Time Series Analysis

Venkatramani Rajgopal

Department of Mathematics
University of Applied Sciences, Mittweida

3 Nov 2016

Outline

1. Introduction
2. Classical regression in time series context
3. Preliminaries
   - White noise and Moving averages
   - Stationary Series
   - Autocorrelation function (ACF)
   - Partial - Autocorrelation function (PACF)
4. First-order Autoregression Model (AR(1))
5. Moving Average (MA) model
6. ARMA model
7. Fitting the model (Summary)

Introduction

Time Series

A collection of random variables indexed according to the order they are obtained in time.

Introduction

Time Series

A collection of random variables indexed according to the order they are obtained in time.

- Takes into account the internal structure of data points, such as *autocorrelation, trend or seasonal variations*
- Modeling relationships using data collected over time. For eg Stock Price, Index Closings, GDP etc.

Introduction

Motivation

In the context of;

- Statistics, econometrics, quantitative finance and geophysics the primary goal of time series analysis is forecasting.
- Signal processing, it is used for signal detection and estimation,
- Data mining, pattern recognition and machine learning time series analysis can be used for clustering and classification.

Introduction

Methods

Methods for time series analysis may be divided into two classes: frequency-domain methods and time-domain methods.

Introduction

Methods

Methods for time series analysis may be divided into two classes: frequency-domain methods and time-domain methods.

- *Frequency domain:* include spectral analysis and wavelet analysis;
- *Time domain:* include auto-correlation and cross-correlation analysis.

Models in time domain:

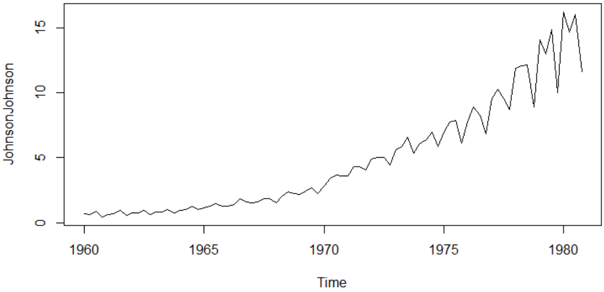
Three broad model classes of practical importance are the autoregressive (AR) models, the moving average (MA) models. These two classes depend linearly on previous data points.

Combinations of these ideas produce autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) models.

Introduction

Nature of Time series data

We will try to handle some *ts* data in R using the inbilt dataset in R (JohnsonJohnson).



data(JohnsonJohnson)

plot(JohnsonJohnson,type="l")

Classical regression in time series context

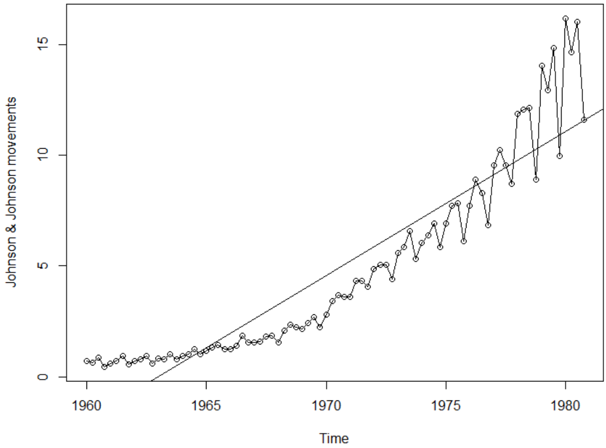
Linear regression in the time series context can be done by assuming some output or *time dependent series*, say x_t for $t = 1, 2 \ldots n$ which are being influenced by a collection of possible inputs or independent series, say $z_{t1}, z_{t2} \ldots z_{tq}$.
Expressed as,

$$x_t = \alpha + \beta_1 z_{t1} + \beta_2 z_{t2} + ...\beta_q z_{tq} + w_t \tag{1}$$

where,
$\beta_1, \beta_2....\beta_q$ are the regression coefficients
w_t is the is the random error.

Classical regression in time series context

A time series plot of the JohnsonJohnson Stock price and the estimated trend line obtained via simple linear regression.



Preliminaries

White noise and Moving averages

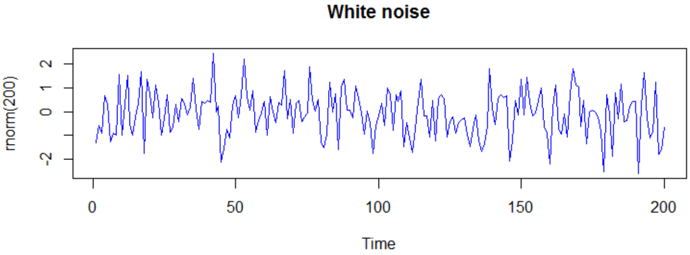
- In discrete time, white noise is a discrete signal whose samples are regarded as a sequence of serially *uncorrelated random variables*, w_t , with mean 0 and variance σ^2.
- Example: A single realization of white noise is a random shock.
- A white noise process is one which has no correlation between its values at different times.

Preliminaries

White noise and Moving averages

- In discrete time, white noise is a discrete signal whose samples are regarded as a sequence of serially *uncorrelated random variables*, w_t , with mean 0 and variance σ^2.
- Example: A single realization of white noise is a random shock.
- A white noise process is one which has no correlation between its values at different times.

plot.ts(rnorm(200),col="blue",main="White noise") in R.



Preliminaries

White noise and Moving averages

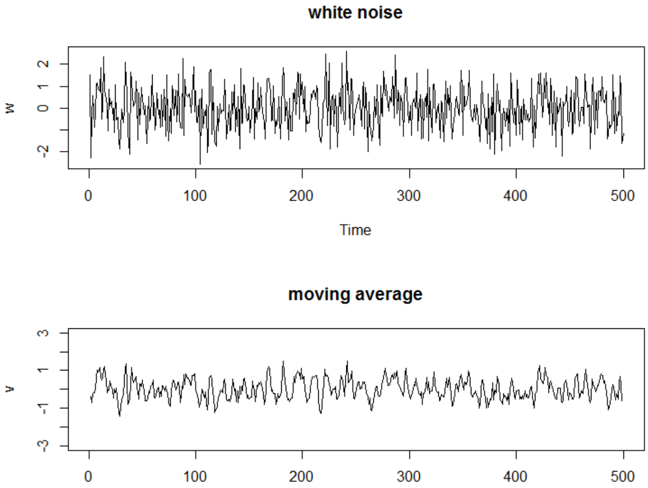
The white noise w_t can be replaced by a moving average that *smooths* the series.
Consider w_t as a average of its current value and its immediate neighbours.

$$v_t = \frac{1}{3}(w_{t-1} + w_t + w_{t+1}) \tag{2}$$

Preliminaries

White noise and Moving averages

Gaussian white noise series and three-point moving average of the Gaussian white noise series could like this.



Equation (3) can be plotted in R using the *filter* function in R.

```
w = rnorm(500,0,1)  # 500 N(0,1) variates
v = filter(w, sides=2, rep(1/3,3))  # moving average
par(mfrow=c(2,1))
plot.ts(w, main="white noise")
plot.ts(v, ylim=c(-3,3), main="moving average")
```

Modelling a time series (ARMA) model requires stationarity.

Modelling a time series (ARMA) model requires stationarity.

Weak Stationary

A series x_t is said to be (weakly) stationary if it satisfies the following properties:

- The mean $E(x_t)$ is same for all t
- The variance of (x_t) is same for all t
- The covariance and the correlation between x_t and x_{t-h} is the same for all t.

Preliminaries

Stationary Series

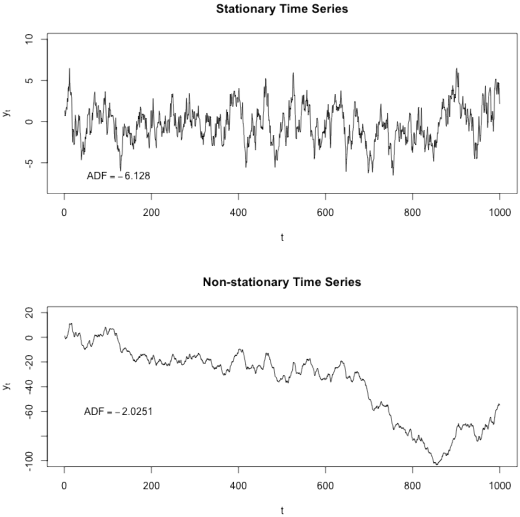


Figure: Stationary vs Non-stationary

Preliminaries

Stationary Series

Addressing non-stationary series.

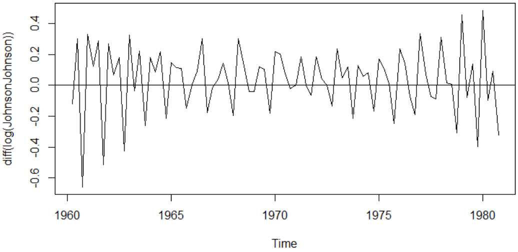
Preliminaries

Stationary Series

Addressing non-stationary series.

- Remove unequal variances. We do this using log of the series.
- We need to address the trend component. We do this by taking *difference* of the series.
- Differenced variable: $\Delta x_t = x_t - x_{t-1}$

This can be done using diff(log(JohnsonJohnson)) in R.



Augmented Dickey–Fuller test **(ADF)** tests the null hypothesis of whether a unit root is present in a time series sample. The alternative hypothesis is usually stationarity or trend-stationarity.
The testing procedure for the ADF test is applied to the AR(1) model;

$$x_t = \rho x_{t-1} + e_t \tag{3}$$

$$\begin{aligned} x_t &: \text{is the variable of interest} \\ \rho &: \text{coefficient on a time trend} \\ e_t &: \text{error term} \end{aligned}$$

A unit root is present if $\rho = 1$. The model would be non-stationary in this case.

Preliminaries

Augmented Dickey–Fuller test

Estimation : We estimate if null-hypothesis is not rejected, then x_t is not stationary.

Testing ADF on our data JohnsonJohnson, in R, we have the following result.

```
> adf.test(diff(log(JohnsonJohnson)), alternative="stationary", k=0)

        Augmented Dickey-Fuller Test

data:  diff(log(JohnsonJohnson))
Dickey-Fuller = -15.613, Lag order = 0, p-value = 0.01
alternative hypothesis: stationary
```

Preliminaries

Autocorrelation function (ACF)

The ACF is a way to measure the linear relationship between an observation at time t and the observations at previous times.

- x_t denote the value of a time series at time t.
- The ACF of the series gives correlations between x_t and x_{t-k} for $k = 1, 2, 3$, etc.

Theoretically, the autocorrelation between x_t and x_{t-k} is,

$$ACF(k) = \frac{\text{Covariance}(x_t, x_{t-k})}{\sigma_{x_t}.\sigma_{x_{t-k}}} = \frac{\text{Covariance}(x_t, x_{t-k})}{\text{Variance}(x_t)}$$

Preliminaries

Autocorrelation function (ACF)

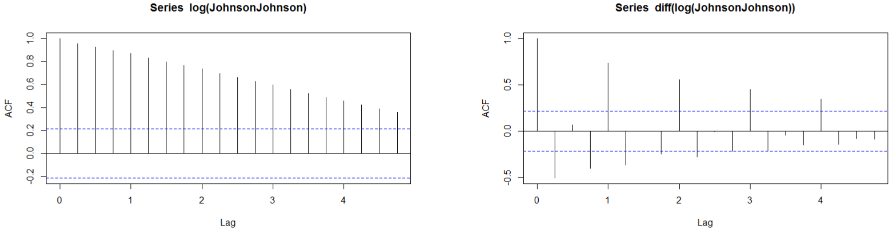


Figure: ACF of a non-stationary vs stationary series

The above ACF is “decaying”, or decreasing, very slowly, and remains well above the significance range (dotted blue lines). This is indicative of a non-stationary series.

On the other hand ACF of a stationary series shows exponential decay. This is indicative of a stationary series.

Preliminaries

Partial - Autocorrelation function (PACF)

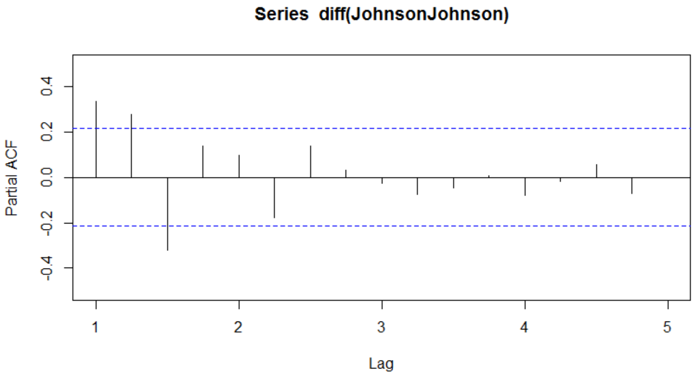
PACF is a simple correlation between x_t and x_{t-k}, minus the part explained by the intervening lags.

$$\rho_k = Corr(x_t - E(x_t|x_{t-1}....x_{t-k+1}), x_{t-k})$$

where;
$E(x_t|x_{t-1}....x_{t-k+1})$ is the minimum squared error predictor.
The 2nd order (lag) partial autocorrelation is;

$$\frac{\text{Covariance}(x_t, x_{t-2}|x_{t-1})}{\sqrt{\text{Variance}(x_t|x_{t-1})\text{Variance}(x_{t-2}|x_{t-1})}} \tag{4}$$

PACF Plots



Series diff(JohnsonJohnson)
Partial ACF
-0.4
-0.2
0.0
0.2
0.4
1
2
3
4
5
Lag

First-order Autoregression Model (AR(1))

AR(1) model is a linear model that predicts the present value of a time series using the immediately prior value in time. In this model, the value of x at time t is a linear function of the value of x at time t–1. Represented as;

$$x_t = \beta_0 + \beta_1 x_{t-1} + \epsilon_t$$

First-order Autoregression Model (AR(1))

AR(1) model is a linear model that predicts the present value of a time series using the immediately prior value in time. In this model, the value of x at time t is a linear function of the value of x at time t–1. Represented as;

$$x_t = \beta_0 + \beta_1 x_{t-1} + \epsilon_t$$

Assumptions:
$\epsilon_t \overset{iid}{\sim} N(0, \sigma^2)$ i.e errors are independently distributed with a normal distribution that has mean 0 and constant variance.
An autoregressive model of order p, AR(p), is of the form;

$$x_t = \beta_0 + \beta_1 x_{t-1} + \beta_2 x_{t-2} + ...\beta_p x_{t-p} + \epsilon_t \tag{5}$$

First-order Autoregression Model (AR(1))

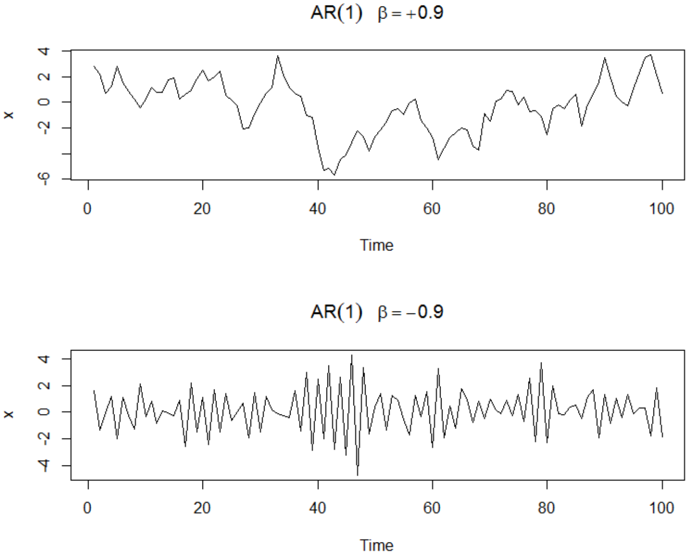
The mean of x_t is zero. If the mean μ of x_t is not zero, we replace x_t by $x_{t-\mu}$ in Equation(4) and write as;

$$x_t = \alpha + \beta_1 x_{t-1} + \beta_2 x_{t-2} + ...\beta_p x_{t-p} + \epsilon_t \tag{6}$$

where $\alpha = \mu(1 - \beta_1..... - \beta_p)$.
We note that this equation is similar to the regression model of (1) and hence the term *auto (or self) regression.*

First-order Autoregression Model (AR(1))

Sample path of an AR(1) process



First-order Autoregression Model (AR(1))

Sample path of an AR(1) process

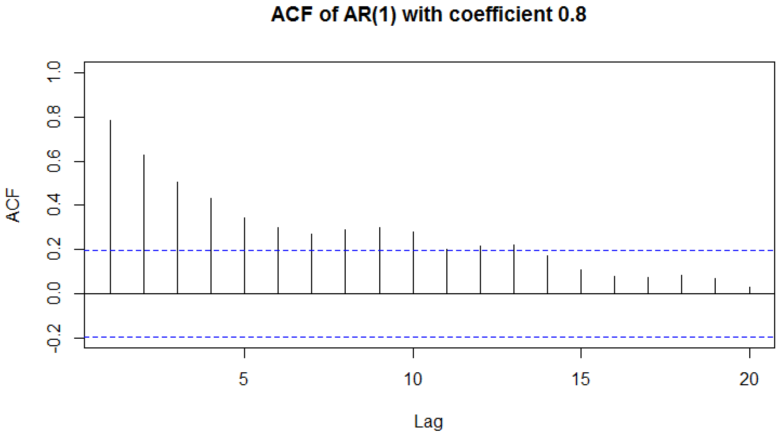
Plotted using the below R codes.

```
par(mfrow=c(2,1))
plot(arima.sim(list(order=c(1,0,0), ar=.9), n=100), ylab="x",main=(expression(AR(1)~~~beta==+.9)))
plot(arima.sim(list(order=c(1,0,0), ar=-.9), n=100), ylab="x",main=(expression(AR(1)~~~beta==-.9)))
```

First-order Autoregression Model (AR(1))

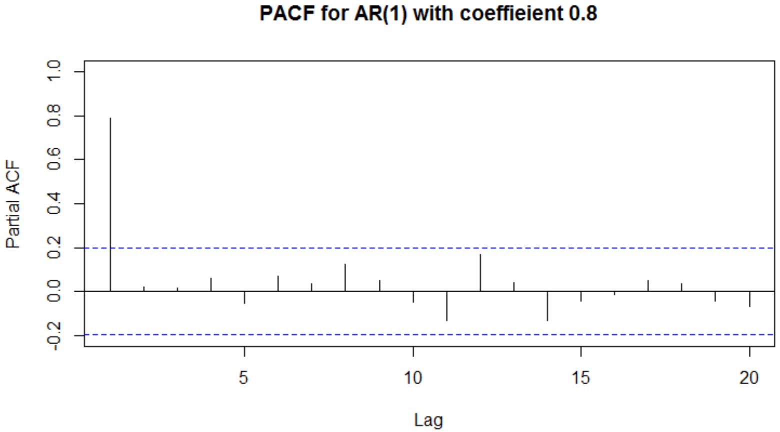
ACF for AR(1)

For an AR(1) model, the ACF is $\text{ACF(k)} = \rho_k = \beta^k$.
We say this function tails off.



First-order Autoregression Model (AR(1))

PACF for AR(1)



Moving Average (MA) model

The moving-average model specifies that the output variable depends linearly on the current and various past values.
A first order moving average model, denoted by MA(1) is given by;

$$x_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1}$$

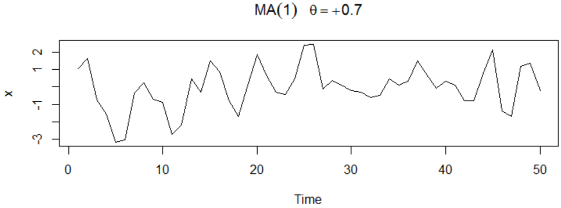
and the *qth* order moving average model with q lags, **MA(q)** is given by;

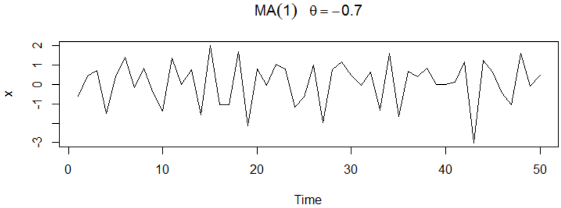
$$x_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q}$$

where there are q lags in the moving average and $\theta_1, \theta_2 .... \theta_q$ are parameters.

Moving Average (MA1) process

When for e.g $\theta = 0.7$, x_t and x_{t-1} are positively correlated and when $\theta = -0.7$, they are negatively correlated.





The above plot shows that the series is *smoother* when $\theta = 0.7$.

Moving Average (MA1) process

R code.

```
par(mfrow = c(2,1))
plot(arima.sim(list(order=c(0,0,1), ma=.7), n=50), ylab="x",main=(expression(MA(1)~~~theta==+.7)))
plot(arima.sim(list(order=c(0,0,1), ma=-.7), n=50), ylab="x",main=(expression(MA(1)~~~theta==-.7)))
```

ARMA Model

Autoregressive moving average (ARMA) models combine both p autoregressive terms and q moving average terms, also called ARMA(p,q) given by,

$$x_t = \beta_1 x_{t-1} + ... + \beta_p x_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q} \tag{7}$$

Fitting the model (Summary)

Three items should be considered to determine a first guess at an ARIMA model: a time series plot of the data, the ACF, and the PACF.

- Plot the data. Identify any unusual observations.
- If the data are non-stationary: take first differences of the data until the data are stationary.
- For data with a curved upward trend accompanied by increasing variance, consider transforming the series with either a logarithm or a square root.
- Examine the ACF/PACF: Is an AR(p) or MA(q) model appropriate?

References I

- Robert H. Shumway , David S. Stoffer
  Time Series analysis and its applications
- Rob J Hyndman, George Athanasopoulos
  Forecasting: Principles and practice
- The Pennsylvania State University - STAT 510 - Applied Time Series analysis
  https://onlinecourses.science.psu.edu/stat510/
- R Documentation
  https://www.rdocumentation.org/

Thank You for your attention.