Time Series in Mathematical Finance

Cláudia Nunes

Instituto Superior Técnico (IST, Portugal) and CEMAT
cnunes@math.ist.utl.pt
European Summer School in Industrial Mathematics
Universidad Carlos III de Madrid

July 2013

Outline

The objective of this short course is to provide some knowledge of time series, with applications in finance. For that we need to understand the particular characteristics of these kinds of series, introduce some statistical tools and gain experience.

- Part I: Introductory principles
- Part II: Linear Time Series Models
- Part III: More Fancy Models...

Main goal of the course

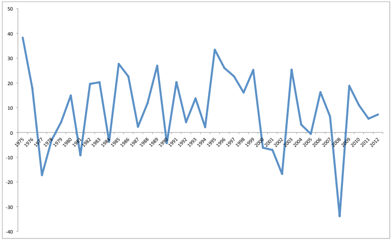


Figura : Dow Jones Industrial Average Yearly Returns, reporting to 1975-2012

Can we predict what will happen to the Dow Jones Industrial Average Yearly Returns for the next 5 years?

Time Series

We want to use *time series* in order to answer questions like this one.
What is a time series? It is a set of repeated observations of the same variable (???? is this true???) (such as the GNP or a stock return). We can write a time series as:

$$\{X_1, X_2, \ldots, X_T\}$$

or simply as $\{X_t\}$. The subscript t indicates time...
The word time series is used interchangeably to denote a sample, $\{x_t\}$, and a probability model for that sample.

Models

Time-series consists of interesting parametric models for the joint distribution of $\{X_t\}$.

The models impose structure, which we must evaluate to see if it captures the features that we think are present in the data.

Also we need to reduce the estimation problem to the estimation of a few parameters of the time-series model.

Independence

Usually in statistics one assumes that the random variables are independent over time. However time series are typically not i.i.d., which is what makes them interesting.
For example, if the return of a stock is today unusually high, then probably tomorrow will also be unusually high, until it drops down.

Part I: Introductory principles

The theory of time series includes the following aspects:

- stationarity of the data
- dynamic dependence
- autocorrelation
- modeling/estimation
- forecasting

Linear models means that the model attempt to capture the linear relationship between X_t (the data) and information available prior to time t.

Stationarity

- A time-series $\{X_t\}$ is strictly stationary if:

$$(X_{t_1}, X_{t_2}, \dots, X_{t_k}) =^{\mathcal{D}} (X_{t_1+t}, X_{t_2+t}, \dots, X_{t_k+t}), \quad \forall t, t_1, t_2, t_k, k$$

(joint distribution is invariant under time shift)

- This is a very strong condition, hard to verify empirically.
- Weaker version: A time-series $\{X_t\}$ is weakly-stationary if:

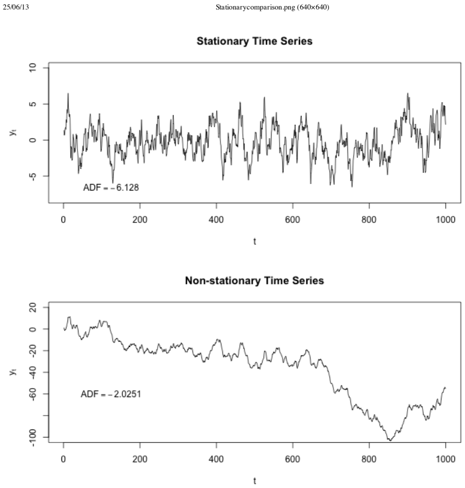
$$E[X_t] = \mu, \quad Cov(X_t, X_{t-l}) = \gamma_l, \quad \forall t, l$$

Properties of stationarity

- Strong stationarity does *not* $\Rightarrow$ weak stationarity, as $E[X_t^2]$ must be finite (e.g., the Cauchy distribution... not even the first order moment is finite!)
- Strong stationarity+finite second order moments $\Rightarrow$ weak stationarity.
- Weak stationarity does *not* $\Rightarrow$ strong stationarity; the exception occurs for gaussian processes, so:

Weak stationarity + normality $\Rightarrow$ strong stationarity.

Two simulated time series processes, one weakly-stationary the other non-stationary.



From now one, when we say:

Stationary

we mean

Weakly-stationary

Covariance and Correlation Functions

- $\gamma_l = Cov(X_t, X_{t-l})$ is called the lag-l autocovariance, which has the following properties:
  - $\gamma_0 = Var[X_t]$ (assuming stationarity)
  - $\gamma_l = \gamma_{-l}$
- $\rho_l = \frac{\gamma_l}{\gamma_0}$ is called the lag-l autocorrelation.
  - $\rho_0 = 1$
  - $\rho_l = \rho_{-l}$
- $\hat{\rho}_l = \frac{\sum_{t=l+1}^{T}(X_t - \bar{X})(X_{t-l} - \bar{X})}{\sum_{t=l+1}^{T}(X_t - \bar{X})^2}$ is the lag-l sample autocorrelation / empirical autocorrelation.

Tests for the Significance of Correlation

- Testing individual ACF: to test $H_0 : \rho_l = 0$ vs $H_1 : \rho_l \neq 0$:

$$Z = \frac{\hat{\rho}_l}{\sqrt{(1 + 2\sum_{i=1}^{l-1} \hat{\rho}_i^2)/T}} \sim N(0,1)$$

  Reject H_0 if $|z_{obs}| > \phi^{-1}(1 - \alpha/2)$.
- Portmanteau test: to test $H_0 : \rho_1 = \rho_2 = \ldots = \rho_m = 0$:

$$Q(m) = T \sum_{l=1}^{m} \hat{\rho}_l^2 \sim_a \chi^2_{(m)}$$

- Ljung and Box test: similar to the previous one, with:

$$Q(m) = T(T+2) \sum_{l=1}^{m} \frac{\hat{\rho}_l^2}{T - l} \sim_a \chi^2_{(m)}$$

Note: the selection of m may affect the performance of the $Q(m)$ statistics in the last two tests. Simulation studies suggest the $m \approx \ln(T)$.

- The correlogram analysis is a key tool to explore the interdependency of the observation values.
- It can also be used as a tool to identify the model and the estimate the orders of its components.

Example: data from TUI

- TUI AG is a German multinational travel and tourism company, dedicated to tourism, shipping, and logistics.
- Today it is one of the world's largest tourist firms with interests across Europe. It owns travel agencies, hotels, airlines, cruise ships and retail stores.
- It is quoted in the LSE (TUI Travel PLC (TT.L) -LSE)
- Data available online include TUI stock values, in particular, daily values, quoting open, high, low, close and volume values.
- Data reports from 13/01/2000 until 14/05/2002 (more recent data! check http://uk.finance.yahoo.com/q/hp?s=TT.L)

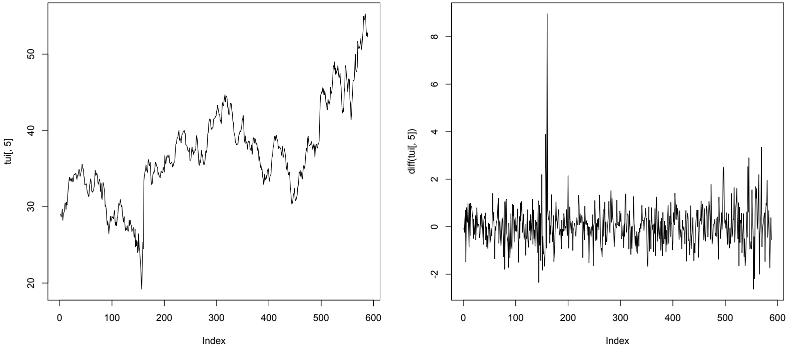


Figura : Plot of close values and difference of order 1 of close values

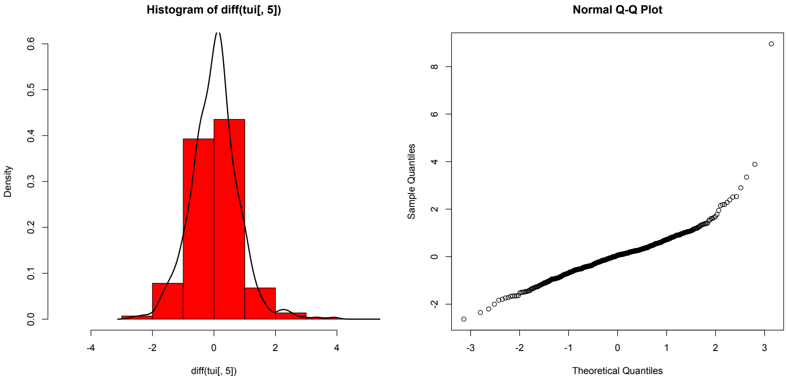


Figura : Histogram and QQPlot for difference of order 1 of close values

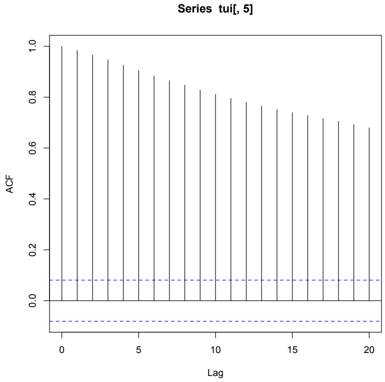


Figura : Empirical ACF for close values

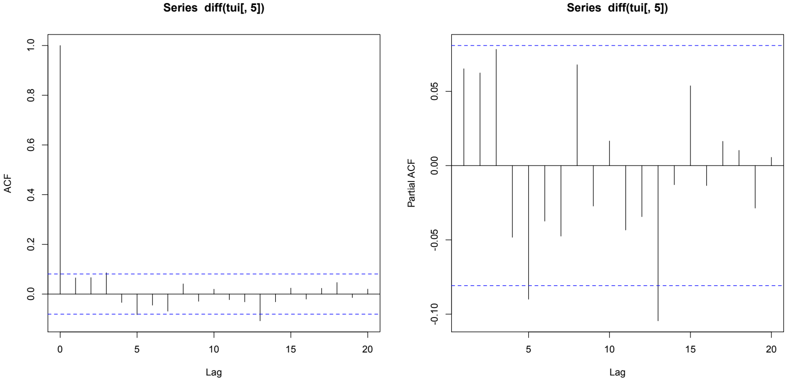


Figura : Empirical ACF and PACF for difference of order 1 of close values

Partial autocorrelation function

In addition to the autocorrelation between X_t and X_{t+k}, we may want to investigate the correlation between them after their dependency on $X_{t+1}, \ldots, X_{t+k-1}$ has been removed.

$$\phi_{kk} = Corr\left(X_t, X_{t+k} | X_{t+1}, \ldots, X_{t+k-1}\right)$$

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & \rho_1 & 1 \end{vmatrix}}$$

Filters

- Recall that in the simple linear regression model, we assume the following:
$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$
with $E[\varepsilon_i] = 0$, $Var[\varepsilon_i] = \sigma^2$ and $Cov(\varepsilon_i, \varepsilon_j) = 0$, for $i \neq j$
- So we decompose the response variable, Y, in a linear trend plus an error term.
- A key concept in traditional time series analysis is the decomposition of a given time series $\{X_t\}$ into a trend $\{T_t\}$, a seasonal component $\{S_t\}$ and the remainder, $\{\varepsilon_t\}$.
- A common method for obtaining the trend is to use linear filters on given time series:
$$T_t = \sum_{i=-\infty}^{\infty} \lambda_i X_{t+i}$$

For the TUI data:

- $\lambda_i = \{\frac{1}{5}, \ldots, \frac{1}{5}\}$ (5 times) $\rightarrow$ weekly averages
- $\lambda_i = \{\frac{1}{25}, \ldots, \frac{1}{25}\}$ (25 times) $\rightarrow$ monthly averages
- $\lambda_i = \{\frac{1}{81}, \ldots, \frac{1}{81}\}$ (81 times) $\rightarrow$ quaterly averages

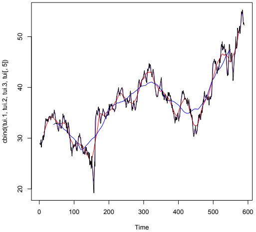


Figura : Linear filters for weekly, monthly and quaterly averages

The irregularity decreases as we filter more, ie, if we weight more observations

Decomposition of time series

It is often assumed that many macroeconomic time series are subject to two sorts of forces:

- those that influence the long-run behavior of the series
- those that influence the short-run behavior of the series.

So, for example, growth theory focuses on the forces that influence long-run behavior whereas business cycle theory focuses on the forces that influence short-run behavior.
Other reasons - to transform a nonstationary series into a stationary series by removing the trend!

Components

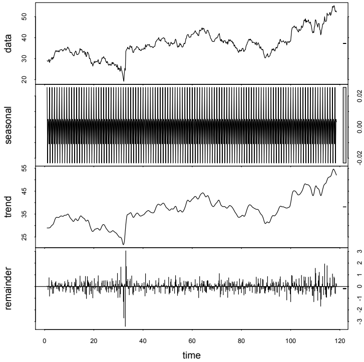
Another possibility for evaluating the trend of a time series is to use nonparametric regression techniques:

$$X_t = T_t + C_t + S_t + \varepsilon_t$$

- T_t is the Trend Component, that reflects the long term progression of the series;
- C_t is the Cyclical Component, that describes repeated but non-periodic fluctuations, possibly caused by the economic cycle;
- S_t, the Seasonal Component, reflecting seasonal variation;
- ε_t, the Irregular Component (or "noise"), that describes random, irregular influences. It represents the residuals of the time series after the other components have been removed.

Weekly decomposition

For the TUI data: weekly, monthly or quarterly seasonal decomposition?



Monthly decomposition

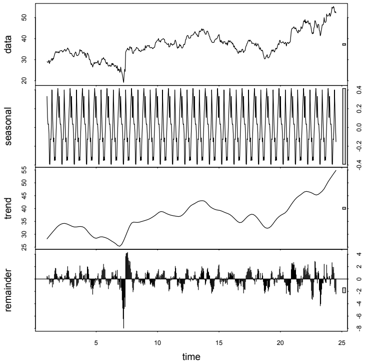


Figura : Monthly decomposition

Quaterly decomposition

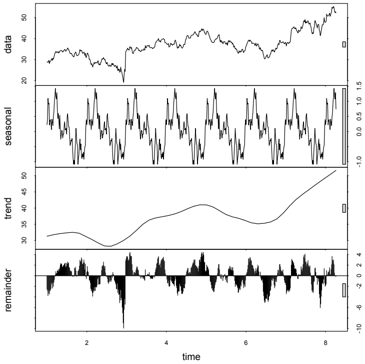
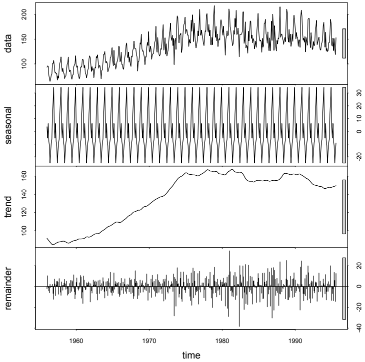


Figura : Quaterly decomposition

When seasonal decomposition really works...

Data: monthly beer production in Australia from Jan. 1956 to Aug. 1995



Decomposition based on differences

An alternative to the trend stationary assumption to account for trend behavior in a time series is to assume that the series is difference stationary, i.e., $\{X_t\}$ is stationary in differenced form.

A time series $\{X_t\}$ is difference stationary of order d, if

- $\{\Delta^d X_t\}$ is stationary
- $\{\Delta^{d-1} X_t\}$ is not-stationary

Note:

- $\Delta X_t = X_t - X_{t-1}$
- $\Delta^2 X_t = X_t - 2X_{t-1} + X_{t-2}$
- $\Delta^d X_t = \sum_{i=0}^{d} (-1)^i \binom{d}{i} X_{t-i}$ (prove this!)

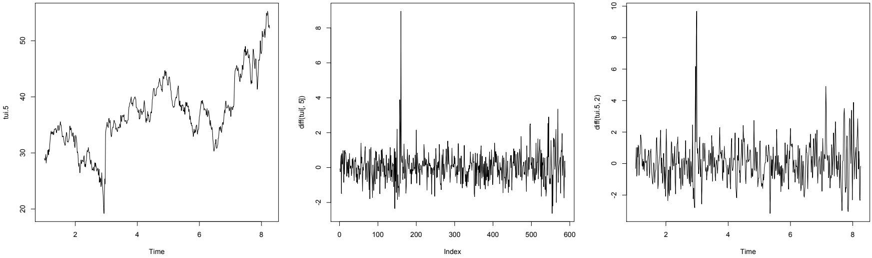


Figura : Tui, ΔTui and Δ^2Tui series

ΔTui looks already stationary; there is no need for a second order difference...

Identifying the order of differencing

Beware of over-differencing! This is one of the common errors that practioners do!

- Rule 1: If the series has positive autocorrelations out a high number of lags, then it probably need a higher order of differencing. Differencing tends to introduce *negative* correlation. If you apply more than once, lag-1 autocorrelation will be driven even further in the negative direction.
- Rule 2: If the lag-1 autocorrelation is zero or even negative, then the series does not need further differencing. If the lag-1 autocorrelation is more negative than -0.5, this may mean the series has been overdifferenced. Check if there are patterns (like up-down-up) in the transformed series.

- Rule 3: IThe optimal order of differencing is often the order of differencing at which the standard deviation is lowest.
- Rule 4: A model with no orders of differencing assumes that the original series is stationary. A model with one order of differencing assumes that the original series has a constant average trend...

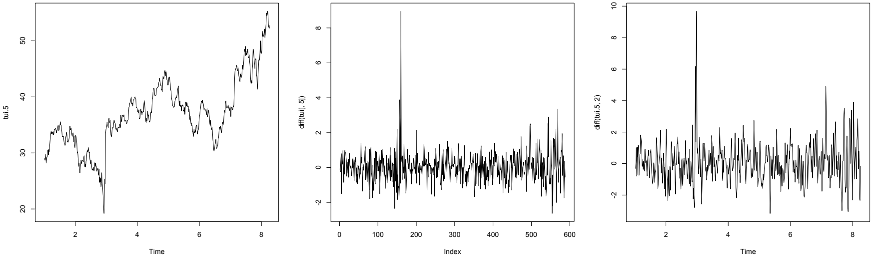


Figura : Tui, ΔTui and Δ^2Tui series

ΔTui looks already stationary; there is no need for a second order difference...

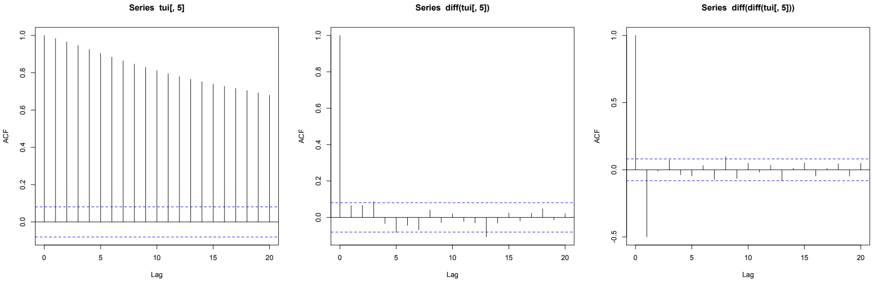


Figura : Empirical ACF of Tui, ΔTui and Δ^2Tui series

Box-Cox transformations

What happens if even with differences, we cannot stabilize the data? In particular, we cannot stabilize the variance...

Box-Cox transformations is a family of power transforms, used precisely to stabilize variance, make the data more normal distribution-like, improve the validity of measures of association such as the Pearson correlation between variables and for other data stabilization procedures.

$$Y_t^\lambda = \begin{cases} \frac{X_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln X_t & \lambda = 0 \end{cases}$$

Some commonly used transformations

Values of λ	Transformation
-1.0	$\frac{1}{X_t}$
-0.5	$\frac{1}{\sqrt{X_t}}$
0	$\ln X_t$
0.5	$\sqrt{X_t}$
1.0	X_t

The Box-Cox Power transformation only works if all the data is positive

This can be achieved by adding a constant c to all data such that it all becomes positive before it is transformed.

The Box-Cox Power transformation and Differencing

- A variance stabilizing transformation, if needed, should be performed before any other analysis, such as diferencing.
- Frequently, the transformation not only stabilizes the variance, but also improves the approximation of the distribution by a normal distribution.

Example

Data that is not normal at all...

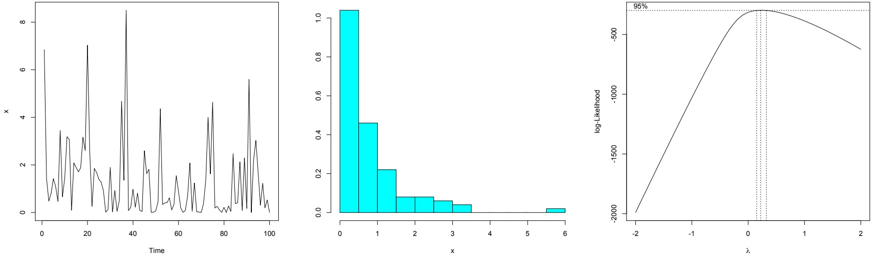


Figura : Data, histogram and log-likelihood as a function of the power λ

Best Box-Cox transformation: $\lambda = 0.22$

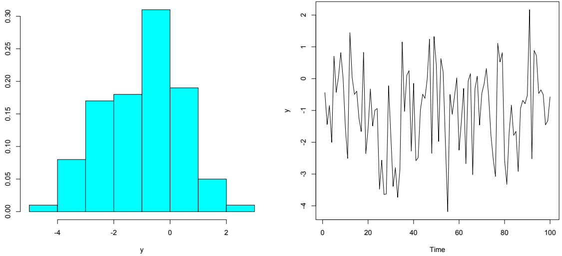


Figura : Histogram and plot of $\frac{X_t^{0.222}-1}{0.22}$

Linear Regression model

In many applications, the relationship between two time series is of major interest.

- The market model in finance is an example that relates the return of an individual stock to the return of a market index.
- The term structure of interest rates is another example in which the time evolution of the relationship between interest rates with different maturities is investigated.

These examples lead to the consideration of a linear regression in the form

$$Y_t = \beta_0 + \beta_1 Y_t + \varepsilon_t$$

where $\{X_t\}$ and $\{Y_t\}$ are two time series.

Correlated error terms

In the usual LRS, we assume that $Cov(\varepsilon_i, \varepsilon_j) = 0$ for $i \neq j$. But in real applications this is hardly the case, and so we end up with an error term that is serially correlated!

Consequence: the LS estimates of β_0 and β_1 may not be consistent!

This is one of the most commonly misued econometric models, because the serial dependence on the errors is overlloked...

Regression for TUI data

Model: $Tui_t = \beta_0 + \beta_1 t + \varepsilon_t$

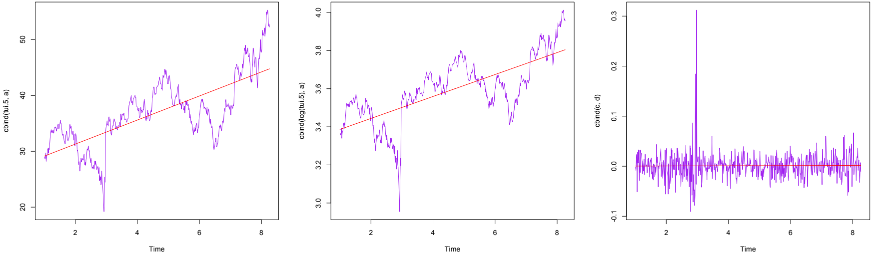


Figura : Regression models adjusted to data Tui, Log(Tui) and Diff(Log(Tui))

Summary statistics for the regression of Tui

```
Call:
lm(formula = log(tui.5) ~ t)

Residuals:
    Min       1Q   Median       3Q      Max
-0.54237 -0.08303  0.02603  0.08575  0.21146

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 3.386e+00  9.832e-03  344.33   <2e-16 ***
t           7.112e-04  2.888e-05   24.63   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1192 on 587 degrees of freedom
Multiple R-squared: 0.5082, Adjusted R-squared: 0.5073
F-statistic: 606.5 on 1 and 587 DF,  p-value: < 2.2e-16
```

Analysis of residuals

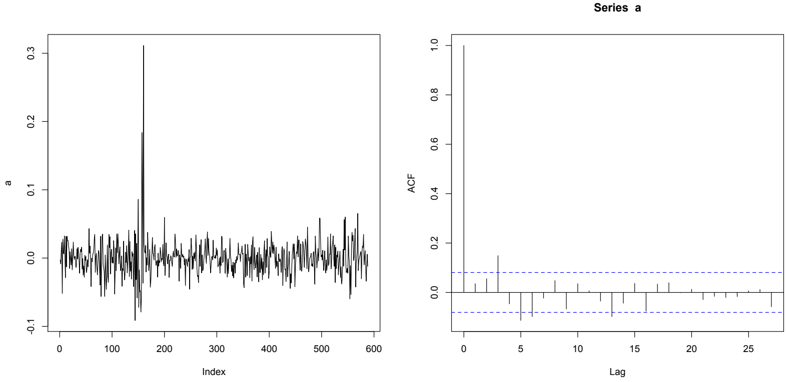


Figura : Analysis of residuals for *Diff*(*Tui*)

Regression for Beer data

Model 1: $log(Beer_t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \varepsilon_t$

Model 2:

$$log(Beer_t) = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 cos(\frac{2\pi}{12}) + \gamma sin(\frac{2\pi}{12}) + \varepsilon_t$$

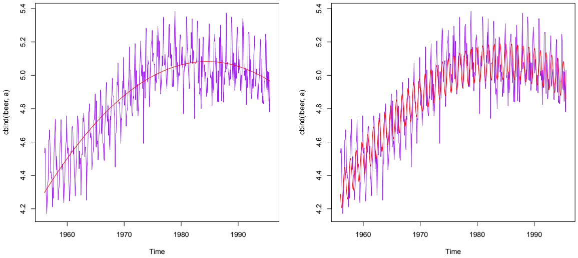


Figura : Regression models 1 and 2

Summary statistics for the regression of beer

```
Call:
lm(formula = lbeer ~ t + t2 + sin.t + cos.t)

Residuals:
     Min       1Q   Median       3Q      Max
-0.33191 -0.08655 -0.00314  0.08177  0.34517

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.833e+03  1.841e+02 -20.815   <2e-16 ***
t            3.868e+00  1.864e-01  20.751   <2e-16 ***
t2          -9.748e-04  4.718e-05 -20.660   <2e-16 ***
sin.t       -1.078e-01  7.679e-03 -14.036   <2e-16 ***
cos.t       -1.246e-02  7.669e-03  -1.624    0.105
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1184 on 471 degrees of freedom
Multiple R-squared: 0.8017,Adjusted R-squared:   0.8
F-statistic: 476.1 on 4 and 471 DF,  p-value: < 2.2e-16
```

Analysis of residuals

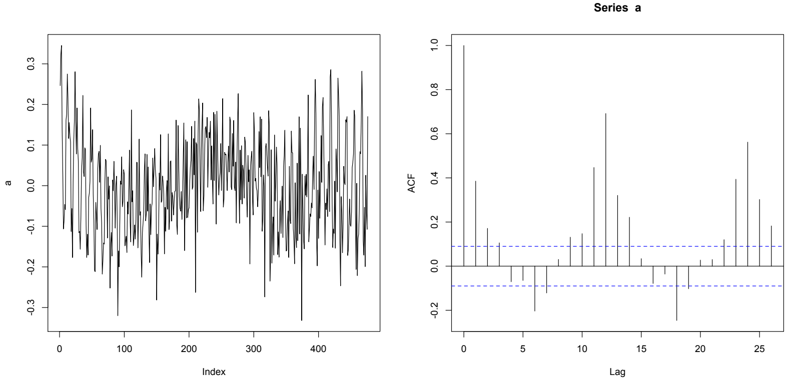


Figura : Analysis of residuals of model 2 for $Log(Beer)$

Statistical Inference

Consider a univariate time series $\{X_t\}$, such that:

$$X_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$
$$\mu_{t+1} = \mu_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2)$$

so that $\{\mu_t\}$ is the trend of the series (not observable) and $\{X_t\}$ is the (observable) data.
There are three types of inference commonly discussed in the literature:

- Filtering
- Prediction
- Smoothing

Types of Inference

Let $\mathcal{F}_t = \sigma(X_1, X_2, \ldots, X_t)$ the information available at time t (inclusive) and assume that the model, as well as its parameters, is known. Then

- Filtering: recover the state variable $\{\mu_t\}$, given $\mathcal{F}_t$ (i.e., remove the measurement errors from the data)
- Prediction: forecast μ_{t+h} or X_{t+h}, given $\mathcal{F}_t$ (where t is the forecast origin)
- Smoothing: estimate μ_t given $\mathcal{F}_T$, where $T > t$.

One simple analogy...

Imagine that you are reading a handwritten note. Then

- Filtering: figure out the word you are reading based on knowledge accumulated from the beginning of the note
- Prediction: guess the next word
- Smoothing: deciphering a particular word once you have read through the note.

In this section...

... We deal, for now, with smoothing!
Notation: $\hat{X}_t(h)$ is the estimator of X_{t+h} given the information up to time t, $\mathcal{F}_t$.

- Idea: each value should be a weighted sum of past observations
- $\hat{X}_t(1) = \sum_{i=0}^{\infty} \lambda_i X_{t-i}$
- One possibility: use geometric weights:

$$\lambda_i = \alpha(1-\alpha)^i, \quad 0 < \alpha < 1$$

- We call *exponential smoothing* because the weights decay exponentially.

Holt-Winters method

... or double exponential smoothing. Advised to use when there is a trend in the data.

$$\hat{X}_t(1) = S_t + \gamma B_t$$

$$S_t = \alpha X_t + (1-\alpha)(S_{t-1} + B_{t-1}), \quad \text{with } S_1 = X_1$$

$$B_t = \beta(S_t - S_{t-1}) + (1-\beta)B_{t-1}, \quad \text{with} B_1 = X_1 - X_0$$

Smoothing of beer data

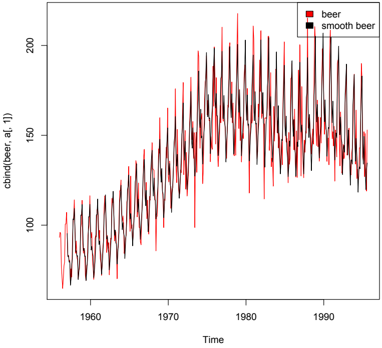


Figura : Double exponential smoothing of beer data, with $\alpha = 0.07532, \beta = 0.07435$ and $\gamma = 0.14388$

Prediction using smoothing

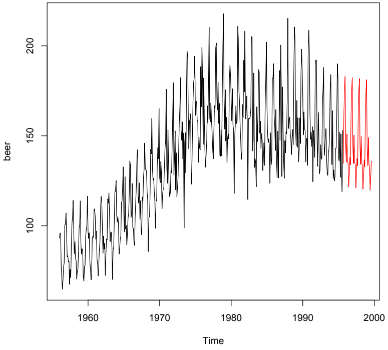


Figura : Prediction of the next 12 monthly values

Linear Time Series Models

Building block: white noise

The building block for the analysis of time series is the white noise process, hereby denoted by $\{\varepsilon_t\}$. In the least general case:

$$\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$$

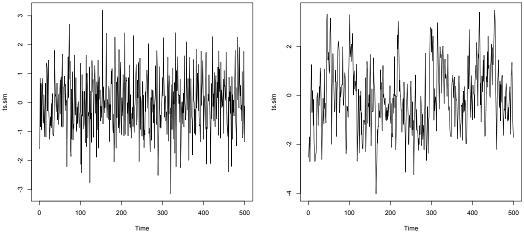
which has the following trivial but important consequences:

- $E[\varepsilon_t | \mathcal{F}_{t-1}] = 0, \forall t$ (stationary in mean)
- $E[\varepsilon_t \varepsilon_{t-j}] = Cov(\varepsilon_t, \varepsilon_{t-j}) = 0$
- $Var[\varepsilon_t | \mathcal{F}_{t-1}] = \sigma_\varepsilon^2, \forall t$ (conditional homoskedacity)

White noise in practice

- The sample ACFs should be close to zero
- The sample PACFs should be close to zero
- The sequence of observations should be random, without any visible pattern

Can you check which one is a sample-path of a white noise?



An example of a white noise

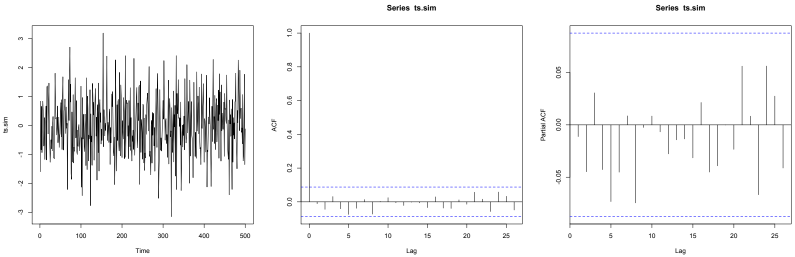


Figura : Data, sample acf and sample pacf

An example of a non-white noise

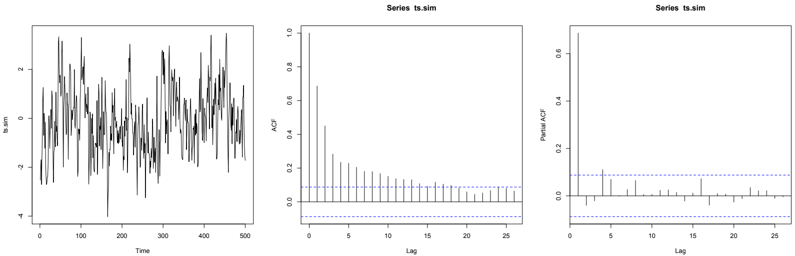


Figura : Data, sample acf and sample pacf

Linear time series

By itself, $\{\varepsilon_t\}$ is a pretty boring process. If ε_t is unusually high, there is no tendency for ε_{t+1} to be unusually high or low, so it does not capture the interesting property of persistence that motivates the study of time series. More realistic models are constructed by taking combinations of $\{\varepsilon_t\}$.

In particular, one such class of models are the *Linear Time Series*.

A time series $\{X_t\}$ is said to be linear if it can be written as:

$$X_t = \mu + \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

Some properties

- Stationarity

$$E[X_t] = \mu; \quad Var[X_t] = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i^2$$

- Because $Var[X_t] < \infty$, $\{\psi_i^2\}$ must be a convergent sequence, i.e., $\psi_i^2 \to 0$ when $i \to \infty$.
- Therefore, for such a stationary series, the impact of the remote shock ε_{t-i} on X_t vanishes as i increases.
- the lag-l autocovariance and autocorrelation:

$$\gamma_l = Cov(X_t, X_{t+l}) = \ldots = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+l}$$

$$\rho_l = Corr(X_t, X_{t+l}) = \frac{\sum_{i=0}^{\infty} \psi_i \psi_{i+l}}{\sum_{i=0}^{\infty} \psi_i^2}$$

Basic ARMA models

There are special linear time series models, the usually called ARMA models; for example:

$$
\begin{aligned}
\mathrm{AR}(1): &\quad X_t = \phi X_{t-1} + \varepsilon_t \\
\mathrm{MA}(1): &\quad X_t = \varepsilon_t + \theta \varepsilon_{t-1} \\
\mathrm{AR}(p): &\quad X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t \\
\mathrm{MA}(q): &\quad X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \ldots + \theta_q \varepsilon_{t-q} \\
\mathrm{ARMA}(p,q): &\quad X_t = \underbrace{\phi_1 X_{t-1} + \ldots + \phi_p X_{t-p}}_{\mathrm{AR}(p)} + \underbrace{\varepsilon_t + \ldots + \theta_q \varepsilon_{t-q}}_{\mathrm{MA}(q)}
\end{aligned}
$$

All these models are zero-mean, and are used to represent deviations of the series about a mean.
If a series has mean $\bar{X}$ and if it is an AR(1), then:

$$(X_t - \bar{X}) = \phi(X_{t-1} - \bar{X}) + \varepsilon_t \Leftrightarrow X_t = (1-\phi)\bar{X} + \phi X_{t-1} + \varepsilon_t$$

So we can assume mean zero, since adding means and deterministic trends is trivial...

Lag operators and polynomials

It is easier to represent and manipulate ARMA models in lag-operation notation.

- L operator: $LX_t = X_{t-1}$ (moves the index back one time unit)
- L^j operator: $L^j X_t = X_{t-j}$
- Δ operator: $\Delta X_t = X_t - X_{t-1} = (1-L)X_t$

Using this notation:

$$\begin{aligned} \text{AR}: &\quad a(L)X_t = \varepsilon_t \\ \text{MA}: &\quad X_t = b(L)\varepsilon_t \\ \text{ARMA}: &\quad a(L)X_t = b(L)\varepsilon_t \end{aligned}$$

Relationship between AR and MA models

We can manipulate the models as follows:

- AR(1) to MA(∞) by recursive substitution:

$$\begin{aligned} X_t &= \phi X_{t-1} + \varepsilon_t = \phi(\phi X_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \phi^2 X_{t-2} + \phi\varepsilon_{t-1} + \varepsilon_t \\ &= \dots \\ &= \phi^k X_{t-k} + \sum_{j=0}^{k-1} \phi^j \varepsilon_{t-j} \\ &= \sum_{j=0}^{\infty} \phi^j \varepsilon_{t-j} \end{aligned}$$

if $|\phi| < 1$.

- AR(1) to MA(∞) with lag polynomials:

$$\begin{aligned}(1-\phi L)X_t = \varepsilon_t \Leftrightarrow X_t &= (1-\phi L)^{-1}\varepsilon_t \\ &= (1+\phi L+\phi^2 L^2+\ldots)\varepsilon_t \\ &= \sum_{j=0}^{\infty}\phi^j \varepsilon_{t-j}\end{aligned}$$

This operation is not always admissible. If $|\phi| < 1$ then we say that the model is *stationary*.

- AR(2) to MA(∞):

$$\begin{aligned} X_t &= \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t \Leftrightarrow (1-\phi_1 L - \phi_2 L^2) X_t = \varepsilon_t \\ &\Leftrightarrow (1-\lambda_1 L)(1-\lambda_2 L) X_t = \varepsilon_t \Leftrightarrow X_t = (1-\lambda_1 L)^{-1}(1-\lambda_2 L)^{-1}\varepsilon_t \\ &= \left(\sum_{j=0}^{\infty} \lambda_1^j L^j\right)\left(\sum_{j=0}^{\infty} \lambda_2^j L^j\right)\varepsilon_t \end{aligned}$$

if $|\phi_1| < 1$, as well as for ϕ_2.
Is there a simpler way to compute this? YES...

Note that:

$$\frac{1}{(1-\lambda_1 L)(1-\lambda_2 L)} = \frac{a}{(1-\lambda_1 L)} + \frac{b}{(1-\lambda_2 L)}$$

with $a + b = 1 \Leftrightarrow \lambda_2 a + \lambda_1 b = 0$. So:

$$a = \frac{\lambda_1}{\lambda_1 + \lambda_2}; \quad b = \frac{\lambda_2}{\lambda_1 + \lambda_2}$$

and therefore:

$$X_t = \sum_{j=0}^{\infty} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2} \lambda_1^j + \frac{\lambda_2}{\lambda_1 + \lambda_2} \lambda_2^j \right) \varepsilon_{t-j}$$

Due to the latter, one can prove that the AR(2) is if and only if

$$\phi_2 + \phi_1 < 1$$
$$\phi_2 - \phi_1 < 1$$
$$-1 < \phi_2 < 1$$

- MA(q) to AR(∞):
  Using the same procedure as in the AR case:

$$X_t = b(L)\varepsilon_t \Leftrightarrow b^{-1}(L)X_t = \varepsilon_t$$

When the operation is possible, we say that the model is invertible.

Summary of allowed lag manipulations

You can play with these lag-polynomials as if L would be a number!

- You can multiple them: $a(L)b(L) = a_0 b_0 + (a_1 b_0 + a_0 b_1)L + \ldots$
- They commute: $a(L)b(L) = b(L)a(L)$
- You can raise them to positive powers: $a^2(L) = a(L)a(L)$
- You can invert them, starting by factoring:$a^{-1}(L) = \frac{c_1}{1-\lambda_1 L} + \frac{c_2}{1-\lambda_2 L} + \ldots$

AR(1)

$$X_t = \phi X_{t-1} + \varepsilon_t \Leftrightarrow X_t = \sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}$$

- $E[X_t] = E[\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}] = 0$
- $Var[X_t] = Var[\sum_{i=0}^{\infty} \phi^i \varepsilon_{t-i}] = \frac{\sigma_\varepsilon^2}{1-\phi^2}$
- $\rho_l = Corr(X_t, X_{t+l}) = \rho^l$.
- PACF: for $k \geq 2$, it is zero! (so only $\phi_{11} = \rho_1 = \phi \neq 0$)

AR(1)

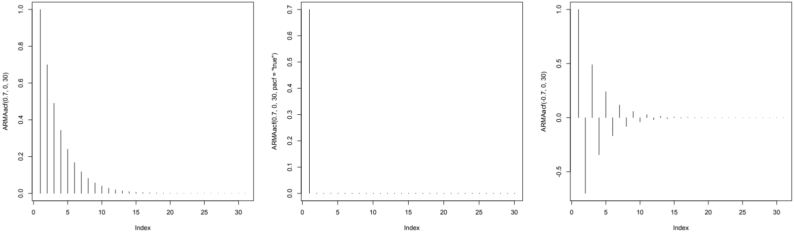


Figura : ACF and PACF for an AR(1), with $\phi = 0.7$, and ACF for an AR(1), with $\phi = -0.7$

AR(2): $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$

- As the model is stationary (under some assumptions...), we have that $E[X_t] = E[X_{t-1}] = E[X_{t-2}]$. Therefore:

$$E[X_t] = \phi_1[X_{t-1}] + \phi_2 E[X_{t-2}] \Rightarrow E[X_t] = 0$$

- Regarding the covariance:
  - $\rho_1 = Corr(X_t, X_{t-1}) = \phi_1 + \phi_2\rho_1 \Leftrightarrow \rho_1 = \frac{\phi_1}{1-\phi_2}$
  - $\rho_k = Corr(X_t, X_{t-k}) = \phi_1\rho_{k-1} + \phi_2\rho_{k-2}, k \geq 2$ (the moment equation)

Can we solve it in a simple way?

Linear Difference Equations

Linear difference equations play an important role in the time series models because the relationship between $\{X_t\}$ and $\{\varepsilon_t\}$ is in the form of a

$\rightarrow$ *Linear Difference Equation Models*.

- General nth-order linear difference equation:

$$C_0X_t + C_1X_{t-1} + \ldots + C_nX_{t-n} = \varepsilon_t$$

  (C_i are constants; set, w.l.g, $C_0 = 1$) The above equation is said to be nonhomogeneous if $\varepsilon_t = 0$; otherwise it is said homogeneous.

- Let $C(L) = (1 + C_1L + \ldots + C_nL^n)$. The equation

$$C(L)X_t = 0$$

  is called the auxiliary equation.

Solution of Linear Difference Equations

Case 1: the auxiliary equation has only one root, with multiplicity m

- Let $(1-L)^m X_t = 0$. Then a general solution is given by

$$X_t = \sum_{j=0}^{m-1} b_j t^j$$

- Let $(1-aL)^m X_t = 0$, with $a \neq 1$. Then a general solution is given by

$$X_t = \sum_{j=0}^{m-1} b_j t^j a^t$$

Case 2: the auxiliary equation has has more than one root

- Let $C(L)X_t = 0$. If $C(L) = \prod_{i=1}^{N}(1 - R_i L)^{m_i}$, with $\sum_{i=1}^{M} m_i = n$, then

$$X_t = \sum_{i=1}^{N} \sum_{j=0}^{m_i - 1} b_{ij} t^j R_i^t$$

In particular, if $m_i = 1$ and R_i are all distint, then

$$X_t = \sum_{i=1}^{n} b_i R_i^t$$

Solution of the autocorrelation-function

How do we solve the second order difference equation:

$$(1 - \phi_1 L - \phi_2 L^2)\rho_l = 0?$$

- Factorize the polynomial
  $(1 - \phi_1 L - \phi_2 L^2) = (1 - \lambda_1 L) \times (1 - \lambda_2 L)$
- In general, the solution to the difference equation is given by:

$$\rho_k = a_1 \lambda_1^k + a_2 \lambda_2^k = \frac{\lambda_1(1 - \lambda_2^2)\lambda_1^k - \lambda_2(1 - \lambda_1^2)\lambda_2^k}{(\lambda_1 - \lambda_2)(1 + \lambda_1 \lambda_2)}$$

In case λ_1 and λ_2 are real valued and different, the ACF consists of two damped exponentials. If they are complex, the series is said to behave pseudo-periodic.

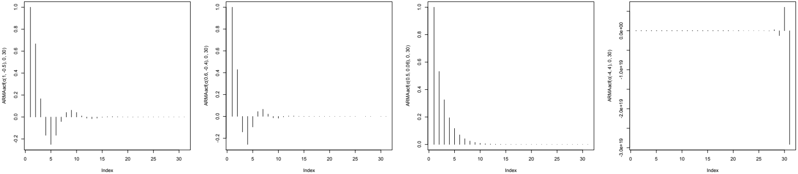


Figura : ACF AR(2): (i)$(1-0.5L)^2$, (ii) complex roots; (iii) stationary; (iv) non-stationary

PACF

Easy: it is zero for all $k \geq 3$!

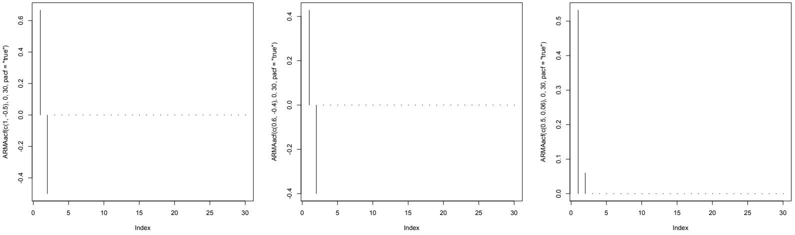


Figura : PACF AR(2): (i)$(1-0.5L)^2$, (ii) complex roots; (iii) stationary

MA(k): $X_t = (1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_k L^k)\varepsilon_t$

- $E[X_t] = 0$
- The ACF is zero for all lags $\geq (k+1)$:

$$\rho_q = \begin{cases} cc\frac{-\theta_q + \sum_{i=1}^{k}\theta_i\theta_{i+q}}{1+\sum_{i=1}^{k}\theta_i^2} & k \leq q \\ 0 & k > q \end{cases}$$

- The partial correlation function behaves as the corresponding ACF for the AR(k) (expressions are too involved!; see the Yule-Walker equations)

So the behaviours of the ACF and PACF of the MA(k) are the same as the ones for the PACF and ACF of the AR(k).

ARMA(1): $X_t = \phi X_{t-1} + \theta_1 \varepsilon_{t-1} + \varepsilon_t$

Properties of ARMA(1,1) models are generalizations of those of AR(1) models, with some minor modifications to handle the impact of the MA(1) component,

- $E[X_t] = 0$
- $Var[X_t] = \frac{(1+\theta_1^2 - 2\phi_1\theta_1)\sigma_\varepsilon^2}{1-\phi_1^2}$
- $\rho_1 = \phi_1 - \frac{\theta_1 \sigma_\varepsilon^2}{Var[X_t]}$ and $\rho_l = \phi_1 \rho_{l-1}$ Thus the ACF behaves very much like that of an AR(1).
- The PACF does not cut off at any finite lag, and it behaves very much like that of an MA(1).

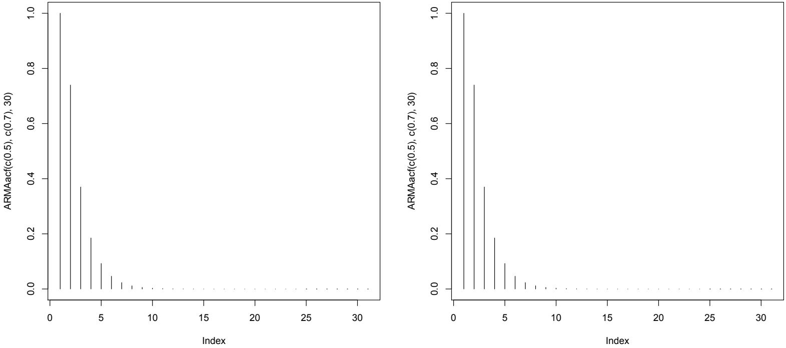


Figura : ACF and PACF of an ARMA(1,1): $X_t = 0.5X_{t-1} + 0.7\varepsilon_{t-1} + \varepsilon_t$

Identifying ARMA models in practice

The order p of an AR and q of the MA time series are unknown, and must be specified empirically (order determination).

There are two ways of doing so:

- Use empirical ACF and PACF
- Use some information criterion function (AIC or BIC)

Use of ACF and PACF

- To identify a reasonable appropriate ARMA model, we need a minimun of 50 observations;
- The number of sample ACF and PACF should be about $\frac{T}{4}$;
- Match the patterns in the sample ACF and PACF with the theoretical ones.

Process	ACF	PACF
AR(p)	Tails of as exponential decay or damped sine wave	Cuts off after lag p
MA(q)	Cuts off after lag q	Tails of as exponential decay or damped sine wave
ARMA(p, q)	Tails off after lag $(p-q)$	Tails off after lag $(p-q)$

Some good advices

- In the initial model identification, concentrate on the general broad features of the sample ACF and PACF without focusing on details.
- Use a conservative threshold of 1.5 standard deviations in the significance of short-term lags ACF and PACF, specially whe using short series.

Information Criteria

There are two information criteria usually used, both likelihood based...

- *Akaike information criterion* (AIC):

$$AIC = \ln(\widehat{\sigma_\varepsilon}^2) + \frac{2(p+q)}{T}$$

(where $\widehat{\sigma_\varepsilon}^2$) is the m.l.e. of σ_ε^2, T is the sample size))

- *Bayesian information criterion* (BIC):

$$BIC = \ln(\widehat{\sigma_\varepsilon}^2) + \frac{(p+q)\ln(T)}{T}$$

The penalty for each parameter used is 2 for AIC and $\ln(T)$ for BIC. Thus BIC tends to select a lower ARMA model, when the sample size is moderate or large.

Going back to the TUI data....

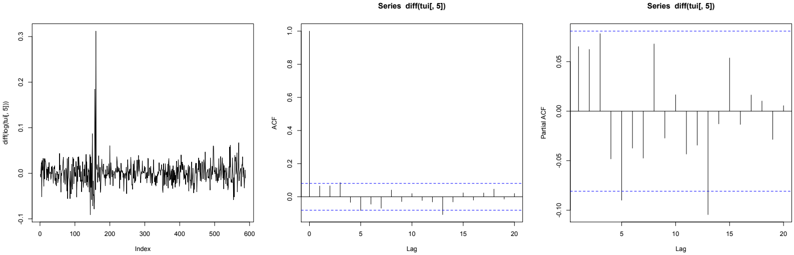


Figura : Series of first difference, sample ACF and sample PACF of TUI data

Going back to the TUI data....

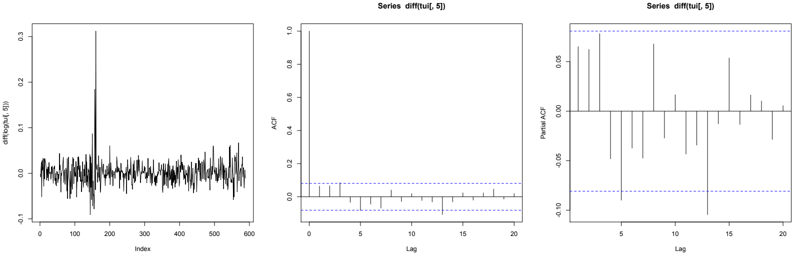
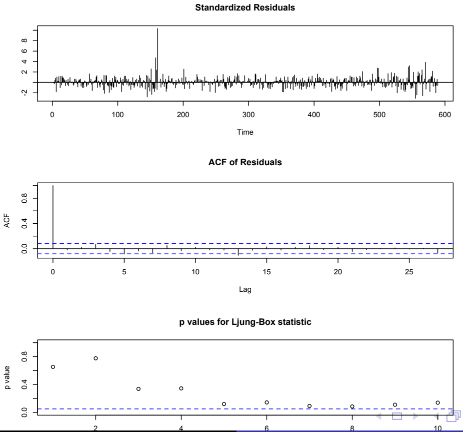


Figura : Series of first difference, sample ACF and sample PACF of TUI data

We try three models: ARMA(1,1), AR(1), AR(3)

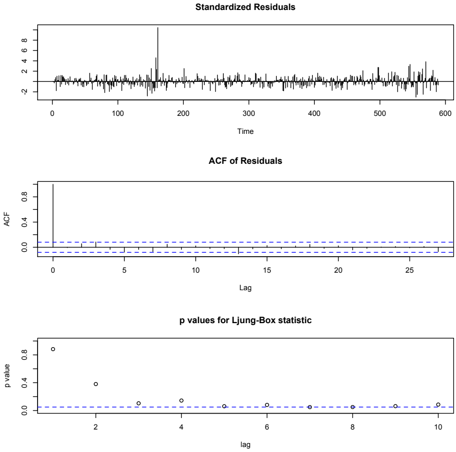
ARMA(1,1):

$$X_t = 0.544X_{t-1} - 0.466\varepsilon_{t-1} + \varepsilon_t$$



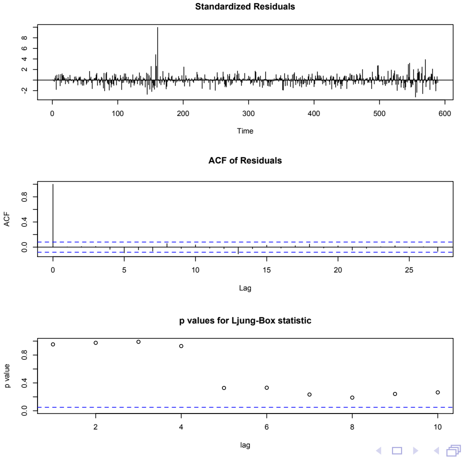
AR(1):

$$X_t = 0.067X_{t-1} + \varepsilon_t$$



AR(3):

$$X_t = 0.058X_{t-1} + 0.058X_{t-2} + 0.080X_{t-3}\varepsilon_t$$



ARMA(1,1): AIC=1500.3
AR(3): AIC=1500.68
AR(1): AIC=1492.52

Notation

One of the most interesting things to do with an ARMA model is to predict future values, given the past.
Notation:

- Predictor of X_{t+l}: $\hat{X}_t(l) = E[X_{t+l} | X_t, X_{t-1}, \ldots, \varepsilon_t, \varepsilon_{t-1}, \ldots]$
- Variance of the predictor: $Var_t(l) = Var[X_{t+l} | X_t, X_{t-1}, \ldots, \varepsilon_t, \varepsilon_{t-1}, \ldots]$

Let $\mathcal{F}_t = \sigma(X_t, X_{t-1}, \ldots, \varepsilon_t, \varepsilon_{t-1}, \ldots)$.

Confidence Intervals for the forecast

The forecast error is given by:

$$e_t(l) = X_{t+l} - \hat{X}_t(l)$$

with variance given by:

$$Var[e_t(l)] = \sigma_\varepsilon^2 \sum_{j=0}^{l-1} \psi_j^2$$

where ψ_j is the of ε_{t+l-j} in the MA(∞) representation.
Then the $(1-\alpha) \times 100\%$ forecast limits are:

$$\hat{X}_t(l) \pm \Phi^{-1}(1 - \frac{\alpha}{2)})\sqrt{Var[e_t(l)]}$$

Eventual forecast functions

- Solve the usual auxiliary function $a(L)\hat{X}_t(l) = 0$ (the same one as you need to solve in order to get the MA(∞) and the ACF).
- As we have seen previously, the general solution to this equation is (see case 2, for instance):

$$\hat{X}_t(l) = \sum_{i=1}^{N} \left(\sum_{j=0}^{m_i - 1} c_{ij} l^j R_i^l \right.$$

Predicting AR(1)

$$X_{t+1} = \phi X_t + \varepsilon_t$$

Now apply conditionl expected value:

- $\hat{X}_t(1) = \phi X_t$
- $\hat{X}_t(2) = E[\phi X_{t+1} | \mathcal{F}_t] = \phi \hat{X}_t(1) = \phi^2 X_t$
- $\hat{X}_t(k) = \phi \hat{X}_t(k-1) = \phi^k X_t$

Notice that

$$\lim_{k \to \infty} \hat{X}_t(k) = 0$$

Regarding the variance of the predictors:

- $Var_t(1) = Var[X_{t+1}|\mathcal{F}_t] = Var[\phi X_t + \sigma_\varepsilon^2] = \sigma_\varepsilon^2$
- $Var_t(2) = Var[\phi^2 X_t + \phi X_{t+1} + \varepsilon_{t+1}|\mathcal{F}_t] =$ $Var[\phi X_{t+1} + \varepsilon_{t+1}] = (1+\phi^2)\sigma_\varepsilon^2$
- $Var_t(k) = \sum_{i=0}^{k-1} \phi^2 \sigma_\varepsilon^2$

Predicting MA

$$X_{t+1} = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \varepsilon_t$$

Forecasting MA models is similarly easy as in the AR models.

- $\hat{X}_t(1) = \sum_{i=1}^{\infty} \theta_i \varepsilon_{t-i+1}$
- $\hat{X}_t(k) = \sum_{i=k}^{\infty} \theta_i \varepsilon_{t-i+k}$
- $Var_t(1) = \sigma_\varepsilon^2$
- $Var_t(k) = (1 + \theta_1^2 + \ldots + \theta_{k-1}^2)\sigma_\varepsilon^2$

Predicting AR's and MA

$$X_{t+1} = \phi_1 X_t + \phi_2 X_{t-2} + \ldots + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \varepsilon_t$$

Exploit the facts that:

$$E[\varepsilon_{t+j}|\mathcal{F}_t] = 0, \quad Var[\varepsilon_{t+j}|\mathcal{F}_t] = \sigma_\varepsilon^2$$

For that, express X_{t+j} as the following sum:

$$X_{t+j} = \{\text{function of}\varepsilon_{t+j}, \varepsilon_{t+j-1}, \ldots\} + \{\text{function of}\varepsilon_t, \varepsilon_{t-1}, \ldots\}$$

It is easier to express forecasts of AR's and ARMA's by inverting to their MA(∞).

Which ARMA processes are stationary?

- MA processes: $X_t = \sum_{j=0}^{\infty} \theta_j \varepsilon_{t-j} \Rightarrow Var[X_t] = \sum_{j=0}^{\infty} \theta_j^2 \sigma_\varepsilon^2$ then

$$\text{MA processes are stationary} \Leftrightarrow \sum_{j=0}^{\infty} \theta_j^2$$

- AR(1): converting to an MA(∞), we get $X_t = \sum_{j=0}^{k} \phi^j \varepsilon_{t-j} + \phi^k X_{t-k}$. So

$$\text{AR(1) is stationary} \Leftrightarrow |\phi| < 1$$

- AR processes: considering the MA(∞) representation, given by $(1-\lambda_1 L)(1-\lambda_2 L)\ldots X_t = \varepsilon_t$, then:

$$\text{AR(p) is stationary} \Leftrightarrow |\lambda_i| < 1, \forall i$$

ARIMA models

- A homogeneous nonstationary time series can be reduce to a stationary time series by taking a proper degree of differencing.
- This leads to a new class of models: the autoregressive integrated moving average models, the ARIMA class.
- Model equation for an ARIMA (p, d, q):

$$a_p(L)(1-L)^d X_t = b_q(L)\varepsilon_t$$

- Then the new series $\{Y_t = (1-L)^d X_t\}$ is an ARMA(p, q) model. Work this series and then go back to the original one, by simple transformation.

SARIMA models

- Many business and economic time series contain a seasonal phenomenon that repeats itself after a regular period of time.
- The smallest period for this repetitive phenomenon is called the seasonal period.
- If one believes that the seasonal, trend and irregular are additive, then one may perform a decomposition of the data, as illustrated before.

$$X_t = T_t + S_t + \varepsilon_t$$

- But usually the seasonal component is not independent of the other nonseasonal components. So we need to extend the ARIMA models in order to take into account the seasonality.

SARIMA$(p, d, q)(P, D, Q)_s$:

$$a_P(L^s)a_p(L)(1-L)^d(1-L^s)^D X_t = b_Q(L^s)b_q(L)\varepsilon_t$$

Example

A realization of a series with seasonality 12...

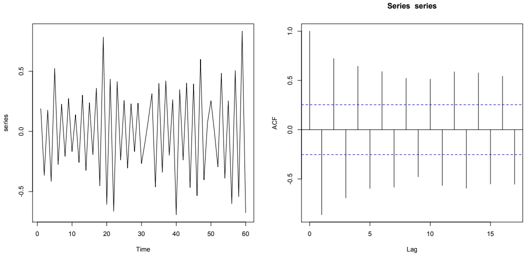


Figura : Series and sample ACF

If we fit a nonseasonal ARIMA model...

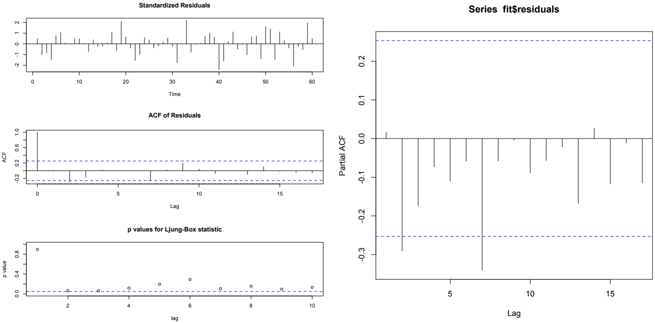


Figura : Analysis of residuals

Model identification

SARIMA$(p, d, q)(P, D, Q)_s$:

$$a_P(L^s)a_p(L)(1-L)^d(1-L^s)^D X_t = b_Q(L^s)b_q(L)\varepsilon_t$$

- Start by identifying the period s; luckily, it will be evident from the data and from the sample ACF...
- Probably you need to perform regular differencing $(1-L)$ and seasonal differencing $(1-L^s)$.
- Compute the ACF and PACF from $\{(1-L)^d(1-L^s)^D X_t\}$. Check pics for the seasonal component and for the non-seasonal component, to help identifying the order of the AR and MA components.
- Proceed as with the non-seasonal models.

Heterocedascity

- One of the main assumption assumed so far is that the variance of the erros is constant.
- In many practical applications, this assumption is not realistic (for example, in financial investment, stock markets volatility is rarely constant over time)
- We need models that incorporate the possibility of a nonconstant error variance.
- Homocedastic model: ε_t, with $Var[\varepsilon_t] = \sigma^2$
- Heterocedastic model: $\eta_t = \sigma_t \varepsilon_t$, with $Var[\epsilon_t] = 1$, and

$$\sigma_t^2 = 1 + \theta_1 \eta_{t-1}^2 + \theta_2 \eta_{t-2}^2 + \ldots + \theta_s \eta_{t-s}^2$$

Therefore in these models, the error term, that in the homocedastic model follows a white noise process, follows an AR(s) model.

Autoregressive conditional heterocedasticity model, ARCG(s)

GARCH

It may also happen that the noise follows an ARMA model:

$$\sigma_t^2 = 1 + \phi_1 \sigma_{t-1}^2 + \ldots, \phi_r \sigma_{t-r}^2 + \theta_1 \eta_{t-1}^2 + \theta_2 \eta_{t-2}^2 + \ldots + \theta_s \eta_{t-s}^2$$

In that case we have a GARCH(r, s).