PREO Journal of Business and Management ISSN: 2619-6824
Vol. 2, Issue 2, October 2021

Powertrain Solutions Forecasting for Thailand's Automotive Industry

Natale Noppon Ghezzi (Corresponding author)
Faculty of Industrial Engineering, Kasetsart University
Bangkok, Thailand
natale.n@ku.th

Naraphorn Paoprasert
Faculty of Industrial Engineering, Kasetsart University
Bangkok, Thailand
fengnpp@ku.ac.th

Abstract

This study was conducted to forecast the demand for powertrain technologies used in passenger vehicles for the market of Thailand. The main objective of this study was to evaluate the best methods to forecast vehicles technologies used in Thailand. Powertrain technologies existing in the market and used in this study consisted of port fuel, gasoline direct injection, diesel, hybrid vehicles and electric vehicles. Since previous studies tended to find various methods to initiate their forecasting models in the original types of the vehicles, in this study, we classified the number of vehicles from the raw data into segments prior to fitting the forecasting model. One advantage of doing this fitting of forecasting model was that it could be dynamically updated even with changes in engine sizes, vehicle prices and the other physical characteristics of the vehicles. Another advantage was that we were able to classify the vehicles within the same competitive market range, reflecting the real demand for each vehicle type and obtaining a more accurate powertrain system forecast model. Results showed better overall forecast performance indices (MAD, MSE, MAPE) when compared with the results from using the data without segment classification.

Keywords: *forecasting model, powertrain solutions, data management, electric vehicles, hybrid vehicles, direct-injected gasoline injection, Internal combustion engines*

1. Introduction

Combustion technologies and internal combustion engines (ICEs) powered by fossil fuels such as gasoline, diesel and ethanol have been and are still playing a big role in the automotive industry. Today many automotive vehicles are also being paired with electric powered motors also known as "hybrid engine vehicles" (HEVs) using fossil fuels as their main energy source or a battery pack large enough to have them categorized as "full hybrid vehicles" (FHVs) pairing the energy sources of fossil fuels and battery charge together. In the latest years there is a large interest and development in vehicles using batteries as their only energy source known as "electric vehicles" (EVs). They have only electric motors as their powertrain system getting them completely independent from fossil fuels as a source of energy. Since the introduction of internal combustion vehicles in the late 1800 many improvements have been done to the technology, and until today the efficiency and capacities of ICEs have greatly improved since the first years. One of the big differences in gasoline powered ICEs used today is the implementation of the "direct-injection" system, a technology in the market since the early 2000's performing a clearer picture of the technical improvements versus engine reliability after around 2010 when compared to the previous dominant technology of port-fuel engines. However, in highly developed countries and countries with a severe interest in reducing their CO2 footprint, the usage of internal combustion engines is being steeply reduced, with many countries aiming to totally ban the use of this technology in commercial passenger cars as the process in combustion of fossil fuels to provide energy to their engines emits various pollutants as their by-products.

However, in Thailand, as in many other Southeast Asian countries, the usage of internal combustion engines is still widespread due to lower emission restrictions when compared to European countries, and less interest in electric vehicles such as in China and the US. For hybrid vehicles, the market segmentation had been in place before 2010, but a visible successive increase in the new vehicle sold amounts began only after 2017. Lastly, for electric vehicles, the true beginning in market segmentation began only after 2018, making this technology with very low sell amount numbers, but a very high percentage in the amount increase from year to year.

In summary, we will analyze data for new sold passenger cars in Thailand by categorizing the powertrain system technologies into five main technology categories:

1. Gasoline ICEs using port-fuel injection (PFI)
2. Gasoline ICEs using direct-fuel injection (DI)
3. Diesel ICEs
4. Hybrid vehicles
5. Electric vehicles (EVs)

The aims of this study were to provide a detailed categorized dataset for all passenger vehicles sold in Thailand, to analyze the information from this raw dataset, and to provide a set of different forecast models to study the level of accuracy for the performed models using the segmented database. This type of information can be helpful for both stakeholders (from suppliers to retailers) related to the automotive industry. Moreover, public citizens looking for an auto vehicle in the future could also benefit from this forecast information as an additional information for their decision making. Anyone with interest and/or has a connection with forecasting analysis may also benefit from this study. The relationship from dataset management and forecasting method in this research could provide other forecast analysts with a set of data-forecast relation and results to be implemented with their models.

The final results in this study were supposed to give a facilitated overview of the database from raw numbers, organized in a new specific structure to perform the forecast model and a summary of results and accuracy level of various forecasting models performed for that exact organized database.

The questions we wanted to precisely answer in this research in order to achieve our objective to provide the abovementioned benefits were:

 ISSN: 2619-6824

1. Is categorizing the database for an automotive forecast into vehicle segments going to affect the forecast model accuracy or not?
2. Will categorizing the data also affect smoothing constants of a forecasting model?
3. Will each type of powertrain solution technology act differently from a forecast with categorized versus non-categorized data?

We expected some differences in the results by categorizing vehicles into segments prior to performing any type of forecasting model. Since there was another organization with categorized data, we also expected to see better forecast results and smoothing constants. How each different type of forecasting model and powertrain technology would react was hard to foresee, but this would be tried to get answered by analyzing the various obtained results.

As all forecasting models are considered a theory in statistics, there is a large amount of forecasting models, each with different pros and cons. Before getting into more specific and unique models, we looked into a study forecasting automobile sales in the US using basic statistics methods by Shahabuddin (2009). The variables used in this forecast model are the following:

1. Durable industrial demand
2. Durable personal consumption
3. Discount rate
4. Non-durable industrial goods demand
5. Personal consumption
6. GNP
7. GDP
8. Population
9. Leading economic indicators: M1 (money that is liquid), M2 (M1+short-term invested assets), M2+all institutional funds.

The variables used in Shahabuddin (2009) paper as a group of forecasting parameters were highly related to economic factors, which were normally taking longer time to show their effect. Therefore, the forecasting model method used was the Statistical Package for the Social Science (SPSS) as it was able to identify lag relationship between variables. In addition, Durban-Watson statistic was used to detect the presence of serial correlation between different input variables and analyze the presence of heteroscedasticity errors.

The pro of proceeding a forecast for automotive sales with these parameters was that researchers did not have a major need in obtaining exact historical data of vehicles sold because the selected economic parameters in combination with the selected forecasting methods would be able to lead up to the final amounts of vehicle demands in the future market. However, the main drawback of this approach was that the mix usage of various parameters to forecast another set of forecasted numbers not directly related to the input data (economic factors to auto vehicle sold in this case) was more likely to have a higher level of uncertainty in the forecasted model.

Sa-ngasoongsong, et al. (2012) conducted a study of multi-step forecasting in automotive industry based on structural relationship identification. To conduct this type of forecasting model, a technique to characterize joint dynamic behavior of variables without strong restriction needs was used. Therefore, the use of Vector Autoregression (VAR), which was a natural extension of Autoregression (AR), was adapted as it is able to identify the underlying structural parameters of the overmentioned joint variables, a common method used in time-series modeling. Successively, Vector Error Correlation Model (VECM) was used to add error correction into the previously performed VAR.

Sa-ngasoongsong, et al. (2012) used a monthly sequence of new automobile sales in the United States between January 1975 to December 2010, categorizing the vehicle types by:

1. Large vehicles
2. Small vehicles

In addition to the historical sales numbers, economic indicators were applied. This study selected four economic indicators, which were:

1. Consumer Price Index (CPI) having a linear over-time trend
2. Unemployment rate at cycling pattern according to economic ups and downs
3. Gas prices with exponentially growing trend
4. Housing starts at cycling pattern according to economic ups and downs

Similarly, Homolka, et al. (2020) conducted a study in European countries on short- and medium-term car registration forecasting by using macro and socio-economic factors. In addition to the previous method of Sa-ngasoongsong, et al. (2012), a large number of economic variables were added using stepwise regression to choose the most influential economic inputs, and Adaptive Network Based Fuzzy Inference System (ANFIS) was used to make the forecast. The economic factors used in this study were divided into five main sectors, namely:

1. Macroeconomic factors indicated by unemployment rate and car registrations in the economy to which the analyzed country has the biggest exports
2. Travel demand management indicated by the public transportation costs
3. Socio-economic factors indicated by the total confidence indicator (major purchases over the next 12 months)
4. Financial conditions indicated by the short-term interest rates for households
5. Fuel and energy prices indicated by the fuel price at the gas station

The data transformation in order to fit the VAR model was similar to the previous method except that the impulse response function (IRF) was performed on stationary data, and the Augmented Dickey-Fuller test (ADF test) was used to determine whether the analyzed time series contained a unit root, and if so, how many differences were required to achieve stationary status.

Conducted by Fantazzini and Toktamysova (2015) forecasting German car sales, another interesting research using VAR is one using Google search data. In this study, Google search data refers to a tool called Google trends in offer since 2004 and is used as a leading indicator for long-term forecasting, and multivariate models. They began by dividing the economic factors into three main groups incorporating:

1. Technological aspects of the product, including innovation and technology, performance and economy of the engine, functionality, safety, space management, design, and aesthetics.
2. Promotion and sales factors, including wholesale and retail prices, customer service, advertising campaigns, and brand image.
3. Political, economic, and social environmental factors, such as organizational issues, political issues, global economic growth, ecological and physical forces, socio-cultural effects and consumer behavior.

This study was a set of multivariate models, which includes both Google data and economic variables. Google data has been used in many literatures as an exogenous variable in univariate models for short-term forecasting. The results of this study showed that models including car sales, Google data and economic variables outperformed competing models in the medium term for most of the car brands, while multivariate models outperformed other models for long-term forecast up to 24-steps ahead.

Grace Haaf, et al. (2014) used discrete choice model to predict the sensitivity of vehicle market share for the US light duty new vehicle market. Each model used utility function of a specific vehicle design that could be related to a specific group of customers.

However, in this study, many variables were omitted as they did not have enough data sources to be included, but we still considered those omitted variables for a further understanding on possibly valuable resources from the data we could have access to. The omitted variables include:

1. Indirect vehicle attributes, like consumer report rating for handling and safety as these would not be known at the time of prediction
2. Vehicle and battery maintenance cost as they did not vary substantially across conventional and hybrid powertrains. (But with today's rising battery electric vehicles technology this variable might have a different effect.)
3. Acceleration time; a covariate of horsepower and weight was used instead.
4. Range; a related fuel economy covariate was used instead since only conventional internal combustion engines and hybrid electric vehicles was considered in the study, and both did not vary substantially. (Similar to vehicle and battery maintenance cost, this point might have a different effect in the present.)
5. Top speed; a covariant of horsepower and weight was used instead.
6. Number of seats; vehicle class was used instead.
7. Two-year retained value; like consumer rating data, this is also unknown at the time of prediction.
8. Attributes specific to alternative-vehicles; they were also not relevant as this study included conventional vehicles and only a limited number of hybrid powertrains.

Massiani and Gohs (2015) used Bass model coefficients to forecast diffusion for innovative products as an empirical investigation for new automotive technologies using market data from the German automotive market. Approaches based on the diffusion model à la Bass underlie with coefficients for innovation and imitation parameters known as p and q subsequently, which represent the intrinsic driving forces together with market potential, M.

Since these parameters are essential for this type of forecasting models, this literature especially focuses on those p and q values.

The Massiani and Gohs (2015) study provided a systematic examination of the p and q parameters as well as parameters estimates based on real market data. The adoption of new technologies has often been found to follow an S-shaped curve, characterized by three different growth phases:

1. Slow take up phase where innovators are the main first users of the technology
2. Phase of more rapid growth as the technology becomes widespread
3. Slowing growth, where technology reaches its saturation point.

On the other hand, the bass model divides the potential buyers of an innovation into two groups:

1. The innovators, people who buy the product first, influenced only by external communication such as advertisement or mass media.
2. The imitators, people who buy if others have already bought the product, influenced by word-of-mouth or so called 'internal communication'.

From the previous point using the Bass model, we are able to understand that when forecasting new technologies, combining economic cost models with growth models is essential. Muraleedharakurup, et al. (2010) used Gompertz Growth Model and Logistic Growth model for predicting the market of hybrid vehicles in the UK. They mentioned that the use of Bass model will need data past the peak sales in order to achieve stable estimates and meaningful predictions, so for relative new technologies, not enough historical data was provided to adopt the model effectively. Since Logistic Model is used when initial sales make the sale of subsequent units easier and the Gompertz model is more appropriate in forecasting market penetration where initial sales do not make subsequent sales easier, they used a comparison of these two models in their study, finalizing in the use of the Gompertz curve as it presented better values of R2 and Mean Absolute Percentage Error (MAPE).

2. Methods

The methodology used in this study is characterized into two main sections, data management and forecasting models. Since we executed the forecasting models by comparing different strategies of data management, both methodology sections had a direct influence; in specific, the data management procedure had an important effect to the simplicity and correctness of the execution of the forecasting models.

2.1 Data management

The data used in this study were mainly derived from Thailand's Department of Land Transportation statistics. The availability of total registered vehicles data in details from a government organization accredits the reliability of the base data used to execute the forecasting model, which was also a very important part of a model.

For big datasets or even small ones, managing and arranging your database before executing a forecast model is essential. This step was required to facilitate the creation and modifications

 ISSN: 2619-6824

of the forecast model and it created order of data to minimize errors and mistakes. Each party might have different methods to organize data, which depended on each individual and type of data being organized. For this study, Microsoft excel was used as a database. The database structure is shown below:

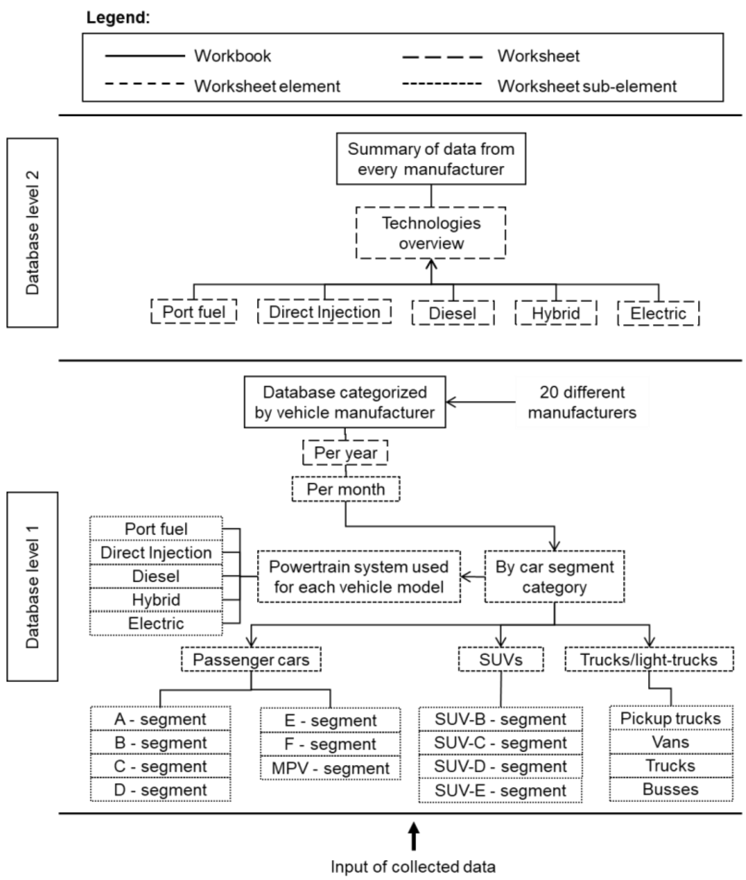

Figure 1. Database Configuration for Data Collection

Referring to the above figure, we introduced the two first database levels, where level 1 is the stage for detailed vehicle characteristics for each manufacturer, and level 2 is the overview of all collected data from each manufacturer for data storing purposes; for the following database levels, the purpose is for forecasting model execution. The reason to divide our database into sections was to prevent errors with the program by having too much information and data execution within a single file. This method also prevents data loss and facilitates debugging when executing the forecasting model.

In database level 1, we arranged our data as much as possible as each workbook had sales number for only one manufacturing brand. There was still plenty of space to insert extra information and conditions to categorize the data and prepare them to be imported for the next database level. The evaluated data were collected from the following 20 car manufacturers, which take part of 99.31%, 99.13%, 98.77% of total registered vehicles for years 2018, 2019, and 2020 (January – June), respectively. These car manufacturers are Toyota, Honda, Isuzu, Mitsubishi, Mazda, Nissan, Ford, MG, Suzuki, Mercedes-Benz, BMW, Chevrolet, Subaru, Hyundai, Tata, Proton, Kia, Ssangyong, FOMM, and Tesla.

In database level 2, we first summarized all available information and prepared to export the data according to the required information to execute the forecasting model. In this part, we also analyzed the information in interest, such as ratio level for used technology, manufacturer market share, and total sales ration by segment. Here, we only imported data of passenger vehicles from database 1; we thus excluded trucks/light-trucks information as they might mislead the scope and accuracy of the forecast model.

2.2 Forecasting models

In this study, we performed the following three forecasting models: moving average, exponential smoothing, and double exponential (Holt's method) since we evaluated that the characteristics of available dataset were compatible with these models. Also, behaviors of different data management techniques seem to be easier to analyze through basic forecasting models. We used the Mean Averaged Percentage Error (MAPE) as forecast accuracy indicator as the value given from this method was in percentage to preserve the magnitude of forecasted data, thus making it easier to compare with other datasets from different countries or even on different topics.

Moving average is a simple yet common model widely used to create a forecast. It is simply the arithmetic average of the most recent observations to forecast a value for the next upcoming period. The following formula was used:

$$F_t = (1/N) \sum_{i=t-N}^{t-1} Di = (1/N)(D_{t-1} + D_{t-2} + \cdots + D_{t-n})$$

Where:
F_t is the forecasted value of period t
D_t is the original data value of period t
N is the amount of averaged datapoints selected to forecast the upcoming value

From the above equations, it is visible that the datapoints selected are flexible, meaning that the model can be executed on averaging different ranges of period selection. As an example, varying the number of selected months or years to forecast the upcoming value is possible.
Unlike moving average, exponential smoothing creates a forecast model by weighting the raw data value and forecasted value of the previous data point to generate the value of the actual data point. The below equation was used:

$$\text{New forecast} = \alpha \text{ (Current observation of demand)} + (1-\alpha) \text{ (Last forecast)}$$
$$F_t = \alpha D_{t-1} + (1-\alpha) F_{t-1}$$

Where:
α = smoothing constant, and $0 \leq \alpha \leq 1$

Holt's method was used in this study as a double exponential forecast model. This double exponential smoothing method is designed to track time series with linear trend, using two smoothing constants:

α for the values of the series (the intercept) similarly to single exponential smoothing β for the trend (slope) of the series

as expressed by the following equation:

$$S_t = \alpha D_t + (1-\alpha)(S_{t-1} + G_{t-1})$$
$$G_t = \beta(S_t - S_{t-1}) + (1-\beta)G_{t-1}$$
$$F_{t,t+\tau} = S_t + \tau G_t$$

Where:
S_t = value of the intercept at time t. This variable is similar to single exponential smoothing, where the most current observation demand Dt is included in the equation
G_t = value of the slope at time t. This variable is the revision for the new slope which is revised by new estimations of $S_t - S_{t-1}$

3. Results

In this section, we explain the obtained results from all datasets after categorization and the results of the executed forecasting models.

3.1 Collected data

After collecting and organizing all necessary information, we evaluated all data before starting the forecast. This preliminary step was important for researchers to understand the behavior and important properties of the data used to evaluate the forecast model in order to accurately evaluate and understand the forecast results, and also be able to detect for occurring errors throughout the model.

Table 1. Correlation of Technologies

Correlation	PFI	DI	Diesel	Hybrid	EV
PFI	1	-0.10892	0.482227	0.036239	-0.28731
DI	-0.10892	1	0.694899	0.779815	0.623204
Diesel	0.482227	0.694899	1	0.519262	0.13658
Hybrid	0.036239	0.779815	0.519262	1	0.707933
EV	-0.28731	0.623204	0.13658	0.707933	1

The above table shows the correlation levels between technologies in our dataset of vehicle sales numbers in Thailand from year 2006 to June 2020. The numbers show expected values of correlations, starting with the most common used technology in Thailand, Port Fuel Injection (PFI), which could be seen as an older technology of Direct Injection (DI). A correlation of -0.1 shows that these two technologies are slightly negatively correlated; the reason for this confirms that the DI is a subsequent of PFI, and the low correlation value shows that this change still has a lot of room to enter the Thailand's market. We can see that DI is positively correlated with Hybrid and Electric vehicle (EV) technologies, with a value of 0.78 and 0.62, respectively; these values show that all these three technologies, which are considered "new" technologies for Thailand's market, are making their entry in the market with a similar pace. For Diesel, there is no major influence on other technologies as diesel technologies in Thailand are more used for their own dedicated market such as big size SUVs or big vehicles using only diesel engines. For Hybrid, there are no negative correlation values, showing that this technology has not been taking any clear place of other technologies yet, but its high correlation values with DI shows that hybrid technology is incoming to Thailand's market with a good pace. Lastly, EV has the highest negative value of correlation in the table with PFI, and EV is having highest correlation with hybrid technology; this explains that EV has strongest effect in taking place vehicles numbers from PFI, and its entry in Thailand's market shares a most similar behavior with hybrid technologies, which is understandable as both technologies are related with vehicles electrifications.

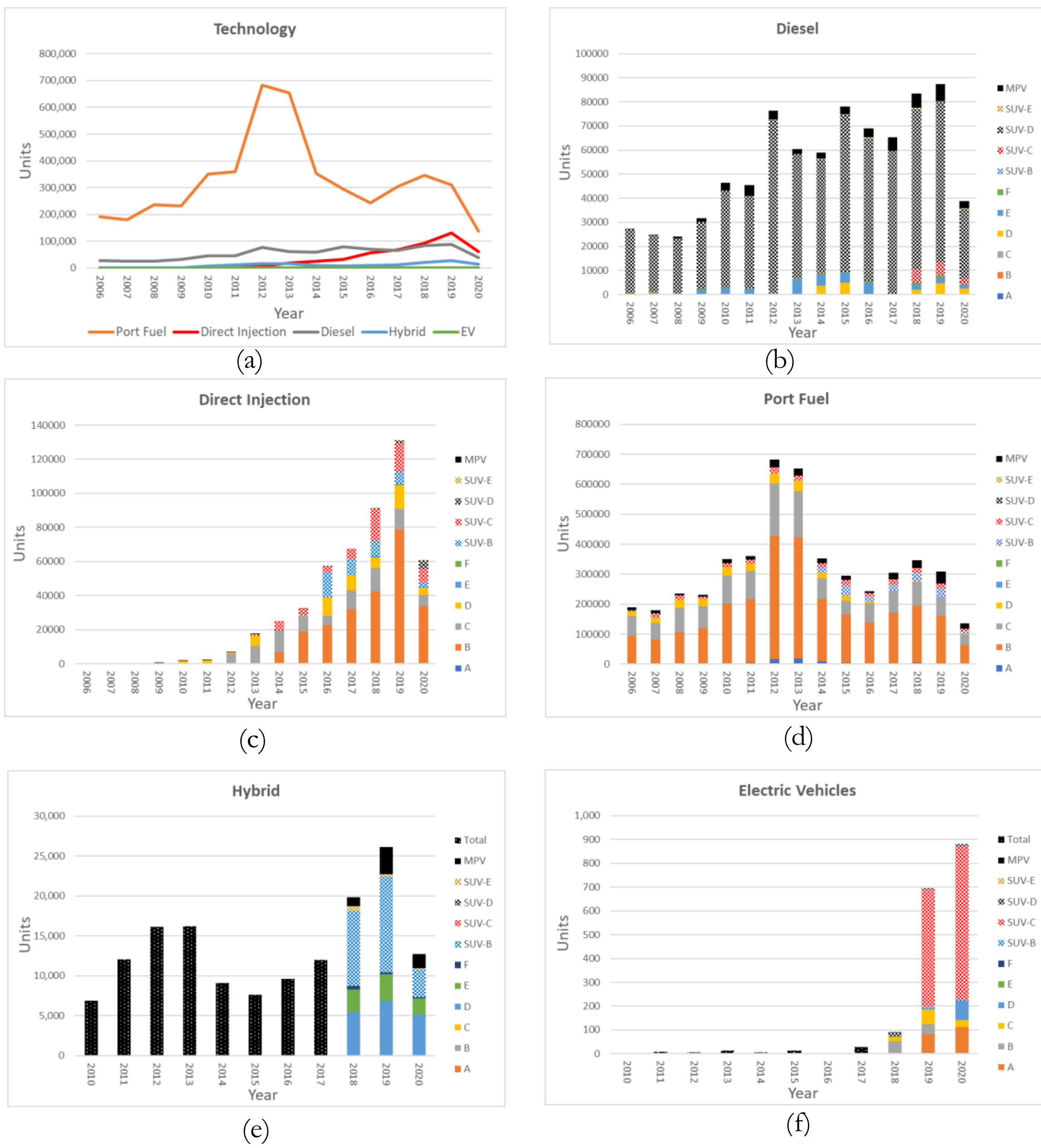

Figure 2. Visualization of Vehicles by Technologies and Segment

PREO Journal of Business and Management
Vol. 2, Issue 2, October 2021
ISSN: 2619-6824

In the graphs with data from 2006 to July 2020, we are able to understand the effects of technologies and car segments behaviors and see how much one is affected from another. In graph (a), there is a steep sales increase in years 2012 and 2013. This is mostly caused by a "first car policy" issued by the Thai government, returning a maximum amount of 100,000 THB of taxes for citizen buying their first car under 1,500 cc engine displacement, which is visible in graph (c) where there is a major increase with b and c-segment vehicles. We can see in graph (b) that, for Thailand's market passenger vehicles, diesel engines are mostly dedicated for SUV-D vehicles; the most common models used are Toyota Fortuner, Isuzu Mu-x, Nissan x-trail which originally are Pick-up Passenger Vehicles (PPVs), but we categorized them as SUV-D for their vehicle and engine sizes. In graph (c) and (d) we can see that DI technology is continuously increasing and starting to take market share from PFI technology as explained in the correlation table. Lastly, for electrification technologies, we are starting to see a constantly increasing behavior in hybrid vehicles (e) since 2015 with major values in SUV-B segment since 2018, which is mainly coming from sales numbers of Toyota CH-R model; and for Electric Vehicles in graph (f), we can see that the units are still in very low magnitude, and how big the increasing ration reached in 2019 and 2020 due to the introduction of SUV-C segment model MG-ZS EV from MG car manufacturer.

3.2 Forecast model

In this section, we observe the results of all three executed forecasting models. Here, we selected vehicles sales number from January 2018 up to June 2020 due to data availability and reasonable period of observation as values from EV technology have been available only after 2018.

The forecasting structure used in every model for categorized and non-categorized data models was to use the whole numbers for a single technology in non-categorized forecasts, and, for categorized forecasts, was to actually execute the forecast on each vehicle segment for every type of technology, and then sum up the obtained values from each segment to the total amount for each specific technology.

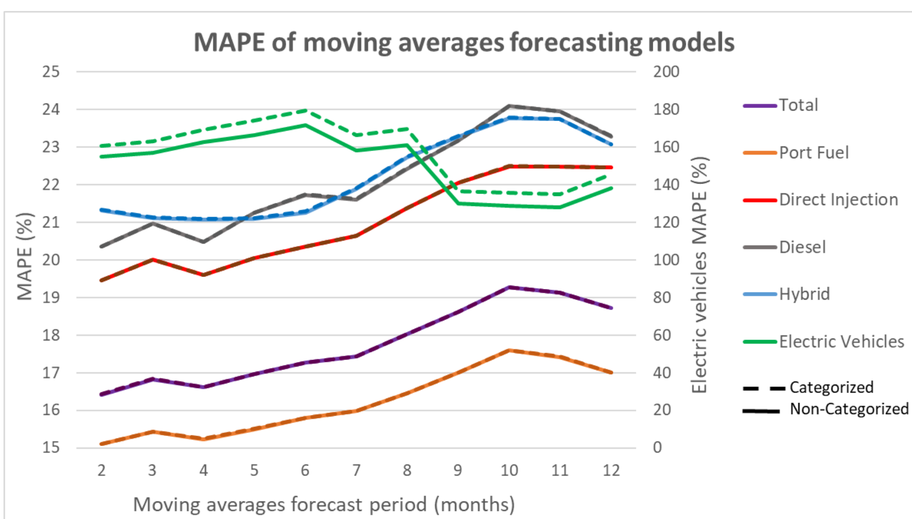

Figure 3. MAPE Values for Moving Average Forecasting Model

ISSN: 2619-6824

Starting from moving average model, we can observe that there is no visible difference in MAPE values between categorized and non-categorized data. But with EV, we can clearly see that categorized data gives worse MAPE results, and this is because there is a big difference and instability of sales numbers between each month, making the difference visible when averaging numbers directly from the total vehicles amount in that technology versus averaging from each segment and then summing it up again, causing a bigger error difference using this process. Lastly, we can also observe that with moving averages, a forecasting model using periods of 2 or 4 months to evaluate the moving averages provides best results for our dataset.

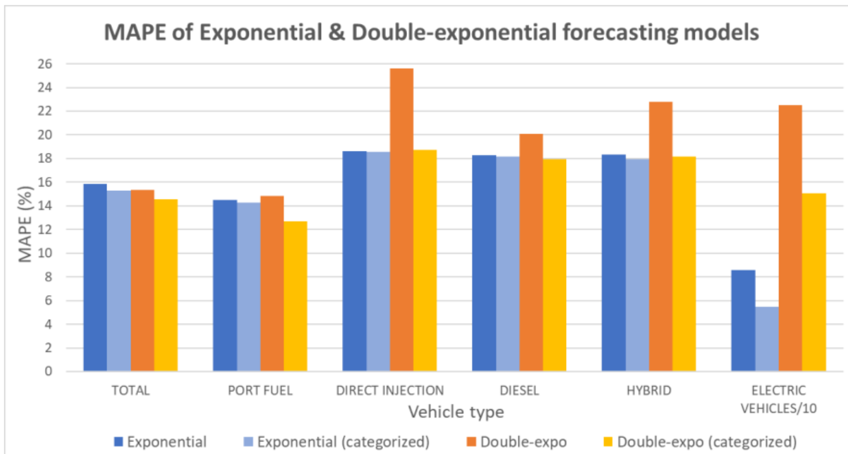

Figure 4. MAPE Values for Exponential Smoothing Forecasting Models

Unlike in moving averages models, we can see from figure 4 that with exponential smoothing, categorizing the dataset provides improvements in MAPE values. This improvement was derived from the opportunity of smoothing the forecasted data in every vehicle segment before summing everything up to a single technology system. This was why with categorized data we could obtain a more accurate forecast with exponential smoothing as the forecasting model was able to execute on smaller multiple sections. For electric vehicles, we could still see that the MAPE results were still very high, but this was still expected as by looking to the raw data for electric vehicles, monthly sales number were brutally swinging up and down due to its very low magnitude.

Table 2. α and β Values Used in Exponential Smoothing Models

Segment	Constant	Total	PFI		Diesel	Hybrid	EV
Exponential	α	0.56112	0.56745	0.58833	0.93445	0.76715	0
Exponential (categorized)	α	0.58708	0.79158	0.49905	0.72772	0.69809	0.49918
	α	0.30490	0.74833	0.70342	0.49528	0.69681	0.83374
Double-expo	β	0.14024	0.01117	0.06611	0.18850	0.00002	0.0000017
Double-expo (categorized)	α	0.39062	0.41259	0.62023	0.50063	0.58374	0.42670
	β	0.58936	0.63711	0.29927	0.22587	0.52566	0.45050

Another important indicator for exponential smoothing forecasting models is the smoothing constant. A smoothing constant with higher value leads to a forecast more dependent on previous actual data rather than previous forecasted data, which is not a good situation for making plans due to high uncertainty. From table 2, focusing on α, we can see the lower value achieved for inconstant new technologies to the market (DI, Hybrid, EV), while for constant technologies dominating and existing for a long time in the market (PFI), non-categorized data will give a lower α value, which is more desirable for planning purposes.

4. Discussion

From the above results, we obtained a clearer picture of the powertrain solutions behaviors in Thailand's automotive market. Starting from the analysis of collected data, we could observe how much effect car manufacturers had on the usage of newer powertrain solution technologies such as Hybrid and EVs. These results supported the idea of having a manufacturer-driven behavior to the future of the market as it showed that car types and technologies used for new car models directly affect the usage amounts of powertrain systems, depending on vehicle segment and manufacturer brand.

For the forecasting models results, we were able to analyze the performance and behavior for each method, on each type of powertrain solution.

With Port Fuel Injection, Holt's method of double-exponential forecasting showed best performance as expected since PFI is considered as the oldest and most commonly used technology, thus having considerably stable values and historical data slope.

Following with Direct Injection, the technological successor of PFI, double-exponential forecast performed with surprisingly high difference of MAPE between organized and non-organized data, which was derived from high uncertainty in slope behavior within each year as the models were executed from data on a monthly basis.

For diesel, there was not much difference from each type of forecast since, similarly to PFI, diesel is also an old technology with its own vehicle segments market that is still not yet being replaced by any other kind of powertrain solution in Thailand's market.

Lastly, for electrification powertrain systems (Hybrid and EV), we could observe a noticeable difference with exponential models using categorized versus non-categorized data. The reason for this behavior is the fact that the amount of vehicle segments using these technologies is still increasing. These new types of technologies market penetration are processing through manufacturers and vehicle types. This behavior causes a high fluctuation in the total number of vehicles using electrified powertrain systems, which could be more accurately analyzed when categorizing the data before performing the forecast model giving us the shown improvement in results.

In summary, categorizing the data into segments showed noticeable improvements with exponential smoothing forecasts. Traditional types of powertrain systems (PFI and diesel) perform better compared with double-exponential as both the data numbers and slope had higher stability. But for newer technologies (DI, Hybrid, EV), exponential smoothing model was recommended from all three types of performed forecasting models.

5. Conclusion

As mentioned at the introduction of this study, categorizing data for an automotive forecasting model into segments could deliver interesting results and benefits, and from a thorough evaluation in our study, we achieved the following points of conclusion:

1. Categorizing data into vehicle segments can provide a noticeable improvement in the indicators of a forecasting model.
2. Smoothing constants of forecasting models are also affected by the data segmentation process by showing more accurate and stable values since forecasts are separately executed into sub-devised elements with closer characteristics.
3. For each type of powertrain solution, data segmentation has different effects by having a stronger influence on newer technologies still on market penetration, and less effect on older technologies that have for long time been existing in the market and have a more stabilized behavior.
4. Powertrain technologies used in the automotive industry market of Thailand are highly dependent on manufacturers' initiative of introducing the technologies into their mostly sold vehicle types, such as B-segment and C-segment passenger cars.
5. Direct Injection technologies have already started replacing market shares from traditional Port Fuel Injected engines.
6. Hybrid technology is not yet replacing other powertrain solutions, but it is starting to achieve a constant increase in market penetration.
7. Electric Vehicle technology has the highest fluctuation and uncertainty in Thailand's market, but results have shown that it has high potential from customers interest side. The future behavior depending on manufacturers and practicality of having an electric vehicle in Thailand might show interesting results with steep increase for the usage of EVs.

From these findings, readers are able to obtain an overview and understand the effects and influence each vehicle segment and manufacturer can cause to Thailand's automotive industry. Moreover, forecasting and data management researchers are able to see the achievable difference in forecasting models performance by categorizing data into reasonable sub-divisions depending on type of forecast and available information.

Further improvements could be obtained by applying this type of data segmentation into other different models. Another interesting point is by applying other external factors such as applying regression methods to introduce population, economic factors, government policies and laws into the forecasting model.

References

Fantazzini, D. and Z. Toktamysova (2015). Forecasting German car sales using Google data and multivariate models. *International Journal of Production Economics 170: 97-135*. https://doi.org/10.1016/j.ijpe.2015.09.010

Grace Haaf, C., et al. (2014). Sensitivity of Vehicle Market Share Predictions to Discrete Choice Model Specification. *Journal of Mechanical Design 136*(12). https://doi.org/10.1115/1.4028282

Homolka, L., et al. (2020). Short- and medium-term car registration forecasting based on selected macro and socio-economic indicators in European countries. *Research in Transportation Economics 80*. https://doi.org/10.1016/j.retrec.2019.100752

Massiani, J. and A. Gohs (2015). The choice of Bass model coefficients to forecast diffusion for innovative products: An empirical investigation for new automotive technologies. *Research in Transportation Economics 50: 17-28*. https://doi.org/10.1016/j.retrec.2015.06.003

Muraleedharakurup, G., et al. (2010, March 25-28). *Building a better business case: the use of non-linear growth models for predicting the market for hybrid vehicles in the UK.* Proceedings of the Fifth International Conference and Exhibition on Ecological Vehicles and Renewable Energies, Monaco. http://wrap.warwick.ac.uk/72456/

Sa-ngasoongsong, A., et al. (2012). Multi-step sales forecasting in automotive industry based on structural relationship identification. *International Journal of Production Economics 140*(2): 875-887. https://doi.org/10.1016/j.ijpe.2012.07.009

Shahabuddin, S. (2009). Forecasting automobile sales. *Management Research News 32*(7): 670-682. https://doi.org/10.1108/01409170910965260

Steven N., Tava L. O. (2015). *Production and Operations Analysis*. Waveland Press.

Thailand's Department of Land Transportation statistics office. https://web.dlt.go.th/statistics/