Major Research project

Forecasting on Intermittent Demand

Submitted By:

Paili Rajesh

2K18/MBA/115

Under the guidance of:

Mr. Dhiraj Kumar Pal

Assistant Professor

DELHI SCHOOL OF MANAGEMENT

Delhi Technological University

Bawana Road Delhi 110042

ACKNOWLEDGEMENT

I offer my sincere thanks and humble regards to DELHI SCHOOL OF MANAGEMENT, DELHI TECHNOLOGICAL UNIVERSITY for imparting us valuable professional training in MBA.

I pay my gratitude and sincere regards to Mr. Dhiraj Kumar Pal, as he has been a constant source of motivation and inspiration. I am also thankful to him for giving his suggestions and encouragement throughout the research report. with out his continuous support and guidance this project wouldn’t have been such an amazing learning experience.

I am also thankful to my family and friends for constantly motivating me and providing me an environment which enhanced my knowledge.

And last but not the least, I would like to thank my University – Delhi technological University for providing me with this wonderful opportunity to work on this project and giving me complete support and exposure that I needed to complete my project.

Paili Rajesh

2K18/MBA/115

CERTIFICATE FROM GUIDE

This is to certify that the project titled **“Forecasting on Intermittent Demand”** is an academic work done by **“Paili Rajesh”** submitted in partial fulfilment of the requirement for the award of the degree of **“Master of Business Administration”** from **Delhi School of Management,** conducted under my guidance.

To the best of my knowledge and belief the data and information presented by her in the project has not been submitted earlier elsewhere.

Mr. Dhiraj Kumar Pal
Assistant Professor
Delhi School of Management
Delhi Technological University

TABLE OF CONTENTS

ABSTRACT:

Forecasting methods are often valued by means of simulation studies. For intermittent demand items there are often very few non–zero observations, so it is hard to check any assumptions, because statistical information is often too weak to determine, for example, distribution of a variable. Therefore, it seems important to verify the forecasting methods on the basis of real data. The main aim of the article is an empirical verification of several forecasting methods applicable in case of intermittent demand. Some items are sold only in specific subperiods (in given month in each year, for example), but most forecasting methods (such as Croston's method) give non–zero forecasts for all periods. For example, summer work clothes should have non–zero forecasts only for summer months and many methods will usually provide non–zero forecasts for all months under consideration. This was the motivation for proposing and testing a new forecasting technique which can be applicable to seasonal items. In the article eight methods were applied to construct separate forecasting systems such as,

- CROSTON
- TSB_CROSTON
- Hyperbolic Exponential Smoothening Model
- SBA
- Holt's Linear
- Error trend Seasonality
- L-STAR (logistic smooth transition Auto regressive)
- ARMA (Auto Regressive Moving Average)

The presented analysis might be helpful for enterprises facing the problem of forecasting intermittent items (and seasonal intermittent items as well).

1. INTRODUCTION

1.1 Background:

In Manufacturing sector, Some product items have an intermittent demand pattern that makes them all but impossible to forecast with traditional, smoothing-based forecasting methods. Items with intermittent demand – also known as lumpy, volatile, variable or unpredictable demand – have many zero or low volume values interspersed with random spikes of demand that are often many times larger than the average. This problem is especially prevalent in companies that manage large inventories of service and spare parts in industries such as aviation, aerospace, automotive, high tech, and electronics, as well as in MRO (Maintenance, Repair and Overhaul).

In these businesses, as much as 80% of the parts and product items may have intermittent or lumpy demand. Intermittent demand makes it difficult to accurately estimate the safety stock and service level inventory requirements needed for successful supply chain planning. Because forecasts of intermittent and lumpy demand have been so unreliable, most companies forecast inventory requirements relying primarily on subjective business knowledge, forecast only a fraction of their higher volume inventory, use simple “rule of thumb” estimates, or traditional statistical forecasting that incorrectly assumes a particular type demand distribution for inventory control. The result is that billions of dollars are wasted every year because of either excess inventory costs or poor customer service due to stock-outs. So that intermittent demand forecasting comes into play to resolve this issue. Manufacturers perceive the forecasting of intermittent data to be an important problem. In practice, the standard method of forecasting intermittent demand is single exponential smoothing, although some production management texts suggest the lesser-known alternative of Croston's method [Croston J.D., 1972, Forecasting and stock control for intermittent demands, Operational Research Quarterly.

1.2 Problem Statement:

Intermittent demand is a random demand, which the demand for a product or service appears sporadically, and lots of zero values exist in the demand data. (Croston,1972)

Expect the continuous production process, lots of demand events (materials, services…) are appear as intermittence. Although some of them can be converted to a continuous demand, for example, the sales of a product record per hour or minute might be intermittent, but it can be convert to the sales per day or month, and it still make sense.

However there also a certain amount of demand events that can not be converted form intermittent data to continuous data , otherwise they will loss the original purpose of the data. For example, the important equipment for manufacture, they can be used for quite a long period, but it can not be shortage; the spare parts of current working facility; the spare parts for transportation, like airplane. These intermittent demand materials often present as a significant part in company's operation, if shortage, it might lead to a huge loss, for example, if a aircraft can not fly properly, each hour it stay on the ground will incur a cost more than $50,000.

Unfortunately, according to Johnston et al. (2003) these intermittent demand items can constitute up to 60% of the total stock value, and because those items often present as low consumption rate and long demand interval, therefore, the risk of obsolescence is much higher than other items.

Therefore, a accurate intermittent demand forecasting is a necessary for achieving better inventory management and company operation. Here will introduce few intermittent demand forecasting methods.

Statstical Models:

1. CROSTON
2. TSB_CROSTON
3. Hyperbolic Exponential Smoothening Model
4. SBA
5. Error trend Seasonality

Mathmatical Models:

6. Holt's Linear
7. L-STAR (logistic smooth transition Auto regressive)
8. ARMA (Auto Regressive Moving Average)

1.3 Objective of the Study:

Forecasting methods are often valued by means of simulation studies. For intermittent demand items there are often very few non–zero observations, so it is hard to check any assumptions, because statistical information is often too weak to determine, for example, distribution of a variable. Therefore, it seems important to verify the forecasting methods on the basis of real data. The main aim of the article is an empirical verification of several forecasting methods applicable in case of intermittent demand. Some items are sold only in specific subperiods (in given month in each year, for example), but most forecasting methods (such as Croston's method) give non–zero forecasts for all periods. For example, summer work clothes should have non–zero forecasts only for summer months and many methods will usually provide non–zero forecasts for all months under consideration. This was the motivation for proposing and testing a new forecasting technique which can be applicable to seasonal items. In the article six methods were applied to construct separate forecasting systems: Croston's, SBA (Syntetos–Boylan Approximation), TSB (Teunter, Syntetos, Babai), MA (Moving Average), SES (Simple Exponential Smoothing) and Holts linear, L-Star, ARMA. A data set from the real company was used to apply all the above forecasting procedures. That data set contained monthly time series for about sixteen thousand products. The forecasts accuracy was tested by means of both parametric and non–parametric measures. The mean absolute scaled error and the shares of best forecasts were estimated. The general conclusion is that in the analyzed company a forecasting system should be based on two forecasting methods: TSB and ARMA, but the latter method should be applied only to seasonal items (products sold only in specific subperiods). It also turned out that Croston's and SBA methods work worse than much simpler methods, such as MA. The presented analysis might be helpful for enterprises facing the problem of forecasting intermittent items.

1.4 Scope of the Study:

The main focus of this study was the forecasting for intermittent demand for industrial warehouses & supply chain fields to predict the upcoming demand accurately so that they can plan their inventories & order levels efficiently.

In this research, we take various factors into consideration to maximizing efficiency. Factors include data patterns i.e frequency of demand & size of the demand which gives us a type of demand and how fast it is moving.

Demand has been classified based on its patterns because few items need not necessarily follow the particular pattern so those items should never be predicted.

The proposed research comprises multiple meta-models which calculate previous demand patterns. Does the resulting MAPE value determine whether the model is efficient?

2. LITERATURE REVIEW

Forecasting demand

In today's organizations, which are subject to abrupt and enormous changes that affect even the most established of structures and where all requirements of business sector need accurate and practical reading into future, the forecasts are becoming very crucial since they are the sign of survival and the language of business in the world. A forecast is a science of estimating the future level of some variables. The variable is most often demand, but it can also be something else, such as supply or price. Forecasting is the operation of making assumption about the future values of studied variables.

In manufacturing, forecasting demands is among the most crucial issues in inventory management; it can be used in various operational planning activities during the production process: capacity planning, used-product acquisition management

For both types of supply chain processes "push/pull," the demand forecasts are considered the ground of supply chain's planning. The pull processes in the supply chain are realized with reference to customer demand, while all push processes are realized in anticipation of customer demand. A company must take into consideration such factors before selecting a suitable forecasting methodology because the choice of a methodology is not as simple as it seems. Forecasting methods are categorized according to four types: qualitative, time series, causal, and simulation.

A time series is nothing but observations according to the chronological order of time.17 Time series forecasting models use mathematical techniques that are based on historical data to forecast demand. It is founded on the hypothesis that the future is an expansion of the past; that's why we can definitely use historical data to forecast future demand.

Many studies about demand forecasting by time series analysis have been done in several domains. They encircle demand forecasting for food product sales, tourism, maintenance repair parts, electricity, automobile, and some other products and services.

By time series analysis, the forecasting accuracies depend on the characteristics of time series of demand. If the transition curves show stability and periodicity, we will reach high forecasting accuracies, whereas we can't expect high accuracies if the curves contain highly irregular patterns.

Croston's method the first ID-specific method was proposed by Croston. His insight was that estimating demand probability (via interval size) and demand size separately was more intuitive and accurate. Let Zt be the estimate of mean non-zero demand size for time t, Vt the estimate of mean interval size between non-zero demands. Xt again denotes actual demand observed at time t, and q is the current number of consecutive zero-demand periods. Yt will denote an estimate of mean demand size (ie. taking zero demands into the calculation). Then,

$If\ z_t = 0, then$

$z'_t = z'_{t-1}$,

$p'_t = p'_{t-1}$

$Otherwise$

$z'_t = \alpha z_t + (1\text{-}\alpha) z'_{t-1}$

$p'_t = \alpha p_t + (1\text{-}\alpha) p'_{t-1}$ $\qquad where\ 0 < \alpha < 1$

Y'_t — Average demand per period

z_t — Actual demand at period t

z'_t — Time between two positive demand

p — Demand size forecast for next period

p'_t — Forecast of demand interval

α — Smoothing constant

SBA (Syntetos-Boylan Approximation):

Many adaptations of Croston's method have been suggested to deal with some of the aforementioned issues. In [6], the authors propose an adjustment, known as the Syntetos-Boylan Approximation (SBA), to Croston's forecast Yt , namely that it should be multiplied by a factor of(1-(alpha/2)) and claim that the new forecast will be approximately unbiased, since

$$\begin{aligned} E\left[\left(1-\frac{\alpha}{2}\right)\left(\frac{Z_t}{V_t}\right)\right] &= \left(1-\frac{\alpha}{2}\right) E\left[\frac{Z_t}{V_t}\right] \\ &= \left(1-\frac{\alpha}{2}\right)\left(\frac{\mu}{p}+\frac{1}{2}\frac{\partial^2(\frac{\mu}{p})}{\partial p^2}\operatorname{Var}(p)\right) \\ &= \left(\frac{2-\alpha}{2}\right)\left(\frac{\mu}{p}+\frac{\alpha}{2-\alpha}\mu\frac{p-1}{p^2}\right) \\ &= \frac{\mu}{p}\left(\frac{2-\alpha}{2}+\frac{\alpha}{2}\frac{p-1}{p}\right) \approx \frac{\mu}{p} \end{aligned}$$

TSB (Teunter, Syntetos and Babai):

$$\text{If } D_t = 1 \text{ then } \begin{cases} P_{t+1} &= \beta(1) + (1-\beta)P_t \\ Z_{t+1} &= \alpha X_t + (1-\alpha)Z_t \\ Y_{t+1} &= P_{t+1}Z_{t+1} \end{cases}$$

$$\text{If } D_t = 0 \text{ then } \begin{cases} P_{t+1} &= (1-\beta)P_t \\ Z_{t+1} &= Z_t \\ Y_{t+1} &= P_{t+1}Z_{t+1} \end{cases}$$

The key advantage of this method is that updating the forecasts at each time period (whether demand occurs or not) allows the estimate Pt to approach zero if there is a long run of periods without demand. By contrast, the interval estimate calculated by Croston's would remain unchanged. In a practical setting, TSB allows decisions to be taken over whether to continue to stock items or not. One idea might be to set some threshold for p, such that if it is exceeded (just once or for a number of consecutive periods) it would be decided that the product is now obsolete

Hyperbolic Exponential Smoothing (HES):

Different method dealing with obsolescence was proposed recently by Prestwich. and is known as Hyperbolic Exponential Smoothing (HES). The method combines the Croston method with a Bayesian approach to derive a new forecast, where if Zt and Vt are the smoothed estimates of demand size and interval length, and Tt is the current number of periods since a demand was last seen, then

$$F_t = \frac{Z_t}{V_t + \beta \frac{T_t}{2}}$$

is the forecast for time t. Note β is the smoothing parameter for interval length. The main difference between TSB and HES is that HES decays hyperbolically over a series of zeros, whereas TSB decays only exponentially.

ARMA model:

An ARMA model, or Autoregressive Moving Average model, is used to describe weakly stationary stochastic time series in terms of two polynomials. The first of these polynomials is for autoregression, the second for the moving average.

Often this model is referred to as the ARMA(p,q) model; where:

p is the order of the autoregressive polynomial,

q is the order of the moving average polynomial.

The equation is given by:

arma model

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}.$$

Where:

φ = the autoregressive model's parameters,

θ = the moving average model's parameters.

c = a constant,

ε = error terms (white noise).

Holt's Linear Trend Method (Double Exponential Smoothing):

Holt's two-parameter model, also known as linear exponential smoothing, is a popular smoothing model for forecasting data with trend. Holt's model has three separate equations that work together to generate a final forecast. The first is a basic smoothing equation that directly adjusts the last smoothed value for last period's trend. The trend itself is updated over time through the second equation, where the trend is expressed as the difference between the last two smoothed values. Finally, the third equation is used to generate the final forecast. Holt's model uses two parameters, one for the overall smoothing and the other for the trend smoothing equation. The method is also called double exponential smoothing or trend-enhanced exponential smoothing.

(Level) $L_t = \alpha * (Y_t - S_{t-s}) + (1 - \alpha) * (L_{t-1} + b_{t-1})$

(Trend) $b_t = \beta * (L_t - L_{t-1}) + (1 - \beta) * b_{t-1}$

(Seasonal) $S_t = \gamma * (Y_t - L_t) + (1 - \gamma) * S_{t-s}$

(Forecast for period m) $F_{t+m} = L_t + m{*}b_t + S_{t+m-s}$

where the parameters are:

α—Alpha

β—Beta

γ—Gamma

m—Number of periods ahead to forecast

s—Length of the seasonality

Lt—Level of the series at time t

bt—Trend of the series at time t

St—Seasonal component at time t

LSTAR model: The observations yt are generated either from the first regime when yt−d is smaller than the threshold, or from the second regime when yt−d is greater than the threshold. If the binary indicator function is replaced by a smooth transition function 0 < G(zt) < 1 which depends on a transition variable zt (like the threshold variable in TAR models), the model becomes a smooth transition autoregressive (STAR) model :

$$yt = Xt\varphi(1)(1 - G(zt)) + Xt\varphi(2)G(zt) + et$$

3. RESEARCH METHODOLOGY

Much inventory management is done subjectively. However, among objective approaches, the most popular appears to be the theory of economic order quantities (EOQ), which determines two quantities for each item: a reorder point and an order quantity. When on-hand inventory reaches the reorder point, one orders an amount equal to the order quantity to replenish the stock. Calculating the order quantity normally requires forecasts only of the average demand per period. In contrast, calculating the correct reorder point requires estimates of the entire distribution of demand over the interval, known as the lead time, between the generation of a replenishment order and its arrival in inventory. (For example, a specific percentile of the predicted demand distribution, e.g. 99%, indicates the level at which the inventory reorder point should be set to ensure a corresponding likelihood of not stocking out of that item during the lead time.) Textbooks simplify this problem by assuming that demands in each time period are independent and normally distributed; neither assumption is valid for intermittent demand, Our approach improves on theirs by dealing with autocorrelation between successive demands and by incorporating a variant of the smoothing that they speculated would be helpful and also, we designed Meta (Boot strapping with different models) model which will give better results.

Industrial datasets:

We assessed the relative accuracy of the various forecasting methods using industrial data from 16000 different parts which further different demand patterns with different sizes. The items provided to us were deliberately biased toward series considered most difficult to forecast. In all cases, the items were ‘live’ rather than at the very end of their life cycles (as indicated, for example, by end-of-life items having 99% zero values. (Data set: A Major common carrier and its Aircraft service parts)

Dec-16	Jan-17	Feb-17	Mar-17	Apr-17	May-17	Jun-17	Jul-17	Aug-17	Sep-17
41.00	0.00	0.00	10.00	0.00	30.00	12.00	30.00	15.00	14.00
3.00	10.00	2.00	5.00	6.00	5.00	7.00	8.00	4.00	9.00
17.00	0.00	2.00	0.00	11.00	5.00	4.00	5.00	3.00	8.00
4.00	7.00	0.00	2.00	9.00	3.00	0.00	2.00	1.00	1.00
1.00	2.00	2.00	0.00	2.00	2.00	2.00	6.00	1.00	2.00
1.00	4.00	2.00	3.00	2.00	2.00	2.00	6.00	1.00	1.00
21.00	15.00	18.00	24.00	12.00	22.00	17.00	12.00	12.00	27.00
18.00	21.00	13.00	22.00	7.00	26.00	17.00	6.00	17.00	26.00
0.00	30.00	20.00	0.00	110.00	110.00	50.00	80.00	0.00	50.00
30.00	233.00	37.00	10.00	27.00	110.00	56.00	230.00	29.00	107.00
8.00	15.00	5.00	13.00	38.00	8.00	10.00	5.00	18.00	10.00
0.00	10.00	2.00	0.00	41.00	12.00	15.00	15.00	0.00	15.00
10.00	6.00	0.00	19.00	10.00	9.00	20.00	5.00	0.00	16.00
5.00	21.00	22.00	8.00	0.00	5.00	0.00	10.00	1.00	0.00
1.00	0.00	2.00	3.00	4.00	0.00	0.00	5.00	1.00	1.00
18.00	4.00	8.00	16.00	8.00	10.00	11.00	16.00	13.00	0.00
40.00	0.00	30.00	70.00	40.00	40.00	30.00	120.00	20.00	40.00
70.00	90.00	120.00	60.00	180.00	20.00	130.00	80.00	89.00	120.00
90.00	68.00	152.00	37.00	88.00	83.00	103.00	92.00	60.00	32.00

Given data set

Count	No of Zero Demar	No of demands	ADI	Std Dev	Average Pop	CV^2	Type of demand	Smoothe		Movement
36	7	29	1.24137931	11.783631	15.06	0.6125827	Erratic Demand	0	19%	Fast Moving
36	3	33	1.090909091	2.9454027	5.19	0.3215223	Smooth Demand	0	8%	Fast Moving
36	4	32	1.125	3.7984959	4.83	0.6176321	Erratic Demand	2	11%	Fast Moving
36	8	28	1.285714286	2.9691536	2.61	1.2930479	Erratic Demand	2	22%	Fast Moving
36	8	28	1.285714286	1.6561573	1.67	0.9874286	Erratic Demand	0	22%	Fast Moving
36	6	30	1.2	1.7698982	1.81	0.9608926	Erratic Demand	0	17%	Fast Moving
36	0	36	1	7.7660229	21.44	0.1311498	Smooth Demand	0	0%	Fast Moving
36	0	36	1	8.5033607	21.08	0.1626682	Smooth Demand	2	0%	Fast Moving
36	5	31	1.161290323	94.184469	99.17	0.9020428	Erratic Demand	2	14%	Fast Moving
36	0	36	1	75.332027	96.67	0.6073036	Erratic Demand	2	0%	Fast Moving
36	2	34	1.058823529	12.0321	16.83	0.5109079	Erratic Demand	0	6%	Fast Moving
36	5	31	1.161290323	12.300987	11.67	1.1116968	Erratic Demand	2	14%	Fast Moving
36	8	28	1.285714286	10.221476	10.75	0.9040872	Erratic Demand	0	22%	Fast Moving
36	11	25	1.44	7.4018445	7.11	1.0834403	Lumpy Demand	0	31%	Fast Moving
36	9	27	1.333333333	1.5884004	1.64	0.939336	Lumpy Demand	2	25%	Fast Moving
36	8	28	1.285714286	5.2396534	6.56	0.6388312	Erratic Demand	0	22%	Fast Moving
36	4	32	1.125	38.005848	48.89	0.6043388	Erratic Demand	0	11%	Fast Moving

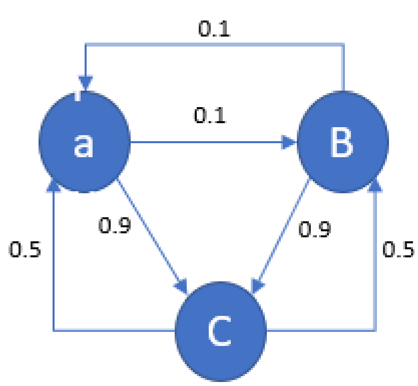
Classification of the data based on it's frequency & demand gaps (non-zero demand)

Meta model:

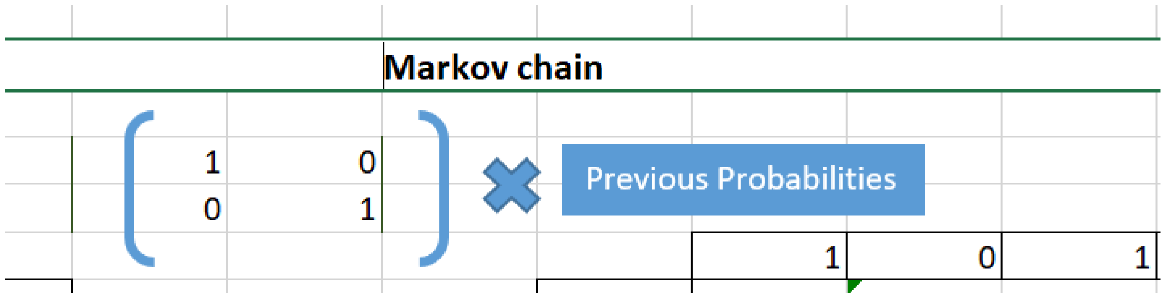
Bootstrapping with markov probability model- Bootstrapping is a statistical technique involving random sampling with replacement. In 2004 Willemain proposed a method using bootstrapping on previous observations of non-zero demand to forecast demand over some lead time (the interval between replenishment ordering and arrival). Two key ideas are present in the method. Firstly, to avoid making forecasts that can only take the same values as have occurred previously, a jittering process is used, allowing for more variation. Let Y be the selected previous value of demand, and Z be a standard normal random variable.

$$Y_{jittered} = 1 + INT\{Y + Z\sqrt{Y}\}$$

Secondly, to model autocorrelation that might be present in the demand, a two-stage Markov Chain model is used, with the states corresponding to zero and non-zero demand observations. The idea is to forecast a series of zero and non-zero demands first, with the chance of seeing either in the next step dependent on the transition matrix probabilities, which are to be estimated.



$$P = \begin{bmatrix} 0 & 0.1 & 0.9 \\ 0.1 & 0.0 & 0.9 \\ 0.5 & 0.5 & 0.0 \end{bmatrix}$$




```
=IF(AND(IF(B3>0,1,0)=1,IF(B2>0,1,0)=1),1,IF(AND(IF(B3>0,1,0)=0,IF(B2>0,1,0)=0),2,IF(AND(IF(B3>0,1,0)=0,IF(B2>0,1,0)=1),3,IF(AND(IF(B3>0,1,0)=1,IF(B2>0,1,0)=0),4,0))))
=IF(AND(IF(B4>0,1,0)=1,IF(B3>0,1,0)=1),1,IF(AND(IF(B4>0,1,0)=0,IF(B3>0,1,0)=0),2,IF(AND(IF(B4>0,1,0)=0,IF(B3>0,1,0)=1),3,IF(AND(IF(B4>0,1,0)=1,IF(B3>0,1,0)=0),4,0))))
=IF(AND(IF(B5>0,1,0)=1,IF(B4>0,1,0)=1),1,IF(AND(IF(B5>0,1,0)=0,IF(B4>0,1,0)=0),2,IF(AND(IF(B5>0,1,0)=0,IF(B4>0,1,0)=1),3,IF(AND(IF(B5>0,1,0)=1,IF(B4>0,1,0)=0),4,0))))
=IF(AND(IF(B6>0,1,0)=1,IF(B5>0,1,0)=1),1,IF(AND(IF(B6>0,1,0)=0,IF(B5>0,1,0)=0),2,IF(AND(IF(B6>0,1,0)=0,IF(B5>0,1,0)=1),3,IF(AND(IF(B6>0,1,0)=1,IF(B5>0,1,0)=0),4,0))))
=IF(AND(IF(B7>0,1,0)=1,IF(B6>0,1,0)=1),1,IF(AND(IF(B7>0,1,0)=0,IF(B6>0,1,0)=0),2,IF(AND(IF(B7>0,1,0)=0,IF(B6>0,1,0)=1),3,IF(AND(IF(B7>0,1,0)=1,IF(B6>0,1,0)=0),4,0))))
=IF(AND(IF(B8>0,1,0)=1,IF(B7>0,1,0)=1),1,IF(AND(IF(B8>0,1,0)=0,IF(B7>0,1,0)=0),2,IF(AND(IF(B8>0,1,0)=0,IF(B7>0,1,0)=1),3,IF(AND(IF(B8>0,1,0)=1,IF(B7>0,1,0)=0),4,0))))
=IF(AND(IF(B9>0,1,0)=1,IF(B8>0,1,0)=1),1,IF(AND(IF(B9>0,1,0)=0,IF(B8>0,1,0)=0),2,IF(AND(IF(B9>0,1,0)=0,IF(B8>0,1,0)=1),3,IF(AND(IF(B9>0,1,0)=1,IF(B8>0,1,0)=0),4,0))))
=IF(AND(IF(B10>0,1,0)=1,IF(B9>0,1,0)=1),1,IF(AND(IF(B10>0,1,0)=0,IF(B9>0,1,0)=0),2,IF(AND(IF(B10>0,1,0)=0,IF(B9>0,1,0)=1),3,IF(AND(IF(B10>0,1,0)=1,IF(B9>0,1,0)=0),4,0))))
=IF(AND(IF(B11>0,1,0)=1,IF(B10>0,1,0)=1),1,IF(AND(IF(B11>0,1,0)=0,IF(B10>0,1,0)=0),2,IF(AND(IF(B11>0,1,0)=0,IF(B10>0,1,0)=1),3,IF(AND(IF(B11>0,1,0)=1,IF(B10>0,1,0)=0),4,0))))
=IF(AND(IF(B12>0,1,0)=1,IF(B11>0,1,0)=1),1,IF(AND(IF(B12>0,1,0)=0,IF(B11>0,1,0)=0),2,IF(AND(IF(B12>0,1,0)=0,IF(B11>0,1,0)=1),3,IF(AND(IF(B12>0,1,0)=1,IF(B11>0,1,0)=0),4,0))))
=IF(AND(IF(B13>0,1,0)=1,IF(B12>0,1,0)=1),1,IF(AND(IF(B13>0,1,0)=0,IF(B12>0,1,0)=0),2,IF(AND(IF(B13>0,1,0)=0,IF(B12>0,1,0)=1),3,IF(AND(IF(B13>0,1,0)=1,IF(B12>0,1,0)=0),4,0))))
=IF(AND(IF(B14>0,1,0)=1,IF(B13>0,1,0)=1),1,IF(AND(IF(B14>0,1,0)=0,IF(B13>0,1,0)=0),2,IF(AND(IF(B14>0,1,0)=0,IF(B13>0,1,0)=1),3,IF(AND(IF(B14>0,1,0)=1,IF(B13>0,1,0)=0),4,0))))
=IF(AND(IF(B15>0,1,0)=1,IF(B14>0,1,0)=1),1,IF(AND(IF(B15>0,1,0)=0,IF(B14>0,1,0)=0),2,IF(AND(IF(B15>0,1,0)=0,IF(B14>0,1,0)=1),3,IF(AND(IF(B15>0,1,0)=1,IF(B14>0,1,0)=0),4,0))))
=IF(AND(IF(B16>0,1,0)=1,IF(B15>0,1,0)=1),1,IF(AND(IF(B16>0,1,0)=0,IF(B15>0,1,0)=0),2,IF(AND(IF(B16>0,1,0)=0,IF(B15>0,1,0)=1),3,IF(AND(IF(B16>0,1,0)=1,IF(B15>0,1,0)=0),4,0))))
=IF(AND(IF(B17>0,1,0)=1,IF(B16>0,1,0)=1),1,IF(AND(IF(B17>0,1,0)=0,IF(B16>0,1,0)=0),2,IF(AND(IF(B17>0,1,0)=0,IF(B16>0,1,0)=1),3,IF(AND(IF(B17>0,1,0)=1,IF(B16>0,1,0)=0),4,0))))
```

Markov matrix and it's formulas

	Chain of events in the next month
1	Demand to Demand
3	Demand to No demand
2	No demand to No demand
4	No demand to Demand

	1	0
1	=(COUNTIF($C$3:C40,1))/(COUNTIF($C$3:C40,1)+COUNTIF($C$3:C40,3))	=1-W6
0	=COUNTIF($C$3:$C40,4)/(COUNTIF($C$3:$C40,2)+COUNTIF($C$3:$C40,4))	=1-W7
1	=IF(W6<0.4,0,IF(W6<0.6,RANDBETWEEN(0,1),IF(W6>0.6,1,"Error")))	=IF(W9=1,0,1)
0	=IF(W7<0.4,0,IF(W7<0.6,RANDBETWEEN(0,1),IF(W7>0.6,1,"Error")))	=IF(W10=1,0,1)

Markov matrix and it's probability criteria

The full bootstrap approach as described goes as follows:

• With the historical data, estimate the transition probabilities for the two-state Markov Chain.

• Generate a sequence of events from the Markov Chain over the desired lead time.

Evaluation measures:

The performance of any forecasting method needs to be evaluated by some metric, to measure how closely the forecasted value matches the true value. Intermittent demand series turn out to be unusually tricky to evaluate; typical forecasting accuracy metrics are often either inappropriate or even impossible to apply. In this we used MEAN (Mean absolute percentage error) to evaluate the best and optimized results. The mean absolute percentage error (MAPE), also known as mean absolute percentage deviation (MAPD), is a measure of prediction accuracy of a forecasting method in statistics, for example in trend estimation, also used as a loss function for regression problems in machine learning. It usually expresses accuracy as a percentage, and is defined by the formula:

$$\mathrm{M} = \frac{100\%}{n} \sum_{t=1}^{n} \left| \frac{A_t - F_t}{A_t} \right|$$

4. CASE INTRODUCTION - INTERMITTENT DEMAND

What is intermittent demand:

Intermittent demand or ID (also known as sporadic demand) comes about when a product experiences several periods of zero demand. Often in these situations, when demand occurs it is small, and sometimes highly variable in size. ID is often experienced in industries such as aviation, automotive, defense and manufacturing; it also typically occurs with products nearing the end of their life cycle. Some companies operating in these areas observe ID for over half the products in their inventories. In such situations there is a clear financial incentive to inventory control and retaining proper stock levels, and therefore to forecasting demand for these items.

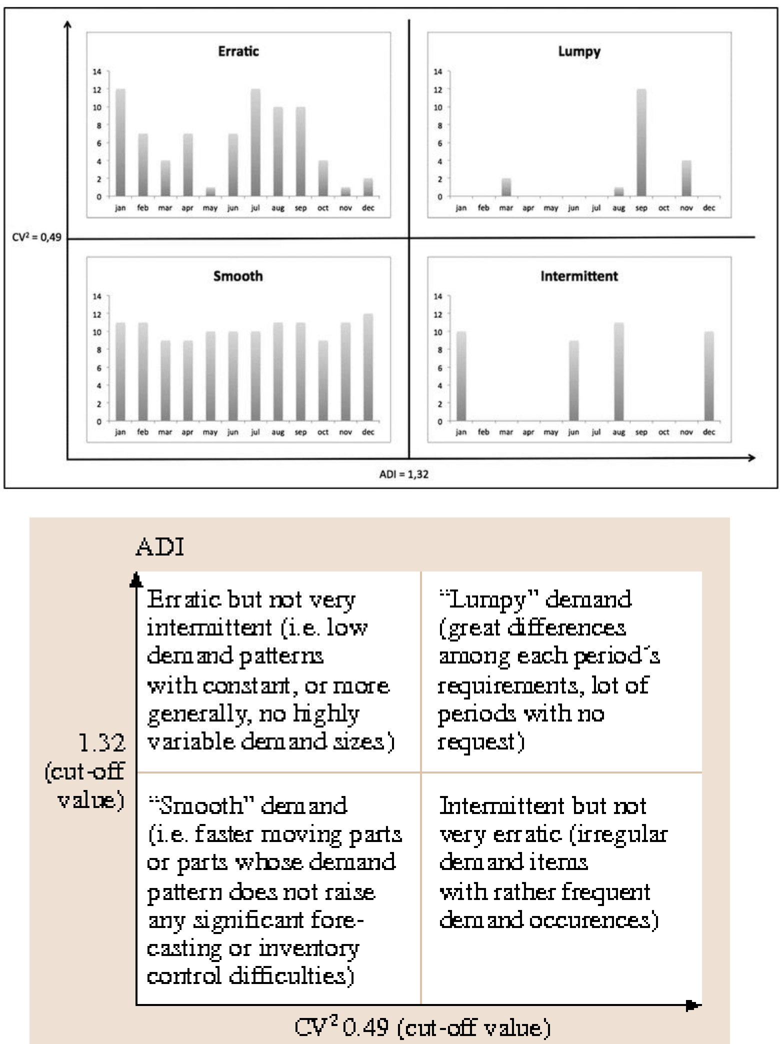


Fig 5.1 Demand classification

Forecasting difficulties:

The many zero values in ID time-series render usual forecasting methods difficult to apply. For example, single exponential smoothing (SES), proposed in 1956, was the first forecasting method to be applied to intermittent demand. The forecast of demand in the next period is a weighted average between two quantities, defined:

$$F_{t+1} = \alpha X_t + (1 - \alpha) F_t$$

where Ft denotes the forecast for time period t, Xt denotes the actual demand observed in period t, and α is a smoothing parameter which can be adjusted between 0 and 1. Higher α will produce a forecast which is more responsive to recent changes in the data, whilst also being less robust to noise. Unfortunately, SES is known to perform poorly in forecasting for ID, since there is an upward bias in the forecast in the period directly after a non-zero demand. Can we develop specific forecasting methods for ID that do better?

This report explores a few different classes of method, and also discusses some error metrics used to evaluate them.

Approach:

<u>Data Description:</u>

The dataset used was provided by an undisclosed industrial partner. It contains 16000 time-series of intermittent demand for unknown items (Aviation parts), with each time-series representing the demand of a distinct item. These time-series are observed in monthly frequency. There are three features in the original data: series number, time, and value.

Aug-16	Sep-16	Oct-16	Nov-16	Dec-16	Jan-17	Feb-17	Mar-17	Apr-17	May-17	Jun-17	Jul-17	Aug-17	Sep-17	Oc
15.00	8.00	31.00	13.00	41.00	0.00	0.00	10.00	0.00	30.00	12.00	30.00	15.00	14.00	
5.00	3.00	6.00	4.00	3.00	10.00	2.00	5.00	6.00	5.00	7.00	8.00	4.00	9.00	
1.00	3.00	4.00	4.00	17.00	0.00	2.00	0.00	11.00	5.00	4.00	5.00	3.00	8.00	
14.00	0.00	2.00	2.00	4.00	7.00	0.00	2.00	9.00	3.00	0.00	2.00	1.00	1.00	
2.00	1.00	0.00	2.00	1.00	2.00	2.00	0.00	2.00	2.00	2.00	6.00	1.00	2.00	
2.00	2.00	0.00	2.00	1.00	4.00	2.00	3.00	2.00	2.00	2.00	6.00	1.00	1.00	
19.00	31.00	21.00	24.00	21.00	15.00	18.00	24.00	12.00	22.00	17.00	12.00	12.00	27.00	
18.00	22.00	18.00	31.00	18.00	21.00	13.00	22.00	7.00	26.00	17.00	6.00	17.00	26.00	
150.00	100.00	200.00	400.00	0.00	30.00	20.00	0.00	110.00	110.00	50.00	80.00	0.00	50.00	
16.00	77.00	116.00	105.00	30.00	233.00	37.00	10.00	27.00	110.00	56.00	230.00	29.00	107.00	
20.00	24.00	12.00	27.00	8.00	15.00	5.00	13.00	38.00	8.00	10.00	5.00	18.00	10.00	
11.00	3.00	19.00	4.00	0.00	10.00	2.00	0.00	41.00	12.00	15.00	15.00	0.00	15.00	
6.00	0.00	15.00	1.00	10.00	6.00	0.00	19.00	10.00	9.00	20.00	5.00	0.00	16.00	
11.00	2.00	3.00	14.00	5.00	21.00	22.00	8.00	0.00	5.00	0.00	10.00	1.00	0.00	
2.00	0.00	1.00	1.00	1.00	0.00	2.00	3.00	4.00	0.00	0.00	5.00	1.00	1.00	
7.00	11.00	10.00	0.00	18.00	4.00	8.00	16.00	8.00	10.00	11.00	16.00	13.00	0.00	
20.00	20.00	20.00	110.00	40.00	0.00	30.00	70.00	40.00	40.00	30.00	120.00	20.00	40.00	
20.00	110.00	150.00	50.00	70.00	90.00	120.00	60.00	180.00	20.00	130.00	80.00	89.00	120.00	
91.00	25.00	80.00	104.00	90.00	68.00	152.00	37.00	88.00	83.00	103.00	92.00	60.00	32.00	
32.00	30.00	35.00	0.00	5.00	16.00	14.00	25.00	5.00	0.00	5.00	5.00	21.00	0.00	

Fig 5.2 Unprocessed Data from the organization

Date preprocessing & Classification:

Data preprocessing is a data mining technique that involves transforming raw data into an understandable format. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues.

To determine a product forecast ability, we apply two coefficients:

- The Average Demand Interval (ADI). It measures the demand regularity in time by computing the average interval between two demands.
- The square of the Coefficient of Variation (CV^2). It measures the variation in quantities.

  Based on these 2 dimensions, the literature classifies the demand profiles into 4 different categories:

  - Smooth demand (ADI < 1.32 and $CV^2 < 0.49$). The demand is very regular in time and in quantity. It is therefore easy to forecast and you won't have trouble reaching a low forecasting error level.

  - Intermittent demand (ADI >= 1.32 and $CV^2 < 0.49$). The demand history shows very little variation in demand quantity but a high variation in the interval between two demands. Though specific forecasting methods tackle intermittent demands, the forecast error margin is considerably higher.

  - Erratic demand (ADI < 1.32 and $CV^2 >= 0.49$). The demand has regular occurrences in time with high quantity variations. Your forecast accuracy remains shaky.

  - Lumpy demand (ADI >= 1.32 and $CV^2 >= 0.49$). The demand is characterized by a large variation in quantity and in time. It is actually impossible to produce a reliable forecast, no matter which forecasting tools you use. This particular type of demand pattern is unforecastable.

No of demands	ADI	Std Dev	Average Population	CV^2	Type of demand	Smoothening
=AL2-AM2	=AL2/AN2	=STDEV(B2:AK2)	=AVERAGE(B2:AK2)	=(AP2/AQ2)^2	=IFERROR(IF(AND(AO2<1.32,AR2<0.49),"	=FORECAST.ETS.SEASON
=AL3-AM3	=AL3/AN3	=STDEV(B3:AK3)	=AVERAGE(B3:AK3)	=(AP3/AQ3)^2	=IFERROR(IF(AND(AO3<1.32,AR3<0.49),"	=FORECAST.ETS.SEASON
=AL4-AM4	=AL4/AN4	=STDEV(B4:AK4)	=AVERAGE(B4:AK4)	=(AP4/AQ4)^2	=IFERROR(IF(AND(AO4<1.32,AR4<0.49),"	=FORECAST.ETS.SEASON
=AL5-AM5	=AL5/AN5	=STDEV(B5:AK5)	=AVERAGE(B5:AK5)	=(AP5/AQ5)^2	=IFERROR(IF(AND(AO5<1.32,AR5<0.49),"	=FORECAST.ETS.SEASON
=AL6-AM6	=AL6/AN6	=STDEV(B6:AK6)	=AVERAGE(B6:AK6)	=(AP6/AQ6)^2	=IFERROR(IF(AND(AO6<1.32,AR6<0.49),"	=FORECAST.ETS.SEASON
=AL7-AM7	=AL7/AN7	=STDEV(B7:AK7)	=AVERAGE(B7:AK7)	=(AP7/AQ7)^2	=IFERROR(IF(AND(AO7<1.32,AR7<0.49),"	=FORECAST.ETS.SEASON
=AL8-AM8	=AL8/AN8	=STDEV(B8:AK8)	=AVERAGE(B8:AK8)	=(AP8/AQ8)^2	=IFERROR(IF(AND(AO8<1.32,AR8<0.49),"	=FORECAST.ETS.SEASON
=AL9-AM9	=AL9/AN9	=STDEV(B9:AK9)	=AVERAGE(B9:AK9)	=(AP9/AQ9)^2	=IFERROR(IF(AND(AO9<1.32,AR9<0.49),"	=FORECAST.ETS.SEASON
=AL10-AM10	=AL10/AN10	=STDEV(B10:AK10)	=AVERAGE(B10:AK10)	=(AP10/AQ10)^2	=IFERROR(IF(AND(AO10<1.32,AR10<0.49	=FORECAST.ETS.SEASON
=AL11-AM11	=AL11/AN11	=STDEV(B11:AK11)	=AVERAGE(B11:AK11)	=(AP11/AQ11)^2	=IFERROR(IF(AND(AO11<1.32,AR11<0.49	=FORECAST.ETS.SEASON
=AL12-AM12	=AL12/AN12	=STDEV(B12:AK12)	=AVERAGE(B12:AK12)	=(AP12/AQ12)^2	=IFERROR(IF(AND(AO12<1.32,AR12<0.49	=FORECAST.ETS.SEASON
=AL13-AM13	=AL13/AN13	=STDEV(B13:AK13)	=AVERAGE(B13:AK13)	=(AP13/AQ13)^2	=IFERROR(IF(AND(AO13<1.32,AR13<0.49	=FORECAST.ETS.SEASON
=AL14-AM14	=AL14/AN14	=STDEV(B14:AK14)	=AVERAGE(B14:AK14)	=(AP14/AQ14)^2	=IFERROR(IF(AND(AO14<1.32,AR14<0.49	=FORECAST.ETS.SEASON
=AL15-AM15	=AL15/AN15	=STDEV(B15:AK15)	=AVERAGE(B15:AK15)	=(AP15/AQ15)^2	=IFERROR(IF(AND(AO15<1.32,AR15<0.49	=FORECAST.ETS.SEASON
=AL16-AM16	=AL16/AN16	=STDEV(B16:AK16)	=AVERAGE(B16:AK16)	=(AP16/AQ16)^2	=IFERROR(IF(AND(AO16<1.32,AR16<0.49	=FORECAST.ETS.SEASON

Apr-19	5/19/20	Jun-19	Jul-19	Cou		No of demand	ADI	Std Dev	Average Populatio	CV^2	Type of demand
780	1180	920	520	38	0	38	1	363.72767	281.4210526	1.670473209	Erratic Demand
740	685	990	5	38	0	38	1	339.16568	346.8684211	0.956080104	Erratic Demand
9	13	14	2	38	0	38	1	17.968698	20.86842105	0.741402493	Erratic Demand
225	275	225	150	38	0	38	1	148.1534	166.7631579	0.789265289	Erratic Demand
450	50	100	125	38	0	38	1	105.54093	112.6315789	0.87805451	Erratic Demand
10	55	21	34	38	0	38	1	21.111032	41	0.265125328	Smooth Demand
295	234	254	162	38	0	38	1	121.84775	224.7894737	0.293821157	Smooth Demand
546	334	364	286	38	0	38	1	207.03348	337.0526316	0.377299059	Smooth Demand
10	6	5	2	38	0	38	1	2.4950876	5.131578947	0.236412033	Smooth Demand
5130	3220	4000	1580	38	0	38	1	1957.9321	1487.657895	1.732164366	Erratic Demand
130	87	113	45	38	0	38	1	50.033559	76.31578947	0.429827297	Smooth Demand
52	60	49	6	38	0	38	1	23.813584	43.68421053	0.297166968	Smooth Demand
1185	660	590	400	38	0	38	1	533.23061	648.7631579	0.675550459	Erratic Demand
625	490	50	200	38	0	38	1	290.19858	343.8421053	0.712315609	Erratic Demand

Fig 5.3 Data Preprocessing

<u>Sequential Data Partitioning:</u>

- Each series was trained on the individual level to capture the unique profile of each item. We used.
- sequential data partitioning to split each series into training and testing sets, with 80% of total.
- observations (starts at the 4th observation) in the training set, and 20% in the

Trained data-80%

Jun-16	1	1	0	0	0	0	0	0	0	Fast Moving
Jul-16	3	3	0.11	1	0.10	0.90	1	0.20	0.90	
Aug-16	4	4	0.39	1	0.39	0.99	1	0.76	0.99	
Sep-16	4	4	0.75	1	0.75	1.00	1	1.41	1.00	
Oct-16	17	6	1.08	1	1.08	1.00	1	1.93	1.00	
Nov-16	0	0	1.57	1	1.57	1.00	1	2.74	1.00	
Nov-18	8	8	4.39	1	4.39	1.00	1	4.73	1.00	
Dec-18	0	0	4.75	1	4.75	1.00	1	5.38	1.00	
Jan-19	4	4	4.75	2	4.75	1.00	2	5.38	1.00	
Feb-19	0	0	4.67	1	4.67	1.00	1	5.11	1.00	4.67
Mar-19	3	3	4.67	2	4.67	1.00	2	5.11	1.00	0.56
Apr-19	1	1	4.50	1	4.50	1.00	1	4.68	1.00	3.50
May-19	2	2	4.15	1	4.15	1.00				
Jun-19	36		3.94	1	3.94	1.00				
Jul-19	4	4	3.55	1	3.55	1.00				
Aug-19		4	3.59	1	3.59	1.00	1	3.08	1.00	

Tested data-20%

Fig 5.4 Data partition (Train & Test data)

Model selection:

Best model (Statistical performance: The statistical performance measures adopted here were Mean Absolute Scaled Error (MASE). MAE is selected because it is easy to interpret and understand, and it treats errors equally. However, it cannot be used to compare across time-series because it is scale dependent. Therefore, the research also utilized MASE to provide a more holistic perspective by comparing accuracy across different time-series). will automatically pickup based on the results or output using meta model, In this case we used markov probability for unbiased results.

- CROSTON
- TSB_CROSTON
- "Hyperbolic Exponential
- Smoothening Model"
- SBA
- Holt's Linear
- "Error trend Seasonality"
- "L-STAR (logistic smooth transition Auto regressive)"
- ARMA

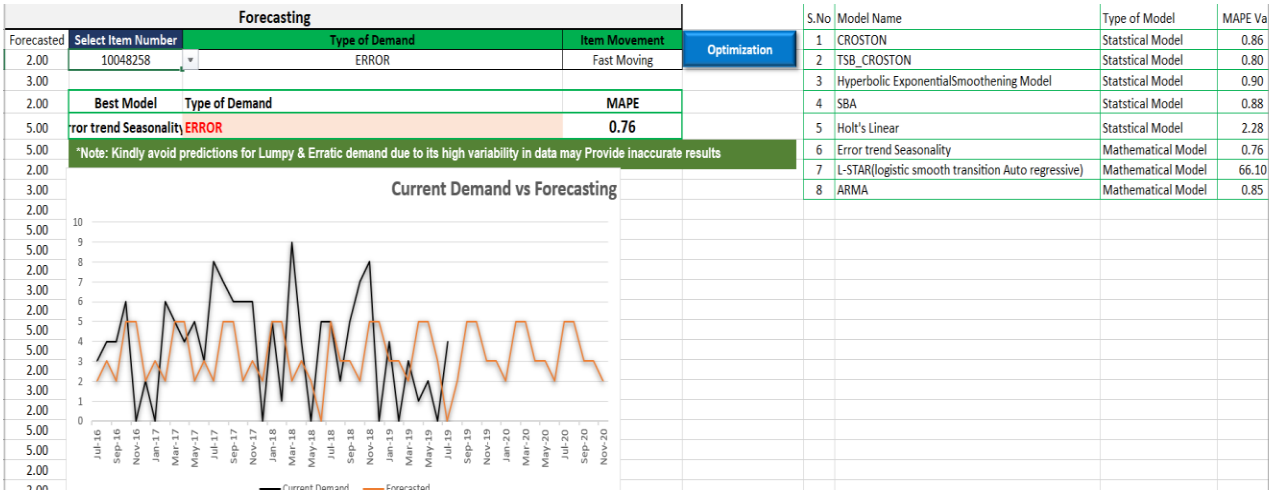
IMPLICATIONS AND LIMITATIONS:

This model will only applicable to intermittent, lumpy, smoothening demand. For erratic demand this model may give wrong results due to its nature (Demand move up against time varied in such a way that model can't predict with accurately)

Forecasted	Select Item Number	Type of Demand	Item Movement
0.20	10048258	ERROR	Fast Moving
0.69			
1.27	Best Model	Type of Demand	MAPE
1.73	SBA	ERROR	0.93
2.47	*Note: Kindly avoid predictions for Lumpy & Erratic demand due to its high variability in data may Provide inaccurate		
2.47			

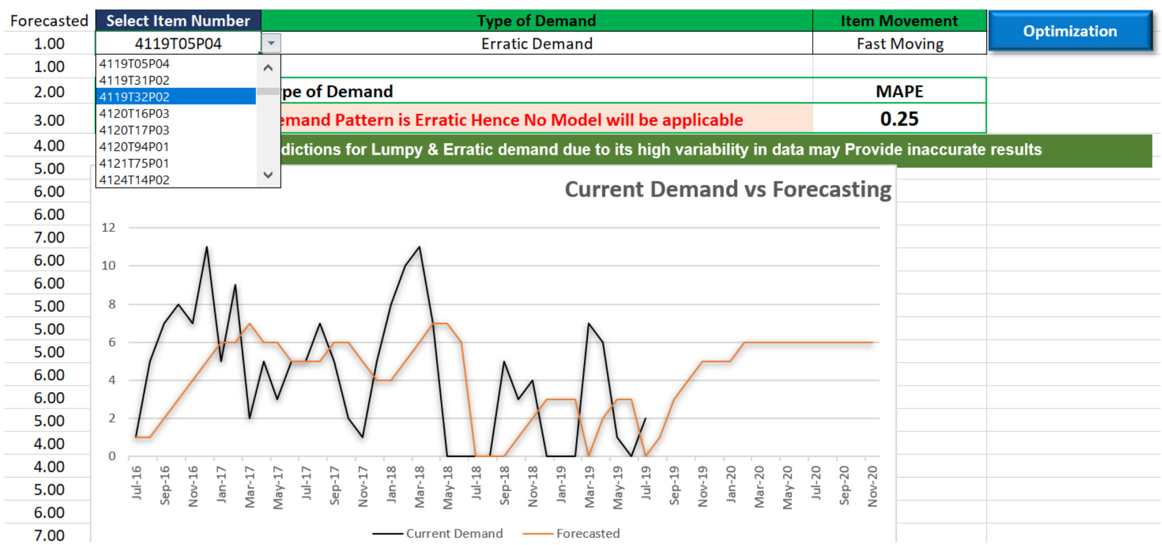
Despite ERROR demand MAPE gives us 0.93. ERROR demand many times throws unmeaningful results.

Step by Step by process to get the final results from the tool:



Step:1

Given image, we can select requires item number from the 'select item number' option.



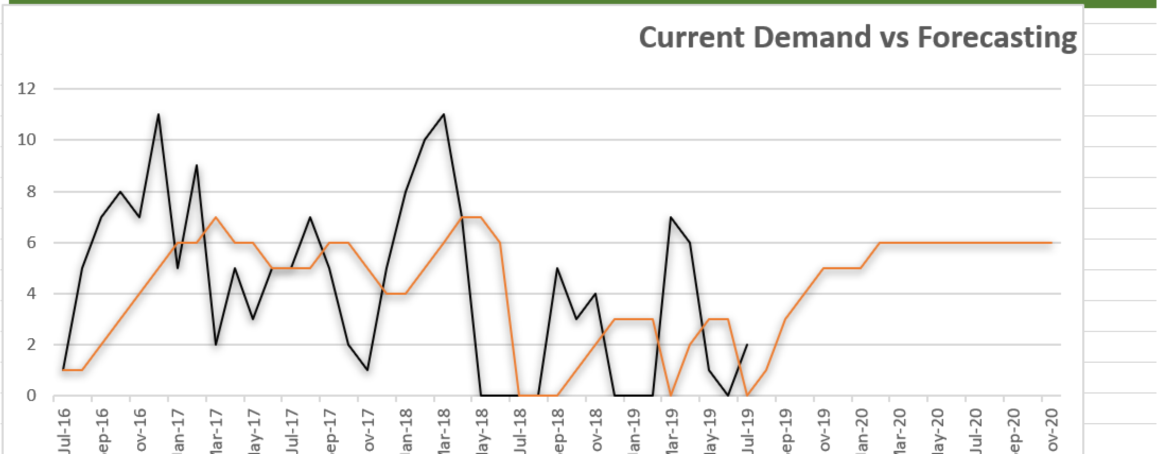
Step:2

We need to check whether selected item number is feasible to forecast or not, after clicking on optimization buttom which ensures to fit the line in the ARMA model. If the type of demand falls under erratic or lumpy, we can't proceed further due to its high irregularity in the pattern. Once we find the type of demand either smooth or intermittent then we can proceed to next step.

Select Item Number	Type of Demand	Item Movemen
4119T05P04	Erratic Demand	Fast Moving

Best Model	Type of Demand	MAPE
ARMA	Demand Pattern is Erratic Hence No Model will be applicable	0.25

***Note: Kindly avoid predictions for Lumpy & Erratic demand due to its high variability in data may Provide inaccu**



Step:3

Analyze other options like best model & item movement (number of non-zero demands in the given duration). We can approve the forecasting based on our MAPE value. More than 60% s highly recommendable. Less than 60 mayn't appropriate to consider for further evaluation.

In this graph test & Train data divided into 75 & 25% ratio. Which means predicting the next 12 months demand by using past 36 months data)

5. CONCLUSION:

The limitations of usual forecasting methods, such as SES, in the intermittent demand case has prompted a number of different ID-specific approaches to forecasting. The first method developed was Croston's; its aptitude for ID forecasting was examined, as well as its drawbacks. Methods that attempted to address some of these drawbacks were considered. More theoretical, model-based approaches were introduced as a promising alternative avenue, as well as methods based in bootstrapping and temporal aggregation. Finally, examples of error metrics that can be used to measure accuracy of forecasts were considered, along with their suitability in the ID setting. We need to realize the type of demand we are working on. If the demand falls under lumpy or erratic, predictions won't be accurate due to it's high variability. The outcome of these results can mislead us.

Instead of best fit model from various models, we should be sure of single model only. When various algorithms involved and unrelated design of each model might effect on the final results.

After doing innumerous trails, I have noticed mathematical model would be superior than statistical models. In statistical models let's say croston or TSB where they considered only few previous values or past data instead of whole, which can again mislead the outcomes of the model.

The biggest everlasting issue in forecasting is unknown or hidden variables which we need to address. These unknown variables can be distinguished into past and furure ones. usually forecasting demand based on single variable only. I.e the amount of demand, but we didn't have any information regarding what causes had got this demand. Lets say if the competitor didn't fulfill the past orders or any internal issues can automatically reflect on the performance of the lead company or any political & technological interferences also causes the fluctuations in demand. So instead if single variable, it would be suggested that consider multiple variables can give better predictions. On the other hand future forecasting need not necessarily get the accurate results due to uncertainty of future. Despite strong analytical & mathematical functions, there can be lot black swans may mislead final end results. So in this context, I would recommend for careful analysis and factors/variables need to consider to mitigate the error rate.

Over all, there are lot of limitations involved in this research. Other researchers suggested deep learning can be better alternative for intermittent demand. It might be promising in few cases, but according to my research I found deep learning had been sending over accurate (over fitting) which highly hazardous in predictions due to a rigid characteristic which further leads to high variation in the testing to future data. So for better results, considering the high number of relatable factors(multi variables) and its weightage of reference might get promising results.

6. REFERENCES:

https://www.lancaster.ac.uk/pg/waller/pdfs/Intermittent_Demand_Forecasting.pdf

https://www.sciencedirect.com/science/article/pii/S0925527318300562

https://onlinelibrary.wiley.com/doi/10.1002/for.2575

https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/

http://www.statsoft.com/textbook/time-series-analysis