Jurnal
Teknologi

Full Paper

THE EFFECTS OF SPECIAL EVENTS ON REGRESSION FOR SUBCOMPACT CAR SALES IN THAILAND

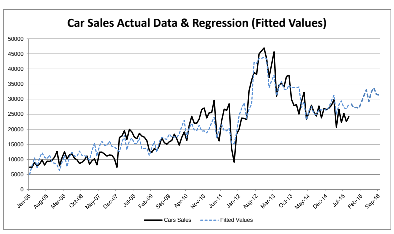
Witchaya Rattanametawee[a], Chartchai Leenawong[a]*, Ponrudee Netisopakul[b]

[a]Department of Mathematics, Faculty of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520 Thailand
[b]Knowledge Management and Knowledge Engineering Lab., Faculty of Information Technology, King Mongkut's Institute of Technology Ladkrabang, Bangkok, 10520, Thailand

Article history
Received
17 June 2016
Received in revised form
20 September 2016
Accepted
18 October 2016

*Corresponding author
chartchai.le@kmitl.ac.th

Graphical abstract



Abstract

This research proposes a method to dealing with multiple linear regression that integrates the seasonality as well as the effects of some special or unanticipated events for sales figures. The method is then applied to the car sales figures in Thailand after having been through the 2011 national big flood and the 2011-2012 government's initiative tax-incentive program for boosting the automobile industry. Besides Thailand's Gross Domestic Products (GDP) and the 12-month Loan's Interest Rate as explanatory variables, seasonal dummy variables along with the proposed special event variables and appropriate event tagging are incorporated. The statistical results obtained from the proposed regression model with seasons and events, compared to the models with neither seasons nor both yields highest adjusted coefficient of determination (R-squre) and accuracy (MAPE).

Keywords: Multiple Regression, Forecasting Event Modelling, Car Sales, Thailand

1.0 INTRODUCTION

In the past several years, subcompact car sales figures in Thailand have been fluctuating considerably due mostly to two crucial events. One is the severe flooding across the country that occurred approximately during the later months of the year 2011 and the other is the tax-incentive first-car buyer program [1] that followed just slightly later for another 1-plus year.

The program was supposed to support the almost devastated automobile industry by reimbursing car taxes to Thai people who reserved for buying their first-owned subcompact cars with not over 1,500-litre engines during the time period 16 September 2011 to 31 December 2012. The reimbursement limit was 100,000 baht per car per person.

The monthly subcompact car sales figures (in Million Baht) in Thailand from January 2005 to September 2015 obtained from the Office of Industrial Economics, Ministry of Industry, Thailand [2], [3] are shown in Figure 1. All four components of the time series [4], namely, trend, seasonal, cyclical, and irregular components are obviously revealed in the figure as well.

Multiple regression [5] has been used in model fitting and forecasting for over a century in various areas such as water resources planning [6], Korean box-office revenue [7], gold prices [8], or even automobile sales in the US [9] and Taiwan [10]. Recently, Thailand car sales forecasting [11] has been conducted using the time-series decomposition method with special event flags. As the name of the method suggested, it considered just one contributing factor, that is, time. Other possible economic causes have been discarded.

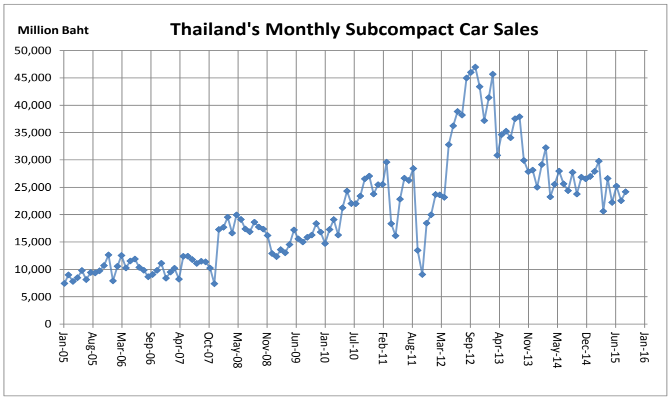


Figure 1 Thailand's monthly car sales (in Million Baht) from January 2005 to September 2015 [2]

In this research, it is, therefore, aimed to find the economic causal factors that influence the dependent variable, i.e., subcompact car sales in Thailand. In addition, other dummy variables are added to the regression so as to account for the seasonality. Besides, the effects of unanticipated events such as the case of Thailand's big flood and the effects of some short-noticed promotional programs such as the case of the tax-incentive program will be accounted for via the proposed special (or unanticipated) event variables. The big flood and the tax-incentive program caused the subcompact car sales data to fluctuate extremely during the period from September 2011 up until January 2014 as seen also in Figure 1 Thailand's monthly car sales (in Million Baht) from January 2005 to September 2015 [2]

The rest of this paper is organized as followed. In Section 2, the methodology to use in this research including multiple linear regression, how to deal with seasonality, and how the special event variables are constructed, is described. In Section 3, the experimental design is explained in detail followed by the experimental results in Section 4. Finally, the conclusion of this research is provided in Section 5.

2.0 METHODOLOGY

In this section, the methodology to be used in this research as well as the experimental design are presented. The methodology includes multiple linear regression, seasonality in regression, and the proposed method for special events. Then, applying the above methodology to the real data, namely, Thailand's car sales figures is explained.

2.1 Multiple Linear Regression

Linear regression is a modeling approach for examining the linear relationship between a variable of interest called the dependent variable y and an independent variable x (simple regression) or a set of independent variables $x_1, x_2, \ldots, x_k$ (multiple regression) that can explain the amount of variation in y. Equation (1) displays the probabilistic linear equation of multiple regression.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon \qquad (1)$$

where

k is the number of independent variables,

β_i is the contribution of the independent variable x_i,

and ε is the error or residual term.

The Ordinary Least Squares (OLS) method used for the parameter estimation minimizes the sum of squared errors for the sample. Subsequently, b_i is used in place of β_i and the deterministic prediction equation becomes Equation (2) as follows.

$$y = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k. \qquad (2)$$

In our case here, the b_i coefficients are obtained through a statistical package.

Turning now to the choices of the independent variables, since GDP is a good indicator of a nation's overall economic condition [12], a better economy positively affects the overall consumption and definitely stimulates more car sales. In contrast, since many car buyers do not pay in cash, the interest rate of the loan to pay for the car negatively affects the buying decision [13]. The higher the loan's interest rate is, the lower the car sales will be. Confirmed by the correlation coefficients between the variables with the correct signs shown in Table 1, Thailand's GDP and the loan's interest rate are considered good predictors for the car sales.

Revealed also by Table 1, the two independent variables, GDP and the loan rate, are not highly linearly correlated to each other. Therefore, the two predictors can be included in the multiple regression models.

Table 1 Correlation coefficient table

Correlation	Car Sales	GDP	Loan Rate
Car Sales	1.00	0.64	-0.46
GDP	0.64	1.00	-0.51
Loan Rate	-0.46	-0.51	1.00

2.2 Seasonality in Regression

For seasonal data such as our car sales figures here, seasonality [14] must be taken into account for. One could compute a seasonal index for each and every season and use these indices to deseasonalize all of the original data before undertaking regression.

However, in this research, another method of dealing with seasonality in regression [15] is employed. It is done so by adding to the regression n–1 dummy variables, probably named $s_2, s_3, \ldots, s_n$, for the n seasons. Each dummy variable s_i is treated as another independent variable except that its values can only be either 1 if season i is observed or 0 otherwise. Thus, the regression equation for the data with k predictors and n seasons simply becomes Equation (3) shown here.

$$y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k + a_2s_2 + a_3s_3 + \ldots + a_ns_n. \quad (3)$$

where a_i represents the coefficient of each seasonal variable.

Since we have monthly data here, 11 dummy variables are required; for example, $s_2, s_3, \ldots, s_{12}$, with s_2 equal to 1 if it is February, s_3 equal to 1 if it is March. Apparently, when all the 11 seasonal dummy variables are equal to 0, it refers to January.

2.3 The Proposed Method for Special Events

During some sales periods, unexpected or short-noticed events that have enormous impacts on sales figures can happen. National disasters such as floods, Tsunami floods, earthquakes, and windstorms are extreme examples that can cause unguarded businesses millions of dollars. Other examples of special events are some short-noticed gigantic promotional campaigns having also massive impacts on some industries. It is easy just to remove the data affected by these special events from the regression model fitting. Nevertheless, keeping those data intact and finding a way to tag them appropriately will be more useful, especially if in the future, these events can somehow be predicted to happen again. Businesses then can prepare for what it is worth, either favorable or unfavorable to them.

Analogous to the dummy variables for seasonality, special event variables are created so that one variable refers to one specific event that can last for more than one period. For example, e_1 refers to the flooding period that lasts from September through November 2011, resulting in e_1 taking the values of 1's in these flooding months and 0's in all others. However, for one real gigantic event that can cause different impacts for different periods, that one event could be split into more than one sub-event and thus more than one event variable.

Consequently, the multiple regression equation having k predictors, n seasons, and m special events eventually becomes

$$y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_kx_k + a_2s_2 + a_3s_3 + \ldots + a_ns_n + \gamma_1e_1 + \gamma_2e_2 + \ldots + \gamma_me_m. \quad (4)$$

where γ_i represents the coefficient of each special event variable.

In this research, two real gigantic events that affect the car sales tremendously are the 2011 Thailand big flood and the government's tax-incentive first-car buyer program in 2011-2012 [1]. First came the big flood in the third quarter of 2011, then a few months later, the incentive program started. When the big flood was starting to emerge, most people had not been impacted yet and therefore the car sales remained roughly normal. The panic commenced later after it was certain that the capital and financial city, Bangkok, would be flooded too. Even though the tax incentive program was already begun during the flood period, the car sales had not turned back high up as expected. Until the flood ended and the people turned calm, the incentive program worked its way up to the top before the promotional campaign ended. With all the situations explained above, four event variables are constructed, each of which represents a sub-event as follows.

- e_1 refers to the flood period alone.
- e_2 refers to the flood with the incentive program period.
- e_3 refers to the incentive program period alone.
- e_4 refers to the aftermath period.

2.4 The Experiments

The monthly data used in our multiple regression are collected [2], [3] from the period January 2005 to September 2015, totaling to 129 data. The original quarterly GDPs were modestly transformed into monthly by first extracting out their quarterly seasonal indices and then non-linear curve fitting them to obtain the monthly seasonal indices which eventually lead to the final monthly GDPs.

To summarize all the gathered data for multiple regression, we have the following.

y = Thailand's subcompact car sales
x_1 = GDP
x_2 = the loan's interest rate
feb, mar, apr,…, dec = eleven seasonal dummy variables for the observed months February through December
e_1, e_2, e_3, e_4 = the four dummy event variables,

and so the regression equation can be constructed as shown in Equation (5).

$$y = b_0 + b_1x_1 + b_2x_2 + a_2\,feb + a_3\,mar + \ldots + a_{12}\,dec + \gamma_1e_1 + \gamma_2e_2 + \gamma_3e_3 + \gamma_4e_4 \quad (5)$$

where γ_i represents the coefficient of each event variable.

The four special events are assigned to the relevant monthly data according to the values in Table 2. When e_i is 1, the event e_i is in effect; 0 otherwise. For example, e_3 for the month July 2012 equals 1 simply referring to the period where the tax-incentive program was in full effect and the flood was already gone. Most car buyers just started to realize the benefits of this incentive program and thus kept pouring in car reservations for purchase before the program ended. Note also that besides these months in the table, the values of the four event variables in all other months are equal to 0 meaning that normal situations are observed.

Table 2 Event assignment to monthly data (either 0 or 1)

Month	e_1	e_2	e_3	e_4	Month	e_1	e_2	e_3	e_4
Sep-11	1	0	0	0	**Nov-12**	0	0	1	0
Oct-11	1	0	0	0	**Dec-12**	0	0	1	0
Nov-11	1	0	0	0	**Jan-13**	0	0	0	1
Dec-11	0	1	0	0	**Feb-13**	0	0	0	1
Jan-12	0	1	0	0	**Mar-13**	0	0	0	1
Feb-12	0	1	0	0	**Apr-13**	0	0	0	1
Mar-12	0	1	0	0	**May-13**	0	0	0	1
Apr-12	0	1	0	0	**Jun-13**	0	0	0	1
May-12	0	1	0	0	**Jul-13**	0	0	0	1
Jun-12	0	1	0	0	**Aug-13**	0	0	0	1
Jul-12	0	0	1	0	**Sep-13**	0	0	0	1
Aug-12	0	0	1	0	**Oct-13**	0	0	0	1
Sep-12	0	0	1	0	**Nov-13**	0	0	0	1
Oct-12	0	0	1	0	**Dec-13**	0	0	0	1
					Jan-14	0	0	0	1

In the next section, results from different multiple regression models regarding Thailand's car sales figures are compared and discussed.

3.0 RESULTS AND DISCUSSION

In this section, various multiple regression models are examined. First, the regression model with the two independent variables only (Model 1), the regression model with added seasonality (Model 2), and regression model with added seasonality and special events (Model 3) are compared for their adjusted R^2's and Mean Absolute Percentage Errors (MAPEs) to see which regression model is best fitted and best accurate among these three. Then, the seasonal regression model with events or Model 3 is explored in detail.

3.1 Comparison among the Three Multiple Regression Models

According to Table 3, the multiple regression Model 1 that does not take into account the seasonal and the special event effects has a low adjusted R^2 of just 42.84% and a high MAPE of 29.82%. The multiple regression Model 2 that incorporates the seasonal effect has a higher adjusted R^2 of 66.65% and a lower MAPE of 22.04%. The last regression Model 3 that includes both seasons and special events is shown to be the most promising model with the highest adjusted R^2 of 85.89% and the lowest MAPE of 15.80%.

Table 3 Three multiple regression models' statistics

	Adjusted R^2	MAPE
Model 1	42.84%	29.82%
Model 2	66.65%	22.04%
Model 3	85.89%	15.80%

As a result, the proposed special event variables coupled with the appropriate tagging used in our seasonal regression model with events are proved effective and also make the model more explainable and more accurate. This Model 3 is hence worth a closer look in the next section.

3.2 The Seasonal Model with Events (Model 3)

By any statistical package, the regression results obtained are equivalent. The coefficients for the two independent variables, eleven seasonal dummy variables, and four event variables of Model 3 are shown in Table 4 with corrected signs as expected. For example, the sign of the loan's interest rate is negative indicating intuitively that people tend to buy less cars when the loan rate is high. In contrast, the GDP's sign is positive as one would expect the car sales to go together with the general economic conditions. Furthermore, the 698.48 absolute coefficient value for the loan rate is much more than the 0.08 value for the GDP indicating also that the interest rate of the loan to pay for the car is a more influential factor on Thailand subcompact car sales.

As for the signs of the monthly seasonal dummy variables, it appears that the eleven variables are all positive. This means that the based month of January is the lowest season of the compact car sales in Thailand, possibly due to the money spending elsewhere during the Christmas and New Year celebrations. The high seasons, on the contrary, for the car sales, according also to Table 4, are from the months of May through September.

Table 4 Coefficient table for Model 3

Description	Coefficient	Description	Coefficient
Y-Intercept	-42,685.58	*aug*	14,137.11
GDP	0.08	*sep*	15,612.88
Loan Rate	-698.48	*oct*	9,911.42
feb	4,966.20	*nov*	8,956.51
mar	9,035.47	*dec*	7,121.69
apr	8,674.60	e_1	-1,098.98
may	13,013.80	e_2	3,282.95
jun	15,300.08	e_3	19,562.81
jul	14,222.91	e_4	8,988.80

Let us now consider the last set of variables, namely, the special event variables, the sign for the flood period alone, i.e., event e_1, is negative confirming that the big flood really affected the car sales inversely. The signs for all other event variables, even the aftermath period, e_4, are positive. For events e_2, e_3 where the tax incentive campaign was active, it is quite normal to see positive coefficients. Especially in the period e_3 when the campaign was in full effect without the flood, the e_3 coefficient reached the peak. As for the reason why the sign of the coefficient in the aftermath event e_4 (where the tax-incentive campaign already ended) is still positive, it is because some people did not actually buy the cars during the campaign period but what they did was just to reserve the cars for later buy. Consequently, the purchase could really occur sometime later, or more specifically, in the event e_4 period. Afterwards, the sales seasons resume their normal patterns as before the two gigantic events occurred and thus all the event variables e_i turn back to 0s. Figure 2 plots the actual and fitted data of the Thailand's monthly car sales figures from January 2005 to September 2015.

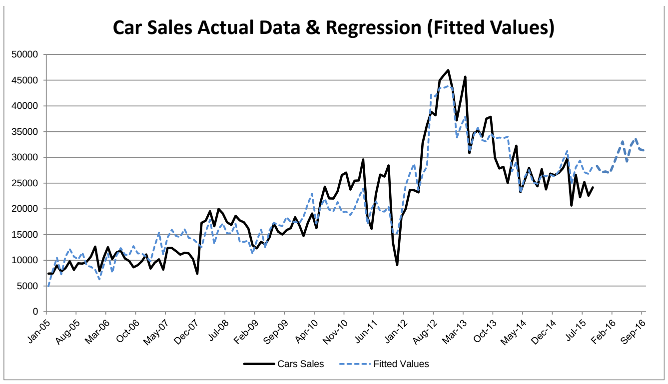


Figure 2 The actual data and its regression of Thailand's monthly compact car sales

4.0 CONCLUSION

In this research, by looking at the correlation coefficients between the subcompact car sales in Thailand as the dependent variable and the economic causal factors examined, it is evidenced that GDP and the loan's interest rate are good explanatory variables. However, these two predictors even better explain the variation of the car sales when seasonal and special event effects are added into the multiple regression model. In our case here, great fluctuations in car sales for some certain periods are due to these two particular events, namely, Thailand's 2011 big flood and the 2011-2012 tax-incentive first-car buyer program. An approach to handling such special events can be done through constructing the proposed special event variables together with proper event tagging. By the high adjusted R^2 and high MAPE obtained, this multiple regression model with seasonality and event variables is proved to be suitably fitted and accurate.

To make the multiple regression model even more fitted and more accurate, other explanatory variables that influence the car industry either positively or negatively should be explored in more detail.

References

[1] Bank of Thailand. 2013. *Thailand's Economic Conditions in 2012*. Bank of Thailand, Bangkok.

[2] Office of Industrial Economics. 2015. "Industrial Indices" [Online]. Available: http://www.oie.go.th/. [Accessed 15 December 2015].

[3] Bank of Thailand. 2015. Thailand's Macro Economic Indicators [Online]. Available: http://www2.bot.or.th/. [Accessed 15 December 2015].

[4] Dagum, E. B. and S. Bianconcini. 2016. *Seasonal Adjustment Methods and Real Time Trend-Cycle Estimation*. London: Springer-Verlag.

[5] Pearson, K. 1908. On the Generalized Probable Error in Multiple Normal Correlation. *Biometrika*. 6: 59-68.

[6] Billings, R. and D. Agthe. 1998. State-space Versus Multiple Regression for Forecasting Urban Water Demand. *Journal of Water Resources Planning and Management*. 124(2): 113-117.

[7] Kim, T., J. Hong, and H. Koo. 2013. Forecasting Box-Office Revenue by Considering Social Network Services in the Korean Market. *Jurnal Teknologi*. 64(2): 97–101.

[8] Ismail, Z., A. Yahya, and A. Shabri. 2009. Forecasting Gold Prices Using Multiple Linear Regression Method. *American Journal of Applied Sciences*. 6(8): 1509-1514.

[9] Shahabuddin, S. 2009. Forecasting Automobile Sales. *Management Research News*. 32(7): 670-682.

[10] Wang, F., K. Chang, and C. Tzeng. 2011. Using Adaptive Network-Based Fuzzy Inference System To Forecast Automobile Sales. *Expert Systems with Applications*. 38: 10587-10593

[11] Leenawong, C. and P. Netisopakul. 2016. Forecasting Car Sales in Thailand with the Impact of the National Flood and the First-Car Buyer Program. *Proceedings of the 4th International Conference on Business and Social Sciences*. Kyoto, Japan. 29-31 March 2016. 1006-1008.

[12] Haugh, D., A. Mourougane, and O. Chatal. 2010. The Automobile Industry In and Beyond the Crisis. *OECD Economics Department, Working Papers*745.

[13] G. Riley. 2011. "tutor2u" [Online]. Available: http://www.tutor2u.net/economics/blog/economics-qa-what-economic-factors-affect-the-demand-for-new-cars. [Accessed 15 December 2015].

[14] Lovell, M. C. 1963. Seasonal Adjustment of Economic Time Series and Multiple Regression. *Journal of the American Statistical Association*. 58(304): 993-1010.

[15] Wilson, J. H. and B. Keating. 2009. *Business Forecasting with ForecastX*. McGraw-Hill Education.