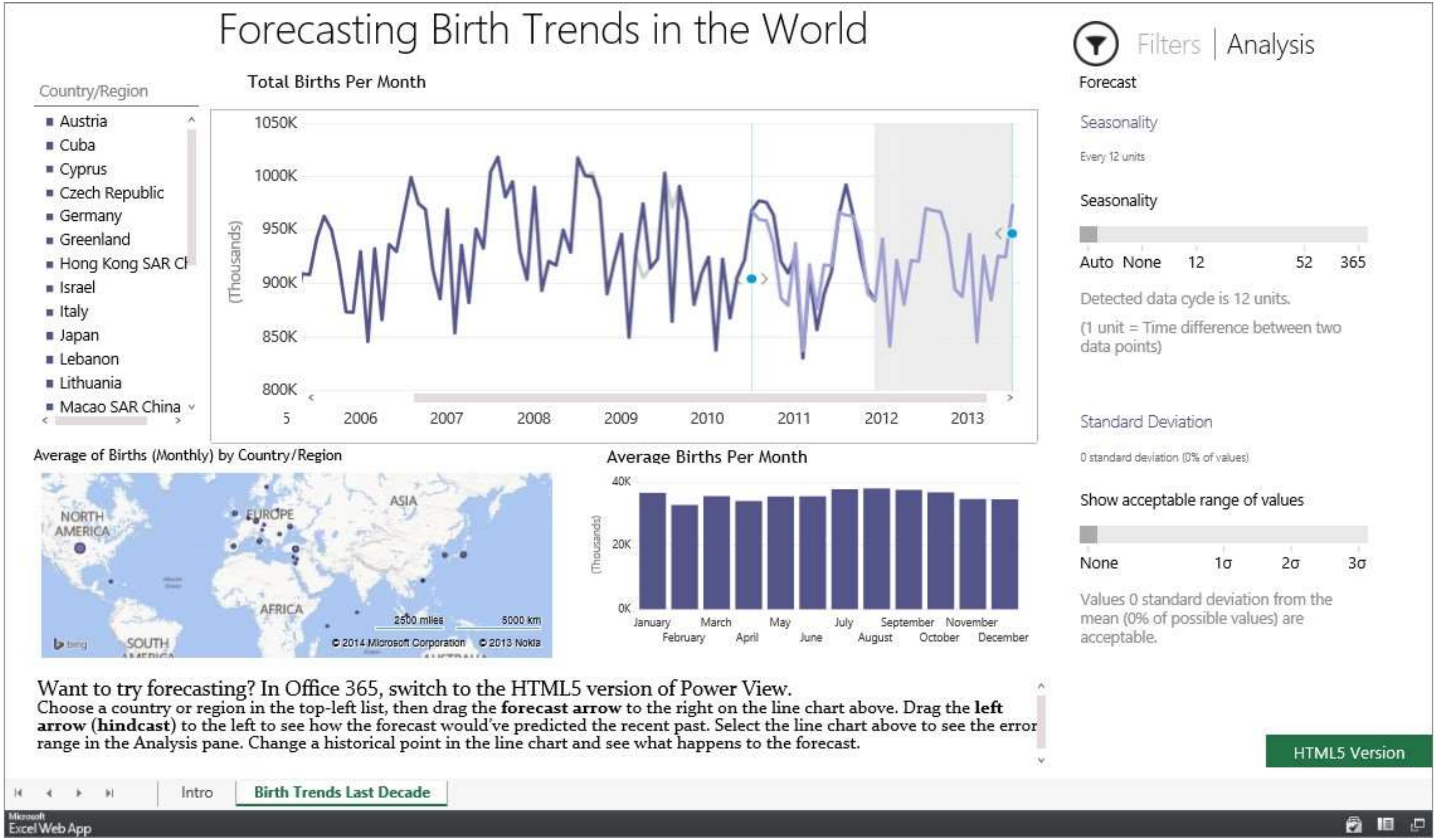
TIME SERIES

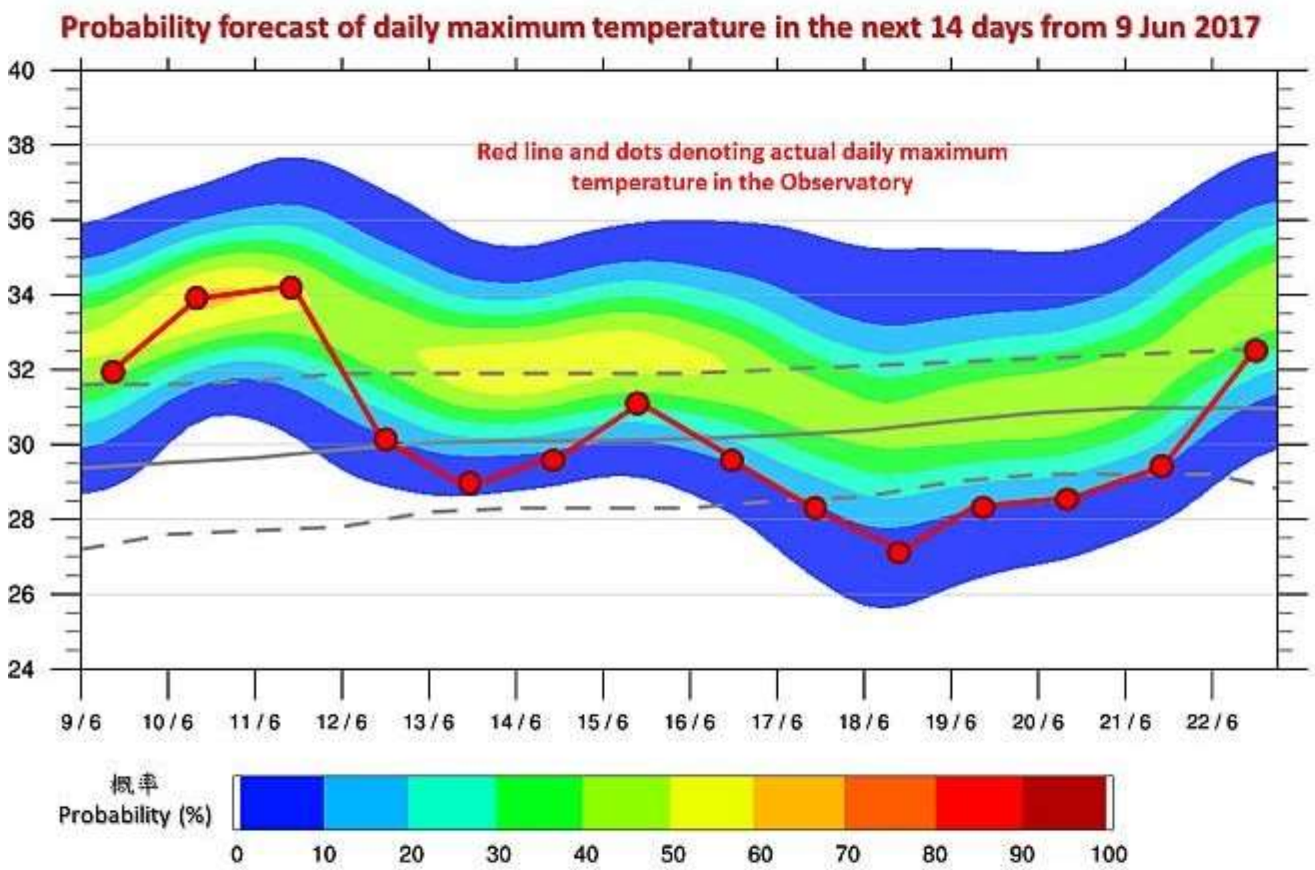
Birth Forecast

proschool
An IMS Initiative

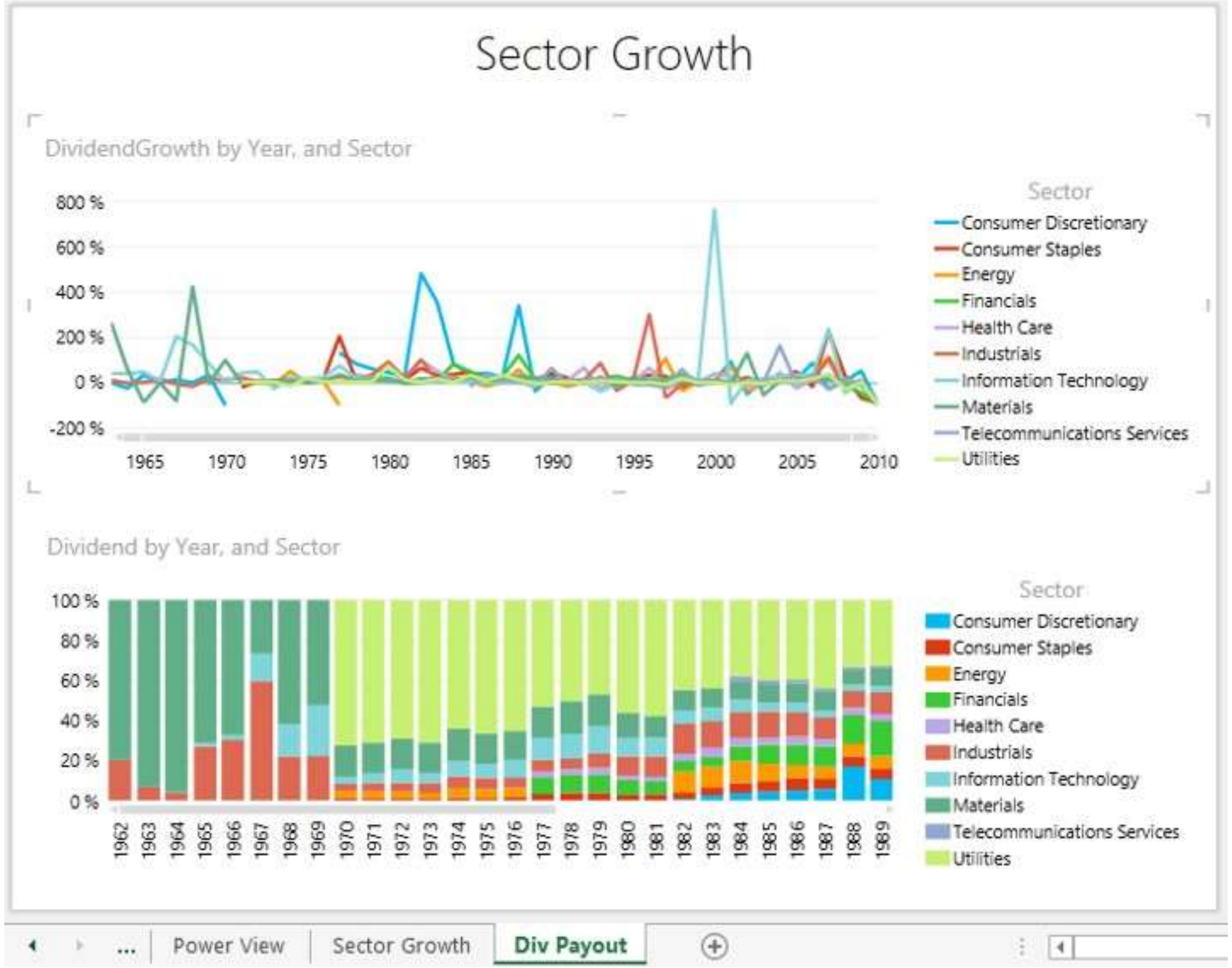


Southern California Daily Energy Report

Daily Maximum Temperature



Sector Growth

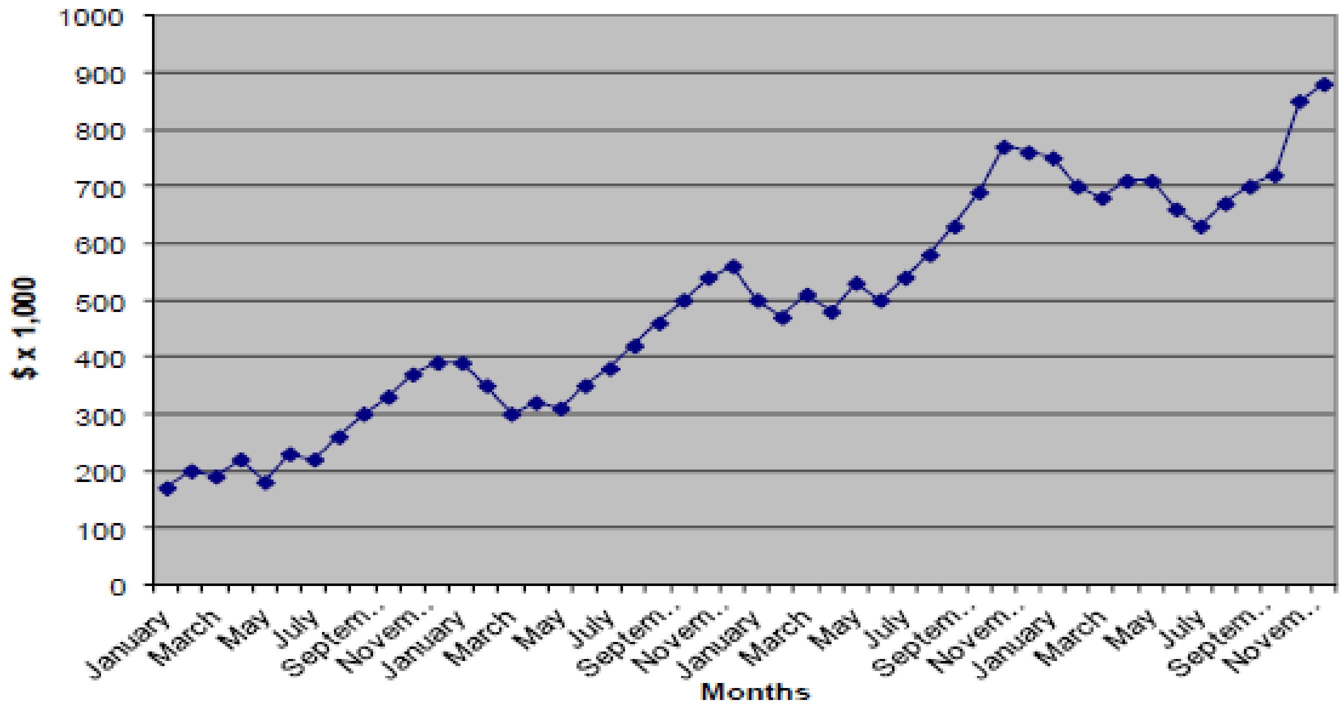


What is Time Series Forecasting?

- **Time series forecasting** is the use of a model to predict future values based on previously observed values.
- A **time series** is a series of data points indexed or listed or graphed in time order.
- Forecasting involves taking models fit on historical data and using them to predict future observations.
- The purpose of time series analysis is generally two fold to understand or model the stochastic mechanisms that gives rise to an observed series and to predict or forecast the future values of a series based on the history of that series.
- Time-Series Forecasting is the science and art of predicting events or data measurements in the future.
- Time-Series Forecasting eliminates outright guessing by using mathematics, probability and statistics.
- **Assumptions in Time Series**
- Time series data should be equally spaced over time
- Patterns of past data will propagate into future
- Can not be used to predict random events(i.e. next tsunami etc.)

Time Series Plot

- A time series plot is a two dimensional plot of variable values against time period. The vertical axis measures the variables of interest and the horizontal axis corresponds to the time periods.



Application of Time Series

- Future safety
- Sales Forecasting
- Budgetary Analysis
- Stock Market Analysis
- Process and Quality Control
- Inventory Studies
- Economic Forecasting
- Risk Analysis & Evaluation of changes.
- Census Analysis
- Profit of experience.

Components of Time Series

proschool
An IMS Initiative

- **Trend –** Secular trend or simple trend indicates the general tendency of the data to increase or decrease over a long period of time.
- **Seasonality –** Seasonal variations are the variations in time series due to rhythmic forces which operate in a repetitive, predictable and periodic manner in a time span of one year or less.
- **Cyclical –** Any pattern showing an up and down movement around a given trend is identified as a cyclical pattern. The duration of a cycle depends on the type of business or industry being analyzed.
- **Irregular –** This component is unpredictable. Every time series has some unpredictable component that makes it a random variable. In prediction, the objective is to “model” all the components to the point that the only component that remains unexplained is the random component.

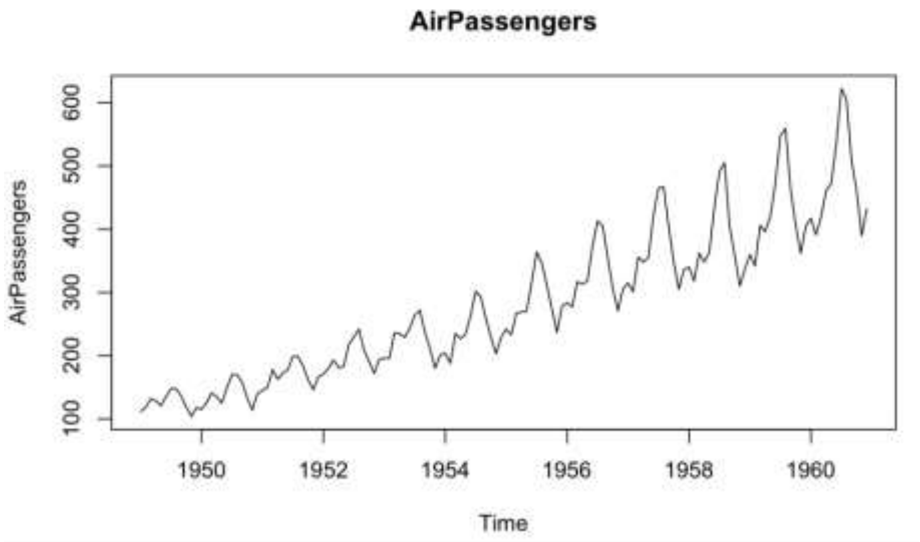
Trend

proschool
An IMS Initiative

- A trend is the general direction of a market or the price of an asset, and trends can vary in length from short to intermediate, to long term.
- The term long period of time is a relative term and cannot be defined exactly. In some cases, 2 weeks may or may not be a long period of time. For example, to control an epidemic, 1 Week is considered as a long period of time.
- The values of the time series are plotted on a graph and these values cluster more or less around a straight line then the trend shown by the straight line is termed as linear otherwise it is termed as a non linear.
- **Examples of secular upward trend are:** Prices, Pollution, literacy rate, Production, etc.
- **Examples of secular downward trend are:** Rate of infant mortality, death rate due to epidemics owing to advancements in medical facilities
- Taking averages over a certain period is a simple way of detecting trend in seasonal data.
- Change in averages with time is evidence of a trend in the given series, though there are more formal tests for detecting trend in time series.

Seasonality

- The seasonal component consists of effects that are reasonably stable with respect to timing, direction and magnitude.
- The seasonal fluctuations can be measured only if the date are recorded hourly, daily, weekly, monthly, quarterly.
- The seasonal variations the time period should not exceed one year.
- **There are three reasons for the study of seasonal variations**:
  1. We can establish the pattern of the past changes,
  2. The projection of past patterns into the future is a useful technique of prediction,
  3. The effects of seasonality can be eliminated from the time series after their presence is established.



Seasonality

- **Types of seasonal variation**
  - ✓ **Seasonal variations due to natural forces:** There are variations in time series due to weather conditions and climatic changes.

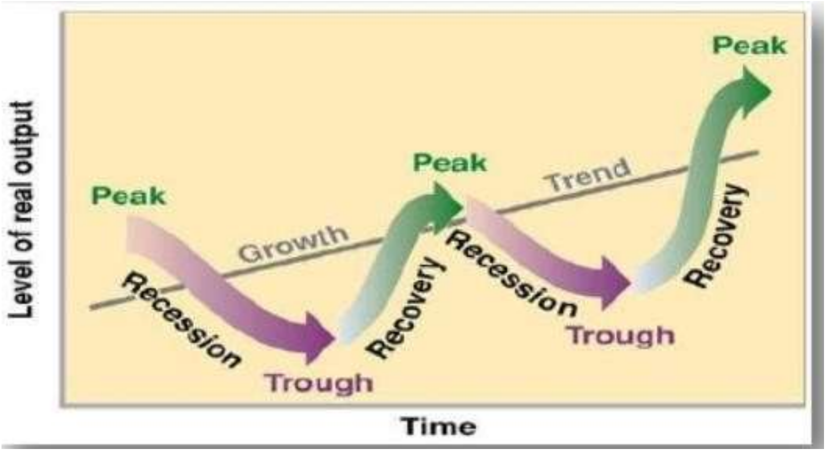
  For Example: Sales of umbrellas zoom up during rainy season, Sale of ice cream zoom up in summer.

  - ✓ **Seasonal variations due to customs:** There are variations due to customs, habits, lifestyle, and conventions of the people in society.

  For Example: Sale of sweets goes up during festival.

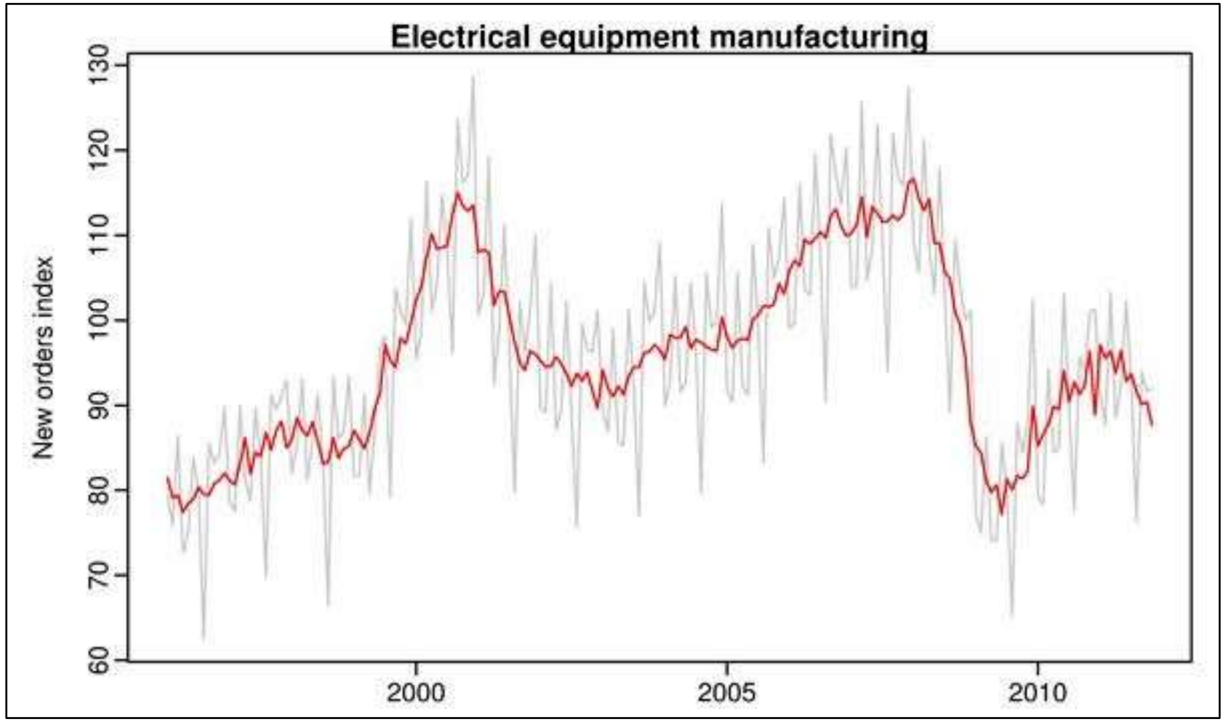
Cyclical Variation or Cyclic Fluctuations

- Time series exhibits **Cyclical Variations** at a fixed period due to some other physical cause, such as daily variation in temperature.
- **Cyclical variation** is a non-seasonal component which varies in recognizable cycle. Sometime series exhibits oscillation which do not have a fixed period but are predictable to some extent.
- The period from the peak of one boom to the peak of the next boom is called a complete cycle.
- Cyclical variations are not periodic but more or less regular in nature.
- For example, economic data affected by business cycles with a period varying between about 5 and 7 years.
- In weekly or monthly data, the **cyclical component** may describes any regular variation (fluctuations) in time series data.
- The **cyclical variation** are periodic in nature and repeat themselves like business cycle, which has four phases
    1. Peak, 2. Recession, 3. Trough/Depression, 4. Expansion.

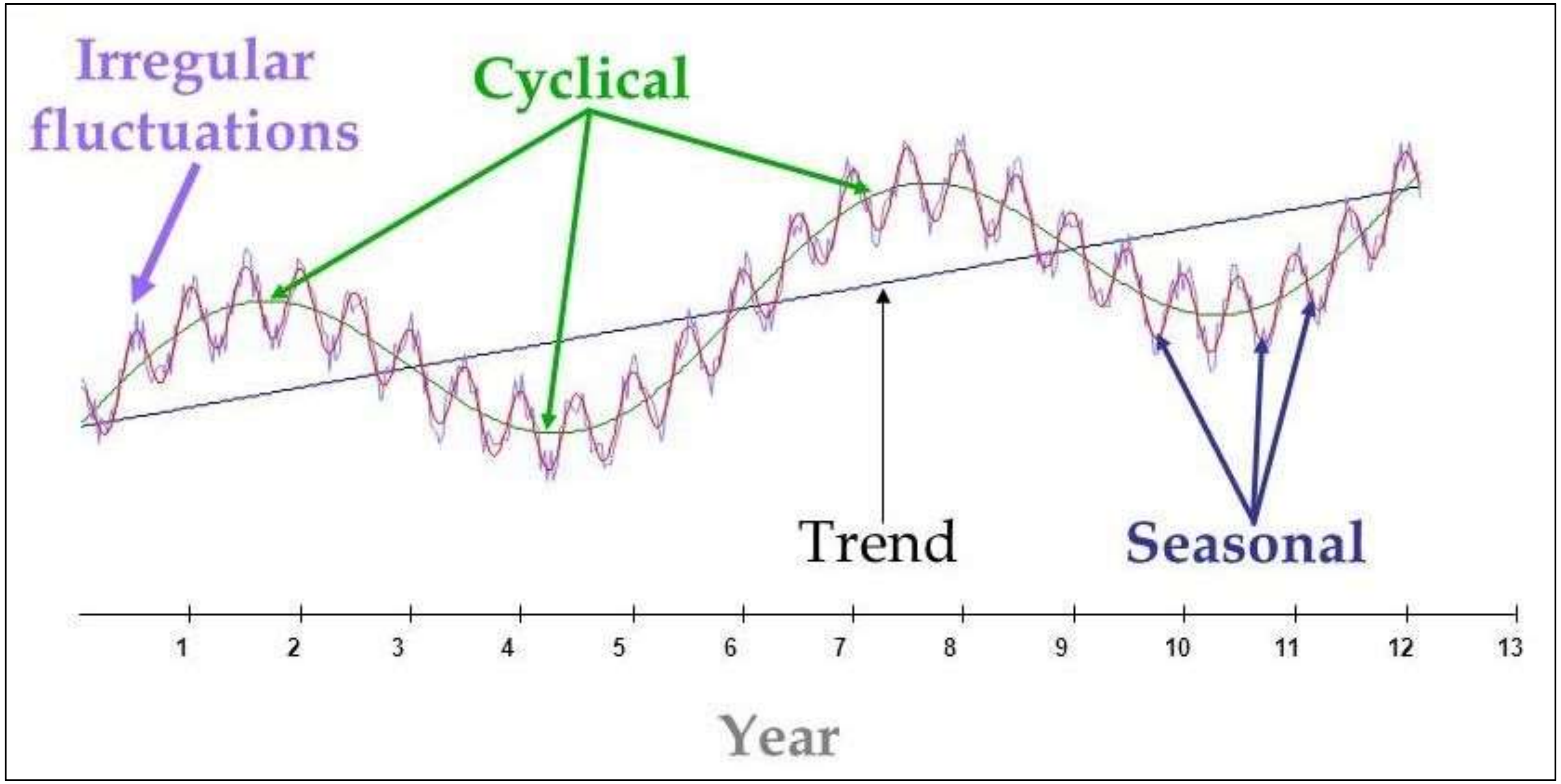


Irregular Components

- Irregular component is the residual variation remaining after the trend-cycle and seasonality have been extracted from original time series.
- This component appears as short-term, unsystematic fluctuations around the trend that do not follow any systematic or repeated pattern, which could be captured by seasonal component.
- Irregularity is formed from all unpredictable effects that affect time series, i.e.:
  - ✓ Unseasonable weather;
  - ✓ Sampling errors;
  - ✓ other errors.
- Examples of Irregular components
  - ✓ Floods,
  - ✓ Earthquakes,
  - ✓ Wars,
  - ✓ Famines

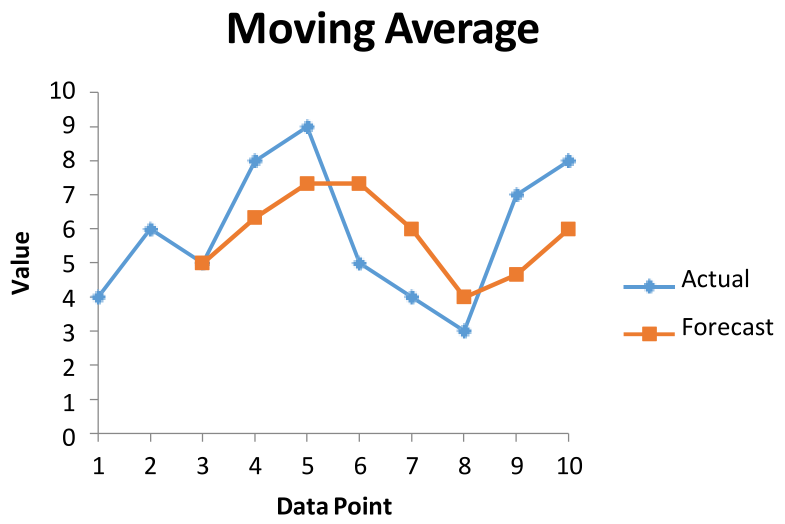


Components of Time Series



Smoothing

- Smoothing is a statistical technique for removal of short term irregularities in a time-series data to improve the accuracy of forecasts.
- Smoothing is a very common statistical process used in time series analysis.
- When you use an average to describe something, the number we are using is a smoothed number.
- For example, we observed a coldest winter. but, How do we find out this as a coldest winter?
- We observed daily low and high temperatures for winter period of each year on recorded history. Now, we have a group of numbers that jump around quite a bit. We need a number to remove all this jumping.
- The process of removing all these jumping around in the data is called smoothing. In this case we just use a simple average to accomplish the smoothing.



Smoothing

- In demand forecasting, we use smoothing to remove random variation (noise) from our historical demand.
- This allows us to better identify demand patterns (primarily trend and seasonality) and demand levels that can be used to estimate future demand
- The most common way to remove noise from demand history is to use different statistical methods to normalize the effect of time series component:
  - ✓ Simple Moving Average
  - ✓ Weighted Moving average.
  - ✓ Simple Exponential Smoothing
  - ✓ Holts Exponential Smoothing/ Double Exponential Smoothing
  - ✓ Winters Exponential Smoothing/Holts Winter Smoothing/ Triple Exponential Smoothing

Simple Moving Average Example 1

Bharat Heavy Electricals Ltd. was incorporated as a government – owned organization in 1959 to domestically manufacture power plant equipment's. Initially, BHEL was predominantly a power equipment company engaged in the manufacture of steam turbines, generators, transformers and motors. Now, the company wants to forecast the sales for coming year based on the past data. So, we calculate a five-year moving average from the following data set:

Year	Sales ($M)
2003	4
2004	6
2005	5
2006	8
2007	9
2008	5
2009	4
2010	3
2011	7
2012	8

Simple Moving Average

What is a Simple Moving Average?

- A simple moving average (SMA) is an arithmetic moving average calculated by adding the number of time periods and then dividing this total by the number of time periods.
- A simple moving average smoothens out volatility, and makes it easier to view the trend.
- If the simple moving average goes up, this means that the trend is increasing and If it is going down it means that the trend is decreasing.
- The longer the timeframe for the moving average, the smoother the simple moving average.
- A short term moving average is more volatile, but its reading is closer to the source data.

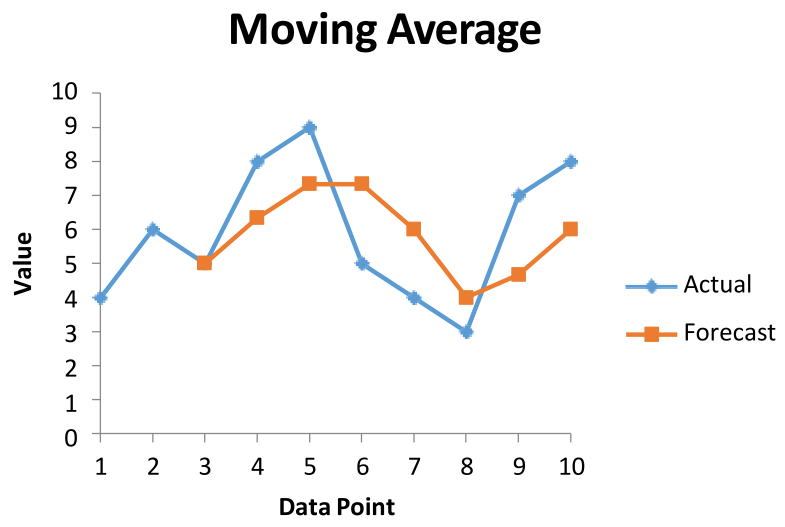
Analytical Significance

- Moving averages are an important analytical tool used to identify trends and the potential for a change in an established trend.

Simple Moving Average

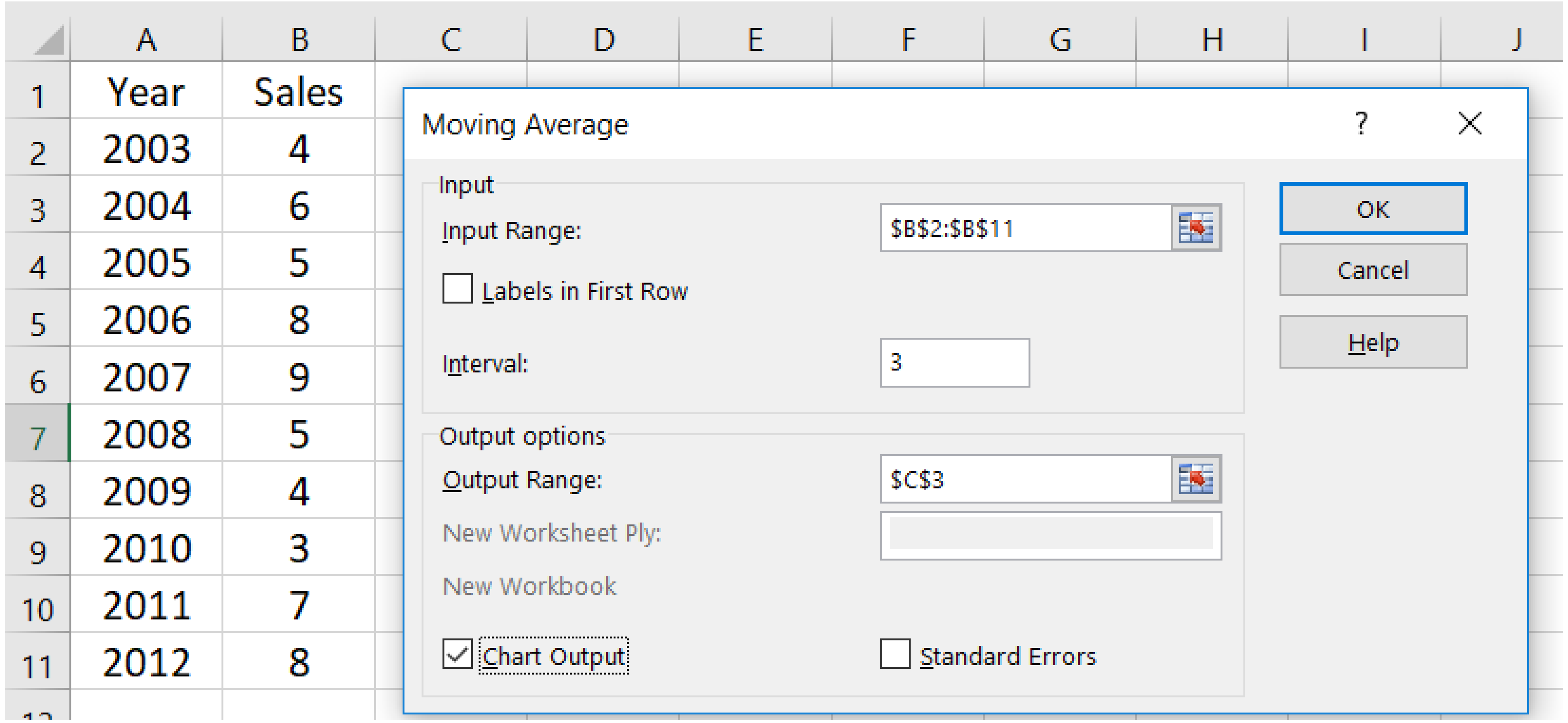
- The Moving average sales for 2007 is calculated by finding the average for the first five year (2003-2007).

Year	Sales	Simple Moving Average Process	Simple Moving Average
2003	4		
2004	6		
2005	5		
2006	8	$=\frac{4+6+5}{3}$	5.00
2007	9	$=\frac{6+5+8}{3}$	6.33
2008	5	$=\frac{5+8+9}{3}$	7.33
2009	4	$=\frac{8+9+5}{3}$	7.33
2010	3	$=\frac{9+5+4}{3}$	6
2011	7	$=\frac{5+4+3}{3}$	4
2012	8	$=\frac{4+3+7}{3}$	4.67
		$=\frac{3+7+8}{3}$	6



Steps to calculate Simple Moving Average

- Data -> Data Analysis -> Moving Average -> Select Input range -> interval



	A	B
1	Year	Sales
2	2003	4
3	2004	6
4	2005	5
5	2006	8
6	2007	9
7	2008	5
8	2009	4
9	2010	3
10	2011	7
11	2012	8

Advantages and Disadvantages

Advantages of Simple Moving Average

- Easily understood
- Easily computed
- Provides stable forecasts

Disadvantages of Simple Moving Average

- Requires saving lots of past data points: at least the N periods used in the moving average computation
- Lags behind a trend
- Ignores complex relationships in data

Weighted Moving Average

- The WMA is a weighted average of the last n periods, where the weighting decreases with each previous period.
- In SMA we have given equal weighting to the sales. For example, in the 3 year moving average (ref: previous example), each year represented 33.33% of the moving average.

$$\hat{Y}_{t+1} = w_1 Y_t + w_2 Y_{t-1} + \cdots + w_k Y_{t-k-1} \qquad \text{where } 0 \le w_i \le 1 \text{ and } \sum w_i = 1$$

- When using demand history to project future demand, it’s logical to come to the conclusion that you would like more recent history to have a greater impact on your forecast.
- We can adapt our moving-average calculation to apply various “weights” to each period to get our desired results. We express these weights as percentages, and the total of all weights for all periods must add up to 100%.

Weighted Moving Average

- Therefore, if we decide we want to apply 45% as the weight for the nearest period in our 3 – year "weighted moving average", we can subtract 45% from 100% to find we have 55% remaining to split over the other 2 periods.
- For example, we may end up with a weighting of 32% & 23% respectively for the 2 years (45 + 32 + 23)= 100).

$$\hat{Y}_{t+1} = 0.15Y_t + 0.2Y_{t-1} + 0.3Y_{t-2} + 0.35Y_{t-3}$$

- The weighted moving average is more customizable than the SMA and EMA.
- The most recent price points are given more weight, but it also work the other way, where historical prices are given more weight

Weighted Moving Average Example

proschool
An IMS Initiative

- Let's continue example 1 using excel, consider a weight for year 2003 to 2005 is 0.23, 0.32 and 0.45.

Year	Sales	3Yr WMA Sales	3Yr WMA Sales
2003	4		
2004	6		
2005	5		
2006	8	=0.45*5 + 0.32*6 + 0.23*4	5.09
2007	9	=0.45*8 + 0.32*5 + 0.23*6	6.58
2008	5	=0.45*9 + 0.32*8 + 0.23*5	7.76
2009	4	=0.45*5 + 0.32*9 + 0.23*8	6.97
2010	3	=0.45*4 + 0.32*5 + 0.23*9	5.47
2011	7	=0.45*3 + 0.32*4 + 0.23*5	3.78
2012	8	=0.45*7 + 0.32*3 + 0.23*4	5.03
		=0.45*8 + 0.32*7 + 0.23*3	6.53

Weighted Moving Average Advantages and Disadvantages

Advantages

- It smooths out random noise in the data
- It is useful for trending data with an ideal value of n

•It is flexible since weights of periods are arbitrary (WMA)

Disadvantages

- For a large value of n, the number of time period are averaged and can over smooth the data.
- For a small value of n, can make the moving average very sensitive.
- Moving averages lag the actual values.
- Historical data should be accurate.

Simple Exponential Method

- Assume, in a weighted average where we consider all of the data points, while assigning exponentially smaller weights as we go back in time.
- For example, if we started with a weight of 0.8, our weights will go back in time it approaching a big zero.

Weights in exponential Smoothing	(0.8)^1	(0.8)^2	(0.8)^3	(0.8)^4	(0.8)^5	(0.8)^6	(0.8)^7	……….
	0.8	0.64	0.512	0.4096	0.3277	0.26214	0.209715	………….
Weights in Weighted moving average	.45	.32	.23					

- In some way exponential smoothing is very similar to the weighted moving average, only the weights dictated are decaying uniformly. The smaller the starting weight, the faster it approaches zero.
- The only problem with this is that the weights do not add up to 1.
- If we take a sum of 3 weights in weighted moving average it will sum to 1
- But in this case the sum of first 3 weights is 1.952 its not equal to 1.

Simple Exponential Smoothing

- To solve the above problem Poisson, Holts or Roberts derived a formula

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

- Initialize L1 is equal to forecast at some point

F1 = L1 = Y1.

The components of the formula

- $\boldsymbol{\alpha}$ = Smoothing coefficient or smoothing factor
- Y_t = The Actual sale of the particular year
- L_{t-1} = The predicted sale of the previous year

Assumptions:

- Simple Exponential Smoothing assumes that our series only contains level and error term.
- If we consider level(L_t) only, then the assumption of the exponential smoothing is that this level at time t will be "stay put" and does not move.

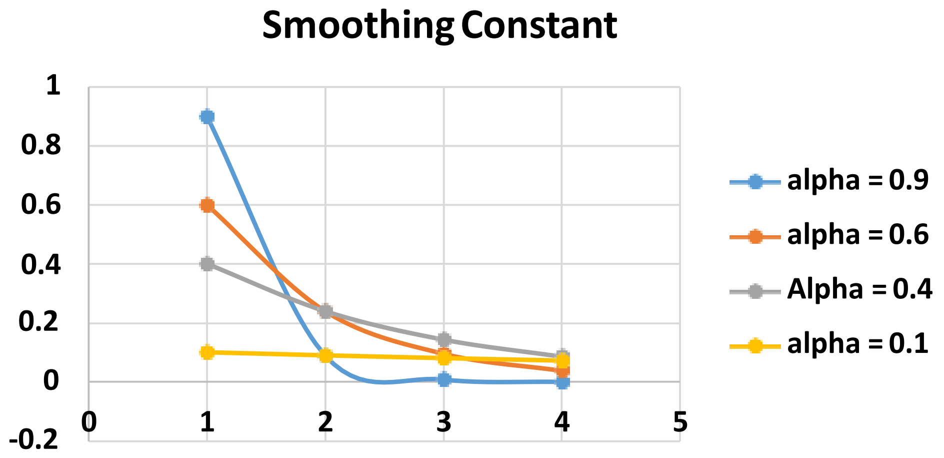
Forecast = Estimated level at most recent time point

$$\mathbf{F_{t+k} = L_t}$$

Smoothing Constant(α)

- Smoothing constant(α) = 1 : The past values have no effect on forecast (Under-Smoothing)
- Smoothing constant(α) = 0 : The past values have equal effect on forecast (Over-Smoothing)
- In this case, we can't give more weight to more recent information Hence, we are typically choosing values between 0 and 1.
- Lets look at a table that shows the different values of α along with the most recent and older periods.
- We have shown a line graph for different value of α.
- From the graph we can observe that lines are decaying exponentially as we are going toward older periods.
- For large value of α i.e.0.9 decay is very fast as compared to smaller value of α

alpha	alpha(1-alpha)	alpha(1-alpha)^2	alpha(1-alpha)^3
0.9	0.09	0.009	0.0009
0.6	0.24	0.096	0.0384
0.4	0.24	0.144	0.0864
0.1	0.09	0.081	0.0729
1	2	3	4



Exponential Smoothing in Excel

- In the below example the smoothing coefficient is 0.5
- $\mathbf{L_t = \alpha Y_t + (1 - \alpha) L_{t-1}}$
- For example, If we have to calculate the sales for year 2004 ,2005 and 2006.
- To calculate the forecast lets consider the first value of forecast column same as the sales i.e. 4.
- For 2004, $\boldsymbol{\alpha}$ = 0.5, $\mathbf{Y_t}$ is the sale of 2003, $\mathbf{L_{t-1}}$ is the forecast value (4)
- For 2005, $\boldsymbol{\alpha}$ = 0.5, $\mathbf{Y_t}$ is the sale of 2004, $\mathbf{L_{t-1}}$ is the forecast value of year 2004
- For 2006, $\boldsymbol{\alpha}$ = 0.5, $\mathbf{Y_t}$ is the sale of 2005, $\mathbf{L_{t-1}}$ is the forecast value of year 2005

Year	Sales ($M)	Forecast	Exponential Moving Average
2003	4	4	
2004	6	0.5*4+(1-0.5)*4	4
2005	5	0.5*6+(1-0.5)*4	5
2006	8	0.5*5+(1-0.5)*5	5
2007	9	0.5*8+(1-0.5)*5	6.5
2008	5	0.5*9+(1-0.5)*6.5	7.75
2009	4	0.5*5+(1-0.5)*7.75	6.375
2010	3	0.5*4+(1-0.5)*6.375	5.188
2011	7	0.5*3+(1-0.5)*5.1875	4.094
2012	8	0.5*7+(1-0.5)*4.09375	5.547

Performance Metrics of forecasting method

- How do we know which forecasting method is better?
- There are difference performance metrics will tell us which forecasting method to use

MAPE –

- The **Mean Absolute Percentage Error** (**MAPE**) or **Mean Absolute Percentage Deviation** (**MAPD**) is a measure of prediction accuracy of a forecasting method in statistics,
- It usually expresses accuracy as a percentage, and is defined by the formula:

$$MAPE(M) = \frac{1}{n} \sum \frac{|Actual - Forecast|}{|Actual|} * 100$$

- MAPE is scale sensitive
- It is not useful for low volume data
- MAPE is undefined when actual demand is zero. Furthermore, when the actual value is quite small, it will take on extreme values.
- This scale sensitivity renders the MAPE close to an error measure for low-volume data.

MAPE

		A	B	C = A – B	D = C/A	F = |D|
	Year	Sales ($M)	Predicted value	Error	Error / Actual Sales	Absolute of Error/Actual Sales
	2003	4	4	0	0	0
	2004	6	4	2	0.3333	0.3333
	2005	5	4.2	0.8	0.16	0.16
	2006	8	4.28	3.72	0.465	0.465
	2007	9	4.652	4.348	0.4831	0.4831
	2008	5	5.0868	-0.09	-0.017	0.01736
	2009	4	5.07812	-1.08	-0.27	0.2695
	2010	3	4.970308	-1.97	-0.657	0.6568
	2011	7	4.773277	2.227	0.3181	0.3181
	2012	8	4.995949	3.004	0.3755	0.3755
G	**Sum**					3.079

N = Total number of Observations
= 10

Alpha	0.1
Total Absolute Error/ Actual Sales	3.079
MAPE (G/ N)*100	31%

Performance Metrics of forecasting method

RMSE –

- Root Mean Square Error (RMSE) is the standard deviation of the prediction errors(Residuals).
- Residuals are a measure of how far data points are from the regression line;
- RMSE is a measure of how these residuals are spread out.
- It tells you how concentrated the data is around the line of best fit.
- RMSE is commonly used in forecasting, and regression analysis to verify experimental results.

$$RMSE = \left[\sum_{i=1}^{N} (Forecast - Actual)^2 / N\right]^{1/2}$$

RMSE

		A	B	C = A – B	D = C*C			
	Year	Sales ($M)	Predicted value	Error	Square of Error		N = Total number of Observations = 10	
	2003	4	4	0	0		Sum of Square of Error	56.42
	2004	6	4	2	4		Alpha	0.1
	2005	5	4.2	0.8	0.64		RMSE (Sqrt(E/N))	2.38
	2006	8	4.28	3.72	13.84			
	2007	9	4.652	4.348	18.91			
	2008	5	5.0868	-0.09	0.008			
	2009	4	5.0781	-1.08	1.162			
	2010	3	4.9703	-1.97	3.882			
	2011	7	4.7733	2.227	4.958			
	2012	8	4.9959	3.004	9.024			
E	Sum				56.42			

Performance Metrics of forecasting method

MSE –

- Mean Squared Error (MSE) is a measure of dispersion of forecast errors
- The smaller the value of MSE, model is more stable
- The MSE value is misleading, more noticeable MSE will lead to large error term.

$$MSE = \frac{1}{n} \sum (Actual - Predicted)^2$$

MSE

		A	B	C = A – B	D = C*C			
	Year	Sales ($M)	Predicted value	Error	Square of Error		N = Total number of Observations = 10	
	2003	4	4	0	0		Sum of Square of Error	56.4181
	2004	6	4	2	4		Alpha	0.1
	2005	5	4.2	0.8	0.64		MSE (E/N)	5.64
	2006	8	4.28	3.72	13.8384			
	2007	9	4.652	4.348	18.9051			
	2008	5	5.0868	-0.0868	0.0075			
	2009	4	5.0781	-1.0781	1.1623			
	2010	3	4.9703	-1.9703	3.8821			
	2011	7	4.7733	2.2267	4.9583			
	2012	8	4.9959	3.0041	9.0243			
E	Sum				56.4181			

Performance metrics for different values of α

- Lets check the performance metrics for different values of α.
- We can see that the MAPE is low for minimum (0.1) and maximum values (0.9) of α.
- Here, we have found MAPE for 4 different values of α. The objective of finding different α value is to minimize the error (MAPE/RMSE etc). Therefore, instead of doing this iteration multiple times to find the least error , we can directly use the "R" function: Holt-Winter. In this function, R automatically selects the α value which minimizes the error.

Alpha = 0.1	MAPE	31%
	RMSE	2.38
	MSE	56.40%
Alpha = 0.3	MAPE	36%
	RMSE	2.27
	MSE	51.30%
Alpha = 0.6	MAPE	36%
	RMSE	2.27
	MSE	51.50%
Alpha = 0.9	MAPE	32%
	RMSE	2.24
	MSE	50.10%

Steps to calculate EMA using excel Data Analysis Tool Pack

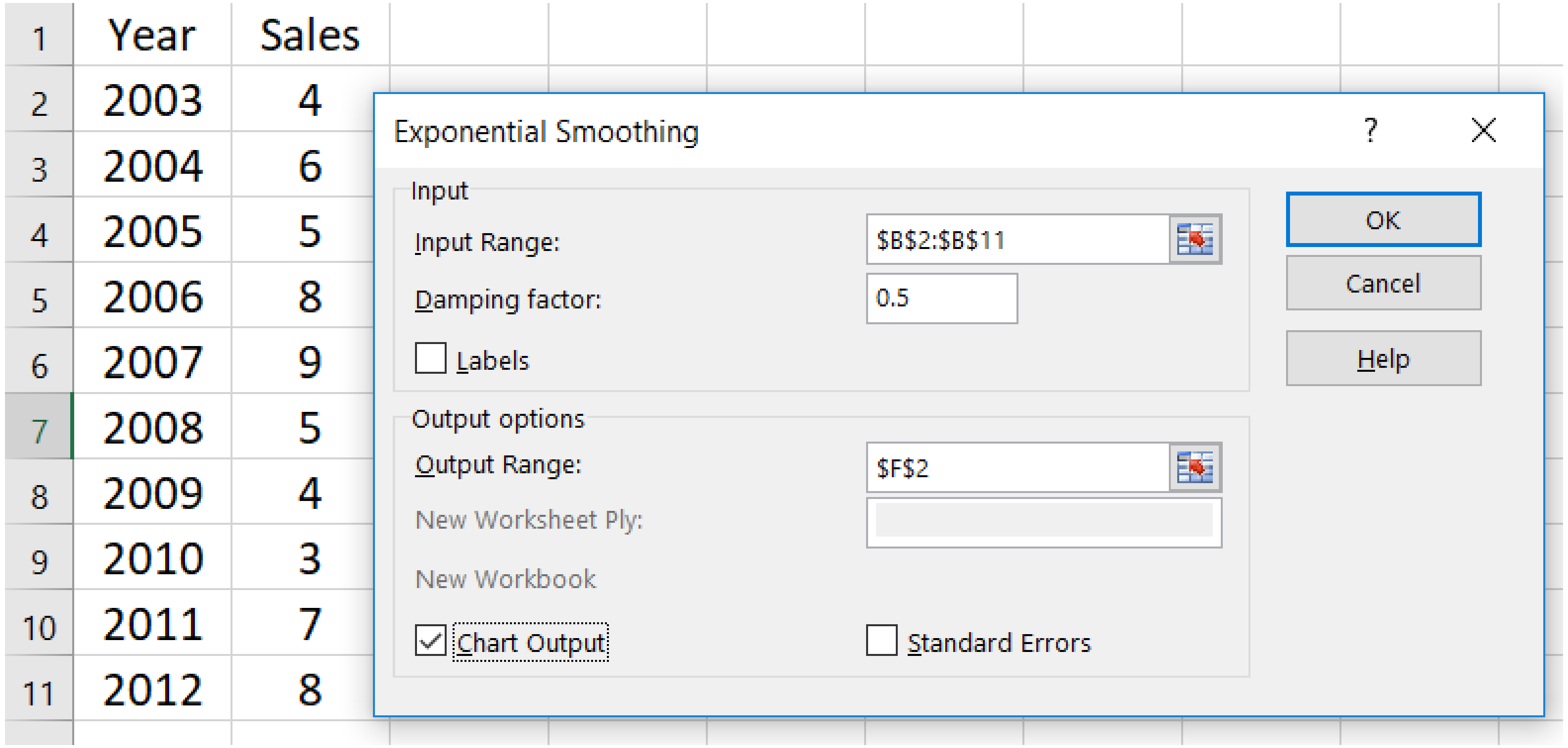
Calculate a five-year moving average from the following data set:

Here, consider smoothing constant or damping factor as 0.5.

Steps to solve in Excel

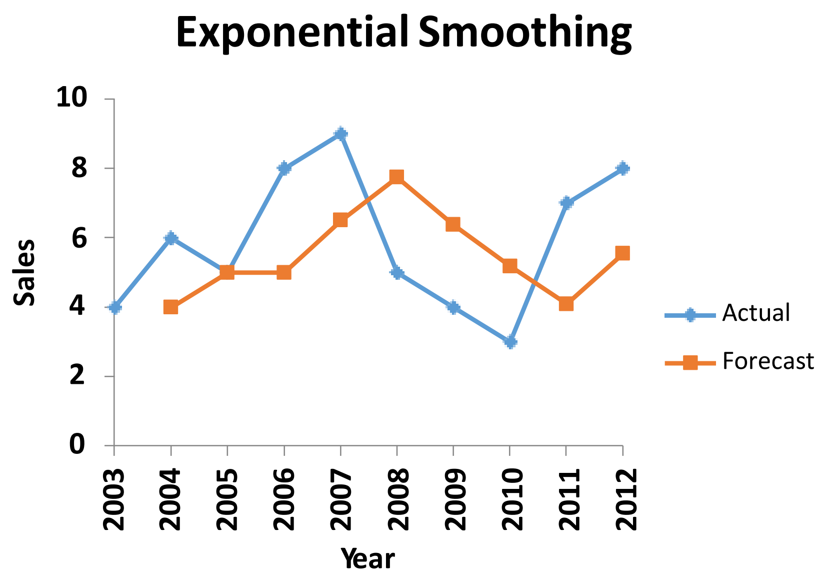
Data -> Data Science -> Exponential Smoothing -> Ok

Year	Sales ($M)
2003	4
2004	6
2005	5
2006	8
2007	9
2008	5
2009	4
2010	3
2011	7
2012	8



Exponential Moving Average

Year	Sales	Exponential Moving Average
2003	4	
2004	6	4
2005	5	5
2006	8	5
2007	9	6.5
2008	5	7.75
2009	4	6.375
2010	3	5.188
2011	7	4.094
2012	8	5.547



Advantages and Disadvantages of EMA

Advantages

- It smooths out random noise in the data
- Smoothing constant is flexible since it chooses an arbitrary value.
- Last period forecast and actual value is required.

Disadvantages

- The α should be very low
- Large value of α turn exponential smoothing into last periods value (α near to 1) or last period's forecast
- It also lags the actual value
- It is partially dependent on last period's forecast

Advanced Exponential Smoothing

Simple exponential smoothing is to forecast future values by applying a smoothing coefficients to all the previous values in our series

Apart from Simple Exponential Smoothing, there are other two types of advanced exponential smoothing

Holt's Exponential smoothing is used for forecasting a series that has a trend but no seasonality

Holt Winters or Winters Exponential smoothing used for forecasting a series that have both trend and seasonality.

Box-Jenkins (ARIMA) Models

Three basic ARIMA models for a stationary time series y_t :

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

i.e., y_t depends on its *p* previous values

(2) Moving Average model of order *q* (MA(*q*))

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

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

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

Box-Jenkins (ARIMA) Models

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

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

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

Box-Jenkins (ARIMA) Models

In an ARIMA model, the random disturbance term ε_t is typically assumed to be "white noise"; i.e., it is identically and independently distributed with a mean of 0 and a common variance σ^2 across all observations.

We write $\varepsilon_t \sim$ i.i.d.$(0, \sigma^2)$

A five-step iterative procedure

1) Stationarity Checking and Differencing
2) Model Identification
3) Parameter Estimation
4) Diagnostic Checking
5) Forecasting

Step One: Stationarity checking

Stationarity

"Stationarity" is a fundamental property underlying almost all time series statistical models.
A time series y_t is said to be stationary if it satisfies the following conditions:

$$(1)\ \ E(y_t) = u_y \quad for\ all\ t.$$

$$(2)\ Var(y_t) = E[(y_t - u_y)^2] = \sigma_y^2 \quad for\ all\ t.$$

$$(3)\ Cov(y_t, y_{t-k}) = \gamma_k \quad for\ all\ t.$$

Stationarity

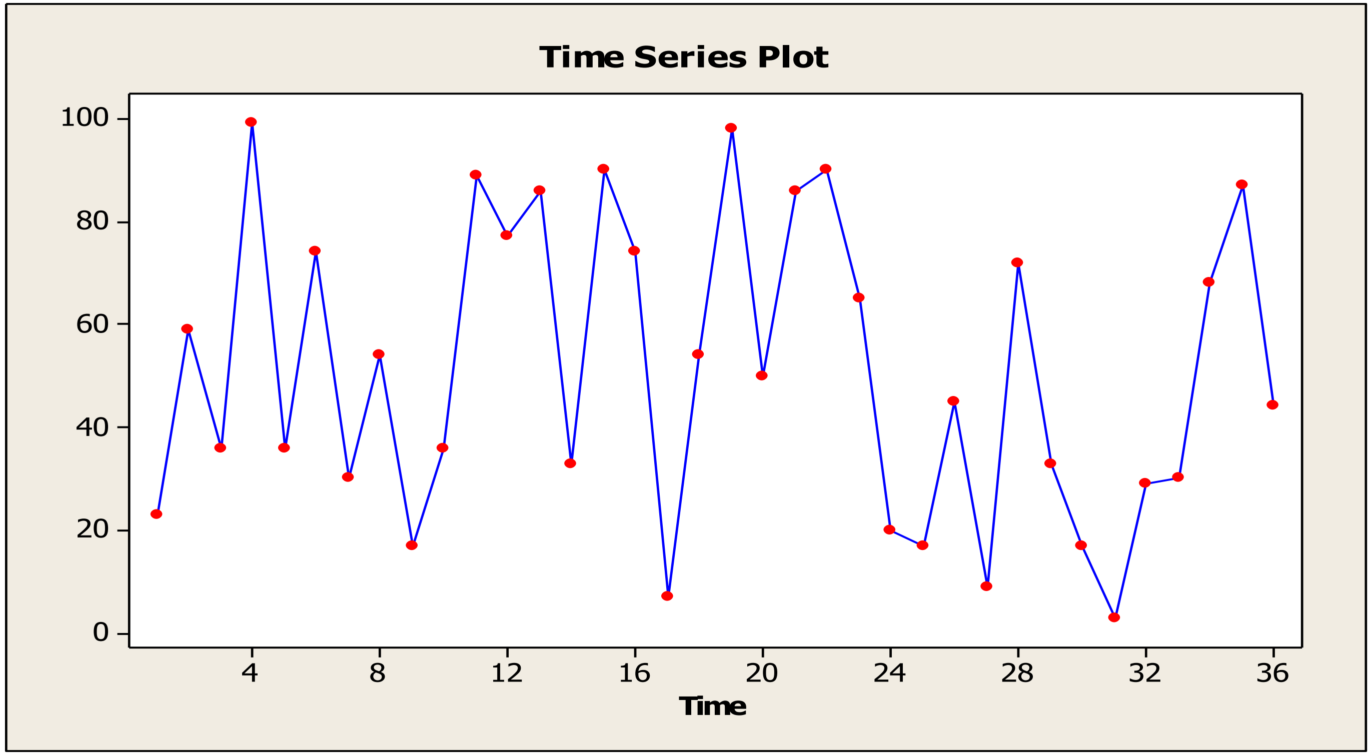
The white noise series ε_t satisfies the stationarity condition because

(1) E(ε_t) = 0

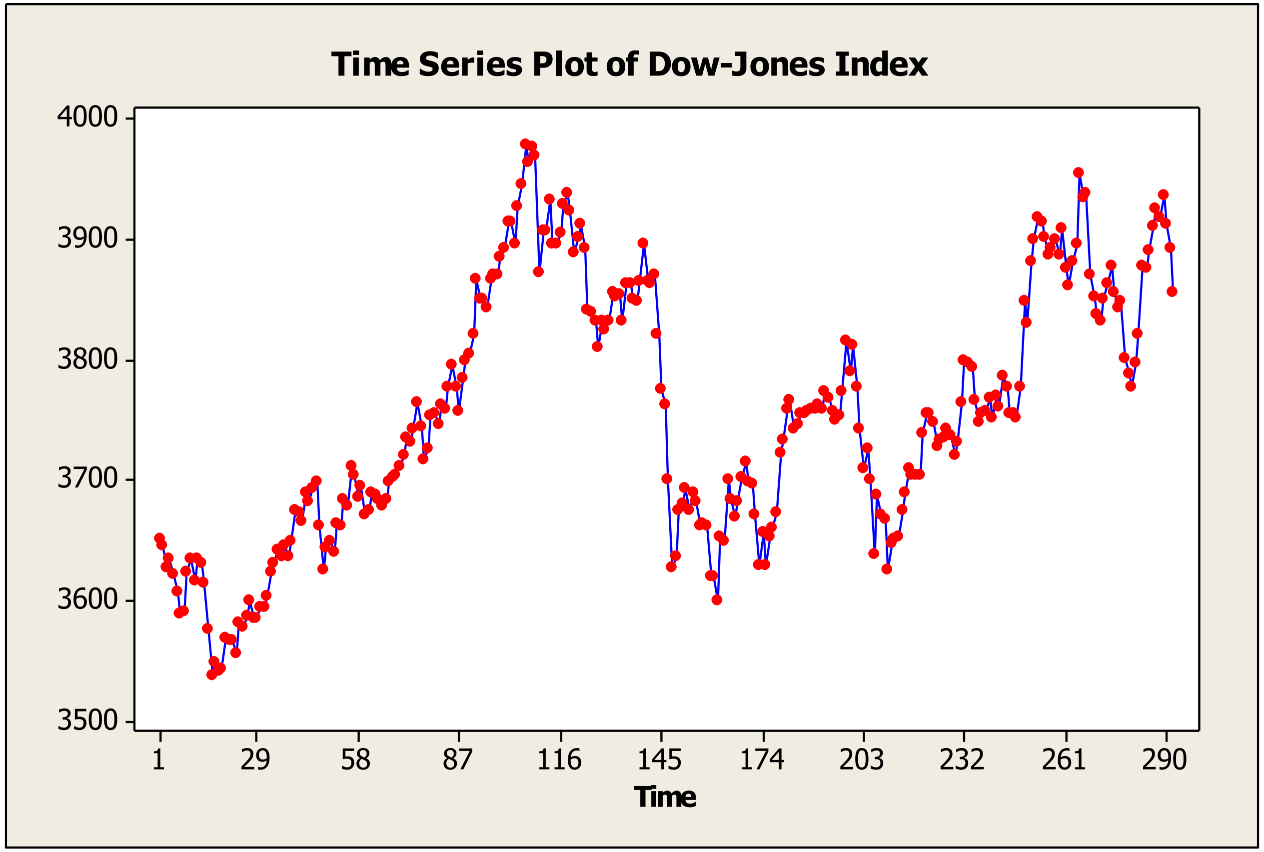
(2) Var(ε_t) = σ^2

(3) Cov($\varepsilon_t \varepsilon_{t-s}$) = for all s $\neq$ 0

Example of a white noise series



Example of a non-stationary series



Stationarity

Suppose y_t follows an AR(1) process without drift.

Is y_t stationarity?

Note that

$$
\begin{aligned}
y_t &= \phi_1 y_{t-1} + \varepsilon_t \\
&= \phi_1(\phi_1 y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\
&= \varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + \phi_1^3 \varepsilon_{t-3} + \ldots\ldots\ldots\ldots + \phi_1^t y_o
\end{aligned}
$$

Stationarity

Without loss of generality, assume that $y_o = 0$. Then $E(y_t)=0$.
Assuming that t is large, i.e., the process started a long time ago, then

$$\operatorname{var}(y_t) = \frac{\sigma^2}{(1-\phi_1^2)}, \text{ provided that } |\phi_1| < 1.$$

It can also be shown that provided that the same condition is satisfied,

$$\operatorname{cov}(y_t y_{t-s}) = \frac{\phi_1^s \sigma^2}{(1-\phi_1^2)} = \phi_1^s \operatorname{var}(y_t)$$

Stationarity

Special Case: $\phi_1 = 1$

$$y_t = y_{t-1} + \varepsilon_t \,.$$

It is a "random walk" process. Now,

$$y_t = \sum_{j=0}^{t-1} \varepsilon_{t-j} \,.$$

Thus,

$$(1)\quad E(y_t) = 0 \quad for \;\; all \;\; t.$$

$$(2)\; Var(y_t) = t\sigma_\varepsilon^2 \quad for \;\; all \;\; t.$$

$$(3)\; Cov(y_t\,, y_{t-s}) = (t-s)\sigma_\varepsilon^2 \quad for \;\; all \;\; t.$$

Stationarity

Suppose the model is an AR(2) without drift, i.e.,

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t$$

It can be shown that for y_t to be stationary,

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

The key point is that AR processes are not stationary unless appropriate prior conditions are imposed on the parameters.

Stationarity

Consider an MA(1) process without drift:

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

It can be shown, regardless of the value of θ_1, that

$$E(y_t) = 0$$

$$\text{var}(y_t) = \sigma^2 (1 + \theta_1^2)$$

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

Stationarity

For an MA(2) process

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2}$$

$$E(y_t) = 0$$

$$\operatorname{var}(y_t) = \sigma^2 (1 + \theta_1^2 + \theta_2^2)$$

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

Stationarity

In general, MA processes are stationarity regardless of the values of the parameters, but not necessarily “invertible”.

An MA process is said to be invertible if it can be converted into a stationary AR process of infinite order.

The conditions of invertibility for an MA(k) process is analogous to those of stationarity for an AR(k) process. More on invertibility in tutorial.

Differencing

Often non-stationary series can be made stationary through differencing.
Examples:

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

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

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

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

Differencing

Differencing continues until stationarity is achieved.

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

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

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

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

Differencing

Backward shift operator, *B*

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

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

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

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

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

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

Differencing

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

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

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

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

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

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

Population autocorrelation function

The question is, in practice, how can one tell if the data are stationary?

Let $\gamma_k = \text{cov}(y_t, y_{t-k})$ so that

$\rho_k = \gamma_k / \gamma_o$ is the autocorrelation coefficient at lag k.

Consider an AR(1) process,

$$\rho_k = \phi_1^k$$

If the data are non-stationary (i.e., random walk), then $\phi_1 = 1$ for all values of k

If the data are stationary (i.e., $|\phi_1| < 1$), then the magnitude of the autocorrelation coefficient "dies down" as k increases.

Population autocorrelation function

Consider an AR(2) process without drift :

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \,.$$

The autocorrelation coefficients are

$$\rho_1 = \frac{\phi_1}{1-\phi_2},$$

$$\rho_2 = \phi_2 + \frac{\phi_1^2}{1-\phi_2}$$

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} \qquad for\ k \geq 2.$$

Population autocorrelation function

Then the autocorrelation function dies down according to a mixture of damped exponentials and/or damped sine waves.

- In general, the autocorrelation of a stationary AR process dies down gradually as k increases.

Population autocorrelation function

Moving Average (MA) Processes

Consider a MA(1) process without drift :

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}.$$

Recall that

$$(1)\ E(y_t) = 0 \quad for\ all\ t.$$

$$(2) Var(y_t) = \gamma_0 = \sigma_\varepsilon^2 (1 + \theta_1^2) \quad for\ all\ t.$$

$$(3) Cov(y_t, y_{t-s}) = \gamma_s = \begin{cases} -\theta_1 \sigma_\varepsilon^2 & s = 1 \\ 0 & s > 1. \end{cases}$$

Population autocorrelation function

Therefore the autocorrelation coefficient of the MA(1) process at lag k is

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

$$= \begin{cases} \dfrac{-\theta_1}{1+\theta_1^2} & k = 1 \\ 0 & k > 1. \end{cases}$$

The autocorrelation function of the MA(1) process "cuts off" after lag k=1.

Population autocorrelation function

Similarly, for an MA(2) process :

$$\rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2},$$

$$\rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2},$$

$$\rho_k = 0 \quad for\ k > 2.$$

The autocorrelation function of a MA(2) process cuts off after 2 lags.

Population autocorrelation function

In general, all stationary AR processes exhibit autocorrelation patterns that "die down" to zero as k increases, while the autocorrelation coefficient of a non-stationary AR process is always 1 for all values of k. MA processes are always stationary with autocorrelation functions that cut off after certain lags.

Question: how are the autocorrelation coefficients "estimated" in practice?

Sample Autocorrelation Function (SAC)

For the series $y_1, y_2, \ldots, y_n$, the sample autocorrelation at lag k is

$$r_k = \frac{\sum_{t=1}^{n-k} (y_t - \bar{y})(y_{t+k} - \bar{y})}{\sum_{t=1}^{n} (y_t - \bar{y})^2}$$

where

$$\bar{y} = \frac{\sum_{t=1}^{n} y_t}{n}$$

Sample Autocorrelation Function (SAC)

r_k measures the linear relationship between time series observations separated by a lag of k time units

The standard error of r_k is

$$s_{r_k} = \sqrt{\frac{1 + 2\sum_{j=1}^{k-1} r_j^2}{n}}.$$

The t_{r_k} statistic is

$$t_{r_k} = \frac{r_k}{s_{r_k}}.$$

Sample Autocorrelation Function (SAC)

Behaviour of SAC

(1) The SAC can cut off. A spike at lag k exists in the SAC if r_k is statistically large. If

$$\left| t_{r_k} \right| > 2$$

Then r_k is considered to be statistically large. The SAC cuts off after lag k if there are no spikes at lags greater than k in the SAC.

Sample Autocorrelation Function (SAC)

(2) The SAC is said to die down if this function does not cut off but rather decreases in a 'steady fashion'. The SAC can die down in

(i) a damped exponential fashion
(ii) a damped sine-wave fashion
(iii) a fashion dominated by either one of or a combination of both (i) and (ii).

The SAC can die down fairly quickly or extremely slowly.

Sample Autocorrelation Function (SAC)

The time series values $y_1, y_2, \ldots, y_n$ should be considered stationary if the SAC of the time series values either cuts off fairly quickly or dies down fairly quickly.

However if the SAC of the time series values $y_1, y_2, \ldots, y_n$ dies down extremely slowly, and r_1 at lag 1 is close to 1, then the time series values should be considered non-stationary.

Stationarity Summary

Stationarity of data is a fundamental requirement for all time series analysis.
MA processes are always stationary

AR and ARMA processes are generally not stationary unless appropriate restrictions are imposed on the model parameters.

Stationarity Summary

Population autocorrelation function behaviour:

1) stationary AR and ARMA: dies down gradually as k increases
2) MA: cuts off after a certain lag
3) Random walk: autocorrelation remains at one for all values of k

Stationarity Summary

Sample autocorrelation function behaviour:

1) stationary AR and ARMA: dies down gradually as k increases
2) MA: cuts off after a certain lag
3) Random walk: autocorrelation functions dies down very slowly and the estimated autocorrelation coefficient at lag 1 is close to 1.

What if the data are non-stationary?

Perform differencing until stationarity is accomplished.

A random walk process

```
                              The ARIMA Procedure
                        Name of Variable = y

                 Mean of Working Series    16.79147
                 Standard Deviation         9.39551
                 Number of Observations          98

                                Autocorrelations

 Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

   0     88.275614        1.00000    |                    |********************|              0
   1     85.581769        0.96948    |                .   |******************* |       0.101015
   2     81.637135        0.92480    |             .      |******************  |       0.171423
   3     77.030769        0.87262    |           .        |*****************   |       0.216425
   4     72.573174        0.82212    |          .         |****************    |       0.249759
   5     68.419227        0.77506    |         .          |****************    |       0.275995
   6     64.688289        0.73280    |        .           |***************     |       0.297377
   7     61.119745        0.69237    |       .            |**************      |       0.315265
   8     57.932253        0.65627    |       .            |*************       |       0.330417
   9     55.302847        0.62648    |      .             |*************.      |       0.343460

                                     "." marks two standard errors
```

Random walk process after first difference

```
                    The ARIMA Procedure
                    Name of Variable = y

     Period(s) of Differencing                            1
     Mean of Working Series                        0.261527
     Standard Deviation                            1.160915
     Number of Observations                              97
     Observation(s) eliminated by differencing            1

                         Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0      1.347723        1.00000    |                    |********************|              0
  1      0.712219        0.52846    |                .   |***********         |       0.101535
  2      0.263094        0.19521    |               .    |****.               |       0.126757
  3     -0.043040        -.03194    |               .   *|    .               |       0.129820
  4     -0.151081        -.11210    |               .  **|    .               |       0.129901
  5     -0.247540        -.18367    |               .****|    .               |       0.130894
  6     -0.285363        -.21174    |               .****|    .               |       0.133525
  7     -0.274084        -.20337    |               .****|    .               |       0.136943
  8     -0.215508        -.15991    |              .  ***|     .              |       0.140022
  9     -0.077629        -.05760    |              .    *|     .              |       0.141892
                                        "." marks two standard errors
```

Step Two: Model Identification

Model Identification

When the data are confirmed stationary, one may proceed to tentative identification of models through visual inspection of both the SAC and partial sample autocorrelation (PSAC) functions.

Partial autocorrelation function

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

The partial autocorrelations at lags 1, 2, 3, …, make up the partial autocorrelation function or PACF.

Summary of the Behaviour of autocorrelation and partial autocorrelation functions

Behaviour of autocorrelation and partial autocorrelation functions

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

Summary of the Behaviour of autocorrelation and partial autocorrelation functions

Behaviour of AC and PAC for specific non-seasonal models

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

Summary of the Behaviour of autocorrelation and partial autocorrelation functions

Behaviour of AC and PAC for specific non-seasonal models

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

Summary of the Behaviour of autocorrelation and partial autocorrelation functions

Behaviour of AC and PAC for specific non-seasonal models

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

Step Three: Parameter Estimation

Parameter Estimation

The method of least squares can be used. However, for models involving an MA component, there is no simple formula that can be applied to obtain the estimates.

Another method that is frequently used is maximum likelihood.

Parameter Estimation

Given n observations y_1, y_2, …, y_n, the likelihood function L is defined to be the probability of obtaining the data actually observed.

The maximum likelihood estimators (m.l.e.) are those value of the parameters for which the data actually observed are most likely, that is, the values that maximize the likelihood function L.

Parameter Estimation

Given n observations $y_1, y_2, \ldots, y_n$, the likelihood L is the probability of obtaining the data actually observed.

For non-seasonal Box-Jenkins models, L is a function of δ, ϕ's, θ's and σ_ε^2 given $y_1, y_2, \ldots, y_n$.

The maximum likelihood estimators (m.l.e.) are those values of the parameters which make the observation of the data a most likely event, that is, the values that maximize the likelihood function L.

Parameter Estimation

To test whether the drift term should be included in the model, i.e., H_0: $\delta = 0$ vs H_a: $\delta \neq 0$

If $\left| \dfrac{\bar{z}}{s_z / \sqrt{n'}} \right| > 2$, reject H_0, where $\bar{Z}$ is the mean of

the working series, s_z is the standard deviation, and n' is the sample size of the working series.

Parameter estimation

Refers to the MA(1) example seen before

```
              The ARIMA Procedure
              Name of Variable = y

         Mean of Working Series    0.020855
         Standard Deviation        1.168993
         Number of Observations          98
```

Here,

$$t = 0.020855/(1.168993/\sqrt{98})$$

$$= 0.176608 < 2$$

The drift term should not be included.

Step Four: Diagnostic Checking

Diagnostic Checking

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

(1) t-tests for coefficient significance
(2) residual analysis
(3) model selection criteria

Diagnostic checking (t-tests)

Consider the data series of Example 4. First, the data appear to be stationary. Second, the drift term is significant. The SAC and SPAC functions indicate that the an AR(2) model probably best fits the data. For illustration purpose, suppose an MA(1) and ARMA(2,1) are fiitted in addition to the AR(2).

Diagnostic checking (t-tests)

```
data MA2;
input y;
cards;
4.0493268
7.0899911
4.7896497
.
.
2.2253477
2.439893;
proc arima data=ma2;
identify var=y;
estimate p=2 method=ml printall;
estimate q=1 method=ml printall;
estimate p=2 q=1 method=ml printall;
run;
```

Diagnostic checking (t-tests)

The AR(2) model estimation results produce

$$\hat{y}_t = 4.68115 + 0.35039 y_{t-1} - 0.49115 y_{t-2}$$

The t test statistic for $H_o : \psi_1 = 0$ is 8.93, indicating significance.

Similarly, the t test also indicates that ψ_2 is significantly different from zero.

Diagnostic checking (t-tests)

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

$$\mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \ldots\ldots\phi_p}$$

Hence for the present AR(2) model, we have

4.10353 = 4.68115/(1-0.35039+0.49115)

Diagnostic checking (t-tests)

The MA(1) model produces

$$y_t = 4.10209 + e_t + 0.48367 e_{t-1}$$

The t test indicates significance of the coefficient of

- θ_1

Note that for any MA models, $\mu = \delta$.

Diagnostic checking (t-tests)

For the ARMA(2,1) model,

$$\hat{y}_t = 4.49793 + 0.40904 y_{t-1} - 0.50516 y_{t-2} + e_t - 0.07693 e_{t-1}$$

t tests results indicate that the two AR coefficients are significant, while the MA coefficient is not significant.

For any ARMA model,

$$\mu = \frac{\delta}{1 - \phi_1 - \phi_2 - \dots\dots - \phi_p}$$

Residual Analysis

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

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

Residual Analysis

Let's take the AR(2) model as an example, let u_t be the model's residual. It is assumed that

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots\dots\dots\dots$$

Suppose we want to test, up to lag 6,

$$H_o : \rho_1 = \rho_2 = \dots\dots\rho_6 = 0$$

The portmanteau test statistic gives 3.85 with a p-value of 0.4268. Hence the null cannot be rejected and the residuals are uncorrelated at least up to lag 6.

Residual Analysis

The results indicate that AR(2) and ARMA(2, 1) residuals are uncorrelated at least up to lag 30, while MA(1) residuals are correlated.

Model Selection Criteria

Akaike Information Criterion (AIC)

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

Schwartz Bayesian Criterion (SBC)

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

where L = likelihood function

k = number of parameters to be estimated,

n = number of observations.

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

Model Selection Criteria

For the AR(2) model,

AIC = 1424.66, SBC = 1437.292

For the MA(1) model,

AIC = 1507.162, SBC = 1515.583

For the ARMA(2,1) model,

AIC = 1425.727, SBC = 1442.57

Model Selection Criteria

	t-test	Q-test	AIC	SBC
AR(2)	√	√	1424.66	1437.292
MA(1)	√	×	1507.162	1515.583
ARMA(2,1)	√ (partial)	√	1425.727	1442.57

Judging these results, it appears that the estimated AR(2) model best fits the data.

Step Four: Forecasting

Forecasting

The h-period ahead forecast based on an ARMA(p,q) model is given by

$$\hat{y}_{t+h} = \hat{\delta} + \hat{\phi}_1 y_{t+h-1} + \ldots..\hat{\phi}_p y_{t+h-p} + e_{t+h} - \hat{\theta}_1 e_{t+h-1} - \ldots..\hat{\theta}_q e_{t+h-q}$$

where elements on the r.h.s. of the equation may be replaced by their estimates when the actual values are not available.

Forecasting

For the AR(2) model,

$$\hat{y}_{498+1} = 4.68115 + 0.35039 \times 2.4339893 - 0.49115 \times 2.2253477 = 4.4410$$

$$\hat{y}_{498+2} = 4.68115 + 0.35039 \times 4.4410 - 0.49115 \times 2.4339893 = 5.0418$$

For the MA(1) model,

$$\hat{y}_{498+1} = 4.10209 + 0 + 0.48367 \times (-0.7263) = 3.7508$$

$$\hat{y}_{498+2} = 4.10209 + 0 + 0.48367 \times (0) = 4.10209$$

Thank You.