Specialist Courses

Daniele Zago

February 7, 2022

Contents

Part I

Time Series

INSTRUCTOR: prof. Luisa Bisaglia

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. Finally, we will cover some aspects of long-memory and integer autoregressive models for count data.

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 information about the observed past data.

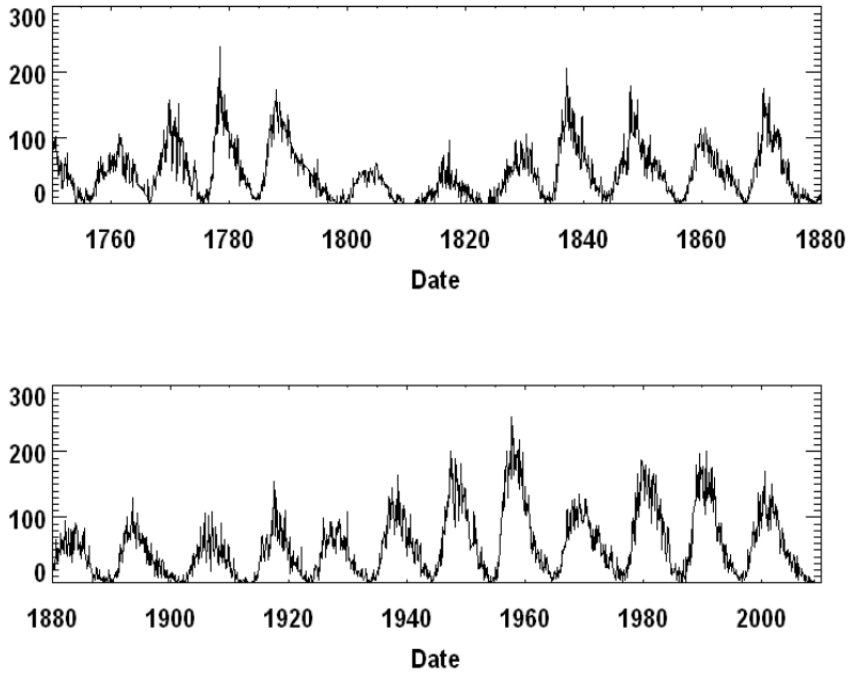

Figure 1: Classical linear time series models are not able to explain this behaviour, since cyclic components are not constant in amplitude over time.

There are several approaches in modern time series, namely

- *Classical approach*: trend + cycle + seasonality.
- *Modern approach*: Box and Jenkins procedure with ARIMA models.
- *State-space approach*: follows Durbin and Koopman (2012), we will not treat it here.

1.1 Classical approach

We assume a data-generating process given by a basic deterministic function of time plus additive noise,

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

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

- Additive: TREND + SEASONALITY + CYCLES: $f(t) = T_t + S_t + C_t$

› *Multiplicative*: TREND · SEASONALITY · 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

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

On the other hand, in order to model S_t we could use dummy variables with sine/cosine transform to promote cyclic behaviour.

Problem Empirical time series contain both deterministic trends and stochastic trends, which cannot be modeled by stationary processes.

$$\text{DETERMINISTIC TREND} \quad \mathbb{E}[X_t] = f(t)$$

$$\text{STOCHASTIC TREND} \quad \sum_{i=1}^{t} \varepsilon_t$$

1.2 Modern approach

We can consider the data-generating process (DGP) as a stochastic process which yields the observed time series as a sample path over time. 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

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

The autocorrelation function is then defined as

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

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

$$(X_{t_1}, \dots, X_{t_n}) \overset{\text{d}}{=} (X_{t_1+k}, \dots, X_{t_n+k})$$

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. $\text{Cov}(X_t, X_{t-k}) = \gamma(k)$ 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 $\text{Cov}(X_t, X_{t-k}) = \gamma(k)$ 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.

In addition to autocorrelation, we also consider the correlation between X_t and X_{t+k} after controlling for the effect of the intermediate values $X_{t+1}, \dots, X_{t+k-1}$ using a linear regression (projection). We call this dependence the *partial autocorrelation* of X.

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,

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

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

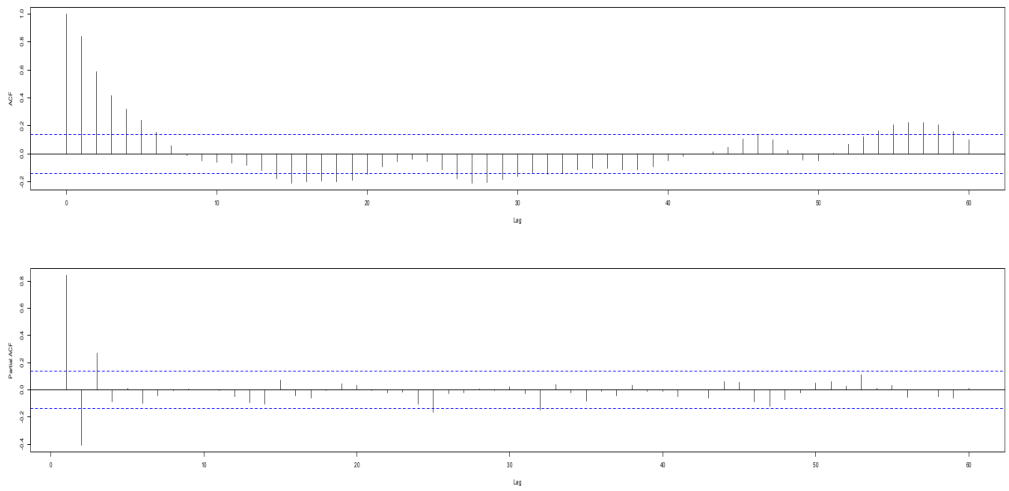

Figure 2: Autocorrelation (top) and partial autocorrelation (bottom) for a simulated time series.

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: Stochastic processes in time-series analysis

2021-11-29

In this lecture we review some of the fundamental processes used in time-series analysis, starting from the simplest process (white noise) and then moving towards standard but more complicated construction (ARIMA).

2.1 White noise process

The white-noise process serves as the building block for defining more complex linear time series processes and reflects information that is not directly observable. In general, any sequence X_t of i.i.d random variables such that $\mathbb{E}[X_t] = 0$ and $\mathbb{V}[X_t] = \sigma^2 < \infty$ is a white noise process.

In general it's convenient to write a stochastic process as a sum of white noise terms,

$$X_t = \sum_{j=1}^{\infty} \psi_j \varepsilon_j,$$

since we can leverage standard proof techniques to prove theorems related to the process behaviour.

In the white-noise case, the probability behavior (law) of X is completely determined by all of its finite-dimensional distributions. When all of the finite-dimensional distributions are Gaussian, the process is called a Gaussian process.

Since uncorrelated normal random variables are also independent, a Gaussian white-noise 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 weight both from new and from past observations. Moreover, $\mathbb{E}[X_t] = t\mu$ and $\mathbb{V}[X_t] = t\sigma^2$. Applying the first-difference operator $(1 - B)$, where B is such that $BX_t = X_{t-1}$, 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

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

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 and correlation functions are

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

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

A moving average model of the form $\mathrm{MA}(q)$ is the above model truncated to the first q components:

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

These models are easily tractable since they are stationary by definition and estimation is very simple in the Gaussian case.

On the other hand, an autoregressive model AR is such that

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

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

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

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 $k = 1, \ldots, p$. These equations can be used to compute the model coefficients when solving them in terms of the unknown $\boldsymbol{\varphi}$ and sample autocorrelation $\widehat{\rho}_k$.

For an AR model we have that the sample autocorrelation is exponentially decaying in k, depending on the model parameters, and 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

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

which usually allows us to model more complicated correlation structures using 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 (1) is stationary, we can find an AR(∞) representation for it by solving the equality

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

and a $MA(\infty)$ representation by solving

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

Example (AR(1))

Consider the AR(1) model, then

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

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

$$\begin{aligned} (1 - \varphi)\Psi(B) &= 1 - \vartheta B \\ &\Updownarrow \\ (1 - \varphi B)(1 + \psi_2 B + \psi_2 B^2 + \dots) &= 1 - \vartheta B \\ &\Updownarrow \end{aligned}$$

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 $d > 2$.

Lecture 3: Transfer function models

2021-12-06

Theorem 1 (Wold decomposition)
Let $(X_t)_t$ be a non-deterministic stationary time series with $\mathbb{E}[X_t] = 0$, then

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} + V_t,$$

where V_t is deterministic and

1. *$\psi_0 = 1$ and $\sum_{j=1}^{\infty} \psi_j^2 < \infty$.*
2. *$a_t = WN(0, \sigma^2)$.*
3. *$\mathbb{E}[a_t a_s] = 0$ for all $s, t = 0, \pm 1, \pm 2, \ldots$*

With this decomposition we can approximate any stationary time series using a linear process of the form

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j},$$

where $\sum_{j=1}^{\infty} |\psi_j| < \infty$.

3.1 Transfer function models

Transfer function models are models where an output series y_t is related to one or more input series x_t. We link the two series by the following ***transfer function model*** (TFM)

$$y_i = \nu(B)x_t + \eta_t, \tag{2}$$

where the relationship is linear, $\nu(B) = \sum_{j=-\infty}^{\infty} \nu_j B^j$, and is called the *transfer function of the linear filter* that transform x_t into y_t. η_t is a noise series independent of x_t. The weights ν_j are called ***impulse response weights*** and the TFM is called ***stable*** if

$$\sum_{j=-\infty}^{\infty} |\nu_j| < \infty,$$

and in particular this yields a BIBO (Bounded Input Bounded Output) relationship. The TFM is said to be ***causal*** if $\nu_j = 0$ for $j < 0$, since the present output is affected only by the system current and past values,

$$y_t = \sum_{j=0}^{\infty} \nu_j B^j.$$

The purpose of TF models is to identify the TF $\nu(B)$ and the noise model, possibly using a simpler representation which is similar to an ARIMA model

$$\delta(B)y_t = \omega(B)B^b x_t,$$

where

$$\delta(B) = 1 - \delta_1 B - \delta_2 B^2 - \ldots \delta_r B^r$$

$$\omega(B) = \omega_0 - \omega_1 B - \ldots - \omega_s B^s$$

and b is a delay parameter that tells us the lag that elapses before the impulse of the input variable produces an effect on the output variable.

With the above representation, we can rearrange the terms so that y_t has an explicit representation in terms of x_t,

$$y_t = \frac{\omega(B)}{\delta(B)} x_{t-b} + \eta_t, \tag{3}$$

and by equating Equation (3) to Equation (2) we can write the transfer function $\nu(B)$ as

$$\nu(B) = \frac{\omega(B)B^b}{\delta(B)}. \tag{4}$$

and the orders s, r, b of the model in Equation (3) can be found by equating the coefficients of B^j to both sides in Equation (4)

$$\delta(B)\nu(B) = \omega(B)B^b,$$

which yields the following equation

$$(1 - \delta_1 B - \delta_2 B^2 - \ldots - \delta_r B^r)(\nu_0 + \nu_1 B + \ldots) = (\omega - \omega_1 B - \ldots - \omega_s B^s)B^b,$$

and we obtain the following relationships between the components of the model

$$\begin{aligned}
&\nu_j = 0 && \text{if } j < b \\
&\nu_j = \delta_1 \nu_{j-1} + \delta_2 \nu_{j-2} + \ldots + \delta_r \nu_{j-r} + \omega_0 && \text{if } j = b \\
&\nu_j = \delta_1 \nu_{j-1} + \delta_2 \nu_{j-2} + \ldots + \delta_r \nu_{j-r} - \omega_{j-b} && \text{if } j = b+1, \ldots, b+s \\
&\nu_j = \delta_1 \nu_{j-1} + \delta_2 \nu_{j-2} + \ldots + \delta_r \nu_{j-r} && \text{if } j > b+s
\end{aligned}$$

By observing the behaviour of the cross-correlation function between x_t and y_t – similarly to what we do with ACF and PACF for estimating p, d, q in an ARIMA model) – we can find the appropriate values of s, r, b.

Def. (Cross-correlation function)
We say that X_t and Y_t are ***jointly stationary*** if they are univariate stationary and $\text{Cov}(X_t, Y_s) = f(|s-t|)$, and in this case we define the ***cross-correlation function*** between X_t and Y_t as the function

$$\gamma_{XY}(k) = \mathbb{E}[(X_t - \mu_X)(Y_{t+k} - \mu_Y)],$$

Marginals By definition we have that $\rho_{XX}(k) = \rho_X(k)$.

Symmetry It's relevant the order in which we compute the cross-correlation function, since unlike the ACF the CCF is not symmetric around the origin,

$$\rho_{XY}(k) \neq \rho_{XY}(-k),$$

instead we have that

$$\rho_{XY}(k) \neq \rho_{YX}(-k).$$

However, we have a way of obtaining the direction of association between the time series by inspecting the graph of the ACF. The direction depends on the software implementation of the function.

Example (AR(1) model)
Let $Y_t \sim \text{AR}(1)$, then we have $(1-\varphi B)Y_t = X_t$ and for time $t+k$ we can write

$$Y_{t+k} = \frac{1}{1-\varphi B}X_{t+k} = X_{t+k} + \varphi X_{t+k-1} + \varphi^2 X_{t+k-2} + \dots,$$

therefore the cross-covariance function between X_t and Y_t are

$$\gamma_{XY}(k) = \mathbb{E}[X_t Y_{t+k}] = \begin{cases} \varphi^k \sigma_k^2 & \text{if } k \geq 0 \\ 0 & \text{if } k \leq 0 \end{cases}$$

ARMA model In general, the ARMA(p,q) model can be written as a transfer function model without the white noise term η_t, and where X_t is a white noise itself uncorrelated with Y_t.

3.2 Cross-correlation function and TF models

Let x_t and y_t be stationary series with $\mu_x = \mu_y = 0$, then the transfer function at time $t+k$ is

$$y_{t+k} = \nu_0 x_{t+k} + \nu_1 x_{t+k-1} + \nu_2 x_{t+k-2} + \ldots + \eta_{t+k},$$

therefore if we multiply both left and right by x_t and take expectations we have

$$\gamma_{xy}(k) = \nu_0 \gamma_x(k) + \nu_1 \gamma_x(k-1) + \nu_2 \gamma_x(k-2) + \dots,$$

hence the CCF in the doubly stationary case has the following simple representation:

$$\rho_{xy}(k) = \frac{\sigma_x}{\sigma_y}\left[\nu_0 \rho_x(k) + \nu_1 \rho_x(k-1) + \nu_2 \rho_x(k-2) + \ldots\right]. \tag{5}$$

Therefore, by Equation (5) we observe that the relationship between the CCF and IRF ν_j is contaminated by the fact that they are not white noise, and therefore display the correlations at previous times. However, for a ***white noise model*** x_t we would see $\rho_x(k) = 0$ for all $k \neq 0$ and therefore we would have a direct way of estimating ν_k by letting

$$\gamma_{xy}(k) = \nu_k \sigma_k^2,$$

hence we can estimate the covariance function nd obtain an impulse response function which is directly proportional to the CCF,

$$\rho_{xy}(k) = \frac{\sigma_x}{\sigma_y}\nu_k \implies \nu_k = \frac{\sigma_y}{\sigma_x}\rho_{xy}(k). \tag{6}$$

Idea Therefore, our goal for estimating a TF model is to reduce the problem to a whitened series for x_t, and then apply the estimation procedure above.

In the general TF model given by

$$y_t = \nu(B)x_t + \eta_t,$$

if we assume $x_t \sim \text{ARMA}(p, q)$ we can calculate the ***pre-whitened input series***

$$\alpha_t = \frac{\varphi_x(B)}{\vartheta_x(B)}x_t,$$

and applying this transformation to both y_t and η_t we can obtain the ***filtered series***

$$\begin{cases} \beta_t = \frac{\varphi_x(B)}{\vartheta_x(B)}y_t \\ \varepsilon_t = \frac{\varphi_x(B)}{\vartheta_x(B)}\eta_t \end{cases}$$

Finally, the TF model becomes

$$\beta_t = \nu(B)\alpha_t + \varepsilon_t,$$

where the input series is $\alpha_t \sim \text{WN}(0, \sigma^2)$ and we can estimate the transfer function using Equation (6) between β_t and α_t.

3.2.1 General procedure for the identification of a TF model

1. Identify an $\text{ARMA}(p, q)$ model for the input x_t,

$$\varphi_x(B)x_t = \vartheta_x(B)\alpha_t$$

2. Prewhiten $x_t \to \alpha_t = \frac{\varphi_x(B)}{\vartheta_x(B)}x_t$ and apply the same filter to $y_t \to \beta_t = \frac{\varphi_x(B)}{\vartheta_x(B)}y_t$.

3. Calculate the CCF between the whitened input series and the residuals of the model for y_t,

$$\widehat{\nu}_k = \frac{\widehat{\sigma}_\beta}{\widehat{\sigma}_\alpha} \ldots,$$

   to get a preliminary estimation of the transfer function ν_k.

4. Identify the order b, r, s of the TF model by inspecting the estimated TF (or equivalently the CCF) and estimate the transfer function using the fact that

$$\widehat{\nu}_j = \frac{\widehat{\omega}(B)}{\widehat{\delta}(B)}B^b,$$

   which is of course done by nonlinear least squares or other methods.

5. Identify a model for the estimated residuals $\widehat{\eta}_t$ given by

$$\widehat{\eta}_t = y_t - \widehat{\nu}(B)x_t.$$

6. Estimate the model and check goodness-of-fit, generally by checking both $\widehat{\varepsilon}_t$ and $\widehat{\alpha}_t$ are white noise. Moreover, since we assume that $\varepsilon_t \sim \text{WN}$ and $\eta_t \perp\!\!\!\perp x_t$. we need to check that $\widehat{\rho}_{\alpha,\widehat{\varepsilon}}(k)$ is non significant.

For checking the last step, there are test statistics which are based on Portmanteau tests.

Lecture 4: Spectral analysis

2021-12-13

In this lecture we introduce spectral analysis, which transforms the data from the time domain to the frequency domain by decomposing the time series into a Fourier basis of coefficients. The idea is to decompose X_t in terms of combination of sinusoids with random and uncorrelated coefficients.

4.1 Periodicity

Consider a periodic process of the form

$$X_t = C \cdot \cos(2\pi\omega t + \varphi), \quad t = \pm 1, \pm 2, \ldots,$$

where ω is a *frequency* index, C the *amplitude* and φ the *phase* of the process. We can introduce random variation in X_t by allowing the amplitude and phase to vary, since by the usual sine and cosine rules we can write X_t as

$$X_t = A\cos(2\pi\omega t) + B\sin(2\pi\omega t). \tag{7}$$

In the above equation, $A = C\cos\varphi$ and $B = -C\sin\varphi$, and $A, B \sim \mathcal{N}(0, \sigma^2)$. We have that

1. $C = \sqrt{A^2 + B^2}$
2. $\varphi = \tan^{-1}(-B/A)$
3. X_t is a stationary process with $\mu_t = 0$ and

$$\gamma(h) = \mathrm{Cov}(X_t, X_{t+h}) = \sigma^2\cos(2\pi\omega h).$$

We consider a generalization of (7) given by a mixture of periodic series with multiple frequencies and amplitudes,

$$X_t = A_0 + \sum_{i=1}^{q} \left\{A_i\cos(2\pi\omega_i t) + B_i\sin(2\pi\omega_i t)\right\}, \tag{8}$$

where $A_i, B_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$ and the ω_i are distinct frequencies. In this case,

$$\gamma(0) = \sum_{i=1}^{q}\sigma_i^2, \quad \gamma(h) = \sum_{i=1}^{q}\sigma_i^2\cos(2\pi\omega_i t).$$

The main objective of spectral-based time-series analysis is to sort out the essential frequency components ω_i of a time series, including their relative contribution to the total power of the signal.

For a sample $x_1, \ldots, x_n$ from X_t we can write the following representation

$$X_t = A_0 + \sum_{j=1}^{\frac{n-1}{2}} \left\{A_j\cos(2\pi tj/n) + B_j\sin(2\pi tj/n)\right\}, \tag{9}$$

for $t = 1, 2, \ldots, n$ and suitably chosen coefficients. If n is even we can modify the above equation and an additional component. Equation (9) holds for any sample and can be interpreted as an approximation to (8) with some coefficients possibly close to zero.

Our problem is now to estimate the A_j's and B_j's using a linear model given the frequencies which are relevant to the observed model. We do so by plotting the ***periodogram***, i.e. the estimates of the variance explained by the j^{th} component $P(j/n) = \frac{1}{2}(\widehat{A}_j^2 + \widehat{B}_j^2)$.

By inspecting the periodogram we can observe which frequencies $\omega_j = j/n$ are predominant over the others and eventually observe frequencies which are "hidden" inside the time series.

Theorem 2 (Parseval's theorem)

The sample variance is the sum of the contribution of the observed periodogram

$$\frac{1}{n}\sum_{t=1}^{n}(X_t - \overline{X})^2 = \frac{1}{2}\sum_{j=1}^{\frac{n-1}{2}}(A_j^2 + B_j^2) = \sum_{j=1}^{\frac{n-1}{2}} P(j/n)$$

Lecture 5: Spectral analysis (cont.)

2021-12-20

If a stationary process X_t has auto covariance function which is absolutely summable,

$$\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty,$$

then it has the representation in terms of its Fourier transform given by

$$\gamma(h) = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(\omega) e^{2\pi i \omega h} f(\omega)\, \mathrm{d}\omega, \quad h = 0, \pm 1, \pm 2, \ldots,$$

and we can define the spectral density of the process as

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega h}, \quad -\frac{1}{2} \leq \omega \leq \frac{1}{2}.$$

Properties of f

1. $f(\omega) \geq 0$ for all ω
2. $f(\omega) = f(-\omega)$, therefore we only consider $\omega > 0$.
3. $\gamma(0) = \mathbb{V}[X_t] = \int_{-1/2}^{1/2} f(\omega)\, \mathrm{d}\omega$, which is the total variance of the process.

Since the ARMA processes satisfy absolute summability, we can represent them in terms of their spectral densities.

Example (White noise)
White noise is such that $X_t \sim \mathrm{WN}(0, \sigma^2)$ has a constant frequency spectrum, since no frequency dominates over any other.

Example (AR(1))
We can write

$$f(\omega) = \sum_{-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega h} = \frac{\sigma_\varepsilon^2}{1 - 2\varphi \cos(2\pi\omega) + \varphi^2},$$

and if $\varphi > 0$ the spectrum is dominated by low frequencies, whereas if $\varphi < 0$ the dominating frequencies are high.

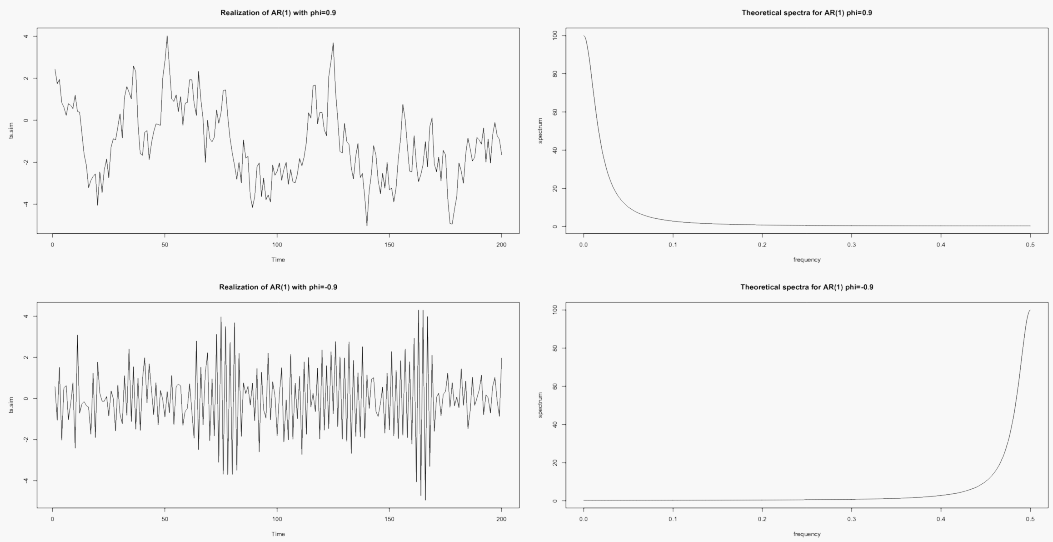

Figure 3: Frequency spectrum for AR(1) models with $\varphi > 0$ (above) and $\varphi < 0$ (below).

Example (MA(1))

The same behaviour can be seen for a moving average model, which is however less visible than the AR process. Indeed, the MA component is useful to model the component of the process which has not been explained by the autoregressive part.

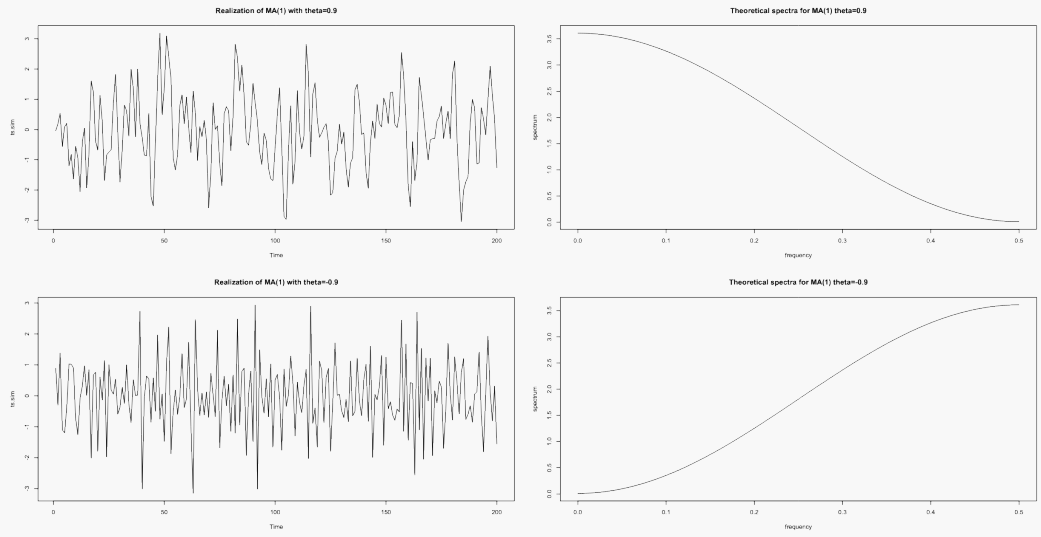

Figure 4: Frequency spectrum for a MA(1) model with $\vartheta > 0$ (above) and $\vartheta < 0$ (below).

Example (`ARMA`(p, q))

For a general $X_t \sim \text{ARMA}(p, q)$ we can prove that

$$f(\omega) = \sigma_\varepsilon^2 \frac{|\vartheta(e^{-2\pi i\omega})|^2}{|\varphi(e^{-2\pi i\omega})|^2} \tag{10}$$

5.1 Estimation of the spectral density

Estimating the spectral density can be done similarly to what we do for the histogram of a continuous density. Let X_t be a zero-mean stochastic process, we define

$$\widehat{f}(\omega) = \widehat{\gamma}_0 + 2\sum_{k=1}^{n-1} \widehat{\gamma}(k)\cos(2\pi\omega k),$$

which is a way in which we can write the spectral density of X_t. The periodogram as we defined above is an inconsistent estimate of the true spectral density, which has poor sample properties.

For a periodogram we can see that

$$\frac{\widehat{A}_\omega^2 + \widehat{B}_\omega^2}{\gamma_0} = \frac{2\widehat{f}(\omega)}{f(\omega)} \sim \chi_2^2,$$

hence $\mathbb{E}[\widehat{f}(\omega)] = f(\omega)$ and $\mathbb{V}[\widehat{f}(\omega)] = f^2(\omega)$. Hence, the variance of the estimator $\not\to 0$ as $n \to \infty$, which means that the estimation strategy yields a bad result.

Alternative methods for constructing estimators include the following approaches:

1. *Nonparametric estimators*: using a moving average to smooth the estimate,

$$\overline{f}(\omega) = \sum_{k=-m}^{m} w_m(k)\widehat{f}(\omega_j + \frac{k}{n}),$$

   where $w_m(k) \geq 0$, $w_m(k) = w_m(-k)$ and $\sum_{k=-m}^{m} w_m(k) = 1$. Using the smoothed spectrum we have

$$\frac{2(2m+1)\overline{f}(\omega)}{f(\omega)} \sim \chi^2_{2(2m+1)},$$

   consequently $\mathbb{E}[\overline{f}(\omega)] \approx f(\omega)$ and $\mathbb{V}[\overline{f}(\omega)] \approx f^2(\omega)/(2m+1)$ and we have consistency if $m \to \infty$ while $m/n \to 0$. We usually choose something like $m = \sqrt{n}$ to get some insights into the shape of the true spectrum.

2. *Parametric estimators* based on a fitted AR model. Since any ARMA time series admits an AR(∞) representation, we can use the theoretical estimate of such a model estimated using AIC, AICC, BIC, ...

   The disadvantage in this case can be the fact that p can be very large. Instead, we can use an ARMA model and use the theoretical spectral density in Equation (10) to represent the estimated $\widehat{f}$.

Lecture 6: Nonlinear time series models

2022-01-11

A lot of research in the last 30 years has been spent to develop time-series models which can detect nonlinear patterns in the data. We refer again to Wold's theorem (**??**) and consider a time series of the form

$$X_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} + V_t,$$

where $\psi_0 = 1$ and $\sum_{j=1}^{\infty} \psi_j^2 < \infty$, $a_t \sim \text{WN}(0, \sigma^2)$ and a_j's are uncorrelated.

Def. (General linear process)
X_t is said to be ***linear*** if if can be written as

$$X_t = \mu + \sum_{j=0}^{\infty} \psi_j a_{t-j},$$

such that $\sum |\psi_j| < \infty$ and $a_j \overset{\text{iid}}{\sim} \text{WN}(0, \sigma^2)$. The above process is sometimes called ***general linear process***.

Limitations Linear models are limited in the sense that they cannot model ***strong asymmetries*** in data, ***irregular jumps***, and ***switching regimes***.

6.1 Nonlinear framework

Whereas linearity is well-defined, non-linearity is hardly definite and the early development of nonlinear time series focused on various parametric forms: ARCH, GARCH, threshold models, ...

In this lectures we emphasize simple parametric models which are applicable without overly-complex specifications. There are examples of *explicit* and *implicit* approaches, which differ in the way they are represented:

- *Implicit*: ARMA model with non-gaussian innovations.
- *Explicit*: $X_t = h(X_{t-1}, \ldots, X_{t-k})$.

Explicit models have surpassed implicit modelling since it is in general difficult to identify the correct distribution of the white-noise terms.

Attention In the nonlinear setting, the tools of standard analysis (ACF, PCF, ...) are not helpful since they only detect linear patterns.

Def. (Nonlinear model)
The genreal representation is

$$X_t = f(a_t, a_{t-1}, a_{t-2}),$$

where $a_t \overset{\text{iid}}{\sim} \text{WN}(0, \sigma^2)$ and $f(\cdot)$ is some nonlinear function.

Remark We could linearize the above model by considering the Taylor series around zero (Volterra series)

$$X_t = \mu + \sum_{i,j} b_{ij} a_{i,t-j} + \sum_{ijkl} b_{ij,kl} a_{i,t-j} a_{k,t-l} + \dots \tag{11}$$

The Volterra series (11) is usually too complicated unless severely truncated.

Other very general models are ARMA models with time-dependent parameters.

The model for nonlinear time series can be written in terms of the conditional mean and variance,

$$\mu_t = \mathbb{E}[X_t|\mathcal{F}_{t-1}] = g(F_{t-1})$$

$$\sigma_t^2 = \mathbb{V}[X_t|\mathcal{F}_{t-1}] = h(F_{t-1})$$

where g and h are well-defined nonlinear function with $h(\cdot) > 0$. If $g(\cdot)$ is nonlinear and $h(\cdot)$ is constant, then X_t is nonlinear in mean, for example

$$X_t = \varepsilon_t + \alpha\varepsilon_{t-1}^2.$$

Otherwise, if $h(\cdot)$ is time-variant then X_t is nonlinear in variance. All GARCH models are of this type.

6.2 NLAR(1)

We consider the simplest conditional mean model given by

$$X_t = g(X_{t-1}, \vartheta) + a_t,$$

where ϑ is a vector of parameters and $a_t \sim$ IID. It's natural to consider functions which are nearly linear as first candidates, but also more extreme nonlinear functions if needed.

There are few papers on nonlinear autoregressive processes, since the fact that there are too many nonlinear functions that we can consider renders this class of processes unusable for statistical analysis.

6.3 Reversibility

Def. (Reversibility)
The stationary sequence X_t is ***time-reversible*** if the finite-dimensional distributions of $(X_1, X_2, \dots, X_n)$ and $(X_n, \dots, X_2, X_1)$ is the same for any n.

Usage Since i.i.d sequences and ARMA models are time-reversible, we can use time-reversibility to detect deviations from the Gaussianity-linearity hypothesis. For example, Chen et al (2000) look at the test statistic

$$\mathbb{E}[\sin(\omega(X_t - X_{t-k}))],$$

and if the value is zero then the process is time-reversible. Other tests rely for example on differences between backward and forward autocorrelation.

6.4 Threshold AR models

Consider the change-point model

$$X_t = \begin{cases} \varphi_1 X_{t-1} + a_t & \text{if } X_{t-1} < r \\ \varphi_2 X_{t-1} + a_t & \text{if } X_{t-1} \geq r \end{cases}$$

we call this model the *self-exciting* threshold-AR model. If X_{t-1} is replaced by an exogenous variable Z_{t-d}, then this model is called *threshold-autoregressive* (TAR). Using piecewise linear models we can obtain a better approximation of the conditional mean equation.

Def. (SETAR model)
We define the ***self-exciting TAR model*** (***SETAR***) with threshold X_{t-d} if

$$X_t = \varphi_0^{(j)} + \varphi_1^{(j)} X_{t-1} + \ldots + \varphi_p^{(j)} X_{t-p} + a_t^{(j)} \quad \text{if } \gamma_{j-1} \leq X_{t-d} \leq \gamma_j,$$

where $j = 1, \ldots, k$.

Remark In the above model, γ_j's are the thresholds and X_{t-d} is the threshold variable.

6.4.1 Estimation

Suppose that we have an observed time series $X_1, \ldots, X_n$ and we fix the order k of the SETAR model. We alternate a two-step procedure:

› First we assume that the partition A_j and orders p_j are known, so that we can use the least squares procedure to minimize the loss function

$$\mathcal{S}(\vartheta) = \sum_{j=1}^{k} S^{(j)} = \sum_{\substack{X_{t-d} \in A_j \\ p < t \leq n}} \left[X_t - (\varphi_0^{(j)} + \varphi_1^{(j)} X_{t-1} + \ldots + \varphi_p^{(j)} X_{t-p}) \right]^2$$

› We find the partition $\widehat{A}_j$ such that $\mathcal{S}(\widehat{\vartheta})$ is minimized.

Otherwise, to determine the autoregressive orders $p^{(j)}$'s we might use an information criteria such as BIC, AICC, ...

Testing for linearity becomes essential to check non-linearity when fitting nonlinear models. In general, there are

a) Tests for departure from linear models towards general nonlinear models (less powerful).

b) Tests for departure from linear models towards threshold autoregressive (more powerful).

6.4.2 Smooth Transition Autoregressive models

A generalization of the TAR model is the ***smooth transition autoregressive model*** (STAR) which allows for a smoother transition between the two regimes via a cumulative distribution function instead of a jump function.

Lecture 7: Markov-switching model

2022-01-13

A Markov-switching (MS) model changes the rules of the switching regimes from the threshold-autoregressive model, which are not deterministic anymore. A MS(p) model with two regimes can be defined as

$$X_t = \begin{cases} \alpha_1 + \sum_{i=1}^{p} \varphi_{1,i} X_{t-i} + a_{1,t} & \text{if } s_t = 1 \\ \alpha_2 + \sum_{i=1}^{p} \varphi_{2,i} X_{t-i} + a_{2,t} & \text{if } s_t = 2 \end{cases} \tag{12}$$

where $a_{i,t} \sim \text{IID}(0, \sigma_i^2)$. The state variable s_t is unobservable and we assume that it follows a first-order Markov chain with transition probabilities

$$P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix},$$

and a particular choice of P drives the model behaviour. This is crucially different from the SETAR model, since the regimes are defined by the Markov chain and not simply determined by the past values of X_t. Thus, *forecast* of a Markov-switching model are linear combinations of forecasts produced by the sub-models.

Estimating a MS model is much harder since the states are not observable, therefore we need a ***filtering*** approach. The log-likelihood can be constructing recursively from some initial conditions since

$$f_{it} = f(X_t | s_t = i, X_{t-1}, \vartheta_i), \quad i = 1, 2,$$

and under normality this is a Gaussian density with parameters from Equation (12). Now, we can write the contribution by marginalizing s_t as

$$g(X_t | X_{t-1}, \vartheta_1, \vartheta_2) = f_{1t}\rho_{1t|t-1} + f_{2t}\rho_{2t|t-1}, \tag{13}$$

and Hamilton have shown that the optimal inference and forecast can be determined from the conditional likelihood in (13).

7.1 Bilinear models

This class of models is not very used/useful even though they are a natural extension of the ARMA model. A general bilinear model BL(p, q, r, s) can be written as

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + a_t + \sum_{i=1}^{q} \vartheta_j a_{t-j} + \underbrace{\sum_{i=1}^{r} \sum_{j=1}^{s} \beta_{ij} X_{t-i} a_{t-j}}_{\text{bilinear component}},$$

where $a_t \sim \text{IID}(0, \sigma_a^2)$. This model is too complex and therefore people usually study the Lower Triangular Bilinear Model

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + a_t + \sum_{i=1}^{q} \vartheta_j a_{t-j} + \underbrace{\sum_{i=1}^{r} \sum_{j=1}^{s} \beta_{ij} X_{t-i-j} a_{t-i}}_{\text{bilinear component}},$$

where X has only past values w.r. to a in the bilinear component.

› These models can model occasional outbursts in time series.

› BL model can have conditional heteroscedasticity, although GARCH-type models are better in this regard.

› ML procedures are used but asymptotic distribution is unknown.

› Probabilistic properties are often derived using the state-space representation.

7.2 Long-memory models

ARMA models are said to be short-term models, since their autocorrelation function usually tends to zero with an exponential decrease. Thus the $d \in \{0, 1\}$ parameter in an ARIMA(p, d, q) model controls the transition from short-memory to infinite memory.

Granger introduced an extension to the ARIMA model by considering $d \in [0, 1]$ yielding ARFIMA models. These models are necessary if we want to take into account series that show memory between $I(0)$ and $I(1)$ processes.

Properties ARFIMA models can take into account

› Presence of long-range cycles

› Slowly-decaying autocorrelation structures.

There are different definitions of long-memory processes:

1. In the time domain, a long-memory process is such that its autocorrelation function decays like a power function, i.e. if $\alpha \in (0, 1)$ and $c_\rho > 0$

$$\rho(k) = c_\rho k^{-\alpha}, \quad k \longrightarrow \infty.$$

2. In the frequency domain, a long-memory process is such that its spectral density is unbounded at zero, i.e.

$$f(\omega) \sim c_f \omega^{-\alpha}, \quad \omega \longrightarrow 0^+.$$

We can define the fractional difference operator using the Gamma function as

$$(1 - B)^d = \sum_{j=0}^{\infty} \frac{\Gamma(j - d)}{\Gamma(j + 1)\Gamma(-d)} B^j,$$

and when $d \in (0, 0.5)$ the ARFIMA(p, d, q) process is stationary with $\rho(k) \sim k^{2d-1}$. When $d \in (-0.5, 0)$ the process is stationary with intermediate memory, although in practice this is never used. For $d \in [0.5, 1)$ the process is mean-reverting even though it is not covariance-stationary.

In the following we will focus on ARFIMA(p, d, q) processes with $d \in (0, 0.5)$ which yields the most interesting type of process.

7.2.1 Estimation

Estimation approaches are mainly divided into two broad classes:

1. ML estimation, which requires specifying both p and q.
2. Semi-parametric or nonparametric approaches, where we assume that the ARMA component is relatively unimportant.

One of the best-known method is the semi-parametric GPH estimator introduced by Geweke and Porter-Hudack and developed by Robinson. This method approximates the spectral density near the origin,

$$f(\omega) \sim c_f \big(4 \sin^2(\omega/2)\big)^{-d},$$

and therefore we can apply the least-squares method to

$$\log l(\omega_j) = \log c_f - d \log \big(4 \sin^2(\omega_j/2)\big) + u_j,$$

where u_j are i.i.d error terms and ω_j are the Fourier frequencies. The problem with this method is its high variance in the estimates.

7.3 Integer autoregressive (INAR) models

Integer autoregressive (INAR) models can be used to model time series of counts, which are of particular interest in practice. In some cases, the discrete values can be approximated by Gaussian models, however for small values we need a more proper model.

Notation We introduce the ***thinning operator*** $\circ$ which substitutes the multiplication operator. Let $\alpha \in [0, 1]$, then we define

$$\alpha \circ X = \sum_{i=1}^{X} Y_i, \quad Y_i \text{ i.i.d r.v.'s with } \mathbb{E}[Y_i] = \alpha. \tag{14}$$

Typically, Y_i's are assumed to be i.i.d $\text{Ber}(\alpha)$, and therefore we have

$$Y_i \overset{\text{iid}}{\sim} \text{Ber}(\alpha) \implies \alpha \circ X | X \sim \text{Bin}(X, \alpha).$$

The INAR(1) model is defined as

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t,$$

where $\alpha \in [0, 1)$ and ε_t are i.i.d discrete random variables with mean $\mu_\varepsilon > 0$ and variance σ_ε^2. Usually we consider Poisson-distributed errors, but more flexible discrete distributions are possible.

The INAR(1) process is non-linear due to the thinning operator, but it's a member of the conditional linear first-order AR models,

$$\mathbb{E}[X_t | X_{t-1}] = \alpha X_{t-1} + \mu_\varepsilon,$$

$$\mathbb{V}[X_t | X_{t-1}] = \alpha(1 - \alpha) X_{t-1} + \sigma_\varepsilon^2.$$

If $\varepsilon_t \sim \text{Pois}(\lambda)$, then $X_t \sim \text{Pois}\left(\frac{\lambda}{1-\alpha}\right)$.

Generalizing to the INAR(p) is not straightforward and depends on the definition of the thinning operator, e.g.

$$X_t = \alpha_1 \circ X_{t-1} + \ldots + \alpha_p \circ X_{t-p} + \varepsilon_t,$$

where the thinning operators are applied with independent Y_j's from (14).

Estimation is simple for Poisson innovations and $p = 1$, whereas for all other cases we have some problems. As for the forecast we use the median, since we want an integer value for our predictions and the mean is usually a real number.

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.