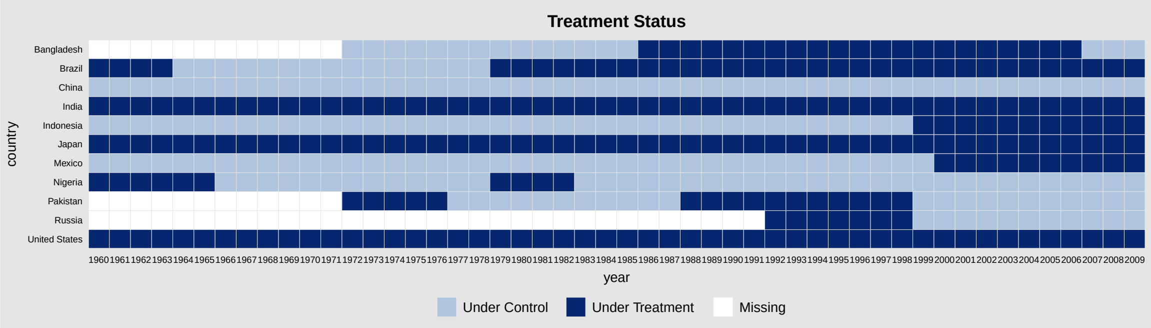
Linear Models Lecture 18: Random Effects

Robert Gulotty

University of Chicago

February 26, 2026

panelView


Treatment Status
country
Bangladesh
Brazil
China
India
Indonesia
Japan
Mexico
Nigeria
Pakistan
Russia
United States
1960 1961 1962 1963 1964 1965 1966 1967 1968 1969 1970 1971 1972 1973 1974 1975 1976 1977 1978 1979 1980 1981 1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009
year
Under Control
Under Treatment
Missing

Motivation for Random Effects/ Partial Pooling

- Given a model $y_{it} = \beta_0 + \beta_1 x_{it} + \alpha_i + e_{it}$, assume that
  1. Strict exogeneity holds: $E[e_i|\alpha_i, \boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots \boldsymbol{x}_{iT}] = 0 \quad \forall i$
  2. The unobserved features are uncorrelated with $\boldsymbol{X}$. $E[\alpha_i|\boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots \boldsymbol{x}_{iT}] = \alpha_0 \quad \forall i$.
- Pooled OLS is consistent under 1), but we can do more (derive a more efficient estimator) with assumption 2).
- Under 2) we can use the group structure to
  - rebalance the between and within estimator, dragging estimates toward group means,
  - use information about variables that do not vary within unit,
  - make predictions about unobserved groups.
- This comes at the cost of introducing bias if 2) is false, ie, we failed to control for some time-invariant characteristic.

Aitken (1935) Generalized Least Squares

- We will use group structure to model dependence between observations.

$$\text{var}[\boldsymbol{e}|\boldsymbol{X}] = \Sigma\sigma^2$$

- If we know Σ, we can do no better than to premultiply our linear model with $\Sigma^{-1/2}$,

$$\begin{aligned}
\boldsymbol{Y} &= \boldsymbol{X}\beta + \boldsymbol{e} \\
\Sigma^{-1/2}\boldsymbol{Y} &= \Sigma^{-1/2}\boldsymbol{X}\beta + \Sigma^{-1/2}\boldsymbol{e} \\
\tilde{\boldsymbol{Y}} &= \tilde{\boldsymbol{X}}\beta + \tilde{\boldsymbol{e}} \\
\tilde{\beta}_{GLS} &= (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} \\
&= ((\Sigma^{-1/2}\boldsymbol{X})'(\Sigma^{-1/2}\boldsymbol{X}))^{-1}(\Sigma^{-1/2}\boldsymbol{X})'(\Sigma^{-1/2}\boldsymbol{Y}) \\
&= (\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\Sigma^{-1}\boldsymbol{Y}
\end{aligned}$$

Here we have $E[\tilde{\beta}_{GLS}] = \beta$ and $var(\tilde{\beta}_{GLS}) = \sigma^2(\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X})^{-1}$

Aitken (1935) Generalized Least Squares

- We will use group structure to model dependence between observations.

$$\text{var}[\boldsymbol{e}|\boldsymbol{X}] = \Sigma\sigma^2$$

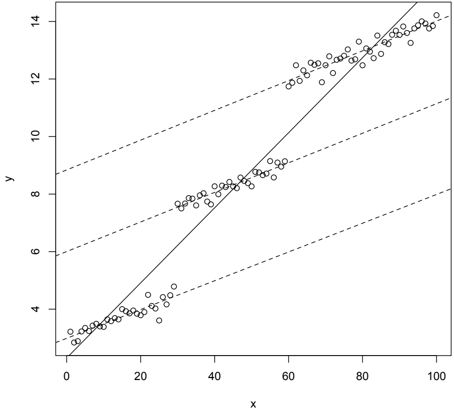
- If we know Σ, we can do no better than to premultiply our linear model with $\Sigma^{-1/2}$,

$$\begin{aligned}
\boldsymbol{Y} &= \boldsymbol{X}\beta + \boldsymbol{e} \\
\Sigma^{-1/2}\boldsymbol{Y} &= \Sigma^{-1/2}\boldsymbol{X}\beta + \Sigma^{-1/2}\boldsymbol{e} \\
\tilde{\boldsymbol{Y}} &= \tilde{\boldsymbol{X}}\beta + \tilde{\boldsymbol{e}} \\
\tilde{\beta}_{GLS} &= (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{Y}} \\
&= ((\Sigma^{-1/2}\boldsymbol{X})'(\Sigma^{-1/2}\boldsymbol{X}))^{-1}(\Sigma^{-1/2}\boldsymbol{X})'(\Sigma^{-1/2}\boldsymbol{Y}) \\
&= (\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\Sigma^{-1}\boldsymbol{Y}
\end{aligned}$$

Here we have $E[\tilde{\beta}_{GLS}] = \beta$ and $var(\tilde{\beta}_{GLS}) = \sigma^2(\boldsymbol{X}'\Sigma^{-1}\boldsymbol{X})^{-1}$

Random Effects Assumption

- The random effects assumption is that $Cov[\alpha_i, \boldsymbol{x}_{it}] = 0$.
- That is, we can consistently estimate our results in the cross-section.
- Here is a picture where that is violated.



Random Effects Assumption

- Define a composite error term $\nu_{it} = \alpha_i + e_{it}$.

$$y_{it} = \boldsymbol{\beta}' \boldsymbol{x}_{it} + \alpha_i + e_{it} = \boldsymbol{\beta}' \boldsymbol{x}_{it} + \nu_{it}$$

$$\begin{aligned} Var(\nu_{it}) &= E[(\alpha_i + e_{it})^2] = E[\alpha_i^2 + 2\alpha_i e_{it} + e_{it}^2] = \sigma_\alpha^2 + \sigma_e^2 \\ Cov(\nu_{it}\nu_{is}) &= E[\nu_{it}\nu_{is}] = E(\alpha_i + e_{it})(\alpha_i + e_{is}) \\ &= E(\alpha_i^2 + \alpha_i e_{it} + \alpha_i e_{is} + e_{it}e_{is}) = E(\alpha_i^2) = \sigma_\alpha^2 \quad \forall t \neq s \end{aligned}$$

- Recall correlation is $\rho = cov(x, y)/\sqrt{var(x)var(y)}$

$$\rho_{\nu_{it}\nu_{is}} = \frac{\sigma_\alpha^2}{\sqrt{(\sigma_\alpha^2 + \sigma_e^2)(\sigma_\alpha^2 + \sigma_e^2)}} = \frac{\sigma_\alpha^2}{(\sigma_\alpha^2 + \sigma_e^2)}$$

Random Effects Assumption

- Define a composite error term $\nu_{it} = \alpha_i + e_{it}$.

$$y_{it} = \boldsymbol{\beta}' \mathbf{x}_{it} + \alpha_i + e_{it} = \boldsymbol{\beta}' \mathbf{x}_{it} + \nu_{it}$$

$$Var(\nu_{it}) = E[(\alpha_i + e_{it})^2] = E[\alpha_i^2 + 2\alpha_i e_{it} + e_{it}^2] = \sigma_\alpha^2 + \sigma_e^2$$

$$Cov(\nu_{it}\nu_{is}) = E[\nu_{it}\nu_{is}] = E(\alpha_i + e_{it})(\alpha_i + e_{is})$$
$$= E(\alpha_i^2 + \alpha_i e_{it} + \alpha_i e_{is} + e_{it}e_{is}) = E(\alpha_i^2) = \sigma_\alpha^2 \quad \forall t \neq s$$

- Recall correlation is $\rho = cov(x,y)/\sqrt{var(x)var(y)}$

$$\rho_{\nu_{it}\nu_{is}} = \frac{\sigma_\alpha^2}{\sqrt{(\sigma_\alpha^2 + \sigma_e^2)(\sigma_\alpha^2 + \sigma_e^2)}} = \frac{\sigma_\alpha^2}{(\sigma_\alpha^2 + \sigma_e^2)}$$

Random Effects Assumption

- Define a composite error term $\nu_{it} = \alpha_i + e_{it}$.

$$y_{it} = \boldsymbol{\beta}'\boldsymbol{x}_{it} + \alpha_i + e_{it} = \boldsymbol{\beta}'\boldsymbol{x}_{it} + \nu_{it}$$

$$\begin{aligned} Var(\nu_{it}) &= E[(\alpha_i + e_{it})^2] = E[\alpha_i^2 + 2\alpha_i e_{it} + e_{it}^2] = \sigma_\alpha^2 + \sigma_e^2 \\ Cov(\nu_{it}\nu_{is}) &= E[\nu_{it}\nu_{is}] = E(\alpha_i + e_{it})(\alpha_i + e_{is}) \\ &= E(\alpha_i^2 + \alpha_i e_{it} + \alpha_i e_{is} + e_{it}e_{is}) = E(\alpha_i^2) = \sigma_\alpha^2 \quad \forall t \neq s \end{aligned}$$

- Recall correlation is $\rho = cov(x, y)/\sqrt{var(x)var(y)}$

$$\rho_{\nu_{it}\nu_{is}} = \frac{\sigma_\alpha^2}{\sqrt{(\sigma_\alpha^2 + \sigma_e^2)(\sigma_\alpha^2 + \sigma_e^2)}} = \frac{\sigma_\alpha^2}{(\sigma_\alpha^2 + \sigma_e^2)}$$

Random Effects Assumption

- Define a composite error term $\nu_{it} = \alpha_i + e_{it}$.

$$y_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \alpha_i + e_{it} = \boldsymbol{\beta}'\mathbf{x}_{it} + \nu_{it}$$

$$\begin{aligned} Var(\nu_{it}) &= E[(\alpha_i + e_{it})^2] = E[\alpha_i^2 + 2\alpha_i e_{it} + e_{it}^2] = \sigma_\alpha^2 + \sigma_e^2 \\ Cov(\nu_{it}\nu_{is}) &= E[\nu_{it}\nu_{is}] = E(\alpha_i + e_{it})(\alpha_i + e_{is}) \\ &= E(\alpha_i^2 + \alpha_i e_{it} + \alpha_i e_{is} + e_{it}e_{is}) = E(\alpha_i^2) = \sigma_\alpha^2 \quad \forall t \neq s \end{aligned}$$

- Recall correlation is $\rho = cov(x, y)/\sqrt{var(x)var(y)}$

$$\rho_{\nu_{it}\nu_{is}} = \frac{\sigma_\alpha^2}{\sqrt{(\sigma_\alpha^2 + \sigma_e^2)(\sigma_\alpha^2 + \sigma_e^2)}} = \frac{\sigma_\alpha^2}{(\sigma_\alpha^2 + \sigma_e^2)}$$

Building Blocks

Suppose that $\boldsymbol{\nu}_i = \begin{pmatrix} \nu_{i1} \\ \nu_{i2} \\ \vdots \\ \nu_{iT} \end{pmatrix}$ for a generic individual i.

$$\underset{(T\times T)}{E(\boldsymbol{\nu}_i\boldsymbol{\nu}_i')} = \begin{pmatrix} \sigma_\alpha^2+\sigma_e^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2+\sigma_e^2 & \dots & \sigma_\alpha^2 \\ \dots & & & \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2+\sigma_e^2 \end{pmatrix}$$
$$= \sigma_e^2\boldsymbol{I} + \sigma_\alpha^2\boldsymbol{ii}'$$
$$= \boldsymbol{\Omega}$$

Stacking people

Stacking individuals $\boldsymbol{\nu} = \begin{pmatrix} \boldsymbol{\nu}_1 \\ \boldsymbol{\nu}_2 \\ \vdots \\ \boldsymbol{\nu}_t \end{pmatrix}$:

$$\underset{(NT \times NT)}{E[\boldsymbol{\nu}\boldsymbol{\nu}']} = \begin{pmatrix} \boldsymbol{\Omega} & \mathbf{0} & \mathbf{0} & \dots \\ \mathbf{0} & \boldsymbol{\Omega} & \mathbf{0} & \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\Omega} & \\ \dots & & & \end{pmatrix} = \boldsymbol{\Omega} \otimes \boldsymbol{I}$$

Theoretical GLS: Random Effects Estimation

Using the fact that $var(\boldsymbol{\nu}) = \boldsymbol{\Omega} \otimes \boldsymbol{I}$, we can apply GLS

$$\hat{\beta}_{RE} = (\boldsymbol{X}'(\boldsymbol{\Omega} \otimes \boldsymbol{I})^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{\Omega} \otimes \boldsymbol{I})^{-1}\boldsymbol{y}$$

But we need estimates of σ^2_ν, σ^2_e and σ^2_α

$$\sigma^2_\nu = \sigma^2_e + \sigma^2_\alpha$$

FGLS for Random Effects

- Run OLS on pooled data and estimate residuals: $\hat{\nu}$

$$\widehat{\sigma^2_\nu} = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\nu}_{it}^2}{NT - k}$$

- We still need estimate of either σ^2_α or σ^2_e.
- Wooldridge approach: take an average of observed covariances:

$$E\left(\frac{\sum_{s=1}^{T-1} \sum_{t=s+1}^{T} \hat{\nu}_{is}\hat{\nu}_{it}}{N\frac{(T-1)T}{2} - k}\right) = \sigma^2_\alpha$$

$$\widehat{\sigma}^2_\nu - \widehat{\sigma}^2_\alpha = \widehat{\sigma}^2_e$$

Partial Pooling

- Random effects models are partial pooling:

$$\boldsymbol{b}_{RE} = \boldsymbol{\lambda} \boldsymbol{b}_W + (\boldsymbol{I} - \boldsymbol{\lambda}) \boldsymbol{b}_B$$

$$\boldsymbol{\lambda} = (\boldsymbol{S}_{xx}^W + (1 - \theta^2) \boldsymbol{S}_{xx}^B)^{-1} \boldsymbol{S}_{xx}^W$$

- Here θ is between 0 and 1. Under random effects:

$$\theta = 1 - \sqrt{\frac{\sigma_e^2}{\sigma_e^2 + T\sigma_\alpha^2}}$$

$$\theta = 1 - \sqrt{\frac{1}{1 + T(\sigma_\alpha^2 / \sigma_e^2)}}$$

Comparing Fixed vs Random Effects

- Random effects
  - Is efficient.
  - Allows analysis of time-constant variables.
  - Allows extrapolation to unobserved groups.
  - Assumes $\boldsymbol{X}$ is independent of α
  - plm(y $\sim$ x, data = dta, index = c("id", "year"), model = "random")
- Fixed effects
  - Washes out time-constant variables.
  - Uses many n degrees of freedom.
  - Allows any sort of relationship between $\boldsymbol{X}$ and α
  - plm(y $\sim$ x, data = dta, index = c("id", "year"), model = "within")

Hausman (1978) Specification Test

- **Idea:** compare FE and RE estimates of β.
  - Under H_0 (RE valid): both consistent, RE more efficient $\Rightarrow$ difference is small.
  - Under H_1 (RE invalid): RE inconsistent, FE still consistent $\Rightarrow$ systematic difference.
- **Test statistic**

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})' \left[\widehat{\text{Var}}(\hat{\beta}_{FE}) - \widehat{\text{Var}}(\hat{\beta}_{RE}) \right]^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE}) \xrightarrow{d} \chi^2_K$$

  where $K = \dim(\beta)$. Reject H_0 when H is large.
- **GMM connection:** This is a special case of the GMM endogeneity test (Hansen, Thm 13.16, Lecture 16). The GMM version replaces the classical variance difference with a robust sandwich estimator, making it valid under heteroskedasticity.

Hausman (1978) Specification Test

- **Idea:** compare FE and RE estimates of β.
  - Under H_0 (RE valid): both consistent, RE more efficient $\Rightarrow$ difference is small.
  - Under H_1 (RE invalid): RE inconsistent, FE still consistent $\Rightarrow$ systematic difference.
- **Test statistic**

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})' \left[\widehat{\mathrm{Var}}(\hat{\beta}_{FE}) - \widehat{\mathrm{Var}}(\hat{\beta}_{RE})\right]^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE}) \xrightarrow{d} \chi^2_K$$

where $K = \dim(\beta)$. Reject H_0 when H is large.

- **GMM connection:** This is a special case of the GMM endogeneity test (Hansen, Thm 13.16, Lecture 16). The GMM version replaces the classical variance difference with a robust sandwich estimator, making it valid under heteroskedasticity.

Hausman (1978) Specification Test

- **Idea:** compare FE and RE estimates of β.
  - Under H_0 (RE valid): both consistent, RE more efficient $\Rightarrow$ difference is small.
  - Under H_1 (RE invalid): RE inconsistent, FE still consistent $\Rightarrow$ systematic difference.
- **Test statistic**

$$H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})' \left[\widehat{\text{Var}}(\hat{\beta}_{FE}) - \widehat{\text{Var}}(\hat{\beta}_{RE})\right]^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE}) \xrightarrow{d} \chi^2_K$$

  where $K = \dim(\beta)$. Reject H_0 when H is large.
- **GMM connection:** This is a special case of the GMM endogeneity test (Hansen, Thm 13.16, Lecture 16). The GMM version replaces the classical variance difference with a robust sandwich estimator, making it valid under heteroskedasticity.

Hausman Test in Practice

```
fe_mod <- plm(log(gsp) ~ log(pcap) + log(hwy) + log(util),
              data = pdata, model = "within")
re_mod <- plm(log(gsp) ~ log(pcap) + log(hwy) + log(util),
              data = pdata, model = "random")

phtest(fe_mod, re_mod)   # classical Hausman chi-squared
```

- **Caveat:** This is a *specification test*, not a model-selection tool.
- Using it to choose FE vs. RE makes your estimator a *pretest estimator* $\Rightarrow$ biases subsequent inference.
- Better alternatives when unsure:
  - Use FE (conservative).
  - Use **Correlated Random Effects** (Mundlak), which nests both and supplies a robust version of the same test.

Hausman Test in Practice

```
fe_mod <- plm(log(gsp) ~ log(pcap) + log(hwy) + log(util),
              data = pdata, model = "within")
re_mod <- plm(log(gsp) ~ log(pcap) + log(hwy) + log(util),
              data = pdata, model = "random")

phtest(fe_mod, re_mod)   # classical Hausman chi-squared
```

- **Caveat:** This is a *specification test*, not a model-selection tool.
- Using it to choose FE vs. RE makes your estimator a *pretest estimator* $\Rightarrow$ biases subsequent inference.
- Better alternatives when unsure:
  - Use FE (conservative).
  - Use **Correlated Random Effects** (Mundlak), which nests both and supplies a robust version of the same test.

Time-Invariant Regressors: the Identification Problem

Panel baseline

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_i'\gamma + \alpha_i + e_{it}, \qquad i = 1, \ldots, N, \; t = 1, \ldots, T$$

- $\boldsymbol{x}_{it}$: time-varying covariates
- $\boldsymbol{z}_i$: *time-invariant* covariates
- α_i : unit fixed effect, unobserved and potentially *correlated* with $\boldsymbol{x}_{it}, \boldsymbol{z}_i$

- The within transformation eliminates α_i but also wipes out $\boldsymbol{z}_i$ (no variation over t) $\Rightarrow$ standard FE cannot identify coefficients on $\boldsymbol{z}_i$.
- E.g. we want to know the coefficient on colonial legacy, ethnic fractionalization, or distance to Russia.

Time-Invariant Regressors: the Identification Problem

Panel baseline

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_i'\gamma + \alpha_i + e_{it}, \qquad i = 1, \ldots, N, \ t = 1, \ldots, T$$

- $\boldsymbol{x}_{it}$: time-varying covariates
- $\boldsymbol{z}_i$: *time-invariant* covariates
- α_i : unit fixed effect, unobserved and potentially *correlated* with $\boldsymbol{x}_{it}, \boldsymbol{z}_i$

- The within transformation eliminates α_i but also wipes out $\boldsymbol{z}_i$ (no variation over t) $\Rightarrow$ standard FE cannot identify coefficients on $\boldsymbol{z}_i$.
- E.g. we want to know the coefficient on colonial legacy, ethnic fractionalization, or distance to Russia.

Model-Based Remedies for z_i

- **Correlated Random Effects (Mundlak, 1978)**
  - Model the dependence of α_i on the *time-varying* regressors: $\alpha_i = \delta_0 + \bar{\boldsymbol{x}}_i' \boldsymbol{\delta} + \zeta_i$, $\bar{\boldsymbol{x}}_i \equiv T^{-1} \sum_t \boldsymbol{x}_{it}$.
  - Conditional on $\bar{\boldsymbol{x}}_i$, treat ζ_i as *random*: $\mathbb{E}[\zeta_i \mid \boldsymbol{x}_{it}, \boldsymbol{z}_i] = 0$.
  - Estimation via GLS delivers FE-consistent $\hat{\beta}$ and identifies γ on $\boldsymbol{z}_i$.
- **Hausman–Taylor (1981)**
  - Partition regressors $\boldsymbol{x}_{it} = (\boldsymbol{x}_{1it}, \boldsymbol{x}_{2it})$, $\boldsymbol{z}_i = (\boldsymbol{z}_{1i}, \boldsymbol{z}_{2i})$ such that $\boldsymbol{x}_{1it}, \boldsymbol{z}_{1i}$ are exogenous, $\boldsymbol{x}_{2it}, \boldsymbol{z}_{2i}$ possibly endogenous.
  - Use $\boldsymbol{x}_{1it}$ (within variation) and $\boldsymbol{z}_{1i}$ (between variation) as instruments for $\boldsymbol{x}_{2it}, \boldsymbol{z}_{2i}$.
  - Two-step IV–GLS attains efficiency under random-effects covariance.

Model-Based Remedies for $\boldsymbol{z}_i$

- **Correlated Random Effects (Mundlak, 1978)**
  - Model the dependence of α_i on the *time-varying* regressors: $\alpha_i = \delta_0 + \bar{\boldsymbol{x}}_i' \boldsymbol{\delta} + \zeta_i$, $\bar{\boldsymbol{x}}_i \equiv T^{-1} \sum_t \boldsymbol{x}_{it}$.
  - Conditional on $\bar{\boldsymbol{x}}_i$, treat ζ_i as *random*: $\mathbb{E}[\zeta_i \mid \boldsymbol{x}_{it}, \boldsymbol{z}_i] = 0$.
  - Estimation via GLS delivers FE-consistent $\hat{\beta}$ and identifies γ on $\boldsymbol{z}_i$.
- **Hausman–Taylor (1981)**
  - Partition regressors $\boldsymbol{x}_{it} = (\boldsymbol{x}_{1it}, \boldsymbol{x}_{2it})$, $\boldsymbol{z}_i = (\boldsymbol{z}_{1i}, \boldsymbol{z}_{2i})$ such that $\boldsymbol{x}_{1it}, \boldsymbol{z}_{1i}$ are exogenous, $\boldsymbol{x}_{2it}, \boldsymbol{z}_{2i}$ possibly endogenous.
  - Use $\boldsymbol{x}_{1it}$ (within variation) and $\boldsymbol{z}_{1i}$ (between variation) as instruments for $\boldsymbol{x}_{2it}, \boldsymbol{z}_{2i}$.
  - Two-step IV–GLS attains efficiency under random-effects covariance.

Correlated Random Effects (Mundlak)

- **Structural restriction**

$$\alpha_i \;=\; \delta_0 \;+\; \bar{\boldsymbol{x}}_i'\boldsymbol{\delta} \;+\; \zeta_i, \qquad \mathbb{E}[\zeta_i \mid \boldsymbol{x}_{it}, \boldsymbol{z}_i] = 0$$

where $\bar{\boldsymbol{x}}_i = T^{-1}\sum_t \boldsymbol{x}_{it}$. All correlation between α_i and the time-varying covariates is absorbed by $\bar{\boldsymbol{x}}_i'\boldsymbol{\delta}$.

- Substituting into the panel model:

$$y_{it} \;=\; x_{it}'\beta \;+\; z_i'\gamma \;+\; \delta_0 + \bar{x}_i'\delta \;+\; \underbrace{\zeta_i + e_{it}}_{\text{composite error}}$$

- Apply GLS (random-effects weighting) to the augmented model.
  - $\hat{\beta}$ identical to FE (we will see why via FWL).
  - $\hat{\gamma}$ on z_i is now identified.
  - Testable implication: H_0: $\delta = \mathbf{0}$ reduces CRE to conventional RE.

Correlated Random Effects (Mundlak)

- **Structural restriction**

$$\alpha_i \;=\; \delta_0 \;+\; \bar{\boldsymbol{x}}_i'\boldsymbol{\delta} \;+\; \zeta_i, \qquad \mathbb{E}[\zeta_i \mid \boldsymbol{x}_{it}, \boldsymbol{z}_i] = 0$$

  where $\bar{\boldsymbol{x}}_i = T^{-1}\sum_t \boldsymbol{x}_{it}$. All correlation between α_i and the time-varying covariates is absorbed by $\bar{\boldsymbol{x}}_i'\boldsymbol{\delta}$.

- Substituting into the panel model:

$$y_{it} \;=\; \boldsymbol{x}_{it}'\boldsymbol{\beta} \;+\; \boldsymbol{z}_i'\gamma \;+\; \delta_0 + \bar{\boldsymbol{x}}_i'\boldsymbol{\delta} \;+\; \underbrace{\zeta_i + e_{it}}_{\text{composite error}}$$

- Apply GLS (random-effects weighting) to the augmented model.
  - $\hat{\beta}$ identical to FE (we will see why via FWL).
  - $\hat{\gamma}$ on $\boldsymbol{z}_i$ is now identified.
  - Testable implication: $H_0: \; \boldsymbol{\delta} = \mathbf{0}$ reduces CRE to conventional RE.

Correlated Random Effects (Mundlak)

- **Structural restriction**

$$\alpha_i = \delta_0 + \bar{\boldsymbol{x}}_i'\boldsymbol{\delta} + \zeta_i, \qquad \mathbb{E}[\zeta_i \mid \boldsymbol{x}_{it}, \boldsymbol{z}_i] = 0$$

where $\bar{\boldsymbol{x}}_i = T^{-1}\sum_t \boldsymbol{x}_{it}$. All correlation between α_i and the time-varying covariates is absorbed by $\bar{\boldsymbol{x}}_i'\boldsymbol{\delta}$.

- Substituting into the panel model:

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \boldsymbol{z}_i'\gamma + \delta_0 + \bar{\boldsymbol{x}}_i'\boldsymbol{\delta} + \underbrace{\zeta_i + e_{it}}_{\text{composite error}}$$

- Apply GLS (random-effects weighting) to the augmented model.
  - $\hat{\boldsymbol{\beta}}$ identical to FE (we will see why via FWL).
  - $\hat{\gamma}$ on $\boldsymbol{z}_i$ is now identified.
  - Testable implication: $H_0: \ \boldsymbol{\delta} = \boldsymbol{0}$ reduces CRE to conventional RE.

Why $\hat{\beta}_{CRE} = \hat{\beta}_{FE}$: FWL Intuition

- The CRE regression includes $\boldsymbol{x}_{it}$, $\bar{\boldsymbol{x}}_i$, $\boldsymbol{z}_i$, and a constant.
- By the **Frisch–Waugh–Lovell theorem**, the coefficient on $\boldsymbol{x}_{it}$ is obtained by first *partialing out* $\bar{\boldsymbol{x}}_i$ (and $\boldsymbol{z}_i$, constant) from both y_{it} and $\boldsymbol{x}_{it}$.
- Partialing $\bar{\boldsymbol{x}}_i$ from $\boldsymbol{x}_{it}$:

$$\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i = \text{within-demeaned data}$$

- So $\hat{\beta}_{CRE}$ uses *only within-unit variation* — exactly the same variation that FE uses.
- **Bonus:** CRE also identifies γ (the coefficient on $\boldsymbol{z}_i$) because ζ_i is uncorrelated with all regressors by assumption.

Why $\hat{\beta}_{CRE} = \hat{\beta}_{FE}$: FWL Intuition

- The CRE regression includes $\boldsymbol{x}_{it}$, $\bar{\boldsymbol{x}}_i$, $\boldsymbol{z}_i$, and a constant.
- By the **Frisch–Waugh–Lovell theorem**, the coefficient on $\boldsymbol{x}_{it}$ is obtained by first *partialing out* $\bar{\boldsymbol{x}}_i$ (and $\boldsymbol{z}_i$, constant) from both y_{it} and $\boldsymbol{x}_{it}$.
- Partialing $\bar{\boldsymbol{x}}_i$ from $\boldsymbol{x}_{it}$:

$$\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i = \text{within-demeaned data}$$

- So $\hat{\beta}_{CRE}$ uses *only within-unit variation* — exactly the same variation that FE uses.
- **Bonus:** CRE also identifies γ (the coefficient on $\boldsymbol{z}_i$) because ζ_i is uncorrelated with all regressors by assumption.

Why $\hat{\beta}_{CRE} = \hat{\beta}_{FE}$: FWL Intuition

- The CRE regression includes $\boldsymbol{x}_{it}$, $\bar{\boldsymbol{x}}_i$, $\boldsymbol{z}_i$, and a constant.
- By the **Frisch–Waugh–Lovell theorem**, the coefficient on $\boldsymbol{x}_{it}$ is obtained by first *partialing out* $\bar{\boldsymbol{x}}_i$ (and $\boldsymbol{z}_i$, constant) from both y_{it} and $\boldsymbol{x}_{it}$.
- Partialing $\bar{\boldsymbol{x}}_i$ from $\boldsymbol{x}_{it}$:

$$\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i \;=\; \text{within-demeaned data}$$

- So $\hat{\beta}_{CRE}$ uses *only within-unit variation* — exactly the same variation that FE uses.
- **Bonus:** CRE also identifies γ (the coefficient on $\boldsymbol{z}_i$) because ζ_i is uncorrelated with all regressors by assumption.

The Mundlak Test = Hausman Test

- From the CRE regression, test

$$H_0: \ \boldsymbol{\delta} = \mathbf{0} \qquad \text{(unit means have no additional explanatory power)}$$

- Under H_0: α_i uncorrelated with $\boldsymbol{x}_{it} \Rightarrow$ standard RE is valid.
- Under H_1: RE is inconsistent, but CRE (= FE for $\boldsymbol{\beta}$) remains consistent.
- A standard F-test (or Wald test) on $\delta = \mathbf{0}$ is **numerically equivalent** to the classical Hausman χ^2.
- **Advantage:** the CRE version
  - is easy to compute (just an F-test on extra regressors),
  - works with *robust* / *clustered* standard errors (the classical Hausman test does not).

The Mundlak Test = Hausman Test

- From the CRE regression, test

$$H_0: \boldsymbol{\delta} = \mathbf{0} \qquad \text{(unit means have no additional explanatory power)}$$

- Under H_0: α_i uncorrelated with $\boldsymbol{x}_{it} \Rightarrow$ standard RE is valid.
- Under H_1: RE is inconsistent, but CRE (= FE for $\boldsymbol{\beta}$) remains consistent.
- A standard F-test (or Wald test) on $\boldsymbol{\delta} = \mathbf{0}$ is **numerically equivalent** to the classical Hausman χ^2.
- **Advantage:** the CRE version
  - is easy to compute (just an F-test on extra regressors),
  - works with *robust* / *clustered* standard errors (the classical Hausman test does not).

R implementation: Mundlak CRE

```
library(plm)                        # panel-data toolkit
data("Produc", package = "Ecdat")# US state production data

dffrost<-data.frame( state = str_replace(toupper(rownames(state.x77)),"\\s","_"),
frost = state.x77[,"Frost"])%>%mutate(state=ifelse(state=="TENNESSEE", "TENNESSE",state))

Producweather<-Produc%>%left_join(dffrost)
pdata <- pdata.frame(Producweather, index = c("state", "year"))

pdata$pcap_bar <- Between(log(pdata$pcap))   # == ave(pcap, state)
pdata$hwy_bar <- Between(log(pdata$hwy))
pdata$util_bar<- Between(log(pdata$util))

cre_mod <- plm(log(gsp) ~ log(pcap)+log(hwy)+log(util)+pcap_bar+hwy_bar+util_bar+frost,
               data    = pdata,
               model   = "random")      # GLS w/ error-components

fe_mod <- plm(log(gsp) ~ log(pcap)+ log(hwy)+log(util),
               data    = pdata,
               model   = "within")
```

Regression results

	Model 1	Model 2
Intercept	0.14	
	(0.65)	
Time-varying covariates		
log(pcap)	2.26***	2.26***
	(0.15)	(0.15)
log(hwy)	−0.81***	−0.81***
	(0.11)	(0.11)
log(util)	−0.47***	−0.47***
	(0.08)	(0.08)
Means		
pcap_bar	−1.19	
	(0.81)	
hwy_bar	0.87	
	(0.45)	
util_bar	0.41	
	(0.40)	
Time invariant		
frost	−0.00	
	(0.00)	
σ_{idios}	0.08	
σ_{id}	0.18	
R^2	0.81	0.71
Adj. R^2	0.81	0.69
Observations	816	816

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Two-Way Mundlak (Wooldridge, 2021)

- With unit *and* time fixed effects: $y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \alpha_i + \lambda_t + e_{it}$
- **Two-way Mundlak:** include both *unit means* $\bar{\boldsymbol{x}}_{i\cdot}$ and *time means* $\bar{\boldsymbol{x}}_{\cdot t}$ as additional regressors in a pooled or RE regression.
- Delivers TWFE-equivalent $\hat{\beta}$ while allowing
  - time-invariant regressors ($\boldsymbol{z}_i$) and
  - unit-invariant regressors ($\boldsymbol{w}_t$).
- **Key application:** staggered difference-in-differences with heterogeneous treatment effects.
  - Standard TWFE can assign *negative weights* to some treatment effects.
  - Two-way Mundlak avoids this problem by explicitly modeling the dependence structure.

Two-Way Mundlak (Wooldridge, 2021)

- With unit *and* time fixed effects: $y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + \lambda_t + e_{it}$
- **Two-way Mundlak:** include both *unit means* $\bar{\boldsymbol{x}}_{i\cdot}$ and *time means* $\bar{\boldsymbol{x}}_{\cdot t}$ as additional regressors in a pooled or RE regression.
- Delivers TWFE-equivalent $\hat{\boldsymbol{\beta}}$ while allowing
  - time-invariant regressors ($\boldsymbol{z}_i$) and
  - unit-invariant regressors ($\boldsymbol{w}_t$).
- **Key application:** staggered difference-in-differences with heterogeneous treatment effects.
  - Standard TWFE can assign *negative weights* to some treatment effects.
  - Two-way Mundlak avoids this problem by explicitly modeling the dependence structure.

Two-Way Mundlak (Wooldridge, 2021)

- With unit *and* time fixed effects: $y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha_i + \lambda_t + e_{it}$
- **Two-way Mundlak:** include both *unit means* $\bar{\boldsymbol{x}}_{i\cdot}$ and *time means* $\bar{\boldsymbol{x}}_{\cdot t}$ as additional regressors in a pooled or RE regression.
- Delivers TWFE-equivalent $\hat{\boldsymbol{\beta}}$ while allowing
  - time-invariant regressors ($\boldsymbol{z}_i$) and
  - unit-invariant regressors ($\boldsymbol{w}_t$).
- **Key application:** staggered difference-in-differences with heterogeneous treatment effects.
  - Standard TWFE can assign *negative weights* to some treatment effects.
  - Two-way Mundlak avoids this problem by explicitly modeling the dependence structure.

Robust Standard errors

- Newey-West, like White with a specified lag, adjusts for autocorrelation and heteroskedasticity. `vcovNW`
- Beck-Katz, use large-T asymptotics to correct for cross-sectional dependence: good for if T/N is not too small. `vcovBK`
- Driscoll-Kraay applies Newey-West to cross-sectional dependence: good if if T/N is small. `vcovSCC`

The Dynamic Panel Problem

- Many panel models include a lagged dependent variable:

$$Y_{it} = \rho Y_{it-1} + \boldsymbol{X}_{it}'\beta + \alpha_i + e_{it}$$

- **Problem:** the within transformation creates correlation between the transformed lagged DV and the transformed error (Nickell, 1981):

$$\tilde{Y}_{it-1} = (Y_{it-1} - \bar{Y}_i) \quad \text{is correlated with} \quad \tilde{e}_{it} = (e_{it} - \bar{e}_i)$$

because $\bar{Y}_i$ contains Y_{it} which depends on e_{it}, and $\bar{e}_i$ contains e_{it-1} which affects Y_{it-1}.
- The bias is $O(1/T)$: severe when T is small (typical in micro panels), vanishes as $T \to \infty$.
- **FE is inconsistent** for fixed T with a lagged dependent variable.

The Dynamic Panel Problem

- Many panel models include a lagged dependent variable:

$$Y_{it} = \rho Y_{it-1} + \boldsymbol{X}_{it}'\boldsymbol{\beta} + \alpha_i + e_{it}$$

- **Problem:** the within transformation creates correlation between the transformed lagged DV and the transformed error (Nickell, 1981):

$$\tilde{Y}_{it-1} = (Y_{it-1} - \bar{Y}_i) \quad \text{is correlated with} \quad \tilde{e}_{it} = (e_{it} - \bar{e}_i)$$

  because $\bar{Y}_i$ contains Y_{it} which depends on e_{it}, and $\bar{e}_i$ contains e_{it-1} which affects Y_{it-1}.
- The bias is $O(1/T)$: severe when T is small (typical in micro panels), vanishes as $T \to \infty$.
- **FE is inconsistent** for fixed T with a lagged dependent variable.

The Dynamic Panel Problem

- Many panel models include a lagged dependent variable:

$$Y_{it} = \rho Y_{it-1} + \boldsymbol{X}_{it}'\boldsymbol{\beta} + \alpha_i + e_{it}$$

- **Problem:** the within transformation creates correlation between the transformed lagged DV and the transformed error (Nickell, 1981):

$$\tilde{Y}_{it-1} = (Y_{it-1} - \bar{Y}_i) \quad \text{is correlated with} \quad \tilde{e}_{it} = (e_{it} - \bar{e}_i)$$

  because $\bar{Y}_i$ contains Y_{it} which depends on e_{it}, and $\bar{e}_i$ contains e_{it-1} which affects Y_{it-1}.
- The bias is $O(1/T)$: severe when T is small (typical in micro panels), vanishes as $T \to \infty$.
- **FE is inconsistent** for fixed T with a lagged dependent variable.

Arellano-Bond: First Differencing + Instruments

- **Step 1:** First-difference to eliminate α_i:

$$\Delta Y_{it} = \rho \Delta Y_{it-1} + \Delta \boldsymbol{X}_{it}' \beta + \Delta e_{it}$$

- **Step 2:** ΔY_{it-1} is correlated with Δe_{it} (since Y_{it-1} appears in both). Use lagged levels as instruments: Y_{is} for $s \leq t-2$.
- **Moment conditions:**

$$E[Y_{is} \cdot \Delta e_{it}] = 0 \quad \text{for } s \leq t-2$$

- At $t = 3$: one instrument (Y_{i1}). At $t = 4$: two (Y_{i1}, Y_{i2}). At $t = T$: $T - 2$ instruments.
- Total moment conditions: $\frac{(T-1)(T-2)}{2}$ — the system is **overidentified** $\Rightarrow$ GMM.

Arellano-Bond: First Differencing + Instruments

- **Step 1:** First-difference to eliminate α_i:

$$\Delta Y_{it} = \rho \Delta Y_{it-1} + \Delta \boldsymbol{X}_{it}' \beta + \Delta e_{it}$$

- **Step 2:** ΔY_{it-1} is correlated with Δe_{it} (since Y_{it-1} appears in both). Use lagged levels as instruments: Y_{is} for $s \leq t-2$.
- **Moment conditions:**

$$E[Y_{is} \cdot \Delta e_{it}] = 0 \quad \text{for } s \leq t-2$$

- At $t = 3$: one instrument (Y_{i1}). At $t = 4$: two (Y_{i1}, Y_{i2}). At $t = T$: $T-2$ instruments.
- Total moment conditions: $\frac{(T-1)(T-2)}{2}$ — the system is **overidentified** $\Rightarrow$ GMM.

Arellano-Bond: First Differencing + Instruments

- **Step 1:** First-difference to eliminate α_i:

$$\Delta Y_{it} = \rho \Delta Y_{it-1} + \Delta \boldsymbol{X}_{it}' \boldsymbol{\beta} + \Delta e_{it}$$

- **Step 2:** ΔY_{it-1} is correlated with Δe_{it} (since Y_{it-1} appears in both). Use lagged levels as instruments: Y_{is} for $s \leq t-2$.
- **Moment conditions:**

$$E[Y_{is} \cdot \Delta e_{it}] = 0 \quad \text{for } s \leq t-2$$

- At $t = 3$: one instrument (Y_{i1}). At $t = 4$: two (Y_{i1}, Y_{i2}). At $t = T$: $T-2$ instruments.
- Total moment conditions: $\frac{(T-1)(T-2)}{2}$ — the system is **overidentified** $\Rightarrow$ GMM.

Arellano-Bond as GMM

- This is exactly the GMM framework from Lectures 15–16:
  1. **Step 1:** Estimate with a preliminary weight matrix (e.g., $\boldsymbol{W} = (\boldsymbol{Z}'\boldsymbol{H}\boldsymbol{Z})^{-1}$ where $\boldsymbol{H}$ is a first-difference covariance structure).
  2. **Step 2:** Re-estimate using the optimal weight matrix from step 1 residuals (two-step GMM).
- ***J*-test** (Hansen/Sargan): test the $\frac{(T-1)(T-2)}{2} - k$ overidentifying restrictions.
  - df = moments − parameters
  - Rejection $\Rightarrow$ serial correlation in e_{it} or invalid instruments.
- **Key diagnostic:** AR(2) test on differenced residuals.
  - AR(1) in Δe_{it} is expected by construction.
  - AR(2) in Δe_{it} implies AR(1) in e_{it} $\Rightarrow$ instruments invalid.

Arellano-Bond as GMM

- This is exactly the GMM framework from Lectures 15–16:
  1. **Step 1:** Estimate with a preliminary weight matrix (e.g., $\boldsymbol{W} = (\boldsymbol{Z}'\boldsymbol{H}\boldsymbol{Z})^{-1}$ where $\boldsymbol{H}$ is a first-difference covariance structure).
  2. **Step 2:** Re-estimate using the optimal weight matrix from step 1 residuals (two-step GMM).
- *J*-**test** (Hansen/Sargan): test the $\frac{(T-1)(T-2)}{2} - k$ overidentifying restrictions.
  - df = moments − parameters
  - Rejection ⇒ serial correlation in e_{it} or invalid instruments.
- **Key diagnostic:** AR(2) test on differenced residuals.
  - AR(1) in Δe_{it} is expected by construction.
  - AR(2) in Δe_{it} implies AR(1) in e_{it} ⇒ instruments invalid.

Arellano-Bond as GMM

- This is exactly the GMM framework from Lectures 15–16:
  1. **Step 1:** Estimate with a preliminary weight matrix (e.g., $\boldsymbol{W} = (\boldsymbol{Z}'\boldsymbol{H}\boldsymbol{Z})^{-1}$ where $\boldsymbol{H}$ is a first-difference covariance structure).
  2. **Step 2:** Re-estimate using the optimal weight matrix from step 1 residuals (two-step GMM).
- J-**test** (Hansen/Sargan): test the $\frac{(T-1)(T-2)}{2} - k$ overidentifying restrictions.
  - df = moments − parameters
  - Rejection $\Rightarrow$ serial correlation in e_{it} or invalid instruments.
- **Key diagnostic:** AR(2) test on differenced residuals.
  - AR(1) in Δe_{it} is expected by construction.
  - AR(2) in Δe_{it} implies AR(1) in $e_{it} \Rightarrow$ instruments invalid.

Arellano-Bond as GMM

- This is exactly the GMM framework from Lectures 15–16:
  1. **Step 1:** Estimate with a preliminary weight matrix (e.g., $\boldsymbol{W} = (\boldsymbol{Z}'\boldsymbol{H}\boldsymbol{Z})^{-1}$ where $\boldsymbol{H}$ is a first-difference covariance structure).
  2. **Step 2:** Re-estimate using the optimal weight matrix from step 1 residuals (two-step GMM).
- *J*-**test** (Hansen/Sargan): test the $\frac{(T-1)(T-2)}{2} - k$ overidentifying restrictions.
  - df = moments − parameters
  - Rejection $\Rightarrow$ serial correlation in e_{it} or invalid instruments.
- **Key diagnostic:** AR(2) test on differenced residuals.
  - AR(1) in Δe_{it} is expected by construction.
  - AR(2) in Δe_{it} implies AR(1) in e_{it} $\Rightarrow$ instruments invalid.

Extensions and Practical Guidance

- **System GMM** (Blundell-Bond, 1998): augments the differenced equations with *level equations* instrumented by lagged differences.
  - Additional moments: $E[\Delta Y_{is} \cdot (\alpha_i + e_{it})] = 0$ for $s = t - 1$.
  - Especially useful when the series is persistent (ρ near 1), where Arellano-Bond instruments are weak.
- **Practical rules:**
  - Designed for "large N, small T" panels.
  - Too many instruments $\Rightarrow$ overfitting. Collapse or limit lag depth.
  - Always report: (1) J-test p-value, (2) AR(2) test, (3) number of instruments vs. groups.
- **Software:** `plm::pgmm()` in R; `xtabond2` in Stata.
  - `pgmm(y ~ lag(y,1) + x | lag(y, 2:99), data=pdata, effect="twoways")`

Extensions and Practical Guidance

- **System GMM** (Blundell-Bond, 1998): augments the differenced equations with *level equations* instrumented by lagged differences.
  - Additional moments: $E[\Delta Y_{is} \cdot (\alpha_i + e_{it})] = 0$ for $s = t - 1$.
  - Especially useful when the series is persistent (ρ near 1), where Arellano-Bond instruments are weak.
- **Practical rules:**
  - Designed for "large N, small T" panels.
  - Too many instruments $\Rightarrow$ overfitting. Collapse or limit lag depth.
  - Always report: (1) J-test p-value, (2) AR(2) test, (3) number of instruments vs. groups.
- **Software:** `plm::pgmm()` in R; `xtabond2` in Stata.
  - `pgmm(y ~ lag(y,1) + x | lag(y, 2:99), data=pdata, effect="twoways")`

Extensions and Practical Guidance

- **System GMM** (Blundell-Bond, 1998): augments the differenced equations with *level equations* instrumented by lagged differences.
  - Additional moments: $E[\Delta Y_{is} \cdot (\alpha_i + e_{it})] = 0$ for $s = t - 1$.
  - Especially useful when the series is persistent (ρ near 1), where Arellano-Bond instruments are weak.
- **Practical rules:**
  - Designed for "large N, small T" panels.
  - Too many instruments $\Rightarrow$ overfitting. Collapse or limit lag depth.
  - Always report: (1) J-test p-value, (2) AR(2) test, (3) number of instruments vs. groups.
- **Software:** `plm::pgmm()` in R; `xtabond2` in Stata.
  - `pgmm(y ~ lag(y,1) + x | lag(y, 2:99), data=pdata, effect="twoways")`

Application: Does Income Cause Democracy?

- **Acemoglu, Johnson, Robinson & Yared (2008):** Classic question — does economic development promote democratization?
- **Model:** $\text{Democracy}_{it} = \rho\,\text{Democracy}_{it-1} + \beta\,\text{Income}_{it-1} + \alpha_i + \lambda_t + e_{it}$
- **The problem with FE:** Democracy is highly persistent ($\rho \approx 0.7$).
  - Within-transformation induces Nickell bias in $\hat{\rho}$ (biased toward zero).
  - This bias contaminates $\hat{\beta}$ as well.
  - With $T \approx 10$ (five-year panels), the bias is substantial.
- **Arellano-Bond solution:** First-difference to eliminate α_i, then instrument $\Delta\text{Democracy}_{it-1}$ with lagged levels.
- **Finding:** Once dynamic panel bias is corrected, the income effect on democracy becomes insignificant — country fixed effects explain most of the cross-national correlation.

Application: Does Income Cause Democracy?

- **Acemoglu, Johnson, Robinson & Yared (2008):** Classic question — does economic development promote democratization?
- **Model:** $\text{Democracy}_{it} = \rho \, \text{Democracy}_{it-1} + \beta \, \text{Income}_{it-1} + \alpha_i + \lambda_t + e_{it}$
- **The problem with FE:** Democracy is highly persistent ($\rho \approx 0.7$).
  - Within-transformation induces Nickell bias in $\hat{\rho}$ (biased toward zero).
  - This bias contaminates $\hat{\beta}$ as well.
  - With $T \approx 10$ (five-year panels), the bias is substantial.
- **Arellano-Bond solution:** First-difference to eliminate α_i, then instrument $\Delta\text{Democracy}_{it-1}$ with lagged levels.
- **Finding:** Once dynamic panel bias is corrected, the income effect on democracy becomes insignificant — country fixed effects explain most of the cross-national correlation.

Application: Does Income Cause Democracy?

- **Acemoglu, Johnson, Robinson & Yared (2008):** Classic question — does economic development promote democratization?
- **Model:** $\text{Democracy}_{it} = \rho\, \text{Democracy}_{it-1} + \beta\, \text{Income}_{it-1} + \alpha_i + \lambda_t + e_{it}$
- **The problem with FE:** Democracy is highly persistent ($\rho \approx 0.7$).
  - Within-transformation induces Nickell bias in $\hat{\rho}$ (biased toward zero).
  - This bias contaminates $\hat{\beta}$ as well.
  - With $T \approx 10$ (five-year panels), the bias is substantial.
- **Arellano-Bond solution:** First-difference to eliminate α_i, then instrument $\Delta\text{Democracy}_{it-1}$ with lagged levels.
- **Finding:** Once dynamic panel bias is corrected, the income effect on democracy becomes insignificant — country fixed effects explain most of the cross-national correlation.

Arellano-Bond in R: Tax Progressivity Example

```
library(plm); library(haven)
dta <- read_dta("progressTax.dta")
pdta <- pdata.frame(subset(dta, year >= 1850),
                    index = c("ccode", "year"))

# Arellano-Bond: two-step GMM
ab_mod <- pgmm(
  topratep ~ lag(topratep, 1) + himobpopyear2p
    | lag(topratep, 2:99),
  data = pdta,
  effect = "twoways",     # unit + time effects
  model = "twosteps")     # two-step GMM
summary(ab_mod)

# Key diagnostics
summary(ab_mod)$sargan   # J-test (overid)
# Check AR(2) test: p > 0.05 validates
#   the moment conditions
```

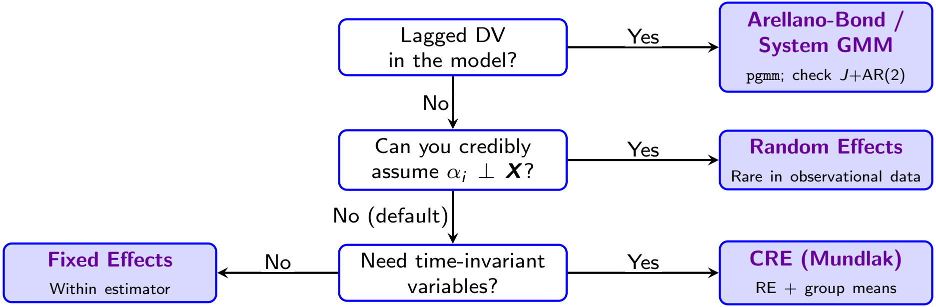
Reading Arellano-Bond Output

```
summary(ab_mod)
# Coefficients:
#                    Estimate Std. Error z-value Pr(>|z|)
# lag(topratep,1)     0.xxx     0.xxx     x.xx    0.xxx
# himobpopyear2p      0.xxx     0.xxx     x.xx    0.xxx
#
# Sargan test: chisq(df) = xx, p-value = 0.xxx
#   (large p -> instruments valid)
# Autocorrelation test (2):  z = x.xx, p = 0.xxx
#   (large p -> no AR(2) in levels)
```

Checklist for reporting:

1. $\hat{\rho}$: persistence of lagged DV (compare to FE estimate)
2. Sargan/Hansen J-test p-value > 0.05
3. AR(2) test p-value > 0.05
4. Number of instruments $<$ number of groups

Choosing a Panel Estimator



The choice of estimator should follow from your research design, not from a post-hoc test.

Why Not Let the Hausman Test Decide?

- A common workflow: estimate FE and RE, run the Hausman test, pick the "winner."
- This is pre-test estimation — your final estimator is selected by a preliminary test.
- **Problem:** the sampling distribution of the reported estimate is a *mixture* of the FE and RE distributions, weighted by the probability the test selects each one.
- Standard errors and confidence intervals ignore the selection step ⇒ undercoverage.
- **Better approach:** let your **research design** determine the estimator:
  - Is selection into units plausibly correlated with covariates? (Almost always yes with observational country-, state-, or individual-panels.) ⇒ FE or CRE.
  - Use the Hausman test as a *diagnostic*, not a decision rule. Report it alongside your preferred specification.

Why Not Let the Hausman Test Decide?

- A common workflow: estimate FE and RE, run the Hausman test, pick the "winner."
- This is pre-test estimation — your final estimator is selected by a preliminary test.
- **Problem:** the sampling distribution of the reported estimate is a *mixture* of the FE and RE distributions, weighted by the probability the test selects each one.
- Standard errors and confidence intervals ignore the selection step ⇒ undercoverage.
- **Better approach:** let your **research design** determine the estimator:
  - Is selection into units plausibly correlated with covariates? (Almost always yes with observational country-, state-, or individual-panels.) ⇒ FE or CRE.
  - Use the Hausman test as a *diagnostic*, not a decision rule. Report it alongside your preferred specification.

Why Not Let the Hausman Test Decide?

- A common workflow: estimate FE and RE, run the Hausman test, pick the "winner."
- This is pre-test estimation — your final estimator is selected by a preliminary test.
- **Problem:** the sampling distribution of the reported estimate is a *mixture* of the FE and RE distributions, weighted by the probability the test selects each one.
- Standard errors and confidence intervals ignore the selection step $\Rightarrow$ undercoverage.
- **Better approach:** let your **research design** determine the estimator:
  - Is selection into units plausibly correlated with covariates? (Almost always yes with observational country-, state-, or individual-panels.) $\Rightarrow$ FE or CRE.
  - Use the Hausman test as a *diagnostic*, not a decision rule. Report it alongside your preferred specification.

When Is Random Effects Reasonable?

RE assumes $\alpha_i \perp \boldsymbol{X}_{it}$: unit heterogeneity is unrelated to covariates.

Plausible settings in political science:

1. **Randomized/quasi-random assignment across units** — e.g., audit experiments in randomly sampled municipalities. Treatment is exogenous.
2. **Surveys with cluster sampling** — individuals within randomly selected clusters. The cluster effect is a nuisance, not a confounder.
3. **Meta-analysis** — studies are "units"; heterogeneity is modeled as random for population-average inference.

When to be skeptical (most observational panels):

- Countries self-select into trade agreements, wars, institutions.
- States adopt policies endogenously.
- Individuals sort into groups based on characteristics that also affect Y.

Default: assume α_i is correlated with $\boldsymbol{X}$ and use FE or CRE unless you can argue otherwise.

When Is Random Effects Reasonable?

RE assumes $\alpha_i \perp \boldsymbol{X}_{it}$: unit heterogeneity is unrelated to covariates.

Plausible settings in political science:

1. **Randomized/quasi-random assignment across units** — e.g., audit experiments in randomly sampled municipalities. Treatment is exogenous.
2. **Surveys with cluster sampling** — individuals within randomly selected clusters. The cluster effect is a nuisance, not a confounder.
3. **Meta-analysis** — studies are "units"; heterogeneity is modeled as random for population-average inference.

When to be skeptical (most observational panels):

- Countries self-select into trade agreements, wars, institutions.
- States adopt policies endogenously.
- Individuals sort into groups based on characteristics that also affect Y.

Default: assume α_i is correlated with $\boldsymbol{X}$ and use FE or CRE unless you can argue otherwise.

When Is Random Effects Reasonable?

RE assumes $\alpha_i \perp \boldsymbol{X}_{it}$: unit heterogeneity is unrelated to covariates.

Plausible settings in political science:

1. **Randomized/quasi-random assignment across units** — e.g., audit experiments in randomly sampled municipalities. Treatment is exogenous.
2. **Surveys with cluster sampling** — individuals within randomly selected clusters. The cluster effect is a nuisance, not a confounder.
3. **Meta-analysis** — studies are "units"; heterogeneity is modeled as random for population-average inference.

When to be skeptical (most observational panels):

- Countries self-select into trade agreements, wars, institutions.
- States adopt policies endogenously.
- Individuals sort into groups based on characteristics that also affect Y.

Default: assume α_i is correlated with $\boldsymbol{X}$ and use FE or CRE unless you can argue otherwise.

Discussion and Course Integration

- Panel estimators fit into the **unified GMM framework** from Lectures 15–16:
    - FE: within-group moment conditions (just-identified).
    - RE: adds between-group moments via GLS weighting (efficiency gain under H_0).
    - CRE (Mundlak): augmented RE that nests FE; Mundlak test $=$ Hausman test.
    - Arellano-Bond: lagged levels as instruments in first-differenced equations (overidentified $\Rightarrow$ J-test).
- You can use panel structure to improve efficiency (not discarding cross-unit variation) at the cost of bias if assumptions fail.
- Many studies focus on slow-moving processes that cannot be estimated with pure fixed effects.
- Dynamic panels require GMM-based approaches; the J-test provides a specification check that was not available with FE alone.

Discussion and Course Integration

- Panel estimators fit into the **unified GMM framework** from Lectures 15–16:
  - FE: within-group moment conditions (just-identified).
  - RE: adds between-group moments via GLS weighting (efficiency gain under H_0).
  - CRE (Mundlak): augmented RE that nests FE; Mundlak test = Hausman test.
  - Arellano-Bond: lagged levels as instruments in first-differenced equations (overidentified $\Rightarrow$ J-test).
- You can use panel structure to improve efficiency (not discarding cross-unit variation) at the cost of bias if assumptions fail.
- Many studies focus on slow-moving processes that cannot be estimated with pure fixed effects.
- Dynamic panels require GMM-based approaches; the J-test provides a specification check that was not available with FE alone.