Volume 10, Issue 2, 2016

Influence of Factors on Clothing Sales and Its Future Trend: Regression Analysis and Time Series Forecast of Clothing Sales

Avoni Gharde
Ph.D. Program in Textile Technology Management
College of Textiles
North Carolina State University
Raleigh, North Carolina
27695-8301
USA

ABSTRACT

Forecasting is crucial for many industries. Sales forecasting for retail is critical to understand the volatile product demand and supply chain lifecycle. This paper is a study conducted on clothing store sales with data of previous 20 years. Variables such as GDP, GDP per capita, population, marriage rates and their effect on clothing store sales were studied. Regression analysis, ARIMA model for time series analysis with JMP software were used to forecast sales for the next 20 and 50 years.

Keywords: Sales forecasting, clothing stores, Regression analysis, ARIMA

J T A T M

Introduction

The apparel supply chain has one purpose – to provide an appealing and desirable product to satisfy consumer needs, wants, or aspirations. When successful, the connection results in a sale. Because this context is the purpose of the process, every forecast begins with the customer – by observing the client's adjustments to the marketplace and in the unexpected ways the customer adjusts the market to his or her lifestyle and preferences. Forecasting, in the apparel and textiles industry, is a process that spans shifts in color and styles, changes in lifestyles and buying patterns, and different ways of doing business (Brannon, 2000). What appeals to be near random activity is, in fact, a process of negotiation between the fashion industry and the consumer and between the various segments in the supply side chain. In the narrowest sense, forecasting attempts to project past trends into the future.

Cost reduction, customer requirements, globalization and technology, have impacted the textiles, apparel and fashion industry and have contributed to the varying enablers of seasonal looks in the market.

Sales Forecasting

Sales forecasting refers to predicting future demand (or sales), assuming that the factors which affected demand in the past and are affecting the present will still have an influence in the future. It is an important task but is tough to accomplish (Liu, Ren, Choi, Hui, & Ng, 2013). In the fashion retailing industry, which is defined as the retailing business of fashion products including apparel, shoes, and fashion beauty products, forecasting itself can be treated as a "service" which represents the set of analytical tools that facilitate the companies to make the best decisions for predicting the future. Undoubtedly, a good forecasting service system assists in avoiding understocking or overstocking in retail inventory planning. It further relates to critical operations of the whole supply chain management, production planning, pricing (C.-H.Chiu, T.-M.Choi, & Li, 2009; Chiu & Choi, 2010), and achieving high customer service level (Frank, Garg, Raheja, & Sztandera, 2003).

Over the past decade, some research studies have been reported in the literature. However, each forecasting method has its limits and drawbacks. For example, the traditional statistical methods depend highly on the time series data's features, and this will affect the forecasting accuracy a lot. Artificial intelligence (AI) methods can perform better regarding accuracy than the traditional statistical forecasting models, but they usually require a much longer time and a larger requirement on computational power. Thus, many researchers propose to combine multiple methods together to form a new “hybrid method” to achieve an efficient and effective forecasting task (Liu, Ren, Choi, Hui, & Ng, 2013).

Statistical Models for Sales Forecasting

Traditionally, fashion sales forecasting is accomplished by the statistical methods. A lot of statistical methods have been used for sales forecasting, which includes linear regression, moving average, weighted average, exponential smoothing, exponential smoothing with a trend, double exponential smoothing, Bayesian analysis, and so forth. Statistical time series analysis tools such as ARIMA and SARIMA are also widely employed in sales forecasting. The production planning of autumn-winter textiles from creation to distribution is shown in figure 1.

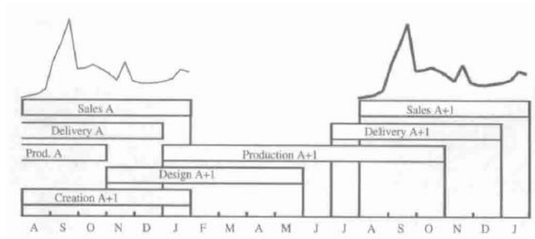

Figure 1. Production Planning of Autumn-Winter Textile Items

The purchasing managers need to know almost one year in advance the quantities of raw materials to order which will correspond to the forecast total volume of each textile product range. Production managers also require item quantities to manufacture particularly early in the case of imported items from distant countries. Moreover, it is necessary to fit in the forecasts after the launch of the textile items according to the restock achieved by local manufacturers. Therefore the textile industry requires both mean (one season) term and short term (one to six weeks) forecasting models (Vroman, Happiette, & Vasseur, 2001).

In the literature, Green & Harrison (1973) apply a Bayesian approach to explore forecasting for a mail order company which sells ladies dresses. After that, Thomassey, Happiette & Castelain (2003), used item classification to examine the accuracy of sales forecasting for new items. They found that a larger number of item families and appropriate classification criteria are required in the particular prediction procedure to achieve an improved forecasting precision. They concluded that product family and aggregated forecasting are more accurate than the individual item’s forecasting. It is accepted that there is no forecasting technique appropriate to an all sales time series (Ferrer, Hoyo, & Arroyo, 1997).

Witt & Witt (1994), have reported that different forecasting techniques might perform differently in handling stable vs. unstable data. It is well accepted that no forecasting technology is appropriate to all

J T A T M

data patterns. But no research has been investigated to compare the effects of various methods on different sales data patterns from apparel retailing, therefore, leaving room for further exploration.

Time-Critical Decision Modeling and Analysis

The ability to model, perform decision modeling and analysis is an important feature of many real-world applications. These range from medical treatments to military systems. Existing formalisms and steps of inference have not been capable of real-time applications where trade-offs between decision quality and computational tractability are essential. In practice, a practical approach to time-critical dynamic decision modelling should provide explicit support for the modelling of temporal processes and for dealing with time-critical situations (Arsham, n.d.). Almost all managerial decisions depend on forecasts. A decision becomes operational at a stage in the future, so ideally, it should be based on forecasts of future conditions.

Two approaches to forecasting are the estimate of future value, based on an analysis of factors that influence future values, i.e., the explanatory method, or else the prediction builds on an inferred study of past data behavior over time, i.e., the extrapolation method. Both approaches can lead to a creation of accurate and meaningful forecasts, however, even for a modest degree of the desired accuracy, the former method is often difficult to implement and validate.

J
T
A
T
M

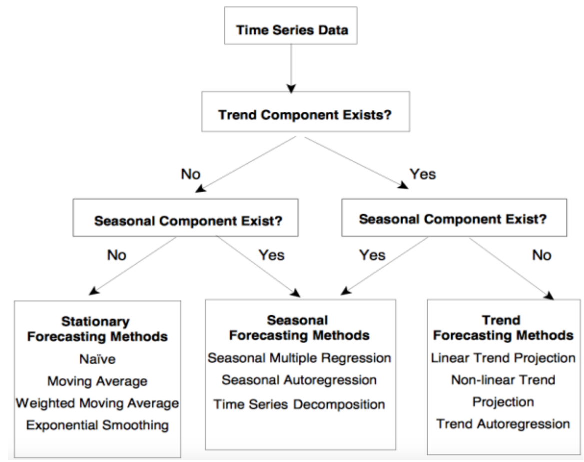

Figure 2. Potential Choices of Forecasting Methods

Data Collection

US Department of Commerce's database provided with the data of estimated annual sales of US Retail and Food services with various categories including clothing and accessories stores around the US. Population, marriage rates, divorce rates as well as GDP and GDP per capita were gathered from US Census Bureau and CDC (Centers for Disease Control and Prevention). The variables chosen to study their effect on sales were GDP, GDP per capita, population and marriage rates of USA.

Influence of Factors on Expenditures

It is quite obvious to the fact that as the household income shall increase, costs shall go up. About clothing, consumer behavior is highly unpredictable and variables such as income, season, political climate, environmental factors have no effect on the shopping pattern. No matter what, going shopping is a major source of relaxation as well as a household chore (Badar, 2008).

According to the study conducted by Badar, decision making is a critical component of consumer behavior that significantly affects the choices made. In his study among the sampled respondents, more than 57 percent were married, around 41 percent were singles while less than one percent were divorced and widows. He found that family size contributed significantly in shopping pattern of the consumers; larger the family size, higher will be the shopping expenditure. In another study by Smith (2004), there were many single households in the US than married households with children. Smith found that shopping as a family was three times more likely in single mother households than married ones. (Martin & Ramsaran-Fowdar, 18 - 20 November, 2013).

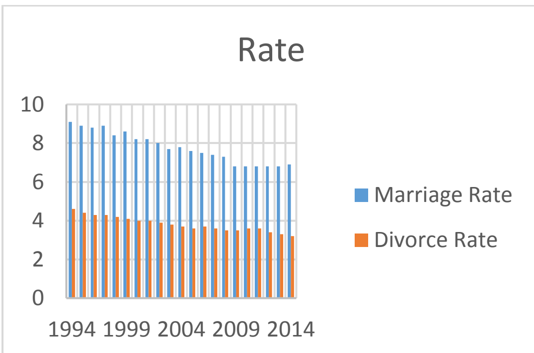

Figure 3. Rates of marriage and divorce (USA)

In the case of married couples, an interesting observation was made. An increase in the wife's income relative to her husband's income was associated with greater expenditure on food, childcare, and children's and women's clothing, and reduced expenditure on transportation (Phipps and Burton, 1998). Lundberg, Startz, and Stillman (2003) noted that wives preferred to save more than their husbands and that relative control over household decisions are affected by control over market income (ZIOL-GUEST, DELEIRE, & KALIL, 2006).

From these research studies, a prominent hypothesis which formed was the influence of marital status on the consumer shopping behavior – in this case, whether or not it is directing affecting the clothing store sales. To start with, a regression analysis was done on identifying factors which significantly affected clothing store sales in the period of 1994-2014.

JTATM

Regression Analysis

Regression analysis is a statistical technique to analyze quantitative data to estimate model parameters and make forecasts. The X-axis is the horizontal line, and the vertical line is referred to as the Y-axis. Regression analysis looks for a relationship between the X variable (sometimes called the "independent" or "explanatory" variable) and the Y variable (the "dependent" variable) (Guerard Jr., 2013). Forecasts using the regression line, assume that the relationship which existed in the past between the two variables will continue to live in the future.

Consumption makes up for a majority of real Gross Domestic Product, known as GDP (Guerard Jr., 2013). GDP per Capita compares GDP on a purchasing power parity basis divided by population as of 1 July for the same year. From the figure below, we see that the real GDP has increased substantially with a slight dip during the recession of 2009.

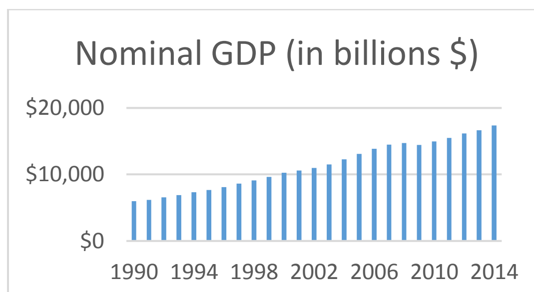

Figure 4. Incremental GDP (in billions $) of United States

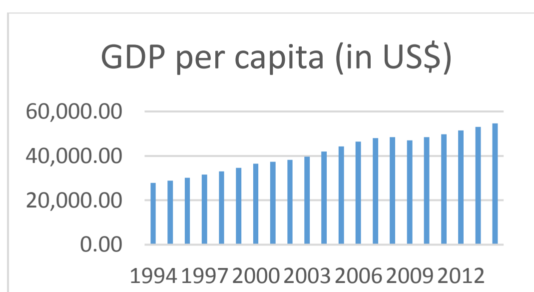

Figure 5. Incremental GDP per Capita (in $) of United States

Figure 6. Clothing Store Sales

This corresponds to the trend shown by GDP per Capita and clothing sales during the specified period.

Using JMP software, after analyzing the data, the best-fit line and a regression model were developed with the clothing sales by GDP of USA plot; for the period 1994-2014. As we can see by the plot, the increase in GDP through the years has guided the simultaneous incremental clothing sales during 1994-2014.

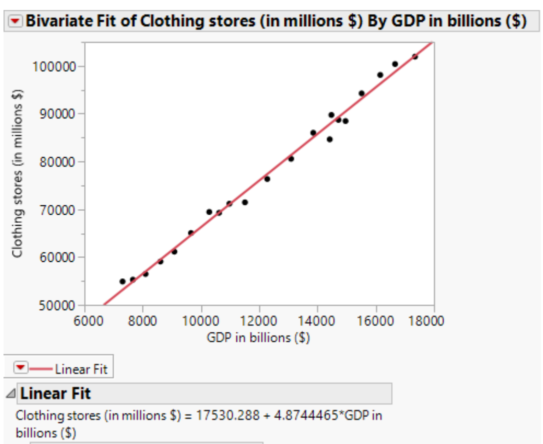

Figure 7. Best fit of Clothing Sales by GDP (USA)

An estimated regression model equation generated by the software is as follows:

Clothing stores (in millions $) = 17530.288 + 4.8744465*GDP in billions ($)

The predictor variable here being GDP (in billions $) and the response variable is clothing sales (in millions $) and R-square as well as R-square adjacent values as 0.9918 and 0.9913 respectively.

Regression models were also developed for clothing store sales by other variables like marriage rate and GDP per Capita. Equation generated was as follows:

*Clothing stores = 220509.66-8421.855*Marriage Rate*

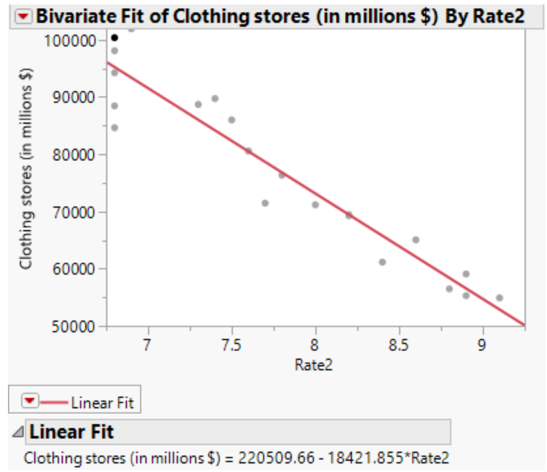

Figure 8. Best fit of Clothing Sales by Marriage Rate (USA)

R-square as well as R-square adjacent values were produced as 0.91195 and 0.9073 respectively. There is a very high negative correlation amongst the two with inverse proportionality. The probability value was found to be <0.001 for marriage rates, which proves a large amount of interaction. The equation generated for GDP per Capita by clothing store sales was:

*Clothing stores = 1930.1785+1.8162133*GDP per capita*

Figure 9. Best Fit of Clothing Sales by GDP per Capita

J T A T M

R-square as well as R-square adjacent values were produced as 0.98958 and 0.9890 respectively for figure 9. There is a high interaction between GDP per capita and clothing sales, so as the income increases, expenditure on clothing stores shall proportionally rise.

For figure 10, R-square, as well as R-square adjacent values, were produced as 0.9641 and 0.9622 respectively. The estimated regression equation generated by the software was:

*Clothing stores = - 153624.8 +0.0007939*Population*

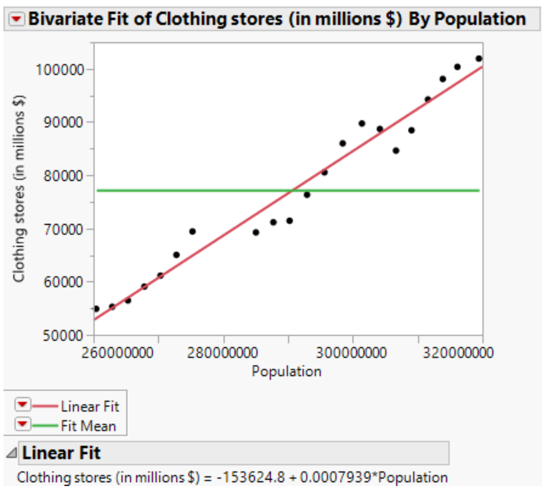

Figure 10. Best fit of Clothing Sales by Population

About the hypothesis presented earlier of marital status affecting clothing store sales, as seen in the previous analysis, there is a high negative correlation. The next step was to figure out the impact on income and GDP since sales and GDP were proportionally related to one another. At a fundamental level, every marriage creates a new household, an independent economic unit that generates income, spends, saves, and invests. The vast majority of these new homes produces babies and transforms what are mostly teens, into responsible adults, contributing to the necessary next generation of human capital to the economy.

According to the U.S. Bureau of Labor Statistics, Consumer Expenditure Survey (2007-2008), irrespective of their marital status, consumers with ages 27 to 29 years, earn more, spend more, are more likely to have a bachelor's degree. Also, they are more likely to own a home than their early twenties counterparts. Singles in their 20 to 24 years earn less in income than early-twenties married couples but spend about the same amount per capita. However, late-twenties singles earn per-capita incomes similar to those of late-twenties married couples but spend significantly more per capita. Income appreciably affects GDP of the nation, and this is now in accordance with the analysis done below with independent variable as marriage rate.

JTATM

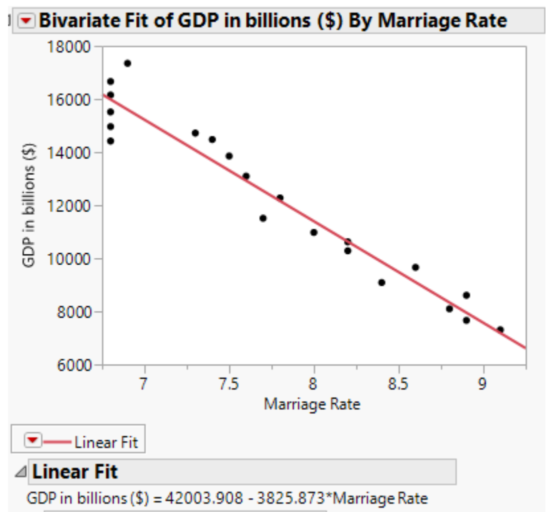

Figure 11. Best fit of GDP by Marriage Rate

The equation generated was:

*GDP in billions ($) = 42003.908 - 3825.873*Marriage Rate*

R-square as well as R-square adjacent values were produced as 0.9423 and 0.9392 respectively. Also, the probability value was <0.001 which shows a high significance.

Using the Graph Builder tool on JMP software, a visualizing correlation between data becomes easier. The figure given below shows us clothing store sales in relation to population and GDP per Capita of USA (period of 1994-1014).

Figure 12. Correlation between Clothing Store Sales by Population and GDP per Capita

As population and GDP per capita were increased, there was a significant rise in the clothing sales as reported, which is, again, in correlation to the analysis done above.

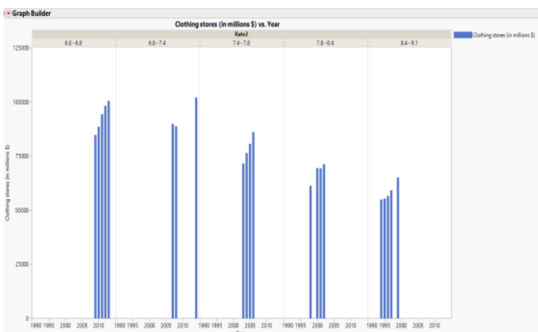

Figure 13. Marriage Rate vs. Clothing Store Sales

As per the graphs, there has been a decrease in marriage rates during the period 1994-2014. The trend seems to be the same with an increase in clothing store sales despite lower rates of marriage.

Time Series Forecast Model on JMP

The Time Series platform allows you to explore, analyze, and forecast univariate time series. A time series is a set y_1, y_2, ... , y_N of observations taken over a series of equally-spaced time periods. The analysis begins with a plot of the points in the time series (SAS Institute Inc., n.d.). Also, the platform displays graphs of the autocorrelations and partial autocorrelations of the series. These indicate how and to what degree each point in the series is correlated with earlier values in the series.

Using this platform, data was studied, and a growing trend pattern was discovered in annual clothing store sales from 1994-2014 while a decreasing trend was seen in the marriage rate vs. GDP (in billions $).

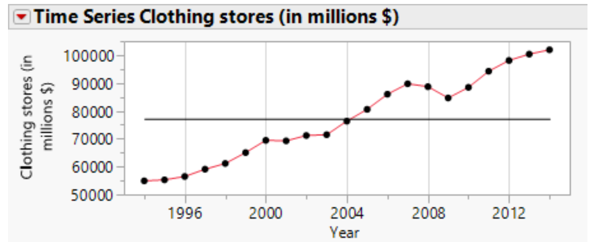

Figure 14. Trend of Clothing Store Sales in the US (1994-2014)

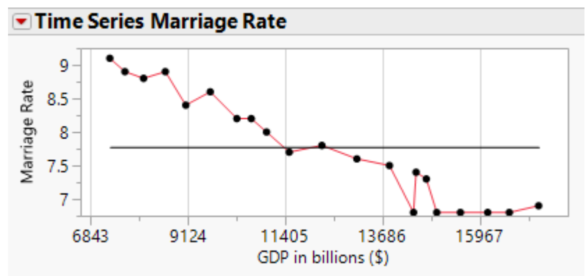

Figure 15. Trend of Marriage Rate by GDP in the US (1994-2014)

ARIMA

Theoretically, ARIMA models are the most general class of models for forecasting a time series which can be made "stationary" by differencing, perhaps in conjunction with nonlinear transformations such as logging or deflating. A random variable, like a time series, is stationary if its statistical properties are all constant over time. *A stationary series has no trend as its variations around its mean have a constant amplitude, and it wiggles in a consistent fashion*, i.e., its short-term random time patterns always look the same in a statistical sense. The latter condition means that its *autocorrelations* (correlations with its prior deviations from the mean) remain constant over time remains constant over time (Duke University, n.d.). A random

JTATM

variable of this type can be seen as a combination of signal and noise, and the signal could be a pattern of fast or slow mean reversion, sinusoidal oscillation, or a rapid alternation in sign. A seasonal component can be implemented, if needed. An ARIMA model is used as a "filter" that tries to separate the signal from the noise, and the signal is then extrapolated into the future to obtain forecasts.

To choose an appropriate model, first, a determination of the order of differencing is required to stationarize the series. Normally, the correct amount of differencing is the lowest order of differencing that yields a time series, which fluctuates around a well-defined mean value and whose autocorrelation function (ACF) plot decays fairly rapidly to zero (either from above or below). Second is to determine the orders *p* and *q* of the integrated time series of order *d* by studying ACF and partial ACF i.e. PACF.

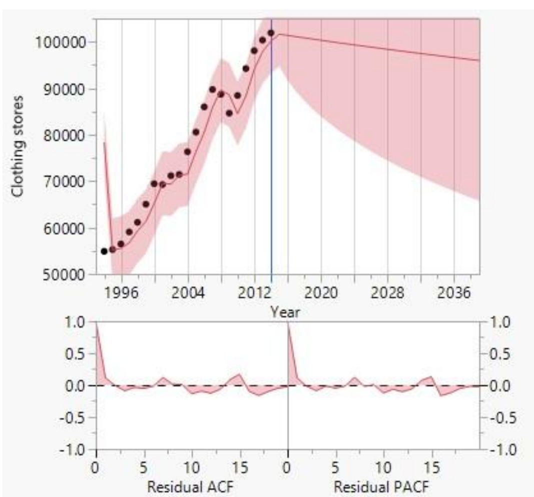

Figure 16. AR (1) Model for Clothing Store Sales

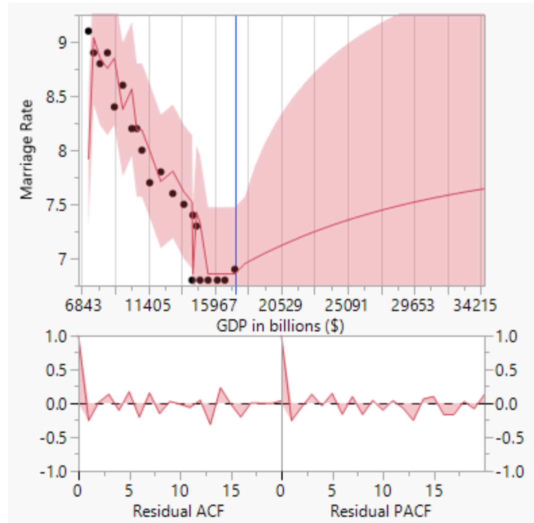

Figure 17. AR (1) Model for Marriage Rate

The plots of ACF and PACF can be interpreted as follows:

a) If the ACF abruptly stops at some point, after *q* spikes, this can be an MA, i.e., Moving Average (q) type model.
b) If the PACF abruptly stops after *p* spikes too, here AR (p) model would work best.
c) Had neither been present but a decline towards zero was evident in some fashion, the appropriate model would have been ARMA (*p,q)*.

JMP software helps to work with this model as well as SARIMA. With the given input data, a forecast was generated as follows:

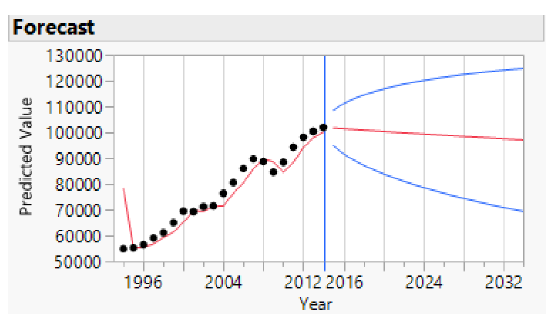

Figure 18. Predicted Sales for Next 20 Years (2015-2034) with AR (1) Model

J T A T M

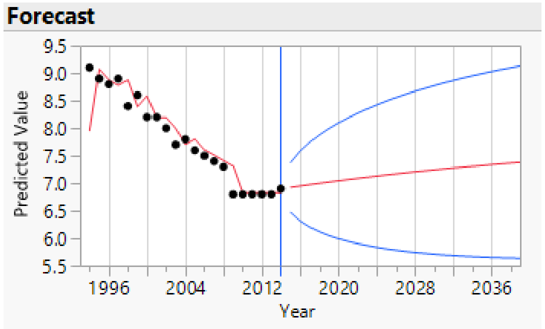

Figure 19. Predicted Change in Marriage Rates for Next 20 Years (2015-2034) with AR (1) Model

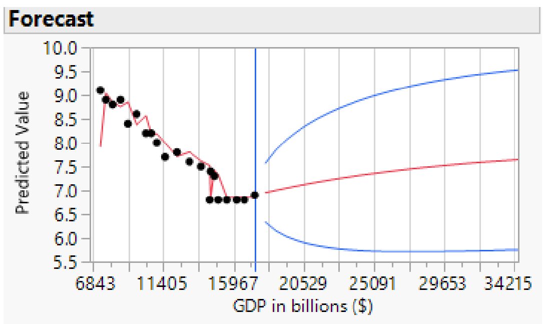

Figure 20. Predicted Change in marriage Rate as GDP Increases with AR (1) Model

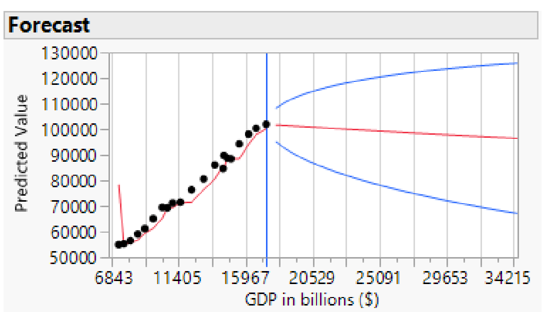

Figure 21. Predicted Rise in Clothing Store Sales as GDP Increases with AR (1) Model

AR (1) models (depicted above) were plotted with probability values and 0.05 (95% confidence intervals). In figure 18, R-square, as well as R-square adjacent values, were found as 0.83867 and 0.83018 respectively for clothing store sales. For marriage rates (figure 19), R-square, as well as R-square adjacent values produced, were 0.8227 and 0.81339 respectively while for marriage rates vs. GDP (figure 20), R-square as well as R-square adjacent values were 0.7554 and 0.7425 respectively. Finally, we see a predicted rise in clothing store sales (figure 21) as GDP rise and R-square as well as R-square adjacent values were found as 0.84211 and 0.8338 respectively. Once the plots are made, predicted values give us accurate annual sales data with a confidence interval range of values.

All the probability values obtained were less than 0.001 which proved that they are significant; as there is a rise in marriage rate with increase in GDP, there should be a rise in the clothing store sales as well. This proves the hypothesis of marriage rate being a factor influencing the clothing store sales!

JTATM

Conclusion and Future Scope

In this paper, a short literature review regarding the tools and models that are used for forecasting is provided. Real-time data from US Department of Commerce (retail-clothing stores sales) as well as demographics and economic survey data from US Census Bureau was studied to analyze and predict future sales for a span of 20 years and 50 years.

From the regression analysis done on the data, there is high significance and interaction present between GDP, GDP per Capita, population to clothing store sales and also marriage rates by GDP in the USA in the specified period of 1994-2014. Also, in the trend observed (of clothing store sales), there seems to be an increase in these sales while there is a decrease in marriage rates.

With time series analysis, there is a growth in expenditure as the household income per capita or GDP of USA rises per year; or so we expect this. However, the forecasts show a dip in the overall sale and a large confidence interval is plotted with variance. We also found there will be a rise in marriage rates as GDP increase which in turn should lead to a rise in clothing store sale. Just forecasting based on one factor isn't yielding accurate results.

The AR (1) model produced a large confidence interval between predicted values. Newer models such as pure AI and hybrid models could be tried to predict more refined forecasting. Gathering many historical data creates noise, therefore optimizing the data, although a hassle, can lead to improved results.

References

Arsham, H. (n.d.). *Time-Critical Decision Making*. Retrieved from http://home.ubalt.edu/ntsbarsh/stat-data/forecast.htm#rgintroduction

Badar, H. (2008). An Exploration of Female Shopping Behaviour: A Case Study of City Faisalabad (Pakistan). *Pakistan Journal of Life and Social Sciences*, 6 (2), 75-79.

Brannon, E. (2000). *Color Forecasting In Fashion Forecasting.* New York: Fairchild Publication Inc.

C.-H.Chiu, T.-M.Choi, & Li, D. (2009). Pricewall orwar: the pricing strategies for retailers. *IEEE Transactions on Systems,Man, and Cybernetics Part A*, 39 (2), 331–343.

Chiu, C.-H., & Choi, T.-M. (2010). Optimal pricing and stocking decisions for newsvendor problem with value-at-risk consideration. *IEEE Transactions on Systems,Man, and Cybernetics Part A*, 40 (5), 1116–1119.

Choi, T. M., Hui, C. L., & Yu, Y. (2003). *Intelligent Fashion Forecasting Systems: Models and Applications.* New York, NY, USA: Springer.

Duke University. (n.d.). *Introduction to ARIMA models*. Retrieved from Forecasting: http://people.duke.edu/~rnau/411arim.htm

Ferrer, A. G., Hoyo, J. D., & Arroyo, A. M. (1997). Univariate forecasting comparisons: The case of the Spanish automobile industry. *Journal of Forecasting*, 16, 1-17.

Frank, C., A. G., Raheja, A., & Sztandera, L. (2003). Forecasting women's apparel sales using mathematical modeling. *International Journal of Clothing Science and Technology*, 15 (2), 107–125.

Green, M., & Harrison, P. J. (1973). Fashion forecasting for a mail order company using a Bayesian approach. *Operational Research Quarterly*, 24 (2), 193–205.

Guerard Jr., J. (2013). Regression Analysis and Forecasting Models. In J. Guerard Jr., *Introduction to Financial Forecasting in Investment Analysis* (pp. 19-45). Springer.

Liu, N., Ren, S., Choi, T.-M., Hui, C.-L., & Ng, S.-F. (2013). Sales Forecasting for Fashion Retailing Service Industry: A Review. *Mathematical Problems in Engineering*.

Lundberg, Shelly, Richard Startz, and Steven Stillman. (2003). The Retirement Consumption Puzzle: A Marital Bargaining Approach. Journal of Public Economics, 87 (5–6): 1199–1218.

Martin, K., & Ramsaran-Fowdar, R. R. (18 - 20 November, 2013). An Examination of the Consumer Buying Behaviour of Single-Parent Households. *Proceedings of 23rd International Business Research Conference.* Melbourne, Australia.

Phipps, Shelley A. and Peter S. Burton. 1998. What's Mine Is Yours? The Influence of Male and Female Incomes on Patterns of Household Expenditure. Economics, 65 (260), 599–613.

SAS Institute Inc. (n.d.). *Time Series Analysis*. Retrieved from JMP SAS: http://www.jmp.com/support/help/Time_Series_Analysis.shtml

Smith, J. (2004). A single-minded marketplace. *Marketing Management*, *Vol. 13, Issue 4, p. 52.*

J T A T M

Somer, E., & Ruvio, A. (2014). The Going Gets Tough, So Let's Go Shopping: On Materialism, Coping, and Consumer Behaviors Under Traumatic Stress. *Journal of Loss and Trauma*, 19, 426–441.

Thomassey, S., Happiette, M., & Castelain, J. M. (2003). Mean-term textile sales forecasting using families and items classification. *Studies in Informatics and Control*, 12 (1). 41–52.

Vroman, P., Happiette, M., & Vasseur, C. (2001). A Hybrid Neural Model for Mean-Term Sales Forecasting of Textile Items. *Studies in Informatics and Control*, 10 (2), 149-167.

Witt, C. A., & Witt, S. F. (1994). Forecasting International Tourist Flows. *Annals of Tourism Research*, 21, 612-628.

ZIOL-GUEST, K. M., DELEIRE, T., & KALIL, A. (2006). The Allocation of Food Expenditure in Married- and Single-Parent Families. *The Journal of Consumer Affairs*, 40 (2), 347-371.

J
T
A
T
M