Empirical exercise – The RE estimator as the FGLS estimator

1. Aim of the exercise

A main linear regression model that can accommodate panel data is the random effects model. We first study the assumptions of this model and then investigate why the error of this model is not spherical. We then use the feasible generalised least squares estimator to consistently estimate the coefficients of this model. The feasible generalised least squares estimator we use for this model is called the random effects estimator.

The theoretical context is as follows. Consider the linear model $y_{it} = \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}$ where $\varepsilon_{it} = \mu_i + \nu_{it}$ is a composite error term. There are T observations available for each individual i. There are N individuals. $\mathbf{x}'_{it}$ contains k independent variables for individual i at time t. It is a $1 \times k$ row vector. $\boldsymbol{\beta}$ is the vector of coefficients. It is a $k \times 1$ column vector.

The model imposes the following assumptions on the terms of the composite error of the regression. μ_i is a random effect with distribution $N(0, \sigma^2_\mu)$. It is individual-specific and hence it is the 'individual' or 'permanent' component of the error term. μ_i are uncorrelated across i. μ_i is uncorrelated with $\mathbf{x}'_{it}$. This is a main assumption of the model and it can be a strong assumption depending on the empirical relationships analysed. ν_{it} is a noise term with distribution $N(0, \sigma^2_\nu)$. It is not only individual-specific but also time-specific and hence it is the 'idiosyncratic' component of the error term. ν_{it} are uncorrelated within and across i. They are also uncorrelated with μ_i and with $\mathbf{x}'_{it}$.

The assumptions made on μ_i and ν_{it} imply that ε_{it} are serially correlated within i due to the permanent component μ_i. In fact, $\mathrm{Cov}\left[\varepsilon_{it}, \varepsilon_{is}\right] = \sigma^2_\mu$ for every i and $t \neq s$. On the other hand, ε_{it} are uncorrelated across i. That is, $\mathrm{Cov}\left[\varepsilon_{it}, \varepsilon_{js}\right] = 0$ for every $i \neq j$. It holds that $\mathrm{Var}\left[\varepsilon_{it}\right] = \sigma^2_\mu + \sigma^2_\nu$. The stated covariance and variance terms fully define the variance-covariance matrix of ε_{it}.

Since ε_{it} are serially correlated within i, the OLS estimator is not efficient. We can transform the terms of the regression equation so that the errors become serially uncorrelated. The random effects transformation subtracts from each variable of the regression a fraction of the time average of that variable to obtain $y_{it} - \lambda_i \bar{y}_i = \left(\mathbf{x}'_{it} - \lambda_i \bar{\mathbf{x}}'_i\right)\boldsymbol{\beta} + \varepsilon_{it} - \lambda_i \bar{\varepsilon}_i$ where $\lambda_i = 1 - \left(\sigma^2_\nu / \left(\sigma^2_\nu + T_i \sigma^2_\mu\right)\right)^{1/2}$. $T_i = T$ if the panel data of interest is balanced. It can be shown that in this model the transformed error is not serially correlated so that $\mathrm{Cov}\left[\varepsilon^*_{it}, \varepsilon^*_{is}\right] = 0$ for every i and $t \neq s$ where $\varepsilon^*_{it} = \varepsilon_{it} - \lambda_i \bar{\varepsilon}_i$ and $\bar{\varepsilon}_i = 1/T_i \sum_t^{T_i} \varepsilon_{it}$. This means that we can estimate the transformed model using the OLS estimator. However, λ_i depends on the unknown parameters σ^2_ν and σ^2_μ which need to be estimated. There are different methods for estimating these parameters. Here we consider the Swamy-Arora method also used by Stata. The relevant formulas are presented in `https://www.stata.com/manuals13/xtxtreg.pdf`. Given the estimates of the variance components, $\hat{\lambda}_i$ can be constructed. Hence, estimation of the transformed equation becomes 'feasible'. The feasible generalised least squares estimator is then the ordinary least squares estimator on the transformed y_{it} and $\mathbf{x}'_{it}$. The feasible generalised least squares estimator for the random effects model is called the random effects estimator.

The model as presented above is for individual i at time t. We can generalise the notation so that it accommodates N individuals and T_i observations available for each individual. If we stack all observations, the model becomes $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. $\mathbf{y}$ is $NT \times 1$. $\mathbf{X}$ is $NT \times k$. $\boldsymbol{\varepsilon}$ is $NT \times 1$. The variance-covariance matrix of the error can be expressed as $\mathrm{E}\left[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \mathbf{X}\right] =$

$\mathbf{\Omega} = \sigma_\mu^2 (\mathbf{I}_N \otimes \mathbf{J}_T) + \sigma_\nu^2 \mathbf{I}_{NT}$. $\mathbf{J}_T = \boldsymbol{\iota}_T \boldsymbol{\iota}_T'$ where $\boldsymbol{\iota}_T$ is a $T \times 1$ column of ones so that $\mathbf{J}_T$ is a $T \times T$ matrix filled with ones. $\otimes$ is the Kronecker product. It can be shown that the variance-covariance matrix of the error can be rewritten as $\mathbf{\Omega} = \left(T\sigma_\mu^2 + \sigma_\nu^2\right) \mathbf{P} + \sigma_\nu^2 \mathbf{M}$. $\mathbf{M} = \mathbf{I}_{NT} - \mathbf{P}$. $\mathbf{P} = \mathbf{D} \left(\mathbf{D}'\mathbf{D}\right)^{-1} \mathbf{D}'$ is the projection matrix for $\mathbf{D}$ and it is $NT \times NT$. $\mathbf{D} = \mathbf{I}_N \otimes \boldsymbol{\iota}_T$ which represents a matrix with N block diagonal elements where each block is given by $\boldsymbol{\iota}_T$. It is $NT \times N$. Using the Cholesky decomposition $\mathbf{\Omega}^{-1} = \mathbf{\Psi}'\mathbf{\Psi}$, it can be shown that $\mathbf{\Psi} = \mathbf{I}_{NT} - \lambda_i \mathbf{P}$. Then, $\mathbf{\Psi y} = \mathbf{I}_{NT}\mathbf{y} - \lambda_i \mathbf{Py}$ is just the above introduced transformation $y_{it} - \lambda_i \bar{y}_i$ in matrix form for N individuals with T_i observations available for each individual. Since $\mathbf{\Psi}$ is unknown, we can estimate it with $\hat{\mathbf{\Psi}} = \mathbf{I}_{NT} - \hat{\lambda}_i \mathbf{P}$. The feasible generalised least squares estimator is then just the ordinary least squares estimator on the transformed $\mathbf{y}$ and $\mathbf{X}$. The feasible generalised least squares estimator for the random effects model is the random effects estimator and takes the form

$$\hat{\boldsymbol{\beta}}_{RE} = \left(\left(\hat{\mathbf{\Psi}}\mathbf{X}\right)' \left(\hat{\mathbf{\Psi}}\mathbf{X}\right)\right)^{-1} \left(\hat{\mathbf{\Psi}}\mathbf{X}\right)' \left(\hat{\mathbf{\Psi}}\mathbf{y}\right).$$

The empirical context is as follows. '*Francis Vella and Marno Verbeek, 1998. Whose wages do unions rise? A dynamic model of unionism and wage rate determination for young men, Journal of Applied Econometrics, 13(2), 163-183.*' analyse how union membership affects hourly wages. To this purpose data were collected from the National Longitudinal Survey which contains observations on 545 males for the years 1980-1987. The data can be obtained from the Data Archive of the Journal of Applied Econometrics at `http://econ.queensu.ca/jae/1998-v13.2/vella-verbeek/`. A description of the subset of the variables used in this exercise is as follows. 'nr' is the identification number of an agent. The data contains 545 agents. 'year' is the year of observation. The data contains 8 years of data for each individual. Therefore, there are 4360 observations available. 'wage' is hourly wage. 'union' indicates whether wage is set by collective bargaining. 'exper' is age-6-school. 'school' is the years of schooling. 'black' and 'hisp' are indicators of ethnic origin. 'mar' indicates agent is married. 'hlth' indicates whether agent is disabled. 'rur' indicates whether living in rural area. Our aim is to use the random effects regression model to explain hourly wage with union membership controlling for a set of independent variables and controlling for unobserved individual heterogeneity that we believe is not correlated with the independent variables.

2. Load the data

Load the data to the memory.

```
clear;
load 'M:\exercisere.mat';
```

3. Define the dependent variable

Name the wage variable as `y`.

```
y = wage;
```

4. Determine the number of units of observation, number of units of time, number of observations

Use the code presented at the end of the section to determine the number of men in the data, the number of years data is available for each agent, and the total number of observations for all years and individuals. In this exercise the panel data is balanced so that $T_i = T$. That is, T years of data is available for all individuals in the panel data.

```
uniq_ind = unique(nr);
uniq_tim = unique(year);
N = size(uniq_ind,1);
T = size(uniq_tim,1);
NT = size(nr,1);
```

5. The variance-covariance matrix of the error

Section 1 of this exercise has presented the structure of the variance-covariance matrix of the errors of the random effects regression model. First, the covariance of the errors across time but within individuals, across time and across individuals, as well as the variance of the errors across time and individuals are shown to be equal to particular variance parameters. The variance-covariance matrix of the errors is then expressed in matrix notation in terms of these parameters.

Take a moment and assume that $\sigma_\mu^2 = 0.11$ and $\sigma_\nu^2 = 0.15$. The code presented at the end of the section constructs the variance-covariance matrix of the errors using the notation for this matrix introduced in Section 1, and the assumed values for the variance parameters. Looking at this matrix, are the errors heteroskedastic? Are they serially correlated?

```
O = 0.11*kron(eye(N),ones(T,1)*ones(T,1)')+0.15*eye(NT);
```

6. Create a matrix containing dummy variables for panel units

Use the code presented at the end of the section to create the **D** matrix introduced in Section 1.

```
D = kron(eye(N),ones(T,1));
```

7. Create projection matrices

Use the code presented at the end of the section to create the projection matrices **P** and **M** for matrix **D**.

```
I = eye(NT);
P_D = D*inv(D'*D)*D';
M_D = I-P_D;
```

8. Create the systematic component of the regression equation

Create the systematic component of the regression equation.

```
x_0 = ones(NT,1);
X = [x_0 union exper exper.*exper school black hisp mar hlth rur];
```

9. Obtain the estimated transformation matrix

The code presented at the end of the section takes three steps to construct the transformation matrix $\hat{\boldsymbol{\Psi}}$ introduced in Section 1. First, the variance parameters σ_μ^2 and σ_ν^2 are estimated using the functions `exerciserefunbgs` and `exerciserefunwgs`. These functions use the panel data within-group and between-group estimators to estimate the variance parameters. Second, $\hat{\boldsymbol{\lambda}}$ is constructed using the estimated variance parameters. Finally, the $\hat{\boldsymbol{\Psi}}$ matrix is constructed.

```
BGS = exerciserefunbgs(y,X,NT,T,P_D);
WGS = exerciserefunwgs(y,X,NT,N,P_D,M_D);
sigma_hat_sq_bgs = BGS.sigma_hat_sq;
sigma_hat_sq_nu = WGS.sigma_hat_sq;
sigma_hat_sq_mu = sigma_hat_sq_bgs-sigma_hat_sq_nu/T;
Lambda_hat = 1-sqrt(sigma_hat_sq_nu/(sigma_hat_sq_nu+T*sigma_hat_sq_mu));
Psi_hat = eye(NT)-Lambda_hat*P_D;
```

10. Obtain the RE coefficient estimates as the FGLS coefficient estimates

Use the $\hat{\boldsymbol{\Psi}}$ matrix to transform $\mathbf{y}$ and $\mathbf{X}$ and respectively name the transformed variables as `y_t` and `X_t`. Apply the OLS estimator on the transformed variables.

```
y_t = Psi_hat*y;
X_t = Psi_hat*X;
B_hat_RE = inv(X_t'*X_t)*(X_t'*y_t);
```