**Chapter 1a**

**Predictive Analytics I**

**Time Series Analysis- Moving Average**

***If you can look into the seeds of time and say which grain will grow and which will not, then unto me. William Shakespeare, 1564-1616.***

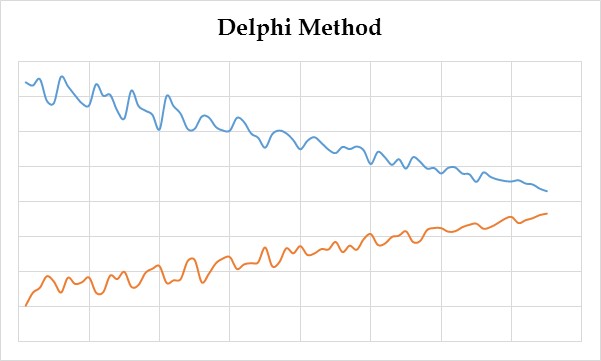
# **Introduction.** Marketing, finance, and operations are manufacturing and service systems' three fundamental building blocks. Planning, organizing, budgeting, executing, and controlling are the primary responsibilities of the three key managers. Operations Managers need forecasting for capacity planning, break-even analysis, inventory management, aggregate planning, etc. Financial Managers need forecasting for investment analysis, revenue and cost analysis, cash flow planning, etc. Marketing Managers need forecasting for pricing, sales force planning, promotions, etc.

# Marketing, finance, and operation functions need long-term and short-term forecasts for strategic, tactical, and operational decisions. They need forecasting to develop integrated resources and processes to support multi-dimensional, flexible, and structurally integrated capabilities. Capabilities to understand the revolving business eco-system, continually create value, and re-shape itself to survive, exist, and grow. While the three key managers may be interested in forecasting different variables, they have a common interest in the sales forecast.

There are two general forecasting methods, **quantitative** and **qualitative**.

**Quantitative forecast** uses historical data such as previous sales, revenues, production mix, production volumes, and financial reports. The two main types of quantitative forecasting are time series analysis and regression or associative analysis.

**Qualitative forecasting.**Qualitative forecast relies on the subjective opinion and intuition of experts in the field.  An example of the qualitative forecasting technique is the **Delphi method**, which relies on a panel of experts who answer questionnaires in several rounds. After each round, a [facilitator](https://en.wikipedia.org/wiki/Facilitator) panel provides an anonymous summary of the experts’ forecasts in the most recent round and the reasons for each class of judgments. The experts are encouraged to revise their earlier responses in the light of the categorized responses provided by the others in the opposite opinion sides. It is believed that the gaps between different opinions will decrease during this process, and the whole group will converge on the most logical conclusions. The Delphi method is pictorially represented in the following figure.

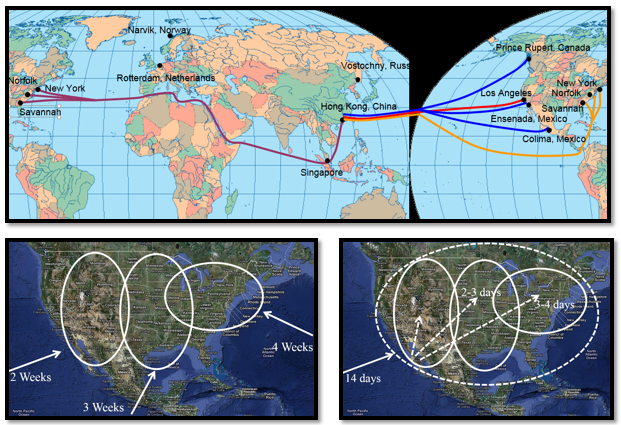


We devote the rest of this chapter to quantitative forecasting. While our variable of interest throughout the example is the sales volume, the ideas, concepts, and methods can be applied to any other variable.

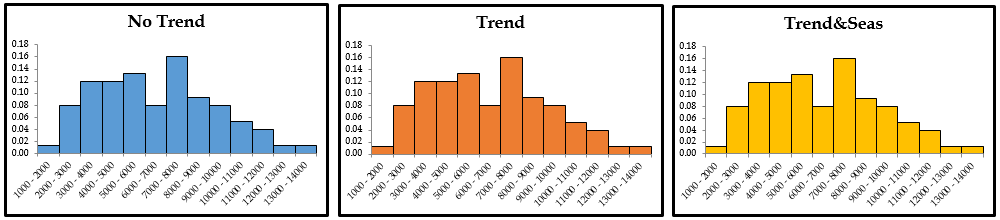
**Characteristics of Forecasting Techniques.** All forecasting techniques have three main characteristics in common.

1. **Forecasts are usually wrong**. Since the world is not deterministic – at least to us- all forecasts are almost always incorrect. Forecasts provide the expected value or average (μ) for the variable of interest -sales or demand in our case. Demand is a random variable usually following Poisson or Normal distribution. Thus, besides the average demand, we should always accompany our forecast with a measure of variability- standard deviation, variance, or the coefficient of variations. If our average forecast for next year is F, and the standard deviation of F is s, the coefficient of variations of our forecast CV= s/F. The lower the coefficient of variations, the better the forecast. Which one of the following distributions shows more variability? A distribution with a standard deviation of 100, or a distribution with a standard deviation of 100,000? We do not know. It depends on the averages. If the average of the first distribution is 100 and the average of the other distribution is 1,000,000, then they have CV1 = 100/100=1 and CV2=100000/1000000=0.1. The second distribution shows less variability.
2. **Forecasts for aggregate items are more accurate than the forecast for individual items.** Aggregate forecasts reduce the amount of variability. Forecasts for the entire U.S. economy is more accurate than the forecasts for the manufacturing, agriculture, and service sectors. A forecast for manufacturing is more accurate than a forecast for the car industry. A forecast for the car industry is more accurate than a forecast for Ford Motors. A forecast for Ford Motors is more accurate than a forecast for Mustangs, and a forecast for a Mustang is more accurate than the forecast for a Mustang convertible. Aggregate forecasts reduce the range of variability. We can intuitively understand that the forecast for the summation of the two products is more accurate than the forecast for each product since the low demand for one product can compensate for the high demand for the other. Therefore, the summation of the demand for the two products will show less variability around the average that we forecast. From a mathematical point of view, suppose we have a random variable of x, which shows the demand for Ford Fiesta, and a random variable z showing the demand for Ford EcoSport. For simplicity, suppose the average and standard deviation of z are the same as that of x.
3. Suppose both cars use the same engine, and we define a new random variable, y, as the summation of the two variables (y=x+z) to estimate the required number of engines. For simplicity, suppose the average (μ) and the standard deviation (σ) of these two random variables are equal. When two random variables are added, then μ(y) = μ(x) + μ(z), and since μ(x) = μ(z) = μ, therefore, μ(y) = μ+ μ=2 μ. But σ(y) is NOT σ(x) + σ(z). In other words, σ(y) ≠ σ(x) + σ(z). The variances of the two variables are added, not their standard deviation. VAR(y) =VAR(x) +VAR (z). Therefore, σy2 = σ2+σ2= 2σ2  🡺 σy2 =SQRT (2) σ 🡺 σy2= 1.41σ. Therefore, the standard deviation of a variable defined as the summation of two other variables is less than twice the standard deviation of each variable. The same is true if 2 is replaced by 3, 4, n.
4. Long-term forecasts are less accurate than short-term forecasts. Forecast accuracy diminishes as we look further into the future. As time passes, we get better information and make better predictions.

**How enterprises may benefit from forecasting techniques characteristics.** The combined ports of Los Angeles and Long Beach, also known as San Pedro Bay Ports, are ranked 5th in the world with respect to container handling, after the Port of Singapore and three ports in China. More than 50 percent of containers coming to the U.S. pass through San Pedro Bay ports, and more than 1/3 of the containerized products consumed in all other states pass through these ports. The total value of the trade is about $300 billion, creating around $30 billion in state and local taxes and 3 million full-time equivalent jobs. San Pedro Bay ports and Southern California need to retain their competitive edge. Otherwise, other potential routes will take business from Southern California and its ports. What are the competing edges of Southern California ports? Deepwater facilities for Post-Panama ships, which may contain more than 8,000 containers; state-of-the-art on-dock facilities to transfer containers between ship and train; intermodal transfer between ship, truck, and train; consolidation and distribution facilities for trans-loading from 20-foot containers and 40-foot containers to 56-foot containers, which are allowed to move on California roads, but as crucial as this capability is and maybe more important than this capability, are the common characteristics of all forecasting techniques.



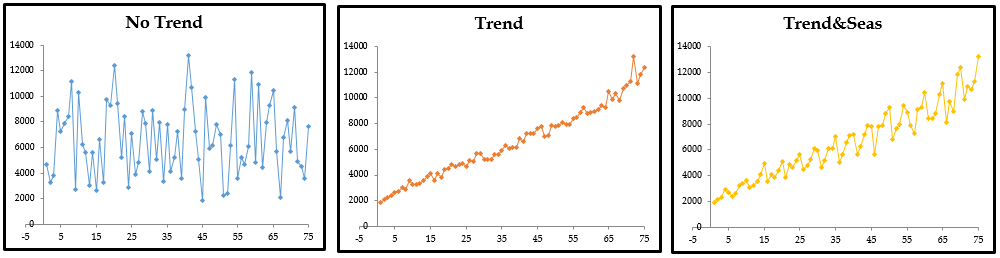
If we want to transfer the load from the Far East to the East Coast, it will take four weeks. From the Far East to West Coast takes two weeks, and from the Far East to the mid-U.S. takes something between 2 to 4 weeks. Now, if I am going to shiploads from the Far East to East Coast, I should forecast the demand of the East Coast 4 weeks in advance. If I ship from the Far East to West Coast, I should estimate the demand for the West Coast 2 weeks in advance. Estimates of the West Coast, which require a forecast of 2 weeks, are more accurate than for the East Coast, which requires a 4-week forecast. Shorter time provides more accuracy. Look at the other property. All of the forecasts for the East Coast, West Coast, and mid-U.S. are less accurate than the forecast for the total demand in the U.S. So instead of forecasting for the East Coast alone for 4 weeks, the West Coast alone for 2 weeks, and the mid-U.S. for 3 weeks, I forecast the total demand for the U.S. for 14 days, 2 weeks in advance. Then when I send the container here, in one day, I may transfer it to anywhere in California, in 2-3 days, I may transfer it to somewhere in the mid-U.S., and in 3 to 4 days, I may transfer it to somewhere in the East Coast. Now instead of estimating the demand of the East Coast alone, which is less accurate than the demand for the whole U.S., and instead of forecasting it for four weeks from now, I can forecast it for 14 days plus three days, which is 17 days from now. The forecast for the whole U.S. between 14 days and 17 days in advance is much more accurate than the forecast for the East Coast 4 weeks in advance and the forecast for the mid-U.S. 3 weeks in advance.

**Time Series Analysis.** TheHistograms for three sets of data are shown below. They look the same. 

The descriptive statistics of these three sets are the same with all decimal points.



But, they are time series. They have happened over a set of equal time intervals. While the descriptive statistics as a snapshot of each set are the same, the sets are very different if we draw them as they have been observed along a time-related axis. Time series analysis is a quantitative forecasting technique applying a series of historical data collected regularly (e.g., hourly, daily, weekly, monthly, and yearly).



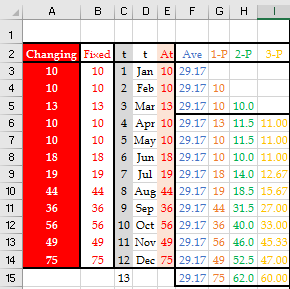
**Time series** is based on analyzing past data to identify system components and extend them into the future for better forecasting. To gain the capability to extend those into the future and show some random components. This is done concerning any variable of interest, but because we are discussing forecasting for demand, a value of interest always demands unless stated otherwise. Time series may contain three systematic components.

* **Level.** Where the demand is.
* **Trend.** In what direction the demand will move. Upward or downward?
* **Seasonality.** Predictable fluctuations. Does the demand peak (or valley) around specific times?

We can identify and quantify systematic components, but there is no way to predict or control random components. That is why forecasts are never precise.

There are two well-known time series analysis methods, (*i*) Moving Average and (*ii*) Exponential Smoothing. Let's compare two extreme points to set the stage for the moving average. On the extreme left, we may decide to have our forecast for the next period be equal to the actual for this period. On the extreme right, we may set our forecast for the next period as equal to the average of all the data for all earlier periods. Two extreme points, one relies only on one piece of data, and the other relies on all pieces of data.

**Problem 1- All Periods vs. Last Period.** In some of the examples I use, the input data are random. That means the input data changes when we hit enter on the keyboard. Therefore, one can extract an almost infinite number of problems from a single problem. However, in order not to allow the data to change when solving a problem, we usually copy the random data and paste it by value into a different part of the Excel sheet. We then work on the paste-by value data, which are now fixed at specific values. In the current example, random data are in column A, and we have pasted them by value in column E. In this example, the random data is in column A, where we have pasted it by value in column E. You may click on the figure below to access the Excel sheet.

[](http://www.csun.edu/~aa2035/CourseBase/Forecasting/Slides/1.MovingAverage2018.xlsx#Page1)

Suppose we have the data for the past 12 periods (days, weeks, months, etc.), and we want to forecast for the next period.In this chapter, we may interchangeably refer to actual values in period t as at (where is a subscript) or simply as At. Whenever possible, we avoid subscript in or notations. Similarly, we interchangeably refer to the forecast values in period t as Ft and Ft.

Given the past data, now we want to forecast for the next period. On the extreme right, we may stay conservative and assume that the forecast for the next period is equal to the average of the last 12 periods. Therefore,

F13 = (A12+A11+A10+………+A3+A2+A1)/12.

In that case, in cell F3, we may type = **AVERAGE (E3: E14).** However, since we need to copy it down to all cells, we may push F4, and the formula will change to = AVERAGE ($E$3:$E$14). Now we can copy it down to all periods. Following this conservative procedure, our forecast for period 13 is 29.17.

**Last Period Forecast.** From the extreme left,we may assume that the forecast for the next period is equal to the actual for this period.

Ft+1 = At

Under the last period forecast, we can forecast starting from period 2. The forecast for period 13 is equal to the actual in period 12, F13=A12 =75. The two extreme approaches are pictorially compared in the following figure. As we can see, one forecast is smooth, with no response to the current changes, and the other purely reflects the most recent changes.



The last period forecast and the average of all periods are simple, straightforward, and inexpensive. We can always use them as bases to compare the quality of other forecasting techniques. Suppose the quality of other forecasting techniques is lower. In that case, we can always use one of the last period forecasts or the average of all period’s forecasts- the better of the two.

However, as in many other situations, perhaps neither extreme right nor extreme left provides the best solution. Staying somewhere on the continuum between the two extremes may provide a better solution. That is the core idea behind the moving average approach.

One step to the right from the last period forecast is to compute the average of the last two periods. We may average every two consecutive periods and assume it as the forecast for the next period. In that case, our first forecast is for period 3 where F3= MA22 = (A2+A1)/2. Then we follow the same logic for period 4 and so on. We may also move one more step towards a smoother and less aggressive forecast and compute 3-period-MA. In that case, we cannot forecast for periods 1-3, and our first forecast is for period 4, where F4= MA33 = (A3+A2+A1)/3. For example, forecasts for period 8, using 2-period and 3-period-MAs are F8= MA27 = (A7+A6)/2, and F8= MA37 = (A7+A6+A7)/3, respectively. Using 2- and 3- period-MAs, our forecasts for period 13 are F13= (A12+A11)/2= (49+75)/2 = 62, and F13= (A12+A11+ A10)/3= (49+75+56)/3=60, respectively.

**Moving Average.** This forecasting techniqueuses the average of the most recent actual data in a certain period to forecast the next period. The moving average calculated from the current period becomes the next period's forecast, a revolving average covering ***n*** periods -but not all the periods. As time passes, the forecast is renovated based on the most recent data set. Moving Average (M.A.) calculates the forecast using the newest pieces of data and removes the oldest data.

Before writing the general moving average formula, let’s compute a 4-period-MA for period 10. We need to start from A10 and include four pieces of the actual data. What is the index of the oldest piece of data? Is it 10-4=6? No. Let’s write the equation, MA410 = (A10+A9+A8+A87)/4. Therefore, the index of the last piece of data is 10-4+1. In a 25-periods-MA, in period 100, if the youngest piece of data belongs to period 100, the oldest piece of data belongs to what period? 100-25+1 = 76

This knowledge can lead us to the general moving average forecasting formula as

Ft+1= MAnt = (At+At-1+……+At-n+1)/n

The Excel formulas in row 6 of the table for columns F to I are as follows:



Using all 1-, 2-, 3- period-MAs, our forecast for period 13 came out 29.17, 75, 62, and 64, respectively. Which one should we follow?

**Measuring Forecast Accuracy and Variability.** It’s now the time to talk about defining the measure of effectiveness to find the suitability of the forecasting technique to our specific set of data. How could I measure a forecasting technique's suitability for my specific data set? How can I compare the quality of several forecasting techniques for my specific data set?

Suppose the green dots in the following figures are our actual data, and the red dots are our forecasts based on a given method. As we can see, two forecasts under-estimate, and the other two over-estimate the demand. For simplicity, suppose the difference between actual and forecast in all periods have the same value except that two of the values are positive, and the other two are negative.



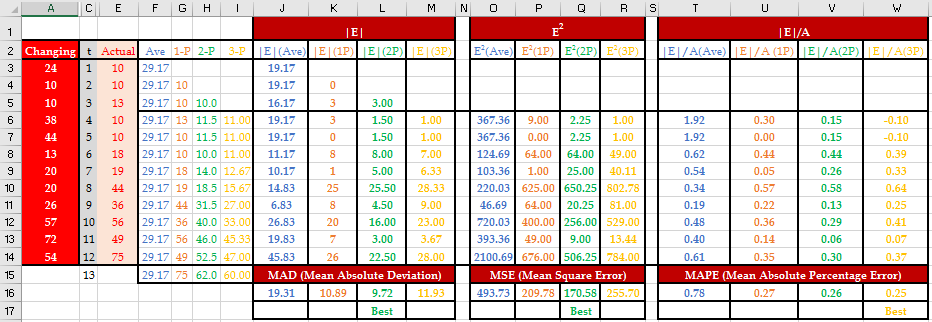
One measure of effectiveness to evaluate if a forecasting technique fits our data is to measure the gaps between actual and forecast Et=At-Ft. If we add the errors in the figure, the two negative values cross out the two positive ones, and we get an average of zero. If the average gap is 0, one may conclude that the forecasting method is perfect.

Going back to our 12-period data set, in the last period forecast, we need to compute Et= At-Ft, for t = 2 to 12 (we do not have a forecast for period 1). For 2-period-MAs, we need to compute errors for all periods t=3 to12, and for 3-period-MAs, we need to compute Et= At-Ft, for periods 4-12. Then we need to compute the summation or averages of the deviations for each method, which could be misleadingly close to 0. Negative and positive components may cross each other out, while there may be significant gaps in both positive and negative sides. Since the variations on the negative side (forecasts being greater than actual) are undesirable, the first task is to remove the negative signs. Two well-known procedures to remove negative signs; are absolute value and square. If I have a value of -2, its absolute value is +2, and its square is +4. We can compute the absolute value of columns F to I or square them. Then we can average absolute deviations or squared deviations for each of the procedures over the number of periods where Etexists for the corresponding procedure. That would be the averages of 12, 11, 10, and 9 absolute or squared errors for average, last period, 2-period, and 3-period-MAs, respectively. We formally define Mean Absolute Deviation (MAD) and Mean Squared Error (MSE).

MAD = AVERAGE(E) over the periods for which both At and Ft exist.

MSE= AVERAGE(E2) over the periods for which both At and Ft exist.

MAD for 1-period-MA is computed over 11 periods, while 3-period-MA is computed over nine periods. Suppose we compute MAD or MSE to compare two or more forecasting procedures. In that case, they must be computed only over the periods for which actual data and forecasts are available for all procedures. Therefore if we are comparing 1- and 3-period-MA, we need to compare them over periods 4 to 12. MAD and MSE are computed for all the methods over periods 4-12 in the following Excel table. Clicking on the table will take you to the Excel sheet, where you can check the different formulas in the corresponding cells.

[](http://www.csun.edu/~aa2035/CourseBase/Forecasting/Slides/MA-Main-ReplaceEveryThing.xlsx#Page1)

There are several points to mention. First, The MAD computation was more straightforward when implemented about a century ago. However, working with an absolute component in mathematical expressions is much more difficult. It is not difficult to deal with squatted functions in mathematical expressions. Second, 1.25MAD is an estimate for the standard deviation of the forecast, and MSE is also an estimate for the standard deviation of the forecast. Third, sometimes we may assign a higher weight to a positive gap (where the forecast is less than the actual demand) than a negative gap (where the forecast is estimated at more than the actual demand). In the latter case, we are overstock, while in the first case, we have lost sales. Usually, the cost of the overstock is less than the cost of lost sales. In these cases, we may assign a coefficient greater than 1 to negative values of Et=At-Ft. Alternatively, we may use the newsvendor problem insight – discussed to assign different weights to over-estimate and underestimate. Fourth, Besides MAD and MSE, there is a third method that we refer to as a relative mean absolute deviation or R-MAD. Instead of averaging |Et| values in this third method, we average |Et|/At values. For example, a |Et| of 10 states there were ten units of deviations between At and Ft. If At is 200, then ten relative to 200 is a .05 (or 5% gap). Therefore, the relative absolute gaps (relative to the demand) are considered instead of the absolute gaps. Fifth, there is another measure, the maximum absolute deviation, where the preferred method is the method with a minimal maximum absolute gap between actual and forecast.

**Mean absolute deviation serves two essential purposes.** I can use it to evaluate a single forecasting method or compare two or more methods to identify the best one. MAD also provides an estimate of the standard deviation of the demand forecast. The standard deviation of the forecast is equal to 1.25 MAD. Therefore, we can have Ft+1 as the average demand for the next period and 1.25 of the most recent MAD as the standard deviation of the forecast.

**Now we are prepared to watch the lectures on excel based moving average embedded in the excel file at**

<https://www.csun.edu/~aa2035/CourseBase/Forecasting/MA-2020/1.1.MA-Game50.xlsx>

<https://www.csun.edu/~aa2035/CourseBase/Forecasting/MA-2020/1.2.MA.YouTube.xlsx>

After understanding all the functions in the Excel sheet, please set column E equal to column A, and then go to an empty cell and click delete bottom several times. Each time you hit delete, your data changes, and there are two interesting observations. First, the three methods do not necessarily get the same results, as one may prefer a 2-period-MA and the other a 3-period-MA. Second, if we only consider one of the three measures, it will not always point to a specific method as the best method. For example, as we change data, for one set of data, in the framework of MAD measure of effectiveness, each time we may find a different method (average, 1-, 2, or 3-period-MA) preferable. That means the choice of the method depends on the nature of the data.

**The tracking signal** is defined as the summation of the difference between actual and forecast, so we keep the positive and negative signs.

If actual is greater than forecast, the sign is positive. Otherwise, it is negative divided by MAD. And because MAD by itself is a summation of the absolute value of the difference between actual and forecast divided by the number of periods, then if I replace MAD with this equation, I will get this formula. Note that MAD is always positive because it is a summation of the actual forecast with absolute value signs divided by the number of periods. Therefore, MAD is always positive, but the numerator of the tracking signal can be positive or negative. So tracking signal by itself could be positive or negative.

TS=SUM(Et)/MAD

We also define an LCL and a UCL for T.S. In some textbooks, it is assumed that the tracking signal should change between -4 and +4. But one can mathematically prove that these numbers are not correct.

**Age of Data.** What is the age of data in a five-period moving average? The most recent piece of data is one period old, the second piece of data is two periods old, the third piece of data is three periods old, the fourth piece of data is four periods old, and the fifth piece of data is five periods old. On average, data is (1+2+3+4+5)/=15/5=3 periods old. We could have reached the same number by stating that the newest piece of data is one period old, the oldest piece of data is five periods old, therefore, on average, the data is (1+5)/2 = 3 periods old. In general, in an n-period moving average, the data is (1+n)/2 periods old.

**Weighted Moving Average.** A weighted moving average allows assigning different weights to the data used in a moving average. Usually, the weights go down as the age of the data in the moving average is increased. For example, usually the weight of the most recent piece of data, i.e., this period’s data, is higher than the weight of all other data contributing to the moving average. The sum of the weights must be equal to 1 (100%). If it is not, we need to transform it to 1 by defining each new weight to be equal to the old weight divided by the sum of the weights. Forecasting based on an n period weighted moving average is stated as follows

Ft = w1At-1 + w2At-2 + w3At-3 + ……….+ wnAt-n

Example: Determine the forecast for period 13 in the following data using a four-period moving average where the weights are w1=0.4, w2=0.3, w3=0.2, and w4=0.1. We need to multiply the demand of period 12 (the most recent period) by the weights of 0.5, the demand of period 11 (the second most recent period) by the weights of 0.3, the demand of period 10 (the third most recent period) by the weights of 0.2, and the demand of period 9 (the oldest period) by the weights of 0.1, and add them together.



Use solver to find the best weights were we should satisfy the w4≥ w3≥ w2≥ w1 constraint.

