

CHAPTER 6. PANEL DATA

Joan Llull

Quantitative Statistical Methods II
Barcelona GSE

INTRODUCTION

Panel data

The term **panel data** refers to data sets with **repeated observations** over time (or other dimensions) for a given cross-section of units.

Units 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.

Cigarette Consumption

We will use the same **example** throughout the chapter.

Consider the estimation of the demand for cigarettes:

$$\ln C_{it} = \beta_0 + \beta_1 \ln P_{it} + \beta_2 \ln Y_{it} + \eta_i + v_{it},$$

where:

- $\ln C_{it}$ is log consumption of cigarettes,
- $\ln P_{it}$ is log price of the cigarettes,
- $\ln Y_{it}$ is log income of the individual,
- $\eta_i + v_{it}$ is unobserved.

STATIC MODELS

General notation

We consider the following **model**:

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

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

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

$$\mathbf{y}_i \equiv \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}, X_i = \begin{pmatrix} \mathbf{x}_{i1}' \\ \vdots \\ \mathbf{x}_{iT}' \end{pmatrix}, \boldsymbol{\eta}_i = \eta_i \boldsymbol{\iota}_T, \text{ and } \mathbf{v}_i = \begin{pmatrix} v_{i1} \\ \vdots \\ v_{iT} \end{pmatrix},$$
$$\mathbf{y} \equiv \begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_N \end{pmatrix}, X = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix}, \boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_N \end{pmatrix}, \text{ and } \mathbf{v} = \begin{pmatrix} \mathbf{v}_1 \\ \vdots \\ \mathbf{v}_N \end{pmatrix},$$

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

Hence, we can use **compact notation**: $\mathbf{y}_i = X_i\boldsymbol{\beta} + (\boldsymbol{\eta}_i + \mathbf{v}_i)$, and $\mathbf{y} = X\boldsymbol{\beta} + (\boldsymbol{\eta} + \mathbf{v})$.

General assumptions for static models

For static models, we assume:

- **Fixed effects:** $\mathbb{E}[\mathbf{x}_{it}\eta_i] \neq 0$ or **random effects:** $\mathbb{E}[\mathbf{x}_{it}\eta_i] = 0$.
- **Strict exogeneity:** $\mathbb{E}[\mathbf{x}_{it}v_{is}] = 0 \ \forall s, t$. This assumption rules out effects of past v_{is} on current \mathbf{x}_{it} (e.g. \mathbf{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: $\mathbf{u} \equiv \boldsymbol{\eta} + \mathbf{v}$ and estimate $\boldsymbol{\beta}$ by OLS:

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

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

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

Pooled OLS in Cigarette Demand

In our previous example, we would **redefine** our regression as:

$$\ln C_j = \beta_0 + \beta_1 \ln P_j + \beta_2 \ln Y_j + u_j,$$

where I used subindex j to emphasize that each observation *it* is considered as one **independent** observation.

OLS		
$\ln \text{ Prices } (\beta_1)$	-0.083	(0.015)
$\ln \text{ Income } (\beta_2)$	-0.032	(0.006)

Potential problems:

- preferences,
- education,
- parental smoking behavior,...

**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 $\bar{y}_i \equiv T^{-1} \sum_{t=1}^T y_{it}$:

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

Given the previous **assumptions**:

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

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 $\mathbb{E}[\mathbf{x}_{it} \eta_i] \neq 0$ or $\mathbb{E}[\mathbf{x}_{it} \eta_i] = 0$.

Strict exogeneity is a crucial assumption (see next slide).

The role of strict exogeneity

In the case where $N \rightarrow \infty$ and T is fixed, consistency depends on **strict exogeneity**.

To see it, recall that:

$$\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \frac{1}{T}(\mathbf{x}_{i1} + \dots + \mathbf{x}_{iT}) \text{ and } \tilde{v}_{it} = v_{it} - \frac{1}{T}(v_{i1} + \dots + v_{iT}).$$

Therefore $\mathbb{E}[\tilde{\mathbf{x}}_{it}\tilde{v}_{it}] = 0$ requires $\mathbb{E}[\mathbf{x}_{it}v_{is}] = 0 \ \forall \ s, t$ unless $T \rightarrow \infty$.

This has motivated the development of **dynamic panel data models**, to relax this assumption.

Pros and cons of within groups estimator

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

Limitations:

- **Not efficient:**

- When $N \rightarrow \infty$ but T is fixed, less efficient than e.g. $\hat{\beta}_{GLS}$ if $\mathbb{E}[\mathbf{x}_{it}\eta_i] = 0$.
- Even if $\mathbb{E}[\mathbf{x}_{it}\eta_i] \neq 0$, differencing introduces correlation in the errors.

- It does not allow to identify coefficients for **time-invariant regressors**, and identification is through **switchers**.

Within Groups in Cigarette Demand

In our example:

	OLS		WG	
ln Prices (β_1)	-0.083	(0.015)	-0.292	(0.023)
ln Income (β_2)	-0.032	(0.006)	0.107	(0.019)

Least Squares Dummy Variables

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

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

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. $\hat{\boldsymbol{\beta}}_{LSDV}$).

LSDV in Cigarette Demand

We can generate individual (firm) dummies and estimate by OLS, to check that it delivers the same results:

	OLS	WG	LSDV
ln Prices (β_1)	-0.083 (0.015)	-0.292 (0.023)	-0.292 (0.023)
ln Income (β_2)	-0.032 (0.006)	0.107 (0.019)	0.107 (0.019)
Individual 1 ($\beta_0 + \eta_1$)			2.804 (0.288)
Individual 2 ($\beta_0 + \eta_2$)			3.455 (0.398)
Individual 3 ($\beta_0 + \eta_3$)			2.891 (0.416)
Individual 4 ($\beta_0 + \eta_4$)			2.908 (0.384)
Individual 5 ($\beta_0 + \eta_5$)			3.490 (0.433)
Individual 6 ($\beta_0 + \eta_6$)			2.092

First-Differenced Least Squares

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

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

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 $\mathbb{E}[\Delta \mathbf{x}_{it} \Delta v_{it}] = 0$ 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 ($\Delta v_{it} = \varepsilon_{it} \sim iid(0, \sigma_\varepsilon^2)$).

FDLS in Cigarette Demand

To get FDLS we generate first differences and estimate by OLS:

	OLS	WG	LSDV	FDLS
ln Prices (β_1)	-0.083 (0.015)	-0.292 (0.023)	-0.292 (0.023)	-0.413 (0.035)
ln Income (β_2)	-0.032 (0.006)	0.107 (0.019)	0.107 (0.019)	0.178 (0.055)
Individual 1 ($\beta_0 + \eta_1$)			2.804 (0.288)	
Individual 2 ($\beta_0 + \eta_2$)			3.455 (0.398)	
Individual 3 ($\beta_0 + \eta_3$)			2.891 (0.416)	
Individual 4 ($\beta_0 + \eta_4$)			2.908 (0.384)	
...			...	

**The random effects model.
Error components.**

Uncorrelated effects

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

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

The inefficiency is provided by the **serial correlation** induced by η_i :

$$\mathbb{E}[u_{it}u_{is}] = \mathbb{E}[(\eta_i + v_{it})(\eta_i + v_{is})] = \mathbb{E}[\eta_i^2] = \sigma_\eta^2.$$

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:

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

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

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

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 $\mathbb{E}[\mathbf{x}_{it}\eta_i] = 0$:

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

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

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

Theta-differencing

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

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

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

Two special cases:

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

Feasible GLS

$\hat{\beta}_{GLS}$ 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**:

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

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

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

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

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

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

Feasible GLS in Cigarette Demand

In our example, if we now estimate $\hat{\beta}_{FGLS}$, we get:

	OLS	WG	FGLS
ln Prices (β_1)	-0.083 (0.015)	-0.292 (0.023)	-0.122 (0.014)
ln Income (β_2)	-0.032 (0.006)	0.107 (0.019)	-0.012 (0.004)

Testing for correlated individual effects.

Testing for correlated effects (Hausman test)

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

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

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

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

$$h = \hat{\mathbf{q}}' [\text{avar}(\hat{\mathbf{q}})]^{-1} \hat{\mathbf{q}} \stackrel{a}{\sim} \chi^2(K),$$

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

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

and:

$$\text{avar}(\hat{\mathbf{q}}) = \text{avar}(\hat{\beta}_{WG}) - \text{avar}(\hat{\beta}_{FGLS}).$$

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

Hausman test in Cigarette Demand

Output from software packages often includes the Hausman test. In our example:

	Statistic	P-value
Hausman test	24.661	0.000

DYNAMIC MODELS

Autoregressive models with individual effects

Autoregressive 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:** $\mathbb{E}[v_{it} v_{is}] = 0 \quad \forall s \neq t.$
- **Predetermined initial cond.:** $\mathbb{E}[y_{i0} v_{it}] = 0 \quad \forall t = 1, \dots, T.$

Properties of pooled OLS and WG estimators

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

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

$$\text{because } \mathbb{E}[y_{it-1}\eta_i] = \sigma_\eta^2 \left(\frac{1-\alpha^{t-1}}{1-\alpha} \right) + \alpha^{t-1} \mathbb{E}[y_{i0}\eta_i] > 0.$$

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

$$\text{plim}_{N \rightarrow \infty} \hat{\alpha}_{WG} < \alpha,$$

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

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

Supposedly consistent estimators with $\hat{\alpha} \gg \alpha_{OLS}$ or $\hat{\alpha} \ll \hat{\alpha}_{WG}$ 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**: $\mathbb{E}[\Delta y_{it-2} \Delta y_{it-1}] \neq 0$, $\mathbb{E}[y_{it-2} \Delta y_{it-1}] \neq 0$.
- **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:

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

where:

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

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

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

AR(1) Cigarette Demand (No Covariates)

In our example, we redefine the model to be an AR(1) process (for now without regressors):

$$\ln C_{it} = \alpha \ln C_{it-1} + \eta_i + v_{it}.$$

The Anderson-Hsiao results (together with OLS and WG) are:

	OLS	WG	Anderson-Hsiao
Lagged cons. ($\ln C_{it-1}$)	0.982 (0.003)	0.884 (0.061)	1.395 (0.090)

Differenced GMM estimation.

GMM in 3 slides (I): the setup

GMM finds parameter estimates that come as close as possible to satisfy **orthogonality conditions** (or moment conditions) in the sample.

E.g., consider K regressors \mathbf{x}_i and L “instruments” \mathbf{z}_i :

$$u_i = y_i - f(\mathbf{x}_i, \boldsymbol{\beta}) \quad \mathbf{z}_i = g(\mathbf{x}_i).$$

The model specifies L **moment conditions**: $\mathbb{E}[\mathbf{z}_i u_i] = \mathbf{0}$.

Sample analogue:

$$\mathbf{b}_N(\boldsymbol{\beta}) = \frac{1}{N} \sum_{i=1}^N \mathbf{z}_i u_i(\boldsymbol{\beta}).$$

GMM in 3 slides (II): the estimation

Two possible cases **cases**:

- $L = K$ (**just identified**): $\mathbf{b}_N(\hat{\beta}_{GMM}) = \mathbf{0}$.
- $L > K$ (**overidentified**): $\hat{\beta}_{GMM} = \arg \min_{\beta} \mathbf{b}_N(\beta)' W_N \mathbf{b}_N(\beta)$.

W_N is a **positive definite** weighting matrix.

Optimal GMM (efficient) uses $\left(\frac{1}{N} \sum_{i=1}^N \mathbb{E}[z_i u_i u_i' z_i'] \right)^{-1}$, the inverse of the variance-covariance matrix, as weighting matrix.

GMM in 3 slides (III): asymptotic properties

$\hat{\beta}_{GMM}$ is a **consistent** estimator of β .

It is **asymptotically normal**, with the following variance:

$$avar(\hat{\beta}_{GMM}) = (D'WD)^{-1}D'WS_0WD(D'WD)^{-1},$$

where:

$$D \equiv \text{plim}_{N \rightarrow \infty} \frac{\partial \mathbf{b}_N(\beta)}{\partial \beta'},$$

$$W \equiv \text{plim}_{N \rightarrow \infty} W_N,$$

$$S_0 \equiv \frac{1}{N} \sum_{i=1}^N \mathbb{E}[z_i u_i u_i' z_i'].$$

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
$\Delta y_{iT} = \alpha \Delta y_{iT-1} + \Delta v_{iT}$	$y_{i0}, y_{i1}, y_{i2}, \dots, y_{iT-2}$	$\mathbb{E} \left[\Delta v_{iT} \begin{pmatrix} y_{i0} \\ y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT-2} \end{pmatrix} \right] = \mathbf{0}$

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

Moment conditions in matrix form

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

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

and the **sample analogue** is:

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

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

$$\begin{aligned} \hat{\alpha}_{GMM} &= \arg \min_{\alpha} \left(\frac{1}{N} \sum_{i=1}^N \Delta \mathbf{v}_i'(\alpha) Z_i \right) W_N \left(\frac{1}{N} \sum_{i=1}^N Z_i' \Delta \mathbf{v}_i(\alpha) \right) = \\ &= (\Delta \mathbf{y}_{-1}' Z W_N Z' \Delta \mathbf{y}_{-1})^{-1} \Delta \mathbf{y}_{-1}' Z W_N Z' \Delta \mathbf{y}. \end{aligned}$$

Optimal weighting matrix

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

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

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

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

Windeijer (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 $\mathbb{E}[\Delta \mathbf{v}_i \Delta \mathbf{v}_i']$:

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

GMM in Cigarette Demand

GMM results in our example are:

	Coefficient	Standard Error
Least Squares (OLS)	0.982	(0.003)
Within Groups (WG)	0.884	(0.061)
Anderson-Hsiao	1.395	(0.090)
One-step GMM	1.023	(0.104)
Two-step GMM	0.994	(0.040)
Two-step GMM small sample	0.994	(0.121)

Potential limitations of Arellano-Bond

Weak instruments:

- When $\alpha \rightarrow 1$, **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.

Dynamic linear model

Once we include **regressors**, the model is:

$$y_{it} = \alpha y_{it-1} + \mathbf{x}_{it}'\boldsymbol{\beta} + \eta_i + v_{it} \quad |\alpha| < 1.$$

We maintain the **previous assumptions**: error components, serially uncorrelated shocks, and predetermined initial conditions.

Therefore, moment conditions of the kind:

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

are **still valid**.

Assumptions on regressors

Different assumptions regarding \mathbf{x}_{it} will enable different additional **orthogonality conditions**:

- \mathbf{x}_{it} can be correlated or uncorrelated with η_i .
- \mathbf{x}_{it} can be endogenous, predetermined, or strictly exogenous with respect to v_{it} .

For instance, if assumptions are analogous to those for y_{it-1} , we may use \mathbf{x}_{it-1} (and longer lags) as instruments:

$$Z_i = \begin{pmatrix} y_{i0} & \mathbf{x}'_{i0} & \mathbf{x}'_{i1} & \dots & 0 & \dots & 0 & \mathbf{0}' & \dots & \mathbf{0}' \\ \vdots & \vdots & \vdots & \ddots & \vdots & \dots & \vdots & \vdots & \vdots & \vdots \\ 0 & \mathbf{0}' & \mathbf{0}' & \dots & y_{i0} & \dots & y_{iT-2} & \mathbf{x}'_{i0} & \dots & \mathbf{x}'_{iT-1} \end{pmatrix}.$$

Dynamic Cigarette Demand with Regressors

In our example, we rewrite the employment equations as follows:

$$\ln C_{it} = \alpha \ln C_{it-1} + \beta_1 \ln P_{it} + \beta_2 \ln Y_{it} + \eta_i + v_{it}.$$

Results are:

	OLS	WG	GMM
Lagged dep (α)	0.947 (0.011)	0.528 (0.064)	0.495 (0.127)
ln Prices (β_1)	0.010 (0.006)	-0.501 (0.098)	-0.607 (0.143)
ln Income (β_2)	0.049 (0.011)	0.369 (0.044)	0.338 (0.051)

System GMM estimation

Additional orthogonality conditions

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

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

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

$$\mathbb{E}[\Delta y_{it}\eta_i] = 0, \quad \forall t$$

or, alternatively:

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

The System GMM estimator

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

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

where:

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

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

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

System-GMM in Cigarette Demand

GMM results in our example are:

	Coefficient	Standard Error
Least Squares (OLS)	0.982	(0.003)
Within Groups (WG)	0.884	(0.061)
Anderson-Hsiao	1.395	(0.090)
One-step GMM	1.023	(0.104)
Two-step GMM	0.994	(0.040)
Two-step GMM small sample	0.994	(0.121)
One-step System-GMM	0.926	(0.023)
Two-step System-GMM small	0.911	(0.032)

Specification tests

Specification tests

There are several relevant aspects for the validity of the estimation that can be **tested formally**.

- **Orthogonality conditions:** are they small enough to not reject that they are zero (overidentifying restrictions).
- **Validity** of a subset of more restrictive assumptions (difference Sargan test, Hausman test).
- **Serial correlation in the data:** vital for the validity of the instruments (Arellano-Bond test).

Sargan/Hansen overidentifying restrictions test

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:

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

where $\hat{\mathbf{u}}$ are those predicted residuals from the first step and $\hat{\hat{\mathbf{u}}}$ are predicted from the second stage, and:

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

Testing stronger assumptions

Some of the assumptions that we make are **stronger** than others.

If the problem is overidentified, we can **test** whether results change if we **include or exclude** the orthogonality conditions generated by them.

If they are true, increase **efficiency**, but if not, **inconsistent**!

Two ways of testing it:

- **Overidentifying restrictions** (difference in Sargan should be close to zero: $\Delta S = S - S^A \stackrel{a}{\sim} \chi^2(L - L^A)$).
- **Differences in coefficients** (Hausman test: $\hat{q} = \hat{\delta}_{GMM}^A - \hat{\delta}_{GMM}$).

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:

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

where $\Delta \mathbf{v}_{-2}$ is the second lagged residual in differences, and $\Delta \mathbf{v}_*$ 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**.