fORged by Machines

Date of Submission: 10/10/2019

Priyadarshan Patil (Priyadarshan@utexas.edu)
Endorsing Faculty: Erhan Kutanoglu (erhank@mail.utexas.edu)

Problem Statement

Provided the monthly demand data for a hardware component from 1996-2005, design a mathematical model to predict the demand and come up with an inventory scheme to minimize costs.

Exploratory Data Analysis

Figure 1 below shows the data plotted per year, with months on the X-axis. The most notable observation here is the seasonal nature of the demand. This is also corroborated by Figure 2, showing the % change in demand from each month. This mandated the inclusion of these two factors for a demand prediction model. While interpretations might be drawn for the reasons for the seasonal behavior, without further details, such an effort would be futile, and is hence omitted. Another observation is that demand peaked around 2000-2001, and the demand curves for other years lie below this period.

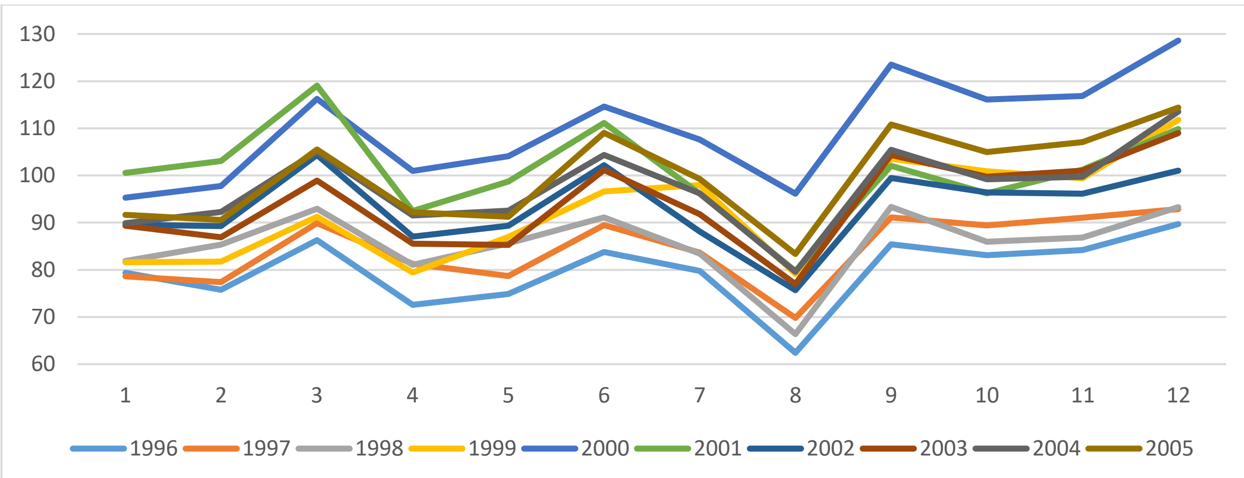

Figure 1: Variation of monthly demand with month number

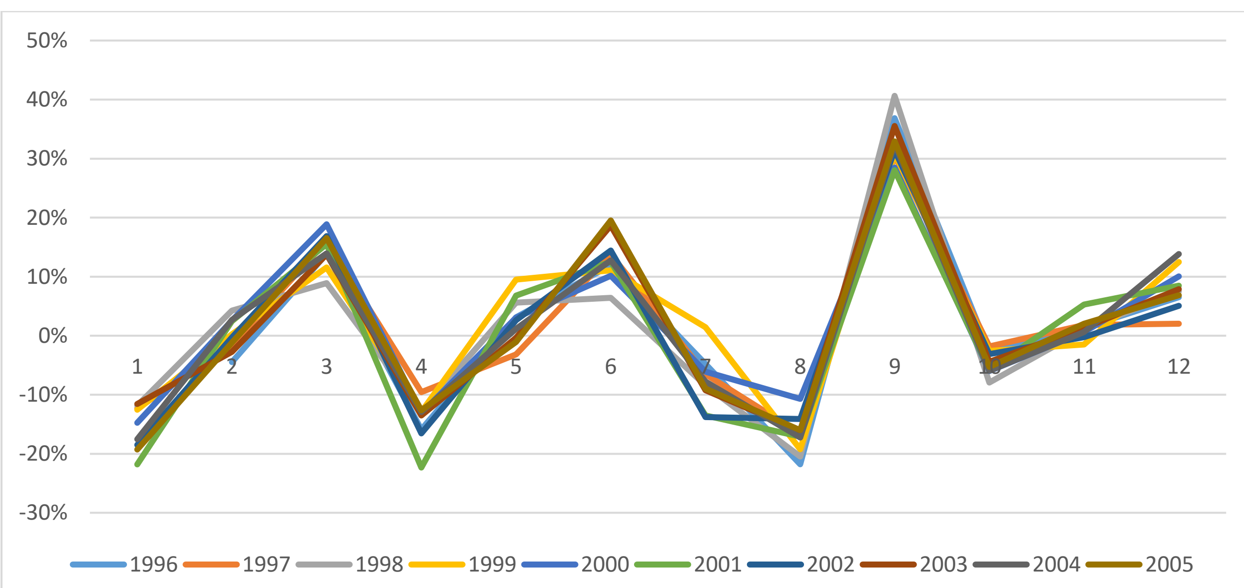

Figure 2: Percent change in demand from previous month

Additional data exploration was performed by manipulating the data exponentially, logarithmically and in polynomial manner, but no meaningful insights were drawn.

Mathematical Model for Demand Estimation

Keeping replicability, reproducibility, and best data practices in mind, we chose the regression model shown below in Equation (1). The coefficients for the regression equation are provided in Table 1 below, along with the t-statistics. This model allows for coefficient interpretation easily. From the first estimation, March and July coefficients could not be said to be statistically significantly different from zero, so they were dropped, and December has been dropped to prevent multicollinearity.

$$Demand_{Month} = c_0 + Demand_{PriorMonth} * (c_1 + c_2 * Jan + c_3 * Feb + c_4 * Apr + c_5 * May + c_6 * Jun + c_7 * Aug + c_8 * Sep + c_9 * Oct + c_{10} * Nov) + \epsilon \tag{1}$$

where Jan, Feb, etc. are binary indicator variables which take the value 1 if the previous month is that specific month and 0 otherwise.

Name	*Variable*	*Coefficients*	*t Stat*
c_0	Intercept	4.80141	1.68
c_1	Demand	0.79453	27.46
c_2	Jan	0.15474	11.70
c_3	Feb	0.29946	22.79
c_4	Apr	0.17688	13.10
c_5	May	0.28245	21.75
c_6	Jun	0.07740	6.81
c_7	Aug	0.46590	28.65
c_8	Sep	0.11187	9.95
c_9	Oct	0.16832	14.43
c_{10}	Nov	0.24003	20.79

Table 1: Regression model coefficients

Goodness of fit measures: $R^2 = 0.9402$; adjusted $R^2 = 0.9347$

Key considerations and assumptions for this model

1. **Small number of observations:** We follow the one-in-ten rule [1] closely, and, being provided 120 data points, we use them to estimate 11 coefficients. This has a slight risk of overfitting but, owing to lack of data and goodness of fit, this is an acceptable risk.
2. **Multicollinearity:** Due to dropping three month-indicator variables, there is no multicollinearity in the model.
3. **Autocorrelation:** Time series data often runs the risk of autocorrelation, but this was addressed by ensuring that the potential correlations are grouped together as a single set for independent variables. Additionally, the Durbin-Watson test [2][3] was conducted and the statistic indicated that there is no statistical evidence that the error terms are positively or negatively correlated.
4. **Multivariate normality:** Mardia's test [4] was performed to test for multivariate normality. With calculated p-values for skewness and kurtosis being greater than our p-value of 0.05, we can conclude that the sample comes from a multivariate normal distribution.

For all the statistical work above, the level of confidence was set at 95%. The intercept does not meet this criterion, but as its purpose is to capture the unexplained uncertainty in the model, we do not omit it. All the other coefficients are statistically significant at this confidence level. This provides us a demand prediction model based on historic data, and if the seasonal trends for demand hold, this model will function well. This concludes the first part of this problem statement, the demand prediction part. The second part of the report discusses the inventory management policy given these demand predictions.

Inventory Management Policy

The first key to cost reduction is via safety stock. A good demand prediction model obviates a large safety stock. For example, the starting inventory in 1996 was 60 units ($60 holding cost), and in December 2005 it was 73 units ($73 holding cost). According to my interpretation of the process, each month progresses as follows:

- Receipt of units ordered in previous month
- Fulfilment of backorders
- Fulfilment of new orders
- Placing an order for units to be received at the start of next month (lead time = 1 month)
- Month passes
- Holding and inventory costs are calculated

Thus, we receive a new shipment before having to fulfil next month's demand, implying that the safety stock does not need to be so high. If the demand prediction is within 20 units of the actual demand, we are still better off in terms of actual costs than holding 60 units in inventory, as both cases cost us $60. Using this observation and testing the regression model on randomly generated demand curves following the pattern shapes observed in the data, we set the safety stock level to 5 units. Note that depending on the decision makers' risk-aversion, this may be set higher or lower to observe tradeoff between costs and other factors (potential customer goodwill, etc.). The threshold of 5 units allowed us to fulfil more than 99% of the generated demand in the same month and was therefore chosen. As an observation, setting safety stock to zero lowers total costs, but keeps the firm vulnerable to higher backorder risks, affecting performance metrics.

The inventory management policy is a modified version of just-in-time (JIT) and base stock policy. It only focuses on the current inventory level, backorder status, and projected demand for next month, and aims to reach the safety stock level next month. Thus, the orders placed each month follow Equation (2) below:

$$Number\ of\ units\ ordered = Predicted\ demand - Current\ inventory + Backorders + 5 \quad (2)$$

For reference, keeping a safety stock of 5 units increases the total cost by an average of 7.5% as compared to a safety stock of 0 units, while reducing backorder costs and increasing fulfilment percentage for randomly generated demand curves.

Conclusions

In this report, we present a demand prediction and inventory policy for a hardware component based on provided historical data. The demand prediction model is a multivariate regression considering the holding and backorder costs alongside the demand from the previous month. The underlying assumptions for this model choice were tested rigorously to ensure compliance. A high goodness of fit measure and ease of implementation and interpretability are key features of this model. This model also does not suffer from the lack of data availability, in contrast to other methods such as neural networks and constraint programming. This method depends on the demand pattern being similar in 2006-07 and significant deviations in the demand pattern will throw this forecast method. The lack of contextual information and more data leads us to believe that the assumption of similar demand pattern is a reasonable one.

The inventory management policy is a modified JIT and base stock policy, focusing on maintaining a safety stock of 5 units per month via predicted demand. The inventory policy only requires the current inventory level, backorder level, and predicted demand as inputs. While the safety stock level can also be determined dynamically as the 2006-07 data is provided, we assume that the demand pattern will not alter drastically from the previous years, leading to a static safety stock level being appropriate. Overall, while this competition is supposed to be a data driven competition, these “older” techniques have been chosen due to lack of a large dataset and proven track record.

Readme

Download the Excel file fORged_by_Machines.xlsx from the GitHub repository found here: https://github.com/PriyadarshanPatil/fORged-comp. This file has two sheets: data generation and summary statistics. The data generation sheet has space on lines 122-145 for the data from years 2006-2007, in the same format as the provided datasets. Paste your demand data in cells C122-C145. The results from our forecasts and policy will be populated automatically in cells D122-L145. These columns will provide the following data sequentially (column number in parenthesis): forecast demand (D), beginning inventory (E), order quantity which is fulfilled at start of next month (F), ending inventory (G), backorders, if any (H), holding costs (I), backorder costs (J), and total costs for that time period (K).

The summary statistics sheet provides the requested summary data: total cost (2006-2007), total holding cost (2006-2007), total backorder cost (2006-2007), average holding cost (2006-2007), and average backorder cost (2006-2007).

References

[1] Harrell Jr, F. E., Lee, K. L., & Mark, D. B. (1996). Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Statistics in medicine*, *15*(4), 361-387.

[2] Durbin, J., & Watson, G. (1950). Testing for Serial Correlation in Least Squares Regression: I. *Biometrika, 37*(3/4), 409-428. doi:10.2307/2332391

[3] Durbin, J., & Watson, G. (1951). Testing for Serial Correlation in Least Squares Regression. II. *Biometrika, 38*(1/2), 159-177. doi:10.2307/2332325

[4] Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*(3), 519-530.