A Study on Inventory Optimization

Vaidiyanathan Lalgudi Venkatesan

September 27, 2019

Abstract

This paper presents an academic study about the policies and techniques involved in inventory management. Supply chain challenges, especially the replenishment problem involving optimized, non-optimized approaches, and inventory policies (r, Q) and (s, S) are discussed. The continuous review (r, Q) inventory policy places an order of Q quantities when the inventory level reaches r. On the other hand, the (s, S) policy involves periodic review of inventory: if inventory is found below level s, it is brought back to the level S by placing an order. The paper also discusses the concept of Demand forecasting and techniques such as Exponential Smoothing, rolling forecast using Winter's method (optimized using Excel solver) and ARIMA (in R). Demand on a sample dataset is forecasted and the results are used to calculate the safety stock level for a given Lead time. Finally, a multiperiod linear programming model is defined to optimize material flows and inventory levels in order to reduce out of stock and excess inventory along with reduced costs. The model proposed allow supply chain planning through simultaneous optimization of safety stock levels in tandem with material flows.

1. Introduction

Over the years, Supply chain management has evolved from an initial focus on improving relatively simple, but very labor-intensive processes to the present-day engineering and managing of extraordinarily complex global networks. Inventory management is one of the central problems in retail. Frequently inventory managers need to decide how many items of each product they need to order from suppliers. **A stock out of a fast-moving product leads to a loss in sales whereas**

carrying excess inventory of not so popular products increases holding costs. The typical cost of carrying inventory is at least 10.0 percent of the inventory value. So, the median company spends over 1 percent of revenues carrying inventory, although for some companies the number is much higher [2]. Therefore, managing a company’s inventory becomes critical.

Inventory optimization is a method of balancing capital investment constraints or objectives and service-level goals over a large assortment of stock-keeping units (SKUs) while taking demand and supply volatility into account [3]. The first step of this involves predicting the demand of a product. While stable demand can be predicted, a highly volatile demand causes the **“Bullwhip effect”**, which causes a small change in actual demand to a large change in the perceived demand, and hence major changes in inventory than required.

1.1. Deterministic and Stochastic Inventory Optimization

There are both **optimized** and **non-optimized** approaches to inventory management. The non-optimized approach involves single stage calculations to predict the demand for a particular item in a store, given a number of days of demand. Coming to optimized approaches, it can be either Deterministic or Stochastic. The **deterministic** has every variable state uniquely determined by the parameters, whereas the variables states in a **stochastic** way are defined by probability distributions and hence are able to handle volatile demand in a better way.

1.2. Single and Multi-Echelon Inventory Planning

A Single echelon approach considers each echelon in a supply chain separately and forecasts demand and determines required inventory for that level. Whereas a Multi echelon approach is a more holistic approach, modeling multiple stages of the supply chain to continually update and optimize the safety stock across all echelons.

2. Inventory Policies

The Inventory management policies have different decision parameters that are either calculated based on average demand and the lead times or obtained from historical data. We will discuss two inventory policies, the Continuous review (r, Q) and Periodic review (s, S) policy.

2.1. Continuous review (r, Q) policy

This inventory is continuously reviewed to keep track of the products, under this policy. Whenever the inventory level reaches ***r***, an order of ***Q*** units is placed for a product. This policy is better suited for continuous time models. The condition needs to be changed as ordering when the inventory levels reach less than or equal to ***r***, in case of discrete time planning methods.

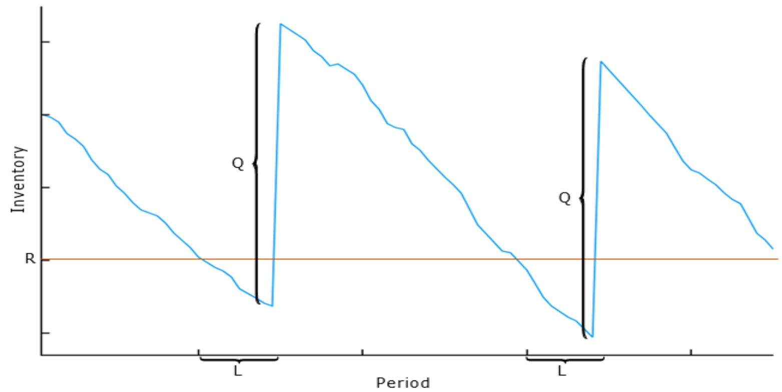

Figure 1 (r,Q) Policy

2.2. Periodic review (s, S) policy

Under this, the inventory is reviewed in a predefined time interval. This comes in places where continuous review of inventory is not possible. During the review, if the inventory levels are identified to be less than a threshold ***s***, orders are placed to bring the inventory to a specified base stock level ***S***. The decision of placing a replenishment order or not, and the amount of the order depend on the inventory level at the moment the inventory is reviewed.

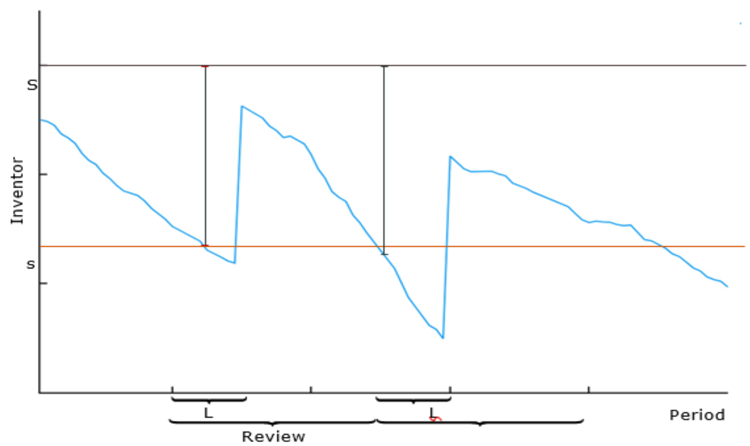

Figure 2 (s,S) Policy

3. Demand Forecasting – Case Study

Demand forecasting plays an important role in Inventory optimization. Having collaborative demand forecasts for an entire supply chain help plan and optimize inventory accordingly. The quality of a forecast depends on the amount of variation. A useful forecast should include more than a single number (mean, forecast error, “probabilistic forecast”). Companies must identify the factors that influence future demand and then ascertain the relationship between these factors and future demand: past demand, lead time of product replenishment, planned advertising or marketing efforts, price discounts, state of the economy, actions that competitors have taken.

Typically, there are 3 types of Forecasts: **Long-term (Qualitative), Medium-term (Causal) and Short-term (Time series)**. The long-term forecasts are usually subjective and involve expert opinions whereas the medium-term forecasts are causal with the demand highly correlated with some environmental factors (e.g. Demand and pricing). The short-term use time series to forecast future demand using past demand. There are also simulation-based forecasts, where customer choices that give rise to demand are imitated and AI/ Machine learning / Ensemble methods.

The observed demand in a forecast can be defined as:

$$\text{Observed demand} = \text{Systematic component} + \text{Random component}$$

Systematic component is the expected value of demand having the levels, trend and seasonality. **Level** refers to the current de-seasonalized demand, **Trend** is the growth or decline in demand, and the **Seasonality** is the predictable seasonal fluctuation in demand. **Random component** is the part of forecast that deviates from systematic part. The difference between forecast and actual demand gives us the **forecast error**. Different demand forecasting techniques are discussed through forecast of a single product demand. The same can be scaled for multiple products, stores and distribution centers.

3.1. Exponential Smoothing

Exponential smoothing technique for forecasting demand includes all past observations with a smoothing parameter **Alpha (α)**. The alpha is chosen by the user. It is used to weigh recent observations much more heavily than very old observations. It follows **no trend and seasonality**.

$$L_0 = \frac{1}{n}\sum_{i=1}^{n} D_i$$

$$L_t = \alpha D_t + (1-\alpha)L_{t-1}$$

$$F_{t+1} = L_t$$

$$F_{t+n} = L_t$$

Where L_t is the estimate of level at time t, D is the actual demand observed, Ft is the forecast of demand for period t. Since this method does not include the seasonality factor, let us explore other models which take them into account.

3.2. Holt-Winter's Model

HW Model was first proposed in the early 1960s and is an extension of exponential smoothing method. All data values in a series contribute to the calculation of the prediction model [4]. This method is used when there is trend and seasonality in the data set.

In general, it can be said that the HW technique is a complex expansion of the exponential smoothing method, since it sums up this approach to manage trend and seasonality.

$$L_p = \frac{\sum_{t=1}^{p} D_t}{p}$$

$$S_i = \frac{D_i}{L_p} \quad \textit{for } i = 1, \ldots, p$$

$$T_p = 0$$

Where level L in period p is the average of first p demands, with the initial trend as zero. Once the initialization step is complete, we use the below formulas to determine estimates of level, trend and seasonality components in each remaining period for which we have the demand data:

Level $: L_t = \alpha\left(\frac{D_t}{S_{t-p}}\right) + (1-\alpha)(L_{t-1} + T_{t-1})$

Trend $: T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1}$

Seasonality: $S_t = \gamma\left(\frac{D_t}{L_t}\right) + (1-\gamma)S_{t-p}$

Forecast $: F_{t+n} = (L_t + nT_t)S_{t+n-p}$

Where the smoothing parameters $0 < \alpha, \beta, \gamma < 1$. We can tune the smoothing parameters for optimized demand estimate.

In this paper, we test our Winter's method for optimizing demand using a data set containing the monthly demand of a product for 12 years. After we calculate the forecast using the above formulas, we can formulate an optimization model to assist with determining smoothing parameters (α, β, γ).

Objective: Minimize error measure (MSE)

Constraints: Smoothing parameters (α, β, γ) must be between 0 and 1

Data: Information for forecasts and actual demand

This optimization problem was solved using Excel solver (Check Excel file attached) and the optimized MSE was 257.11, MAD was 11.91 and the **optimal values of smoothing parameters (α, β, γ) were (0.29,0.05,0.88)**. The demand forecast can be used to identify the safety stock and reorder calculation for a given lead time

3.3. ARIMA model

The **A**uto**R**egressive **I**ntegrated **M**oving **A**verage model is a popular technique used for time series forecasting. The model can be defined as:

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} + e_t$$

Where y'_t is the difference series. The predictors on the right-hand side include both lagged values of y_t and lagged errors. We call this an ARIMA(p, d, q) model, where:

• **p** = order of the autoregressive part, that is deals with the previous values called lags;

• **d** = degree of first differencing involved;

• **q** = order of the moving average part, that deals on shocks or errors;

The Autoregressive models (AR) are long term models, the effect of an observation slowly disappears as time goes on, whereas the Moving average (MA) models are short term models. The integrated part is what makes the data stationary through differencing [5].

In the data set used in this paper, the demand has an **Upward trend** and there is also a **Seasonal component**, as shown in Figure 3a. Therefore, in order to make the time series stationary on both mean and variance, we take the log transformation (to remove the variance) and perform first order differencing (to remove the upward trend). (Figure 3b)

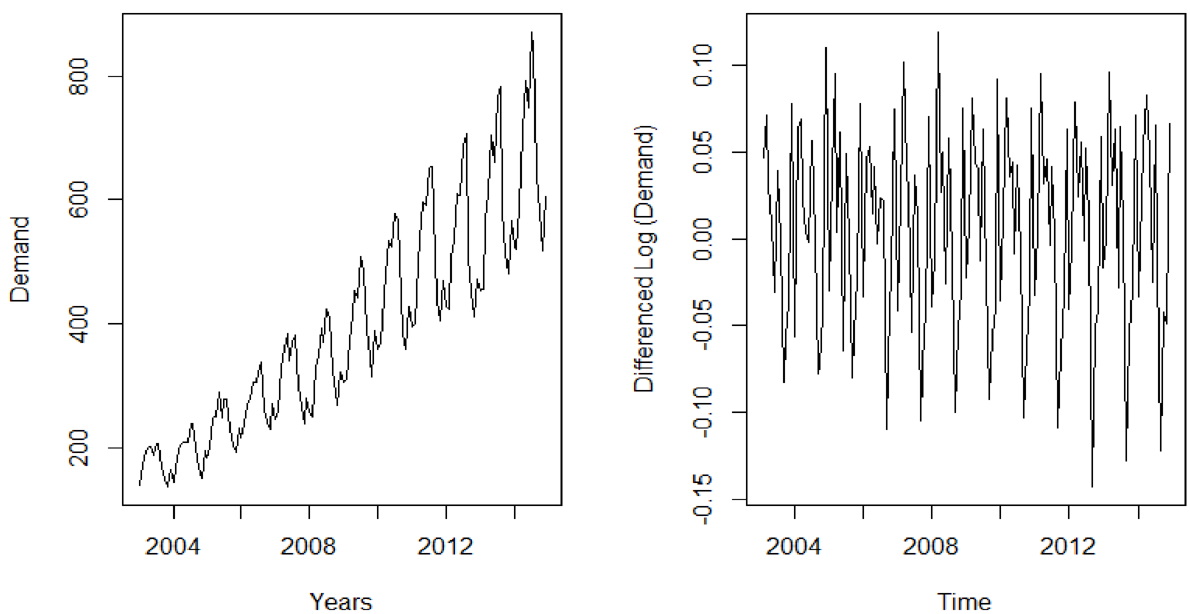

Figure 3(a) Upward trend and seasonal component in demand; (b) demand after taking log and differentiating

Next, the autocorrelation factor (ACF) and partial autocorrelation factor (PACF) are created in order to identify the presence of AR and MA components. We can see that there are spikes outside the insignificant zone and confirm information in residuals for AR and MA models.

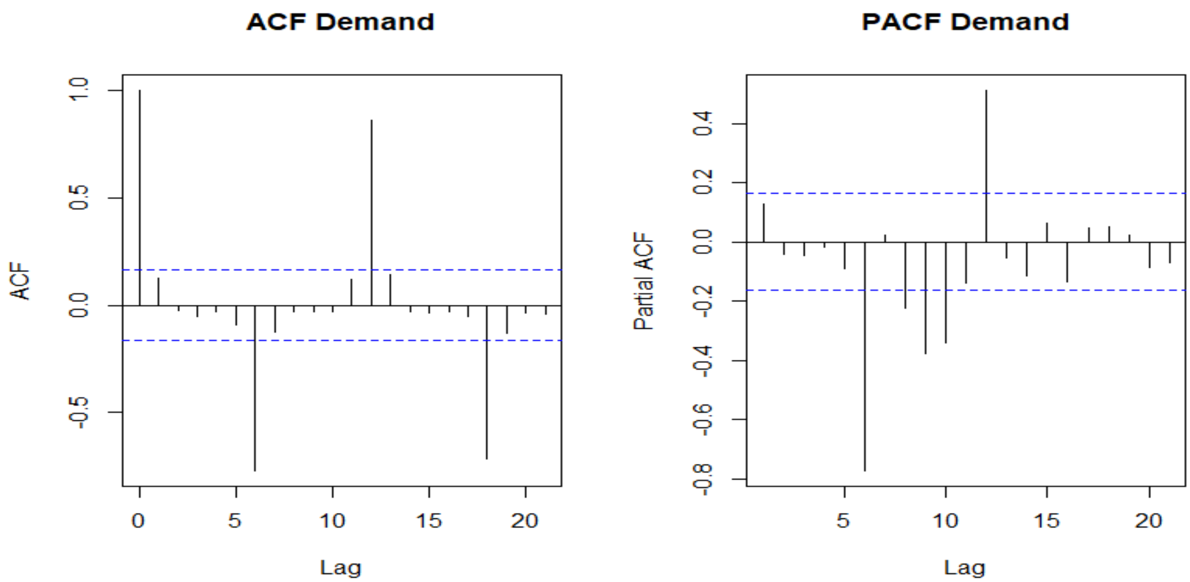

Figure 4 ACF and PACF for demand. Spikes are seen above insignificant line indicating information

The best fit model has Integrated component equal to 1 and MA as 1. There is also a seasonal MA with lag 12 of order 1. Finally, we forecast the demand for next 3 years using our model and this is shown below. The blue line indicates the forecast and red lines, the confidence interval. The ACF and PACF are checked again for residuals to confirm there are no more information to be extracted.

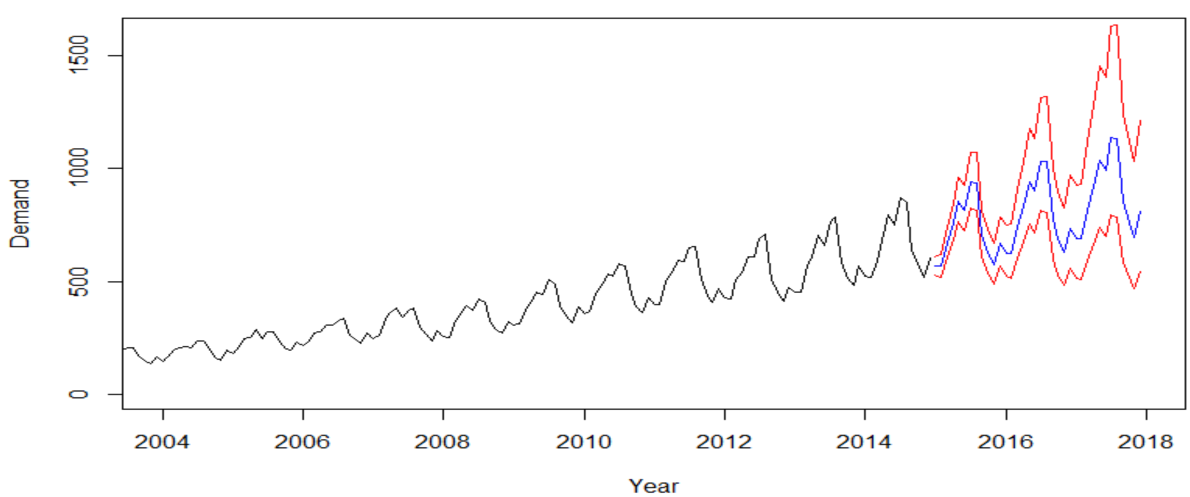

Figure 5 ARIMA model prediction with 95% CI

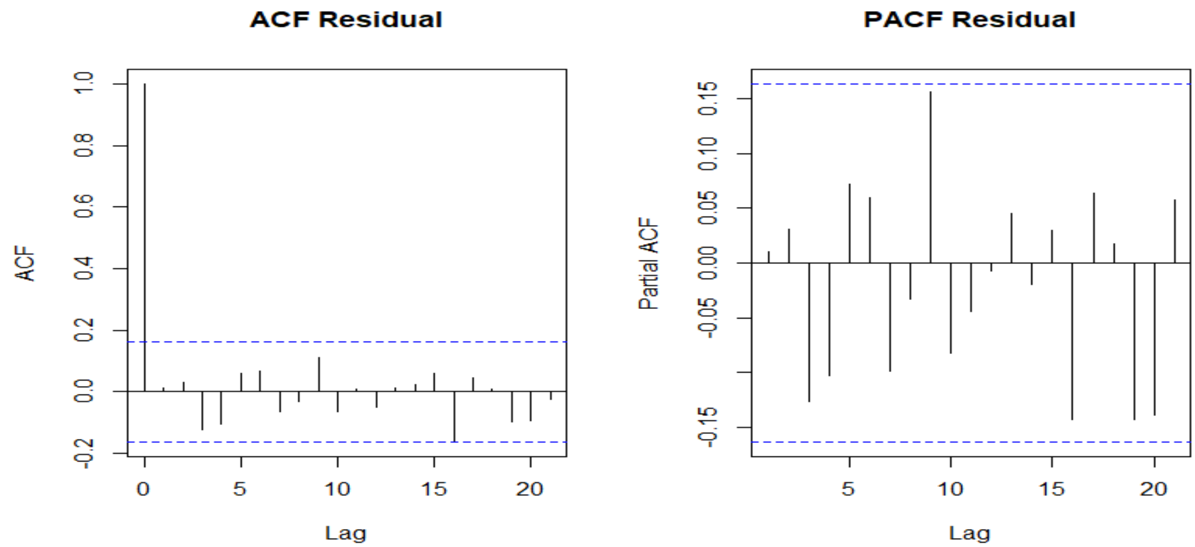

Figure 6 ACF and PACF indicating no more information to extract

3.4. Safety Stock and Reorder point

Once the Demand (D) for a particular time is forecasted, then it is necessary to identify the safety stock (SS), reorder point and quantity for a given lead time. The demand in our data set is approximately normally distributed. Customer service level is the confidence level given for our safety stock, i.e. the business can tolerate stockouts no more than 5% of the replenishment cycles for a 95% service level [6]. This can be calculated using Z-score. For example, to satisfy demand with a 95% confidence interval, according to statistical analysis, it is necessary to carry extra inventory equal to 1.65 standard deviations of demand variability. This is equivalent to a Z-score of 1.65. The safety stock is given by

$$\text{Safety stock} = Z \times \sigma_{LT} \times D_{avg}$$

Where σ_{LT} is the standard deviation of lead time and D_{avg} is the average demand.

For the data considered in this paper, the average demand is 144 units per month, that is 4.8 units per day. Considering a lead time of 7 days with a standard deviation of 1.1 days, the safety stock required for our data with a desired cycle service level of 95 is approximately 9 units.

4. Inventory Optimization Model

Till now, the various inventory policies were discussed along with Demand forecast using a sample data. Now, let us define the inventory optimization model which can be used for a given supply chain network, including suppliers, warehouses, retailers, and a number of customers.

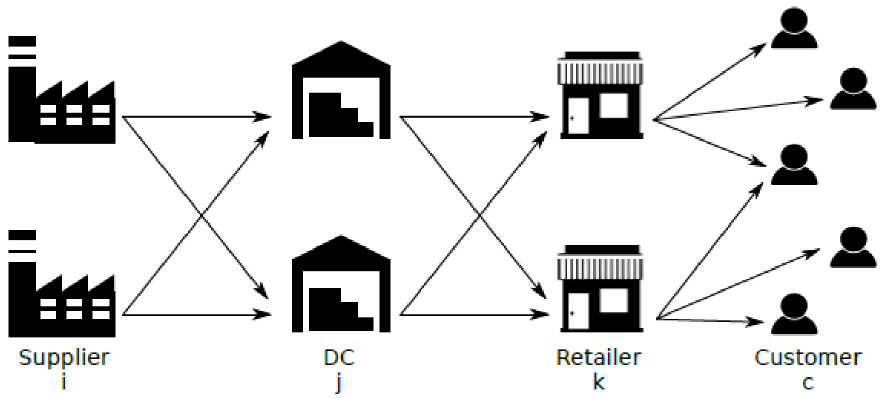

Figure 7 Supply Chain Network Structure

An optimal supply chain plan defines the amount of material transported between facilities at any given time period within the planning horizon. When determining these flows, the inventory levels at the storage facilities are simultaneously determined, because the multiperiod planning models employed include inventory balance constraints [7]. This indicates that given a demand forecast it is possible to determine the exact timing and amount of inventory replenishments. It is required to determine the optimal material flows and inventory levels to satisfy the demand forecast. Hence our objective function is to minimize the transportation and inventory costs.

$$Max \sum_i \sum_j \sum_t TC_{ij} x_{ijt} + \sum_j \sum_k \sum_t TC_{jk} x_{jkt} + \sum_k \sum_c \sum_t TC_{kc} x_{kct}$$

$$+ \sum_j \sum_t HC_j inv_{jt} + \sum_k \sum_t HC_k inv_{kt}$$

Such that

$$\sum_k x_{kct} = D_{ct} \qquad \forall\, k, t$$

$$inv_{jt} = inv_{jt-1} + \sum_i x_{ijt} - \sum_k x_{jkt} - \sum_c x_{jct} \; \forall\, j, t$$

$$inv_{kt} = inv_{kt-1} + \sum_j x_{jkt} - \sum_c x_{kct} \qquad \forall\, i,t$$

$$inv_{j0} = InitInv_j \qquad \forall\, j$$

$$inv_{jT} = InitInv_j \qquad \forall\, j$$

$$x_{ijt}, x_{jkt}, x_{kct} \geq 0, inv_{jt} \leq MaxInvj, inv_{kt} \leq MaxInv_k$$

Where In the model, i is a supplier, j is a distribution center, k is a retailer, c is a customer, and t is the time period. The parameter TC is the transportation cost, HC is the inventory holding cost, MaxInv is the warehouse capacity, and D is the demand. The variables inv and x represent inventory and flow, respectively. The initial inventory and final inventory are set as InitInv to reduce the impact of initial inventory.

The above shows the situation of Just-in-time operation. To give a more practical solution, we need to think about the cost of placing an order as well. For this, if we include binary variables $y_{\mathbf{ijt}}$ and $y_{\mathbf{ijk}}$ to indicate when an order is placed. Now the ordering costs are added as constraints to our model and in order to consider this,

$$x_{ijt} \leq MaxInv_j y_{ijt} \quad \forall\, i,j,t$$

$$x_{jkt} \leq MaxInv_k y_{jkt} \quad \forall\, j,k,t$$

$$y_{jkt}, y_{ijt} \in \{0,1\}$$

$$OrderCost = \sum_t \left(\sum_i \sum_j CO_{ij} y_{ijt} + \sum_j \sum_k CO_{jk} y_{jkt} \right)$$

Where CO is the cost of placing an order. Solving the multiperiod linear programming model defined above with the constraints will give us the parameters to Optimize inventory for different suppliers, warehouses, stores and customers.

5. Conclusions

Supply chain decision making is critical with respect to the inventory management. Various approaches in inventory optimization were discussed such as the Continuous (r,Q) and Periodic review (s,S). Various demand forecasting methods such as Exponential Smoothing, Winter's method and ARIMA were discussed. Exponential Smoothing can be used in case of no trend or seasonality while Winter's method was found to capture the seasonal component as well. Winter's method and ARIMA were used to accurately forecast the demand for a sample data and were tested using Mean Absolute Deviation and Mean Square Error and it was found that the ARIMA model performing better. After the demand was forecasted, the concept of safety stock was introduced and calculated for the average demand and a given lead time. Finally, a multiperiod linear programming optimization model was defined using various cost and flow constraints to optimize inventory across suppliers, warehouses, retailers and customers.

This study takes an important step in understanding the concepts and policies involved in Supply chain inventory management, finally defining a mathematical model bringing us closer to inventory optimization. There are many additional inventory management concepts in demand forecasting and cost optimization, which could be considered in the future develop improved supply chain planning models.

6. Future Work

- Analyzing Inventory optimization segments such as Stage Optimization, Mix Optimization, Lot Size optimization, Prebuild Optimization
- Explore other demand forecasting techniques such as Prophet and Deep Neural Network to optimize inventory
- Compare the performance of Inventory policies (r,Q) and (s,S)
- Study safety stock modeling techniques: proportional to throughput, proportional to throughput with risk-pooling effect, explicit risk-pooling, and guaranteed service time

References

[1] Data: https://github.com/vaidilv/Inventory-Optimization "Demand Data.xlsx"

[2] Marisa Brown, "Inventory Optimization: Show Me the Money," Supply Chain Management Review, July 19, 2011

[3] https://en.wikipedia.org/wiki/Inventory_optimization

[4] Güzin Tirkeş, Cenk Güray, Neş'e Çelebi (2018): Demand Forecasting: a comparison between the Holt winter's, trend analysis and decomposition models

[5] M.W.T. Gemmink; A Model to Prevent Stockouts in Retail using Time Series Sales Forecasting

[6] https://web.mit.edu/2.810/www/files/readings/King_SafetyStock.pdf

[7] Bradley and Arntzen, 1999: The Simultaneous Planning of Production, Capacity, and Inventory in Seasonal Demand Environments

[8] Braulio Brunaud, Jose M. Lainez-Aguirre, Jose M. Pinto, and Ignacio E. Grossmann (2018): Inventory Policies and Safety Stock Optimization for Supply Chain Planning

[9] Machine learning for inventory optimization: https://medium.com/opex-analytics/machine-learning-for-inventory-optimization-38a9ac86a80a

[10] Deep learning approach to inventory optimization: https://www.scnsoft.com/blog/inventory-optimization-with-data-science

[11] Afshin Oroojlooyjadid, Lawrence Snyder, Martin Takáč (2016): Applying Deep Learning to the Newsvendor Problem