Error Component Model

Model setup as before:

$$\boldsymbol{y} = \boldsymbol{X}\,\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

Error term has a special structure:

$$\varepsilon_{it} = u_i + v_t + w_{it}$$

$$u_i \sim N(0, \sigma_u^2) \quad , \quad v_t \sim N(0, \sigma_v^2) \quad , \quad w_{it} \sim N(0, \sigma_w^2)$$

all components are uncorrelated, hence

$$Var(\varepsilon_{it}) = \sigma^2 = \sigma_u^2 + \sigma_v^2 + \sigma_w^2$$

Note:

$$E(u_i v_t) = E(u_i w_{it}) = E(v_t w_{it}) = 0$$
$$E(u_i u_j) = 0, \quad i \neq j$$
$$E(v_t v_s) = 0, \quad t \neq s$$
$$E(w_{it} w_{is}) = E(w_{it} w_{jt}) = E(w_{it} w_{js}) = 0 \quad i \neq j \; , \quad t \neq s$$

and therefore the errors are homoskedastic.

$$corr(\varepsilon_{it}, \varepsilon_{jt}) = \frac{Cov(\varepsilon_{it}, \varepsilon_{jt})}{\sqrt{Var(\varepsilon_{it})\ Var(\varepsilon_{jt})}} = \frac{\sigma_v^2}{\sigma_u^2 + \sigma_v^2 + \sigma_w^2} \qquad i \neq j$$

$$corr(\varepsilon_{it}, \varepsilon_{is}) = \frac{Cov(\varepsilon_{it}, \varepsilon_{is})}{\sqrt{Var(\varepsilon_{it})\ Var(\varepsilon_{is})}} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_v^2 + \sigma_w^2} \qquad t \neq s$$

$$corr(\varepsilon_{it}, \varepsilon_{js}) = \frac{Cov(\varepsilon_{it}, \varepsilon_{js})}{\sqrt{Var(\varepsilon_{it})\ Var(\varepsilon_{js})}} = 0 \qquad i \neq j\ , \quad t \neq s$$

$$\boldsymbol{V} = \begin{bmatrix} \sigma_u^2[\boldsymbol{A_T}] & \sigma_v^2[\boldsymbol{I_T}] & \cdots & \sigma_v^2[\boldsymbol{I_T}] \\ \sigma_v^2[\boldsymbol{I_T}] & \sigma_u^2[\boldsymbol{A_T}] & \cdots & \sigma_v^2[\boldsymbol{I_T}] \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_v^2[\boldsymbol{I_T}] & \sigma_v^2[\boldsymbol{I_T}] & \cdots & \sigma_u^2[\boldsymbol{A_T}] \end{bmatrix}$$

dimensions: NT x NT

$$\boldsymbol{A_T} = \begin{bmatrix} \frac{\sigma^2}{\sigma_u^2} & 1 & \cdots & 1 \\ 1 & \frac{\sigma^2}{\sigma_u^2} & \cdots & 1 \\ \vdots & \vdots & \cdots & \vdots \\ 1 & 1 & \cdots & \frac{\sigma^2}{\sigma_u^2} \end{bmatrix}$$

T xT

More often the error term has a simplified structure:

$$\varepsilon_{it} = u_i + w_{it}$$

$$u_i \sim N(0, \sigma_u^2) \quad , \quad w_{it} \sim N(0, \sigma_w^2)$$

all components are uncorrelated, hence

$$Var(\varepsilon_{it}) = \sigma^2 = \sigma_u^2 + \sigma_w^2$$

Note:

$$E(u_i w_{it}) = 0$$

$$E(u_i u_j) = 0, \quad i \neq j$$

$$E(w_{it} w_{is}) = E(w_{it} w_{jt}) = E(w_{it} w_{js}) = 0 \quad i \neq j \ , \quad t \neq s$$

and therefore the errors are homoskedastic.

$$corr(\varepsilon_{it}, \varepsilon_{jt}) = \frac{Cov(\varepsilon_{it}, \varepsilon_{jt})}{\sqrt{Var(\varepsilon_{it})\ Var(\varepsilon_{jt})}} = \frac{\sigma_v^2}{\sigma_u^2 + \sigma_w^2} \quad i \neq j$$

$$corr(\varepsilon_{it}, \varepsilon_{is}) = \frac{Cov(\varepsilon_{it}, \varepsilon_{is})}{\sqrt{Var(\varepsilon_{it})\ Var(\varepsilon_{is})}} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_w^2} \quad t \neq s$$

$$corr(\varepsilon_{it}, \varepsilon_{js}) = \frac{Cov(\varepsilon_{it}, \varepsilon_{js})}{\sqrt{Var(\varepsilon_{it})\ Var(\varepsilon_{js})}} = 0 \quad i \neq j \ , \quad t \neq s$$

$$
\boldsymbol{V}=\begin{bmatrix}
\sigma_u^2[\boldsymbol{A_T}] & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & \sigma_u^2[\boldsymbol{A_T}] & \cdots & \mathbf{0} \\
\vdots & \vdots & \cdots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & \sigma_u^2[\boldsymbol{A_T}]
\end{bmatrix}
$$

$$
\boldsymbol{A_T}=\begin{bmatrix}
\frac{\sigma^2}{\sigma_u^2} & 1 & \cdots & 1 \\
1 & \frac{\sigma^2}{\sigma_u^2} & \cdots & 1 \\
\vdots & \vdots & \cdots & \vdots \\
1 & 1 & \cdots & \frac{\sigma^2}{\sigma_u^2}
\end{bmatrix}
$$

T xT

dimensions: NT x NT

Transformation matrix (Cholesky decomposition) is then

$$
\boldsymbol{P}=\begin{bmatrix}
[\boldsymbol{P_T}] & \mathbf{0} & \cdots & \mathbf{0} \\
\mathbf{0} & [\boldsymbol{P_T}] & \cdots & \mathbf{0} \\
\vdots & \vdots & \cdots & \vdots \\
\mathbf{0} & \mathbf{0} & \cdots & [\boldsymbol{P_T}]
\end{bmatrix}
$$

Where

$$
\boldsymbol{P_T} = \boldsymbol{I_T} - \frac{\theta}{T}\, \boldsymbol{i_T}\, \boldsymbol{i_T^T}
$$

$$
\theta = 1 - \frac{\sigma_w}{\sqrt{T\sigma_u^2+\sigma_w^2}} \quad , \quad \boldsymbol{i_T^T} = [1\ \ 1\ \ 1\ \ 1\ \ 1\ \ \ldots\ \ 1]
$$

However, more often the representation is

$$V = \begin{bmatrix} \mathbf{\Omega} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Omega} & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Omega} \end{bmatrix} \qquad \mathbf{\Omega} = \begin{bmatrix} \sigma_w^2+\sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_w^2+\sigma_u^2 & \cdots & \sigma_u^2 \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_w^2+\sigma_u^2 \end{bmatrix}$$

dimension NT x NT T x T

Note, that

$$\mathbf{\Omega} = \sigma_w^2 \, \boldsymbol{I_T} + \sigma_w^2 \, \boldsymbol{i_T} \, \boldsymbol{i_T^T} \quad , \quad \boldsymbol{i_T^T} = [1 \;\; 1 \;\; 1 \;\; 1 \;\; 1 \; \ldots \; 1]$$

For GLS the decomposition is $\boldsymbol{V}^{-\frac{1}{2}} = \boldsymbol{I_T} \otimes \mathbf{\Omega}^{-\frac{1}{2}}$

$$\mathbf{\Omega}^{-\frac{1}{2}} = \boldsymbol{I_T} - \frac{\theta}{T} \, \boldsymbol{i_T} \, \boldsymbol{i_T^T}$$ and

$$\theta = 1 - \frac{\sigma_w}{\sqrt{T\sigma_u^2+\sigma_w^2}}$$ as before.

The transformation is therefore:

$$\boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{y_i} = \begin{bmatrix} y_{i1} - \theta \bar{y}_i \\ y_{i2} - \theta \bar{y}_i \\ y_{i3} - \theta \bar{y}_i \\ \dots \\ y_{iT} - \theta \bar{y}_i \end{bmatrix} \qquad \boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{x_{i,k}} = \begin{bmatrix} X_{i1,k} - \theta \bar{X}_{i,k} \\ X_{i2,k} - \theta \bar{X}_{i,k} \\ X_{i3,k} - \theta \bar{X}_{i,k} \\ \dots \\ X_{iT,k} - \theta \bar{X}_{i,k} \end{bmatrix}$$

where **k** denotes the k-th regressor in of all **X**'s, and bar denotes the group mean

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{it} \qquad \bar{X}_{i,k} = \frac{1}{T} \sum_{t=1}^{T} X_{it,k}$$

$$\left(Y_{it} - \theta \bar{Y}_i\right) = \beta_1 (1-\theta) + \beta_2 \left(X_{it,2} - \theta \bar{X}_{i,2}\right) +$$

$$+ \beta_3 \left(X_{it,3} - \theta \bar{X}_{i,3}\right) + \dots + \beta_K \left(X_{it,K} - \theta \bar{X}_{i,K}\right) + w_{it}^{*}$$

It can be shown that the GLS estimator is a matrix-weighted average of BETWEEN unit estimator and WITHIN unit estimator.
Illustration for one variable regression:

$$\bar{Y}_i = \beta_B \bar{X}_i + \bar{\varepsilon}_i \qquad \text{BETWEEN}$$

$$\left(Y_{it} - \bar{Y}_i\right) = \beta_W \left(X_{it} - \bar{X}_i\right) + \left(w_{it} - \bar{w}_i\right) \qquad \text{WITHIN}$$

Panel Data Models

Start with generic model where the **x** does not contain a constant:

$$y_{it} = \alpha_i + \boldsymbol{\beta}^T \boldsymbol{x}_{it} + \varepsilon_{it}$$

α_i can be Constant and the same for all i . . . . POOLED OLS
Constant but different across i . . . . FIXED EFFECT
Random draw for each i . . . . . RANDOM EFFECT

Pooled OLS

All α_i are the same and $\alpha_i = \alpha$
you already know how to deal with, apply SUR methods

Fixed effects

Each individual has a specific intercept α_i ,
LSDV - Least Square Dummy Variable method - use constant term dummy for each goup.
We have K regression coefficients and N dummies to estimate

$$\boldsymbol{y}_i = \boldsymbol{i}\alpha_i + \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\varepsilon}_i$$

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{i} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \boldsymbol{i} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix} \beta + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_N \end{bmatrix}$$

Testing fixed effects vs. pooled OLS

Classical F-test (or any other test for zero-restriction):

Hypothesis: $\alpha_i = \alpha$ for all i

$$F_{(N-1, NT-N-K)} = \frac{\dfrac{R_u^2 - R_r^2}{(N-1)}}{\dfrac{1 - R_u^2}{(NT-N-K)}}$$

WITHIN and BETWEEN group estimators

Original formulation (TOTAL):

$$y_{it} = \alpha_i + \boldsymbol{\beta}^{\boldsymbol{T}} \boldsymbol{x}_{\boldsymbol{it}} + \varepsilon_{it}$$

Deviations from the group mean (WITHIN)

$$y_{it} - \bar{y}_i = \boldsymbol{\beta}^{\boldsymbol{T}} (\boldsymbol{x}_{\boldsymbol{it}} - \bar{\boldsymbol{x}}_{\boldsymbol{i}}) + \varepsilon_{it} - \bar{\varepsilon}_i$$

In terms of the group means (BETWEEN)

$$\bar{y}_i = \alpha + \boldsymbol{\beta}^{\boldsymbol{T}} \bar{\boldsymbol{x}}_{\boldsymbol{i}} + \bar{\varepsilon}_i$$

All three models could be estimated by OLS, lets focus on β.

$$S_{xx}^{t} = \sum_{i=1}^{N} \sum_{t=1}^{T} (\boldsymbol{x}_{\boldsymbol{it}} - \bar{\boldsymbol{x}})(\boldsymbol{x}_{\boldsymbol{it}} - \bar{\boldsymbol{x}})^T$$

$$S_{xy}^{t} = \sum_{i=1}^{N} \sum_{t=1}^{T} (\boldsymbol{x}_{\boldsymbol{it}} - \bar{\boldsymbol{x}})(\boldsymbol{y}_{\boldsymbol{it}} - \bar{\boldsymbol{y}})$$

$$S_{xx}^{w} = \sum_{i=1}^{N} \sum_{t=1}^{T} (\boldsymbol{x}_{\boldsymbol{it}} - \bar{\boldsymbol{x}}_{\boldsymbol{i}})(\boldsymbol{x}_{\boldsymbol{it}} - \bar{\boldsymbol{x}}_{\boldsymbol{i}})^T$$

$$S_{xy}^{w} = \sum_{i=1}^{N} \sum_{t=1}^{T} (\boldsymbol{x}_{\boldsymbol{it}} - \bar{\boldsymbol{x}}_{\boldsymbol{i}})(\boldsymbol{y}_{\boldsymbol{it}} - \bar{\boldsymbol{y}}_{\boldsymbol{i}})$$

$$S_{xx}^{b} = \sum_{i=1}^{N} \sum_{t=1}^{T} (\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{x}})^T = \sum_{i=1}^{N} T(\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{x}})(\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{x}})^T$$

$$S_{xy}^{b} = \sum_{i=1}^{N} \sum_{t=1}^{T} (\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{x}})(\bar{\boldsymbol{y}}_{\boldsymbol{i}} - \bar{\boldsymbol{y}}) = \sum_{i=1}^{N} T(\bar{\boldsymbol{x}}_{\boldsymbol{i}} - \bar{\boldsymbol{x}})(\bar{\boldsymbol{y}}_{\boldsymbol{i}} - \bar{\boldsymbol{y}})$$

Obviously

$$S_{xy}^{t} = S_{xy}^{w} + S_{xy}^{b} \qquad S_{xx}^{t} = S_{xx}^{w} + S_{xx}^{b}$$

The estimators are:

$$b^{t} = \left[S_{xx}^{t}\right]^{-1} S_{xy}^{t} = \left[S_{xx}^{w} + S_{xx}^{b}\right]^{-1} \left[S_{xy}^{w} + S_{xy}^{b}\right]$$

$$b^{w} = \left[S_{xx}^{w}\right]^{-1} S_{xy}^{w} \qquad b^{b} = \left[S_{xx}^{b}\right]^{-1} S_{xy}^{b}$$

From this we can derive that

$$S_{xx}^{w}\, b^{w} = S_{xy}^{w} \qquad S_{xx}^{b}\, b^{b} = S_{xy}^{b}$$

Inserting these into b^t:

$$b^{t} = F^{w} b^{w} + F^{b} b^{b}$$

$$F^{w} = \left[S_{xx}^{w} + S_{xx}^{b}\right]^{-1} S_{xx}^{w} = I - F^{b}$$

The total estimator is a weighted average of BETWEEN and WITHIN

RANDOM effect model

$$y_{it} = \alpha + \boldsymbol{\beta}^{T} \boldsymbol{x}_{it} + u_i + w_{it}$$

Variance-covariance matrix is

$$\boldsymbol{V} = \begin{bmatrix} \boldsymbol{\Omega} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Omega} & \cdots & \mathbf{0} \\ \vdots & \vdots & \cdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Omega} \end{bmatrix} \qquad \boldsymbol{\Omega} = \begin{bmatrix} \sigma_w^2+\sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_w^2+\sigma_u^2 & \cdots & \sigma_u^2 \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_w^2+\sigma_u^2 \end{bmatrix}$$

dimension NT x NT T x T

Note, that

$$\boldsymbol{\Omega} = \sigma_w^2 \boldsymbol{I_T} + \sigma_w^2 \boldsymbol{i_T} \boldsymbol{i_T^T} \quad , \quad \boldsymbol{i_T^T} = [1 \; 1 \; 1 \; 1 \; 1 \; \ldots \; 1]$$

For GLS the decomposition is $\boldsymbol{V}^{-\frac{1}{2}} = \boldsymbol{I_T} \otimes \boldsymbol{\Omega}^{-\frac{1}{2}}$

$$\boldsymbol{\Omega}^{-\frac{1}{2}} = \boldsymbol{I_T} - \frac{\theta}{T} \boldsymbol{i_T} \boldsymbol{i_T^T}$$ and

$$\theta = 1 - \frac{\sigma_w}{\sqrt{T\sigma_u^2+\sigma_w^2}}$$ as before.

The Choleski decomposition gives

$$\boldsymbol{\Omega}^{-\frac{1}{2}} \boldsymbol{y_i} = \begin{bmatrix} y_{i1} - \theta \bar{y}_i \\ y_{i2} - \theta \bar{y}_i \\ y_{i3} - \theta \bar{y}_i \\ \ldots \\ y_{iT} - \theta \bar{y}_i \end{bmatrix}$$

Note, that θ=1 leads to LSDV (like if σ_w=0 and only u_i are present and there would be no difference between LSDV and RE)

Now we can get the GLS estimator as before

$$\hat{\boldsymbol{\beta}}^{RE} = \hat{\boldsymbol{F}}^{w} \boldsymbol{b}^{w} + \hat{\boldsymbol{F}}^{b} \boldsymbol{b}^{b}$$

$$\hat{\boldsymbol{\beta}}^{RE} = \hat{\boldsymbol{F}}^{w} \boldsymbol{b}^{w} + (\boldsymbol{I} - \hat{\boldsymbol{F}}^{w}) \boldsymbol{b}^{b}$$

$$\boldsymbol{F}^{w} = \left[\boldsymbol{S}_{xx}^{w} + \lambda \boldsymbol{S}_{xx}^{b} \right]^{-1} \boldsymbol{S}_{xx}^{w} = \boldsymbol{I} - \boldsymbol{F}^{b}$$

$$\lambda = (1 - \theta)^2 = \frac{\sigma_w^2}{T\sigma_u^2 + \sigma_w^2}$$

The RE estimator is again weighted average of BETWEEN and WITHIN

λ = 1 => θ =0 is OLS since σ_u=0 and hence u_i are not present
λ = 0 => θ =1 is LSDV since σ_w=0 and hence only u_i are present
for T->infinity u_i become observable

Feasible GLS

$$y_{it} = \alpha + \boldsymbol{\beta}^{\boldsymbol{T}} \boldsymbol{x}_{\boldsymbol{it}} + u_i + w_{it}$$

$$y_{it} - \bar{y}_i = \boldsymbol{\beta}^{\boldsymbol{T}} (\boldsymbol{x}_{\boldsymbol{it}} - \bar{\boldsymbol{x}}_{\boldsymbol{i}}) + w_{it} - \bar{w}_i$$

since

$$E\left(\sum_{t=1}^{T} (w_{it} - \bar{w}_i)^2 \right) = (T-1)\ \sigma_w^2$$

$$\sigma_{wi}^2 = \frac{\sum_{t=1}^{T} (w_{it} - \bar{w}_i)^2}{T - 1}$$

We do not know β, hence using sample errors:

$$\hat{\sigma}_{wi}^2 = \frac{\sum_{t=1}^{T} (w_{it} - \bar{w}_i)^2}{T - K - 1}$$

We have N such estimators, for each group one and we can average them:

$$s_w^2 = \frac{1}{N} \sum_{i=1}^{N} \hat{\sigma}_{wi}^2 = \frac{1}{N} \sum_{i=1}^{N} \frac{\sum_{t=1}^{T} (w_{it} - \bar{w}_i)^2}{T - K - 1} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (w_{it} - \bar{w}_i)^2}{NT - NK - N}$$

This assumes different coefficient for each group, having N means of y and K parameters

$$\hat{\sigma}_w^2 = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} (w_{it} - \bar{w}_i)^2}{NT - N - K}$$

Since we know that

$$Var(\varepsilon_{it}) = \sigma^2 = \sigma_u^2 + \sigma_w^2$$

we can estimate $\hat{\sigma}^2 = \dfrac{\boldsymbol{e}^T\boldsymbol{e}}{NT} = \hat{\sigma}_u^2 + \hat{\sigma}_w^2$

Testing FE vs. RE

Hypothesis: $\sigma_u^2=0$

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}\left[\sum_{t=1}^{T} w_{it}\right]^2}{\sum_{i=1}^{N}\sum_{t=1}^{T} w_{it}^2} - 1\right]^2 = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}\left(T\bar{w}_i\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T} w_{it}^2} - 1\right]^2$$

$$LM = \frac{NT}{2(T-1)}\left[\frac{\boldsymbol{e}^T\boldsymbol{D}\boldsymbol{D}^T\boldsymbol{e}}{\boldsymbol{e}^T\boldsymbol{e}} - 1\right]^2 \sim \chi_1^2$$

Hausman test:

$$W = [\boldsymbol{b} - \hat{\boldsymbol{\beta}}]^T\left[Var[\boldsymbol{b}] - Var[\hat{\boldsymbol{\beta}}]\right]^{-1}[\boldsymbol{b} - \hat{\boldsymbol{\beta}}] \sim \chi_K^2$$

where **b** are FE, β are RE coefficients and Var() appropriate v-c matrix

Dynamic Models

$$y_{it} = \alpha_i + \boldsymbol{\beta}^{\boldsymbol{T}} \boldsymbol{x}_{\boldsymbol{it}} + \delta y_{i,t-1} + \varepsilon_{it}$$

under all standard previous assumptions the ε are correlated with the fixed effect or random effect since the lagged y is correlated with ε by definition.

More transparent it is in the RE model, where **u** (which is part of α) is correlated with lagged y.

Take first differences

$$y_{it} - y_{i,t-1} = \boldsymbol{\beta}^{\boldsymbol{T}} \left(\boldsymbol{x}_{\boldsymbol{it}} - \boldsymbol{x}_{\boldsymbol{i,t-1}} \right) + \delta (y_{i,t-1} - y_{i,t-2}) + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

u (which is part of α in RE) vanished, there is still correlation between lagged y and ε (and there is MA1 process present) but all more lagged y and their differences are valid instruments.

Conlusion:
Use IV or GMM technique.

Autocorrelation

$$y_{it} = \alpha + \boldsymbol{\beta}^{\boldsymbol{T}} \boldsymbol{x}_{\boldsymbol{it}} + u_i + \varepsilon_{it}$$

and $\varepsilon_{it} = \rho\varepsilon_{i,t-1} + v_{it}$

$$y_{it} - \rho y_{i,t-1} = \alpha(1-\rho) + \boldsymbol{\beta}^{\boldsymbol{T}}\left(\boldsymbol{x}_{\boldsymbol{it}} - \rho\boldsymbol{x}_{\boldsymbol{i,t-1}}\right) + (\varepsilon_{it} - \rho\varepsilon_{i,t-1}) + u_i(1-\rho)$$

$$y_{it} - \rho y_{i,t-1} = \alpha(1-\rho) + \boldsymbol{\beta}^{\boldsymbol{T}}\left(\boldsymbol{x}_{\boldsymbol{it}} - \rho\boldsymbol{x}_{\boldsymbol{i,t-1}}\right) + v_{it} + u_i(1-\rho)$$

Having ρ we can use T-1 observations for estimation of β and adjust variances by $(1-\rho)^2$.
So estimate ρ and then use P-W or C-O transformation.