Chapter 2. Dynamic panel data models

School of Economics and Management - University of Geneva

Christophe Hurlin, Université of Orléans

University of Orléans

April 2018

1. Introduction

Definition (Dynamic panel data model)

We now consider a dynamic panel data model, in the sense that it contains (at least) one lagged dependent variables. For simplicity, let us consider

$$y_{it} = \gamma y_{i,t-1} + \beta^{'} x_{it} + \alpha_i^* + \varepsilon_{it}$$

for $i = 1, .., n$ and $t = 1, .., T$. α_i^* and λ_t are the (unobserved) individual and time-specific effects, and ε_{it} the error (idiosyncratic) term with $\mathrm{E}(\varepsilon_{it}) = 0$, and $\mathrm{E}(\varepsilon_{it}\varepsilon_{js}) = \sigma_\varepsilon^2$ if $j = i$ and $t = s$, and $\mathrm{E}(\varepsilon_{it}\varepsilon_{js}) = 0$ otherwise.

1. Introduction

Remark

In a dynamic panel model, the choice between a fixed-effects formulation and a random-effects formulation has implications for estimation that are of a **different nature** than those associated with the static model.

1. Introduction

Dynamic panel issues

1. If lagged dependent variables appear as explanatory variables, strict exogeneity of the regressors no longer holds. The **LSDV is no longer consistent when** n tends to infinity and T is fixed.
2. The **initial values of a dynamic process** raise another problem. It turns out that with a random-effects formulation, the interpretation of a model depends on the assumption of initial observation.
3. The consistency property of the MLE and the GLS estimator also depends on the way in which T and n tend to infinity.

Introduction

The outline of this chapter is the following:

- **Section 1:** Introduction
- **Section 2:** Dynamic panel bias
- **Section 3:** The IV (Instrumental Variable) approach
  - **Subsection 3.1:** Reminder on IV and 2SLS
  - **Subsection 3.2:** Anderson and Hsiao (1982) approach
- **Section 4:** The GMM (Generalized Method of Moment) approach
  - **Subsection 4.1:** General presentation of GMM
  - **Subsection 4.2:** Application to dynamic panel data models

Section 2

The Dynamic Panel Bias

2. The dynamic panel bias

Objectives

1. Introduce the **AR(1)** panel data model.
2. Derive the semi-asymptotic bias of the **LSDV estimator**.
3. Understand the sources of the **dynamic panel bias** or **Nickell's bias**.
4. Evaluate the **magnitude** of this bias in a simple AR(1) model.
5. Asses this bias by **Monte Carlo** simulations.

Dynamic panel bias

1. The LSDV estimator is consistent for the static model whether the effects are fixed or random.
2. On the contrary, the LSDV is inconsistent for a dynamic panel data model with individual effects, whether the effects are fixed or random.

2. The dynamic panel bias

Definition (Nickell's bias)

The biais of the LSDV estimator in a dynamic model is generaly known as dynamic panel bias or Nickell's bias (1981).

Nickell, S. (1981). Biases in Dynamic Models with Fixed Effects, *Econometrica*, 49, 1399–1416.

Anderson, T.W., and C. Hsiao (1982). Formulation and Estimation of Dynamic Models Using Panel Data, *Journal of Econometrics*, 18, 47–82.

2. The dynamic panel bias

Definition (AR(1) panel data model)

Consider the simple AR(1) model

$$y_{it} = \gamma y_{i,t-1} + \alpha_i^* + \varepsilon_{it}$$

for $i = 1, .., n$ and $t = 1, .., T$. For simplicity, let us assume that

$$\alpha_i^* = \alpha + \alpha_i$$

to avoid imposing the restriction that $\sum_{i=1}^{n} \alpha_i = 0$ or $\mathbb{E}(\alpha_i) = 0$ in the case of random individual effects.

Assumptions

1. The autoregressive parameter γ satisfies
$$|\gamma| < 1$$
2. The initial condition y_{i0} is observable.
3. The error term satisfies with $\mathbb{E}(\varepsilon_{it}) = 0$, and $\mathbb{E}(\varepsilon_{it}\varepsilon_{js}) = \sigma_{\varepsilon}^2$ if $j = i$ and $t = s$, and $\mathbb{E}(\varepsilon_{it}\varepsilon_{js}) = 0$ otherwise.

2. The dynamic panel bias

Dynamic panel bias

In this AR(1) panel data model, we will show that

$$\underset{n \to \infty}{\text{plim}} \; \widehat{\gamma}_{LSDV} \neq \gamma \quad \textbf{dynamic panel bias}$$

$$\underset{n,T \to \infty}{\text{plim}} \; \widehat{\gamma}_{LSDV} = \gamma$$

2. The dynamic panel bias

The LSDV estimator is defined by (cf. chapter 1)

$$\widehat{\alpha}_i = \overline{y}_i - \widehat{\gamma}_{LSDV} \overline{y}_{i,-1}$$

$$\widehat{\gamma}_{LSDV} = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{i,t-1} - \overline{y}_{i,-1} \right)^2 \right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{i,t-1} - \overline{y}_{i,-1} \right) \left(y_{it} - \overline{y}_i \right) \right)$$

$$\overline{x}_i = \frac{1}{T} \sum_{t=1}^{T} x_{it} \quad \overline{y}_i = \frac{1}{T} \sum_{t=1}^{T} y_{it} \quad \overline{y}_{i,-1} = \frac{1}{T} \sum_{t=1}^{T} y_{i,t-1}$$

2. The dynamic panel bias

Definition (bias)

The bias of the LSDV estimator is defined by:

$$\widehat{\gamma}_{LSDV} - \gamma = \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{i,t-1} - \overline{y}_{i,-1} \right)^2 \right)^{-1} \left(\sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{i,t-1} - \overline{y}_{i,-1} \right) \left(\varepsilon_{it} - \overline{\varepsilon}_{i} \right) \right)$$

2. The dynamic panel bias

The bias of the LSDV estimator can be rewritten as:

$$\widehat{\gamma}_{LSDV} - \gamma = \frac{\sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{i,t-1} - \overline{y}_{i,-1} \right) \left(\varepsilon_{it} - \overline{\varepsilon}_{i} \right) / \left(nT \right)}{\sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{i,t-1} - \overline{y}_{i,-1} \right)^{2} / \left(nT \right)}$$

2. The dynamic panel bias

Let us consider the numerator. Because ε_{it} are (1) uncorrelated with α_i^* and (2) are independently and identically distributed, we have

$$\operatorname*{plim}_{n\to\infty} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{i,t-1} - \overline{y}_{i,-1}\right)\left(\varepsilon_{it} - \overline{\varepsilon}_i\right)$$

$$= \underbrace{\operatorname*{plim}_{n\to\infty} \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} y_{i,t-1}\varepsilon_{it}}_{\mathrm{N}_1} - \underbrace{\operatorname*{plim}_{n\to\infty} \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} y_{i,t-1}\overline{\varepsilon}_i}_{\mathrm{N}_2}$$

$$- \underbrace{\operatorname*{plim}_{n\to\infty} \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \overline{y}_{i,-1}\varepsilon_{it}}_{\mathrm{N}_3} + \underbrace{\operatorname*{plim}_{n\to\infty} \frac{1}{nT} \sum_{t=1}^{T} \sum_{i=1}^{n} \overline{y}_{i,-1}\overline{\varepsilon}_i}_{\mathrm{N}_4}$$

2. The dynamic panel bias

Theorem (Weak law of large numbers, Khinchine)

If $\{X_i\}$, for $i = 1, .., m$ is a sequence of i.i.d. random variables with $\mathbb{E}(X_i) = \mu < \infty$, then the sample mean converges in probability to μ:

$$\frac{1}{m}\sum_{i=1}^{m} X_i \xrightarrow{p} \mathbb{E}(X_i) = \mu$$

or

$$\underset{m\rightarrow\infty}{plim}\ \frac{1}{m}\sum_{i=1}^{m} X_i = \mathbb{E}(X_i) = \mu$$

2. The dynamic panel bias

By application of the WLLN (Khinchine's theorem)

$$\mathrm{N}_1 = \underset{n \to \infty}{\text{plim}} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} y_{i,t-1} \varepsilon_{it} = \mathbb{E}\left(y_{i,t-1} \varepsilon_{it}\right)$$

Since (1) $y_{i,t-1}$ only depends on $\varepsilon_{i,t-1}$, $\varepsilon_{i,t-2}$ and (2) the ε_{it} are uncorrelated, then we have

$$\mathbb{E}\left(y_{i,t-1} \varepsilon_{it}\right) = 0$$

and finally

$$\mathrm{N}_1 = \underset{n \to \infty}{\text{plim}} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} y_{i,t-1} \varepsilon_{it} = 0$$

2. The dynamic panel bias

For the second term N_2, we have:

$$\begin{aligned}
\mathrm{N}_2 &= \operatorname*{plim}_{n\to\infty} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} y_{i,t-1} \overline{\varepsilon}_i \\
&= \operatorname*{plim}_{n\to\infty} \frac{1}{nT} \sum_{i=1}^{n} \overline{\varepsilon}_i \sum_{t=1}^{T} y_{i,t-1} \\
&= \operatorname*{plim}_{n\to\infty} \frac{1}{nT} \sum_{i=1}^{n} \overline{\varepsilon}_i \, T \overline{y}_{i,-1} \qquad \text{as} \quad \overline{y}_{i,-1} = \frac{1}{T} \sum_{t=1}^{T} y_{i,t-1} \\
&= \operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \overline{\varepsilon}_i \overline{y}_{i,-1}
\end{aligned}$$

2. The dynamic panel bias

In the same way:

$$N_3 = \underset{n\to\infty}{\text{plim}} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \overline{y}_{i,-1} \varepsilon_{it} = \underset{n\to\infty}{\text{plim}} \frac{1}{nT} \sum_{i=1}^{n} \overline{y}_{i,-1} \sum_{t=1}^{T} \varepsilon_{it} = \underset{n\to\infty}{\text{plim}} \frac{1}{n} \sum_{i=1}^{n} \overline{y}_{i,-1} \overline{\varepsilon}$$

$$N_4 = \underset{n\to\infty}{\text{plim}} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \overline{y}_{i,-1} \overline{\varepsilon}_{i} = \underset{n\to\infty}{\text{plim}} \frac{1}{nT} T \sum_{i=1}^{n} \overline{y}_{i,-1} \overline{\varepsilon}_{i} = \underset{n\to\infty}{\text{plim}} \frac{1}{n} \sum_{i=1}^{n} \overline{y}_{i,-1} \overline{\varepsilon}_{i}$$

2. The dynamic panel bias

The numerator of the bias expression can be rewritten as

$$
\begin{aligned}
& \operatorname*{plim}_{n\to\infty} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{i,t-1} - \overline{y}_{i,-1} \right) \left(\varepsilon_{it} - \overline{\varepsilon}_{i} \right) \\
= \quad & \underbrace{0}_{\mathrm{N}_1} - \underbrace{\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \overline{\varepsilon}_{i} \overline{y}_{i,-1}}_{\mathrm{N}_2} - \underbrace{\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \overline{y}_{i,-1} \overline{\varepsilon}_{i}}_{\mathrm{N}_3} + \underbrace{\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \overline{y}_{i,-1} \overline{\varepsilon}_{i}}_{\mathrm{N}_4} \\
= \quad & - \operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \overline{y}_{i,-1} \overline{\varepsilon}_{i}
\end{aligned}
$$

2. The dynamic panel bias

Solution

The numerator of the expression of the LSDV bias satisfies:

$$\underset{n\rightarrow\infty}{plim}\ \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{i,t-1}-\overline{y}_{i,-1}\right)\left(\varepsilon_{it}-\overline{\varepsilon}_{i}\right)=-\ \underset{n\rightarrow\infty}{plim}\ \frac{1}{n}\sum_{i=1}^{n}\overline{y}_{i,-1}\overline{\varepsilon}_{i}$$

2. The dynamic panel bias

Remark

$$\widehat{\gamma}_{LSDV} - \gamma = \frac{\sum\limits_{i=1}^{n} \sum\limits_{t=1}^{T} \left(y_{i,t-1} - \overline{y}_{i,-1} \right) \left(\varepsilon_{it} - \overline{\varepsilon}_{i} \right) / \left(nT \right)}{\sum\limits_{i=1}^{n} \sum\limits_{t=1}^{T} \left(y_{i,t-1} - \overline{y}_{i,-1} \right)^{2} / \left(nT \right)}$$

$$\underset{n \to \infty}{\text{plim}} \; \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{i,t-1} - \overline{y}_{i,-1} \right) \left(\varepsilon_{it} - \overline{\varepsilon}_{i} \right) = - \underset{n \to \infty}{\text{plim}} \; \frac{1}{n} \sum_{i=1}^{n} \overline{y}_{i,-1} \overline{\varepsilon}_{i}$$

If this plim is not null, then the LSDV estimator $\widehat{\gamma}_{LSDV}$ is **biased** when n tends to infinity and T is fixed.

2. The dynamic panel bias

Let us examine this plim

$$\underset{n\to\infty}{\text{plim}} \ \frac{1}{n}\sum_{i=1}^{n} \overline{y}_{i,-1}\overline{\varepsilon}_i$$

We know that

$$\begin{aligned}
y_{it} &= \gamma y_{i,t-1} + \alpha_i^* + \varepsilon_{it} \\
&= \gamma^2 y_{i,t-2} + \alpha_i^* \left(1+\gamma\right) + \varepsilon_{it} + \gamma\varepsilon_{i,t-1} \\
&= \gamma^3 y_{i,t-3} + \alpha_i^* \left(1+\gamma+\gamma^2\right) + \varepsilon_{it} + \gamma\varepsilon_{i,t-1} + \gamma^2\varepsilon_{i,t-2} \\
&= \ldots \\
&= \gamma^t y_{i0} + \frac{1-\gamma^t}{1-\gamma}\alpha_i^* + \varepsilon_{it} + \gamma\varepsilon_{i,t-1} + \gamma^2\varepsilon_{i,t-2} + \ldots + \gamma^{t-1}\varepsilon_{i1}
\end{aligned}$$

2. The dynamic panel bias

For any time t, we have:

$$\begin{aligned} y_{it} &= \varepsilon_{it} + \gamma \varepsilon_{i,t-1} + \gamma^2 \varepsilon_{i,t-2} + ... + \gamma^{t-1} \varepsilon_{i1} \\ &\quad + \frac{1-\gamma^t}{1-\gamma} \alpha_i^* + \gamma^t y_{i0} \end{aligned}$$

For $y_{i,t-1}$, we have:

$$\begin{aligned} y_{i,t-1} &= \varepsilon_{i,t-1} + \gamma \varepsilon_{i,t-2} + \gamma^2 \varepsilon_{i,t-3} + ... + \gamma^{t-2} \varepsilon_{i1} \\ &\quad + \frac{1-\gamma^{t-1}}{1-\gamma} \alpha_i^* + \gamma^{t-1} y_{i0} \end{aligned}$$

2. The dynamic panel bias

$$y_{i,t-1} = \varepsilon_{i,t-1} + \gamma\varepsilon_{i,t-2} + \gamma^2\varepsilon_{i,t-3} + ... + \gamma^{t-2}\varepsilon_{i1} + \frac{1-\gamma^{t-1}}{1-\gamma}\alpha_i^* + \gamma^{t-1}y_{i0}$$

Summing $y_{i,t-1}$ over t, we get:

$$\begin{aligned} \sum_{t=1}^{T} y_{i,t-1} &= \varepsilon_{i,T-1} + \frac{1-\gamma^2}{1-\gamma}\varepsilon_{i,T-2} + ... + \frac{1-\gamma^{T-1}}{1-\gamma}\varepsilon_{i1} \\ &\quad + \frac{(T-1) - T\gamma + \gamma^T}{(1-\gamma)^2}\alpha_i^* + \frac{1-\gamma^T}{1-\gamma}y_{i0} \end{aligned}$$

2. The dynamic panel bias

$$y_{i,t-1} = \varepsilon_{i,t-1} + \gamma\varepsilon_{i,t-2} + \gamma^2\varepsilon_{i,t-3} + ... + \gamma^{t-2}\varepsilon_{i1} + \frac{1-\gamma^{t-1}}{1-\gamma}\alpha_i^* + \gamma^{t-1}y_{i0}$$

Proof: We have (each lign corresponds to a date)

$$\begin{aligned}
\sum_{t=1}^{T} y_{i,t-1} &= y_{i,T-1} + y_{i,T-2} + .. + y_{i,1} + y_{i,0} \\
&= \varepsilon_{i,T-1} + \gamma\varepsilon_{i,T-2} + .. + \gamma^{T-2}\varepsilon_{i1} + \frac{1-\gamma^{T-1}}{1-\gamma}\alpha_i^* + \gamma^{T-1}y_{i0} \\
&\quad +\varepsilon_{i,T-2} + \gamma\varepsilon_{i,T-3} + ... + \gamma^{T-3}\varepsilon_{i1} + \frac{1-\gamma^{T-2}}{1-\gamma}\alpha_i^* + \gamma^{T-2}y_{i0} \\
&\quad +.. \\
&\quad +\varepsilon_{i,1} + \frac{1-\gamma^{1}}{1-\gamma}\alpha_i^* + \gamma y_{i0} \\
&\quad +y_{i0}
\end{aligned}$$

2. The dynamic panel bias

Proof (ct'd): For the individual effect α_i^*, we have

$$
\begin{aligned}
& \frac{\alpha_i^*}{1-\gamma}\left(1-\gamma+1-\gamma^2+...+1-\gamma^{T-1}\right) \\
= & \frac{\alpha_i^*}{1-\gamma}\left(T-1-\gamma-\gamma^2-..-\gamma^{T-1}\right) \\
= & \frac{\alpha_i^*}{1-\gamma}\left(T-\frac{1-\gamma^T}{1-\gamma}\right) \\
= & \frac{\alpha_i^*\left(T-T\gamma-1+\gamma^T\right)}{(1-\gamma)^2}
\end{aligned}
$$

2. The dynamic panel bias

So, we have

$$\begin{aligned}
\overline{y}_{i,-1} &= \frac{1}{T}\sum_{t=1}^{T} y_{i,t-1} \\
&= \frac{1}{T}\left(\varepsilon_{i,T-1} + \frac{1-\gamma^2}{1-\gamma}\varepsilon_{i,T-2} + ... + \frac{1-\gamma^{T-1}}{1-\gamma}\varepsilon_{i1} \right. \\
&\quad \left. + \frac{\left(T - T\gamma - 1 + \gamma^T\right)}{\left(1-\gamma\right)^2}\alpha_i^* + \frac{1-\gamma^T}{1-\gamma} y_{i0}\right)
\end{aligned}$$

2. The dynamic panel bias

Finally, the plim is equal to

$$
\begin{aligned}
& \operatorname*{plim}_{n\rightarrow\infty} \frac{1}{n}\sum_{i=1}^{n} \overline{y}_{i,-1}\overline{\varepsilon}_{i} \\
= \quad & \operatorname*{plim}_{n\rightarrow\infty} \frac{1}{n}\sum_{i=1}^{n} \left\{ \frac{1}{T} \left(\varepsilon_{i,t-1} + \frac{1-\gamma^{2}}{1-\gamma}\varepsilon_{i,t-2} + ... + \frac{1-\gamma^{T-1}}{1-\gamma}\varepsilon_{i1} \right.\right. \\
& \left.\left. + \frac{\left(T - T\gamma - 1 + \gamma^{T}\right)}{\left(1-\gamma\right)^{2}}\alpha_{i}^{*} + \frac{1-\gamma^{T}}{1-\gamma}y_{i0} \right) \times \frac{1}{T}\left(\varepsilon_{i1} + ... + \varepsilon_{iT}\right) \right\}
\end{aligned}
$$

2. The dynamic panel bias

Because ε_{it} are i.i.d, by a law of large numbers, we have:

$$
\begin{aligned}
& \underset{n\to\infty}{\text{plim}} \ \frac{1}{n}\sum_{i=1}^{n} \overline{y}_{i,-1}\overline{\varepsilon}_i \\
= \ & \underset{n\to\infty}{\text{plim}} \ \frac{1}{n}\sum_{i=1}^{n} \left\{ \frac{1}{T}\left(\varepsilon_{i,T-1} + \frac{1-\gamma^2}{1-\gamma}\varepsilon_{i,T-2} + ... + \frac{1-\gamma^{T-1}}{1-\gamma}\varepsilon_{i1} \right.\right. \\
& \left.\left. + \frac{\left(T - T\gamma - 1 + \gamma^T\right)}{\left(1-\gamma\right)^2}\alpha_i^* + \frac{1-\gamma^T}{1-\gamma}y_{i0} \right) \times \frac{1}{T}\left(\varepsilon_{i1} + ... + \varepsilon_{iT}\right) \right\} \\
= \ & \frac{\sigma_\varepsilon^2}{T^2}\left(\frac{1-\gamma}{1-\gamma} + \frac{1-\gamma^2}{1-\gamma} + ... + \frac{1-\gamma^{T-1}}{1-\gamma} \right) \\
= \ & \frac{\sigma_\varepsilon^2}{T^2}\frac{\left(T - T\gamma - 1 + \gamma^T\right)}{\left(1-\gamma\right)^2}
\end{aligned}
$$

2. The dynamic panel bias

Theorem

If the errors terms ε_{it} are i.i.d.$\left(0, \sigma_{\varepsilon}^{2}\right)$, we have:

$$\begin{aligned}
& \underset{n \to \infty}{plim} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{i,t-1} - \overline{y}_{i,-1}\right) \left(\varepsilon_{it} - \overline{\varepsilon}_{i}\right) \\
= \quad & -\underset{n \to \infty}{plim} \frac{1}{n} \sum_{i=1}^{n} \overline{y}_{i,-1} \overline{\varepsilon}_{i} \\
= \quad & -\frac{\sigma_{\varepsilon}^{2}}{T^{2}} \frac{\left(T - T\gamma - 1 + \gamma^{T}\right)}{\left(1-\gamma\right)^{2}}
\end{aligned}$$

2. The dynamic panel bias

By similar manipulations, we can show that the denominator of $\widehat{\gamma}_{LSDV}$ converges to:

$$\begin{aligned}
& \underset{n \to \infty}{\text{plim}} \frac{1}{nT} \sum_{i=1}^{n} \sum_{t=1}^{T} \left(y_{i,t-1} - \overline{y}_{i,-1} \right)^2 \\
= & \frac{\sigma_\varepsilon^2}{1-\gamma^2} \left(1 - \frac{1}{T} - \frac{2\gamma}{\left(1-\gamma\right)^2} \times \frac{\left(T - T\gamma - 1 + \gamma^T \right)}{T^2} \right)
\end{aligned}$$

2. The dynamic panel bias

So, we have :

$$
\begin{aligned}
& \underset{n\rightarrow\infty}{\text{plim}} \left(\widehat{\gamma}_{LSDV} - \gamma\right) \\
= \quad & \underset{n\rightarrow\infty}{\text{plim}} - \frac{\frac{1}{nT}\sum\limits_{i=1}^{n}\sum\limits_{t=1}^{T}\left(y_{i,t-1} - \overline{y}_{i,-1}\right)\left(\varepsilon_{it} - \overline{\varepsilon}_{i}\right)}{\frac{1}{nT}\sum\limits_{i=1}^{n}\sum\limits_{t=1}^{T}\left(y_{i,t-1} - \overline{y}_{i,-1}\right)^{2}} \\
= \quad & -\frac{\frac{\sigma_{\varepsilon}^{2}}{T^{2}}\frac{\left(T-T\gamma-1+\gamma^{T}\right)}{(1-\gamma)^{2}}}{\frac{\sigma_{\varepsilon}^{2}}{1-\gamma^{2}}\left(1-\frac{1}{T}-\frac{2\gamma}{(1-\gamma)^{2}}\times\frac{\left(T-T\gamma-1+\gamma^{T}\right)}{T^{2}}\right)}
\end{aligned}
$$

2. The dynamic panel bias

This semi-asymptotic bias can be rewriten as:

$$\begin{aligned}
& \operatorname*{plim}_{n\to\infty} \left(\widehat{\gamma}_{LSDV} - \gamma\right) \\
= & -\frac{\left(T - T\gamma - 1 + \gamma^T\right)}{\left(\frac{1-\gamma}{1+\gamma}\right)\left(T^2 - T - \frac{2\gamma}{(1-\gamma)^2} \times \left(T - T\gamma - 1 + \gamma^T\right)\right)} \\
= & -\frac{(1+\gamma)\left(T - T\gamma - 1 + \gamma^T\right)}{(1-\gamma)\left(T^2 - T - \frac{2\gamma}{(1-\gamma)^2} \times \left(T - T\gamma - 1 + \gamma^T\right)\right)}
\end{aligned}$$

2. The dynamic panel bias

Fact

If T also tends to infinity, then the numerator converges to zero, and denominator converges to a nonzero constant $\sigma_{\varepsilon}^{2} / \left(1-\gamma^{2}\right)$, hence the LSDV estimator of γ and α_i are consistent.

Fact

If T is fixed, then the denominator is a nonzero constant, and $\widehat{\gamma}_{LSDV}$ and $\widehat{\alpha}_i$ are inconsistent estimators when n is large.

2. The dynamic panel bias

Theorem (Dynamic panel bias)

In a dynamic panel AR(1) model with individual effects, the semi-asymptotic bias (with n) of the LSDV estimator on the autoregressive parameter is equal to:

$$\underset{n\rightarrow\infty}{plim}\left(\widehat{\gamma}_{LSDV}-\gamma\right)=-\frac{(1+\gamma)\left(T-T\gamma-1+\gamma^{T}\right)}{(1-\gamma)\left(T^{2}-T-\frac{2\gamma}{(1-\gamma)^{2}}\times\left(T-T\gamma-1+\gamma^{T}\right)\right)}$$

2. The dynamic panel bias

Theorem (Dynamic panel bias)

For an AR(1) model, the dynamic panel bias can be rewriten as :

$$\begin{aligned}\underset{n\rightarrow\infty}{plim}\left(\widehat{\gamma}_{LSDV}-\gamma\right) &= -\frac{1+\gamma}{T-1}\left(1-\frac{1}{T}\frac{1-\gamma^{T}}{1-\gamma}\right)\\ &\times\left(1-\frac{2\gamma}{\left(1-\gamma\right)\left(T-1\right)}\left(1-\frac{1-\gamma^{T}}{T\left(1-\gamma\right)}\right)\right)^{-1}\end{aligned}$$

2. The dynamic panel bias

Fact

The dynamic bias of $\widehat{\gamma}_{LSDV}$ is caused by having to eliminate the individual effects α_i^* from each observation, which creates a correlation of order $(1/T)$ between the explanatory variables and the residuals in the transformed model

$$\begin{aligned}(y_{it} - \overline{y}_i) &= \gamma \left(y_{i,t-1} - \underbrace{\overline{y}_{i,-1}}_{\text{depends on past value of } \varepsilon_{it}} \right) \\ &\quad + \left(\varepsilon_{it} - \underbrace{\overline{\varepsilon}_i}_{\text{depends on past value of } \varepsilon_{it}} \right)\end{aligned}$$

2. The dynamic panel bias

Intuition of the dynamic bias

$$\left(y_{it}-\overline{y}_{i}\right)=\gamma\left(y_{i,t-1}-\overline{y}_{i,-1}\right)+\left(\varepsilon_{it}-\overline{\varepsilon}_{i}\right)$$

with $cov\left(\overline{y}_{i,-1},\overline{\varepsilon_{i}}\right)\neq 0$ since

$$\begin{aligned}
cov\left(\overline{y}_{i,-1},\overline{\varepsilon_{i}}\right) &= cov\left(\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1},\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\right)\\
&= cov\left(\frac{1}{T}\sum_{t=1}^{T}y_{i,t-1},\frac{1}{T}\sum_{t=1}^{T}\varepsilon_{it}\right)\\
&= \frac{1}{T^{2}}cov\left(\left(y_{i1}+...+y_{iT-1}\right),\left(\varepsilon_{i1}+...+\varepsilon_{iT}\right)\right)
\end{aligned}$$

Intuition of the dynamic bias

$$(y_{it} - \overline{y}_i) = \gamma (y_{i,t-1} - \overline{y}_{i,-1}) + (\varepsilon_{it} - \overline{\varepsilon}_i) \quad \text{with } cov (\overline{y}_{i,-1}, \overline{\varepsilon_i}) \neq 0$$

If we approximate y_{it} by ε_{it} (in fact y_{it} also depend on $\varepsilon_{it-1,}$ $\varepsilon_{t-2}, ...$) then we have

$$\begin{aligned} cov (\overline{y}_{i,-1}, \overline{\varepsilon_i}) &= \frac{1}{T^2} cov ((y_{i1} + ... + y_{iT-1}), (\varepsilon_{i1} + ... + \varepsilon_{iT})) \\ &\simeq \frac{1}{T^2} (cov (\varepsilon_{i,1}, \varepsilon_{i,1}) + ... + (cov (\varepsilon_{i,T-1}, \varepsilon_{i,T-1}))) \\ &\simeq \frac{(T-1)\, \sigma^2_{\varepsilon}}{T^2} \neq 0 \end{aligned}$$

Intuition of the dynamic bias

$$(y_{it} - \overline{y}_i) = \gamma \left(y_{i,t-1} - \overline{y}_{i,-1} \right) + (\varepsilon_{it} - \overline{\varepsilon}_i) \quad \text{with } cov\left(\overline{y}_{i,-1}, \overline{\varepsilon_i} \right) \neq 0$$

If we approximate y_{it} by ε_{it} then we have

$$cov\left(\overline{y}_{i,-1}, \overline{\varepsilon_i} \right) = \frac{(T-1)\,\sigma_\varepsilon^2}{T^2}$$

By taking into account all the interaction terms, we have shown that

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \overline{y}_{i,-1} \overline{\varepsilon}_i = cov\left(\overline{y}_{i,-1}, \overline{\varepsilon_i} \right) = \frac{\sigma_\varepsilon^2}{T^2} \frac{\left((T-1)\,\gamma - 1 + \gamma^T \right)}{(1-\gamma)^2}$$

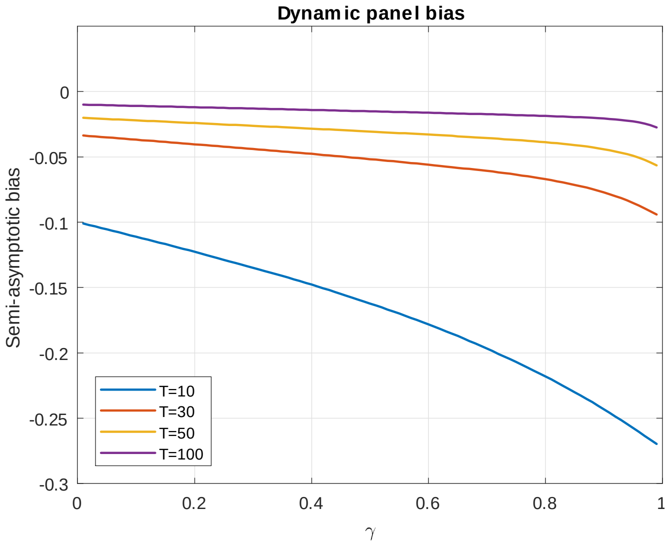
2. The dynamic panel bias

Remarks

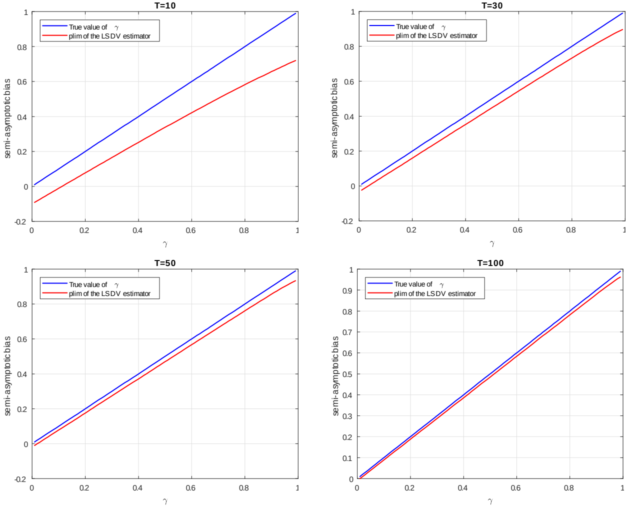
$$\begin{aligned}\operatorname*{plim}_{n\to\infty}\left(\widehat{\gamma}_{LSDV}-\gamma\right) &= -\frac{1+\gamma}{T-1}\left(1-\frac{1}{T}\frac{1-\gamma^{T}}{1-\gamma}\right)\\ &\quad\times\left(1-\frac{2\gamma}{\left(1-\gamma\right)\left(T-1\right)}\left(1-\frac{1-\gamma^{T}}{T\left(1-\gamma\right)}\right)\right)^{-1}\end{aligned}$$

1. When T is large, the right-hand-side variables become asymptotically uncorrelated.
2. For small T, this bias is always negative if $\gamma > 0$.
3. The bias does not go to zero as γ goes to zero.

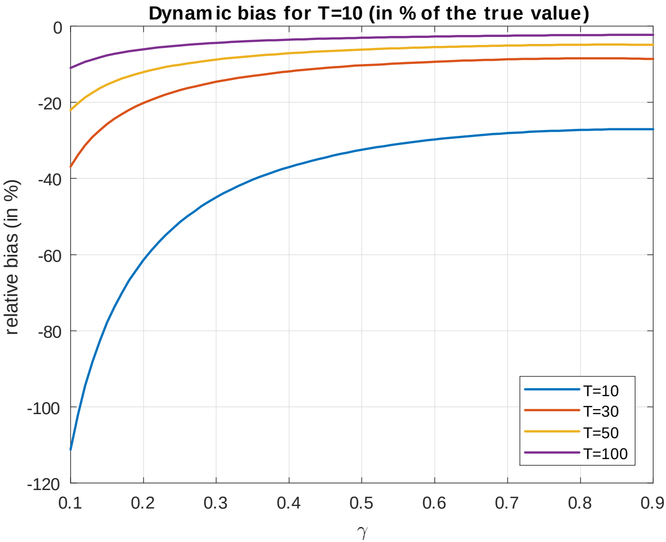
2. The dynamic panel bias



2. The dynamic panel bias



2. The dynamic panel bias



Monte Carlo experiments

How to check these semi-asymptotic formula with Monte Carlo simulations?

2. The dynamic panel bias

Step 1: parameters

- Let assume that $\gamma = 0.5$, $\sigma_{\varepsilon}^{2} = 1$ and $\varepsilon_{it} \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$.
- Simulate n individual effects α_i^* once at all. For instance, we can use a uniform distribution

$$\alpha_i^* \sim U_{[-1,1]}$$

Step 2: Monte Carlo pseudo samples

- Simulate n (typically $n = 1,000$) i.i.d. sequences $\{\varepsilon_{it}\}_{t=1}^{T}$ for a given value of T (typically $T = 10$)
- Generate n sequences $\{y_{it}\}_{t=1}^{T}$ for $i = 1, .., n$ with the model:

$$y_{it} = \gamma y_{i,t-1} + \alpha_i^* + \varepsilon_{it}$$

- Repeat S times the step 2 in order to generate $S = 5,000$ sequences $\left\{y_{it}^{(s)}\right\}_{t=1}^{T}$ for $s = 1, .., S$ for each cross-section unit $i = 1, ..., n$

2. The dynamic panel bias

Step 3: LSDV estimates on pseudo series

- For each pseudo sample $s = 1, ..., S$, consider the empirical model

$$y_{it}^{s} = \gamma y_{i,t-1}^{s} + \alpha_i + \mu_{it} \quad i = 1, .., n \quad t = 1, ... T$$

and compute the LSDV estimates $\widehat{\gamma}_{LSDV}^{s}$.

- Compute the **average bias** of the LSDV estimator $\widehat{\gamma}_{LSDV}$ based on the S Monte Carlo simulations

$$av.bias = \frac{1}{S} \sum_{s=1}^{S} \widehat{\gamma}_{LSDV}^{s} - \gamma$$

Step 4: Semi-asymptotic bias

1. Repeat this experiment for various **cross-section dimensions** n: when n increases,the average bias should converge to

$$\begin{aligned}\operatorname*{plim}_{n\to\infty}\left(\widehat{\gamma}_{LSDV}-\gamma\right) &= -\frac{1+\gamma}{T-1}\left(1-\frac{1}{T}\frac{1-\gamma^{T}}{1-\gamma}\right)\\ &\times\left(1-\frac{2\gamma}{(1-\gamma)(T-1)}\left(1-\frac{1-\gamma^{T}}{T(1-\gamma)}\right)\right)^{-1}\end{aligned}$$

2. Repeat this this experiment for various **time dimensions** T: when T increases,the average bias should converge to 0.

2. The dynamic panel bias

```
%================================================
%=== Chapter 2: Monte Carlo simulations
%=== Dynamic Panel semi-asymptotic Bias
%=== C. Hurlin, University of Orleans, 2018
%================================================
clc, clear, close all

% === Step 1: Parameters ===
gam=0.5;                              % Autogressive parameter
T=10;                                 % Time dimension T
n=1000;                              % Cross-sectional dimension n
alpha=unifrnd(-1,1,1,n);              % Individual effects
S=5000;                               % Number of Monte Carlo replications
```

2. The dynamic panel bias

```
% === Step 2: Simulations ===
ys=zeros(T+1,n,S);                  % Initialization o ys(it)
gam_hat=NaN(S,1);                   % Vector of parameters
eps=normrnd(0,1,T+1,n,S);           % Simulation on the errors terms

for s=1:S

    for t=2:T+1
        ys(t,:,s)=alpha+gam*ys(t-1,:,s)+eps(t,:,s);
    end

    % === Step 3: LSDV estimation ===
    ysb0=ys(2:end,:,s)-mean(ys(2:end,:,s));         % ysb(it)=y(it)-y_bar(i)
    ysb0=ysb0(:);                                   % Vector of ysb(i,t)
    ysb1=ys(1:end-1,:,s)-mean(ys(1:end-1,:,s));     % ysb(i,t-1)=y(i,t-1)-y_bar(i
    ysb1=ysb1(:);                                   % Vector of ysb(i,t-1)
    gam_hat(s)=pinv(ysb1'*ysb1)*ysb1'*ysb0;         % LSDV estimates

end
```

2. The dynamic panel bias

```
disp(' '), disp('Adjusted sample size T ')
disp(T)

theoretical_bias=-((T-1)-T.*gam+gam.^T).*(1+gam)./...
    ((T.^2-T-(2.*gam./((1-gam).^2)).*((T-1)-T.*gam+gam.^T)).*(1-gam));
disp(' '), disp('Theoretical bias')
disp(theoretical_bias)

disp(' '), disp('Average bias (Monte Carlo simulation)')
disp(mean(gam_hat)-gam)

figure
hist(gam_hat)
title('Histogram of the LSDV estimates for \gamma=0.5, T=10 and n=50')
grid('on')
xlabel('\gamma_hat')
ylabel('Number of simulations')
print 'Figure_Monte_Carlo_1' -depsc2;
```

2. The dynamic panel bias

Command Window

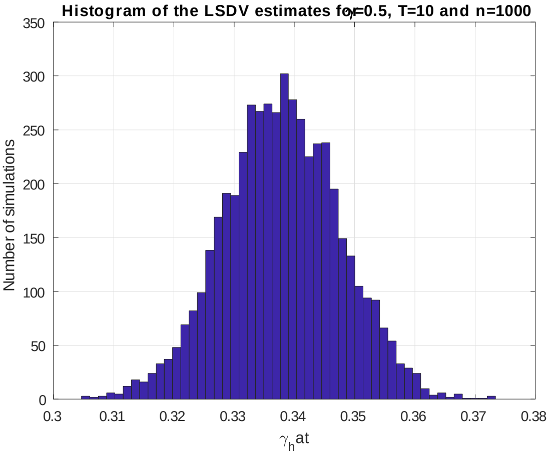
```
Adjusted sample size T
    10


Theoretical bias
   -0.1622


Average bias (Monte Carlo simulation)
   -0.1623

fx >>
```

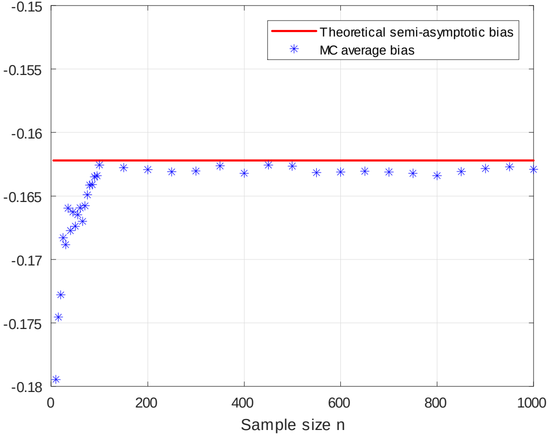
2. The dynamic panel bias



2. The dynamic panel bias

Click me!

2. The dynamic panel bias



2. The dynamic panel bias

Question: What is the importance of the dynamic bias in micro-panels?

> *"Macroeconomists should not dismiss the LSDV bias as insignificant. Even with a time dimension T as large as 30, we find that the bias may be equal to as much 20% of the true value of the coefficient of interest." (Judson et Owen, 1999, page 10)*

Judson R.A. and Owen A. (1999), Estimating dynamic panel data models: a guide for macroeconomists. *Economics Letters*, 1999, vol. 65, issue 1, 9-15.

2. The dynamic panel bias

Finite Sample results (Monte Carlo simulations)

n	T	γ	Avg. $\widehat{\gamma}_{LSDV}$	Avg. bias
10	10	0.5	0.3282	-0.1718
50	10	0.5	0.3317	-0.1683
100	10	0.5	0.3338	-0.1662
10	50	0.5	0.4671	-0.0329
50	50	0.5	0.4688	-0.0321
100	50	0.5	0.4694	-0.0306

2. The dynamic panel bias

Finite Sample results (Monte Carlo simulations)

n	T	γ	Avg. $\widehat{\gamma}_{LSDV}$	Avg. bias
10	10	-0.3	-0.3686	-0.0686
50	10	-0.3	-0.3743	-0.0743
100	10	-0.3	-0.3753	-0.0753
10	50	-0.3	-0.3134	-0.0134
50	50	-0.3	-0.3133	-0.0133
100	50	-0.5	-0.3142	-0.0142

2. The dynamic panel bias

Fact (smearing effect)

The LSDV for dynamic individual-effects model remains biased with the **introduction of exogenous variables** *if T is small; for details of the derivation, see Nickell (1981); Kiviet (1995).*

$$y_{it} = \alpha + \gamma y_{i,t-1} + \beta' x_{it} + \alpha_i + \varepsilon_{it}$$

In this case, **both estimators $\widehat{\gamma}_{LSDV}$ and $\widehat{\beta}_{LSDV}$ are biased**.

2. The dynamic panel bias

What are the solutions?

Consistent estimator of γ can be obtained by using:

1. ML or FIML (but additional assumptions on y_{i0} are necessary)
2. Feasible GLS (but additional assumptions on y_{i0} are necessary)
3. LSDV bias corrected (Kiviet, 1995)
4. IV approach (Anderson and Hsiao, 1982)
5. GMM approach (Arenallo and Bond, 1985)

2. The dynamic panel bias

What are the solutions?

Consistent estimator of γ can be obtained by using:

1. ML or FIML (but additional assumptions on y_{i0} are necessary)
2. Feasible GLS (but additional assumptions on y_{i0} are necessary)
3. LSDV bias corrected (Kiviet, 1995)
4. **IV approach (Anderson and Hsiao, 1982)**
5. **GMM approach (Arenallo and Bond, 1985)**

Key Concepts Section 2

1. AR(1) panel data model
2. Semi-asymptotic bias
3. Dynamic panel bias (Nickell's bias)
4. Monte Carlo experiments
5. Magnitude of the dynamic panel bias

Section 3

The Instrumental Variable (IV) approach

Subsection 3.1

Reminder on IV and 2SLS

3.1 Reminder on IV and 2SLS

Objectives

1. Define the **endogeneity bias** and the **smearing effect**.
2. Define the notion of **instrument** or **instrumental variable**.
3. Introduce the **exogeneity** and **relevance** properties of an instrument.
4. Introduce the notion of **just-identified** and **over-identified** systems.
5. Define the **IV** estimator and its asymptotic variance.
6. Define the **2SLS** estimator and its asymptotic variance.
7. Define the notion of **weak instrument**.

3.1 Reminder on IV and 2SLS

Consider the (population) **multiple linear regression model**:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

- $\mathbf{y}$ is a $N \times 1$ vector of observations y_j for $j = 1, .., N$
- $\mathbf{X}$ is a $N \times K$ matrix of K explicative variables $\mathbf{x}_{jk}$ for $k = 1, ., K$ and $j = 1, .., N$
- $\boldsymbol{\beta} = (\beta_1 .. \beta_K)'$ is a $K \times 1$ vector of parameters
- $\boldsymbol{\varepsilon}$ is a $N \times 1$ vector of error terms ε_i with (spherical disturbances)

$$\mathbb{V}\left(\boldsymbol{\varepsilon} \middle| \mathbf{X} \right) = \sigma^2 \mathbf{I}_N$$

3.1 Reminder on IV and 2SLS

Endogeneity we assume that the assumption A3 (**exogeneity**) is violated:

$$\mathbb{E}\left(\boldsymbol{\varepsilon} \middle| \mathbf{X}\right) \neq \mathbf{0}_{N \times 1}$$

with

$$\text{plim} \frac{1}{N} \mathbf{X}' \boldsymbol{\varepsilon} = \mathbb{E}\left(\mathbf{x}_j \varepsilon_j\right) = \gamma \neq \mathbf{0}_{K \times 1}$$

3.1 Reminder on IV and 2SLS

Theorem (Bias of the OLS estimator)

If the regressors are endogenous, i.e. $\mathbb{E}\left(\varepsilon|\,\mathbf{X}\right) \neq 0$, *the OLS estimator of* $\boldsymbol{\beta}$ *is biased*

$$\mathbb{E}\left(\widehat{\boldsymbol{\beta}}_{OLS}\right) \neq \boldsymbol{\beta}$$

where $\boldsymbol{\beta}$ *denotes the true value of the parameters. This bias is called the* **endogeneity bias**.

3.1 Reminder on IV and 2SLS

Theorem (Inconsistency of the OLS estimator)

If the regressors are **endogenous** *with plim* $N^{-1}\mathbf{X}'\varepsilon = \gamma$*, the OLS estimator of* $\boldsymbol{\beta}$ *is inconsistent*

$$plim\ \widehat{\boldsymbol{\beta}}_{OLS} = \boldsymbol{\beta} + \mathbf{Q}^{-1}\gamma$$

where $\mathbf{Q} =$ *plim* $N^{-1}\mathbf{X}'\mathbf{X}$.

3.1 Reminder on IV and 2SLS

Proof: Given the definition of the OLS estimator:

$$\begin{aligned}
\widehat{\beta}_{OLS} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} \\
&= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{X}\beta+\varepsilon\right) \\
&= \beta+\left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\varepsilon\right)
\end{aligned}$$

We have:

$$\begin{aligned}
\text{plim } \widehat{\beta}_{OLS} &= \beta+\text{plim}\left(\frac{1}{N}\mathbf{X}'\mathbf{X}\right)^{-1}\times\text{plim}\left(\frac{1}{N}\mathbf{X}'\varepsilon\right) \\
&= \beta+\mathbf{Q}^{-1}\gamma\neq\beta
\end{aligned}$$

Remarks

$$\text{plim } \widehat{\beta}_{OLS} = \beta + \mathbf{Q}^{-1}\gamma$$

1. The implication is that even though only one of the variables in $\mathbf{X}$ is correlated with ε, **all of the elements of $\widehat{\beta}_{OLS}$ are inconsistent**, not just the estimator of the coefficient on the endogenous variable.

2. This effects is called **smearing effect**: the inconsistency due to the endogeneity of the one variable is smeared across all of the least squares estimators.

Example (Endogeneity, OLS estimator and smearing)

Consider the multiple linear regression model

$$y_i = 0.4 + 0.5x_{i1} - 0.8x_{i2} + \varepsilon_i$$

where ε_i is $i.i.d.$ with $\mathbb{E}(\varepsilon_i)$. We assume that the vector of variables defined by $\mathbf{w}_i = (x_{i1} \ : x_{i2} : \varepsilon_i)$ has a multivariate normal distribution with

$$\mathbf{w}_i \sim N\left(\mathbf{0}_{3\times 1}, \boldsymbol{\Delta}\right)$$

with

$$\boldsymbol{\Delta} = \begin{pmatrix} 1 & 0.3 & 0 \\ 0.3 & 1 & 0.5 \\ 0 & 0.5 & 1 \end{pmatrix}$$

It means that $\mathbb{C}ov(\varepsilon_i, x_{i1}) = 0$ (x_1 is **exogenous**) but $\mathbb{C}ov(\varepsilon_i, x_{i2}) = 0.5$ (x_2 is endogenous) and $\mathbb{C}ov(x_{i1,}x_{i2}) = 0.3$ (x_1 is correlated to x_2).

3.1 Reminder on IV and 2SLS

Example (Endogeneity, OLS estimator and smearing (cont'd))

Write a Matlab code to (1) generate $S = 1,000$ samples $\{y_i, x_{i1}, x_{i2}\}_{i=1}^N$ of size $N = 10,000$. (2) For each simulated sample, determine the OLS estimators of the model

$$y_i = \beta_1 + \beta_2 x_{i1} + \beta_3 x_{i2} + \varepsilon_i$$

Denote $\widehat{\boldsymbol{\beta}}_s = \left(\widehat{\beta}_{1s} \ \widehat{\beta}_{2s} \ \widehat{\beta}_{3s}\right)'$ the OLS estimates obtained from the simulation $s \in \{1, ..S\}$. (3) compare the true value of the parameters in the population (DGP) to the average OLS estimates obtained for the S simulations

3.1 Reminder on IV and 2SLS

```
clear all; clc ; close all

% Data Generating Process
beta0=[0.4 0.5 -0.8]';                    % True value of the parameters
K=length(beta0);                          % Number of explicative variables
nsim=1000;                                % Number of simuations
N=10000;                                  % Sample size
gam=0.5;                                  % corr(x(:,end),eps)=gam
Sigma=[1 0.3 0 ; 0.3 1 gam; 0 gam 1];     % Matrix Sigma
beta=zeros(K,nsim);                       % Initialisation
for i=1:nsim
    R = mvnrnd(zeros(3,1),Sigma,N);       % Multivariate normal
    X=[ones(N,1) R(:,1:2)];               % Explicative variables N(0,4)
    eps=R(:,3);                           % Error term with E(x(:,end).eps)=gam
    y=X*beta0+eps;                        % Dependent variable
    beta(:,i)=X\y;                        % OLS estimates
end
disp('    beta0     beta_ols (average)')
disp([beta0 mean(beta,2)])
```

3.1 Reminder on IV and 2SLS

beta0	beta_ols (average)
0.4000	0.3999
0.5000	0.3353
-0.8000	-0.2504

3.1 Reminder on IV and 2SLS

Question: What is the solution to the endogeneity issue?

The use of **instruments**..

3.1 Reminder on IV and 2SLS

Definition (Instruments)

Consider a set of H variables $\mathbf{z}_h \in \mathbb{R}^N$ for $h = 1, ..N$. Denote $\mathbf{Z}$ the $N \times H$ matrix $\left(\mathbf{z}_1 : .. : \mathbf{z}_H\right)$. These variables are called **instruments** or **instrumental variables** if they satisfy two properties:

(1) **Exogeneity**: They are uncorrelated with the disturbance.

$$\mathbb{E}\left(\varepsilon | \mathbf{Z}\right) = \mathbf{0}_{N \times 1}$$

(2) **Relevance**: They are correlated with the independent variables, $\mathbf{X}$.

$$\mathbb{E}\left(x_{jk} z_{jh}\right) \neq 0$$

for $h \in \{1, .., H\}$ and $k \in \{1, .., K\}$.

3.1 Reminder on IV and 2SLS

Assumptions: The instrumental variables satisfy the following properties.

Well behaved data:

$$\text{plim} \frac{1}{N} \mathbf{Z}'\mathbf{Z} = \mathbf{Q}_{ZZ} \text{ a finite } H \times H \text{ positive definite matrix}$$

Relevance:

$$\text{plim} \frac{1}{N} \mathbf{Z}'\mathbf{X} = \mathbf{Q}_{ZX} \text{ a finite } H \times K \text{ positive definite matrix}$$

Exogeneity:

$$\text{plim} \frac{1}{N} \mathbf{Z}'\boldsymbol{\varepsilon} = \mathbf{0}_{K \times 1}$$

3.1 Reminder on IV and 2SLS

Definition (Instrument properties)

We assume that the H instruments are **linearly independent**:

$$\mathbb{E}\left(\mathbf{Z}'\mathbf{Z}\right) \quad \text{is non singular}$$

or equivalently

$$\text{rank}\left(\mathbb{E}\left(\mathbf{Z}'\mathbf{Z}\right)\right) = H$$

3.1 Reminder on IV and 2SLS

The **exogeneity condition**

$$\mathbb{E}\left(\left.\varepsilon_j\right|\mathbf{z}_j\right)=0\Longrightarrow\mathbb{E}\left(\varepsilon_j\mathbf{z}_j\right)=0_H$$

can expressed as an **orthogonality condition** or **moment condition**

$$\mathbb{E}\left(\underset{(H,1)}{\mathbf{z}_j}\underset{(1,1)}{\left(y_j-\mathbf{x}_j'\beta\right)}\right)=\underset{(H,1)}{0_H}$$

So, we have H **equations and** K **unknown parameters** β

3.1 Reminder on IV and 2SLS

Definition (Identification)

The system is **identified** if there exists a unique vector $\boldsymbol{\beta}$ such that:

$$\mathbb{E}\left(\mathbf{z}_j\left(y_j - \mathbf{x}_j'\boldsymbol{\beta}\right)\right) = 0$$

where $\mathbf{z}_j = \left(z_{j1}..z_{jH}\right)'$. For that, we have the following conditions:

(1) If $H < K$ the model is **not identified**.

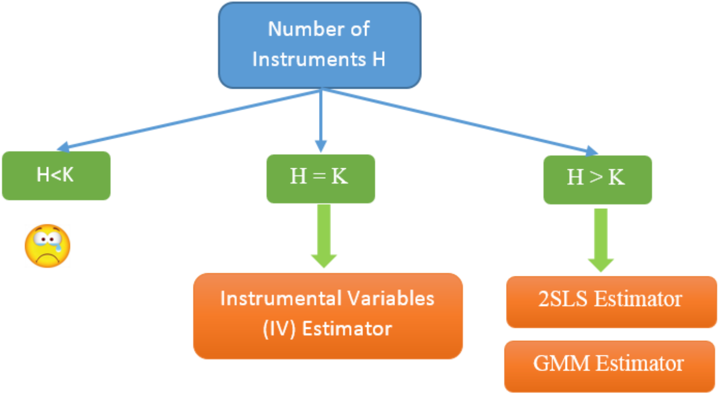
(2) If $H = K$ the model is **just-identified**.

(3) If $H > K$ the model is **over-identified**.

Remark

1. **Under-identification**: less equations (H) than unknowns (K)....
2. **Just-identification**: number of equations equals the number of unknowns (unique solution)...=> **IV estimator**
3. **Over-identification**: more equations than unknowns. Two equivalent solutions:
   1. Select K linear combinations of the instruments to have a unique solution)...=> **Two-Stage Least Squares (2SLS)**
   2. Set the sample analog of the moment conditions as close as possible to zero, i.e. minimize the distance between the sample analog and zero given a metric (optimal metric or optimal weighting matrix?) => **Generalized Method of Moments (GMM)**.

3.1 Reminder on IV and 2SLS



3.1 Reminder on IV and 2SLS

Assumption: Consider a **just-identified** model

$$H = K$$

3.1 Reminder on IV and 2SLS

Motivation of the IV estimator

By definition of the instruments:

$$\text{plim}\frac{1}{N}\mathbf{Z}'\boldsymbol{\varepsilon} = \text{plim}\frac{1}{N}\mathbf{Z}'\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right) = \mathbf{0}_{K\times 1}$$

So, we have:

$$\text{plim}\frac{1}{N}\mathbf{Z}'\mathbf{y} = \left(\text{plim}\frac{\mathbf{1}}{\mathbf{N}}\mathbf{Z}'\mathbf{X}\right)\ \boldsymbol{\beta}$$

or equivalently

$$\boldsymbol{\beta} = \left(\text{plim}\frac{\mathbf{1}}{\mathbf{N}}\mathbf{Z}'\mathbf{X}\right)^{-1}\text{plim}\frac{1}{N}\mathbf{Z}'\mathbf{y}$$

3.1 Reminder on IV and 2SLS

Definition (Instrumental Variable (IV) estimator)

If $H = K$, the **Instrumental Variable (IV) estimator** $\widehat{\beta}_{IV}$ of parameters $\boldsymbol{\beta}$ is defined as to be:

$$\widehat{\beta}_{IV} = \left(\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{Z}'\mathbf{y}$$

3.1 Reminder on IV and 2SLS

Definition (Consistency)

Under the assumption that plim $N^{-1}\mathbf{Z}'\varepsilon = \mathbf{0}$, the IV estimator $\widehat{\boldsymbol{\beta}}_{IV}$ is **consistent**:

$$\widehat{\boldsymbol{\beta}}_{IV} \xrightarrow{p} \boldsymbol{\beta}$$

where $\boldsymbol{\beta}$ denotes the true value of the parameters.

3.1 Reminder on IV and 2SLS

Proof: By definition:

$$\widehat{\beta}_{IV} = \left(\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{Z}'\mathbf{y} = \beta + \left(\frac{1}{N}\mathbf{Z}'\mathbf{X}\right)^{-1}\left(\frac{1}{N}\mathbf{Z}'\varepsilon\right)$$

So, we have:

$$\text{plim}\widehat{\boldsymbol{\beta}}_{IV} = \boldsymbol{\beta} + \left(\text{plim}\frac{1}{N}\mathbf{Z}'\mathbf{X}\right)^{-1}\left(\text{plim}\frac{1}{N}\mathbf{Z}'\varepsilon\right)$$

Under the assumption of exogeneity of the instruments

$$\text{plim}\frac{1}{N}\mathbf{Z}'\varepsilon = \mathbf{0}_{K\times 1}$$

So, we have

$$\text{plim}\ \widehat{\boldsymbol{\beta}}_{IV} = \boldsymbol{\beta} \quad \square$$

3.1 Reminder on IV and 2SLS

Definition (Asymptotic distribution)

Under some regularity conditions, the IV estimator $\widehat{\beta}_{IV}$ is **asymptotically normally distributed**:

$$\sqrt{N}\left(\widehat{\beta}_{IV}-\beta\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0}_{K\times 1},\sigma^2\mathbf{Q}_{ZX}^{-1}\mathbf{Q}_{ZZ}\mathbf{Q}_{ZX}^{-1}\right)$$

where

$$\underset{K\times K}{\mathbf{Q}_{ZZ}}=\text{plim}\frac{1}{N}\mathbf{Z}'\mathbf{Z}\qquad\underset{K\times K}{\mathbf{Q}_{ZX}}=\text{plim}\frac{1}{N}\mathbf{Z}'\mathbf{X}$$

3.1 Reminder on IV and 2SLS

Definition (Asymptotic variance covariance matrix)

The **asymptotic variance covariance matrix** of the IV estimator $\widehat{\beta}_{IV}$ is defined as to be:

$$\mathbb{V}_{asy}\left(\widehat{\beta}_{IV}\right) = \frac{\sigma^2}{N}\mathbf{Q}_{ZX}^{-1}\mathbf{Q}_{ZZ}\mathbf{Q}_{ZX}^{-1}$$

A consistent estimator is given by

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_{IV}\right) = \widehat{\sigma}^2\left(\mathbf{Z}'\mathbf{X}\right)^{-1}\left(\mathbf{Z}'\mathbf{Z}\right)\left(\mathbf{X}'\mathbf{Z}\right)^{-1}$$

Remarks

1. If the system is just identified $H = K$,

$$\left(\mathbf{Z}'\mathbf{X}\right)^{-1} = \left(\mathbf{X}'\mathbf{Z}\right)^{-1}$$

$$\mathbf{Q}_{ZX} = \mathbf{Q}_{XZ}$$

   the estimator can also written as

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_{IV}\right) = \widehat{\sigma}^2 \left(\mathbf{Z}'\mathbf{X}\right)^{-1} \left(\mathbf{Z}'\mathbf{Z}\right) \left(\mathbf{Z}'\mathbf{X}\right)^{-1}$$

2. As usual, the estimator of the variance of the error terms is:

$$\widehat{\sigma}^2 = \frac{\widehat{\boldsymbol{\varepsilon}}'\widehat{\boldsymbol{\varepsilon}}}{N-K} = \frac{1}{N-K}\sum_{i=1}^{N}\left(y_i - \mathbf{x}_i'\widehat{\beta}_{IV}\right)^2$$

3.1 Reminder on IV and 2SLS

Relevant instruments

1. Our analysis thus far has focused on the "**identification**" condition for IV estimation, that is, the "**exogeneity assumption**," which produces

$$\text{plim}\frac{1}{N}\mathbf{Z}'\varepsilon = \mathbf{0}_{K\times 1}$$

2. A growing literature has argued that greater attention needs to be given to the relevance condition

$$\text{plim}\frac{1}{N}\mathbf{Z}'\mathbf{X} = \mathbf{Q}_{ZX} \text{ a finite } H \times K \text{ positive definite matrix}$$

with $H = K$ in the case of a just-identified model.

3.1 Reminder on IV and 2SLS

Relevant instruments (cont'd)

$$\text{plim}\frac{1}{N}\mathbf{Z}'\mathbf{X} = \mathbf{Q}_{ZX} \text{ a finite } H \times K \text{ positive definite matrix}$$

1. While strictly speaking, this condition is sufficient to determine the asymptotic properties of the IV estimator
2. However, the common case of "**weak instruments**," is only barely true has attracted considerable scrutiny.

3.1 Reminder on IV and 2SLS

Definition (Weak instrument)

A **weak instrument** is an instrumental variable which is only slightly correlated with the right-hand-side variables $\mathbf{X}$. In presence of weak instruments, the quantity $\mathbf{Q}_{ZX}$ is close to zero and we have

$$\frac{1}{N}\mathbf{Z}'\mathbf{X} \simeq \mathbf{0}_{H\times K}$$

3.1 Reminder on IV and 2SLS

Fact (IV estimator and weak instruments)

In presence of **weak instruments**, *the IV estimators* $\widehat{\beta}_{IV}$ *has a poor precision (great variance). For* $\mathbf{Q}_{ZX} \simeq \mathbf{0}_{H \times K}$, *the asymptotic variance tends to be very large, since:*

$$\mathbb{V}_{asy}\left(\widehat{\beta}_{IV}\right) = \frac{\sigma^2}{N}\mathbf{Q}_{ZX}^{-1}\mathbf{Q}_{ZZ}\mathbf{Q}_{ZX}^{-1}$$

As soon as $N^{-1}\mathbf{Z}'\mathbf{X} \simeq \mathbf{0}_{H \times K}$, *the estimated asymptotic variance covariance is also very large since*

$$\widehat{\mathbb{V}}_{asy}\left(\widehat{\beta}_{IV}\right) = \widehat{\sigma}^2\left(\mathbf{Z}'\mathbf{X}\right)^{-1}\left(\mathbf{Z}'\mathbf{Z}\right)\left(\mathbf{X}'\mathbf{Z}\right)^{-1}$$

3.1 Reminder on IV and 2SLS

Assumption: Consider an **over-identified** model

$$H > K$$

3.1 Reminder on IV and 2SLS

Introduction

If $\mathbf{Z}$ contains more variables than $\mathbf{X}$, then much of the preceding derivation is unusable, because $\mathbf{Z}'\mathbf{X}$ will be $H \times K$ with

$$\text{rank}\left(\mathbf{Z}'\mathbf{X}\right) = K < H$$

So, the matrix $\mathbf{Z}'\mathbf{X}$ has no inverse and we cannot compute the IV estimator as:

$$\widehat{\beta}_{IV} = \left(\mathbf{Z}'\mathbf{X}\right)^{-1} \mathbf{Z}'\mathbf{y}$$

Introduction (cont'd)

The crucial assumption in the previous section was the **exogeneity assumption**

$$\text{plim}\frac{1}{N}\mathbf{Z}'\varepsilon = \mathbf{0}_{K\times 1}$$

1. That is, every column of $\mathbf{Z}$ is asymptotically uncorrelated with ε.
2. That also means that every **linear combination** of the columns of $\mathbf{Z}$ is also uncorrelated with ε, which suggests that one approach would be to choose K linear combinations of the columns of $\mathbf{Z}$.

3.1 Reminder on IV and 2SLS

Introduction (cont'd)

Which linear combination to choose?

A choice consists in using is the **projection of the columns** of $\mathbf{X}$ in the column space of $\mathbf{Z}$:

$$\widehat{\mathbf{X}} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}$$

With this choice of instrumental variables, $\widehat{\mathbf{X}}$ for $\mathbf{Z}$, we have

$$\begin{aligned}
\widehat{\beta}_{2SLS} &= \left(\widehat{\mathbf{X}}'\mathbf{X}\right)^{-1}\widehat{\mathbf{X}}'\mathbf{y} \\
&= \left(\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{y}
\end{aligned}$$

3.1 Reminder on IV and 2SLS

Definition (Two-stage Least Squares (2SLS) estimator)

The **Two-stage Least Squares (2SLS)** estimator of the parameters $\boldsymbol{\beta}$ is defined as to be:

$$\widehat{\beta}_{2SLS} = \left(\widehat{\mathbf{X}}'\mathbf{X}\right)^{-1} \widehat{\mathbf{X}}'\mathbf{y}$$

where $\widehat{\mathbf{X}} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}$ corresponds to the projection of the columns of $\mathbf{X}$ in the column space of $\mathbf{Z}$, or equivalently by

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

3.1 Reminder on IV and 2SLS

Remark

By definition

$$\widehat{\beta}_{2SLS} = \left(\widehat{\mathbf{X}}'\mathbf{X}\right)^{-1}\widehat{\mathbf{X}}'\mathbf{y}$$

Since

$$\widehat{\mathbf{X}} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X} = \mathbf{P}_Z\mathbf{X}$$

where $\mathbf{P}_Z$ denotes the projection matrix on the columns of $\mathbf{Z}$. Reminder: $\mathbf{P}_Z$ is symmetric and $\mathbf{P}_Z\mathbf{P}_Z' = \mathbf{P}_Z$. So, we have

$$\begin{aligned}
\widehat{\beta}_{2SLS} &= \left(\mathbf{X}'\mathbf{P}_Z'\mathbf{X}\right)^{-1}\widehat{\mathbf{X}}'\mathbf{y} \\
&= \left(\mathbf{X}'\mathbf{P}_Z'\mathbf{P}_Z\mathbf{X}\right)^{-1}\widehat{\mathbf{X}}'\mathbf{y} \\
&= \left(\widehat{\mathbf{X}}'\widehat{\mathbf{X}}\right)^{-1}\widehat{\mathbf{X}}'\mathbf{y}
\end{aligned}$$

3.1 Reminder on IV and 2SLS

Definition (Two-stage Least Squares (2SLS) estimator)

The **Two-stage Least Squares (2SLS)** estimator of the parameters $\boldsymbol{\beta}$ can also be defined as:

$$\widehat{\boldsymbol{\beta}}_{2SLS} = \left(\widehat{\mathbf{X}}'\widehat{\mathbf{X}}\right)^{-1}\widehat{\mathbf{X}}'\mathbf{y}$$

It corresponds to the OLS estimator obtained in the regression of $\mathbf{y}$ on $\widehat{\mathbf{X}}$. Then, the 2SLS can be computed in two steps, first by computing $\widehat{\mathbf{X}}$, then by the least squares regression. That is why it is called the two-stage LS estimator.

3.1 Reminder on IV and 2SLS

A **procedure** to get the 2SLS estimator is the following

Step 1: Regress each explicative variable x_k (for $k = 1, ..K$) on the H instruments.

$$x_{kj} = \alpha_1 z_{1j} + \alpha_2 z_{2j} + .. + \alpha_H z_{Hj} + v_j$$

Step 2: Compute the OLS estimators $\widehat{\alpha}_h$ and the fitted values $\widehat{x}_{kj}$

$$\widehat{x}_{kj} = \widehat{\alpha}_1 z_{1j} + \widehat{\alpha}_2 z_{2j} + .. + \widehat{\alpha}_H z_{Hj}$$

Step 3: Regress the dependent variable y on the fitted values $\widehat{x}_{ki}$:

$$y_j = \beta_1 \widehat{x}_{1j} + \beta_2 \widehat{x}_{2j} + .. + \beta_K \widehat{x}_{Kj} + \varepsilon_j$$

The 2SLS estimator $\widehat{\boldsymbol{\beta}}_{2SLS}$ then corresponds to the OLS estimator obtained in this model.

3.1 Reminder on IV and 2SLS

Theorem

If any column of $\mathbf{X}$ *also appears in* $\mathbf{Z}$*, i.e. if one or more explanatory (exogenous) variable is used as an instrument, then that column of* $\mathbf{X}$ *is reproduced exactly in* $\widehat{\mathbf{X}}$*.*

Example (Explicative variables used as instrument)

Suppose that the regression contains K variables, only one of which, say, the K^{th}, is correlated with the disturbances, i.e. $\mathbb{E}\left(x_{Ki}\varepsilon_i\right) \neq 0$. We can use a set of instrumental variables z_1,..., z_J plus the other $K-1$ variables that certainly qualify as instrumental variables in their own right. So,

$$\mathbf{Z} = \left(z_1 : .. : z_J : x_1 : .. : x_{K-1}\right)$$

Then

$$\widehat{\mathbf{X}} = \left(x_1 : .. : x_{K-1} : \widehat{x}_K\right)$$

where $\widehat{x}_K$ denotes the projection of x_K on the columns of $\mathbf{Z}$.

3.1 Reminder on IV and 2SLS

Key Concepts SubSection 3.1

1. Endogeneity bias and smearing effect.
2. Instrument or instrumental variable.
3. Exogeneity and relavance properties of an instrument.
4. Instrumental Variable (IV) estimator.
5. Two-Stage Least Square (2SLS) estimator.
6. Weak instrument.

Subsection 3.2

Anderson and Hsiao (1982) IV approach

3.2 Anderson and Hsiao (1982) IV approach

Objectives

1. Introduce the **IV** approach of **Anderson and Hsiao (1982)**.
2. Describe their 4 steps **estimation procedure**.
3. Introduce the **first difference** transformation of the dynamic model.
4. Describe their choice of **instruments**.

3.2 Anderson and Hsiao (1982) IV approach

Consider a **dynamic panel data model** with **random individual effects**:

$$y_{it} = \gamma y_{i,t-1} + \beta^{'} x_{it} + \rho^{'} \omega_i + \alpha_i + \varepsilon_{it}$$

- α_i are the (unobserved) individual effects,
- x_{it} is a vector of K_1 time-varying explanatory variables,
- ω_i is a vector of K_2 time-invariant variables.

3.2 Anderson and Hsiao (1982) IV approach

Assumption: we assume that the component error term $v_{it} = \varepsilon_{it} + \alpha_i$

- $\mathbb{E}(\varepsilon_{it}) = 0$, $\mathbb{E}(\alpha_i) = 0$
- $\mathbb{E}(\varepsilon_{it}\varepsilon_{js}) = \sigma_\varepsilon^2$ if $j = i$ and $t = s$, 0 otherwise.
- $\mathbb{E}(\alpha_i\alpha_j) = \sigma_\alpha^2$ if $j = i$, 0 otherwise.
- $\mathbb{E}(\alpha_i x_{it}) = 0$, $\mathbb{E}(\alpha_i \omega_i) = 0$ (**exogeneity** assumption for ω_i)
- $\mathbb{E}(\varepsilon_{it} x_{it}) = 0$, $\mathbb{E}(\varepsilon_{it} \omega_i) = 0$ (**exogeneity** assumption for x_{it})

3.2 Anderson and Hsiao (1982) IV approach

The $K_1 + K_2 + 3$ **parameters** to estimate are

$$y_{it} = \gamma y_{i,t-1} + \beta^{'} x_{it} + \rho^{'} \omega_i + \alpha_i + \varepsilon_{it}$$

1. γ the autoregressive parameter,
2. β is the $K_1 \times 1$ vector of parameters for the time-varying explanatory variables,
3. ρ is the $K_2 \times 1$ vector of parameters for the time-invariant variables,
4. σ_{ε}^2 and σ_{α}^2 the variances of the error terms.

3.2 Anderson and Hsiao (1982) IV approach

Remark

If the vector ω_i includes a constant term, the associated parameter can be interpreted as the **mean** of the (random) individual effects

$$y_{it} = \gamma y_{i,t-1} + \beta^{'} x_{it} + \rho^{'} \omega_i + \alpha_i + \varepsilon_{it}$$

$$\alpha_i^* = \mu + \alpha_i \qquad \mathbb{E}\left(\alpha_i\right) = 0$$

$$\underset{(K_2,1)}{\omega_i} = \begin{pmatrix} 1 \\ z_{i2} \\ \cdots \\ z_{iK_2} \end{pmatrix} \qquad \underset{(K_2,1)}{\rho} = \begin{pmatrix} \mu \\ \rho_2 \\ \cdots \\ \rho_{K_2} \end{pmatrix}$$

3.2 Anderson and Hsiao (1982) IV approach

Vectorial form:

$$y_i = y_{i,-1}\gamma + X_i\beta + \omega_i^{'}\rho e + \alpha_i e + \varepsilon_i$$

- ε_i, y_i and $y_{i,-1}$ are $T \times 1$ vectors (T is the **adjusted sample size**),
- X_i a $T \times K_1$ matrix of time-varying explanatory variables,
- ω_i is a $K_2 \times 1$ vector of time-invariant variables,
- e is the $T \times 1$ unit vector, and

$$\mathbb{E}\left(\alpha_i\right) = 0 \quad \mathbb{E}\left(\alpha_i x_{it}^{'}\right) = 0 \quad \mathbb{E}\left(\alpha_i \omega_i^{'}\right) = 0$$

3.2 Anderson and Hsiao (1982) IV approach

In the dynamic panel data models context:

- The **Instrumental Variable (IV)** approach was first proposed by **Anderson and Hsiao (1982)**.
- They propose an IV procedure with 2 choices of instruments and **4 steps** to estimate γ, β, ρ and σ_{ε}^2.

Anderson, T.W., and C. Hsiao (1982). Formulation and Estimation of Dynamic Models Using Panel Data, *Journal of Econometrics*, 18, 47–82.

3.2 Anderson and Hsiao (1982) IV approach

The Anderson and Hsiao (1982) IV approach

1. **First step:** first difference transformation
2. **Second step:** choice of instruments and IV estimation of γ and β
3. **Third step:** estimation of ρ
4. **Fourth step:** estimation of the variances σ_α^2 and σ_ε^2

3.2 Anderson and Hsiao (1982) IV approach

The Anderson and Hsiao (1982) IV approach

1. **First step:** first difference transformation
2. **Second step:** choice of instruments and IV estimation of γ and β
3. **Third step:** estimation of ρ
4. **Fourth step:** estimation of the variances σ_{α}^{2} and σ_{ε}^{2}

3.2 Anderson and Hsiao (1982) IV approach

First step: first difference transformation

Taking the **first difference** of the model, we obtain for $t = 2, .., T$.

$$\left(y_{it} - y_{i,t-1}\right) = \gamma\left(y_{i,t-1} - y_{i,t-2}\right) + \beta^{'}\left(x_{it} - x_{i,t-1}\right) + \varepsilon_{it} - \varepsilon_{i,t-1}$$

- The first difference transformation leads to "lost" one observation.
- But, it allows to eliminate the individual effects (as the Within transformation).

3.2 Anderson and Hsiao (1982) IV approach

The Anderson and Hsiao (1982) IV approach

1. **First step:** first difference transformation
2. **Second step:** choice of instruments and IV estimation of γ and β
3. **Third step:** estimation of ρ
4. **Fourth step:** estimation of the variances σ_{α}^{2} and σ_{ε}^{2}

3.2 Anderson and Hsiao (1982) IV approach

Second step: choice of the instruments and IV estimation

$$(y_{it} - y_{i,t-1}) = \gamma (y_{i,t-1} - y_{i,t-2}) + \beta' (x_{it} - x_{i,t-1}) + \varepsilon_{it} - \varepsilon_{i,t-1}$$

A valid instrument z_{it} should satisfy

$$\mathbb{E} (z_{it} (\varepsilon_{it} - \varepsilon_{i,t-1})) = 0 \quad \textbf{Exogeneity} \text{ property}$$

$$\mathbb{E} (z_{it} (y_{i,t-1} - y_{i,t-2})) \neq 0 \quad \textbf{Relevance} \text{ property}$$

3.2 Anderson and Hsiao (1982) IV approach

Anderson and Hsiao (1982) propose two valid instruments:

1. **First instrument:** $z_{i,t} = y_{i,t-2}$

$$\mathbb{E}\left(y_{i,t-2}\left(\varepsilon_{it} - \varepsilon_{i,t-1}\right)\right) = 0 \quad \textbf{Exogeneity} \text{ property}$$

$$\mathbb{E}\left(y_{i,t-2}\left(y_{i,t-1} - y_{i,t-2}\right)\right) \neq 0 \quad \textbf{Relevance} \text{ property}$$

2. **Second instrument:** $z_{i,t} = \left(y_{i,t-2} - y_{i,t-3}\right)$

$$\mathbb{E}\left(\left(y_{i,t-2} - y_{i,t-3}\right)\left(\varepsilon_{it} - \varepsilon_{i,t-1}\right)\right) = 0 \quad \textbf{Exogeneity} \text{ property}$$

$$\mathbb{E}\left(\left(y_{i,t-2} - y_{i,t-3}\right)\left(y_{i,t-1} - y_{i,t-2}\right)\right) \neq 0 \quad \textbf{Relevance} \text{ property}$$

3.2 Anderson and Hsiao (1982) IV approach

Remarks

- The initial first differences model includes $K_1 + 1$ regressors.
- The regressor $(y_{i,t-1} - y_{i,t-2})$ is endogeneous.
- The regressors $(x_{it} - x_{i,t-1})$ are assumed to be exogeneous.

3.2 Anderson and Hsiao (1982) IV approach

Definition (Instruments)

Anderson and Hsiao (1982) consider two sets of $K_1 + 1$ instruments, in both cases the system is **just identified** (IV estimator):

$$\underset{(K_1+1,1)}{\mathbf{z}_i} = \left(\underset{(1,1)}{y_{i,t-2}} \; : \; \underset{(1,K_1)}{(x_{it} - x_{i,t-1})'} \right)'$$

$$\underset{(K_1+1,1)}{\mathbf{z}_i} = \left(\underset{(1,1)}{(y_{i,t-2} - y_{i,t-3})} \; : \; \underset{(1,K_1)}{(x_{it} - x_{i,t-1})'} \right)'$$

3.2 Anderson and Hsiao (1982) IV approach

IV estimator with the first set of instruments

$$\begin{pmatrix} \widehat{\gamma}_{IV} \\ \widehat{\beta}_{IV} \end{pmatrix} = \left(\mathbf{Z}'\mathbf{X}\right)^{-1} \mathbf{Z}'\mathbf{y} =$$

$$\left(\sum_{i=1}^{n} \sum_{t=2}^{T} \begin{pmatrix} \left(y_{i,t-1} - y_{i,t-2}\right) y_{i,t-2} & y_{i,t-2} \left(x_{it} - x_{i,t-1}\right)' \\ \left(x_{it} - x_{i,t-1}\right) y_{i,t-2} & \left(x_{it} - x_{i,t-1}\right) \left(x_{it} - x_{i,t-1}\right)' \end{pmatrix} \right)^{-1}$$

$$\times \left(\sum_{i=1}^{n} \sum_{t=2}^{T} \begin{pmatrix} y_{i,t-2} \\ x_{it} - x_{i,t-1} \end{pmatrix} \left(y_{i,t} - y_{i,t-1}\right) \right)$$

3.2 Anderson and Hsiao (1982) IV approach

IV estimator with the second set of instruments

$$\begin{pmatrix} \widehat{\gamma}_{IV} \\ \widehat{\beta}_{IV} \end{pmatrix} = \left(\mathbf{Z}'\mathbf{X}\right)^{-1} \mathbf{Z}'\mathbf{y} =$$

$$\left(\sum_{i=1}^{n} \sum_{t=3}^{T} \begin{matrix} \left(y_{i,t-1} - y_{i,t-2}\right)\left(y_{i,t-2} - y_{i,t-3}\right) & \left(y_{i,t-2} - y_{i,t-3}\right)\left(x_{it} - x_{i,t-1}\right. \\ \left(x_{it} - x_{i,t-1}\right)\left(y_{i,t-2} - y_{i,t-3}\right) & \left(x_{it} - x_{i,t-1}\right)\left(x_{it} - x_{i,t-1}\right)' \end{matrix} \right.$$

$$\times \left(\sum_{i=1}^{n} \sum_{t=3}^{T} \begin{pmatrix} y_{i,t-2} - y_{i,t-3} \\ x_{it} - x_{i,t-1} \end{pmatrix} \left(y_{i,t} - y_{i,t-1}\right) \right)$$

3. Instrumental variable (IV) estimators

Remarks

1. The first estimator (with $z_{it} = y_{i,t-2}$) has an advantage over the second one (with $z_{it} = y_{i,t-2} - y_{i,t-3}$), in that the minimum number of time periods required is two, whereas the first one requires $T \geq 3$.

2. In practice, if $T \geq 3$, the choice between both depends on the correlations between $(y_{i,t-1} - y_{i,t-2})$ and $y_{i,t-2}$ or $(y_{i,t-2} - y_{i,t-3})$ $=>$ **relevance assumption.**

Anderson, T.W., and C. Hsiao (1981). Estimation of Dynamic Models with Error Components, *Journal of the American Statistical Association*, 76, 598–606

3.2 Anderson and Hsiao (1982) IV approach

The Anderson and Hsiao (1982) IV approach

1. **First step:** first difference transformation
2. **Second step:** choice of instruments and IV estimation of γ and β
3. **Third step:** estimation of ρ
4. **Fourth step:** estimation of the variances σ_{α}^{2} and σ_{ε}^{2}

3.2 Anderson and Hsiao (1982) IV approach

Third step

$$y_{it} = \gamma y_{i,t-1} + \beta^{'} x_{it} + \rho_i^{'} \omega_i + \alpha_i + \varepsilon_{it}$$

- Given the estimates $\widehat{\gamma}_{IV}$ and $\widehat{\beta}_{IV}$, we can deduce an estimate of ρ, **the vector of parameters for the time-invariant variables ω_i.**
- Let us consider, the following equation

$$\overline{y}_i - \widehat{\gamma}_{IV} \overline{y}_{i,-1} - \widehat{\beta}_{IV}^{\prime} \overline{x}_i = \rho^{\prime} \omega_i + v_i \qquad i = 1, ..., n$$

  with $v_i = \alpha_i + \overline{\varepsilon}_i$.

- The parameters vector ρ can simply be estimated by **OLS**.

3.2 Anderson and Hsiao (1982) IV approach

Definition (parameters of time-invariant variables)

A consistent estimator of the parameters ρ is given by

$$\underset{(K_2,1)}{\widehat{\rho}} = \left(\sum_{i=1}^{n} \omega_i \omega_i' \right)^{-1} \left(\sum_{i=1}^{n} \omega_i h_i \right)$$

with $h_i = \overline{y}_i - \widehat{\gamma}_{IV} \overline{y}_{i,-1} - \widehat{\beta}_{IV}' \overline{x}_i$.

3.2 Anderson and Hsiao (1982) IV approach

The Anderson and Hsiao (1982) IV approach

1. **First step:** first difference transformation
2. **Second step:** choice of instruments and IV estimation of γ and β
3. **Third step:** estimation of ρ
4. **Fourth step:** estimation of the variances σ_{α}^2 and σ_{ε}^2

3.2 Anderson and Hsiao (1982) IV approach

Fourth step: estimation of the variances

Definition

Given $\widehat{\gamma}_{IV}$, $\widehat{\beta}_{IV}$, and $\widehat{\rho}$, we can estimate the **variances** as follows:

$$\widehat{\sigma}_{\varepsilon}^{2} = \frac{1}{n\left(T-1\right)} \sum_{t=2}^{T} \sum_{i=1}^{n} \widehat{\varepsilon}_{it}^{2}$$

$$\widehat{\sigma}_{\alpha}^{2} = \frac{1}{n} \sum_{i=1}^{n} \left(\overline{y}_{i} - \widehat{\gamma}_{IV} \overline{y}_{i,-1} - \widehat{\beta}_{IV}^{\prime} \overline{x}_{i} - \widehat{\rho}^{\prime} z_{i} \right)^{2} - \frac{1}{T} \widehat{\sigma}_{\varepsilon}^{2}$$

with

$$\widehat{\varepsilon}_{it} = \left(y_{i,t} - y_{i,t-1}\right) - \widehat{\gamma}_{IV} \left(y_{i,t-1} - y_{i,t-2}\right) - \widehat{\beta}_{IV}^{\prime} \left(x_{i,t} - x_{i,t-1}\right)$$

3.2 Anderson and Hsiao (1982) IV approach

Theorem

The instrumental-variable estimators of γ, β, and σ_ε^2 are consistent when n (correction of the **Nickell bias***), or T, or both tend to infinity.*

$$\underset{n,T\to\infty}{plim}\ \widehat{\gamma}_{IV} = \gamma \qquad \underset{n,T\to\infty}{plim}\ \widehat{\beta}_{IV} = \beta \qquad \underset{n,T\to\infty}{plim}\ \widehat{\sigma}_\varepsilon^2 = \sigma_\varepsilon^2$$

The estimators of ρ and σ_α^2 are consistent only when n goes to infinity.

$$\underset{n\to\infty}{plim}\ \widehat{\rho} = \rho \qquad \underset{n\to\infty}{plim}\ \widehat{\sigma}_\alpha^2 = \sigma_\alpha^2$$

3.2 Anderson and Hsiao (1982) IV approach

Key Concepts SubSection 3.2

1. Anderson and Hsiao (1982) IV approach.
2. The 4 steps of the estimation procedure.
3. First difference transformation of the dynamic panel model.
4. Tow choices of instrument.

Section 4

Generalized Method of Moment (GMM)

4. The GMM approach

Let us consider the same **dynamic panel data model** as in section 3:

$$y_{it} = \gamma y_{i,t-1} + \beta^{'} x_{it} + \rho^{'} \omega_i + \alpha_i + \varepsilon_{it}$$

- α_i are the (unobserved) individual effects,
- x_{it} is a vector of K_1 time-varying explanatory variables,
- ω_i is a vector of K_2 time-invariant variables.

4. The GMM approach

Assumptions: we assume that the component error term $v_{it} = \varepsilon_{it} + \alpha_i$

- $\mathbb{E}(\varepsilon_{it}) = 0$, $\mathbb{E}(\alpha_i) = 0$
- $\mathbb{E}(\varepsilon_{it}\varepsilon_{js}) = \sigma_{\varepsilon}^2$ if $j = i$ and $t = s$, 0 otherwise.
- $\mathbb{E}(\alpha_i\alpha_j) = \sigma_{\alpha}^2$ if $j = i$, 0 otherwise.
- $\mathbb{E}(\alpha_i x_{it}) = 0$, $\mathbb{E}(\alpha_i \omega_i) = 0$ (**exogeneity** assumption for ω_i)

4. The GMM approach

Definition (First difference model)

The GMM estimation method is based on a model in **first differences**, in order to swip out the individual effects α_i and th variables ω_i:

$$\left(y_{it}-y_{i,t-1}\right)=\gamma\left(y_{i,t-1}-y_{i,t-2}\right)+\beta^{'}\left(x_{it}-x_{i,t-1}\right)+\varepsilon_{it}-\varepsilon_{i,t-1}$$

for $t=2,..,T$.

4. The GMM approach

Intuition of the moment conditions

- Notice that $y_{i,t-2}$ and $(y_{i,t-2} - y_{i,t-3})$ are not the only valid instruments for $(y_{i,t-1} - y_{i,t-2})$.
- All the **lagged variables** $y_{i,t-2-j}$, for $j \geq 0$, satisfy

$$\mathbb{E}\left(y_{i,t-2-j}\left(\varepsilon_{i,t} - \varepsilon_{i,t-1}\right)\right) = 0 \quad \textbf{Exogeneity property}$$

$$\mathbb{E}\left(y_{i,t-2-j}\left(y_{i,t-1} - y_{i,t-2}\right)\right) \neq 0 \quad \textbf{Relevance property}$$

- Therefore, they all are legitimate **instruments** for $(y_{i,t-1} - y_{i,t-2})$.

4. The GMM approach

Intuition of the moment conditions

The $m+1$ conditions

$$\mathbb{E}\left(y_{i,t-2-j}\left(\varepsilon_{i,t}-\varepsilon_{i,t-1}\right)\right)=0 \quad \text{for} \quad j=0,1,..,m$$

can be used as **moment conditions** in order to estimate

$$\theta=\left(\beta,\gamma,\rho,\sigma_{\alpha}^{2},\sigma_{\varepsilon}^{2}\right)$$

Arellano, M., and S. Bond (1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

4. The GMM approach

Remark: The moment conditions

$$\mathbb{E}\left(y_{i,t-2-j}\left(\varepsilon_{i,t}-\varepsilon_{i,t-1}\right)\right)=0 \quad \text{for} \quad j=0,1,..,m$$

mean that there exists a vector of parameters (true value)

$$\theta_0=\left(\beta_0^{'},\gamma_0,\rho_0^{'},\sigma_{\alpha 0}^2,\sigma_{\varepsilon 0}^2\right)^{'}$$

such that

$$\mathbb{E}\left(y_{i,t-2-j}\times\left(\Delta y_{it}-\gamma_0\Delta y_{i,t-1}-\beta_0^{'}\Delta x_{it}\right)\right)=0$$

where $\Delta=(1-L)$ and L denotes the lag operator .

4. The GMM approach

We consider two alternative assumptions on the explanatory variables x_{it}

1. **The explanatory variables x_{it} are strictly exogeneous.**
2. **The explanatory variables x_{it} are pre-determined.**

4. The GMM approach

We consider two alternative assumptions on the explanatory variables x_{it}

1. **The explanatory variables x_{it} are strictly exogeneous.**
2. **The explanatory variables x_{it} are pre-determined.**

4. The GMM approach

Assumption: exogeneity

We assume that the time varying explanatory variables x_{it} are **strictly exogeneous** in the sense that:

$$\mathbb{E}\left(x_{it}^{'}\varepsilon_{is}\right) = 0 \quad \forall\left(t, s\right)$$

4. The GMM approach

Definition (moment conditions)

For each period, we have the following orthogonal conditions

$$\mathbb{E}\left(q_{it}\Delta\varepsilon_{it}\right) = 0, \quad t = 2, .., T$$

$$\underset{(t-1+TK_1,1)}{q_{it}} = \left(y_{i0}, y_{i1}, .., y_{i,t-2}, x_i^{'}\right)^{'}$$

where $x_i^{'} = \left(x_{i1}^{'}, .., x_{iT}^{'}\right)$, $\Delta = (1 - L)$ and L denotes the lag operator

4. The GMM approach

Example (moment conditions)

The condition $\mathbb{E}\left(q_{it}\Delta\varepsilon_{it}\right) = 0$, $q_{it} = \left(y_{i0}, y_{i1}, .., y_{i,t-2}, x_i'\right)'$ at time $t = 2$ becomes

$$\mathbb{E}\left(\underset{(1+TK_1,1)}{q_{i2}}\;\underset{(1,1)}{\Delta\varepsilon_{i2}}\right) = \mathbb{E}\left(\begin{pmatrix} y_{i0} \\ x_i' \end{pmatrix}\left(\varepsilon_{i2} - \varepsilon_{i1}\right)\right) = \underset{(1+TK_1,1)}{0}$$

where $x_i' = \left(x_{i1}', .., x_{iT}'\right)$. At time $t = 3$, we have

$$\mathbb{E}\left(\underset{(2+TK_1,1)}{q_{i3}}\;\underset{(1,1)}{\Delta\varepsilon_{i3}}\right) = \mathbb{E}\left(\begin{pmatrix} y_{i0} \\ y_{i1} \\ x_i' \end{pmatrix}\left(\varepsilon_{i3} - \varepsilon_{i2}\right)\right) = \underset{(2+TK_1,1)}{0}$$

4. The GMM approach

Under the exogeneity assumption, what is the number of moment conditions?

$$\mathbb{E}\left(q_{it}\Delta\varepsilon_{it}\right)=0, \quad t=2,..,T$$

Time	Number of moment conditions
$t=2$	$1+TK_1$
$t=3$	$2+TK_1$
...	...
$t=T$	$T-1+TK_1$
total	$T\left(T-1\right)\left(K_1+1/2\right)$

Proof: the total number of moment conditions is equal to

$$
\begin{aligned}
r &= 1 + TK_1 + 2 + TK_1 .. + TK_1 + (T-1) \\
&= T(T-1)K_1 + 1 + 2 + .. + (T-1) \\
&= T(T-1)K_1 + \frac{T(T-1)}{2} \\
&= T(T-1)\left(K_1 + \frac{1}{2}\right)
\end{aligned}
$$

4. The GMM approach

Stacking the $T-1$ first-differenced equations in matrix form, we have

$$\underset{(T-1,1)}{\Delta y_i} = \underset{(T-1,1)}{\Delta y_{i,-1}}\underset{(1,1)}{\gamma} + \underset{(T-1,K_1)}{\Delta X_i}\underset{(K_1,1)}{\beta} + \underset{(T-1,1)}{\Delta \varepsilon_i} \quad i = 1, .., N$$

where :

$$\underset{(T-1,1)}{\Delta y_i} = \begin{pmatrix} y_{i2} - y_{i1} \\ y_{i3} - y_{i2} \\ .. \\ y_{iT} - y_{i,T-1} \end{pmatrix} \quad \underset{(T-1,1)}{\Delta y_{i,-1}} = \begin{pmatrix} y_{i1} - y_{i0} \\ y_{i2} - y_{i1} \\ .. \\ y_{iT-1} - y_{i,T-2} \end{pmatrix}$$

4. The GMM approach

Stacking the $T-1$ first-differenced equations in matrix form, we have

$$\underset{(T-1,1)}{\Delta y_i} = \underset{(T-1,1)}{\Delta y_{i,-1}} \underset{(1,1)}{\gamma} + \underset{(T-1,K_1)}{\Delta X_i} \underset{(K_1,1)}{\beta} + \underset{(T-1,1)}{\Delta \varepsilon_i} \quad i = 1,..,N$$

where :

$$\underset{(T-1,K_1)}{\Delta X_i} = \begin{pmatrix} x_{i2} - x_{i1} \\ x_{i3} - x_{i2} \\ .. \\ x_{iT} - x_{i,T-1} \end{pmatrix} \quad \underset{(T-1,1)}{\Delta \varepsilon_i} = \begin{pmatrix} \varepsilon_{i2} - \varepsilon_{i1} \\ \varepsilon_{i3} - \varepsilon_{i2} \\ .. \\ \varepsilon_{iT} - \varepsilon_{i,T-1} \end{pmatrix}$$

4. The GMM approach

Definition (moment conditions)

The conditions $\mathbb{E}\left(q_{it}\Delta\varepsilon_{it}\right) = 0$ for $t = 2, .., T$, can be written as

$$\mathbb{E}\left(\underset{(r,T-1)}{W_i}\underset{(T-1,1)}{\Delta\varepsilon_i}\right) = \underset{(m,1)}{0}$$

$$W_i = \begin{pmatrix} \underset{(1+TK_1,1)}{q_{i2}} & 0 & ... & 0 \\ 0 & \underset{(2+TK_1,1)}{q_{i3}} & & \\ & & .. & \\ 0 & & .. & \underset{(T-1+TK_1,1)}{q_{iT}} \end{pmatrix}$$

where $r = T\left(T-1\right)\left(K_1 + 1/2\right)$ is the number of moment conditions.

4. The GMM approach

Example (moment conditions, vectorial form)

At time $t = 2$, we have

$$\mathbb{E}\left(q_{i2}\Delta\varepsilon_{i2}\right) = \mathbb{E}\left(\left(\begin{array}{c} y_{i0} \\ x_i' \end{array}\right)\left(\varepsilon_{i2} - \varepsilon_{i1}\right)\right) = 0$$

In a vectorial form we have the first set of $1 + TK_1$ moment conditions

$$\mathbb{E}\left(W_i\Delta\varepsilon_i\right) = \mathbb{E}\left(\left(\begin{array}{cccc} \underset{(1+TK_1,1)}{q_{i2}} & 0 & \ldots & 0 \end{array}\right)\left(\begin{array}{c} \varepsilon_{i2} - \varepsilon_{i1} \\ \varepsilon_{i3} - \varepsilon_{i2} \\ .. \\ \varepsilon_{iT} - \varepsilon_{i,T-1} \end{array}\right)\right) = 0$$

4. The GMM approach

Example (moment conditions, vectorial form)

At time $t = 3$, we have

$$\mathbb{E}\left(q_{i3}\Delta\varepsilon_{i3}\right) = \mathbb{E}\left(\begin{pmatrix} y_{i0} \\ y_{i1} \\ x_i' \end{pmatrix}\left(\varepsilon_{i3} - \varepsilon_{i2}\right)\right) = 0$$

In a vectorial form we have the second set of $2 + TK_1$ moment conditions

$$\mathbb{E}\left(W_i\Delta\varepsilon_i\right) = \mathbb{E}\left(\begin{pmatrix} 0 & \underset{(2+TK_1,1)}{q_{i3}} & \dots & 0 \end{pmatrix}\begin{pmatrix} \varepsilon_{i2} - \varepsilon_{i1} \\ \varepsilon_{i3} - \varepsilon_{i2} \\ .. \\ \varepsilon_{iT} - \varepsilon_{i,T-1} \end{pmatrix}\right) = 0$$

4. The GMM approach

Example

For $T = 10$ et $K_1 = 0$ (without explicative variable), we have

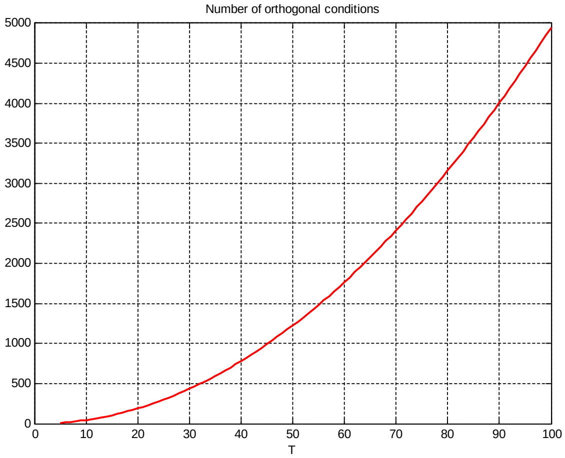
$$r = \frac{T(T-1)}{2} = 45 \text{ orthogonal conditions}$$

Example

For $T = 50$ et $K_1 = 0$ (without explicative variable), we have

$$r = \frac{T(T-1)}{2} = 1225 \text{ orthogonal conditions !!}$$

4. The GMM approach



4. The GMM approach

We consider two alternative assumptions on the explanatory variables x_{it}

1. **The explanatory variables x_{it} are strictly exogeneous.**
2. **The explanatory variables x_{it} are pre-determined.**

4. The GMM approach

We consider two alternative assumptions on the explanatory variables x_{it}

1. **The explanatory variables x_{it} are strictly exogeneous.**
2. **The explanatory variables x_{it} are pre-determined.**

4. The GMM approach

Assumption: pre-determination

We assume that the time varying explanatory variables x_{it} are **pre-determined** in the sense that:

$$\mathbb{E}\left(x_{it}'\varepsilon_{is}\right) = 0 \text{ if } t \leq s$$

4. The GMM approach

In this case, we have

$$\mathbb{E}\left(q_{it}\Delta\varepsilon_{it}\right) = 0, \quad t = 2, .., T$$

$$\underset{(t-1+tK_1,1)}{q_{it}} = \left(y_{i0}, y_{i1}, .., y_{i,t-2}, x'_{i1}, .., \underbrace{x'_{i,t-2}}_{\text{not } T}\right)'$$

4. The GMM approach

Definition

The conditions $\mathbb{E}\left(q_{it}\Delta\varepsilon_{it}\right) = 0$ for $t = 2, .., T$, can be written as

$$\mathbb{E}\left(\underset{(r,T-1)}{W_i}\underset{(T-1,1)}{\Delta\varepsilon_i}\right) = \underset{(m,1)}{0}$$

$$W_i = \begin{pmatrix} \underset{(1+K_1,1)}{q_{i2}} & 0 & ... & 0 \\ 0 & \underset{(2+2K_1,1)}{q_{i3}} & & \\ & & .. & \\ 0 & & .. & \underset{(T-1+(T-1)K_1,1)}{q_{iT}} \end{pmatrix}$$

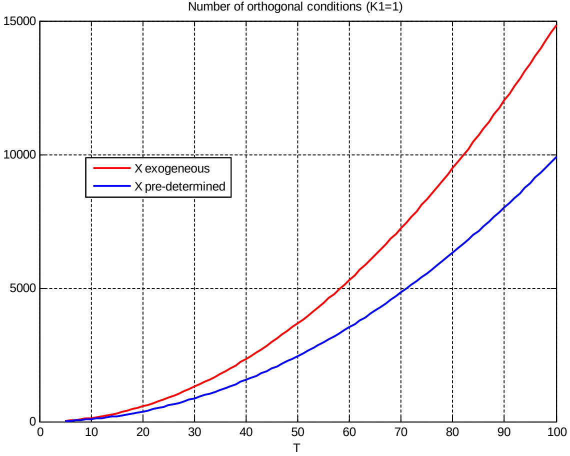
where $r = T\left(T-1\right)\left(K_1+1\right)/2$ is the number of moment conditions.

4. The GMM approach

Proof: the total number of moment conditions is equal to

$$\begin{aligned} r &= 1 + K_1 + 2 + K_1 .. + (T-1) K_1 + (T-1) \\ &= (1 + K1)(1 + 2 + ... + (T-1)) \\ &= (1 + K_1) \frac{T(T-1)}{2} \end{aligned}$$

4. The GMM approach



4. The GMM approach

Fact

Whatever the assumption made on the explanatory variable, the number of othogonal conditions (moments) r is much larger than the number of parameters, e.g. $K_1 + 1$. Thus, Arellano and Bond (1991) suggest a generalized method of moments (GMM) estimator.

Arellano, M., and S. Bond (1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, 58, 277–297.

4. The GMM approach

We will exploit the moment conditions

$$\mathbb{E}\left(W_i \Delta \varepsilon_i\right) = 0$$

to estimate by GMM the parameters $\theta = \left(\gamma, \beta'\right)'$ in

$$\Delta y_i = \Delta y_{i,-1}\gamma + \Delta X_i \beta + \Delta \varepsilon_i \quad i = 1, .., n$$

Subsection 4.1

GMM: a general presentation

4.1 GMM: a general presentation

Definition

The standard method of moments estimator consists of solving the unknown parameter vector θ by equating the theoretical moments with their empirical counterparts or estimates.

4.1 GMM: a general presentation

1. Suppose that there exist relations $m(y_t;\theta)$ such that

$$\mathbb{E}\left(m\left(y_t;\theta_0\right)\right)=0$$

   where θ_0 is the true value of θ and $m(y_t;\theta_0)$ is a $r\times 1$ vector.

2. Let $\widehat{m}(y,\theta)$ be the sample estimates of $\mathbb{E}(m(y_t;\theta))$ based on n independent samples of y_t

$$\widehat{m}(y,\theta)=\frac{1}{n}\sum_{t=1}^{n}m\left(y_t;\theta\right)$$

3. Then the method of moments consit in finding $\widehat{\theta}$, such that

$$\widehat{m}\left(y,\widehat{\theta}\right)=0$$

4.1 GMM: a general presentation

Intuition of the GMM

Consider the moment conditions such that

$$\mathbb{E}\left(m\left(y_t;\theta_0\right)\right)=0$$

Under some regularity assumptions, $\forall\theta\in\Theta$

$$\widehat{m}\left(y,\theta\right)=\frac{1}{n}\sum_{t=1}^{n}m\left(y_t;\theta\right)\xrightarrow{p}\mathbb{E}\left(m\left(y_t;\theta\right)\right)$$

In particular

$$\widehat{m}\left(y,\theta_0\right)\xrightarrow{p}\mathbb{E}\left(m\left(y_t;\theta_0\right)\right)=0$$

So, the GMM consists in finding $\widehat{\theta}$ such that

$$\widehat{m}\left(y,\widehat{\theta}\right)=0\quad\Longrightarrow\quad\widehat{\theta}\xrightarrow{p}\theta_0$$

4.1 GMM: a general presentation

Fact (just identified system)

If the number r of equations (moments) is equal to the dimension a of θ, it is in general possible to solve for $\widehat{\theta}$ uniquely. The system is **just identified**.

4.1 GMM: a general presentation

Example (classical method of moment)

Consider a random variable $y_t \sim t\left(v\right)$. We want to estimate v from an i.i.d. sample $\{y_1, ..y_n\}$. We know that:

$$\mu_2 = \mathbb{E}\left(y_t^2\right) = \mathbb{V}\left(y_t\right) = \frac{v}{v-2}$$

If μ_2 is known, then v can be identified as:

$$v = \frac{2\mathbb{E}\left(y_t^2\right)}{\mathbb{E}\left(y_t^2\right) - 1}$$

4.1 GMM: a general presentation

Example (classical method of moment)

Now let us consider the sample variance $\widehat{\mu}_{2,T}$

$$\widehat{\mu}_2 = \frac{1}{n}\sum_{t=1}^{n} y_t^2 \xrightarrow{p} \mu_2$$

So, a consistent estimate (classical method of moment) of v is defined by:

$$\widehat{v} = \frac{2\widehat{\mu}_2}{\widehat{\mu}_2 - 1}$$

4.1 GMM: a general presentation

Example (classical method of moment)

Another way to write the problem consists in defining

$$m(y_t; v) = y_t^2 - \frac{v}{v-2}$$

By definition, we have:

$$\mathbb{E}(m(y_t; v)) = \mathbb{E}\left(y_t^2 - \frac{v}{v-2}\right) = 0$$

4.1 GMM: a general presentation

Example (classical method of moment)

The moment condition ($r = 1$) is

$$\mathbb{E}\left(m\left(y_t; v\right)\right) = \mathbb{E}\left(y_t^2 - \frac{v}{v-2}\right) = 0$$

The empirical counterpart is

$$\widehat{m}\left(y; v\right) = \frac{1}{n}\sum_{t=1}^{n} m\left(y_t; v\right) = \frac{1}{n}\sum_{i=1}^{n}\left(y_t^2 - \frac{v}{v-2}\right)$$

So, the estimator $\widehat{v}$ of the classical method of moment is defined by:

$$\widehat{m}\left(y; \widehat{v}\right) = 0 \quad \Leftrightarrow \quad \widehat{v} = \frac{2\widehat{\mu}_2}{\widehat{\mu}_2 - 1} \xrightarrow{p} v = \frac{2\mathbb{E}\left(y_t^2\right)}{\mathbb{E}\left(y_t^2\right) - 1}$$

4.1 GMM: a general presentation

Definition (over-identified system)

If the number of moments r is greater than the dimension of θ, the system of non linear equation $\widehat{m}\left(y;\widehat{v}\right)=0$, in general, has no solution. The system is **over-identified**.

4.1 GMM: a general presentation

If the system is over-identified, it is then necessary to minimize some norm (or **distance measure**) of $\widehat{m}(y;\theta)$ in order to find $\widehat{\theta}$:

$$q(y,\theta) = \widehat{m}(y;\theta)' S^{-1} \widehat{m}(y;\theta)$$

where S^{-1} is a positive definite (weighting) matrix.

4.1 GMM: a general presentation

Example (weigthing matrix)

Consider a random variable $y_t \sim t(v)$. We want to estimate v from an i.i.d. sample $\{y_1, ..y_n\}$. We know that:

$$\mu_2 = \mathbb{E}\left(y_t^2\right) = \frac{v}{v-2}$$

$$\mu_4 = \mathbb{E}\left(y_t^4\right) = \frac{3v^2}{(v-2)(v-4)}$$

The two moment conditions $(r = 2)$ can be written as

$$\mathbb{E}\left(m\left(y_t; v\right)\right) = \mathbb{E}\left(\begin{array}{c} y_t^2 - \frac{v}{v-2} \\ y_t^4 - \frac{3v^2}{(v-2)(v-4)} \end{array}\right) = \left(\begin{array}{c} 0 \\ 0 \end{array}\right)$$

4.1 GMM: a general presentation

Example (weigthing matrix)

It is impossible to find a unique value $\widehat{v}$ such that

$$\widehat{m}\left(y;\widehat{v}\right)=\frac{1}{n}\sum_{t=1}^{n}m\left(y_t;\widehat{v}\right)=\left(\begin{array}{c}\frac{1}{n}\sum_{t=1}^{n}y_t^2-\frac{\widehat{v}}{\widehat{v}-2}\\ \frac{1}{n}\sum_{t=1}^{n}y_t^2-\frac{3\widehat{v}^2}{(\widehat{v}-2)(\widehat{v}-4)}\end{array}\right)=\left(\begin{array}{c}0\\0\end{array}\right)$$

So, we have to impose some weights to the two moment conditions

$$\widehat{m}\left(y;v\right)'S^{-1}\ \widehat{m}\left(y;v\right)$$

4.1 GMM: a general presentation

Example (weigthing matrix)

Let us assume that

$$S^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

then we have

$$\begin{aligned} \widehat{m}(y; v)' S^{-1} \; \widehat{m}(y; v) \;\; = \;\; & \left(\frac{1}{n} \sum_{t=1}^{n} y_t^2 - \frac{v}{v-2} \right)^2 \\ & + 2 \left(\frac{1}{n} \sum_{t=1}^{n} y_t^2 - \frac{3v^2}{(v-2)(v-4)} \right)^2 \end{aligned}$$

It is now possible to find $\widehat{v}$ such that $\widehat{m}(y; v)' S^{-1} \; \widehat{m}(y; v) = 0$

4.1 GMM: a general presentation

Definition (GMM estimator)

The GMM estimator $\widehat{\theta}$ minimizes the following criteria

$$\widehat{\theta} = \underset{\theta \in \mathbb{R}^a}{\arg\min}\ \underset{(1,1)}{q(y,\theta)} = \underset{\theta \in \mathbb{R}^a}{\arg\min}\ \underset{(1,r)}{\widehat{m}(y;\theta)'}\underset{(r,r)}{S^{-1}}\underset{(r,1)}{\widehat{m}(y;\theta)}$$

where S^{-1} is the optimal weighting matrix.

4.1 GMM: a general presentation

What is the optimal weigthing matrix?

$$\widehat{\theta} = \underset{\theta \in \mathbb{R}^a}{\arg\min} \underset{(1,1)}{q(y,\theta)} = \underset{\theta \in \mathbb{R}^a}{\arg\min} \underset{(1,r)}{\widehat{m}(y;\theta)'} \underset{(r,r)}{S^{-1}} \underset{(r,1)}{\widehat{m}(y;\theta)}$$

The optimal choice (if there is **no autocorrelation** of $m(y;\theta_0)$) of S turns out to be

$$\underset{(r,r)}{S} = \mathbb{E}\left(\underset{(r,1)}{m(y;\theta_0)} \underset{(1,r)}{m(y;\theta_0)'} \right)$$

The matrix S corresponds to **variance-covariance matrix** of the vector $m(y;\theta_0)$.

4.1 GMM: a general presentation

Definition (Optimal weighting matrix)

In the general case, the optimal weighting matrix is the inverse of the long-run variance covariance matrix of $m(y_t; \theta_0)$.

$$\underset{(r,r)}{S} = \sum_{j=-\infty}^{\infty} \mathbb{E}\left(\underset{(r,1)}{m(y_t; \theta_0)} \; \underset{(1,r)}{m(y_{t-j}; \theta_0)'} \right)$$

Remark

The optimal weighting matrix is

$$S = \sum_{j=-\infty}^{\infty} \mathbb{E}\left(m\left(y_t;\theta_0\right) m\left(y_{t-j};\theta_0\right)'\right)$$

We can replace the unknow value θ_0 by the GMM estimator $\hat{\theta}$ and the optimal weighting matrix becomes

$$S = \sum_{j=-\infty}^{\infty} \mathbb{E}\left(m\left(y_t;\widehat{\theta}\right) m\left(y_{t-j};\widehat{\theta}\right)'\right)$$

4.1 GMM: a general presentation

Problem 1 How to estimate S?

$$S = \sum_{j=-\infty}^{\infty} \mathbb{E}\left(m\left(y_t; \widehat{\theta}\right) m\left(y_{t-j}; \widehat{\theta}\right)' \right)$$

A first solution (too) simple solution consits in using the empirical counterparts of variance and covariances

$$\widehat{S} = \sum_{j=-(n-2)}^{n-2} \widehat{\Gamma}_j$$

$$\widehat{\Gamma}_j = \frac{1}{n} \sum_{t=j+2}^{n} m\left(y_t; \widehat{\theta}\right) m\left(y_{t-j}; \widehat{\theta}\right)'$$

But, this estimator may be no positive definite...

4.1 GMM: a general presentation

Solution (Non-parametric kernel estimators)

A positive definite kernel estimator for S has been proposed by Newey and West (1987) and is defined as

$$\widehat{S} = \widehat{\Gamma}_0 + \sum_{j=1}^{q} \left(1 - \frac{j}{q+1}\right) \left(\widehat{\Gamma}_j + \widehat{\Gamma}'_j\right)$$

$$\widehat{\Gamma}_j = \frac{1}{n} \sum_{t=j+2}^{n} m\left(y_t; \widehat{\theta}\right) m\left(y_{t-j}; \widehat{\theta}\right)'$$

where q is a bandwidth parameter and $K(j) = 1 - j/(q+1)$ a Bartlett kernel function.

4.1 GMM: a general presentation

Example (Newey and West kernel estimator)

$$\widehat{S} = \widehat{\Gamma}_0 + \sum_{j=1}^{q} \left(1 - \frac{j}{q+1}\right) \left(\widehat{\Gamma}_j + \widehat{\Gamma}'_j\right)$$

If $q = 2$ then we have

$$\widehat{S} = \widehat{\Gamma}_0 + \frac{2}{3} \left(\widehat{\Gamma}_1 + \widehat{\Gamma}'_1\right) + \frac{1}{3} \left(\widehat{\Gamma}_2 + \widehat{\Gamma}'_2\right)$$

4.1 GMM: a general presentation

Other estimators => other kernel functions

$$\widehat{S} = \widehat{\Gamma}_0 + \sum_{j=1}^{q} K\left(\frac{j}{q+1}\right)\left(\widehat{\Gamma}_j + \widehat{\Gamma}_j'\right)$$

1. **Gallant (1987):** Parzen kernel

$$K(u) = \begin{cases} 1 - 6|u|^2 + 6|u|^3 & \text{if } 0 \leq |u| \leq 1/2 \\ 2(1-|u|)^3 & \text{if } 1/2 \leq |u| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

2. **Andrews (1991)**: quadratic spectral kernel

$$K(u) = \frac{3}{(6\pi u/5)^2}\left(\frac{\sin(6\pi u/5)}{(6\pi u/5)} - \cos(6\pi u/5)\right)$$

4.1 GMM: a general presentation

Problem 2

$$\widehat{\theta} = \underset{\theta \in \mathbb{R}^a}{\arg\min}\ \widehat{m}(y;\theta)' S^{-1} \widehat{m}(y;\theta)$$

$$S = \sum_{j=-\infty}^{\infty} \mathbb{E}\left(m\left(y_t;\widehat{\theta}\right) m\left(y_{t-j};\widehat{\theta}\right)' \right)$$

1. In order to compute $\widehat{\theta}$, we have to know S^{-1}.
2. In order to compute S, we have to know $\widehat{\theta}$... **a circularity issue**

Solutions

1. **Two-step GMM:** Hansen (1982)
2. **Iterative GMM:** Ferson and Foerster (1994)
3. **Continuous-updating GMM:** Hansen, Heaton and Yaron (1996), Stock and Wright (2000), Newey and Smith (2003), Ma (2002).

Solutions

1. **Two-step GMM:** Hansen (1982)
2. **Iterative GMM:** Ferson and Foerster (1994)
3. **Continuous-updating GMM:** Hansen, Heaton and Yaron (1996), Stock and Wright (2000), Newey and Smith (2003), Ma (2002).

Two-step GMM

Step 1: put the same weight to the r moment conditions by using an identity weighting matrix

$$S_0 = I_r$$

Compute a first consistent (but not efficient) estimator $\widehat{\theta}_0$

$$\begin{aligned}\widehat{\theta}_0 &= \arg\min_{\theta \in \mathbb{R}^a} \widehat{m}(y;\theta)' S_0^{-1} \widehat{m}(y;\theta) \\ &= \arg\min_{\theta \in \mathbb{R}^a} \widehat{m}(y;\theta)' \widehat{m}(y;\theta)\end{aligned}$$

4.1 GMM: a general presentation

Two-step GMM

Step 2: Compute the estimator for the optimal weighting matrix $\widehat{S}_1$

$$\widehat{S}_1 = \widehat{\Gamma}_0 + \sum_{j=1}^{q} K\left(\frac{j}{q+1}\right)\left(\widehat{\Gamma}_j + \widehat{\Gamma}_j'\right)$$

$$\widehat{\Gamma}_j = \frac{1}{n}\sum_{t=j+2}^{n} m\left(y_t; \widehat{\theta}_0\right) m\left(y_{t-j}; \widehat{\theta}_0\right)'$$

Finally, compute the efficient GMM estimator $\widehat{\theta}_1$ as

$$\widehat{\theta}_1 = \underset{\theta \in \mathbb{R}^a}{\arg\min}\ \widehat{m}\left(y;\theta\right)' \widehat{S}_1^{-1} \widehat{m}\left(y;\theta\right)$$

Subsection 4.2

Application to dynamic panel data models

4.2 Application to dynamic panel data models

Various GMM estimators (i.e. moment conditions) have been proposed for dynamic panel data models

1. **Arellano and Bond (1991): GMM estimator**
2. **Arellano and Bover (1995): GMM estimator**
3. **Ahn and Schmidt (1995): GMM estimator**
4. **Blundell and Bond (2000): a system GMM estimator**

4.2 Application to dynamic panel data models

Various GMM estimators (i.e. moment conditions) have been proposed for dynamic panel data models

1. **Arellano and Bond (1991): GMM estimator**
2. **Arellano and Bover (1995): GMM estimator**
3. **Ahn and Schmidt (1995): GMM estimator**
4. **Blundell and Bond (2000): a system GMM estimator**

4.2 Application to dynamic panel data models

Problem

- Let us consider the dynamic panel data model in first differences

$$\Delta y_i = \Delta y_{i,-1}\gamma + \Delta X_i \beta + \Delta \varepsilon_i \quad i = 1, .., n$$

- We want to estimate the $K_1 + 1$ parameters $\theta = \left(\gamma, \beta'\right)'$.
- For that, we have $r = T\left(T-1\right)\left(K_1 + 1/2\right)$ moment conditions (if x_{it} are strictly exogeneous)

$$\mathbb{E}\left(W_i \Delta \varepsilon_i\right) = \mathbb{E}\left(W_i \times \left(\Delta y_i - \Delta y_{i,-1}\gamma - \Delta X_i \beta\right)\right) = 0_r$$

4.2 Application to dynamic panel data models

Let us denote

$$m\left(y_i, x_i; \theta\right) = W_i \times \left(\Delta y_i - \Delta y_{i,-1}\gamma - \Delta X_i \beta\right)$$

with

$$\mathbb{E}\left(m\left(y_i, x_i; \theta\right)\right) = 0_r$$

4.2 Application to dynamic panel data models

Definition (Arellano and Bond (1991) GMM estimator)

The Arellano and Bond GMM estimator of $\theta = (\gamma, \beta')'$ is

$$\widehat{\theta} = \underset{\theta \in \mathbb{R}^{K_1+1}}{\arg\min} \left(\frac{1}{n} \sum_{i=1}^{n} m(y_i, x_i; \theta) \right)' S^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} m(y_i, x_i; \theta) \right)$$

or equivalently

$$\widehat{\theta} = \underset{\theta \in \mathbb{R}^{K_1+1}}{\arg\min} \left(\frac{1}{n} \sum_{i=1}^{n} \Delta\varepsilon_i' W_i' \right) S^{-1} \left(\frac{1}{n} \sum_{i=1}^{n} W_i \Delta\varepsilon_i \right)$$

with $S = \mathbb{E}\left(m(y; \theta_0)\ \ m(y; \theta_0)\right)'$.

4.2 Application to dynamic panel data models

Under the assumption of non-autocorrelation, the optimal weighting matrix can be expressed as

$$S = \mathbb{E}\left(\frac{1}{n^2}\sum_{i=1}^{n} W_i \Delta\varepsilon_i \Delta\varepsilon_i' W_i'\right)$$

In the general case, S is the long-run variance covariance matrix of $n^{-2}\sum_{i=1}^{n} W_i \Delta\varepsilon_i \Delta\varepsilon_i' W_i'$.

4.2 Application to dynamic panel data models

Various GMM estimators (i.e. moment conditions) have been proposed for dynamic panel data models

1. **Arellano and Bond (1991): GMM estimator**
2. **Arellano and Bover (1995): GMM estimator**
3. **Ahn and Schmidt (1995): GMM estimator**
4. **Blundell and Bond (2000): a system GMM estimator**

4.2 Application to dynamic panel data models

In addition to the previous moment conditions, Arellano and Bover (1995) also note that $\mathbb{E}\left(\overline{v}_i\right) = \mathbb{E}\left(\overline{\varepsilon}_i + \alpha_i\right) = 0$, where

$$\overline{v}_i = \overline{y}_i - \gamma \overline{y}_{i,-1} - \beta' \overline{x}_i - \rho' \omega_i$$

Therefore, if instruments $\widetilde{q}_i$ exist (for instance, the constant 1 is a valid instrument) such that

$$\mathbb{E}\left(\widetilde{q}_i \overline{v}_i\right) = 0$$

then a more efficient GMM estimator can be derived by incorporating this additional moment condition.

Arellano, M., and O. Bover (1995). "Another Look at the Instrumental Variable Estimation of Error-Components Models," Journal of Econometrics, 68, 29–51.

4.2 Application to dynamic panel data models

Definition

Arellano and Bond (1991) consider the following moment conditions

$$\mathbb{E}\left(m\left(y_i, x_i; \theta\right)\right) = \mathbb{E}\left(W_i\left(\Delta y_i - \Delta y_{i,-1}\gamma - \Delta X_i \beta\right)\right) = 0$$

Definition

Arellano and Bover (1995) consider additional moment conditions

$$\mathbb{E}\left(m\left(y_i, x_i; \theta\right)\right) = \mathbb{E}\left(\widetilde{q}_i\left(\overline{y}_i - \gamma \overline{y}_{i,-1} - \beta' \overline{x}_i - \rho' \omega_i\right)\right) = 0$$

4.2 Application to dynamic panel data models

Various GMM estimators (i.e. moment conditions) have been proposed for dynamic panel data models

1. **Arellano and Bond (1991): GMM estimator**
2. **Arellano and Bover (1995): GMM estimator**
3. **Ahn and Schmidt (1995): GMM estimator**
4. **Blundell and Bond (2000): a system GMM estimator**

4.2 Application to dynamic panel data models

Apart from the previous linear moment conditions, Ahn and Schmidt (1995) note that the homoscedasticity condition on $\mathbb{E}\left(\varepsilon_{it}^{2}\right)$ implies the following $T-2$ linear conditions

$$\mathbb{E}\left(y_{it}\Delta\varepsilon_{i,t+1}-y_{i,t+1}\Delta\varepsilon_{i,t+1}\right)=0 \quad t=1,..,T-2$$

Combining these restrictions to the previous ones leads to a more efficient GMM estimator.

Ahn, S.C., and P. Schmidt (1995). "Efficient Estimation of Models for Dynamic Panel Data," *Journal of Econometrics*, 68, 5–27.

4.2 Application to dynamic panel data models

Various GMM estimators (i.e. moment conditions) have been proposed for dynamic panel data models

1. **Arellano and Bond (1991): GMM estimator**
2. **Arellano and Bover (1995): GMM estimator**
3. **Ahn and Schmidt (1995): GMM estimator**
4. **Blundell and Bond (2000): a system GMM estimator**

4.2 Application to dynamic panel data models

Definition (system GMM)

The system GMM (Blundell and Bond) was invented to tackle the weak instrument problem. It consists in considering both the equation in level and in first differences

$$\mathbb{E}\left(y_{it,-s}\Delta\varepsilon_{it}\right)=0 \quad \mathbb{E}\left(x_{i,t-s}\Delta\varepsilon_{it}\right)=0 \quad \text{Difference equation}$$

Following additional moments are explored:

$$\mathbb{E}\left(\Delta y_{it,-s}\left(\alpha_i^*+\varepsilon_{it}\right)\right)=0 \quad \mathbb{E}\left(\Delta x_{i,t-s}\left(\alpha_i^*+\varepsilon_{it}\right)\right)=0 \quad \text{Level equation}$$

Blundell and Bond, S. (2000): GMM Estimation with persistent panel data: an application to production functions. *Econometric Reviews*,19(3), 321-340.

4.2 Application to dynamic panel data models

Remarks

1. While theoretically it is possible to add additional moment conditions to improve the asymptotic efficiency of GMM, it is doubtful how much efficiency gain one can achieve by using a huge number of moment conditions in a finite sample.

2. Moreover, if higher-moment conditions are used, the estimator can be very sensitive to outlying observations.

Remarks

1. Through a simulation study, Ziliak (1997) has found that the downward bias in GMM is quite severe as the number of moment conditions expands, outweighing the gains in efficiency.

2. The strategy of exploiting all the moment conditions for estimation is actually not recommended for panel data applications. For further discussions, see Judson and Owen (1999), Kiviet (1995), and Wansbeek and Bekker (1996).

End of Chapter 2

Christophe Hurlin (University of Orléans)