Determination Forecasting Sporadic Demand in Supply Chain Management

*1Mehmet Gülşen and 2Kunter İpek
1,2Başkent University, Faculty of Engineering, Ankara, Turkey

Abstract:

With increasing number of SKUs(stock keeping units) in the supply chain, demand data for many products have become more sporadic with few nonzero observations. Almost binary-like pattern of the demand data makes forecasting difficult. There are several established methods such as Croston's exponential smoothing or SBA with limited general success. Most of the time forecasting of a sporadic demand requires modification of a standard method to incorporate some domain specific information into the forecasting model. In this research, we present a demand forecasting approach for a local candy manufacturer. The company has a portfolio of products with highly sporadic demand. Our approach includes standard methods and an alternative model that is based on the dynamic aggregation of the demand data.

Key words: forecasting, sporadic demand, croston

1. Introduction

In consumer driven market economy of today, the customer needs have become more sophisticated. The product life cycles are shortened while the variety of products gets wider, and customers expect fast availability of products in the market place. This put pressure on all actors of the supply chain to procure and produce expanding variety of products in a short notice. The number of SKU's (stock keeping units) in the supply chain has increased significantly in recent years. According to FMI, average number of SKUs in a supermarket is 39,500[1]. For big stores, the average moves up to 60,000 items[2]. The increasing number of SKU's brings several challenges to the management of supply chains. As more items move inside the network, the proper management tools need to be developed for demand forecasting and inventory management. With increasing number of products, the data becomes more granular and sparse. In a department store, a simple polo shirt is categorized according to color and size. There may be other categorization factors, such as: availability of breast pocket, packaging, gender specific cutting and half sizes. The eventual product variety will be determined by the product of those categories, and it may easily go up to several hundred with the inclusion of other categories. The supply chain management system needs to generate demand forecast and develop stock keeping strategies for each such sub-category. One challenging issue is the lower granularity of data at sub-category level. There may be periods with no activity that are registered as zero demand/production periods in historical records. The data stream looks like a binary series with zeros and low-valued positive numbers. Such data sets are called intermittent or sporadic data, and they present challenges when directly used in many of the standard forecasting tools. Based on the research and practices in the real-life applications, several forecasting methods are favored

*Corresponding author: Address: Faculty of Engineering, Department of Industrial Engineering, Başkent University, Ankara, TURKEY. E-mail address: mgulsen@baskent.edu.tr, Phone: +90-312-246-6666

over others for intermittent and sporadic data[3]. The short list includes simple moving averages(SMA), simple exponential smoothing(SES) and Croston's method. These tools have gained wide acceptance of practitioners because of their simplicity, and they can work with data sets that include zero-valued observations.

2. Literature

Any time series type of forecasting method can be used for intermittent data. Some of the common approaches have gained popularity because of their simplicity and better accuracy. The brief explanation and review of these methods are given in this section.

2.1 Simple Moving Average (SMA)

The underlying assumption in simple moving average is that the best estimate for the future demand is the average of what were observed recently. In "Simple Moving Average" method, demand forecast for the next period is given by:

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + \ldots + Y_{t-m+1}}{m} \quad (1)$$

The future value is based on the average of past *m* observation. Each observation has an equal weight of (*1/m*). One critical factor in simple moving average method is the length of history. If *m* is set to higher value, a longer history will be used. This will be more effective in terms of filtering noise, but the model will be slow in responding to changes in data trends. If *m* is smaller, the forecast model will be more responsive. However, that will increase the risk that the model is responding a noise, not an underlying trend change.

2.2 Simple Exponential Smoothing (SES) Method

Exponential smoothing method is similar to simple moving average as both methods aims to catch the underlying behavior of the past data based on the mean value of the past observations. The main difference between two methods is in how they calculate the historical mean. In simple moving average, each observation has an equal weight whether the event occurred recently or far in the past. Every single observation of history has equal impact on the future forecast. In exponential smoothing the historical average is calculated in a *biased* manner. A recent event is assumed to have more relevance as compared to something happened long ago. Thus, a higher weight is given to a recent observation when calculating the mean. In SES approach, geometrically declining weights are assigned to each historical data point, starting from the most recent observation to the oldest one.

The exponential smoothing is probably the most common forecasting tool used in business applications. This primarily due to simplicity of the method, and due to fact that SES produces robust forecasts under variety of conditions. In a business environment, thousands of forecast need to be generated automatically every single day. Under such circumstances, having a robust system has ultimate importance, as there may be no chance for a manual intervention.

The SES model could be expressed in different forms. One common way to define the forecasted demand is to use interpolation between the most recent actual observation and the most recent forecast.

$$\hat{Y}_{t+1} = \alpha X_t + (1-\alpha)\hat{Y}_t \qquad (2)$$

where α is the *smoothing constant* between 0 and 1. The SES is a dynamic and robust approach to forecasting. As soon as we have the first actual observation we can start generating forecasts. Periodically the forecast is updated as more actual data become available. The alpha determines the responsiveness of the model. If it is set to a lower value the changes in forecasts from one period to another will be less volatile. This will smooth out the noise effect in the data, but it will be slow down the model's ability to respond to trend changes in the data set.

The SES approach is based on the exponentially weighted average of past observations. To illustrate this point, we can expand the equation in (2) to include all past observations.

$$\begin{aligned}
\hat{Y}_{t+1} &= \alpha X_t + (1-\alpha)\hat{Y}_t \\
&= \alpha X_t + (1-\alpha)\left[\alpha X_{t-1} + (1-\alpha)\hat{Y}_{t-1}\right] = \alpha X_t + \alpha(1-\alpha)X_{t-1} + (1-\alpha)^2\hat{Y}_{t-1} \\
&= \alpha X_t + \alpha(1-\alpha)X_{t-1} + (1-\alpha)^2\left[\alpha X_{t-2} + (1-\alpha)\hat{Y}_{t-2}\right] = \alpha X_t + \alpha(1-\alpha)^1 X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + (1-\alpha)^3\hat{Y}_{t-2} \qquad (3) \\
&= \ldots\ldots.. \\
&= \alpha X_t + \alpha(1-\alpha)^1 X_{t-1} + \alpha(1-\alpha)^2 X_{t-2} + \alpha(1-\alpha)^3 X_{t-3} + \alpha(1-\alpha)^4 X_{t-4} + \ldots..
\end{aligned}$$

Unlike SMA, the exponential moving average use the entire history, and the weight of an observation is discounted by a factor of 1-α for each period of movement on the time frame.

2.3 Croston's Method

Croston's approach to forecasting can be considered as extension of the SES method[4]. The primary difference is in the data handling. Croston divides the input data into two sub-series. The first set includes only non-zero observations. The second set stores time durations between the non-zero observations. By using the traditional SES method two sets of forecasts are generated: (1) magnitude of non-zero demand, (2) time interval between two non-zero events. The smoothing factor alpha is kept identical in both forecasts.

The final resulting forecast is the ratio of non-zero demand forecast to time interval forecast. This number is not the forecast of the demand, but a forecast for the demand rate[5][6]. In Croston's method the forecast is updated only after a positive demand is observed. Otherwise it remains constant.

Y_t original series

Z_t non - zero demand series

T_t time internal series (4)

$$\hat{Y}_{t+1} = \frac{\hat{Z}_{t+1}}{\hat{T}_{t+1}}$$

The method proposed by Croston's has been heavily researched and has been a subject of some criticism. In 1973, Rao published a corrected version of Croston's method[7]. A modified version is developed to eliminate bias due to value of smoothing parameter. The original forecast of the Croston's is multiplied with a de-biasing factor[8][9].

$$\hat{Y}_{t+1} = (1 - \frac{\alpha_p}{2}) \frac{\hat{Z}_{t+1}}{\hat{T}_{t+1}} \qquad (5)$$

3. Empirical Study

The actual data from a confectionary producer is used to evaluate the performance of the various forecasting approaches. The data set include historical orders placed by local and international customers. The history covers entire 2016 and the first four months of 2017. In a period of 16 months the company received over two thousand orders for 217 different confectionary items. Although the company has a portfolio of over 300 hundred products, the top 10 items in the order history makes almost half of confectionary ordered by volume. On the other end of the spectrum, out of 227 products ordered within 16 months, 64 items only ordered once. As it can be seen in Table 1, the order data exhibit typical characteristics of an intermittent data. Within the observation period, there are 2007 orders for 227 different products. The average ordering frequency is 2007/227 = 9. The distribution is highly skewed to the lower end. The median ordering frequency is three.

Table 1: Distribution of ordering frequency for 227 products

Number of orders placed	**Frequency**
1	64
2	33
3	12
4+	118
	227(total)

To test the effectiveness of different approaches, the data set is divided into two sections. We remove the last 30 days of the history and keep it as holdout period. The original data set minus last 30 days is used as training data. The forecasting model is developed on the truncated data.

Once the model is ready, a 30-day projection of demand is generated and compared with the holdout data.

Figure 1 shows the history and forecast for the most frequently order item (product id: 152 01 121) in the product portfolio. The black and blue lines represent training and holdout portions of the original data set. The green line represents the back-fitting of the model. The red line shows the 30-day projection. The forecasting performance of each model is measured over the holdout period.

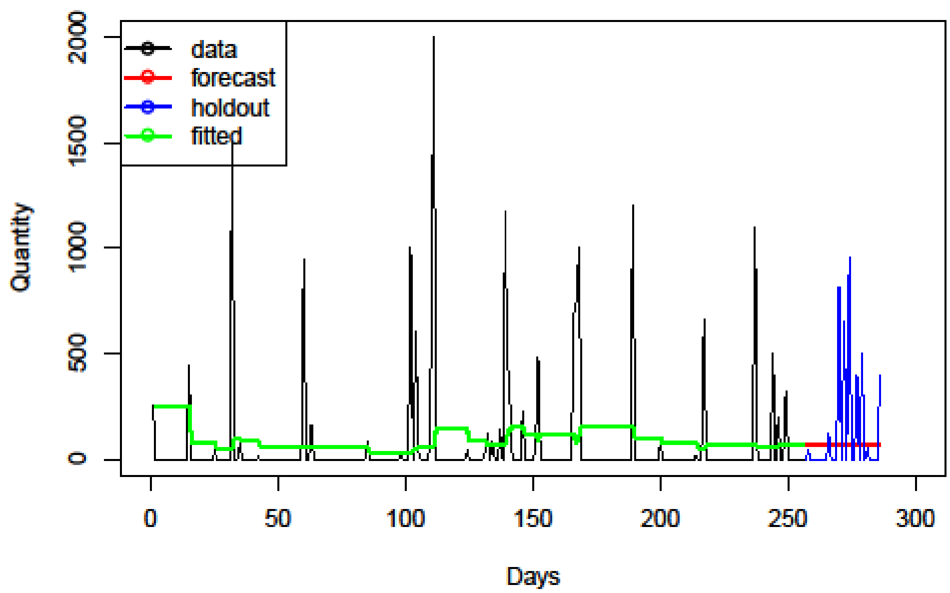

Figure 1: Order history and forecast for the most frequently ordered item

One challenging area for sporadic data is the selection of proper error metrics. For comparison purposes, scale-free metrics such as MAPE (mean absolute percent error) are frequently used in evaluating forecast performance of different data series. However, MAPE is based on percentage error, $100e_t / Y_t$, requires division to Y_t, which is commonly equal to zero in sporadic data sets. Several alternatives metrics are proposed in the literature[10][11].

The MASE, mean absolute scaled error, an alternative error metrics used frequently in evaluating forecast performance in sporadic data models. In MASE, the performance of the proposed method is evaluated with respect to the performance of a simple benchmark. Let's assume that a naïve forecast is completely based on the last observed value ($\hat{Y}_{t+1} = Y_t$) is used as benchmark. Then the scaled error will be:

$$q_t = \frac{e_t}{\frac{1}{n-1}\sum_{i=1}^{n-1}\left|Y_{t+1} - Y_t\right|}$$

The mean absolute scale error is calculated by averaging the absolute values of q_t's. A value greater 1 indicates that the performance is worse than the performance of a naïve forecasting method.

3. Results

Three different forecasting methods are used in this study. First one is a simple average which is based on historical average demand. The calculated average value is used as a static projection of the future demand. The other two approaches used in this study, exponential smoothing and Croston are dynamic methods which means forecasts are updated with each new observation. The simple average is the worst performing method as compared to other two. This is probably due to static nature of the approach. Table 2 shows performance of each approach on 10 different data sets. These are the most frequently ordered products within the last 16 months. The table includes one more item, 000 00 000, which is the aggregate of previous 10 items. The forecast performance for the aggregated data is relatively better than individual products. Considering that lot of items in the data sets are ordered just once or twice a year, aggregation could be reasonable approach. Similar products with extremely sporadic demand could be pooled together to form a single product.

Table 2: Forecast comparison for 10 products

Product ID	Simple Average	Exponential Smoothing	Croston
152 01 117	1.054	1.033	1.009
152 01 121	1.098	1.026	0.914
152 02 02	0.962	0.938	0.843
152 02 06	1.197	1.033	0.970
152 03 01	1.020	0.931	0.890
152 03 34	1.232	1.066	1.494
152 03 44	1.178	1.007	0.959
152 03 45	1.196	1.006	0.908
152 03 57	0.996	0.946	1.022
152 03 84	1.262	1.021	0.969
000 00 000	1.095	0.952	0.961

3. Conclusion and future work

Among the three alternatives used in this study, only two, exponential smoothing and Croston's methods look like reasonable approaches. They do better than the benchmark (i.e., naïve) on average. We are planning to extend this study to include methodology to pick up the best forecasting model. For that purpose, the data should be tested for in-sample (i.e., training) and out-sample(i.e., holdout) data. Based on the performance on the training data a proper model should be selected for the holdout period.

References

[1] FMI: Food Marketing Institute, industry group for food retailers. https://www.fmi.org/our-research/supermarket-facts. Accessed on 03-July-2017

[2] http://www.foodretailworld.com/Definitions.htm. Accessed on 03-July-2017

[3] Petropoulos, F., Kourentzes, N., Nikolopoulos, K., Another look at estimators for intermittent demand. International Journal of Production Research: 2016; 181, p. 154-161

[4] Croston, J.D. Forecasting and stock control for intermittent demands. Operational Research Quarterly: 1972; 23(3), p. 289-303

[5] Kourentas, N. Intermittent demand forecasts with neural networks. International Journal of Production Economics: 2013; 143, p 198-206

[6] Petropoulos, F., Kourentzes, N. Forecast combinations for intermittent demand. Journal of Operations Research Society: 2015; 66, p. 914-924

[7] Rao, A. V. A comment on: Forecasting and stock control for intermittent demands. Operational Research Quarterly: 1973; 24(4), p. 639-640

[8] Syntetos, A. A., Boylan, J.E. On the bias of intermittent demand estimates. International Journal of Production Economics: 2001; 71, p. 457-466

[9] Syntetos, A. A., Boylan, J.E. The accuracy of intermittent demand estimates. International Journal of Forecasting: 2005; 21(2), p. 304-314

[10] Makridakis, S. & Hibon, M., 2000; The M3-competition: Results, conclusions and implications, International Journal of Forecasting, 16, 451–476.

[11]Hyndman, R. J. & Koehler, A. B., 2006; Another look at measures of forecast accuracy, International Journal of Forecasting.