{it} \quad i = 1, \dots, n \quad t = 1, \dots, T \quad (1)$$

where, for example,

- $y_{it}$  = log wage of person  $i$  at time  $t$ .
- $x_{it}$  = time varying characteristics at time  $t$  like age, experience, health, ...
- $z_i$  = time invariant characteristics at time  $t$  like education, race, sex, ...
- $\alpha_i$  = unobserved individual effect like spunk, ability
- $u_{it}$  = everything else.

we will stack the model so that all  $T$  observations on person 1 comes first, and then person 2, and so on.

Now consider the matrix,

$$P = I_n \otimes T^{-1} \mathbf{1}_T \mathbf{1}_T^\top \equiv I_n \otimes J_T$$

where the latter matrix is  $T^{-1}$  times a matrix of ones. It is easy to see that  $P$  represents an orthogonal projection, it is symmetric and idempotent. What does it do? Consider

$$Py = \begin{bmatrix} J_T & & & \\ & \ddots & & \\ & & \ddots & \\ & & & J_T \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} \bar{y}_1 \mathbf{1}_T \\ \vdots \\ \vdots \\ \bar{y}_n \mathbf{1}_T \end{bmatrix}$$

And therefore,

$$Qy \equiv (I - P)y = y - \bar{y}$$

is a deviation-from-individual-means vector. Note that if we wanted to view  $P$  as representing a least squares projection, we can think of it as arising from a model in which there are dummy variables for just the individual effects,

$$y_{it} = \alpha_i + u_{it}$$

We might write this as,

$$y = Z\alpha + u.$$

It is a useful exercise to show that  $P = Z(Z^\top Z)^{-1}Z^\top$  where  $\hat{y} = Py$  would be the least squares fit and  $\hat{u} = Qy$  would be the residual vector. Clearly applying  $Q$  to  $Z$  yields,

$$QZ = 0$$

since there is no temporal variability in  $Z$  by hypothesis. A common estimator of (1) for at least the  $\beta$  component is

$$\hat{\beta}_W = (X^\top QX)^{-1}X^\top Qy$$

which is frequently called the “within group” estimator. As long as we assume

$$Ex_{it}u_{it} = 0$$

$\hat{\beta}_W$  is consistent for  $\beta$ . But as the name suggests,  $\hat{\beta}_W$  uses only some of the information available. We also have the “between groups” information which is obtained by multiplying (1) by  $P$

$$\bar{y}_i = \bar{x}_i\beta + z_i\gamma + \alpha_i + \bar{u}_i$$

Note here that we can delete the  $n(T - 1)$  redundant observations. Let's denote OLS estimators of  $(\beta, \gamma)$  as  $(\hat{\beta}_B, \hat{\gamma}_B)$  for “between.” Can we combine  $\hat{\beta}_B, \hat{\beta}_W$  somehow?

#### *A Simple Measurement Error Problem (Revisited)*

1. Suppose that  $y_i \sim \mathcal{N}(\mu, \sigma_i^2)$  for  $i = 1, 2$ . The GLS estimator of  $\mu$  is:

$$\hat{\mu} = (X^\top \Omega^{-1} X)^{-1} X^\top \Omega^{-1} y$$

where

$$\Omega = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix} \quad X = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

so

$$\hat{\mu} = (1/\sigma_1^2 + 1/\sigma_2^2)^{-1} [y_1/\sigma_1^2 + y_2/\sigma_2^2]$$

2. Matrix Case – interpret as two independent estimates

$$y_i \sim \mathcal{N}_p(\mu, \Omega_i) \quad i = 1, 2$$

$$\hat{\mu} = (\Omega_1^{-1} + \Omega_2^{-1})^{-1}[\Omega_1^{-1}y_1 + \Omega_2^{-1}y_2]$$

here

$$\Omega = \begin{bmatrix} \Omega_1 & 0 \\ 0 & \Omega_2 \end{bmatrix} \quad X = \begin{bmatrix} I_p \\ I_p \end{bmatrix}$$

Note that if we put in values for  $\sigma_2$  or  $\Omega_2$ , and let them tend to infinity, then we get just first component.

To apply this to get the expression at the top of page 1381 of *HT* note that if

$$\hat{\delta}_i \sim \mathcal{N}(\delta, V_i) \quad i = W, B$$

our simple weighted least squares approach gives,

$$\hat{\delta} = (V_B^{-1} + V_W^{-1})^{-1}[V_B^{-1}\hat{\delta}_B + V_W^{-1}\delta_W]$$

*HT* rewrite this incorrectly so *be careful!* Note that

$$V_B^{-1} + V_W^{-1} = V_B^{-1}(V_B + V_W)V_W^{-1}$$

so

$$(V_B^{-1} + V_W^{-1})^{-1} = V_W(V_B + V_W)^{-1}V_B$$

so we may write,

$$\hat{\delta} = \Delta\hat{\delta}_B + (I - \Delta)\delta_W$$

where

$$\Delta = V_W(V_B + V_W)^{-1}$$

Note *HT* write  $\Delta = (V_B + V_W)^{-1}V_W$ ! Note also that  $V_B$  and  $V_W$  are not the same dimension so this adds to the complication. But this can be reconciled by the prior comment about letting covariance elements tend to infinity. Or better, by restricting the whole weighting exercise to the elements of  $\delta$  for which both estimates for  $W$  and  $B$  are available, that is the the  $\beta$  component. Exercise: Verify that  $\hat{\beta}_B$  and  $\hat{\beta}_W$  have covariance zero in order to justify the application of the above result.

**Generalized Least Squares** Now let's consider GLS estimation of the model

$$y = X\beta + Z\alpha + u$$

By treating the  $\alpha_i$ 's as random and  $u_i$  and  $\alpha_i$  as independent. we can compute the covariance matrix of  $\epsilon = u + Z\alpha$  as,

$$\begin{aligned}\Omega = E\epsilon\epsilon^\top &= E(Z\alpha + u)(Z\alpha + u)^\top \\ &= \sigma_\alpha^2 Z Z^\top + \sigma_u^2 I_{nT} \\ &= \sigma_u^2 I_{Tn} + \sigma_\alpha^2 (I_n \otimes T J_T) \\ &= \sigma_u^2 I_{nT} + T\sigma_\alpha^2 P\end{aligned}$$

Thus we may write the GLS estimator with  $\tilde{X} = [X:Z]$ ,

$$\hat{\delta} = \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = (\tilde{X}^\top \Omega^{-1} \tilde{X})^{-1} \tilde{X}^\top \Omega^{-1} y$$

This is sometimes called the Balestra-Nerlove estimator. Using the following Lemma it can be reformulated as yielding a simple least squares estimator after a preliminary transformation of the data.

*Lemma* (Nerlove) Let  $\sigma_\epsilon^2 = \sigma_u^2 + T\sigma_\alpha^2$ , then  $\Omega^{-1/2} = \sigma_\epsilon^{-1}P + \sigma_u^{-1}Q$ .

*Proof:* We will show this by computing directly that  $\Omega^{-1/2}\Omega\Omega^{-1/2} = I_{nT}$ , Noting that  $PQ = 0$ , we have,

$$\begin{aligned}(\sigma_\epsilon^{-1}P + \sigma_u^{-1}Q)[\sigma_u^2 I + T\sigma_\alpha^2 P](\sigma_\epsilon^{-1}P + \sigma_u^{-1}Q) &= \sigma_\epsilon^{-2}(\sigma_u^2 + T\sigma_\alpha^2)P + \sigma_u^{-2}\sigma_u^2 Q \\ &= P + Q \\ &= I_{nT}\end{aligned}$$

*Remark:*  $\Omega$  has only 2 distinct eigenvalues  $\sigma_u^2 + T\sigma_\alpha^2$  and  $\sigma_u^2$  and corresponding eigenvectors  $P$  and  $Q$ .

Having computed  $\Omega^{-1/2}$  we can transform (1) by  $\Omega^{-1/2}$  to obtain a spherical error,  $HT$  use  $\sigma_u\Omega^{-1/2}$  to get.

$$\sigma_u\Omega^{-1/2}y = (\theta P + Q)y = y - (1 - \theta)\bar{y}$$

where  $\theta = \sigma_u/(\sigma_u^2 + T\sigma_\alpha^2)^{1/2}$ , and similarly for the other variables. Here we are doing a form of ‘‘partial deviations from means’’ analogous to partial differencing in autocorrelation correction. Under our assumption, such estimates are efficient. Note that in unbalanced cases the transformations are slightly more complicated.

A third approach to deriving the classical panel data estimation methods for random effects is to regard the individual effects,  $\alpha_i$ , as parameters that need to be estimated, but ones on which you have some prior information

and proceed in a Bayesian fashion. To this end, we can consider, minimizing a penalized log likelihood expression,

$$\min_{\alpha, \beta, \gamma} \sum_i \sum_t (y_{it} - x_{it}\beta - z_i\gamma - \alpha_i)^2 + \lambda \sum_i \alpha_i^2.$$

where the  $\lambda$  represents the strength of the prior. The effect of the (second) penalty term is to shrink the fixed effect estimates – that would be obtained if the model were estimated with  $\lambda = 0$  – toward zero. Ideally, one would choose  $\lambda = \sigma_u^2/\sigma_\alpha^2$ . It is not particularly obvious from the linear algebra, but it turns out that this approach is equivalent to the two prior approaches that have been described.

### *Specification Tests*

Intuitively, if our assumption is violated, then  $\hat{\beta}_W$  is still consistent for  $\beta$ , but inefficient relative to the optimal  $\hat{\beta} = \Delta_{11}\hat{\beta}_B + (I - \Delta_{11})\hat{\beta}_W$ . This seems to be ideally suited for the  $H$ -test. We have an efficient estimator under  $H_0$  which is inconsistent under  $H_A$ :  $\hat{\beta}$  and a consistent estimator under  $H_A$ :  $\hat{\beta}_W$ .

There are three obvious options for testing:  $\omega_1 = \hat{\beta} - \hat{\beta}_W$ ,  $\omega_2 = \hat{\beta} - \hat{\beta}_B$ , and  $\omega_3 = \hat{\beta}_W - \hat{\beta}_B$ .

HT show that the three tests are asymptotically equivalent. As in other  $H$ -tests we can use the fact that under  $H_0$ , e.g.,  $V(\hat{\beta} - \hat{\beta}_W) = V(\hat{\beta}) - V(\hat{\beta}_W)$ .

*Estimation of  $\gamma$ .* Recall that we still have problems with estimation of  $\gamma$  in the fixed effects model and we might want to use fixed effects if we believed that there were endogeneity problems. We can think of distinguishing

$$\begin{aligned} X &= [X_1 : X_2] \\ Z &= [Z_1 : Z_2] \end{aligned}$$

where as in  $HT$   $X_2$  and  $Z_2$  are to be treated as *endogenous* and  $[X_1 : Z_1]$  as exogenous. Then we write

$$\tilde{y} = \tilde{X}\beta + \tilde{Z}\gamma + \tilde{\varepsilon}$$

where  $\Omega^{-1/2}y = \tilde{y}$  and so forth, and we have the two reduced form equations

$$[X_2 : Z_2] = [X_1 : Z_1]\Pi$$

For slightly esoteric reasons 2SLS and 3SLS are equivalent here – basically because of the fact that the “other equations” are *exactly* identified.

### A General Approach to Computation

The simplest, but perhaps not most memory efficient means of estimation is to take

$$\tilde{y} = \tilde{\beta} + \tilde{Z}\gamma + \tilde{\varepsilon}$$

and define the instruments

$$W = (QX_1, PX_1, QX_2, Z_1)$$

An interesting aspect of this approach is that it makes clear that  $X_1$  plays two roles. (i) estimation of  $\beta$ , and (ii) instrumental variable for  $Z_2$ .

This formulation also clarifies the conditions under which it is possible to estimate (identify) both  $\beta$  and  $\gamma$ . Clearly  $[QZ_1:QZ_2:Z_1]$  all serve as successful “instruments for themselves”. So the question reduces to: are there available IV’s for  $Z_2$ , the endogenous time invariant variables? This is easily seen to be answered by comparing the number of columns of  $PX_1$  to the number of columns of  $Z_2$ . There need to be at least as many columns of  $PX_1$  as the number of columns of  $Z_2$ .

*Estimating  $\sigma_\alpha^2$  and  $\sigma_u^2$ .* Finally we should address the question of estimating the variances in the matrix  $\Omega$ . I have two suggestions on this.

1. *Between approach.* In the  $B$ -data we have

$$\bar{\varepsilon}_i = \alpha_i + T_i^{-1} \sum_{t=1}^{T_i} u_{it}$$

so

$$V(\bar{\varepsilon}_i) = \sigma_\alpha^2 + T_i^{-1} \sigma_u^2$$

so we have a simple model for heteroscedasticity in this equation, and we can *estimate* by fitting the model

$$\bar{\varepsilon}_i^2 = \sigma_\alpha^2 + \sigma_u^2(1/T_i)$$

to the squared residuals from the between model.

2. Using the within data as a check of this, we have,

$$\tilde{u}_{it} = u_{it} - \bar{u}_{it}$$

so

$$V(\tilde{u}_{it}) \approx \sigma_u^2$$

and we can then compare  $\hat{\sigma}_u^2$  with what we get in the first approach based on the between data.

## References

Hausman J. and W.E. Taylor (1981). Panel Data and Unobservable Individual Effects, *Econometrica*, pp. 1377-98.