General method of moments

Hugo Harari-Kermadec

EPOG
Econometrics

Nov. 21, 2019

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 : ε_i errors have 0 expectation.
- H_3 : ε_i errors are uncorrelated with X_i.
- H_4 : ε_i errors are uncorrelated with common variance σ^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
$\Rightarrow$ remove a variable (set a ref if you have dummies)

not H_2 : ε_i errors have non 0 expectation
$\Rightarrow$ add an intercept

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

not H_2 : ε_i errors have non 0 expectation
$\Rightarrow$ add an intercept

not H_3 : ε_i errors are correlated with X_i
$\Rightarrow$ use instrumental variables

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

not H_2 : ε_i errors have non 0 expectation
$\Rightarrow$ add an intercept

not H_3 : ε_i errors are correlated with X_i
$\Rightarrow$ use instrumental variables

H_4 : ε_i errors are correlated and present heteroskedasticity
$\Rightarrow$ 2 steps estimation (2SLS).

OLS : opening the black box

Let's center Y and X : $Y_i - \overline{Y}$ and $X_i - \overline{X}$ to get rid of the intercept. The OLS estimator $\hat{b}$ minimizes :
$S = \min_b \sum_i (Y_i - bX_i)^2 = \min_b \sum_i (\varepsilon_i)^2$.

OLS : opening the black box

Let's center Y and X : $Y_i - \overline{Y}$ and $X_i - \overline{X}$ to get rid of the intercept. The OLS estimator $\hat{b}$ minimizes :
$S = \min_b \sum_i (Y_i - bX_i)^2 = \min_b \sum_i (\varepsilon_i)^2$. Let's derive S with respect to b :

$$\frac{\partial S}{\partial b} = -2b \sum_i X_i(Y_i - bX_i) = -2b \sum_i X_i \varepsilon_i.$$

OLS : opening the black box

Let's center Y and X : $Y_i - \overline{Y}$ and $X_i - \overline{X}$ to get rid of the intercept. The OLS estimator $\hat{b}$ minimizes :
$S = \min_b \sum_i (Y_i - bX_i)^2 = \min_b \sum_i (\varepsilon_i)^2$. Let's derive S with respect to b :

$$\frac{\partial S}{\partial b} = -2b \sum_i X_i (Y_i - bX_i) = -2b \sum_i X_i \varepsilon_i.$$

Here is where we use that $X_i \perp \varepsilon_i$.

$$\frac{\partial S}{\partial b} = 0 \leftrightarrow \sum_i X_i Y_i = b \sum_i X_i X_i \Rightarrow b = \frac{\sum_i X_i Y_i}{\sum_i X_i X_i}$$

OLS : opening the black box

Let's center Y and X : $Y_i - \overline{Y}$ and $X_i - \overline{X}$ to get rid of the intercept. The OLS estimator $\hat{b}$ minimizes :
$S = \min_b \sum_i (Y_i - bX_i)^2 = \min_b \sum_i (\varepsilon_i)^2$. Let's derive S with respect to b :

$$\frac{\partial S}{\partial b} = -2b \sum_i X_i (Y_i - bX_i) = -2b \sum_i X_i \varepsilon_i.$$

Here is where we use that $X_i \perp \varepsilon_i$.

$$\frac{\partial S}{\partial b} = 0 \leftrightarrow \sum_i X_i Y_i = b \sum_i X_i X_i \Rightarrow b = \frac{\sum_i X_i Y_i}{\sum_i X_i X_i}$$

The actual OLS estimator for b is the correlation :

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

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 correlated **across time** ? H_4 fails : they have memory, include lags.

The OLS estimator $\hat{b}$ minimizes : $S = \min_b \sum_i (Y_i - bX_i)^2$ Let's derive S with respect to b :

$$\frac{\partial S}{\partial b} = -2b \sum_i X_i \varepsilon_i .$$

If $X_i \not\perp \varepsilon_i$, then the estimator will be biased. The technical solution is to replace X_i (here only, not in the model) by a variable linked with X_i but not with ε_i.
For example, the age of a compagny is correlated to its size, but not perfectly.
The main limitation are weak instruments : the link is to 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

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

The denominator is the estimator of var(ε) under the assumption of homoskedasticity (and independence).
The 2SLS estimates first this denominator, and then plug it in the definition of $\hat{b}$.

Heteroskedasticity

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

The denominator is the estimator of var(ε) 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 ε

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

AutoRegressive model : AR(p)

In AutoRegressive models, ε_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 ε_t is a AR(p), then $A(x)$ has no unit root and ε_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 ε_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 ε_t is a AR(p) then $PACF(p) \neq 0$ and for any $k > p$, $PACF(k) = 0$.

Orthogonality

Theorem

If ε_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 ε_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 θ can be defined as solutions of an equation $\mathbb{E}[m(\theta, X_i)] = 0$ with $dim(m) \geq dim(\theta)$.
The simple linear model writes :

$$Y_i = a + bX_i + \varepsilon_i \text{ with } \mathbb{E}[\varepsilon_i] = 0$$

that is

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

and 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.

PGMM function

not the package PGMM !
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,index=c("Company_Name","year"),
effect = "twoways", model = "twosteps")
summary(output)
```