Specialist Courses

Daniele Zago

November 29, 2021

Contents

Part I

Time Series

This is a short course aimed at giving an introduction to time series models and their application to modelling and prediction of data collected sequentially in time. The aim is to provide specific techniques for handling data and at the same time to provide some understanding of the theoretical basis for the techniques. Topics covered will include univariate linear and non linear models (both in mean and variance) and some basics of spectral analysis.

Textbook references

Brockwell and Davis (2016)	Introduction to Time Series and Forecasting
Fan and Yao (2005)	Nonlinear Time Series: Nonparametric and Parametric Methods
Shumway and Stoffer (2017)	Time Series Analysis and Its Applications: With R Examples
Tsay (2013)	Multivariate Time Series Analysis: With R and Financial Applications
Wei (2019)	Multivariate Time Series Analysis and Applications
Brockwell (2009)	Time Series: Theory and Methods
Douc et al. (2014)	Nonlinear Time Series: Theory, Methods and Applications with R Examples

Course contents

- Introduction to linear time series models
- Introduction to spectral analysis
- Transfer function models
- Introduction to nonlinear time series models
  - Threshold AR models
  - Markov Switching models
  - Bilinear models
  - ARCH type models
- Long memory models
- INAR models

Lecture 1: Introduction to time series

2021-11-12

In general, in time series we are interested in *a)* understanding the stochastic mechanism that gives rise to an observed series and *b)* to forecast future values of a series based on the observed history. As this course is introductory, we will restrict our analysis to *univariate* time series.

Assumption We assume that future behaviour is equal to previous behaviour, i.e. we are able to forecast the future based on the observed past data.

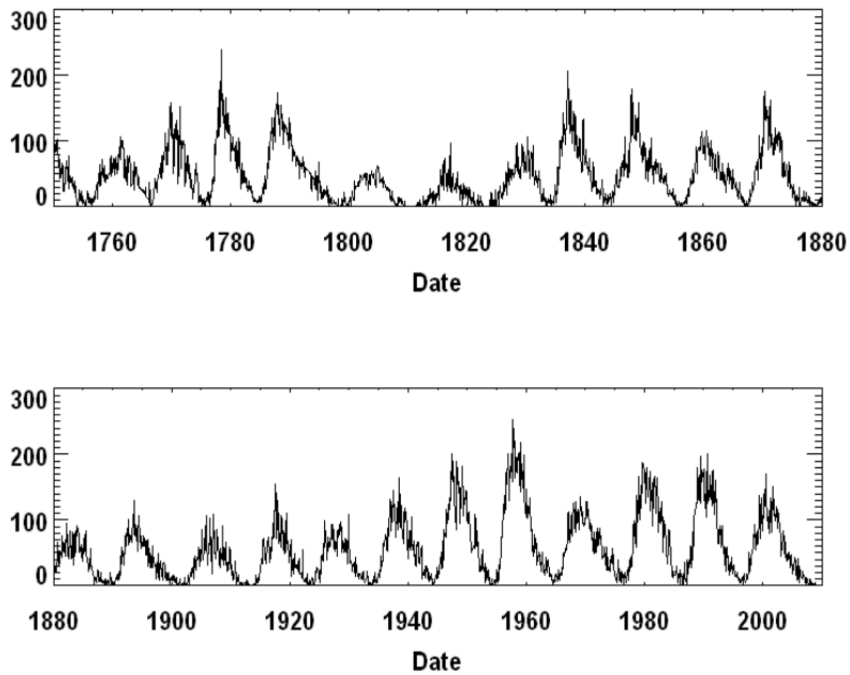

Figure 1: Linear time series models are not able to explain this behaviour, since cyclic components are unequal over time.

There are several approaches in modern time series, namely

› Classical approach

› Modern approach of Box and Jenkins.

› State-space approach of Durbin and Koopman (2012).

1.1 Classical approach

We assume a data-generating process given by the sum of different components

$$Y_t = f(t) + \varepsilon_t, \quad \varepsilon_t \sim \mathrm{WN}(0, \sigma_\varepsilon^2),$$

such that $\mathbb{E}[\varepsilon_t] = 0$, $\mathbb{V}[\varepsilon_t] = \sigma_\varepsilon^2$, $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$. Assuming different shapes of $f(t)$ lets us obtain different types of time series, i.e.

› Additive TREND + SEASONAL + CYCLES: $f(t) = T_t + S_t + C_t$

› Multiplicative TREND · SEASONAL · CYCLES: $f(t) = T_t \cdot S_t \cdot C_t$

The classical approach establishes that trend, seasonal, and cyclic components should be estimated separately with simple models and then combined. For example, we can use a linear model for the trend such as

$$T_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \ldots + \alpha_g t^g.$$

In order to model S_t we could use a dummy variables with sine/cosine transform to include cyclic behaviour.

Problem Empirical time series contain deterministic component and stochastic trend components, which cannot be modeled by stationary processes.

1.2 Modern approach

We can consider a DGP as a stochastic process, with time series seen as a realization from a stochastic process. To perform statistical inference, we need to assume that at least some features of the underlying probability law are *stationary* over the time period of interest.

Def. (Stochastic process)
A collection of random variables $X = (X_t)_t$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is called a ***stochastic process***

Remark A stochastic process is therefore a function of two arguments $X : \mathcal{T} \times \Omega \to X$, $(t, \omega) \mapsto X_t(\omega)$ and for a fixed value of ω we obtain a *path* from the stochatstic process.

Sample We only observe a portion of the infinite path of the stochastic process,

$$\ldots, X_{-t}, X_{-t-1}, \ldots, X_0, \underbrace{X_1, X_2, \ldots, X_t}_{x_1, x_2, \ldots, x_t}, \ldots,$$

therefore if we want to make inference over the DGP we must make some strong assumptions.

Def. (Mean function)
For a stochastic process X_t, the ***mean function*** is

$$\mu_t = \mathbb{E}[X_t] \quad \text{for } t \in \mathcal{T}.$$

Def. (Autocovariance function)
For a stochastic process X_t, the ***autocovariance function*** is

$$\gamma_{t,s} = \mathrm{Cov}(X_t, X_s) = \mathbb{E}[(X_t - \mu_t)(X_s - \mu_s)], \quad \text{for } t, s = 0, \pm 1, \pm 2, \ldots$$

The autocorrelation function is then defined as

$$\operatorname{Cor}(X_t, X_s) = \frac{\operatorname{Cov}(X_t, X_s)}{\sqrt{\mathbb{V}[X_t]}\sqrt{\mathbb{V}[X_s]}}.$$

We need to make some strong assumptions on the structure of the process in order to make inference possible.

Def. (Strong stationarity)
A process $(X_t)_t$ is ***strictly stationary*** if it is invariant under time shifts, i.e. if

$$(X_{t_1}, \dots, X_{t_n}) \stackrel{\mathrm{d}}{=} (X_{t_1+k}, \dots, X_{t_n+k})$$

for any $n \geq 1$, any choice of $t_1, \dots, t_n$ and al time shifts $k \in \mathbb{Z}$.

Marginals Choosing for instance $n = 1$ means that the marginal distribution of X_t the same as that of X_{t-k} for all t and k.

Def. (Weak stationarity)
A process $(X_t)_t$ is ***weakly stationary*** if

1. $\mathbb{E}[X_t] = \mu < \infty$ for all t.
2. $\mathbb{V}[X_t] = \sigma^2 < \infty$ for all t.
3. $\operatorname{Cov}(X_t, X_{t-k}) = \gamma(k)$ is independent of t for each k.

Weaker Rather than imposing conditions on all possible distributions, we impose conditions only on the first two moments of the series.

Implications

› Strong stationarity + $\mathbb{E}[X_t]^2 < \infty \implies$ weak stationarity.

› Weak stationarity $\not\Longrightarrow$ strong stationarity.

› Weak stationarity + Gaussian $\implies$ Strong stationarity.

Example (Random walk)
For a random walk $Y_t = Y_{t-1} + \varepsilon_t$, we have that $\mathbb{V}[Y_t] = t\sigma^2$ and the process is therefore non-stationary.

Since our objective is to find a model which is able to take into account the ***linear*** dependence between the observations, two very important functions are the autocorrelation and autocovariance functions.

Since for a stationary time series X_t we have $\mathrm{Cov}(X_t, X_{t-k}) = \gamma(k)$ for all k, we can therefore define the ACF as

$$\rho(k) = \frac{\gamma(k)}{\gamma(0)}, \quad k = 0, \pm 1, \pm 2, \dots$$

from which we can see that γ and ρ are even functions, namely

$$\gamma(-k) = \gamma(k), \quad \rho(-k) = \rho(k).$$

Def. (Sample autocorrelation function)
We define the ***sample autocorrelation function*** (ACF) as

$$\widehat{\rho}(k) = \frac{\widehat{\gamma}(k)}{\widehat{\gamma}(0)},$$

where

$$\widehat{\gamma}(k) = \frac{1}{n} \sum_{t=1}^{n-|k|} (X_t - \overline{X})(X_{t+|k|} - \overline{X}).$$

Bias Even if this estimator is biased in finite samples, this is preferred to the unbiased estimator since when dividing by n we have a nonnegative-definite estimator.

When fitting a model that relies on correlation between X_t and X_{t-k}, conditioned on immediate variables. We will only include a further lagged variable X_{t-k} in the model for X_t if X_{t-k} makes a genuine and additional contribution to X_t in addition to those from $X_{t-1}, \dots, X_{t-k+1}$. We measure this dependence using the *partial autocorrelation*.

Def. (Partial autocorrelation)
The ***partial autocorrelation*** at lag k is the autocorrelation between z_t and z_{t+k} with the linear dependence of z_t on $z_{t+1}, \dots, z_{t+k-1}$ removed. Namely,

$$\begin{cases} \alpha(1) = \mathrm{Cor}(z_{t+1}, z_t) \\ \alpha(k) = \mathrm{Cor}\left(z_{t+k} - \pi_{t,k}(z_{t+k}), z_t - \pi_{t,k}(z_t)\right) \quad \text{if } k \geq 2 \end{cases}$$

where $\pi_{t,k}(x)$ is the orthogonal projection (regression) of x onto $z_{t+1}, \dots, z_{t+k-1}$.

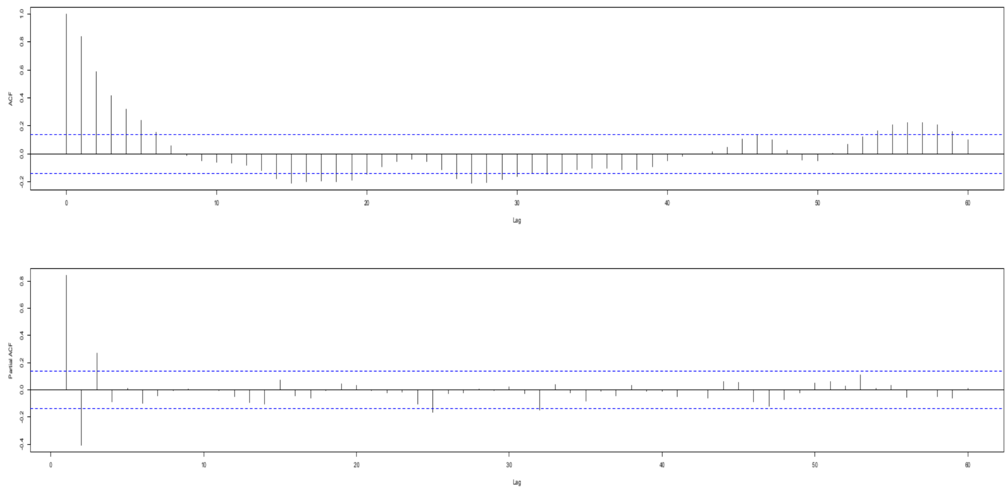

Figure 2: Autocorrelation (top) and partial autocorrelation (bottom) for a simulated time series.

The ACF and PACF are the main instruments that we use for choosing the most appropriate model for the DGP under the modern approach to time series (Box-Jenkins procedure).

Lecture 2

2021-11-29

2.1 White noise process

It serves as a building block in defining more complex linear time series processes and reflects information that is not directly observable. I It is easy to see that a sequence of independent and identically distributed (i.i.d.) random variables with mean 0 and finite variance σ^2 is a special white noise process.
In general it's convenient to write a stochastic process as a sum of white noise terms, since it's easier to prove theorems related to them.
The probability behavior (law) of a stochastic process is completely determined by all of its finite-dimensional distributions.
When all of the finite-dimensional distributions are Gaussian (normal), the process is called a Gaussian process.
Since uncorrelated normal random variables are also independent, a Gaussian WN process is, in fact, a sequence of i.i.d. normal random variables.

2.2 Random walk

Whereas the white noise is a simple process with no memory, the random walk has infinite memory and is nonstationary,

$$X_t = \mu + X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma^2),$$

where $X_0 = 0$ by convention. The process is such that by recursion,

$$\begin{aligned} X_t &= \mu + (\mu + X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \dots \\ &= \underbrace{t\mu}_{\text{drift}} + \underbrace{\sum_{i=1}^{t} \varepsilon_i}_{\text{stoch. trend}}. \end{aligned}$$

The last sum is called ***stochastic trend***, since every error ε enters with the same error even from past observations. Moreover, $\mathbb{E}[X_t] = t\mu$ and $\mathbb{V}[X_t] = t\sigma^2$. Applying the first-difference operator to the process yields

$$(1 - B)X_t = X_t - X_{t-1} = \mu + \varepsilon_t,$$

which is a stationary model.

2.3 Linear time series

We introduce the ARMA model, the most famous type of linear time-series model which is used even for non-linear data. Forecasts from these models in these case have been empirically shown to be more accurate than forecasts from more complicated models.

A general linear process is of the form

$$X_t = \varepsilon_t + \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i},$$

where $\sum_{i=1}^{\infty} \psi_i^2 < \infty$. This type of process is such that

i. $\mathbb{E}[X_t] = 0$ for each t

ii. $\mathrm{Cov}(X_t, t_{t-k}) = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$ for $k \geq 0$ and $\psi_0 = 1$.

An example is when the weights are an exponentially decaying sequence $\psi_j = \varphi^j$ with $|\varphi| < 1$, and in this case

$$X_t = \varepsilon_t + \varphi\varepsilon_{t-1} + \varphi^2\varepsilon_{t-2} + \ldots$$

The variance of this process can be written as a geometric series

$$\mathbb{V}[X_t] = \sigma_\varepsilon^2 \cdot \sum_{i=0}^{\infty} \varphi^k = \frac{\sigma_\varepsilon^2}{1-\varphi^2},$$

moreover, the covariance functions are

$$\mathrm{Cov}(X_t, X_{t-k}) = \frac{\varphi^k \sigma_\varepsilon^2}{1-\varphi^2}$$

$$\mathrm{Cor}(X_t, X_{t-k}) = \varphi^k$$

A moving average model of the form MA(q) =

$$X_t = \vartheta_1 \varepsilon_{t-1} + \ldots + \vartheta_q \varepsilon_{t-q} + \varepsilon_t.$$

These models are easily tractable since they are stationary by definition and
An autoregressive model AR is such that

$$X_t = c + \varphi_1 X_1 + \ldots + \varphi_p X_{t-p} + \varepsilon_t,$$

where X_t's could also be random variables each uncorrelated with the next value X_{t+1}. The expected value of the process can be calculated in terms of the autoregressive coefficients, by assuming the process to be stationary

$$\mathbb{E}[X_t] = \mathbb{E}[c + \varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p} + \varepsilon_t] \implies \mathbb{E}[X_t] = \frac{c}{1 - \sum_{i=1}^{p} X_{t-i}}.$$

Again, by assuming the process to be stationary we observe an autocovariance of the form

$$\gamma_k = \begin{cases} \varphi_1\gamma_1 + \varphi_2\gamma_2 + \ldots + \varphi\gamma_p + \sigma_\varepsilon^2 & k = 0 \\ \varphi_1\gamma_1 + \varphi_2\gamma_2 + \ldots + \varphi\gamma_p & k > 0 \end{cases}$$

$$\rho_k = \varphi_1\rho_{k-1} + \varphi_2\rho_{k-2} + \ldots + \varphi_p\rho_{k-p}, \quad k > 0,$$

which yield the Yule-Walker equations when considering them for $k = 1, \dots, p$. These equations can be used to compute the model coefficients when solving them in terms of the unknown $\boldsymbol{\varphi}$ and sample autocorrelation $\widehat{\rho}_k$.
We have that the sample autocorrelation is exponentially decaying in k, depending on the model parameters.

Its partial autocorrelation function is null for $k > p$.

2.4 ARMA model

We introduce the combined ARMA model in order to model more complicated dynamics of time series, yielding the ARMA(p, q) defined as

$$X_t = \varphi_1 X_{t-1} + \ldots + \varphi_p X_{t-p} + \vartheta_1 \varepsilon_{t-1} + \ldots + \vartheta_q \varepsilon_{t-q} + \varepsilon_t,$$

which usually allows us to model correlation structures with a smaller number of parameters.

Using a backshift operator $B^k X_t = X_{t-k}$ we can write this model as

$$\varphi(B) X_t = \vartheta(B) \varepsilon_t,$$

where the polynomials in B are defined as

$$\varphi(B) = 1 - \varphi_1 B - \ldots - \varphi_p B^p$$
$$\vartheta(B) = 1 + \vartheta B + \ldots + \vartheta_q B^q$$

and are extremely important since we can determine the properties of an ARMA model in terms of $\vartheta(\cdot)$ and $\varphi(\cdot)$. Moreover, if the ARMA model is stationary, we can write an AR(∞) representation as

$$\vartheta(B)^{-1} \varphi(B) X_t = \varepsilon_t,$$

and a $MA(\infty)$ representation as

$$X_t = \varphi(B)^{-1} \vartheta(B) \varepsilon_t.$$

Example (AR(1))
Consider the AR(1) model, then

$$\begin{aligned} Y_t &= \varphi Y_{t-1} + \varepsilon_t \\ &= \varphi(\varphi Y_{t-2} + \varepsilon_{t+1}) + \varepsilon_t \\ &= \ldots \\ &= \varepsilon_t + \varphi \varepsilon_{t-1} + \varphi^2 \varepsilon_{t-2} + \ldots \end{aligned}$$

Example (General procedure)
We write the relationship

$$(1-\varphi B)y_t = (1-\vartheta B)\varepsilon_t,$$

for which we can write

$$y_t = \frac{1-\vartheta B}{1-\varphi B}\varepsilon_t.$$

Our goal now is to obtain a relationship of the form

$$Y_t = \Psi(B)\varepsilon_t = \sum_{i=0}^{\infty}\psi_i B^i,$$

and therefore $\Psi(B) = \frac{1-\vartheta B}{1-\varphi B}$, from which

$$(1-\varphi)\Psi(B) = 1-\vartheta B$$

$$\Updownarrow$$

$$(1-\varphi B)(1+\psi_2 B+\psi_2 B^2+\ldots) = 1-\vartheta B$$

$$\Updownarrow$$

Example (General AR(∞))
The procedure is the same, except we now have to find a relationship of the form

$$\varphi(B)Y_t = \vartheta(B)\varepsilon_t \longrightarrow$$

$$\frac{\varphi(B)}{\vartheta(B)} = \Xi(B),$$

and find the parameters in terms of φ and ϑ.

In order to check for invertibility of the process, we need to invert the MA operator which is doable if the characteristic equation $\vartheta(B) = 0$ has solutions $|B_i| > 1$.
In order to check for of the process, we need to invert the AR operator which is doable if the characteristic equation $\vartheta(B) = 0$ has solutions $|B_i| > 1$.

Time series models only work if the data is stationary, therefore in general it's recommended to check for the evidence of trend or seasonality before applying an ARMA model. We can remove nonstationarity either via regression or via simple differentiation.

Even though the model might not be invertible, it's still better to have it stationary and not invertible. In general, it's advised to differentiate the series rather than risking for the time series to be nonstationary.

Testing whether the trend is deterministic or stochastic can be performed via a unit root test, which is usually not very powerful.

We obtain the ARIMA(p, d, q) class of models by applying a d-order to X_t and modeling the result as an ARMA(p, q) model, i.e.

$$\varphi(B)(1-B)^d Y_t = \vartheta(B)\varepsilon_t.$$

- In general, this process is simply an ARMA$(p+d, q)$ with d unit roots in the autoregressive polynomial.
- In general, we don't see time series such that $d > 2$.

References

Brockwell, P. J. (2009). *Time Series: Theory and Methods*. 2° edizione. New York, N.Y: Springer.

Brockwell, P. J. and Davis, R. A. (2016). *Introduction to Time Series and Forecasting*. Third. Springer Texts in Statistics. New York: Springer-Verlag.

Douc, R. et al. (2014). *Nonlinear Time Series: Theory, Methods and Applications with R Examples*. 1° edizione. Boca Raton: Chapman and Hall/CRC.

Durbin, T. l. J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. 2 edizione. Oxford: OUP Oxford.

Fan, J. and Yao, Q. (2005). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.

Shumway, R. H. and Stoffer, D. S. (2017). *Time Series Analysis and Its Applications: With R Examples*. 4th ed. New York, NY: Springer.

Tsay, R. S. (2013). *Multivariate Time Series Analysis: With R and Financial Applications*. 1. edizione. Hoboken, New Jersey: John Wiley & Sons Inc.

Wei, W. W. S. (2019). *Multivariate Time Series Analysis and Applications*. 1. edizione. Hoboken, NJ: John Wiley & Sons Inc.