Heterogeneous panel data models

Jakub Mućk
SGH Warsaw School of Economics

Heterogeneity in the slope coefficients

- In the standard linear panel data model we control for unobserved heterogeneity:

$$y = \beta' X + u, \tag{1}$$

  Where u is the sum of individual-specific component (in the RE model) and the idiosyncratic component. In the FE model, the individual-specific intercepts are introduced while the u contains only the idiosyncratic shock.
- At the same time, we have assumed that all slope coefficients (vector β) are the same for all unit and all periods. **In the above formulation, we don't allow for any interaction between individual effects and explanatory variable.**
- Consider the following formulation:

$$y_{it} = \beta'_{it} X_{it} + u_{it}, \tag{2}$$

  where all slope coefficients captured by β_{it} are now time-varying and individual-specific.
  - Although the above general formulation seems to be more realistic it lacks any explanatory power and is not useful for prediction.
  - The above model is not estimable since the number of parameters exceeds the number of observations.
- More applicable formulations:

$$y_i = \beta'_i X_i + u_i, \tag{3}$$

$$y_t = \beta'_t X_t + u_t \tag{4}$$

Heterogeneity in the slope coefficients II

SGH

- Which kind of heterogeneity in the slopes should introduce? In general, we pay more attention to individual effects but it depends on
  - T and N,
  - the research question.
- To account for the individuals differences in the slope coefficients we will introduce:
  - Seemingly Unrelated Regression (SUR),
  - Swamy's random coefficient model,
  - Mean group estimation.

Seemingly Unrelated Regression (SUR)

- **Seemingly Unrelated Regression (SUR)** is estimation method that is designed to estimate a system of linear equation (with potentially different set of explanatory variables) and which accounts for the cross-equation correlation of the error term.
- Consider the following set of equations:

$$y_i = X_i \beta_i + \varepsilon_i \quad \text{for } i \in \{1, \ldots, m\} \tag{5}$$

  where the index i denotes the i-th equation in the considered system.
- In the matrix form:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} X_1 & 0 & \ldots & 0 \\ 0 & X_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & X_m \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_m \end{bmatrix}. \tag{6}$$

- In i-th equation, K_i parameters are estimated. It yields the total number of coefficient $K = \sum_{i=1}^{m} K_i$. In addition, the $K_i > T_i$.
- Strictly exogeneity is assumed, i.e., $\mathbb{E}(\varepsilon | X_1, \ldots, X_m) = 0$.
- In the SUR framework, it is possible to assume that the covariance matrix of the error term is not diagonal:

$$\Omega = \mathbb{E}\left(\varepsilon\varepsilon|' X_1, \ldots, X_m\right) = \begin{bmatrix} \sigma_{11}^2 I & \sigma_{12}^2 I & \ldots & \sigma_{1m}^2 I \\ \sigma_{21}^2 I & \sigma_{22}^2 I & \ldots & \sigma_{2m}^2 I \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1}^2 I & \sigma_{m2}^2 I & \ldots & \sigma_{mm}^2 I \end{bmatrix}. \tag{7}$$

- Given the above structure of the variance-covariance matrix of the error term, the system of equations can be estimated with FGLS (feasible generalized least squares). Conventionally, the two-step estimation includes the following steps
  1. Running the OLS regression for the considered system of equations to get consistent and unbiased estimates of the variance-covariance matrix of the error term ($\hat{\Omega}$).
  2. Based on the estimates of the $\hat{\Omega}$, standard GLS estimator can be applied:

$$\hat{\beta}^{SUR} = \left(X'\hat{\Omega}^{-1}X\right)^{-1} X'\hat{\Omega}^{-1}y. \tag{8}$$

- Note that if Ω is diagonal then β^{SUR} will be close to the OLS estimator.

The SUR estimation and panel data

- In the context of long and narrow panel data, the SUR can be applied to account for a potential heterogeneity in the slopes.
- Consider the case of long (relatively large T) and narrow (not so large N) panel. Then, the standard linear model can be expressed as a set of equations:

$$
\begin{aligned}
y_1 &= \beta_1' X_1 + \varepsilon_1, \\
y_2 &= \beta_2' X_2 + \varepsilon_2, \\
\vdots &= \vdots \\
y_N &= \beta_N' X_N + \varepsilon_N,
\end{aligned}
$$

  where β_N is the individual-specific vector of the structural parameters.
- The SUR method accounts for cross-equation correlation. In the above case, this correlation is equivalent to cross-sectional dependence.
- It's possible to test heterogeneity of slopes. The standard Wald test can be used to verify the hypothesis about:
  - homogeneity of all slopes, i.e., $\mathcal{H}_0 : \quad \beta_1 = \ldots = \beta_N$, where β_i stands for vector of parameters for i-th unit.
  - homogeneity of some slopes, i.e., $\mathcal{H}_0 : \quad \beta_{1,j} = \ldots = \beta_{N,j}$, where $\beta_{i,j}$ stands for j-th parameter for i-th unit.

- **The SUR method** provides more efficient estimates since it accounts for cross-equation dependence.
- Cross-equation dependence can be tested with the LM statistic (Breusch and Pagan, 1980):

$$LM = T \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{i,j}^2, \tag{9}$$

where $\rho_{i,j}$ is cross-sectional correlation coefficient:

$$\hat{\rho}_{i,j} = \frac{\sum_{t=1}^{T} \hat{\varepsilon}_{it} \hat{\varepsilon}_{jt}}{\left(\sum_{t=1}^{T} \hat{\varepsilon}_{it}\right)^{\frac{1}{2}} \left(\sum_{t=1}^{T} \hat{\varepsilon}_{jt}\right)^{\frac{1}{2}}}. \tag{10}$$

The LM statistic is valid for fixed N as $T \to \infty$ and is asymptomatically distributed as χ^2 with $N(N-1)/2$ degrees of freedom.

Swamy's random coefficient model

Swamy's random coefficient model I

SGH

- Swamy (1970) proposes the random coefficient model. Consider the following model:

$$y_i = X_i \beta_i + \varepsilon_i \tag{11}$$

where the individual-specific slope β_i is the sum of common (β) and unit-specific (α_i) components:

$$\beta_i = \beta + \alpha_i, \tag{12}$$

where

1. $\mathbb{E}(\alpha_i) = 0$,
2. $\mathbb{E}(\alpha_i \alpha_i') = \Sigma$.

- **Question:** how to estimate β and Σ?
- The dependent variable can be expressed:

$$y_i = X_i \beta_i + \varepsilon_i = X_i \beta + X_i \alpha_i + \varepsilon_i = X_i \beta + \nu_i,$$

where $\nu_i = X_i \alpha_i + \varepsilon_i$ and $\mathbb{E}(\nu_i) = 0$.

- The variance-covariance of the error term ν_i for i-th unit is the following:

$$\mathbb{E}\left(\nu_i \nu_i'\right) = \mathbb{E}\left((X_i \alpha_i + \varepsilon_i)(X_i \alpha_i + \varepsilon_i)'\right) = \mathbb{E}\left(\varepsilon_i \varepsilon_i'\right) + X_i \mathbb{E}\left(\alpha_i \alpha_i'\right) X_i'.$$

- If the idiosyncratic error term is spherical then:

$$\mathbb{E}\left(\nu_i \nu_i'\right) = \sigma_i^2 I + X_i \Sigma X_i' = \Pi_i.$$

- The Π variance-covariance matrix for the error term will be block-diagonal.
- Finally, the GLS estimator can be applied:

$$\hat{\beta}^{RC} = \left(\sum_i X_i' \Pi_i^{-1} X_i\right)^{-1} \sum_i X_i' \Pi_i^{-1} y_i = \sum_i W_i \hat{\beta}_i^{OLS}. \tag{13}$$

where $\hat{\beta}_i^{OLS}$ is the unit-specific OLS estimates and W_i:

$$W_i = \left[\sum_i (\Sigma + V_i)^{-1}\right]^{-1} (\Sigma + V_i)^{-1} \tag{14}$$

where V_i is the panel-specific variance-covariance of $\hat{\beta}_i^{OLS}$, i.e., $\hat{V}_i = \sigma_i^2 \left(X_i' X_i\right)^{-1}$.

- The variance of β can be calculated as:

$$Var(\beta) = \sum_i (\Sigma + V_i)^{-1}. \tag{15}$$

Swamy's random coefficient model III

SGH

- Finally, the remainder element of the variance-covariance components which captures the variation of the slope coefficients, i.e., Σ, can be estimated based on the variation in the panel-specific β_i^{OLS} estimates:

$$\hat{\Sigma} = \frac{1}{N-1}\left(\sum_i \hat{\beta}_i^{OLS}(\hat{\beta}_i^{OLS})' - N\bar{\beta}^{OLS}(\bar{\beta}^{OLS})'\right) - \frac{1}{N}\sum_i \hat{V}_i$$

  where $\bar{\beta}^{OLS}$ is the average from the OLS estimates.

- Swamy (1970) postulates to omit the last component because it is negligible in large samples and it can be not positive definite.

Testing homogeneity in slopes

- To test whether the random coefficient model is statistically motivated one might compare the panel-specific estimates with their weighted (by V_i^{-1}) average.
- Test statistic:

$$\chi = \sum_{i=1}^{N} \left(\hat{\beta}_i^{OLS} - \tilde{\beta}\right)' \hat{V}_i^{-1} \left(\hat{\beta}_i^{OLS} - \tilde{\beta}\right), \tag{16}$$

where

$$\tilde{\beta} = \left(\sum_{i=1}^{N} \hat{V}_i^{-1}\right)^{-1} \sum_{i=1}^{N} \hat{V}_i^{-1} \hat{\beta}_i^{OLS}.$$

- The null hypothesis:

$$\mathcal{H}_0 \quad \beta_1 = \beta_2 = \ldots = \beta_N.$$

- The test statistic χ is asymptotically χ^2 distributed with $k \times (m-1)$ degrees of freedom.

Mean group estimation

Mean group estimation I

- The Mean Group estimator (MG) was proposed by Pesaran and Smith (1995) to deal with dynamic random coefficient model.
- The MG estimator is defined as the average of the unit-specific OLS estimators $\hat{\beta}_i^{OLS}$:

$$\hat{\beta}^{MG} = \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_i^{OLS}, \tag{17}$$

where

$$\hat{\beta}_i^{OLS} = \left(X_i' X_i\right)^{-1} X_i' y_i. \tag{18}$$

- It is assumed that all explanatory variables are strictly exogenous.
- The MG estimation is possible when both T and N are sufficiently large.
- The MG estimation can be applied irrespectively of the nature of heterogeneity in the slope coefficient. It can be applied if
  - the differences in slopes are random (as in the Swamy estimator),
  - diversity in the slopes can be captured by the fixed effects.
- The variance of the MG estimator:

$$Var(\hat{\beta}^{MG}) = \frac{1}{N(N-1)} \sum_{i=1}^{N} \left(\hat{\beta}_i^{OLS} - \hat{\beta}^{MG}\right) \left(\hat{\beta}_i^{OLS} - \hat{\beta}^{MG}\right)'. \tag{19}$$

- The MG estimator will be very close to the Swamy's estimator if T tends to infinity and there is some heterogeneity in the slopes:

$$\lim_{T\to\infty}\left(\hat{\beta}^{MG}-\hat{\beta}^{RC}\right)=0$$

Pooled MG

- In the **pooled mean group** estimation all coefficients are pooled, i.e, they are constrained to be identical:

$$\forall_i \beta_i = \beta. \tag{20}$$

- However, one might pooled only subset of coefficients.
- To test assumption abut homogeneity of coefficients one might used the standard Hausman test comapring mean group and pooled mean group estimates:
  - Under null both estimates are consistent while under alternative only mean group estimates are consistent.
  - By pooling we increase efficiency of estimates.

- The cross-sectional dependence leads to endogeneity in the mean group estimation. Recalling the least square estimator:

$$\hat{\beta}_i^{LS} = \beta_i + \left(X'X\right)^{-1} X'\mathbb{E}(\varepsilon_i), \tag{21}$$

  one might observed that presence of common factors lead to endogeneity.
- In the CCE (common correlated effects) estimation we control for multi-factor structure of the error term. To account for common factors individual-specific regressions are extended by cross-sectional averages of dependent variable.
- Given a high degree of uncertainty about the structure of the error term one might use also cross-sectional averages of explanatory variable. In addition, the lags of cross-sectional averages of both dependent and explanatory variables can be also included (see Chudik and Peseran, 2015).