Time Series Analysis

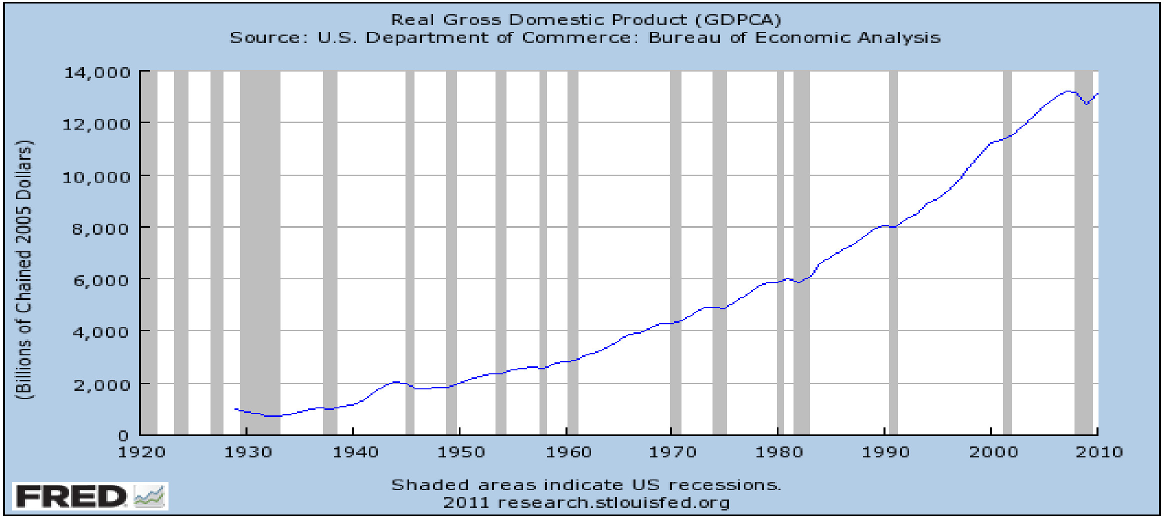
Samarjit Das
Economic Research Unit
Indian Statistical Institute
Kolkata-700108

What is Time Series?

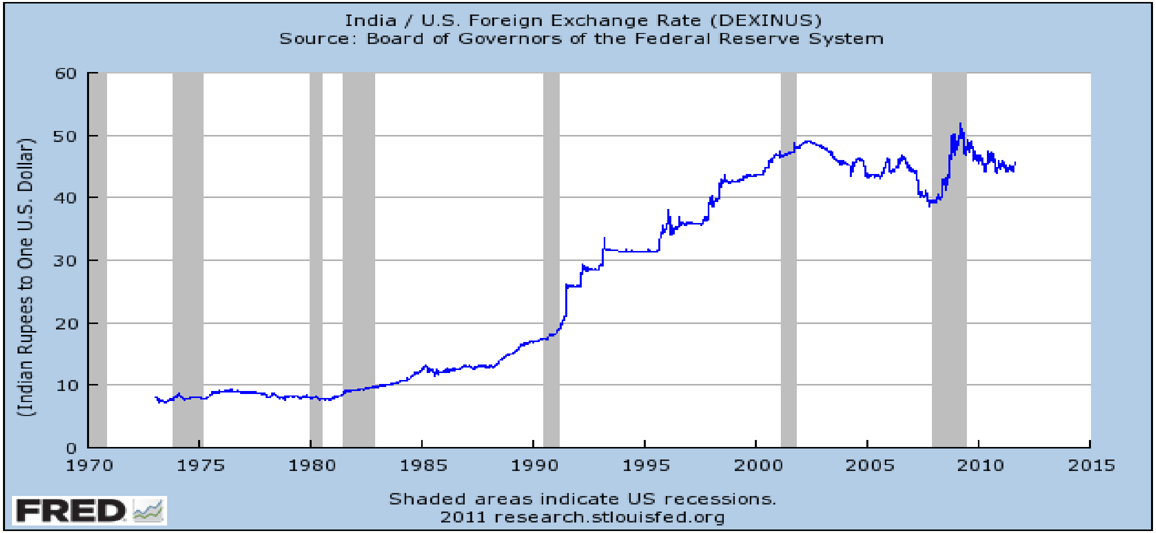
- *A Time series* is a set of observations, each one being recorded at a specific time. (Annual GDP of a country, Sales figure, etc)
- *A discrete time series* is one in which the set of time points at which observations are made is a discrete set. (All above including irregularly spaced data)
- *Continuous time series* are obtained when observations are made continuously over some time intervals. *It is a theoretical Concept*. (Roughly, ECG graph).
- *A discrete valued time series* is one which takes discrete values. (No of accidents, No of transaction etc.).

Few Time series Plots

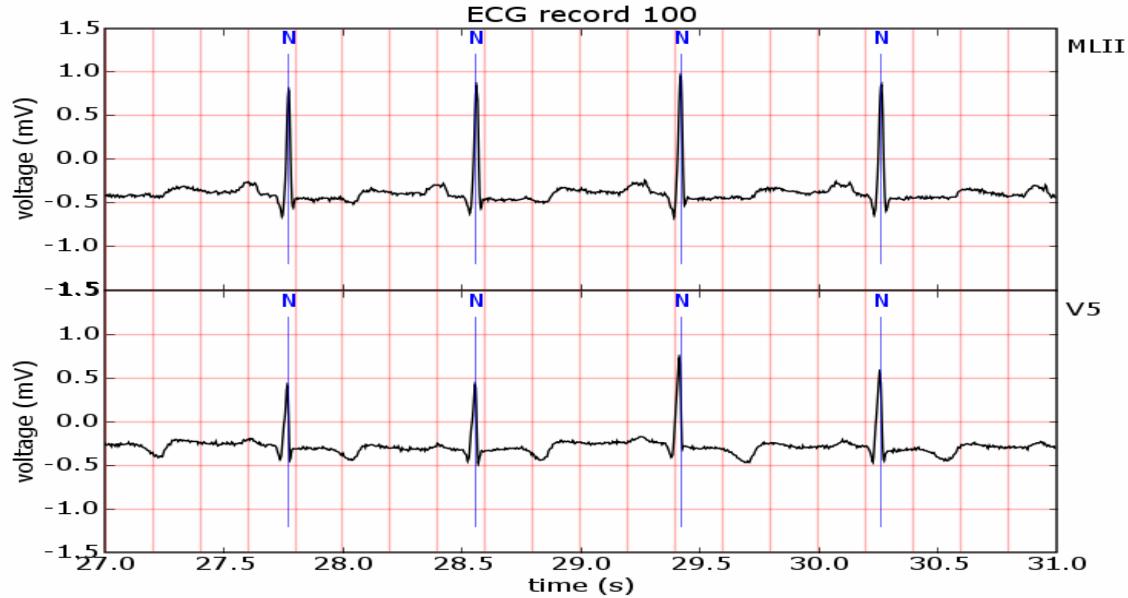
Annual GDP of USA



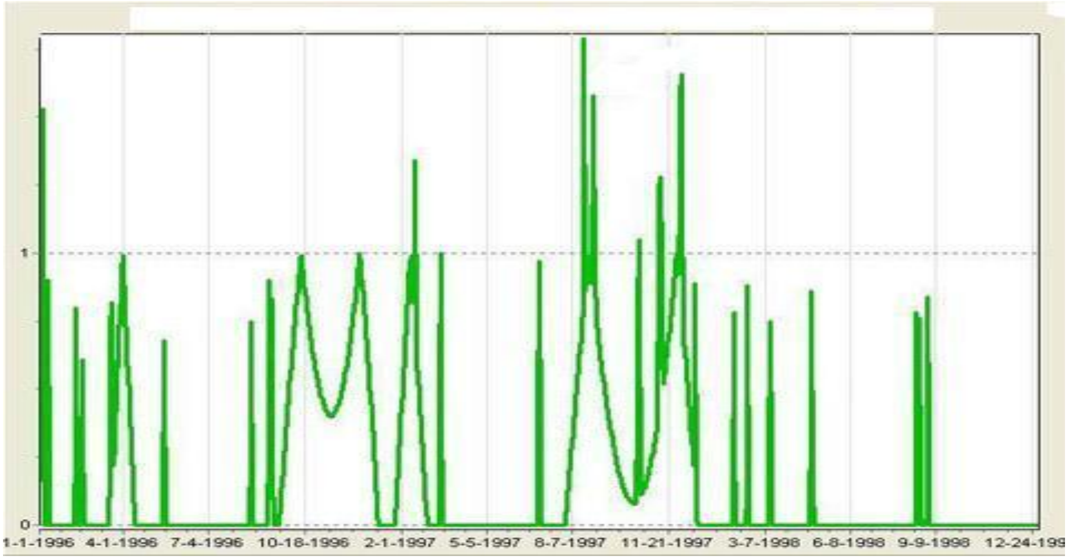
A discrete time series is one in which the setof time points at which observations are made is a discrete set. (All above including irregularly spaced data)



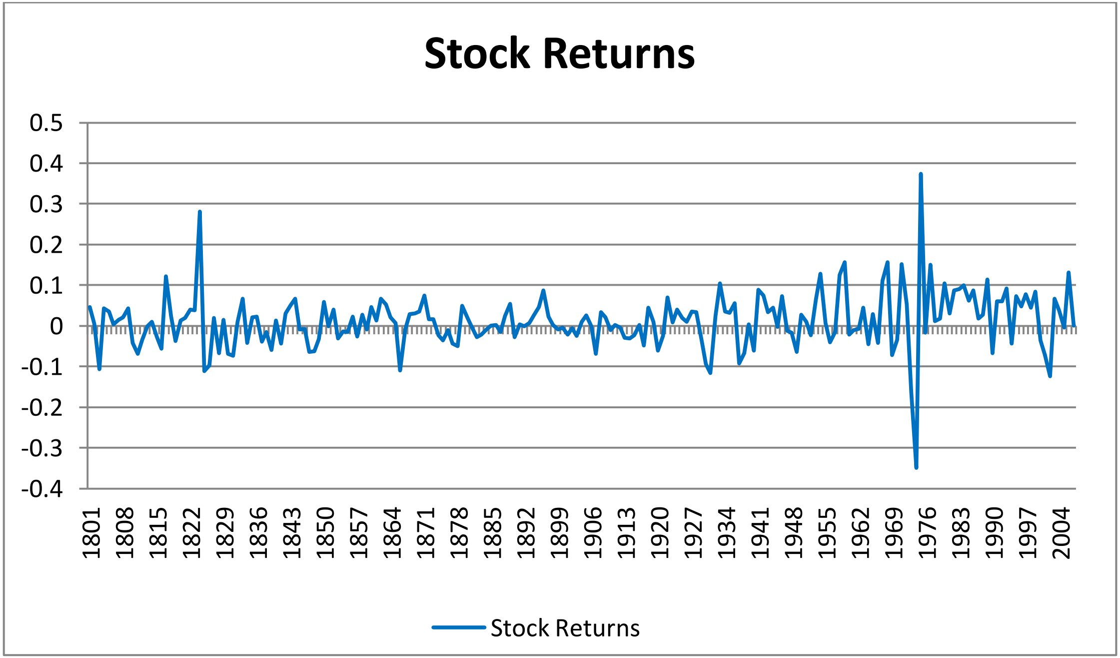
Continuous time series are obtained when observations are made continuously over some time intervals. (ECG graph).



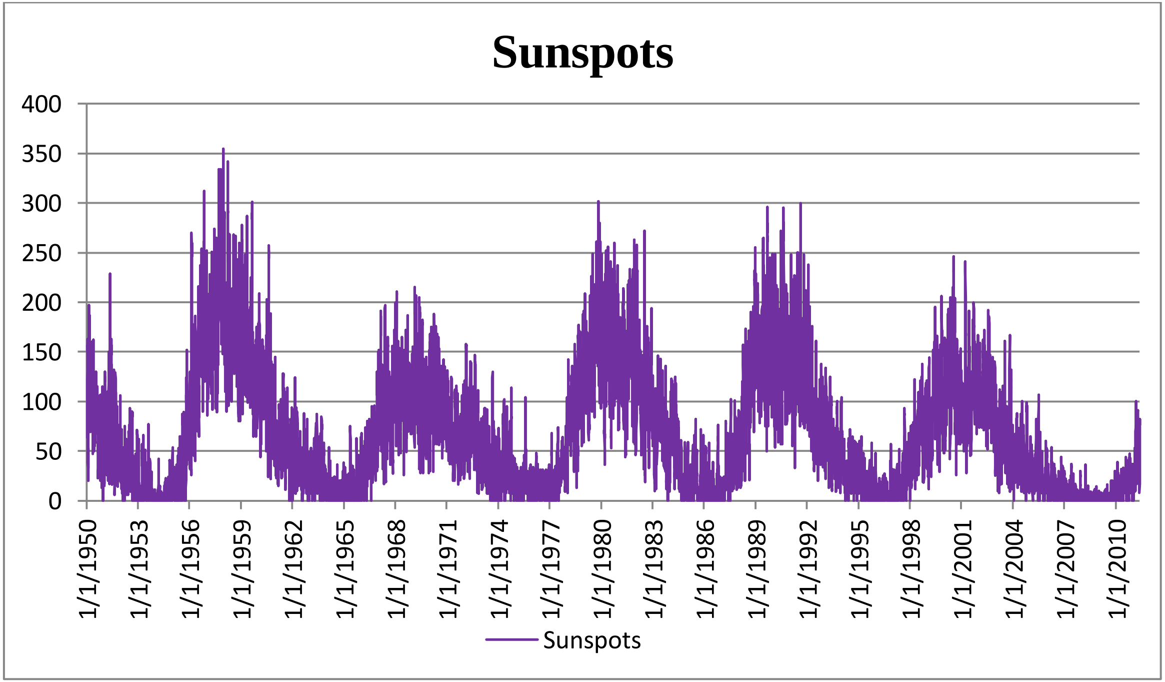
A discrete valued time series is one which takes discrete values.
(No of accidents, No of transaction etc.).
Time series plot on car accident in U.K.



Continuous time series data (Stock returns):

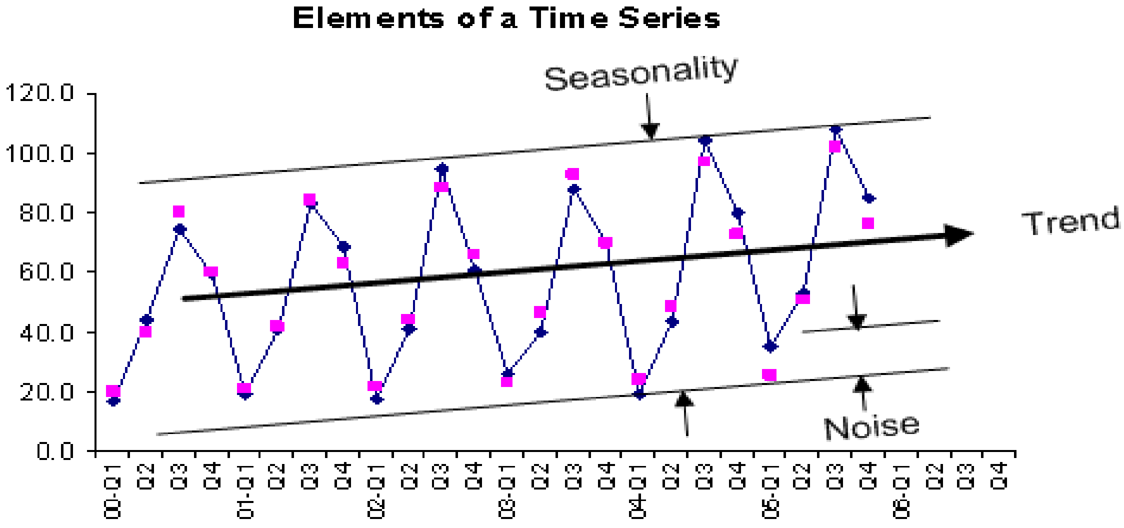


Time series data (Number of sunspots) showing cycles:



Quarterly Sales of Ice-cream

Q1-Dec-Jan



Objective of Time Series Analysis

- Forecasting (Knowing future is our innate wish).
- Control (whether anything is going wrong, think of ECG, production process etc)
- Understanding feature of the data including seasonality, cycle, trend and its nature. Degree of seasonality in agricultural price may indicate degree of development. Trend and cycle may mislead each other (Global temperature may be an interesting case)

Objective

- Description: Plot the data. Try to feel the data. Some descriptive statistics may be calculated to get some ideas about the data.
- Explanation: Deeper understanding of the mechanism that generated the time series.

Stochastic processes Approach

- Time series are an example of a stochastic or random process
- A stochastic process is a statistical phenomenon that evolves in time according to probabilistic laws.
- Mathematically, a stochastic process is an indexed collection of random variables

$$\{y_t : t \in T\}$$

Stochastic processes

- We are concerned only with processes indexed by time, either discrete time or continuous time processes such as

$$\{y_t : t \in (-\infty, \infty)\} = \{y_t : -\infty < t < \infty\}$$

Or

$$\{y_t : t \in (1,2,3, \dots)\} = \{y_{1,} y_{2,} y_3, \dots\}$$

Stochastic Process

- A stochastic process $\{y_t\}_{t=-\infty}^{\infty}$ is a collection of random variables or a process that develops in time according to probabilistic laws.
- The theory of stochastic processes gives us a formal way to look at time series variables.

DEFINITION

$$y(w, t): stochastic\ process$$

Sample Space   Index Set

- For a fixed t, $y(w, t)$ is a random variable.
- For a given w, $y(w, t)$ is called a sample function or a realization as a function of t.

Stochastic Process

- Time series is a realization or sample function from a certain stochastic process.
- A time series is a set of observations generated sequentially in time. Therefore, they are dependent to each other. This means that we do NOT have random sample.
- We assume that observations are equally spaced in time.
- We also assume that closer observations might have stronger dependency.

JOINT PDF OF A TIME SERIES

- Remember that

$$F_{y_1}(y_1): the\ marginal\ cdf$$

$$f_{y_1}(y_1): the\ marginal\ pdf$$

$$F_{y_1,y_2,\ldots,y_n}(y_1, y_2, \ldots, y_n): the\ joint\ cdf$$

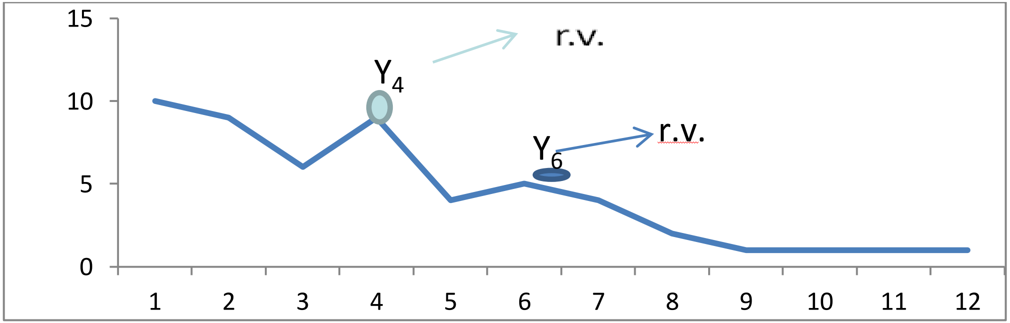
$$f_{y_1,y_2,\ldots,y_n}(y_1, y_2, \ldots, y_n): the\ joint\ pdf$$

JOINT PDF OF A TIME SERIES

- For the observed time series, say we have two points, t and s.
- The marginal pdfs: $f_{Y_t}(y_t) and\ f_{Y_s}(y_s)$
- The joint pdf: $f_{Y_t,Y_s}(y_t, y_s) \neq f_{Y_t}(y_t).f_{Y_s}(y_s)$

JOINT PDF OF A TIME SERIES

- Since we have only one observation for each r.v. Y_t, inference is too complicated if distributions (or moments) change for all t (i.e. change over time). So, we need a simplification.



JOINT PDF OF A TIME SERIES

- To be able to identify the structure of the series, we need the joint pdf of $y_1, y_2, \ldots, y_T$. However, we have only one sample (realization). That is, one observation from each random variable.
- This is in complete contrast to that of a cross-section/survey data. For cross section data, for a given population, we have a random sample. Based on the sample we try to infer about the population.

JOINT PDF OF A TIME SERIES

- In Time series, each random variable has one distribution/population. And from each population we have just one observation. So inference is not feasible unless we have some strong restrictive assumptions.
- Therefore, it is very difficult to identify the joint distribution. Hence, we need an assumption to simplify our problem. This simplifying assumption is known as **STATIONARITY**.

STATIONARITY

- The most vital and common assumption in time series analysis.
- The basic idea of stationarity is that the probability laws governing the process **do not** change with time.
- The process is in statistical equilibrium.

Why does Stationarity Assumption work?

- Now, suppose each distribution has same mean. In that case the common mean could be estimated based on the realization of size 'n'.
- We can visualize the fact in the following way---

Suppose we have 10 identical machines producing some item, *say*, bulb. Suppose each machine is run for one hour. Now it is easy to visualize that total (average) output by 10 machines is same as that of total (average) output by a single machine running for 10 hours.

TYPES OF STATIONARITY

- STRICT (STRONG OR COMPLETE) STATIONARY PROCESS: Consider a finite set of r.v.s. $Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n}$ from a stochastic process $\{Y(w,t); t = 0, \pm 1, \pm 2, \ldots\}$.

- The n-dimensional distribution function is defined by $F_{Y_{t_1}, Y_{t_2}, \ldots, Y_{t_n}}\left(y_{t_1}, y_{t_2}, \ldots, y_{t_n}\right) = P(w: Y_{t_1} < y_1, \ldots, Y_{t_1} < y_n)$

where y_i, $i = 1,2, \ldots, n$ are any real numbers.

STRONG STATIONARITY

- A process is said to be **first order stationary** in distribution, if its one dimensional distribution function is time-invariant, i.e., $F_{Y_{t_1}}(y_1) = F_{Y_{t_{1+k}}}(y_1)$ for any t_1 and k.
- **Second order stationary** in distribution if $F_{Y_{t_1},Y_{t_2}}(y_1, y_2) = F_{Y_{t_{1+k}},Y_{t_{2+k}}}(y_1, y_2)$ for any t_1, t_2 and k.
- **nth order stationary** in distribution if $F_{Y_{t_1},Y_{t_2},\ldots,Y_{t_n}}(y_1, y_2, \ldots, y_n) = F_{Y_{t_{1+k}},Y_{t_{2+k}},\ldots,Y_{t_{n+k}}}(y_1, y_2, \ldots, y_n)$ for any $t_1, \ldots, t_n$ and k.

STRONG STATIONARITY

n^{th} order stationarity in distribution = strong stationarity

$\rightarrow$ Shifting the time origin by an amount "k" has no effect on the joint distribution, which must therefore depend only on time intervals between $t_1, t_2 \ldots, t_n$ not on absolute time, t.

STRONG STATIONARITY

- So, for a strong stationary process

i. $f_{Y_{t_1},Y_{t_2},\dots,Y_{t_n}}(y_1, y_2, \dots, y_n) = f_{Y_{t_{1+k}},Y_{t_{2+k}},\dots,Y_{t_{n+k}}}(y_1, y_2, \dots, y_n)$

ii. $E(Y_t) = E(Y_{t+k}) \Rightarrow \mu_t = \mu_{t+k} = \mu \;\; \forall t, k$

Expected value of a series is constant over time, not a function of time

iii. $Var(Y_t) = Var(Y_{t+k}) \Rightarrow \sigma_t^2 = \sigma_{t+k}^2 = \sigma^2 \;\; \forall t, k$

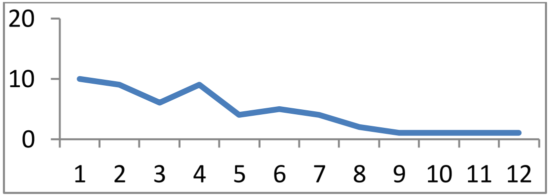
The variance of a series is constant over time, homoscedastic

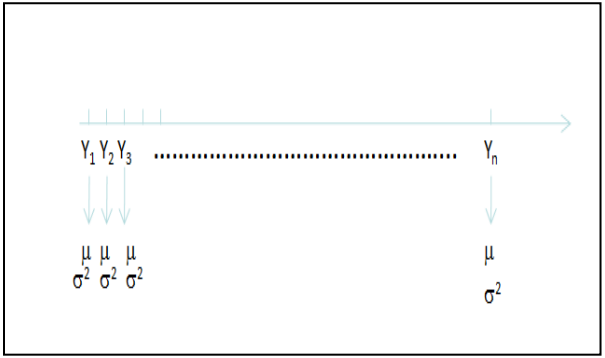
iv. $cov(Y_t, Y_s) = cov(Y_{t+k}, Y_{s+k}) \Rightarrow \gamma_{t,s} = \gamma_{t+k,s+k}, \forall t, k \;\; \Rightarrow$

$\gamma_{|t-s|} = \gamma_{|t+k-s-k|} = \gamma_h$

Not constant, not depend on time, depends on time interval, which we call “**lag**”, k.

STRONG STATIONARITY





$$cov(y_2 y_1) = \gamma_{2-1} = \gamma_1$$

$$cov(y_3 y_2) = \gamma_{3-2} = \gamma_1$$

$$cov(y_n y_{n-1}) = \gamma_{n-(n-1)} = \gamma_1$$

$$cov(y_3 y_1) = \gamma_{3-1} = \gamma_2$$

$$cov(y_1 y_3) = \gamma_{1-3} = \gamma_{-2}$$

Affected from time lag, k.

STRONG STATIONARITY

v. $corr[y_t y_s] = corr[y_{t+k} y_{s+k}] \Rightarrow \rho_{t,s} = \rho_{t+k,s+k} \ \forall t,k$

$$\Rightarrow \rho_{|t-s|} = \rho_{|t+k-s-k|} = \rho_h$$

Let $t = t - k$ and $s = t$,

$$\rho_{t,t-k} = \rho_{t+k,t} = \rho_k \quad \forall t, k$$

Remark: We have assumed the existence of 2nd order moments.

- It is usually impossible to verify a distribution particularly a joint distribution function from an observed time series. So, we use weaker sense of stationarity.

WEAK STATIONARITY

- WEAK (COVARIANCE) STATIONARITY OR STATIONARITY IN WIDE SENSE: A time series is said to be **covariance stationary** if its first and second order moments are unaffected by a change of time origin.
- That is, we have constant mean and variance with covariance and correlation beings functions of the time difference only.

WEAK STATIONARITY

$$E[y_t] = \mu, \quad \forall t$$

$$var[y_t] = \sigma^2 < \infty, \quad \forall t$$

$$cov[y_t, y_{t-k}] = \gamma_k, \quad \forall t$$

$$corr[y_t, y_{t-k}] = \rho_k, \quad \forall t$$

From, now on, when we say “stationary”, we imply weak stationarity.

EXAMPLE

- Consider a time series $\{Y_t\}$ where

$$Y_t = e_t$$

and $e_t \sim iid(0, \sigma^2)$. Is the process stationary?

EXAMPLE

- MOVING AVERAGE: Suppose that $\{Y_t\}$ is constructed as

$$y_t = \frac{e_t + e_{t-1}}{2}$$

And $e_t \sim iid(0, \sigma^2)$. Is the process $\{y_t\}$ stationary?

EXAMPLE

- RANDOM WALK

$$y_t = e_1 + e_2 + \cdots + e_t$$

where $e_t \sim iid(0, \sigma^2)$.. Is the process $\{y_t\}$ stationary?

EXAMPLE

- Suppose that time series has the form

$$y_t = a + bt + e_t$$

where a and b are constants and e_t is a weakly stationary process with mean 0 and autocovariance function γ_k. Is $\{y_t\}$ stationary?

EXAMPLE

$$y_t = (-1)^t e_t$$

where $e_t \sim iid(0, \sigma^2)$.. Is the process $\{y_t\}$ stationary?

STRONG VERSUS WEAK STATIONARITY

- Strict stationarity means that the joint distribution only depends on the 'difference' *h*, not the time $(t_1, \ldots, t_k)$.
- Finite variance is not assumed in the definition of strong stationarity, therefore, strict stationarity does not necessarily imply weak stationarity. For example, processes like i.i.d. Cauchy is strictly stationary but not weak stationary.
- A nonlinear function of a strict stationary variable is still strictly stationary, but this is not true for weak stationary. For example, the square of a covariance stationary process may not have finite variance.
- Weak stationarity usually does not imply strict stationarity as higher moments of the process may depend on time t.

STRONG VERSUS WEAK STATIONARITY

- If process $\{y_t\}$ is a Gaussian time series, which means that the distribution functions of $\{y_t\}$ are all multivariate Normal, weak stationary also implies strict stationary. This is because a multivariate Normal distribution is fully characterized by its first two moments.

STRONG VERSUS WEAK STATIONARITY

- For example, a white noise is stationary but may not be strict stationary, but a Gaussian white noise is strict stationary. Also, general white noise only implies uncorrelation while Gaussian white noise also implies independence. Because if a process is Gaussian, uncorrelation implies independence. Therefore, a Gaussian white noise is just $iid\ N(0, \sigma^2)$.

Measure of Dependence--Autocovariance

- Because the random variables comprising the process are not independent, we must also specify their covariance

$$\gamma_{t_1,t_2} = cov(y_{t_1}, y_{t_2})$$

Autocorrelation

- It is useful to standardize the autocovariance function (acvf)
- Consider stationary case only
- Use the autocorrelation function (acf)

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

Autocorrelation

- More than one process can have the same acf
- Properties are:

$\rho_0 = 1$

$\rho_k = \rho_{-k} \longrightarrow$ for stationary series

$|\rho_t| \leq 1$

Autocorrelation

Autocorrelation refers to the correlation of a time series with its own past and future values.

Autocorrelation is also sometimes called "lagged correlation" or "serial correlation", which refers to the correlation between members of a series of numbers arranged in time.

Positive autocorrelation might be considered a specific form of "persistence", a tendency for a system to remain in the same state from one observation to the next.

For example, the likelihood of tomorrow being rainy is greater if today is rainy than if today is dry.

Autocorrelations (contd.)

- A graph of the correlation values is called a "correlogram"
- Ideally, to obtain a useful estimate of the autocorrelation function, at least 50 observations are needed
- Generally, The estimated autocorrelations would be calculated up to lag no larger than N/4

Partial Autocorrelation(PAC)

As a complementary to ACF tool, we introduce the partial autocorrelation function, $\pi(t)$ which denotes the partial correlation between y_0 and y_t after adjusting for $y_1, \dots, y_{t-1}$. Let $e_j = y_j - E\left[y_j \middle| y_{j-1}, y_{j-2}, \dots, y_1\right]$, $j = 3,4, \dots$.

where $E[.]$ denotes linear regression of y_j on $y_{j-1}, y_{j-2}, \dots, y_1$. The quantity e_j are the residuals, i.e. what's left, after linear regression using the lagged observations.

- $\pi(0) = \mathrm{corr}(y_0, y_0) = 1.$
- $\pi(1) = \mathrm{corr}(y_1, y_0) = \rho(1)$
- $\pi(2) = \mathrm{corr}(y_2 - E[y_2|y_1], y_0 - E[y_0|y_1])$
- $\pi(3) = \mathrm{corr}(y_3 - E[y_3|y_2, y_1], y_0 - E[y_0|y_1, y_2])$
- $\pi(t) = corr(y_t - E[y_t|y_{t-1}, \dots, y_1], y_0 - E[y_0|y_1, y_2, \dots, y_{t-1}])$

PACF

- PACF is the correlation between y_t and y_{t-k} after their mutual linear dependency on the intervening variables $y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}$ has been removed.

- The conditional correlation

$$Corr(y_t, y_{t-k} | y_{t-1}, y_{t-2}, \ldots, y_{t-k+1}) = \phi_{kk}$$

is usually referred as the partial autocorrelation in time series.

e.g.,

$$\phi_{11} = corr(y_t, y_{t-1}) = \rho_1$$

$$\phi_{22} = corr(y_t, y_{t-2} | y_{t-1})$$

CALCULATION OF PACF

1. **REGRESSION APPROACH:** Consider a model

$$y_{t-k} = \phi_{k1} y_{t-k+1} + \phi_{k2} y_{t-k+2} + \cdots + \phi_{kk} y_t + e_{t-k}$$

from a zero mean stationary process where ϕ_{ki} denotes the coefficients of y_{t-k+i} and e_{t-k} is the zero mean error term which is uncorrelated with y_{t-k+i} , $i = 0,1, \dots, k$

- Multiply both sides by y_{t-k+j}

$$y_{t-k} y_{t-k+j} = \phi_{k1} y_{t-k+1} y_{t-k+j} + \cdots + \phi_{kk} y_t y_{t-k+j} + e_{t-k} y_{t-k+j}$$

CALCULATION OF PACF

and taking the expectations

$$\gamma_j = \phi_{k1}\gamma_{j-1} + \phi_{k2}\gamma_{j-2} + \cdots + \phi_{kk}\gamma_{j-k}$$

diving both sides by γ_0

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \cdots + \phi_{kk}\rho_{j-k}$$

Pacf

CALCULATION OF PACF

- For *j=1,2,…,k,* we have the following system of equations

$$\rho_1 = \phi_{k1} + \phi_{k2}\rho_1 + \cdots + \phi_{kk}\rho_{k-1}$$

$$\rho_2 = \phi_{k1}\rho_1 + \phi_{k2} + \cdots + \phi_{kk}\rho_{k-2}$$

$$\cdots\cdot$$

$$\rho_k = \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \cdots + \phi_{kk}$$

CALCULATION OF PACF

- Using Cramer's rule successively for $k = 1,2, \ldots$

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & \rho_2 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{vmatrix}} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

CALCULATION OF PACF

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

CALCULATION OF PACF

2. Levinson and Durbin’s Recursive Formula:

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\, \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\, \rho_{k-j}}$$

Where

$$\phi_{kj} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j}, \quad j = 1,2,\ldots k-1$$

Some Popular Stochastic Processes

1. White Noise:

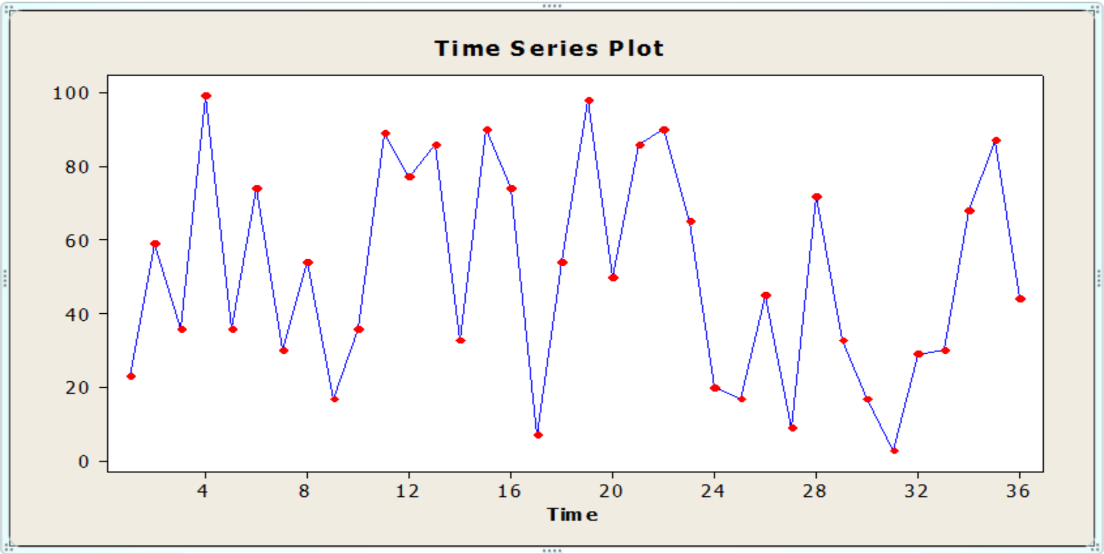
White noise

- This is a purely random process, a sequence of uncorrelated random variables
- Has constant mean and variance
- Also

$$\gamma_k = cov(y_t, y_{t+k}) = 0, \quad k \neq 0$$

$$\gamma_k = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}$$

An Illustrative plot of a white noise series



2. Random Walk -- A Non-stationary Process

Random walk

- Start with $\{y_t\}$ being white noise or purely random
- $\{y_t\}$ is a random walk if

$$y_0 = 0$$

$$y_t = y_t + e_t = \sum_{k=0}^{t} e_t$$

Random walk

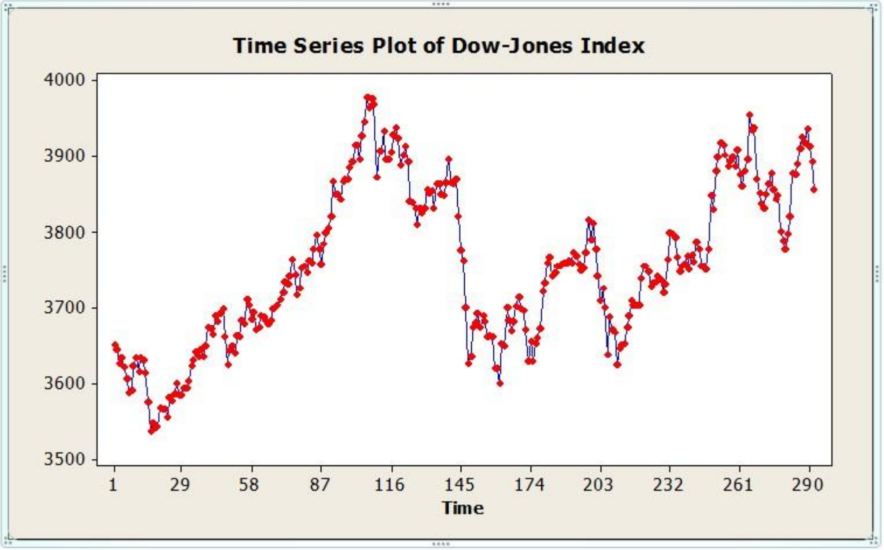
- The random walk is not stationary

$$E(y_t) = 0, Var(y_t) = t\sigma^2$$

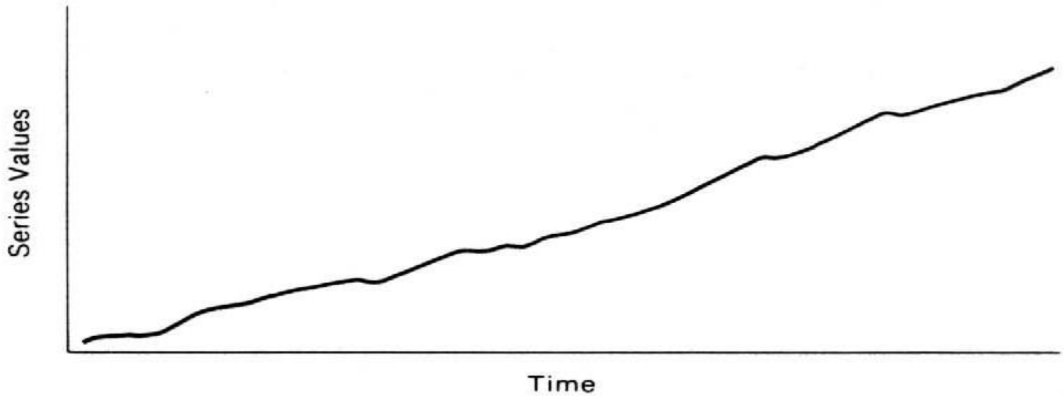
- First differences are stationary

$$\Delta y_t = y_t - y_{t-1} = e_t$$

An Illustrative plot of a Random Walk

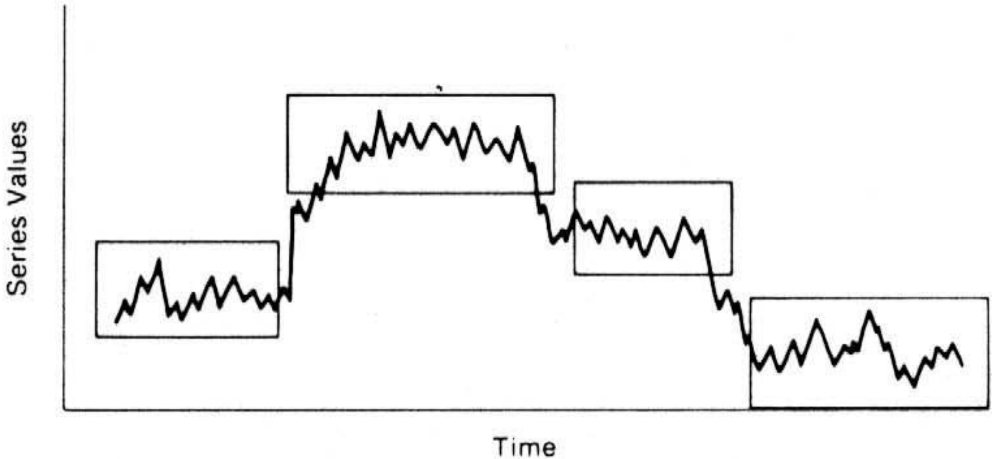


Some Other nonstationary series



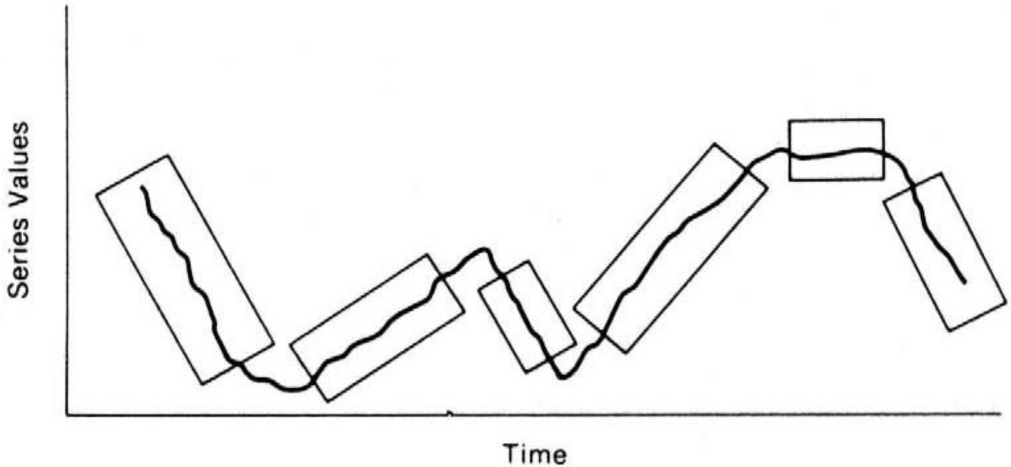
A Nonstationary Series: Overall Trend

Some nonstationary series (cont.)



A Nonstationary Series: Random Changes in Level

Some nonstationary series (cont.)



A Nonstationary Series: Random Changes in Both Level and Slope

3. Moving Average Processes

MOVING AVERAGE PROCESSES

- Suppose you win 1 Dollar if a fair coin shows a head and lose 1 Dollar if it shows tail. Denote the outcome on toss t by a_t.

$$e_t = \begin{cases} 1, head\ shows\ up \\ -1, tail\ shows\ up \end{cases}$$

- The average (y_t)winning from the 4 tosses:

$y_t = \frac{1}{2}e_t + \frac{1}{2}e_{t-1} + \frac{1}{2}e_{t-2} + \frac{1}{2}e_{t-3}$ ⇒ Moving average process

MOVING AVERAGE PROCESSES

- Notice that the observed series (y_t) is autocorrelated even though the generating series e_t is uncorrelated.
- The series $(y_t)$is the weighted aggregation of some uncorrelated random variables.
- In Economics, the generating series, e_t, is called the *random shock*.
- Random shocks are generally unobserved and are thought to be some unobserved economic activity.

MOVING AVERAGE PROCESSES

Consider a simple example: $y_t = e_t + \theta e_{t-1}$

Let y_t be the return in stock market. Assume *theta* (θ) is positive. So a good news from yesterday or a positive activity in yesterday has a positive impact on today's return.

Moving average processes

- Start with being $\{e_t\}$ white noise or purely random, mean zero, s.d. σ_e

- $\{y_t\}$ is a moving average process of order q (written MA(q) if for some constants $\theta_0, \theta_1, \ldots, \theta_q$ we have

$$y_t = \theta_0 e_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}$$

Usually $\theta_0 = 1$.

Moving average processes

- The mean and variance are given by

$$E(y_t) = 0, \quad var(y_t) = \sigma_e^2 \sum_{k=0}^{q} \theta_k^2$$

Moving average processes

- If the e_t 's are normal then so is the process, and it is then strictly stationary.

- The autocorrelation is

$$\rho_k = \begin{cases} 1 \;\; if\; k = 0 \\ \sum_{i=0}^{q-k} \theta_i \theta_{i+k} / \sum_{i=0}^{q} \theta_i^2 \;\; if\; k = 1,2,\dots,q \\ 0 \;\; if\; k > q \\ \rho_{-k} \;\; if\; k < 0 \end{cases}$$

The process is weakly stationary because the mean is constant and the covariance does not depend on t.

Moving average processes

- Note the autocorrelation cuts off at lag q
- For the MA(1) process with $\theta_0 = 1$.

$$\rho_k = \begin{cases} 1 & if\ k = 0 \\ \theta_1 / 1 + \theta_1^2 & if\ k = \pm 1 \\ 0 & otherwise \end{cases}$$

Moving average processes

- In order to ensure there is a unique MA process for a given acf, we impose the condition of invertibility

- This ensures that when the process is written in series form, the series converges

- For the MA(1) process $y_t = e_t + \theta e_{t-1}$, the condition is $|\theta| < 1$

Check that

$y_t = e_t + \frac{1}{5} e_{t-1}$ and

$y_t = e_t + 5e_{t-1}$

Both have the same autocorrelation function

The value of $\theta_1 /\, 1 + \theta_1^2$ is same for $\theta_1 = 5 \; and \, \frac{1}{5}$. The 1st one is invertible but 2nd one is NOT.

Moving average processes

- For general processes introduce the backward shift operator B.

$$B^j y_t = y_{t-j}$$

- Then the MA(q) process is given by

$$y_t = \left(\theta_0 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^{2q}\right) e_t = \theta(B) e_t$$

Moving average processes

- The general condition for invertibility is that all the roots of the equation $\theta(B) = 0$ lie outside the unit circle (have modulus less than one)

MA: Stationarity

- Consider an MA(1) process without drift:

$$y_t = e_t + \theta e_{t-1}$$

- It can be shown, regardless of the value of, that

$$E(y_t) = 0$$

$$var(y_t) = \sigma_e^2(1 + \theta^2)$$

$$cov(y_t y_{t-s}) = \begin{cases} -\theta\sigma_e^2 & if\ s = 1 \\ 0 & otherwise \end{cases}$$

MA: Stationarity

- For an MA(2) process

$$y_t = e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}$$

$$E(y_t) = 0$$

$$var(y_t) = \sigma_e^2(1 + \theta_1^2 + \theta_2^2)$$

$$cov(y_t y_{t-s}) = \begin{cases} -\theta_1 \sigma_e^2 (1 - \theta_2) & if\ s = 1 \\ -\theta_2 \sigma_e^2 & if\ s = 2 \\ 0 & otherwise \end{cases}$$

MA: Stationarity

- In general, MA processes are stationarity regardless of the values of the parameters, but not necessarily “invertible”.
- An MA process is said to be invertible if it can be converted into a stationary AR process of infinite order.
- In order to ensure there is a unique MA process for a given acf, we impose the condition of **invertibility.**
- Therefore, invertibility condition for MA process servers two purposes: (a) it is useful to represent an MA process as an (infinite order) AR process; and (b) it ensures that for a given ACF, there is an unique MA process.

4. Autoregressive Process

Autoregressive processes

- Assume $\{e_t\}$ is purely random with mean zero and s.d. σ_e
- Then the autoregressive process of order p or AR(p) process is

$$y_t = \varphi_1 y_t + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + e_t$$

Autoregressive processes

- The first order autoregression is

$$y_t = \varphi y_t + e_t$$

- Provided $|\varphi| < 1$ it may be written as an infinite order MA process
- Using the backshift operator we have

$$(1 - \varphi B)y_t = e_t$$

Autoregressive processes

- From the previous equation we have

$$y_t = \frac{e_t}{(1 - \varphi B)}$$

$$y_t = (1 + \varphi B + \varphi^2 B^2 + \cdots)e_t$$

$$y_t = e_t + \varphi e_{t-1} + \varphi^2 e_{t-2} + \cdots$$

Autoregressive processes

- Then $E(y_t) = 0$, and if $|\varphi| < 1$

$$var(y_t) = \sigma_y^2 = \sigma_e^2/(1 - \varphi^2)$$

$$\gamma_k = \varphi^k \sigma_e^2/(1 - \varphi^2)$$

$$\rho_k = \varphi^k$$

Autoregressive processes

- The AR(p) process can be written as

$$\left(1 + \varphi_1 B + \varphi_2 B^2 + \cdots + \varphi_p B^p\right) y_t = e_t$$

$$or$$

$$y_t = e_t / \left(1 + \varphi_1 B + \varphi_2 B^2 + \cdots + \varphi_p B^p\right) = f(B) e_t$$

Autoregressive processes

- This is for

$$f(B) = \left(1 + \varphi_1 B + \varphi_2 B^2 + \cdots + \varphi_p B^p\right)^{-1}$$
$$f(B) = \left(1 + \beta_1 B + \beta_2 B^2 + \cdots + \beta_p B^p\right)$$

for some $\beta_1, \beta_2, \ldots$.
This gives y_t as an infinite MA process, so it has mean zero

Autoregressive processes

- Conditions are needed to ensure that various series converge, and hence that the variance exists, and the autocovariance can be defined

- Essentially these are requirements that the β_i become small quickly enough, for large i

Autoregressive processes

- The β_i may not be able to be found however.
- The alternative is to work with the φ_i
- The acf is expressible in terms of the roots $\pi_i\ i = 1,2,\ldots,p$of the auxiliary equation

$$y^p - \varphi_1 y^{p-1} - \cdots \varphi_p = 0$$

Autoregressive processes

- Then a necessary and sufficient condition for stationarity is that for every i $|\pi_i| < 1$
- An equivalent way of expressing this is that the roots of the equation

$$f(B) = \left(1 + \varphi_1 B + \varphi_2 B^2 + \cdots + \varphi_p B^p\right)$$

must lie outside the unit circle.

AR: Stationarity

- Suppose y_t follows an AR(1) process without drift.
- Is y_t stationarity?
- Note that

$$y_t = \varphi_1 y_{t-1} + e_t$$
$$y_t = \varphi_1(\varphi_1 y_{t-2} + e_{t-1}) + e_t$$

$$y_t = e_t + \varphi_1 e_{t-1} + \varphi_1^2 e_{t-2} + \varphi_1^3 e_{t-3} + \cdots + \varphi_1^t y_0$$

Stationarity

- Without loss of generality, assume that $y_0 = 0$. Then $E(y_t) = 0$.
- Assuming that t is large, i.e., the process started a long time ago, then $\mathrm{var}(y_t) = \frac{\sigma^2}{(1-\phi_1^2)}$, provided that $|\phi_1| < 1$. It can also be shown that provided that the same condition is satisfied, $\mathrm{cov}(y_t y_{t-s}) = \frac{\phi_1^s \sigma^2}{(1-\phi_1^2)} = \phi_1^s \ \mathrm{var}(y_t)$

Stationarity

- Suppose the model is an AR(2) without drift,

  i.e., $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$
- It can be shown that for y_t to be stationary,
- The key point is that AR processes are not stationary unless appropriate prior conditions are imposed on the parameters.

$$\phi_1 + \phi_2 < 1, \phi_2 - \phi_1 < 1 \text{ and } |\phi_2| < 1$$

5. Autoregressive and Moving Average (ARMA) Processes

ARMA processes

- Combine AR and MA processes
- An ARMA process of order (p,q) is given by

$$y_t = \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + e_t + \beta_1 e_{t-1} + \cdots + \beta_1 e_{t-q}$$

ARMA processes

- Alternative expressions are possible using the backshift operator

$$\varphi(B)y_t = \theta(B)e_t$$

Where

$$\varphi(B) = 1 + \alpha_1 B + \cdots + \alpha_p B^p$$

$$\theta(B) = 1 + \beta_1 B + \cdots + \beta_q B^q$$

ARMA processes

- An ARMA process can be written in pure MA or pure AR forms, the operators being possibly of infinite order

$$y_t = \psi(B)e_t$$

$$\pi(B)y_t = e_t$$

- Usually the mixed form requires fewer parameters

6. ARIMA—Integrated ARMA

ARIMA processes

- General autoregressive integrated moving average processes are called ARIMA processes
- When differenced say d times, the process is an ARMA process
- Call the differenced process W_t. Then Wt is an ARMA process and

$$W_t = \Delta^d y_t = (1 - B)^d y_t$$

ARIMA processes

- Alternatively specify the process as

$$\varphi(B)W_t = \theta(B)e_t$$

Or

$$\varphi(B)(1 - B)^d y_t = \theta(B)e_t$$

- This is an ARIMA process of order (p,d,q)

ARIMA processes

- The model for y_t is non-stationary because the AR operator on the left hand side has d roots on the unit circle
- d is often 1
- Random walk is ARIMA(0,1,0)
- Can include seasonal terms

Non-zero mean

- We have assumed that the mean is zero in the ARIMA models
- There are two alternatives
  - mean correct all the W_t terms in the model
  - incorporate a constant term in the model

ACF and PACF for some useful Models

Summary of the Behavior of autocorrelation and partial autocorrelation functions

Behavior of autocorrelation and partial autocorrelation functions

Model	AC	PAC
Autoregressive of order p $y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ \ldots . + \phi_p y_{t-p} + e_t$	Dies down	Cuts off after lag p
Moving Average of order q $y_t = \delta + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ \ldots . - \theta_q e_{t-q}$	Cuts off after lag q	Dies down
Mixed Autoregressive-Moving Average of order (p,q) $y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ \ldots . + \phi_p y_{t-p}$ $+ e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ \ldots . - \theta_q e_{t-q}$	Dies down	Dies down

Summary of the Behavior of autocorrelation and partial autocorrelation functions

Behavior of AC and PAC for specific non-seasonal models

Model	AC	PAC
First-order autoregressive $y_t = \delta + \phi_1 y_{t-1} + e_t$	Dies down in a damped exponential fashion; specifically: $\rho_k = {\phi_1}^k \; for \; k \geq 1$	Cuts off after lag 1
Second-order autoregressive $y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$	Dies down according to a mixture of damped exponential and /or damped sine waves; specifically: $\rho_1 = \dfrac{\phi_1}{1 - \phi_2},$ $\rho_2 = \phi_1 + \frac{{\phi_1}^2}{1-\phi_2};$ $\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \; for \; k \geq 3$	Cuts off after lag 2

Summary of the Behavior of autocorrelation and partial autocorrelation functions

Behavior of AC and PAC for specific non-seasonal models

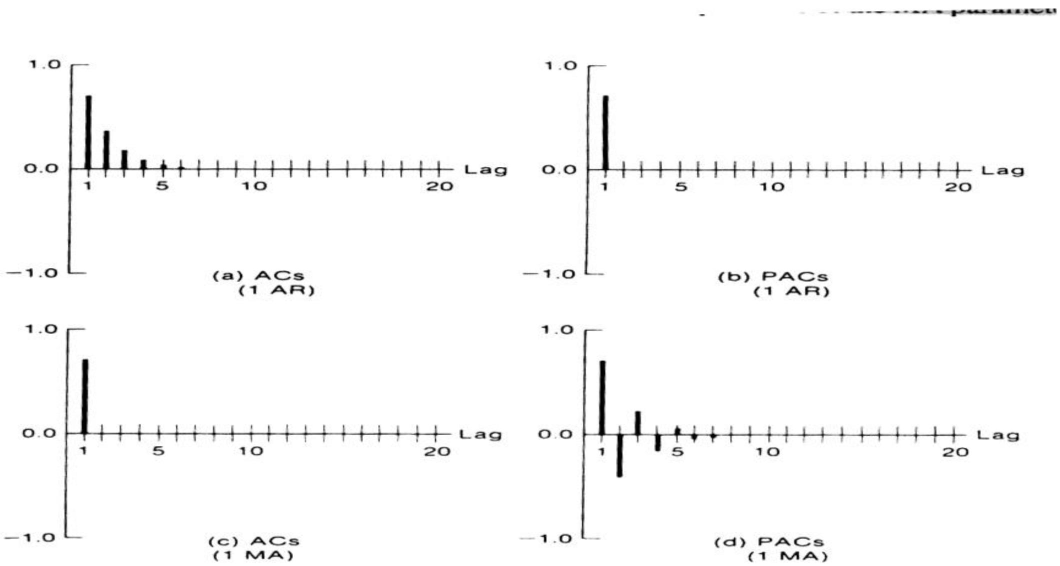
Model	AC	PAC
First-order moving average $y_t = \delta + e_t - \theta_1 e_{t-1}$	Cuts off after lag 1; specifically: $\rho_1 = \frac{-\theta_1}{1 + {\theta^2}_1}$ $\rho_k = 0 \quad for\ k \geq 2$	Dies down in a fashion dominated by damped exponential decay
Second-order moving average $y_t = \delta + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$	Cuts off after lag 2; specifically: $\rho_1 = \frac{-\theta_1(1-\theta_2)}{1 + {\theta^2}_1 + {\theta^2}_2}$, $\rho_2 = \frac{-\theta_2}{1 + {\theta^2}_1 + {\theta^2}_2}$; $\rho_k = 0 \quad for\ k > 2$	Dies down according to a mixture of damped exponentials and /or damped sine waves

Summary of the Behavior of autocorrelation and partial autocorrelation functions

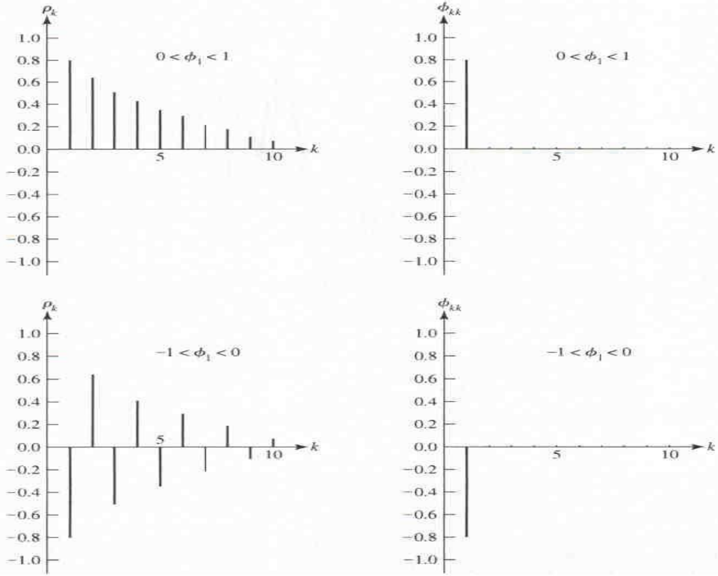
Behavior of AC and PAC for specific non-seasonal models

Model	AC	PAC
Mixed autoregressive-moving average of order (1,1) $y_t = \delta + \phi_1 y_{t-1} + e_t - \theta_1 e_{t-1}$	Dies down in a damped exponential fashion; specifically: $\rho_1 = \frac{(1-\phi_1\theta_1)(\phi_1-\theta_1)}{1+\theta^2{}_1-2\phi_1\theta_1}$, $\rho_k = \phi_1\rho_{k-1} \quad for\ k \geq 2$	Dies down in a fashion dominated by damped exponential decay

Theoretical ACs and PACs (cont.)



AR(1) PROCESS



ρ_k
$0 < \phi_1 < 1$
ϕ_{kk}
$0 < \phi_1 < 1$
ρ_k
$-1 < \phi_1 < 0$
ϕ_{kk}
$-1 < \phi_1 < 0$
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4
−0.6
−0.8
−1.0
5
10
k

AR(2) PROCESS

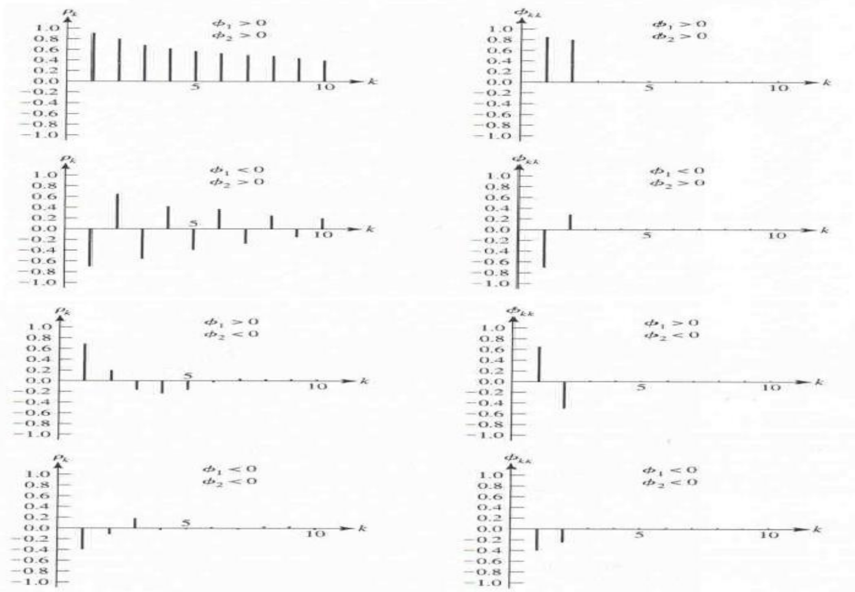


FIGURE ACF and PACF of AR(2) process: $(1 - \phi_1 B - \phi_2 B^2)\dot{Z}_t = a_t$.

MA(1) PROCESS

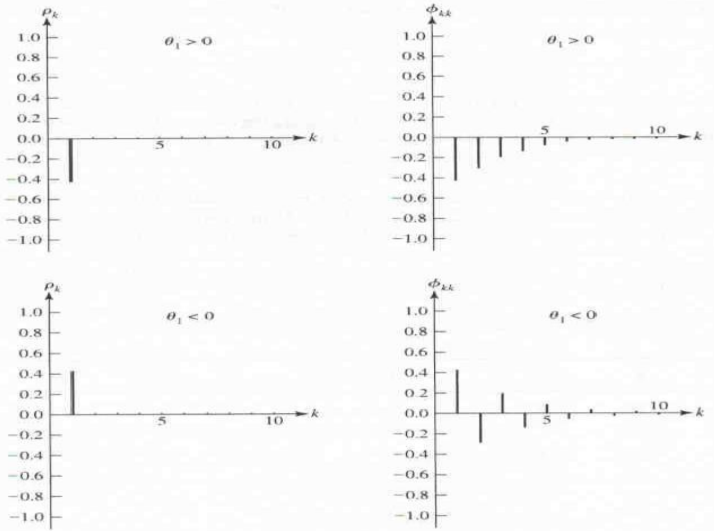


FIGURE ACF and PACF of MA(1) processes: $\dot{Z}_t = (1 - \theta_1 B)a_t$.

MA(2) PROCESS

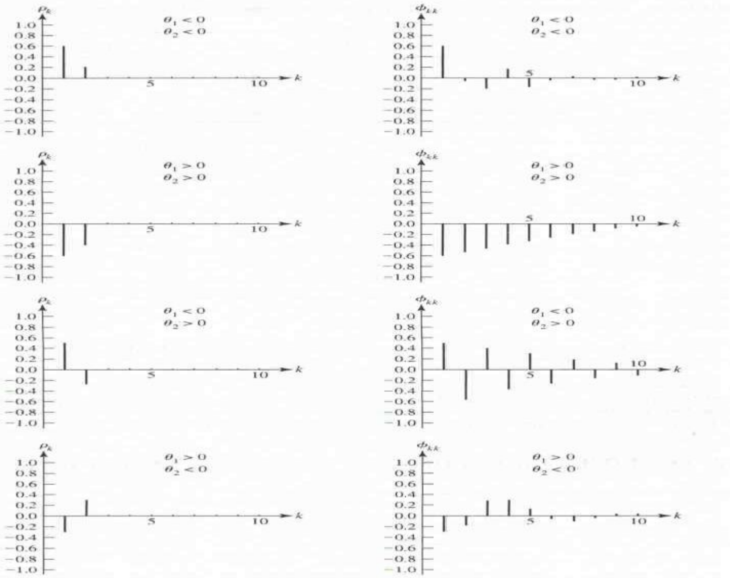


FIGURE ACF and PACF of MA(2) processes: $Z_t = (1 - \theta_1 B - \theta_2 B^2)a_t$.

ARMA(1,1) PROCESS

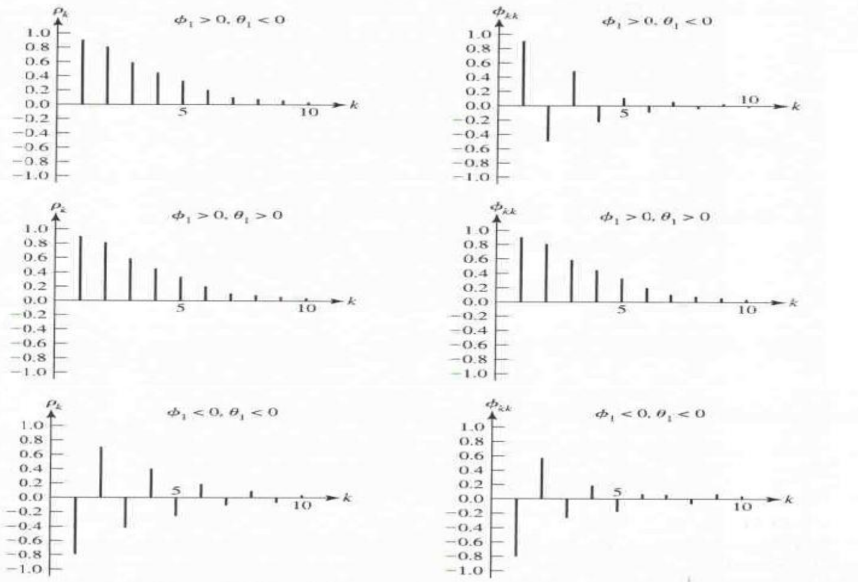


FIGURE ACF and PACF of ARMA(1, 1) model $(1 - \phi_1 B)\dot{Z}_t = (1 - \theta_1 B)a_t$.

ARMA(1,1) PROCESS (contd.)

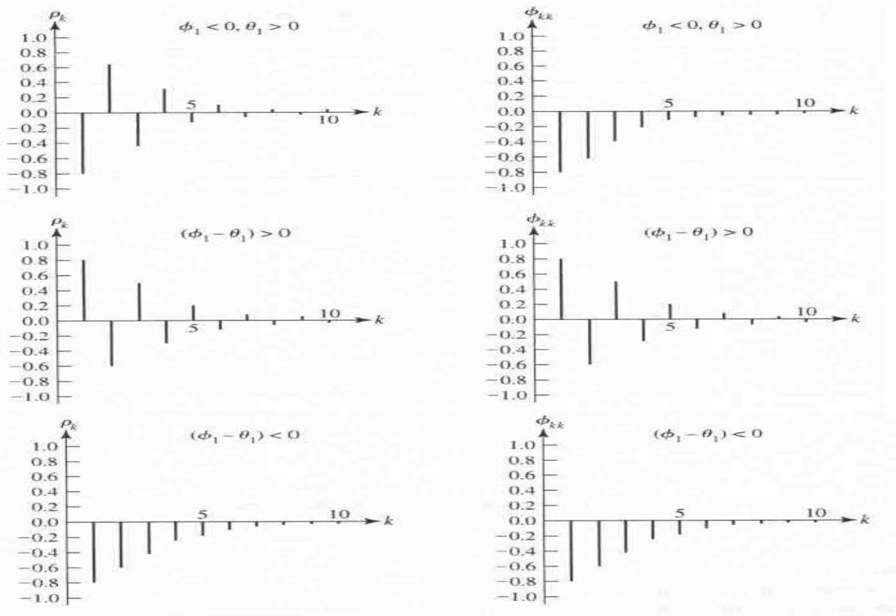
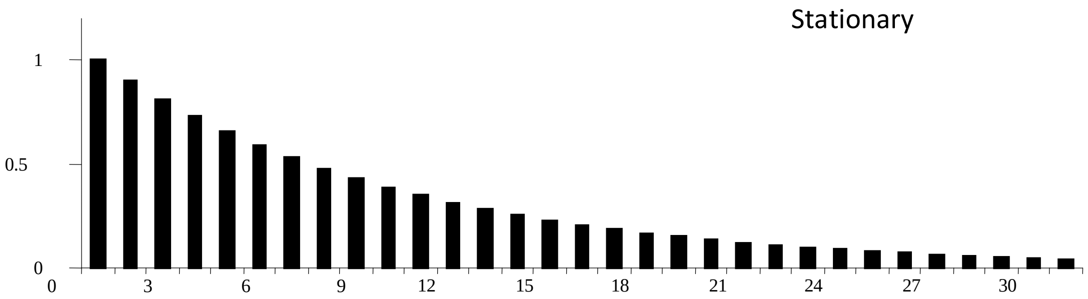
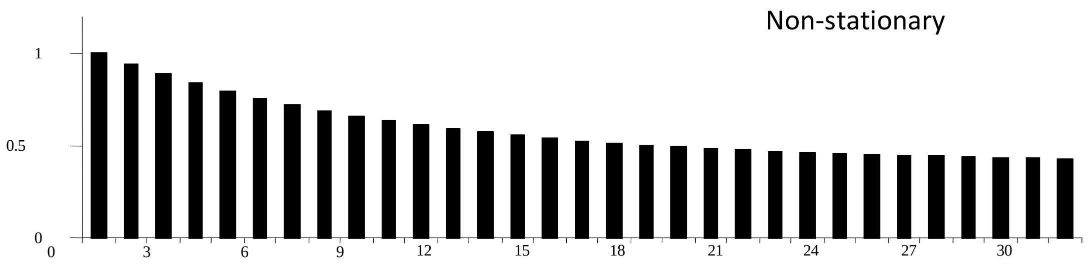


FIGURE (*Continued*)


Stationary
1
0.5
0
0
3
6
9
12
15
18
21
24
27
30


Non-stationary
1
0.5
0
0
3
6
9
12
15
18
21
24
27
30

THE SAMPLE AUTOCORRELATION FUNCTION

$$\hat{\rho}_k = r_k = \frac{\sum_{t=1}^{n-k}\left(Y_t - \overline{Y}\right)\left(Y_{t-k} - \overline{Y}\right)}{\sum_{t=1}^{n}\left(Y_t - \overline{Y}\right)^2}, k = 0,1,2,..$$

I. A plot $\hat{\rho}_k$ versus $k \rightarrow$ a sample correlogram.

II. For large sample sizes, $\hat{\rho}_k$ is normally distributed with mean ρ_k and variance is approximated by Bartlett's approximation for processes in which $\rho_k = 0$ for $k>m$.

THE SAMPLE AUTOCORRELATION FUNCTION

$$Var(\hat{\rho}_k) \approx \frac{1}{n}\left(1 + 2\rho_1^2 + 2\rho_2^2 + \cdots + 2\rho_m^2\right)$$

I. In practice, ρ_i's are unknown and replaced by their sample estimates, $\hat{\rho}_i$. Hence, we have the following large-lag standard error of $\hat{\rho}_k$:

$$s_{\hat{\rho}_k} = \sqrt{\frac{1}{n}\left(1 + 2\hat{\rho}_1^2 + 2\hat{\rho}_2^2 + \cdots + 2\hat{\rho}_m^2\right)}$$

THE SAMPLE AUTOCORRELATION FUNCTION

I. For a WN process, we have

$$s_{\hat{\rho}_k} = \sqrt{\frac{1}{n}}$$

II. The ~95% confidence interval for ρ_k:

$$\hat{\rho}_k \pm 2\frac{1}{\sqrt{n}}$$

For a WN process, it must be close to zero.

III. Hence, to test the process is WN or not, draw a $\pm 2/n^{1/2}$ lines on the sample correlogram. If all $\hat{\rho}_k$ are inside the limits, the process could be WN (we need to check the sample PACF, too).

THE SAMPLE PARTIAL AUTOCORRELATION FUNCTION

$$\hat{\phi}_{11} = \hat{\rho}_1$$

$$\hat{\phi}_{kk} = \frac{\hat{\rho}_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j}\,\hat{\rho}_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j}\,\hat{\rho}_{k-j}}$$

$$where\ \hat{\phi}_{kj} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk}\hat{\phi}_{k-1,k-j},\ j = 1,2,\cdots,k-1.$$

I. For a WN process, $Var\left(\hat{\phi}_{kk}\right) \approx \frac{1}{n}$

II. $\pm 2/n^{1/2}$ can be used as critical limits on ϕ_{kk} to test the hypothesis of a WN process.

Sample Partial Autocorrelation Function (SPAC)

I. r_{kk} may intuitively be thought of as the sample autocorrelation of time series observations separated by a lag k time units with the effects of the intervening observations eliminated.

II. The standard error of r_{kk} is $S_{r_{kk}} = \sqrt{\frac{1}{n}}$.

III. The $t_{r_{kk}}$ statistic is $t_{r_{kk}} = \frac{r_{kk}}{S_{r_{kk}}}$.

Box-Jenkins Methodology (ARIMA Models)

Box-Jenkins Methodology (ARIMA Models)

I. The Box-Jenkins methodology refers to a set of procedures for identifying and estimating time series models within the class of autoregressive integrated moving average (ARIMA) models.

II. ARIMA models are regression models that use lagged values of the dependent variable and/or random disturbance term as explanatory variables.

III. ARIMA models rely heavily on the autocorrelation pattern in the data

IV. This method applies to both non-seasonal and seasonal data.

Box-Jenkins Methods--A five-step iterative procedure

I. Stationarity Checking and Differencing
II. Model Identification
III. Parameter Estimation
IV. Diagnostic Checking
V. Forecasting

Step One: Stationarity checking

Non-Stationary

- Not-stationary = Non-stationary, when distribution (parameters) changes over time.

  Various important examples are:

  Deterministic trend and Stochastic trend.

Deterministic Trend (TSP)

$$y_t = \alpha + \beta t + e_t$$
$$E(y_t) = \alpha + \beta t$$

See that mean changes over time.

One can apply OLS to estimate the model parameters.

Stochastic Trend (DSP)—Unit Root Process

$$y_t = y_{t-1} + e_t$$

$$y_t = y_0 + e_1 + e_2 + .. + e_{t-1+} e_t$$

$$E(y_t) = y_0 + E(e_1) + \cdots + E(e_t) = y_0$$

$$\sigma_{y_t}^2 = population\ variance\ of\ (y_0 + e_1 + e_2 + .. + e_{t-1+} e_t)$$

$$= population\ variance\ of\ (e_1 + e_2 + .. + e_{t-1+} e_t)$$

$$= \sigma_e^2 + \sigma_e^2 + \cdots + \sigma_e^2$$

$$= t\sigma_e^2$$

This process is known as random walk.

Stochastic Trend

$$y_t = \mu + y_{t-1} + e_t$$

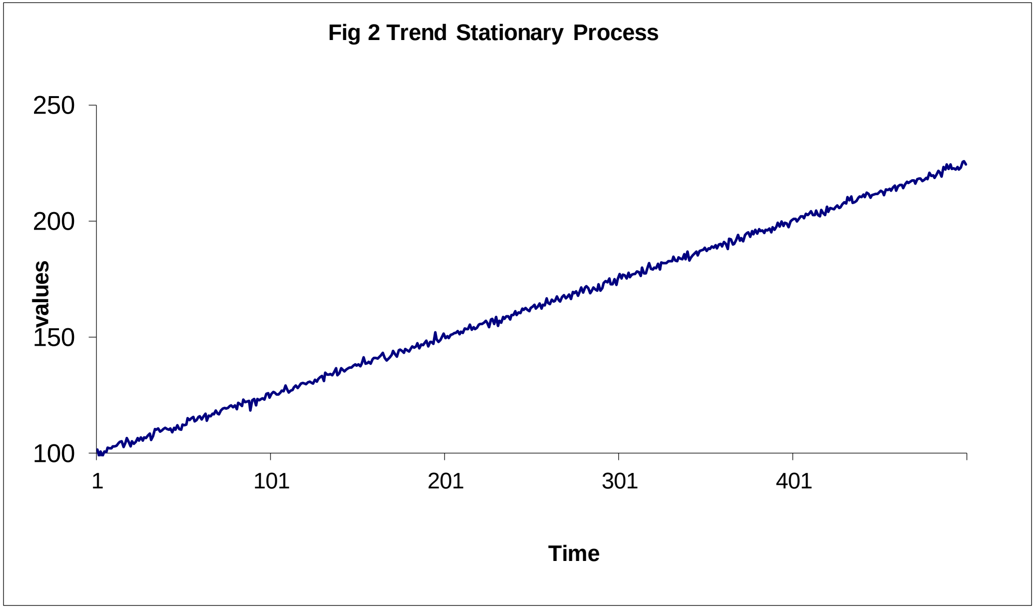
$$y_t = \mu t + y_0 + e_1 + e_2 + \cdots + e_{t-1} + e_t$$

$$E(y_t) = \mu t + y_0$$

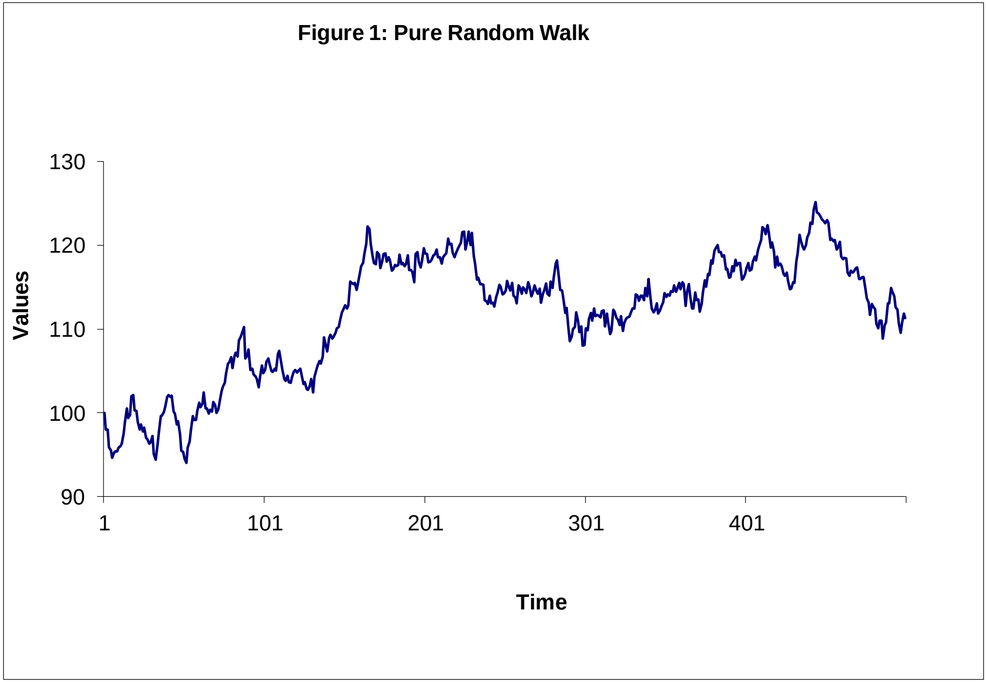
$$\sigma_{y_t}^2 = t\sigma_e^2$$

This process is known as random walk with drift.

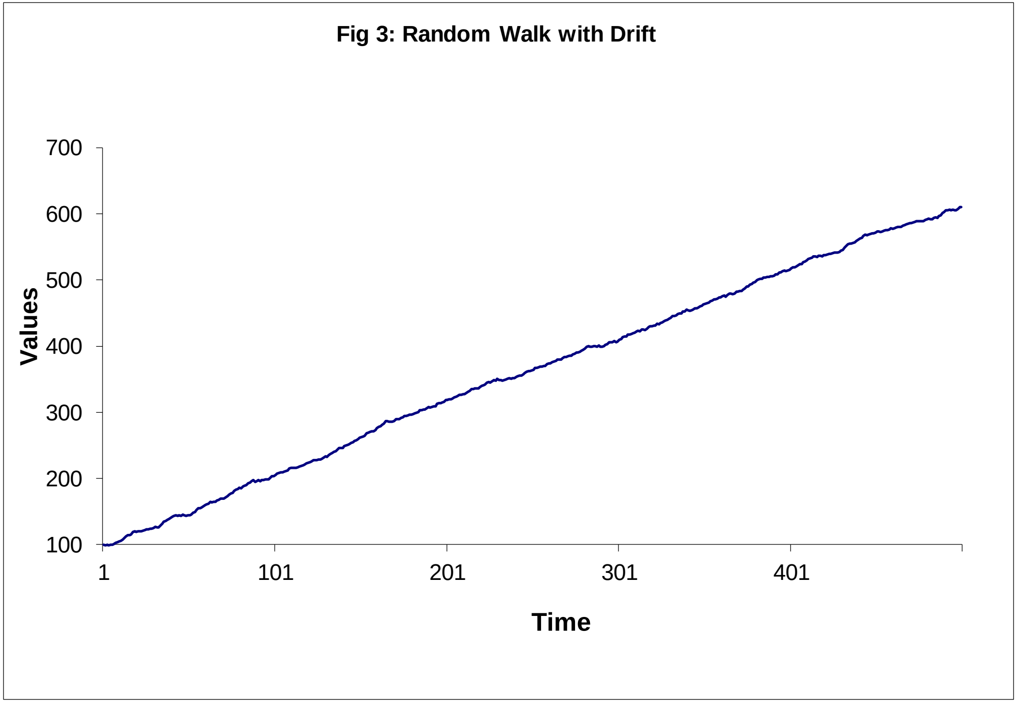
Deterministic Trend



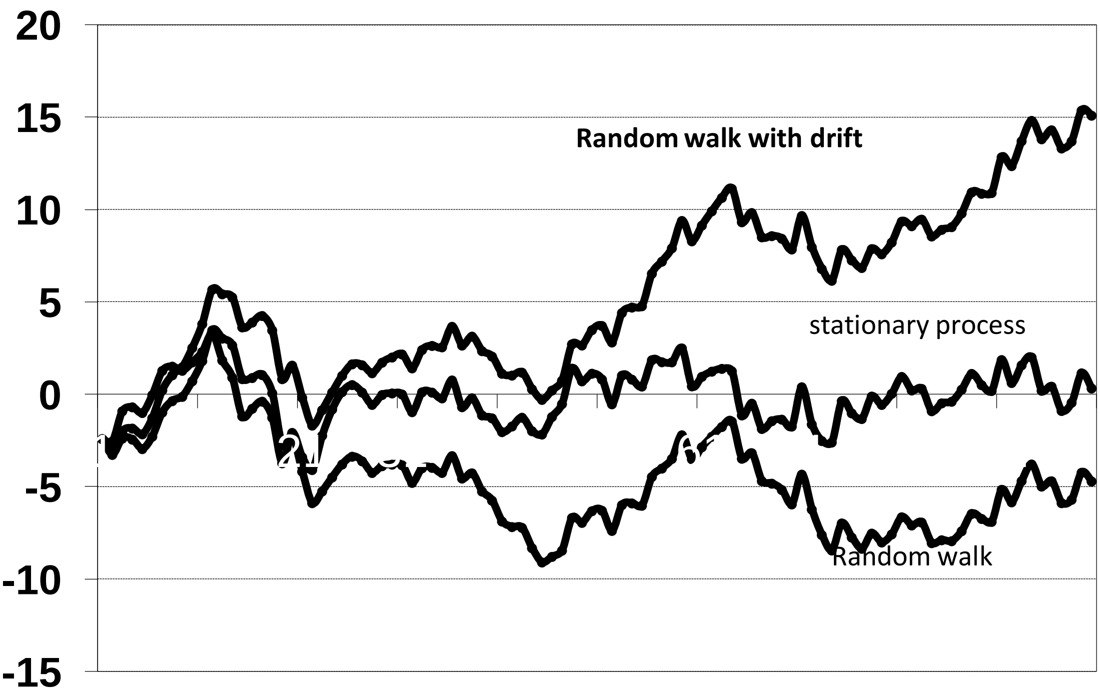
Stochastic Trend



Stochastic Trend



Non Stationary Process



TSP VS DSP

$$\Delta y_t = \alpha + \beta t + \phi y_{t-1} + \sum_{j=1}^{m} \delta_j \Delta y_{t-j} + e_t \, .$$

$$t_{\hat{\rho}} \Longrightarrow \frac{\int_0^1 W_{dm}(r)\,dW(r)}{\left(\int_0^1 [W_{dm}(r)]^2\,dr\right)^{1/2}},$$

Distribution (under the null of unit root) is non-standard, NOT t-distribution or normal

This test is known as Augmented Dickey-Fuller Test (ADF).

Decision

I. $\beta = 0$ and $\rho < 1$ implies series is purely stationary.

II. $\beta \neq 0$ and $\rho < 1$ implies series is purely non-stationary, non-stationary is due to deterministic trend.

III. $\beta = 0$ and $\rho = 1$ implies series is non-stationary, and non-stationary is due to stochastic trend.

Differencing

I. Often non-stationary series can be made stationary through differencing.

Examples:

1) $y_t = y_{t-1} + e_t$ is not stationary, but

$w_t = y_t - y_{t-1} = e_t$ is stationary

2) $y_t = 1.7 y_{t-1} - 0.7 y_{t-2} + e_t$ is not stationary, but

$w_t = y_t - y_{t-1} = 0.7 w_{t-1} + e_t$ is stationary

Differencing

I. Differencing continues until stationarity is achieved.

$$\Delta y_t = y_t - y_{t-1}$$

$$\Delta^2 y_t = \Delta(\Delta y_t) = \Delta(y_t - y_{t-1}) = y_t - 2y_{t-1} + y_{t-2}$$

The differenced series has *n*-1 values after taking the first-difference, *n*-2 values after taking the second difference, and so on.

II. The number of times that the original series must be differenced in order to achieve stationarity is called the *order of integration*, denoted by *d*.

III. In practice, it is almost never necessary to go beyond second difference, because real data generally involve only first or second level non-stationarity.

Differencing

I. Backward shift operator, B

$$By_t = y_{t-1}$$

II. B, operating on y_t, has the effect of shifting the data back one period.

III. Two applications of B on y_t shifts the data back two periods.

$$B(By_{t)} = B^2 y_t = y_{t-2}$$

IV. m applications of B on y_t shifts the data back m periods.

$$B^m y_t = y_{t-m}$$

Differencing

I. The backward shift operator is convenient for describing the process of differencing.

$$\Delta y_t = y_t - y_{t-1} = y_t - By_t = (1-B)y_t$$

$$\Delta^2 y_t = y_t - 2y_{t-1} + y_{t-2} = (1-2B+B^2)y_t = (1-B)^2 y_t$$

II. In general, a *d*th-order difference can be written as

$$B^d y_t = \ (1-B)^d y_t$$

III. The backward shift notation is convenient because the terms can be multiplied together to see the combined effect.

I. If the process is non-stationary then first differences of the series are computed to determine if that operation results in a stationary series.

II. The process is continued until a stationary time series is found.

III. This then determines the value of *d*.

IV. Sometimes, transformations, like log or some variance stabilizing transformations are made before ‘Differencing.

Step Two: Model Identification

Identification

Determination of the values of *p* and *q*.

To determine the value of *p* and *q* we use the graphical properties of the autocorrelation function and the partial autocorrelation function.

Again recall the following.

Properties of the ACF and PACF of MA, AR and ARMA Series

Process	MA(q)	AR(p)	ARMA(p,q)
Auto-correlation function	Cuts off	Infinite. Tails off. Damped Exponentials and/or Cosine waves	Infinite. Tails off. Damped Exponentials and/or Cosine waves after q-p.
Partial Autocorrelation function	Infinite. Tails off. Dominated by damped Exponentials & Cosine waves.	Cuts off	Infinite. Tails off. Dominated by damped Exponentials & Cosine waves after p-q.

Summary: To determine *p* and *q*.

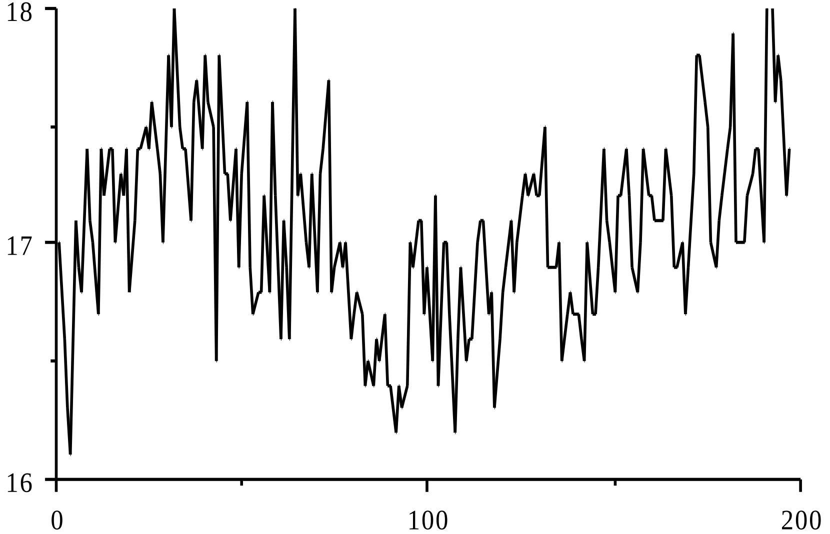
Use the following table.

	MA(q)	AR(p)	ARMA(p,q)
ACF	Cuts after q	Tails off	Tails off
PACF	Tails off	Cuts after p	Tails off

Note: Usually $p + q \leq 4$. There is no harm in over identifying the time series. (Allowing more parameters in the model than necessary. We can always test to determine if the extra parameters are zero.)

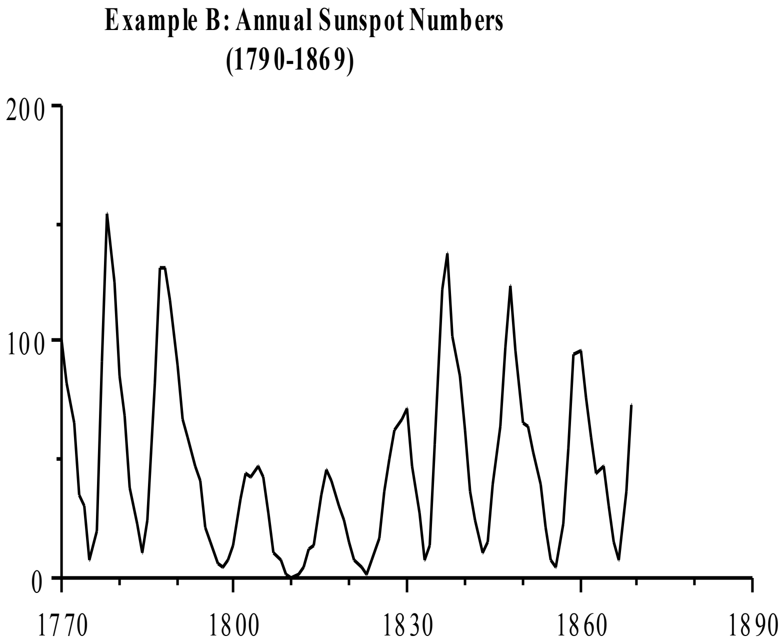
Examples

Example A: "Uncontrolled" Concentration, Two-Hourly Readings: Chemical Process



The data:

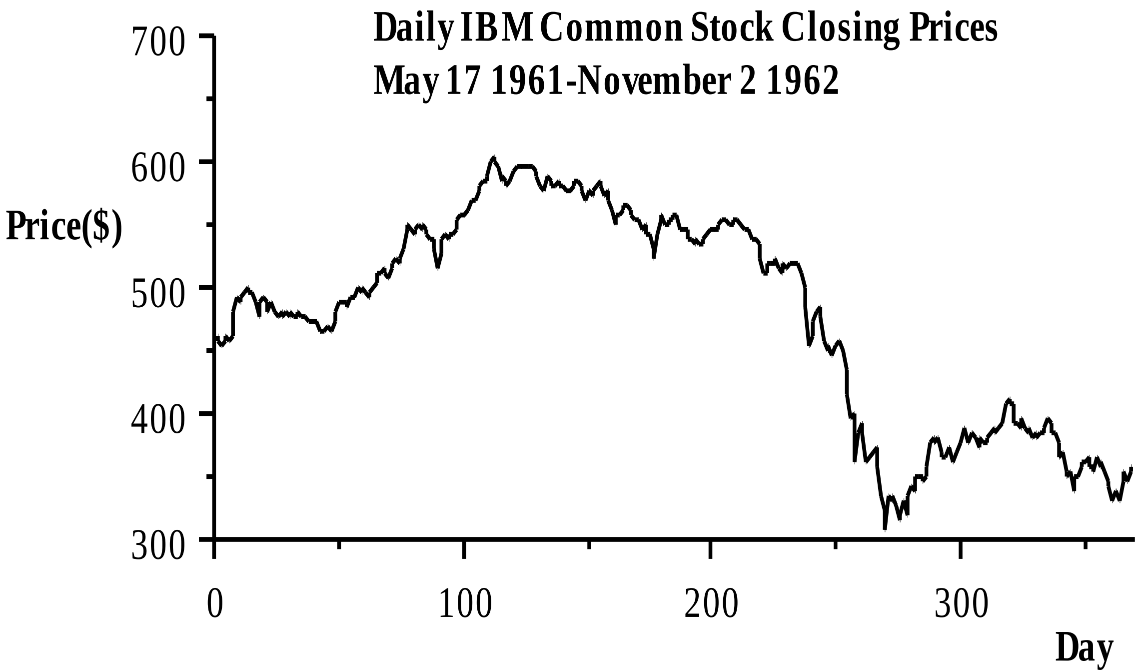
1	17.0	41	17.6	81	16.8	121	16.9	161	17.1
2	16.6	42	17.5	82	16.7	122	17.1	162	17.1
3	16.3	43	16.5	83	16.4	123	16.8	163	17.1
4	16.1	44	17.8	84	16.5	124	17.0	164	17.4
5	17.1	45	17.3	85	16.4	125	17.2	165	17.2
6	16.9	46	17.3	86	16.6	126	17.3	166	16.9
7	16.8	47	17.1	87	16.5	127	17.2	167	16.9
8	17.4	48	17.4	88	16.7	128	17.3	168	17.0
9	17.1	49	16.9	89	16.4	129	17.2	169	16.7
10	17.0	50	17.3	90	16.4	130	17.2	170	16.9
11	16.7	51	17.6	91	16.2	131	17.5	171	17.3
12	17.4	52	16.9	92	16.4	132	16.9	172	17.8
13	17.2	53	16.7	93	16.3	133	16.9	173	17.8
14	17.4	54	16.8	94	16.4	134	16.9	174	17.6
15	17.4	55	16.8	95	17.0	135	17.0	175	17.5
16	17.0	56	17.2	96	16.9	136	16.5	176	17.0
17	17.3	57	16.8	97	17.1	137	16.7	177	16.9
18	17.2	58	17.6	98	17.1	138	16.8	178	17.1
19	17.4	59	17.2	99	16.7	139	16.7	179	17.2
20	16.8	60	16.6	100	16.9	140	16.7	180	17.4
21	17.1	61	17.1	101	16.5	141	16.6	181	17.5
22	17.4	62	16.9	102	17.2	142	16.5	182	17.9
23	17.4	63	16.6	103	16.4	143	17.0	183	17.0
24	17.5	64	18.0	104	17.0	144	16.7	184	17.0
25	17.4	65	17.2	105	17.0	145	16.7	185	17.0
26	17.6	66	17.3	106	16.7	146	16.9	186	17.2
27	17.4	67	17.0	107	16.2	147	17.4	187	17.3
28	17.3	68	16.9	108	16.6	148	17.1	188	17.4
29	17.0	69	17.3	109	16.9	149	17.0	189	17.4
30	17.8	70	16.8	110	16.5	150	16.8	190	17.0
31	17.5	71	17.3	111	16.6	151	17.2	191	18.0
32	18.1	72	17.4	112	16.6	152	17.2	192	18.2
33	17.5	73	17.7	113	17.0	153	17.4	193	17.6
34	17.4	74	16.8	114	17.1	154	17.2	194	17.8
35	17.4	75	16.9	115	17.1	155	16.9	195	17.7
36	17.1	76	17.0	116	16.7	156	16.8	196	17.2
37	17.6	77	16.9	117	16.8	157	17.0	197	17.4
38	17.7	78	17.0	118	16.3	158	17.4		
39	17.4	79	16.6	119	16.6	159	17.2		
40	17.8	80	16.7	120	16.8	160	17.2		


Example B: Annual Sunspot Numbers
(1790-1869)
200
100
0
1770
1800
1830
1860
1890

Example B: Sunspot Numbers: Yearly

The Data:

1770	101	1795	21	1820	16	1845	40
1771	82	1796	16	1821	7	1846	64
1772	66	1797	6	1822	4	1847	98
1773	35	1798	4	1823	2	1848	124
1774	31	1799	7	1824	8	1849	96
1775	7	1800	14	1825	17	1850	66
1776	20	1801	34	1826	36	1851	64
1777	92	1802	45	1827	50	1852	54
1778	154	1803	43	1828	62	1853	39
1779	125	1804	48	1829	67	1854	21
1780	85	1805	42	1830	71	1855	7
1781	68	1806	28	1831	48	1856	4
1782	38	1807	10	1832	28	1857	23
1783	23	1808	8	1833	8	1858	55
1784	10	1809	2	1834	13	1859	94
1785	24	1810	0	1835	57	1860	96
1786	83	1811	1	1836	122	1861	77
1787	132	1812	5	1837	138	1862	59
1788	131	1813	12	1838	103	1863	44
1789	118	1814	14	1839	86	1864	47
1790	90	1815	35	1840	63	1865	30
1791	67	1816	46	1841	37	1866	16
1792	60	1817	41	1842	24	1867	7
1793	47	1818	30	1843	11	1868	37
1794	41	1819	24	1844	15	1869	74



Daily IBM Common Stock Closing Prices
May 17 1961-November 2 1962
Price($)
700
600
500
400
300
0
100
200
300
Day

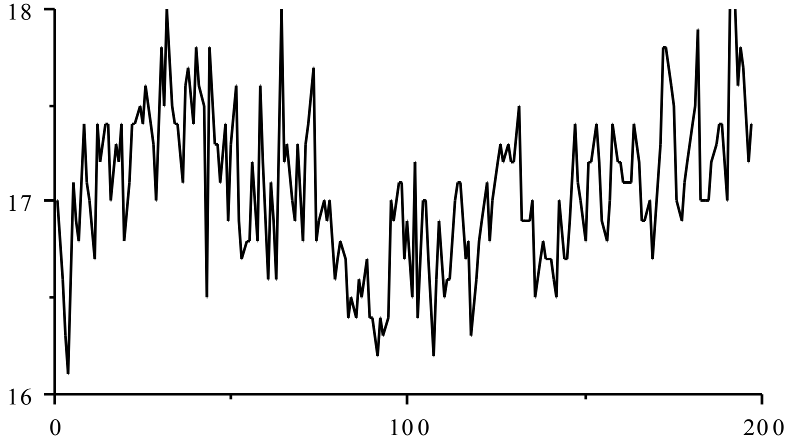
Example C: IBM Common Stock Closing Prices: Daily (May 17 1961- Nov 2 1962)

460	471	527	580	551	523	333	394	330
457	467	540	579	551	516	330	393	340
452	473	542	584	552	511	336	409	339
459	481	538	581	553	518	328	411	331
462	488	541	581	557	517	316	409	345
459	490	541	577	557	520	320	408	352
463	489	547	577	548	519	332	393	346
479	489	553	578	547	519	320	391	352
493	485	559	580	545	519	333	388	357
490	491	557	586	545	518	344	396	
492	492	557	583	539	513	339	387	
498	494	560	581	539	499	350	383	
499	499	571	576	535	485	351	388	
497	498	571	571	537	454	350	382	
496	500	569	575	535	462	345	384	
490	497	575	575	536	473	350	382	
489	494	580	573	537	482	359	383	
478	495	584	577	543	486	375	383	
487	500	585	582	548	475	379	388	
491	504	590	584	546	459	376	395	
487	513	599	579	547	451	382	392	
482	511	603	572	548	453	370	386	
487	514	599	577	549	446	365	383	
482	510	596	571	553	455	367	377	
479	509	585	560	553	452	372	364	
478	515	587	549	552	457	373	369	
479	519	585	556	551	449	363	355	
477	523	581	557	550	450	371	350	
479	519	583	563	553	435	369	353	
475	523	592	564	554	415	376	340	
479	531	592	567	551	398	387	350	
476	547	596	561	551	399	387	349	
478	551	596	559	545	361	376	358	
479	547	595	553	547	383	385	360	
477	541	598	553	547	393	385	360	
476	545	598	553	537	385	380	366	
475	549	595	547	539	360	373	359	
473	545	595	550	538	364	382	356	
474	549	592	544	533	365	377	355	
474	547	588	541	525	370	376	367	
474	543	582	532	513	374	379	357	
465	540	576	525	510	359	386	361	
466	539	578	542	521	335	387	355	
467	532	589	555	521	323	386	348	
471	517	585	558	521	306	389	343	

Read downwards

Chemical Concentration data:

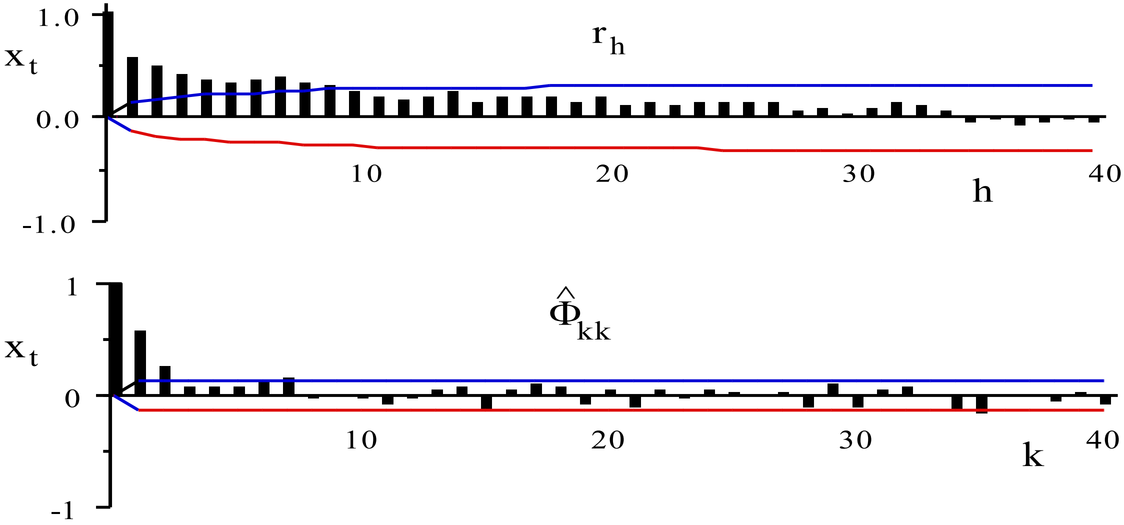
Example A: "Uncontrolled" Concentration, Two-Hourly Readings: Chemical Process



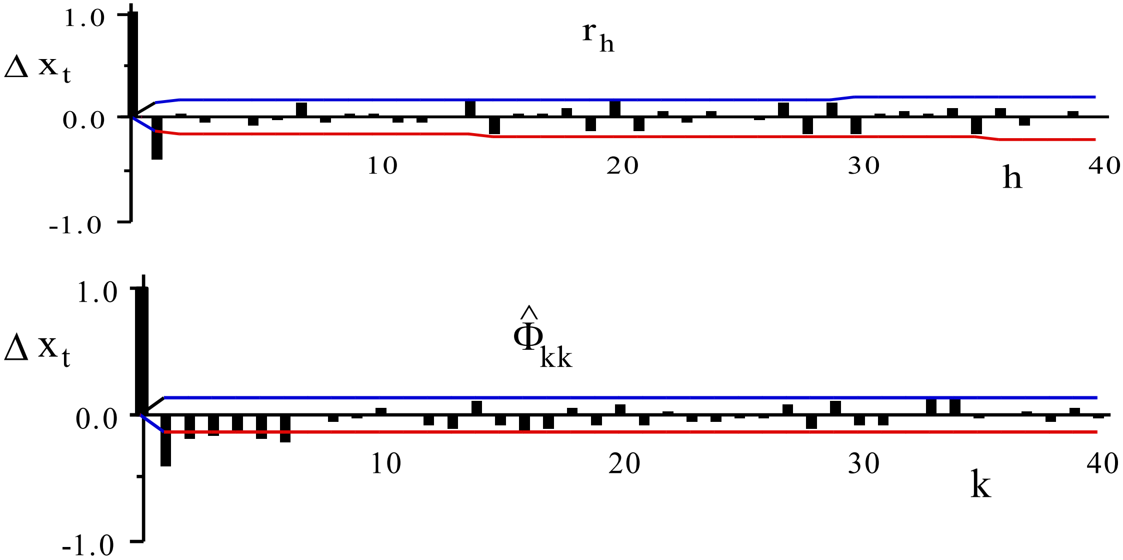
par **Summary Statistics**

d	N	Mean	Std. Dev.
0	197	17.062	0.398
1	196	0.002	0.369
2	195	0.003	0.622

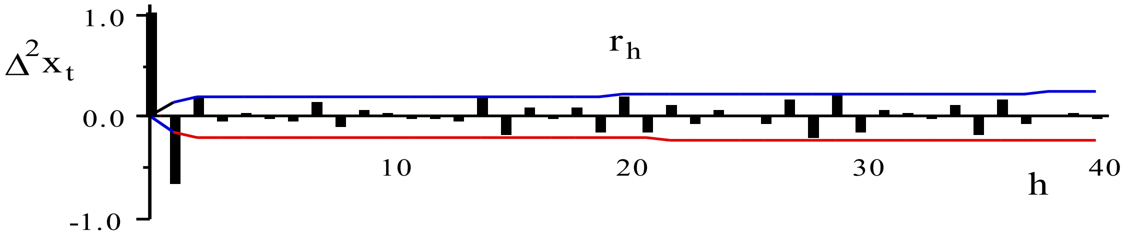
ACF and PACF for $X_t, \Delta X_t \; and \; \Delta^2 X_t$ Chemical concentration DATA

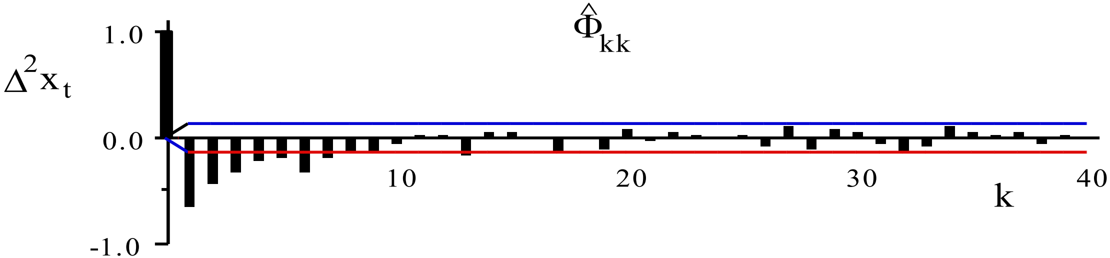


ACF and PACF for $X_t, \Delta X_t$ *and* $\Delta^2 X_t$ Chemical concentration DATA



ACF and PACF for $y_t, \Delta y_t$ *and* $\Delta^2 y_t$ Chemical concentration DATA

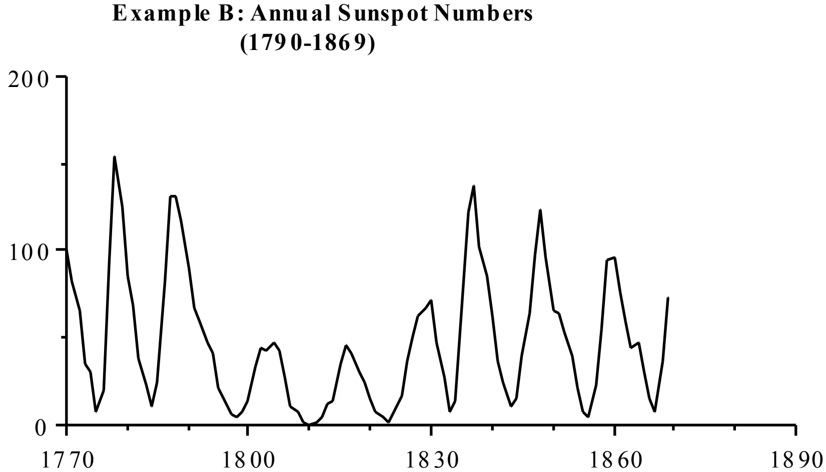




Possible Identifications

1. $d = 0$, $p = 1$, $q = 1$
2. $d = 1$, $p = 0$, $q = 1$

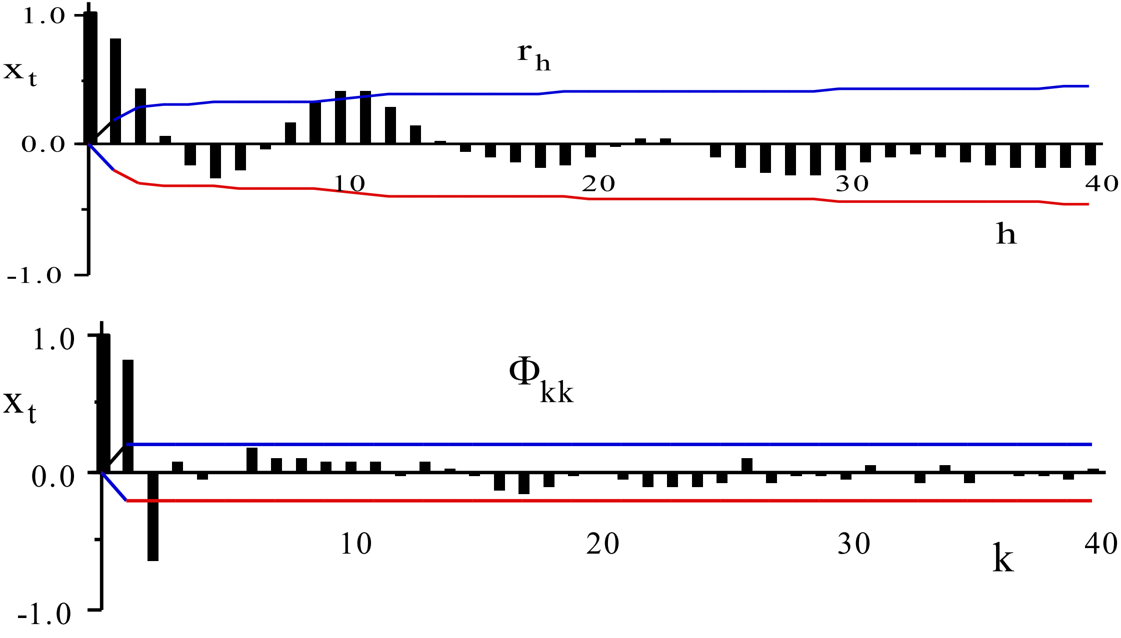
Sunspot Data:



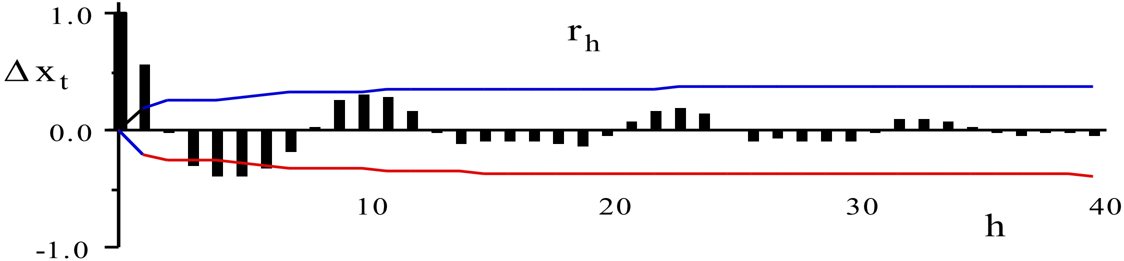
Summary Statistics for the Sunspot Data

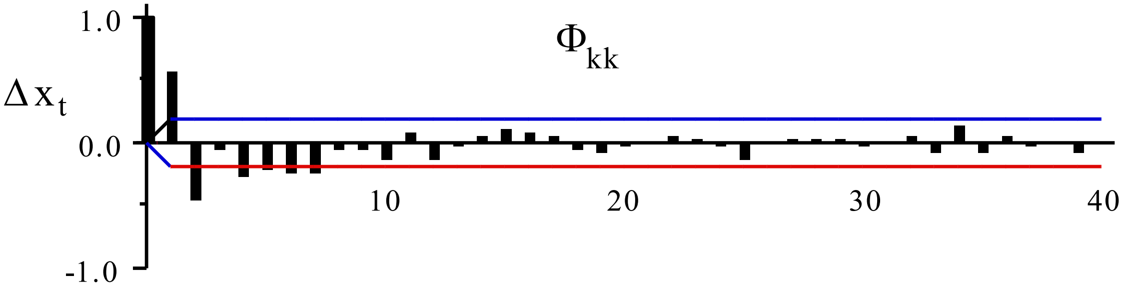
d	N	mean	Std. Dev.
0	100	46.950	37.186
1	99	-0.273	22.440
2	98	0.571	20.198

ACF and PACF for $y_t, \Delta y_t$ *and* $\Delta^2 y_t$ Sunspot Data

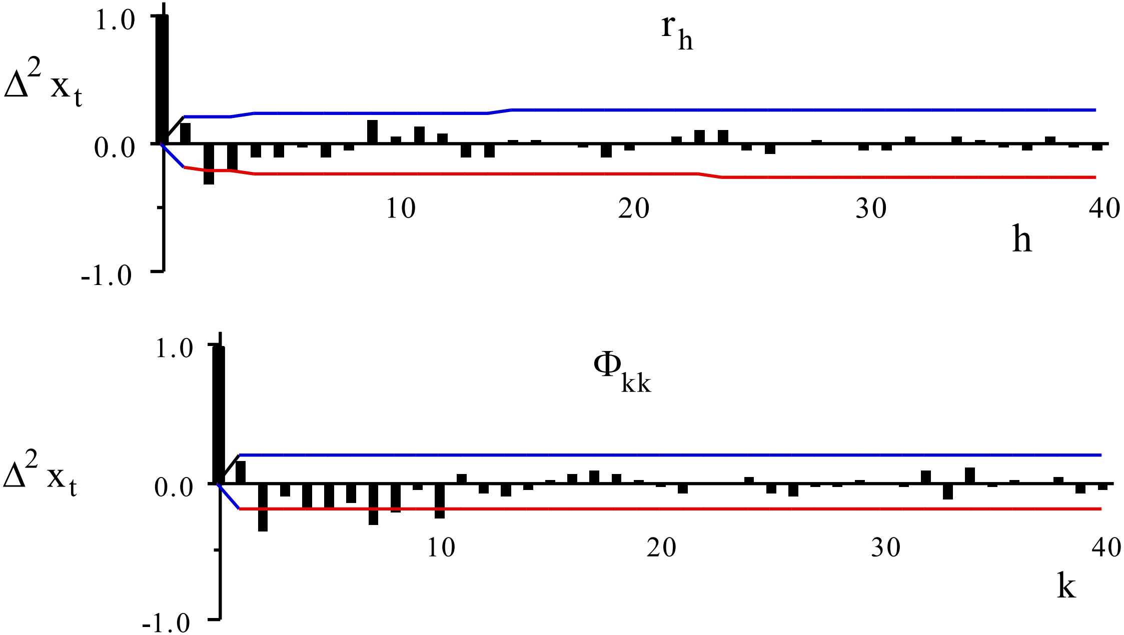


ACF and PACF for $y_t, \Delta y_t$ and $\Delta^2 y_t$ Sunspot Data





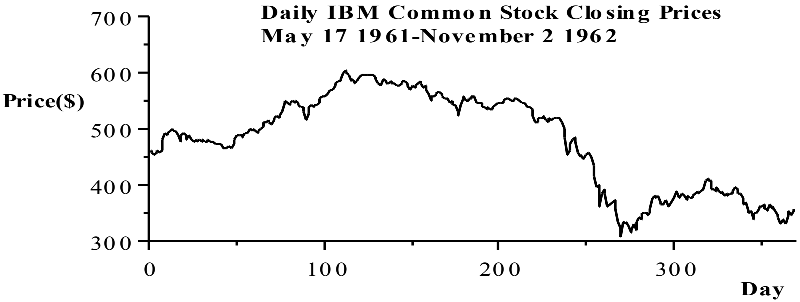
ACF and PACF for $y_t, \Delta y_t$ *and* $\Delta^2 y_t$ Sunspot Data



Possible Identification

1. $d = 0$, $p = 2$, $q = 0$

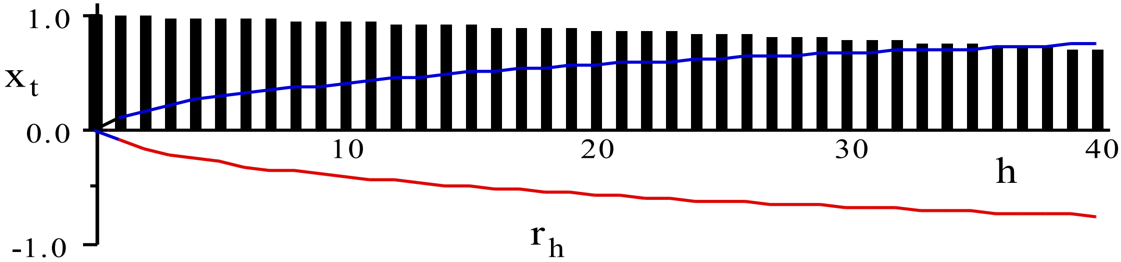
IBM stock data:

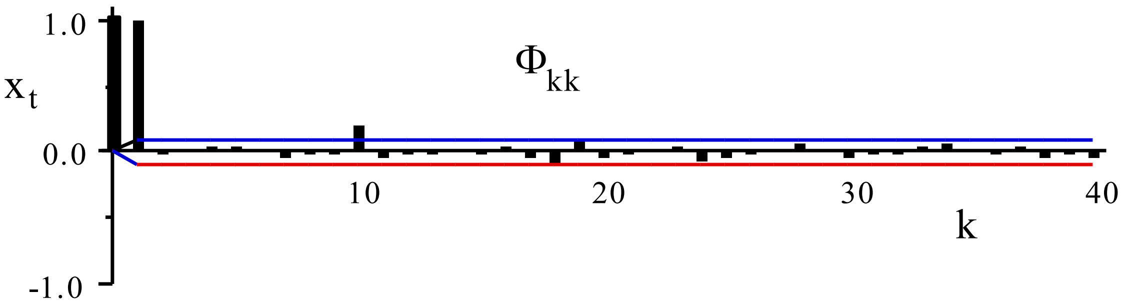


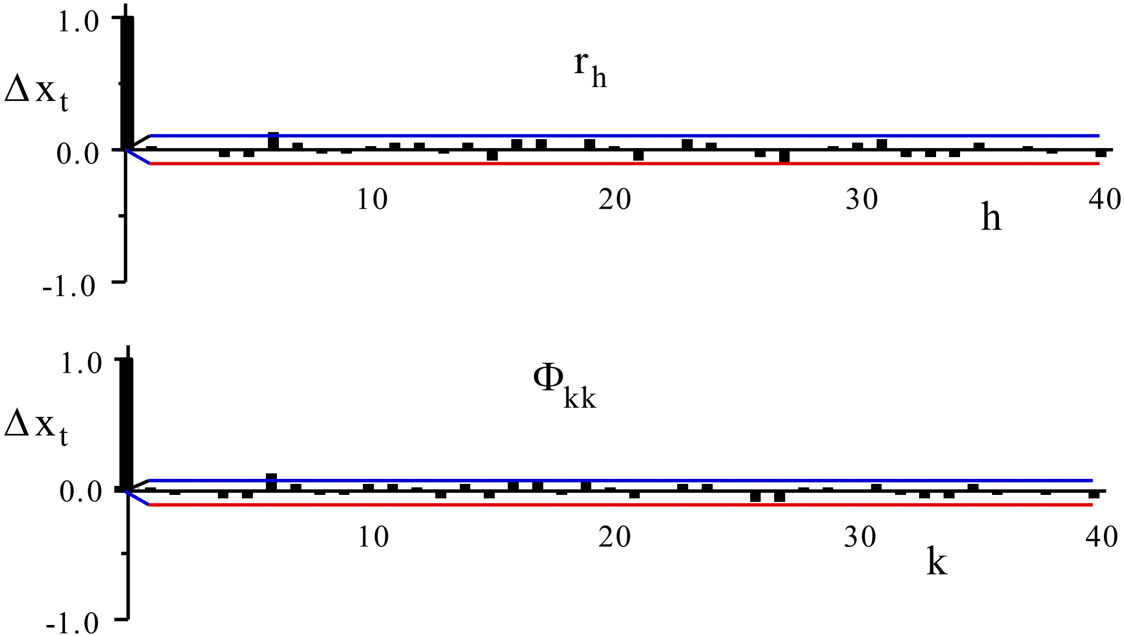
Summary Statistics

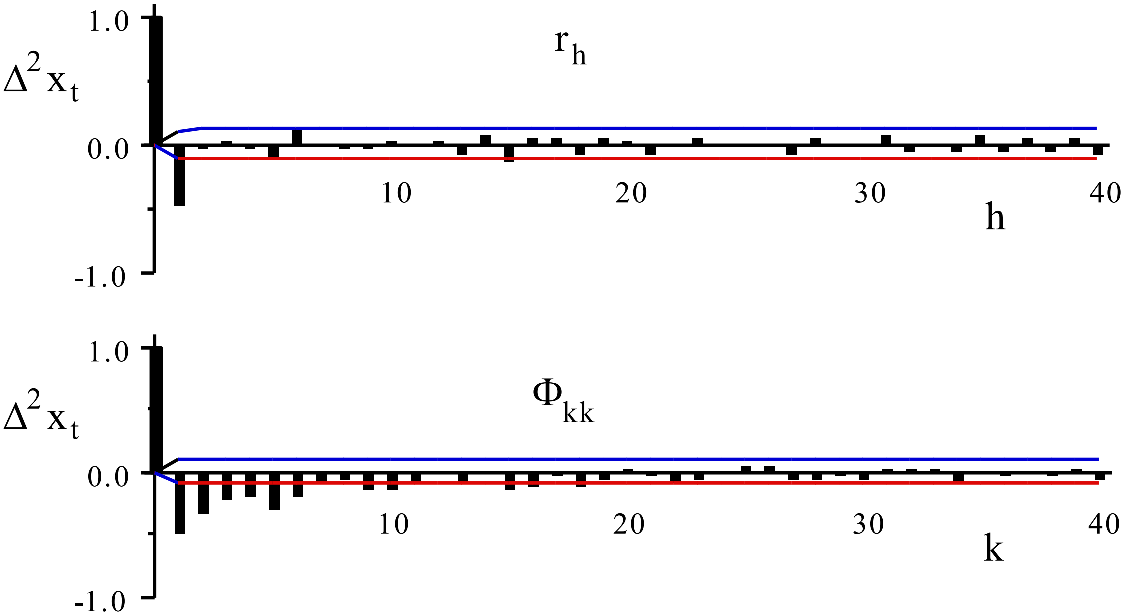
d	N	Mean	Std Dev
0	414	465.899	87.530
1	413	-0.249	7.816
2	412	0.019	10.895

ACF and PACF for $y_t, \Delta y_t$ and $\Delta^2 y_t$ (IBM Stock Price Data)






1.0
0.0
-1.0
Δx_t
r_h
10
20
30
40
h
1.0
0.0
-1.0
Δx_t
Φ_{kk}
10
20
30
40
k


1.0
$\Delta^2 x_t$
r_h
0.0
10
20
30
40
h
-1.0
1.0
$\Delta^2 x_t$
Φ_{kk}
0.0
10
20
30
40
k
-1.0

Possible Identification

1. $d = 1$, $p = 0$, $q = 0$

Step Three: Parameter Estimation

Preliminary Estimation

<u>Using the Method of moments</u>

Equate sample statistics to population parameters

Estimation of parameters of an MA(q) series

The theoretical autocorrelation function in terms the parameters of an MA(q) process is given by.

$$\rho_h = \begin{cases} \dfrac{\alpha_h + \alpha_1\alpha_{h+1} + \cdots + \alpha_{q-h}\alpha_q}{1 + \alpha_1^2 + \alpha_2^2 + \cdots + \alpha_q^2} & 1 \le h \le q \\ 0 & h > q \end{cases}$$

To estimate $\alpha_1, \alpha_1, \ldots, \alpha_1$, we solve the system of equations:

$$r_h = \frac{\hat{\alpha}_h + \hat{\alpha}_1\hat{\alpha}_{h+1} + \cdots + \hat{\alpha}_{q-h}\hat{\alpha}_q}{1 + \hat{\alpha}_1^2 + \hat{\alpha}_2^2 + \cdots + \hat{\alpha}_q^2} \quad 1 \le h < q$$

This set of equations is non-linear and generally very difficult to solve For $q = 1$ the equation becomes:

$$r_1 = \frac{\hat{\alpha}_1}{1 + \hat{\alpha}_1^2}$$

Thus $(1 + \hat{\alpha}_1^2)r_1 = \hat{\alpha}_1 = 0$ Or $r_1\hat{\alpha}_1 - \hat{\alpha}_1 + r_1 = 0$

This equation has the two solutions

$$\hat{\alpha}_1 = \frac{1}{2r_1} \pm \sqrt{\frac{1}{4r_1^2} - 1}$$

One solution will result in the MA(1) time series being invertible

For $q = 2$ the equations become:

$$r_1 = \frac{\hat{\alpha}_1 + \hat{\alpha}_1\hat{\alpha}_2}{1 + \hat{\alpha}_1 + \hat{\alpha}_2}$$

$$r_1 = \frac{\hat{\alpha}_1}{1 + \hat{\alpha}_1 + \hat{\alpha}_2}$$

Estimation of parameters of an ARMA(p, q) series

We use a similar technique.

Namely Obtain an expression for r_h in terms $\beta_1, \beta_2, \ldots, \beta_p$; $\alpha_1, \alpha_2, \ldots, \alpha_q$ of and set up $q + p$ equations for the estimates of $\beta_1, \beta_2, \ldots, \beta_p$; $\alpha_1, \alpha_2, \ldots, \alpha_q$ by replacing ρ_h by r_h.

Estimation of parameters of an ARMA(p,q) series

Example: The ARMA(1,1) process

The expression for ρ_1 and ρ_2 in terms of β_1 and α_1 are: $\rho_1 = \frac{(1+\alpha_1\beta_1)(\alpha_1+\beta_1)}{1+\alpha_1^2+2\alpha_1\beta_1}$

$$\rho_2 = \rho_1\beta_1$$

Further $\sigma^2 = var(e_t) = \frac{1-\beta_1^2}{1+\alpha_1^2+2\alpha_1\beta_1}\sigma_y(0)$

Thus the expression for the estimates of $\beta_1, \alpha_1,$

and σ^2 are : $r_1 = \frac{(1+\hat{\alpha}_1\hat{\beta}_1)(\hat{\alpha}_1+\hat{\beta}_1)}{1+\hat{\alpha}_1^2+2\hat{\alpha}_1\hat{\beta}_1}$

$$r_2 = r_1\hat{\beta}_1$$

And $\sigma^2 = \frac{1-\hat{\beta}_1^2}{1+\hat{\alpha}_1^2+2\hat{\alpha}_1\hat{\beta}_1}\sigma_y(0)$

Hence $\hat{\beta}_1 = \frac{r_2}{r_1}$ and

$$r_1\left(1 + \hat{\alpha}_1^2 + 2\hat{\alpha}_1\hat{\beta}_1\right) = \left(1 + \hat{\alpha}_1\hat{\beta}_1\right)(\hat{\alpha}_1 + \hat{\beta}_1)$$

Or

$$r_1\left(1 + \hat{\alpha}_1^2 + 2\hat{\alpha}_1\frac{r_2}{r_1}\right) = \left(1 + \hat{\alpha}_1\frac{r_2}{r_1}\right)(\hat{\alpha}_1 + \frac{r_2}{r_1})$$

$$\left(r_1 - \frac{r_2}{r_1}\right)\hat{\alpha}_1^2 + \left(2r_2 - 1 - \frac{r_2^2}{r_1^2}\right)\hat{\alpha}_1\left(r_1 + \frac{r_2}{r_1}\right) = 0$$

This is a quadratic equation which can be solved

Example (Chemical Concentration Data)

the time series was identified as either an ARIMA(1,0,1) time series or an ARIMA(1,0,1) series.

If we use the first identification then series y_t is an ARMA(1,1) series.

Identifying the series y_t is an ARMA(1,1) series.

The autocorrelation at lag 1 is $r_1 = 0.570$ and the autocorrelation at lag 2 is $r_2 = 0.495$. Thus the estimate of β_1 is $0.495/0.570 = 0.87$. Also the quadratic equation

$$\left(r_1 - \frac{r_2}{r_1}\right)\hat{\alpha}_1^2 + \left(2r_2 - 1 - \frac{r_2^2}{r_1^2}\right)\hat{\alpha}_1\left(r_1 + \frac{r_2}{r_1}\right) = 0$$

$$0.298\,\hat{\alpha}_1^2 + 0.7642\hat{\alpha}_1 + 0.2984 = 0$$

which has the two solutions -0.48 and -2.08. Again we select as our estimate of a_1 to be the solution -0.48, resulting in an **invertible** estimated series.

Since $\delta = \mu(1 - \beta_1)$ the estimate of δ can be computed as follows:

$$\hat{\delta} = \bar{y}\left(1 - \hat{\beta}_1\right) = 17.062(1 - 0.87) = 2025$$

Thus the identified model in this case is

$$y_t = 0.87\, y_{t-1} + e_t - 0.48\, e_t + 2.25$$

If we use the second identification then series $\Delta y_t = y_t - y_{t-1}$ is an MA(1) series. Thus the estimate of α_1 is:

$$\hat{\alpha}_1 = \frac{1}{2r_1} \pm \sqrt{\frac{1}{4r_1^2} - 1}$$

The value of $r_1 = -0.413$.
Thus the estimate of α_1

$$\hat{\alpha}_1 = \frac{1}{2(-0.413)} \pm \sqrt{\frac{1}{4(-0.413)^2} - 1} = \begin{cases} -1.89 \\ -0.53 \end{cases} \text{ is:}$$

The estimate of $\alpha_1 = -0.53$, corresponds to an invertible time series. This is the solution that we will choose.

The estimate of the parameter μ is the sample mean. Thus the identified model in this case is:

$$\Delta y_t = e_t - 0.53 e_{t-1} + 0.002 \quad \text{Or}$$

$$y_t = y_{t-1} + e_t - 0.53 e_{t-1} + 0.002$$

(An ARIMA(0,1,1) model).

This compares with the other identification:

$$y_t = 0.87\, y_{t-1} + e_t - 0.48\, e_{t-1} + 2.25$$

(An ARIMA(1,0,1) model)

Preliminary Estimation

of the Parameters of an AR(p) Process

The regression coefficients $\beta_1, \beta_2, \dots, \beta_p$ and the auto correlation function ρ_h satisfy the ***Yule-Walker equations***:

$$\rho_1 = \beta_1 1 + \cdots + \beta_p \rho_{p-1}$$

$$\rho_2 = \beta_1 \rho_1 + \cdots + \beta_p \rho_{p-2}$$

$$\dots$$

$$\rho_p = \beta_1 \rho_{p-1} + \cdots + \beta_p 1$$

And $$\sigma(0) = \frac{\sigma^2}{1 - \beta_1 \rho_1 - \cdots - \beta_p \rho_p}$$

The Yule-Walker equations can be used to estimate the regression coefficients $\beta_1, \beta_2, \ldots, \beta_p$ using the sample auto correlation function r_h by replacing ρ_h with r_h.

$$r_1 = \hat{\beta}_1 1 + \cdots + \hat{\beta}_p r_{p-1}$$

$$r_1 = \hat{\beta}_1 1 + \cdots + \hat{\beta}_p r_{p-1}$$

...

$$r_1 = \hat{\beta}_1 1 + \cdots + \hat{\beta}_p r_{p-1}$$

And $$\hat{\sigma}^2 = C_x(0) \times (1 - \hat{\beta}_1 r_1 + \cdots + \hat{\beta}_p r_p)$$

Example

Considering the data in example 1 (Sunspot Data) the time series was identified as an $AR(2)$ time series.
The autocorrelation at lag 1 is $r_1 = 0.807$ and the autocorrelation at lag 2 is $r_2 = 0.429$.
The equations for the estimators of the parameters of this series are

$$1.00\hat{\beta}_1 + 0.807\hat{\beta}_2 = 0.807$$
$$0.807\hat{\beta}_1 + 1.00\hat{\beta}_2 = 0.429$$

which has solution $\hat{\beta}_1 = 1.321$

$$\hat{\beta}_2 = -0.637$$

Since $\delta = \mu(1 - \beta_1 - \beta_2)$ then it can be estimated as follows:

$$\hat{\delta} = \bar{y}\left(1 - \hat{\beta}_1 - \hat{\beta}_2 \right) = 46.590(1 - 1.321 + 0.637) = 14.9$$

Thus the identified model in this case is

$$y_t = 1.321\, y_{t-1} - 0.637\, y_{t-2} + e_t + 14.9$$

Maximum Likelihood Estimation

of the parameters of an

ARMA(p,q) Series

The method of Maximum Likelihood Estimation selects as estimators of a set of parameters $\theta_1, \theta_2, \dots, \theta_k$, the values that maximize

$$L(\theta_1, \theta_2, \dots, \theta_k) = f(y_1, y_2, \dots, y_N; \theta_1, \theta_2, \dots, \theta_k)$$

where $f(y_1, y_2, \dots, y_N; \theta_1, \theta_2, \dots, \theta_k)$ is the joint density function of the observations $y_1, y_2, \dots, y_N$.

$L(\theta_1, \theta_2, \dots, \theta_k)$ is called the ***Likelihood function***.

It is important to note that:

finding the values -$\theta_1, \theta_2, \ldots, \theta_k$- to maximize $L(\theta_1, \theta_2, \ldots, \theta_k)$ is equivalent to finding the values to maximize

$$l(\theta_1, \theta_2, \ldots, \theta_k) = \ln L(\theta_1, \theta_2, \ldots, \theta_k)$$

$l(\theta_1, \theta_2, \ldots, \theta_k)$ is called the log-Likelihood function.

Again let $\{e_t : t \in T\}$ be identically distributed and uncorrelated with mean zero. In addition assume that each is normally distributed.

Consider the time series $\{y_t : t \in T\}$ defined by the equation:

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + \delta + e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2} + \ldots + \alpha_q e_{t-q}$$

Assume that $y_1, y_2, \dots, y_n$ are observations on the time series up to time $t = N$.

To estimate the $p + q + 2$ parameters $\beta_1, \beta_2, \dots, \beta_p\ \alpha_1, \alpha_2, \dots, \alpha_q;\ , \sigma^2$ by the method of Maximum Likelihood estimation we need to find the joint density function of $y_1, y_2, \dots, y_n$

$$f(y_1, y_2, \dots, y_n \,|\beta_1, \beta_2, \dots, \beta_p\,;\ \alpha_1, \alpha_2, \dots, \alpha_q, \delta, \sigma^2)$$
$$= \ f(\boldsymbol{y}|\ \boldsymbol{\beta}, \boldsymbol{\alpha}, \delta\ , \sigma^2).$$

We know that $e_1, e_2, \dots, e_n$ are independent normal with mean zero and variance σ^2.

Thus the joint density function of $e_1, e_2, \dots, e_n$ is $g(e_1, e_2, \dots, e_n\,;\,\sigma^2) = g(\boldsymbol{u}\,;\,\sigma^2)$ is given by.

$$g(e_1, e_2, \dots, e_n\,;\,\sigma^2) = g(\boldsymbol{u}\,;\,\sigma^2)$$
$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{N} e_t^2\right\}$$

It is difficult to determine the exact density function of $y_1, y_2, \dots, y_n$ from this information however if we assume that p starting values on the y process $\boldsymbol{y}^* = (y_{1-p}, y_{2-p}, \dots, y_0)$ and q starting values on the e $-$process $e^* = (e_{1-p}, e_{2-p}, \dots, e_0)$ have been observed then the conditional distribution of $\boldsymbol{y} = (y_{1-p}, y_{2-p}, \dots, y_0)$ given $\boldsymbol{y}^* = (y_{1-p}, y_{2-p}, \dots, y_0)$ and $e^* = (e_{1-p}, e_{2-p}, \dots, e_0)$ can easily be determined.

The system of equations :

$$y_1 = \beta_1 y_0 + \beta_2 y_{-1} + \ldots + \beta_p y_{1-p} + \delta + e_1 + \alpha_1 e_0 + \alpha_2 e_{-1} + \ldots + \alpha_q e_{1-q}$$

$$y_2 = \beta_1 y_1 + \beta_2 y_0 + \ldots + \beta_p y_{2-p} + \delta + e_2 + \alpha_1 e_1 + \alpha_2 e_0 + \ldots + \alpha_q e_{2-q}$$

$$\ldots$$

$$y_N = \beta_1 y_{N-1} + \beta_2 y_{N-2} + \ldots + \beta_p y_{N-p} + \delta + e_N + \alpha_1 e_{N-1} + \alpha_2 e_{N-2} + \ldots + \alpha_q e_{N-q}$$

can be solved for:

$$e_1 = e_1\,(\boldsymbol{y}, \boldsymbol{y}^*, \boldsymbol{u}^*;\ \boldsymbol{\beta}, \boldsymbol{\alpha}, \delta)$$

$$e_2 = e_2\,(\boldsymbol{y}, \boldsymbol{y}^*, \boldsymbol{u}^*;\ \boldsymbol{\beta}, \boldsymbol{\alpha}, \delta)$$

$$\cdots$$

$$e_N = e_N\,(\boldsymbol{y}, \boldsymbol{y}^*, \boldsymbol{u}^*;\ \boldsymbol{\beta}, \boldsymbol{\alpha}, \delta)$$

(The jacobian of the transformation is 1)

Then the joint density of **x** given $\boldsymbol{y}^*$ and $\boldsymbol{e}^*$ is given by:

$$f(y|y^*e^*, \beta, \alpha, \delta, \sigma^2)$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{N} e_t^2(y^*e^*, \beta, \alpha, \delta)\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n exp\left\{-\frac{1}{2\sigma^2} s^*(\boldsymbol{\beta}, \boldsymbol{\alpha}, \delta)\right\}$$

$$where\ \ s^*(\beta, \alpha, \delta) = \sum_{t=1}^{N} e_t^2(\boldsymbol{y}^*\boldsymbol{e}^*, \boldsymbol{\beta}, \boldsymbol{\alpha}, \delta)$$

Let:

$$L_{y|y^*,e^*}(\beta,\alpha,\delta,\sigma^2) =$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n exp\left\{-\frac{1}{2\sigma^2}\sum_{t=1}^{N} e_t^2(y^*e^*,\beta,\alpha,\delta)\right\}$$

$$= \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n exp\left\{-\frac{1}{2\sigma^2}s^*(\boldsymbol{\beta},\boldsymbol{\alpha},\delta)\right\}$$

= "conditional likelihood function"

$$where\ \ s^*(\beta,\alpha,\delta) = \sum_{t=1}^{N} e_t^2(\boldsymbol{y}^*\boldsymbol{e}^*,\boldsymbol{\beta},\boldsymbol{\alpha},\delta)$$

“conditional log likelihood function” =

$$l_{y|y^*,e^*}(\beta,\alpha,\delta,\sigma^2) = ln\ \ L_{y|y^*,e^*}(\beta,\alpha,\delta,\sigma^2)$$

$$= \frac{n}{2} - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{N} e_t^2(y^* e^*, \beta, \alpha, \delta)$$

$$= \frac{n}{2} - \ln(2\pi) - \frac{n}{2}\ln(2\sigma^2) - \frac{1}{2\sigma^2} s^*(\boldsymbol{\beta}, \boldsymbol{\alpha}, \delta)$$

The values that maximize

$l_{y|y^*,e^*}(\beta,\alpha,\delta,\sigma^2)$ and $L_{y|y^*,e^*}(\beta,\alpha,\delta,\sigma^2)$

$$are\ the\ values$$

$$\hat{\beta},\hat{\alpha},\hat{\delta}$$

That minimize

$$s^*(\boldsymbol{\beta},\boldsymbol{\alpha},\delta)=\sum_{t=1}^{N}e_t^2(y^*e^*,\beta,\alpha,\delta)$$

With $\hat{\sigma}^2=\frac{1}{N}\sum_{t=1}^{N}e_t^2(y^*e^*,\beta,\alpha,\delta)=\frac{1}{N}s^*(\boldsymbol{\beta},\boldsymbol{\alpha},\delta)$

Comment:

The minimization of:

$$s^*(\boldsymbol{\beta}, \boldsymbol{\alpha}, \delta) = \sum_{t=1}^{N} e_t^2(y^* e^*, \beta, \alpha, \delta)$$

Requires a iterative numerical minimization procedure to find:

$$\hat{\beta}, \hat{\alpha}, \hat{\delta}$$

- Steepest descent
- Simulated annealing
- etc

Comment:

The computation of:

$$s^*(\boldsymbol{\beta}, \boldsymbol{\alpha}, \delta) = \sum_{t=1}^{N} e_t^2(y^* e^*, \beta, \alpha, \delta)$$

for specific values of

$$\boldsymbol{\beta}, \boldsymbol{\alpha}, \delta$$

can be achieved by using the forecast equations

$$e_t = y_t - \hat{y}_{t-1}(1)$$

Comment:

The minimization of :

$$s^*(\boldsymbol{\beta}, \boldsymbol{\alpha}, \delta) = \sum_{t=1}^{N} e_t^2(y^* e^*, \beta, \alpha, \delta)$$

assumes we know the value of starting values of the time series $\{y_t | \ t \ \in T\}$ and $\{e_t | \ t \ \in T\}$

Namely y^* and e^*.

Approaches:

1. Use estimated values

$$\bar{y}\ for\ the\ components\ of\ y^*$$

$$0 for\ the\ components\ of\ e^*$$

2. Use forecasting and backcasting equations to estimate the values:

Backcasting:

If the time series $\{y_t | t \in T\}$ satisfies the equation:

$$y_t = \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + \delta + e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2} + \ldots + \alpha_q e_{t-q}$$

It can also be shown to satisfy the equation:

$$y_t = \beta_1 y_{t+1} + \beta_2 y_{t+2} + \ldots + \beta_p y_{t+p} + \delta + e_t + \alpha_1 e_{t+1} + \alpha_2 e_{t+2} + \ldots + \alpha_q e_{t+q}$$

Both equations result in a time series with the same mean, variance and autocorrelation function:

In the same way that the first equation can be used to forecast into the future the second equation can be used to backcast into the past:

Approaches to handling starting values of the series $\{y_t | t \in T\}$ and $\{e_t | t \in T\}$

1. Initially start with the values:

$$\bar{y} \; for\ the\ components\ of\ y^*$$
$$0 \; for\ the\ components\ of\ e^*$$

2. Estimate the parameters of the model using Maximum Likelihood estimation and the conditional Likelihood function.
3. Use the estimated parameters to backcast the components of **x***. The backcasted components of **u*** will still be zero.

4. Repeat steps 2 and 3 until the estimates stablize.

This algorithm is an application of the ***E-M* algorithm**

This general algorithm is frequently used when there are missing values.

The *E* stands for Expectation (using a model to estimate the missing values)

The *M* stands for Maximum Likelihood Estimation, the process used to estimate the parameters of the model.

Some Examples using:

- Minitab
- Statistica
- S-Plus
- SAS

Step Four: Diagnostic Checking

Diagnostic Checking

- Often it is not straightforward to determine a single model that most adequately represents the data generating process, and it is not uncommon to estimate several models at the initial stage. The model that is finally chosen is the one considered best based on a set of diagnostic checking criteria. These criteria include

  (1) t-tests for coefficient significance

  (2) residual analysis

  (3) model selection criteria

Diagnostic checking (t-tests)

- Note that for any AR model, the estimated mean value and the drift term are related through the formula

$$\mu = \frac{\delta}{1 - \varphi_1 - \varphi_2 - \cdots - \varphi_p}$$

Portmanteau test

- Box and Peirce proposed a statistic which tests the magnitudes of the residual autocorrelations as a group
- Their test was to compare Q below with the Chi-Square with $K - p - q$ $d.f.$ when fitting an ARMA(p, q) model

$$Q = N \sum_{k=1}^{K} r_k^2$$

Portmanteau test

- Box & Ljung discovered that the test was not good unless n was very large
- Instead use modified Box-Pierce or Ljung-Box-Pierce statistic—reject model if Q^* is too large

$$Q^* = N(N+2)\sum_{k=1}^{K}\frac{r_k^2}{N-k}$$

Residual Analysis

- If an ARMA(p,q) model is an adequate representation of the data generating process, then the residuals should be uncorrelated.
- Portmanteau test statistic:

$$Q^*(k) = (N-d)(N-d+2)\sum_{k=1}^{K}\frac{r_k^2(e)}{N-d-l} \sim \chi^2_{k-p-q}$$

Model Selection Criteria

- Akaike Information Criterion (AIC)

$$AIC = -2\, ln(L) + 2k$$

- Schwartz Bayesian Criterion (SBC)

$$SBC = -2\, ln(L) + k\, ln(n)$$

where $L =$ likelihood function

$k =$ number of parameters to be estimated,

$n =$ number of observations.

- Ideally, the AIC and SBC should be as small as possible

AIC

- The Akaike Information Criterion is a function of the maximum likelihood plus twice the number of parameters
- The number of parameters in the formula penalizes models with too many parameters

Parsimony

- Once principal generally accepted is that models should be parsimonious—having as few parameters as possible
- Note that any ARMA model can be represented as a pure AR or pure MA model, but the number of parameters may be infinite

Parsimony

- AR models are easier to fit so there is a temptation to fit a less parsimonious AR model when a mixed ARMA model is appropriate
- Ledolter & Abraham (1981) *Technometrics* show that fitting unnecessary extra parameters, or an AR model when a MA model is appropriate, results in loss of forecast accuracy

REASONS FOR USING A PARSIMONIOUS MODEL

- Fewer numerical problems in estimation.
- Easier to understand the model.
- With fewer parameters, forecasts less sensitive to deviations between parameters and estimates.
- Model may applied more generally to similar processes.
- Rapid real-time computations for control or other action.
- Having a parsimonious model is less important if the realization is large.

REASONS NEEDING A LONG REALIZATION

- Estimate correlation structure (i.e., the ACF and PACF) functions and get accurate standard errors.
- Estimate seasonal pattern (need at least 4 or 5 seasonal periods).
- Approximate prediction intervals assume that parameters are known (good approximation if realization is large).
- Fewer estimation problems (likelihood function better behaved).
- Possible to check forecasts by withholding recent data .
- Can check model stability by dividing data and analyzing both sides.

Step Four: Forecasting

FORECASTING

$$y_t = \varphi y_{t-1} + e_t$$

$$\Downarrow$$

$$\hat{\varphi}(estimates\ of\ \varphi)$$

$$\Downarrow$$

$$\hat{y}_t = \varphi \hat{y}_{t-1}$$

$$\Downarrow$$

$$\hat{y}_{t+1}\ (forecast)$$

FORECASTING FROM AN ARMA MODEL

THE MINIMUM MEAN SQUARED ERROR FORECASTS

Observed time series, $y_1, y_2, \dots, y_n$. n: the forecast origin

Observed sample

Y_1 Y_2 .... Y_n | Y_{n+1}? Y_{n+2}?

$$\hat{y}_n(1) \rightarrow the forecast\ value\ of\ y_{n+1}$$
$$\hat{y}_n(2) \rightarrow the forecast\ value\ of\ y_{n+2}$$
$$\hat{y}_n(l) \rightarrow the forecast\ value\ of\ y_{n+l}$$
$$\rightarrow l\ step\ ahed\ forecast\ of\ y_{n+1}$$
$$\rightarrow minimum\ MSE\ forecast\ \ of\ y_{n+1}$$

FORECASTING FROM AN ARMA MODEL

$$\hat{y}_n(l) = E(y_{n+l}|y_n, y_{n-1}, \dots, y_1)$$
$$= the\ conditional\ expectation\ of\ y_{n+l}$$
$$given\ the\ observed\ sample$$

FORECASTING FROM AN ARMA MODEL

- The stationary ARMA model for y_t *is*
  $\varphi_p(B)y_t = \theta_q(B)e_t$ *Or*
  $$y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \ldots + \beta_p y_{t-p} + \delta + e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2} + \ldots + \alpha_q e_{t-q}$$
- Assume that we have data $y_1, y_2, \ldots, y_n$ *and we want to* forecast y_{n+l} *(i.e., l steps ahead from forecast origin n).* Then the actual value is

$$y_{n+l} = \beta_0 + \beta_1 y_{n+l-1} + \beta_2 y_{n+l-2} + \ldots + \beta_p y_{n+l-p} + \delta + e_t + \alpha_1 e_{n+l-1} + \alpha_2 e_{n+l-2} + \ldots + \alpha_q e_{n+l-q}$$

FORECASTING FROM AN ARMA MODEL

- Considering the Random Shock Form of the series

$$y_{n+l} = \beta_0 + \psi(B)e_{t+l} = \beta_0 + \frac{\theta_q(B)}{\varphi_p(B)}e_{t+l},$$

$$= \beta_0 + \ e_{n+l} \ + \psi_1 e_{n+l-1} \\ + \ \psi_2 e_{n+l-2} \ + \ldots + \psi_n e_n + \cdots$$

FORECASTING FROM AN ARMA MODEL

- Taking the expectation of Y_{n+l}, we have

$$\begin{aligned}\hat{y}_n(l) &= E(y_{n+l}|y_n, y_{n-1}, \dots, y_1)\\ &= \psi_l e_n + \psi_l e_{n-1} + \cdots\end{aligned}$$

Where

$$E(e_{n+j}|y_n, y_{n-1}, \dots, y_1) = \begin{cases} 0 & if\ j > 0 \\ e_{n+j} & if\ j \le 0 \end{cases}$$

FORECASTING FROM AN ARMA MODEL

- The forecast error:

$$\begin{aligned}\varepsilon_n(l) &= y_{n+l} - \hat{y}_n(l) \\ &= e_{n+l} + \psi_1 e_{n+l-1} + \cdots + \psi_{l-1} e_{n+1} \\ &= \sum_{i=0}^{l-1} \psi_i e_{n+l-i}\end{aligned}$$

The expectation of the forecast error: $E\left(\varepsilon_n(l)\right) = 0$

So, the forecast in unbiased.

The variance of the forecast error:

$$var\left(\varepsilon_n(l)\right) = var\left(\sum_{i=1}^{l-1} \psi_i e_{n+l-i}\right) = \sigma_e^2 \sum_{i=1}^{l-1} \psi_i^2$$

FORECASTING FROM AN ARMA MODEL

One step-ahead ($l = 1$)

$$y_{n+1} = \beta_0 + e_{n+1} + \psi_1 e_n + \psi_2 e_{n-1} + \cdots$$

$$\hat{y}_n(1) = \beta_0 + \psi_1 e_n + \psi_2 e_{n-1} + \cdots$$

$$\varepsilon_n(1) = y_{n+1} - \hat{y}_n(1) = e_{n+1}$$

$$var\left(\varepsilon_n(1)\right) = \sigma_e^2$$

FORECASTING FROM AN ARMA MODEL

Two step-ahead ($l = 2$)

$$y_{n+2} = \beta_0 + e_{n+2} + \psi_1 e_{n+1} + \psi_2 e_n + \cdots$$

$$\hat{y}_n(2) = \beta_0 + \psi_2 e_n + \cdots$$

$$\varepsilon_n(2) = y_{n+2} - \hat{y}_n(2) = e_{n+2} + \psi_1 e_{n+1}$$

$$var\big(\varepsilon_n(2)\big) = \sigma_e^2(1 + \psi_1^2)$$

FORECASTING FROM AN ARMA MODEL

Note that,

$$\lim_{l \to \infty} \hat{y}_n(l) - \mu = 0$$

$$\lim_{l \to \infty} var\big(\varepsilon_n(l)\big) = \gamma_0 < \infty$$

That's why ARMA (or ARIMA) forecasting is useful only for short-term forecasting.

PREDICTION INTERVAL FOR Y_{n+l}

A 95% prediction interval for Y_{n+l} $(l\ \ steps\ ahead)$ is

$$\hat{y}_n(l) \pm 1.96\sqrt{var\big(\varepsilon_n(l)\big)}$$

$$\hat{y}_n(l) \pm 1.96\sqrt{var\big(\varepsilon_n(l)\big)}$$

For one step-ahead this simplifies to

$$\hat{y}_n(1) \pm 1.96\sigma_e$$

For one step-ahead this simplifies to

$$\hat{y}_n(2) \pm 1.96\sigma_e\sqrt{(1 + \psi_1^2)}$$

UPDATING THE FORECASTS

- Let's say we have n observations at time $t = n$ and find a good model for this series and obtain the forecast for y_{n+1}, y_{n+2} and so on. At $t = n + 1$, we observe the value of y_{n+2} Now, we want to update our forecasts using the original value of y_{n+l} and the forecasted value of it.

UPDATING THE FORECASTS

The forecast error is

$$\varepsilon_n(l) = y_{n+l} - \hat{y}_n(l) = \sum_{i=1}^{l-1} \psi_i e_{n+l-i}$$

We can also write this as

$$\begin{aligned}
\varepsilon_n(l) &= y_{n+l} - \hat{y}_n(l+1) \\
&= \sum_{i=1}^{l} \psi_i e_{n-1+l+1-i} \\
&= \sum_{i=1}^{l-1} \psi_i e_{n+l-i} + \psi_l e_n
\end{aligned}$$

UPDATING THE FORECASTS

$$y_{n+l} - \hat{y}_{n-1}(l+1) = y_{n+l} - \hat{y}_n(l) + \psi_l e_n$$

$$\hat{y}_n(l) = \hat{y}_{n-1}(l+1) + \psi_l e_n$$

$$\hat{y}_n(l) = \hat{y}_{n-1}(l+1) + \psi_l \{y_n - \hat{y}_{n-1}(1)\}$$

$$\hat{y}_{n+1}(l) = \hat{y}_n(l+1) + \psi_l \{y_{n+1} - \hat{y}_n(1)\}$$

$$n = 100$$

$$\hat{y}_{101}(l) = \hat{y}_{100}(2) + \psi_1 \{y_{101} - \hat{y}_{100}(1)\}$$

Forecast of an AR(1) process

$$y_t = \phi y_{t-1} + e_t \longrightarrow y_n(l) = ?$$

$$l = 1 \quad y_{n+1} = \phi y_n + e_{n+1}$$

$$E(y_{n+1} \mid \mathrm{I}_n) = \phi y_n$$

$$l = 2 \quad y_{n+2} = \phi y_{n+1} + e_{n+2}$$

$$E(y_{n+2} \mid \mathrm{I}_n) = \phi^2 y_n$$

$$\text{for any } l \quad y_n(l) = \phi^l y_n$$

The forecast decays geometrically as l increases

Forecast of an AR(p) process

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots\ldots..\phi_p y_{t-p} + e_t \rightarrow$$

$$\hat{y}_n(l) = E(y_{n+l} \mid y_n, y_{n-1}, \ldots..) = ?$$

$$l = 1 \quad y_{n+1} = \phi_1 y_n + \phi_2 y_{n-1} + \ldots\ldots..\phi_p y_{n-p+1} + e_{n+1}$$

$$\hat{y}_n(1) = E(y_{n+1} \mid \mathrm{I}_n) = \phi_1 y_n + \phi_2 y_{n-1} + \ldots\ldots..\phi_p y_{n-p+1}$$

$$l = 2 \quad y_{n+2} = \phi_1 y_{n+1} + \phi_2 y_n + \ldots\ldots..\phi_p y_{n-p+2} + e_{n+2}$$

$$\hat{y}_n(2) = E(y_{n+2} \mid \mathrm{I}_n) = \phi_1 \hat{y}_n(1) + \phi_2 y_n + \ldots\ldots..\phi_p y_{n-p+2}$$

$$\text{for any } l \quad \hat{y}_n(l) = \phi_1 \hat{y}_n(l-1) + \phi_2 \hat{y}_n(l-2) + \ldots\ldots..\phi_p \hat{y}_n(l-p)$$

You need to calculate the previous forecasts *l-1*,*l-2*,….

Forecast of a MA(1)

$$y_t = e_t + \theta e_{t-1}$$

$$\hat{y}_n(l) = E(y_{n+l} \mid I_n) = ?$$

$l = 1$ $$y_{n+1} = e_{n+1} + \theta e_n$$

$$\hat{y}_n(1) = E(y_{n+1}) = \theta e_n$$

$$e_n = (1+\theta L)^{-1} y_n$$

$l = 2$ $$y_{n+2} = e_{n+2} + \theta e_{n+1}$$

$$\hat{y}_n(2) = 0$$

$l > 1$ $\hat{y}_n(l) = 0$

That is the mean of the process

Forecast of a MA(q)

$$y_t = (1 + \theta_1 B + \theta_2 B^2 + \ldots\ldots\theta_q B^q) e_t$$

$$\hat{y}_n(l) = E(y_{n+l} \mid I_n) = \begin{cases} (\theta_l + \theta_{l+1} B + \theta_{l+2} B^2 + \ldots\theta_q B^{q-l}) e_n & l \le q \\ 0 & l > q \end{cases}$$

where $\quad e_n = \dfrac{1}{1 + \theta_1 B + \ldots\theta_q B^q} y_n$

Forecast of an ARMA(1,1)

$$(1-\phi B)y_t = (1+\theta B)e_t$$

$$y_{n+l} = \phi y_{n+l-1} + e_{n+l} + \theta e_{n+l-1}$$

$$l = 1 \quad \hat{y}_n(1) = \phi y_n + \theta e_n \qquad \text{where} \quad e_n = \frac{1-\phi B}{1+\theta B} y_n$$

$$l = 2 \quad \hat{y}_n(2) = \phi \hat{y}_n(1) = \phi(\phi y_n + \theta e_n)$$

$$l \geq 2 \quad \hat{y}_n(l) = \phi \hat{y}_n(l-1) = \phi^2 \hat{y}_n(l-2) = \ldots \phi^{l-1} \hat{y}_n(1)$$

Forecast of an ARMA(p,q)

$$\Phi_p(B)y_t = \Theta_q(B)e_t$$

$$y_{n+l} = \phi_1 y_{n+l-1} + ...\phi_p y_{n+l-p} + e_{n+l} + \theta_1 e_{n+l-1} + ... + \theta_q e_{n+l-q}$$

$$\hat{y}_n(l) = \phi_1 \hat{y}_n(l-1) + ... + \phi_p \hat{y}_n(l-p) + \hat{e}_n(l) + \theta_1 \hat{e}_n(l-1) + ... + \theta_q \hat{e}_n(l-q)$$

$$\hat{y}_n(j) = E(y_{n+j} \mid y_n, y_{n-1}.....) \qquad j \geq 1$$

$$\hat{y}_n(j) = y_{n+j} \qquad j \leq 0$$

$$\hat{e}_n(j) = 0 \qquad j \geq 1$$

$$\hat{e}_n(j) = e_{n+j} = e_{n+j} - \hat{y}_{n+j-1}(1) \qquad j \leq 0$$

Example: ARMA(2,2)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2}$$

$$l = 1 \quad y_{n+1} = \phi_1 y_n + \phi_2 y_{n-1} + e_{n+!} + \theta_1 e_n + \theta_2 e_{n-1}$$

$$\hat{y}_n(1) = E(y_{n+1} \mid I_n) = \phi_1 y_n + \phi_2 y_{n-1} + \theta_1 \hat{e}_n + \theta_2 \hat{e}_{n-1}$$

where

$$\hat{y}_n(0) = y_n$$

$$\hat{y}_n(-1) = y_{n-1}$$

$$\hat{e}_n = \frac{\Phi_2(B)}{\Theta_2(B)} y_n$$

$$\hat{e}_{n-1} = y_{n-1} - \hat{y}_{n-2}(1)$$

Updating forecasts

Suppose you have information up to time *n*, such that

$$\hat{y}_n(1), \hat{y}_n(2), \ldots\ldots \hat{y}_n(l)$$

When new information comes, can we update the previous forecasts?

1. $$\varepsilon_n(l) = y_{n+l} - \hat{y}_n(l) = \sum_{j=0}^{l-1} \psi_j e_{n+l-j}$$

2. $$\varepsilon_{n-1}(l+1) = \sum_{j=0}^{l+1-1} \psi_j e_{n-1+l+1-j} = \sum_{j=0}^{l} \psi_j e_{n+l-j}$$

$$\varepsilon_{n-1}(l+1) = \sum_{j=0}^{l-1} \psi_j e_{n+l-j} + \psi_l e_n = \varepsilon_n(l) + \psi_l e_n$$

3. $$y_{n+l} - \hat{y}_{n-1}(l+1) = y_{n+l} - \hat{y}_n(l) + \psi_l e_n$$

$$\hat{y}_n(l) = \hat{y}_{n-1}(l+1) + \psi_l e_n$$

$$\hat{y}_{n+1}(l) = y_n(l+1) + \psi_l e_{n+1}$$

Problems

P1: For each of the following models:

$$(i)\ (1-\phi_1 B)y_t = e_t$$

$$(ii)\ (1-\phi_1 B-\phi_2 B^2)y_t = \mathrm{e}_t$$

$$(iii)\ (1-\phi_1 B)(1-B)y_t = e_t$$

(a)Find the *l*-step ahead forecast of Z_{n+l}

(b) Find the variance of the *l*-step ahead forecast error for *l*=1, 2, and 3.

P2: Consider the IMA(1,1) model : $(1-B)y_t = (1-\theta B)e_t$

(a)Write down the forecast equation that generates the forecasts

(b) Find the 95% forecast limits produced by this model

(c)Express the forecast as a weighted average of previous observations.

FORECASTS OF THE TRANSFORMED SERIES

- If you use variance stabilizing transformation, after the forecasting, you have to convert the forecasts for the original series.
- If you use log-transformation, you have to consider the fact that

$$E\left[y_{n+\ell} \middle| y_1, \cdots, y_n\right] \geq \exp\left\{E\left[\ln\left(y_{n+\ell}\right) \middle| \ln\left(y_1\right), \cdots, \ln\left(y_n\right)\right]\right\}$$

FORECASTS OF THE TRANSFORMED SERIES

- If X has a normal distribution with mean μ and variance σ^2,

$$E[\exp(y)] = \exp\left(\mu + \frac{\sigma^2}{2}\right).$$

- Hence, the minimum mean square error forecast for the original series is given by

$$\exp\left[\hat{y}_n(\ell) + \frac{1}{2}Var(\varepsilon_n(\ell))\right] \text{ where } \mathrm{y}_{\mathrm{n}+\ell} = \ln(\mathrm{y}_{\mathrm{n}+\ell})$$

$$\mu = E(y_{n+\ell} | y_1, \cdots, y_n) \qquad \sigma^2 = Var(y_{n+\ell} | y_1, \cdots, y_n)$$

MEASURING THE FORECAST ACCURACY

Technique	Abbrev	Measures
Mean Squared Error	MSE	The average of squared errors over the sample period
Mean Error	ME	The average dollar amount or percentage points by which forecasts differ from outcomes
Mean Percentage Error	MPE	The average of percentage errors by which forecasts differ from outcomes
Mean Absolute Error	MAE	The average of absolute dollar amount or percentage points by which a forecast differs from an outcome
Mean Absolute Percentage Error	MAPE	The average of absolute percentage amount by which forecasts differ from outcomes

MEASURING THE FORECAST ACCURACY

1. Mean Squared Error

The formula used to calculate the mean squared error is:

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(a_t - f_t)^2$$

2. Mean Percentage Error

The formula used to calculate the mean percentage error is:

$$MPE = \frac{1}{n}\sum_{t=1}^{n}\frac{(a_t - f_t)}{a_t} \times 100$$

3. Mean Absolute Error

The formula used to calculate the mean absolute error is:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}\left|(a_t - f_t)\right|$$

MEASURING THE FORECAST ACCURACY

4. Mean Absolute Percentage Error

The formula used to calculate the mean absolute percentage error is:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\frac{|(a_t - f_t)|}{a_t} \times 100$$

5. Theil's U Statistic

The formulas used to calculate Theil's U statistics are:

$$U1 = \frac{\sqrt{\sum_{t=1}^{n}(a_t - f_t)^2}}{\sqrt{\sum_{t=1}^{n}a_t^2} + \sqrt{\sum_{t=1}^{n}f_t^2}}, \qquad U2 = \sqrt{\frac{\sum_{t=1}^{n-1}\left(\frac{f_{t+1} - a_{t+1}}{a_t}\right)^2}{\sum_{t=1}^{n-1}\left(\frac{a_{t+1} - a_t}{a_t}\right)^2}}$$

To interpret the U statistics the general guide is:

- U1 is bound between 0 and 1, with values closer to 0 indicating greater forecasting accuracy.
- if U2 = 1, there is no difference between a naïve forecast and the technique used
- if U2 < 1 the technique is better than a naïve forecast; and
- if U2 > 1 the technique is no better than a naïve forecast.

References

1. Analysis of time series, J. D. Hamilton
2. Introduction to time series analysis, Brockwell and Davis.
3. Time series analysis, Brockwell and Davis.
4. Time Series Analysis: Forecasting and Control, Box and Jenkins, G.C. Reinsel

Softwares

1. SAS
2. SPSS
3. STATA
4. Eviews
5. TSP
6. R

Thank You