Time Series Forecasting

Day 1

Time Series Forecasting

- Prediction is the most important technique in data analysis for decision making.
- Several statistical and data mining methods have been developed to predict an unknown value of a variable based available knowledge base.
- Regression and data mining techniques like CART, ANN etc. are some examples of predictive methods.
- In regression, the value of an unknown response y is predicted based on some known regressors (Xs).
- A practical application is predicting the possibility of loan default by examining an applicant's demographic and economic features.

Cross Sectional data

- In the bank loan example, the historical data for the regression model was all collected at the same instant of time.
- It is important that this is so. If the data were collected over period of time, the regression possibly would not work as the data is not coherent.
- Such data which is collected all at a time is called cross sectional data.
- For cross sectional data, order of observations does not matter.

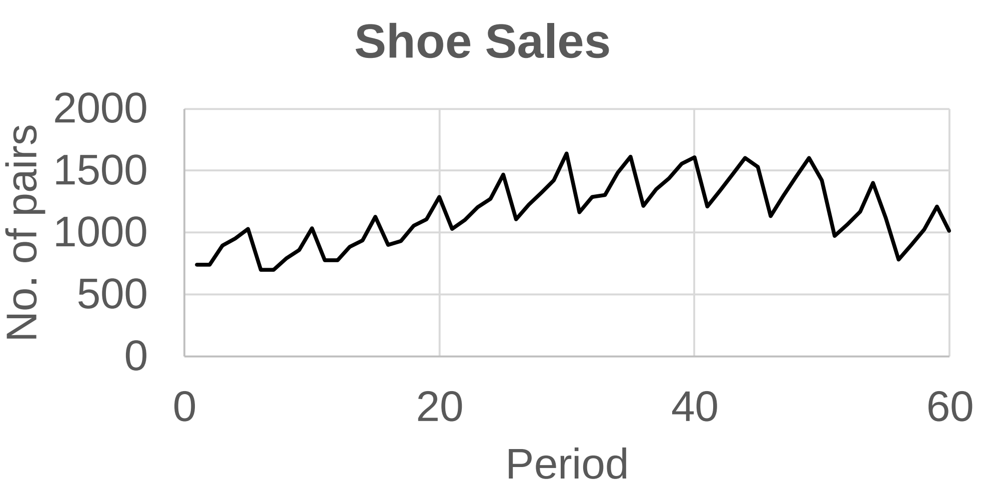
model	mpg	hp	wt	gear
Mazda RX4	21	110	2.62	4
Datsun 710	22.8	93	2.32	4
Merc 240D	24.4	62	3.19	4
Cadillac Fleetwood	10.4	205	5.25	3
Honda Civic	30.4	52	1.615	4
Toyota Corolla	33.9	65	1.835	4

model	mpg	hp	wt	gear
Cadillac Fleetwood	10.4	205	5.25	3
Merc 240D	24.4	62	3.19	4
Honda Civic	30.4	52	1.615	4
Toyota Corolla	33.9	65	1.835	4
Mazda RX4	21	110	2.62	4
Datsun 710	22.8	93	2.32	4

Time Series data

- In contrast, look at the following data.
- The data is collected over time.
- There is only one variable, sales. This variable is ordered in ascending sequence of time in months.
- In time series data the present value of the variable is dependent on its value in the previous periods.
- The ordering of observations is very important.

Year	Month	No. of Pairs
2011	Jan	742
2011	Feb	741
2011	Mar	896
…	…	…
2015	Oct	1023
2015	Nov	1209
2015	Dec	1013



Characteristics of time series data

- There is only one variable which is both input and response variable.
- The variable is a continuous variable.
- Each observation belongs to different time interval which is regular and same through out.
- The observations are in a time sequence.
- The ordering of observations is very important.
- In between periods can not be missed.
- Typical periods are yearly, quarterly, monthly, weekly, daily, hourly.
- Treatment of outliers, missing values is different in cross-sectional and time series data.

Applications of Time Series Forecasting

Function-wise

- Financial – Forecasting cash requirements, interest rates etc.
- Marketing – Demand Forecasting.
- Production – Capacity planning, Production scheduling.
- Investment – Stock Market and Bond rate predictions.
- HR – Manpower planning.

Industry-wise

- Airline industry – Ticket (surge) pricing.
- Government Policy – GDP, infrastructure and citizen welfare planning.
- Retail – Inventory planning, Staffing plans.

Time Series Forecasting – Time Horizon

Airline Industry Example

- Strategic Planning – 5 to 10 years – New fleet acquisition.
- Tactical Planning – 1 to 3 years – New routes, staffing plans.
- Operational Plans - < 1 year – Ticket pricing, maintenance, onboard catering.

Time Series Analysis Considerations

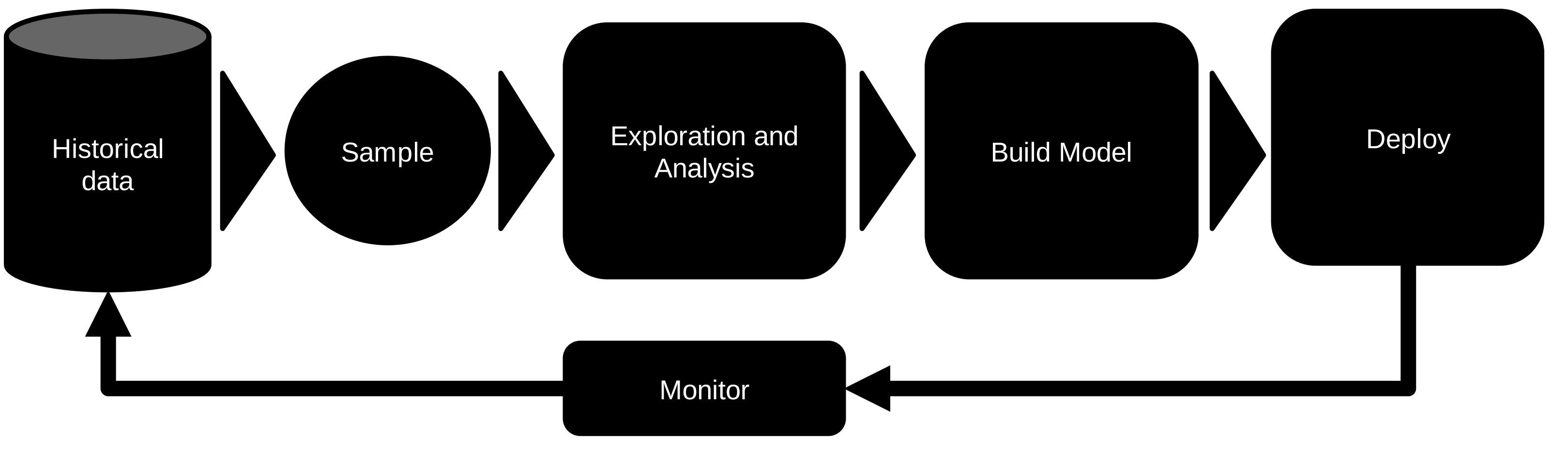
- How much data is available?
- How truly does past data reflect future scenarios?
- Stability of past data, especially sudden shocks and level jumps.
- How well are we able to understand data?
- Is the environment volatile with frequent, drastic changes?
- Is forecast self-correcting i.e. does it itself interfere with future values?
- Required length of forecast; long, medium and short term.

Overview of this module

- Forecasting is an important tool in the hands of investors in future to manage risk.
- It is the art and science of predicting future values of a variable using data-based methods.
- The main methods discussed in this course are
  - Regression Analysis.
  - Decomposition.
  - Smoothing.
  - Auto-Regressive Moving Average Methods.
- Advantages – Fact-based, well developed tools, software support, objective and accurate if used properly.
- Warnings – Lack of clear thinking, questionable data, no management oversight, improper use of techniques.

Forecasting Steps

- Problem formulation.
- Data collection.
- Model building and assessment.
- Model implementation.
- Forecasting, monitoring and continuous evaluation.



Time Series Patterns

Components of Time Series data

- Time Series is characterized by patterns in distribution of data.
- These patterns are called components of time series . There are mainly of two types of components, namely systematic and random.
- Time Series components.
  - Trend
  - Seasonal
  - Cyclical
  - Irregular

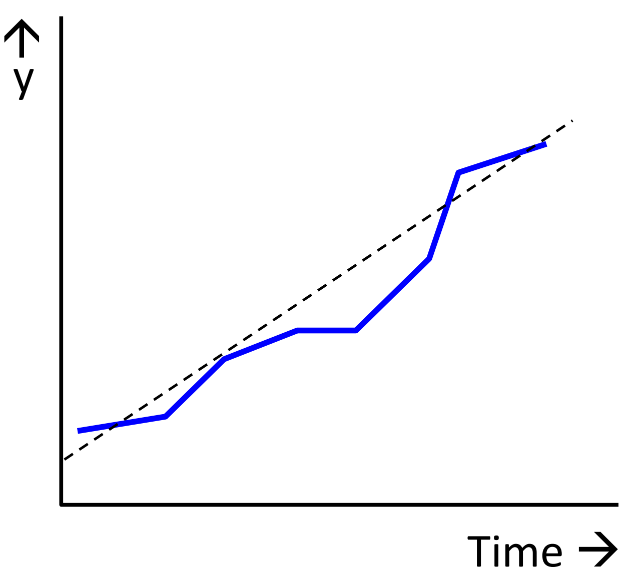
Systematic

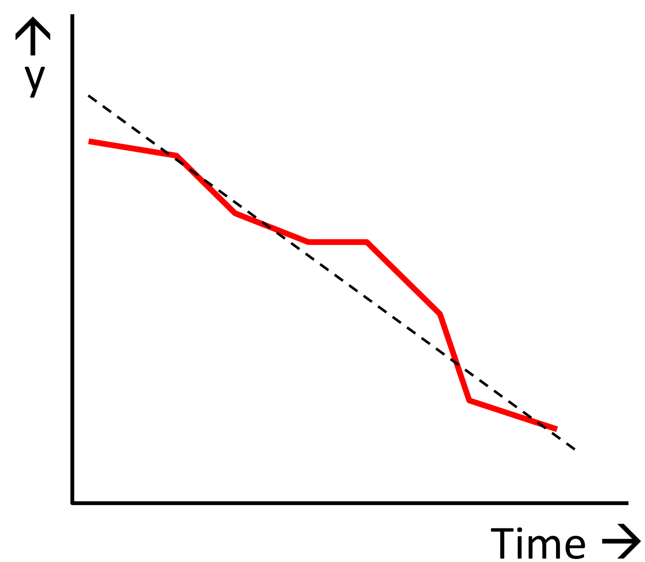
Random

- Trend, Seasonal and Cyclic components are capable of being analyzed and estimated.
- The irregular components represents noise (uncertainty) in the system and is the unexplained part.
- Larger the irregular component, less precise is our estimate and larger is the confidence interval of the estimate.

Trend Component

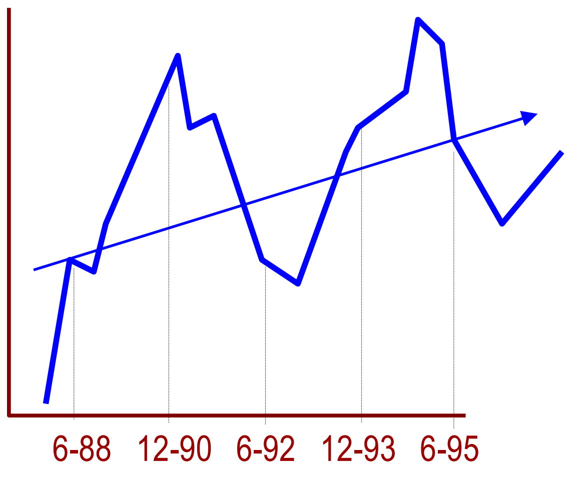
- A **trend** is a long-term, relatively smooth pattern or direction that the series exhibits.
- Its duration is more than one year.
- For example, the population of India during the past several decades has exhibited a trend of relatively steady growth. In this case, the trend is linear.
- Trend can be either be upward or downward.





Cyclic Component

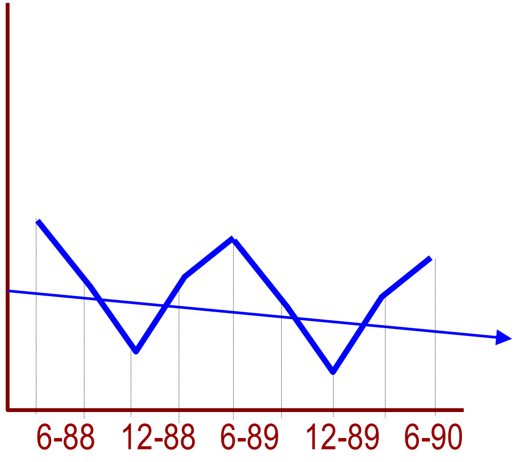
- A **cycle** is a wavelike pattern about a long-term trend that is generally apparent over a number of years.
- By definition, it has a duration of more than one year.
- Examples of cycles include the well-known business cycles that record periods of economic boom, recession and inflation, long-term product-demand cycles, and cycles in the monetary and financial sectors.



A **cycle** is a wavelike pattern describing a long term behavior (for more than one year). Cycles are seldom regular, and often appear in combination with other components.

Seasonal Component

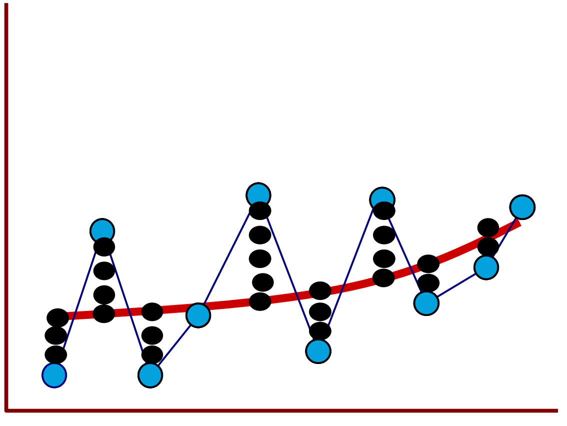
- **Seasonal variations** are like cycles, but they occur over short repetitive calendar periods and by definition, have durations of less than one year.
- The term **seasonal variation** may refer to the four traditional seasons, or to systematic patterns that occur during the period of one week or even over the course of one day.
- Stock-market prices, for example, often show highs and lows at particular times during the day.



The **seasonal component** of the time-series exhibits a short term (less than one year) calendar repetitive behavior.

Irregular Component

- **Random variation** comprises the irregular changes in a time series that are not caused by any other components.
- It tends to hide the existence of the other, more predictable components.
- As **random variation** exists in almost all time series, we need to look at ways to remove the random variation, thereby allowing us to describe and measure the other components and ultimately, to make accurate forecasts.



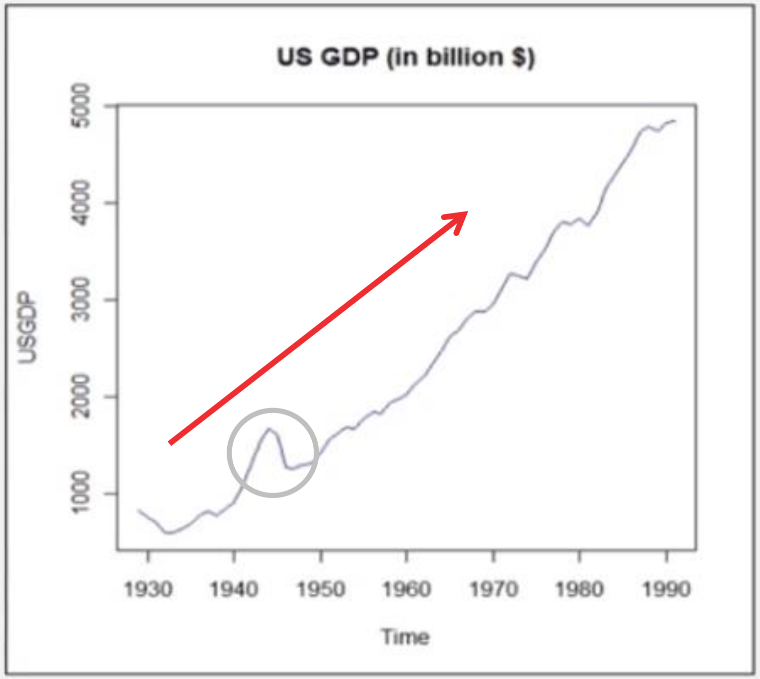
Random variation comprises of unpredictable changes in time series. It tends to hide other (more predictable) components. We try to remove random variation, which then allows us to identify the other components.

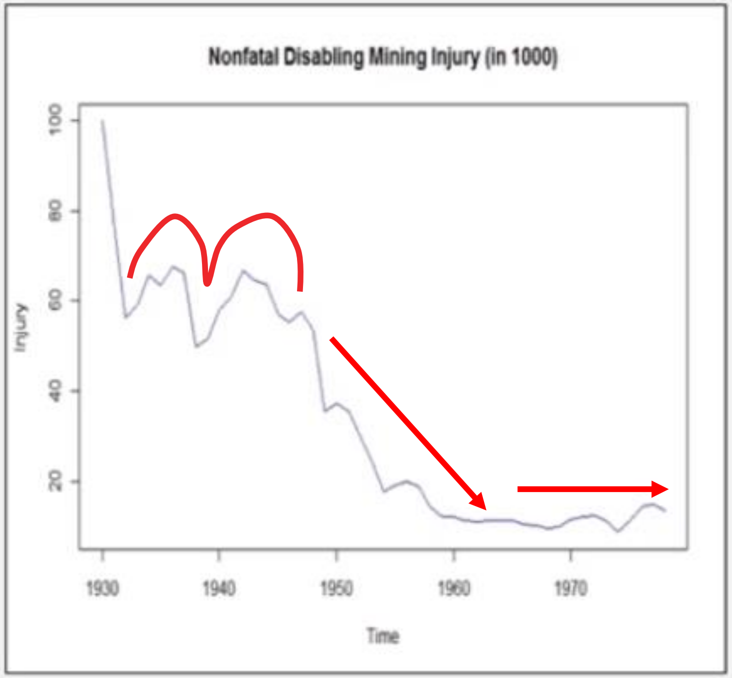
Time Series Visualization

Analyze the Graph – Example 1

- Upward trend
- Some small jump in 40’s

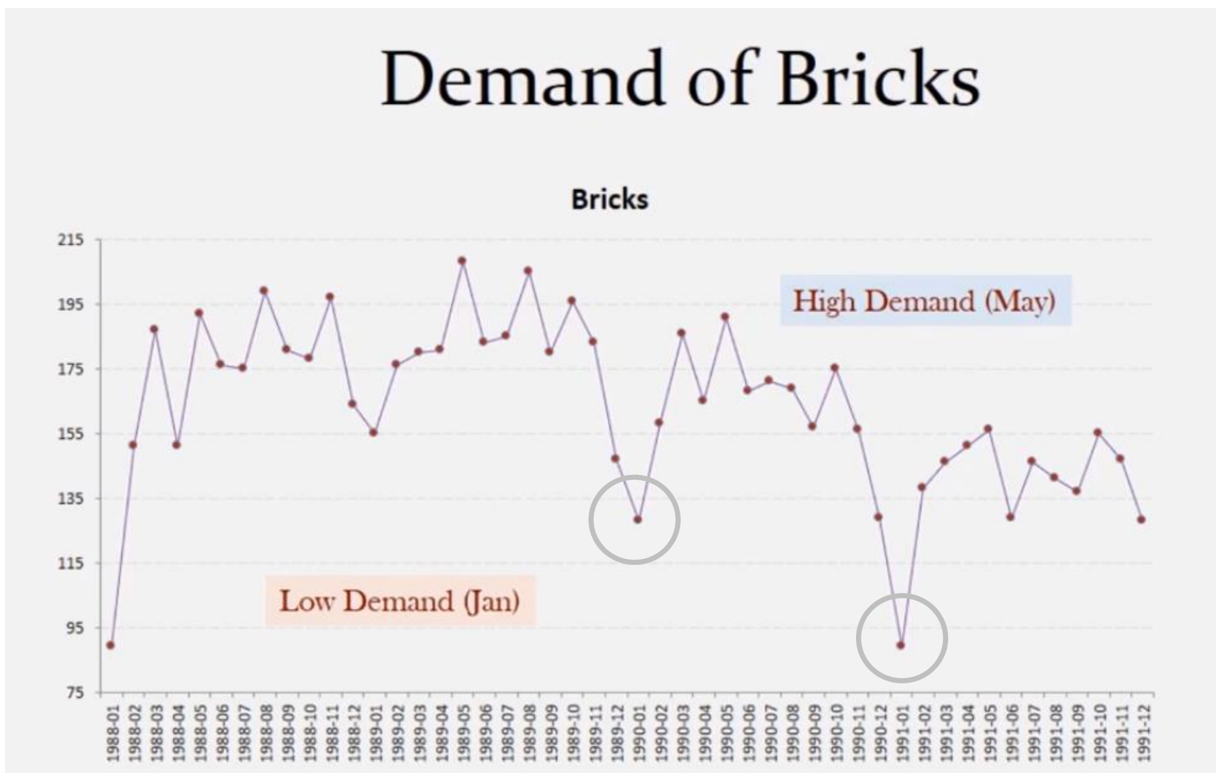
- Initial fluctuations.
- Later downward trend.
- Flat after 60’s.
- Initial period unstable.





Analyze the Graph – Example 2

- No trend.
- Repetition – Seasonality present.
- Seasonal period – 1 year.
- High demand in May, Low in Dec, others fluctuations.
- Outliers – 2 suspected outliers.

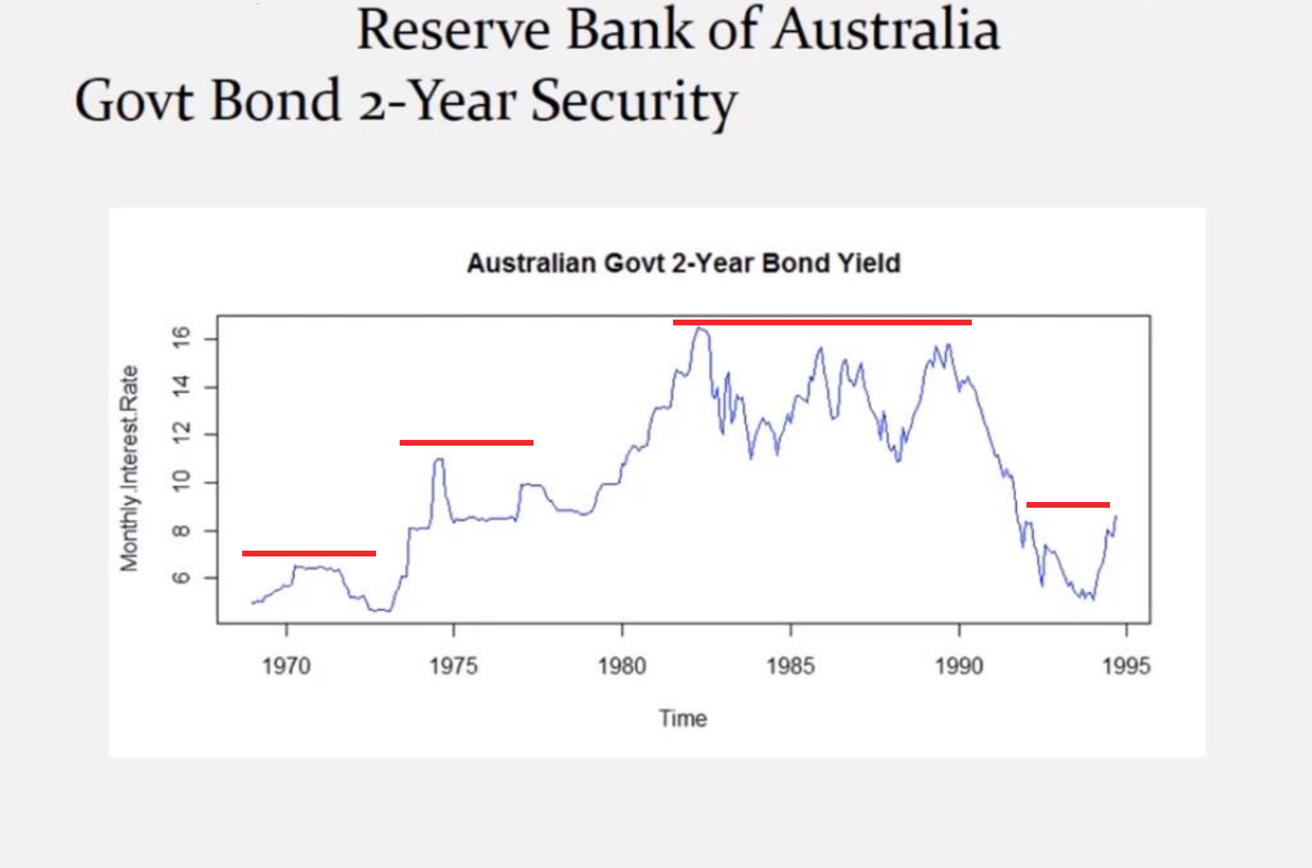


Analyze the Graph – Example 3

- Initial upward trend, period of stability and then downward trend.
- Seasonality.
- No abrupt changes, no outliers.

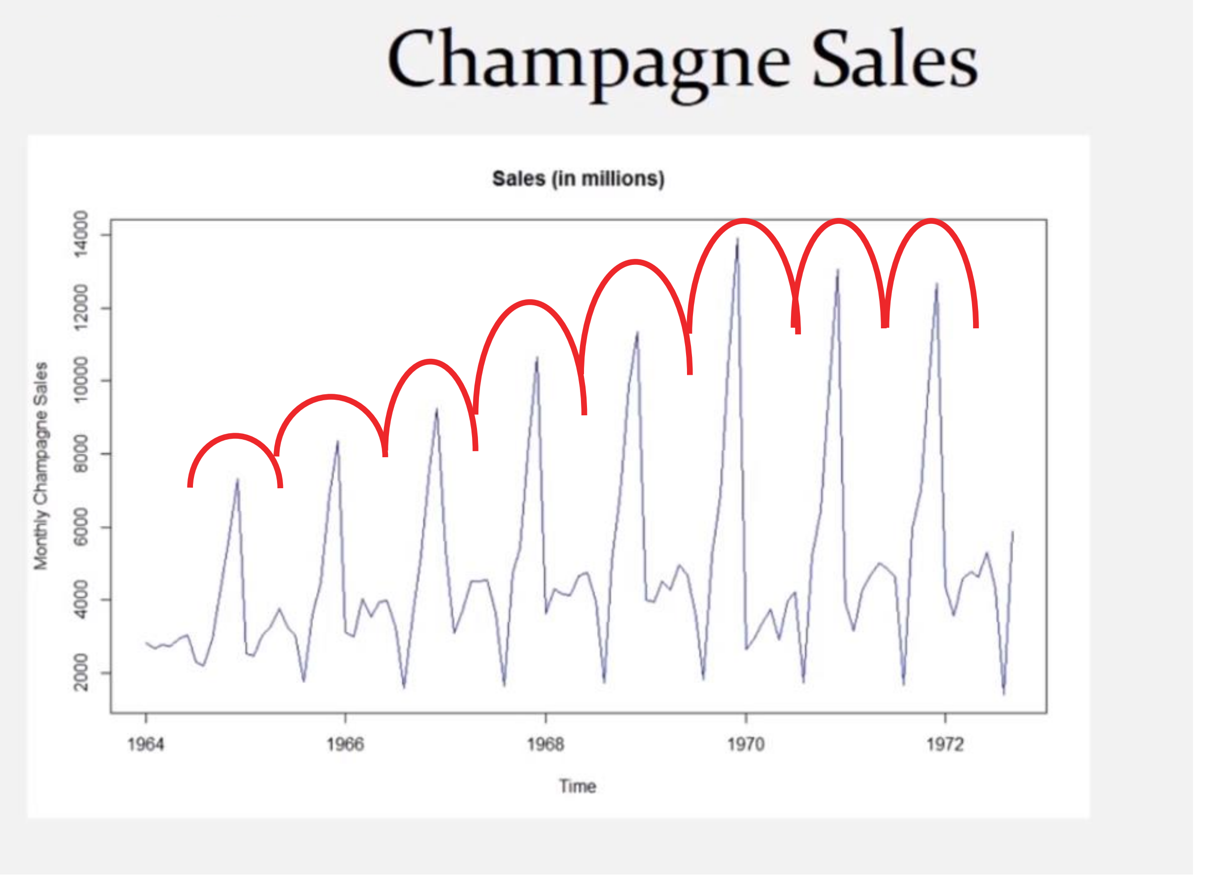
Analyze the Graph – Example 4

- No trend, no seasonality, no cycles, no outliers
- Abrupt jumps



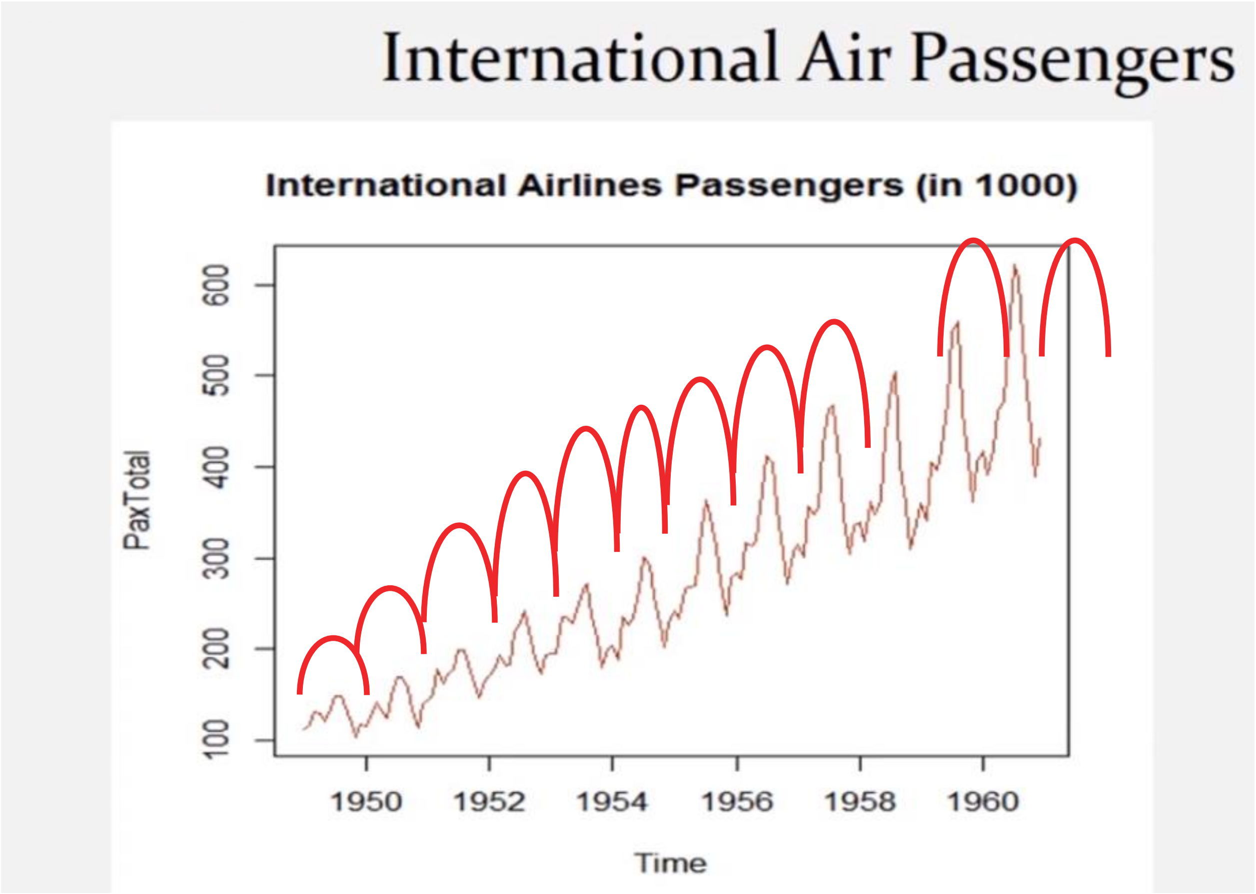
Analyze the Graph – Example 5

- Seasonality.
- Non-constant variance.
- No trend, no cycles, no abrupt changes, no outliers.



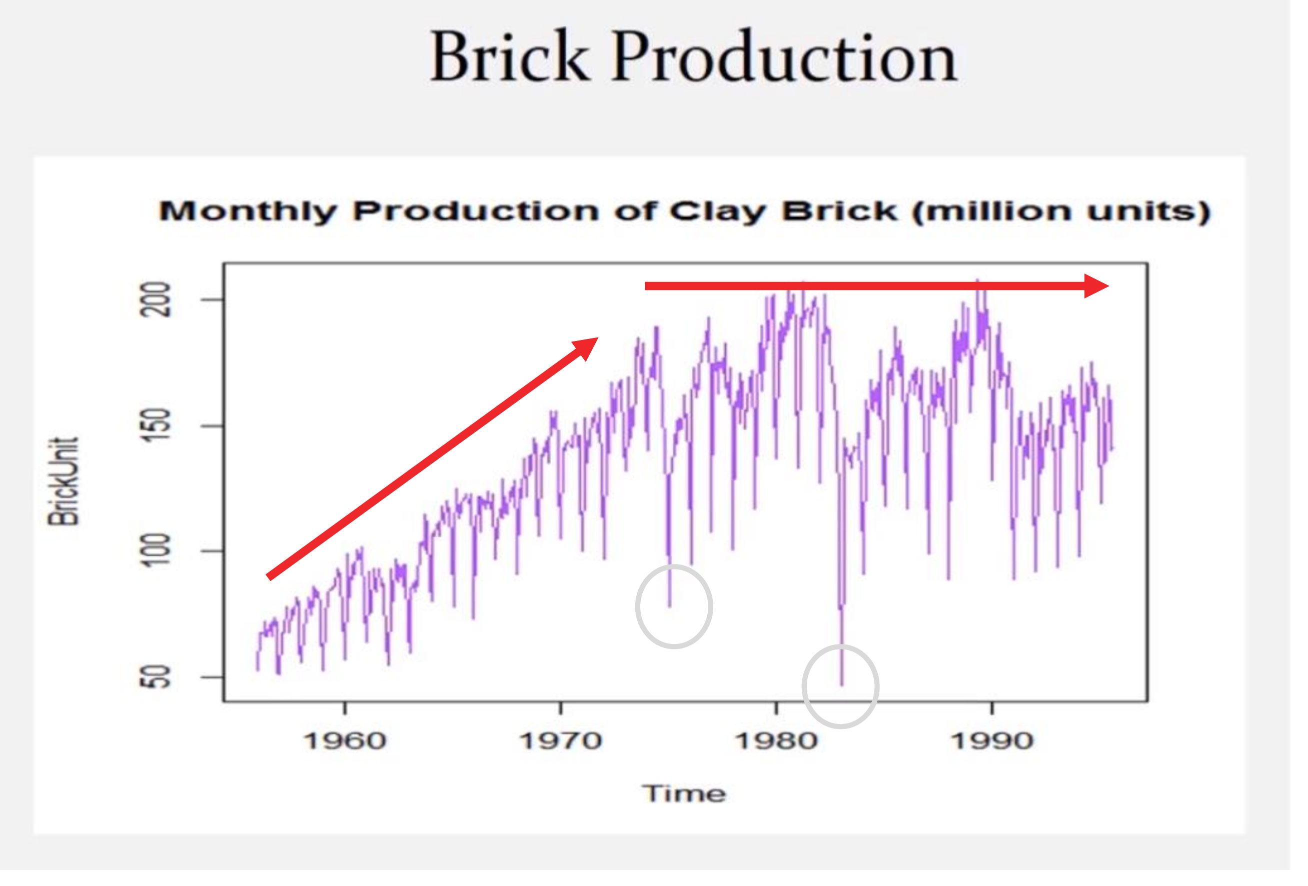
Analyze the Graph – Example 6

- Trend, Seasonality.
- No cycles, No abrupt changes, No outliers



Analyze the Graph – Example 7

- Initial trend, later stable.
- Seasonality.
- Outliers.



Missing Values and Outliers

Missing Values in Time Series Data

- Missing value create holes in the time series.
- No missing value is allowed as data is an ordered set.
- It is not possible to shift the series after discarding missing value.
- Possible reasons for missing values:
    - Data is not collected.
    - Data is corrupted.
    - Data never existed.

Marking and Imputing Missing Values

- Marking missing values:
  - NaN is the default marker for missing values.
  - Methods available to identify missing values: isna(), notna()
- Methods to input missing values:
  - When there is no seasonality, take the mean of adjacent values on either side.
  - When data has seasonality, take the mean of all periods with same seasonal index.
  - When the series has only linear trend, use linear interpolation.

Outliers in Time Series Data

- Outliers are extreme values that affect estimation of statistics.
- Detection of outliers in a time series is difficult.
- In a series with trend, usual methods of z-score and box-plot would not work.
- Use of decomposition helps to identify outliers.
- Once identified use imputation methods previously explained to replace outliers.

Forecasting Accuracy

Forecasting Methods

- Many methods are available for time series forecasting. They are listed below. Choice of a method depends on the underlying time series pattern.
  - Naïve: Use immediate previous value to forecast new value.
  - Average: Use average of all previous values to forecast new value.
  - Moving Average: Use fixed number of previous values to forecast new value.
  - Exponential: Use the equation $F_{t+1} = \alpha Y_t + (1 - \alpha)F_t$.
  - Decomposition: Use equation $Y_t = Trend_t \times Seasonal_t \times Irregular_t$
  - ARIMA: Use combination of AR and MA models
- Not all models can be used in all situations. Later we explain the suitability of each of above models when a specific pattern is observed.

Forecasting Accuracy

- Accuracy of forecasting of the different methods listed in previous slide can be compared using forecasting error measurements.
- These measures are:

Name	Description	Formula
Forecast Error	Compare actual value to forecast value	$actual\ value\ - forecast\ value$
MAE	Mean Absolute Error	$mean(abs(actual\ - forecast))$
MSE	Mean Square Error	$mean((actual\ - forecast)^2)$
MAPE	Mean Absolute Percentage Error	$Mean\left(abs\left(\frac{(actual\ - forecast)}{actual} \times 100\right)\right)$
RMSE	Square root of MSE.	$\sqrt{mean((actual\ - forecast)^2)}$

Naïve and Average Methods

- These are simplest of forecasting methods.
- They are only used when under lying pattern is horizontal.
- They are not appropriate when significant trend, seasonal or cyclical components are present.
- They are used for short term forecasting like 1 or 2 periods ahead.
- When they are level changes naïve method quickly adjusts itself, where as average method takes several periods to smoothen out the level change.
- Average method has higher accuracy when the pattern is stable and in general.

Forecasting Using Naïve Method

- This method uses most recently observed value to forecast next observation value.

Period	Observed	Forecast	Error	Abs(Error)	Squared Error	% Error	Abs(% error)
1	17						
2	21	17	4	4	16	19.05	19.05
3	19	21	-2	2	4	-10.53	10.53
4	23	19	4	4	16	17.39	17.39
5	18	23	-5	5	25	-27.78	27.78
6	16	18	-2	2	4	-12.50	12.50
7	20	16	4	4	16	20.00	20.00
8	18	20	-2	2	4	-11.11	11.11
9	22	18	4	4	16	18.18	18.18
10	20	22	-2	2	4	-10.00	10.00
11	15	20	-5	5	25	-33.33	33.33
12	22	15	7	7	49	31.82	31.82
Total			5	41	179	1.19	211.69

MAE = 3.73 MSE = 16.27 MAPE = 19.24%

Forecasting Using Average Method

- This method uses average of all previous values to forecast next observation value.

Period	Observed	Forecast	Error	Abs(Error)	Squared Error	% Error	Abs(% error)
1	17						
2	21	17.00	4.00	4.00	16.00	19.05	19.05
3	19	19.00	0.00	0.00	0.00	0.00	0.00
4	23	19.00	4.00	4.00	16.00	17.39	17.39
5	18	20.00	-2.00	2.00	4.00	-11.11	11.11
6	16	19.60	-3.60	3.60	12.96	-22.50	22.50
7	20	19.00	1.00	1.00	1.00	5.00	5.00
8	18	19.14	-1.14	1.14	1.31	-6.35	6.35
9	22	19.00	3.00	3.00	9.00	13.64	13.64
10	20	19.33	0.67	0.67	0.44	3.33	3.33
11	15	19.40	-4.40	4.40	19.36	-29.33	29.33
12	22	19.00	3.00	3.00	9.00	13.64	13.64
Total			4.53	26.81	89.07	2.76	141.34

MAE = 2.44 MSE = 8.10 MAPE = 12.85%

Moving Average and Exponential Smoothing

- These are more sophisticated methods than naïve methods considered earlier.
- These methods also require the time series pattern to be horizontal.
- They are not appropriate when significant trend, seasonal or cyclical components are present.
- However, they adopt much better to changing levels, while maintaining forecast accuracy.
- Therefore they are recommended for short term forecasting for one or two periods ahead.

Moving Averages

- The moving average method uses the average of most recent k periods for forecasting values for next period.

$$MOVING\ AVERAGE\ FORECAST\ OF\ ORDER\ k$$

$$F_{t+1} = \frac{\sum most\ recent\ k\ data\ values}{k} = \frac{Y_t + Y_{t-1} + \cdots + Y_{t-k+1}}{k}$$

- When a new observation becomes available it replaces the oldest observation in the equation.
- Because of this the average value keeps changing as new observations become available.

Moving Average Method

- The table shows calculations for 3 period (k=3) moving average.

Period	Observed	Forecast	Error	Abs(Error)	Squared Error	% Error	Abs(% error)
1	17						
2	21						
3	19						
4	23	19	4	4	16	17.39	17.39
5	18	21	-3	3	9	-16.67	16.67
6	16	20	-4	4	16	-25.00	25.00
7	20	19	1	1	1	5.00	5.00
8	18	18	0	0	0	0.00	0.00
9	22	18	4	4	16	18.18	18.18
10	20	20	0	0	0	0.00	0.00
11	15	20	5	5	25	-33.33	33.33
12	22	21	3	3	9	13.64	13.64
Total			0	24	92	-20.79	129.21

MAE = 2.67 MSE = 10.22 MAPE = 14.36%

- Best value of k is 6 when MSE = 6.79

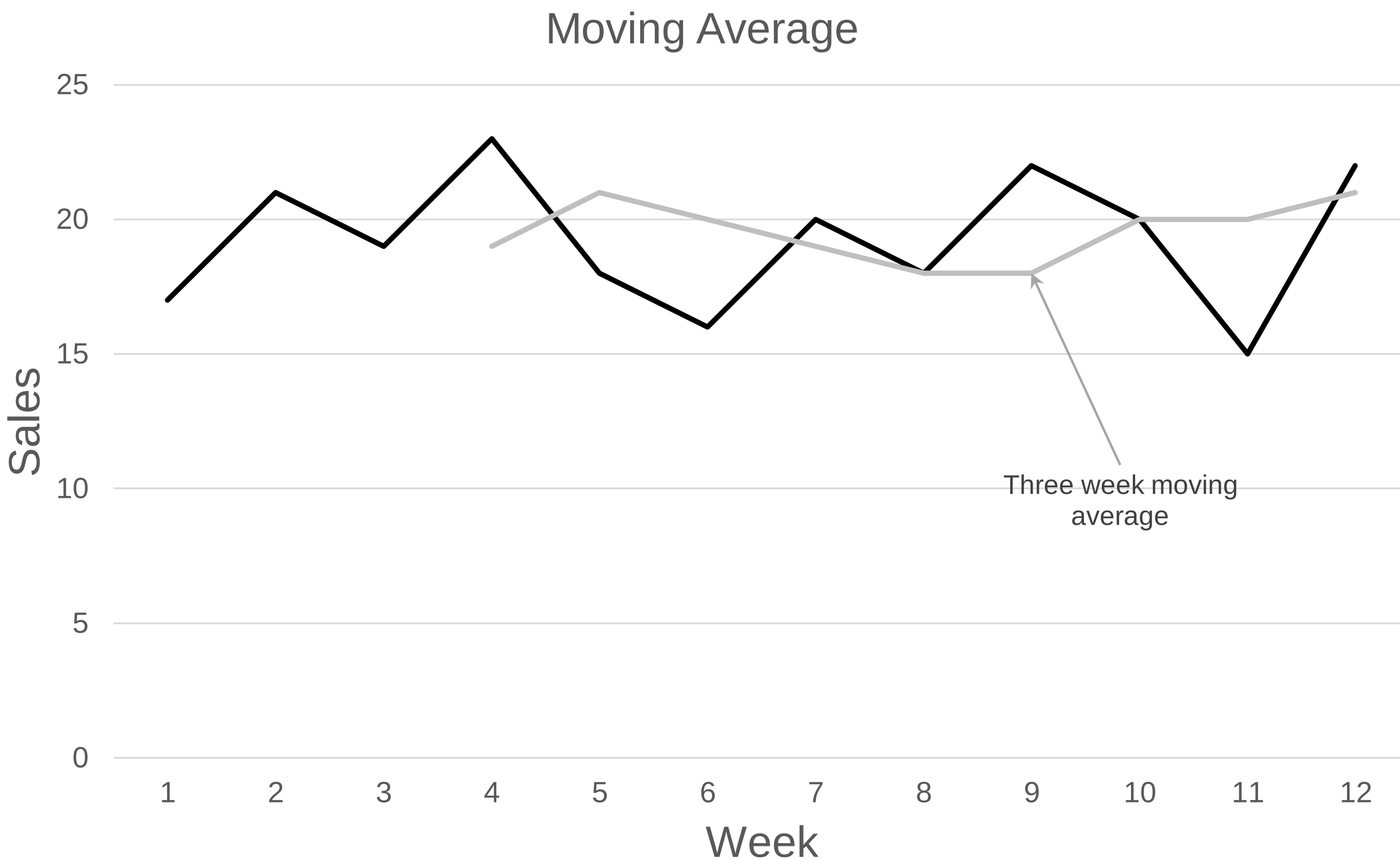
Moving Average Method – Best k

- The table shows calculations for 6 period (k=6) moving average.

Week	Sales	F-MA6	FC Error	Abs Error	Sq Error	% Error	Abs % Err
1	17						
2	21						
3	19						
4	23						
5	18						
6	16						
7	20	19	1	1	1	5	5
8	18	19.5	-1.5	1.5	2.25	-8.33	8.33
9	22	19	3	3	9	13.64	13.64
10	20	19.5	0.5	0.5	0.25	2.50	2.50
11	15	19	-4	4	16	-26.67	26.67
12	22	18.5	3.5	3.5	12.25	15.91	15.91

MAE = 2.25 MSE = 6.79 MAPE = 12.01%

Graph of 3 Period Moving Average



Exponential Smoothing

greatlearning

- Exponential smoothing is a weighted average of past values.
- The exponential smoothing method provides smoothed values for all the time periods observed.
- When smoothing time series at time t, the exponential smoothing method considers all the data available at t, $(y_t, y_{t-1}, \ldots)$
- Exponential Smoothing forecast

$$F_{t+1} = \alpha Y_t + (1-\alpha) F_t$$

Y_t is observed value at period t.
F_t is forecast value at period t.

$$F_t = \alpha Y_{t-1} + \alpha(1-\alpha) Y_{t-2} + \cdots + (1-\alpha)^{t-1} Y_1$$

- The value of α is chosen on the basis of how much smoothing is required.
- A small value provides a lot of smoothing and a large value, little smoothing.

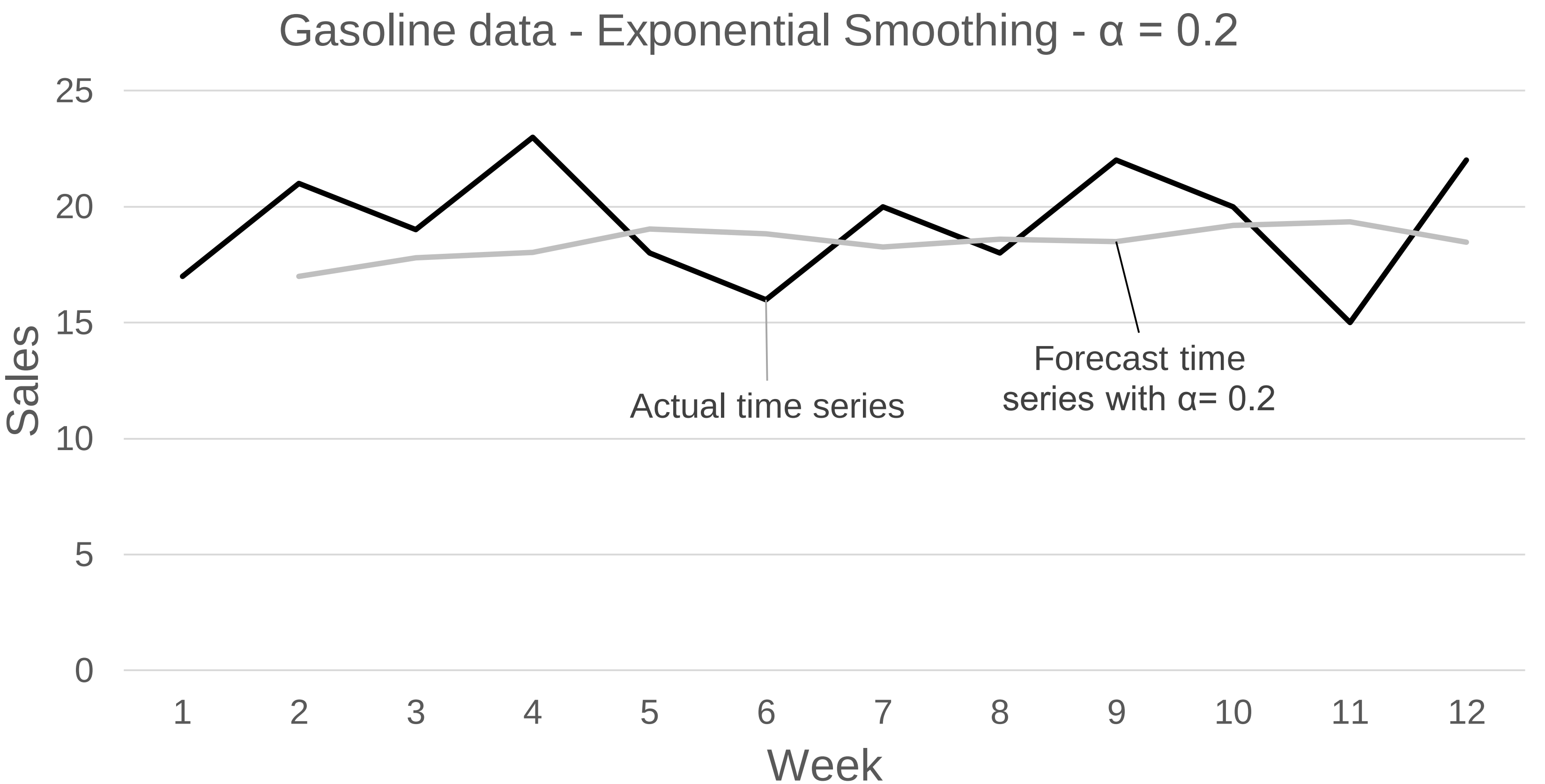
Exponential Smoothing for gasoline data

- When $\alpha = 0.2$

Week	Sales	Forecast	Forecast Error	Squared Error
1	17			
2	21	17.00	4	16.00
3	19	17.80	1.2	1.44
4	23	18.04	4.96	24.60
5	18	19.03	-1.03	1.06
6	16	18.83	-2.83	8.01
7	20	18.26	1.74	3.03
8	18	18.61	-0.61	0.37
9	22	18.49	3.51	12.32
10	20	19.19	0.81	0.66
11	15	19.35	-4.35	18.92
12	22	18.48	3.52	12.39

MSE = 8.98

Exponential Smoothing Graph



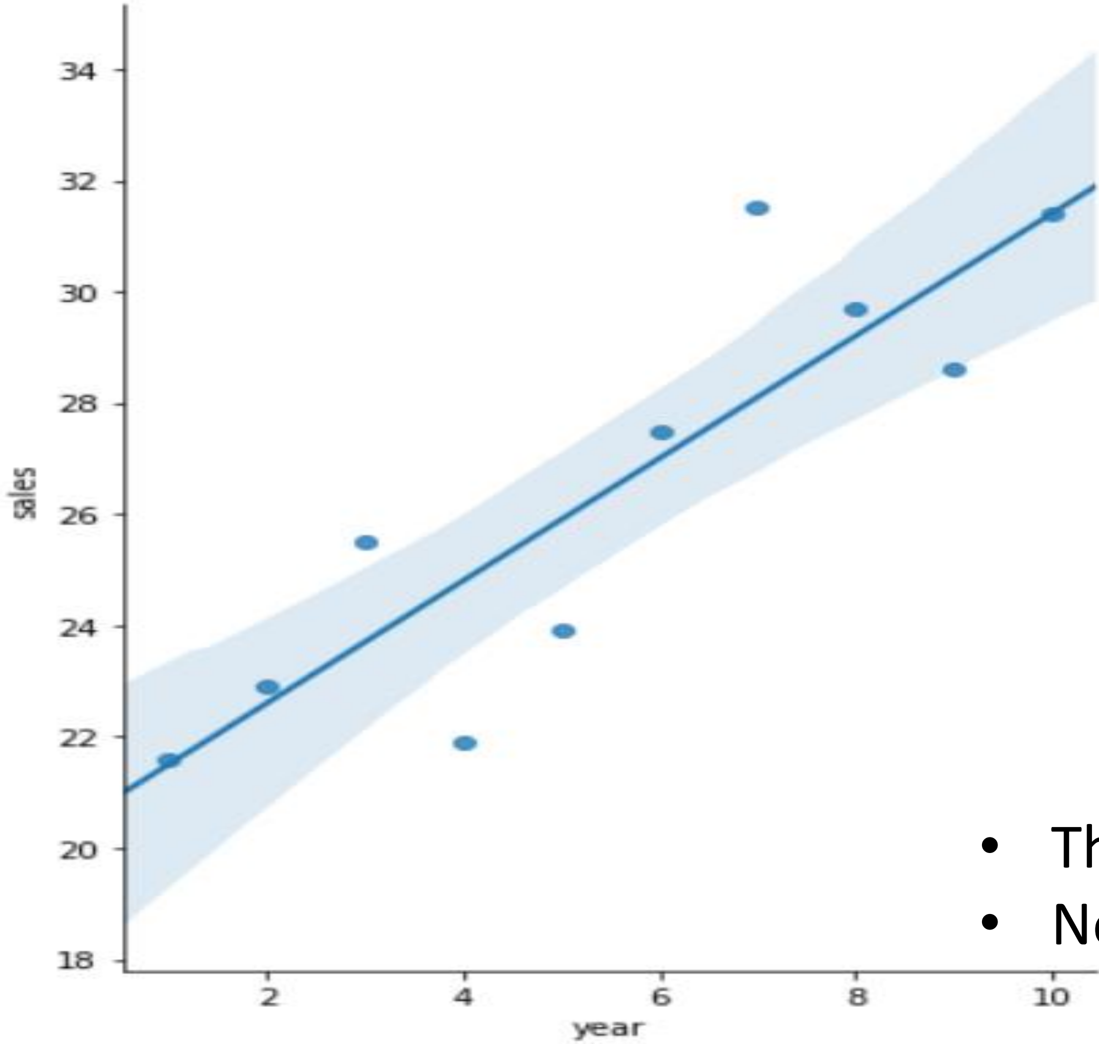
Trend Projection using Regression

- The methods considered so for, required the time series pattern to be horizontal.
- We will now consider when linear trend component is only present.
- In this case linear regression may be used for forecasting.
- We will use following bicycle sales data for the purpose:

Period	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Sales	21.6	22.9	25.5	21.9	23.9	27.5	31.5	29.7	28.6	31.4

- This is a simple linear regression model with time period considered as independent variable.
- The steps are
  1. Draw line plot to verify linearity and absence of seasonality and cyclic components.
  2. Run regression.
  3. Analyze residuals for model validation.

Linear Regression of Bicycle Data.



- The graph is fairly linear.
- No unusual observations.

Linear Regression of Bicycle Data

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  sales   R-squared:                       0.765
Model:                            OLS   Adj. R-squared:                  0.735
Method:                 Least Squares   F-statistic:                     26.01
Date:                Wed, 21 Aug 2019   Prob (F-statistic):           0.000930
Time:                        15:44:27   Log-Likelihood:                -19.798
No. Observations:                  10   AIC:                             43.60
Df Residuals:                       8   BIC:                             44.20
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     20.4000      1.338     15.244      0.000      17.314      23.486
year           1.1000      0.216      5.100      0.001       0.603       1.597
==============================================================================
Omnibus:                        0.173   Durbin-Watson:                   1.824
Prob(Omnibus):                  0.917   Jarque-Bera (JB):                0.126
Skew:                           0.151   Prob(JB):                        0.939
Kurtosis:                       2.539   Cond. No.                         13.7
==============================================================================
```

Day 2

Holt's Linear Exponential Smoothing

- Charles Holt developed an enhanced version of simple exponential smoothing (SES), which can be used for forecasting time series having a trend component.

- Recall the SES model with smoothing constant α.

$$F_{t+1} = \alpha Y_t + (1-\alpha)F_t$$

- The Holt's Model has 2 constants α and β and 3 equations.

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + b_{t-1})$$
$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)(b_{t-1})$$
$$F_{t+k} = L_t + b_t k$$

$$L_t = estimate\ of\ the\ level\ of\ series\ at\ period\ t.$$
$$b_t = estimate\ of\ the\ slope\ of\ series\ at\ period\ t.$$
$$\alpha, \beta = smoothing\ constants\ for\ level\ and\ slope\ respectively.$$
$$F_{t+k} = Forecast\ for\ k\ periods\ ahead.$$

Holt’s Model – Bike sales example

- Let us apply Holt’s model to bike sales example previously considered with α = 0.1 and β = 0.2.

Period	1	2	3	4	5	6	7	8	9	10
Sales	21.6	22.9	25.5	21.9	23.9	27.5	31.5	29.7	28.6	31.4

- To get started we fix the initial values as follows:

$$L_1 = Y_1 = 21.6 \; and \; b_1 = (Y_2 - Y_1) = (22.9 - 21.6) = 1.3$$

- With this information we can calculate forecast values for next period.

$$L_2 = .1(22.9) + .9(21.6 + 1.3) = 22.9$$

$$b_2 = .2\,(22.9 - 21.6) + (1 - .2)(1.3) = 1.3$$

We can continue and prepare the table in the next page.

Holt's Model – Bike sales example continued

$\alpha =$ 0.1	$\beta =$ 0.2	k = 1

Period	Sales	Estimated Level	Estimated Trend	Forecast	Forecast Error	Squared Error
1	21.6	21.600	1.300			
2	22.9	22.900	1.300	22.900	0.000	0.000
3	25.5	24.330	1.326	24.200	1.300	1.690
4	21.9	25.280	1.251	25.656	-3.756	14.108
5	23.9	26.268	1.198	26.531	-2.631	6.924
6	27.5	27.470	1.199	27.466	0.034	0.001
7	31.5	28.952	1.256	28.669	2.831	8.016
8	29.7	30.157	1.245	30.207	-0.507	0.257
9	28.6	31.122	1.189	31.402	-2.802	7.851
10	31.4	32.220	1.171	32.311	-0.911	0.830
						39.678

$F_{11} = 32.220 + 1.171 = 33.391$ MSE = 39.678/9 = 4.41

Holt Winter's Model

- Used to make predictions when time series data has trend and seasonality.
- Additive or Multiplicative model can be used depending on:
  - Additive model is preferred when the seasonal component's magnitude is approximately constant.
  - Multiplication model is preferred when magnitude of seasonal component is non-constant.

Holt Winter's Model Equation

$$L_t = \alpha(y_t - S_{t-m}) + (1-\alpha)(L_{t-1} + T_{t-1}),$$

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

$$S_t = \gamma(y_t - L_t) + (1-\gamma)S_{t-m}.$$

- where α, β, and γ are the three smoothing parameters
- α - level pattern
- β – trend
- γ - seasonality
- m – seasonal period

Time Series Decomposition

- The method described in this section is useful when time series data contains all the components i. e. trend, seasonal, cyclic and irregular.
- The method separates (or decomposes) the series into the above components.
- This method can be used for understanding the series and as well as forecast future values.
- Since cyclic patterns are difficult to model, we will exclude (or combine with trend) this component from our analysis.
- How the components are to generate the observed values depends on our assumptions about the model.
- The relationships can be defined as either additive or multiplicative model.

Additive or Multiplicative model?

- The additive model is given in the form:

$$Y_t = \text{Trend}_t + \text{Sesonal}_t + \text{Irregular}_t$$

- The values of three components are added to derive the observed value of the time series.
- The irregular component is the unexplained variability of the series.
- The additive model is appropriate when seasonal fluctuations do not depend on the level of the time series.
- If the seasonal components increase with the trend, then the multiplicative model is more appropriate.

Multiplicative model?

- The multiplicative model is given in the form:

  $Y_t = Trend_t \times Sesonal_t \times Irregular_t$

- The values of three components are multiplied to derive the observed value of the time series.
- Trend is measured in the units of the y variable, but the seasonal and irregular components are in relative terms.
- A value of S_t or I_t greater than 1 indicates that the value is above trend line and vice-versa.
- The multiplicative model is most popular and is the most preferred decomposition model.

Umbrella Sales Case

It is common knowledge that umbrella sales is very seasonal and depending on specific brand can have a trend too.

The following data pertains to quarterly sales of a particular company.

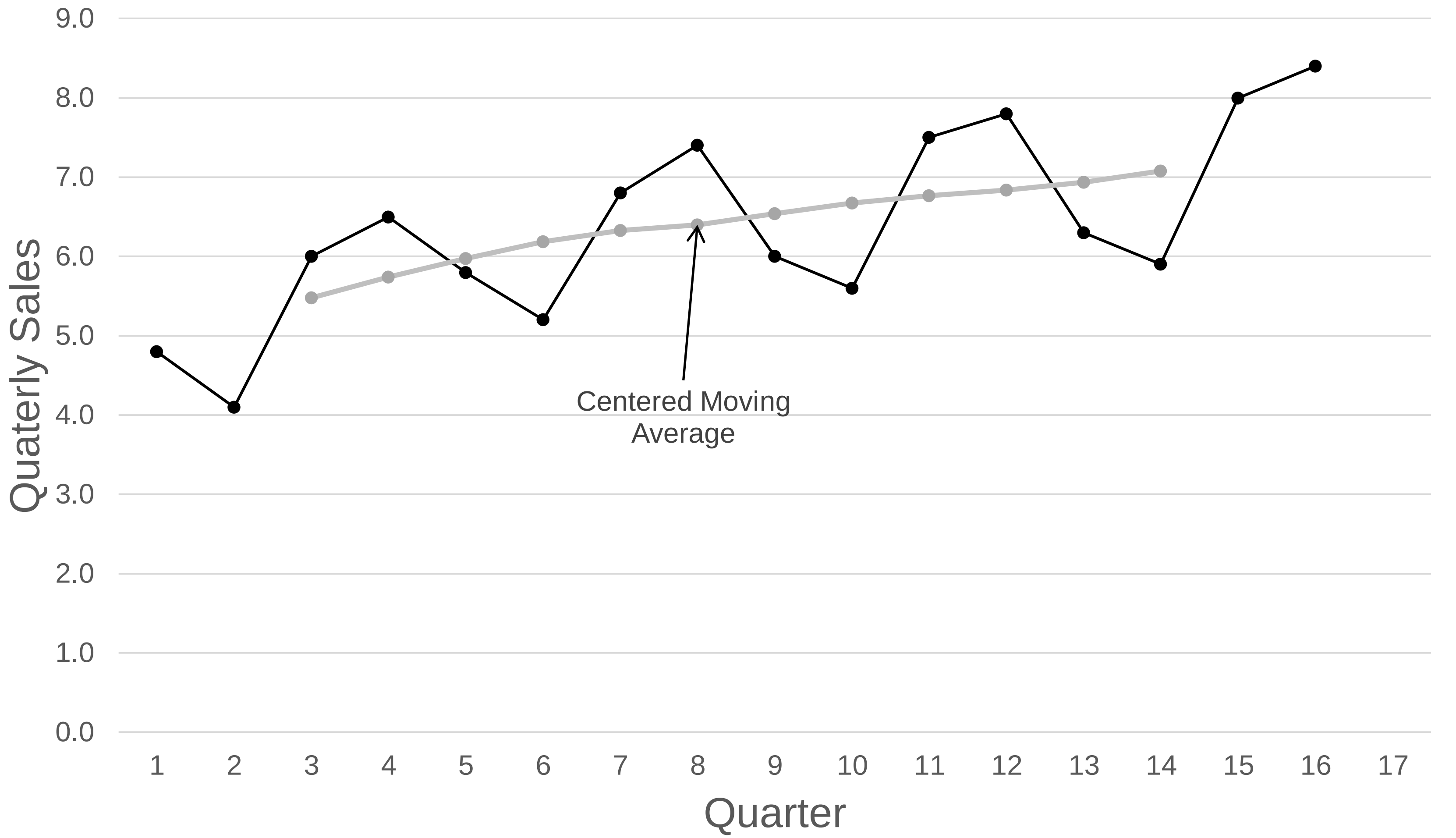
Year	1				2				3			
Quarter	1	2	3	4	1	2	3	4	1	2	3	4
Sales	4.8	4.1	6.0	6.5	5.8	5.2	6.8	7.4	6.0	5.6	7.5	7.8

Year	4			
Quarter	1	2	3	4
Sales	6.3	5.9	8.0	8.4

Time Series Decomposition Steps

1. Plot the graph and analyze the pattern for trend, seasonality and unusual observations.
2. Identify repetition value (no. of periods) for seasonality.
3. Compute the moving average with above repetition value as k.
4. Centre the moving average series to align with original time series.
5. Calculate SI (Seasonality*Irregular) index.
6. Average the SI index to compute seasonal index.
7. Using linear regression, extract trend of the series T.
8. Multiply S*T to estimate forecast value.

Umbrella Sales Time Series Plot



Calculating Seasonal Index - 1

Year	Quarter	Sales	4 Quarter MA	Centred MA	Seasonal Irregular Index
1	1	4.8			
1	2	4.1	5.350		
1	3	6.0	5.600	5.475	1.096
1	4	6.5	5.875	5.7375	1.133
2	1	5.8	6.075	5.975	0.971
2	2	5.2	6.300	6.1875	0.840
2	3	6.8	6.350	6.325	1.075
2	4	7.4	6.450	6.4	1.156
3	1	6.0	6.625	6.5375	0.918
3	2	5.6	6.725	6.675	0.839
3	3	7.5	6.800	6.7625	1.109
3	4	7.8	6.875	6.8375	1.141
4	1	6.3	7.000	6.9375	0.908
4	2	5.9	7.150	7.075	0.834
4	3	8.0			
4	4	8.4			

Calculating Seasonal Index

Quarter	Seasonal Irregular Index			Seasonal Index	Adjusted Seasonal Index
1	0.971	0.918	0.908	0.932	0.931
2	0.840	0.839	0.834	0.838	0.836
3	1.096	1.075	1.109	1.093	1.092
4	1.133	1.156	1.141	1.143	1.141

New Observation (Period 17):

Trend (Period: 17) = 7.91

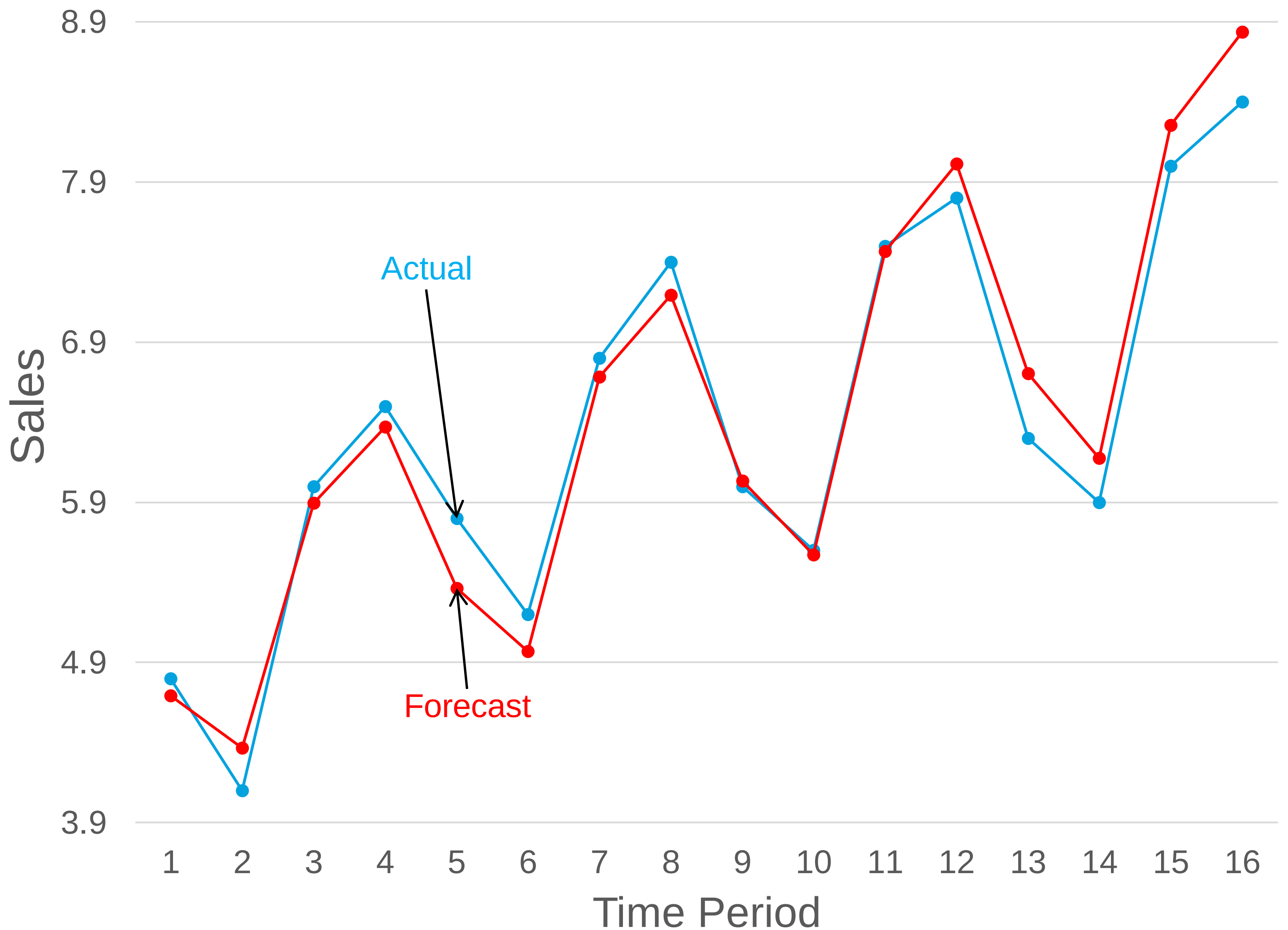
Seasonal Index (Quarter: 1) = 0.931

Forecast Value (S*I) = 7.91 * 0.931 = 7.36

Forecasting Sales

Time Period	Sales	Seasonal Index	Trend	Forecast Sales
1	4.8	0.932	5.03	4.691
2	4.1	0.838	5.21	4.367
3	6.0	1.093	5.39	5.895
4	6.5	1.143	5.57	6.370
5	5.8	0.932	5.75	5.362
6	5.2	0.838	5.93	4.969
7	6.8	1.093	6.11	6.682
8	7.4	1.143	6.29	7.193
9	6.0	0.932	6.47	6.032
10	5.6	0.838	6.65	5.572
11	7.5	1.093	6.83	7.469
12	7.8	1.143	7.01	8.015
13	6.3	0.932	7.19	6.703
14	5.9	0.838	7.37	6.175
15	8.0	1.093	7.55	8.255
16	8.4	1.143	7.73	8.838
17			7.91	7.374

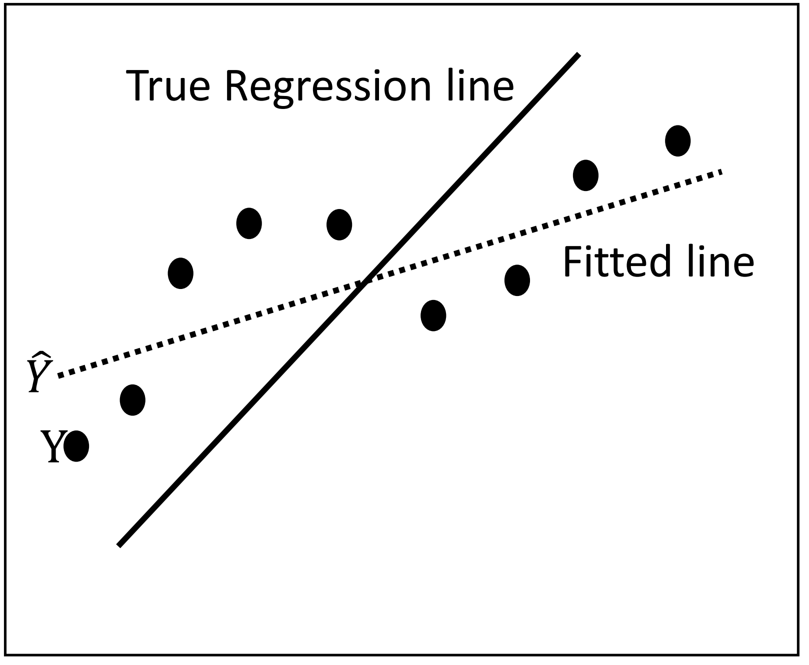
Actual versus Forecast Graph



ARIMA Concepts

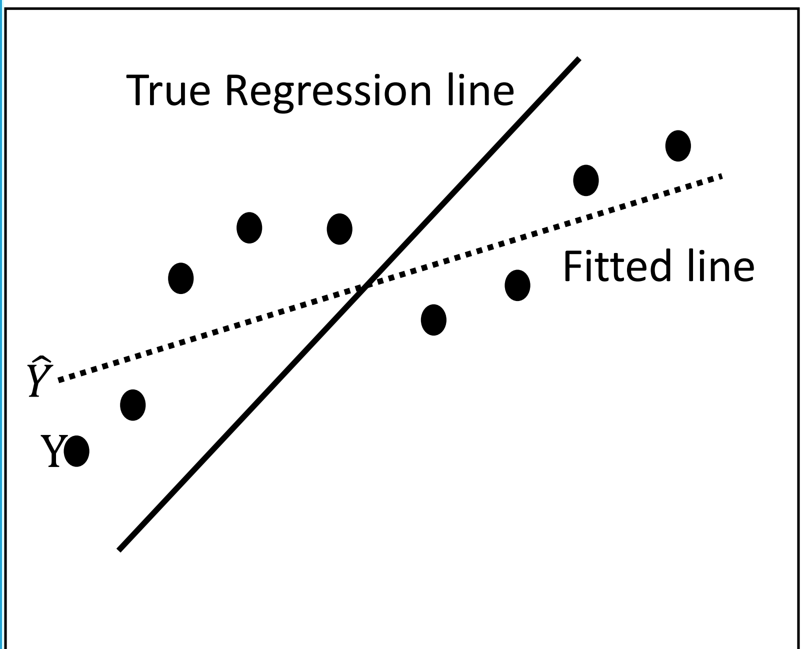
Stationarity

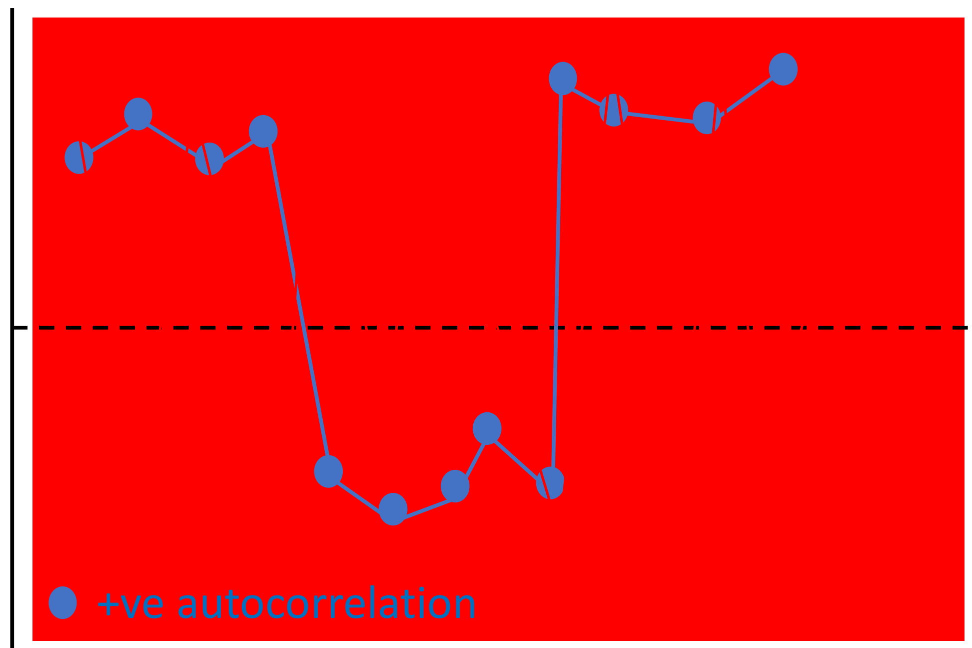
- Cross Sectional regression models assume that residuals are independent.
- If error terms are not independent the basic assumption of regression is violated.
- In time series data the previous value of error term influences it next value
- This leads to fitted regression to deviate from real relationship.
- The regression model also underestimated the variance of the error term.
- If the error terms of a time series are uncorrelated, the series is said to be stationary.



Stationarity

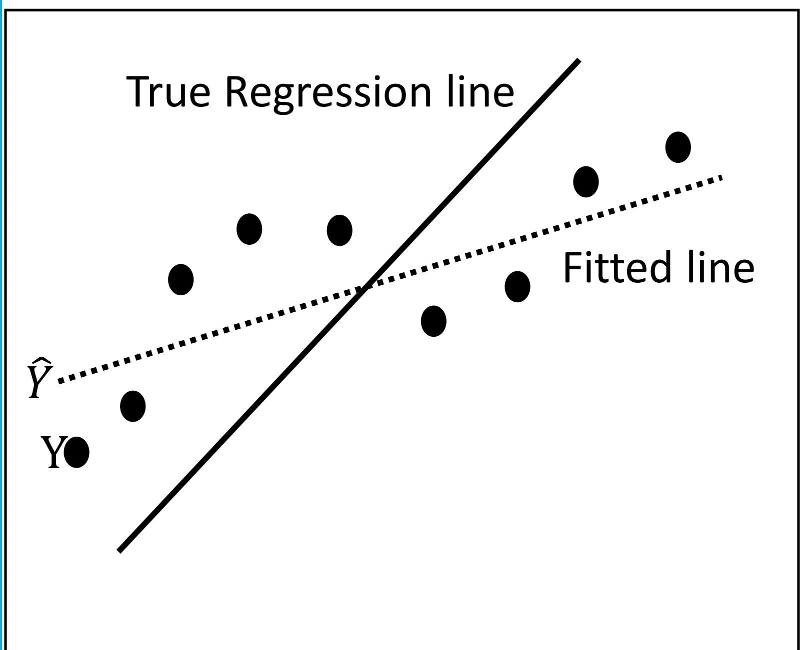
- The condition under which error term is influenced by its previous value is called auto-correlation.
- Auto correlation can be positive or negative.
- Positive correlation makes error terms to sequentially align to the same side of the mean.
- Negative correlation makes error terms to alternate on either side of the mean line.

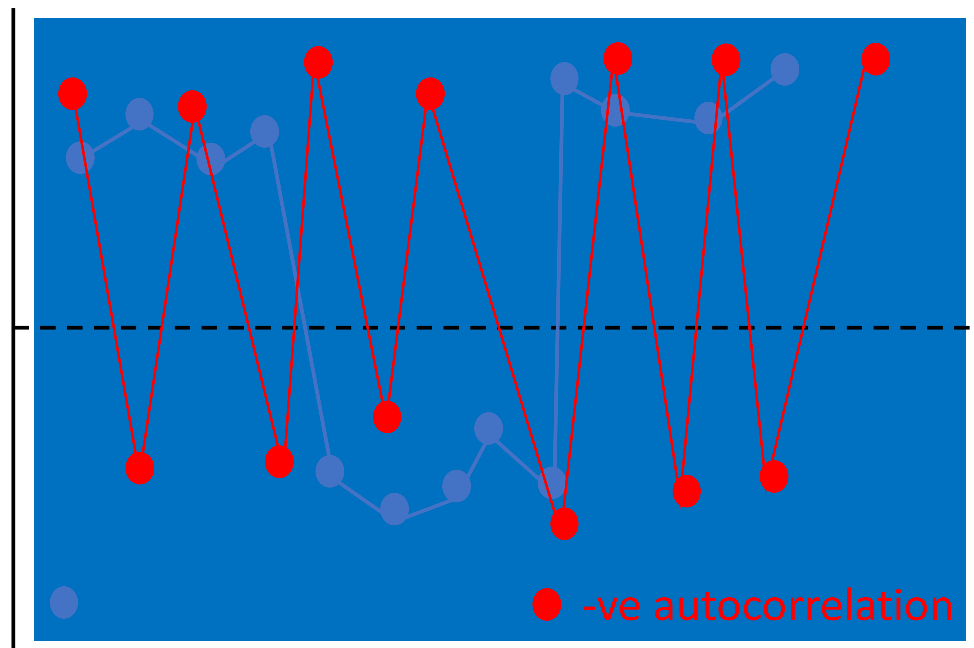




Stationarity

- The condition under which error term is influenced by its previous value is called auto-correlation.
- Auto correlation can be positive or negative.
- Positive correlation makes error terms to sequentially align to the same side of the mean.
- Negative correlation makes error terms to alternate on either side of the mean line.

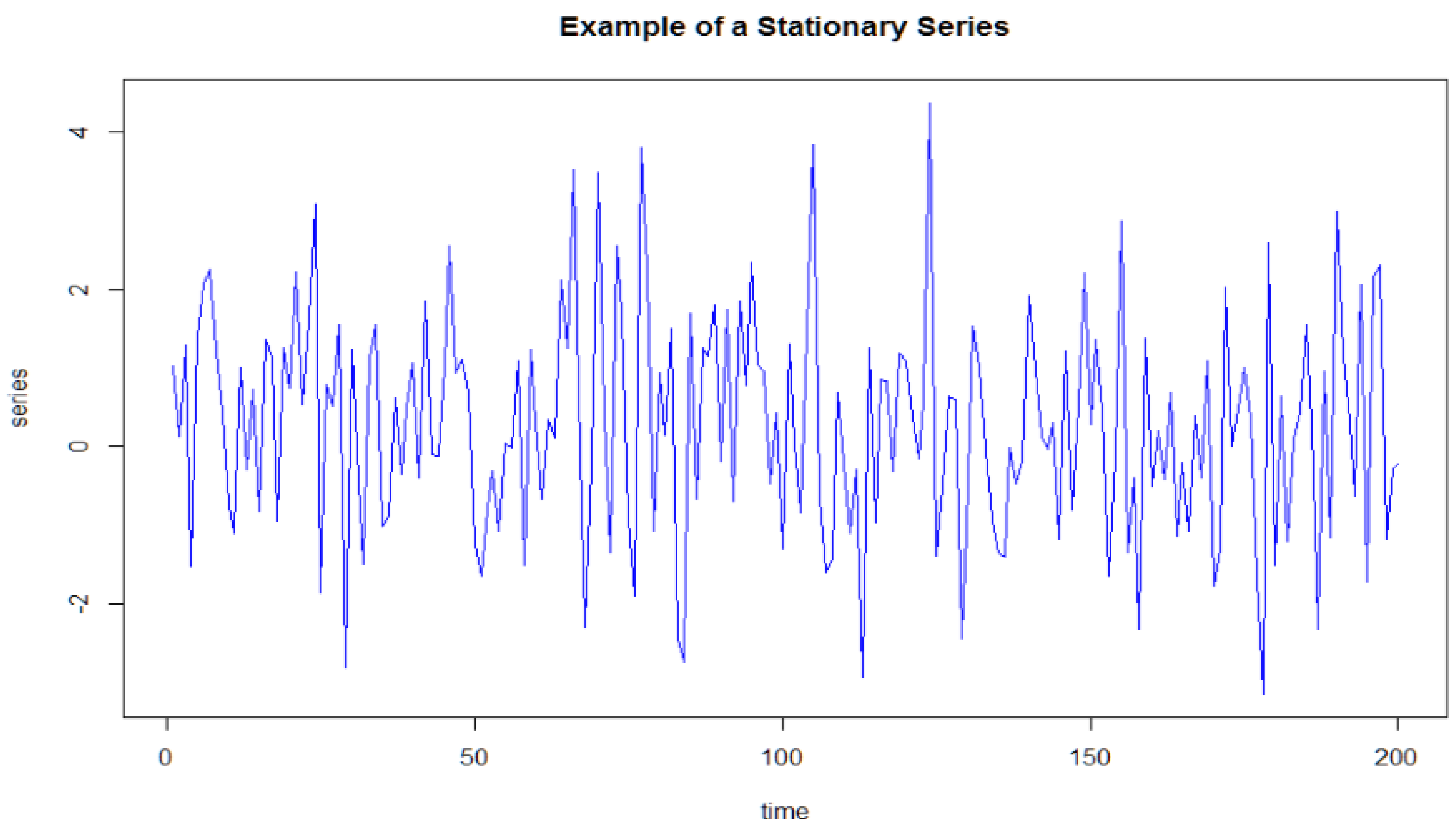




ARIMA and Stationarity

- Most widely-used approaches to time series forecasting.
- Acronym for **A**uto-**R**egressive **I**ntegrated **M**oving **A**verages
- Only Stationary Series can be forecasted!!
- If Stationarity condition is violated, the first step is to stationarize the series.
- A stationary time series is one whose properties do not depend on the time at which the series is observed.
- A stationary series will not have any predictable pattern. It is also called White Noise.

Plot of Stationary Series

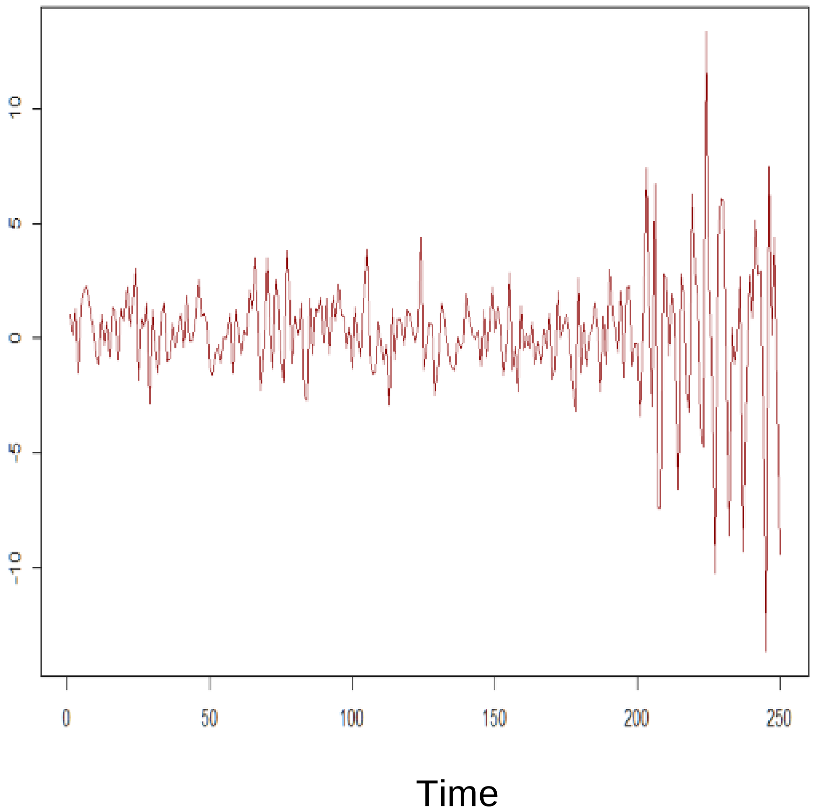


Properties of Stationary Series

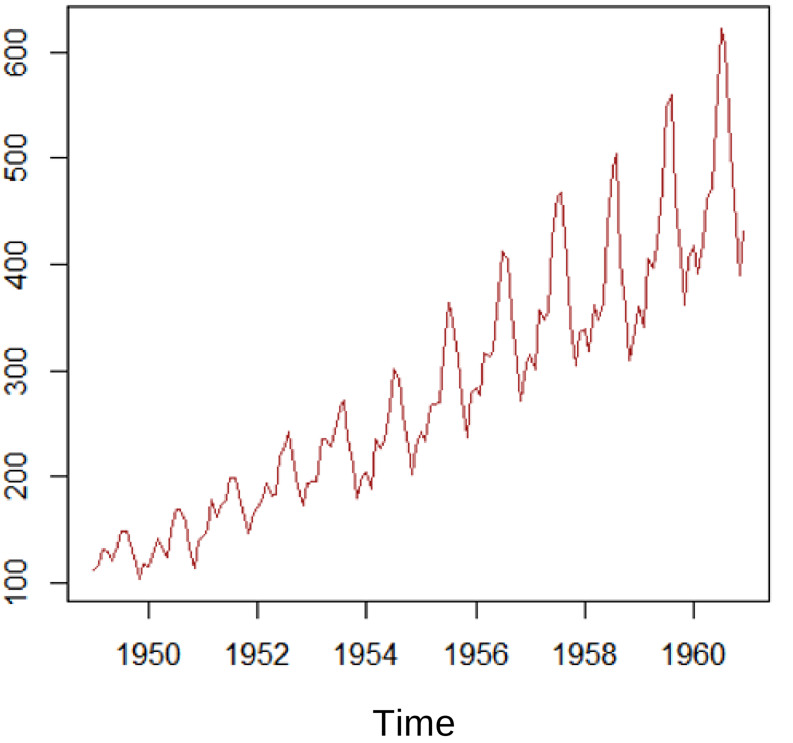
- Mean of the time series will be a constant.
  - Time series with trends, or with seasonality, are not stationary.
  - Trend and seasonality will affect the value of the time series at different times.
- Variance of the time series will be a constant
- The correlation between the t^{th} term in the series and the $t+m^{th}$ term in the series is constant for all time periods and for all m.

Examples of Non-Stationary Series

Example of non-constant variance



Example of trend and non-constant variance



ARIMA Pre-processing Steps

1. Visualization.
2. Stationarization.
   - Do a formal test of hypothesis.
   - If series non-stationary, stationarize.

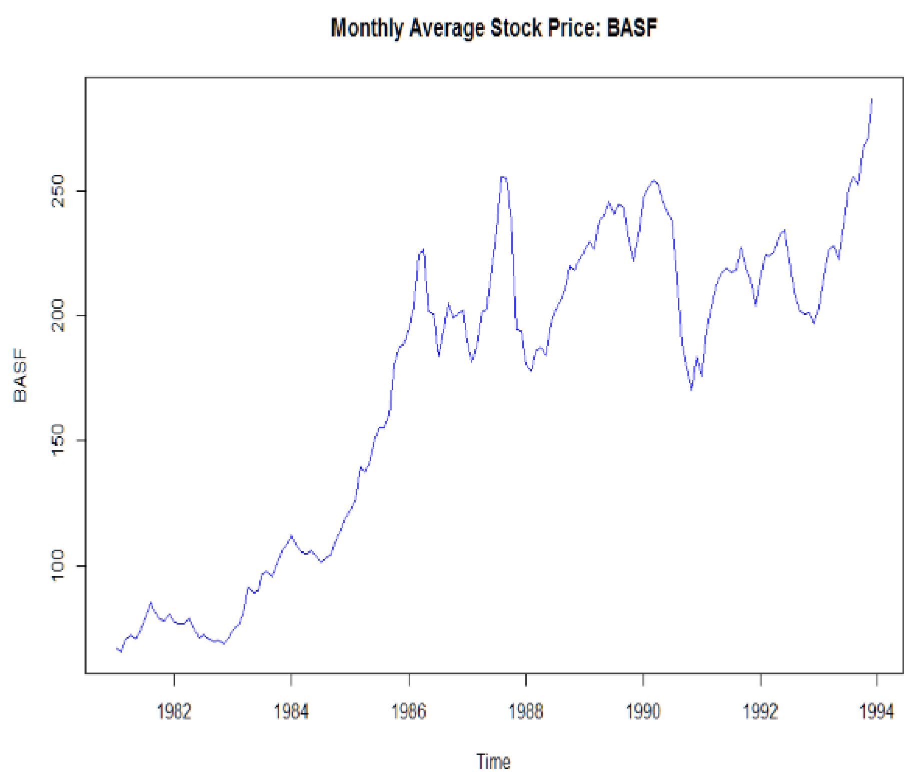
1. Explore Autocorrelations and Partial Autocorrelations.
2. Build ARIMA Model.
   - Identify training and test periods.
   - Decide on model parameters.
   - Compare models using accuracy measures.
   - Make prediction.

Test of Stationarity

- Augmented Dickey-Fuller Test.
- Tests whether a time series is NON-STATIONARY.
    - Null hypothesis H_0: Time series non-stationary.
    - Alternative hypothesis H_a: Time series is stationary.
- Rejection of null hypothesis implies that the series is stationary.

Test of Stationarity – BASF Stock Price

Test	Augmented Dicky Fuller
Python Function	adfuller()
P-Value	0.4713
Result	Do not Reject H_0
Conclusion	Series is non-stationary



Stationarizing Time Series

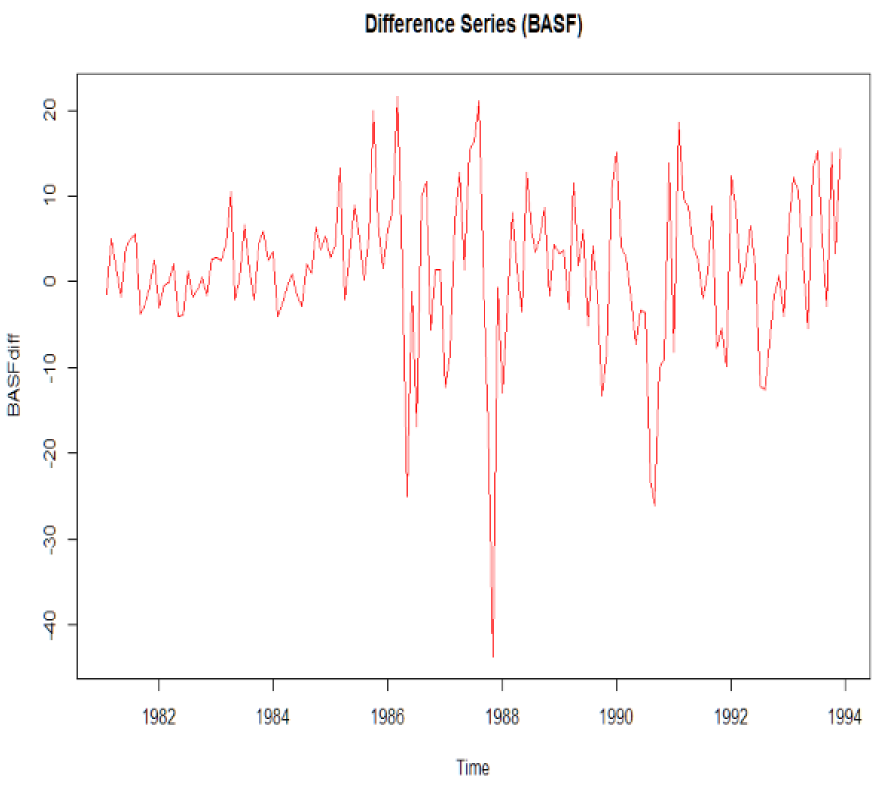
- It is possible to make a non-stationary series stationary by taking differences between consecutive observations.
- There is a simple but effective method to stationarize a time series.
  - Take difference of consecutive terms in a series.
  - Known as a Difference Series of Order 1.
- Python function diff().

BASF Stock Price

	1981 Jan	1981 Feb	1981 Mar	1981 Apr	1981 May	1981 Jun	1981 Jul	1981 Aug	1981 Sep	1981 Oct
Original	67.27	65.86	70.8	72.38	70.61	74.62	79.56	85.08	81.39	78.73
Difference		-1.4	4.94	1.57	-1.77	4.02	4.94	5.52	-3.69	-2.67

Stationarizing Time Series

Test	adfuller(BASF.diff())
p-Value	0.01
Result	Reject H_0
Conclusion	Series is Stationary



Autocorrelation

- Autocorrelation: Correlation in the series with itself. It is abbreviated by ACF.
- If the observed value at period t has correlation with values at a previous period, then autocorrelation it said to exist.
- Recall that in linear regression we made an assumption that observations are independent. This assumption is violated when autocorrelation is present.
- ACF can be of different orders.
- The correlation with immediate predecessor is called ACF of order 1 or ACF(1).
- Similarly, autocorrelation with past value 2 periods ago, is called ACF(2) and so on.
- The value of ACF ranges between -1 and 1.

$$-1 \le ACF \le 1$$

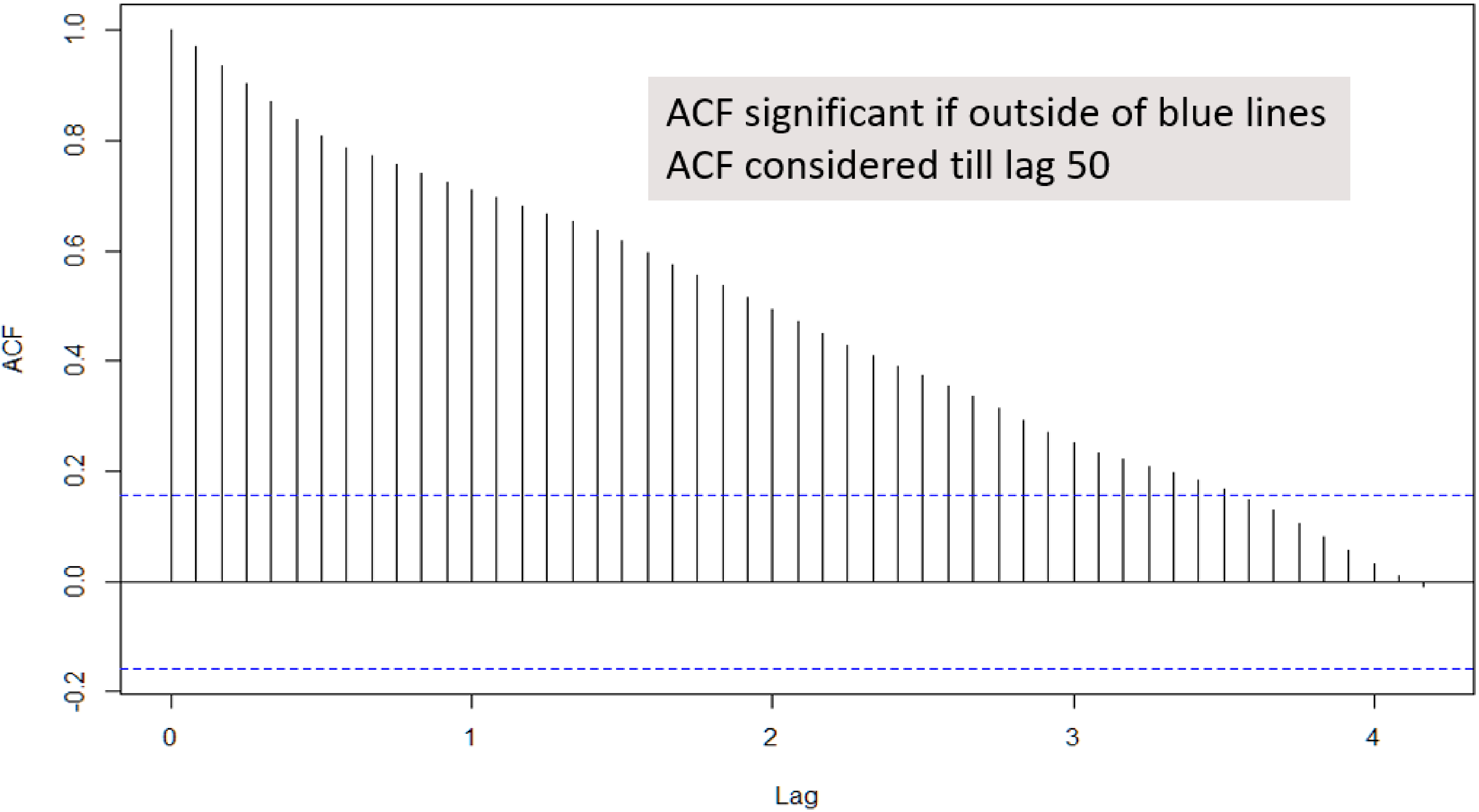
$$ACF(0) = 1$$

Autocorrelation

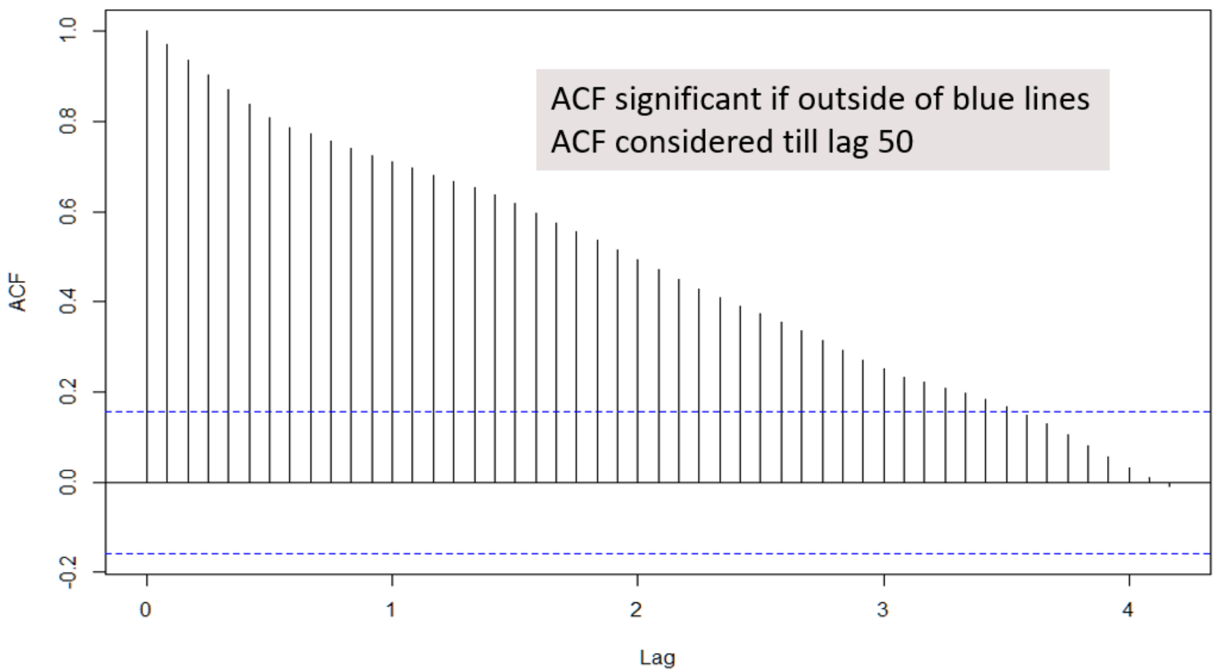
- Autocorrelation of different orders gives inside information regarding time series
- To determines order p of the series, we compute series with different lags as show below.

Year	Month	Original Series	Lag(1) Series	Lag(2) Series	Lag(3) Series
1981	Jan	67.27			
1981	Feb	65.86	67.27		
1981	Mar	70.80	65.86	67.27	
1981	Apr	72.38	70.80	65.86	67.27
1981	May	70.61	72.38	70.80	65.86
1981	Jun	74.62	70.61	72.38	70.80
1981	Jul	79.56	74.62	70.61	72.38
1981	Aug	85.08	79.56	74.62	70.61
1981	Sep	81.39	85.08	79.56	74.62
1981	Oct	78.73	81.39	85.08	79.56
1981	Nov	78.01	78.73	81.39	85.08
1981	Dec	80.55	78.01	78.73	81.39
1982	Jan	77.55	80.55	78.01	78.73
1982	Feb	77.18	77.55	80.55	78.01
1982	Mar	77.02	77.18	77.55	80.55
1982	Apr	79.12	77.02	77.18	77.55
1982	May	75.04	79.12	77.02	77.18

Order of Autocorrelation (ACF)



Plot of ACF at different lags



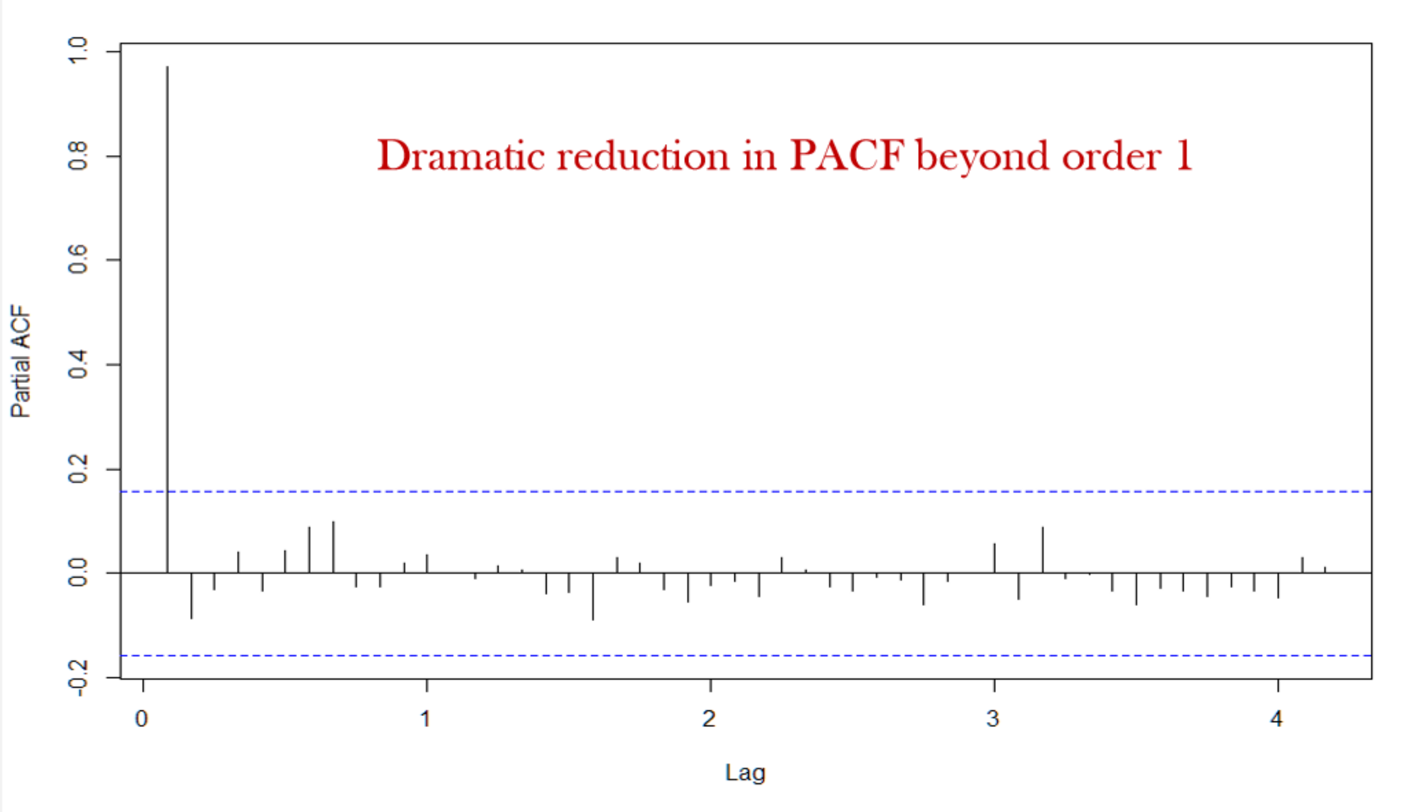
The blue lines show 95% confidence interval for the observed ACF.

- Autocorrelations are seen to decrease as lag increases.
- Autocorrelations are significant till high order 36 or so.
- Significant autocorrelations imply observations of long past influences current observation.
- Is there really such a dependency or it is only apparent?

Partial Autocorrelations (PACF)

- Partial autocorrelation adjusts for the intervening periods.
- PACF(1) = ACF(1).
- PACF(2) is the correlation between Original and Lag(2) series AFTER the influence of Lag(1) series has been eliminated.
- PACF(3) is the correlation between Original and Lag(3) series AFTER the influence of Lag(1) and Lag(2) series has been eliminated.
- PACF(50) is the correlation between Original and Lag(50) series AFTER the influence of Lag(1) through Lag(49) series has been eliminated.

Partial Autocorrelations (PACF)



- The PACF dramatically reduces the dependency depicted by ACF.
- BASF stock value depends on the immediate previous day's value

First order Autocorrelation

- The order of autocorrelation is the number of successive previous time intervals that the present value is related to.
- In case of first order autocorrelation the regression equation may be written as

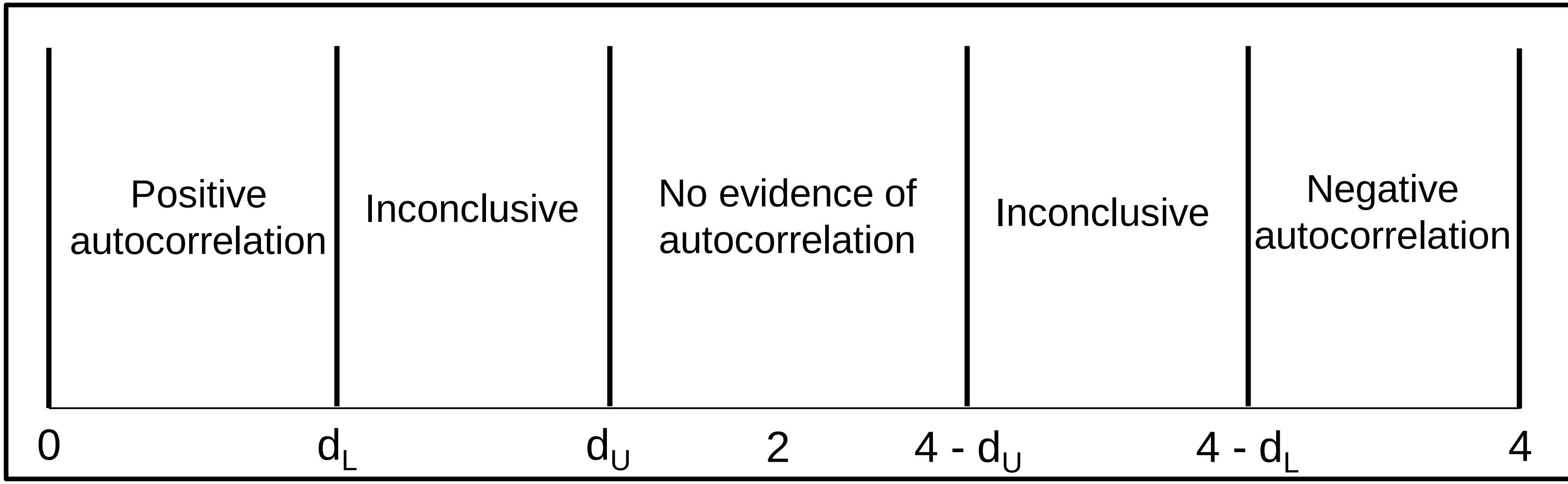
$$Y_t = \beta_0 + \beta_1 X_t + \varepsilon_t$$

with

$$\varepsilon_t = \rho \varepsilon_{t-1} + v_t$$

- If $\rho > 0$, we have +ve autocorrelation and if $\rho < 0$, we have –ve autocorrelation.
- The Durbin-Watson statistic computed as below may be used to test for presence of first order autocorrelation
- Durbin-Watson $$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum e_t^2}$$

Durbin-Watson test



Significant points of dL and du: α = 0.05

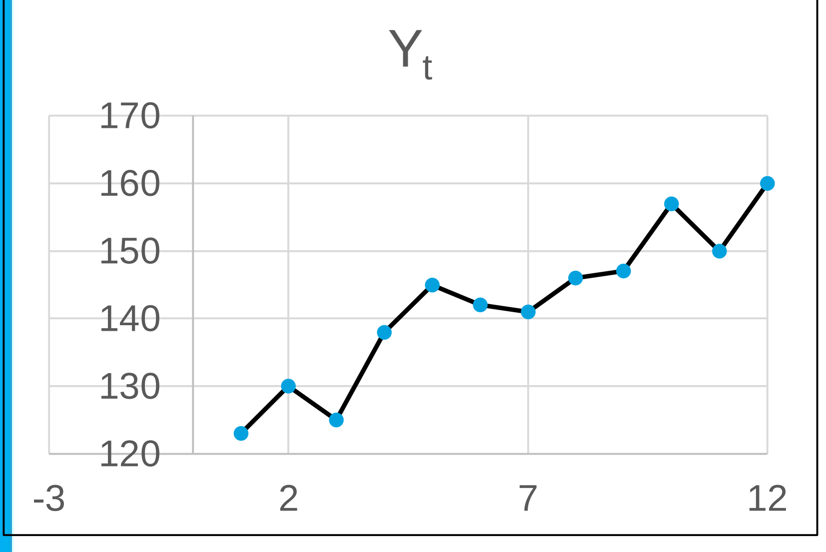
	Number of Independent Variables									
	1		2		3		4		5	
n	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	.95	1.54	0.82	.175	0.69	1.97	0.56	2.21
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89

Computing Auto-Correlations

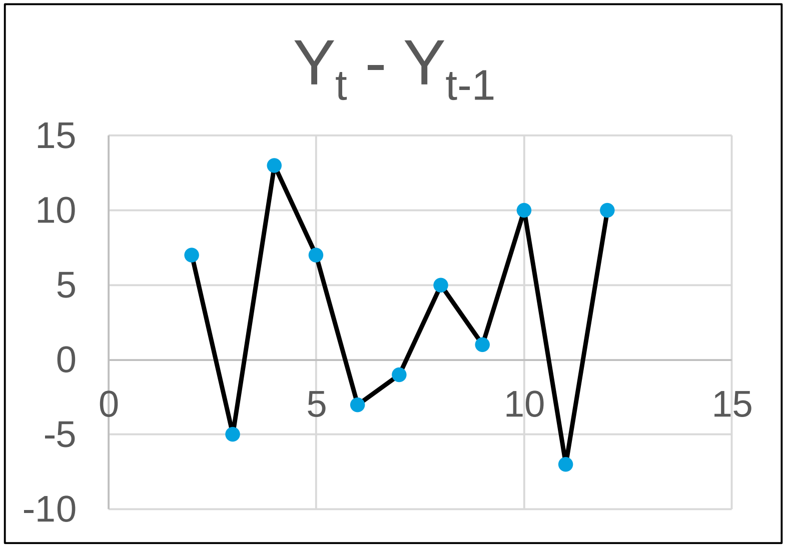
The formula for computing Lag k autocorrelation is

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum(Y_t - \bar{Y})^2} \qquad SE_{r_1} = \frac{1}{\sqrt{n}}$$

Time Series with Trend



Time Series De-Trended

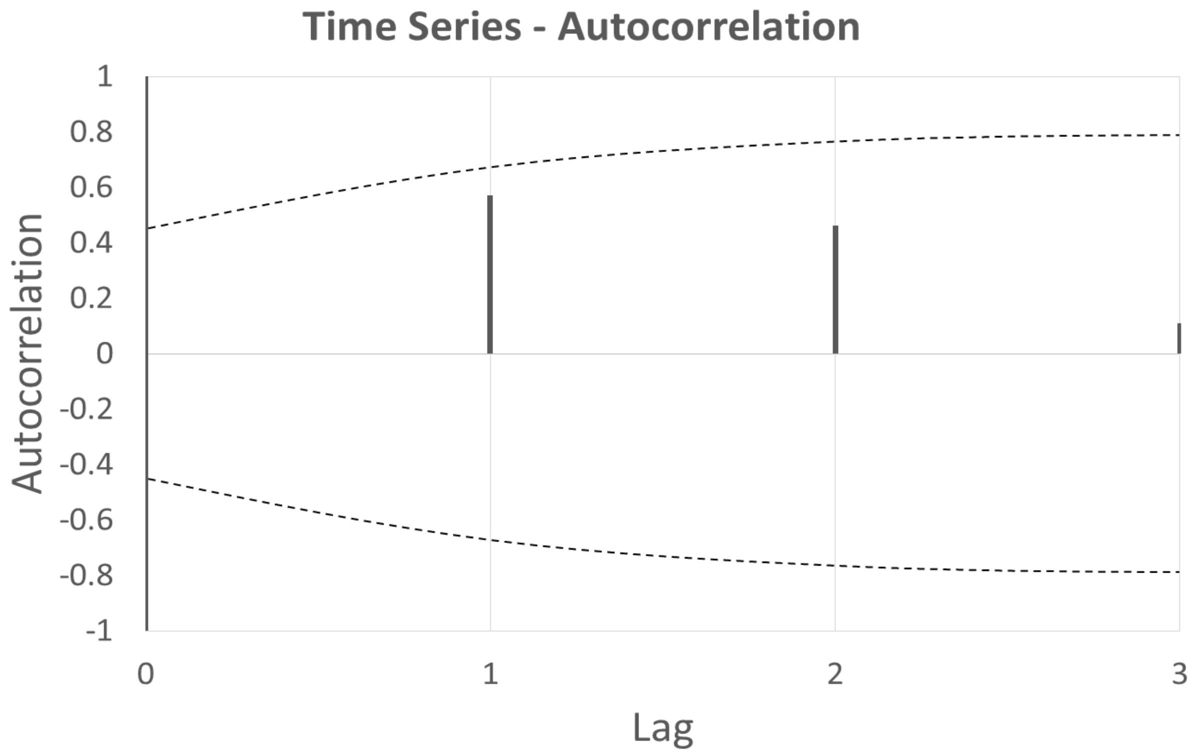


The Correlogram

Lag	ACF	T	LBQ*
1	0.571	1.98	5.00
2	0.463	1.25	8.59
3	0.110	0.27	8.82

* LBQ = Ljung-Box Q

$$Q = n(n+2)\sum_{k=1}^{m}\frac{r_k^2}{n-k}$$

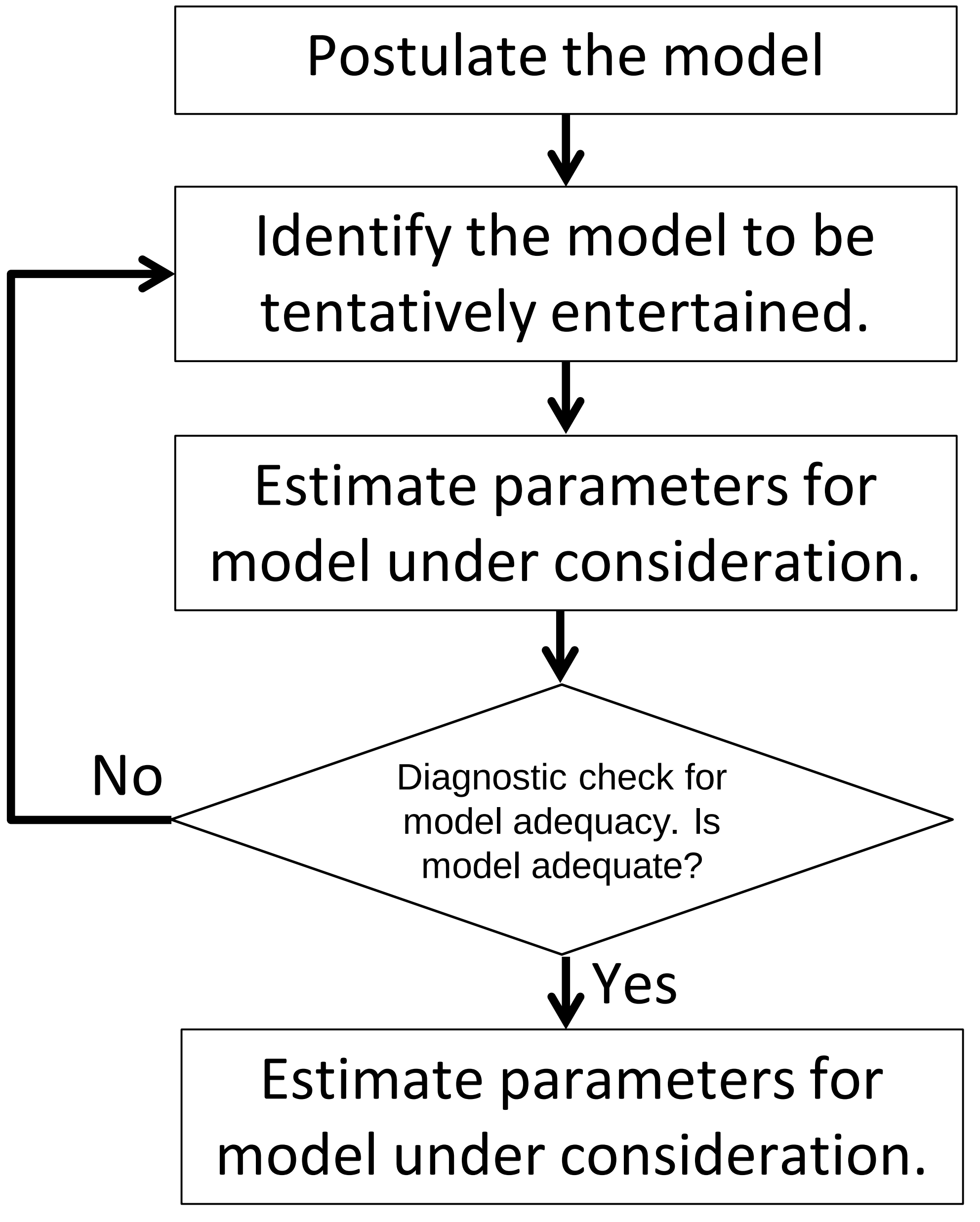


ARIMA

ARIMA

- The term ARIMA stands for Autoregressive Integrated Moving Average.
- It is also known as Box-Jenkin Methodology.
- The methodology does not assume any particular pattern and can be applied to any time series data.
- It is an iterative approach - i.e. The model parameters are varied and several alternative fits are generated.
- The residuals are analysed for best fit such that:
  - The residuals are randomly distributed.
  - The mean square error is minimum.

ARIMA Flow Diagram



Autoregression

- In autoregression, the present value of the series is regressed with the previous values within the series.
- It can depend on several past observations - AR(p) process–p: parameter (to be determined from data)
- $Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 Y_{t-3} + \ldots + \beta_p Y_{t-p} + \varepsilon_t$
- $\beta_1, \beta_2, \ldots \beta_p$: autoregressive parameters of various orders
- $\varepsilon(t)$: White noise, iid r.v. with mean 0, variance σ^2

Autoregressive model

<u>AR(p) Process</u>

- Y_t depends only on its own past values Y_{t-1}, Y_{t-2}, Y_{t-3},
- A common representation: Y_t depends on p of its past values called as AR(p) model

$$Y_t = \mu + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \alpha_3 Y_{t-3} + \ldots.. + \alpha_p Y_{t-p} + \varepsilon_t$$

- Example: This year's income depends on past incomes

Moving Average Model

MA(q) Process

- Y_t depends only on the random error terms which follow a white noise process
- A common representation: Y_t depends on q of its past values called as MA(q) model

$$Y_t = \mu + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} - \ldots - \beta_q \varepsilon_{t-q}$$

- Example: Number of patients discharged on a day depends on how many patients get admitted and stay for one day, two days or three days.

Integrated

I(d) Process

- A non-stationary series is made stationary by differencing it as many number as the number of previous (lagged) values it depends on.
- If the series is dependent only on immediate previous value i.e. lag is 1 then it is i(1) process.
- When the series depends on previous two values it is differenced twice. A series is diffenced d times until it becomes stationary.
- The stationarity can be checked by Dicky-Fuller test or by plotting ACF, PACF correlograms.

ARIMA

- AR: autoregressive
  - weighted moving average over past observations.
- I: integration
  - linear trend.
- MA: moving average
  - weighted moving average over past errors.

Initial ARIMA model identification

Process	ACF	PACF
White noise	No significant Spikes	No significant Spikes
Non Stationary	Spikes damp out very slowly	
AR(p)	Damps out	Spikes cut off after lag p
MA(q)	Spikes cut off after lag q	Damps out
ARMA(p,q)	Damps out	Damps out

Model Selection Criteria

Data Condition			Model Choice
Seasonality	Trend	Autocorrelation	
Absent	Absent	Absent	Simple Moving Average Simple Exponential Smoothing
Absent	Present	Absent	Linear Regression Holt's Model
Present	Present	Absent	Decomposition Holts-Winter's Model
Absent	Absent	Present	ARIMA
Present	Present	Present	Decomposition ARIMA