Journal of Advances in Developmental Research (IJAIDR)

E-ISSN: 0976-4844 ● Website: www.ijaidr.com ● Email: editor@ijaidr.com

Investigating the Learning Effect of Users Driving a Plug-In Hybrid Vehicle

Zaid Thanawala

Introduction

Hybrid electric vehicles (HEVs), battery electric vehicles (BEVs) and plug-in hybrid electric vehicles (PHEVs) are becoming more popular. HEVs cannot be charged using an electric outlet and only have an on-board energy recovery system which charges a small battery. BEVs are charged only by plugging into electric outlets. PHEVs are charged by either plugging into electric outlets or by means of on-board electricity generation. These vehicles can drive at full power in electric-only mode over a limited range. As such, PHEVs offer valuable fuel flexibility (Raskin & Shah, 2006). PHEVs may have a larger battery and a more powerful motor compared to a HEV, but their range is still very limited (Anderman, 2004). These vehicles would also enable meeting the demands for electrical power due to the increasing use of the electronic features to improve vehicle performance, fuel economy, emissions, passenger comfort, and safety (Emadi, et al., 2005). Reducing dependence on foreign oil and emissions of CO_2 and particulates are among the leading reasons that plug-in hybrid electric vehicles (PHEVs) are increasing in popularity. Most PHEVs are planned to have a full electric range between 10–40 miles. This is within the daily commute distance of the average American driver (Schneider, et al., 2008).

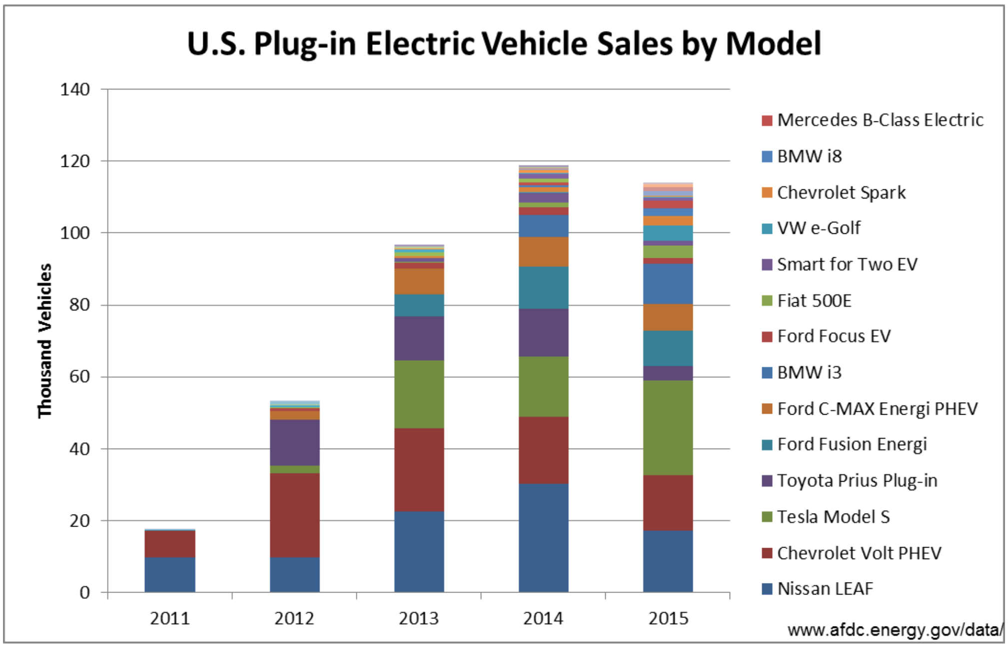


Figure 1: US PHEV Sales by Model (Energy, 2016)

Figure 1 shows an increasing trend in people purchasing PHEVs per year. According to the U.S. DOE, the sales in year 2015 were slightly lower due to introduction of new BEVs.

Most economists conceptualize learning by firms as a costless and automatic process, a joint output of production activity, which results in reductions in average production (labor) costs. In the simplest model of learning, average cost reductions are expressed as a function of cumulated output, investment or time (Malerba, 1992). Learning by firms is one of the most significant dynamic processes taking place within industries. There is a long tradition of empirical case studies on learning curves and yield improvements at both the firm and the industry levels usually associated to the effects of learning by doing. A learning model based on learning by doing and learning curves has been used in analyses of productivity growth at both the sectorial and the macro levels. In such research, learning is not the center of attention, but only one of the variables that are used to explain productivity growth(Wright, 1936).

Performance gains from the learning curve effect are common but not universal. There is no physical law that requires individuals, work groups, companies or industries to learn from their experience. Performance gains come from a variety of mechanisms discussed below but behind these mechanisms is a willingness to learn, an ability to learn and, in many cases, an investment in learning. Individuals, work groups, companies and industries that do not have the willingness, ability or investment may find their costs declining very little or, even increasing. The most common cause of slow learning is the willingness component that results from corporate or industry arrogance.

Another concept that arises from learning curves are experience curves. Experience Curves are an expansion of the Learning Curve idea from individual and group learning to factories, companies or entire industry sectors. Companies can use Experience Curves to develop marketing and manufacturing strategy. Experience Curves are usually established over longer time periods than Learning Curves. In addition, market price is often used as a substitute for actual cost since costs for such a wide-ranging study are often unavailable. Experience curves are similar in behavior and are often represented by the same formulae as Learning Curves. But, there are some differences:

- The cost improvements are often the result of macro-level changes in systems, technologies and culture rather than individual or group experience.
- Improvements from individuals and work groups certainly can contribute when widespread and accumulated over time.

These differences between Learning and Experience curves result in differences in their use and application. Experience curves apply to Manufacturing, Marketing and Business strategy. This contrasts with Learning Curves which are most useful for tactical applications such as evaluating work group performance or estimating product cost. The concept was first proposed by Bruce Henderson of the Boston Consulting Group (BCG) (Quarterman Lee, 2014).

For this study we look at households (owning Chevy Volts) as firms and the output of production can be thought of as the money spent on gas over time. Thus, using the simplest model of learning, reduction in average money spent on gas can be expressed as a function of time as defined by Malerba *et al.*(Malerba, 1992). Learning curves of users will describe how efficient they become at a particular task through constant repetition. The goal of this study is to collect data from real PHEV users and analyze how their expenditure on gas will reduce over time by observing the difference in the Total Miles driven and EV Miles driven.

Methodology

Time Series

To observe a trend over time, a time series was run on the data. A univariate time series is a sequence of measurements of the same variable collected over time. The measurements are made at regular time intervals (monthly in this case). One difference from standard linear regression is that the data are not necessarily independent and not necessarily identically distributed. One defining characteristic of time series is that this is a list of observations where the ordering matters. Ordering is very important because there is dependency and changing the order could change the meaning of the data.

The basic objective usually is to determine a model that describes the pattern of the time series. Uses for such a model are:

- To describe the important features of the time series pattern.
- To explain how the past affects the future or how two time series can “interact”.
- To forecast future values of the series.

There are two basic types of time domain models:

- Models that relate the present value of a series to past values and past prediction errors - these are called ARIMA models (for Autoregressive Integrated Moving Average). This is the model that is used in this study.
- Ordinary regression models that use time indices as x-variables. These can be helpful for an initial description of the data and form the basis of several simple forecasting methods.

Since ARIMA models were used for this study, they are explained further. ARIMA models, also called Box-Jenkins models, are models that may possibly include autoregressive terms, moving average terms, and differencing operations.

Various abbreviations are used:

- When a model only involves autoregressive terms it may be referred to as an AR model. When a model only involves moving average terms, it may be referred to as an MA model
- When no differencing is involved, the abbreviation ARMA may be used.

Three items should be considered to determine a first guess at an ARIMA model: a time series plot of the data, the ACF (autocorrelation function), and the PACF (partial autocorrelation function).

- An AR has a sinusoidal ACF that converges to 0.
- MA models have theoretical ACFs with non-zero values at the MA terms in the model and zero values elsewhere
- ARMA models (including both AR and MA terms) have ACFs and PACFs that both tail off to 0.

- If the ACF and PACF do not tail off, but instead have values that stay close to 1 over many lags, the series is non-stationary and differencing will be needed.
- If all autocorrelations are non-significant, then the series is random (white noise; the ordering matters, but the data are independent and identically distributed.)

After a model has been estimated, The Ljung-Box statistic, also called the modified Box-Pierce statistic, is a function of the accumulated sample autocorrelations, r_j, up to any specified time lag m is performed. This statistic can be used to examine residuals from a time series model in order to see if all underlying population autocorrelations for the errors may be 0 (up to a specified point). The next step in this case is forecasting the data.

Clustering

Since there are so many discreet users, clustering was thought of as a way to divide the users into groups to be analyzed further. Clustering is the process of organizing objects into groups whose members are similar in some way. A cluster is therefore a collection of objects which are similar between them and are dissimilar to the objects belonging to other clusters. The goal of clustering is to determine the intrinsic grouping in a set of unlabeled data.

Data collection

To accomplish the goal of this study, data was collected from real Chevrolet Volt users. Chevrolet Volt is one of the highest selling plug-in car that is currently sold in the U.S. The Volt comes with an OnStar subscription with allows users to easily manage the capabilities of their electric vehicle with extended range including setting charge alerts, checking on current battery charge level and contacting a dedicated Volt advisor. A user created website (www.voltstats.net) not affiliated with Chevrolet in any way, allows Volt owners to sync their OnStar performance data to compare to other Volt users. The users are not required to enter any personally identifying information.

Volt Stats obtains the following information through OnStar:

- Volt Identification Number (different from the vehicle identification number/VIN)
- Reading Timestamp
- Odometer Reading
- EV Odometer Reading
- Total Gallons of Gas Burned

All information that is obtained by Volt Stats is considered public and viewable by third parties. Thus, using an SQL script the website was mined for the aforementioned data. Data was collected for the 50 oldest owners of the vehicle so as to get a better understanding of how driving patterns change over time.

Data filtering

The data collected needed to be filtered so as to import it in R and run a time series analysis on it. Since the users had different time periods of data, a new data set was created with people having ownership of

more than 35 months. This data was organized in ascending order of months and a time series analysis was performed on it.
A supplementary dataset was made with the same users but this time, data was organized for the latest 35 months of ownership and a similar analysis was performed on this data set.

Preliminary analysis

Four random users were chosen and the difference between total miles and ev miles driven were plotted against months of ownership. Using the regression function in Excel, a trend line was also plotted to check whether it was feasible to hypothesize the existence of a learning curve for PHEV owners.

User – 1

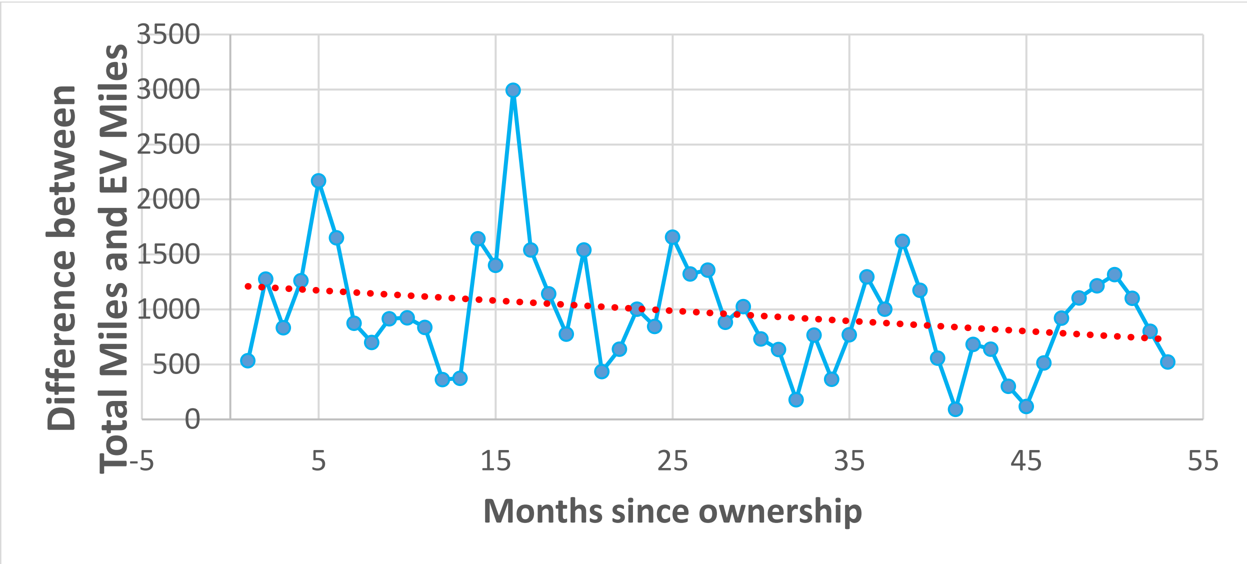


Figure 2: Preliminary analysis-User 1

User – 2

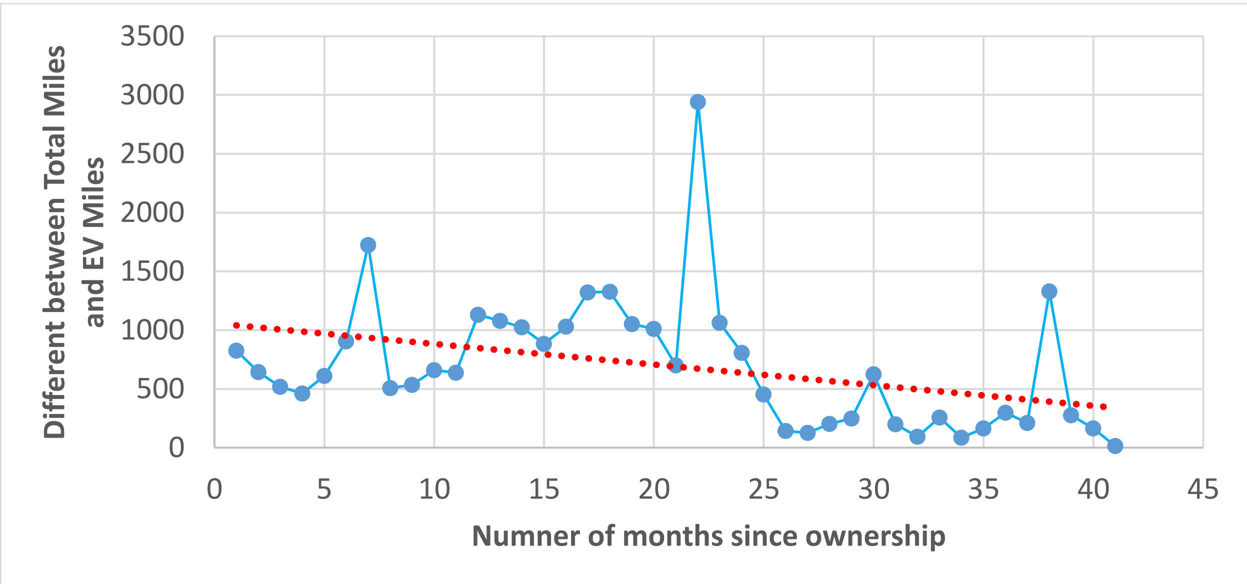


Figure 3: Preliminary analysis: User 2

User – 3

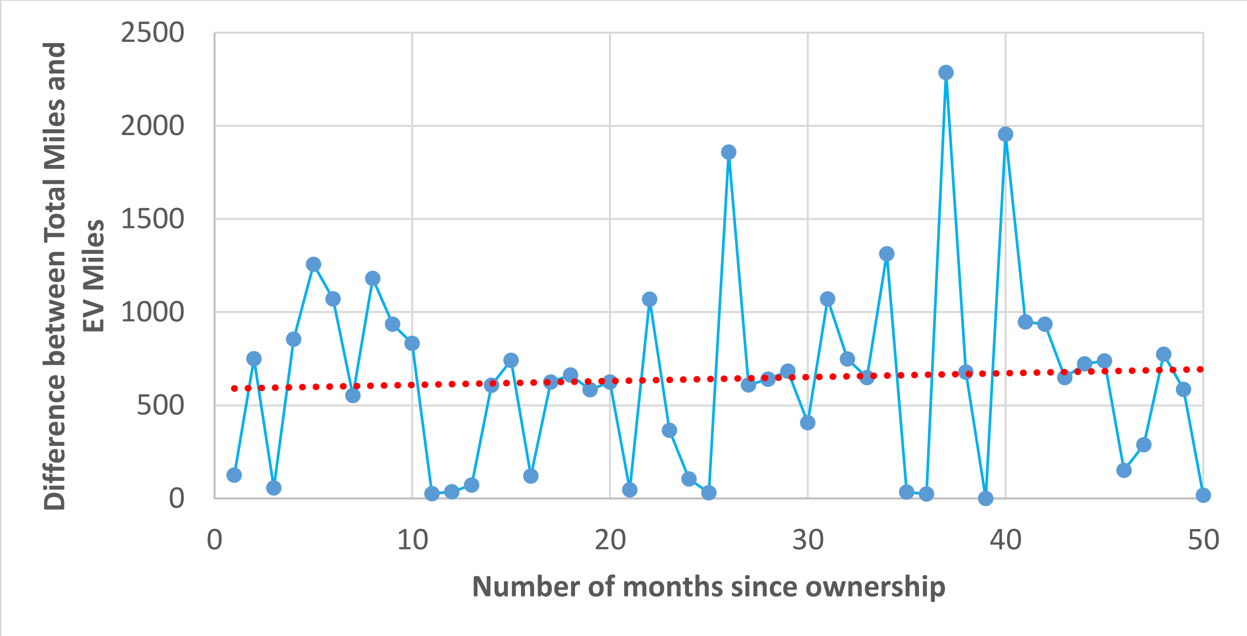


Figure 4: Preliminary analysis: User 3

User – 4

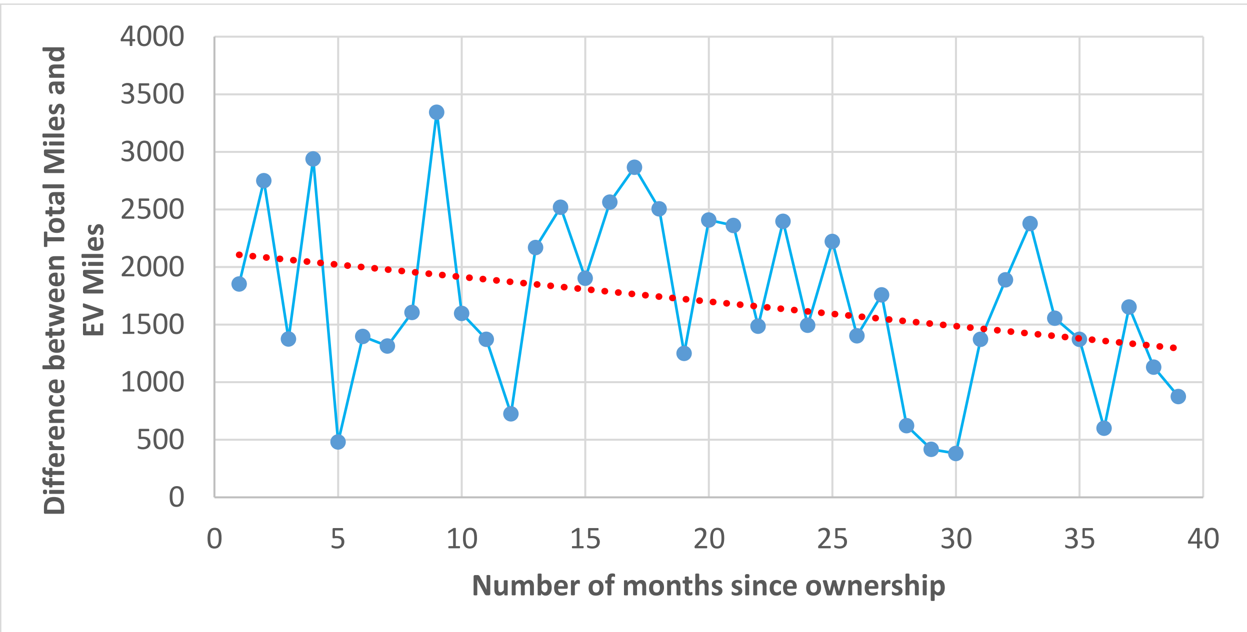


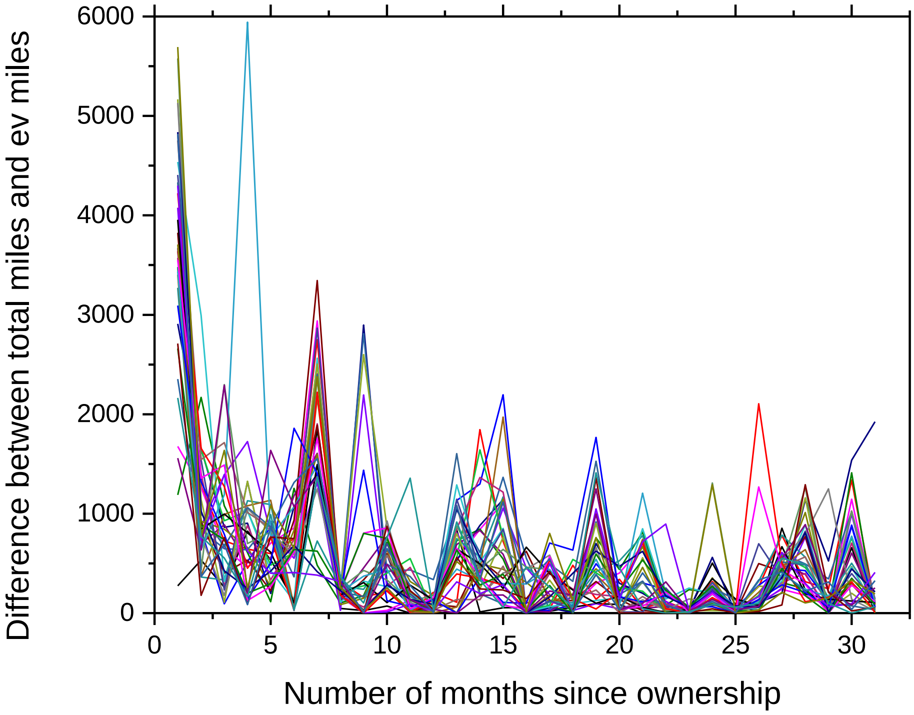
Figure 5: Preliminary analysis: User 4

The above graphs show that there is trend which shows that the difference between total miles and EV miles is reducing. This implies that the users are driving more electric miles and using less gas which in turn supports our hypothesis that they spend less money on gas over time. User-3 has an unusual graph in which the trend is slightly increasing/almost constant, but that could be due to geo-spatial reasons and not attributed to learning. There is not enough data to pinpoint the exact problem. But overall, the preliminary analysis shows that there is a learning curve in users who buy a PHEV. Thus, a time series analysis was performed on the data to have a closer look.

Journal of Advances in Developmental Research (IJAIDR)
E-ISSN: 0976-4844 ● Website: www.ijaidr.com ● Email: editor@ijaidr.com

There are over 31 users in the new datasets that were created but R can perform a time-series analysis only on one dataset. Thus, clustering was thought of as a way to average out data for similar users.

Clustering



The graph above is a plot of 31 unique individuals’ difference between total miles and EV miles driven v/s the number of months since ownership. It can be seen that there are no visible clusters since the users have very similar graphs. The ‘kmeans’ function in R also determined that there were no clusters in the dataset. Due to this, all the users’ data were averaged out to run a time series analysis to observe the learning effect over time.

Time Series

1st data set

First data set was created with aggregated and averaged out data for 31 unique individuals over their first 35 months of ownership of the Volt.

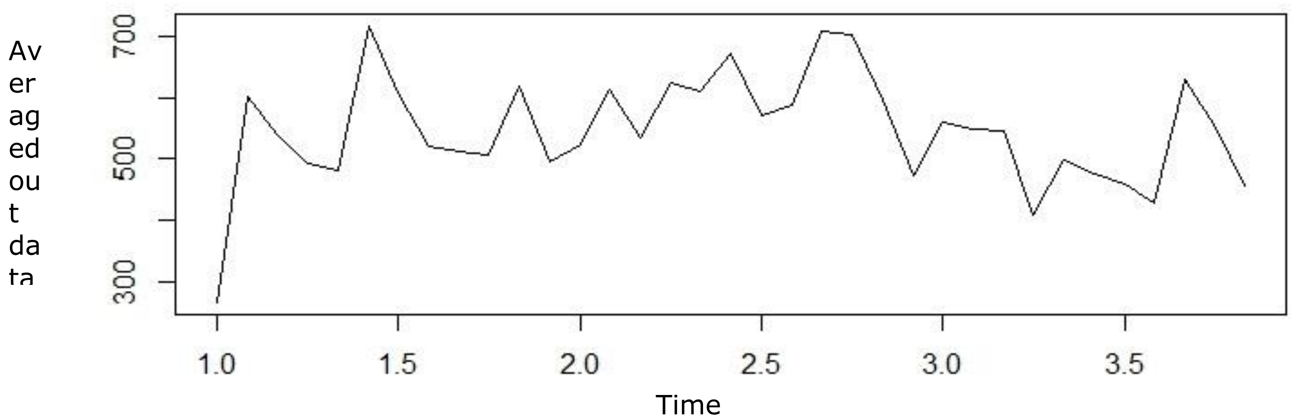


Figure 6: Time series for 1st dataset

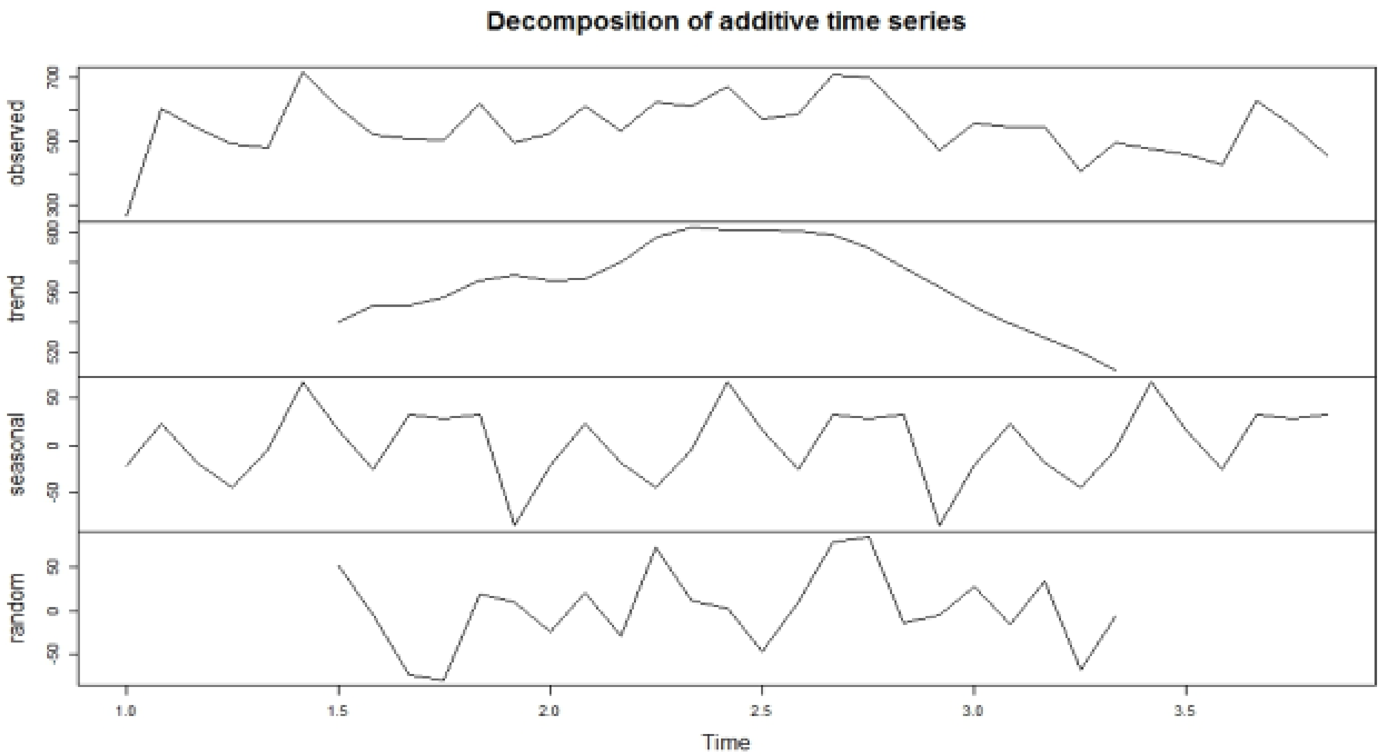


Figure 7: Decomposition of time series for the 1st dataset

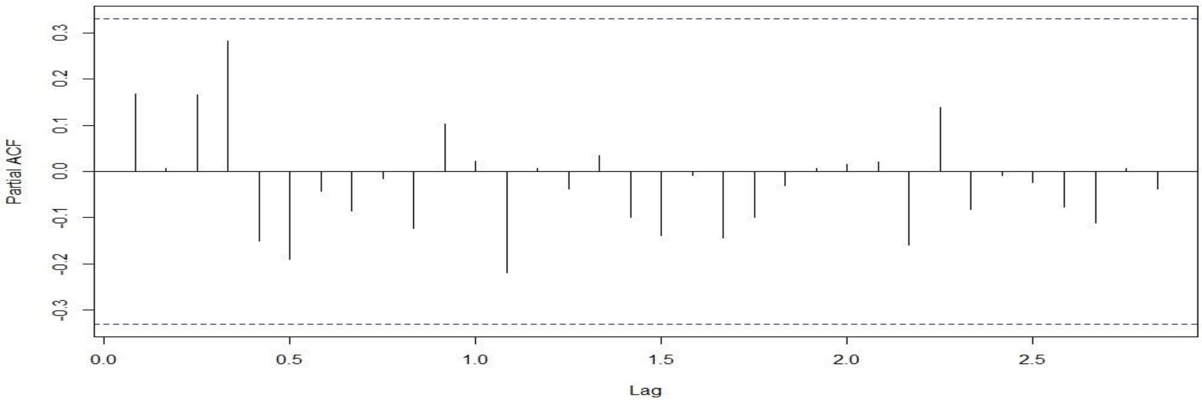


Figure 8: PACF for time series of 1st dataset

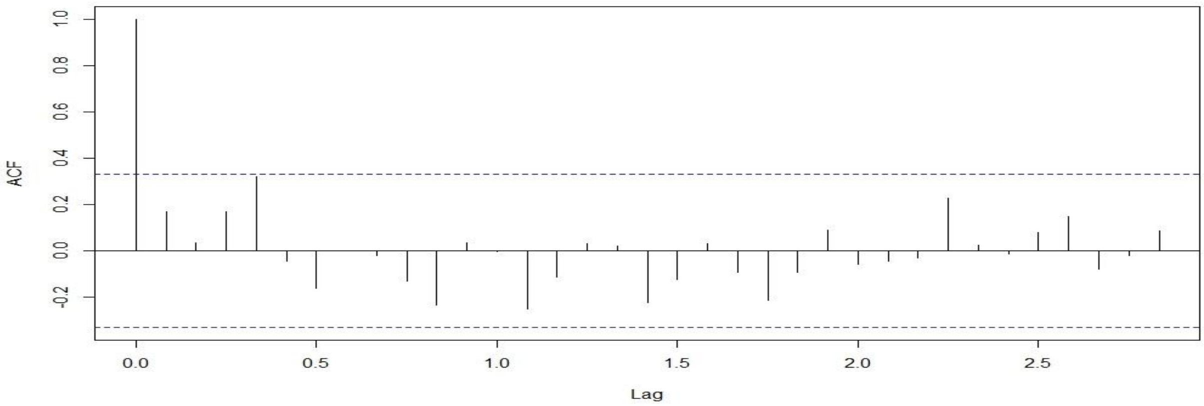


Figure 9: ACF for time series of 1st dataset

Looking at the PACF and ACF from figure 8 and 9 we cannot choose a model for the time series because they are very random and not representative of the actual data. A difference filtering with a lag of 1 is used to remove the trend component observed in Figure 7 since that is a requirement for the ARIMA model. The inherent trend could also cause distortion of the data. To find the orders of the ARIMA (p,d=1,q) model we try different values of p and q, and choose the combination which gives lowest AIC and BIC

p	**d**	**q**	**AIC**	**BIC**
0	1	0	273.81	274.9
0	1	1	273.78	277.05
1	1	0	276.34	279.61
1	1	1	277.48	282.93

Table 1: p,d,q values for the ARIMA model

From Table 1 it is clear that the p and q values are 0 and 0.

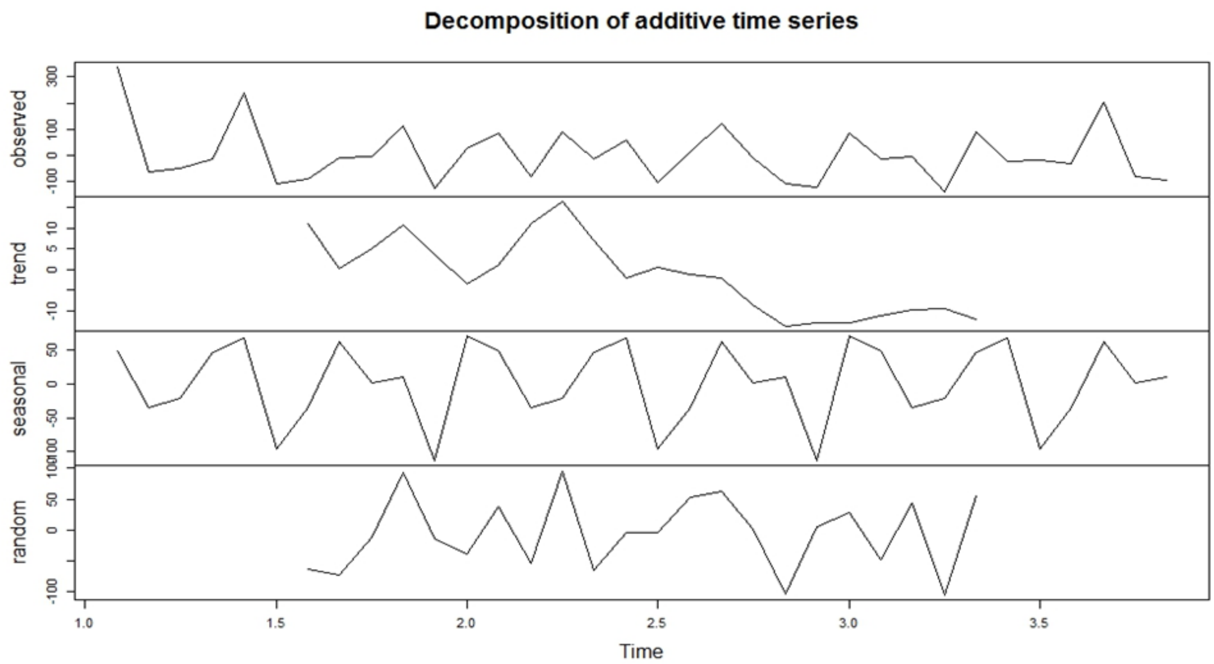


Figure 10: Decomposition of time series for 1st dataset with lag=1

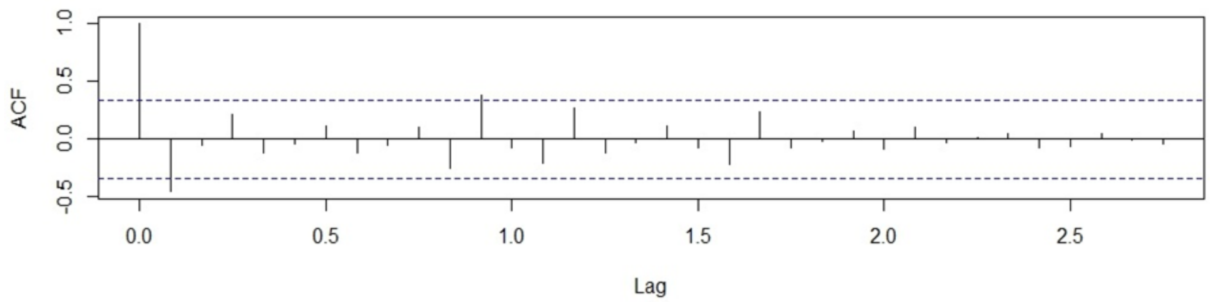


Figure 11: ACF for 1st time series with lag=1

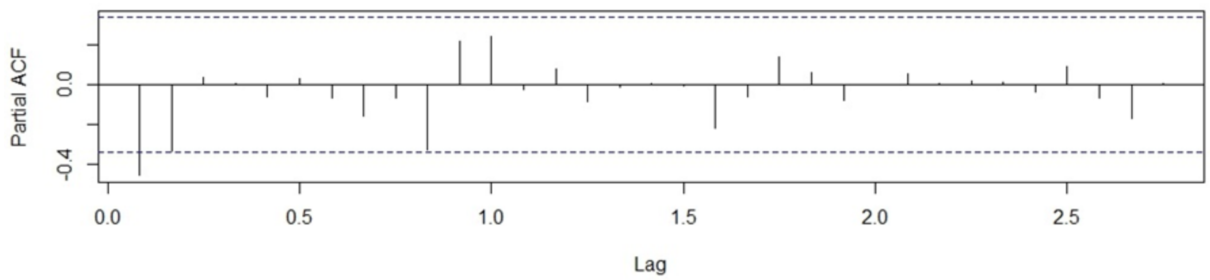


Figure 12: PACF for 1st time series with lag =1

Journal of Advances in Developmental Research (IJAIDR)

E-ISSN: 0976-4844 ● Website: www.ijaidr.com ● Email: editor@ijaidr.com

From Figure 11 and Figure 12 we see that the ACF goes to zero exponentially and the PACF cuts off, we can choose a moving average model with p=0. We run an ARIMA model and perform a Ljung-Box test.

Null hypothesis: $\varphi_1 = \varphi_2 = \cdots \varphi_L = 0$ which implies that the model does not exhibit lack of fit

Alternate hypothesis: $\varphi_1 \neq \varphi_2 \neq \cdots \varphi_L \neq 0$ which implies that the model exhibits lack of fit

$X^2 = 11.709$, df = 1

$X^2_{critical}$ (97.5% confidence) = 5.023

Since $X^2_{critical}$ is less than X^2, we accept the null hypothesis which means that the residuals do not have any correlation and they are just random data or white noise. Thus the model does not exhibit lack of fit and is perfect.

In the decomposed time series data there is an inherent trend of the difference between total miles and EV miles reducing over time. This implies that people are driving more electric miles than on gas miles. So there is a learning effect that is observed over the first 35 months of ownership. There is a seasonal effect that is also observed that is because the batteries operate better at around 25°C. During the winter when it is cold, energy is lost to heat up the battery before it can operate optimally. Thus, the electric miles that are obtained from the vehicle are less than the miles obtained in the spring/summer when it is not too cold (Ramadass, et al., 2002; Choi & Lim, 2002).

2nd data set

A second data set was created with aggregated and averaged data for 31 unique individuals over their latest 35 months (whenever - April 2016) of ownership of the Volt. There is an overlap of the first 10 points between the 1st and the 2nd data set. A similar analysis was performed as before to observe if the trend continues.

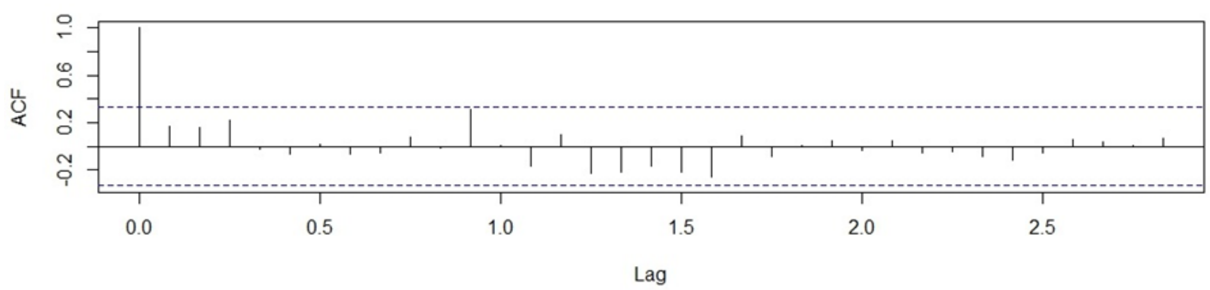


Figure 13: ACF for 2nd time series data

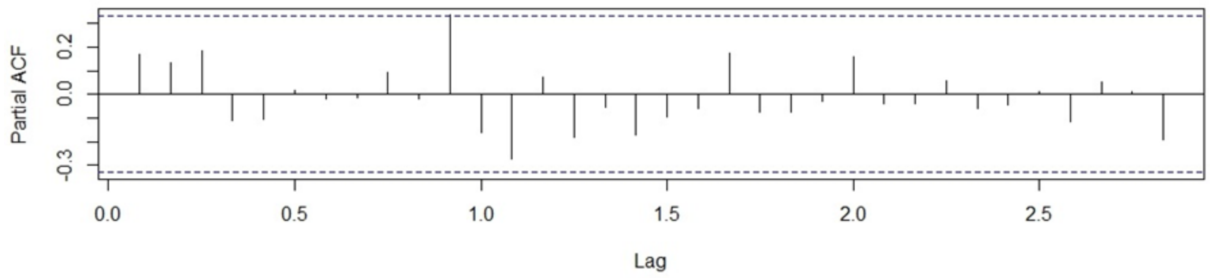


Figure 13: PACF for 2nd time series data

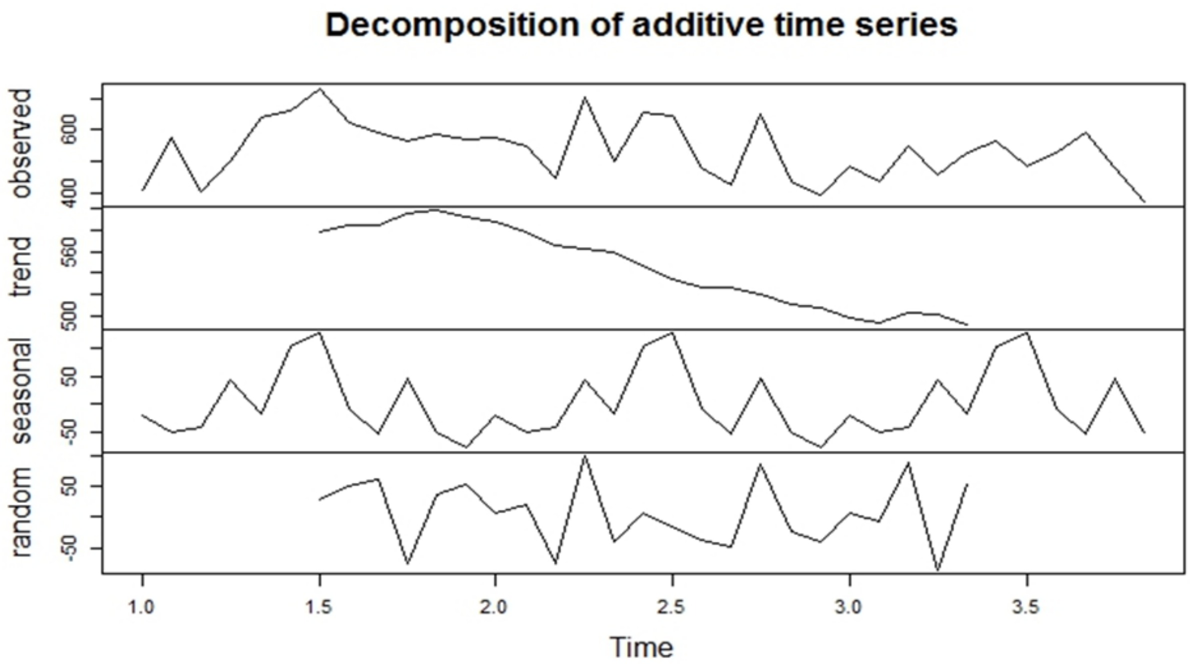


Figure 14: Decomposition of time series for 2nd data set

Looking at the PACF and ACF from figure 8 and 9 we cannot choose a model for the time series because they are very random and not representative of the actual data. A difference filtering with a lag of 1 is used to remove the trend component observed in Figure 7 since that is a requirement for the ARIMA model. The inherent trend could also cause distortion of the data. To find the orders of the ARIMA (p,d=1,q) model we try different values of p and q, and choose the combination which gives lowest AIC and BIC

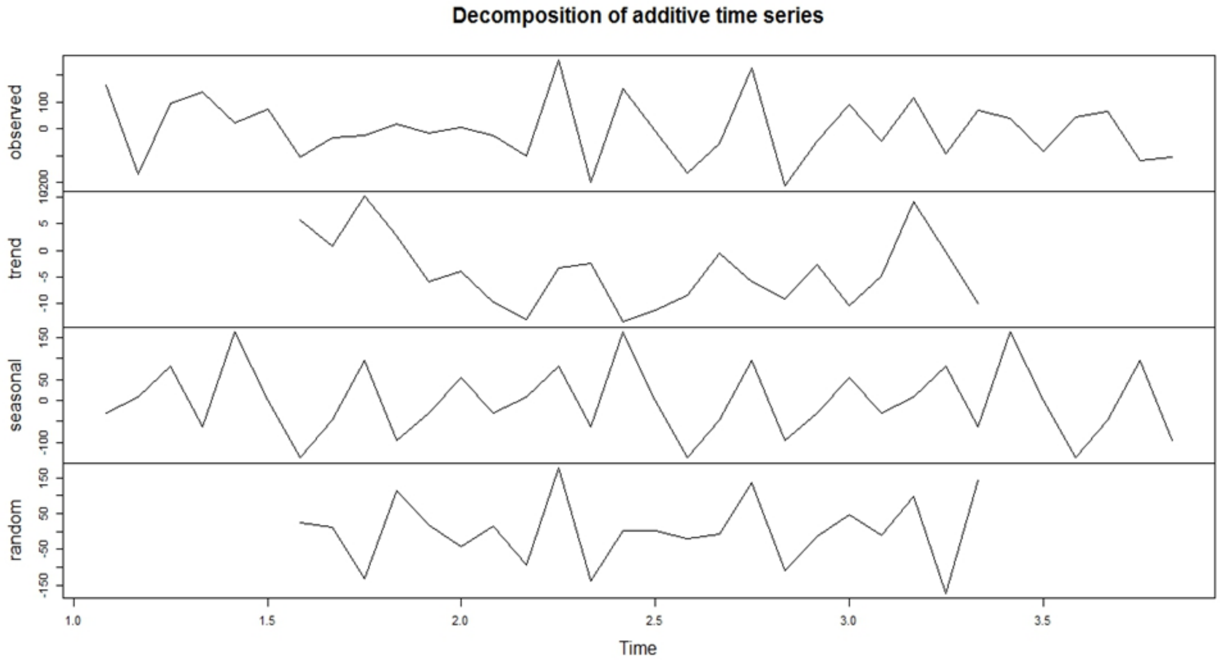


Figure 15: Decomposition of time series for 2nd data set with lag = 1

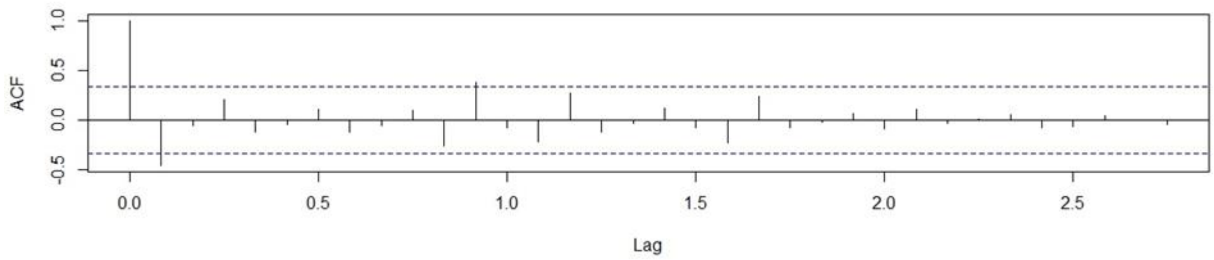


Figure 16: ACF for 2nd time series data with lag=1

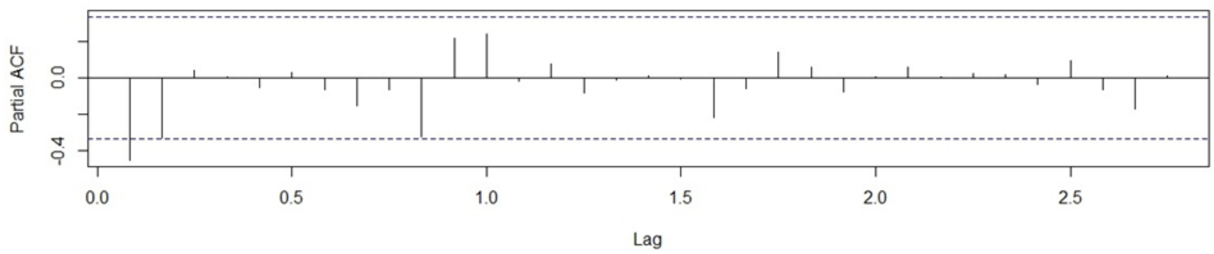


Figure 17: PACF for 2nd time series data with lag=1

From Figure 11 and Figure 12 we see that the ACF goes to zero exponentially and the PACF cuts off, we can choose a moving average model with p=0. We run an ARIMA model and perform a Ljung-Box test.

Null hypothesis: $\varphi_1 = \varphi_2 = \cdots \varphi_L = 0$ which implies that the model does not exhibit lack of fit

Alternate hypothesis: $\varphi_1 \neq \varphi_2 \neq \cdots \varphi_L \neq 0$ which implies that the model exhibits lack of fit

$X^2 = 1.954$, df = 1

$X^2_{critical}$ (97.5% confidence) = 5.023

We accept the null hypothesis which means that the residuals do not have any correlation and they are just random data or white noise. Thus the model does not exhibit lack of fit and is perfect.

Similar trends of learning and seasonal effects are seen in this data set just as in the first one.

Conclusion

In the model of learning, average cost reductions are expressed as a function of cumulated output, investment or time (Malerba, 1992). A learning model based on learning by doing is observed in our study. Constant repetition of the same task by the users enables them to gain experience and hence perform the task better. In this case it was driving the PHEV Chevy Volt and get the difference between total miles and electric miles as small as possible.

From the time series analysis on both the data sets it is seen that performance gains from the learning curve effect are observed in this case. The longer the car is owned, people actually drive more electric miles than gasoline miles thus proving our hypothesis that over time the amount of money spent on gasoline will reduce. There needs to be a more comprehensive data set to actually be able to attribute this reduction in amount of money spent on gas. But this can serve as a starting point for a more rigorous and extensive analysis into this concept.

References

1. Anderman, 2004. The challenge to fulfil electrical power requirements of advanced vehicles. J. Power Sources, 127(1-2), pp. 2-7.
2. Choi, S. S. & Lim, H. S., 2002. Factors that affect cycle-life and possible degradation mechanisms of a Li-ion cell based on LiCoO2. Journal of Power Sources, 111(1), pp. 130-136.
3. Emadi, A., Rajashekara, K., Williamson, S. S. & Luki, S. M., 2005. Topological overview of hybrid electric and fuel cell vehicular power system architectures and configurations. IEEE Trans. Veh. Technol, 54(3), pp. 763-770.
4. Energy, U. D. o., 2016. Maps and Data for Vehicles. [Online]
5. Available at: http://www.afdc.energy.gov/data/categories/vehicles--2
6. [Accessed May 2016].
7. Malerba, F., 1992. Learning by Firms and Incremental Technical Change. The Economic Journal, 102(413), pp. 845-859.
8. Quarterman Lee, P., 2014. Learning & Experience Curves In Manufacturing, Kansas City: Strategos, Inc. .

9. Ramadass, P., Haran, B., White, R. & Popov, B. N., 2002. Capacity fade of Sony 18650 cells cycled at elevated temperatures: Part I. Cycling performance. Journal of Power Sources, 112(2), pp. 606-613.
10. Raskin, A. & Shah, S., 2006. The Emerge of Hybrid Electric Vehicles, New York: Alliance Bernstein.
11. Schneider, K. P., Gerkensmeyer, C. E., Kintner-Meyer, M. C. W. & Fletcher, R., 2008. Impact assessment of plug-in hybrid vehicles on pacific northwest distribution systems. Proc. Power Energy Soc. Gen. Meet., pp. 1-6.
12. Wright, T., 1936. Factors affecting the cost of airplanes. Journal of Aeronautical Sciences, Volume 3, pp. 122-128.