Introduction to Time Series Econometrics

Romain Lafarguette, Ph.D.

Quant & IMF External Consultant

Singapore Training Institute, 08 November 2022

Outline

1. **Data concepts**: population, sample, data types, data generating process, etc.
2. **Estimation strategy**

NB: this slide-deck is heavily inspired by the excellent course of `ChristopheHurlin`https://sites.google.com/view/christophe-hurlin/teaching-resources

Overview

Financial econometrics (including time-series econometrics) are based on four main elements:

1. A sample of data
2. An econometric model, based on a theory or not
3. An estimation method to estimate the coefficients of the model
4. Inference/testing approach to validate the estimation

Population vs. Sample

Definition: Population

A **population** is defined as including all entities (e.g. banks or firms) or all the time periods of the processus that has to be explained

- In most cases, it is impossible to observe the entire statistical population, due to constraints (recording period, cost, etc.)

- A researcher would instead observe a **statistical sample** from the population. He will estimate an econometric model to understand the **properties on the population as a whole**.

Data Generating Process

Definition: Data Generating Process

A **Data Generating Process (DGP)** is a process in the real world that "generates" the data (or the sample) of interest

Example: Data Generating Process

Let us assume that there is a linear relationship between interest rates in two countries (R, R^*), their forward (F) and their spot exchange rate (S).

$$\frac{F}{S} = \frac{1+R}{1+R^*}$$

This non-arbitrage relationship (CIP) can be used in the foreign exchange market to determine the forward exchange rate

$$\mathbb{E}[\mathbb{F}|\mathbb{S} = \sim, \mathbb{R} = \diagdown, \mathbb{R}^* = \diagdown^*] = \sim * \frac{\nVdash + \diagdown}{\nVdash + \diagdown^*}$$

This relationship is the **Data Generating Process** for F

The equivalent of population for time series econometrics is the

Econometrics Challenge

The challenge of econometrics is to draw conclusions about a DGP (or population), after observing only one realization $\{x_1, \ldots X_N\}$ of a random sample (the dataset).

Data Types

In econometrics, sets can be mainly distinguished in three types:

1. Cross-sectional data
2. Time series data
3. Panel data

Cross-Sectional Data

Cross-sectional data are the most common type of data encountered in statistics and econometrics.

- Data at the entities level: banks, countries, individuals, households, etc.
- **No time dimension**: only one "wave" or multiple waves of different entities
- Order of data does not matter: no time structure

Time Series Data

Time series data are very common in financial econometrics and central banking. They entail specific estimation methods to do the **time-dependence**.

- Data for a single entity (person, bank, country, etc.) collected at multiple time periods. Repeated observations of the same variables (interest rate, GDP, prices, etc.)
- Order of data is important!
- The observations are typically not independent over time
- In this case, the notion of population corresponds to the **Data Generating Process (DGP)**

Panel Data

Also called longitudinal data. They contain the most information and allow for more complex estimation and analysis.

- Data for multiple entities (individuals, firms, countries, banks, etc.) in which outcomes and characteristics of each entity are observed at multiple points in time
- Combine cross-sectional and time-series information
- Present several advantages with respect to cross-sectional and time series data, depending on the topic at hands

Econometric Model

Definition: Econometric Model

An econometric model specificies the statistical relationship between different economic variables, that are expected to be stable over time

1. **Parametric model:** fully characterization of the relationship by a **set of parameters** θ and a **link function** f supposed to be known; the specification can be linear or non linear, and includes some randomness ϵ

$$Y = f(X; \theta) + \epsilon$$

2. **Non-parametric and semi-parametric models**: the link function can not be described using a finite number of parameters. The link function is assumed to be unknown and has to be estimated

Empirical Strategy

The general approach of (financial) econometrics is as follows:

1. Specification of the model

2. Estimation of the parameters

3. Diagnostic tests
   - Significance tests
   - Specification tests
   - Backtesting tests
   - etc.

4. Interpretation and use of the model

Random variable

- A **random variable** is a function $f : \Omega \mapsto \mathcal{R}$ that assigns to a set of outcome Ω a **value**, often a real number.
- The probability of an outcome is equal to its **measure** divided by the measure of all possible outcomes
  - Example: obtaining an even number by rolling a dice: $\{2, 4, 6\}$
  - Probability to obtain an even number by rolling a dice: $m(\{2, 4, 6\})/m(\{1, 2, 3, 4, 5, 6\}) = \frac{1}{2}$ *(here, the measure simply "counts" the outcomes with equal weights)*
- Random variables are the "building block" of statistics:
  - Random variables are characterized by their distribution (generating function, moments, quantiles, etc.)
  - The behavior of two or more random variables can be characterized by their dependence/independence, matrix of variance-covariance, joint distribution, etc.
  - The main theorem of statistics (law of large numbers, central limit theorem, etc.) leverages the properties of random variables

Stochastic Process

- A stochastic process is a sequence of random variables indexed by time (t):

$$\ldots, Y_1, Y_2, \ldots, Y_t, Y_{t+1}, \ldots =$$

Stationarity

Ergodicity

Moments

Estimator

Convergence

Biais

Efficiency