INFORMS 2019
A Contest fORged by Amazon Web Services and Computational OR Exchange

Order-Up-To Policy for Stochastic Time Series Demand

Srinivasan Balan *
Atchyuta Bharadwaj Manda †

October 12^{th}, 2019

Advisor
Dr. Reha Uzsoy

Edward P. Fitts Department of Industrial and Systems Engineering
NC State University, Raleigh

NC STATE UNIVERSITY

*sbalan@ncsu.edu
†amanda@ncsu.edu

Contents

List of Figures

List of Tables

1 Introduction

In any supply chain, there is considerable optimization scope to find the optimal order quantity, optimal shipment size, minimum cost, the best route, and optimal service level. Hi-tech industries for hardware components are interesting in terms of uncertainties and stochastic events occurrence in day to day operations which result in demand patterns that exhibit trend, seasonality and cyclical behaviours. There are many methods available to work with time series data. Many authors have presented various statistical and machine learning techniques to predict the future demand. On the other hand, inventory optimization (IO) is a well-known topic in supply chain optimization which has drawn the attention of many researchers over the last six decades. Many attempts have been made in the past to find the exact solution using Markov Decision Processes, Stochastic Dynamic Programming and Mixed Integer Linear Programming based models. All these methods work on the basis of expectation of average cost criterion by keeping demand in the form of a probability distribution. These methods do not work efficiently when the demand is correlated and seasonal in the form of a time series. In the current contest, we provided an order upto level policy for the stochastic time series monthly demand. We used Auto Regressive Integrated Moving Average (ARIMA) modeling using R to find out the demand prediction for the years 2006 and 2007. We used the forecast errors as the basis for finding the safety stock norms to add to the mean forecasted demand. Thus an order up to level will be calculated based on the opening inventory levels for Jan 2006. We also developed a Genetic Algorithm based simulation optimization model to witness the optimal order upto policy for our solution in MATLAB. A comparison of order upto levels is shared at the end to illustrate the basestock levels and simulation optimization levels by using the demand forecast and its standard error.

The remainder of this report is organized as follows: Section 2 gives an overview of the problem. A short review of the relevant literature is provided in Section 3. Section 4 explains our approach to predict stochastic parameters. The solution approach and simulation optimization are discussed in Section 5. Section 6 provides the directions to check the solution for the contest. Eventually, we provide a summary of the report in section 7.

2 Problem Description

In this competition, we were given a problem statement pertaining to monthly demand of a hardware device from 1995 to 2005. We are asked to set up an ongoing inventory system starting from Jan 2006. We are given with costs such as inventory holding costs (h) and backorder costs (b) for any excess inventory or backorder inventory levels at the end of each period. Holding costs are charged at \$1 for inventory less than 90 units and \$2 for inventory exceeding 90 units per month. If the demand is not satisfied, then we incur a backorder costs of \$3 per piece. A review period is considered in the inventory system. The goal is to determine the optimal inventory policy using a data driven analytical model and validation which will be used to test the actual demand from Jan 2006 to Dec 2007. The opening stock for Jan 1995 is 60 units and the opening stock for Jan 2006 is 73 units. We observed that the demand pattern is highly volatile during 2000 and a sharp dip in 2001. We know this may be due to the dot-com bubble for hardware devices. Hence this input was considered to be critical and should be taken in the prediction model.

3 Literature Review

There are many methods available to predict the time series data based on the historical trends, seasonality and correlation components. Auto Regressive Integrated Moving Average methods are common to model such time series data and predict based on the causal relationships of demand patterns with respect to time [2]. We can convert the non-stationary data into stationary data by decomposing the data and then fit ARIMA models to find the best fit.This method handle issues with seasonality, cross sectional data, introducing lags and transformations to convert to a stationary distribution [3] [2]. Then, it will be easy to predict the demand for the future period. But if the time series can be looked as a sequential data, then there are several methods available to predict demand. First, Hidden Markov models are very popular for several years. Second, If the data is in the form of temporal data, recurrent neural networks will be effective. Third, a complicated method known as the Gaussian processes can also be effective in time series modeling. A time series can be decomposed into a standard regression problem when the current predicted value depends upon a block of historical values [4].

In our competition, the problem described and data provided shows that the historical demand data exhibits trend, seasonality and correlated demand pattern. This helped us to use ARIMA models to check the feasibility and predict the future demand data based on the best fit parameters [1].

4 Development of Stochastic Prediction

In this section we describe our approach to predict stochastic parameters in the demand prediction using time series forecasting. We used the following steps in predicting the future demand for 2006 and 2007.

Table 1: Steps for Demand Prediction

S.No.	Steps for time series demand prediction
1	Identify through a visual inspection whether the data has seasonality or trends
2	Identify whether the decomposition technique required is additive or multiplicative
	Log transform the multiplicative if needed
3	Test appropriate additive algorithm
	a. Simple Moving Average
	b. Seasonal Adjustment
	c. Simple Exponential Smoothing
	d. Holts Exponential Smoothing
	e. Holts Winters Exponential Smoothing
	f. ARIMA
4	Perform statistical tests to verify correct model selected
	Box-Ljung test with forecast errors follows Normal distribution with mean 0 and variance σ^2
	Use Auto correlation function and Partial auto correlation function
	to determine the p,d,q parameters for ARIMA models.
5	Repeat steps 3 and 4 if necessary

When we plotted the given data, we observed that the time series possess trend, seasonality, cyclicity and correlations among each periods. We know that if the data follows all three patterns such as trend, seasonality and correlations, then ARIMA models will be helpful in finding the prediction. By considering the difference of lag 1, we incorporated the dot-com bubble outlier into our ARIMA model which helps to take this into consideration.

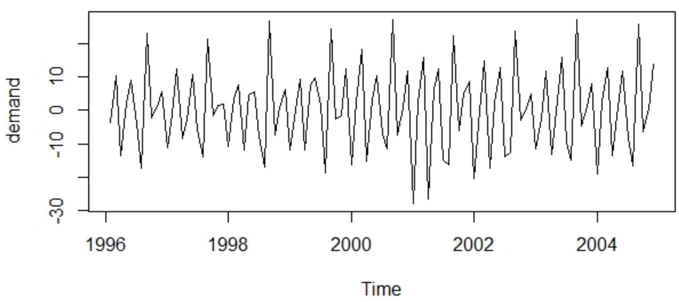
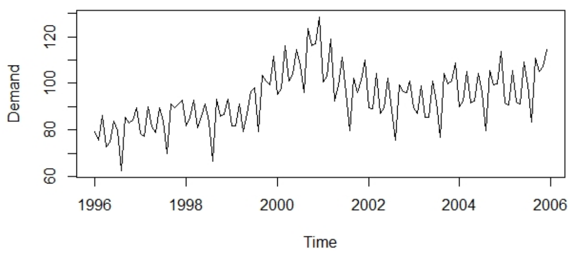


Figure 1: Seasonal difference of 1

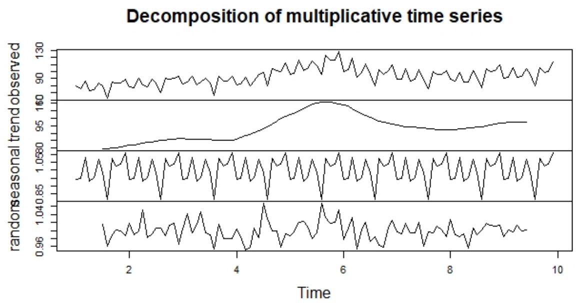
Table 1 shows the steps taken for time series demand prediction during statistical learning. Table 2 gives the different algorithms used to model various combinations of trend, seasonality and correlations. It is evident that the selection of the model is determined by the three parameters namely trend, seasonality and cyclicity. If all the three parameters are present in the data set, ARIMA(p,d,q) models will fit the data and provide good results.

Table 2: Appropriate test algorithms

Types of algorithms	Seasonal	Trend	Correlations
Simple Moving Average	no	no	no
Seasonal adjustment	yes	yes	no
Simple Exponential Smoothing	no	yes	no
Multiple Regression	no	yes	no
Holts Exponential Smoothing	no	yes	no
Holt-Winters Exponential Smoothing	yes	yes	no
ARIMA	yes	yes	yes



(a) Given demand data

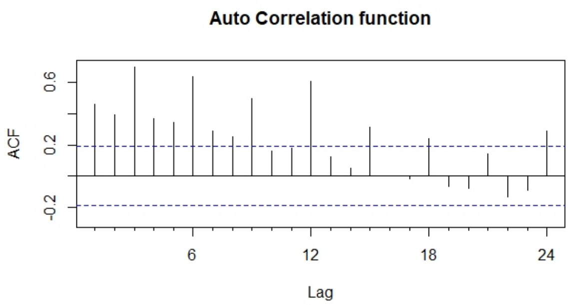


(b) Decomposition of demand

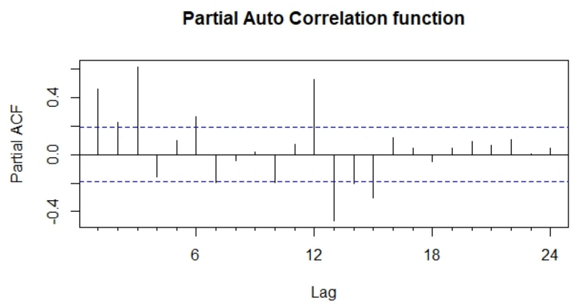
Figure 2: Demand of hardware device 1996-2005

4.1 ARIMA Model for Demand

The given time series data was plotted in R and shown in Figure 2a. We decomposed the data to visualise the trend, seasonality, cyclicity and random components which is shown in 2b. We divided the given data into training set and testing set. Training set consists of data from 1995 to 2004 (108 data points). Testing set consists of 12 data points from 2005 data (12 data points). Using auto correlation function (ACF) and partial auto correlation function (PACF), we developed various ARIMA models to find out the model parameters and estimated significance test using Box-Ljung test. Figure 3a and 3b show the ACF and PACF functions for maximum lags upto 25. Using ACF and PACF, different ARIMA models were tested using forecast R package. We tried to fit auto.arima to find the best fit based on AIC and BIC values. We also tried to find the best fit models based on AR and MA terms from ACF and PACF plots.



(a) ACF



(b) PACF

Figure 3: Correlation, Co-variance and Partial analysis

We created different ARIMA models by varying the auto regressive terms (p), difference terms (d) and moving average terms (q). The best fit model will have the highest p-value from Box-Ljung test when we compare to 95% confidence levels. Here we assumed $\alpha = 5\%$ which is a standard value of risk. We used MSE and MAPE as error measurement to test the forecast values for the test data set. The results are tabulated in Table 3.

The Ljung Box test is described as follows:

H_0: The data are independently distributed. Correlations in the population taken from the samples are zero.

H_1: The data are not independently distributed, there exhibits a serial correlation among demand points.

Table 3: Comparison of error terms and Box-Ljung test

Model #	ARIMA (p,d,q) (p,d,q)[lag]	MAPE	MSE	RMSE	Ljung-Box test χ^2	P-value ($\alpha = 5\%$)
Model 1	(2,1,4)(0,1,1)[12]	2.08%	7.02	2.65	5.56	1.84%
Model 2	(1,1,2)(2,1,1)[12]	2.53%	8.35	2.89	18.89	0.00%
Model 3	(2,0,1)(1,0,1)[12]	3.31%	12.84	3.58	39.39	0.00%
Model 4	(11,1,1)	5.90%	41.59	6.45	26.53	0.00%
Model 5	(0,1,2) Fourier K=5	2.40%	9.08	3.01	0.10	75.02%
Model 6	(1,1,2) Fourier K=4	2.87%	11.28	3.36	0.36	54.86%
Model 7	(2,1,3) Fourier K=3	3.26%	14.54	3.81	14.76	0.01%
Model 8	(3,1,0) Fourier K=6	2.26%	7.42	2.72	0.05	82.77%

For significance level α, the critical region for rejection of the hypothesis of randomness is $Q > \chi^2_{1-\alpha,h}$, where Q is the test statistic. If $p < 0.05$: we can reject the null hypothesis assuming a 5% chance of making a mistake. So we can assume that the values are showing dependence on each other. But If $p > 0.05$: we don't have enough statistical evidence to reject the null hypothesis. So we cannot assume that the values are dependent. This could mean that the values are dependent anyway or it can mean that the values are independent. This shows the white noise component to show that the residuals are uncorrelated (Acf=0) and the residuals are identical, independent and normally distributed with mean 0 and variance σ^2. From the Table 3, it is clear that model 8 is the best fit based on MAPE term and it gives a p-value above 82%. Therefore, we used this model for predicting the forecast for 2006 and 2007.

4.2 Model Validation

We used these models for the testing data set to test the mean square error (MSE) and root mean square error (RMSE) as the parameters for validation. The best fit model with the highest p-value from Ljung test and low MSE is selected to predict the demand for 2006 and 2007 years respectively.

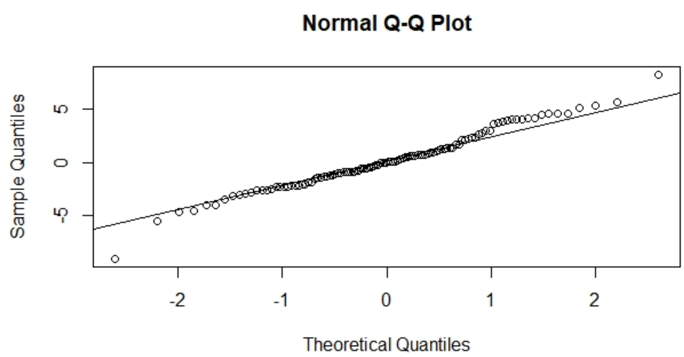


Figure 4: Normal Q-Q plot

Residuals follow Normal distribution with mean 0.17 and variance 7.451. The goodness of fit statistics using Anderson-Darling statistic shows 0.6431 as p value is greater than 0.05 significance level. Therefore, the residuals follow normal distribution.

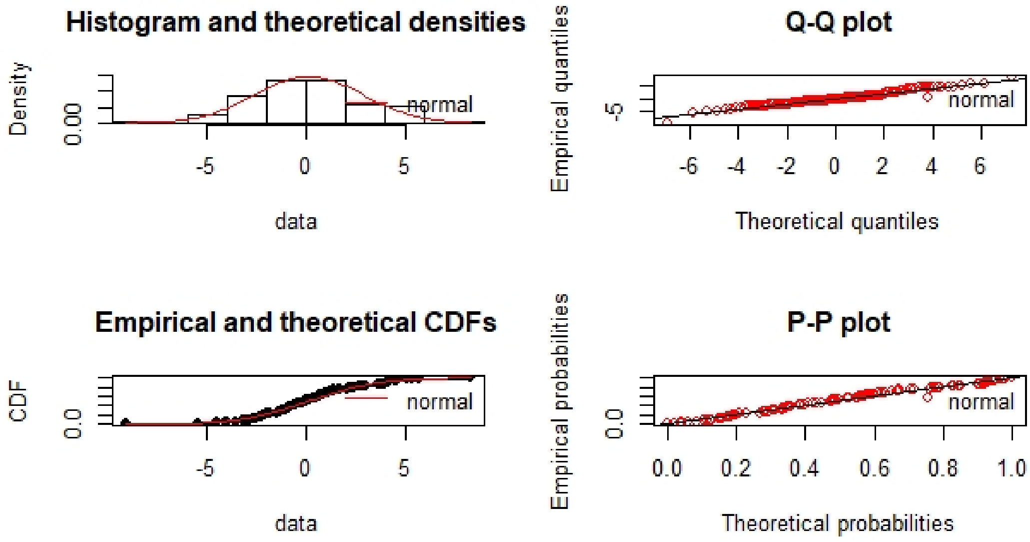


Figure 5: Normally distributed residuals

The best fit model for predicting demand for 2006 and 2007 is given by ARIMA(3,1,0) with Fourier K=6 having RMSE of 2.68 with mean 0.179 and variance 7.2307. The model coefficients are ar1: -0.3816, ar2: -0.1514, ar3: 0.3710, S1-12: -0.7138, C1-12: 3.3623, S2-12:-3.4239, C2-12: 2.0264, S3-12: -2.0717, C3-12: -2.832, S4-12: 1.7888, C4-12: 8.3960, S5-12: -1.3349, C5-12: 1.9632, C6-12: -1.3118. R code is shared in the Appendix to get all the forecast values for each period from Jan 2006 to Dec 2007. The predicted forecasts for the years 2006 and 2007 are shown in the Table 4 and 5.

Table 4: Forecasts for the year 2006

Months	Jan-06	Feb-06	Mar-06	Apr-06	May-06	Jun-06	Jul-06	Aug-06	Sep-06	Oct-06	Nov-06	Dec-06
Forecast	97.18	98.03	110.39	95.58	98.14	109.28	101.25	85.76	110.33	105.65	106.63	114.29

Table 5: Forecasts for the year 2007

Month	Jan-07	Feb-07	Mar-07	Apr-07	May-07	Jun-07	Jul-07	Aug-07	Sep-07	Oct-07	Nov-07	Dec-07
Forecast	97.55	97.74	110.40	95.76	97.97	109.32	101.32	85.66	110.37	105.67	106.57	114.32

5 Order Upto Policy

Having forecasted the demand for the next two years, the goal now is to determine the optimal order quantities. To determine these quantities we begin by using the demand forecast for period t as the mean demand ($\hat{\mu}_t$) for that period. The standard deviation of the forecast error ($\hat{\sigma}_{e,t}$) is used to capture the uncertainty in the demand forecast. This procedure is in line with the literature [6]. We can use these demand parameters to figure out the optimal base stock level in each period t. The idea here is that these base stock levels in each period will help the firm in managing the inventory system as long as the demand is inline with the forecasts. If the inventory position at the end of period t is x, then the firm simply needs to place an order to bring its inventory position equal to the base stock level (order up level) of the next period. Synder and Shen (2011), showed that a base stock policy is optimal in every period of a finite-horizon model [6].

The optimal base stock level (S^*) for a single period problem is given by:

$$S^* = \hat{\mu}_t + \hat{\sigma}_{e,t} * \Phi^{-1}(\frac{b}{b+h}) \tag{1}$$

where b is the back order cost, h is the holding cost and Φ^{-1} is the commutative probability density function.

However, due the presence of variable inventory costs we resort to simulation optimization as both a more flexible approach and a validation tool.

5.1 Simulation Model

To begin the procedure, we set up the simulation for a given set of base stock quantities for each period t (S_t). In a particular simulation run, we begin in period one (Jan 2006) with the ending inventory from previous period (given). The demand for this period is randomly generated with mean given by the sum of forecasted demand for period 1 ($\hat{\mu}_t$) and mean of the forecast errors and standard deviation from the forecast errors. As mentioned in the contest guidelines, we assume the entire demand for period t is known at the beginning of the period. using this information we calculate the inventory position after the demand is met. An order is then placed to bring the inventory position to base stock level of period t. We also calculate any back orders or finished goods inventory and the associated costs.

The order placed at the beginning of the first period arrives at the beginning of the second period and becomes part of the starting inventory which is used to meet the demand of the second period. The procedure mentioned above is then repeated starting with demand generation followed by order quantity calculation and finished goods inventory and back order calculation along with the associated costs for that period.

At the end of the horizon, the simulation returns the total cost. The above simulation model is thus capable of returning the total cost for a given set of base stock values. Since there is variance in each simulation run due to the randomly generated demand, for a given vector of base stocks we perform the simulation a 5000 times and take the Monte Carlo estimate of the total cost. The number of replications was chosen based on experimentation to balance simulation run time and reduce variance in the total cost values. The 95% confidence interval is given in the result of the simulation optimization model.

5.2 Optimization Procedure

The simulation optimization procedure requires the use of a optimization algorithm and we use a Genetic Algorithm (GA) for our simulation optimization engine in this work due to its simplicity of implementation. The ga function of the MATLAB Global Optimization Toolbox [5] was used to implement the genetic algorithm. The total population size was set to 100. To increase the convergence of the GA, we seed the initial population of base stock vectors with the demand forecast enabling the GA to have a good starting point. The elite count was set to 10 and the crossover fraction was set to 0.6. This means that of the 90 remaining individuals, on average 0.6*90 = 54 are from crossover children and the remaining 36 are generated through mutation. The mutation, crossover, migration, scale and shrink options were left at their default values. The default mutation function, `Gaussian`, and the default crossover function, `Scattered`, were used for this implementation. We refer the reader to the Matlab global optimization toolbox documentation for a detailed explanation of the rest of the default functions and values of the Matlab ga parameters.

Table 6: Parameter values

Parameters	End Inv (Dec 2004)	Back order cost	holding cost (< 90)	holding cost ($>= 90$)
Values	73	3	1	2

The use of the simulation optimization using a meta heuristic like GA gives us a flexible optimization framework. For example, the inventory holding cost is mentioned as $1 per unit per period until the ending inventory exceeds 90 units and $2 per period per units after that. Such conditions can be easily incorporated into the simulation function using this approach.

Table 7: Comparison Between GA and Theoretical Base Stock Solution.

	Theoretical Base Stock Solution	GA Solution
Average of 5000 Runs	151.69	152.06
95% Confidence Interval	[151.26 152.14]	[151.63 152.48]

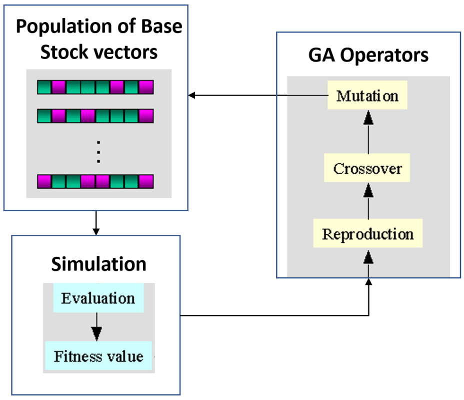


Figure 6: Simulation Optimization Workflow

5.3 Results

The simulation optimization procedure described above was implemented using the cost and inventory parameters given in Table 6. The final GA solution was compared with the theoretical base stock values obtained by using Equation 1. Even though the inventory holding cost changes based on the inventory held, this quantity rarely exceeds the given limit of 90 units. This means that a holding cost $1 per unit per period can be used in Equation 1 to find the theoretical base stock values for each period. Figure 7 shows the comparison between the optimal answer provided by the GA and the theoretical optimal base stock values for each period. Table 7 shows total cost comparison between the two methods. Both these comparisons show very good agreement between the two methods. Multiple GA runs were performed and the solutions were all found to be in agreement with one another.

Thus following this simulation optimization procedure, we obtain the base stock levels for each period. The next section provides details to find the beginning inventory, order quantity, ending inventory, holding cost, and back order cost.

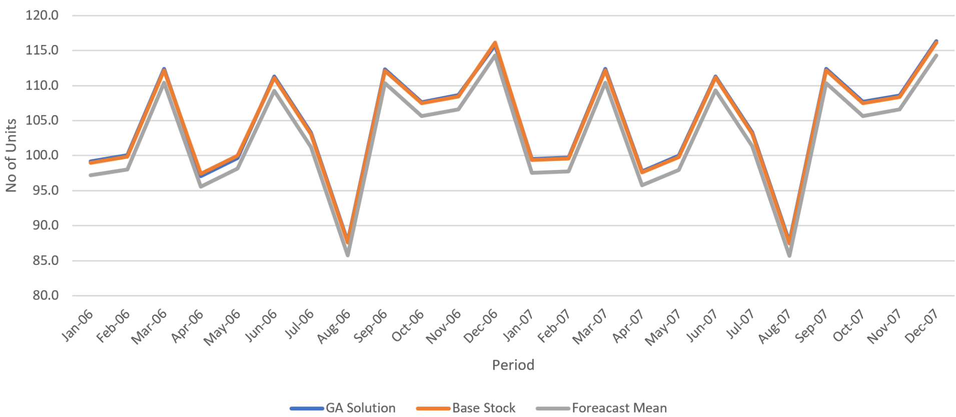


Figure 7: GA Solution vs Theoretical Base Stock Solution.

6 Read Me

By following the procedure detailed in this report we can first forecast the demand data and then obtain the optimal base stock values. The R-script named 'RScriptDemandForecasting.R' is used to predict the demand for 2006 and 2007. Each line of code consists of comments and R-packages to be installed. Training data 'train.csv' and testing data is divided and stored separately in the set directory path. There are clipboard "copy paste" options provided to feed input data directly to R-script. The output forecast and residuals are printed in .csv file with names 'residual.csv' and 'forecast.csv'.

The simulation optimization procedure can be implemented using MATLAB software with the simulation file,'SIMFUN2.mat', and the actual GA being implemented using the file, 'SIMOPT1.m'.

After obtaining optimal base stock values, the firm can use a simple Excel sheet ('InventoryManagement.xlsx') to determine the order qualities, opening and closing inventory and costs. The actual demand values need to be inputted into the sheet in column 'D' of the sheet starting in cell 'D4' for Jan 2005. This base stock values from simulation optimization procedure need to be inputted into 'F' starting in 'F4'. This will ensure that the beginning inventory, order quantity, ending inventory position will be populated in columns 'C','H' and 'I' respectively. The inventory holding cost and the back orders costs are also calculated starting in L32:L36. For the purposes of this two year (Jan 2005 to Dec 2006) window, the base stock values have already been populated in the excel sheet. All the fields mentioned in this section can be found in the GitHub repository named, 'INFORMS-2019-Inventory-Contest'.

7 Summary

We approached the problem using ARIMA models to predict the demand forecast for the years 2006 and 2007 by dividing training and testing set. We incorporated Fourier based ARIMA model for predicting demand which have pretty good p-value from Ljung-box test and low MAPE value (2.26%). The optimal order upto level will take the forecasted value as the mean order quantity and safety stock from the forecast error. Using the forecast from ARIMA model and the standard error from residuals, we determined the optimal order upto level for the future period from Jan 2006 to Dec 2007. To verify the order upto level and check how best we can optimize the order quantity based on stochastic error component, we developed a GA based simulation optimization model to find out the optimal order upto level. We observed that the simulation optimization results are in agreement with the order upto levels which we determined from the basestock calculation. A comparison of the order quantity based on basestock and simulation optimization was shown. The total cost, average inventory holding and backorder costs were calculated in the Excel sheet. Both R code and MATLAB codes are added to the Github repository.

References

[1] George EP Box and David R Cox. An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243, 1964.

[2] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[3] T Hastie, G James, D Witten, and R Tibshirani. An introduction to statistical learning springer. *New York*, 2013.

[4] Rob J Hyndman, Yeasmin Khandakar, et al. *Automatic time series for forecasting: the forecast package for R*. Number 6/07. Monash University, Department of Econometrics and Business Statistics ..., 2007.

[5] MATLAB. *Global Optimization Toolbox*. MATLAB, Natick, Massachusetts, r2016a edition, 2016. `https://www.mathworks.com/help/gads/`.

[6] Lawrence V Snyder and Zuo-Jun Max Shen. *Fundamentals of supply chain theory*. Wiley Online Library, 2011.