

## 13.3.1 TESTING THE SIGNIFICANCE OF THE GROUP EFFECTS

The  $t$  ratio for  $a_i$  can be used for a test of the hypothesis that  $\alpha_i$  equals zero. This hypothesis about one specific group, however, is typically not useful for testing in this regression context. If we are interested in differences across groups, then we can test the hypothesis that the constant terms are all equal with an  $F$  test. Under the null hypothesis of equality, the efficient estimator is pooled least squares. The  $F$  ratio used for this test is

$$F(n-1, nT-n-K) = \frac{(R_{LSDV}^2 - R_{Pooled}^2)/(n-1)}{(1 - R_{LSDV}^2)/(nT-n-K)}, \quad (13-9)$$

where *LSDV* indicates the dummy variable model and *Pooled* indicates the pooled or restricted model with only a single overall constant term. Alternatively, the model may have been estimated with an overall constant and  $n-1$  dummy variables instead. All other results (i.e., the least squares slopes,  $s^2$ ,  $R^2$ ) will be unchanged, but rather than estimate  $\alpha_i$ , each dummy variable coefficient will now be an estimate of  $\alpha_i - \alpha_1$  where group "1" is the omitted group. The  $F$  test that the coefficients on these  $n-1$  dummy variables are zero is identical to the one above. It is important to keep in mind, however, that although the statistical results are the same, the interpretation of the dummy variable coefficients in the two formulations is different.<sup>8</sup>

## 13.3.2 THE WITHIN- AND BETWEEN-GROUPS ESTIMATORS

We can formulate a pooled regression model in three ways. First, the original formulation is

$$y_{it} = \mathbf{x}'_{it}\beta + \alpha + \varepsilon_{it}. \quad (13-10a)$$

In terms of deviations from the group means,

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \beta + \varepsilon_{it} - \bar{\varepsilon}_i, \quad (13-10b)$$

while in terms of the group means,

$$\bar{y}_i = \bar{\mathbf{x}}'_i \beta + \alpha + \bar{\varepsilon}_i. \quad (13-10c)$$

All three are classical regression models, and in principle, all three could be estimated, at least consistently if not efficiently, by ordinary least squares. [Note that (13-10c) involves only  $n$  observations, the group means.] Consider then the matrices of sums of squares and cross products that would be used in each case, where we focus only on estimation of  $\beta$ . In (13-10a), the moments would accumulate variation about the overall means,  $\bar{y}$  and  $\bar{\mathbf{x}}$ , and we would use the total sums of squares and cross products,

$$\mathbf{S}_{xx}^{total} = \sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}})(\mathbf{x}_{it} - \bar{\mathbf{x}})' \quad \text{and} \quad \mathbf{S}_{xy}^{total} = \sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}})(y_{it} - \bar{y}).$$

For (13-10b), since the data are in deviations already, the means of  $(y_{it} - \bar{y}_i)$  and  $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$  are zero. The moment matrices are **within-groups** (i.e., variation around group means)

<sup>8</sup>For a discussion of the differences, see Suits (1984).

sums of squares and cross products,

$$\mathbf{S}_{xx}^{within} = \sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \quad \text{and} \quad \mathbf{S}_{xy}^{within} = \sum_{i=1}^n \sum_{t=1}^T (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)(y_{it} - \bar{y}_i).$$

Finally, for (13-10c), the mean of group means is the overall mean. The moment matrices are the **between-groups** sums of squares and cross products—that is, the variation of the group means around the overall means;

$$\mathbf{S}_{xx}^{between} = \sum_{i=1}^n T(\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}})(\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}})' \quad \text{and} \quad \mathbf{S}_{xy}^{between} = \sum_{i=1}^n T(\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}})(\bar{y}_i - \bar{\bar{y}}).$$

It is easy to verify that

$$\mathbf{S}_{xx}^{total} = \mathbf{S}_{xx}^{within} + \mathbf{S}_{xx}^{between} \quad \text{and} \quad \mathbf{S}_{xy}^{total} = \mathbf{S}_{xy}^{within} + \mathbf{S}_{xy}^{between}.$$

Therefore, there are three possible least squares estimators of  $\beta$  corresponding to the decomposition. The least squares estimator is

$$\mathbf{b}^{total} = [\mathbf{S}_{xx}^{total}]^{-1} \mathbf{S}_{xy}^{total} = [\mathbf{S}_{xx}^{within} + \mathbf{S}_{xx}^{between}]^{-1} [\mathbf{S}_{xy}^{within} + \mathbf{S}_{xy}^{between}]. \quad (13-11)$$

The within-groups estimator is

$$\mathbf{b}^{within} = [\mathbf{S}_{xx}^{within}]^{-1} \mathbf{S}_{xy}^{within}. \quad (13-12)$$

This is the LSDV estimator computed earlier. [See (13-4).] An alternative estimator would be the between-groups estimator,

$$\mathbf{b}^{between} = [\mathbf{S}_{xx}^{between}]^{-1} \mathbf{S}_{xy}^{between} \quad (13-13)$$

(sometimes called the **group means estimator**). This least squares estimator of (13-10c) is based on the  $n$  sets of groups means. (Note that we are assuming that  $n$  is at least as large as  $K$ .) From the preceding expressions (and familiar previous results),

$$\mathbf{S}_{xy}^{within} = \mathbf{S}_{xx}^{within} \mathbf{b}^{within} \quad \text{and} \quad \mathbf{S}_{xy}^{between} = \mathbf{S}_{xx}^{between} \mathbf{b}^{between}.$$

Inserting these in (13-11), we see that the least squares estimator is a **matrix weighted average** of the within- and between-groups estimators:

$$\mathbf{b}^{total} = \mathbf{F}^{within} \mathbf{b}^{within} + \mathbf{F}^{between} \mathbf{b}^{between}, \quad (13-14)$$

where

$$\mathbf{F}^{within} = [\mathbf{S}_{xx}^{within} + \mathbf{S}_{xx}^{between}]^{-1} \mathbf{S}_{xx}^{within} = \mathbf{I} - \mathbf{F}^{between}$$

The form of this result resembles the Bayesian estimator in the classical model discussed in Section 16.2. The resemblance is more than passing; it can be shown [see, e.g., Judge (1985)] that

$$\mathbf{F}^{within} = \{[\text{Asy. Var}(\mathbf{b}^{within})]^{-1} + [\text{Asy. Var}(\mathbf{b}^{between})]^{-1}\}^{-1} [\text{Asy. Var}(\mathbf{b}^{within})]^{-1},$$

which is essentially the same mixing result we have for the Bayesian estimator. In the weighted average, the estimator with the smaller variance receives the greater weight.

where  $\mathbf{i}_T$  is a  $T \times 1$  column vector of 1s. Since observations  $i$  and  $j$  are independent, the disturbance covariance matrix for the full  $nT$  observations is

$$\Omega = \begin{bmatrix} \Sigma & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Sigma & \mathbf{0} & \cdots & \mathbf{0} \\ & & & \ddots & \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \Sigma \end{bmatrix} = \mathbf{I}_n \otimes \Sigma. \quad (13-21)$$

13.4.1 GENERALIZED LEAST SQUARES

The generalized least squares estimator of the slope parameters is

$$\hat{\beta} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{y} = \left( \sum_{i=1}^n \mathbf{X}_i'\Omega^{-1}\mathbf{X}_i \right)^{-1} \left( \sum_{i=1}^n \mathbf{X}_i'\Omega^{-1}\mathbf{y}_i \right)$$

To compute this estimator as we did in Chapter 10 by transforming the data and using ordinary least squares with the transformed data, we will require  $\Omega^{-1/2} = [\mathbf{I}_n \otimes \Sigma]^{-1/2}$ . We need only find  $\Sigma^{-1/2}$ , which is

$$\Sigma^{-1/2} = \frac{1}{\sigma_\varepsilon} \left[ \mathbf{I} - \frac{\theta}{T} \mathbf{i}_T \mathbf{i}_T' \right],$$

where

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

The transformation of  $\mathbf{y}_i$  and  $\mathbf{X}_i$  for GLS is therefore

$$\Sigma^{-1/2}\mathbf{y}_i = \frac{1}{\sigma_\varepsilon} \begin{bmatrix} y_{i1} - \theta \bar{y}_i \\ y_{i2} - \theta \bar{y}_i \\ \vdots \\ y_{iT} - \theta \bar{y}_i \end{bmatrix}, \quad (13-22)$$

and likewise for the rows of  $\mathbf{X}_i$ .<sup>11</sup> For the data set as a whole, then, generalized least squares is computed by the regression of these partial deviations of  $y_{it}$  on the same transformations of  $\mathbf{x}_{it}$ . Note the similarity of this procedure to the computation in the LSDV model, which uses  $\theta = 1$ . (One could interpret  $\theta$  as the effect that would remain if  $\sigma_\varepsilon$  were zero, because the only effect would then be  $u_i$ . In this case, the fixed and random effects models would be indistinguishable, so this result makes sense.)

It can be shown that the GLS estimator is, like the OLS estimator, a matrix weighted average of the within- and between-units estimators:

$$\hat{\beta} = \hat{\mathbf{F}}^{within} \mathbf{b}^{within} + (\mathbf{I} - \hat{\mathbf{F}}^{within}) \mathbf{b}^{between}, \quad (13-23)$$

<sup>11</sup>This transformation is a special case of the more general treatment in Nerlove (1971b).

<sup>12</sup>An alternative form of this expression, in which the weighing matrices are proportional to the covariance matrices of the two estimators, is given by Judge et al. (1985).

where now,

$$\hat{\mathbf{F}}^{within} = [\mathbf{S}_{xx}^{within} + \lambda \mathbf{S}_{xx}^{between}]^{-1} \mathbf{S}_{xx}^{within},$$

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

To the extent that  $\lambda$  differs from one, we see that the inefficiency of least squares will follow from an inefficient weighting of the two estimators. Compared with generalized least squares, ordinary least squares places too much weight on the between-units variation. It includes it all in the variation in  $\mathbf{X}$ , rather than apportioning some of it to random variation across groups attributable to the variation in  $u_i$  across units.

There are some polar cases to consider. If  $\lambda$  equals 1, then generalized least squares is identical to ordinary least squares. This situation would occur if  $\sigma_u^2$  were zero, in which case a classical regression model would apply. If  $\lambda$  equals zero, then the estimator is the dummy variable estimator we used in the fixed effects setting. There are two possibilities. If  $\sigma_\varepsilon^2$  were zero, then all variation across units would be due to the different  $u_i$ s, which, because they are constant across time, would be equivalent to the dummy variables we used in the fixed-effects model. The question of whether they were fixed or random would then become moot. They are the only source of variation across units once the regression is accounted for. The other case is  $T \rightarrow \infty$ . We can view it this way: If  $T \rightarrow \infty$ , then the unobserved  $u_i$  becomes observable. Take the  $T$  observations for the  $i$ th unit. Our estimator of  $[\alpha, \beta]$  is consistent in the dimensions  $T$  or  $n$ . Therefore,

$$y_{it} - \mathbf{x}'_{it}\beta - \alpha = u_i + \varepsilon_{it}$$

becomes observable. The individual means will provide

$$\bar{y}_i - \bar{\mathbf{x}}'_i\beta - \alpha = u_i + \bar{\varepsilon}_i.$$

But  $\bar{\varepsilon}_i$  converges to zero, which reveals  $u_i$  to us. Therefore, if  $T$  goes to infinity,  $u_i$  becomes the  $\alpha_i \mathbf{d}_i$  we used earlier.

Unbalanced panels add a layer of difficulty in the random effects model. The first problem can be seen in (13-21). The matrix  $\Omega$  is no longer  $\mathbf{I} \otimes \Sigma$  because the diagonal blocks in  $\Omega$  are of different sizes. There is also groupwise heteroscedasticity, because the  $i$ th diagonal block in  $\Omega^{-1/2}$  is

$$\Omega_i^{-1/2} = \mathbf{I}_{T_i} - \frac{\theta_i}{T_i} \mathbf{i}_{T_i} \mathbf{i}'_{T_i}, \quad \theta_i = 1 - \frac{\sigma_\varepsilon}{\sqrt{\sigma_\varepsilon^2 + T_i \sigma_u^2}}.$$

In principle, estimation is still straightforward, since the source of the groupwise heteroscedasticity is only the unequal group sizes. Thus, for GLS, or FGLS with estimated variance components, it is necessary only to use the group specific  $\theta_i$  in the transformation in (13-22).

### 13.4.2 FEASIBLE GENERALIZED LEAST SQUARES WHEN $\Sigma$ IS UNKNOWN

If the variance components are known, generalized least squares can be computed as shown earlier. Of course, this is unlikely, so as usual, we must first estimate the

disturbance  
mation of tl

and

Therefore,

Since

if  $\beta$  were 0  
 $i$  would be

Since  $\beta$  is  
indeed, ur  
LSDV res

We have

$$\hat{s}_\varepsilon^2 =$$

The  $\beta$   
 $\beta$  are rec  
slopes. Th

This is th  
degrees of

It re  
spite of t  
the ordi  
most ca

<sup>13</sup>A form.