EASR

Engineering and Applied Science Research

https://www.tci-thaijo.org/index.php/easr/index

Published by the Faculty of Engineering, Khon Kaen University, Thailand

Disclosing fast moving consumer goods demand forecasting predictor using multi linear regression

Farizal F*, Muhammad Dachyar, Zarahmaida Taurina, and Yusuf Qaradhawi

Department of Industrial Engineering, Faculty of Engineering, Universitas Indonesia, Kampus Baru UI Depok 16424, Indonesia

Received 1 September 2020
Revised 2 March 2021
Accepted 8 March 2021

Abstract

Demand forecasting accuracy undoubtedly influences a company perfomance. With an accurate forecast, the company will be able to utilize its resources efficiently. In practical, most companies only utilize historical selling data as predictor to forecast their product demand either using qualitative forecasting method or time series. However, in this study on a-fast moving consumer goods (FMCG), i.e., insecticide product, these methods do not give good results as expected. The methods produce Mean Absolute Percentage Errors (MAPEs) above 20%. To provide a more accurate forecasting, this study proposes a Multi Linear Regression (MLR) model that uses predictors including climate, promotion, cannibalization, holiday, product prices, number of retail stores, population, and income. The result shows that the MLR gives the best accurate forecast compare to time series methods and simple linear regressions. Using five predictors, i.e., product price, cannibalism, price disparity, fest day and weather, the proposed MLR model gives more accurate forecast with MAPE 8.66%.

Keywords: FMCG, Insecticide product, Time series, Multi linear regression, Forecasting

1. Introduction

The initial stage of push-based production system is forecasting. Forecasting is essential to a company since with a good forecast, the company can ensure effective use of its resources. Forecasting is a combination of science and art [1]. The scientific part of forecasting is predicting the future using statistics as tools for increasing forecast accuracy. Forecasting is an art refers to user intuition to choose the best equation that represents the most prediction accuracy of empirical data; getting the best equation is rare.

Demand forecasting is considered as one of the critical functions that affects companies worldwide across all industries, including heavy manufacturing, retail, pharmaceutical, automotive, electronics, telecommunication, financial, Fast Moving Consumer Goods and others [2]. Fast Moving Consumer Goods (FMCG) or Consumer Packed Goods (CPG) refer to products needed by consumer for daily or frequent use [3]. FMCG include soft drinks, personal care products, household products, stationery, insecticides, etc. One of FMCG characteristics is a shorter shelf life time caused by high consumer demand or perishable product [4]. FMCG have low profit margin but due to large quantity of products being sold, they bring enormous cumulative profits [5]. Reference [6] listed five characteristics of FMCG:

- The user is coming from wide spectrum of consumers
- The product is consumed frequently
- Demand may change rapidly
- The product has low consumer's loyalty
- The consumer use the product for convenience

Learning from the characteristics mentioned above, predicting FMCG's demand is not easy. The selling patterns of FMCG brand are not the result of time (the moment), but due to some activities of marketing, advertisement, weather, etc. According to [7], based on their investigation to 50 companies, company's forecast does not perform well as they wish. Most companies utilize their historical sales data to predict their future demand. Excluding sales promotion or weather as predictor affects company's forecast accuracy. These problems can be overcome by using causal model. Causal model attempts to determine the cause-effect relationship between a set of variables. This approach consists of a variety of techniques such as linear programming, simulation, and multiple linear regression [8].

To build an accepted causal model is not easy due to some reasons including inability to identify leading factors or unavailable data. To overcome the problems, time series can be an alternative solution. Multiple linear regression (MLR) or artificial neural network (ANN) might be a good alternative if leading variables can be identified and data for the variables are available. In this study, an accurate multi linear forecasting model to predict FMCG, i.e. insecticide product was developed. Some combination of time series were also investigated.

*Corresponding author. Tel.: +62 813 8524 0730
Email address: farizal@eng.ui.ac.id
doi: 10.14456/easr.2021.64

2. Methods

2.1 Forecasting method

Forecasting is predicting the future event(s) based on past and current data [9]. Forecasting spans in many fields such as engineering, business and economy, environment, social sciences, including politics. A variety of forecasting methods are available to used. The methods span from the simplest one such as using the last observation to forecast to the advanced approaches such as ANN and simultaneous equation. In 1975, [10] already listed 150 forecasting approaches.

In general, forecasting method can be classified into four groups [11]:

- Qualitative method

Qualitative forecasting method is subjective approach since it depends on human judgment. It is usually used when historical data are not available or incomplete. To get a better forecast, this method will be best exercised by the expert on the area. One well known forecasting method belongs to this group is the Delphi Method that introduced by the RAND Corporation on the mid-sixties [9]. Delphi Method employs repetitive rounds of some panels until they come up with a consensus or justified opinion differences.

- Time series method

Time series forecasting method uses historical data to predict future observations [12]. This method is based on the assumption that past events will lead to future events. In this matter since the forecast is for predicting future demand, historical data used is item selling data. Time Series Method is more appropriate to used when the demand pattern does not vary significantly from time to time. Time series method is a popular method since it is simple to be implemented. It can be used as a good starting point for demand forecasting. Belong to this group are methods such as Moving Average, Exponential Smoothing, Naïve, Trend Projection, and Seasonality [12, 13].

- Simulation forecasting method

Simulation forecasting method is replicating consumer's choice to a simulation model. The model is tested as the forecast. Performance of this method depends on how well the simulation model replicates the real consumer situation.

In qualitative forecasting and time series methods that only use one variable, product sales historical data are usually used to predict the future demand. The predicted values are generated according to the method selected to process the data. Moving Average forecasting technique is calculating an average of the last "n" values for predicting the next period value. Weighted Moving Average is assigning different weight to each period with the sum of all weights equals one. This technique puts more weight to the recent data than the old ones. Exponential Smoothing technique is assigning exponential weighted to the observed data so that the most recent data are more heavily weighted than previous data. Naïve method is a forecasting technique based on simple approach in which the previous period actual data (t-1) is used as a forecast for the next period. Averaging past value is similar to the moving average. The difference is in the number of periods used to calculate the average. In this technique, averages are calculated using all previous data. Trend projection forecasting technique is using trend line from some past data to predict the future.

- Causal method

Causal forecasting method is based on the assumption that demand forecasting has correlation with measurable circumstance factors such as economic conditions, weather, number of population and others. On Causal method, demand, factor to be predicted, is called as dependent variable. Its value depends on one or more other factors. While factor(s) that steers dependent variable is called as independent variable, regressor, or predictor. Economic conditions such as gross domestic product (GDP), income, and interest rate drives selling rate. If economic flourish people will spend more; the selling increases. In a big city (the population is big), the number of buyers is also big, thus the selling increases. Relating the predictor(s) and the dependent variable, this method formulates the correlation in form of a model. If the correlation is assumed as linear, the model is called as the multiple linear regression (MLR).

Generic model of MLR is formulated as equation (1) [14]:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 X_3 + ... + \alpha_n X_n + \varepsilon \quad (1)$$

Where:

Y: the dependent variable (variable to be predicted)

α_0: a constant (intercept)

$\alpha_1, \alpha_2, ..., \alpha_n$: regression coefficient (slope)

$X_1, X_2, ..., X_n$: the independent variables or the predictors

ε: error estimated

Based on Equation (1), every successful effort to use this model is relied on determining which independent variable(s) to used and the form of the model. The predictors can be determined from related previous studies and/or experts' suggestion. Furthermore, to use this multilinear model, a set of requirements, namely residual randomness, normally, and non-collinearity, should be fulfilled and tested for model validation [15]. MLR method requires that the predictors should have no correlation one another. To detect the degree or strength of any correlations between the independent variables, multi-collinearity tests such as Variance Inflation Factor (VIF), Pearson Correlation between variables, or Eigen Values correlation coefficient, Red indicator, condition number (CN) and condition index (CI) are used. According to [16] as mentioned on [17], as multicollinearity is considered as a sample phenomenon, there is no single method able to detect its occurrence. However, there are three primary methods, VIF, correlation coefficient, and eigenvalue method that usually used.

VIF measures how much the variance of the estimated regression coefficient is inflated if the independent variables are correlated. VIF is calculated as $\frac{1}{(1-R_j^2)}$ If calculated VIF is equal to 1 the variables are not correlated one another. If the value is between 5 to 10, it indicates moderate correlation and it may be problematic to use the variables. And if VIF is more than 10, it indicates the regression coefficients are weakly estimated due to the presence of multicollinearity [18].

Pearson correlation between two variables is calculated with equation (2).

$$r = \frac{n\left(\sum X_1 X_2\right) - \left(\sum X_1\right)\left(\sum X_2\right)}{\sqrt{\left[n \sum X_1^2 - \left(\sum X_1\right)^2\right]\left[n \sum X_2^2 - \left(\sum X_2\right)^2\right]}} \quad (2)$$

Where:
r: the coefficient of correlation
n: the number of observations
X_1 and X_2: the first and the second predictors.
If the value of r is close to 0.8 or more, collinearity is likely to exist [19].
Eigenvalue λ represents the variance of the linear combination of the variables. A very small eigenvalue (close to 0.05) indicates multicollinearity. Sometime Conditional Index CI is used to represent eigenvalue as shown on Equation (3).

$$CI = \sqrt{\frac{\lambda_{(p-1)}}{\lambda_{(1)}}} \tag{3}$$

Where:
$(p - 1)$: the number of predictors
$\lambda(p - 1)$: the maximum eigenvalue
$\lambda(1)$: the minimum value
If the value of CI is less than 15, it means weak multicollinearity exist. CI between 15 to 30 indicates moderate multicollinearity. CI for more than 30 shows an evidence of strong multicollinearity [19].

2.1.1 Test for regression significance

The test is to determine whether there is a linear relationship between the dependent variable Y and a subset of independent variables $X_1, X_2, ..., X_n$. F test is used for testing the significance with the appropriate hypothesis testing format are followed:

- The null hypotheses H_0: $\beta_1, \beta_2, \ldots, \beta_n = 0$

Which means the proposed MLR model can not be used for predicting the demand vs.

- The alternative hypotheses H_1: at least $\beta_1, \beta_2, \ldots, \beta_n \neq 0$

Which means the proposed MLR model can be used for predicting demand.
If H_0 is rejected it implies that the proposed MLR model is accepted since at least one independent variable influences the dependent variable. To carry out significance test, analysis of variance (ANOVA) is calculated. The $F_{\text{statistics}}$ result compares to F_{table} to determine whether reject or fail to reject H_0.
A measure to tell how well the model fits the sample data is Coefficient of Determination R^2. R^2 value is calculated by squaring r on Equation (2) but for all used predictors. The value of R^2 is between 0 to 1. If the model well fits the data, means its total variability is small, R^2 will close to 1. On the contrary if R^2 closes to 0, the model does not fit the data. In general, R^2 around 0.8 is considered as good model [18]. R^2_{adj} is the normalized value of R^2.

2.2 Forecasting error

With many available forecasting methods, one way to determine which method to used is evaluating the method performance through calculating its forecasting error [9, 13]. Many forecasting error measures have been developed such as Mean Absolute Deviation (MAD), Mean Square Error (MSE), Mean Squared Deviation (MSD), Root Mean Square Error (RMSE), and Mean Absolute Percent Error (MAPE). Due to the weaknesses related to the measures, some variants and newer measures have been proposed such as Median Absolute Percentage Error (MdAPE), Mean Absolute Scaled Error (MASE), Average Relative Mean Absolute Error (AvgRelMAE). However, since data to be forecasted are varied and diverse, it is very hard to claim one measure is the best to used [20-22]. But, for each measure, the lower the value the better the forecasting method.
Among the measures, one popular and easy to used measure is MAPE [23]. MAPE expresses the error as a percentage. It means MAPE is dimensionless or unit-free. As addition to easiness to interpret its result, MAPE can be used to compare the accuracy of the same or even different forecasting models on two distinctly different series [13, 24]. MAPE is calculated using Equation (4) below:

$$MAPE = \sum_{i=1}^{n} \frac{\frac{|Y_i - \hat{Y}_i|}{Y_i} x\ 100\%}{n} \tag{4}$$

Where:
Y_i: Actual dependent variable values
$\hat{Y}_i$: Predicted dependent variable results given by the model
n: The number of observations
A forecasting method is considered good if its MAPE is between 10% to 20%. If the error is less than 10% the method is considered as excellent [25-28].

3. Results and discussion

3.1 Problem setting

An FMCG company produces insecticide product utilizes qualitative method to forecast its future demand. Forecasting is conducted by its production staffs. Figure 1 presents the company's actual sales vs. the company's demand forecast. As clearly shown on Figure 1, the company's forecast most of the time has similar pattern with the actual sales' pattern.
However, the company's forecasting values (predicted sales) are quite different with the actual sales' values. Using Equation (4), company's forecast shown on Figure 1 has MAPE 28.9%. This error is considered big. According to [25], a very good forecast should have MAPE less than 10%. A forecast is considered good if its MAPE is less than 20%. Since the company's MAPE is more than 20%, its qualitative forecasting is considered not good enough. This research is proposing a forecasting model that has error smaller than the error the company's forecast has.

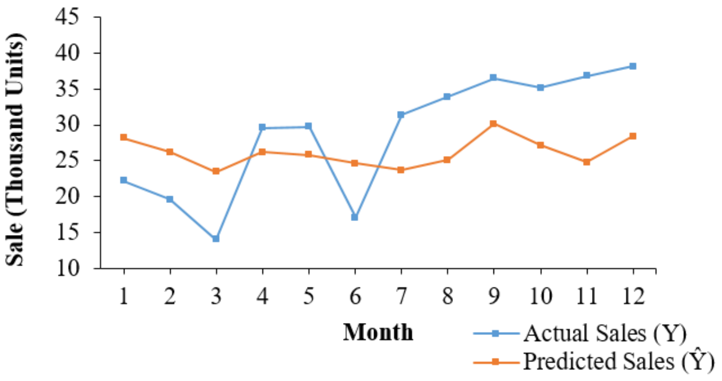

Figure 1 Actual sales and company's forecast

3.2 Time series and causal forecasting methods

In the first step of this study, forecasting is conducted using single predictor with single factor forecasting methods. The objective is to investigate the accuracy of the forecast compare to the qualitative method. Since most of the time, predicting demand is based on its past demand (historical sales data), the same predictor is used with time series methods. On this study, the time series methods used are moving average, weighted moving average, and exponential smoothing.

It makes sense to say that the number of product sales depends on factors such as the number of population, income, and the product price itself. Using these predictors, single factor causal method (i.e. simple linear regression with predictor of number of population, income, and price) is used to forecast the future demand. Figure 2 shows the summary of the forecasting results for time series and simple casual methods while MAPE for each method is displayed on Table 1.

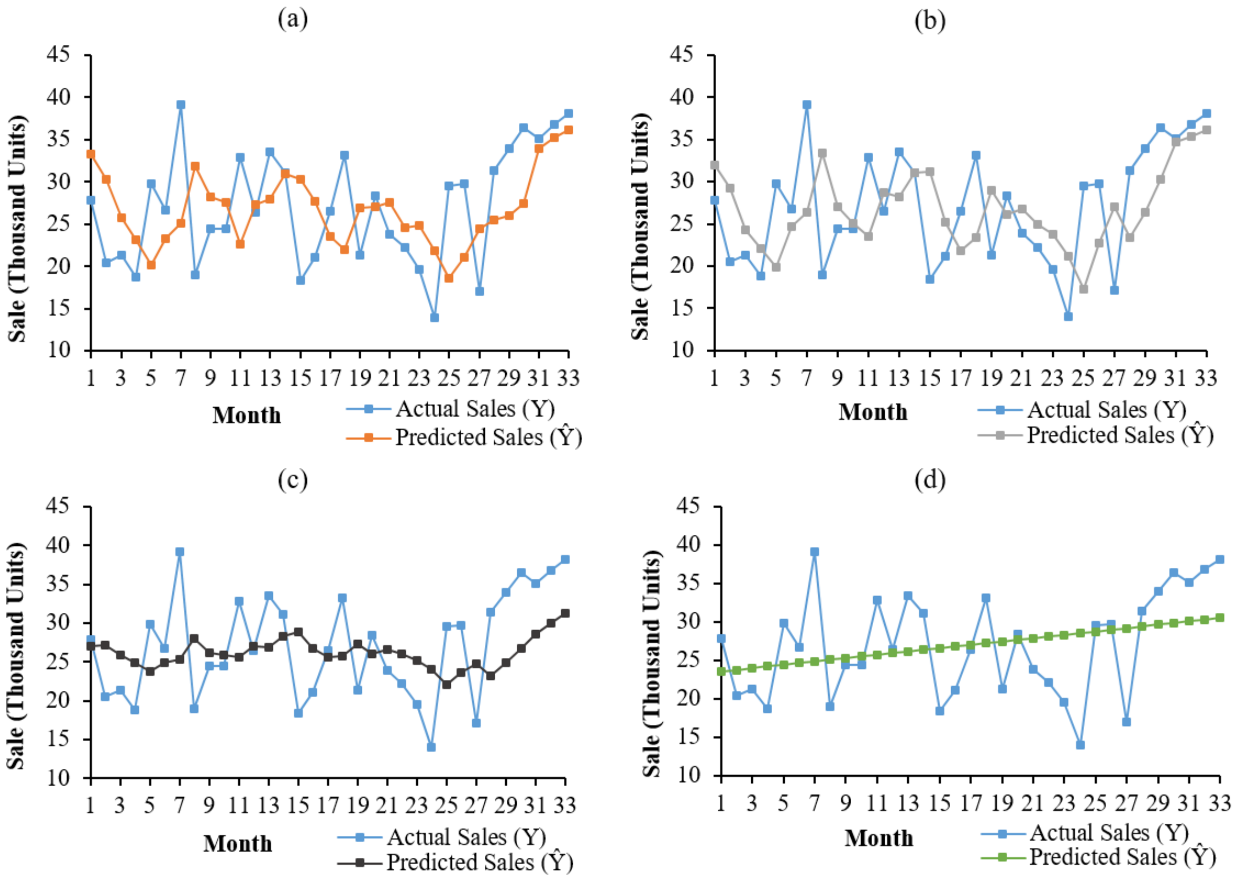

Figure 2 Single factor forecasting result by: (a) moving average; (b) weighted moving average; (c) exponential smoothing; (d) simple linear regression on sale

Figure 2 shows that forecasting results of time series methods has similar pattern with the actual sales' patterns but with one-step behind. However, the values predicted most of the time are very much different. These differences are indicated by their MAPE. As clearly shown in Table 1, MAPE from all time-series forecasting methods and causal methods are above 20%. The only method that gives MAPE slightly less than 20% is the one that uses Price as its predictor. Its MAPE is 19.36%. These results show that for insecticide product, the best single predictor to used when predicting future product demand is Price. However, these single factor forecasting methods are not good enough. MAPE that represents high accuracy forecasting technique is under 10% [25-28]. As a result, either time series methods or single factor causal methods are not accurate enough to predict the forecast.

Table 1 Single variable forecasting method accuracy

Forecasting method		MAPE (%)
Time series (based on sale as predictor)	Moving average	24.60
	Weighted moving average	23.66
	Exponential smoothing	23.89
Causal method: simple linear regression (based on predictor)	Sale	21.74
	Population	21.76
	Income	20.05
	Price	19.36

3.3 Multiple linear regression forecasting method

The first step to use multiple linear regression (MLR) method for forecasting is determining the independent variables of the model. For the purpose, in this study the predictors are collected through literature study and retrieved from experts. Two types of literatures were used; research papers and text-books. As the results 14 factors were found, i.e. 1. the number of consumers, 2. product price, 3. type of consumer, 4. cannibalism, 5. quality of product, 6. population income, 7. price disparity, 8. fest day, 9. weather, 10. forecast, 11. advertisement, 12. product shelf life, 13. the number of retails to sell the product, and 14. type of sales. Tables 2 and 3 show citing of the factors from text-book sources and research papers, respectively. Both Tables 2 and 3 clearly show that the most cited factor to include to the model is product price. From 20 sources, 16 of them included the price as their factor. The next most cited factors after price are the number of population, population income, and price disparity. They have been used by 8 different studies. As for experts involved on this study, they are the company's marketing director and research and development director. Their opinion on which factors influence the demand were asked.

Table 2 Forecasting factors from text books

No.	Author	Factor										
		1	2	3	4	5	6	7	8, 9	10	11	12
1	Reference [29]	v	v	v			v	v	v		v	v
2	Reference [11]		v		v			v	v		v	
3	Reference [30]		v		v	v	v	v				
4	Reference [31]	v	v	v			v	v				
5	Reference [32]	v	v				v		v	v		

Table 3 Forecasting factors from previous studies

No.	Author	Factor													
		1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	Reference [8]	v			v				v						
2	Reference [33]				v					v					
3	Reference [34]	v													
4	Reference [35]		v	v		v			v		v		v		
5	Reference [36]		v		v	v									
6	Reference [37]		v					v							
7	Reference [38]		v					v							
8	Reference [39]	v	v							v					
9	Reference [40]	v	v				v								
10	Reference [41]		v				v					v		v	
11	Reference [42]						v	v		v				v	
12	Reference [43]		v					v						v	
13	Reference [44]		v		v				v			v			
14	Reference [6]		v												
15	Reference [45]	v	v				v			v			v	v	v
	Total count	5	11	1	4	2	4	4	3	4	1	2	2	4	1

Combining literature studies and expert opinion, as the result, from 14 factors, 8 factors are considered as independent variable for the model to be developed. The eight factors are the number of consumers, price, cannibalism, income, price disparity, fest day, weather, and the number of retails. Data for the variables gathered from the past 3 years (2016-2018) are shown in Table 4.

In this research, to detect multicollinearity among the eight independent variables, VIF test and Pearson Correlation Coefficient were used. The calculation results with IBM SPSS are shown on Tables 5 and 6. Table 5 clearly shows that 3 out of 8 predictors have VIF values greater than 10. Number of consumers or Population, income, and number of retailers have VIF as 71.041, 11.608, and 84.315, respectively. These values indicate that those 3 predictors are collinear. These results are in line with correlation coefficient calculation results shown on Table 6. The coefficient for population, number of retailers and income are close to or even more than 0.8. As the conclusion, these three predictors are omitted for further consideration. While, the rest five predictors have acceptable VIF values since they are less than 10. Even most of them are less than 5. These 5 predictors are not collinear one another [18, 19].

Table 4 Input data

Month	Sale (Item)	Population	Price (IDR)	Cannibal (Box)	Income (IDR)	Delta (IDR)	Fest (Day)	Weather (mm)	Retailer (Shop)
Jan-16	36,789	255,731,975	30,120	-1,403.50	13,299,651	-1,000	20	150	25,513
Feb-16	34,125	256,002,250	30,120	-1,403.50	13,299,651	-500	20	150	25,776
Mar-16	28,916	256,272,525	33,450	-1,403.50	14,770,031	2,000	21	150	25,926
Apr-16	27,817	256,542,800	33,450	-1,403.50	14,201,138	1,500	21	300	26,038
May-16	20,456	256,813,075	34,120	-1,403.50	14,485,586	1,000	20	400	26,278
Jun-16	21,321	257,083,350	34,600	-1,403.50	14,689,369	1,000	22	400	26,514
Jul-16	18,780	257,353,625	35,500	-2,807.00	14,613,500	2,000	15	500	26,761
Aug-16	29,801	257,623,900	32,450	-1,403.50	13,357,974	500	22	300	27,157
Sep-16	26,718	257,894,175	34,740	-1,403.50	14,300,648	500	21	200	27,397
Oct-16	39,144	258,164,450	28,540	-1,403.50	12,627,942	-1,000	21	200	27,518
Nov-16	18,965	258,434,725	34,740	-1,403.50	14,564,897	2,000	22	400	28,015
Dec-16	24,425	258,705,000	34,740	-1,403.50	14,564,897	1,000	20	400	28,487
Jan-17	24,474	258,970,492	27,680	-2,807.00	11,639,557	1,000	21	300	28,842
Feb-17	32,859	259,235,983	28,540	-1,403.50	12,001,191	1,000	19	100	29,293
Mar-17	26,458	259,501,475	30,508	-1,403.50	12,828,743	1,500	22	200	29,477
Apr-17	33,497	259,766,967	28,528	-1,403.50	11,533,766	500	18	100	29,721
May-17	31,130	260,032,458	27,756	-1,403.50	11,221,649	500	20	350	29,922
Jun-17	18,388	260,297,950	32,450	-561.4	12,633,452	2,000	11	300	30,125
Jul-17	21,106	260,563,442	30,871	-561.4	12,095,227	3,000	21	200	30,479
Aug-17	26,499	260,828,933	32,450	1,403.50	11,553,494	3,000	22	100	30,663
Sep-17	33,185	261,094,425	30,832	-561.4	12,080,077	1,000	19	50	30,960
Oct-17	21,310	261,359,917	31,689	-1,403.50	12,630,911	2,000	22	400	31,088
Nov-17	28,385	261,625,408	31,597	-1,403.50	12,594,241	1,000	22	300	31,257
Dec-17	23,865	261,890,900	30,033	-1,403.50	11,970,854	2,000	18	500	31,376
Jan-18	22,198	262,151,267	32,860	-1,403.50	13,151,953	3,000	22	400	31,523
Feb-18	19,586	262,411,633	33,831	-1,403.50	13,540,455	3,000	19	500	31,622
Mar-18	13,977	262,672,000	33,639	-1,403.50	13,463,608	3,000	21	500	31,705
Apr-18	29,550	262,932,367	32,924	-561.4	12,644,848	1,000	21	200	31,788
May-18	29,744	263,192,733	33,436	-561.4	12,841,488	1,000	20	50	31,871
Jun-18	17,063	263,453,100	32,199	-1,403.50	12,366,275	3,000	9	50	31,972
Jul-18	31,364	263,713,467	30,529	-1,403.50	11,373,779	-1,000	22	20	32,102
Aug-18	33,934	263,973,833	31,220	-1,403.50	12,335,958	500	21	20	32,175
Sep-18	36,470	264,234,200	30,033	-561.4	12,143,224	-500	19	50	32,283
Oct-18	35,162	264,494,567	32,439	-561.4	12,292,522	500	23	50	32,454
Nov-18	36,814	264,754,933	31,220	-561.4	12,643,174	500	22	50	32,636
Dec-18	38,141	265,015,300	31,220	-561.4	12,566,879	-1,000	19	100	32,866

Table 5 Multicollinearity results with VIF

Model		Unstandardized coefficients		Standardized	t	Sig.	Collinearity statistics	
		B	Std. error	Coeff. beta			Tolerance	VIF
1	(Constant)	-75934.832	297878.553		-0.255	0.801		
	Pop	0.001	0.001	0.211	0.373	0.712	0.014	71.041
	Price	-1.742	0.596	-0.543	-2.924	0.007	0.130	7.668
	Cann	2.147	0.941	0.218	2.282	0.031	0.491	2.037
	Income	0.003	0.001	0.408	1.786	0.085	0.086	11.608
	Delta	-2.735	0.625	-0.493	-4.373	0.000	0.354	2.827
	Fest	340.729	167.958	0.147	2.029	0.052	0.859	1.164
	Whether	-12.869	4.566	-0.301	-2.818	0.009	0.395	2.534
	Retailer	-0.223	1.774	-0.077	-0.126	0.901	0.012	84.315

Table 6 Pearson correlation coefficient between predictors

	Pop	Price	Cann	Income	Delta	Fest	Whether	Retailer
Pop	1	-0.124	0.397*	-0.557**	0.051	-0.057	-0.339*	0.981**
Price		1	0.034	0.794**	0.423*	-0.012	0.388*	-0.178
Cann			1	-0.319	0.068	0.106	-0.455**	0.410*
Income				1	0.161	0.070	0.475**	-0.626**
Delta					1	-0.219	0.469**	0.144
Fest						1	0.005	-0.068
Whether							1	-0.284
Retailer								1

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Using the five left over predictors, the VIF values are recalculated. As Table 7 shows the highest VIF value of the 5 predictors is 2.1 which indicates no collinearity exist.

Table 7 Recalculated VIF of five predictors[(a)]

Model		Unstandardized coefficients		Standardized	t	Sig.	Collinearity statistics	
		B	Std. error	Coef. beta			Tolerance	VIF
1	(Constant)	48416.016	8544.189		5.667	0.000		
	Price	-0.642	0.253	-0.200	-2.542	0.016	0.752	1.329
	Cann	1.336	0.840	0.136	1.589	0.122	0.638	1.566
	Delta	-3.147	0.496	-0.567	-6.338	0.000	0.582	1.718
	Fest	382.952	168.437	0.165	2.274	0.030	0.886	1.129
	Whether	-12.500	4.242	-0.292	-2.947	0.006	0.474	2.108

[(a)]Dependent variable: demand

MLR required that the data set should be approximately normally distributed. To check this requirement, Normal Probability Plot on standardize residual of demand can be used. The plot on demand is shown on Figure 3. According to [18], Fat Pencil Test can be used to check the normality on a normal probability plot. If a fat pencil can cover all the observations on the plot, it can safely conclude that the data set is approximately normally distributed. Figure 3 shows that all observations can very much be covered by fat pencil. The data set used for MLR model is approximately normal.

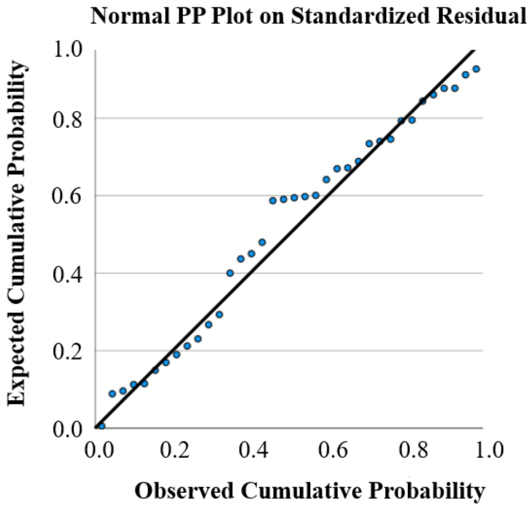

Figure 3 Normal probability plot on standardized residual

The next is checking for residual randomness requirement. For the purpose Scatter Plot of the residuals can be used. If the scatter plot of the residuals is randomly scattered around zero for the entire range of fitted value, the residual randomness requirement is fulfilled. However, if the plot resembles a certain pattern like bowl or double bowl, the requirement is not fulfilled [18]. Scatter plot with bowl or double bowl pattern is suitable for nonlinear regression. Scatter plot for FMCG forecasting model is displayed on Figure 4. The figure clearly shows no particular pattern but random. Residual randomness requirement for the MLR model is fulfilled.

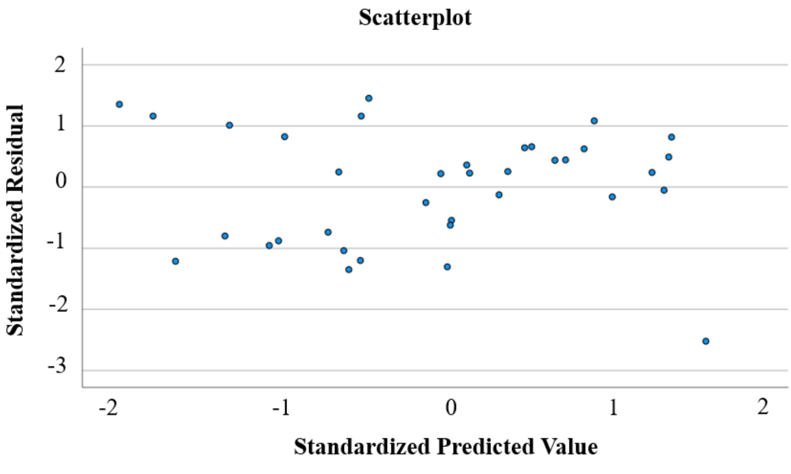

Figure 4 Scatter plot for residual

Based on information from Table 7, MLR model for the insecticide demand forecasting is shown on equation (5).

$$Y = 48,416 - 0.64\ X_1 + 1.34\ X_2 - 3.15\ X_3 + 382.95\ X_4 - 12.5\ X_5 \quad (5)$$

Or Demand = 48,416 – 0.64 Price + 1.34 Cannibalism –3.15 Price Disparity + 382.95 Fest Day – 12.5 Weather

By inspection analysis on the proposed MLR model of Equation (5):

Variable X_1 is the product price. Price has a direct impact on sale or customer willingness to buy. The higher the product price, the less the customers want to buy [46, 47]. The proposed model makes sense since the coefficient for product price is negative. The negative sign means the price has the opposite impact to the demand; as the price is higher, the demand is lesser. For insecticide product, price contributes negatively to the sale.

Variable X_2 is cannibalism. Cannibalism means introducing a new product in a simple way such as using a new casing, new color, etc. but the content itself is still the same; the old one will be rebranded. Cannibalism will lift the sale up since customers are encouraged to buy a new fresh-better looking product [48, 49]. According to [48], the relationship of Apple's iPad and Mac is an example of advantage from market cannibalization. The proposed model makes sense since the cannibalism coefficient is positive. The positive sign means the variable has positive impact on the sale.

Variable X_3 is price disparity. Price disparity is defined as the difference between the own price and the competitor price. Price disparity is positive means the company product is sold more expensive than its competitor. While if it is negative, it is sold cheaper. Thus, if the disparity is positive it will discourage the consumers to buy [50, 51]. The coefficient for price disparity in the proposed model is negative. From this point, the model makes sense.

Variable X_4 is fest day. Fest day refers to national holiday where offices and businesses are closed. In Indonesia, fest day is included Eidul-Fitr, Eidul-Adha, Christmas, etc. Fest day clearly affects retail trade due to sales increment in almost all categories include food, cloth, gift, etc. [52, 53]. The coefficient for the fest in the proposed model is positive; the model makes sense.

Variable X_5 is weather. That the weather has direct impact on sales is well studied [54-56]. The weather can have both positive (increases sale) or negative (reduces sale) impact on a product sale. Weather aspect can be in many forms, for instance temperature, climate, sun intensity, etc. In this study the weather is on the form of precipitation. The model said that the precipitation has negative impact on the insecticide sale. The more raining the less the sale. This relationship makes sense since during the rainy season the number of mosquitoes is less than during the dry time. The less the mosquito the less the family uses the insecticide the less the sale.

Equation (5) is used to calculate the predicted sales. Figure 5 plots the actual sales and the predicted (forecasted) ones. Figure 5 shows that the predicted sales calculated using MLR model is very much closed to the actual sales. Using Equation (4), the MAPE of this MLR forecasting is 8.66%. This value is much lower than MAPE from previous methods mentioned on Table 1. The MLR model gives high accuracy demand forecast for insecticide product [25-28].

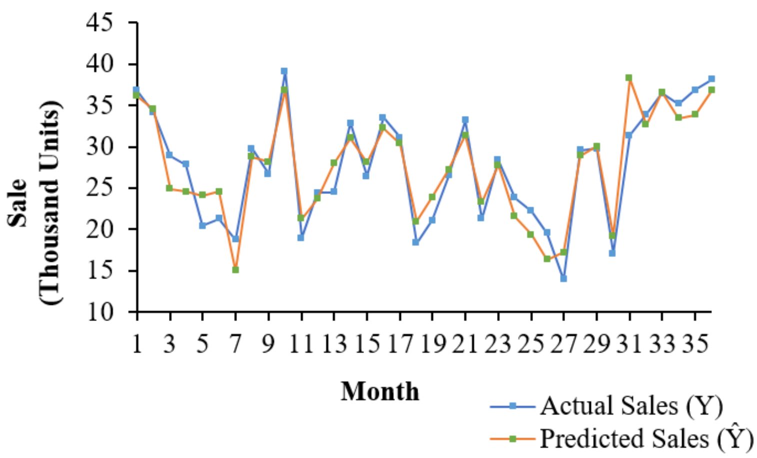

Figure 5 MLR Predicted sales vs the actual sales.

Providing accurate forecasting results, the followings are tests and analysis to evaluate the validity of the MLR model listed on Equation (5).

3.3.1 F Test

To conduct *F*-test, two hypotheses statements are set up:

- H_0: $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5 = 0$ vs. H_1: at least one of $\beta_1, \beta_2, \beta_3, \beta_4, \beta_5 \neq 0$

The procedures to obtain the conclusion of hypotheses testing are listed below:

- $F_{Statistic}$ = 36.85 (Calculation result shown on Table 8)
- F_{Table} (from Table F)

$F_{Table} = F_{\{(\alpha)\ (df\text{numerator} = \text{m})(df\text{denominator} = \text{n-m-1})\}}$ where $m = 5$, $n = 36$, $\alpha = 0{,}05$, dfdenominator = m-n-1 = 36-5-1 = 30. So $F_{Table} = F_{\{(0{,}05)(5{,}30)\}} = 2.53$

- Conclusion: $F_{Statistic} \geq F_{Table}$, 36.85 > 2.53

Since $F_{Statistic}$ is greater than F_{Table}, reject the null hypotheses (H_0) and conclude that at least one predictor influences the demand. This result means the proposed MLR model statistically can be used for predicting insecticide product demand. Hypotheses testing for each individual effect is not necessary due to the rejection of null hypotheses.

Table 8 Analysis of variance (ANOVA) result

Model	Sum of squares	df	Mean square	Fstatistic	Sig
Regression	1.38.E+09	5	2.76.E+08	36.85	0.00
Residual	2.25.E+08	30	7.49.E+06		
Total	1.61.E+09	35			

Although, *F*-test is frequently used to test of the hypotheses, statistical test such as *P*- value can also be used for hypotheses testing. As shown in Table 8, the significant *P*-value is 0.00. This value is smaller than the significance level $\alpha = 0.05$, so the conclusion is to reject the null hypotheses. The *F*-test and the *P*-value test give the same result, i.e. rejecting the null hypotheses. It can be concluded that product price, cannibalism, price disparity, fest day and weather can confidentially be used as factors to predict insecticide product demand.

3.3.2 Coefficient of determination R^2 and R^2_{adj}

The value of R^2 and R^2_{adj} are shown in Table 9. Value of R^2 and R^2_{adj} are 86% and 83.7% respectively. These results show that 86 % of the variation in variability of insecticide's product demand can be explained by the MLR model. This high percentage of R^2 indicates that the proposed model can be used for predicting insecticide product's demand [18].

Table 9 Summary

Model	R	R2	Adj. R2	Std. error of Y	Observations
1	0.927	0.860	0.837	2,736.61	36

4. Conclusion

From this study some important findings are as followed: forecasting future FMCG demand by just using historical sales data as the predictor does not give accurate demand prediction; the MAPEs of Moving Average, Weighted Moving Average, Exponential Smoothing, and simple linear regression on sales methods are above 20%. However, in case one predictor forecasting is utilized, it is better to pick price as the independent variable compare to use historical sales. Using more than one predictor, the proposed multi linear regression model is $Y = 48,416 - 0.64 X_1 + 1.34 X_2 -3.15 X_3 + 382.95 X_4 – 12.5 X_5$. The model has a low MAPE of 8.66% and a high R^2 of 86%. This quite accurate model uses five predictors to forecast demand, i.e., product price, cannibalism, price disparity, fest day and weather.

5. Acknowledgements

This work has been supported partially by PITTA 2019 Grant funded by DRPM Universitas Indonesia under contract No: NKB-0732/UN2.R3.1/HKP.05.00/2019.

6. References

[1] Evans MK. Practical business forecasting. USA: Blackwell; 2003.
[2] Chase CW. Demand-driven forecasting: a structured approach to forecasting [Internet]. 2013 [cited 2021 Jan 27]. Available from: http://ebookcentral.proquest.com.
[3] Pongiannan K, Chinnasamy J. Do advertisements for fast moving consumer goods create response among the consumers?-an analytical assessment with reference to India. Int J Innovat Tech Manag. 2014;5(4):249-54.
[4] Ramanuj M. Product management in India. 3rd ed. New Delhi: PHI Learning; 2004.
[5] Sean B. The advertising handbook. 2nd ed. UK: Routledge; 2002.
[6] Liu Y, Li M, Zhu Z. Simulated annealing sales combining forecast in FMCG. 10th International Conference on e-Business Engineering; 2013 Sep 11-13; Coventry, UK. New York: IEEE; 2013. p. 230-5.
[7] Adebanjo D, Mann R. Identifying problems in forecasting consumer demand in the fast moving consumer goods sector. Benchmark Int J. 2014;7(3):223-30.
[8] Caruana A. Steps in forecasting with seasonal regression: a case study from the carbonated soft drink market. J Prod Brand Manag. 2001;10:94-102.
[9] Montgomery DC, Jennings CL, Kulahci M. Introduction to time series analysis and forecasting. 2nd ed. New Jersey: Wiley; 2008.
[10] Mitchell A, Dodge BH, Kruzic PG, Miller DC, Schwartz P, Suta BE. Handbook of forecasting techniques. Springfield: National technical information service, USA; 1975.
[11] Chopra S, Meindl P. Supply chain management, strategy, planning and operation. 6th ed. Northwestern: Pearson; 2016.
[12] Hyndman RJ, Athanasopoulos G. Forecasting: principles and practice [Internet]. 2019 [cited 2021 Jan 27]. Available from: https://robjhyndman.com/uwafiles/fpp-notes.pdf.
[13] Farnum NR, Stanton L-VW. Quantitative forecasting method. USA: PWS-KENT; 1989.
[14] Chatterjee S, Simonoff JS. Handbook of regression analysis. New Jersey: Wiley; 2013.
[15] Farizal, alRasyid H, Rahman A. Model peramalan konsumsi bahan bakar jenis premium di indonesia dengan regresi linier berganda. Jurnal Ilmiah Teknik Industri. 2014;13(2):166-76.
[16] Kmenta, MG. Elements of econometrics. 2nd ed. New York: Macmillan; 1980.
[17] Imdad, MU, Aslam, M, Altaf, S, Ahmed, M. Some new diagnostics of multicollinearity in linear regression model. Sains Malaysiana. 2019;48(9):2051-60.
[18] Vining GG. Statistical methods for engineers. 2nd ed. Pacific Grove: Duxbury Press; 1998.
[19] Young DS. Handbook of regression methods. Boca Raton: CRC Press; 2017.

[20] Amstrong JS, Collopy F. Error measures for generalizing about forecasting methods: empirical comparisons. Int J Forecast. 1992;8:69-80
[21] Makridakis S. Accuracy measures: theoretical and practical concerns. Int J Forecast. 1993;9(4):527-9.
[22] Joseph M. Forecast error measures: Understanding them through experiments [Internet]. 2020 [cited 2021 Jan 27]. Available from: https://towardsdatascience.com/forecast-error-measures-understanding-them-through-experiments-da7ddcb0b035.
[23] Kim S, Kim H. A new metric of absolute percentage error for intermittent demand forecasts. Int J Forecast. 2016;32:669-79.
[24] Byrne RF. Beyon traditional time-series: using demand sensing to improve foecasts in volatile times. J Bus Forecas. 2012;31(2):13-9.
[25] Makridakis SG, Wheelwright SC, McGee VE. Metode dan aplikasi peramalan. Jakarta: Erlangga; 1999.
[26] George KM, Park N, Yang Z. A reliability measure for time series forecasting predictor. IFAC-PapersOnline. 2015;48:850-5.
[27] Fotios P, Konstantinos N, Georgios PS, Vassilis A. Empirical heuristics for improving intermittent demand forecasting. Ind Manag Data Syst. 2013;113:683-96.
[28] Gilliland M. The business forecasting deal: exposing myths, eliminating bad practices, providing practical solutions. New Jersey: Wiley; 2010.
[29] Dinar M, Hasan M. Pengantar ekonomi: teori dan aplikasi. Jakarta: CV. Nur Lina; 2018.
[30] Feigin G. Supply chain planning and analytics: the right product in the right place at the right time. New York: Business Expert Press; 2011.
[31] Daniel M. Pengantar ekonomi pertanian. Jakarta: Bumi Aksara; 2004.
[32] Mankiw NG. Teori makro ekonomi (terjemahan). Jakarta: Gramedia Pustaka Utama; 2003.
[33] Rahman MA, Sarker BR, Escobar LA. Peak demand forecasting for a seasonal product using bayesian approach. J Oper Res Soc. 2011;62:1019-28.
[34] Vhatkar S, Dias J. Oral-care goods sales forecasting using artificial neural network model. Proc Comp Sci. 2016;79:238-43.
[35] Aksoy A, Ozturk N, Sucky E. Demand forecasting for apparel manufacturers by using neuro-fuzzy techniques. J Model Manag. 2014;9:18-35.
[36] Schneider MJ, Gupta S. Forecasting sales of new and existing products using consumer reviews: a random projections approach. J Forecast. 2016;32:243-56.
[37] Ma S, Fildes R, Huang T. Demand forecasting with high dimensional data: the case of SKU retail sales forecasting with intra- and inter- category promotional information. Eur J Oper Res. 2016;249:245-57.
[38] Park YS, Han, E, Kim, J, Lee EK. Factors influencing the difference between forecasted and actual drug sales volumes under the price-volume agreement in South Korea. Health Policy; 2016;120(8):867-74.
[39] Kros JF, Keller CM. Seasonal regression forecasting in the U.S. beer import market. Adv Bus Manag Forecast. 2015;7:73-96.
[40] Fan ZP, Che, YJ, Chen ZY. Product sales forecasting using online reviews and historical sales data: A method combining the bass model and sentiment analysis. J Bus Res. 2017;2:90-100.
[41] Stoops GT. A comparison of forecasting methods for estimating the sales of retail firm. USA: Bowling Green State University; 1986.
[42] Kavitha S, Varuna S, Ramya R. A comparative analysis on linear regression and support vector regression. International Conference on Green Engineering and Technologies; 2016 Nov 19; Coimbatore, India. New York: IEEE; 2017. p. 1-5.
[43] Kumar A, Adlakha A, Mukherjee K. Modelling of product sales promotion and price discounting strategies using fuzzy logic in a retail organization. Ind Manag Data Syst. 2016;116:1-35.
[44] Thiesing FM, Vornberger O. Sales forecasting using neural networks. International conference on neural networks (ICNN'97); 1997 Jun 12; Houston, USA. New York: IEEE; 1997. p. 2125-8.
[45] Guo ZX, Li M, Wong WK. Intelligent multivariate sales forecasting using wrapper approach and neural networks. 10th International Conference on Industrial Informatics; 2012 Jul 25-27; Beijing, China. New York: IEEE; 2012. p. 145-50.
[46] Dodds WB, Monroe KB, Grewal D. Effects of price, brand, and store information on buyers' product evaluations. J Market Res. 1991;28:307-19.
[47] Menon RGV, Sigurdsson V, Larse NM, Fagerstrom A, Foxall GR. Consumer attention to price in social commerce: eye tracking patterns in retail clothing. J Bus Res. 2016;69:5008-13.
[48] Anthony SD. Combating cannibalization concerns. Harvard Business Review [Internet]. 2011 [cited 2021 Jan 27]. Available from: https://hbr.org/2011/02/combating-cannibalization-conc.
[49] Zhang Q, Zhao Q, Zhao X, Tang L. On the introduction of green product to a market with environmentally conscious consumers. Comput Ind Eng. 2020;139:1-16.
[50] Wu JH, Lin YC, Hsu FS. An empirical analysis of synthesizing the effects of service quality, perceived value, corporate image and customer satisfaction on behavioral intentions in the transport industry: a case of Taiwan high-speed rail. Innovat Market. 2011;7:83-100.
[51] Porral CC, Levy-Mangin JP. Store brands' purchase intention: examining the role of perceived quality. Eur Res Manag Bus Econ. 2017;23:90-5.
[52] Basker E. I was four weeks before Christmas: retail sales and the length of the Christmas shopping season. Econ Lett. 2005;89:317-22.
[53] Alper CE, Aruoba SB. Moving holidays and seasonal adjustment: the case of Turkey. Rev Middle East Econ Finance. 2004;2:203-9.
[54] Stulec I, Petljak K, Naletina D. Weather impact on retail sales: how can weather derivatives help with adverse weather deviations?. J Retailing Consum Serv. 2019;49:1-10.
[55] Keleş B, Gomez-Acevedo P, Shaikh NI. The impact of systematic changes in weather on the supply and demand of beverages. J Prod Econ. 2018;195:186-97.
[56] Murray KB, Muro FD, Finn A, Leszczyc PP. The effect of weather on consumer spending. J Retailing Consum Serv. 2010;17:512-20.