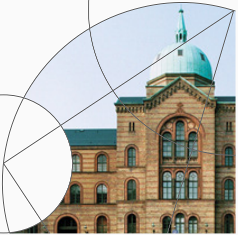
Dynamics in Linear Panel Data

Advanced Microeconometrics

Anders Munk-Nielsen

2022



Intro

Where are we in the course?

Part	Topic	Parameterization non-linear	Estimation non-linear	Dimension $\dim(x)$	Numerical optimization	M-estimation (Part III)	Outcome (y_i)	Panel (c_i)
I	OLS	$\div$	$\div$	low	$\div$	$\checkmark$	$\mathbb{R}$	$\checkmark$
II	LASSO	$\div$	$\checkmark$	high	$\checkmark$	$\div$	$\mathbb{R}$	$\div$
IV	Probit	$\checkmark$	$\checkmark$	low	$\checkmark$	$\checkmark$	$\{0,1\}$	$\div$
	Tobit	$\checkmark$	$\checkmark$	low	$\checkmark$	$\checkmark$	$[0;\infty)$	$\div$
	Logit	$\checkmark$	$\checkmark$	low	$\checkmark$	$\checkmark$	$\{1,2,...,J\}$	$\div$
	Sample selection	$\checkmark$	$\checkmark$	low	$\checkmark$	$\checkmark$	$\mathbb{R}$ and $\{0,1\}$	$\div$
	Simulated Likelihood	$\checkmark$	$\checkmark$	low	$\checkmark$	$\checkmark$	Any	$\checkmark$
	Quantile Regression	$\div$	$\checkmark$	(low)	$\checkmark$	$\checkmark$	$\mathbb{R}$	$\div$
	Non-parametric	$\checkmark$	$(\checkmark)$	∞	$\div$	$\div$	$\mathbb{R}$	$\div$

Discussion

The case for active labor market policies

Data shows that there is a *correlation* between low earnings and exposure to long spells of unemployment. Many countries have programmes to push unemployed into work. Randomly pick a side and argue with your neighbor:

1. ALPs prevent a self-reinforcing negative spiral (e.g. self-confidence, human capital, network, ...)
2. ALPs are wasteful due to selection (e.g. meaningless bureaucratic tasks for those unable to work)

Relate the discussion to the Dynamic FE model below.

Dynamic FE Model

$$y_{it} = \rho y_{it-1} + c_i + u_{it}, \quad t = 1, ..., T.$$

Discussion

Dynamic FE Model

$$y_{it} = \rho y_{it-1} + c_i + u_{it}, \quad t = 1, ..., T.$$

Discuss

What's observably different between

1. High σ_c and low ρ, ("unobserved heterogeneity")
2. High ρ, low σ_c ("true state dependence").

What is the role of unemployment insurance depending on which regime is true?

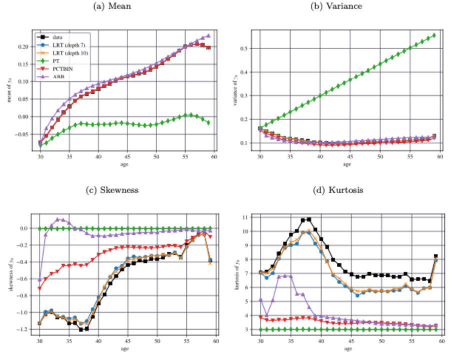
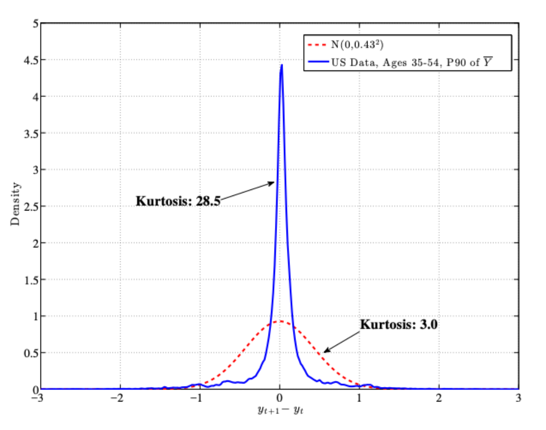


Figure 3.1: Log income over the life-cycle

- First four moments of the cross-sectional income distribution over the life-cycle.
- Black = data, green = model used in nearly all macro models.

Source: Druedahl & Munk-Nielsen (2020)

New insights into old issues

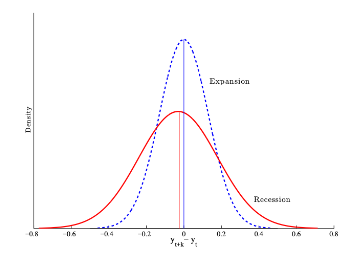


- Income growth is a lot more heavy-tailed than a log normal can produce.

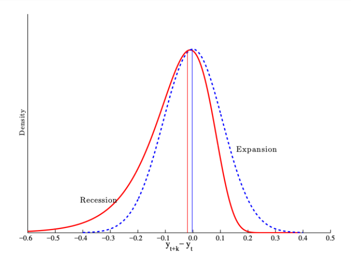
Source: Guvenen 2016

Business cycle

Myth: countercyclical variance



Fact: procyclical skewness

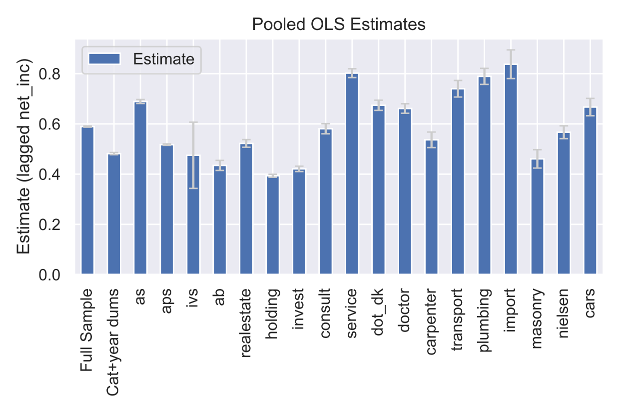


- **Interpretation:**
  - Myth: in recessions, the (symmetric) income risk increases.
  - Fact: in recessions, you risk a large drop and probably will not get a bonus.

Source: Guvenen 2016

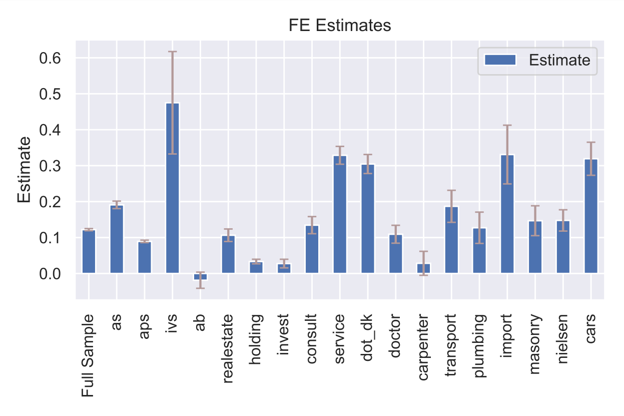
Firm Profit Persistence

Figure 1: AR(1) Estimate: POLS



Firm Profit Persistence: FE

Figure 2: AR(1) Estimate: FE



Agenda

- **Dynamic panel data models:** when we introduce a lagged outcome, y_{it-1}, (AR(1)) we cannot use FE or FD.
- **Introduction** to panel data GMM: how to instrument in panel data settings.
- **The Arellano-Bond Estimator:** FD estimation of an AR(1) model with fixed effects using instruments.

Dynamic Models

Dynamic RE Model

$$y_{it} = \rho y_{it-1} + \mathbf{x}_{it}\beta + c_i + u_{it}.$$

- **What's new?** Lagged outcome variable, regressors are $\mathbf{z}_{it} = (y_{it-1}, \mathbf{x}_{it})$
- **100$ question:** does this invalidate the RE / FE models?
- **RE:** if $\mathrm{E}(c_i \mathbf{x}_{it}) \neq \mathbf{0}$, we cannot use RE or POLS.
- **Strict exogeneity:** $\mathrm{E}(u_{it}|\mathbf{Z}_i, c_i) = 0 \Rightarrow$ FE/FD are consistent.
  - Generally, we assume $c_i \perp\!\!\!\perp u_{it}$, so $\mathrm{E}(u_{it}|\mathbf{Z}_i, c_i) = \mathrm{E}(u_{it}|\mathbf{Z}_i)$ ("$\perp\!\!\!\perp$" signifies independence)
- **Problem:** y_{it} is in $\mathbf{z}_{it+1}$, and y_{it} is a function of u_{it}, so future regressors have information on the current outcome...
  - ... proved mathematically on next slide $\rightarrow$

Strict Exogeneity Breakdown

- **Model:** consider the dynamic panel model

$$y_{it} = \rho y_{it-1} + \mathbf{x}_{it}\beta + c_i + u_{it},$$

defining $\mathbf{z}_{it} \equiv (y_{it-1}, \mathbf{x}_{it})$.

- **Proposition:** strict exogeneity (SE), $\mathrm{E}(u_{it}|\mathbf{z}_{i1}, ..., \mathbf{z}_{iT}) = 0$, breaks down when y_{it-1} is a regressor.
- **Proof:** Note that if SE holds, then in particular $\mathrm{E}(u_{it}\mathbf{z}_{is}) = \mathbf{0}$ must hold for all (t, s).
- **Consider** $s = t + 1$. Then $\mathbf{z}_{it+1} = (y_{it}, \mathbf{x}_{it+1})$, so

$$\begin{aligned} \mathrm{E}(u_{it}y_{it}) &= \mathrm{E}\left[u_{it}\left(\rho y_{it-1} + \mathbf{x}_{it+1}\beta + c_i + u_{it}\right)\right] \\ &= \mathrm{E}(u_{it}^2) = \sigma_u^2 > 0. \quad \blacksquare \end{aligned}$$

Dynamic RE Model

$$y_{it} = \rho y_{it-1} + \mathbf{x}_{it}\beta + c_i + u_{it}.$$

- **Insight:** Strict exogeneity is violated by construction with a lagged regressor.
- **Moreover:** $\mathrm{E}(y_{it-1}c_i) \neq 0$ (write out the equation for period $t-1$)
  - $\Rightarrow$ unobserved effects induce **spurious state dependence**:
    "Even with $\rho = 0$, y_{it} will be persistent due to c_i."
- $\Rightarrow$ POLS is inconsistent.

Feedback effects

Static model with feedback

$$\begin{aligned} y_{it} &= \mathbf{z}_{it}\beta + \delta h_{it} + c_i + u_{it} \\ h_{it} &= \mathbf{z}_{it}\xi + \eta y_{it-1} + r_{it} \end{aligned}$$

- **Claim:** Strict exogeneity is violated by h_{it}.
- **Proof:** take h_{it+1} and u_{it}:

$$\begin{aligned} \mathrm{E}(u_{it} h_{it+1}) &= \mathrm{E}\left[(\mathbf{z}_{it+1}\xi + \eta y_{it} + r_{it+1}) u_{it}\right] \\ &= \ldots + \eta \underbrace{\mathrm{E}(u_{it} u_{it})}_{\neq 0} + \ldots \quad \blacksquare \end{aligned}$$

Recall: Exogeneity Conditions for FE

FE: requires $\mathrm{E}(u_{it}|\mathbf{z}_i) = 0$ (strict exogeneity) for consistency.

- **Why?** The transformed error term, $\ddot{u}_{it} \equiv u_{it} - \overline{u}_i$, must be exogenous to transformed regressors, $\ddot{\mathbf{z}}_{it} \equiv \mathbf{z}_{it} - \overline{\mathbf{z}}_i$.
- Writing out:

$$\begin{aligned} \mathrm{E}(\ddot{\mathbf{z}}_{it}' \ddot{u}_{it}) &= \mathrm{E}\left[(\mathbf{z}_{it} - \overline{\mathbf{z}}_i)' (u_{it} - \overline{u}_i) \right] \\ &= \mathrm{E}(\mathbf{z}_{it}' u_{it}) + \mathrm{E}(\overline{\mathbf{z}}_i' \overline{u}_i) - \mathrm{E}(\overline{\mathbf{z}}_i' u_{it}) - \mathrm{E}(\mathbf{z}_{it}' \overline{u}_i). \end{aligned}$$

- Take e.g. $\mathrm{E}(\overline{\mathbf{z}}_i' u_{it}) = T^{-1} \sum_{s=1}^{T} \mathrm{E}(\mathbf{z}_{is}' u_{it})$.
  - Hence, if $\mathrm{E}(u_{it} \mathbf{z}_{is}) \neq 0$ for some s, then $\mathrm{E}(\overline{\mathbf{z}}_i' u_{it}) \neq 0$.

FD exogeneity condition

FD: only requires $\mathrm{E}(u_{it}|\mathbf{z}_{it+1}, \mathbf{z}_{it}, \mathbf{z}_{it-1}) = 0$.

- **Why?** Same idea – transformed errors, $\Delta\mathbf{z}_{it}$ and Δu_{it}, must be uncorrelated.
- Writing out

$$\begin{aligned}\mathrm{E}(\Delta\mathbf{z}'_{it}\Delta u_{it}) &= \mathrm{E}(\mathbf{z}'_{it}u_{it}) - \mathrm{E}(\mathbf{z}'_{it}u_{it-1}) \\ &\quad -\mathrm{E}(\mathbf{z}'_{it-1}u_{it}) + \mathrm{E}(\mathbf{z}'_{it-1}u_{it-1}).\end{aligned}$$

- $\Rightarrow$ it is enough to have

$$\mathrm{E}(u_{it}|\mathbf{z}_{it+1}, \mathbf{z}_{it}, \mathbf{z}_{it-1}) = 0,$$

  - because $\mathrm{E}(u_{it}|\mathbf{z}_{it}, \mathbf{z}_{it-1}) = 0 \Rightarrow \mathrm{E}(u_{it}\mathbf{z}_{it}) = \mathbf{0} \wedge \mathrm{E}(u_{it}\mathbf{z}_{it-1}) = \mathbf{0}$
  - and $\mathrm{E}(u_{it-1}|\mathbf{z}_{it}) = 0 \Rightarrow \mathrm{E}(u_{it-1}\mathbf{z}_{it}) = \mathbf{0}$
  - **Contemporaneous exogeneity** only implies $\mathrm{E}(\mathbf{z}_{it}u_{it}) = \mathbf{0}$ and $\mathrm{E}(\mathbf{z}_{it-1}u_{it-1}) = \mathbf{0}$.

FD Example

AR(1) model

$$y_{it} = \rho y_{it-1} + c_i + u_{it}, \quad u_{it} \sim \text{IID}(0, \sigma_u^2), c_i \sim \text{IID}(0, \sigma_\alpha^2).$$

- **Transforming** with FD, we get

$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it}.$$

- **Check exogeneity:**

$$\begin{aligned} \text{E}(\Delta y_{it-1} \Delta u_{it}) &= \text{E}(y_{it-1} u_{it}) - \text{E}(y_{it-1} u_{it-1}) \\ &\quad - \text{E}(y_{it-2} u_{it}) + \text{E}(y_{it-2} u_{it-1}). \end{aligned}$$

  - Enough to show that $\text{E}(y_{it-1} u_{it-1}) \neq 0$ to show $\text{E}(\Delta y_{it-1} \Delta u_{it}) \neq 0$.

FD Example

- **Proposition:** $\mathrm{E}(y_{it-1}u_{it-1}) \neq 0$
- **Proof:** note that

$$\begin{aligned}
\mathrm{E}(y_{it-1}u_{it-1}) &= \mathrm{E}\left[(\rho y_{it-2} + c_i + u_{it-1})u_{it-1}\right] \\
&= \underbrace{\mathrm{E}(\rho y_{it-2}u_{it-1})}_{=0 \quad (*)} + \underbrace{\mathrm{E}(c_i u_{it-1})}_{=0} + \underbrace{\mathrm{E}(u_{it-1}^2)}_{=\sigma_u^2}
\end{aligned}$$

  - where $\mathrm{E}(c_i u_{it}) = 0$ for all t (by independence assumption)
- **In conclusion:** $\mathrm{E}(y_{it-1}u_{it-1}) = \sigma_u^2 > 0$. ■

(*) **Bonus:** consider the first term

$$\begin{aligned}
\mathrm{E}(\rho y_{it-2}u_{it-1}) &= \rho\mathrm{E}\left[(\rho y_{it-3} + c_i + u_{it-2})u_{it-1}\right] \\
&= \rho^2\mathrm{E}(y_{it-3}u_{it-1}) = \ldots = \rho^{t-1}\mathrm{E}(y_{i1}u_{it-1}).
\end{aligned}$$

- **The initial condition** (y_{i1}) must be treated *differently*, since there is no y_{i0}.
  - Either $y_{i1} = \rho y_{i0} + \mathbf{x}_{i1}'\beta + c_i + u_{i1}$, where y_{i0} is just an IID draw,
  - or $y_{i0} = 0$.
  - Either way, $\mathrm{E}(y_{i1}u_{it-1}) = 0$.

Conclusion: Consistency

Assumption	Consistency of FE	FD	AB
$\mathrm{E}(u_{it}\|\mathbf{z}_i) = 0$ (strict exogeneity)	✓	✓	✓
$\mathrm{E}(u_{it}\|\mathbf{z}_{it}) = 0$ (contemporaneous exogeneity)	÷	÷	÷
$\mathrm{E}(u_{it}\|\mathbf{z}_{it}, \mathbf{z}_{it-1}, ..., \mathbf{z}_{i1}) = 0$ (sequential exogeneity)	÷	÷	✓
$\mathrm{E}(u_{it}\|\mathbf{z}_{it-1}, \mathbf{z}_{it}, \mathbf{z}_{it+1}) = 0$	÷	✓	✓

- Here, ✓ means "consistent" and ÷ means "inconsistent."
- **Arellano Bond (AB):** possible to recover consistency by using instruments...
  - A GMM estimator.

Sequential Exogeneity

Assumption: Sequential Exogeneity

$$\mathrm{E}(u_{it}|\mathbf{z}_{i1}, ..., \mathbf{z}_{it}) = \mathbf{0}.$$

- **Proposition:** Both the FE and FD estimators are

Weak dependence

- **Def.: Weak dependence:** $\{(\mathbf{z}_{it}, u_{it})\}_{t=1}^{\infty}$ is weakly dependent if correlations die out.
- **FE:** the inconsistency of FE is of order $O(T^{-1})$ as $T \to \infty$
  - That is, it dies out.
- **FD**: the inconsistency of FD does not die out.
- **However:** In this course, we assume *fixed* T *but* $N \to \infty$, so we do not care much about this (apart from mathematical joy)

GMM

Panel GMM

- **Model:**

$$y_{it} = \mathbf{x}_{it}\beta + u_{it}$$

- **Stacking** over T gives

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{u}_i.$$

- **Assume** we have access to a set of **instruments**, $\mathbf{Z}_i$ ($T \times r$ with $r \geq K$), that are **exogenous**, i.e.

$$\mathrm{E}(\mathbf{Z}_i'\mathbf{u}_i) = \mathbf{0}_{r\times 1}.$$

  - These are our **moment conditions**.
  - **Example:** Suppose $K = r = 1$. Then the moment conditions are $\mathrm{E}(z_{it}u_{it}) = 0$ for $t = 1, ..., T$.

- **Idea:** Replace "E" with "$N^{-1}\sum_i$" and $\mathbf{u}_i$ with $\mathbf{y}_i - \mathbf{X}_i\beta$, and choose β to minimize

$$\|\textstyle\sum_{i=1}^{N}\mathbf{Z}_i(\mathbf{y}_i - \mathbf{X}_i\beta) - \mathbf{0}_{r\times 1}\|_W,$$

where $\|\mathbf{a}\|_\mathbf{W} \equiv \mathbf{a}'\mathbf{W}\mathbf{a}$.

GMM Estimator

Panel GMM Estimator

$$\hat{\beta}_{\text{PGMM}} = \arg\min_{\boldsymbol{\beta}} Q_N(\beta),$$

$$\text{where } Q_N(\beta) = \left[\sum_{i=1}^{N} \mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\beta)\right]' \mathbf{W}_N \left[\sum_{i=1}^{N} \mathbf{Z}_i'(\mathbf{y}_i - \mathbf{X}_i\beta)\right].$$

- $\mathbf{W}_N$ $(r \times r)$ is a ***weighting matrix***
  - **Default** (in practice): $\mathbf{W}_N = \mathbf{I}_{r\times r}$ (identity matrix).
  - **Implication:** affects **efficiency** but **not consistency**.
- **Note:** OLS is $\min \sum_i (\cdot)^2$; GMM is $\min[\sum_i (\cdot)]^2$ (when $\mathbf{W}_N = \mathbf{I}_{r\times r}$)
- **Argmin?** the argument (input) that minimizes
  - Here: turns out to be solvable in closed form
  - Later: solve more general problems

Panel GMM Estimator

$$\hat{\beta}_{\text{PGMM}} = \arg\min_{\boldsymbol{\beta}} \left[\sum_{i=1}^{N} \mathbf{Z}_i(\mathbf{y}_i - \mathbf{X}_i\beta)\right]' \mathbf{W}_N \left[\sum_{i=1}^{N} \mathbf{Z}_i(\mathbf{y}_i - \mathbf{X}_i\beta)\right].$$

- **Turns out** that the linearity implies a **closed form solution** to the minimization problem.
- **Example:** if $K = r = T = 1$ and $\mathbf{W}_N = 1$

$$\min_{\beta} \left[\sum_{i=1}^{N} z_i(y_i - x_i\beta)\right]^2$$

$$\begin{aligned} \Rightarrow \text{FOC} : 2\left[\sum_i z_i(y_i - x_i\hat{\beta})\right]\left[\sum_i z_i(-x_i)\right] &= 0 \\ \Leftrightarrow \left(\sum_i z_i x_i\right)\left(\sum_i z_i x_i\right)\hat{\beta} &= \left(\sum_i z_i x_i\right)\left(\sum_i z_i y_i\right) \end{aligned}$$

$$\Leftrightarrow \hat{\beta} = \left[\left(\sum_i z_i x_i\right)\left(\sum_i z_i x_i\right)\right]^{-1}\left(\sum_i z_i x_i\right)\left(\sum_i z_i y_i\right).$$

Linear GMM

Linear Panel GMM Estimator

$$\hat{\beta}_{\text{PGMM}} = \left[\left(\sum_i \mathbf{X}_i'\mathbf{Z}_i\right) \mathbf{W}_N \left(\sum_i \mathbf{Z}_i'\mathbf{X}_i\right)\right]^{-1} \left(\sum_i \mathbf{X}_i'\mathbf{Z}_i\right) \mathbf{W}_N \left(\sum_i \mathbf{Z}_i'\mathbf{y}_i\right).$$

- **In matrix** form, with $\overbrace{\mathbf{X}}^{NT\times K}, \overbrace{\mathbf{Z}}^{NT\times r}, \overbrace{\mathbf{W}_N}^{r\times r}, \overbrace{\mathbf{Y}}^{NT\times 1}$

$$\hat{\beta}_{\text{PGMM}} = (\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{Y}).$$

1-step GMM (2SLS)

1-step GMM

Estimate $\hat{\beta}_{PGMM}$ using $\mathbf{W}_N = (N^{-1} \sum_i \mathbf{Z}_i'\mathbf{Z}_i)^{-1}$.

- **Motivation:** Identical to running 2-stage least squares (2SLS):
  1. Regress $\mathbf{X}_i$ on $\mathbf{Z}_i$, compute prediction $\hat{\mathbf{X}}_i = \hat{\gamma}\mathbf{Z}_i = (\sum_i \mathbf{Z}_i'\mathbf{Z}_i)^{-1} \left(\sum_i \mathbf{Z}_i'\mathbf{X}_i\right) \mathbf{Z}_i$,
  2. Regress $\mathbf{y}_i$ on $\hat{\mathbf{X}}_i$.
- **Efficiency** occurs when $\mathbf{u}_i$ are IID (conditional on $\mathbf{Z}_i$), i.e. $\mathbf{u}_i|\mathbf{Z}_i \sim \text{IID}(0, \sigma^2 \mathbf{I}_{T\times T})$.
- **Fun linear algebra:** Let $\mathbf{P}_Z \equiv \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ (projection matrix, idempotent and symmetric)
  - Then $\hat{\mathbf{X}}_i = \mathbf{P}_Z\mathbf{Z}_i$ and the estimator becomes

$$\hat{\beta}_{\text{PGMM}} = \hat{\beta}_{\text{2SLS}} = (\mathbf{X}'\mathbf{P}_Z\mathbf{X})^{-1}(\mathbf{X}'\mathbf{P}_Z\mathbf{Y}).$$

  - **Note:** if $\mathbf{Z}_i = \mathbf{X}_i$ (no instruments), then $\mathbf{P}_Z\mathbf{X} = \mathbf{X}$ (predicting $\mathbf{X}$ with itself) and $\mathbf{X}'\mathbf{P}_Z = (\mathbf{P}_\mathbf{Z}'\mathbf{X}) = \mathbf{X}$, so it simplifies to OLS.
  - **Problem:** with $NT > 100,000$, $\mathbf{P}_Z$ takes more memory than is available on a typical laptop to store.

2-step GMM

1. Obtain a consistent (but inefficient) estimate of β using $\hat{\beta}_{\text{1step}}$ and compute $\hat{\mathbf{u}}_i \equiv \mathbf{y}_i - \mathbf{X}_i\beta$.
2. Compute the estimate

$$\hat{\beta}_{\text{2step}} = \left[\left(\textstyle\sum_i \mathbf{X}_i'\mathbf{Z}_i\right)\hat{\mathbf{S}}^{-1}\left(\textstyle\sum_i \mathbf{Z}_i'\mathbf{X}_i\right)\right]^{-1}\left(\textstyle\sum_i \mathbf{X}_i'\mathbf{Z}_i\right)\hat{\mathbf{S}}^{-1}\left(\textstyle\sum_i \mathbf{Z}_i'\mathbf{y}_i\right).$$

$$\text{where } \hat{\mathbf{S}} = N^{-1}\textstyle\sum_{i=1}^{N} \mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i \quad (r \times r).$$

- **Efficiency?** It turns out that we minimize $\mathrm{V}(\hat{\beta}_{\text{PGMM}})$ by setting $\mathbf{W}_N = \mathbf{S}^{-1}$.
  - **Intuition:** put less weight on imprecise instruments.
  - **However:** it might be possible to improve on a 2-step procedure with an imprecise 1st stage...
    [not covered here]
- **Hence:** if $\hat{\mathbf{S}}$ is consistent for $\mathbf{S} \equiv \mathrm{E}(\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i)$, then using $\mathbf{W}_N = \hat{\mathbf{S}}^{-1}$ works best.

PGMM Variance

- **The PGMM** estimator is defined as

$$\hat{\beta}_{\text{PGMM}} = \left[\left(\sum_i \mathbf{X}_i'\mathbf{Z}_i\right)\mathbf{W}_N\left(\sum_i \mathbf{Z}_i'\mathbf{X}_i\right)\right]^{-1}\left(\sum_i \mathbf{X}_i'\mathbf{Z}_i\right)\mathbf{W}_N\left(\sum_i \mathbf{Z}_i'\mathbf{y}_i\right).$$

- Recall our moment conditions, $\text{E}(\mathbf{Z}_i'\mathbf{u}_i) = \mathbf{0}_{r\times 1}$.
- **CLT** gives

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbf{Z}_i'\mathbf{u}_i \xrightarrow{d} \mathcal{N}(\mathbf{0},\mathbf{S}),$$

  - where independence over i gives $\mathbf{S} = \text{E}(\mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i)$
- **Intuition** for variance: let
$\mathbf{C} \equiv \left[\left(\sum_i \mathbf{X}_i'\mathbf{Z}_i\right)\mathbf{W}_N\left(\sum_i \mathbf{X}_i'\mathbf{Z}_i\right)\right]^{-1}\left(\sum_i \mathbf{X}_i'\mathbf{Z}_i\right)\mathbf{W}_N.$

$$\text{V}(\hat{\beta}_{\text{PGMM}}|\mathbf{X},\mathbf{Z}) = \mathbf{CSC}'.$$

Variance Estimation

Panel-robust PGMM Variance Estimator

$$\hat{\mathrm{V}}(\hat{\beta}_{\text{PGMM}}) = N\left(\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\mathbf{W}_N\hat{\mathbf{S}}\mathbf{W}_N\mathbf{Z}'\mathbf{X}\left(\mathbf{X}'\mathbf{Z}\mathbf{W}_N\mathbf{Z}'\mathbf{X}\right)^{-1}$$

- **2-step GMM:** with $\mathbf{W}_N^{\text{opt}} = \hat{\mathbf{S}}^{-1} \equiv N^{-1}\sum_{i=1}^{N}\mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i$, this simplifies

$$\hat{\mathbf{V}}(\hat{\beta}_{2\text{SGMM}}) = N\left(\mathbf{X}'\mathbf{Z}\mathbf{W}_N^{\text{opt}}\mathbf{Z}'\mathbf{X}\right)^{-1}.$$

  - Note: if $\hat{\mathbf{S}}$ is not normalized by N^{-1}, the N in front disappears.
- **Efficiency:** 2-step GMM is optimal *given* $E(\mathbf{Z}'\mathbf{u}) = \mathbf{0}_{T\times 1}$.
- **More generally:** we might have $\mathrm{E}(\mathbf{u}|\mathbf{Z}) = \mathbf{0}_{T\times 1}$...
  - ... this condition implies $\mathrm{E}(\mathbf{Z}'\mathbf{u}) = \mathbf{0}$, but also $\mathrm{E}[f(\mathbf{Z})'\mathbf{u}] = \mathbf{0}$
  - I.e. $\mathbf{u}$ is uncorrelated with any function of $\mathbf{Z}$.

Overidentifying Restrictions

- **Question:** can we test whether $\mathrm{E}(\mathbf{Z}_i'\mathbf{u}_i) = \mathbf{0}_{r\times 1}$ is satisfied?
- **Just identified case:** when $r = K$, we are able to set $\frac{1}{N}\sum_i \mathbf{Z}_i'\hat{\mathbf{u}}_i = \mathbf{0}_{r\times 1}$ exactly.
  - I.e. we can make the instruments and the residual exactly uncorrelated.
    - (like how the OLS residual is uncorrelated with the residual by construction)
- When $r > K$, we have $r - K$ instruments "too many."
  - We say that "we have $r - K$ **overidentifying restrictions**"
  - $\Rightarrow$ we might be unable to make $\frac{1}{N}\sum_i \mathbf{Z}_i'\hat{\mathbf{u}}_i$ zero
- **The null:** $\mathcal{H}_0 : \mathrm{E}(\mathbf{Z}_i'\mathbf{u}_i) = \mathbf{0}_{r\times 1}$, so all moments are valid.
- **Alternative:** many things can cause $\mathrm{E}(\mathbf{Z}_i'\mathbf{u}_i) \neq \mathbf{0}_{r\times 1}$, including
  - Incorrect functional form,
  - Endogeneity for one or more of our instruments,
  - etc.
- **Intuition:** if all moments are valid but some are not needed, the true β can satisfy them all... up to sampling noise.

The Sargen Test of Overidentifying Restrictions

Under the null hypothesis, $\mathcal{H}_0 : \mathrm{E}(\mathbf{Z}_i'\mathbf{u}_i) = \mathbf{0}$, and for $\mathbf{Z}$ satisfying the rank condition, the test statistic

$$\mathrm{OIR} \equiv \left(\sum_i \mathbf{Z}_i'\hat{\mathbf{u}}_i\right)' (N\hat{\mathbf{S}})^{-1} \left(\sum_i \mathbf{Z}_i'\hat{\mathbf{u}}_i\right)$$

satisfies that

$$\mathrm{OIR} \xrightarrow{d} \chi^2(r-K).$$

- **Robustness:** $\hat{\mathbf{S}} = N^{-1}\sum_{i=1}^N \mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i$ is robust to
  - Heteroskedasticity,
  - Serial correlation of u_{it} over t (within a given i).

Arellano-Bond

Arellano Bond

Model

$$y_{it} = \rho y_{it-1} + \mathbf{x}_{it}\beta + c_i + u_{it}.$$

We assume that $u_{it} \sim \text{IID}(0, \sigma_u^2)$.

- **Motivation:** we showed that strict exogeneity is invalidated *by construction*
- **Goal:** recover consistency for FD under *sequential* exogeneity
  - FE cannot be saved, strict exogeneity is required.
  - FD is salvageable.

FD Estimator

Model in First Differences

$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta \mathbf{x}'_{ti} \beta + \Delta u_{it}.$$

- **The problem:** by construction, $\Delta y_{it-1} \equiv y_{it-1} + y_{it-2}$ is correlated with $\Delta u_{it} \equiv u_{it} - u_{it-1}$.
- **Solution:** use y_{it-2} as an instrument in period it.
  - I.e. use $\mathrm{E}\left(y_{it-2}\Delta u_{it}\right) = 0$ as an orthogonality condition.
- **Validity** comes from the fact that u_{it} are IID and thus serially uncorrelated:

$$\begin{aligned}\mathrm{E}\left(y_{it-2}\Delta u_{it}\right) &= \mathrm{E}\left[\left(\rho y_{it-3} + \mathbf{x}_{it-2}\beta + c_i + u_{it-2}\right)\left(u_{it} - u_{it-1}\right)\right] \\ &= \rho \mathrm{E}\left(y_{it-3}\Delta u_{it}\right) = \ldots = \rho^{t-2-1}\mathrm{E}\left(y_{i1}\Delta u_{it}\right) = 0\end{aligned}$$

  - since our data has no y_{i0}.

More Instruments for Δy_{it-1}?

- **So far:** shown that $\mathrm{E}(y_{it-2}\Delta u_{it}) = 0$ is valid.
- **Turns out:** $\mathrm{E}(y_{is}\Delta u_{it}) = 0$ for all $s \leq t-2$.
  - No instruments available for $t = 1, 2$.
- **Telescoping** list of instruments:

$$\begin{aligned}
\mathbf{z}_{i3} &= \begin{pmatrix} y_{i1} \end{pmatrix} \\
\mathbf{z}_{i4} &= \begin{pmatrix} y_{i1} & y_{i2} \end{pmatrix} \\
\mathbf{z}_{i5} &= \begin{pmatrix} y_{i1} & y_{i2} & y_{i3} \end{pmatrix} \\
\mathbf{z}_{i6} &= \begin{pmatrix} y_{i1} & y_{i2} & y_{i3} & y_{i4} \end{pmatrix} \\
&\vdots \\
\mathbf{z}_{iT} &= \begin{pmatrix} y_{i1} & y_{i2} & y_{i3} & y_{i4} & \cdots & y_{iT-2} \end{pmatrix}
\end{aligned}$$

Instruments for other regressors

- **Model:**
$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta \mathbf{x}_{it}\beta + \Delta u_{it}.$$
- **What about** instruments relating to $\Delta \mathbf{x}_{it}$?
  - $\Rightarrow$ depends on what we are willing to assume about $\mathbf{x}_{it}$.
- **Strictly exogenous:** if $\mathrm{E}(u_{it}|\mathbf{x}_{i}) = 0$, we can use the full vector $\mathbf{x}_{i} \equiv (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT})'$.
- **Sequentially exogenous** (or **predetermined**): if only $\mathrm{E}(u_{it}|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}) = 0$, we can use only $(\mathbf{x}_{i1}, \ldots, \mathbf{x}_{it})$ as an instrument for $\Delta \mathbf{x}_{it}$.

Instruments for Δx_{it}

Example: Female Labor Supply and Fertility

$$y_{it} = \rho y_{it-1} + \beta_0 + \beta_1 k_{it} + \beta_2 f_{it} + c_i + u_{it},$$

where $y_{it} = \mathbf{1}\{\text{work in year } t\}$, $k_{it} = \mathbf{1}\{\text{kids aged } [2;6]\}$ and $f_{it} = \mathbf{1}\{\text{gives birth in year } t\}$.

- **Fertility** (f_{it}) is likely
  - contemporaneously endogenous: $\mathrm{E}(f_{it}u_{it}) \neq 0$
    - requires **external instrument**
  - correlated with individual effects: $\mathrm{E}(f_{it}c_i) \neq 0$
    - solved by **FD**
- **Kids** (k_{it}) is likely
  - not strictly exogenous, since $u_{it} \curvearrowright f_{it+1} \curvearrowright k_{it+2}$
  - but a *predetermined* variable (assuming no child murdering!)
    - we can use $(k_{i1}, \ldots, k_{it-1})$ in period t
  - correlated with individual effects, $\mathrm{E}(k_{it}c_i) \neq 0$.
    - solved by **FD**

Full instrument matrix

- **Full matrix** when y_{it-1} is the only regressor

$$\mathbf{Z}_i = \begin{pmatrix} y_{i1} & 0 & 0 & 0 & 0 & 0 & \cdots & \cdots \\ 0 & y_{i1} & y_{i2} & 0 & 0 & 0 & \cdots & \cdots \\ 0 & 0 & 0 & y_{i1} & y_{i2} & y_{i3} & \cdots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \end{pmatrix}$$

- With **vector notation:**

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{z}_{i3} & \mathbf{0}_{1\times 2} & \mathbf{0}_{1\times 3} & \cdots & \mathbf{0}_{1\times T-2} \\ 0 & \mathbf{z}_{i4} & \mathbf{0}_{1\times 3} & \cdots & \mathbf{0}_{1\times T-2} \\ 0 & \mathbf{0}_{1\times 2} & \mathbf{z}_{i5} & \cdots & \mathbf{0}_{1\times T-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \mathbf{0}_{1\times 2} & \mathbf{0}_{1\times 3} & \cdots & \mathbf{z}_{iT} \end{pmatrix}.$$

  - **Note:** the $\mathbf{Z}_i$ has $T-2$ rows and $\sum_{t=3}^{T}(t-2)$ rows.

Arellano-Bond Estimator

$$\hat{\beta}_{\text{AB}} = \left[\left(\sum_i \tilde{\mathbf{X}}_i'\mathbf{Z}_i\right)\mathbf{W}_N\left(\sum_i \mathbf{Z}_i'\tilde{\mathbf{X}}_i\right)\right]^{-1}\left(\sum_i \tilde{\mathbf{X}}_i'\mathbf{Z}_i\right)\mathbf{W}_N\left(\sum_i \mathbf{Z}_i'\tilde{\mathbf{y}}_i\right).$$

- where $\tilde{\mathbf{X}}_i$ is the $T-2\times K+1$ matrix with t'th row $(\Delta y_{it-1}, \Delta \mathbf{x}_{it}')$, and $\tilde{\mathbf{y}}_i$ is the $T-2\times 1$ vector with rows Δy_{it}
- **Problem:** how to estimate the weighting Matrix?
  1. Initial guess: $\mathbf{W}_N = \left(N^{-1}\sum_i \mathbf{Z}_i'\mathbf{Z}_i\right)^{-1}$
  2. Optimal weight: should be an estimate of $\mathbf{S}^{-1}$, where $\mathbf{S} = \mathrm{E}(\mathbf{Z}_i'\mathbf{u}\mathbf{u}'\mathbf{Z})$.
  - **Problem:** to estimate $\mathbf{S}$, we need "residuals", so we need parameters.
  - **Solution:** get a first-stage *consistent* but *inefficient* estimate, and use those in a second stage.

Two-stage estimation

Arellano-Bond Two-step Estimator

1. Compute $\hat{\beta}_{\text{AB}}$ setting $\mathbf{W}_N = (N^{-1} \sum_i \mathbf{Z}_i' \mathbf{Z}_i)^{-1}$.
   (this is the 2SLS estimator)
2. Re-estimate $\hat{\beta}_{\text{AB}}$ with $\mathbf{W}_N = \left(N^{-1} \sum_i \mathbf{Z}_i' \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i' \mathbf{Z}_i\right)^{-1}$.

- **Assuming** that u_{it} are IID, the two-step estimator is asymptotically efficient.

Instrument relevance?

- **Instruments** in general have to satisfy
  1. **Exogeneity**: follows from model structure and the assumption that u_{it} is serially uncorrelated.
  2. **Relevance**: the instrument must predict
- **Relevance:** can be a problem in many situations...
  - ... after all, we are hoping for the level of y_{it} to predict future changes, Δy_{it+2}
- **However**: relevant e.g. for the *heterogeneous income profiles* (HIP) model:

$$y_{it} = c_i + \beta_i t + p_{it} + u_{it}$$
$$p_{it} = p_{it-1} + \epsilon_{it},$$

where $(c_i, \beta_i)' \sim \mathcal{N}(\mathbf{0}, \Sigma)$, and ϵ_{it}, u_{it} are IID, and p_{it} is unobserved.

  - Here, the *level* and *changes* in y_{it} become increasingly correlated over the life-cycle.
  - See Druedahl & Munk-Nielsen (2018).

Levels and growth rates