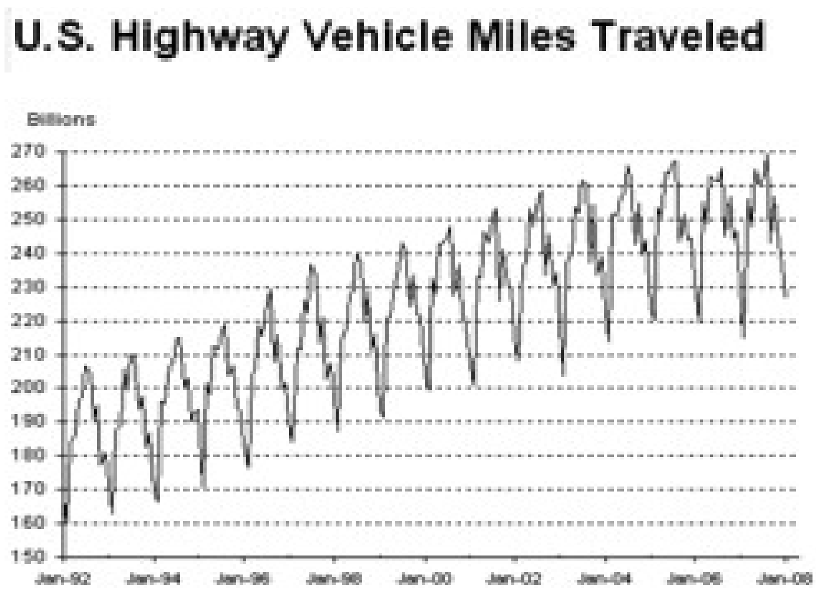
Forecasting Time Series

SLIDE NUMBER: 1



SLIDE NUMBER: 2

My Introduction

Name: **Bharani Kumar**

Education: **IIT Hyderabad ,Indian School of Business**

Professional certifications:

PMP → Project Management Professional

PMI-ACP → Agile Certified Practitioner

PMI-RMP→ Risk Management Professional

CSM → Certified Scrum Master

LSSGB → Lean Six Sigma Green Belt

LSSBB → Lean Six Sigma Black Belt

SSMBB → Six Sigma Master Black Belt

ITIL → Information Technology Infrastructure Library

Agile PM→ Dynamic System Development Methodology Atern

- **DATA SCIENTIST**
- **RESEARCH in ANALYTICS, DEEP LEARNING & IOT**

Deloitte

Driven using US policies

Infosys

Driven using Indian policies under Large enterprises

ITC Infotech

Driven using Indian policies SME

HSBC

Driven using UK policies

SLIDE -4

AGENDA

Why Forecasting

Learn about the various examples of forecasting

Forecasting Strategy
Learn about decomposing, forecasting & combining

EDA & Graphical Representation

Learn about exploratory data analysis, scatter plot, time plot, lag plot, ACF plot

Forecasting components
Learn about Level, Trend, Seasonal, Cyclical, Random components

Forecasting Models & Errors
Learn about various forecasting models to be discussed & the various error measures

SLIDE -5

Why Forecasting?

- Why Forecasting Why forecast, when you would know the outcome eventually?
- *Early* knowledge is the key, even if that knowledge is imperfect
  - For setting production schedules, one needs to forecast sales
  - For staffing of call centers, a company needs to forecast the demand for service
  - For dealing with epidemic emergencies, nations should forecast the various flu

RICHTER SCALE of earthquake energy:

Each level is 10 times stronger than the previous level

	Description	Occurrence	In Population	Movement
1	SMALL	DAILY	every minute	small
2	SMALL	DAILY	every hour	small
3	SMALL	DAILY	every day	small
4	SMALL	DAILY	every week	moderate sudden
5	MODERATE	MONTHLY	every 10 years	strong sudden
6	MODERATE	MONTHLY	every 30 years	strong sudden
7	MAJOR	MONTHLY	every 50 years.	severe sudden
8	GREAT	YEARLY	every 100 years	very severe
9	GREAT	YEARLY	every 300 years	very severe
10	SUPER	RARELY	every 1000 years	extreme

- Earthquake – Over forecasting & under forecasting. E.g. Italian scientists imprisoned
- The Richter magnitude scale (also Richter scale) assigns a magnitude number to quantify the energy released by an earthquake. The Richter scale, developed in the 1930s, is a base-10 logarithmic scale, which defines magnitude as the logarithm of the ratio of the amplitude of the seismic waves to an arbitrary, minor amplitude.
- As measured with a seismometer, an earthquake that registers 5.0 on the Richter scale has a shaking amplitude 10 times that of an earthquake that registered 4.0, and thus corresponds to a release of energy 31.6 times that released by the lesser earthquake.[1] The Richter scale was succeeded in the 1970s by the moment magnitude scale. This is now the scale used by the United States Geological Survey to estimate magnitudes for all modern large earthquakes.[2]

SLIDE - 6

Types of forecast

- Short Term or Long Term
- Micro Scale or Macro Scale
- Data or Judgment
- Qualitative or Quantitative

Forecasting Classification

- Point Forecast

- Density Forecast
- Interval Forecast

SLIDE-7

Who generates Forecast?

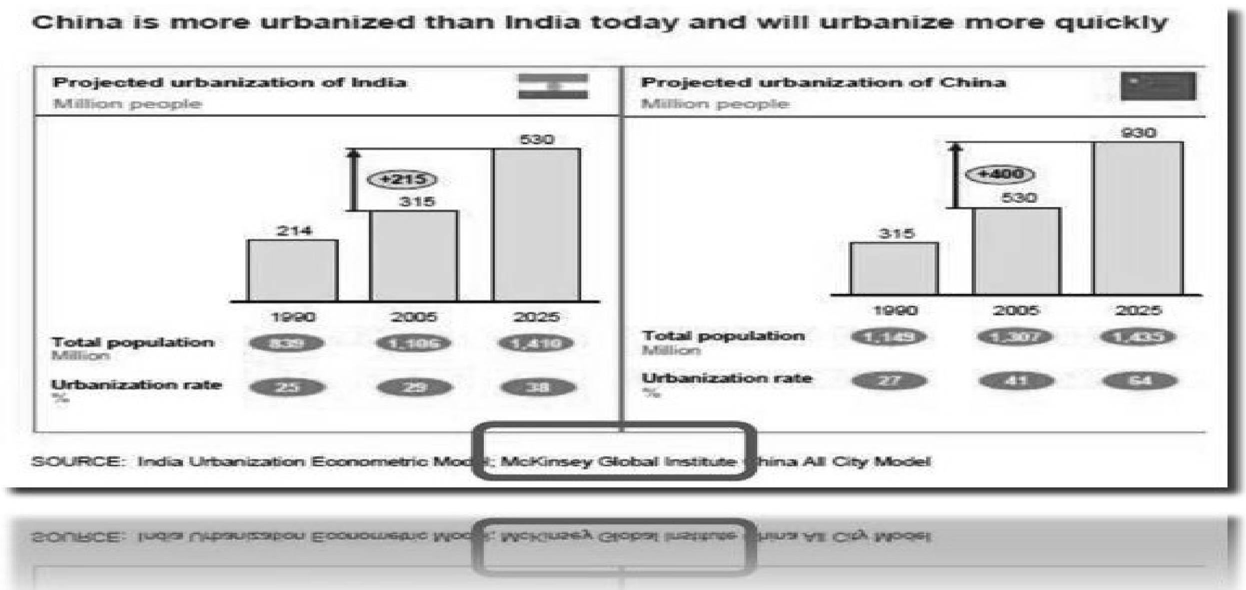
- Governments
- Corporates
- Private Sectors
- Consulting Firms
- NGO'S
- Thomas Watson, founding father of IBM, has forecasted that there will be world market for 5 computers in 1943. At that time, it was right not now. Computer occupies size of a building, only used by govt. for defense research, super costly, AC
- A famous person has also quoted that "I cannot foresee any one to have a computer at home. This statement was done in 1977. DEC (Digital Equipment Corporation) president did this. Forecasting can be difficult even for super intelligent people.
- Consulting firms – forecast and sell reports. For e.g., Gartner, Forrester,
- Govt – In transportation – to predict how many people travel by a train via a route.
- Retail – Inventory for right products based on fashion. Otherwise, we may need to provide additional discounts. For e.g., 80% discount and then you go bankrupt

– Marketing – Sales

– Accounting – What will you forecast

– Bank – Nonperforming assets (NPA) forecasting

SLIDE – 8



SLIDE-9

Time series vs Cross-sectional data

- Cross sectional data
  - Cross sectional – Single point in time, you get all information (Stores in Chennai, Mumbai, Hyderabad, and Delhi) Got sales data on Jan 1st. Sales explained using Location, Advertisement, etc.
- Time series Data
  - Times series – Months & Sales. If we have sales for Chennai for last 1 year then we can forecast.

SLIDE-10

Month	Ridership ('000)
Jan-91	1709
Feb-91	1621
Mar-91	1973
Apr-91	1812
May-91	1975
Jun-91	1862
Jul-91	1940
Aug-91	2013
Sep-91	1596
Oct-91	1725
Nov-91	1676
Dec-91	1814
Jan-92	1615
Feb-92	1557
Mar-92	1891
Apr-92	1956
May-92	1885
Jun-92	1623

Dataset for further discussion

- Monthly Ridership of passengers from Jan 1991 to March 2004

t = 1, 2, 3,…....= time period index

Y_t = value of the series at time period t

Y_{t+k} = forecast for time period t+k, given data until time t

e_t = forecast error for period t

Monthly data t = 1 for Jan, 2 for Feb….

Yearly data t = 1 for 1991, 2 for 1992.......

Ridership measured in 1000s

K = 7 for forecasting weekly data

Error can be positive or negative

SLIDE-11

Forecasting Strategy

- Define Goal
- Data Collection
- Explore & Visualize Series
- Pre-Process Data
- Partition Series
- Apply Forecasting Method(s)
- Evaluate & Compare Performance
- Implement Forecasts / System

Already available data is secondary data

Survey etc. is primary data

Pre-process = error free + good Quality because garbage in garbage out

Partition series = Training & Validation data

AR, ES, MA, ARMA, ARIMA

SLIDE-12

Forecasting Strategy – Step 1

#1 Is the goal descriptive or predictive?

- Descriptive = Time Series Analysis
- Predictive = Time Series Forecasting

- **#2 What is the forecast horizon?**

    - How far into the future? k in Y_{t+k}
    - Rolling forward or at single time point?
- **#3 How will the forecast be used?**
    - Who are the stakeholders?
    - Numerical or event forecast?
    - Cost of over-prediction & under-prediction
- **#4 Forecasting expertise & automation**
    - In-house forecasting or consultants?
    - How many series? How often?
    - Data & software availability

- # Define goal:

- 1^{st} – Retrospective/Descriptive -> Rainy season train travel is less & during summer vacation train travel might be more. Looking at past data & establish relationship between variables
- 2^{nd} – Prospective / Predictive goal -> Forecasting the price of ridership
- Revenue management – Empty seat in airline for a flight on one day cannot be sold the next day. How can we sell yesterday's seat today.
- However, a bag of rice if not sold today can be sold tomorrow. Ola cabs etc.
- 2. a. We can forecast for 3 months for mobile phone sales. Because we get new phones every 2-3 months. Cycle time is compressed. In 2001, whole telecom industry went for a melt

down. Lucent technologies, vendors who were financed by company. Stock market crashed & vendors told you can take the equipment because we don't have money to pay. Who will buy old equipment. $80 stock to $.59 cents.

- 2.b. Predict sales for June, then we want to update the data & predict for July. It will cost more money because more data, more processing, more consultants.
- If you are Flipkart then we forecast the warehouse for a delivery period.
- 3. Who are the stakeholders? Will marketing team use or finance team use the forecasted results. Numerical or event forecast (sales) or Whether there will be strike or not, earthquake or not, to invest or not to invest, to open a plant or not to plant
- Italian Scientist jailed for under prediction.
- 4. Startup will have consultants & large companies with huge revenues / profits can have in-house team
- Forecasting expertise & domain expertise is needed

SLIDE- 13

Forecasting Strategy – Step 2

#1 Data Quality

  - Typically small sample, so need good quality
  - Data same as series to be forecasted

- **#2 Temporal Frequency**
  - Should we use real-time ticket collection data?

- Balance between signal & noise
- Aggregation / Disaggregation

- **#3 Series Granularity?**
  - Coverage of the data – Geographical, population, time,..
  - Should be aligned with goal
- **#4. Domain expertise**
  - Necessary information source
  - Affects modeling process from start to end
  - Level of communication/coordination between forecasters & domain experts
- No missing data; incorrect data due to data entry error, transformation error, measurement error

Data series same – forecast sales of coke for next 1 year then we need data for last 3 years of Pepsi. If we need to forecast for a new product then we need to look at qualitative forecast techniques such as Delphi, Judgmental techniques, Benchmarking,

2. If goal is to predict daily sales for next 3 days then I have sales for every month (Ecommerce Company) 30* 24 hours. I will aggregate the data to daily and then use it to forecast. If I want to predict every hour then hourly data can be considered.

If putting money in ATM, in weekend people might draw more money, depends on location etc. Daily amount aggregation may be used to forecast or may be for weekends we might disaggregate and then forecast

3. Zero/low counts vs mixed populations

Particular routes. Particular populations (Senior citizen, children) Particular days (weekends), Rush hours ticket price is more in US for Amtrak.

4. Domain expertise – The most important aspect. In ordinary business process, we use simple techniques because communication the model to business people might be extremely challenging

SLIDE-14

Forecasting Strategy Step3 (Explore Series)

<u>Additive:</u>

Y_t = Level + Trend + Seasonality + Noise

<u>Multiplicative:</u>

Y_t = Level x Trend x Seasonality x Noise

SYSTEMATIC PART

- Level
- Trend
- Seasonal Patterns

NON-SYSTEMATIC PART

- Noise

Time series is divided into:

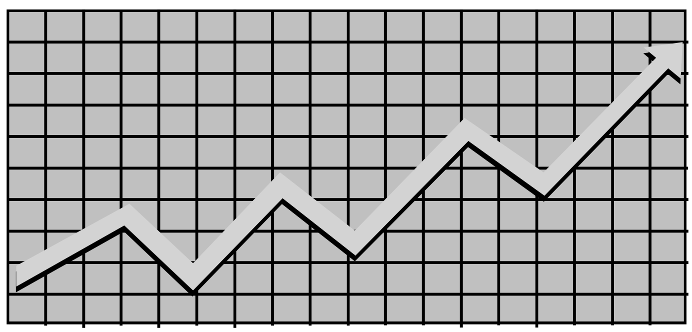
Level = simple average

Noise is called as God's particle by some people

SLIDE-15

Trend Component

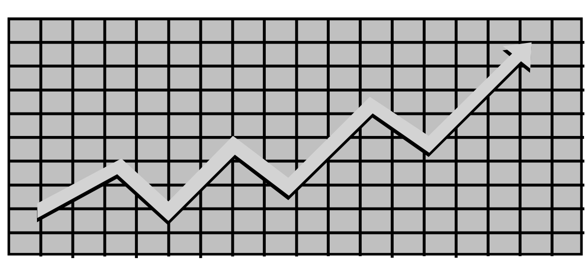
- Persistent, overall upward or downward pattern
- Due to population, technology etc.
- Overall Upward or Downward Movement
- Several years duration



SLIDE-16

Seasonal Component

- Regular pattern of up & down fluctuations
- Due to weather, customs etc.
- Occurs within one year
- Example: Passenger traffic during 24 hours



Efficient running of a shop; keeping inventory ready for the boom season; employing more number of employee for the boom period.

Seasonal variation in vegetable price series is less prominent for developed country as compared to developing countries!

Infrastructure---cold-storage, food processing industry.

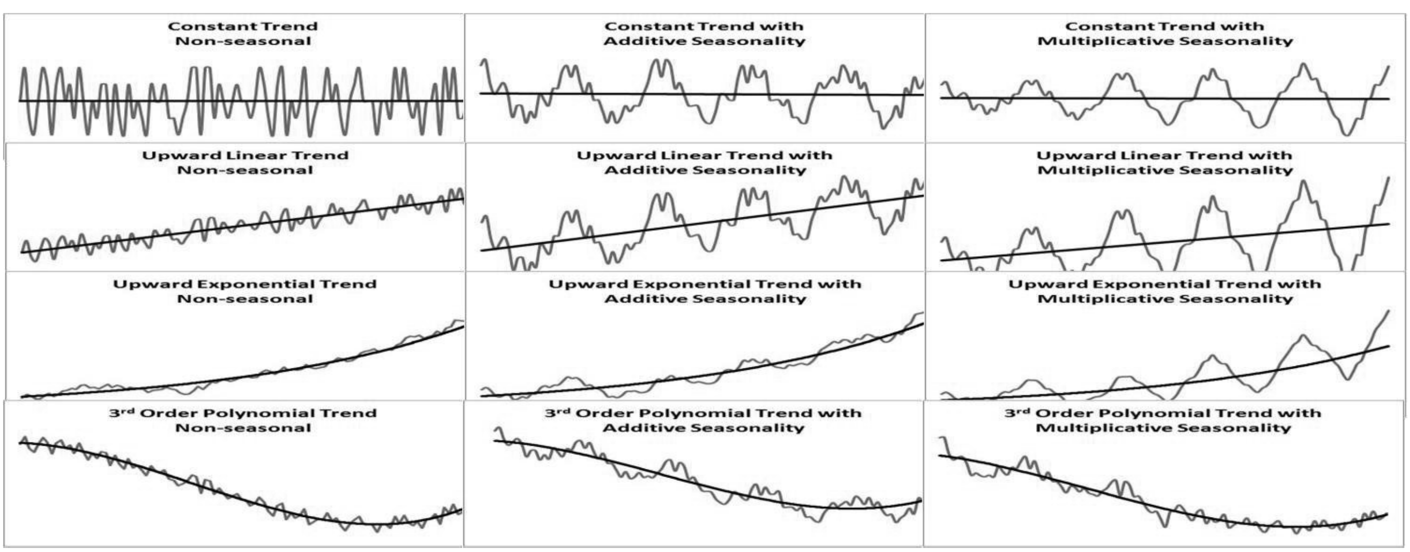
SLIDE- 17

Irregular/Random/Noise Component

- Erratic, unsystematic, 'residual' fluctuations
- Due to random variation or unforeseen events
  - Union strike
  - War
- Short duration & nonrepeating

SLIDE-18

Time Series Components

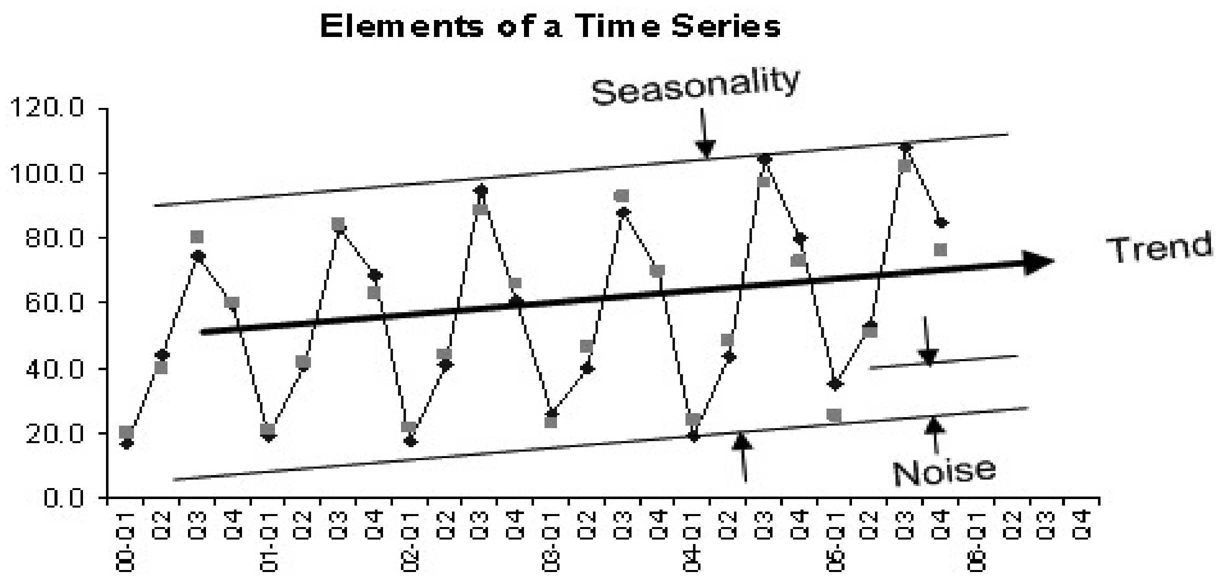


In Seasonality, we will have Additive & Multiplicative seasonality

1^{st} column = Trend

SLIDE-19

Quarterly Sales of Ice-cream



SLIDE-20

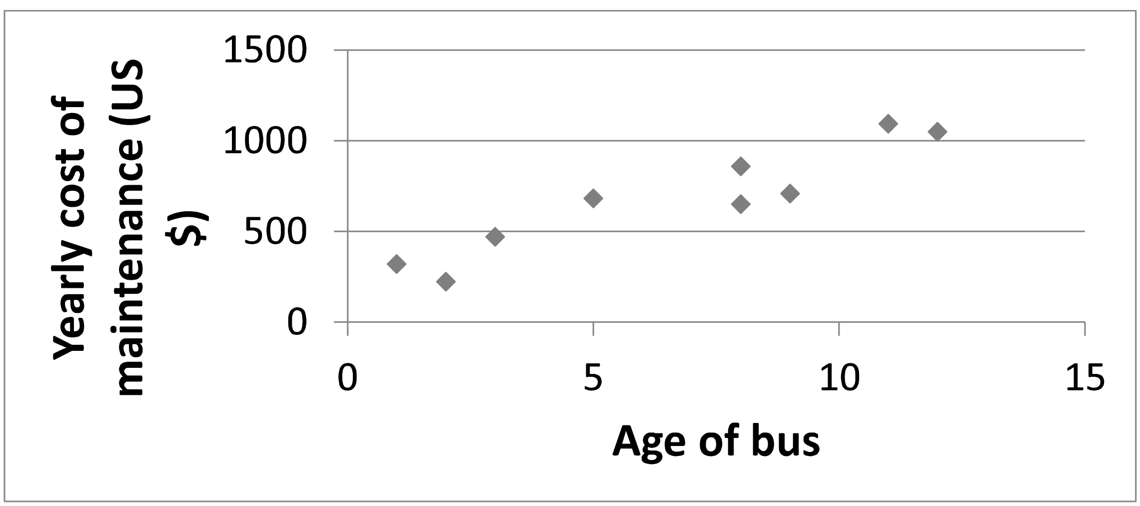
Scatter Diagram

- Plots one variable against another
- One of the simplest tools for visualization
- Example: Maintenance cost and Age for nine buses (Spokane Transit)
- This is an example of cross-sectional data (observations collected in a single point of time)

Cost	Age
859	8
682	5
471	3
708	9
1094	11
224	2
320	1
651	8
1049	12

SLIDE-21

SCATTER PLOT



Scatter plot of Age of bus Vs maintenance cost(yearly)

SLIDE-22

Observations:

- Older buses have higher cost of maintenance
- There is some variation (case to case)
- The rise in cost is about $ 80 per year of age
- It may be possible to use 'age' to forecast maintenance cost
- Forecast would not be a 'certain' prediction – there would be some error

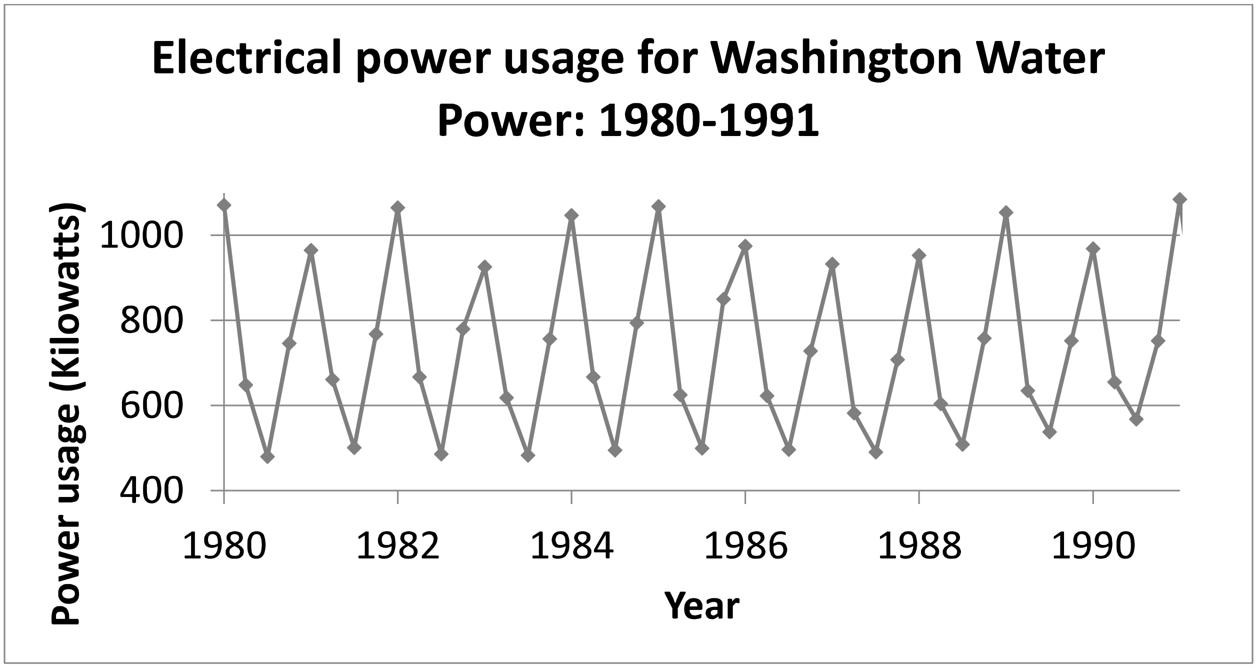
SLIDE-23

Time Plot

- Plots a variable against time index
- Appropriate for visualizing serially collected data (time series)
- Brings out many useful aspects of the structure of the data
- Example: Electrical usage for Washington Water Power (Quarterly data from 1980 to 1991)

SLIDE - 24

Electrical power usage for Washington Water Power: 1980-1991:



SLIDE - 25

Observations

- There is a cyclic trend
- Maximum demand in first quarter; minimum in third quarter
- There may also be a slowly increasing trend (to be examined)
- Any reasonable forecast should have cyclic fluctuations
- Trend (if any) need to be utilized for forecasting
- Forecast would not be exact – there would be some error

SLIDE-26

Lag plot:

- Plots a variable against its own lagged sample
- Brings out possible association between successive samples
- Example: Monthly sale of VCRs by a music store in a year

- Y_t = Number of VCRs sold in time period *t*
- Y_{t-k} = Number of VCRs sold in time period *t – k*

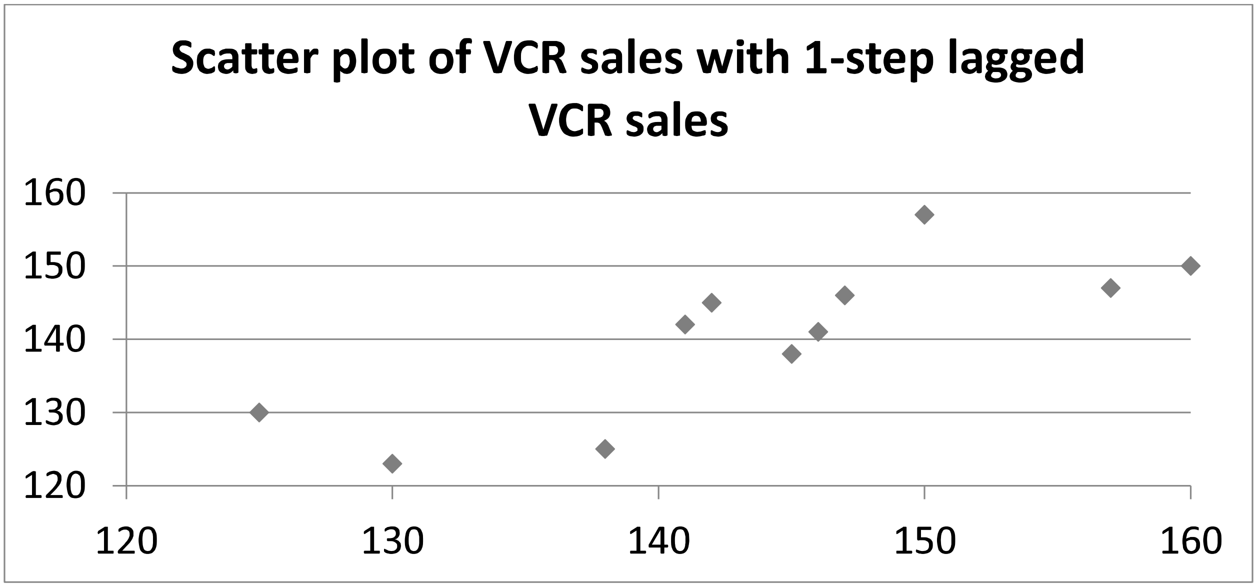
SLIDE-27

Example of lagged variables Number of VCR’s sold in month

Time	Original	Lagged one step	Lagged two steps
1	123		
2	130	123	
3	125	130	123
4	138	125	130
5	145	138	125
6	142	145	138
7	141	142	145
8	146	141	142
9	147	146	141
10	157	147	146
11	150	157	147
12	160	150	157

SLIDE - 28

Lag plot (k = 1)



SLIDE - 29

Observations:

- There is a reasonable degree of association between the original variable and the lagged one
- Value of lagged variable is known beforehand, so it is useful for prediction
- Association between original and lagged variable may be *quantified* through a correlation

SLIDE -30

Autocorrelation:

- Correlation between a variable and its lagged version (one time-step or more)

$r_k = (\sum_{t=k+1}^{n}(Y_t - \overline{Y})(Y_{t-k} - \overline{Y}))/(\sum_{t=1}^{n}(Y_t - \overline{Y})^2)$

k=0,1,2,3…

Y_t= Observation in time period t

Y_{t-k}= Observation in time period $t - k$

$\overline{Y}$= Mean of the values of the series

r_k= Autocorrelation coefficient for k-step lag

SLIDE-31

Standard error of r_k:

- **The standard error is**

  The standard error of the mean estimates the variability between samples whereas the standard deviation measures the variability within a single sample.

  SE(r_k) = sqrt((1+2$\sum_{i=1}^{k=1} r_i^2$)/(n))

- Increases progressively with k, but eventually reaches a maximum value
- If the 'true' autocorrelation is 0, then the estimate r_k should be in the interval (– 2SE(r_k), 2SE(r_k)) 95% of the time
- Sometimes SE(r_k) is approximated by sqrt(1/n)

SLIDE-32

Correlogram or ACF plot:

– Plots the ACF or Autocorrelation function (r_k) against the lag (k)

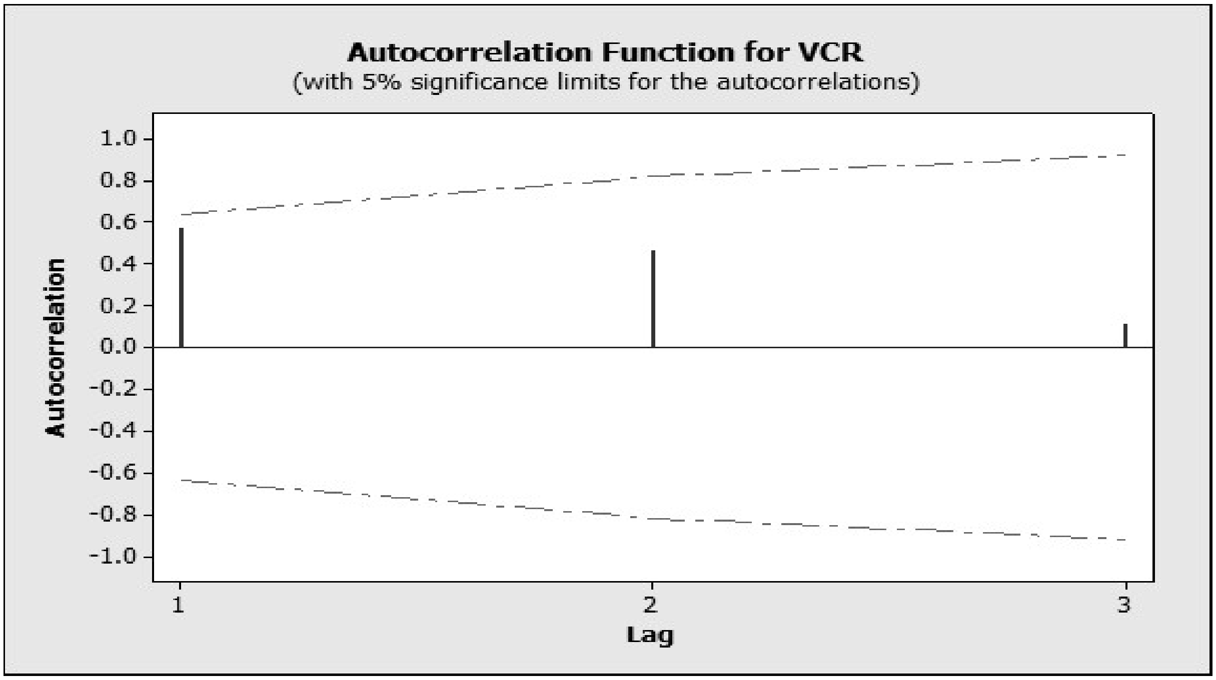
– Plus-and-minus two-standard errors are displayed as limits to be exceeded for statistical significance

– Reveals lagged variables that can be potentially useful for forecasting

SLIDE-33

Correlogram for VCR data:

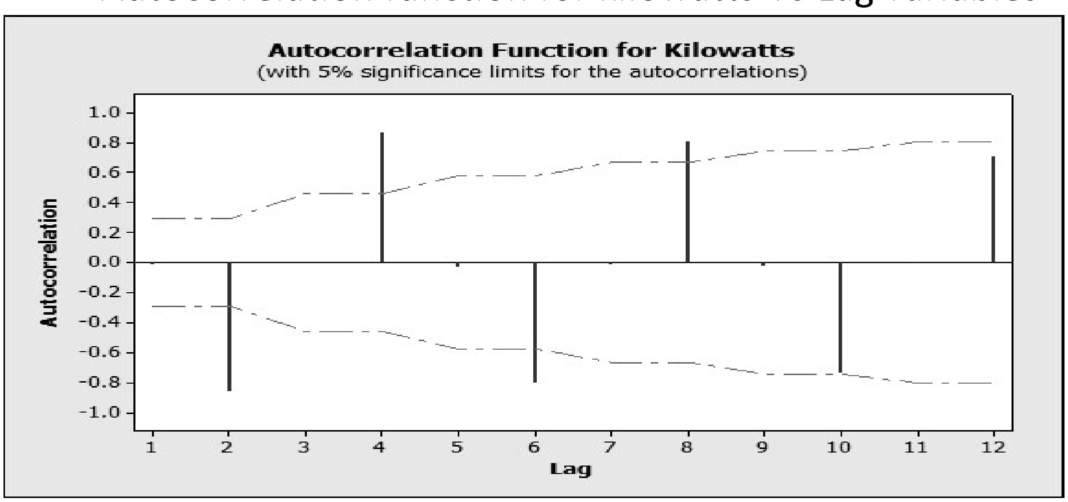
- Autocorrelation function for VCR Vs Lag variables



SLIDE-34

ACF plot for electricity usage data:

- Autocorrelation function for kilowatts Vs Lag variables



SLIDE-35

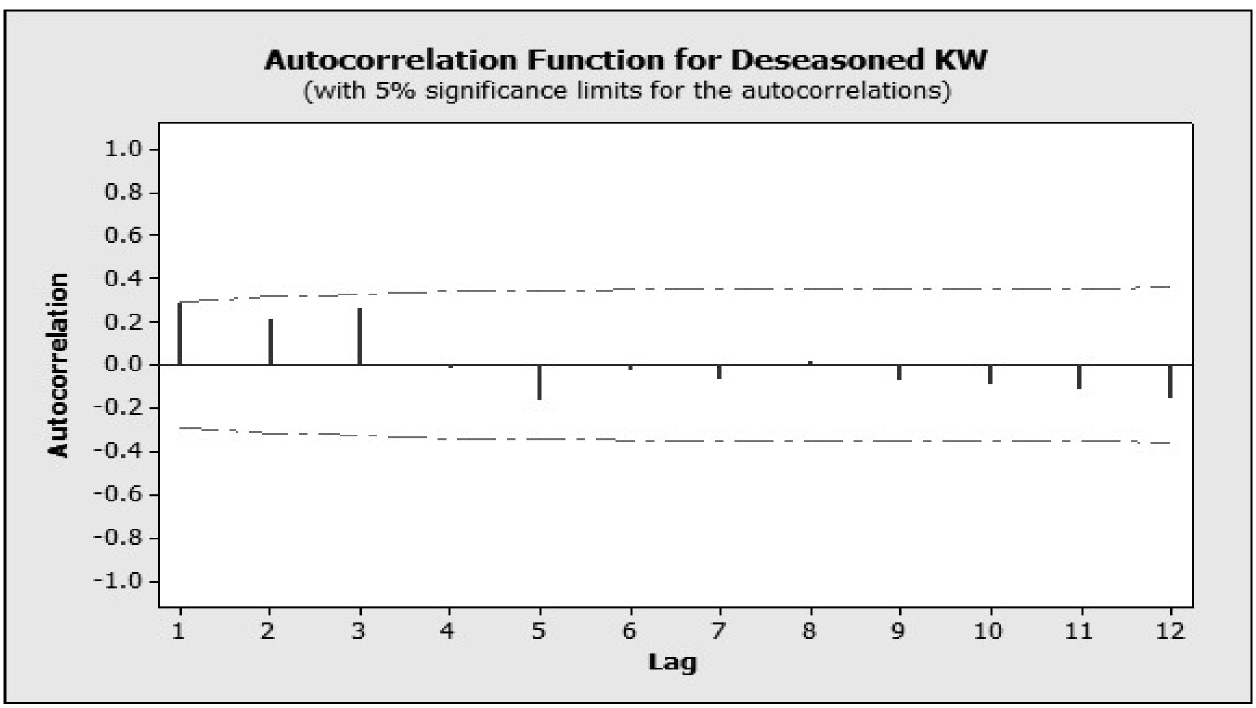
Observations:

- Every alternate sample is large, many of them statistically significant also
- ACFs at lags 4, 8, 12, etc are positive
- ACF at lags 2,6,10 etc are negative
- All these pick up the seasonal aspect of the data
- The data may be re-examined after ‘removing’ seasonality

SLIDE-36

ACF of de-seasoned KW data:

- Autocorrelation Function for DE seasoned KW Vs Lag variables.



SLIDE-37

Observations:

- De-seasoned series has small ACFs
- This part of the data has little forecasting value

SLIDE-38

Typical questions in exploratory analysis:

- Is there a TREND?
- Is there a SEASONALITY?
- Are the data RANDOM?

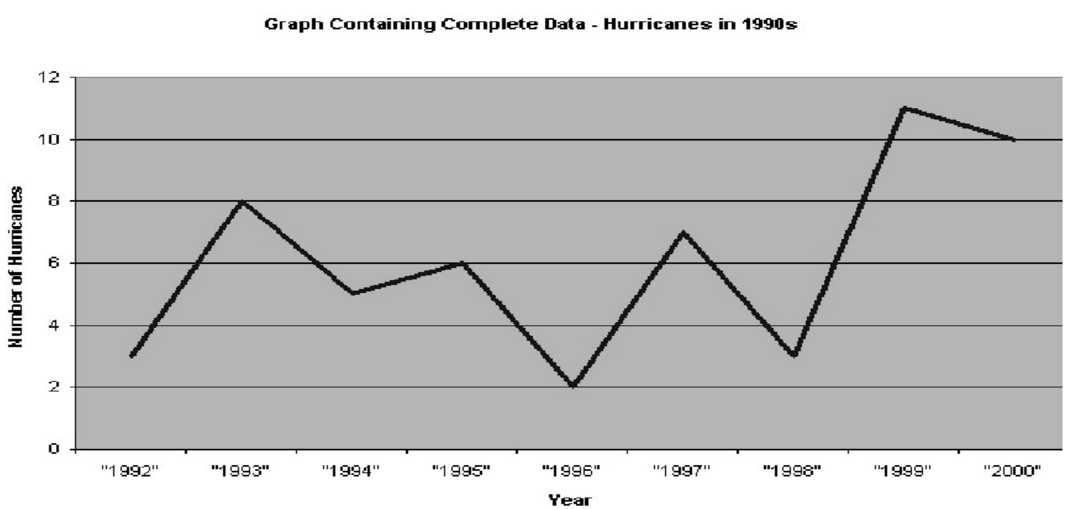
All the plots contains information regarding these questions

SLIDE-39

Time series plots:

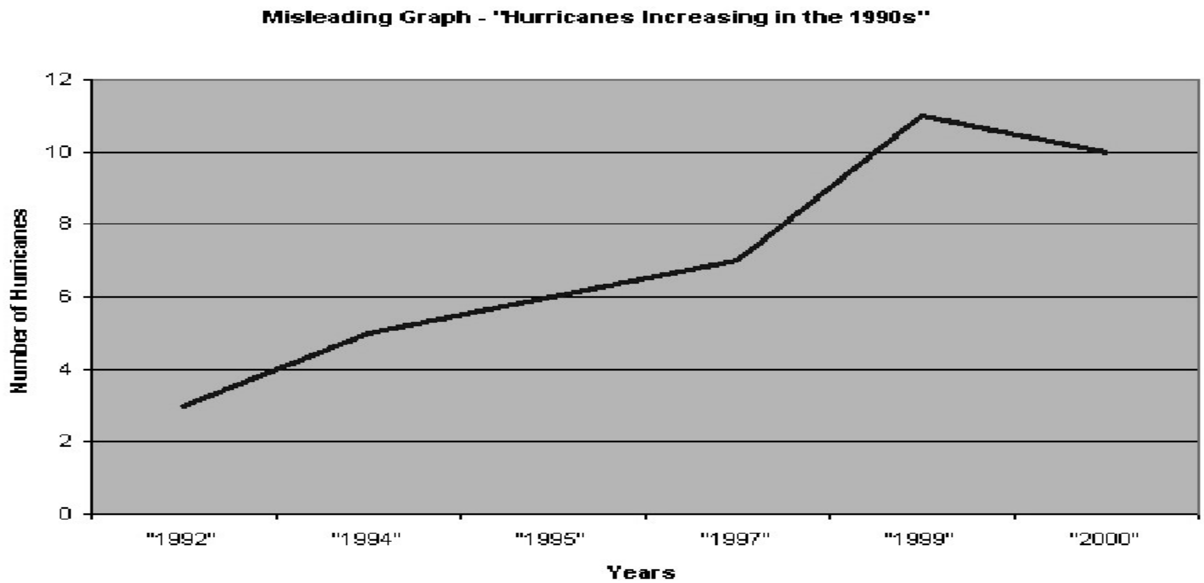
Graph containing complete data- Hurricane in 1990's

- Number of Hurricanes Vs Year



SLIDE- 40

Effect of omission of data on the Time series plot

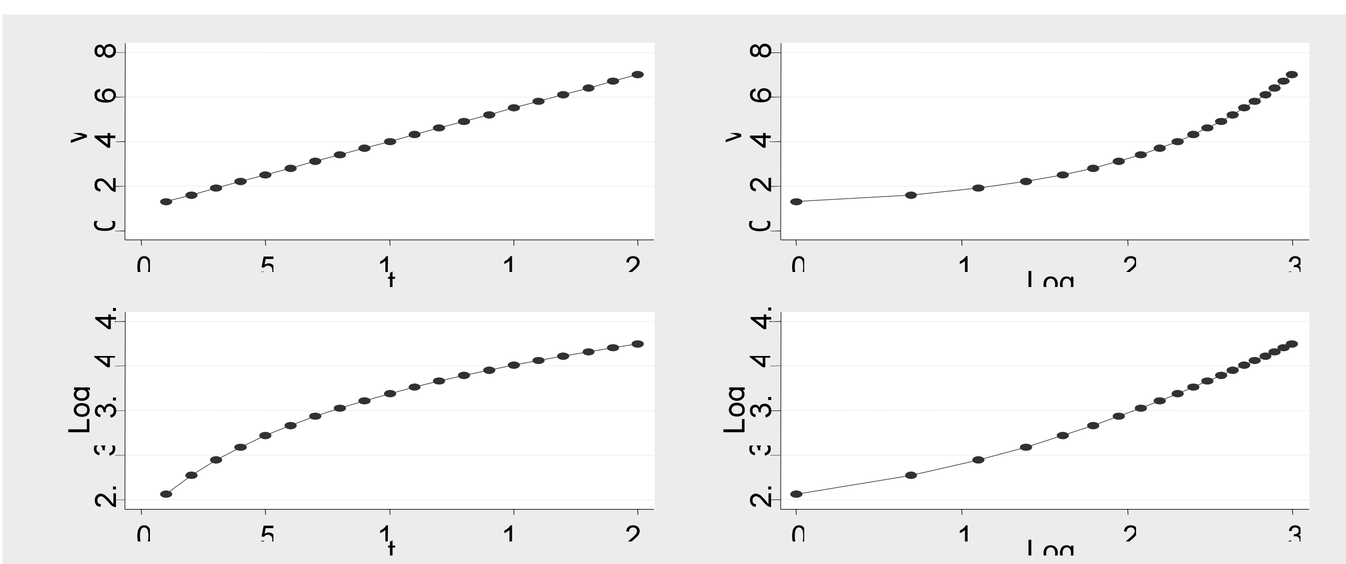


Misleading graph- Hurricanes increasing in 1990's

- Number of Hurricanes Vs year

SLIDE-42

Confusing kind of trend due to other type of scaling



SLIDE-43

Few points on Plots

- Plot helps us to summarize & reveal patterns in data
- Graphics help us to identify anomalies in data
- Plot helps us to present a huge amount of data in small space & makes huge data set coherent
- To get all the advantages of plot, the "Aspect Ratio" of plot is very crucial
- The ratio of Height to Width of a plot is called the ***ASPECT RATIO***

SLIDE-44

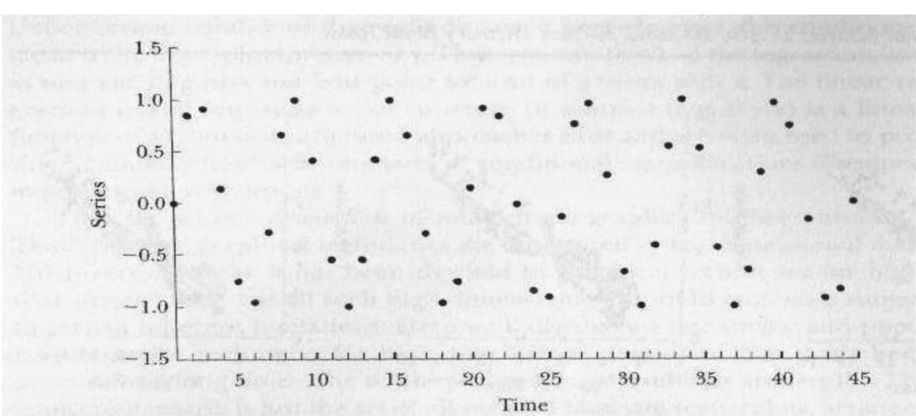
Aspect Ratio:

- Generally aspect ratio should be around 0.618
- However, for long time series data aspect ratio should be around 0.25. To understand the impact of aspect ratio see the two plots in the next two slides

SLIDE-45

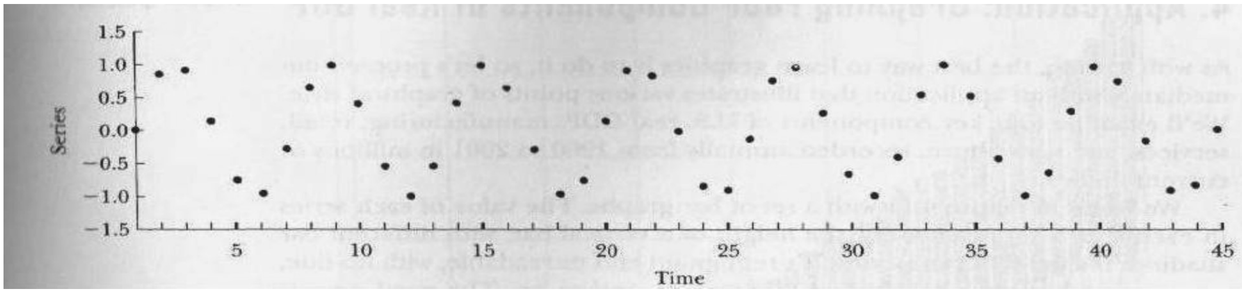
Aspect ratio:

Time series plot:



-Series Vs Time

SLIDE-46



SLIDE-47

Preliminaries for Step 3 of 8-Step forecasting strategy

- Should we use all historical data for forecasting
- Solution = DATA PARTIONING

Training Data - Fit the model only to TRAINING period

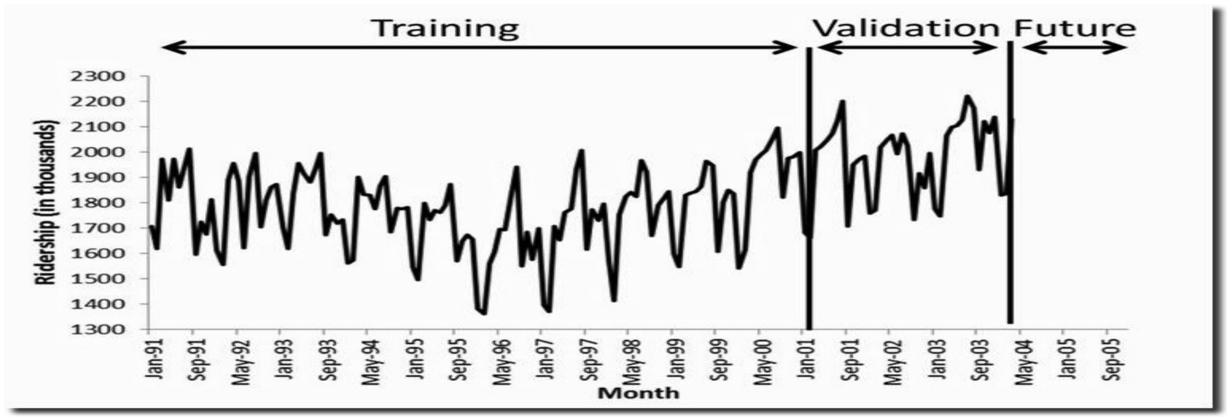
Validation Data- Assess performance on VALIDATION period

. How to choose validation period data. If you want to forecast for the next 12 months then validation data should have at least 12 months.

SLIDE-48

Partitioning:

- Training and Validation and future



Ridership (in thousands) Vs Month

- Deploy model by joining Training + Validation to forecast the Future

SLIDE NUMBER: 49

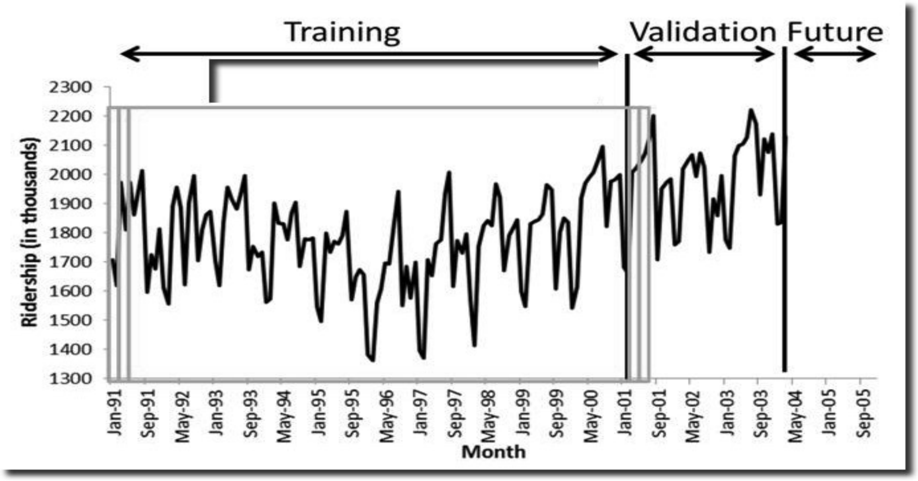
How to choose a Validation Period?

- Forecast Horizon
- Length of series
- Seasonality
- Underlying conditions affecting series
- Strategy to choose Validation Data Period

SLIDE NUMBER: 50

Rolling –Forward forecasts

Ridership (thousands) VS Month



SLIDE NUMBER: 51

NAÏVE Forecasts:

Forecast method: Last sample $\widehat{Y}_{t+1} = Y_t$

k-step ahead $F_{t+k} = Y_t$

Seasonal series (M series) $F_{t+k} = Y_{t-M+K}$

SLIDE NUMBER: 52

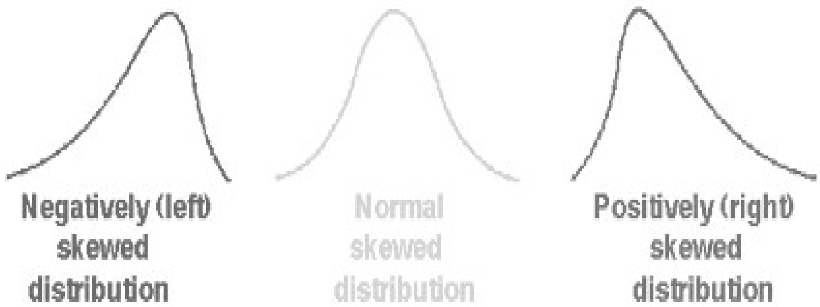
Forecast error:

- Forecast error is $e_t = Y_t - \widehat{Y}_t$
- If model is adequate, forecast error should contain no information
- Plots of e_t should resemble that of 'white noise' or uncorrelated random numbers with 0 mean and constant variance

  (There should be NO PATTERN)

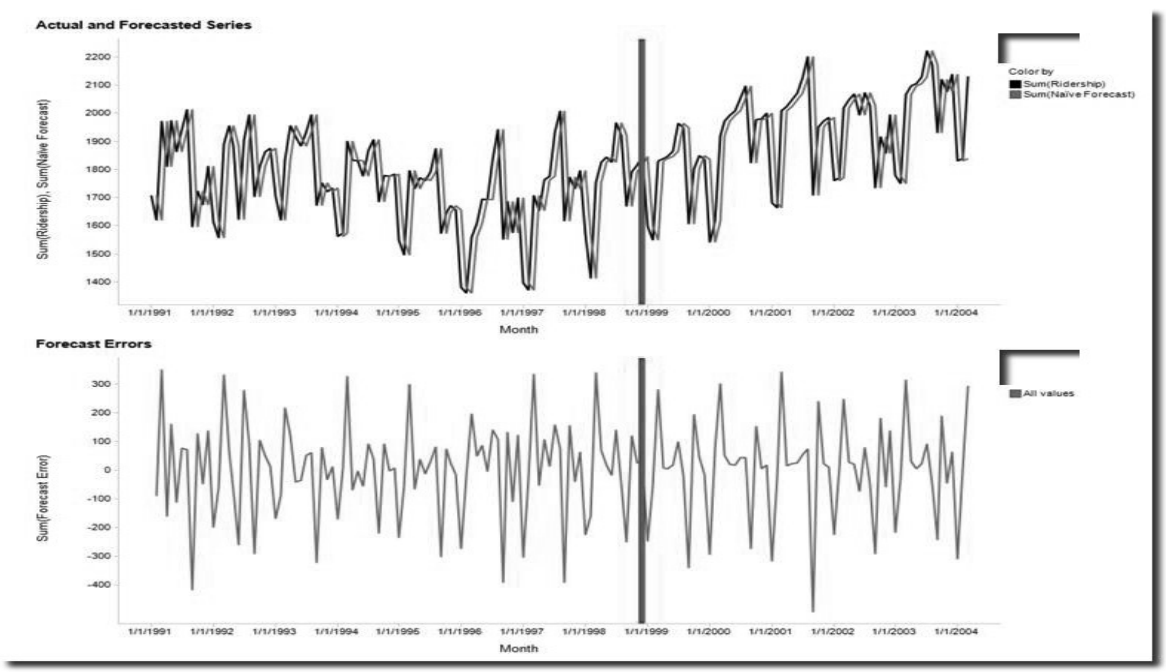
SLIDE NUMBER: 53

- Forecast error can follow different distributions based on business context



SLIDE NUMBER: 54

Forecasting Errors:



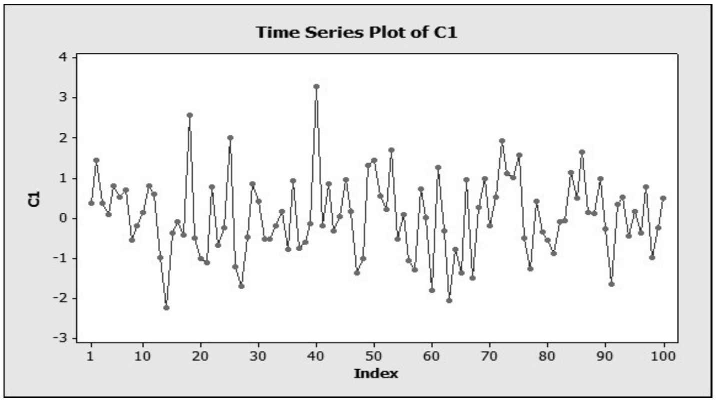
SLIDE NUMBER: 55

Evaluating Predictive Accuracy:

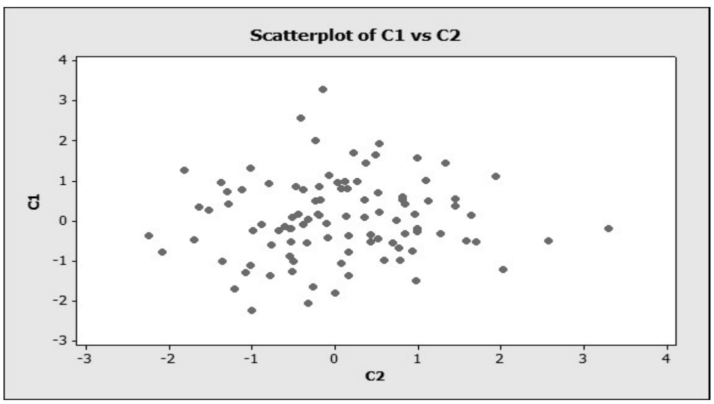
Mean error	$$ME=\frac{1}{T}\sum_{t=1}^{n}e_t$$
Mean absolute deviation	$$MAD=\frac{1}{n}\sum_{t=1}^{n}\lvert e_t\rvert$$
Mean squared error	$$MSE=\frac{1}{n}\sum_{t=1}^{n}e_t^2$$
Root mean squared error	$$RMSE=\sqrt{\frac{1}{n}\sum_{t=1}^{n}e_t^2}$$
Mean percentage error	$$MPE=\frac{1}{n}\sum_{t=1}^{n}\frac{e_t}{Y_t}$$
Mean absolute percentage error	$$MAPE=\frac{1}{n}\sum_{t=1}^{n}\left\lvert\frac{e_t}{Y_t}\right\rvert$$

SLIDE NUMBER: 56

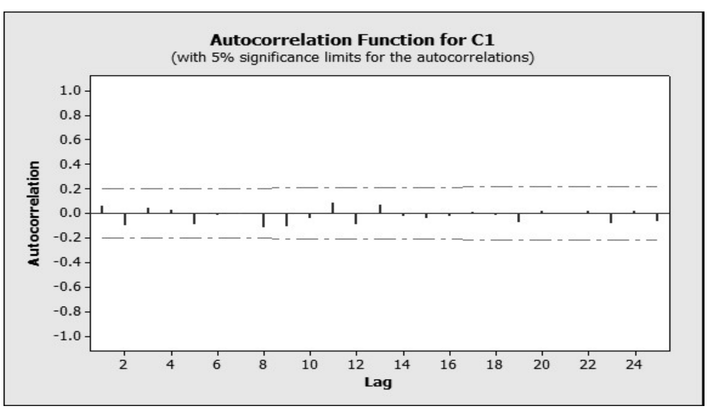
Typical plots of ‘White noise’



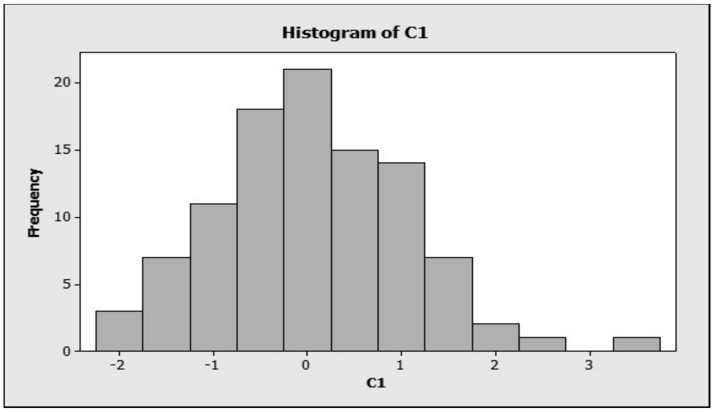
Time plot



Lag plot



ACF Plot



Histogram

SLIDE NUMBER: 57

Mean error (ME)

- If the ME is around zero, forecasts are called unbiased. Model is unbiased to overestimation or the underestimation. Certainly this is a desirable property of a model

Actual data	Forecast based on Model-1	Error from model-1	Forecast based on Model-2	Error from model-2
100	101	1	110	10
200	199	-1	190	-10
300	301	1	310	10
400	399	-1	390	-10
ME		0		0

SLIDE NUMBER: 58

- Mean error has the disadvantage that small amount and large amount of error may have same effect
- To overcome this problem we may define two different forecast performance measure
- 1. Mean Absolute Deviation:

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

- 2. Mean Square Error:

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

SLIDE NUMBER: 59

MAD & MSE:

Actual data	Forecast based on model-1	Error from model-1	Forecast based on model-2	Error from model-2
100	101	1	110	10
200	199	-1	190	-10
300	301	1	310	10
400	399	-1	390	-10
MAD		1		10
MSE		1		100
ME		0		0

SLIDE NUMBER: 60

Problem with ME, MAD, MSE:

- All these three measures are not unit free and also not scale free
- Just think of a case that one is forecasting sales figures. Someone in India using rupee figure, and somebody else in USA is expressing the same sales figure in dollar. Both are using the same model. However, forecast measure will differ. This is a very awkward situation
- MSE has the added disadvantage that its unit is in square. RMSE does not have this added disadvantage
- So we need unit free measure

SLIDE NUMBER: 61

MPE and MAPE---Unit free measure

$$MPE = \frac{1}{n}\sum_{t=1}^{n}\frac{e_t}{Y_t} \qquad MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{e_t}{Y_t}\right|$$

- Both expressed in percentage form
- Both are unit free

SLIDE NUMBER: 62

Last Sample: Number of customers requiring repair work

Customers Y_t	Fitted Value	Residual e_t	|e_t|	e_t^2	e_t/Y_t	|e_t/Y_t|
58						
54	58	-4	4	16	-0.07407	0.074074
60	54	6	6	36	0.1	0.1
55	60	-5	5	25	-0.09091	0.090909
62	55	7	7	49	0.112903	0.112903
62	62	0	0	0	0	0
65	62	3	3	9	0.046154	0.046154
63	65	-2	2	4	-0.03175	0.031746
70	63	7	7	49	0.1	0.1
		MAD	MSE	RMSE	MPE	MAPE
		4.25	23.5	4.85	0.0203	0.0695

Forecast method: Last sample $\hat{Y}_{t+1} = Y_t$

SLIDE NUMBER: 63

MA: Number of customers requiring repair work

Customers Y_t	Fitted Value	Residual e_t	|e_t|	e_t^2	e_t/Y_t	|e_t/Y_t|
58						
54	57.3333	-2.3333	2.3333	5.4444	-0.0424	0.0424
60	56.3333	5.6667	5.6667	32.1111	0.0914	0.0914
55	59.0000	3.0000	3.0000	9.0000	0.0484	0.0484
62	59.6667	5.3333	5.3333	28.4444	0.0821	0.0821
62	63.0000	0.0000	0.0000	0.0000	0.0000	0.0000
65	63.3333	6.6667	6.6667	44.4444	0.0952	0.0952
63	57.3333	**MAD**	**MSE**	**RMSE**	**MPE**	**MAPE**
70	56.3333	**3.83**	**19.91**	**4.46**	**0.0458**	**0.0599**
		-2.3333	2.3333	5.4444	-0.0424	0.0424
		5.6667	5.6667	32.1111	0.0914	0.0914

Forecast method: 3-point moving average

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + Y_{t-2}}{3}$$

SLIDE NUMBER: 64

Challenges:

Missing values

- Compute average metrics
- Exclude missing values

Zero Counts

MAE/RMSE: no problem

Cannot compute MAPE

- Exclude Zero count
- Use alternate measure – MASE

SLIDE NUMBER: 65

Forecast / Prediction Interval

Probability of 95% that the value will be in the range [a,b]

If the forecast errors are normal, prediction interval is

σ = estimated standard deviation of forecast errors

k = some multiple

(k=2 corresponds to 95% probability)

$$F_{t+k} \pm k\,\sigma$$

Challenges to formula

•Errors often non-normal

•If model is biased (over/under-forecasts), symmetric interval around F_{t+k}?

• Estimating the error standard deviation is tricky

One solution is transforming errors to normal

SLIDE NUMBER: 66

Forecast / Prediction Interval – Non-Normal

To construct prediction interval for 1-step-ahead forecasts

1. Create roll-forward forecasts (F_{t+1}) on validation period
2. Compute forecast errors
3. Compute percentiles of error distribution

   ($e^{(5)}$=5th percentile; $e^{(95)}$=95th percentile)
4. Prediction interval:

   $[\ F_{t+1} + e^{(5)}\ ,\ F_{t+1} + e^{(95)}\ ]$

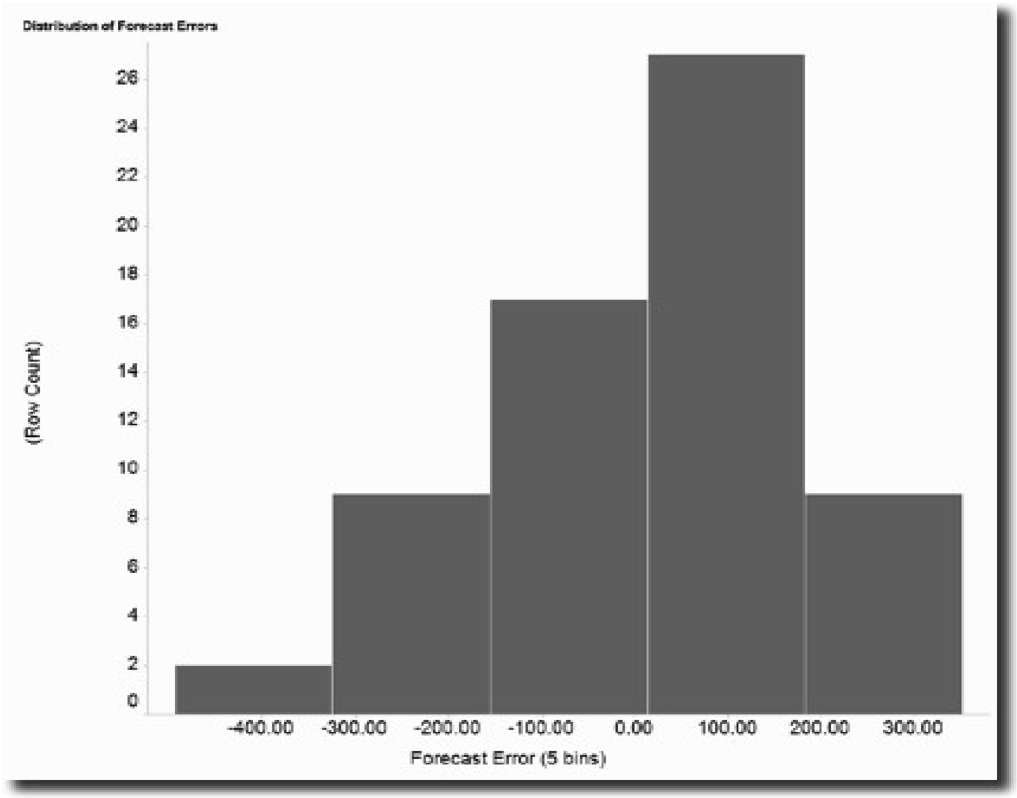
In Excel *=percentile*

5th percentile = -307.0

95th percentile = 292.8

95% prediction interval for 1-step ahead forecast F_{t+1}:

$[F_{t+1} - 307\ ,\ F_{t+1} + 292.8]$



SLIDE NUMBER: 67

Forecasting Different Methods:

Model Based:

- **Linear regression**
- **Autoregressive models**
- **ARIMA**
- **Logistic regression**
- **Econometric models**

Data driven:

- **Naïve forecasts**
- **Smoothing**
- **Neural nets**

Linear regression combined with economics = econometric

Auto regression means self-regression; Yt regresses on Yt-1, Yt-2 etc It is used on top of Linear regression because we assume that Yt, Yt-1, Yt-2 are dependent

ARIMA = Combination of model based & data driven, AR is model based approach; MA is data driven. It tends to be very complex but very useful. Explaining to stakeholders will become difficult

Econometric models: Macroeconomics (GDP etc.) Application of regression technique for economic scenario.

If intuitive knowledge is better, then we can go with Model based approach. Data based is done using black box techniques.

SLIDE NUMBER: 68

Forecasting Different Methods:

Linear Model: $Y_t = \beta_o + \beta_1 t + \varepsilon$

Exponential Model: $\text{Log}(Y_t) = \beta_o + \beta_1 t + \varepsilon$

Quadratic Model: $Y_t = \beta_o + \beta_1 t + \beta_2 t^2 + \varepsilon$

Additive Seasonality: $Y_t = \beta_o + \beta_1 D_{Jan} + \beta_2 D_{Feb} + \beta_3 D_{Mar} + \ldots\ldots + \beta_{11} D_{Nov} + \varepsilon$

Additive Seasonality with Quadratic Trend:

$Y_t = \beta_o + \beta_1 t + \beta_2 t^2 + \beta_3 D_{Jan} + \beta_4 D_{Feb} + \beta_5 D_{Mar} + \ldots\ldots + \beta_{13} D_{Nov} + \varepsilon$

Multiplicative Seasonality:

$\text{Log}\ (Y_t) = \beta_o + \beta_1 D_{Jan} + \beta_2 D_{Feb} + \beta_3 D_{Mar} + \ldots\ldots + \beta_{11} D_{Nov} + \varepsilon$

Linear Model: Simple & Multiple; after showing SLR, go to MLR_training score & add the month details beside Predicted value. Jan 91 …etc

SLIDE NUMBER: 69

Irregular Component:

Irregular Components	**Solutions**
− **Outliers** − **Special Events** − **Interventions**	− **Remove unusual periods from the model** − **Model Separately** − **Keep in the model, and use dummy variable.**

T20 match, football world cup match – traffic along the stadium

Interventions (promotion, policy change).

SLIDE NUMBER: 70

External Information

- Forecasting Airline Ticket Price

- Forecasting Internet Sales
- $Airfare_t$ = b0 + b1 (Petrol Price)$_t$ + e → Must be forecasted

- Fuel price impacts the airline ticket
- Amount spend in advertisements
- Sales(t) =
- *g*{ *f*(sales(t-1, t-2, ... , t-6), a1*SQRT[AdSpend(t-1)] + ... + a6*SQRT[AdSpend(t-6)] }

SLIDE NUMBER: 71

Linear Regression for forecasting

Global Trend

- Linear Trend (constant growth)
- Exponential Trend (% growth)

Seasonality

- Additive (Y)
- Multiplicative (log(Y)
- Irregular Patterns

- Global trend
- linear trend (constant growth) - use time index as predictor
- exponential trend (% growth) - use log(Y) as response and time index as predictor

- Additive or multiplicative seasonality (or other shape)
- Use dummy variables for seasons
- Y for additive model or log(y) for multiplicative model

SLIDE NUMBER: 72

Autoregressive (AR) Models

- AR model is used to forecast errors
- AR model captures autocorrelation directly
- Autocorrelation measures how strong the values of a time series are related to their own past values
- Lag(1) autocorrelation = correlation between

  $(y_1, y_2,\dots, y_{t-1})$ and $(y_2,y_3,\dots, y_t)$

- Lag(k) autocorrelation = correlation between

  $(y_1, y_2,\dots\dots,y_{t-k})$ and $(y_{k+1},y_{k+2},\dots,y_t)$

  - Technically: compute the **correlation** between the series and the lagged series (approximately)
  - Lag(1) autocorrelation = correlation between (y1, y2, …, yt-1) and (y2,y3,…, yt)
  - Lag(k) autocorrelation = correlation between (y1, y2, …, yt-k) and (yk+1,yk+2,…,yt)
  - Note: autocorrelation measures **linear** relationship

SLIDE NUMBER: 73

Autocorrelation & its uses

- Check forecast errors for independence
- Model remaining information
- Evaluate predictability
- Positive lag-1 autocorrelation ("stickiness")?:
- high values usually immediately follow _______ values, and low values usually immediately follow _______ values
- Negative lag-1 autocorrelation ("swings"?):
- high values usually immediately follow _______ values, and low values usually immediately follow _________ values
- High positive autocorrelation at multiples of a certain lag (e.g. lags 4, 8, 12…) indicates ____________.

SLIDE NUMBER: 74

Autoregressive Model

- Multi-layer model
- Model the forecast errors, by treating them as a time series
- Then examine autocorrelation of "errors of forecast errors" ?
  - ✓ If autocorrelation exists, fit an AR model to the forecast errors series
  - ✓ If autocorrelated, continue modeling the level-2 errors (not practical)
  - ✓ AR model can also be used to model original data

$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \varepsilon_t$ -> AR(2), order = 2

1-step ahead forecast : $F_{t+1} = \alpha + \beta_1 Y_t + \beta_2 Y_{t-1}$

2-steps ahead: $F_{t+2} = \alpha + \beta_1 F_{t+1} + \beta_2 Y_t$

3-steps ahead: $F_{t+3} = \alpha + \beta_1 F_{t+2} + \beta_2 F_{t+1}$

SLIDE NUMBER: 75

Autoregressive Model

- Use level 1 to forecast next value of series F_{t+1}
- Use AR to forecast next forecast error (residual) E_{t+1}
- Combine the two to get an improved forecast F^*_{t+1}

  $F^*_{t+1} = F_{t+1} + E_{t+1}$

SLIDE NUMBER: 76

Random Walk

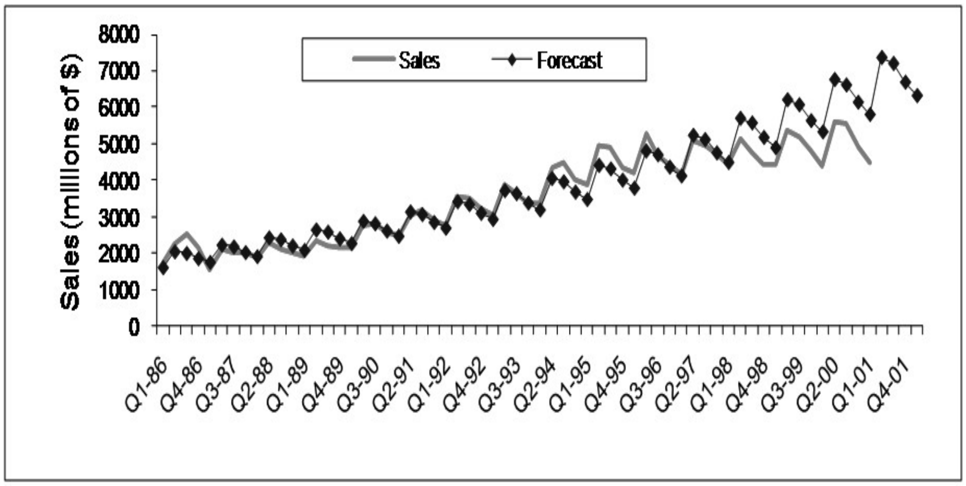
- Specific case of AR(1) model
- If $\beta_1 = 1$ in AR(1) model then it is called as Random Walk
- Equation will be $Y_t = a + Y_{t-1} + \varepsilon_t$
  - a = drift parameter
  - σ(std of ε) = volatility
- Changes from one period to the next are random
- How to find out whether there in random walk to not in the data?
  - Run AR(1) model & check for the value of β_1
  - Do a differenced series and run ACF plot
  - How to estimate drift & volatility?

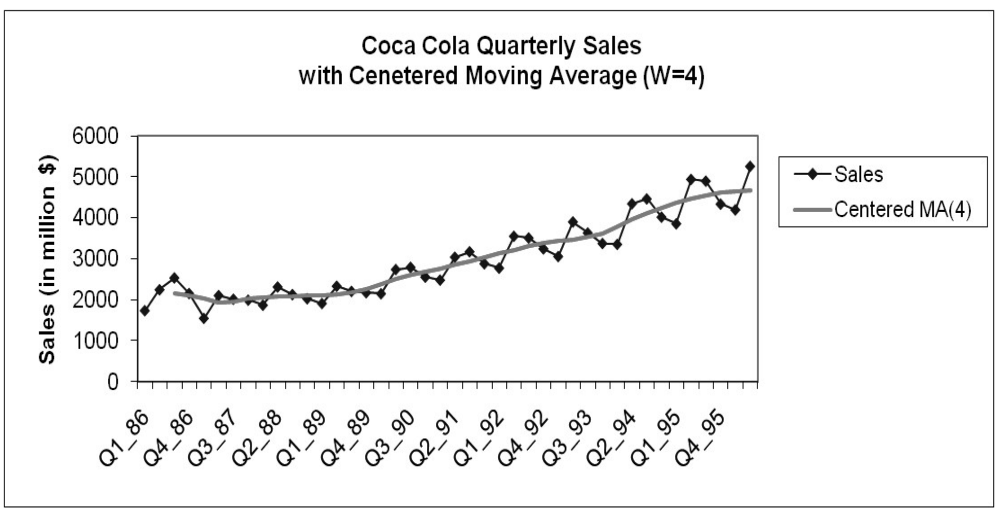
SLIDE NUMBER: 77

- One-step-ahead forecast: $F_{t+1} = a + Y_t$
- Two-step-ahead forecast: $F_{t+2} = a + Y_{t+1} = 2a + Y_t$
- k-step-ahead forecast : $F_{t+k} = ka + Y_t$
- If the drift parameter is 0, then the k-step-ahead forecast is $F_{t+k} = Y_t$ for all k

SLIDE NUMBER: 78

Model based approaches & drawbacks





SLIDE NUMBER: 79

Model vs Data based approaches

- Model Based Approach
  - Past is SIMILAR to Future
- Data Based Approach
  - Past is NOT SIMILAR to Future

- Model based vs data based depends on complexity of dataset & how many types of seasonality we have. Is complexity constant throughout or is it changing.
- If focus is Indian market then I use data driven approach because market is very volatile. Model based approach is very good for one time forecast & ongoing basis then use databased. Use combination of model & data based approaches

SLIDE NUMBER: 80

Forecast methods based on smoothing

There are two major forecasting techniques based on smoothing

- Moving averages

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + \cdots + Y_{t-k+1}}{k}$$

- Exponential smoothing

$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\,\hat{Y}_t$$

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

– Success depends on choosing window width

– Balance between over & under smoothing

– If no seasonality then use narrow window. In financial data, look for random walk

SLIDE NUMBER: 81

Smoothing – Moving Average

➔ Smoothing Noise

- Forecast future points by using an average of several past points
- More suitable for series with no Trend & no seasonality

➔ Forecasting
➔ Removing Seasonality & Computing seasonal indexes
➔ Data Visualization

- A time-plot of the MA reveals the Level & Trend of a series
- It filters out the seasonal & random components

Windows = no. of seasonality; coke = 4 & Amtrak = 12;

SLIDE NUMBER: 82

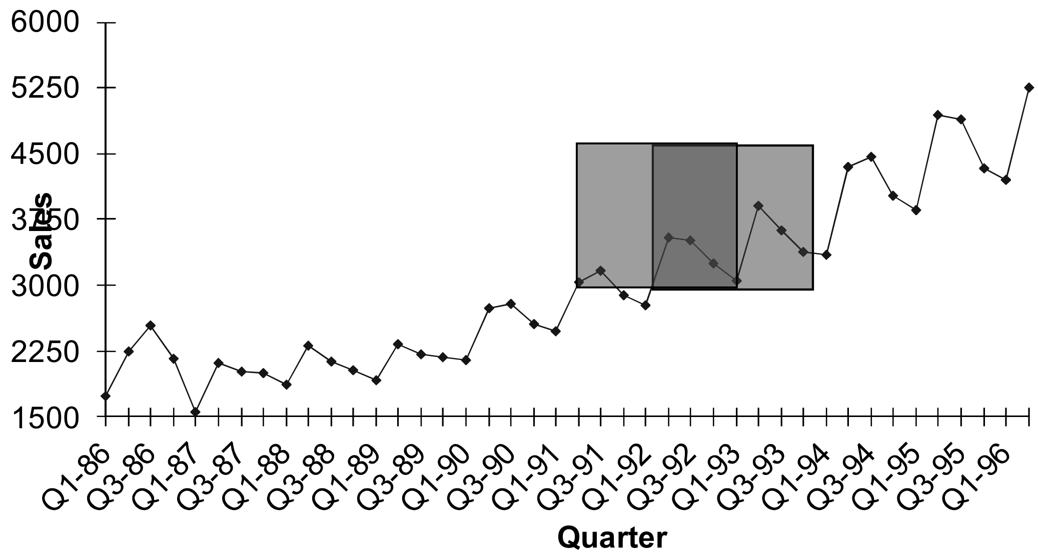
Moving Average - Calculations

Centered Moving Average

- It is calculated based on a window centered around time 't'

Trailing Moving Average

- It is calculated based on a window from time't' & backwards



SLIDE NUMBER: 83

1. Calculation – Trailing MA Choose window width (W)
2. For MA at time t, place window on time points *t-W+1,…,t*

W=5

t-4 t-3 t-2 t-1 t

3. Compute average of values in the window:

$$MA_t = \frac{y_{t-W+1} + y_{t-W+2} + \cdots + y_{t-1} + y_t}{W}$$

SLIDE NUMBER: 84

Calculation – Centered MA

Compute average of values in window (of width *W*), which is centered at *t*

Odd width: center window on time *t* and average the values in the window

Even width: take the two “almost centered” windows and average the values in them

W=4 w=4

W=5

t-2 t-1 t t+1 t+2

$$MA_t = \left(\frac{y_{t-2} + y_{t-1} + y_t + y_{t+1}}{4} + \frac{y_{t-1} + y_t + y_{t+1} + y_{t+2}}{4} \right) / 2$$

SLIDE NUMBER: 85

Moving Average Hands On

SLIDE NUMBER: 86

Exponential Smoothing

Simple Exponential Smoothing

- No Trend
- No Seasonality
- Level
- Noise (cannot be modeled)

Holt’s method

- Also called double exponential

- Trend
- No Seasonality

Winter's method

- Trend
- Seasonality
- Variants are possible

➢ Assigns more weight to most recent observations
➢ Assigns less weight to farthest observations

SLIDE NUMBER: 87

Simple Exponential Smoothing:

Forecasts = *estimated level* at most recent time point: $F_{t+k} = L_t$

Adaptive algorithm: adjusts most recent forecast (or level) based on the actual data:

$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\,\hat{Y}_t$$

$$L_t = \alpha Y_t + (1-\alpha) L_{t-1}$$

α = the *smoothing constant* ($0 < \alpha \leq 1$)

Initialization: $F_1 = L_1 = Y_1$

SLIDE NUMBER: 88

Simple Exponential Smoothing

The formula: $L_t = \alpha Y_t + (1- \alpha) L_{t-1}$

Substitute L_t with its own formula:

$L_t \quad = \alpha Y_t + (1- \alpha)[\alpha Y_{t-1} + (1- \alpha) L_{t-2}] =$

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

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

SLIDE NUMBER: 89

Simple Exponential Smoothing

The formula: $L_t = \alpha Y_t + (1- \alpha) L_{t-1}$

$Y_{t+1} = L_t = L_{t-1} + \alpha (Y_t - L_{t-1})$

$= Y_t \;\; + \alpha (Y_t - Y_t)$

$= Y_t + \alpha E_t$

Update previous forecast → Y_t

E_t → By an amount that depends on the error in the previous forecast

α → controls the degree of "learning"

SLIDE NUMBER: 90

Smoothing Constant 'α'

α determines how much weight is given to the past

α =1: past observations have no influence over forecasts (under-smoothing)

$\alpha \to 0$: past observations have large influence on forecasts (over-smoothing)

Selecting α

"Typical" values: 0.1, 0.2

Trial & error: effect on visualization

Minimize RMSE or MAPE of training data

If alpha = 0; today's value = tomorrow; If huge volatility then we can have high alpha value;

If less volatility then use low alpha value

SLIDE NUMBER: 91

Exponential Smoothing Hands On

SLIDE NUMBER: 92

MA Vs ES

MA	ES
• Assigns equal weights to all past observations • Better to forecast when data & environment is not volatile • Window width is key to success	• Assigns more weight to recent observations than past observations • Better to forecast when data & environment is volatile • Smoothing constant (α) value is key to success

SLIDE NUMBER: 93

De-trending & De-seasoning

Regression

- To remove trend and/or seasonality, fit a regression model with trend and/or seasonality
- Series of forecast errors should be de-trended & de-seasonalized

Differencing

- Simple & popular for removing trend and / or seasonality from a time series
- Lag-1 difference: Y_t – Yt_{-1} (For removing trend) ; Lag-M difference: $Y_t – Y_{t-M}$ (For removing seasonality)
- Double – differencing: difference the differenced series

Ratio to Moving average

- Uses moving average to remove seasonality
- Generates seasonal indexes as a byproduct

SLIDE NUMBER: 94

Seasonal Indexes

For a series with M seasons:

S_j = seasonal index for the j^{th} season

Indicates the exceedance of Y on season *j* above/below the average of Y in a complete cycle of seasons

Example: Daily sales at retail store shows that Friday has a seasonal index of 1.30 and Monday has an index of 0.65

Meaning: Friday sales is 30% higher than the weekly average, and Monday sales is 35% lower than the weekly average sales

Average of the M seasonal indexes is 1 (they must sum to M).

SLIDE NUMBER: 95

Seasonal Indexes

1. Construct the series of *centered* moving averages of span *M.*
2. For each *t*, compute the *raw seasonals* = Y_t / MA_t
3. S_j = average of raw seasonals belonging to season *j* (normalize to ensure that seasonal indexes have average=1)

Quarter	Sales	Centered MA with W=4			
			raw seasonal	s(j)	seasonal index
Q1_86	1734.83				
Q2_86	2244.96				
Q3_86	2533.80	2143.76	1.18194269	1.063872992	**1.062660535**
Q4_86	2154.96	2102.82	1.02479749	0.96423378	**0.963134878**
Q1_87	1547.82	2020.32	0.76612585	0.879530967	**0.878528598**
Q2_87	2104.41	1934.99	1.08755859	1.096926116	**1.09567599**
Q3_87	2014.36	1954.74	1.03050222		
Q4_87	1991.75	2021.05	0.9855033		
Q1_88	1869.05	2061.44	0.90667088		
Q2_88	2313.63	2080.07	1.11228431		
Q3_88	2128.32	2089.65	1.01850452		
Q4_88	2026.83	2097.04	0.96651998		
Q1_89	1910.60	2109.01	0.90592533		
Q2_89	2331.16	2137.18	1.09076712		

SLIDE NUMBER: 96

Seasonal Indexes

De-seasonalized (=seasonally-adjusted) series:

$DSY_t = Y_t$ / appropriate seasonal index

- If done appropriately, de-seasonalized series will not exhibit Seasonality
- If so, examine for trend and fit a model
- This model will yield de-seasonalized forecasts
- Convert forecasts by re-seasonalizing, i.e. multiply them by the appropriate seasonal index

SLIDE NUMBER: 97

The seasonally-adjusted sales for Q1-86 are in the range

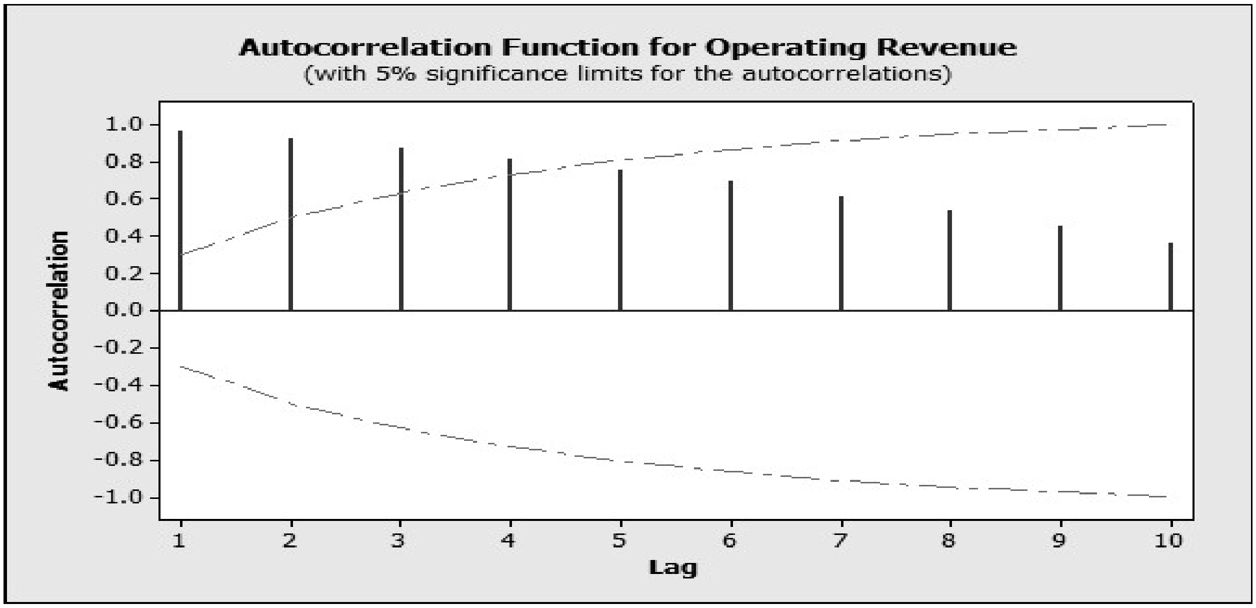
1. \$1500-\$1700 (million)
2. \$1700-\$1800 (million)
3. \$1800-\$1900 (million)
4. \$1900-\$2000 (million)

1734.83 / 0.8785 = \$1974.8 mil

Quarter	Sales	Centered MA with W=4	raw seasonal	s(j)	seasonal index
Q1_86	1734.83				
Q2_86	2244.96				
Q3_86	2533.80	2143.76	1.18194269	1.063872992	**1.062660535**
Q4_86	2154.96	2102.82	1.02479749	0.96423378	**0.963134878**
Q1_87	1547.82	2020.32	0.76612585	0.879530967	**0.878528598**
Q2_87	2104.41	1934.99	1.08755859	1.096926116	**1.09567599**
[illegible]	[illegible]	[illegible]	[illegible]		

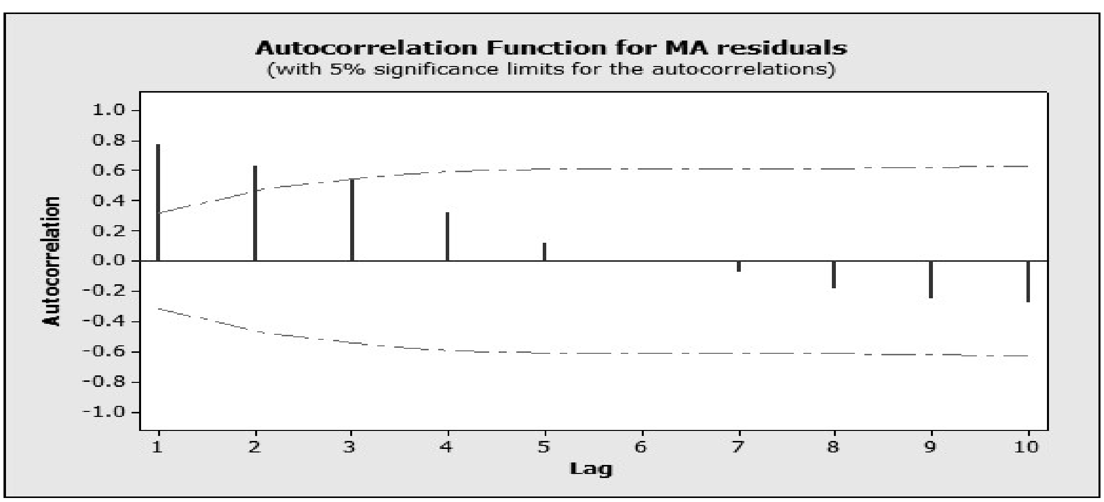
SLIDE NUMBER: 98

Example: Forecasting 2005 sales of Sears from 1955-2004 data (Table 3-4)



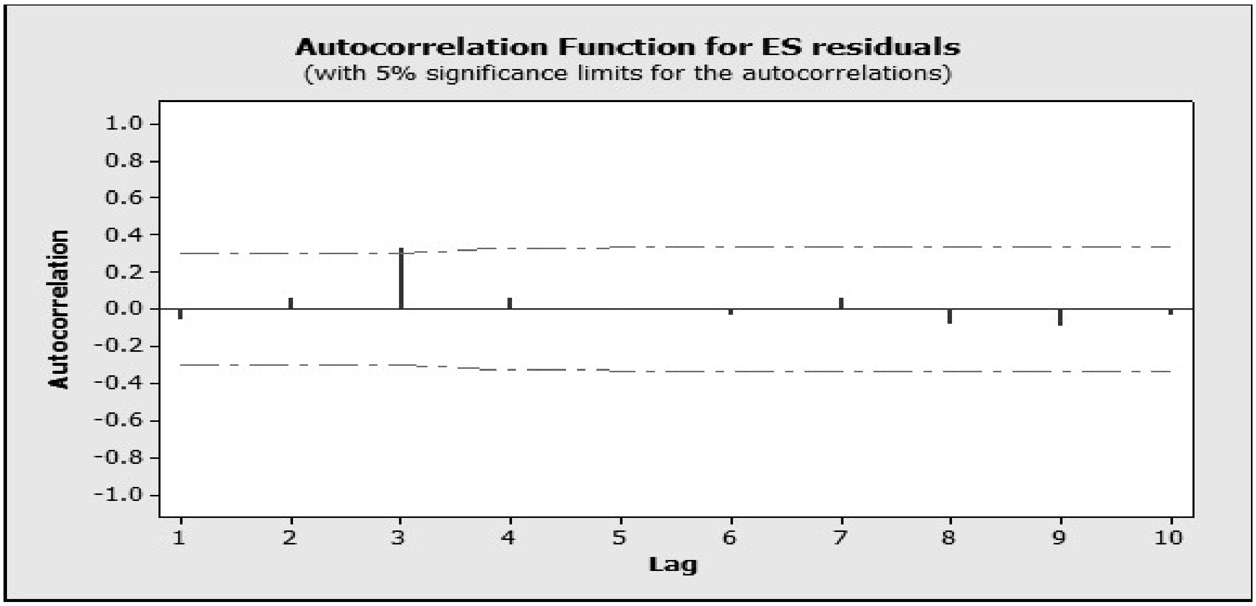
SLIDE NUMBER: 99

Sears sales data: ACF of 5-point MA residuals



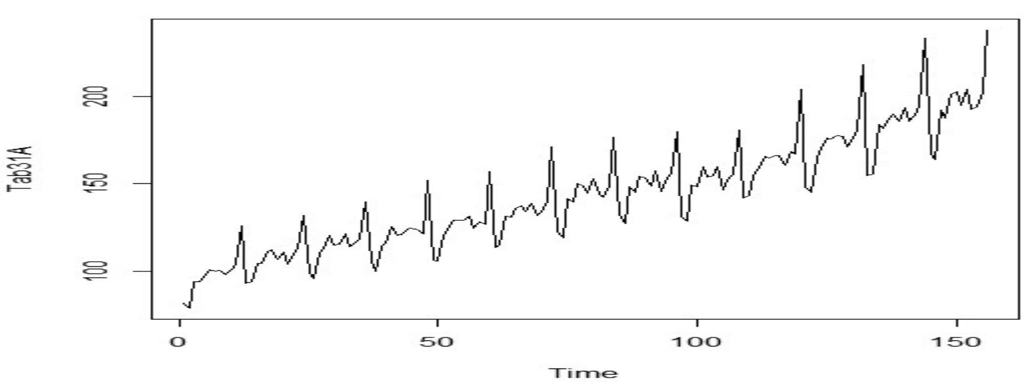
SLIDE NUMBER: 100

Sears sales data: ACF of Exponential smoothing residuals



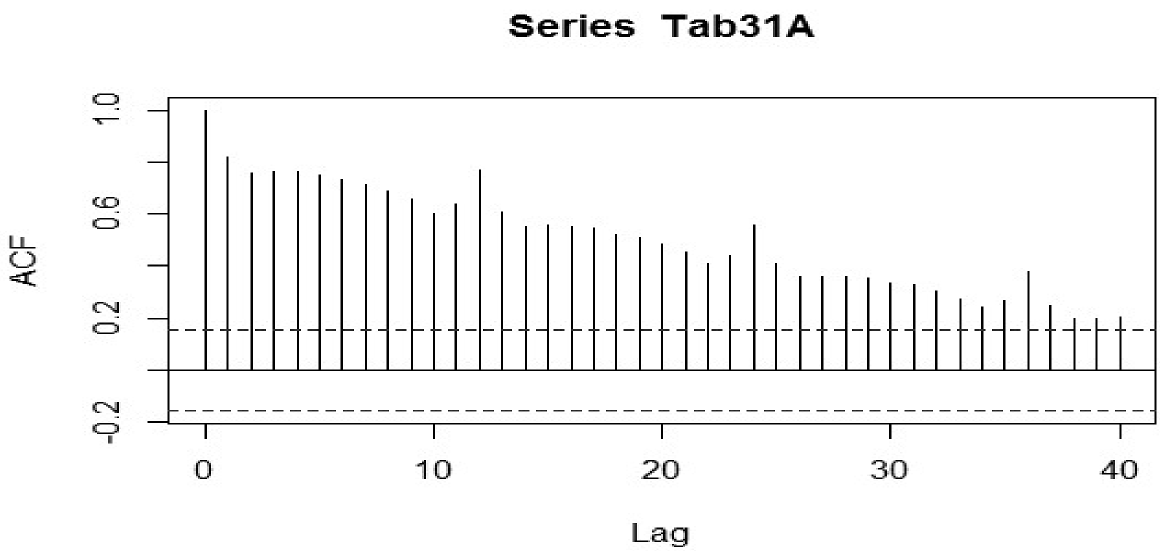
SLIDE NUMBER: 101

Example: Monthly sales for All US retail stores,1983-1995 (Table 3.8)



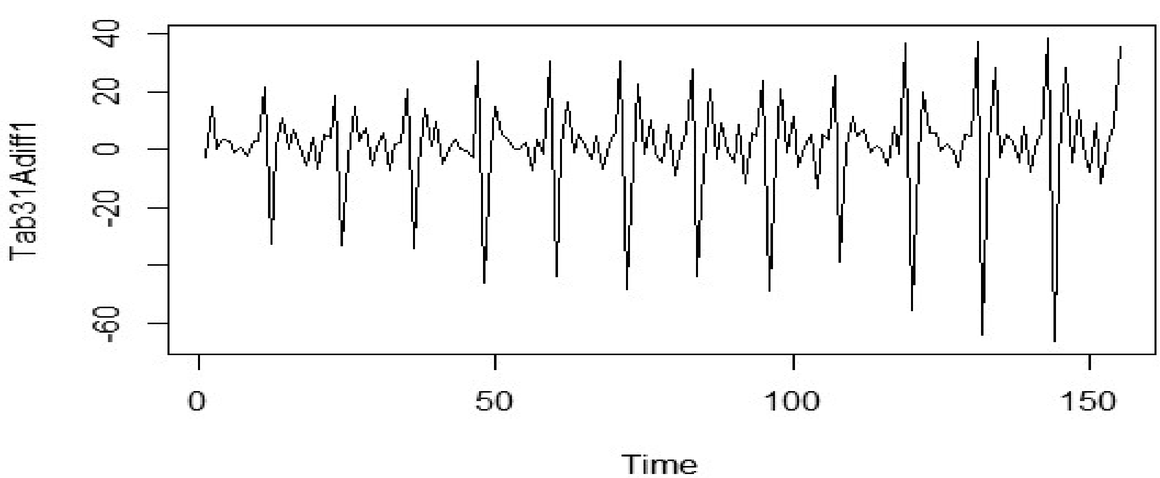
SLIDE NUMBER: 102

Autocorrelation function for retail sales



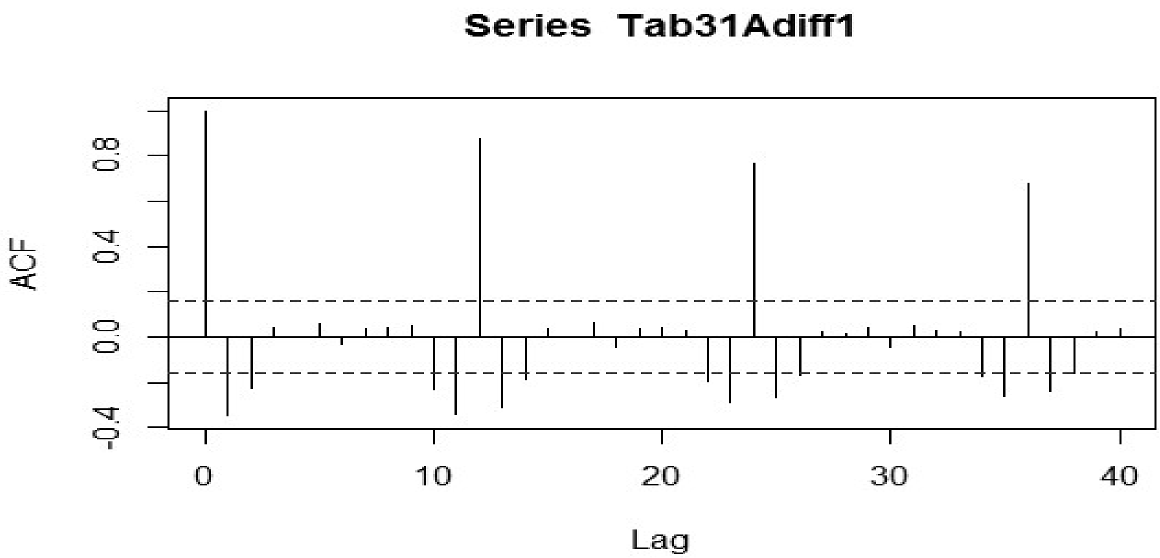
SLIDE NUMBER: 103

Time plot of US Monthly retail sales: differenced



SLIDE NUMBER: 104

ACF of US Monthly retail sales: differenced



SLIDE NUMBER: 105

Formal (classical) Treatment of Time Series Data

Decomposition of a time series (Y_t) into its components:

1) **Trend (T_t)**
2) **Cycle (C_t)**
3) **Seasonal (S_t)**
4) **Irregular (I_t)**

Different combinations of the above components:

A: $Y_t = T_t \times C_t \times S_t \times I_t \longrightarrow$ Multiplicative Model (Most Popular)

B: $Y_t = T_t + C_t + S_t + I_t \longrightarrow$ Additive Model (Less Popular)

C: $Y_t = T_t \times C_t \times S_t + I_t \longrightarrow$ Mixed Model

D: Other Combinations $\longrightarrow$ Mixed Model

SLIDE NUMBER: 106

Classical Treatment of Time Series Data When to use Additive Model

- **In some time series, the amplitude of both the seasonal and irregular variations do not change as the level of the trend rises or falls.**
- **In such cases, an additive model is appropriate. In the additive model, the observed time series is considered to be the sum of three independent components.**
- **Each of the three components has the same units as the original series.**

SLIDE NUMBER: 107

Formal (classical) Treatment of Time Series Data When to use Multiplicative Model

- **In many time series, the amplitude of both the seasonal and irregular variations increase as the level of the trend rises.**
- **In this situation, a multiplicative model is usually appropriate.**
- **Under this model, the trend has the same units as the original series, but the seasonal and irregular components are unit less factors, distributed around 1.**

SLIDE NUMBER: 108

ADDITIVE VS MULTIPLICATIVE SEASONALITY

- **Seasonal components can be additive in nature or multiplicative. For example, during the month of December the sales for a particular toy may increase by 1 million dollars every year. Thus, we could *add* to our forecasts for every December the amount of**

1 million dollars (over the respective annual average) to account for this seasonal fluctuation. In this case, the seasonality is *additive*.

- **Alternatively, during the month of December the sales for a particular toy may increase by 40%, that is, increase by a *factor* of 1.4. Thus, when the sales for the toy are generally weak, than the absolute (dollar) increase in sales during December will be relatively weak (but the percentage will be constant); if the sales of the toy are strong, than the absolute (dollar) increase in sales will be proportionately greater. Again, in this case the sales increase by a certain *factor*, and the seasonal component is thus *multiplicative* in nature (i.e., the multiplicative seasonal component in this case would be 1.4).**

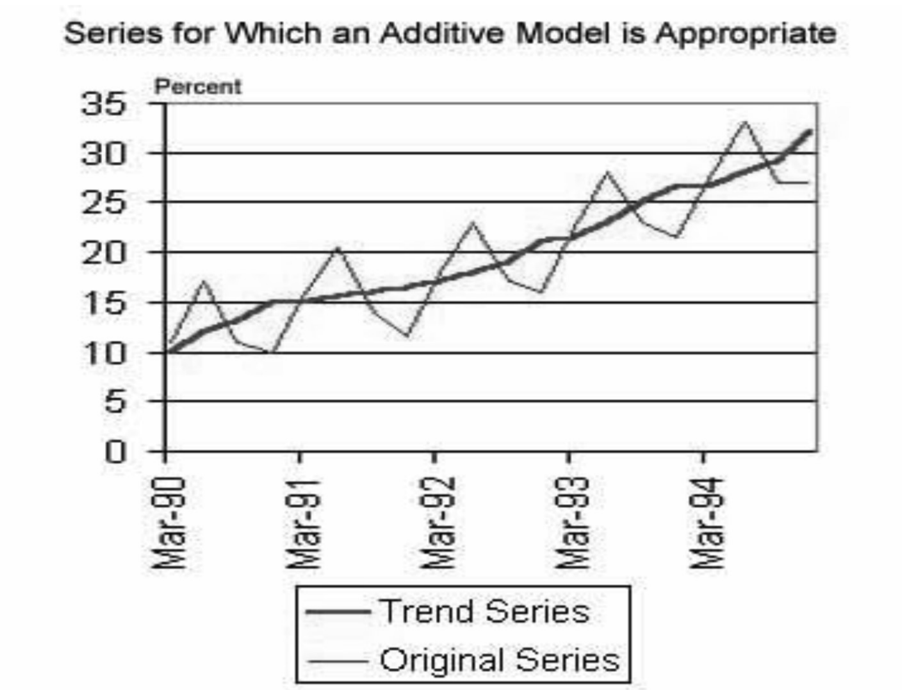
SLIDE NUMBER: 109

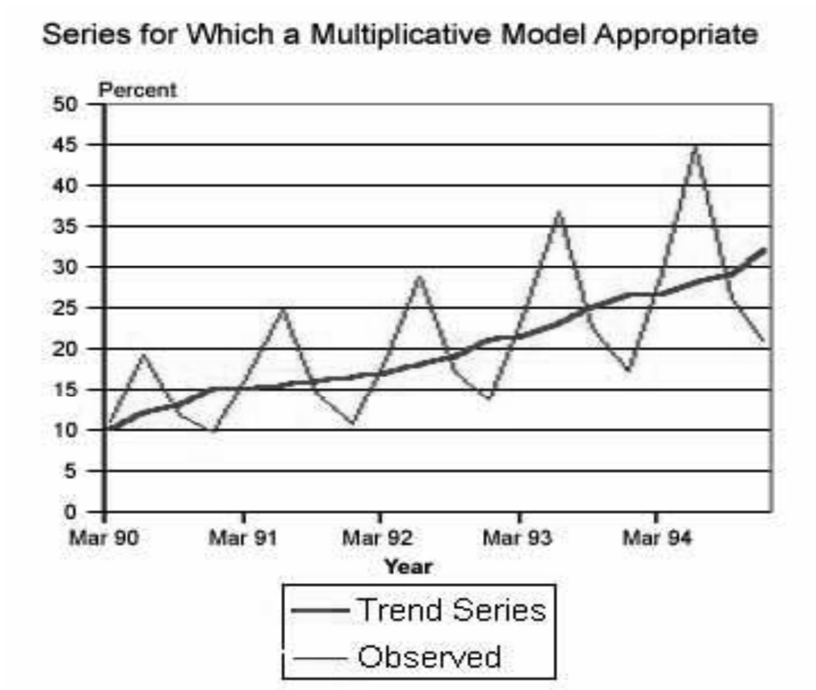
ADDITIVE VS MULTIPLICATIVE SEASONALITY

- **In plots of the series, the distinguishing characteristic between these two types of seasonal components is that in the additive case, the series shows steady seasonal fluctuations, regardless of the overall level of the series; in the multiplicative case, the size of the seasonal fluctuations vary, depending on the overall level of the series.**

SLIDE NUMBER: 110

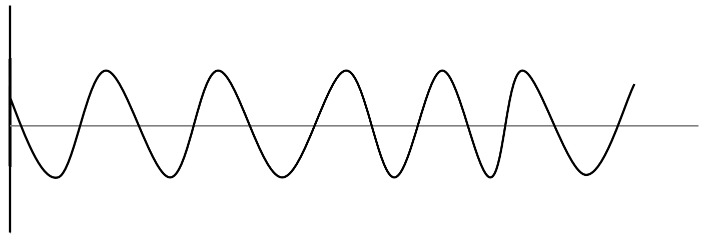
ADDITIVE VS MULTIPLICATIVE SEASONALITY



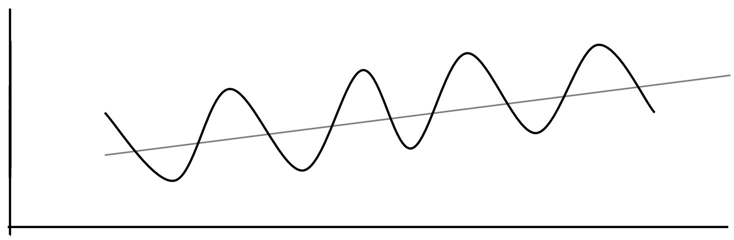


SLIDE NUMBER: 111

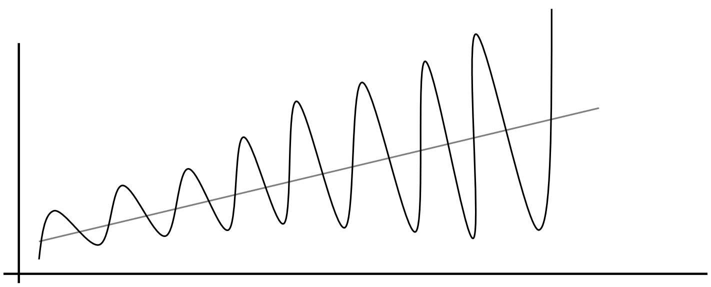
Additive and Multiplicative Models



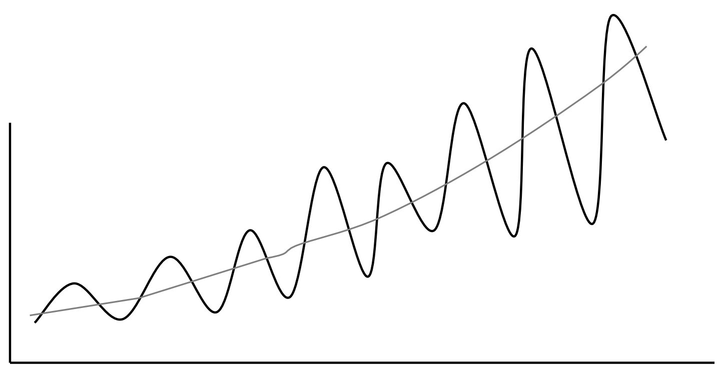
1. No trend and additive seasonal variability



2. Additive seasonal variability with trend



3. Multiplicative seasonal variability with a trend.



4. Multiplicative seasonality variability with non-linear trend

SLIDE NUMBER: 112

Typical models

- Additive model

$Y_t = T_t + S_t + C_t + I_t$

- Multiplicative model

$Y_t = T_t \times S_t \times C_t \times I_t$

$\log(Y_t) = \log(T_t) + \log(S_t) + \log(C_t) + \log(I_t)$

A multiplicative model can be regarded as an additive model for log-transformed time series.

SLIDE NUMBER: 113

Curve Fitting:

- From the above plot, different people may fit different line based on their naked eye judgment.
- Which one is the best 'in some sense'?
- We introduce a method known as 'Least Square Method'.

SLIDE NUMBER: 114

Curve Fitting

. This method assumes a particular form of trend, say, linear trend, in the form as:

$T_t = a + b \times t$

Where *a* and *b* are unknown, to be estimated from the data.

- Note that we know the value of *t* in the above equation but we do not have any data on T_t. But we have data on Y_t where:

$Y_t = T_t + S_t + I_t$

SLIDE NUMBER: 115

Curve Fitting:

Note that we assume an additive model. In case of multiplicative model, we can take the log transformation:

Yt= Tt x St x It   or

logYt= logTt + logSt + logIt

Thus, we reduce it to the additive form.

SLIDE NUMBER: 116

Curve Fitting

- Least Square Method as the name suggests tells us to look for values of '*a'* and '*b'* such that the sum of the square of the '*error*' is minimized.

- What is '*error*'?
- Given any values for 'a' and 'b' for each time point – 't' we can calculate the value of the trend – T_t

SLIDE NUMBER: 117

- But we have some data value - Y_t
- The error for each time point e_t is defined as the difference between the actual data and the calculated trend i.e.:

$$e_t = Y_t - T_t$$

- The sum of the square of the '*error*' is given by:

$$\Sigma_t e_t^2 = \Sigma_t (Y_t - T_t)^2 = \Sigma_t (Y_t - a + b \times t)^2$$

SLIDE NUMBER: 118

Curve Fitting

- The solution to the minimization is given by:

$$b = (n\Sigma_t tY_t - \Sigma_t Y_t \Sigma_t t) / ({}_n \Sigma_t t^2 {}_- \ \Sigma_t t)$$

$$a = (\Sigma_t Y_t - b\Sigma_t t) / n$$

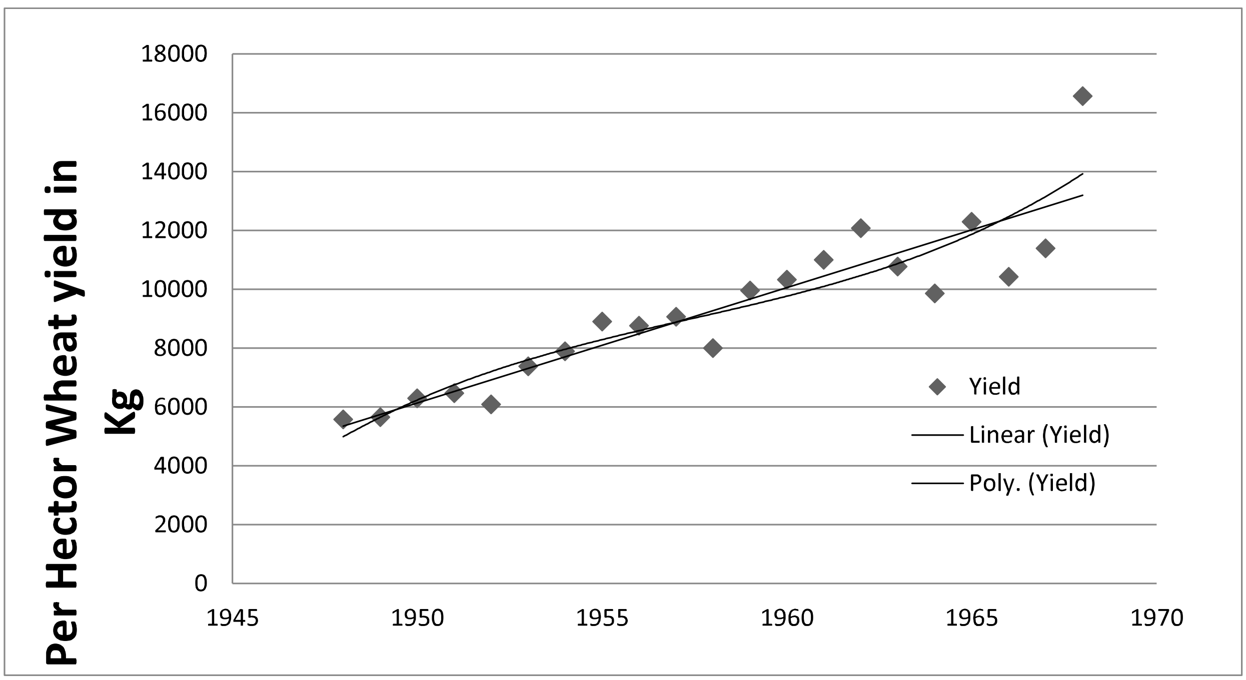
SLIDE NUMBER: 119

- This curve fitting technique is not restricted to linear trends. It can be obtained for polynomial, exponential as well as other types of non – linear trends.

- One advantage of this technique is that forecasts for any particular time point can be readily obtained from the calculated coefficients of the fitted curve.
- One disadvantage though is that the choice of the curve to be fitted is subjective

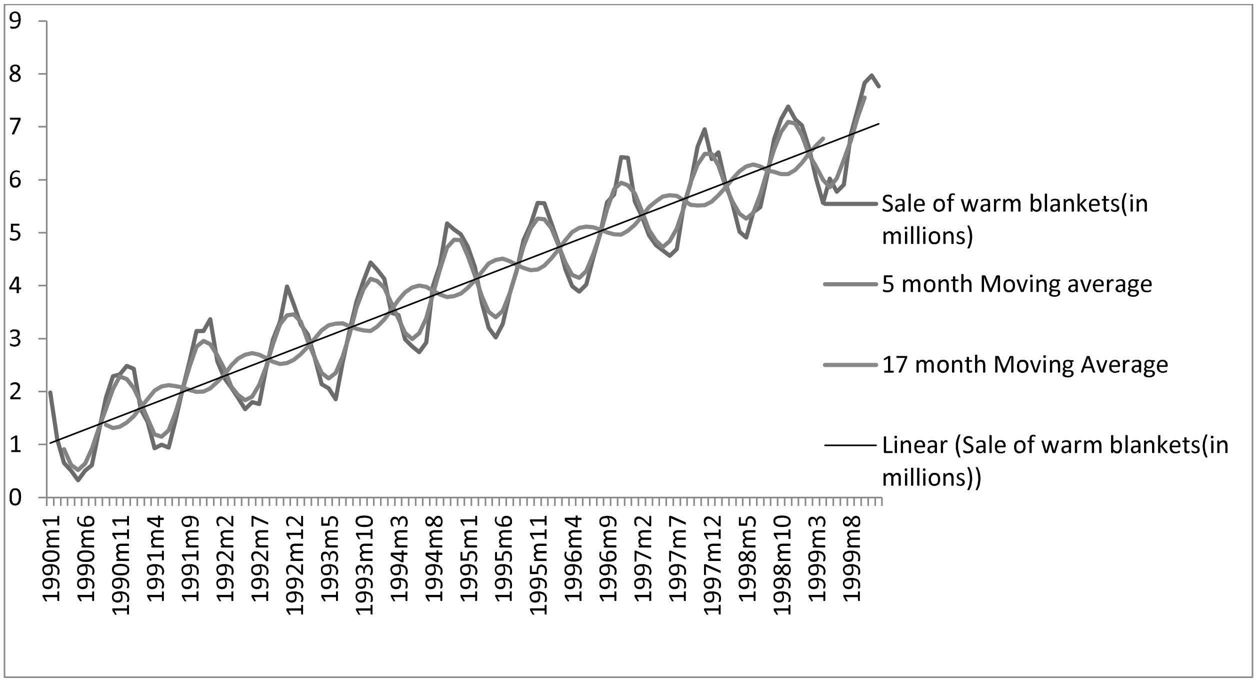
SLIDE NUMBER: 120

Curve Fitting

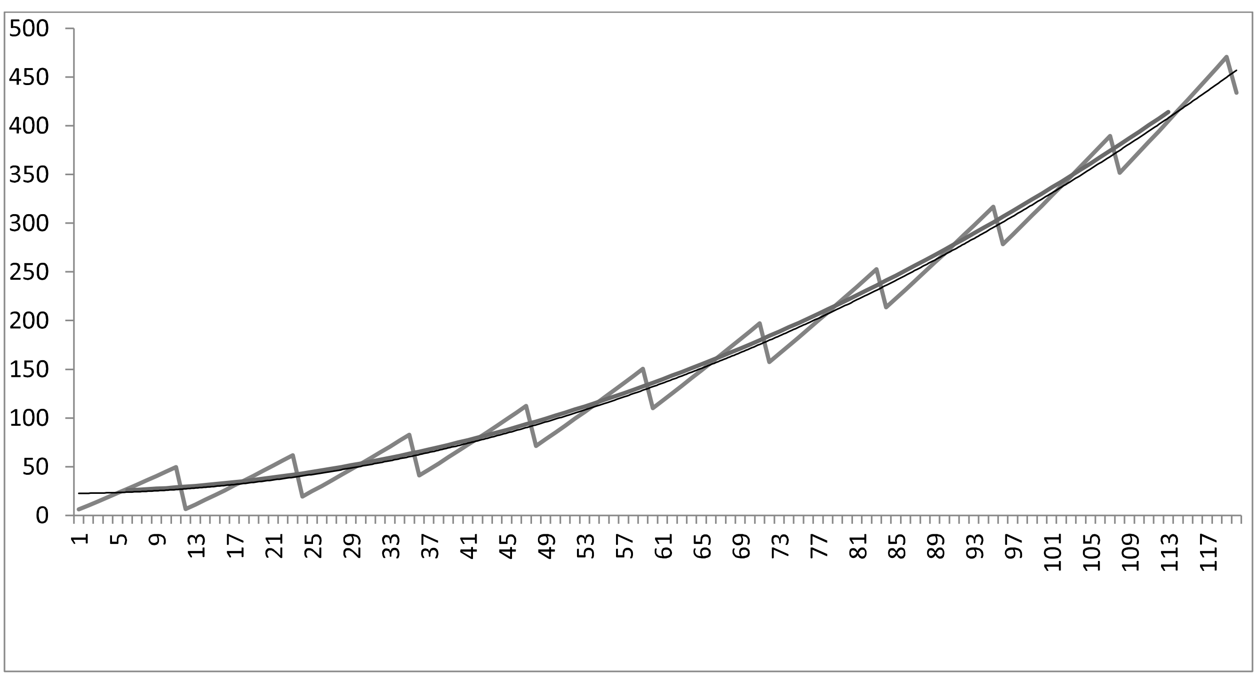


SLIDE NUMBER: 121

Curve Fitting



SLIDE NUMBER: 122 – Curve Fitting



SLIDE NUMBER: 123

SIMPLE EXPONENTIAL SMOOTHING (SES)

- Suppressing short-run fluctuation by smoothing the series
- Weighted averages of all previous values with more weights on recent values
- No trend, No seasonality

SLIDE NUMBER: 124

SIMPLE EXPONENTIAL SMOOTHING (SES)

- **Observed time series**

$Y_1, Y_2, \ldots, Y_n$

- **The equation for the model is**

$$S_t = \alpha Y_{t-1} + (1-\alpha) S_{t-1}$$

Whereα: the smoothing parameter, $0 \leq \alpha \leq 1$

***Y_t*: the value of the observation at time t**

***S_t*: the value of the smoothed obs. at time t.**

SLIDE NUMBER: 125

SIMPLE EXPONENTIAL SMOOTHING (SES)

- The equation can also be written as

$$S_{t+1} = \alpha Y_t + (1-\alpha) S_t$$

$$= S_t + \alpha(Y_t - S_t)$$

$$S_t = S_{t-1} + \underbrace{\alpha(Y_t - S_t)}_{\text{the forecast error}}$$

SLIDE NUMBER: 126

SIMPLE EXPONENTIAL SMOOTHING (SES)

- Why Exponential?: For the observed time series $Y_1, Y_2, \ldots, Y_n$, Y_{n+1} can be expressed as a weighted sum of previous observations.

$$\hat{Y}_t(1) = c_0 Y_t + c_1 Y_{t-1} + c_2 Y_{t-2} + \cdots$$

Where c_i's are the weights.

Giving more weights to the recent observations, we can use the geometric weights (decreasing by a constant ratio for every unit increase in lag).

$$\Rightarrow c_i = \alpha(1-\alpha)^i; i = 0,1,\ldots; 0 \le \alpha \le 1.$$

SLIDE NUMBER: 127

Forecasting strategy

- Forecast each part separately
  - Forecasting of regular components is easier
- Need to decompose/extract different parts
- Steps
  - Separate
  - Forecast
  - Combine

- Cyclic component is often most difficult to handle
- Long-term cyclic movements may come to be clubbed with trend in a decomposition process

SLIDE NUMBER: 128

Decomposition with R

- Splits data into trend, seasonal component and random components
- Only one seasonal component can be handled
- Smoothing is used to extract trend
- Long-term cyclic movements get clubbed with trend
- Short-term cyclic movements get clubbed with random component

SLIDE NUMBER: 129

R code for decomposition of US retail sales data (Case 3-1A)

- For data input

data1<-read.table(file="case3-1a.txt",header=TRUE)

datawide <- ts(data1, frequency=12, start=c(1983,1))

- For additive decomposition

library("TTR")

datacomp <- decompose(datawide)

plot(datacomp)

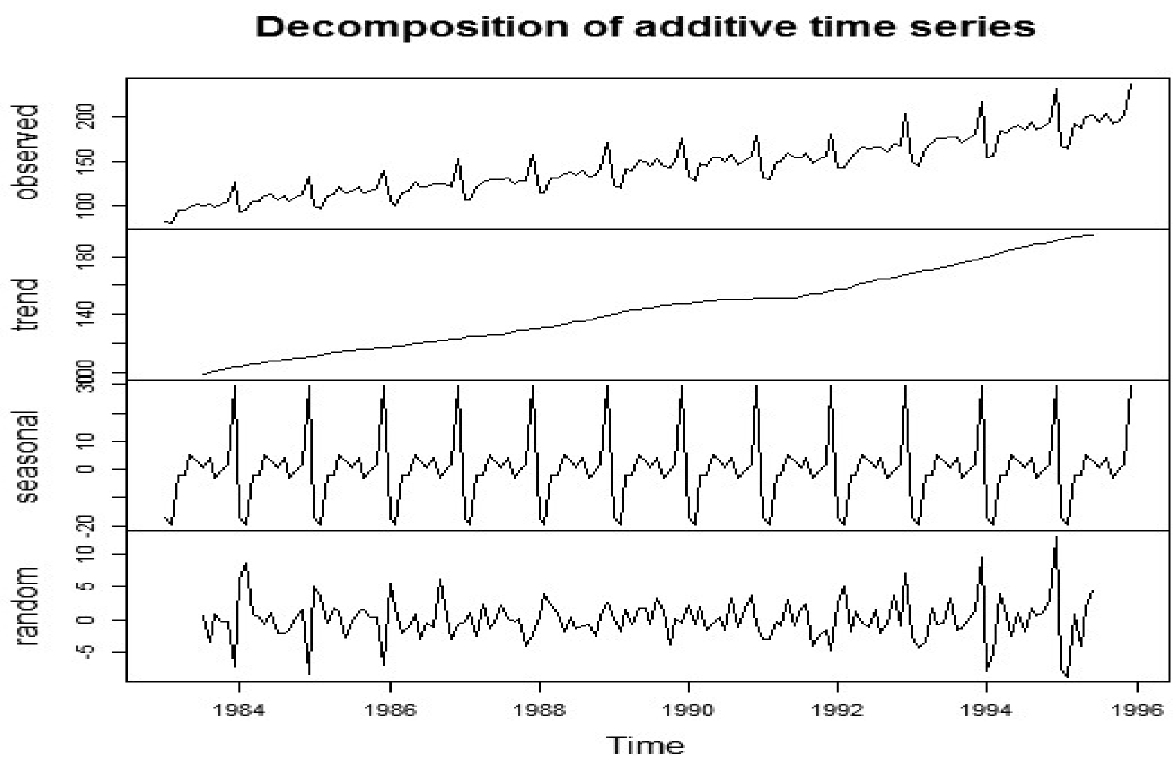
For multiplicative decomposition

datacomp < decompose(datawide,type="multiplicative")

plot(datacomp)

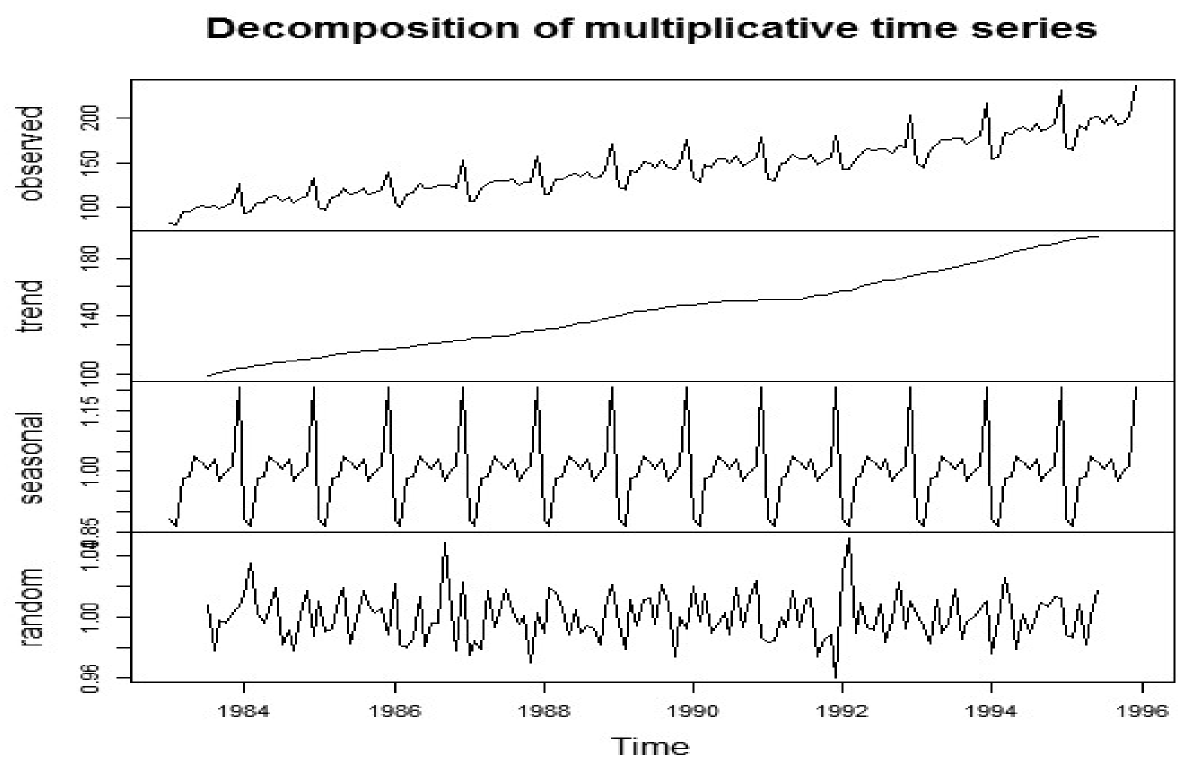
SLIDE NUMBER: 130

Additive decomposition produced by R



SLIDE NUMBER: 131

Multiplicative decomposition produced by R



SLIDE NUMBER: 132

Generating forecasts after Adjusting for trend through differencing

- ‘Difference’ original series

$W_t = Y_t - Y_{t-1}$

- Model it; generate fit/forecast it ($\hat{W}_1$, $\hat{W}_2$,…, $\hat{W}_n$, $\hat{W}_{n+1}$ etc.)
- Forecast original series

$\hat{Y}_t = Y_{t-1} + \hat{W}_t$

- A second round of differencing is also possible (need to be ascertained through time plot)

SLIDE NUMBER: 133

Adjustment for trend through explicit modeling

- Split time series into trend and 'other' parts
  - $Y_t = T_t + W_t$
- Model W_t; Generate 'forecasts' ($\hat{W}_1$, $\hat{W}_2$,…, $\hat{W}_n$, $\hat{W}_{n+1}$ etc.)
- Model T_t; Generate 'forecasts' ($\check{T}_1$, $\check{T}_2$,…, $\check{T}_n$, $\check{T}_{n+1}$ etc.)
- Put everything together
- $\hat{Y}_t = \check{T}_t + \hat{W}_t$
- T_t may be modeled directly from Y_t also.
- For multiplicative situation, use log transform

SLIDE NUMBER: 134

Modeling of trend

- Smooth the original time series
  - Moving average with high k
  - Exponential smoothing with small α
- Fit a curve to the trend component or to the original series
  - Linear curve
  - Quadratic curve
  - Exponential growth curve

SLIDE NUMBER: 135

R code for fitting linear trend to US retail sales data (Case 3-1A)

- For obtaining fit

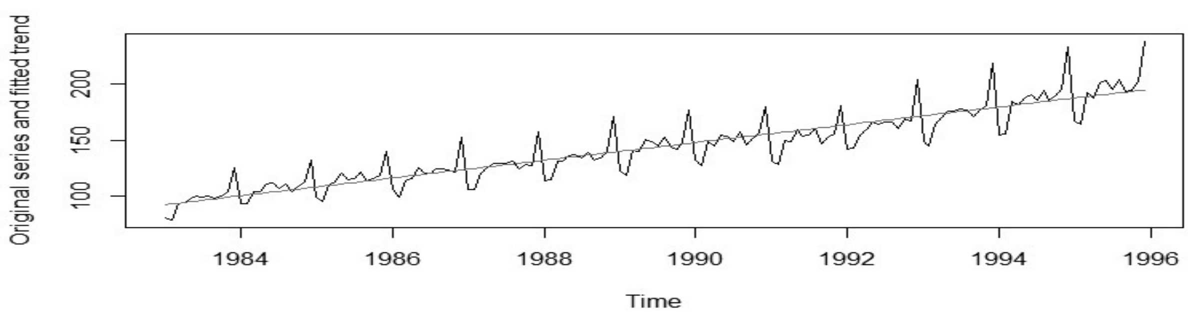
```
time <- 1983 + (0:155)/12

data1<-cbind(data1,time)

olsdata<-lm(Sales~time,data=data1)

residdata<-olsdata$resid

yhatdata<-data1[[1]]-residdata

plot(datawide,ylab="Original series and fitted trend")

lines (time,yhatdata,col="3“)
```

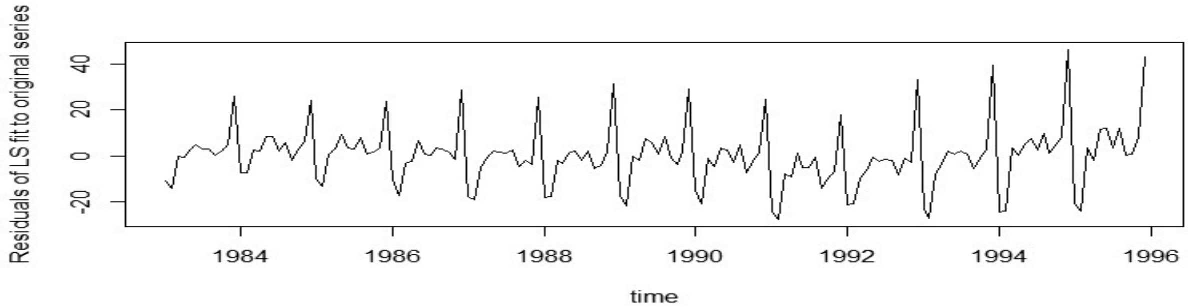
- For plotting residuals

```
plot(time,residdata,type="l", ylab="Residuals of LS fit to original series")
```

SLIDE NUMBER: 136

Line fit to US retail sales data and residuals





SLIDE NUMBER: 137

R code for fitting straight line to trend part of US retail sales data

- For obtaining fit

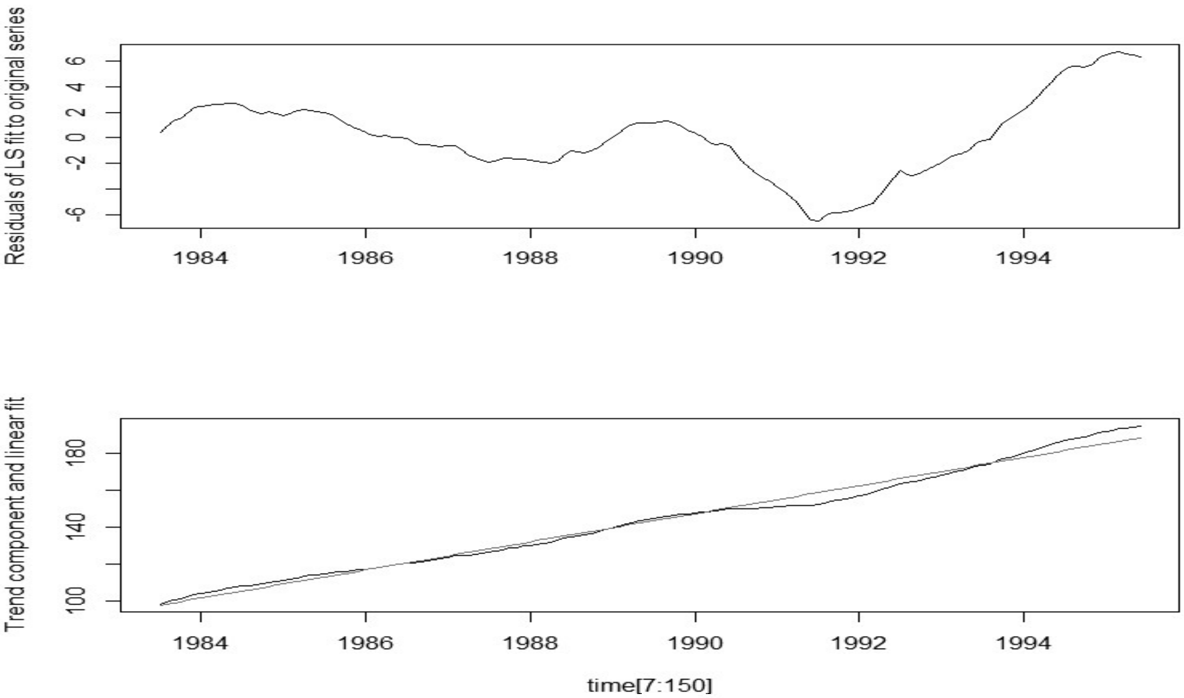
```
olsyhatdata<-lm(datacomp$trend[7:150]~time[7:150])

FittedTrend <- datacomp$trend[7:150] - olsyhatdata$resid

plot(time[7:150],datacomp$trend[7:150], ylab="Trend component and linear fit",type="l")

lines(time[7:150],FittedTrend,col="2")
```

- For plotting residuals

```
plot(time[7:150],olsyhatdata$resid,type="l")
```

SLIDE NUMBER: 138

Line fit to US retail sales trend and residuals



SLIDE NUMBER: 139

Seasonal components

- If the period length is s, identifying the seasonal component amounts to identifying s numbers that would indicate deviation from the mean
- In an additive model, these components would be centered around 0

- In a multiplicative model, these components would be centered around 1
- Methods for identifying seasonal component range from simple averaging to calculations based on several steps of moving averages
- In the multiplicative model, the seasonal multipliers are called *index* numbers

SLIDE NUMBER: 140

Smoothing methods

Smoothing can be used for

- Understanding the ups and downs in a time series, disregarding random fluctuations
- Separating 'trend' from other components
- Forecasting

There are two major approaches to smoothing

- Moving averages
- Exponential smoothing

SLIDE NUMBER: 141

Moving average models

- Naïve model

$\hat{Y}_{n+1} = Y_n$

- Naïve trend model

$\hat{Y}_{n+1} = Y_n + (Y_n - Y_{n-1})$

- Simple average model

$\hat{Y}_{n+1} = (Y_1 + Y_2 + \ldots + Y_n)/n$

- Moving average over *k* time periods

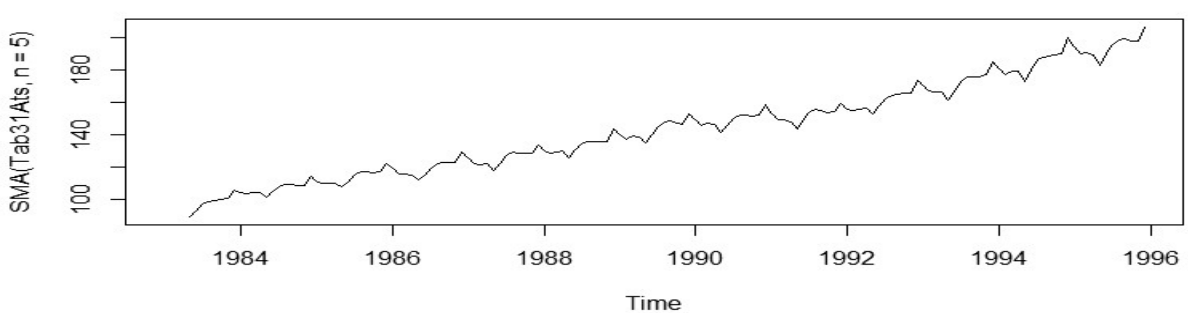
$\hat{Y}_{n+1} = (Y_n + Y_{n-1} + \ldots + Y_{n-k+1})/k$

SLIDE NUMBER: 142

- Moving average smoothing with R R code

```
plot(time,SMA(data1[[1]],n=12),type="l")
lines(time,data1[[1]],col="2")
```

- Output



SLIDE NUMBER: 143

Exponential smoothing

- Simple exponential smoothing

$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha)\ \hat{Y}_t$

- Holt's version: includes a trend component

$L_t = \alpha Y_t + (1 - \alpha)\ (L_{t-1} + T_{t-1})$

$T_t = \beta\ (L_t - L_{t-1}) + (1 - \beta)\ T_{t-1}$

$\hat{Y}_{t+p} = L_t + pT_t$

L_t : Local (current) level

T_t : Local (current) trend

α : Smoothing parameters for level

β : Smoothing parameters for trend

SLIDE NUMBER: 144

Exponential smoothing (contd.)

- Winter's modification: seasonal component

$L_t = \alpha\ (Y_t / S_{t-s}) + (1 - \alpha)\ (L_{t-1} + T_{t-1})$

$T_t = \beta\ (L_t - L_{t-1}) + (1 - \beta)\ T_{t-1}$

$T_t = \gamma\ (Y_t / L_t) + (1 - \gamma)\ S_{t-s}$

$\hat{Y}_{t+p} = L_t + pT_t$

L_t : Local (current) level

T_t : Local (current) trend

S_t : Local (current) seasonal component

α :: Smoothing parameters for level

β :: Smoothing parameters for trend

γ :: Smoothing parameters for seasonal component (can ::have s separate values for different seasons)

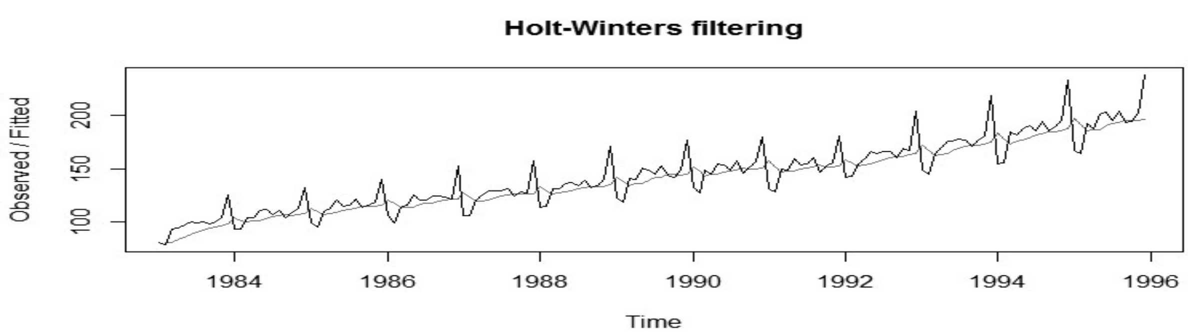
SLIDE NUMBER: 145

Exponential smoothing with R

- R code

plot(HoltWinters(data1[[1]],beta=False gamma=FALSE))

- Output



SLIDE NUMBER: 146

Modeling the random component

A time series is said to be stationary if

- It has a constant mean
- Autocorrelation (ACF) at lag *k* is the same at all parts of the series

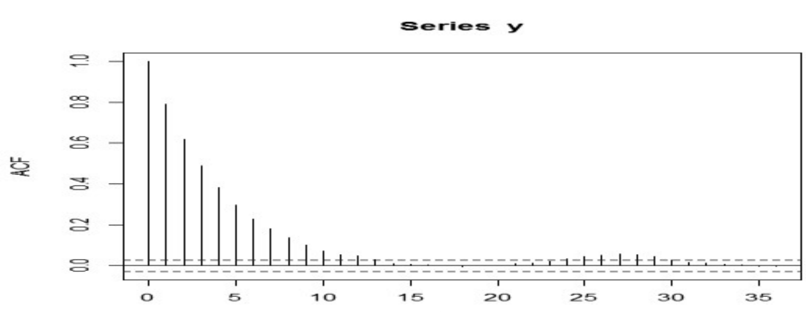
Basic models for stationary time series

- Autoregressive (AR) model
- Moving average (MA) model
- Autoregressive Moving Average (ARMA) model

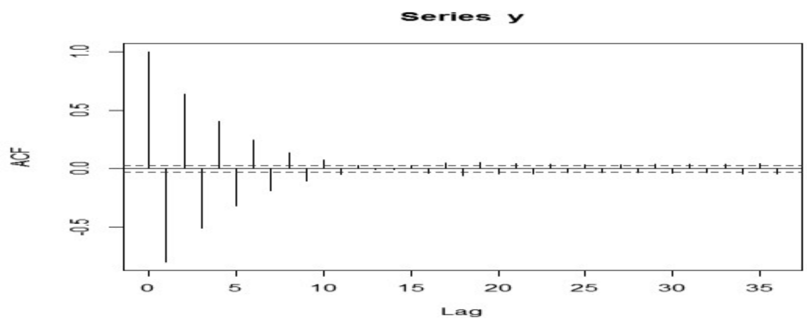
SLIDE NUMBER: 147

- **AR(1) model**
- $Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varepsilon_t$, ε_t **white noise**

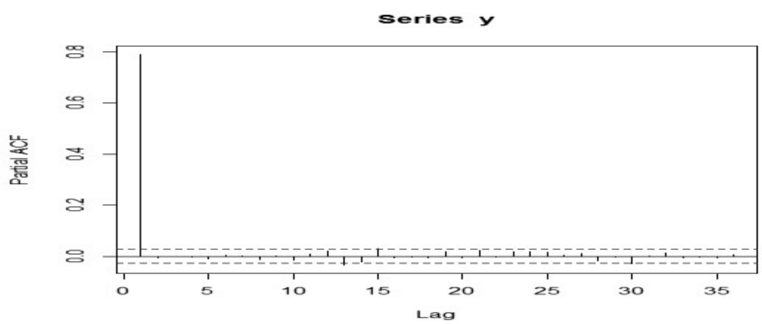
$\varphi_1 = 0.8$ **ACF plots**

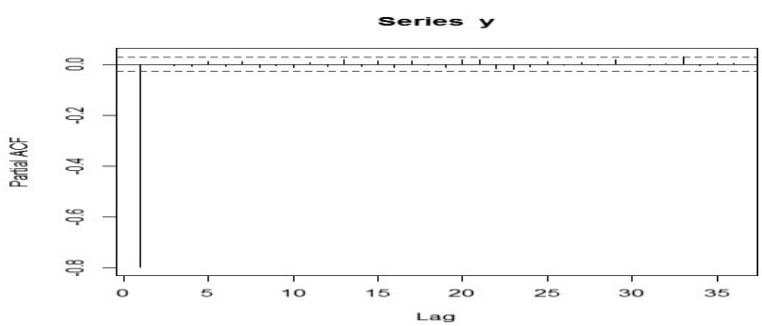


φ_1 = – 0.8



PACF(Partial ACF) Plot





SLIDE NUMBER: 148

AR(p) model

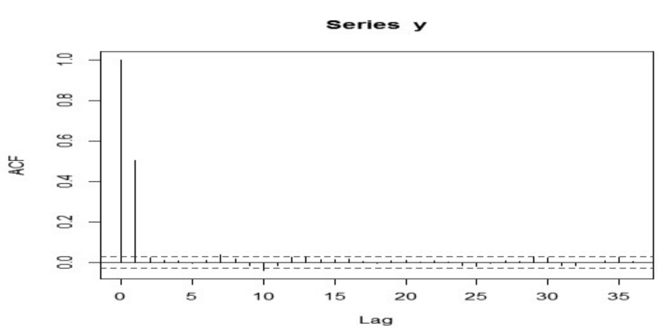
- $Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \varepsilon_t$, ε_t white noise
- Such a model has non-zero ACF at all lags
- However, only the first p PACFs are non-zero; the rest are zero
- If PACF plot shows large PACFs only at a few lags, then AR model is appropriate
- If an AR model is to be fitted, the parameters φ_0, φ_1, φ_2,…, φ_p have to be estimated from the data, under the restriction that the estimated values should guarantee a stationary process

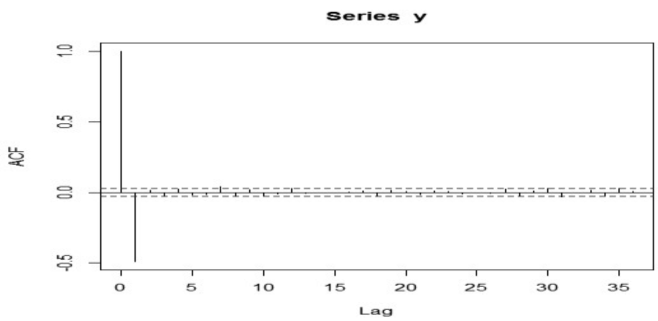
SLIDE NUMBER: 149

MA(1) model

- $\boldsymbol{Y_t = \theta_0 + \varepsilon_t + \theta_1\ \varepsilon_{t-1}}$ **,** $\boldsymbol{\varepsilon_t}$ **white noise**

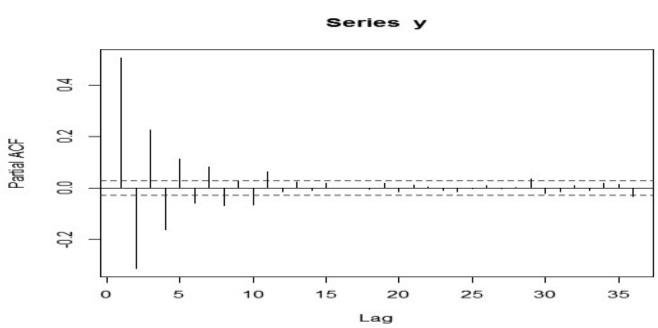
ACF Plots $\vartheta_1 = 0.8$

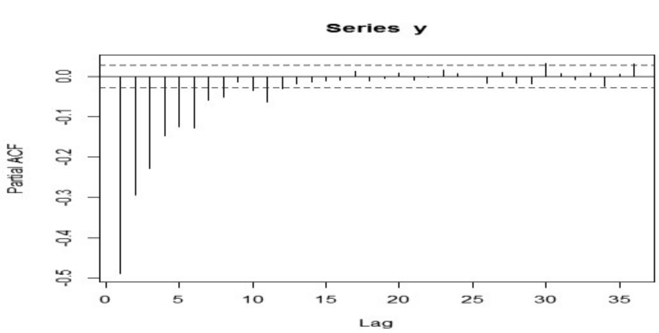




PACF Plots

$$\vartheta_1 = -0.8$$





SLIDE NUMBER: 150

MA(q) model

- $Y_t = \vartheta_0 + \varepsilon_t + \vartheta_1\, \varepsilon_{t-1} + \vartheta_2\, \varepsilon_{t-2} + \ldots + \vartheta_q\, \varepsilon_{t-q}$, ε_t white noise
- Such a model has non-zero PACF at all lags
- However, only the first q ACFs are non-zero; the rest are zero
- If ACF plot shows large ACFs only at a few lags, then MA model is appropriate
- If an MA model is to be fitted, the parameters $\vartheta_0, \vartheta_1, \vartheta_2, \ldots, \vartheta_q$ have to be estimated from the data

SLIDE NUMBER: 151

ARMA(p,q) model

- $Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \ldots + \varphi_p Y_{t-p} + \varepsilon_t + \vartheta_1\, \varepsilon_{t-1} + \vartheta_2\, \varepsilon_{t-2} + \ldots + \vartheta_q\, \varepsilon_{t-q}$, ε_t white noise
- Such a model has non-zero ACF and non-zero PACF at all lags
- If an ARMA(p,q) model is to be fitted, the parameters φ_0, φ_1, $\varphi_2, \ldots, \varphi_p$, ϑ_1, $\vartheta_2, \ldots, \vartheta_q$ have to be estimated from the data, under the restriction that the estimated values produce a stationary process
- AR(p) is ARMA(p,0)
- MA(q) is ARMA(0,q)

SLIDE NUMBER: 152

ARIMA(p,d,q) model

- If d-times differenced series is ARMA(p,q), then original series is said to be ARIMA(p,d,q).
- ARIMA stands for 'Autoregressive Integrated Moving average'.
- If W_t is the differenced version of Y_t, i.e., $W_t = Y_t - Y_{t-1}$,

  then Y_t can be written as

  $$Y_t = W_t + W_{t-1} + W_{t-2} + W_{t-3} + \dots .$$

  Thus, the series Y_t is an 'integrated' (opposite of 'differenced') version of the series W_t.
- If Y_t is ARIMA(p,d,q), it is non-stationary.
- However, its d-times differenced version, an ARMA(p,q) process, can be stationary.

SLIDE NUMBER: 153

Box-Jenkins ARIMA model-building

- Model identification
  - If the time plot 'looks' non-stationary, difference it until the plot looks stationary
  - Look at ACF and PACF plots for possible clue on model order (p, q)
  - When in doubt (regarding choice of p and q), use the principle of *parsimony*: A simple model is better than a complex model
- Estimate model parameters
- Check residuals for health of model

- Iterate if necessary
- Forecast using the fitted model

SLIDE NUMBER: 154

THANK YOU

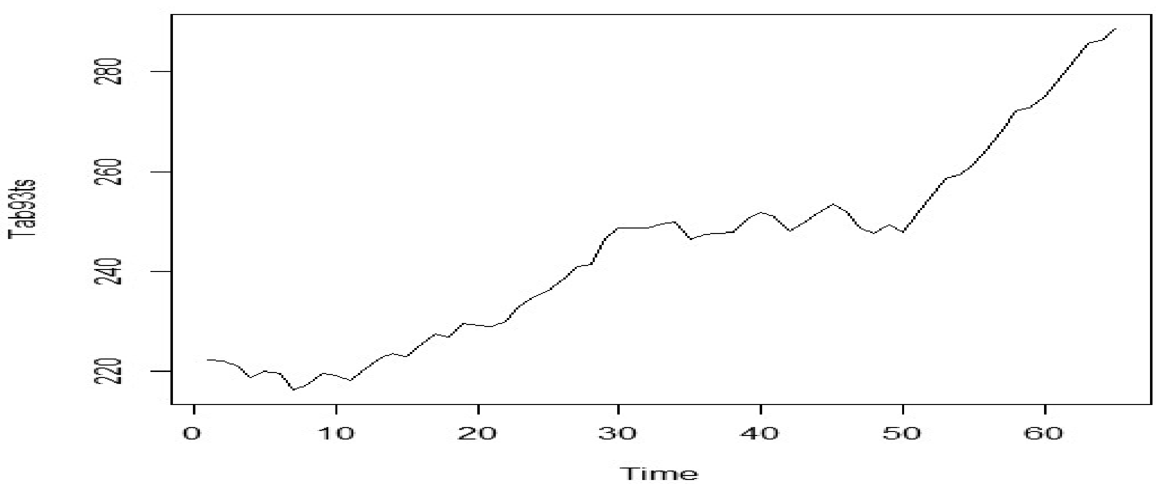
SLIDE NUMBER: 155

Appendix

- Read for additional information & R code

SLIDE NUMBER: 156

Example: Dow-Jones transportation index, daily closing averages (Tab9-3)

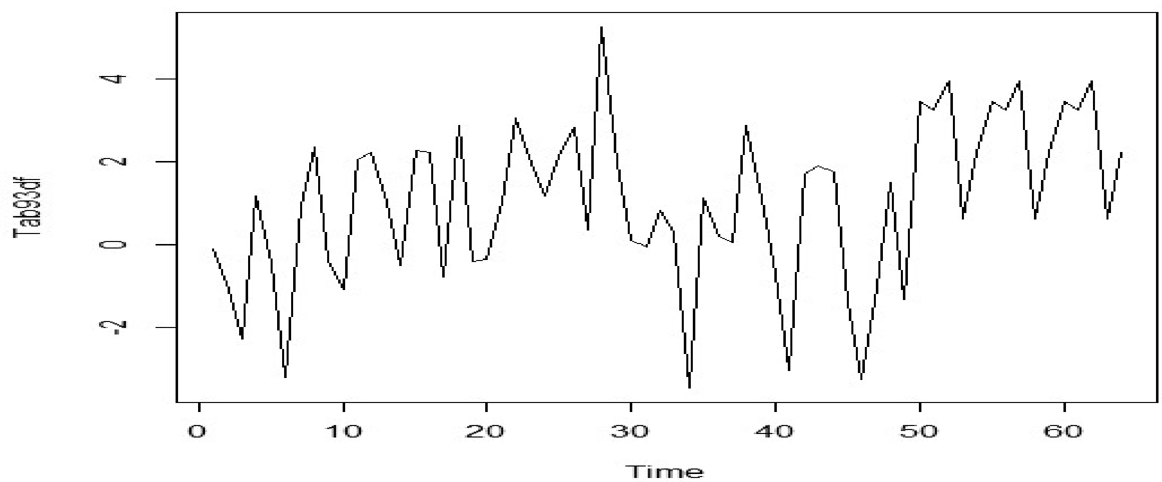


Plot ‘looks’ non-stationary (mean is not stable)

Don’t bother to plot ACF or PACF; try differencing

SLIDE NUMBER: 157

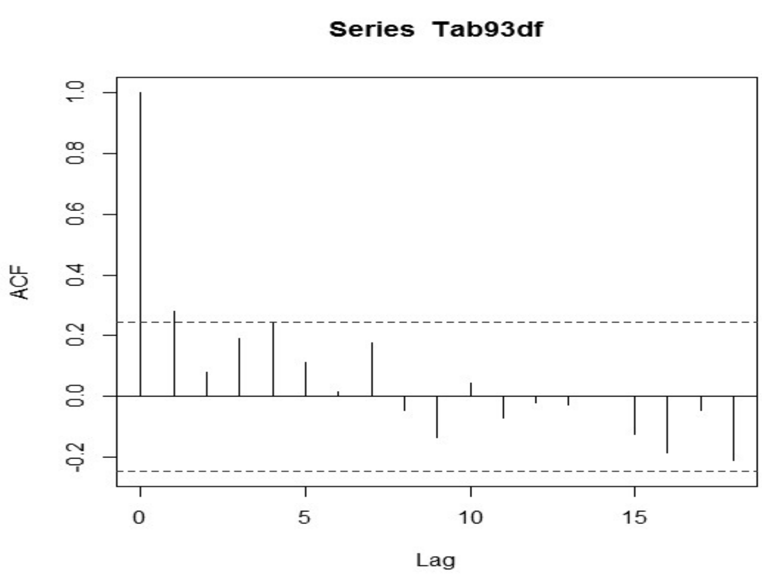
Time plot of transportation index data, differenced once

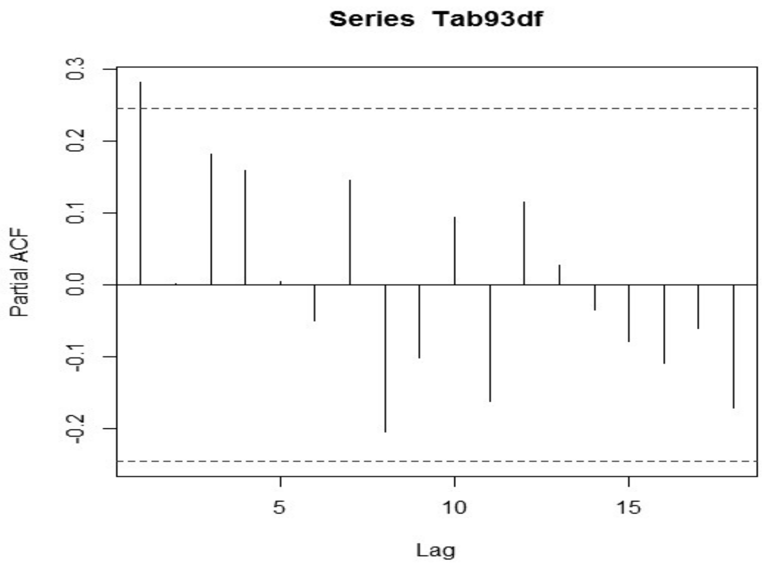


- Looks more stationary, settle for this for now
- Plot ACF/PACF

SLIDE NUMBER: 158

ACF/PACF plot of transportation index data, differenced once

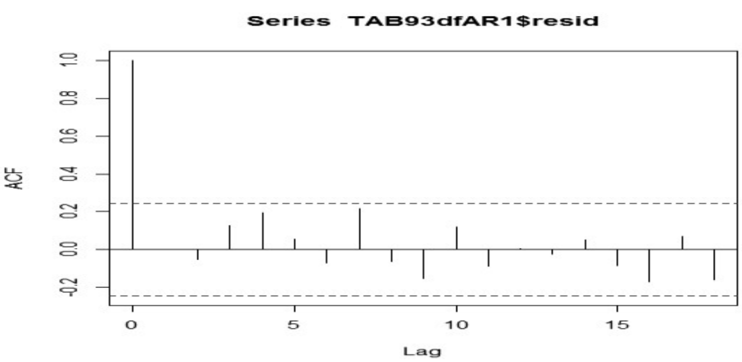


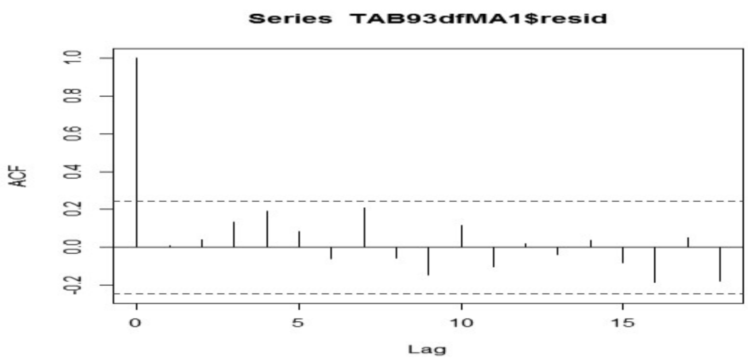


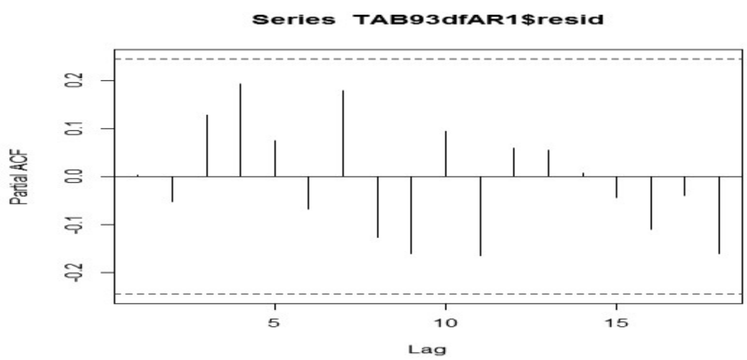
- Only one significant ACF, suggesting MA(1) model
- Only one significant PACF, suggesting AR(1) model

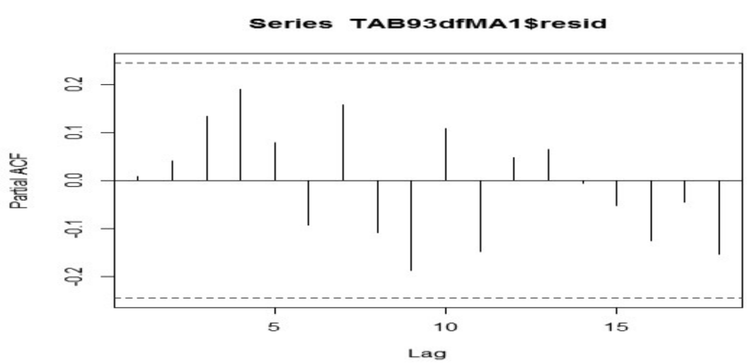
SLIDE NUMBER: 159

Transportation index data: ARIMA(1,1,0) and ARIMA(0,1,1) model fitting









SLIDE NUMBER: 160

Transportation index data: ARIMA(1,1,0) and ARIMA(0,1,1) model fitting (contd.)

Read data

ts<-read.table(file="H:\\tab9-3.txt",header=TRUE)

Differencing the data

dts=ts[[1]][2:65]-ts[[1]][1:64]

One line R code, assuming differenced data in Tab93df:

```
r1<-arima(dts,order=c(1,0,0),method="ML")
```

AR(1) model

$$W_t = \underset{(0.319)}{1.035} + \underset{(0.119)}{0.280}W_{t-1} + \varepsilon_t$$

Estimated variance of ε_t is 3.43

AIC = 266.5

MA(1) model

$$W_t = \underset{(0.297)}{1.038} + \varepsilon_t + \underset{(0.120)}{0.287}\,\varepsilon_{t-1}$$

Estimated variance of ε_t is 3.43

AIC = 266.6

Akaike Information Criterion is used to compare parametric models (smaller AIC is better)

The models produce comparable results

SLIDE NUMBER: 161

Transportation index data: ARIMA(1,1,0) and ARIMA(0,1,1) forecasts

R code for forecasting

```
predict(r1,n.ahead=1)
```

R code for forecasting

predict(r1,n.ahead=1)

MA(1) model

W_{66} = 1.474

(1.852)

Y_{66} = 288.57 + 1.474

= 290.04

(1.852)

95% prediction interval: 290.04 ± 3.70

SLIDE NUMBER: 162

Transportation index data (continued)

- If constant term in differenced series is not used, differencing, model fitting and forecast generation can be combined into a single step

predict(arima(ts,order=c(1,1,0)), n.ahead=1)

predict(arima(ts,order=c(0,1,1)), n.ahead=1)

SLIDE NUMBER: 163

Revisit US retail sales data (random component)

- Code to retrieve the random component

```
data1<-read.table(file="case3-1a.txt",header=TRUE)

datawide <- ts(data1, frequency=12, start=c(1983,1))

datacomp <- decompose(datawide,type="multiplicative")
```

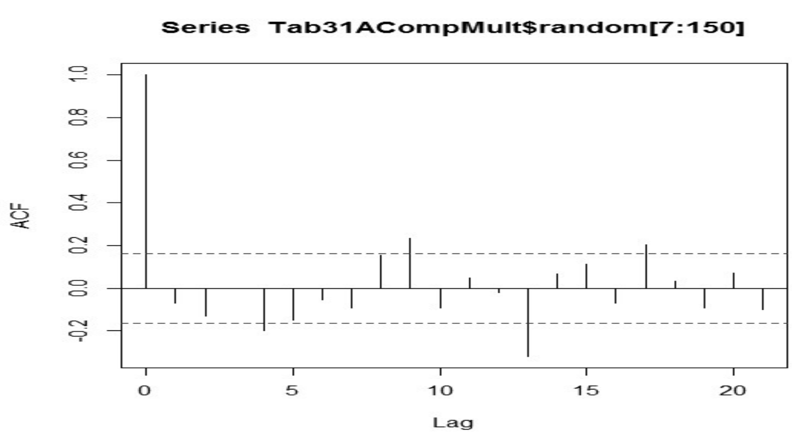
- Code to fit an ARMA model automatically

```
library("forecast")

auto.arima(datacomp$random)
```

SLIDE NUMBER: 164

Analysis of US retail sales data (random component)

- Why does this happen?
- Correlogram is not benign



- Even Ljung-Box test (code given below) indicates that there are some non-zero ACFs

```
Box.test(datacomp$random[7:150], lag=20, type="Ljung-Box")
```

SLIDE NUMBER: 165

Analysis of US retail sales data (random component)

- There is no clear way to use the information in the random component.
- Exponential smoothing:

```
plot(HoltWinters(datacomp$random[7:150],beta=FALSE,
gamma=FALSE))
```

produces a alpha = 0.03 (a lot of smoothing).

- Forecast contribution, obtained by the code:

```
predict(HoltWinters(datacomp$random[7:150],
beta=FALSE, gamma=FALSE),n.ahead=1)
```

is negligible (a factor of 1.0003).

- Eventual forecast has to rely on the trend and seasonal components only.

SLIDE NUMBER: 166

Building seasonality into ARIMA models

- Consider monthly data with annual seasonality

- The seasonal phenomena may not be restricted to a fixed periodic pattern
- In fact, anything that happens in a non-seasonal series (governed by UNIT LAG) can happen in a seasonal series at a larger time scale (LAG 12).
- Such happenings can be represented by an ARIMA-like model that operates as lag 12.
- Call it ARIMA(P,D,Q).

SLIDE NUMBER: 167

A combined ARIMA model

- A combined model would have two levels:
  - one ARIMA(p,d,q) for lag 1,
  - another ARIMA(P,D,Q) for lag 12.
- The two layers can co-exist.
- It does not matter which one comes first, the model happens to be the same either way.
- Such a model is called a seasonal ARIMA model.
- It can even be implemented in R.

SLIDE NUMBER: 168

Example: Seasonal ARIMA modeling of US retail sales data

- Read data

```
ts<-read.table(file="H:\\Case3-1A.txt",header=TRUE)
ats <- ts(ts, frequency=12, start=c(1983,1))
```

- Code to automatically fit seasonal ARIMA model

```
library("forecast")
auto.arima(log(ats))
```

- Produces ARIMA(1,1,3)(2,1,1)$_{12}$ model, but has computational problem
- Code to fit ARIMA(1,1,3)(2,1,1)$_{12}$ model

```
atsfit <-
arima(log(ats),order=c(1,1,3),seasonal=list(order=c(2,1,1),perio
d=12))
summary(atsfit)
```

- Produces appropriate estimates and diagnostic information

SLIDE NUMBER: 169

Forecast of US retail sales from Seasonal ARIMA modeling

- Code to generate ARIMA forecast

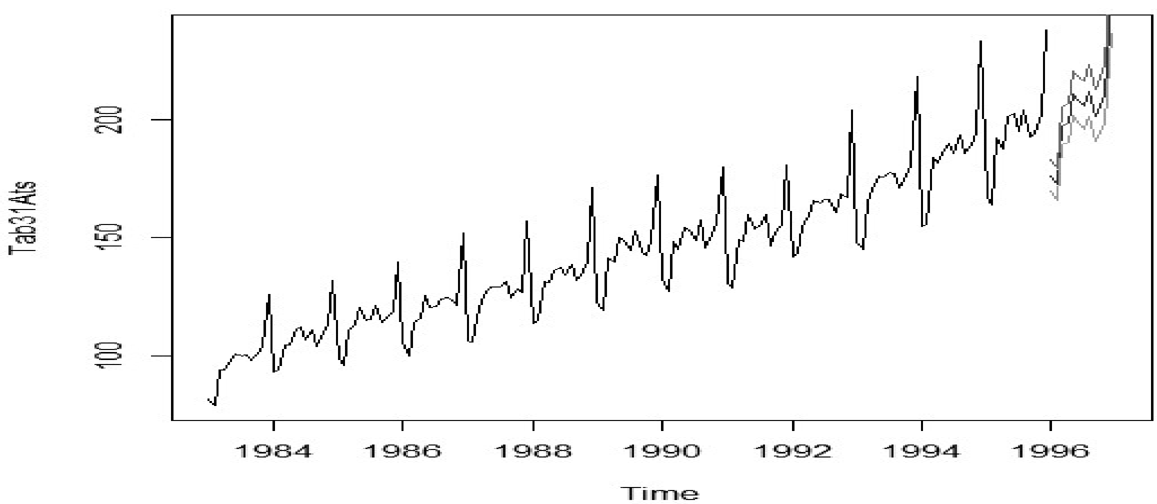
```
atsforecast <- predict(atsfit, n.ahead=12)
lpl <- atsforecast$pred-2*atsforecast$se
upl <- atsforecast$pred+2*atsforecast$se
cbind(exp(atsforecast$pred), exp(lpl), exp(upl))
```

- Produces forecasts and prediction intervals
- Additional code to generate plots

```
plot.ts(ats,xlim=c(1983,1997))

lines((1996+(0:11)/12),exp(atsforecast$pred),col=4)

lines((1996+(0:11)/12),exp(lpl),col=3)

lines((1996+(0:11)/12),exp(upl),col=2)
```

SLIDE NUMBER: 170

Forecast of US retail sales from Seasonal ARIMA modeling



SLIDE NUMBER: 171

Revisit decomposition approach: How to recombine components?

- In an additive model
    - add the forecasts of the components

  - add the variances of the components (generally an approximate but conservative approach)
- In a multiplicative model, the same strategy can be taken in log-scale
- Anti-log or exponential function brings back the forecasts to original scale
- Upper and lower prediction limits can be transformed similarly; they would continue to have the same coverage probability.

SLIDE NUMBER: 172

Importance of error estimates

- In order to provide an estimate of the overall forecast error, or to provide prediction intervals, error estimates for each component is needed.
- Simple decomposition may provide an idea of what to expect, but it does not give error estimates.
- Attempt a fresh decomposition, while looking for error estimates all the way.
- One route will be illustrated; not unique.

SLIDE NUMBER: 173

Step 1: Fit trend model, generate forecast & error estimates from regression model

- Code for this purpose (working with log series)

```
time <- 1983 + (0:155)/12

logats <- log(ats)

ats.df <- as.data.frame(cbind(time, logats))

data.new <- data.frame(time = c(1996+(0:11)/12))

logtrend <- predict(lm(logats~time), newdata=data.new,
se.fit=TRUE)

logtrendfit <- logats - lm(logats~time)$resid
```

- logtrend has the prediction information
- logtrendfit has the fitted curve within the time span of available data

SLIDE NUMBER: 174

Step 2: Get seasonal component with error estimates by averaging de-trended series

- Code for this purpose

```
atsdetrended <- logats - logtrendfit

atsmat <- matrix(atsdetrended,ncol=12,byrow=T)

logseason <- colMeans(atsmat)

logseason.var <- rep(0,12)
```

```
for (i in 1:12) logseason.var[i] <- var(atsmat[,i])/(156/12)
```

- logseason has the seasonal estimate
- logseason.var has the estimated variance

SLIDE NUMBER: 175

Step 3: Get forecast of random part, with error estimates, by ARIMA modeling

- Code for obtaining random component and modeling

```
atsrandom <- atsdetrended - rep(logseason,13)
auto.arima(atsrandom)
```

- This produces ARIMA(0,1,1) model, but has computational problem
- Use arima directly with specified model order:

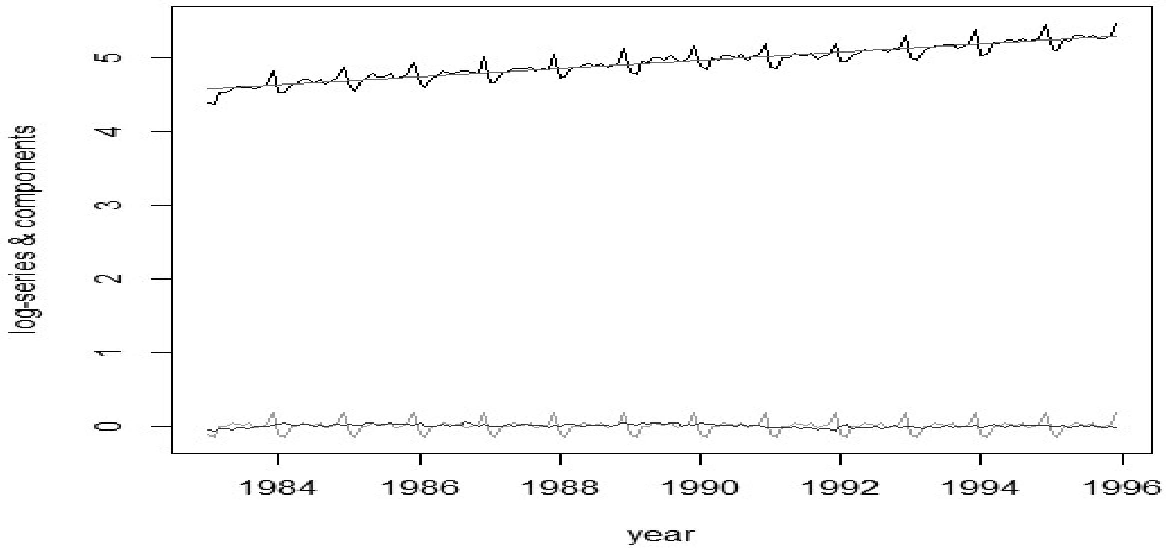
```
atsranfit <- arima(atsrandom,order=c(0,1,1))
atsranforecast <- predict(atsranfit, n.ahead=12)
```

- atsranforecast has prediction information
- Plotting the series:

```
plot.ts(logats,xlim=c(1983,1997),ylim=c(0,6))
lines((1983+(0:155)/12),logtrendfit,col=4)
lines((1983+(0:155)/12),atsdetrended,col=3)
lines((1983+(0:155)/12),atsrandom,col=2)
```

SLIDE NUMBER: 176

The original log-series and the separated components



SLIDE NUMBER: 177

Step 4: Re-combine forecasts, add variances, generate prediction intervals

- Code for this purpose

```
atslogcombpred <- logtrend$fit + logseason + atsranforecast$pred

atslogcombse <- sqrt((logtrend$se.fit)^2 + logseason.var + (atsranforecast$se)^2)

lfl <- atslogcombpred -2* atslogcombse

ufl <- atslogcombpred +2* atslogcombse
```

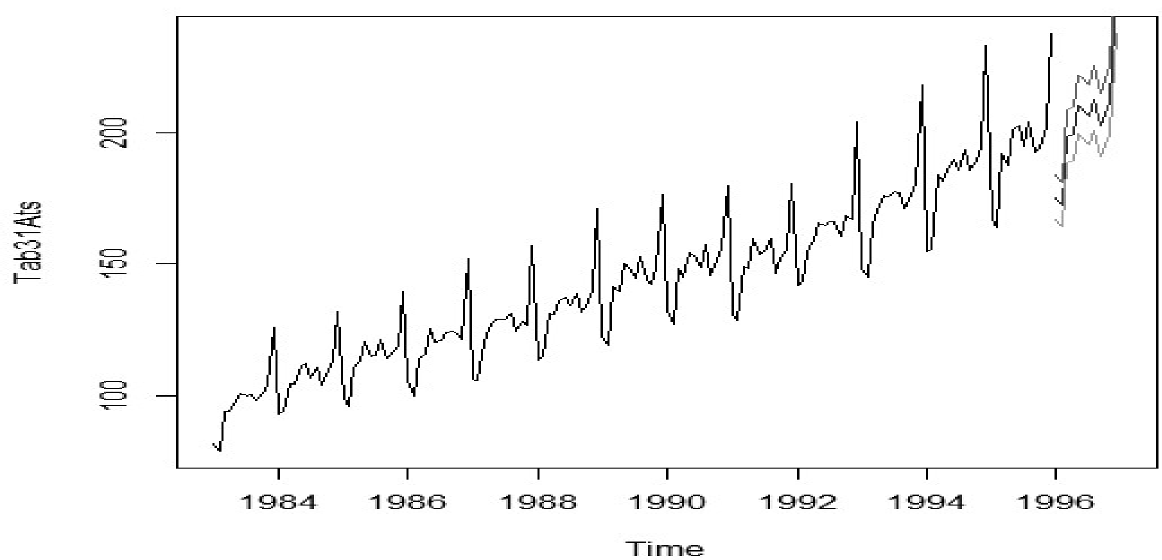
```
cbind(exp(atslogcombpred), exp(lfl), exp(ufl))
```

- Code for producing forecast plot

```
plot.ts(ats,xlim=c(1983,1997))
lines((1996+(0:11)/12),exp(atslogcombpred),col=4)
lines((1996+(0:11)/12),exp(lfl),col=3)
lines((1996+(0:11)/12),exp(ufl),col=2)
```

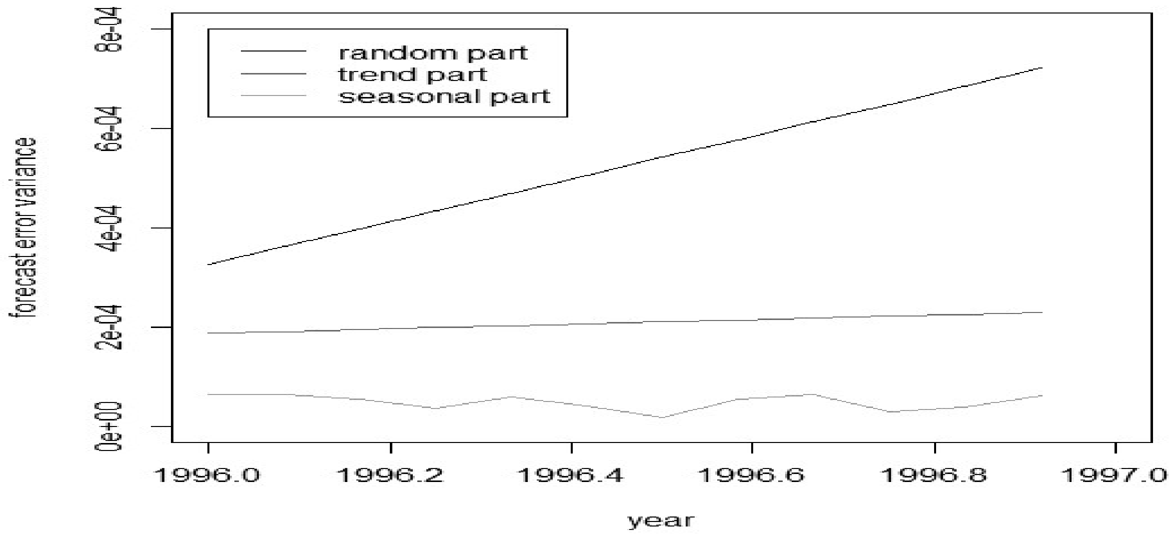
SLIDE NUMBER: 178

The original series, one-year ahead forecast and prediction intervals



SLIDE NUMBER: 179

Comparison of different components of forecast error variance



SLIDE NUMBER: 180

Forecast from generalized exponential smoothing

- Code for this purpose

```
atsHWfit <- HoltWinters(ats, beta=TRUE, gamma=TRUE)
lhwl <- (forecast(atsHWfit, h=12)$lower)[,2]
uhwl <- (forecast(atsHWfit, h=12)$upper)[,2]
cbind(forecast(atsHWfit, h=12)$mean, lhwl, uhwl)
```

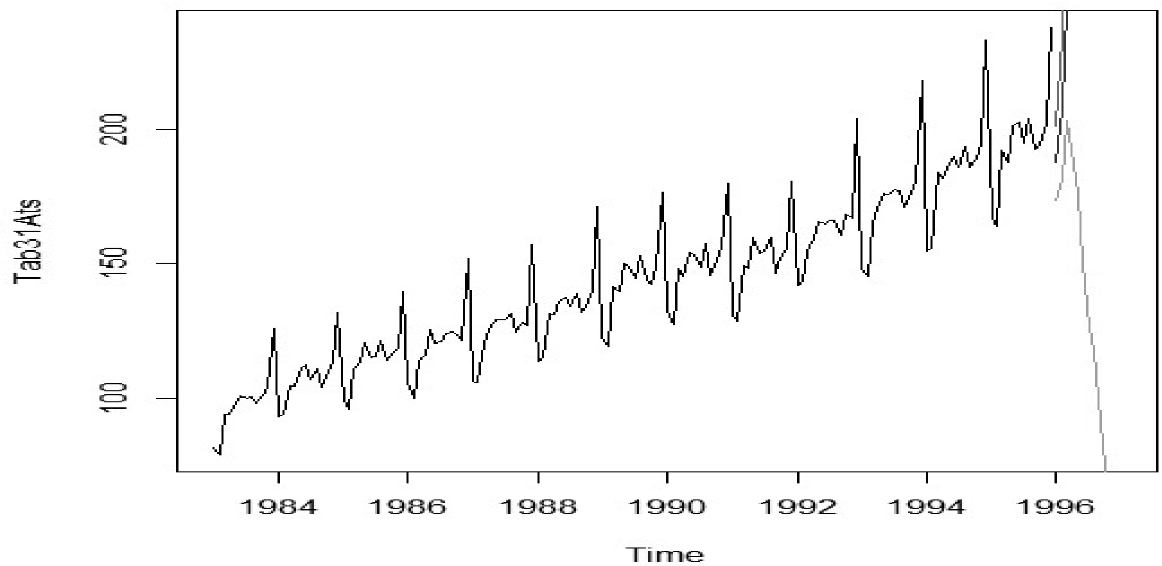
- Code for producing forecast plot

```
plot.ts(ats,xlim=c(1983,1997))
lines((1996+(0:11)/12),forecast(atsHWfit, h=12)$mean,col=4)
lines((1996+(0:11)/12),lhwl,col=3)
```

```
lines((1996+(0:11)/12),uhwl,col=2)
```

SLIDE NUMBER: 181

Forecast from generalized exponential smoothing



SLIDE NUMBER: 182

Thank you