STA6856

Time Series Analysis

Achraf Cohen
acohen@uwf.edu

Department of Mathematics and Statistics
The University of West Florida

- This course covers Time Series models.
- We will use R/RStudio
- Assignments 3-4
- Two take-home exams

These slides are being updated! if you catch a typo/error please send an email to acohen@uwf.edu! Thank you!

Definitions

A Time Series is a collection of observations x_t made sequentially in time.
A discrete-time time series is a collection of observations x_t in which the set T_0 of times at which observations are made is a discrete set.
A continuous-time time series is a collection of observations x_t are made continuously over some time interval.

Definitions

A Time Series is a collection of observations x_t made sequentially in time.

A discrete-time time series is a collection of observations x_t in which the set T_0 of times at which observations are made is a discrete set.

A continuous-time time series is a collection of observations x_t are made continuously over some time interval.

This course covers discrete-time time series

Introduction

Time

We mean by *time*:

- Seconds, hours, years,...

Introduction

Time

We mean by *time*:

- Seconds, hours, years,...
- Spatial: 1^{st} machine in a row, 2^{nd} machine,...

Introduction

Time

We mean by *time*:

- Seconds, hours, years,...
- Spatial: 1^{st} machine in a row, 2^{nd} machine,...
- Depth: one millimetre down, two millimetre down,...

Time

We mean by *time*:

- Seconds, hours, years,...
- Spatial: 1^{st} machine in a row, 2^{nd} machine,...
- Depth: one millimetre down, two millimetre down,...

The important point here is to have an *ordered variable like time* that there is a meaning of direction in its values. Then from a given observation, *past, present and future have a meaning.*

Introduction

Plotting a time series is an important early step in its analysis
In general, a plot can reveal:

- *Trend*: upward or downward pattern that might be extrapolated into the future

Introduction

Plotting a time series is an important early step in its analysis
In general, a plot can reveal:

- *Trend*: upward or downward pattern that might be extrapolated into the future
- *Periodicity*: Repetition of behavior in a regular pattern

Introduction

Plotting a time series is an important early step in its analysis
In general, a plot can reveal:

- *Trend*: upward or downward pattern that might be extrapolated into the future
- *Periodicity*: Repetition of behavior in a regular pattern
- *Seasonality*: Periodic behavior with a known period (hourly, monthly, every 2 months...)

Introduction

Plotting a time series is an important early step in its analysis
In general, a plot can reveal:

- *Trend*: upward or downward pattern that might be extrapolated into the future
- *Periodicity*: Repetition of behavior in a regular pattern
- *Seasonality*: Periodic behavior with a known period (hourly, monthly, every 2 months...)
- *Heteroskedasticity*: changing variance

Introduction

Plotting a time series is an important early step in its analysis
In general, a plot can reveal:

- *Trend*: upward or downward pattern that might be extrapolated into the future
- *Periodicity*: Repetition of behavior in a regular pattern
- *Seasonality*: Periodic behavior with a known period (hourly, monthly, every 2 months...)
- *Heteroskedasticity*: changing variance
- *Dependence*: positive (successive observations are similar) or negative (successive observations are dissimilar)

Introduction

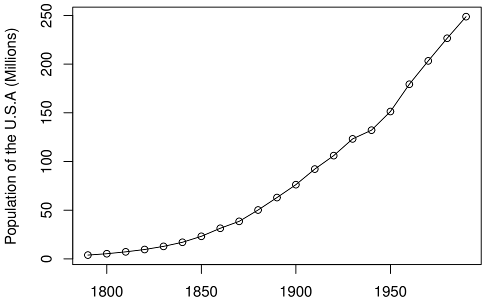
Plotting a time series is an important early step in its analysis
In general, a plot can reveal:

- *Trend*: upward or downward pattern that might be extrapolated into the future
- *Periodicity*: Repetition of behavior in a regular pattern
- *Seasonality*: Periodic behavior with a known period (hourly, monthly, every 2 months...)
- *Heteroskedasticity*: changing variance
- *Dependence*: positive (successive observations are similar) or negative (successive observations are dissimilar)
- Missing data, outliers, breaks...

Introduction

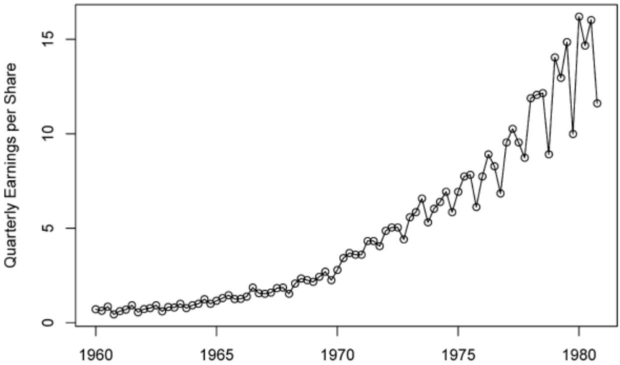
Example 1: U.S.A. population at ten year intervals from 1790-1990

- There is a upward trend
- There is a slight change in shape/structure
- Nonlinear behavior



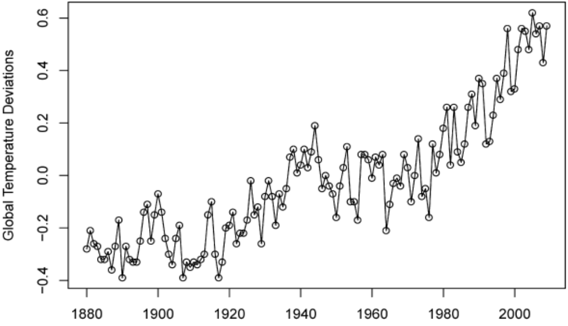
Example 2: Johnson & Johnson Quarterly Earnings

There are 84 quarters (21 years) measured from the 1st quarter of 1960 to the last quarter of 1980.
Note the gradually increasing underlying trend and the rather regular variation superimposed on the trend that seems to repeat over quarters.



Example 3: Global Warming

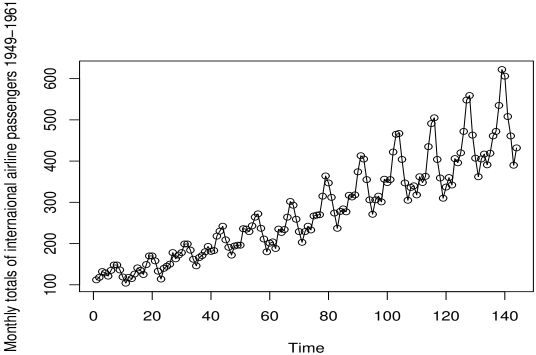
The data are the global mean land-ocean temperature index from 1880 to 2009. We note an *apparent upward trend* in the series during the latter part of the 20th century that has been used as an argument for the global warming hypothesis (whether the overall trend is natural or whether it is caused by some human-induced interface)



Example 4: Airline passengers from 1949-1961

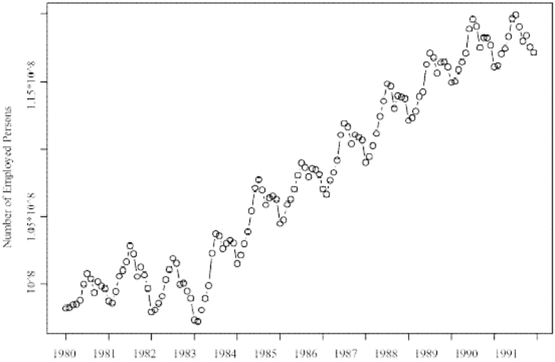
Trend? Seasonality? Heteroskedasticity? ...

Upward trend, seasonality on a 12 month interval, increasing variability



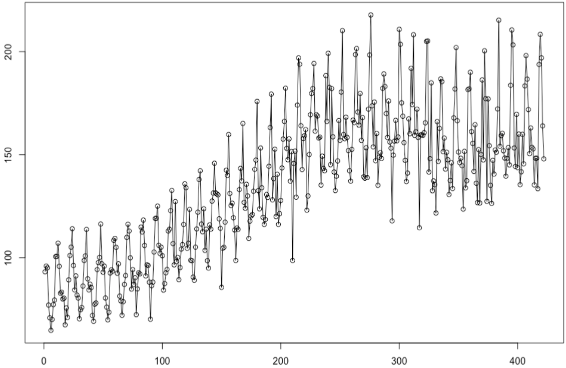
Example 5: Monthly Employed persons from 1980-1991

Trend? Seasonality? Heteroskedasticity? ...
Upward trend, seasonality with a structural break



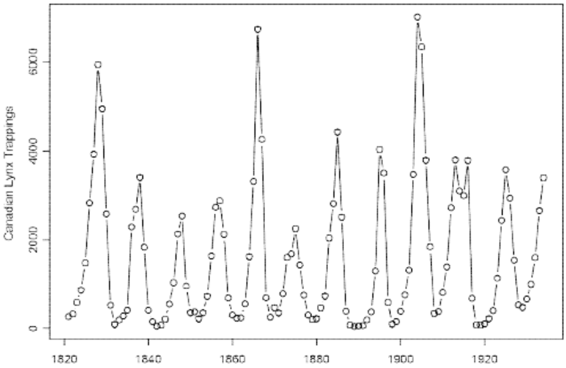
Example 6: Monthly Beer Production in Australia

Trend? Seasonality? Heteroskedasticity? breaks?... **no trend in last 100 Months, no clear seasonality**



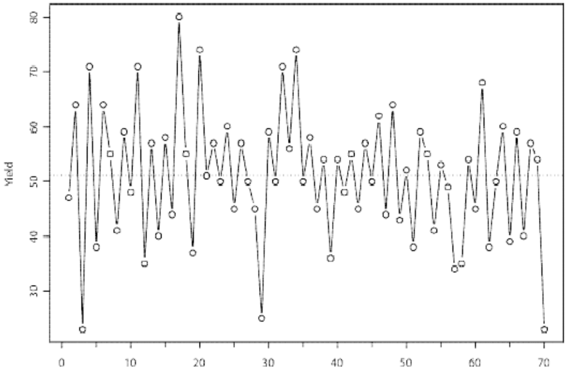
Example 7: Annual number of Candadian Lynx trapped near McKenzie River

Trend? Seasonality? Heteroskedasticity? breaks?... **no trend, no clear seasonality as it does correspond to a known period, periodicity**



Example 8: Yield from a controlled chemical batch process

Trend? Seasonality? Heteroskedasticity? breaks?... **Negative dependence: successive observations tend to lie on opposite sides of the mean.**



Example 9: Monthly real exchange rates between U.S and Canada

Trend? Seasonality? Heteroskedasticity? breaks?...

Example 9: Monthly real exchange rates between U.S and Canada

Trend? Seasonality? Heteroskedasticity? breaks?...

No obvious seasonality or trend. Hard to make a long range prediction. Positive Dependence: successive observations tend to lie on the same side of the mean.

Remarks

The issue of distinguishing between dependence and trend is difficult: There is no unique decomposition of a series into trend and dependence behaviors.
The issue that tampers this question: we have only one realization. If we had many realizations, we might be able to average to determine trend.

Introduction

Objectives

What do we hope to achieve with time series analysis?

- Provide a model of the data (testing of scientific hypothesis, etc.)

Objectives

What do we hope to achieve with time series analysis?

- Provide a model of the data (testing of scientific hypothesis, etc.)
- Predict future values (very common goal of analysis)

Introduction

Objectives

What do we hope to achieve with time series analysis?

- Provide a model of the data (testing of scientific hypothesis, etc.)
- Predict future values (very common goal of analysis)
- Produce a compact description of the data (a good model can be used for "data compression")

Introduction

Modeling

We take the approach that the data is a realization of random variable. However, many statistical tools are based on assuming any R.V. are IID.

In Times Series:

- R.V. are usually not independent (affected by trend and seasonality)

Introduction

Modeling

We take the approach that the data is a realization of random variable. However, many statistical tools are based on assuming any R.V. are IID.

In Times Series:

- R.V. are usually not independent (affected by trend and seasonality)
- Variance may change significantly

Introduction

Modeling

We take the approach that the data is a realization of random variable. However, many statistical tools are based on assuming any R.V. are IID.

In Times Series:

- R.V. are usually not independent (affected by trend and seasonality)
- Variance may change significantly
- R.V. are usually not identically distributed

The first goal in time series modeling **is to reduce the analysis needed to a simpler case**: Eliminate Trend, Seasonality, and heteroskedasticity then we model the remainder as dependent but Identically distributed

Introduction

The probabilistic model

A complete probabilistic model/description of a time series X_t observed as a collection of n random variables at times t_1,t_2,. . . , t_n for any positive integer n is provided by the joint probability distribution,

$$F(C_1, C_2, ..., C_n) = P(X_1 \leq C_1, ..., X_n \leq C_n)$$

- This is generally difficult to write, unless the case the variables are jointly normal.

Introduction

The probabilistic model

A complete probabilistic model/description of a time series X_t observed as a collection of n random variables at times $t_1, t_2, \ldots, t_n$ for any positive integer n is provided by the joint probability distribution,

$$F(C_1, C_2, ..., C_n) = P(X_1 \leq C_1, ..., X_n \leq C_n)$$

- This is generally difficult to write, unless the case the variables are jointly normal.
- Thus, we look for other statistical tools => quantifying dependencies

Introduction

Recall the basic concepts

X and Y are r.v.'s with finite variance

$$Cov(X,Y) = E((X - E(X))(X - E(X)))$$

and correlation

$$corr(X,Y) = \frac{Cov(X,Y)}{S_X S_Y}$$

- r.v.'s with zero correlation are uncorrelated
- Uncorrelated does not imply Independence
- Linear combination?

Some properties of Expectation and Variances/Covariances

X, Y, W, and Z are r.v.'s

- Cov(Y,Z) = E(YZ) - E(Y)E(Z)
- Var(X) = Cov(X,X) = E(X^2) - $(E(X))^2$
- Var(a+bX)= b^2 Var(X)
- Cov(aX+bY,cZ+dW)= ac Cov(X,Z) + ad Cov(X,W) + bc Cov(Y,Z) +bd Cov(Y,W)
- $E(\sum X_i) = \sum E(X_i)$

We will make use of these rules a lot! Remember them :)

Introduction

A time series model for the observed data x_t

- The mean function $\mu_X = E(X_t)$
- The Covariance function $\gamma_X(r, s) = E((X_r - \mu_X(r))(X_s - \mu_X(s)))$ for all integers r and s

The focus will be to determine the mean function and the Covariance function to define the time series model.

Some zero-Mean Models

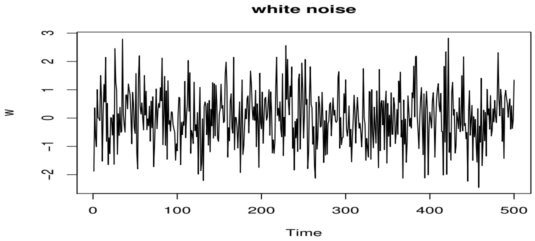
iid Noise

The simplest model for a times series: no trend or seasonal component and in which the observations are IID with zero mean.

- We can write, for any integer n and real numbers x_1, x_2,...,x_n,

$$P(X_1 \leq x_1, ..., X_n \leq x_n) = P(X_1 \leq x_1)...P(X_n \leq x_n)$$

It plays an important role as a building block for more complicated time series models



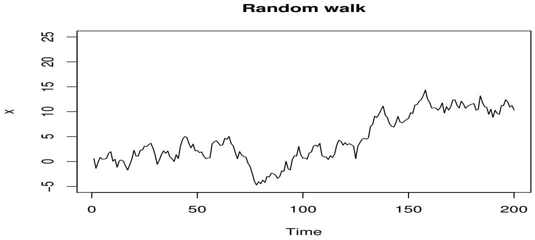
Some zero-Mean Models

Random Walk

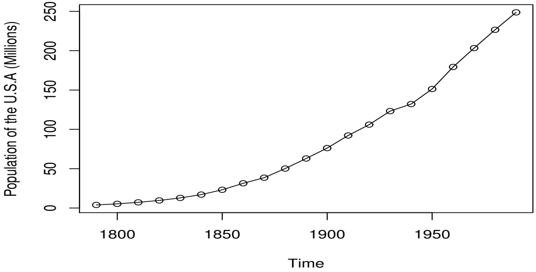
The random walk $\{S_t\}, t = 0, 1, 2, ...$ is obtained by cumulatively summing iid random variables, $S_0 = 0$

$$S_t = X_1 + X_2 + \cdots + X_t, \qquad t = 1, 2, ...$$

where X_t is iid noise. It plays an important role as a building block for more complicated time series models



Models with Trend



In this case a zero-mean model for the data is clearly inappropriate. The graph suggests trying a model of the form:

$$X_t = m_t + Y_t$$

where m_t is a function known as the trend component and Y_t has a zero mean. Estimating m_t?

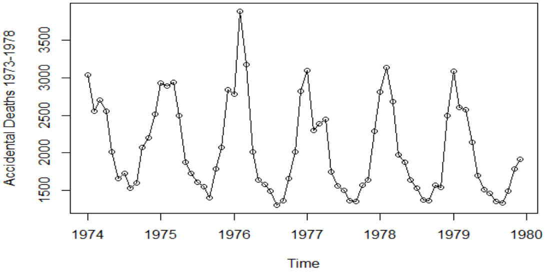
Models with Trend

m_t can be estimated using a least squares regression procedure (quadratic regression)
The estimated trend component $\hat{m}_t$ provides a natural predictor of future values of X_t

$$\hat{X}_t = \hat{m}_t + Y_t$$

Example with R

Models with Seasonality



In this case a zero-mean model for the data is clearly inappropriate. The graph suggests trying a model of the form:

$$X_t = S_t + Y_t$$

where S_t is a function known as the season component and Y_t has a zero mean. Estimating S_t?

Time series Modeling

- Plot the series => examine the main characteristics (trend, seasonality, ...)
- Remove the trend and seasonal components to get stationary residuals/models
- Choose a model to fit the residuals using sample statistics (sample autocorrelation function)
- Forecasting will be given by forecasting the residuals to arrive at forecasts of the original series X_t

Stationary and Autocorrelation function

Let X_t ba a time series:

- The mean function

$$\mu_X(t) = E(X_t)$$

- The Covariance function

$$\gamma_X(r,s) = Cov(X_r, X_s) = E((X_r - \mu_X(r))(X_s - \mu_X(s)))$$

for all integers r and s

Stationary and Autocorrelation function

Definitions

1. X_t is **strictly** stationary if $\{X_1, \dots X_n\}$ and $\{X_{1+h}, \dots X_{n+h}\}$ have the same joint distributions for all integers h and $n > 0$.

Stationary and Autocorrelation function

Definitions

1. X_t is **strictly** stationary if $\{X_1, \ldots X_n\}$ and $\{X_{1+h}, \ldots X_{n+h}\}$ have the same joint distributions for all integers h and $n > 0$.
2. X_t is **weakly** stationary if
   - $\mu_X(t)$ is independent of t.
   - $\gamma_X(t+h, t)$ is independent of t for each h.

Stationary and Autocorrelation function

Definitions

1. X_t is **strictly** stationary if $\{X_1, \dots X_n\}$ and $\{X_{1+h}, \dots X_{n+h}\}$ have the same joint distributions for all integers h and $n > 0$.
2. X_t is **weakly** stationary if
   - $\mu_X(t)$ is independent of t.
   - $\gamma_X(t+h, t)$ is independent of t for each h.
3. Let X_t be a stationary time series. **The autocovariance function** (ACVF) of X_t at lag h is
$$\gamma_X(h) = Cov(X_{t+h}, X_t)$$
   The **autocorrelation function** (ACF) of X_t at lag h is
$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Cor(X_{t+h}, X_t)$$

Stationary and Autocorrelation function

iid Noise

If X_t is iid noise and $E(X_t^2) = \sigma^2 < \infty$, then The process X_t is strictly stationary since the joint distribution can be written, for any integer n and real numbers c_1, c_2,...,c_n, as follows:

$$\begin{aligned} P(X_1 \leq c_1, ..., X_n \leq c_n) &= P(X_1 \leq c_1)...P(X_n \leq c_n) \\ &= F(c_1) \dots F(c_n) \\ &= \prod_{i=1}^{n} F(c_i) \end{aligned}$$

This does not depend on t. $X_t \sim IID(0, \sigma^2)$

The autocovariance function

$$\gamma_X(t+h, h) = \begin{cases} \sigma^2, & \text{if } h = 0 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

Stationary and Autocorrelation function

White Noise

If X_t is a sequence of uncorrelated random variables, each with zero mean and variance σ^2, then clearly X_t is stationary with the same autocovariance function as the iid noise. We can write

$$X_t \sim WN(0, \sigma^2)$$

Clearly, every $IID(0, \sigma^2)$ sequence is $WN(0, \sigma^2)$ but not conversely

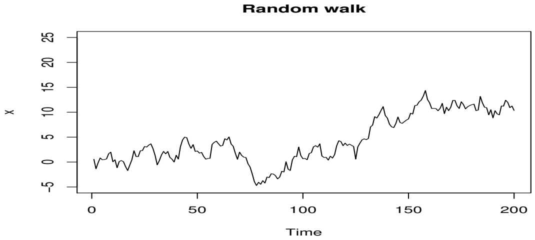
Stationary and Autocorrelation function

Random Walk

If $\{S_t\}$ is the random walk with X_t is a $IID(0, \sigma^2)$ sequence, then The random walk $\{S_t\}, t = 0, 1, 2, ...$ is obtained by cumulatively summing iid random variables, $S_0 = 0$

$$S_t = X_1 + X_2 + \cdots + X_t, \qquad t = 1, 2, ...$$

where X_t is iid noise. It plays an important role as a building block for more complicated time series models



Stationary and Autocorrelation function

First-Order Moving Average or MA(1) Process

Consider the series defined by the equation

$$X_t = Z_t + \theta Z_{t-1}, \qquad t = 0, \pm 1, \ldots.$$

where Z_t is $WN(0, \sigma^2)$ noise and θ is a real-valued constant.

- $E(X_t)$?
- $\gamma_X(t+h, h)$?

Stationary and Autocorrelation function

First-Order AutoRegressive AR(1) Process

Let us assume now that X_t is a stationary series satisfying the equation

$$X_t = \Phi X_{t-1} + Z_t, \qquad t = 0, \pm 1, \ldots.$$

where Z_t is $WN(0, \sigma^2)$ noise, $|\Phi| < 1$, and Z_t is uncorrelated with X_s for each $s < t$.

- $E(X_t)$?
- $\gamma_X(t+h, h)$?

The Sample Autocorrelation function

In practical problems, we do not start with a model, but with observed data $(x_1, x_2, \dots, x_n)$. **To assess the degree of dependence** in the data and to **select a model** for the data, one of the important tools we use is the sample autocorrelation function (Sample ACF).

Definition

Let $x_1, x_2, \dots, x_n$ be observations of a time series. The sample mean of $x_1, x_2, \dots, x_n$ is

$$\overline{x} = 1/n \sum_{t=1}^{n} x_i$$

The sample autocovariance function is

$$\hat{\gamma}(h) := 1/n \sum_{t=1}^{n-|h|} (x_{t+|h|} - \overline{x})(x_t - \overline{x}), \quad -n < h < n$$

The sample autocorrelation function is

$$\hat{\rho} = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

The Sample Autocorrelation function

Remarks

1. The sample autocorrelation function (ACF) can be computed for any data set and is not restricted to observations from a stationary time series.
2. For data containing a Trend, $|\hat{\rho}(h)|$ will display slow decay as h increases.
3. For data containing a substantial deterministic periodic component, $|\hat{\rho}(h)|$ will exhibit similar behavior with the same periodicity.

The Sample Autocorrelation function

We may recognize the sample autocorrelation function of many time series:

Remarks

1. White Noise => Zero
2. Trend => Slow decay
3. Periodic => Periodic
4. Moving Average (q) => Zero for $|h| > q$
5. AutoRegression (p) => Decay to zero exponentially

The Sample Autocorrelation function

Examples with R

Estimation and Elimination of Trend and Seasonal Components

The first step in the analysis of any time series is to plot the data. Inspection of the graph may suggest the possibility of representing the data as follows (the classical decomposition):

$$X_t = m_t + s_t + Y_t$$

where

- m_t is the trend component
- s_t is the seasonal component
- Y_t random noise component / Residuals

if seasonal and noise fluctuations appear to increase with the level of the process => Eliminate by using a preliminary transformation of the data (natural log,...).

Estimation and Elimination of Trend and Seasonal Components

Approaches

1. Estimate and eliminate the trend and the seasonal components in the hope that the residual Y_t will turn out to be a stationary time series => Find a Model using stationary process theory.
2. Box and Jenkins (1976) proposed to apply differencing operators to the series until the differenced observations resemble a realization of some stationary time series.

Estimation and Elimination of Trend and Seasonal Components

Trend Estimation

Moving average and spectral smoothing are an essentially nonparametric methods for trend (or signal) estimation

$$X_t = m_t + Y_t, \quad E(Y_t) = 0$$

1. Smoothing with a finite moving average filter
2. Exponential smoothing
3. Smoothing by eliminating the high-frequency components (Fourier series)
4. Polynomial fitting (Regression)

Estimation and Elimination of Trend and Seasonal Components

Trend Estimation: Smoothing Moving Average

Let q be a nonnegative integer and consider the two-sided moving average

$$W_t = \frac{\sum_{j=-q}^{q} X_{t-j}}{2q+1}$$

It is useful to think about $\hat{m}_t$ as a process obtained from X_t by application of a linear operator or linear filter $\hat{m}_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j}$ with weights $a_j = \frac{1}{2q+1}$. This particular filter is a low-pass filter in the sense that it takes the data and removes from it the rapidly fluctuating (high frequency) components $\hat{Y}_t$ to leave slowly varying estimated trend term $\hat{m}_t$.

With R (smooth.ma{itsmr})

Estimation and Elimination of Trend and Seasonal Components

Trend Estimation: Exponential Smoothing

For any fixed $\alpha \in [0, 1]$, the one-sided moving averages $\hat{m}_t$ defined by:

$$\hat{m}_t = \alpha X_t + (1 - \alpha)\hat{m}_{t-1}; t = 2, ...n$$

and $\hat{m}_1 = X_1$

With R (smooth.exp{itsmr})

Estimation and Elimination of Trend and Seasonal Components

Trend Estimation: Smoothing by eliminating of high frequency

Using Fourier Transform we can delete some high frequency.

With R (smooth.fft{itsmr})

Trend Estimation: Polynomial fitting

Using regression procedures

R function (we can use lm{stats} or trend{itsmr})

Nonseasonal Model With Trend: Estimation

Remark: Smoothing Moving Average

- There are many filters that could be used for smoothing!
- Large q will allow linear trend function $m_t = c_0 + c_1 t$ to pass
- We must be aware of choosing q to be too large if m_t is not linear (Example with R)
- Clever choice of the weights a_j can design a filter that will not only be effective in attenuating noise in the data, but that will also allow a larger class of trend functions (for example all polynomials of degree ≤ 3) to pass.
- Spencer 15-point moving average is a filter that passes polynomials of degree ≤ 3 without distortion. Its weights are: $1/320[-3, -6, -5, 3, 21, 46, 67, 74, 67, 46, 21, 3, -5, -6, -3]$

Estimation and Elimination of Trend and Seasonal Components

Trend Elimination: Differencing

Instead of attempting to remove the noise by smoothing as in Method 1, we now attempt to eliminate the trend term by differencing. We define the lag-1 difference operator ∇ by:

$$\nabla X_t = X_t - X_{t-1} = (1 - B)X_t$$

where B is the backward shift operator : $BX_t = X_{t-1}$

With R (diff{base})

Estimation and Elimination of Trend and Seasonal Components

Trend Elimination: Differencing

Powers of the operators B and ∇ are defined as follows:

$$\nabla^j(X_t) = \nabla(\nabla^{j-1}(X_t))$$

where $j >= 1, \nabla^0(X_t) = X_t$ and

$$B^j(X_t) = X_{t-j}$$

Example: $\nabla^2 X_t$?

If the operator ∇ is applied to a linear trend function $m_t = c_0 + c_1 t$, then we obtain the constant function $\nabla m_t = m_t - m_{t-1} = c_1$. In the same way we can show any polynomial trend of degree k can be reduced to a constant by application of the operator ∇^k.

It is found in practice that the order k of differencing required is quite small; frequently 1 or 2.

Estimation and Elimination of Trend and Seasonal Components

Both Trend and Seasonal

The classical Decomposition model

$$X_t = m_t + s_t + Y_t$$

where $E(Y_t) = 0$, $s_{t+d} = s_t$, and $\sum_{j=1}^{d} s_j = 0$

d is the period of the seasonal component.
The methods described previously (for the trend) can be adapted in a natural way to eliminate both the trend and seasonality components.

Estimation: Both Trend and Seasonal [Method 1]

Suppose we have observations $x_1, x_2, \ldots x_n$

1. The trend is first estimated by applying a moving average filter specially to eliminate the seasonal component of period d
   - if the period is even, say d=2q then $\hat{m}_t = (0.5x_{t-q} + x_{t-q+1}, \ldots, 0.5x_{t+q})/d \;\; q < t \leq n-q$
   - if the period is odd, say d=2q+1 then we use the simple moving average
2. Estimate the seasonal component; for each k=1,..., d, we compute the average w_k of the deviation $x_{k+jd} - \hat{m}_{k+jd}$, $q < k + jd \leq n - q$; and we estimate the seasonal component as follows:

$$\hat{s}_k = w_k - \frac{\sum_{i=1}^{d} w_i}{d}; \quad k = 1, \ldots, d$$

   and $\hat{s}_k = \hat{s}_{k-d}; \quad k > d$
3. The *deseasonalized* data is then $d_t = x_t - \hat{s}_t$ t=1,...n
4. Reestimate the trend from the *deseasonalized* data using one of the methods already described.

Eliminating: Both Trend and Seasonal [Method 2]

We can use the Differencing operator to eliminate the trend and the seasonal component

1. Eliminate the seasonality of period d using $\nabla_d X_t = X_t - X_{t-d}$
   $X_t = m_t + s_t + Y_t$ Applying $\nabla_d X_t = m_t - m_{t-d} + Y_t - Y_{t-d}$
2. Eliminate the trend $m_t - m_{t-d}$ using $\nabla^k X_t$

Testing Noise squence

IID Null Hypothesis

1. Q_{LB} Ljung-Box Test (H_0 : data are Independent vs. H_1 : data are not independent)
2. Q_{ML} McLeod- Li Test (autocorrelations of squared data)
3. Turning point Test (IID vs. not IID)
4. The difference-sign Test (randomness)
5. The rank Test (detecting a linear trend)

Remember that as you increase the number of tests, the probability of at least one rejects the null hypothesis when it is true increases.

Summary

Trend and Seasonality

- Estimation and Elimination of Trend and Seasonal Components
  1. Smoothing methods
  2. Differencing Operator ∇
  3. Properties of the operator ∇
  4. Procedure to estimate Both Trend and Seasonality (with R)
- Testing Noise sequence (test(residuals) with R)

Stationary Processes

Remarks

1. A key role in time series analysis is given by processes whose properties **do not vary with time**.
2. If we wish to make predictions then clearly we must assume that *somehting* does not vary with time
3. In time series analysis, our goal is to predict a series that contains a random component, if this random component is stationary (*weakly*) then we can develop powerful techniques to forecast its future values.

Stationary Processes

Basic Properties

1. The autocovariance function (ACVF)
   $\gamma(h) = Cov(X_{t+h}, X_t), \quad h = 0, \pm 1, \pm 2, \ldots$
2. The autocorrelation function (ACF) $\rho(h) = \frac{\gamma(h)}{\gamma(0)}$
3. $\gamma(0) \geq 0$
4. $|\gamma(h)| \leq \gamma(0)$, for all h ($\rho(h) \leq 1$) or (Cauchy-Schwarz inequality $E(XY)^2 \leq E(X^2)E(Y^2)$)
5. $\gamma(h) = \gamma(-h)$

Stationary Processes

Prediction

The ACF and ACVF provide a useful measure of **the degree of dependence** among the values of a time series at different times => Very important if we consider the prediction of future values of the series in terms of the past and present values.

Stationary Processes

Prediction

The ACF and ACVF provide a useful measure of **the degree of dependence** among the values of a time series at different times => Very important if we consider the prediction of future values of the series in terms of the past and present values.

Question

What is the role of the autocorrelation function in prediction?

Stationary Processes

The role of autocorrelation in prediction

- Consider X_t a stationary Gaussian time series (all of its joint distributions are Multivariate Normal)
- We observed X_n and we would like to find the function of X_n that gives us *the best predictor* of X_{n+h} (the value of the series after h time units).
- **The best predictor** will be given by the function of X_n that minimizes the mean squared error (MSE).

Question

What is function of X_n that gives us the best predictor of X_{n+h}?

Answer

The best predictor of X_{n+h} in terms of MSE is given by
$E(X_{n+h}|X_n) = \mu + \rho(h)(X_n - \mu)$

The role of autocorrelation in prediction

For time series with non-normal joint distributions the calculation are in general more complicated => we look at **the best linear predictor** ($l(X_n) = aX_n + b$), then our problem becomes finding a and b that minimize $E((X_{n+h} - aX_n - b)^2)$.

Answer

The best linear predictor of $l(X_n)$ in terms of MSE is given by

$$l(X_n) = \mu + \rho(h)(X_n - \mu)$$

The fact that the best linear predictor depends only on the mean and the ACF of the series X_t means that it can be calculated without more detailed knowledge of the series X_t

Stationary Processes: Examples

The MA(q) process

X_t is a moving-average process of order q if

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

where $Z_t \sim WN(0, \sigma^2)$ and $\theta_1, \ldots, \theta_q$ are constants.

Remarks

- If X_t is a stationary q-correlated time series with mean 0, then it can be represented as the MA(q) process

Stationary Processes: Examples

The ARMA(1,1) process

The time series X_t is an ARMA(1,1) process if it is stationary and satisfies (for every t)

$$X_t - \Phi X_{t-1} = Z_t + \theta Z_{t-1}$$

where $Z_t \sim WN(0, \sigma^2)$

Stationary Processes: Properties

The MA(q) process

The simplest ways to construct a time series that is strictly stationary is to "filter" an iid sequence of random variables. Consider $Z_t \sim IID$, we define:

$$X_t = g(Z_t, Z_{t-1}, \ldots, Z_{t-q})$$

for some real-valued function $g(., \ldots, .)$. We can say that X_t is **q-dependent**.

Remarks

- IID is 0-dependent
- WN is 0-correlated
- A stationary time series is q-correlated if $\gamma(h) = 0$ whenever $lhl > q$
- MA(1) is 1-correlated
- MA(q) is q-correlated

Stationary Processes: Properties

The MA(q) process

X_t is a **moving-average process** of order q if

$$X_t = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

where $Z_t \sim WN(0, \sigma^2)$ and $\theta_1, \ldots, \theta_q$ are constants.

If X_t is a stationary q-correlated time series with mean 0, then it can be represented as the MA(q) process given by the equation above.

Stationary Processes: Properties

The class of Linear time series models, which includes the class of Autoregressive Moving-Average (ARMA) models, provides a general framework for studying stationary processes.

Linear processes

The time series X_t is a **Linear process** if it has the representation:

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j} \tag{2}$$

for all t, where $Z_t \sim WN(0, \sigma^2)$ and ψ_j is a sequence of constants with $\sum_{j=-\infty}^{\infty} \mid \psi_j \mid < \infty$

1. The condition $\sum_{j=-\infty}^{\infty} \mid \psi_j \mid < \infty$ ensures the infinite sum in (2) converges
2. If $\psi_j = 0$ for all $j < 0$ then a linear process is called a moving average or MA(∞)

Stationary Processes: Linear Process

Proposition

Let Y_t be a stationary series with mean 0 and autocovariance function γ_y. If $\sum_{j=-\infty}^{\infty} \mid \psi_j \mid < \infty$, then the time series

$$X_t = \sum_{j=-\infty}^{\infty} \psi_j Y_{t-j} = \sum_{j=-\infty}^{\infty} \psi_j B^j Y_t = \psi(B) Y_t \tag{3}$$

is stationary with mean zero and autocorrelation function $\gamma_x(h)$

1. $\gamma_x(h) = \sum \sum \psi_j \psi_k \gamma x(h - k + j)$

Stationary Processes: Properties

The AR(1) process

The time series X_t is an AR(1) process was found to be stationary:

$$X_t - \Phi X_{t-1} = Z_t + \theta Z_{t-1} \tag{4}$$

if $Z_t \sim WN(0, \sigma^2)$ and $\mid \Phi \mid < 1$, and Z_t is uncorrelated with X_s for each $s \leq t$

1. X_t is called a causal or future-independent function of Z_t.
2. If $\mid \Phi \mid > 1$ (the series does not converge, we can rewrite then AR(1))
3. If $\Phi = \pm 1$, there is no stationary solution of (4)

Stationary Processes: Properties

The ARMA(1,1) process

The time series X_t is an ARMA(1,1) process if it is stationary and satisfies (for every t)

$$X_t - \Phi X_{t-1} = Z_t + \theta Z_{t-1}$$

where $Z_t \sim WN(0, \sigma^2)$ and $\theta + \Phi \neq 0$

1. A stationary the ARMA(1,1) exists if and only if $\phi \neq \pm 1$
2. If $| \phi | < 1$, then the unique stationary solution is $X_t = Z_t + (\theta + \phi) \sum \phi^{j-1} Z_{t-j}$. This X_t is causal, since X_t can be expressed in terms of the current and past values of Z_s, $s \leq t$
3. If $| \phi | > 1$, then the unique stationary solution is $X_t = -\theta\psi^{-1} Z_t + (\theta + \phi) \sum \phi^{-j-1} Z_{t+j}$. This X_t is noncausal, since X_t can be expressed in terms of the current and future values of Z_s, $s \geq t$. (unatural solution)

Stationary Processes: Properties

The ARMA(1,1) process

The time series X_t is an ARMA(1,1) process if it is stationary and satisfies (for every t)

$$X_t - \Phi X_{t-1} = Z_t + \theta Z_{t-1}$$

where $Z_t \sim WN(0, \sigma^2)$ and $\theta + \Phi \neq 0$

1. if $\theta = \pm 1$, then ARMA(1,1) process is invertible in the more general sense that Z_t is a mean square limit of finite linear combinations of X_s, $s \leq t$.
2. If $| \theta | < 1$, then the ARMA(1,1) is invertible, since Z_t can be expressed in terms of the current and past values of X_s, $s \leq t$
3. If $| \theta | > 1$, then the ARMA(1,1) is noninvertible, since Z_t can be expressed in terms of the current and future values of X_s, $s \geq t$.

Stationary Processes: Properties

The Sample Mean and Autocorrelation Function

A weakly stationary time series X_t is charachterized by its mean μ, its autocovariance $\gamma(.)$, and its autocorrelation $\rho(.)$. The estimation of these statistics play a crucial role in problems of inference => constructing an appropriate model for the data. We examine here some properties of the sample estimates $\overline{X}$ and $\hat{\rho}(.)$.

1. $E(\overline{X}_n) = \mu$
2. $Var(\overline{X}_n) = n^{-1} \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right) \gamma(h)$
3. If the time series is Gaussian, then $(\overline{X}_n - \mu)/\sqrt{n} \sim N(0, (1 - \frac{|h|}{n})\gamma(h))$
4. The confidence Intervals for μ are given $\overline{X}_n \pm Z_{1-\alpha/2} \frac{\sqrt{\hat{v}}}{\sqrt{n}}$, where $\hat{v} = \sum_{|h|<\sqrt{n}} (1 - \frac{|h|}{\sqrt{n}}) \hat{\gamma}(h)$. For ARMA processes, this is a good approximation of v for large n.

Stationary Processes: Properties

The Sample Mean: Asymptotic distribution

If we know the asymptotic distribution of $\overline{X}_n$, we can use it to infer about μ (e.g. is μ=0?). Similarly for $\hat{\rho}(h)$

1. $$(\overline{X}_n - \mu)\sqrt{n} \sim AN(0, \sum_{|h|<n} (1 - \frac{|\ h\ |}{n})\gamma(h))$$

2. In this case, the confidence Intervals for μ are given $\overline{X}_n \pm Z_{1-\alpha/2}\frac{\sqrt{\hat{v}}}{\sqrt{n}}$, where $\hat{v} = \sum_{|h|<\sqrt{n}} (1 - \frac{|h|}{\sqrt{n}})\hat{\gamma}(h)$. For ARMA processes, this is a good approximation of v for large n.

Example

What are the approximate 95% confidence intervals for the mean of AR(1)?
"AN" means Asymptotically Normal

Stationary Processes: Properties

The Estimation of $\gamma(.)$ and $\rho(.)$

The sample autocovariance and autocorrelation functions are defined by:

$$\hat{\gamma}(h) = \frac{\sum_{t=1}^{n-|h|}(X_{t+|h|} - \overline{X}_n)(X_t - \overline{X}_n)}{n}$$

and

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$$

1. For h slightly smaller than n, the estimate of $\gamma(h)$ and $\rho(h)$ are unreliable. Since there few pairs (X_{t+h}, X_t) available (only one if h=n-1). **A practical guide is to have at least n=50 and h $\leq$ n/4**

Stationary Processes: Properties

The Estimation of $\gamma(.)$ and $\rho(.)$: Asymptotic distribution

The sampling distribution of $\rho(.)$ can usually be approximated by a normal distribution for large sample sizes. For Linear Models (ARMA):

$$\hat{\boldsymbol{\rho}} = (\hat{\rho}(1), \ldots, \hat{\rho}(k))' \sim AN(\boldsymbol{\rho}, \frac{W}{n})$$

where $\boldsymbol{\rho} = (\rho(1), \ldots, \rho(k))$, and W is the covariance matrix whose (i, j) element is given by **Bartlett's formula**:

$$w_{ij} = \sum_{k=1}^{\infty} \{\rho(k+i) + \rho(k-i) - 2\rho(i)\rho(k)\} \times \{\rho(k+j) + \rho(k-j) - 2\rho(j)\rho(k)\}$$

1. IID Noise?

Stationary Processes: Properties ACF

The sample ACF Examination - Results

- If $\mid \hat{\rho}(h) \mid < \frac{1.96}{\sqrt{n}}$, for all $h \geq 1$, then assume $MA(0)$, WN sequence.
- If $\mid \hat{\rho}(1) \mid > \frac{1.96}{\sqrt{n}}$, then we should look at the rest $\hat{\rho}(h)$ with $\pm\frac{1.96\sqrt{1+2\rho^2(1)}}{\sqrt{n}}$; we can replace $\rho(1)$ by its estimate (you can also remark that $2\hat{\rho}^2(1)/n \sim 0$, for large n).
- In general, if $\mid \hat{\rho}(h_0) \mid > \frac{1.96}{\sqrt{n}}$ and $\mid \hat{\rho}(h) \mid < \frac{1.96}{\sqrt{n}}$, for $h \geq h_0$, then assume MA(q) model with $q = h_0$.

Forecasting P_nX_{n+h}

Now we consider the problem of predicting the values X_{n+h}; $h > 0$. Let's assume X_t is a stationary time series with μ and γ.
The goal is to find the linear combination of $1, X_n, \ldots, X_1$ that minimizes the mean squared error. We will denote

$$P_nX_{n+h} = a_0 + a_1X_n + \cdots + a_nX_1$$

It remains to find the coefficients a_i that minimizes:

$$E((X_{n+h} - a_0 - a_1X_n - \cdots - a_nX_1)^2)$$

Forecasting $P_n X_{n+h}$

We can show that $P_n X_{n+h}$ is given by:

$$P_n X_{n+h} = \mu + \sum_{i=1}^{n} a_i (X_{n+1-i} - \mu)$$

and

$$E((X_{n+h} - P_n X_{n+h})^2) = \gamma(0) - \boldsymbol{a}_n' \gamma_n(h)$$

where $\boldsymbol{a}_n$ satisfies

$$\Gamma_n \boldsymbol{a}_n = \gamma_n(h)$$

where $\boldsymbol{a}_n = (a_1, \ldots, a_n)'$, $\Gamma_n = \begin{bmatrix} \gamma(0) & \gamma(1) & \ldots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \ldots & \gamma(n-2) \\ \vdots & \vdots & \ddots & \vdots \\ \gamma(n-1) & \gamma(n-2) & \ldots & \gamma(0) \end{bmatrix}$;

$\gamma_n(h) = (\gamma(h), \gamma(h+1), \ldots, \gamma(h+n-1))'$; and $a_0 = \mu(1 - \sum_{i=1}^{n} a_i)$

Prediction Algorithms

The following prediction algorithms use the idea of one-step predictor $P_n X_{n+1}$ based on n previous observations would be used to calculate $P_{n+1} X_{n+2}$. This said to be recursive.

1. The Durbin-Levinson Algorithm (well suited to forecasting $AR(p)$)
2. The Innovations Algorithm (well suited to forecasting $MA(q)$)

ARMA(p,q) models

ARMA(p, q) process

X_t is an ARMA(p, q) process if X_t is stationary and if for every t,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q}$$

where $Z_t \sim WN(0, \sigma^2)$ and the polynomials $(1 - \phi_1 z - \cdots - \phi_p z^p)$ and $(1 + \theta_1 z + \cdots + \theta_q z^q)$ have no common factors.

The process X_t is said to be an ARMA(p,q) process with mean μ if $X_t - \mu$ is an ARMA(p,q) process.

ARMA Models

We can write:

$$\phi(B)X_t = \theta(B)Z_t$$

where $\phi(.)$ and $\theta(.)$ are the *pth* and *qth* degree polynomials:

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$$

and

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$$

and B is the backward shift operator $B^j X_t = X_{t-j}$, $j = 0, \pm 1, ...$

ARMA models

$ARMA(1,1)$ process

X_t is an ARMA$(1,1)$ process if X_t is stationary and if for every t,

$$X_t - \phi_1 X_{t-1} = Z_t + \theta_1 Z_{t-1}$$

where $Z_t \sim WN(0, \sigma^2)$

Lecture notes on ARMA(1,1).

ARMA models

$ARMA(p,q)$ process: Existence, Causality, and Invertibility

- A stationary solution (existence, uniqueness) X_t exists if and only if:

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \quad \text{for all } | z |= 1$$

- An ARMA(p,q) process X_t is **causal** if there exist constant ψ_j such that $\sum_{j=0}^{\infty} | \psi_j |< \infty$ and $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$, for all t. Causality is equivalent to the condition:

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \quad \text{for all } | z |\leq 1$$

- An ARMA(p,q) process X_t is **invertible** if there exist constant π_j such that $\sum_{j=0}^{\infty} | \pi_j |< \infty$ and $Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j}$, for all t. Invertibility is equivalent to the condition:

$$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \neq 0 \quad \text{for all } | z |\leq 1$$

More details can be found in Section 3.1, Chapter 3. Similar to ARMA(1,1).

ARMA models

$ARMA(p,q)$ process: Autocorrelation Function (ACF)

The autocorrelation function (ACF) of the causal ARMA(p,q) process X_t can be found using the fact that $X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}$ (causal MA(∞) process) and

$$\gamma_X(h) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}$$

Find the ACF of ARMA(1,1), using the fact that

$$\psi_0 = 1; \psi_j = (\phi + \theta)\phi^{j-1}, \quad j \geq 1$$

ARMA models

$ARMA(1,1)$ process: Autocorrelation Function (ACF)

$$\gamma_X(0) = \frac{(1+\phi)^2}{1-\phi^2}\sigma^2$$

$$\gamma_X(1) = \frac{(\theta+\phi)(1+\phi\theta)}{1-\phi^2}\sigma^2$$

$$\gamma_X(h) = \phi^{h-1}\gamma_x(1)$$

Section 3.2.1 has details about the calculation of the ACVF.

ARMA models

Definition: Partial Autocorrelation Function (PACF)

The partial autocorrelation function (PACF) of ARMA process X_t is the function $\alpha(.)$ defined by

$$\alpha(0) = 1$$

and

$$\alpha(h) = \phi_{hh}, \quad h \geq 1$$

where ϕ_{hh} is the last component of $\boldsymbol{\phi_h} = \Gamma_h^{-1}\boldsymbol{\gamma}_h$

Think about it as a conditional correlation $Cor(X_t, X_{t+h} \mid X_{t+1}, \ldots, X_{t+h-1})$

Modeling ARMA models

Practical facts about PACF and ACF

- The PACF is often best used to identify the AR(p) models
- For AR(p) models, the theoretical PACF are equal to zero after $h = p$
- The MA(q) models are better identified using ACF.

Examples with R

Modeling ARMA models (Chapter 5)

To determine an appropriate ARMA(p,q) model to represent an observed stationary process, we need to:

- Choose the orders p and q (order selection)
- Estimate the mean
- Estimate the coefficients $\{\phi_i, i = 1, \ldots, p\}$ and $\{\theta_i, i = 1, \ldots, q\}$
- Estimate the white noise variance σ^2
- Select a model

Modeling ARMA models (Chapter 5)

To determine an appropriate ARMA(p,q) model to represent an observed stationary process, we need to:

- Choose the orders p and q (order selection) Use ACF and PACF plots.
- Estimate the mean Use the mean-corrected process $X_t - \overline{X}_n$.
- Estimate the coefficients $\{\phi_i, i = 1, \ldots, p\}$ and $\{\theta_i, i = 1, \ldots, q\}$
- Estimate the white noise variance σ^2
- Select a model

Modeling ARMA models (Chapter 5)

When p and q are **known** and the time series is **mean-corrected**, good estimators of vectors $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ can be found imaging the data to a stationary Gaussian time series and **maximizing the likelihood** with respect to the $p+q+1$ parameters ($\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ and σ^2). We can estimate these parameters using:

- The Yule-Walker and Burg procedures for pure autoregressive models AR(p) ($yw\{itsmr\}, burg\{itsmr\}$)
- The Innovations and Hannan-Rissanen procedures for ARMA(p,q) ($ia\{itsmr\}, hannan\{itsmr\}$)

Modeling ARMA models (Chapter 5)

Properties

- The Burg's algorithm usually gives higher likelihoods than the Yule-Walker equations for AR(p)
- For pure MA processes, the Innovations algorithm usually gives higher likelihoods than the Hannan-Rissanen procedure
- For ARMA models, the Hannan-Rissanen is more successful in finding causal models

These preliminary estimations are required for initialization of the likelihood maximization.

Yule-Walker Estimation

The Sample Yule-Walker equations are:

$$\hat{\phi} = (\hat{\phi}_1, \ldots, \hat{\phi}_p)' = \hat{R}_p^{-1}\hat{\rho}_p$$

and

$$\hat{\sigma}^2 = \hat{\gamma}(0)\left[1 - \hat{\rho}_p'\hat{R}_p^{-1}\hat{\rho}_p\right]$$

where $\hat{\rho}_p = (\hat{\rho}(1), \ldots, \hat{\rho}(p))'$

- Large-Sample Distribution of $\hat{\phi} \sim N(\phi, \frac{\sigma^2\Gamma_p^{-1}}{n})$
- Confidence intervals of ϕ_{pj} are $\hat{\phi}_{pj} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{\sigma}^2\hat{\Gamma}_p^{-1}}{n}}$

We can use the Innovations algorithm in order to estimate parameters of MA(q). The Confidence regions of the coefficients:

MA(q) Estimation

- Confidence intervals of θ_j are $\hat{\theta}_{mj} \pm z_{1-\alpha/2}\sqrt{\frac{\sum_{i=0}^{j-1}\hat{\theta}_{mi}^2}{n}}$

Estimation ARMA Models (Chapter 5)

Suppose X_t is a Gaussian time series with mean zero

Maximum Likelihood Estimators (Section 5.2 from the Textbook)

$$\hat{\sigma}^2 = \frac{S(\hat{\phi}, \hat{\theta})}{n}$$

where $S(\hat{\phi}, \hat{\theta}) = \sum_{j=1}^{n} \frac{(X_j - \hat{X}_j)^2}{r_{j-1}}$,
and $\hat{\phi}, \hat{\theta}$ are the values of ϕ, θ that minimize:

$$l(\phi, \theta) = \ln n^{-1} S(\phi, \theta) + n^{-1} \sum_{j=1}^{n} \ln r_{j-1}$$

Minimization of $l(\phi, \theta)$ must be done numerically. Initial values for $\phi, \theta)$ can be obtained from preliminary estimation algorithms (Yule-walker, Burg, Innovations, and Hannan).

ARMA Models

Order selection

The Akaike Information criterion bias-corrected (AICC) is defined as follows:

$$AICC := -2\ln(L_X(\beta, S_X(\beta)/n) + \frac{2(p+q+1)n}{n-p-q-2}$$

It was designed to be an approximately unbiased estimate of the Kullback-Leibler index of the fitted model relative to the true model.

We select p and q values for our fitted model to be those that minimize $AICC(\hat{\beta})$.

Modeling ARMA models

Paramater Redundancy

Consider a white noise process $X_t = Z_t$. We can write this as

$$0.3X_{t-1} = 0.3Z_{t-1}$$

By subtracting the two representation we have:

$$X_t - 0.3X_{t-1} = Z_t - 0.3Z_{t-1}$$

which looks like an ARMA(1,1) model. Of course X_t is still white noise. We have this problem because of the parameter redundancy or over-parameterization.
We can solve this problem by looking at the common factors of the two polynomials $\phi(B)$ and $\theta(B)$.

Forecasting ARMA models

Given $X_1, X_2, \ldots, X_n$ observations, we want to predict X_{n+h}. We know that the best linear predictor is given by:

$$P_n X_{n+h} = \hat{X}_{n+h} = \mu + \sum_{i=1}^{n} a_i (X_{n+i-1} - \mu)$$

where the vector a_n satisfies $\Gamma_n a_n = \gamma_n(h)$. The mean squared errors:

$$E(X_{n+h} - P_n X_{n+h})^2 = \gamma(0) - a_n' \gamma(h) = \gamma(0)(1 - a_n' \rho(h))$$

Use Durbin-Levinson and Innovations algorithms to solve these equations.

Forecasting ARMA models

Examples

- For AR(1): $P_n X_{n+1} = \phi X_n$
- For AR(1) with nonzero-mean: $P_n X_{n+h} = \mu + \phi^h (X_n - \mu)$
- For AR(p): if n>p then $P_n X_{n+1} = \phi_1 X_n + \cdots + \phi_p X_{n+1-p}$
- For MA(q): $P_n X_{n+1} = \sum_{j=1}^{\min(n,q)} \theta_{nj} (X_{n+1-j} - \hat{X}_{n+1-j})$
- For ARMA(p,q): if $n > m = \max(p, q)$ then for all $h \geq 1$

$$P_n X_{n+h} = \sum_{i=1}^{p} \phi_i P_n X_{n+h-i} + \sum_{j=h}^{q} \theta_{n+h-1,j} (X_{n+h-j} - \hat{X}_{n+h-j})$$

The calculations are performed automatically using *forecast{itsmr}* function into R

ARIMA models

We have already seen the importance of the class of ARMA models for representing stationary time series. A generalization of this class, which includes a wide range of nonstationory series, is provided by the ARIMA (AutoRegressive Integrated Moving Average)models.

Definition

If d is a nonnegative integer, then X_t is an **ARIMA(p,d,q) process** if

$$Y_t := (1 - B)^d X_t$$

is a causal ARMA(p,q) process.
B is the backward shift operator.

- *ARIMA processes reduce to ARMA processes when differenced finitely many times*
- *X_t is stationary if and only if $d = 0$*

ARIMA models

The definition means that X_t satisfies a difference equation of the form:

$$\phi(B)(1-B)^d X_t = \theta(B)Z_t, \quad Z_t \sim WN(0, \sigma^2)$$

The polynomial $\phi^*(B) = \phi(B)(1-B)^d$ has now a zero of order d at $z = 1$.

Example

Consider X_t is an ARIMA(1,1,0) process of for $\phi \in (-1, 1)$,

$$(1-\phi B)(1-B)X_t = Z_t, \quad Z_t \sim WN(0, \sigma^2)$$

ARIMA models

ACF for ARIMA models

- A distinctive feature of the data that suggests the appropriateness of an ARIMA model is **the slowly decaying positive sample autocorrelation function**.
- In order to fit an ARIMA model to the data, we would apply the operator $\nabla = 1 - B$ repeatedly in the hope of some j, $\nabla^j X_t$ will have a **rapidly decaying sample autocorrelation function**, that is compatible with that of an ARMA process with no zeros of the autoregressive polynomial near the unit circle.

Example with R with $\phi = 0.8$, n=200, and $\sigma^2 = 1$

ARIMA models

Modeling of ARIMA models

- Deviations (e.g. trend, seasonality, heteroskedasticity) from stationary may be suggested by the graph of the the series itself or by the sample autocorrelation function or both.
- We have seen how to handle the Trend and Seasonality components
- Logarithmic Transformation is appropriate whenever the series whose variance increases linearly with the mean. A general class of variance-stabilizing transformations is given by **Cox-Box transformation** f_λ:

$$f_\lambda(X_t) = \begin{cases} \lambda^{-1}(X_t^\lambda - 1), & X_t \geq 0, \lambda > 0 \\ \ln X_t, & X_t > 0, \lambda = 0 \end{cases}$$

- In practice, λ is often 0 or 0.5
- We can use $powerTransform\{car\}$ to estimate λ.

Example with R using *wine* data

ARIMA models

Units Roots in Time Series

- The unit root problem arises when either the AR or MA polynomial of ARMA model has a root on or near to the unit circle.
- *A root near to 1 of the AR* polynomial suggest that the data **should be differenced** before fitting an ARMA model
- *A root near to 1 of the MA* polynomial suggest that the data **were overdifferenced.**

ARIMA models

Units Roots in Time Series:

- Augmented Dickey-Fuller Test for AR processes
- For MA process is more complicated (general case not fully resolved)

SARIMA models

We have already seen the how difference the series X_t at lag s is convenient way of eliminating a seasonal component of period s.

Definition

If d and D are nonnegative integers, then X_t is a seasonal $ARIMA(p, d, q)x(P, D, Q)_s$ process with period s if the difference series $Y_t = (1 - B)^d(1 - B^s)^D X_t$ is a causal ARMA process defined by

$$\phi(B)\Phi(B^s)Y_t = \theta(B)\Theta(B^s)Z_t$$

where B is the backward shift operator. $Z_t \sim WN(0, \sigma^2)$.
$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$, $\Phi(z) = 1 - \Phi_1 z - \cdots - \Phi_P z^P$,
$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$, and $\Theta(z) = 1 + \Theta_1 z + \cdots + \Theta_Q z^Q$

- *The process Y_t is causal if and only if $\phi(z) = 0$ and $\Phi(z) = 0$ for $\mid z \mid > 1$*

SARIMA models

A nonstationary process has often a seasonal component that repeats itself after a regular period of time. The seasonal period can be:

- Monthly: s=12 (12 observations per a year)
- Quarterly: s=4 (4 observations per a year)
- Daily: s=365 (365 observations per a year)
- Daily per week: s=5 (5 working days)
- Weekly: s=52 (observations per a year)

Example lecture note

Forecasting Techniques

We focused up to this point on fitting time series models for stationary and nonstationary series. We present now 2 techniques for forecasting:

- The ARAR algorithm: use AR-ARMA modeling and shortening.
- The Holt-Winters (HW) algorithm: use of exponential smoothing for forecasting.

See R code for examples.

Regression with ARMA errors

The regression model defines with ARMA errors is given by:

$$Y_t = \beta X_t + R_t \text{ where,}$$
$$R_t = \phi_1 R_{t-1} + \cdots + \phi_p R_{t-p} - \theta_1 z_{t-1} - \cdots - \theta_q z_{t-q} + z_t$$

The regression interpretations is the same as in the regression models.

STL decomposition

The **S**easonal and **T**rend decomposition using **L**OESS is a robut method for decomposing time series

- The seasonal component is allowed to change over time.
- It can be robust to outliers.

Cleveland, R. B., Cleveland, W. S., McRae, J. E., Terpenning, I. J. (1990). STL: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics, 6(1), 3-33.

Financial time series models

- Financial time series data are special because of their features, which include tail heaviness, asymmetry, volatility, and serial dependence without correlation.
- Let consider P_t the price of a stock of other financial asset at time t, then we can define the log return by $Z_t = \log(P_t) - \log(P_{t-1})$
- A model of the return series should support the fact that this data has a conditional variance h_t of Z_t is not independent of past value of Z_t.

ARCH model

The idea of the ARCH (autoregressive conditional heteroscedasticity) model is to incorporate the sequence h_t in the model by:

$$Z_t = \sqrt{h_T} e_t, \qquad e_t \sim IIDN(0,1) \tag{5}$$

h_t is known as the **volatility** and related to the past values of Z_t^2 via the **ARCH(p)** model:

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2, \tag{6}$$

The **GARCH(p,q)** (generalized ARCH) is given by:

$$h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Z_{t-i}^2 + \sum_{i=1}^{q} \beta_i h_{t-i}, \tag{7}$$

$\alpha_0 > 0$ and $\alpha_i \geq 0$, $\beta_i \geq 0$.