Fueling Alternatives: Evidence from Real-World Driving Data

Ashley Langer, University of Arizona, and Shaun McRae, ITAM
Proposal for NBER's "Economics of Energy Use in Transportation"

Project Overview

Gasoline is currently the overwhelming fuel of choice for personal vehicles. While vehicles that run on ethanol, natural gas, or electricity are present on the roads, their adoption is generally constrained by the relative scarcity of fueling stations.[1] The economics literature has begun to show that electric charging stations have a causal effect on electric vehicle adoption in a relatively aggregated sense (e.g. Li and Zhou (2016), Zhou and Li (2016), Li, Jing (2017)), but evidence on the relationship between fueling station infrastructure and household preferences for vehicle adoption using micro-data is scarce. Understanding the density of alternative fueling stations required to encourage households to adopt alternative fuel vehicles is crucial for public policies that aim to encourage the transition of personal transportation away from gasoline.

In this project, we will use novel, high-frequency data on individuals' driving and fueling behavior to understand how dense a network of alternative fuel stations would need to be to make alternative fuel vehicles attractive to a larger share of the population. We have access to data on where and how a group of people drive for 40 days per person at a frequency of ten hertz (10x/second). In addition to data on the vehicle's location and speed, we observe fuel consumption, driving behavior, and cameras that allow us to observe the vehicle's surroundings. This data allows us to very clearly identify not only where drivers stop to purchase gasoline, but also the full set of routes and trips that the drivers make surrounding this refueling decision.

We will use this data to estimate drivers' willingness to trade off traveling farther from their routes for lower fuel prices or desirable station attributes such as an attractive brand or a right-hand-turn entrance. In particular, by estimating how willing drivers are to travel out of their way, we will be able to then calculate the difference in consumer surplus that they would receive from driving an alternative fuel vehicle that they could only fuel at a subset of fueling stations. We will then be able to estimate surfaces that show the range of consumer surplus loss from alternative fuel vehicles relative to gasoline vehicles across different relative station densities and relative fuel costs.

To our knowledge, this will generate the first estimates of consumers' valuation of fueling station density that rely on micro-level data on drivers' real-world decisions rather than either aggregated data or calibrated models (e.g. Greaker and Heggedal (2010)). This information is critical for understanding the density of alternative fuel stations necessary for widespread adoption of alternative fuel vehicles. While there have been extensive subsidies of vehicles that run on compressed natural gas or electricity, public policies to encourage fueling infrastructure have been more heterogeneous. The combination of high costs to construct stations and deliver fuels like compressed natural gas and the low level of alternative fuel vehicle adoption has likely limited the development of this technology. However, with the increasing speed of electric vehicle charging, we are approaching a situation where electric charging will look increasingly like gasoline refueling, so an understanding of the value of the gasoline station network and the required density for an alternative fueling network to be competitive is both important and timely.

Neither coauthor has any conflict of interest that may bear on the proposed work.

[1] Retail gasoline in the U.S. is generally a 10% ethanol blend, but in this project we will focus on vehicles that run on 100% alternative fuel. "Flex-fuel" vehicles that can run on ethanol or gasoline are widely adopted in the United States, but are overwhelmingly fueled by gasoline rather than ethanol in real-world driving (Anderson and Sallee, 2011).

Data

The data we will use for this project comes from the Integrated Vehicle-Based Safety Systems (IVBSS) study conducted by the University of Michigan's Transportation Research Institute (UMTRI) from April 2009 to May 2010. During this study, identical vehicles were provided to 108 drivers in southeast Michigan for about seven weeks each, who drove over 200,000 total miles during the study. The drivers used the vehicles as if they were their own (including purchasing their own gasoline) and UMTRI collected a detailed dataset that included driving data such as location, speed, acceleration, heading, weather, and instantaneous fuel use. Cameras in the vehicles captured video of the driver and the surrounding roadway.

We have already done considerable work to identify every time a driver in the sample stops to purchase gasoline and to get an estimate of how much gasoline is purchased. To identify gasoline purchases, we matched the vehicle locations from the driving data to a database of gasoline stations from OPIS (Oil Price Information Service) and contain the name, brand, address, approximate geographic coordinates, and daily gasoline prices for every gas station in Michigan and Ohio. For gas stations in southeast Michigan, we supplemented this information using aerial photography from Google Earth to add the exact latitude and longitude of the gas pumps and each of the station entrances, and we identified every vehicle stop within a 100 meter radius of a gas pump. We then reviewed the left-side camera images for all of these potential stops, and if the camera showed that the vehicle was stopped beside a gas pump then the stop was coded as a gasoline refueling stop.

One variable not recorded by the monitoring equipment was the fuel tank level. The amount of fuel remaining in the tank is the major factor that determines whether a driver stops for gasoline. We recovered an estimate of the fuel tank level using images from an in-car "over-the-shoulder" camera directed at the steering wheel and dashboard, combined with second-by-second fuel consumption data. With this data we find that drivers very rarely stop to purchase gasoline unless they have less than a quarter of a tank remaining. About one third of drivers fill the tank at every gas stop, while only about 5 percent of drivers leave the gas station with their fuel tank less than half full. The mean quantity of gasoline purchased at each stop is 8.3 gallons. The mean gas price paid by drivers in the sample is \$2.59 per gallon, with a range from \$1.97 to \$2.96.

Finally, combining the fueling stop and tank level data with the high-frequency location data, we use Dijkstra's algorithm applied to the street network to calculate a measure of the excess time for the driver to arrive at every gas station on every trip. The median excess time to the chosen gas station is 0.2 minutes, and slightly more than half of gas stops have an excess time of 30 seconds or less. In most cases, these represent stops at gas stations along the route that the driver was on, with no deviation from the optimal route. On its own, this would suggest that the network of alternative fueling stations would need to be extremely dense to make alternative fuel vehicles attractive to drivers, but on its own does not allow us to say anything about the cost of a more limited fueling network.

Empirical approach

To understand how drivers would react to a more limited set of alternative fueling locations, we have run some preliminary regressions that explore how drivers choose between gasoline fueling stations with different characteristics. We have currently estimated two full-information discrete choice models of driver's choice of gas station: a mulitnomial logit conditional on stopping on a particular trip and a nested logit that first allows the driver to choose whether to stop and, if so, which station to stop at.

On a given trip $t = 1, ..., T$ each driver $i = 1, ...N$ has a choice of whether to stop at each of $k = 1, ..., K$ stations and the choice $k = 0$ of not stopping, which is in its own next. The utility

that each driver receives for each choice is given by:

$$U_{ikt} = C + \alpha p_{kt} + X_{ikt}\beta + \xi_{i0t} + \varepsilon_{ikt}$$

where p_{kt} is the price of gasoline at station k on the day of trip t, X_{ikt} are characteristics of the driver, station, and trip, such as whether the station requires the driver to turn left to enter, and ξ_{0t} is the value of not stopping on this trip, which we model as:

$$\xi_{i0t} = \gamma W_{i0t}$$

where W_{0t} is a vector of characteristics of the trip including the type of trip (commute or non-commute) and the amount of gasoline remaining in the fuel tank at the start of the trip. Not stopping is assumed to have a price of zero and a time driven out of the way of zero.

While this model gives us an initial understanding of the station choices drivers make in the data, there is still a substantial amount of work required to estimate a model that we can take to the counterfactual simulations. In particular we are worried about two strong assumptions in this model: full information and static decision-making. The model currently assumes that drivers know the price at every potential station on every trip, an assumption that seems quite unlikely in reality.[2] We plan to investigate two ways of relaxing this assumption: the first is to use the history of driving to identify which stations the driver has passed recently and then use the price at the time the driver passed the station as the driver's assumption about the current price at the station. We could then assume that the driver assumes that prices at stations she hasn't passed recently are equal to a prediction of that station's price given current average prices and the station's average deviation from the overall average price. An alternative approach would be to use the recent innovation in Abaluck and Adams (2017) that seeks to empirically identify consideration sets using variation in the data. A combination of these two approaches would potentially allow us to get a better sense of how drivers trade off lower gas prices for driving further out of the way, which is the critical parameter for understanding how dense a network of alternative fueling stations would need to be to make alternative fuel vehicles attractive to drivers.

Additionally, it is likely that drivers' choice of whether to stop for gasoline on a given trip has a dynamic component insofar as a driver who knows she will be passing an inexpensive station later in the day is unlikely to stop for gas now. We are unsure whether incorporating some of these dynamics will require a full dynamic estimation or whether we can include additional controls in the "upper" nest of the nested logit that determines whether the driver stops for gas, but we need to explore dynamic issues substantially if we are going to believe our estimates.

Preliminary Results

With the above caveats, we can show the results our initial analysis of gas station choice, which suggest that we will get reasonable estimates of stopping behavior in a more thorough analysis. Table 1 presents the results of both the multinomial logit and the nested logit, where drivers are assumed to only consider stations that are less than 10 minutes out of the way when deciding where to stop. Each regression also includes brand fixed effects for the twelve largest brands and an "other" category for all other brands. The key result in Table 1 is that drivers do tend to choose stations with lower gas prices and that are less far out of their way. Distance driven out of the way is even more important during commute trips, and price is slightly less important during commutes, although not statistically significantly so. The nested logit regressions also suggest that drivers are less likely to stop for gas during a commute and more likely to stop as the tank level drops, although this effect is non-linear.

[2] In 2009 when this data was collected, applications such as "Gas Buddy" that provide information on current gas prices at different stations were quite rare.

Table 1: Gas Station Choice Results

	Conditional on stopping		Nested logit	
	(1)	(2)	(3)	(4)
Gas price ($/gallon)	-6.142*	-6.587*	-1.379*	-1.436*
	(0.656)	(0.727)	(0.199)	(0.217)
× Commute (0/1)		2.797		0.404
		(1.464)		(0.513)
Excess time (minutes)	-0.832*	-0.794*	-0.392*	-0.376*
	(0.045)	(0.047)	(0.030)	(0.029)
× Commute (0/1)		-0.322*		-0.163*
		(0.161)		(0.069)
Right-side arrive (0/1)	0.387*	0.387*	0.181*	0.182*
	(0.081)	(0.081)	(0.042)	(0.042)
No stop equation				
Commute (0/1)			0.153	1.090
			(0.101)	(1.345)
Tank level (liters)			-0.019	-0.019
			(0.015)	(0.015)
Tank level squared			0.0027*	0.0027*
			(0.0003)	(0.0003)
Observations	85200	85200	1549938	1549938
No. of choice groups	794	794	24529	24529

These results have the expected signs, but in order to understand the magnitudes of the estimates, we can scale the excess time coefficients by the price coefficient to calculate drivers implied value of time. These calculations suggest that drivers have extremely high values of time: $58 -$125/hour for non-commute driving. While these values of time are high, Houde (2012) also estimates a high value of time of $54/hour, and we would expect that these estimates may decline as we relax the assumption that drivers have perfect information about gasoline prices at all stations. If, however, this is actually the value of time that drivers use when considering whether to drive out of their way to purchase fuel, this suggests that the density of alternative fueling stations would have to be very high before drivers will prefer alternative fuel vehicles to gasoline vehicles.

Continuing Work

Once we have finalized our station choice estimates, we will conduct counterfactual simulations to understand how our drivers would value having the choice of a less extensive network of fueling stations. We can calculate the consumer surplus cost of having access to a less extensive fueling network, $a = \{0, 1, ..., A\}$ rather than the current gasoline fueling network, $j = \{0, 1, ..., J\}$, as:

$$E[CS(alt)_i] - E[CS(gas)_i] = \frac{1}{\alpha}\left[\sum_{t=1}^{T} \ln \sum_{a=0}^{A} V_{iat} - \sum_{t=1}^{T} \ln \sum_{j=0}^{J} V_{ijt}\right]$$
$$\text{where} \quad V_{iat} = \alpha p_{at} + X_{iat}\beta + \xi_{i0t}$$
$$\text{and} \quad V_{ijt} = \alpha p_{jt} + X_{ijt}\beta + \xi_{i0t}$$

We will this change in consumer surplus for many draws of different percentages of existing stations that will offer the alternative fuel to get the expected change in consumer surplus for driver i for

different alternative fueling station penetration rates. We can then create plots of how much less expensive an alternative fuel vehicle would need to be relative to a conventional vehicle to get $x\%$ of drivers in our sample to adopt at different alternative fueling station penetration rates.

This approach to understanding how drivers value fueling station density require a few substantial assumptions. First, we assume that existing gasoline station locations would offer the alternative fuel and continue to also offer gasoline, although this assumption can be relaxed if we had proposed alternative fueling station locations. Second, we need to make an assumption about what price (in terms of cost per mile) the alternative fuel would sell at. We will conduct our analysis with different alternative fuel costs, but given that we do not observe the universe of potential trips, we cannot calculate optimal markups for alternative fuel stations at different alternative fueling station densities. Finally, we are assuming that the experience of filling a vehicle with the alternative fuel will be similar to the current experience of filling a vehicle with gasoline. While this is likely true for alternative fuels like natural gas, it is currently less true for electric vehicle charging, which takes substantially longer than filling a tank with gasoline. Given substantial recent declines in the time required for a substantial vehicle charge, our estimates still seem relevant to electric vehicles if charging times continue to decline and more charging is done outside the vehicle owner's home.

Funding Need

Because of the novel data, detailed empirical analysis, extensive counterfactual simulations, and important policy question, we anticipate that this project will result in a paper that we will submit to at least a journal like the *American Economic Journal: Policy*. Both co-authors have published work in environmental economics broadly and the economics of fuel consumption specifically at journals of this level. A research assistant is necessary to move this project forward because of the substantial data work remaining to improve our measures of each driver's information set, to estimate models with alternative information sets and dynamic considerations, and to run the extensive counterfactual simulations necessary. While earlier support has allowed us to make as much progress as we have, funding from the *National Bureau of Economic Research* would be critical to completing this project in a reasonable timeframe.

References

Abaluck, Jason and Abi Adams, "What Do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses," NBER Working Paper #23566 2017.

Anderson, Soren T. and James M. Sallee, "Using Loopholes to Reveal the Marginal Cost of Regulation: The Case of Fuel-Economy Standards," *The American Economic Review*, June 2011, *101* (4), 1375–1409.

Greaker, Mads and Tom-Reiel Heggedal, "Lock-In and the Transition to Hydrogen Cars: Should Governments Intervene?," *The B.E. Journal of Economic Analysis & Policy*, 2010, *10* (1).

Houde, Jean-François, "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline," *American Economic Review*, August 2012, *102* (5), 2147–2182.

Li, Jing, "Compatibility and Investment in the U.S. Electric Vehicle Market," October 2017.

Shanjun, Tong Lang Xing Jianwei Li and Yiyi Zhou, "The Market for Electric Vehicles: Indirect Network Effects and Policy Design," *JAERE*, 2016, *4* (1), 89–133.

Zhou, Yiyi and Shanjun Li, "Technology Adoption and Critical Mass: The Case of U.S. Electric Vehicle Market," SSRN Scholarly Paper August 2016.