This is our first lecture on forecasting. All functions and processes of every organization, profit and nonprofit, needs forecasting. For example, accounting needs forecasting for break analysis and estimating profit and loss. Finance needs forecasting to have an estimate of cash flow, finding the best sources for cash deficit, and the best uses for cash surplus. Human resource needs forecasting for hiring and training. Marketing needs forecasting to know what is the stomach of the market, what is the trend, and the seasonalities of the market, so one could properly use them for pricing and promotion policies. Operations needs forecasting for production planning for event management and for scheduling.

In all forecasting lectures, we talk about qualitative and quantitative techniques. Regarding qualitative techniques, we just mentioned a technique called Delphi. For quantitative techniques, we talk about time series analysis and causal relationship forecasting. In time series analysis, we talk about moving average and exponential smoothing. Causal relationships consist of linear and non-linear, single and multi-variable regression. Finally, we will talk about some measures to find out which forecasting technique is the best for our specific set of data.

All forecasting techniques have four characteristics. The first characteristic of forecasts is that they are usually always inaccurate and they are, most of the time, wrong. There is a very tiny possibility that a forecast comes out to what really happens in reality; therefore, we should always accompany our forecast with a measure of variability, and that measure of variability is standard of deviation or variance. Our forecast should say, "our forecast for next year is this much, and the standard deviation for this forecast is that much." The lower the standard deviation, the better the forecast. Four characteristics of forecasting techniques is forecasts for aggregate items are more accurate than forecasts for individual items. Forecasts for the total use economy is more accurate than forecasts for subtotals of manufacturing, agriculture, and such. An example is that 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, and the forecast for Ford Motors is more accurate than a forecast for Mustangs. Aggregate forecasts reduce the amount of variability with respect to the average forecast. We will later show these mathematical characteristics, but right now we can intuitively accept that the forecast for the summation of two products is more accurate than the forecast for individual products, because one product can be compensated by the other product and all forecasts come out closer to the average forecast.

The fourth characteristic of forecasting techniques states that long-range forecasts are less accurate than short-range forecasts. Forecasts further into the future tend to be less accurate than forecasts for the near future. As time passes, we get better information, and make better predictions.

Now, I would like to give you an example on how enterprises may benefit from the characteristics of all forecasting techniques. The combined ports of Los Angeles and Long Beach, also known as San Pedro Bay Port, from the point of container handling are ranked 5th in the world, after the Port of Singapore and three ports in China. More than 50 percent of containers coming to the United States 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 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? Deep water 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 important as this capability is and maybe more important than this capability, are the two last characteristics of all forecasting techniques.

If we want to transfer the load from Far East to East Coast, it will take 4 weeks. From Far East to West Coast, it takes 2 weeks, and from Far East to the mid-United States it takes something between 2 to 4 weeks. Now, if I am going to ship loads from Far East to East Coast, I should forecast the demand of the East Coast 4 weeks in advance. If I am going to ship from Far East to West Coast, I should estimate the demand for the West Coast 2 weeks in advance. Estimates of West Coast, which requires a forecast of 2 weeks, is more accurate than for the East Coasts, which is 4 weeks in advance. Shorter time provides more accuracy. Look at the other property. The forecast for East Coast, West Coast, and mid-United States, all of them are less accurate than the forecast for the total demand in the United States. So instead of forecasting for East Coast alone for 4 weeks and West Coast alone for 2 weeks and mid-United States for 3 weeks, I forecast the demand for all the United States 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 to somewhere in the mid-United States and 3 to 4 days 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 United States, and instead of forecasting it for 4 weeks from now, I can forecast it for 14 days plus 3 days, which is 17 days from now. The forecast for the United States for the whole United States between 14 days and 17 days in advance is much more accurate than the forecast for the East Coast, 4 weeks in advance and forecast for the mid-United States, which is 3 weeks in advance. We said we would talk about different forecasting techniques, and first we talked about qualitative techniques. Delphi is one of the techniques in which we ask opinions of experts with respect to something. For example, we collect data with respect to the demand of our product. Some of them make estimated demands. We summarize the point of departures of these experts and send it to these people, and summarize points of departure of these people and send it to the pessimistic people and try to collect the next round of opinions. In the next round these people may become a little bit – may bring their forecast a little bit lower. And these people, after hearing the points of departures of optimistic people, may bring the estimates a little bit up. Again, we summarize these people, send it to the optimistic people and summarize the ideas of optimistic people and send it to pessimistic people, and after several rounds these people will convert somewhere, and this level will be the task demand for the coming year.

Time series is based on analysis of the past data and trying to find a trend for future. This is done with respect to any variable of interest, but because we are talking about forecasting for demand, always a value of interest is demand unless stated otherwise. Time series can show some systematic component and some random components. You cannot do anything about random components, but we can identify three types of systematic components. The first one is level; where we think our demand is, the second one is trend; how it will change over time upward or downward, and the third one if there is any seasonality. We can identify and quantify systematic components. Nothing can be done about random components. That is why no forecast can come out the same as actual because we are living in a probabilistic war and can’t be 100 percent predicted.

Regarding time series, we talk about three techniques; naïve forecast, moving average, and exponential smoothing. In naïve forecast, forecast for next period is equal to actual for this period. We use F for forecast and A for actual. T is this period, can be anything 1, 2, 3, 4, 5. T plus 1 is the next period. So if T is 1, T plus one is 2. If T is 5, T plus 1 is 6. At, actual demand in period t. F (t + 1) is forecast of demand for period t + 1. Naïve technique is very cheap, very simple, and easy. We can always use it as a base to compare quality of other forecasting techniques. If quality of other forecasting technique is lower than naïve technique, then always use naïve technique, which is the cheapest, easiest, and performs better.

Before going to moving average, let’s compare two extreme points. In one extreme point we may say my forecast for next period is equal to actual for this period. The other extreme point is my forecast for next period is average of all pieces of 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. Obviously between these two extremes, we can have some other techniques, which rely on say 3 most recent pieces of data, 7 most recent pieces of data, or 100 most recent pieces of data. Three period moving average in period 7 is A7 + A6 + A5 divided by 3. We use this notation. This shows 3 period moving average and shows we are in period 7, so the most recent period data is 7. Then this is one period old. This is two period old. Then we remove these three because when I write here 3, that means we are talking about 3 period moving average. Three period moving average in period T, the most recent piece of data is At. One period old is At-1. Two period old is At-2. So I have 3 pieces of data divided by 3. That would be 3 period moving average in period T. 10 period moving average in period t is At + At-1 + At-2 + plus what? What is the last piece of data? So we are talking about moving average in period t. And we are talking about 10 period moving average. What would be the last piece of data? At-10 + 1, which is At-9.

N period moving average in period t is At + At-1 + At-2 + At-n + 1 divided by N. Then we assume the forecast for period t + 1 is equal to moving average for period t. For example, if you are in period 10, and we are talking about 5 periods moving average of period 10, we assume as forecast for the next period, which is 10 + 1, which is 11, which is equal to A10 + A10-1, which is 9 + 10-2 which is 8 + A10-5 + 1, which is A6. And here I have A7 + A6 divided by 5.

Now suppose we have data for the 12th period. We want to compute 3 periods moving average and 6 periods moving average for this data, and finally we want to forecast for the next period using 3 periods moving average and 6 periods moving average, and would like to compare the quality between these two forecasts. On the left-hand side of the screen I have 3 period moving average, and on the right-hand side I have 6 period moving average. After 3 period I can compute 3 period moving average, which is 1366, and I have gotten these numbers by adding the first 3 periods and dividing them by 3. For 6 period moving average I add up the first 6 periods and divide them by 6 and then continue for all other periods. So if I am going to compute 3 period moving average for period 4, I can simply add these 3 numbers and divide by 3, and I will get 1458. However, there is also a smarter way to compute 1458. I know that I have 1366 as moving average of the first 3 periods. So what is the difference between the moving average in the second 3 periods compared to the moving average in the first 3 periods? These two numbers are the same in both computations, but in the second 3 period moving average, I have added 1576 while I have dropped 1300. In other words, in the first 3 periods, I have 1300 + 1356 + 1442 divided by 3. In the second 3 period moving average, still I have 1356 and still I have 1442, but there is 1576 also over there, and I have divided all of them by 3. So this part is common between the first 3 period moving average and the second 3 period moving average. In the second 3 period moving average I have dropped 1300 divided by 3, and I have added 1576 divided by 3. In order to obtain 1458 using 1366, all I need is to take 1366 add 1576 to it divided by 3 and subtract 1300 divided by 3. In other words, I have 1366 here + 1576 minus 1300 divided by 3, 276 divided by 3 is 92. Therefore, 92 + 1366 is equal to 1458. In order to obtain, for example, 2154, all I need is 2092 and the most recent piece of data, which I have observed in period 11, and that is 2214, which is the most recent piece of data. And in my previous computation for 2092 I had these 3 numbers added to each other divided by 3; therefore, the oldest piece of data is 2214 minus 2028 divided by 3.186 divided by 3. And 186 divided by 3 is 62. And 62 + 2092 is 2154.

Now I can compute any new N period moving average using the previous N period moving average by a simple computation. I write down the previous moving average, add the most recent piece of data, subtract the oldest piece of data and divide it by the number of periods. Here I have computed the first 6 period moving average by adding the first 6 pieces of data I have and dividing them by 6. Now in order to compute the next 6 period moving average, all I need to do is to take the previous 6 period moving average, which is 1537, and then add to it something and divide it by 6. What I should add is the most recent piece of data, 1996, drop the oldest piece of data, 1300, 1537, 696 divided by 6. 1537 plus 116, and this is equal to 1653. When I am talking about period 7, the oldest piece of data is for period 1, but when I talk about period 11, the oldest piece of data now is A5. It’s the time to talk about defining the measure of effectiveness to find the suitability of the forecasting technique to our specific set of data. Here I have a new set of data, and I have computed 3 period moving average and have assumed it as forecast for the next period. These are my forecasts, and these are my actuals. How could I know if this is a good forecasting technique? Let me go through a simple example. Suppose these are our actual values for 4 periods. These are my forecasts for these periods. And for simplicity, suppose the difference between actual and forecast in all periods have the same absolute fact. This value and this value and this value are all equal. However, here and here forecasts are graded in actual while here and here the actuals are more than forecast. One measure of effectiveness to find if a forecasting technique is good for our data is to add the differences, to add the errors. So if I add this value and to this and add to this, because two of these values are negative and two are positive, the summation would be equal to 0. So that would tell me the forecasting technique is excellent because the summation of the errors have come out 0. Adding the actual values is two negative values and these two positive values are not correct. I should find a way to remove the sign no matter if these values are positive or negative. I should remove the sign and add them together. There are two ways to remove the sign. The first one, if I have a value of A, I can put absolute value of A, and that will remove the sign. If I have negative 2, the absolute value of negative 2 is positive 2. The other way of removing a sign is if I have a value of A, I can square it. This will also remove the sign, because if I have negative 2, when I square it, it becomes 4. We have seen some of the absolute values and defined mean absolute deviation, which is the summation of the absolute value of the difference between actual and forecast divided bythe number of observations. Here are actual demands. Here are 3 period moving averages, and here are my forecasts. So my forecasts start from period 4 up. In period 4, my forecast is 1382. My actual is 1470. The difference is 88. An absolute value of that difference is 88. In the second period, the actual is 1008. Forecast is 1416. The difference is negative. If I compute the actual value of that negative, I put it right here. I will do the same for all pieces of data. Then I will add all these numbers. All of them are positive values. And I have them for 9 periods. I add up these numbers, divide it by 9, and I will get 202. This is my MAD, mean absolute deviation. My forecast for period 13 is 1450, and mean absolute deviation for this forecasting technique is 202. Mean absolute deviation serves two important purposes. I can use it to compare two or more forecasting techniques and find out which technique is better is standard deviation of forecast is equal to 1.25 times MAD. Now I have my Ft as my average forecast for the next period, and also I have MAD multiplied by 1.25, which is the standard deviation of my forecast. So using 3 period moving average, my forecast for next period is 1450, and my standard deviation of forecast is 202 multiplied by 1.25, which is equal to 250.25. Let’s just write 250. My forecast for the next period has normal distribution with mean of Ft, and standard deviation of 1.25 MAD, which in this specific case my forecast for period 13 has average of 1450 with standard deviation of 250, therefore, I can expect that the actual demand for the next period with more than 99 percent probability to be between 1450-750 to 1450 + 750. And that is a characteristic of normal distribution. More than 99 percent of observations will be between mean plus 3 standard deviation and mean minus 3 standard deviation.

Let’s summarize, for each period I will find the difference between actual and forecast and then remove its sign. Then I will add them up, and then I will divide it by the number of observations. This would be my MAD. Standard deviation of my forecast would be 1.25 times MAD. I can use MAD to serve two purposes. The first one to find out which forecasting technique 3 period moving average, 6 period moving average, 9 period moving average, 13 period moving average, which technique is the most appropriate technique for a specific set of data? The most proper technique which has the mean of MAD. And the second application of MAD is standard deviation of my forecast of 1.25 MAD. However, when I compare two or more forecasting techniques, I should compare them over the same number of periods. Now I have this set of data. You may want to know which technique is 3 period moving average or is 6 period moving average is more appropriate for these specific set of data. So I compute 3 period moving average, and in each period I have moving average for that period equal to the forecast for the next period. I also compute 6 period moving averages and moving average of each period is set to the forecast for the next period. For 3 period forecasts I have my forecast from period 4 on. For 6 period forecasts I have my forecast from period 7 on, but when I compare two techniques, I should compare them over the same number of periods because here I don’t have 6 period forecasts, so I drop this set of data, and I will compare the two techniques from period 7 to period 12. For each period I will find the difference between actual and forecast, and if it is negative, I remove the sign. I will do the same for 6 period moving average. I will find the difference between actual and forecast, and if I have a negative sign, I remove it. Using 3 period moving average, my forecast for next period is 1450. Using 6 period moving average my forecast for next period is 1519. Which one is better? The forecast cannot tell us which one is better. It is MAD that can tell us which technique is better. 3 period moving average has a MAD of almost 160. 6 period moving average has a MAD of 195. 160 is less than 195. For this specific set of data, 3 period moving average is better than 6 period moving average. Our forecast for period 13 has a mean of 1450, and the standard deviation of this forecast is 1.25 times 160, which is 200. Therefore, forecast for the next period is normally distributed with mean of 1450, and the standard deviation of 200. Here is our data, actual data, 3 period moving average and 6 period moving average. The longer the number of periods, the smoother the data. The smaller the number of period, the more reactive the forecasting technique to the percent changes in the actual data. For this specific example, 3 period moving average performs better than 6 period moving average. Should we always use as many periods as we can? We don’t know. The mission of effectiveness is MAD. The forecasting technique with lower MAD is the better technique.

Now let’s look at another set of data. This is a different set of data for which we have computed 6 period moving average and 3 period moving average, and we can compare them from period 7 on. 3 period moving average has a forecast of 1590, 6 period moving average 1540. These are just average forecasts, and we cannot judge two different techniques based on their forecasts. We should look at – we need to look at MAD. For 3 period MAD is 293. For 6 period, MAD is 254. Therefore, for this specific set of data, 6 period moving average forecasting technique is a better fit because it will end up with a lower standard. Because it ends up with a lower MAD. Our forecast for next period has an average of 1540, and the standard deviation of 1.25 times 254, almost 254. 317. Average 1540, standard deviation 317.

Again, as we see, 6 period moving average is smoother and 3 period moving average is more reactive. The larger the number of periods, the smoother the curve. If you are a long term in the store, you look for larger number of periods to smooth the forecast and see what the general trend is. If you are a day trader, you stay with smaller number of periods to be more reactive to the most percent changes in data. Here I have shown 5-day, 20-day, 50-day, and 100-day moving average for Microsoft stock. As you see, 5 days moving average is more reactive to the most percent changes. 100 days moving average is more smoother and shows the general trend. For example, if you are here, and you want to put some money in your 401K, and you realize Microsoft is not a good investment, but as you move in these periods, you see that it is not a bad idea if you invest a fraction of a 401K in Microsoft. Again, here you may think it’s not good to invest more. We said our measure of effectiveness to select a technique among several techniques for forecasting is MAD. The lower the MAD, the better of the forecasting technique. Now we define a second measure of effectiveness to help us find out if our forecasting technique is reliable. Reliability means quality over time. Reliability means that the forecasting technique is good, and it will remain good over time. We define tracking signal as summation of the difference between actual and forecast, so we keep the sign. If actual is greater than forecast, the sign is positive, otherwise it is negative divided by MAD. And because MAD by itself is 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 summation of actual forecast with absolute value signs divided by the number of periods. Therefore, MAD is always positive, but the numerator of tracking signal can be positive or negative. So tracking signal by itself could be positive, could be negative.

Here I have a set of data, and here are my forecasts. To come up with these forecasts, I have simply assumed that forecast for period 2 is equal to actual for period 1, and then in each of the future periods, I just have added forecast and actual and divided by 2. For example, I have 685 by adding the actual of the previous period to the forecast of the previous period and divided by. But that is not our concern right now. Suppose using another approach I have come up with some forecasts; therefore, here are my forecasts and here are my actuals. These are the absolute value between actual and forecast. 720, 685, the absolute value of the difference is 35 and so. Then I have added these two numbers, and I have added all these numbers, and I have put it here. Here I have divided 10 by 1, and I have 1 as a MAD in the first period. The summation of the absolute value of the difference between actual and forecast for the second period is 245 divided by 2, 22.5. Summation of the absolute value of the difference between actual and forecast for all 8 periods for period 2 to period 9 is 261.5. I have divided by 8, and I have 32.7. This is MAD for the last period. This is MAD for the first period, which is indeed by period 2. Here I have done the same computation, but I have not removed the sign, so the difference between actual and forecast is minus 10. And then I have found the difference between actual and forecast for all periods. This 5.6, I have found it by adding all these numbers. This negative 26.8 I have found it by adding all these numbers. If I divide 5.6 by MAD of that period, which is 32.4, I will get 0.2. If I divide negative 26.8 by MAD of this period, I will get negative 0.8. What do you expect tracking signal to be? What value? 1,000? 5,000? 100? Negative 100? What? Positive or negative? The expected value of tracking signal is sometimes actuals are greater than forecast, sometimes forecasts are greater than actual. Therefore, the expected value of a set of random positives and negatives is 0. 0 divided by a positive number is 0. This is my tracking signal for period one, for period two, and period three, and so on. We have an upper control to limit and a lower control to limit. Tracking should have two characteristics, first, it should remain between upper control limits and lower control limits. Second, it shouldn’t show any specific pattern. When we look at tracking signal over time, it should completely look random with no specific pattern. So this tracking signal is okay. This forecasting technique is reliable because the tracking signal is between the upper and lower control limits and because I cannot extract any pattern. If tracking signal is out of upper control limit or lower control limit, then the forecasting technique is not reliable. However, if it is within control limits, it may not be reliable. For example, in this case, tracking signal is positive and remains positive; it is always greater than zero. Tracking signal is the difference between actual and forecast, so we divide it by positive number. So this is always positive. If tracking signal is always positive, that means actual is always greater than forecast like this case. Summation of the differences between actual and forecast is always positive. Actual is always greater than forecast. We are underestimating the demand in a case like this. We clearly see a pattern, a seasonality pattern. In some periods actuals are greater than forecasts. In some periods forecasts are greater than actual, so either there is a seasonality in the real world, in our actual demand, which we have not incorporated into our forecasting technique, or we have incorporated the seasonality into our forecasting technique, which doesn’t exist in reality, is not a systematic component of our demand. Seasonality doesn’t need the actual season. You can also have seasons of less than three months. For example, a demand for movie theaters is seasonal. We have high season and Friday, Saturday, and Sunday and low season in other times. Demand for public transportation is seasonal. It is high season, for example, between 6:30 and 9:30 in the morning and 3:30 to 6:30 in the afternoon and low season in other times.

So we talked about MAD and tracking signal. To make sure that our forecasting technique has good quality and that quality retains itself over time, it is reliable, MAD is mainly served for two purposes. One of the reasons is to select one technique over several forecasting techniques and to test standard deviation of forecasts. Tracking signal is mainly used for two purposes again, to make sure that the summation of the differences between actual and forecast divided by MAD remains within upper control limit and lower control limit, and upper control limit and lower control limit in reality is set to +5 and -5, so it wants to make sure that tracking signal remains between +5 and -5. And the second main purpose of tracking signal is to make sure that while the tracking signal is within upper and lower control limits, it also doesn’t show any systematic component. It is completely random. I think this is enough for our first introductory session on forecasting, and I really encourage you to solve the assignments. Thank you for your patience.