Forecasting in supply chains

Role of demand forecasting

Effective transportation system or supply chain design is predicated on the availability of accurate inputs to the modeling process. One of the most important inputs are the **demands** placed on the system. Forecasting techniques are used to predict, in the face of uncertainty, what the demands on the system will be in the future so that appropriate designs and operating plans can be devised.

What else might we forecast?

Demands are not the only uncertain parameters that may require forecasting when building decision support models for supply chain and transportation systems. Operational parameters on the *supply side* may be just as important to forecast. For example, suppose one would like to build a model for routing trucks between various customer locations. The objective of the model is to determine minimum travel time tours. Clearly, travel time parameters between customer locations and depots are required. Forecasting techniques can be used.

Ideas to remember

- Forecasts are usually wrong. (Do not treat as known information!).
- Forecasts must include analysis of their potential errors.
- The longer the forecast horizon, the less accurate the forecast will generally be.
- Benefits can often be obtained from transforming a forecasted quantity into a known quantity (when this can be done, it comes at a cost however).
- Systems that are agile in responding to change are less dependent on accuracy in forecasts.

Forecasting Methodology Tree

Reference: Armstrong, J.S. *Long-range Forecasting*, Second Edition, 1985.

Forecasting time series data

Frequently, problems in forecasting for logistics systems require the analysis of univariate time series data; often we are interested in the evolution of customer demand for a single product over time, and what the future demand will be for that product. If demand for that product is independent of demand for other products, univariate techniques can be utilized. (Recall: univariate means of a single variable, like demand). Suppose that the observed demand values are given by $X_t, t = 1, 2, ...$. Our goal is to develop a forecasted demand value Y_t for a number of future periods of interest.

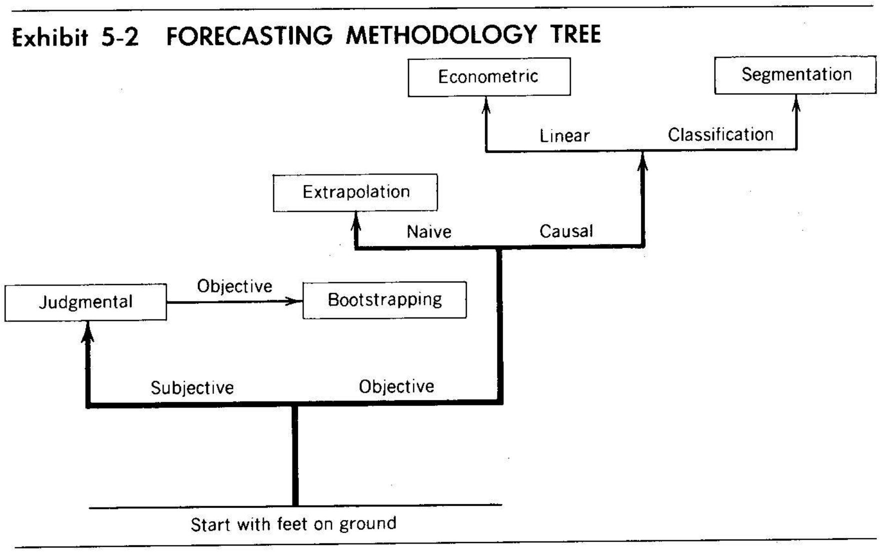

Figure 1: Forecasting methodology tree.

Historical Data

Assume that we have collected a number of historical observations X_t of a value we wish to forecast, for $t = 1, 2, ..., n$ periods up to and including the present. Further, suppose that these observations have been made at uniform intervals; for example, once a day, once a week, once a month, etc.

Forecast Notation

Suppose that we are currently in some period $\tau - 1$ and we wish to develop a forecast for the next period. We will use the notation Y_τ to represent that forecast. Additionally, we may also have to generate a multiple-period-ahead forecasts sometimes. In this case, we will use the notation $Y_T(\tau - 1)$ to represent the forecast for period T determined in period $\tau - 1$.

Evaluating quality

Any forecasting method can and should be evaluated by analyzing the errors produced by the forecasts; this is easily done once observations are made. Suppose that e_t are the errors over a set of n time periods:

$$e_t = Y_t - X_t \quad t = 1, 2, ...n$$

Errors should be compared against any assumptions made in the modeling process; for example, mean zero is common. In regression models, we might assume that the errors are normal; this should be verified.

Definition 1 (Unbiased Errors) *A set of errors is said to be unbiased if and only if:*

$$E[e_t] = 0$$

The errors also can give a measure of the quality of the forecast, but this may be confounded with the uncertainty in the quantity to be forecast. One way to do this is to calculate the MSE:

$$MSE = \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

Note of course that this value has units (something^2), and thus the performance of a method for forecasting demand in barrels of oil is not directly comparable to forecasting number of microprocessors. We will refer to the quantity $\sqrt{MSE}$ as the *standard error* of the forecast.

Another common measure of quality is the *mean absolute deviation* (MAD), which is useful since its units are (something) rather than (something^2):

$$MAD = \frac{1}{n}\sum_{i=1}^{n} |e_i|$$

Also useful is a relative performance error of a set of forecasts that is unitless. One such measure is the mean absolute percentage error (MAPE), which yields the average relative error of a set of forecasts:

$$MAPE = \left[\frac{1}{n}\sum_{i=1}^{n}\left|\frac{e_i}{X_i}\right|\right] \times 100$$

Error Normality Assumption and its Implications

In most situations, we will assume that the errors from a set of forecasts are independent and identically-distributed according to a normal random variable with mean 0 and variance $\sigma_E^2 = MSE$: $e_t \sim N(0, \sigma_E^2)$. This should make sense intuitively. The idea behind forecasting is to create a statistical model of the data to be forecast, and we intend that the errors of that statistical model exhibit "reasonable" behavior. Errors with mean zero indicate that our model is unbiased, and constant-variance errors without any correlation over time indicate that we have captured most of the signal of the data in our model.

Extrapolation forecasting for time series data

One of the simplest and most prevalent means for forecasting time series data is through the use of extrapolation-based techniques. Assume for now that for our time series data, we have collected a number of historical observations X_t, for $t = 1, 2, ..., n$ periods up to and including the present. Further, suppose that these observations have been made at uniform intervals; for example, once a day, once a week, once a month, etc.

Constant-pattern (stationary) time-series data

In some cases, you will be forecasting quantities that you expect to remain relatively constant over time. Since these quantities are still likely to fluctuate somewhat around a mean, and since that mean might move from one level to another over time, forecasting techniques have been developed for constant pattern data.

An appropriate statistical model for the observed quantity at time t might be:

$$X_t = \mu + \epsilon_t$$

where μ is some unknown mean, and ϵ_t are i.i.d. random variables with mean zero and variance σ^2. When it is reasonable to model data in this way, but we expect that μ might be varying slowly over time, it is appropriate to use *moving-average* or *exponential smoothing* methods.

m-period Moving Average

m-period moving averages are the simplest type of extrapolation forecast, and are useful for constant-pattern data. An m-period moving average is simply the arithmetic mean of the m most-recent observations. The forecast made in period $t-1$ for period t is thus:

$$Y_t = \frac{1}{m} \sum_{i=t-m}^{t-1} X_i$$

In moving average models, an implicit assumption is that the data have some constant mean. Thus, the multiple-step-ahead forecast $Y_T(t-1)$ is also equal to the single step-ahead forecast: $Y_T(t-1) = Y_t$ for all $T > t$.

Pros:

- Simple to calculate, implement
- Easy update procedure: $Y_{t+1} = Y_t + \frac{1}{m}(X_t - X_{t-m})$
- Outlier observations eliminated from consideration after m periods

Cons:

- If the data contains a trend, this method will always lag behind it.

Simple exponential smoothing

Another prevalent and popular forecasting method is exponential smoothing, in which the most recent forecast is combined with the most recent observation to generate a new forecast for a stationary series:

$$Y_t = \alpha X_{t-1} + (1-\alpha)Y_{t-1}$$

$$Y_t = Y_{t-1} + \alpha(X_{t-1} - Y_{t-1}) = Y_{t-1} - \alpha e_{t-1}$$

where $\alpha \in [0,1]$, is a smoothing constant. The idea behind exponential smoothing is that your previous forecast was probably pretty good, but now you have one new piece of observed data that you can use to update your forecast. The smoothing constant will control this update. Large values of α will place more weight on the most recent observation (a *reactive* step), and less weight on the previous forecast. Of course, if $\alpha = 1$, your new forecast is simply the observed value of demand in the previous period!

Exponential smoothing methods with $\alpha < 1$ essentially "use" all previous periods of observed demand in generating the forecast. Since the specification equation in this case defines a recursion, we can expand it:

$$Y_t = \alpha X_{t-1} + (1-\alpha)\alpha X_{t-2} + (1-\alpha)^2 Y_{t-2}$$

$$Y_t = \sum_{i=0}^{\infty}(1-\alpha)^i \alpha X_{t-i-1} = \sum_{i=0}^{\infty} a_i X_{t-i-1}$$

Hence, each previous demand observation is included to produce the newest forecast, but the weights given to each of these observations decreases exponentially (hence, the name). $\alpha > \alpha(1-\alpha) > \alpha(1-\alpha)^2 > ...$

Pros:

- Simple to calculate, implement
- Easy update procedure
- Infinite degree of control over data age using α

Cons:

- If the data contains a trend, this method will always lag behind it.
- Outliers will always have some impact

Exponential smoothing methods will still lag the trend, but less so since the average age of the data used in the forecast can be made younger with an appropriate choice of α.

An exponential smoothing example:

Example: Weekly orders of Atlanta Gateway CountryStore for a computer model: 200, 250, 175, 186, 225, 285, 305, 190. 3-period moving averages in periods 4,5,6,7,8: 208, 204, 195, 232, 272. Errors: 22, -21, -90, -73, 82.

Now, let's try ES with $\alpha = 0.3$: Suppose that the initial forecast is right on the money, $Y_1 = 200$. Forecast for period 2 is now (0.3)(200) + (0.7)(200) = 200. Period 3 is (0.3)(250) + (0.7)(200) = 215. Period 4: (0.3)(175) + (0.7)(215) = 203. 5,6,7,8 are 198, 206, 230, 252.

Age of Data Comparison Between MA and ES

One way to compare MA and ES is to consider the average "age" of the data included in the forecast. The average data age for an m-period moving averages can be calculated as:

$$\frac{1+2+...+m}{m} = \frac{1}{m}\frac{m(m+1)}{2} = \frac{m+1}{2}$$

where the data from period $t-1$ is 1 period old, from $t-2$ is two periods old, etc. The average age of the data in exponential smoothing models can be calculated by assuming we have an infinite history of historical observations. Note from the recursive formula developed earlier, the weight given to data that is i periods old is $\alpha(1-\alpha)^{i-1}$. Therefore, we can calculate the average age as follows:

$$\sum_{i=1}^{\infty} i\alpha(1-\alpha)^{i-1} = \frac{1}{\alpha}$$

Thus, exponential smoothing and moving average methods use data with the same average age when:

$$\frac{m+1}{2} = \frac{1}{\alpha}$$

Suppose we would like to "tune" our smoothing constant so that an exponential smoothing method used data with the same average age as an m-period moving average. We would simply let $\alpha = \frac{2}{m+1}$ to accomplish this. For example, ES(0.4) is equivalent from a data age point-of-view to a 4-period moving average, MA(4).

Forecasting Data with a Trend: Double-Exponential Smoothing (Holt's Method)

As described earlier, simple exponential smoothing and moving averages are suitable for forecasting data that is varying around a constant mean. If the mean changes every once in a while, these methods will also do a good job of forecasting around the new mean once they "catch up" to it. However, if the demand data to be forecast includes a positive or negative growth trend, MA and ES methods will tend to produce forecasts that always lag behind the trend, thus giving poor forecast errors. Fortunately, exponential smoothing methods have been developed to specifically deal with data with trends.

An extension of exponential smoothing, known as Holt's method, is designed for data with trends. In Holt's method, we will calculate two values over time, and use them to create our forecast. The first, R_t, is an estimate of the expected *level* of the series in period t. The second, G_t, is an estimate of the trend (or slope) in period t.

Given these two values, the standard next period forecast is given by:

$$Y_t = R_{t-1} + G_{t-1}$$

The idea here is to add our best estimate of the slope to our best estimate of the level (at period $t-1$) to create the forecast. The slope estimate represents the change in demand we expect from one period to the next; if we imagine that our trend is constant for the next τ time periods, the multiple-step-ahead forecast for period $t+\tau$ is given by:

$$Y_{t+\tau}(t-1) = R_{t-1} + (1+\tau)G_{t-1}$$

Let's now consider how to update our estimates R_t and G_t each period after the the data X_t becomes known. We will now employ two smoothing constants, $\alpha \in [0,1]$ for the level and $\beta \in [0,1]$ for the slope. After observing the demand X_t, we first update the estimate of the level by combining X_t with the previous period's forecast of demand:

$$R_t = \alpha X_t + (1-\alpha)(R_{t-1} + G_{t-1}) = \alpha X_t + (1-\alpha)Y_t$$

Next, we re-estimate the slope by combining a new slope estimate given by the difference between the current level and the previous level with our previous slope estimate.

$$G_t = \beta(R_t - R_{t-1}) + (1-\beta)G_{t-1}$$

Note in this case that we do **not** use $(X_t - X_{t-1})$ as the new slope estimate, since the smoothed value $(R_t - R_{t-1})$ is likely a better estimate of the true slope. (Based on what you know about random fluctuations, you should think about this claim.)

To use Holt's method, it is necessary to choose values for the two parameters α and β, and to choose initial values for R_1 and G_1 (assuming that the first period for which data is available is period 1). In contrast, to initialize simple exponential smoothing, it is necessary only to choose values for α and Y_1. A reasonable choice for R_1 might be X_1, and a reasonable choice for G_1 might be $\frac{X_N - X_1}{N}$, where N is the latest period for which we have already observed data, or perhaps is another earlier period selected so that the trend in $[1, N]$ is representative and constant.

Equivalence of Holt's Method to Simple Exponential Smoothing when $\beta = 0$ and $G_1 = 0$

Holt's method reduces to simple exponential smoothing when the parameter β is set to zero and the initial slope $G_1 = 0$. This should be clear; in each period, G_t will now be zero. Therefore, $Y_t = S_{t-1} = \alpha X_{t-1} + (1 - \alpha)Y_{t-1}$, which is the simple exponential smoothing model.

Simple linear regression methods

When we suspect trend in a time-series, we might postulate that the time series data fits a linear model, as follows:

$$X_t = a + bt + \epsilon_t$$

where a and b are unknown constants, and ϵ_t are the error random variables that we assume are i.i.d. normal random variables with mean 0 and variance σ^2. Given, therefore, some estimates of $\widehat{a}$ and $\widehat{b}$, we can use this type of model for forecasting:

$$Y_t = \widehat{a} + \widehat{b}t$$

with the multiple step-ahead forecast given by:

$$Y_{t+\tau}(t-1) = \widehat{a} + \widehat{b}(t + \tau)$$

Given a time series, $X_i, i = 1, 2, ..., t-1$, we can create a prediction for the next period t by determining reasonable estimates of $\widehat{a}$ and $\widehat{b}$. The most logical way to do so is to use a set of m most recent observations of X_i, in fact as far back as you believe the current trend should hold.

It can be shown that the *least-squares* method can be employed with this data to develop good, unbiased estimates of a and b. In this method, we attempt to minimize the sum of the squared errors in our data. Let $\widehat{X}_i$ be the predicted value for observation i, $\widehat{X}_i = \widehat{a} + \widehat{b}i$. The goal then is to minimize:

$$g(\widehat{a}, \widehat{b}) = \sum_{i=t-m}^{t-1} (X_i - \widehat{X}_i)^2 = \sum_{i=t-m}^{t-1} [X_i - (\widehat{a} + \widehat{b}i)]^2$$

It turns out this is easy to do by simply setting the partial derivatives of this function with respect to $\widehat{a}$ and $\widehat{b}$ equal to zero and solving simultaneously. The details of this process are quite simple, and will be skipped in this section.

Forecasting with Seasonality

Often, time series data display behavior that is *seasonal*. Seasonality is defined to be the tendency of time-series data to exhibit behavior the repeats itself every q periods. We use the term *season* to represent the period of time before behavior begins to repeat itself; q is therefore the season length in periods.

For example, air travel is known to exhibit several types of seasonal behavior. For example, if we example passenger travel demand week after week, we may see that certain days of the week always exhibit relatively higher demands than others; for example, travel on Fridays, Sundays might be highest, followed by Mondays, followed by Thursdays, and finally Sats, Tues, Weds. In this case, the season length is 7 periods (days). In another example, suppose we look at monthly demand for PCs. This may also exhibit a seasonal pattern in which the months of August (back-to-school), October, November, and December (winter holidays) exhibit strong demand, while January and February exhibit especially weak demand. In this example, the season length is 12 periods (months).

Forecasting Seasonal Data with Constant Multiplicative Factors

In many cases, it may make sense to model seasonal behavior with a multiplicative factor. For example, perhaps a statement like "air travel demand is 20% greater than average on Fridays" is reasonable. Note that implicit in this statement is that the seasonal effect on demand is *relative*; if overall demand grows (due to a positive trend), then the multiplicative seasonal factor would predict 20% more demand of the new average.

If we do not believe that seasonal factors change over time, one simple method of dealing with seasonality is to forecast using a *deseasonalized* time series, $\overline{X}_i \quad i = 1, 2, ..., t-1$, defined as follows:

$$\overline{X}_i = \frac{X_i}{S_j} \quad j = i \mod q$$

where X_i is the original time series, and S_j is the seasonal factor associated with the j-th period in each season.

Note that some method must be developed to determine a set of seasonal factors $S_j \quad j = 1, ..., q$. Good practice also dictates that these factors have an average value of 1, therefore:

$$q = \sum_{j=1}^{q} S_j$$

To estimate a set of factors using multiple (at least 2) seasons of historical data, you should first calculate a deseasonalized level $\widehat{X}_t$ for each period. If the data contains no trend, the deseasonalized level might just be the average value over all periods. If the data contains a trend, you might fit a simple linear regression model to the data, and use it to generate predicted values for each period; the predicted value for t would be a good estimate for $\widehat{X}_t$. Each factor S_j can then be estimated by averaging over like periods for each season for which data is available:

$$S_j = \text{average}\left(\frac{X_j}{\widehat{X}_j}, \frac{X_{j+q}}{\widehat{X}_{j+q}}, \frac{X_{j+2q}}{\widehat{X}_{j+2q}}, ...\right)$$

Given a set of factors, we can forecast using moving averages, exponential smoothing, or Holt's method using the deasonalized series exactly as before. Then, suppose $\overline{Y}_t$ is the forecast generated for period t using the deasonalized series. Then, the true forecast for t is:

$$Y_t = \overline{Y}_t S_j \quad \text{where } j = t \mod q$$

Triple Exponential Smoothing with Multiplicative Seasonal Factors: Winters' Method

For some problems, we may wish to continually update our seasonal factors as we receive new data. Winters' Method uses a triple exponential smoothing approach to extend Holt's Method to additionally update seasonal factors.

In this method, we will calculate three values over time: an estimate of the deseasonalized level $\overline{R}_t$, an estimate of the trend $\overline{G}_t$, and an estimate of a multiplicative seasonal factor S_t. Suppose that the season length is q (for example, if a season is a week of days, then $q = 7$). For notation purposes, we will now explicitly record a seasonal factor S_t for each period t, rather than just for the first q seasonal periods.

Given these three values, our forecast for the next period is:

$$Y_t = (\overline{R}_{t-1} + \overline{G}_{t-1}) S_{t-q}$$

In this expression, we simply multiply the deseasonalized forecast (given by Holt's equation) by the seasonal factor to yield the new forecast. Note that we use the best estimate of the seasonal factor for this time period in the season, which was last updated q periods ago. To make multiple-step-ahead forecasts (for $\tau < q$), the following expression should be used:

$$Y_{t+\tau}(t-1) = (\overline{R}_{t-1} + (1+\tau)\overline{G}_{t-1}) S_{t+\tau-q}$$

Now let's look at how to update the level, trend, and seasonal factor each time period. First, the current deseasonalized level is updated as follows:

$$\overline{R}_t = \alpha \frac{X_t}{S_{t-q}} + (1-\alpha)(\overline{R}_{t-1} + \overline{G}_{t-1})$$

Note that this update is identical to Holt's, except that we use a deseasonalized demand observation in the first term, and our deseasonalized forecast in the second term.

Next, we update the trend:

$$\overline{G}_t = \beta(\overline{R}_t - \overline{R}_{t-1}) + (1-\beta)\overline{G}_{t-1}$$

Again, we blend a deseasonalized trend estimate with our previous trend estimate.

Finally, we update our estimate of the seasonal factor. Here we introduce a new smoothing constant $\gamma \in [0, 1]$, and create a combination of the most recently observed seasonal factor given by the demand X_t divided by the deasonalized series level estimate $\overline{R}_t$ and the previous best seasonal factor estimate for this time period:

$$S_t = \gamma \frac{X_t}{\overline{R}_t} + (1-\gamma) S_{t-q}$$

Since seasonal factors represent deviations above and below the average, the average of any q consecutive seasonal factors should always be 1. Thus, after estimating S_t, it is good practice (but not necessary) to renormalize the q most recent seasonal factors such that:

$$\sum_{i=t-q+1}^{t} S_i = q$$

Variations on Winters' Method

First, it should be clear that when $\gamma = 0$ and $q = 1$ and the first seasonal factor is S_1 is set equal to one, Winters' method is identical to Holt's method. However, it is sometimes useful to create a forecast that uses seasonality as a component but does not include a trend. This can also be easily modeled in Winters' method, by simply setting $\beta = 0$ and $\overline{G}_1 = 0$. The model now reduces to a regular exponential smoothing model, with multiplicative seasonal factors!

Advanced: Initializing Winters' Method with Multiplicative Factors

Unlike simple exponential smoothing and Holt's method, Winters' method requires more data for initialization. As a general rule of thumb, a minimum of two full seasons (or $2q$ periods) of historical data is needed to initialize a set of seasonal factors. Suppose the current time period is period 0, and we have two seasons of historical data (if you want to initialize with more than two seasons of historical data, a similar process can be easily developed). Let the historical data be given by $X_{-2q+1}, X_{-2q+2}, ..., X_0$. Here are the steps to initialize Winters' for period 0:

1. Calculate the averages for the two separate seasons of historical data:

$$V_1 = \frac{1}{q} \sum_{i=-2q+1}^{-q} X_i$$

$$V_2 = \frac{1}{q} \sum_{i=-q+1}^{0} X_i$$

2. Define the initial slope estimate $\overline{G}_0$ to be $\frac{V_2 - V_1}{q}$. This is the slope of the line generated by placing V_1 at the center of the first season's periods, and V_2 at the center of the second season's periods, and connecting the two points.

3. Let $\overline{R}_0$, the estimated series level in period 0, be given as $V_2 + \overline{G}_0 \frac{q-1}{2}$.

4. (a) Given $\overline{R}_0$ and $\overline{G}_0$ now, the estimated series level at any time t during the two previous seasons is given by $\overline{R}_0 + t\overline{G}_0$, $t = -2q+1, -2q+2, ..., 0$ (e.g.,

for period -1 the value would be $\overline{R}_0 - \overline{G}_0$, period -2 $\overline{R}_0 - 2\overline{G}_0$, etc.). The initial seasonal factors are generated by taking the observed demand in t and dividing by the estimated level:

$$S_t = \frac{X_t}{\overline{R}_0 + t\overline{G}_0} \quad \forall t = -2q+1, -2q+2, ..., 0$$

(b) Now we average the factors for the two corresponding periods in each season:

$$S_{-q+1} = \frac{S_{-2q+1} + S_{-q+1}}{2}, ..., I_0 = \frac{I_{-q} + I_0}{2}$$

(c) Finally, we normalize the seasonal factors:

$$S_t = q\left[\frac{S_t}{\sum_{i=-q+1}^{0} S_i}\right] \quad \forall t = -q+1, -q+2, ..., 0$$

Causal Forecasting with Linear Regression Models

The time series extrapolation methods that we've discussed up to this point are all reasonable forecasting models, and they can be used to effectively model stationary, trend, and seasonal data. However, there are often additional factors that may play an important role in the demand that will be seen in any given period, and we would like a forecasting methodology that can capture these factors. Typically, we have a number of *seasonal* factors that need to be incorporated. For example, daily demand for trucking services may have day of week and week of month effects which could be captured with multiple seasonal factors. Also, they may have day of year effects that reflect certain holidays; these may not repeat in a strictly seasonal way since the number of days between holidays changes.

Linear regression models provide a useful framework for developing causal demand forecasts that allow us to capture multiple seasonal effects as well as other causal effects. As you already know, a linear regression model is a statistical model that assumes that a value to be estimated (such as a forecast) can be decomposed into a weighted sum of a number of causal *factors*, plus a random error term:

$$
\begin{aligned}
Y_t &= A_0 + A_1 f_{1,t} + A_2 f_{2,t} + ... + A_n f_{n,t} + \epsilon_t \\
&= A_0 + \sum_{j=1}^{n} A_i f_{j,t} + \epsilon_t
\end{aligned}
$$

In this expression, the values $A_1, ..., A_n$ are the values of the weights (coefficients) given to specific factors f, and ϵ_t is the random error term, while A_0 is an additive constant. It will be assumed that the error in each time period is independently and identically distributed according to a normal random variable with mean 0 and some variance σ_E^2. Each factor $f_{j,t}$ represents a piece of known data that is available in period t that might be useful to create a forecast.

Estimating Coefficients in Regression Models

Before we discuss some factors that might be useful in demand forecasting models, we should first review how to estimate the coefficients of a regression model. Here, unlike the case with the extrapolation models presented previously, we can explicitly use the forecast errors during a historical calibration period to estimate values for the coefficients that give the best "fit" to the data. In essence, we are using an analytical process to "tune" the parameters of the model such that our forecasts over the historical calibration period have as small errors as possible.

You should recall from your statistics courses that the *least-squares* method can be employed estimates for the values A. In least-squares, we use a calculus methodology to minimize the sum of the squared errors over our historical calibration period. For each period, the actual error value e_t is given by $e_t = Y_t - X_t$, where X_t is the observed value of demand. Thus, the squared error in t is $e_t^2 = (Y_t - X_t)^2$. Plugging in our formula for

Y_t yields:

$$e_t^2 = \left(A_0 + \sum_{j=1}^{n} A_j f_{j,t} - X_t \right)^2$$

Now suppose we have N periods of historical data for use in estimation, and suppose for simplicity that these periods have numbers $1, ..., N$. To determine the best estimates for the factors A, we can look to minimize the sum of squared errors over this historical data:

$$\min_A SSE = \min_A \sum_{t=1}^{N} e_t^2 = \min_A \sum_{t=1}^{N} \left(A_0 + \sum_{j=1}^{n} A_j f_{j,t} - X_t \right)^2$$

Without developing the mathematics formally, it is a simple exercise in calculus and linear algebra to determine the values of $A_0, A_1, ..., A_n$ that minimize this expression. The details will be skipped here; you can consult your statistics notes for more information. The important point is that the least-squares estimates for the coefficients are easy to determine, and they stem from minimizing the sum of the squared errors observed over a calibration period.

Choice of Factors in Regression Models

Given that regression might be a useful technique for developing a forecasting model, what might be some useful factors to include?

1. *Model Seasonal Effects with Indicator Variables*: Seasonal effects are best modeled using indicator variables. For example, suppose you think the month of the year has an important effect on demand. For each month that you think is significantly different from average, you could introduce an indicator variable that is equal to one if the period is in that month, and 0 if not. Note: if you use indicator variables in this way, and you plan on using an indicator for each month, do not use an indicator for the last month since it is not necessary and may confound the estimation procedure (because of linear dependence). For example, if you had an indicator variable for all months of the year except December, it is clear that if a period had zeros for all indicators, it would have to be in December!

2. *Model Other Important Exceptional Periods with Indicator Variables*: Sometimes exceptional periods occur from a forecasting point of view that are not necessarily seasonal, but still have an effect on demand. Model such effects with indicator variables. For example, perhaps demand for accounting services spikes the two weeks prior to April 15 each year. Or perhaps passenger vehicle-miles traveled spikes on holiday weekends.

3. *Capturing Moving Average Effects*: In regression models, you may think that past period demands should still be factored into forecasts with a proper coefficient.

This is simple to do with a regression model. However, when doing so, you must be careful to verify that the errors are not serially-correlated since they must remain independent over time.

4. *Capturing Trend*: Trends in data over time can be captured by including a factor t.

Evaluating Regression Model Quality

Again, this is not a course on statistics but I want you to think back to what you've learned regarding linear regression models. Recall that the R^2 value (the ratio of the sum of squares of the regression terms to the total sum of squares of the data) of a regression model gives you an idea of the quality of a regression model; as R^2 approaches 1, more and more of the total variation in the model is explained by the factors and less is explained by simple random noise. Thus, larger values of R^2 tend to indicate better "fit." However, remember that using more predictive factors will always increase the value of R^2. It is equally important to check the statistical significance of the estimated coefficients; they should each have t values that indicate they have a high probability that they are significantly different from zero. Finally, you should ensure that there is no inherent bias in your regression forecasting model by analyzing your errors (residuals). For example, you should verify that the errors appear to be approximately normally distributed (using a normal probability plot or a statistical test). Also, you should look for errors that are homoskedastic and do not exhibit any curvilinear patterns when plotted against (1) time, (2) predicted values, and (3) values of the factors. Homoskedastic errors exhibit a constant variance or spread when plotted against a dependent variable.