Econometric Methods for Panel Data

Based on the books by BALTAGI: *Econometric Analysis of Panel Data* and by HSIAO: *Analysis of Panel Data*

Robert M. Kunst
robert.kunst@univie.ac.at

University of Vienna
and
Institute for Advanced Studies Vienna

May 4, 2016

Outline

Unbalanced panels: basic issues

Often, the number of available time points varies across individuals, for individual i there are T_i observations, T_i may vary with i.

Most available software programs handle this generalization well. Some issues must be modified relative to balanced panels, but changes are minor in many respects. It is definitely not recommended to discard some observations in order to make the panel balanced.

Fixed effects in unbalanced panels

The one-way fixed-effects (FE) estimator (or LSDV or within estimator) is obtained by transforming all variables to deviations from their individual-specific average

$$\tilde{y}_{it} = y_{it} - T_i^{-1} \sum_{s=1}^{T_i} y_{is}, \quad \tilde{X}_{it} = X_{it} - T_i^{-1} \sum_{s=1}^{T_i} X_{is}.$$

It is easily seen that OLS on the transformed variables is again equivalent to a regression among the untransformed variables after including individual dummies (hence, LSDV). The FE estimator can again be written as

$$\hat{\beta} = (\mathbf{X}'\mathbf{Q}\mathbf{X})^{-1}\mathbf{X}'\mathbf{Q}y,$$

as in the balanced case, but with a slightly less simple 'purging' matrix **Q**.

The purging matrix in the unbalanced case

The matrix that describes this transformation is still block-diagonal, as in the balanced case

$$\mathbf{Q} = \begin{bmatrix} \mathbf{I}_{T_1} - {T_1}^{-1}\mathbf{J}_{T_1} & 0 & \cdots & 0 \\ 0 & \mathbf{I}_{T_2} - {T_2}^{-1}\mathbf{J}_{T_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & \mathbf{I}_{T_N} - {T_N}^{-1}\mathbf{J}_{T_N} \end{bmatrix},$$

but it cannot be written as $\mathbf{I}_N \otimes (\mathbf{I}_T - T^{-1}\mathbf{J}_T)$. One may write $\mathrm{diag}(\mathbf{I}_{T_1} - {T_1}^{-1}\mathbf{J}_{T_1}, \ldots, \mathbf{I}_{T_N} - {T_N}^{-1}\mathbf{J}_{T_N})$.

Random effects in unbalanced panels

The one-way random-effects panel ('error component') model

$$y_t = \alpha + X_{it}'\beta + u_{it}, \quad u_{it} = \mu_i + \nu_{it}, \quad i = 1, \ldots, N; t = 1, \ldots, T_i,$$

has a block-diagonal error covariance matrix

$$\mathrm{E}uu' = \mathbf{\Omega} = \mathrm{diag}(\sigma_\nu^2 \mathbf{I}_{T_1} + \sigma_\mu^2 \mathbf{J}_{T_1}, \ldots, \sigma_\nu^2 \mathbf{I}_{T_N} + \sigma_\mu^2 \mathbf{J}_{T_N}),$$

thus a simple structure, although it cannot be represented any more as $\sigma_\nu^2 \mathbf{I}_{NT} + \sigma_\mu^2 \mathbf{I}_N \otimes \mathbf{J}_N$, as in the balanced case.

Inverting the covariance matrix

The matrix $\mathbf{\Omega}$ can be inverted blockwise, with blocks inverted like in the balanced case:

$$(\sigma_\nu^2 \mathbf{I}_{T_i} + \sigma_\mu^2 \bar{\mathbf{J}}_{T_i})^{-1} = \sigma_\nu^{-2}(\mathbf{I}_{T_i} - \bar{\mathbf{J}}_{T_i}) + (T_i \sigma_\mu^2 + \sigma_\nu^2)^{-1} \bar{\mathbf{J}}_{T_i},$$

such that $\mathbf{\Omega}^{-1}$ is a block-diagonal matrix with N such blocks. In practice, the only critical problem is how to estimate the variances $\sigma_\nu^2, \sigma_\mu^2$.

Note that this scheme implies N *different* values for variance ratios such as

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

How to estimate component variances

The most popular methods are based on ANOVA (analysis of variance) and rely on the expectations of expressions such as $\hat{u}'\mathbf{Q}\hat{u}$ and $\hat{u}'\mathbf{P}\hat{u}$ for various residuals $\hat{u}$:

- OLS residuals yield the Wallace and Hussain estimators;
- FE residuals yield the Amemiya estimators;
- using both FE and between residuals yields the Swamy and Arora estimators.

For all these cases, expectations are known (though not trivial) expressions in the variance parameters, and a simple linear equation system must be solved.

Alternative but complex methods, such as minimum-norm or maximum-likelihood, are not necessarily better than the ANOVA algorithms. Most of them require nonlinear optimization.

More aspects of unbalanced panels

- Two-way panel analysis becomes much more complicated.
- Testing for the existence of (fixed or random) effects is straightforward.
- Hausman tests can be applied in unbalanced panels. It is less easy to justify FE generally, as even asymptotically it may be that not necessarily $T_i \to \infty, i = 1, \ldots, N$.