

# General method of moments

Hugo Harari-Kermadec

EPOG<sub>2</sub><sup>+</sup>  
Heterodox econometrics

2021

- 1 Instruments
- 2 AutoRegressive model
- 3 Generalized method of moments

- 1 Instruments
  - Linear model
  - Instruments
  - Heteroskedasticity
- 2 AutoRegressive model
- 3 Generalized method of moments

## Ordinary least squares

$$Y_i = a + bX_i + \varepsilon_i$$

Under the following hypothesis

$H_1$  : explanatory variables ( $X$ 's) are linearly independent.

$H_2$  :  $\varepsilon_i$  errors have 0 expectation.

$H_3$  :  $\varepsilon_i$  errors are uncorrelated with  $X_i$ .

$H_4$  :  $\varepsilon_i$  errors are uncorrelated with common variance  $\sigma^2$ .

The OLS estimator is unbiased and as minimum variance among linear estimators.

not  $H_1$  : explanatory variables are multicollinear  
 $\Rightarrow$  remove a variable (set a ref if you have dummies)

- not  $H_1$  : explanatory variables are multicollinear  
⇒ remove a variable (set a ref if you have dummies)
- not  $H_2$  :  $\varepsilon_i$  errors have non 0 expectation  
⇒ add an intercept

- not  $H_1$  : explanatory variables are multicollinear  
 $\Rightarrow$  remove a variable (set a ref if you have dummies)
- not  $H_2$  :  $\varepsilon_i$  errors have non 0 expectation  
 $\Rightarrow$  add an intercept
- not  $H_3$  :  $\varepsilon_i$  errors are correlated with  $X_i$   
 $\Rightarrow$  use instrumental variables

- not  $H_1$  : explanatory variables are multicollinear  
⇒ remove a variable (set a ref if you have dummies)
- not  $H_2$  :  $\varepsilon_i$  errors have non 0 expectation  
⇒ add an intercept
- not  $H_3$  :  $\varepsilon_i$  errors are correlated with  $X_i$   
⇒ use instrumental variables
- $H_4$  :  $\varepsilon_i$  errors are correlated and present heteroskedasticity  
⇒ 2 steps estimation (2SLS).



## OLS : opening the black box

Let's center  $Y$  and  $X$  :  $Y_i - \bar{Y}$  and  $X_i - \bar{X}$  to get rid of the intercept. The OLS estimator  $\hat{b}$  minimizes :

$$LS = \min_b \sum_i (Y_i - bX_i)^2 = \min_b \sum_i (\varepsilon_i)^2.$$

## OLS : opening the black box

Let's center  $Y$  and  $X$  :  $Y_i - \bar{Y}$  and  $X_i - \bar{X}$  to get rid of the intercept. The OLS estimator  $\hat{b}$  minimizes :

$LS = \min_b \sum_i (Y_i - bX_i)^2 = \min_b \sum_i (\varepsilon_i)^2$ . Let's derive  $LS$  with respect to  $b$  :

$$s(b) = \frac{\partial LS}{\partial b} = -2b \sum_i X_i (Y_i - bX_i)$$

If the model holds, there exists a value  $b_0$  for  $b$  such as  $\forall i, Y_i - b_0 X_i = \varepsilon_i$  and then

$$s(b_0) = -2b_0 \sum_i X_i (Y_i - b_0 X_i) = -2b_0 \sum_i X_i \varepsilon_i.$$

If the model holds, there exists a value  $b_0$  for  $b$  such as  $\forall i, Y_i - b_0 X_i = \varepsilon_i$  and then

$$s(b_0) = -2b_0 \sum_i X_i (Y_i - b_0 X_i) = -2b_0 \sum_i X_i \varepsilon_i.$$

Here is where we use the assumption  $H_3 : X_i \perp \varepsilon_i$ .

$$\mathbb{E}[s(b_0)] = -2b_0 \sum_i \mathbb{E}[X_i \varepsilon_i] = 0$$

If the model holds, there exists a value  $b_0$  for  $b$  such as  $\forall i, Y_i - b_0 X_i = \varepsilon_i$  and then

$$s(b_0) = -2b_0 \sum_i X_i (Y_i - b_0 X_i) = -2b_0 \sum_i X_i \varepsilon_i.$$

Here is where we use the assumption  $H_3 : X_i \perp \varepsilon_i$ .

$$\begin{aligned} \mathbb{E}[s(b_0)] &= -2b_0 \sum_i \mathbb{E}[X_i \varepsilon_i] = 0 \\ &= -2nb_0 \text{Cov}(X, \varepsilon) = 0 \end{aligned}$$

If the model holds, there exists a value  $b_0$  for  $b$  such as  $\forall i, Y_i - b_0 X_i = \varepsilon_i$  and then

$$s(b_0) = -2b_0 \sum_i X_i(Y_i - b_0 X_i) = -2b_0 \sum_i X_i \varepsilon_i.$$

Here is where we use the assumption  $H_3 : X_i \perp \varepsilon_i$ .

$$\begin{aligned} \mathbb{E}[s(b_0)] &= -2b_0 \sum_i \mathbb{E}[X_i \varepsilon_i] = 0 \\ &= -2nb_0 \text{Cov}(X, \varepsilon) = 0 \end{aligned}$$

Then we define the OLS estimator as the value of  $b$  that cancels the empirical counterpart (by the Law of Large Numbers) :

$$s(\hat{b}) = -2\hat{b} \sum_i X_i(Y_i - \hat{b}X_i) = 0$$

If the model holds, there exists a value  $b_0$  for  $b$  such as  $\forall i, Y_i - b_0 X_i = \varepsilon_i$  and then

$$s(b_0) = -2b_0 \sum_i X_i (Y_i - b_0 X_i) = -2b_0 \sum_i X_i \varepsilon_i.$$

Here is where we use the assumption  $H_3 : X_i \perp \varepsilon_i$ .

$$\begin{aligned} \mathbb{E}[s(b_0)] &= -2b_0 \sum_i \mathbb{E}[X_i \varepsilon_i] = 0 \\ &= -2nb_0 \text{Cov}(X, \varepsilon) = 0 \end{aligned}$$

Then we define the OLS estimator as the value of  $b$  that cancels the empirical counterpart (by the Law of Large Numbers) :

$$\begin{aligned} s(\hat{b}) = -2\hat{b} \sum_i X_i (Y_i - \hat{b} X_i) &= 0 \\ \Leftrightarrow \sum_i X_i Y_i &= \hat{b} \sum_i X_i X_i \end{aligned}$$

If the model holds, there exists a value  $b_0$  for  $b$  such as  $\forall i, Y_i - b_0 X_i = \varepsilon_i$  and then

$$s(b_0) = -2b_0 \sum_i X_i(Y_i - b_0 X_i) = -2b_0 \sum_i X_i \varepsilon_i.$$

Here is where we use the assumption  $H_3 : X_i \perp \varepsilon_i$ .

$$\begin{aligned} \mathbb{E}[s(b_0)] &= -2b_0 \sum_i \mathbb{E}[X_i \varepsilon_i] = 0 \\ &= -2nb_0 \text{Cov}(X, \varepsilon) = 0 \end{aligned}$$

Then we define the OLS estimator as the value of  $b$  that cancels the empirical counterpart (by the Law of Large Numbers) :

$$\begin{aligned} s(\hat{b}) = -2\hat{b} \sum_i X_i(Y_i - \hat{b}X_i) &= 0 \\ \Leftrightarrow \sum_i X_i Y_i &= \hat{b} \sum_i X_i X_i \\ \Leftrightarrow \hat{b} &= \frac{\sum_i X_i Y_i}{\sum_i X_i X_i} \end{aligned}$$

The OLS estimator for  $b$  is then the correlation.



## Panel data

$Y_{i,t}$  is very dependent of  $Y_{i,t+1}$ . We should not use a standard linear model.

$$Y_{i,t} = a + bX_{i,t} + \varepsilon_{i,t} \Leftrightarrow Y_{i,t} - a - bX_{i,t} = \varepsilon_{i,t}$$

- What if the residuals are positives for some compagnies? Correlation with  $X_{i,t} \Rightarrow H_3$  fails : instruments are needed.
- What if the residuals are larger for some compagnies?  $H_4$  fails : heteroskedasticity.
- What if the residuals are correlated **across time**?  $H_4$  fails : they have memory, include lags.

The OLS estimator  $\hat{b}$  is based on :

$$s(b) = \frac{\partial LS}{\partial b} = -2b_0 \sum_i X_i(Y_i - b_0 X_i) = -2b_0 \sum_i X_i \varepsilon_i = 0$$

If  $X_i \not\perp \varepsilon_i$ , then this equation does not hold and the OLS estimator is biased.

The OLS estimator  $\hat{b}$  is based on :

$$s(b) = \frac{\partial LS}{\partial b} = -2b_0 \sum_i X_i(Y_i - b_0 X_i) = -2b_0 \sum_i X_i \varepsilon_i = 0$$

If  $X_i \not\perp \varepsilon_i$ , then this equation does not hold and the OLS estimator is biased.

The technical solution is to replace  $X_i$  in this estimating equation (not in the model) by a variable  $Z_i$  linked with  $X_i$  but not with  $\varepsilon_i$ .

$$-2b_0 \sum_i Z_i(Y_i - b_0 X_i) = -2b_0 \sum_i Z_i \varepsilon_i = 0$$

The OLS estimator  $\hat{b}$  is based on :

$$s(b) = \frac{\partial LS}{\partial b} = -2b_0 \sum_i X_i(Y_i - b_0 X_i) = -2b_0 \sum_i X_i \varepsilon_i = 0$$

If  $X_i \not\perp \varepsilon_i$ , then this equation does not hold and the OLS estimator is biased.

The technical solution is to replace  $X_i$  in this estimating equation (not in the model) by a variable  $Z_i$  linked with  $X_i$  but not with  $\varepsilon_i$ .

$$-2b_0 \sum_i Z_i(Y_i - b_0 X_i) = -2b_0 \sum_i Z_i \varepsilon_i = 0$$

For example, the age of a company is correlated to its size, but not perfectly.

The main limitation are weak instruments : the link is too weak to estimate well the effect of  $X_i$ .

You can give instruments to the `plm` function :

```
plm(Y~X, data = db, instruments = ~Z)
```

## Heteroskedasticity

$$\hat{b} = \frac{Cov(X, Y)}{Var(X)} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

The denominator is the estimator of  $\text{var}(\varepsilon)$  under the assumption of homoskedasticity (and independence).

The 2SLS estimates first this denominator, and then plug it in the definition of  $\hat{b}$ .

## Heteroskedasticity

$$\hat{b} = \frac{\text{Cov}(X, Y)}{\text{Var}(X)} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$$

The denominator is the estimator of  $\text{var}(\varepsilon)$  under the assumption of homoskedasticity (and independence).

The 2SLS estimates first this denominator, and then plug it in the definition of  $\hat{b}$ .

## Lags

Include lags in the model and/or estimate a second model for  $\varepsilon$

$$Y_{i,t} = rY_{i,t-1} + a + bX_{i,t} + \varepsilon_{i,t}$$

## AutoRegressive model : AR(p)

In AutoRegressive models,  $\varepsilon_t$  depends directly on its past.

### AR model

$$\varepsilon_t = a_1 \varepsilon_{t-1} + a_2 \varepsilon_{t-2} + \cdots + a_p \varepsilon_{t-p} + e_t$$

$$\Leftrightarrow (1 - a_1 L - a_2 L^2 + \cdots - a_p L^p) \varepsilon_t = A(L) \varepsilon_t = e_t$$

- $a_0 = 1$  by definition and  $e_t$  is iid.
- $p$  is the order of the AR : how long the past affects the present.
- the polynomial  $A(x)$  has no unit root.

### Theorem

*If  $\varepsilon_t$  is a AR(p), then  $A(x)$  has no unit root and  $\varepsilon_t$  is stationary.*



## Partial autocorrelation function

Suppose we don't know the order  $p$  of the model :

$$\varepsilon_t = a_1 \varepsilon_{t-1} + a_2 \varepsilon_{t-2} + \cdots + a_p \varepsilon_{t-p} + e_t$$

How to find  $p$ ? We consider all the nested models of order  $k \in \mathbb{N}^*$ . PACF( $k$ ) is the coefficient of the last variable of the linear model explaining  $\varepsilon_t$  by its  $k$  past values.

$$\varepsilon_t = \hat{a}_1 \varepsilon_{t-1} + \hat{a}_2 \varepsilon_{t-2} + \cdots + PACF(k) \varepsilon_{t-k} + e_t$$

### Theorem

*If  $\varepsilon_t$  is a  $AR(p)$  then  $PACF(p) \neq 0$  and for any  $k > p$ ,  $PACF(k) = 0$ .*

# Orthogonality

## Theorem

If  $\varepsilon_t$  is a  $AR(p)$  then  $PACF(p) \neq 0$  and for any  $k > p$ ,  $PACF(k) = 0$ .

$$\varepsilon_t = \hat{a}_1 \varepsilon_{t-1} + \hat{a}_2 \varepsilon_{t-2} + \cdots + \hat{a}_p \varepsilon_{t-p} + e_t$$

$$\varepsilon_t - \hat{a}_1 \varepsilon_{t-1} - \hat{a}_2 \varepsilon_{t-2} - \cdots - \hat{a}_p \varepsilon_{t-p} = e_t$$

## Orthogonality

### Theorem

If  $\varepsilon_t$  is a  $AR(p)$  then  $PACF(p) \neq 0$  and for any  $k > p$ ,  $PACF(k) = 0$ .

$$\varepsilon_t = \hat{a}_1 \varepsilon_{t-1} + \hat{a}_2 \varepsilon_{t-2} + \cdots + \hat{a}_p \varepsilon_{t-p} + e_t$$

$$\varepsilon_t - \hat{a}_1 \varepsilon_{t-1} - \hat{a}_2 \varepsilon_{t-2} - \cdots - \hat{a}_p \varepsilon_{t-p} = e_t$$

then any further lag is orthogonal to  $e_t$

### Further lags are instruments

$$e_t \perp \varepsilon_{t-p-k} \quad \forall k > 0$$

- 1 Instruments
- 2 AutoRegressive model
- 3 Generalized method of moments

## Generalized method of moments

GMM are designed to estimate any model where the parameters  $\theta$  can be defined as solutions of an equation  $\mathbb{E}[m(\theta, X_i)] = 0$  with  $\dim(m) \geq \dim(\theta)$ .  $m$  replaces  $s$  in a more general setting than replacing  $Y$  by an instrument.

## Generalized method of moments

GMM are designed to estimate any model where the parameters  $\theta$  can be defined as solutions of an equation  $\mathbb{E}[m(\theta, X_i)] = 0$  with  $\dim(m) \geq \dim(\theta)$ .  $m$  replaces  $s$  in a more general setting than replacing  $Y$  by an instrument.

The simple linear model writes :  $Y_i = a + bX_i + \varepsilon_i$

## Generalized method of moments

GMM are designed to estimate any model where the parameters  $\theta$  can be defined as solutions of an equation  $\mathbb{E}[m(\theta, X_i)] = 0$  with  $\dim(m) \geq \dim(\theta)$ .  $m$  replaces  $s$  in a more general setting than replacing  $Y$  by an instrument.

The simple linear model writes :  $Y_i = a + bX_i + \varepsilon_i$

That is, under  $H_3$ ,

$$\mathbb{E}[1(Y_i - a + bX_i)] = 0 \text{ and } \mathbb{E}[X_i'(Y_i - a + bX_i)] = 0$$

and, if  $H_3$  does not hold, with instruments :

$$\mathbb{E}[1(Y_i - a + bX_i)] = 0 \text{ and } \mathbb{E}[Z_i'(Y_i - a + bX_i)] = 0$$

You can also add more instruments than  $X$ 's, or use any linear form.

## PGMM function

**Not the PGMM package!** The pgmm function of PLM package.  
 Suppose you want to estimate the model

$$Y_{i,t} = a + bY_{i,t-1} + cX_{i,t} + dX_{i,t-1} + \varepsilon_{i,t}$$

using the lags of Y and X further than 1 (for example from 2 to 10) as GMM instruments and some classical instrument Z.

The pgmm function has 3 parts for the model, separated by |

- ① the model itself `Y ~lag(Y)+X+lag(X)`
- ② the gmm instruments `lag(Y+X,2:10)`
- ③ the classical instruments Z

```
output <- pgmm(Y ~lag(Y)+X+lag(X) |lag(Y+X,2:10) |Z,
data=db_toy,index=c("Company","Year"),
effect = "twoways", model = "twosteps")
summary(output)
```