Chapter 2. Dynamic panel data models

Master of Science in Economics - University of Geneva

Christophe Hurlin, Université d'Orléans

Université d'Orléans

April 2010

Introduction

Definition

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

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

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 $E(\varepsilon_{it}) = 0$, and $E(\varepsilon_{it}\varepsilon_{js}) = \sigma_\varepsilon^2$ if $j = i$ and $t = s$, and $E(\varepsilon_{it}\varepsilon_{js}) = 0$ otherwise.

Introduction

Fact

It turns out that in this circumstance 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**.

Introduction

1. If lagged dependent variables also 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.

Introduction

1. The consistency property of the MLE and the generalized leastsquares estimator (GLS) also depends on this assumption and on the way in which the number of time-series observations (T) and the number of crosssectional units (N) tend to infinity.

The dynamic panel bias

Section 1. The dynamic panel bias

The dynamic panel bias

1. The LSDV (or CV) estimator is consistent for the static model whether the effects are fixed or random.
2. In this section we show that the LSDV (or CV) is inconsistent for a dynamic panel data model with individual effects, whether the effects are fixed or random.

The dynamic panel bias

Definition

The biais of the LSDV estimator in a dynamix 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.

The dynamic panel bias

Definition

Consider the simple dynamic 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 $E(\alpha_i) = 0$ in the case of random individual effects).

The dynamic panel bias

We also assume that:

1. The autoregressive parameter γ satisfies

$$|\gamma| < 1$$

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

The dynamic panel bias

In this model, the LSDV estimator is defined by

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

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

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

The dynamic panel bias

Definition

The bias of the LSDV estimator is given by:

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

The dynamic panel bias

Let us consider the numerator. Because $\varepsilon_{it}$are uncorrelated with α_i^* are independently and identically distributed, we have

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

$$= \underset{N\to\infty}{plim}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}y_{i,t-1}\varepsilon_{it}-\underset{N\to\infty}{plim}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}y_{i,t-1}\overline{\varepsilon}_i$$

$$-\underset{N\to\infty}{plim}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\overline{y}_{i,-1}\varepsilon_{it}+\underset{N\to\infty}{plim}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\overline{y}_{i,-1}\overline{\varepsilon}_i$$

The dynamic panel bias

By definition

$$\underset{N \rightarrow \infty}{plim} (NT)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} y_{i,t-1} \varepsilon_{it} = E\left(y_{i,t-1} \varepsilon_{it}\right) = 0$$

since $y_{i,t-1}$ only depends on $\varepsilon_{i,t-1}$, $\varepsilon_{i,t-2}$, etc.

The dynamic panel bias

Besides, we have:

$$\begin{aligned}
\underset{N\rightarrow\infty}{plim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}y_{i,t-1}\overline{\varepsilon}_i &= \underset{N\rightarrow\infty}{plim}\frac{1}{NT}\sum_{i=1}^{N}\overline{\varepsilon}_i\sum_{t=1}^{T}y_{i,t-1} \\
&= \underset{N\rightarrow\infty}{plim}\frac{1}{NT}\sum_{i=1}^{N}\overline{\varepsilon}_i\, T\overline{y}_{i,-1} \\
&= \underset{N\rightarrow\infty}{plim}\frac{1}{N}\sum_{i=1}^{N}\overline{\varepsilon}_i\overline{y}_{i,-1}
\end{aligned}$$

The dynamic panel bias

In the same way:

$$\underset{N\rightarrow\infty}{plim}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\overline{y}_{i,-1}\varepsilon_{it}=\underset{N\rightarrow\infty}{plim}\frac{1}{N}\sum_{i=1}^{N}\overline{y}_{i,-1}\overline{\varepsilon}_{i}$$

$$\underset{N\rightarrow\infty}{plim}\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\overline{y}_{i,-1}\overline{\varepsilon}_{i}=\underset{N\rightarrow\infty}{plim}\frac{1}{N}\sum_{i=1}^{N}\overline{y}_{i,-1}\overline{\varepsilon}_{i}$$

The dynamic panel bias

Fact

The numerator of the expression of the LSDV bias can be simplified as follows:

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

The dynamic panel bias

Let us examine this plim. By continous substitution, we have:

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

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

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

The dynamic panel bias

Fact

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

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

The dynamic panel bias

Example

Demonstration: we have (each lign corresponds to a date)

$$\sum_{t=1}^{T} y_{i,t-1} = y_{i,T-1} + y_{i,T-2} + .. + y_{i,1} + y_{i,0}$$

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

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

$$+..$$

$$+\varepsilon_{i,1} + \frac{1-\gamma^{1}}{1-\gamma}\alpha_i^* + \gamma y_{i0}$$

$$+y_{i0}$$

The dynamic panel bias

Example

For the individual effect α_i^*, we have

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

The dynamic panel bias

So, we have

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

The dynamic panel bias

Definition

Because ε_{it} are uncorrelated, independently and identically distributed, by a law of large numbers, we have:

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

The dynamic panel bias

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

$$\begin{aligned} & \underset{N\rightarrow\infty}{plim} \frac{1}{(NT)} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(y_{i,t-1} - \overline{y}_{i,-1}\right)^2 \\ = & \frac{\sigma_{\varepsilon}^2}{1-\gamma^2} \left[1 - \frac{1}{T} - \frac{2\gamma}{(1-\gamma)^2} \times \frac{(T-1) - T\gamma + \gamma^T}{T^2} \right] \end{aligned}$$

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 no matter how large N is.

The dynamic panel bias

So, we have :

$$\underset{N\to\infty}{plim}\left(\widehat{\gamma}_{LSDV}-\gamma\right)=-\frac{\frac{\sigma_{\varepsilon}^{2}}{T^{2}}\frac{(T-1)-T\gamma+\gamma^{T}}{(1-\gamma)^{2}}}{\frac{\sigma_{\varepsilon}^{2}}{1-\gamma^{2}}\left[1-\frac{1}{T}-\frac{2\gamma}{(1-\gamma)^{2}}\times\frac{(T-1)-T\gamma+\gamma^{T}}{T^{2}}\right]}$$

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

The dynamic panel bias

Definition

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

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

The dynamic panel bias

Fact

The bias of $\widehat{\gamma}_{LSDV}$ is caused by having to eliminate the unknown 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

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

The dynamic panel bias

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. Nor does the bias go to zero as γ goes to zero

The dynamic panel bias

Example

For instance, when $T = 2$, the asymptotic bias is equal to $-(1+\gamma)/2$, and when $T = 3$, it is equal to $-(2+\gamma)(1+\gamma)/2$. Even with $T = 10$ and $\gamma = 0.5$, the asymptotic bias is -0.167.

The dynamic panel bias

Fact

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**.

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)

The Instrumental Variable (IV) approach

Section 2.

The Instrumental Variable (IV) approach

The Instrumental Variable (IV) approach

We now consider a **dynamic panel data model with random individual effects**:

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

for $i = 1, .., N$ and $t = 1, .., T$. α_i^* are the (unobserved) individual effects, x_{it} is a vector of K_1 time-invariant explanatory variables , z_i is a vector of K_2 time-invariant explanatory variables and ε_{it} the error (idiosyncratic) term with $E(\varepsilon_{it}) = 0$, and $E(\varepsilon_{it}\varepsilon_{js}) = \sigma_\varepsilon^2$ if $j = i$ and $t = s$, and $E(\varepsilon_{it}\varepsilon_{js}) = 0$ otherwise.We assume that :

$$E(\alpha_i) = 0 \quad E(\alpha_i x_{it}) = 0 \quad E(\varepsilon_{it} x_{it}) = 0$$

The Instrumental Variable (IV) approach

In a vectorial form, we have:

$$y_i = y_{i,-1}\gamma + X_i\beta + ez_i'\rho + e\alpha_i + \varepsilon_{it}$$

with X_i now denoting the $T \times K_1$ time-varying explanatory variables, z_i being the $1 \times K_2$ time-invariant explanatory variables including the intercept term, and

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

The Instrumental Variable (IV) approach

Example

Reminder: The standard approach in cases where **right-hand side variables are correlated with the residuals** is to estimate the equation using **instrumental variables regression**.

1. The idea behind instrumental variables is to find a set of variables, termed **instruments,** that are both **(1) correlated with the explanatory variables in the equation**, and **(2) uncorrelated with the disturbances**.
2. These instruments are used to **eliminate the correlation between right-hand side variables and the disturbances**.

The Instrumental Variable (IV) approach

Example

Two-stage least squares (TSLS) is a special case of instrumental variables regression. As the name suggests, there are two distinct stages in two-stage least squares.

1. In the first stage, TSLS finds the portions of the endogenous and exogenous variables that can be attributed to the instruments.This stage involves estimating an OLS regression of each variable in the model on the set of instruments.
2. The second stage is a regression of the original equation, with all of the variables replaced by the fitted values from the first-stage regressions. The coefficients of this regression are the TSLS estimates.

The Instrumental Variable (IV) approach

Example

More formally, let be Z the matrix of instruments, and let y and X be the dependent and explanatory variables. Then the coefficients computed in two-stage least squares are given by

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

The Instrumental Variable (IV) approach

In the dynamic panel data models context, the **Instrumental Variable (IV)** approach was first proposed by **Anderson and Hsiao (1982)**.

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

The Instrumental Variable (IV) approach

This estimation method consists of the following procedure.

1. **First step:** Taking the **first difference of the model**, we obtain

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

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

2. **Second step: Because $y_{i,t-2}$ or $(y_{i,t-2} - y_{i,t-3})$ are correlated with $(y_{i,t-1} - y_{i,t-2})$ but are uncorrelated with $\varepsilon_{it} - \varepsilon_{i,t-1}$ they can be used as an instrument for $(y_{i,t-1} - y_{i,t-2})$and estimate γ and by the instrumental-variable method.**

The Instrumental Variable (IV) approach

Then, the following estimator is consistent.

$$\begin{pmatrix} \widehat{\gamma}_{IV} \\ \widehat{\beta}_{IV} \end{pmatrix} = \left[\sum_{i=1}^{N} \sum_{t=3}^{T} \right.$$

$$\times \begin{pmatrix} (y_{i,t-1} - y_{i,t-2})(y_{i,t-2} - y_{i,t-3}) & (y_{i,t-2} - y_{i,t-3})(x_{it} - x_{i,t-1})' \\ (x_{it} - x_{i,t-1})(y_{i,t-2} - y_{i,t-3}) & (x_{it} - x_{i,t-1})(x_{it} - x_{i,t-1})' \end{pmatrix}$$

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

The Instrumental Variable (IV) approach

Another consistent estimator is:.

$$\begin{pmatrix} \widehat{\gamma}_{IV} \\ \widehat{\beta}_{IV} \end{pmatrix} = \left[\sum_{i=1}^{N} \sum_{t=2}^{T} \right.$$

$$\left. \times \begin{pmatrix} \left(y_{i,t-1} - y_{i,t-2}\right) y_{i,t-2} & y_{i,t-2}\left(x_{it} - x_{i,t-1}\right)' \\ \left(x_{it} - x_{i,t-1}\right) y_{i,t-2} & \left(x_{it} - x_{i,t-1}\right)\left(x_{it} - x_{i,t-1}\right)' \end{pmatrix} \right]^{-1}$$

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

The Instrumental Variable (IV) approach

1. The second estimator has an advantage over the first one, 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 $(y_{i,t-2} - y_{i,t-3})$.

The Instrumental Variable (IV) approach

1. **Third step:** 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) for** $i = 1, .., N$. Let us consider, the following equation

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

with

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

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

The Instrumental Variable (IV) approach

1. **Step four:** given $\widehat{\gamma}_{IV}$, $\widehat{\beta}_{IV}$, and $\widehat{\rho}$, we can estimate the variances as follows:

$$\begin{aligned}\widehat{\sigma}_{\varepsilon}^2 &= \frac{1}{N(T-1)}\sum_{t=2}^{T}\sum_{i=1}^{N}\Big[(y_{i,t}-y_{i,t-1})-\widehat{\gamma}_{IV}(y_{i,t-1}-y_{i,t-2}) \\ &\quad -\widehat{\beta}_{IV}^{'}(x_{i,t}-x_{i,t-1})\Big]^2\end{aligned}$$

$$\widehat{\sigma}_{\alpha}^2=\frac{1}{N}\sum_{i=1}^{N}\left[\overline{y}_i-\widehat{\gamma}_{IV}\overline{y}_{i,-1}-\widehat{\beta}_{IV}^{'}\overline{x}_i-\widehat{\rho}'z_i\right]^2-\frac{1}{T}\widehat{\sigma}_{\varepsilon}^2$$

The Instrumental Variable (IV) approach

Theorem

The instrumental-variable estimators of γ, β, and σ_{ε}^2 are consistent when N or T or both tend to infinity. The estimators of ρ and σ_{α}^2 are consistent only when N goes to infinity. They are inconsistent if N is fixed and T tends to infinity.

The GMM approach

Section 3. The GMM approach

The GMM approach

Fact

We note that $y_{i,t-2}$ *or* $(y_{i,t-2} - y_{i,t-3})$ *is not the only instrument for* $(y_{i,t-1} - y_{i,t-2})$*. In fact, it is noted by Arellano and Bond (1991) that all* $y_{i,t-2-j}$, $j = 0, 1, ..,$ *satisfy the conditions*

$$E\left[y_{i,t-2-j}\left(y_{i,t-1} - y_{i,t-2}\right)\right] \neq 0$$

$$E\left[y_{i,t-2-j}\left(\varepsilon_{i,t} - \varepsilon_{i,t-1}\right)\right] = 0$$

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

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.

The GMM approach

Let us consider the same **dynamic panel data model** as before

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

for $i = 1, .., N$ and $t = 1, .., T$. α_i^* are the (unobserved) individual effects, x_{it} is a vector of K_1 time-invariant explanatory variables , z_i is a vector of K_2 time-invariant explanatory variables and ε_{it} the error (idiosyncratic) term with $E(\varepsilon_{it}) = 0$, and $E(\varepsilon_{it}\varepsilon_{js}) = \sigma_\varepsilon^2$ if $j = i$ and $t = s$, and $E(\varepsilon_{it}\varepsilon_{js}) = 0$ otherwise.We assume that :

$$E(\alpha_i) = 0 \quad E(\alpha_i x_{it}) = 0$$

The GMM approach

Definition

The GMM estimation method is also based on the **first difference of the model** (to swip ou the individual effects α_i and the time unvariant variables z_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 $t = 2, .., T$.

The GMM approach

Fact

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

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

The GMM approach

Definition

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

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

where $x_i' = (x_{i1}', .., x_{iT}')$, for each période we have the following orthogonal conditions

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

The GMM approach

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

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

where :

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

The GMM approach

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

The GMM approach

Definition

The $T(T-1)(K_1+1/2)$ orthogonality (or moment) conditions can be represented as

$$E(W_i \Delta \varepsilon_i) = 0$$

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

is of dimension $[T(T-1)(K_1+1/2)] \times (T-1)$.

The GMM approach

Proof.

The total number of ligns of the matrix W_i (equal to the number of orthognal conditions) is equal to:

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

□

The GMM approach

Definition

The number of othogonal conditions (moments), e.g. $T(T-1)(K_1+1/2)$, is much larger than the number of parameters,e.g. $K1+1$. Thus, Arellano and Bond (1991) suggest a generalized method of moments (GMM) estimator.

The GMM approach

Now we relax the assumption of exogeneity, and we assume only that explanatory variables are pre-determined, that is:

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

The GMM approach

In this case, we have

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

$$\underset{(t-1+tK_1,1)}{q_{it}}=\left(y_{i0},y_{i1},..,y_{i,t-2},x_{i1}',..,x_{it}'\right)'$$

The GMM approach

Definition

If the explanatory variables are pre-determined, there are $T\left(T-1\right)\left(K_1+1\right)/2$ orthogonality (or moment) conditions that can be represented as

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

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

is of dimension $\left[T\left(T-1\right)\left(K_1+1\right)/2\right] \times \left(T-1\right)$.

The GMM approach

Proof.

The total number of ligns of the matrix W_i (equal to the number of orthognal conditions) is equal to:

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

□

The GMM approach

Definition

Whatever the assumption made on the explanatory variable, we have largely more orthogonal conditions than parameters to estimate:

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

and we can use GMM to estimate $\theta = \left(\gamma, \beta'\right)$ in

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

The GMM approach

3.1. GMM: a general presentation

The GMM approach

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.

The GMM approach

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

$$E\left[m\left(y;\theta_0\right)\right]=0$$

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

2. Let $\widehat{m}(y,\theta)$ be the sample estimates $m(y,\theta)$ of based on N independent samples of y_i

$$\widehat{m}(y,\theta)=(1/N)\sum_{i=1}^{N}m\left(y_i;\theta\right)$$

3. Then the method of moments estimator of θ is to find $\widehat{\theta}$, such that

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

The GMM approach

Fact

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

The GMM approach

Example

Let us assume that y have a t-distrubtion with v degree of freedom. We want to estimate v from a sample $\{y_1, ..y_N\}$ where the y_i are $i.i.d.$ and have the same distribution as y. We know that the two first uncentered moments are:

$$\mu_1 = E(y) = 0 \qquad \mu_2 = E\left(y^2\right) = var(y) = \frac{v}{v-2}$$

If μ_2 is known, then v can be derived from :

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

The GMM approach

Example

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

$$\widehat{\mu}_{2,T} = \frac{1}{T} \sum_{i=1}^{T} y_t^2$$

we know that

$$\widehat{\mu}_{2,T} \xrightarrow[T \to \infty]{p} \mu_2$$

The GMM approach

Example

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

$$\widehat{v} = \frac{2\widehat{\mu}_{2,T}}{\widehat{\mu}_{2,T} - 1}$$

The GMM approach

Example

Another way to write the problem. Let us define

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

By definition, we have:

$$E[m(y;v)] = E\left(y^2 - \frac{v}{v-2}\right) = 0$$

The GMM approach

Example

Let us define

$$\widehat{m}\left(y;v\right)=\frac{1}{N}\sum\nolimits_{i=1}^{N}m\left(y_i;v\right)=\frac{1}{N}\sum\nolimits_{i=1}^{N}\left(y_i^2-\frac{v}{v-2}\right)$$

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

$$\widehat{m}\left(y;\widehat{v}\right)=0$$

The GMM approach

Definition

If the number r of equations (moments) is greater than the dimension a of θ, the system of non linear equation $\widehat{m}\left(y;\widehat{v}\right)=0$, in general has no solution. It is then necessary to minimize some norm (or distance measure) of $\widehat{m}\left(y;\widehat{\theta}\right)$, say

$$\widehat{m}\left(y;\widehat{\theta}\right)' A \ \widehat{m}\left(y;\widehat{\theta}\right)$$

where A is is some positive definite matrix. The system is **sur-identified**.

The GMM approach

What is the optimal weigthing matrix? The property of the estimator thus obtained depends on A. The optimal choice (if there is no autocorrelation of $m(y;\theta_0)$) of A turns out to be

$$A^* = \left\{ E\left[m(y;\theta_0) \;\; m(y;\theta_0) \right]' \right\}^{-1}$$

Definition

The GMM estimation of is to choose $\widehat{\theta}$ such that it minimizes the criteria

$$\widehat{\theta} = \underset{\{\theta \in \mathbb{R}^a\}}{ArgMin} q(y,\theta) = \underset{\{\theta \in \mathbb{R}^a\}}{ArgMin} \widehat{m}(y;\hat{\theta})' A^* \widehat{m}\left(y;\widehat{\theta}\right)$$

The GMM approach

Definition

Remark: if we assume that thre is autocorrelation in $m(y;\theta_0)$, the optimal weighting matrix is given by by the inverse of the long run variance covariance matrix of $m(y;\theta_0)$:

$$\underset{(r,r)}{A^*} = \left[\sum_{j=-\infty}^{\infty} E\left\{ h(\theta_0, w_t)\, h(\theta_0, w_{t-v})' \right\} \right]^{-1}$$

The GMM approach

3.2. Application to dynamic panel data models

The GMM approach

Fact

Let us consider the model

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

We want to estimate the $K_1 + 1$ *parameters of the* $\theta = (\gamma, \beta')'$ *vector. For that, we have* $T(T-1)(K_1 + 1/2)$ *moment conditions (if* x_{it} *are strictly exogeneous) that can be represented as*

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

The GMM approach

Let us denote

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

with

$$E\left[m(y_i, x_i; \theta) \right] = 0$$

The GMM approach

Definition

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

$$\widehat{\theta} = \underset{\left\{\theta \in \mathbb{R}^{K_1+1}\right\}}{ArgMin} \left(\frac{1}{N} \sum_{i=1}^{N} m\left(y_i, x_i; \theta\right) \right)' A^* \left(\frac{1}{N} \sum_{i=1}^{N} m\left(y_i, x_i; \theta\right) \right)$$

or equivalently

$$\widehat{\theta} = \underset{\left\{\theta \in \mathbb{R}^{K_1+1}\right\}}{ArgMin} \left(\frac{1}{N} \sum_{i=1}^{N} \Delta \varepsilon_i' W_i' \right) A^* \left(\frac{1}{N} \sum_{i=1}^{N} W_i \Delta \varepsilon_i \right)$$

with $A^* = \left\{ E\left[m\left(y; \theta_0\right) \; m\left(y; \theta_0\right) \right]' \right\}^{-1}$

The GMM approach

Definition

The optimal weighting matrix can be estimated by

$$\widehat{A}^{*} = \left[E\left(\frac{1}{N^{2}} \sum_{i=1}^{N} W_i \Delta\varepsilon_i \Delta\varepsilon_i' W_i' \right) \right]^{-1}$$

The GMM approach

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

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

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

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

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.

The GMM approach

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

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

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.

The GMM approach

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.

The GMM approach

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).