Silky Silk & Cottony Cotton Corp

Business Optimization Exercise

2/9/2012

Data_Miners_Anonymous

Naveen Kumar	**61210144**
Akshay Sethi	**61210413**
Karthik Vemparala	**61210505**
Sruthi Yalaka	**61210416**

Contents

Executive Summary

This report focuses on the business application of forecasting techniques. We discuss how accurately forecasting retail demand for women's apparel in the United States (primary market of the company) helps the company reduce operating expenditure and rationalize capital expenditure in the future.

We have tried to forecast and model both the seasonality and the trend using the women apparel sales data gathered. We discuss different methods for forecasting, before choosing a modified Multiple Regression (using polynomial function of time) as the most appropriate model. Our decision takes into account the business requirements as well as the efficacy of forecast. While the seasonality helps us reduce operating costs, the trend is used to forecast average sales 5 years later thereby helping rationalize capital expenditure.

We sourced women apparel sales data from 1992 till 2011 from the US census data collected using monthly and annual retail trade surveys and administrative records. However, we were able to utilize data from 2008 onwards only. The financial crisis of 2008, significantly impacted demand and changed the retail landscape. For accurate forecasts, we used 4 years data starting from Jan 2008. We have used 2008 to 2010 data as training and 2011 data for validation. The data plot in Figure 1 of the appendix clearly shows seasonality in sales. Further, we can distinctly note the change in trend post year 2008.

We finally settled for a modified multiple regression with an equation shown below:

$$(MA)_t = 271.88 - (1.924)^*t + (0.049)^*t^2 + (1.65)^*(MA)_{(t-1)} - (0.734)^*MA_{(t-2)}$$

This particular model provides us with the cleanest forecast that meets the business mandate and has the least error. We have discussed the model in detail under the Section: *Methodology*.

While the model is fairly accurate it suffers from sudden/significant changes in the economic outlook. We noticed this in the year 2008. We also tried to incorporate the S&P500 index in our regression model but it did not yield much. Since the S&P500 and the retail demand are simultaneous indicators hence the stock index can't be used in the regression. Further, forecasting or estimating the S&P index in the future is difficult. The model also suffers from regulatory changes, developments in trade policy and rate of adoption of new technology by the industry players. Since we have used only 4 years data, we believe that forecasting into the future will have a lot of errors. Hence we shall discount any numbers forecasted into the distant future.

Using forecast of seasonality, SSCCC can manage its procurement, inventory and production schedule to reduce operating costs. Further, we forecasted the trend in the demand data post 2008 provides us an average demand in 2016 of US$79.3bn. On this basis we recommend that the company not invest in new machinery and plant as the current capacity will suffice.

Stakeholder

Silky Silk & Cottony Cotton Corp (SSCCC), based in the Guangzhou, China is one of the largest exporters of apparels to the US. As of 2010, US imported more than US $28.8billion worth of apparels from China. SSCCC being one of the largest apparel companies makes up for 20% of this trade. SSCCC is primarily involved in manufacturing of cotton and silk apparels and serves a wide clientele of customers from high end fashion houses to the aisle of Wal-Mart!

We, Data_Miners_Anonymous, have been mandated by SSCCC to help them optimize their operations and rationalize their operating and capital expenditure in the wake of the financial crisis.

Business Model

The value chain in the finished apparel business is buyer driven. The three main parts include procurement, production/inventory and retail sales. The picture below depicts this value chain.

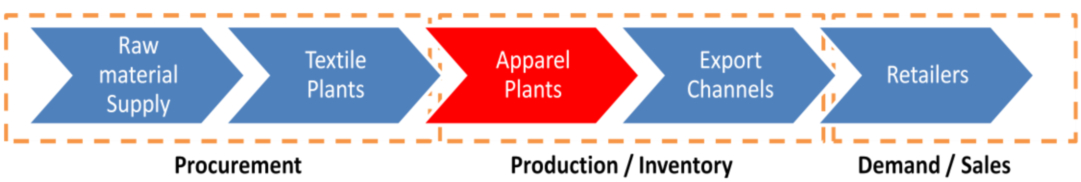


Figure 1: Value Chain

Sales Contracts 80% of SSCCC's orders received are regular orders which follow the standard lead time of 16 weeks while the remaining 20% are ad hoc orders that arise due to demand fluctuations. Lead time for ad hoc orders is typically 4 weeks and is met out of the finished goods inventory. This inventory buffer ensures that our model does not need to accurately capture peaks and slight errors in forecast can be easily managed.

Costs and Influencing Factors The total cost for a garment consists of direct costs associated with production (e.g. fabric, labor, and shipping), policy-related costs and supply chain costs. As SSCCC serves customers who follow a lean retailer model, any increase in its lead time in procurement or fluctuations in demand will impact costs associated with WIP inventory, Finished Goods Inventory and Inventory at Risk.

The Mandate

Over the last 3 years, SSCCC has seen an unprecedented drop in demand from their retailers in the United States. In the year 2008, SSCCC had finished increasing their capacity to meet the increasing demand from retailers. But this sudden fall caught them off guard and the company suffered severe losses. As of 2008, SSCCC plants were 33% utilized. They are now in the process of drafting a 5 year plan which optimizes operating cost by modifying the production plan taking into account seasonality and rationalizing capacity over the next few years.

To this affect we, Data_Miners_Anonymous (DMA) have been mandated to:

1. Identify the reasons for the unprecedented drop in demand.
2. Identify & Forecast seasonality to optimize procurement, inventory and production schedule. This shall help the SSCCC rationalize their operating expenditure
3. Forecast average demand up till 2016 to help SSCCC manage their capacity expansion and capital expenditure.

Methodology

Benchmark: In order to set the benchmark to evaluate the performance and usefulness of various forecasting methods that are being used, Naïve forecasting was performed and the performance metrics were measured. There is a clear seasonality that can be observed in the data (Appendix – Fig1). The year 2011 was used as a validation period. The results were forecasted and measured against the actual data to evaluate the errors. The MAPE was 2.64% and the RMSE was around 94 million dollars. (Appendix – Table 1). The plot of actual versus the Naïve forecast values shows that for the most part the method has been under-forecasting which can be explained by the fact that the trend is not being considered in the Naïve forecasts.

Data Exploration: The data plot (Appendix – Figure 1) clearly displays both trend and seasonality. Monthly seasonality can be clearly observed (Appendix – Figure 2) and the trend is visible too.

However there is a drastic change or dip during 2008 which can be attributed to the economic shock. Hence in order to utilize the data for forecasting 2012, we will have to break up the data into 2 parts, pre-2008 and post-2008. The complete data would be utilized to analyze the seasonality as the seasonality does not seem to be affected by the economic shock in 2008.

Seasonal Indexes: The seasonal indexes were calculated and we observed that December sales are about 50% higher than the average sales during the year. The sales are 20-25% lower during the first two months of the calendar year. The indexes are shown in Appendix - Table 2.

Methods used: Firstly, we de-seasonalized the data by using Moving Averages over a 12-month period. The de-seasonalized data can be seen in the Appendix – Figure 3. To model the trend in the data, we used a MLR with the time variable as the predictor. However, this model was not completely capturing the trend. We thus needed to model the trend using a polynomial function of time. Further, we also noticed that the residuals were heavily auto-correlated. To capture the trend in the residuals, we used a Lag-1 and Lag-2 variables as inputs to the model. Here, we needed to use the first 36 months as the training period and the remaining 11 months as the validation period. The performance significantly improved and easily beat the benchmark set by the Naive model. The average error was just 14.3 million dollars and the RMSE was around 15.9 million dollars. These are errors which are less than 1% of the sales values. This model provides a roll-forward forecast on a monthly basis. However, we can use the predicted values to provide forecasts for the next 6 months. The results are shown in Figure 4 and Figure 5 of the appendix.

The second method we used was Holt-Winter exponential smoothing. We again used a Training period of 36 months (January 2008 to December 2010) and a Validation period of 11 months (January 2011 to November 2011). We used the default values of the learning factors: $\alpha = 0.2$, $\beta = 0.15$ and $\gamma = 0.05$. This model also gave us a good performance. However, it over-forecasted for the most part and the graph of the residuals shown in Figure 7 of the appendix clearly illustrates this. We also tried using various different values of the learning factors. However, the default values provided the best results.

In sum, the multiple regression method using a polynomial function of time on the de-seasonalized data provided the best forecast performance. This can be used by the business on a roll-forward basis to predict monthly or quarterly demand. The summary of the models' performance can be seen in the appendices.

Limitations: The model does not account for changes in economic outlook, policy and regulation and technology adoption, as explained below:

- As was seen during the 2008 financial crisis, the consumption and demand patterns slumped and the trend changed significantly thereafter. While we explored regression using market movements as a dependant variable, but the accuracy of prediction remains questionable as the market moves simultaneously with demand, and cannot be predicted.
- The garment export industry is governed by trade regulations, currency fluctuations and related government policies. Changes in these policies are known to impact demand significantly for overseas clients.
- As Western retailers adopt just-in-time inventory models, there is a pressing need for SSCCC and other Asian players to adapt to advanced information technology to reduce production times. The frequency of data updates and the associated time lag will help to improve the model significantly.

Conclusion & Recommendation

Based on the above analysis and forecasts, the outlook projected for the women's retails sales are positive and the demand is expected to grow over the next few years. Based on the available projections currently, retail demand would be worth US$79.3b (using 2008 prices) in 2016. At this rate SSCCC's plant would be 74% utilized and hence we recommend no capital expenditure to build capacity.

The model (Multiple Regression) that has been created would help generate forecasts on a roll forward basis that will help the company decide the resource allocation and optimize operating expenses. Even with some over or under prediction that the model will generate, the inventory handling can be done efficiently as the pattern is understood and the planning could incorporate the seasonality and the errors to certain extent in to the buffer.

We note that the US$79.3b sales in 2012 is a large number and should be discount. This is error in forecast is because of the limited training and validation we have. More data addition by periodically updates will incorporate the actual data going forward. As the data size increases the performance of the model is expected to be better and the processes will be more efficient.

APPENDICES

Figure 1: Overlay of S&P 500 index and the Women's Retail Sales from 2003 until 2011

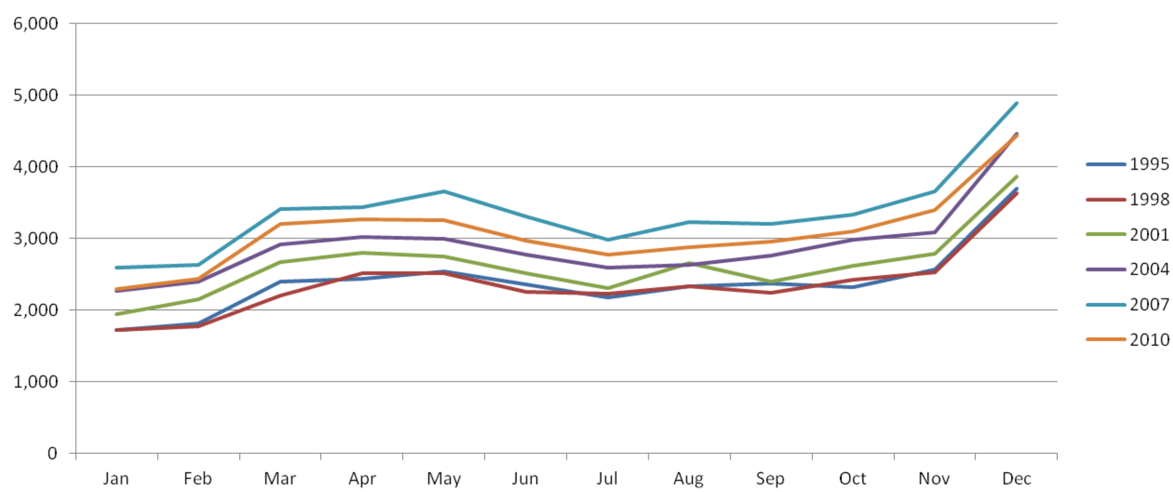
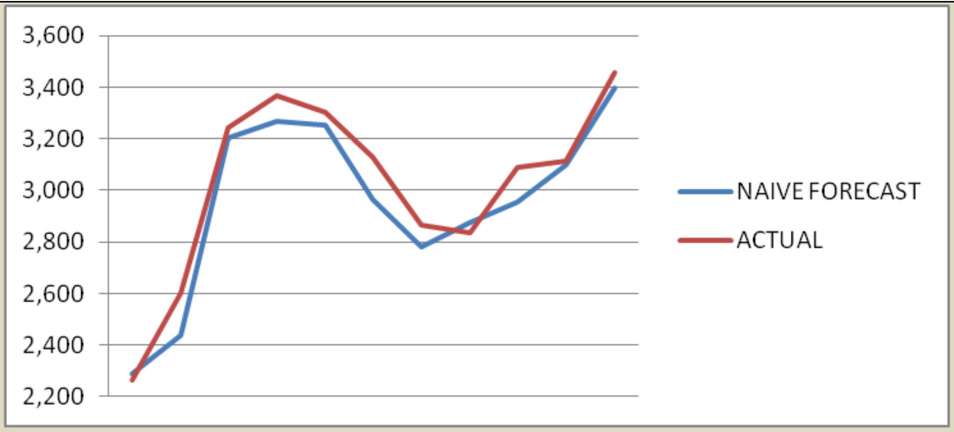


Figure 2: Women's retail sales figures displaying consistent monthly seasonality



NAIVE FORECAST ACCURACY MEASURES	
Average Error	67.91
MAE	79.18
MSE	8,846.27
RMSE	94.05
MAPE	2.64%

**** Numbers are in Millions of Dollars**

Table 1: Naive Forecast performance indicators

January	0.742175277
February	0.791155726
March	1.001283573
April	1.036938503
May	1.060040452
June	0.970777155
July	0.903434002
August	0.965914436
September	0.959841893
October	0.999256854
November	1.070287557
December	1.498894573

Table 2: Calculated Seasonal Indexes

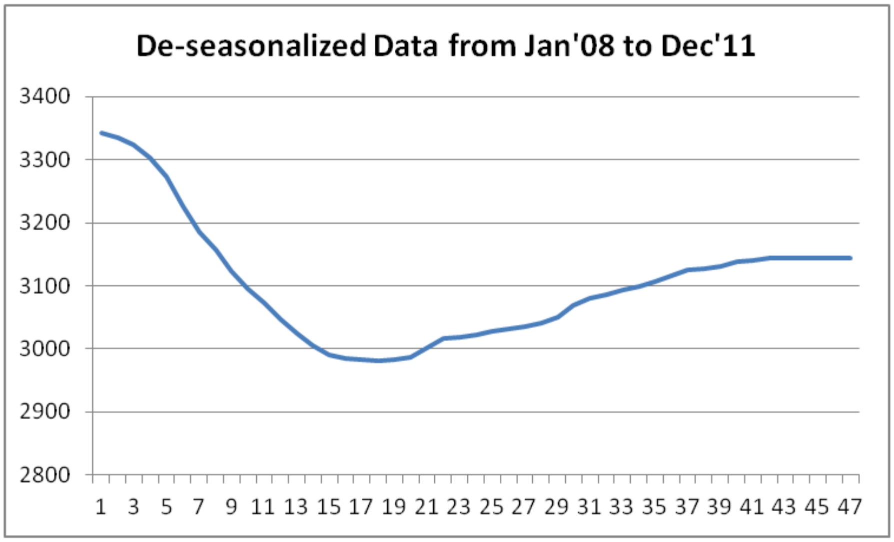


Figure 3: Figure showing the trend of the de-seasonalized data

---

The Regression Model

Input variables	Coefficient	Std. Error	p-value	SS
Constant term	271.8802185	153.7242127	0.08747561	322010900
t	-1.92398322	1.98304105	0.33996055	57706.80469
t squared	0.0488452	0.04318242	0.26726553	201731.6875
MA (Lag 1)	1.64951682	0.09594848	0	19003.04688
MA (Lag 2)	-0.73350829	0.10060847	0.00000001	1688.44751

Residual df	29
Multiple R-squared	0.996722375
Std. Dev. estimate	5.6360302
Residual SS	921.1802368

Training Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
921.1697172	5.205115462	4.58676E-06

Validation Data scoring - Summary Report

Total sum of squared errors	RMS Error	Average Error
2769.964231	15.86867764	-14.2737043

$$(MA)_t = 271.88 - (1.924)*t + (0.049)*t^2 + (1.65)*(MA)_{(t-1)} - (0.734)*MA_{(t-2)}$$

Figure 4: The Regression Model

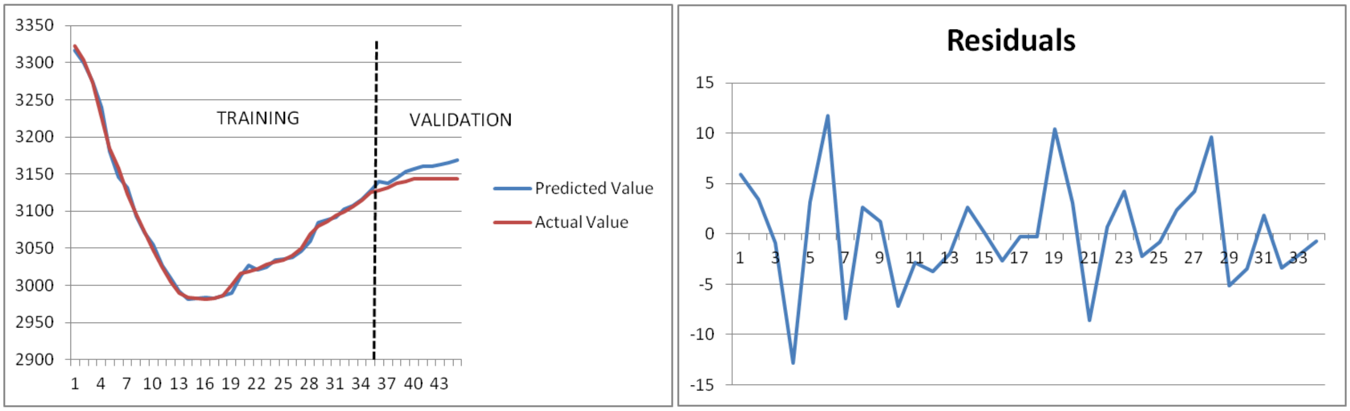


Figure 5: Graphs illustrating the performance of the MLR model

---

Error Measures (Training)

MAPE	2.8428922
MAD	87.717358
MSE	10441.209

Forecast

Month	Actual	Forecast	Error	LCI	UCI
Saturday, January 01, 2011	2266	2384.3703	-118.37033	2184.0932	2584.6475
Tuesday, February 01, 2011	2601	2556.9313	44.068685	2356.6541	2757.2085
Tuesday, March 01, 2011	3243	3212.2025	30.797474	3011.9254	3412.4797
Friday, April 01, 2011	3368	3376.3902	-8.3901761	3176.113	3576.6674
Sunday, May 01, 2011	3304	3465.3881	-161.3881	3265.1109	3665.6653
Wednesday, June 01, 2011	3130	3125.4595	4.5405174	2925.1823	3325.7367
Friday, July 01, 2011	2867	2920.6271	-53.627082	2720.3499	3120.9043
Monday, August 01, 2011	2838	3101.7465	-263.74648	2901.4693	3302.0237
Thursday, September 01, 2011	3087	3115.0506	-28.050621	2914.7734	3315.3278
Saturday, October 01, 2011	3114	3249.2386	-135.23863	3048.9615	3449.5158
Tuesday, November 01, 2011	3456	3507.1402	-51.140198	3306.863	3707.4174

Error Measures (Validation)

MAPE	2.8219775
MAD	81.759845
MSE	12469.901

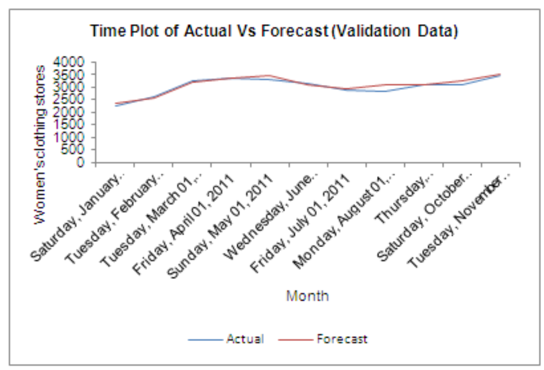


Figure 6: Performance of the Holt-Winter model

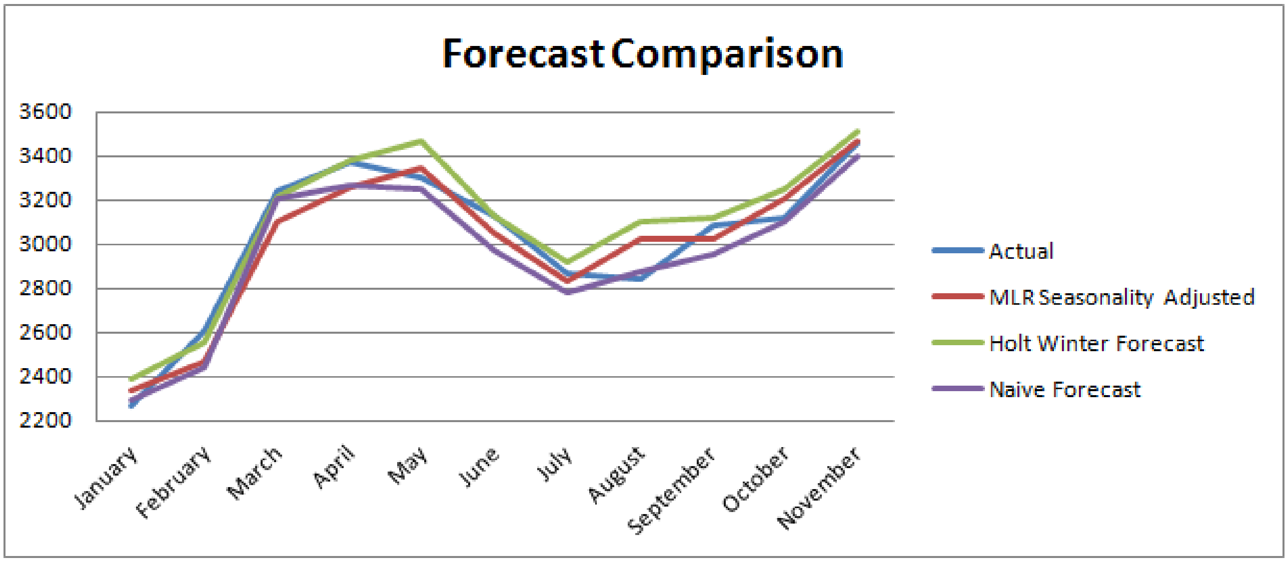


Figure 7: Performance Comparison between forecasts from Naive, Holt Winter and Regression against the actual data