Forecasting Model for Sporadic Distributor-Based Market

By Name(s): Ahmed Elazzamy and Stanley Park
Advisor: Dr.Tugba Efendigil

Topic Areas: Demand Planning, Forecasting, Supply Chain Planning

Summary:
In this research we investigate the benefits and validity of linking downstream distributor data in a fast-moving consumer goods company to improve forecast accuracy for intermittent demand. We used multi-tier regression analysis to link distributor sellout data to a retailer in order to predict shipment volume, and compared that against Croston method. We concluded that using multi-tier regression analysis has improved forecast accuracy in some cases; however, the success of this method is subject to data availability that could be a constraint in certain situations.

Prior to MIT, Ahmed worked at Oracle Dubai office for 6 years as Principal SC Solution. he received BSc in Computer Science, Misr International University 2004

Prior to MIT, Stanley worked at A.T. Kearney Seoul Office for 5.5 years. He received his B.S. Industrial Engineering, Korea University 2012

KEY INSIGHTS

1. The multi-tier regression analysis has a benefit of integrating the downstream distributor data, and has shown increase of forecast accuracy on SKUs level with an average of +20%.

2. Distributors are encouraged to log their inventory balance each month for better accuracy instead of calculating it backward from the current period.

3. Croston's method maintains a stable level across the time horizon which will positively impact the upstream manufacturing in terms of planning and setting inventory levels.

Introduction

Inaccurate demand forecasting impacts a company's business and operation in several ways. Generally, if the company forecasts demand higher than the actual demand, it will lead to higher inventory cost and failure in capacity management, causing the company's profitability to decrease. On the other hand, if the forecasts are lower than the actual demand, the situation will cause longer delivery time and more stockouts, which will eventually lead to missed revenue opportunities and, more severely, a loss of customers. In more extreme cases, inaccurate forecasting could even affect the survival of a company, as seen in the case of an aircraft manufacturer that went bankrupt because the company failed to manage the customer distress caused by overbookings. Therefore, many companies, regardless of the industry, have put effort into improving their demand forecasting. A well-known fast-moving consumer goods (FMCG) company and Fortune 500 company, which is the sponsor of the project, is also putting effort into improving its forecasting method through analyzing its shipment history and applying moving average to the forecast model.

The sponsor company is facing highly sporadic demand as shown in Figure 1. If we look at the actual shipment volume of a single SKU (Stock keeping unit), we could easily identify that sporadic demand occurs over time. The shipment volumes have lots of periods of zero demand, and even when there is a demand, it is highly fluctuating, ranging from 3,000 units to 8,000 units as shown in Figure 1.

Therefore, with the current forecasting model that the company utilizes, it was not able to accurately capture the sporadic demand and forecast the demand accurately.
Sporadic demand (also known as intermittent demand) is often characterized by two factors. First, the demand pattern shows several periods with zero demand. Second, when the demand occurs, it is highly variable. Therefore, the pattern usually occurs in a product when it is at the end of the cycle or where it is in the off-seasonal cycle. For example, the spare parts industry is well known for its sporadic demand. Spare parts for aviation and defense industry are kept in stock, but those parts usually have zero demand over a period of time because customers only order when they require maintenance or have to repair the products.

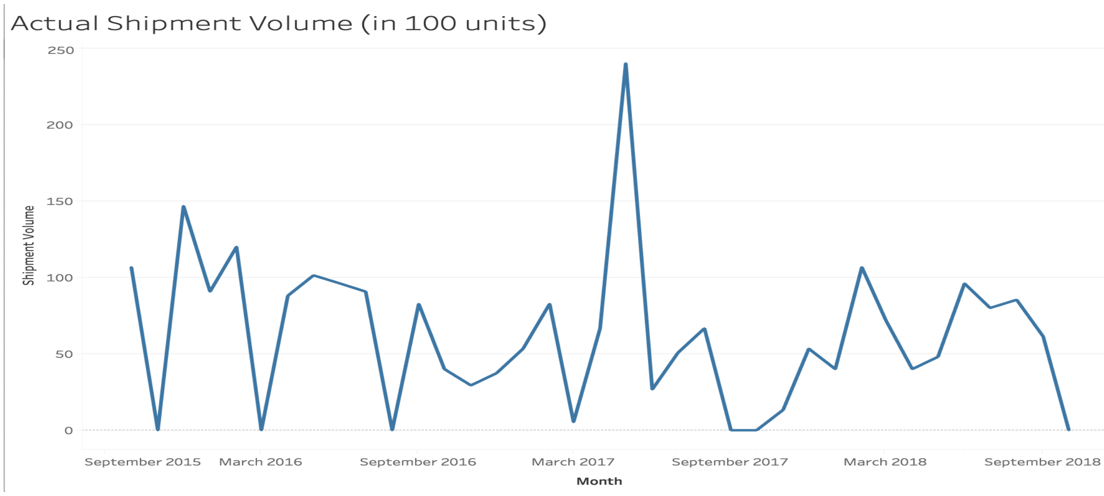


Figure1 Sporadic demand of the company. Data from the sponsor company

Methodology

As defined in the introduction we investigated the factors that drive the sporadic demand in order to improve the company's current forecasting method. We will link those identified drivers into the forecasting model by transitioning the causal relations into a mathematical model. This procedure uses information from downstream data such as point-of-sale transaction, product location in the store, and other information that shapes consumer demand. Then once demand is predicted another model is built in conjunction with some other variables to be able to predict the shipments. For example, if we want to build a model that predicts the number of shipments for a certain product, then we would first predict the demand, and use the predicted demand as an independent variable to predict number of shipments. The second approach is to directly forecast the sporadic demand by applying the intermittent forecasting method. These methods accuracy will be measured and evaluated through metrics such as A-MAPE (Mean Absolute Percent Error), root mean square error (RMSE), and mean deviation (MD).
Also, the categorization method proposed in literature are utilized since our forecasting methods include various SKUs with different behaviors. That is, we will categorize the SKUs into meaningful categories with similar characteristics and apply different forecasting methods to enhance accuracy and exclude bias.

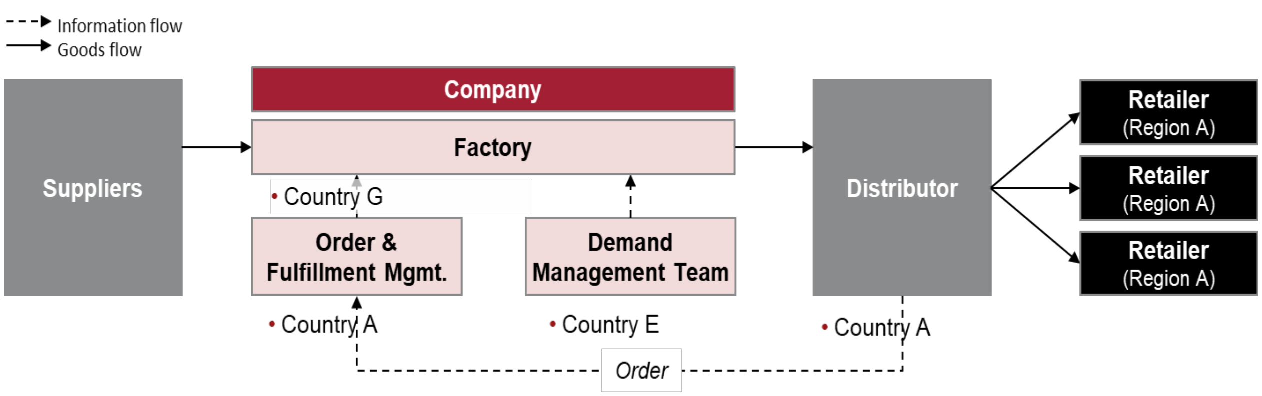


Figure2 Overall process of information and goods from suppliers to retailers. Data from the sponsor company from the sponsor company

Scope

In the case examined for this research, a distributor, which has a direct relationship with retailers in various regions, places orders through a fulfillment management team in Country D. Then, Country D team will manage the orders from distributors and deliver the information to the factories in Europe. Receiving materials from suppliers of many countries, the factory fulfills the order to the distributor in Country A. In the process, Demand Management team located in Country E forecasts shipments based on historical shipment data to analyze the available capacity and to provide forecasts to fulfillment management team (see Figure 2). Even though the company has set up firm process across the countries, due to the sporadic demand we believe there is a huge opportunity for improvements.

Even though the analysis requires more data points, we considered the basic data set.

i) Shipments data represents the actual shipment volume for each SKU that was sent from the company to various distributors. The data was defined weekly and monthly, as the system of the company was not able to gather data at a more detailed level.

ii) Forecast data that includes forecasted volume for each item was analyzed by the company. This data will be the starting point to develop the forecasting method and later it will be used to compare various metrics such as RMSE (Root Mean Square Error) or A-MAPE (Adjusted Mean Absolute Percent Error) with developed model.

iii) Inventory data represents end of month inventory balance at the distributor. Due to the system's limitation, it was only received as monthly figures. Through analyzing the inventory data, we could examine the stock-outs and how those events could impact the sales and shipments of the company.

iv) Sellout data represents the material flow that was sent from distributors to the retailers. The data was gathered to understand the behavior of distributor market, as the demand from retail side continuously affects the volume of distributors' shipments.

v) Building Block Data: building blocks represents the budget and promotional activities shapes the distributor demand forecast.

Results

We first developed a simulation experiment to compare the result with the internal forecast. As we initially assumed that the company will replenish the products that fall below the target inventory level, we conducted the simulation to verify this behavior. The target inventory level was calculated by the average of inventory level for the past one year (in SKU level). Then we estimated the shipment volumes by using the below equation:

Estimated Shipment = Target Inventory – (Last Month Ending Inventory – Sell-out)

That is, if the last month inventory minus sellout falls under the target inventory level, the company will replenish. If it does not fall below that level, the company will not replenish it.

Then we ran the regression model with the mentioned five predictors on different categories and also investigated seasonality effect on the shipment volume. We found that Inventory and sellout of the period prior to shipment are more correlated with shipment volume than the current period figures. Building blocks and seasonality were also not significant with high p-value greater than 0.05. The final model was able to capture 76% of the variation. We also validated the distribution of the difference between prediction and actual shipment volume and we found that it is normally distributed. The output of this process was two models one to predict the expected sellout, and the other model predicted expected supply using expected demand and other explanatory variables.

In order to forecast at the SKU level, we predicted the shipments at the category level and broke them down to the SKU level. Shipment volumes for each category were split based on each SKU's contribution to the total category volume.

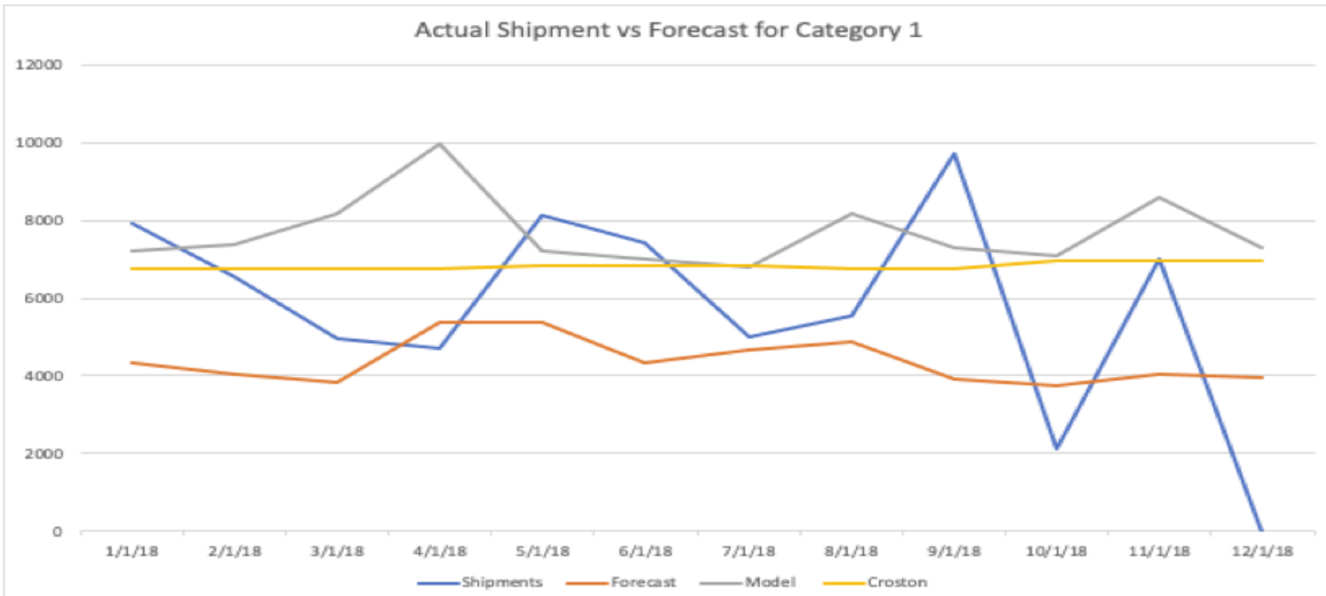
As for the results, 78 SKUs were improved in terms of RMSE (see Table 1), while 37 SKUs were not improved.

	Oct	Nov	Average
20%~	71	63	67
10~20%	3	5	4
0~10%	4	10	7
No Improv	37	37	37
Total	**115**	**115**	**115**

Table 1 Improvement of Forecasting Accuracy in SKU.

Conclusions

After we conducted the analysis, we noticed that some forecasting methods outperform other techniques in some categories. Each method has its own benefits and limitations. For example, as shown in Figure 3, Croston's method maintains a stable level across the time horizon which will positively impact the upstream manufacturing in terms of planning and setting inventory levels, as it doesn't get affected by the lumpiness of the demand. The current forecasting method of moving average also has a benefit of higher accuracy in some of the categories, however, moving average gets affected by periods that has no demand and may cause some disruption in upstream if the there are multiple periods with zero demand. The multi-tier regression analysis has a benefit of integrating the downstream distributor data, and it has improved the accuracy in 4 out of 5 categories in terms of A-MAPE as shown in section 6.1.4 Table 9. However, it's high dependent on the accuracy and the availability of these data.



[Figure3 Shipment vs forecast techniques comparison for category 1 on 2018]