Supply Chain Fundamentals: Forecasting

Notes from “Supply Chain & Logistics Fundamentals”

CTL.SC1X (offered by MITx)

weeks 2-4

Contents

[High-Level Context 2](#_Toc445126771)

[Fundamental Forecasting Approaches 3](#_Toc445126772)

[Forecast Measurements 3](#_Toc445126773)

[Forecasting Level Data 5](#_Toc445126774)

[Cumulative Forecast & Naïve Forecast 5](#_Toc445126775)

[Moving Averages 5](#_Toc445126776)

[Simple Exponential Smoothing 6](#_Toc445126777)

[Forecasting Data with a Level and Trend 6](#_Toc445126778)

[Double Exponential Smoothing (Holt Model) 7](#_Toc445126779)

[Damped Trends 7](#_Toc445126780)

[Forecasting Data with a Level, Trend, and Seasonal Component 8](#_Toc445126781)

[Triple Exponential Smoothing (Holt-Winter Method) 8](#_Toc445126782)

[Suggested Smoothing Coefficients 9](#_Toc445126783)

[Additional Forecasting Methods 9](#_Toc445126784)

[Causal Forecasting 9](#_Toc445126785)

[Intermittent Demand 9](#_Toc445126786)

[Bass Diffusion Model 9](#_Toc445126787)

# High-Level Context

There are three types of questions typically associated with forecasting:

|  |  |
| --- | --- |
| Question | Forecast Types |
| What should we do to shape and create demand for our product? | Demand Planning* Product & Packaging
* Promotions
* Pricing
* Place
 |
| What should we expect demand to be given the demand plan in place? | Demand Forecasting* Strategic, Tactical, Operational
* Considers internal & external factors
* Baseline, unbiased, & unconstrained
 |
| How do we prepare for and act on demand when it materializes? | Demand Management* Balances demand & supply
* Sales & operations planning (S&OP)
* Bridges both sides of a firm
 |

Forecasting can also be done at three different time horizons:

|  |  |  |
| --- | --- | --- |
| Level | Horizon | Purposes |
| Strategic | Year/Years | * Business planning
* Capacity planning
* Investment strategies
 |
| Tactical | Quarterly | * Brand plans
* Budgeting
* Sales Planning
* Manpower Planning
 |
| Months/Weeks | * Short-term capacity planning
* Master planning
* Inventory Planning
 |
| Operational | Days/Hours | * Transportation planning
* Production planning
* Inventory deployment
 |

Forecasting Truisms:

|  |  |
| --- | --- |
| Truism: | Suggestion |
| 1. Forecasts are always wrong | Use range forecasts and track forecast error |
| 2. Aggregated forecasts are more accurate | Risk-pooling reduces coefficient of variation |
| 3. Shorter horizon forecasts are more accurate | Postpone customization until as late as possible |

# Fundamental Forecasting Approaches

|  |  |
| --- | --- |
| Subjective | Objective |
| Judgmental:* Sales force survey
* Jury of experts
* Delphi techniques
 | Causal / Relational* Econometric models
* Leading indicators
* Input-Output models
 |
| Experimental:* Customer surveys
* Focus group sessions
* Test marketing
 | Time Series* “Black Box” approach
* Past predicts the future
* Identify patterns
 |

Oftentimes, you will need to use a combination of approaches. The subjective methods are most useful when data is not available (example: new products) or when the environment is changing and we do not expect the past to be a good predictor of the future. When data is available, the objective methods are used to mine through that data and identify patterns and causal factors. This class focuses mainly on the mathematics of Time Series forecasting.

# Forecast Measurements

Forecast measurements measure two things:

1. Accuracy – how close are the forecasts to the actual values?
2. Bias – are we typically over- or under-forecasting actual values?

Forecast error is defined as:

$e\_{t}=A\_{t}-F\_{t}$ Forecast Error

$A\_{t}$ = actual value (at time *t*)

$F\_{t}$ = forecasted value (at time *t*)

Note that positive error indicates actuals are higher than the forecast and negative error indicates that actuals are lower than the forecast. The equation above is sometimes calculated in the reverse manner (forecast minus actuals, rather than actuals minus forecast), so it’s important to be clear about which method is being used. The method used above is more common in academic texts and is used in this course. It follows naturally when we define a model as:

$$X\_{t}=\hat{X}\_{t}+e\_{t}$$

That is, the actual values $X\_{t}$ are represented by some estimate $\hat{X}\_{t}$ plus an error term. In more advanced techniques, the distribution of error terms and the changes in this distribution over time may also be modeled.

The following measures are commonly used to measure forecasts:

|  |  |  |  |
| --- | --- | --- | --- |
| Mean Deviation(MD) | $$MD=\frac{1}{T}\sum\_{t=1}^{T}e\_{t}$$ | Mean Absolute Deviation(MAD) | $$MAD=\frac{1}{T}\sum\_{t=1}^{T}\left|e\_{t}\right|$$ |
| Mean Squared Error (MSE) | $$MSE=\frac{1}{T}\sum\_{t=1}^{T}e\_{t}^{2}$$ | Root Mean Squared Error (RMSE) | $$RMSE=\sqrt{\frac{1}{T}\sum\_{t=1}^{T}e\_{t}^{2}}$$ |
| Mean Percent Error(MPE) | $$MPE=\frac{1}{T}\sum\_{t=1}^{T}\frac{e\_{t}}{A\_{t}}$$ | Mean Absolute Percent Error (MAPE) | $$MAPE=\frac{1}{T}\sum\_{t=1}^{T}\frac{\left|e\_{t}\right|}{A\_{t}}$$ |

The equations above have errors subscripted by *t*, which implies that we are measuring forecasts for one item at different points in time. We can also measure forecasts at one point in time across multiple items using these measures. In that case it would make more sense to subscript by *n* to indicate each unique item. Also note:

* MD and MAD are measured in units. MD measures bias, and MAD measures forecast accuracy. These metrics may be useful for one item, but they are difficult to interpret when comparing multiple items at different sales levels.
* MSE and RMSE both measure accuracy. RMSE is more commonly used as its dimension is units. These are similar to the variance and the standard deviation (respectively) of the forecast errors. The only different is that instead of being centered on the mean error they are centered at zero.
* MPE and MAPE convert errors into percentages. Like MD and MAD they measure bias and accuracy (respectively), but in a way that allows easier comparisons across items with different sales levels.

There are three other measures I find useful which were not discussed in class: MAD/MEAN, WMPE, and WMAPE (or WAPE). MAD/MEAN just takes the MAD score and divides by the mean. This may be useful for single items and produces results similar to the MAPE (but more stable since we use the same denominator across all points in time). WMPE and WMAPE weight the percent errors by sales so that fast-moving items are more important. They also simplify very nicely so that we need only calculate the sum (or the average) of errors, absolute errors, and sales.

|  |  |
| --- | --- |
| Weighted Mean Percent Error(WMPE) | $$WMPE=\frac{1}{\sum\_{n}^{}A\_{n}}\sum\_{n=1}^{N}A\_{n}\left(\frac{e\_{n}}{A\_{n}}\right)=\frac{\sum\_{n}^{}e\_{n}}{\sum\_{n}^{}A\_{n}}=\frac{Sum of Error}{Sum of Sales}$$ |
| Weighted Mean Absolute Percent Error(WMAPE or WAPE) | $$WMAPE=\frac{1}{\sum\_{n}^{}A\_{n}}\sum\_{n=1}^{N}A\_{n}\left(\frac{\left|e\_{n}\right|}{A\_{n}}\right)=\frac{\sum\_{n}^{}\left|e\_{n}\right|}{\sum\_{n}^{}A\_{n}}=\frac{Sum of Abs. Error}{Sum of Sales}$$ |

# Forecasting Level Data

## Cumulative Forecast & Naïve Forecast

If we assume that a forecast is level (that it has no trend or seasonal components), all that we need to do to produce a good forecast is determine that level. All of these models take the form:

$$x\_{t}=a+e\_{t}$$

where *a* indicates the level of the forecast and:

$$e\_{t} \~ N(μ=0,σ^{2}=V\left[e\right])$$

The two simplest ways to do this are the Cumulative Forecast (simply average all of the data), and the Naïve Forecast (just predict the last value).

|  |  |
| --- | --- |
| Cumulative Forecast | Naïve Forecast |
| Average all observations (from 1 to *t*):$$\hat{x}\_{t,t+1}=\hat{a}\_{t}=\frac{1}{t}\sum\_{i=1}^{t}x\_{i}$$ | Use the last observation as the next forecast:$$\hat{x}\_{t,t+1}=\hat{a}\_{t}=x\_{t}$$ |

The Cumulative Forecast uses all data points and is suitable if the level never changes. The Naïve Forecast is more robust if the level changes, as it will change immediately with it. However, the Naïve forecast will also be more volatile or “jumpy” whereas the cumulative forecast will be more steady over time.

## Moving Averages

Moving averages combine the good properties of the cumulative and naïve forecasts. They average data so that they are more steady and less volatile over time. They also can respond to level changes (although it may take a few observations before they completely adjust). The simplest way to do this is to use a moving average:

|  |
| --- |
| Moving Average |
| Take the average of the last *N* observations$$\hat{x}\_{t,t+1}=\hat{a}\_{t}=\frac{1}{N}\sum\_{i=t+1-N}^{N}x\_{i}$$ |

*N* can be set as needed. Note that if $N=1 $this produces the naïve forecast and if $N=t$ this produces the cumulative forecast.

## Simple Exponential Smoothing

An alternate approach is to use exponential smoothing. This differs from a moving average in that a moving average places equal value on each observation. The exponential smoothing uses weights that place less value on older values. This implies that older data is less relevant to today’s forecast – which typically makes sense. The interesting thing about exponential smoothing is that it never really throws away old data though. It just weights it less heavily than more recent data.

|  |
| --- |
| Simple Exponential Smoothing |
| Take a weighted average of the last observation $x\_{t}$ and the previous estimate of the level $\hat{a}\_{t-1}$ using weights $α$ and $(1- α)$. $α$ is the blending coefficient and determines how quickly new data is incorporated into the model.$$\hat{x}\_{t,t+1}=\hat{a}\_{t}$$$$\hat{a}\_{t}=αx\_{t}+(1-α)\hat{a}\_{t-1}$$$0\leq α\leq 1$ Smoothing coefficient. (In practice, use: $0.1\leq α\leq 0.3$) |

This model can be initialized with a single value:

$$\hat{a}\_{1}=x\_{1}$$

Alternate methods may also be used. For example, we could use the simple average of the first few values if we want a more stable starting point. Over time, the initial value will become less and less relevant as new data is added.

This method is called exponential smoothing because it is equivalent to taking a weighted average of all available data using exponentially decaying weights (after sufficient data is available):

$$\hat{a}\_{t}=\sum\_{i=0}^{t-1}x\_{t-i}α\left(1-α\right)^{i}$$

# Forecasting Data with a Level and Trend

If a time series has both level and trend, we assume it has the form:

$$x\_{t}=\left(a+bt\right)+e\_{t}$$

*a* = level, b = trend

$$e\_{t} \~ N(μ=0,σ^{2}=V\left[e\right])$$

We could build a model simply by applying a linear regression to our data over time, but if the level of trend change over time this would not work. Again, this is where exponential smoothing can be used to build a model where both level and trend can adjust over time.

## Double Exponential Smoothing (Holt Model)

Double exponential smoothing uses an estimate of the current level and trend to produce a forecast of the form:

$$\hat{x}\_{t,t+τ}=\hat{a}\_{t}+τ\hat{b}\_{t}$$

The method for updating both level and trend is given below:

|  |  |
| --- | --- |
| Updating Equation | Interpretation |
| $$\hat{a}\_{t}=αx\_{t}+(1-α)\left(\hat{a}\_{t-1}+\hat{b}\_{t-1}\right)$$ | The new level is an average of:1. The last observation
2. The expected level, taking into account the trend
 |
| $$\hat{b}\_{t}=β\left(\hat{a}\_{t}-\hat{a}\_{t-1}\right)+(1-β)\hat{b}\_{t-1}$$ | The new trend is an average of:1. The difference between the current and last level
2. The old trend
 |

Note that these equations should be applied in order. The new estimate of the level is used in the updating equation for the level.

## Damped Trends

Trend models should be used with caution because they will produce trends that continue indefinitely. If sales are increasing this year, they will continue to increase forever. Or if sales are decreasing, we may lower the forecast until we are forecasting negative quantities. A damped trend model overcomes this by forecasting:

|  |  |
| --- | --- |
| $$\hat{x}\_{t,t+τ}=\hat{a}\_{t}+\sum\_{i=1}^{τ}ϕ^{i} \hat{b}\_{t}$$ | $$0\leq ϕ\leq 1$$ |

If $ϕ=1$ this is a regular trend model. With values less than 1, the trend will be scaled down as we forecast further out, eventually removing the trend altogether.

The damped trend changes the updating equations to:

|  |
| --- |
| $$\hat{a}\_{t}=αx\_{t}+(1-α)\left(\hat{a}\_{t-1}+ϕ\hat{b}\_{t-1}\right)$$ |
| $$\hat{b}\_{t}=β\left(\hat{a}\_{t}-\hat{a}\_{t-1}\right)+(1-β)ϕ\hat{b}\_{t-1}$$ |

#  Forecasting Data with a Level, Trend, and Seasonal Component

## Triple Exponential Smoothing (Holt-Winter Method)

If the data has a level, trend, and seasonal component we need a forecast that takes all of these into account:

$$x\_{t}=\left(a+bt\right)F\_{t}+e\_{t}$$

*a* = level

*b* = trend

$F\_{t}$ = seasonal coefficient at time *t*

NOTE: This represents a multiplicative seasonal model (since we are multiplying our estimates underlying estimates by $F\_{t}$. We may also use an additive model where $F\_{t}$ is added to the underlying estimate, but this is less common.

Errors are again assumed to be normally distributed:

$$e\_{t} \~ N(μ=0,σ^{2}=V\left[e\right])$$

As we might expect, the updating method for both level, trend, and seasonal coefficients uses exponential smoothing:

|  |  |
| --- | --- |
| Updating Equation | Interpretation |
| $$\hat{a}\_{t}=α\left(\frac{x\_{t}}{\hat{F}\_{t-P}}\right)+(1-α)\left(\hat{a}\_{t-1}+\hat{b}\_{t-1}\right)$$ | The new level is an average of:1. The last observation (de-seasonalized)
2. The expected level, taking into account the trend
 |
| $$\hat{b}\_{t}=β\left(\hat{a}\_{t}-\hat{a}\_{t-1}\right)+(1-β)\hat{b}\_{t-1}$$ | The new trend is an average of:1. The difference between the current level and the new level
2. The old trend
 |
| $$\hat{F}\_{t}=γ\left(\frac{x\_{t}}{\hat{a}\_{t}}\right)+(1-γ)\hat{F}\_{t-P}$$(*P* is the period of the seasonal cycle) | The new seasonal coefficient is an average of:1. The coefficient we would estimate based on the last observation
2. The last coefficient
 |

Note that these equations should be applied in order. The new estimate of the level is used in the updating equation for both the level and the trend. Also, these equations can be modified to use a damped trend if we desire.

## Suggested Smoothing Coefficients

While the smoothing coefficients are allowed to range from 0 to 1, they are typically restricted to the following ranges:

Stationary Models

|  |  |  |
| --- | --- | --- |
| Level (α) | 0.01 to 0.30 | (0.1 reasonable) |

Level / Trend Models:

|  |  |  |
| --- | --- | --- |
| Level (α) | 0.02 to 0.51 | (0.19 reasonable) |
| Trend (β) | 0.005 to 0.176 | (0.053 reasonable) |
| Seasonality (γ) | 0.05 to 0.50 | (0.10 reasonable) |

# Additional Forecasting Methods

## Causal Forecasting

Causal and promotional forecasting was demonstrated using linear regression. Simple linear regression and multiple linear regression were both used, and an emphasis was placed on checking the R2 value and examining the residuals to see if the model makes sense.

## Intermittent Demand

## Bass Diffusion Model