Chapter 2: Panel Data

Joan Llull

Microeconometrics.
IDEA PhD Program. Fall 2017

joan.llull [at] movebarcelona [dot] eu

Introduction

Panel data

The term **panel data** refers to data sets with **repeated observations** over time for a given cross-section of individuals.

Individuals can be persons, households, firms, countries,...

It is different from **repeated cross-sections**.

Main **advantages** of panel data:

- Permanent unobserved heterogeneity
- Dynamic responses and error components

Micro and macro panel data

Micro panel data usually has large N, small T (e.g. household surveys).

Macro panel data usually have longer T, but smaller N (e.g. daily stock market returns for three composites).

Our interest here: **fixed** T**,** $N \to \infty$ (micro panels).

Approaches are **closer to cross-section approaches** than to time series.

Static Models

General notation

We consider the following **model**:

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + (\eta_i + v_{it})$$

where y_{it} and $\boldsymbol{x}_{it}$ are **observed**, and $\eta_i + v_{it}$ is **unobserved**.

Let $\{y_{it}, \boldsymbol{x}_{it}\}_{i=1,\ldots,N}^{t=1,\ldots,T}$ be our **sample**. We define:

$$\boldsymbol{y}_i \equiv \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}, X_i = \begin{pmatrix} \boldsymbol{x}_{i1}' \\ \vdots \\ \boldsymbol{x}_{iT}' \end{pmatrix}, \boldsymbol{\eta}_i = \eta_i \boldsymbol{\iota}_T, \text{ and } \boldsymbol{v}_i = \begin{pmatrix} v_{i1} \\ \vdots \\ v_{iT} \end{pmatrix},$$

$$\boldsymbol{y} \equiv \begin{pmatrix} \boldsymbol{y}_1 \\ \vdots \\ \boldsymbol{y}_N \end{pmatrix}, X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}, \boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\eta}_1 \\ \vdots \\ \boldsymbol{\eta}_N \end{pmatrix}, \text{ and } \boldsymbol{v} = \begin{pmatrix} \boldsymbol{v}_1 \\ \vdots \\ \boldsymbol{v}_N \end{pmatrix},$$

where $\boldsymbol{\iota}_T$ is a size T **vector of ones**.

Hence, we can use the following **compact notation**:

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta} + (\boldsymbol{\eta}_i + \boldsymbol{v}_i), \quad \text{and} \quad \boldsymbol{y} = X\boldsymbol{\beta} + (\boldsymbol{\eta} + \boldsymbol{v})$$

General assumptions for static models

For static models, we assume:

- **Fixed effects**: $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] \neq 0$ or **random effects**: $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] = 0$.
- **Strict exogeneity**: $\mathbb{E}[\boldsymbol{x}_{it}v_{is}] = 0\, \forall s, t$. This assumption rules out effects of past v_{is} on current $\boldsymbol{x}_{it}$ (e.g. $\boldsymbol{x}_{it}$ cannot include lagged dependent variables).
- **Error components**: $\mathbb{E}[\eta_i] = \mathbb{E}[v_{it}] = \mathbb{E}[\eta_i v_{it}] = 0$.
- **Serially uncorrelated shocks**: $\mathbb{E}[v_{it}v_{is}] = 0\; \forall s \neq t$.
- **Homoskedasticity and i.i.d. errors**: $\eta_i \sim iid(0, \sigma_\eta^2)$ and $v_{it} \sim iid(0, \sigma_v^2)$, which does not affect any crucial result, but simplifies some derivations.

Pooled OLS

A simple approach: define: $\boldsymbol{u} \equiv \boldsymbol{\eta} + \boldsymbol{v}$ and estimate $\boldsymbol{\beta}$ by OLS:

$$\hat{\boldsymbol{\beta}}_{OLS} = (X'X)^{-1}X'\boldsymbol{y}.$$

The **properties** of $\hat{\boldsymbol{\beta}}_{OLS}$ depend on $\mathbb{E}[\boldsymbol{x}_{it}\eta_i]$, as $\mathbb{E}[\boldsymbol{x}_{it}v_{it}] = 0$:

- If $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] = 0 \underset{\mathbb{E}[\boldsymbol{x}_{it}v_{it}]=0}{\Rightarrow} \mathbb{E}[\boldsymbol{x}_j u_j] = 0$ (**random effects**):
  - $\hat{\boldsymbol{\beta}}_{OLS}$ is **consistent** as $N \to \infty$, or $T \to \infty$, or both.
  - it is **efficient** only if $\sigma_\eta^2 = 0$.
- If $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] \neq 0 \Rightarrow \mathbb{E}[\boldsymbol{x}_j u_j] \neq 0$ (**fixed effects**):
  - $\hat{\boldsymbol{\beta}}_{OLS}$ is **inconsistent** as $N \to \infty$, or $T \to \infty$, or both.
  - **cross-section** results are also inconsistent (but panel helps in constructing a consistent alternative).

The fixed effects model.
Within groups estimation.

Within groups estimator

Write the model in **deviations from individual means**, $\tilde{y}_{it} \equiv y_{it} - \bar{y}_i$, where $\bar{y}_i \equiv T^{-1} \sum_{t=1}^{T} y_{it}$:

$$\tilde{y}_{it} = (\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)'\boldsymbol{\beta} + (\eta_i - \bar{\eta}_i) + (v_{it} - \bar{v}_i) = \tilde{\boldsymbol{x}}'_{it}\boldsymbol{\beta} + \tilde{v}_{it}.$$

Given the previous **assumptions**:

$$\mathbb{E}[\tilde{\boldsymbol{x}}_{it}\tilde{v}_{it}] = 0.$$

Therefore, **OLS on the transformed model**:

$$\hat{\boldsymbol{\beta}}_{WG} = \left(\tilde{X}'\tilde{X}\right)^{-1} \tilde{X}'\tilde{y},$$

is a **consistent** estimator either if $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] \neq 0$ or $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] = 0$.

Strict exogeneity is a crucial assumption.

Pros and cons of within groups estimator

Advantage: consistent either if $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] \neq 0$ or $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] = 0$.

Limitations:

- **Not efficient**:
  - When $N \to \infty$ but T is fixed, less efficient that e.g. $\hat{\boldsymbol{\beta}}_{GLS}$ if $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] = 0$.
  - If $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] \neq 0$, it is only efficient when all regressors are correlated with η_i.
- Does not allow identification of coefficients for **time-invariant regressors**, and identification of other coefficients is provided by **switchers**.

Least Squares Dummy Variables

The Within Groups estimator can also be obtained by including a set of N individual **dummy variables**:

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \eta_1 D_{1i} + ... + \eta_N D_{Ni} + v_{it},$$

where $D_{hi} = \mathbb{1}\{h = i\}$ (e.g. D_{1i} takes the value of 1 for the observations on individual 1 and 0 for all other observations).

OLS estimation of this model gives **numerically equivalent** estimates to WG (that's why $\hat{\boldsymbol{\beta}}_{WG}$ is a.k.a. $\hat{\boldsymbol{\beta}}_{LSDV}$).

This gives intuition on why WG is **not very efficient** if there is only limited time-series variation (degrees of freedom are $NT - K - N = N(T-1) - K$)

First-Differenced Least Squares

Another transformation we can consider is **first differences**:

$$\Delta y_{it} = \Delta \boldsymbol{x}'_{it}\boldsymbol{\beta} + \Delta v_{it}, \text{ for } i = 1, ..., N; t = 2, ..., T$$

where $\Delta y_{it} = y_{it} - y_{it-1}$.

Takes out time-invariant individual effects ($\Delta\eta_i = \eta_i - \eta_i = 0$), so OLS on the differenced model is **consistent**.

Consistency requires $\mathbb{E}[\Delta \boldsymbol{x}_{it} \Delta v_{it}] = 0$ which is implied by but weaker than strict exogeneity.

WG more efficient than FDLS under **classical assumptions**.

FDLS more efficient if v_{it} random walk ($\Delta v_{it} = \varepsilon_{it} \sim iid(0, \sigma_\varepsilon^2)$).

The random effects model.
Error components.

Uncorrelated effects

Now we assume uncorrelated or **random effects**: $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] = 0$.

In this case, OLS is **consistent**, but **not efficient**.

The inefficiency is provided by the **serial correlation** induced by the presence of η_i in the error term:

$$\mathbb{E}[u_{it}u_{is}] = \mathbb{E}[(\eta_i + v_{it})(\eta_i + v_{is})] = \mathbb{E}[\eta_i^2] = \sigma_\eta^2, \quad \text{for } s \neq t.$$

The **variance** of the unobservables (under classical assumptions) is:

$$\mathbb{E}[u_{it}^2] = \mathbb{E}[\eta_i^2] + \mathbb{E}[v_{it}^2] = \sigma_\eta^2 + \sigma_v^2$$

Error structure

Therefore, the variance-covariance matrix of the unobservables is:

$$\mathbb{E}[\boldsymbol{u}_i \boldsymbol{u}_i'] = \begin{pmatrix} \sigma_\eta^2 + \sigma_v^2 & \sigma_\eta^2 & \dots & \sigma_\eta^2 \\ \sigma_\eta^2 & \sigma_\eta^2 + \sigma_v^2 & \dots & \sigma_\eta^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\eta^2 & \sigma_\eta^2 & \dots & \sigma_\eta^2 + \sigma_v^2 \end{pmatrix} = \Omega_i,$$

whose dimensions are $T \times T$, and $\mathbb{E}[\boldsymbol{u}_i \boldsymbol{u}_h'] = 0 \ \forall \ i \neq h$, or:

$$\mathbb{E}[\boldsymbol{u}\boldsymbol{u}'] = \begin{pmatrix} \Omega_1 & 0 & \dots & 0 \\ 0 & \Omega_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Omega_N \end{pmatrix} = \Omega,$$

which is block-diagonal with dimension $NT \times NT$.

Generalized Least Squares

Under the classical assumptions, GLS (Balestra-Nerlove) estimator is **consistent and efficient** if $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] = 0$:

$$\hat{\boldsymbol{\beta}}_{GLS} = \left(X'\Omega^{-1}X\right)^{-1} X'\Omega^{-1}\boldsymbol{y}.$$

If $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] \neq 0$ GLS is **inconsistent** as $N \to \infty$ and T is fixed.

This estimator is **unfeasible** because we do not know σ_η^2 and σ_v^2.

Theta-differencing

$\hat{\boldsymbol{\beta}}_{GLS}$ is **equivalent** to OLS on the theta-differenced model:

$$y_{it}^* = \boldsymbol{x}_{it}^{*\prime}\boldsymbol{\beta} + u_{it}^*,$$

where:

$$y_{it}^* = y_{it} - (1-\theta)\bar{y}_i,$$

and:

$$\theta^2 = \frac{\sigma_v^2}{\sigma_v^2 + T\sigma_\eta^2}.$$

Consistency relies on $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] = 0$ (as η_i not eliminated).

Two **special cases**:

- If $\sigma_\eta^2 = 0$ (i.e. no individual effect), OLS is efficient.
- If $T \to \infty$, then $\theta \to 0$, and $y_{it}^* \to \tilde{y}_{it} = y_{it} - \bar{y}_i$: WG is efficient.

Feasible GLS

$\hat{\boldsymbol{\beta}}_{GLS}$ is **unfeasible** because we do not know σ_η^2 and σ_v^2.

A consistent estimator of σ_v^2 is provided by the **WG residuals**:

$$\hat{\tilde{v}}_{it} \equiv \tilde{y}_{it} - \tilde{\boldsymbol{x}}_{it}' \hat{\boldsymbol{\beta}}_{WG}$$

$$\hat{\sigma}_v^2 = \frac{\hat{\tilde{\boldsymbol{v}}}' \hat{\tilde{\boldsymbol{v}}}}{N(T-1) - K}$$

Then, a consistent estimator of σ_η^2 is given by the **BG residuals**:

$$\bar{y}_i = \bar{\boldsymbol{x}}_i' \boldsymbol{\beta} + \bar{\eta}_i + \bar{v}_i, \quad i = 1, ..., N \Rightarrow \hat{\boldsymbol{\beta}}_{BG}$$

$$\hat{\bar{u}}_i \equiv \bar{y}_i - \bar{\boldsymbol{x}}_i' \hat{\boldsymbol{\beta}}_{BG}$$

$$\hat{\sigma}_{\bar{u}}^2 = (\widehat{\sigma_\eta^2 + \frac{1}{T}\sigma_v^2}) = \frac{\hat{\bar{\boldsymbol{u}}}' \hat{\bar{\boldsymbol{u}}}}{N - K} \quad \Rightarrow \quad \hat{\sigma}_\eta^2 = \hat{\sigma}_{\bar{u}}^2 - \frac{1}{T}\hat{\sigma}_v^2.$$

Testing for correlated individual effects

Testing for correlated effects (Hausman test)

$\hat{\boldsymbol{\beta}}_{WG}$ is **consistent regardless** of $\mathbb{E}[\boldsymbol{x}_{it}\eta_i]$ being equal to zero or not.

$\hat{\boldsymbol{\beta}}_{FGLS}$ is **consistent only** if $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] = 0$.

$\Rightarrow$ we can test whether both **estimates are similar!**

The **Hausman test** does exactly this comparison:

$$h = \hat{\boldsymbol{q}}'[avar(\hat{\boldsymbol{q}})]^{-1}\hat{\boldsymbol{q}} \overset{a}{\sim} \chi^2(K)$$

under the **null hypothesis** $\mathbb{E}[\boldsymbol{x}_{it}\eta_i] = 0$, where:

$$\hat{\boldsymbol{q}} = \hat{\boldsymbol{\beta}}_{WG} - \hat{\boldsymbol{\beta}}_{FGLS},$$

and:

$$avar(\hat{\boldsymbol{q}}) = avar\left(\hat{\boldsymbol{\beta}}_{WG}\right) - avar\left(\hat{\boldsymbol{\beta}}_{FGLS}\right).$$

Requires **classical assumptions** (FGLS to be more efficient than WG).

Dynamic Models

Autoregressive models with individual effects

Autorregressive panel data model

We consider the following model:

$$y_{it} = \alpha y_{it-1} + \eta_i + v_{it} \quad |\alpha| < 1.$$

Other regressors can be included, but main results unaffected.

We assume:

- **Error components**: $\mathbb{E}[\eta_i] = \mathbb{E}[v_{it}] = \mathbb{E}[\eta_i v_{it}] = 0.$
- **Serially uncorrelated shocks**: $\mathbb{E}[v_{it} v_{is}] = 0 \; \forall \; s \neq t.$
- **Predetermined initial cond.:** $\mathbb{E}[y_{i0} v_{it}] = 0 \; \forall \; t = 1, ..., T.$

Properties of pooled OLS and WG estimators

Even assuming $\mathbb{E}[y_{it-1}v_{it}] = 0$, still **OLS** delivers:

$$\underset{N\to\infty}{\text{plim}}\ \hat{\alpha}_{OLS} > \alpha,$$

because $\mathbb{E}[y_{it-1}\eta_i] = \sigma_\eta^2\left(\frac{1-\alpha^{t-1}}{1-\alpha}\right) > 0.$

Doing the **within groups** transformation we see that:

$$\underset{N\to\infty}{\text{plim}}\ \hat{\alpha}_{WG} < \alpha$$

because $\mathbb{E}[\tilde{y}_{it-1}\tilde{v}_{it}] = -A\sigma_v^2 < 0.$ $\left(A = \frac{(1-\alpha)\left(1+T\left(1-\alpha^{t-1}-\alpha^{T-1-t}\right)\right)+\alpha T\left(1-\alpha^{T-1}\right)}{T^2(1-\alpha)^2}\right)$

WG bias vanishes as $T\to\infty$ (bias not small even with $T = 15$).

Supposedly consistent estimators that give $\hat{\alpha} >> \alpha_{OLS}$ or $\hat{\alpha} << \hat{\alpha}_{WG}$ should be **seen with suspicion**.

Anderson and Hsiao

Consider the model in **first differences**:

$$\Delta y_{it} = \alpha \Delta y_{it-1} + \Delta v_{it}.$$

OLS in first differences is **inconsistent**: $\mathbb{E}[\Delta y_{it-1} \Delta v_{it}] = -\sigma_v^2 < 0$.

However, y_{it-2} or Δy_{it-2} are valid **instruments** for Δy_{it-1}:

- **Relevance**: $\mathbb{E}[\Delta y_{it-2} \Delta y_{it-1}] \neq 0$ and $\mathbb{E}[y_{it-2} \Delta y_{it-1}] \neq 0$.
- **Orthogonality**: $\mathbb{E}[\Delta y_{it-2} \Delta v_{it}] = \mathbb{E}[y_{it-2} \Delta v_{it}] = 0$.

Anderson and Hsiao (1981) proposed this **2SLS estimators**:

$$\hat{\alpha}_{AH} = \left(\widehat{\Delta \boldsymbol{y}'_{-1}} \widehat{\Delta \boldsymbol{y}_{-1}} \right)^{-1} \widehat{\Delta \boldsymbol{y}'_{-1}} \Delta \boldsymbol{y},$$

where:

$$\widehat{\Delta \boldsymbol{y}_{-1}} = Z \left(Z'Z \right)^{-1} Z' \Delta \boldsymbol{y}_{-1},$$

where Z can be $\boldsymbol{y}_{-2}$ or $\Delta \boldsymbol{y}_{-2}$.

Requires min. **three periods** ($T = 2$ and y_{i0}). Only **efficient** if $T = 2$.

Differenced GMM estimation

Moment conditions

Given previous assumptions, several **moment conditions**:

Equation	**Instruments**	**Orthogonality cond.**
$\Delta y_{i2} = \alpha \Delta y_{i1} + \Delta v_{i2}$	y_{i0}	$\mathbb{E}[\Delta v_{i2} y_{i0}] = 0$
$\Delta y_{i3} = \alpha \Delta y_{i2} + \Delta v_{i3}$	y_{i0}, y_{i1}	$\mathbb{E}\left[\Delta v_{i3} \begin{pmatrix} y_{i0} \\ y_{i1} \end{pmatrix}\right] = \mathbf{0}$
$\Delta y_{i4} = \alpha \Delta y_{i3} + \Delta v_{i4}$	y_{i0}, y_{i1}, y_{i2}	$\mathbb{E}\left[\Delta v_{i4} \begin{pmatrix} y_{i0} \\ y_{i1} \\ y_{i2} \end{pmatrix}\right] = \mathbf{0}$
$\vdots$	$\vdots$	$\vdots$
$\Delta y_{iT} = \alpha \Delta y_{iT-1} + \Delta v_{iT}$	$y_{i0}, y_{i1}, y_{i2}, ..., y_{iT-2}$	$\mathbb{E}\left[\Delta v_{iT} \begin{pmatrix} y_{i0} \\ y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT-2} \end{pmatrix}\right] = \mathbf{0}$

We end up with $(T-1)T/2$ **moment conditions**.

Moment conditions in matrix form

We can write these **moment conditions** as $\mathbb{E}[Z_i'\Delta \boldsymbol{v}_i] = 0$, where:

$$Z_i = \begin{pmatrix} y_{i0} & 0 & 0 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & y_{i0} & y_{i1} & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \dots & \vdots \\ 0 & 0 & 0 & 0 & \dots & y_{i0} & y_{i1} & \dots & y_{iT-2} \end{pmatrix} \text{ and } \Delta \boldsymbol{v}_i = \begin{pmatrix} \Delta v_{i2} \\ \Delta v_{i3} \\ \vdots \\ \Delta v_{iT} \end{pmatrix},$$

and the **sample analogue** is:

$$\boldsymbol{b}_N(\alpha) = \frac{1}{N} \sum_{\mathrm{i=1}}^{\mathrm{N}} Z_i' \Delta \boldsymbol{v}_i(\alpha)$$

Hence, the **GMM estimator** (proposed by Arellano and Bond, 1991) is:

$$\hat{\alpha}_{GMM} = \arg\min_{\alpha} \left(\frac{1}{N} \sum_{\mathrm{i=1}}^{\mathrm{N}} \Delta \boldsymbol{v}_i'(\alpha) Z_i \right) W_N \left(\frac{1}{N} \sum_{\mathrm{i=1}}^{\mathrm{N}} Z_i' \Delta \boldsymbol{v}_i(\alpha) \right) =$$

$$= (\Delta \boldsymbol{y}_{-1}' Z W_N Z' \Delta \boldsymbol{y}_{-1})^{-1} \Delta \boldsymbol{y}_{-1}' Z W_N Z' \Delta \boldsymbol{y}$$

Optimal weighting matrix

The **optimal weighting matrix** (efficient GMM) is:

$$W_N = \left(\frac{1}{N} \sum_{\text{i=1}}^{\text{N}} \mathbb{E}[Z_i' \Delta \boldsymbol{v}_i \Delta \boldsymbol{v}_i' Z_i] \right)^{-1}$$

The **sample analogue** is obtained in a **two-step** procedure:

$$W_N = \left(\frac{1}{N} \sum_{\text{i=1}}^{\text{N}} [Z_i' \widehat{\Delta \boldsymbol{v}_i(\hat{\alpha})} \widehat{\Delta \boldsymbol{v}_i'(\hat{\alpha})} Z_i] \right)^{-1}$$

Windmeijer (2005) proposes a **finite sample correction** of the variance that accounts for α being estimated.

The most common **one-step** (and first-step) matrix uses the structure of $\mathbb{E}[\Delta \boldsymbol{v}_i \Delta \boldsymbol{v}_i']$:

$$\mathbb{E}[\Delta \boldsymbol{v}_i \Delta \boldsymbol{v}_i'] = \sigma_v^2 \begin{pmatrix} 2 & -1 & 0 & 0 & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \\ 0 & 0 & 0 & 0 & \dots & 2 \end{pmatrix}$$

Potential limitations of Arellano-Bond

Weak instruments:

- When $\alpha \to 1$, **relevance** of the instrument decreases.
- Instruments are still valid, but have **poor small sample** properties.
- **Monte Carlo evidence** shows that with $\alpha > 0.8$, estimator behaves poorly unless huge samples available.
- There are **alternatives** in the literature.

Overfitting:

- **"Too many" instruments** if T relative to N is relatively large.
- We might want to **restrict** the number of instruments used.
- It is always good practice to check **robustness** to different combinations of instruments.

System GMM estimation

Additional orthogonality conditions

Recall our $(T-1)T/2$ **moment conditions**:

$$\mathbb{E}[y_{it-s}\Delta v_{it}] = 0 \quad t = 2, ..., T; s \geq 2$$

System GMM (Arellano and Bover, 1995) uses the assumption $\mathbb{E}[y_{i0}|\eta_i] = \frac{\eta_i}{1-\alpha}$, which implies:

$$\mathbb{E}[\Delta y_{is}\eta_i] = 0,$$

or, alternatively:

$$\mathbb{E}[\Delta y_{iT-s}u_{iT}] = 0, \quad u_{iT} \equiv \eta_i + v_{iT}, \quad s = 1, ..., T-1.$$

The System GMM estimator

Analogously to the first-differenced GMM, the estimator is given by $\mathbb{E}[(Z^*)' u_i^*] = 0$:

$$\hat{\alpha}_{Sys-GMM} = \left((X^*)' Z^* W_N (Z^*)' X^*\right)^{-1} X^* Z^* W_N (Z^*)' \boldsymbol{y}^*,$$

where:

$$Z_i^* \equiv \begin{pmatrix} Z_i & \mathbf{0} & \dots & \mathbf{0} \\ 0 & \Delta y_{i1} & \dots & \Delta y_{iT-1} \end{pmatrix}, \boldsymbol{u}_i^* \equiv \begin{pmatrix} \Delta \boldsymbol{v}_i \\ \eta_i + v_{iT} \end{pmatrix}, X_i^* \equiv \begin{pmatrix} \Delta \boldsymbol{y}_{-1i} \\ y_{iT-1} \end{pmatrix} \text{ and } \boldsymbol{y}_i^* \equiv \begin{pmatrix} \Delta \boldsymbol{y}_i \\ y_{iT} \end{pmatrix}$$

This estimator is **more efficient**, as it uses additional moment conditions.

It **reduces small sample bias**, especially when $\alpha \rightarrow 1$

Specification Tests

Overidentifying restrictions

The null hypothesis is whether the **orthogonality** conditions are **satisfied** (i.e. moments are equal to zero).

The test can only be implemented if the problem is **overidentified** (otherwise the sample moments are exactly zero by construction).

The **test** is:

$$S = NJ_N(\beta) = N\left(\frac{1}{N}\sum_{\mathrm{i=1}}^{\mathrm{N}}\hat{\hat{\mathbf{u}}}_i' Z_i \left(\frac{1}{N}\sum_{\mathrm{i=1}}^{\mathrm{N}} Z_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i' Z_i\right)^{-1}\frac{1}{N}\sum_{\mathrm{i=1}}^{\mathrm{N}} Z_i'\hat{\hat{\mathbf{u}}}_i\right),$$

where $\hat{\boldsymbol{u}}$ are predicted residuals from the first step and $\hat{\hat{\boldsymbol{u}}}$ are those predicted from the second stage. **Under the null**,

$$S \overset{a}{\sim} \chi^2(L-K).$$

If some assumptions are stronger than others: **include or exclude** the orthogonality conditions generated by them.

If they are true, increase **efficiency**, but if not, **inconsistent**! $\Rightarrow$ **Difference in Sargan** or **Hausman test**.

Direct test for serial correlation

The test was proposed by Arellano-Bond (1991).

Tests for the presence of **second order autocorrelation** in the first-differenced residuals.

If differences in residuals are second-order correlated, some **instruments would not be valid**!

The **test** is:

$$m_2 = \frac{\widehat{\Delta \boldsymbol{v}_{-2}}' \widehat{\Delta \boldsymbol{v}_*}}{se} \overset{a}{\sim} \mathcal{N}(0,1),$$

where $\Delta \boldsymbol{v}_{-2}$ is the second lagged residual in differences, and $\Delta \boldsymbol{v}_*$ is the part of the vector of contemporaneous first differences for the periods that overlap with the second lagged vector.

Values close to zero do not reject the hypothesis of **no serial correlation**.