Fuel Econonmy Analysis for Vehicular Fuel Efficiency

Rahul Agrawal, Deepak Gupta
CSE
Jaypee Int. of Inf. Tech.
Noida, India
{rahul17a1997, deepakg578}@gmail.com

Dr. Adwitiya Sinha, Megha Rathi
CSE
Jaypee Int. of Inf. Tech.
Noida, India
{mailtoadwitiya, megharathi.cs}@gmail.com

***Abstract*—Current U.S. fuel economy standards require new vehicles to achieve at least 34.1 miles per gallon on average by 2020. Significant changes are required to achieve this target. New vehicles must have better engine, drive-type, transmission, and fuel type combination such that number of vehicle in which greater average mileage achieved will be more advanced. These changes will not only increase average mileage, but also reduce fuel usage, and thus reducing its corresponding production energy demands as well. Using XGBoost and Random Forest classifier we created a model using vehicle specific performance parameters to predict Green House score and combined average mileage.**

Keywords— Fuel efficiency; Green House score; Average MPG; Random Forest; XGBoost

I. INTRODUCTION

Keeping environmental issues in mind, fuel efficiency is the major factor which buyers look upon to decide which vehicle to purchase, especially when fuel prices are sky rocketing. Since, fuel efficiency is regularly monitored by federal and state governments of U.S. to minimize the impact of vehicle, to the environment.

In U.S. fuel efficiency, measured in terms of distance travelled per amount of gasoline used. The standard unit of measuring in the U.S. is miles per gallon or MPG. The Environmental Protection Agency (EPA) requires automakers to report MPG ratings for city driving, highway driving, and combined.

The MPG rating and fuel economy of a vehicle depends on various attributes. In this paper, we used data as provided by U.S. government, to describe factors affecting a vehicle's MPG rating and created a model to predict MPG and greenhouse score based on various vehicle attributes. The goal is to find and study some of these attributes and to determine their effect upon fuel efficiency.

The combined miles per gallon reflects fuel efficiency and is our response variable - in other words, we described, tested, and modelled the relationships between combined MPG and the rest of the variables. We hypothesized that the following factors may affect a vehicle's fuel efficiency. These factors were tested and presented in the analyses part of this report.

By using vehicle attributes that are fixed in the medium run, such as vehicular dimensions, power train architecture, and drive type, effects of factors like fuel prices, car prices and maintenance can be accounted for. Various omitted variable bias's also have a significant role to play in behavior pattern of customers, in purchasing any particular model of vehicle.

For instance, many vehicle characteristics that are known to have strong correlation with each other:

- vehicle size and fuel economy,
- fuel economy and vehicle power,

are negatively correlated. Due to these factors being strongly associated, including these parameters in the regression analysis may lead to unsatisfactory, inconsistent results because of collinearity. As a result, the findings of regression model would also be difficult to interpret as whole.

Taking the case of model having variable cost (in $) per mile and distance (in miles) per gallon in regression, we find that the coefficient of cost per mile is negatively related. This further directs to the fact that people will less likely chose a vehicle that has higher cost per mile value, all other parameters being same. Same is the case for MPG, which is also negatively correlated, that everyone would prefer to buy vehicles with higher mpg.

II. LITERATURE

More recent literature addresses endogeneity in vehicle choice and technology decisions in novel ways. Incorporating not only the decision of what new vehicle to purchase, but also the decision on how long to hold and whether to buy a used vehicle, and how many miles to drive a vehicle each year. Market prices for the vehicles are also treated as endogenous variables in these models.

The fact that a manufacturer often uses the same engine platform across different vehicle models is used to develop an instrument for fuel economy, since vehicles in different classes that use the same engine tend to exhibit similar power and fuel economy characteristics.

[**] Similar papers on vehicular demand which also demand no use discrete choice models, also contend with endogeneity issues. For example, model of new and old vehicles, indicate that more number of new, more fuel efficient vehicles will be sold in coming years where prices of gasoline are expected to rise, as supply diminishes slowly.

Nowadays, quite a few different vehicle manufacturers have started adjusting vehicular parameters as and when required to meet both consumer's demands and other keep in pace with other manufacturers' automobile strategies [**].

III. METHODOLOGY

The study will examine the hybrid/non-hybrid passenger vehicles in the U.S., which consists around 300 million cars on road. We have collected data for following years 2011 - 2018, from the www.fueleconomy.gov website and collated them using R program.

After fetching vehicle parametric dataset [2] a model of MPG and greenhouse has been created. In this, we have developed the model using the following approach:

- Situation analysis is being used to determine what all technologies changes can be adopted and plausible changes that can be done in vehicle parameters and sales are required, to meet directed fuel economy goals by 2020.
- Automotive material demand is being monitored regularly under the various future vehicle scenarios, based on different vehicle class sales and how each vehicle class's sales is likely to reach.

The outputs of the model are the Green house score and combined MPG in each scenario. These are the important factors associated with environment, which are being affected. (Figure 1)

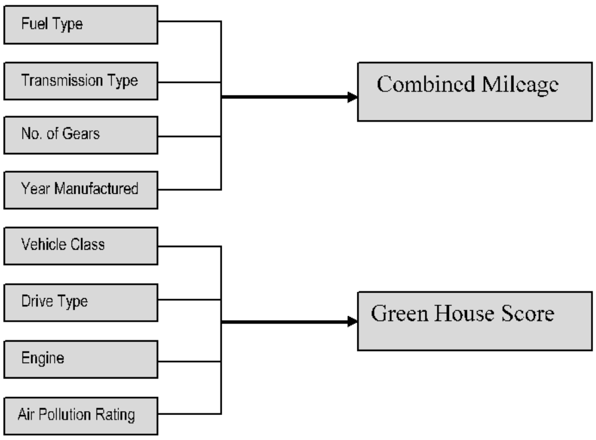


Figure 1. Model Overview

The factors associated with car such as weight, design specification, aerodynamic efficiency, etc. Are not included in our analyses. Also, the mileage of plug-in hybrids and electric vehicle is not a viable option for this study, due to different mileage unit.

III A. ALGORTIHM USED

RANDOM FOREST:

Random forests is an ensemble learning method that used in various data science problems like regression, classification. The selection of single hyper-parameter tree from multiple decision trees at the time of training will result the best model of Random Forest. Random Forest gives the class that is mode of the other classes or mean prediction of the single tree. It chooses the best model from various decision tree and also avoid over fitting up to some extent.

Random forest is similar to bootstrap algorithm which contains various decision trees with different variables at root node. If our dataset consists 5000 observations with complete population along with 20 variables then random forest tries to build a different model with different size of sample and having different variables at root node.

Let's understand the process, Random forest will take a random stratified sample of 1000 observation and 10 anonymous initial variables at root node to build a predictive model. The whole process will repeat up to n times (n estimators) and then it will make a final prediction on each observation of validation dataset. Final prediction is a result of each prediction of every decision tree. You can say that the final prediction is the mean of each prediction of decision tree.

XTREME GRADIENT BOOST

XGBoost is an advanced algorithm that has recently been using in various data science challenges.

XGBoost is best for its high accuracy and performance. It is an enactment of boosted decision trees. XGBoost is an ensemble of decision trees where in weighted combinations of predictors is taken. XGBoost works on the same lines of Random Forest, but there is a difference in working procedures. The similarities are that the features extracted in both the cases is completely random in nature.

Steps that algorithmic carried throughout the process.

Step 1: Learning of a regression independent variables.
Step 2: Calculation of Residuals.
Step 3: Prediction of Residuals.

Parameters mentioned below is used for calculating error rate. Error in prediction is given by:

$$J = (z, \hat{z}) \tag{1}$$

where

$$J(.) = X(z[i] - \hat{z}[i])^2 \tag{2}$$

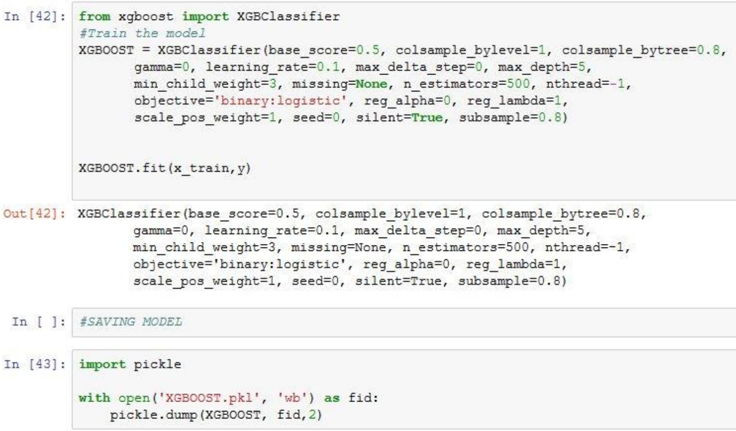
$\hat{z}$ can be adjusted to reduce the error, by using the following formula:

$$z[i] = z[i] + \alpha f[i] \tag{3}$$

where

$$f[i] \approx \nabla J(z, \hat{z}) \tag{4}$$

Each learner estimates the gradient of the loss function. Gradient Descent is used to take sequence of steps to re-duce sum of predictors weighted by step size α.



```
In [42]: from xgboost import XGBClassifier
         #Train the model
         XGBOOST = XGBClassifier(base_score=0.5, colsample_bylevel=1, colsample_bytree=0.8,
                gamma=0, learning_rate=0.1, max_delta_step=0, max_depth=5,
                min_child_weight=3, missing=None, n_estimators=500, nthread=-1,
                objective='binary:logistic', reg_alpha=0, reg_lambda=1,
                scale_pos_weight=1, seed=0, silent=True, subsample=0.8)


         XGBOOST.fit(x_train,y)

Out[42]: XGBClassifier(base_score=0.5, colsample_bylevel=1, colsample_bytree=0.8,
                gamma=0, learning_rate=0.1, max_delta_step=0, max_depth=5,
                min_child_weight=3, missing=None, n_estimators=500, nthread=-1,
                objective='binary:logistic', reg_alpha=0, reg_lambda=1,
                scale_pos_weight=1, seed=0, silent=True, subsample=0.8)

In [ ]: #SAVING MODEL

In [43]: import pickle

         with open('XGBOOST.pkl', 'wb') as fid:
             pickle.dump(XGBOOST, fid,2)
```

Figure 2. XGBoost Model to Predict Combined MPG

III B. FLOWCHART

We started by collecting data for different years, and applied various data processing methods as per the data requirement. Various steps applied:

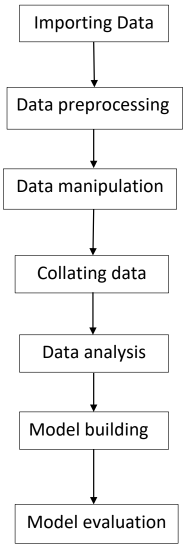


Figure 2: Flowchart

- First we removed some of the unnecessary columns such as Sales Area, Standards, Standard Description, Under hood ID's and source file names.
- Then separated transmission type and number of transmissions from the transmission column, substituting CVT transmission for a '1' in transmission type variable.
- Removed hydrogen and electricity operated cars as we are interested in gas based cars (including hybrids) or vehicles with an MPG rating.
- Took only MPG of gasoline engines from the hybrid fuel cars, such that we compare similar fuel type i.e. gas and thus keeping the response variable in Miles per Gallon.
- For duplicate entries we took a mean of numerical variables to gauge the average values of our variable of interest (Combined MPG)

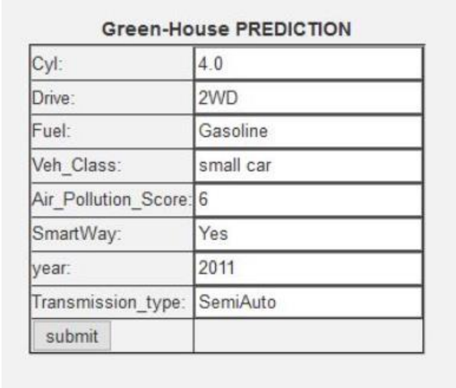


Figure 3. Predictive Model for Greenhouse Score

This was performed on data for years from 2011 - 2018, and finally collated them to make one single data file. Next step was to create a predictive model. We created a model and trained it on the dataset. Thus, a model was successfully created to predict Green House Score and let user evaluate his vehicle.

III C. CURRENT PRODUCTION SITUATION

While assessing the possible impacts of fuel efficiency of vehicular data, it is important to understand how changes in real time technologies could affect the number of vehicles purchased and types of it. Exploring this factual question need the use of models that analyze the consumer's choices of vehicles.

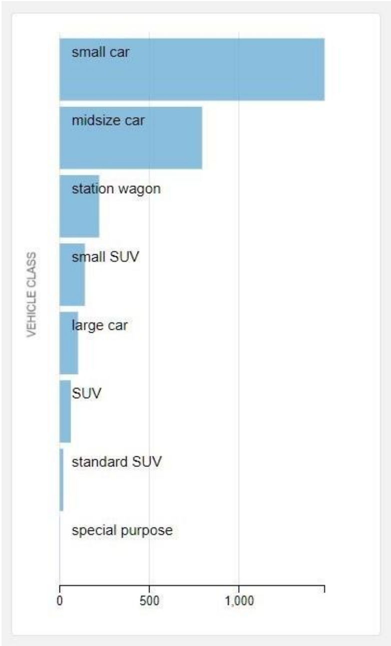


Figure 4. Vehicle Distribution Over Vehicle Class Type

The factual literature has spent comparatively less time analyzing the efficiency of the production rate of the latest vehicle in the market. It is usually assumed that vehicle manufacturers take advantage of all profit making chances. Although, we analyze that it disconnects between the actuality of low cost opportunities to increase fuel economy and the non-fulfillment of the market to invest in such development may not be entirely due to ambiguity in customers decision making process.

Over the years many vehicles with different fuel type launched in market. Gasoline based cars continue to dominate market share over the years. But slowly and steadily, the electric vehicles are increasing and taking market share from Gasoline cars due to being more cost effective in long term.

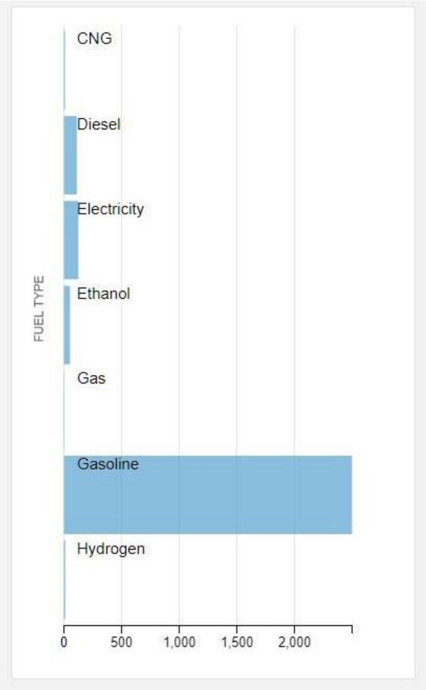


Figure 5. Vehicle Distribution Over Fuel Type

As previously mentioned, several analyses show that recent improvements being done in engine efficiency, have resulted into improvement of other vehicle parameters, such as power and acceleration, rather than fuel economy.

Car manufactures are maximizing sell profits by constantly improving small and minor details of vehicle parameter, rather they should add improvements such that both fuel economy and other vehicle characteristics are improved.

III D. USING PREDICTIVE MODEL FOR ANALYSIS

Using the model build we predicted the values of Greenhouse Score, for each fuel type and keeping other predictor variables fixed at values in Figure 6.

Green-House PREDICTION

CYLINDER SIZE:	6.0
DRIVE TYPE:	2WD
FUEL:	Gasoline
VEHICLE CLASS:	small car
AIR POLLUTION SCORE:	6
SMART WAY:	Yes
YEAR:	2013
TRANSMISSION TYPE:	SemiAuto
Predict	

Figure 6. Input Values for Other Variables

The result we found were interesting, and as follows:

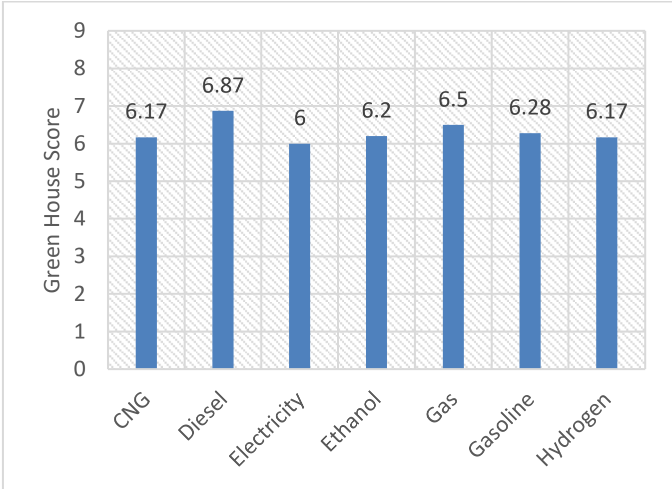


Figure 7. Greenhouse Score Over Fuel Type

Using the model build we also predicted the values of Greenhouse Score, for each transmission type and keeping other predictor variables fixed at values in Figure 8.

Green-House PREDICTION

CYLINDER SIZE:	6.0
DRIVE TYPE:	2WD
FUEL:	Gasoline
VEHICLE CLASS:	small car
AIR POLLUTION SCORE:	6.28
SMART WAY:	Yes
YEAR:	2013
TRANSMISSION TYPE:	Manual
Predict	

Figure 8. Input Values for Other Variables

The result we found are as follows:

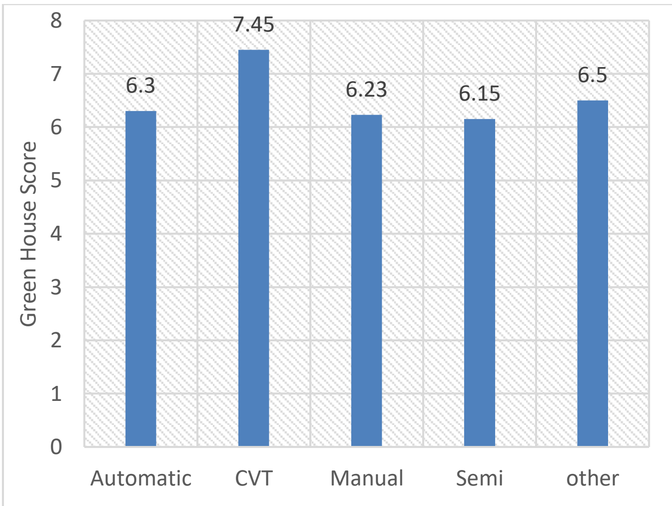


Figure 9. Greenhouse Score Over Transmission Type

IV. RESULTS

We have the necessary information to deduce the passenger vehicle fuel usage and the impact it has on environment under the various circumstances. Results have been projected for years up to 2020, keeping in mind the changes required to meet the 2018 targets, and then fuel consumption and other vehicle characteristics remaining constant after. Under the diverse situations, we expect no drastic changes, but rather changes will happen linearly, as seen from 2011 up to 2018 and will continue in similar fashion now onwards.

In addition, some findings were unexpected:

A. More the number of transmission, lower is the car's MPG.

B. Although it is believed that manual transmission is more fuel efficient than automatic transmissions, during our analysis, the study showed no significant advantages of manual transmission. Instead engines with continuously variable transmission (CVT) are the more fuel efficient transmission type, which are often used in hybrid vehicles.

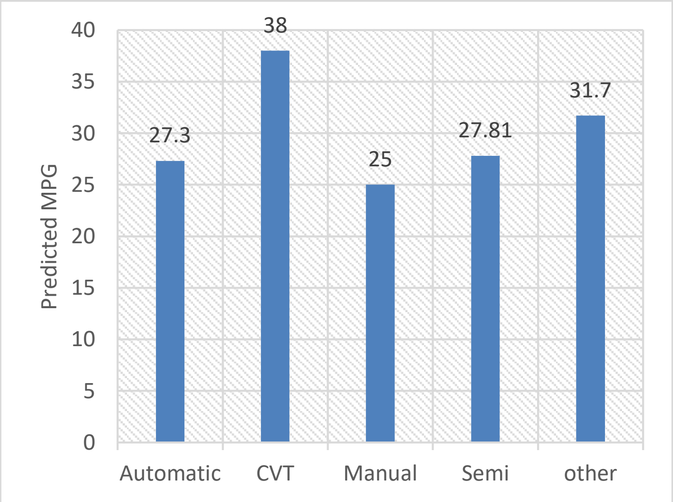
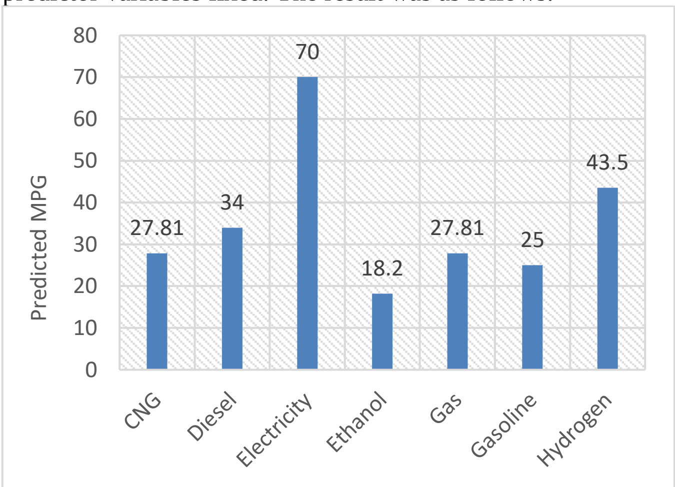


Figure 10. Predicted MPG Keeping Other Fixed

Using the predictive model build, we predicted the values of average mileage, for each fuel type and keeping other predictor variables fixed. The result was as follows.



The results show drastic difference mileage for electric vehicle. EV seem to have at least double the mileage as compared to any other fuel type. This can also be a major reason for shift to EV. As technology gets better, the battery get more optimized, better motor engine are available, thus more reliable and better performance from cars.

Sales of electric vehicle is increasing rapidly, moreover by 2020, it is estimated the cost of electric vehicles are expected to be the same as conventional fuel powered equivalents. While this will steal revenue currently being generated from conventional fuel powered cars, it will also open new pool of opportunities for new car manufacturers in the area of new mobility services, new electric charging infrastructure and for battery manufacturers.

Electric vehicle sales across all major regional markets are on the upswing. In 2017, in terms of sales, the US had up to 200,000 units sold. Most US policy proposals indicate that support for greenhouse gas reduction will be scaled back. Despite this, the plug-in vehicle sales grew 27% to almost 200,000 units, and market penetration finally crossed the 1 % mark.

IVA. 2018 VEHICLE SCENARIO

Vehicle launches grew rapidly in the last few years. Small sedans continue to dominate market share over the years. But slowly and steadily, the share of medium size sedans, SUVs kept increasing and take market share from small sedans. (Figure 12)

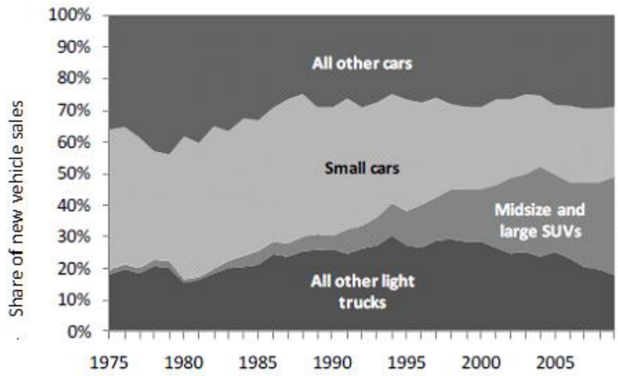


Figure 11. Trend of SUV's Sales In The U.S. [1]

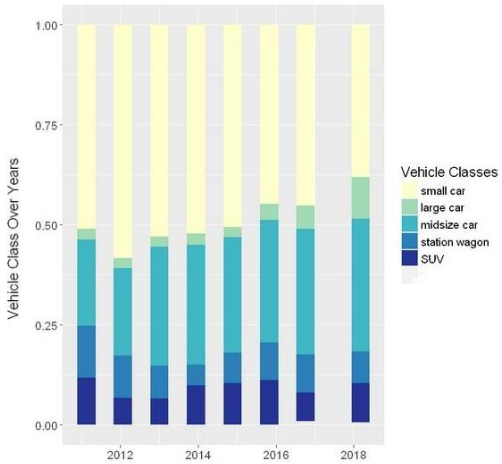


Figure 12. Class Wise Percentage Over the Years

Two-wheel drive remains the more preferred drivetrain type over the years, but slowly more and more four-wheel drive vehicles are coming to the market in recent years. It is a trend that manufacturers are adopting four-wheel drive in their SUVs to meet the needs for various road conditions. (Figure 13)

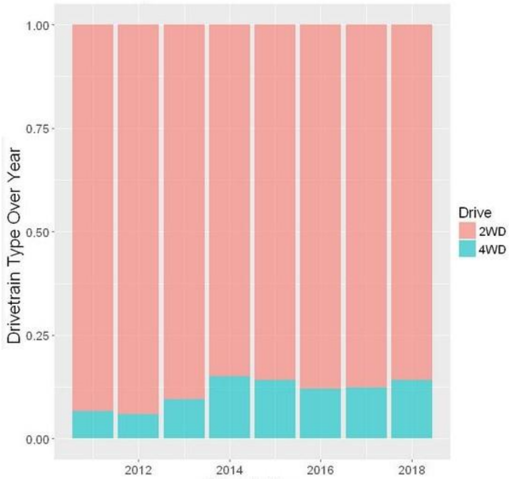


Figure 13. Share of Drive Type in SUVs

Predicted state of future vehicle sales figure are not proposed as sales estimates, but a way to discover potential in an indeterminate future. Few situations that highlight different, near future approaches for improving fuel economy are:

A. *Vehicle size reduction* – SUVs are in general larger, heavier and thus less fuel-efficient vehicles of all vehicle type, yet there sales continue to increase, mainly due to luxury. Thus a need to invoke shift to smaller vehicle like small sedans, with higher mpg is required.

B. *Initializing improved, more fuel efficient powertrains* – Standard internal kindling engines are presumed to continue enhance in terms of fuel efficiency, with lower friction and revolution like direct injection. The normally fuel consumption of the upcoming vehicles can reduce further by launching more electricity,gasoline, diesel, and hybrid and plugin hybrid electric vehicles.

V. CONCLUSION

In this exercise, we studied factors affecting fuel economy by analyzing different vehicle features. Our results confirm trends in fuel economy observed in the U.S. automobile industry. However, there are additional factors that were not captured in our analysis such as weight, design specification, etc. Some of the assumptions behind the statistical techniques we employed may not be valid and can be looked at in more detail in further studies.

However, almost all précised advancement in vehicles have been counterbalance by advancement in horsepower and acceleration performance of vehicles, rather than being obtain in actual fuel consumption reduction. For upcoming vehicles in this analysis, it is reckoning that this trend will move backward given more rigid fuel economy regulation. There will be some advancement in performance in upcoming vehicles in market, but not at former rates. Partially of the upcoming vehicle efficiency advancement will be devoted to improving fuel economy of vehicular market.

VI. REFERENCES

[1] U.S. Environmental Protection Agency, 2009. Light-Duty Automotive Technology, Carbon Dioxide Emissions, and Fuel Economy Trends:1975 through 2009, EPA420-R-09-014, November 2009.

[2] https://www.fueleconomy.gov/feg/download.shtml

[3] Evaluating the Consumer Response to Fuel Economy: A Review of the Literature, Gloria Helfand and Ann Wolverton, Working Paper # 09-04 August, 2009

[4] Atkinson, S F and Halvorsen, R., A new hedonic technique for estimating attribute demand an application to the demand for automobile fuel efficiency, The Review of Economics and Statistics, 66 (1984), 417-26.

[5] Battese, George E., A note on the estimation of Cobb-Douglas Production Functions when some explanatory variables have zero values, Journal of Agricultural Economics, 48(1997), 250-252.

[6] U.S. Environmental Protection Agency, 2009. Draft Regulatory Impact Analysis: Proposed Rulemaking to Establish Light-Duty Vehicle Greenhouse Gas Emission Standards and Corporate Average Fuel Economy Standards, Assessment and Standards Division, Office of Transportation and Air Quality, EPA-420-D-09-003, September 2009.

[7] Goldberg, Pinelopi K., The Effects of the Corporate Average Fuel Efficiency Standards in the US, The Journal of Industrial Economics XLVI (March 1998).

[8] Greene, David L., CAFE OR PRICE?: An Analysis of the Effects of Federal Fuel Economy Regulations and Gasoline Price on New Car MPG, 1978-89, International Association Energy Economics, 11(3), 37-58 (1990).

[9] Executive Summary, Light-Duty Automotive Technology and Fuel Economy trends: 1975 through 2008, United States Environmental Protection Agency (2008