Econometrics

Spring semester 2024

Lecturer: Nicolás Blampied

Contents

- Time-series econometrics
    - Lecture 1: Basic regression analysis with time series data
    - Lecture 2: Trends and seasonality
    - Lecture 3: Further Issues in Using OLS with Time Series Data
    - Lecture 4: Serial Correlation and Heteroskedasticity in Time Series Regressions
    - Lecture 5 & 6: Advanced Time Series Topics

Practical information

- Literature

    - WOOLDRIDGE, Jeffrey M. Introductory econometrics : a modern approach. Seventh edition. Boston: Cengage Learning, 2020. xxi, 826. ISBN 9781337558860.

    - HEISS, Florian. Using R for introductory econometrics. 2nd edition. Düsseldorf: Florian Heiss, 2020. 368 stran. ISBN 9788648424364.

- Assessment methods

    Class assignments: 50 %

    Class assignments will account for 50 % of the final grade. These exercises will enhance your problem-solving skills and prepare you for midterms (or the final exam).

Practical information

- Assessment methods

Midterm exams: 50 %

Two midterm exams will take place during regular seminar time on March 29th (topics related to the time-series econometrics, 25 % of the final grade) and on May 17th (topics related to the panel data econometrics and limited-dependent variable models, 25 % of the final grade). Students who fail to attend the midterms or want to improve their final grade may take the final exam. It will include all topics related to the course. The outcomes of the final exam also replace the results of both midterm exams).

Grade distribution:

A: 90 – 100	D: 66 – 73
B: 82 – 89	E: 60 – 65
C: 74 – 81	F: 0 – 59

Lecture 1:

Basic regression analysis with time series data

Contents

Lecture 1

- About time series
- Static and finite distributed lag models
- Finite sample properties of OLS
- Gauss-Markov theorem
- Normality and inference

What is time series data?

- **If we compare time-series data with cross-sectional data, we could say that the main difference is that the data follows a temporal order.** This implies that data corresponding to a period t is preceded by data at time t-1.

- **We can think about these variables as random, since we cannot foresee the outcome of variables in the future.** For instance, do we know what the unemployment rate or GDP growth will be next year?

- **A sequence of random variables indexed by time is called a stochastic process or a time series process**. When we gather the data, we only see one realization of the process.

- If we could go back in time and change anything, we could get a different one. But we cannot, this is why we should take time series as random (like random sampling when working with cross-sectional data).

What is time series data?

Time series vs Cross sectional data

Time series

- Observation of a single subject at different time spans.
- The idea is to track the evolution of a variable along a certain period of time.
- There are infinite examples, can you think of one?

Cross sectional data

- Observation of multiple subjects at the same time.
- The idea is to undertand the difference among these diverse subjects at a particular time.
- There are infinite examples, can you think of one?

What is time series data?

Cross sectional data

Country	Year	GDP
Australia	2023	5
Austria	2023	4
Czech Republic	2023	7
Nigeria	2023	4

Panel data (second part of the course)

Country	Year	GDP
Australia	2021	3
Australia	2022	4
Australia	2023	5
Austria	2021	5
Austria	2022	2
Austria	2023	4
Czech Republic	2021	8
Czech Republic	2022	6
Czech Republic	2023	7
Nigeria	2021	4
Nigeria	2022	4
Nigeria	2023	4

What is time series data?

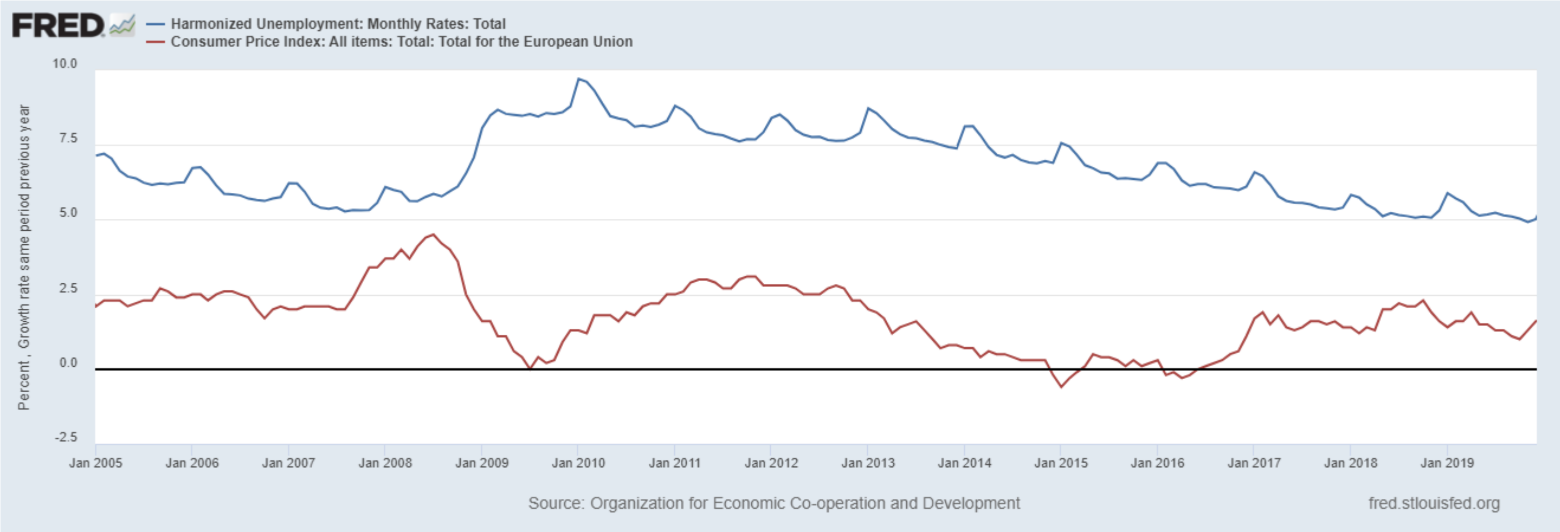


Figure 1: Unemployment and inflation in the European Union. **Not seasonally adjusted.**

Static and finite distributed lag models

Static model:

Let us assume we have time series data on two variables, y and z and the following ‘static model’ where z determines y.

$$y_t = \beta_0 + \beta_1 z_t + \mu_t, \ t = 1, 2, \dots, n. \qquad (1)$$

The model is called ‘static’ because it assumes that changes in z have an immediate effect on y.

Can you think of an example of a ‘static model’. Do you remember the Phillips curve?

Static and finite distributed lag models

So, how would Eq. (1) change?

$$inflation_t = \beta_0 + \beta_1\, unemployment_t + \mu_t, \quad t = 1, 2, \ldots, n. \qquad (2)$$

What do you think a drawback of this simple model is?

Do you think the relationship between unemployment and inflation is contemporaneous?

If past unemployment rates affect inflation today, how would you modify Eq. (2)?

Static and finite distributed lag models

What about including lags into the model?

$$inflation_t = \beta_0 + \beta_1\, unem_t + \beta_2\, unem_{t-1} + \beta_3\, unem_{t-2} + \mu_t \qquad (3)$$

This would be a 'finite distributed lag model' (FDL), where we allow past unemployment rates to affect inflation today.

This FDL model is of order 2, since it includes two lags of the explanatory variable.

How do we measure the impact of a change in unemployment on inflation? Two important concepts…

Static and finite distributed lag models

Impact propensity or impact multiplier

Let us assume that unemployment=c at all periods, but only at time t unemployment = c +1 (one-time increment). Let us also assume, for simplicity, that the residuals are zero.

$$inflation_{t-1} = \beta_0 + \beta_1 c + \beta_2 c + \beta_3 c$$

$$inflation_{t} = \beta_0 + \beta_1 (c+1) + \beta_2 c + \beta_3 c$$

$$inflation_{t+1} = \beta_0 + \beta_1 c + \beta_2 (c+1) + \beta_3 c$$

$$inflation_{t+2} = \beta_0 + \beta_1 c + \beta_2 c + \beta_3 (c+1)$$

$$inflation_{t+3} = \beta_0 + \beta_1 c + \beta_2 c + \beta_3 c$$

Note that $inflation_t$ - $inflation_{t-1}$ = β_1

β_1 is called the impact multiplier and measures the instant change in inflation due to a one-unit change in unemployment at time t.

Equivalently: $inflation_{t+1}$ - $inflation_{t-1}$ = β_2 ; $inflation_{t+2}$ - $inflation_{t-1}$ = β_3.

At t+3, $inflation_{t+3}$ = $inflation_{t-1}$ (we are back to the initial level of inflation).

Static and finite distributed lag models

Long-run propensity (LRP) or long-run multiplier

What if the increment was permanent?

$$inflation_{t-1} = \beta_0 + \beta_1 c + \beta_2 c + \beta_3 c$$

$$inflation_t = \beta_0 + \beta_1(c+1) + \beta_2 c + \beta_3 c$$

$$inflation_{t+1} = \beta_0 + \beta_1(c+1) + \beta_2 (c+1) + \beta_3 c$$

$$inflation_{t+2} = \beta_0 + \beta_1(c+1) + \beta_2(c+1) + \beta_3(c+1)$$

Note that $inflation_t$ - $inflation_{t-1} = \beta_1$

However, now:

$inflation_{t+1}$ - $inflation_{t-1} = \beta_1 + \beta_2$

$inflation_{t+2}$ - $inflation_{t-1} = \beta_1 + \beta_2 + \beta_3$.

Since we only have two lags:

$$\textbf{LRP} = \boldsymbol{\beta_1 + \beta_2 + \beta_3}$$

which we interpret as the long run impact it has a one-unit permanent increase in unemployment.

Finite sample properties of OLS

Unbiasedness of OLS

Under certain assumptions, we can be sure that OLS is unbiased.

<u>Assumption 1: linear in parameters</u>

The stochastic process $\{(x_{t1}, x_{t2}, \ldots, x_{tk}, y_t): t = 1, 2, \ldots, n\}$ follows a linear model in its parameters:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_k x_{tk} + U_t$$

where $\{u_{t1}: t = 1, 2, \ldots, n\}$ is the sequence of errors and n is the number of observations (time periods).

Time period

Indicates each explanatory variable

Finite sample properties of OLS

Which one of this is non-linear in parameters?

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + U_t \qquad (a)$$

Variables

$$y_t = \beta_0 + {\beta_1}^2 x_{t1} + \beta_2 x_{t2} + U_t \qquad (c)$$

Finite sample properties of OLS

Assumption 2: No perfect collinearity

In the sample (and therefore in the underlying time series process), no independent variable is constant nor a perfect linear combination of the others.

Explanatory variables may be correlated, but they cannot be perfectly correlated!

It would be impossible to perform our estimation if two variables were perfectly correlated.

Finite sample properties of OLS

This is how a time series dataset looks like:

t	Inflation (y)	Unemployment (x1)	Real interest rate (x2)
1	5%	6%	3%
2	6%	5%	2%
3	7%	5%	1%
4	4%	4%	1%
5	5%	4%	0%
6	6%	4%	1%
7	8%	3%	1%
8	7%	4%	2%
9	6%	5%	3%
10	5%	5%	4%
11	4%	6%	4%
12	3%	7%	5%

Table 1

Do you see any problem in table 2?

x2 = 2x1

t	Inflation (y)	Unemployment (x1)	Real interest rate (x2)
1	5%	6%	12%
2	6%	5%	10%
3	7%	5%	10%
4	4%	4%	8%
5	5%	4%	8%
6	6%	4%	8%
7	8%	3%	6%
8	7%	4%	8%
9	6%	5%	10%
10	5%	5%	10%
11	4%	6%	12%
12	3%	7%	14%

Table 2

Here inflation is our dependent variable (y_t), while unemployment and the real interest rate are our explanatory variables (x_t1 and x_t2). In this case, we assume that our series extend for 12 periods (n=12).

Finite sample properties of OLS

$A =$

1	2	4
4	1	2
6	2	4

$\det|A| = 4 + 32 + 24 - 24 - 4 - 32 = 0$ → **NON INVERTIBLE**

Finite sample properties of OLS

Assumption 3: Zero conditional mean

For each t, the expected value of the error u_t, given the explanatory variables for all time periods, is zero. Mathematically:

$$E(u_t / X) = 0, \qquad t = 1, 2, \dots n.$$

Intuitively, this implies that the explanatory variables should be independent from the residual term. In simple words, u_t should be uncorrelated with each explanatory variable at each t.

Or, if you want to put differently, we could say that the average value of u_t is unrelated to X.

Finite sample properties of OLS

What could make Assumption 3 to fail?

- Omitted variables

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + U_t \quad (a)$$

$$y_t = \beta_0 + \beta_1 x_{t1} + U_t \quad (b)$$

x_{t2},
U_t,
x_{t1},
l.

- Measurement error of explanatory variables.

If variables are badly measured, then some information may fall on the residuals.

Finite sample properties of OLS

Theorem

Under Assumptions 1, 2 and 3, the OLS estimators are unbiased conditional on X.

Reminder: For those who forgot it, remember that the bias of an estimator is the difference between the expected value of the estimator and the true value of the parameter estimated.

If we think of a parameter θ, unbiasedness means that $Bias\left[\hat{\theta}\right] = E\left[\hat{\theta}\right] - \theta = 0$.

Simply put, an estimator is unbiased if its estimates are on average correct.

Gauss-Markov theorem

So, we have an unbiased estimator, but is this the best estimator we can have?
What else do we need to be sure this is the case?

Gauss-Markov theorem

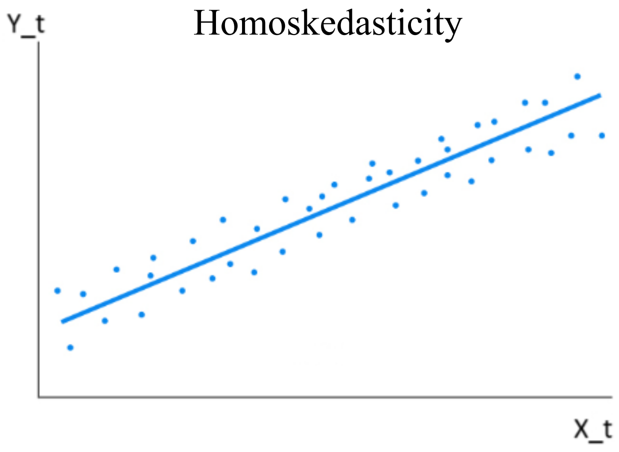
Homoskedasticity

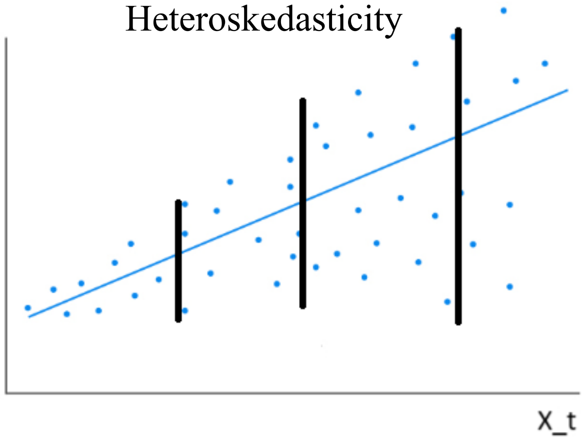
No serial correlation

Gauss-Markov theorem

Assumption 4: Homoskedasticity

Conditional on X, the variance of u_t is the same for all t: Var(u_t/X) = Var(u_t/X) = σ^2, t = 1, 2, …, n.





Gauss-Markov theorem

Assumption 5: No serial correlation or autocorrelation

Conditional on X, the errors in two different time periods are uncorrelated: $Corr(u_t, u_s / X) = 0$, for all $t \neq s$.

https://www.youtube.com/watch?v=ZjaBn93YPWo&t=604s

Gauss-Markov theorem

<u>OLS sampling variances</u>

Under Assumptions 1, 2, 3, 4 and 5, the variance of $\widehat{\beta_j}$, conditional on X, is:

$Var(\beta_j/X = \sigma^2 / [\text{SST}_j\ (1 - R_j^2)], j = 1, \dots k.$

where SST_j is the total sum of squares of x_{tj} and R_j^2 is the R-squared from the regression of x_j.

Under Assumptions 1, 2, 3, 4 and 5, the estimator $\hat{\sigma} = \frac{SSR}{n-k-1}$ is an unbiased estimator of σ^2.

Gauss-Markov theorem

Zero conditional mean

Linear in parameters

No perfect collinearity

Under the five assumptions we just reviewed, the OLS estimator is BLUE

(best linear unbiased estimator conditional on X).

Homoskedasticity

No autocorrelation

Normality and inference

In order to use the OLS std. errors to make inference, t statistics and F statistics, we need to add the assumption of normality.

Assumption 6: Normality

The errors u_t are independent of X and are independently and identically distributed as Normal (0, σ^2). Assumption 6 clearly implies previous assumptions (no autocorrelation, homoskedasticity, and zero conditional mean).

If also Assumption 6 holds, **then our OLS estimators are normally distributed** conditional on X. Under the null hypothesis, each t statistic has a t distribution, and each F statistic has a F distribution (**we can construct our confidence intervals**).

Main takeaways from Lecture 1

1. What is a time series? And the difference with cross sectional data?
2. What is a static model and what a FDL?
3. How do we calculate the impact multiplier? And how the long-run multiplier?
4. What are the five assumptions of the G-M theorem?
5. Under which assumptions is OLS unbiased? And under which is BLUE?
6. Can you define the five assumptions of the G-M theorem?
7. What other assumption do we need to be able to make inference?

Seminar 1

Exercise for the seminar

- The objective of the first seminar is that you get acquainted with an econometric software of your preference (I will provide R codes) and start running the first regressions.

- You can work in pairs.

- This exercise will be assessed, and you will have the chance to get points from it (a max of 5 points will be awarded).

- You will have the chance to deliver the exercise until next week.

- You have to deliver the replication codes and the report before Thursday 29^{th}, February 2024.

Exercise for the seminar

What do you have to do?

1. For those who do not have R installed, download it from https://cran.r-project.org/bin/windows/base/.

2. Install the package 'wooldridge' and go to example 10.2 on the following webpage: https://justinmshea.github.io/wooldridge/articles/Introductory-Econometrics-Examples.html

3. Replicate the exercise by yourself.

4. In addition to generating the baseline graphs and estimations, estimate the regression **tbill_model** and **updated_model** including 1 lag for the explanatory variables cpi and deficit.

Exercise for the seminar

Deliverables:

You should submit a code that allows replicating:

1. tbill_model and updated model (both with the summary for the regressions). 1 point

2. Both figures presented in Chapter 10, example 10.2. 1 point

3. Idem point 1 but performing both estimations including 1 lag for each explanatory variable. 2 points.

4. A brief report (1 page maximum) interpreting the coefficient in the regressions and mentioning some differences between the baseline estimation and the one including lags. Report the impact multiplier and the long-term impact multiplier. 1 point.

Exercise for the seminar

Some help with R:

- You might find useful the following packages: "wooldridge" ; "xts" ; "dplyr" ; "quantmod"
- Remember to install them and call them with Library!
- Remember that an easy way to create lags uses the package "dplyr“.
- Remember to bring the summary of the regression with summary().
- Do not hesitate to ask questions if there is something you do not understand (I am here to help).