One-way error component models. Hausman test.

Jakub Mućk
SGH Warsaw School of Economics

One-way error component model

	Random effects	Fixed effects
Individual effects	$\mu_i \sim \mathcal{N}\left(0, \sigma^2_\mu\right)$ drawn from the random sample $\Longrightarrow$ we can estimate the parameter of distribution, i.e., σ^2_μ	α_i α_i are assumed to be constant over time
Assumptions:	(i) $\mathbb{E}\left(\mu_i \vert \varepsilon_{it}\right) = 0$ (ii) $\mathbb{E}\left(\mu_i \vert x_{it}\right) = 0$ individual effects are independent of the explanatory variable x_{it}	(i) $\mathbb{E}\left(\alpha_i \vert \varepsilon_{it}\right) = 0$
Estimation	GLS	OLS (within or LSDV)
Efficiency	higher	lower
Additional:		impossible to use time invariant regressors (collinearity with α_i)

- Random and Fixed effects models:

$$\begin{aligned} \text{RE:} \quad y_{it} &= \alpha + \beta_1 x_{1it} + \ldots + \beta_1 x_{kit} + \mu_i + \varepsilon_{it}, \\ \text{FE:} \quad y_{it} &= \alpha_i + \beta_1 x_{1it} + \ldots + \beta_1 x_{kit} + u_{it} \end{aligned}$$

  where $\mu_i \sim \mathcal{N}(0, \sigma_\mu^2)$, $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ and $u_{i,t} \sim \mathcal{N}(0, \sigma_u^2)$.
- We can test significance of individual effects:
  - **The random effect model:** the Langrange multiplier statistic:

$$\mathcal{H}_0: \quad \sigma_\mu^2 = 0$$

  - **The fixed effect model:** test for poolability:

$$\mathcal{H}_0: \quad \alpha = \alpha_1 = \ldots = \alpha_N.$$

- Can we compare the FE and RE models?

- **Cobb-Douglas production function**. Consider the following model:

$$y_{it} = \beta x_{it} + u_{it} + \eta_i \tag{1}$$

  - y_{it} – the log output.
  - x_{it} – the log of a variable input;
  - η_i – an firm input that is constant over time, e.g., managerial skills.
  - u_{it} – a stochastic input that is outside framer's control.
  - β - the technological parameter.
- η_i is the **unobserved** individual effect.
- Let's assume that the input (x_{it}) depends on the quality of managerial skills (individuals effects η_i)
  $\implies \mathbb{E}(\eta_i | x_{it}) \neq 0$
  $\implies$ **the RE estimator is not consistent.**

- The null $\mathcal{H}_0$ in the Hausman test is that **both the random and fixed effect estimates are consistent**.
- If the alternative hypothesis $\mathcal{H}_1$ holds then **the random effect estimates are inconsistent**.
- Test statistics:

$$H = \left[\hat{\beta}^{FE} - \hat{\beta}^{RE}\right]' \left(Var\hat{\beta}^{FE} - Var\hat{\beta}^{RE}\right)^{-1} \left[\hat{\beta}^{FE} - \hat{\beta}^{RE}\right] \tag{2}$$

- The statistics H is distributed χ^2 with degrees of freedom determined by K, i.e., the dimension of the coefficient vector β.

Hausman Test- some remarks

SGH

- In general, the Hausman test asks whether the fixed effects and random effects estimates of β are **significantly** different.
- We can test only models with the same set of explanatory variables:
  - We cannot compare the random effects estimates corresponding to time-invariant regressors due to their collinearity with individual intercept.
- The rejection of the null hypothesis indicates that the random effect estimates of β are not consistent or **that the model is wrongly specified (misspecification error)**.
- It is assumed that the fixed effect model is consistent under both null and alternative.
  - What if regressors are not strictly exogenous?
- The Hausman may be used in more general context.

Between Estimator

- **The between estimator** uses **just cross-sectional variation**.
- Averaging over time yields:

$$\bar{y}_i = \alpha^{Between} + \beta_1^{Between}\bar{x}_{1i} + \ldots + \beta_k^{Between}\bar{x}_{ki} + \bar{u}_i, \tag{3}$$

where $\bar{y}_i = T^{-1}\sum_t y_{it}$, $\bar{x}_{1i} = T^{-1}\sum_t x_{1it}$, …, $\bar{x}_{ki} = T^{-1}\sum_t x_{kit}$, $\bar{u}_i = T^{-1}\sum_t u_{it}$.

- or in the matrix form:

$$\bar{y} = \alpha^{Between} + \bar{x}\beta^{Between} + u. \tag{4}$$

- The parameters $\alpha^{Between}$, $\beta_1^{Between}$, …, $\beta_k^{Between}$ can be estimated with the OLS estimator.

Coefficient of determination (R^2)

- **Coefficient of determination (R^2):** is an indicator showing how well data fit a statistical model.
- General definition:

$$R^2 = 1 - \frac{ESS}{TSS}, \tag{5}$$

where

  - ESS - is the residuals sum of squares:

$$ESS = \sum_i (\hat{y}_i - y)^2, \tag{6}$$

  - TSS - is the total (observed) sum of square,

$$TSS = \sum_i (y_i - \bar{y})^2. \tag{7}$$

- R^2 is between 0 and 1.
- In the context of panel data, the conventional coefficient of determination is maximized by **the pooled estimator**.

- **The within R^2:**

$$R^2_{Within} = 1 - \frac{ESS_{Within}}{TSS_{Within}}, \quad (8)$$

where

- ESS_{Within} - is the residuals sum of squares:

$$ESS_{Within} = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(\hat{y_{it}} - y_{it}\right)^2, \quad (9)$$

- TSS_{Within} - is the total (observed) sum of square,

$$TSS_{Within} = \sum_{i=1}^{N}\sum_{t=1}^{T}\left(y_{it} - \bar{y}_i\right)^2. \quad (10)$$

- The within R^2 is maximized by the within estimator.

- **The between R^2:**

$$R^2_{Between} = 1 - \frac{ESS_{Between}}{TSS_{Between}}, \tag{11}$$

where

- $ESS_{Between}$ - is the residuals sum of squares:

$$ESS_{Between} = \sum_{i=1}^{N} \left(\hat{\bar{y}}_i - \bar{y}_i\right)^2, \tag{12}$$

- $TSS_{Between}$ - is the total (observed) sum of square,

$$TSS_{Between} = \sum_{i=1}^{N} \left(\bar{y}_i - \bar{y}\right)^2. \tag{13}$$

- The between R^2 is maximized by the between estimator.

Testing linear hypotheses

Testing linear hypotheses

- Statistical models are usually formulated in order to verify testable implications of the economic theory.
- A set of linear hypotheses can be described as follows:

$$R\beta = q, \tag{14}$$

where β is the vector of parameters, the matrix R and vector q describe the linear constraints.

- R is $m \times (K+1)$ and q is $m \times 1$ when m is the number of linear constraints.
- Example. Consider the case when $k = 4$ and the following three hypotheses:

$$\begin{aligned} \beta_1 + \beta_3 &= 1 \\ \beta_2 - \beta_4 &= 0 \\ \beta_1 + 0.5\beta_4 &= 2 \end{aligned}$$

then the matrix R and vector q:

$$R = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 1 & 0 & 0 & 0.5 \end{bmatrix} \quad \text{and} \quad q = \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix}.$$

- **Restricted Least Squares (RLS) estimator:**

$$\hat{\beta}^{RLS} = \hat{\beta}^{OLS} + \left(X'X\right)^{-1} R' \left(R\left(X'X\right)^{-1} R'\right)^{-1} \left(q - R\hat{\beta}^{OLS}\right), \qquad (15)$$

where $\hat{\beta}^{OLS}$ is the OLS estimator.

- **Restricted Least Squares (RLS) estimator:**

$$\hat{\beta}^{RLS} = \hat{\beta}^{OLS} + \left(X'X\right)^{-1} R' \left(R \left(X'X\right)^{-1} R'\right)^{-1} \left(q - R\hat{\beta}^{OLS}\right), \tag{15}$$

where $\hat{\beta}^{OLS}$ is the OLS estimator.

- **Wald test:**

$$\begin{aligned} \mathcal{H}_0 : \quad R\beta &= q \\ \mathcal{H}_1 : \quad R\beta &\neq q \end{aligned}$$

- Test statistics for m linear hypotheses:

$$\mathcal{F} = \frac{(R\beta - q)' \left(R \left(X'X\right)^{-1} R'\right)^{-1} (R\beta - q)}{m} \tag{16}$$

- if true then $\mathcal{F} \sim \mathcal{F}(m, NT - K)$ where K is the number of regressors and NT refers to the number of observations.
- Alternatively:

$$\mathcal{W} = (R\beta - q)' \left(R \left(X'X\right)^{-1} R'\right)^{-1} (R\beta - q) = m\mathcal{F} \tag{17}$$

- The $\mathcal{W}$ is χ^2 distributed with m degree of freedoms.

- The rank of the matrix R should equal the number of restrictions (m).
- The test statistics $\mathcal{F}$ is preferable in smaller samples.
- The test statistics, both $\mathcal{F}$ and $\mathcal{W}$, are sometimes very sensitive to the estimates of the variance-covariance of the error terms.
  - Viewed from the empirical perspective, it's essentially to investigate whether the obtained results are not sensitive to a choice of the variance-covariance estimator.
- Alternative procedures can be used here, e.g. the Lagrange multiplier test (*score test*) or the likelihood-ratio test.
  - Engle (1983) shows that these tests (Wald, score & LR) are asymptotically equivalent but in finite samples they can lead to different conclusions.