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

April 25, 2018

Outline

The hypotheses of poolability tests

These tests help select the panel model to be estimated, within the framework of fixed-effects models. For example:

- Are there individual effects or is it preferable to ignore them and to estimate by pooled OLS?
- Are there time effects on top of the individual effects?
- Are the coefficients β really constant across individuals?

These tests follow a simple principle: restricted and unrestricted models are compared via LR or Wald statistics. In most cases, null distributions are F or χ^2, with degrees of freedom matching the number of restrictions.

Testing for individual effects

Here, the null model (restricted) is

$$y_{it} = \alpha + \beta' X_{it} + \nu_{it},$$

with *iid* errors, and the alternative (unrestricted) model is

$$y_{it} = \alpha + \beta' X_{it} + \mu_i + \nu_{it},$$

assuming $\sum_{i=1}^{N} \mu_i = 0$. The null hypothesis may be written as:

$$H_0 : \mu_i = 0, i = 1, \ldots, N.$$

Testing for individual effects: the statistic

The traditional restriction-test statistic is

$$F_{1-way} = \frac{(ESS_R - ESS_U)/(N-1)}{ESS_U/((T-1)N-K)},$$

as there are $N-1$ restrictions from the maintained hypothesis to the null. There are $(T-1)N-K$ degrees of freedom in the unrestricted model. Under H_0, this statistic will be distributed as $F_{N-1,N(T-1)-K}$, assuming Gaussian errors.

Other analogous tests

For the null of the pooled model and the alternative of the two-way model, the distribution of the F–statistic will be distributed as $F_{(N+T-2,NT-N-T+1-K)}$ under the null, assuming Gaussian errors.

For the null of the one-way individual-effects model and the alternative of the two-way model, the F–statistic will be distributed as $F_{(T-1,NT-N-T+1-K)}$ under the null, assuming Gaussian errors.

For large NT, χ^2 versions may be used instead, which are independent of Gaussian error assumptions.

Testing homogeneity of individual slopes

In one specification, the unrestricted model is

$$y_{it} = \alpha + \beta_i' X_{it} + \mu_i + \nu_{it},$$

and the null can be expressed as

$$H_0 : \beta_1 = \ldots = \beta_N = \beta, \mu_1 = \ldots = \mu_N = 0,$$

with $(K+1)(N-1)$ restrictions. The unrestricted model has only $N(T-K-1)$ degrees of freedom. If this test rejects, either the one-way model should be considered or individuals must be modelled separately. It may make more sense just to test the slopes, with $K(N-1)$ restrictions.

Constellations in panel models and degrees of freedom

		coefficients β	
effects	name	β	β_i
α	OLS	$NT - K - 1$	$N(T - K) - 1$
α_i	one-way	$N(T - 1) - K$	$N(T - K - 1)$
α_{it}	two-way	$N(T - 1) - T + 1 - K$	$N(T - K - 1) - T + 1$

The test by Roy and Zellner

Some researchers (among them, Baltagi) criticize that the usual *F*–test checks poolability in an otherwise perfect Gauss-Markov regression with $\mathrm{E}\nu\nu' = \sigma_\nu^2 \mathbf{I}$. To cope with this problem, one may also test for $\beta_i = \beta$ and/or $\beta_t = \beta$ in a (one-way or two-way) RE model.

This test is called the Roy-Zellner test. Essentially, it tests for ‘fixed-type’ poolability of slope coefficients in a random-effects model.

Likelihood-ratio tests for variance parameters

Assume L_u is the likelihood of an unrestricted model, L_r is the likelihood of a restricted model. A lemma says that the *likelihood-ratio* (LR) statistic

$$LR = 2\left(\log L_u - \log L_r\right),$$

will be, under the null of the restricted model, asymptotically distributed χ^2 with degrees of freedom matching the number of restrictions.

The lemma requires several *regularity conditions*. In testing a variance parameter for zero, these conditions are violated and the property is not guaranteed to hold.

LR test for random effects in two-way panels

The unrestricted RE model

$$y_{it} = \alpha + \beta' X_{it} + u_{it},$$
$$u_{it} = \mu_i + \lambda_t + \nu_{it},$$

with the restricted null $H_0 : \sigma_\mu^2 = \sigma_\lambda^2 = 0$ violates the regularity conditions. The LR test statistic has the non-standard distribution

$$\frac{1}{4}\chi^2(0) + \frac{1}{2}\chi^2(1) + \frac{1}{4}\chi^2(2),$$

where $\chi^2(0)$ denotes point mass at zero.

LR test for random effects in one-way panels

The unrestricted RE model

$$y_{it} = \alpha + \beta' X_{it} + u_{it},$$
$$u_{it} = \mu_i + \nu_{it},$$

with the restricted null $H_0 : \sigma_\mu^2 = 0$ again violates the regularity conditions. The LR test statistic has the non-standard distribution

$$\frac{1}{2}\chi^2(0) + \frac{1}{2}\chi^2(1).$$

An analogous test can be used for random time effects. These tests are due to Gourieroux, Holly & Monfort.

LM tests for random effects

Lagrange-multiplier (LM) tests have standard χ^2 asymptotics. They have usually less power. One may test for two-way random effects using pooled-OLS residuals $\tilde{u}$ by LM:

$$
\begin{aligned}
LM &= LM_1 + LM_2, \\
LM_1 &= \frac{NT}{2(T-1)} \left\{ 1 - \frac{\tilde{u}' (\mathbf{I}_N \otimes \mathbf{J}_T) \tilde{u}}{\tilde{u}'\tilde{u}} \right\}^2, \\
LM_2 &= \frac{NT}{2(N-1)} \left\{ 1 - \frac{\tilde{u}' (\mathbf{J}_N \otimes \mathbf{I}_T) \tilde{u}}{\tilde{u}'\tilde{u}} \right\}^2,
\end{aligned}
$$

or one may test for one-way effects by using LM_1 or LM_2. This test is due to Breusch & Pagan.

The Hausman test principle

Hausman tests can be used in all situations where two model specifications and two estimators are available with the following properties:

1. In the restricted model (null), the estimator $\hat{\theta}$ is efficient, the estimator $\tilde{\theta}$ is consistent though typically not efficient;
2. in the unrestricted model (alternative), the estimator $\hat{\theta}$ is inconsistent, the estimator $\tilde{\theta}$ is consistent.

Then, the difference $q = \hat{\theta} - \tilde{\theta}$ should diverge under the alternative and it should converge to zero under the null. Moreover, under the null q and $\hat{\theta}$ should be uncorrelated.

The RE and the FE model

The null of the RE and the alternative of the FE model correspond to the Hausman situation:

1. In the RE model, the GLS-type RE estimator is efficient by construction for Gaussian errors, the FE estimator and even the OLS estimator are consistent;
2. in the FE model, the RE estimator is inconsistent, because of the omitted-variable effect, while FE is consistent by construction.

Two views on the RE inconsistency in the FE model

1. The estimator $(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}y$ is inconsistent, if the true model is $y = X\beta + Z_\mu\mu + \nu$, as the regressors Z_μ are omitted;
2. Mundlak showed that the RE estimator for a stochastic regression model with X and Z_μ correlated is identical to the FE estimator: the RE estimator imposes independence of effects and covariates.

Some argue, however, that Mundlak's alternative is not really the same concept as the fixed-effects model.

The Hausman test statistic

The Hausman test statistic is defined as

$$m = q'(\mathrm{var}\hat{\beta}_{FE} - \mathrm{var}\hat{\beta}_{RE})^{-1} q,$$

with $q = \hat{\beta}_{FE} - \hat{\beta}_{RE}$. Under RE, the matrix difference in brackets is positive, as the RE estimator is efficient and any other estimator has a larger variance.

The statistic m is distributed χ^2 under the null of RE, with degrees of freedom determined by the dimension of β, K.

R^2 in a panel model

Should the variation due to effects be part of the explained variation or not? If yes, the R^2 has little to say on the importance of the substantial covariates β.

There is no unanimous agreement on which R^2 to report in a panel. Some programs (STATA etc.) follow the suggestion by Wooldridge and report three measures: within R^2, between R^2, and overall R^2.

The within R^2

For the *within* R^2, the *total sum of squares* TSS is defined as

$$TSS = \sum_{i=1}^{N} \sum_{t=1}^{T} (y_{it} - \bar{y}_i)^2 ,$$

i.e. after adjusting for effects. Then, also residuals in the *residual sum of squares* ESS rely on the 'purged' model, and the statistic

$$1 - \frac{ESS}{TSS}$$

is maximized by LSDV or the FE estimator.

The between R^2

For the *between* R^2, the *total sum of squares* TSS is defined as

$$TSS = \sum_{i=1}^{N} (\bar{y}_i - \bar{y})^2 ,$$

i.e. just N individual time averages. Residuals in the *residual sum of squares* ESS also rely on the between model with N observations, and the statistic

$$1 - \frac{ESS}{TSS}$$

is maximized by the between estimator.

The overall R^2

Finally, the *overall* R^2 relies on the usual TSS definition

$$TSS = \sum_{t=1}^{T} \sum_{i=1}^{N} (y_{it} - \bar{y})^2 ,$$

and on residuals calculated without accounting for effects. It is maximized by pooled OLS.

Values for the three R^2 may be compared. For example, if within and overall R^2 are close, this is evidence for individual effects being not so important etc.