Excel Forecasting

Trends and Forecasts using Charts

Charts can be used for visualising forecast and data analysis. Trends can be easily spotted and forecasts can be made directly on the chart.

The most common charts used to visualise data analysis are Column charts, Line charts, Scatter charts and Pie charts.

Creating Trendlines

Before you start creating charts and trendlines make sure that you have sufficient data to show a meaningful chart and enough data for Excel to be able to calculate a reliable trend.

Insufficient data might twist the results. For example, seasonal fluctuations might be mistaken for long-term trends if the data is taken from only one year. The chart must be a 2-D chart. Excel cannot create trendlines on 3-D charts.

- The data is ordered from earliest to most recent.
- No data is missing. If data is unavailable for a period, enter an estimate.

Trendlines are graphical representations of trends in data that you can use to analyse problems of prediction. Such analysis is called regression analysis. By using regression analysis, you can extend a trendline in a chart beyond the actual data to predict future values. For example, this chart uses a simple linear trendline that is forecast ahead three periods to clearly show a trend toward rising revenue.
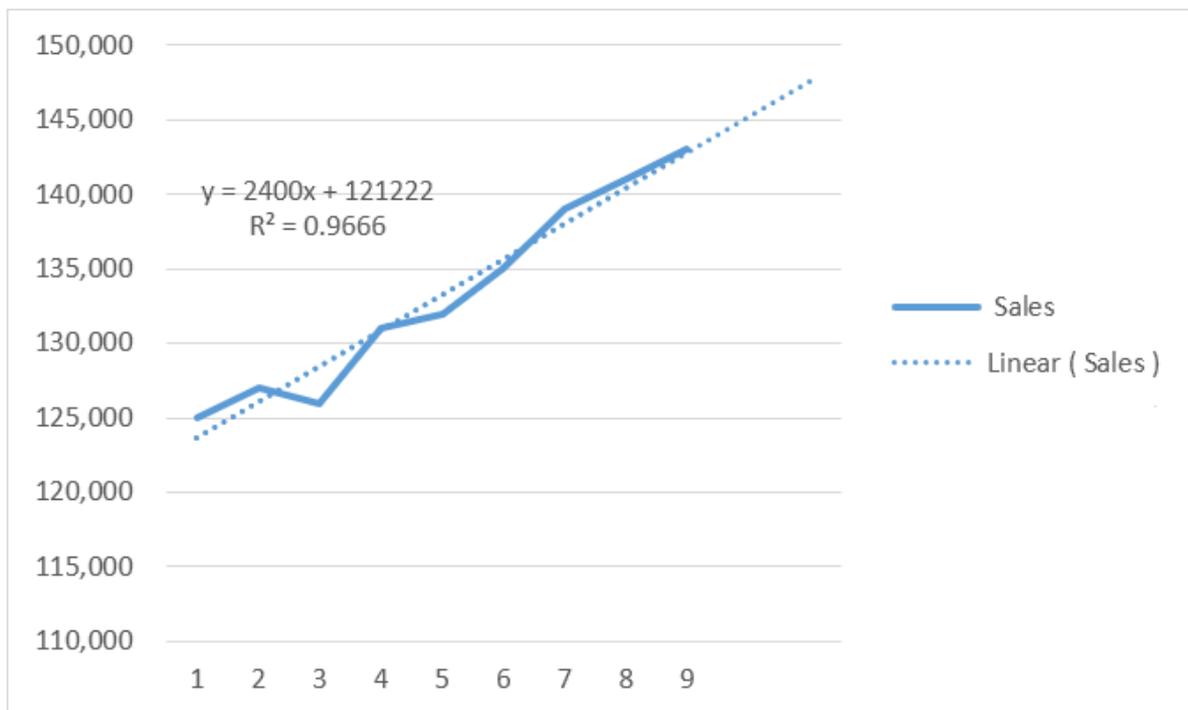
How reliable are these trendlines? The answer is in the value of R^2 (R-squared) the coefficient of determination, which is a number between 0 and 1.

A trendline is most reliable when its R-squared value is at or near 1. When you fit a trendline to your existing data, Excel automatically calculates its R-squared value based on a formula. The more closely a trendline match your existing data, the more accurate a forecast that uses this trendline is likely to be. If you want, you can display this value on your chart.

Correlation

When we are forecasting we need to determine x, the independent variable and y, the dependant variable. With time series data it is obvious, x is the time series. With other types of data you need to establish whether there is a correlation between the data.

The Correl function determines the relationship (the correlation coefficient) between two arrays of data. For example, you may need to find out if your company's increased advertising costs also increase your company's sales figures or if higher temperatures in summer increases the sales of ice cream.

CORREL function

The Correl function consists of two required arguments, in the following order: Array1, Array2. Array1 and Array2 are the arrays of data you want to examine to see if there is any relationship. In the Array1 box enter the first range of data. In the Array2 box enter the second range of data.

CORREL(Array1, Array2)

	A	B	C	D	E	F	G
1	Period	Sales	Av Temp				
2	Jan	50000	10				
3	Feb	45000	9		=CORREL(B2:B13,C2:C13)		
4	Mar	50000	11		CORREL(array1, array2)		
5	Apr	65000	12				
6	May	75000	15				
7	Jun	125000	17				
8	Jul	155000	20				
9	Aug	176500	22				
10	Sep	95000	18				
11	Oct	65000	12				
12	Nov	40000	9				
13	Dec	45000	8				
14							

The Correl function will return a result between -1 and 1. 0.95439.

Between 1 and 0:

If the result is close to 1 then it is because there is a close positive relationship between the two arrays (If the temperature rises the ice cream sales increase). Close to 0 then there is not a close relationship between the two arrays (the sales do not increase because of increased advertising costs).

Between 0 and -1:

If the result is close to -1 then there is a close negative relationship between the two arrays (if the price goes up the sales goes down). Close to 0 but a negative value then there is not a close relationship between the two arrays (the sales does not decrease if the price goes up).

Creating a scatter chart to display correlation

Scatter charts are very useful to visualise correlation.

In the example below it is easy to spot the trend in the chart. There is a close relationship between average temperature and sales. Of course, you can also see that the Correl function returns 93% which also emphasise that there is a close relationship between the two arrays.
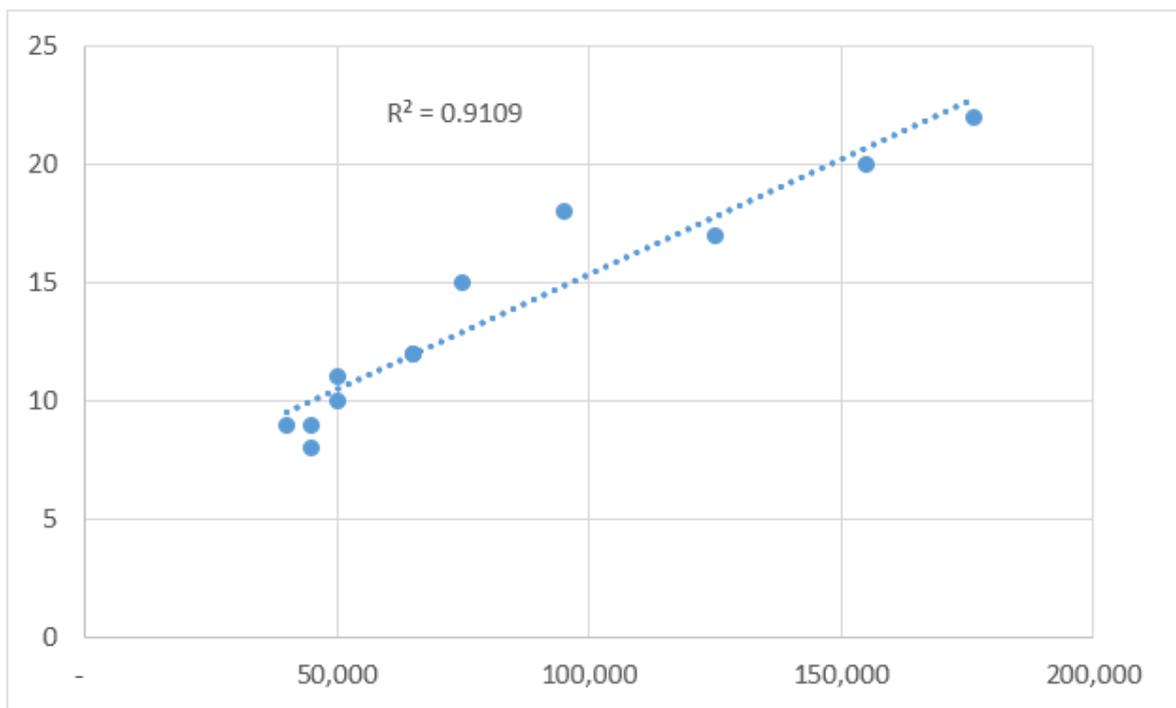
To make it even easier to spot the trend you can add a linear trendline to the chart and if you add the R^2 value to the chart you can see how accurate your trendline is. A value close to 1 shows that the trendline is very accurate. Close to 0 there is not really any trend.

Linear Regression

Linear regression analysis is the standard statistical technique for estimating the relationships among variables. For example, how do the sales figures change over time?

If the goal is prediction, or forecasting, linear regression can be used to fit a predictive model to an observed data set of y and x values or known actual data (y) over time (x) Typically it is time series data. After developing such a model, if an additional value of x (a new period) is then given without its accompanying value of y, the fitted model can be used to make a prediction of the value of y.

The sales figures (y) are known for a number of periods (x) it makes it possible forecasting sales (y) for future periods (x).

Linear regression can be calculated using the FORECAST function, the TREND function or by calculating the standard equation: **Y = mX + c** and then using the SLOPE and INTERCEPT functions.
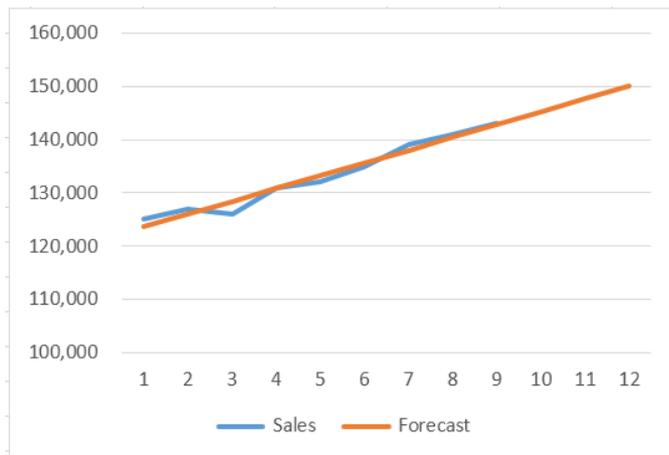
FORECAST function

The Forecast function can calculate a trend, a linear forecast, or linear regression for as many future periods needed based on data from previous periods.

FORECAST(x, Known_y's, Known_x's)

	A	B	C	D	E	F
1	Period	Sales	Forecast			
2	1	125,000	=FORECAST(A2,$B$2:$B$10,$A$2:$A$10)			
3	2	127,000	FORECAST(x, known_ys, known_xs)			
4	3	126,000	128,422			
5	4	131,000	130,822			
6	5	132,000	133,222			
7	6	135,000	135,622			
8	7	139,000	138,022			
9	8	141,000	140,422			
10	9	143,000	142,822			
11	10		145,222			
12	11		147,622			
13	12		150,022			

The Forecast function consists of three required arguments, in the following order: X, Known_y's, and Known_x's. X is the period for which you want to calculate the forecast. Known_y's is the array with the known values. Known_x's is the array with the known periods (must be a numeric range and not dates, months or years).

Sales actuals plotted against forecast.

TREND function

The Trend function can calculate a trend, a linear forecast, or linear regression for as many future periods needed based on data from previous periods.

TREND(Known_y's, Known_x's, New_x's, Const)

	A	B	C	D	E
1	**Period**	**Sales**	**Trend**		
2	1	125,000	=TREND(B2:B10,A2:A10,A2:A13)		
3	2	127,000	TREND(known_ys, [known_xs], [new_xs], [const])		
4	3	126,000	128,422		
5	4	131,000	130,822		
6	5	132,000	133,222		
7	6	135,000	135,622		
8	7	139,000	138,022		
9	8	141,000	140,422		
10	9	143,000	142,822		
11	10		145,222		
12	11		147,622		
13	12		150,022		

The Trend function consists of one required argument and three optional arguments, in the following order: Known_y's, Known_x's, New_x's, and Const. Known_y's is the array with the known values. Known_x's is the array with the known periods (must be a numeric range and not dates, months or years). New_x's is the array with future periods if the Trend function is used for forecasting. Const is a logical value specifying whether to force the constant b to equal 0.

The Trend function <u>must</u> be entered as an array formula, thus:

- Select the range where you want the results
- Enter the function and it's required arguments
- Press the keys Ctrl+Shift+Enter

The formula will return the results (the new y's) for the selected range.

SLOPE and INTERCEPT functions

To calculate linear regression the standard equation **Y = mX + c** you need the slope and y interceptor values.

SLOPE(Known_y's, Known_x's)
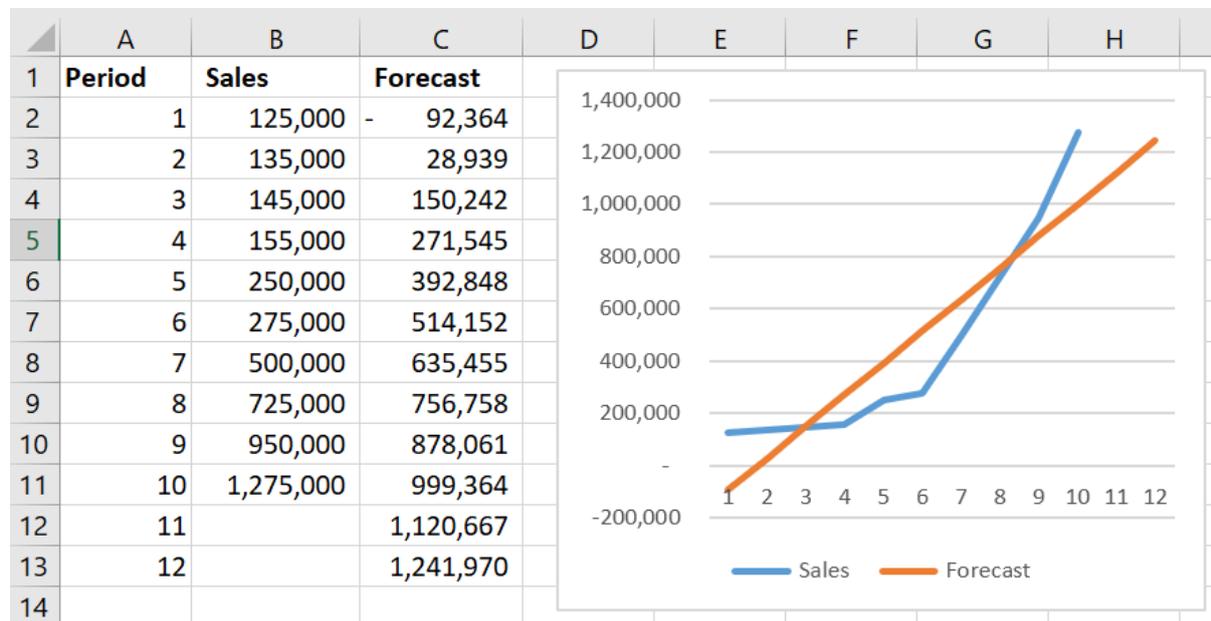
INTERCEPT(Known_y's, Known_x's)

	A	B	C	D	E	F	G	H
1	**Period**	**Sales**	**y=mx+b**		**Slope**	2400		
2	1	125,000	=($F$1*A2)+$F$2			121222		
3	2	127,000	126022					
4	3	126,000	128422		**mx+b**	(Slope*First_x)+Y Intercept		
5	4	131,000	130822					
6	5	132,000	133222					
7	6	135,000	135622					
8	7	139,000	138022					
9	8	141,000	140422					
10	9	143,000	142822					
11	10		145222					
12	11		147622					
13	12		150022					

Both functions require the following arguments: Known_y's, and Known_x's. Known_y's is the array with the known values. Known_x's is the array with the known periods (must be a numeric range and not dates, months or years)

The regression line is calculated by multiplying the value of the slope by the first x value and then adding the value of the Y intercept.

Exponential Regression

Sometimes the growth in a model is not linear, but exponential. If the growth is exponential then you should not use the previous methods which only work for linear regression.

	A	B	C
1	Period	Sales	Forecast
2	1	125,000	- 92,364
3	2	135,000	28,939
4	3	145,000	150,242
5	4	155,000	271,545
6	5	250,000	392,848
7	6	275,000	514,152
8	7	500,000	635,455
9	8	725,000	756,758
10	9	950,000	878,061
11	10	1,275,000	999,364
12	11		1,120,667
13	12		1,241,970
14			

In the example above the sales has an exponential growth rate. The Forecast function has been used to forecast the next few periods but this is clearly not a realistic forecast. The known sales are already higher than the forecast.

GROWTH function

The Growth function can calculate exponential growth and exponential growth forecast.
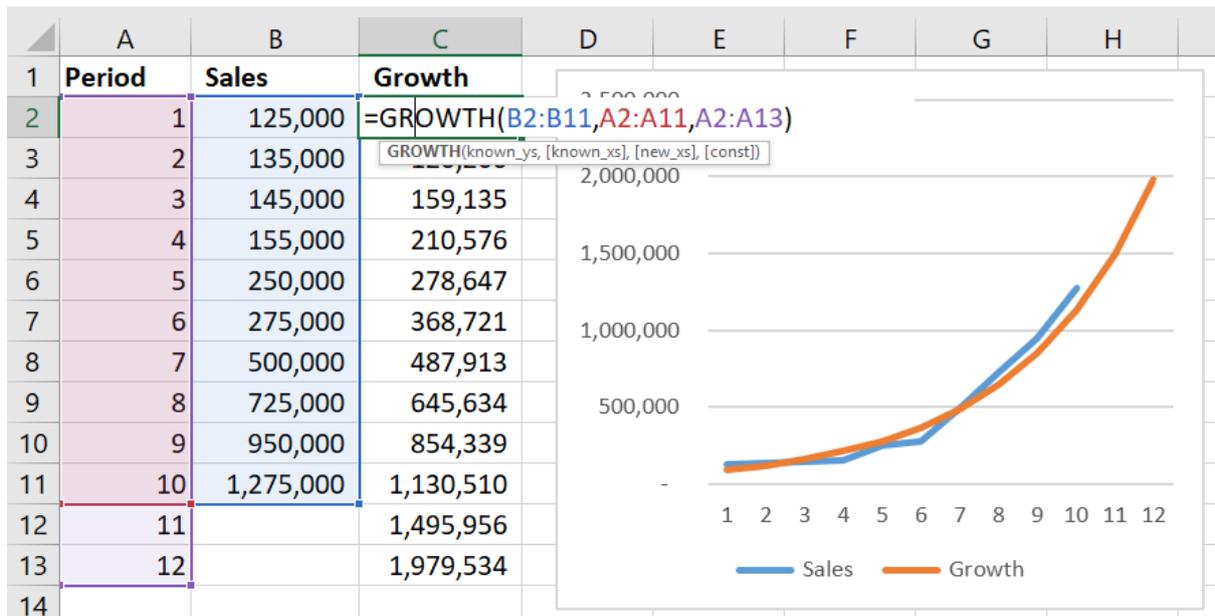
GROWTH(Known_y's, Known_x's, New_x's, Const)

The Growth function consists of one required argument and three optional arguments, in the following order: Known_y's, Known_x's, New_x's, and Const. Known_y's is the array with the known values. Known_x's is the array with the known periods (must be a numeric range and not dates, months or years). New_x's is the array with future periods if the Growth function is used for forecasting. Const is a logical value specifying whether to force the constant b to equal 0.

The Growth function <u>must</u> be entered as an array formula, thus:

- Select the range where you want the results
- Enter the function and it's required arguments
- Press the keys Ctrl+Shift+Enter

The formula will return the results (the new y's) for the selected range.

	A	B	C	D	E	F	G	H
1	**Period**	**Sales**	**Growth**					
2	1	125,000	=GROWTH(B2:B11,A2:A11,A2:A13)					
3	2	135,000						
4	3	145,000	159,135					
5	4	155,000	210,576					
6	5	250,000	278,647					
7	6	275,000	368,721					
8	7	500,000	487,913					
9	8	725,000	645,634					
10	9	950,000	854,339					
11	10	1,275,000	1,130,510					
12	11		1,495,956					
13	12		1,979,534					
14								

GROWTH(known_ys, [known_xs], [new_xs], [const])

If you compare the forecast made with the Growth function with that made by the Forecast function in a chart it is obvious that you get a much more accurate forecast using the Growth function.

Averaging

Moving Average

The Moving Average forecast is based on the average of known values from a number of periods.

The two months Moving Average forecast is based on 2 months known values. It can be the two previous month's values or it can be based on two months the year before whatever gives the most accurate forecast.

If you forecast one period using the average of the previous two periods will you only be able to forecast one future period. If you use the average of two periods one period back, you can forecast two future periods using this method.

	A	B	C	D	E
1	Period	Sales	Moving Average 2 periods back		
2	1	125,000			
3	2	127,000			
4	3	126,000	=AVERAGE(B2:B3)		
5	4	131,000	AVERAGE(number1, [number2], ...)		
6	5	132,000	128,500		
7	6	135,000	131,500		
8	7	139,000	133,500		
9	8	141,000	137,000		
10	9	143,000	140,000		
11	10		142,000		
12	11				
13	12				

In the example above the forecast for March (Period 3) is the average of January and February.

Weighted Moving Average

Another averaging method is the Weighted Moving Average. The Weighted Moving Average is similar to the Moving Average except that it assumes that the most recent periods are the best predictors. Each period that is used for the forecast is given a weighting. The largest weighting is assigned to the most recent period. Each preceding period is given successively less and less weight.

For example, during the calculation of a three-period Weighted Moving Average, the most recent three periods are used. The results from the previous period might be given the weight

of 60%, the middle period might be assigned a weight of 30%, and the 3rd most distant period might be assigned a weight of 10%.

	A	B	C	D	E	F	G	H	I
1	**Period**	**Sales**			**Weighted Moving Average**				
2	1	125,000			i.e. most recent performance is the best predictor				
3	2	135,000			Period 1	10%			
4	3	145,000			Period 2	30%			
5	4	155,000	=(B4*$F$5)+(B3*$F$4)+(B2*$F$3)			60%			
6	5	250,000	150,000						
7	6	225,000	211,000						
8	7	245,000	225,500						
9	8	220,000	239,500						
10	9	235,000	228,000						
11	10		231,500						

There is no specific function to calculate the Weighted Moving Average, the formula is constructed as follows:

Forecast (period four) = (Value of period three * Weight of period three) + (Value of period two * Weight of period two) + (Value of period one * Weight of period one)

Seasonal Forecasting

If you have seasonal actual data (for example each year you are selling more in December and June) you can use seasonal forecast to forecast future data. A weight can be added to the periods to increase the accuracy in the forecast.
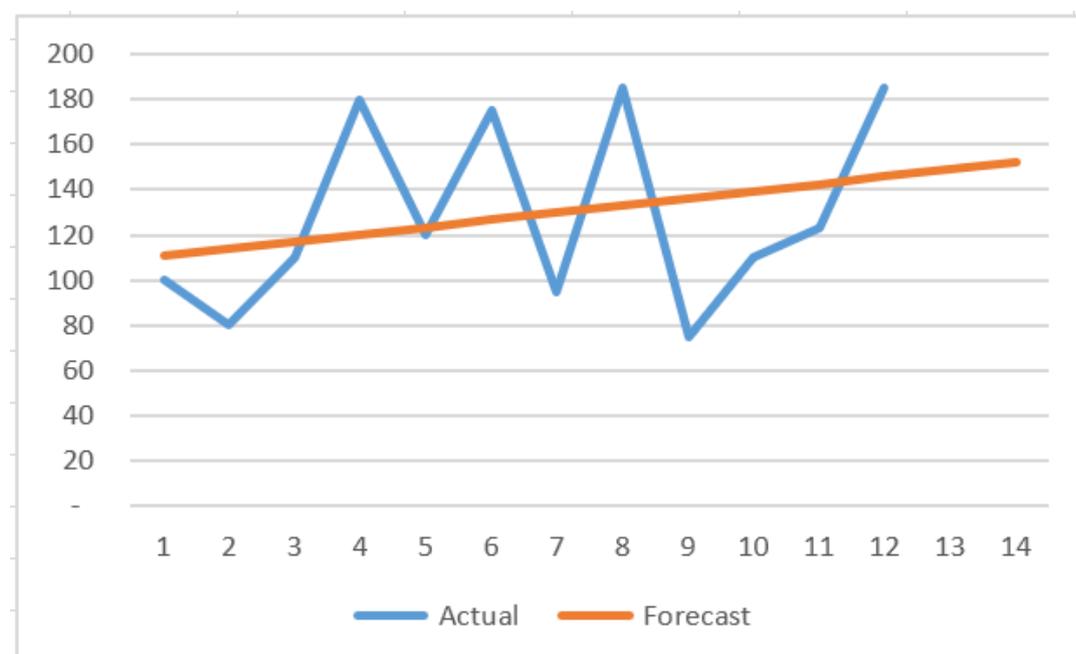

If you use the Forecast function to calculate a linear forecast you can see periods over forecast and under forecast. In the example above you have seasonal known actual data. The forecast values are wildly inaccurate as the line of linear regression averages and smooths the variances in the actuals. But we know that this is predicable behaviour.

We need to add an index value to each quarter to get a more accurate forecast. You have the data in quarters (4 periods). The sum of the index values must be 4 (12 if you are working with months!)

	A	B	C	D	E	F	G	H
1	Period	Quarter	Actual	Forecast	Seasonal Forecast			
2	1	Q1	100	111	=D2*VLOOKUP(B2,$G$5:$H$8,2,FALSE)			
3	2	Q2	80	114	80			
4	3	Q3	110	117	176			
5	4	Q4	180	120	108		Q1	0.9
6	5	Q1	120	123	111		Q2	0.7
7	6	Q2	175	127	89		Q3	1.5
8	7	Q3	95	130	195		Q4	0.9
9	8	Q4	185	133	120			
10	9	Q1	75	136	123		For Quarters, seasonal	
11	10	Q2	110	139	98		indices must sum to 4.	
12	11	Q3	123	143	214			
13	12	Q4	185	146	131			
14	13	Q1		149	134			

You can use error measurement calculations and the SOLVER tool to help you increase the accuracy of your seasonal forecasts when you are working with seasonal historical data. In the example above the seasonal index for quarter 1 is set to 0.9, set to 0.7 for quarter 2, set to 1.5 for quarter 3, and set to 0.9 for quarter 4. (0.9+0.7+1.5+0.9=4)
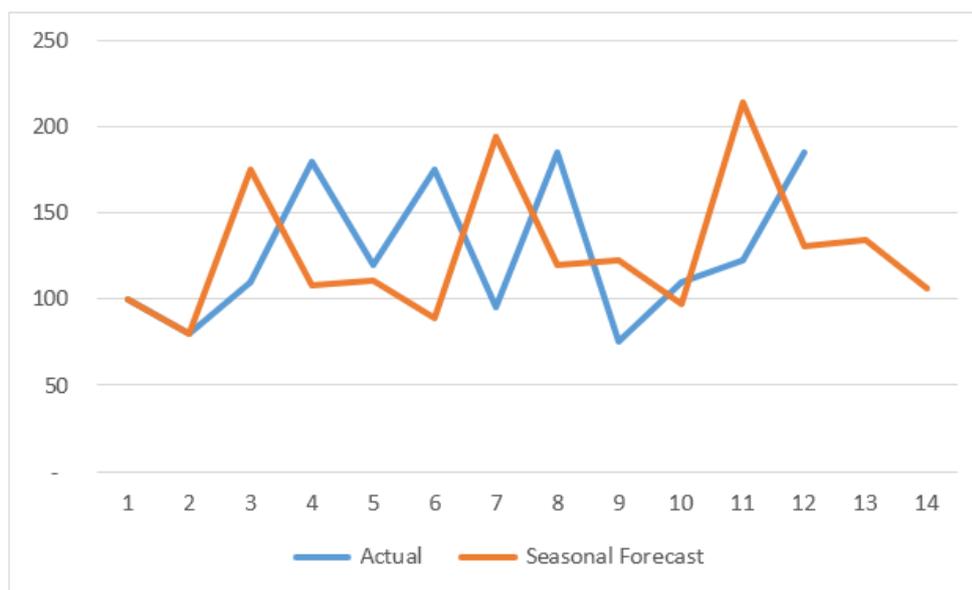
To get the seasonal forecast you just have to multiply the seasonal index value by the Forecast function's forecast and if you compare the actual data with the seasonal forecast you will see that the seasonal forecast is much more accurate than the Forecast function's forecast. You can either include a separate column in your worksheet populated with the relevant seasonal indices or you can enter them into a range of cells and use the VLOOKUP function to return the required value for the seasonal multiplier.

Measuring Forecast Accuracy

A key question in any forecasting is how to measure performance. Such measures are very important when you come to select a forecasting method and create a forecasting model, since you may compare alternatives and choose the method with the highest level of accuracy.

Then, once the method is being used on a regular basis, you need similar measures to confirm whether the forecasts are maintaining their accuracy. If not, you will need to get things back on track by putting improvements in place such as improved data capture, better statistical methods and maybe changing the forecasting method. To keep up getting more accuracy forecasts you could work with more than one forecasting method so you always can compare the accuracy.

To measure how accurate the forecast is you can use different methods. To be able to compare different forecasting methods you will look at the mean errors but first you must calculate the errors. Later you compare the different forecast methods to find out which one is the most accurate for our data right now, because this can change over time.

Absolute Deviation

It is a good idea to calculate absolute error/deviation to get the right picture of the error. If you calculate the average from positive and negative errors the result will not be very useful. To calculate the average you need to summarise all the values and then divide the sum with the number of data. You can have negative and positive errors which will neutralise or even out each other and the average result will not tell us anything about the forecast accuracy.

You need to calculate the absolute error/deviation to get a useful result in an average calculation. Use Excel's ABS function to calculate absolute (unsigned) values. ABS always returns a positive value whether the input is positive or negative.

To calculate the absolute error enter a formula like this:

=ABS(Actual Value – Forecast Value)

Mean Absolute Deviation (MAD)

The MAD/MAE is very important measurement to compare forecast methods. You need an overall result to compare methods. MAD/MAE is the average of the absolute deviation/error.

	A	B	C	D	E	F	G
1	Period	Sales	Forecast	Absolute Deviation			
2	1	125,000	139,689	14,689			
3	2	135,000	141,857	6,857			
4	3	145,000	144,025	975		MAD	Mean Absolute Deviation
5	4	155,000	146,193	8,807			
6	5	167,000	148,361	18,639		=AVERAGE(D2:D10)	
7	6	178,000	150,528	27,472		AVERAGE(number1, [number2], ...)	
8	7	145,000	152,696	7,696			
9	8	128,905	154,864	25,959			
10	9	156,340	157,032	692			
11	10		159,200				
12	11		161,368				
13	12		163,536				

HOW TO MEASURE FORECAST ACCURACY?

Forecast accuracy is a measure used for judging the efficiency of the forecasting process. There are several different measures, the most commonly used are:

- Mean Absolute Deviation (MAD)
- Mean Absolute Percentage Error (MAPE)
- Mean Percentage Error (MPE)

To determine which of these is the most appropriate measure to use, answer the following four basic questions in your business context:

Error Sign - Do you want to treat Positive and Negative Forecast Error the same or different? Will it make any difference to you if Forecast is more than Actual Sales or less than Actual Sales? This depends on what business you are in. If you are in business of perishable products then you will always prefer Forecast to be less than Actual as surplus production is as good as loss. On the contrary if you are in the business of selling products which have reasonable shelf life then Forecast should be always greater than Actual as loss of sales is more costly than holding stock.

Reference to Base - Do you want reference to base value while measuring Forecasting Accuracy? Does a forecast error of 2 over a forecast value of 100 and an error or 2 over a forecast value of 10 mean the same to you? The answer to this question depends of what you are forecasting. If you are forecasting dispatches from factory where KPIs are "fulfilment" related like On Time In Full (OTIF), Order Fulfil Rate, Revenue Value Fulfilled on time etc. then reference to base is important since you know 100% perfection is never possible. On the other hand, if you are forecasting for Point of Sales (POS) then every unit of stock out costs you the same. 100% shelf availability is a must and there may not be any compromise on this count. In such case reference to base does not mean anything. For example, there's no point in achieving shelf availability of 98% and missing 2% availability during pick times.

Error Spread - Does it matter to you if forecast error is concentrated on few points or do you want forecast error to be spread evenly? 100% forecast accuracy is impossible, some error is bound to happen. The question is what is your preference for distribution of error? Do we want it to be evenly distributed over all points or you don`t mind if you go horribly wrong on few points and look good if calculated for the entire horizon. The answer to this depends upon type of products you are dealing in. If you are dealing with products that have very long shelf life which means unsold stock from one period can be effectively put into use in subsequent periods then we don`t mind going horribly wrong with the forecast in one period as long as we make up for that in the horizon. Forecast Accuracy of the complete horizon is more important in such case than accuracy of an individual month or a week.

Weightage to History Data Points - Do you want to treat error on every point equally or you want to differentiate between individual data points based on their place in history? For example, when you are analysing forecast accuracy do you care if total forecast error is contributed more by recent past than by distant past? In a fashion business it does make lot of difference. Fitment over recent past is more important than fitment over distant past. Again, the choice of this decision depends on what you are forecasting. If forecast is meant for long term investment planning in planning for facility then all data points mean the same. However, if forecast is for POS and influenced greatly by consumer behaviour then close fitment over recent history is more important than fitment over distant history.

Map the standard Forecast Accuracy measures to each of the previous criteria.

| Measure	MAD	MAPE	MPE
Measure Sensitive to Sign of Error?	No	No	Yes
Measure has reference to Base?	No	Yes	Yes
Measure gives equal importance to all data points?	No	Yes	Yes
Measure sensitive to distribution of over data points?	No	No	No

Using Solver to Optimise Forecasts

Installing Solver

Solver is part of the Excel application but as it is an Add-In you have to install it in order to use it. You should only have to do this once. In Excel for Windows click the File tab, then Options and then Add-ins. Solver is then available on the Data tab. In Excel for Mac, click the Tools menu, then Add-ins. Solver is then available on the Tools menu.

Using Solver to Optimise Weighted Moving Average Forecasts

The Weighted Moving Average forecasting model needs a weight/percentage value for previous periods. Instead of guesswork analysing the data with Solver can determine the right weights to get the most accurate forecast.

	A	B	C	D	E	F	G	H	I	J	K
1	Period	Sales	Forecast	Error			Weighted Moving Average				
2	1	125,000					i.e. most recent performance is the best predictor				
3	2	135,000					Period 1	10%			
4	3	145,000					Period 2	30%			
5	4	155,000	140,000	=ABS(B5-C5)			Period 3	60%			
6	5	250,000	150,000	ABS(number)			SUM	100%			
7	6	225,000	211,000	14,000							
8	7	245,000	225,500	19,500		MAD	31,667				
9	8	220,000	239,500	19,500			Need the lowest MAD				
10	9	206,000	228,000	22,000			Constraint is that the sum of the weights is 100%				
11	10		214,100								
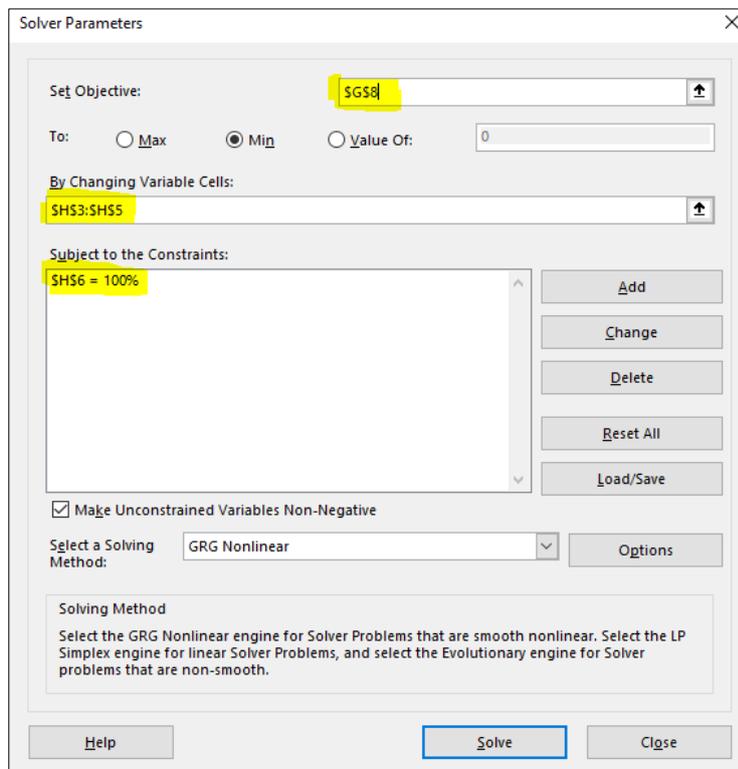
The Weighted Moving Average model uses a weight for 1st period of 10%, 30% for the 2nd period and 70% for the 3rd period. These percentages give a MAD of 31,667. The formula in cell C5 for the first forecast period is:

=(B4*$H$5)+(B3*$H$4)+(B2*$H$3)

Solver parameters:

The **Set Objective**: is the cell with MAD. To: **Min** because you want the lowest possible MAD to get the most accurate forecast.

By Changing Variable Cells: is the range of cells containing the weights.

Subject to the Constraints: The constraint is entered as the sum of the weights must be equal to 1 (100%).

Solver calculates the new weighting required to minimise MAD:

	A	B	C	D	E	F	G	H	I	J	K
1	Period	Sales	Forecast	Error			Weighted Moving Average				
2	1	125,000					i.e. most recent performance is the best predictor				
3	2	135,000					Period 1	7%			
4	3	145,000					Period 2	19%			
5	4	155,000	141,750	13,250			Period 3	74%			
6	5	250,000	151,750	98,250			SUM	100%			
7	6	225,000	225,000	0							
8	7	245,000	224,828	20,172		MAD	28,716				
9	8	220,000	241,611	21,611			Need the lowest MAD				
10	9	206,000	225,014	19,014			Constraint is that the sum of the weights is 100%				
11	10		211,311								

And the forecast can now be plotted:
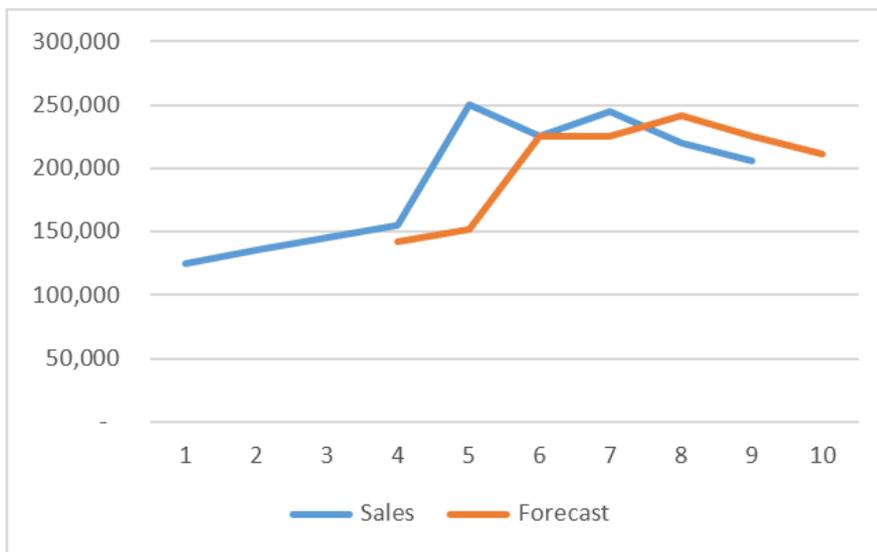
Using Solver to Optimise Seasonal Forecasts

The Seasonal forecasting model needs a weighted value for each period. The weights must equal the number of periods. Solver will calculate the right weights to get the most accurate forecast by minimising the Mean Absolute Deviation.

In the example below the Seasonal forecast model uses the following weights:

Q1   0.9
Q2   0.7
Q3   1.5
Q4   0.9

These weights sum to 4 and return a MAD of 50.

The formula in cell I10, the MAD, is =AVERAGE(F2:F13)

The formula in cell I11, the SUM, is =SUM(I2:I5)

	A	B	C	D	E	F	G	H	I	
1	Period	Quarter	Actual	Forecast	Seasonal Forecast	Deviation				
2	1	Q1	100	111	100	0		Q1	0.9	
3	2	Q2	80	114	80	0		Q2	0.7	
4	3	Q3	110	117	176	66		Q3	1.5	
5	4	Q4	180	120	108	72		Q4	0.9	
6	5	Q1	120	123	111	9				
7	6	Q2	175	127	89	86		*For Quarters, seasonal*		
8	7	Q3	95	130	195	100		*indices must sum to 4.*		
9	8	Q4	185	133	120	65				
10	9	Q1	75	136	123	48		MAD	50	
11	10	Q2	110	139	98	12		SUM	4	
12	11	Q3	123	143	214	91				
13	12	Q4	185	146	131	54				
14	13	Q1		149	134					
15	14	Q2		152	106					

The Solver parameters:

**Set Objective**: I10 is the cell with the MAD.

To: Min because you want the lowest possible MAD to get the most accurate forecast.

By Changing Variable Cells: the range I2:I5 are the weights that Solver must change to minimise MAD.

Subject to the Constraints: I11 must equal 4. The sum of the quarterly weights.
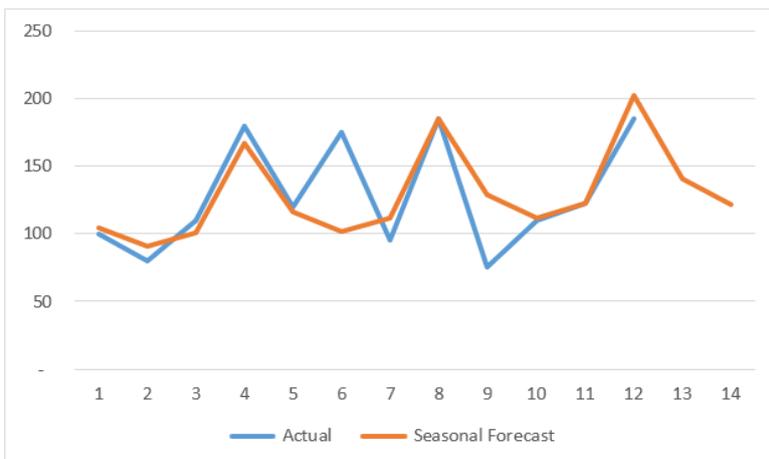
	A	B	C	D	E	F	G	H	I
1	Period	Quarter	Actual	Forecast	Seasonal Forecast	Deviation			
2	1	Q1	100	111	105	5		Q1	0.94531
3	2	Q2	80	114	91	11		Q2	0.80012
4	3	Q3	110	117	101	9		Q3	0.86306
5	4	Q4	180	120	167	13		Q4	1.39151
6	5	Q1	120	123	117	3			
7	6	Q2	175	127	101	74		*For Quarters, seasonal*	
8	7	Q3	95	130	112	17		*indices must sum to 4.*	
9	8	Q4	185	133	185	0			
10	9	Q1	75	136	129	54		MAD	17
11	10	Q2	110	139	111	1		SUM	4
12	11	Q3	123	143	123	0			
13	12	Q4	185	146	203	18			
14	13	Q1		149	141				
15	14	Q2		152	122				
16	15	Q3							

Solver has calculated new weighting and MAD is reduced to 17.
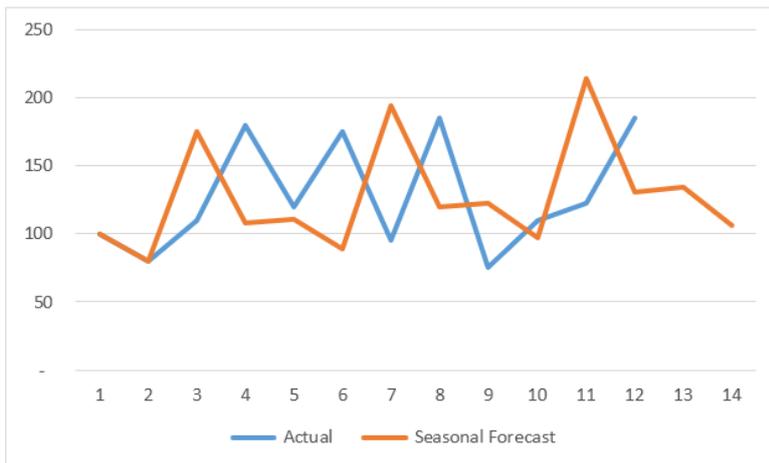
Compare the plots of the seasonal forecast before and after the optimising.

Before

After