Chapter 5

The linear fixed-effects estimators: matrix creation

In this chapter three basic models and the data matrices needed to create estimators for them are defined[1]. The first is termed the cross-section model: although it incorporates some panel aspects, it is little different from a dummy variable cross-section. The second model allows for individual heterogeneity, treating this as a fixed effect to be removed by the covariance transformation; this is called the fixed-effects model. An alternative approach to the heterogeneity problem is time-differencing of the data. Section 5.3 and 5.4 consider differencing in balanced and unbalanced panels[2]. These are the differencing models.

For each model, three cases are considered:

unrestrictedSlopes and intercepts vary over time

pooledSlopes and intercepts are constant over time

restrictedSlopes are constant, but intercepts may vary over time.

5.1Cross-sections: the simple panel model

This model has no individual heterogeneity but allows for slopes and intercepts to vary over time.

---

[1] This chapter involves a large amount of matrix algebra which is straightforward but extensive. Although the different models are defined by similar equations, the basic equations are described in some detail here as they are directly implemented in the software and so form part of the validation of the programs. A shorter version of this chapter is in preparation as a discussion paper.

[2] The issue of panel balance does not materially affect the cross-section or fixed effects, although it simplifies the data requirements for the latter. As this can be done post extraction, we ignore the issue here and outline the adjustments in the next chapter. However, the balance of the dataset will determine whether the matrices for a differencing model have to be formed during extraction or whether they can be created post-extraction.

5.1.1 The unrestricted case

The "unrestricted" regression is

$$y_{it} = x_{it}\beta_t + \lambda_t + u_{it}$$
$$E(u_{it}) = 0 \qquad E(u_{it}u_{is}) = \sigma_{ts} \tag{5.1}$$

where x_{it} is a $1xK_x$ row vector, β_t is a $K_x x1$ column vector, and the other terms are all scalars. Stacked over all individuals for time t,

$$y_t = X_t\beta_t + J_t\lambda_t + u_t \tag{5.2}$$

where J_t is an N_t vector of ones and X_t is the $N_t x K_x$ matrix of the x_{it} stacked. Define

$$Q_t \equiv I_t - \frac{1}{N_t} J_t J_{t'} \tag{5.3}$$

where I_t is the $N_t x N_t$ identity matrix. Note that

$$Q_t = Q_{t'} = Q_t Q_{t'} \qquad Q_t J_t = 0 \tag{5.4}$$

Then premultiplying (5.2) by Q_t will remove the time effects:

$$\begin{aligned} Q_t y_i &= Q_t X_t \beta_t + Q_t J_t \lambda_t + Q_t u_t \\ &= Q_t X_t \beta_t + Q_t u_t \end{aligned} \tag{5.5}$$

For the system of equations over all t and n, the equivalent of (5.5) is

$$PY = PZ\zeta + PU \tag{5.6}$$

where

$$Y \equiv \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} \quad Z \equiv \begin{bmatrix} X_1 & 0 & & \\ 0 & X_2 & & \\ & & \vdots & \\ & & & X_T \end{bmatrix} \quad U \equiv \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix} \quad \zeta \equiv \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_T \end{bmatrix} \tag{5.7}$$

and P is the system equivalent of Q_t, namely

$$P \equiv \begin{bmatrix} Q_1 & 0 & & \\ 0 & Q_2 & & \\ & & \vdots & \\ & & & Q_T \end{bmatrix} \tag{5.8}$$

P is also symmetric and idempotent. The OLS solution to this will be to minimise

$$U'PU = Y'PY + \zeta'Z'PZ\zeta - 2\zeta'Z'PY \qquad \textbf{(5.9)}$$

which gives the normal equations

$$\hat{\zeta} = (Z'PZ)^{-1} Z'PY \qquad \textbf{(5.10)}$$

The constituents of (5.10) are block-diagonal:

$$Z'PZ = \begin{bmatrix} S_1 & 0 & & 0 \\ 0 & S_2 & & 0 \\ & & \vdots & \\ 0 & 0 & & S_T \end{bmatrix} \quad Z'PY = \begin{bmatrix} R_1 & 0 & & 0 \\ 0 & R_2 & & 0 \\ & & \vdots & \\ 0 & 0 & & R_T \end{bmatrix} \qquad \textbf{(5.11)}$$

where

$$S_t = X_{t'} Q_t X_t \qquad R_t = X_{t'} Q_t y_t \qquad \textbf{(5.12)}$$

and therefore (5.10) is equivalent to T separate estimations of

$$\hat{\beta}_t = S_t^{-1} R_t = (X_{t'} Q_t X_t)^{-1} X_{t'} Q_t y_t \qquad \textbf{(5.13)}$$

Note that

$$\begin{aligned} X_{t'} Q_t X_t &= X_{t'} X_t - \frac{1}{N_t} X_{t'} J_t J_{t'} X_t \\ &= X_{t'} X_t - N_t \overline{X}_{t'} \overline{X}_t \end{aligned} \qquad \textbf{(5.14)}$$

where

$$\overline{X}_{t'} = \frac{1}{N_t} J_{t'} X_t = \frac{1}{N_t} \sum_i x_{it} \qquad \textbf{(5.15)}$$

that is, the mean of x_{it} over all individuals for period t. Similarly,

$$X_{t'} Q_t y_t = X_{t'} y_t - N_t \overline{X}_{t'} \overline{y}_t \qquad \textbf{(5.16)}$$

and

$$\bar{y}_{t'} = \frac{1}{N_t} J_{t'} y_t = \frac{1}{N_t} \sum_i y_{it} \qquad \textbf{(5.17)}$$

Create a matrix $v_i'v_i$ by summing over cross-products for individuals for a period t:

$$v_t = [\ X_t \ \ J_t \ \ y_t\ ]$$

$$v_{t'} v_t = \begin{bmatrix} \sum_i x_{i't} x_{it} & \sum_i x_{it} & \sum_i x_{i't} y_{it} \\ \sum_i x_{i't} & N_t & \sum_i y_{it} \\ \sum_i y_{it} x_{it} & \sum_i y_{it} & \sum_i y_{it}^2 \end{bmatrix} \qquad \textbf{(5.18)}$$

Clearly $X_t'X_t$ is the top-left corner of $v_t'v_t$ and $X_t y_t$ the top right, but the matrix also contains all the other information necessary to calculate the terms in (5.14) and (5.16).

The OLS solution requires the summation of $v_t'v_t$ over i for each separate t. This is the format created by the extraction software.

Finally, note that the value of the time effects can be calculated from the mean for each period:

$$\begin{aligned} \frac{1}{N_t} \sum_i y_{it} &= \frac{1}{N_t} \sum_i x_{it} \beta_t + \frac{1}{N_t} \sum_i \lambda_t + \frac{1}{N_t} \sum_i u_{it} \\ \bar{y}_t &= \bar{x}_t \beta_t + \lambda_t + \bar{u}_t \\ \lambda_t &= \bar{y}_t - \bar{x}_t \beta_t - \bar{u}_t \\ \hat{\lambda}_t &= \bar{y}_t - \bar{x}_t \hat{\beta}_t \end{aligned} \qquad \textbf{(5.19)}$$

This information is readily obtained from the cross-product matrix.

5.1.2The pooled case

The "pooled" model constrains both slopes and intercepts to be constant over all periods:

$$y_{it} = x_{it}\beta + \lambda + u_{it} \tag{5.20}$$

Stacking over N and T:

$$Y = X\beta + J_{NT}\lambda + U \tag{5.21}$$

where X = [X_1' X_2' ... X_T']' and NT = ΣN_t. Premultiplying by the NT x NT matrix Q_{NT} (as defined above) gives

$$\begin{aligned} Q_{NT}Y &= Q_{NT}X\beta + Q_{NT}J_{NT}\lambda + Q_{NT}U \\ &= Q_{NT}X\beta + Q_{NT}U \end{aligned} \tag{5.22}$$

The time effect has been removed. The transformation matrix takes means over the whole regression, as all observations are treated alike. The OLS solution is

$$\hat{\beta} = (X'Q_{NT}X)^{-1}X'Q_{NT}Y \tag{5.23}$$

This differs from the unrestricted version in that the regressor matrices are no longer block diagonal and the summation is taken over the whole regression. This time the regressors split into

$$\begin{aligned} X'Q_{NT}X &= X'X - \frac{1}{NT}X\,J_{NT}J_{N'T}X \\ &= X'X - NT\,\overline{X'}\,\overline{X} \\ &= \sum_t\sum_i x_{i't}x_{it} - \frac{1}{NT}\sum_t\sum_i x_{i't}\sum_t\sum_i x_{it} \end{aligned} \tag{5.24}$$

In other words, the mean is this time taken from the whole set of observations. Summing the raw cross-product matrix gives

$$\sum_t v_{t'}v_t = \begin{bmatrix} \sum_t\sum_i x_{i't}x_{it} & \sum_t\sum_i x_{it} & \sum_t\sum_i x_{i't}y_{it} \\ \sum_t\sum_i x_{i't} & NT & \sum_t\sum_i y_{it} \\ \sum_t\sum_i y_{it}x_{it} & \sum_t\sum_i y_{it} & \sum_t\sum_i y_{it}^2 \end{bmatrix} \tag{5.25}$$

and so the cross-product matrix once more provides all the information necessary to calculate the estimator. In this case the intercept is found from

$$\frac{1}{NT}\sum_t\sum_i y_{it} = \frac{1}{NT}\sum_t\sum_i x_{it}\beta_t + \frac{1}{NT}\sum_t\sum_i \lambda_t + \frac{1}{NT}\sum_t\sum_i u_{it}$$

$$\hat{\lambda} = \bar{y} - \bar{x}\hat{\beta} \tag{5.26}$$

with the means taken over all variables.

5.1.3The restricted case

In the time-specific intercept case with constant slopes (the "within" estimator) the model is

$$y_{it} = x_{it}\beta + \lambda_t + u_{it} \tag{5.27}$$

To remove the time effect, stack over all individuals and premultiply by Q_t as for the unrestricted model

$$\begin{aligned} Q_t y_i &= Q_t X_t \beta + Q_t J_t \lambda_t + Q_t u_t \\ &= Q_t X_t \beta + Q_t u_t \end{aligned} \tag{5.28}$$

For the system of TxN equations, the appropriate transformation matrix is P, above:

$$PY = PX\beta + PU \tag{5.29}$$

but note that $X\beta$ is as defined for the pooled model, instead of the $Z\zeta$ in the unrestricted model. Again, the normal equations are

$$\hat{\beta} = (X'PX)^{-1} X'PY \tag{5.30}$$

Unlike the unrestricted model, these terms are no longer block-diagonal; however,

$$X'PX = \sum_t X_t' Q_t X_t \qquad X'PY = \sum_t X_t' Q_t y_t \tag{5.31}$$

and from (5.14) and (5.16) it may be observed that

$$\sum_t X_{t'} Q_t X_t = \sum_t X_{t'} X_t - \sum_t N_t \overline{X}_{t'} \overline{X}_t$$
$$\sum_t X_{t'} Q_t y_t = \sum_t X_{t'} y_t - \sum_t N_t \overline{X}_{t'} \overline{y}_t \qquad \textbf{(5.32)}$$

Therefore, the elements of the regression in this case are the sum of those in the unrestricted case after the latter has been adjusted to take deviations from time means (in the pooled case, the relevant matrices were summed before taking deviations). Thus the within estimator is also achievable from the $v_t'v_t$ cross-product matrix.

For the within case, the estimates of λ_t are given by

$$\frac{1}{N_t}\sum_i y_{it} = \frac{1}{N_t}\sum_i x_{it}\beta + \frac{1}{N_t}\sum_i \lambda_t + \frac{1}{N_t}\sum_i u_{it}$$
$$\textbf{(5.33)}$$
$$\hat{\lambda}_t = \overline{y}_t - \overline{x}_t\hat{\beta}$$

that is, by taking means for each period.

This method of taking deviations from time means is a one-way analysis-of-covariance approach. Clearly the coefficients could also be estimated by using time dummies, so why bother taking deviations to remove these dummies? The main reason is that it simplifies testing the different specifications, as the tests are carried out on the same number of coefficients in all three models. A second reason is that the analysis-of-covariance method merely tests for constancy over time; using the standard F-tests in the time-dummy specification tests for both constancy and a common level of the intercept in all three models[3].

5.1.4Variances and testing in the simple model

[3] A levels cross-section with time dummies included is available in the software, although this option does not calculate the specification tests. The structure of the input matrix is block-diagonal as before, and the basic data requirement is still the matrix $v_i'v_i$.

Consider variances for the unrestricted model first. Defining e_{it} as the residual error and E its NTx1 system equivalent, then

$$e_{it} = y_{it} - x_{it}\hat{\beta}_t - \hat{\lambda}_t \qquad \textbf{(5.34)}$$

or

$$PE = PY - PZ\hat{\zeta} \qquad \textbf{(5.35)}$$

using the same notation of (5.6). The residual sum of squares is given by

$$\begin{aligned} E'PE &= Y'PY - 2Y'PZ\hat{\zeta} + \hat{\zeta}'Z'PZ\hat{\zeta} \\ &= Y'PY - (2Y'PZ - Y'PZ(Z'PZ)^{-1}Z'PZ)\hat{\zeta} \\ &= Y'PY - (2Y'PZ - Y'PZ)\hat{\zeta} \\ &= Y'PY - Y'PZ\hat{\zeta} \end{aligned} \qquad \textbf{(5.36)}$$

or RSS=TSS-ESS. Substituting the estimated coefficients again,

$$\begin{aligned} E'PE &= Y'PY - Y'PZ(Z'PZ)^{-1}Z'PY \\ &= Y'P(I_{NT} - PZ(Z'PZ)^{-1}Z')PY \\ &= (U'P + Z'P)(I_{NT} - PZ(Z'PZ)^{-1}Z')(PZ + PU) \end{aligned} \qquad \textbf{(5.37)}$$

where NT is $\Sigma_t N_t$. As the middle PZ terms drop out,

$$\begin{aligned} E'PE &= U'P(I_{NT} - PZ(Z'PZ)^{-1}Z')PU \\ &= U'PU - U'PZ(Z'PZ)^{-1}Z'PU \end{aligned} \qquad \textbf{(5.38)}$$

E'PE is a scalar, and so the solution to (5.38) is the trace of E'PE. Taking expected values,

$$\begin{aligned} \_(E'PE) &= \_(tr[U'PU - U'PZ(Z'PZ)^{-1}Z'PU\ ]) \\ &= \_(tr[P'UU - PZ(Z'PZ)^{-1}Z'P'UU\ ]) \\ &= tr[P\ \_(U'U) - PZ(Z'PZ)^{-1}Z'P\ \_(U'U)\ ] \end{aligned} \qquad \textbf{(5.39)}$$

On the assumption that $\_(UU') = \sigma_u^2 I_{NT}$,

$$
\begin{aligned}
\_(E'PE) &= \sigma_u^2 tr[P - PZ(Z'PZ)^{-1} Z'P] \\
&= \sigma_u^2 (tr\, P - tr((Z'PZ)^{-1} Z'PZ)) \\
&= \sigma_u^2 (tr\, P - tr(I_K)) \\
&= \sigma_u^2 (\sum_t N_t (1 - \frac{1}{N_t}) - K) \\
&= \sigma_u^2 (\sum_t (N_t - 1) - K)
\end{aligned} \tag{5.40}
$$

Therefore

$$
\hat{\sigma}_u^2 = \frac{E'PE}{\sum_t N_t - T - K} \tag{5.41}
$$

A similar result holds for the pooled and within estimators. The main differences are the value of "K" and the trace of the first matrix. If K_x is the number of variables in x_{it}, then

$$
\begin{aligned}
\hat{\sigma}_u^2 &= \frac{E' PE_u}{\sum_t N_t - T - TK_x} \\
\hat{\sigma}_p^2 &= \frac{E' PE_p}{\sum_t N_t - 1 - K_x} \\
\hat{\sigma}_r^2 &= \frac{E' PE_r}{\sum_t N_t - T - K_x}
\end{aligned} \tag{5.42}
$$

where the u, p, and r subscripts refer to the unrestricted, pooled and restricted models. On these error assumptions, F-statistics for testing hypotheses of the latter two specifications are

$$F_{\substack{up \\ u\ vs.\ p}} = \frac{(E'PE_p - E'PE_u)/((T-1)(K_x+1))}{E'PE_u/(\sum_T N_t - T - TK_x)}$$

$$F_{\substack{ur \\ u\ vs.\ r}} = \frac{(E'PE_r - E'PE_u)/(K_x(T-1))}{E'PE_u/(\sum_T N_t - T - TK_x)} \tag{5.43}$$

$$F_{\substack{rp \\ r\ vs.\ p}} = \frac{(E'PE_p - E'PE_r)/(T-1)}{E'PE_r/(\sum_t N_t - T - K_x)}$$

Large values imply a rejection of the more restricted hypothesis.

One refinement is to note that, as ζ is being estimated for the unrestricted model over T separate regressions, it is relatively easy to calculate separate estimates of the variance for each period. In this case, the time-heteroscedastic errors for period t are (from equation (5.41)):

$$\hat{\sigma}^2_{ut} = \frac{e_{t'} Q_t e_t}{N_t - 1 - K} \tag{5.44}$$

These are the errors reported by the regression program. However, the F-tests in (5.43) make the assumption that the variance is homoscedastic; the potential benefit of time-heteroscedastic errors appears small relative to the additional complexity of the corrected F-statistic.

5.2Fixed-effects: allowing for individual heterogeneity

The models in this section are the more usual "panel" models in that they allow for individual heterogeneity. this is treated as a fixed effect and removed by taking deviations from individual means. Time dummies are left in the regression. This is because the transformation matrix which removes individual dummies cannot remove time dummies, and vice versa. It is possible to construct a matrix which removes both effects, but the structure of the resulting matrices are too complicated for our purposes. In any case, it will be demonstrated that it is not necessary to remove the time dummies to construct estimators for the stability of coefficients over time[4].

5.2.1The unrestricted case

Let the fundamental equation be

$$y_{it} = x_{it}\beta_t + \alpha_i + \lambda_t + u_{it}$$

$$E(u_{it}) = 0 \qquad E(u_{it}u_{is}) = \sigma_{ts} \qquad \textbf{(5.45)}$$

or, stacked for individual i,

$$y_i = w_i\beta + J_i\alpha_i + L_i\lambda + u_i \qquad \textbf{(5.46)}$$

where J_i is a T_i vector of ones, λ is a vector of the time effects, the T_ixT matrix L is a TxT identity matrix with the rows removed for which an observation is missing, and

$$w_i \equiv \begin{bmatrix} x_i1 & 0 & & \\ 0 & x_i2 & & \\ & & \vdots & \\ & & & x_{iT} \end{bmatrix} \quad \beta \equiv \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_T \end{bmatrix} \qquad \textbf{(5.47)}$$

with rows of w_i removed where observations are missing. Define

$$Q_i \equiv I_i - \frac{1}{T_i} J_i J_{i'} \qquad \textbf{(5.48)}$$

where I_i is the T_ixT_i identity matrix. Note that

$$Q_i = Q_{i'} = Q_i Q_{i'} \qquad Q_i J_i = 0 \qquad \textbf{(5.49)}$$

4 Perfect collinearity between the time dummies and the individual dummies means that one time dummy should be dropped to remove the linear dependency. However, this can easily be done post extraction, and it does not change the qualitative results of this chapter at all. Thus, although the X'X of this section is actually singular, this is ignored solely to simplify the exposition. The next chapter discusses appropriate corrections.

Assume the individual effects α_i are fixed (which enables single-stage regression). To remove them from the equation, premultiply by Q_i:

$$\begin{aligned} Q_i y_i &= Q_i w_i \beta + Q_i J_i \alpha_i + Q_i L_i \lambda + Q_i u_i \\ &= Q_i w_i \beta + Q_i L_i \lambda + Q_i u_i \\ &= Q_i z_i \zeta + Q_i u_i \end{aligned} \quad (5.50)$$

with $z_i=[w_i\ L_i]$ and $\zeta=[\beta'\ \lambda']'$. The OLS solution to this will be to minimise

(5.51)

The normal equations for this are

$$\hat{\zeta} = \left(\sum_i^N z_{i'} Q_i z_i \right)^{-1} \sum_i^N z_{i'} Q_i y_i \quad (5.52)$$

The breakdown of the first element here is:

$$\begin{aligned} z_{i'} Q_i z_i &= z_{i'} (I_i - \frac{1}{T_i} J_i J_{i'}) z_i \\ &= z_{i'} z_i - \frac{1}{T_i} z_{i'} J_i J_{i'} z_i \\ &= z_{i'} z_i - T_i \bar{z}_{i'} \bar{z}_i \end{aligned} \quad (5.53)$$

Similarly,

$$z_{i'} Q_i y_i = z_{i'} y_i - T_i \bar{z}_{i'} \bar{y}_i \quad (5.54)$$

where

$$\bar{z}_i \equiv \frac{1}{T_i} J_{i'} z_i \qquad \bar{y}_i \equiv \frac{1}{T_i} \sum_t y_{it} \quad (5.55)$$

are the mean values of the elements of z_i and y_i. Because of the block nature of w_i and L_i, this is equivalent to creating a vector of $1/T_i$ times all variables:

$$\bar{z}_i = \left[\frac{1}{T_i} \frac{1}{T_i} \cdots \frac{1}{T_i} \right] \begin{bmatrix} x_i 1 & 0 & & 1 & 0 & \\ 0 & x_i 2 & & 0 & 1 & \\ & & \vdots & & & \vdots \\ & & x_{iT} & & & 1 \end{bmatrix} \quad \textbf{(5.56)}$$

$$= \frac{1}{T_i} [\, x_i 1 \; x_i 2 \ldots x_{iT} \;\; 1 1 \ldots 1]$$

It can be seen that

(5.57)

Therefore, if defining

$$v_i \equiv [\, x_i 1 \; x_i 2 \ldots x_{iT} \;\; 1 1 \ldots 1] \quad \textbf{(5.58)}$$

to give the cross-product matrix

$$v_{i'}v_i = \begin{bmatrix} x_i1'x_i1 & x_i1'x_i2 & & & x_i1' & x_i2' & \\ x_i2'x_i1 & x_i2'x_i2 & & & x_i1' & x_i2' & \\ & & \vdots & & & \vdots & \\ & & & x_{i'T}\,x_{iT} & & & x_{i'T} \\ & & & & & & \\ x_i1 & x_i2 & & & 1 & 1 & \\ x_i1 & x_i2 & & & 1 & 1 & \\ & & \vdots & & & \vdots & \\ & & & x_{iT} & & & 1 \end{bmatrix} \qquad \textbf{(5.59)}$$

then it is clear that $T_{izi}'z_i = T_i^{-1}v_i'v_i$ and $z_i'z_i$ is the block diagonal of $v_i'v_i$. A similar story holds for $z_i'Q_iy_i$. As y_i is a T_i x 1 column vector,

(5.60)

Then $T_{izi}'y_i$ is the horizontal summation of $T_i^{-1}v_i'y_i'$ and $z_i'y_i$ are the diagonal terms $x_{i1}'y_{i1}..x_{i1}'y_{i1}$ and $y_{i1}..y_{i1}$ of $v_i'y_i'$.

Unlike the simple case, the time effects λ are now directly estimated, rather than being extracted from the time means as in Section 5.1. This makes no real difference to the outcome. The issue of testing for levels and constancy of the intercepts does not arise as all three models are based around deviations from the mean of the whole regression; therefore the tests for a constant intercept in all periods amount to testing for zero intercepts in all periods.

5.2.2The pooled case

For the pooled model, the hypothesis is that β and λ are constant over time:

$$y_{it} = x_{it}\beta + \alpha_i + u_{it} \tag{5.61}$$

or, stacked for individual i,

$$y_i = X_i\beta + J_i\alpha_i + u_i \tag{5.62}$$

where

$$X_i \equiv [\,'x_i 1' \;\; x_i 2' \;\ldots\; x_{i'T}\,] \tag{5.63}$$

and x_{it} contains the constant term[5]. Using Q_i as above and still assuming the individual effects α_i are fixed, the latter are removed by premultiplying by Q_i:

$$\begin{aligned} Q_i y_i &= Q_i X_i\beta + Q_i J_i\alpha_i + Q_i u_i \\ &= Q_i X_i\beta + Q_i u_i \end{aligned} \tag{5.64}$$

The OLS minimisation problem is

(5.65)

giving

$$\hat{\beta} = \left(\sum_{i}^{N} X_{i'} Q_i X_i\right)^{-1} \sum_{i}^{N} X_{i'} Q_i y_i \tag{5.66}$$

Again, this breaks down into:

$$\begin{aligned} X_{i'} Q_i X_i &= X_{i'} X_i - \frac{1}{T_i} X_{i'} J_i J_{i'} X_i \\ &= X_{i'} X_i - T_i \overline{X}_{i'} \overline{X}_i \\ \\ X_{i'} Q_i y_i &= X_{i'} y_i - T_i \overline{X}_{i'} \overline{y}_i \end{aligned} \tag{5.67}$$

[5] Including the constant term within the x_{it} at this stage is merely a simplification and has no bearing on the results.

with

$$\overline{X}_i \equiv \frac{1}{T_i} J_{i'} X_i \qquad \overline{y}_i \equiv \frac{1}{T_i} \sum_t y_{it} \qquad \textbf{(5.68)}$$

the mean values of the elements of X_i and y_i. Note that the mean of X_i is different from the mean of z_i because of the block nature of the latter. In this case

$$X_{i'} X_i = \sum_t x_{i't} x_{it} \qquad \textbf{(5.69)}$$

and, summing over i,

$$\sum_i X_{i'} X_i = \sum_i \sum_t x_{i't} x_{it} = \sum_t \sum_i x_{i't} x_{it} \qquad \textbf{(5.70)}$$

In addition

$$\begin{aligned} \overline{X}_i &= \left[\frac{1}{T_i} \frac{1}{T_i} \cdots \frac{1}{T_i} \right] \begin{bmatrix} x_i 1 \\ x_i 2 \\ \vdots \\ x_{iT} \end{bmatrix} \\ &= \frac{1}{T_i} \sum_t x_{it} \end{aligned} \qquad \textbf{(5.71)}$$

Summing over i gives

$$\sum_i \frac{1}{T_i} \sum_t x_{i't} \sum_t x_{it} = \sum_i \frac{1}{T_i} \sum_t \sum_s x_{i't} x_{is} = \sum_t \sum_s \sum_i \frac{1}{T_i} x_{i't} x_{is} \qquad \textbf{(5.72)}$$

All the summations over time are from 1 to T. Although T_i can vary from individual to individual, missing values are set to zero and so the mean calculations are correct whether the summation is over T or T_i.

Using the definition of v_i from (5.58), it is apparent that $X_i'X_i$ is the sum of the diagonal blocks of v_iv_i and $T_i x_i'X_i$ is the sum of the all of the blocks of the means matrix. $X_i'Q_iy_i$ has a similar structure:

$$\sum_i X_{i'} y_i = \sum_i \sum_t x_{i't} y_{it} = \sum_t \sum_i x_{i't} y_{it}$$
$$\sum_i \frac{1}{T_i} \sum_t x_{i't} \sum_t y_{it} = \sum_i \frac{1}{T_i} \sum_t \sum_s x_{i't} y_{is} = \sum_t \sum_s \sum_i \frac{1}{T_i} x_{i't} y_{is} \tag{5.73}$$

5.2.3The restricted case

Now consider β constant and λ varying over time:

$$y_{it} = x_{it} \beta + \alpha_i + \lambda_t + u_{it} \tag{5.74}$$

or, stacked,

$$y_i = X_i \beta + J_i \alpha_i + L_i \lambda + u_i \tag{5.75}$$

where all the terms are as defined above. As for Section 5.2.1, the constant term is separated out from x_{it}. This is the commonest form of linear panel estimator found in applied work, even though it involves significant (and often not tested) restrictions on the basic model (5.45). Premultiplying by Q_i to remove the individual effects:

$$\begin{aligned} Q_i y_i &= Q_i X_i \beta + Q_i J_i \alpha_i + Q_i L_i \lambda + Q_i u_i \\ &= Q_i X_i \beta + Q_i L_i \lambda + Q_i u_i \\ &= Q_i C_i \chi + Q_i u_i \end{aligned} \tag{5.76}$$

with C_i=[X_i L_i] and χ=[β' λ']'. The OLS solution is

$$\hat{\chi} = \left(\sum_i^N C_{i'} Q_i C_i \right)^{-1} \sum_i^N C_{i'} Q_i y_i \tag{5.77}$$

Breaking this down,

$$\begin{aligned} C_{i'} Q_i C_i &= C_{i'} C_i - \frac{1}{T_i} C_{i'} J_i J_{i'} C_i \\ &= C_{i'} C_i - T_i \overline{C}_{i'} \overline{C}_i \end{aligned}$$
$$C_{i'} Q_i y_i = C_{i'} y_i - T_i \overline{C}_{i'} \overline{y}_i \tag{5.78}$$

with

$$\overline{C}_i \equiv \frac{1}{T_i} J_i' C_i \qquad \overline{y}_i \equiv \frac{1}{T_i} \sum_t y_{it} \tag{5.79}$$

for the mean values. The mean of y_i is the same in all these alternative hypotheses, but again, the mean of C_i has some slightly different elements:

(5.80)

Therefore

$$C_i' C_i = \begin{bmatrix} \sum_t x_{i't} x_{it} & 0 & 0 & & 0 \\ & & & & \\ 0 & 1 & 0 & & \\ 0 & 0 & 1 & & \\ & & & \vdots & \\ 0 & & & & 1 \end{bmatrix} \tag{5.81}$$

and

$$T_i \overline{C}_i' \overline{C}_i = \frac{1}{T_i} \begin{bmatrix} \sum_t x_{i't} \sum_t x_{it} & \sum_t x_{i't} & \dots & \sum_t x_{i}2' \\ & & & \\ \sum_t x_{it} & 1 & & \\ \vdots & \vdots & & \\ \sum_t x_{it} & & & 1 \end{bmatrix} \tag{5.82}$$

As for the pooled model, $\Sigma x_{it}'\Sigma x_{it}$ can be calculated post extraction. For the simple means of x_{it} over t, if column c of x_{it} contains the constant term then

$$\frac{1}{T}\sum_t x_{i't}\, x_{itc} = \frac{1}{T}\sum_t x_{i't} 1 = \frac{1}{T_i}\sum_t x_{it} \qquad \textbf{(5.83)}$$

From the definition of v_i it is apparent that $C_i'C_i$ is the block diagonal of v_iv_i and $T_{i}C_i'C_i$ is T_i^{-1} times one of the constant intersections of $v_i'v_i$. However, in both cases the non-constant sections (the x_{it} bits) must be summed. Again, $C_i'Q_iy_i$ has a similar structure, with

$$C_{i'} y_i = \begin{bmatrix} \sum_t x_{i't} y_{it} \\ y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} \qquad T_i \bar{C}_{i'} \bar{y}_i = \frac{1}{T_i}\begin{bmatrix} \sum_t x_{i't} \sum_t y_{it} \\ \sum_t y_{it} \\ \sum_t y_{it} \\ \vdots \\ \sum_t y_{it} \end{bmatrix} \qquad \textbf{(5.84)}$$

Clearly all that is needed to estimate all models are the two matrices $\Sigma v_i'v_i$ and $\Sigma T_i^{-1}v_i'v_i$. As these can be created piecewise (that is, for each individual i, $v_i'v_i$ is found and then summed into a totalling cross-product matrix), and as this matrix can be created without reference to the particular variables used in a particular regression (that is, $\Sigma v_i'v_i$ contains all variables of interest of which a subset are used in any particular regression), this presents no real difficulties to the extraction or analysis software.[6]

It is clear that the data matrices for the pooled and restricted versions are constructed merely by summing over time the relevant elements from the unrestricted matrix. This is what might be expected. In the simple cross-section models, taking deviations for each period led to different matrices being needed for different hypotheses about variation over time, as the mean used depended upon the model in question. In the fixed effects models, the transformation is

[6] The fact that the constant terms are sometimes included in x_{it} and sometimes represented as separate elements is merely for notational convenience and makes little difference to the analysis. However, for analytical purposes it is easier if the constant terms are always included in x_{it} rather than being grouped separately, and so this is the structure the extraction software uses.

being made with respect to deviations from individual means. In that context, different assumptions about time-varying slopes and intercepts merely amounts to a rearrangement of the variables; the same mean (over all periods for one individual) is used. Had assumptions been made about coefficients varying over individuals, then testing the different models would have involved calculating different means - as the cross-section models necessitated.

Note that there are two significant disadvantages to the fixed-effects formulation used here. Firstly, each cross-product means matrix $\Sigma T_i^{-1} v_i' v_i$ is dependent upon a particular value of T: as the divisor in the means matrix is T_i, and as $T_i \leq T$, a different value of T alters the range of acceptable T_is. Only in the case of the balanced matrix can the means for several years be calculated from one $\Sigma v_i' v_i$ matrix, because in this case the constant divisor T can be taken outside all the summations over i. In the general case of an unbalanced panel, it is no longer possible to run regressions on multiple combinations of years using the same cross-product means matrix. Separate means matrices need to be created for each selection of years.

Secondly, creation of these matrices involves much computer time and memory - significantly more than for the simple panel model. However, note that $\Sigma v_i' v_i$ for a given T contains $\Sigma v_i' v_i$ for any smaller T. As the two matrices can be created independently, it is sensible to extract $\Sigma v_i' v_i$ in one run and then the means matrix $\Sigma T_i^{-1} v_i' v_i$ separately - possibly for several different values of T. This is a relatively efficient solution to the extraction problem, and is implemented in the extraction software.

In fact, the data requirement is less onerous than stated above. Consider the two matrices requested for this section and the previous one. It is clear that each of the diagonal blocks in $\Sigma v_i' v_i$ (5.59) corresponds to one of the $v_t' v_t$ matrices in (5.18). There is no necessity to calculate the whole matrix (5.59); the time-effects cross-products matrices will do equally well. The analysis software takes account of this. However, as the practical difference between creating $\Sigma v_i' v_i$ and $\Sigma T_i^{-1} v_i' v_i$ is negligible, the extraction software writer allows for the creation of full $\Sigma v_i'v_i$ matrices. Such a matrix can contain useful information on interperiod correlations; in addition, it allows for time-differencing models to be constructed, as will be shown in Section 5.3[7].

5.2.4 Variances and testing in the heterogeneous model

Consider variances for the general model first. The sum of squared residuals is

$$\begin{aligned} \Sigma e_{i'} Q_i e_i &= \Sigma\left[y_{i'} Q_i y_i - 2\, y_{i'} Q_i z_i \hat{\zeta} + \hat{\zeta}\, z_{i'} Q_i z_i \hat{\zeta} \right] \\ &= \Sigma\, y_{i'} Q_i y_i - \Sigma\, y_{i'} Q_i z_i \left(\Sigma\, z_{i'} Q_i z_i \right)^{-1} \Sigma\, z_{i'} Q_i y_i \\ RSS &= TSS + ESS \end{aligned} \quad (5.85)$$

where e_i is the residual error. Define the symmetric, idempotent matrix P:

$$P \equiv \begin{bmatrix} Q_1 & 0 & & 0 \\ 0 & Q_2 & & 0 \\ & & \vdots & \\ 0 & 0 & & Q_N \end{bmatrix} \quad (5.86)$$

Then

$$\sum_i^N e_{i'} Q_i e_i = E'PE \quad \sum_i^N y_{i'} Q_i y_i = Y'PY \quad \sum_i^N z_{i'} Q_i z_i = Z'PZ \quad \sum_i^N z_{i'} Q_i y_i = Z'PY \quad (5.87)$$

where $E=[e_1' \; e_2' \; ... \; e_N']'$, $Y=[y_1' \; y_2' \; ... \; y_N']'$, and $Z=[z_1' \; z_2' \; ... \; z_N']'$. Define $U=[u_1' \; u_2' \; ... \; u_N']'$ and I_{NT} as the identity matrix with ΣT_i rows. Substituting (5.87) in (5.85) gives

$$\begin{aligned} E'PE &= Y'PY - Y'PZ(Z'PZ\,)^{-1} Z'PY \\ &= (U'P + Z'P)(\,I_{NT} - PZ(Z'PZ\,)^{-1} Z')(PZ + PU) \\ &= U'PU - U'PZ(Z'PZ\,)^{-1} Z'PU \end{aligned} \quad (5.88)$$

As for section 5.1, E'PE is a scalar with its solution equal to its trace. Taking expected

---

[7] The full $\Sigma v_i'v_i$ matrix should also allow for a "minimum-distance" model to be generated, which allows for a more general error structure (Hsiao (1986) ch.3; Chamberlain (1984)). This is currently being investigated.

values,

$$\begin{aligned} \_(E'PE) &= \_(tr[U'PU - U'PZ(Z'PZ)^{-1}Z'PU]) \\ &= tr[P\,\_(U'U) - PZ(Z'PZ)^{-1}Z'P\,\_(U'U)] \end{aligned} \quad \textbf{(5.89)}$$

Maintaining the earlier assumption that $\_(UU')=\sigma_u^2 I_{NT}$ leads to

$$\begin{aligned} \_(E'PE) &= \sigma_u^2 tr[P - PZ(Z'PZ)^{-1}Z'P] \\ &= \sigma_u^2(tr\,P - tr(I_K)) \\ &= \sigma_u^2(\sum_i T_i - N - K) \end{aligned} \quad \textbf{(5.90)}$$

Therefore

$$\hat{\sigma}_u^2 = \frac{E'PE}{\sum_i T_i - N - K} \quad \textbf{(5.91)}$$

This result holds for all three models, as the structure of P is identical in all three. The only difference between them is the value of K. If K_x is the number of variables in x_{it} (excluding any constant term), then

$$\begin{aligned} \hat{\sigma}_u^2 &= \frac{E'PE_u}{\sum_i T_i - N - T(K_x + 1)} \\ \hat{\sigma}_p^2 &= \frac{E'PE_p}{\sum_i T_i - N - (K_x + 1)} \\ \hat{\sigma}_r^2 &= \frac{E'PE_r}{\sum_i T_i - N - (K_x + T)} \end{aligned} \quad \textbf{(5.92)}$$

where the u, p, and r subscripts refer to the unrestricted, pooled and restricted models.

On the error assumptions, F-statistics for testing hypotheses of unrestricted, pooled and time-dummy are

$$F_{\substack{up\\u\,vs.\,p}} = \frac{(E'PE_p - E'PE_u)/((T-1)(K_x+1))}{E'PE_u/(\sum_i T_i - N - T(K_x+1))}$$

$$F_{\substack{ur\\u\,vs.\,r}} = \frac{(E'PE_r - E'PE_u)/(K_x(T-1))}{E'PE_u/(\sum_i T_i - N - T(K_x+1))} \tag{5.93}$$

$$F_{\substack{pt\\r\,vs.\,p}} = \frac{(E'PE_p - E'PE_r)/(T-1)}{E'PE_r/(\sum_i T_i - N - (K_x+T))}$$

One difficulty is the calculation of N and ΣT_i. However, from the bottom-right hand corner of $T_i^{-1}v_i'v_i$ we have the $T_i x T_i$ block

$$\frac{1}{T_i}\begin{bmatrix} 1 & 1 & & 1 \\ 1 & 1 & & 1 \\ & \vdots & & \\ 1 & 1 & & 1 \end{bmatrix} \tag{5.94}$$

This will actually be stored as a TxT block with zeroes in the appropriate places, but this is qualitatively the same result. Summing the diagonal gives

$$\frac{1}{T_i}\sum_{t=1}^{T_i} 1 = \frac{T_i}{T_i} = 1 \tag{5.95}$$

Therefore, summing the constant diagonal for the whole matrix gives

$$\sum_{i=1}^{N}\left(\frac{1}{T_i}\sum_{t=1}^{T_i} 1\right) = \sum_{i=1}^{N} 1 = N \tag{5.96}$$

Meanwhile, for the total number of observations, let d_{it} be a marker, 1 if individual i was observed in period t and 0 otherwise. Then

$$\sum_{i=1}^{N} T_i = \sum_{i=1}^{N}\sum_{t=1}^{T_i} 1 = \sum_{i=1}^{N}\sum_{t=1}^{T} d_{it} = \sum_{t=1}^{T}\sum_{i=1}^{N} d_{it} = \sum_{t=1}^{T} N_t \tag{5.97}$$

Clearly ΣT_i is then the sum of the diagonal of the units section of $\Sigma v_i'v_i$, as this gives the number of observations in each period.

Observe that N comes from the means-matrix and thus is time dependent. This is correct: if an individual only has observations outside the range of years used for a particular regression, then he may not be included in the "N" for that regression. the total number of observations is just a straightforward tally of the number of observations in each period, and so the total number of times an individual has been observed is irrelevant.

5.3Differencing models (complete observations sets)

An alternative approach to individual heterogeneity is to take time differences. This also removes the individual effect; however, for unbalanced panels the arithmetic is more complicated. This section considers the case where all individuals have the same number of observations.

5.3.1The unrestricted case

The unrestricted case is the same as for the heterogenous model of Section 5.2:

$$y_i = w_i \beta + J_i \alpha_i + L_i \lambda + u_i \qquad \textbf{(5.98)}$$

where the matrices are as defined for equation (5.46). To remove heterogeneity by time-differencing, the transformation matrix is

$$\underset{T_i-1 \times T_i}{Q_i} \equiv \begin{bmatrix} -1 & 1 & 0 & & & \\ 0 & -1 & 1 & & & \\ & & & \vdots & & \\ & & & 0 & -1 & 1 \end{bmatrix} \qquad \textbf{(5.99)}$$

We still have $Q_i J_i = 0$; however, Q_i is no longer idempotent. Instead,

$$\underset{T_i \times T_i}{Q_{i'}Q_i} = \begin{bmatrix} 1 & -1 & 0 & & & \\ -1 & 2 & -1 & & & \\ & & & \vdots & & \\ & & & -1 & 2 & -1 \\ & & & 0 & -1 & 1 \end{bmatrix} = I_i + S_i \qquad \textbf{(5.100)}$$

where

$$S_i \equiv \begin{bmatrix} 0 & -1 & 0 & & & \\ -1 & 1 & -1 & & & \\ & & & \vdots & & \\ & & & -1 & 1 & -1 \\ & & & 0 & -1 & 0 \end{bmatrix} \qquad \textbf{(5.101)}$$

The normal equations for the coefficients are (from (5.52)):

$$\hat{\zeta} = \left(\sum_i^N z_{i'} Q_{i'} Q_i z_i \right)^{-1} \sum_i^N z_{i'} Q_{i'} Q_i y_i \qquad \textbf{(5.102)}$$

Note that one time dummy will have to be deleted because the matrix is not of full rank. Section 5.4 discusses the issue of identification of time dummies in more detail. Breaking down the components of (5.102),

$$z_{i'} Q_{i'} Q_i z_i = z_{i'} z_i + z_{i'} S_i z_i \qquad z_{i'} Q_{i'} Q_i y_i = z_{i'} y_i + z_{i'} S_i y_i \qquad \textbf{(5.103)}$$

Definex $_{it}$=(-x_{it}); that is, the negative of x_{it}. Then, from the definition of S_i and z_{it},

(5.104)

It can be seen that $z_i'S_iz_i$ is just the Hadamard product of $z_i'z_i$ and the (K+1)x(K+1) equivalent of S_i:

$$z_{i'} S_i z_i = z_{i'} z_i \circ H_i \qquad \textbf{(5.105)}$$

where J is a K-vector of ones and

$$H_i \equiv \begin{bmatrix} 0_{KK} & -J'J & 0_{KK} & 0_{KK} & & 0_K1 & -J & 0_K1 & 0_K1 & \\ -J'J & J'J & -J'J & 0_{KK} & & -J & J & -J & 0_K1 & \\ 0_{KK} & -J'J & J'J & -J'J & & 0_K1 & -J & J & -J & \\ & \vdots & & & & \vdots & & & & \\ & -J' & J' & -J' & 0_{1K} & & -1 & 1 & -1 & 0 \\ & 0_{1K} & -J' & J' & -J' & & 0 & -1 & 1 & -1 \\ & 0_{1K} & 0_{1K} & -J' & 0_{1K} & & 0 & 0 & -1 & 0 \end{bmatrix} \qquad \textbf{(5.106)}$$

Therefore, all the information to calculate this specification of the time-differencing model is in the matrix $v_i'v_i$. As the matrix H_K is a constant matrix depending only on the years of the analysis and not on any individual variables, then $\Sigma_i v_i'v_i$ can be built up and used for regressions on multiple years.

Finally, calculating $z_i'Q_i'Q_iy_i$ requires

$$z_{i'} S_i y_i = \begin{bmatrix} \bar{x}_i 1' y_i 2 \\ \bar{x}_i 2' y_i 1 + x_i 2' y_i 2 + \bar{x}_i 2' y_i 3 \\ \vdots \\ \bar{x}_{i'T-1} y_{iT-2} + x_{i'T-1} y_{iT-1} + \bar{x}_{i'T-1} y_{iT} \\ \bar{x}_{i'T} y_{iT-1} \\ - y_i 2 \\ - y_i 1 + y_i 2 - y_i 3 \\ \vdots \\ - y_{iT-2} + y_{iT-1} - y_{iT} \\ - y_{iT-1} \end{bmatrix} \qquad \textbf{(5.107)}$$

which has a similar structure to $z_i'S_iz_i$ and can also be calculated from $v_i'v_i$.

Restrictions can be placed on the specification in a manner analogous to the earlier sections.

5.3.2The pooled case

Let β and λ be constant over time:

$$y_{it} = x_{it} \beta + \alpha_i + u_{it} \qquad \textbf{(5.108)}$$

x_{it} contains the constant term. Stacking over t and using Q_i as above and still assuming the individual effects α_i are fixed, the latter is removed by premultiplying by Q_i:

$$\begin{aligned} Q_i y_i &= Q_i X_i \beta + Q_i J_i \alpha_i + Q_i u_i \\ &= Q_i X_i \beta + Q_i u_i \end{aligned} \qquad \textbf{(5.109)}$$

The OLS normal equations are

$$\hat{\beta} = \left(\sum_i^N X_{i'} Q_{i'} Q_i X_i \right)^{-1} \sum_i^N X_{i'} Q_{i'} Q_i y_i \qquad \textbf{(5.110)}$$

Again, this breaks down into:

$$X_{i'} Q_{i'} Q_i X_i = X_{i'} X_i + X_{i'} S_i X_i \qquad X_{i'} Q_{i'} Q_i y_i = X_{i'} y_i + X_{i'} S_i y_i \qquad \textbf{(5.111)}$$

But in this case

$$X_{i'} S_i X_i = \bar{x}_i 2' x_i 1 + (\bar{x}_i 1' + x_i 2' + \bar{x}_i 3') x_i 2 + (\bar{x}_i 2' + x_i 3' + \bar{x}_i 4') x_i 3 + \dots$$

(5.112)

$$= \sum_{t=2}^{T_i-1} \left(x_{i't} x_{it} - x_{i't} x_{it-1} - x_{i't} x_{it+1} \right) - x_i 2' x_i 1 - x_{i'T-1} x_{iT}$$

wherex $_{it}$=(-x$_{it}$), as before. $X_i'S_iy_i$ has a similar structure:

$$X_{i'} S_i y_i = \bar{x}_i 2' y_i 1 + (\bar{x}_i 1' + x_i 2' + \bar{x}_i 3') y_i 2 + (\bar{x}_i 2' + x_i 3' + \bar{x}_i 4') y_i 3 + \dots$$

(5.113)

$$= \sum_{t=2}^{T_i-1} \left(x_{i't} y_{it} - x_{i't} y_{it-1} - x_{i't} y_{it+1} \right) - x_i 2' y_i 1 - x_{i'T-1} y_{iT}$$

Once more the result is a constant matrix multiple of the constructed $\Sigma v_i'v_i$; in fact, the result is merely the summation over t of the x'x section of the unrestricted case. Thus the pooled case can be feasibly constructed from a correctly formed cross-product.

5.3.3The restricted case

Finally, consider β constant and λ varying over time:

$$y_{it} = x_{it} \beta + \alpha_i + \lambda_t + u_{it} \qquad \textbf{(5.114)}$$

or, stacked,

$$y_i = X_i \beta + J_i \alpha_i + L_i \lambda + u_i \qquad \textbf{(5.115)}$$

where all the terms are as defined above. Again, the constant term is made explicit. Premultiplying by Q_i removes the individual effects:

$$Q_i y_i = Q_i X_i \beta + Q_i L_i \lambda + Q_i u_i \qquad \textbf{(5.116)}$$

with $C_i=[X_i\ L_i]$ and $\chi=[\beta'\ \lambda']'$, as in Part II. The OLS solution is

$$\hat{\chi} = \left(\sum_{i}^{N} C_{i'} Q_{i'} Q_i C_i \right)^{-1} \sum_{i}^{N} C_{i'} Q_{i'} Q_i y_i \qquad \textbf{(5.117)}$$

Breaking this down,

$$C_{i'} Q_{i'} Q_i C_i = C_{i'} C_i + C_{i'} S_i C_i \qquad C_{i'} Q_{i'} Q_i y_i = C_{i'} y_i + C_{i'} S_i y_i \qquad \textbf{(5.118)}$$

where, as would be expected from previous cases

$$C_{i'} C_i = \begin{bmatrix} X_{i'} X_i & 0 & 0 & & 0 \\ & & & & \\ 0 & 1 & 0 & & \\ 0 & 0 & 1 & & \\ & & & \vdots & \\ 0 & & & & 1 \end{bmatrix} \quad C_{i'} S_i C_i = \begin{bmatrix} X_{i'} S_i X_i & & X_{i'} S_i & & \\ & & & & \\ & 0 & -1 & & \\ S_i X_i & -1 & 1 & & \\ & & & \vdots & \\ & & & & 0 \end{bmatrix} \textbf{(5.119)}$$

with $X_i'S_iX_i$ as defined above and

$$X_{i'} S_i = \left[\bar{x}_i 2' \; (\bar{x}_i 1' + x_i 2' + \bar{x}_i 3') \; (\bar{x}_i 2' + x_i 3' + \bar{x}_i 4') \; \ldots \right] \qquad \textbf{(5.120)}$$

Again, this presents no especial difficulties in the construction of the OLS estimator from a cross-product matrix. The information requirements are less than for the estimators in section 5.2, as all that is needed to estimate all models is the matrix $\Sigma v_i' v_i$, which can be created piecewise[8].

Thus, although the time-differencing model requires more computer power than the simple panel models of Section 5.1, it is less of a burden than the deviations models. In addition, the cross-product matrix used for the time-differencing model ($\Sigma v_i' v_i$,) is not dependent on the actual number of years used in regression.

[8] Again, the representation of the constant terms sometimes as being included in x_{it} and sometimes as separate elements is merely for notational convenience.

On the other hand, the time-differencing model is a less efficient solution than the deviations model for two reasons. Firstly, it does not take full account of all observations, as the end observations only contribute once to the model; secondly, the deviations estimators use the cross-correlation between all the explanatory variables to determine the coefficients, whereas the time differencing approach merely uses correlations between two periods.

Finally, the differencing estimator described in this section suffers from the problem of missing data. The above solution is only appropriate where all individuals have a full set of observations for the period of interest. A balanced sample is very much the exception in NES extractions, and so section 5.4 considers the issue of time differencing from an unbalanced panel.

5.3.4. Variances and testing in the time-differencing model

Again, consider variances for the general model first. As the system is still block-diagonal, the system equivalent of $Q_i'Q_i$ is P:

$$P \equiv \begin{bmatrix} Q_{1'}Q_1 & 0 & 0 \\ 0 & Q_{2'}Q_2 & 0 \\ & \vdots & \\ 0 & 0 & Q_{N'}Q_N \end{bmatrix} \qquad \textbf{(5.121)}$$

Using the logic and notation of the earlier models

$$\begin{aligned} E'PE &= Y'PY - Y'PZ(Z'PZ\ )^{-1} Z'PY \\ &= (U'P + Z'P)(\,I_{NT} - PZ(Z'PZ\ )^{-1} Z')(PZ + PU) \\ &= U'PU - U'PZ(Z'PZ\ )^{-1} Z'PU \end{aligned} \qquad \textbf{(5.122)}$$

The solution to this scalar is the trace of E'PE. Taking expected values,

$$\begin{aligned} \_(E'PE) &= \_(tr[U'PU - U'PZ(Z'PZ\ )^{-1} Z'PU\ ]) \\ &= tr[P\ \_(U'U) - PZ(Z'PZ\ )^{-1} Z'P\ \_(U'U)\ ] \end{aligned} \qquad \textbf{(5.123)}$$

and assuming that_(UU')=$\sigma_u^2 I_{NT}$

$$\begin{aligned} \_(E'PE) &= \sigma_u^2 tr[P - PZ(Z'PZ)^{-1} Z'P] \\ &= \sigma_u^2 (tr\, P - tr(I_K)) \\ &= \sigma_u^2 (tr(I_{NT}) + \sum_i tr(S_i) - tr(I_K)) \\ &= \sigma_u^2 (NT + \sum_i (T-2) - K) \\ &= \sigma_u^2 (2N(T-1) - K) \end{aligned} \quad \textbf{(5.124)}$$

Thus

$$\hat{\sigma}_u^2 = \frac{E'PE}{2N(T-1) - K} \quad \textbf{(5.125)}$$

Again, the results is common to all three models, with only the value of K changing. Taking K_x as the number of variables in x_{it} (excluding the constant), the relevant adjustments are

$$\begin{aligned} \hat{\sigma}_u^2 &= \frac{E'PE_u}{2N(T-1) - T(K_x + 1)} \\ \hat{\sigma}_p^2 &= \frac{E'PE_p}{2N(T-1) - (K_x + 1)} \\ \hat{\sigma}_r^2 &= \frac{E'PE_r}{2N(T-1) - (K_x + T)} \end{aligned} \quad \textbf{(5.126)}$$

where the u, p, and r subscripts now refer to the unrestricted, pooled and restricted models. Note that the inefficiency of the differencing approach is to some extent reflected in the higher denominators in the estimation of the standard error.

The appropriate F-tests for this specification are

$$\begin{aligned} \underset{u\ vs.\ p}{F_{up}} &= \frac{(E'PE_p - E'PE_u)/((T-1)(K_x + 1))}{E'PE_u/(2N(T-1) - T(K_x + 1))} \\ \underset{u\ vs.\ r}{F_{ur}} &= \frac{(E'PE_r - E'PE_u)/(K_x(T-1))}{E'PE_u/(2N(T-1) - T(K_x + 1))} \\ \underset{r\ vs.\ p}{F_{rp}} &= \frac{(E'PE_p - E'PE_r)/(T-1)}{E'PE_r/(2N(T-1) - (K_x + T))} \end{aligned} \quad \textbf{(5.127)}$$

The value of N can be easily extracted from the cross-product matrix. T and K are known.

5.4Time differencing with missing observations

If the panels are unbalanced, then the approach of the previous section will not work. This is because the matrix H_i is different for each individual, and so the post facto creation of the necessary data matrices from $\Sigma v_i'v_i$ is not possible. Observations may only be used where both T and T-1 are observed, and these cannot be identified from any of the group matrices.

In this case, matrices must be constructed initially in first differences. Define

$$\begin{aligned} \hat{x}_{it} &= x_{it} - x_{it-1} \qquad x_{it}, x_{it-1} \text{ observed} \\ \hat{x}_{it} &= 0 \qquad\qquad\qquad \text{otherwise} \end{aligned} \tag{5.128}$$

and similar terms for y_{it} and u_{it}. Clearly, as the individual-specific term α_{it} is constant ($\alpha_{it}=\alpha_i$ for all t), it drops out of the final equation. With no individual heterogeneity in the equation for y_{it}, the approach of Section 5.1 can be used, and so all the results of that section apply here. The fact that the "means transformation" (the matrix Q_t) is being applied to differenced data does not have any implications for the estimation of the slope coefficients.

However, there may be some confusion over the equation to be estimated. Consider a straight differencing of the stacked equation (5.2):

$$\begin{aligned} y_t - y_{t-1} &= X_t\beta_t - X_{t-1}\beta_{t-1} + J_t\lambda_t - J_{t-1}\lambda_{t-1} + u_t - u_{t-1} \\ &= X_t\beta_t - X_{t-1}\beta_{t-1} + J_t(\lambda_t - \lambda_{t-1}) + u_t - u_{t-1} \\ &= (X_t - X_{t-1})\beta_t + X_{t-1}(\beta_t - \beta_{t-1}) + J_t(\lambda_t - \lambda_{t-1}) + u_t - u_{t-1} \end{aligned} \tag{5.129}$$

where the requirement that an individual must be observed in T and T-1 to be included ensures that $J_t=J_{t-1}$. To estimate this equation requires that both X_t and X_{t-1} be included explicitly as regressors. This is feasible for extraction and analysis goes, but it requires a degree of carefulness in ensuring that the correct variables are included.

To simplify the analysis, the NES extraction software makes the assumption that $\beta_t=\beta_{t-1}$ in (5.129)[9]. Although there is no conceptual necessity for this, it is justified on the grounds of practicality. The form of (5.129) with this assumption is

$$\begin{aligned} y_t - y_{t-1} &= (X_t - X_{t-1})\beta_t + J_t(\lambda_t - \lambda_{t-1}) + u_t - u_{t-1} \\ \hat{y}_t &= \hat{X}_t\beta_t + J_t\hat{\lambda}_t + \hat{u}_t \end{aligned} \qquad \textbf{(5.130)}$$

Clearly the form of (5.130) is identical to (5.2), with dependent and explanatory variables and a unit vector. Premultiplication by the matrix Q_t defined in (5.3) will still remove the time dummies and so, once the matrices have been constructed correctly, all the mathematics of the first section can be applied. The presence of the unit vector in (5.130) means that N_t can be recovered from the cross-product matrix.

This time-differencing approach implies a restriction on the slope coefficients that is not present in the earlier models. A necessary assumption for the differencing approach of (5.130) is that the slopes do not change significantly from one year to the next, although over the whole period the slopes may shift. The calculated βs now represent the best-fitting slope for any two consecutive periods; that is, instead of the ζ of equation (5.7) being estimated slopes for 1975, 1976, 1977... and so on, it now represents a separate slope coefficient for two years at a time (1975/6, 1976/7, 1977/8...).

This has implications for the interpretation of these coefficients. If the slopes do change over time, then the differencing model will show less variation than the basic time-effects model, simply because the differencing approach estimates average slopes. A larger degree of autocorrelation may also be expected, with the slopes evolving over time. Models of evolving coefficients have been developed by some authors, but they are outwith the scope of this work. The F-tests described will only check for parameter constancy, not systematic change.

[9] This means that the automatic software procedures produce data appropriate for (5.130). Obviously, there is no reason why a cross-product matrix should not contain the data to estimate (5.129).

We should remark that, although the practice described above involves a limitation on the values of the coefficients, it remains a more general specification than is often found in differencing models. In most applied analysis, all coefficients save the intercept are kept constant over time; here only constancy over any two consecutive periods is assumed.

For the time dummies, taking time differences means that the change in λ over any two periods is now being estimated. This holds for both (5.129) and (5.130), although few authors seem to recognise this point. As this is a reparameterisation and not a restriction (such as that implied by moving from (5.129) to (5.130)), this is unaffected by any evolution of the intercept, in the sense that the estimation of any change is unbiased. However, in contrast to the slope coefficients, we no longer have an absolute measure of the level of the intercept, and so estimates of the intercept are no longer directly comparable with the results from models in levels[10].

5.4.1. Variances and testing in the differenced unbalanced panels

Taking differences before calculating the matrix means that the expected error term is no longer the same as in section 5.1.4. In the differenced model, the error term is $(u_t\text{-}u_{t-1})$. Retaining the assumption that _(UU')=$I_{NT}\sigma^2$, the correct variance for this term is therefore

$$\_\left[(u_t - u_{t-1})(u_t - u_{t-1})'\right] = \_(u_{t'} u_{t-1}) + \_(u_t u_{t'-1}) = 2\, I_t \sigma^2 \qquad \textbf{(5.131)}$$

In this case, then the expected value of the error term from (5.40) becomes

$$\_(E'PE) = 2\sigma_u^2 (\sum_t (N_t - 1) - K) \qquad \textbf{(5.132)}$$

and so the estimated error terms for the three models are now

[10] The levels of the intercepts can be recovered from the means of the regression.

$$\hat{\sigma}_u^2 = \frac{E' PE_u}{2(\sum_t N_t - T - TK_x)}$$

$$\hat{\sigma}_p^2 = \frac{E' PE_p}{2(\sum_t N_t - 1 - K_x)} \qquad \textbf{(5.133)}$$

$$\hat{\sigma}_r^2 = \frac{E' PE_r}{2(\sum_t N_t - T - K_x)}$$

The F-statistics are unchanged as the extra 2s cancel out.

As in the previous sections, note that estimating ζ for the unrestricted model over T separate regressions enables the calculation of time heteroscedastic errors. Allowing for heteroscedasticity in (5.133) gives

$$\_\left[(u_t - u_{t-1})(u_t - u_{t-1})'\right] = I_t(\sigma_t^2 + \sigma_{t-1}^2) \equiv I_t \sigma_{ut}^2 \qquad \textbf{(5.134)}$$

and therefore

$$\hat{\sigma}_{ut}^2 = \frac{e_{t'} Q_t e_t}{N_t - 1 - K} \qquad \textbf{(5.135)}$$

Note that this estimated variance is a compound term, and so there is no need to divide the RSS by two. These are the reported errors, but again, the F-tests for model restrictions are based on homoscedastic errors.