APRIL 23, 2021

ASSESSING THE IMPACT OF GASOLINE PRICES ON HYBRID AND ELECTRIC VEHICLE (HEV) SALES IN THE UNITED STATES (US) FROM 1999-2018

SUHAAS RAMANI
THE GEORGE WASHINGTON UNIVERSITY

Introduction

Hybrid Electric Vehicles (HEV) have been in production in the US market since 1999, with the Honda Insight being the first mass-produced car. They combine the gasoline operated internal combustion engine with at least one battery operated electric motor to function.[1] Electric motors are highly efficient in comparison to their gasoline counterparts and use a third of the energy of a gasoline-powered vehicle to power 59 to 62 percent of the car.[2] This assessment is quite telling and is possibly the unique selling point for HEVs, however the issue with electric batteries is their storage capacity. Even though they might be more efficient, they still do not have the capacity to power vehicles to the same extent as gasoline tanks. Hence why the hybrid alternative became attractive and existed prior to the success of the electric vehicle, which not only provides a cost-effective alternative to gasoline but also reduces the risk of a lack of power from the poor storage capacity of the battery. From an initial average sale of 17 cars in 1999 there has been a rapid expansion in sales to 343,219 by 2018. This is an approximate growth of more than two million percent in nearly ten years. More importantly, this forceful change in tastes and preferences could be attributed to consumers seeking more cost-effective and energy-efficient alternatives to their gasoline powered vehicles, an increase in the production and distribution of electric and hydrogen fuel stations, and from a societal shift in thinking to a more environmentally conscious narrative with rising temperatures and global warming, among others.

Historically there had always been interest in creating an all-electric mode of transport that reduced the dependence on oil. This especially took form in the US after the oil embargo in 1973, that was placed by the members of the Organization of Arab Petroleum Exporting Countries on nations that were perceived as supporting Israel during the Yom Kippur War.[3] There was a complete halt on oil exports entering the US market, which caused a spike in the global oil prices and thus propelled the existing car industries to consider alternatives to gasoline powered vehicles. Billions of dollars were spent by auto manufacturers on researching and developing a commercially viable alternative to the gasoline powered car. From both an environmental perspective and governmental initiative, programs were launched in the US to push automakers into investing and creating gasoline-alternatives. In the 1990s for instance, California enacted a program to promote zero-emission vehicles which gave rise to GM's EV1 in 1996 which was not only expensive to produce but also failed to garner the necessary traction in the market due to its limiting driving range and was therefore dropped.[4] In 1993, the Clinton administration propelled the Partnership for a New Generation of Vehicles (PNGV), which tasked the big three automakers in the US, GM, Fiat-Chrysler and Ford and other smaller firms to develop a car that can travel 80 miles per gallon of gasoline. The target of these initiatives was to produce energy-efficient cars that ran with no cost to the environment.

[1] US Department of Energy
[2] Usher (2019)
[3] Alhaiji (2005)
[4] Usher (2019)

It is oil which sparked the very thought process for US automakers to consider producing HEVs, however now as HEVs gather more appeal, one must evaluate how much of the consumption is dependent on oil. The narrative has now shifted from the producer needing alternatives to sell due to governmental and environmental pressure, to consumers switching their consumption patterns. This paper aims to evaluate to what extent the impact of HEV sales has been because of changing gasoline prices.

Research Approach

Prior to the statistical analysis that was conducted to examine the relationship, the data for each variable was collected from governmental databases from 1999 to 2018. The data for HEV sales and the number of hydrogen and electric charging stations were attained from the US Department of Energy, gasoline price data was taken from the US Energy Information Administration, and the data for CO_2 emissions was attained from the Environmental Protection Agency.

Once the data was obtained, the main hypothesis was set up to analyze the impact.

$$H_o: The\ true\ difference\ in\ means\ between\ gasoline\ prices\ and\ HEV\ sales\ equals\ 0$$
$$H_a: The\ true\ difference\ in\ means\ between\ gasoline\ prices\ and\ HEV\ sales\ does\ not\ equal\ 0$$

An initial Shapiro-Wilk test was run on the HEV sales and gasoline price variables to test their normality in order to conduct the two-sample t-distribution test, evaluate the significance of the results and understand whether the null should be rejected or not. Thereafter, the linearity was examined for each independent and dependent variable before running the linear regression model. The other independent variables that were included in this model are the percentage of hydrogen fuel stations, percentage of electric charging stations and percentage of CO_2 emissions. After deriving the linear regression model, summary statistics were produced to assess the impact of each variable within the model, their significance, the Variance Inflation Factor (VIF) and the Multiple R-squared value. Finally, diagnostic plots were produced to understand the pattern of the residuals, normality of the residuals in the model and identify whether there are any outliers and leverage points.

Results

In order to test the hypothesis, it was necessary to test the normality of the variables being analyzed for the t-distribution test to apply. From the Shapiro-Wilk test on HEV sales and gasoline prices, the results of the p-values were 0.277 and 0.05959, respectively[5]. Given that the values for both are greater than 0.05, it can be concluded that they are indeed normal. Upon establishing this basis, the conditions for the t-distribution test were satisfied and run to assess whether the null hypothesis should be rejected. From the test, the p-value attained was miniscule, with a value of

[5] Appendices F and E

0.000001343, indicating that there is sufficient evidence at the 0.05 significance level to reject the null hypothesis and conclude that the true difference between the means of gasoline prices and HEV sales is not 0. Rather, there is a relationship that exists which can be explored further.[6]

The linearity for each independent and dependent variable were assessed prior to running the linear regression model. In the process, what was observed was that the variables for the number of hydrogen fuel stations, number of electric charging stations and amount of CO_2 emissions were non-linear in nature, and therefore had to be logged.

Characteristic	Beta	95% CI[1]	p-value	VIF[1]
gas_price	93,936	21,575, 166,296	0.015	2.5
log_hydrogen	18,770	184, 37,357	0.048	1.4
log_electric	70,733	1,763, 139,703	0.045	3.2
log_co2	456,257	-576,877, 1,489,391	0.4	1.7

[1] CI = Confidence Interval, VIF = Variance Inflation Factor

$R^2 = 0.861$

Table 1: Multilinear Regression Table

From the results displayed in the table above, it can be seen that all variables are significant at the 0.05 significance level other than the percentage of CO_2 emissions. The beta values for the independent variables are all positively linear in relationship with the gasoline prices. More importantly, there is a very high positive linear relationship between the variable of interest gasoline price and the HEV sales. For every one unit increase in the gasoline price, the model indicates that there will be an increase in HEV sales by 93936 units. This is quite a steep rise and showcases how impactful gasoline prices are in determining HEV sales in the US car market. Similarly, if there is a one percent increase in either the hydrogen fuel stations, electric charging stations, or even the CO_2 emissions, then the number of HEV sales will increase by 18770, 70733 and 456257 units, respectively. For all the independent variables that are highlighted within the model, what can be seen is that the percentage of CO_2 emissions has the greatest effect on the HEV sales. This follows the current global narrative regarding global warming and climate change, where consumers and governmental bodies are becoming more cognizant of the rising temperatures. Furthermore, through the VIF, the multicollinearity of the model can be checked and since the values are quite low with the maximum being 3.2 for the percentage of electric vehicles, it can be safe to assume that one predictor is not being linearly predicted from the other predictors by much. Finally, from this table it can be seen that the R-squared value is quite high at 0.861. Which indicates

[6] Appendix G

that the independent variables in the model are able to successfully predict most of the impact on the dependent variable.

In order to get a better understanding of the linear regression analysis of the model, core diagnostic plots were run. In examining each plot, one can appreciate how the residuals are operating and whether there are any influential cases affecting the outcome of the model. In Appendix A, the residuals versus fitted plot is graphed. Here one can visually see that there is no distinct pattern of the fit of the residuals along the line, indicating that there are no non-linear trends in the residuals or of non-constant variances. They are randomly spread out about the line. In Appendix B, the residuals appear to be normally distributed in the Q-Q plot that has been graphed, thereby indicating that the linear regression model is good enough to contain most of the variability in the data. In Appendix C, the Scale-Location plot shows that the residuals are for the most part equally spread out about a nearly horizontal line, which satisfies the condition for homoskedasticity. Finally, in Appendix D, the residuals versus leverage plot provides an indication of whether there are any influential cases that might be affecting the regression model. It can be clearly seen from the graph that point 11 might be skewing the data upwards but not by much, similarly points 13 and 20 appear to be leveraging the data downwards. However, even though these points might be influential enough to distort the model, they still lack merit to be excluded from the model as they fail to cross the Cook's Distance lines which are demarcated by broken red lines on the plot. Overall, by observing these diagnostic plots, the regression model satisfies the necessary assumptions required and allows for further interpretation to be made of the results with existing literature.

Discussion

The impact of gasoline prices as shown through the linear regression analysis is significant, where if there is an increase in one unit of the gasoline price, there will be an increase in HEV sales by 93936 units. Through previous literature one can see this impact is not novel, but rather a recurring trend in academic research. Beresteanu and Li (2011) in their paper *"Gasoline Prices, Government Support, and the Demand for Hybrid Vehicles in the United States"*, argue as per their research that hybrid vehicle sales in 2006 would be 37% lower had gasoline prices stayed at the 1999 levels.[7] Furthermore even in research that is indirectly related to HEVs, much has been done in assessing how gasoline prices affect consumer decisions in purchasing cars with lower fuel economies. In a 2016 paper, Barla, Couture and Samano (2016) identified that there is a short-run impact of a rise in gasoline prices on consumer decisions to purchase vehicles with lower fuel economies. They discovered that a gasoline price increase will instigate registration of new vehicles with fuel economies less than 27.2 MPG and will simultaneously lower the purchase of vehicles above that limit.[8] This assessment is still relevant in our discussion of HEV sales, for HEVs are vehicles with lower gasoline consumption and hence have a much lower fuel economy in comparison to their gasoline counterparts. Similarly, Diamond (2008) also ran a multilinear regression

[7] Beresteaunu and Li (2011)
[8] Barla, Couture and Samano (2016)

model in his policy paper for three of the most sold hybrid models in the US and identified that gas prices were the strongest and most significant predictor of state market share for HEVs.[9] From previous research it can be clearly seen that gasoline prices have a very significant impact on HEV sales or in sales with vehicles with a lower fuel economy, this corroborates the results identified in this paper.

Limitations and Scope for Improvements

Naturally there will be limitations within the model, and there will be avenues through which the current research can be improved for more detailed assessment of the impact that gasoline prices have on HEV sales.

This research process can contain omitted variables, a threat to internal validity, because the R-squared value is not perfect, there are still certain variables if included can explain the relationship better. This can affect the causal mechanism and variables such as lithium-battery prices or raw aluminum prices could have been included to assess whether the internal components that make up HEVs can affect the production value and therefore the sale value of HEVs. Another major threat to internal validity would be the political economy. This is because there are definitely endogenous policy shocks and legislation that have been made and are targeted towards HEVs, oil usage and CO_2 emissions that would impact the nature in which the independent variables that are associated with these policies respond. It would be essential to standardize for these policy impacts in order to exclusively assess the individual relationships between each independent variable and HEV sales. There is also one key threat to external validity that is present within this research that needs to be accounted for. This is the threat to interaction in setting. We do not know how accurately these results hold in different environments. Such as other countries, which have different economies and varied levels of car usages, or even availability of HEVs and charging stations to support the relationships. It would be wise to standardize for developmental factors across borders in order to assess how gasoline prices affect countries with varied levels of growth and development.

Other ways in which this model can be improved is instead of assessing on a national scale, we can observe on a state-wide basis in the US, where different states have varied levels of impacts. This can be based on their individual developmental stages, governmental policies or even geography. Doing a state-wide analysis would make a more holistic and sound study in order to assess the individual effect that gasoline prices have on HEV sales. Finally, this study has a very small sample size and a limited number of observations per sample size, which raises the issue of statistical power. This means that there can be an issue of low statistical power, where the probability of committing a type II error, a false negative, is high. Essentially, the null hypothesis might have been incorrectly rejected when it is in fact true. A way to resolve this issue would be to increase the sample size by more years and take a span of thirty to fifty years instead of nineteen.

[9] Diamond (2008)

References

Alhajji, A. F. "THREE DECADES AFTER THE OIL EMBARGO: WAS 1973 UNIQUE?" *The Journal of Energy and Development* 30, no. 2 (2005): 223-37. Accessed April 13, 2021. http://www.jstor.org/stable/24812665.

BARLA, PHILIPPE, ETIENNE COUTURE, and MARIO SAMANO. "Gasoline Prices and Fuel Economy of New Vehicles in Quebec." *Canadian Public Policy / Analyse De Politiques* 42, no. 2 (2016): 181-93. Accessed April 15, 2021. http://www.jstor.org/stable/24883711.

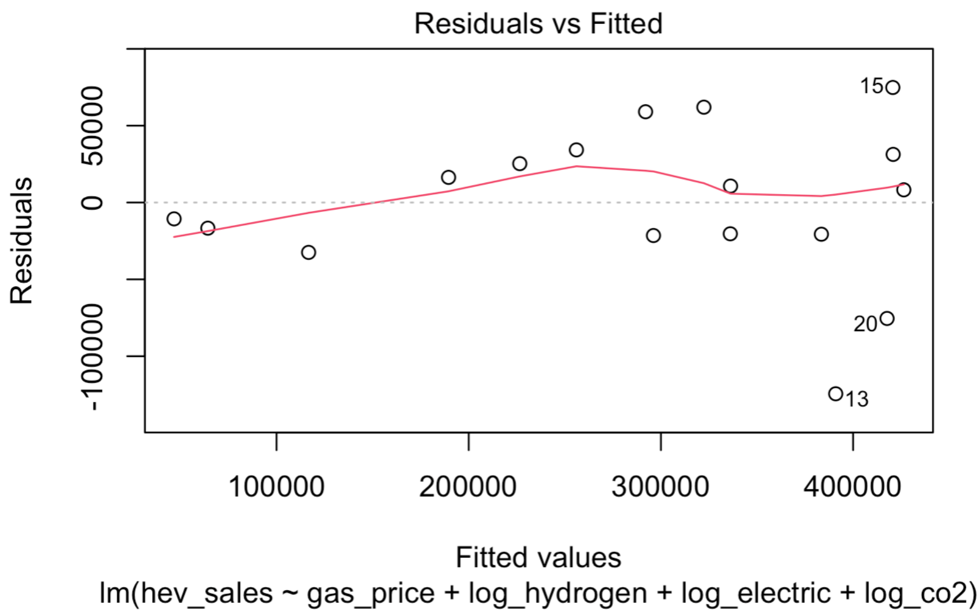
Beresteanu, Arie, and Shanjun Li. "GASOLINE PRICES, GOVERNMENT SUPPORT, AND THE DEMAND FOR HYBRID VEHICLES IN THE UNITED STATES." *International Economic Review*52, no. 1 (2011): 161-82. Accessed April 15, 2021. http://www.jstor.org/stable/23016626.

Diamond, David B. 2008. Public policies for hybrid -electric vehicles: The impact of government incentives on consumer adoption. Ph.D. diss., George Mason University, http://proxygw.wrlc.org/login?url=https://www-proquest-com.proxygw.wrlc.org/dissertations-theses/public-policies-hybrid-electric-vehicles-impact/docview/250882287/se-2?accountid=11243 (accessed April 15, 2021).

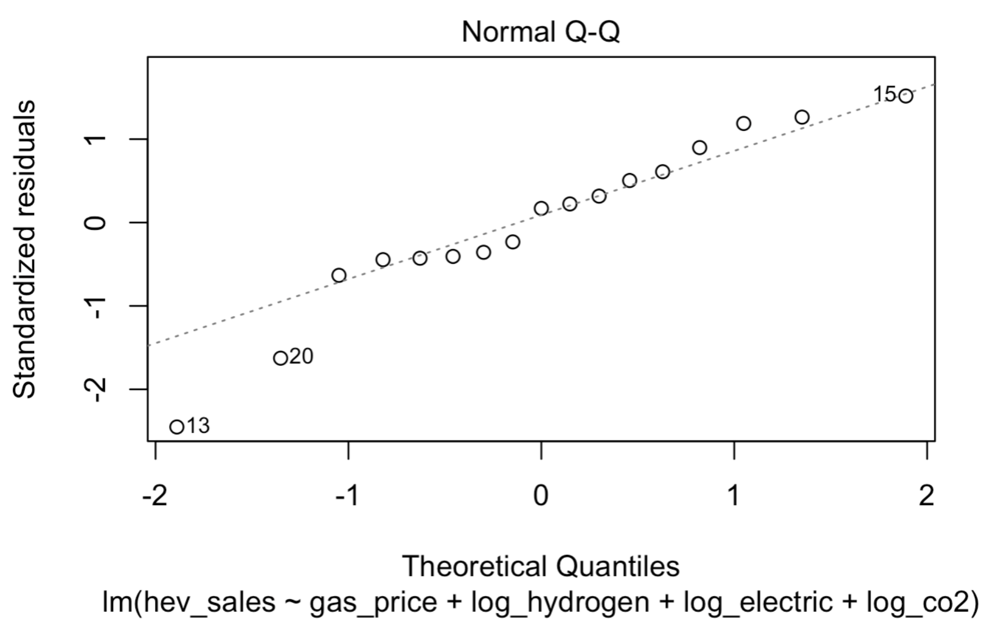
Hybrid Electric Vehicles. Accessed April 15, 2021. afdc.energy.gov/vehicles/electric_basics_hev.html.

USHER, BRUCE. "THE RISE OF ELECTRIC VEHICLES." In *Renewable Energy: A Primer for the Twenty-First Century*, 81-95. New York; Chichester, West Sussex: Columbia University Press, 2019. Accessed April 13, 2021. http://www.jstor.org/stable/10.7312/ushe18784.11.

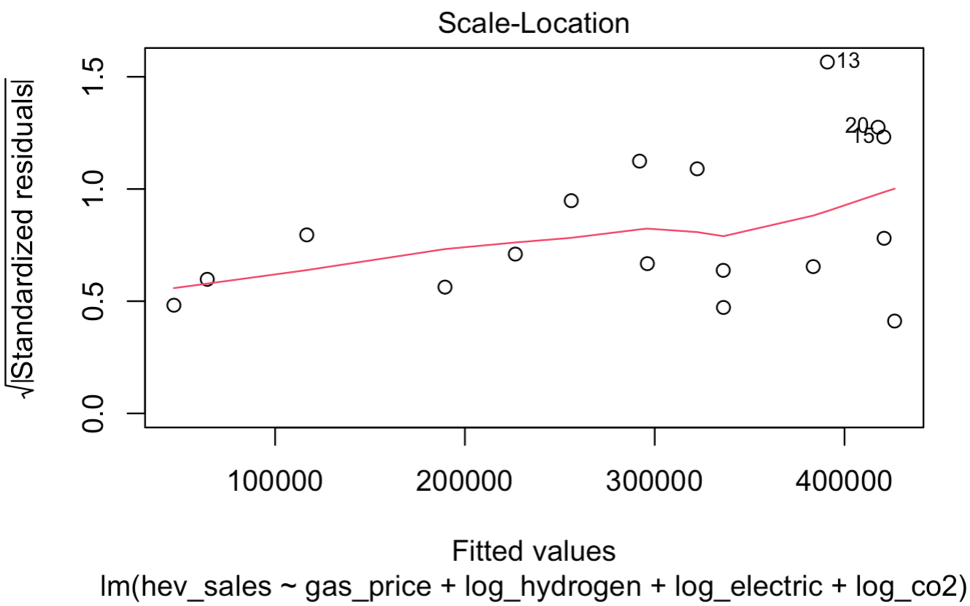
Appendices



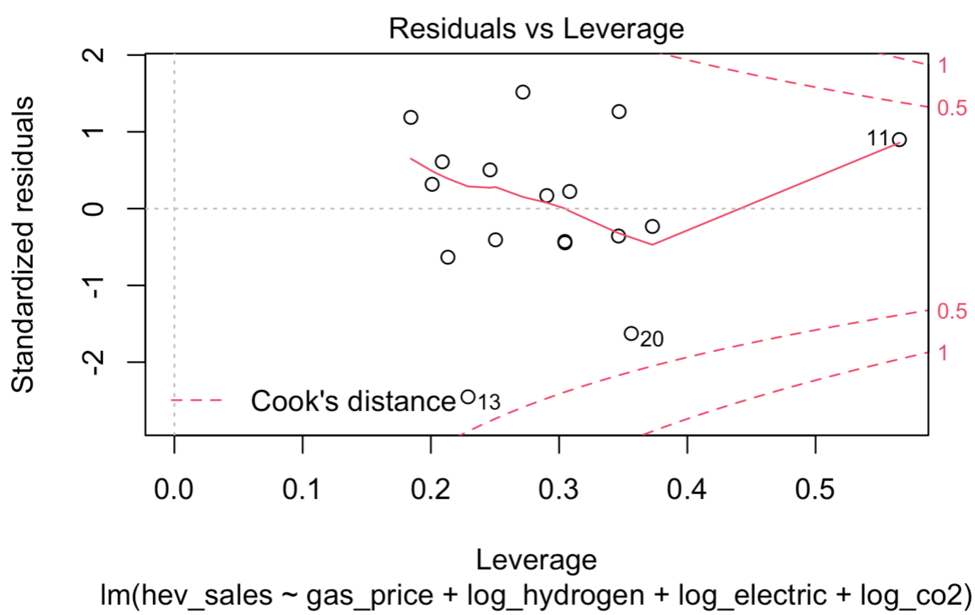
Appendix A: Residuals vs. Fitted Plot



Appendix B: Normal Q-Q Plot of the residuals



Appendix C: Scale-Location Plot of the residuals



Appendix D: Residuals vs. Leverage Plot

```
        Shapiro-Wilk normality test

data:  df$hev_sales
W = 0.90847, p-value = 0.05959
```

Appendix E: S-W Normality Test for HEV Sales

```
        Shapiro-Wilk normality test

data:  df$gas_price
W = 0.94333, p-value = 0.277
```

Appendix F: S-W Normality Test for Gasoline Prices

```
        Welch Two Sample t-test

data:  df$hev_sales and df$gas_price
t = 6.921, df = 19, p-value = 1.343e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 173427.4 323795.3
sample estimates:
   mean of x    mean of y 
248613.85000      2.48685 
```

Appendix G: Welch Two Sample t-test for HEV Sales and Gasoline Prices

R code used

```
pacman :: p_load(pacman, tidyverse, dplyr, ggplot2, cowplot, stringr, pastecs,
         tidyr, BSDA, readxl)

options(scipen = 9999)

#HEV Purchases

hev_sales <- read_excel("hev_sales.xlsx", "Sheet1")

colnames(hev_sales)[2] <- "Total HEV Sales"

#co2 emissions

co2_emissions <- read_csv("co2_emissions.csv")
co2_emissions <- t(co2_emissions)
co2_emissions <- co2_emissions[-c(1:10), ]
co2_emissions <- co2_emissions[, -c(1, 3, 4, 5)]
co2_emissions <- as.data.frame(co2_emissions)
colnames(co2_emissions)[1] <- "Fossil fuel combustion: carbon dioxide"
co2_emissions$`Fossil fuel combustion: carbon dioxide` <- as.numeric(co2_emissions$`Fossil
fuel combustion: carbon dioxide`)
log_co2 <- log(co2_emissions$`Fossil fuel combustion: carbon dioxide`)
log_co2 <- as.data.frame(log_co2)

#Oil Prices

oil_prices <- read_excel("oil_prices.xls", sheet = "Data 1", skip = 2)
oil_prices <- oil_prices %>%
  select(Date, `U.S. All Grades All Formulations Retail Gasoline Prices (Dollars per Gallon)`)
oil_prices <- oil_prices[-c(1:9), ]
oil_prices <- oil_prices[-c(21:22),]


# Alternative Fueling Stations

alt_fueling_stations <- read_excel("alt_fueling_stations.xlsx")
alt_fueling_stations <- alt_fueling_stations %>%
  select(Year, `Electric*`, Hydrogen)
```

```
alt_fueling_stations <- alt_fueling_stations[-c(1:7, 28), ]
log_electric <- log(alt_fueling_stations$`Electric*`)
log_hydrogen <- log(alt_fueling_stations$Hydrogen)
```

 # Combining datasets

```
df <- cbind(hev_sales, log_co2, log_hydrogen, log_electric, oil_prices)
df <- df[, -c(6)]
row.names(df) <- NULL
colnames(df) <- c("Year", "hev_sales", "log_co2", "log_electric", "log_hydrogen", "gas_price")
```

 # Understanding whether the independent data in question is normal

```
shapiro.test(df$hev_sales)
shapiro.test(df$gas_price)
```

 # Both are normal given that their p-values are above 0.05.

 # Performing t-test on the hypothesis

```
t.test(df$hev_sales, df$gas_price)
```

 # The data is extremely significant at the 0.05 sig level, indicating that there is a relationship between retail gasoline prices and HEV sales.

 # Conducting a multilinear regression

```
model <- lm(hev_sales ~ gas_price + log_hydrogen + log_electric + log_co2, data = df)
model
summary(model)
plot(model)
```

 # Producing an aesthetic display of the regression analysis

```
library(gtsummary)
model %>%
  tbl_regression() %>%
  add_vif() %>%
  add_glance_source_note(include = r.squared)
```