Time Series Econometrics, 2ST111

Lecture 1. Introduction and Overview

Yukai Yang

Department of Statistics, Uppsala University

Teachers

Yukai Yang

(YY) Floor 3, B322
yukai.yang@statistik.uu.se

Starting Point

Prerequisites

- knowledge about the probability theory and statistics
- knowledge of linear regression
- one programming language: R and RStudio

Broad Outline of the Course

- We use Hamilton's book "Time Series Analysis".
- For the outline, see the schedule.

Theory and Practice

The ultimate goal of the course is:

We want to develop the tools necessary
for analyzing relevant problems in real time series data.

Lectures and classes are:

- partly theoretical
  Deal with the mathematical structure of the models and explore the properties.
  Necessary for understanding the tools (and for being critical towards them!).
- partly empirical
  Feeling for real data. Hands-on experience.
  Promote an interest for doing empirical analyses.
  Introduce practical tools to perform analyses for e.g. MA theses.

The Exam

The ultimate goal of the exam is

to test whether you understand.

- the main results and the underlying intuition
- the tools and how they should be applied
- the details of specific econometric models

Today's lecture is about

- what time series econometrics is?
- overview
- readings, software and some resources

What is a time series?

A time series is a set of observations ordered by time.

Time series can be found in

- Economics (price indices, unemployment measurements, ...)
- Finance (stock prices, stock market indices, exchange rates, ...)
- Meteorology (temperature and precipitation records, ...)

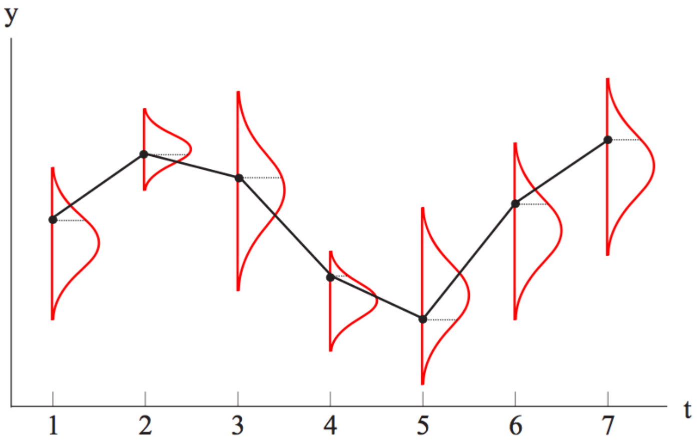
and in many other fields.

These observations are presumed equidistant over the time interval in most cases, but it is not always the case.

Time series data is a realization of a stochastic processes.

What is a time series?

Observation y_t is a realization of a random variable y_t.
Only one observation per random variable!



Example: S&P 500 stock market index

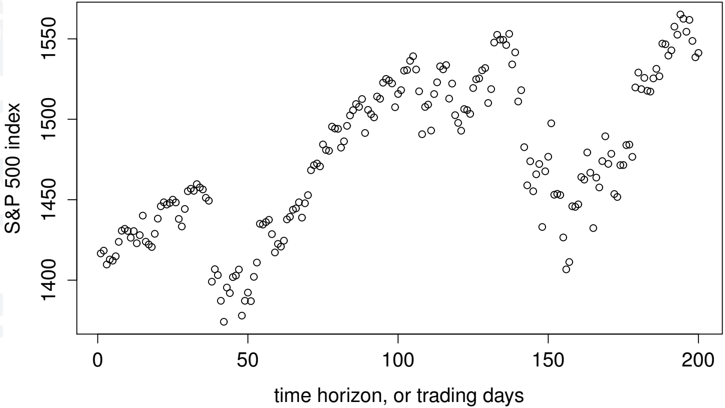
The sequence of the S&P 500 stock market index is a typical time series. It is

- a portfolio of 500 stocks in the US stock market.
- chosen by some committee based on the market capitalization, liquidity, industry grouping and some other factors.
- supposed to be a leading indicator for the US market equities, and one of the most commonly used benchmarks for it. These companies (500 stocks) form a representative of the industries in the US economy.

Data source: finance.yahoo.com

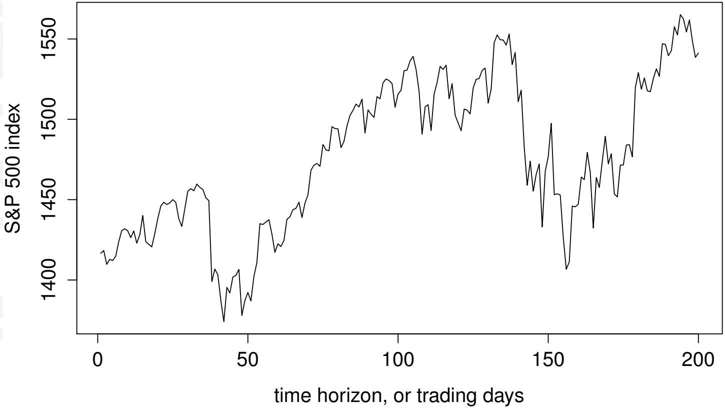
Example: S&P 500 stock market index

S&P daily closing price indices in circles, from 1 Jan to 17 Oct in 2007
not equidistant, 200 trading days



Example: S&P 500 stock market index

S&P daily closing price indices (time series plot), from 1 Jan to 17 Oct in 2007



Example: RW Model

- An example is the random walk (RW) model given by

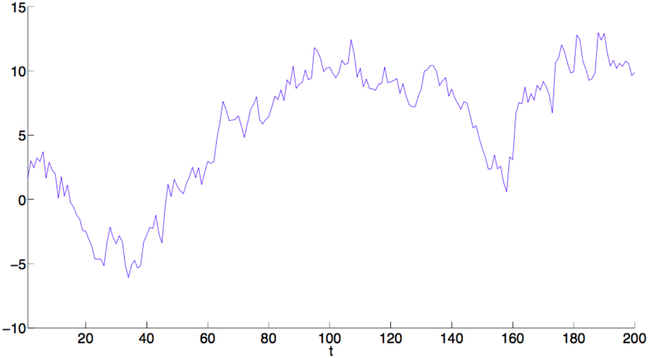
$$y_t = y_{t-1} + u_t \tag{1}$$

for $t = 1, 2, 3, \ldots$.

- The error term u_t are random variables assumed to be (mutually) independent, identically distributed (*i.i.d.*)
- u_t is independent of y_{t-1}, y_{t-2}, ..., y_0.
- u_t is normally assumed to be normally (Gaussian) distributed with mean zero.

RW Model (one sample path)

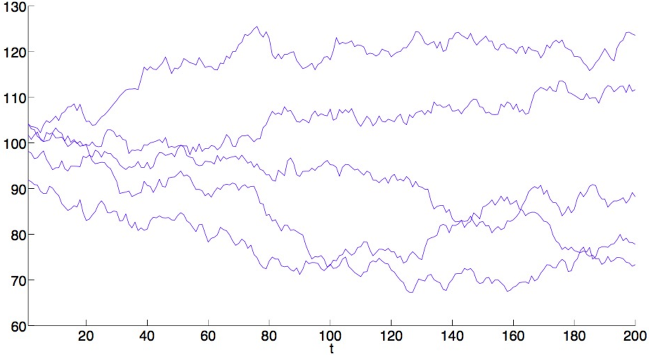
$$y_t = y_{t-1} + u_t$$



Sample path of a RW model with $u_t \overset{iid}{\sim} \mathcal{N}(0,1)$ and $y_0 = 0$.

RW Model (5 sample path)

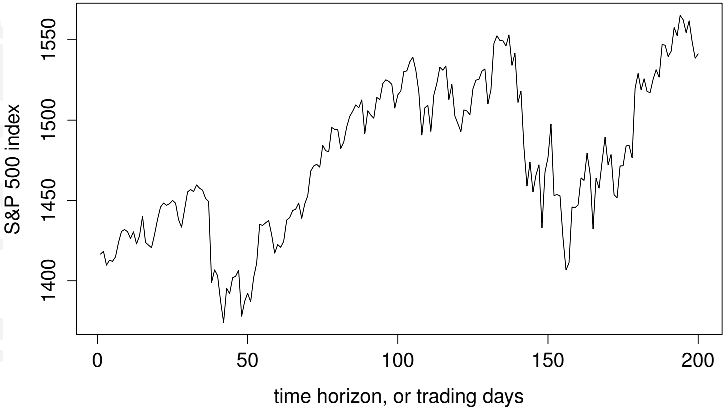
$$y_{i,t} = y_{i,t-1} + u_{i,t}, \quad i = 1, ..., 5$$



Sample paths of a RW model with $u_{i,t} \overset{iid}{\sim} \mathcal{N}(0,1)$ and $y_{i,0} \overset{iid}{\sim} \mathcal{N}(100,25)$.

Example: S&P 500 stock market index

Is this similar?



Time Series Analysis

- Time series analysis is important because it concerns the extraction of the useful information from the historical data.
- How to conduct the time series analysis given a particular time series?

  1. set up a hypothetical statistical model to represent the series in order to obtain insights into the mechanism that generates the data.
  2. once a satisfactory model has been formulated, to extrapolate from the model in order to anticipate (forecast) the future values of the time series.
  3. control future events via intervention

Example: RW Model

- Now we propose the RW model as the suitable one for daily S&P data.
- The best forecast of tomorrow's value is the current value. That is

$$E_t(y_{t+1}) = E_t(y_t) + E_t(u_{t+1}) = y_t \tag{2}$$

  as $E_t(u_{t+1}) = 0$, where $E_t(\cdot)$ is the conditional expectation given the information available at time t.
- A time series econometrician faces the task to construct models capable of forecasting, interpreting, and testing hypothesis concerning economic data.

Model Diagnostics

- Having selected a time series model, the parameters of the model need to be estimated and its goodness of fit to the data can be checked.
- Some fundamental assumptions have to checked as well, for example, no autocorrelation in residuals, no heteroskedasticity (implied by *i.i.d.* errors), and etc.
- If there are several suitable models for the data, we need to choose the best one based on some information criterion or criteria.
- If the model is satisfactory it may be used for forecasting.

Forecast Evaluation

- Once a time series has been analyzed and its future values have been forecast, it is reasonable to question how good the forecasts are. Typically, there will be several plausible models to extrapolate from in order to forecast the series.
- With forecasts from several models it is inevitable that the sample will show differences in forecast accuracy between the models.
- Because of this it is important to investigate how likely this outcome is due to pure chance, that is, whether the observed difference is statistically significant or not.

"Modelling Economic Series" by C.W.J. Granger

The basic objective is to affect the beliefs - and hence the behaviour - of other research workers and possibly other economic agents. These beliefs may be about the size or signs of certain coefficients (relating to a hypothesis from an economic theory), the quality of a forecasting technique or the relevance of some policy strategy, for example. The degree of belief, about the correctness of some hypothesis say, may be measured as a probability and can be affected by the outcome of an empirical investigation.

For example an economist may say that his or her belief about the statement 'inflation can be controlled by using a policy of money supply being attached to pre-announced slowly growing monetary targets is 0.4'. A carefully conducted empirical study may obtain results that change this economists' belief to 0.6 and consequentially change his or her behaviour. Of course, this prupose is not limited to econometric modelling and can also apply to a new economic theory at all levels of sophistication.

Nonlinear/Nonstationary Time Series Models

- Nonlinear and nonstationary time series models have gained more and more attention in the last two decades. The fact is that there are empirical evidences that many realistic time series are non-Gaussian and have a structure that change over time.
- For example, many economic time series are known to show a large number of nonlinear features such as cycles, asymmetries, jumps, thresholds, heteroskedasticity and combinations thereof, that additionally need to be taken into account.

Characteristics of Economic Time Series

Many economic time series do not have a constant mean, and most exhibit phases of relative tranquility followed by periods of high volatility.

Stylized facts:

- a clear trend
- a high degree of persistence, especially from the shocks
- varying volatility
- co-movements between series

The Window

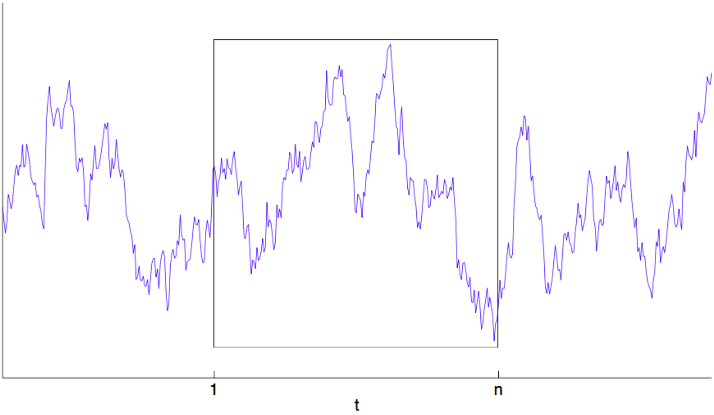
We think of the sample

$$\{y_t\}_{t=1}^{n} = y_1, y_2, \ldots, y_n$$

as a 'window' out of an infinite past and infinite future:

$$\{y_t\}_{t=-\infty}^{\infty} = \ldots, y_{-1}, y_0, \underbrace{y_1, y_2, \ldots, y_n}_{\text{sample}}, y_{n+1}, y_{n+2}, \ldots$$

The Window



Sample paths of a stochastic process $\{y_t\}_{t=-\infty}^{\infty}$.

LLN and CLT (iid case)

Let $y_1, \ldots, y_n$ be a sequence of *iid* random variables with finite mean and variance μ and σ^2, respectively, and let $\bar{y}_n = \frac{1}{n}\sum_{t=1}^{n} y_t$.

- Law of Large Numbers (LLN)

$$\bar{y}_n \xrightarrow{p} \mu \quad \text{as} \quad n \to \infty$$

- Central Limit Theorem (CLT)

$$\sqrt{n}\,\frac{\bar{y}_n - \mu}{\sigma} \xrightarrow{d} \mathcal{N}(0,1) \quad \text{as} \quad n \to \infty$$

The assumption of independence tends to be made rather casually, even though it is often inappropriate.

LLN and CLT for dependent random variables

Question:
Does the LLN and CLT apply when $y_1, ..., y_n$ are dependent?

Answer:
Yes, but under certain conditions.

Example: MA(1) Model

A simple time series model with temporal dependence is the moving average model of order 1 (MA(1)).

$$y_t = \theta u_{t-1} + u_t,$$

where $u_0, ..., u_n$ is a sequence of *iid* random variables with finite mean and variance μ_u and σ_u^2, respectively.

Thus, we have $\mu = E(y_t) = (\theta + 1)\mu_u$ and $Cov(y_t, y_{t-1}) = \theta\sigma_u^2$ for all integers t.

$y_1, ..., y_n$ are dependent random variables if $\theta \neq 0$.

Example: MA(1) Model

Moreover,

$$\bar{y}_n = \frac{1}{n}\sum_{t=1}^{n}(\theta u_{t-1} + u_t) = \theta \times \underbrace{\frac{1}{n}\sum_{t=1}^{n} u_{t-1}}_{\xrightarrow{p} \mu_u} + \underbrace{\frac{1}{n}\sum_{t=1}^{n} u_t}_{\xrightarrow{p} \mu_u} \xrightarrow{p} \mu$$

as $n \to \infty$. The LLN still applies.

It can be shown that

$$\sqrt{n}\,\frac{\bar{y}_n - \mu}{(\theta + 1)\sigma_u} \xrightarrow{d} \mathcal{N}(0, 1) \quad \text{as} \quad n \to \infty.$$

The CLT applies.

Example: MA(1) Model, Proof for the CLT

By defining $\bar{u}_n = \frac{1}{n}\sum_{t=1}^n u_t$, we have $\bar{y}_n = \theta(\bar{u}_n + \frac{1}{n}u_0 - \frac{1}{n}u_n) + \bar{u}_n$.

By rearranging, we have

$$\bar{y}_n = (\theta+1)\bar{u}_n + \underbrace{\frac{\theta}{n}(u_0 - u_n)}_{\xrightarrow{p} 0}.$$

Remember that $\sqrt{n}(\bar{u}_n - \mu_u)/\sigma_u \xrightarrow{d} \mathcal{N}(0,1)$.

It follows that

$$\sqrt{n}\,\frac{\bar{y}_n - \mu}{(\theta+1)\sigma_u} = \sqrt{n}\,\frac{\bar{u}_n - \mu_u}{\sigma_u} + \underbrace{\frac{\theta(u_0 - u_n)}{\sqrt{n}(\theta+1)\sigma_u}}_{\xrightarrow{p} 0} \xrightarrow{d} \mathcal{N}(0,1).$$

Central Question

How far can we relax the independence assumption of the classical theory and still preserve the validity of the LLN and CLT?

Recommended Readings

- Hayashi (2000), "Econometrics", Princeton University Press.
- Fuller (1995), "Introduction to Statistical Time Series", Wiley.
- Brockwell & Davis (2009), "Time Series: Theory and Methods", Springer.
- Lütkepohl (2010), "New Introduction to Multiple Time Series Analysis", Springer.
- White (2000), "Asymptotic Theory for Econometricians", Academic Press.

Web Resources

- MIT video courses in mathematics,
  `http://ocw.mit.edu/courses/most-visited-courses/`

To be continued! Thank you!