Time Series and Forecasting
Introduction

Maria Eduarda Silva
mesilva@fep.up.pt

School of Economics, University of Porto

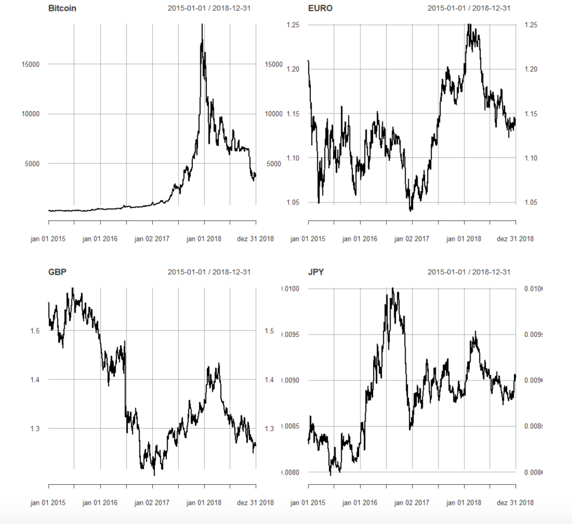
LMU, July 1 2022

What is a time series?

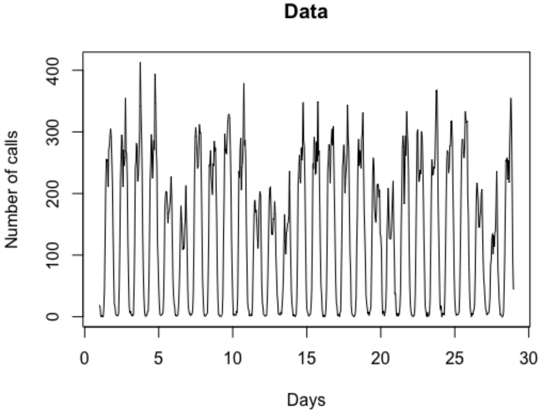
- *a damned thing following another (R.A. Fisher ?)*

- Observations made at regular intervals in time : economy, industry, finance, environment, etc

Criptocurrency vs usual currencies



Mobile calls



Time series data vs Cross-Section data

Types of data:

- Time series data collected at regular intervals over time
- Cross-sectional: data collected at a single point in time

	Fertility	Agriculture	Examination	Education	Catholic	Infant.Mortality
Courtelary	80.2	17.0	15	12	9.96	22.2
Delemont	83.1	45.1	6	9	84.84	22.2
Franches-Mnt	92.5	39.7	5	5	93.40	20.2
Moutier	85.8	36.5	12	7	33.77	20.3
Neuveville	76.9	43.5	17	15	5.16	20.6

Time series data vs Cross-Section data

- Classical statistics: independent observations.
- Classical statistics:
  - iid data $X_1, \ldots, X_n$

Time series data vs Cross-Section data

- Classical statistics: independent observations.
- Classical statistics:
  - iid data $X_1, \ldots, X_n$
  - sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and sample variance $s^2 = \frac{1}{(n-1)}\sum_{1}^{n}(x_i - \overline{x})^2$ capture univariate properties

Time series data vs Cross-Section data

- Classical statistics: independent observations.
- Classical statistics:
  - iid data $X_1, \ldots, X_n$
  - sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and sample variance $s^2 = \frac{1}{(n-1)}\sum_1^n (x_i - \overline{x})^2$ capture univariate properties
  - $\overline{X}$ and s^2 do not tell us how X_t and X_{t+k} are related.

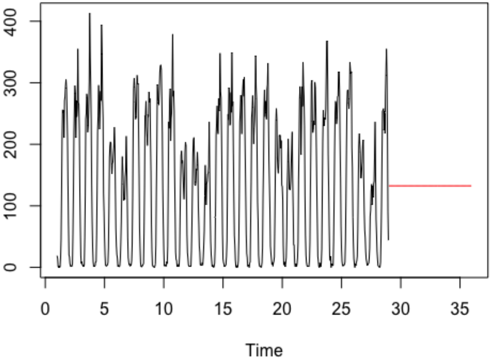
Time series data vs Cross-Section data

- Classical statistics: independent observations.
- Classical statistics:
  - iid data $X_1, \ldots, X_n$
  - sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and sample variance $s^2 = \frac{1}{(n-1)}\sum_1^n (x_i - \overline{x})^2$ capture univariate properties
  - $\overline{X}$ and s^2 do not tell us how X_t and X_{t+k} are related.
  - Saying that the series X_t is independent means that knowing previous values does not help to predict the next value.

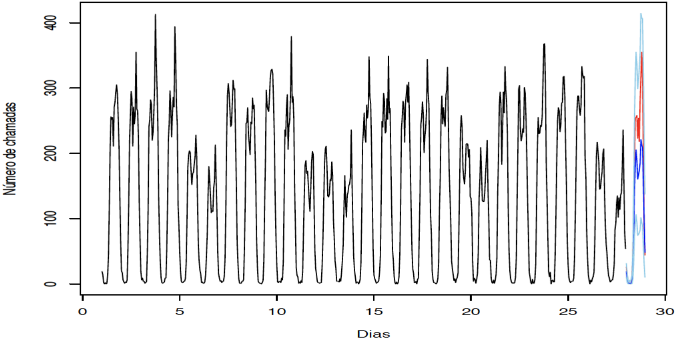
Time series data vs Cross-Section data

- Classical statistics: independent observations.
- Classical statistics:
  - iid data $X_1, \ldots, X_n$
  - sample mean $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ and sample variance $s^2 = \frac{1}{(n-1)}\sum_1^n (x_i - \overline{x})^2$ capture univariate properties
  - $\overline{X}$ and s^2 do not tell us how X_t and X_{t+k} are related.
  - Saying that the series X_t is independent means that knowing previous values does not help to predict the next value.
  - If $X_1, \ldots, X_N$, are known, the prediction for X_{N+1} is the average of all X_t's.

Forecasts for the number of calls?????



Forecasts for the number of calls!!!!



How to achive this result? Time series methods!

Why special methods to analyse time series?

- Time series data have a natural ordering- the time of observation
- Time series data are correlated data and that correlation is called serial correlation

Consequence: most of usual methods to analyse data are not appropriate for time series

Effects of serial correlation

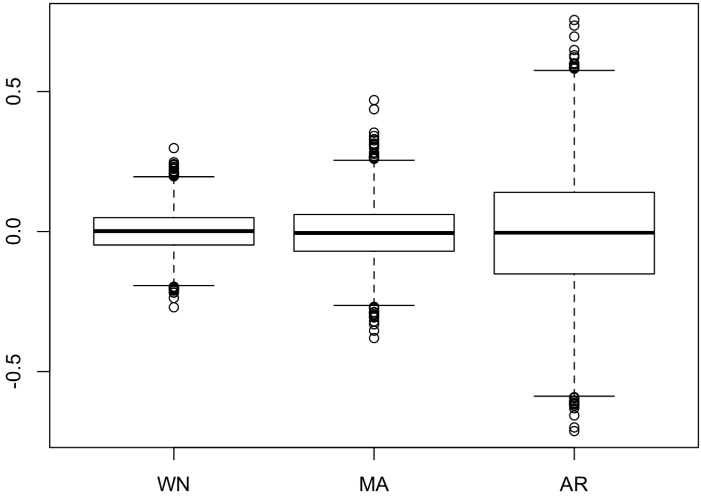
- Time series data are correlated data and that correlation is called serial correlation.
- Serial correlation affects the inferential procedures usual under iid settings.
- The simplest example is the variance of the sample mean.
- Recall that when we have $X_1, \ldots, X_n$ iid observations from a random variable X with $E(X) = \mu$ and $var(X) = \sigma^2$ the sample mean $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$
- When $X_1, \ldots, X_n$ are serially correlated $var(\overline{X}) > \frac{\sigma^2}{n}$

Effects of serial correlation: a simulation exercise

The plots below illustrate how serial correlation affects the variance of the sample mean $\overline{X}$. The boxplots represent the means of 5000 samples of size 1000 observations from

- iid N(0,5) - WN
- a moving-average model -MA- implying that adjacent observations are correlated
- an autoregressive model- AR- implying that observations are correlated with previous observations, although the correlation becomes negligible once observations become far apart.

Under iid the boxplot represents 5000 observations from $\overline{X} \sim N(0, 5/1000)$
Under serially correlated data the variance is inflated.


0.5
0.0
-0.5
WN
MA
AR

Effects of serial correlation: conclusion

If you use results obtained under iid to make inference for non iid data you under estimate the variance, that is your uncertainty and that must be avoided

Contents and References

Content

- Time series characteristics
- Linear, ARMA, models for time series
- Time series modelling
- Forecasting
- Multivariate Time Series

References

- Forecasting: Principles and Practice (3rd ed) . Rob J Hyndman and George Athanasopoulos `https://otexts.com/fpp3/`
- Time Series: A Data Analysis Approach Using R. Robert Shumway, David Stoffer, CRC press, 2019.