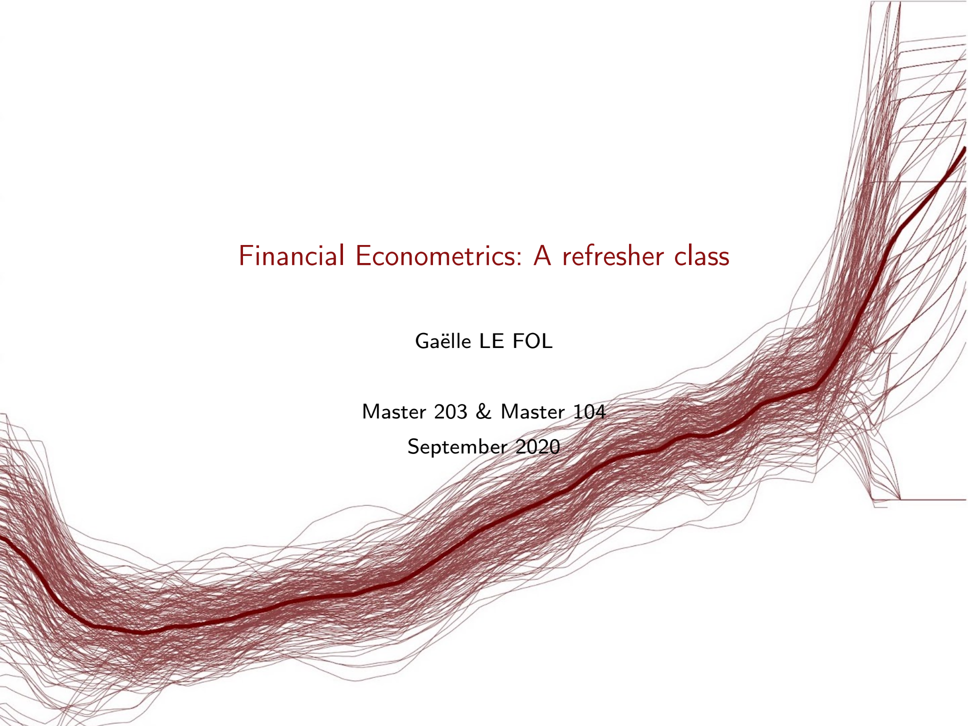

Financial Econometrics: A refresher class

Gaëlle LE FOL

Master 203 & Master 104

September 2020

Outline

The two stage least squares method (2SLS), also called instrumental variables (IV) regression, is an alternative to the OLS when at least one of the explanatory variables is endogenous. In that case, the variable in question is correlated with the error term violating one of the OLS hypotheses.This situation might occur when one of the variable is measured with error, if there are omitted variables in the model - that are correlated with the dependent variable and the endogenous explanatory variable-, or if the two variables are simultaneously determined.

The idea of the 2SLS is to use intrumental variables that are not correlated with the error term to estimate the parameters of the model. These instrumental variables are correlated to the endogenous variables but not to the error term.
For example, if we have the following model :

$$\begin{aligned} y_{1,t} &= \beta_0 + \beta_1 x_{1,t} + \beta_{22} x_{2,t} + \beta_3 y_{2,t} + \varepsilon_{1,t}, & (64) \\ y_{2,t} &= \gamma_0 + \gamma_1 x_{1,t} + \gamma_2 x_{2,t} + \gamma_3 x_{3,t} + \varepsilon_{2,t}, & (65) \end{aligned}$$

where $y_{1,t}$ and $y_{2,t}$ are endogenous variables, $x_{1,t}$, $x_{2,t}$ and $x_{3,t}$ are exogenous variables and $\varepsilon_{1,t}$ and $\varepsilon_{2,t}$ independent white noise.

Two stage least squares II

The 2SLS method consists in :

Step 1 : Regress y_2 on the exogenous variables,

$$y_{2,t} = \hat{\gamma}_0 + \hat{\gamma}_1 x_{1,t} + \hat{\gamma}_2 x_{2,t} + \hat{\gamma}_3 x_{3,t} + \hat{\varepsilon}_{2,t} \tag{66}$$

Step 2 : Replace $y_{2,t}$ by $\hat{y}_{2,t} + \hat{\varepsilon}_{2,t}$ in the equation of y_1 (64),

$$\begin{aligned} y_{1,t} &= \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 \left(\hat{y}_{2,t} + \hat{\varepsilon}_{2,t}\right) + \varepsilon_{1,t} \\ &= \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 \hat{y}_{2,t} + \varepsilon_{1,t} + \beta_3 \hat{\varepsilon}_{2,t} \\ &= \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \beta_3 \hat{y}_{2,t} + \tilde{\varepsilon}_t, \end{aligned} \tag{67}$$

with $\tilde{\varepsilon}_t = \varepsilon_{1,t} + \beta_1 \hat{\varepsilon}_{2,t}$.

The 2SLS are filtering out the endogenous part of the stochastic explanatory variable $y_{2,t}$ to keep the exogenous part thanks to the regressions (66) and (67).

- Consider the model

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t.$$

y_t is n.i.d. with mean $\beta_0 + \beta_1 x_t$ and variance σ^2.

- The joint probability density function of $y_1, y_2, \ldots, y_T$ given the mean and variance is

$$f(y_1, y_2, \ldots, y_T | \beta_0 + \beta_1 x_t, \sigma^2)$$

- The independence of the y's gives

$$\begin{aligned} f(\quad y_1 \quad & , y_2, \ldots, y_T | \beta_0 + \beta_1 x_t, \sigma^2) \\ = \quad & f(y_1 | \beta_0 + \beta_1 x_t, \sigma^2) f(y_2 | \beta_0 + \beta_1 x_t, \sigma^2) \ldots f(y_T | \beta_0 + \beta_1 x_t, \sigma^2), \end{aligned}$$

where

$$f(y_t) = \frac{1}{\sigma\sqrt{2\pi}} exp\left\{-\frac{1}{2}\frac{(y_t - \beta_0 - \beta_1 x_t)^2}{\sigma^2}\right\}$$

- Finally,

$$\begin{aligned} f(y_1, y_2, \ldots, \quad y_T \quad & | \beta_0 + \beta_1 x_t, \sigma^2) \\ = \quad & \frac{1}{\sigma^T(\sqrt{2\pi})^T} exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - \beta_0 - \beta_1 x_t)^2}{\sigma^2}\right\} \end{aligned}$$

- If the y's are given and the parameters β_0, β_1, and σ^2 are unknown, this function is called the likelihood function, denoted $L(\beta_0, \beta_1, \sigma^2)$ and written as :

$$L(\beta_0, \beta_1, \sigma^2) = \frac{1}{\sigma^T(\sqrt{2\pi})^T} exp\left\{-\frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - \beta_0 - \beta_1 x_t)^2}{\sigma^2}\right\} \quad (68)$$

- The method of maximum likelihood consists in estimating the unknown parameters such that the log likelihood function, observing the given y's, is maximum.
- In order to get these parameters, we differentiate the log likelihood :

$$lnL(\beta_0, \beta_1, \sigma^2) = -\frac{T}{2}ln(\sigma^2) - \frac{T}{2}ln(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\frac{(y_t - \beta_0 - \beta_1 x_t)^2}{\sigma^2} \quad (69)$$

with respect to β_0, β_1 and σ^2.

The ML estimators

- The ML estimators of the regression coefficients $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ are identical to the OLS estimators.
- The ML estimator of σ^2 is $\sum e_t^2/T$. This estimator is biased contrary to the OLS estimator $\sum e_t^2/(T-2)$.
- As T increases, the two estimators tend to be equal. Asymptotically, the ML estimator of σ^2 is also unbiased.

Exercise 25 : Maximum Likelihood estimators

Show that the three above properties hold.

What if we have another type of distribution ?

Exercise 26 : ML estimators and other distribution

Consider a variable z exponentially distributed with parameter λ. Use the ML method to estimated λ.

We already have seen that the generalized least square are a method to correct problems of autocorrelation and/or heteroscedasticity of the residuals, i.e. when the variance-covariance matrix of the residuals Ω is not a scalar but is known.
In practice, this matrix is usually unknown and has to be estimated. One possible solution is to use the estimated GLS of feasible GLS (FGLS), by doing :

1. We estimated the model using OLS ignoring the structure of the variance-covariance matrix, and thus the problems of autocorrelation and/or heteroskedasticity of the residuals. We save the estimated residuals.
2. We take the following as for variance-covariance matrix estimator

$$\hat{\Omega}_\varepsilon = \begin{pmatrix} \hat{\varepsilon}_1^2 & \hat{\varepsilon}_1\hat{\varepsilon}_2 & \dots & \hat{\varepsilon}_1\hat{\varepsilon}_n \\ \hat{\varepsilon}_1\hat{\varepsilon}_2 & \hat{\varepsilon}_2^2 & \dots & \hat{\varepsilon}_2\hat{\varepsilon}_n \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\varepsilon}_1\hat{\varepsilon}_n & \hat{\varepsilon}_2\hat{\varepsilon}_n & \dots & \hat{\varepsilon}_n^2 \end{pmatrix}$$

3. The estimated parameters are $\hat{\beta}_{FGLS} = (X'\hat{\Omega}_\varepsilon^{-1}X)^{-1}X'\hat{\Omega}_\varepsilon^{-1}Y$ and the estimator of the variance of the parameters is $\hat{V}(\hat{\beta}) = (X'\hat{\Omega}_\varepsilon^{-1}X)^{-1}$

The OLS are focusing on modelling the conditional mean of the dependent variable, - i.e. the relationship between the average explained variable and average values of the explanatory variables. The model gives the conditional expectation of the explained variable, knowing the explanatory variables. The estimation gives the conditional average.

$$\hat{\beta}_\tau = argmin \sum_{i=1}^{N} (y_i - X_i\beta)^2, \quad \text{, and we have} \quad \sum_{i=1}^{N} \hat{\varepsilon_i} = \sum_{i=1}^{N} (y_i - X_i\hat{\beta}) = 0.$$

Moreover, the OLS needs quite strong hypotheses to be BLUE that may not be statisfied. The quantile regression needs no assumption on the distribution of the errors and can explore different aspect of the relationship (not only the average) between the explained variable and the explanatory variables. It is far more robust to outliers and non-normality. Still in quantile regression, the independent variable, usually called the response variable, is assume to be non autocorrelated and homoskedastic.

For example, if the explained variable is multimodal or if the relationship is not linear, the quantile regression will be appropriate while the OLS won't. Obviously in any of those cases, there some alternative ways to estimate the relationship (like adding polynomial terms in the model equation for example) but the quantile regression may be more natural.

Finally, if we are more interested in what is going on in queues of the distribution than in the middle (average), we will favor the quantile regression rather than the OLS regression.

Typically, we can try to understand returns but we can be more interested in extrem returns as there are the ones that will impose particular protections to asset managers. The evaluation of Value-at-risk model is of that kind. In fact, as soon as we are more interested in anything that is not the conditional mean (the conditional median or any other conditional quantile), we will use quantile regression.

Definition

The τ^{th} quantile of the distribution of a random variable Y is such that

$$Q_Y(\tau) = inf\{Y : F_Y(y) \geq \tau\},$$

where $F_Y(y) = P(Y \leq y)$ is the cumulated distribution function and *inf* refere to the infinimum and means the smallest y satisfaying the inequality. All quantiles are between 0 and 1.

Quantile regressions effectively model the entire conditional distribution of y (and not only the mean as is the case in the OLS) and explore their impact on the location, scale and shape of the distribution of y.

We can estimate the quantile q_τ in the sample, by solving the minimization problem :

$$\hat{q}_\tau = argmin \sum_{i=1}^{N} \{\rho_\tau(y_i - q)\},$$

where ρ is the loss function : $\rho_\tau(y) = y(\tau - \mathbb{1}_{y<0})$.

How would you write the developped equation ?

$$\hat{q}_\tau = argmin \left[(\tau - 1) \sum_{y_i<q}^{N} (y_i - q) + \tau \sum_{y_i \geq q}^{N} (y_i - q) \right],$$

Applied to the regression problem, instead of looking for $\bar{Y} = \bar{X}\beta$, we want $Q_{Y|X}(\tau) = X\beta_\tau$ and we get

$$\hat{\beta}_\tau = argminE\{\rho_\tau(Y - X\beta)\}$$

Here again, in the sample we have :

$$\hat{\beta}_\tau = argmin \sum_{i=1}^{N} \rho_\tau(y_i - X_i\beta)$$

How would you write the developped equation ?

$$\hat{\beta}_\tau = argmin \left[(\tau - 1) \sum_{y_i < X\beta}^{N} (y_i - X_i\beta) + \tau \sum_{y_i \geq X\beta}^{N} (y_i - X_i\beta) \right],$$

Finally, we have

$$\hat{\beta}_\tau = argmin \left[(1 - \tau) \sum_{y_i < X\beta}^{N} |y_i - X_i\beta| + \tau \sum_{Y_i \geq X\beta}^{N} |y_i - X_i\beta| \right],$$

We can select the quantile of interest by setting τ to the corresponding value but the most common choices are $0.05, 0.01, 0.25, 0.5, 0.75, 0.9$, or 0.95. Setting τ 0.5, well be fitting the median and the weights will be symmetric. Fo all the other quantiles, the weights are not symmetric.

We can transform this minimization problem into a linear problem that a simplex method can solve. The Generalised Method of Moments (GMM) can also be a adequate tool to estimate the parameters of interest but is out of the scope of this course.

Exercise 27 : Quantile regression

Get the article of Ram (2008) from the library and read it. Use the data set called **mrw.gdt** and reproduce his Table 1, panel A by doing three regressions :

1. An OLS regression.
2. A quantile regression using the first quartile.
3. A quantile regression using the last quartile.

What does that quantile regression add to the understanding ?

R. RAM (2008) :" Parametric variability in cross-country growth regressions : An application of quantile-regression methodology", Economics Letter 99, p. 387-389.