Two-way error component models. GLS estimation. Cross-sectional dependence.

Jakub Mućk
SGH Warsaw School of Economics

Two-way Error Component Model

Two-way Error Component Model

$$y_{it} = \alpha + X'_{it}\beta + u_{it} \qquad i \in \{1, \ldots, N\}, t \in \{1, \ldots, T\}, \tag{1}$$

the composite error component:

$$u_{i,t} = \mu_i + \lambda_t + \varepsilon_{i,t}, \tag{2}$$

where:

- **μ_i – the unobservable individual-specific effect**;
- **λ_t – the unobservable time-specific effect**;
- $u_{i,t}$ – the remainder disturbance.

- **Two-way fixed effects model**:

$$y_{it} = \alpha_i + \lambda_t + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + u_{it}, \tag{3}$$

  where
  - α_i - the individual-specific intercept;
  - λ_i - the period-specific intercept;
  - $u_{it} \sim \mathcal{N}\left(0, \sigma_u^2\right)$.
- As in the case of the one-way fixed effect model, independent variables x_1, ..., x_k cannot be time invariant (for a given unit).
- The estimation:
  1. the least squares dummy variable estimator (LSDV);
  2. the within estimator;
  3. combination of the above methods.

The LSDV estimator

- In the analogous fashion to the one-way fixed effect model, we define the following dummy variables:
  - for the j-th unit:
$$\mathcal{D}_{jit} = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$
  - for the τ-th period:
$$\mathcal{B}_{\tau it} = \begin{cases} 1 & t = \tau \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

- In the analogous fashion to the one-way fixed effect model, we define the following dummy variables:
  - for the j-th unit:

$$\mathcal{D}_{jit} = \begin{cases} 1 & i = j \\ 0 & \text{otherwise} \end{cases}, \tag{4}$$

  - for the τ-th period:

$$\mathcal{B}_{\tau it} = \begin{cases} 1 & t = \tau \\ 0 & \text{otherwise} \end{cases}. \tag{5}$$

- The LSDV estimator can be obtained by estimating the following model:

$$y_{it} = \sum_{j=1}^{N} \alpha_j \mathcal{D}_{jit} + \sum_{\tau=1}^{N} \lambda_\tau \mathcal{B}_{\tau it} + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + u_{i,t}. \tag{6}$$

  where $u_{i,t} \sim \mathcal{N}\left(0, \sigma_u^2\right)$.
- The parameters in equation (7) can be estimated with the OLS.

- **The LSDV estimator:**

$$y_{it} = \sum_{j=1}^{N} \alpha_j \mathcal{D}_{jit} + \sum_{\tau=1}^{N} \lambda_\tau \mathcal{B}_{\tau it} + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + u_{i,t}, \quad (7)$$

where $u_{i,t} \sim \mathcal{N}\left(0, \sigma_u^2\right)$.

- We can test whether the individual-specific and period-specific effects are significantly different:

$$\mathcal{H}_0 : \quad \alpha_1 = \ldots = \alpha_k \quad \wedge \quad \lambda_1 = \ldots = \lambda_T = 0. \quad (8)$$

- To verify the null described by (8) we compare the sum of squared errors from the pooled model with (with restrictions, SSE_R) with the sum of squared errors from the LSDV model (without restrictions, SSE_R). The test statistics $\mathcal{F}$:

$$\mathcal{F} = \frac{(SSE_R - SSE_U)/(N - T - 1)}{SSE_U/(NT - K - T)} \quad (9)$$

if null is emph then $\mathcal{F} \sim \mathcal{F}_{(N-T-1, NT-K-T)}$.

- **The LSDV estimator:**

$$y_{it} = \sum_{j=1}^{N} \alpha_j \mathcal{D}_{jit} + \sum_{\tau=1}^{N} \lambda_\tau \mathcal{B}_{\tau it} + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + u_{i,t}, \tag{7}$$

where $u_{i,t} \sim \mathcal{N}\left(0, \sigma_u^2\right)$.

- We can test whether the individual-specific and period-specific effects are significantly different:

$$\mathcal{H}_0 : \quad \alpha_1 = \ldots = \alpha_k \quad \wedge \quad \lambda_1 = \ldots = \lambda_T = 0. \tag{8}$$

- To verify the null described by (8) we compare the sum of squared errors from the pooled model with (with restrictions, SSE_R) with the sum of squared errors from the LSDV model (without restrictions, SSE_R). The test statistics $\mathcal{F}$:

$$\mathcal{F} = \frac{(SSE_R - SSE_U)/(N-T-1)}{SSE_U/(NT-K-T)} \tag{9}$$

if null is emph then $\mathcal{F} \sim \mathcal{F}_{(N-T-1, NT-K-T)}$.

- One might test only individual or period effects:

$$\mathcal{H}_0 : \qquad \alpha_1 = \ldots = \alpha_k, \tag{10}$$

$$\mathcal{H}_0 : \qquad \lambda_1 = \ldots = \lambda_k. \tag{11}$$

The construction of test for poolability is the same but we use different degrees of freedom.

- Consider the following transformations:
  - averaging over time (for each unit i): $\bar{y}_i$,
  - averaging over unit (for each period t): $\bar{y}_t$,
  - averaging over time and unit: $\bar{y}$,

  for the dependent variable y_{it}:

$$\bar{y}_i = \frac{1}{T}\sum_{t}^{T} y_{it}, \quad \bar{y}_t = \frac{1}{N}\sum_{i}^{N} y_{it}, \quad \bar{y} = \frac{1}{NT}\sum_{t}^{T}\sum_{i}^{N} y_{it}.$$

The within estimator

- Consider the following transformations:
  - averaging over time (for each unit i): $\bar{y}_i$,
  - averaging over unit (for each period t): $\bar{y}_t$,
  - averaging over time and unit: $\bar{y}$,

  for the dependent variable y_{it}:

$$\bar{y}_i = \frac{1}{T}\sum_t^T y_{it}, \quad \bar{y}_t = \frac{1}{N}\sum_i^N y_{it}, \quad \bar{y} = \frac{1}{NT}\sum_t^T\sum_i^N y_{it}.$$

- Given the above transformations we can eliminate both the individual- and period specific effects:

$$(y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}_{it}) = \beta' (x_{it} - \bar{x}_i - \bar{x}_t + \bar{x}_{it}) + (u_{it} - \bar{u}_i - \bar{u}_t + \bar{u}_{it}). \quad (12)$$

- **The within estimator:**

$$\tilde{y}_{it} = \beta_1 \tilde{x}_{1it} + \ldots + \beta_k \tilde{x}_{kit} + \tilde{u}_{it} \tag{13}$$

where

$$\begin{aligned}
\tilde{y}_{it} &= y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}_{it} \\
\tilde{x}_{1it} &= x_{1it} - \bar{x}_{1i} - \bar{x}_{1t} + \bar{x}_{1it} \\
\vdots &\ \vdots\ \vdots \\
\tilde{x}_{kit} &= x_{kit} - \bar{x}_{ki} - \bar{x}_{kt} + \bar{x}_{kit} \\
\tilde{u}_{it} &= u_{it} - \bar{u}_i - \bar{u}_t + \bar{u}_{it}.
\end{aligned}$$

- The parameters β_1, ..., β_k can be estimated by the OLS.
- The independent variables, i.e., x_{1it}, ..., x_{kit}, cannot be time (unit) invariant.

- **Two-way random effects model**:

$$y_{it} = \alpha + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + u_{it}, \tag{14}$$

where the error component is as follows:

$$u_{it} = \mu_i + \lambda_t + \varepsilon_{it}, \tag{15}$$

where

  - μ_i is the individual-specific error component and $\mu_i \sim \mathcal{N}\left(0, \sigma^2_\mu\right)$;
  - λ_t is the period-specific error component and $\lambda_t \sim \mathcal{N}\left(0, \sigma^2_\lambda\right)$;
  - ε_{it} is the idiosyncratic error component and $\varepsilon_{it} \sim \mathcal{N}\left(0, \sigma^2_\varepsilon\right)$.

- **Two-way random effects model**:

$$y_{it} = \alpha + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + u_{it}, \quad (14)$$

where the error component is as follows:

$$u_{it} = \mu_i + \lambda_t + \varepsilon_{it}, \quad (15)$$

where
  - μ_i is the individual-specific error component and $\mu_i \sim \mathcal{N}\left(0, \sigma_\mu^2\right)$;
  - λ_t is the period-specific error component and $\lambda_t \sim \mathcal{N}\left(0, \sigma_\lambda^2\right)$;
  - ε_{it} is the idiosyncratic error component and $\varepsilon_{it} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$.
- The independent variables can be time invariant.
- Individual-specific and period-specific effects are independent:

$$\begin{aligned} \mathbb{E}(\mu_i, \mu_j) &= 0 \quad \text{if} \quad i \neq j, \\ \mathbb{E}(\lambda_s, \lambda_t) &= 0 \quad \text{if} \quad s \neq t \\ \mathbb{E}(\mu_i, \lambda_t) &= 0 \end{aligned}$$

- Estimation method: GLS (*generalized least squares*).

- It is assumed that the error term in the two-way RE model can be described as follow:

$$u_{it} = \mu_i + \lambda_t + \varepsilon_{it}, \tag{16}$$

where $\mu_i \sim \mathcal{N}\left(0, \sigma_\mu^2\right)$, $\lambda_t \sim \mathcal{N}\left(0, \sigma_\lambda^2\right)$ and $\varepsilon_{it} \sim \mathcal{N}\left(0, \sigma_\varepsilon^2\right)$.

- The variance-covariance matrix of the error term is not diagonal!.
- The diagonal elements of the variance covariance matrix of the error term:

$$\begin{aligned}
\mathbb{E}\left(u_{it}^2\right) &= \mathbb{E}\left(\mu_i + \lambda_t + \varepsilon_{it}\right)^2 \\
&= \underbrace{\mathbb{E}\left(\mu_i^2\right)}_{=\sigma_\mu^2} + \underbrace{\mathbb{E}\left(\lambda_t^2\right)}_{=\sigma_\lambda^2} + \underbrace{\mathbb{E}\left(\varepsilon_{it}^2\right)}_{=\sigma_\varepsilon^2} + \underbrace{2\text{cov}\left(\mu_i, \lambda_t\right)}_{=0} + \underbrace{2\text{cov}\left(\mu_i, \varepsilon_{it}\right)}_{=0} + \underbrace{2\text{cov}\left(\lambda_t, \varepsilon_{it}\right)}_{=0} \\
&= \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\varepsilon^2.
\end{aligned}$$

- For a given unit, non-diagonal elements of the variance covariance matrix of the error term $(t \neq s)$:

$$\begin{aligned}
\text{cov}\left(u_{it}, u_{is}\right) &= \mathbb{E}\left[\left(\mu_i + \lambda_t + \varepsilon_{it}\right)\left(\mu_i + \lambda_s + \varepsilon_{is}\right)\right], \\
&= \underbrace{\mathbb{E}\left(\mu_i^2\right)}_{=\sigma_\mu^2} + \underbrace{\mathbb{E}\left(\lambda_t \lambda_s\right)}_{=0} + \underbrace{\mathbb{E}\left(\varepsilon_{it}\varepsilon_{is}\right)}_{=0} + \underbrace{\mathcal{COV}_1}_{=0} \\
&= \sigma_\mu^2,
\end{aligned}$$

where $\mathcal{COV}_1 = \text{cov}(\lambda_t, \mu_i) + \text{cov}(\lambda_s, \mu_i) + \text{cov}(\lambda_t, \varepsilon_{it}) + \text{cov}(\lambda_s, \varepsilon_{it}) + 2\text{cov}(\mu_i, \varepsilon_{it})$.

- Similarly, for a a given period, non-diagonal elements of the variance covariance matrix of the error term ($i \neq j$):

$$\begin{aligned} \operatorname{cov}\left(u_{it}, u_{jt}\right) &= \mathbb{E}\left[\left(\mu_i + \lambda_t + \varepsilon_{it}\right)\left(\mu_j + \lambda_t + \varepsilon_{jt}\right)\right], \\ &= \underbrace{\mathbb{E}\left(\mu_i \mu_j\right)}_{=0} + \underbrace{\mathbb{E}\left(\lambda_t^2\right)}_{=\sigma_\lambda^2} + \underbrace{\mathbb{E}\left(\varepsilon_{it}\varepsilon_{jt}\right)}_{=0} + \underbrace{\mathcal{COV}_2}_{=0} \\ &= \sigma_\lambda^2. \end{aligned}$$

where $\mathcal{COV}_2 = \operatorname{cov}(\mu_i, \lambda_t) + \operatorname{cov}(\mu_i, \lambda_t) + \operatorname{cov}(\mu_i, \varepsilon_{it}) + \operatorname{cov}(\mu_j, \varepsilon_{it}) + 2\operatorname{cov}(\lambda_t, \varepsilon_{it})$.

- Finally, the variance-covariance matrix of the error term can be described:

$$\mathbb{E}\left(u_{it}, u_{js}\right) = \begin{cases} \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\varepsilon^2 & \text{if} \quad i = j, t = s, \\ \sigma_\mu^2 & \text{if} \quad i = j, t \neq s, \\ \sigma_\lambda^2 & \text{if} \quad i \neq j, t = s, \\ 0 & \text{if} \quad i \neq j, t \neq s. \end{cases} \tag{17}$$

- Unlike the one-way RE model, the variance-covariance matrix $\mathbb{E}\left(u_{it}, u_{js}\right)$ is not block-diagonal because there is correlation between units. This correlation is caused by the the period-specific effects and equals $\sigma_\lambda^2/(\sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\varepsilon^2)$.
- Akin to the one-way RE model there is equicorrelation which is implied by the unit-specific effects and equals $\sigma_\mu^2/(\sigma_\mu^2 + \sigma_\lambda^2 + \sigma_\varepsilon^2)$.

- Consider the following transformation:

$$z_{it}^* = z_{it} - \theta_1 \bar{z}_i - \theta_2 \bar{z}_t + \theta_3 \bar{z}_{it}, \tag{18}$$

where $z_{it} \in \{y_{it}, x_{1it}, \ldots, x_{kit}, u_{it}\}$ and

$$\bar{z}_i = \frac{1}{T}\sum_t^T z_{it}, \quad \bar{y}_t = \frac{1}{N}\sum_i^N z_{it}, \quad \bar{z} = \frac{1}{NT}\sum_t^T\sum_i^N z_{it},$$

and

$$\begin{aligned}
\theta_1 &= 1 - \frac{\sigma_\varepsilon}{\sqrt{T\sigma_\mu^2 + \sigma_\varepsilon^2}} \\
\theta_2 &= 1 - \frac{\sigma_\varepsilon}{\sqrt{N\sigma_\lambda^2 + \sigma_\varepsilon^2}} \\
\theta_3 &= \theta_1 + \theta_2 + \frac{\sigma_\varepsilon}{\sqrt{T\sigma_\mu^2 + N\sigma_\lambda^2 + \sigma_\varepsilon^2}} - 1.
\end{aligned}$$

.

- **Swamy and Aurora (1972)** propose running three least squares regressions to estimate σ_ε^2, σ_μ^2 and σ_λ^2 which allow us to calculate θ_1, θ_2 and θ_3.
  1. The within regression allows to estimate the variance of the idiosyncratic component, i.e., $\hat{\sigma}_\varepsilon^2$.
  2. The between (**individuals**) regression allows to estimate the σ_μ which is the estimated variance of the error term from the following regression:

     $$\bar{y}_i - \bar{y} = \beta_1 \left(\bar{x}_{1i} - \bar{x}_1\right) + \ldots + \beta_k \left(\bar{x}_{ki} - \bar{x}_k\right) + \bar{u}_i - \bar{u},$$

     and $\hat{\sigma}_\mathcal{I}^2$ is the estimated variance of the error term from the above regression. Then, the estimated variance of the individual specific-error term can be described as follow

     $$\hat{\sigma}_\mu^2 = \frac{\hat{\sigma}_\mathcal{I}^2 - \hat{\sigma}_\varepsilon^2}{T}.$$

  3. The between (**periods**) regression allows to estimate the σ_λ which is the estimated variance of the error term from the following regression:

     $$\bar{y}_t - \bar{y} = \beta_1 \left(\bar{x}_{1t} - \bar{x}_1\right) + \ldots + \beta_k \left(\bar{x}_{kt} - \bar{x}_k\right) + \bar{u}_t - \bar{u},$$

     and $\hat{\sigma}_\mathcal{T}^2$ is the estimated variance of the error term from the above regression. Then, the estimated variance of the individual period-error term can be described as follow

     $$\hat{\sigma}_\lambda^2 = \frac{\hat{\sigma}_\mathcal{T}^2 - \hat{\sigma}_\varepsilon^2}{N}.$$

- Alternative estimators of σ_ε^2, σ_μ^2 and σ_λ^2 are proposed by Wallace and Hussain (1969), Amemiya (1971), Fuller and Battese (1974) and Nerlove (1971).
- In general, estimates of σ_ε^2, σ_μ^2 and σ_λ^2 allow to calculate θ_1, θ_2 and θ_3 and estimate model for transformed variables:

$$y_{it}^* = \beta_1 x_{1it}^* + \ldots + \beta_k x_{kit}^* + u_{it}^* \tag{19}$$

  where $z_{it}^* = z_{it} - \theta_1 \bar{z}_i - \theta_2 \bar{z}_t + \theta_3 \bar{z}_{it}$ and $z_{it} \in \{y_{it}, x_{1it}, \ldots, x_{kit}, u_{1it}\}$.
- **Computational warning:** sometimes estimates of σ_μ^2 or σ_λ^2 could be negative. This is due to the fact that we use two-step strategy rather than jointly estimation.

$$y_{it} = \alpha + \beta_1 x_{1it} + \beta_k x_{kit} + \mu_i + \lambda_t + \varepsilon_{it} \quad (20)$$

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

Non-spherical variance-covariance matrix of the error term

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

$$\mathbb{E}\left(uu'\right) = \sigma_u^2 I \tag{21}$$

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

$$\mathbb{E}\left(u_{i.}u_{i.}'\right) = \Sigma_{i,i} = \left(\sigma_\mu^2 + \sigma_\varepsilon^2\right)\begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \vdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix}, \tag{22}$$

where

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

where σ_μ^2 and σ_ε^2 stands for the variance of the individual-specific and idiosyncratic error term.

- More general case:

$$\mathbb{E}\left(uu'\right) = \Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \dots & \Sigma_{1,N} \\ \Sigma_{2,1} & \Sigma_{2,2} & \dots & \Sigma_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{N,1} & \Sigma_{N,2} & \dots & \Sigma_{N,N} \end{bmatrix}, \tag{23}$$

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:

$$\mathbb{E}\left(uu'\right) = \Sigma = \begin{bmatrix} \Sigma_{1,1} & \Sigma_{1,2} & \dots & \Sigma_{1,N} \\ \Sigma_{2,1} & \Sigma_{2,2} & \dots & \Sigma_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma_{N,1} & \Sigma_{N,2} & \dots & \Sigma_{N,N} \end{bmatrix}, \tag{23}$$

  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).

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

$$\hat{\beta}^{GLS} = \left(X'\hat{\Sigma}^{-1}X\right)^{-1} X'\hat{\Sigma}^{-1}y, \tag{24}$$

and the variance-covariance estimator:

$$Var\left(\hat{\beta}^{GLS}\right) = \left(X'\hat{\Sigma}^{-1}X\right)^{-1}. \tag{25}$$

- In the presence of the non-spherical disturbances the GLS estimator is BLUE.
- Key challenge: Σ !.

The variance-covariance matrix of the error term can be non-spherical due to:

1. Autocorrelation (serial correlation).
2. Heteroskedasticity.
3. Cross-sectional dependence.
4. Combination of the above cases.

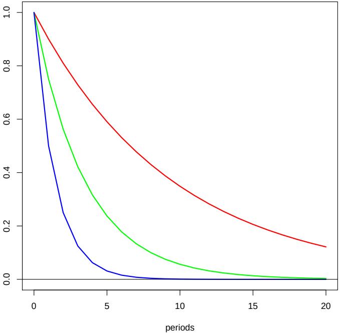
- **Autocorrelation** or **Serial correlation** is the correlation of the error term with its past values.
- **AR(1) example:**

$$u_{it} = \rho u_{it-1} + \eta_{it} \tag{26}$$

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

- Autocorrelation signals that disturbances displays some memory.
- One possible explanation for autocorrelation is that relevant factors are omitted $\implies$ omitted variables bias.

Response to a unit shock in period $t = 0$.



$\rho = 0.95$, $\rho = 0.75$, $\rho = 0.5$.

Detecting autocorrelation

- Visual inspection: plotting residuals.
- Simple regressions.
- Test proposed by Baltagi and Wu (1999).
  - The null hypothesis is about no autocorrelation:

$$\mathcal{H}_0 \quad \rho = 0. \tag{27}$$

  - The test investigates only the first-order autocorrelation.
  - The test statistics is the Durbin-Watson statistic tailored to the panel data.

General idea:

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

$$z_{it}^* = \begin{cases} (1-\hat{\rho}^2)^{\frac{1}{2}} z_{it} & \text{if} \quad t=1 \\ (1-\hat{\rho}^2)^{\frac{1}{2}} \left(z_{it} \left(\frac{1}{1-\hat{\rho}^2} \right)^{\frac{1}{2}} - z_{it-1} \left(\frac{\hat{\rho}^2}{1-\hat{\rho}^2} \right) \right) & \text{if} \quad t>1 \end{cases} \quad (28)$$

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:

$$\mathbb{E}\left(uu'\right) = \Sigma = diag\left(I\sigma^2_{u1}, \ldots, I\sigma^2_{u,N}\right) \neq I\sigma^2_u, \tag{29}$$

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.

- 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:

$$Var(\hat{\beta}) = \left(X'X\right)^{-1}\left(X'\hat{\Sigma}X\right)\left(X'X\right)^{-1} \tag{30}$$

  where $\hat{\Sigma} = diag(\hat{u}_1^2, \ldots, \hat{u}_N^2)$.
- The clustered robust standard errors:
  - All observations are divided into G groups:

$$Var(\hat{\beta}) = \left(X'X\right)^{-1}\left(\sum_{i=1}^{G} x_i'\hat{u}_i\hat{u}_i'x_i\right)\left(X'X\right)^{-1}. \tag{31}$$

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

$$\mathbb{E}\left(u_{it}u_{jt}\right) \neq 0 \quad \text{if} \quad i \neq j. \tag{32}$$

- 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.

- Consider the static panel data model:

$$y_{it} = \alpha + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + u_{it}. \tag{33}$$

- The cross-sectional correlation coefficient:

$$\hat{\rho}_{i,j} = \frac{\sum_{t=1}^{T} \hat{u}_{it}\hat{u}_{jt}}{\left(\sum_{t=1}^{T} \hat{u}_{it}\right)^{\frac{1}{2}} \left(\sum_{t=1}^{T} \hat{u}_{jt}\right)^{\frac{1}{2}}}. \tag{34}$$

  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:

$$\mathcal{H}_0 : \quad \rho_{i,j} = 0 \quad \text{if} \quad i \neq j. \tag{35}$$

$$\mathcal{H}_1 : \quad \rho_{i,j} \neq 0 \quad \text{if} \quad i \neq j \tag{36}$$

- The LM statistic (Breusch and Pagan, 1980):

$$LM = T \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{i,j}^2. \tag{37}$$

  The above 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.

- The LM statistic exhibits substantial size distortion when N is relatively large due to fact that it is not correctly centered for fixed T.

- Pesaran (2004) proposes the following test statistic:

$$CD = \sqrt{\frac{2T}{N\left(N-1\right)}}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\hat{\rho}_{i,j}\right). \tag{38}$$

  Under the null about no cross sectional dependence, the CD statistic is normally distributed, i.e., $CD \sim \mathcal{N}(0,1)$, 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:

$$CD = \sqrt{\frac{2}{N\left(N-1\right)}}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sqrt{T_{ij}}\hat{\rho}_{i,j}\right). \tag{39}$$

- Friedman's statistics:
$$FR = \frac{T-1}{(N-1)R_{ave}+1} \tag{40}$$
where R_{ave} is the average Spearman's cross-sectional correlation.
- The FR statistic is asymptotically χ^2 distributed with with $T-1$ degrees of freedom, for fixed T and and sufficiently large N.

- Friedman's statistics:

$$FR = \frac{T-1}{(N-1)R_{ave}+1} \quad (40)$$

where R_{ave} is the average Spearman's cross-sectional correlation.

- The FR statistic is asymptotically χ^2 distributed with with $T-1$ degrees of freedom, for fixed T and and sufficiently large N.
- Frees' statistics bases on the sum of the squared rank correlation coefficients R^2_{ave}:

$$FRE = N\left(R^2_{ave} - \frac{1}{T-1}\right), \quad (41)$$

where

$$R^2_{ave} = \sqrt{\frac{2}{N(N-1)}}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sqrt{T_{ij}}\hat{r}_{i,j}\right). \quad (42)$$

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 ot two χ^2 random variables).
- Both Friedman's and Frees' statistic are designed to static panel data models.