

Statistical Model For Short-Term Forecasting Sparkling Beverage Sales In Southern Thailand

Wassana Suwanvijit, Thaksin University, Thailand
Chamnein Choonpradub, Songkhla University, Thailand
Nittaya McNeil, Songkhla University, Thailand

ABSTRACT

This study developed a simple statistical model for forecasting a company's sparkling beverage sales in the 14 provinces of Southern Thailand. Data comprised sales revenue from January 2000 to December 2006 obtained from the company. We fitted an observation-driven multiple regression model to log-transformed monthly revenue containing season of year (month), location and beverage flavour as factors, as well as lagged observations for the preceding four months. The model gave a r-squared of 0.95 and was effective for forecasting revenues for up to 12 future months. Using such models for forecasting sales revenue can assist company managers with planning more effectively.

Keywords: sales revenue, multiple regression model, Sales forecasting, Southern Thailand, Sparkling beverages.

INTRODUCTION

Forecasts are basic inputs for many kinds of decisions in business organizations. Forecasts help managers by reducing some of the uncertainty, thereby enabling them to develop more useful plans. Business forecasting involves more than predicting demand. Forecasts are also used to predict profits, revenues, costs, productivity changes, prices and law materials (Kran 2008). Sales forecasting helps to set sales targets and to plan production, marketing, distribution, etc. (Lingham 2004). Accurate sales forecasts facilitate effective and efficient allocation of scarce resources. Over-estimates of demand lead to several problems. First, excess inventory uses up valuable shelf space and leads to obsolescence. Next, scarce working capital blocked up in inventory carrying charges cannot be used for other purposes, such as for R&D or promotional expenses. Third, storage charges are incurred to store excess inventory in public or private warehouses. Finally, margins are reduced when excess inventory is removed through end-of-year clearance sales. Under-estimates of demand lead to a different set of problems. First, stock-outs lead to wasted shelf space. Next, insufficient inventory leads to lost sales and consequent lost margins. Third, failure to keep up with customer demand may necessitate the use of limited and expensive overtime production leading to lower profitability. Finally, and most importantly, the firm may lose customers when prospects facing an empty store shelf try an alternative brand or go to an alternative store and are satisfied by the competitive offering. Given the detrimental impact of inaccurate forecasts, marketers use a variety of sales forecasting techniques in order to forecast sales accurately. Lin and Hsu (2002) applied the Grey model to forecast sales of eight subcategory non-alcoholic beverages in Taiwan between 2001 and 2003. The advantages of the Grey model are that it can use only a few data to estimate an unknown system and it can use a first-order differential equation to characterize the unknown system's behavior. TSO (2002) described how to use the translation model (a mathematical model in which the process of human-language translation) for forecasting long-term electricity demand for the Republic of Ireland. Celia (2004) used a fuzzy forecasting model for women's casual sales. RNCOS (2007) is studying the South Korean food, beverages and tobacco market forecast until 2011 by using an in-depth study and evaluation of the past, current, and future market trends in the food, beverage and tobacco industry of South Korea. Newberne (2007) demonstrated the use of the Holt-Winters model (a model used when

data exhibits both trend and seasonality) on common healthcare data series. Dodds (2008) used the Bass model (applied from Riccati equation with constant coefficients) in long-term new product forecasting.

The purposes of this paper are to apply multiple linear regression analysis to short-term forecasting sparkling beverages’ sales revenue, to illustrate methods for graphing and statistical modeling available to businesses, and to demonstrate how each method can be implemented using freely available software.

METHODS

Steps to forecast are: 1) decide what to forecast, 2) evaluate & analyze appropriate data, 3) select & test the forecasting model, and 4) generate the forecast and monitor forecast accuracy over time.

In this study, the outcome was sales revenue. Determinants were flavour, branch location and seasonal period. Flavours were identified for five types of products, namely “cola product” (cola flavour), “colour products” (orange, red and green flavour) and “lime product” (lime flavour). There were two types of returnable packages (10 oz and 1 liter) and three types of non-returnable packages (1.25 liter, 2 liter and 325 ml). There are 20 branches in 14 provinces. Figure 1 gives an illustration of a path diagram for the data described above involving outcomes (sales revenue) and determinants composed of branch location, flavour, season (month), and lagged outcomes for preceding months.

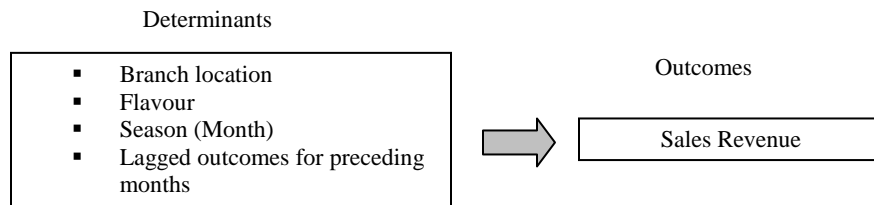


Figure 1: Data Path Diagram

Data for each month were available in computer files with records for sales revenue separated by flavour and branch location. After correcting or imputing data entry errors, records from years 2000 to 2006 were stored in an MySQL database. MySQL and Excel programs were used to create sales revenue in baht by month, flavour and branch location. All graphical and statistical analyses were performed using R (Venables and Smith 2004).

This quantitative research focused on using statistical graphics and statistical models, including multiple linear regression for forecasting sales revenue.

Sales revenues generally have skewed distributions, so it is essential to transform them by taking logarithms. Log-transformations can also ensure that statistical assumptions of symmetry and variance homogeneity of errors are satisfied. Sales revenue by month and branch location are sometimes zero for small branches and some adjustment is needed to avoid taking logarithms of 0.

The time series of total monthly sales revenue was plotted to assess trend and seasonal effects before choosing the best method and model for forecasting.

We fitted a multiple linear regression model to the data and compared results using R software (Venables and Smith 2004). Then, the model of log-transformed sales revenue, which contains seasonal effects and time-lagged terms, was applied for 12-month forecasting.

STATISTICAL MODEL

We fitted multiple linear regression models to the data, after transforming, using natural logarithms to ensure that statistical assumptions of normality and constant variance were satisfied.

The predictor variables compressed (a) the interactions between branch and flavour, (b) month of the year, and (c) the (log-transformed) sales revenues in the preceding four months. If Y_t is the sales revenue in branch i , flavour j , of year y , month t , s is the “season-month” (January, February,...) and ε_t is a series of independent normally distributed errors with mean 0, we write

$$\ln Y_t = \alpha + A_{ij} + \gamma_s + \delta_1 \ln Y_{t-1} + \delta_2 \ln Y_{t-2} + \delta_{11} \ln Y_{t-11} + \delta_{12} \ln Y_{t-12} + \varepsilon_t \tag{1}$$

where α , β and $(\delta_1, \delta_2, \delta_{11}, \delta_{12})$ are parameters in the model denoting an initial value, a trend, and two further coefficients denoting the influence of the sales in the previous four months, respectively, and $\gamma_1 = 0, \gamma_2, \gamma_{11}, \gamma_{12}$ is a set of seasonal effects indicating how the sales revenue varied with month of the year. Forecasts for $\ln Y_{t+k}$ (k months in the future) are obtained by substituting the estimated values for the coefficients into the right-hand side of (1), using the forecast values themselves for values of $k > 1$. However, to obtain forecasts for Y_{t+k} (1) must be transformed back by exponentiation and the forecast is then the mean of Y_t which has a log-normal distribution with expected value

$$E[Y_t] = \exp\left(\mu + \frac{1}{2}\sigma^2\right) \tag{2}$$

where μ and σ are the mean and standard deviation, respectively, of. Thus if σ is the standard deviation of the errors in the fitted regression model, the forecast of Y_{t+k} is

$$E(Y_{t+k}) = \exp\left(\alpha + A_{ij} + \gamma_s + \delta_1 \ln Y_{t+k-1} + \delta_2 \ln Y_{t+k-2} + \delta_{11} \ln Y_{t+k-11} + \delta_{12} \ln Y_{t+k-12} + \frac{1}{2}\sigma^2\right) \tag{3}$$

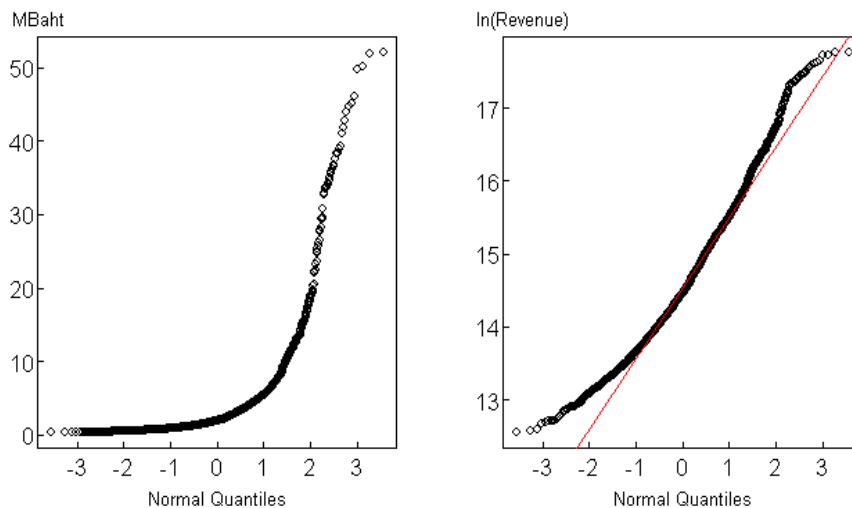


Figure 2: Sales revenue distribution before and after transforming to ln (Baht)

We used associative models that used explanatory variables to predict future sales revenue. The model is a multiple regression model since more than one predictor variable is used to predict sales. The goodness-of-fit of the sales forecasting model is checked with such statistics as r-squared and the standard error of regression relative to the mean and standard deviation of the response variable sales. Later, the partial explanatory power of each predictor variable is checked for expected sign and significance. The error terms are scanned for potential heteroskedasticity (serial auto correlation of the error term) in order to satisfy the assumptions underlying the use of multiple regression analysis.

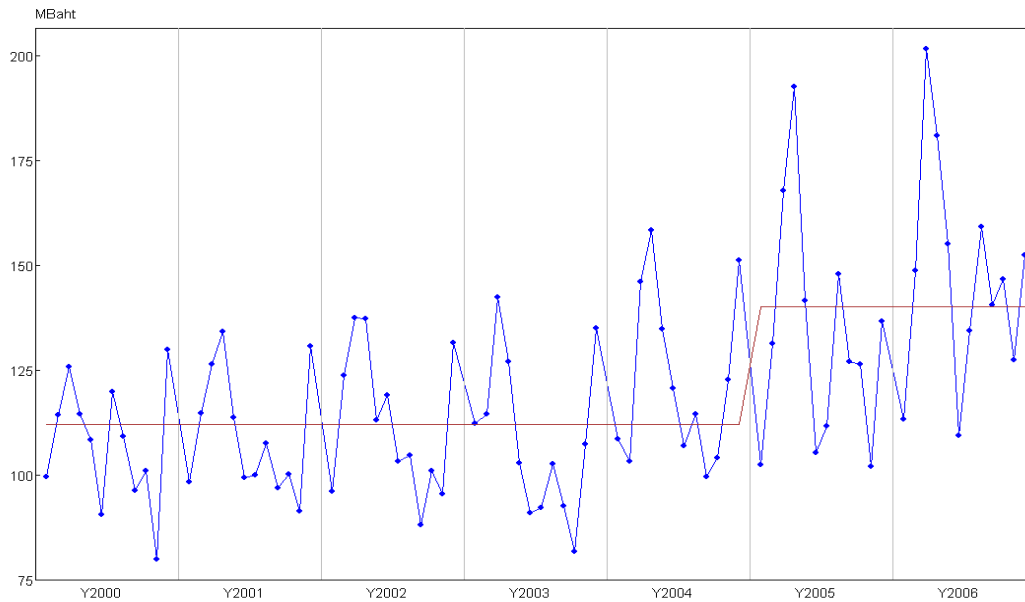


Figure 3: Time series of total sales revenue

PRELIMINARY ANALYSIS

Figure 2 shows the overall distribution before and after transforming the data by taking natural logarithms of sales revenue. It shows that the distribution is more symmetric after transforming the data. This log-normal distribution can be used to provide an estimate of consumption rate in each month.

Figure 3 shows a time series plot of the total monthly sales revenue from the years 2000 to 2006. The sales varied between 79.9 and 201.6 MBaht (average value for 2000–2006 is 120.1 MBaht). The sales continuously increased to peak at 201.6 MBaht in 2006. The data contains seasonal effects and there was an upward trend in the last few years.

RESULTS

Figure 4 shows plots of the observed total sales revenue with the natural log-transformed (left panel) and sales revenue (right panel) versus fitted value after fitting the model given by (3).

Figure 5 shows a plot of the time series of data based on the model given by (3). The monthly forecasts of sales revenue in the company during the twelve months of 2007 were 120.7, 142.7, 172.9, 179.8, 139.8, 121.4, 119.8, 136.2, 128.4, 128.1, 126.0 and 146.5 MBaht.

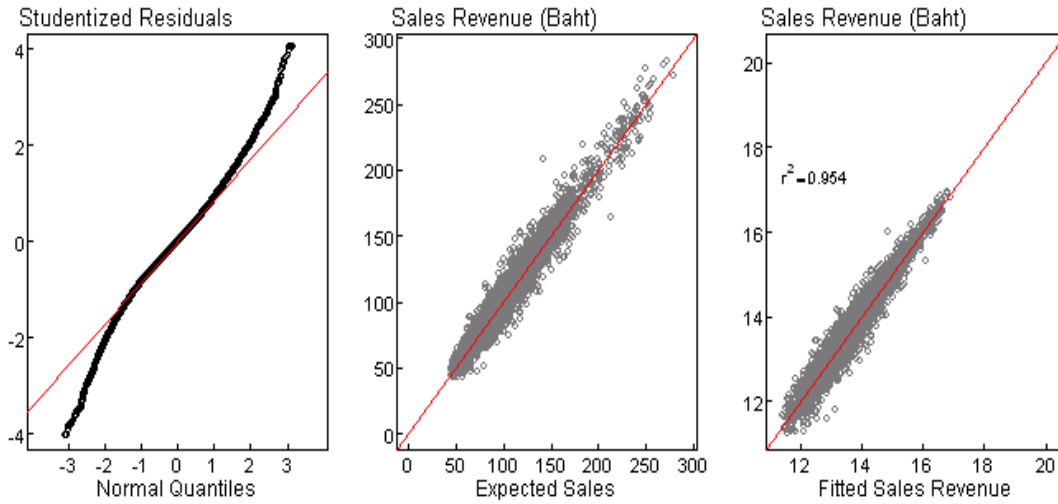


Figure 4: Plots of sales revenue and fitted values from fitting multiple linear regression model to logarithms of monthly sales revenue

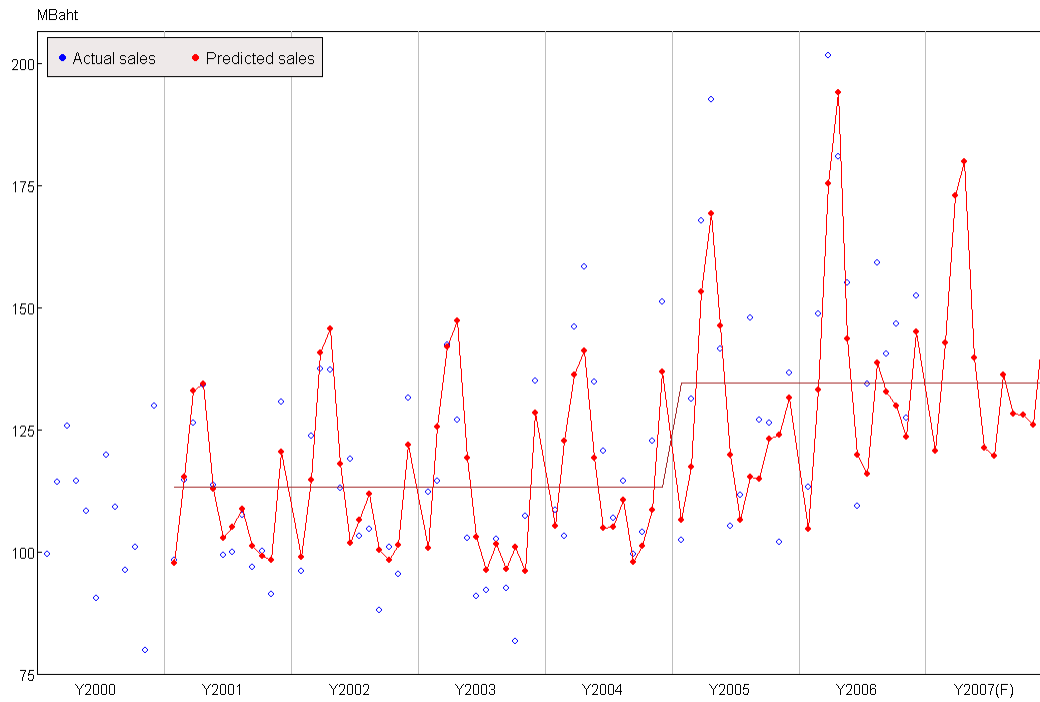


Figure 5: Comparison between actual and forecast total sales revenue

Table 1 shows the forecast and actual values of sales revenue (MBaht) and percentage errors in 2006. The average percentage error of predicted sales revenue was relatively small.

Table 1: Forecast and actual sales revenue of sales revenue (MBaht) in year 2006

Month	Forecast	Actual	% error
January	104.8	113.4	7.6%
February	133.2	148.7	10.4%
March	175.5	201.6	12.9%
April	193.9	181.0	7.1%
May	143.6	155.1	7.4%
June	120.0	109.4	9.7%
July	116.1	134.5	13.7%
August	138.8	159.1	12.8%
September	132.8	140.6	5.5%
October	129.9	146.7	11.5%
November	123.6	127.4	3.0%
December	145.1	152.6	4.9%
Average	138.1	147.5	6.4%

Figure 6 shows a plot of the time series of data (all flavours) with the forecasts based on the model given by (3) comparison by package types (returnable and non-returnable packages).

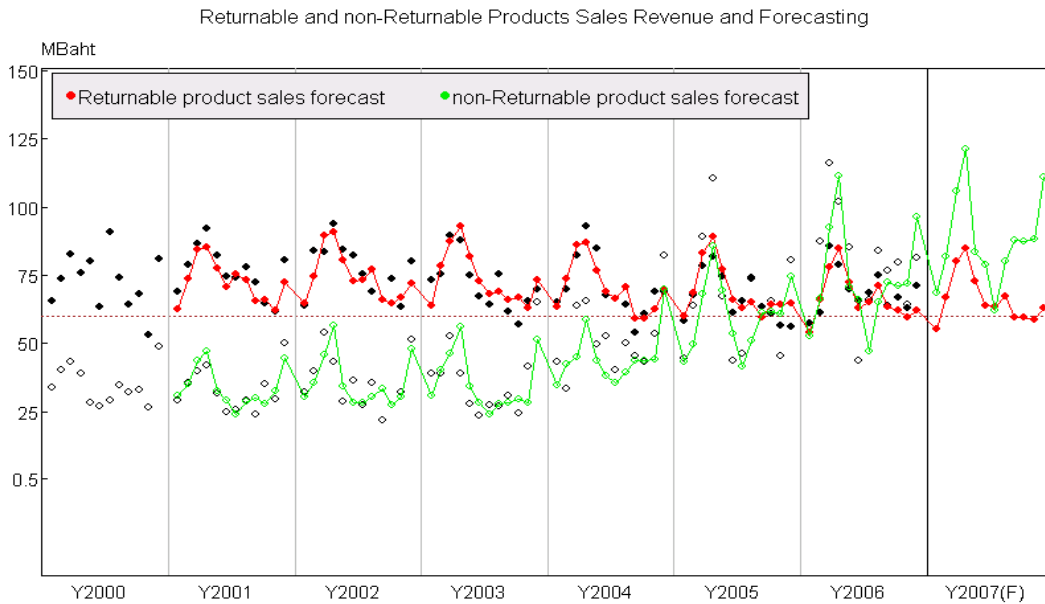


Figure 6: Comparison between actual and forecasting sales revenue grouped by package types

Figure 7 shows a plot of the time series of returnable products sales revenue with the forecasts based on the model given by (3) grouped by flavour (cola, orange, red, green and lime). There was a slightly downward trend in sales revenues.

Figure 8 shows a plot of the time series of non-returnable products sales revenue with the forecasts based on the model given by (3) grouped by flavour (cola, orange, red, green and lime). There was a moderate upward trend in sales revenue of each flavour.

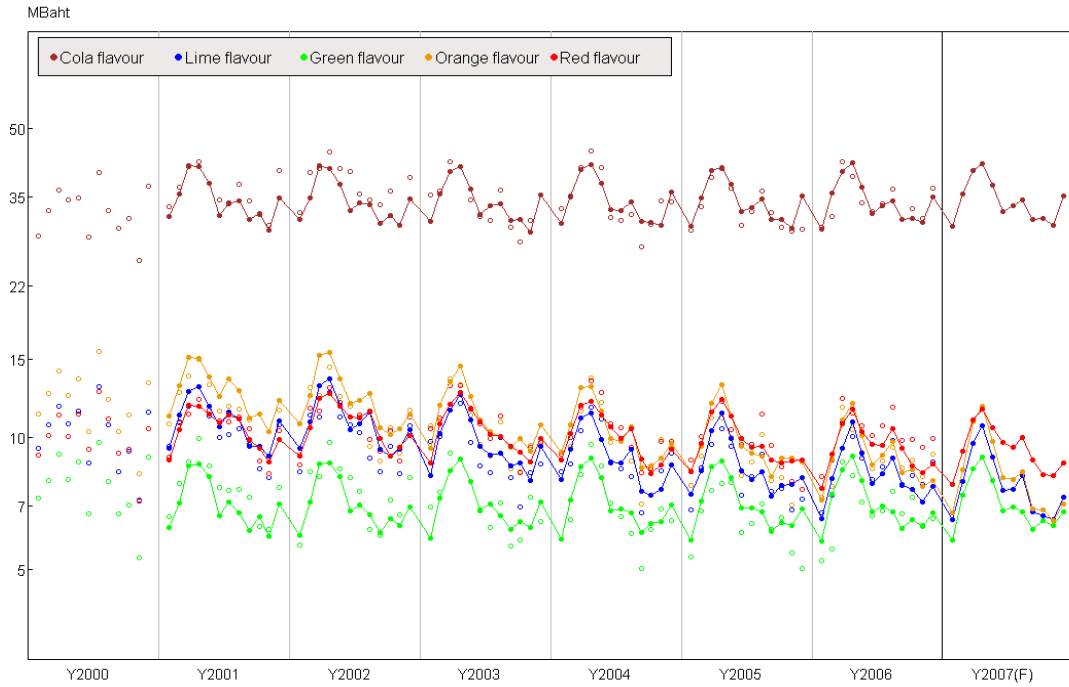


Figure 7: Comparison between actual and forecasting sales revenue of returnable products grouped by Flavour

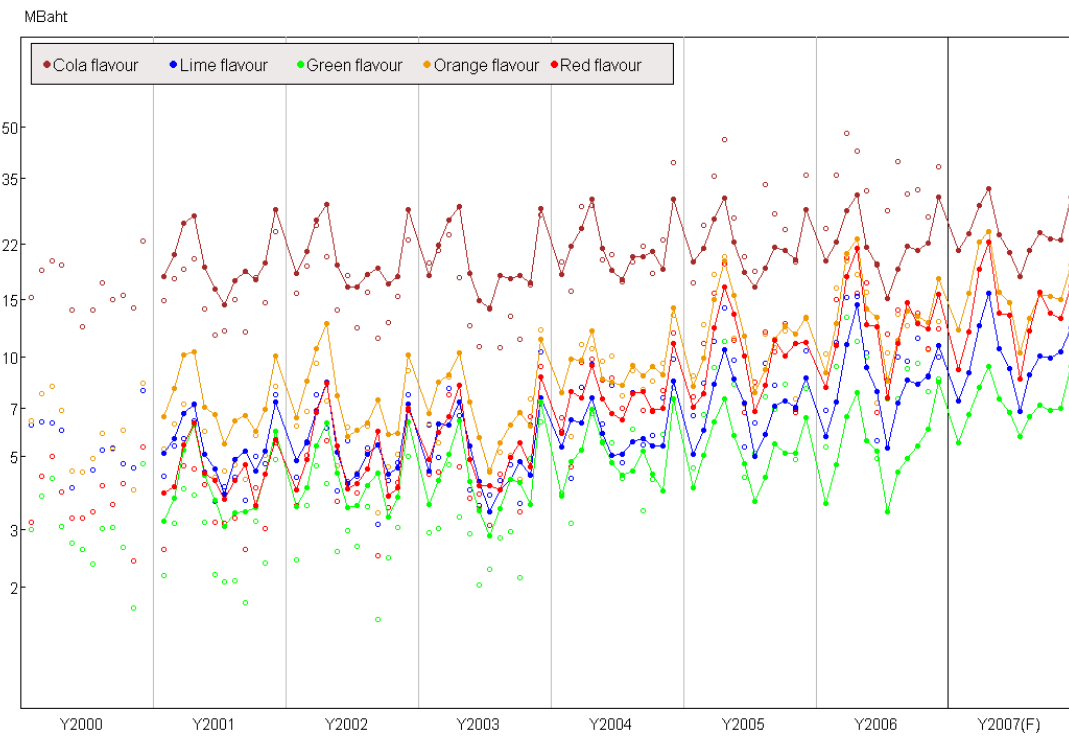


Figure 8: Comparison between actual and forecasting sales revenue of non-returnable products grouped by Flavour

DISCUSSION

Figure 3 reveals that the sales have increased substantially in the last few years. This could be due to an expansion in modern trade affecting consumer behaviour and life style of Southerners. The seasonal effects found in our study could be related both to regional climatic changes and to human activities. In the dry season (extending from February to April), hot weather and long holidays lead to greater consumption of sparkling beverages.

Figure 4 shows plots of sales after fitting an observation-driven multiple regression model to log-transformed monthly revenue containing season of year (month), location and beverage flavour as factors, as well as lagged observations for the preceding four months. . The model predicts the proportions in the 8,400 cells very well, giving a r-squared of 0.95.

Figure 5 shows a plot of the time series of data with the forecasts based on the model given by (3). The results from Table 1 indicate that the percent error of sales forecasting in year 2006, compared with actual sales, was approximately only 6.4%. This indicates that the model was effective for forecasting revenues for up to 12 future months.

Figure 6 shows a plot of the time series of data with the forecasts based on the model given by (3), grouped by package types (returnable and non-returnable packages). As we can see, the returnable products have a negative trend. However, the non-returnable product sales have a moderate growth of revenue, possibly due to consumer behaviour change. Most consumers prefer convenience packages such as polyethylene terephthalate plastic (PET) bottles and cans.

Figure 7 shows a plot of the time series of returnable products sales with the forecasts based on the model given by (3), grouped by flavour (cola, orange, red, green and lime). There was a slightly downward trend in sales revenues of all flavours, except the Cola product. Red flavour had more value than orange in recent years. Now there are a lot of orange juice products on the market, so consumers have more opportunities to buy this kind of product.

Figure 8 shows a plot of the time series of non-returnable product sales with forecasts based on the model given by (3), grouped by flavour (cola, orange, red, green and lime). There is a slight growth trend with the Cola flavour, while there is moderate growth with lime and colour flavours.

CONCLUSIONS

The observation-driven multiple regression model was effective for forecasting total sales revenues, including sales revenue grouped by flavour and package types for up to 12 future months. It can be applied for forecasting other business data. Using such models for forecasting sales revenue can assist company managers in planning more effectively.

ACKNOWLEDGEMENTS

We are grateful for Prof. Don McNeil for his helpful advice and suggestions. We also would like to thank Khun Dumrongrugs Apibalsawasdi for his guidance. This research was supported by a grant from the Thai Doctoral degree program of the Commission on Higher Education, Thailand.

AUTHOR INFORMATION

Wassana Suwanvijit is a lecturer at faculty of Economics and Business Administration, Thaksin University, Thailand, and a marketing advisor at Sparkling Beverage Company. She also is currently a PhD candidate in Research Methodology.

Chamnein Choonpradub PhD, Assistant Professor) is a lecturer for the Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkhla University (Pattani campus), Thailand.

Nittaya McNeil PhD is a lecturer for the Department of Mathematics and Computer Science, Faculty of Science and Technology, Prince of Songkhla University (Pattani campus), Thailand.

REFERENCES

1. Bureau., Indian Food and Beverages Forecast (2007-2011), June, 2007.
2. Caruana, Albert., Step in forecasting with seasonal regression: A case study from the carbonated soft drink market, Vol. 10, No. 2, pp.94-102, 2001.
3. Celia, Frank., Les, Sztandera., Balaji, Vemulapali., Asish, Garg., Amar, Raheja., Fuzzy Forecasting Model for Women's Casual Sales. National Textile Center Annual Report, November, 2004.
4. Cox, D.R. and Wermuth, N., *Multivariate Dependencies-Models, Analysis and Interpretation*, London: Chapman & Hall, 1996.
5. Dodds, Wellesley., An Application of the Bass Model in Long-Term New Product Forecasting. *Journal of Marketing Research* (JMR), 2008.
6. Hanke, John E., Wichern, Dean W., Business Forecasting. Pearson Education, New Jersey, 2005.
7. Higgins, Richard S., Kaplan, David R. and McDonald, Michel J., Residual Demand Analysis of the Carbonated Soft Drink Industry. *Empirica*, Vol.22, No.2, pp.115-126, 2005.
8. Kran, T.A., Equipment Productivity, LMDA, 2008.
9. Lin, Chin-Tsai., Hsu, Pi-Fang., Forecast of non-alcoholic beverage sales in Taiwan using the Grey theory. *Asia Pacific Journal of Marketing and Logistics*, Vol.14, No.4, pp.3-12, 2002.
10. Lingham, Leo., Marketing-Acquisition & Analysis of information, 2004.
<http://en.allexperts.com/q/Marketing-1090/Marketing-Acquisition-Analysis-information.htm>. [March 20, 2009]
11. Newbeme, Joan H., Captain, USAF., MSCHolt-Winters Forecasting in Healthcare, Mike O'Callaghan Federal Hospital, 2007.
12. R Development Core Team., R: A language and environment for statistical computing, Vienna, 2007.
13. RNCOS., South Korean Food, Beverages and Tobacco Market Forecast till 2011, December, 2007.
14. Tinney, Mary-Colleen., Retail Sales Analysis: Champagne and Sparkling Wine Sales Rising, Wine Communications Group, 2007.
15. TSO., A New Methodology for forecasting long term electricity demand for the Republic of Ireland. *Electricity Demand Model Review*, Generation Capacity Planning, September, 2002.
16. Venables, W.N., Ripley B.D., Modern Applied Statistics with S. Springer, Queensland, 2002.
17. Venables, W. N., Smith, D.M., An Introduction to R, 2004.
<http://cran.r-project.org/doc/manuals/R-intro.pdf> [March 30, 2007]
18. Williams., Developing Sales Forecasts, 2007. <http://web.csustan.edu/market/williams/Chapter%2010.htm> [March 14, 2008]

NOTES