Econometric Methods fo Panel Data — Part II

Robert M. Kunst
University of Vienna

April 2009

1 Tests in panel models

Whereas restriction tests within a specific panel model follow the usual principles, based on t and F statistics, in this section we consider the tricky model selection problem of choosing among the various panel models.

1.1 Poolability test

The simplest *poolability test* has as its null hypothesis the OLS model

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

and as its alternative the FE model

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

In other words, we test for the presence of individual effects. Formally one may write

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

Intuitively, one may consider the F statistic according to the construction principle

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

Here ESS_R denotes the residual sum of squares under the null hypothesis, ESS_U the residual sum of squares under the alternative. Under H_0, the statistic F_{1-way} is distributed as F with $(N-1, (T-1)N - K)$ degrees of freedom. The two sums of squares evolve as intermediate results from OLS and from FE estimation. In most empirical panels, this test will reject its null hypothesis. If the statistic remains insignificant, no panel models need to be specified, as all individuals are sufficiently homogeneous.

Similarly, one may construct a *poolability test* with the alternative of a *two-way* model

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

where there are $N+T-2$ testable restrictions. One may also use an analogous test for the null hypothesis of an FE–*one-way* and the alternative of an FE–*two-way* model. In large samples, it is preferable to use the χ^2 approximation instead of the F distribution.

Baltagi suggests another important *poolability test*, with an even more general alternative

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

and the null hypothesis of an OLS model. Here the unrestricted model has $N(K+1)$ estimable parameters, while the restricted model has only $K+1$. The numerator degrees of freedom increase to $N(K+1) - (K+1) = (N-1)(K+1)$. If this test rejects, this indicates that *some* panel method should be used. It is unclear, however, whether modeling of the 'effects' suffices to cope with the observed heterogeneity across individuals.

Some software executes these *poolability* tests automatically. If they are not run by default but likelihoods are calculated, then two nested models can always be compared by the LR statistic (the *likelihood ratio*)

$$2(\log L_U - \log L_R).$$

This statistic is always positive, as the unrestricted model must always have the larger likelihood, and it is asymptotically distributed as χ^2 under H_0—the restricted model. The degrees of freedom correspond to the number of tested restrictions.

Table 1 may serve to summarize the diverse poolability hypotheses. *Poolability* tests allow testing all nested pairs in the table. The specification β_{it} does not make sense, as it exhausts the existing degrees of freedom. Of main interest are the test of the pooled OLS model versus the *one-way* model and also versus the *two-way* model and/or the test *one-way* versus *two-way*. Tests with alternatives in the β_i column are tests for mis-specification rather than tools for a specification search. One might also investigate whether some elements of β are constant, whereas others may depend on i.

Example. In the Baltagi/Griffin data, the statistic F_{1-way} attains the impressive value of 83.96, which implies a rejection of the null hypothesis of constant effects for all countries in the data set at every useful significance level. The test of the null of a *one-way* model versus the alternative of a *two-way* model attains a statistic of 6.23, certainly much less but still in the *tails*

Table 1: Various 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	$N(T-1)-T+1-K$	$N(T-K-1)-T+1$

of an $F(18,303)$ distribution. The LR variant for the former test yields the value of 579.68, similarly in the *tails* of the $\chi^2(17)$ distribution. For the latter test, we obtain the value of 107.747, again highly significant if compared to the $\chi^2(18)$ qunatiles. The data set contains significant time effects as well as significant individual effects. The evidence for individual effects is much stronger. These tests operate vertically within the first column of Table 1.

The other suggested tests for the *poolability* of regression coefficients work as follows. The test statistic for common coefficients *and* constant effects versus individual-specific β_i and μ_i is highly significant at 129.32. This test operates *diagonally* in Table 1. Unfortunately, even the test statistic for constant β_i in the FE model is significant at 27.34 (in Table 1 horizontally in the second row). This is unfortunate, as it means that the basic assumption of the FE model, i.e. that regression coefficients are independent of country i, is rejected, albeit not at the extreme significance levels of other tests. Even for the model with individually varying β_i the absence of time effects is rejected (vertically in the second column of Table 1). The statistically recommended model with β_i, μ_i, and λ_t is unattractive, however.

1.2 Tests for the presence of random effects

In the RE model, an obvious suggestion is to test whether there exist any individual effects μ_i or time effects λ_t. Possible null hypotheses are a pooled model without any effects or a model with only one of the two possible types of unobserved level effects. In the model

$$\begin{aligned} y_{it} &= \alpha + \beta' X_{it} + u_{it}, \\ u_{it} &= \mu_i + \lambda_t + \nu_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \end{aligned} \tag{1}$$

with the two variance parameters σ_λ^2 and σ_μ^2, the null hypothesis of the pooled model is given by $\sigma_\lambda^2 = \sigma_\mu^2 = 0$, which appears to be a quite simple restriction. Following the customary statistical likelihood-ratio (LR) principle, the likelihood can be evaluated under all hypotheses of concern, i.e. for the pooled regression and for the RE–GLS estimator. It can be shown, however, that the resulting LR statistic

$$2\left(\log L_u - \log L_r\right)$$

is not—as one may have expected—distributed as $\chi^2(2)$, because of a violation of some regularity conditions in the formulation of the hypotheses. In detail, they are violated by the fact that the null hypothesis is situated on the boundary of the alternative. Here, L_u denotes the likelihood for the unrestricted (general) model, i.e. for the RE–GLS estimator, and L_r denotes the likelihood for the restricted (specific) model, i.e. the OLS estimator. The distribution under $H_0 : \sigma_\lambda^2 = \sigma_\mu^2 = 0$ is given as

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

which is a mixture of three χ^2 distributions. Note that $\chi^2(0)$ formally denotes a χ^2 distribution with zero degrees of freedom, which means a unit mass at zero. By contrast, $\chi^2(1)$ and $\chi^2(2)$ are simply χ^2 distributions with one and two degrees of freedom. Another problem is that the evaluation of L_u requires estimating the *two-way* RE model. Unfortunately, many programs do not provide this estimate or at least they do not provide the likelihood. If one wishes to test for one of the two effects only, i.e. for $\sigma_\lambda^2 \neq 0$ *or* for $\sigma_\mu^2 \neq 0$, the distribution simplifies to

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

In any case, special tables must be consulted, or alternatively significance points are determined by simulations. The main reference is Gourieroux *et al.* (1982).

Most of these problems are avoided by the Lagrange-Multiplier (LM) test according to Breusch & Pagan. LM tests are approximations to LR tests and they only require an evaluation under the null hypothesis. Breusch & Pagan proof that the LM statistic has the simple form

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

The vector $\tilde{u}$ represents the residuals from pooled OLS estimation. One may also use LM_1 alone to test for the presence of individual effects $\sigma_\mu^2 \neq 0$ or

LM_2 to test for time effects $\sigma_\lambda^2 \neq 0$. Both terms are distributed as $\chi^2(1)$, therefore the statistic LM is distributed as $\chi^2(2)$, in all cases under $H_0 : \sigma_\lambda^2 = \sigma_\mu^2 = 0$. Therefore, the tests are not appropriate checks for the presence of time effects in an RE model with individual effects. Another drawback of this very popular test is that it is constructed as a two-sided test, as if $\sigma_\lambda^2 < 0$ were possible. Therefore, it does not attain the power properties of the LR tests. One-sided modifications have been suggested in the literature but they did not have a major impact on software programmers.

Example. The program STATA yields the statistic LM_1 on request. In our gasoline panel, LM_1 has the highly significant value of 1465.55, which is to be compared with the $\chi^2(1)$ distribution. There is strong evidence on country-specific level effects in gasoline demand. Some personal programming yields the value for LM_2 at 5.80. This is much less but still significant at the 5% level, with a p–value of 0.016.

1.3 The Hausman Test

Some textbooks introduce the Hausman test—sometimes 'Hausman-Wu test'—as another statistical testing principle, in analogy to the LR and LM tests. Then, the Hausman test of panel econometrics is only a specific application of this general principle. It is, however, an application that was mentioned explicitly by Hausman in his original contribution.

The Hausman principle can be applied to all hypothesis testing problems, in which two different estimators are available, the first of which $\hat{\beta}$ is *efficient* under the null hypothesis, however *inconsistent* under the alternative, while the other estimator $\tilde{\beta}$ is consistent under *both* hypotheses, possibly without attaining efficiency under any hypothesis. Hausman had the intuitive idea to construct a test statistic based on $q = \hat{\beta} - \tilde{\beta}$. Because of the consistency of both estimators under the null, this difference will converge to zero, while it fails to converge under the alternative. Furthermore, one may exploit the fact that the difference and $\hat{\beta}$ are uncorrelated under the null, otherwise the estimator $\hat{\beta}$ could be improved, which would contradict the assumption of efficiency.

These ideas motivate that Hausman suggested the statistic

$$m = q' (\mathrm{var} q)^{-1} q,$$

where $\mathrm{var} q = \mathrm{var}\tilde{\beta} - \mathrm{var}\hat{\beta}$ follows from the known properties of both estimators under the null hypothesis and from uncorrelatedness. The statistic m is distributed as χ^2 under the null hypothesis, with degrees of freedom corresponding to the dimension of β.

In the concrete case of panel models, we know that the FE estimator is consistent in the RE model as well as in the FE model. In the FE model it is even efficient, in the RE model it has good asymptotic properties. By contrast, the RE–GLS estimator cannot be used in the FE model, while it is efficient by construction in the RE model. The *inconsistency of the RE estimator in the FE model* follows from the fact that, as $T \to \infty$, the individual fixed effects μ_i are not estimated but are viewed as realizations of random variables with mean zero. The violation of the assumption $\mathrm{E}u = 0$ for the regression model leads to an inconsistency.

Historically, this effect was treated extensively by MUNDLAK, who interpreted the FE model as a generalized RE model with possible correlation of total error $\nu_{it} + \mu_i$ and regressor. Thus, individual heterogeneity in observed characteristics causes heterogeneous error terms. MUNDLAK's model succeeds in explaining the paradox that the FE model is more 'general' than the RE model, while one instinctively may tend to view randomness as a generalization of non-randomness.

Example. In the empirical example, the Hausman statistic for the *one-way* model is given by STATA as $m = 302.80$. This value is to be compared with a $\chi^2(3)$ distribution, it is highly significant. The recommended model is the FE model. Heterogeneity across countries cannot be interpreted as random variables with mean zero. If the individual dimension N is small, rejection by the Hausman test is the typical case.

RE or FE? Occasionally (HSIAO, 2004), the problem of discriminating FE and RE models attains an almost philosophical depth. Basically, RE appears to be attractive if the individual dimension N is large relative to the time dimension T, such that individual effects can be viewed as random. Additionally, RE offers the technical advantage that time-constant explanatory variables (covariates, attributes) can be used, while they are annulled in the LSDV procedure. In MUNDLAK's interpretation, the FE model is a model with *random effects* that correlate with the explanatory variables, while the RE model is a specific case with zero correlation. Correspondingly, many programs (e.g. STATA) provide the descriptive measure of correlation of effects and covariates in the standard output of every FE estimation. If this correlation is very different from zero, as it is typical for small individual dimension N, this indicates that one should rather use the FE–LSDV method. In these cases, it is also recommended to consult the Hausman statistic. In the data example, the correlation of effects and covariates in the *one-way* LSDV regression is given as -0.247.

2 Coefficients of determination in panel models

There is no generally accepted standard for coefficients of determination (measures of goodness of fit, R^2) in panel models. The software STATA follows a suggestion by WOOLDRIDGE (2002) and provides three different versions of R^2 with every FE estimation output:

1. R^2 *within*,

2. R^2 *between*,

3. R^2 *overall*.

R^2 *within* describes the goodness of fit for the observations that have been adjusted for their individual means. In principle, we know that R^2 is defined as a ratio

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

where ESS denotes the residual sum of squares, TSS denotes the sum of squares for the dependent variable (traditionally adjusted for a mean). For the R^2 *within*, the denominator is defined by

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

Thus, the R^2 *within* is the measure that is maximized by LSDV regression.

The R^2 *between* describes the goodness of fit for the N different individual means. Here, the denominator is simply formed from

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

while the numerator contains the N different 'residuals' that evolve from using the FE coefficients to model individual means. Because this regression problem is certainly not optimized by the FE estimator, the value for R^2 *between* will be often quite unsatisfactory. For the so called between regression, this R^2 should attain a maximum. Because of some details of definitions (see WOOLDRIDGE), the STATA values do not coincide exactly with this concept, except for the between regression itself. Usually, they are slightly larger. In the gasoline demand example, direct calculation would yield a value of 0.542, while STATA displays 0.576.

Finally, the R^2 *overall* corresponds to the usual R^2 of OLS regression. Thus, the FE estimator is fitted to a pooled regression. If the R^2 *overall* is very close to the R^2 *within*, then individual heterogeneity is low, and one obtains an indication that one could work with the pooled regression. The comparison among R^2 measures complements the formal *poolability* tests. It would not be surprising, however, if R^2 *overall* in the LSDV regression were much lower than in the pooled OLS regression, which attempts to balance out artificially existing heterogeneity. In the example, the LSDV regression attaines an R^2 *overall* of 0.615 (via direct calculation 0.592), while the R^2 of a pooled OLS regression attains 0.855. Even here, there is a sizeable discrepancy between calculated and reported values (see WOOLDRIDGE).

These R^2 can also be used in *two-way* panels in an analogous manner.

The software EViews uses the usual denominator TSS even in the FE model, while it uses for RSS a numerator that has been reduced by fitting in the individual means. This induces a relatively high value for R^2, the effects are simply interpreted as $N-1$ additional regressors. Thus, EViews attains the high value of 0.973. RE models, however, report a lower R^2, as the effects are interpreted as belonging to the errors structure. In the STATA concept, the goodness of fit measures remain comparable for RE and FE, while no R^2 at all is displayed if the RE estimation is performed using maximum-likelihood.

Acknowledgement

I appreciate the comments of Victor Waters who pointed out a numerical error in an earlier version of these lecture notes. Any further comments are definitely welcome.

References

[1] ARELLANO, M. (2003) *Panel Data Econometrics*, Oxford University Press.

[2] BALTAGI, B. (2005) *Econometric Analysis of Panel Data*, 3rd Edition, Wiley.

[3] CARRARO, C., PERACCHI, F. AND G. WEBER, G. (1993) (eds.) 'The Econometrics of Panels and Pseudo Panels', *Journal of Econometrics* **59**, 1/2, 1993.

[4] Diggle, P., Heagerty, P., Liang, K.Y., and S. Zeger (2002) *Analysis of Longitudinal Data*, 2nd Edition, Oxford University Press.

[5] Gourieroux, C., Holly, A., and A. Monfort (1982) 'Likelihood Ratio test, Wald test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters'. *Econometrica* **50**, 63–80.

[6] Hausman, J.A. (1978) 'Specification tests in econometrics'. *Econometrica* **46**, 1251–1271.

[7] Hsiao, C. (1986, 2004) *Analysis of Panel Data*, Cambridge University Press.

[8] Mundlak, Y. (1978) 'On the pooling of time series and cross section data,' *Econometrica* **46**, 69–85.

[9] Wallace, T.D., and A. Hussain (1969) 'The use of error components models in combining cross-section with time series data,' *Econometrica* **37**, 55–72.

[10] Wansbeek, T.J., and A. Kapteyn (1982) 'A class of decompositions of the variance-covariance matrix of a generalized error components model,' *Econometrica* **50**, 713–724.

[11] Wooldridge, J.M. (2002) *Econometric Analysis of Cross Section and Panel Data*. MIT Press.