US NEV Sales Regression Analysis

Final Report Section B

Team 3: Ariana Richardson, Berkay Becu, Rameen Gauher, Sen Yang, Yuanting Fan

December 3, 2024

1 Introduction

An electric vehicle (EV) primarily uses electricity for propulsion, rather than gasoline or diesel. This includes plug-in electric vehicles (PEVs), battery electric vehicles (BEVs), and hybrid vehicles (HEVs) or plug-in hybrid vehicles (PHEVs) [1]. EVs are more sustainable than combustion engines, with lower environmental impacts, including fewer tailpipe emissions and reduced carbon footprint [2]. They also offer better energy conversion efficiency and quieter operation, leading to lower costs per mile and maintenance [3].

While EVs date back to the 19th century, mass production began in the 2010s with models like the Nissan Leaf in 2011 [3]. This spurred significant growth, with China leading the market in plug-in vehicles [4]. In the U.S., highway-capable EV sales exceeded 4.6 million units by December 2023, making it the third-largest EV market globally [5].

2 Problem Statement

All these developments motivate us to research EV sales in the U.S. using historical data dating back to the 2010s. We will focus on NEVs (New Energy Vehicles), a term that encompasses BEVs, PHEVs, and FCEVs (Fuel Cell Electric Vehicles). Our goal is to perform a time series multiple linear regression analysis of NEV sales to develop a regression model that predicts NEV sales based on key macroeconomic and market factors. The model will also analyze the purchase patterns of NEVs over time, identifying trends, seasonal patterns, and potential lag effects.

3 Data Description

3.1 Variables of Interest

To perform a time series multiple linear regression analysis for NEV sales prediction, we first identified relevant variables for both the response and predictors. The analysis focuses on the period from December 2010 to December 2023 (13 years) due to the relatively recent development of the NEV market in the U.S. The response variable is the monthly number of NEVs sold (the sum of PHEV and BEV sales), with data collected from Argonne National Laboratory reports [6].

For the predictors, we categorized the variables into four main groups: macroeconomic, market-specific, demographic, and sector-specific factors.

Macroeconomic Factors: These standard economic statistics describe how macroeconomic conditions affect EV sales. Monthly or quarterly data from the Federal Reserve Economic Data (FRED) website [7] was used. Key variables include:

- GDP (U.S.): Measured quarterly in billions of dollars.
- Money Supply (M1, M2): Monthly data in billions of dollars.
- Federal Funds Effective Rate: Monthly percentage rates.
- Durable & Nondurable Goods Consumption: Monthly data in billions of dollars.
- Unemployment Rate: Monthly percentage.
- Residential & Non-residential Fixed Investment: Quarterly data in billions of dollars.
- Per Capita Disposable Income: Monthly data in dollars.
- PPIBC/PPIBM/PCE: Indices measuring changes in battery and electrical component costs.

Demographic Factors: We focused on factors affecting the primary target group for NEVs, based on research indicating that middle-aged, highly-educated employees (25-55) are the main adopters [8]. Variables included:

- U.S. population: Quarterly data.
- Percentage of employees with a bachelor's degree or higher: Monthly data.
- Percentage of middle-aged employees (25-55): Monthly data.

Data was accessed from the Bureau of Labor Statistics (BLS).

Market-specific and Sector-specific Factors: These factors include both market dynamics and competition. We considered:

- Light-duty vehicle production and inventory: Monthly data from FRED [7].
- Fuel prices: Gasoline prices for ICE vehicles and electric retail prices for NEVs, collected monthly via FRED [7].
- Infrastructure and government incentives: Number of electric ports and new energy policies, collected monthly via the Alternative Fuels Data Center website [9, 10].
- ICE vehicle sales: Monthly data from Argonne [6].
- Battery cost and Tesla Model S price: Monthly data from the Alternative Fuels Data Center and car price tracker websites. Missing Tesla S prices were imputed using the nearest available data.

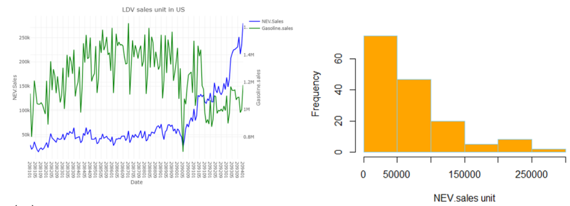
For sector-specific factors, we considered the impact of the COVID-19 pandemic, particularly the 2020–2023 chip shortage, which disrupted the automobile industry. This was treated as a categorical variable (1 for the pandemic period, 0 otherwise), with the pandemic period defined from January 2020 to May 2023.

In total, we selected 26 predictors, including one categorical and 25 numerical variables. We also introduced time-related variables (year, month, and date) to account for temporal trends. This resulted in a data set covering 157 months (157 data points) with 29 variables, with most data collected monthly. For quarterly data, linear interpolation was applied.

3.2 Variables: Quantitative Analysis

3.2.1 Response Variable

We began our data analysis by examining the response variable, monthly NEV sales, to understand its distribution and the effects of time (trend and seasonality). The response variable, NEV sales, exhibited a left-skewed distribution (Figure 1-b), indicating an initial phase of smaller sales followed by rapid growth in the market after 2019. This upward trend is further visualized in Figure 1-a, which shows the growing popularity of NEVs since their mass adoption in 2010. To assess the impact of time on the response, we constructed a small regression model that considered only the date and month (as categorical) variables. Figure 2 presents the model summary, which indicates that the date is a significant factor (with a p-value less than 2e-16). However, there is no statistical evidence of seasonality, as the month variables are not statistically significant (the smallest p-value belongs to December at 0.0533).



(a) LDV Sales Unit in United States

(b) NEV Sales Histogram

Figure 1: Response Variable Statistics

```
Call:
lm(formula = data$Total.NEV.Sales ~ data$Date + as.factor(data$Month))

Residuals:
   Min     1Q Median     3Q    Max 
-79738 -25424  -1119  18530 122044 

Coefficients:
                          Estimate Std. Error t value Pr(>|t|)    
(Intercept)             -1.262e+06  8.664e+04 -14.565   <2e-16 ***
data$Date                3.077e+01  2.013e+00  15.284   <2e-16 ***
as.factor(data$Month)2   3.076e+03  1.362e+04   0.226   0.8217    
as.factor(data$Month)3   1.923e+04  1.362e+04   1.412   0.1601    
as.factor(data$Month)4   1.356e+04  1.362e+04   0.996   0.3211    
as.factor(data$Month)5   1.809e+04  1.362e+04   1.328   0.1862    
as.factor(data$Month)6   1.675e+04  1.362e+04   1.229   0.2209    
as.factor(data$Month)7   1.651e+04  1.362e+04   1.212   0.2276    
as.factor(data$Month)8   1.741e+04  1.363e+04   1.278   0.2034    
as.factor(data$Month)9   1.526e+04  1.363e+04   1.120   0.2647    
as.factor(data$Month)10  1.295e+04  1.363e+04   0.950   0.3437    
as.factor(data$Month)11  1.143e+04  1.363e+04   0.838   0.4032    
as.factor(data$Month)12  2.607e+04  1.338e+04   1.949   0.0533 .  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 34720 on 144 degrees of freedom
Multiple R-squared:  0.6273,	Adjusted R-squared:  0.5963 
F-statistic:  20.2 on 12 and 144 DF,  p-value: < 2.2e-16
```

Figure 2: Regression Results under Time Variables

3.2.2 Response versus Predictors

For the analysis of predictors, we wanted to visualize the relationship with the response to validate our initial variable selection. The complete set of scatter plots can be found in Appendix A. To summarize this analysis, we found predictors negatively correlated with response such as battery cost and unemployment rate (see Figure 3). This makes sense since lower battery cost means cheaper NEVs prices which would increase the sales. If unemployment increases, fewer people would be buying expensive goods such as cars which would explain negative relationship. We also found predictors that are positively correlated with response such as gasoline and electric prices (see Figure 4). The second one is counter-intuitive but it is evident that the electric prices increases across time due to inflation which coincides with rising trend of NEVs sales. Furthermore, charge cost of electric vehicles is becoming cheaper despite that electric price increases compared to fuel(gasoline) cost so we observe this counter-intuitive positive relation.

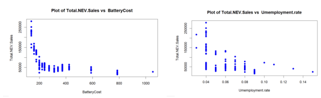


Figure 3: Negatively correlated predictors

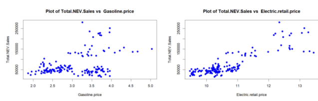


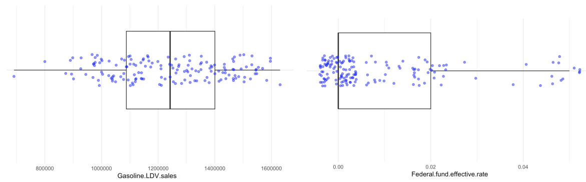
Figure 4: Positively correlated predictors

3.2.3 Correlation Analysis: Predictors

To observe the relation between the predictors, we conducted a correlation analysis. The full table can be found in Appendix B. The analysis showed that some macroeconomic variables are very strongly positively correlated such as PPIBM,PCE,GDP consumption and investment variables, correlation coefficient is higher than 0.9. We also observed unusual strong negative correlation between the battery cost and demographic variables such as population and percentage high-educated employees, middle-aged employees (coefficient less than -0.9). This may be coming from the fact that there is a negative trend of battery cost opposed to positive trend in demographic variables. However, this high correlation among predictors may result in multicollinearity where we address this issue in the following sections.

3.2.4 Data Points: Outlier Detection

We used three methods for outlier detection: Z-score, IQR (Interquartile Range), and Modified Z-score. The Z-score method identifies points which lie beyond three standard deviations from the mean. The IQR method classifies data points as outliers which lie outside Q1 - 1.5×IQR and Q3 + 1.5×IQR range. The Modified Z-score uses median instead of mean and MAD instead of standard deviation and values exceeding 3.5 are flagged as outliers. Figure 5a and 5b shows the outlier results analysis for two of our most important variables. We further found that during the COVID-19 period, several economic indicators showed outlier values, but these were retained as they represented legitimate market conditions rather than data anomalies. Since our dataset was relatively small to begin with, we did not remove the remaining outliers since that could be very well be just noise in the data, nevertheless, we made sure to use robust regression methods so our results are not affected by them.



(a) Outlier detection for Gasoline LDV Sales

(b) Outlier detection for Federal Fund Rate

Figure 5: Outlier Detection Examples

4 Regression Analysis

4.1 First Approach: Stepwise MLR

With 29 variables and only 157 data points, the number of predictors is disproportionately large compared to the sample size. To achieve meaningful insights while avoiding overfitting, it is necessary to employ statistical methods for variable selection instead of including all variables in the regression model. To address the challenge of high dimensionality, we employed stepwise regression for variable selection. The stepwise regression is an iterative approach that systematically selects or removes variables based on specific criteria, such as p-values or changes in model fit, and is widely used for simplifying regression models. In the initial step, we included all 29 variables in the stepwise regression model. A variance inflation factor (VIF) analysis was then conducted to assess multicollinearity among predictors. The initial regression results are presented in Appendix C. After performing a both-sides stepwise regression, 18 variables were retained in the model. Most of the selected variables were statistically significant, and the overall regression performance was statistically significant, with an F-statistic of 464.3 and a p-value less than 0.001. The model achieved an adjusted R-squared value of 0.9816, indicating a strong fit to the data.

4.1.1 Multicolinearity Correction

Next, we assessed the multicollinearity of the retained variables using VIF. The results are summarized in Appendix D. The VIF results revealed significant multicollinearity issues, as several variables exhibited VIF values exceeding acceptable thresholds. To address this, we excluded variables with VIF values greater than 62, based on established guidelines for multicollinearity diagnostics. However, given the exploratory nature of our analysis and the observed time series patterns, we retained "time" (represented by "Month" and "Date") as control variables despite their high VIF values. This decision was made to account for temporal effects that could impact the outcomes. After removing several variables, we run the stepwise regression again, resulting in a model with 12 variables. The summary of the model's performance is in Appendix E. Most of the selected variables were statistically significant, and the overall regression performance was statistically significant, with an F-statistic of 221 and a p-value less than 0.001. The residual standard error was 12,910, with an adjusted R-squared of 0.9442, indicating a strong model fit. Then, we performed the The VIF analysis, which showed manageable levels of multicollinearity for most variables, but still, Date and PPIBM had high VIF values (see Appendix F).

4.1.2 Autocorrelation Correction

We then tested autocorrelation issues for our second-iteration stepwise model. The Durbin-Watson test showed that the DW value of the model was 0.92266, with a p-value of nearly 0. This means we reject the H0, indicating true autocorrelation is greater than 0. To fix this issue, we decided to add lagged variables into the model. We first used the ACF test to determine which lagged variables should be added (see Figure 6).From the ACF of Residuals, the residual autocorrelation at lag orders 1, 2 and 3 is significant, so we created these lagged variables and added them into the regression model. Also, we removed Data and M2SL for their high VIFs and potential higher multicollinearity after adding lagged variables. After that, we ran the stepwise model again, and the final regression model contained 18 predictors(15 original variables plus 3 lagged variables, (see Figure 7), 6 of which are significant.

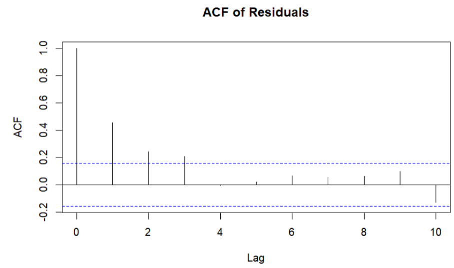


Figure 6: ACF of Residuals

```
Residuals:
     Min       1Q   Median       3Q      Max
-25173.0  -4280.0   -141.2   4555.6  27542.0

Coefficients:
                                                 Estimate Std. Error t value Pr(>|t|)
(Intercept)                                    -2.893e+05  1.867e+05  -1.549  0.12362
Month                                           3.843e+01  2.504e+02   0.153  0.87826
Covid.19                                        1.067e+04  4.351e+03   2.452  0.01547 *
Number.of.new.energy.policies                  -4.377e+01  2.772e+01  -1.579  0.11663
Production                                      8.577e-03  1.823e-02   0.470  0.63878
Inventory                                      -6.654e-03  9.340e-03  -0.712  0.47745
Gasoline.LDV.sales                              5.741e-02  5.484e-03  10.468  < 2e-16 ***
Federal.fund.effective.rate                     1.644e+05  2.021e+05   0.813  0.41751
Umemployment.rate                               1.960e+05  6.749e+04   2.903  0.00431 **
PPIBC                                          -3.383e+02  5.127e+02  -0.660  0.51055
PPIBM                                           7.253e+02  4.568e+02   1.588  0.11467
Tesla.model.S.price                             1.286e+00  8.995e-01   1.429  0.15518
BatteryCost                                     1.138e+01  2.354e+01   0.484  0.62948
Percentage.of.employees.who.are.middle.aged..25.55.  2.024e+05  2.490e+05   0.813  0.41784
Gasoline.price                                 -5.091e+02  2.151e+03  -0.237  0.81321
Electric.retail.price                          -4.284e+03  1.980e+03  -2.164  0.03226 *
lagged_value1                                   5.052e-01  6.457e-02   7.823 1.32e-12 ***
lagged_value2                                   1.191e-01  7.384e-02   1.613  0.10908
lagged_value3                                   3.717e-01  6.645e-02   5.594 1.19e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 7: Stepwise Regression Results: multicolinearity and autocorrelation corrected

The overall regression performance was statistically significant, with an F-statistic of 354.9 and a p-value of less than 0.001. The residual standard error was 8,384, with an adjusted R-squared of 0.9765, indicating a strong model fit. We then retested the autocorrelation and multicollinearity issues. The DW test showed that the DW value of the model was 1.7772 (close to 2), with a p-value nearly 0.01, indicating significantly reduced autocorrelation issues as compared to before. Furthermore, the VIF values of the predictors were not excessively high, suggesting that multicollinearity was within acceptable levels.

4.1.3 Goodness of fit

Next, we tested the goodness of fit for the final stepwise regression model. To examine the constant variance assumption, we analyzed scatterplots of residuals versus fitted values (see Figure 8). The scatterplots indicated that the assumption of constant variance seems to be violated.

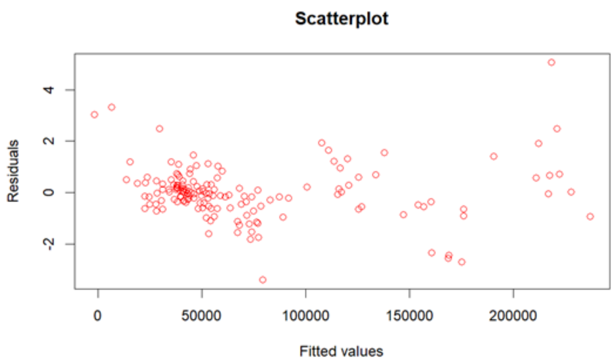


Figure 8: Scatterplot for Constant Variance Assumption

To test the linearity assumption, we examined scatterplots of residuals versus selected variables where the plots can be found in Appendix G. The results suggested that the linearity assumption was nearly violated, particularly in the plot of Residuals vs. Battery Cost. To test the normality assumption, we used both a histogram and a QQ-plot (see Figure 9). These visualizations suggested that the normality assumption is generally satisfied. However, some outliers were identified by observing Cook's distance (see Figure 10). Given the limited number of data points, we decided to retain these outliers in our analysis.

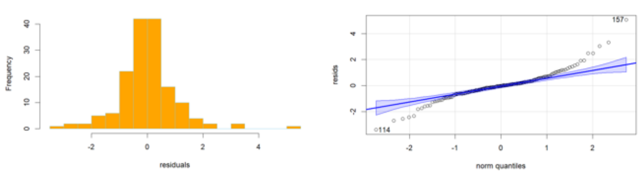


Figure 9: Plots for Normality Assumption

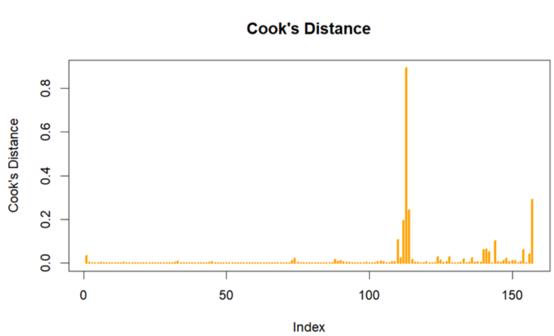


Figure 10: Cook's Distance

The goodness of fit results indicated that our final stepwise regression model does not seem to be a good fit to our dataset.

4.2 Second Approach: BoxCox MLR

Since we found that the linearity assumption was not met for the original model, we decided to apply the BoxCox transformation to address this issue and better fit our dataset. First, we calculated the log-likelihood values for a range of transformation parameters lambda using the Box-Cox function (see Figure 11), and identified the optimal lambda value, which was 0.42.

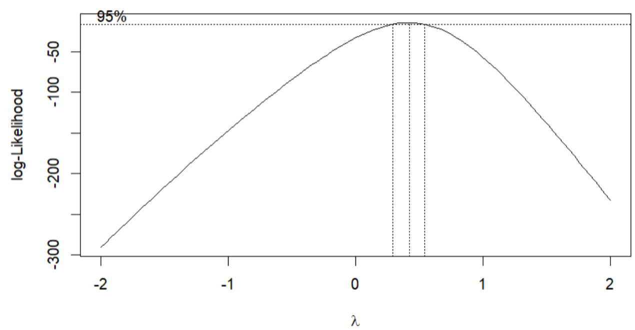


Figure 11: BoxCox Result

Then we transformed "Total.NEV.Sales" using the value of best lambda, that is, if lambda = 0, we use log transformation, otherwise we use (y**lambda - 1) / lambda. After applying the transformation, we re-estimated the BoxCox regression model and observed the results found in Appendix H. We found that the overall regression performance was statistically significant, with an F-statistic of 280.9 and a p-value of less than 0.001. The residual standard error was 10.26, and the adjusted R-squared value was 0.9805, indicating a strong model fit. However, the number of predictors was large, and several of them were not statistically significant.

4.2.1 Multicolinearity Correction

We then tested the multicollinearity issue by VIF test and the results are in the Appendix I. The VIF results revealed significant multicollinearity issues, as several variables exhibited VIF values exceeding acceptable thresholds. To address this, we excluded variables with VIF values greater than 62, based on established guidelines for multicollinearity diagnostics. And then we reran the BoxCox regression model, the results are in Appendix J. We found that overall regression performance was statistically significant, with an F-statistic of 217.5 and a p-value of less than 0.001. The residual standard error was 15.72, and the adjusted R-squared value was 0.9542, indicating a strong model fit. Also, most of the predictors were statistically significant. We noticed that "Date" was removed in this model because of the high VIF value. To examine whether "Date" is important for the regression model, we then added it to the model again and used anova to test its performance. The results showed that the F-statistic is 0.0373, with a p-value of larger than 0.1, indicating we should not keep Date in the BoxCox transformed model by force since it will not improve the performance of the model. For multicollinearity, the BoxCox model produced the following results from the VIF test:

```
                                            Month                          Covid.19
                                         1.413901                          5.954011
                      Number.of.new.energy.policies                        Production
                                         4.727917                          5.825524
                                        Inventory                 Gasoline.LDV.sales
                                        33.351361                          2.802633
                      Federal.fund.effective.rate                 Umemployment.rate
                                         4.345362                          3.322561
                                            PPIBC                               PCE
                                        10.165497                         26.632445
                              Tesla.model.S.price                       BatteryCost
                                        10.094932                         21.775594
Percentage.of.employees.who.are.middle.aged..25.55.                    Gasoline.price
                                        17.370260                          3.849206
                             Electric.retail.price
                                         7.153222
```

Figure 12: VIF Values of Variables

4.2.2 Autocorrelation Correction

As for the autocorrelation, the Durbin-Watson test showed the DW value of the model was 1.0071, with a p-value that is nearly 0. This means we reject the H0, indicating true autocorrelation is greater than 0. To fix this issue, we decided to add lagged variables into the model. We first used the ACF test to determine which lagged variables should be added(Figure 13).

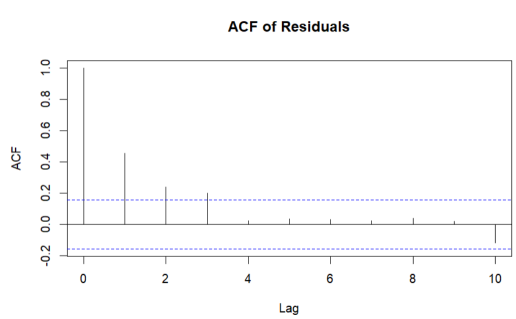


Figure 13: ACF of Residuals

From the ACF of Residuals, we found that the residual autocorrelation at lag orders 1, 2 and 3 is significant, so we created these lagged variables and added them into the regression model. The final BoxCox regression model had the following results:

```
Residuals:
    Min      1Q  Median      3Q     Max 
-28.368  -6.616   0.174   6.915  33.049 

Coefficients:
                                                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                        -3.651e+01  2.255e+02  -0.162 0.871612    
Month                                               5.022e-01  3.306e-01   1.519 0.131028    
Covid.19                                            7.873e+00  5.324e+00   1.479 0.141530    
Number.of.new.energy.policies                      -5.218e-02  3.433e-02  -1.520 0.130872    
Production                                          2.938e-05  2.433e-05   1.208 0.229268    
Inventory                                           1.436e-05  1.375e-05   1.044 0.298196    
Gasoline.LDV.sales                                  7.343e-05  7.841e-06   9.365 2.29e-16 ***
Federal.fund.effective.rate                         8.083e+02  1.389e+02   5.821 4.05e-08 ***
Umemployment.rate                                   5.483e+02  8.258e+01   6.640 7.03e-10 ***
PPIBC                                               1.285e-01  6.419e-01   0.200 0.841607    
PCE                                                 1.038e-01  1.637e-02   6.342 3.17e-09 ***
Tesla.model.S.price                                 2.477e-03  1.026e-03   2.414 0.017113 *  
BatteryCost                                         1.405e-02  2.869e-02   0.490 0.625175    
Percentage.of.employees.who.are.middle.aged..25.55. -2.839e+02  3.260e+02  -0.871 0.385327    
Gasoline.price                                      1.093e+01  3.005e+00   3.639 0.000389 ***
Electric.retail.price                              -8.083e+00  2.646e+00  -3.055 0.002710 ** 
lagged_value1                                       3.494e-01  5.892e-02   5.929 2.41e-08 ***
lagged_value2                                       7.277e-02  5.991e-02   1.215 0.226641    
lagged_value3                                       2.369e-01  5.316e-02   4.457 1.73e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 14: BoxCox Regression:Final Iteration

The final regression performance was statistically significant, with an F-statistic of 357.4 and a p-value of less than 0.001. The residual standard error was 11.15, and the adjusted R-squared value was 0.9767, indicating a strong model fit. Also, most of the predictors were statistically significant. We then retested the autocorrelation and multicollinearity issues. The DW test showed that the DW value of the model was 1.8235 (close to 2), with a p-value greater than 0.01, indicating significantly reduced autocorrelation issues when compared to before. Furthermore, the VIF values of the predictors were not excessively high, suggesting that multicollinearity was within acceptable levels.

4.2.3 Goodness of fit

Next, we tested the goodness of fit for the final BoxCox regression model. To examine the constant variance assumption, we analyzed scatterplots of residuals versus fitted values (see Figure 15). The scatterplots indicated that the assumption of constant variance appeared to hold.

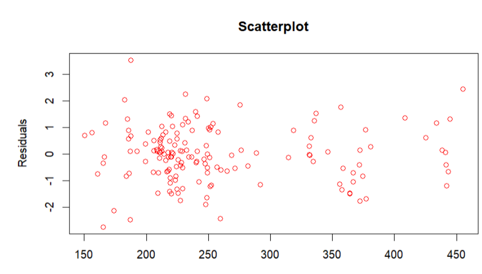


Figure 15: Scatterplot for Constant Variance Assumption

To test the linearity assumption, we examined scatterplots of residuals versus selected variables (see Figure 31). The results suggested that the linearity assumption almost held (see Appendix K). To test the normality assumption, we used both a histogram and a QQ-plot (see Figure 16). These visualizations suggested that the normality assumption is generally satisfied. However, some outliers were identified by observing Cook's distance (see Figure 17). Given the limited number of data points, we decided to retain these outliers in our analysis.

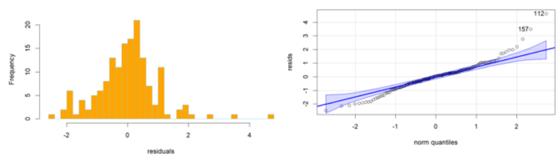


Figure 16: Plots for Normality Assumption

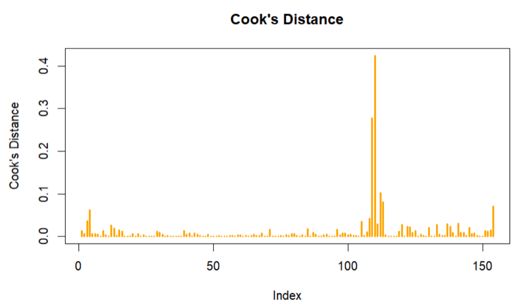


Figure 17: Cook's Distance

Results of goodness of fit indicated that our final BoxCox fitted the dataset well.

4.3 Third Approach: Ridge MLR

Considering the dataset has a high multicollinearity issue, we also tried Ridge regression to fix that problem. The Ridge process won't remove predictors. We fitted the ridge regression model using alpha of 0, visualizing the coefficients of the ridge model as a function of the regularization parameter(see Figure 18), and then using cross-validation to find the best lambda based on model performance (minimizing prediction error). The model performance with different values of lambda were shown in Figure 18.

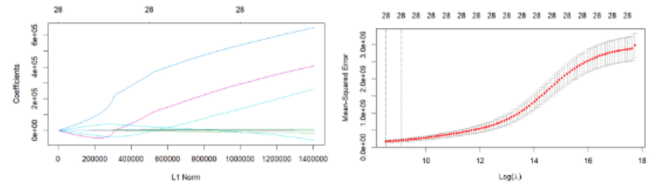


Figure 18: Ridge Regression Results

We finally obtained the best lambda, which was 5003.271, and applied it to the final ridge model. The coefficients of the predictors in this model are:

```
28 x 1 sparse Matrix of class "dgCMatrix"
                                                                               s0
Month                                                                5.622378e+02
Date                                                                -8.760754e-02
Covid.19                                                            -1.707478e+04
Number.of.new.energy.policies                                       -1.088210e+02
Production                                                           4.564174e-02
Inventory                                                            4.621963e-03
Gasoline.LDV.sales                                                   2.751609e-02
M1SL                                                                 1.263975e+00
M2SL                                                                 5.417634e-01
Federal.fund.effective.rate                                          6.405528e+05
Umemployment.rate                                                    2.701036e+05
PPIBC                                                               -1.103099e+02
PPIBM                                                                2.166538e+02
PCE                                                                  2.478844e+01
GDP                                                                  1.960865e+00
Durable.goods.consumption                                            2.536275e+01
Nondurable.goods.consumption                                         1.631283e+01
Nonresidential.fixed.Investment                                      1.062938e+01
Residential.fixed.Investment                                        -3.438602e+00
Tesla.model.S.price                                                  2.411760e+00
ElectricPorts                                                        2.122293e-01
BatteryCost                                                         -2.308608e+00
Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education -5.744401e+04
Percentage.of.employees.who.are.middle.aged..25.55.                  4.043770e+05
Population                                                          -6.745629e-01
Per.capita.disposable.income                                         9.417733e-01
Gasoline.price                                                       7.844055e+03
Electric.retail.price                                               -1.098814e+01
```

Figure 19: Coefficients of Predictors in Ridge Regression Model

4.4 Last Approach: Poisson Regression

Since the constant variance assumption is violated, we also tried using Poisson regression to address the issue. From the results (see Figure 20), all the predictors were significant. The overall regression test showed a p-value of 0, indicating that at least one predictor had a non-zero coefficient. However, the goodness of fit test showed a p-value of 0, suggesting that Poisson regression does not fit our dataset well.

```
Coefficients:
                                                                       Estimate Std. Error  z value Pr(>|z|)    
(Intercept)                                                           1.826e+01  4.784e-01   38.160  < 2e-16 ***
Month                                                                 1.461e-02  1.210e-04  120.767  < 2e-16 ***
Date                                                                  2.785e-03  3.674e-05   75.795  < 2e-16 ***
Covid.19                                                             -1.597e-02  2.282e-03   -6.997 2.61e-12 ***
Number.of.new.energy.policies                                        -9.604e-05  1.145e-05   -8.391  < 2e-16 ***
Production                                                           -9.801e-08  1.109e-08   -8.835  < 2e-16 ***
Inventory                                                             4.324e-07  7.678e-09   56.313  < 2e-16 ***
Gasoline.LDV.sales                                                    5.841e-07  3.133e-09  186.473  < 2e-16 ***
M1SL                                                                 -3.417e-05  8.053e-07  -42.430  < 2e-16 ***
M2SL                                                                  2.034e-05  3.996e-06    5.091 3.56e-07 ***
Federal.fund.effective.rate                                           8.011e+00  1.381e-01   57.993  < 2e-16 ***
Umemployment.rate                                                    -3.328e+00  6.389e-02  -52.087  < 2e-16 ***
PPIBC                                                                 7.212e-03  3.134e-04   23.015  < 2e-16 ***
PPIBM                                                                -7.119e-04  2.873e-04   -2.478   0.0132 *  
PCE                                                                   4.180e-04  1.053e-05   39.698  < 2e-16 ***
GDP                                                                  -1.826e-04  5.858e-06  -31.173  < 2e-16 ***
Durable.goods.consumption                                             1.610e-03  2.199e-05   73.204  < 2e-16 ***
Nondurable.goods.consumption                                         -7.930e-04  3.298e-05  -24.047  < 2e-16 ***
Nonresidential.fixed.Investment                                      -6.890e-04  1.925e-05  -35.801  < 2e-16 ***
Residential.fixed.Investment                                         -1.671e-03  2.722e-05  -61.394  < 2e-16 ***
Tesla.model.S.price                                                   3.786e-05  7.865e-07   48.144  < 2e-16 ***
ElectricPorts                                                         8.340e-06  2.163e-07   38.562  < 2e-16 ***
BatteryCost                                                          -1.642e-03  1.460e-05 -112.495  < 2e-16 ***
Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education -4.855e+00  9.943e-02  -48.827  < 2e-16 ***
Percentage.of.employees.who.are.middle.aged..25.55.                  -4.713e+00  1.291e-01  -36.509  < 2e-16 ***
Population                                                           -3.633e-04  3.918e-06  -92.727  < 2e-16 ***
Per.capita.disposable.income                                         -2.683e-06  6.154e-07   -4.359 1.31e-05 ***
Gasoline.price                                                        1.920e-01  1.955e-03   98.197  < 2e-16 ***
Electric.retail.price                                                -2.330e-02  1.030e-03  -22.625  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 5123903  on 156  degrees of freedom
Residual deviance:   72056  on 128  degrees of freedom
AIC: 74128

Number of Fisher Scoring iterations: 4
```

Figure 20: Poisson Regression

4.5 Model Comparison

Here, we adopt ***10-fold cross-validation*** to compare the performance of the five models in terms of prediction accuracy. The models include:

- Model 1: Stepwise multicollinearity corrected model
- Model 2: Stepwise multicollinearity & autocorrelation corrected model
- Model 3: Box-Cox multicollinearity corrected model
- Model 4: Box-Cox multicollinearity & autocorrelation corrected model
- Model 5: Ridge regression model

The prediction accuracy measures are as follows: MSPE, MAE, MAPE, PM, R-squared, Mean residual and Residual Sum of Squares. The comparison table is shown in Table 1.

The comparison table emphasizes that:

(1) Once the autocorrelation issue has been resolved, either stepwise or Box-Cox models' predictive power has been strengthened. For example, compared to Model 1, Model 2 shows lower prediction errors (MSPE, MAE, MAPE, PM, residual sum of squares) and higher R-squared (97. 66% vs. 92. 01%). This also applies to Model 4 compared to Model 3.

(2) Whether or not the autocorrelation is corrected, Box-Cox transformed multiple linear regression performs significantly better than stepwise multiple linear regression, with lower prediction errors (MSPE, MAE, MAPE, PM, Residual Sum of Squares) and higher R-squared. This can be derived from the comparison of Model 3 vs Model 1, and Model 4 vs. Model 2.

(3) Without any treatment (i.e., no multicollinearity or autocorrelation correction), Model 5 (ridge multiple linear regression) does not perform poorly in prediction. It achieves lower prediction errors (MSPE, MAE, MAPE, PM, Residual Sum of Squares) and higher R-squared compared to Model 1. This indicates that, under the constraint of no careful treatment, using ridge regression with all features is not the worst option because ridge itself can handle multicollinearity and reduce the severity of autocorrelation.

4.6 Variables Selected

We compared the statistically significant variables and their coefficients at $\alpha = 0.05$ from Model 2 (Stepwise multicollinearity and autocorrelation corrected model) and Model 4 (Box-Cox multicollinearity and autocorrelation corrected model). Both models successfully address multicollinearity and autocorrelation problems and are considered 'best' for identifying the predictors for the response variable.

The significant variables identified are as follows:

Significant predictors	**Coefficient**	
	Model 2	**Model 4**
Gasoline.LDV.sales	4.820e-02	7.343e-05
Federal.fund.effective.rate	7.529e+05	8.083e+02
Unemployment.rate	1.444e+05	5.483e+02
Gasoline.price	4.228e+03	1.093e+01
Electric.retail.price	-6.482e+03	-8.083e+00
lagged value1	3.965e-01	3.494e-01
lagged value3	3.786e-01	2.369e-01
PCE	/	1.038e-01
Tesla.model.S.price	/	2.477e-03
M1SL	2.272e+00	/

Table 2: Significant variables selected

We can observe that:

Model	MSPE	MAE	MAPE	PM	R-squared	Mean Residual	Residual Sum of Squares
Model 1	1.529660e+08	8.586711e+03	1.491908e+01	7.987972e-02	9.201203e-01	-1.477414e+01	2.399444e+09
Model 2	5.281100e+07	5.343426e+03	8.998258e+00	2.339006e-02	9.766099e-01	-1.935032e+01	8.143157e+08
Model 3	2.218659e+02	1.091726e+01	4.475114e+00	4.908420e-02	9.509158e-01	5.622677e-02	3.484675e+03
Model 4	1.082560e+02	8.089410e+00	3.355103e+00	2.205188e-02	9.779481e-01	4.811473e-02	1.678848e+03
Model 5	1.281676e+08	7.844014e+03	1.313336e+01	4.450664e-02	9.554934e-01	-1.444706e+01	2.021927e+09

Table 1: Comparison of 5 models by 7 prediction accuracy measures

(1) Fewer than 10 variables are statistically significant in both models.

(2) The two models identify 7 common variables, and their coefficients have the same sign. These variables are:

   (a) **Macroeconomic Variables (Market Demand)**: The Federal Fund Effective Rate and Unemployment Rate. Both models show positive coefficients, which is counterintuitive. Typically, higher interest rates increase borrowing costs, reducing consumer spending. However, one possible explanation is that a rise in the Federal Reserve's interest rates may signal a healthy economy, encouraging larger purchases. In addition, government stimulus programs for car manufacturers and customers may have boosted car sales during periods of high unemployment, with incentives such as discounts, low-interest loans, and special offers.

   (b) **Industry Variables (Cost of Owning a NEV)**: The Electric Retail Price and Gasoline Price. The Electric Retail Price is negatively associated with the response, while the Gasoline Price is positively associated. This is reasonable because higher electricity prices increase the operational cost of running a NEV, whereas higher gasoline prices encourage customers to purchase NEVs as a more economical alternative.

   (c) **Sector-Specific Variables**: Gasoline LDV Sales, 1-Month Lagged Sales, and 3-Month Lagged Sales. All three variables are positively associated with the response, which is justifiable. Gasoline LDV Sales represent the broader car market, as gasoline vehicles remain dominant. Stronger LDV sales suggest a higher demand for cars. Furthermore, 1-month and 3-month lagged NEV sales reflect strong autocorrelation in the NEV market.

(3) Model 4 identified two additional variables: The PCE Index (Personal Consumption Expenditures Price Index) and the Tesla Model S Price. Both variables are positively associated with NEV sales, which is counterintuitive and warrants further investigation. The result may be specific to the dataset's limitations, as the NEV market is relatively new, having emerged around 2010.

(4) Model 2 identifies another macroeconomic variable: M1 money supply. M1 represents the most liquid forms of money in an economy and is positively associated with NEV sales, indicating that when consumers have more disposable income, NEV sales tend to increase.

(5) As mentioned previously, after correcting for autocorrelation, no seasonality or time trend is detected in both models, even the effect of covid-19 is not significant either.

5 Conclusion and Recommendations

In summary, we began with an exploratory data analysis, including visualizations such as boxplots, correlation plots showing the relationship between the response and each predictor, trend and seasonality analysis, and outlier detection. Given the large number of variables, we employed a stepwise model (both directions) for feature selection, followed by corrections for multicollinearity and autocorrelation. Due to violations of the assumptions of constant variance of residuals and linearity (particularly with respect to residuals vs. battery cost), we applied a Box-Cox transformation, which also corrects multicollinearity and autocorrelation. In addition, we experimented with Ridge regression, which is effective in handling multicollinearity and autocorrelation by shrinking the coefficients of predictors, and Poisson regression, due to the violation of the constant variance assumption. However, Poisson is not found to be a good fit for the data.

Finally, we compared the models based on prediction accuracy and analyzed the variables selected. Through this regression analysis, we found that autocorrelation has a significant impact on the model, and the Box-Cox transformation is crucial for improving prediction accuracy. Ridge regression, while not offering the best results, is still a reasonable choice when no thorough corrections are applied.

We identified nine variables that are statistically significant, providing insightful explanations. However, because of the short history of the NEV market and the limited data size, the reliability of the predicting power of these variables may be questionable. It is advisable to revisit this regression analysis after the market has developed further, ideally after another ten years of data accumulation.

References

[1] Xiaoli Sun, Zhengguo Li, Xiaolin Wang, and Chengjiang Li. Technology development of electric vehicles: A review. *Energies*, 13(1):90, 2019.

[2] Fayez Alanazi. Electric vehicles: benefits, challenges, and potential solutions for widespread adaptation. *Applied Sciences*, 13(10):6016, 2023.

[3] Eckard Helmers and Patrick Marx. Electric cars: technical characteristics and environmental impacts. *Environmental Sciences Europe*, 24:1–15, 2012.

[4] Shawn Ou, Zhenhong Lin, Zhixin Wu, Jihu Zheng, Renzhi Lyu, Steven V Przesmitzki, and Xin He. A study of china s explosive growth in the plug-in electric vehicle market. Technical report, Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States). National ..., 2017.

[5] Nic Lutsey, Peter Slowik, and Lingzhi Jin. Sustaining electric vehicle market growth in us cities. *International Council on Clean Transportation*, 2016.

[6] Ldv total sales of pev and hev by month. `https://www.anl.gov/sites/www/files/2023-12/Total%20Sales%20for%20Website_November%202023.pdf`.

[7] Fred economic data: Series by categories. `https://fred.stlouisfed.org/categories`.

[8] Deepak Jaiswal, Arun Kumar Deshmukh, and Park Thaichon. Who will adopt electric vehicles? segmenting and exemplifying potential buyer heterogeneity and forthcoming research. *Journal of Retailing and Consumer Services*, 67:102969, 2022.

[9] Alternative fuels data center:federal and state laws and incentives. `https://afdc.energy.gov/laws`.

[10] Alternative fuels data center:alternative fueling station counts by state. `https://afdc.energy.gov/stations/states`.

A Response vs. Predictors Figures

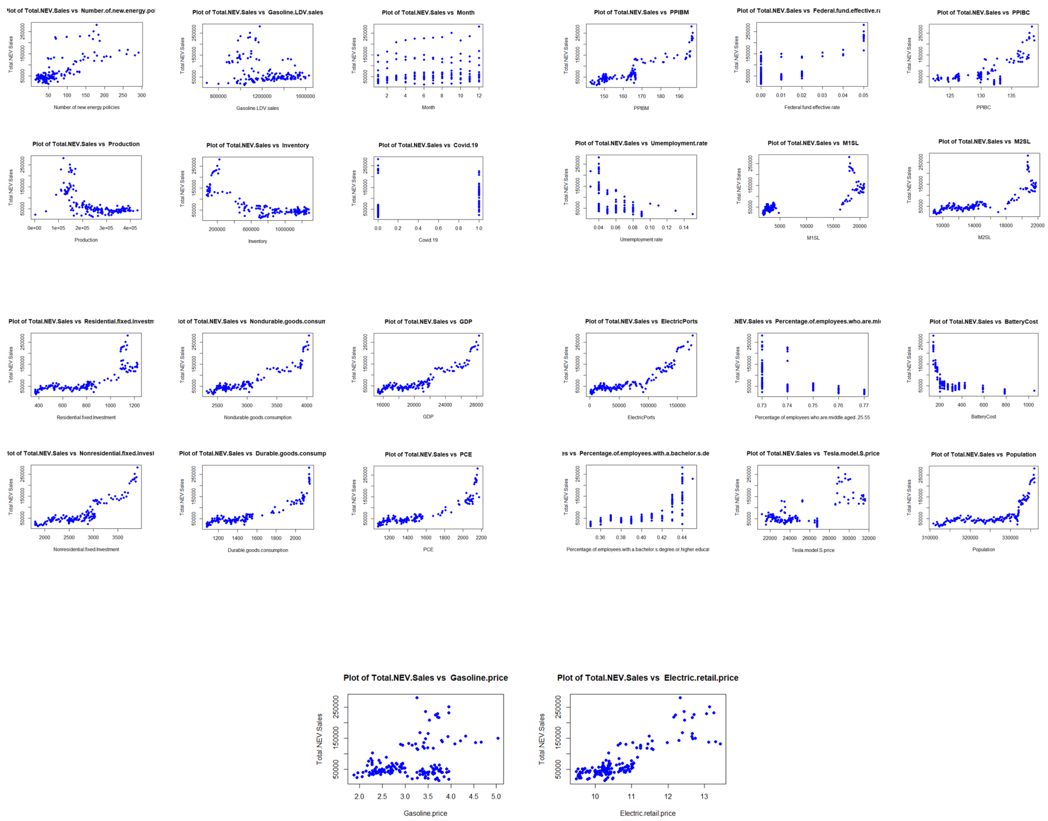


Figure 21: Response vs. Predictors Figures

B Correlation Matrix For Predictors

	Covid.19	Number.of	Productior	Inventory	Gasoline.L	M1SL	M2SL	Federal.fu	Umemploy	PPIBC	PPIBM	PCE	GDP
Covid.19	1.000	0.586	-0.676	-0.766	-0.443	0.839	0.794	0.116	-0.055	0.563	0.617	0.720	0.632
Number.of	0.586	1.000	-0.634	-0.781	-0.410	0.761	0.789	0.587	-0.415	0.682	0.853	0.795	0.818
Productior	-0.676	-0.634	1.000	0.857	0.428	-0.737	-0.789	-0.462	0.194	-0.695	-0.729	-0.744	-0.747
Inventory	-0.766	-0.781	0.857	1.000	0.559	-0.877	-0.858	-0.515	0.214	-0.890	-0.853	-0.852	-0.822
Gasoline.L	-0.443	-0.410	0.428	0.559	1.000	-0.436	-0.314	-0.129	-0.268	-0.603	-0.379	-0.275	-0.268
M1SL	0.839	0.761	-0.737	-0.877	-0.436	1.000	0.942	0.400	-0.247	0.735	0.844	0.935	0.853
M2SL	0.794	0.789	-0.789	-0.858	-0.314	0.942	1.000	0.530	-0.462	0.654	0.907	0.979	0.956
Federal.fu	0.116	0.587	-0.462	-0.515	-0.129	0.400	0.530	1.000	-0.548	0.516	0.785	0.602	0.734
Umemploy	-0.055	-0.415	0.194	0.214	-0.268	-0.247	-0.462	-0.548	1.000	-0.093	-0.492	-0.533	-0.599
PPIBC	0.563	0.682	-0.695	-0.890	-0.603	0.735	0.654	0.516	-0.093	1.000	0.763	0.681	0.661
PPIBM	0.617	0.853	-0.729	-0.853	-0.379	0.844	0.907	0.785	-0.492	0.763	1.000	0.926	0.967
PCE	0.720	0.795	-0.744	-0.852	-0.275	0.935	0.979	0.602	-0.533	0.681	0.926	1.000	0.966
GDP	0.632	0.818	-0.747	-0.822	-0.268	0.853	0.956	0.734	-0.599	0.661	0.967	0.966	1.000
Durable.gc	0.716	0.804	-0.757	-0.857	-0.299	0.926	0.982	0.617	-0.538	0.689	0.937	0.993	0.978
Nondurabl	0.685	0.843	-0.750	-0.869	-0.357	0.906	0.962	0.689	-0.522	0.735	0.977	0.975	0.986
Nonreside	0.561	0.781	-0.689	-0.741	-0.179	0.785	0.921	0.771	-0.647	0.581	0.950	0.932	0.988
Residentia	0.720	0.751	-0.758	-0.789	-0.203	0.871	0.980	0.536	-0.593	0.563	0.876	0.962	0.956
Tesla.mod	0.389	0.618	-0.341	-0.630	-0.597	0.608	0.452	0.319	-0.088	0.748	0.582	0.512	0.470
ElectricPo	0.713	0.801	-0.791	-0.864	-0.318	0.907	0.981	0.661	-0.487	0.686	0.951	0.975	0.985
BatteryCo	-0.496	-0.533	0.475	0.436	-0.092	-0.602	-0.798	-0.482	0.659	-0.186	-0.693	-0.764	-0.803
Percentag	0.708	0.703	-0.766	-0.760	-0.191	0.820	0.951	0.579	-0.507	0.517	0.861	0.913	0.931
Percentag	-0.518	-0.545	0.566	0.501	-0.069	-0.592	-0.796	-0.429	0.606	-0.214	-0.663	-0.746	-0.781
Populatior	0.608	0.670	-0.716	-0.675	-0.064	0.736	0.914	0.604	-0.631	0.425	0.827	0.884	0.922
Per.capita	0.722	0.776	-0.809	-0.859	-0.309	0.893	0.972	0.664	-0.460	0.678	0.937	0.960	0.972
Gasoline.p	0.098	0.299	-0.028	-0.329	-0.414	0.253	0.087	0.166	0.085	0.548	0.297	0.160	0.126
Electric.re	0.505	0.775	-0.585	-0.696	-0.298	0.779	0.810	0.694	-0.481	0.629	0.893	0.842	0.868

	Durable.gc	Nondurabl	Nonreside	Residentia	Tesla.mod	ElectricPo	BatteryCo	Percentag	Percentag	Populatior	Per.capita	Gasoline.p	Electric.re
Covid.19	0.716	0.685	0.561	0.720	0.389	0.713	-0.496	0.708	-0.518	0.608	0.722	0.098	0.505
Number.of	0.804	0.843	0.781	0.751	0.618	0.801	-0.533	0.703	-0.545	0.670	0.776	0.299	0.775
Productior	-0.757	-0.750	-0.689	-0.758	-0.341	-0.791	0.475	-0.766	0.566	-0.716	-0.809	-0.028	-0.585
Inventory	-0.857	-0.869	-0.741	-0.789	-0.630	-0.864	0.436	-0.760	0.501	-0.675	-0.859	-0.329	-0.696
Gasoline.L	-0.299	-0.357	-0.179	-0.203	-0.597	-0.318	-0.092	-0.191	-0.069	-0.064	-0.309	-0.414	-0.298
M1SL	0.926	0.906	0.785	0.871	0.608	0.907	-0.602	0.820	-0.592	0.736	0.893	0.253	0.779
M2SL	0.982	0.962	0.921	0.980	0.452	0.981	-0.798	0.951	-0.796	0.914	0.972	0.087	0.810
Federal.fu	0.617	0.689	0.771	0.536	0.319	0.661	-0.482	0.579	-0.429	0.604	0.664	0.166	0.694
Umemploy	-0.538	-0.522	-0.647	-0.593	-0.088	-0.487	0.659	-0.507	0.606	-0.631	-0.460	0.085	-0.481
PPIBC	0.689	0.735	0.581	0.563	0.748	0.686	-0.186	0.517	-0.214	0.425	0.678	0.548	0.629
PPIBM	0.937	0.977	0.950	0.876	0.582	0.951	-0.693	0.861	-0.663	0.827	0.937	0.297	0.893
PCE	0.993	0.975	0.932	0.962	0.512	0.975	-0.764	0.913	-0.746	0.884	0.960	0.160	0.842
GDP	0.978	0.986	0.988	0.956	0.470	0.985	-0.803	0.931	-0.781	0.922	0.972	0.126	0.868
Durable.gc	1.000	0.986	0.944	0.969	0.525	0.982	-0.768	0.917	-0.748	0.889	0.967	0.168	0.851
Nondurabl	0.986	1.000	0.956	0.940	0.585	0.982	-0.737	0.896	-0.707	0.864	0.963	0.253	0.888
Nonreside	0.944	0.956	1.000	0.936	0.377	0.960	-0.853	0.930	-0.817	0.939	0.952	0.060	0.856
Residentia	0.969	0.940	0.936	1.000	0.359	0.962	-0.861	0.959	-0.862	0.955	0.952	0.001	0.782
Tesla.mod	0.525	0.585	0.377	0.359	1.000	0.458	0.017	0.213	0.068	0.126	0.403	0.765	0.607
ElectricPo	0.982	0.982	0.960	0.962	0.458	1.000	-0.786	0.950	-0.781	0.917	0.992	0.109	0.841
BatteryCo	-0.768	-0.737	-0.853	-0.861	0.017	-0.786	1.000	-0.884	0.935	-0.928	-0.785	0.247	-0.617
Percentag	0.917	0.896	0.930	0.959	0.213	0.950	-0.884	1.000	-0.895	0.975	0.958	-0.130	0.738
Percentag	-0.748	-0.707	-0.817	-0.862	0.068	-0.781	0.935	-0.895	1.000	-0.935	-0.784	0.319	-0.548
Populatior	0.889	0.864	0.939	0.955	0.126	0.917	-0.928	0.975	-0.935	1.000	0.923	-0.196	0.715
Per.capita	0.967	0.963	0.952	0.952	0.403	0.992	-0.785	0.958	-0.784	0.923	1.000	0.056	0.816
Gasoline.p	0.168	0.253	0.060	0.001	0.765	0.109	0.247	-0.130	0.319	-0.196	0.056	1.000	0.348
Electric.re	0.851	0.888	0.856	0.782	0.607	0.841	-0.617	0.738	-0.548	0.715	0.816	0.348	1.000

Figure 22: Correlation Matrix for Predictors

C Initial Stepwise Regression Results

```
Call:
lm(formula = data$Total.NEV.Sales ~ Month + Date + Covid.19 +
    Number.of.new.energy.policies + Inventory + Gasoline.LDV.sales +
    M2SL + Umemployment.rate + PCE + Durable.goods.consumption +
    Nondurable.goods.consumption + Nonresidential.fixed.Investment +
    Residential.fixed.Investment + Tesla.model.S.price + ElectricPorts +
    BatteryCost + Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education +
    Population, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-23665.1  -3936.7    146.7   3616.7  22207.9

Coefficients:
                                                                       Estimate Std. Error t value Pr(>|t|)
(Intercept)                                                          -2.260e+05  6.211e+05  -0.364 0.716547
Month                                                                 7.202e+02  2.193e+02   3.283 0.001300 **
Date                                                                  1.657e+02  5.764e+01   2.874 0.004689 **
Covid.19                                                             -1.253e+04  4.596e+03  -2.727 0.007226 **
Number.of.new.energy.policies                                        -1.265e+02  2.333e+01  -5.423 2.55e-07 ***
Inventory                                                             3.758e-02  8.642e-03   4.349 2.64e-05 ***
Gasoline.LDV.sales                                                    3.695e-02  5.178e-03   7.135 4.98e-11 ***
M2SL                                                                 -1.769e+01  4.651e+00  -3.803 0.000214 ***
Umemployment.rate                                                     1.598e+05  1.024e+05   1.560 0.121025
PCE                                                                   4.052e+01  1.881e+01   2.154 0.032941 *
Durable.goods.consumption                                             9.803e+01  3.106e+01   3.156 0.001961 **
Nondurable.goods.consumption                                          6.942e+01  2.902e+01   2.393 0.018078 *
Nonresidential.fixed.Investment                                      -7.402e+01  2.564e+01  -2.886 0.004525 **
Residential.fixed.Investment                                         -1.212e+02  4.047e+01  -2.994 0.003263 **
Tesla.model.S.price                                                   2.816e+00  1.150e+00   2.449 0.015569 *
ElectricPorts                                                         9.869e-01  3.611e-01   2.733 0.007104 **
BatteryCost                                                          -6.182e+01  1.802e+01  -3.432 0.000792 ***
Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education -3.573e+05  1.670e+05  -2.140 0.034152 *
Population                                                           -2.065e+01  6.372e+00  -3.241 0.001495 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 7405 on 138 degrees of freedom
Multiple R-squared:  0.9838,    Adjusted R-squared:  0.9816
F-statistic: 464.3 on 18 and 138 DF,  p-value: < 2.2e-16
```

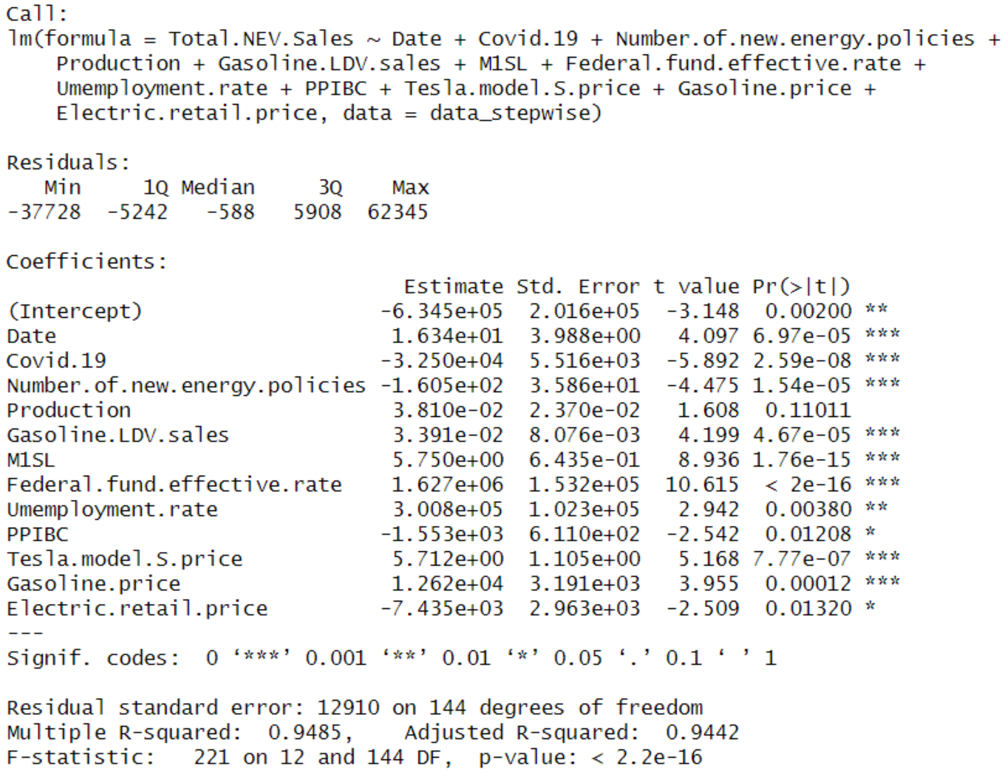
Figure 23: Initial Stepwise Regression Results

D VIF Values of Variables: Stepwise MLR model

Variable	VIF Value
Month	1.657330
Date	18101.439629
Covid.19	11.668300
Number.of.new.energy.policies	5.166106
Inventory	34.059699
Gasoline.LDV.sales	3.063568
M2SL	1049.702132
Umemployment.rate	12.613786
PCE	122.058974
Durable.goods.consumption	323.540754
Nondurable.goods.consumption	582.539717
Nonresidential.fixed.Investment	515.587131
Residential.fixed.Investment	309.945416
Tesla.model.S.price	31.408739
ElectricPorts	888.792974
BatteryCost	32.582692
Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education	70.251108
Population	6278.272065

Figure 24: VIF Values of Variables: Stepwise MLR model

E Stepwise Regression Results:Second Iteration



```
Call:
lm(formula = Total.NEV.Sales ~ Date + Covid.19 + Number.of.new.energy.policies +
    Production + Gasoline.LDV.sales + M1SL + Federal.fund.effective.rate +
    Umemployment.rate + PPIBC + Tesla.model.S.price + Gasoline.price +
    Electric.retail.price, data = data_stepwise)

Residuals:
   Min     1Q Median     3Q    Max
-37728  -5242   -588   5908  62345

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                  -6.345e+05  2.016e+05  -3.148  0.00200 **
Date                          1.634e+01  3.988e+00   4.097 6.97e-05 ***
Covid.19                     -3.250e+04  5.516e+03  -5.892 2.59e-08 ***
Number.of.new.energy.policies -1.605e+02  3.586e+01  -4.475 1.54e-05 ***
Production                    3.810e-02  2.370e-02   1.608  0.11011
Gasoline.LDV.sales            3.391e-02  8.076e-03   4.199 4.67e-05 ***
M1SL                          5.750e+00  6.435e-01   8.936 1.76e-15 ***
Federal.fund.effective.rate   1.627e+06  1.532e+05  10.615  < 2e-16 ***
Umemployment.rate             3.008e+05  1.023e+05   2.942  0.00380 **
PPIBC                        -1.553e+03  6.110e+02  -2.542  0.01208 *
Tesla.model.S.price           5.712e+00  1.105e+00   5.168 7.77e-07 ***
Gasoline.price                1.262e+04  3.191e+03   3.955  0.00012 ***
Electric.retail.price        -7.435e+03  2.963e+03  -2.509  0.01320 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12910 on 144 degrees of freedom
Multiple R-squared:  0.9485,    Adjusted R-squared:  0.9442
F-statistic:   221 on 12 and 144 DF,  p-value: < 2.2e-16
```

Figure 25: Stepwise Regression Results: Second Iteration

F VIF Values of Variables under Stepwise MLR model:Second Iteration

Variable	VIF Value
Date	28.509174
Covid.19	5.532017
Number.of.new.energy.policies	4.016289
Production	4.561375
Gasoline.LDV.sales	2.451947
M1SL	20.328002
Federal.fund.effective.rate	4.689438
Umemployment.rate	4.137900
PPIBM	6.983916
Tesla.model.S.price	9.551348
Gasoline.price	3.862100
Electric.retail.price	6.924861

Figure 26: VIF Values of Variables

G Linearity Assumption: Final Stepwise MLR Model

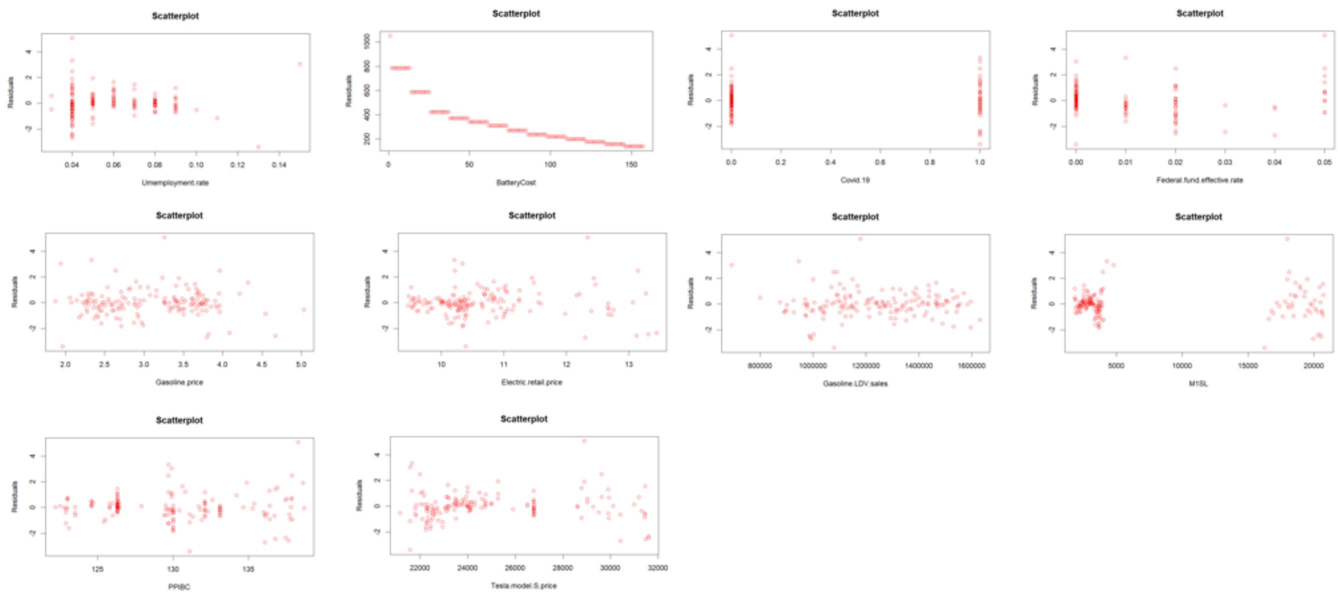


Figure 27: Scatterplots for Linearity Assumption

H BoxCox Regression Results:First Iteration

```
Residuals:
     Min       1Q   Median       3Q      Max 
-28.1193  -4.9063  -0.0071   5.9883  25.3090 

Coefficients:
                                                                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                                        1.298e+03  1.272e+03   1.020 0.309459    
Month                                                              1.531e+00  3.270e-01   4.682 7.14e-06 ***
Date                                                               2.772e-01  1.019e-01   2.720 0.007435 ** 
Covid.19                                                          -6.494e+00  7.125e+00  -0.911 0.363789    
Number.of.new.energy.policies                                     -6.342e-02  3.699e-02  -1.714 0.088880 .  
Production                                                        -9.753e-06  2.634e-05  -0.370 0.711829    
Inventory                                                          6.016e-05  1.812e-05   3.320 0.001172 ** 
Gasoline.LDV.sales                                                 6.175e-05  7.729e-06   7.990 6.82e-13 ***
M1SL                                                              -3.752e-03  1.907e-03  -1.967 0.051320 .  
M2SL                                                              -1.266e-03  1.045e-02  -0.121 0.903708    
Federal.fund.effective.rate                                        9.869e+02  3.762e+02   2.623 0.009761 ** 
Umemployment.rate                                                 -1.455e+02  1.545e+02  -0.942 0.348104    
PPIBC                                                              4.446e-01  7.196e-01   0.618 0.537778    
PPIBM                                                              1.397e-01  7.528e-01   0.186 0.853059    
PCE                                                                4.624e-02  3.277e-02   1.411 0.160723    
GDP                                                               -2.351e-02  1.650e-02  -1.424 0.156780    
Durable.goods.consumption                                          2.024e-01  6.194e-02   3.268 0.001391 ** 
Nondurable.goods.consumption                                      -6.756e-02  8.059e-02  -0.838 0.403425    
Nonresidential.fixed.Investment                                   -4.324e-02  4.954e-02  -0.873 0.384383    
Residential.fixed.Investment                                      -1.340e-01  8.153e-02  -1.644 0.102630    
Tesla.model.S.price                                                5.022e-03  2.269e-03   2.214 0.028631 *  
ElectricPorts                                                      1.536e-03  6.151e-04   2.497 0.013786 *  
BatteryCost                                                       -1.192e-01  3.249e-02  -3.670 0.000355 ***
Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education -4.502e+02  2.641e+02  -1.705 0.090668 .  
Percentage.of.employees.who.are.middle.aged..25.55.               -3.308e+02  3.347e+02  -0.988 0.324973    
Population                                                        -3.780e-02  1.073e-02  -3.522 0.000595 ***
Per.capita.disposable.income                                      -8.256e-04  1.634e-03  -0.505 0.614190    
Gasoline.price                                                     1.602e+01  5.386e+00   2.974 0.003514 ** 
Electric.retail.price                                             -4.115e+00  2.776e+00  -1.482 0.140758    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 10.26 on 128 degrees of freedom
Multiple R-squared:  0.984,	Adjusted R-squared:  0.9805 
F-statistic: 280.9 on 28 and 128 DF,  p-value: < 2.2e-16
```

Figure 28: BoxCox Regression Results

I VIF Values of Variables under BoxCox model:First Iteration

```
                                                            Month                                                Date
                                                         1.920368                                        29501.153805
                                                         Covid.19                         Number.of.new.energy.policies
                                                        14.618061                                            6.767884
                                                       Production                                           Inventory
                                                         8.927369                                           78.038416
                                               Gasoline.LDV.sales                                                M1SL
                                                         3.557467                                          282.814081
                                                             M2SL                        Federal.fund.effective.rate
                                                      2759.810876                                           44.756856
                                                Umemployment.rate                                               PPIBC
                                                        14.968166                                           15.339518
                                                            PPIBM                                                 PCE
                                                       202.623852                                          193.148413
                                                              GDP                          Durable.goods.consumption
                                                      5172.768085                                          670.782082
                                       Nondurable.goods.consumption                    Nonresidential.fixed.Investment
                                                      2342.557766                                         1002.806743
                                      Residential.fixed.Investment                                  Tesla.model.S.price
                                                       655.749898                                           63.749847
                                                    ElectricPorts                                         BatteryCost
                                                      1344.055530                                           55.226187
Percentage.of.employees.with.a.bachelor.s.degree.or.higher.education      Percentage.of.employees.who.are.middle.aged..25.55.
                                                        91.605819                                           24.533040
                                                       Population                         Per.capita.disposable.income
                                                      9287.203160                                          221.534172
                                                    Gasoline.price                               Electric.retail.price
                                                        17.426933                                            9.631857
```

Figure 29: VIF Values of Variables, BoxCox First Iteration

J BoxCox Regression Results:Second Iteration

```
Residuals:
    Min      1Q  Median      3Q     Max 
-33.628  -8.070   0.667   7.233  61.779 

Coefficients:
                                                     Estimate Std. Error t value Pr(>|t|)    
(Intercept)                                        -2.115e+02  3.036e+02  -0.697 0.487176    
Month                                               1.577e+00  4.301e-01   3.666 0.000348 ***
Covid.19                                           -7.225e+00  6.970e+00  -1.037 0.301642    
Number.of.new.energy.policies                      -1.184e-01  4.739e-02  -2.498 0.013639 *  
Production                                          4.312e-05  3.262e-05   1.322 0.188309    
Inventory                                           1.688e-05  1.815e-05   0.930 0.354127    
Gasoline.LDV.sales                                  4.868e-05  1.051e-05   4.630 8.22e-06 ***
Federal.fund.effective.rate                         1.424e+03  1.796e+02   7.925 6.25e-13 ***
Umemployment.rate                                   7.793e+02  1.116e+02   6.983 1.05e-10 ***
PPIBC                                              -5.618e-01  8.978e-01  -0.626 0.532499    
PCE                                                 2.052e-01  1.865e-02  11.003  < 2e-16 ***
Tesla.model.S.price                                 5.312e-03  1.384e-03   3.839 0.000186 ***
BatteryCost                                        -7.419e-02  3.127e-02  -2.373 0.018997 *  
Percentage.of.employees.who.are.middle.aged..25.55.  2.466e+01  4.317e+02   0.057 0.954524    
Gasoline.price                                      2.206e+01  3.880e+00   5.685 7.24e-08 ***
Electric.retail.price                              -9.746e+00  3.667e+00  -2.657 0.008783 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.72 on 141 degrees of freedom
Multiple R-squared:  0.9586,	Adjusted R-squared:  0.9542 
F-statistic: 217.5 on 15 and 141 DF,  p-value: < 2.2e-16
```

Figure 30: BoxCox Regression Results,Second Iteration

K Linearity Assumption: Final BoxCox Model

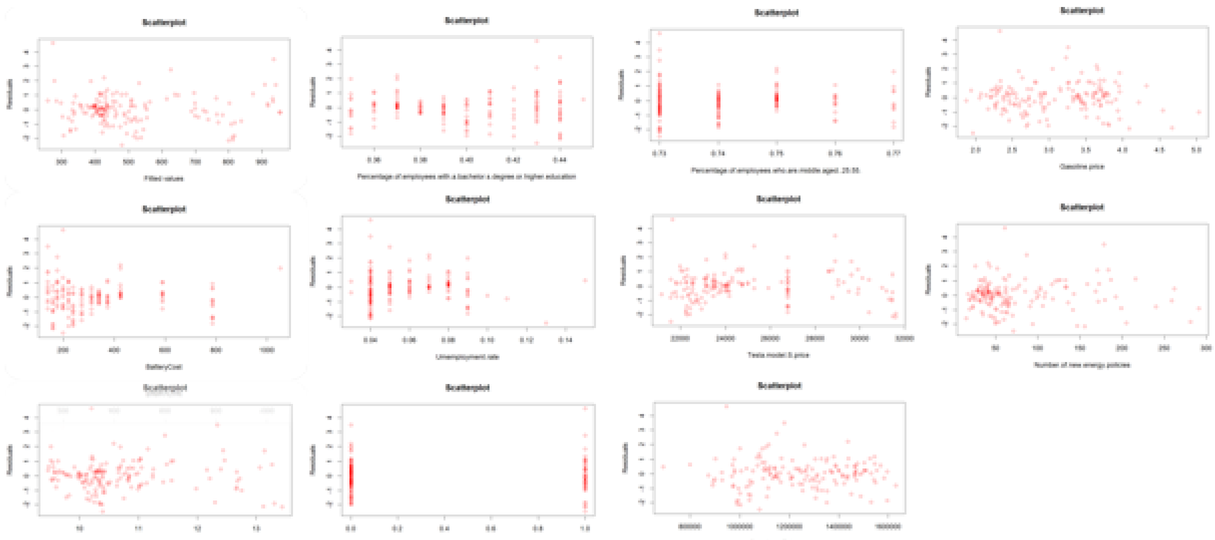


Figure 31: Scatterplots for Linearity Assumption