Suplementary Material: Exploring EV Driver Charging Substitution Behavior

Aspen Fryberger Underwood*

September 8, 2023

Appendix A: Machine Learning Model

We worked with a variety of machine learning models to predict which Level 2 stations remain free. This was done due to concerns about underlying characteristics causing bias in which stations remained free. Ultimately, a random forest machine learning model [1] is used to predict which level 2 stations remain free in 2018 and which become not free using data station and 2017 usage data. The random forest was tuned to 11 decision trees and a .66-.33 train-test split.

To predict if a station became not free, I used the number of station visits in 2017, average charging sessions in 2017 (minutes), number of ports, unique drivers, visit frequency, station distance to downtown, station density within 1 mile, distance to nearest interstate, station business type: entertainment, grocery, apartment, hotel, medicine, other, parking garage, school, shopping, work, or industrial.

Table 1 displays the confusion matrix. While the accuracy score is quite high at 70.2%, the precisions score of 50% indicates a high false positive rate. This can be explained by the small total number of stations that remained free. Figure 1 shows the receiver operator characteristic curve for the random forest machine learning model showing only a small improvement over random assignment. It appears unlikely there is significant bias.

Table 1: Confusion Matrix

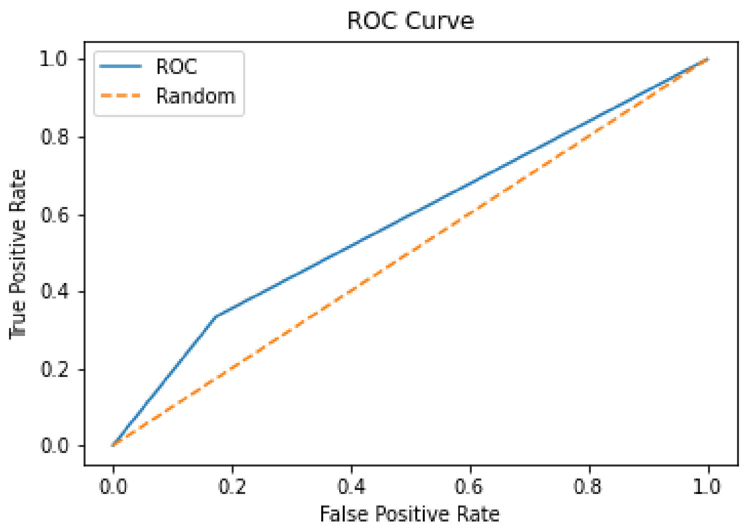
	Not Free	**Remained Free**
Predicted Not Free	53	6
Predicted Remained Free	19	6

Note: This table shows the Confusion Matrix from the Random Forest model. **Source:** Author's calculation.

*aspenfryberger@gmail.com, John E. Walker Department of Economics, Clemson University

[1] Implementation from sklearn `https://scikit-learn.org/stable/modules/ensemble.html#forest`

Figure 1: Random Forest ROC



Note: The Figure shows the receiver operating characteristic curve (ROC) from the random forest machine learning. ROC shows how well the model predicts which stations remain free and which station is not free. The closer to random the more random the prediction. **Source:** Author's calculation.

Appendix B: Data Description

This appendix provides additional description about how the data was cleaned and formulated for this paper. A charging station is defined as one or more ports at a single street address offering identical charging capability. Each station is one or more ports in a single location that offer identical charging capability. If a single charger exists in a parking lot it is a station; if five chargers are in a parking lot, this is also a single station. However, if there are both level 2 and level 3 chargers in the same location they are treated as different stations because they offer different charging capabilities, but this only occurs in two locations. Level 3 stations are excluded from this analysis.

In addition to station information, the data includes driver specifics which allow me to track the charging behavior of drivers across stations over time. Each driver is identified by a unique code. This allows me to observe changes in driver charging behavior at stations across the network and know what stations they frequent before and after the subsidy ends. The data also includes the home zip code for each driver, allowing for a general calculation of the distance between the driver's home zip code and each station. Although the number of drivers increased over 2017 and 2018, to reliably observe how behavior changes before and after the subsidy ends, I only include drivers who began using the network before May 1, 2017.

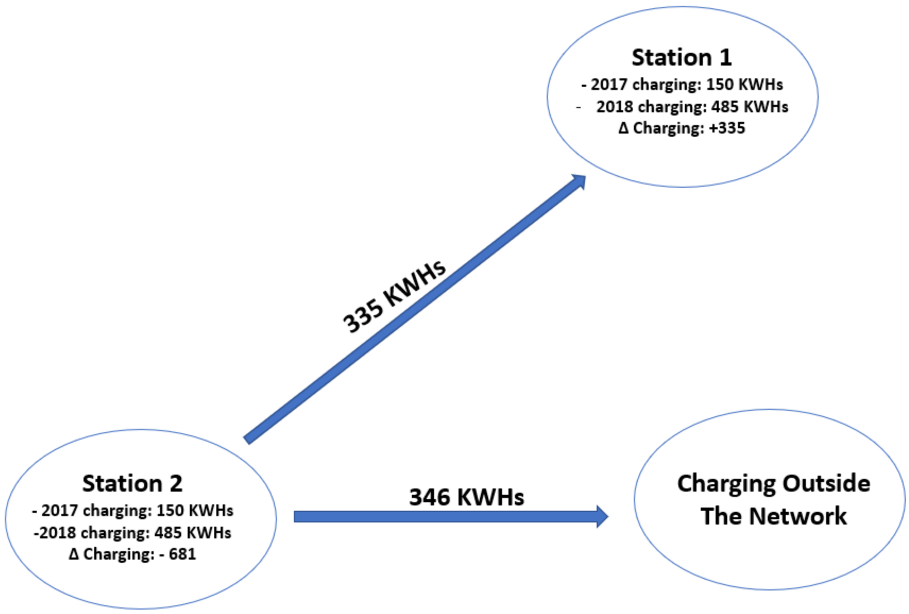
In addition to information on stations and drivers, the data includes specifics about each charging transaction. Every time a driver plugs their vehicle into a charger, the data records the starting and ending date and time of the event, station-specific identification, driver-specific identification, and the KWHs charged in the session. This allows me to observe how much each driver charges on the network, the unique number of stations they visit, and the frequency of visits for each driver. For stations it allows me to calculate the number of unique drivers, average length of charging time, and total usage in KWs over time.

Station and driver identifiers allow me to calculate how each driver's charging moves across stations by observing where drivers decrease and increase charging after the subsidy ends. The movement of charging across stations is calculated by observing the total amount of charging each driver charges at every station in 2017 and 2018. If a driver decreases charging at a station in 2018, then charging is leaving that station, but if the charging at a station increases, charging moving towards that station.

If charging decreases at a station then charging leaves an "origin" station and if the charging increases at a station, then that station is the "destination" station. If there are two destination stations the charge from the the origin goes to each based on the percentage increase at each destination station. For the purpose of this paper I am assuming that total charging does not change as a result of the price change so the total amount of charging that leaves the origin stations must equal the total charging going to destination stations. If more charging leaves stations on the network then increases at the destination on the network then charging has left the network and goes to the destination “outside.” It is reasonable to assume that the total amount of EV charging does not change between the two years because even after the subsidy ended, charging an EV is still cheaper than buying gasoline, and given the significant driver investment in an EV it is unlikely drivers would make a significant switch back to gasoline. Additionally, the cost of charging an EV is very low, and the average charging session costs around a dollar. It is somewhat unlikely that a price change this small would lead to significant changes in the amount someone drives.

An example of this is shown in Figure 2 where the charging for station 2 is less in 2018 than in 2017, but there is an increase in charging at station 1. However, because of the total decrease in charging at stations on the network is greater than the increase, some charging moves “outside” the network, and the charging coming from station 2 is proportionally allocated to station 1 and outside. This process allows for the analysis of how charging moves across stations.

Figure 2: Example of Driver Transfer of Charging



Note: This figure shows how movement across stations is calculated. **Source:** Author’s calculation.