University of Illinois — Department of Economics
Fall 1998 — Roger Koenker

Economics 472

Lecture 13

Panel Data

Consider a model of the form

$$y_{it} = x_{it}\beta + z_i\gamma + \alpha_i + u_{it} \quad i = 1, \ldots, n \;\; t = 1, \ldots T \tag{1}$$

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, ...
α_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,

$$P = I_n \otimes T^{-1} 1_T 1_T' \equiv I_n \otimes J_T$$

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

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

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 $P = Z(Z'Z)Z'$ 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 β component is

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

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

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

$\hat{\beta}_w$ is consistent for β. 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 (β, γ) as $(\hat{\beta}_B, \hat{\gamma}_B)$ for "between." Can we combine β_B, β_W somehow?

A Simple Measurement Error Problem

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

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

   where

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

   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

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

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

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 σ_2 or Ω_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

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

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.

Now let's consider GLS estimation of this model by treating the α_i's as random and z_i and α_i as independent. Let

$$\Omega = E\varepsilon\varepsilon' \;\; = \;\; E(Z\alpha + u)(Z\alpha + u)'$$

$$
\begin{aligned}
&= \sigma_\alpha^2 ZZ' + \sigma_u^2 In_T \\
&= \sigma_u^2 I_{Tn} + \sigma_\alpha^2 (I_n \otimes TJ_T) \\
&= \sigma_u^2 I_{nT} + T\sigma_\alpha^2 P \\
&= (\sigma_u^2 | T\sigma_\alpha^2)P + \sigma_u^2 Q
\end{aligned}
$$

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

$$
\begin{aligned}
\hat{\delta} &= \begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = (\tilde{X}'\Omega^{-1}\tilde{X})^{-1}\tilde{X}'\Omega^{-1}y \\
\hat{\delta} &= \Delta\hat{\delta}_B + (I - \Delta)\hat{\delta}_W
\end{aligned}
$$

where $\Delta = V_W(V_B + V_W)^{-1}$ and $V_i = V(\hat{\delta}_i)$. This is often called the Balestra Nerlove estimator. This is eminently sensible in light of our simple measurement error model. We can formulate this as a preliminary transformation of the data and acheive some added insight using the following Lemma.

Lemma (Nerlove) $\Omega^{-1/2} = \sigma_\epsilon^{-1} P + \sigma_u^{-1} Q$
Proof: We will show computing directly that $\Omega^{-1/2}\Omega\Omega^{-1/2} = I_{nT}$, Noting that $PQ = 0$, we have,

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

Remark: Ω has only 2 distinct eigenvalues $\sigma_u^2 + T\sigma_\alpha^2$ and σ_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.

Specification Tests

Intuitively, if our assumption is violated, then β_W is still consistent for β, 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 γ. Recall that we still have problems with estimation of γ in the fixed effects model and we might want to use fixed effects if we believed that there were endogoneity problems. We can think of distinguishing

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

where as in *HT* X_2 and Z_2 are to be treated as *endogenous* and $[X_1 \vdots Z_1]$ as exogenous. The 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

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

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 β, and (ii) instrumental variable for Z_2.

This formulation also clarifies the conditions under which it is possible to estimate (identify) both β and γ. Clearly $[QZ_1 \vdots QZ_2 \vdots Z_1]$ 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 σ^2_α *and* σ^2_u. Finally we should address the question of estimating the variances in the matrix Ω. 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.

Reference:

Hausman J. and W.E. Taylor (1981). *Econometrica*, pp. 1377-98.