Quantitative Methods
Regression models and panel data
Part III: Statistical tests for panel data models

BIg DAta MAnagement (BIDAMA) - Cycle XXVII

DIpartimento di Scienze Aziendali - Management e Innovation Systems (DISA-MIS)

Rodolfo Metulini

✉ rmetulini@unisa.it

Department of Economics and Statistics (DISES) - University of Salerno

Outline

Tests on individual and/or time effects

in order to test whether either individual or time effects are present, two approaches exist:

1. to start from estimating the model which wipes out the effects (e.g. the within) and then a restricted model (e.g. pooled OLS)
2. to start from the pooled OLS and to infer about the presence of such effects analyzing the OLS estimated residuals

1. the **F test** is based on the first approach, and it tests the presence of **fixed effects**
2. the **Breusch-Pagan** test is based on the second approach and it is devoted to test the presence of **random effects**

F test for fixed effects: one way

- One could test the joint significance of these dummies, i.e. $H_0 : \mu_1 = \mu_2 = ... = \mu_{N-1} = 0$, by performing an F-test.
- Under H_0 the statistic:

$$\frac{\hat{e}_{OLS}^T \hat{e}_{OLS} - \hat{e}_{within}^T W \hat{e}_{within}}{\hat{e}_{within}^T W \hat{e}_{within}} \frac{NT - K - N + 1}{N - 1}$$

follows a Fisher F with $N - 1$ and $NT - K - N + 1$ degrees of freedom.[1]

  - the sum of squared residuals of the within model is $\hat{e}_{within}^T W \hat{e}_{within}$ and $NT - K - N + 1$ are the degrees of freedom
  - the sum of squared residuals of the OLS model is $\hat{e}_{OLS}^T \hat{e}_{OLS}$ and $NT - K - 1$ are the degrees of freedom
  - $N - 1$ is the difference between the two degrees of freedom

[1] A difference between two χ^2 is a χ^2 with d.o.f. the difference between d.o.f. of the first and d.o.f. of the second.

F test for fixed effects: two way

- One could test the joint significance of these dummies, i.e. $H_0 : \mu_1 = \mu_2 = ... = \mu_{N-1} = 0, \lambda_1 = \lambda_2 = ... = \lambda_{T-1} = 0$, by performing an F-test.
- Under H_0 the statistic:

$$\frac{\hat{e}_{OLS}^T \hat{e}_{OLS} - \hat{e}_{2W,within}^T W^2 \hat{e}_{2W,within}}{\hat{e}_{2W,within}^T W^2 \hat{e}_{2W,within}} \frac{NT - K - N - T + 1}{N + T - 1}$$

follows a Fisher F with $N + T - 1$ and $NT - K - N - T + 1$ degrees of freedom.

  - the sum of squared residuals of the two way within model is $\hat{e}_{2W,within}^T W \hat{e}_{2W,within}$ and $NT - K - N - T + 1$ are the degrees of freedom

F test for fixed effects: conditional

- One could test the significance of the time dummies, conditional to the presence of individual dummies, i.e. $H_0 : \lambda_1 = \lambda_2 = ... = \lambda_{N-1} = 0$ with $\mu_i \neq 0, i = 1, ..., T-1$, by performing an F-test.
- Under H_0 the statistic:

$$\frac{\hat{e}_{within}^T W \hat{e}_{within} - \hat{e}_{2W,within}^T W^2 \hat{e}_{2W,within}}{\hat{e}_{2W,within}^T W^2 \hat{e}_{2W,within}} \frac{NT-K-N-T+1}{T-1}$$

follows a Fisher F with $T-1$ and $NT-K-N-T+1$ degrees of freedom.

Lagrange Multiplier (LM) test

- Breusch-Pagan (BP, 1980) test is a Lagrange multipliers test based on restricted model residuals
- In its original formulation, LM considers the score vector $g(\theta) = \frac{\delta lnL}{\delta\theta}$ (i.e. the vector of the partial derivatives of the log-likelihood function from a restricted model)[2] and the variance of the score vector, which is the information matrix $I(\theta) = E(-\frac{\delta lnL}{\delta\theta\delta\theta^T})(\theta)$
- Under the H_0: "the restricted is the true model", $g(\hat{\theta}) \sim N(0, V(\hat{\theta}))$
- by denoting with $\hat{g}$ and $\hat{V}$ the score and the estimated variance in the restricted model:

$$\hat{g}^T \hat{V}^{-1} \hat{g} \sim \chi^2_n,$$

where n is the number of restrictions.

[2] Intuitively, if the restricted estimator is near the maximum of the likelihood function, the score should be close to 0.

BP test for random effects (i)

- To test for the absence of individual (random) effect we impose a null with the following constraint: $H_0 : \sigma^2_\mu = 0$
- In this case, the estimator for the parameters $\hat{\theta}$ is the OLS and $\hat{e}^T_{OLS}\hat{e}_{OLS}/NT$ is the estimator for $\hat{\sigma}^2_\upsilon$ (since $\sigma^2_\mu = 0$)
- The statistic $LM_\mu = \frac{NT}{2(T-1)}(T\frac{\hat{e}^T_{OLS}B_\mu\hat{e}_{OLS}}{\hat{e}^T_{OLS}\hat{e}_{OLS}} - 1)^2$ asymptotically distributes as a χ^2 with 1 degree of freedom (only 1 constraint, $\sigma^2_\mu = 0$)
- To reject (or not) the null hypothesis, we compare the LM_μ terms with the one on χ^2 table.

BP test for random effects (ii)

- Likewise, the test for the absence of time (random) effect impose the null $H_0 : \sigma^2_\lambda = 0$. The estimator for the parameters $(\hat{\theta})$ is the OLS and $\hat{e}^T_{OLS}\hat{e}_{OLS}/NT$ is the estimator for $\hat{\sigma}^2_\upsilon$ (since $\sigma^2_\lambda = 0$)
- The statistic $LM_\lambda = \frac{NT}{2(N-1)}(N\frac{\hat{e}^T_{OLS}B_\lambda\hat{e}_{OLS}}{\hat{e}^T_{OLS}\hat{e}_{OLS}} - 1)^2$ asymptotically distributes as a χ^2 with 1 degree of freedom.
- The BP test extends to the two-ways error component model and the related statistic is computed as the sum of the previous statistics: $LM_{\mu\lambda} = LM_\mu + LM_\lambda$ and distributes as a χ^2 with 2 degrees, under the null that both individual and time (random) effects are absent.

Application

Application in R: example 4.1 (Croissant, Millo)

Tests for correlated effects

- Let assume $E(X^T \upsilon) = 0$
- A critical assumption in error component regression model is that $E(X^T \mu) = 0$
- If both (μ and υ) the model errors are not correlated with the explanatory variables, both fixed effects (within) and random effects (GLS) estimators are consistent, but:
    1. if $E(X^T \mu) = 0$: both estimators are consistent, random effects is more efficient than fixed effects
    2. if $E(X^T \mu) \neq 0$: just the within fixed effect is consistent (since it wipes out μ).

The Hausman test

- Given π a quantity measuring the relation between X and μ, the Hausman test (1978) is based on testing null hypotesis $H_0 : \pi = 0$ (i.e., uncorrelation)
- We have that $\hat{\pi}^T \hat{V}(\hat{\pi})^{-1}\hat{\pi}$ is distributed as a χ^2 with K degrees of freedom, where $\hat{\pi} = \hat{\beta}_A - \hat{\beta}_B$ and $V(\hat{\pi}) = V(\beta_A) + V(\beta_B)$
- Principle: to compare models A and B, where:
  1. under H_0: A and B are both consistent, but B more efficient than A
  2. under H_1: only A is consistent
- In our case,"A" is the within, "B" is the GLS. $\hat{\pi} = \hat{\beta}_{within} - \hat{\beta}_{GLS}$.

Application

Application in R: example 4.2 (Croissant, Millo)

Poolability

- The question of whether to pool the data or not naturally arises with panel data.
- The restricted model is the pooled model given by OLS representing a behavioural equation with the same parameters over time and across individuals (just one β, constant along all the sample observations).
- The unrestricted model, however, is the same behavioural equation but with different parameters across individuals (or across time).

Poolability: examples

- Balestra and Nerlove (1966) considered a dynamic demand equation for natural gas across 36 states over 6 years. The question is whether the parameters of this demand equation vary from one year to the other.
- Baltagi and Griffin (1983) considered panel data on motor gasoline demand for 18 OECD countries. They were interested in testing whether the behavioral relationship predicting demand is the same across OECD countries
- Holly et al. (2010) considered whether the effect of income on house prices in 49 US states using a T=28 years' panel is varying along states

Chow test - poolability on individuals

- the null hypotesis H_0 is that all β's are constant along individuals.
- unrestricted estimation consists in estimating (by OLS) one different model for each individual (a total of N models)
- the sum of squares of the residuals for the N (unrestricted) models (stacked) is $\hat{e}_N^T \hat{e}_N$. The degrees of freedom are $NT - NK - N$, because, totally, $NK + N$ parameters need to be estimated[3].
- The restricted model can be either the OLS ($\hat{e}_{OLS}^T \hat{e}_{OLS}$, $NT - K - 1$) or the within ($\hat{e}_{within}^T \hat{e}_{within}$, $NT - N - K$)
- The statistic

$$\frac{\hat{e}_{within}^T \hat{e}_{within} - \hat{e}_N^T \hat{e}_N}{\hat{e}_N^T \hat{e}_N} \frac{N(T-K-1)}{(N-1)K}$$

is distributed as a F with $(N-1)K$ and $N(T-K-1)$ degrees of freedom (Chow, 1960).

[3] NK different slopes and N different intercepts.

Application

Application in R: example 8.4 (Croissant, Millo)

References

- Balestra, P., & Nerlove, M. (1966). Pooling cross section and time series data in the estimation of a dynamic model: The demand for natural gas. Econometrica: Jrn. of the econometric society, 585-612.
- Baltagi, B. H., & Griffin, J. M. (1983). Gasoline demand in the OECD: An application of pooling and testing procedures. European Economic Review, 22(2), 117-137.
- Breusch, T. S., & Pagan, A. R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. The review of economic studies, 47(1), 239-253.
- Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. Econometrica: Journal of the Econom. Society, 591-605.
- Feng, Q., & Horrace, W. C. (2012). Alternative technical efficiency measures: skew, bias and scale. Jrn. of Appl. Econometr., 27(2), 253-268.
- Hausman, J. A. (1978). Specification tests in econometrics. Econometrica: Jrn. of the econometric society, 1251-1271.
- Holly, S., Pesaran, M. H., & Yamagata, T. (2010). A spatio-temporal model of house prices in the USA. Journal of Econometrics, 158(1), 160-173.
- Honda, Y. (1985). Testing the error components model with non-normal disturbances. The Review of Economic Studies, 52(4), 681-690.

- Maximum Likelihood Estimation (MLE) is a method of estimating the parameters of a distribution, given observed data.
- This is achieved by maximizing a likelihood function so that, under the assumed statistical model, the observed dataset is the most probable.
- Given the vector $\theta = [\theta_1, \theta_2, ..., \theta_k]^T$, the problem to solve is that of find an estimate for θ maximizing the joint density at the observed data sample $y = (y_1, y_2, ..., y_n)$:
  $\hat{\theta} = argmax_{\theta \in \Theta} \hat{L}_n(\theta, y)$,
  where $L_n(\theta) = L_n(\theta, y) = f_n(Y, \theta)$ is the likelihood function (and $lnL_n(\theta)$ is the log likelihood function).
- The method consists on let to 0 the partial first derivative of the distribution function f_n with respect to each element of θ $(\frac{\delta lnL}{\delta \theta})$ (e.g. in the simple linear regression model with β_0 and β_1 the solution is a system of two equations).

Go back to slide