Estimates of the Demand for E85
Using Stated-Preference Data off Revealed-Preference Choices

Kenneth Liao
Iowa State University
kliao@iastate.edu

Sébastien Pouliot
Iowa State University
pouliot@iastate.edu

Selected Paper prepared for presentation at the 2016 Agricultural & Applied Economics Association Annual Meeting, Boston, Massachusetts, July 31-August 2

Estimates of the Demand for E85
Using Stated-Preference Data off Revealed-Preference Choices

Kenneth Liao*

and

Sébastien Pouliot†

May 2016

Abstract: This paper estimates the relative preferences of motorists for E10 and E85 from an intercept survey of motorists with flex-fuel vehicles at E85 fuel stations in Arkansas, California, Colorado, Iowa, and Oklahoma. The information collected includes prices observed at fuel stations, fuel choices by flex motorists, and responses to a series of opinion questions about ethanol and gasoline. We also proposed a hypothetical scenario to each motorist in which either the price of the fuel selected was increased or the price of the fuel not selected was decreased. We first estimate fuel preferences using the revealed preference data from the observed choices. We then use the stated preference data from the hypothetical price scenario to estimate preferences in empirical models that correct for endogeneity from unobservable demand shifters that carry over to the stated preference empirical model. We find that motorists significantly discount E85 compared to E10 even when accounting for the different energy content of the two fuels and that the distribution of willingness to pay for E85 does not vary significantly between regions, except for California where motorists are willing to pay significantly more for E85.

Keywords: Ethanol, Gasoline, Renewable Fuel Standard, Willingness to pay, Intercept survey.
JEL codes: Q18, Q41, Q42.
Acknowledgements: We acknowledge support from the Biobased Industry Center at Iowa State University and the National Science Foundation under Grant Number EPS-1101284. The views in this paper are attributable solely to the authors.

* Postdoctoral Research Associate, Department of Economics, Iowa State University. *Email*: kliao@iastate.edu.
† Associate Professor, Department of Economics, Iowa State University. *Email*: pouliot@iastate.edu.

1 Introduction

The second iteration of the Renewable Fuels Standard (RFS2) requires increasing quantities of ethanol and other biofuels to be blended into the motor fuel consumed in the United States each year. So far, meeting the ethanol requirement has been relatively easy because the vast majority of gasoline consumed in the United States is E10, which contains about 10 percent ethanol. The maximum quantity of ethanol that can be blended into the total pool of motor fuel through E10 is commonly referred to as the 'E10 blend wall'. In 2015, the United States consumed nearly 140 billion gallons of retail gasoline which means a blend wall of about 14 billion gallons of ethanol. The quantity of ethanol mandated by RFS2 is now reaching the point where it is set to surpass the blend wall.

The implied corn-ethanol mandate in RFS2 was originally scheduled to be 14.4 billion gallons in 2014, and 15 billion gallons in 2015 and 2016. [3] The Environmental Protection Agency (EPA) is responsible for setting the required biofuel volumes. On November 30, 2015, EPA released its final rule for the 2014, 2015, and 2016 renewable fuel volumes, lowering the implied corn-ethanol mandate to 13.61 billion gallons in 2014, 14.05 billion gallons in 2015, and 14.5 billion gallons in 2016. In May 2016, EPA proposed an ethanol volume of 14.8 billion gallons for 2017.

One solution to the blend wall is to use alternative gasoline blends that contain more than 10 percent ethanol such as E85, a gasoline blend that contains between 51 and 83 percent ethanol. On average, a gallon of E85 contains 74 percent ethanol, therefore each gallon of E85 consumed as a substitute for E10 increases aggregate ethanol consumption by 0.64 gallons (EIA 2015). Thus, ethanol consumption could exceed the blend wall if even a small fraction of motorists use E85 instead of E10. However, E85 consumption has historically been scant due to high prices and limited availability. The question is whether E85 provides a feasible pathway for compliance with the expanding biofuel mandates, and if so, how low would the E85 price have to be to entice enough consumption? This study estimates the relative preferences of motorists for E10 and E85 to better understand the aggregate demand for ethanol in the United States.

[3] Corn-ethanol refers to first-generation ethanol produced primarily from corn in the United States. The implied corn-ethanol mandate is the amount of total renewable fuel required by RFS2 minus the required amounts of cellulosic biofuel, biomass-based diesel, and advanced biofuel.

This study provides an important piece for policy analysis of the biofuel mandates. The results allow prediction of the share of motorists who choose E85 instead of E10 given fuel prices, a crucial part of understanding the demand for ethanol beyond the E10 blend wall. Estimates of motorists' willingness to pay (WTP) for E85 as a substitute for E10 can be used to understand the feasibility of expanding the mandates (e.g., Pouliot and Babcock 2014), and evaluate the welfare impacts of RFS2 (e.g., Anderson 2012, Pouliot and Babcock 2016).

Relatively little is known about the preferences of US motorists when it comes to using E85 as a substitute for E10 despite the importance for policy analysis. There is no comprehensive source of nationwide E85 sales data, E85 is only available at a limited number of retail fuel stations, and consumption is restricted to motorists who drive flexible-fuel vehicles (FFVs). Previous studies have estimated relative preferences for ethanol and gasoline for flex motorists in Brazil and Minnesota (where sales data are available) or have conducted nationwide mail and online surveys to collect stated preference (SP) data. A brief review of the literature is provided in the next section.

In this study, we collected data from E85 fuel stations in different regions across the United States by performing an intercept survey similar to the one conducted by Salvo and Huse (2013) in Brazil. With the intercept survey, we obtained revealed preference (RP) data by observing actual fuel purchases that we augment with SP data obtained by asking motorists questions while they refueled their vehicles. The SP data are used to increase the variation in the observed E85 and E10 prices. Specifically, we asked motorists if they would make the same fuel choice if the relative E85-E10 price had been some amount less favorable towards their chosen fuel.

We estimate fuel preferences using a random utility logit model given fuel prices, motorist characteristics, and station characteristics. Flex motorists with high willingness to pay for E85 may have selected themselves into our sample by choosing to drive out of their way to visit the E85 fuel station. We are able to inform whether motorists self-selected in our survey based on their responses to questions about fuel station choice and distance driven to access E85. We use the SP data with the wider range of prices to add precision. The empirical model for the SP data is different than the empirical model for the RP data because the hypothetical fuel prices that were presented to each motorist depended on the motorist's RP fuel choice. This creates an endogeneity problem because the motorists' unobservable demand shifters are correlated with the hypothetical prices.

We estimate fuel preferences based on the difference in fuel prices (the E85 premium) and the quotient of fuel prices (the E85 ratio). Both approaches yield similar results. We find that the distribution of WTP is spread over a wide range of relative fuel prices, and when E85 and E10 are priced equally on a cost-per-mile basis, 25 percent of flex motorists choose E85. The probability that a flex motorist chooses E85 is not significantly different between the regions where we conducted our survey, other than for stations in California, where we find that 75 percent of flex motorists choose E85 when it is priced evenly on a cost-per-mile basis. However, in California, several confounding factors, such as a different retail model, are apparent.

The next section of this paper offers background information about E85 and a review of the related literature. Section 3 contains details on the intercept survey design. We describe the theoretical models and estimation technique in Section 4. We discuss the survey data in Section 5, and we present our empirical models and estimation results in Section 6. Our conclusions are in Section 7.

2 Background and Literature

Most automobiles cannot accommodate gasoline blends with more than 10 or 15 percent ethanol by volume. FFVs can operate using a range of gasoline blends including E10, E85, and any combination of the two. Most FFVs are alternate versions of conventional vehicle models.[4] The operation of an FFV is identical to a conventional vehicle except that E85 yields lower fuel economy than E10 because ethanol has lower energy content per volume than gasoline. Ethanol contains about two-thirds of the energy of gasoline so an FFV running on E85 gets between 75 and 80 percent as many miles per gallon compared to E10, depending on the specific vehicle and the exact concentration of ethanol in the E85, which can vary across states and seasons. In many cases, consumers are not able to acquire a certain vehicle make and model in anything but the

[4] Until 2014, automobile manufacturers received a substantial credit from the US Corporate Average Fuel Economy (CAFE) standards for producing FFVs. The credit started declining in 2015 and will cease in 2020. Under the rule, up to an annual limit, FFVs were treated as though they were operated partially on E85, but the fuel economy was calculated as the total miles the vehicle could travel per gallon of gasoline input (the ethanol fuel input was excluded in the fuel economy calculation). The result is that the majority of FFVs in the United States today are large sedans, SUVs, pickup trucks, and minivans, and they are mostly manufactured by American automobile companies.

FFV version or are initially unaware that they have purchased an FFV. Thus a motorist's decision to purchase an FFV is often independent of ethanol preference and price.

Most retail fuel stations do not supply E85 because it requires a dedicated underground storage tank and the pumps that dispense E85 require modifications to withstand the greater corrosive properties of ethanol. Currently less than 3 percent (about 2,700) of retail fuel stations offer E85 in the United States, and the highest concentration of E85 stations is in the Midwest (AFDC 2015).

Efforts to understand the demand for E85 in the United States have been somewhat hindered by the lack of data on the consumption of E85. One potential alternative is data for Brazil where more than half of vehicles are FFVs, and retail fuel stations offer both pure ethanol and a gasoline-ethanol blend called gasohol. Salvo and Huse (2013) collect fuel preference data using an intercept survey of flex motorists in Brazil. Salvo and Huse (2013) find that after adjusting for the difference in energy, about 20 percent of flex motorists choose ethanol when ethanol is priced 20 percent higher than gasoline, and 20 percent of flex motorists choose gasoline when gasoline is priced 20 percent higher than ethanol.

For US motorists, the best available data on E85 sales come from a monthly survey of E85 stations in Minnesota conducted by the Minnesota Department of Commerce. Anderson (2012) estimates the distribution of preferences for E85 using Minnesota data from between 1997 and 2006. During that time period, the energy-adjusted price of E85 was almost always greater than the price of E10. As a result, Anderson (2012) is unable to recover the full distribution of willingness to pay for E85 and instead estimates the upper tail of the distribution where the energy-adjusted price of E85 is higher than the price of E10, and only flex motorists with high WTP for E85 use it.

Corts (2010) recognizes that most of the early Minnesota data represent E85 use by government fleet vehicles and tests whether government fleet FFV mandates encourage retail fuel stations to invest in E85 fueling infrastructure and whether increased availability of E85 increases motorist demand for FFVs. Corts (2010) shows that government fleet adoption of FFVs led to an increase in the number of retail E85 stations, but concedes that the second hypothesis cannot be tested due to limitations of the data. Specifically, Corts (2010) notes that most FFVs in the dataset were purchased prior to the widespread availability of E85 and that motorists may not even know of the vehicles' capabilities.

Recent studies have used nationwide mail and online surveys to obtain stated preference data on WTP for E85. An advantage of hypothetical choice experiments is that they allow researchers to 'observe' consumer choices between products and/or attribute levels that may not exist in the market. Jensen et al. (2010) emphasize in their study the feedstock used to produce the ethanol and estimate motorists' WTP for E85 from corn, E85 from switchgrass, and E85 from wood. Jensen et al. (2010) find that consumers are willing to pay a premium to use E85 from switchgrass instead of E10 made with corn-ethanol. When it comes to E85 from corn versus E10 from corn, Jensen et al. (2010) find that some motorists discount E85 for perceived 'food versus fuel' reasons, while other motorists discount E10 for concerns about fuel security and the environment.

Petrolia et al. (2010) use a nationwide contingent valuation survey to identify the drivers of the demand for E85. Petrolia et al. (2010) find that the overall perception of ethanol is positive, and the majority of motorists perceive ethanol to have a positive influence on the environment, the economy, and on national security. Aguilar et al. (2015) use a discrete-choice experiment to estimate motorist preferences for E0, E20, and E85. Aguilar et al. (2015) find that the average motorist prefers to refuel with ethanol and that if the cost per mile were the same for E85 and E10, then E85 would dominate the market. Their study also found that about 20 percent of motorists surveyed indicated strong unwillingness to buy fuel with any ethanol.

3 Intercept Survey Design

We designed the intercept survey to obtain data on a broad range of factors that might affect flex motorists' WTP for E85 as a substitute for E10. The survey collected motorists' fuel choices, and motorists were interviewed while they were refueling. We completed each interview in about two minutes. The complete questionnaire and appendices accompanying this paper are available to the interested reader by contacting the authors.

Intercept Survey Method

For each station we visited, we recorded station-level data such as the station name and brand, the station address, the prices of the E10 fuels (usually regular, midgrade, and premium), and the price of E85. The procedure used to choose which flex motorists to interview was that whenever the interviewer was idle, the interviewer targeted the next flex motorist to pull alongside any of the station's pumps. If a second flex motorist pulled up to a pump while an interview was being conducted, then the interviewer did not interview the second flex motorist. Instead, when the interviewer completed the first interview, the interviewer reset and once again waited to target the next flex motorist to pull alongside any of the station's pumps. This sequencing rule avoided possible selection bias by the interviewer. In practice, the share of the vehicle fleet that are FFVs is small and the survey was quick, and as such, despite the strict sequencing rule, we managed to capture virtually all flex vehicles refueling at the E85 stations during our visits.

We visually identified FFVs in two ways. First, many newer FFVs have some sort of badge on the back (or in rare cases the side) of the vehicle indicating that they are FFVs. Second, most FFVs have a yellow gas cap, a yellow ring, or a yellow sticker inside the gas door indicating that the vehicle is an FFV capable of using E85. In practice, identifying FFVs required the interviewer to pace around the pumps and closely inspect vehicles as they were refueling, but over the entire course of the data collection, it was never a problem. In general, a third way to tell whether a vehicle was an FFV was if the motorist chose E85, but it could be that the motorist was making a mistake (by choosing E85 for a conventional vehicle not equipped to use it) or had a vehicle with aftermarket modifications to use E85.[5]

Survey Questions

Before intercepting a motorist, the interviewer passively observed the motorist's fuel choice and characteristics. The recorded motorist's characteristics were: the vehicle make, the vehicle model, the vehicle type (car, truck, SUV, or van), the state on the license plate, whether the vehicle had an FFV badge, whether the vehicle had a yellow gas cap, and the gender of the motorist. The interviewer also recorded the transaction volume and expenditure when the motorist finished refueling.

Once a motorist began refueling, the interviewer approached and asked whether the motorist was willing to participate in a short survey. The interviewer then followed with a series

[5] Over the course of conducting the survey, we learned that a small share of motorists have aftermarket modifications to conventional vehicles (not originally manufactured as FFVs) to use E85 because the higher octane content can improve the vehicles' performance. In most cases, the vehicles are modified so that they can use either E85 or E10, but in rare cases the vehicles are configured so that they can only use E85, and switching back to E10 requires modifying the vehicle.

of questions about the motorist's awareness of E85 and opinions on topics that might explain the motorist's fuel choice. Details about these questions are provided in the appendix.

We wanted to know if the motorists we surveyed were randomly drawn from the general population of flex motorists or if they ended up in our sample only because they sought out E85. For the motorists who chose E10, we know that they did not come specifically for the E85, and we assume that they would still have chosen to refuel at the station even if every station offered E85. These flex motorists were randomly sampled from the local population of flex motorists. But for motorists who chose E85 we asked, "Did you choose to fuel at this station because it offers E85?" If they responded positively, we followed by asking, "How far out of your way did you have to drive?" We use responses to these questions to sort the motorists who chose E85.

We obtain SP data to complement the RP data by proposing a hypothetical price scenario and asking whether motorists would have (hypothetically) made the same fuel choice. If the motorist refueled with E10, the scenario was one in which either the price of E10 was increased or the price of E85 was decreased. If the motorist refueled with E85, the scenario was one in which either the price of E85 was increased or the price of E10 was decreased. The amount of the price change was \$0.25, \$0.50, or \$0.75 per gallon.

4 Theoretical Framework and Estimation Technique

The theoretical model motivates our empirical approach and corrects for the endogeneity of the hypothetical prices in the SP setting. We develop a model in which motorists make their fuel decisions based on the premium for E85 measured as the difference between the nominal price of E85 and the price of E10. In an alternative model, not shown in the interest of space, motorists make their fuel decisions based on the ratio of the nominal price of E85 and the price of E10.[6] We estimate both the price premium and the price ratio models and compare the fit of the models to determine whether one of the two decision rules dominate.

Model of Motorists' Fuel Choice in the RP Setting

The theoretical model is concerned with the fuel choices flex motorists make rather than the quantity of fuel they purchase. We let the demand for fuel be perfectly inelastic in the short run

[6] Consistent with the theoretical model, the price variable that enters the empirical model in the version with the E85 ratio is the log of the E85 ratio.

so that motorists choose either E85 or E10 based on relative prices, but the amount of fuel they purchase is price-independent.

Throughout, we use subscript e to denote E85 and subscript g to denote conventional E10. The indirect utility that motorist i derives from consumption of fuel $j \in \{e, g\}$ is

$$V_{ij}\left(I_i, p_{ij}, d_{ij}, \boldsymbol{x_i}, \varepsilon_{ij}\right) = v_{ij}\left(I_i, p_{ij}, d_{ij}, \boldsymbol{x_i}\right) + \varepsilon_{ij}. \tag{1}$$

where I_i is the motorist's income, p_{ij} is the nominal (not energy-adjusted) price of fuel j for motorist i, d_{ij} is the distance motorist i drives to access fuel j, $\boldsymbol{x_i}$ is a vector of characteristics about the motorist and the fueling station, and ε_{ij} is an unobservable stochastic demand shifter specific to the motorist and fuel choice. We normalize the distance driven to access E10 to zero to reflect that E10 is available at every fuel station and finding it involves zero search cost. On the other hand, E85 is not available at every fuel station, and motorists might have to incur some driving cost to access E85. We assume that ε_{ij} is a type 1 generalized extreme value random variable so that the difference between ε_{ig} and ε_{ie} follows a logistic distribution.

We choose a linear functional form for $v_{ij}(\cdot)$ whereby

$$v_{ij}\left(I_i, p_{ij}, \boldsymbol{x_i}\right) = \gamma_j I_i + \alpha_j p_{ij} + \delta_j d_{ij} + \boldsymbol{x_i}'\boldsymbol{\beta_j}.$$

We assume $\gamma_e = \gamma_g \equiv \gamma$ meaning that additional income affects the indirect utility in the same way regardless of fuel choice. We also assume that $\alpha_e = \alpha_g \equiv \alpha$. The intuition is that motorists' fuel choices do not respond to individual fuel prices but rather to the difference in the fuel prices. We let $\boldsymbol{\beta} \equiv \boldsymbol{\beta_e} - \boldsymbol{\beta_g}$, and we define $p_i \equiv p_{ie} - p_{ig}$ to be the E85 premium observed by motorist i. A motorist chooses E85 if $V_{ie}(\cdot) \geq V_{ig}(\cdot)$ which amounts to

$$\alpha p_i + \delta_e d_{ie} + \boldsymbol{x_i}'\boldsymbol{\beta} \geq \varepsilon_i,$$

where $\varepsilon_i \equiv \varepsilon_{ig} - \varepsilon_{ie}$ is symmetric with a mean of zero and follows a logistic distribution. To simplify the notation, we define $\boldsymbol{w_i} \equiv (p_i, d_{ie}, \boldsymbol{x_i})$ and $\boldsymbol{\theta} \equiv (\alpha, \delta_e, \boldsymbol{\beta})$. Then,

$$\Pr(\text{E85}_i) = \Lambda(\boldsymbol{w_i}'\boldsymbol{\theta}), \tag{2}$$

and

$$\Pr(\text{E10}_i) = 1 - \Lambda(\boldsymbol{w_i}'\boldsymbol{\theta}), \tag{3}$$

where $\Lambda(\cdot)$ is the cumulative logistic distribution.

Likelihood Equation

We estimate the coefficients in equations (2) and (3) by maximum likelihood. We define the dependent variable for motorist i:

$$y_i = \begin{cases} 1, & \text{if fuel choice is E85;} \\ 0, & \text{if fuel choice is E10.} \end{cases}$$

Under standard assumptions, the components of $\boldsymbol{w_i}$ are exogenous with respect to y_i so we can consistently estimate $\boldsymbol{\theta}$ using the conditional MLE that maximizes

$$\text{L} = \prod_{i=1}^{n} f(y_i|\boldsymbol{w_i}, \boldsymbol{\theta}) = \prod_{i=1}^{n} [1 - \Lambda(\boldsymbol{w_i}'\boldsymbol{\theta})]^{1-y_i}[\Lambda(\boldsymbol{w_i}'\boldsymbol{\theta})]^{y_i}.$$

The log-likelihood equation is

$$\ln \text{L} = \sum_{i=1}^{n} \ln f(y_i|\boldsymbol{w_i}, \boldsymbol{\theta}) = \sum_{i=1}^{n} \{(1 - y_i) \ln[1 - \Lambda(\boldsymbol{w_i}'\boldsymbol{\theta})] + y_i \ln[\Lambda(\boldsymbol{w_i}'\boldsymbol{\theta})]\}. \qquad (4)$$

We have ignored the sample selection problem. We sample from flex motorists who refuel at E85 stations rather than the general population of flex motorists. Thus parameter estimates that maximize (4) using all of our observations will be biased to mimic the sample and not the population.

Identification

The survey was not conducted on a random sample of motorists. Rather, the survey was conducted at stations that offered E85 with flex motorists who chose to refuel at that particular station. This causes two closely related identification issues for our empirical model.

The first issue is about the identification of the coefficient for the distance driven to access E85. Recall that a motorist chooses E85 if $\alpha p_i + \delta_e d_{ie} + \boldsymbol{x_i}'\boldsymbol{\beta} \geq \varepsilon_i$. The distance driven by motorists who choose E10 is zero because it is accessible at every fuel station. The distance is positive only for some motorists who select E85. The implication is that in our dataset, distance is positively correlated with motorists' decisions to refuel with E85. In a random sample of motorists, we would observe motorists who do not choose to refuel with E85 because the distance to access E85 is too great, and the correlation between the distance and the probability of refueling with E85 would be negative.

The second related identification issue is that motorists might select to drive to a fuel station because it offers E85. This means that the probability that a randomly drawn flex motorist from the population chooses E85 is less than the probability that we observe a flex motorist choose E85 in our sample.

A solution to the two identification issues is to use only observations from motorists who are representative of the population of flex motorists because their patronage at the fuel stations where we surveyed was not conditional on the offering of E85. These motorists did not incur any costs for traveling to the fuel station such that excluding the variable for the distance driven will not bias estimated coefficients.

The question we must answer is, which motorists in our dataset who chose E85 were randomly sampled? The survey contains two questions that inform self-selection. The first question asked whether the motorists came to the fuel station because it offered E85. The second question (asked if motorists answered yes to the previous question) asked how far out of their way they drove. A possible problem with these questions is their interpretation by motorists. It is possible that motorists answer that they came to the fuel station because of E85, but they would have chosen the same fuel station if every station offered E85. Indeed, several motorists indicated that they specifically chose the fuel station because it offered E85 but that the distance driven out of their way was zero. Nonetheless, some of the motorists who answered a distance equal to zero may have chosen a different fuel station if every fuel station offered E85.

The two questions that inform self-selection offer two possible ways to select a potentially random sample of motorists. In the strictest sample, we exclude all observations from E85 motorists who responded that they came to the station because of E85. In a more inclusive sample, we add observations from E85 motorists who say they came for the E85 but did not drive out of their way at all or drove out of their way by some short distance no more than one mile. Estimates from these different samples inform upper and lower bounds on the actual population shares depending on how motorists interpreted the questions.

Model Incorporating Stated Preferences

SP data have been used to complement RP data in previous studies, typically to increase the number of observations and expand the choice set to include some alternative(s) not actually available in the market. In our study the RP data have only modest variation in fuel prices, and the SP data are generated to expand that variation. The traditional method for estimating models on combined RP and SP data used in the transportation and the environmental economic literature has been described by Ben-Akiva and Morikawa (1990), Hensher and Bradley (1993), Adamowicz et al. (1994), and Hensher et al. (1999).

The intuition behind the traditional approach to combining RP and SP data is that the unobserved factors are different for the two types of data. This is typically implemented by allowing for different intercept and scale parameters for the distributions of error terms in the SP setting and the RP setting. The traditional approach is appropriate for combining RP data from observed market transactions with SP data from a survey when the attributes of the hypothetical choices are independent of the RP choice so that respondents' unobservable characteristics are not correlated with the hypothetical options.

Train and Wilson (2008) and Train and Wilson (2009) considered SP data constructed from RP choices, called 'SP-off-RP', and presented a model that is appropriate for our study. The important distinction from the traditional approach is that the same unobserved factors that affect a motorist's fuel choice in the RP setting are present in the SP setting. Since a motorist's initial RP fuel choice depends on both observed characteristics and unobservable factors, the unobserved factors are not independent of the motorist's choice and subsequent hypothetical prices offered. The non-independent unobserved factors persist in the hypothetical SP scenario, and we need to account for this when incorporating the SP data.

We define the utility that flex motorist i derives from fuel j in the SP setting as

$$W_{ij}\left(I_i, \tilde{p}_{ij}, \boldsymbol{x_i}, \varepsilon_{ij}, \eta_{ij}\right) = V_{ij}\left(I_i, \tilde{p}_{ij}, \boldsymbol{x_i}, \varepsilon_{ij}\right) + \eta_{ij} = \gamma_j I_i + \alpha_j \tilde{p}_{ij} + \boldsymbol{x_i}'\boldsymbol{\beta_j} + \varepsilon_{ij} + \eta_{ij},$$

where $\tilde{p}_{ij}$ is the hypothetical price of fuel j presented to motorist i, η_{ij} is a generalized extreme random variable with scale $(1/\zeta)$ that captures additional unobservable aspects of the SP setting not present in the RP setting. Note that in the SP setting, the relationships between both the observable and unobservable factors that determine the RP choice $V_{ij}(\cdot)$ in equation (1) are preserved. This means the SP data help inform the coefficient estimates, and the same unobservable ε_{ij} term for motorist i that affects the RP setting carries forward to the SP setting. The total unobservable error term in the SP setting is $\varepsilon_{ij} + \eta_{ij}$, where ε_{ij} also affects the motorist's RP choice which is used by the interviewer to generate the hypothetical prices offered to the motorist $\tilde{p}_{ij}$. Thus the hypothetical prices in the SP setting are endogenous because the hypothetical prices are correlated with the total error term given by $\varepsilon_{ij} + \eta_{ij}$. This can also be thought of as a missing variable problem where the missing correlated variable is the motorist's unobservable demand shifter from the RP setting.

The motorist chooses E85 in the SP setting if $W_{ie}(\cdot) \geq W_{ig}(\cdot)$ which we can re-write as

$$\eta_i \leq \zeta\big[V_{ie}(I_i, \tilde{p}_{ie}, \boldsymbol{x_i}, \varepsilon_{ie}) - V_{ig}\big(I_i, \tilde{p}_{ig}, \boldsymbol{x_i}, \varepsilon_{ig}\big)\big],$$

where $\eta_i \equiv \zeta\big(\eta_{ig} - \eta_{ie}\big)$ is symmetric with a mean of zero and follows a logistic distribution. The ζ term normalizes the logistic distribution of η_i to have a scale of one. Then the probability that a motorist chooses E85 in the SP setting is

$$\Pr(\text{SP E85}_i) = \Lambda\big(\zeta(\alpha\tilde{p}_i + \boldsymbol{x_i}'\boldsymbol{\beta} - \varepsilon_i)\big),$$

and the probability that a motorist chooses E10 in the SP setting is

$$\Pr(\text{SP E10}_i) = 1 - \Lambda\big(\zeta(\alpha\tilde{p}_i + \boldsymbol{x_i}'\boldsymbol{\beta} - \varepsilon_i)\big).$$

For a single motorist, the joint probability of a specific RP and SP choice combination is the product of the marginal probability of the RP choice and the probability of the SP choice, conditional on the RP choice. Let $y_i = (y_{i1}, y_{i2})$, where

$$y_{i1} = \begin{cases} 1 & \text{if RP fuel choice is E85;} \\ 0 & \text{if RP fuel choice is E10,} \end{cases}$$

and

$$y_{i2} = \begin{cases} 1 & \text{if SP fuel choice is E85;} \\ 0 & \text{if SP fuel choice is E10.} \end{cases}$$

The SP-off-RP likelihood function for the entire sample is

$$\text{L} = \prod_{y_i=(0,0)} \Pr(\text{RP E10}_i)\Pr(\text{SP E10}_i|\text{RP E10}_i) \prod_{y_i=(0,1)} \Pr(\text{RP E10}_i)\Pr(\text{SP E85}_i|\text{RP E10}_i)$$
$$\prod_{y_i=(1,0)} \Pr(\text{RP E85}_i)\Pr(\text{SP E10}_i|\text{RP E85}_i) \prod_{y_i=(1,1)} \Pr(\text{RP E85}_i)\Pr(\text{SP E85}_i|\text{RP E85}_i).$$

The ε_i's that enter the conditional SP logits are not observed but we know their conditional distributions so we can integrate over the density to calculate the expected value of the logits given the correlated errors. For example, the logit probability of a motorist choosing E85 in the SP setting conditional on that motorist choosing E85 in the RP setting is

$$\Pr(\text{SP E85}_i|\text{RP E85}_i) = \int \Lambda\big(\delta(\alpha\tilde{p}_i + \boldsymbol{x_i}'\boldsymbol{\beta} - \varepsilon_i)\big) f(\varepsilon_i|\varepsilon_i \leq \alpha p_i + \boldsymbol{x_i}'\boldsymbol{\beta}) d\varepsilon_i.$$

The integrals are evaluated using simulation where draws of ε_i are taken from its conditional density, the logit probability $\Lambda(\cdot)$ is calculated for each draw, and the results are averaged. We generate draws for ε_i following the method described by Train and Wilson (2009).

The parameters are estimated using maximum likelihood estimation over the joint probability log-likelihood function.

$$\begin{aligned}\text{RP SP MLE} = \sum_i \{ &(1-y_{i1})(1-y_{i2})\ln[\Pr(\text{RP E10}_i)\cdot\Pr(\text{SP E10}_i|\text{RP E10}_i)] \\ &+(1-y_{i1})y_{i2}\ln[\Pr(\text{RP E10}_i)\cdot\Pr(\text{SP E85}_i|\text{RP E10}_i)] \\ &+y_{i1}(1-y_{i2})\ln[\Pr(\text{RP E85}_i)\cdot\Pr(\text{SP E10}_i|\text{RP E85}_i)] \\ &+y_{i1}y_{i2}\ln[\Pr(\text{RP E85}_i)\cdot\Pr(\text{SP E85}_i|\text{RP E85}_i)]\}.\end{aligned} \tag{5}$$

5 Data Collection and Summary Statistics

We obtained the cooperation of two E85 retailers to conduct our survey. We collected a total of 972 observations from 17 E85 stations in 6 urban areas between October 2014 and April 2015.[7] In chronological order, the urban areas we visited were Des Moines, IA (DM), Colorado Springs, CO (CS), Tulsa, OK (TS), Little Rock, AR (LR), Sacramento, CA (SAC), and Los Angeles, CA (LA). We personally collected most of the observations, and we also used a small team of undergraduate students, trained by the authors, to help collect some of the observations in Des Moines. In each urban area, we visited between two and four different stations and collected around 100 or more observations. Unfortunately, we observed little variation in fuel prices within an urban area during the time we were there. All of the E85 stations we visited in DM, CS, TS, and LR were operated by a retailer we will call 'Retailer A', and all of the E85 stations we visited in SAC and LA were operated by a retailer we will call 'Retailer B'.

Observed Data and Survey Responses

From the initial 972 observations of motorists refueling their FFVs, we remove 91 observations where the motorists chose not to/were unable to complete/participate in the survey, a total non-response rate of 9 percent. That leaves us with an initial estimation sample of 881 observations. Table 1 summarizes the fuel choice data broken down by station, region and retailer. In the entire sample of 881 flex motorists, the average E85 price was \$2.19 per gallon, and the average E10 price was \$2.58 per gallon. Therefore, the average E85 premium (defined as the E85 price per gallon minus the E10 price per gallon) was −\$0.39. The average E85 ratio (defined as the E85

[7] We collected a total of 994 observations, but 22 were from vehicles with aftermarket modifications to use E85. These observations are excluded because the vehicles can only be identified as flex when motorists choose E85.

price divided by the E10 price) was 0.85. Overall, 431 (49 percent) of the flex motorists chose E85 while 450 (51 percent) chose E10.[8] Fuel prices, the share of flex motorists who chose E85, and opinions on ethanol varied considerably across the different regions we visited. For the most part, motorists in DM, LA, and SAC had a relatively high opinion of ethanol and E85 while opinions were less positive in CS, LR, and TS.

In general, fuel prices were most favorable towards E85 at Retailer B's stations in California where on average the E85 premium was −$0.54, and the E85 ratio was 0.83. We observed 231 flex motorists refueling at Retailer B's locations, and 89 percent chose E85. Retailer B's E85 prices were not drastically more favorable than Retailer A's E85 prices, but each of Retailer B's pumps served a larger share of the local E85-choosing community of flex motorists because E85 stations were less common in California. Also, Retailer B ran promotions providing special fuel cards and other incentives to local flex motorists, marketing E85 as a clean-burning, high-performance fuel.

For Retailer A, the fuel prices were most favorable towards E85 in DM, where the average E85 premium was −$0.47, and the average E85 ratio was 0.83, the same as the average price ratio observed at Retailer B's stations. Absolute fuel prices were higher in LA and SAC, so the E85 premium in those areas was larger in magnitude. The share of flex motorists who chose E85 among DM flex motorists was 42 percent, less than half of what we observed at E85 stations in LA or SAC. We suspect that one reason for the difference is that stations that offer E85 are more common in Retailer A's areas. Thus local flex motorists with high WTP for E85 can choose between a few E85 stations and will not all be observed in the sample.

Figure 1 plots the E85 premiums faced by flex motorists in our sample and the share of motorists who chose E85 at each premium, and Figure 2 plots the same share of motorists, but the relative prices are the E85 ratios. The figures are similar and show a notable shift between Retailer A and Retailer B. The downward-sloping curves suggest that flex motorists do indeed respond to relative fuel prices, and that price effects could dominate any potential regional effects. We formally investigate and discuss these notions in the next section.

Of the 450 motorists who chose E10, 87 (19 percent) stated they would switch to E85 in the SP setting if given the more favorable relative E85 price. Of the 431 motorists who chose

[8] Among the 450 flex motorists who chose E10, 414 (92 percent) chose regular grade 87 octane (85 octane in CS), 24 (5 percent) chose midgrade, and 12 (3 percent) chose premium.

E85, 201 (47 percent) stated they would switch to E10 if given the less favorable relative E85 price. Table 2 summarizes the RP and SP fuel choice data. One factor that is possibly driving the asymmetric switching behavior is ignorance of E85 by flex motorists who chose E10. Many had never used E85, and some were not aware they had an FFV or that the station offered E85. Thus many of the flex motorists who chose E10 were not responding to the relative E85 price.

Figure 3 shows the share of motorists who chose E85 in the SP setting given the hypothetical E85 premium offered to them and given their choice in the RP setting. Figure 4 shows the analogous SP E85 shares with respect to the E85 ratio. The figures show that in both the RP and SP settings, in general, a higher share of motorists chose E85 when the E85 price was more favorable. The figures also show that motorists who chose E10 in the RP setting were more likely to choose E10 in the SP setting even when presented with a relatively favorable E85 price. Likewise, motorists who chose E85 in the RP setting were more likely to choose E85 in the SP setting even when presented with a relatively unfavorable E85 price.

6 Empirical Models and Results

In this section we describe the empirical models, their results, and the implications. We also compare how well the models fit the data.

Empirical Models

We begin by estimating the RP-only versions of the E85 premium and E85 ratio models with four different estimation samples. Models 1.1 to 1.4 use the E85 premium, and Models 2.1 to 2.4 use the E85 ratio. In Models 1.1 and 2.1, we include all observations and do not correct for sample selection. In Models 1.2 and 2.2, we use the strictest sample selection rule where we remove observations from motorists who chose E85 and responded that 'Yes', they came to the station because it offered E85. More than 90 percent of the motorists who chose E85 answered 'Yes' to this question. The resulting estimation sample is 479 observations where 29 (6 percent) chose E85. In Models 1.3 and 2.3, we include observations from motorists who choose E85 and say that they came for the E85 as long as they did not drive out of their way at all. This increases the sample size to 670 observations. In Models 1.4 and 2.4, we extend the sample further to a total of 714 observations to include motorists who came for E85 and drove out of their way as long as they did not drive out of their way a distance greater than one mile.

Each of the eight models (viz., Models 1.1 to 2.4) uses the following vehicle, motorist, and station characteristics as explanatory variables: vehicle ownership (personal, government, company, other), vehicle type (car, truck, SUV, van), whether the vehicle had an FFV badge, number of miles driven per year, gender, age, opinions about which fuel is better for the environment, the engine, the economy, or national security, opinion on which fuel yields more miles per gallon, and the state where the station was located (AR, CA, CO, IA, and OK). Variables that describe the characteristics of the fuel stations are not added to the model but the state dummies summarize most of that information.

Models 1.1 to 2.4 are informative for identifying the various drivers of E85 demand in different samples. However we also want to learn whether preferences differ across locations without controlling for motorists' characteristics. If we find that the distribution of preferences is not significantly different, then we can conclude that estimates of E85 preferences from one state inform E85 preferences from other states. Thus we estimate Models 3.1 to 4.4 which are analogous to Models 1.1 to 2.4 but include only the relative E85 price variable and the state dummies (and not vehicle ownership, age, gender, opinions about fuels, etc.).

We use the SP data to estimate three variations of the RP-data-only models where the sample of motorists is based on a distance driven "out of their way" to the fuel station equal to zero (Models 1.3 and 2.3). The new models are 5.1, 5.2, and 5.3 for the E85 premium and 6.1, 6.2, and 6.3 for the E85 ratio. The first two variations are to view and compare bias, and the third variation is our preferred SP-off-RP model. In the first variation of the model, we estimate the logit model from equation (4), but we use only the SP data. In the second variation, we pool the SP and RP data together, treating all observations the same, and estimate the same model again. In these first two models, all of the differences between the RP data and SP data are ignored, and so the endogeneity problem is prominent.

The third variation is the model of Train and Wilson (2008) given in equation (5) and our preferred method where we appropriately model the nature of the SP-off-RP data-generating process. This variation of the model allows the SP data to inform the parameter estimates and to add precision while still allowing separate error terms and accounting for the correlation between the hypothetical prices that motorists are offered and the unobservable factors that influence their RP and SP choices. To estimate the model, we simulate one thousand conditional draws of the unobservable error terms for each motorist using the method described in Section 4.

Results

We focus the discussion of the results on the marginal effects of price and location variables and properties of the WTP distribution. The full sets of coefficient estimates, marginal effects, standard errors, and p-values for each model estimated are available in the appendix. Robust standard errors for the coefficients are calculated using a sandwich estimator as described by Cameron and Trivedi (2005 p. 828). In brief, we find that the significant factors affecting the probability a flex motorist chooses E85 (other than the fuel prices) are: vehicle ownership (personal, government, company, other), motorist opinions about which fuel is better for the environment, the engine, the economy, miles per gallon, and whether the motorist is in one of Retailer A's areas or one of Retailer B's areas. Other factors like the motorist's age, gender, vehicle type, and how many miles driven per year are not significant.

Table 3 shows average marginal effects and standard errors calculated using the delta method for selected coefficients for Models 1.1 to 2.4. In all eight models, the variable for the relative fuel prices (either the E85 premium or the log of the E85 ratio) is negative and statistically significant. The marginal effect estimates of the E85 premium in Models 1.1 to 1.4 are -0.282, -0.117, -0.246, and -0.255 respectively. When our estimation sample includes motorists who incur some cost to access E85, the magnitude of the E85 premium coefficient decreases. The interpretation of the coefficient for the E85 premium variable in Model 1.3 (for example) is that increasing the E85 premium by 10 cents decreases the probability that a motorist chooses E85 by 2.46 percent.

Figure 5 shows the average estimated probability that motorists choose E85 given the E85 premium, and Figure 6 shows the analogous probability estimates given the E85 ratio. Table 3 shows that the regional effects for California are considerable and that California motorists are about 50 percent more likely to choose E85 than motorists elsewhere. Thus the estimated demand curves are calculated and displayed for Retailer A and Retailer B separately. By contrast, the state effects for Retailer A's locations in Arkansas, Colorado, and Oklahoma are not significantly different from Iowa and suggest that the share of motorists who choose E85 given prices does not vary significantly across those regions. When we estimate the models with only the price and region variables in Models 3.1 to 4.4, the magnitude of the effect of the price variable increases compared to the models with all independent variables, and only the California effect is significant. Results of Models 3.1 to 4.4 are summarized in Table 4.

Table 5 shows results of Models 5.1 to 6.3, i.e., the models that incorporate the SP data. As anticipated, there is significant bias in the estimates from Models 5.1 and 6.1 that only use the SP data. Specifically, the sign of the price variable marginal effect becomes positive implying that motorists presented with high E85 prices are more likely to choose E85 than motorists presented with low E85 prices. This is because unfavorable hypothetical E85 prices were primarily offered to motorists with revealed preference for E10 and vice versa. Models 5.2 and 6.2 stack the SP and RP data together and yield similarly biased estimates, though not quite to the degree of the price variable effects reversing signs.

Models 5.3 and 6.3 are our preferred SP-off-RP models. The estimated marginal effects for most of the independent variables are virtually identical to the RP-only models of 1.3 and 2.3. The largest change is to the estimates of the price variable effects where the estimated standard errors are greatly reduced. For example, the results of Model 1.3 indicate that increasing the E85 premium by $0.10 lowers the share of motorists who choose E85 by 2.46 percent with a standard error of 0.84 percent, while the results of Model 5.3 indicate that a similar increase in the E85 premium lowers the share of motorists who choose E85 by 2.77 percent with a standard error of 0.24 percent.

Figure 7 shows the estimated share of motorists who choose E85 given the E85 premium from Models 1.3 and 5.3, and Figure 8 shows the share given the E85 ratio from Models 2.3 and 6.3. The figures compare the distributions of WTP for E85 from the RP-only models with the distributions from the SP-off-RP models and show that the resulting distributions are remarkably similar both in mean and variance. The most notable change between the RP-only models and the SP-off-RP models is that the estimate of the mean willingness to pay from among the motorists at Retailer B's stations shifts downward in the SP-off-RP model.

Tables 3, 4 and 5 also show parameters of the logistic distribution that summarize our findings regarding willingness to pay for E85 compared to E10. The logistic distribution has two parameters: its mean (μ) and its scale (s). By reporting values for these parameters that are implied from our estimates, we can summarize the whole distribution of preferences. For the E85 ratio models, we report values for coefficients both in logs and in levels.

The mean willingness to pay is especially informative of preferences. In Model 1.2, we find a mean willingness to pay of $-$\$1.88/gal by motorists at Retailer A's fuel stations and mean willingness to pay of $-$\$1.07/gal by motorists at Retailer B's fuel stations. For the corresponding price ratio model, we find mean values of 0.51 and 0.69. As expected, we find that the more E85 motorists we include in our estimation sample, the greater the estimate of mean WTP for E85.

Our results indicate that, on average, motorists discount E85 by more than the discount necessary for E10 and E85 to be equal on a cost-per-mile basis. Based on the sample selection criterion separating Models 2.2 and 2.3, we find a mean WTP for E85 in Retailer A's fuel stations between 51 and 63 percent of the price of E10. These are discounts in excess of the cost-per-mile parity discount of between 14 and 27 percent. In California, from Models 2.2 and 2.3, we estimate a mean willingness to pay between 69 and 116 percent in terms of the E85 ratio, suggesting that a large share of the general population of flex motorists in California are willing to pay a premium for E85 in excess of the cost-per-mile parity price.

Goodness of Model Fit

We use McFadden's pseudo R-squared to measure goodness of fit and compare how well our models fit the data. McFadden's pseudo R-squared is a transformation of the log-likelihood value and a measure of how much of the observed variation in fuel choices is explained by the model. The pseudo R-squared values tend to be about half of traditional R-squared values from OLS estimation, and values of 0.2 to 0.4 represent excellent fit (Domencich and McFadden 1975).

The E85 premium models and the E85 ratio models fit the data similarly well. The log-likelihood values and pseudo R-squared values are almost identical, and neither model has consistently higher values. Therefore on the basis of pseudo R-squared, we cannot conclude in favor of the E85 premium models or the E85 ratio models.

7 Conclusion

In this study, we estimate the distribution of preferences for E85 relative to E10 among flex motorists. With the collaboration of two E85 retailers, we conducted an intercept survey at E85 stations to collect both revealed fuel preferences and stated fuel opinions from motorists with FFVs. We visited E85 stations in the urban areas of Colorado Springs, Des Moines, Little Rock, Tulsa, Los Angeles and Sacramento. We find much stronger preferences for E85 in California than in the other regions covered by our survey. When the nominal E85 price per gallon was about 80 percent of the nominal E10 price per gallon, less than half flex motorists outside of California chose E85, whereas nearly 90 percent of flex motorists in California chose E85. The difference could be because motorists around Los Angeles and Sacramento are genuinely willing to pay more to refuel with E85 as a substitute for E10, or it could be because of the California retailer's marketing techniques to promote biofuels to local flex motorists. Another confounding factor is that E85 stations are less common in California than in the other regions covered by our survey.

We find that the significant factors affecting the probability flex motorists choose E85 are the fuel prices, vehicle ownership (personal, government, company, other), motorist opinions about which fuel is better for the environment, the engine, the economy, miles per gallon, and whether the observation was collected in the state of California. We did not find differences in willingness to pay in between states other than California. This is a key result and it means that all else equal, the probability that a motorist chooses E85 is not significantly different in Des Moines than it is in Colorado Springs, Little Rock, or Tulsa, despite the fact that the general opinion of ethanol among flex motorists in our sample is much higher in Des Moines than the other regions. Extrapolating to other regions of the United States, this result indicates that we may be able to apply estimation results from one state to project national demand, though we would need to make adjustments for California and other states with highly significant confounders.

In the four states excluding California, we find a mean willingness to pay for E85 between 51 and 63 percent of the price of E10. In California, we find a mean willingness to pay for E85 between 69 and 116 percent of the price of E10. Estimates from the SP-off-RP models that incorporate the stated preference data are similar to the estimates from the RP-only models. The SP data feature greater variation in 'observed' fuel prices and allow for more precise estimation of the distribution of willingness to pay. In particular, the estimated standard errors of the price variable coefficients are about 70 percent smaller in the SP-off-RP models than they are in the RP-only models.

Our models are relatively successful in fitting the observed survey data as measured by McFadden's pseudo R-squared values. We estimate models where the motorists respond to the absolute difference in fuel prices (the E85 premium) as well as models where the motorists respond to the relative difference in fuel prices (the E85 ratio). We find practically no difference in how well these two models fit the data. Thus we cannot say whether motorists are responding to the E85 ratio or the E85 premium when they make fuel choices.

References

Adamowicz, W, J. Louviere, and M. Williams. 1994. Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management, 26*(3), 271-292.

Alternative Fuels Data Center. 2015. *Alternative fueling station locator.* Retrieved from http://www.afdc.energy.gov/fuels/ethanol_locations.html.

Aguilar, F., Z. Cai, P. Mohebalian, and W. Thompson. 2015. Exploring the drivers' side of the "blend wall". US consumer preferences for ethanol blend fuels. *Energy Economics, 49*, 217-226.

Anderson, S. 2012. The demand for ethanol as a gasoline substitute. *Journal of Environmental Economics and Management, 63*(2), 151-168.

Ben-Akiva, M. and T. Morikawa. 1990. Estimation of switching models from revealed preferences and stated intentions. *Transportation Research Part A: General, 24*(6), 485-495.

Cameron, A. and P.Trivedi. 2005. Microeconometrics: methods and applications. Cambridge University Press.

Corts, K. 2010. Building out alternative fuel retail infrastructure: Government fleet spillovers in E85. *Journal of Environmental Economics and Management, 59*(3), 219-234.

Domencich, T. and D. McFadden. 1975. Urban Travel Demand-A Behavioral Analysis (No. Monograph).

Hensher, D. and M. Bradley. 1993. Using stated response choice data to enrich revealed preference discrete choice models. *Marketing Letters, 4*(2), 139-151.

Hensher, D., J. Louviere, and J. Swait. 1999. Combining sources of preference data. *Journal of Econometrics, 89*(1), 197-221.

Jensen, K., C. Clark, B. English, R. Menard, D. Skahan, and A. Marra. 2010. Willingness to pay for E85 from corn, switchgrass, and wood residues. *Energy Economics, 32*(6), 1253-1262.

Petrolia, D., S. Bhattacharjee, D. Hudson, and C. Herndon. 2010. Do Americans want ethanol? A comparative contingent-valuation study of willingness to pay for E-10 and E-85. *Energy Economics, 32*(1), 121-128.

Pouliot, S. and B. Babcock. 2014. The demand for E85: Geographical location and retail capacity constraints. *Energy Economics, 45*, 134-143.

Pouliot, S. and B. Babcock. 2016. Compliance path and impact of ethanol mandates on retail fuel market in the short run. *American Journal of Agricultural Economics*, Forthcoming.

Salvo, A. and C. Huse. 2013. Build it, but will they come? Evidence from consumer choice between gasoline and sugarcane ethanol. *Journal of Environmental Economics and Management, 66*(2), 251-179.

Train, K. and W. Wilson. 2008. Estimation on stated-preference experiments constructed from revealed-preference choices. *Transportation Research Part B: Methodological, 42*(3), 191-203.

Train, K. and W. Wilson. 2009. Monte Carlo analysis of SP-off-RP data. *Journal of Choice Modelling, 2*(1), 101-117.

Tables

Table 1 Observed E85 and E10 prices and shares of motorists who choose E85 by station, region, and retailer

Urban area and station	Number of observations	Avg. E85 price ($/gal)	Avg. E10 price ($/gal)	Avg. E85 premium (E85 - E10)	Avg. E85 ratio (E85/E10)	Share of motorists using E85
Co. Springs 1	11	2.00	2.00	0.00	1.00	9.1%
Co. Springs 2	33	2.00	2.02	-0.02	0.99	30.3%
Co. Springs 3	54	2.00	2.06	-0.06	0.97	13.0%
CS total	**98**	**2.00**	**2.04**	**-0.04**	**0.98**	**18.4%**
Des Moines 1	117	2.16	2.72	-0.56	0.79	46.2%
Des Moines 2	50	2.30	2.64	-0.35	0.87	28.0%
Des Moines 3	27	2.32	2.81	-0.49	0.82	51.9%
Des Moines 4	114	2.29	2.69	-0.39	0.85	40.4%
DM total	**308**	**2.25**	**2.70**	**-0.46**	**0.83**	**41.6%**
Little Rock 1	26	1.84	2.18	-0.34	0.84	34.6%
Little Rock 2	23	1.83	2.13	-0.30	0.86	34.8%
Little Rock 3	60	1.83	2.18	-0.35	0.84	31.7%
LR total	**109**	**1.83**	**2.17**	**-0.34**	**0.84**	**33.0%**
Tulsa 1	58	1.80	2.09	-0.29	0.86	41.4%
Tulsa 2	12	1.80	2.10	-0.30	0.86	66.7%
Tulsa 3	65	1.80	2.04	-0.24	0.88	18.5%
TS Total	**135**	**1.80**	**2.07**	**-0.27**	**0.87**	**32.6%**
Retailer A total	**650**	**2.05**	**2.38**	**-0.34**	**0.86**	**34.8%**
Los Angeles 1	85	2.61	3.20	-0.59	0.82	95.3%
Los Angeles 2	52	2.63	3.10	-0.47	0.85	84.6%
LA total	**137**	**2.62**	**3.16**	**-0.54**	**0.83**	**91.2%**
Sacramento 1	43	2.57	3.23	-0.66	0.79	81.4%
Sacramento 2	51	2.48	2.92	-0.44	0.85	88.2%
SAC total	**94**	**2.52**	**3.06**	**-0.54**	**0.82**	**85.1%**
Retailer B total	**231**	**2.58**	**3.12**	**-0.54**	**0.83**	**88.7%**
Sample total	**881**	**2.19**	**2.58**	**-0.39**	**0.85**	**48.9%**

Data are from 17 stations in six urban areas: Colorado Springs (CS), Des Moines (DM), Los Angeles (LA), Little Rock (LR), Sacramento (SAC), and Tulsa (TS). We cooperated with Retailer A in CS, DM, LR, and TS, and with Retailer B in LA and SAC. We conducted surveys around DM over the course of two months before spending one week at each other area. Prices are in nominal, non-energy-adjusted terms and are averaged over the observations in the sample for each station/region/retailer. The E85 premium is the E85 price minus the E10 price. The E85 ratio is the E85 price divided by the E10 price.

Table 2 Summary of SP choices and fuel switching

		SP choice		
		E10	E85	Total
RP choice	E10	363	87	450
	E85	201	230	431
	Total	564	317	881
Proportion of motorists who switched			0.327	
Proportion who switched from E85 to E10			0.466	
Proportion who switched from E10 to E85			0.193	

After observing motorists' RP choices, motorists were asked if they would still have made the same choice if the fuel they chose had been more expensive by an amount or they were asked if they would still have made the same choice if the fuel they did not choose had been less expensive by an amount. The amount that the hypothetical fuel prices varied from the actual fuel prices was $0.25, $0.50, or $0.75.

Table 3 Results of models with only RP data and all variables

	Model 1.1		Model 1.2		Model 1.3		Model 1.4	
	Marginal FX (SE)		Marginal FX (SE)		Marginal FX (SE)		Marginal FX (SE)	
E85 Premium	-0.282	(0.080)	-0.117	(0.066)	-0.246	(0.090)	-0.255	(0.087)
Arkansas (AR)	0.027	(0.048)	0.036	(0.034)	0.029	(0.052)	0.017	(0.051)
Colorado (CO)	-0.017	(0.063)	0.060	(0.048)	0.019	(0.067)	0.003	(0.066)
Oklahoma (OK)	0.092	(0.044)	0.088	(0.035)	0.110	(0.051)	0.097	(0.051)
California (CA)	0.447	(0.039)	0.243	(0.025)	0.522	(0.047)	0.504	(0.045)
Observations	881		479		670		714	
McFadden's R^2	0.319		0.312		0.323		0.327	
Ret. A μ (prem.)	-0.78		-1.82		-1.22		-1.12	
Ret. A s (prem.)	0.66		0.49		0.68		0.68	
Ret. B μ (prem.)	0.77		-1.07		0.40		0.55	
Ret. B s (prem.)	0.63		0.54		0.67		0.68	

	Model 2.1		Model 2.2		Model 2.3		Model 2.4	
	Marginal FX (SE)		Marginal FX (SE)		Marginal FX (SE)		Marginal FX (SE)	
Ln E85 Ratio	-0.777	(0.215)	-0.342	(0.186)	-0.681	(0.241)	-0.699	(0.233)
Arkansas (AR)	0.004	(0.046)	0.028	(0.041)	0.010	(0.050)	-0.003	(0.049)
Colorado (CO)	-0.009	(0.063)	0.073	(0.084)	0.026	(0.069)	0.009	(0.067)
Oklahoma (OK)	0.072	(0.043)	0.082	(0.053)	0.093	(0.049)	0.080	(0.049)
California (CA)	0.462	(0.038)	0.273	(0.075)	0.538	(0.046)	0.520	(0.044)
Observations	881		479		670		714	
McFadden's R^2	0.320		0.307		0.323		0.328	
Ret. A μ (ln ratio)	-0.30		-0.68		-0.46		-0.43	
Ret. A μ (ratio)	0.74		0.51		0.63		0.65	
Ret. A s (ln ratio)	0.23		0.18		0.24		0.24	
Ret. B μ (ln ratio)	0.29		-0.37		0.15		0.21	
Ret. B μ (ratio)	1.34		0.69		1.16		1.23	
Ret. B s (ln ratio)	0.23		0.18		0.24		0.25	

The E85 premium is the E85 price minus the E10 price, and Ln E85 ratio is the log of the E85 price divided by the E10 price. State effects are compared to Iowa. Not all variables are shown. The model samples are: 1)All observations, 2) only motorists who did not come for E85, 3) including motorists who came for E85 but did not drive out of their way, and 4) including motorists who came for E85 but drove no more than 1 mile out of their way. Parameters of the logistic WTP distribution (μ and s) are calculated for each model and each retailer.

Table 4 Results of models with only RP data and state variables

	Model 3.1		Model 3.2		Model 3.3		Model 3.4	
	Marginal FX (SE)		Marginal FX (SE)		Marginal FX (SE)		Marginal FX (SE)	
E85 Premium	-0.365	(0.089)	-0.137	(0.070)	-0.305	(0.097)	-0.320	(0.094)
Arkansas (AR)	-0.020	(0.045)	0.011	(0.042)	-0.001	(0.050)	-0.025	(0.048)
Colorado (CO)	-0.055	(0.063)	0.060	(0.080)	-0.009	(0.069)	-0.035	(0.066)
Oklahoma (OK)	0.002	(0.044)	0.027	(0.047)	0.008	(0.050)	-0.018	(0.049)
California (CA)	0.463	(0.038)	0.187	(0.080)	0.502	(0.052)	0.489	(0.049)
Observations	881		479		670		714	
McFadden's R^2	0.211		0.109		0.199		0.208	
Ret. A μ (prem.)	-0.67		-1.53		-1.03		-0.94	
Ret. A s (prem.)	0.50		0.39		0.54		0.53	
Ret. B μ (prem.)	0.50		-0.90		0.21		0.30	
Ret. B s (prem.)	0.94		0.57		0.99		0.99	

	Model 4.1		Model 4.2		Model 4.3		Model 4.4	
	Marginal FX (SE)		Marginal FX (SE)		Marginal FX (SE)		Marginal FX (SE)	
Ln E85 Ratio	-0.996	(0.238)	-0.395	(0.199)	-0.842	(0.259)	-0.877	(0.250)
Arkansas (AR)	-0.049	(0.043)	0.001	(0.037)	-0.025	(0.047)	-0.050	(0.046)
Colorado (CO)	-0.045	(0.064)	0.075	(0.093)	0.001	(0.071)	-0.026	(0.068)
Oklahoma (OK)	-0.023	(0.042)	0.019	(0.043)	-0.012	(0.047)	-0.039	(0.046)
California (CA)	0.484	(0.036)	0.217	(0.086)	0.524	(0.050)	0.511	(0.047)
Observations	881		479		670		714	
McFadden's R^2	0.212		0.112		0.201		0.209	
Ret. A μ (ln ratio)	-0.27		-0.56		-0.40		-0.37	
Ret. A μ (ratio)	0.76		0.57		0.67		0.69	
Ret. A s (ln ratio)	0.19		0.14		0.20		0.20	
Ret. B μ (ln ratio)	0.19		-0.32		0.08		0.12	
Ret. B μ (ratio)	1.21		0.73		1.08		1.12	
Ret. B s (ln ratio)	0.18		0.13		0.19		0.20	

The E85 premium is the E85 price minus the E10 price, and Ln E85 ratio is the log of the E85 price divided by the E10 price. State effects are compared to Iowa. The model samples are: 1) All observations, 2) only motorists who did not come for E85, 3) including motorists who came for E85 but did not drive out of their way, and 4) including motorists who came for E85 but drove no more than 1 mile out of their way. Parameters of the logistic WTP distribution (μ and s) are calculated for each model and each retailer.

Table 5 Results of models with SP and RP data

	Model 5.1		Model 5.2		Model 5.3	
	Marginal FX	(SE)	Marginal FX	(SE)	Marginal FX	(SE)
E85 Premium	0.024	(0.037)	-0.085	(0.032)	-0.277	(0.024)
Arkansas (AR)	-0.047	(0.055)	-0.014	(0.040)	0.028	(0.046)
Colorado (CO)	-0.085	(0.052)	-0.054	(0.040)	0.027	(0.049)
Oklahoma (OK)	-0.000	(0.056)	0.046	(0.039)	0.084	(0.040)
California (CA)	0.100	(0.055)	0.343	(0.037)	0.319	(0.029)
Observations	670		670		670	
McFadden's R^2	0.103		0.169		0.199	
Ret. A μ (prem.)					-1.19	
Ret. A s (prem.)					0.59	
Ret. B μ (prem.)					-0.04	
Ret. B s (prem.)					0.57	

	Model 6.1		Model 6.2		Model 6.3	
	Marginal FX	(SE)	Marginal FX	(SE)	Marginal FX	(SE)
Ln E85 Ratio	0.025	(0.084)	-0.185	(0.072)	-0.599	(0.058)
Arkansas (AR)	-0.042	(0.056)	-0.024	(0.040)	-0.014	(0.048)
Colorado (CO)	-0.078	(0.052)	-0.060	(0.039)	-0.023	(0.051)
Oklahoma (OK)	0.008	(0.056)	0.035	(0.038)	0.047	(0.041)
California (CA)	0.119	(0.056)	0.351	(0.038)	0.325	(0.029)
Observations	670		670		670	
McFadden's R^2	0.101		0.167		0.191	
Ret. A μ (ln ratio)					-0.54	
Ret. A μ (ratio)					0.58	
Ret. A s (ln ratio)					0.28	
Ret. B μ (ln ratio)					0.04	
Ret. B μ (ratio)					1.04	
Ret. B s (ln ratio)					0.27	

The E85 premium is the E85 price minus the E10 price, and Ln E85 ratio is the log of the E85 price divided by the E10 price. State effects are compared to Iowa. Not all variables are shown. The models are: 1) Only SP observations, 2) SP and RP observations pooled together, and 3) SP-off-RP model correcting for endogenous hypothetical prices. The sample includes motorists who came for E85 but did not drive out of their way. Parameters of the logistic WTP distribution (μ and s) are calculated for each model and each retailer.

Figures

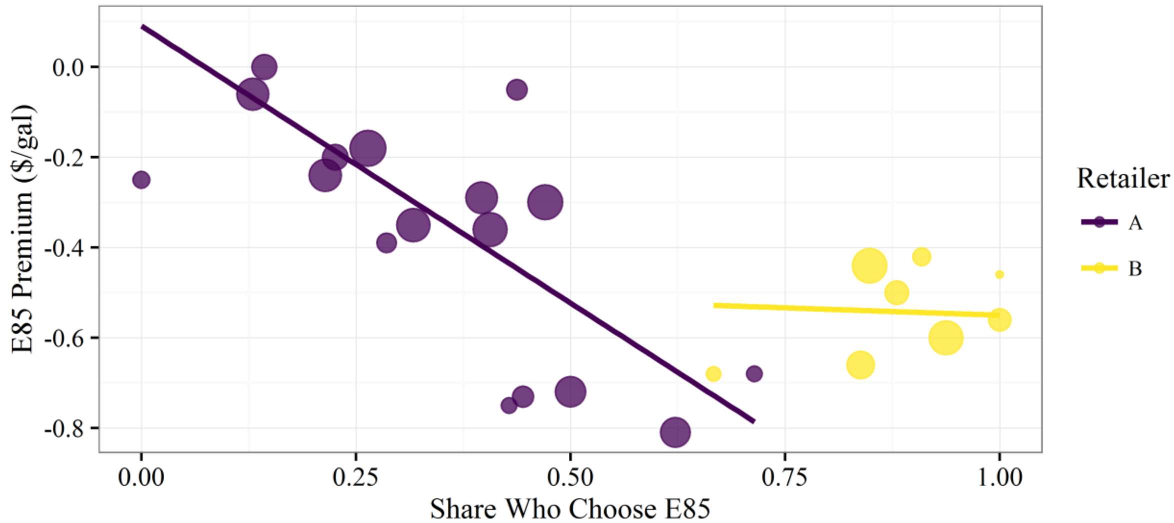

Figure 1 E85 premium and share who choose E85 in RP setting by retailer

Data are observations of 881 flex motorists refueling at E85 stations across the United States. Retailer A is in the Midwest and Retailer B is in California. The E85 premium is the nominal E85 price minus the E10 price. The size of the bubbles represents the number of observations at the given E85 premium. The lines are weighted OLS regressions.

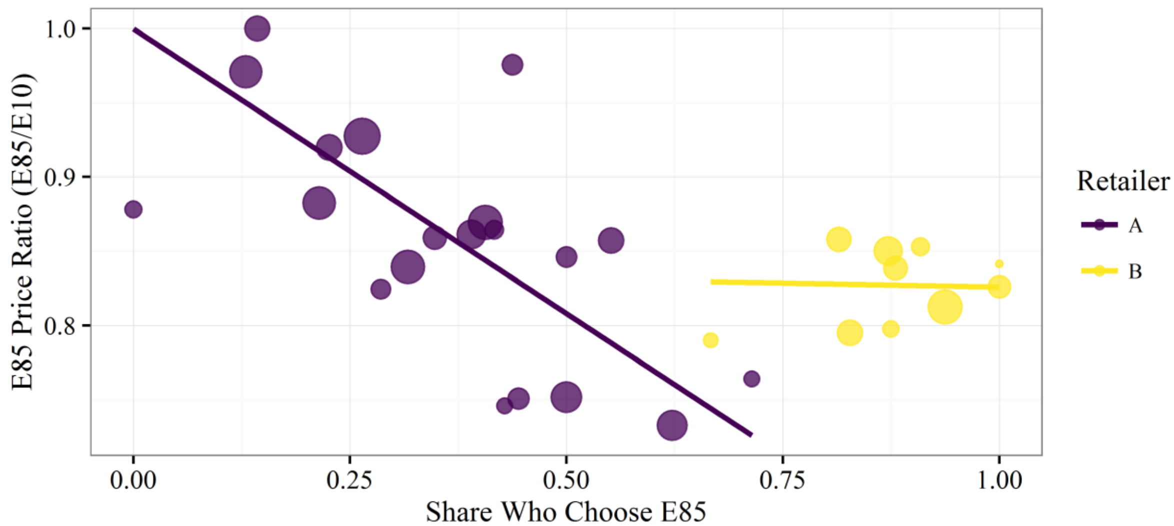

Figure 2 E85 ratio and share who choose E85 in RP setting by retailer

Data are observations of 881 flex motorists refueling at E85 stations across the United States. Retailer A is in the Midwest and Retailer B is in California. The E85 ratio is the nominal E85 price divided by the E10 price. The size of the bubbles represents the number of observations at the given E85 ratio. The lines are weighted OLS regressions.

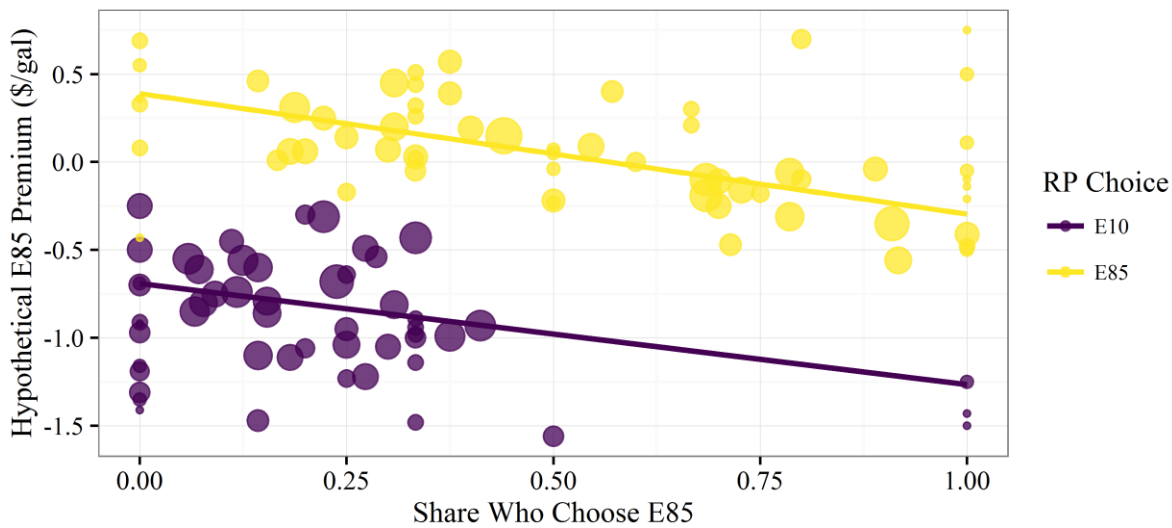

Figure 3 E85 premium and share who hypothetically choose E85 in SP setting

Data are responses from 881 flex motorists to a hypothetical price scenario. Motorists who chose E10 in the RP setting were offered favorable E85 prices and vice versa. Retailer A is in the Midwest and Retailer B is in California. The E85 premium is the nominal E85 price minus the E10 price. The size of the bubbles represents the number of observations at the given hypothetical E85 premium. The lines are weighted OLS regressions.

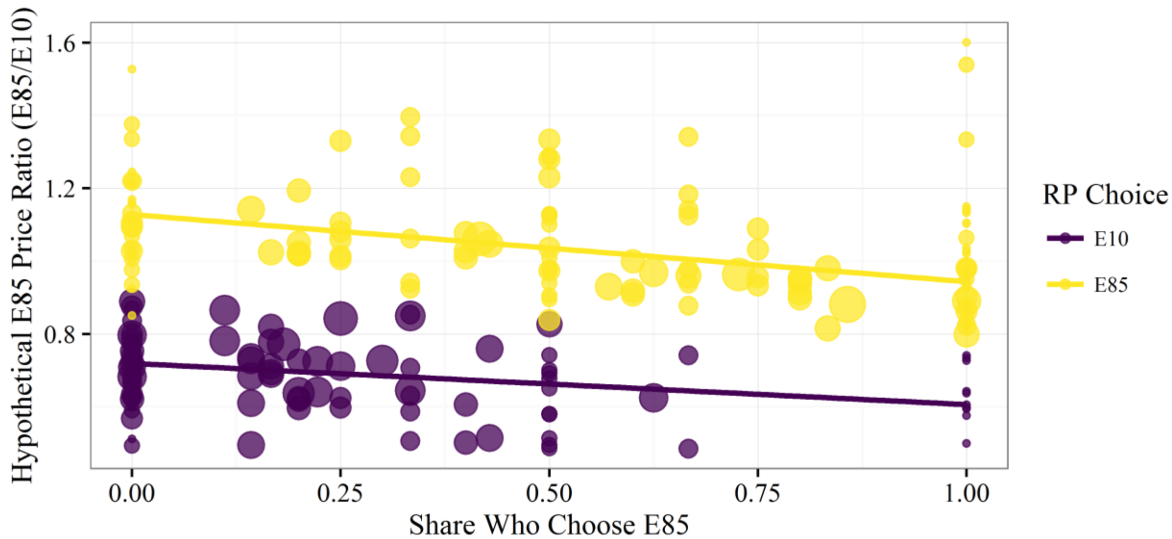

Figure 4 E85 ratio and share who hypothetically choose E85 in SP setting

Data are responses from 881 flex motorists to a hypothetical price scenario. Motorists who chose E10 in the RP setting were offered favorable E85 prices and vice versa. Retailer A is in the Midwest and Retailer B is in California. The E85 ratio is the nominal E85 price divided by the E10 price. The size of the bubbles represents the number of observations at the given hypothetical E85 ratio. The lines are weighted OLS regressions.

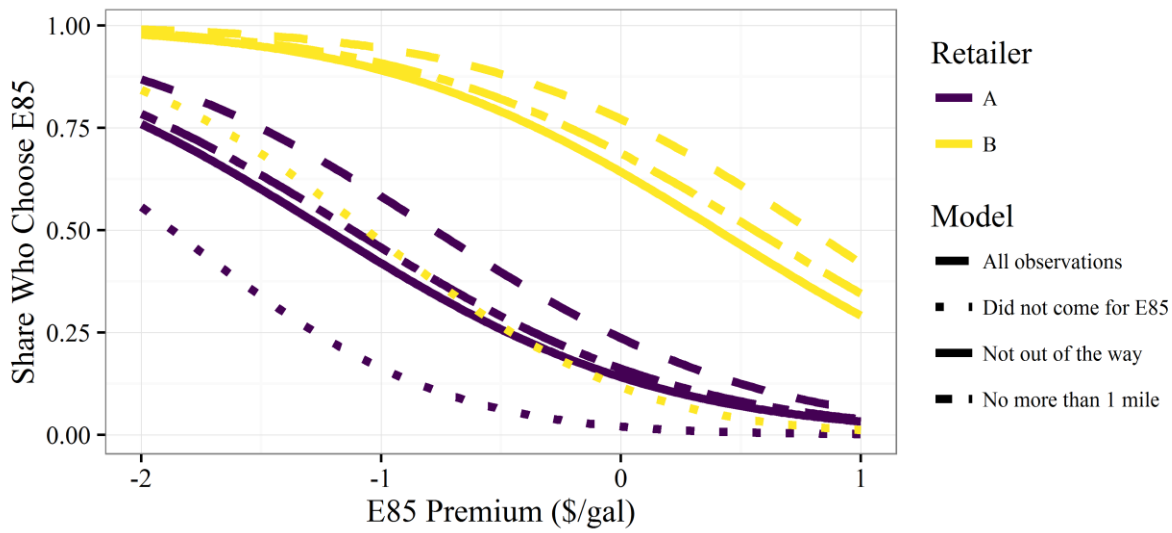

Figure 5 Estimated E85 demand from RP models with E85 premium

Estimated shares are logit probabilities calculated at each E85 premium for each motorist in the sample and averaged for each retailer. The model samples are: All observations (no correction), only motorists who did not come for E85, including motorists who came for E85 but did not drive out of their way, and including motorists who came for E85 but drove no more than 1 mile out of their way.

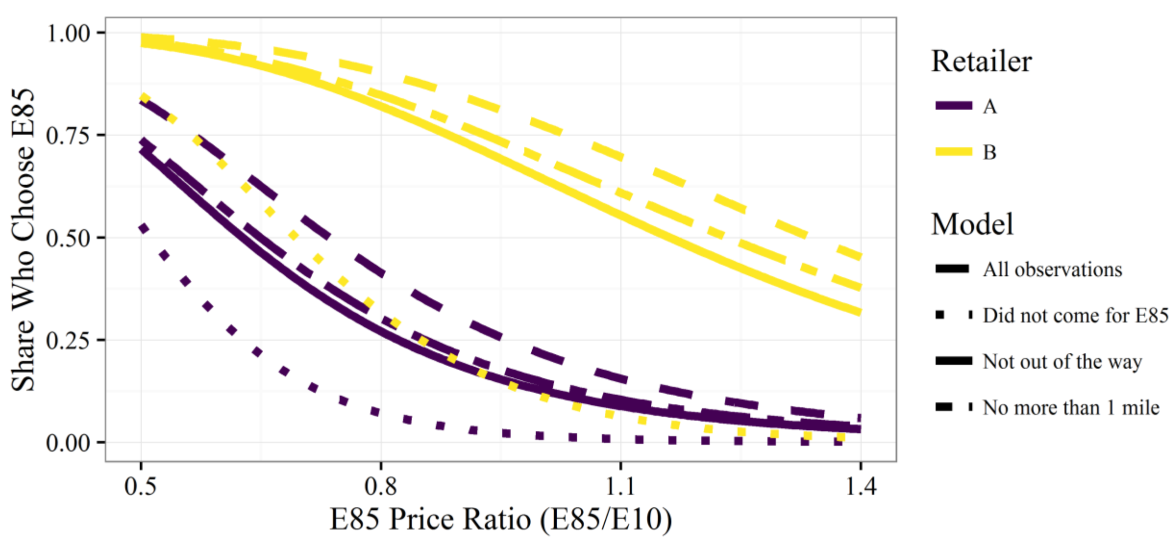

Figure 6 Estimated E85 demand from RP models with E85 ratio

Estimated shares are logit probabilities calculated at each E85 ratio for each motorist in the sample and averaged for each retailer. The model samples are: All observations (no correction), only motorists who did not come for E85, including motorists who came for E85 but did not drive out of the way, and including motorists who came for E85 but drove no more than 1 mile out of their way.

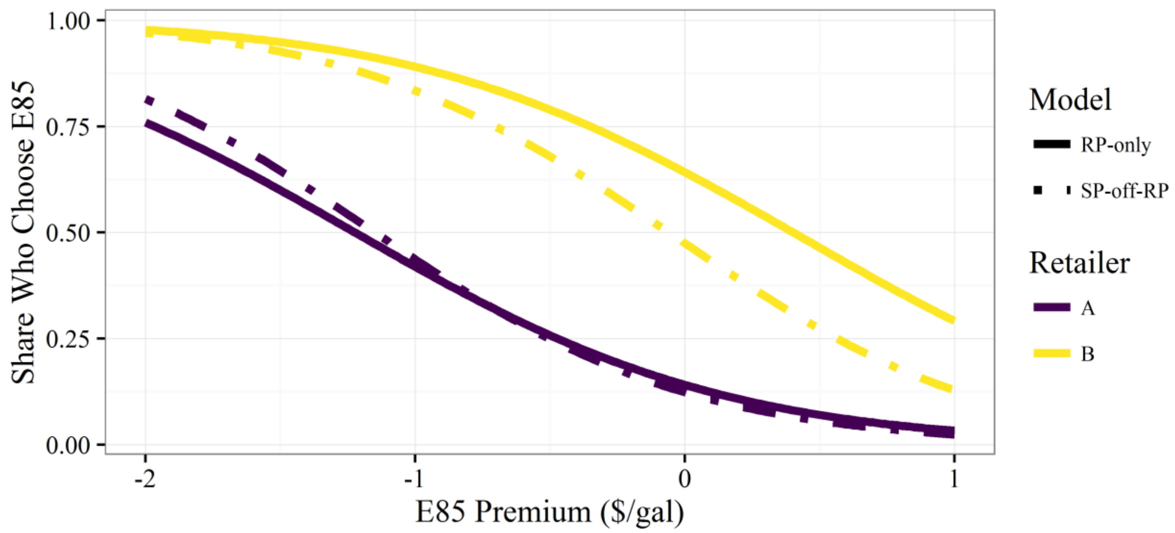

Figure 7 Estimated E85 demand from SP and RP models with E85 premium

Estimated shares are logit probabilities calculated at each E85 premium for each motorist in the sample and averaged for each retailer. The model sample includes motorists who came for E85 but did not drive out of their way. The RP-only model uses only RP data and the SP-off-RP model is uses both SP and RP data to obtain more precise estimates.

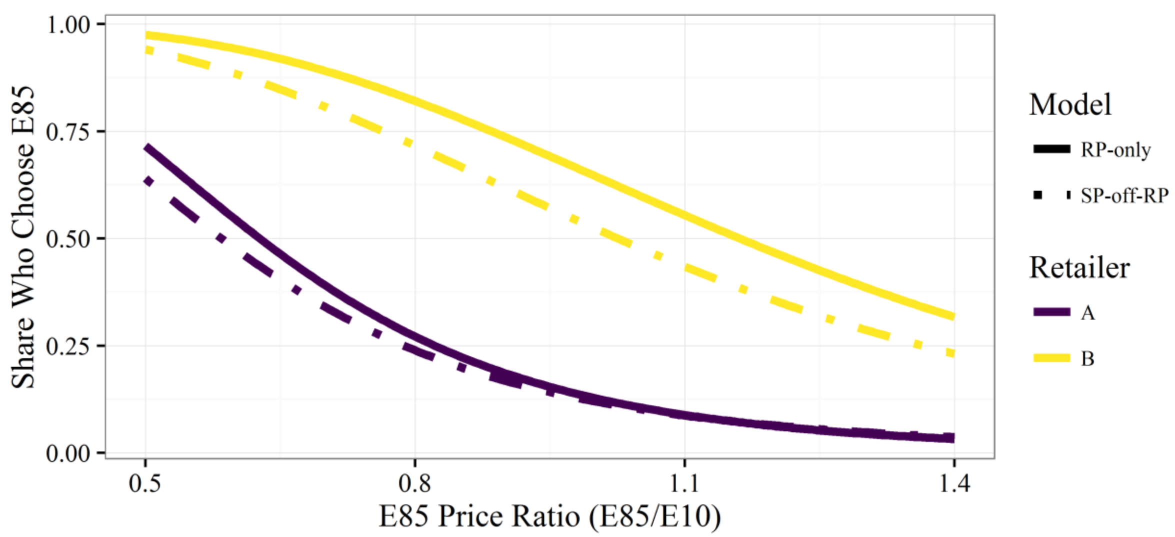

Figure 8 Estimated E85 demand from SP and RP models with E85 ratio

Estimated shares are logit probabilities calculated at each E85 ratio for each motorist in the sample and averaged for each retailer. The model sample includes motorists who came for E85 but did not drive out of their way. The RP-only model uses only RP data and the SP-off-RP model is uses both SP and RP data to obtain more precise estimates.