Chapter 3

Data analysis and models for sales revenue

The analysis and forecasting in this chapter are based on conventional statistical methods. Graphical method is also used for preliminary and summarizing the data as well as communicating the results of an analysis and forecasting. All of the graphical, statistical models fitting and assessing the forecasting results were carried out using programs written in *R* software.

3.1 Characteristics of the data and area of study

In this study, data was collected routinely in 14 provinces of Southern Thailand by the sparkling beverages company. The Southern Thailand market can be grouped by branches location and consumer preferences into tourist area, Muslim area and other areas as shown in Figure 2. There are 20 branches in 14 provinces. Provinces with more than one branch were Surat Thani, Nakorn Sri Thammarat, Phang-nga, Chumphon and Narathiwat.

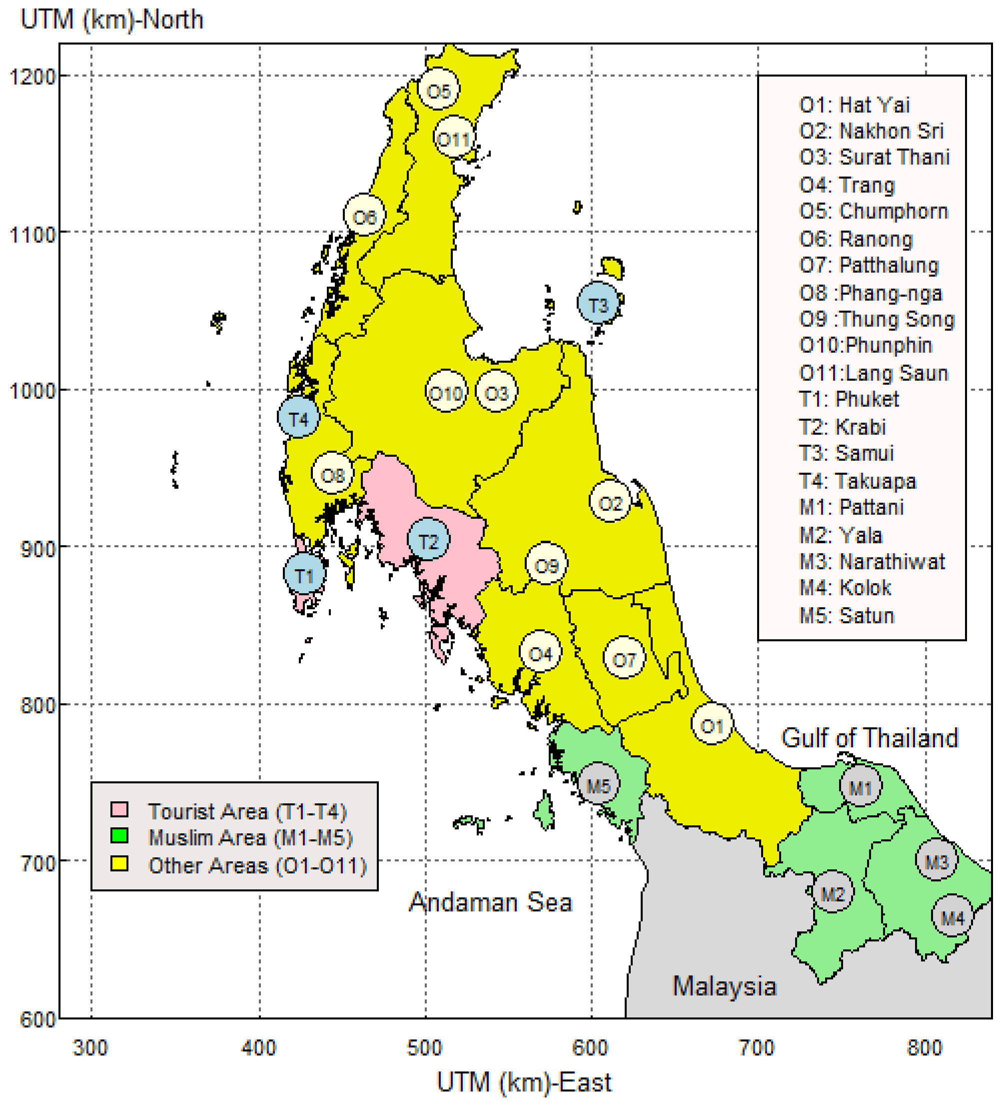

Figure 2: Branch locations and their definitions as tourist, Muslim and other areas

Flavours were identified for three types of products, namely "cola flavour", "colour flavour" (orange, red and green) and "lime flavour". There were two types of returnable packages (10 oz and 1 liter) and three types of non-returnable packages (1.25 liters, 2 liters and 325 ml).

We created sales revenue grouped by branch location, flavour, monthly (84 months from January 2000 to December 2006), quarterly (28 quarters from January-March 2000 to October-December 2006) and yearly (7 years from 2000 to 2006).

Sales revenue data have positively skewed distributions. There was a few of data entry errors. The data were available in computer files with records for sales revenue. After correcting or imputing the errors, records were stores in MySQL database. MySQL and Microsoft Excel were used to code and create sales revenue in each group.

3.2 Distribution of sales revenue

Linear models produce estimates with good statistical properties when the relationships are, in fact, linear, and the errors are normally distributed. In some cases, when the distribution of the response is skewed, we need to transform the response, using, such as square root, logarithm, or reciprocal transformations in order to produce a better fit.

Sales revenue data generally have positively skewed distributions so it is conventional to transform them by taking natural logarithms to ensure that statistical assumptions of normality and constant variance were satisfied. Cells with zero counts were adjusted to avoid the problems of taking the logarithm of zero. Figure 3 shows the overall distribution before and after transforming the data by taking natural logarithms of sales revenue. The sales revenue data were symmetry and variance homogeneously for residually. This log-normal distribution can be used to provide an estimate of consumption rate and sales forecasting.

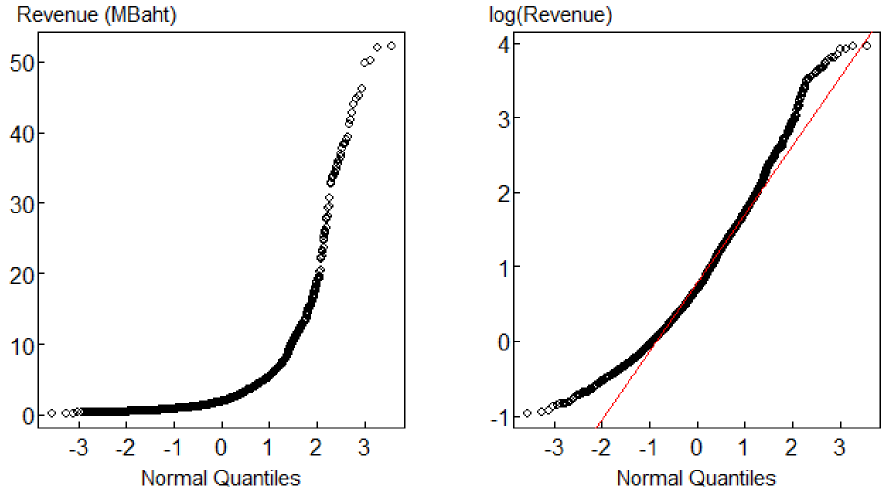

Figure 3: Sales revenue distribution before and after transforming to natural logarithms

3.3 Distribution of per capita consumption rates

Per capita consumption rates were calculated based on the sales revenue (in 1,000 Baht) per 1,000 population (exclude tourist), according to the 2000 Population and Housing Census of Thailand, for each branch location. The per capita consumption analysis is based on the graphical method and statistical methods. First, the characteristics of the data are described. After that, distributions of per capita consumption rates are explained. Finally, the model (Equations 2-4, Chapter 2) was used to fit the data.

Table 1 shows an annual per capita consumption rate in each branch location compared with population in the corresponding catchment area. The average annual per capita consumption rate was 297 Baht for the total of 8.087 million population.

Table 1: Annual per capita consumption rates grouped by branch location

Branch	Symbol	Population*	Consumption**
Nakhon Sri Thammarat (excl. Thung Song)	Nk	1,375	92
Hat Yai	Hy	1,258	267
Surat Thani (excl. Phunphin & Samui)	Sr	746	183
Narathivat (excl. Kolok)	Nv	600	69
Pattani	Pn	599	92
Trang	Tr	597	142
Phatthalung	Pl	501	84
Yala	Yl	418	172
Chumphorn (excl. Lang Saun)	Cp	380	76
Krabi	Kb	338	169
Phuket	Pk	250	697
Satun	St	219	144
Phang-nga (excl. Takuapa)	Pg	192	105
Ranong	Rn	162	208
Thung Song	Ts	148	325
Phunphin	Pp	91	428
Lang Saun	Ls	69	291
Kolok	Kl	65	386
Takuapa	Tk	44	583
Samui	Sm	35	1,424
Total		8,087	
Average annual per capita consumption (Baht per population)			297

* Population (x1000) ** Annual per capita consumption (Baht per 1,000 population)

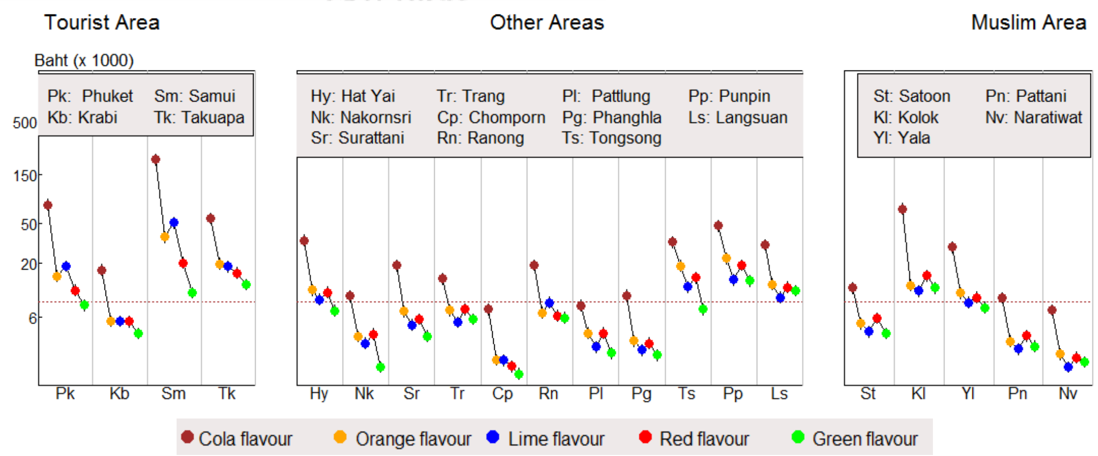

Figure 4: Per capita consumption rates grouped by branch, area and flavour

The Southern Thailand market can be grouped by branches location and consumer preferences into tourist area, Muslim area and other areas. Figure 4 thus shows the per capita consumption rates per 1,000 population of each flavour grouped by branch location and consumer preferences for Muslim, tourist, and other areas. Areas with different proportions of their population being Muslim, had different beverage preferences due to consumers preferring colour and lime products. Samui (Sm) and Phuket (Pk) branches had higher rates than the other branches, while Nakhon Sri Thammarat (Nk) branch had the lowest per capita consumption rate.

3.4 Fitting models for sales revenue

Based on the sales revenue data routinely reported in each branch location from 2000-2006, multiple linear regression models of log-transformed per capita consumption rates (Equations 2-4, Chapter 2) were used to assess the effects of branch location, flavour and season of the year. The results presented in this section also appear in Appendix I.

Table 2: Result of fitting multiple linear regression model

Model	R^2	df	RSS
1	0.913	2767	211.9
2	0.959	2691	100.7
3	0.973	2577	65.1

Table 2, the results indicate that Model (3) is the best regression model to use in this study because it has the highest r-squared value, which was very close to 1.

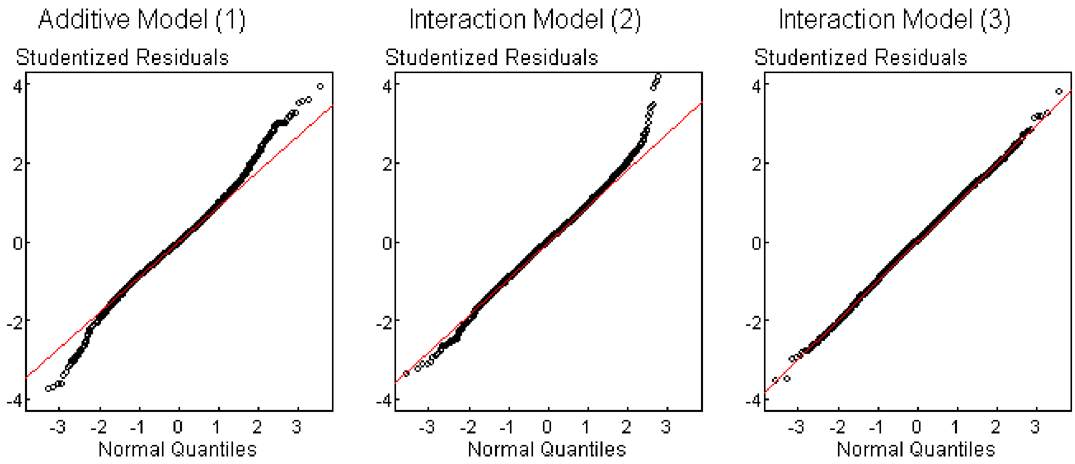

Figure 5: Comparison between residual plots for three models

Figure 5, it is clear that the multiple linear regression model (3) fits the data very well as evident by the linear in the residuals plot (right panel).

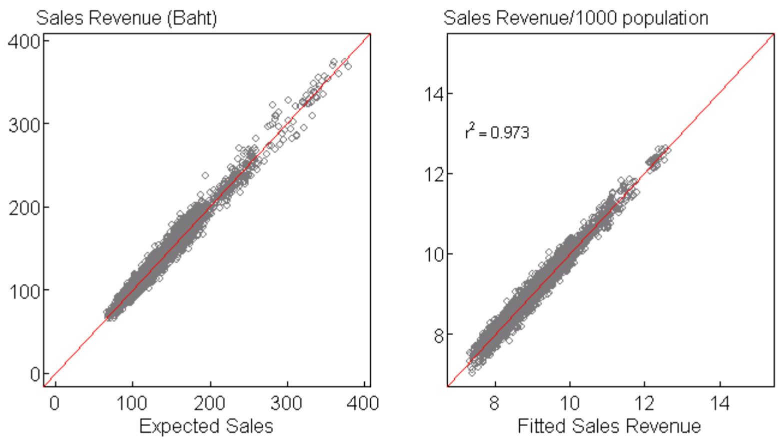

Figure 6: Plots of sales revenue and fitted values for interaction model

Figure 6 shows plots of sales revenue in Baht (left panel) and per 1,000 population (right panel) versus fitted value from using Model 3. The model predicts the proportions in the 2800 cells very well.

3.5 Time series of sales revenue

The time series of total monthly log-transform sales revenue was plotted to assess trend and seasonal effects before choosing the best method and model for forecasting. During the study period from January 2000 to December 2006, the time series of total sales revenue are plotted in Figure 7. The data contained seasonal effects and there was an upward trend in the last few years due to an impact of the economic growth. Consumers had more purchasing power to buy more beverages. An expansion in modern trade outlets also affected consumers buying behaviors and their life style.

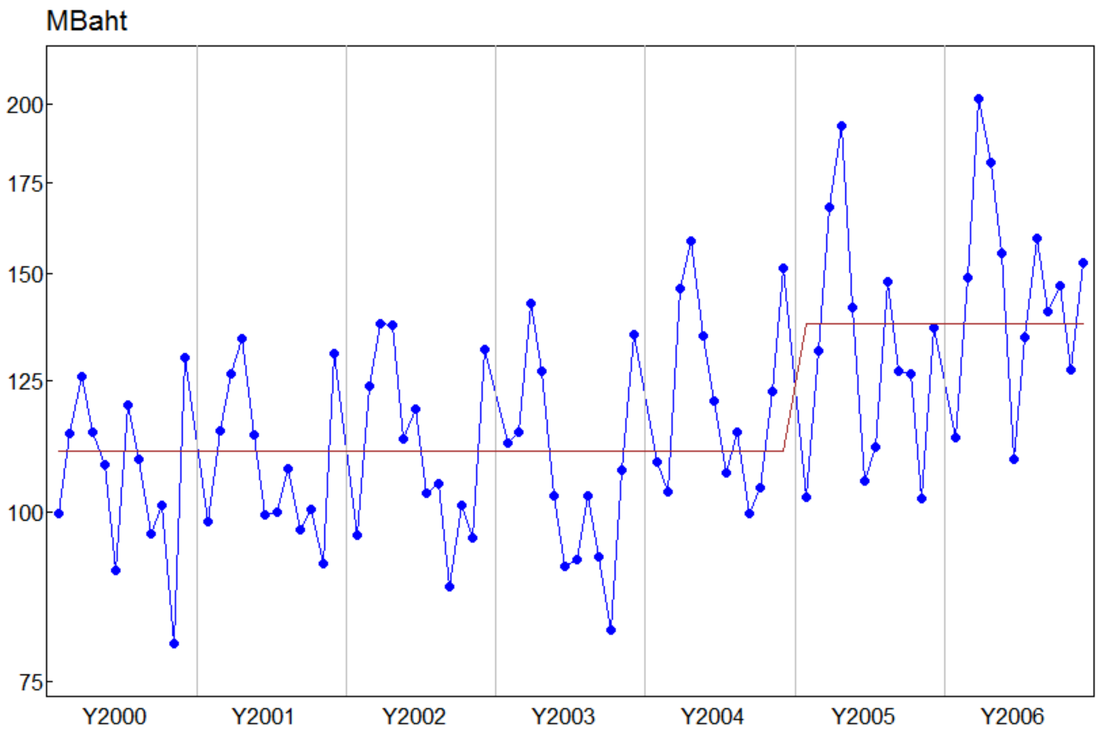

Figure 7: Time series of sales revenue during years 2000-2006

3.6 Short-term sales forecasting

Based on the sales revenue data routinely reported in each flavour and package type from 2000-2006, observation-driven multiple linear regression models of log-transformed sales revenue (Equations 5-7, Chapter 2) were used for short-term total sales forecasting, including the sales grouped by flavour and package type. The results presented in this section also appear in Appendix II.

Figure 8 shows a comparison between the total actual sales revenue during years 2000-2006 and 12 future months forecasting results during years 2001-2007. The model predicts the proportions in the 8,400 cells very well, giving a r-squared of 0.95**.** In addition, the model can give the 12-months sales revenue in year 2007 forecasting that contains the seasonality. In the dry season (extending from February to April), hot weather and long holidays lead to greater consumption of sparkling beverages.

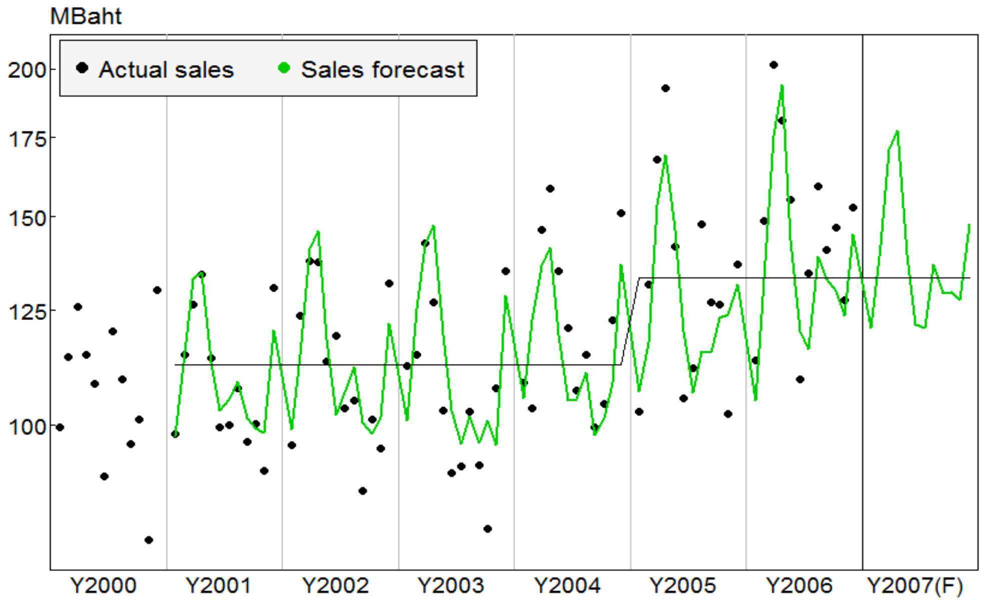

Figure 8: Comparison between total actual sales and forecasting results

Figure 9 shows a comparison between the total actual sales revenue during years 2000-2006 and 12 future months forecasting results during years 2001-2007 grouped by package type (returnable and non-returnable packages). The returnable package product was in a negative situation trend whereas the non-returnable package product had a moderate growth of sales revenue. Most consumers preferred convenience packages such as PET bottles and can.

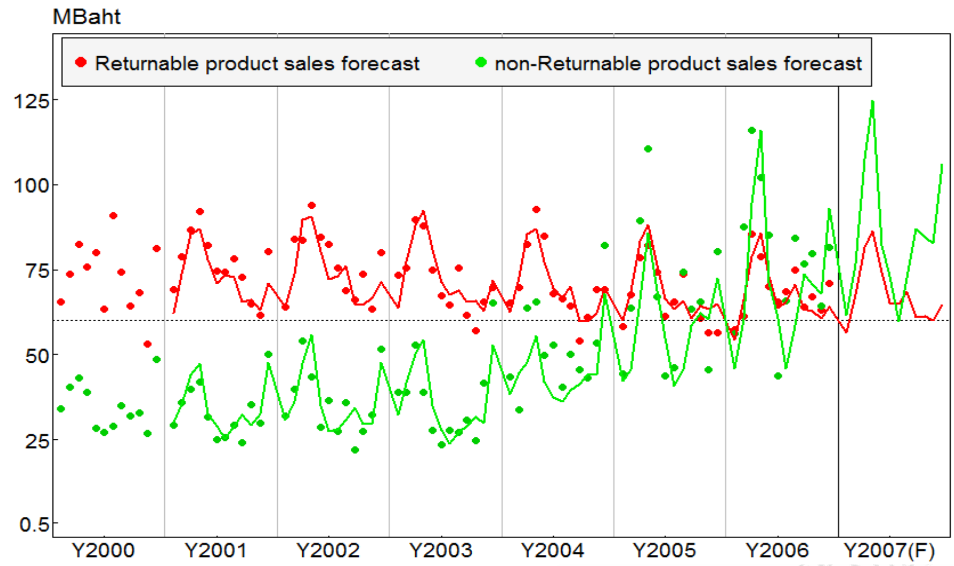

Figure 9: Comparison between actual sales and forecasts by package type

Figure 10 shows a comparison between the total actual sales revenue during years 2000-2006 and 12 future months forecasting results during years 2001-2007 grouped by flavour (Cola, Red, Orange, Green and Lime) and package type (returnable and non-returnable packages). There was a slight downward trend in returnable package product sales revenues of all flavour. In the early years orange flavour had the most value in colour products but red flavour became more value in the last few years. It is possible that there were lot of orange juice products in the market, so consumers had more opportunities to buy this kind of product. There was a moderate upward trend in non-returnable package product sales revenues of all flavours due to PET and can packages became more popular than the glass bottle one.

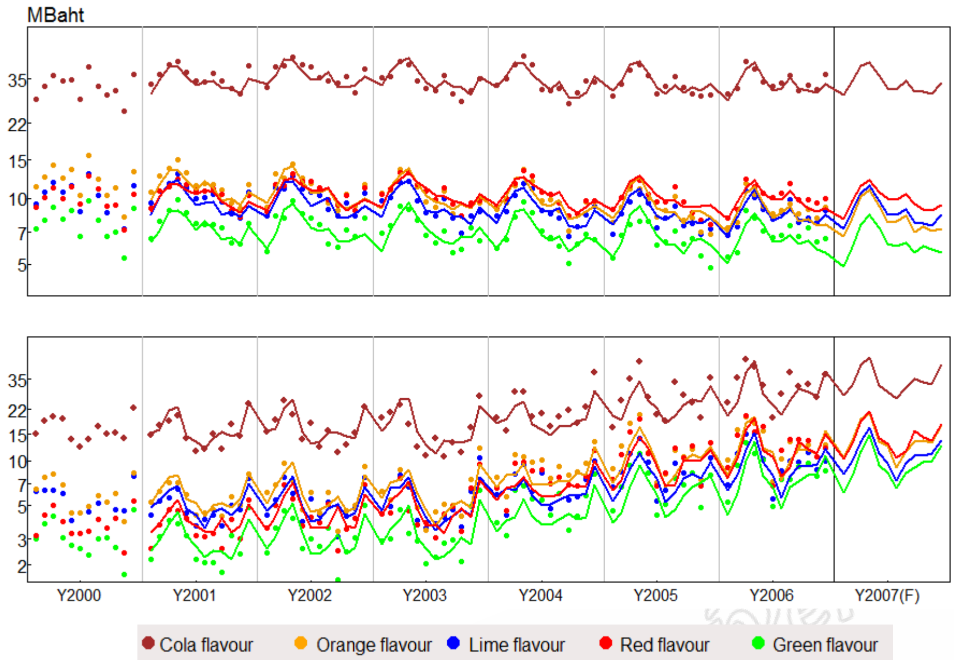

Figure 10: Comparison between actual sales and forecasts by flavour and package type (the returnable package in top panel and the non-returnable package in bottom panel)

Table 3 reveals the comparison between forecasting errors and actual sales, including the absolute percent errors in each month in year 2006. The model gave a r-squared of 0.95 and mean absolute percent error 6%.

Table 3: Forecasting errors from short-term forecasts

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

3.7 Sales trend and branch effects

There are 20 branches in 14 provinces. Figure 11 shows the monthly (60 months from January 2000 to December 2004) sales revenue trend and branch effects.

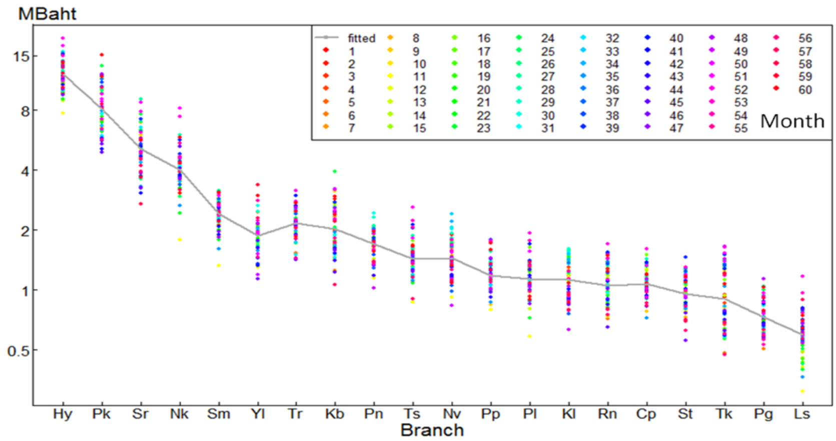

Figure 11: Monthly sales revenue plot by branch location

Figure 11 indicates that the monthly sales in each branch location have increased substantially. However, each branch has different patterns in the growth rate compared with the others. So, the branch location is a factor of interest since it affects the sales trend.

3.8 Long-term sales forecasting

Based on the sales revenue data routinely reported in each branch location from 2000-2004, Lee-Carter 1-3 components extension and Holt-Winter models of log-transformed sales revenue (Equations 8-15, Chapter 2) were used for long-term sales forecasting grouped by branch location, including the separate forecasting. The results presented in this section also appear in Appendix III.

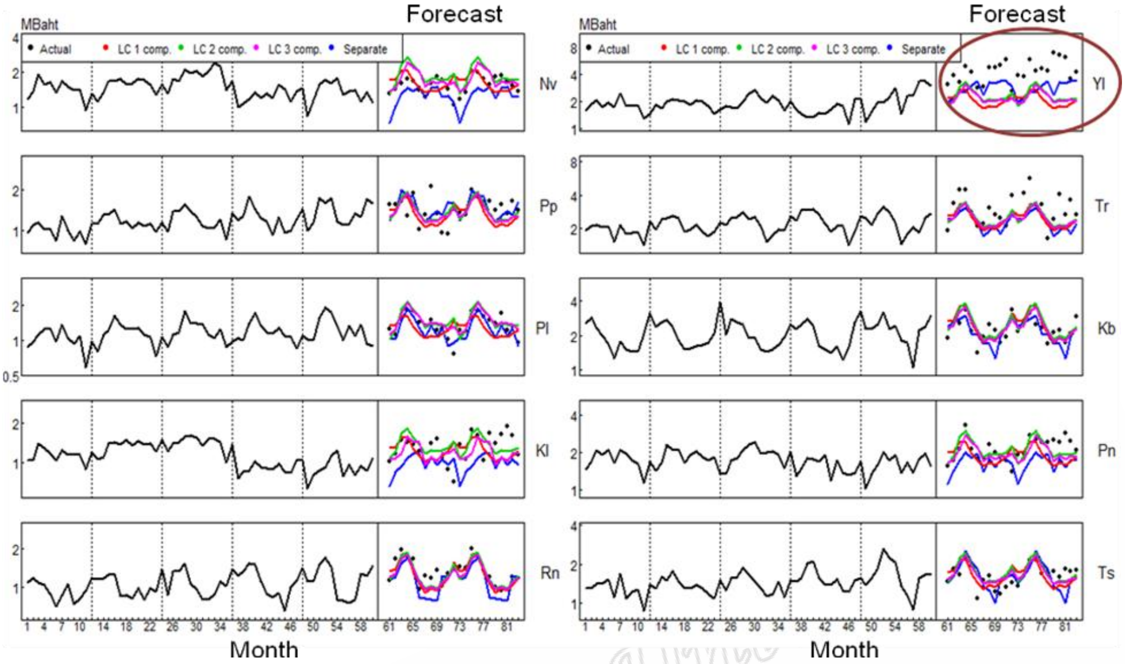

Figure 12: Comparison of actual sales and forecasting results in each model grouped by branch location

Figure 12 shows a comparison between the actual sales revenue during years 2000-2004 and 24 future months forecasting results during years 2005-2006 in each model. Branch locations were labeled using the symbol described in Table 1. As can be seen that there is a rapid trend in Yala branch during the years 2005-2006 due to the market execution and distribution penetration by the company. This trend may caused high forecasting errors. So, we evaluated the forecast accuracy by comparison both the total forecasting error and the errors excluding the Yala (Yl) branch.

Table 4 shows the forecasting error analysis by MAD, MSE and MAPE comparison between each model. It also compares total errors and the errors without forecast the Yala branch. The forecasting errors using Lee-Carter model with 3 components extension are less than using the other models. The forecasting errors exclude Yala branch are less than the forecasting errors from all branches.

Table 4: Forecasting errors from long-term forecasts

Methods	MAD		MSE		MAPE	
	Total	Excl. Yala	Total	Excl. Yala	Total	Excl. Yala
Lee-Carter 1 component	0.22	0.20	0.09	0.06	1.52	1.35
Lee-Carter 2 components	0.20	0.17	0.07	0.05	1.35	1.20
Lee-Carter 3 components	0.19	0.17	0.07	0.05	1.32	1.36
Separate forecast	0.22	0.20	0.08	0.07	1.47	1.39

Figure 13 shows the distribution of total forecasting errors in each model in the box plot compares with the distribution of forecasting errors excluding Yala branch. It is clear that the Lee-Carter model with 3 components extension is a suitable model since the median value of error (as can be noticed from the line in box plot) is very small. The model also gives the less far out of errors than other models.

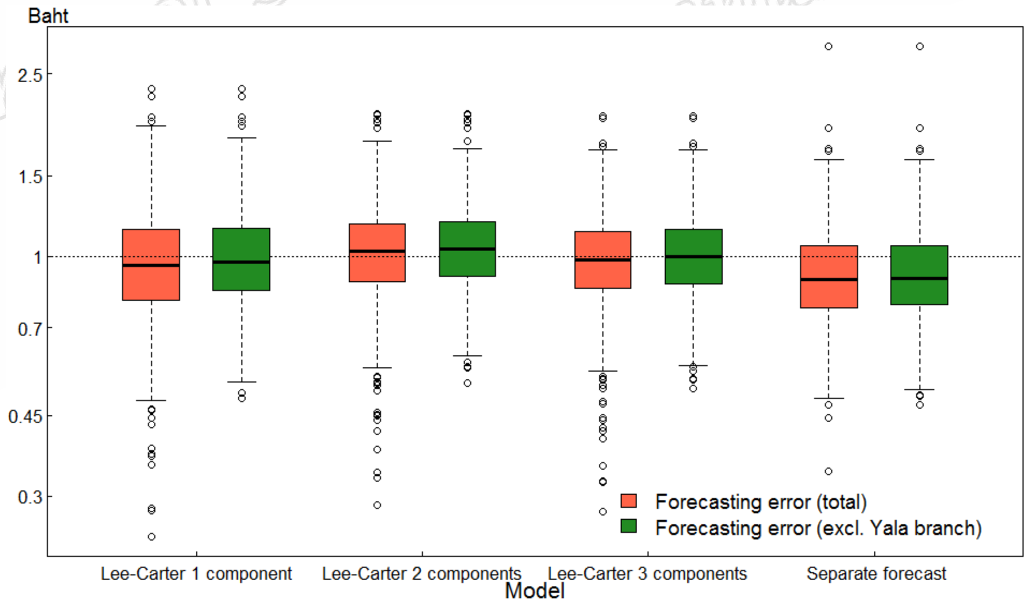

Figure 13: Comparison of forecasting errors between each method from total sales forecast and the forecast exclude Yala branch

Figure 14 shows time series of the actual sales during years 2000 - 2004 (60 months) by branch and the sales forecast from fitting Lee-Carter model with 3 components extension during years 2005 – 2006 (24 months ahead). The plot reveals that the forecasting sales is appropriate fitted to the last period trends and the model can give the reasonable trend in the 24 future months.

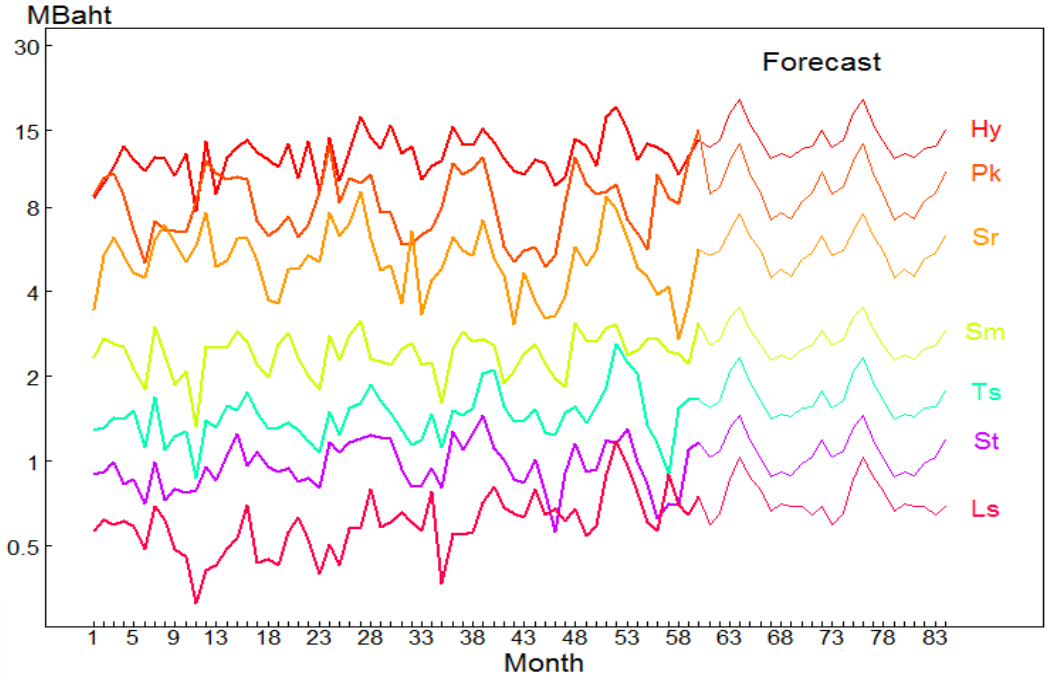

Figure 14: Sales forecast during years 2005-2006

It is useful to consider the 95% prediction intervals with the point forecast. Figure 15 shows some of the forecasting results with 95% prediction intervals comparison between Lee-Carter model with 3 components extension and separate forecasts. As can be seen that although there are some of the observed data are outside of the predicted lined but they are well-covered by the prediction intervals. The plot displays that the forecasting results using the Lee-Carter model with 3 components extension are well-covered by the prediction intervals than the separate method.

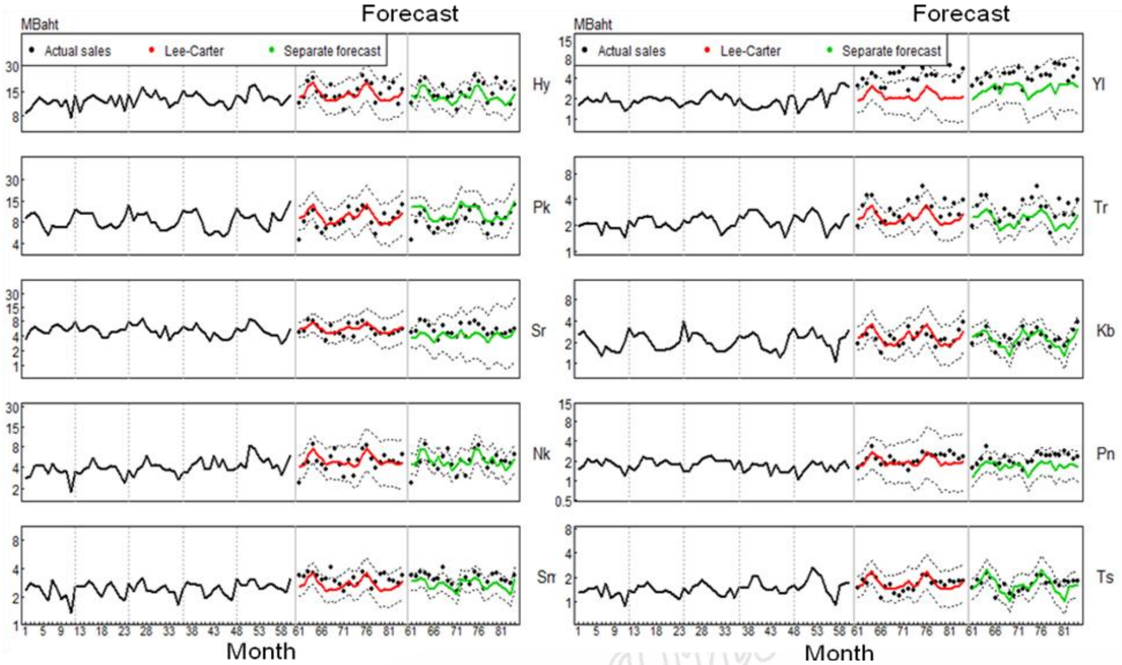

Figure 15: Sales forecast by branch location with 95% prediction interval