Econometrics Lecture Notes-Panel Data Analysis

Qingfeng Liu

Otaru University of Commerce

assumptions for OLS

$$
\begin{aligned}
y_i &= \alpha + \beta x_i + \varepsilon_i \\
\hat{\beta} &= \left(\mathbf{x}'\mathbf{x}\right)^{-1} \mathbf{x}' y
\end{aligned}
$$

1. $Var\left(\varepsilon_i\right) = \sigma^2$ for $i = 1, \cdots, n$. Homogeneity.
2. $Cov\left(\varepsilon_i \varepsilon_j\right) = 0$ for $i \neq j$. No series correlation.
3. $E\left(\varepsilon_i | x_i\right) = 0$ $\left(E\left(\varepsilon_i x_i\right) = 0\right)$. Exogeneity.

properties of OLS

1. $E\left(\hat{\beta}\right) = \beta$, unbiased.
2. $\lim_{n\to\infty} \hat{\beta} = \beta$, consistency.
3. $Var\left(\hat{\beta}\right) \leq Var\left(\tilde{\beta}\right)$, where is any other linear estimator, efficiency.

What happens without assumptions 1.-3.?

1. Without assumption 1. (heteroscedasticity), $\hat{\beta}$ will be inefficient.
2. Without assumption 2. (series correlation), $\hat{\beta}$ will be inefficient.
3. Without assumption 3. (endogenity), $\hat{\beta}$ will be inconsistent.

Solutions

1. GLS for models with heteroscedasticity or series correlation.
2. IV (2SLS, GMM) for models with endogenity.

What is Panel Data

Panel data repeated observations on the same cross section, observed for several time periods. (Longitudinal data). Some examples.

1. More observations increase precision in estimation.
2. Consistent estimation of the fixed effects model, solving problem of omitted variables bias.
3. Learning more time series dynamics.

Outline

- Several models and corresponding estimation methods.
- Hausman Test.
- Empirical analysis examples.

Panel Data Models

General Expression

$$y_{it} = \alpha_{it} + \mathbf{x}_{it}\beta + \varepsilon_{it}, \ \ i = 1, \cdots, n, \ \ t = 1, \cdots, T.$$

Pooled Models

$$y_{it} = \alpha + \mathbf{x}_{it}\beta + \varepsilon_{it},$$

$$E\left(\varepsilon | \mathbf{X}\right) = \mathbf{0}$$

Estimation method for Pooled Models OLS

Panel Data Models

For now, we assume **Exogeneity**

$$E\left[\varepsilon_{it}|\alpha_i, \mathbf{x}_{i1}, \cdots, \mathbf{x}_{iT}\right] = 0, \ t = 1, \cdots, T.$$

Rondom Effect Model (RE) α_i (**Individual Effect**) is random variable and uncorrelated with $\mathbf{x}_{it}$,

$$y_{it} = \alpha_i + \mathbf{x}_{it}\beta + \varepsilon_{it},$$

Estimatin Method for RE Model Pooled OLS works well for RE model.

Panel Data Models

Fixed Effect Model (FE) α_i is random variable and correlated with $\mathbf{x}_{it}$,

$$y_{it} = \alpha_i + \mathbf{x}_{it}\beta + \varepsilon_{it},$$

Estimation method for FE Model Pooled OLS is inconsistent for FE, and does not work well for FE model.

Basic Estimation Methods for RE Model

- **Pooled OLS**
- **Between estimator**, between estimator is the OLS estimator of following equation

$$\bar{y}_i = \alpha + \bar{\mathbf{x}}_i \beta + (\alpha_i - \alpha + \bar{\varepsilon}_i)$$

where $\bar{y}_i = 1/T \sum_{t=1}^{T} y_{it}, \bar{\mathbf{x}}_i = 1/T \sum_{t=1}^{T} \mathbf{x}_{it}$ and $\bar{\varepsilon}_i = 1/T \sum_{t=1}^{T} \varepsilon_{it}$.

Basic Estimation Methods for FE Model

- **Pooled OLS** and **Between estimators** do not work well for FE model (inconsistent).
- **Within estimator,** subtract $y_{it} = \alpha_i + \mathbf{x}_{it}'\beta + \varepsilon_{it}$ from $\bar{y}_i = a_i + \overline{\mathbf{x}}_i'\beta + \bar{\varepsilon}_i$ yields

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \overline{\mathbf{x}}_i)'\,\beta + (\varepsilon_{it} - \bar{\varepsilon}_i) \tag{1}$$

LSDV

$$\begin{bmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{bmatrix} = \begin{bmatrix} \mathbf{e} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{e} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} + \begin{bmatrix} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_N \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

Within estimator is the OLS estimator of eq.(1). Notice that no α_i in eq.(1).

- Within estimator of FE model is consistent.

Basic Estimation Method for FE Model

- **First-Differences estimator,** subtracting $y_{i,t-1}$ from y_{it} yield

$$y_{it} - y_{i,t-1} = \left(\mathbf{x}_{it} - \mathbf{x}_{i,t-1}\right)' \beta + \left(\varepsilon_{it} - \varepsilon_{i,t-1}\right) \tag{2}$$

  First-Differences estimator is the OLS estimator of the above equation.

- First-Differences estimator of FE model is consistent.

Consistency and efficiency

Table 21.1. *Linear Panel Model: Common Estimators and Models*[a]

	Assumed Model		
Estimator of β	**Pooled (21.1)**	**Random Effects (21.3) and (21.5)**	**Fixed Effects (21.3) Only**
Pooled OLS (21.1)	Consistent	Consistent	Inconsistent
Between (21.7)	Consistent	Consistent	Inconsistent
Within (or Fixed Effects) (21.8)	Consistent	Consistent	Consistent
First Differences (21.9)	Consistent	Consistent	Consistent
Random Effects (21.10)	Consistent	Consistent	Inconsistent

[a] This table considers only consistency of estimators of β. For correct computation of standard errors see Section 21.2.3.

GLS Estimator for RE Models (RE Estimator) I.

RE Model

$$\begin{aligned} y_{it} &= \alpha_i + \mathbf{X}'_{it}\beta + \varepsilon_{it} \\ y_{it} &= \mathbf{X}'_{it}\beta + (\alpha_i + \varepsilon_{it}) \\ y_{it} &= \mathbf{X}'_{it}\beta + u_{it} \end{aligned}$$

where $u_{it} = (\alpha_i + \varepsilon_{it})$

u_{it} is series correlated over t OLS is not efficient

$$\text{Cov}[(\alpha_i + \varepsilon_{it}), (\alpha_i + \varepsilon_{is})] = \begin{cases} \sigma_\alpha^2, & t \neq s, \\ \sigma_\alpha^2 + \sigma_\varepsilon^2, & t = s. \end{cases}$$

GLS Estimator for RE Models (RE Estimator) II.

GLS Estimator Use GLS to deal with series correlation

$$\Omega^{-1/2}\mathbf{y}_i = \Omega^{-1/2}\mathbf{W}_i\delta + \Omega^{-1/2}\left(\alpha_i + \boldsymbol{\varepsilon}_i\right)$$

$$y_{it} - \hat{\lambda}\bar{y}_i = \left(1 - \hat{\lambda}\right)\mu + \left(\mathbf{x}_{it} - \hat{\lambda}\bar{\mathbf{x}}_i\right)'\beta + v_{it}, \qquad (3)$$

where $v_{it} = \left(1 - \hat{\lambda}\right)\alpha_i + \left(\varepsilon_{it} - \hat{\lambda}\hat{\varepsilon}_i\right)$ and $\hat{\lambda}$ is consistent estimator for $\lambda = 1 - \sigma_\varepsilon / \left(T\sigma_\alpha^2 + \sigma_\varepsilon^2\right)^{1/2}$.

- The feasible **GLS (FGLS) estimator** is the OLS estimator of β in model (3).
- **FGLS** is consistent and full efficient for RE models, but is inconsistent for FE models.

Hausman Test for Panel Data Models I.

How to distinguish between RE and FE Use **Hausman Test**.

Idea Using the fact that **Within estimator** is consistent and **FGLS** is inconsistent for FE models.

Hausman Test for Panel Data Models II.

Hausman Test statistics

1. When α_i and ε_{it} both are *i.i.d.*

$$\mathrm{H} = \left(\widetilde{\boldsymbol{\beta}}_{1,\mathrm{RE}} - \widehat{\boldsymbol{\beta}}_{1,\mathrm{W}}\right)' \left[\widehat{\mathrm{V}}[\widehat{\boldsymbol{\beta}}_{1,\mathrm{W}}] - \widehat{\mathrm{V}}[\widetilde{\boldsymbol{\beta}}_{1,\mathrm{RE}}]\right]^{-1} \left(\widetilde{\boldsymbol{\beta}}_{1,\mathrm{RE}} - \widehat{\boldsymbol{\beta}}_{1,\mathrm{W}}\right),$$

2. Otherwise, use Robust Hausman Test

$$\mathrm{H}_{\mathrm{Robust}} = \left(\widetilde{\boldsymbol{\beta}}_{1,\mathrm{RE}} - \widehat{\boldsymbol{\beta}}_{1,\mathrm{W}}\right)' \left[\widehat{\mathrm{V}}_{\mathrm{Boot}}[\widetilde{\boldsymbol{\beta}}_{1,\mathrm{RE}} - \widehat{\boldsymbol{\beta}}_{1,\mathrm{W}}]\right]^{-1} \left(\widetilde{\boldsymbol{\beta}}_{1,\mathrm{RE}} - \widehat{\boldsymbol{\beta}}_{1,\mathrm{W}}\right),$$

$$\widehat{\mathrm{V}}_{\mathrm{Boot}}[\widetilde{\boldsymbol{\beta}}_{1,\mathrm{RE}} - \widehat{\boldsymbol{\beta}}_{1,\mathrm{W}}] = \frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\boldsymbol{\delta}}_b - \overline{\widehat{\boldsymbol{\delta}}}\right)\left(\widehat{\boldsymbol{\delta}}_b - \overline{\widehat{\boldsymbol{\delta}}}\right)', \tag{4}$$

3. Ignoring whether α_i and ε_{it} both are *i.i.d.*, just use Robust Hausman Test at all times.

Within vs. First Differences Estimator (FD) for FE Models I.

Within

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)' \beta + (\varepsilon_{it} - \bar{\varepsilon}_i) \tag{5}$$

$$\begin{aligned} v_t &= (\varepsilon_{it} - \bar{\varepsilon}_i) \\ cov(v_t, v_{t-1}) &= E((\varepsilon_{it} - \bar{\varepsilon}_i)(\varepsilon_{i,t-1} - \bar{\varepsilon}_i)) \\ &= 0 - \frac{\sigma_\varepsilon^2}{T} - \frac{\sigma_\varepsilon^2}{T} + \frac{\sigma_\varepsilon^2}{T} \\ &= -\frac{\sigma_\varepsilon^2}{T} \end{aligned}$$

Within vs. First Differences Estimator (FD) for FE Models II.

FD

$$y_{it} - y_{i,t-1} = \left(\mathbf{x}_{it} - \mathbf{x}_{i,t-1}\right)' \beta + \left(\varepsilon_{it} - \varepsilon_{i,t-1}\right) \tag{6}$$

$$\begin{aligned} \varpi_t &= \left(\varepsilon_{it} - \varepsilon_{i,t-1}\right) \\ cov\left(\varpi_t, \varpi_{t-1}\right) &= E\left(\left(\varepsilon_{it} - \varepsilon_{i,t-1}\right)\left(\varepsilon_{i,t-1} - \varepsilon_{i,t-2}\right)\right) \\ &= 0 - 0 - \sigma_\varepsilon^2 - 0 \\ &= \sigma_\varepsilon^2 \end{aligned}$$

Time-invariant variable in FE models

- $$y_{it} = \alpha_i + \beta_1 x_{1t} + \beta_2 d_{it} + \varepsilon_{it}$$

  when d_{it} is time-invariant,$d_{it} = d_i$, then

  $$y_{it} = \alpha_i + \beta_1 x_{1t} + \beta_2 d_i + \varepsilon_{it}$$

- The β_2 can not be estimated by Within or FD.
- If we want to estimate β_2, other methods are required (see following slides).

Panel robust estimate of asymptotic variance

Genaral expression of Panel Data models

$$\tilde{\mathbf{y}}_i = \tilde{\mathbf{W}}_i\theta + \mathbf{u}_i$$

Three panel robust estimates of asymptotic variance

$E\left(u_{it}|w_{it}\right) = 0$, u_{it} are independent over i, $V\left(u_{it}\right)$ and $cov\left(u_{it}, u_{is}\right)$ both are time-variant (means heteroscedastic and series dependent)

1.

$$\widehat{\mathrm{V}}[\widehat{\boldsymbol{\theta}}_{\mathrm{OLS}}] = \left[\sum_{i=1}^{N} \widetilde{\mathbf{W}}_i{}'\widetilde{\mathbf{W}}_i\right]^{-1} \sum_{i=1}^{N} \widetilde{\mathbf{W}}_i{}'\widehat{\mathbf{u}}_i\widehat{\mathbf{u}}_i'\widetilde{\mathbf{W}}_i \left[\sum_{i=1}^{N} \widetilde{\mathbf{W}}_i{}'\widetilde{\mathbf{W}}_i\right]^{-1},$$

Panel robust estimate of asymptotic variance

2.

$$\widehat{\mathrm{V}}[\widehat{\boldsymbol{\theta}}_{\mathrm{OLS}}] = \left[\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\mathbf{w}}_{it}\widetilde{\mathbf{w}}_{it}'\right]^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{s=1}^{T}\widetilde{\mathbf{w}}_{it}\widetilde{\mathbf{w}}_{is}'\widehat{u}_{it}\widehat{u}_{is}\left[\sum_{i=1}^{N}\sum_{t=1}^{T}\widetilde{\mathbf{w}}_{it}\widetilde{\mathbf{w}}_{it}'\right]^{-1},$$

Panel robust estimate of asymptotic variance III.

3. Bootstrap estimator

$$\widehat{\mathrm{V}}_{\mathrm{Boot}}[\widehat{\theta}] = \frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\theta}_b - \overline{\widehat{\theta}}\right)\left(\widehat{\theta}_b - \overline{\widehat{\theta}}\right)',$$

Conditional MLE

Assuming normal distribution, **MLE** gets the same results of **Within** estimator

$$\begin{aligned} \mathrm{L}_{\mathrm{COND}}(\boldsymbol{\beta},\sigma^2,\boldsymbol{\alpha}) &= \prod_{i=1}^{N} f(y_{i1},\ldots,y_{iT}|\bar{y}_i) \\ &= \prod_{i=1}^{N} \frac{f(y_{i1},\ldots,y_{iT},\bar{y}_i)}{f(\bar{y}_i)} \\ &= \prod_{i=1}^{N} \frac{(2\pi\sigma^2)^{-T/2}}{(2\pi\sigma^2/T)^{-1/2}} \exp\left\{\sum_{t=1}^{T} -[(y_{it}-\mathbf{x}_{it}'\boldsymbol{\beta})^2+(\bar{y}_i-\bar{\mathbf{x}}_i'\boldsymbol{\beta})^2]/2\sigma^2\right\}. \end{aligned} \tag{21.38}$$

Estimate alpha using Within Estimate

$$\widehat{\alpha}_i = \bar{y}_i - \bar{\mathbf{x}}_i' \widehat{\beta}_{\mathrm{W}}, \quad i = 1, \ldots, N.$$

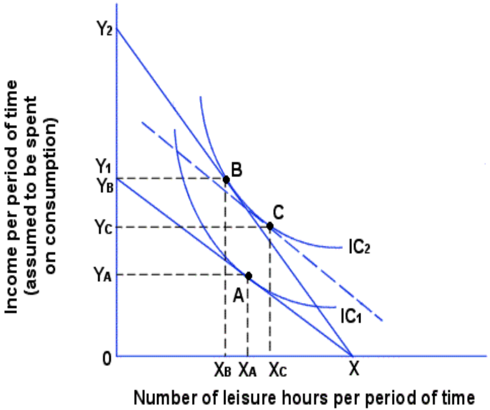
T must be large for consistency of $\hat{\alpha}_i$.

Other Issues

- Unbalanced panel data
- Measurement error

Example of application, labor supply to wages

- The data on 532 males for each of the 10 years from 1979 to 1988. $\ln hrs_{it} = \alpha_i + \beta \ln wgs_{it} + \varepsilon_{it}$



	POLS	Between	Within	First Diff	RE–GLS	RE–MLE
α	7.442	7.483	7.220	.001	7.346	7.346
β	.083	.067	.168	.109	.119	.120
Robust se[b]	(.030)	(.024)	(.085)	(.084)	(.051)	(.052)
Boot se	[.030]	[.019]	[.084]	[.083]	[.056]	[.058]
Default se	{.009}	{.020}	{.019}	{.021}	{.014}	{.014}
R^2	.015	.021	.016	.008	.014	.014
RMSE	.283	.177	.233	.296	.233	.233
RSS	427.225	0.363	259.398	417.944	288.860	288.612
TSS	433.831	17.015	263.677	420.223	293.023	292.773
σ_α	.000		.181		.161	.162
σ_ε	.283		.232		.233	.233
λ	0.000	–	1.000	–	.585	.586
N	5320	532	5320	4788	5320	5320

Assumption

Endogenous Variables or lagged dependent variables as regressors

$$E\left[\varepsilon_{it} \middle| \alpha_i, \mathbf{x}_{i1}, \cdots, \mathbf{x}_{iT}\right] \neq 0, \ \ t = 1, \cdots, T.$$

Hereinafter, ε_{it} will be denoted as u_{it},

$$E\left[u_{it} \middle| \alpha_i, \mathbf{x}_{i1}, \cdots, \mathbf{x}_{iT}\right] \neq 0, \ \ t = 1, \cdots, T.$$

This assumption cause inconsistencies with some aforementioned estimators.

Estimation method under Endogenous setting

Panel GMM without α_i

For simplicity, consider about models without individual effect α_i

$$\begin{aligned} y_{it} &= x_{it}\beta + u_{it} \\ \mathbf{y}_i &= \mathbf{X}_i\beta + \mathbf{u}_i. \end{aligned}$$

$$\mathbf{y}_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix}; \quad \mathbf{X}_i = \begin{bmatrix} \mathbf{x}'_{i1} \\ \vdots \\ \mathbf{x}'_{iT} \end{bmatrix}; \quad \mathbf{u}_i = \begin{bmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{bmatrix}.$$

Insrumental Variable Assume Z_i satisfying $E\left[\mathbf{Z}_i'\mathbf{u}_i\right]' = 0$, $E\left[\mathbf{Z}_i'\mathbf{X}_i\right]' \neq 0$ exist, Z_i can be used as instrumental variables.

Moment Condition

$$Q_N(\beta) = \left[\sum_{i=1}^{N} \mathbf{Z}_i'\mathbf{u}_i\right]' \mathbf{W}_N \left[\sum_{i=1}^{N} \mathbf{Z}_i'\mathbf{u}_i\right],$$

GMM Estimator

$$\hat{\beta}_{PGMM} = \arg\min Q_N(\beta)$$

$$\widehat{\beta}_{\mathrm{PGMM}} = \left[\left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{Z}_i\right)\mathbf{W}_N\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{X}_i\right)\right]^{-1}\left(\sum_{i=1}^{N}\mathbf{X}_i'\mathbf{Z}_i\right)\mathbf{W}_N\left(\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{y}_i\right)$$

$$\widehat{\mathrm{V}}[\widehat{\beta}_{\mathrm{PGMM}}] = [\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_N(N\widehat{\mathbf{S}})\mathbf{W}_N'\mathbf{Z}'\mathbf{X}[\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}]^{-1}$$

where $\hat{S}$ is a consistent estimator of

$$\mathbf{S} = \mathrm{plim}\,\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i,$$

2SLS (One-Step GMM)

$$\widehat{\beta}_{2SLS} = [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}.$$

2-Step GMM (2SGMM)

$$\widehat{\beta}_{2SGMM} = [\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}\widehat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{y}.$$

where $\hat{S}$ is a estimator of S from 2SLS. 2SGMM is **more efficient** than One-Step GMM.

GMM for Models with Individual Effects

First-Differences GMM

Model assume Weak Exogeneity Assumption:
$E\left[z_{is}\varepsilon_{it}\right] = 0, s \leq t,$

$$y_{it} = \alpha_i + x_{it}'\beta + \varepsilon_{it}.$$

1. Take first-differences

$$y_{it} - y_{i,t-1} = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})'\boldsymbol{\beta} + (\varepsilon_{it} - \varepsilon_{i,t-1}), \quad t = 2, \ldots, T.$$

2.

$$y_{it}^* = x_{it}^{*\prime}\beta + \varepsilon_{it}^* \tag{7}$$

3. Carry out *GMM* with model (7).

Remark: For $s \leq t-1$, we have $E\left(Z_{is}\left(\varepsilon_{it} - \varepsilon_{i,t-1}\right)\right) = E\left(Z_{is}\varepsilon_{it}^*\right) = 0$, but for $s = t$, $E\left(Z_{is}\varepsilon_{it}^*\right) \neq 0$, hence, $z_{i,t-1}, z_{i,t-2}, \cdots$ can be used as instrument for x_{it}, however, z_{it} can not.

GMM for Models with Individual Effects

Within GMM

- Under the same assumption in last slide, since $E\left[z_{it}\bar{\varepsilon}_i\right] \neq 0$, GMM does not work.
- If $E\left[z_{is}\varepsilon_{it}\right] = 0$ for all s and t, GMM works. But the assumption $E\left[z_{is}\varepsilon_{it}\right] = 0$ for all s and t, is too strong.

GMM for Models with Individual Effects

GMM for Random Effect Model

New meaning of "Random Effect Model" If Z_i satisfying $E\left(\mathbf{Z}_i'\left(a_i+\varepsilon_i\right)\right)=0$ exists, we call the model as **random effect model**.

GMM for RE models Defining $\mathbf{u}_i=a_i+\varepsilon_i$, we can use $E\left(\mathbf{Z}_i'\mathbf{u}_i\right)=E\left(\mathbf{Z}_i'\left(a_i+\varepsilon_i\right)\right)=0$ as moment condition and carry out GMM estimation.

Table 22.1. *Panel Exogeneity Assumptions and Resulting Instruments*

Exogeneity Assumption	Moment Condition	Instrument Vector[a]
Summation	$\mathrm{E}\left[\sum_t \mathbf{z}_{it} u_{it}\right] = \mathbf{0}$	$[\mathbf{z}_{it}]$
Contemporaneous	$\mathrm{E}[\mathbf{z}_{it} u_{it}] = \mathbf{0}$, all t	$[\mathbf{0}'_{r_1} \cdots \mathbf{0}'_{r_{t-1}}\ \mathbf{z}'_{it}\ \mathbf{0}'_{r_{t+1}} \cdots \mathbf{0}'_{r_T}]$
Weak	$\mathrm{E}[\mathbf{z}_{is} u_{it}] = \mathbf{0}$, $s \leq t$, all t	$[\mathbf{0}'_{r_1} \cdots \mathbf{0}'_{r_{t-1}}\ (\mathbf{z}^t_{it})'\ \mathbf{0}'_{r_{t+1}} \cdots \mathbf{0}'_{r_T}]$
Strong	$\mathrm{E}[\mathbf{z}_{is} u_{it}] = \mathbf{0}$, all s and t	$[\mathbf{0}'_{r_1} \cdots \mathbf{0}'_{r_{t-1}}\ (\mathbf{z}^T_{it})'\ \mathbf{0}'_{r_{t+1}} \cdots \mathbf{0}'_{r_T}]$

Application Example

Hours and Wages

$$\Delta \text{lnhrs}_{it} = \beta_1 \Delta \text{lnwg}_{it} + \beta_2 \Delta \text{kids}_{it} + \beta_3 \Delta \text{age}_{it} + \beta_4 \Delta \text{agesq}_{it} + \beta_5 \Delta \text{disab}_{it} + \Delta u_{it}.$$

Table 22.2. *Hours and Wages: GMM-IV Linear Panel Model Estimators*[a]

		Base Case		Stacked	
	OLS	**2SLS**	**2SGMM**	**2SLS**	**2SGM**
β_1	0.112	0.209	0.547	0.543	0.330
Panel se	(.096)	(.374)	(.327)	(.209)	(.110
Het se	[.079]	[.423]	[–]	[.226]	[–]
Default se	{.023}	{.389}	{ – }	{.169}	{ – }
RMSE	.283	.296	.307	.307	.298
Instruments	5	9	9	72	72
OIR Test	–	–	5.45	–	69.51
dof	–	–	4	–	67
p-value	–	–	.244	–	.393
N	4256	4256	4256	4256	4256

Dynamic Panel Models

Dynamic Panel Model

$$\begin{aligned} y_{it} &= \gamma y_{i,t-1} + \mathbf{x}_{it}\beta + \alpha_i + \varepsilon_{it} \\ i &= 1, \cdots, N, \quad t = 1, \cdots, T. \ |\gamma| < 1 \end{aligned}$$

Endogenity Since $y_{i,t-1}$ is a function of α_i

$$y_{i,t-1} = \gamma y_{i,t-2} + \mathbf{x}_{i,t-1}\beta + \alpha_i + \varepsilon_{i,t-1}$$

$$E\left(y_{i,t-1}\alpha_i\right) \neq 0$$

Estimator for Dynamic Panel Models

Arellano–Bond Estimator

- Take difference

$$(y_{it} - y_{i,t-1}) = \gamma (y_{i,t-1} - y_{i,t-2}) + (\mathbf{x}_{it} - \mathbf{x}_{i,t-1}) \beta + (\varepsilon_{it} - \varepsilon_{i,t-1})$$

- Use $y_{i,t-2}$ as instrumental variable
- Perform GMM

$$\widehat{\beta}_{\mathrm{AB}} = \left[\left(\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i' \mathbf{Z}_i \right) \mathbf{W}_N \left(\sum_{i=1}^{N} \mathbf{Z}_i' \widetilde{\mathbf{X}}_i \right) \right]^{-1} \left(\sum_{i=1}^{N} \widetilde{\mathbf{X}}_i' \mathbf{Z}_i \right) \mathbf{W}_N \left(\sum_{i=1}^{N} \mathbf{Z}_i' \widetilde{\mathbf{y}}_i \right).$$