1. Non-spherical variance-covariance matrix of the error term

2. The First-Difference (FD) estimator
Spherical variance-covariance matrix

In the previous meeting we’ve assumed that the variance covariance matrix of the error term is spherical:

\[ \mathbb{E}(uu') = \sigma_u^2 I \]  \hspace{1cm} (1)

or, at least, block diagonal (in the RE model):

\[
\mathbb{E}(u_i u_i') = \Sigma_{i,i} = \begin{pmatrix}
1 & \rho & \ldots & \rho \\
\rho & 1 & \ddots & \rho \\
\vdots & \vdots & \ddots & \vdots \\
\rho & \rho & \ldots & 1
\end{pmatrix}.
\]  \hspace{1cm} (2)

where

\[ \rho = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\varepsilon^2}, \]

where \( \sigma_\mu^2 \) and \( \sigma_\varepsilon^2 \) stands for the variance of the individual-specific and idiosyncratic error term.
Non-spherical variance-covariance matrix

More general case:

$$E(\mathbf{uu}') = \Sigma = \begin{bmatrix}
    \Sigma_{1,1} & \Sigma_{1,2} & \cdots & \Sigma_{1,N} \\
    \Sigma_{2,1} & \Sigma_{2,2} & \cdots & \Sigma_{2,N} \\
    \vdots & \vdots & \ddots & \vdots \\
    \Sigma_{N,1} & \Sigma_{N,2} & \cdots & \Sigma_{N,N}
\end{bmatrix},$$

where $\Sigma_{i,j}$ is the variance-covariance matrix of the error term between $i$-th and $j$-th (cross-sectional) unit.
More general case:

\[ E(uu') = \Sigma = \begin{bmatrix}
\Sigma_{1,1} & \Sigma_{1,2} & \ldots & \Sigma_{1,N} \\
\Sigma_{2,1} & \Sigma_{2,2} & \ldots & \Sigma_{2,N} \\
\vdots & \vdots & \ddots & \vdots \\
\Sigma_{N,1} & \Sigma_{N,2} & \ldots & \Sigma_{N,N}
\end{bmatrix}, \tag{3}\]

where \( \Sigma_{i,j} \) is the variance-covariance matrix of the error term between \( i \)-th and \( j \)-th (cross-sectional) unit.

Implications: the least squares estimator is still consistent (if other assumptions are satisfied) but it is no longer BLUE (best linear unbiased estimator).
The Generalized Least Squares estimator

- To overcome the problem of non-spherical variance-covariance matrix of the error term
- If we know $\Sigma$ and the other assumptions are satisfied one might apply the GLS (Generalized Least Squares) estimator:

$$\hat{\beta}^{\text{GLS}} = \left( X' \hat{\Sigma}^{-1} X \right)^{-1} X' \hat{\Sigma}^{-1} y, \quad (4)$$

and the variance-covariance estimator:

$$\text{Var} \left( \hat{\beta}^{\text{GLS}} \right) = \left( X' \hat{\Sigma}^{-1} X \right)^{-1}. \quad (5)$$

- In the presence of the non-spherical disturbances the GLS estimator is BLUE.
- Key challenge: $\Sigma$ !.
The variance-covariance matrix of the error term can be non-spherical due to:

1. Autocorrelation (serial correlation).
2. Heteroskedasticity.
4. Combination of the above cases.
Autocorrelation or Serial correlation is the correlation of the error term with its past values.

AR(1) example:

\[ u_{it} = \rho u_{it-1} + \eta_{it} \]  

(6)

where \( \rho \in (0, 1) \) and \( \eta_{it} \sim \mathcal{N}(0, \sigma^2) \).

Autocorrelation signals that disturbances displays some memory.

One possible explanation for autocorrelation is that relevant factors are omitted \( \implies \) omitted variables bias.

\[ \rho = 0.95, \rho = 0.75, \rho = 0.5. \]
Detecting autocorrelation

- Visual inspection: plotting residuals.
- Simple regressions.
- Test proposed by Baltagi and Wu (1999).
  - The null hypothesis is about no autocorrelation:
    \[ H_0 \quad \rho = 0. \] (7)
  - The test investigates only the first-order autocorrelation.
  - The test statistics is the Durbin-Watson statistic tailored to the panel data.
The GLS estimator assuming AR(1) error term

General idea:

1. The autocorrelation coefficient is estimated from the OLS (within) residuals.
2. All variables are transformed:

\[
\begin{align*}
  z^*_it &= \begin{cases} 
    (1 - \hat{\rho}^2)\frac{1}{2} z_{it} & \text{if } t = 1 \\
    (1 - \hat{\rho}^2)^{\frac{1}{2}} \left( z_{it} \left( \frac{1}{1 - \hat{\rho}^2} \right)^{\frac{1}{2}} - z_{it-1} \left( \frac{\hat{\rho}^2}{1 - \hat{\rho}^2} \right) \right) & \text{if } t > 1
  \end{cases}
\end{align*}
\]  

Note that the above transformation is quite similar to the Prais-Winters transformation.

3. The first observation of each panel should be removed and then it is possible to apply within (FE) estimator to transformed data.
4. Baltagi and Wu propose the GLS estimator of the RE model with the AR error term. The main idea is quite similar to basic RE model.
Heteroskedasticity refers to the situation in which the variance of the error term is not constant.

Example for panel data:

\[ \mathbb{E} (uu') = \Sigma = \text{diag} \left( I \sigma^2_{u1}, \ldots, I \sigma^2_{u,N} \right) \neq I \sigma^2_u, \quad (9) \]

when \( \sigma^2_{u,1} \neq \ldots \neq \sigma^2_{u,N} \).

General intuition: uncertainty associated with the outcome \( y \) (captured by the variance of the error term) is not constant for various values of independent variables \( x \).
The Robust standard errors

- When the error term is heteroskedastic the robust estimator of the variance-covariance can be used to obtain consistent estimates of the standard errors.

- **White’s heteroscedasticity-consistent estimator:**

  \[ \text{Var}(\hat{\beta}) = (X'X)^{-1} \left( X'\hat{\Sigma}X \right) (X'X)^{-1} \]  
  
  where \( \hat{\Sigma} = \text{diag}(\hat{u}_1^2, \ldots, \hat{u}_N^2) \).

- **The clustered robust standard errors:**
  - All observations are divided into G groups:

  \[ \text{Var}(\hat{\beta}) = (X'X)^{-1} \left( \sum_{i=1}^{G} x'_i \hat{u}_i \hat{u}_i' x_i \right) (X'X)^{-1}. \]  

  (11)
Cross-sectional dependence

- **Cross-sectional dependence** takes place when the error terms between individuals at the same time period $t$ are correlated:

$$\mathbb{E}(u_{it}u_{jt}) \neq 0 \text{ if } i \neq j.$$  \hspace{1cm} (12)

- The cross-sectional dependence may arise due to the presence of common or/and unobserved factors that become a part of the error term. For instance:
  - common business cycles fluctuations,
  - spillovers,
  - neighborhood effects, herd behavior, and interdependent preferences.
Testing for the cross-sectional dependence

Consider the static panel data model:

\[ y_{it} = \alpha + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + u_{it}. \]  

(13)

The cross-sectional correlation coefficient:

\[ \hat{\rho}_{i,j} = \frac{\sum_{t=1}^{T} \hat{u}_{it} \hat{u}_{jt}}{\left( \sum_{t=1}^{T} \hat{u}_{it} \right)^{\frac{1}{2}} \left( \sum_{t=1}^{T} \hat{u}_{jt} \right)^{\frac{1}{2}}}. \]  

(14)

Note that \( \rho_{i,j} = \rho_{j,i} \). For panel consisting of \( N \) unit we get \( N(N-1)/2 \) pair-wise correlation coefficients.

The hypothesis of interest:

\[ H_0 : \rho_{i,j} = 0 \text{ if } i \neq j. \]  

(15)

\[ H_1 : \rho_{i,j} \neq 0 \text{ if } i \neq j \]  

(16)

The LM statistic (Breusch and Pagan, 1980):

\[ LM = T \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{i,j}^2. \]  

(17)

The above statistic is valid for fixed \( N \) as \( T \to \infty \) and is asymptotically distributed as \( \chi^2 \) with \( N(N-1)/2 \) degrees of freedom.

The LM statistic exhibits substantial size distortion when \( N \) is relatively large due to fact that it is not correctly centered for fixed \( T \).
Testing for the cross-sectional dependence

- Pesaran (2004) proposes the following test statistic:

\[ CD = \sqrt{\frac{2T}{N(N - 1)}} \left( \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{i,j} \right). \]  \hfill (18)

Under the null about no cross sectional dependence, the CD statistic is normally distributed, i.e., \( CD \sim N(0, 1) \), for \( N \to \infty \) and sufficiently large \( T \).

- The CD statistic can be used in a wide range of panel-data models, e.g., basic static models, homogeneous/heterogeneous dynamic model, nonstationary model.

- Unbalanced panels:

\[ CD = \sqrt{\frac{2}{N(N - 1)}} \left( \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{T_{ij}} \hat{\rho}_{i,j} \right). \]  \hfill (19)
Testing for the cross-sectional dependence

- **Friedman’s statistics:**
  \[ FR = \frac{T - 1}{(N - 1)R_{\text{ave}} + 1} \]  \hspace{2cm} (20)
  where \( R_{\text{ave}} \) is the average Spearman’s cross-sectional correlation.

- The FR statistic is asymptotically \( \chi^2 \) distributed with \( T - 1 \) degrees of freedom, for fixed \( T \) and sufficiently large \( N \).
Testing for the cross-sectional dependence

- **Friedman’s statistics:**
  \[
  FR = \frac{T - 1}{(N - 1)R_{ave} + 1}
  \]  
  (20)
  where \( R_{ave} \) is the average Spearman’s cross-sectional correlation.
- The FR statistic is asymptotically \( \chi^2 \) distributed with \( T - 1 \) degrees of freedom, for fixed \( T \) and sufficiently large \( N \).
- **Frees’ statistics** bases on the sum of the squared rank correlation coefficients \( R^2_{ave} \):
  \[
  FRE = N \left( R^2_{ave} - \frac{1}{T - 1} \right)
  \]  
  (21)
  where
  \[
  R^2_{ave} = \sqrt{\frac{2}{N (N - 1)}} \left( \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \sqrt{T_{ij}\hat{r}_{i,j}} \right)
  \]  
  (22)
  where \( r_{i,j} \) is the Spearman’s correlation between residuals for \( i \) and \( j \) unit.
- The null is rejected when \( R^2_{ave} > 1/(T - 1) + Q_b/N \), where \( Q_b \) is the \( b \)-th quantile of the \( Q \) distribution (the \( Q \) distribution is the weighted sum of two \( \chi^2 \) random variables).
- Both Friedman’s and Frees’ statistic are designed to static panel data models.
Outline

1. Non-spherical variance-covariance matrix of the error term

2. The First-Difference (FD) estimator
The First-Difference (FD) estimator is an alternative estimation technique that eliminates the fixed effect as well as time invariant regressors.

Note that

\[ y_{it} = \alpha_i + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + u_{it} \quad \text{for} \quad t = 1, \ldots T, \tag{23} \]

\[ y_{it-1} = \alpha_i + \beta_1 x_{1it-1} + \ldots + \beta_k x_{kit-1} + u_{it-1} \quad \text{for} \quad t = 2, \ldots T, \tag{24} \]

and differencing both equations yields:

\[ \Delta y_{it} = \beta_1 \Delta x_{1it} + \ldots + \beta_k \Delta x_{kit} + \Delta u_{it}, \tag{25} \]

where \( \Delta \) is the well-known (from time series analysis) first-difference operator, i.e. \( \Delta z_t = z_t - z_{t-1} \).

The parameters in (25) can be estimated with the least squares. In the matrix form:

\[ \beta^{FD} = (\Delta X' \Delta X)^{-1} \Delta X' \Delta y. \tag{26} \]
The estimates of fixed effects can be also recovered:

\[
\hat{\alpha}_{i}^{FD} = \bar{y}_{i} - \bar{x}_{i}\hat{\beta}^{FD}.
\]  
(27)

If the error term in (25) is not correlated with independent variable (weak exogeneity) then the least squares estimator is unbiased and consistent.

The above assumption is less restrictive than in standard FE model:

\[
E(\Delta u_{it} | \Delta x_{it}) = E(u_{it} - u_{it-1} | x_{it} - x_{it-1}) = 0.
\]  
(28)

The FE estimator is more efficient when the disturbances are not serially correlated and homoskedastic.

But If \(u_{it}\) is driven by random walk (autocorrelation with \(\rho = 1\)) then the FD estimator is more efficient.