Developing Intelligent Systems for Performing Business Analytics

Time Series Analysis & Forecasting

Dr. Rita Chakravarti
Institute of Systems Science
National University of Singapore
Email: rita@nus.edu.sg

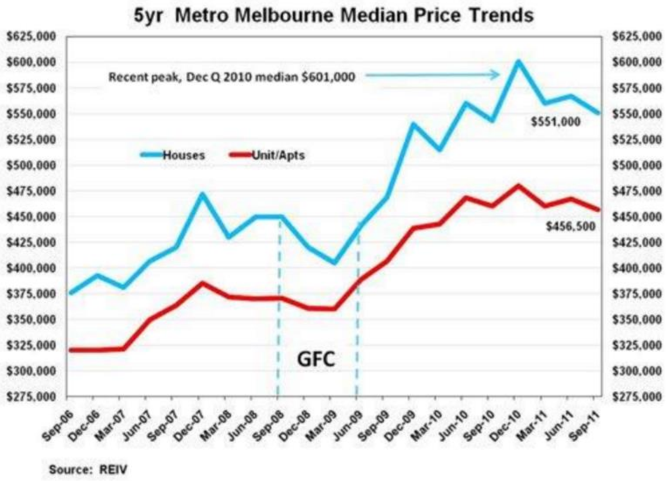
Objective

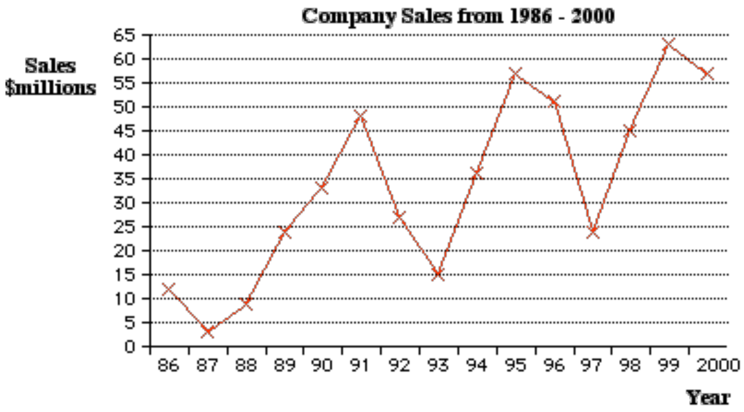
- To introduce the concepts of time series (TS) forecasting
- To observe different types of time series components
- To discuss different methodologies of time series (TS) forecasting
- To discuss the practical concerns in modeling time series

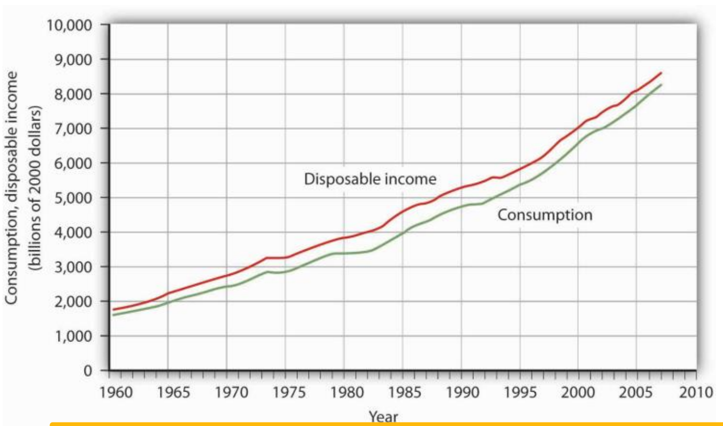
Time Series Forecasting: *where?*

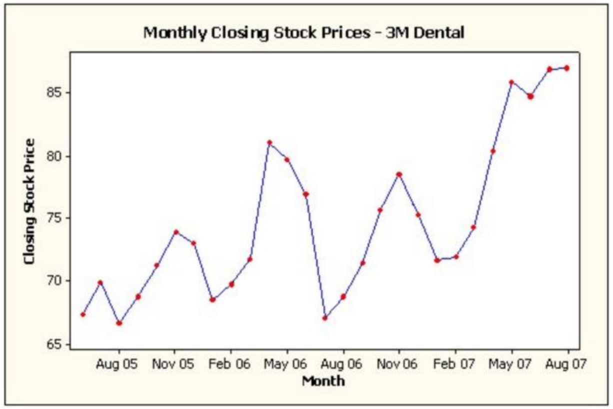
- Time series forecasting is performed in nearly every organization that works with quantifiable data. Examples of use of forecasting for decision support
  - Retail stores: sales
  - Energy companies: reserve, production, demand, prices
  - Education institutes: enrollment
  - Governments: tax receipts and spending
  - Inter. financial organization: inflation and economic activity
  - Transportation companies: future travel
  - Banks: new home purchases
  - Venture capital firms: market potential to evaluate business plan

Examples of Time Series









Important features of time series : Usually the observations are taken at regular intervals (e.g.: hours, days, weeks, months, years) - Time Series Data

The influence among **neighboring observations** is informative for prediction in Time Series

Time Series Forecast: *as predictive model*

- The goal of building a time series model is the same as the goal for other types of predictive models, e.g.: linear regression
  - To create a model such that the error between the predicted value of the target variable and the actual value is as small as possible
    - "fitting" the time series data (training data)

Values,
TS patterns

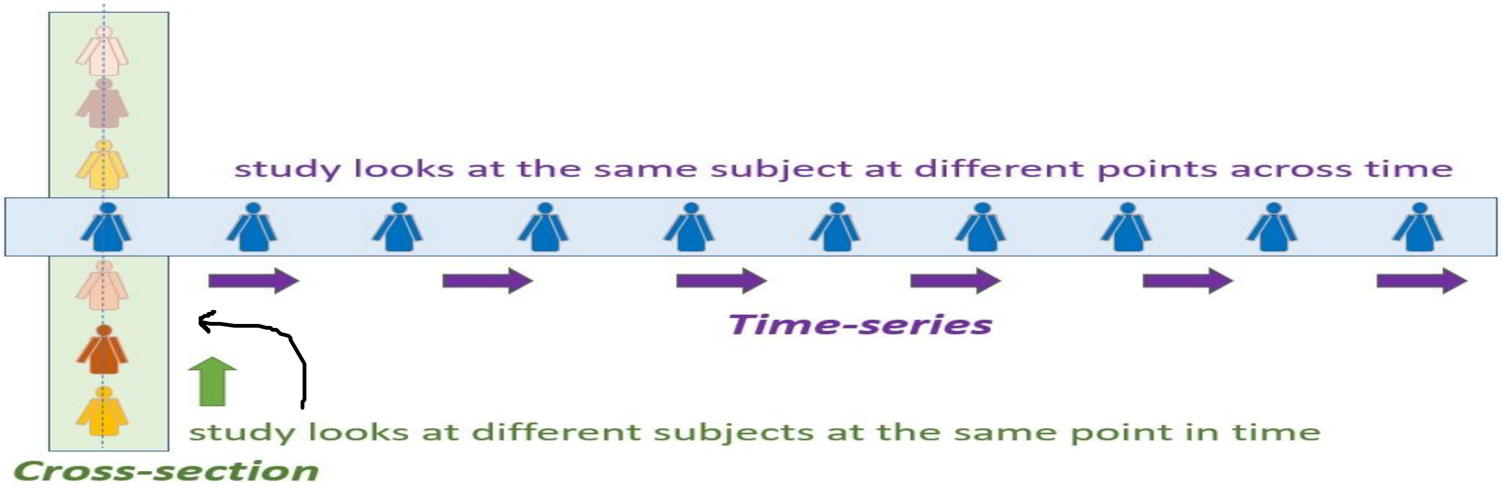
- The primary difference
  - In time series models, early observations of the "target" variable (Y) are used to predict future values
  - Whereas other types of models use other (X-s) variables as predictors

What is Time Series Data

- In our previous discussion of prediction (Regression Theory)
  - The order of measurements in time does not matter

    These are called *cross-sectional data*

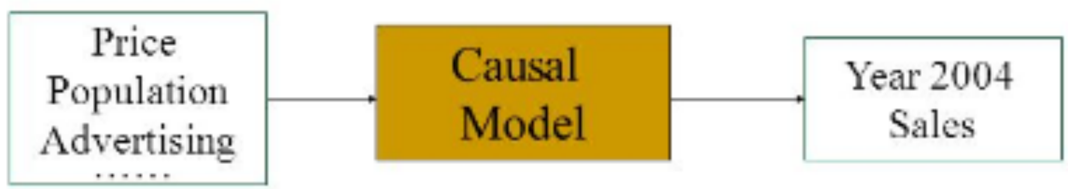
- ***Time series*** is a chronological sequence of observations on a particular variable.



Time Series vs Causal Models

Forecasting technique is Time Series Technique

Forecasting technique is Regression



lot of parameters req.- gathering data tougher.
takes other external factors into consideration while predicting.

TS Forecasting Methods

- Forecasting methods that are popular in business applications can be roughly divided into two kinds
  - Model-based methods (parametric)
    - The training data are used to estimate the parameters of the model, and then the model is used to generate forecasts.
      - Predictors / regressors are to be explicitly indicated
      - E.g.: Multiple linear regression, (Neural networks, Decision Trees)still non-para as u cant get equat
  - Algorithm-based methods (non-parametric)
    - The algorithms "capture" patterns from the data and construct the functions of integration

focus of our discussion

The result of learning is sometimes called "Time series ***model***"

- E.g.: Smoothing methods

TS Forecasting: *key of success*

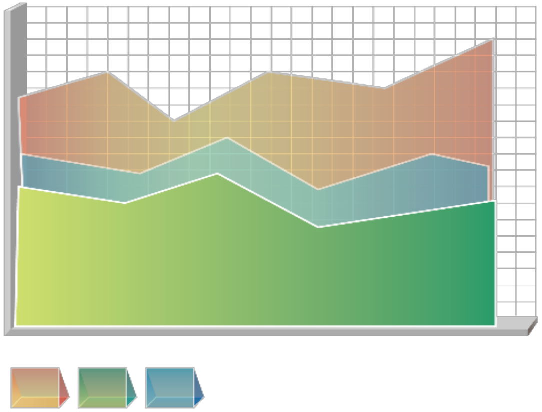
- The key idea of data-driven TS forecasting methods is to uncover the ***patterns*** from the data
  - E.g.: trend, seasonality, …

and

find an appropriate ***integration*** of the components

  - E.g.: additive, multiplicative, …

to "fit" the actual series



Time Series Analysis

Time Series Patterns & Components

Time series data is assumed to consist of systematic pattern/components and random noise (error). Hence de-noise/noise filtering is required in most time series analysis technique in order to observe the pattern.

The Components of a Time Series are:

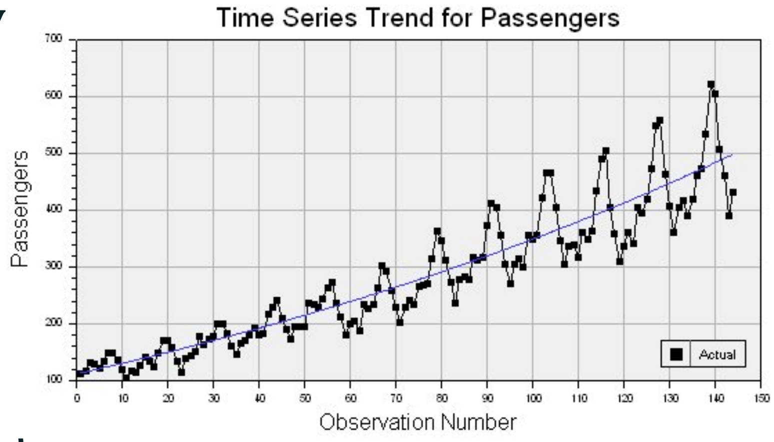
1) Trends
2) Cycle
3) Seasonal variation and
4) Irregular fluctuations

Trend

Trend refers to upward or downward movement that characterizes a time series over a period of time. Thus trend reflects the long-run growth or decline in the time series

Trend movements can represent a variety of factors. For example, long run movements in the sales of a particular industry might be determined by one, some or all of the following factors:

- Technological change in the industry
- Changes in consumer tastes
- Increases in per capita income
- Increases in total population
- Market growth
- Inflation or deflation (price changes)



Cycle

Cycle refers to up and down movements around trend levels. These fluctuations can have a duration of anywhere from two to ten years or even longer measured from peak to peak or trough to trough

One of the common cyclical fluctuations found in time series data is the "business cycle" The business cycle is represented by fluctuations in the time series caused by recurrent periods of prosperity alternating with recession

(increase or decrease in economic activity)

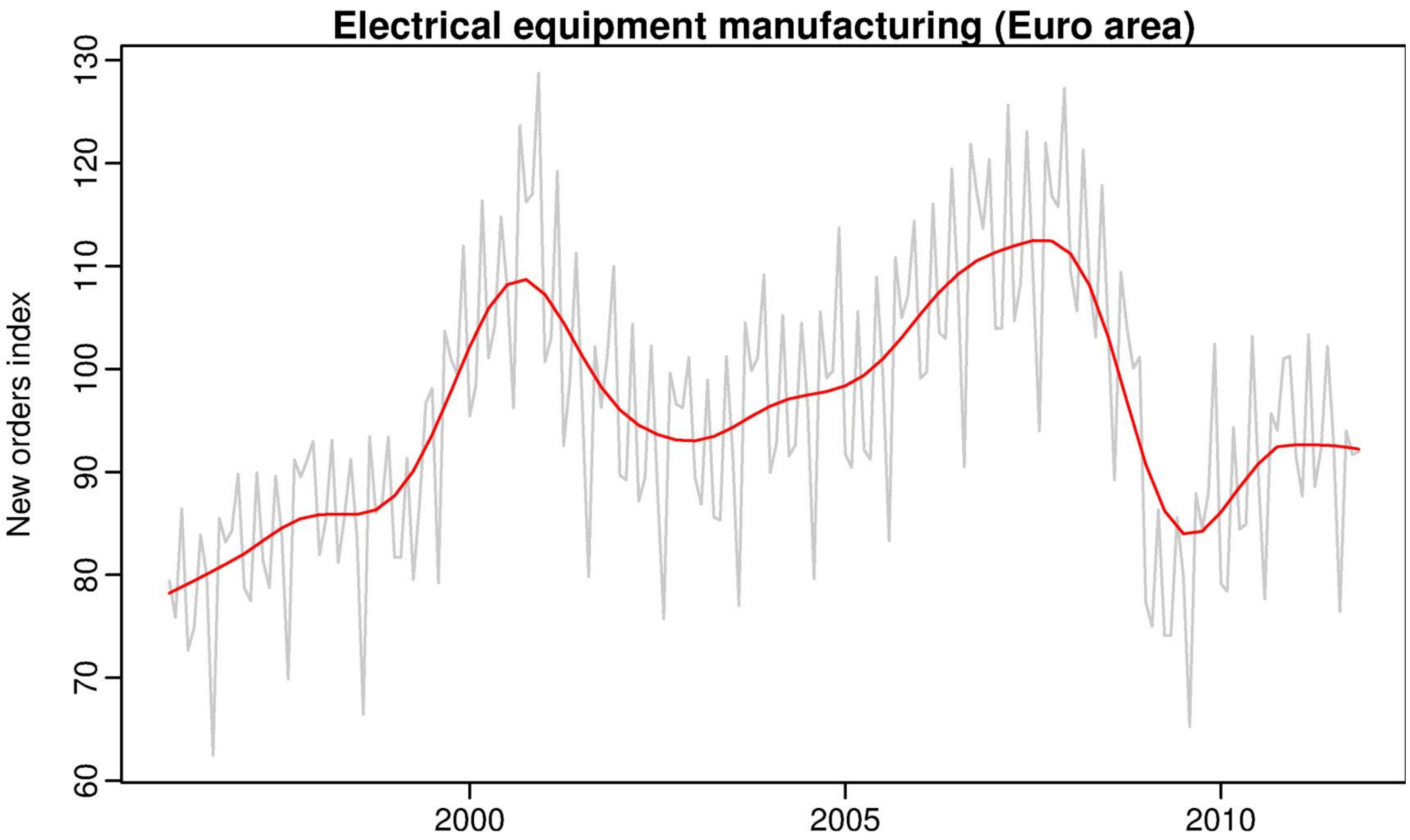
However it is not necessary that it always happens due to changes in economic factors e.g.

- Cyclical fluctuations in agricultural yields due to weather cycles
- Cyclical fluctuations in demand of a type of clothing due to fashion change

Because there is no single explanation for cyclical fluctuations the vary greatly in both length and magnitude

duration small- Seasonal
long duration - cycles

Cycle


Electrical equipment manufacturing (Euro area)
New orders index
130
120
110
100
90
80
70
60
2000
2005
2010

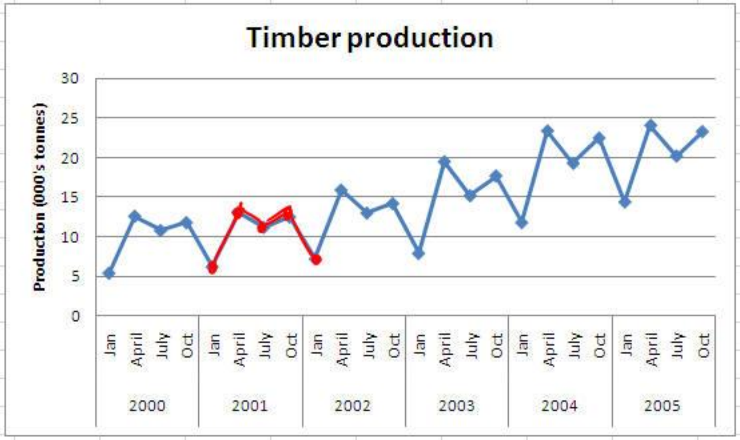
Seasonal Variation

Seasonal variations are periodic patterns in a time series that complete themselves within a calendar year and are then repeated on a yearly basis

Seasonal variations are typically caused by factors such as weather or customs e.g.

- Average monthly temperature
- Monthly sales volume in a departmental stores

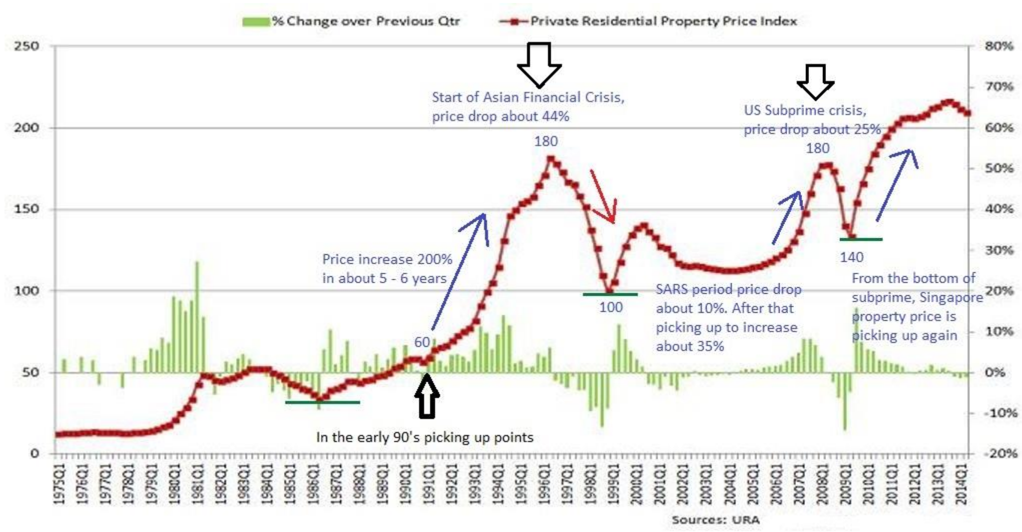
Ordinarily series of monthly or quarterly data are used to examine seasonal variations. Clearly a single yearly observation would not reveal variations that occur during the year



Irregular Fluctuation

Irregular fluctuations are erratic movements in a time series that follow no recognizable regular pattern. Such movements represent what is “left over” in a time series after trend, cycle and seasonal variations have been accounted for.

Many irregular fluctuations in a time series are caused by ‘unusual’ events that cannot be forecasted – earthquakes, accidents, hurricanes, wars etc.



FORECASTING

Forecasting Methods

Two types of forecasting methods for TS data : **Qualitative** & **Quantitative**

Qualitative : This type uses the opinion of experts to predict future event subjectively. Such methods are often used when the historical data is either not available or scarce

Example :

1. New product launch where no historical data is available. In such a case for forecasting the company has to rely on expert opinion
2. Another case where one needs to predict if and when new technologies will be discovered and adopted

The commonly used techniques are :

- Subjective curve fitting
- **Delphi method**
- Time independent technological comparisons

They are frequently called judgmental forecasting methods

Qualitative Technique: Delphi Methods

- The Delphi method is based on the assumption that group judgments are more valid than individual judgments.
- Predictions and reasoning from experts are collected and summarized by a moderator.
- Experts are now told the predictions of other experts and also provided feedback by the moderator
- Experts revise as they see fit
- After at least two rounds of feedback the forecasts are combined.
- The Delphi method was developed at the beginning of the Cold War to forecast the impact of technology on warfare – **relevant for policy making etc.**

Advantages of Qualitative Forecasts

- Takes advantage of the users' domain knowledge
- Can be used when we have no history, e.g., when a new product with no past analogue is introduced
- Can be used when available data is informal or qualitative and quantitative data is missing or sparse

Limitations of Qualitative Forecasts

- Impractical or too expensive when thousands of forecasts are required on a regular basis
  - Airlines passenger forecasts
  - Inventory planning
  - Economic Planning

- Expert judgement used for forecasting is subjective
  - It is lost when the expert leaves

- Expert judgement is often biased due to
  - Recency : recent occurrences are given greater weight
  - Risk aversion; human forecasters consistently under-forecast cancellation rates to avoid overbooking;
  - Hopes and fears of the expert
  - Confirmation bias

Approaching Time Series Analysis

- There are many, many different quantitative time series techniques.

- It is usually impossible to know which technique will be best for a particular data set.

- It is customary to try out several different techniques and select the one that seems to work best.

- To be an effective time series modeler, you need to keep several time series techniques in your “tool box.”

Limitations of Forecasting

- Many managers have an unrealistic attitude towards forecasting – some of are pessimistic about what can be accomplished while others are unduly optimistic
- The future is neither completely knowable nor totally obscure. High quality forecasts should be seen as having a very direct impact on "bottom line". Some uncertainty (which cannot be dealt with by forecasting methods) will inevitably remain
- Some quantities are notoriously more difficult to forecast that others
- Forecasting techniques are not capable of eliminating uncertainty about the future, but usually they do a good job in predicting potential "trends" and "characterizing" residual uncertainties

Principles of Forecasting

- Forecasts at a higher level of granularity are (almost always) more accurate than forecasts at a lower level of granularity
  - Sales forecasts for SKUs are likely to be worse than sales forecasts for brands; sales forecasts for brands are likely to be worse than sales forecasts for products
  - Sales forecasts for stores are likely to be worse than sales forecasts for cities; sales forecasts for cities are likely to be worse than sales forecasts for sales territories
  - It is important to determine the level(s) at which data should be aggregated so that the forecasts are good & meaningful

Principle of forecasting contd..

- Near term forecasts are (almost always) more accurate than distant term forecasts

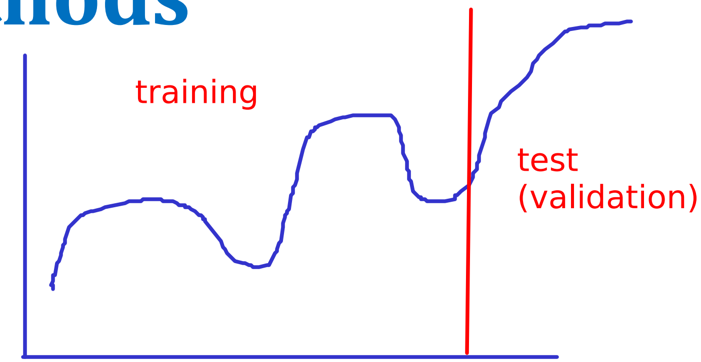
  - 1 week ahead forecasts will usually be better than 1 month ahead forecasts;

  - 1 month ahead forecasts will usually be better than 2 month ahead forecasts;

  - 2 month ahead forecasts will usually be better than 1 year ahead forecasts

  - It is important to determine the forecast horizon(s) required for operational, tactical and strategic decisions

Evaluating Methods



- Forecasting method is selected - many times by intuition, previous experience, or computer resource availability

1. Divide the data into two parts - an initialization part (training) and a test (validation) part
2. Use the forecast technique to determine the fitted values for the initialization data set
3. Use the forecast technique to forecast the test data set and determine the forecast errors
4. Evaluate errors (MAD, MPE, MSE, MAPE)
5. Use the technique, modify, or develop new model

Forecasting Accuracy

- We need a way to compare different time series techniques for a given data set.
- Four common techniques are the:
  - mean absolute deviation, $$\text{MAD} = \sum_{i=1}^{n} \frac{\left| Y_i - \hat{Y}_i \right|}{n}$$
  - mean absolute percentage error, $$\text{MAPE} = \frac{100}{n} \sum_{i=1}^{n} \frac{\left| Y_i - \hat{Y}_i \right|}{Y_i}$$
  - the mean squared error, $$\text{MSE} = \sum_{i=1}^{n} \frac{\left(Y_i - \hat{Y}_i \right)^2}{n}$$
  - root mean squared error. $$\text{RMSE} = \sqrt{\text{MSE}}$$

Two types of forecasts are considered : a) Point Forecast b) Prediction Interval Forecast

What do these Measure?

- Bias – Does the forecast tend to underestimate or overestimate the actual
- MAD – Measures the error, i.e., the gap between forecasts and actuals
- MAPE – Measures the error as a percentage
- MSE – Also measures the error
  - Strongly penalizes large errors

indirectly variance

- RMSE – Scales down the MSE

indirectly standard deviation

Forecasting Philosophies

- Forecasting "Philosophies":
  - Some forecasting systems use many different models to obtain each forecast and then determines the best according to some statistical model
  - Others use a single model
  - Some combine forecasts from different models

- Classes of Models
  - Regression
  - Time series

Quantitative Forecasting Techniques

Causal

- Regression Analysis

Univariate

- Time Series Regression (t is the independent variable)
- Dummy Variable Regression (to predict seasonality)
- Classical Decomposition Methods
- Smoothing Techniques
- Box-Jenkins Methods (ARIMA)

TIME SERIES REGRESSION

Trend Models (Time Series Regression)

$$y_t = TR_t + \varepsilon_t$$

Where y_t = value of the time series in period t

TR_t = the trend in time period t

ε_t = the error term in time period t

No Trend : $TR_t = \beta_0$

Linear Trend : $TR_t = \beta_0 + \beta_1 t$

Quadratic Trend : $TR_t = \beta_0 + \beta_1 t + \beta_2 t^2$

The pth – order polynomial trend model is :

$$TR_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \dots + \beta_p t^p$$

Application

The State University Credit Union, a savings institution open to faculty and staff of the university handles savings accounts and makes loans to members.

In order to plan its investment strategies, the credit union requires both point predictions and prediction intervals of monthly loan requests (in ‘000 dollars) to be made by the faculty and staff in future months

Loan Requests (‘000s)

Month	Year 1	Year2
Jan	297	808
Feb	249	809
Mar	340	867
Apr	406	855
May	464	965
Jun	481	921
Jul	549	956
Aug	553	990
Sep	556	1019
Oct	642	1021
Nov	670	1033
Dec	712	1127

Regression Statistics	
Multiple R	0.989028
R Square	0.978177
Adjusted R Square	0.977185
Standard Error	39.67949
Observations	24

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	261.1232	16.71897	15.61838	2.17E-13	226.4502	295.7962	226.4502	295.7962
X Variable 1	36.74348	1.170084	31.40242	9.08E-20	34.31687	39.17008	34.31687	39.17008

Use the data and do the necessary analysis to help the credit union in their decision making. Can you improve upon the above result?

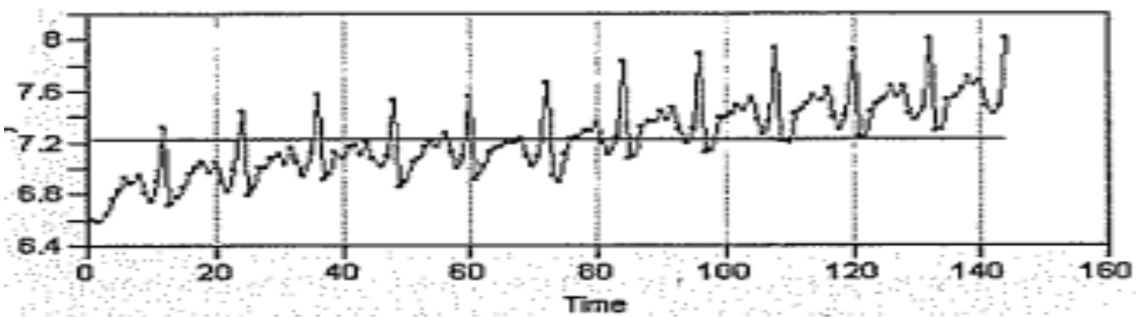
DUMMY VARIABLE REGRESSION

To Predict Seasonality Pattern

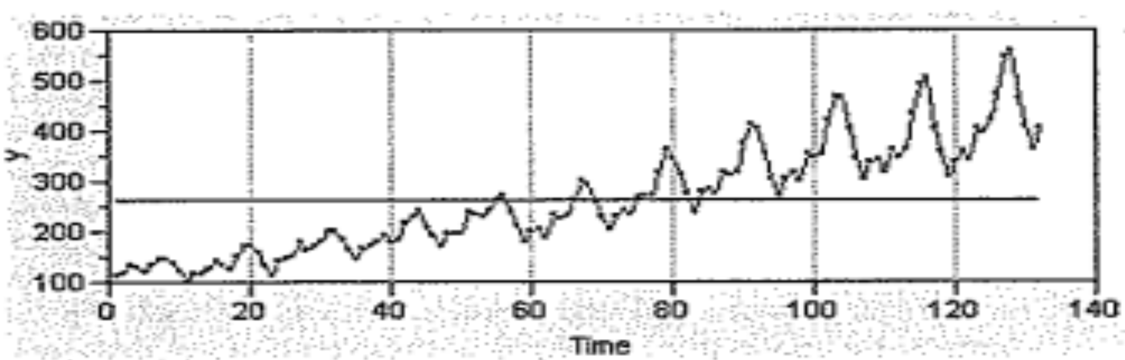
Time Series with Seasonal Variation

Two Types of seasonal variation for Time Series :

1) Constant Seasonal Variation : If the magnitude of seasonal swing does not depend on the level of the time series



2) Increasing Seasonal Variation : If the magnitude of seasonal swing depends on the level of the time series



In such case it is a common practice to convert **increasing** variation to **constant** variation using the Transformation

$y_t^* = y_t^\lambda$

where $0 < \lambda < 1$

As λ approaches 0 one uses $y_t^* = \ln y_t$

Modeling Seasonal Variation By Using Dummy Variables

While analyzing a time series that exhibits constant seasonal variation, a model of the following form is often used:

$$y_t = TR_t + SN_t + \varepsilon_t$$

Where

y_t = the observed value of time series in time period t
TR_t = the trend in time period t
SN_t = the seasonal factor in time period t
ε_t = the error term(irregular factor) in time period t

One way to model seasonal patterns is to employ **dummy variables.** Assuming there are L seasons (months, quarters etc.) per year the seasonal factor SN_t is expressed as:

$$SN_t = \beta_{s1} x_{s1,t} + \beta_{s2} x_{s2,t} + \ldots\ldots\ldots\ldots\ldots\ldots\ldots + \beta_{s(L-1)} x_{s(L-1),t}$$

Where $x_{s1,t}$, $x_{s2,t}$, ……… $x_{s(L-1),t}$ are **dummy variables** that are defined as follows:

$x_{s1,t}$ = 1 if time period t is season 1
0 otherwise

$x_{s2,t}$ = 1 if time period t is season 2
0 otherwise

.

$x_{s(L-1),t}$ = 1 if time period t is season L-1
0 otherwise

Modeling Seasonal Variation By Using Dummy Variables contd…

When L = 12, we have monthly data. If the trend is linear the model becomes :

$$y_t = TR_t + SN_t + \varepsilon_t$$

$$= (\beta_0 + \beta_1 t) + (\beta_2 M_1 + \beta_3 M_2 \ldots\ldots\ldots\ldots\ldots\ldots + \beta_{12} M_{11}) + \varepsilon_t$$

Application

Traveler's Rest Inc operates 4 hotels in a Midwestern city. The analysts in the operations division of the organization were asked to develop a model that could be used to obtain short-term forecasts (up to one year) of the number of occupied rooms in the hotels. These forecasts were needed by various personnel to assist in decision making with regard to hiring additional help during the summer months, ordering materials that have long delivery lead times, budgeting of local advertising expenditure and so on.

Monthly average of hotel room occupancy y_t for 14 years have been provided. Explore the pattern of y_t and y_t^{λ} for various values of λ. Fit a seasonal model using dummy variables.

CLASSICAL DECOMPOSITION METHOD

less than 12 months data cant use dummy or ARIMA, therefore use this......

Classical Decomposition of Time Series

- **Trend** – does not necessarily imply a monotonically increasing or decreasing series but simply a lack of constant mean, though in practice, a linear or quadratic function is often to predict the trend;
- **Cycle** – refers to patterns or waves in the data that are repeated after approximately equal intervals with approximately equal intensity. For example, some economists believe that “business cycles” repeat themselves every 4 or 5 years;
- **Seasonal** – refers to a cycle of one year duration;
- **Random (irregular)** – refers to the (unpredictable) variation not covered by the above

Decomposition Methods

Decomposition Methods are used to forecast time series that exhibit trend and seasonal effects

These models have no theoretical basis – they are strictly an intuitive approach

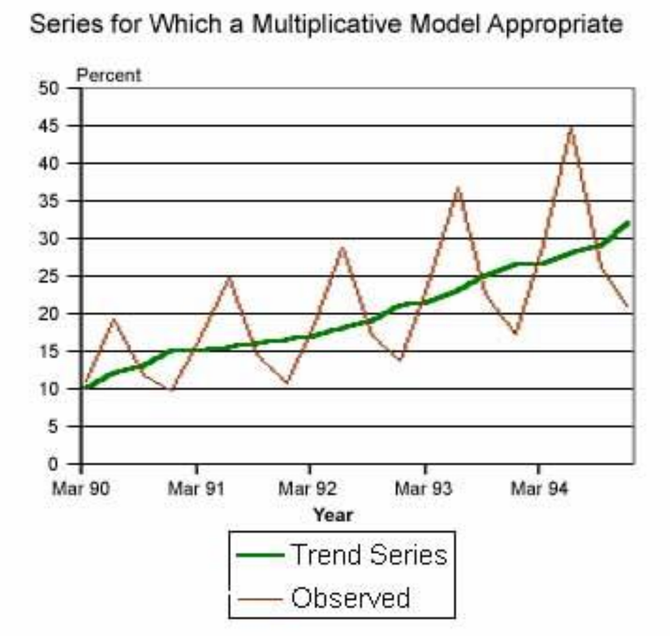
They are useful when the parameters describing a time series is not changing over time

The basic idea behind these models is to decompose the time series into several factors : 1) Trend 2) Seasonal 3) Cyclical and 4) Irregular (error)

Two Types of Decomposition Methods

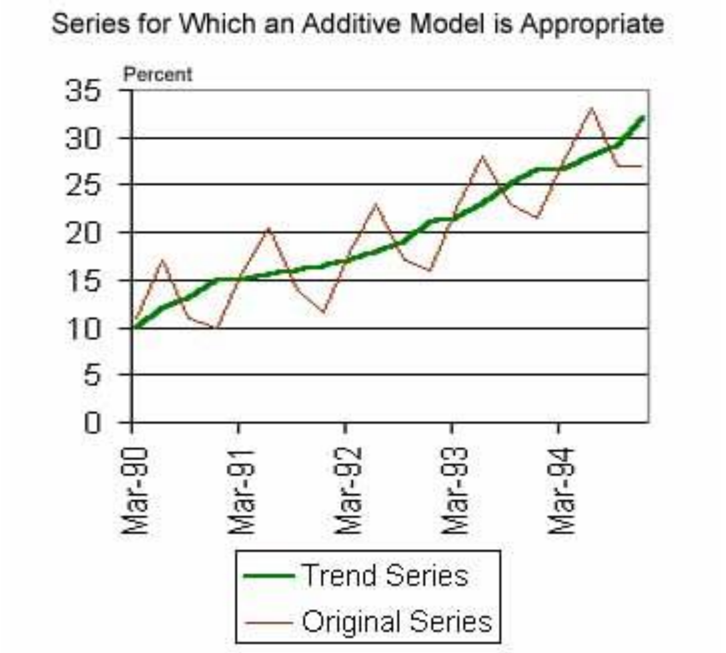
Multiplicative Models

$$Y_t = TR_t \times SN_t \times CL_t \times IR_t$$



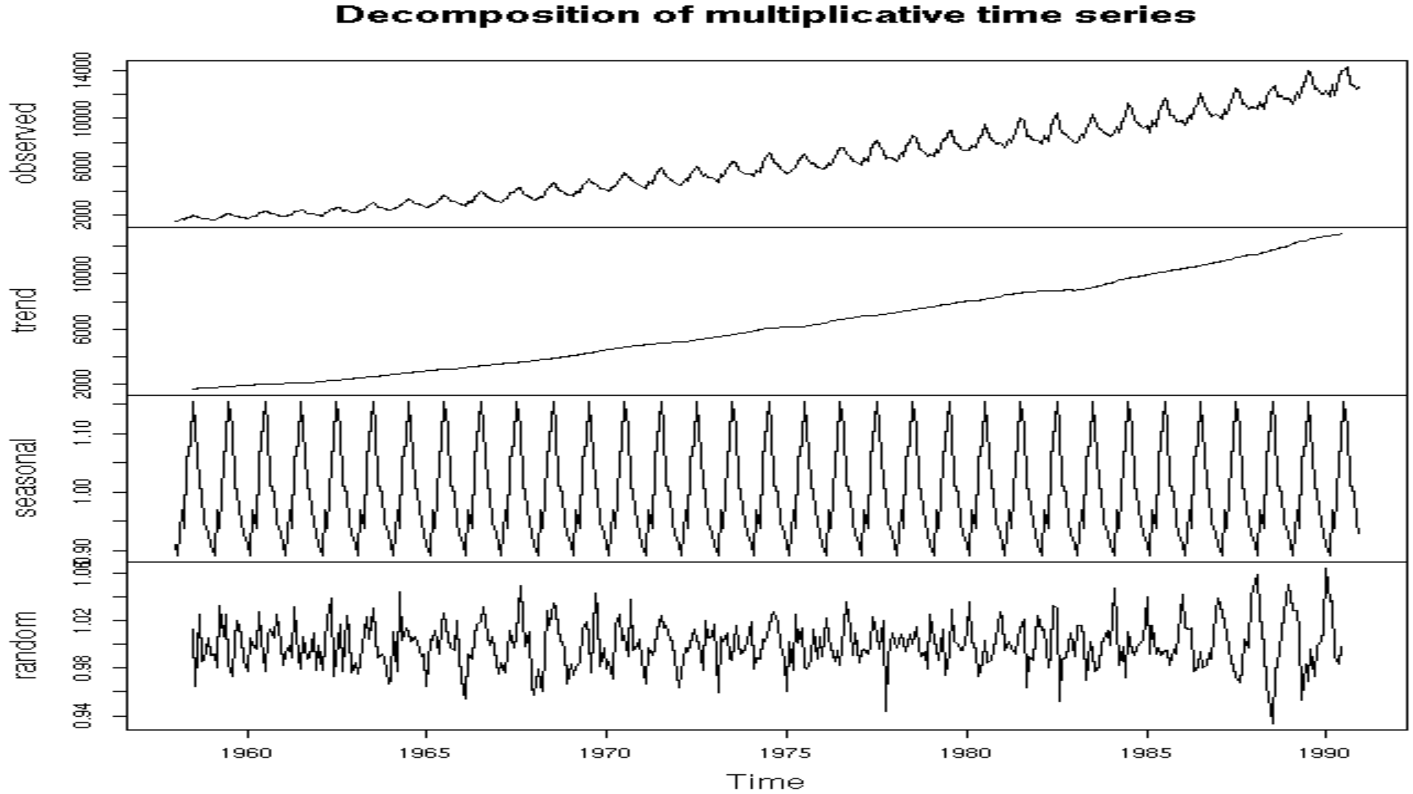
Additive Models

$$Y_t = TR_t + SN_t + CL_t + IR_t$$



Multiplicative Decomposition

Consider a Time Series that exhibits increasing or decreasing Seasonal Variation



Multiplicative Decomposition Method

Multiplicative Models

$$y_t = TR_t \times SN_t \times CL_t \times IR_t$$

Where.

y_t = the observed value of the time series at time t

TR_t = the trend component (or factor) at time t

SN_t = the seasonal component (or factor) at time t

CL_t = the cyclical component (or factor) at time t

IR_t = the irregular component (or factor) at time t

Multiplicative Decomposition Method

This decomposition method employs multiplicative seasonal factor

i.e. seasonal factor is multiplied by trend (instead of added to the trend like dummy variable regression)

Example Sales (in case the model is multiplicative) :

Let the trend equation be : $TR_t = 500 + 50t$, $t = 0$, Q_4 2002

So in 2003 the forecasts using trend equation only is :

$TR_1 = 550$ Q_1 2003 $TR_2 = 600$ Q_2 2003

$TR_3 = 650$ Q_3 2003 $TR_4 = 700$ Q_4 2003

Suppose it is known that sales are lowest in the Q_1 highest in the Q_2, moderately high in the Q_3 and moderately low in Q_4

As we know the sales are seasonal and past record shows that

$SN_{Q1} = 0.4$ $SN_{Q2} = 1.6$ $SN_{Q3} = 1.2$ $SN_{Q4} = 0.8$

Multiplicative Decomposition Method

So assuming multiplicative model, if we consider both trend and seasonal effects (trend * seasonal)

So in 2003 the forecasts using trend and seasonality would be :

$TR_1 X\, SN_{Q1} = 220$ $\quad Q_1\ 2003$ $\quad TR_2 X\, SN_{Q2} = 960$ $\quad Q_2\ 2003$

$TR_3 X\, SN_{Q3} = 780$ $\quad Q_3\ 2003$ $\quad TR_4 X\, SN_{Q4} = 560$ $\quad Q_4\ 2003$

If the multiplicative seasonal factors remain constant over time what would be the forecasts in 4 quarters in 2004?

Multiplication of the trend by seasonal factors implies that the size of the seasonal swing is proportional to the trend. So if the trend increases so does the swing.

However the assumption is that the seasonal factors are constant over time. If not other techniques need to be used.

Multiplicative Decomposition Method

The seasonal factor SN_t models cyclical pattern in a time series that are completed within a calendar year. If a time series displays a cycle that has a longer duration, a cyclical factor can be defined. Suppose in our sales data all 4 quarters of 2003 are included in the boom period of a business cycle. Also suppose CL_1 = 1.08, CL_2 = 1.09, CL_3 = 1.09, CL_4 = 1.10. If we consider trend, seasonal & cyclical factors, sales in the four quarters of 2003 would be :

$TR_1 X SN_{Q1} X CL_1 = 238$ Q_1 2003 $TR_2 X SN_{Q2} X CL_2 = 1046$ Q_2 2003

$TR_3 X SN_{Q3} X CL_3 = 850$ Q_3 2003 $TR_4 X SN_{Q4} X CL_4 = 616$ Q_4 2003

If the multiplicative cyclical factors in 2004 are 1.09, 1.05, 1.01, 0.99 in four quarters what would be the forecasts in 4 quarters in 2004?

Multiplicative Decomposition Method

Work shop

- A cola company owns and operates ten drive in soft drink stores. They were selling "diet cola" that was introduced in the market just three years ago and has been gaining popularity. They order the supply of the "diet cola" from a regional distributor. The company has an inventory policy that attempts to meet practically all of the demand for "diet cola", while at the same time ensuring that the company does not tie up its money needlessly by ordering much more than the actual demand. In order to implement its inventory policy the company needs to forecast monthly "diet cola" sales(in hundreds of cases).
- At the end of each month the cola company desires point forecasts and prediction interval forecasts of "diet cola" in future months
- The company has recorded monthly "diet cola" sales for previous 3 years in the data set "Cola_data.xls"

Multiplicative Decomposition Method

Step by step approach

1) Calculate **12** month moving average (MA_t) - to eliminate **seasonal** & irregular fluctuation --- Question : If it was quarterly data what would you do???

2) Next calculate Centered Moving Average CMA_t :Why??

Clue : we are calculating moving average of even order. If it was of odd order, this step was not necessary

Since the model is $y_t = TR_t \times SN_t \times CL_t \times IR_t$, this implies that

$$SN_t \times IR_t = y_t / TR_t \times CL_t = y_t / CMA_t$$

Because CMA_t is considered to be an estimate of $TR_t \times CL_t$, as averaging process is assumed to have removed seasonal and short term irregular fluctuations

3) Calculate $SN_t \times IR_t$

Multiplicative Decomposition Method

Step by step approach

4) Next step calculate average SN_t for each season (Jan, Feb, ...)
5) Then find normalization factor for SN_t . As L=12 (number of seasons) the normalization factor is $12/\sum_1^L SN_t$
6) By multiplying average SN_t by normalizing factor one would get estimate of SN_t.
7) Next step is to calculate the de-seasonalized observation d_t at time t:

$$d_t = y_t / \text{ estimated } SN_t$$

Deseasonalized observations are computed in order to better estimate the trend components. By dividing y_t by seasonality we can have better understanding of trend

8) Next step is to estimate TR_t . Look at plot of d_t , if it is linear please fit

$$TR_t = \beta_0 + \beta_1 t$$

This completes the estimation of SN_t (step 6) and TR_t

Multiplicative Decomposition Method

Step by step approach

9) So from the equation $y_t = TR_t \times SN_t \times CL_t \times IR_t$

it implies that $CL_t \times IR_t = y_t / (TR_t \times SN_t)$ where one can use the estimate of TR_t and SN_t

10) It is observed empirically that when considering monthly or quarterly data we can average out ir_t by taking a three period moving average

11) Finally we calculate the estimate of IR_t by using the formula,

$$cl_t \times ir_t \ / \ cl_t$$

Note : Traditionally the estimates tr_t, sn_t, cl_t and ir_t are obtained by using multiplicative decomposition method and used to describe the time series.

However we can also use these estimates to forecast the future values of time series. If there is no pattern in irregular component, we predict IR_t to be 1.

Multiplicative Decomposition Method

Forecasting using the method: In case $IR_t = 1$,

A) The point forecast of $y_t = tr_t \times sn_t \times cl_t$, if a well defined cycle exists and can be predicted

B) The point forecast of $y_t = tr_t \times sn_t$, if a well defined cycle does not exist or if CL_t can not be predicted

For our Cola example where

$$tr_t = b_0 + b_1 t$$

We can find the forecast value of TR_t by taking $t = 37, 38, \ldots\ldots$ Multiply each TR_t by corresponding monthly SN_t to obtain the point estimate

Although there is no theoretically correct prediction interval for y_t Bowerman , O'Connell & Koehler have found that a fairly accurate (approx.) $100(1-\alpha)\%$ prediction interval for y_t is

$[y_t \pm B_t[100(1-\alpha)]]$ where $B_t[100(1-\alpha)]$ is a error bound in a $100(1-\alpha)\%$ prediction interval $[tr_t \pm B_t[100(1-\alpha)]]$ for the deseasonalized observation

$$d_t = TR_t + \varepsilon_t = \beta_0 + \beta_1 t + \varepsilon_t$$

Multiplicative Decomposition – Sales – Homework – slides 52-59

$$Y_t = TC_t \times SN_t \times IR_t$$

Examples:

Quarterly sales from 2010 to 2013 – Assumption no cyclical trend

Period (t)	Year	Quarter	Sales
1	1	1	72
2		2	110
3		3	117
4		4	172
5	2	1	76
6		2	112
7		3	130
8		4	194
9	3	1	78
10		2	119
11		3	128
12		4	201
13	4	1	81
14		2	134
15		3	141
16		4	216

Estimation of seasonal component (SN_t)

$$Y_t = TC_t \times SN_t \times IR_t$$

$$\therefore \qquad \hat{S}N_t = \frac{Y_t}{TC_t \times IR_t}$$

Follow similar steps 1, ….9 articulated on slides 12-14 , adjusting for quarterly (and not monthly) data and ignoring CL_t

Calculation of MA & CMA (steps 1-2)

Period (t)	Year	Quarter	Sales	MA (4)	CMA
1	1	1	72		
2		2	110		
3		3	117	T=2.5 117.75	T=3 118.25
4		4	172	T=3.5 118.75	119
5	2	1	76	119.25	120.875
6		2	112	122.5	125.25
7		3	130	128	128.25
8		4	194	128.5	129.375
9	3	1	78	130.25	130
10		2	119	129.75	130.625
11		3	128	131.5	131.875
12		4	201	132.25	134.125
13	4	1	81	136	137.625
14		2	134	139.25	141.125
15		3	141	143	
16		4	216		

Calculating SN_t (steps 3-6)

Quarter	Sales	MA(4)	TR	SNXIR(t)	Avg(SN(t))	Exp(SN(t))
1	72				0.606	0.60661
2	110				0.918	0.91892
3	117	117.75	118.25	0.989	0.991	0.99199
4	172	118.75	119	1.445	1.481	1.48248
1	76	119.25	120.875	0.629	0.606	0.60661
2	112	122.5	125.25	0.894	0.918	0.91892
3	130	128	128.25	1.014	0.991	0.99199
4	194	128.5	129.375	1.5	1.481	1.48248
1	78	130.25	130	0.6	0.606	0.60661
2	119	129.75	130.625	0.911	0.918	0.91892
3	128	131.5	131.875	0.971	0.991	0.99199
4	201	132.25	134.125	1.499	1.481	1.48248
1	81	136	137.625	0.589	0.606	0.60661
2	134	139.25	141.125	0.95	0.918	0.91892
3	141	143			0.991	0.99199
4	216				1.481	1.48248
					3.996	
					1.001001001	

S1. TR(3) = 117/118.25=0.9894

S2. Average SN(t)
= average(yr1, yr2, yr3)

S3. Normalizing factor
= 4 /sum(average(SN(t))

S4. Expected SN(t)
= S2 * S3

Computation of d_t (step 7)

Period(t)	Year	Quarter	Sales	Exp(SN(t))	d=deseasonalize Y	
1	1	1	72	0.60661	118.69	
2		2	110	0.91892	119.71	
3		3	117	0.99199	117.94	
4		4	172	1.48248	116.02	
5	2	1	76	0.60661	125.29	
6		2	112	0.91892	121.88	
7		3	130	0.99199	131.05	
8		4	194	1.48248	130.86	
9	3	1	78	0.60661	128.58	
10		2	119	0.91892	129.50	
11		3	128	0.99199	129.03	
12		4	201	1.48248	135.58	
13	4	1	81	0.60661	133.53	
14		2	134	0.91892	145.82	
15		3	141	0.99199	142.14	
16		4	216	1.48248	145.70	

Calculating TR_t(step 8)

Period(t)	d=deseasonalize Y	tr=113.685+1.856t	estimate Y = TRXSN
1	118.69	115.54	70.09
2	119.71	117.40	107.88
3	117.94	119.25	118.30
4	116.02	121.11	179.54
5	125.29	122.97	74.59
6	121.88	124.82	114.70
7	131.05	126.68	125.66
8	130.86	128.53	190.55
9	128.58	130.39	79.09
10	129.50	132.25	121.52
11	129.03	134.10	133.03
12	135.58	135.96	201.55
13	133.53	137.81	83.60
14	145.82	139.67	128.34
15	142.14	141.53	140.39
16	145.70	143.38	212.56

Regression for Estimating TR_t

SUMMARY OUTPUT

Regression Statistics	
Multiple F	0.930985
R Square	0.866734
Adjusted I	0.857215
Standard I	3.585966
Observati	16

ANOVA

	df	*SS*	*MS*	*F*	*gnificance F*
Regressio	1	1170.863	1170.863	91.05289	1.66E-07
Residual	14	180.0281	12.85915		
Total	15	1350.891			

	Coefficients	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	113.6851	1.880497	60.45481	2.47E-18	109.6518	117.7183	109.6518	117.7183
X Variable	1.855725	0.194476	9.542164	1.66E-07	1.438615	2.272835	1.438615	2.272835

Measuring Forecast Accuracy :

Let $e_t = Y_t - \hat{Y}_t$ be the errors of forecast.

1) <u>Mean Squared Error</u>

$$MSE = \sum_{t=1}^{n} e_t^2 / n$$

$$RMSE = \sqrt{MSE}$$

MSE = 11.93
RMSE = 3.45

2) <u>Mean Absolute Deviation</u>

$$MAD = \sum_{t=1}^{n} |e_t| / n$$

$$RMAD = \sqrt{MAD}$$

MAD = 2.89
RMAD = 1.70

Additive Decomposition Method

Additive Models

$$y_t = TR_t + SN_t + CL_t + IR_t$$

Where.

y_t = the observed value of the time series at time t

TR_t = the trend component (or factor) at time t

SN_t = the seasonal component (or factor) at time t

CL_t = the cyclical component (or factor) at time t

IR_t = the irregular component (or factor) at time t

Additive Decomposition Method

Step by step approach

1) Calculate **12** month moving average (MA_t) - to eliminate **seasonal** & irregular fluctuation --- Question : If it was quarterly data what would you do???

2) Next calculate Centered Moving Average CMA_t :Why??

Clue : we are calculating moving average of even order. If it was of odd order, this step was not necessary

Since the model is $y_t = TR_t + SN_t + CL_t + IR_t$, this implies that

$$SN_t + IR_t = y_t - (TR_t + CL_t) = y_t - CMA_t$$

Because CMA_t is considered to be an estimate of $TR_t + CL_t$, as averaging process is assumed to have removed seasonal and short term irregular fluctuations

3) Calculate $SN_t + IR_t$

Additive Decomposition Method

Step by step approach

4) Next step calculate average SN_t for each season (Jan, Feb, ….)

5) Then do normalization of average SN_t . As L is the number of seasons the normalization is achieved by subtracting $\sum_1^L SN_t/\mathrm{L}$ from average SN_t so that the sum of the normalized $SN_t = 0$

6) Next step is to calculate the de-seasonalized observation d_t at time t:

$$d_t = y_t - \text{estimated } SN_t$$

Deseasonalized observations are computed in order to better estimate the trend components. By subtracting seasonality from y_t we can have better understanding of trend

7) Next step is to estimate TR_t . Look at plot of d_t , if it is linear please fit $TR_t = \beta_0 + \beta_1 t$ or if quadratic fit $TR_t = \beta_0 + \beta_1 t + \beta_2 t^2$

This completes the estimation of SN_t (step 6) and TR_t

Additive Decomposition Method

Step by step approach

8) So from the equation $y_t = TR_t + SN_t + CL_t + IR_t$

it implies that $CL_t + IR_t = y_t - (TR_t + SN_t)$ where one can use the estimate of TR_t and SN_t

9) It is observed empirically that when considering monthly or quarterly data we can average out ir_t by taking a three period moving average

10) Finally we calculate the estimate of IR_t by using the formula,

$$(cl_t + ir_t) - cl_t$$

Note : Traditionally the estimates tr_t, sn_t, cl_t and ir_t are obtained by using additive decomposition method and used to describe the time series.

However we can also use these estimates to forecast the future values of time series. If there is no pattern in irregular component, we predict IR_t to be 0.

Additive Decomposition Method

Forecasting using the method: In case $IR_t = 0$,

A) The point forecast of $y_t = tr_t + sn_t + cl_t$, if a well defined cycle exists and can be predicted

B) The point forecast of $y_t = tr_t + sn_t$, if a well defined cycle does not exist or if CL_t can not be predicted

Although there is no theoretically correct prediction interval for y_t Bowerman , O'Connell & Koehler have found that a fairly accurate (approx.) $100(1\text{-}\alpha)\%$ prediction interval for y_t is

$[y_t \pm B_t[100(1\text{-}\alpha)]]$ where $B_t[100(1\text{-}\alpha)]$ is a error bound in a $100(1\text{-}\alpha)\%$ prediction interval $[tr_t \pm B_t[100(1\text{-}\alpha)]]$ for the deseasonalized observation

$$d_t = y_t - sn_t$$

EXPONENTIAL SMOOTHING

Smoothing Methods

Smoothing

- Respond to the most recent behavior of the series
- Employ the idea of weighted averages
- They range in the degree of sophistication

1. Simple Averages - quick, inexpensive (should only be used on stationary data)
2. Moving Averages - a constant number specified at the outset and a mean computed for the *most recent observations* - such as a 3 or 4 period moving average.
   - Works best with stationary data.
   - The larger the order of the moving average, the greater the smoothing effect. Larger *n* when there are wide, infrequent fluctuations in the data.
   - By smoothing recent actual values, removes randomness.

Smoothing Methods – Formula Summary

- Simple average
  - m ~ most recently observation

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + .. + Y_{t-m+1}}{m}$$

- Moving average Smoothing (Trailing moving average)
  - w ~ window / width or interval

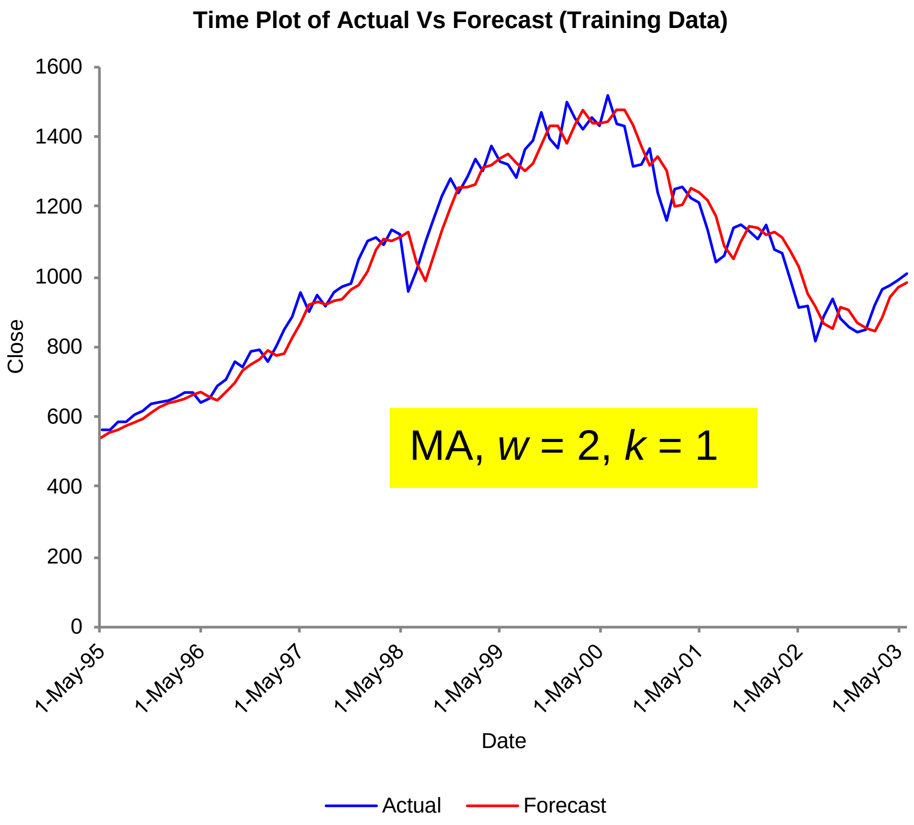
$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + .. + Y_{t-w+1}}{w}$$

- Simple Exponential Smoothing
  - Smoothing parameter $0 < \alpha < 1$

$$\hat{Y}_{t+1} = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} + \ldots +$$

Simple Moving Average Smoothing

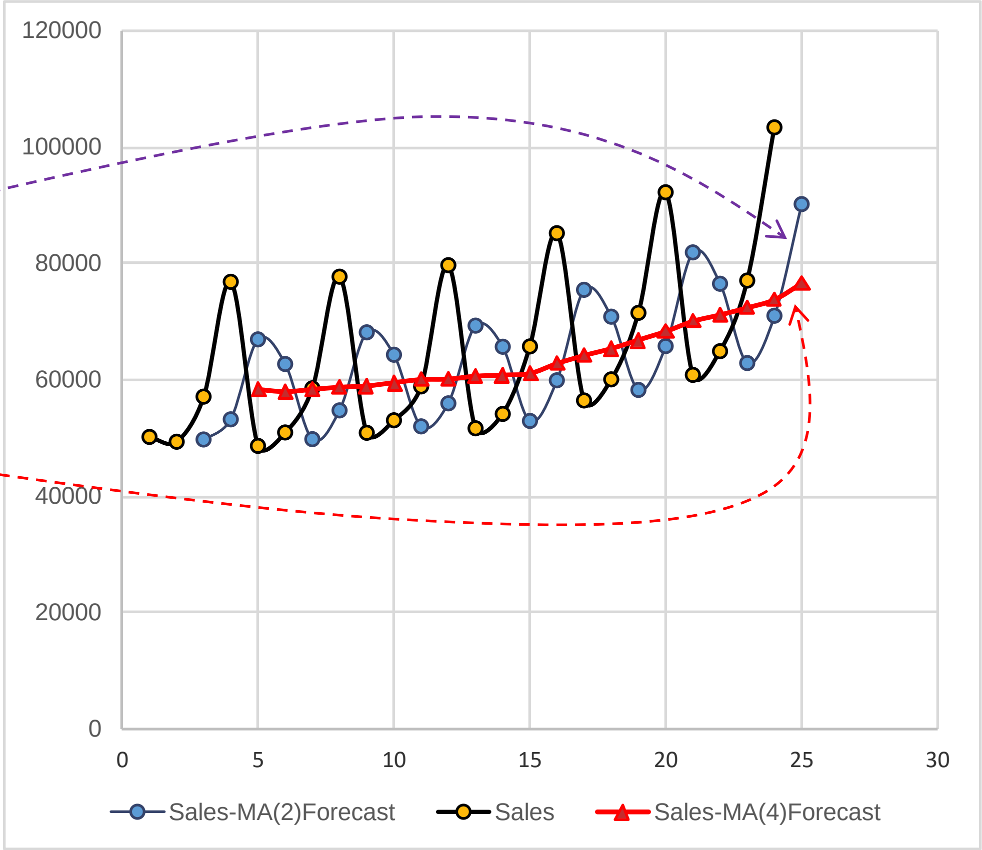
- Example: *SP500 Index*
  - At current time point *t*, called ***forecast origin***
  - Forecast ***k***-step-ahead, *k* is the ***forecast horizon***
  - MA with *w*, called ***window width***, is the average of the most recent available *w* values of the series, i.e.: the average of a ***moving window***



$$Y_{t+k} = (Y_t + Y_{t-1} \dots + Y_{t-w+1})/w$$

Simple Moving Average Smoothing (cont.)

- Example: *Department Store Quarterly Sales*
  - The original time series *Sales*
  - 1-step-ahead forecast with $w = 2$
  - 1-step-ahead forecast with $w = 4$



Moving Averages

Observation	Demand	3 Period Moving Average
1	7	-
2	14	-
3	11	-
4	19	10.6667
5	9	14.6667
6	8	13
7	12	12
8	11	9.6667
9	7	10.3333
10	10	10
11	10	9.3333
12		9

- Simply forecast the next period observation to be the average of m previous observations
- Underlying assumption – the time series has a constant level (over the long term)
  - Very bad for time series with trend or seasonality
- How large should m be?
  - Large m (say 20) tends to smooth out the variability
  - Small m (say 5) useful for capturing recent changes

Exponential Smoothing

- Exponential Smoothing provides a forecasting method that is most effective when the components (trend and seasonal factors) of the time series may be changing over time
- More recent observations are weighted more heavily than more remote observations
- The unequal weighting is accomplished by using one or more smoothing constants, which determine how much weight is given to each observation

Simple Exponential Smoothing

- Instead of taking simple average over the *w* most recent values, ***simple exponential smoothing*** takes a ***weighted average*** of ***all*** past values in the given sample series

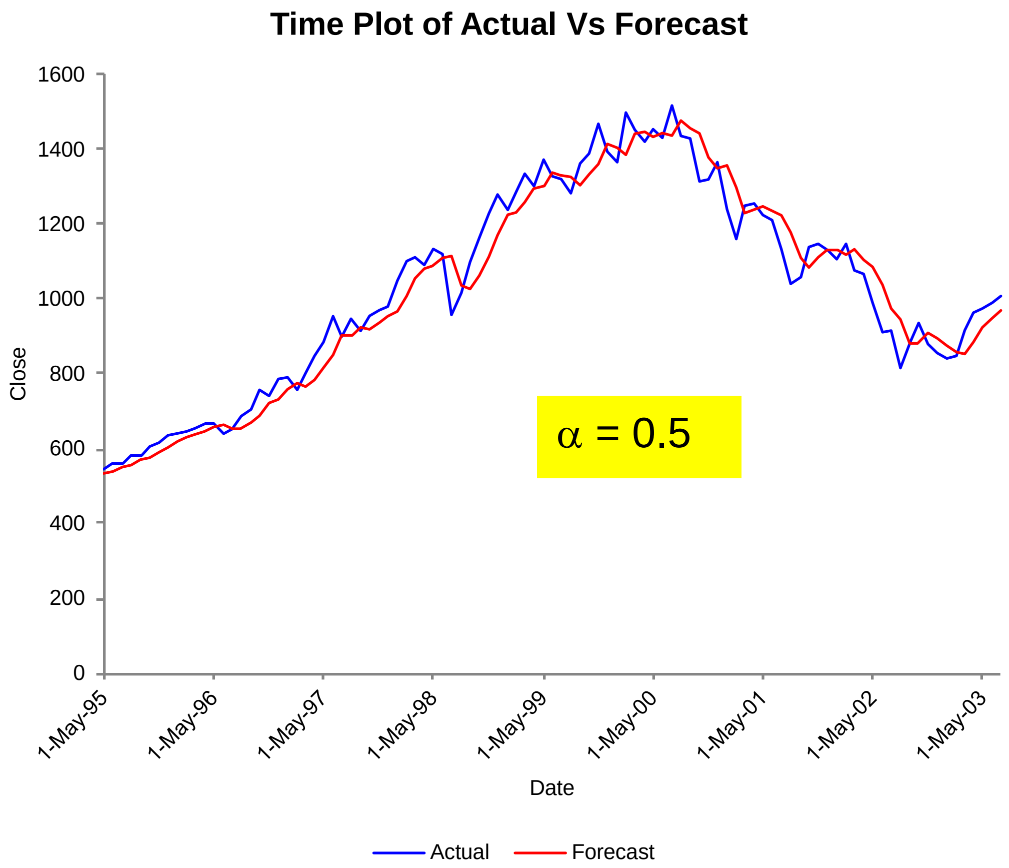
$$Y_{t+1} = \alpha Y_t + \alpha(1-\alpha)Y_{t-1} + \alpha(1-\alpha)^2 Y_{t-2} \ldots + \ldots$$

smoothing parameter $0 < \alpha < 1$

- Impact of α?
  - **Close to 1:** only the most recent observations have influence on the forecast
  - **Close to 0:** more (almost all) past observations have an influence on the forecast

Simple Exponential Smoothing: *parameter α*

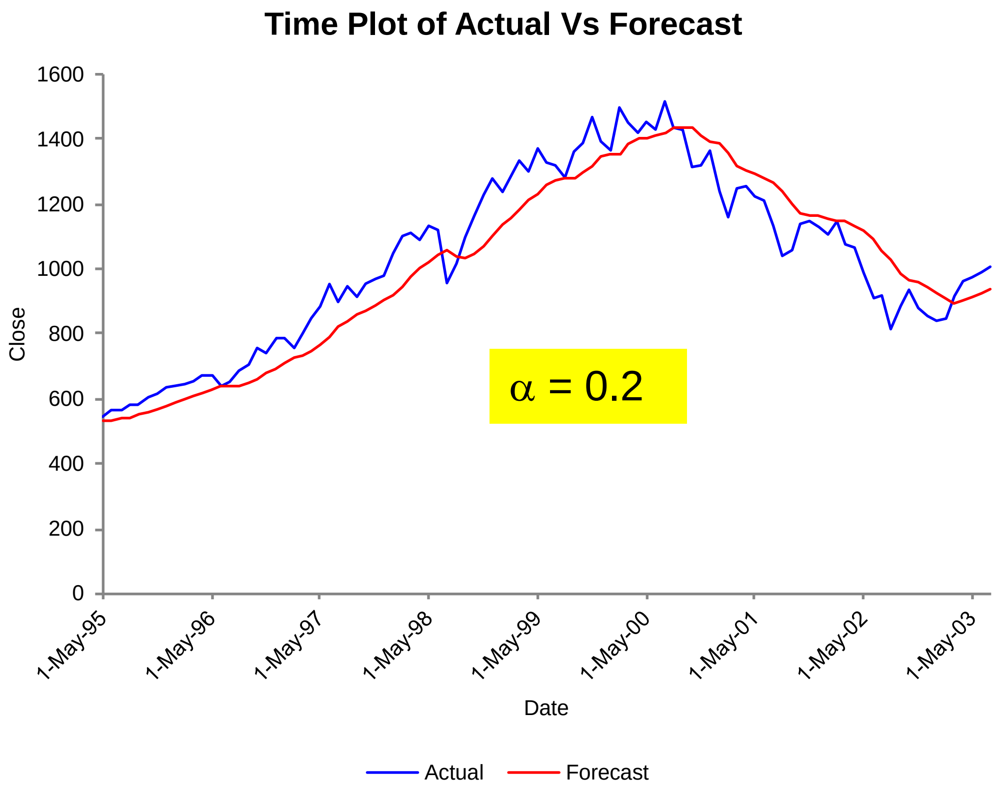
- The same series from SP500



Simple Exponential Smoothing: *parameter α*

- The same series from SP500
  - It appears an over smoothing

    (more TS methods are to be introduced later)



Exponential Smoothing - Formulation

- If $y(1), y(2), \ldots$, is a time series of observations in periods 1, 2, … and $F(2), F(3), \ldots$ the series of forecasts in periods 2, 3, … then

$$F(t+1) = \alpha y(t) + (1-\alpha)F(t)$$

$$= \alpha y(t) + (1-\alpha)\{\alpha\, y(t-1) + (1-\alpha)F(t-1)\}$$

$$= \alpha y(t) + \alpha(1-\alpha)y(t-1) + (1-\alpha)\,\{\alpha\, y(t-2) + (1-\alpha)F(t-2)\}$$

$$= \alpha y(t) + \alpha(1-\alpha)y(t-1) + \alpha(1-\alpha)\,\{y(t-2) + \ldots$$

- Underlying assumption – the time series has a constant or slowly changing mean
  - No trend or seasonality
- How large should α be?
  - If $\alpha = 1$, the forecast is the last observed value
  - If $\alpha = 0$, the forecast is the last forecast
  - Large α (say 0.8) tends to capture changes in the underlying process fast but are sensitive to noise (random fluctuations)
  - A compromise is required; values between 0.1 and 0.25 often used

Exponential Smoothing

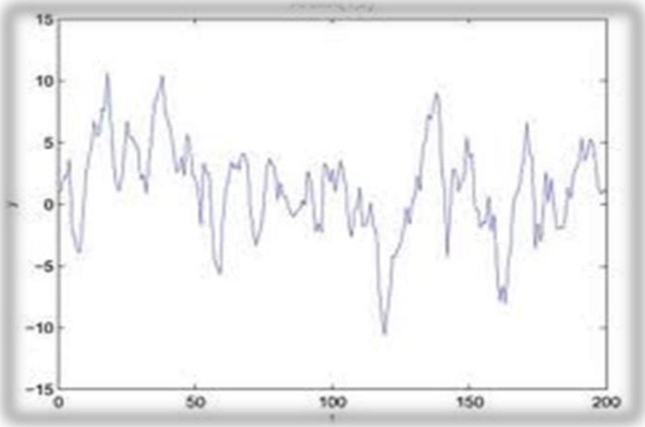
Observation	Demand	Exponential Smoothing (Damping Factor .3)
1	7	-
2	14	7
3	11	11.9
4	19	11.27
5	9	16.681
6	8	11.3043
7	12	8.99129
8	11	11.097387
9	7	11.0292161
10	10	8.20876483
11	10	9.462629449
12		

- All observations are **NOT** weighted equally
- Greater weight is given to more recent observations.
- Suppose that the damping factor is 0.3 (α =0.7)
- 70% weight is given to the most recent observation.
  - 70% * 30% = 21% is given to the 2nd most recent, etc.

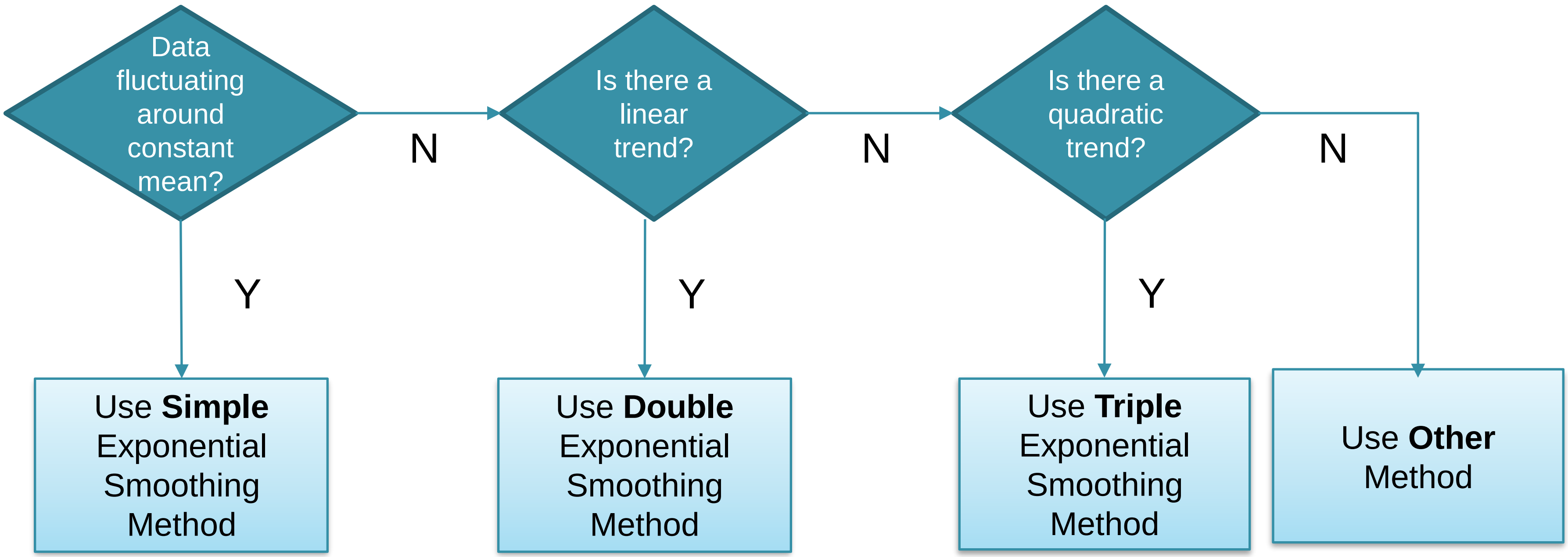
Other Exponential Smoothing Methods

- Holt's Method
  - Generalizes exponential smoothing
  - Can be used for series which show trend
  - Two smoothing parameters

- Winter's Method
  - Can be used for series which show trend and seasonality
  - Three smoothing parameters



Three types of Exponential smoothing methods



Simple Exponential Method

- Data fluctuating a constant mean, there is no trend .
- Consider **no trend equation**, where β_0 is changing over time.

$$y_t = \beta_0 + \varepsilon_t$$

Step 1 : Find Initial estimate $a_0(0)$

$$a_{0(0)} = \bar{y} = \frac{\sum y_t}{n}$$

Step 2 : Compute $a_0(T)$ with an arbitrary α

$$a_{0(T)} = \alpha y_T + (1-\alpha) a_{0(T-1)}$$

Step 3 : Compute MSE/RMSE/MAD/MAPE

$$\text{MSE} = \sum_{i=1}^{n} \frac{\left(Y_i - \hat{Y}_i\right)^2}{n}$$

Step 4 : Select α that yield the lowest in step 3.

Double Exponential Smoothing

Double exponential smoothing can be used to apply unequal weightings when β_0 and β_1 are slowing changing over time.

$$Y_t = \beta_0 + \beta_{1t} + \varepsilon_t$$

There are two versions of double exponential smoothing.

1) One parameter double exponential smoothing, using one smoothing constants. *(Brown's linear exponential smoothing)*

2) Holt two parameter exponential smoothing, using two smoothing constants.

$$\hat{y}_{t+h|t} = \ell_t + hb_t$$
$$\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1} + b_{t-1})$$
$$b_t = \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)b_{t-1}$$

Holt's Trend Corrected Exponential Smoothing (2 parameters)

Suppose a time series displays a linear trend. Then the series can be described by the linear trend model :

$$y_t = \beta_0 + \beta_1 t + \varepsilon_t$$

Therefore level (or mean) at time T is $\beta_0 + \beta_1$ T and

level (or mean) at time T-1 is $\beta_0 + \beta_1$ (T-1)

So increase or decrease in the level of time series from time T-1 to T is:

$$(\beta_0 + \beta_1 T) - (\beta_0 + \beta_1 (T-1)) = \beta_1$$

This fixed rate of increase or decrease β_1 is called growth rate

Holt's trend corrected exponential smoothing is appropriate when both the level and the growth rate are changing

Holt's Trend Corrected Exponential Smoothing (2 parameters)

A model different from the linear model is needed to describe the changing level and growth rate

To implement Holt's trend corrected exponential smoothing we let ℓ_{T-1} denote the estimate of the level of the time series at time period T-1 and we denote b_{T-1} denote the estimate of the growth rate of the time series in time T-1. Then if we observe a new time series value y_T in time period T, we use two smoothing equations to update the estimates ℓ_{T-1} and b_{T-1}

$$\ell_T = \alpha y_T + (1-\alpha)(\ell_{T-1} + b_{T-1})$$

$$b_T = \gamma [\ell_T - \ell_{T-1}] + (1-\gamma) b_{T-1}$$

where α and γ are smoothing constants between 0 and 1

Additive Holt-Winters Method (3 parameters)

Suppose a time series displays a linear trend locally and has a seasonal pattern with constant (additive) seasonal variation and that the level, growth rate and seasonal pattern may be changing. Then the estimate ℓ_T for the level, the estimate b_T for the growth rate, and the estimate sn_T for the seasonal factor of the time series in time T is given by the smoothing equations :

$$\ell_T = \alpha (y_T - sn_{T-L}) + (1-\alpha)(\ell_{T-1} + b_{T-1})$$

$$b_T = \gamma [\ell_T - \ell_{T-1}] + (1-\gamma) b_{T-1}$$

$$sn_T = \delta [y_T - \ell_T] + (1-\delta) sn_{T-1}$$

where α , γ and δ are smoothing constants between 0 and 1

AUTOCORRELATION

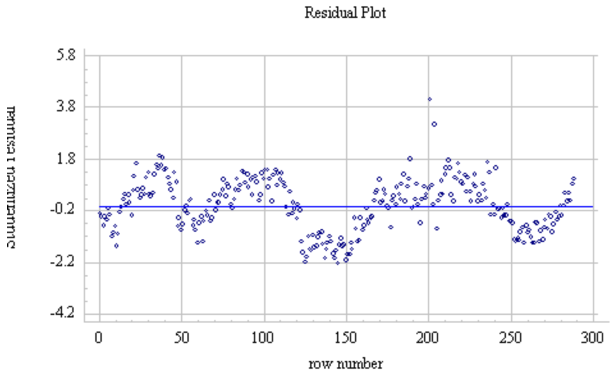
Detecting Autocorrelation

The validity of the regression method requires that the errors are independent. However when time series data are being analyzed this assumption is often violated. It is quite common that the time ordered error terms to be auto correlated

Just as correlation (r) measures the extent of linear relationship between two variables; **Auto-correlation** measures the linear relationship between lagged values of a time series. It can be of two types : positive and negative

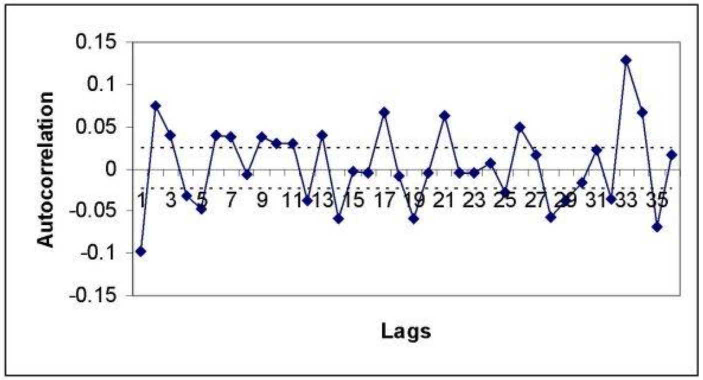
(exactly like simple correlation)

Positive Autocorrelation



Cyclical Pattern

Negative Autocorrelation



Alternating Pattern

Detecting Autocorrelation

Error terms occurring over time have **positive auto-correlation**
1) if a positive error term in time period t tends to produce, or be followed by another positive error term in time period t+k ,

& have **negative auto-correlation**

2) if a negative error term in time period t tends to produce, or be followed by another negative error term in time period t+k

\+ Autocorrelation : Error sign + + + + - - - + + + - - - - - + + + +

\- Autocorrelation : Error sign + - + - - + - + - + + - + - + - + +

Random : Error sign + - + + + - + - - + - + - - +

If the error term $\epsilon_t = \Phi_1 \epsilon_{t-1} + a_t$ can be expressed in this form where ϵ_t = error term in time period t , Φ_1 = correlation co-efficient and a_1 , a_2 , ... are i.i.d Normal with mean 0 and a variance independent of time then there exists **first-order autocorrelation**

Durbin-Watson Test for (First Order) +ve/-ve Autocorrelation Detection (n observations)

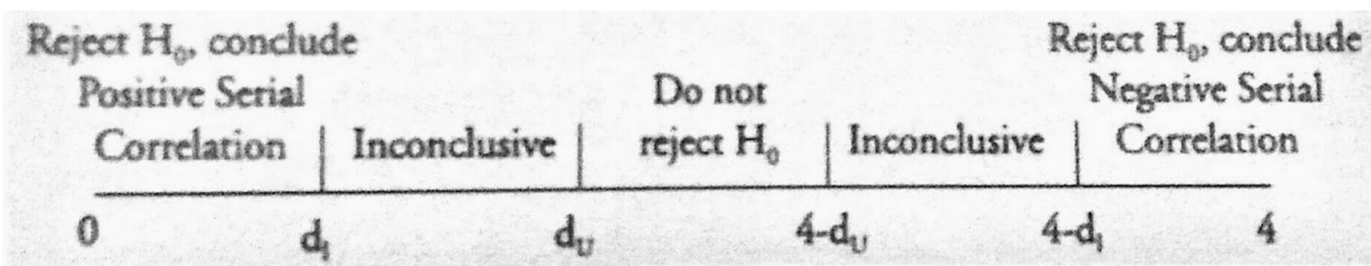
The Durbin-Watson Statistic

H_0: positive autocorrelation does not exist

H_1: positive autocorrelation is present

- Calculate the Durbin-Watson test statistic = D
  (The Durbin-Watson Statistic can be found using Excel)
- Find the values d_L and d_U from the Durbin-Watson table

$$D = \frac{\sum_{t=2}^{n} (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^{n} \varepsilon_t^2}$$



Autocorrelation Revisited

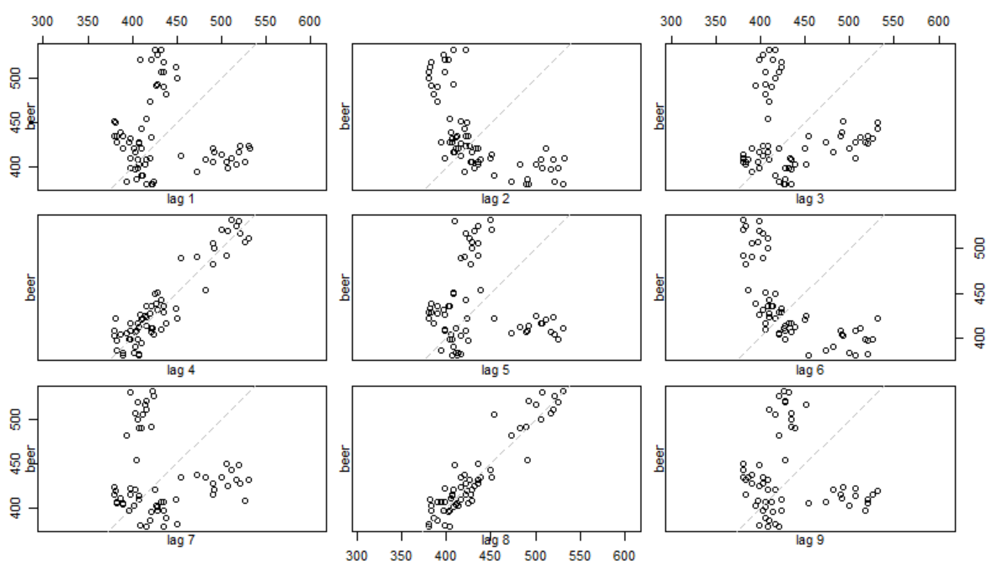
- Co-variances are often difficult to interpret because they depend on the units of measurement of the data. Correlations can be obtained through computing the autocorrelations of a time series.
- **Autocorrelations** are statistical measures that indicate how a time series is related to itself over time
- The autocorrelation at **lag** 1 is the correlation between the original series Y_t and the same series moved forward one period (represented as Y_{t-1})

Covariance & correlation

$$\rho\sigma_X\sigma_Y = \sigma_{XY}$$

- Correlation

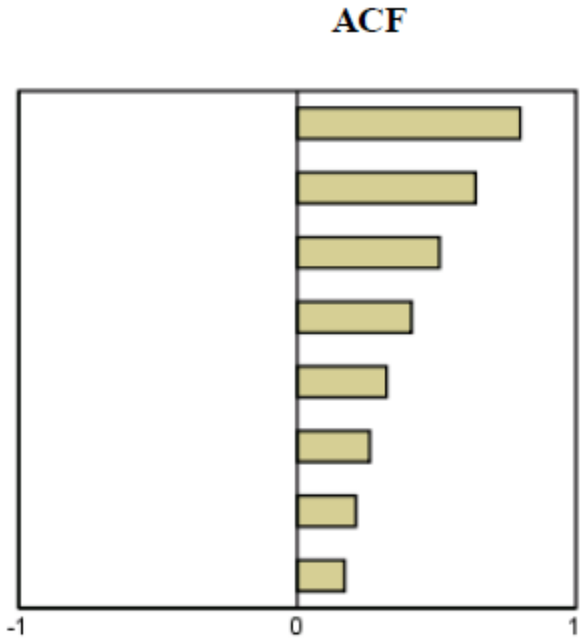
$$\rho = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{\mathrm{cov}(X,Y)}{\mathrm{StandardDev}(X)\times\mathrm{StandardDev}(Y)}$$



Autocorrelations (cont.)

- The theoretical autocorrelation function

$$\rho_k = \frac{E[(y_t - \mu)(y_{t+k} - \mu)]}{\sigma_y^2} = \frac{\text{cov}[y_t y_{t+k}]}{\text{var}(y_t)}$$



- The sample autocorrelation

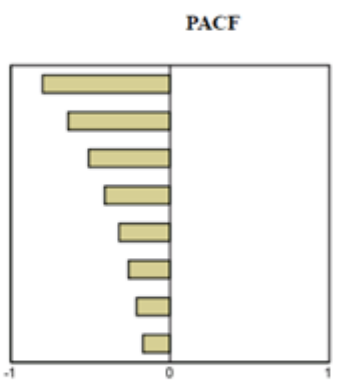
$$r_k = \frac{\sum_{t=1}^{N-k}(y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{N}(y_t - \bar{y})^2} \qquad k = 0,1,2,\ldots k$$

- In practice, to obtain a useful estimate of the autocorrelation function, at least 50 observations are needed
- The estimated autocorrelations r_k would be calculated up to lag no larger than N/4

consider only till 12 periods.. beyond that pointless..

Partial-Autocorrelations (PACs)

- Partial-autocorrelations are another set of statistical measures used to identify time series models
- It measures the strength of the relationship between observations in a series controlling for the effect of the intervening time periods.
- PAC is Similar to AC, except that when calculating it, the ACs with all the elements within the lag are partialled out (Box & Jenkins, 1976)



<u>Partial autocorrelations</u> are used to measure the degree of association between Y_t and Y_{t-k}, when the effects of other time lags (1, 2, 3, …, k – 1) are removed.

$$\frac{\text{Covariance}(y, x_3 | x_1, x_2)}{\sqrt{\text{Variance}(y | x_1, x_2)\text{Variance}(x_3 | x_1, x_2)}}$$

correlate the “parts” of y and x_3 that are not predicted by x_1 and x_2.

AUTOREGRESSIVE MODELS

Autoregressive (AR) Models

- Recall in linear regression, auto-correlation must not exist when applying SLR.

- In the case of AR Model, the target variable is expressed as a linear regression of its history.

$$Y_t = b_0 + b_1 Y_{t-1} + b_2 Y_{t-2} + b_3 Y_{t-3} + \ldots\ldots\ldots\ldots$$

AR Models (cont'd)

$$Y_t = \Phi_0 + \Phi_1 Y_{t-1} + \Phi_2 Y_{t-2} + \Phi_3 Y_{t-3} + \ldots\ldots\ldots\ldots$$

- Auto-Regressive ~ Regression on itself (Predicting *Y* using *Y* history)

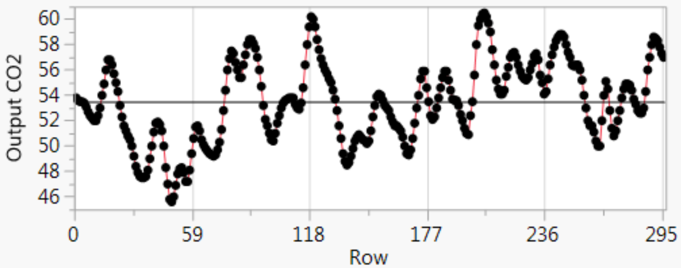
- Lets assume that Y_t can be predicted base on Y_{t-1}

- Error = e_t= actual - expected = $y_t - \boxed{y_t} = y_t - \Phi_0 - \Phi_1 y_{t-1}$
- Full model : $y_t = \Phi_0 + \Phi_1 y_{t-1} + e_t$

- Φ_0 & Φ_1 are estimated based on minimizing SSE

Example

Time Series Output CO2

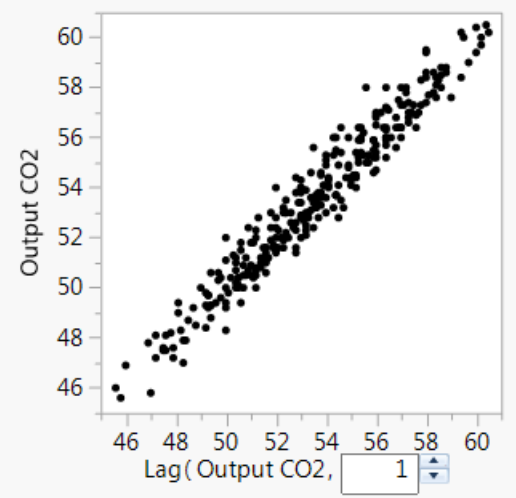


Mean	53.509122
Std	3.1967072
N	296
Zero Mean ADF	0.137862
Single Mean ADF	-1.863957
Trend ADF	-2.257128

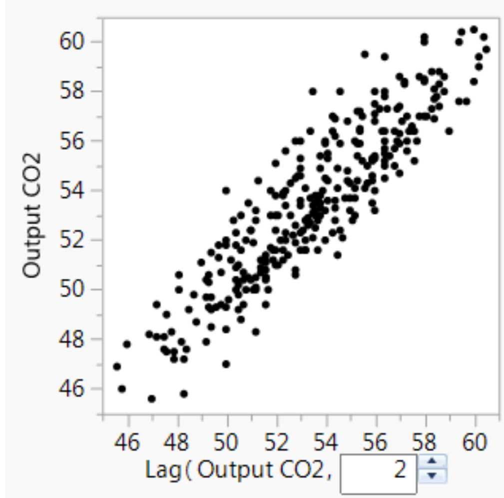
Time Series Basic Diagnostics

Lag	AutoCorr	-.8-.6-.4-.2 0 .2 .4 .6 .8	Ljung-Box Q	p-Value	Lag	Partial	-.8-.6-.4-.2 0 .2 .4 .6 .8
0	1.0000		.	.	0	1.0000	
1	0.9708		281.778	<.0001*	1	0.9708	
2	0.8960		522.666	<.0001*	2	-0.8039	
3	0.7925		711.766	<.0001*	3	0.1883	
4	0.6800		851.434	<.0001*	4	0.2600	
5	0.5745		951.471	<.0001*	5	0.0595	
6	0.4854		1023.15	<.0001*	6	-0.0626	
7	0.4161		1075.99	<.0001*	7	-0.0144	
8	0.3656		1116.93	<.0001*	8	0.0549	
9	0.3304		1150.48	<.0001*	9	0.0055	
10	0.3065		1179.46	<.0001*	10	0.0314	
11	0.2880		1205.12	<.0001*	11	-0.1166	
12	0.2693		1227.65	<.0001*	12	-0.0430	
13	0.2473		1246.71	<.0001*	13	0.0511	
14	0.2215		1262.05	<.0001*	14	0.0538	
15	0.1930		1273.74	<.0001*	15	-0.0461	
16	0.1649		1282.30	<.0001*	16	0.0339	
17	0.1398		1288.48	<.0001*	17	-0.0033	
18	0.1210		1293.12	<.0001*	18	0.0855	
19	0.1103		1297.00	<.0001*	19	0.0165	
20	0.1078		1300.72	<.0001*	20	-0.0258	
21	0.1112		1304.68	<.0001*	21	-0.0492	
22	0.1171		1309.10	<.0001*	22	0.0097	
23	0.1228		1313.96	<.0001*	23	0.0491	
24	0.1259		1319.10	<.0001*	24	-0.0012	
25	0.1259		1324.26	<.0001*	25	0.0043	

Lag Plot



Lag Plot



Example (Cont'd)

Model: AR(3)

Model Summary

DF	292	Stable	Yes
Sum of Squared Errors	34.2952484	Invertible	Yes
Variance Estimate	0.11744948		
Standard Deviation	0.34270903		
Akaike's 'A' Information Criterion	216.272318		
Schwarz's Bayesian Criterion	231.033756		
RSquare	0.98864058		
RSquare Adj	0.98852387		
MAPE	0.46466563		
MAE	0.24796464		
-2LogLikelihood	208.272318		

Parameter Estimates

Term	Lag	Estimate	Std Error	t Ratio	Prob>|t|	Constant Estimate
AR1	1	2.19621	0.0512770	42.83	<.0001*	1.46625292
AR2	2	-1.68410	0.0963793	-17.47	<.0001*	
AR3	3	0.46054	0.0513752	8.96	<.0001*	
Intercept	0	53.61491	0.7062130	75.92	<.0001*	

Important Assumptions

- Linear relationship between successive values
- Y_t is a Stationary process
- Errors are normal and independent

STATIONARY PROCESS

Stochastic (random) Process

- A stochastic process is a collection $\{Y_t : t = 1, 2, \ldots, T\}$ of random variables ordered in time. Example : the error term in a linear regression model is assumed to be a stochastic process.
- A stochastic process is weakly stationary if for all t values

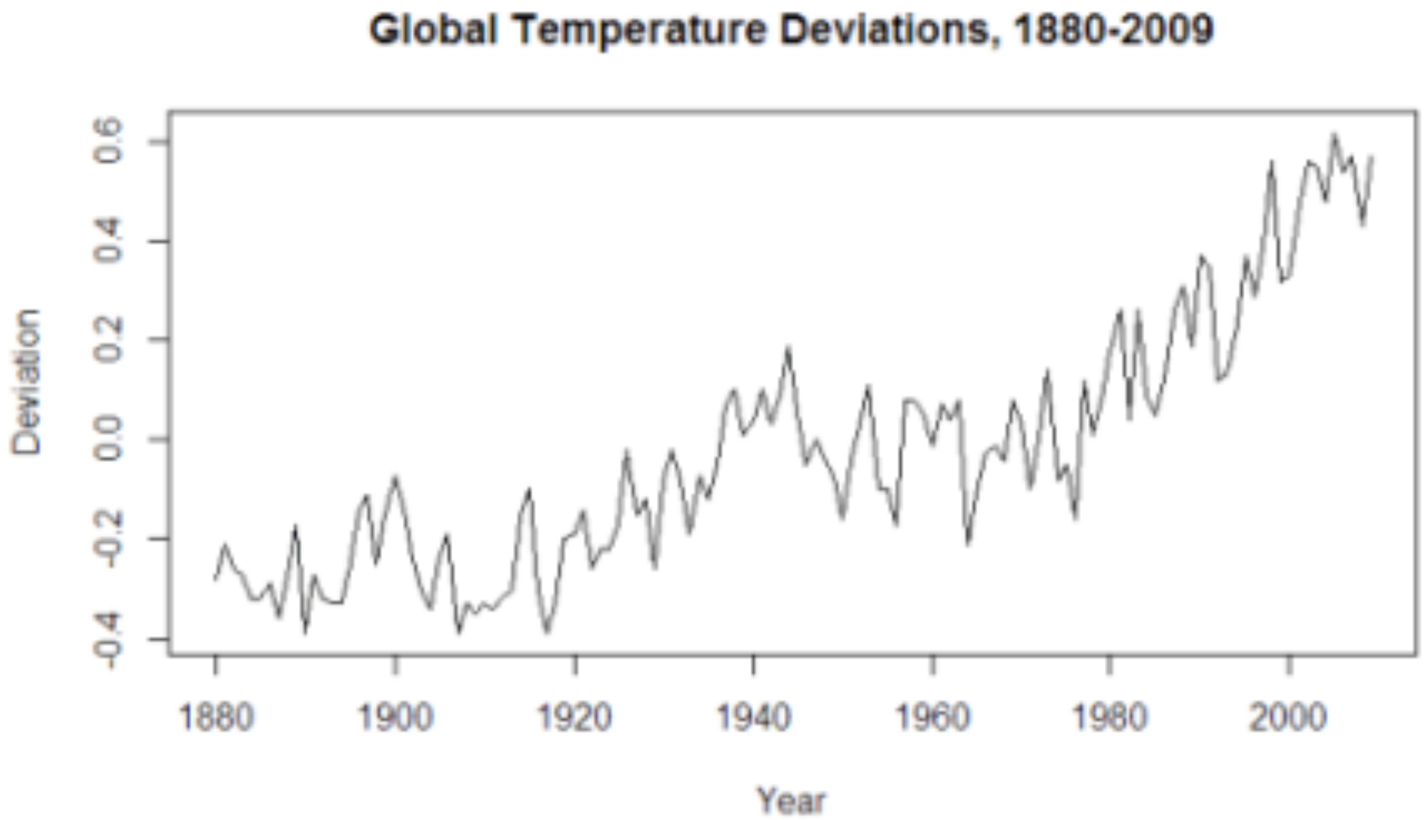
$$
\begin{aligned}
E[Y_t] &= \mu \\
\mathrm{var}(Y_t) &= \sigma^2 \\
\mathrm{cov}(Y_t Y_{t-k}) &= \gamma_k \qquad \forall \qquad t
\end{aligned}
$$

i.e. its statistical properties do not change over time

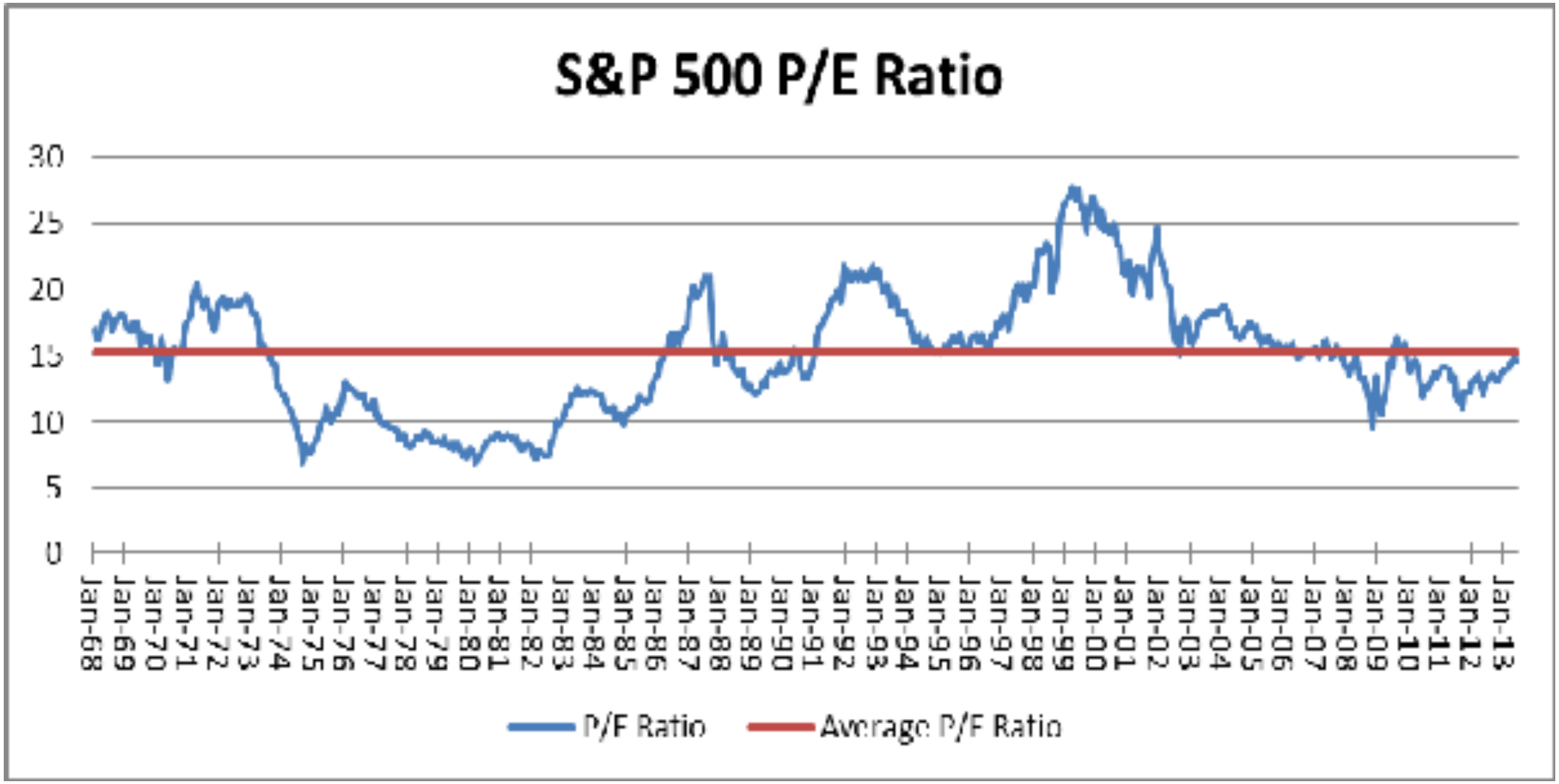
Stationary Stochastic Processes

- In order to model a time series with the Box-Jenkins approach, the series **has to be stationary**
- Thus in order to tentatively identify a Box-Jenkins model we must first determine whether the series is stationary
- Intuitively a time series is stationary if the statistical properties (e.g. mean and variance) of the series are essentially constant over time
- If not then it is non stationary
- Most time series data are **nonstationary**

Some Nonstationary series



Some nonstationary series (cont.)



Achieving Stationarity - Differencing

- If a series is nonstationary, a standard procedure to transform it to a stationary process is to take the differences of consecutive observations

- Regular differencing (RD)

(1st order) $\nabla y_t = y_t - y_{t-1}$

(2nd order) $\nabla^2 y_t = y_t - 2y_{t-1} + y_{t-2}$ =yt - yt-1 - (yt-1 - yt-2)

- It is unlikely that more than two regular differencing would ever be needed

- Sometimes regular differencing by itself **is not** sufficient and prior transformation is also needed

for non-linear series

Backshift Operator (B) – Elegant Mathematical Expression used to Denote Differencing

The backshift operator B is useful when working with time series model expression.

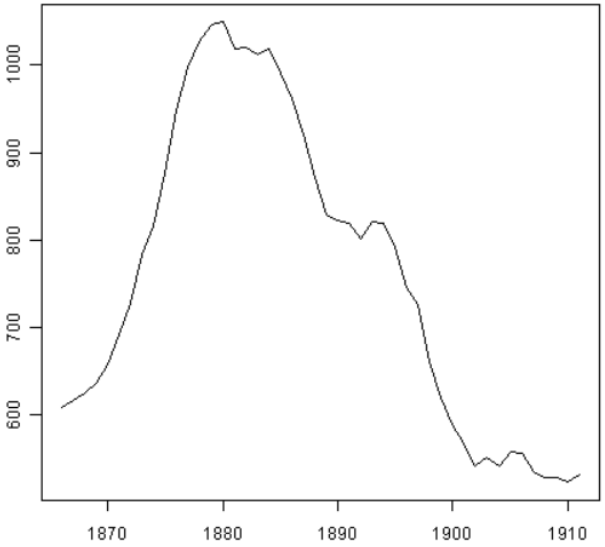
$$By_t = y_{t-1}$$
$$B(By_t) = B^2 y_t = y_{t-2}$$

Hence, 1st order differencing can be written as

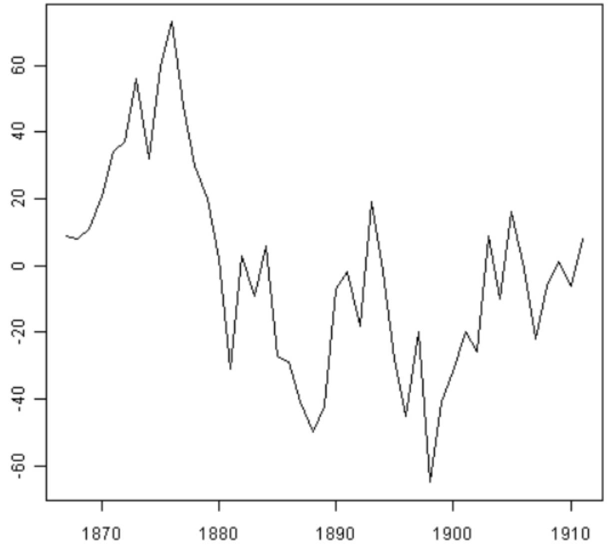
$$y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t$$

& 2nd order differencing : $(1 - B)^2 y_t$

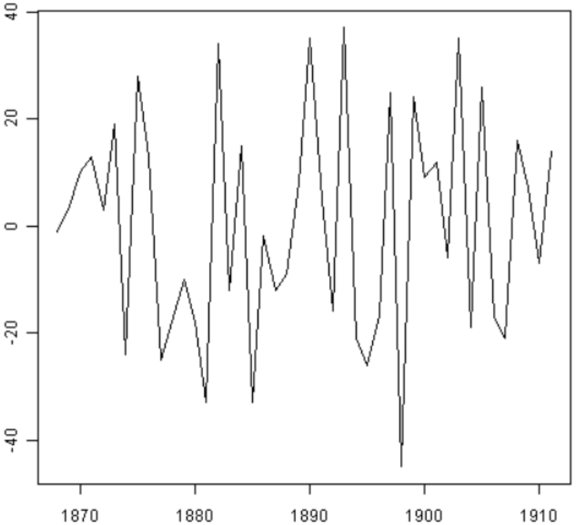
What happens to the series after differencing?



y_t



$y'_t = y_t - y_{t-1} = y_t - By_t = (1- B)y_t$



$y''_t = (1- B)^2 y_t$

MOVING AVERAGE(MA) MODELS

Moving Average Models (MA)

- Moving Average model of order q (MA(q))

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

i.e., y_t depends on q previous random error terms.

- Using backshift notation

$$y_t = \mu + \theta_q(B)\varepsilon_t$$

where

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_q B^q$$

- The model is called a moving average because it is defined as a moving average of the error series, $\boxed{\varepsilon_t.}$
- Here we use *moving average* only in reference to a model of the above form.

BOX-JENKINS MODEL - ARIMA

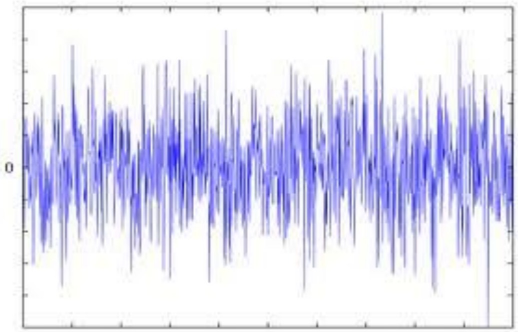
The White Noise Process

- The Box-Jenkins models are based on the idea that a time series can be usefully regarded as generated from (driven by) a series of **uncorrelated independent** "**shocks**" $\boxed{e_t}$

$$E[e_t] = 0 \quad var[e_t] = \sigma_e^2$$

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

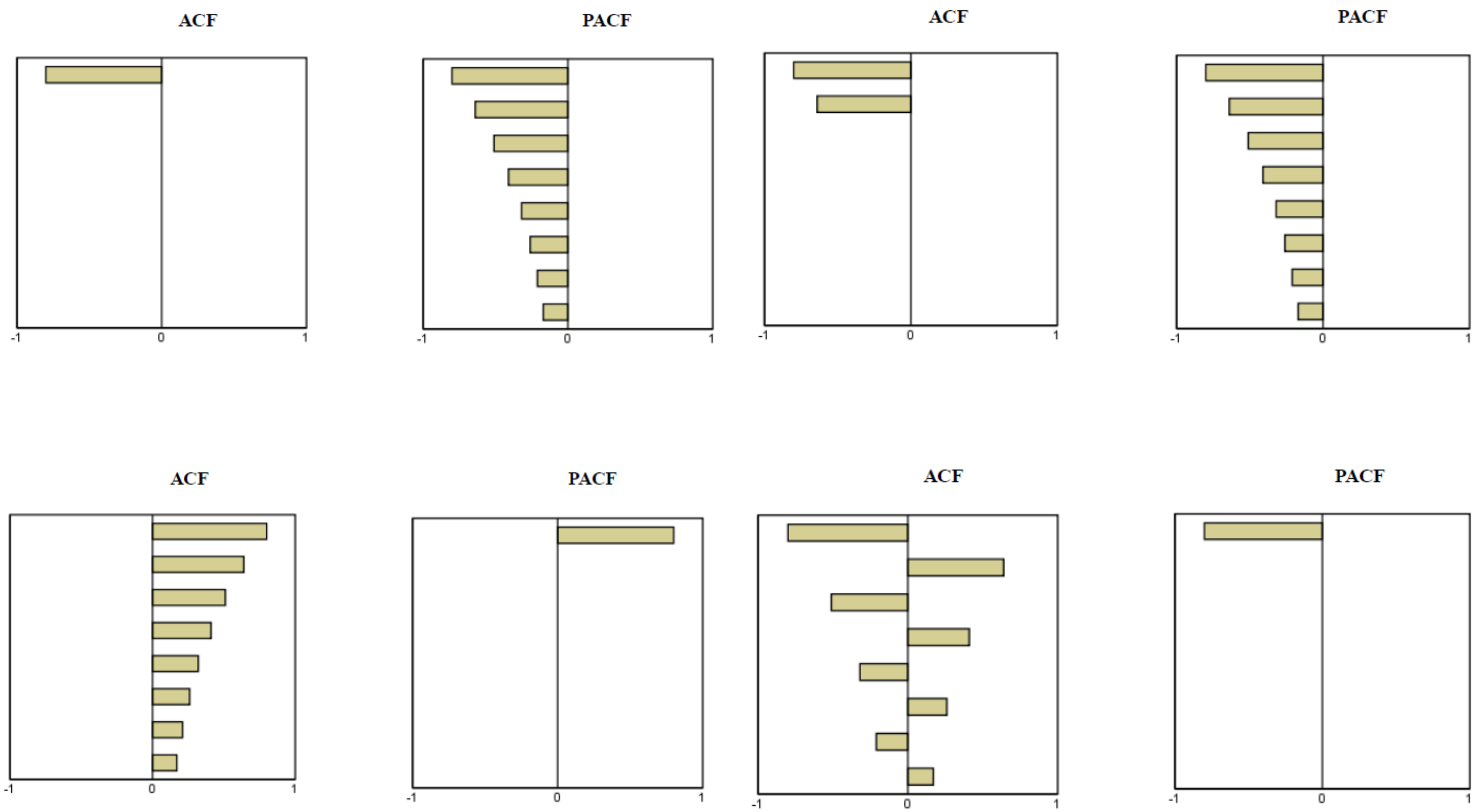
- Such a sequence e_t, e_{t-1}, e_{t-2},… is called a "white noise process"



Note: Assessing Stationarity

- Box-Jenkins forecasting models are tentatively identified by examining the behavior of Autocorrelation function (ACF) and Partial Autocorrelation function (PACF)
- The **ACF and PACF** plots can readily expose non-stationarity in the mean.
  - The autocorrelations of a stationary series drop to zero relatively quickly.
  - For a typical pattern of non-stationary series
    - ρ_1 is very large and positive (ACF plot).
    - ρ_k's are relatively large and positive, until k gets big enough (ACF plot).
    - PACF plot displays a large spike close to 1 at lag 1.

Various Patterns of ACF & PACF



AR Model Building

Autoregressive Models (AR)

- Autoregressive model of order *p* (AR(*p*))

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

  i.e., y_t depends on its *p* previous values
- Using backshift notation

$$\phi_p(B) y_t = \phi_0 + \varepsilon_t$$

  where

$$\phi_p(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$$

- Autoregression should be treated differently from ordinary regression models.
  - The basic assumption of independence of the error terms can easily be violated, since the explanatory variables usually have a built-in dependence relationship.
  - Determining the number of past values of y_t is not always straightforward.

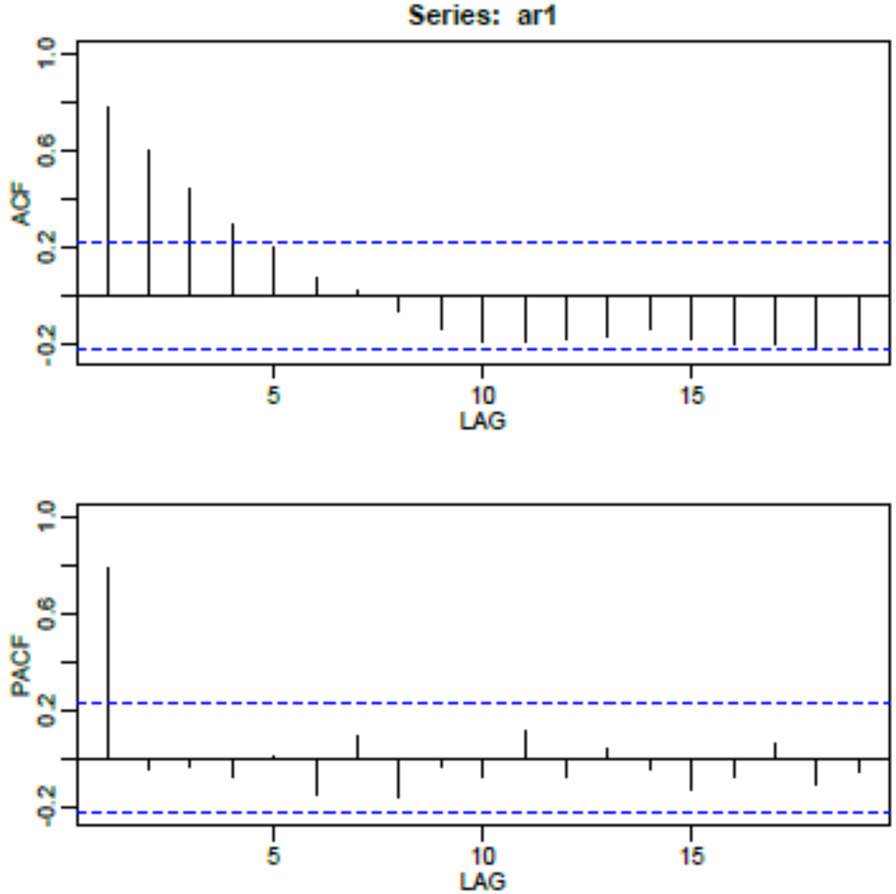
An Autoregressive Model of Order one AR(1)

- The basic form of AR(1) is:
  - Observation y_t depends on y_{t-1}.
  - The value of autoregressive coefficient ϕ_1 is between –1 and 1.

$$y_t = C + \phi_1 y_{t-1} + e_t$$

- This model is also known as ARIMA (1,0,0)

Theoretical ACF and PACF for AR(1)



- Characteristics:
  - ACF dies down
  - PACF cuts off for lag > 1

AR(p)

- AR(2) process

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \qquad (where\ \varepsilon_t\ is\ white\ noise)$$

where $|\phi_2| < 1$, $\phi_2 + \phi_1 < 1$, and $\phi_2 - \phi_1 < 1$, which are the stationarity requirement for an AR(2) process.

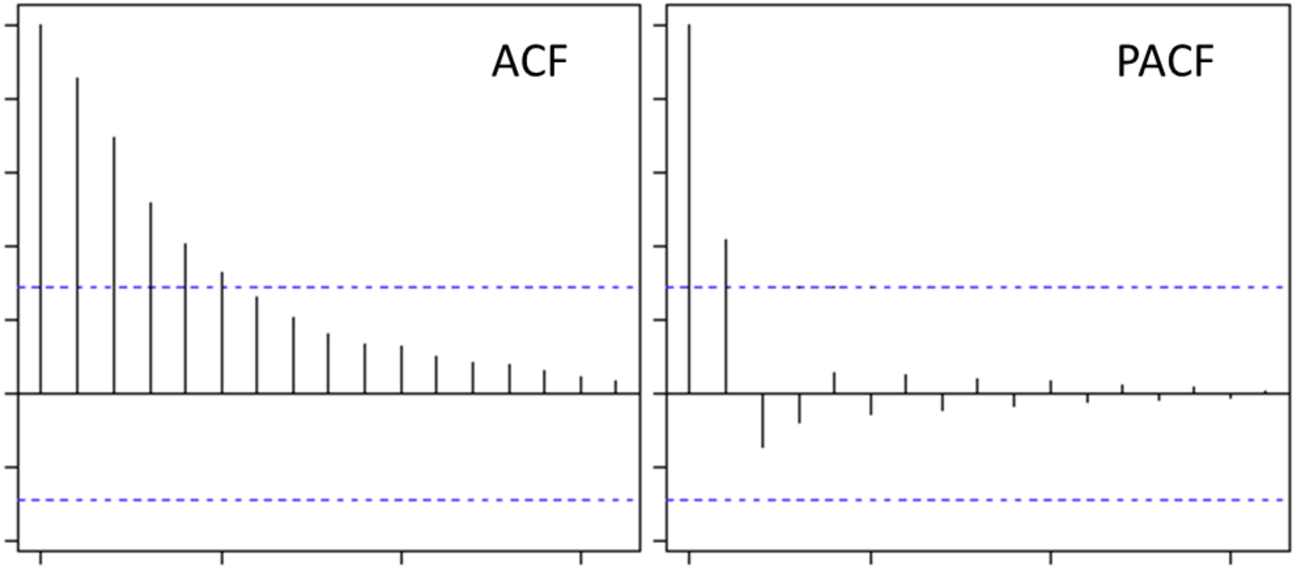
- AR(p) process

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

$$where\ \varepsilon_t\ is\ white\ noise$$

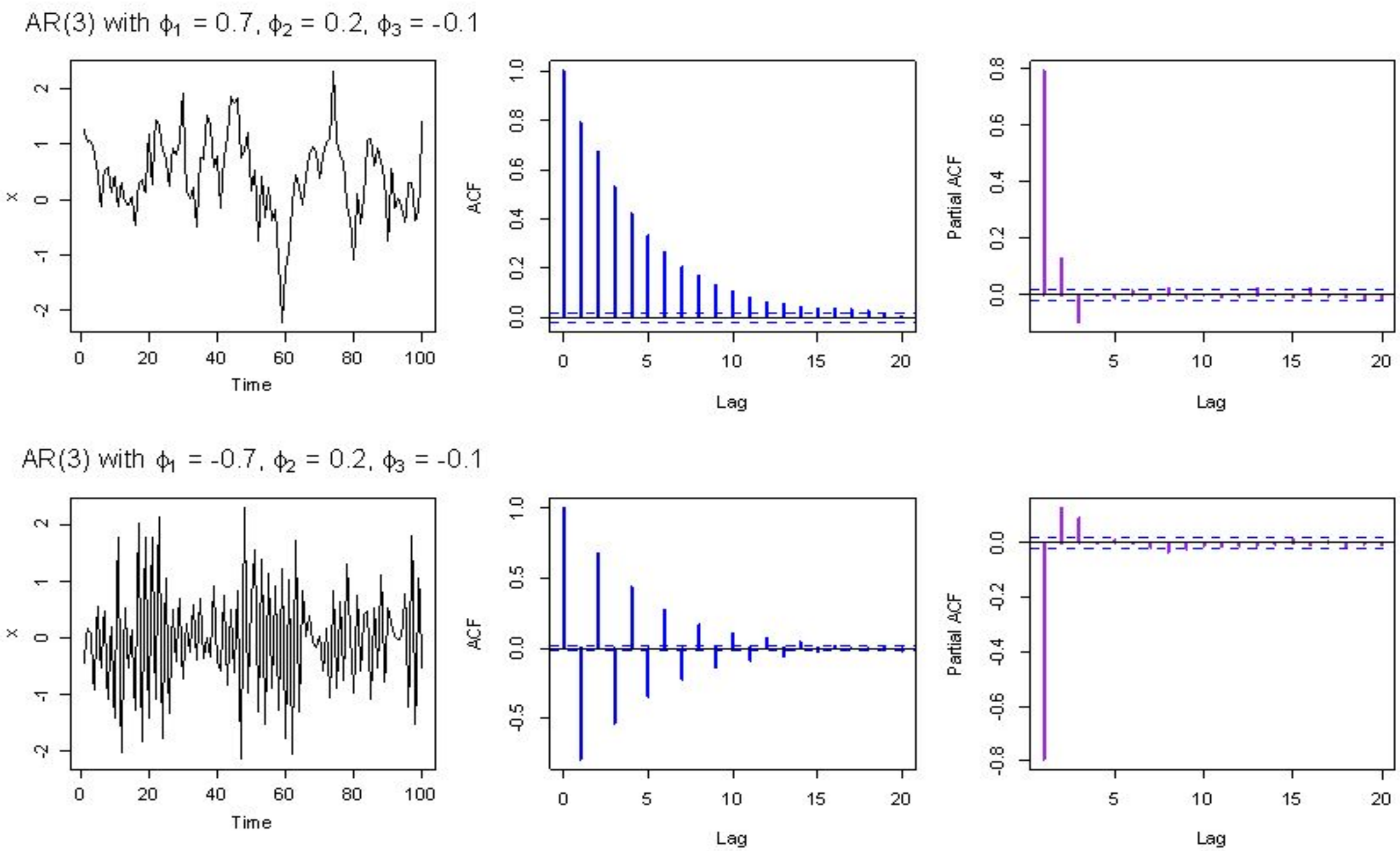
More complicated stationarity requirement of ϕ_i's holds for $p \geq 3$.

Theoretical ACF and PACF for AR(2)



- Characteristics:
  - ACF dies down
  - PACF cuts off for lag > 2

Theoretical ACF and PACF for AR(3)



- Characteristics:
  - ACF dies down.
  - PACF cuts off after lag *3*.

BUILDING MOVING AVERAGE(MA) MODELS

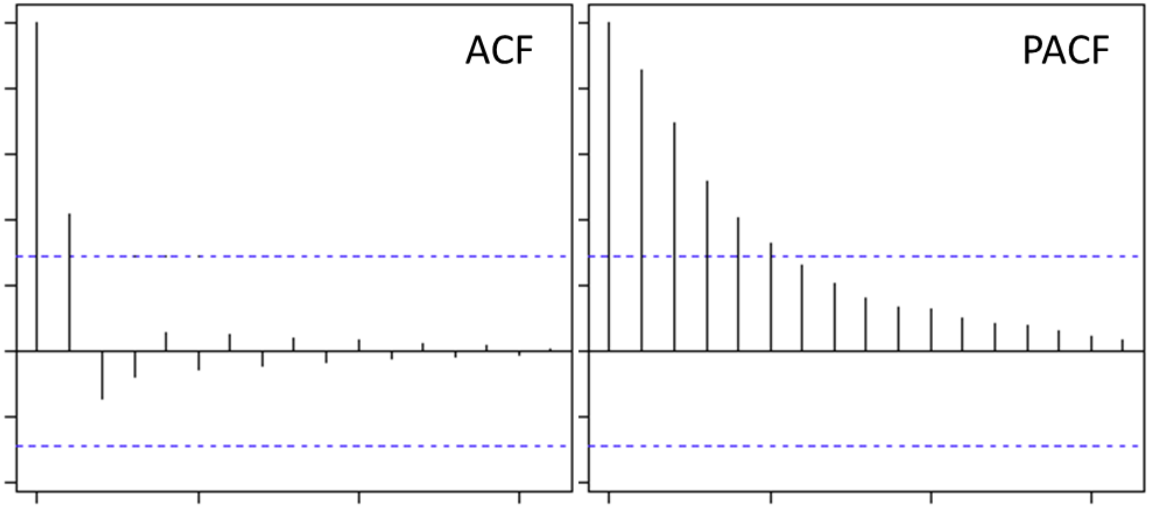
MA(1)

- MA(1) process:

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} \quad (where\ \varepsilon_t\ is\ white\ noise)$$

where $|\theta_1| < 1$, which is the stationarity requirement for an MA(1) process.

Theoretical ACF and PACF for MA(1)



- Characteristics:
  - ACF cuts off after lag 1
  - PACF dies down

- Note that there is only one significant autocorrelation at time lag 1.
- The partial autocorrelations decay exponentially, but because of random error components, they do not die out to zero as do the theoretical autocorrelation.

MA(q)

- MA(2) process

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} \qquad (\textit{where } \varepsilon_t \textit{ is white noise})$$

where $|\theta_2| < 1$, $\theta_2 + \theta_1 < 1$, and $\theta_2 - \theta_1 < 1$, which is the stationarity requirement for an MA(2) process.

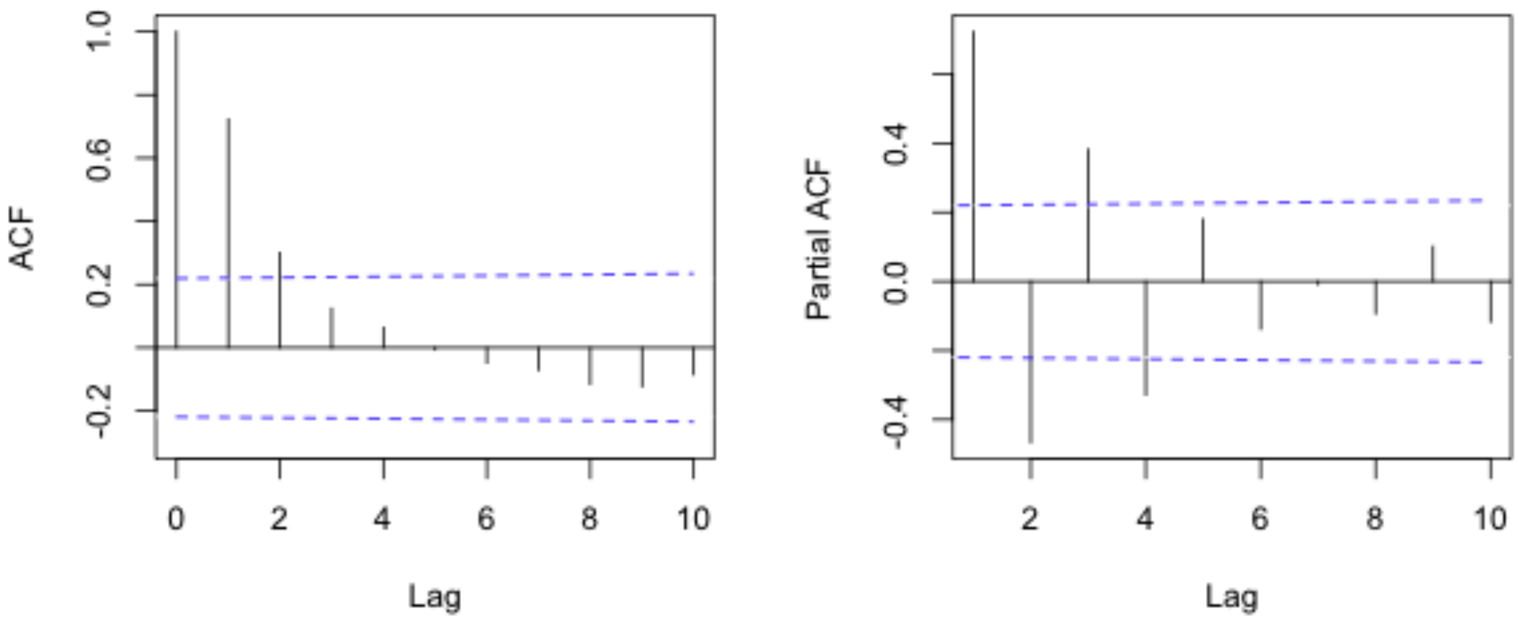
- MA(q) process

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \mathrm{L} \; - \theta_q \varepsilon_{t-q}$$

$$\textit{where } \varepsilon_t \textit{ is white noise}$$

More complicated stationarity requirement of θ_i's holds for $q \geq 3$.

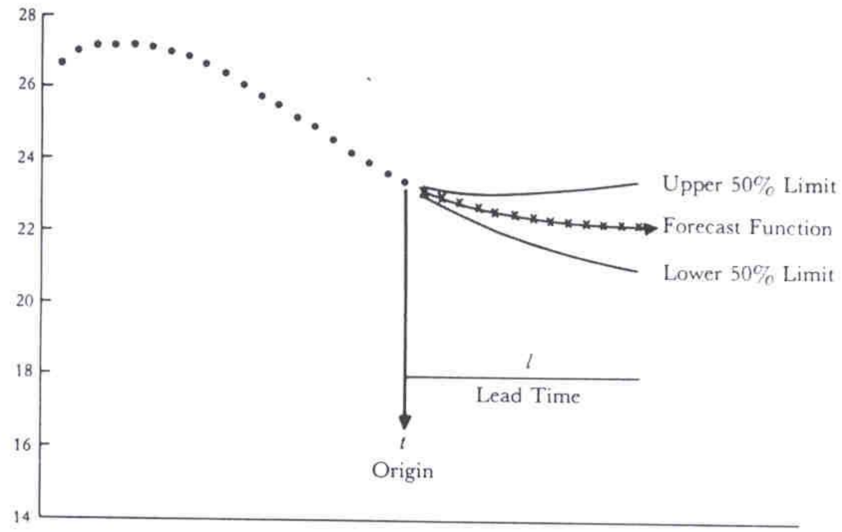
Theoretical ACF and PACF for MA(*2*)



- Characteristics:
  - ACF cuts off after lag *2*.
  - PACF dies down.

AR & MA are Integrated to Obtain ARIMA

- Combine AR(p) & MA(q) Models to Obtain ARMA (p,q) & ARIMA (p,d,q)
- **A**utoregressive **I**ntegrated **M**oving-**a**verage
- Can represent a wide range of time series
- A "**stochastic**" modeling approach that can be used to calculate the probability of a future value lying between two specified limits



ARIMA models (Cont.)

- In the 1960's Box and Jenkins recognized the importance of these models in the area of economic forecasting
- "Time series analysis - forecasting and control"
  - George E. P. Box   Gwilym M. Jenkins
  - 1st edition was published in 1976

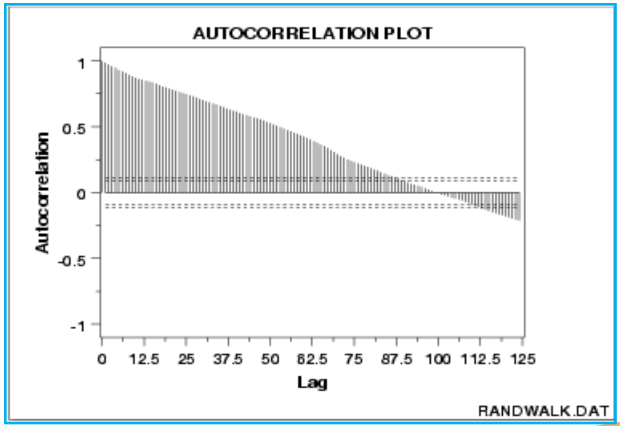
George E. P. Box

ARIMA models (Cont.)

- ARIMA models rely heavily on **autocorrelation** patterns in the data.

- ARIMA methodology of forecasting is different from most methods because it does not assume any particular pattern in the historical data of the series to be forecast.

- It uses an interactive approach of identifying a possible model from a general class of models. The chosen model is then checked against the historical data to see if it accurately describe the series.



Box-Jenkins(ARIMA) Methodology

- Identification ~Determine, given a sample of time series observations, what is the model of the [stationary] data.
- Estimation ~Estimate the parameters of the chosen model

$$AR(p): X_t = \alpha + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{t-p} + \varepsilon_t; \qquad \varepsilon_t \overset{i.i.d.}{\sim} N\left(\mu, \sigma^2\right)$$

$$MA(q): X_t = \alpha + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}; \qquad \varepsilon_t \overset{i.i.d.}{\sim} N\left(\mu, \sigma^2\right)$$

$$\begin{aligned} ARMA(p,q): X_t &= \alpha + \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \varepsilon_t \\ &\quad - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}; \end{aligned} \qquad \varepsilon_t \overset{i.i.d.}{\sim} N\left(\mu, \sigma^2\right)$$

Autoregressive Moving Average Models (ARMA)

- Autoregressive-moving average model of order *p* and *q* (ARMA(*p*,*q*))

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

  i.e., y_t depends on its *p* previous values and *q* previous random error terms

- Using backshift notation

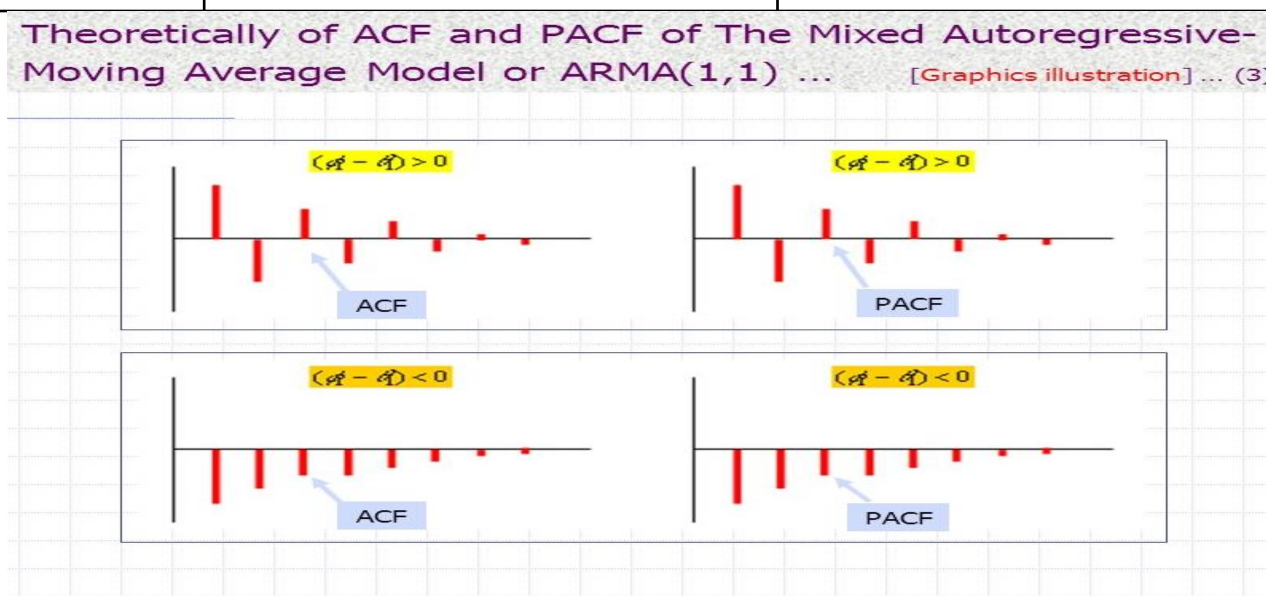
$$\phi_p(B) y_t = \phi_0 + \theta_q(B) \varepsilon_t$$

- When $q = 0$, the ARMA(p,0) model reduces to AR(p); when $p = 0$, the ARMA(0,q) model reduces to MA(q).

Summary of the Behaviors of ACF and PACF

Behaviors of ACF and PACF for general non-seasonal models

Process	ACF	PACF
AR(*p*)	Dies down.	Cuts off after lag *p*.
MA(*q*)	Cuts off after lag *q*.	Dies down.
ARMA(*p*,*q*)	Dies down.	Dies down.



Non-seasonal Autoregressive Integrated Moving Average models (ARIMA)

- ARMA models can only be used for stationary data. This class of models can be extended to non-stationary series by allowing differencing the data series. $\Rightarrow$ ARIMA models
- Backshift notation

$$\underbrace{\phi_p(B)}_{AR}\underbrace{(1-B)^d}_{I}\, y_t = \phi_0 + \underbrace{\theta_q(B)}_{MA}\varepsilon_t$$

and $\delta = \mu\phi_p(B)\Phi_P(B^L)$

e.g. ARIMA(1,1,1) $\quad (1-\phi_1 B)(1-B)y_t = \phi_0 + (1-\theta_1 B)\varepsilon_t$

- The general non-seasonal model: ARIMA(p,d,q)
  - AR: p = order of the autoregressive part
  - I: d = order of integration
  - MA: q = order of the moving average part

Mixtures ARIMA models

- If non-stationarity is added to a mixed ARMA model, then the general ARIMA (p, d, q) is obtained.
- The equation for the simplest ARIMA (1, 1, 1) is given below.

$$y_t = C + (1 + \phi_1) y_{t-1} - \phi_1 y_{t-2} + e_t - \theta_1 e_{t-1}$$

Mixtures ARIMA models

- The general ARIMA (p, d, q) model gives a tremendous variety of patterns in the ACF and PACF, so it is not practical to state rules for identifying general ARIMA models.
- In practice, it is seldom necessary to deal with values p, d, or q that are larger than 0, 1, or 2.
- It is remarkable that such a small range of values for p, d, or q can cover such a large range of practical forecasting situations.

Diagnostic Checking

Diagnostic Checking

- Often it is not straightforward to determine a single model that most adequately represents the data generating process. The suggested tests include

  (1) residual analysis,

  (2) model selection criteria.

Residual Analysis

- The residuals left over after fitting the model should be white noise.
  - ACF and PACF of the residuals show no significant autocorrelations or partial autocorrelations.
  - Residual autocorrelations as a group should be consistent with those produced by random errors.
    - Portmanteau test

White Noise Series

- White noise series $\{\varepsilon_t\}$
  - ε's are iid (independent and identically distributed) random variables with finite mean and variance.

$$(1)\ E(\varepsilon_t) = c \quad for\ all\ t.$$

$$(2)\ Var(\varepsilon_t) = b \quad for\ all\ t.$$

$$(3)\ Cov(\varepsilon_t, \varepsilon_{t+s}) = 0 \quad for\ all\ t,\ \ s \neq t.$$

Portmanteau Test

- Rather than study the autocorrelation values one at a time, an alternative approach is to consider a whole set of autocorrelation values all at one time, and test to see whether the set is significantly different from a zero set.
- Ljung-Box test

$$Q = n(n+2)\sum_{k=1}^{m}\frac{r_k^2(e)}{n-k} \sim \chi^2_{m-r}$$

where $r_k(e)$ = the residual autocorrelation at lag k

n = number of residuals

k = time lag

m = number of time lags to be tested

r = number of parameters estimated in the model

Residual Analysis

- If the portmanteau test is insignificant, the model is adequate.

- If the portmanteau test is significant, the model is inadequate. Then we need to go back and consider other ARIMA models.

- The pattern of significant spikes in the ACF and PACF of the residuals may suggest how the model can be improved.
  - Significant spikes at low lags suggest the nonseasonal AR or MA components of the model.

Model Selection Criteria

- Akaike Information Criterion (AIC)

  $$\text{AIC} = -2\ln(L) + 2k$$

- Schwartz Bayesian Criterion (SBC)

  $$\text{SBC} = -2\ln(L) + k\ln(n)$$

  where L = likelihood function

  k = number of parameters to be estimated,

  n = number of observations.

- Ideally, the AIC and SBC will be as small as possible.
- Usually the model with the smallest AIC and SBC will have residuals which resemble white noise.

Special Cases of ARIMA

• White noise	ARIMA(0,0,0)
• Random Walk	ARIMA(0,1,0) with no constant
• Random Walk with drift	ARIMA(0,1,0) with constant
• Autoregression	ARIMA(p,0,0)
• Moving Average	ARIMA(0,0,q)

Process Steps to Consider

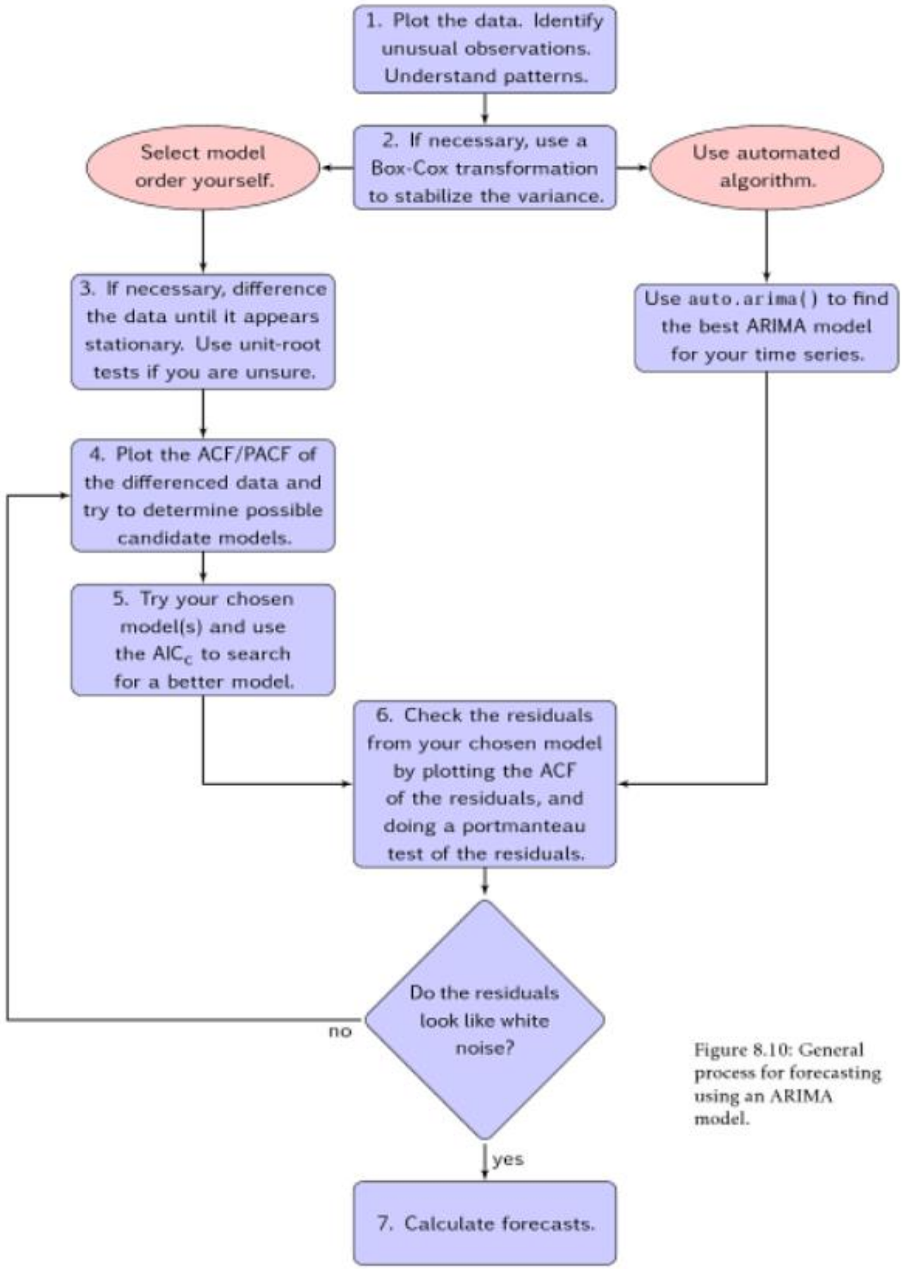


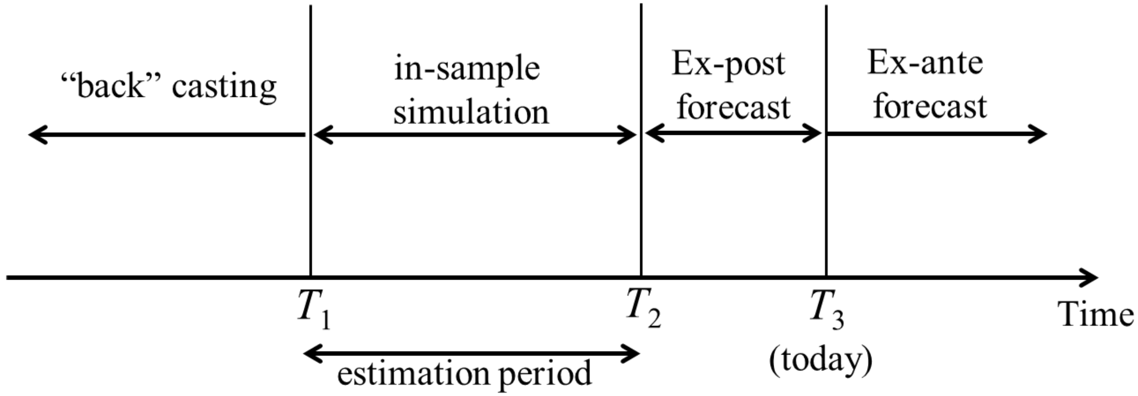
Figure 8.10: General process for forecasting using an ARIMA model.

Source : Forecasting Principles and practice – Rob J.H. & George A.

Out-of-Sample Forecasts – Applicable to All Techniques

1) Ex post forecast
   - Prediction for the period in which actual observations are available

2) Ex ante forecast
   - Prediction for the period in which actual observations are not available.



Summary

- Concept of time series data and its components.
- Concept of autocorrelations
- Uncovering patterns from time series data
- Key techniques
  - ✓ Time Series Regression
  - ✓ Dummy Variable Regression to Predict Seasonality
  - ✓ Decomposition methods
  - ✓ Smoothing methods
  - ✓ ARIMA

Data Issues

- How much data is required for forecasting?
  - Enough data to identify seasonal patterns (normally this should be at least 2 years)
  - For regression methods there should be at least 4 times as many data points as there are independent variables
  - Too much data may also be bad as old data tends to become “stale” or “irrelevant”
- Bad data should be marked and excluded

User Intervention

- It is necessary for users to “influence” forecasts to correct for
  - Changes in market conditions
  - Competitor actions
  - Political problems
  - Special events (conventions, strikes, etc.)
- Many forecasting systems allow users to add a number to a forecast or multiply it by a percentage
- Users may use existing products to “sponsor” new ones.
- For best forecasts statistical techniques should be combined with domain knowledge

Accuracy Measures

Time	Actual	Monthly Forecasts	Bias	Absolute	Squared	AbsolutePercentageE
38749	977.5	991.882	-14.3816458	14.382	206.83	0.0147127
38777	963.5	971.426	-7.925543963	7.9255	62.814	0.0082258
38808	1032.5	974.965	57.53476989	57.535	3310.2	0.0557237
38838	1098.75	1042.8	55.94696936	55.947	3130.1	0.0509187
38869	1244	1092.49	151.5117438	151.51	22956	0.121794
38899	1345	1240.23	104.7717919	104.77	10977	0.0778972
38930	1493.13	1320.87	172.2566844	172.26	29672	0.1153666
38961	1383.5	1486.74	-103.2430715	103.24	10659	0.0746246
38991	1098.75	1360.65	-261.895459	261.9	68589	0.2383576
39022	1093.13	1133.67	-40.54655036	40.547	1644	0.0370923
39052	1157.5	1172.71	-15.21146997	15.211	231.39	0.0131417
39083	1146.25	1177.87	-31.61601431	31.616	999.57	0.0275821
Average			5.600183701	84.737	12703	0.0696198
					112.71	

REFERENCE BOOK

- Forecasting, Time Series, and Regression by Bruce L. Bowerman, | Richard T O'Connell & Anne B Koehler, 4th Edition