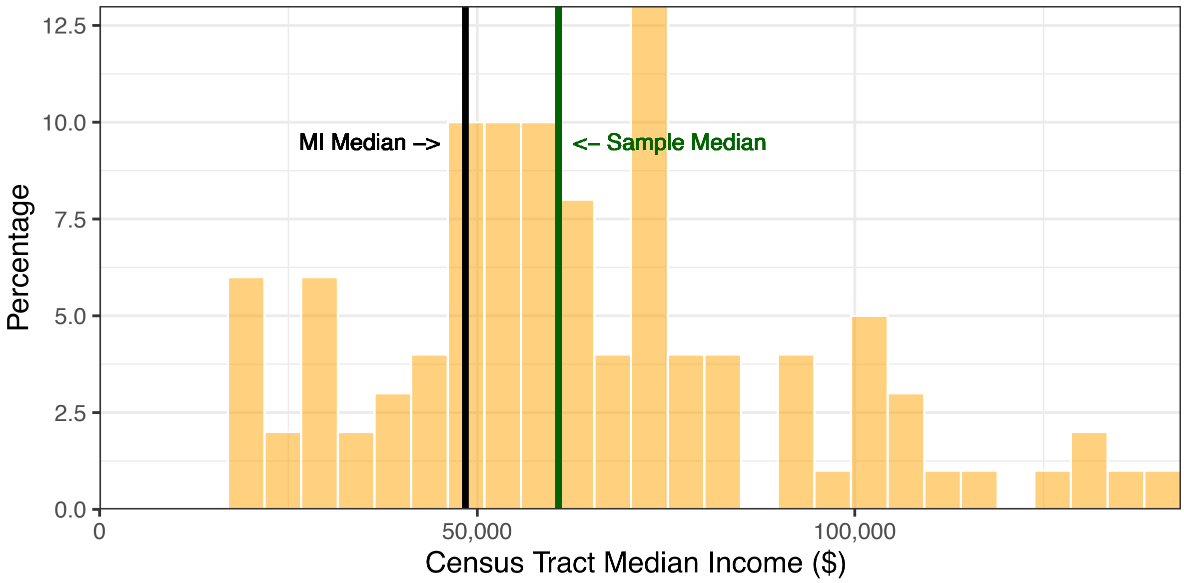
Online Appendix

Title: Fueling Alternatives: Gas Station Choice and the Implications for Electric Charging

Authors: Jackson Dorsey, Ashley Langer, and Shaun McRae

A Additional Tables & Figures Referenced in Main Paper

Figure A.1: Driver Income Distribution



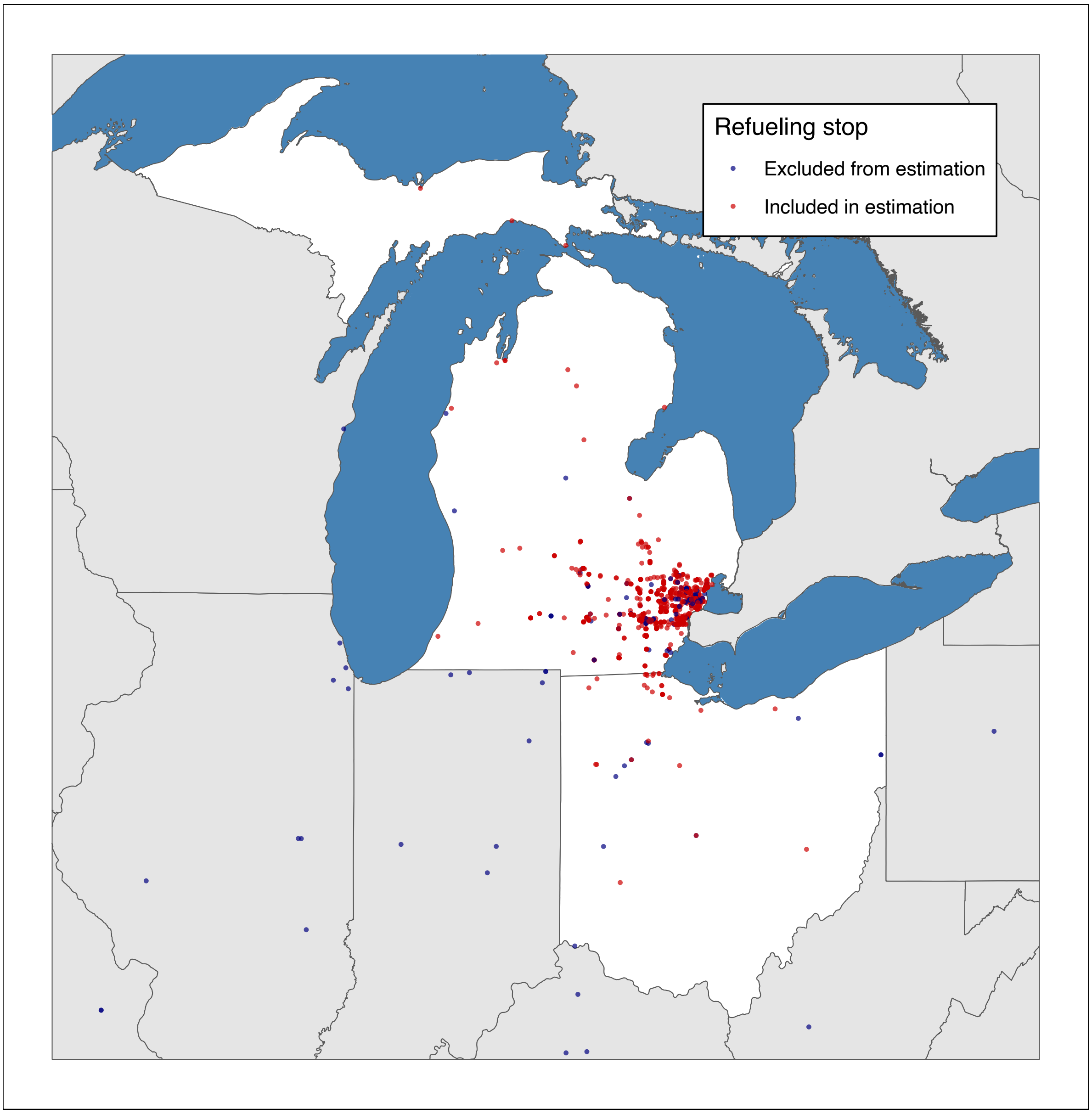
Notes: Distribution of drivers' income based on the Census tract in the 2010 American Community Survey.

Table A.1: Station Brand Choice Shares

Brand	Choice Share (%)
BP	17.69
Citgo	5.07
Costco	2.71
Marathon	12.38
Meijer	6.25
Mobil	8.61
Other	18.40
Shell	5.90
Speedway	14.86
Sunoco	8.14

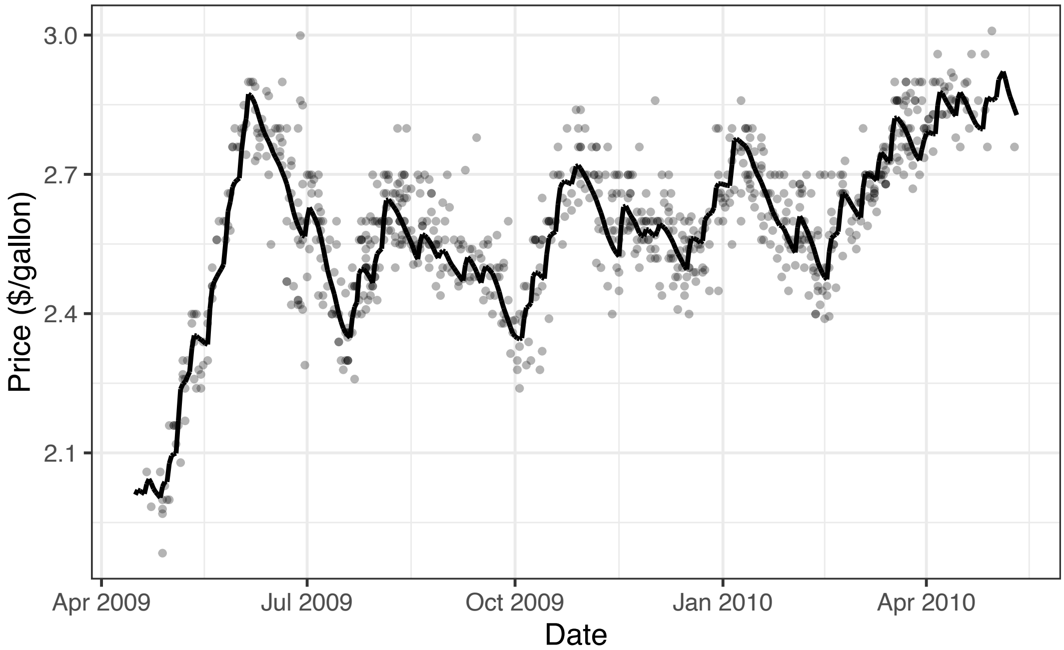
Notes: Empirical choice shares across all refueling stops made by all drivers during the experiment.

Figure A.2: Map of Observed Refueling Locations



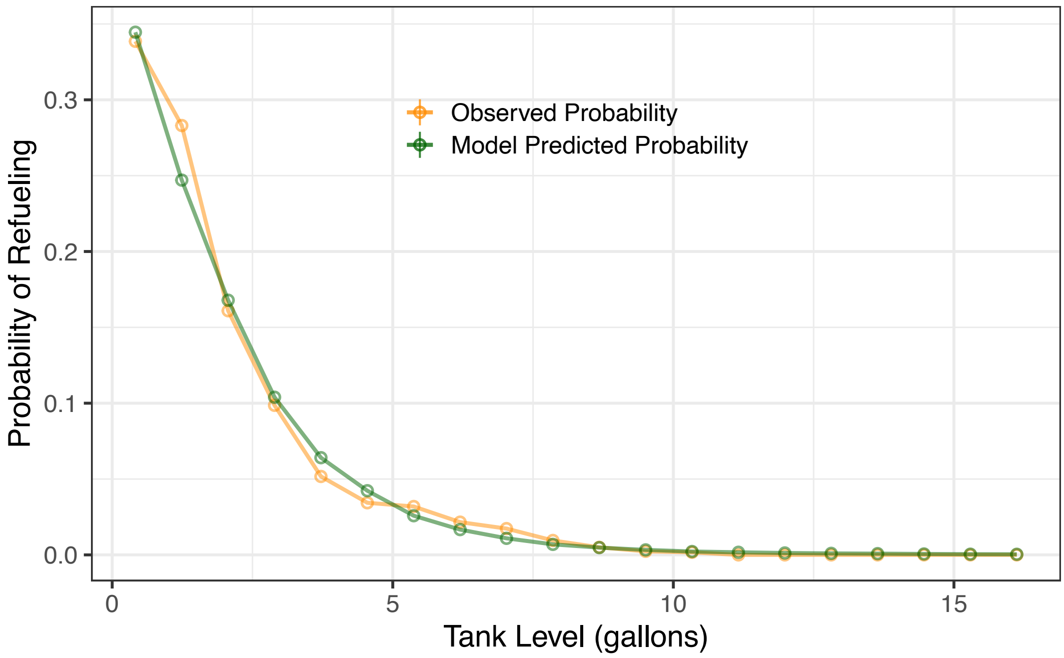
Notes: Each point is a refueling stop that we identified using vehicle locations and the in-car cameras. The red points are the refueling stops that are included in the estimation dataset. Reasons for excluding a stop from estimation include missing price information, missing location data at the start or end of the trip, and missing information on tank levels. Our OPIS price data only covers Michigan and Ohio, so all stops outside of those states are excluded from the estimation data. Background map data is from U.S. Census Bureau (2018), Esri (2019) and Government of Canada (2016).

Figure A.3: Average gasoline price and observed purchase price through sample period



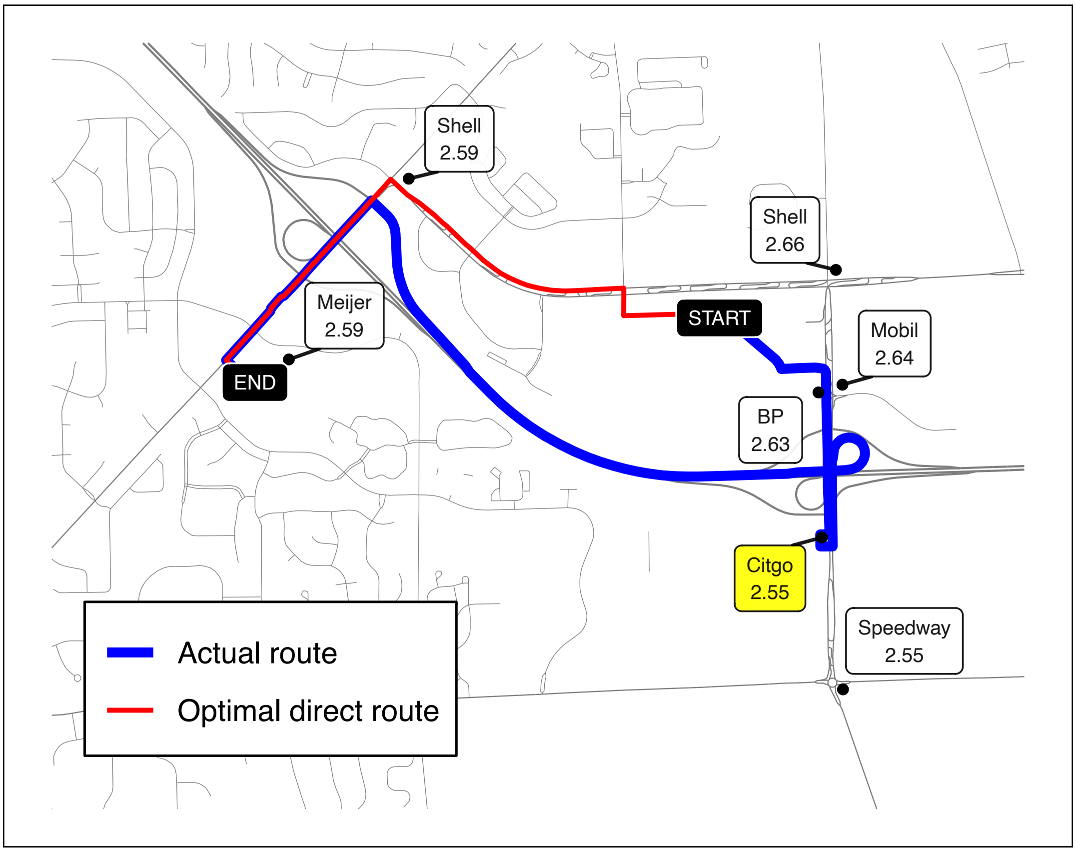
Notes: The black line shows the daily average gasoline price in MI and OH from the OPIS data. Each dot represents the date and price associated with one of the refueling stops that we identify. Posted prices exhibit greater variation within a period.

Figure A.4: Probability of Stopping by Tank Level Model Fit



Notes: The orange circles show the empirical probability of stopping by trip starting tank level in the data. The green circles show the model fit of these probabilities using our baseline specification.

Figure A.5: Example Trip Route including Nearby Refueling Options



Notes: The red line shows the optimal direct route from origin to destination for a trip taken by a driver in Ann Arbor, MI. The blue line shows the route taken by the driver to stop at the Citgo station. Each point shows the location of gas stations near the driver's route along with the prices on the day of the trip.

Table A.2: Comparison of Non-Passed and Passed Stations

	Not Passed (N=121790)		Passed (N=52601)			
	Mean	Std. Dev.	Mean	Std. Dev.	Diff. in Means	Std. Error
Excess Time (minutes)	11.37	5.18	8.50	5.82	-2.87	0.03
Current Price ($/gallon)	2.59	0.17	2.62	0.15	0.03	0.00

Notes: The table compares prices and excess time for all previously passed and not previously passed stations. The sample includes all potential stations that are within a 20 minute deviation from a trip.

Table A.3: Driver Preference Estimates - Impact of Station Brand

	Full Price Information				Imperfect Price Information			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Value of Not Stopping								
1[No Stop] $\times$ Constant	-5.862	-4.856	-6.394	-5.406	-12.302	-10.792	-12.472	-12.472
	(0.703)	(0.802)	(0.746)	(0.860)	(1.219)	(1.463)	(1.625)	(1.625)
1[No Stop] $\times$ Tank Level	0.976	0.932	1.102	1.037	1.163	1.102	1.325	1.325
	(0.057)	(0.058)	(0.065)	(0.068)	(0.065)	(0.069)	(0.090)	(0.090)
1[No Stop] $\times$ (Tank Level)^2	-0.002	-0.005	-0.011	-0.013	0.008	0.006	-0.008	-0.008
	(0.005)	(0.005)	(0.006)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)
Station Choice								
α - Expected Expenditure ($)	-0.323	-0.243	-0.335	-0.251	-0.574	-0.464	-0.511	-0.511
	(0.031)	(0.033)	(0.033)	(0.035)	(0.051)	(0.057)	(0.062)	(0.062)
θ - Weight on Current Price					0.336	0.355	0.313	0.313
					(0.060)	(0.081)	(0.076)	(0.076)
γ - Excess Time (minutes)	-0.192	-0.255	-0.139	-0.206	-0.154	-0.227	-0.168	-0.168
	(0.018)	(0.019)	(0.017)	(0.019)	(0.016)	(0.018)	(0.018)	(0.018)
Nesting Parameter								
λ	0.507	0.570	0.480	0.557	0.518	0.568	0.557	0.557
	(0.035)	(0.036)	(0.041)	(0.042)	(0.034)	(0.035)	(0.041)	(0.041)
Choice Set	All	All	Passed	Passed	All	All	Passed	Passed
Station Brand Fixed Effects	N	Y	N	Y	N	Y	N	Y
Number of Stops	848	848	771	771	848	848	771	771
Number of Trips	19825	19825	19588	19588	19825	19825	19588	19588
Observations	2285082	2285082	834208	834208	2285082	2285082	834208	834208
McFadden R^2	0.299	0.312	0.269	0.282	0.303	0.410	0.274	0.285
Implied Value of Time ($/hour)								
	35.65	63.04	24.83	49.11	16.13	29.33	9.51	19.73
	(4.92)	(10.38)	(3.99)	(8.74)	(2.26)	(4.58)	(1.80)	(3.45)

Notes: Table reports pseudo maximum likelihood estimates of driver preferences. The expected purchase quantities are predicted from the regressions in Table 4. The full information models assume drivers know current gas prices at each station and the imperfect information models allow drivers' price perception to be a weighted average of current price and station average price. λ is the nested logit correlation parameter. In models with brand fixed effects, the constant in the decision to stop represents small brand and unbranded stations. Choice Set = "All" indicates that all stations within 20 minutes of the driver's route are included in the choice set. Choice Set = "Passed" means stations that the driver has previously passed that are within 20 minutes of the route are included in the choice set. MLE standard errors are reported in parentheses. The implied value of time (per hour) is calculated as $60 \cdot \frac{\gamma}{\alpha}$.

Table A.4: Purchase Quantity Regression - Evidence for Exclusion of Current Prices

	Purchase Quantity (gallons)				
	(1)	(2)	(3)	(4)	(5)
Tank Level	-0.429	-0.449	-0.415	-0.423	0.212
	(0.203)	(0.203)	(0.200)	(0.201)	(0.157)
$(\text{Tank Level})^2$	-0.002	-0.0009	-0.008	-0.007	-0.064
	(0.023)	(0.023)	(0.022)	(0.022)	(0.018)
Current Price ($/gallon)		-1.50		-0.849	-1.02
		(0.785)		(1.19)	(0.914)
Driver's Mean Purchase (gallons)					0.978
					(0.033)
Observations	751	751	751	751	751
R^2	0.121	0.125	0.185	0.186	0.563
Choice Set	Passed	Passed	Passed	Passed	Passed
Station Brand FEs	Y	Y	Y	Y	Y
Excess Time Control	Y	Y	Y	Y	Y
Month-Year FEs	N	N	Y	Y	Y

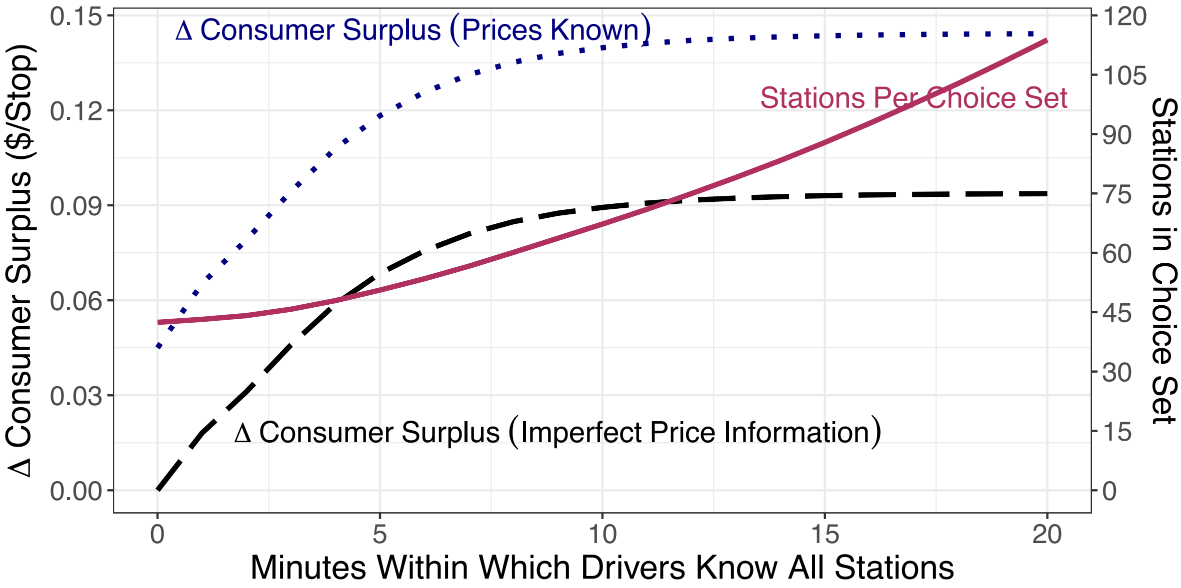
Notes: The dependent variable is the imputed purchase quantity associated with each observed refueling stop. The regressions estimates are used to predict expected purchase quantity conditional on stopping for each trip conditional on initial tank level (gallons). The regression in Column (3) is used to predict expected purchase quantity for the baseline model in Table 5.

Table A.5: Driver Access to Gasoline Station Network vs. Electric Charging Station Network

	Gas (All)	Gas (Passed)	Electric (All)
Stations within 5 Minutes of Route	20.99	12.62	8.19
Closest Station (minutes)	0.99	1.12	4.10

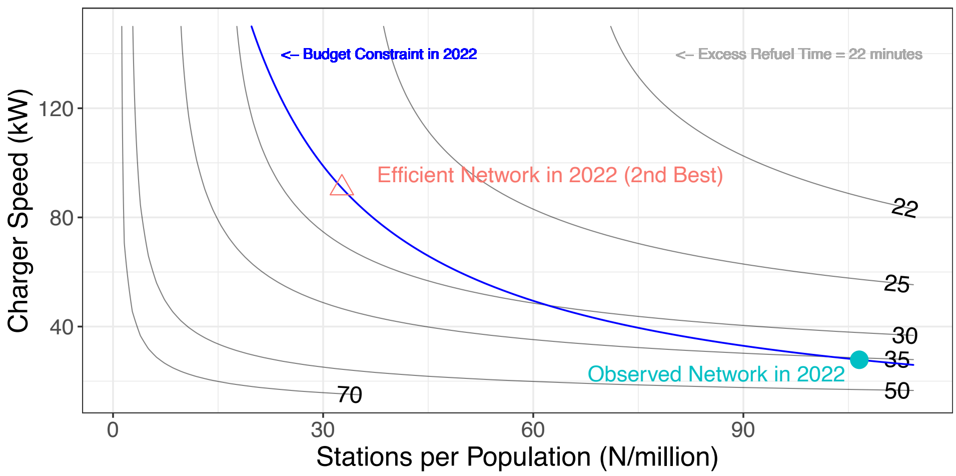
Notes: Column (1) includes all gas stations in OPIS and Column (2) only counts the stations that each driver has previously passed. Column (3) summarizes all public electric chargers in 2022 based on DOE data. The first row reports the average number of stations located within five minutes deviation from a driver's route, across all trips in our data. The second row shows that average time to the closest station relative to driver's routes across all trips in the data.

Figure A.6: Consumer Surplus Gains from Adding Stations to Choice Set



Notes: The solid maroon line plots the average number of stations available to drivers if all the previously unpassed stations within X minutes of the drivers' optimal route were added to the choice set. The dashed black line shows the change in consumer surplus in dollars per gas stop from adding additional (unpassed) stations to the choice set, assuming that drivers remain imperfectly informed about current gas prices, relative to the baseline case where only passed stations are in the choice set. The dotted blue line shows the change in consumer surplus in dollars per gas stop if drivers were perfectly informed about current prices and additional unpassed stations are added to the choice set, relative to the baseline case with only passed stations in the choice set and imperfect information about current prices.

Figure A.7: Excess Refuel Time Contour Map as a Function of Station Density and Charger Speed with Budget Constraint



Notes: The thin grey lines show contours representing the estimated excess refueling time per EV refueling stop across different counterfactual combinations of station density (number of stations) and charging speed of the network (kW). The excess refueling time for each counterfactual network configuration is determined by the location of charging stations relative to drivers routes and behavioral assumptions that are described in Section 4. The blue line plots the available budget set based on the number of stations and average charge speed of the observed 2022 network.

B Data Construction

In this section, we describe the procedure for constructing our dataset of gasoline refueling stops, including the station location and purchase quantity. This procedure relied on the following multimodal data from the IVBSS experiment (Sayer et al., 2010):

- second-by-second latitude and longitude of the vehicle
- second-by-second gasoline consumption of the vehicle
- left-side camera video feed
- over-the-shoulder camera video feed showing the dashboard (from which we extracted images at a five-minute frequency)

Unfortunately, the onboard computers used in the IVBSS experiment did not record the fuel tank level, so we were not able to calculate the fuel purchase quantities directly.

We combined the above IVBSS data with the latitude and longitude of the gas stations in Michigan and Ohio (Oil Price Information Service, 2012). For all of the stations in southeast Michigan, we confirmed the location of each station using aerial photographs from an online mapping service. We adjusted these locations when necessary to correspond to the exact latitude and longitude of the gas pumps at each station.

Our procedure for constructing the refueling stop dataset is as follows:

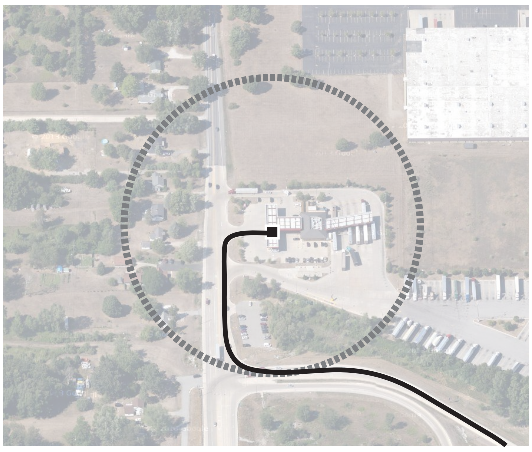
1. We coded every vehicle stop within a 200-meter radius of a gas pump as a potential refueling stop (left panel of Figure B.1). Because some of these stops may have been for reasons other than purchasing gasoline, we then checked the left-side video images during all of these potential stops (right panel). If the camera showed that the vehicle was stopped beside a gas pump, the stop was coded as a refueling stop.

2. We extracted a sample of cabin photos from the over-the-shoulder video feed at a five-minute frequency. From the cabin photos, we cropped a 30x25 pixel rectangle at the fuel gauge location on the dashboard. Based on an analysis of the pixels in this rectangle, we classified the image into one of four types: good, underexposed, overexposed, and low contrast. The image was rescaled and smoothed.

3. We then uploaded the "good" gauge photos to Amazon S3 for classification by Mechanical Turk workers (Dorsey et al., 2013). The Mechanical Turk workers completed a series of tasks. Each task consisted of the classification of three fuel gauge photos. The gauge level classifications ranged from 0 (empty) to 8 (full), with a value of 9 corresponding to an illegible gauge level. At least three workers classified each photo.

4. We calculated the gauge level associated with each five-minute interval (*i.e.* each photo) as the mean gauge classification (across the Amazon Turk workers) between 0 and 8.

Figure B.1: Procedure for identifying gas station stops



Notes: All vehicle stops within a 200-meter radius of gasoline station pumps were considered as possible refueling stops (left image). Images from the driver's side camera were used to confirm that the car was stopped at a gas pump (right image).

5. We combined the five-minute observations of the fuel gauge level with the second-by-second data on gasoline consumption in milliliters. We used the gasoline consumption data to calculate the cumulative fuel consumption between each trip start-point, end-point, refueling event, or five-minute fuel gauge image.

6. To recover the fuel tank levels, we estimated a regression of the cumulative fuel consumption on (i) a quadratic in the fuel gauge level between 0 and 8, and (ii) driver-by-refueling-event fixed effects. From this regression, the driver-by-refueling-event fixed effects correspond to the initial fuel tank level at the start of each refueling event.

7. We combined the initial fuel tank level after each refueling event with the incremental fuel consumption data to calculate the fuel tank level at each second in the data.

8. For some trips and drivers, the imputed fuel tank levels were physically impossible: either greater than the tank capacity of 17.2 gallons, or less than zero. This necessarily implied that we had missed a refueling stop in step (1) of our procedure. In these cases, we used a combination of the fuel gauge photos and the latitude and longitude data to search for the time and location of the missing refueling stop. These stops may not have been identified in (1) because: (i) the stop was outside of Michigan and Ohio, (ii) the station was outside of southeast Michigan and had incorrect coordinates in the OPIS data, (iii) there was a GPS fault in the vehicle and the location was not recorded before and after the refueling stops, (iv) the refueling occurred at a station that was missing from the OPIS data, or (v) the vehicle was left running while refueling so the stop did not occur at the end of a trip. For those cases in which we were not able to identify the exact station where the driver stopped, we still coded the stop as a refueling event to be

able to recover the tank levels. However, without the gasoline prices, we were unable to include these stops in our estimation.

9. We inferred the fuel purchase quantity associated with each refueling event as the change in fuel tank level (in gallons) before and after a refueling event. For imputed purchase quantities of less than two gallons, we went back to the fuel gauge images and confirmed that there was a visible increase in the gauge level after the refueling stop. We deleted potential refueling events with no visible gauge increase. These were likely due to a driver parking beside a gas pump, entering the shop or bathroom, but not purchasing gasoline.

10. We compared the imputed tank levels to the mean gauge level classifications from the Mechanical Turk workers in (4). Where there was a large discrepancy between the two values, we reviewed the fuel gauge images and, in many cases, manually corrected the Mechanical Turk reports.

11. After adding or deleting refueling stops as discussed in (8) and (9), and correcting the reported fuel gauge levels in (10), we repeated the above steps from (5) to (10). We continued this process until there were no imputed fuel tank levels above 17.2 or below 0 gallons, and all refueling stops had a visible increase in the fuel gauge level. This process gave us our final dataset of refueling stops and purchase quantities.

C Details on Station Choice Model

This appendix section provides details on our empirical model. Section C.1 goes through the identification arguments for our model in detail. Section C.2 provides the form of the quasi-maximum likelihood estimator. Section C.3 explains how we calculate the value of information from our model.

C.1 Identification

Identification of our model rests upon a series of assumptions that we outline here. First, recall that each driver forms expectations about the price that they will pay at each station and the quantity of gasoline that they will purchase at the start of each trip. Using this information, drivers choose whether to stop and at which station to stop. We express prices and quantities as the sum of drivers' expectations and prediction errors:

$$p_{ikt} \equiv \mathbb{E}_i[p_{ikt}|Z_{ikt}] + \eta^p_{ikt} \tag{C.1}$$

$$q_{ikt} \equiv \mathbb{E}_i[q_{ikt}|Z_{ikt}] + \eta^q_{ikt}. \tag{C.2}$$

In this formulation, Z_{ikt} are observable characteristics of the driver, trip, and station, such as the driver's tank level at the start of the trip, the long-run average price at each station, and the station brand. We take $\mathbb{E}_i[p_{ikt}|Z_{ikt}]$ and $\mathbb{E}_i[q_{ikt}|Z_{ikt}]$ to be the driver i's (subjective) expected purchase price and quantity given the information they have at the start of the trip. Further, η^p_{ikt} and η^q_{ikt} represent mean-zero unobserved shocks to prices and quantities, respectively, that are not explained by Z_{ikt}.

Given these definitions, the expected fuel expenditure at station k on trip t conditional on variables the driver uses at the start of the trip to form expectations is:

$$\mathbb{E}_i\Big[(\mathbb{E}_i[p_{ikt}|Z_{ikt}] + \eta^p_{ikt}) \cdot (\mathbb{E}_i[q_{ikt}|Z_{ikt}] + \eta^q_{ikt})\Big|Z_{ikt}\Big]. \tag{C.3}$$

Plugging this function into the utility function from Equation (1) in the text, we get:

$$\begin{aligned} U_{ikt} = \alpha\mathbb{E}_i\Big[(\mathbb{E}_i[p_{ikt}|Z_{ikt}] + \eta^p_{ikt}) \cdot (\mathbb{E}_i[q_{ikt}|Z_{ikt}] + \eta^q_{ikt})\Big|Z_{ikt}\Big] \\ +\gamma \text{Excess Time}_{ikt} + X_k'\beta + \varepsilon_{ikt}, \end{aligned} \tag{C.4}$$

where, as in the main text, X_k is a vector of characteristics of the station k such as corporate brand. Also, recall, that the value of not stopping on trip is given by:

$$U_{i0t} = W_{it}'\delta + \varepsilon_{i0t}, \tag{C.5}$$

where W_{it} includes variables—such as tank level—that impact the value of not stopping to refuel for driver i on trip t.

To identify the model, we begin by making an assumption about the relationship between the shocks to purchase price and quantity, η^p_{ikt} and η^q_{ikt}, and the driver's expected purchase price and expected quantity conditional on Z_{ikt}:

Assumption 1 *Unexpected shocks to the quantity of gasoline that a driver purchases at any given station, η^q_{ikt}, are independent of the driver's expected purchase price, $\mathbb{E}_i[p_{ikt}|Z_{ikt}]$, and the shocks to the driver's expected purchase price, η^p_{ikt}, given the information the driver has at the start of the trip, Z_{ikt}:*

$$\begin{aligned} \mathbb{E}_i[p_{ikt}|Z_{ikt}] &\perp\!\!\!\perp \eta^q_{ikt}|Z_{ikt}, \\ \textit{and } \eta^p_{ikt} &\perp\!\!\!\perp \eta^q_{ikt}|Z_{ikt}. \end{aligned}$$

Similarly, unexpected shocks to the prices at any given station, η^p_{ikt} are independent of the driver's expected purchase quantity, $\mathbb{E}_i[q_{ikt}|Z_{ikt}]$ given the information the driver has at the start of the trip, Z_{ikt}

$$\mathbb{E}_i[q_{ikt}|Z_{ikt}] \perp\!\!\!\perp \eta^p_{ikt}|Z_{ikt}.$$

This assumption ensures that the expectations over the gasoline price and purchase quantity that the driver has at the start of their trip are uncorrelated with the shocks that they receive to both purchase price and quantity when they arrive at the station to make the purchase (and that those shocks are also uncorrelated with each other). In practice, the strongest component of this assumption is that that drivers' fuel purchases are inelastic with respect to the price shock, η^p_{ikt}, conditional on Z_{ikt}. One interpretation of this is that drivers decide on their fuel purchase quantity before observing the actual price at the station. This is a weaker version of the assumption in Hastings and Shapiro (2013) that drivers are completely price inelastic when purchasing gasoline, since we do allow the driver's purchase quantity to respond to variables like the station brand, the quantity of fuel in their tank, and month-year fixed effects via the conditioning on Z_{ikt}. We see in Columns (2), (4), and (5) of Appendix Table A.4 that the station's current price is not a statistically significant predictor of observed purchase quantities conditional on the variables we assume the driver uses to predict purchase quantities. This lends support to the assumption that the η^p_{ikt} that the driver observes when they reach the station does not substantially affect the quantity of fuel purchased.

Assumption 1 allows us to rewrite the driver's expected gas expenditure at station k on trip t as:

$$\underbrace{\mathbb{E}_i[p_{ikt} \cdot q_{ikt}|Z_{ikt}]}_{\text{Expected Fuel Expenditure}} = \underbrace{\mathbb{E}_i[p_{ikt}|Z_{ikt}]}_{\text{Expected Price}} \cdot \underbrace{\mathbb{E}_i[q_{ikt}|Z_{ikt}]}_{\substack{\text{Expected Purchase} \\ \text{Quantity}}}, \tag{C.6}$$

which we can plug back into the utility function, Equation C.4 to write utility as:

$$U_{ikt} = \alpha \mathbb{E}_i[p_{ikt}|Z_{ikt}] \cdot \mathbb{E}_i[q_{ikt}|Z_{ikt}] + \gamma \text{Excess Time}_{ikt} + X_k'\beta + \varepsilon_{ikt}. \tag{C.7}$$

The above equation makes clear that our empirical approach will require variation in the price drivers expect to pay independent of the quantity they expect to purchase. This requires an additional assumption. To explain the assumption, we first define $Z^p_{ikt} \subseteq Z_{ikt}$ and $Z^q_{ikt} \subseteq Z_{ikt}$, which are the information that the driver uses to form expectations over purchase price and quantity, respectively. Using this notation, we impose the following assumption:

Assumption 2 *The information the driver uses to form expectations over the purchase quantity at the beginning of a trip, Z^q_{ikt}, is not identical to the information the driver uses to form expectations over the purchase price, Z^p_{ikt}. Further, p_{ikt} is independent of Z^q_{ikt} given Z^p_{ikt} and q_{ikt} is independent of Z^p_{ikt} given Z^q_{ikt},*

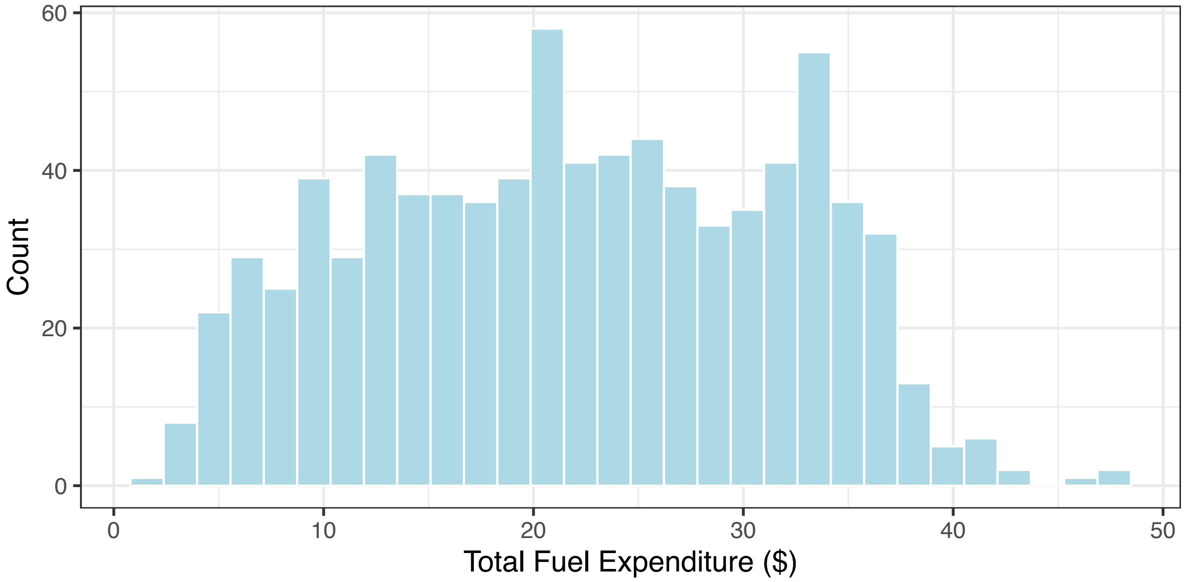
$$\begin{gathered} Z^p_{ikt} \subseteq Z_{ikt} \textit{ and } Z^q_{ikt} \subseteq Z_{ikt} \textit{ such that } Z^p_{ikt} \neq Z^q_{ikt}, \\ \textit{and } \mathbb{E}_i[p_{ikt}|Z^p_{ikt}] = \mathbb{E}_i[p_{ikt}|Z_{ikt}], \textit{ and } \mathbb{E}_i[q_{ikt}|Z^q_{ikt}] = \mathbb{E}_i[q_{ikt}|Z_{ikt}]. \end{gathered}$$

Assumption 2 is an exclusion restriction: it must be the case that there are some variables that are associated with the expected purchase quantity or the expected purchase price, but not both. In addition, the assumption implies that drivers' expectations of purchase prices are only based on the information in Z^p_{ikt} and not the other information in Z_{ikt}, and their expectations of purchase price are based only on the information in Z^q_{ikt} and not other information in Z_{ikt}.

In theory, we could allow, for instance, for Z^p to be a proper subset of Z^q so that all variables in Z^p are included in Z^q. However, we found that there was not sufficient variation in the data to precisely estimate the model parameters under this formulation. Therefore, we specify Z^p_{ikt} and Z^q_{ikt} as completely distinct sets of variables. Specifically, we specify that stations' current prices and stations' long-run average prices affect the price the driver expects to pay but not the quantity of fuel that the driver expects to purchase, conditional on Z^q_{ikt}. Here we are again relying on the evidence in Table A.4 that prices do not significantly affect expected purchase quantities conditional on the driver's tank level, the station's brand, ExcessTime$_{ikt}$, and the month-year fixed effects. Relatedly, we assume that the driver's tank level, the station's brand, ExcessTime$_{ikt}$, and the month-year fixed effects are not affecting the price the driver expects to pay, conditional on the station's long-run average price and the station's current station price.

Another concern might be that drivers' expectations about their purchase quantity could be correlated with the actual price at the station if drivers budget a fixed dollar amount for gasoline (*e.g.* \$20). In this case, the expected purchase quantity would be mechanically related to the realized price at the station where the driver chooses to refuel in violation of Assumptions 1 and 2. To provide evidence that this type of budgeting is not an important determinant

Figure C.1: Distribution of Total Fuel Expenditure



Notes: The graphic shows a histogram of total fuel expenditure in dollars across all refueling stops in our sample. Total expenditure is calculated as the current price (in $ per gallon) at the driver's selected station multiplied by the implied purchase quantity (gallons) for the trip. Our procedure for recovering purchase quantities is described in the appendix.

of purchase quantities, Figure C.1 plots a histogram of the implied fuel expenditure across all refueling stops and shows that there is only weak evidence of a discrete jump in expenditure at any specific dollar amount.[46]

Assumption 2 allows us to rewrite Equation (C.7) as:

$$U_{ikt} = \alpha \mathbb{E}_i[p_{ikt}|Z^p_{ikt}] \cdot \mathbb{E}_i[q_{ikt}|Z^q_{ikt}] + \gamma \text{Excess Time}_{ikt} + X'_k\beta + \varepsilon_{ikt}. \tag{C.8}$$

This equation illustrates how our estimation approach, where we estimate $\mathbb{E}_i[q_{ikt}|Z^q_{ikt}] = (Z^q_{ikt})'\phi$ in a first-stage, builds off of Assumptions 1 and 2.

The last step to identify α is to make the standard identification assumption that the expected expenditures conditional on Z^p_{ikt} and Z^q_{ikt} are not correlated with the unobservable quality of the station. We formalize the final identification assumption as follows:

Assumption 3 *The unobservable match quality shocks, ε_{ikt} are independent of the expected purchase quantity and price and the shocks to purchase quantity and price, conditional on the observable characteristics X_k, and W_{it}*

$$\begin{aligned} & \mathbb{E}_i[p_{ikt}|Z^p_{ikt}], \eta^p_{ikt} \perp\!\!\!\perp \varepsilon_{ikt}|X_k, W_{it} \ \ \forall i \\ \textit{and} \quad & \mathbb{E}_i[q_{ikt}|Z^p_{ikt}], \eta^q_{ikt} \perp\!\!\!\perp \varepsilon_{ikt}|X_k, W_{it} \ \ \forall i. \end{aligned}$$

[46]There is a small increase in the likelihood of expenditure around $20, but this increase would imply that only very small percentage of refueling stops are affected by this type of budgeting. There are also similarly-sized increases in expenditures at other, non-round expenditures.

There are two potential concerns with this assumption. The first is that stations with high unobservable quality, ε_{ikt}, also have higher prices. This is the most common identification issue in many discrete logit settings (e.g. Berry et al., 1995). Our inclusion of station brand fixed effects in utility helps increase the plausibility of this assumption. However, it is still possible that this specification does not fully control for variation in station quality within a given brand. In theory we could include station fixed effects to control for station-level quality, but we do not have enough drivers or trips in practice to identify these station fixed effects. Table C.1 provides further evidence in support of this assumption by showing that our results are remarkably robust to controlling for characteristics of the neighborhood where each station is located, including census tract median income and population density.

The second concern is that anything that increases the driver's expected expenditure, $\mathbb{E}_i[p_{ikt}|Z^p_{ikt}] \cdot \mathbb{E}_i[q_{ikt}|Z^q_{ikt}]$ via the expected purchase quantity, $\mathbb{E}_i[q_{ikt}|Z_{ikt}]$, will also likely increase the driver's utility of stopping on this trip by increasing the amount of fuel that the driver will have in the tank after the stop.[47] For example, when a driver's tank level is close to empty, their expected purchase quantity is likely to be high which will tend to increase utility. On the other hand, expected expenditure will also be higher which will tend to decrease utility. To help alleviate this concern, we include every variable in Z^q_{ikt} as controls in the utility function (in either X_k or W_{it}) so that they are not captured by the ε_{ikt}'s. Assumption 3 further rules out, for instance, that a driver will form particularly high expectations of purchase quantity (and therefore expected expenditure) at a station that they prefer, thereby biasing the estimate of α.

Finally, our model imposes an implicit assumption that drivers are not engaging in a sequential search for gas stations. This assumption is supported by the fact that gas prices that drivers observed recently are uncorrelated with the decision to stop to refuel, conditional on tank level. A sequential model of search would have drivers stopping more often if they have recently seen prices that are above the current price (De los Santos et al., 2012; Diamond, 1971).

C.2 Estimation

We estimate the utility model using the two-step estimator discussed in the main text. We estimate the first step regression of purchase quantities on Z^q_{ikt} via ordinary least squares. We then insert the estimated coefficients, $\widehat{\phi}$ in a nested logit quasi likelihood function that we estimate in a second step. Note that there are only two nests in our model, one for all stations, and a second for the choice not to stop at a station on this trip. We label the nesting parameter for the station nest λ_k and the nesting parameter for the choice not to stop is λ_0, which we

[47] In particular, $\mathbb{E}_i[q_{ikt}|Z_{ikt}]$ may be correlated with the value of not stopping, $U_{i0t} = W'_{it}\delta + \varepsilon_{i0t}$.

Table C.1: Driver Preference Estimates - Station Neighborhood Controls

	(1)	(2)	(3)	(4)
Value of Not Stopping				
1[No Stop] $\times$ Constant	-12.472	-12.646	-14.673	-15.393
	(1.625)	(1.635)	(1.881)	(1.961)
1[No Stop] $\times$ Tank Level	1.325	1.347	1.388	1.42
	(0.090)	(0.091)	(0.096)	(0.099)
1[No Stop] $\times$ (Tank Level)^2	-0.008	-0.007	-0.007	-0.009
	(0.006)	(0.006)	(0.006)	(0.006)
Station Choice				
α - Expected Expenditure ($)	-0.511	-0.542	-0.569	-0.583
	(0.062)	(0.065)	(0.070)	(0.071)
θ - Weight on Current Price	0.313	0.312	0.285	0.286
	(0.076)	(0.072)	(0.072)	(0.070)
γ - Excess Time (minutes)	-0.168	-0.171	-0.197	-0.191
	(0.018)	(0.018)	(0.018)	(0.018)
Station Census Tract Median Income ($)		0.096		-0.052
		(0.014)		(0.015)
Station Census Tract Population Density (Pop/Mile^2)			-0.473	-0.505
			(0.054)	(0.058)
Nesting Parameter				
λ	0.557	0.569	0.627	0.627
	(0.041)	(0.041)	(0.044)	(0.044)
Choice Set	Passed	Passed	Passed	Passed
Station Brand Fixed Effects	Y	Y	Y	Y
Month-Year Fixed Effects	Y	Y	Y	Y
Number of Stops	771	771	771	771
Number of Trips	19588	19588	19588	19588
Observations	834208	834208	834208	834208
McFadden R^2	0.285	0.286	0.287	0.288
	Implied Value of Time ($/hour)			
	19.73	18.96	20.82	19.71
	(3.45)	(3.19)	(3.43)	(3.28)

Notes: Table reports pseudo maximum likelihood estimates of driver preferences. The expected purchase quantities are predicted from the regressions in Table 4. Each column shows how estimates change if we add controls for the characteristics for the station's neighborhood. In particular, we add controls for the median household income ($10,000s) and population density (100 inhabitants per square mile) of each station's Census tract according to data from the 2010 American Community Survey. Coefficient standard errors are reported in parentheses. The implied value of time is calculated as $60 \cdot \frac{\gamma}{\alpha}$ and standard errors are reported in parentheses.

normalize to one. The quasi log-likelihood function is therefore:

$$\mathcal{QLL} = \sum_{i=1}^{N}\sum_{t=1}^{T}\sum_{k=0}^{K}\log\left(\frac{\exp(V_{ikt}(\widehat{\phi})/\lambda_k)(\sum_{j\in B_k}\exp(V_{ijt}(\widehat{\phi})/\lambda_k))^{\lambda_k-1}}{1+(\sum_{j=1}^{K}\exp(V_{ijt}(\widehat{\phi})/\lambda_k))^{\lambda_k}}\right)\mathbb{1}\{i \text{ chose } k \text{ on } t\} \tag{C.9}$$

where B_k refers to the nest for choice k, $V_{i0t} = W_{it}'\delta$ for the choice not to stop, and

$$V_{ikt}(\widehat{\phi}) = \alpha\big(\theta p_{ikt} + (1-\theta)\overline{p}_k\big)\cdot\big((Z_{ikt}^q)'\widehat{\phi}\big) + \gamma\text{ExcessTime}_{ikt} + X_k'\beta$$

for the choice to stop at each station $k = 1, \ldots, K$.

C.3 Value of Information

In Section 4.3, we measure the welfare effects of changing drivers' information, assuming consumers make purchase decisions based on imperfect *perceptions* about product attributes, and then ex-post utility depends on *actual* product attributes. Specifically, we compare the welfare difference between these two scenarios, $s \in \{0,1\}$, where the driver perceives the expenditure to refuel at station j as P_j^{0*} if $s = 0$ and as P_j^{1*} if $s = 1$. We calculate the expected change in consumer surplus for driver i on trip t between scenario 1 and scenario 0 as:

$$\begin{aligned}\Delta CS = -\frac{1}{\alpha}&\left[\ln\Big(1+(\sum_{j\in C^1}\exp(\frac{V_j^{1*}}{\lambda}))^{\lambda}\Big) - \ln\Big(1+(\sum_{k\in C^0}\exp(\frac{V_k^{0*}}{\lambda}))^{\lambda}\Big)\right]\\ &-\Big(\sum_{j\in C^1}\pi_j^{1*}(P_j - P_j^{1*}) - \sum_{k\in C^0}\pi_k^{0*}(P_k - P_k^{0*})\Big),\end{aligned} \tag{C.10}$$

where for scenario s, C^s is the choice set, P_j^{s*} is the perceived price at station j, P_j is the actual price at station j, V_j^{s*} is the perceived utility from choosing j, π_j^{s*} is the probability of choosing station j, and λ is the nesting parameter for the station nest. The actual prices at each station are held fixed across the two choice scenarios. The first line of Equation (C.10) is analogous to the standard formula for a change in consumer surplus for the nested logit model (Small and Rosen, 1981), given drivers' ex-ante *perceived* consumer surplus. The second line adjusts consumer surplus to account for the fact that actual prices paid may differ from consumer perceptions.

D Estimation Heterogeneity and Robustness Analysis

The disaggregate nature of our data allows us to further explore the heterogeneity in preferences across different types of drivers in our sample and across different types of trips. Table D.1 shows the results for specifications that allow the expenditure coefficient, α, the weight on current prices, θ, and the disutility from excess time, γ, to vary by age, gender, and Census tract income. Additionally, we allow for heterogeneity in price sensitivity across the time of trip (weekday versus weekend) and whether the driver's home has a garage.[48]

We find evidence of substantial heterogeneity in expenditure sensitivity across age groups. For ease of interpretation, we report the average marginal effects of belonging to each demographic group on the value of time in Table D.2. We find that a driver being in the oldest age category (age 60-70) is associated with a \$5/hour (25%) reduction in the value of time. This is in keeping with the literature on consumption in other settings (e.g. Aguiar and Hurst, 2005) that suggests that retirees may not change their consumption levels, but may engage in time-intensive approaches to reducing costs such as driving farther for less expensive gas or engaging in time-intensive search. Women's implied value of time is nearly \$6/hour (30%) higher than men's. In addition, we see that drivers from high-income Census tracts have a \$7.35/hour (37%) higher value of time compared to drivers in middle-income Census tracts. Lastly, having a home garage is associated with a \$4.65 increase in the value of time. Overall, the heterogeneity estimates—although sometimes statistically imprecise—are generally in line with demographic patterns in value of time that we would expect.

Having established our baseline and heterogeneity analyses, we also perform a series of robustness checks to support the validity of our value of time estimates. Table D.3 shows that our baseline value of time estimates are not particularly sensitive to removing month-by-year fixed effects and controlling instead for either the daily average price across all stations or the driver's mean purchase quantity in the value of not stopping. Table D.4 further shows that our value of time estimates are not sensitive to alternative assumptions about drivers' choice sets, such as only considering stations near home or that have been passed recently (within 7 or 14 days). Since each of these specifications further reduce the number of stops in the data, we maintain the specification with all stations previously passed in the data as the baseline choice set.

Table D.5 shows the impact of alternative specifications of stations' average prices based on shorter durations rather than the average price over our entire data period. We see that estimates of stations' average price based on shorter periods (week, month, quarter, or half-year) all yield somewhat larger weights on current price and higher implied values of time. This is what we would expect if these measures were only noisy estimates of the true long-run

[48] For the "purchase quantity regressions" we fit a flexible function that interacts both tank level and tank level squared with dummy variables for each of the demographic groups or trip types (*e.g.* weekend). We identify whether a driver's home has a garage by looking at images from an online mapping service near the location where the driver most frequently stops.

Table D.1: Heterogeneous Preferences

	(1)	(2)	(3)	(4)	(5)
Expected Expenditure					
α (Constant)	−0.511	−0.516	−0.533	−0.504	−0.563
	(0.062)	(0.063)	(0.066)	(0.063)	(0.066)
× 1[Age 60-70]	0.065				
	(0.017)				
× 1[Female]		0.015			
		(0.015)			
× 1[Low Income (Q1)]			−0.103		
			(0.020)		
× 1[High Income (Q5)]			−0.045		
			(0.021)		
× 1[Weekend]				−0.024	
				(0.017)	
× 1[Has Home Garage]					0.054
					(0.018)
Weight on Current Price					
θ (Constant)	0.393	0.243	0.207	0.322	0.364
	(0.085)	(0.084)	(0.079)	(0.083)	(0.137)
× 1[Age 60-70]	−0.299				
	(0.129)				
× 1[Female]		0.141			
		(0.125)			
× 1[Low Income (Q1)]			0.394		
			(0.153)		
× 1[High Income (Q5)]			0.097		
			(0.118)		
× 1[Weekend]				−0.031	
				(0.121)	
× 1[Has Home Garage]					−0.069
					(0.137)
Excess Time					
γ (Constant)	−0.179	−0.147	−0.128	−0.169	−0.150
	(0.019)	(0.018)	(0.017)	(0.019)	(0.022)
× 1[Age 60-70]	0.060				
	(0.019)				
× 1[Female]		−0.045			
		(0.020)			
× 1[Low Income (Q1)]			−0.064		
			(0.024)		
× 1[High Income (Q5)]			−0.087		
			(0.028)		
× 1[Weekend]				0.007	
				(0.022)	
× 1[Has Home Garage]					−0.025
					(0.020)
Nesting Parameter					
λ	0.536	0.552	0.552	0.557	0.576
	(0.041)	(0.041)	(0.041)	(0.041)	(0.041)
Choice Set	Passed	Passed	Passed	Passed	Passed
Station Brand Fixed Effects	Y	Y	Y	Y	Y
Month-Year Fixed Effects	Y	Y	Y	Y	Y
Number of Stops	771	771	771	771	771
Number of Trips	19588	19588	19588	19588	19588
Observations	834208	834208	834208	834208	834208

Notes: Table reports psuedo maximum likelihood estimates of driver preferences. Standard errors are reported in parentheses. The choice set for each specification includes stations that the driver has previously passed before that are within 20 minutes of the route. First stage expected purchase quantity and the value of not stopping regressions both also include demographics.

average price drivers use to form beliefs about station prices. While these alternative models provide a similar fit to the data (similar McFadden R-squared), we chose the baseline specification that would provide the most conservative estimate of the value of time. The differences

Table D.2: Average Marginal Effects on Value of Time ($/Hour)

1[Age 60-70]	1[Female]	1[Low Income (Q1)]	1[High Income (Q5)]	1[Weekend]	1[Has Home Garage]
-5.00	5.91	3.48	7.35	-1.67	4.65
(2.63)	(2.64)	(2.51)	(3.06)	(2.70)	(2.56)

Notes: Table reports the average marginal effect on the value of time. The marginal effects corresponds to the estimates in Column (5) of Table D.1. Delta method standard errors are reported in parentheses.

in the value of time between the quarter, half-year, and full sample average price specifications (which vary from $19.73/hour to $27.65/hour but are not statistically different from each other) do raise the possibility that drivers may be using information that is related to, but not precisely, the long-run average price as we specify it. Overall, we find evidence across robustness checks that drivers are not *only* using the current prices at stations to choose where to stop for fuel, but instead are forming expectations of station prices based on both current and lagged information. Our preferred specification approximates this lagged information with the average price at each station over our full time period, but future work with larger data sets and variation explicitly related to expectation setting may be able to better specify the form of these expectations.

Finally, Table C.1 in Appendix Section C shows that controlling for additional characteristics of the neighborhoods surrounding stations, such as the Census tract median income or population density, does not substantially affect our value of time estimates. This provides reassurance that our price sensitivity estimates are not affected by unobservable station quality attributes such as neighborhood safety.

Table D.3: Driver Preference Estimates - Market Prices and Heterogeneous Quantities

height	(1)	(2)	(3)	(4)
Value of Not Stopping				
1[No Stop] × Constant	−12.256	−16.857	−12.472	0.205
	(1.593)	(2.552)	(1.625)	(0.185)
1[No Stop] × Tank Level	1.229	1.275	1.325	0.439
	(0.086)	(0.089)	(0.090)	(0.062)
1[No Stop] × (Tank Level)^2	−0.012	−0.012	−0.008	0.059
	(0.006)	(0.006)	(0.006)	(0.011)
1[No Stop] × Daily Average Price ($/gallon)		1.439		
		(0.533)		
1[No Stop] × Driver's Mean Purchase (gallons)				−1.147
				(0.146)
Station Choice				
α - Expected Expenditure ($)	−0.471	−0.500	−0.511	−0.455
	(0.058)	(0.061)	(0.062)	(0.056)
θ - Weight on Current Price	0.180	0.385	0.313	0.162
	(0.061)	(0.096)	(0.076)	(0.059)
γ - Excess Time (minutes)	−0.183	−0.173	−0.168	−0.113
	(0.017)	(0.017)	(0.018)	(0.020)
Nesting Parameter				
λ	0.529	0.534	0.557	0.538
	(0.040)	(0.039)	(0.041)	(0.040)
Choice Set	Passed	Passed	Passed	Passed
Station Brand Fixed Effects	Y	Y	Y	Y
Month-Year Fixed Effects	N	N	Y	N
Number of Stops	771	771	771	771
Number of Trips	19588	19588	19588	19588
Observations	834208	834208	834208	834208
McFadden R^2	0.280	0.280	0.285	0.280
	Implied Value of Time ($/hour)			
	23.33	20.76	19.73	14.93
	(3.70)	(3.43)	(3.45)	(3.87)

Notes: Table reports pseudo maximum likelihood estimates of driver preferences. Coefficient standard errors are reported in parentheses. The implied value of time is calculated as $60 \cdot \frac{\gamma}{\alpha}$ and standard errors are reported in parentheses.

E EV Refueling Choice Model and Counterfactuals

This appendix provides additional details on our EV charging simulation. Section E.1 explains the assumptions underlying the simulation in more detail. Section E.2 expands on the implementation of our simulation model. Section E.3 presents the version of our simple theoretical model that allows the social planner to make a discrete choice between Level 2 and Level 3 charging for each station rather than continuously choosing the speed of the charging network.

E.1 Assumptions Underpinning EV Station Choice Simulation

Excess Refueling Time

When making their EV refueling decision, we allow drivers to either (1) drive to the charg-

Table D.4: Driver Preference Estimates - Varying Specification of Driver Choice Sets

	(1)	(2)	(3)	(4)	(5)
Value of Not Stopping					
1[No Stop] $\times$ Constant	-12.472	-8.761	-8.445	-12.7	-12.48
	(1.625)	(1.720)	(1.867)	(1.660)	(1.704)
1[No Stop] $\times$ Tank Level	1.325	1.485	1.257	1.335	1.333
	(0.090)	(0.144)	(0.116)	(0.092)	(0.094)
1[No Stop] $\times$ (Tank Level)^2	-0.008	-0.021	-0.007	-0.008	-0.011
	(0.006)	(0.008)	(0.009)	(0.006)	(0.006)
Station Choice					
α - Expected Expenditure ($)	-0.511	-0.421	-0.407	-0.520	-0.512
	(0.062)	(0.074)	(0.082)	(0.064)	(0.066)
θ - Weight on Current Price	0.313	0.448	0.386	0.305	0.276
	(0.076)	(0.115)	(0.125)	(0.075)	(0.079)
γ - Excess Time (minutes)	-0.168	-0.142	-0.095	-0.153	-0.147
	(0.018)	(0.021)	(0.023)	(0.018)	(0.018)
Nesting Parameter					
λ	0.557	0.440	0.421	0.560	0.572
	(0.041)	(0.049)	(0.054)	(0.042)	(0.043)
Choice Set	Passed Ever	Home $\leq$ 10 mi. & Passed	Home $\leq$ 5 mi. & Passed	Passed $\leq$ 14 days	Passed $\leq$ 7 days
Station Brand Fixed Effects	Y	Y	Y	Y	Y
Month-Year Fixed Effects	Y	Y	Y	Y	Y
Number of Stops	771	475	383	751	718
Number of Trips	19588	14343	11350	19557	19503
Observations	834208	605680	459354	727205	600454
McFadden R^2	0.285	0.301	0.300	0.279	0.271
	Implied Value of Time ($/hour)				
	19.73	20.26	13.98	17.63	17.19
	(3.45)	(4.73)	(5.15)	(3.36)	(3.52)

Notes: Table reports pseudo maximum likelihood estimates of driver preferences. The expected purchase quantities are predicted from the regressions in Table 4. Each column shows results varying our specification of drivers' choice set. Column (1) shows our base specification where all stations that the driver has previously passed that are within 20 minutes of the route are included in the choice set. Columns (2) and (3) also set the choice set to stations that the driver has previously passed that are within 20 minutes, but further restrict the sample to only trips that started within 10 miles and 5 miles of the driver's home, respectively. Columns (4) and (5) restrict the choice sets to only stations within 20 minutes of the route and that the driver has passed within the last 14 days or 7 days, respectively. Coefficient standard errors are reported in parentheses. The implied value of time is calculated as $60 \cdot \frac{\gamma}{\alpha}$ and standard errors are reported in parentheses.

Table D.5: Driver Preference Estimates - Varying Specification of $\overline{P}$

	(1)	(2)	(3)	(4)	(5)	(6)
Value of Not Stopping						
1[No Stop] $\times$ Constant	-12.256	-12.472	-6.029	-7.087	-8.795	-8.448
	(1.593)	(1.625)	(0.936)	(1.054)	(1.224)	(1.246)
1[No Stop] $\times$ Tank Level	1.229	1.325	1.067	1.118	1.21	1.188
	(0.086)	(0.090)	(0.070)	(0.075)	(0.082)	(0.082)
1[No Stop] $\times$ (Tank Level)^2	-0.012	-0.008	-0.012	-0.011	-0.011	-0.011
	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.006)
Station Choice						
α - Expected Expenditure ($)	-0.471	-0.511	-0.279	-0.329	-0.408	-0.394
	(0.058)	(0.062)	(0.039)	(0.045)	(0.054)	(0.055)
θ - Weight on Current Price	0.180	0.313	0.536	0.555	0.390	0.452
	(0.061)	(0.076)	(0.255)	(0.130)	(0.105)	(0.110)
γ - Excess Time (minutes)	-0.183	-0.168	-0.198	-0.188	-0.177	-0.181
	(0.017)	(0.018)	(0.020)	(0.019)	(0.019)	(0.019)
Nesting Parameter						
λ	0.529	0.557	0.547	0.540	0.541	0.546
	(0.040)	(0.041)	(0.043)	(0.043)	(0.041)	(0.042)
Choice Set	Passed	Passed	Passed	Passed	Passed	Passed
Station Brand Fixed Effects	Y	Y	Y	Y	Y	Y
Month-Year Fixed Effects	N	Y	Y	Y	Y	Y
$\overline{P}$	Full Sample	Full Sample	Week	Month	Quarter	Half-Year
Number of Stops	771	771	771	771	771	771
Number of Trips	19588	19588	19588	19588	19588	19588
Observations	834208	834208	834208	834208	834208	834208
McFadden R^2	0.280	0.285	0.282	0.283	0.283	0.283
	Implied Value of Time ($/hour)					
	23.33	19.73	42.50	34.23	26.03	27.65
	(3.70)	(3.45)	(7.90)	(6.51)	(4.83)	(5.33)

Notes: Table reports pseudo maximum likelihood estimates of driver preferences. The expected purchase quantities are predicted from the regressions in Table 4. Each column shows how estimates change if we specify a different measure of average price ($\overline{P}$) entering the drivers expectations. Column (1) is our base specification that sets $\overline{P}$ as the station's average price over the entire sample, Columns (3-6) set $\overline{P}$ as the mean price at the station over the last week, month, quarter, and half year, respectively, where the moving average window is truncated if the observation is at the beginning of the data period and such that there are not sufficient previous days in the data set. Choice Set = "All" indicates that all stations with 20 minutes of the driver's route are included in the choice set. Choice Set = "Passed" means stations that the driver has previously passed that are within 20 minutes of the route are included in the choice set. Coefficient standard errors are reported in parentheses. The implied value of time is calculated as $60 \cdot \frac{\gamma}{\alpha}$ and standard errors are reported in parentheses.

ing station and wait for their vehicle to recharge or (2) park their vehicle at the charging station and walk to their destination. For each station, we assume that drivers would choose to "walk" or "wait" to minimize the excess time spent refueling.

We define excess time refueling as the additional time that a driver would spend if they choose to refuel on a given trip compared to if they were instead to travel directly from their trip's origin to destination on the optimal route plus the time spent at the destination. We include the time spent at the destination because drivers can recharge their vehicle while they are visiting a destination. For example, a driver that spends many hours at work could park and charge their EV at a station during the work day.

The calculation of excess time refuel time for each EV station on each trip entails several steps:

1. Determine the drivers' refuel quantities conditional on stopping.

    - In our estimation sample, drivers purchase an average of 8.4 gallons which is equivalent to 283 kWh of electricity. Thus, on our baseline counterfactual we assume that drivers would refill their EV with an equivalent amount of "fuel" as we observe the drivers refueling in the gasoline market. However, we also show how the results would change if drivers decided to refill their EV more frequently with smaller quantities.
    - In our main set of counterfactuals, we assume that EV drivers stop for fuel at the same rate as drivers of gasoline vehicles. One interpretation of this assumption is that EVs have the same range as gasoline vehicles, but any interpretation that maintains the relationship between the continuation value (value of not stopping) and the value of stopping at any of the stations would lead to this frequency of stopping. This assumption allows us to better isolate the effects of the charging station network on excess refueling time. However, a limitation is that current EVs may have shorter range than gasoline vehicles, or drivers may feel more uncomfortable with low battery charge than they do with low fuel tank level. Moreover, because EVs refuel at a much slower rate than gasoline vehicles, drivers may choose to refuel EVs more (or less) frequently with smaller (larger) quantities. We run several alternative specifications that vary the assumed frequency and quantities that drivers would refuel their EV. These results are in Table 8.

2. Calculate the technological time required for the driver to refuel.

    - The time required to refill is determined by the assumed speed of the charging technology. For example, a 121 kW charger would take over two hours to charge 283 kWh of electricity.

3. Calculate the excess time associated associated with the driver's two possible refueling options: (1) "wait" and (2) "walk".

- For the "wait" option the total excess refueling time is equal to the sum of the excess driving time to travel to the station plus the technological charging time (from Step 2). The excess drive time is calculated in the same way as we calculate excess travel time for gasoline stations (see Section 2).
- For the "walk" method, excess time is calculated as follows:
  (a) Determine the amount of time that the driver spends at the final destination.
  (b) Calculate the time it would take the driver to walk round-trip from the refueling station to the destination assuming a walking speed of 3 miles per hour.
  (c) Determine how much "additional" waiting time, if any, is needed to complete the charging cycle. Here, we compare the technological refuel time with the sum of the time spent at the destination and the round-trip walking time from the station. If the technological refuel time exceeds the sum, then additional waiting time is added to the total excess refueling time.
    - For instance, suppose a driver requires three hours of technological charging time to achieve the refuel quantity established in Step 1. Further, suppose that the driver spends 2 hours at the destination and it takes 20 minutes to walk to and from the charging station. In this case, 40 minutes of waiting time is added to the total excess time for the "walk" method.
  (d) Calculate the net driving time to drive from the origin to the charging station instead of the origin to the destination.
    - Note that this net driving time could be negative if the station is closer to the origin than the destination.
  (e) The total excess time for the "walk" option is:
  Total Excess Time = Added Walk Time (b) + Added Wait Time (c) + Net Drive Time (d)

4. The excess time for each station on each trip is determined by the option (walk or wait) with the lowest total excess time.
   - In our main set of counterfactuals, we assume that drivers always choose the time-minimizing option whenever they decide to wait at a charging station or to park at the station and walk to their destination. In practice, drivers may have explicit preferences for either walking or waiting when recharging their EV. Therefore, we also solve the counterfactuals under two alternative assumptions: (1) drivers prefer to always walk and (2) drivers prefer to always wait. We also present additional results for scenarios in which drivers value waiting time differently than they value walking time. These results are in Table 8.

Example

As an example, suppose a driver is traveling from origin A to destination C and is considering stopping at station B. The driver has two options: (1) they can wait at station B—this option adds five minutes of driving time to visit station B, or (2) they can leave their car at station B and walk to destination C—this option saves two minutes of total driving time but adds 32 minutes of round-trip walking time. The vehicle will take one hour to recharge and the driver plans to spend 20 minutes at destination C. If the driver waits at the station for the vehicle to charge, they will add 65 minutes of excess time to refuel—five minutes of added driving time plus 60 minutes of waiting time. On the other hand, if they choose to park and walk to the destination they would add only 40 minutes of excess refueling time—two fewer minutes of driving time, 32 minutes of added walking time, and 10 minutes of waiting time. Thus, we specify that the excess time associated with recharging at station B is 40 minutes.

Driver Refueling Choice

Once we have computed each driver's excess refueling time for each station, we simulate drivers' refueling choices. In this step of the simulation, we further assume that EV charging stations have homogeneous prices and brand qualities. This assumption allows us to isolate the impact of changes in the density of charging stations and speed of the charging stations on excess refueling time. In practice, we would expect that charging stations with more convenient locations might charge higher prices. If stations with better locations were to charge higher prices that would lead us to underestimate expected EV refueling times. However, given that the EV network is relatively sparse and that drivers have a high value of time, we do not expect that price heterogeneity would lead to substantial changes in station choices.

To carry out the simulations, we also need to ensure that our model predicts enough charging events so that drivers would obtain a sufficient amount of charge to cover the mileage that we observe in the driving data. To achieve this, we start with an initial guess for the intercept in the value of not stopping equation. We then iterate the EV refueling simulations with different levels of the intercept in the value of not stopping equation until we find an intercept value such that the expected number of EV charging events predicted by the model is equal to the total number of gasoline refueling event observed in the driving data. In robustness checks in which we assume the EV refueling frequency differs from the gasoline refueling frequency, we accordingly solve for an intercept in the value-of-not-stopping equation that would predict the desired EV refueling frequency.

E.2 EV Counterfactual Simulations Implementation Details

As discussed in Section 5.2, we use our model of drivers' behavior, combined with a range of different potential EV charging network configurations, to understand the value to drivers of improving network speed relative to density. We calculate the excess refueling time for each EV station located within a 20-minute drive of each trip in our data and use our refueling choice model outlined in Section 3 to predict which charging stations drivers would choose and the expected time cost associated with those refueling choices. We repeat this simulation for 220 different combinations of EV charging station speed and density, with charger speeds ranging from 15kW to 300kW and for the locations of chargers in each year between 2012 and 2022. We then regress these excess times on a flexible translog functional form of charging network speed and density to understand the expected marginal effect of changing network speed or density on drivers' excess time per stop. This approach mimics the approach in Gowrisankaran et al. (2023) and Butters et al. (2021) where firms form expectations of profits based on a profit surface across model states.

Table E.1: Translog Regression Fit of Excess Refuel Time Surface

	Log(Excess Refuel Time)	
	(1)	(2)
Intercept	6.246***	10.09***
	(0.0783)	(0.1930)
Log(N)	-0.2349***	-0.4840***
	(0.0096)	(0.0493)
Log(S)	-0.2726***	-1.732***
	(0.0103)	(0.0384)
Log^2(S)		0.1615***
		(0.0035)
Log^2(N)		0.0170***
		(0.0036)
Log(N) $\times$ Log(S)		0.0070*
		(0.0037)
Observations	220	220
R^2	0.85574	0.98698
Adjusted R^2	0.85441	0.98667

Notes: Table reports regression estimates of excess refueling time (per stop) on the number of stations (N) and the charger speed of the network. We use 99 different combinations of stations (N) and kW charger speed (S) to fit the regressions. Our preferred specification, Column (2), is used to evaluate the elasticity of excess refueling time with respect to changes in the number of stations (N) and changes to the charging speed (S) of the network.

Table E.1 presents the results of this regression both with and without higher order interactions of network speed (S) and density (N). We see that including squared terms and interactions improves the model fit substantially, with an R^2 of 0.987 in our preferred specification (Column (2)). In both models, increasing speed or density will decrease excess time, and the squared terms and interaction are all positive in Column (2). We use derivatives based on this regression to understand the elasticity of excess time with respect to network speed and density.

E.3 Model of Public Charging Investment - Discrete

This appendix recasts the social planner's problem in Section 5.2 as a discrete choice between faster, more expensive chargers and slower, cheaper chargers. This setting more precisely replicates a planner choosing between installing Level 2 and Level 3 chargers.

Consider a planner that chooses a charging network design to minimize drivers' time costs subject to a budget constraint. The planner can invest in two types of charging technologies—fast chargers (e.g., direct-current chargers) and standard chargers (e.g., alternating-current chargers). The two technologies differ in their power capacity (kW), which determines the recharging rate and the time required to recharge an EV's battery. The power capacity of a fast charger is ρ_F, whereas the power capacity of a standard charger is ρ_S. As such, the planner chooses the number of stations to build, N, and the share of stations that are fast chargers, $S_F \in [0, 1]$. The share of fast chargers implicitly defines the average charging rate (i.e., speed) of the network, $\overline{R}$. Drivers' refueling time, τ, is a decreasing function of both the N and S_F. However, increasing either N or S_F will raise the capital cost of the network. For simplicity, we abstract away from modeling the exact locations of the charging stations and begin by assuming that each additional fast charger and standard charger would be evenly distributed spatially throughout the network. Thus, the planner's investment problem can be written formally as follows:

$$\begin{aligned} \min_{N, S_F} \quad & \tau(N, S_F) \\ \text{s.t.} \quad & \kappa \cdot N \cdot \overline{R} \leq B, \\ & 0 \leq S_F \leq 1, \\ & \overline{R} = S_F \cdot \rho_F + (1 - S_F)\rho_S \end{aligned} \tag{E.1}$$

Here, the first constraint imposes that the total capital cost of the network ($\kappa \cdot N \cdot \overline{R}$) must be weakly less than the budget available to the planner to spend on charging infrastructure, B. The functional form for capital costs is motivated by Nicholas (2019), whose estimates show that capital costs are roughly proportional to the total power capacity ($N \cdot \overline{R}$) of the network. For example, installing four 20 kW standard chargers would cost approximately the same as installing one 80 kW fast charger. Therefore, the κ parameter represents the fixed cost of increasing the power capacity of the network. The second constraint imposes that the share of fast chargers is bounded between zero and one. Finally, the third constraint defines the relationship between the share of fast chargers, S_F, and the average charge speed of the network, $\overline{R}$, which determines the capital cost of the network.

To clarify the exposition of the optimal solution, we recast the planner's problem as a choice of the number of stations, N, and the average charge speed, $\overline{R}$, noting that there is a one-to-one mapping between $\overline{R}$ and S_F, as shown in the last line of E.1. The refueling time function is monotonically decreasing in both arguments, so the planner would choose to use the entire budget. Thus, the following Lagrangian characterizes the solution to the planner's problem:

$$\begin{aligned}\mathcal{L}(N,\overline{R},\lambda,\mu_1,\mu_2) = \tau(N,\overline{R}) + \lambda_1(\kappa\cdot N\cdot\overline{R}-B)\\ +\mu_1\cdot(\tfrac{\overline{R}-\rho_s}{\rho_F-\rho_s}) + \mu_2(\tfrac{\overline{R}-\rho_s}{\rho_F-\rho_s}-1).\end{aligned} \tag{E.2}$$

Above, λ represents the shadow cost of relaxing the budget constraint, and μ_1 and μ_2 represent the Lagrangian multiplier associated with the constraints that the share of fast chargers ($S_F = \frac{\overline{R}-\rho_s}{\rho_F-\rho_s}$) must be weakly greater than zero and weakly less than one.

Broadly, the solution to the planner's problem can be separated into three possible cases: (1) an interior solution in which the optimal share of fast chargers lies between zero and one, (2) a corner solution in which the optimal share of fast chargers equals zero, and (3) a corner solution in which the optimal share of fast chargers equals one. Below, we derive the Karush–Kuhn–Tucker (KKT) conditions for an optimal solution to the problem.

Case 1: Interior solution

The KKT conditions for an interior solution where the share of fast chargers ($S_F = \frac{\overline{R}-\rho_s}{\rho_F-\rho_s}$) lies between zero and one are as follows:

$$\frac{\partial\tau}{\partial\overline{R}} = \lambda\cdot\kappa\cdot N, \tag{E.3}$$

$$\frac{\partial\tau}{\partial N} = \lambda\cdot\kappa\cdot\overline{R}, \tag{E.4}$$

$$\kappa\cdot N\cdot\overline{R} = B. \tag{E.5}$$

Rearranging the first two conditions we can derive the following simple optimality condition:

$$\tfrac{\partial\tau}{\partial N}\cdot\tfrac{N}{\tau} = \tfrac{\partial\tau}{\partial\overline{R}}\cdot\tfrac{\overline{R}}{\tau} \quad\Rightarrow\quad \varepsilon_N = \varepsilon_{\overline{R}}. \tag{E.6}$$

Intuitively, the most efficient charging network for a given level of spending must satisfy the condition that the elasticity of time savings from adding additional stations ε_N should be equal to the elasticity of time-saving from increasing the charging speed of the network $\varepsilon_{\overline{R}}$.

Case 2: Corner solution with no fast chargers

The KKT condition for the case where no fast chargers are built is shown below.

$$\varepsilon_N > \varepsilon_{\overline{R}} \tag{E.7}$$

$$\kappa\cdot N\cdot\overline{R} = B \tag{E.8}$$

In order for it to be optimal to build only standard chargers and build no fast chargers, condition E.7 states that the elasticity of time-savings from increasing the number of stations must be strictly greater than the elasticity of time-savings from increasing the average speed of the network.

Case 3: Corner solution with all fast chargers

The KKT condition for the case where only fast chargers are built are as follows:

$$\varepsilon_N < \varepsilon_{\overline{R}}, \tag{E.9}$$

$$\kappa \cdot N \cdot \overline{R} = B. \tag{E.10}$$

Analogous to Case 2, it is optimal to spend the entire budget on fast chargers if the elasticity of time-savings from increasing the number of stations must be strictly less than the elasticity of time-savings from increasing the average speed of the network.

Discrete Model Implementation

When implementing the simulation where the social planner chooses a share of charging stations to be Level 3 rather than Level 2, the largest difference from our baseline simulation with continuous charging network speed choice is that we need to take a stand on *which* stations would be Level 3 rather than Level 2 if the social planner chooses an interior solution. To do this, we train a machine learning model of Level 3 charging using data on the characteristics of charger locations for chargers in Michigan and Ohio and 8 nearby states: Illinois, Indiana, Kentucky, Minnesota, New York, Pennsylvania, West Virginia, and Wisconsin. These characteristics included: (i) housing and demographic characteristics at a census tract level (U.S. Census Bureau, 2012), (ii) the straight-line distance from the EV charger to the nearest primary road in 2010 (U.S. Census Bureau, 2010a) and 2021 (U.S. Census Bureau, 2021), and (iii) the number of businesses by primary NAICS code in the census block and census tract of the charger (Data Axle, 2010; U.S. Census Bureau, 2010b). We use logistic lasso and importance sampled learning ensemble (ISLE) as suggested in Hara et al. (2021) for prediction tasks for tabular data. We follow the hyperparameter tuning and model evaluation in Hara et al. (2021). Training and test sets are divided randomly by station-level (training, 75%; test, 25%). The best algorithm is the ISLE with subsampling ratio 0.5 and learning rate 0.1 when evaluated by the area under the receiver operating characteristic curve (AUC, 0.802). The ISLE with subsampling ratio 0.5 and learning rate 0.1 is also known as a standard hyperparameter specification for stochastic gradient boosting machine (Friedman, 2002). We then applied the best algorithm to predict Level 3 charger installation probabilities for each EV charging station in Michigan and Ohio. We also calculate Shapley Additive explanation (SHAP, Lundberg and Lee, 2017) values ease interpretability of the model.

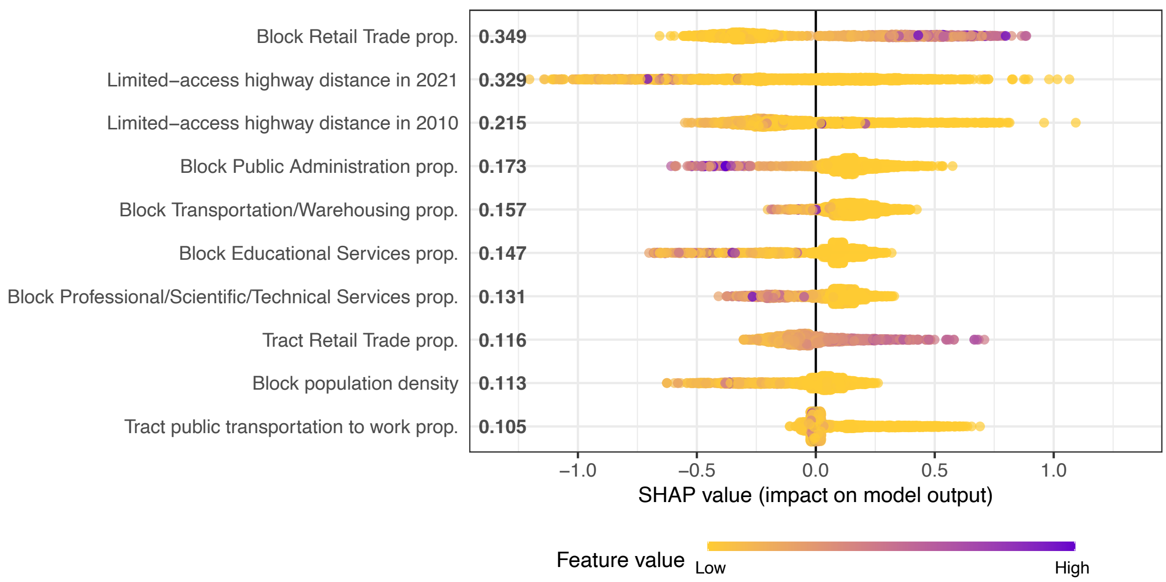
After predicting the probability of each charger in Michigan and Ohio being Level 3, we order chargers by this predicted probability and then assume that the social planner will make chargers Level 3 in this order (from highest probability to lowest probability). Thus, if the social planner chooses 25% of charging stations to be Level 3, then the 25% of stations with the highest probability of being Level 3 as predicted in our model will be Level 3 and the remainder will be Level 2.

Given this ordering of stations, we assume that Level 3 chargers charge at 80 kW and Level 2 chargers charge at 10 kW,[49] and then conduct robustness checks with faster Level 3 charging speeds. We proceed with solving the social planner's optimization in the same way that we did with the continuous choice case: We solve for drivers' excess time given a grid of 121 potential combinations of Level 3 charger share $\{0, 0.1, 0.2...1\}$ and station density (year 2012-2022). We then fit a flexible functional form through these excess time estimations to predict the excess time for any potential combination of Level 3 share and station density.

Discrete Model Results

We present some illustrative results of our machine learning model before turning to the social planner's solution to the discrete optimization problem.

Figure E.1: Impact of Local Characteristics on Level 3 Chargers



Notes: The figure shows the SHAP values for the ten most important variables for explaining whether a given charger is Level 3 rather than Level 2 in Minnesota, Wisconsin, Illinois, Indiana, Kentucky, West Virginia, New York, and Pennsylvania. We plot low feature values plotted in yellow and high feature values plotted in purple.

[49]The U.S. Department of Transportation defines Level 2 chargers as being 7-19 kW and Level 3 chargers as being 50-350 kW (https://www.transportation.gov/rural/ev/toolkit/ev-basics/charging-speeds).

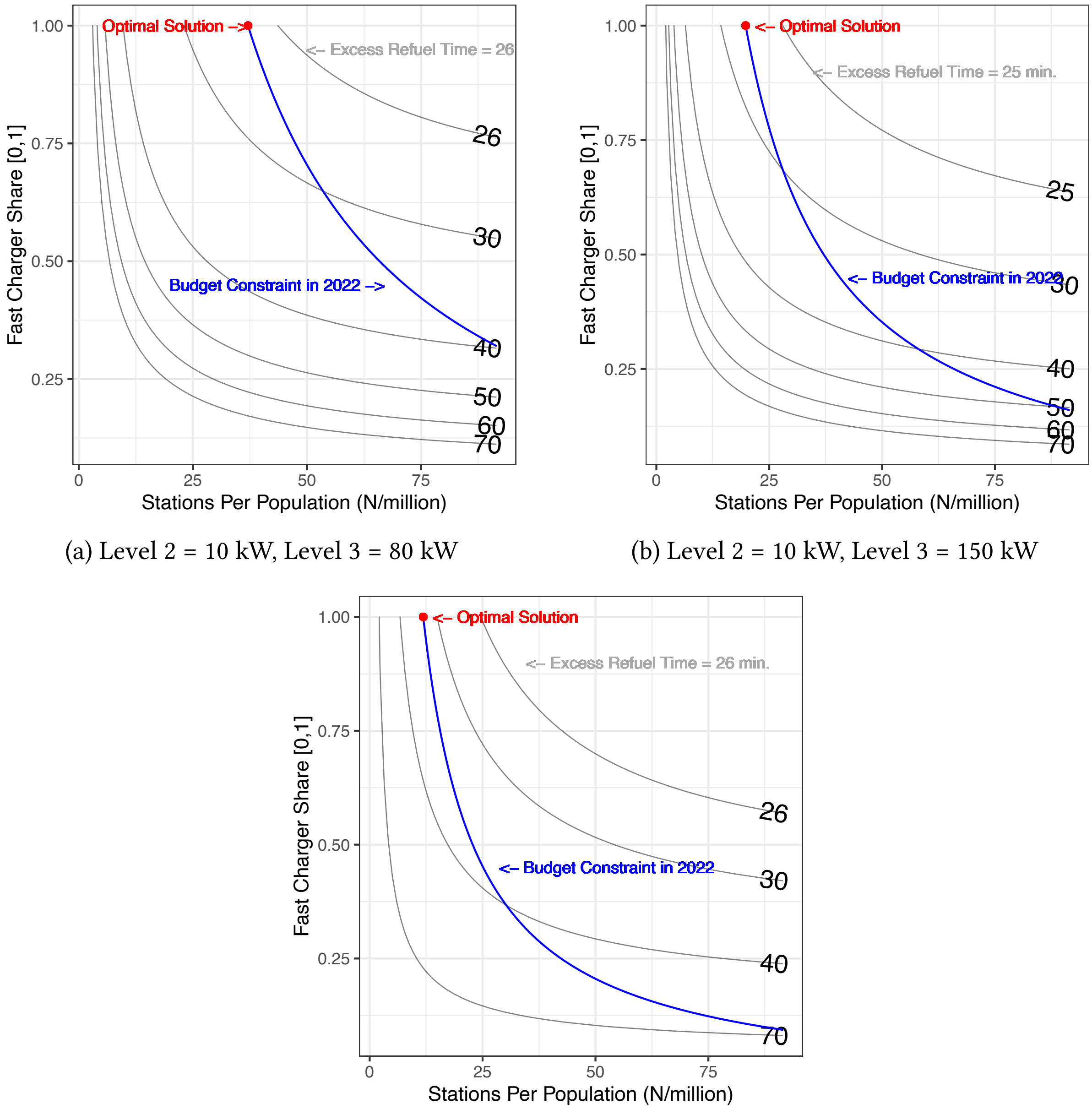
Figure E.1 displays the ten most important variables in explaining whether a given charger is Level 3 rather than Level 2. For the top variable, Census block retail trade proportion, we see the yellow dots concentrated below 0 SHAP value and the purple dots concentrated above 0. This means that if the EV charging station's Census block retail trade proportion is high, the predicted probability of having Level 3 EV chargers increases. On the flip side, it also means that if the EV charging station's Census block retail trade proportion is low, the predicted probability of having Level 3 EV chargers decreases. In contrast, for the bottom variable, the proportion of people who uses public transportation to work in the Census tract, we see the yellow dots concentrated above 0 SHAP value but not many dots below 0. This means that if the EV charging station's Census tract public transportation to work proportion is high, the predicted probability of having Level 3 EV chargers is lower, all else equal. However, the flip side is not true in this case. The predicted probability of having Level 3 EV chargers does not increase when the EV charging station's Census tract public transportation to work proportion is low. The indices on the left are the average absolute values of these SHAP values, which is called global impact in (Lundberg et al., 2018). So, for example, if we average over the absolute values of the plotted dots for the top variable, Census block retail trade proportion, we get 0.349. We include the ten variables with the highest global impact in the order of their global impact in the figure.

Figure E.2 presents information analogous to Figure A.7 in the text, showing the excess time contour lines in grey and the social planner's budget constraint in blue for the discrete social planner's problem in 2022. Panel E.2a shows the solution to the problem if the Level 3 charger technology is assumed to be 80 kW. We see that the excess time-minimizing solution is a corner solution to allocate 100% of charging stations as Level 3.

Figure E.2b and E.2c show that this result–that the social planner would choose to make 100% of stations Level 3–holds across a variety of potential charging speeds for Level 3 chargers. In fact, we found that the social planner would only allocate less than 100% of chargers to Level 3 if we assume that Level 2 chargers are faster than the SAE maximum charge rate for Level 2 chargers of 20 kW.

We take this as substantial support for the baseline result in the paper that investments in charging station speed substantially outperform investments in charging station density in terms of reducing EV drivers' excess time spent refueling. Because these results always find the corner solution–that the social planner would invest 100% of their budget in Level 3 charging–we focus our primary analysis in the main text on the social planner's continuous choice over charging station speed rather than the discrete choice between Level 2 and Level 3 charging.

Figure E.2: Discrete Charger Model Solutions in 2022



(a) Level 2 = 10 kW, Level 3 = 80 kW

(b) Level 2 = 10 kW, Level 3 = 150 kW

(c) Level 2 = 10 kW, Level 3 = 250kW

Notes: The thin grey lines show contours representing the estimated excess refueling time per EV refueling stop across different counterfactual combinations of station density (number of stations) and the share of fast chargers. For these simulations we assume the Level 3 fast chargers have a rate of either 80 kW (a), 150 kW (b), or 250 kW (c). We assume that Level 2 chargers would charge at a rate of 10 kW.

References

Butters, R Andrew, Jackson Dorsey, and Gautam Gowrisankaran, "Soaking up the sun: Battery investment, renewable energy, and market equilibrium," Technical Report, National Bureau of Economic Research 2021.

Data Axle, "ReferenceUSA historical business dataset," 2010. Accessed through the University of Arizona Libraries: https://arizona.figshare.com/articles/dataset/ReferenceUSA_historical_business_datasets/14515541.

De los Santos, Babur, Ali Hortaçsu, and Matthijs R. Wildenbeest, "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, 2012, *102* (6), 2955–80.

Diamond, Peter A, "A model of price adjustment," *Journal of Economic Theory*, 1971, *3* (2), 156–168.

Dorsey, Jackson, Ashley Langer, and Shaun McRae, "Fuel gauge image classifications by Amazon Mechanical Turk workers," 2013. Data collected between April 24, 2013 and May 30, 2013.

Esri, "Great Lakes Shorelines," 2019. https://www.arcgis.com/home/item.html?id=d87347457bc84e5c985db9e904b66b10.

Friedman, Jerome H, "Stochastic gradient boosting," *Computational statistics & data analysis*, 2002, *38* (4), 367–378.

Government of Canada, "Provinces/Territories, Cartographic Boundary File - 2016 Census," 2016. https://open.canada.ca/data/en/dataset/a883eb14-0c0e-45c4-b8c4-b54c4a819edb.

Gowrisankaran, Gautam, Ashley Langer, and Wendan Zhang, "Policy uncertainty in the market for coal electricity: The case of air toxics standards," Technical Report, National Bureau of Economic Research 2023.

Hara, Konan, Yasuki Kobayashi, Jun Tomio, Yuki Ito, Thomas Svensson, Ryo Ikesu, Ung il Chung, and Akiko Kishi Svensson, "Claims-based algorithms for common chronic conditions were efficiently constructed using machine learning methods," *Plos one*, 2021, *16* (9), e0254394.

Hastings, Justine S and Jesse M Shapiro, "Fungibility and consumer choice: Evidence from commodity price shocks," *The Quarterly Journal of Economics*, 2013, *128* (4), 1449–1498.

Lundberg, Scott M and Su-In Lee, "A unified approach to interpreting model predictions," *Advances in neural information processing systems*, 2017, *30.*

__ , Gabriel G Erion, and Su-In Lee, "Consistent individualized feature attribution for tree ensembles," *arXiv preprint arXiv:1802.03888*, 2018.

Nicholas, Michael, "Estimating electric vehicle charging infrastructure costs across major US metropolitan areas," *Int. Counc. Clean Transp*, 2019, *14* (11).

Oil Price Information Service, "Daily gasoline station locations and retail prices for Michigan and Ohio, April 2009 to May 2010," 2012. Dataset purchased in May 2012.

Sayer, James R, Mary Lynn Buonarosa, Shan Bao, Scott E Bogard, David J LeBlanc, Adam D Blankespoor, Dillon S Funkhouser, and Christopher B Winkler, "Integrated vehicle-based safety systems light-vehicle field operational test, methodology and results report: Dataset," Technical Report, University of Michigan, Ann Arbor, Transportation Research Institute 2010. Accessed on Tri-IvbssDB database at UMTRI in May 2012.

Small, Kenneth A and Harvey S Rosen, "Applied welfare economics with discrete choice models," *Econometrica: Journal of the Econometric Society*, 1981, pp. 105–130.

U.S. Census Bureau, "TIGER/Line Shapefiles: Primary Roads," 2010. https://www2.census.gov/geo/tiger/TIGER2010/PRIMARYROADS/.

_ , "TIGER/Line Shapefiles with 2010 Census block geography," 2010. https://www2.census.gov/geo/tiger/TIGER2010BLKPOPHU/.

_ , "American Community Survey 5-year Summary Files, 2008–2012," 2012. https://www2.census.gov/programs-surveys/acs/summary_file/2012/data/5_year_by_state/.

_ , "Cartographic Boundary Files: Shapefiles for County and States," 2018. https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html.

_ , "TIGER/Line Shapefiles: Primary Roads," 2021. https://www2.census.gov/geo/tiger/TIGER2021/PRIMARYROADS/.