Quantitative Methods
Regression models and panel data
Part II: 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

Further examples

$$y_{nt} = \alpha + x_{nt}^T \beta + \mu_n + \upsilon_{nt}$$

1. Earnings equation in labour economics: y_{nt} measures earnings of the head of the household, x_{nt} contains a set of variables like experience, education, union membership, sex, race, etc. Note that μ_n is time-invariant and it accounts for any individual specific effect that is not included in the regression. In this case we could think of it as the individual's unobserved ability. The remainder disturbance υ_{nt} varies with individuals and time and can be thought of as the usual disturbance in the regression.
2. Production function (utilizing data on firms across time): y_{nt} measures output and x_{nt} measures inputs. The unobservable firm specific effects will be captured by the μ_n and we can think of these as the unobservable entrepreneurial or managerial skills of the firm's executives.

The model (i)

- For the observation of individual n at period t we can write the most simple version of the panel model as:

$$y_{nt} = \alpha + x_{nt}^{T}\beta + e_{nt} \tag{1}$$

or, equivalently:

$$y_{nt} = z_{nt}^{T}\gamma + e_{nt} \tag{2}$$

- The error component is the sum of two effects: μ_n, which is the effect for individual n; υ_{nt} is the residual (or idiosyncratic) effect: $e_{nt} = \mu_n + \upsilon_{nt}$

The model (ii)

- For the whole sample of $O = NT$ observations we can use the following matrix notation:

$$y = \alpha j + X\beta + e \tag{3}$$

or, equivalently:

$$y = Z\gamma + e \tag{4}$$

where j is a $NTx1$ vector of ones, y is a vector of length NT, X is a matrix of dimension $NTxK$, Z is $NTx(K+1)$, β is a vector of length K and γ a vector of length $K+1$.

The model (iii)

$$
y = \begin{pmatrix} y_{11} \\ y_{12} \\ .. \\ y_{1T} \\ y_{21} \\ y_{22} \\ .. \\ y_{2T} \\ .. \\ y_{N1} \\ y_{N2} \\ .. \\ y_{NT} \end{pmatrix}
Z = \begin{pmatrix}
1 & x_{11}^1 & x_{11}^2 & \dots & x_{11}^K \\
1 & x_{12}^1 & x_{12}^2 & \dots & x_{12}^K \\
1 & .. & .. & .. & .. \\
1 & x_{1T}^1 & x_{1T}^2 & \dots & x_{1T}^K \\
1 & x_{21}^1 & x_{21}^2 & \dots & x_{21}^K \\
1 & x_{22}^1 & x_{22}^2 & \dots & x_{22}^K \\
1 & .. & .. & .. & .. \\
1 & x_{2T}^1 & x_{2T}^2 & \dots & x_{2T}^K \\
1 & .. & .. & .. & .. \\
1 & x_{N1}^1 & x_{N1}^2 & \dots & x_{N1}^K \\
1 & x_{N2}^1 & x_{N2}^2 & \dots & x_{N2}^K \\
1 & .. & .. & .. & .. \\
1 & x_{NT}^1 & x_{NT}^2 & \dots & x_{NT}^K
\end{pmatrix}
\gamma = \begin{pmatrix} \alpha \\ \beta_1 \\ .. \\ \beta_K \end{pmatrix}
e = \begin{pmatrix} e_{11} \\ e_{12} \\ .. \\ e_{1T} \\ e_{21} \\ e_{22} \\ .. \\ e_{2T} \\ .. \\ e_{N1} \\ e_{N2} \\ .. \\ e_{NT} \end{pmatrix}
$$

Transformation

- Panel data models break the total variation of the dependent variable y in intra- and inter-individual.
- The inter-individual transformation (called *between*) is obtained by post multiplying $B = I_N \otimes J_T/T$ with y (x).
  - For example, for one variable, we will have $(By)^T = (\bar{y}_{1\cdot}, \bar{y}_{1\cdot}, ..., \bar{y}_{1\cdot}, \bar{y}_{2\cdot}, \bar{y}_{2\cdot}, ..., \bar{y}_{2\cdot}, ..., \bar{y}_{N\cdot}, \bar{y}_{N\cdot}, ..., \bar{y}_{N\cdot})$
- The intra-individual transformation (called *within*) is obtained by post multiplying $W = I_{NT} - B$ with y (x)
  - For example, for one variable, we will have $(Wy)^T = (y_{11} - \bar{y}_{1\cdot}, y_{12} - \bar{y}_{1\cdot}, ..., y_{1T} - \bar{y}_{1\cdot}, y_{21} - \bar{y}_{2\cdot}, y_{22} - \bar{y}_{2\cdot}, ..., y_{2T} - \bar{y}_{2\cdot}, ..., y_{N1} - \bar{y}_{N\cdot}, y_{N2} - \bar{y}_{N\cdot}, ..., y_{NT} - \bar{y}_{N\cdot})$
- W and B are:
  1. symmetric, so $B^T = B$ and $W^T = W$
  2. idempotent, so $WxW = W$ and $BxB = B$
  3. They decompose a vector, so $Bxy + Wxy = y$
  4. orthogonal, so $W^TxB = 0$

Random effect's errors assumptions

- the errors, in matrix form, can be written as $e = (I_N \otimes J_T)\mu + \upsilon$, where μ is a vector of individual effects μ_n of length N where each element is repeated T times, υ is a NT-length vector with idiosyncratic terms.
- The estimated model is defined by estimated parameters $\hat{\gamma} = (\hat{\alpha}, \hat{\beta}^T)^T$ and the NT vector of residuals $\hat{e} = e - Z(\hat{\gamma} - \gamma)$
- Assumptions:
  1. the expected value of μ_n and υ_{nt} is 0
  2. the effects μ_n are mutually uncorrelated ($E(\mu_n, \mu_m) = 0$, $\forall m \neq n$) and homoschedastic ($E(\mu_n^2) = \sigma_\mu^2, \forall n = 1, ..., N$)
  3. the idiosyncratic terms υ_{nt} are also mutually uncorrelated and homoschedastic
  4. υ_{nt} and μ_n are uncorrelated each others
- It follows that:
  1. the variance $E(e_{nt}^2) = \sigma_\mu^2 + \sigma_\upsilon^2$
  2. covariance $E(e_{nt}e_{ns}), s \neq t = \sigma_\mu^2$
  3. covariance $E(e_{nt}e_{mt}), n \neq m = 0$

The covariance matrix

- For a given individual n, the variance-covariance matrix for $e_n^T = (e_{n1}, e_{n2}, .., e_{nT})$ is

$$\Omega_{nn} = E(e_n e_n^T) = \sigma_\upsilon^2 I_T + \sigma_\mu^2 J_T \tag{5}$$

- for $n \neq m$ the covariance matrix is 0, given the assumptions of mutual uncorrelation. It follows that:
- The full (for all individuals) variance-covariance matrix Ω is a N-block diagonal matrix, each one is in the form of equation 5:

$$\Omega = \sigma_\upsilon^2 I_{NT} + \sigma_\mu^2 (I_N \otimes J_T)$$

- $\Omega = \sigma_\upsilon^2 (B + W) + T\sigma_\mu^2 B = \sigma_\upsilon^2 W + \sigma_1^2 B$ (the var-covar matrix can be represented as a linear combination of var terms with weights the between and the within matrix).[1]
- A desirable assumption is that both μ and υ are uncorrelated with x, that is $E(\mu \mid x) = 0$ and $E(\upsilon \mid x) = 0$

[1] where $\sigma_1^2 = \sigma_\upsilon^2 + T\sigma_\mu^2$

The covariance matrix - An example

Let N = 2 and T = 2 (a very simple case)

$$\Omega_{n=1,n=1} = \begin{pmatrix} \sigma^2_\upsilon & 0 \\ 0 & \sigma^2_\upsilon \end{pmatrix} + \begin{pmatrix} \sigma^2_\mu & \sigma^2_\mu \\ \sigma^2_\mu & \sigma^2_\mu \end{pmatrix} = \begin{pmatrix} \sigma^2_\mu + \sigma^2_\upsilon & \sigma^2_\mu \\ \sigma^2_\mu & \sigma^2_\mu + \sigma^2_\upsilon \end{pmatrix}$$

$$\Omega = \begin{pmatrix} \Omega_{n=1,n=1} & \mathbf{0} \\ \mathbf{0} & \Omega_{n=2,n=2} \end{pmatrix}$$

We have obtained a 4 x 4 var-covar matrix where elements belonging to a different individual are zero, and elements belonging to the same individual at different time are σ^2_μ

OLS estimators (fixed effects)

- The variability in a panel has two sources:
  1. *between*, or inter-individual, which is the variability of panel variables measured in individual means which is $\bar{y}_n$ (or By)
  2. *within*, or intra individual, which is the variability of panel variables measured in deviation from the means (i.e., $y_{nt} - \bar{y}_n$, or Wy)
- OLS method of estimation (the same adopted in the cross section linear regression) can be applied to 3 models: the *raw data* one (pooling model, as it was a cross-section), the *within* and the *between*

On raw data: The *pooling* model

- The model to be estimated is $y = Z\gamma + e$.[2]
- OLS is based on minimizing $(y^T - \gamma^T Z^T)(y - Z\gamma) = \hat{e}^T\hat{e}$
- the OLS estimator is

$$\hat{\gamma}_{OLS} = (Z^T Z)^{-1} Z^T y \tag{6}$$

  while the difference between the vector of true parameters and estimated parameters is $\hat{\gamma}_{OLS} - \gamma = (Z^T Z)^{-1} Z^T e$
- About β, $\hat{\beta}_{OLS} = (X^T(I - \bar{J})X)^{-1} X^T(I - \bar{J})y$ with $E(\hat{\beta}) = \beta$ (unbiased) only if the error e and X are uncorrelated (it may be not, if unobserved hetrogeneity exists!)
- $V(\hat{\gamma}_{OLS}) = (Z^T Z)^{-1} Z^T \Omega Z (Z^T Z)^{-1} \neq \sigma^2 (Z^T Z)^{-1}$
- So, OLS is unbiased and consistent only if e and X are uncorrelated, but:
  1. the expression for the variance is complex
  2. OLS is not the BLUE: it exist at least one estimator which is unbiased and more efficient.

[2] e here is not the sum of μ and υ

The between estimator

- Between estimator is an OLS applied to the between transformed model (it does not allow to detect intra-individual variation)
$$By = BZ\gamma + Be = \alpha j + BX\beta + Be$$
- The variables (such as, e.g., gender, religion) in the model that do not exhibit intra-individual variation are **unaffected** by this transformation
- The NT observations become in fact N distinct observations with individual mean repeated T times each.
- The between estimator reads as:
$$\hat{\beta}_B = (X^T \bar{B} X)^{-1} X^T \bar{B} y, \tag{7}$$
where $\bar{B} = B - \bar{J}$ is the matrix that transforms the variable in its mean into a deviation from the overall mean
- $V(\hat{\beta}_B) = \sigma_1^2 (X^T \bar{B} X)^{-1}$

The within estimator

- Within estimator is an OLS applied to the within transformed model (μ wipes out)

$$Wy = W(\alpha j + X\beta + e) = WX\beta + W\upsilon$$

- This transformation removes the vector of ones associated to the intercept α but also the matrix associated to the vectors of individual effects μ_n.
- It also **removes** regressors without intra-individual variation
- Applying OLS to the transformed model we have the within estimator:

$$\hat{\beta}_W = (X^T WX)^{-1} X^T Wy \tag{8}$$

with standard expression for the variance (no μ_n)
$V(\hat{\beta}_W) = (X^T WX)^{-1} X^T W \Omega WX (X^T WX)^{-1} = \sigma^2_\upsilon (X^T WX)^{-1}$

Least squares dummy variables (LSDV)

- The model works on raw data and can be written as:

$$y = X\beta + (I_N \otimes J_T)\mu + \upsilon$$

  where μ is the N-length vector of parameters (each repeated T times) to be estimated (μ is now taken into account in a deterministic way)

- Therefore, there are $N + K$ parameters and the model is feasible only if N is not too large compare to T
- β coefficients estimated with within or LSDV are the same (Frisch-Waugh theorem)
- However, LSDV directly estimates the vector of parameters μ while the within just the βs
  - In the *within*, μ_n are estimated in a separate step: $\hat{\alpha}_n = \bar{y}_{n.} - \bar{x}_{n.}^T\hat{\beta}$, $\hat{\alpha} = \bar{\bar{y}} - \bar{\bar{x}}^T\hat{\beta}$, $\hat{\mu}_n = \hat{\alpha}_n - \hat{\alpha}$

Application

Application in R: Example 2.1 (Croissant, Millo)

GLS estimator (i)

- The within estimator is a regression on panel data that have been transformed so that individual effects vanish while LSDV estimates all the effects, but the risk is of having too many coefficients to be estimated.
- On the contrary, GLS consider the individual effect as a random (stochastic) draw from a specific distribution (i.e., normal) and seeks to estimate the parameters (i.e. the variance) of this distribution.
- In all cases, the aim is the same: to obtain an efficient estimate for the slopes (i.e., the β parameters).
- Assume $\mu_n \sim iid(0, \sigma^2_\mu)$, $\upsilon_{nt} \sim iid(0, \sigma^2_\upsilon)$ and the μ_n are independent of the υ_{nt}.
- The random effects model is an appropriate specification if we randomly draw N individuals from a large population where N is large. LSDV would lead to an enormous loss of degrees of freedom.

GLS estimator (ii)

- Assuming independence between the error terms μ_n and υ_{nt}, the covariance matrix is

$$\Omega = E(ee^T) = \sigma_\mu^2(I_N \otimes J_T) + \sigma_\upsilon^2(I_N \otimes I_T)$$

- This implies a homoskedastic variance

$$var(e_{nt}) = \sigma_\mu^2 + \sigma_\upsilon^2, \forall n, t$$

- Ω is equicorrelated and block-diagonal and it exhibits serial correlation over time only between the same individual. In fact,

$$cov(e_{nt}, e_{ms}) = \begin{cases} \sigma_\mu^2 + \sigma_\upsilon^2 & \text{if } n = m, s = t \\ \sigma_\mu^2 & \text{if } n = m, s \neq t \\ 0 & \text{otherwise} \end{cases}$$

GLS estimator (iii)

- The GLS estimator reads as:

$$\hat{\gamma}_{GLS} = (Z^T \Omega^{-1} Z)^{-1} (Z^T \Omega^{-1} y) \quad (9)$$

- $V(\hat{\gamma}_{GLS}) = (X^T \Omega^{-1} X)^{-1}$ [3]
- Ω is also linear combination of two well-known idempotent and orthogonal matrices: $\Omega = \sigma_1^2 B + \sigma_\upsilon^2 W$
- Since the dimension of Ω depends on the sample, may be infeasible to estimate $\hat{\gamma}_{GLS}$ with eq. 9 (it requires the inversion of Ω).
  - An efficient way is to estimate OLS on pre transformed data.
  - Let C such that $C^T C = \Omega^{-1}$, $\tilde{y} = Cy$, $\tilde{Z} = CZ$, then $\hat{\gamma} = (\tilde{Z}^T \tilde{Z})^{-1} (\tilde{Z}^T \tilde{y})$
  - $C = \Omega^{-0.5} = \frac{1}{\sigma_1} B + \frac{1}{\sigma_\upsilon} W$, with σ_1 and σ_υ that need to be estimated in advance.

[3] The constant term has no effect on the variance.

GLS estimator (iv)

- we can denote $\phi = \frac{\sigma_\upsilon}{\sigma_1}$.

  It follows that $\theta = 1 - \phi = 1 - \sqrt{\frac{\sigma_\upsilon^2}{T\sigma_\mu^2 + \sigma_\upsilon^2}}$
- As Ω depends on W and B, it may be clear that GLS produces results in between the within (is equal to the within when σ_μ dominates, $\theta = 1$) and the pooled OLS (is equal to the pooled OLS when σ_υ dominates, $\phi = 1$)

Application

Application in R: Example 2.2 (Croissant, Millo)

Comparison of the estimators

- We have four different estimators for the same model (the one-way error component model): the within and the between just use one source of variance (respectively, intra and inter-individual), the pooled OLS and the GLS use both sources.
- If the assumption on **null correlation between the errors and the regressors** is in place, **all estimators are consistent** and unbiased and then they will give similar results

Relations between estimators (i)

- OLS and GLS gives intermediate results between the *within* and the *between* estimators, as they use both sources of variance.
- It is in fact demonstrated that $\hat{\beta}_{GLS}$ can be expressed as a weighted average of the *within* and the *between*

$$\hat{\beta}_{GLS} = (X^T W X + \phi^2 X^T \bar{B} X)^{-1} (X^T W X \hat{\beta}_W + \phi^2 X^T \bar{B} X \hat{\beta}_B)$$

- Similarly with $\hat{\beta}_{OLS}$, which is the GLS when $\phi = \phi^2 = 1$

$$\hat{\beta}_{OLS} = (X^T W X + X^T \bar{B} X)^{-1} (X^T W X \hat{\beta}_W + X^T \bar{B} X \hat{\beta}_B)$$

- For OLS the weights are exactly the shares of the intra- and inter-individual variance of the regressors.
- For GLS, the weights depends also on the variance of the errors.

Relations between estimators (ii)

- Because $\phi <= 1$, GLS will gives less weight to the between variation, compared to OLS.
- Two special cases can happen:
  1. $\phi \rightarrow 0$: this means that σ_v is very small compared to σ_μ. In this case GLS converges to the within estimator.
  2. $\phi \rightarrow 1$: this means that σ_v is very large compared to σ_μ. In this case GLS converges to the OLS (equal weight to within and between)

Fixed vs. random effects

- The individual effects are not fixed or random by nature and they can be modelled either as random or fixed depending on
  1. the purpose of estimation
  2. the probabilistic structure (of errors)
  3. the correlation between the errors and the regressors
- in micro data, generally random effects are preferred (individuals are randomly drawn from a population), while in macro analysis, fixed effects are more popular (because the individual effects may be of interest per se)
- Assumption of uncorrelated effects (given $E(X^T \upsilon) = 0$). Two situations:
  1. $E(X^T \mu) = 0$: individual effects are not correlated with X. Both models (within and GLS) are consistent, but random effects (GLS) is more efficient than fixed effects (within)
  2. $E(X^T \mu) \neq 0$: individual effects are correlated with X. Only the fixed effects (within) method gives consistent estimates as, with the within transformation, the individual effects vanish.

Application

Application in R: Example 2.3 (Croissant, Millo)

The two ways (error) model (i)

- The two ways error component is obtained by adding a time invariant effect λ_t to the model:

$$y_{nt} = \alpha + x_{nt}^T \beta + e_{nt} \quad (10)$$
$$e_{nt} = \mu_n + \lambda_t + \upsilon_{nt} \quad (11)$$

- Assumptions for the two way error model[4]:
  1. λ has zero mean and it is homoschedastic with variance σ^2_λ
  2. time effects are mutually uncorrelated: $E(\lambda_t \lambda_s) = 0, \forall t \neq s$
  3. time effects are uncorrelated with the individual effects and with the idiosyncratic terms

- The variance-covariance matrix is $\Omega = \sigma^2_\upsilon I_{NT} + \sigma^2_\mu I_N \otimes J_T + \sigma^2_\lambda J_N \otimes I_T$

[4] to be added to the assumptions of the one-way, slide 8

The two ways (error) model (ii)

- The variance-covariance matrix can be also expressed in terms of a combination of the idempotent and mutually orthogonal matrices W, B_μ and B_λ:

$$\Omega = \sigma_\upsilon^2 W + (T\sigma_\mu^2 + \sigma_\upsilon^2)B_\mu + (N\sigma_\lambda^2 + \sigma_\upsilon^2)B_\lambda - \sigma_\upsilon^2 \bar{J}$$

where $B_\mu = I_N \otimes J_T/T$, $B_\lambda = J_T \otimes I_N/N$ and $\bar{J} = \frac{J_{NT}}{NT}$

- B_μ x y computes the individual means; B_λ x y computes the time means; $\bar{J}$ x y computes the overall mean.
- The new within matrix is $W = I_{NT} - B_\mu - B_\lambda + \bar{J}$

Fixed and random two ways models (i)

- As for the one way model, the two ways fixed effects model can be obtained in two different ways:
  1. by estimating OLS on the model that includes individual and time dummies (two ways LSDV)
  2. by estimating OLS on the model where all the variables have been transformed in deviation from time and individual means (pre multiplied by W, $y_{nt} - \bar{y}_{n.} - \bar{y}_{.t} + \bar{\bar{y}}$)

Fixed and random two ways models (ii)

- For the GLS, similarly to what has been proposed in the one way error model, the variables are pre-multiplied by $C = \Omega^{-0.5}$ because of the problem of the inversion of Ω, and then an OLS is applied on the transformed data.
- The role of each stochastic component is evaluated by means of:
  1. $\theta_\mu = 1 - \frac{\sigma_\upsilon}{\sqrt{\sigma_\upsilon^2 + T\sigma_\mu^2}}$ (a measure for the importance of individual variance)
  2. $\theta_\lambda = 1 - \frac{\sigma_\upsilon}{\sqrt{\sigma_\upsilon^2 + N\sigma_\lambda^2}}$ (a measure for the importance of temporal variance)
  3. $\theta_2 = 1 - \frac{\sigma_\upsilon}{\sqrt{\sigma_\upsilon^2 + T\sigma_\mu^2 + N\sigma_\lambda^2}}$ (a measure for the importance of individual + temporal variance)

Application

Application in R: Example 2.8 and 2.9 (Croissant, Millo)

Advanced Arguments

- alternative ways to model the errors in the random effect model
- models to account for endogeneity in the covariates
- models for unbalanced panels (Chapter 3, Croissant & Millo)
- models for count or dichotomous data
- dynamic models accounting for time lagged terms
- models for spatial panels accounting for spatial lagged terms

Take home messages

- Different ways to account for heterogeneity (wipes out that or accounting for that)
- The individual and time effects may be modelled as fixed or random
- Within, Between, pooled OLS or GLS
- GLS and OLS gives intermediate results between within and between, with two extreme cases
- The use of fixed or random depends on the correlation assumption among the regressors and the errors

References

- Amemiya, T. (1971). The Estimation of the Variances in a Variance–Components Model. International Economic Review, 12, 1–13.
- Baltagi, B. (2008). Econometric analysis of panel data. John Wiley & Sons.
- Croissant, Y., & Millo, G. (2019). Panel data econometrics with R. Wiley.
- Nerlove M (1971). Further Evidence on the Estimation of Dynamic Economic Relations from a Time–Series of Cross–Sections. Econometrica, 39, 359–382.
- Schaller, H. (1990). A re-examination of the q theory of investment using us firm data. Journal of applied econometrics, 5(4), 309-325.
- Swamy, P. A. V. B., & Arora, S. S. (1972). The exact finite sample properties of the estimators of coefficients in the error components regression models. Econometrica: journal of the Econometric Society, 261-275.
- Wallace TD, Hussain A (1969). The Use of Error Components Models in Combining Cross Section With Time Series Data. Econometrica, 37(1), 55–72.