Albert-Ludwigs University Freiburg
Department of Empirical Research and Econometrics

Lecture Notes on

Time Series Analysis

by
Assoc. Prof. Dr. Sevtap Kestel

6-10 June, 2011

Contents

Course Outline

The study of the sequence of data points measured at successive times enables us to often either to understand the underlying theory of the data points (where did they come from what generated them), or to make forecasts (predictions). **Time series *prediction*** is the use of a model to predict future events based on known past events: to predict future data points before they are measured.

The objective of the course is to provide students to learn time series modelling in theory and practice. The course will start with reviewing the fundamental concepts in regression analysis. Autocorrelation function, Linear Stationary models: General linear process, Autoregressive, Moving averages, ARMA processes, Non-stationary models: Autoregressive Integrated Moving Average and Integrated Moving Average processes, Forecasting: Minimum Mean Square Error Forecast, updating forecasts, Stochastic Model building: Model identification, Model estimation (maximum likelihood estimation), Model diagnostic checking, Seasonal models, Spectral analysis and filtering, Vector Autoregressive Models, and cointegration will be covered.

Course Schedule

Lecture	Topics	Chapters to be covered
Day 1	**Introduction, classical Time series**	**1-4**
Day 2	**Stochastic Time series**	**4**
Day 3	**Forecasting, Integrated models, unit root**	**5-6**
Day 4	**Multiplicative seasonal models, ARCH (m)**	**7-8**
Day 5	**Vector Autoregressive Models and Cointegration**	**8-9**

<u>Teaching Methods:</u>

Presentation of teaching materials include introduction of the theoretical base with illustrative examples and exercises solved in the class. Tutorials enhance the application of the theory and the interpretation of the results. Application to data set by using Software Eviews will be presented during PC Pool sessions. Students are **strongly** recommended to participate lectures and tutorials.

<u>Grading:</u>

Final Exam (75%)

A comprehensive 90 min. final exam will be given. The test will be in-class and closed-book exam. If you miss the final exam, you will be treated according to the regulations of the University. Students are required to pass at least 50% of the final examination.

Assignments (25%)

A group of two students will submit one set of assignment given during the teaching period. The deadline of all assignments are due to the Final Examination (August 4, 1011, 14:00 h)

Chapter 1
Review of Statistics

Definition: A numerically valued function X of w with domain Ω,

$w \in \Omega \; : \; w \rightarrow X(w)$ is called a **random variable** (r.v).

Proposition: If X and Y are random variables, then any mathematical combination of those, such as, $X+Y$, $X-Y$, XY, $\frac{X}{Y}\,(Y \neq 0)$ *and* $aX+bY$ are also random variables.

	Countable Case	*Density Case*
Range	$V_n \quad n=1,2,\cdots$	$-\infty < u < \infty$
probability	P_n	$f(u)du = dF(u)$
$P(a \le x \le b)$	$\sum_{a \le V_n \le b} P_n$	$\int_a^b f(u)du$
$P(X \le x) = F(x)$	$\sum_{V_n \le x} P_n$	$\int_{-\infty}^{x} f(u)du$
$E(X)$	$\sum P_n V_n$	$\int_{-\infty}^{\infty} uf(u)du$
condition	$\sum_n P_n \lvert V_n \rvert < \infty$	$\int_{-\infty}^{\infty} \lvert u \rvert f(u)du < \infty$
$E(Q(X))$	$\sum \varphi(x) f(x)$	$\int_{-\infty}^{\infty} \varphi(u) f(u)du$
Variance	$\sum (X-\mu)^2 f(x)$	$\int_{-\infty}^{\infty} (X-\mu)^2 f(u)du$
Skewness	$\frac{\sum (X-\mu)^3 f(x)}{\sigma^3}$	$\frac{\int_{-\infty}^{\infty} (X-\mu)^3 f(u)du}{\sigma^3}$
Kurtosis	$\frac{\sum (X-\mu)^4 f(x)}{\sigma^4}$	$\frac{\int_{-\infty}^{\infty} (X-\mu)^4 f(u)du}{\sigma^4}$

<u>Independent Random Variables:</u>

If two variables X and Y are independent, then the following hold

$$F(x,y) = F(x).F(y)$$
$$f(x,y) = f(x).f(y)$$
$$E(xy) = E(x).E(y)$$
$$Cov(x,y) = 0; \quad \rho_{xy} = 0$$

The Normal Distribution

Normal (Gaussian) distribution has the following properties:

- symmetrical
- easy to estimate probabilities
- inferential statistics

The Normal Distribution is the most commonly used distribution in statistics. It has a unique mode. The observations lie around the mean, median and mode value symmetrically. The density, mean, variance of the distribution are:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty < x < \infty$$

$$E(x) = \mu \quad \sigma^2 = Var(x)$$

Standard Normal Distribution is used to define the variables which are originally normally distributed and standardized around the mean.

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Theorem 1: If X is $N(\mu, \sigma^2), \sigma^2 > 0$,then the r.v $W = \frac{X-\mu}{\sigma}$ is $N(0,1)$.

Theorem 2: If the r.v X is $N(\mu, \sigma^2), \sigma^2 > 0$, then the r.v $V = \frac{(X-\mu)^2}{\sigma^2}$ is a Chi-square distribution with degrees of freedom 1, $\chi^2(1)$.

Hypothesis Testing

A statistical hypothesis is a statement about the distribution of X. If the hypothesis completely specifies the distribution, then for a simple hypothesis, we define

$$\begin{aligned} H_0 &: \theta = \theta_0 \\ H_A &: \theta = \theta_1 \end{aligned} .$$

otherwise, a composite hypothesis is defined as

$$H_0 : \theta \geq \theta_0$$
$$H_A : \theta < \theta_0$$

Critical region is the subset of sample space that corresponds to rejecting the null hypothesis.

Type I error refers Rejecting a true H_0; Type II error and Type II error refers to Failing to reject a false H_0 (Accepting a false H_0). The probabilities of those errors define:

$P(Type\ I\ error) = \alpha$ and $P(Type\ II\ error) = \beta$

For simple H_0, the probability, of rejecting a true $H_0(\alpha)$ is referred to as the **significance** level, denoted by α.

For composite H_0, the size of the test (critical region) is the maximum probability of rejecting H_0 when it is true.

Standard approach specified select some acceptable level of α determine the value of critical value. Among all critical regions of size α, we choose the one with smallest β. The power function $K(\theta)$ is the probability of rejecting H_0 when the true value of the parameter is θ.

Example: Suppose the random variable denotes **w**aiting times in a bank queue. The aim is to determine if the mean waiting time is equal to 7 minute. or not is tested based on a random sample of 315 observations as follows:

One-Sample t-Test

	Test Value = 7					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Wait time in min.	-25.490	314	.000	-1.6748	-1.8041	-1.5455

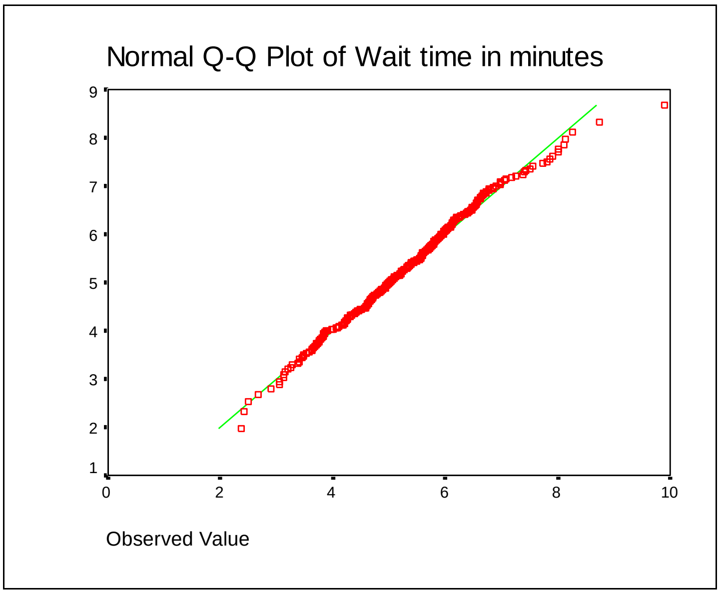
Tests of Normality

1. Parametric approach : Goodness of fit tests
2. Non-parametric approach: Kolmogrov-Smirnov tests

3. Graphical approach: P-P, Q-Q plots are the graphs of percentiles of ordered observations. They should form a linear pattern.

- Jarque-Bera test statistics: depends on the values of sample skewness and kurtosis
- For a normally distributed random variable (r.v.) Skewness S(y)≈0; Kurtosis K(y) ≈3

Example: Q-Q Plot of the example on the waiting times in the queue

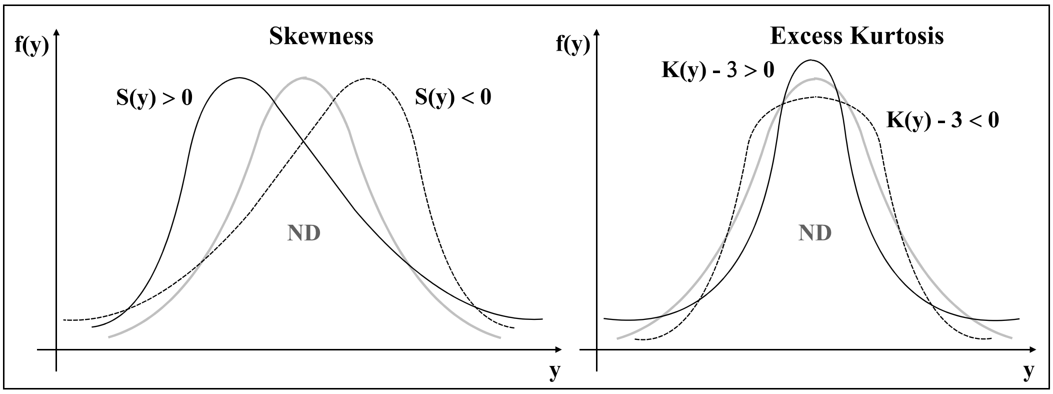


One-Sample Kolmogorov-Smirnov Test Variable: Waiting time

		Wait time in mi utes
N		315
Normal Parameters	Mean	5.3252
	Std. Deviation	1.16614
Most Extreme Differences	Absolute	.039
	Positive	.039
	Negative	-.028
Kolmogorov-Smirnov Z		.686
Asymp. Sig. (2-tailed)		.734

a Test distribution is Normal.
b Calculated from data.

Jargue-Bera Test:



The test statistics is:

$$JB = \left(\frac{T}{6}\right) \cdot \left(\hat{s}^2 + \frac{1}{4}(\hat{k} - 3)^2\right) \sim \chi^2(2)$$

T : the number of the observations in the sample, $\hat{S}$: sample skewness; $\hat{k}$: sample kurtosis,

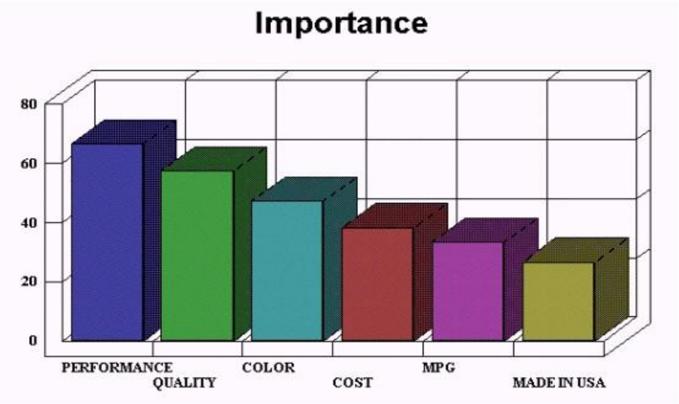
H_0 : Data set is normally distributed versus alternative that data follow a different distribution. Some significance values are: 1% ≈ 9,21 and 5% ≈ 5,99

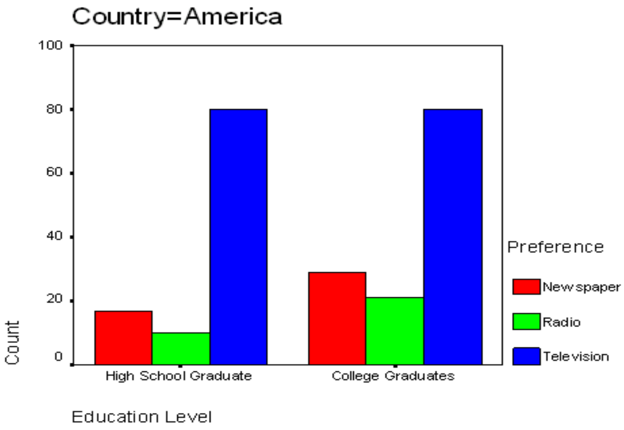
Use of Graphics in Analyses

Illustrstive representation of the observations give researcher an important information on analysing the behavior and the pattern of the data set. There are many graphical methods which depend on the type of the data as well as the choice of selection.

Examples: The following is the illustrations for some graphical representations which are commonly used for quantitative and qualitative data sets.

Bar Chart and clustered bar chart

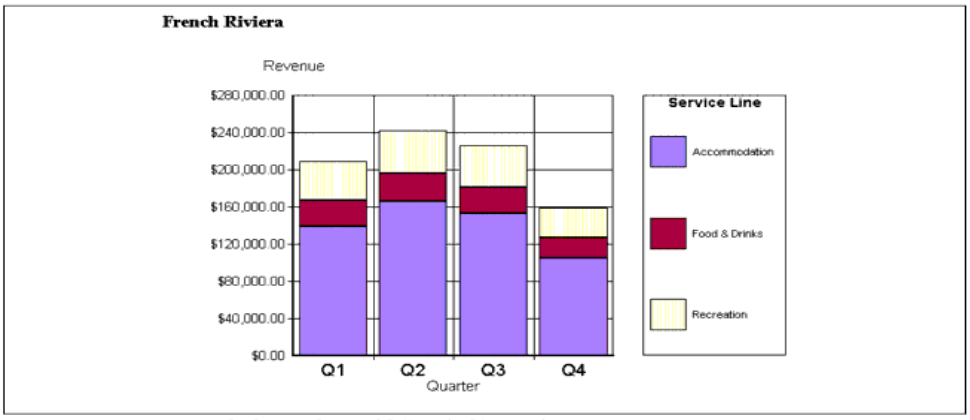




Source: http://www.strategicinit.com/_borders/car_imp_bar_chart.jpg

Stacked bar chart

The variables in this graph are the revenues which result from the different service lines (Accommodation, Food& Drinks and Recreation) on the French Rivera. The categories are the quarters Q1, Q2, Q3 and Q4.

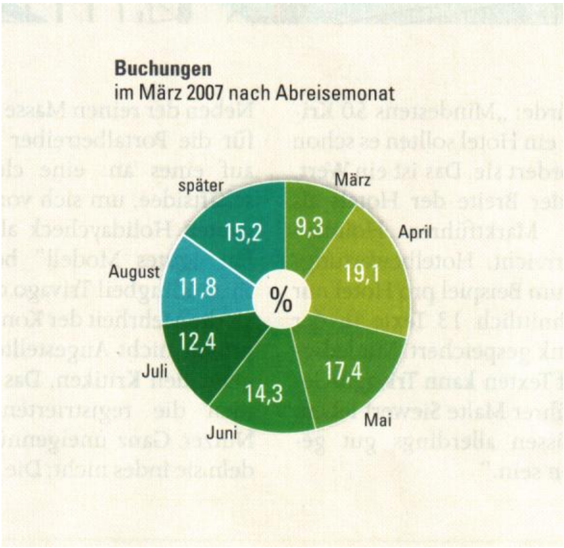


Stacked Bar Chart with Vertical Legend

Source: https://zinc.isc-seo.upenn.edu/wi/help/EN/images/Ewbu26048a.gif

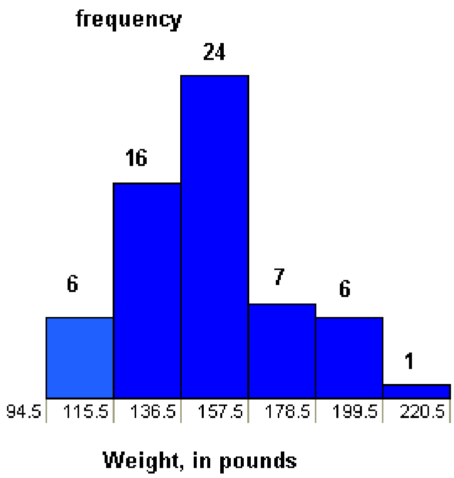
Pie chart

The variables in the pie chart are the numbers of bookings in percentages in march 2007. The categories are the different months in which the holidays will start. The most booked holidays in march are booked for April, May, June, July and August.



Source: FVW Travel Magazine 12/ 2007 page 38

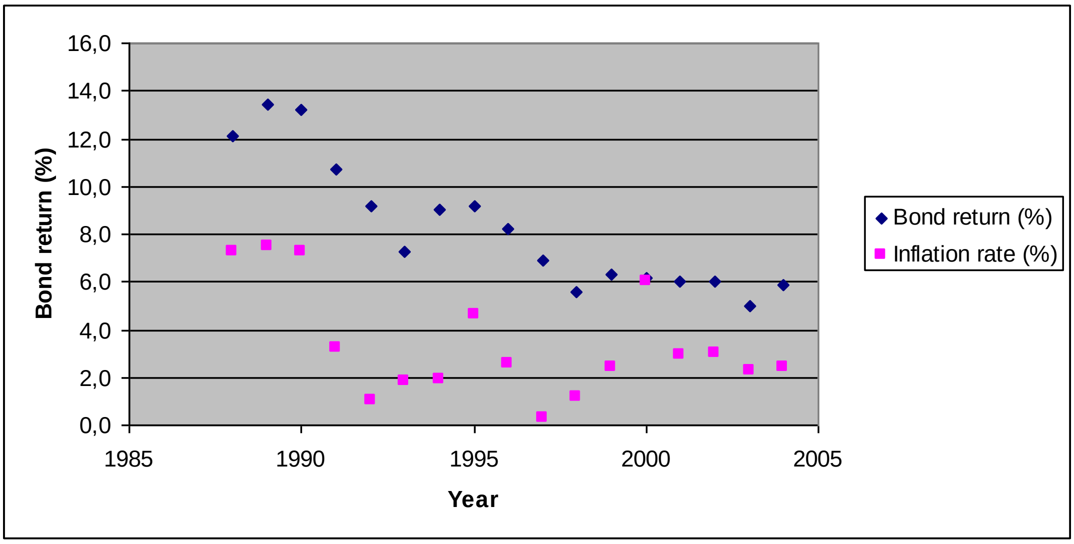
Histogram



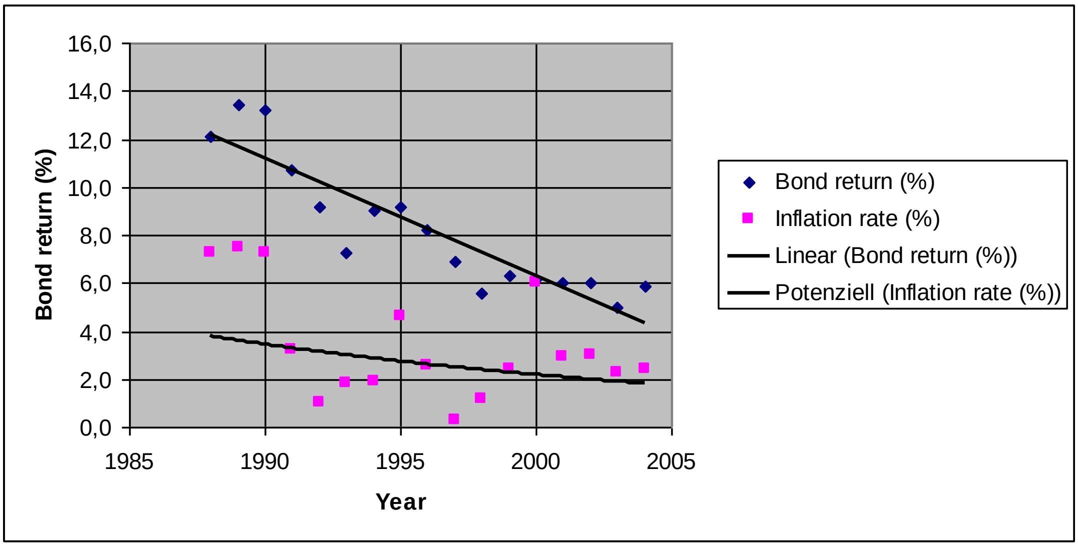
Source: http://sofia.fhda.edu/gallery/statistics/images/histogram1.gif

The variables in this graph are the numbers of frequencies. The categories are the different weights in pounds. The histogram describes a right skewed curve. With 24 has the weight of 157.5 the highest frequency.

Line graphs



Since 1988 the bond return rate and the inflation rate converge. Maybe in the future there will be equal.



The bond returns rate decrease. The inflation rate will be equal. The inflation rate was the highest in the time between 1988 and 1990.

Use of Descriptive Statistics

Example: The series inflation rate and bond return
Measures of Location:

	N	Minimum	Maximum	Mean
Inflation rate (%)	17	0,3	7,5	3,394
Bond return (%)	17	5,0	13,4	8,247

Measures of Dispersion:

	N	Range	Std. Deviation	Variance
Inflation rate (%)	17	7,2	2,3015	5,297
Bond return (%)	17	8,4	2,7178	7,386

Example: Waiting times in a bank queue

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance
Wait time in min.	315	7.54	2.36	9.90	5.3252	1.16614	1.360

	N	Mean	Std. Deviation	Std. Error Mean
Wait time in min.	315	5.3252	1.16614	0.06570

<u>**Association between two variables**</u>

$$\mu_{1,1} = E[(X-\mu_X)(Y-\mu_Y)] = Cov(X,Y) = \sigma_{XY}$$
$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = E[XY] - \mu_X \mu_Y$$

is called the **Covariance** between random variables X and Y.

Theorem: If random variables X and Y are independent, then Cov(X,Y)=0.

Remark: The reverse is not always true.

Theorem: Let $X_1, X_2, .., X_n$ be random variables having finite variances, $\sigma^2_{X_1}, \sigma^2_{X_2}, ..., \sigma^2_{X_n}$, respectively, and covariance, $\sigma_{X_i X_j} \neq 0 \quad i \neq j \quad i, j = 1, 2, .., n$. Define a new random variable $Y = \sum_{i=1}^{n} a_i X_i$ for any set of constants a_i, i=1,..,n. Then

$$\sigma_Y^2 = Var[Y] = Var[\sum_{i=1}^{n} a_i X_i] = \sum_{i=1}^{n} a_i^2 Var[X_i] + 2\sum_{i \neq j} \sum a_i a_j \sigma_{X_i X_j}$$

Example: Let X_1, X_2 be r.v's having means μ_{X_1}, μ_{X_2}, variances $\sigma_{X_1}^2, \sigma_{X_2}^2$ and covariance $\sigma_{X_1 X_2} \neq 0$

$$Y = 4X_1 - 5X_2$$
$$E[Y] = 4\mu_{X_1} - 5\mu_{X_2}$$
$$Var[Y] = 16\sigma_{X_1}^2 + 25\sigma_{X_2}^2 - 40\sigma_{X_1 X_2}$$

Correlation Coefficient

Definition: Let X_1, X_2 be r.v's having means μ_{X_1}, μ_{X_2}, variances $\sigma_{X_1}^2, \sigma_{X_2}^2$ and covariance $\sigma_{X_1 X_2} \neq 0$, the correlation coefficient, ρ; the measure of association between two variables is

$$\rho = \frac{Cov(X_1, X_2)}{\sqrt{Var[X_1]}\sqrt{Var[X_2]}} = \frac{\sigma_{X_1 X_2}}{\sigma_{X_1}\sigma_{X_2}}; \qquad -1 \leq \rho \leq 1$$

Example:
Pearson´s Correlation:

		Bond return (%)	Inflation rate (%)
Bond return (%)	Correlation	1	0,688
	Sig. (2-tailed)	0	0,002*
	N	17	17
Inflation rate (%)	Correlation	0,688	1
	Sig. (2-tailed)	0,002*	0
	N	17	17

** Correlation is significant at the 0.01 level (2-tailed).

The Method of Least Squares:

Given a linear equation $Y = \alpha + \beta X + \epsilon$, the estimated line $\hat{\mu}_{Y|X} = E[Y|X] = \hat{\alpha} + \hat{\beta}X$ requires the estimation of the parameters α and β. To attain the best fit the estimates of α and β should minimize the sum of squared errors as flows:

$\min \sum_{i=1}^{n} \epsilon_i^2$. Letting $q = \sum_{i=1}^{n} \epsilon_i^2$ and

$\frac{\partial q}{\partial \alpha} = 0$, $\frac{\partial q}{\partial \beta} = 0$ result in the normal equations

$$\sum_{i=1}^{n} y_i = n\hat{\alpha} + \hat{\beta} \sum_{i=1}^{n} x_i$$

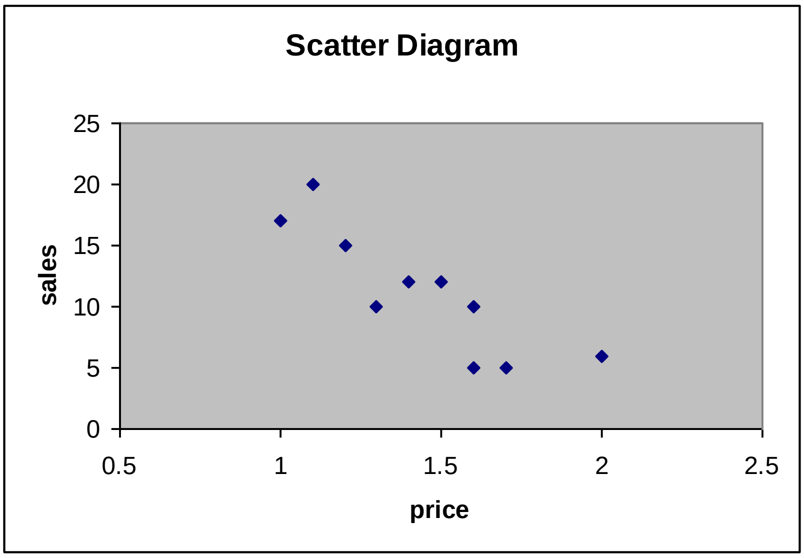
$$\sum_{i=1}^{n} x_i y_i = \hat{\alpha} \sum_{i=1}^{n} x_i + \hat{\beta} \sum_{i=1}^{n} {x_i}^2$$

Solving two equations and two unknowns give:

$$\hat{\beta} = \frac{n\sum_{i=1}^{n} x_i y_i - (\sum_{i=1}^{n} y_i)(\sum_{i=1}^{n} x_i)}{n\sum_{i=1}^{n} {x_i}^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{S_{XY}}{S_{XX}} \; ; \quad \hat{\alpha} = \frac{(\sum_{i=1}^{n} y_i)}{n} - \hat{\beta}\frac{(\sum_{i=1}^{n} x_i)}{n} \Rightarrow \hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

Example: Simple Linear Regression Analysis

Suppose Mr. Bump observes the selling price and sales volume of milk gallons for 10 randomly selected weeks as follows



WEEK	X*	Y*	X^2	Y^2	XY
1	1.3	10	1.69	100	13
2	2	6	4	36	12
3	1.7	5	2.89	25	8.5
4	1.5	12	2.25	144	18
5	1.6	10	2.56	100	16
6	1.2	15	1.44	225	18
7	1.6	5	2.56	25	8
8	1.4	12	1.96	144	16.8
9	1	17	1	289	17
10	1.1	20	1.21	400	22
sum	14.4	112	21.56	1488	149.3

* Thousand of gallons

Normal equation:

$$\sum y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i$$

$$1120 = (10)\hat{\beta}_0 + 14.4\hat{\beta}_1$$

$$S_{xx} = n\sum x^2 - (\sum x)^2 = (10)21.56 - (14.4)^2 = 8.24$$

$$S_{yy} = n\sum y^2 - (\sum y)^2 = (10)1488 - 112^2 = 2336$$

$$S_{xy} = n\sum xy - \sum x \sum y = (10)149 - (14.4)(112) = -119.8$$

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}} = \frac{(10)(149.3) - (14.4)112}{(10)(21.56) - (14.4)^2} = \frac{-119.8}{8.24} = -14.54$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}\ \bar{x} = \frac{112}{10} - (-14.54)\frac{14.4}{10} = 32.14$$

Regression Equation:

$$\hat{y} = 32.14 - 14.54x$$

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} = \frac{(10)149.3 - (14.4)112}{\sqrt{(10)21.56 - 14.4^2}\sqrt{(10)1488 - 112^2}} = -0.86$$

By using bivariate normal approach

$$E_{Y|x} = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1)$$

$$E_{Y|x} = 11.2 - 0.86 \frac{4.833}{0.827}(x - 1.44) = 32.05 - 14.48x$$

Standard error of estimate:

week	X	Actual Y	Estimated Y	Error e	e^2
1	1.3	10	13.238	-3.238	10.48464
2	2	6	3.06	2.94	8.6436
3	1.7	5	7.422	-2.422	5.866084
4	1.5	12	10.33	1.67	2.7889
5	1.6	10	8.876	1.124	1.263376
6	1.2	15	14.692	0.308	0.094864
7	1.6	5	8.876	-3.876	15.02338
8	1.4	12	11.784	0.216	0.046656
9	1	17	17.6	-0.6	0.36
10	1.1	20	16.146	3.854	14.85332
sum	14.4	112	112.024	0	59.42482

$$\hat{\sigma}_e = \sqrt{\frac{\sum error^2}{n-2}} = \sqrt{\frac{\sum (y - \hat{y})^2}{n-2}} = \sqrt{\frac{59.42}{8}} = 2.72$$

Predicting Y:

Suppose Mr. Bump wished to forecast the quantity of milk sold if the price were set at \$1.63.

$\hat{y} = E_{Y|x=1.63} = 32.14 - (14.54)1.63 = 8.44$ or 8,440 gallons.

Standard error of the forecast measures when x=1.63 is

$$\sigma_p = \sigma_e \sqrt{1 + \frac{1}{n} + \frac{(x - \bar{x})^2}{\sum (x_i - \bar{x})^2}} = 2.72\sqrt{1 + \frac{1}{10} + \frac{(1.63 - 1.44)^2}{0.8240}} = 2.90$$

Then 95% prediction interval is

$8.44 \pm t_{n-2,\,0.025}\, 2.90$ => $8.44 \pm 2.306(2.90)$ => (1.753, 15.121)

Inferences on β_1:

Standard error of estimator of β_1

$$\sigma_{\beta_1} = \frac{\sigma_e}{S_{xx}} = \frac{\sigma_e}{\sqrt{\sum (x-\bar{x})^2}} = \frac{2.72}{\sqrt{0.824}} = 3.00$$

The 95% confidence interval is:

-14,54 $\pm\, t_{n-2,\,0.025}(3.00)$ => -14.54 $\pm$ 6.918 => (-21.458, -7.622)

Hypothesis Testing

H_0: $\beta=0$ vs H_0: $\beta\neq 0$

$$t = \frac{-14.54-0}{3.00} = -4.8 < -2.306$$ Reject Ho.

Chapter 2
Introduction to Time Series

Examples of Time Series

1. Economic Time Series: share prices on successive days, export totals on successive days, average incomes in successive months, company profits in successive years, Annual growth rate, Seasonal ice cream consumption, Weekly traffic volume
2. Physical Time Series: Meteorology, marine science and geophysics
   Hourly temperature readings, Rainfall in successive days, air temperature in successive hours, Electrical signals
3. Marketing Time series: Advertising expenditure in successive time periods, the analysis of sales figures in successive weeks/months
4. Process Control: to detect the changes in the performance of a manufacturing process by measuring variable which shows the quality of the process.
5. Binary process: Observations can take one of only two values, usually denoted by 0 and 1. Particularly in communication theory.
6. Point process: A series of events ocurring randomly in time. Dates of a major railway disasters. Distribution of the no. of events and the time intervals between events are concerned.
7. Demographic Time Series: in the study of population to predict the changes in population
8. Data in business, economics, engineering, environment, medicine, earth sciences, and other areas of scientific investigations are often collected in the form of time series, i.e. Daily stock prices

Objectives of Time Series

1. Description: to detect the data and to obtain simple descriptive measures, to detect outlers and adjust to its expected value, to detect turning points (i.e. upward trend suddenly changed to downward trend)
2. Explanation: the variation in one series to explain the variation in another series. Multiple linear regression and linear systems are useful
3. Prediction: Given the observed time series to predict future values of the values
4. Control the Process : control charts

2.1. Time Series Components

Definition: A Stochastic Process is a process that developes in time according to probabilistic laws. Let X_t a string of random variables, X_t is the observation at time t. We usually observe only one realization of a stochastic process over a finite period of time.

Definition: A Time Series is a set of observations generated sequentially in time. They are statistically dependent observations. It is a particular realization of Stochastic Processes. If t is continuous, we have continuous time series. If the realizations are taken at specific time points it is discrete type of time series.

Types of variation
TS=TrendxSeasonalxCyclicalxIrregular

Seasonal: A pattern of change that repeats itself period after period.
Cyclical: Variation at a fixed period due to some other physical cause. i.e. business cycles with a period of 5 and 7 years.
Trend: Lond-term change in the mean which represents the growth or decline in the time series over extended period of time
Irregular: Series of residuals. It measures the variability of the time series after the other components are removed.

Definition: Time Series is said to be stationary if there is no systematic change in mean (no trend), no systematic change in variance and if strictly periodic variations have ben removed.

Examples: The following graphs illustrate different series plotted against different time periods.

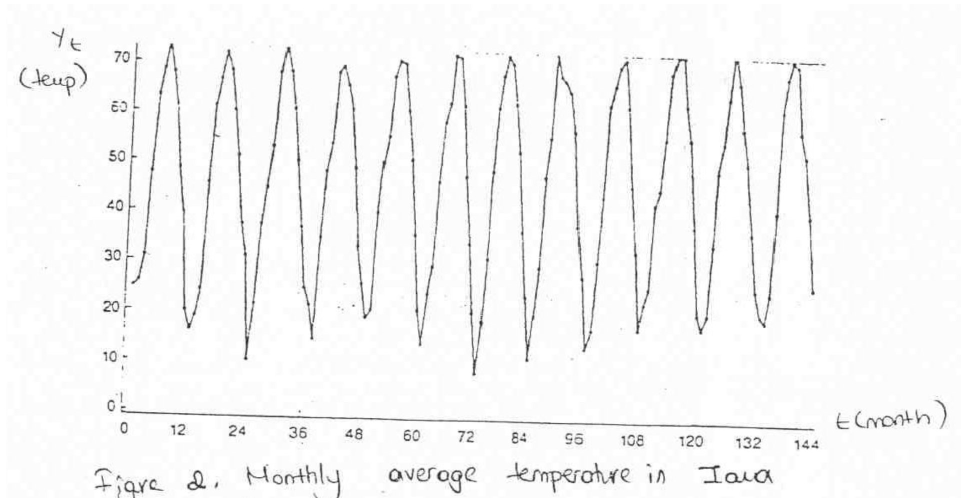


Figure 2. Monthly average temperature in Iowa (January 1964 – December 1975)

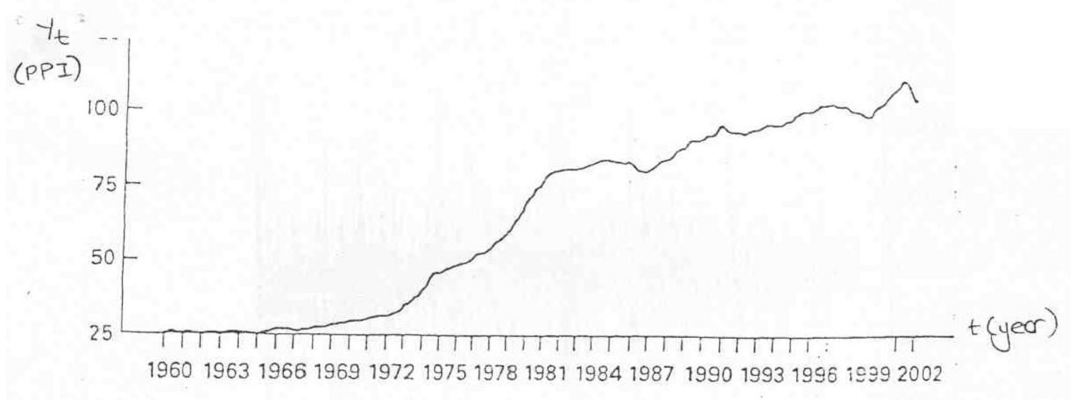


Figure 3. Producer Price Index (base 1995=100)

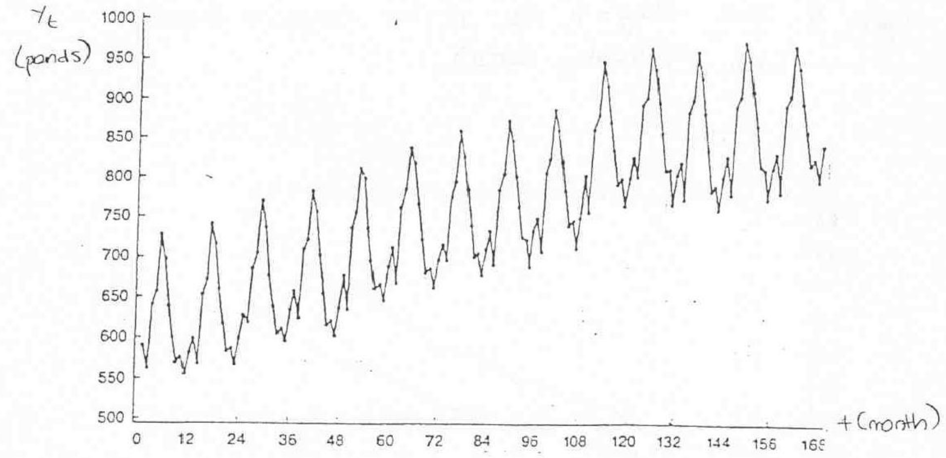


Figure 4. Monthly Milk Production Per Cow
(Jan. 1962 – Dec. 1975

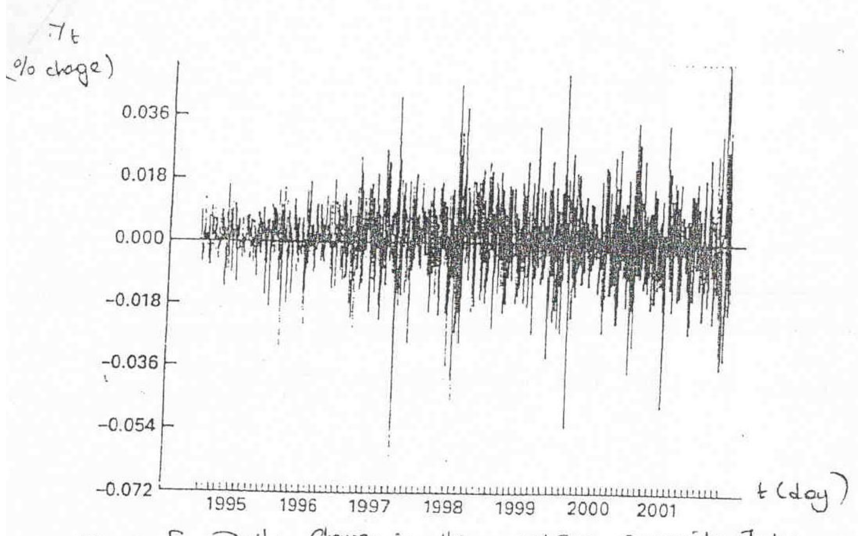
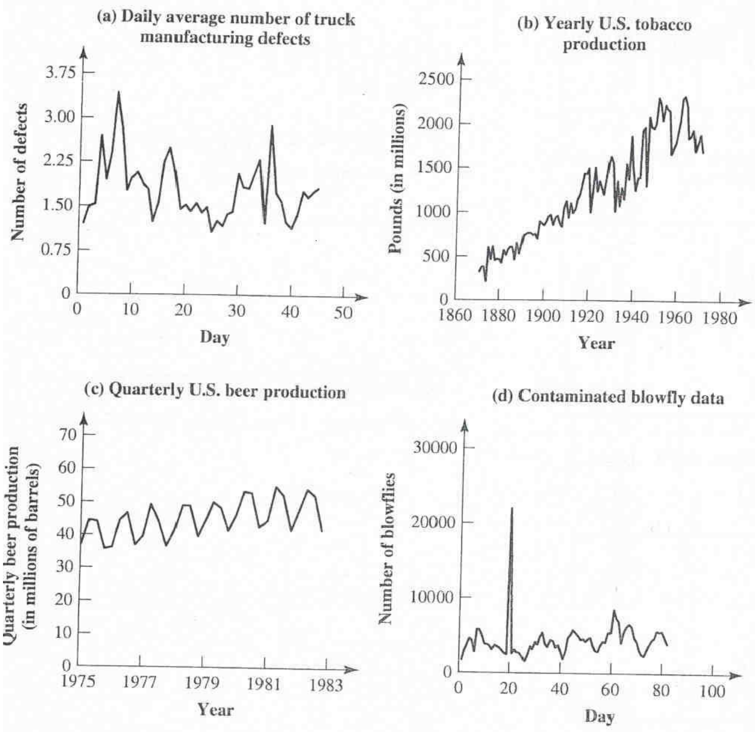
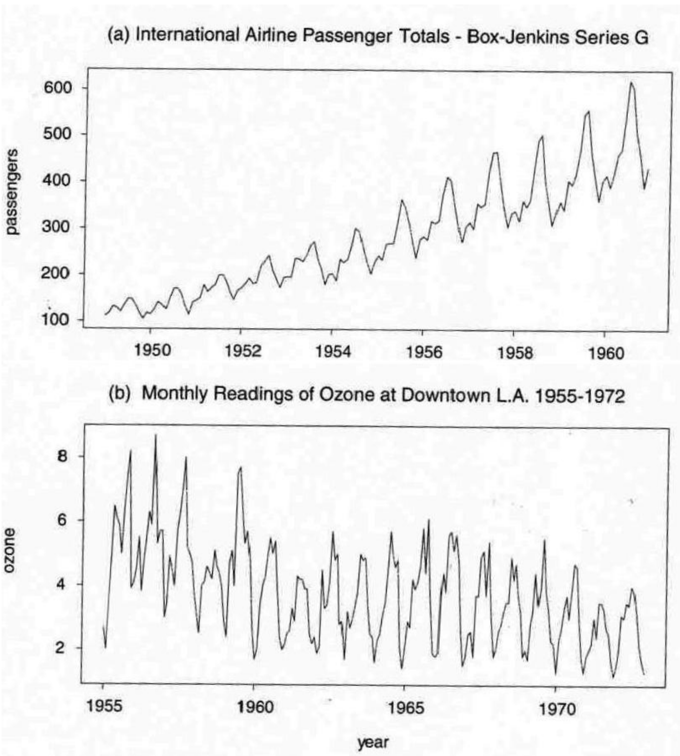


Figure 5. Daily Changes in the NYSE Composite Index

CHAPTER 1 Overview



'IGURE 1.1 Examples of some time series.



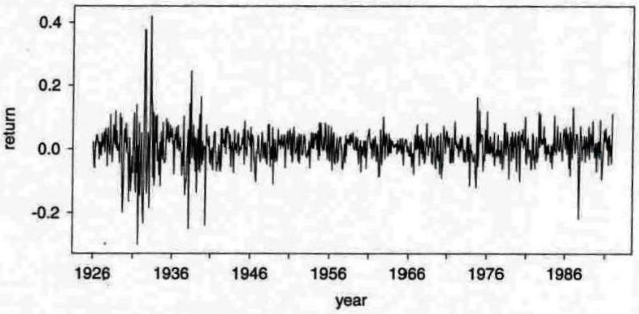
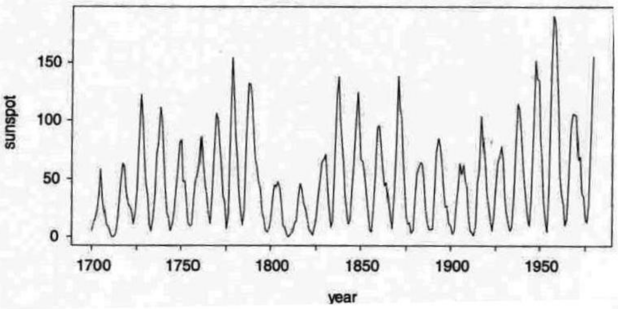
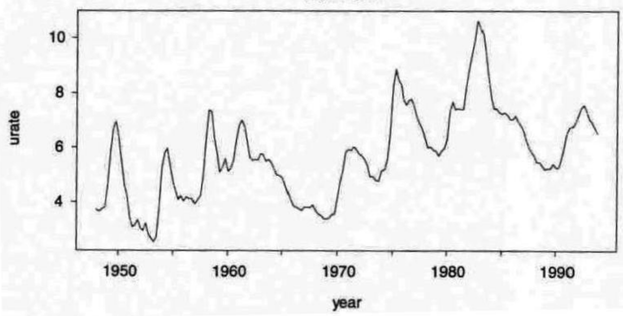


FIGURE 1.8 Value-weighted S&P 500 returns 1926–1991.

(a) Annual Sunspot Numbers 1700-1979



(b) Seasonally Adjusted Quarterly US Unemployment Rates 1948-1993



2.2. The steps in Time Series:

- Model Identification
  - Time Series plot of the series
  - Check for the existence of a trend or seasonality
  - Check for the sharp changes in behavior
  - Check for possible outliers
- Remove the trend and the seasonal component to get stationary residuals.
- Estimation
  - Method of Moments Estimation (MME)
  - Maximum Likelihood Estimation (MLE)
- Diagnostic Checking
  - Normality of error terms
  - Independency of error terms
  - Constant error variance (Homoscedasticity)
- Forecasting
  - Exponential smoothing methods
  - Minimum mean square error (MSE) forecasting

Transformation

The reasons to use transformation on data are to stabilize the variance, to make the seasonal effect additive and to make the data normally distirbuted. For example, when the standard deviation is proportional to the mean, logarithmic transformation is useful. If there is a trend in the series and the size of the seasonal effect appears to increase with the mean, transformation is required. If seasonal effect is directly proportional to the mean, seasonal effect is said to be multiplicative and logarithmic transformation is used.

There are three types of seasonal models:

a. $X_t=\mu_t+S_t+\varepsilon_t$ additive model, no transformation is needed.

b. $X_t=\mu_t S_t \varepsilon_t$ logarithmic transformation

c. $X_t=\mu_t S_t+\varepsilon_t$

where μ_t is the mean; S_t is the seasonality effect and ε_t is the irregular effect

Analysing series which contain a trend

We measure the trend and/or remove the trend in order to analyze the fluctuations. With seasonal data start with calculating the successive yearly averages. The techniques used are polynomial fitting, difference filters.

1. **Polynomial fitting** such as a polynomial curve (linear, quadratic etc.) , logistic function or Gompertz function etc.

Logistic function:

$$f(t)=\frac{a}{1+be^{-ct}}; \quad t\in\mathbb{R}, a,b,c\in\mathbb{R}/\ 0$$

$$\lim_{t\to\infty} f_{\log}(t)=a \quad if\ c>0$$

A resembles the maximum growth of the system.

Mitscherlich Curve

This function models the long term growth of a system.

$$f(t)=a+be^{-ct}; \quad t\geq 0, a,b\in\mathbb{R}, c<0$$

$$\lim_{t\to\infty} f_{\log}(t)=a$$

where a is the saturation value of the system. The initial value of the system is f(t)=a+b.

Gompertz Function

To model the increas or decrease of the system.

$$f(t)=\exp\ a+bc^{t}\ ; \quad t\geq 0, a,b\in\mathbb{R}, c\in(0,1)$$

$$\log\ f(t)=a+be^{t\log c}$$

Allometric Function

Used commonly to model the trend function in biometry and economics.

$$f(t)=bt^{a}; \quad t\geq 0, a\in\mathbb{R}, b>0$$

$$\log\ f(t)=\log b+a\log t\rightarrow \text{Cobb-Douglas Function}$$

Fitted function provides of the trend, and the esiduals provide an estimate of local fluctuations.

2. **Linear Filters**

Let $a_{-h}, a_{-h+1},\ldots,a_s$ be arbitrary numbers having h, s≥0, h+s+1≤n, The linear transformation $X_t=\sum_{i=-h}^{s}a_iX_{t-i}$; $t=s+1,..,n-h$, is linear filter with weights $a_{-h}, a_{-h+1},\ldots,a_s$.

For $a_{-h}, a_{-h+1},\ldots,a_s$ satisfying the condition $\sum_{i=-h}^{s}a_i=1$, then the process is called Moving Average of order s.

Difference Filters

Lemma: For a polynomial $f(t) = a_0 + a_1 t + .. + a_p t^p$ of degree p, the difference $\Delta f(t) = f(t) - f(t-1) = (a_0 + a_1 t + .. + a_p t^p) - (a_0 + a_1(t-1) + .. + a_p (t-1)^p)$ is a polynomial of degree at most p-1.

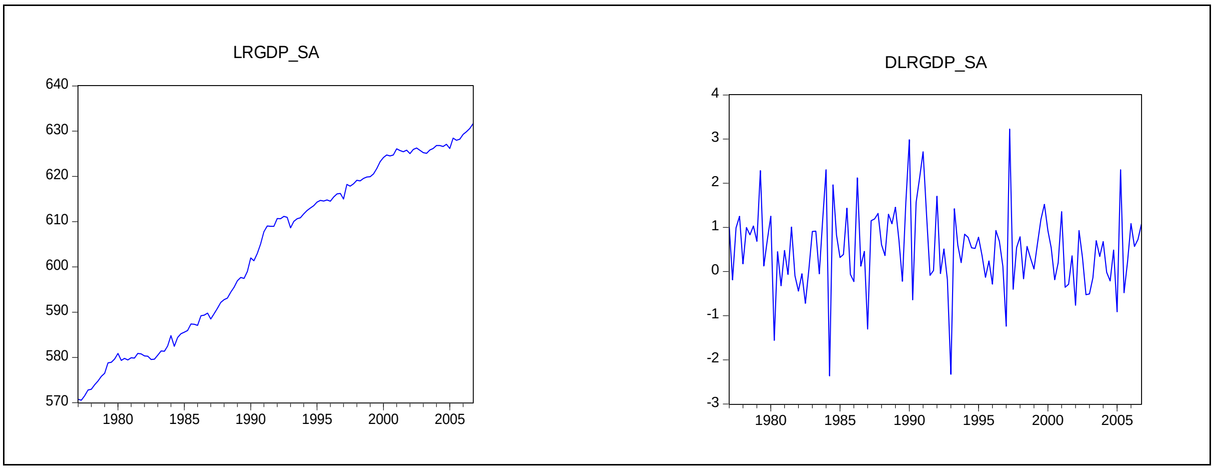
For a linear filter $\Delta X_t = X_t - X_{t-1}$ of order 1.

$\Delta^2 X_t = \Delta(\Delta X_t) = \Delta X_t - \Delta X_{t-1} = X_t - 2X_{t=1} + X_{t-2}$ of order 2.

$\Delta^p X_t = \Delta(\Delta^{p-1} X_t)$ of order p.

If a time series has a polynomial trend of order p, then the difference filter of order p removes the trend up to a constant.

Example: The series having an upward trend (left) is detrended by taking the difference once (p=1) (plot given on the right)



Backward shift Operator (B or L)

$X_{t-1} = BX_t, \qquad X_{t-2} = B^2 X_t,$

$\Delta^2 X_t = \Delta(\Delta X_t) = \Delta X_t - \Delta X_{t-1} = X_t - 2X_{t=1} + X_{t-2} = (1 - 2B - B^2) X_t = (1-B)^2 X_t$

$\Delta^d X_t = (1-B)^d X_t$

Chapter 3
Classical Time Series Modeling

According to classical time-series analysis an observed time series is the combination of some pattern and random variations. The aim is to separate them from each other in order to describe to historical pattern in the data, and to prepare forecasts by projecting the revealed historical pattern into the future. Traditionally, there are two types of methods for identifying the pattern. (i). Smoothing: The random fluctuations are removed from the data by smoothing the time series.(ii).Decomposition**:** The time series is broken into its components and the pattern is the combination of the systematic parts.

The pattern itself is likely to contain some, or all, of the following three components: trend, seasonal and cyclical.

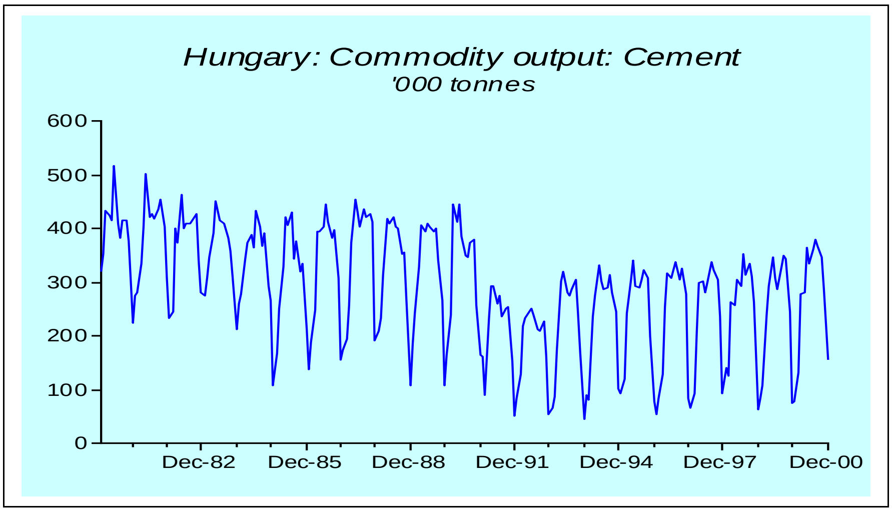
Trend: The long-term general change in the level of the data with a duration of longer than a year. it can be linear (straight line) or non-linear (smooth curve), like e.g. exponential, quadratic.

Seasonal variations: Regular wavelike fluctuations of constant length, repeating themselves within a period of no longer than a year. Seasonal variations are usually associated with the four seasons of the year, but they may also refer to any systematic pattern that occurs during a month, a week or even a single day.

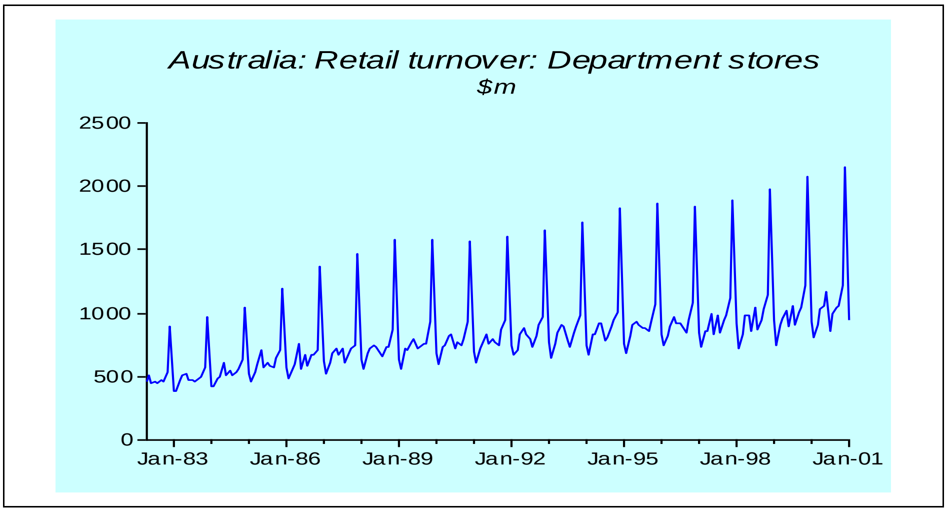
Cyclical variations: Wavelike movements, quasi regular fluctuations around the long-term trend, lasting longer than a year.

Some examples to illustrate these patterns are given in the following figures as follows:

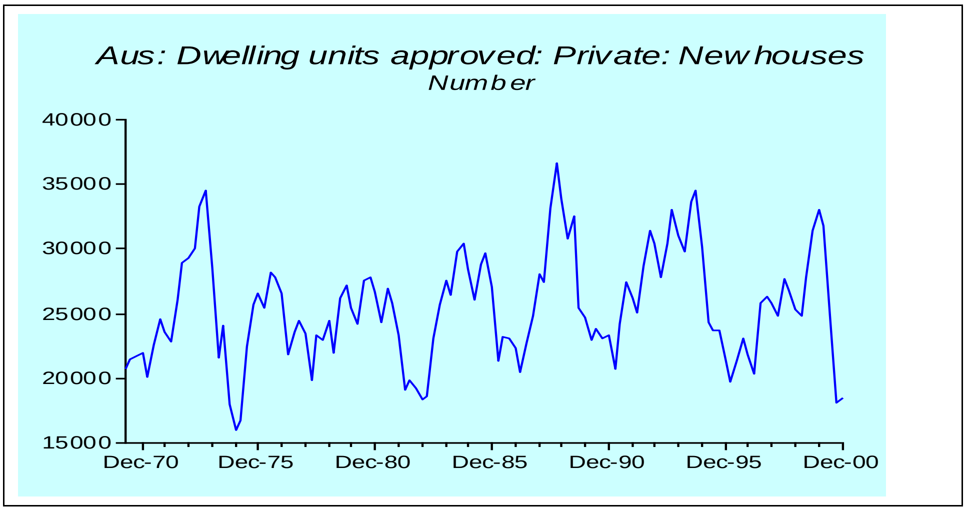
a. Hourly earnings of manufacturing in major seven countries



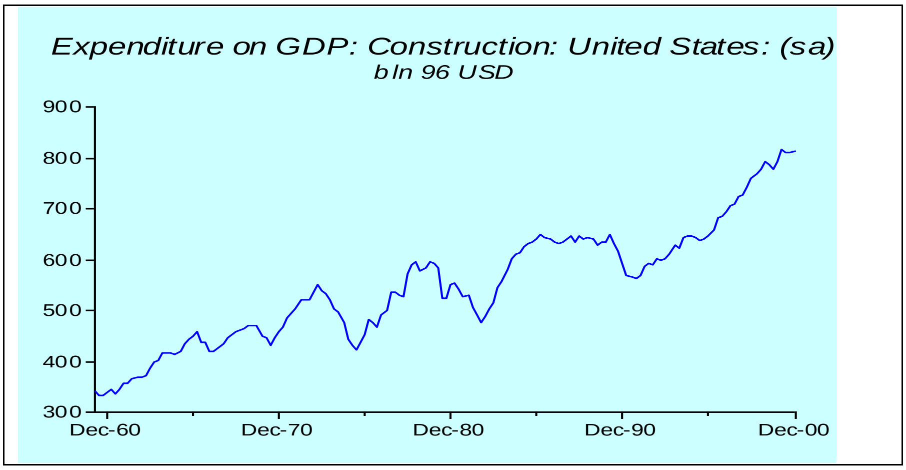
b. Quarterly amount of Cement as commodity output for Hungary between1991-2000



c. Monthly Retail turnover for Australia between 1983-2001



d. Number of new houses approved in Australia (1970-2000)



e. Construction expenditure in USA (1960-2000)

Figure 3.1: Figures a-e are the examples to the components of a time series data (source: Selvanathan et al. 2004)

The time period between the beginning trough and the peak is called expansion phase, while the period between the peak and the ending trough is termed contraction phase. Cyclical variations are often attributed to business cycles, i.e. to the ups and downs in the general level of business activity. Seasonal and cyclical variations might be very similar in their appearance. However, while seasonal variations are absolutely regular and occur over calendar periods no longer than a year, cyclical variations might and do change both in their intensity (amplitude) and duration, and they last longer than a year. It is far more difficult to study and predict the cyclical component than the seasonal component.

The random variations of the data comprise the deviations of the observed time series from the underlying pattern. When this irregular component is strong compared to the (quasi-) regular components, it tends to hide the seasonal and cyclical variations, and it is difficult to be detached from the pattern. However, if we manage to capture the trend, the seasonal and cyclical variations, the remaining changes do not have any discernible pattern, so they are totally unpredictable.

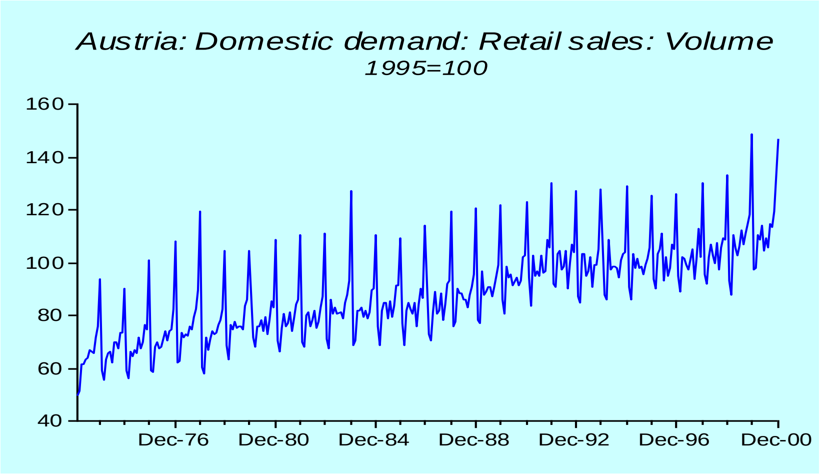
The four components of a time series (T: trend, S: seasonal, C: cyclical, R: random) can be combined in different ways. Accordingly, the time series model used to describe the observed data (Y) can be either

Additive $Y_t = T_t + S_t + C_t + R_t$, or Multiplicative: $Y_t = T_t \times S_t \times C_t \times R_t$

For example, if the trend is linear, these two models look as follows:

$Y_t = (a + bt) + S_t + C_t + R_t$ $Y_t = (a + bt) \times S_t \times C_t \times R_t$

In an additive model the seasonal, cyclical and random variations are absolute deviations from the trend. They do not depend on the level of the trend, whereas in a multiplicative model the seasonal, cyclical and random variations are relative (percentage) deviations from the trend. The higher the trend, the more intensive these variations are.



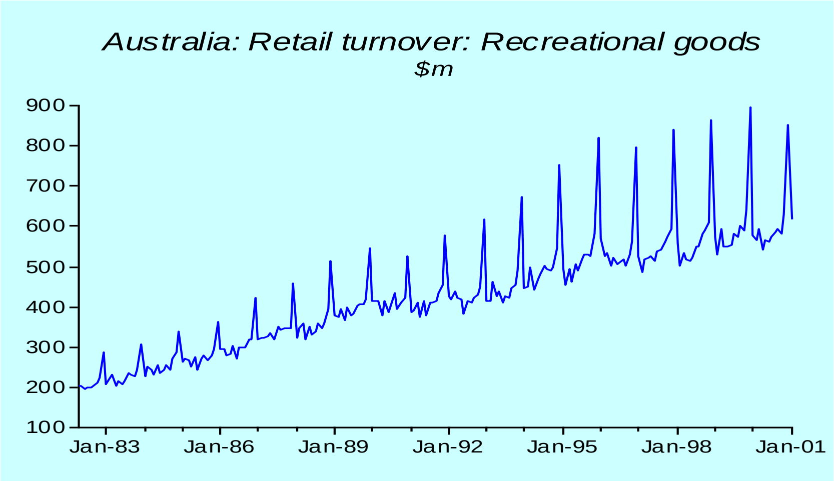


Figure 3.2: Examples to the additive and multiplicative time series (Source: Selvanathan et al. 2004)

These time series have an increasing linear trend component, but the fluctuations around this trend (the first figure above) have the same intensity; the fluctuations around this trend (the second above) are more and more intensive. Though in practice the multiplicative model is the more popular, both models have their own merits and, depending on the nature of the time series to be analysed, they are equally acceptable.

3.1. Smoothing Techniques

They are used to remove, or at least reduce, the random fluctuations in a time series so as to more clearly expose the existence of the other components. There are two types of smoothing methods: (i) **Moving averages:** A moving average for a given time period is the (arithmetic) average of the values in that time period and those close to it. (ii) **Exponential Smoothing:** The exponentially smoothed value for a given time period is the weighted average of all the available values up to that period.

Example: Moving Average

Given the sales per day in the following table,

Day	Sales	3-day moving sum	3-day moving average
1	43		
2	45	110.0	36.7
3	22	92.0	30.7
4	25	78.0	26.0
5	31	107.0	35.7
6	51	etc.	etc.

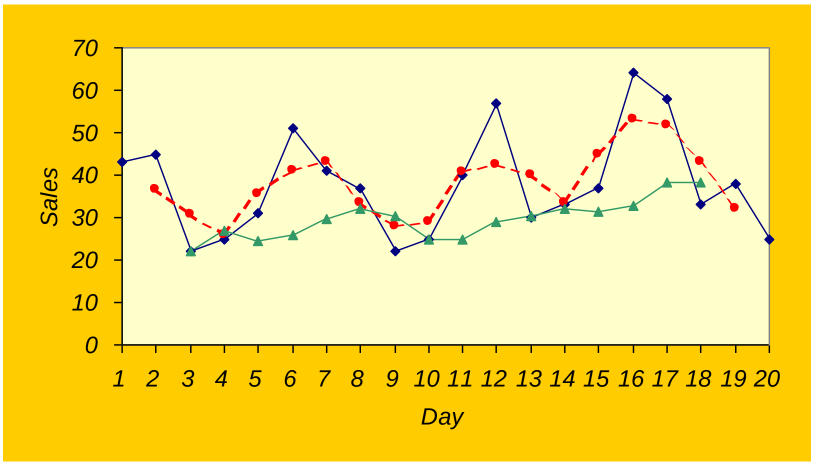


Figure 3.3: Plots of original, MA(3) and MA(5) series

3-day Moving averages (MA(3)) and 5-day Moving averages (MA(5)) are given in Figure 3.3. In this figure blue diamond, red dots and green triangles represent the original, MA(3) and MA(5) series, respectively. This figure suggests that the longer the moving average period the stronger the smoothing effect, the shorter the smoothed series. When the moving average period is relatively large, along with the random variations, the seasonal and cyclical variations are also removed and only the long-term trend can be revealed.

Exponential Smoothing

Let St : exponentially smoothed value for time period t ;

$$S_t = wY_t + (1-w)S_{t-1}$$

where

S_{t-1} : exponentially smoothed value for time period t -1;

Y_t : observed value for time period t ;

w : smoothing constant, $0 < w < 1$.

Note: Assuming that Y has been observed from t =1, this formula can be applied only from the second time period. For t = 1 we set the smoothed value equal to the observed value, i.e. $S_1 = Y_1$. The smoothing constant determines the strength of smoothing, the larger the value of w the weaker the smoothing effect.
The formula for the exponentially smoothed series can be expanded as follows:

$$S_t = wY_t + (1-w)S_{t-1} = wY_t + (1-w)(wY_{t-1} + (1-w)S_{t-2})$$

$$S_t = wY_t + w(1-w)Y_{t-1} + (1-w)^2 S_{t-2}$$

$$= wY_t + w(1-w)Y_{t-1} + (1-w)^2 Y_{t-2} + w(1-w)^3 Y_{t-3} + .. + w(1-w)^{t-1} Y_1$$

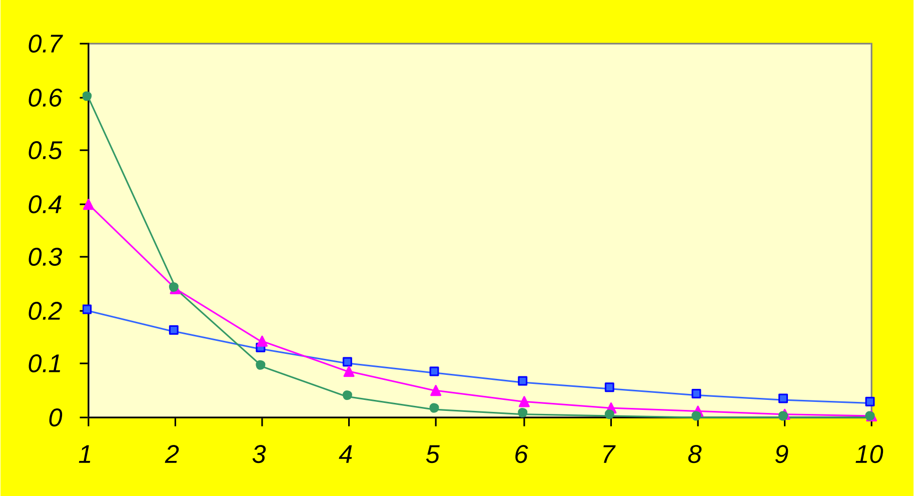


Figure 3.4. The impact of the value of w on convergence

The exponentially smoothed value for period t depends on all available observations from the first period through period t, but the weights assigned to past observations, w(1-w) decline geometrically with the age of the observations (Figure 3.4). Beyond a certain age the observations do not really count since they do not have measurable effects on the exponentially smoothed value.

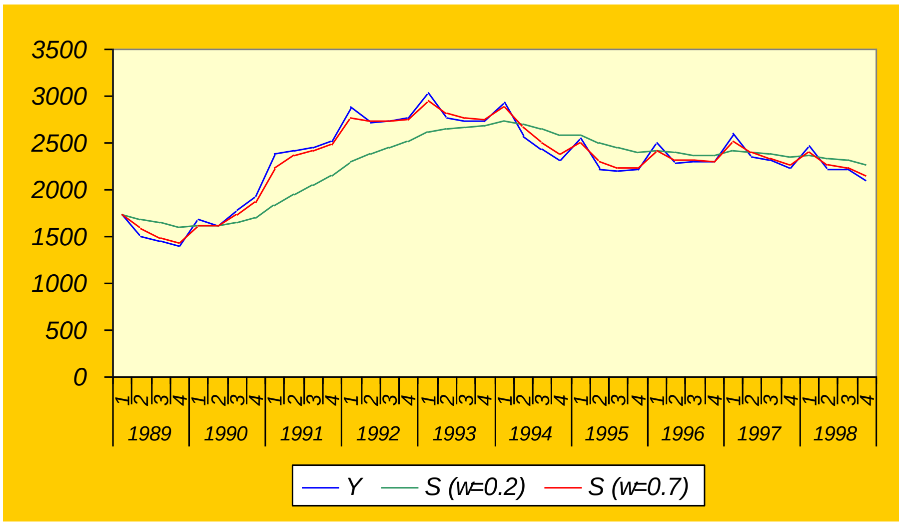


Figure 3.5. Exponentially smoothed series for different w.

Figure 3.4 shows that if w = 0.2, $w(1-w)^t$ approaches zero relatively slowly and even $w(1-w)^{10}$ is substantial. On the other hand, if w = 0.6, $w(1-w)^t$ approaches zero much faster and at t = 6 it is already negligible.

		Y	S (w=0.2)	S (w=0.7)
1989	1	1735.6	1735.6	1735.6
	2	1507.9	1690.1	1576.2
	3	1450.2	1642.1	1488.0
	4	1402.7	1594.2	1428.3
1990	1	1689.9	1613.3	1611.4
	2	1621.4	1615.0	1618.4
		etc.	etc.	etc.

$$S_2 = wY_2 + (1-w)S_1 = 0.2x1507.9 + 0.8x1735.6 = 1690.1$$
$$S_3 = wY_3 + (1-w)S_2 = 0.2x1450.2 + 0.8x1690.1 = 1642.1$$

for w = 0.7, S_t is quite similar to Y_t, i.e. there is very little smoothing. However, if w = 0.2, S_t does not have the seasonal pattern of Y_t, i.e. there is far more smoothing.

3.2 Capturing the Components

Smoothing procedures are used to facilitate the identification of the systematic components of the time series. If we manage to decompose the time series into the trend, seasonal and cyclical components, then we can construct a forecast by projecting these parts into the future.

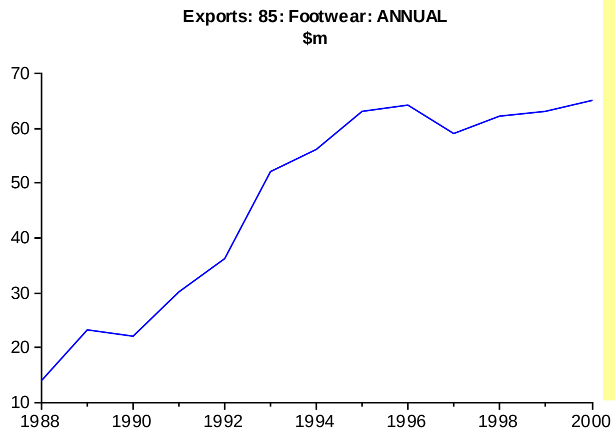
Trend analysis: The easiest way of isolating a long-term linear trend is by simple linear regression, where the independent variable is the t time variable.

$Y_t = \beta_0 + \beta_1 t + \varepsilon_t$ and t is equal to 1 for the first time period in the sample and increases by one each period thereafter. After having created this variable, this linear time trend model can be estimated as any other simple linear regression model. It should be noted that this model is not appropriate if the trend is likely to be non-linear.

Example: The graph below shows exports of footwear ($m) from 1988 through 2000. This time series has an upward trend, which is perhaps linear perhaps not. We fit first a linear regression model to the data and test the significance if the model is appropriate. To estimate a linear trend line, first you have to create a time variable *t* and then regress *fwexport* on *t*.

1988	1	14
1989	2	23
1990	3	22
1991	4	30
1992	5	36
1993	etc	etc

$$\hat{y} = 15.308 + 4.505t$$



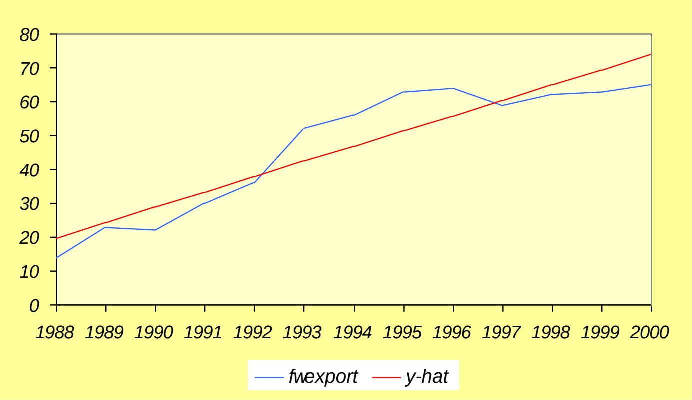


Figure 3.6: The plot of the original series and the trend analysis

Therefore, for example, in 1988 (t = 1) it is $\hat{y} = 15.308 + 4.505 \times 1 = 19.813$ \$m

and in 1999 (t = 12) it is $\hat{y} = 15.308 + 4.505 \times 12 = 69.368$ \$m

Measuring the cyclical effect

Assume that the time series model is multiplicative and consists of only two parts: the trend and the cyclical components so that

$Y_t = T_t x C_t \rightarrow C_t = \dfrac{Y_t}{T_t}$. Under these assumptions the cyclical effect can be measured by

expressing the actual data as the percentage of the trend: $\dfrac{Y_t}{\hat{Y}_t} x100$.

Example continued: Calculate and plot the percentage of trend.

year	t	fwexport	y-hat	y/y-hat*100
1988	1	14	19,81	70,66
1989	2	23	24,32	94,58
1990	3	22	28,82	76,32
1991	4	30	33,33	90,01
1992	5	etc.	etc.	etc.

$14/19.81*100 \approx 71$

So in 1988 the actual exports of footwear were about 29% below the trend line.

Note: We have assumed that the time series pattern does not have a seasonal component and that the random variations are negligible. The first of these assumptions is certainly satisfied since the data is annual. However, when these assumptions are invalid, we should remove the seasonal and random variations before attempting to identify the trend and cyclical components.

Measuring the seasonal effect

Depending on the nature of the time series, the seasonal variations can be captured in different ways.

i. Assume, for example, that the time series does not contain a discernible cyclical component and can be described by the following multiplicative model

$$Y_t = T_t x S_t x R_t \rightarrow \frac{Y_t}{T_t} = S_t x R_t$$

This suggests that dividing the estimated trend component ($\hat{Y}$) into the time series we obtain an estimate for the product of the seasonal and random variations.

Seasonal Factor: $\frac{Y_t}{\hat{Y}_t} x100$. In order to remove the random variations from this ratio, we average the seasonal factors for each season and adjust these averages to ensure that they add up to the number of seasons. This can be achieved by seasonal indices.

Example: The graph below shows retail turnover for households goods ($m) from the second quarter of 1982 through the fourth quarter of 2000.

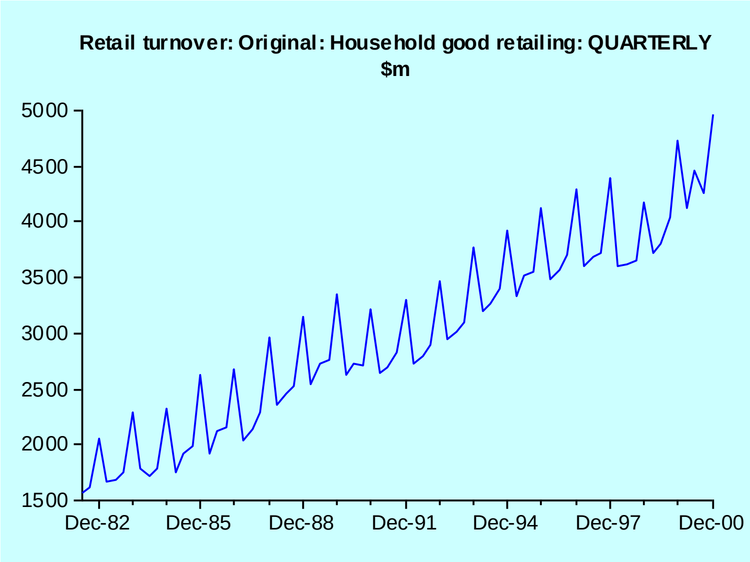


Figure 3.7 Quarterly retail turnover on household good retailing (1982-2000)

This time series has an upward linear trend and quarterly seasonal variations. It probably has some cyclical variations too, but this third component seems to be less significant than the other two. Estimated trend line is

$$\hat{y} = 1589.189 + 36.604t$$

We estimate seasonal factors and the indices:

quarter	t	retail	y-hat	y/y-hat
Jun-82	1	1553.2	1625.8	0.955
Sep-82	2	1601.9	1662.4	0.964
Dec-82	3	2052.2	1699.0	1.208
Mar-83	4	1666.0	1735.6	0.960
Jun-83	5	1680.4	1772.2	0.948

$1553.2/1625.8 = 0.955$

In order to find the seasonal indices the seasonal factors ($Y/\hat{Y}$) have to be grouped, averaged and, if necessary, adjusted.

Year	Q1	Q2	Q3	Q4
1982		0.955	0.964	1.208
1983	0.960	0.948	0.962	1.240
1984	0.948	0.890	0.905	1.163
etc.	etc.	etc.	etc.	etc.
1998	0.914	0.908	0.909	1.031
1999	0.909	0.922	0.971	1.129
2000	0.973	1.043	0.990	1.144

$0.929 \times \frac{4.000}{3.997} = 0.930$

Sum	16.728	18.062	18.283	21.945	Total
Average	0.929	0.951	0.962	1.155	3.997
Index	0.930	0.951	0.963	1.156	4.000

$I_{Mar} = 93.0\%$ $I_{Sep} = 96.3\%$

$I_{Jun} = 95.1\%$ $I_{Dec} = 115.6\%$

These seasonal indices suggest that in the March, June and September quarters retail turnover is expected to be 7.0, 4.9 and 3.7% below its trend value, while in the December quarter retail turnover is expected to be 15.6% above its trend value.

ii. When the time series model is multiplicative and has all four parts, i.e. a trend, a cyclical component, a seasonal component and random variations,

$$Y_t = T_t x C_t x S_t x R_t \rightarrow \frac{Y_t}{CMA_t} = \frac{Y_t}{T_t x C_t} = S_t x R_t$$

the data is first divided by (centered) moving averages, which are supposed to capture the trend and cyclical components, then the seasonal factors and indices are calculated from these ratio-to-moving averages and the trend and cyclical components are estimated from the centered moving averages, instead of the original data.

Note: The order of the centered moving average must be equal to the number of seasons. For example, we use 4-quarter CMA if the data is quarterly and seasonality has 4 phases a year, and we use 12-month CMA if the data is monthly and seasonality has 12 phases a year.

Example continued: Re-estimate the seasonal component using the ratio-to-moving average instead of the original data.

quarter	t	retail	cma(4)
Jun-82	1	1553.2	MISSING
Sep-82	2	1601.9	MISSING
Dec-82	3	2052.2	1734.2
Mar-83	4	1666.0	1767.3
Jun-83	5	1680.4	1814.0
Sep-83	6	etc.	etc.

Following the same steps than in part (b) we get the following seasonal indices:
I_{March}=93%, I_{June}=94.8%, $I_{Sept.}$=96.5%, I_{Dec}=115.7%.

This time there is not much difference between the indices computed from the original data and the indices computed from the centered moving averages. The seasonal indices can be used to deseasonalise a time series, i.e. to remove the seasonal variations from the data. The seasonally adjusted data (in publications usually denoted as *sa*) is obtained by dividing the observed, unadjusted data by the seasonal indices.

For example: For the June quarter of 1982 the seasonally adjusted retail turnover is

$$1553.2 / 94.8 \times 100 = 1638.2 \text{ \$m}$$

3.3 Forecasting

After having studied the historical pattern of a time series, if there is reason to believe that the most important features of the variable do not change in the future, we can project the revealed pattern into the future in order to develop forecasts.

If a time series exhibits no (or hardly any) trend, cyclical and seasonal variations, exponential smoothing can provide a useful forecast for one period ahead: $F_{t+1} = S_t$

Example: Assume that exponential smoothing with w = 0.2 and w = 0.7 on quarterly Australian unemployed persons (in thousands) is applied. Since this time series does have some seasonal variations, exponential smoothing cannot be expected to forecast unemployment reasonably well. Nevertheless, just for illustration, let us forecast unemployment for the first quarter of 1999.

		unemployed	S (w=0.7)
1998	1	2461,4	2402,8
	2	2210,9	2268,5
	3	2221,3	2235,5
	4	2102,6	2142,5

This is the smoothed value for the fourth quarter of 1998, and thus the forecast for the first quarter of 1999.

If a time series exhibits a long-term (linear) trend and seasonal variations, we can use regression analysis to develop forecasts in two different ways.

1. We can forecast using the estimated trend and seasonal indices as:

$$F_t = T_t \times S_t = (\hat{\beta}_0 + \hat{\beta}_1 t) \times I_t$$

2. Alternatively, we can forecast using the estimated multiple regression model with a time variable and seasonal dummy variables.

Example: Forecast retail turnover for households goods for the first quarter of 2001 applying the first approach can be implemented as follows.

Obtain the trend estimate from part *a* and the March seasonal index from part *b so that*
t = 76, I_{76} = *IMar* = 0.930 and $\hat{y} = 1589.189 + 36.604t$

$$F_{76} = \hat{y}_{76} = (1589.2 + 36.6 \times 76) \times 0.930 = 4064.8$$

We have predicted retail turnover for households goods for the first quarter of 2001. Suppose we had another forecast value of 4203.4 for the same data and the same time period using a different forecasting model. How would we decide which forecast is more accurate? However, this does not imply by any means that Model 2 would produce more accurate forecast for all time periods than Model 1.

How can we decide which forecasting model is the most accurate in a given situation?

Forecast the variable of interest for a number of time periods using alternative models and evaluate some measure(s) of forecast accuracy for each of these models. Among a number of possible criteria that can be used for this purpose the two most commonly used are mean absolute devition (MAD) and Sum of squares of forecast error (SSFE). These are as follows:

$$MAD = \frac{1}{n}\sum_{t=1}^{n} |y_t - F_t|$$

$$SSFE = \frac{1}{n}\sum_{t=1}^{n} (y_t - F_t)^2$$

Example: Based on the example above two models are proposed. The forecasts with respect to the actual values are compared and MAD, SSFE are calculated as follows:

Actual	Forecats		error		Squared error	
Value	Model1	Model 2	Model1	Model 2	Model1	Model 2
6.0	7.5	6.3	-1.5	-0.3	2.25	0.09
6.6	6.3	6.7	0.3	-0.1	0.09	0.01
7.3	5.4	7.1	1.9	0.2	3.61	0.04
9.4	8.2	7.5	1.2	1.9	1.44	3.61
					7.39	3.75

Model 1 : MAD = 4.9/4=1.225 and $SSFE$ = 7.39/4=1.8475

Model 2 : MAD = 2.5/4=0.625 and $SSFE$ = 3.75/4=0.9375

According to both criteria Model 2 is the more accurate.

Exercises (Assignment 1)

1. Given the data below and Excel

a. Plot the series and comment on the possible components it might have.
b. Find the smoothed series by using 4-weekMoving Average technique plot the series.
c. Apply exponential smoothing technique with w=0.3
d. Decompose the series into its components and calculate the seasonal indices for every quarters
e. Predict the amount of shipments for the second quarter of 1989.

Year	Quarter	Private Residential investments (billions dollar)	Trend	Year	Quarter	Private Residential investments (billions dollar)	Trend
1980	1	34.2	38.5	1983	1	63.8	47.2
	2	34.3	39.2		2	62.3	47.9
	3	37.7	39.9		3	48.2	48.6
	4	42.5	40.6		4	42.2	49.4
1981	1	43.1	41.4	1984	1	51.2	50.1
	2	42.7	42.1		2	60.7	50.8
	3	38.2	42.8		3	62.4	51.5
	4	37.1	43.5		4	59.1	52.3
1982	1	43.1	44.3	1985	1	47.1	53.0
	2	43.6	45.0		2	44.7	53.7
	3	41	45.7		3	37.8	54.4
	4	53.7	46.4		4	52.7	55.2

2.The following data provide the unemployment rates during 10 years from 1990 to 1999 together with an index of industrial production from Federal Reserve Board.

Year, X2	Unemployment, Y	Index of production, X1
1990	3.1	113
1991	1.9	123
1992	1.7	127
1993	1.6	138
1994	3.2	130
1995	2.7	146
1996	2.6	151
1997	2.9	152
1998	4.7	141
1999	3.8	159

Fit a multiple regression model by using software to express the change in unemployment in terms of year and the index of production.

Chapter 4
Stochastic Time Series Modeling

4.1 Stationary Models

1. Strictly stationary process: If the joint dist. of $(X_{t1},...X_{tn})$ is the same as the joint distribution of $(X_{t1+h},...X_{tn+h})$

2. Weak stationary process: $\{X_t\}$ is weakly stationary (second-order stationary) if

i) $\mu_x(t)=E(X_t)$ is independent of t.

ii) $Cov(X_r,X_s)=\gamma_x(r,s)=E[(X_r-\mu_x(r))(X_s-\mu_x(s))]$ is independent of t for each h.

Autocovariance (ACVF) Function

Let $\{X_t\}$ be a stationary time series with $E(X_t)=\mu$ and $V(X_t)=\sigma^2$. Autocovariance Function of the series is

$$\gamma(h)=Cov(X_{t+h},X_t)$$
$$=E\left[(X_{t+h}-E[X_{t+h}])(X_t-E[X_t]\right]=E\left[(X_{t+h}-\mu_{t+k})(X_t-\mu_t\right]=E\left[X_{t+h}X_t\right]-\mu_{t+h}\mu_t$$

i.e. for t = 1,2,3,... $\gamma(h)=Cov(X_{1+h},X_1)=Cov(X_{2+h},X_2)=Cov(X_{3+h},X_3)=...$

$$\gamma(0)=Var[X_t]=\sigma_t^2$$

Auto-correlation (ACF) Function is $\rho(h)=\dfrac{\gamma_x(h)}{\gamma_x(0)}=\dfrac{C\text{ov}(X_{t+h},X_t)}{Var(X_t)}$

Autocorrrelation (ACF) Function measures the dependency between variables in a series.

$$\rho(h)=\frac{\gamma_x(h)}{\gamma_x(0)}=\frac{C\text{ov}(X_{t+h},X_t)}{Var(X_t)}$$

Remark: For a weak stationary time series, mean is constant and covariance depends only on lag.

Properties:

1. ACF is an even function of the lag $\rho(h)=\rho(-h)$

Proof:

$$\gamma(-h) = Cov(X_{t-h}, X_t)$$
$$= E[X_{t-h}X_t] - \mu_{t-h}\mu_t \quad \textbf{by symmetricity}$$
$$= E[X_{t+h}X_t] - \mu_{t+h}\mu = \gamma(h)$$

2. $|\rho(h)| \le 1$

Proof: Consider the linear function $a_1X_t + a_2X_{t+h}$ where a_i's, i=1,2 are any constants. By the property of variance $Var[a_1X_t + a_2X_{t+h}] \ge 0$

$$Var[a_1X_t] + Var[a_2X_{t+h}] + 2a_1a_2Cov[X_t, X_{t+h}] \ge 0$$
$$a_1^{\,2}Var[X_t] + a_2^{\,2}Var[X_{t+h}] + 2a_1a_2\gamma(h) \ge 0$$
$$Var[X_t](a_1^{\,2} + a_2^{\,2}) + 2a_1a_2\gamma(h) \ge 0$$

If $a_1 = a_2 = 1$, $\quad \gamma(h) \ge -Var[X_t] \Rightarrow \dfrac{\gamma(h)}{Var[X_t]} \ge -1$

If $a_1 = a_2 = -1$, $\quad \gamma(h) \le Var[X_t] \Rightarrow \dfrac{\gamma(h)}{Var[X_t]} \le 1$

Sample Autocovariance and Autocorrelation Function:

Let $x_1, x_2 ..., x_n$ be observations of a series. Given the sample mean is $\overline{X} = \frac{1}{n}\sum_{t=1}^{n} x_t$, the sample autocovariance function $\hat{\gamma}(h) = \frac{1}{n}\sum_{t=1}^{n-|h|}(x_{t+|h|} - \overline{X})(x_t - \overline{X})$ and the sample autocorrelation function $\hat{\rho}(h) = \dfrac{\hat{\gamma}(h)}{\hat{\gamma}(0)}$ $\quad -n < h < n$ where

$$\hat{\rho}(h) = \frac{\sum_{allt}(x_t - \overline{x})(x_{t+h} - \overline{x})}{\sum_{allt}(x_t - \overline{x})^2} = \frac{\sum_{allt}(x_t x_{t+h}) - \sum_{allt} x_t \sum_{allt} x_{t+h}}{\sum_{allt} x^2{}_t - (\sum_{allt} x_t)^2}$$

Lemma: For standard normally distributed data, i.e. if $\{X_t\}$ independent and identically distributed, N (0,1), then sampling distribution of $\hat{\rho}(h) \sim N(0, \frac{1}{n})$.

Correlogram is an aid to interpret a set of ACF where $\hat{\rho}(h)$, sample autocorrelations are plotted against lag h.

Remarks:

For data containing trend $|\hat{\rho}(h)|$ will exhibit slow decay as h increases. For data with a periodic component $|\hat{\rho}(h)|$ will exhibit similar behavior with the same periodicity. If the series is random then for large n, $\hat{\rho}(h) \approx 0$ and $\hat{\rho}_k \sim N(0, \frac{1}{n})$. This leads us to find a 95% confidence interval for the population correlation coefficient. Therefore, we can conclude that if 95% of $\hat{\rho}(h)$ values lie within $\mp \frac{2}{\sqrt{n}} \Rightarrow$ time series is random.

When there exists a short-term correlation, fairly large value of $\hat{\rho}(1)$ is followed by 2 or more coefficients which is significantly smaller than zero, tend to get successively smaller and $\hat{\rho}(h)$ gets to zero for large h. In alternating series correlogram also tends to alternate.

For a non-stationary series: If the series contains a trend, $\hat{\rho}(h)$ values will not come down to zero except very large h. Trend should be removed first. In seasonal fluctuations: Correlogram exhibit an oscillation at the same frequency. If X_t follows a sinusoidal pattern, then so does $\hat{\rho}(h)$.

Tests of serial correlation

Durbin Watson statistic (DW)

DW is used to detect the serial correlation in error process. It is an informative statistics for the regression estimations. The test statistics in Durbin Watson is

$DW = \frac{\sum_{t=2}^{n} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n} \varepsilon_t^2}$ where ε_t is the residual from the estimated equation. It can be shown that DW≈2-2ρ where ρ is the first order serial correlation coefficient. When there is no serial correlation, ρ=0 and DW statistic takes a value close to 2. Positive serial correlation produces a DW<2, while negative serial correlation produces a DW>2. This test can also be generalized to tests of higher orders.

Portmanteau Test:

An important source of information in detecting the presence and form of serial correlation is the correlogram. Qualitative examination of the correlogram is an important diagnostic tool but it does not constitute a formal statistical test. The Box-Pierce and its related test the Ljung-Box test are both portmanteu tests which allow us to test the hypothesis that the first h points in the correlogram are random with a true value of zero.

Box-Pierce test statistics is defined as $Q = n\sum_{i=1}^{h}\hat{\rho}_i^2$

Q is asymptotically distributed as Chi-square distribution with degrees of freedom being h. A modified sample statistics is Ljung-Box statistics is

$$Q^* = n(n+2)\sum_{i=1}^{h}\frac{1}{(n-i)^i}\hat{\rho}_i^2 .$$

Q^* is also distributed Chi-square with degrees of freedom of h. Under the null hypothesis of no serial correlation, large Q or Q^* value indicates the presence of serial correlation.

Example: Given the table below choose the best fitting model.

p	q	$\hat{\sigma}^2$	SIC	AIC
0	1	1.033	-9.149	-8.155
0	2	0.962	-9.191	-8.215
0	3	0.955	-9.169	-8.210
2	0	0.984	-9.168	-8.191
3	0	0.973	-9.149	-8.177
3	1	0.971	-9.122	-8.181
1	2	0.964	-9.158	-8.20

Based on the AIC and SIC values the model chosen is ARIMA(0,1,2)

Example: Determine which of the series whose corrlograms are given beow, do have serial correlation.

Correlogram of D4BIP

Date: 06/08/05 Time: 10:15
Sample: 1960Q1 1998Q4
Included observations: 151

Lag	AC	PAC	Q-Stat	Prob
1	-0.129	-0.129	2.5762	0.108
2	-0.046	-0.064	2.9089	0.234
3	0.156	0.145	6.7181	0.081
4	-0.276	-0.250	18.702	0.001
5	-0.033	-0.086	18.877	0.002
6	0.008	-0.054	18.888	0.004
7	-0.045	0.016	19.215	0.008
8	-0.142	-0.226	22.472	0.004
9	-0.033	-0.119	22.645	0.007
10	0.002	-0.057	22.645	0.012
11	-0.029	-0.017	22.782	0.019
12	-0.076	-0.219	23.754	0.022
13	0.096	-0.014	25.291	0.021
14	-0.015	-0.065	25.330	0.031
15	0.036	0.029	25.549	0.043
16	0.148	0.001	29.281	0.022
17	-0.094	-0.096	30.813	0.021
18	0.040	-0.018	31.089	0.028
19	0.019	-0.006	31.151	0.039
20	-0.131	-0.142	34.197	0.025
21	0.134	0.074	37.391	0.015
22	-0.064	-0.076	38.133	0.018
23	-0.034	0.001	38.340	0.023
24	0.031	-0.079	38.515	0.031
25	-0.039	0.018	38.795	0.039
26	-0.086	-0.169	40.176	0.037
27	-0.001	-0.040	40.176	0.049
28	0.076	-0.002	41.249	0.051
29	-0.003	0.031	41.251	0.065
30	0.061	-0.033	41.963	0.072
31	-0.012	-0.080	41.993	0.090
32	0.092	0.089	43.622	0.083
33	-0.040	0.023	43.940	0.097
34	0.013	-0.050	43.973	0.118
35	0.070	0.040	44.949	0.121
36	-0.142	-0.056	48.999	0.073

Correlogram of DLOGBIP

Date: 06/08/05 Time: 10:04
Sample: 1960Q1 1998Q4
Included observations: 155

Lag	AC	PAC	Q-Stat	Prob
1	-0.316	-0.316	15.732	0.000
2	-0.284	-0.426	28.598	0.000
3	-0.322	-0.794	45.249	0.000
4	0.922	0.713	182.19	0.000
5	-0.312	-0.153	197.98	0.000
6	-0.267	-0.026	209.63	0.000
7	-0.319	-0.077	226.35	0.000
8	0.876	0.075	353.29	0.000
9	-0.297	-0.007	368.01	0.000
10	-0.248	-0.016	378.34	0.000
11	-0.319	-0.065	395.54	0.000
12	0.827	-0.061	511.89	0.000
13	-0.266	0.086	523.98	0.000
14	-0.219	0.056	532.26	0.000
15	-0.313	0.051	549.29	0.000
16	0.791	0.104	658.71	0.000
17	-0.254	-0.039	670.09	0.000
18	-0.202	0.015	677.34	0.000
19	-0.313	-0.009	694.90	0.000
20	0.751	-0.007	796.48	0.000
21	-0.231	0.069	806.13	0.000
22	-0.194	-0.065	813.01	0.000
23	-0.316	-0.022	831.44	0.000
24	0.721	0.046	928.09	0.000
25	-0.220	-0.057	937.17	0.000
26	-0.182	0.009	943.40	0.000
27	-0.319	-0.068	962.77	0.000
28	0.694	-0.028	1054.9	0.000
29	-0.203	-0.021	1062.9	0.000
30	-0.154	0.108	1067.5	0.000
31	-0.309	0.092	1086.2	0.000
32	0.657	-0.030	1171.6	0.000
33	-0.188	0.078	1178.6	0.000
34	-0.129	0.045	1182.0	0.000
35	-0.304	0.063	1200.7	0.000
36	0.607	-0.053	1276.1	0.000

4.2. The models

1.White noise (WN) Process (Random shock)

$\{X_t\}$ is a sequence of independent and identical random variables with zero mean and finite variance, σ^2, $X_t \sim WN(0,\sigma^2)$, $\{X_t\}$ is stationary with

$$\gamma_x(t+h,t) = \begin{cases} \sigma^2 & h=0 \\ 0 & h \neq 0 \end{cases}; \qquad \rho_k = \begin{cases} 1, k=0 \\ 0, k \neq 0 \end{cases}; \qquad \phi_{kk} = \begin{cases} 1, k=0 \\ 0, k \neq 0 \end{cases}$$

White Noise process is a purely random process where all autocorrelation functions for every h are close to zero. White noise (in spectral analysis): white light is produced in which all frequencies (i.e., colors) are present in equal amount. It is a memoryless process, builds block from which we can construct more complicated models and it plays the role of an orthogonal basis in the general vector and function analysis.

Example: White Noise process is a purely random process where all autocorrelation functions for every h are close to zero.

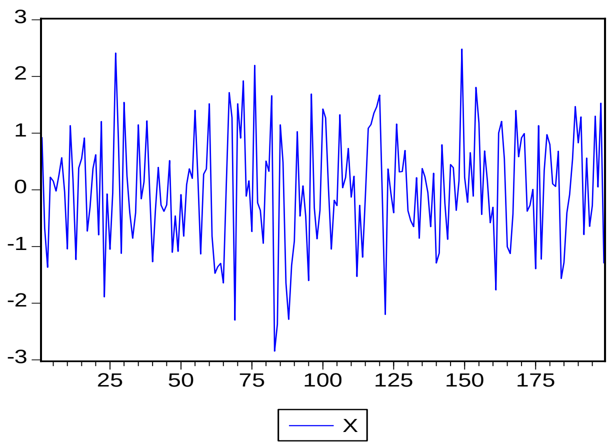


Figure 4.1. The plot of index numbers having White Noise model

2. Random Walk

Let $\{S_t, t=1,2,\cdots\}$ be a process with $S_t = X_1 + \cdots + X_t$ where $X_t \sim WN(0,\sigma^2)$. Then,

$E(S_t) = 0; Var(S_t) = t\sigma^2; \gamma(t+h,t) = t\sigma^2$

Since $\gamma(t+h,t)$ depends on t, S_t is not stationary. However, $Z_t = X_t - X_{t-1}$ is stationary.

3. Linear Process

Let $\{Z_t\}$ be a WN process with mean 0 and variance σ^2. $\{X_t\}$ is a Linear process if

$X_t = Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + ... = Z_t + \sum_{i=1}^{\infty} \psi_i Z_{t-i}$ having

$$E[X_t] = 0; Var[X_t] = \sigma_Z^2$$

$$\gamma(h) = \begin{cases} \sigma_Z^2 & h = 0 \\ 0 & h \neq 0 \end{cases}; \quad \rho(h) = \begin{cases} 1 & h = 0 \\ 0 & h \neq 0 \end{cases}$$

and $\sum_{i=1}^{\infty} |\psi_i| < \infty$ as the stationarity condition.

4.Moving Average Process MA(q)

Let $\{Z_t\}$ be a WN process with mean 0 and variance σ^2. $\{X_t\}$ is a Moving Average of order q if $X_t = \theta_0 Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} = \sum_{i=0}^{q} \theta_i Z_{t-i}$ where $\{\theta_i\}$ are the constants and usually $\theta_0 = 1$.

For q=1, MA(1) is $X_t = \theta Z_{t-1} + Z_t$

Then $E(X_t) = 0; V(X_t) = \sigma^2(1+\theta^2) < \infty$

$$\gamma_x(t+h,t) = \begin{cases} \sigma^2(1+\theta^2) & h = 0 \\ \sigma^2\theta & h = \mp 1 \\ 0 & |h| > 1 \end{cases}; \quad \rho_x(h) = \begin{cases} 1 & h = 0 \\ \dfrac{\theta}{1+\theta^2} & h = \mp 1 \\ 0 & |h| > 1 \end{cases}$$

i. MA(q) process is second-order stationary for all values of $\{\theta_i\}$.

ii. If Z_t's are Normally distributed, so are the X_t's $\Rightarrow$ Normal Process $\Rightarrow$ Strictly stationary.

Example:

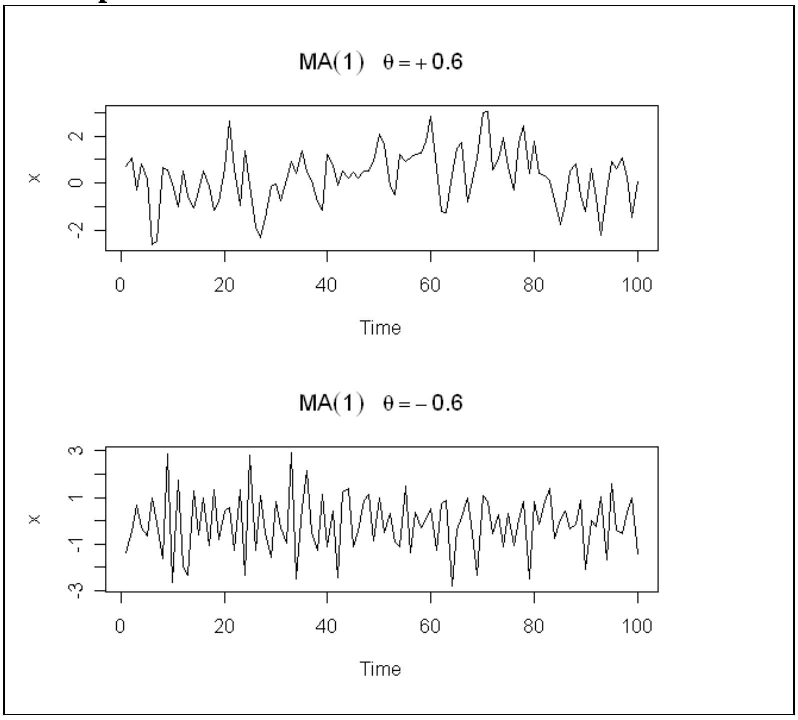


Figure 4.2. An example to MA process

5. Autoregressive Process: AR(p)

Let $\{X_t\}$ is a stationary series and $\{Z_t\}$ is a White Noise with mean 0 and variance σ^2.

$\{X_t\}$ is said to be AR(p) if

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t \quad t = 0, \mp 1, \cdots; \ where \ |\phi_i| < 1$$

AR(1) Process

$$X_t = \phi X_{t-1} + Z_t \quad \text{with} \quad E(X_t) = 0, and \qquad V(X_t) = \frac{\sigma^2}{1-\phi^2} \qquad |\phi_1| < 1$$

Proof: AR(p) can be expressed in terms of MA(∞) by successive substitution.

Take p=1

$$X_t = \phi X_{t-1} + Z_t = \phi[\phi X_{t-2} + Z_{t-1}] + Z_t = \phi^2 X_{t-2} + \phi Z_{t-1} + Z_t = \phi^2[\phi X_{t-3} + Z_{t-2}] + \phi X_{t-1} + Z_t$$

$$\vdots$$

$$X_t = Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \phi^3 Z_{t-3} + \cdots \quad iff \quad |\phi| < 1$$

$$E(X_t) = E[Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \phi^3 Z_{t-3} + \cdots\ ] = 0$$

$$V(X_t) = V[Z_t + \phi Z_{t-1} + \phi^2 Z_{t-2} + \phi^3 Z_{t-3} + \cdots\ ] = \sigma^2[1 + \phi^2 + \phi^4 + \cdots]$$

$$V(X_t) = \frac{\sigma^2}{1-\phi^2} \quad where \quad \sum_{i=0}^{\infty} r^i = \frac{1}{1-r} \quad if\ |r| < 1$$

Example: AR(2) process

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t \qquad Z_t \sim WN(0, \sigma^2)$$

Example:

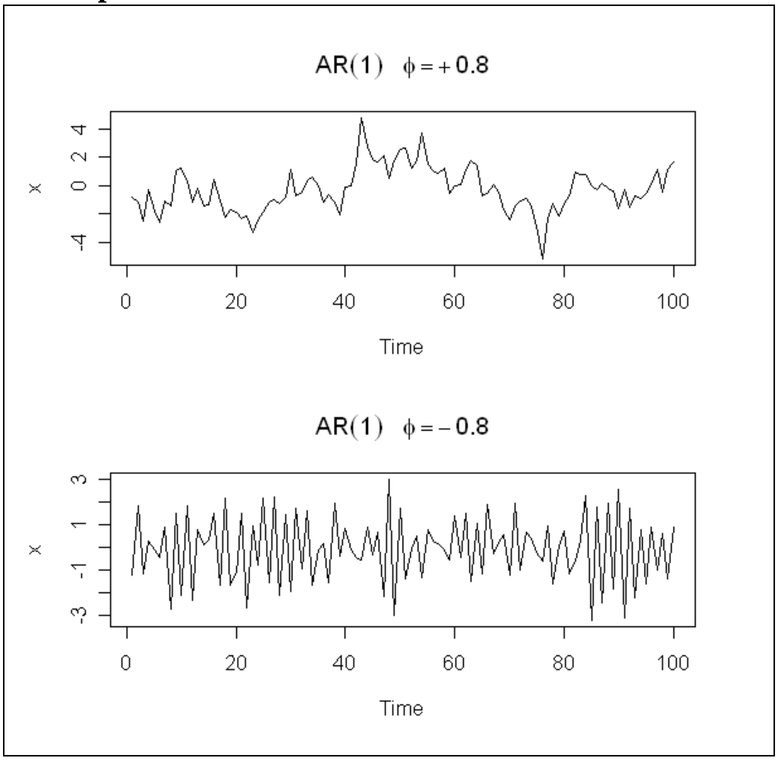


Figure 4.3 An example to AR(1) process

Yule Walker Equations

The recursive computation of the autocorrelation function of an AR(p) model satisfying the stationary condition.

Consider AR(p) process

$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + Z_t$ where $Z_t \approx WN(0, \sigma^2)$ satisfying the stationary conditions. The autocorrelation function of AR(p) satisfies for any h:

$$\rho(h) = \sum_{j=1}^{p} \phi_j \rho(h-j) = \rho(-s)$$

$$\gamma(h) = \sum_{j=1}^{p} \phi_j \gamma(h-j)$$

Proof is straightforward by taking the expectation of AR(p) process is multiplied both sides of the equation by X_{t+h} .

These equations can be expressed as

$$\begin{pmatrix} \rho(1) \\ \rho(2) \\ \rho(3) \\ . \\ .\rho(p) \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & \rho(2) & ...... & \rho(p-1) \\ \rho(1) & 1 & & & \\ \rho(2) & & 1 & & \\ . & . & . & .... & . \\ \rho(p-1) & & & & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ . \\ .\phi_p \end{pmatrix}$$

Replacing the parameters by empirical estimators

$$\begin{pmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \\ \hat{\rho}(3) \\ . \\ .\hat{\rho}(p) \end{pmatrix} = \begin{pmatrix} 1 & \hat{\rho}(1) & \hat{\rho}(2) & ...... & \hat{\rho}(p-1) \\ \hat{\rho}(1) & 1 & & & \\ \hat{\rho}(2) & & 1 & & \\ . & . & . & .... & . \\ \hat{\rho}(p-1) & & & & 1 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ . \\ .\phi_p \end{pmatrix} \Rightarrow \boldsymbol{\rho} = \mathbf{R\Phi}$$

$$\begin{pmatrix} \hat{\phi}_1 \\ \hat{\phi}_2 \\ \hat{\phi}_3 \\ . \\ .\hat{\phi}_p \end{pmatrix} = \begin{pmatrix} 1 & \hat{\rho}(1) & \hat{\rho}(2) & ...... & \hat{\rho}(p-1) \\ \hat{\rho}(1) & 1 & & & \\ \hat{\rho}(2) & & 1 & & \\ . & . & . & .... & . \\ \hat{\rho}(p-1) & & & & 1 \end{pmatrix}^{-1} \begin{pmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \\ \hat{\rho}(3) \\ . \\ .\hat{\rho}(p) \end{pmatrix} \Rightarrow \hat{\mathbf{\Phi}} = \mathbf{R}^{-1}\boldsymbol{\rho}$$

Asymptotic distribution of Yule-Walker estimators

For a causal AR(p) process, the asymptotic distribution of

$\sqrt{n}(\hat{\Phi} - \Phi) \xrightarrow{d} N(0, \sigma^2 R^{-1})$ as $\hat{\sigma}^2 \xrightarrow{p} \sigma^2$

Partial autocorrelation Function (PACF)

Correlogram is useful for identifying a pure moving average model, since there will tend to be cut-off significant points on the correlogram after appropriate lag depth. For autoregressive or mixed processes, the order of the autoregressive component may be harder to determine from the correlogram. For this reason, it is usual to use a complementary procedure which involves plotting the estimated coefficient of X_{t-k}, from an Least Square estimate of an AR(p) model. If the observations are generated by an AR(p) process, then the theoretical partial autocorrelations are zero at lags beyond p. Since any invertible MA process can be represented as an AR process with geometrically decreasing coefficients, the partial autocorrelation function for an MA process should decay slowly. The identification of a mixed model may be more difficult to determine.

Under the assumption of normality the partial correlation of X and Y conditional on W is

$$\rho_{x,y,w} = \frac{E[(X-E(X|W))(Y-E(Y|W))]}{\{E[(X-E(X|W))^2]E[(Y-E(Y|W))^2]\}^{1/2}} = \frac{\rho_{xy}-\rho_{xw}\rho_{yw}}{[(1-\rho_{xw}^2)(1-\rho_{yw}^2)]^{1/2}}$$

For an AR(p) process PAC, ϕ_{hh} is the correlation coefficient between X_t and X_{t-h} controlling the effect of X_{t-h-1}

$$\phi_{hh} = \frac{\begin{vmatrix} 1 & \rho(1) & \cdots & \rho(h-2) & \rho(1) \\ \rho(1) & 1 & \cdots & \rho(h-3) & \rho(2) \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \rho(h-1) & \rho(h-2) & \cdots & \rho(1) & \rho(h) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \cdots & \cdots & \rho(h-1) \\ \rho(1) & 1 & \cdots & \cdots & \rho(h-2) \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ \vdots & \vdots & \cdots & \vdots & \rho(1) \\ \rho(h-1) & \cdots & \cdots & \rho(1) & \rho(h) \end{vmatrix}}$$

Equivalently, Levinson and Durbin’s Recursive Formula gives

$$\phi_{hh} = \frac{\rho(h) - \sum_{j=1}^{h-1}\phi_{h-1,j}\quad\rho_{h-j}}{1-\sum_{j=1}^{h-1}\phi_{h-1,j}\quad\rho_j} \qquad h = 1,2,3,\cdots$$

$$\phi_{hj} = \phi_{h-1} - \phi_{hh}\phi_{h-1,h-j} \qquad j = 1,2,\ldots,h-1$$

Example: For an AR(2) process find the partial autocorrelation function.

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t \qquad Z_t \sim WN(0,\sigma\ \ )$$

$$\phi_{11} = \rho(1); \quad \phi_{22} = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}; \quad \phi_{pp} = \phi_p; \quad \phi_{hh} = 0 \quad h > p$$

Partial Autocorrelation for MA(1) process

$$\phi_{hh} = \frac{\theta^h (1 - \theta^2)}{1 - \theta^{2(h+1)}} \quad for \quad h > 0$$

$$\phi_{11} = \frac{\theta(1 - \theta^2)}{1 - \theta^4} \quad \phi_{22} = \frac{\theta^2 (1 - \theta^2)}{1 - \theta^6} \quad \phi_{33} = \frac{\theta^3 (1 - \theta^2)}{1 - \theta^8}$$

Asymptotic distribution of Partial Autocorrelations

For a causal AR(p) process, the asymptotic distribution of

$$\sqrt{n}\hat{\alpha}_{kk} \xrightarrow{d} N(0,1)$$

6. Combined Autoregressive Moving Average (ARMA) processes

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

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t + \theta_1 X_{t-1} \cdots + \theta_q Z_{t-q}$$

<u>ARMA(1,1) process</u>

$$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t + \theta Z_{t-1}$$

<u>Properties of the ACF and PACF for various ARMA Models</u>

Model	**ACF**	**PACF**
AR(1)	Exponential or oscillatory decay	ϕ_{hh} =0 for h>1
AR(2)	Exponential or sine wave decay	ϕ_{hh} =0 for h>2
AR(p)	Exponential or sine wave decay	ϕ_{hh} =0 for h>p
MA(1)	$\rho_h = 0$ =0 for h>1	Dominated by damped exponential
MA(2)	$\rho_h = 0$ =0 for h>2	Dominated by damped exponential or sine wave
MA(q)	$\rho_h = 0$ =0 for h>q	Dominated by linear combination of damped exponential and/or sine waves
ARMA(1,1)	Tails off. Exponential decay from lag 1	Tails off. Dominated by exponential decay from lag 1
ARMA(p,q)	Tails off after (q-p) lags. Exponential and/or sine wave decay after (q-p) lags	Tails off after (p-q) lags. Dominated by damped exponentials and or sine waves after (p-q) lags

Backward Shift Operator, B:

- $BX_t = X_{t-1}$
- $B(BX_t) = BX_{t-1} = X_{t-2}$

  $B^2 X_t = X_{t-2}$

- $B^j X_t = X_{t-j} \quad j \geq 0 \quad B^0 = 1$

Example:

Random Walk process can be expressed as

$$Y_t = Y_{t-1} + e_t \Rightarrow Y_t - Y_{t-1} = e_t \Rightarrow Y_t - BY_t = e_t \Rightarrow (1 - B)Y_t = e_t$$

Example:

The following series contain some part of the 159 observations on the monthly differences between the yield mortgages and the yield on government loans in Netherlands From Jan. 1961 to Dec. 1973

Year	Jan	Feb	March	April	May	June	July	Aug	Sept	Oct	Nov	Dec
1961	0.66	0.70	0.74	0.63	0.70	0.66	0.61	0.52	0.60	0.61	0.7	1.1
1962	1.17	1.23	0.85	0.78	0.71	0.55	0.56	0.74	0.80	0.75	0.74	0.79
1963	0.78	1	1.05	1.09	1.05	0.75	0.73	0.77	0.77	0.84	0.66	0.68
……												

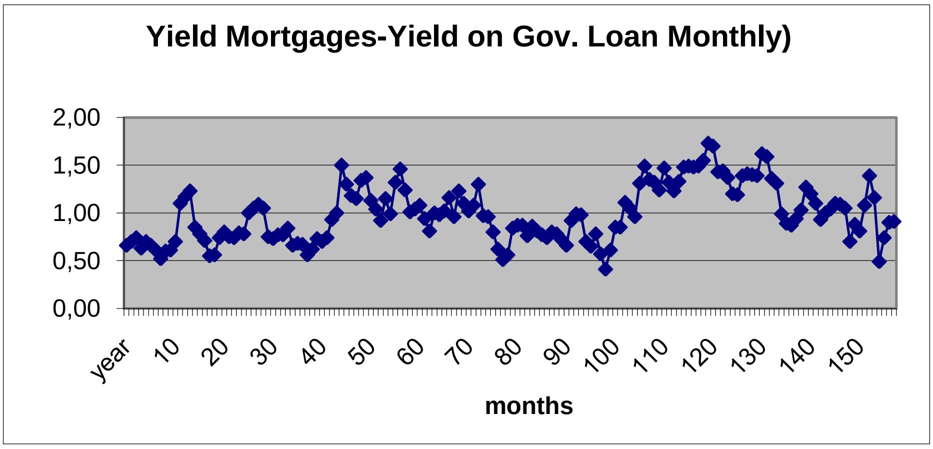


Figure 4.4 Plot of the series

Sample Average $\bar{x} = \dfrac{0.66 + 0.70 + 0.74 + \ldots}{159} = 0.993$

The Variance is

$$\hat{\gamma}(0) = Cov(X_t, X_t) = Var(X_t) = \frac{(0.66-0.993)^2 + (0.70-0.993)^2 + (0.74-0.993)^2 + ...}{158} = 0.085$$

<u>Correlation coefficient for lag h</u>

h=1

$$\hat{\rho}(1) = \frac{Cov(X_t, X_{t+1})}{Var(X_t)} = \frac{\hat{\gamma}(X_t, X_{t+1})}{\hat{\gamma}(X_t, X_t)} = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} = \frac{(0.66-0.993)(0.70-0.993)+(0.74-0.993)(0.63-0.993)+....}{(0.66-0.993)^2+(0.70-0.9)^2+...} = 0.84$$

h=2

$$\hat{\rho}(2) = \frac{\hat{\gamma}(2)}{\hat{\gamma}(0)} = \frac{(0.66-0.993)(0.74-0.993)+(0.70-0.993)(0.63-0.993)+....}{(0.66-0.993)^2+(0.70-0.9)^2+...} = 0.6$$

h=3

$$\hat{\rho}(3) = \frac{\hat{\gamma}(3)}{\hat{\gamma}(0)} = \frac{(0.66-0.993)(0.63-0.993)+(0.70-0.993)(0.70-0.993)+....}{(0.66-0.993)^2+(0.70-0.9)^2+...} = 0.584$$

<u>Autocorrelation and partial auticorrelation Functions</u> for lag h=1,2,3,…,20

ACF

h	1	2	3	4	5	6	7	8	9	10
ρ	0.841	0.683	0.584	0.515	0.457	0.427	0.405	0.386	0.361	0.321
h	11	12	13	14	15	16	17	18	19	20
ρ	0.329	0.338	0.337	0.294	0.231	0.166	0.126	0.062	0.047	0.042

PACF

h	1	2	3	4	5	6	7	8	9	10
φ	0.841	-0.083	0.111	0.036	0.018	0.091	0.025	0.035	0.003	-0.044
h	11	12	13	14	15	16	17	18	19	20
φ	0.168	0.001	0.027	-0.110	-0.08	-0.057	0.007	-0.152	0.122	-0.071

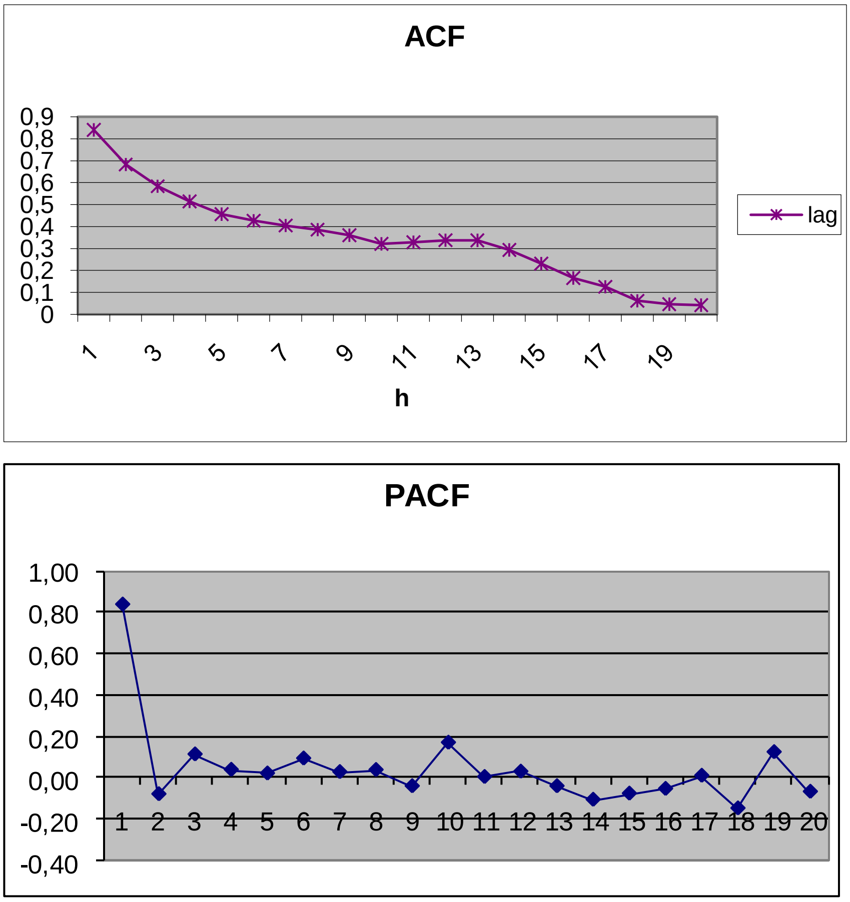


Figure 4.5 Plots of the ACF and PACF values

Dependent Variable: YIELD_DATA
Method: Least Squares
Date: 01/21/08 Time: 00:14
Sample (adjusted): 2 158
Included observations: 157 after adjustments
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.003013	0.079014	12.69410	0.0000
AR(1)	0.841074	0.042868	19.62019	0.0000

R-squared	0.712937	Mean dependent var	0.994586
Adjusted R-squared	0.711085	S.D. dependent var	0.292556
S.E. of regression	0.157251	Akaike info criterion	-0.849287
Sum squared resid	3.832832	Schwarz criterion	-0.810354
Log likelihood	68.66904	F-statistic	384.9517
Durbin-Watson stat	1.860085	Prob(F-statistic)	0.000000

4.3 Stationary Conditions:

Invertibility:

Consider a MA(q) process.

$$X_t = \theta_0 Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q} = (1 + \theta_1 B + \theta_2 B^2 + .. + \theta_q B^q) Z_t$$

$$X_t = \Theta(B) Z_t$$

The model can be written in infinite order autoregressive form with drift

$$X_t = Z_t + \pi_1 X_{t-1} + \pi_2 X_{t-2} + .. = \sum_{i=1}^{\infty} \pi_i X_{t-i} + Z_t \; ; \qquad \sum_{j=1}^{\infty} |\pi_j| < \infty$$

We can express the series as

$$\Pi(B) X_t = Z_t \Rightarrow \Pi(B)\Theta(B) = 1$$

The series is stationary if the the roots of $\Pi(B) = 0$ lies outside of the unit circle, i.e. $|B| > 0$.

This condition is satisfied when $|\theta_i| < 0; i = 1,..,q$. The process is invertible if the coefficients of the MA(q) lie within the unit circle.

Causality: Characteristic equation

Consider AR(p) process given as

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} + Z_t$$

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} - \cdots - \phi_p X_{t-p} = Z_t$$

$$(1 - \phi_1 B + \phi_2 B^2 + \cdots + \phi_p B^p) X_t = Z_t$$

$$\Phi(B) X_t = Z_t$$

Expressing the series as an infinite order MA process yields

$$X_t = Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \cdots = (1 + \psi_1 B + \psi_2 B^2 + \cdots) Z_t = \Psi(B) Z_t$$

$$\Phi(B)\Psi(B) = 1$$

$\Phi(B) = 0$ is called the characteristic equation of the series. To ensure the stationary condition, the roost of the characteristic equation should lie outside the unit circle.

4.4. Estimation

Consider AR(1) process having drift

$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t$ where $|\phi| < 1,\ \mu \in \Re, Z_t \approx N(0,\sigma^2)$, given x_t, t=1,2,..n, the likelihood function

$$L(\mu,\phi,\sigma) = f(x_1) f(x_2|x_1)\ldots f(x_2|x_{n-1})$$

As $X_2|X_1 \approx N(\mu + \phi(X_{t-1} - \mu),\sigma^2)$, $f(x_t|x_{t-1}) = f_Z((x_t - \mu) - \phi(x_{t-1} - \mu))$ and

$$X_1 \approx N(\mu, \frac{\sigma^2}{1-\phi^2})$$

Then the likelihood is

$$L(\mu,\phi,\sigma^2) = f(x_1)\prod_{t=2}^{n} f_Z((x_t - \mu) - \phi(x_{t-1} - \mu))$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} (1-\phi^2)^{\frac{1}{2}} \exp\left\{-\frac{S(\mu,\phi)}{2\sigma^2}\right\}$$

$$S(\mu,\phi) = (1-\phi^2)(x_1 - \mu)^2 + \sum_{t=2} \left[(X_t - \mu) - \phi(X_{t-1} - \mu)\right]^2$$ **Unconditional sum of squares**

$$\ln L(\mu,\phi,\sigma^2) = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln 2\pi + \frac{1}{2}(1-\phi^2) - \frac{S(\mu,\phi)}{2\sigma^2}$$

$$\frac{\partial \ln L(\mu,\phi,\sigma^2)}{\partial\sigma^2} = 0 \Rightarrow \hat{\sigma}^2 = \frac{S(\hat{\mu},\hat{\phi})}{n}$$

AR models are linear models conditional on initial values. Therefore, dropping the term in the likelihood that causes non-linearity, the "**Conditional Likelihood**" is found.

$$L(\mu,\phi,\sigma^2|x_1) = \prod_{t=2}^{n} f_Z((x_t - \mu) - \phi(x_{t-1} - \mu))$$

$$= (2\pi\sigma^2)^{-\frac{n-1}{2}} (1-\phi^2)^{\frac{1}{2}} \exp\left\{-\frac{S_C(\mu,\phi)}{2\sigma^2}\right\}$$

$$S(\mu,\phi) = \sum_{t=2} \left[(X_t - \mu) - \phi(X_{t-1} - \mu)\right]^2$$ **Conditional sum of squares**

$$\frac{\partial \ln L(\mu,\phi,\sigma^2)}{\partial\sigma^2} = 0 \Rightarrow \hat{\sigma}^2 = \frac{S(\hat{\mu},\hat{\phi})}{n-1}$$

Estimates of μ and ϕ are:

Let $\beta = \mu(1-\phi) \Rightarrow S_C(\mu,\phi) = \sum_{t=2}^{n} \left[X_t - (\beta + \phi X_{t-1})\right]^2$. By LSE technique

$$\hat{\beta} = \overline{X}_2 - \hat{\phi}\overline{X}_1 \Rightarrow \hat{\mu} = \frac{\overline{X}_2 - \hat{\phi}\overline{X}_1}{1-\phi^2}; \quad \hat{\phi} = \frac{\sum_{t=2}^{n}(x_t - \overline{X}_2)(x_{t-1} - \overline{X}_1)}{\sum_{t=2}^{n}(x_{t-1} - \overline{X}_1)^2}$$

$$\overline{X}_1 = \frac{1}{n-1}\sum_{t=1}^{n-1} x_t; \quad \overline{X}_2 = \frac{1}{n-1}\sum_{t=2}^{n} x_t$$

Maximum Likelihood of MA(1) Process

Given $Z_t = (X_t - \mu) - \theta Z_{t-1}$

$$L(\mu,\phi,\sigma^2) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \left(\frac{1-\theta^2}{1-\theta^{2(n+1)}}\right)^{\frac{1}{2}} \exp\left\{-\frac{S(\mu,\theta)}{2\sigma^2}\right\}$$

Maximum Likelihood of ARMA(1,1) Process

Given $Z_t = X_t - \phi X_{t-1} - \theta Z_{t-1}$

$$L(\mu,\phi,\sigma^2) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} \left|Z'Z\right|^{-\frac{1}{2}} \exp\left\{-\frac{S(\theta,\phi)}{2\sigma^2}\right\}$$

$$\left|Z'Z\right| = \frac{(1-\phi^2)(1-\theta^2) + (1-\theta^{2n})(\theta-\phi)^2}{(1-\theta^2)(1-\phi^2)}$$

Example

Consider a stationary MA(1) model with zero mean given below. Derive the likelihood function and obtain the unconditional least squares estimates of the parameters.

$X_t = \theta Z_{t-1} + Z_t; \quad Z_t \approx N(0,\sigma^2), i.i.d.$ Given t=1,2

$$L(\phi,\sigma^2) = f(x_1)\, f(x_2 \mid x_1)$$

$$\left.\begin{array}{l} X_1 = Z_1 \\ X_2 = \theta Z_1 + Z_2 \end{array}\right\} \quad Z_1 = X_1 \qquad Z_2 = X_2 - \theta X_1 \Rightarrow |J| = \begin{vmatrix} 1 & 0 \\ \theta & 1 \end{vmatrix}$$

$$L(\phi,\sigma^2) = f(z_1)\, f(z_2 \mid x_1)\, |J|$$

$$L(\phi,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} \exp\left\{-\frac{z_1^2}{2\sigma^2}\right\} \left(\frac{1}{2\pi\sigma^2}\right)^{1/2} \exp\left\{-\frac{z_2^2}{2\sigma^2}\right\}$$

$$L(\phi,\sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right) \exp\left\{-\frac{\left[z_1^2 + (x_2 - \theta x_1)^2\right]}{2\sigma^2}\right\}$$

$$\log L(\phi,\sigma^2) = -\log(2\pi\sigma^2) + \frac{\left[z_1^2 + (x_2 - \theta x_1)^2\right]}{2\sigma^2}$$

$$\frac{\partial}{\partial\sigma^2} \log L(\phi,\sigma^2) = 0 \Rightarrow -\frac{1}{\sigma^2} + \frac{\left[z_1^2 + (x_2 - \theta x_1)^2\right]}{2\left(\sigma^2\right)^2} = 0$$

$$\hat{\sigma}^2 = \frac{\left[z_1^2 + (x_2 - \hat{\theta} x_1)\right]}{2}$$

$$\frac{\partial}{\partial\theta} \log L(\phi,\sigma^2) = 0 \Rightarrow \frac{\left[(x_2 - \theta x_1) x_1\right]}{2\left(\sigma^2\right)^2} = 0$$

$$\hat{\theta} = \frac{x_2}{x_1}$$

Exercises

1. Consider the model $X_t = \phi X_{t-2} + Z_t + \theta Z_{t-1}$ where $Z_t \sim WN$

a. Is the model stationary? Invertible? State the conditions for stationarity of this model.
b. Write the model interms of a linear process, specify ψ_1, ψ_2, ψ_3 in terms of (ϕ,θ)
c. For ϕ=0.2, what constant should be added to the right hand side of the model so that $E(X_t)=5$?
d. Suppose $E(Z_t)=1$. What would be $E(X_t)$ in terms of (ϕ,θ)

2. Consider the following model with $\sigma^2=1.44$.

$$X_t = 0.4X_{t-1} + 0.2X_{t-2} + Z_t; \qquad Z_t \sim WN$$

a. Find ρ_1, ρ_2, ρ_3
b. ϕ_{11}, ϕ_{22}, ϕ_{33} and $\gamma(0)$

3. Consider the model

$$X_t = X_{t-1} + \phi X_{t-2} + Z_t$$

a. For what values of ϕ is the model stationary?
b. Find $\rho(1)$ in terms of ϕ.
c. Find ψ_1, ψ_2, ψ_3, ψ_4.

4. Consider the model

$$X_t = \phi X_{t-2} + Z_t + \theta Z_{t-3}$$

Assuming the model is stationary, find $\gamma(0)$, $\rho(1)$, $\rho(2)$, ϕ_{22}.

5. Let $\{Z_t\}$ be zero-mean white noise. Find the autocorrelation function for the following two processes:

a. $X_t = Z_t + \frac{1}{3}Z_{t-1}$
b. $X_t = Z_t + 3Z_{t-1}$
c. You should have discovered that both series are stationary and have the same autocorrelation functions. Do you think that these models could be distinguished on the basis of observations of Z_t.

6. Suppose $X_t = 5 + 2t + Z_t$ where $\{Z_t\}$ is a zero-mean stationary series with autocovariance function $\gamma(h)$.

a. Find the mean function for $\{X_t\}$
b. Find the autocovariance function for $\{X_t\}$
c. Is $\{X_t\}$ stationary? (Why or why not)

7. Suppose $X_t = \beta_0 + \beta_1 t + Z_t$ where $\{Z_t\}$ is stationary. Show that $\{X_t\}$ is not stationary but that $\nabla X_t = X_t - X_{t-1}$ is stationary.

8. Calculate $V(X_t)$ in terms of σ^2 for the following stochastic process

$$X_t - \mu = Z_t + 0.4Z_{t-1} + (0.4)^2 Z_{t-2} + (0.4)^3 Z_{t-3} + ...$$

9. Find ψ_1, ψ_2, ψ_3 for the following models

a. $(1-0.8B)(X_t-\mu)=Z_t$

b. $X_t=0.8X_{t-1}-0.1X_{t-2}+Z_t$

c. Find $\rho(1)$, $\rho(2)$ and $\rho(3)$ for models (a) and (b).

10. Determine which of the models to be chosen to model the series.

Model 1

Dependent Variable: SERIES01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.842172	0.075479	11.15765	0.0000
MA(1)	-0.302824	0.133818	-2.262957	0.0259

R-squared	0.493503	Mean dependent var	-0.174202
Adjusted R-squared	0.488282	S.D. dependent var	1.638134
S.E. of regression	1.171830	Akaike info criterion	3.175006
Sum squared resid	133.1991	Schwarz criterion	3.227433
Log likelihood	-155.1628	Durbin-Watson stat	1.966866

Inverted AR Roots	.84
Inverted MA Roots	.30

Model 2

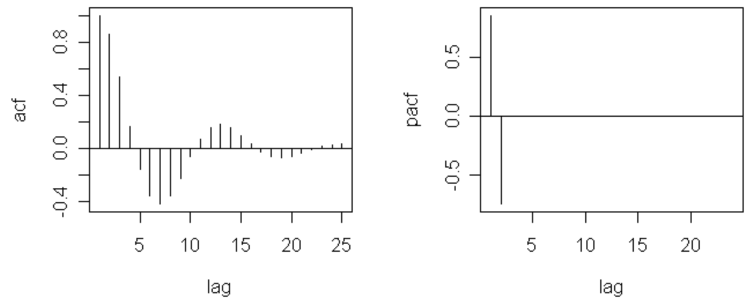
Dependent Variable: SERIES01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.692724	0.073036	9.484638	0.0000

R-squared	0.472652	Mean dependent var	-0.174202
Adjusted R-squared	0.472652	S.D. dependent var	1.638134
S.E. of regression	1.189592	Akaike info criterion	3.195148
Sum squared resid	138.6827	Schwarz criterion	3.221361
Log likelihood	-157.1598	Durbin-Watson stat	2.221013

Inverted AR Roots	.69

10. Write the order of the process based on the ACF and PACF plots below.



11. Determine which of the coefficients to be chosen to model the series.

Model 1

Dependent Variable: SERIES01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.961179	0.436832	2.200340	0.0302
AR(2)	-0.090880	0.329854	-0.275515	0.7835
MA(1)	-0.412044	0.414466	-0.994155	0.3227

R-squared	0.495243	Mean dependent var	-0.175980
Adjusted R-squared	0.484616	S.D. dependent var	1.646460
S.E. of regression	1.181997	Akaike info criterion	3.202423
Sum squared resid	132.7262	Schwarz criterion	3.281554
Log likelihood	-153.9187	Durbin-Watson stat	1.965834

Inverted AR Roots	.85	.11
Inverted MA Roots	.41	

Model 2

Dependent Variable: SERIES01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.854304	0.086835	9.838303	0.0000
MA(1)	-0.294178	0.135943	-2.163984	0.0329
MA(2)	-0.048162	0.121424	-0.396641	0.6925

R-squared	0.494160	Mean dependent var	-0.174202
Adjusted R-squared	0.483622	S.D. dependent var	1.638134
S.E. of regression	1.177154	Akaike info criterion	3.193911
Sum squared resid	133.0264	Schwarz criterion	3.272551
Log likelihood	-155.0986	Durbin-Watson stat	1.997229

Inverted AR Roots	.85	
Inverted MA Roots	.41	-.12

12. Write the estimated models for the following series.

Series 1

Dependent Variable: SERIES01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-78.16783	501.0114	-0.156020	0.8760
AR(1)	1.610020	0.008702	185.0209	0.0000
AR(2)	-0.616464	0.008702	-70.84448	0.0000

R-squared	0.995075	Mean dependent var	-58.50484
Adjusted R-squared	0.995074	S.D. dependent var	4162.713
S.E. of regression	292.1641	Akaike info criterion	14.19287
Sum squared resid	6.99E+08	Schwarz criterion	14.19544
Log likelihood	-58116.82	F-statistic	827096.1
Durbin-Watson stat	2.208966	Prob(F-statistic)	0.000000

Inverted AR Roots	.98	.63

Series 2

Dependent Variable: Difference(SERIES01)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.774125	0.010924	70.86718	0.0000
MA(1)	-0.265242	0.016638	-15.94151	0.0000

R-squared	0.392578	Mean dependent var	-0.305861
Adjusted R-squared	0.392504	S.D. dependent var	371.3930
S.E. of regression	289.4713	Akaike info criterion	14.17423
Sum squared resid	6.86E+08	Schwarz criterion	14.17595
Log likelihood	-58041.49	Durbin-Watson stat	1.997692

Inverted AR Roots	.77
Inverted MA Roots	.27

Series 3

Dependent Variable: SERIES01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(2)	0.988475	0.001648	599.7436	0.0000
MA(1)	0.120766	0.010562	106.1123	0.0000
MA(3)	-0.122652	0.010562	-11.61255	0.0000

R-squared	0.992215	Mean dependent var	-58.50484
Adjusted R-squared	0.992213	S.D. dependent var	4162.713
S.E. of regression	367.3250	Akaike info criterion	14.65074
Sum squared resid	1.10E+09	Schwarz criterion	14.65331
Log likelihood	-59991.77	Durbin-Watson stat	0.892310

Inverted AR Roots	.99	-.99		
Inverted MA Roots	.29	-.42	-1.00	

Series 4

Dependent Variable: SERIES01

Variable	Coefficient	Std. Error	t-Statistic	Prob.
AR(1)	0.823213	0.013572	134.3349	0.0000
AR(2)	-0.828464	0.013416	-61.75404	0.0000
MA(1)	-0.334644	0.018897	-17.70896	0.0000
MA(3)	-0.078778	0.013710	-5.746101	0.0000
MA(4)	0.025864	0.012300	2.102713	0.0355

R-squared	0.995252	Mean dependent var	-58.50484
Adjusted R-squared	0.995250	S.D. dependent var	4162.713
S.E. of regression	286.8942	Akaike info criterion	14.15671
Sum squared resid	6.74E+08	Schwarz criterion	14.16099
Log likelihood	-57966.74	Durbin-Watson stat	1.986695

Inverted AR Roots	.96	.86		
Inverted MA Roots	.44	.32	-.21-.37i	-.21+.37i

Chapter 5
Forecasting

Consider the process $\phi(B)X_t = \theta(B)Z_t$, $Z_t \approx WN(0,\sigma^2)$

Aim is to predict X_{n+l}, where l is the forecast horizon, with minimum mean squared error. Define the function

$P_n X_{n+l} = \alpha_0 + \alpha_1 X_n + \alpha_2 X_{n-1} + .. + \alpha_n X_1$ which predicts X_{n+l} with minimum mean squared error. Then,

$$\min q(\alpha_0, \alpha_1, \alpha_2, ..., \alpha_n) = q(\alpha) = E\left[X_{n+l} - P_n X_{n+l}\right]^2 \Rightarrow$$

$$\frac{\partial q(\alpha)}{\partial \alpha_j} = 0 \Rightarrow \frac{\partial q(\alpha)}{\partial \alpha_0} = 0 \Rightarrow E\left[X_{n+l}\right] = \alpha_0 + \sum_{i=1}^{n} \alpha_i E\left[X_{n+1-i}\right] \Rightarrow \hat{\alpha}_0 = \mu(1 - \sum_{i=1}^{n} \alpha_i)$$

$$\frac{\partial q(\alpha)}{\partial \alpha_j} = 0 \Rightarrow E\left[(X_{n+l} - P_n X_{n+l})X_{n+1-i}\right] = 0 \Rightarrow E\left[X_{n+l} - X_{n+1-i}\right] = \alpha_0 + \sum_{i=1}^{n} \alpha_i E\left[X_{n+l} X_{n+1-i}\right]$$

$\Rightarrow \gamma_n(l) = P_n \alpha_n$, from the system of equations, α_n can be solved simultaneously. Here,

$$P_n = \left[\gamma(i-j)\right]_{i,j=1}^{n}$$
$$\alpha_n = (\alpha_0, ...\alpha_n)$$
$$\gamma_n(l) = \left[\gamma(l)\ \gamma(l+1)\ ...\ \gamma(l+n-1)\right]$$

Therefore, the mean square prediction error is

$$E\left[X_{n+l} - P_n X_{n+l}\right]^2 = \gamma(0) - 2\sum_{i=1}^{n} \alpha_i \gamma(l+i-1) + \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j \gamma(j-i)$$
$$= \gamma(0) - \alpha_n' \gamma_n(l)$$

$X_n(l)$ provides a good approximation to $P_n X_{n+l}$. Hence,

$$X_n(l) = P_n X_{n+l} = E\left[X_{n+l} \mid X_n X_{n-1} .. X_1\right]$$

X_{n+l} can be written in its causal and invertible forms:

$$X_{n+l} = \sum_{j=0}^{\infty} \psi_j Z_{n+l-j} , \quad \psi_0 = 1$$

$$Z_{n+l} = \sum_{j=0}^{\infty} \pi_j X_{n+l-j} , \quad \pi_0 = 0$$

Then

$$X_n(l) = E\left[X_{n+l} \mid X_n X_{n-1}..X_1\right] = \sum_{j=0}^{\infty} \psi_j E\left[Z_{n+l-j} \mid X_n X_{n-1}..X_1\right]$$

Note that

$$E\left[Z_{n+j} \mid X_n X_{n-1}..X_1\right] = Z_{n+j} \quad \text{for} \quad j \le 0; \quad \text{zero otherwise} .$$

<u>The error in the forecast</u> $e_n(l)$ is:

$$e_n(l) = X_{n+l} - X_n(l)$$

$$e_n(l) = Z_{n+l} - \psi_1 Z_{n+l-1} + .. + \psi_{l-1} Z_{n+1}$$

<u>The Mean Square Prediction Error is</u>

$$Var\left[e_n(l)\right] = Var\left[X_{n+l} - X_n(l)\right]$$

$$Var\left[e_n(l)\right] = Var\left[Z_{n+l} - \psi_1 Z_{n+l-1} + .. + \psi_{l-1} Z_{n+1}\right] = \sigma^2 \sum_{j=0}^{l-1} \psi_j^2$$

<u>The auto-covariance among the prediction errors is</u>

$$\gamma(h) = E\left[(X_{n+l} - X_n(l))(X_{n+l+h} - X_{n+h}(l))\right] = \sigma^2 \sum_{j=0}^{l-1} \psi_j \psi_{j+k}$$

<u>Example</u>: Given AR(1) process with drift μ, predict X_{n+l} .

$$X_t - \mu = \phi(X_{t-1} - \mu) + Z_t$$

$$X_n(l) = E\left[\mu + \phi(X_{n+l-1} - \mu) + Z_{n+l} \mid X_n X_{n-1}..X_1\right] = \sum_{j=0}^{\infty} \phi^j E\left[Z_{n+l-j} \mid X_n X_{n-1}..X_1\right]$$

<u>For l=1</u>

$$X_n(1) = E\left[\mu + \phi(X_n - \mu) + Z_{n+1} \mid X_n X_{n-1}..X_1\right] = \mu \sum_{j=0}^{\infty} \phi^j E\left[Z_{n+1-j} \mid X_n X_{n-1}..X_1\right]$$

$$X_n(1) = \mu(1-\phi) + E\left[\phi X_n \mid X_n X_{n-1}..X_1\right] + E\left[Z_{n+1} \mid X_n X_{n-1}..X_1\right] = \mu + \phi(X_n - \mu)$$

or

$$X_n(1) = \mu + \sum_{j=0}^{\infty} \phi^j E\left[Z_{n+1-j} \mid X_n X_{n-1}..X_1\right]$$

$$X_n(1) = \mu + \phi^0 E\left[Z_{n+1} \mid X_n X_{n-1}..X_1\right] + \phi^1 E\left[Z_n \mid X_n X_{n-1}..X_1\right] + \phi^2 E\left[Z_{n-1} \mid X_n X_{n-1}..X_1\right] + ...$$

$$X_n(1) = \mu + \phi^1 Z_n + \phi^2 Z_{n-1} + \phi^3 Z_{n-2} + \phi^4 Z_{n-3}......$$

$$X_n(1) = \mu + \phi(Z_n + \phi\ Z_{n-1} + \phi^2 Z_{n-2} + \phi^3 Z_{n-3}......) = \mu + \phi \sum_{j=0}^{\infty} \phi^j Z_{n-j} = \mu + \phi(X_n - \mu)$$

.

.

<u>For *l*=k</u>

$$X_n(k) = E\left[\mu + \phi(X_{n+k} - \mu) + Z_{n+k} \mid X_n X_{n-1}..X_1\right] = \mu + \sum_{j=0}^{\infty}\phi^j E\left[Z_{n+k-j} \mid X_n X_{n-1}..X_1\right]$$

$$X_n(k) = \mu + \phi^k (X_n - \mu)$$

<u>Prediction Error and its variance</u>

<u>For *l*=1</u>

$$e_n(1) = X_{n+1} - X_n(1) = \{\mu + \phi(X_n - \mu) + Z_{n+1}\} - \{\mu + \phi(X_n - \mu)\} = Z_{n+1}$$

$$Var\left[e_n(1)\right] = \sigma^2$$

<u>For *l*=2</u>

$$e_n(2) = X_{n+2} - X_n(2) = \{\mu + \phi(X_{n+1} - \mu) + Z_{n+2}\} - \{\mu + \phi^2 (X_n - \mu)\} = Z_{n+2} + \phi Z_{n+1}$$

or

$$e_n(2) = Z_{n+1} + \psi_1 Z_n = Z_{n+1} + \phi Z_n$$

$$Var\left[e_n(2)\right] = \sigma^2 (1 + \phi^2)$$

<u>For *l*=k</u>

$$e_n(k) = X_{n+k} - X_n(k) = \{\mu + \phi(X_{n+k-1} - \mu) + Z_{n+k}\} - \{\mu + \phi^k (X_n - \mu)\}$$

or

$$e_n(k) = Z_{n+k} + \psi_1 Z_{n+k-1} + \psi_2 Z_{n+k-1} + \ldots + \psi_{k-1} Z_{n+1} = Z_{n+k} + \phi Z_{n+k-1} + \phi^2 Z_{n+k-1} + \ldots + \phi^{k-1} Z_{n+1}$$

$$Var\left[e_n(k)\right] = \sigma^2 (1 + \phi^2 + \phi^4 + \ldots. + \phi^{2(l-1)}).\frac{(1-\phi^2)}{(1-\phi^2)}$$

$$Var\left[e_n(k)\right] = \frac{\sigma^2}{(1-\phi^2)}(1 + \phi^2 + \phi^4 + \ldots. + \phi^{2(l-1)} - \phi^2 - \phi^4 - \ldots. - \phi^{2(l-1)} - \phi^{2l}) = \frac{\sigma^2}{(1-\phi^2)}(1-\phi^{2l})$$

(1-α)x100% Prediction Limits

$$X_n(l) \pm z_{\alpha/2}\left\{Var\left[e_n(l)\right]\right\}^{1/2}$$

For a 95% confidence interval for the prediction we take $z_{\alpha/2}$ rounded to 2.

5.1. Forecast Updating

As the new observations become available, the forecasts have to be updated. Suppose we are at time n and predicting (*l*+1) steps ahead. Then

$$X_n(l+1) = E\left[X_{n+l+1} \mid X_n X_{n-1} .. X_1\right] = \psi_{l+1} Z_n + \psi_{l+2} Z_{n-1} + \psi_{l+3} Z_{n-2} + \ldots$$

After (n+1)st observation become available, we update the prediction of X_{n+l+1} as

$$X_{n+1}(l) = \psi_l Z_{n+1} + \psi_{l+1} Z_n + \psi_{l+2} Z_{n-1} + \ldots = \psi_l Z_{n+1} + X_n(l+1)$$

$$X_{n+1}(l) = \psi_l \left[X_{n+1} - X_n(1)\right] + X_n(l+1)$$

Out of Sample Forecasts

To assess the forecasting performance of two proposed models, a holdback period of k is forecasted. The efficiency of the forecast for each model is performed based on the comparison of the Mean Square Prediction errors of those out of sample forecasts.

Example Yield data example. Forecast the series for $l = 1,2,3$ periods starting if n=156.

$X_{156} = 0.49; \quad \mu = 1.003$

Dependent Variable: YIELD_DATA
Method: Least Squares
Date: 01/21/08 Time: 00:14
Sample (adjusted): 2 158
Included observations: 157 after adjustments
Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1.003013	0.079014	12.69410	0.0000
AR(1)	0.841074	0.042868	19.62019	0.0000

R-squared	0.712937	Mean dependent var	0.994586
Adjusted R-squared	0.711085	S.D. dependent var	0.292556
S.E. of regression	0.157251	Akaike info criterion	-0.849287
Sum squared resid	3.832832	Schwarz criterion	-0.810354
Log likelihood	68.66904	F-statistic	384.9517
Durbin-Watson stat	1.860085	Prob(F-statistic)	0.000000

$$\hat{X}_{156}(1) = 1.003 + 0.85(0.49 - 1.003) = 0.56$$

$$\hat{X}_{156}(2) = 1.003 + 0.85^2(0.49 - 1.003) = 0.62$$

$$\hat{X}_{156}(3) = 1.003 + 0.85^3(0.49 - 1.003) = 0.68$$

$\hat{\sigma}^2 = 0.157251^2 = 0.025$

$Var[e_{156}(1)] = 0.025$

$Var[e_{156}(2)] = 0.025\left(\dfrac{1-\hat{\phi}^4}{1-\hat{\phi}^2}\right) = 0.025\left(\dfrac{1-0.85^4}{1-0.85^2}\right) = 0.041$

$Var[e_{156}(3)] = 0.025\left(\dfrac{1-\hat{\phi}^6}{1-\hat{\phi}^2}\right) = 0.025\left(\dfrac{1-0.85^6}{1-0.85^2}\right) = 0.054$

Suppose that $X_{157} = 0.7$ is observed. The Prediction updates are

<u>for l=1</u> $\hat{X}_{n+1}(1) = \psi_1 [X_{157} - \hat{X}_{156}(1)] + \hat{X}_{156}(2) = 0.85(0.74 - 0.56) + 0.62 = 0.77$

<u>for l=2</u> $\hat{X}_{n+1}(2) = \psi_1 [X_{157} - \hat{X}_{156}(1)] + \hat{X}_{156}(3) = 0.85(0.74 - 0.56) + 0.68 = 0.81$

5.2. Efficiency of Forecasting

One method of evaluating a forecasting technique uses the summation of the absolute errors. The mean absolute deviation (MAD) measure forecast accuracy by averaging the magnitudes of the forecast errors (absolute values of each error).

$$MAD = \frac{\sum_{i=1}^{n} \left| X_i - \hat{X}_i \right|}{n} = \frac{\sum |e_i|}{n}$$

The mean square prediction error (MSPE) is an alternative method for evaluating a forecasting technique. This approach provides a penalty for large forecasting errors as it squares each.

$$MSPE = \frac{\sum_{i=1}^{n} (X_i - \hat{X}_i)^2}{n} = \frac{\sum_{i=1}^{n} e_i^2}{n}$$

Mean absolute percentage error (MAPE) expresses errors in terms of percentages. This approach is useful when the size or the magnitude of the forecast variable is important in evaluating the accuracy of the forecast. MAPE provides and indication of how large the forecast errors are in comparison to the actual values in the series.. MAPE can also be used to compare the accuracy of the same different techniques on two entirely different series.

Sometimes it is necessary to determine whether a forecasting methods is biased (consistently forecasting low or high). The mean percentage error (MPE) is used in these cases. If the forecasting approach is unbiased MPE produce a percentage that is close to zero. If the result is a large negative percentage, the forecasting method is consistently overestimating. If the result is a large positive percentage, the forecasting method is consistently underestimating.

Example: Suppose for the Yield data analyzed previously the following statistics calculated from residuals
MAD=1.3, MSPE= 13.5, MAPE=6.95%, MPE=2.03%,

MAD indicates that each forecast deviated by an average of 1.3 amount. The MSE and MAPE would be compared to the MSE and MAPE of an alternative model. The one which yields the minimum would be preferred model. Finally, the small MPE 2.03% indicates that the technique is not biased.

To determine statistically if the MSPE of two models are different from the other, we use the statistic

$$F = \frac{\sum_{i=1}^{n} e_{1i}^2}{\sum_{i=1}^{n} e_{2i}^2} \approx F_{n,n}$$ with the assumptions

i. The forecast errors have zero mean and are normally distributed
ii. The forecast errors are serially uncorrelated
iii. The forecast errors are contemporaneously uncorrelated with each other

These assumptions may not be realized for the series as the multi-step forecasts produce serially correlated values. This leads the assumption on the distribution of the proportion of sum squared errors fails.
In order to overcome the contemporaneously correlated forecast errors, Granger-Newbold Test is used.

The Granger-Newbold Test: The null hypothesis claims that the forecast accuracy of linear combinations of the residuals are uncorrelated.

Let $Y_t = e_{1t} + e_{2t}$ and $W_t = e_{1t} - e_{2t}$ and $\rho_{YW} = E(e_{1t}^2 - e_{2t}^2)$

The model 1 has larger MSPE if ρ_{YW} is positive and model 2 has a larger MSPE

otherwise. Given the sample correlation coefficient $\hat{\rho}_{YW}$,

$$\frac{\hat{\rho}_{YW}}{\sqrt{(1-\hat{\rho}_{YW}^2)/n-1}} \approx t_{n-1}$$

.

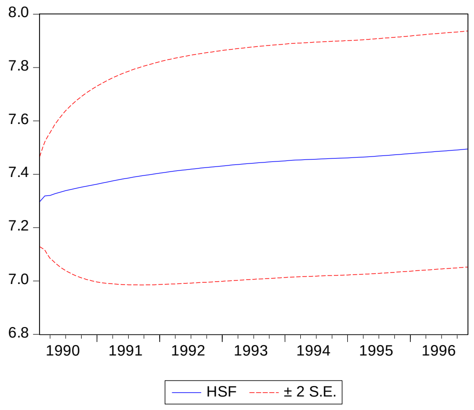
Let $d_i = g(e_{1i}) - g(e_{2i})$, g).= being any function of the residuals and

Theil`s Inequality Coefficient (TIC)

Variation in forecast lies between zero and one. A value of TIC being zero indicates a perfect fit. TIC and MAPE are scale invariant statistics.
Variation in the forecast can be decomposed into three parts: Bias, Variance and Covariance whose proportions relative to the variance sums up to one.
The bias proportion tells us how far the mean of the forecast is from the mean of the actual series. The variance proportion is the variation among the variances of forecast and actual series and covariance proportion measures the remaining unsystematic forecasting errors.

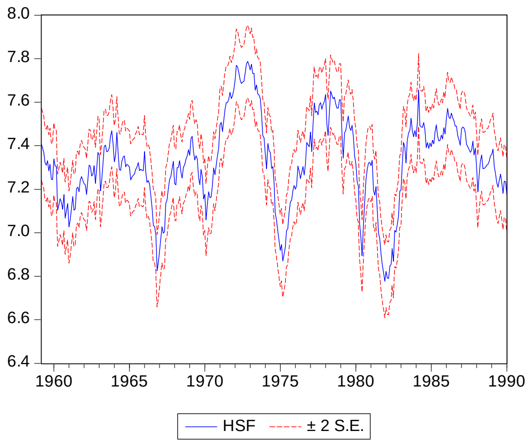
$$TIC = \frac{\sqrt{\frac{1}{h+1}\sum_{t=n+1}^{n+h} X_t - \hat{X}_t{}^{2}}}{\sqrt{\frac{1}{h+1}\sum_{t=n+1}^{n+h} X_t{}^{2}}\sqrt{\frac{1}{h+1}\sum_{t=n+1}^{n+h} \hat{X}_t{}^{2}}}$$

Example:



Forecast: HSF
Actual: HS
Forecast sample: 1990M02 1996M12
Included observations: 72

Root Mean Squared Error	0.318700
Mean Absolute Error	0.297261
Mean Abs. Percent Error	4.205889
Theil Inequality Coefficient	0.021917
Bias Proportion	0.869982
Variance Proportion	0.082804
Covariance Proportion	0.047214



Forecast: HSF
Actual: HS
Forecast sample: 1959M01 1990M01
Adjusted sample: 1959M03 1990M01
Included observations: 371

Root Mean Squared Error	0.082172
Mean Absolute Error	0.062917
Mean Abs. Percent Error	0.864011
Theil Inequality Coefficient	0.005607
Bias Proportion	0.000000
Variance Proportion	0.037292
Covariance Proportion	0.962708

Example: The series contains the logarithm of monthly housing starts (HS) over the period 1959M01-1996M01, logarithm of the S&P index (SP) from 1959M01-1996M12.
Estimation: HS on C, SP, lag of HS with an AR (1) using data from 1959M01-1990M01

Dependent Variable: HS
Method: Least Squares
Date: 10/19/97 Time: 21:59
Sample(adjusted): 1959:03 1990:01
Included observations: 371 after adjusting endpoints
Convergence achieved after 4 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.321924	0.117278	2.744975	0.0063
HS(-1)	0.952653	0.016218	58.74157	0.0000
SP	0.005222	0.007588	0.688249	0.4917
AR(1)	-0.271254	0.052114	-5.205027	0.0000

R-squared	0.861373	Mean dependent var	7.324051
Adjusted R-squared	0.860240	S.D. dependent var	0.220996
S.E. of regression	0.082618	Akaike info criterion	-2.138453
Sum squared resid	2.505050	Schwarz criterion	-2.096230
Log likelihood	400.6830	Hannan-Quinn criter.	2.013460
F-statistic	0.000000		

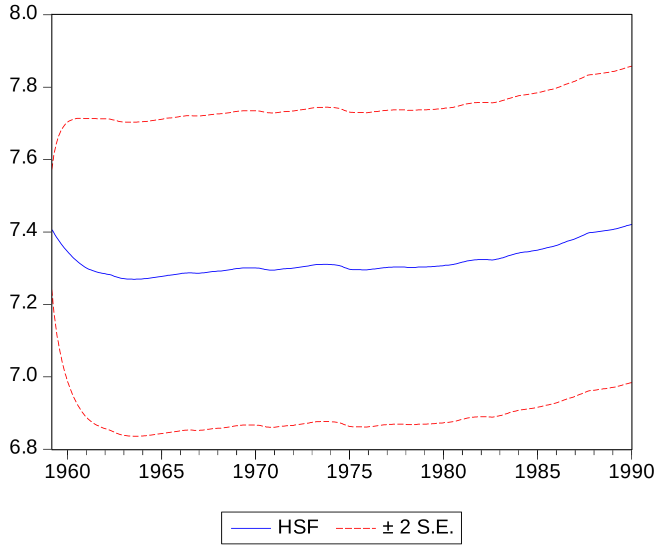
Inverted AR Roots	-.27

The fitted model is

HS = 0.321923974124 + 0.952652670606*HS(-1) + 0.00522248778209*SP + [AR(1)=-0.271254015478]

1. **Dynamic forecast**: dynamic, multi-step forecasts starting from the 1st period in the forecast sample. Previously forecasted values for the lagged dependent variables are used in forming forecasts of the current value. This choice will only be available when the estimated equation contains dynamic components, e.g. lagged dependent variables or ARMA terms.

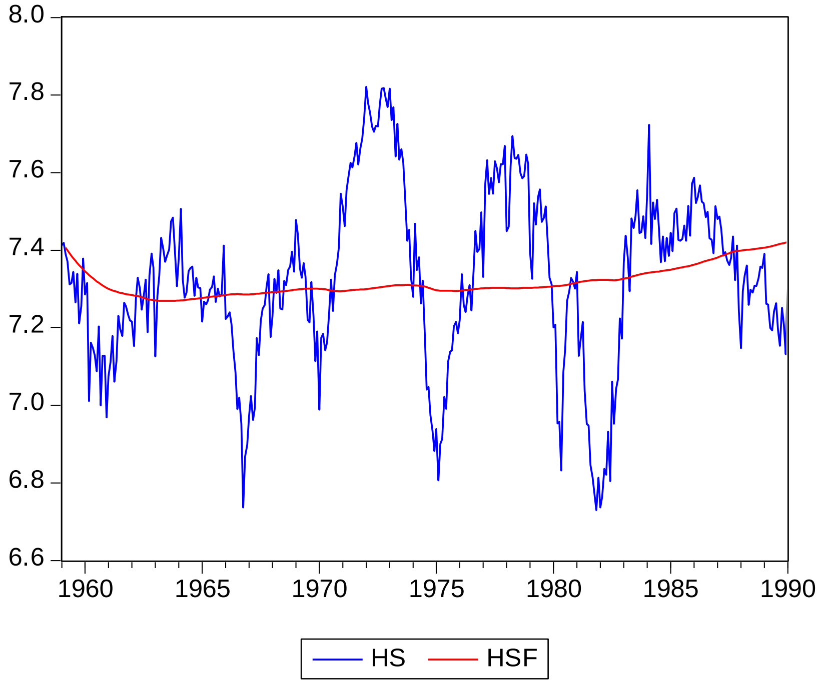
For the example we estimate the equation using data from 1959:01 to 1990:01.
In sample dynamic forecast: from **1959:01 to 1990:01**



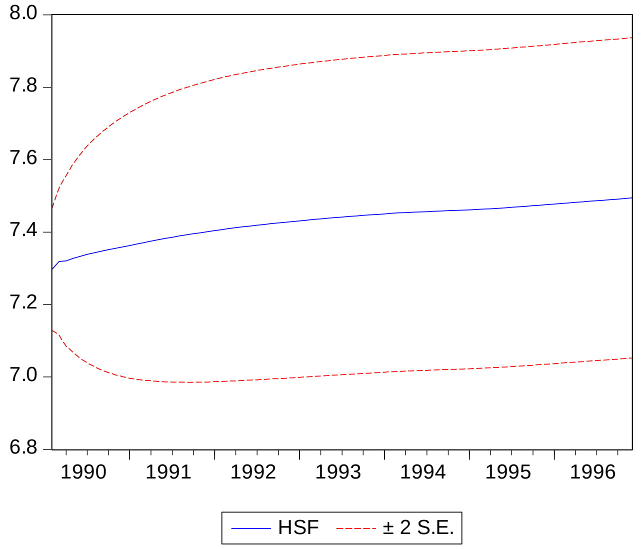
Forecast: HSF
Actual: HS
Forecast sample: 1959M01 1990M01
Adjusted sample: 1959M03 1990M01
Included observations: 371

Root Mean Squared Error	0.221735
Mean Absolute Error	0.170050
Mean Abs. Percent Error	2.328604
Theil Inequality Coefficient	0.015141
Bias Proportion	0.000848
Variance Proportion	0.678102
Covariance Proportion	0.321050

Sample adjustment: because we introduce the 1st lag and AR term in the residuals, we can only estimate from 1959M03. The loss of 2 observations occurs because the residual loses one observation due to the lagged endogenous variable so that the forecast for the error term can begin only from the 3rd observation.

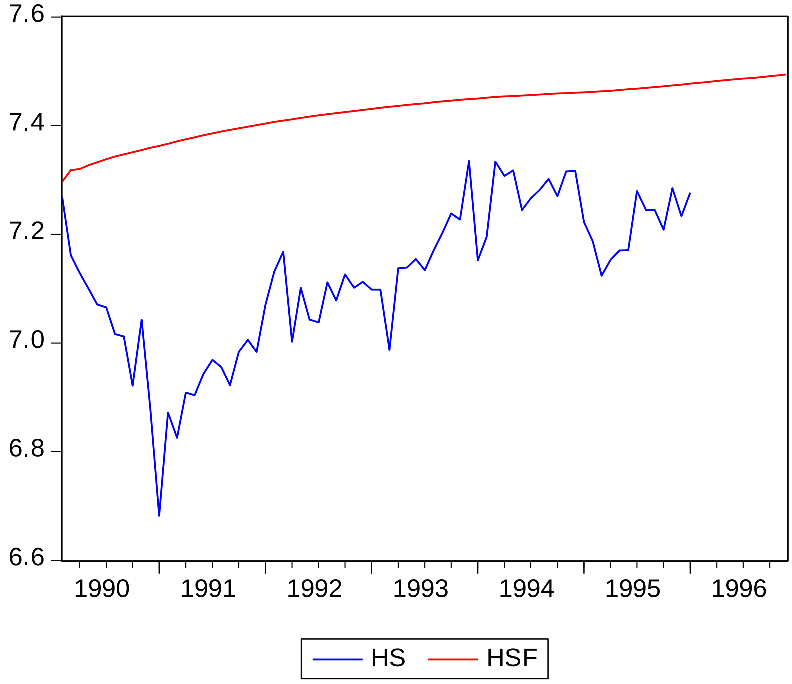


2. **Out of sample dynamic forecast**: from **1990:02 to 1996:12**



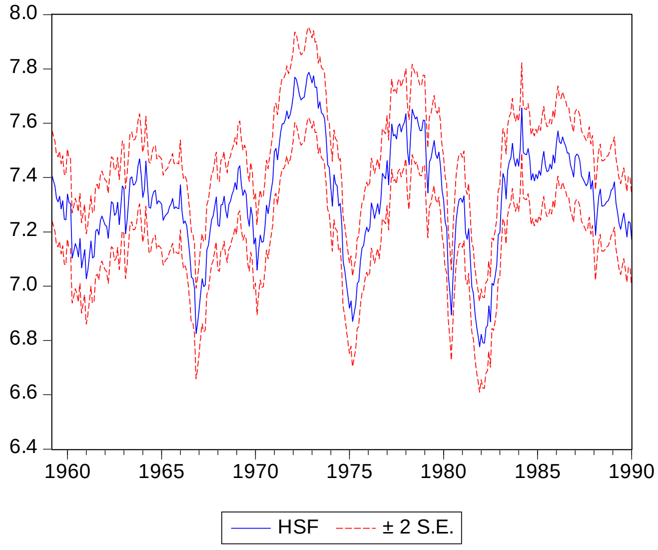
Forecast: HSF
Actual: HS
Forecast sample: 1990M02 1996M12
Included observations: 72

Root Mean Squared Error	0.318700
Mean Absolute Error	0.297261
Mean Abs. Percent Error	4.205889
Theil Inequality Coefficient	0.021917
Bias Proportion	0.869982
Variance Proportion	0.082804
Covariance Proportion	0.047214



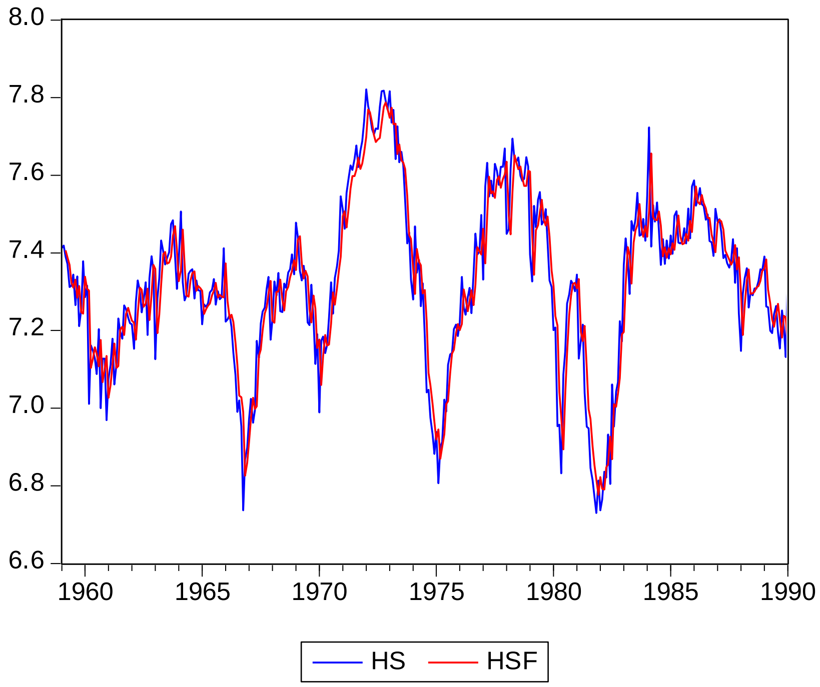
3. **Static forecast**: one-step-ahead forecasts, using the actual, rather than forecasted values for lagged dependent variables.

3.1 In sample static forecast:

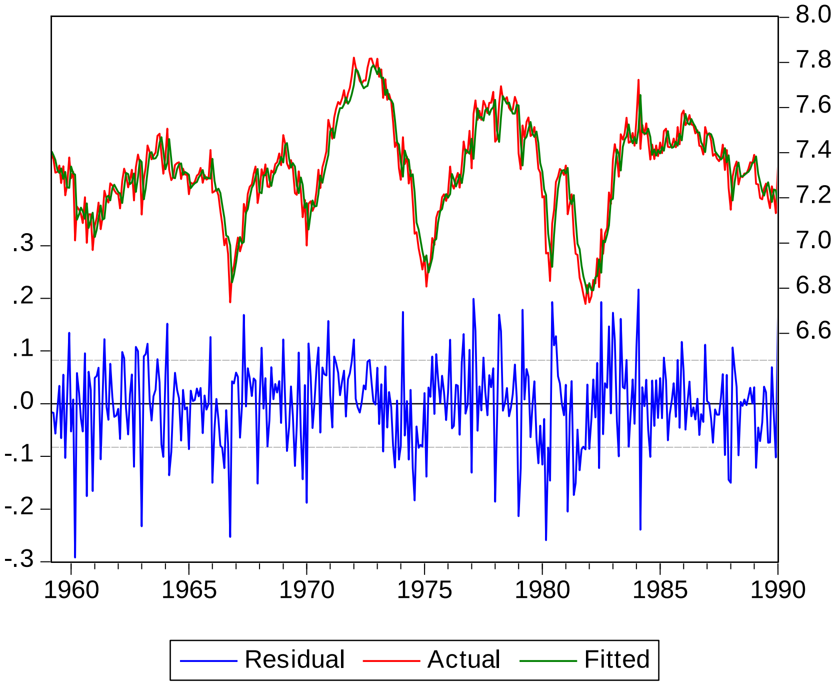


Forecast: HSF
Actual: HS
Forecast sample: 1959M01 1990M01
Adjusted sample: 1959M03 1990M01
Included observations: 371

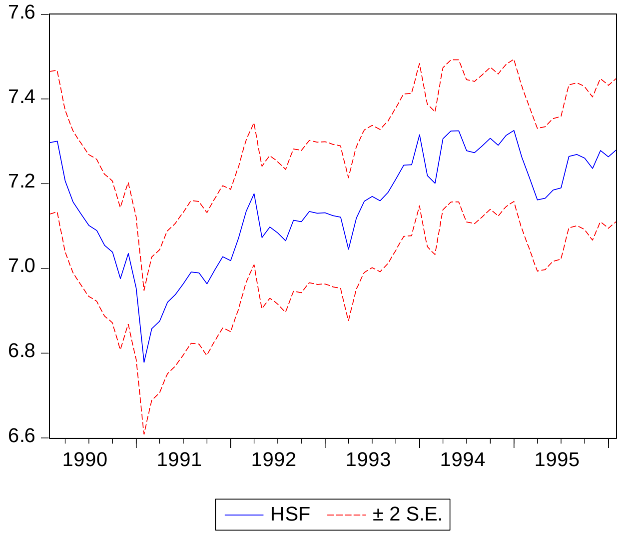
Root Mean Squared Error	0.082172
Mean Absolute Error	0.062917
Mean Abs. Percent Error	0.864011
Theil Inequality Coefficient	0.005607
Bias Proportion	0.000000
Variance Proportion	0.037292
Covariance Proportion	0.962708



The one-step ahead static forecasts are more accurate than the dynamic forecasts since the actual value of the lagged dependent variable is used in forming the forecast of HS. These **one-step ahead static forecasts are the same forecasts used in the Actual, Fitted, Residual Graph** displayed for the equation estimation below, i.e. the fitted value by estimation.

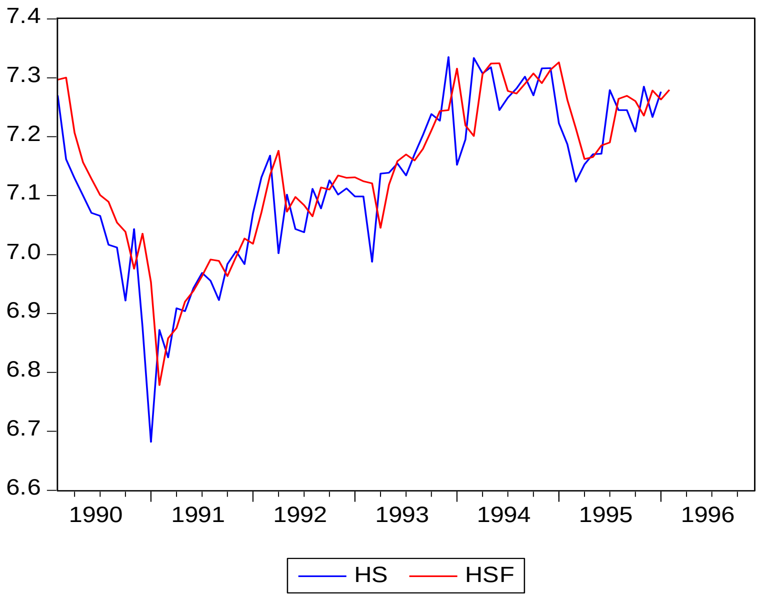


3.2. Out of sample static forecast:



Forecast: HSF
Actual: HS
Forecast sample: 1990M02 1996M12
Adjusted sample: 1990M02 1996M02
Included observations: 72

Root Mean Squared Error	0.070691
Mean Absolute Error	0.051155
Mean Abs. Percent Error	0.723547
Theil Inequality Coefficient	0.004955
Bias Proportion	0.091461
Variance Proportion	0.022643
Covariance Proportion	0.885896



Sample adjustment: For static forecast, we are responsible for the supply of the actual value for the lagged dependent variables. Since we only have data for HS until 1996M01, we can only do static forecast until 1996M02.

Exercises

1. Consider the model $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + Z_t + \theta Z_{t-1}$; $\quad Z_t \sim WN$

a. Obtain the l-step forecasts, l=1,2 recursively
b. Obtain the forecast errors $e_n(l)$, l=1,2 and find their variances.

2. For the stationary and invertible ARMA(2,2) model with zero mean,
a. Obtain recursively a. the first three forecasts
b. The forecast errors and their variances

Chapter 6
Model Selection and Non-Stationary Time Series Model

Determination of the best suitable model for given observed series and choosing the appropriate model on order of p and q are very important as the forecasting and prediction will rely on the model chosen. ACF and PACF show specific properties for specific models. Hence, they can be used as a criteria to identify the suitable model. With messy data sample ACF and PACF plots become complicated and harder to interpret. The ideal is to choose the model having few parameters as possible. It will be seen that many different models can fit to the same data so that we should choose the most appropriate one. Box Jenkins approach gives a systematic algorithm to determine the best model.

6.1. The Box-Jenkins Approach

Box and Jenkins (1976) suggest a three-stage approach to pure time series modeling. These are identification, estimation and diagnostic checking.

At the identification stage, a tentative ARIMA model is specified that may approximate the data generating process for the given sample, through examination of the correlogram and the partial autocorrelation functions. Once a model has been tentatively identified, the next stage is to estimate its parameters. Once the tentative model has been estimated, a set of estimated residuals are automatically generated.

For example for an AR(1) model the estimated residuals are $\hat{Z}_t = X_t - \psi X_{t-1}$. If the fitted model is correct, then this residual series should be approximately white noise. One the test of adequacy of the model thus includes testing for the whiteness of the fitted residuals using diagnostic checks such as the Box-Pierce or Ljung-Box portmanteau statistics.

If the estimated parameters of the fitted model are significantly different from zero and the fitted residuals appear to be approximate white noise, then the fitted model may be held to be adequate. If the model fails on either of these counts, then the identification stage should be returned to.

6.2. Testing the Dynamic Modeling (Diagnostic Checking)

1. Residual Analysis: As described above, residuals should have a random pattern and be modeled as white noise. If the Normal P-P plot of the residuals follow a normal distribution, the series is called a Gaussian process.
2. Overfitting : Having identified what is believed to be the correct model, we actually fit a more elaborate one. To conclude which model explains the series better we Akaike's Information Criterion (AIC) and Schwarz's Information Criterion (SIC) are compared for each model. The model having smaller value of AIC or SIC proposes a better fit.

Akaike's Information Criterion is

$$AIC = \ln \hat{\sigma}_k^2 + \frac{n+2k}{n}$$

where $\hat{\sigma}_k^2$ is the sample variance, k is the number of the parameters in the models and n is the number of observations. The value of k yielding the minimum AIC specifies the best model. Corrected AIC (AICc) is a modified AIC for eliminating the bias.

$$AICc = \ln \hat{\sigma}_k^2 + \frac{n+k}{n-k-2}$$

Schwarz's Information Criterion (SIC) (Bayesian Information Criterion (BIC))

$$SIC = \ln \hat{\sigma}_k^2 + \frac{k \ln n}{n}$$

SIC does well getting the correct order in large samples, whereas AIC tends to be superior in small samples where the relative number of parameters is large.

6.3. Non-Stationary Processes

Many time series like stock prices behave as through they have no fixed mean. Even so, they exhibit homogeneity in the sense that apart from local level and/or trend, one part of the series behave like any other part. Models that describe such homogeneous nonstationary behavior can be obtained by supposing some suitable difference of the process to be stationary. These models are called Autoregressive Integrated Moving Average (ARIMA) processes.

Differencing operator, Δ **:**

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

$$\Delta^2 X_t = \Delta(\Delta X_t) = X_t - X_{t-2}$$

$$= \Delta(X_t - X_{t-1}) = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2})$$

$$= (X_t - 2X_{t-1} + X_{t-2}) = (1 - 2B + B^2)X_t = (1-B)^2 X_t$$

For d differencing

$$\Delta^d X_t = (1-B)^d X_t$$

Integrated Models

Definition: $X_t \quad {}_{t \in N}$ is said to be ARIMA(p,d,q) if $\nabla^d X_t = (1-B)^d X_t$ is ARMA(p,q).

In another words, $\Phi(B)(1-B)^d X_t = \Theta(B) Z_t$

Then ARIMA(p,d,q) is an ARMA(p,q) series differenced d times.

ARIMA(1,1,1)

$$X_t - X_{t-1} = \psi(X_{t-1} - X_{t-2}) + Z_t + \theta Z_{t-1}$$
$$(1-B)X_t = \psi(B - B^2)X_t + Z_t + \theta Z_{t-1}$$
$$((1-B) - \psi(B - B^2))X_t = Z_t + \theta Z_{t-1}$$
$$(1-\psi B)(1-B)X_t = Z_t + \theta Z_{t-1}$$

ARIMA (p,d,q)

$$(1-\psi_1 B - \psi_2 B^2 - \ldots - \psi_p B^p)(1-B)^d X_t = \sum_{i=0}^{q} \theta_i Z_{t-i}$$

Example:

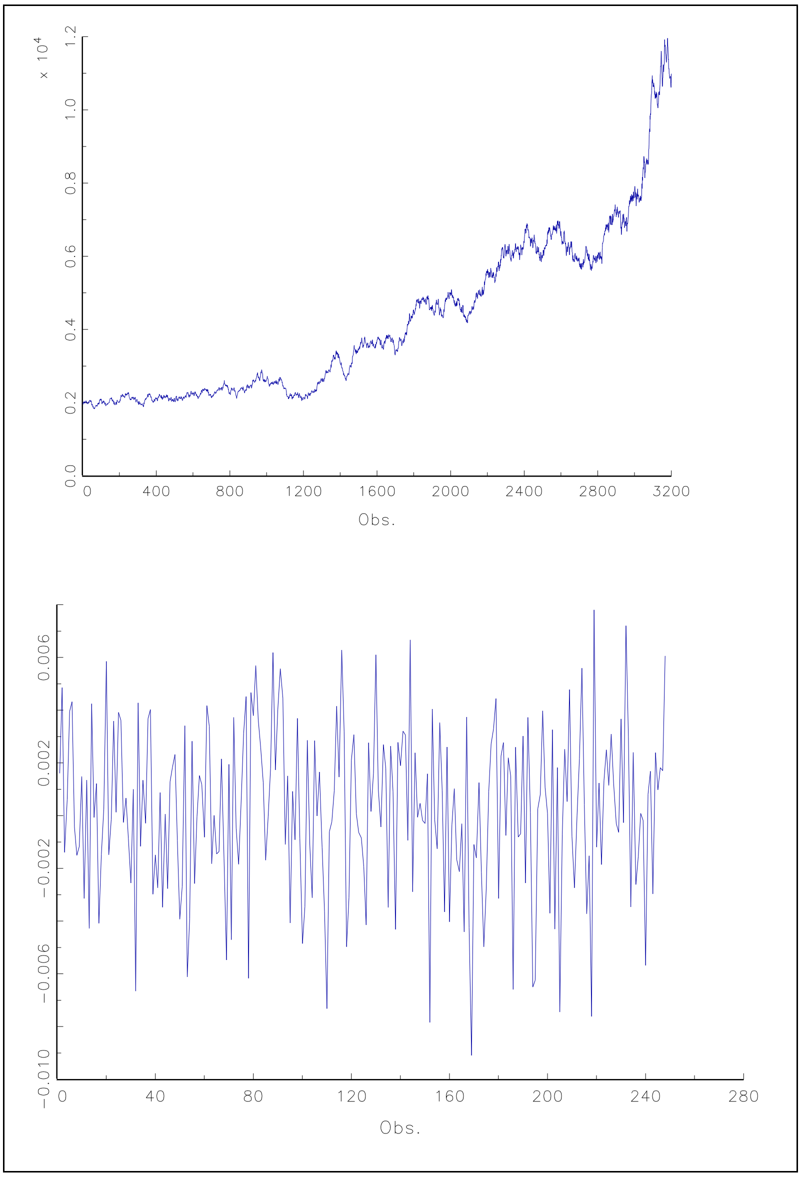


Figure 6.1. A series with trend on the left and the differenced data on the right

Example: Let $\{X_t\}_{t \in N}$ be ARIMA(1,1,1). Then

$$(X_t - X_{t-1}) = \phi(X_{t-1} - X_{t-2}) + Z_t + \theta Z_{t-1}$$
$$\Phi(B)(1-B)X_t = \Theta(B)Z_t$$
$$\Phi(B)(1-B) = (1 - B - \phi B - \phi B^2)$$
$$\Theta(B) = (1+\theta B)$$

ARIMA(1,1,1) in causal form:

Let $X_t = \Psi(B)Z_t$

$$\Phi(B)(1-B)\Psi(B) = \Theta(B)$$
$$(1 - B - \phi B - \phi B^2)(1 + \psi_1 B + \psi_2 B^2 + \ldots) = (1+\theta B)$$
$$B: \quad \psi_1 - 1 - \phi = \theta \Rightarrow \psi_1 = (1+\phi) + \theta$$
$$B^2: \quad \psi_2 - \psi_1 - \phi\psi_1 + \phi = 0 \Rightarrow \psi_2 = \psi_1(1+\phi) + \phi$$
$$B^3: \quad \psi_3 - \psi_1 - \phi\psi_2 + \phi\psi_1 = 0 \Rightarrow \psi_3 = \psi_1(1-\phi) + \phi\psi_2$$

....

ARIMA(1,1,1) in invertible form:

Let $Z_t = \Pi(B)X_t$

$$\Phi(B)(1-B) = \Theta(B)\Pi(B)$$
$$(1 - B - \phi B - \phi B^2) = (1+\theta B)(1 + \pi_1 B + \pi_2 B^2 + \ldots)$$
$$B: \quad -1 - \phi = \pi_1 + \theta \Rightarrow \pi_1 = -(1 + \phi + \theta)$$
$$B^2: \quad \phi = \pi_2 + \pi_1\theta \Rightarrow \pi_2 = \phi - \pi_1\theta$$
$$B^3: \quad \pi_3 + \theta\pi_2 = 0 \Rightarrow \pi_3 = -\theta\pi_2$$

....

Example: Find the l-step ahead forecast for an ARIMA (1,1,1) process.

$$X_t = (1+\phi)X_{t-1} - \phi X_{t-2} + Z_t + \theta Z_{t-1}$$

l-step ahead forecasts:

$$E\left[X_{n+l} \,|\, X_n, X_{n-1}, ..X_1\right] = E\left[(1+\phi)X_{n+l-1} - \phi X_{n+l-2} + Z_{n+l} + \theta Z_{n+l-1}) \,|\, X_n, X_{n-1}, ..X_1\right]$$

For *l*=1

$$X_n(1) = E\left[X_{n+1} \middle| X_n, X_{n-1}, ..X_1\right] = E\left[((1+\phi)X_{n+1-1} - \phi X_{n+1-2} + Z_{n+1} + \theta Z_{n+1-1}) \middle| X_n, X_{n-1}, ..X_1\right]$$

$$X_n(1) = (1+\phi)E\left[X_n \middle| X_n, X_{n-1}, ..X_1\right] - \phi E\left[X_{n+1} \middle| X_n, X_{n-1}, ..X_1\right]$$

$$+ E\left[Z_{n+1} \middle| X_n, X_{n-1}, ..X_1\right] + \theta E\left[Z_n \middle| X_n, X_{n-1}, ..X_1\right]$$

$$X_n(1) = (1+\phi)X_n - \phi X_{n-1} + \theta Z_n$$

For l=2

$$X_n(2) = E\left[X_{n+2} \middle| X_n, X_{n-1}, ..X_1\right] = E\left[((1+\phi)X_{n+2-1} - \phi X_{n+2-2} + Z_{n+2} + \theta Z_{n+2-1}) \middle| X_n, X_{n-1}, ..X_1\right]$$

$$X_n(2) = (1+\phi)E\left[X_{n+1} \middle| X_n, X_{n-1}, ..X_1\right] - \phi E\left[X_n \middle| X_n, X_{n-1}, ..X_1\right]$$

$$+ E\left[Z_{n+2} \middle| X_n, X_{n-1}, ..X_1\right] + \theta E\left[Z_{n+1} \middle| X_n, X_{n-1}, ..X_1\right]$$

$$X_n(2) = (1+\phi)X_n(1) - \phi X_n$$

For l=k

$$X_n(k) = E\left[X_{n+k} \middle| X_n, X_{n-1}, ..X_1\right]$$

$$X_n(k) = (1+\phi)X_n(k-1) - \phi X_n(k-2)$$

Example: Given ARIMA(1,1,0) process, the model is fitted to the past 50 observations and it is found that $\hat{\phi} = 0.40, \hat{\sigma} = 0.18$. Last two observations are $X_{49} = 33.4, X_{50} = 33.9$

a. Calculate the minimum mean squared forecasts and 95% prediction intervals for the next 5 periods.
b. A new observation $X_{51} = 34.2$ is observed. Update the forecasts.

a. $X_t = (1+\phi)X_{t-1} - \phi X_{t-2} + Z_t$

For l=1

$$X_n(1) = (1+\phi)E\left[X_n \middle| X_n, X_{n-1}, ..X_1\right] - \phi E\left[X_{n+1} \middle| X_n, X_{n-1}, ..X_1\right]$$

$$+ E\left[Z_{n+1} \middle| X_n, X_{n-1}, ..X_1\right] = (1+\phi)X_n - \phi X_{n-1}$$

For l=2

$$X_n(2) = (1+\phi)E\left[X_{n+1} \middle| X_n, X_{n-1}, ..X_1\right] - \phi E\left[X_n \middle| X_n, X_{n-1}, ..X_1\right]$$

$$+ E\left[Z_{n+2} \middle| X_n, X_{n-1}, ..X_1\right] = (1+\phi)X_n(1) - \phi X_n$$

For l=3,4,5,….

$$X_n(k) = E\left[X_{n+k} \middle| X_n, X_{n-1}, ..X_1\right]$$

$$X_n(k) = (1+\phi)X_n(k-1) - \phi X_n(k-2)$$

Estimates are: For l=1,2,3,4,5

$$\hat{X}_n(1) = 1.4(33.9) - 0.4(33.4) = 34.1$$
$$\hat{X}_n(2) = 1.4(34.1) - 0.4(33.9) = 34.18$$
$$\hat{X}_n(3) = 1.4(34.18) - 0.4(34.1) = 34.212$$
$$\hat{X}_n(4) = 1.4(34.212) - 0.4(34.18) = 34.2248$$
$$\hat{X}_n(5) = 1.4(34.2248) - 0.4(34.212) = 34.22992$$

The ψ coefficients are

$$(1 - B - \phi B - \phi B^2)(1 + \psi_1 B + \psi_2 B^2 + \ldots) = 1$$
$$B: \quad \psi_1 - 1 - \phi = 0 \Rightarrow \psi_1 = (1+\phi)$$
$$B^2: \quad \psi_2 - \psi_1 - \phi\psi_1 + \phi = 0 \Rightarrow \psi_2 = \psi_1(1+\phi) + \phi$$
$$B^k: \quad \psi_k = \psi_{k-1}(1+\phi) - \phi\psi_{k-2}$$

$$\hat{\phi} = 0.40, \hat{\sigma} = 0.18 \Rightarrow \psi_1 = 1.4, \psi_2 = 1.56, \psi_3 = 1.624, \psi_4 = 1.6496$$

95% prediction interval for l-step ahead forecast is:

$$\hat{X}_n(l) \pm (1.96\sqrt{Var(e_n(l))})$$
$$\hat{X}_n(l) \pm (1.96\hat{\sigma}\sqrt{1 + \psi_1^2 + \psi_2^2 + .. + \psi_{l-1}^2})$$

$$\hat{X}_n(1) \pm 1.96\hat{\sigma} \Rightarrow 34.1 \pm 0.3528$$
$$\hat{X}_n(2) \pm 1.96\hat{\sigma}\sqrt{1+\psi_1^2} \Rightarrow 34.18 \pm 0.6070$$
$$\hat{X}_n(3) \pm 1.96\hat{\sigma}\sqrt{1+\psi_1^2+\psi_2^2} \Rightarrow 34.212 \pm 0.8193$$
$$\hat{X}_n(4) \pm 1.96\hat{\sigma}\sqrt{1+\psi_1^2+\psi_2^2+\psi_3^2} \Rightarrow 34.2248 \pm 0.9998$$
$$\hat{X}_n(5) \pm 1.96\hat{\sigma}\sqrt{1+\psi_1^2+\psi_2^2+\psi_3^2+\psi_4^2} \Rightarrow 34.22992 \pm 1.1568$$

b. Forecast Updates are $\hat{X}_{n+1}(l) = \hat{X}_n(l+1) + \psi_l \left[X_{n+1} - \hat{X}_n(1) \right]$

$$\hat{X}_{51}(1) = \hat{X}_{50}(2) + \psi_1 \left[X_{51} - \hat{X}_{50}(1) \right] = 34.32$$
$$\hat{X}_{51}(2) = \hat{X}_{50}(3) + \psi_2 \left[X_{51} - \hat{X}_{50}(1) \right] = 34.368$$
$$\hat{X}_{51}(3) = \hat{X}_{50}(4) + \psi_3 \left[X_{51} - \hat{X}_{50}(1) \right] = 34.3904$$
$$\hat{X}_{51}(4) = \hat{X}_{50}(5) + \psi_4 \left[X_{51} - \hat{X}_n(1) \right] = 34.3949$$

Example:

Based on date of quarterly U.S. GNP from 1947(1) to 2002(3) with 223 observations, the data represent Real U.S. Gross National Product in billions of chained 1996 dollars and they have been seasonally adjusted. The data were obtained from the Federal Reserve Bank of St. Louis.

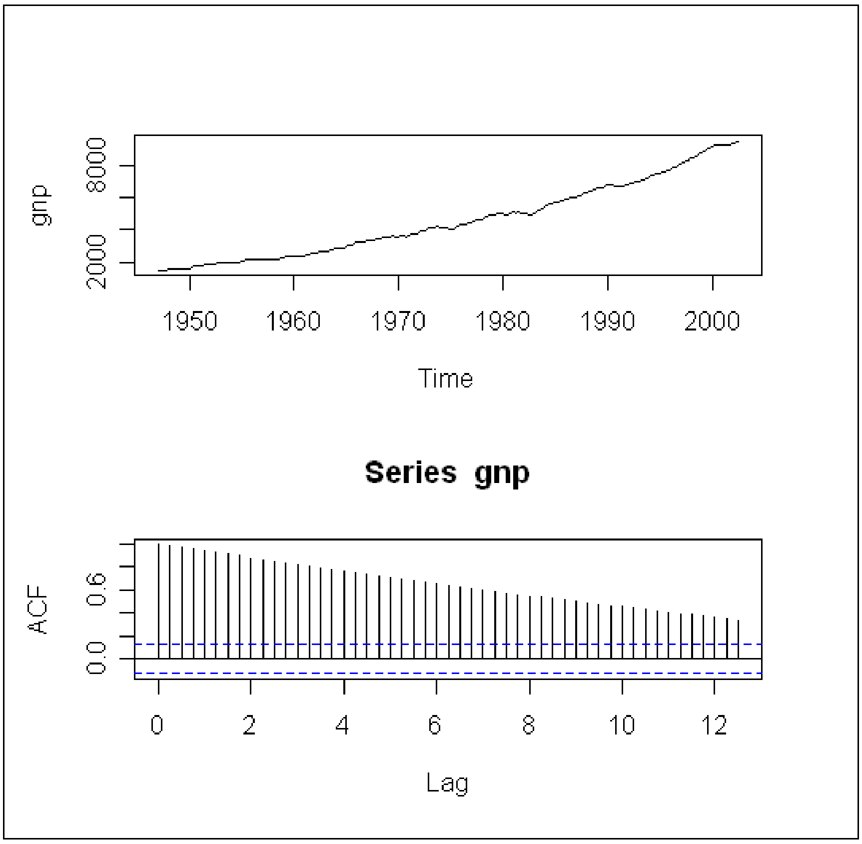


Figure 6.2 Plots of original series and its ACF

Regarding the plot of the original data and the relative ACF, it is not clear from the upper graph that the variance is increasing with time because that the strong trend hides any other effect.

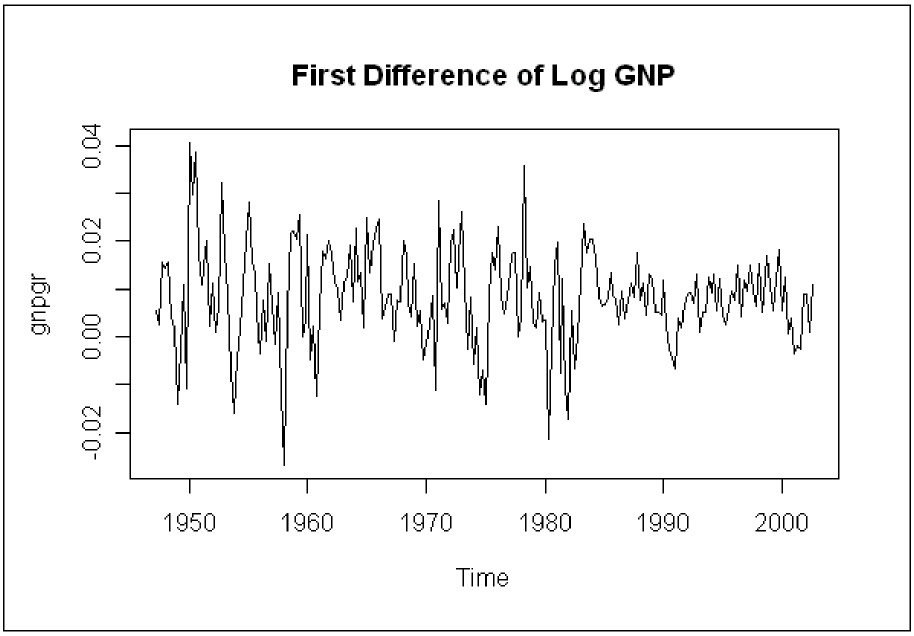


Figure 6.3. Plot of differenced data

For the purpose of the demonstration, the first difference of the logged data is displayed. Now the trend has been removed we are able to notice that the variability in the second half of the data is larger than in the first half of the data. Also, it appears as though a trend is still present after differencing.

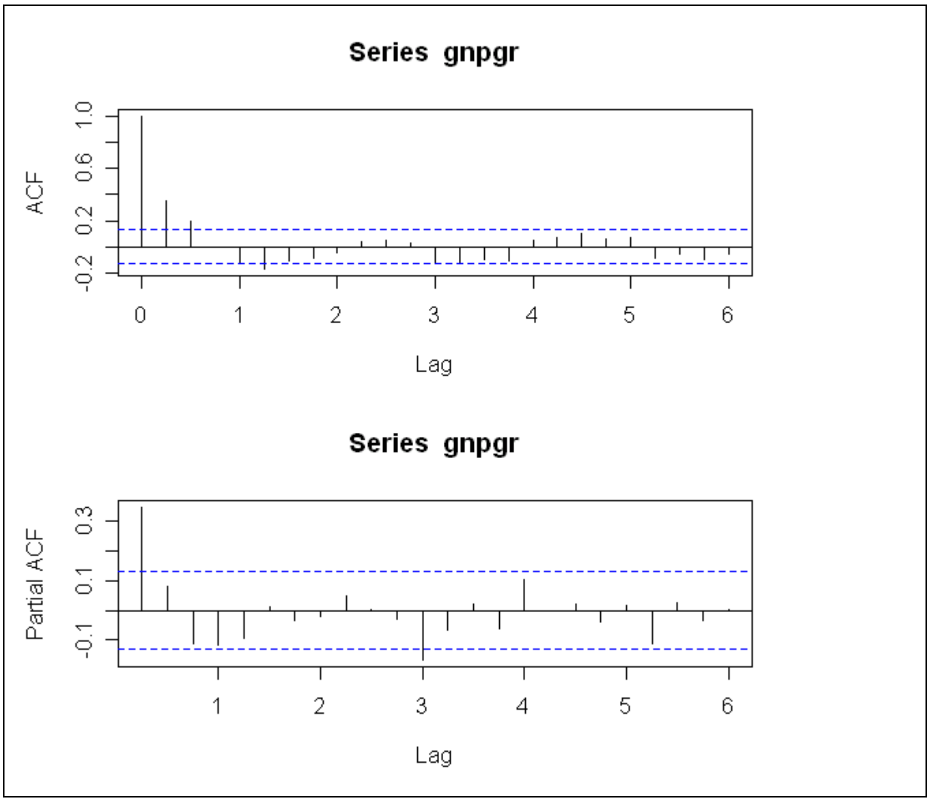


Figure 6.4 ACF of differenced data

The sample ACF and PACF of the quarterly growth rate are plotted in the upper figure. Inspecting the sample ACF and PACF, we might feel that the ACF is cutting off at lag 3 and the PACF is tailing off. This would suggest the GNP growth rate follows an MA(3) process, or log GNP follows an ARIMA(0, 1,3) model. Rather than focus on one model, we will also suggest that it appears that the ACF is tailing off and the PACF is cutting off at lag 1. This suggests an AR(1) model for the growth rate, or ARIMA(1, 1, 0) for log GNP. As a preliminary analysis, we will fit both models.

Using MLE to fit the **MA (3)** model for the growth rate (the first difference of log GNP), gnpgr, the estimated output is:

Coefficients:

	ma1	ma2	ma3	intercept
	0.3208	0.2478	0.0909	0.0083
s.e.	0.0662	0.0718	0.0701	0.0010

The variance is estimated as 8.853e-05: log likelihood = 720.78, aic = -1431.55

The estimated model MA(3)is

$$\hat{Y}_t = .008_{(.001)} + .321_{(.066)} Z_{t-1} + .248_{(.072)} Z_{t-2} + .091_{(.070)} Z_{t-3} + Z_t$$

where $\hat{\sigma}_w$ = *8.853e-05* is based on 218 degrees of freedom. The values in parentheses are the corresponding estimated standard errors. All of the regression coefficients are significant, including the constant.

Using MLE to fit the **AR (1)** model for the growth rate (the first difference of log GNP), gnpgr, the estimated output is:

Coefficients:

	ar1	intercept
	0.3467	0.0083
s.e.	0.0627	0.0010

Variance estimated as 9.03e-05: log likelihood = 718.61, aic = -1431.22
The estimated model AR(1) is:

$$\hat{Y}_t = .008_{(.001)} + .347_{(.063)} Y_{t-1} + Z_t$$

Plots of residuals for the model MA (3) in Figure 6.5 show that standardized residuals follow no obvious patterns. Notice that there are outliers, however, with a few values exceeding 3 standard deviations in magnitude. The ACF of the standardized residuals shows no apparent departure from the model assumptions, and the Q-statistic is never significant at the lags shown.

The upper figure shows a histogram of the residuals (top), and a normal Q-Q plot of the residuals (bottom). Here we see the residuals are somewhat close to normality except for a few extreme values in the tails.

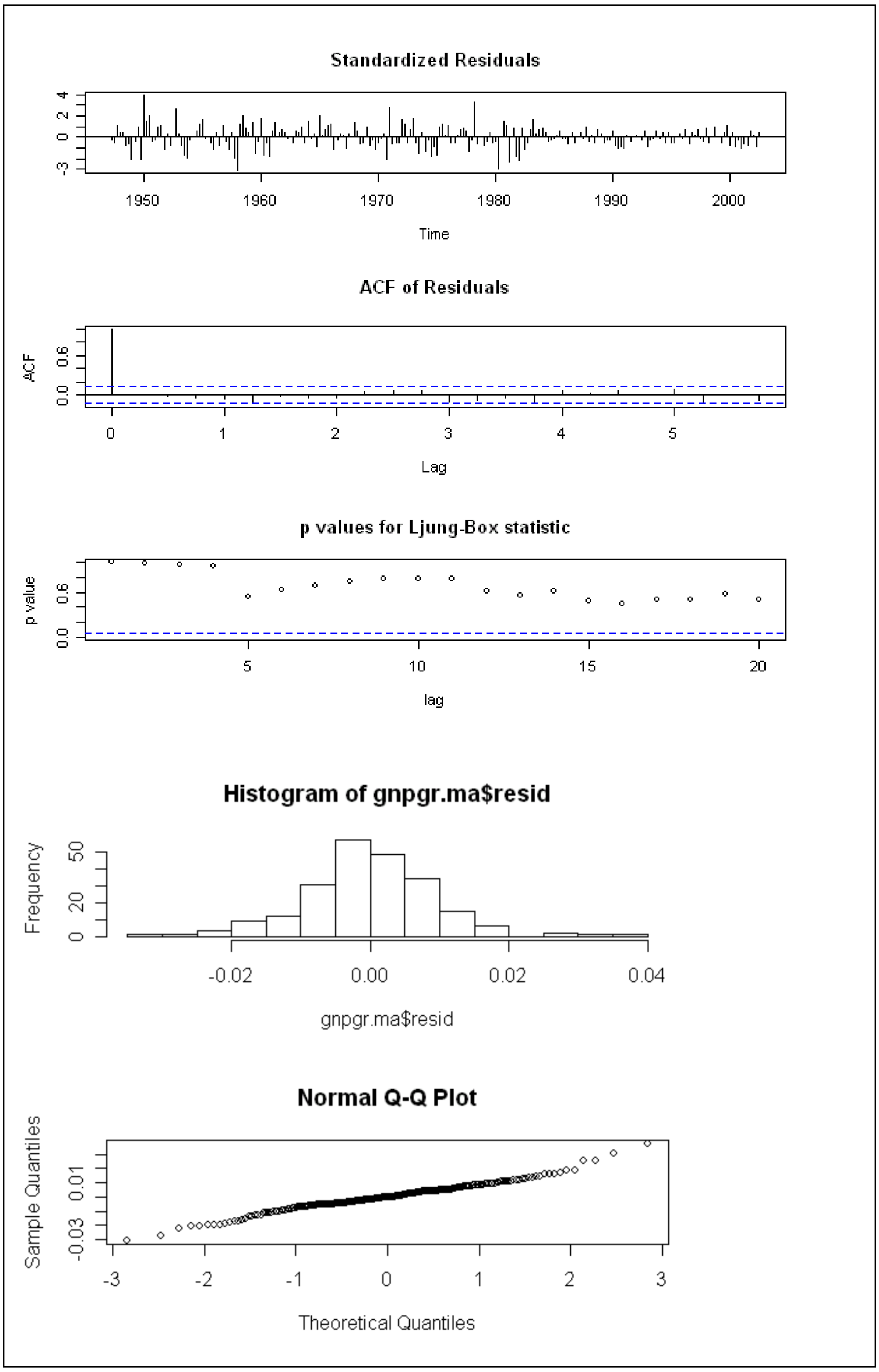


Figure 6.5 Residual Analysis of GDP series

6.4 Unit Roots in Time Series

The problem of unit root arises when either the AR or MA polynomial of ARMA has a root on or near the unit circle. A root near 1 in AR part is and indication that the data should be differenced, a root near 1 in MA part shows data were over-differenced.

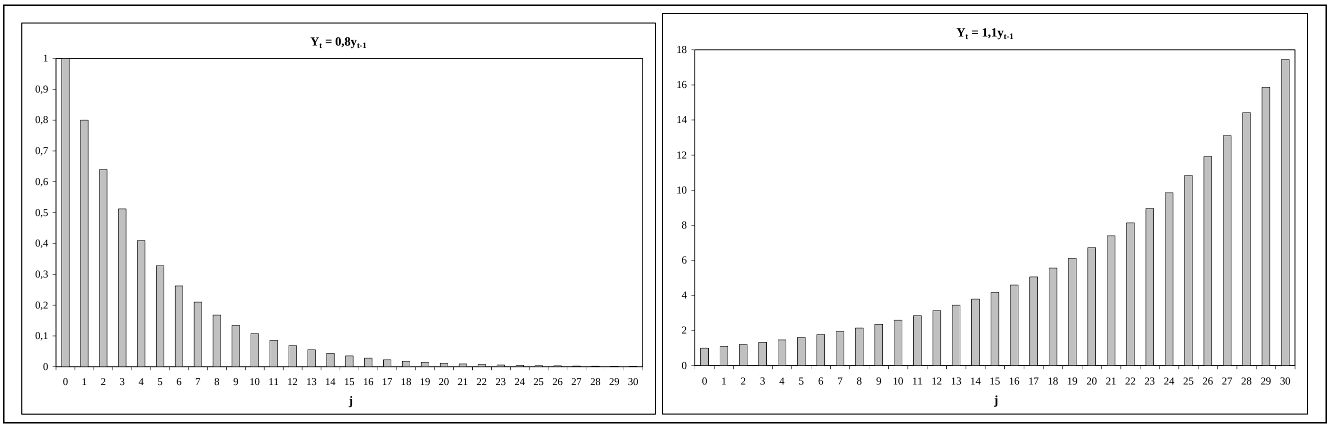


Figure 6.6. The impact of coefficient in AR(1) on the ACF (graph on the left φ=0.8; graph on right φ=1.1)

Let X_t be an AR(1) Process with drift μ.

$$X_t - \mu = \phi\left(X_{t-1} - \mu\right) + Z_t, \qquad Z_t \sim WN(0,\sigma^2)$$

for $|\phi| < 1 \Rightarrow E\left(X_t\right) = \mu < \infty$

For large n, the sampling distribution of $\hat{\phi}_{MLE} \sim N(\phi, \frac{1-\phi^2}{n})$. However, the normality is not applicable when $|\phi| = 1$ or $|\phi| > 1$. Therefore, we test statistically if

$H_0 : \phi = 1$ vs. $H_A : \phi < 1$.

Rewriting the model in I(1) [integrated model with degree 1] gives

$$X_t - \mu = \phi\left(X_{t-1} - \mu\right) + Z_t$$
$$X_t = \mu(1-\phi) + \phi X_{t-1} + Z_t$$

Subtracting X_{t-1} from both sides

$$\nabla X_t = X_t - X_{t-1} = \mu(1-\phi) + \phi X_{t-1} + Z_t - X_{t-1}$$
$$\nabla X_t = \mu(1-\phi) + (\phi - 1) X_{t-1} + Z_t$$
$$\nabla X_t = \alpha + \phi^* X_{t-1} + Z_t$$

where $\alpha = 1-\phi; \quad \phi^* = \phi - 1$. Now the hypothesis become

$H_0 : \phi^* = 0$ vs. $H_A : \phi^* < 0$

Dickey-Fuller Test statistic for AR(1) process

$$\hat{\tau} = \frac{\hat{\phi}^*}{\hat{\sigma}(\hat{\phi}^*)}$$

The limiting distribution of $\hat{\tau}$ has been derived and the tables of the percentiles of the distribution under H_0 is available. The test is rejected when $\hat{\tau}$ is too negative.

Table values by Dickey and Fuller are available, such as the model with drift will have critical values as follows:

Significance level	1%	5%	10%
Critical values without drift	-2.58	-1.95	-1.62
Critical values with drift	-3.43	-2.86	-2.57

Mac Kinnon Table also enables us to find the critical values for the ADF test based on the formula

$$K = \beta_\infty + \beta_1 T^{-1} + \beta_2 T^{-2}$$

In general for AR(p):

$$X_t - \mu = \phi_1 \;\; X_{t-1} - \mu \;\; + \phi_2 \;\; X_{t-2} - \mu \;\; + \ldots\ldots + \phi_p \;\; X_{t-p} - \mu \;\; + Z_t, \qquad Z_t \sim WN(0, \sigma^2)$$
$$\nabla X_t = \alpha + \phi_1^* X_{t-1} + \phi_2^* \nabla X_{t-2} + .. + \phi_p^* \nabla X_{t-p} + Z_t$$
$$\alpha = \mu(1 - \phi_1 - \phi_2 - .. \phi_p); \quad \phi_1^* = (\sum_{i=1}^{p} \phi)_i - 1; \quad \phi_j^* = -\sum_{i=j}^{p} \phi_i, \quad j = 2,..p$$

Having ∇X_t as an AR(p-1) process yields $H_0 : \phi_1^* = 0$ vs. $H_A : \phi_1^* < 0$.

The test statistics $\hat{\tau} = \dfrac{\hat{\phi}_1^*}{\hat{\sigma}(\hat{\phi}_1^*)}$ keeping the same critical values for rejecting null hypothesis.

Example: Given AR(3) with drift

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

$\nabla X_t = 0.1503 - 0.0041 X_{t-1} + 0.9335 \nabla X_{t-2} + 0.1548 \nabla X_{t-3} + Z_t$

$\hat{\sigma}(\alpha) = 0.1135; \quad \hat{\sigma}(\phi_1^*) = 0.0028; \quad \hat{\sigma}(\phi_2^*) = 0.0707; \quad \hat{\sigma}(\phi_3^*) = 0.0708$

$H_0 : \phi_1^* = 0$ vs. $H_A : \phi_1^* < 0$

$\hat{\tau} = \dfrac{-0.0041}{0.0028}$ =-1.464 >-2.57→ At 10% level of significance we fail to reject null hypothesis. There exists Unit root.

<u>Sequential test procedure:</u>

1. Start with a relatively high number of lags, such as 10
2. Subsequently, reduce the number of lags until the last coefficient is significant different from zero at 10 % level of significance.
3. Compare the models (without drift and trend, with drift, and with drift and trend) by looking at the Akaike criterion. Then choose the model having the lowest Akaike criterion.
4. If the value of the test statistic is greater than (or in absolute values lesser than) the critical value, fail to reject the existence of unit root.

<u>The Phillips-Peron Test:</u> A nonparametric method of controlling for higher order serial correlation in the series. The test statistic follows a t-distribution asymptotically. Mac Kinnon table values are also used to test unit root.

Example: Given AR(3) without drift

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

$\nabla X_t = -0.0012 X_{t-1} + 0.9395 \nabla X_{t-2} - 0.1585 \nabla X_{t-3} + Z_t$

$\hat{\sigma}(\phi_1^*) = 0.0018; \quad \hat{\sigma}(\phi_2^*) = 0.0707; \quad \hat{\sigma}(\phi_3^*) = 0.0709$

$H_0 : \phi_1^* = 0$ vs. $H_A : \phi_1^* < 0$

$\hat{\tau} = \dfrac{-0.0012}{0.0018}$ =-0.667 >-1.62→ At 10% level of significance we fail to reject null hypothesis. There exists Unit root.

6.5. Detrended Series

Consider the following random walk model without drift

$$X_t = a + bt + Z_t, \qquad Z_t \sim WN(0,\sigma^2)$$

$$\nabla X_t = (a + bt + Z_t) - (a + b(t-1) + Z_{t-1}) = b + Z_t - Z_{t-1}$$

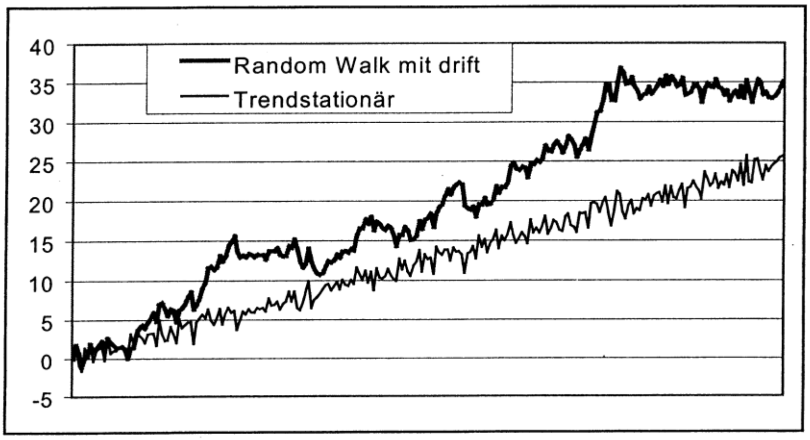
MA(1) part has unit root which destroys the stationarity.

Trend stationary time series are not mean stationary but include a trend. Including a trend component into the regression model, the process is expressed as
$Y_t = a + bt + \beta X_t + Z_t$,

where X_t is a stationary process. Differencing this series increases variance of the error term. Therefore, by LSE of the trend, one can obtain, the stationary series $\hat{X}_t = \gamma \hat{Y}_t - (\hat{\alpha}_1 + \hat{\alpha}_2 t)$.

Difference stationary time series (which are most of economic time series) contain a stochastic trend, differencing results in a stationary time series (see Figure 3.b).

Trend Stationary Process



.Difference stationary process

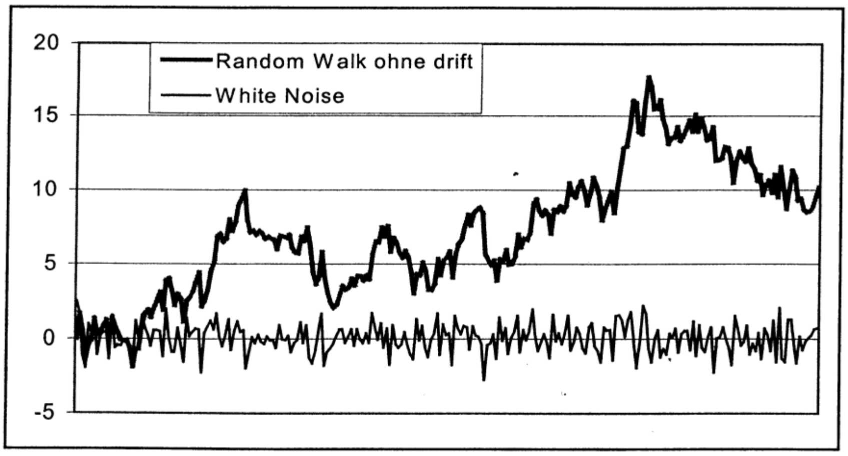


Figure 6.7. Examples to Trend and Difference stationary processes (source R.Fuess)

For stationarity of the error terms of the estimation equation $Y_t = a + \beta X_t + \varepsilon_t$ the following rules are observed:

$Y_t \sim I(0)$ and $X_t \sim I(0)$ $\Rightarrow \varepsilon_t \sim I(0)$,
$Y_t \sim I(1)$ and $X_t \sim I(0)$ $\Rightarrow \varepsilon_t \sim I(1)$,
$Y_t \sim I(1)$ and $X_t \sim I(1)$ $\Rightarrow \varepsilon_t \sim I(1)$, if Y and X are not cointegrated,
$Y_t \sim I(1)$ and $X_t \sim I(1)$ $\Rightarrow \varepsilon_t \sim I(0)$, if Y and X are cointegrated.

The residuals are only then I(0) if both variables Y and X either are I(0) or I(1) and cointegrated. The simplest case of cointegration is given when Y and X are I(1) and the linear combination of both variables is I(0), i.e. the residuals are stationary.

Dickey and Fuller provide the appropriate test statistics to determine whether a series contains a unit root, a unit root plus drift, and/or a unit root plus drift plus a time trend.

For a series having a structural break point results in weakly dependent residuals. In such cases Phillips-Perron Test (1988) can be used to test the existence of unit root.

Example: **S&P 500 index, Period: 02.01.1995 – 31.12.2007 daily data**

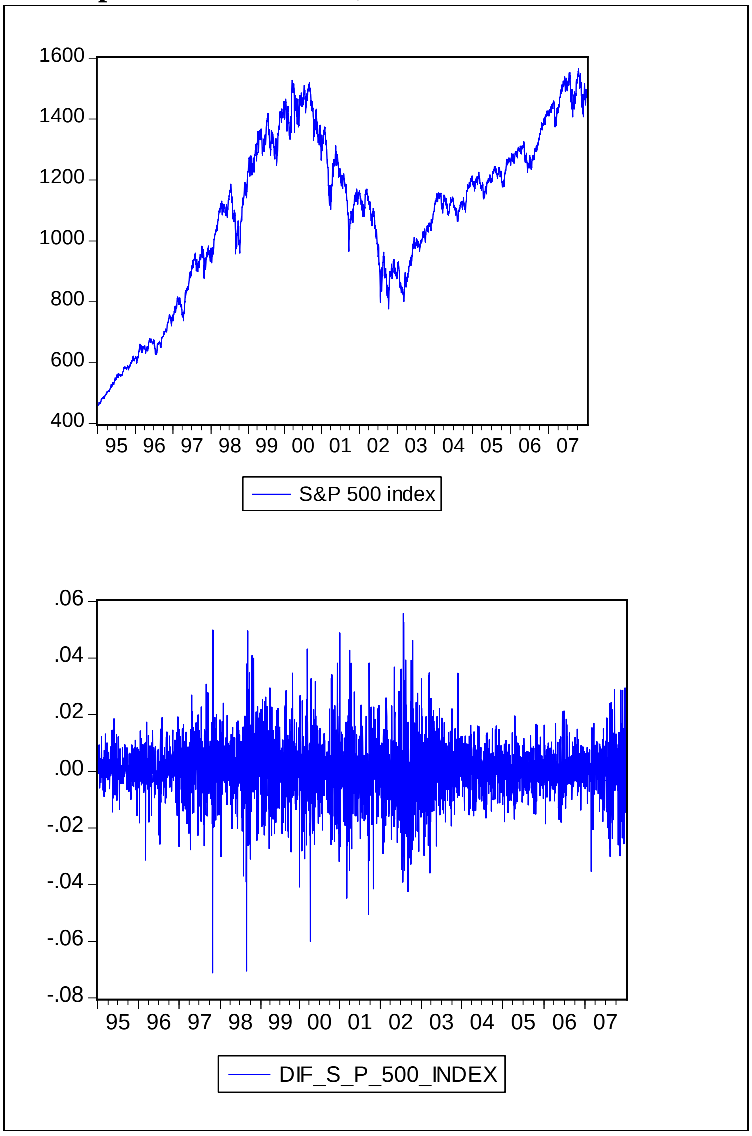


Figure 6.8. Original and differenced series

Let y_t denote index of S&P 500 and Z_t denote white-noise

$$Y_t = \mu + \phi Y_{t-1} + Z_t$$

Unit root test:

Ho: $\phi = 1$ S&P 500 index has unit root (not stationary)

H_1 $\phi < 1$ no unit root (stationary)

Null Hypothesis: S_P_500_INDEX has a unit root				
Exogenous: Constant				
Lag Length: 0 (Automatic based on SIC, MAXLAG=28)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			**-1.926797**	**0.3201**
Test critical values:	1% level		-3.432187	
	5% level		-2.862237	
	10% level		-2.567185	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(S_P_500_INDEX)				
Method: Least Squares				
Date: 06/18/08 Time: 20:01				
Sample (adjusted): 1/03/1995 12/31/2007				
Included observations: 3228 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
S_P_500_INDEX(-1)	-0.001493	0.000775	-1.926797	0.0541
C	1.942350	0.872543	2.226080	0.0261
R-squared	0.001149	Mean dependent var		0.312655
Adjusted R-squared	0.000840	S.D. dependent var		12.18334
S.E. of regression	12.17822	Akaike info criterion		7.837794
Sum squared resid	478444.8	Schwarz criterion		7.841561
Log likelihood	-12648.20	F-statistic		3.712547
Durbin-Watson stat	2.086200	Prob(F-statistic)		0.054093

Null Hypothesis: DIF_S_P_500_INDEX has a unit root				
Exogenous: Constant				
Lag Length: 0 (Automatic based on SIC, MAXLAG=28)				
			t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic			**-59.09627**	**0.0001**
Test critical values:	1% level		-3.432188	
	5% level		-2.862237	
	10% level		-2.567185	
*MacKinnon (1996) one-sided p-values.				
Augmented Dickey-Fuller Test Equation				
Dependent Variable: D(DIF_S_P_500_INDEX)				
Method: Least Squares				
Date: 06/18/08 Time: 21:44				
Sample (adjusted): 1/04/1995 12/31/2007				
Included observations: 3227 after adjustments				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
DIF_S_P_500_INDEX(-1)	-1.039872	0.017596	-59.09627	0.0000
C	0.000375	0.000190	1.973382	0.0485
R-squared	0.519901	Mean dependent var		-2.13E-06
Adjusted R-squared	0.519752	S.D. dependent var		0.015557
S.E. of regression	0.010781	Akaike info criterion		-6.221451
Sum squared resid	0.374839	Schwarz criterion		-6.217683
Log likelihood	10040.31	F-statistic		3492.369
Durbin-Watson stat	2.000633	Prob(F-statistic)		0.000000

We do not reject H_o and can conclude Dif_S&P has no root and it is stationary

Example: A series of data on the Annual GDP of a country is log-transformed and differenced. The graphs of these three series are presented in Figure 6.9

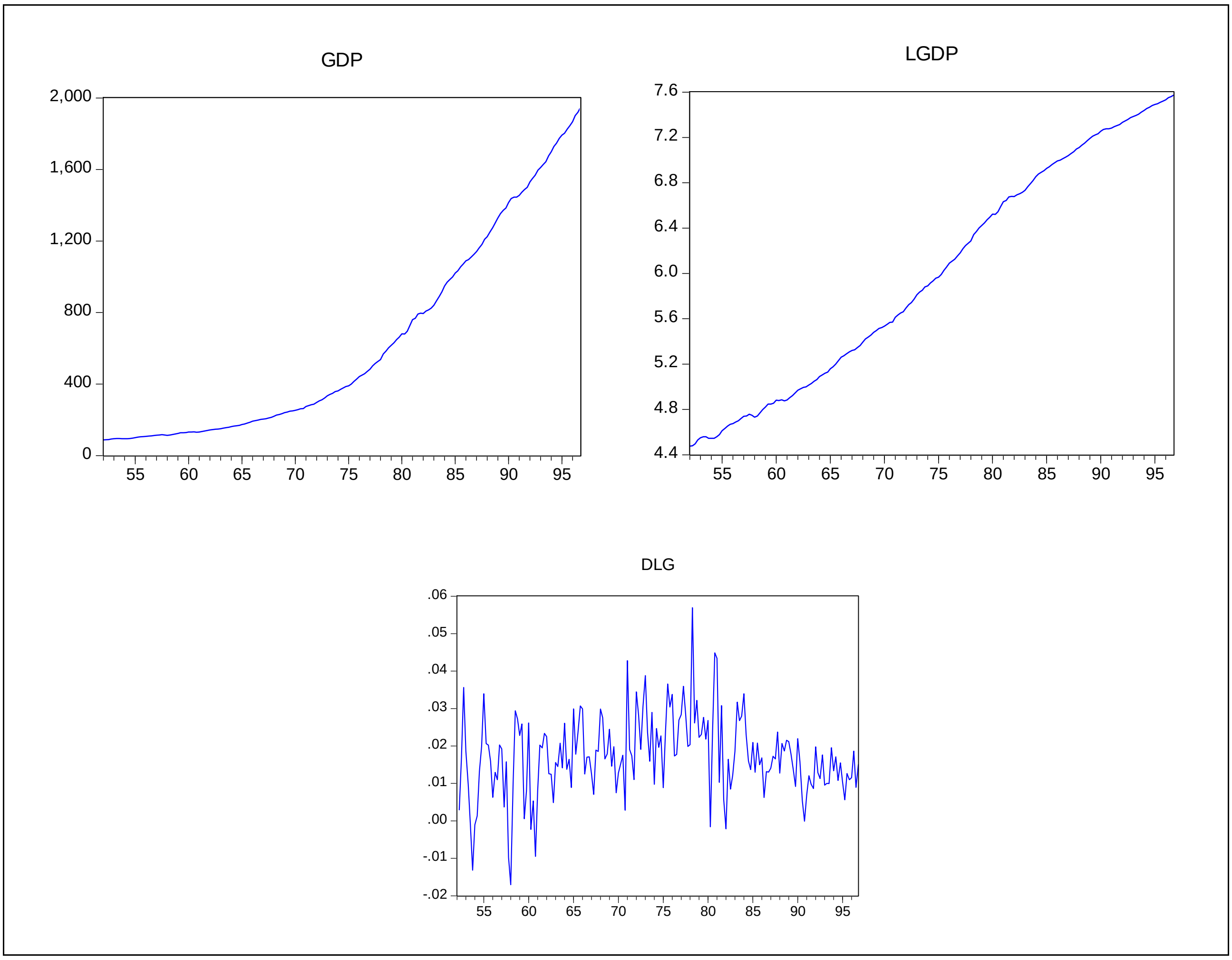


Figure 6.9 Graphs of GDP data having a log transformation and difference filter.

ACF log(GDP)

Date: 05/13/09 Time: 15:39
Sample: 1952Q1 1996Q4
Included observations: 180

	AC	PAC	Q-Stat	Prob
1	0.986	0.986	178.09	0.000
2	0.973	-0.014	352.22	0.000
3	0.959	-0.011	522.39	0.000
4	0.945	-0.001	688.68	0.000
5	0.931	-0.009	851.10	0.000
6	0.918	-0.011	1009.6	0.000
7	0.904	-0.015	1164.2	0.000
8	0.889	-0.021	1314.8	0.000
9	0.875	-0.016	1461.4	0.000
10	0.860	-0.014	1603.8	0.000
11	0.845	-0.012	1742.1	0.000
12	0.830	-0.008	1876.4	0.000
13	0.815	-0.004	2006.8	0.000
14	0.800	-0.006	2133.1	0.000
15	0.785	-0.010	2255.6	0.000
16	0.771	-0.011	2374.2	0.000
17	0.755	-0.015	2488.9	0.000
18	0.740	-0.010	2599.8	0.000
19	0.725	-0.013	2706.8	0.000
20	0.710	-0.009	2810.0	0.000
21	0.695	-0.007	2909.4	0.000
22	0.679	-0.016	3005.1	0.000
23	0.664	-0.012	3097.0	0.000
24	0.648	-0.019	3185.2	0.000
25	0.632	-0.023	3269.7	0.000
26	0.616	-0.018	3350.3	0.000
27	0.599	-0.010	3427.3	0.000
28	0.583	-0.008	3500.6	0.000
29	0.567	-0.007	3570.3	0.000
30	0.551	-0.010	3636.6	0.000
31	0.535	-0.015	3699.5	0.000
32	0.518	-0.012	3759.0	0.000
33	0.502	-0.005	3815.2	0.000
34	0.486	-0.014	3868.1	0.000
35	0.470	-0.014	3917.9	0.000
36	0.453	-0.018	3964.6	0.000

ACF Dlog(GDP)

Date: 05/13/09 Time: 15:41
Sample: 1952Q1 1996Q4
Included observations: 179

	AC	PAC	Q-Stat	Prob
1	0.391	0.391	27.811	0.000
2	0.220	0.079	36.663	0.000
3	0.104	-0.007	38.673	0.000
4	0.062	0.011	39.392	0.000
5	-0.016	-0.058	39.440	0.000
6	0.102	0.142	41.399	0.000
7	0.150	0.097	45.636	0.000
8	0.086	-0.034	47.022	0.000
9	0.181	0.152	53.276	0.000
10	0.251	0.151	65.331	0.000
11	0.230	0.084	75.562	0.000
12	0.104	-0.055	77.675	0.000
13	0.120	0.041	80.485	0.000
14	0.171	0.151	86.220	0.000
15	0.131	0.032	89.615	0.000
16	0.215	0.123	98.801	0.000
17	0.184	0.008	105.55	0.000
18	0.172	0.062	111.48	0.000
19	0.059	-0.055	112.19	0.000
20	0.135	0.047	115.91	0.000
21	0.042	-0.077	116.27	0.000
22	0.077	0.041	117.49	0.000
23	0.029	-0.071	117.66	0.000
24	0.102	0.027	119.84	0.000
25	0.126	0.026	123.18	0.000
26	0.099	-0.054	125.27	0.000
27	0.120	-0.003	128.32	0.000
28	0.110	0.017	130.94	0.000
29	0.097	0.009	132.96	0.000
30	-0.066	-0.198	133.92	0.000
31	0.021	0.024	134.01	0.000
32	-0.017	-0.057	134.07	0.000
33	-0.051	-0.084	134.64	0.000
34	0.049	0.041	135.17	0.000
35	0.050	-0.055	135.74	0.000
36	0.063	0.019	136.63	0.000

Figure 6.10 ACF, PACF of log transformed series (left) and diffrenced log(GDP) (right)

Null Hypothesis: LGDP has a unit root
Exogenous: Constant
Lag Length: 1 (Automatic based on SIC, MAXLAG=13)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		0.206904	0.9724
Test critical values:	1% level	-3.467205	
	5% level	-2.877636	
	10% level	-2.575430	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(LGDP)
Method: Least Squares
Date: 05/13/09 Time: 15:45
Sample (adjusted): 1952Q3 1996Q4
Included observations: 178 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
LGDP(-1)	0.000154	0.000743	0.206904	0.8363
D(LGDP(-1))	0.390070	0.069286	5.629868	0.0000
C	0.009714	0.004596	2.113632	0.0360

R-squared	0.154647	Mean dependent var	0.017393
Adjusted R-squared	0.144985	S.D. dependent var	0.010612
S.E. of regression	0.009813	Akaike info criterion	-6.393542
Sum squared resid	0.016851	Schwarz criterion	-6.339916
Log likelihood	572.0252	Hannan-Quinn criter.	-6.371795
F-statistic	16.00701	Durbin-Watson stat	2.051332
Prob(F-statistic)	0.000000		

Null Hypothesis: DLG has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic based on SIC, MAXLAG=13)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-8.831733	0.0000
Test critical values:	1% level	-3.467205	
	5% level	-2.877636	
	10% level	-2.575430	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(DLG)
Method: Least Squares
Date: 05/13/09 Time: 15:42
Sample (adjusted): 1952Q3 1996Q4
Included observations: 178 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
DLG(-1)	-0.609023	0.068958	-8.831733	0.0000
C	0.010620	0.001402	7.575367	0.0000

R-squared	0.307085	Mean dependent var	6.88E-05
Adjusted R-squared	0.303148	S.D. dependent var	0.011723
S.E. of regression	0.009786	Akaike info criterion	-6.404533
Sum squared resid	0.016855	Schwarz criterion	-6.368783
Log likelihood	572.0034	Hannan-Quinn criter.	-6.390035
F-statistic	77.99951	Durbin-Watson stat	2.052474
Prob(F-statistic)	0.000000		

Figure 6.11: Outputs of the Unit root tests

The Models Fitted

Model 1: AR(1)

Dependent Variable: LGDP
Method: Least Squares
Date: 05/13/09 Time: 16:36
Sample (adjusted): 1952Q2 1996Q4
Included observations: 179 after adjustments
Convergence achieved after 7 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C	-26.26741	48.08515	-0.546269	0.5856
AR(1)	1.000537	0.000800	1250.710	0.0000

R-squared	0.999887	Mean dependent var	5.999972
Adjusted R-squared	0.999886	S.D. dependent var	0.998870
S.E. of regression	0.010655	Akaike info criterion	-6.234537
Sum squared resid	0.020093	Schwarz criterion	-6.198924
Log likelihood	559.9911	Hannan-Quinn criter.	-6.220096
F-statistic	1564277.	Durbin-Watson stat	1.211297
Prob(F-statistic)	0.000000		

Inverted AR Roots 1.00
Estimated AR process is nonstationary

Model 2 ARIMA(1,1,2)

Dependent Variable: DLGDP
Method: Least Squares
Date: 05/13/09 Time: 16:38
Sample (adjusted): 1952Q3 1996Q4
Included observations: 178 after adjustments
Convergence achieved after 29 iterations
MA Backcast: 1952Q1 1952Q2

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.017438	0.001248	13.96996	0.0000
AR(1)	0.331161	0.277345	1.194039	0.2341
MA(1)	0.028161	0.280826	0.100278	0.9202
MA(2)	0.108431	0.121545	0.892113	0.3736

R-squared	0.161954	Mean dependent var	0.017393
Adjusted R-squared	0.147505	S.D. dependent var	0.010612
S.E. of regression	0.009798	Akaike info criterion	-6.390988
Sum squared resid	0.016705	Schwarz criterion	-6.319487
Log likelihood	572.7979	Hannan-Quinn criter.	-6.361992
F-statistic	11.20865	Durbin-Watson stat	1.990762
Prob(F-statistic)	0.000001		

Inverted AR Roots	.33	
Inverted MA Roots	-.01+.33i	-.01-.33i

Figure 6.12: Outputs of the plausible models

Residual Analysis

Model 1 | Model 2

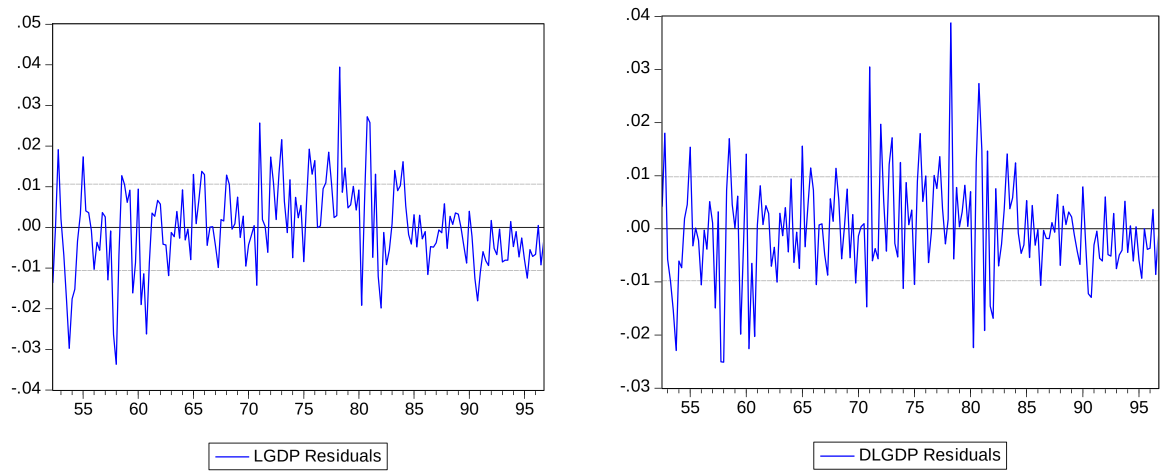


Figure 6.13: Residual plots of two models

ACF of Residuals

MODEL 1

Date: 05/13/09 Time: 16:45
Sample: 1952Q2 1996Q4
Included observations: 179
Q-statistic probabilities adjusted for 1 ARMA term(s)

	AC	PAC	Q-Stat	Prob
1	0.389	0.389	27.610	
2	0.218	0.078	36.277	0.000
3	0.101	-0.009	38.159	0.000
4	0.058	0.009	38.791	0.000
5	-0.020	-0.060	38.866	0.000
6	0.099	0.142	40.704	0.000
7	0.148	0.097	44.808	0.000
8	0.083	-0.034	46.125	0.000
9	0.180	0.152	52.288	0.000
10	0.250	0.151	64.226	0.000
11	0.229	0.084	74.331	0.000
12	0.102	-0.056	76.337	0.000
13	0.117	0.041	79.021	0.000
14	0.168	0.150	84.557	0.000
15	0.128	0.031	87.777	0.000
16	0.212	0.122	96.702	0.000
17	0.181	0.007	103.23	0.000
18	0.169	0.061	108.95	0.000
19	0.056	-0.056	109.57	0.000
20	0.131	0.046	113.10	0.000
21	0.038	-0.077	113.39	0.000
22	0.073	0.040	114.49	0.000
23	0.025	-0.071	114.63	0.000
24	0.100	0.027	116.72	0.000
25	0.124	0.026	119.93	0.000
26	0.096	-0.055	121.87	0.000
27	0.116	-0.005	124.72	0.000
28	0.106	0.017	127.15	0.000
29	0.092	0.008	128.98	0.000
30	-0.071	-0.199	130.08	0.000
31	0.016	0.024	130.14	0.000
32	-0.021	-0.058	130.24	0.000
33	-0.054	-0.084	130.90	0.000
34	0.046	0.040	131.36	0.000
35	0.049	-0.054	131.89	0.000
36	0.061	0.020	132.74	0.000

MODEL 2

Date: 05/13/09 Time: 16:42
Sample: 1952Q3 1996Q4
Included observations: 178
Q-statistic probabilities adjusted for 3 ARMA term(s)

	AC	PAC	Q-Stat	Prob
1	0.004	0.004	0.0030	
2	-0.000	-0.000	0.0030	
3	0.034	0.034	0.2131	
4	0.037	0.037	0.4676	0.494
5	-0.119	-0.120	3.0927	0.213
6	0.064	0.065	3.8458	0.279
7	0.122	0.121	6.6553	0.155
8	-0.058	-0.056	7.2789	0.201
9	0.073	0.080	8.2887	0.218
10	0.176	0.156	14.177	0.048
11	0.162	0.177	19.207	0.014
12	-0.039	-0.016	19.501	0.021
13	0.032	-0.010	19.701	0.032
14	0.114	0.123	22.245	0.023
15	-0.008	0.033	22.257	0.035
16	0.133	0.144	25.780	0.018
17	0.085	0.036	27.209	0.018
18	0.108	0.104	29.550	0.014
19	-0.056	-0.023	30.174	0.017
20	0.113	0.050	32.748	0.012
21	-0.017	-0.059	32.804	0.018
22	0.034	0.026	33.044	0.024
23	-0.067	-0.097	33.980	0.026
24	0.070	0.009	35.003	0.028
25	0.099	0.055	37.036	0.023
26	0.019	-0.029	37.116	0.032
27	0.055	-0.040	37.750	0.037
28	0.075	0.017	38.941	0.037
29	0.082	0.068	40.380	0.036
30	-0.148	-0.173	45.143	0.016
31	0.077	0.009	46.419	0.016
32	-0.018	-0.062	46.493	0.021
33	-0.098	-0.096	48.597	0.017
34	0.043	-0.001	49.001	0.021
35	0.031	-0.074	49.222	0.026
36	0.048	0.015	49.741	0.031

6.6. Structural Breaks-Chow test

The aim is to determine if there exist a structural change in the relationship. Fit models separately for subsample to see if there are significant differences in the estimated equations. The steps are:

1. Partition of data set data into subsamples at times having significant structural breakpoints.
2. Estimate model over whole sample and save residual sum of squares
3. Estimate model with different coefficients before and after the date t_1.
4. Calculate test statistics

$$F(t_1) = \frac{(T-2k)(\varepsilon'\varepsilon - e'e)}{(e'e)k}$$

where ε: residuals for unrestricted model, e: residuals of restricted model, T: no. of observations, k no. of parameters

5. Conclude that the model is stable if F is below the critical value

Example: Given the series below, test if there exists any structural break.

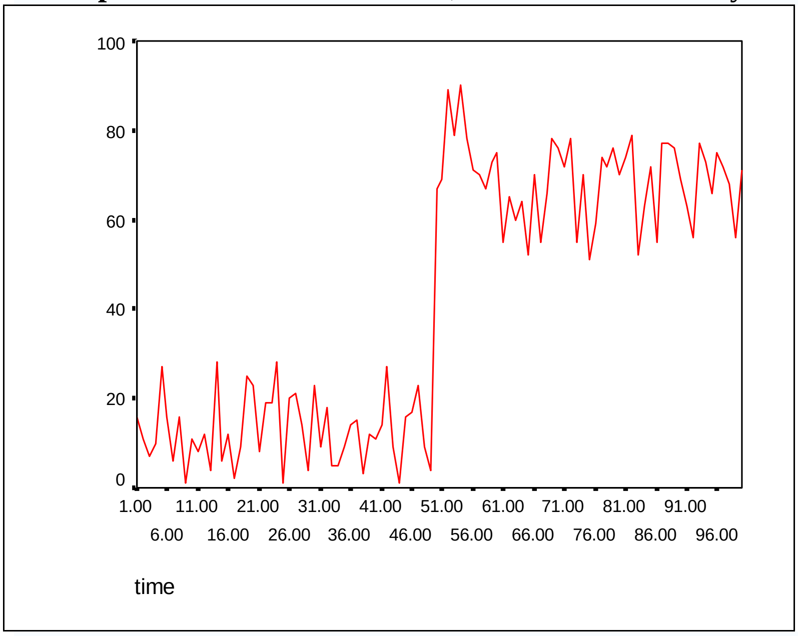


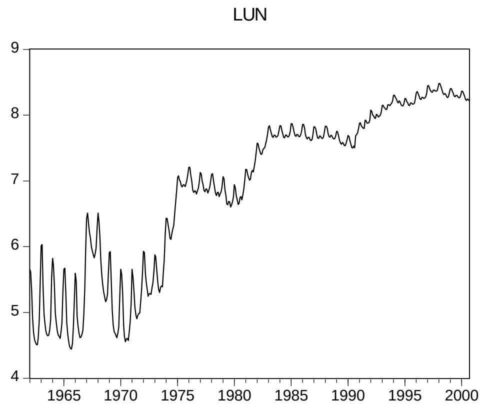
Figure 6.14 Plot of the sries

Chow Breakpoint Test		probability
F-statistics	38.39	0.00
Log Likelihood ratio	65.75	0.00

Therefore, Ho: No structural changes is rejected!

Exercises:

1. Based on the output, determine and test if the following series whose plot is given below contain a structura break.

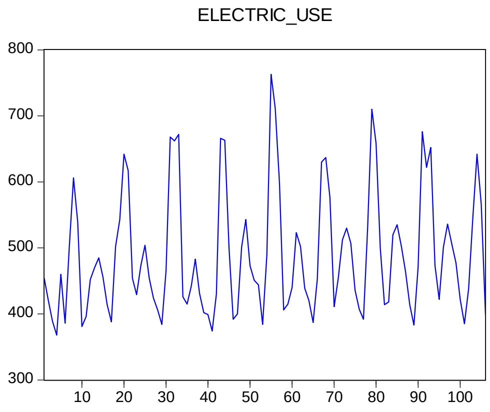


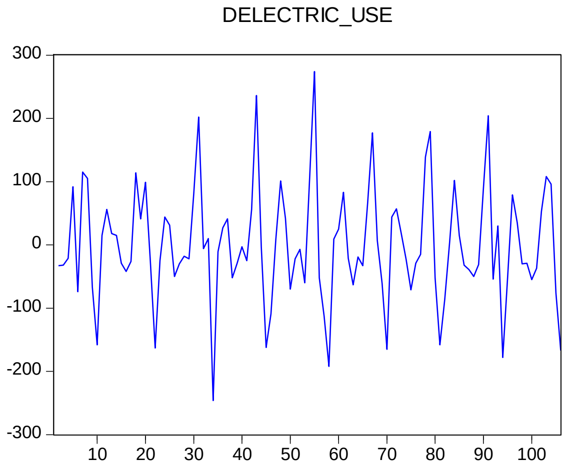
Varying regressors: All equation variables

Equation Sample: 1962M01 2000M09

F-statistic	1848.832	Prob. F(1,463)	0.0000
Log likelihood ratio	747.7519	Prob. Chi-Square(1)	0.0000
Wald Statistic	1848.832	Prob. Chi-Square(1)	0.0000

2. Yearly electric use of a certain plant has been recorded and plotted as below. Based on the graphs and analyses given, test it the unit root exists and find the appropriate model for the series.
3.





ACF

Original series

Date: 05/13/09 Time: 17:07
Sample: 1 106
Included observations: 106

Lag	AC	PAC	Q-Stat	Prob
1	0.524	0.524	29.881	0.000
2	-0.144	-0.576	32.160	0.000
3	-0.476	-0.115	57.346	0.000
4	-0.322	0.063	68.975	0.000
5	0.021	-0.007	69.027	0.000
6	0.210	-0.037	74.067	0.000
7	0.057	-0.215	74.446	0.000
8	-0.299	-0.321	84.895	0.000
9	-0.474	-0.200	111.36	0.000
10	-0.164	0.166	114.55	0.000
11	0.441	0.395	137.95	0.000
12	0.774	0.341	210.94	0.000
13	0.442	-0.112	235.01	0.000
14	-0.148	-0.057	237.73	0.000
15	-0.441	0.068	262.18	0.000
16	-0.296	0.034	273.31	0.000
17	0.003	-0.128	273.32	0.000
18	0.150	-0.160	276.26	0.000
19	0.026	-0.008	276.34	0.000
20	-0.276	0.019	286.52	0.000
21	-0.423	-0.020	310.59	0.000
22	-0.107	0.108	312.16	0.000
23	0.422	0.024	336.66	0.000
24	0.695	0.149	404.08	0.000
25	0.374	-0.076	423.82	0.000
26	-0.144	0.021	426.78	0.000
27	-0.407	-0.036	450.73	0.000
28	-0.241	0.119	459.24	0.000
29	0.012	-0.101	459.26	0.000
30	0.132	-0.005	461.90	0.000
31	0.018	0.057	461.95	0.000
32	-0.243	0.056	471.07	0.000
33	-0.368	-0.024	492.34	0.000
34	-0.079	0.024	493.33	0.000
35	0.385	-0.063	517.28	0.000
36	0.591	0.013	574.48	0.000

Differenced series

Date: 05/13/09 Time: 17:07
Sample: 1 106
Included observations: 105

Lag	AC	PAC	Q-Stat	Prob
1	0.219	0.219	5.1741	0.023
2	-0.349	-0.417	18.450	0.000
3	-0.527	-0.416	49.016	0.000
4	-0.208	-0.224	53.835	0.000
5	0.155	-0.172	56.549	0.000
6	0.370	0.016	72.124	0.000
7	0.226	0.038	77.997	0.000
8	-0.193	-0.169	82.301	0.000
9	-0.517	-0.418	113.60	0.000
10	-0.315	-0.474	125.36	0.000
11	0.281	-0.316	134.82	0.000
12	0.712	0.188	196.10	0.000
13	0.293	0.096	206.61	0.000
14	-0.320	-0.084	219.21	0.000
15	-0.461	0.035	245.70	0.000
16	-0.186	0.151	250.06	0.000
17	0.156	0.182	253.18	0.000
18	0.293	-0.002	264.29	0.000
19	0.198	-0.048	269.39	0.000
20	-0.162	0.002	272.88	0.000
21	-0.493	-0.119	305.39	0.000
22	-0.229	0.008	312.50	0.000
23	0.262	-0.135	321.92	0.000
24	0.632	0.081	377.24	0.000
25	0.215	-0.061	383.76	0.000
26	-0.255	0.039	393.00	0.000
27	-0.450	-0.095	422.12	0.000
28	-0.112	0.114	423.95	0.000
29	0.129	0.004	426.42	0.000
30	0.258	-0.031	436.43	0.000
31	0.158	-0.016	440.25	0.000
32	-0.136	0.063	443.09	0.000
33	-0.448	-0.014	474.42	0.000
34	-0.188	0.070	479.99	0.000
35	0.269	-0.013	491.61	0.000
36	0.545	0.107	540.04	0.000

Null Hypothesis: ELECTRIC_USE has a unit root

Exogenous: Constant

Lag Length: 11 (Automatic based on SIC, MAXLAG=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.817925	0.1096
Test critical values:	1% level	-3.501445	
	5% level	-2.892536	
	10% level	-2.583371	

Null Hypothesis: D(ELECTRIC_USE) has a unit root

Exogenous: Constant

Lag Length: 10 (Automatic based on SIC, MAXLAG=12)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-11.45223	0.0001
Test critical values:	1% level	-3.501445	
	5% level	-2.892536	
	10% level	-2.583371	

MODEL 1:

Dependent Variable: DIF_ELECTRIC_USE

Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.501761	9.145950	-0.054862	0.9564
AR(1)	0.004764	0.076082	0.062621	**0.9502**
AR(2)	-0.085144	0.075745	-1.124090	**0.2640**
AR(3)	-0.202099	0.081909	-2.467361	**0.0155**
AR(12)	0.638938	0.082295	7.764024	**0.0000**

R-squared	0.641363	Mean dependent var	-0.763441
Adjusted R-squared	0.625061	S.D. dependent var	92.56389
S.E. of regression	56.67897	Akaike info criterion	10.96495
Sum squared resid	282700.5	Schwarz criterion	11.10111
Log likelihood	-504.8701	F-statistic	39.34330
Durbin-Watson stat	2.614008	Prob(F-statistic)	0.000000

Inverted AR Roots	.94	.82-.50i	.82+.50i	.49-.86i
	.49+.86i	.02+.97i	.02-.97i	-.47+.83i
	-.47-.83i	-.84+.47i	-.84-.47i	-.97

Model 2:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.081713	1.127498	-0.959393	0.3400
AR(1)	0.346492	0.085281	4.062930	**0.0001**
AR(2)	-0.235234	0.077020	-3.054188	**0.0030**
AR(12)	0.645663	0.073520	8.782100	**0.0000**
MA(1)	-0.973815	0.017299	-56.29357	**0.0000**
R-squared	0.744881	Mean dependent var		-0.763441
Adjusted R-squared	0.733285	S.D. dependent var		92.56389
S.E. of regression	47.80414	Akaike info criterion		10.62437
Sum squared resid	201100.8	Schwarz criterion		10.76053
Log likelihood	-489.0330	F-statistic		64.23428
Durbin-Watson stat	1.803597	Prob(F-statistic)		0.000000

Residual Analyses

Model 1

Date: 05/13/09 Time: 17:13
Sample: 14 106
Included observations: 93
Q-statistic probabilities adjusted for 4 ARMA term(s)

	AC	PAC	Q-Stat	Prob
1	0.083	0.083	0.6636	
2	0.182	0.176	3.8771	
3	-0.048	-0.078	4.1040	
4	-0.004	-0.027	4.1052	
5	-0.063	-0.039	4.4974	0.034
6	0.003	0.013	4.4981	0.105
7	0.025	0.043	4.5608	0.207
8	0.017	0.003	4.5925	0.332
9	-0.095	-0.116	5.5479	0.353
10	-0.246	-0.248	11.993	0.062
11	-0.030	0.049	12.090	0.098
12	-0.209	-0.138	16.840	0.032
13	-0.070	-0.087	17.383	0.043
14	-0.147	-0.114	19.801	0.031
15	0.079	0.078	20.509	0.039
16	-0.059	-0.042	20.913	0.052
17	0.047	0.014	21.175	0.070
18	-0.080	-0.084	21.934	0.080
19	-0.040	-0.112	22.123	0.105
20	-0.018	-0.017	22.164	0.138
21	0.014	0.014	22.187	0.178
22	0.141	0.056	24.653	0.135
23	0.084	-0.011	25.542	0.143
24	0.145	0.044	28.223	0.104
25	-0.029	-0.044	28.329	0.131
26	-0.096	-0.193	29.556	0.130
27	-0.174	-0.148	33.623	0.071
28	0.031	0.029	33.756	0.089
29	-0.022	-0.003	33.825	0.112
30	0.039	-0.061	34.040	0.134
31	-0.030	-0.071	34.164	0.161
32	-0.020	-0.051	34.221	0.194
33	-0.115	-0.098	36.177	0.168
34	-0.027	0.056	36.286	0.199
35	-0.033	-0.045	36.449	0.230
36	0.212	0.180	43.404	0.086

Model 2

Date: 05/13/09 Time: 17:11
Sample: 14 106
Included observations: 93
Q-statistic probabilities adjusted for 4 ARMA term(s)

	AC	PAC	Q-Stat	Prob
1	-0.311	-0.311	9.3092	
2	-0.021	-0.130	9.3504	
3	-0.033	-0.091	9.4595	
4	-0.057	-0.115	9.7822	
5	-0.098	-0.188	10.745	0.001
6	-0.007	-0.149	10.749	0.005
7	0.037	-0.071	10.890	0.012
8	0.037	-0.021	11.029	0.026
9	-0.036	-0.080	11.167	0.048
10	-0.138	-0.254	13.181	0.040
11	0.176	-0.002	16.513	0.021
12	-0.138	-0.154	18.586	0.017
13	0.208	0.115	23.373	0.005
14	-0.175	-0.166	26.780	0.003
15	0.190	0.091	30.885	0.001
16	-0.112	-0.057	32.326	0.001
17	0.035	0.053	32.465	0.002
18	-0.078	-0.070	33.177	0.003
19	-0.016	-0.085	33.207	0.004
20	-0.064	-0.166	33.702	0.006
21	-0.017	-0.135	33.737	0.009
22	0.106	-0.040	35.140	0.009
23	-0.007	-0.015	35.147	0.013
24	0.075	-0.055	35.870	0.016
25	0.079	0.201	36.680	0.018
26	-0.012	0.000	36.700	0.026
27	-0.178	-0.027	40.957	0.012
28	0.126	-0.026	43.115	0.010
29	-0.078	0.045	43.962	0.011
30	0.020	-0.045	44.020	0.015
31	-0.018	0.027	44.067	0.020
32	0.015	-0.020	44.100	0.027
33	-0.090	-0.093	45.280	0.028
34	-0.004	-0.061	45.282	0.036
35	-0.062	-0.152	45.859	0.042
36	0.212	0.102	52.852	0.012

Chapter 7
Seasonal ARIMA Models

Dependence on the past tends to occur most strongly at multiples of seasonal lag s. For example, monthly economical data is expected to have a seasonal effect of lag 12 a strong component, or temperature having a seasons of three months etc.

Definition: $X_t \quad {}_{t\in N}$ is said to be **pure Seasonal** ARMA(P,Q)$_s$ having form

$$\Phi(B^s)X_t = \Theta(B^s)Z_t$$

where

$$\Phi(B^s) = (1-\Phi_1 B^s - \Phi_2 B^{2s} - .. - \Phi_P B^{Ps}); \Theta(B^s) = (1-\Theta_1 B^s - \Theta_2 B^{2s} - .. - \Theta_Q B^{Qs})$$

Example 1: For a stationary ARMA(1,0)$_3$, P=1, Q=0, s=3

$$(1-\Phi B^3)X_t = Z_t \Rightarrow X_t = \Phi X_{t-3} + Z_t$$

State the conditions for stationarity:

$$1-\Phi B^3 = 0 \Rightarrow B = \frac{1}{\sqrt[3]{\Phi}} \Rightarrow |\Phi| < 1$$

Find the first 4 PACF values.

$$\gamma(h) = Cov(\Phi X_{t-3} + Z_t, X_{t-h})$$

By Yule Walker equations

$$\gamma(0) = \Phi\gamma(3) + \sigma^2$$
$$\gamma(1) = \Phi\gamma(2)$$
$$\gamma(2) = \Phi\gamma(1)$$
$$\gamma(3) = \Phi\gamma(0)$$
$$\gamma(4) = \Phi\gamma(1)$$

Replacing $\gamma(2) = \Phi\gamma(1) \quad in \quad \gamma(1)$ we find $\gamma(2) = 0, \quad \gamma(1) = 0, \quad \gamma(4) = 0$

$$\gamma(3) = \Phi\gamma(0) = \frac{\Phi\sigma^2}{1-\Phi^2}$$

Therefore, $\rho(2) = \rho(1) = \rho(4) = 0, \quad \rho(3) = \Phi$

PACF's

$$\phi_{11} = \rho(1) = 0; \quad \phi_{22} = \frac{\rho(2) - \rho(1)^2}{1-\rho(1)^2} = 0$$

$$\phi_{33} = \Phi; \quad \phi_{44} = 0 \quad \text{by cut-off property}$$

Definition: $X_t \quad {}_{t \in N}$ is said to be **Integrated Seasonal** ARIMA (p,d,q)x(P,D,Q)$_s$ having

form $\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D X_t = \theta(B)\Theta(B^s)Z_t$

where

$$\Phi(B^s) = (1-\Phi_1 B^s - \Phi_2 B^{2s} - .. - \Phi_P B^{Ps})$$
$$\Theta(B^s) = (1-\Theta_1 B^s - \Theta_2 B^{2s} - .. - \Theta_Q B^{Qs})$$
$$\phi(B) = (1-\phi_1 B - \phi_2 B^2 - .. - \phi_p B^p)$$
$$\theta(B) = (1-\theta_1 B - \theta_2 B^2 - .. - \theta_q B^q)$$

Example 2: ARMA(1,0)$_3$ can also be expressed as ARIMA(1,0,0)x(0,0,0)$_3$

Example 3: ARIMA(0,0,0)x(1,1,0)$_3$

$$(1-\Phi B^3)(1-B^3)X_t = Z_t \Rightarrow X_t - X_{t-3} = \Phi(X_{t-3} - X_{t-6}) + Z_t$$

Example 4: ARIMA(0,0,0)x(1,1,1)$_3$

$$(1-\Phi B^3)(1-B^3)X_t = (1+\Theta B^3)Z_t \Rightarrow X_t - X_{t-3} = \Phi(X_{t-3} - X_{t-6}) + Z_t + \Theta Z_{t-3}$$

Example 5: ARIMA(0,1,0)x(1,0,1)$_3$

$$(1-\Phi B^3)(1-B)X_t = (1+\Theta B^3)Z_t \Rightarrow X_t - X_{t-1} = \Phi(X_{t-3} - X_{t-4}) + Z_t + \Theta Z_{t-3}$$

Example 7: On Outbord Marine Data quarterly collected from Dec. 1983 to Sept. 1993

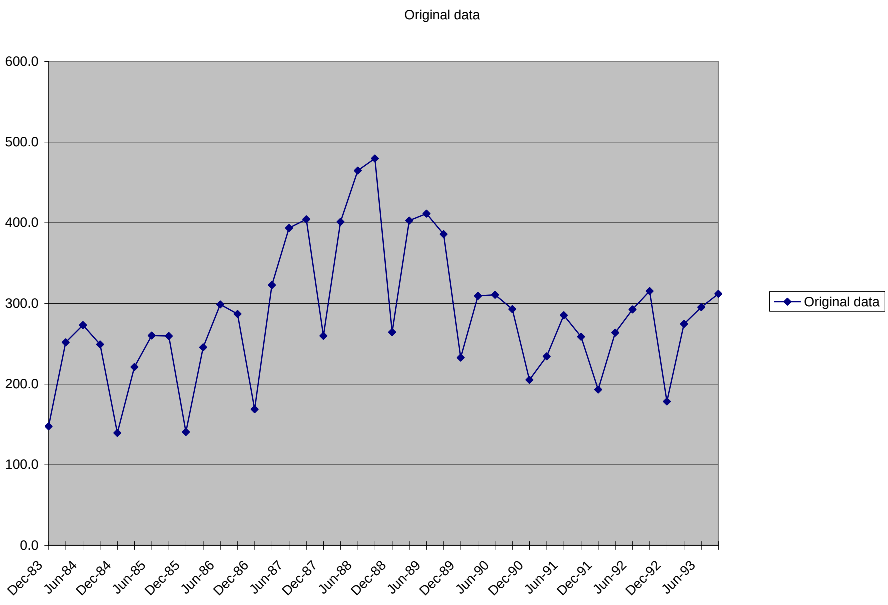


Figure 7.1. Time series graph of the data

Original series

Date: 06/21/08 Time: 16:47
Sample: 1983Q4 1993Q3
Included observations: 40

	AC	PAC	Q-Stat	Prob
1	0.421	0.421	7.6500	0.006
2	0.194	0.020	9.3177	0.009
3	0.312	0.272	13.736	0.003
4	0.733	0.669	38.819	0.000
5	0.177	-0.619	40.328	0.000
6	-0.119	-0.386	41.033	0.000
7	-0.052	-0.216	41.172	0.000
8	0.297	-0.100	45.816	0.000
9	-0.193	-0.143	47.831	0.000
10	-0.442	0.083	58.762	0.000
11	-0.347	0.056	65.722	0.000
12	0.028	0.124	65.769	0.000
13	-0.361	-0.033	73.877	0.000
14	-0.509	0.057	90.624	0.000
15	-0.386	-0.161	100.63	0.000
16	0.014	-0.116	100.65	0.000
17	-0.262	0.011	105.65	0.000
18	-0.341	0.024	114.54	0.000
19	-0.216	-0.026	118.26	0.000
20	0.130	-0.175	119.68	0.000

Differenced series

Date: 06/21/08 Time: 17:02
Sample: 1983Q4 1993Q3
Included observations: 39

	AC	PAC	Q-Stat	Prob
1	-0.275	-0.275	3.1868	0.074
2	-0.307	-0.414	7.2547	0.027
3	-0.282	-0.672	10.780	0.013
4	0.836	0.609	42.694	0.000
5	-0.200	0.101	44.582	0.000
6	-0.316	-0.099	49.406	0.000
7	-0.254	-0.004	52.640	0.000
8	0.712	0.014	78.801	0.000
9	-0.177	-0.181	80.480	0.000
10	-0.288	-0.041	85.055	0.000
11	-0.260	-0.164	88.925	0.000
12	0.640	-0.056	113.15	0.000
13	-0.153	-0.069	114.60	0.000
14	-0.217	0.089	117.61	0.000
15	-0.242	0.070	121.52	0.000
16	0.540	-0.078	141.81	0.000

MULTIPLICATIVE MODEL (Classical)

Year	Data	Seasonal Adjusted	Year		Seasonally Adjusted
12/ 1983	147.6	210.7279544	3/1988	259.7	370.7726948
3/1984	251.8	240.9128147		401.1	383.7574662
	273.1	230.5266085		464.6	392.173791
	249.1	216.0401135		479.7	416.0354977
	139.3	198.8780762	3/1989	264.4	377.4828668
3.1995	221.2	211.6358801		402.6	385.19261
	260.2	219.6375816		411.3	347.1826953
	259.5	225.0598533		385.9	334.684383
	140.5	200.5913116	3/1989	232.7	332.2248982
3/1986	245.5	234.8852105	…..	309.2	295.8309862
	298.8	252.2202513	……	310.7	262.2651676
	287	248.9101268	………	293	254.1138228
	168.8	240.9951131	**Unnormalized**	**Normalized**	
3/1987	322.6	308.6516046	**Seas. Index**	**Seas. Index**	
	393.5	332.1575264	**115.872%**	**116.0504%**	
	404.3	350.6423842	**112.776%**	**112.9497%**	
3/1988	259.7	370.7726948	**68.508%**	**68.6136%**	
	401.1	383.7574662	**102.229%**	**102.3863%**	
			399.385%	**400.0000%**	

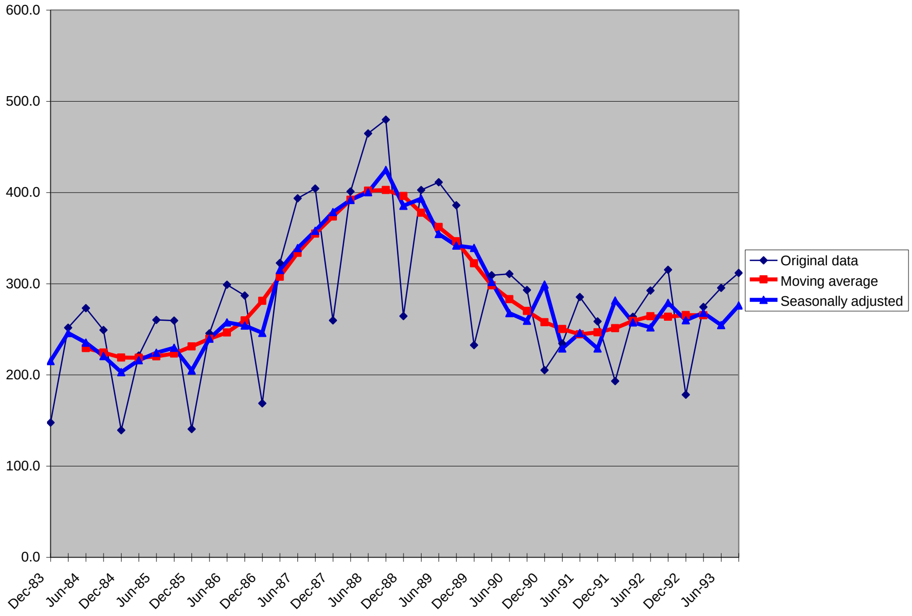


Figure 7.2. Seasonally adjusted series and MA series

Example 2: Monthly marriages in New York for between 2000-2005

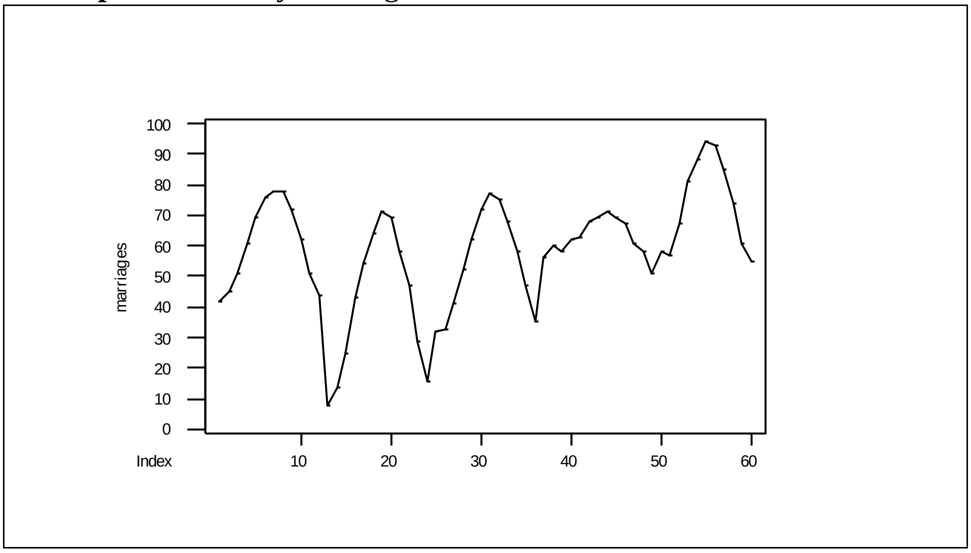
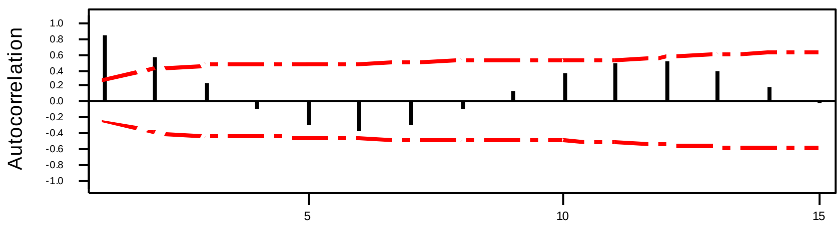


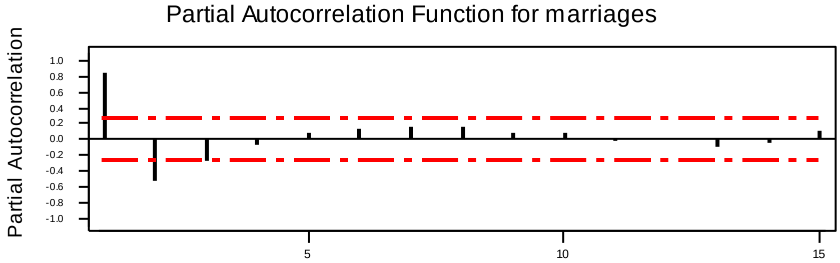
Figure 7.3 Plot of the original series

Autocorrelation Function for marriages



Lag	Corr	T	LBQ	Lag	Corr	T	LBQ	Lag	Corr	T	LBQ
1	0.85	6.55	45.12	8	-0.11	-0.43	93.03	15	-0.03	-0.11	157.47
2	0.57	2.81	65.59	9	0.13	0.52	94.32				
3	0.22	0.99	68.87	10	0.36	1.40	103.87				
4	-0.09	-0.40	69.44	11	0.49	1.87	122.34				
5	-0.31	-1.34	75.84	12	0.51	1.83	142.45				
6	-0.38	-1.60	85.71	13	0.39	1.34	154.66				
7	-0.30	-1.23	92.16	14	0.18	0.61	157.39				

Partial Autocorrelation Function for marriages



Lag	PAC	T	Lag	PAC	T	Lag	PAC	T
1	0.85	6.55	8	0.16	1.27	15	0.11	0.88
2	-0.53	-4.10	9	0.09	0.69			
3	-0.29	-2.23	10	0.08	0.63			
4	-0.08	-0.65	11	-0.03	-0.21			
5	0.07	0.53	12	-0.01	-0.05			
6	0.14	1.07	13	-0.10	-0.79			
7	0.14	1.09	14	-0.06	-0.43			

ARIMA Model: marriages

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0.8201	0.0778	10.53	0.000
SAR 12	0.6469	0.1161	5.57	0.000
Constant	4.003	1.077	3.72	0.000
Mean	62.99	16.95		

Number of observations: 60

Residuals: SS = 3907.02 (backforecasts excluded)

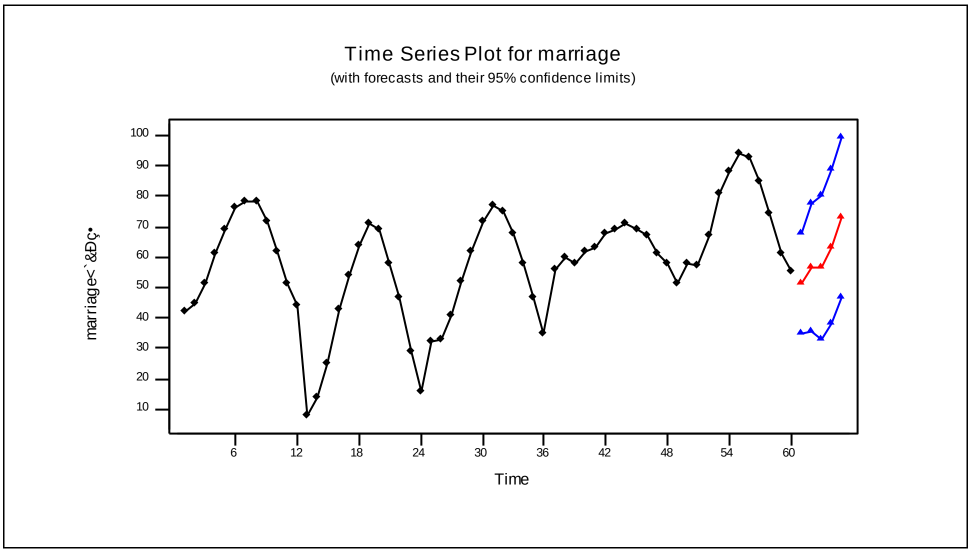
MS = 68.54 DF = 57

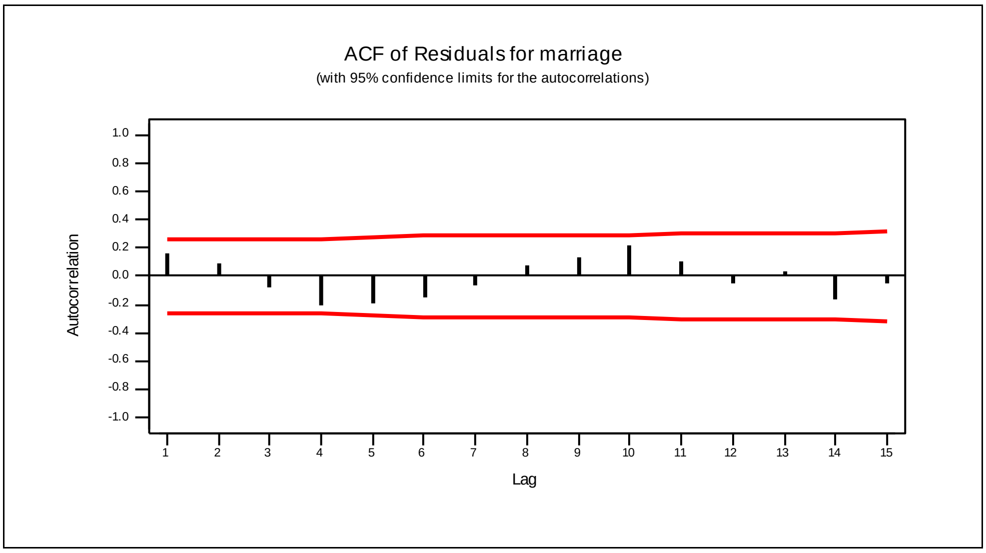
Modified Box-Pierce (Ljung-Box) Chi-Square statistic

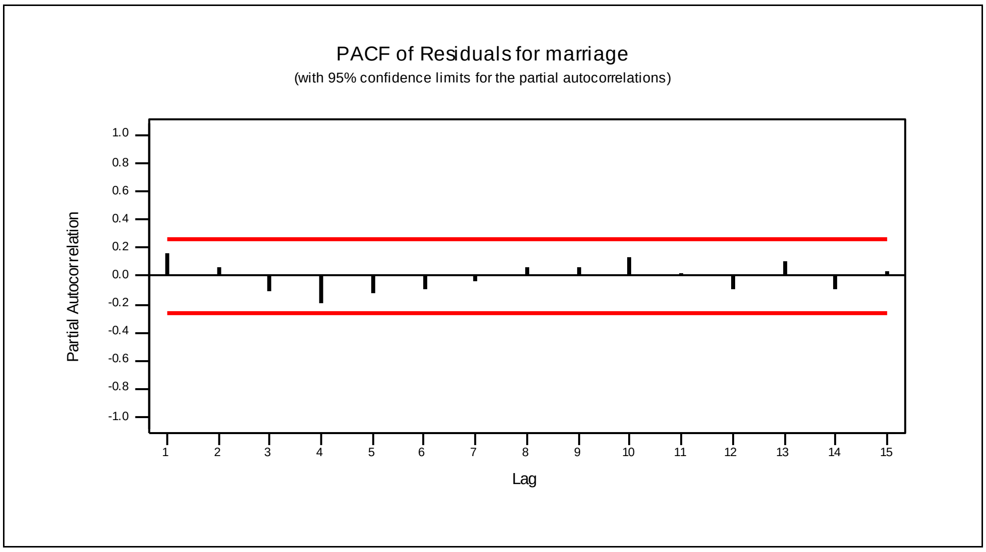
Lag	12	24	36	48
Chi-Square	15.5	33.4	44.8	63.4
DF	9	21	33	45
P-Value	0.079	0.042	0.082	0.037

Forecasts from period 60

		95 Percent Limits		
Period	Forecast	Lower	Upper	Actual
61	51.3288	35.0984	67.5592	
62	56.5594	35.5695	77.5493	
63	56.4888	32.8306	80.1469	
64	63.4298	38.1350	88.7246	
65	72.8732	46.5350	99.2115	







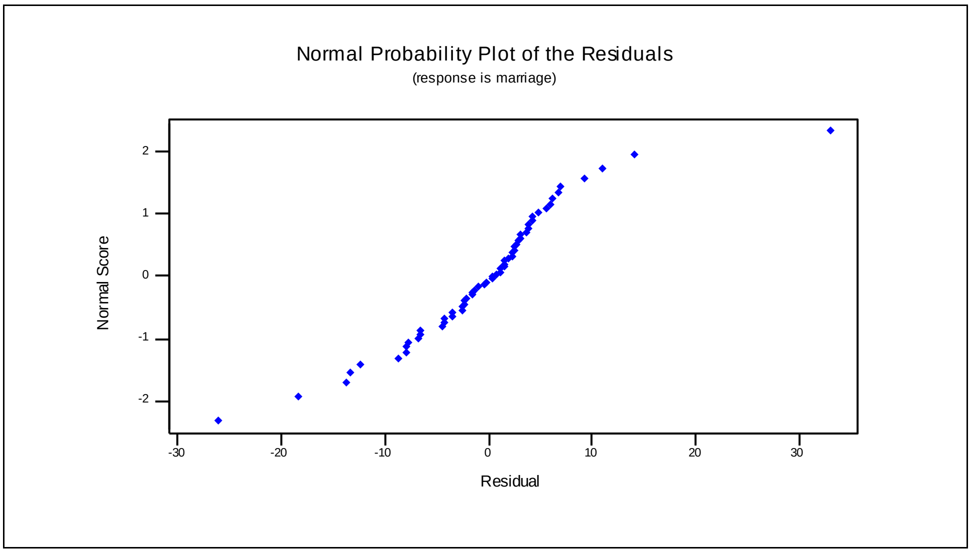


Figure 7.4 Plots of the Residual analyses

Example: Differenced data follow significant dependence on the observations at every lag 4.There is not significant effect of AR or MA parts. Therefore, the model which fits best to the data is ARIMA(0,1,0)x$(0,0,0)_4$.

$$(1-\Phi B^4)(1-B)X_t = Z_t \Rightarrow X_t - X_{t-1} = \Phi(X_{t-4} - X_{t-5}) + Z_t$$

Dependent Variable: DIF_MARINE
Method: Least Squares
Date: 06/21/08 Time: 18:23
Sample (adjusted): 1984Q1 1992Q3
Included observations: 35 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.302565	6.017135	0.050284	0.9602
DIF_MARINE(4)	0.910204	0.068338	13.31915	0.0000

R-squared	0.843156	Mean dependent var	4.788571
Adjusted R-squared	0.838403	S.D. dependent var	88.41493
S.E. of regression	35.54204	Akaike info criterion	10.03475
Sum squared resid	41686.80	Schwarz criterion	10.12363
Log likelihood	-173.6082	F-statistic	177.3999
Durbin-Watson stat	2.620756	Prob(F-statistic)	0.000000

Residual checks: Errors should be uncorrelated

Correlogram of Residuals

Date: 06/21/08 Time: 18:23
Sample: 1984Q1 1992Q3
Included observations: 35

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.321	-0.321	3.9340	0.047
		2	0.197	0.104	5.4515	0.065
		3	0.030	0.136	5.4885	0.139
		4	0.096	0.138	5.8696	0.209
		5	0.031	0.079	5.9116	0.315
		6	-0.100	-0.132	6.3617	0.384
		7	0.001	-0.130	6.3617	0.498
		8	-0.314	-0.409	11.086	0.197
		9	0.149	-0.069	12.196	0.202
		10	-0.374	-0.282	19.444	0.035
		11	0.127	0.033	20.312	0.041
		12	-0.233	-0.049	23.363	0.025
		13	-0.051	-0.093	23.516	0.036
		14	0.044	0.021	23.633	0.051
		15	-0.009	0.044	23.638	0.071
		16	0.001	-0.119	23.638	0.098

Correlogram of Residuals Squared

Date: 06/21/08 Time: 18:04
Sample: 1984Q1 1992Q3
Included observations: 35
Q-statistic probabilities adjusted for 1 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.032	-0.032	0.0393	
		2	-0.123	-0.124	0.6283	0.428
		3	0.368	0.366	6.1233	0.047
		4	-0.174	-0.211	7.3914	0.060
		5	-0.047	0.068	7.4884	0.112
		6	-0.050	-0.291	7.5981	0.180
		7	-0.110	0.094	8.1558	0.227
		8	-0.005	-0.146	8.1570	0.319
		9	-0.192	-0.063	9.9977	0.265
		10	0.059	0.037	10.180	0.336
		11	-0.034	-0.111	10.241	0.420
		12	-0.109	0.049	10.915	0.450
		13	0.265	0.136	15.049	0.239
		14	-0.111	-0.120	15.808	0.260
		15	-0.089	-0.035	16.323	0.294
		16	0.304	0.131	22.608	0.093

Heteroskedasticity refers to non-constant variance. Based on the correlogram of the residuals and squared residuals, we can conclude that there exists no serial correlation on the residual squares. Therefore, the series is homoscedastic.

Forecasting:

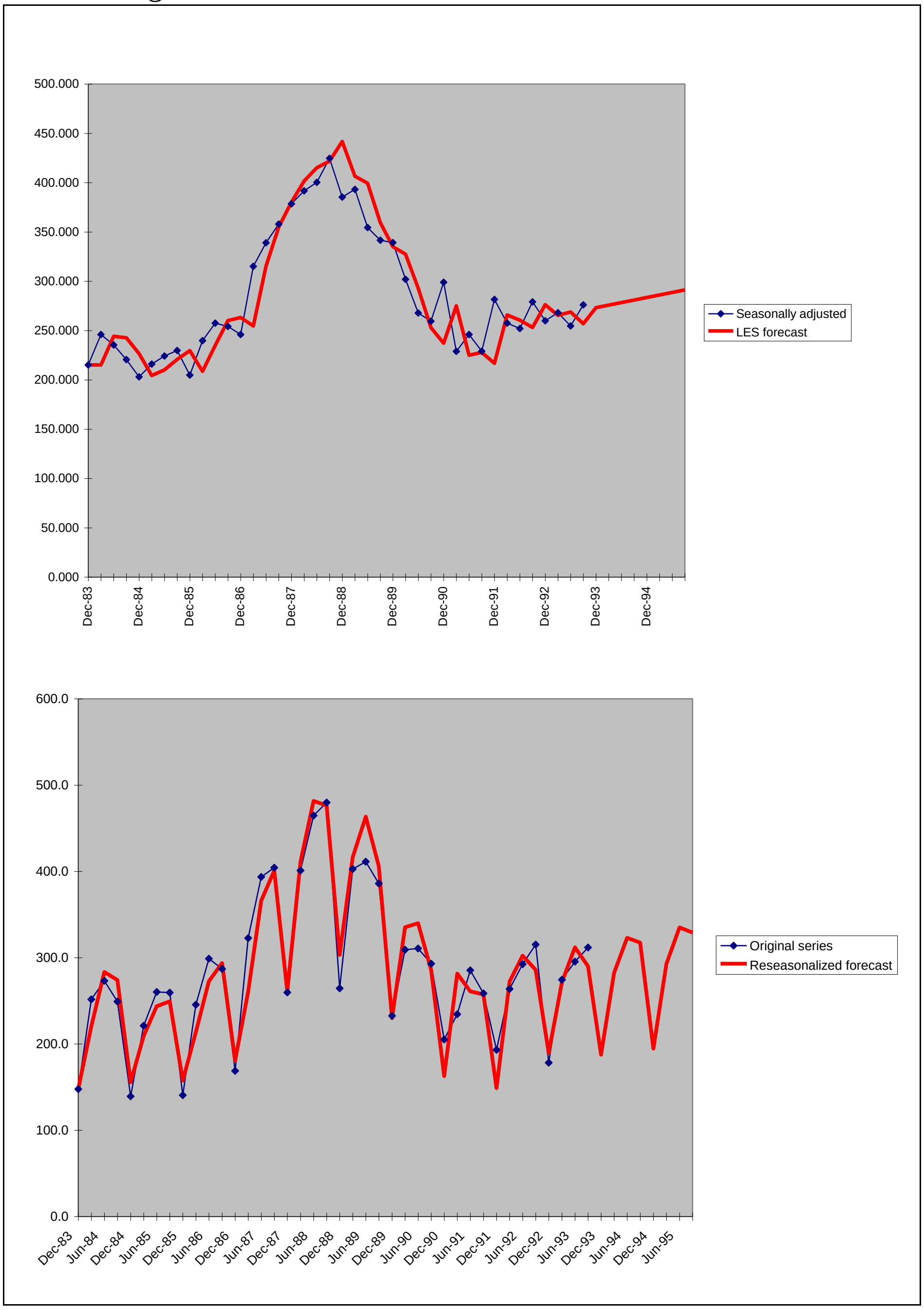


Figure 7.5. Forecast values based on classical approach (above) and the multiplicative approach (below)

Exercises:

1. For the multiplicative seasonal ARIMA(1,2,1)x$(2,1,2)_6$ write the model in original form.
2. Express ARIMA(0,1,0)x$(1,1,1)_3$ in equation form.
3. Express ARIMA(1,1,0)x$(1,1,1)_3$ in equation form.

Chapter 8
ARCH(m) (Autoregressive Conditional Heteroskedasticity) Models

Stationary and nonstationary processes presented till now are assumed to have constant variance (homoskedasticity)-unconditional variance. However, some series exhibit periods of unusually large volatility resulting in non-constant variance. The volatility in the series is modeled by taking into account the conditional variance.

Consider the return or relative gain of a stock at time t is

$X_t = \dfrac{Y_t - Y_{t-1}}{Y_{t-1}}$ where Y_t is the price of the stock at time t.

From here, one can write, $Y_t = (1 + X_t)Y_{t-1}$. Taking the logarithm of both sides and first difference yields

$$\ln(Y_t) = \ln(1 + X_t) + \ln(Y_{t-1})$$

$$\nabla \ln Y_t = \ln(Y_t) - \ln(Y_{t-1}) = \ln(1 + X_t) + \ln(Y_{t-1}) - \ln(Y_{t-1})$$
$$\nabla \ln Y_t = \ln(1 + X_t)$$

If the percent change, X_t, stays relatively small in magnitude, then

$\ln(1 + X_t) \approx p_t$ and $\nabla \ln Y_t \approx p_t$. Therefore, the highly volatile periods tend to be clustered together.

ARCH(1)

Let $X_t = \sigma_t Z_t$ and

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2$$

where $Z_t \sim$ Gaussian WN(0,1)

The conditional distribution of X_t given X_{t-1} is $X_t \mid X_{t-1} \sim \mathrm{N}(0, \alpha_0 + \alpha_1 \mathrm{X}_{t-1}^2)$

Representation of ARCH(1) as AR(1)

$$X_t^2 = \sigma_t^2 Z_t^2$$
$$\underline{-(\alpha_0 + \alpha_1 X_{t-1}^2 = \sigma_t^2)}$$
$$X_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \upsilon_t, \quad \text{where} \quad \upsilon_t = \sigma_t^2 (Z_t^2 - 1) \quad \text{and} \quad Z_t^2 \sim Chi - square(1)$$

The Properties of ARCH process are:

Let the series contain $X_t = \; X_t, X_{t-1}, ...$

1. $E \; X_t \; = E \; \sigma_t Z_t \underset{by\,independence}{=} E \; \sigma_t \; E \; Z_t \; = 0$

2. $Var \; X_t \; = \left[X_t^2 \right] = E\left[\sigma_t^2 Z_t^2 \right] = E\left[\sigma_t^2 \right] E\left[Z_t^2 \right] = E\left[\sigma_t^2 \right]$

$$\sigma^2 = Var \; X_t \; = E\left[\sigma_t^2 \right] = \alpha_0 + \alpha_1 E\left[X_{t-1}^2 \right]$$

$$\sigma^2 = \alpha_0 + \alpha_1 \sigma^2 \Rightarrow \sigma^2 - \alpha_1 \sigma^2 = \alpha_0 \Rightarrow \sigma^2 = \frac{\alpha_0}{1-\alpha_1}$$

3. $$\begin{aligned} &\gamma \; X_{t+h}, X_t \; = E \; X_{t+h} X_t \; = E\left[E\left[X_t X_{t+h} \middle| X_{t+h-1} \right] \right] \\ &\gamma \; X_{t+h}, X_t \; = E\left[X_t E\left[X_{t+h} \middle| X_{t+h-1} \right] \right] = 0 \end{aligned}$$

4. If $\alpha_1 < 1$, the process is White Noise and its unconditional distribution is symmetrical around zero (leptokurtic distribution: see below)

5. If $3\alpha_1^2 < 1$ in addition to property 4, X_t^2 is a causal AR(1) process with

$$\rho(h) = \alpha_1^h, \quad h > 0$$

6. If $3\alpha_1 \geq 1$, in addition to property 5, then X_t^2 is strictly stationary with infinite variance.

GARCH(m,r)

Generalized ARCH model with order m,r is

$$X_t = \sigma_t Z_t$$

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{m} \alpha_j X_{t-j}^2 + \sum_{j=1}^{r} \beta_j \sigma_{t-j}^2$$

GARCH(1,1) is

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta_1 \sigma_{t-1}^2; \quad \alpha_1 + \beta_1 < 1$$

The process can be expressed as ARMA(1,1)

$$X_t^2 = \alpha_0 + (\alpha_1 + \beta_1) X_{t-1}^2 + \sigma_t^2 (Z_t^2 - 1) - \beta_1 (Z_t^2 - 1)$$

$$X_t^2 - \sigma_t^2 = \sigma_t^2 (Z_t^2 - 1)$$

$$\underline{-\beta_1 (X_{t-1}^2 - \sigma_{t-1}^2) = \beta_1 \sigma_{t-1}^2 (Z_{t-1}^2 - 1)}$$

$$\sigma_t^2 - \beta_1 \sigma_{t-1}^2 = \alpha_0 + \alpha_1 X_{t-1}^2$$

Contribution of descriptive statistics on Model determination:
The skewness and the kurtosis of a distribution are

$$S(y) = E\left[\frac{(y-\mu_y)^3}{\sigma_y^3}\right], \quad K(y) = E\left[\frac{(y-\mu_y)^4}{\sigma_y^4}\right]$$

respectively. If the distribution is normal, K(y)=3, S(y)=0. Thereofre, for any distribution, K(y) – 3 is called the excess kurtosis .

Under normality assumption, $\hat{s}(y)$ and $\hat{k}(y)$ are distributed asymptotically as normal with zero mean and variances 6/T and 24/T, respectively.

Financial data often exhibit leptokurtosis, i.e. a kurtosis higher than 3 or an excess kurtosis higher than 0. We consider such return pattern especially for high frequency data, for example daily data. For monthly, quarterly or yearly aggregated data the distribution turns more towards a normal distribution.

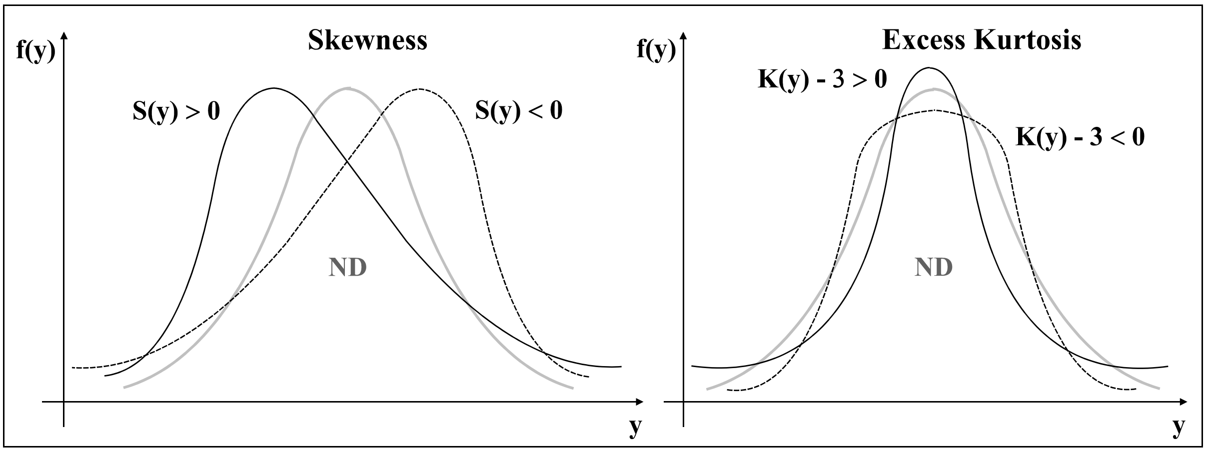


Figure 8.1. The forms of Skewness and Kurtosis for different values

b) Test of Normality

Additional to Q-Q plot and goodness of fit tests Jarque-Bera test statistic measures the difference of the skewness and kurtosis of the series with those from the normal distribution. The statistic is computed as:

$$JB = \left(\frac{T}{6}\right) \cdot \left(\hat{s}^2 + \frac{1}{4}(\hat{k}-3)^2\right) \sim \chi^2(2)$$

Under the null hypothesis of a normal distribution, the Jarque-Bera statistic is distributed as χ^2 with 2 degrees of freedom. [H_0: The distribution is Normal]
1% ≈ 9,21 ;5% ≈ 5,99. The test is only adequate for large samples, whereas for small samples you have to interpret it cautiously.

Example 1: Dax TR returns between 1965-2003 (source R.Fuess)

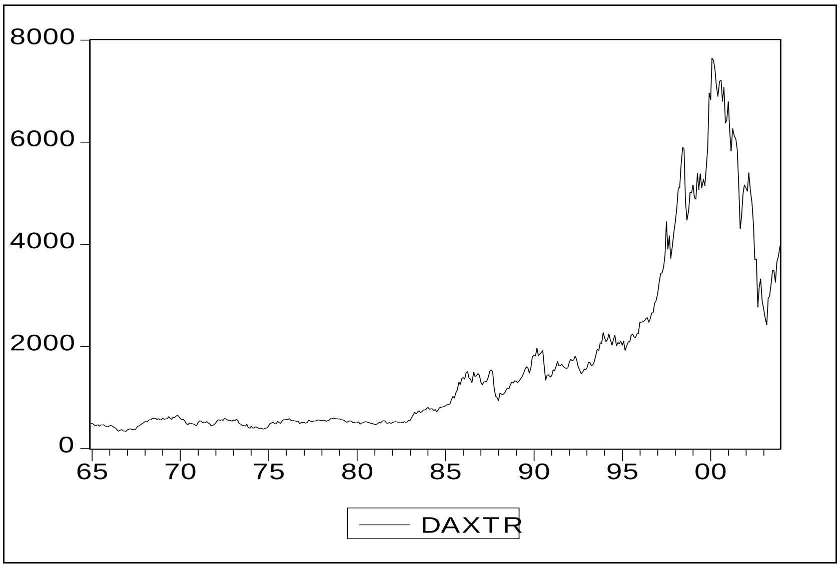


Figure 8.2. Original series

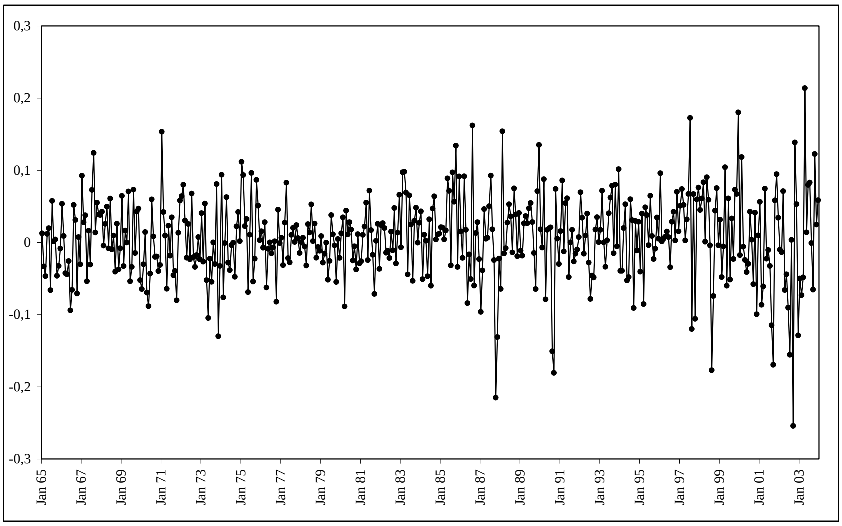


Figure 8.3. Differenced series

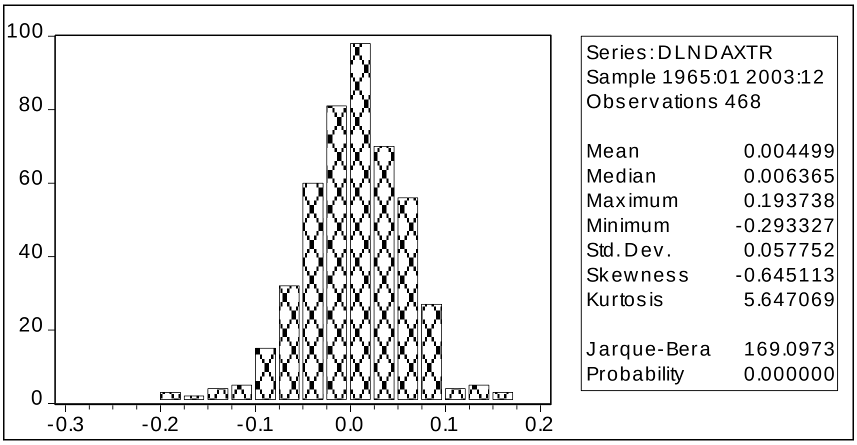


Figure 8.4. Histogram of the differenced series, Normality test and descriptive statistics

Example 2: The series contain observations from Istanbul Stock Exchange (National Defence) daily from 07/03/2000 to 31/12/2007.

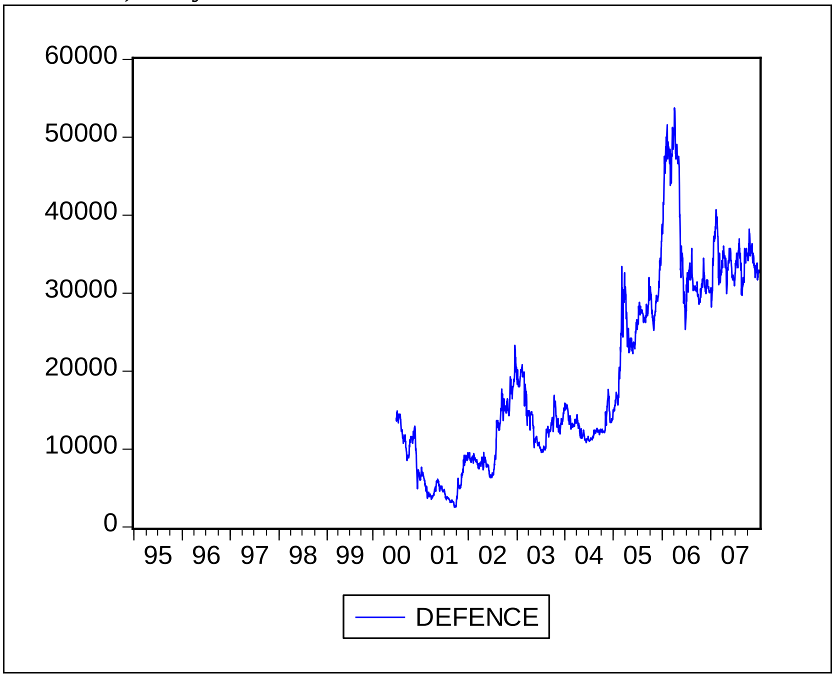
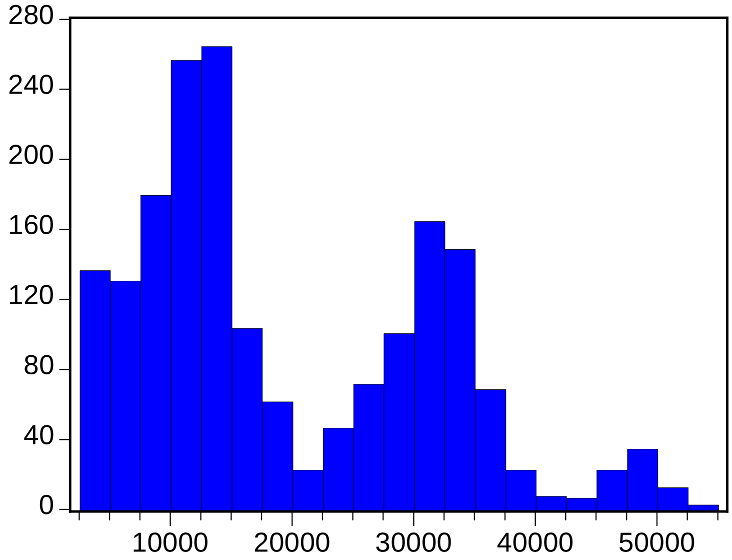


Figure 8.5. The plot of the original series

Correlogram of DEFENCE

Date: 06/22/08 Time: 12:49
Sample: 1/02/1995 12/31/2007
Included observations: 1874

	AC	PAC	Q-Stat	Prob
1	0.998	0.998	1869.1	0.000
2	0.996	-0.014	3731.1	0.000
3	0.994	0.007	5586.2	0.000
4	0.992	0.014	7434.5	0.000
5	0.989	-0.034	9275.6	0.000
6	0.987	0.015	11110.	0.000
7	0.985	-0.021	12937.	0.000
8	0.983	0.006	14756.	0.000
9	0.981	0.020	16569.	0.000
10	0.979	-0.013	18375.	0.000
11	0.976	-0.069	20173.	0.000
12	0.974	-0.006	21963.	0.000
13	0.971	-0.002	23745.	0.000
14	0.969	-0.021	25518.	0.000
15	0.966	-0.007	27283.	0.000
16	0.963	-0.052	29038.	0.000
17	0.960	0.014	30785.	0.000
18	0.958	0.012	32523.	0.000
19	0.955	0.001	34252.	0.000
20	0.953	0.017	35972.	0.000



Series: DEFENCE
Sample 1/02/1995 12/31/2007
Observations 1874

Mean	19480.52
Median	14667.89
Maximum	53761.54
Minimum	2542.530
Std. Dev.	11952.45
Skewness	0.634921
Kurtosis	2.414014
Jarque-Bera	152.7216
Probability	0.000000

Figure 8.6 The histogram and Normality check for the original data. (skewed to right)

Test the stationarity of original data by using ADF Test :

Augmented Dickey-Fuller Unit Root Test on DEFENCE

Null Hypothesis: DEFENCE has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic based on SIC, MAXLAG=24)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.056476	0.7347
Test critical values:	1% level	-3.433640	
	5% level	-2.862880	
	10% level	-2.567530	

Test the stationarity of differenced data by using ADF Test :

Augmented Dickey-Fuller Unit Root Test on D(DEFENCE)

Null Hypothesis: D(DEFENCE) has a unit root
Exogenous: Constant
Lag Length: 0 (Automatic based on SIC, MAXLAG=24)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-42.67318	0.0000
Test critical values:	1% level	-3.433642	
	5% level	-2.862881	
	10% level	-2.567531	

The series becomes stationary after differencing with order 1.

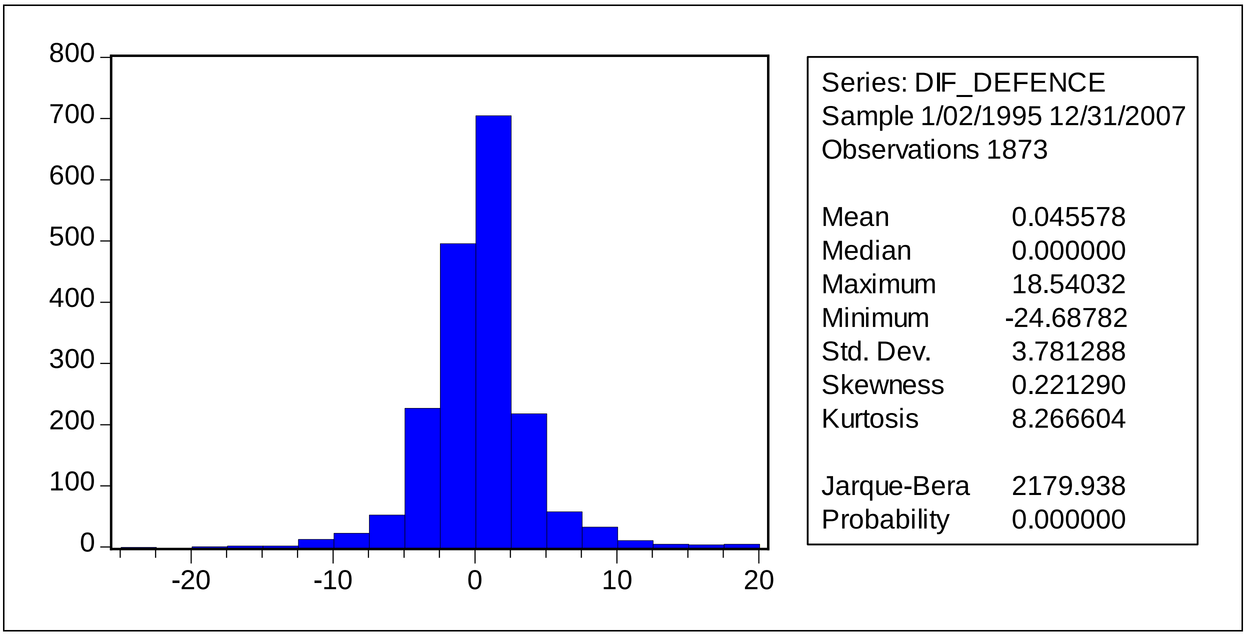


Figure 8.7. The histogram of the difference data show high kurtosis and is symmetric. Normality test is done (Jarque-Bera)

As the Kurtosis is high, this is a sign for ARCH effects

Correlogram of DIF_DEFENCE

Date: 06/21/08 Time: 19:19
Sample: 1/02/1995 12/31/2007
Included observations: 1873

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.024	0.024	1.0900	0.296
		2	0.009	0.009	1.2502	0.535
		3	0.004	0.004	1.2851	0.733
		4	0.005	0.004	1.3273	0.857
		5	-0.007	-0.008	1.4325	0.921
		6	0.009	0.009	1.5696	0.955
		7	0.044	0.044	5.2712	0.627
		8	0.033	0.031	7.3664	0.498
		9	-0.003	-0.005	7.3784	0.598
		10	0.056	0.055	13.270	0.209
		11	-0.054	-0.057	18.674	0.067
		12	-0.006	-0.004	18.736	0.095
		13	-0.015	-0.015	19.161	0.118
		14	0.022	0.021	20.091	0.127
		15	0.014	0.012	20.460	0.155
		16	0.023	0.020	21.485	0.161
		17	0.007	0.002	21.570	0.202
		18	0.000	0.000	21.570	0.252
		19	-0.032	-0.027	23.454	0.218
		20	0.022	0.022	24.399	0.225

The correlogram of the differenced data shows that the returns are not correlated.

Correlogram of SQUARED_DIF_DEFENCE

Date: 06/21/08 Time: 19:20
Sample: 1/02/1995 12/31/2007
Included observations: 1873

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.273	0.273	139.46	0.000
		2	0.205	0.141	218.19	0.000
		3	0.186	0.110	283.05	0.000
		4	0.112	0.020	306.82	0.000
		5	0.172	0.111	362.32	0.000
		6	0.135	0.045	396.66	0.000
		7	0.101	0.016	415.91	0.000
		8	0.111	0.035	439.05	0.000
		9	0.047	-0.029	443.13	0.000
		10	0.171	0.135	498.38	0.000
		11	0.134	0.042	532.39	0.000
		12	0.069	-0.023	541.47	0.000
		13	0.046	-0.039	545.43	0.000
		14	0.014	-0.028	545.80	0.000
		15	0.073	0.044	555.91	0.000
		16	0.060	0.004	562.63	0.000
		17	0.037	-0.007	565.20	0.000
		18	0.070	0.032	574.55	0.000
		19	0.074	0.053	584.82	0.000
		20	0.075	0.019	595.46	0.000

However, the squared returns are correlated. Here, the Coefficient for AR and MA terms are not significant. Therefore, we regress differenced data on constant and the variance.

Dependent Variable: DIF_DEFENCE
Method: ML - ARCH (Marquardt) - Normal distribution
Date: 06/21/08 Time: 19:24
Sample (adjusted): 7/04/2000 12/31/2007
Included observations: 1873 after adjustments
Convergence achieved after 39 iterations
Variance backcast: ON
GARCH = C(2) + C(3)*RESID(-1)^2 + C(4)*GARCH(-1)

	Coefficient	Std. Error	z-Statistic	Prob.
C	0.005147	0.073982	0.069572	0.9445
Variance Equation				
C	0.508804	0.049227	10.33582	0.0000
RESID(-1)^2	0.098621	0.009294	10.61168	0.0000
GARCH(-1)	0.866373	0.009847	87.98028	0.0000

R-squared	-0.000114	Mean dependent var	0.045578
Adjusted R-squared	-0.001720	S.D. dependent var	3.781288
S.E. of regression	3.784538	Akaike info criterion	5.275032
Sum squared resid	26769.18	Schwarz criterion	5.286853
Log likelihood	-4936.067	Durbin-Watson stat	1.951180

The Model is $\sigma_t^2 = 0.508 + 0.0986 X_{t-1}^2 + 0.866 \sigma_{t-1}^2$

Residual Checks:

Correlogram of Standardized Residuals

Date: 06/21/08 Time: 19:25
Sample: 7/04/2000 12/31/2007
Included observations: 1873

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.022	0.022	0.8909	0.345
		2	0.035	0.035	3.1968	0.202
		3	0.009	0.007	3.3346	0.343
		4	0.019	0.018	4.0136	0.404
		5	0.020	0.018	4.7422	0.448
		6	0.007	0.005	4.8381	0.565
		7	-0.014	-0.016	5.1919	0.637
		8	0.039	0.038	7.9917	0.434
		9	0.017	0.015	8.5174	0.483
		10	0.057	0.054	14.740	0.142
		11	-0.022	-0.026	15.673	0.154
		12	-0.010	-0.014	15.881	0.197
		13	-0.011	-0.012	16.110	0.243
		14	0.022	0.020	17.005	0.256
		15	0.014	0.014	17.357	0.298
		16	0.021	0.019	18.171	0.314
		17	0.003	0.002	18.185	0.377
		18	-0.012	-0.019	18.457	0.426
		19	-0.047	-0.049	22.601	0.255
		20	0.015	0.014	23.001	0.289

Correlogram of Standardized Residuals Squared

Date: 06/21/08 Time: 19:26
Sample: 7/04/2000 12/31/2007
Included observations: 1873

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.009	0.009	0.1541	0.695
		2	-0.011	-0.011	0.3803	0.827
		3	0.001	0.001	0.3812	0.944
		4	-0.011	-0.011	0.6173	0.961
		5	0.020	0.020	1.3707	0.927
		6	-0.008	-0.009	1.4953	0.960
		7	-0.006	-0.005	1.5529	0.980
		8	-0.005	-0.005	1.5937	0.991
		9	-0.020	-0.020	2.3836	0.984
		10	0.076	0.076	13.220	0.212
		11	0.009	0.007	13.376	0.269
		12	0.004	0.005	13.400	0.341
		13	-0.008	-0.008	13.517	0.409
		14	-0.014	-0.012	13.890	0.458
		15	-0.001	-0.004	13.892	0.534
		16	-0.008	-0.008	14.028	0.597
		17	-0.017	-0.017	14.591	0.625
		18	-0.010	-0.010	14.796	0.676
		19	0.028	0.031	16.247	0.641
		20	0.002	-0.004	16.255	0.701

Both correlogram show that residuals and squared residuals are white noise. This şeads us to conclude that there exists NO ARCH effects left.

Chapter 9
Vector Autoregressive Analysis

9.1. Vector autoregression (VAR) is an econometric model used to capture the evolution and the interdependencies between multiple time series, generalizing the univariate AR models. All the variables in a VAR are treated symmetrically by including for each variable an equation explaining its evolution based on its own lags and the lags of all the other variables in the model. A VAR model describes the evolution of a set of *k* variables measured over the same sample period ($t \in$T) as a linear function of only their past evolution. The variables are collected in a *k x 1* vector y_t, which has as the ith element $y_{i,}$, the time *t* observation of variable y_i.

For example, if the i^{th} variable is GDP, then $y_{i,t}$ is the value of GDP at t.

A ***(reduced) p-th order VAR***, ***VAR(p)***, is

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \ldots + A_p y_{t-p} + \varepsilon_t$$

where c is a *k x 1* vector of constants (**intercept**), A_i is a *k x k* matrix (for every $i = 1, ..., p$) and ε_t is a *k x 1* vector of error terms satisfying the conditions

1. $E[\varepsilon_t] = 0$. that is, every error term has mean zero;

   The structural, economic shocks which drive the dynamics of the economic variables are assumed to be independent, which implies zero correlation between error terms as a desired property. This is helpful for separating out the effects of economically unrelated influences in the VAR.
   For instance, there is no reason why an oil price shock (as an example of a supply shock) should be related to a shift in consumers' preferences towards a style of clothing (as an example of a demand shock); therefore one would expect these factors to be statistically independent.

2. $E[\varepsilon_t \varepsilon_t'] = \Omega$ the contemporaneous covariance matrix of errors;

   (*n x n* positive definite matrix); This is a desirable feature especially when using low frequency data. For example, an indirect tax rate increase would not affect tax revenues the day the decision is announced, but one could find an effect in that quarter's data.

3. $E[\varepsilon_t \varepsilon_{t-k}'] = 0$ for any k>0 ; there is no correlation across time;

   no serial correlation in individual error terms.

For order p the set of equations becomes

$$
\begin{bmatrix} y_{1t} \\ y_{2t} \\ y_{3t} \\ . \\ y_{kt} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ c_3 \\ . \\ c_k \end{bmatrix} + \begin{bmatrix} a_{1,1}^1 & .. & a_{1,k}^1 \\ .. & .. & .. \\ a_{1,1}^1 & .. & a_{1,k}^1 \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \\ . \\ y_{k,t-1} \end{bmatrix} + \begin{bmatrix} a_{1,1}^2 & .. & a_{1,k}^2 \\ .. & .. & .. \\ a_{1,1}^2 & .. & a_{1,k}^2 \end{bmatrix} \begin{bmatrix} y_{1,t-2} \\ y_{2,t-2} \\ y_{2,t-2} \\ . \\ y_{k,t-2} \end{bmatrix} + .. + \begin{bmatrix} a_{1,1}^p & .. & a_{1,k}^p \\ .. & .. & .. \\ a_{1,1}^p & .. & a_{1,k}^p \end{bmatrix} \begin{bmatrix} y_{1,t-p} \\ y_{2,t-p} \\ y_{3,t-p} \\ . \\ y_{k,t-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ . \\ \varepsilon_{kt} \end{bmatrix}
$$

The *l*-periods back observation y_{t-l} is called the *l*-th ***lag*** of *y*. Thus, a p-th order VAR is also called a ***VAR with p** lags*

Order of integration of the variables

Note that all the variables used have to be of the same order of integration. We have the following cases:

- All the variables are I(0) (stationary):
  one is in the standard case, ie. a VAR in level
- All the variables are I(d) (non-stationary) with d>1:
  - The variables are cointegrated:
    the error correction term has to be included in the VAR. The model becomes a Vector error correction model (VECM) which can be seen as a restricted VAR.
  - The variables are not cointegrated:
    the variables have first to be differenced d times and one has a VAR in difference

Example: VAR(1)

Suppose $\{y_{1t}\}_{t\epsilon T}$ denote real GDP growth, $\{y_{2t}\}_{t\epsilon T}$ denote inflation

$$
\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}
$$

$$
y_{1t} = c_1 + A_{11} y_{1,t-1} + A_{12} y_{2,t-1} + \varepsilon_{1t}
$$

$$
y_{2t} = c_2 + A_{21} y_{1,t-1} + A_{22} y_{2,t-1} + \varepsilon_{2t}
$$

- One equation for each variable in the model.
- The current (time *t*) observation of each variable depends on its own lags as well as on the lags of each other variable in the VAR.

Expressing VAR(p) as VAR(1)

The transformation amounts to merely stacking the lags of the *VAR(p)* variable in the new *VAR(1)* dependent variable and appending identities to complete the number of equations.

Example: *VAR(2)* model

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + e_t$$

can be recast as the *VAR(1)* model

$$\begin{bmatrix} y_t \\ y_{t-1} \end{bmatrix} = \begin{bmatrix} c \\ 0 \end{bmatrix} + \begin{bmatrix} A_1 & A_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \end{bmatrix}$$ where I is the identity matrix.

The equivalent *VAR(1)* form is more convenient for analytical derivations and allows more compact statements.

Structural VAR* (SVAR) *with p lags

$$B_0 y_t = c_0 + B_1 y_{t-1} + B_2 y_{t-2} + ... + B_p y_{t-p} + e_t$$

where c_0 is a $k \times 1$ vector of constants, B_i is a $k \times k$ matrix, $i = 0, ..., p$, and e_t is a $k \times 1$ vector of error terms.
The main diagonal terms of the B_0 matrix (the coefficients on the i^{th} variable in the i^{th} equation) are scaled to 1.
The error terms e_t (***structural shocks***) satisfy the conditions and particularity that all the elements off the main diagonal of the covariance matrix $E(e_t e_t') = \Sigma$ are zero. That is, the structural shocks are uncorrelated.

Example: Two variable structural VAR(1) is:

$$\begin{bmatrix} 1 & b_{01} \\ b_{02} & 1 \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} = \begin{bmatrix} c_{01} \\ c_{02} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}$$

where $Var(e_i) = \sigma_i^2$, i=1,2; $cov(e_1,e_2) = 0$.

Reduced VAR
By premultiplying the structural VAR with the inverse of B_0

$$y_t = B_0^{-1} c_0 + B_0^{-1} B_1 y_{t-1} + B_0^{-1} B_2 y_{t-2} + ... + B_0^{-1} B_p y_{t-p} + B_0^{-1} e_t$$

and denoting

$$B_0^{-1} c_0 = c\,; \quad B_0^{-1} B_i = A_i \text{ , i=1,..,p;} \quad B_0^{-1} e_t = \varepsilon_t$$

one obtains the ***p-th order reduced VAR***

$$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + ... + A_p y_{t-p} + \varepsilon_t$$

Note that in the reduced form all right hand side variables are predetermined at time t. As there are no time t endogenous variables on the right hand side, no variable has a *direct* contemporaneous effect on other variables in the model.

However, the error terms in the reduced VAR are composites of the structural shocks $\varepsilon_t = B_0^{-1} e_t$.

Thus, the occurrence of one structural shock $e_{i,t}$ can potentially lead to the occurrence of shocks in all error terms $\varepsilon_{j,t}$, thus creating contemporaneous movement in all endogenous variables.
Consequently, the covariance matrix of the reduced VAR

$$\Omega = E[\varepsilon_t \varepsilon_t^{'}] = E[B_0^{-1} e_t e_t^{'} (B_0^{-1})'] = B_0^{-1} \Sigma (B_0^{-1})'$$

can have non-zero off-diagonal elements, thus allowing non-zero correlation between error terms.

10.2. Impulse Response Function

The key tool to trace short run effects with an SVAR is the impulse response function.

$y_t = c + A_1 y_{t-1} + A_2 y_{t-2} + \ldots + A_p y_{t-p} + \varepsilon_t$ can be expressed as MA(∞)

$$y_t = c + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \ldots = \Psi(B)\varepsilon_t$$

The matrix ψ_l has the interpretation $\dfrac{\partial y_{t+l}}{\partial \varepsilon_t^{'}} = \psi_l$

i.e. the row i , column j element of ψ_l identifies the consequences of a one-unit increase in the jth variable's innovation at date t (ε_{tj}) for the value of the ith variable at time t+*l*, holding all other innovations at all dates constant.

A plot of the row i, column j element of ψ_l as a function of lag *l* is called the non-orthogonalized impulse response function. It describes the response of $y_{i,t+l}$ to a one-time impulse in y_{it} with all other variables dated t or earlier held constant.

Checking for the lag length

The model should represent the observed processes as precise as possible along with attaining error terms to be at minimum. Therefore, the choice of the number of variables to be included into the model is important.

If the lag length is chosen to be too short, serial correlation among error terms become significant.

A test on the two possible choice of the order

H_0 : the model needs p+1 lags

(the coefficients of $y_{1,t-p}, y_{2,t-p}, \ldots, y_{k,t-p}$ are all zero)

H_a : the model needs p lags

Test statistic: Log Likelihood test

$$\lambda = \frac{Likelihood(\text{restrictedmodel})}{Likelihood(un\text{restrictedmodel})} \sim Chi - square$$

The test is performed to check if choosing the lag p+1 lags improves the power of the test or not.

Other measure for comparison is the Squared Residuals

$$\ln(\frac{\hat{\varepsilon}'\varepsilon}{T})$$

Compared for both models based on the statistics: Akaike Information Criterion, Schwarz information Criterion.

Example: Let following log transformed variables denote
$\{y_{1t}\}_{t\epsilon T}$ consumer price index, $\{y_{2t}\}_{t\epsilon T}$ GDP; $\{y_{2t}\}_{t\epsilon T}$ Money stock M1
$\{y_{2t}\}_{t\epsilon T}$ quarterly average of 3-month interest rate
VAR representation:

$$y_{1t} = c_1 + a^1_{1,1}y_{1,t-1} + .. + a^1_{1,k}y_{k,t-1} + a^2_{1,1}y_{1,t-2} + .. + a^2_{1,k}y_{2,t-2} + a^p_{1,1}y_{3,t-p} + .. + a^p_{1,k}y_{3,t-p} + a^p_{1,1}y_{4,t-p} + .. + a^p_{1,k}y_{4,t-p} + \varepsilon_{1t}$$
$$y_{2t} = c_2 + a^1_{2,1}y_{1,t-1} + .. + a^1_{2,k}y_{k,t-1} + a^2_{2,1}y_{2,t-2} + .. + a^2_{2,k}y_{2,t-2} + a^p_{2,1}y_{3,t-p} + .. + a^p_{2,k}y_{3,t-p} + a^p_{2,k}y_{4,t-p} + .. + a^p_{2,1}y_{4,t-p} + \varepsilon_{2t}$$
$$y_{3t} = c_2 + a^1_{3,1}y_{1,t-1} + .. + a^1_{3,k}y_{k,t-1} + a^2_{3,1}y_{2,t-2} + .. + a^2_{3,k}y_{2,t-2} + a^p_{3,1}y_{3,t-p} + .. + a^p_{3,k}y_{3,t-p} + a^p_{3,k}y_{4,t-p} + .. + a^p_{3,1}y_{4,t-p} + \varepsilon_{3t}$$
$$y_{4t} = c_2 + a^1_{4,1}y_{1,t-1} + .. + a^1_{4,k}y_{k,t-1} + a^2_{4,1}y_{2,t-2} + .. + a^2_{4,k}y_{2,t-2} + a^p_{4,1}y_{3,t-p} + .. + a^p_{4,k}y_{3,t-p} + a^p_{4,k}y_{4,t-p} + .. + a^p_{4,1}y_{4,t-p} + \varepsilon_{4t}$$

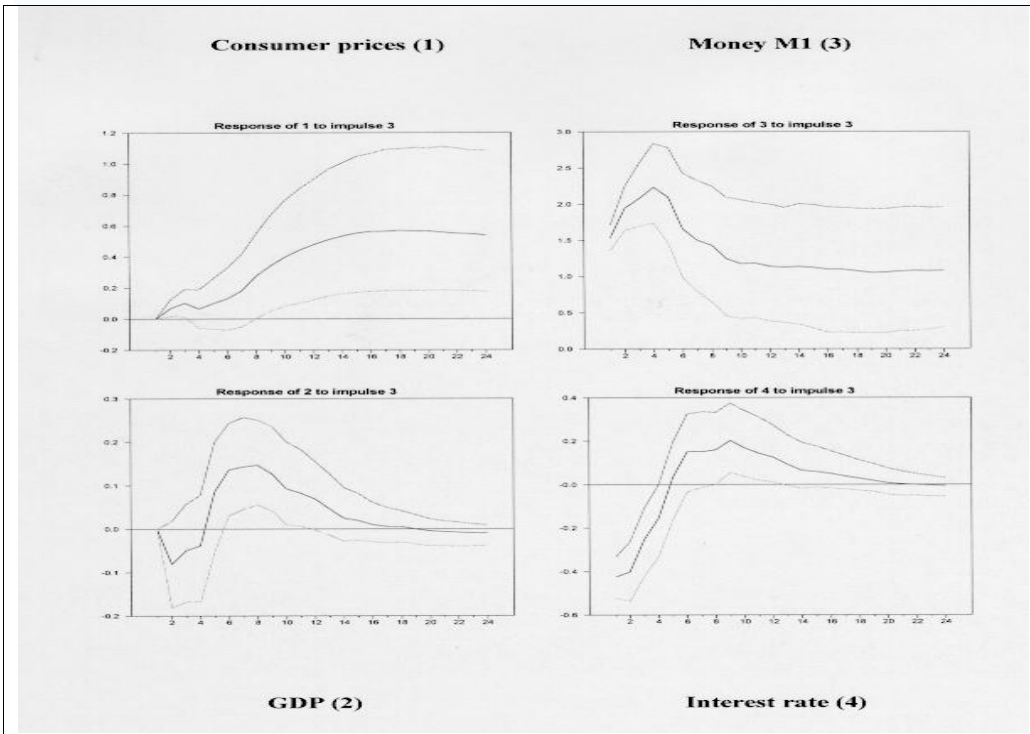


Figure 9.1 Impulse response plots of the variable above

Example: Given

$$y_t = -0.2z_t + 0.6y_{t-1} + 0.4z_{t-1} + \varepsilon_{yt}$$

$$z_t = 0.2y_{t-1} + 0.3z_{t-1} + \varepsilon_{zt}$$

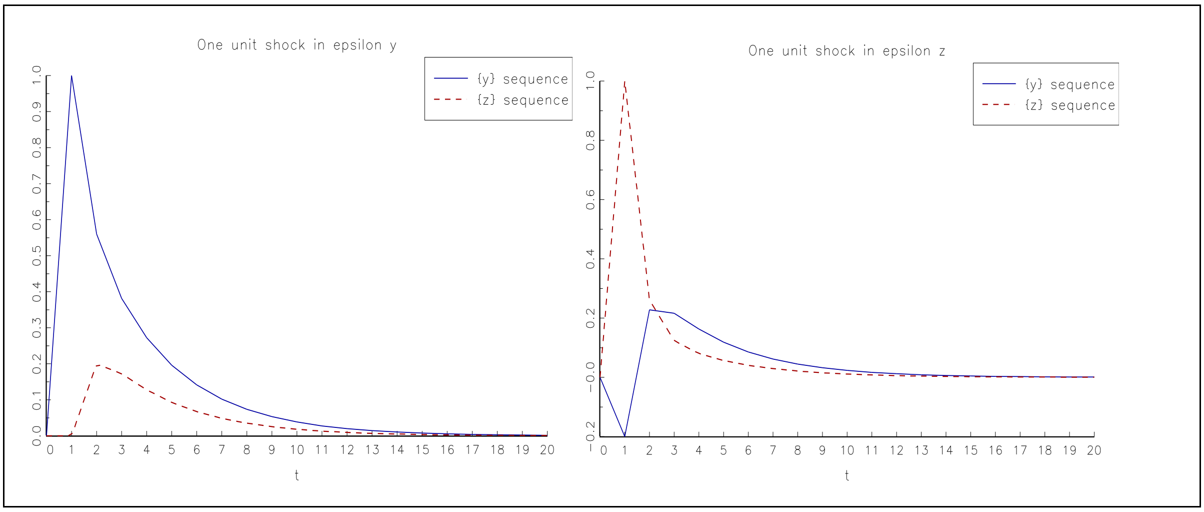


Figure 9.2 Impulse response functions of the example given

Example: Given

$$y_t = 0.6y_{t-1} + 0.4z_{t-1} + \varepsilon_{yt}$$

$$z_t = 0.2y_t + 0.2y_{t-1} + 0.3z_{t-1} + \varepsilon_{zt}$$

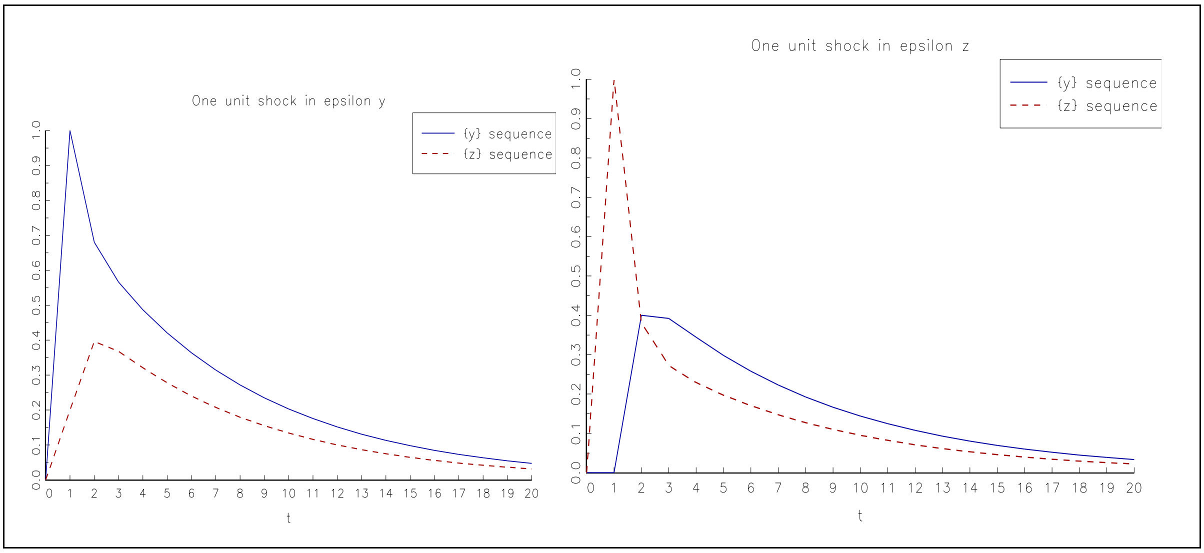


Figure 9.3 Impulse response function of the system of equations above

Example: Exchange Rate and ISE Index are the variable of the concern. VAR(2) model is fitted to the series.

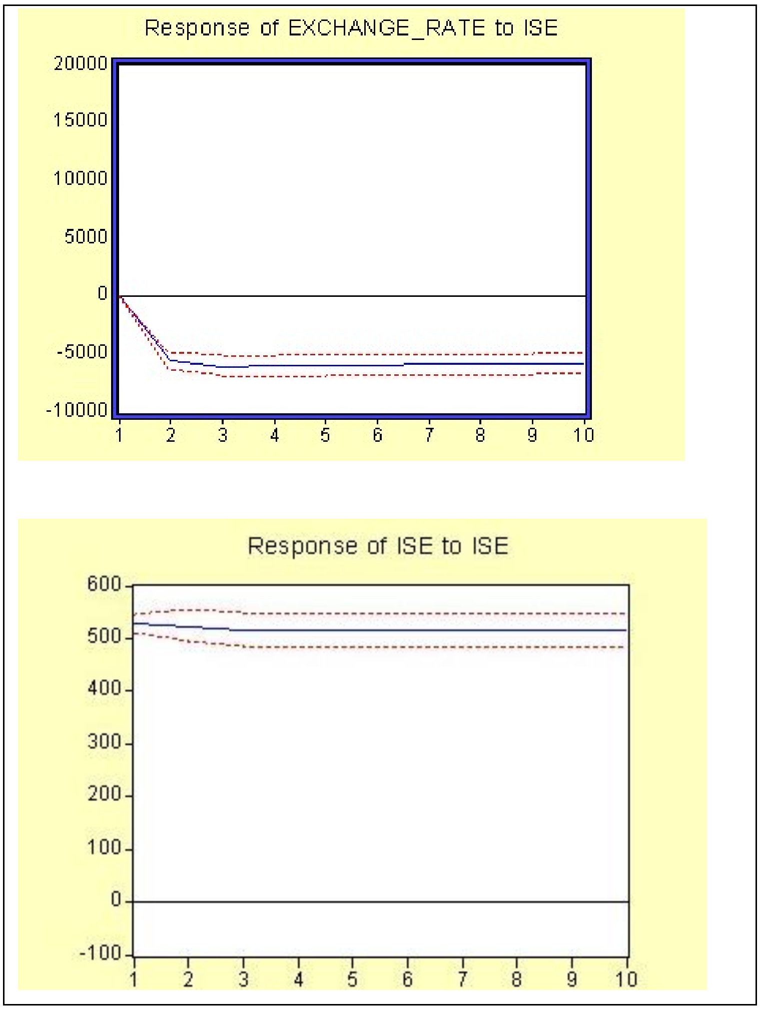


Figure 9.3 Impulse response functions of the selected variables

Vector Autoregression Estimates
Date: 07/20/08 Time: 10:59
Sample (adjusted): 1/04/2001 12/31/2007
Included observations: 1748 after adjustments
Standard errors in ( ) & t-statistics in [ ]

	EXCHANGE_RATE	ISE
EXCHANGE_RATE(-1)	1.082797	0.001270
	(0.02245)	(0.00077)
	[48.2217]	[1.65550]
EXCHANGE_RATE(-2)	-0.092452	-0.001190
	(0.02232)	(0.00076)
	[-4.14189]	[-1.56054]
ISE(-1)	-10.63907	0.993499
	(0.70086)	(0.02394)
	[-15.1799]	[41.4932]
ISE(-2)	10.57968	0.007108
	(0.70180)	(0.02398)
	[15.0751]	[0.29645]
C	15468.75	-100.0376
	(3309.16)	(113.051)
	[4.67452]	[-0.88489]
R-squared	0.991659	0.998669
Adj. R-squared	0.991640	0.998666
Sum sq. resids	4.15E+11	4.84E+08
S.E. equation	15431.88	527.1996
F-statistic	51808.14	326895.9
Log likelihood	-19335.85	-13433.53
Akaike AIC	22.12911	15.37589
Schwarz SC	22.14475	15.39153
Mean dependent	1396230.	24460.15
S.D. dependent	168779.6	14432.87

10.3. Granger causality test

Technique for determining whether one time series is useful in forecasting another.

$X_t \;_{t\varepsilon T}$ is said to Granger-cause $Y_t \;_{t\varepsilon T}$ if it can be shown, usually through a series of F-tests on lagged values of *X* (and with lagged values of *Y* also known), **that those *X* values provide statistically significant** information about future values of *Y*.

The Granger test can be applied only to pairs of variables, and may produce misleading results when the true relationship involves three or more variables.

Example: Let $Y_{1t} \;_{t\in T}$ denote GDP, $Y_{2t} \;_{t\in T}$ denote consumption

H_0 : the coefficients of $y_{1,t-p}, y_{2,t-p}, \ldots, y_{k,t-p}$ are all zero (equivalent of saying y_2 does not Granger-cause y_1)

Pairwise Granger Causality test			
Sample: 1946:1 1995:4			
Lags: 4	Obs		
Null Hypothesis	189	F-Statistic	Probability
GDP doesnot Granger Cause Cons.		1.39156	0.23866
Cons. does not Granger cause GDP		7.11192	2.4E-05

Consumption Granger Cause on GDP.

Example:

Pairwise Granger Causality Tests
Date: 07/20/08 Time: 10:40
Sample: 1/02/2001 12/31/2007
Lags: 5

Null Hypothesis:	Obs	F-Statistic	Probability
INTEREST_RATE does not Granger Cause EXCHANGE_RATE	1745	28.3482	1.1E-27
EXCHANGE_RATE does not Granger Cause INTEREST_RATE		32.1459	2.0E-31
ISE does not Granger Cause EXCHANGE_RATE	1745	58.2545	3.6E-56
EXCHANGE_RATE does not Granger Cause ISE		2.31559	0.04151
GLOBAL does not Granger Cause EXCHANGE_RATE	1745	21.4690	6.7E-21
EXCHANGE_RATE does not Granger Cause GLOBAL		1.16105	0.32611
ISE does not Granger Cause INTEREST_RATE	1745	12.2991	9.7E-12
INTEREST_RATE does not Granger Cause ISE		0.10286	0.99158
GLOBAL does not Granger Cause INTEREST_RATE	1745	2.98831	0.01084
INTEREST_RATE does not Granger Cause GLOBAL		1.48645	0.19105
GLOBAL does not Granger Cause ISE	1745	20.7727	3.3E-20
ISE does not Granger Cause GLOBAL		1.91422	0.08894

9.4 Cointegration

If two or more series are themselves non-stationary, but a linear combination of them is stationary, then the series are said to be cointegrated.

Example:
A stock market index and the price of its associated follow a random walk by time. Testing the hypothesis that there is a statistically significant connection between the futures price and the spot price could now be done by testing for a cointegrating vector.

Example: $Y_{1t\ t\in T}$ Real GDP ; $Y_{2t\ t\in T}$ private investment (real)

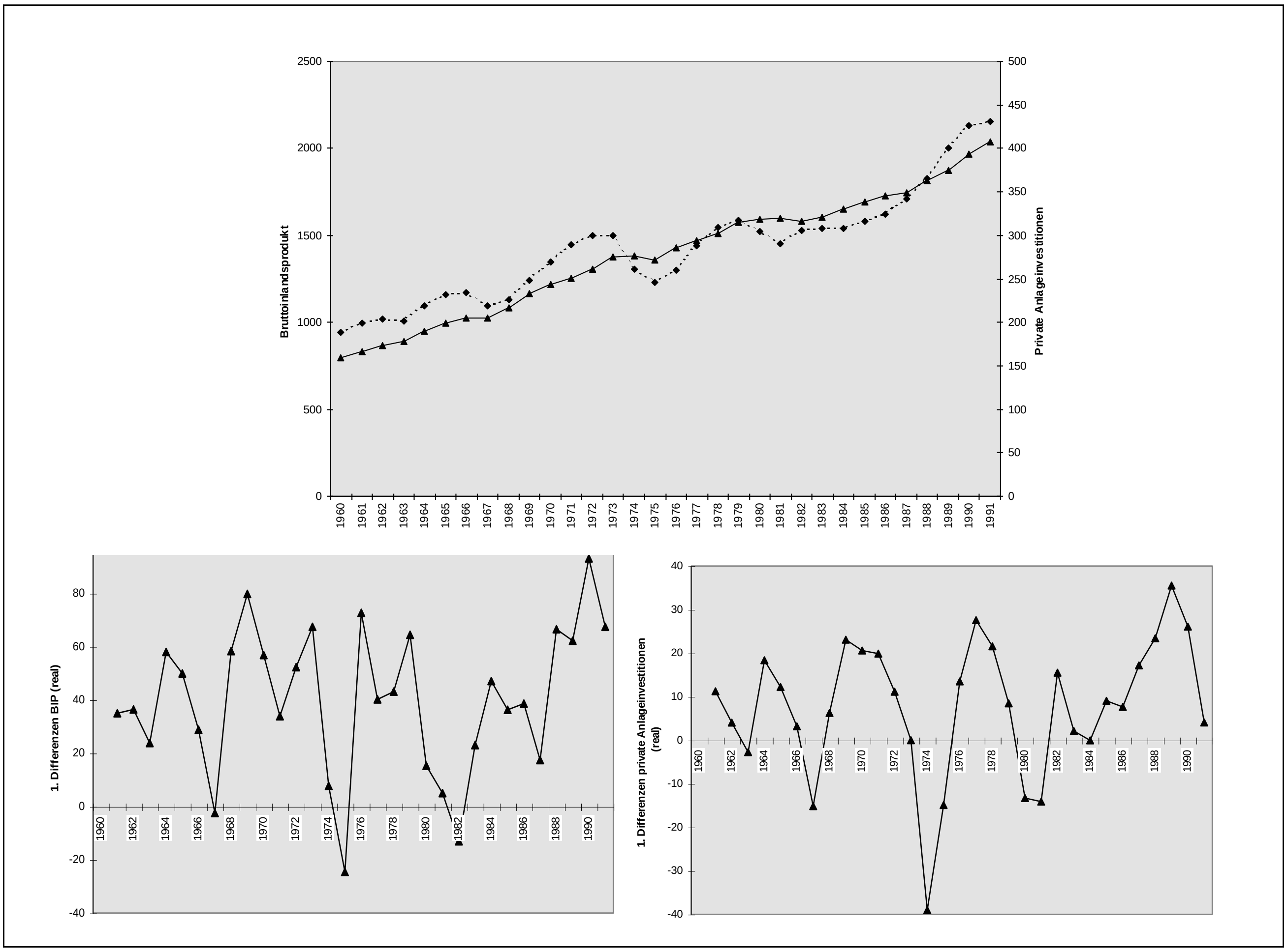


Figure 9.4. The graphs of original series and the plots of differenced series of two variables

The usual procedure for testing hypotheses concerning the relationship between non-stationary variables was to run Ordinary Least Squares (OLS) regressions on data which had initially been differenced.

Although this method is correct in large samples, cointegration provides more powerful tools when the data sets are of limited length, as most economic time-series are.

The two main methods for testing for cointegration are:

1. The Engle-Granger three-step method.
2. The Johansen procedure.

In practise, cointegration is used for such series in typical econometric tests, but it is more generally applicable and can be used for variables integrated of higher order (to detect correlated accelerations or other second-difference effects).

Multicointegration extends the cointegration technique beyond two variables, and occasionally to variables integrated at different orders.
However, these tests for cointegration assume that the cointegrating vector is constant during the period of study. In reality, it is possible that the long-run relationship between the underlying variables change (shifts in the cointegrating vector can occur). The reason for this might be technological progress, economic crises, changes in the people's preferences and behaviour accordingly, policy or regime alteration, and organizational or institutional developments. This is especially likely to be the case if the sample period is long. To take this issue into account Gregory and Hansen (1996) have introduced tests for cointegration with one unknown structural break and Hatemi-J (2007) has introduced tests for cointegration with two unknown breaks.

Example: $Y_t \quad _{t\in T}$ income ; $X_t \quad _{t\in T}$ consumption. Suppose both series are I(1).

Let $Y_t = Y_{t-1} + \varepsilon_{Yt}$ random walk having $Var[Y]_t = t\sigma_{\varepsilon_Y}{}^2$

But in the long run, $X_t - cY \approx 0$ where c is the propensity to consume

$$X_t = cY_t + \varepsilon_t$$

Consider a series of k models

$$y_{1t} = y_{1t-1} + \varepsilon_{1t}$$

$$y_{2t} = y_{2t-1} + \varepsilon_{2t} \qquad \Rightarrow \quad y_{1t}, y_{2t}, .., y_{kt}{}'$$

$$\vdots$$

$$y_{kt} = y_{kt-1} + \varepsilon_{kt}$$

is cointegrated if each series

a. nonstationary (integrated of order one)

b. there exists (at least one) linear combination $a'y_t$ a stationary process

a' is called cointegrating vector.

<u>If cointegration factor is known</u>, then the test of cointegration is reduced to a unit root test
If we can reject the null hypothesis of non-stationarity of linear, then this leads us to

conclude as a combination of I(1) time series. This shows that the data indicates cointegration.

If vector of cointegration factor is unknown, then an estimation of the cointegration relationship is required

Engle-Granger Approach

Estimation of parameters can be done by OLS estimation of the linear regression equation:

$$Y_t = \gamma_0 + \gamma_1 Y_{2t} + .. + \gamma_M Y_{Mt} + \varepsilon_t$$

Dickey-Fuller t test is applied to the OLS residuals $\hat{\varepsilon}_t$. Rejecting the null hypothesis of non-stationarity concludes "cointegration relationship" does exist.

Note: We have to keep in mind that the use of $\hat{\varepsilon}_t$ has consequences for the critical values of the ADF test. In comparison to the critical values of the usual Dickey-Fuller the critical values here are in absolute values higher and depend on the number of included variables M. If the cointegration relation contains a deterministic trend we speak about a deterministic cointegration. The critical values for M (at most equals six) are given by MacKinnon (1991). The critical values of MacKinnon are calculated by:

$$K = \beta_\infty + \beta_1 T^{-1} + \beta_2 T^{-2}$$

Three-step approach

1. Determine the I(d) for every variable

   Dickey Fuller, Perron tests H_0: series is non-stationary
2. Estimate the cointegration relation by OLS regression
3. Test the residuals for stationarity

   $$y_{1t} = \beta_0 + \beta_1 y_{2t} + \varepsilon_t \Rightarrow \varepsilon_t = y_{1t} - \beta_0 - \beta_1 y_{2t}$$

   $$\hat{\varepsilon}_t = y_{1t} - \hat{\beta}_0 - \hat{\beta}_1 y_{2t}$$

   H_0: series are not cointegrated . ADF Test does not give correct critical values because of the OLS residuals. For this reason, we use MacKinnon Table to determine the critical values

9.5.Error Correction Model

Granger Representation Theorem

Determination of the dynamic relationship between cointegrated variables in terms of their stationary error terms.

For bivariate case: Two integrated I(1) variables y_{1t} and y_{2t} yielding one cointegrated combination $\varepsilon_t \sim I(0)$

$$\Delta y_{1t} = \lambda_1 \varepsilon_{t-1} + \sum_{i=1}^{p-1} (a_{11i} \Delta y_{1t-i} + a_{12i} \Delta y_{2t-i}) + \varepsilon_{1t}$$

$$\Delta y_{2t} = \lambda_2 \varepsilon_{t-1} + \sum_{i=1}^{p-1} (a_{21i} \Delta y_{1t-i} + a_{22i} \Delta y_{2t-i}) + \varepsilon_{1t}$$

We estimate parameters by OLS. Regression with only stationary variables on both sides.

Multivariate Cointegration Analysis - Johansen Test

VAR(1) having M I(1) variables can be expressed as:

$$Y_t = \mu + \Gamma Y_{t-1} + \varepsilon_t$$

Where, Y, μ and ε are (Mx1) vectors and Γ is an (MxM) matrix.
By subtracting the lagged vectors Y from both sides of the equation we receive the following relation:

$$Y_t - Y_{t-1} = \mu + \Gamma Y_{t-1} - Y_{t-1} + \varepsilon_t$$

$$\Delta \mathbf{Y_t} = \mu + (\mathbf{A_1} - \mathbf{I})\mathbf{Y_{t-1}} + \varepsilon_t$$

$$\Delta \mathbf{Y_t} = \mu + (\Gamma - \mathbf{I})\mathbf{Y_{t-1}} + \varepsilon_t$$

$\Delta \mathbf{Y_t}$ and are I(0) vectors. Thus, the term (Γ - I)Y_{t-1} must be also I(0). If the variables are not cointegrated, then the matrix Γ is a unit matrix I. If there exists **r** cointegrated relations (ε_t is a (rx1) vector), this term can be written as a I(0) variable:

$$(\Gamma - I)Y_{t-1} = \lambda \gamma' Y_{t-1} = \lambda \varepsilon_{t-1}$$

where γ' is the (rxM) matrix of the cointegration coefficients and λ is a (Mxr) matrix.Multiplying with the cointegration matrix the latter results in the (MxM) matrix (Γ - I). This term is I(0) and λ can be interpreted as the matrix of the M times r error correction coefficients:

$$\Delta Y_t = \mu + \lambda \varepsilon_{t-1} + e$$

This model is a generalization of the ECM in the previous section.If the initial model constitutes a VAR(p) model then the error correction representation contains additionally (p-1) difference terms.
Since the matrix (Γ - I) can be represented by the product of a (rxM) and a (Mxr) matrix, it has the rank r.
This means that the **number of cointegrated relations** is determined by the rank **(r)** of the matrix.
In the marginal case r = 0, i.e Γ = I, the model reduced to a VAR model in differences (M independent random walks). If r equals M we are concerned with M stationary level data, I(0).

Johansen Test

The approach of Johansen is based on the maximum likelihood estimation of the matrix (Γ - I) under the assumption of normal distributed error variables. Following the estimation the hypotheses

$$H_0: r = 0, \quad H_0: r = 1, \ldots, H_0: r = M-1$$

are tested using likelihood ratio (LR) tests.

Example. Variables are: Exchange rate, interest rates, S&P 500(GLOBAL) index, ISE index from 01.01.2000 to 31.12.2007

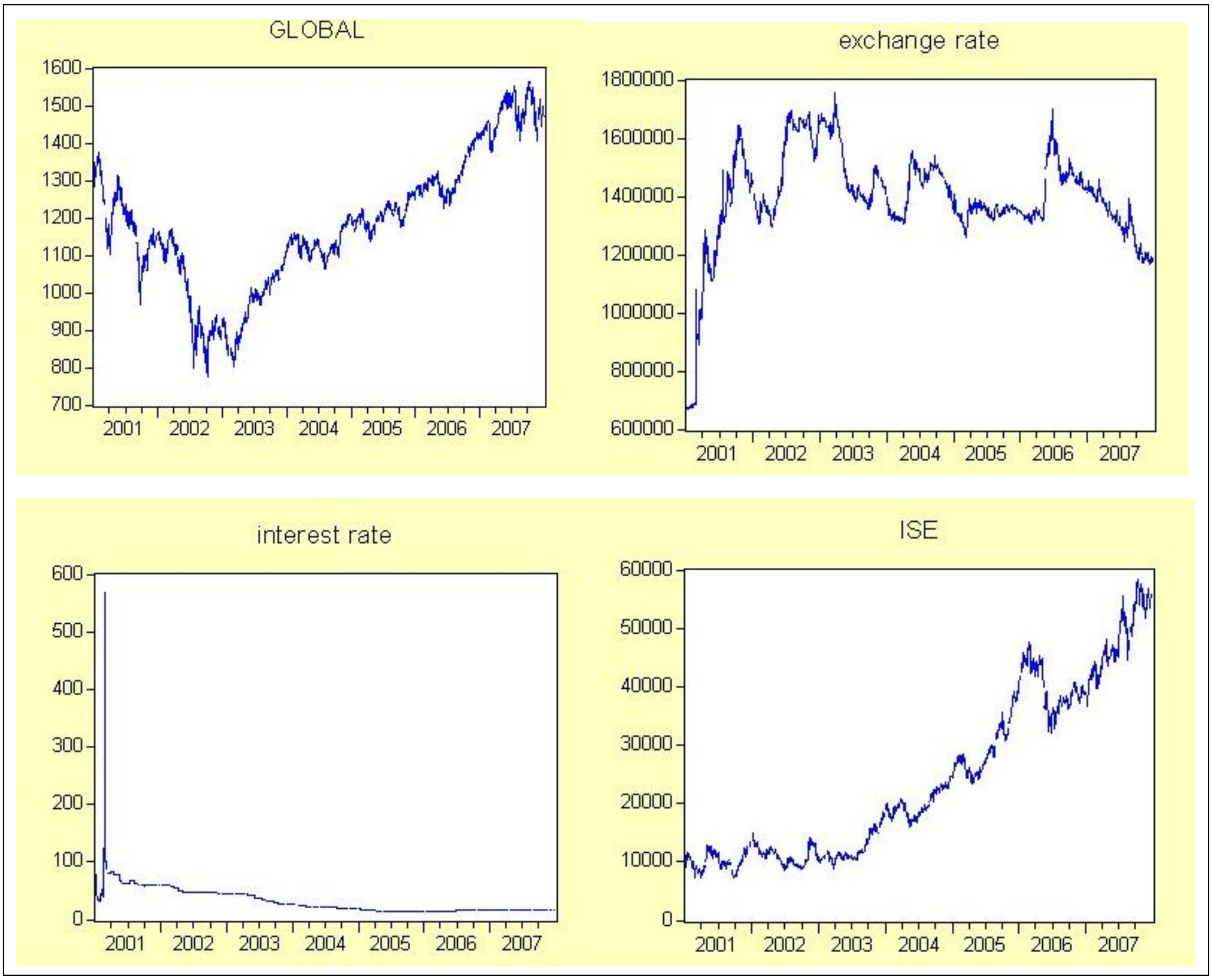


Figure 9.5 Plots of the series

Included observations: 1745 after adjustments
Trend assumption: Linear deterministic trend
Series: GLOBAL EXCHANGE_RATE INTEREST_RATE ISE
Lags interval (in first differences): 1 to 4
Unrestricted Cointegration Rank Test (Trace)

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	0.05 Critical Value	Prob.**
None *	0.065285	156.7717	47.85613	0.0000
At most 1 *	0.017048	38.96042	29.79707	0.0034
At most 2	0.005118	8.955273	15.49471	0.3695
At most 3	1.20E-06	0.002096	3.841466	0.9599

Trace test indicates 2 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values
Unrestricted Cointegration Rank Test (Maximum Eigenvalue)

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	0.05 Critical Value	Prob.**
None *	0.065285	117.8113	27.58434	0.0000
At most 1 *	0.017048	30.00514	21.13162	0.0022
At most 2	0.005118	8.953177	14.26460	0.2901
At most 3	1.20E-06	0.002096	3.841466	0.9599

Max-eigenvalue test indicates 2 cointegrating eqn(s) at the 0.05 level
* denotes rejection of the hypothesis at the 0.05 level
**MacKinnon-Haug-Michelis (1999) p-values

1 Cointegrating Equation(s): Log likelihood -46009.85

Normalized cointegrating coefficients (standard error in parentheses)

GLOBAL	EXCHANGE_RATE	INTEREST_RATE	ISE
1.000000	-0.000120	-16.55210	-0.026394
	(0.00014)	(1.44457)	(0.00218)

2 Cointegrating Equation(s): Log likelihood -45994.85

Normalized cointegrating coefficients (standard error in parentheses)

GLOBAL	EXCHANGE_RATE	INTEREST_RATE	ISE
1.000000	0.000000	-15.66567	-0.024910
		(1.30729)	(0.00194)
0.000000	1.000000	7382.376	12.36210
		(1649.84)	(2.44689)

Therefore, we can conclude that in the long term these three variables are cointegrated and there are 2 cointegration equations.

Pairwise Granger Causality Tests
Date: 07/20/08 Time: 10:40
Sample: 1/02/2001 12/31/2007
Lags: 5

Null Hypothesis:	Obs	F-Statistic	Probability
INTEREST_RATE does not Granger Cause EXCHANGE_RATE	1745	28.3482	1.1E-27
EXCHANGE_RATE does not Granger Cause INTEREST_RATE		32.1459	2.0E-31
ISE does not Granger Cause EXCHANGE_RATE	1745	58.2545	3.6E-56
EXCHANGE_RATE does not Granger Cause ISE		2.31559	0.04151
GLOBAL does not Granger Cause EXCHANGE_RATE	1745	21.4690	6.7E-21
EXCHANGE_RATE does not Granger Cause GLOBAL		1.16105	0.32611
ISE does not Granger Cause INTEREST_RATE	1745	12.2991	9.7E-12
INTEREST_RATE does not Granger Cause ISE		0.10286	0.99158
GLOBAL does not Granger Cause INTEREST_RATE	1745	2.98831	0.01084
INTEREST_RATE does not Granger Cause GLOBAL		1.48645	0.19105
GLOBAL does not Granger Cause ISE	1745	20.7727	3.3E-20
ISE does not Granger Cause GLOBAL		1.91422	0.08894

References:

1. Utts, J., Heckard, R. (2007) Mind on Statistics, 3rd Ed., Thomas.
2. Selvanathan, A *et al*. (2004): *Australian Business Statistics*, 3rd edition (Abridged), Nelson
3. Enders, W., Applied Econometric Time Series, Second Edition, Wiley.
4. Shumway, R.H., Stoffer, D.S., Time Series Analysis and its Applications, Springer, 2000
5. Hamilton, J.D., Time Series Analysis, Princeton University Press, 1994
6. Brockwell P.J. and Davis R.A, Introduction to Time Series and Forecasting, Springer, 1996.
7. Kirchgaessner, G., Wolters, J., Introduction to Modern Time Series Analysis, Springer, 2008.