Sales of Electric Vehicles

Ankita Pal, Chavi Gupta, Liem Luong, Kirti Kharb

Abstract

EV – Electric Vehicle – is not a new acronym. Electric Vehicles first appeared in the mid-19th century. Until the beginning of the 21st century, it became popular and gained interest among drivers. With modern lithium-ion batteries capable of long-distance travel, reduced charging time, advanced power converters with zero emission, etc. These improvements helped lowering the cost of EV ownership versus gasoline cars. It took more than a century for the EV market to innovate itself, but it took only a decade for increasing the EV adoption among the consumers. A milestone of 1 million EVs on U.S. roads was celebrated in October 2018. A report by Edison Electric Institute (EEI) and the Institute for Electric Innovation (IEI) projected the number of EVs on U.S. roads to reach more than 18 million in 2030. Seeing an EV on the road has become as common nowadays and the EV market is becoming competitive among manufacturers. With this aim, our project attempts to ask, and subsequently answer, questions to determine the factors affecting the adoption of EV by car owners. We specifically assess the statistical experiment at the U.S level with a focus on Washington State. We discovered a significant positive trend between electricity price and the count of registered EVs in the state. We also discovered that Per Capita Personal Income plays the most essential role in the EV adoption at the US level and WA state. This report discusses the procedure, methods, findings, and shortcomings of research and analysis on the publicly available data.

Introduction

Electric Vehicles (EV) has been growing in the last decade. Understanding the influential factors contributing to the increasing EV adoptions is critical not only to the car manufacturers but also the government policy makers in supporting this trend. This is our aim of this study to determine the factors that affect the EVs adoption.

Furthermore, according to the US Department of Energy, WA state is the second in the nation in terms of EV registration by state as of January 2019 with almost 20,000 registered EVs. Given nature, our main goal of the analysis will focus on the comparing effect between the US level and the WA state.

There are potentially dozens of factors that could impact the EVs adoption. Although, performing a study with all the potential factors will yield a comprehensive picture of EV adoption, given the limitations of our current dataset, we have constrained our study and analysis to question the factors pertaining to the available dataset.

There are three main questions for our current study:

- Does the electricity price of a state affect the count of registered EVs in the state?
- Do economic indicators have an effect on this EV adoption at the US level?
- Do economic indicators have a similar effect on EV adoption in WA?

Our study to understand the cause-and-effect relationship in EV adoption follows 5 main parts:

- Question
- Analysis Plan
- Assumptions
- Process
- Answer

Data Set Description

Our study is conducted on the following independent datasets:

- EV registration count in the US and in the Washington State in 2017
- Electricity price of different states in US in 2017
- Economic indicators dataset for the US states and WA counties in 2017
- Population dataset for the US states and WA counties in 2018

Since our study assesses EV adoption at the US level with a focus on WA state. Therefore, we group our datasets into two separate group levels:

1. **US-Level**
    a. Electric Vehicle Registration Counts by State
        - Each row of the dataset represents one US state with its total approximate count of EV registration in 2017.
        - URL: https://afdc.energy.gov/data/10962
        - Source: IHS Markit light-duty vehicle registration in 2017

Data Fields	Description
US State	Name of the U.S. state
EV Registration	Number of EV registration by state

    b. Electricity Price by State
        - Each row of the dataset represents one US state with its average retail electricity price (cents/kWh) in 2017.
        - URL: https://www.eia.gov/electricity/state/
        - Source: U.S. Energy Information Administration

Data Field	Description
Name	Name of the U.S. state
Average retail price (cents/kWh)	The average electricity price by state

c. Economic Indicators by State
   - Each row of the dataset represents one US state with its economic indicators: unemployment rate, GDP, and per capita personal income
   - URL: https://fred.stlouisfed.org/categories/27281
   - Source: Federal Reserve Bank of St. Louis

Data Field	Description
Name	Name of the U.S. state
Unemployment Rate	Percent of unemployment rate by state
Gross Domestic Product	Annual GDP (millions of dollars) by state
Per Capita Personal Income	Annual PCPI (dollars) by state

d. Population by State
   - Each row of the dataset represents one US state with its population estimates in 2018
   - URL: https://www.census.gov/quickfacts/geo/chart/US/PST045218
   - Source: U.S. Census Bureau

Data Field	Description
Name	Name of the U.S. state
Population	Population estimates by state

2. County-Level

a. Electric Vehicle Population by WA
   - Each row of the dataset represents one unique EV Vehicle Identification Number (VIN) and its information (city, state, model year, electric range, sale date, etc.) within the Washington state
   - URL: https://data.wa.gov/Transportation/Electric-Vehicle-Title-and-Registration-Activity/rpr4-cgyd
   - Source: WA Department of Licensing

Data Field	Description
WA County	Name of the WA County
EV Registration	Number of EV registration by WA county

b. Economic Indicators by WA counties
   - Each row of the dataset represents one WA county with its economic indicators: unemployment rate, GDP, and per capita personal income.
   - URL: https://fred.stlouisfed.org/categories/30336
   - Source: Federal Reserve Bank of St. Louis

Data Field	Description
Name	Name of the WA county
Unemployment Rate	Percent of unemployment rate by WA county
Gross Domestic Product	Annual GDP (millions of dollars) by WA county

Per Capita Personal Income	Annual PCPI (dollars) by WA county

c. Population by WA counties
   - Each row of the dataset represents one WA county with its population estimates in 2018
   - URL: https://www.census.gov/quickfacts/fact/map/WA/PST045218
   - Source: U.S. Census Bureau

Data Field	**Description**
Name	Name of the WA county
Population	Population estimates by WA county

The following are some limitations and the implemented solutions associated with the given data set:

- **Less number of numeric fields:** From the County-Level group, our Electric Vehicle Population in WA dataset has only 5 numeric data fields out of 26. And most of these numeric fields are not significant enough in our study.

  Solution: To enrich the dataset, we transformed the EV population by WA dataset to keep the relevant information only (WA counties and their number of EV registered). Then we performed the data aggregation and transformation. We matched the economic indicators with the count and population dataset. This yielded a final dataset with many numeric data fields for our hypothesis 3.

  A similar method and solution applied to the US-Level group where we enriched the dataset with more numeric data fields for our hypothesis 2 and hypothesis 1

- **Non-standard data:** One of the challenges when dealing with multiple data sources is the scale. The numeric values of the registered EV counts are small compared to the large population counts. As the variables are measured at different scales, this will affect the analysis and also the interpretation of the coefficients later in the study.

  Solution: We apply the scaled transform method to normalize the number of EV registration by 100000 people per state. This method improved the accuracy of the linear regression model used for the study. Also, the interpretation of the coefficient from the model is meaningful.

  A similar method and solution applied to the County-Level group where we standardized the count of EV sales by WA county population.

We also perform data exploration to help us understand the datasets better in our study. Below are some visualizations that we performed during the data processing step.

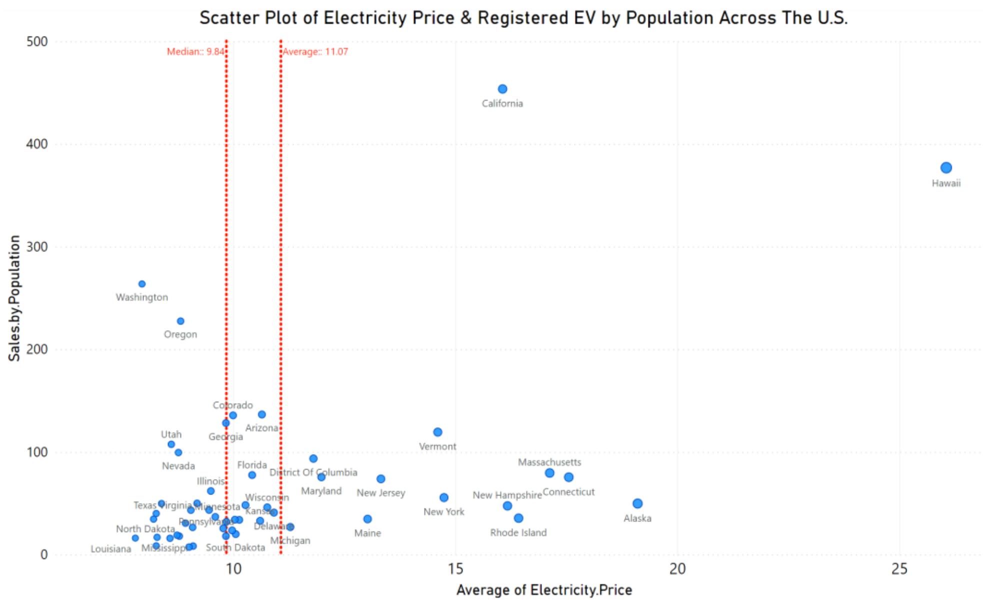

Figure 1: Relationship Between Electricity Price & Registered EV by Population at the US-Level

Observation:

- The cluster of electricity price vs. registered EV by population concentrates around the median electricity price of 9.84 cents/kWh
- Hawaii has the highest electricity price even though its population is small. This is considered an outlier for our study
- Notable point to highlight is that California, Washington, and Oregon are the three states with higher registered EV by population in the mainland

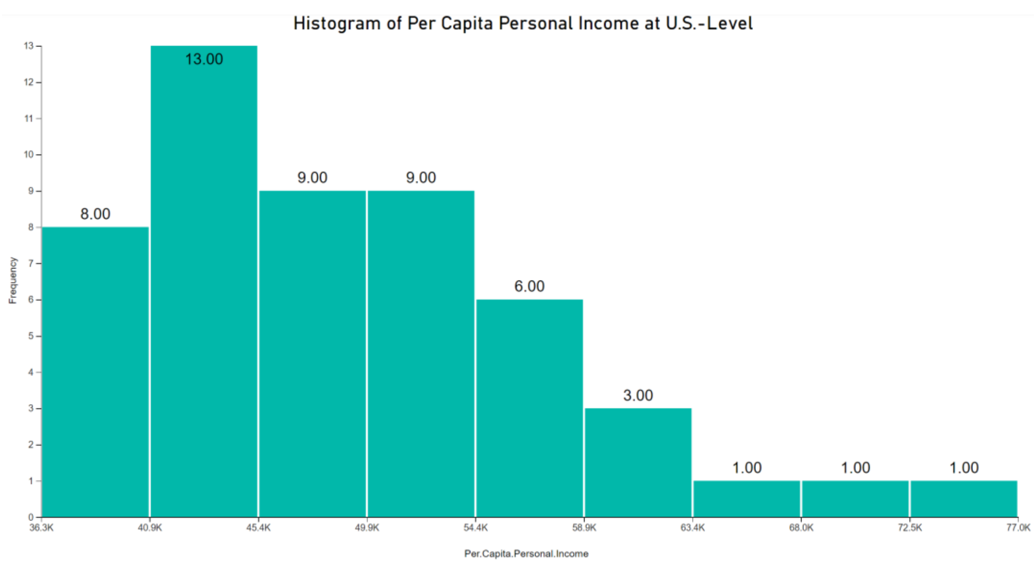

Figure 2: Distribution of Per Capita Personal Income at U.S. Level

Observation:

- Majority of the U.S. State (+ District of Columbia) has around ~ 55K per capita personal income
- The histogram is right skew and has thin tail

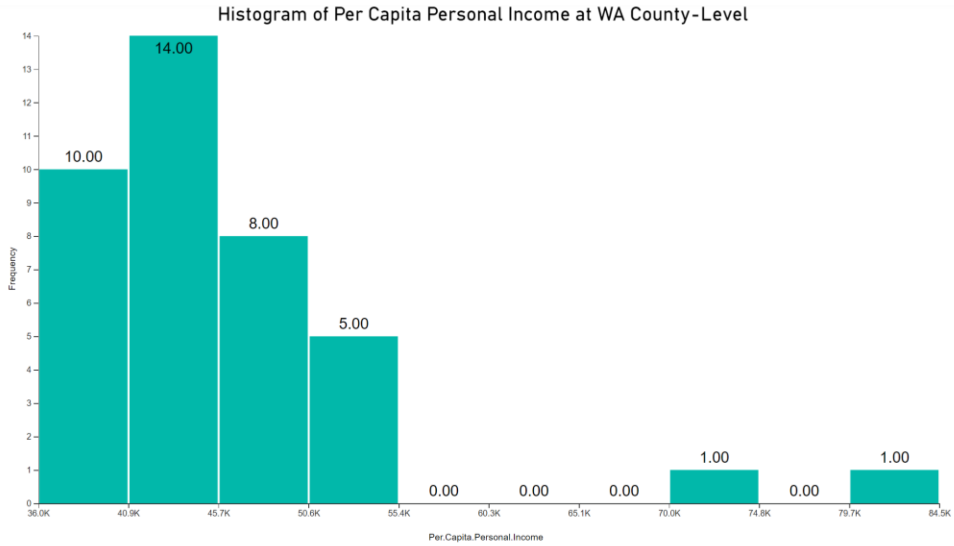

Figure 3: Distribution of Per Capita Personal Income at WA County-Level

Observation:

- At the WA County-Level, the distribution of per capita personal income (PCPI) is not normal across the WA state. King county has the highest PCPI, followed by San Juan county.
- Out of 39 counties, 37 counties have less than 56K PCPI, this account ~ 95% of the WA state.

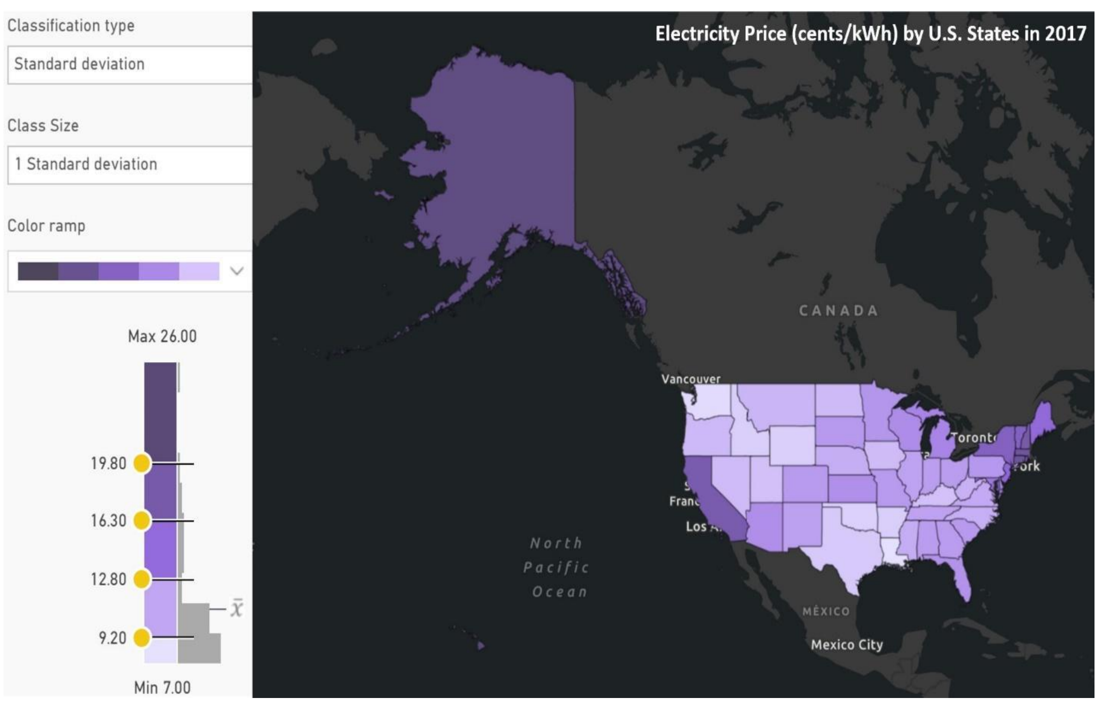

Figure 4: Map of Electricity Price by U.S. State

Observation:

This chart plots the electricity price by U.S. state in 2017. The electricity price values are plotted with color scale of one Standard Deviation

- We can see from the distribution that majority of the state has the low electricity price around the average of ~ 11.07 cents/kWh
- Hawaii is the state with the highest electricity price, followed by Alaska
- Louisiana is the state with the lowest electricity price, followed by Washington

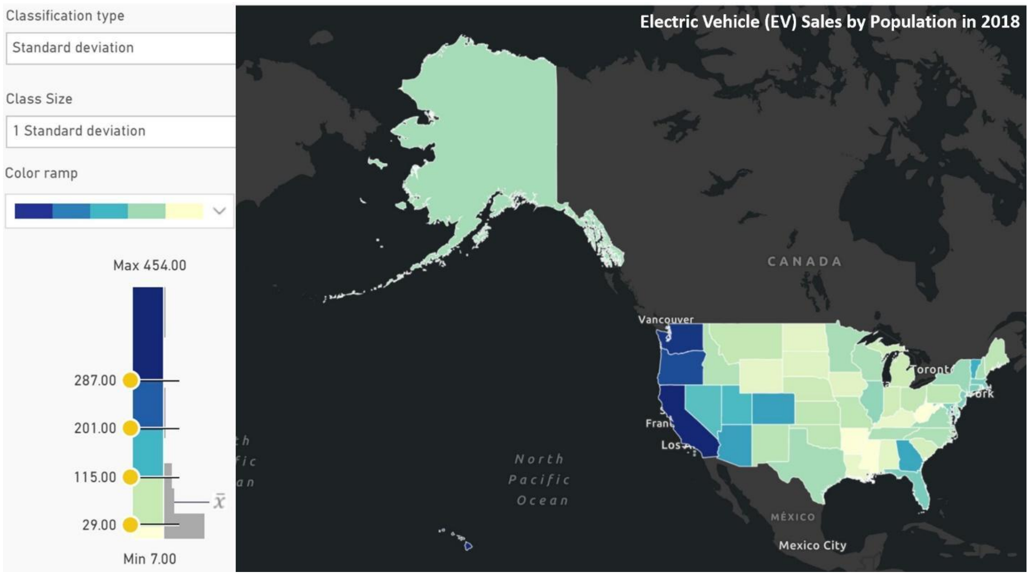

Figure 5: Map of Registered EV by Population by U.S. State

Observation:
This chart plots registered EV sales by population in U.S. state in 2018 with color scale of one Standard Deviation

- We can see from the distribution that the majority of the registered EV sales by population is very small across the U.S. except for the West coast states.
- California is the highest, followed by Washington, and Oregon on the map.
- A notable point to mention here: these three states are on the West Coast Electric Highway where there are EV charging stations for every 25 to 50 miles along the highway from Baja (CA – US) to Vancouver (BC – Canada).

Statistical Methods

There were three questions that we were interested in exploring regarding the sales of EVs in the US. Below is the description of each question and the analysis methods applied for the same.

Question 1: **Is there a relationship between the electricity prices of a state and the number of registered EVs in the state? Further, is there a positive or a negative relationship between the two?**

Since both count of registered EVs and electricity prices are quantitative variables, we used linear regression to answer the question. Furthermore, the coefficients of the linear regression tell us about the strength of the relationship and whether there is a positive or negative effect.
We used the electricity and registered EV counts data from 2017 for 50 US states (+ District of Columbia) for the analysis. Since the data for the states is highly dependent on the demographic population, we used the population data for the states from 2018, to standardize the registered EV counts.

Question 2: Do the economic factors of a state have an effect on the number of EV sales in a state? Which factors have the most significant effect?

To answer this question, we studied the following indicators at a state level:

- GDP
- Per capita personal Income
- Unemployment Rate

To study the effect of each of these quantitative indicators we used linear regression. We also analyzed whether there was a significant interaction between any of the indicators.

We used the economic indicators and registered EV counts data from 2017. We used the population data for the 50 states (+ District of Columbia) from 2018, to standardize the GDP and EV counts since they are dependent on the size of the state.

Question 3: Do the same economic factors (from Question 2) have a similar effect on the EV sales in the counties of Washington State?

To answer this question, we studied the following indicators at a county level:

- GDP
- Per capita personal Income
- Unemployment Rate

To study the effect of each of these quantitative indicators we used linear regression. We used the economic indicators and EV sales data from 2017. We used the population data for 39 counties in Washington from 2018, to standardize the GDP and EV counts since they are dependent on the size of the counties.

While answering all three questions, we used linear regression and thus we will state the assumptions of linear regression before we proceed.

1. **Independent observations:** Based on the data collection methods, the data observations in all our datasets are independent.
2. **Constant variance of residuals:** This will be observed based on the residual plots of the models.
3. **Normality or large sample size :**
   For all three questions, our data was limited to the number of demographic regions (50 states in the US + District of Columbia, 39 counties in Washington). Thus we needed to check for the normality and residual assumptions for our data. While analyzing our response variable, we found that the data was not normally distributed. On trying a few different transformations, we found

that the log-transform made the data quite normal (See Figure 6). This was affirmed by the residual plots (Figure 7).

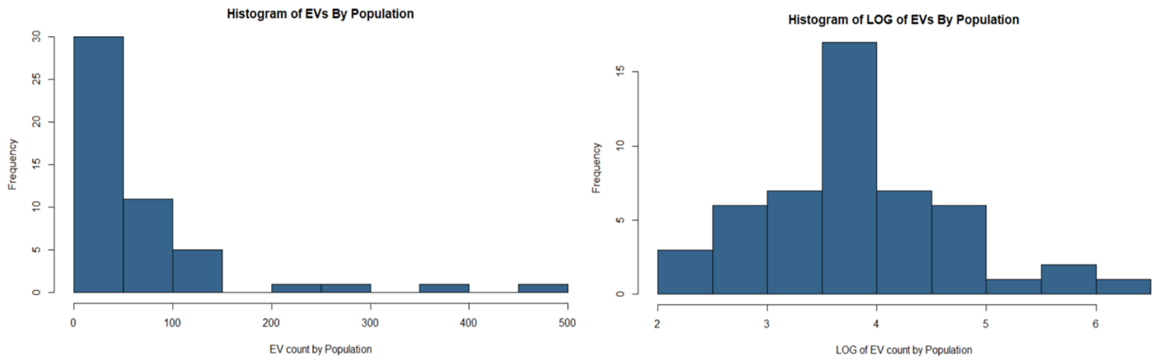

Figure 6: Histogram of EV counts by Population without (left) and with (right) Log-Transformation

Using this result we decide to opt for the log transformed EV count variable for the response variable of all our models.

4. **Linearity of the data points:**
   Although linearity is not a strict requirement for fitting a linear regression, it is good to have and we can see from the scatterplots that the relationship between the predictors i.e. electricity, per capita personal income, etc., have a linear relationship between the variables (Figure 7 & 8).

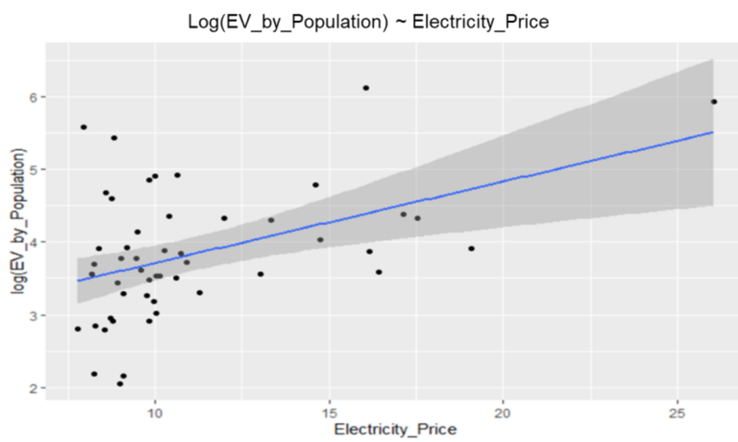

Figure 7: Scatterplots of response vs Electricity Price of US States

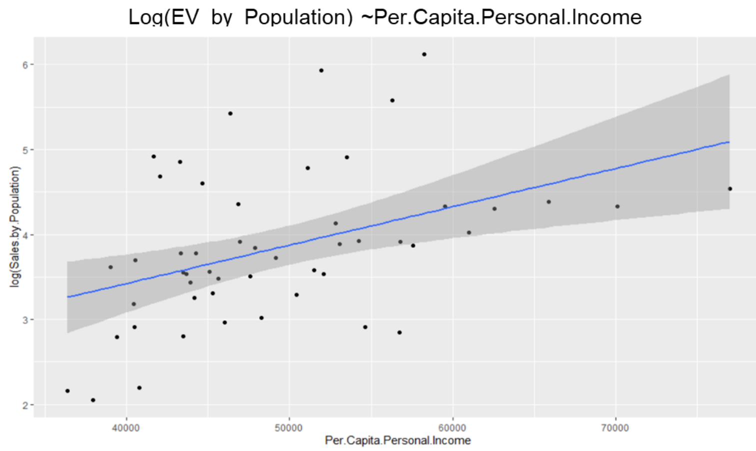

Figure 8: Scatterplots of response vs Per Capita Personal Income of US States

Residual Assumptions

Question 1:

Initially we fit a model of the EV registered vehicles (per 100,000 people) versus the Electricity Prices for the US states, without any transformation to the EV counts. We found that both the constant variance and normality assumptions were violated (Figure 9). On applying the log transformation (Figure 10) the normality assumption was satisfied. We found that the variances improved but were still slightly non-constant.

We find that the non-constant variance is due to the non-uniform distribution of data observations with higher EV counts. The limited size of the dataset is also a cause of concern. Thus the p-values and standard errors for the analysis must be inferred with this limitation in mind.

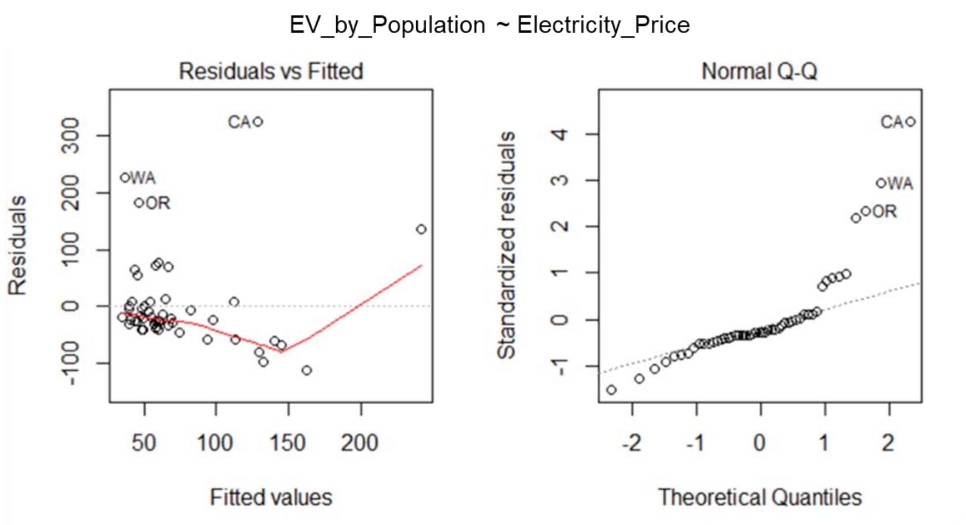

Figure 9: Residual Plots for Model Without Log Transformation

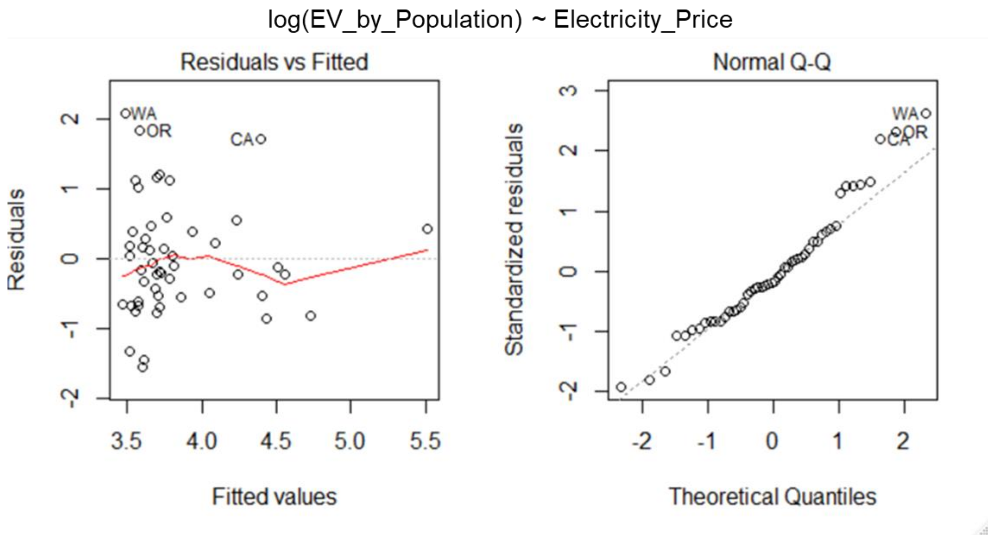

Figure 10: Residual Plots for Model With Log Transformation

Question 2:

Before fitting the model we have performed a few data transformations on the predictor variables.

- Per Capita Personal Income (Per Capita Personal Income/1000): Since a $1 change in income is very trivial, for model interpretability we scale it by $1000.

- Since the EV counts are standardized by population, it is important to scale all other factors at a per capita level as well. Unemployment Rate and Per Capita Personal Income are already at a per capita scale and don't need to be standardized. However GDP needs to be converted to per capita GDP.

We fitted linear regression models for the Log of the EV sales (per 100,000 people) in a state versus each economic factor. We noted that the normality and constant variance assumptions were not satisfied (despite the log transformation). The heavy tails in the data were due to states with high economic indicators like California, Hawaii and Oregon (Figure 11). These states had high residuals. On removal of these points, the models kept producing new outliers at the tails.

The OLS model had limitations because of the small data set size and the high leverage points. To deal with this issue, we decided to use the Weighted Least Squares (WLS) method (see Appendix) for linear regression instead of the OLS method. The WLS method gives more weight to points with smaller residuals and lower weight to points with higher residuals. We used Huber weights for fitting the WLS model.

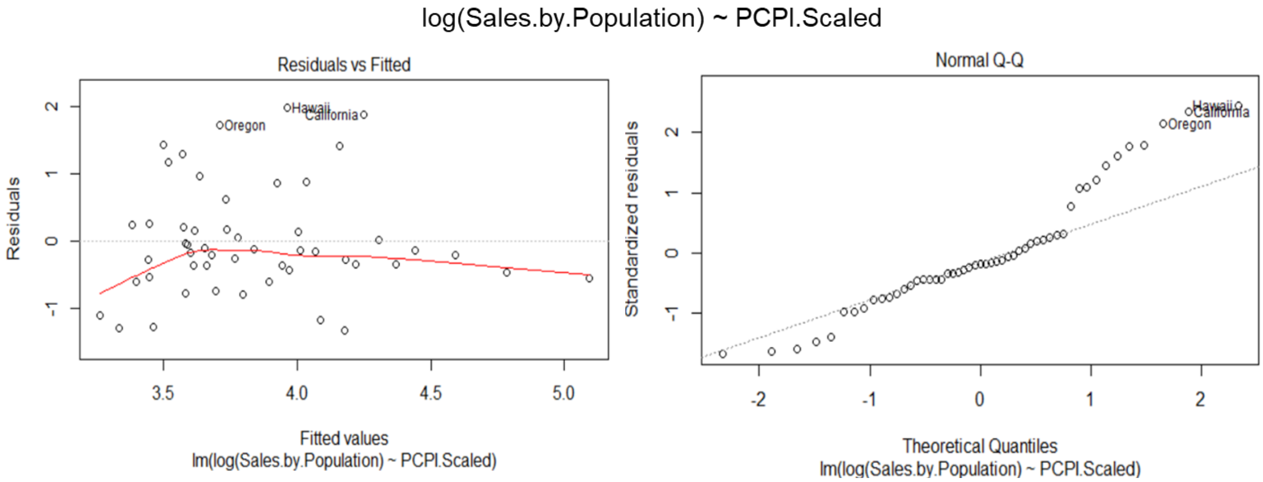

Figure 11: Residual plots for OLS regression of EV Sales by Population vs Per Capita Personal Income

Question 3:

We repeated the same process as was done in Question 2 on the data for Washington counties instead of US States. The limited data size produced a similar effect on the residuals for the OLS model as was seen in Question 2. Thus, we used the WLS with Huber weights in Question 3 as well.

In the next section we will see the results from our analysis and answer the questions posed at the beginning of the analysis.

Results & Discussions

Note: The R-squared values for all the linear regression models are quite poor (<< 0.5). This is because we do not have too many variables in the models and thus the variability in the response i.e. EV counts by Population is not well explained by these models. These models would thus prove to perform quite bad in terms of prediction. However, they will suffice to answer our questions regarding the significance of effect by the coefficients and their test statistics.

Question 1: Is there a relationship between the electricity prices of a state and the number of registered EVs in the state? Further, is there a positive or a negative relationship between the two?

	log(EV_by_Population)		
Predictors	*Estimates*	*std. Error*	*p*
(Intercept)	2.593 ***	0.378	**<0.001**
Electricity_Price	0.112 **	0.033	**0.001**
Observations	50		
R^2 / R^2 adjusted	0.198 / 0.181		

* $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table 1: OLS Summary for log(EV by Population) vs Electricity (US States)

Conclusion: We reject the null hypothesis that $\beta_{Electricity\_Price} = 0$ at 5% significance level.
Thus we conclude that there is a significant effect of Electricity Price on the EV Population. Based on the coefficient of regression, we see that the relationship is positive i.e. for a difference of 1¢/kWh of electricity prices there is an 11.8% increase in the EV sales per 100,000 people.

However, we expected a drop in the EV sales with an increase in the electricity prices. The data shows us otherwise. While trying to examine this result we found that there is a positive correlation between the Per Capita Personal Income and the Electricity Price of a state. Also there is a positive correlation between the Per Capita Personal Income and EV sales of a state. Based on this we infer that states with higher electricity prices tend to have a higher Per Capita Personal Income. Per Capita Personal Income has a significant effect on EV sales which makes sense since people would buy more EV cars when they have a higher purchasing power. Thus, electricity prices also seem to have a positive effect on the EV sales. The correlation between Per Capita Personal Income and Electricity Prices creates a confounding effect.

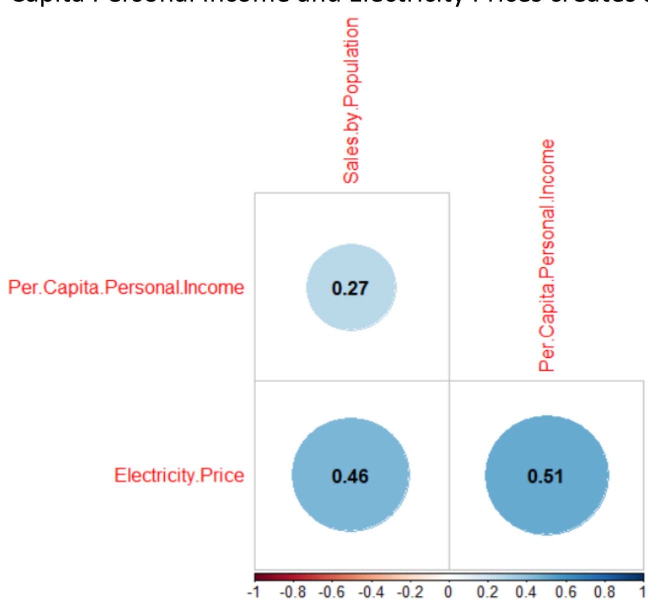

Figure 12: Correlations between Electric Prices, Per Capita Personal Income and Sales by Population

Predictors	log(Sales.by.Population) Estimates	std. Error	p	log(Sales.by.Population) Estimates	std. Error	p	log(Sales.by.Population) Estimates	std. Error	p
(Intercept)	1.56677 *	0.65664	**0.021**	2.59748 ***	0.37602	**<0.001**	1.61933 *	0.67900	**0.021**
PCPI.Scaled	0.02880	0.01522	0.065				0.04512 **	0.01358	**0.002**
Electricity.Price	0.07747 *	0.03664	**0.040**	0.11263 **	0.03240	**0.001**			
Observations	51			51			51		
R^2 / R^2 adjusted	0.253 / 0.222			0.198 / 0.181			0.184 / 0.167		

* p<0.05 ** p<0.01 *** p<0.001

Table 2: OLS Summary for (a) log(EV sales) vs Per Capita Personal Income and Electricity Prices, (b) log(EV sales) vs Electricity Prices, and (c) log(EV sales) vs Per Capita Personal Income (US States)

Question 2: Do the economic factors of a state have an effect on the number of EV sales in a state? Which factors have the most significant effect?

Predictors	log(Sales.by.Population) Estimates	std. Error	p	log(Sales.by.Population) Estimates	std. Error	p
(Intercept)	1.587	0.980	0.112	1.619 *	0.679	**0.021**
Unemployment.Rate	-0.046	0.132	0.731			
Per.Capita.GDP	-0.000	0.000	0.554			
PCPI.Scaled	0.056 *	0.023	**0.017**	0.045 **	0.014	**0.002**
Observations	51			51		
R^2 / R^2 adjusted	0.195 / 0.144			0.184 / 0.167		

* p<0.05 ** p<0.01 *** p<0.001

Table 3: OLS Summary for log(Sales by Population) vs (a) PCPI, Per Capita GDP and Unemployment Rate, (b) PCPI only (US States)

We see from Table 3, that in the full model with all 3 indicators, PCPI is the only significant indicator. In the individual fit as well, PCPI is significant and has a low p-value. However, from Table 4 below, we see that while Per Capita GDP is not significant in the full model, it is significant in the individual fit. The reason that it is not significant in the full model is because there is a correlation between Per Capita GDP and PCPI (Table 5 below). However, there is no correlation between Unemployment Rate and PCPI and thus the estimates and standard errors do not change much for the individual fit of the Unemployment Rate.

	log(Sales.by.Population)			log(Sales.by.Population)			log(Sales.by.Population)		
Predictors	*Estimates*	*std. Error*	*p*	*Estimates*	*std. Error*	*p*	*Estimates*	*std. Error*	*p*
(Intercept)	1.58674	0.97971	0.112	3.14889 ***	0.36524	**<0.001**	4.01824 ***	0.58745	**<0.001**
Unemployment.Rate	-0.04574	0.13245	0.731				-0.04158	0.13699	0.763
Per.Capita.GDP	-0.00006	0.00009	0.554	0.00012 *	0.00006	**0.049**			
PCPI.Scaled	0.05641 *	0.02273	**0.017**						
Observations	51			51			51		
R^2 / R^2 adjusted	0.195 / 0.144			0.077 / 0.058			0.002 / -0.018		

* *p<0.05* ** *p<0.01* *** *p<0.001*

Table 4: OLS Summary for log(Sales by Population) vs (a) PCPI, Per Capita GDP and Unemployment Rate, (b) Per Capita GDP only, and (c) Unemployment Rate only (US States)

Indicator	Correlation
Unemployment Rate	0.069
Per Capita GDP	0.785

Table 5: Correlation between PCPI vs Unemployment Rate and Per Capita GDP (US States)

	log(Sales.by.Population)		
Predictors	*Estimates*	*std. Error*	*p*
(Intercept)	-15.038	19.823	0.452
Unemployment.Rate	2.038	3.764	0.591
Per.Capita.GDP	0.002	0.004	0.605
PCPI.Scaled	0.497	0.369	0.185
Unemployment.Rate * Per.Capita.GDP	-0.000	0.001	0.816
Unemployment.Rate * PCPI.Scaled	-0.066	0.064	0.309
Per.Capita.GDP * PCPI.Scaled	-0.000	0.000	0.374
(Unemployment.Rate * Per.Capita.GDP) * PCPI.Scaled	0.000	0.000	0.536
Observations	51		
R^2 / R^2 adjusted	0.266 / 0.146		

* *p<0.05* ** *p<0.01* *** *p<0.001*

Table 6: Model with Interactions (US States)

Conclusion: Thus, PCPI and Per Capita GDP have an effect on the EV Sales by Population of states. Unemployment Rate does not. There are no significant interactions between the indicators (Table 6).
As discussed in the previous sections we use WLS method for fitting the models due to the limitations of size and high leverage points.

The coefficients for the WLS are not very different from the OLS method as can be seen from Table 7.

	log(Sales.by.Population)			log(Sales.by.Population)		
Predictors	*Estimates*	*std. Error*	*p*	*Estimates*	*std. Error*	*p*
(Intercept)	1.70197 **	0.58304	**0.005**	1.61933 *	0.67900	**0.021**
PCPI.Scaled	0.04106 ***	0.01166	**0.001**	0.04512 **	0.01358	**0.002**
Observations	51			51		
R^2 / R^2 adjusted	NA			0.184 / 0.167		

* *p<0.05* ** *p<0.01* *** *p<0.001*

Table 7: Summary of (a) WLS fit and (b) OLS fit for log(EV Sales) vs PCPI (US States)

For a difference of $1000 in the Per Capita Personal Income, the number of registered EVs per 100,000 people is **4.2%** higher.

Question 3: Do the same economic factors (from Question 2) have a similar effect on the EV sales in the counties of Washington State?

	log(Sales.by.Population)			log(Sales.by.Population)		
Predictors	*Estimates*	*std. Error*	*p*	*Estimates*	*std. Error*	*p*
(Intercept)	4.60104 ***	0.98638	**<0.001**	4.14237 ***	0.50713	**<0.001**
GDP.by.Population	-0.00000 **	0.00000	**0.007**			
Per.Capita.Personal.Income.per.thousand	0.07594 ***	0.01319	**<0.001**	0.06680 ***	0.01083	**<0.001**
Unemployment.Rate	-0.05221	0.08854	0.559			
Observations	39			39		
R^2 / R^2 adjusted	0.602 / 0.568			0.507 / 0.494		

* *p<0.05* ** *p<0.01* *** *p<0.001*

Table 8: OLS Summary for log(Sales by Population) vs (a) PCPI, Per Capita GDP and Unemployment Rate, (b) PCPI only (Washington counties)

From Table 8 we see that in the full model with all 3 indicators, PCPI and GDP by Population are significant indicators. In the individual fit as well, PCPI is significant and has a low p-value. However, from Table 9 below, we see that while Per Capita GDP is significant in the full model, it is not significant in the individual fit. The reason that it is significant in the full model is because there is a correlation between Per Capita GDP and PCPI (Table 10 below). Also, Unemployment Rate (which has a negative correlation with PCPI), is significant in the individual fit but not in the full model. Thus, we can see the confounding effects of the variables in the models.

	log(Sales.by.Population)			log(Sales.by.Population)			log(Sales.by.Population)		
Predictors	*Estimates*	*std. Error*	*p*	*Estimates*	*std. Error*	*p*	*Estimates*	*std. Error*	*p*
(Intercept)	4.60104 ***	0.98638	**<0.001**	7.16054 ***	0.30940	**<0.001**	8.94621 ***	0.56974	**<0.001**
GDP.by.Population	-0.00000 **	0.00000	**0.007**	0.00000	0.00000	0.849			
Per.Capita.Personal.Income.per.thousand	0.07594 ***	0.01319	**<0.001**						
Unemployment.Rate	-0.05221	0.08854	0.559				-0.29821 **	0.09573	**0.004**
Observations	39			39			39		
R^2 / R^2 adjusted	0.602 / 0.568			0.001 / -0.026			0.208 / 0.186		

* $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table 9: OLS Summary for log(Sales by Population) vs (a) PCPI, Per Capita GDP and Unemployment Rate, (b) Per Capita GDP only, and (c) Unemployment Rate only (Washington counties)

Indicator	Correlation
Unemployment Rate	-0.609
Per Capita GDP	0.427

Table 10: Correlation between PCPI vs Unemployment Rate and Per Capita GDP (Washington counties)

Conclusion: Thus, PCPI and Unemployment Rate have a significant effect on the EV Sales by Population of counties. Per Capita GDP does not. Note that at a country level, PCPI and Per Capita GDP were significant but the Unemployment Rate was not.

From the WLS coefficients we conclude:

- For a difference of $1000 in the Per Capita Personal Income, the number of registered EVs per 100,000 people in Washington State is **6.9%** higher.

  Compare this to the country level where % difference of the registered EVs per 100,000 people was **4.2%.**

- For a 1% change in Unemployment Rate, the number of registered EVs per 100,000 people in Washington State is **26.4% lower**.

  Compare this to the country level where there is **no significant effect** of Unemployment rate on the EV sales.

	log(Sales.by.Population)			log(Sales.by.Population)		
Predictors	*Estimates*	*std. Error*	*p*	*Estimates*	*std. Error*	*p*
(Intercept)	4.142 ***	0.507	**<0.001**	4.097 ***	0.595	**<0.001**
Per.Capita.Personal.Income.per.thousand	$\boxed{0.067}$ ***	0.011	**<0.001**	$\boxed{0.069}$ ***	0.013	**<0.001**
Observations	39			39		
R^2 / R^2 adjusted	0.507 / 0.494			NA		

* $p<0.05$ ** $p<0.01$ *** $p<0.001$

	log(Sales.by.Population)			log(Sales.by.Population)		
Predictors	*Estimates*	*std. Error*	*p*	*Estimates*	*std. Error*	*p*
(Intercept)	9.009 ***	0.605	**<0.001**	8.946 ***	0.570	**<0.001**
Unemployment.Rate	$\boxed{-0.307}$ **	0.102	**0.005**	$\boxed{-0.298}$ **	0.096	**0.004**
Observations	39			39		
R^2 / R^2 adjusted	NA			0.208 / 0.186		

* $p<0.05$ ** $p<0.01$ *** $p<0.001$

Table 11: Summary of (a) WLS fit and (b) OLS fit for log(EV Sales) vs PCPI and Unemployment Rate

Future Work

Most of the future work relies on getting a richer and better quality of data. Below are the four desired data points that will greatly benefit our study:

1. Our current datasets only give us the electricity price at the US state level, we don't have this data for the WA county level. We hope to obtain the electricity price for the county level. This will indeed help us perform the test between state and county effectively.

2. According to the US Department of Energy, California (CA) is the leading state in the nation in terms of registered EVs, followed by Washington state (WA). We would like to study the relationship of why these two states have the most registered EVs than any other states. One notable point is worth to discover, CA has Silicon Valley in the South and WA is considered a Silicon Valley in the Pacific Northwest. There are many big tech companies located in these two regions. The number of high-tech workers and their salaries contributed to the economic growth of the whole region. If we have the income salary data of those tech professionals for CA and WA, we can conduct a test for the effect of tech salary on EV sales as well as the interaction of this factor with other economic factors in the linear regression model.

3. Since California (CA), Oregon (OR), and Washington (WA) belong to the West Coast Electric Highway project [http://westcoastgreenhighway.com/electrichighway.htm], there are EV charging stations located every 25 to 50 miles from Baja (CA - USA) to Vancouver (BC- Canada). If we have a better quality of charging station in each state across the US, this could be a contributing factor as well to consider for our hypothesis test.

4. As previously mentioned in our dataset description regarding the limitation of the datapoint, we would like to acquire more historical data points (spanning multiple years). This will allow us to

expand our analysis on trend, study the time-based effect, and increase the sample size in our study.

References

[1] Department of Licensing. "Electric Vehicle Title and Registration Activity: Data.WA: State of Washington." Data.WA.gov, March 5, 2020. https://data.wa.gov/Transportation/Electric-Vehicle-Title-and-Registration-Activity/rpr4-cgyd.

[2] "U.S. Energy Information Administration - EIA - Independent Statistics and Analysis." State Electricity Profiles - Energy Information Administration. Accessed March 17, 2020. https://www.eia.gov/electricity/state/.

[3] "U.S. Energy Information Administration - EIA - Independent Statistics and Analysis." State Electricity Profiles - Energy Information Administration. Accessed March 17, 2020. https://www.eia.gov/electricity/state/.

[4] "Counties." FRED. Federal Reserve Bank of St. Louis. Accessed March 17, 2020. https://fred.stlouisfed.org/categories/30336.

[5] "States." FRED. Federal Reserve Bank of St. Louis. Accessed March 17, 2020. https://fred.stlouisfed.org/categories/27281.

[6] "U.S. Census Bureau QuickFacts: United States." Census Bureau QuickFacts. Accessed March 17, 2020. https://www.census.gov/quickfacts/fact/table/US/PST045219.

[7] "Robust Regression." IDRE Stats. Accessed March 17, 2020. https://stats.idre.ucla.edu/r/dae/robust-regression/.

[8] "Electric Vehicle Sales Forecast and the Charging Infrastructure Required Through 2030", Edison Electric Institute (EEI) Report, November, 2018. https://www.edisonfoundation.net/iei/publications/Documents/IEI_EEI%20EV%20Forecast%20Report_Nov2018.pdf

Appendix

We used a method of robust regression as we had a very small sample size and a number of strong outliers. Even if we removed some outliers, we kept getting new outliers. To account for heteroskedasticity we used a weighted linear regression model.

Weighted Least Squares

Weighted Least Squares is a generalisation of Ordinary Least Squares in which the error covariance matrix is allowed to be different from an identity matrix. Instead of minimizing a residual sum of squares, we minimize the weighted sum of squares.

$$RSS(\beta) = \sum_{i=1}^{n} (y_i - x_i \cdot \beta)^2$$
$$WSS(w, \beta) = \sum_{i=1}^{n} w_i (y_i - x_i \cdot \beta)^2$$

The ordinary least squares is a special case when all $w_i = 1$

The weights depend on the residuals and the equation is solved iteratively and the method is called Iterative reweighted least squares (IRLS). The coefficient matrix at iteration j is given by:

$$\beta_j = [X' W_{j-1} X]^{-1} X' W_{j-1} Y$$

This process continues until it converges.

For our analyses we used Huber weights which are inversely proportional to the residuals. In Huber weighting, observations with small residuals get a weight of 1 and the larger the residual, the smaller the weight. This is defined by the weight function:

$$w(e) = \{1 \; for \; |e| \leq k\}$$
$$w(e) = \left\{\frac{k}{|e|} \; for \; |e| \geq k\right\}$$

The results which we get from WLS were not very far away from the ones we get from OLS as there were only few observations that had high leverage. Using the weights, it helped to normalize the data and to get better estimates of the predictors.

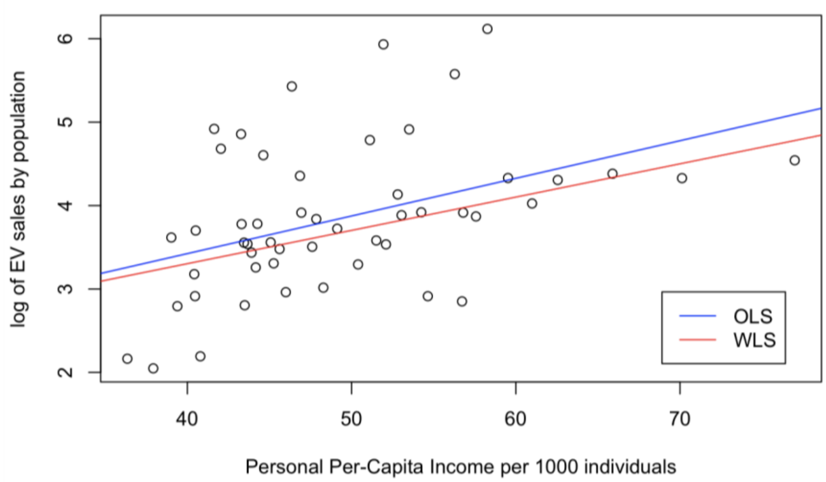

Figure 13: Fitted model comparison for WLS and OLS