Fueling Alternatives: Evidence From Real-World Driving Data

Jackson Dorsey
Indiana University, Kelley School of Business

Ashley Langer
University of Arizona

Shaun McRae
Instituto Tecnológico Autónomo de México (ITAM)

May 2019
Typical American family will spend $1,991 on gas in 2019

Projection - Gas Buddy, Image - Track Gabe Blog
America's New Pollution King

Transportation emissions have surpassed electricity emissions for the first time since 1978

- Electricity emissions (metric tons of CO2)
- Transportation emissions

U.S. Energy Information Administration

Bloomberg
Gasoline, economics, and policy

- Gasoline remains a dominant transportation fuel and transportation now #1 source of CO₂
  - Policy and technology driven changes to the industry
    ▶ Fuel economy standards, gas taxes, rise of EVs/hybrids

- Therefore, researchers and policymakers interested in understanding consumer behavior in this market
  - Many theoretical and empirical works on demand/search
  - Due to data limitations, most of the literature has had to rely on aggregate data or strong modeling assumptions
This paper

- Driver’s choice about where/when to buy gas is complex
  - We use a unique data set to better understand how drivers decision of where/when to purchase gas
- First paper to use high-frequency micro data on drivers’ geographic locations and gasoline purchase behavior
  - We observe 600+ variables including:
    - the last station each driver refueled, stations recently passed, drivers’ current tank level, distance out of the way to each potential station
- We model drivers’ decision as a combination of:
  1. A choice of which stations to consider
  2. Which station to purchase from conditional on the consideration set
This paper

We then use our empirical model of driver behavior to evaluate:

- Drivers’ implied value of time
  - Crucial for knowing the required density an alternative fuel network
- Driver’s demand elasticity w.r.t. current prices vs. average prices
  - Key to understanding implications of fuel taxes and fuel economy standards
- The value of full information in gasoline markets
  - How much are drivers leaving on the table? This also provides an estimate of the cost of search in this mkt.
Literature - choice with imperfect information

- **Search Literature**
  - Online markets, where actual search behavior is observed (De los Santos, Hortacsu, and Wildenbeest, 2012). But, these are often not products that are purchased frequently or in such national volumes.
    - Other empirical search models: Hortacsu, Syverson (2004), Honka (2014), Salz (2017), and more

- **Choice Set Formation**

- **Hybrids**: papers that combine search, rational inattention, and choice set formation
Literature - gasoline demand

- Estimating elasticity of demand for gasoline using aggregate data

- Discrete choice with aggregate data

- Search in gasoline markets
  - Focused on search and consumer price expectations as generating price dispersion and “rockets and feathers” price movements.
    - Yang and Ye (2007), Lewis (2008), Tappata (2009), Chandra and Tappata (2011), and many others.
The IVBSS Experiment

- IVBSS (Integrated Vehicle-Based Safety System) was a $32 million field test of advanced crash-warning technology by the USDOT, industry partners, and the UM Transportation Research Institute (UMTRI)
- Sixteen identical passenger cars were fitted with the technology
- 108 drivers from southeast Michigan were given the vehicles to use for approximately six weeks
What data was collected during the experiments?

- Each car had a computer installed that recorded 600 variables at a rate of 10 times per second
  - Vehicle location, speed, acceleration, fuel use, etc
  - Detailed data from the crash warning systems

- Each car included five cameras (two in-car, three exterior)
Gas pump stops identified using combination of GPS tracks and in-car cameras
We identified over 700 vehicle stops at gas pumps
Pump stops matched to daily station-level price data to obtain gas price paid
People don’t drive out of their way to buy gas

We use this data to calculate the excess distance that driver $i$ would need to travel to get to station $j$ on trip $t$ and how long this would take.
Model of station choice

- On each trip, $t$, driver $i$ can stop at a set, $C$, of potential stations
  - $C$ includes all stations within 3 min. of driver’s route
    - 99.2% of stops are $< 3$ min. away
  - Drivers may not consider all of these stations

- We model the purchase decision in two stages:
  1. Drivers consider a subset $S \subseteq C$ of stations
     - Whether a driver considers a station $j$ can depend on vector $Z_{ijt}$ (i.e. has driver passed stn. recently)
  2. Drivers select a station $j$ from $S$, or the “outside option” of not stopping to maximize utility
     - A driver’s utility from choosing station $j$ depends on a vector $X_{ijt}$ (i.e. current station price)
Probability driver $i$ chooses $j$ on trip $t$:

$$
\text{Prob}_{itj} = \sum_{S \in C_j} \Pr(S|Z_{itj}, \theta) \times \Pr(j|X_{itj}, S, \beta)
$$

Sum over all choice sets that contain $j$

- The probability that driver considers $j$:
  $$
  \phi_{itj}(\theta) = \frac{\exp(Z_{itj}\theta)}{1 + \exp(Z_{itj}\theta)}
  $$

- The probability of consideration set $S$ occurring:
  $$
  \Pr(S|Z_{itj}, \theta) = \prod_{l \in S} \phi_{itl} \prod_{k \notin S} (1 - \phi_{itk})
  $$

- Given $S$, the choice rule follows a standard logit form
Estimation

We estimate the parameters via simulated maximum likelihood

- We find utility parameters, $\beta$, and consideration parameters, $\theta$, that best fit the observed station choices
- Large number of potential consideration sets for each trip
  - Avg. trip has 16 stations nearby, so $2^{16} = 65,536$ possible choice sets
- Therefore, we approximate the probability of a choice at each parameter by averaging over 100 “simulated choice sets”
How can we identify the probability that drivers consider each station?

- Suppose there are 2 stations and “outside option” of not stopping
  - Each station either sets a “high price” or “low price”
  - We see a panel of market shares for each station and the “outside option”
- There are 3 parameters to estimate:
  - $\beta_0$ - the ”constant” utility obtained from stopping at either of the stations
  - $-\beta_1$ - distaste from stopping at a “high price” station
  - $\theta$ - The probability of considering each station

**Identifying Assumption:** The “outside option” is considered with probability 1
Observation 1: low prices

These mkt. shares provide information about drivers’ utility from stopping ($\beta_0$) and how likely they are to consider each station ($\theta$)
Observation 2: differential prices

- These mkt. shares provide information about drivers’ sensitivity to price ($\beta_1$) and how likely they are to consider each station ($\theta$)
Observation 3: high prices

This pins down consideration, \( \theta \), given \( \beta_1, \beta_0 \).

**Intuition:** If fewer drivers substitute to the “outside option” than we would have predicted from observation 2, we infer that many drivers weren’t considering both stations.
Empirical Implementations

- Variables that influence consideration
  - All specifications: constant, tank level, \((\text{tank level})^2\)
  - Specification 1: excess distance to station
  - Specification 2: time since driver last passed station, last station chosen

- Variables that influence choice
  - All specifications: constant, current price, station avg. price, excess dist., right-side arrival
Results: consideration probabilities

- Consideration probabilities fall with distance

Driver 65, Trip 228, Tank level=72%
Results: choice probabilities

Driver 65, Trip 228, Tank level=72%
Consideration probabilities rise as tank level declines

Driver 6, Trip 74, Tank level=42%
Drivers more likely to consider recently passed stations
MUCH more likely to consider last chosen station

Driver 47, Trip 386, Tank level=35%
### Avg. marginal effects of determinants of consideration

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tank Level (L/10)</td>
<td>−0.093</td>
<td>0.004</td>
</tr>
<tr>
<td>(Tank Level)$^2$ (L/10)$^2$</td>
<td>0.004</td>
<td>−0.012</td>
</tr>
<tr>
<td>Excess Distance (min)</td>
<td>−0.033</td>
<td></td>
</tr>
<tr>
<td>Passed Last 7 Days (0/1)</td>
<td></td>
<td>0.014</td>
</tr>
<tr>
<td>Last Station Chosen (0/1)</td>
<td></td>
<td>0.102</td>
</tr>
<tr>
<td>E[Stations Considered]</td>
<td>1.09</td>
<td>0.76</td>
</tr>
<tr>
<td>E[Stations Considered</td>
<td>Purchase]</td>
<td>6.74</td>
</tr>
<tr>
<td>Num. of Trips</td>
<td>22,360</td>
<td>22,360</td>
</tr>
<tr>
<td>Observations</td>
<td>352,449</td>
<td>352,449</td>
</tr>
</tbody>
</table>

In a third specification, we also find that drivers consider more stations when wholesale prices are higher.
## Choice parameter estimates

<table>
<thead>
<tr>
<th>Choice of Station</th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Inside” good</td>
<td>-3.532***</td>
<td>-3.406***</td>
</tr>
<tr>
<td></td>
<td>(0.096)</td>
<td>(0.089)</td>
</tr>
<tr>
<td>Current Station Price ($/gal)</td>
<td>-0.360</td>
<td>-0.081</td>
</tr>
<tr>
<td></td>
<td>(0.322)</td>
<td>(0.347)</td>
</tr>
<tr>
<td>Average Station Price ($/gal)</td>
<td>-7.150***</td>
<td>-6.773***</td>
</tr>
<tr>
<td></td>
<td>(0.936)</td>
<td>(1.031)</td>
</tr>
<tr>
<td>Excess Distance (min)</td>
<td>-0.414***</td>
<td>-0.898***</td>
</tr>
<tr>
<td></td>
<td>(0.081)</td>
<td>(0.059)</td>
</tr>
<tr>
<td>Right-Side Arrival (0/1)</td>
<td>0.268***</td>
<td>0.266***</td>
</tr>
<tr>
<td></td>
<td>(0.091)</td>
<td>(0.097)</td>
</tr>
</tbody>
</table>

| Own Elasticity w.r.t. Current Price        | -0.913      | -0.203      |
| Own Elasticity w.r.t. Avg. Price           | -18.985     | -17.153     |

Drivers very sensitive to avg. prices, but not to current station prices.
Value of time and information

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>Logit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implied Value of Time ($/hr)</td>
<td>10.459</td>
<td>24.825</td>
<td>20.8699</td>
</tr>
<tr>
<td>Annual Value of Full Info ($/driver)</td>
<td>229.435</td>
<td>338.146</td>
<td>-</td>
</tr>
<tr>
<td>$\Delta$ CS from Full Info / Gas Expenditures</td>
<td>0.242</td>
<td>0.357</td>
<td>-</td>
</tr>
</tbody>
</table>

- These values of time are substantially smaller than existing estimates
  - $54$ per hour (Houde, 2012)
- Getting consideration sets right is crucial for value of time estimate
Value of time and information

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>Logit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Implied Value of Time ($/hr)</td>
<td>10.459</td>
<td>24.825</td>
<td>20.8699</td>
</tr>
<tr>
<td>Annual Value of Full Info ($/driver)</td>
<td>229.435</td>
<td>338.146</td>
<td>-</td>
</tr>
<tr>
<td>Δ CS from Full Info / Gas Expenditures</td>
<td>0.242</td>
<td>0.357</td>
<td>-</td>
</tr>
</tbody>
</table>

- Driver welfare would be substantially improved by better information about stations available
  - Lower prices, more convenient stops
  - 2nd col. likely an overestimate of information value if consideration correlated with unobserved quality (more work here)
Policy implications

Chargefox continues expansion of ultra-rapid electric car charging network
Policy implications

Chargefox continues expansion of ultra-rapid electric car charging network

Swiss company ABB has released a DC fast charger capable of recharging an EV nearly three times faster than Tesla's Supercharger... if only there was a car that could handle that kind of electron flow (Credit: ABB)

World's fastest EV charger gives drivers 120 miles in 8 minutes

Loz Blain | April 26th, 2018

April 15, 2019 | 3 MINUTE READ - BRIDIE SCHMIDT
Policy implications

Chargefox continues expansion of ultra-rapid electric car charging network

World’s fastest EV charger gives drivers 120 miles in 8 minutes

EVgo Goes Plaid With New Ultra-Fast Charging Station In Baker, California

Swiss company ABB has released a DC fast charger capable of recharging an EV nearly three times faster than Tesla’s Supercharger... if only there was a car that could handle that kind of power!
Policy implications

- Alternative fueling stations may not need to be as dense as existing stations to be competitive
  - Clear prices would provide a competitive advantage by reducing search costs.
  - Lower value of time than previous estimates reinforces this result (more work to do here).
  - Density can be even lower if alternative fuel is cheaper per mile.

- Information is critically valuable in improving drivers’ welfare.
  - Some of this will come by reducing stations’ profits.
  - Misallocation of drivers across stations causes a pure welfare loss.
  - Not clear how much this has been improved by “Gas Buddy” and the like.
Next steps

- Refine and better understand our estimates.
  - Allow station average price to influence consideration.
  - Improved modeling of unobservable station quality (e.g. last stop, brand, etc).
  - Improved modeling of quantity purchased at each stop: fillers vs. non-fillers.
  - Understand what affects the implied value of time and value of information.

- Potential other counterfactuals? Ideas?
Additional tables and figures
Consideration by tank level

![Graph showing probability of considering station by tank level, with two models: Mod 1 and Mod 2. The graph plots the probability of considering a station at the mean of Xs against tank level at trip start (L/10).]
Drivers are more likely to stop as their tank gets closer to empty.
Choice probabilities

Driver 47, Trip 386, Tank level=35%
## Avg. marginal effects of determinants of consideration

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Tank Level (L/10)</td>
<td>-0.310</td>
<td>-0.093</td>
<td>-0.531</td>
<td>0.004</td>
</tr>
<tr>
<td>Initial Tank Level Squared (L/10)^2</td>
<td>0.025</td>
<td>0.004</td>
<td>0.048</td>
<td>-0.012</td>
</tr>
<tr>
<td>Wholesale Price Rising (0/1)</td>
<td></td>
<td></td>
<td>-0.021</td>
<td></td>
</tr>
<tr>
<td>Wholesale Price ($/gal)</td>
<td></td>
<td></td>
<td>0.104</td>
<td></td>
</tr>
<tr>
<td>Excess Distance (min)</td>
<td></td>
<td>-0.033</td>
<td>0.125</td>
<td></td>
</tr>
<tr>
<td>Ever Passed</td>
<td></td>
<td></td>
<td>-0.001</td>
<td></td>
</tr>
<tr>
<td>Passed Last 7 Days</td>
<td></td>
<td></td>
<td>0.014</td>
<td></td>
</tr>
<tr>
<td>Passed Last 3 Days</td>
<td></td>
<td></td>
<td>0.015</td>
<td></td>
</tr>
<tr>
<td>Last Station Chosen (0/1)</td>
<td></td>
<td></td>
<td>0.102</td>
<td></td>
</tr>
<tr>
<td>E[Stations Considered]</td>
<td>3.05</td>
<td>1.09</td>
<td>5.6</td>
<td>0.76</td>
</tr>
<tr>
<td>E[Stations Considered</td>
<td>Purchase]</td>
<td>17.85</td>
<td>6.74</td>
<td>24.28</td>
</tr>
<tr>
<td>Num. of Trips</td>
<td>22,360</td>
<td>22,360</td>
<td>22,360</td>
<td>22,360</td>
</tr>
<tr>
<td>Observations</td>
<td>352,449</td>
<td>352,449</td>
<td>352,449</td>
<td>352,449</td>
</tr>
</tbody>
</table>
### Value of time and information

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>Logit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Own Elasticity w.r.t. Current Price</td>
<td>-1.015</td>
<td>-0.913</td>
<td>-2.344</td>
<td>-0.203</td>
<td>-0.759</td>
</tr>
<tr>
<td>Implied Value of Time ($/hr)</td>
<td>26.856</td>
<td>10.459</td>
<td>40.921</td>
<td>24.825</td>
<td>20.8699</td>
</tr>
<tr>
<td>Annual Value of Full Info ($/driver)</td>
<td>109.146</td>
<td>229.435</td>
<td>107.127</td>
<td>338.146</td>
<td>-</td>
</tr>
<tr>
<td>Δ CS from Full Info / Gas Expenditures</td>
<td>0.115</td>
<td>0.242</td>
<td>0.113</td>
<td>0.357</td>
<td>-</td>
</tr>
</tbody>
</table>