2016

Essays on the demand for ethanol in the United States: willingness to pay for E85

Kenneth Liao
Iowa State University

Essays on the demand for ethanol in the United States

Willingness to pay for E85

by

Kenneth Liao

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Economics

Program of Study Committee:
Sébastien Pouliot, Major Professor
Georgeanne Artz
Bruce Babcock
Otávio Bartalotti
Chad Hart

Iowa State University

Ames, Iowa

2016

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I thank my major professor, Sébastien Pouliot, for countless hours of thoughtful guidance and for making my time at Iowa State University a wonderful experience. I also thank my committee members, Georgeanne Artz, Otávio Bartalotti, Chad Hart, and especially Bruce Babcock and the Biobased Industry Center for facilitating funding of this research.

I thank Soren Anderson for his advice and encouragement, and I thank the Minnesota Department of Commerce and two E85 retailers who wish to remain anonymous for their cooperation providing data. I thank the team of undergraduate research assistants who helped conduct surveys, and I offer my appreciation to the motorists who participated in them.

Lastly I thank my family, friends, colleagues, and the department faculty and staff for their continued encouragement and support.

ABSTRACT

This dissertation contains three studies that estimate the distribution of willingness to pay (WTP) for E85 as a substitute for E10 among flex motorists in the United States. The results are vital for estimating the demand for ethanol beyond the blend wall and for analysis of the Renewable Fuel Standard. The first study attempts to estimate the distribution of preference for E85 from data generated by a survey of E85 stations in Minnesota. The study uses an extensive sample of recent observations, but estimates of the WTP distribution vary substantially depending on model specification. The conclusion is that the data are not suitable to estimate the distribution of WTP for E85.

The second and third studies collect primary data from E85 stations in different regions of the United States to more accurately estimate preferences for E85 and investigate locational differences. The studies obtain revealed-preference (RP) data from flex motorists refueling at E85 stations and stated-preference (SP) data from surveying the flex motorists and presenting hypothetical scenarios. The second study uses the RP data to estimate relative preferences for E85, and the third study incorporates the SP data to better capture the wide range of fuel-switching behavior.

The estimation sample consists of about nine hundred flex motorists in six urban areas in the Midwest and California. The sample of flex motorists who refuel at E85 stations is endogenously stratified; the probability of a flex motorist appearing in the sample is correlated to the motorist's WTP for E85. The models apply corrective probability weights so estimates reflect the population and not the sample.

The results show that a $0.10 increase in the E85-E10 price difference decreases the probability of motorists choosing E85 by about 2.5 percent, on average, and preferences are spread over a broad range of fuel prices. In general, motorists are willing to pay more for E85 in California than in the Midwest, and when E85 and E10 are priced equally on a cost-per-mile basis, about 25 percent of flex motorists choose E85 in the Midwest compared to 75 percent in California.

CHAPTER 1.

INTRODUCTION

1.1 Motivation and Overview

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[1] in RFS2 was originally scheduled to be 14.4 billion gallons in 2014, and 15 billion gallons in 2015 and 2016. 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 14.50 billion gallons in 2016. The final rule came after EPA received numerous comments from supporters of renewable fuels and supporters of conventional fuels.

[1] 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.

To enforce the mandates, RFS2 provides credits called Renewable Identification Numbers (RINs) that create taxes on conventional fuels and subsidies for biofuels. Taxes and subsidies endogenously adjust to the cost of production, the strength of the demand for biofuels, and the mandated volumes. Therefore, analysis of US biofuel policy requires a description of the demand curve for biofuels and in particular the demand for ethanol beyond the blend wall.

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 no more than 83 and no less than 51 percent ethanol. On average, a gallon of E85 contains about 74 percent ethanol so each gallon of E85 consumed as a substitute for E10 increases aggregate ethanol consumption by about 0.64 gallons (EIA 2015). Thus, ethanol consumption could exceed the blend wall if even a small fraction of motorists refuel with 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 dissertation contains three studies that estimate the relative preferences of motorists for E10 and E85 to better understand the aggregate demand for ethanol in the United States.

These studies provide 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), to predict RIN prices, and to 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). We refer to these motorists as ‘flex motorists’.

Previous studies have estimated relative preferences for ethanol and gasoline for flex motorists in Brazil (e.g., Salvo and Huse 2011, Pouliot 2013) and Minnesota (e.g., Anderson 2012; Corts 2010; Liu and Greene 2013). Other studies have estimated preferences for E85 in the United States using stated-preference (SP) data collected with nationwide mail and online surveys (e.g., Aguilar et. al. 2015; Jensen et. al. 2010; Petrolia et. al. 2010).

The next chapter of this dissertation is a study that expands on the work of Anderson (2012) and uses recent data from a monthly survey of E85 stations in Minnesota. These data are the most comprehensive available data on E85 consumption by US motorists. Using the station-level data, we estimate a model of demand to recover the distribution of motorist preferences. The empirical estimates are highly sensitive to model specification, and we cannot identify a distribution of willingness to pay for E85 from the monthly station data. Furthermore it is unclear whether flex motorists in Minnesota are representative of flex motorists nationwide.

To more accurately estimate WTP for E85 and investigate spatial differences in preferences, we collect primary data from E85 fuel stations in different regions across the United States by performing an intercept survey similar to Salvo and Huse (2013). We obtain revealed-preference (RP) data by observing actual fuel purchases by flex motorists, and we obtain additional SP data by asking motorists a series of short questions while they refuel their FFVs. We find that the share of flex motorists who choose E85 when its price is equal to E10 in energy-equivalent terms is about 25 percent in the Midwest and 75 percent in California, and we find that preferences are spread over a broad range of relative fuel prices.

In Chapter 3 we estimate the distribution of E85 preferences using only the RP data, and in Chapter 4 we incorporate the SP data. Chapter 5 provides a summary and offers conclusions.

1.2 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. Until recently, automobile manufacturers have had incentives from the US Corporate Average Fuel Economy (CAFE) standards to produce FFVs. 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 from American automobile companies.[2]

For motorists, 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

[2] The most common makes for FFVs in the United States are GM (Buick, Cadillac, Chevrolet, GMC, etc.), Ford (Lincoln), and Chrysler (Dodge, Jeep). Toyota and Nissan only manufacture flex versions of their largest pickup trucks (Tundra and Titan) and largest SUVs (Sequoia and Armada). Honda (and Acura), Hyundai (and Kia), Mitsubishi, Subaru, and other major auto brands do not manufacture any FFVs for sale in the United States.

states and seasons. Some motorists choose E85 when its price (in energy-equivalent terms) is at a premium relative to E10 while some motorists choose not to refuel with E85 even when its energy-adjusted price is at a discount relative to E10. In many cases, consumers are not able to acquire a certain vehicle make and model in anything but the 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. The cost to install new fueling infrastructure can be significant for retailers, and they are understandably hesitant to make such an investment without knowing what E85 demand will be. 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.

Pouliot (2013) finds that on average, flex motorists in Brazil slightly discount ethanol relative to gasoline. Motorists treat the two fuels as near-substitutes, and most motorists switch between fuels when their energy-adjusted prices are near parity. About 20 percent of motorists choose ethanol when its price is 10 percent above the energy-equivalent gasoline price, and about 20 percent of motorists choose gasoline when ethanol is discounted by 15 percent.

Salvo and Huse (2013) collect fuel preference data using a consumer intercept survey of flex motorists in Brazil that inspired the survey in this dissertation. 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.

When it comes to US motorists, there is no comprehensive source of data on national E85 sales or prices. The best available data on E85 sales come from a monthly survey of E85 stations in Minnesota conducted by the Minnesota Department of Commerce discussed in Section 2.4.

Anderson (2012) estimates the distribution of preferences for E85 using 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 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. Corts (2010) concludes that data from more recent years is required to estimate a credible model of retail E85 and FFV demand.

Liu and Greene (2013) estimate E85 demand using more recent data which allow a better estimate of non-fleet E85 demand than previous studies. The dependent variable is the share of energy services consumed by flex motorists in Minnesota that is attributable to ethanol, and Liu and Greene (2013) find a high price elasticity of demand for E85.

A limitation of these studies is that the energy-equivalent E85 price was almost always above the E10 price making it difficult to estimate a complete distribution of preferences. Furthermore, these studies raise the question of whether fuel preferences observed in Minnesota are representative of fuel preferences in the rest of the United States.

To estimate E85 demand from motorists outside of Minnesota, recent studies have used nationwide mail and online surveys to obtain stated-preference data on WTP for E85. Jensen et al. (2010) emphasize 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, but about 20 percent of motorists surveyed indicated strong unwillingness to buy fuel with any ethanol.

This dissertation contributes to the literature by estimating the distribution of preferences for E85 using RP and SP data collected from about one thousand flex motorists fueling their FFVs at retail E85 stations in different regions across the United States. We find that when E85 and E10 are priced equally on a cost-per-mile basis, on average, motorists in the Midwest prefer E10 while motorists in California prefer E85. The distribution of WTP is spread over a wide range of relative fuel prices; some motorists choose E85 when it is significantly more expensive than E10 while some motorists choose E10 when it is significantly more expensive than E85.

CHAPTER 2.

ESTIMATING WILLINGNESS TO PAY FOR E85 USING MINNESOTA STATION DATA

2.1 Introduction

This study expands on the work of Anderson (2012). In recent years in Minnesota, the number of fuel stations that offer E85 has increased, E85 prices have fallen relative to E10 prices so that E85 is sometimes offered at a discount, and the majority of E85 sales are now to private motorists rather than government fleet vehicles. Recent data on E85 sales and prices covering a wider range of E85-E10 price differences offer the opportunity to more completely and precisely estimate the distribution of willingness to pay for E85 as a substitute for E10 among flex motorists.

The data used for this study are monthly survey data from E85 fuel stations in Minnesota. E85 stations report their monthly sales of E85 and the volume-weighted average price they charge. We derive a choice model based on Anderson (2012) that provides a theoretical framework for estimating the demand for E85 based on station-level E85 sales.

From anecdotal evidence and findings from prior literature, we expect that Minnesotan flex motorists on average discount E85 relative to E10 on a cost-per-mile basis. This would reflect that, on average, motorists lack knowledge about ethanol or have a negative attitude toward ethanol perhaps because of the food versus fuel debate or because E85 requires refueling more often because of the lower energy content of ethanol. We are not able to verify these priors from the empirical estimates reported in this chapter. Estimates of mean willingness to pay are highly sensitive to model specification. Some specifications yield a positive mean WTP for E85 relative to E10 while others yield a large negative mean WTP. Later in this chapter, we elaborate on the reasons why we cannot identify a distribution of WTP using the survey data from Minnesota E85 stations.

The next section of this chapter provides the details of the theoretical model. In Section 2.3, we explain the empirical model. Section 2.4 describes the data. Section 2.5 contains estimation results, and Section 2.6 concludes.

2.2 Theoretical Model

We derive the demand for E85 based on a choice model described in Anderson (2012). The model is especially useful to formalize the connection between flex motorists' fuel preferences and aggregate market demand for E85. This section contains an overview of the model and the interested reader is referred to Anderson (2012) for additional details.

2.2.1 Motorist Behavior

Each motorist who owns an FFV maximizes the quasi-linear utility function

$$U = v\left(\left(q_e + q_g\right)m\right) + \theta_e q_e + \theta_g q_g + z, \tag{2.1}$$

where q_e is the quantity of E85 in gallons, q_g is the quantity of E10 in gallons, m is the fuel economy of the vehicle in miles per gallon, and z is a numeraire that captures the consumption of all other goods measured in dollars. The first term of the utility function represents the utility gained from driving M miles where $M \equiv \left(q_e + q_g\right)m$, and $v(M)$ is increasing and concave in miles driven. The quantity and the price of E85 are expressed in E10 energy-equivalent gallons. As such, ethanol and gasoline are perfect substitutes in producing miles. The parameters θ_e and θ_g measure the utility from consuming one gallon of E85 or E10 respectively for attributes of the fuel other than its main function to provide vehicle miles.

The parameters θ_e and θ_g are motorist specific and allow fuel choice to affect utility in a way that is unrelated to the cost per mile driven. This means that the motorists will not always choose the fuel with the lowest energy-adjusted price. Motorists receive some direct utility benefit or incur some direct utility cost from fuel consumption unrelated to the fuel's primary use providing energy for the vehicle. For example, some motorists may be willing to pay more for E85 because they value the environmental benefits of using renewable fuels while other motorists may be willing to pay more for E10 to avoid more frequent refueling.

Each motorist faces the budget constraint

$$p_e q_e + p_g q_g + z \leq y,$$

where p_g is the price of E10, p_e is the price of E85 (converted to E10 energy-equivalent dollars), y is the motorist's income, and the price of the composite good z is normalized to 1. By Walras' Law, the budget constraint holds with equality implying

$$z = y - p_e q_e - p_g q_g.$$

Substituting the value of z into equation (2.1), the unconstrained utility maximization problem for flex motorists is

$$\max_{q_e, q_g} U = v\Big(\big(q_e + q_g\big)m\Big) + \theta_e q_e + \theta_g q_g + y - p_e q_e - p_g q_g.$$

Because the two fuels are perfect substitutes, motorists choose either to fuel with E10 or E85, but not both. A motorist chooses to fuel with E85 if the net utility benefit per (energy-equivalent) gallon of E85 is greater than the net utility benefit per gallon of E10:

$$\theta_e - p_e \geq \theta_g - p_g.$$

Following the notation of Anderson (2012), we let $p \equiv p_e - p_g$ be the E85 price premium (or discount if negative) and $\theta \equiv \theta_e - \theta_g$ be the motorist's willingness to pay (or the amount to compensate the motorist if negative) to use E85 as a substitute for E10. Thus, we can restate the decision of motorists to choose E85 if their WTP to use E85 exceeds the price premium they face at the pump, i.e., $\theta \geq p$.

Even though a motorist makes her fuel choice based on the difference in prices and her own preference parameter θ, the quantity of fuel demanded and in turn the motorist's miles driven depend only on the price of the fuel chosen. The first order conditions of the utility maximization problem show that, conditional on the motorist choosing fuel $j \in \{e, g\}$,

$$v'\left({q_j}^* \cdot m\right) \cdot m + \theta_j - p_j = 0$$

To obtain the motorist's choice of miles driven and fuel demand, we re-write the equation:

$$v'\left({q_j}^* \cdot m\right) = \frac{p_{j-}\theta_j}{m}$$

The motorist's choice of miles driven is $M^* = {q_j}^* \cdot m$. Solving the above equation for M^* yields:

$$M^* = v'^{-1}\left(\frac{p_{j-}\theta_j}{m}\right),$$

and the motorist's demand for fuel type j , is ${q_j}^* \equiv M^*/m$.

2.2.2 Station-level aggregate demand

To formally aggregate individual behavior and set up the empirical section, a few more assumptions are employed. The model assumes that each E85 station serves its own market of flex motorists, meaning that each E85 station is a local monopolist for E85, and the price of E85 at other stations does not affect the station's market size. This is not too strong of an assumption as E85 fuel stations are not very common. Motorists in a station's market are aware of the prevailing E85 and E10 prices, and if they choose to refuel with E85, they visit the E85 station. If they choose E10, they may visit the E85 station (all E85 stations in Minnesota supply E10) or they may choose a nearby E10 (only) station. Note that an FFV motorist may be within the market of an E85 station even if there is an E10 station more directly along the motorist's normal driving path. As long as the motorist is aware of the E85 station, and the E85 station is not too far off of the motorist's normal driving path, then the motorist is within the station's market, and if the E85 premium is low enough, the motorist will visit the E85 station and choose E85.

The model assumes that motorist demand for miles is perfectly inelastic in the short run, and, without loss of generality, that motorists are heterogeneous with fuel demand q and willingness to pay for ethanol θ jointly distributed among motorists according to the joint probability density function (pdf) given by $f(q, \theta)$. The total quantity of E85 demanded from an E85 fuel station can be calculated as 1) the number of FFV motorists in the station's market multiplied by 2) the average fuel consumption among those motorists that choose ethanol multiplied by 3) the fraction of those motorists whose willingness to pay for ethanol exceeds the station's E85 price premium. Algebraically, this can be written as

$$Q = N \int_p^{\infty} [\int q f(q, \theta) dq] d\theta = N \int_p^{\infty} \boldsymbol{E}\,(q|\theta) f(\theta) d\theta,$$

where N is the number of flex motorists in the station's market, $\boldsymbol{E}(q|\theta) \equiv \int q f(q|\theta) dq$ is the expected fuel demand conditional on willingness to pay θ, and the expression is simplified using the fact that the joint pdf is the product of the conditional and marginal probability densities: $f(q, \theta) \equiv f(q|\theta) \cdot f(\theta)$. By multiplying and dividing by the unconditional expected fuel demand $\boldsymbol{E}(q)$ the expression can be further simplified:

$$Q = N \cdot \boldsymbol{E}(q) \cdot \int_p^{\infty} \frac{\boldsymbol{E}(q|\theta)}{\boldsymbol{E}(q)} f(\theta) d\theta = N \cdot \boldsymbol{E}(q) \cdot \int_p^{\infty} h(\theta) d\theta,$$

where $h(\theta) \equiv \boldsymbol{E}(q|\theta)/\boldsymbol{E}(q) \cdot f(\theta)$.

Anderson (2012) notes that $h(\theta) \geq 0$ and that $h(\theta)$ integrates to one, making it a proper pdf itself. One can think of $h(\theta)$ as the marginal pdf of willingness to pay for E85 among flex motorists, but instead of using the joint distribution with fuel demand given by $f(q, \theta)$, the distribution $h(\theta)$ puts weights on motorists according to fuel consumption. Defining $H(\theta)$ as the cumulative distribution function (cdf) associated with the pdf $h(\theta)$ allows us to rewrite aggregate ethanol demand:

$$Q = N \cdot \boldsymbol{E}(q) \cdot \int_p^\infty h(\theta) d\theta = N \cdot \boldsymbol{E}(q) \cdot [1 - H(p)]. \tag{2.2}$$

The model provides a direct mapping from the cdf of willingness to pay for E85 among flex motorists (weighted by volume of fuel demanded) to the station-level demand for E85. Taking the natural log of equation (2.2) yields a linear expression that provides the basis for the estimating equation we discuss in the next section:

$$\ln Q = \ln N + \ln \boldsymbol{E}(q) + \ln\big(1 - H(p)\big). \tag{2.3}$$

2.3 Empirical Model

For expositional purposes, we begin this section by re-writing theoretical equation (2.3) as:

$$\ln Q_{e_{kt}} = \ln N_{kt} + \ln \boldsymbol{E}(q_{kt}) + \ln[1 - H(p_{kt})]. \tag{2.4}$$

$Q_{e_{kt}}$ is the quantity of E85 (in E10 energy-equivalent gallons) sold by E85 station k in month t, the product $N_{kt} \cdot \boldsymbol{E}(q_{kt})$ represents the total demand for E10 and E85 by flex motorists in the market of station k in month t, and $[1 - H(p_{kt})]$ is the share of flex motorists (weighted by volume of fuel demanded) in the market of station k in month t who choose E85, given the E85 price premium p_{kt}.

We assume that the volume-weighted distribution of WTP for E85 is the same for all E85 station markets, remains constant over time, and follows a logistic distribution with mean μ and variance σ^2. This is unlike Anderson (2012) who assumes an exponential distribution of willingness to pay and focuses on the upper tail of the distribution.

The share of flex motorists who choose E85 at a given station in a given month is a function of only the station's monthly E85 premium. The logistic distribution has a sensible shape; it is symmetric, unimodal, and its support is all real numbers. Compared to the normal distribution, the logistic distribution has more mass on its tails, which is consistent with previous evidence of a large dispersion of willingness to pay for E85, and the cdf can be written in closed form. Letting $s = \sqrt{3}\sigma/\pi$, the cdf of the logistic distribution is

$$H(p_{kt}; \mu, s) = \frac{1}{1 + \exp\left(-\frac{p_{kt} - \mu}{s}\right)}.$$

Next we model the total demand for E10 and E85 by flex motorists given by $N_{kt} \cdot \boldsymbol{E}(q_{kt})$. The number of FFVs in a given station's market in a given month and the mean fuel demand of those vehicles are not observable. We therefore rely on a set of observable variables to explain the total fuel demand by FFVs in station k's market in month t. Specifically, we express the log of total fuel demand by flex motorists as

$$\ln N_{kt} + \ln \boldsymbol{E}(q_{kt}) = \gamma' \boldsymbol{X}_{kt} + \delta_k + \tau_t + \zeta_t + \omega_k \cdot t,$$

where

$$\gamma' \boldsymbol{X}_{kt} \equiv \gamma_1 \ln(\#E85stations)_{kt} + \gamma_2 M1_{kt} + \gamma_3 M2_{kt} + \gamma_4 M3_{kt} + \gamma_5 M4_{kt}.$$

$\ln(\#E85stations)_{kt}$ is the log of the total number of E85 stations operating in the same county as station k in month t, $M1_{kt}$, $M2_{kt}$, $M3_{kt}$, and $M4_{kt}$ are dummy variables for the first four months that a station sells E85, δ_k is a station fixed effect, τ_t is a month fixed effect, ζ_t is a year fixed effect, and $\omega_k \cdot t$ is a station-specific time trend. We use these measures to estimate the size of the market for each E85 station because we do not have monthly, time-series, local-level data for the number of E10 stations, the number of flex motorists, or other relevant population characteristics. We rely on the station fixed effect and the station-specific time trend variables to capture these and other attributes of the station and surrounding market for fuel.

The station fixed effects control for unobserved station characteristics that remain constant over time. These may include the presence of E85 signage, the prominence and convenience of the station's E85 pump(s), the station's location (distance to a major highway, whether in a big city or small town, etc.), and possibly other demographic characteristics that are potential determinants of local demand such as infrastructures or the availability of public transport. The month fixed effects control for seasonality in motor fuel consumption, and the year effects control for longer-term, market-wide variation in motor fuel consumption, such as the decrease in fuel consumption observed during the last recession. Finally, the station-specific time trends control for effects correlated with time such as growth in the local stock of FFVs or a gradual increase in the local median income.

The model does not control for fuel prices at nearby E85 stations. This is reasonable if E85 search costs are relatively high for consumers and/or E85 stations are relatively spread out. However if there is more than one E85 station in a relatively small area, and prices are displayed prominently, motorists may choose to forego their usual E85 station and choose a neighboring E85 station instead, and this would be problematic for our model. Fortunately, most E85 stations in Minnesota are relatively far from one another, and both E85 and E10 prices are very similar day-to-day among nearby stations, so in general there is not much to be gained by motorists from searching for the station with the lowest fuel prices. Even in cases where fuel stations offering E85 are near one another, the gains that flex motorists can expect from searching are not likely to last long because fuel stations quickly respond to competitors' prices.

The estimating equation is

$$\ln Q_{e_{kt}} = \beta_0 + \gamma' \boldsymbol{X}_{kt} + \delta_k + \tau_t + \zeta_t + \omega_k \cdot t + \ln\left[1 - \frac{1}{1+exp\left(-\frac{p_{kt}-\mu}{s}\right)}\right] + u_{kt}, \qquad (2.5)$$

where u_{kt} is the error term, and β_0, the γ-vector, δ_k, τ_t, ζ_t, ω_i, μ, and s are coefficients to be estimated. The model in (2.5) is similar to the empirical model estimated by Anderson (2012).

2.3.1 Extension of the model

We perform robustness checks and extensions on this model. First, the model in (2.5) assumes that the size of the market for an E85 station is not affected by the station's E85 premium. That is, the model assumes that motorists do not go out of their way to seek out E85 stations when the E85 premium is particularly favorable. Recall that the size of an E85 station's market is the total fuel demand by the flex motorists in the area. The empirical model in (2.5) explains a station's market size with location, signage, brand, and other factors captured by the station fixed effects and other controls in the model, but omits fuel prices.

This assumption potentially misses an important characteristic of retail fuel markets: motorists are not stationary when they are consuming fuel, as pointed out in Houde (2012). Motorists encounter many retail fuel stations along their normal driving route, and may choose one that is further out of their way if the price is favorable or not bother if the discount is not sufficient. If the size of a station's market depends on its E85 premium, then omitting the premium term from that part of the model will bias estimates of the distribution of willingness to pay. In the first extension of the model, we explore the robustness of our basic results to the inclusion of the E85 premium to explain the size of an E85 station's market. In this version of the model, flex motorists may drive out of their way to purchase E85 in months when the E85 premium is low. We assume that the marginal benefit of the money saved is decreasing, and we model the effect of the E85 premium on the size of the market as being linear in logs.

Second, the model in (2.5) assumes perfectly inelastic fuel demand in the short run and as such does not include the price of E85 to explain consumption volumes. If false, this assumption could potentially bias our results. In particular, if consumption volumes are sensitive to fuel prices in the short run and fuel prices are correlated with the E85 premium, then the zero short-run elasticity assumption would bias our estimate of the distribution of WTP. We explore the impact of the short-run elasticity assumption in an alternative specification of the econometric model where we allow the absolute fuel price to affect fuel consumption.

2.3.2 Estimating the parameters of the willingness to pay distribution

We could potentially obtain estimates of μ and s directly by applying a nonlinear estimator to equation (2.5). Unfortunately given the size of our data sample and the number of parameters in the model, estimation of the model's parameters becomes computationally intensive and the results are sensitive to the choice of starting values. We instead use a linear specification of the empirical model which, in addition to making numerical convergence easier, allows us to deal with potentially endogenous fuel prices more conveniently.

To linearize the empirical equation, we use a second-degree Taylor series approximation of $\ln[1 - H(p_{kt})]$. It is reasonable to assume a mean willingness to pay for E85 that is not too far from zero, where the cost per mile is the same for both fuels. If on average motorists' valuation of E85 relative to E10 deviates from the parity price, we do not expect it to deviate by much because E10 and E85 are overall very similar products and the attributes that differentiate them likely represent only a small share of the average motorists' valuation. Taking a second-degree Taylor approximation of $\ln[1 - H(p_{kt})]$ around $p_{kt} = 0$ yields

$$\ln[1 - H(p_{kt})] \approx \ln[1 - H(0)] - \frac{H'(0)}{1-H(0)} \cdot \ p_{kt} + \left(\frac{H'(0)^2}{\left(1-H(0)\right)^2} - \frac{H''(0)}{1-H(0)}\right) \cdot \ \frac{{p_{kt}}^2}{2}. \tag{2.6}$$

Writing the linear and the quadratic terms of the Taylor approximation as β_1 and β_2, the linearized version of equation (2.5) is

$$\ln Q_{e_{kt}} = \widetilde{\beta_0} + \beta_1 p_{kt} + \beta_2 {p_{kt}}^2 + \gamma' \boldsymbol{X}_{kt} + \delta_k + \tau_t + \zeta_t + \omega_k \cdot t + u_{kt}, \tag{2.7}$$

where $\widetilde{\beta_0} = \beta_0 + \ln[1 - H(0)]$. β_1 and β_2 are parameters to be estimated that are functions of the parameters μ and s of the distribution function $H(p_{kt})$. More specifically, given that we use a logistic distribution function, the expressions for the parameters β_1 and β_2 are

$$\beta_1 = \frac{-1}{s(1 + e^{\mu/s})}; \tag{2.8}$$

$$\beta_2 = \frac{-e^{\mu/s}}{2s^2(1 + e^{\mu/s})^2}. \tag{2.9}$$

Solving (2.8) and (2.9) allows us to obtain estimates of μ and s (and in turn σ) from linear estimation.

In the first extension of the empirical model, we use a third-order approximation of $\ln[1 - H(p_{kt})]$, and we use the coefficients on the E85 premium squared and E85 premium cubed to recover estimates of μ and σ. We do this to allow the E85 premium to linearly affect the log of the size of an E85 fuel station's market. That is, if a motorists' decision to enter a particular E85 fuel station's market is a function of the E85 premium that is relatively linear in logs, then the coefficient β_1 captures both the decision of motorists to enter the E85 station's market and the decision of motorists already in the station's market to choose E85 instead of E10. Under this assumption, the coefficients for the E85 premium squared and cubed solely capture willingness to pay. With a third-degree Taylor approximation and a logistic distribution function for willingness to pay, the expression for β_3 is

$$\beta_3 = \frac{(1 - e^{\mu/s})e^{\mu/s}}{6s^3(1 + e^{\mu/s})^3} \tag{2.10}$$

In all of the extensions we perform using the cubic model, we use estimates of β_2 and β_3, and we solve equations (2.9) and (2.10) numerically to estimate values for μ, s, and in turn σ.

2.3.3 Identification and estimation

We estimate the econometric model using both ordinary least squares (OLS) and the generalized method of moments with instrumental variables (IV GMM). In the IV GMM estimation, we instrument for the E85 premium, E85 premium squared, and E85 premium cubed to address the potential endogeneity problem. The IV GMM estimation approach uses supply-side variables to identify the parameters of the distribution of WTP for E85.

We perform OLS estimation because it is possible that the estimates for μ and s are not severely biased. Stations often set E85 fuel prices based on the wholesale E85 price, diminishing the effect of local E85 demand shifts correlating with station-level E85 prices and premiums. However there is a potential that some station-level E85 demand shocks are correlated with the stations' E85 premiums, so we also estimate the model using IV GMM.

Another reason to prefer IV GMM is to correct for endogenous measurement errors in the E85 premiums. As we describe in the next section, we observe each station's E85 price, but we do not observe each station's E10 price or the prevailing E10 price in the local market. Instead, we rely on the statewide monthly average E10 price to calculate the stations' E85 premiums. The measurement error is the difference between the actual local E10 price and the statewide average E10 price. If local E10 prices are correlated with local E85 prices, then the measurement errors are correlated with the E85 premiums. This means the OLS estimates could suffer from attenuation bias, but the IV GMM estimates do not. For example, if the price of E10 and E85 in some local market are both high in a given month, and the local E10 price is higher than the statewide average E10 price, then the 'observed' E85 premium is higher than the actual E85 premium, and the estimates overstate the share of motorists who choose E85 when the premium is high. Alternatively, if the local E10 and E85 prices are low in some month in some market such that the local E10 price is lower than the statewide average E10 price, then the 'observed' E85 premium would be less than the actual premium, and the estimates would understate the share of motorists who choose E85 when the premium is low. Therefore the OLS estimates of the distribution of willingness to pay for E85 could be biased to show a higher variance of preferences.

To instrument for potentially endogenous or mismeasured E85 premiums, we begin with a set of simple instruments that are uncorrelated with local, short-run demand shifts, but correlated with the station's E85 premium. To instrument for a station's E85 premium, based on Anderson (2012), we use the wholesale price of E10, and the wholesale price of E85[3], and we interact these two price series with the number of E85 stations per square mile and the number of all fuel stations per square mile in the same county as the station. These interactions create four variables that capture not only how wholesale prices affect retail prices, but also how local competition affects how retailers respond to those wholesale prices. A retailer in an area that is dense with E85 stations may need to lower the E85 price when the wholesale price drops

[3] The wholesale E85 price is calculated as the weighted average of the wholesale (refiner) E10 price and the wholesale (rack) ethanol price minus the value of the RIN:
$wholesale\ E85\ price = \alpha * wholesale\ E10\ price + (1-\alpha) * (wholesale\ ethanol\ price - RIN\ price)$.
The weights are according to the E85 Handbook's nominal ethanol content of E85 in Minnesota for a given month:
$E85\ ethanol\ percent = \alpha * 0.10 + (1-\alpha)$.

whereas an E85 retailer who faces less competition may be able to keep the E85 price high. In addition to these four instruments, we include the wholesale price of corn, a one-month lag of the log of the station's E85 price, a one-month lag of the log of the station's E85 quantity sold, and a one-month lag of the station's E85 premium.

Next, we use a more complex set of detailed instruments. We generate these instruments in the same manner as Anderson (2012). In addition to the list of instruments described in the previous paragraph, we use the interaction of the wholesale E10 and E85 fuel prices with the station's brand and distance to supplier. Unfortunately, the more complex set of instruments comes at a cost as we do not observe brand or exact geographic location for all fuel stations, thus forcing us to remove observations where station-specific data are not available. We discuss the instruments and estimation sample further in the next section.

2.4 E85 Data in Minnesota

The state of Minnesota has been promoting ethanol production and use with supply-side incentives since the 1980s. As a result, E85's market share in Minnesota is relatively high compared to other states, fuel stations offering E85 are relatively abundant compared to other states, and a majority of sales are to private (non-fleet) motorists. Minnesota was the first state to require that nearly all gasoline blends contain at least 10 percent ethanol and has continued to provide incentives to ethanol producers, blenders, and retailers. Minnesota supplies retail fuel stations with government loans to pay for E85 infrastructure costs. Retailers can have these loans partially or completely forgiven by reporting E85 sales volumes and revenues in a monthly survey conducted by the Minnesota Department of Commerce (MN DoC). This survey is the primary source of the data we use in our estimation.

The Minnesota data start in 1997, when only a handful of E85 stations were operating in the state, and E85 consumers were almost exclusively government fleet vehicles required by law to use E85 whenever possible. Figure 2.1 shows that from 1997 to 2004, the average monthly E85 sales volumes from E85 stations in Minnesota increased steadily from about 200 gallons to about 2,500 gallons. In 2005 and 2006 there was a large increase, and by 2006, the average monthly E85 sales volume had grown to about 7,000 gallons, and it has leveled-off since then. Figure 2.1 also shows a seasonality effect in the E85 sales volumes; Minnesotans drive more in the summer, and so that is when more fuel is sold.

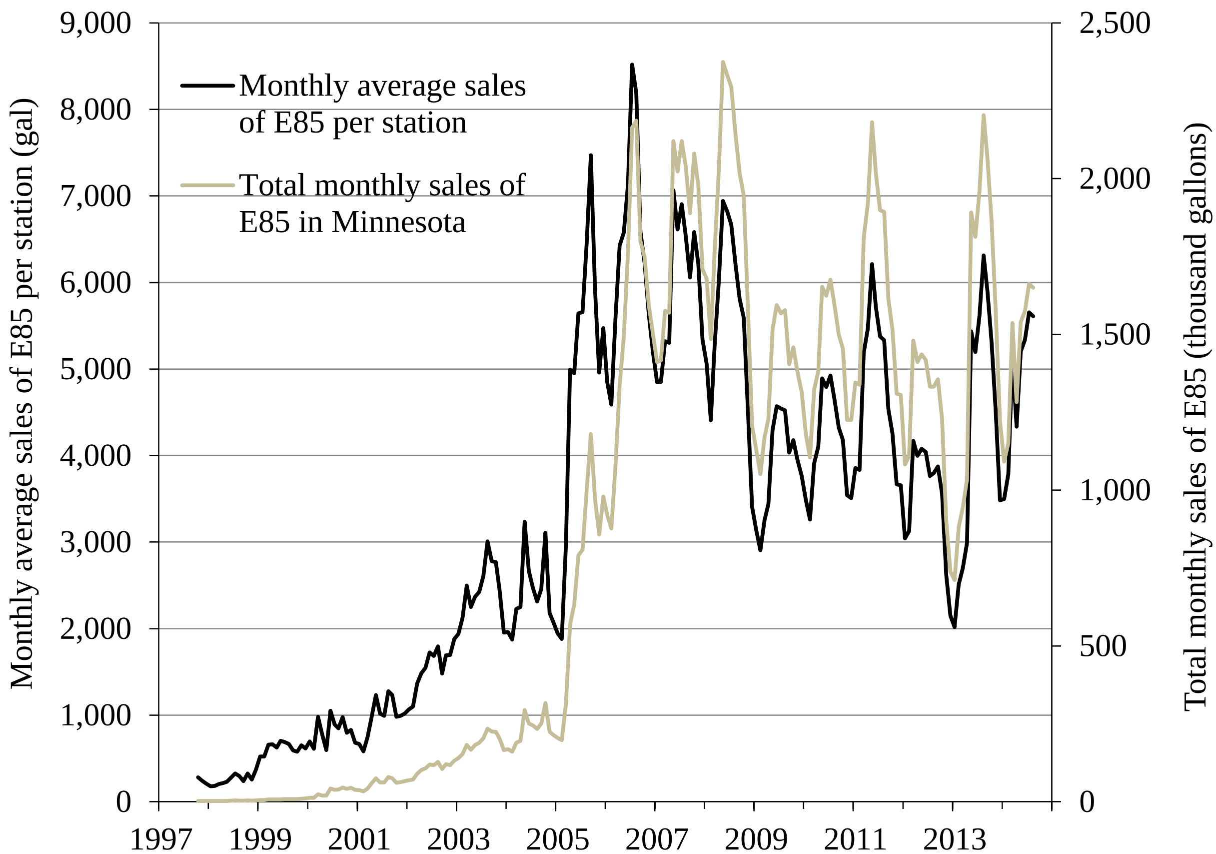

Figure 2.1 Per-station average and statewide total monthly E85 sales volumes

Data are from a survey of E85 stations conducted by the Minnesota Department of Commerce (MN DoC 2014).

Along with the average monthly E85 sales volume per station, Figure 2.1 shows the total monthly consumption of E85 in Minnesota. Even though not all E85 stations report to MN DoC each month, MN DoC keeps track of the total number of operating E85 stations, and the total monthly E85 consumption in Minnesota is calculated as the average E85 sales volume among reporting E85 stations multiplied by the total number of E85 stations operating in Minnesota that month. The total monthly quantity of E85 sold in Minnesota grew steadily from fewer than 2,000 gallons in 1997 to about 250,000 gallons by 2004. Monthly E85 sales increased to over 800,000 gallons in the summer of 2005, and again to over 1,600,000 gallons in the summer of 2006. There is a noticeable seasonal effect, but otherwise total sales seem to have also mostly leveled-off since 2006. Figure 2.1 shows that total E85 sales in Minnesota were low in 2009, 2010, and especially in 2012. That year, the United States experienced a drought that significantly reduced corn yields, causing corn prices to rise and making the ethanol in E85 more expensive.

Figure 2.2 shows that from 1997 to 2004, the number of fuel stations that offered E85 in Minnesota grew steadily from fewer than 10 to about 100, and, like E85 sales volumes, in 2005 and 2006 the number of E85 stations increased significantly, so that there were about 300 E85 stations in Minnesota by the end of 2006. Figure 2.2 also shows that the growth in the number of retail E85 stations plateaued at around 350 in 2009, and there was a small drop in the number of E85 stations in Minnesota at the beginning of 2014. Because both the average E85 sales per station and the number of E85 stations in Minnesota increased sharply between 2004 and 2006, the increase in total statewide E85 consumption in those years was even more prominent.

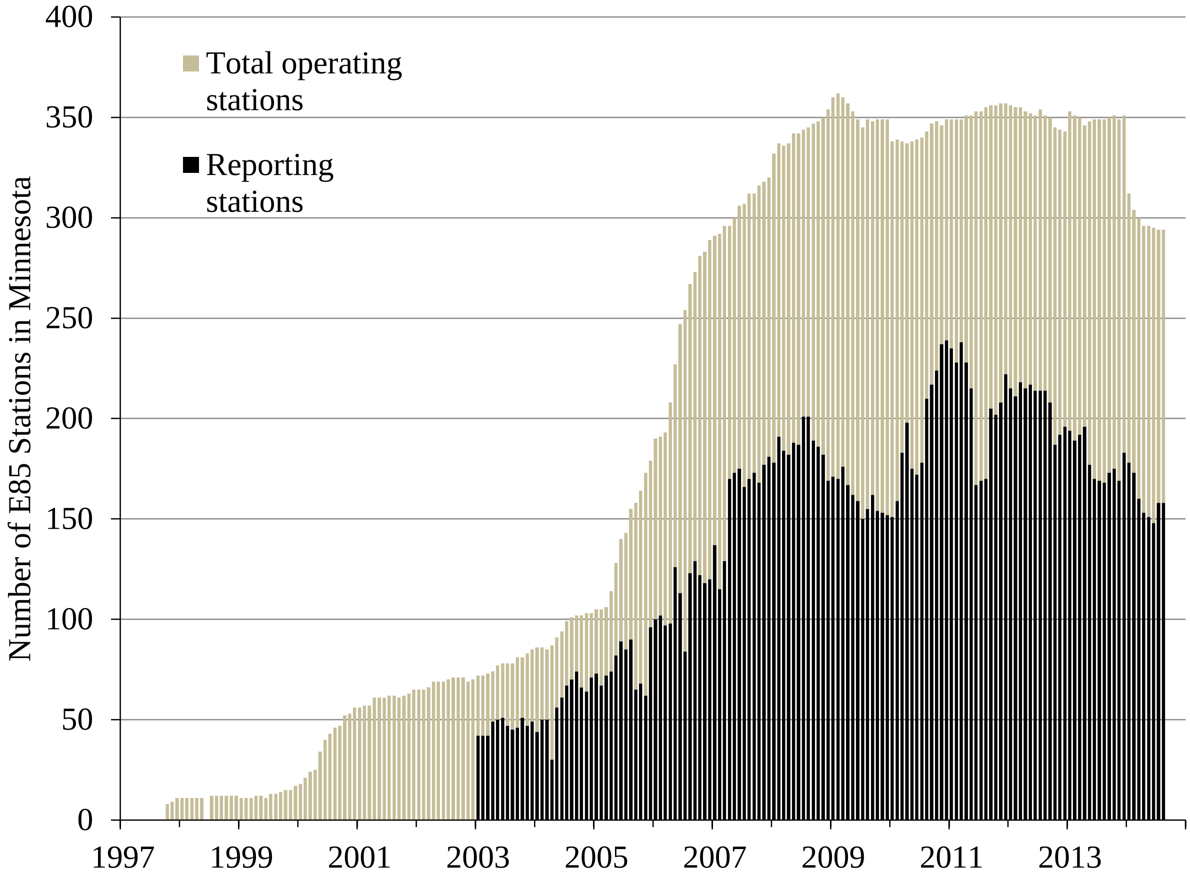

Figure 2.2 Number of retail E85 stations in Minnesota

Data are from a survey of E85 stations conducted by the Minnesota Department of Commerce (MN DoC 2014). MN DoC provides the number of reporting stations starting in January, 2003

2.4.1 Estimation sample

Although the original survey dataset contains more than 21,000 monthly observations from 413 stations, our initial estimation sample consists of 4,891 observations from 58 stations. The reasons for dropping observations are: 1) we remove any E85 price or quantity observations that are extreme outliers likely resulting from reporting error (such as months where the total quantity sold or average price is zero), 2) we use one-month lagged values as instruments so any observation without an observation the preceding month is incomplete, 3) we only use observations from stations with at least forty-eight complete observations to reduce sampling error, 4) we only use observations from the most recent eight years of data – from September 2006 to August 2014, and 5) we only use observations from stations in the Twin-Cities area[4].

We do not include observations from between 1997 and 2005 because almost all of the E85 sales during that period were to government vehicles required by law to use E85. Neither FFVs nor E85 infrastructure were common during that period and data from that time likely misrepresent the preferences of today's FFV motorists. The reason we only use observations from E85 stations in the Twin-Cities area is to minimize potential measurement error between the local E10 prices and the statewide average E10 price. The statewide average E10 price is likely close to the average price in the Twin-Cities area, the only metropolis in Minnesota, and if E10 prices vary in different parts of the state, we avoid the measurement error. Using data from only Twin-Cities stations reduces the number of fixed effects in the model and greatly reduces possible errors from numerical optimization. As we explain in section 2.4.3, we have full station information including brand and exact latitude and longitude for 56 of the 58 stations, accounting for 4,763 of the 4,891 observations.

The model was estimated for data samples of different sizes varying between more than 15,000 observations to around 1,000 observations. Results for these alternative data samples are not presented in this dissertation; conclusions that we draw based on data for the Twin-Cities area are robust to the data sample that we choose.

Table 2.1 shows summary statistics for the initial estimation sample of all the E85 stations, and Table 2.2 shows summary statistics for the subset of stations with full location information. Comparing Table 2.1 to Table 2.2, the summary statistics do not suggest that the data samples are decidedly different from each other. However, to examine the possibility of

[4] The estimation sample only includes E85 stations located in metro-area counties: Anoka, Carver, Dakota, Hennepin, Ramsey, Scott, Wright, and Washington.

sample selection and to see how it impacts our results, we estimate the model with OLS and with IV GMM using both the dataset containing observations from all stations as well as the smaller dataset containing observations from only identified stations.

Table 2.1 Summary statistics for estimation sample with all Twin-Cities area stations

Variable	Mean	Std. Dev	Min.	Max.
Monthly retail E85 sales volume (gal)	5,745.718	3,777.095	7.931	38,955.766
Retail E85 price ($/gal).	3.432	0.623	1.779	4.810
Retail E10 price ($/gal)	3.283	0.515	1.869	4.241
Retail E85 premium ($/gal)	0.149	0.226	-0.599	1.171
Wholesale E85 minus RIN ($/gal)	3.040	0.526	1.956	4.197
Wholesale E10 price ($/gal)	2.601	0.501	1.194	3.636
Wholesale corn price ($/bu)	5.516	1.420	2.913	8.295
Retail E85 station age (months)	75.045	45.034	2.000	178.000
Number of E85 stations in county	18.685	6.446	1.746	38.552
E85 stations per sq mi in county	0.114	0.086	0.039	0.276
All fuel stations per sq mi in county	0.414	0.343	0.088	1.143

Statistics are for 4,891 monthly observations from 58 E85 stations between 9/2006 and 8/2014. The counties included are: Anoka, Carver, Dakota, Hennepin, Ramsey, Scott, Washington, and Wright. E85 prices and volumes are in E10 energy-equivalent terms. Prices are 2014 dollars.

Table 2.2 Summary statistics for estimation sample with identified Twin-Cities area stations

Variable	Mean	Std. Dev	Min.	Max.
Monthly retail E85 sales volume (gal)	5,849.579	3,768.366	105.483	38,955.766
Retail E85 price ($/gal).	3.433	0.622	1.779	4.767
Retail E10 price ($/gal)	3.286	0.514	1.869	4.241
Retail E85 premium ($/gal)	0.147	0.225	-0.599	0.961
Wholesale E85 minus RIN ($/gal)	3.039	0.526	1.956	4.197
Wholesale E10 price ($/gal)	2.604	0.499	1.194	3.636
Wholesale corn price ($/bu)	5.518	1.419	2.913	8.295
Retail E85 station age (months)	75.463	45.380	2.000	178.000
Number of E85 stations in county	18.693	6.501	1.746	38.552
E85 stations per sq mi in county	0.116	0.087	0.039	0.276
All fuel stations per sq mi in county	0.421	0.345	0.088	1.143

Statistics are for 4,763 monthly observations from 56 E85 stations between 9/2006 and 8/2014. The counties included are: Anoka, Carver, Dakota, Hennepin, Ramsey, Scott, Washington, and Wright. E85 prices and volumes are in E10 energy-equivalent terms. Prices are 2014 dollars.

2.4.2 Dependent and independent variables

As explained, the data for E85 prices and sales volumes come from MN DoC. Each month, MN DoC surveys every retail E85 station all over the state. The stations report E85 sales volumes and revenues which are used by MN DoC to calculate volume-weighted monthly average prices. Not all stations report in every month, but E85 stations that received government funding to pay for their infrastructure costs can have those loans partially forgiven by reporting, and many stations participate voluntarily. MN DoC also provide the total number of E85 stations operating in the state each month. The number of E85 stations in Minnesota grew from fewer than 10 to around 350 during the timespan of the data. On average about 54 percent of stations reported sales volumes and prices to MN DoC, as shown in Figure 2.2.

We use the data from MN DoC to tabulate the number of E85 stations in each county that respond to survey in each month, and we divide that number by the fraction of the statewide E85 stations that participate in the survey that month. This variable acts as a proxy for the number of E85 stations operating in each county in each month under the assumption that the proportion of E85 stations that report is the same across counties. Next, we generate dummy variables for the first, second, third, and fourth month after a station begins reporting. We assume the first month a station reports to MN DoC is the first month that the station sold E85. We use these variables to explain the size of a particular E85 station's market. Flex motorists in the area may take some time to learn of the existence of the E85 station and to observe the E85 premium.

We convert the E85 prices and sales volumes into E10 energy-equivalent units. Almost all regular gasoline in Minnesota is E10 and contains roughly 10 percent ethanol during any given month of the year, but the amount of ethanol in the E85 fuel blend depends on the season. In the winters, a higher concentration of gasoline is needed to ensure proper starting in cold conditions. According to the E85 handbook published by the US Department of Energy (DOE 2008), E85 in Minnesota contained between 70 and 79 percent ethanol for most of the duration of the data collection period – 70 percent in the winter months and only reaching 79 percent in July. Using these blend concentrations, and assuming that pure ethanol has two-thirds the energy content per volume as pure gasoline, we calculate conversion factors for each month ranging from 1.26 in January to 1.31 in July. The E85 prices are multiplied and the E85 quantities are divided by the factors to convert to E10 energy equivalence.

To calculate the E85 premium, we obtain monthly data on the retail price of regular unleaded E10 gasoline in Minnesota from the US Energy Information Administration (EIA). EIA surveys around 800 retail locations across the country each week to obtain price data, and it also uses monthly sales reports from petroleum resellers and retailers (EIA 2013 and EIA 2014a). These price data and the E85 price data from MN DoC include all taxes and are the end prices paid by the consumer. EIA combines these price data with other sales and population data to calculate weighted average price estimates at the state level.

We convert the retail E85 prices and the retail E10 prices into August 2014 dollars using monthly CPI data from the US Department of Labor Bureau of Labor Statistics (BLS 2014). Figure 2.3 shows the energy-adjusted real retail price of E85 from each station in each month in our sample along with the statewide average real retail price of E10.

We calculate the E85 premium as the difference between the energy-equivalent real retail price of E85 and the real retail price of E10. Figure 2.4 shows the E85 premiums at the E85 stations in our sample. Each individual dot in Figure 2.4 shows the E85 premium at one station in one month, and the line shows the average E85 premium from among the reporting stations. When the E85 premium is positive, the energy-adjusted price of E85 is higher than the price of

E10. From September 2006 through August 2014, the energy-adjusted E85 premium in Minnesota was mostly positive. Corn and ethanol prices fell in 2013 and 2014, and the E85 premium fell sharply as well. Note that although the average energy-adjusted E85 price has almost always been higher than the average price for E10, there are several instances where individual stations have offered E85 at a discount relative to the average E10 price.

Figure 2.3 Retail E10 and energy-equivalent E85 prices

The data are from EIA (2013, 2014a) and MN DoC (2014). Each dot represents an observation of the volume-weighted monthly average E85 price from an E85 station. The black line is the statewide average E10 price. E85 prices are measured in E10 energy equivalents, and all prices are in real August 2014 dollars per gallon.

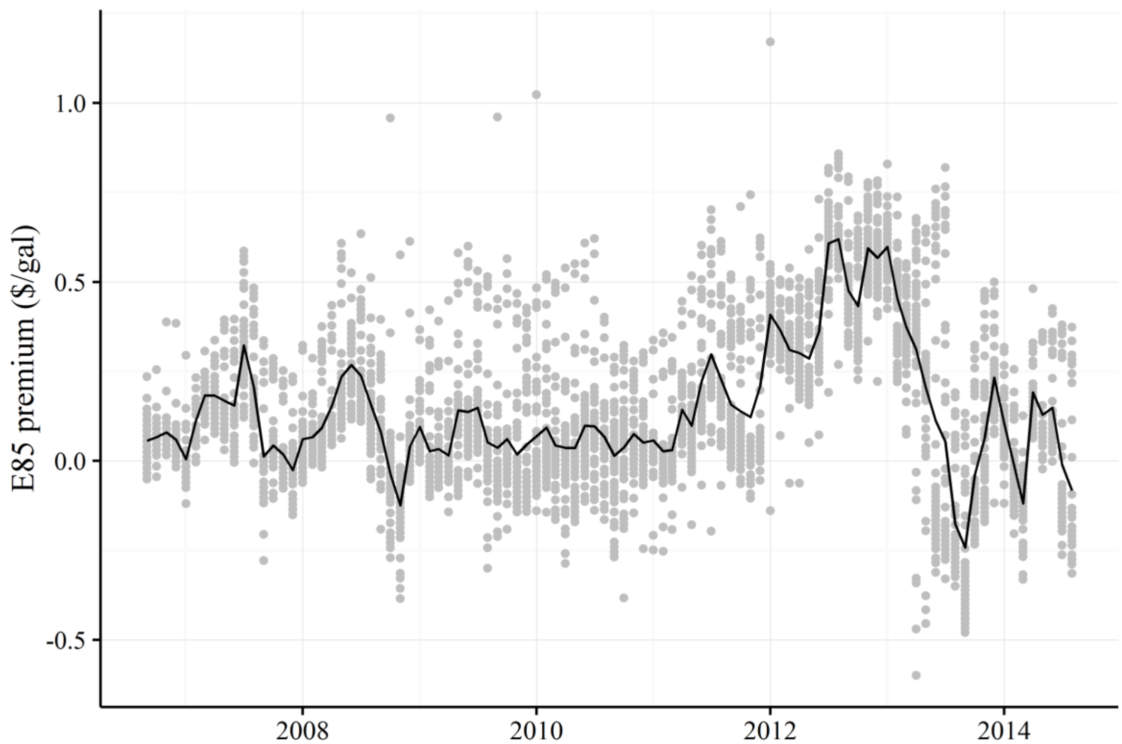

Figure 2.4 Energy-equivalent retail E85 premiums

Data are from EIA (2013, 2014a) and MN DoC (2014). Each dot represents a monthly observation of the E85 premium from an E85 station. The black line is the average E85 premium from among reporting E85 stations. E85 prices are measured in E10 energy equivalents, and all prices are in real August 2014 dollars per gallon.

2.4.3 Instrumental variables

As mentioned briefly in the previous section, our initial set of simple instruments consists of the wholesale prices of E10, E85, corn, and the density of E85 and all fuel stations in the same county, as well as one-month lags of the log of the station's E85 quantity sold, the log of the station's E85 price, and the station's E85 premium. In this section, we provide information about the sources of the instrumental variables.

EIA provides monthly data on the wholesale price of E10 in Minnesota (EIA 2014b). Monthly data for the wholesale price of ethanol are obtained from the Nebraska Energy Office (NEO). NEO reports ethanol average rack prices in Omaha, NE each month. The rack price is the price for truck quantities of pure ethanol charged by ethanol producers to blenders, resellers, and other various clients at the given location (NEO 2014). Because Omaha is relatively close to Minnesota, the Omaha price is likely close to the price paid in Minnesota. We subtract the monthly average RIN price from the rack ethanol price. We obtain RIN price data from the Oil Price Information Service (OPIS). We calculate the wholesale price of E85 in each month as the weighted average of the wholesale E10 price and the rack price of ethanol minus the RIN value. The weights are based on the monthly average ethanol concentration in E85 reported by DOE (2008). We then convert the wholesale E85 price series into E10 energy-equivalent dollars and convert both the E10 and the E85 wholesale fuel price series into August 2014 dollars. We obtain wholesale corn prices from the Chicago Board of Trade (CBOT) by taking the average of the daily prices of the nearest corn futures contract for each month.

As in Anderson (2012), we interact the wholesale E10 and E85 price series with measures of local competition. We calculate the density of E85 stations and the density of all fuel stations in the county where the station is located. We obtain the number of E85 stations in each county from a list maintained by the Alternative Fuels Data Center (AFDC) that provides a snapshot of the E85 retail stations operating in Minnesota in September 2013. The number of E10 retail fuel stations in each county is obtained from MN DoC in a separate dataset and also represents a snapshot of the operating stations in Minnesota in September 2013. We obtain the area (in square miles) of each county in Minnesota from the US Census, and we calculate the E85 and E10 station densities as the number of stations per square mile. The intuition for using these variables as instruments is that retailers facing stiff competition may be more inclined to behave as competitive firms who set their price equal to the marginal cost. On the other hand, E85 retail stations not facing competition may behave as local monopolists, and their retail prices may therefore be less tied to the wholesale prices.

The dataset that MN DoC provided us has the county where each station is located but not the exact geographic location. However, the AFDC's list of E85 stations provides the station's exact geographic coordinates, the station's name, the station's county, and the date the station first started selling E85. By cross-referencing the AFDC list of stations with the data from MN DoC, we are able to infer which E85 price/quantity series belong to which E85 station based on the station's county and the month and year the station began selling E85. Using this method, we are able to positively identify 306 of the 413 stations in the original dataset. The remaining stations could not be identified for one of two reasons. First, we were not able to identify stations that closed before September 2013 and thus were not on the AFDC's list of E85 stations. Second, we were not able to uniquely identify stations from the same county with the same start date (month and year). For reasons discussed in the previous section, we limit the initial estimation sample to 58 of the 413 stations in the dataset. We are able to positively identify 56 of these 58 stations.

For the identified stations, we measure an individual E85 retailer's supplier relationship by calculating the log of the distance (in miles) from the station to the nearest ethanol blending terminal. In addition to capturing a supplier-relation effect, this distance variable also captures the direct, supply-side, transportation cost of supplying the fuel to the station. We create dummy variables for each brand affiliation. Any brand with at least two stations has its own dummy variable, and any station with a unique brand or a brand we could not identify we designate as, 'Other'. This method generates 6 brand categories for the 56 stations.

To construct the more complex set of price instruments that utilizes precise location and brand variables to capture how individual retail stations respond to changes in supply-side costs, we interact the wholesale E85 prices and wholesale E10 prices with the number of E85 stations per square mile in the county, the number of E10 stations per square mile in the county, the logged distance in miles to the nearest ethanol blending terminal, and the 6 brand dummies. These interactions produce a total of 18 instrumental variables. We also keep the rest of the initial instruments: the wholesale price of corn, and the one-month lag values of the station's log of E85 quantity sold, log of E85 price, and E85 premium. The instruments allow us to remedy the endogeneity problem by modeling retail E85 pricing behavior that is exogenous to local, short-run shifts in E85 demand. In addition, instrumenting for the E85 premium in this fashion allows us to correct for the potential measurement errors in the premium discussed in Section 2.3.3.

2.5 Econometric Estimation and Results

We estimate the model given in equation (2.7) under several specifications to verify the robustness of our estimates to the choice of instruments, the estimation sample selected, and the assumptions about motorists' motives to fill at a fuel station. What is common in all specifications is that we apply the standard one-way fixed effects model by subtracting the station's mean value of the variable for all observations and performing estimation on the transformed data (Baltagi 2013). We choose a fixed-effects model over a random-effects model because of the potential correlation between a station's fixed effect and its premium.

We estimate the model using either all 4,891 observations or the 4,763 observations for which we have brand and location information. Both data samples are Twin-Cities area stations and cover the period between September 2006 and August 2014. We label the sample with 58 stations as ‘All’ and the sample with 56 stations as ‘Identified’. In Model 2.1, we estimate the model using OLS and the estimation sample with All stations. In Model 2.2, we estimate the model using OLS and the sample with Identified stations. Then in Models 2.3 and 2.4, we estimate the model using IV GMM and the simple set of instruments with the All and Identified estimation samples respectively. In Model 2.5 we use the complex set of instruments with the Identified sample.

Table 2.3 shows the results. The table shows coefficient estimates for the E85 premium, the E85 premium squared, and the log of the number of E85 stations operating in the same county. Standard errors are in parentheses. The table also contains the estimates of the means and standard deviations of the distribution of willingness to pay implied by the coefficients for the premium and the premium squared. Values of μ and σ are calculated solving equations (2.8) and (2.9) and their standards errors are calculated using the delta method. Recall that we model fixed effects for each station, year effects, month effects, station-specific time trends, and dummy variables for the first four months the station sells E85. We do not report the coefficient estimates for these variables for each of our estimations, but we find E85 demand is highest in the months of May, June, July, and August and lowest in November, December, January, and February. The year effects are the most negative (compared to 2007) in 2012, 2013, and 2014. Appendix A contains complete tables of results.

Table 2.3 Estimation results of initial models

	Model 2.1 All OLS	Model 2.2 Identified OLS	Model 2.3 All IV GMM	Model 2.4 Identified IV GMM	Model 2.5 Identified IV GMM
Stations	58	56	58	56	56
Observations	4,891	4,763	4,891	4,763	4,763
Instruments	n/a	n/a	simple	simple	complex
E85 premium	-0.892	-0.855	-0.175	-0.110	-0.258
	(0.029)	(0.028)	(0.207)	(0.186)	(0.139)
E85 premium^2	0.017	-0.076	-3.095	-2.336	-2.125
	(0.055)	(0.053)	(0.238)	(0.482)	(0.361)
ln(E85 stations)	-0.200	-0.208	-0.238	-0.221	-0.349
	(0.027)	(0.026)	(0.060)	(0.062)	(0.055)
μ	n/a	-1.51	-0.15	0.14	0.25
	n/a	(0.88)	(0.09)	(0.18)	(0.09)
σ	2.13	1.76	-0.05	0.04	0.11
	(0.26)	(0.19)	(0.05)	(0.08)	(0.07)
R-squared	0.668	0.700			

The dependent variable is log E85 sales. Estimates for the month effects, year effects, and other control variables are not shown. Estimates of the mean μ and the standard deviation σ of the WTP distribution are calculated by solving equations (2.8) and (2.9), and their standard errors are calculated using the delta method. Complete estimation results are in Appendix A.

A common result through the models is that the coefficient estimate on the log of the number of E85 stations in the county is negative, and it ranges between -0.20 and -0.35 depending on the estimation sample and model specification. This implies that increasing the number of E85 stations in a county by 10 percent will reduce the E85 sales volumes of the other E85 stations in the county by 2 to 3.5 percent conditional on the location decisions of the new E85 stations. The relatively low estimated decrease in sales volumes from additional E85 stations suggests that E85 markets are relatively distinct or that E85 retailers choose to locate in areas where E85 is not already available even within the Twin-Cities area.

Model 2.1 estimates a positive coefficient on the E85 premium squared, meaning that we cannot solve for μ at the point estimate because μ is a function of the log of the negative premium squared coefficient. Actually, the premium squared coefficient is not statistically different than zero so even if the sign were negative, the resulting estimates of μ and σ would be questionable. This is the case for Model 2.2, where the sign on the premium squared becomes negative, but the coefficient is still not significant. Recall the two models are identical except that Model 2.2 has 128 fewer observations from 2 fewer stations.

As mentioned earlier, there are good reasons, including measurement errors, to suspect that the E85 premiums are correlated with the volumes of E85 sold and should be treated as endogenous. Therefore we estimate the model using IV GMM and the set of basic instruments for both estimation samples in Models 2.3 and 2.4. The results of the IV GMM models are extremely sensitive to the starting values and method for the numerical optimization. This raises questions about the strength and validity of the instruments as well as model specification. If μ, the average willingness to pay is positive, then at energy-equivalent prices, the majority of flex motorists in the Twin Cities would be using E85. We know that this is not the case based on the total E85 sales volumes and separate estimates of the number of FFVs in Minnesota and average fuel demand.

The results of the IV GMM models are drastically different than the results of the OLS models. In the case of the All sample, the estimate of μ rises to $-\$0.15$ per gallon, and in the case of the Identified sample, the estimate of μ is \$0.14 per gallon. Whereas in the OLS models, the squared premium term was not statistically significant, in the IV GMM models, the linear premium term is not significant, so again the calculated values of μ and σ are questionable. In addition, the estimated coefficients in Model 2.3 are such that the calculated standard deviation is negative. In Model 2.5, we expand to the complex set of instruments that require identification of the fuel stations for brand and supplier distance variables, and both the linear premium and squared premium term coefficient estimates are statistically significant. The estimated WTP distribution has a mean of $0.25 and a standard deviation of $0.11. This implies a relatively narrow distribution of preferences and that almost all flex motorists are willing to pay a premium to use E85 which are unrealistic results.

Next the model is extended to allow the possibility that a station's E85 premium influences the size of the station's market. We suppose that if the station's E85 premium is low enough, motorists who would not normally consider visiting the station might choose to visit the station and choose to refuel with E85 simultaneously. For example a flex motorist may be willing to purchase E85 when it is offered at the station's prevailing E85 premium if it were offered at a station along the motorist's normal route, but, for the motorist, visiting the E85 station requires driving out of the way at an additional cost. Then we could imagine that if the E85 premium fell sufficiently, the flex motorist might decide to visit the E85 station. However, the E85 premium required to induce the motorist to travel to the E85 station and refuel with E85 would be lower than the motorist's true willingness to pay for E85. In this way, the station's E85 premium affects the share of motorists who choose E85 among its regular flex motorist patrons, and it also affects the size of the E85 station's market.

To capture this behavior in our model, we assume that the linear premium term affects both the size of the station's E85 market as well as the share of flex motorists within the market who choose E85. We include a term in the regression for the E85 premium cubed, and we use the second and third degree premium coefficients to recover estimates for the mean and variance of the willingness to pay distribution. Note that we assume the coefficients on the premium squared and cubed are not biased by the decision of a flex motorist to drive to a particular fuel station, and this implicitly assumes that the effect of an E85 station's premium on the size of its market is linear in logs, consistent with decreasing marginal utility of money.

Table 2.4 compares the estimates from the empirical specification with only the linear and squared premium terms with the estimates obtained from the third-degree model. Model 2.6 is analogous to Model 2.2 using identified observations and OLS, and Model 2.7 is analogous to Model 2.5, using identified observations and IV GMM with the complex set of instruments.

Model 2.6 is the most promising model estimated based on the results. The coefficient values for the linear, squared, and premium terms are all statistically significant, and the implied parameter values for the WTP distribution are a meaning willingness to pay for E85 of −\$0.74/gallon with a standard deviation of \$0.43. These parameter estimates imply that when the fuels are priced evenly on an energy-equivalent basis, about 15 percent of flex motorists choose E85. Model 2.6 assumes the premium term affects the size of each E85 station's market, which could make sense in the Twin-Cities area if E85 stations are dense and competing for flex motorists. The model also treats the stations' E85 premiums as exogenous from the station-specific, one-month, unobservable E85 demand error terms (conditional on the other control variables such as month effects to capture seasonality). If E85 retailers are setting prices based on E10 prices or are not adjusting E85 prices with demand shocks, then premiums could be exogenous, and the model is properly specified.

Table 2.4 Squared and cubed premium estimation results

	Model 2.2 Squared OLS	Model 2.5 Squared IV GMM	Model 2.6 Cubic OLS	Model 2.7 Cubic IV GMM
Stations	56	56	56	56
Observations	4,763	4,763	4,763	4,763
Instruments	n/a	complex	n/a	simple
E85 premium	-0.855	-0.258	-0.878	-0.220
	(0.028)	(0.139)	(0.028)	(0.131)
E85 premium^2	-0.076	-2.125	-0.362	-0.825
	(0.053)	(0.361)	(0.101)	(0.917)
E85 premium^3	n/a	n/a	0.468	-1.737
	n/a	n/a	(0.141)	(1.252)
ln(E85 stations)	-0.208	-0.349	-0.206	-0.336
	(0.026)	(0.055)	(0.025)	(0.055)
μ	-1.51	0.25	-0.74	0.48
	(0.88)	(0.09)	(0.11)	(3.36)
σ	1.76	0.11	0.43	-0.27
	(0.19)	(0.07)	(0.09)	(38.50)
R-squared	0.700		0.701	

The dependent variable is log E85 sales. Estimates for the month effects, year effects, and other control variables are not shown. For the squared models, estimates μ and σ of are calculated by solving equations (2.8) and (2.9), and their standard errors are calculated using the delta method. For the cubed models, estimates of μ and σ are calculated by solving equations (2.9) and (2.10) numerically, and their standard errors are calculated using Monte Carlo simulation. Complete estimation results are in Appendix A.

In Model 2.7, none of the three E85 premium coefficient estimates are significant at the 5 percent level. The calculated WTP distribution parameters are not precisely estimated, and the values of the coefficients are such that the calculated standard deviation value is negative. For these reasons, Model 2.6 is the preferred model which we use as a baseline to perform extensions to the empirical model.

The extension we make to the baseline Model 2.6 is that we relax the assumption that fuel demand is perfectly inelastic in the short run. We begin by imposing sensible values for the elasticity of E85 fuel other than zero, and we also estimate the price elasticity of demand directly in the OLS log-log model where quantities and prices are (perhaps incorrectly) modeled as exogenous. Table 2.5 shows the results. We find that the greater the magnitude of the elasticity parameter we impose on the log of E85 price, the smaller the magnitude of the coefficient on the E85 premium and E85 premium cubed, but the coefficient on the E85 premium squared remains virtually unchanged. For the WTP distribution this results in marginally lower estimates for the mean willingness to pay: from −\$0.74 with inelastic demand to −\$0.78 when the elasticity is fixed at −0.5. The change in coefficient estimates also results in a higher estimate for the standard deviation: from \$0.43 to \$0.62. The change in the coefficient estimates as the fuel demand elasticity increases also causes the estimated standard errors of the WTP distribution parameters to increase.

We also estimate the short-run elasticity of demand freely using OLS and thus treating prices as exogenous. The estimate for the coefficient on log E85 price is 0.59 which is troubling because theory expects the sign to be negative. This could be an indication of endogeneity where prices and quantities are determined together as market outcomes or it could be an indication of another form of model misspecification. In this extension of the model, the estimated mean willingness to pay for E85 changes to −\$0.65 per gallon with standard deviation \$0.31 per gallon.

Table 2.5 Estimation results relaxing short-run inelastic fuel demand

	Model 2.6 $\eta = 0.00$	Model 2.8 $\eta = -0.10$	Model 2.9 $\eta = -0.30$	Model 2.10 $\eta = -0.50$	Model 2.11 η free
Stations	56	56	56	56	56
Observations	4,763	4,763	4,763	4,763	4,763
E85 premium	-0.878	-0.830	-0.734	-0.636	-1.163
	(0.028)	(0.029)	(0.030)	(0.031)	(0.032)
E85 premium^2	-0.362	-0.362	-0.363	-0.363	-0.359
	(0.101)	(0.102)	(0.105)	(0.109)	(0.098)
E85 premium^3	0.468	0.433	0.362	0.291	0.675
	(0.141)	(0.143)	(0.147)	(0.152)	(0.137)
ln(E85 stations)	-0.206	-0.217	-0.240	-0.263	-0.138
	(0.025)	(0.026)	(0.027)	(0.027)	(0.025)
ln(E85 price)	0.00	-0.10	-0.30	-0.50	0.587
	n/a	n/a	n/a	n/a	(0.033)
μ	-0.74	-0.75	-0.77	-0.78	-0.65
	(0.11)	(0.12)	(0.19)	(0.30)	(0.04)
σ	0.43	0.46	0.53	0.62	0.31
	(0.09)	(0.10)	(0.16)	(0.38)	(0.05)
R-squared	0.70	0.69	0.68	0.67	0.72

The dependent variable is log E85 sales. Estimates for the month effects, year effects, and other control variables are not shown. Estimates of μ and σ are calculated by solving equations (2.9) and (2.10) numerically, and their standard errors are calculated using Monte Carlo simulation. Complete estimation results are in Appendix A.

To summarize, the results are questionable, and suggest that it may not be possible to recover the full distribution of WTP at the motorist level from the E85 station survey data. Our most credible results suggest that the average flex motorist prefers E10 when the two fuels are priced at an energy-equivalent level, but that fuel-switching behavior spans a wide range of prices. It is difficult to recover reasonable estimates of the WTP distribution parameters from the models with the exception of Model 2.6, the cubic, OLS model.

There are several (possibly compounding) explanations for the poor estimation results. The first is that the model is misspecified. For instance, an important assumption of the model is that each station has a monopoly power in selling E85. Although anecdotal evidence supports this assumption, it could nonetheless be the case that nearby E85 stations in the Twin-Cities area compete with each other. Another explanation is biased survey data from the MN DoC survey because fuel stations self-report their E85 prices and volumes. Also, it could be an artifact of the nonlinear estimation and numerical optimization procedure with the large number of factors including fixed effects and station trends, or it could be that E85 premiums are actually exogenous. Another explanation is that there is unobserved heterogeneity that is correlated with the E85 premium. That is, it is possible that our set of explanatory variables does a poor job of explaining expected fuel consumption and the number of flex motorists in a fuel station's market. The observed statewide average E10 price could also be a source of bias. Ideally, we would observe the E10 price at each fuel station, but station-level E10 price data were not made available to us. We tried minimizing the size of this potential bias by using data only for the Twin-Cities area but this might not have been sufficient.

We do not believe that our instruments are problematic. Using two-stage least squares instead of IV GMM models yields F-statistics for the strength of our instruments that are well above 10, the commonly accepted threshold. Wholesale prices are indeed good predictors of retail prices. Tests for over-identification restrictions for the IV GMM show that the instruments are not correlated with the error term so they are arguably exogenous to the demand model.

If we are to believe the results of Model 2.6, the average WTP for E85 among flex motorists in Minnesota is about $-$0.74/gallon after adjusting for the energy difference. The average E10 price in the data is $3.43 meaning that the average WTP is about 22 percent below energy equivalence. Nevertheless, the spread of WTP for E85 is expansive. The estimated standard deviation is \$0.43, meaning that about 15 percent of motorists use E85 when the fuels are priced at parity. Conversely, the implication is that 15 percent of motorists use E10 even when E85 is discounted by \$1.48/gallon.

2.6 Conclusion

The demand for ethanol as a motor fuel in the United States is an important and debated topic. Only a few studies have attempted to estimate the demand for ethanol in the United States beyond the E10 blend wall, and those studies suffer from a lack of available data. The model in this study assumes flex motorists choose between E10 and E85 based on observed prices as well as personal preferences. We use a current and extensive dataset of E85 prices and sales volumes from E85 stations in Minnesota to estimate an empirical model of motorist-level preferences for E85 as a substitute for E10.

Following Anderson (2012), we derive a model to estimate the distribution of willingness to pay for E85 based on station-level survey data. Unlike Anderson (2012), the study in this chapter attempts to estimate the full distribution of willingness to pay. Recently observed prices for E85 below prices for E10 in energy-equivalence should allow recovery of the wider distribution of WTP. Our prior belief was that motorists on average discount E85 relative to E10.

Estimates of WTP for E85 vary greatly across the models considered. We cannot conclude in favor of a specific point estimate of the mean WTP. Several factors could explain why we cannot identify willingness to pay for E85: 1) the theoretical model is inconsistent with E85 market realities; 2) the data from the MN DoC survey are biased; 3) there are problems with numerical optimization because of the large number of fixed effects and station trends; 4) there is unobserved heterogeneity that is correlated with the E85 premiums, and 5) the E85 premiums are biased.

From the work presented in this chapter, we conclude that the Minnesota E85 station survey data are not suitable to estimate a distribution of willingness to pay for E85 at the motorist level based on the methodology proposed. Chapters 3 and 4 estimate WTP for E85 based on data collected from an intercept survey of flex motorists. The chapters show robust estimates that are greatly preferred to the estimates presented in this chapter.

CHAPTER 3.

ESTIMATING WILLINGNESS TO PAY FOR E85 USING DATA FROM AN INTERCEPT SURVEY: EVIDENCE FROM REVEALED PREFERENCE

3.1 Introduction

In this study, we collect E85 sales data from fuel stations in different regions of the United States and estimate the distribution of willingness to pay for E85 as a substitute for E10 among flex motorists. We obtain data by performing an intercept survey in a similar manner to Salvo and Huse (2013). The advantages of an intercept survey over a mail or online survey are: 1) the non-response rate is much lower, 2) all the motorists in our sample drive FFVs, 3) we obtain revealed-preference data by observing actual fuel purchases as well as certain individual characteristics such as the vehicle type or the state on the license plate, and 4) we obtain additional stated-preference data about the motorists by asking them a series of short questions while they refuel their vehicles.

We apply a binary choice, random utility logit model to estimate the probability that a flex motorist chooses E85 given fuel prices, motorist characteristics, and station characteristics. The flex motorists surveyed constitute a choice-based, endogenously stratified sample of the population of flex motorists because we only conduct the survey at stations that sell E85, which make up only a small fraction of retail fuel stations. Flex motorists may select themselves into our sample by choosing to drive to the E85 fuel station specifically because it offers E85, whereas if the motorists were using E10, they could choose a different more convenient E10-only fuel station and would not appear in the sample. We correct for the endogenous stratification in our empirical models so that our estimates represent the general population of flex motorists.

We estimate relative fuel preferences using models where motorists make their fuel choices based on the difference in fuel prices (the E85 premium as in Chapter 2) as well as using models where motorists make their fuel choices based on the ratio of fuel prices. Both approaches yield similar results. We find that about 35 percent of flex motorists choose E85 when its nominal (not energy-adjusted) price is 70 percent of the E10 price, 24 percent of flex motorists choose E85 when its price is 80 percent of the E10 price, and 16 percent of flex motorists choose E85 when its price is 90 percent of the E10 price. The mean of the distribution, when 50 percent of flex motorists choose E85, is when the E85 price is about 63 percent of the E10 price. This means that on a cost per mile basis, flex motorists discount E85 by about 15 percent on average. Furthermore we find that, accounting for fuel prices, the probability that a flex motorist chooses E85 is not significantly different between the regions where we conducted our survey, other than for California, where we find that flex motorists are significantly more likely to choose E85 than elsewhere. However, in California, the E85 retail model is different and there are other confounding factors.

The next section of this chapter offers details on the intercept survey design. In Section 3.3 we describe the theoretical models, sample selection, and estimation technique. We discuss the data in Section 3.4. Section 3.5 presents the empirical models and estimation results. Finally Section 3.6 concludes.

3.2 Intercept Survey Design

The intercept survey was designed to obtain data on a broad range of factors that might affect flex motorists' WTP for E85 as a substitute for E10. The survey was conducted by first observing the motorists' fuel choices from afar and then interviewing motorists while they were refueling. This allowed us to obtain revealed preferences and stated preferences for the fuel choices at observed and hypothetical prices and collect information about the motorists. We completed each interview in about two minutes. The complete survey questionnaire is available as Appendix C.

3.2.1 Intercept survey method

For each station we visit, we begin by recording the following data on a station-level form: the date and start time of the visit, the station name and brand, the station address, the prices of the E10 fuels (usually regular, midgrade, and premium), the price of E85, the number of E10 pumps, the number of E85 pumps, and whether there is E85 price signage. We also record the date and end time of the visit upon leaving a station. If a station changes the price of one or more fuels at some point during the station visit, the interviewer completes the current station-level form and begins a new one with the updated prices.

The procedure used to choose which flex motorists to interview is that whenever the interviewer is idle, the interviewer targets the next flex motorist to pull alongside any of the station's pumps. This is true both at the beginning of the station visit and between each interview. If a second flex motorist pulls up to a pump while an interview is being conducted, then the interviewer does not interview the second flex motorist. Instead, when the interviewer completes the first interview, the interviewer resets and once again waits to target the next flex motorist to pull alongside any of the station's pumps. This sequencing rule avoids possible selection bias by the interviewer. In practice, the share of the vehicle fleet that are FFVs is small and the survey is quick, and as such, despite the strict sequencing rule, we manage to capture virtually all flex vehicles refueling at the E85 stations during our visits.[5]

We visually identify 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 an FFV. 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 is an FFV is if the motorist chooses E85, but it could be that the motorist is making a mistake (by choosing E85 for a conventional vehicle not equipped to use it) or has a vehicle with aftermarket modifications to use E85.[6]

3.2.2 Survey questions

Before intercepting a motorist, the interviewer passively observes characteristics about the motorist and the motorist's fuel choice. The observable characteristics recorded on the motorist-level form are: the vehicle make, the vehicle model, the vehicle type (car,[7] truck, SUV,

---

[5] At the majority of the stations we visited, we observed an average of between two and four flex motorists per hour fueling their vehicles, depending on the location of the station, day of the week, and the weather that day, among other factors.

[6] 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 vehicle's 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.

[7] The vehicle type, 'car' includes most small to midsize vehicles like sedans, coupes, convertibles, hatchbacks, station wagons, and any vehicle that was not broadly a truck, SUV, or van.

or van), the state on the license plate, whether the vehicle has an FFV badge, whether the vehicle has a yellow gas cap, and the gender of the motorist. The interviewer also records the transaction volume and expenditure at the end of the interview when the motorist finishes refueling.

Once a motorist begins refueling, the interviewer approaches and asks whether the motorist is willing to participate in a short survey. Next the interviewer asks, “Is this your personal vehicle?” This question is important to inform whether the motorist is the one making the fuel choice and paying for the fuel. We also want to know whether motorists are aware that their vehicles are capable of using E85. If the motorists choose E85, it is apparent that they are aware of their vehicles’ capabilities, but to the motorists who choose E10, we ask, “Is your vehicle a flex-fuel vehicle capable of using E85?” We already know that the vehicles are FFVs by inspection, but we want to know what share of motorists are unaware of the flex capability of their vehicles. When the E10 motorists respond that their vehicles are indeed FFVs, we ask, “Have you ever fueled this vehicle with E85?” We also ask the E10-using motorists, “Did you know that this station supplies E85 fuel?” We want to know what share of flex motorists who visit the station are aware that the station offers E85.

To correct for the endogenous stratification in the sample, we want to know if the motorists we survey are representative of the general population of flex motorists or if they ended up in our sample only because they sought out E85. We discuss this in more detail in the next section. For the motorists who choose E10, we know that they did not come specifically for the E85, and we assume that they would still choose to refuel at the particular station where we conduct our survey even if it did not offer E85. That is, flex motorists who choose E10 are randomly sampled from the local population of flex motorists. But for motorists who choose E85 we ask, “Did you choose to fuel at this station because it offers E85?” If they respond positively, we follow by asking, “How far out of your way did you have to drive?” We use responses to these questions to sort the motorists who choose E85 and calculate the share of the general population of flex motorists who choose each fuel (as opposed to the share of our sample who choose each fuel). We describe the method in Sections 3.3 and 3.4.

At this point in the survey, we ask a question to obtain SP data on WTP to complement the RP data we observe from the fuel purchase. The interviewer proposes a hypothetical scenario with different fuel prices that may induce the motorist to switch to the other fuel. If the motorist is fueling with E10, the scenario is one where either the price of E10 is increased or the price of E85 is decreased. If the motorist is fueling with E85, the scenario is one where either the price of E85 is increased or the price of E10 is decreased. The models in this chapter use only the RP data, and in Chapter 4 the models incorporate the SP data.

Next we collect data on the various factors that might cause motorists to perceive E85 and E10 as imperfect substitutes and discount one relative to the other. After a motorist responds to the hypothetical fuel price scenario, we ask, “On average, how many miles do you drive per year?” The reason we ask this question is that E85 requires more frequent refueling, which could lead long-distance drivers to discount E85 if the costs of refueling are convex. Alternatively, long-distance drivers may be more conscientious of their fuel choice and its perceived externalities or benefits and instead put a premium on E85.

The next question of the survey is, “How old are you?” Most motorists freely offer a (presumably honest) response, and if the motorist does not want to answer, the interviewer moves on to the next question. The motorist’s age might be a factor because younger motorists may be more likely to adopt new fuel technology and put a premium on E85. On the other hand, older motorists may put a premium on E85 for its perceived role in providing independence from foreign oil.

In the final part of the survey, we ask a series of fuel opinion questions. The motorist is asked to answer either ‘Ethanol’, ‘Gasoline’, or ‘No Difference’. Here the interviewer offers the more colloquial names for the fuels but when prompted or deemed necessary clarifies that ‘Ethanol’ and ‘Gasoline’ refer to the E85 and E10 fuel choices at the station. We then ask these four multiple-choice questions: “Which fuel is better for the environment? Which fuel is better for your engine? Which fuel is better for the economy? Which fuel is better for national security?” If the motorist answers that they do not know the answer for one or more of the questions, the interviewer waits a few seconds to see if the motorist will offer one of the three given responses (ethanol/E85, gasoline/E10, or no difference), but the interviewer can also accept it as a response when the motorists say that they do not know.

The final question of the survey is also a multiple choice question and addresses the energy difference between the fuels. We ask, “Which fuel yields more miles per gallon?” If the motorist answers that either ethanol/E85 or gasoline/E10 yields more miles per gallon (as opposed to ‘no difference’ or ‘don’t know’), we follow-up by asking the motorist to approximate the relative energy difference between the E85 and E10 fuel options. We want to know if the motorists are aware of the energy difference and what they perceive the relative energy difference to be.

3.3 Theoretical Framework, Estimation Technique, and Sample Selection

In this section, we describe the theoretical model that motivates our empirical approach, and we explain how we correct for the selection problem of the endogenously stratified sample.

We develop two versions of the theoretical and the empirical model because in casual conversations with flex motorists we learned that some motorists base their fuel choices on the absolute difference in the two fuel prices (e.g., choose E85 when the price of E85 is $0.50 or more below the price of E10), while others base their fuel choices on the relative difference in prices (e.g., choose E85 when the E85 price divided by the E10 price is 0.8 or lower). Comparison of the two competing models will help determine whether one of the two decision rules dominate.

3.3.1 Model where motorists respond to the absolute difference in prices

The theoretical model is concerned with the fuel choices flex motorists make rather than the quantity of fuel they purchase. In other words, the model focuses on the ‘extensive margin’ rather than the ‘intensive margin’. We let the demand for fuel be perfectly inelastic in the short run so that motorists choose either E85 or E10 based on relative prices, but the amount of fuel they purchase is independent of prices.

The indirect utility that motorist i derives from consumption of fuel $j \in \{e, g\}$ is $V_{ij}(I_i, p_{ij}, \boldsymbol{x}_i, \varepsilon_{ij})$, where e stands for ethanol/E85 and g stands for gasoline/E10 as in Chapter 2, I_i is the motorist’s income, p_{ij} is the nominal (not energy-adjusted) price of fuel j for motorist i, $\boldsymbol{x_i}$ is a vector of characteristics about the motorist and the fueling station, and ε_{ij} is an unobservable demand shifter specific to the motorist and fuel choice. The fuel chosen is the one that yields the greatest indirect utility. For this first version of the model, we introduce the stochastic term additively so that the indirect utility flex motorist i derives from fuel j is

$$V_{ij}(I_i, p_{ij}, \boldsymbol{x}_i, \varepsilon_{ij}) = v_{ij}(I_i, p_{ij}, \boldsymbol{x}_i) + \varepsilon_{ij}.$$

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 prefer the logistic distribution to the normal distribution to model the distribution of WTP for E85 among flex motorists because we know from the literature and from our data that the spread of preferences is broad, and the logistic distribution has more weight on its tails than the normal distribution.

A motorist chooses E85 if $V_{ie}(\cdot) \geq V_{ig}(\cdot)$. We can re-write this decision rule as

$$\varepsilon_i \leq v_{ie}(I_i, p_{ie}, \boldsymbol{x}_i) - v_{ig}\left(I_i, p_{ig}, \boldsymbol{x}_i\right),$$

where $\varepsilon_i \equiv \varepsilon_{ig} - \varepsilon_{ie}$ is symmetric with a mean of zero and follows a logistic distribution. From this, we can write that the probability that a motorist chooses E85 is

$$\Pr(\mathrm{E85}_i) = \Pr\left(\varepsilon_i \leq v_{ie}(\cdot) - v_{ig}(\cdot)\right) = \Lambda\left(v_{ie}(\cdot) - v_{ig}(\cdot)\right), \tag{3.1}$$

where $\Lambda(\cdot)$ is the cdf of the logistic distribution. We choose a linear functional form for $v_{ij}(\cdot)$ whereby

$$v_{ij}\left(I_i, p_{ij}, \boldsymbol{x}_i\right) = \gamma_j I_i + \alpha_j p_{ij} + \boldsymbol{x}_i{}'\boldsymbol{\beta}_j.$$

Substituting the expressions into (3.1) yields

$$\Pr(\mathrm{E85}_i) = \Lambda\left((\gamma_e I_i + \alpha_e p_{ie} + \boldsymbol{x}_i{}'\boldsymbol{\beta}_e) - \left(\gamma_g I_i + \alpha_g p_{ig} + \boldsymbol{x}_i{}'\boldsymbol{\beta}_g\right)\right).$$

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 the same as in the model of Anderson (2012) presented in Chapter 2: motorists do not respond to the fuel prices individually, but rather to the difference in the fuel prices, and if both fuel prices increase or decrease by the same amount, motorists do not switch. Lastly we let $\boldsymbol{\beta} \equiv \boldsymbol{\beta}_e - \boldsymbol{\beta}_g$, which yields

$$\Pr(\mathrm{E85}_i) = \Lambda\left(\alpha\left(p_{ie} - p_{ig}\right) + \boldsymbol{x}_i{}'\boldsymbol{\beta}\right).$$

Finally we define $p_i \equiv p_{ie} - p_{ig}$ to be the difference between the E85 price and the E10 price observed by motorist i. We call this price difference the *E85 premium*. Then for the E85 premium model,

$$\Pr(\mathrm{E85}_i) = \Lambda(\alpha p_i + \boldsymbol{x}_i{}'\boldsymbol{\beta}), \tag{3.2}$$

and

$$\Pr(\mathrm{E10}_i) = 1 - \Lambda(\alpha p_i + \boldsymbol{x}_i{}'\boldsymbol{\beta}). \tag{3.3}$$

Note that unlike Chapter 2 this model does not assume motorists adjust E85 and E10 prices to account for the energy difference, and also note that the E85 premium is negative throughout our sample because in nominal terms the price of E85 was lower than the price of E10 at every station we visited. We provide summary information about the data and details of the estimation sample in Section 3.4.

3.3.2 Model where motorists respond to the relative difference in prices

In the second version of the model, motorists respond to the ratio of the two fuel prices instead of the difference. Unlike the E85 premium, the price ratio captures the cost per mile of E85 relative to E10. The indirect utility flex motorist i derives from fuel j is

$$\tilde{V}_{ij}\left(I_i, p_{ij}, \boldsymbol{x}_i, \varepsilon_{ij}\right) = \tilde{v}_{ij}\left(I_i, p_{ij}, \boldsymbol{x}_i\right) \cdot \exp \varepsilon_{ij},$$

where again $\tilde{v}_{ij}(\cdot)$ is a function of measured variables, and ε_{ij} is a type 1 generalized extreme value random variable whose value is known to the motorist, but unobservable to us.

A motorist chooses E85 if $\tilde{V}_{ie}(\cdot) \geq \tilde{V}_{ig}(\cdot)$. We can re-write this decision rule as,

$$\varepsilon_i \leq \ln \frac{\tilde{v}_{ie}(I_i, p_{ie}, \boldsymbol{x}_i)}{\tilde{v}_{ig}\left(I_i, p_{ig}, \boldsymbol{x}_i\right)}.$$

Again $\varepsilon_i \equiv \varepsilon_{ig} - \varepsilon_{ie}$ follows a logistic distribution. The probability that motorist i chooses E85 is

$$\Pr(\mathrm{E85}_i) = \Pr\left(\varepsilon_i \leq \ln \frac{\tilde{v}_{ie}(I_i, p_{ie}, \boldsymbol{x}_i)}{\tilde{v}_{ig}(I_i, p_{ig}, \boldsymbol{x}_i)}\right) = \Lambda\left(\ln \frac{\tilde{v}_{ie}(I_i, p_{ie}, \boldsymbol{x}_i)}{\tilde{v}_{ig}(I_i, p_{ig}, \boldsymbol{x}_i)}\right). \tag{3.4}$$

We choose a power functional form for $\tilde{v}_{ij}(\cdot)$ whereby

$$\tilde{v}_{ij}(\cdot) = I_i^{\tilde{\gamma}_j} \cdot p_j^{\tilde{\alpha}_j} \cdot \boldsymbol{x}_i^{\tilde{\boldsymbol{\beta}}_j},$$

where, if $\boldsymbol{x}_i$ is $k \times 1$, $\boldsymbol{x}_i^{\tilde{\boldsymbol{\beta}}_j} = x_{i1}^{\tilde{\beta}_{j1}} \cdot x_{i2}^{\tilde{\beta}_{j2}} \cdot \cdots \cdot x_{ik}^{\tilde{\beta}_{jk}}$. Substituting the expressions into (3.4),

$$\Pr(\mathrm{E85}_i) = \Lambda\left(\ln \frac{I_i^{\tilde{\gamma}_e} \cdot p_{ie}^{\tilde{\alpha}_e} \cdot \boldsymbol{x}_i^{\tilde{\boldsymbol{\beta}}_e}}{I_i^{\tilde{\gamma}_g} \cdot p_{ig}^{\tilde{\alpha}_g} \cdot \boldsymbol{x}_i^{\tilde{\boldsymbol{\beta}}_g}}\right). \tag{3.5}$$

Following the same intuition as in the first model, we again assume $\tilde{\gamma}_e = \tilde{\gamma}_g \equiv \tilde{\gamma}$ so income is not a variable that affects motorists' fuel decisions. We assume $\tilde{\alpha}_e = \tilde{\alpha}_g \equiv \tilde{\alpha}$, but this time the implication is that if both fuel prices increase or decrease in a way that maintains the relative price ratio, motorists do not switch. And letting $\tilde{\boldsymbol{\beta}} \equiv \tilde{\boldsymbol{\beta}}_e - \tilde{\boldsymbol{\beta}}_g$ simplifies (3.5) to

$$\Pr(\mathrm{E85}_i) = \Lambda\left(\ln\left(\left(\frac{p_{ie}}{p_{ig}}\right)^{\tilde{\alpha}} \cdot \boldsymbol{x}_i^{\tilde{\boldsymbol{\beta}}}\right)\right) = \Lambda\left(\tilde{\alpha} \ln\left(\frac{p_{ie}}{p_{ig}}\right) + (\ln \boldsymbol{x}_i)'\tilde{\boldsymbol{\beta}}\right). \tag{3.6}$$

In this version of the model, we define $r_i \equiv p_{ie}/p_{ig}$ to be the ratio of the E85 price and the E10 price observed by motorist i. We call this the *E85 ratio*. Then for the E85 ratio model,

$$\Pr(\mathrm{E85}_i) = \Lambda\big(\tilde{\alpha} \ln r_i + (\ln \boldsymbol{x}_i)'\tilde{\boldsymbol{\beta}}\big), \tag{3.7}$$

and,

$$\Pr(\mathrm{E10}_i) = 1 - \Lambda\big(\tilde{\alpha} \ln r_i + (\ln \boldsymbol{x}_i)'\tilde{\boldsymbol{\beta}}\big). \tag{3.8}$$

The probability expressions in (3.7) and (3.8) from the E85 ratio model are analogous to the probability expressions in (3.2) and (3.3) from the E85 premium model except that the ratio model's variables are measured in logs.

3.3.3 Likelihood equation

We estimate the coefficients in equations (3.2) and (3.3) (and also equations (3.7) and (3.8)) by maximum likelihood. To simplify the notation, we let $\boldsymbol{w}_i = (p_i, \boldsymbol{x}_i)$ and $\boldsymbol{\theta} = (\alpha, \boldsymbol{\beta})$ for the price premium model and $\boldsymbol{w}_i = (\ln r_i, \ln \boldsymbol{x}_i)$ and $\boldsymbol{\theta} = (\tilde{\alpha}, \widetilde{\boldsymbol{\beta}})$ for the price ratio model. Then

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

and

$$\Pr(\text{E10}_i) = 1 - \Lambda(\boldsymbol{w}_i{}'\boldsymbol{\theta}).$$

We define the dependent variable for motorist i:

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

The maximum likelihood estimator (MLE) is based on the likelihood formed from the joint distribution of the data $(\boldsymbol{y}, \boldsymbol{W})$. Under standard assumptions, the components of $\boldsymbol{W}$ are exogenous with respect to $\boldsymbol{y}$ 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

$$\ln \text{L} = \sum_{i=1}^{n} \ln f(y_i|\boldsymbol{w}_i, \boldsymbol{\theta}) = \sum_{i=1}^{n} \{(1 - y_i) \ln[1 - \Lambda(\boldsymbol{w}_i{}'\boldsymbol{\theta})] + y_i \ln[\Lambda(\boldsymbol{w}_i{}'\boldsymbol{\theta})]\}. \tag{3.9}$$

However, up until now we have ignored the sample selection problem. We sample from flex motorists who fuel at E85 stations rather than from the general population of flex motorists. Thus parameter estimates that maximize (3.9) will be biased to mimic the endogenously stratified sample and not the population.

3.3.4 Endogenous stratification

We only survey flex motorists who fuel at E85 stations. 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 because most fuel stations do not offer E85. Some motorists in our sample are representative of the population of flex motorists because their patronage at a fuel station where we survey is not caused by the station's offering of E85. However, some of the flex motorists in our sample who choose E85 drive out of their way for it, and if every station offered E85, they would choose a different station and not appear in our sample. The implication is that we oversample flex motorists who choose E85, but we can correct the bias by modeling how the oversampling occurs. We have an endogenously stratified (choice-based) sample.

Recall that we ask each motorist who chooses E85 whether they came specifically for the E85 and, if so, how far out of their way they drove for it. Some motorists indicate that they did not drive out of their way at all. In these cases, either the motorists do not know the station carries E85 until they arrive, or the station is simply their usual station because it is the closest one to their home or work for example. Sometimes the motorist had been fueling at the station regularly since before owning an FFV or before the station started selling E85. Alternatively, other motorists indicate that they did indeed drive out of their way to come to the E85 station. Some motorists drive an extra few blocks and others a few miles or more out of their way, past E10 stations, just to fuel with E85. We use these responses to inform the selection problem.

We assume that all motorists who choose E10 and the motorists who choose E85 but did not drive out of their way for E85 are representative of the general population of flex motorists. We assume these motorists would have fueled at the station even if every retail station in the area offered E85. By removing the motorists who drive out of their way for E85 we have a representative sample of flex motorists. Thus one way to correct our estimates so that they mimic the general population of flex motorists is to remove the oversampled observations of flex motorists who selected themselves into our sample by driving out of their way to fuel with E85.

A second way to correct our estimates without discarding observations is the weighted maximum likelihood estimator (WMLE) proposed by Manski and Lerman (1977). With WMLE, observations where the motorists choose E85 get less weight in the log-likelihood function, and observations where the motorists choose E10 get more weight. The estimator requires that the population proportions of E85-users and E10-users are known and puts inverse probability weights on each observation in the likelihood function. Thus the coefficient estimates from WMLE will be similar to the coefficient estimates from the conditional probability model using the representative sample but WMLE uses all of the observations and offers greater precision.

The derivation of WMLE in this section is based on Cameron and Trivedi (2005 p. 826). The population is divided into two strata. The first stratum is the subset of the population who choose E85, which is defined as C_e. The second stratum is the subset of the population who choose E10, which is defined as C_g. An important distinction must be made between the population probability of a motorist choosing fuel $j \in \{e, g\}$ and the probability of sampling from the subset of motorists who choose fuel j, because the two are different in an endogenously stratified sample. Define H_j as the probability of observing a motorist from subset C_j in the sample of motorists at E85 stations, and define $Q_j(\boldsymbol{\theta})$ as the actual population probability that a flex motorist chooses fuel j. H_j is the observed share of motorists in our sample who choose fuel j, and

$$Q_j(\boldsymbol{\theta}) = \int_{C_j} f(y|\boldsymbol{w},\boldsymbol{\theta})h(\boldsymbol{w})\,dyd\boldsymbol{w},$$

where $h(\boldsymbol{w})$ is the marginal distribution of $\boldsymbol{w}$. We write the joint density of y and $\boldsymbol{w}$ as $g(y,\boldsymbol{w}|\boldsymbol{\theta}) = f(y|\boldsymbol{w},\boldsymbol{\theta}) \cdot h(\boldsymbol{w})$.

To see the problem of endogenous stratification, we begin by obtaining the joint densities over y, $\boldsymbol{w}$, and j, where j is an indicator for the fuel choice stratum from which the observation was obtained. We write the joint densities as the product of the conditional and marginal densities. The population joint density is

$$g(y,\boldsymbol{w},j|\boldsymbol{\theta}) = g(y,\boldsymbol{w}|j,\boldsymbol{\theta}) \cdot Q_j(\boldsymbol{\theta}).$$

Because $g(y,\boldsymbol{w}|j,\boldsymbol{\theta})$ equals the density $g(y,\boldsymbol{w}|\boldsymbol{\theta})$ divided by the population probability of being in fuel choice strata j (so that the density integrates over C_j to one), we can write

$$g(y,\boldsymbol{w}|j,\boldsymbol{\theta}) = \frac{f(y|\boldsymbol{w},\boldsymbol{\theta})h(\boldsymbol{w})}{Q_j(\boldsymbol{\theta})}. \tag{3.10}$$

The joint density for the sample (superscripted s) is

$$g^s(y,\boldsymbol{w},j|\boldsymbol{\theta}) = g(y,\boldsymbol{w}|j,\boldsymbol{\theta}) \cdot H_j. \tag{3.11}$$

Then combining (3.10) and (3.11), we write the joint density for the sample as

$$g^s(y,\boldsymbol{w},j|\boldsymbol{\theta}) = \frac{H_j}{Q_j(\boldsymbol{\theta})} f(y|\boldsymbol{w},\boldsymbol{\theta})h(\boldsymbol{w}). \tag{3.12}$$

Thus the conditional MLE based on the population conditional density $f(y|\boldsymbol{w},\boldsymbol{\theta})$ in (3.9) will be inconsistent for $\boldsymbol{\theta}$ because the estimator ignores the relative sample and population weights.

The WMLE maximizes

$$Q_{WMLE}(\boldsymbol{\theta}) = \sum_i \frac{Q_i}{H_i} \ln f(y_i|\boldsymbol{w}_i,\boldsymbol{\theta}), \tag{3.13}$$

where $Q_i = Q_e$ and $H_i = H_e$ if motorist i chooses E85 and $Q_i = Q_g$ and $H_i = H_g$ if motorist i chooses E10. The estimator multiplies each term from the conditional log-likelihood estimator $\ln f(y_i|\boldsymbol{w}_i, \boldsymbol{\theta})$ by Q_i/H_i giving less weight to all of the E85 observations (whether they drove out of their way or not) and more weight to all of the E10 observations.

The objective function in (3.13) is not formally a likelihood, but we show that the WMLE is consistent following Cameron and Trivedi (2005 p. 828). The WMLE solves the first-order conditions

$$\sum_i \frac{Q_i}{H_i} \frac{\partial \ln f(y_i|\boldsymbol{w}_i, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0}. \tag{3.14}$$

For the estimator to be consistent, the terms in (3.14) must have zero expected value where expectation is with respect to the sampling density $g^s(y, \boldsymbol{w}, j|\boldsymbol{\theta})$ in (3.12). To show this, we first write that

$$\mathrm{E}^s\left[\frac{Q_j}{H_j} \frac{\partial \ln f(y|\boldsymbol{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = \iint \frac{Q_j}{H_j} \frac{\partial \ln f(y|\boldsymbol{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} g^s(y, \boldsymbol{w}, j|\boldsymbol{\theta}) dy d\boldsymbol{w}.$$

After a few manipulations, we can write that

$$\mathrm{E}^s\left[\frac{Q_j}{H_j} \frac{\partial \ln f(y|\boldsymbol{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] = \int \mathrm{E}\left[\frac{\partial \ln f(y|\boldsymbol{w}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right] h(\boldsymbol{w}) d\boldsymbol{w}.$$

Under the usual regularity condition, in the population the specified density satisfies $\mathrm{E}[\partial \ln f(y|\boldsymbol{w}, \boldsymbol{\theta}) / \partial \boldsymbol{\theta}] = \mathbf{0}$, so the WMLE is consistent in the presence of endogenous stratification.

For our application, the population proportion of flex motorists who choose E85 is Q_e and the population proportion who choose E10 is Q_g. Population proportions are calculated for each region by removing the observations of E85 users who drove out of their way to the station. The observed sample proportion of flex motorists who choose E85 is H_e in the region where motorist i fuels, and the sample proportion who choose E10 is H_g. To invoke the WMLE, we apply the probability weights to the log-likelihood function so the expression we maximize to estimate $\boldsymbol{\theta}$ is:

$$\sum_{i=1}^{n} \frac{Q_i}{H_i} \{(1 - y_i) \ln[1 - \Lambda(\boldsymbol{w}_i{}'\boldsymbol{\theta})] + y_i \ln[\Lambda(\boldsymbol{w}_i{}'\boldsymbol{\theta})]\} = \sum_{i=1}^{n} \frac{Q_i}{H_i} \ln f(y_i|\boldsymbol{w}_i, \boldsymbol{\theta}). \quad (3.15)$$

Observe that (3.15) is identical to (3.9) if all probability weights equal 1.

3.4 Data Collection and Summary Statistics

This section discusses how, when, and where the intercept survey was conducted. Summary information for the observed data and survey responses are also provided.

3.4.1 Data collection

We obtained the cooperation of two E85 retailers to conduct our survey. Institutional Review Board (IRB) approval of the survey is provided as Appendix D. We collected a total of 972 observations from 17 E85 stations in 6 urban areas between October 2014 and April 2015. 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 spent the most time at four stations around DM, where we observed significant variation in both nominal and relative fuel prices. Next, we spent one week in each of CS, TS, LR, SAC, and LA. In each location, we visited two or three different stations and collected around 100 observations. Unfortunately, we observed almost no variation in fuel prices within an urban area during the single week we were there. All of the stations we visited in DM, CS, TS, and LR were operated by a retailer we will call ‘Retailer A’, and all of the stations we visited in SAC and LA were operated by a retailer we will call ‘Retailer B’.

Retailer A is an independent fuel retailer who offers E85 in several locations. Similar to how most retail fuel stations have an island or two at one end with an extra nozzle offering diesel, Retailer A’s E85 stations also have an island or two with an extra nozzle offering E85. In almost every case, Retailer A displays prices for its E10 and E85 fuels together prominently on the stations’ main street signs and elsewhere. The E85 stations we visit are in medium-sized urban areas. Each area has several E85 stations no further than ten or fifteen minutes away from one another. The share of stations that offer E85 in these urban areas is relatively high compared to the rest of the country, and each E85 station serves a moderate-sized market of FFVs.

Retailer B’s business model focuses on biofuels by adding special pumps to existing fuel stations. This means stations branded under different names have an E85 pump off on the side owned by Retailer B. Prices for E10 and E85 are likewise displayed prominently. The main difference for Retailer B is that there are far fewer E85 stations per flex motorist in California, and it is not because the share of the vehicles that are FFVs is higher: it is because the share of stations offering E85 is much lower. This means that flex motorists who want to fuel with E85 have relatively little choice of E85 stations and must come from considerable distances.

3.4.2 Observed data and survey responses

From the initial 972 observations of motorists fueling their FFVs, we remove 79 observations where the motorist chose not to/was unable to complete/participate in the survey, a total non-response rate of about 8 percent. That leaves us with an estimation sample of 893 observations. Table 3.1 summarizes the fuel choice data broken down by station, region and retailer. In the entire sample of 893 flex motorists, the average E85 price is \$2.19 per gallon, and the average E10 price is \$2.58. The average E85 premium is −\$0.39 and the average E85 ratio is 0.85. Nevertheless throughout our sample 436 (49 percent) of the flex motorists chose E85, while 457 (51 percent) chose E10.[8] Fuel prices and the share of flex motorists who chose E85 varies considerably across the different areas we visited.

The average fuel prices were most favorable towards E85 at Retailer B's California locations around LA and SAC, where on average the E85 premium was −\$0.54, and the E85 ratio was 0.83. We observed 231 flex motorists fueling at Retailer B's locations, and 89 percent chose E85. Sometimes a 2-, 3-, or even 4-car line formed for the E85 pump while E10 pumps were vacant. 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 more uncommon in California. Also, Retailer B ran promotions providing special fuel cards and other incentives to local residents, 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 about 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.

[8] Among the 457 flex motorists who chose E10, 421 (92 percent) chose regular grade 87 octane (85 octane in CS), 24 (5 percent) chose midgrade, and 12 (3 percent) chose premium.

Even though the average E85 price ratios observed were the lowest in DM, LA, and SAC, the average E85 ratio was still not low enough to favor E85 on a cost-per-mile basis. Recall that E85 has approximately 75 to 80 percent of the energy per volume as E10, so the E85 ratio has to be under 0.80 for E85 to be 'in the money'. Altogether 180 observations (from DM and SAC) were where the E85 ratio was less than 0.80, and for 52 of those observations in DM, the E85 ratio was less than 0.75. Following DM, the average E85 ratio was slightly higher in LR at 0.84, higher still in TS at 0.87, and finally highest in CS where the E85 ratio was 0.98. In fact, for some station visits in CS, the E85 and E10 prices were identical.

Figure 3.1 plots the average E85 premiums and the shares of motorists who chose E85 at each of the 17 E85 stations we visited, and Figure 3.2 plots the same station shares but the relative prices are the stations' average E85 ratios. The figures are similar and show a notable shift between Retailer A and Retailer B. The downward-sloping demand 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 sections.

As described in Section 3.2, we record other observable characteristics about motorists in addition to their fuel choices before approaching with the intercept survey. Recall that motorists' characteristics recorded are: the vehicle make, model, and 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 sex of the motorist. Table 3.2 contains summary statistics for these data. We ended up not using the vehicle license plate data in our regressions and instead we use dummy variables for each of the regions where we survey. There was little variation in license plate states within a region so there is near collinearity with the regional dummy variables.

Table 3.1 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	1.999	1.999	0.000	1.000	9.1%
Co. Springs 2	33	1.999	2.023	-0.024	0.988	30.3%
Co. Springs 3	54	1.999	2.059	-0.060	0.971	13.0%
CS total	**98**	**1.999**	**2.040**	**-0.041**	**0.980**	**18.4%**
Des Moines 1	117	2.158	2.721	-0.563	0.793	46.2%
Des Moines 2	61	2.277	2.690	-0.413	0.846	31.1%
Des Moines 3	28	2.313	2.814	-0.501	0.822	50.0%
Des Moines 4	114	2.294	2.687	-0.392	0.854	40.4%
DM total	**320**	**2.243**	**2.711**	**-0.468**	**0.827**	**41.6%**
Little Rock 1	26	1.838	2.182	-0.344	0.842	34.6%
Little Rock 2	23	1.829	2.129	-0.300	0.859	34.8%
Little Rock 3	60	1.829	2.179	-0.350	0.839	31.7%
LR total	**109**	**1.831**	**2.169**	**-0.338**	**0.844**	**33.0%**
Tulsa 1	58	1.799	2.092	-0.293	0.860	41.4%
Tulsa 2	12	1.799	2.099	-0.300	0.857	66.7%
Tulsa 3	65	1.799	2.040	-0.241	0.882	18.5%
TS Total	**135**	**1.799**	**2.068**	**-0.269**	**0.870**	**32.6%**
Retailer A total	**662**	**2.048**	**2.391**	**-0.343**	**0.857**	**34.9%**
Los Angeles 1	85	2.614	3.204	-0.590	0.816	95.3%
Los Angeles 2	52	2.630	3.099	-0.469	0.849	84.6%
LA total	**137**	**2.620**	**3.164**	**-0.544**	**0.828**	**91.2%**
Sacramento 1	43	2.566	3.229	-0.663	0.795	81.4%
Sacramento 2	51	2.485	2.921	-0.436	0.851	88.2%
SAC total	**94**	**2.522**	**3.062**	**-0.540**	**0.824**	**85.1%**
Retailer B total	**231**	**2.580**	**3.123**	**-0.542**	**0.826**	**88.7%**
Sample total	**893**	**2.186**	**2.580**	**-0.394**	**0.847**	**48.8%**

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.

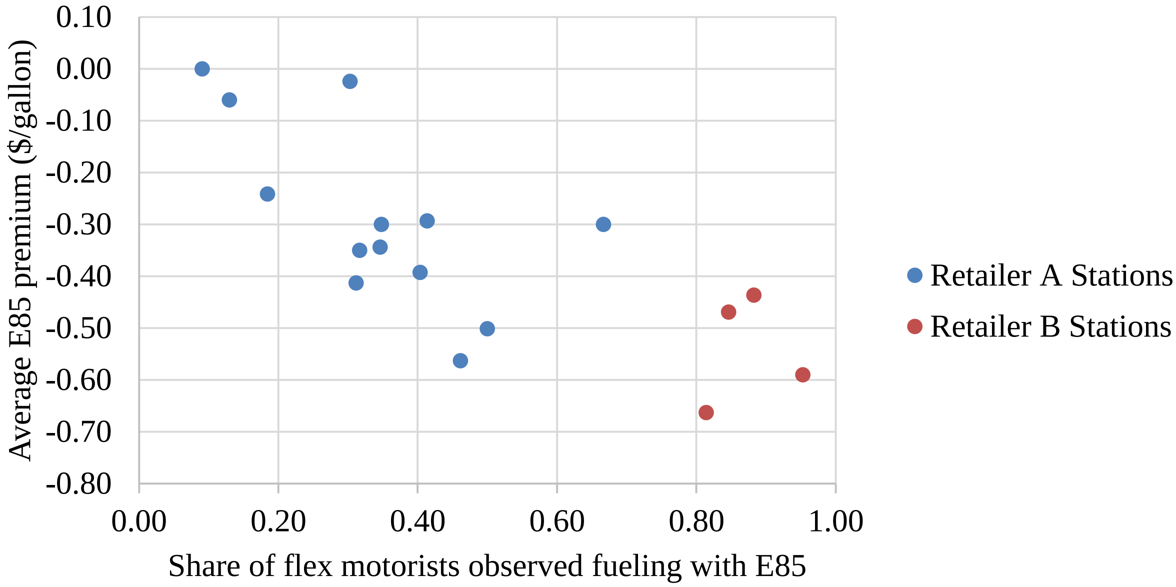

Figure 3.1 Observed shares of flex motorists fueling with E85 and E85 price premium by station and retailer

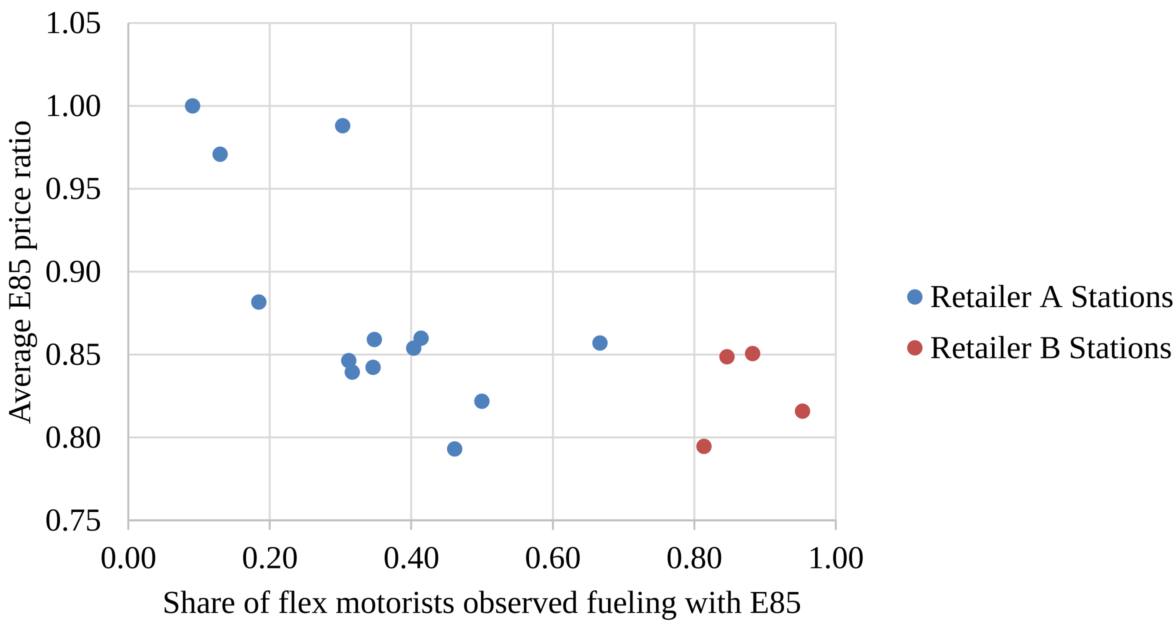

Figure 3.2 Observed shares of flex motorists fueling with E85 and E85 price ratio by station and retailer

Data are from 893 interviews of flex motorists fueling at 17 E85 stations operated by two different retailers. Prices are not energy-adjusted for their relative energy content.

For the vehicle make, the largest share of the vehicles in the sample were Chevrolet, at 46 percent. The most common Chevrolet models were the Silverado, Impala, Tahoe, Suburban, HHR, Equinox and Malibu. The next most common vehicle make was Ford with 18 percent of the sample and common models F150, Explorer, Focus, Fusion, and Taurus. Third in our sample was Dodge with 14 percent and common models Grand Caravan, Ram, and Durango. GMC and Chrysler were fourth and fifth, making up 7 percent of our sample each, and the final 8 percent of the sample represented all of the other vehicle makes.

As for vehicle type, trucks and SUVs each made up about 30 percent of our sample, cars were 25 percent, and vans were the remainder. We were surprised that our sample contained about twice as many men as women. Our initial expectation was that the population of flex motorists would be about half men and half women. It is possible that the types of vehicles that tend to be FFVs (large American-made cars and trucks) are more often driven by men than women. Lastly about 67 percent of the FFVs in our sample had FFV badges, and about 94 percent had some sort of yellow E85 indicator inside the gas door. The noteworthy exceptions are the flexible-fuel Toyotas (Tundra and Sequoia) and Nissans (Titan and Armada), which have badges on the backs, but no yellow gas caps. Other makes and models were also missing the yellow cap/ring/sticker on rare occasions.

Table 3.2 shows that of the 893 flex motorists who completed our survey, 739 (83 percent) responded that they were fueling their personal FFV. Another 80 motorists (9 percent) were fueling company FFVs, 27 (3 percent) were fueling government FFVs, and the remaining 47 motorists (5 percent) were fueling other non-personal vehicles like rentals or FFVs that belonged to friends or family.

Table 3.2 Summary of characteristics of motorists in the sample

Vehicle make	Chevrolet	45.8%
	Ford	18.1%
	Dodge	13.9%
	GMC	7.3%
	Chrysler	7.1%
	Other	7.8%
Vehicle type	Truck	30.5%
	SUV	29.7%
	Car	25.8%
	Van	14.1%
Motorist gender	Male	66.2%
	Female	33.8%
FFV badge	Yes	67.0%
	No	33.0%
Yellow cap/sticker	Yes	94.4%
	No	5.6%
Vehicle ownership (stated)	Personal	82.8%
	Company	9.0%
	Government	3.0%
	Other	5.3%
Age (stated)	Min	18
	1st Qu.	33
	Median	42
	Mean	44.01
	3rd Qu.	54
	Max	88
Miles per year (stated)	Min	500
	1st Qu.	12,000
	Median	17,000
	Mean	21,710
	3rd Qu.	27,000
	Max	120,000

Summary statistics are for 893 observations of flex motorists fueling at E85 stations in the areas of Colorado Springs, Des Moines, Los Angeles, Little Rock, Sacramento, and Tulsa. Vehicle type 'Car' includes coupes, convertibles, sedans, hatchbacks, and station wagons.

In the sample, the range of ages span 18 to 88, and the median age is 42. In some cases, motorists declined to give their age. In these cases the interviewer would move on, and write in an estimate after the interview was completed. However, we decided to exclude these observations from the sample along with the other incomplete observations. Similarly, on rare occasions motorists were unable to answer the question about how intensively they used their vehicle. In most cases, motorists were able to offer an approximation of how many miles they drove per year or per month or per week. Sometimes the motorists would check the odometer and say something like, “Well I’ve driven [odometer reading] miles in [number of years of car ownership] years.” Most of the cases where the motorist was unable to answer was when they were not driving their personal vehicle and were unsure how to respond. Again we excluded these incomplete observations from the sample.

Next, to the motorists who chose E10, we asked questions to measure their knowledge and awareness of E85. The results are in Table 3.3. Of the 457 flex motorists in our sample who fueled with E10, 392 (86 percent) indicated that they were aware that their vehicle was in fact a flexible-fuel vehicle capable of using E85. Of the 392 E10 users who were aware of their vehicles’ capabilities, 148 (38 percent) responded that they had fueled with E85 at least once, while the majority had never tried it. This might be explained by E85 having been historically more expensive than E10 in energy-equivalent terms. Finally, 287 of the 392 responded that they were aware that the station sold E85, and of the remaining 105 who answered they did not know, 80 previously responded that they had never used E85. In general, these are motorists who happen to own FFVs, but know almost nothing about E85. They do not know what it is, they have never used it, and they certainly do not think to look for it.

Table 3.3 Responses to questions to flex motorists who fuel with E10

	Yes	No / Don't know	Total
Is your vehicle an FFV?	392	65	457
Have you ever fueled with E85?	148	244	392
Did you know this station sells E85?	287	105	392

Of the 893 flex motorists in our sample, 457 chose to fuel with one of the E10 blends. We wanted to see if they were responding to the relative fuel prices. To the motorists who responded that their vehicle was an FFV, we asked the follow-up questions shown. Between the 65 motorists who did not know their vehicle was an FFV and the 105 motorists who did not know the station sold E85, there were 170 motorists in our sample who we assume would not have chosen E85 regardless of the relative prices. Only 148/457 (32%) of the flex motorists in our sample who fueled with E10 had ever fueled with E85.

Out of the 457 flex motorists in our sample who chose E10, 65 did not know they were fueling an FFV capable of using E85, and another 105 were not aware that the station sold E85. The implication is that these 170 motorists would not have chosen E85 no matter how low the relative price of E85 would have been. These motorists represent a segment of the population of flex motorists who were not aware of the station’s or the vehicle’s capabilities, though they were not necessarily unwilling to use E85 in the future.

We asked the motorists who chose E85 whether they chose to fuel at the station because of E85, and, if so, how far out of their way they drove. Summary data for these questions is shown in Table 3.4. Out of the 436 flex motorists who chose E85, 407 (93 percent) said that they chose to fuel at the station because it offered E85. And out of those 407 motorists, 195 (48 percent) said that they did not drive out of their way at all. It seems that most motorists drive past a number of fuel stations in their normal routine, and while they may choose to fuel at a particular station due to the station’s unique amenities (e.g., whether it offers E85), most motorists do not consider the station they choose to be ‘out of their way’.[9] We use the responses

[9] In retrospect, a better way to ask this question may have been something along the lines of, “If every gas station in the area offered E85, would you still have chosen to fuel at this station?”

to how far motorists drove to inform about how the general population of flex motorists differs from our sample population. Specifically, we assume that the remaining 212 observations of flex motorists who chose E85 and drove out of their way for it are oversampled. We construct our estimates of the population shares by removing those 212 observations from our sample. The sample and inferred population shares are given in Table 3.5.

Table 3.4 Responses to questions to flex motorists who fuel with E85

Did you choose to fuel at this station because it offers E85?	Yes	407
	No	29
	Total	**436**
How far out of your way did you drive? (miles)	Not at all (zero mi.)	195
	1 mile or less	44
	(1,3] miles	73
	(3,5] miles	42
	(5,10] miles	38
	More than 10 miles	15
	Total	**407**

Statistics are for the 436 observations of flex motorists in our sample who chose to fuel with E85. In total, 407/436 (93%) said they came for the E85, but of those 407 motorists, 195 said that they did not drive out of their way at all. We remove the remaining 212 motorists who drove out of their way for E85 from the sample to calculate population shares.

Responses to the questions we have discussed to this point do not differ significantly by urban area. The measurable differences in the data across regions are in the fuel prices and observed choices, as shown in Table 3.1, but also in the fuel opinion questions shown in Table 3.6 and Table 3.7. Table 3.6 shows the responses to the fuel opinion questions by region from only the 457 motorists who fueled with E10, and Table 3.7 shows the responses from only the 436 motorists who chose E85. In the three questions about which fuel is better for the environment, the engine, and the economy, the differences in opinions across regions are especially apparent. In general, a greater share of flex motorists we surveyed in DM, LA, and

SAC believe that ethanol is better for the environment, for a vehicle's engine, and for the economy, while the average motorist in LR and TS has a much less favorable opinion of ethanol in these same areas, and the average motorist in CS is somewhere in between.

Table 3.5 Sample and population shares of flex motorists who fuel with E85 by region

	Observed sample data			Inferred population subset data		
Region	E85	Total	Share	E85	Total	Share
CS	18	98	18.4%	12	92	13.0%
DM	133	320	41.6%	67	254	26.4%
LA	125	137	91.2%	54	66	81.8%
LR	36	109	33.0%	20	93	21.5%
SAC	80	94	85.1%	48	62	77.4%
TS	44	135	32.6%	23	114	20.2%
Total	**436**	**893**	**48.8%**	**224**	**681**	**32.9%**

The table shows the share of flex motorists who we observed fueling with E85 in each region. The regions are Colorado Springs (CS), Des Moines (DM), Los Angeles (LA), Little Rock (LR), Sacramento (SAC), and Tulsa (TS). In Models 3.2 and 3.5, to generate a subset of the sample data that better represents the population of flex motorists, we remove any observations from motorists who drove out of their way for the E85, leaving 681 of 893 observations. In Models 3.3 and 3.6, we use the inferred population shares to calculate the probability weights.

Table 3.6 Responses to fuel opinion questions by region from flex motorists who fuel with E10

	Region	Observations	Ethanol	Gasoline	No difference	Don't know
Which fuel is better for the environment?	CS	80	64%	5%	13%	19%
	DM	187	78%	6%	11%	6%
	LA	12	50%	25%	8%	17%
	LR	73	52%	18%	14%	16%
	SAC	14	64%	7%	7%	21%
	TS	91	42%	19%	25%	14%
	Total	**457**	**63%**	**11%**	**14%**	**12%**
Which fuel is better for your engine?	CS	80	24%	48%	15%	14%
	DM	187	25%	42%	19%	14%
	LA	12	0%	42%	25%	33%
	LR	73	18%	52%	14%	16%
	SAC	14	14%	43%	7%	36%
	TS	91	13%	69%	9%	9%
	Total	**457**	**20%**	**50%**	**15%**	**14%**
Which fuel is better for the economy?	CS	80	44%	30%	11%	15%
	DM	187	71%	11%	10%	7%
	LA	12	25%	50%	0%	25%
	LR	73	33%	45%	4%	18%
	SAC	14	43%	14%	14%	29%
	TS	91	25%	44%	15%	15%
	Total	**457**	**49%**	**28%**	**10%**	**13%**
Which fuel is better for national security?	CS	80	34%	31%	24%	11%
	DM	187	51%	12%	12%	25%
	LA	12	25%	33%	0%	42%
	LR	73	32%	29%	12%	27%
	SAC	14	21%	21%	14%	43%
	TS	91	21%	29%	30%	21%
	Total	**457**	**37%**	**22%**	**17%**	**23%**
Which fuel yields more miles per gallon?	CS	80	16%	60%	5%	19%
	DM	187	8%	76%	4%	12%
	LA	12	8%	50%	0%	42%
	LR	73	12%	66%	7%	15%
	SAC	14	7%	50%	0%	43%
	TS	91	10%	67%	3%	20%
	Total	**457**	**11%**	**68%**	**4%**	**17%**

Summary statistics are for survey data collected from 457 flex motorists who fueled with E10 at the E85 stations we visited in Colorado Springs (CS), Des Moines (DM), Los Angeles (LA), Little Rock (LR), Sacramento (SAC), and Tulsa (TS).

We separate the responses by fuel choice so we can compare the opinions across regions separately from fuel choices across regions. Table 3.6 shows that even among only the motorists who choose E10, motorists have a much higher opinion of ethanol in DM than they do elsewhere when it comes to the environment and the economy, and the other factors. In DM, 78 percent of E10-using flex motorists responded that ethanol was better for the environment, and 71 percent responded that ethanol was better for the economy. On the other hand, in TS, 42 percent of E10-using flex motorists responded that ethanol was better for the environment, and 25 percent responded that ethanol was better for the economy. Note that of the 231 observations we collected at Retailer B's LA and SAC locations, only 26 chose E10. Also note that we collected fuel opinion data for all of the flex motorists in our sample, even those who did not know they had an FFV or did not know anything about ethanol or E85.

Table 3.7 likewise shows that even among only the motorists in our sample who chose E85, average opinions of ethanol are much higher in DM, LA, and SAC than in CS, LR, and TS. At the extremes are DM and TS. Among the flex motorists who chose to fuel with E85, 84 percent in DM responded that ethanol was better for the environment, compared to 43 percent in Tulsa. And 87 percent of E85-using DM motorists responded that ethanol was better for the economy compared to 55 percent in TS. We model the opinions as explanatory variables in our empirical model. The opinions are especially informative when we compare the DM data with the data from LR and TS. Retailer A operated all the stations in these regions and the E85/E10 price ratios were quite similar. By contrast, opinions of ethanol were drastically different.

Table 3.7 Responses to fuel opinion questions by region from flex motorists who fuel with E85

	Region	Observations	Ethanol	Gasoline	No difference	Don't know
Which fuel is better for the environment?	CS	18	67%	6%	11%	17%
	DM	133	84%	1%	11%	5%
	LA	125	82%	0%	6%	13%
	LR	36	67%	6%	19%	8%
	SAC	80	85%	0%	9%	6%
	TS	44	43%	9%	30%	18%
	Total	**436**	**77%**	**2%**	**11%**	**9%**
Which fuel is better for your engine?	CS	18	39%	33%	6%	22%
	DM	133	43%	30%	17%	10%
	LA	125	64%	8%	10%	18%
	LR	36	44%	28%	17%	11%
	SAC	80	55%	14%	21%	10%
	TS	44	32%	34%	16%	18%
	Total	**436**	**50%**	**21%**	**15%**	**14%**
Which fuel is better for the economy?	CS	18	44%	33%	17%	6%
	DM	133	87%	4%	5%	5%
	LA	125	78%	8%	6%	8%
	LR	36	56%	19%	17%	8%
	SAC	80	66%	18%	10%	6%
	TS	44	55%	20%	16%	9%
	Total	**436**	**73%**	**12%**	**9%**	**7%**
Which fuel is better for national security?	CS	18	28%	11%	28%	33%
	DM	133	69%	7%	8%	17%
	LA	125	45%	10%	6%	39%
	LR	36	44%	28%	11%	17%
	SAC	80	44%	6%	15%	35%
	TS	44	30%	23%	20%	27%
	Total	**436**	**50%**	**11%**	**11%**	**28%**
Which fuel yields more miles per gallon?	CS	18	17%	61%	6%	17%
	DM	133	19%	67%	5%	10%
	LA	125	30%	40%	14%	16%
	LR	36	28%	56%	8%	8%
	SAC	80	15%	55%	18%	13%
	TS	44	23%	52%	5%	20%
	Total	**436**	**22%**	**54%**	**10%**	**13%**

Summary statistics are for survey data collected from 436 flex motorists who fueled with E85 at the E85 stations we visited in Colorado Springs (CS), Des Moines (DM), Los Angeles (LA), Little Rock (LR), Sacramento (SAC), and Tulsa (TS).

For some of the flex motorists we surveyed, the question about national security elicited more confusion rather than an actual response. In 2006, national security and independence from foreign oil were touted as reasons to support the biofuels mandates, but the cause seems to have lost importance with flex motorists in 2015. As with the other questions, motorists in DM, LA, and SAC favor ethanol more than the motorists in the other urban areas, but there were also many more cases where the motorists answered, “No difference” or, “Don’t know”.

The last question of the survey asked which fuel yields more miles per gallon. In DM, about 67 percent of the flex motorists correctly answered E10 yielded more miles per gallon than E85. About 19 percent said that E85 yielded more miles per gallon than E10, 5 percent said there was no difference, and 10 percent answered that they did not know. In other regions, the percentage of motorists who correctly identify that E10 yielded more miles per gallon was even lower. In CS, 61 percent answered correctly, and in LR, SAC, and TS, 56 percent, 55 percent, and 52 percent of motorists respectively correctly answered. Finally in LA, just 40 percent of the flex motorists we surveyed responded that E10 yields more miles per gallon, 30 percent said E85 was better, 14 percent said there was no difference, and 16 percent answered that they did not know. Ignorance about the energy difference of the two fuels likely explains why some motorists drive miles out of their way or wait in line to fuel with E85. We also asked the motorists a follow up question to approximate the percentage the relative energy difference between the two fuels. Some motorists responded with an accurate answer saying that E85 gets about 75-80 percent of the miles per gallon of E10. Some approximated higher energy for E85 in the 90 percent range and some approximated the energy ratio to be as low as 50 percent. Responses were not always in the form of a simple percentage of energy content, but rather some motorists knew the miles per gallon of each, “I get 14 mpg with E85 and 18 mpg with E10,” and others knew how long a tank of each of the two fuels lasted.

Interestingly, many of the flex motorists who chose E85 demonstrated that they understood that E85 was more expensive on an energy-equivalent basis. Some chose E85 for reasons other than the price, while others simply did not bother to calculate the energy-equivalent fuel costs every time they filled up. Many flex motorists said something along the lines of, “I did the math once and figured that I need a $0.60 per gallon discount on E85 for it to be worth it,” and now they make their fuel choice based on some rule-of-thumb or routine.

3.5 Empirical Models and Results

In this section we describe the empirical models, the method for calculating marginal effects and their standard errors, and the results and implications of the models. Lastly we compare how well the models fit the data.

3.5.1 Empirical models

We estimate three versions of the E85 premium model. In Model 3.1, we do not perform any correction for endogenous stratification. We maximize the WMLE expression in (3.15) using our entire sample of 893 observations setting all the weights equal to one. This is equivalent to maximizing the log-likelihood in (3.9). The estimates from Model 3.1 describe the population of flex motorists who fuel at E85 stations rather than the general population of flex motorists. In Model 3.2, we use the same unweighted estimating equation, but we correct for the endogenous stratification by using only the 681 observations from motorists who did not drive out of their way for the E85. In Model 3.3, we use all 893 observations correcting for the endogenous stratification by applying the probability weights from the inferred population shares. Estimates from Model 3.3 will be similar to and more precise than estimates from Model 3.2. We also estimate three analogous versions of the E85 ratio model: in Model 3.4, we do not correct for the endogenous stratification, in Model 3.5, we use only the representative subset, and in Model 3.6, we use the probability weights.

Each of the six models use the following motorist and station characteristics as explanatory variables: vehicle ownership (personal, government, company, other), vehicle type (car, truck, SUV, van), whether the vehicle has an FFV badge, the gender of the motorist, the age of the motorist, how many miles per year the motorist drives, the motorist's opinions about which fuel is better for the environment, the engine, the economy, national security, the motorist's opinion about which fuel yields more miles per gallon, and the region where the station is located (CS, DM, LA, LR, SAC, TS). Variables that describe the characteristics of the fuel stations are not added to the model but the fixed effects and location dummies summarize the information for the station and region.

Models 3.1 to 3.6 are informative for identifying the various drivers of E85 demand. However we also want to learn whether preferences differ across regions when we do not control for motorists' opinions and characteristics. If we find that the regional effects are not significant, then we can conclude that estimates of E85 preferences from one state inform E85 preferences from other states. Thus we estimate Models 3.7 to 3.12 that include only price variables and station location dummies.

3.5.2 Results

We focus the discussion of the results on marginal effects. The full set of coefficient estimates, marginal effects, standard errors, and p-values for each model estimated are available in Appendix B. Robust standard errors for the coefficients are calculated using a sandwich estimator as described in Cameron and Trivedi (2005 p. 828). Tables 3.8 to 3.11 show the average marginal effects calculated as $\frac{1}{N}\sum_{i=1}^{N}\lambda\left(\boldsymbol{w}_i{}'\widehat{\boldsymbol{\theta}}\right)\widehat{\boldsymbol{\theta}}$ and standard errors calculated using the delta-method for Models 3.1 to 3.12. The variables in bold in the results tables are significant at the 5 percent level. We first discuss Models 3.1 to 3.6 in Tables 3.8 and 3.9.

In all 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 price premium in Models 3.1 to 3.3 are −0.267, −0.239, and −0.235 respectively. Corrections for the endogenous sampling reduce the magnitude of the E85 premium coefficient. Models 3.2 and 3.3 yield similar results as expected. The coefficients for the price premium mean that increasing the E85 price premium by 10 cents decreases the probability that a motorist chooses E85 by between 2.35 and 2.67 percent.

Table 3.9 shows marginal effects for the E85 price ratio. A 0.1 increase in the log of the price ratio decreases the probability a motorist chooses E85 by about 7.43 percent in Model 3.4. Like in the models for the price premium, the corrections for endogenous sampling reduce the magnitude of the price ratio coefficient. In Models 3.5 and 3.6, a 0.1 increase in the log of the price ratio decreases the probability a motorist chooses E85 by about 6.7 and 6.5 percent.

Table 3.8 Marginal effects for Models 3.1 to 3.3 with the E85 premium and all variables

Variable	Model 3.1 no sample correction	Model 3.1 standard errors	Model 3.2 general subset	Model 3.2 standard errors	Model 3.3 survey weights	Model 3.3 standard errors
E85 premium	**-0.267**	**0.072**	**-0.239**	**0.083**	**-0.235**	**0.068**
Government FFV	**0.363**	**0.083**	**0.392**	**0.077**	**0.338**	**0.071**
Company FFV	-0.112	0.058	-0.030	0.059	**-0.132**	**0.057**
Other non-personal FFV	0.003	0.061	-0.017	0.072	-0.002	0.057
FFV type: truck	0.010	0.038	0.047	0.038	0.010	0.033
FFV type: SUV	-0.039	0.036	-0.042	0.039	-0.034	0.033
FFV type: van	-0.013	0.047	-0.033	0.055	-0.037	0.045
Badge	0.026	0.030	0.033	0.033	0.029	0.028
Female	0.018	0.031	0.035	0.034	0.023	0.028
Age	0.002	0.001	0.001	0.001	0.002	0.001
Miles per year (thousands)	-0.002	0.001	-0.001	0.001	-0.001	0.001
E85 better for env.	-0.036	0.049	-0.024	0.053	-0.036	0.045
E10 better for env.	-0.170	0.087	-0.148	0.097	**-0.202**	**0.089**
No diff. for environment	-0.016	0.058	-0.042	0.061	-0.022	0.053
E85 better for engine	**0.112**	**0.045**	**0.140**	**0.049**	**0.130**	**0.042**
E10 better for engine	-0.044	0.047	-0.043	0.053	-0.029	0.045
No diff. for engine	0.001	0.051	0.041	0.054	0.014	0.048
E85 better for economy	**0.162**	**0.053**	**0.160**	**0.058**	**0.154**	**0.049**
E10 better for economy	0.017	0.061	0.019	0.068	0.032	0.055
No diff. for economy	0.124	0.065	**0.148**	**0.069**	**0.128**	**0.060**
E85 better for natl. sec.	0.037	0.038	0.034	0.040	0.040	0.034
E10 better for natl. sec.	0.023	0.054	0.049	0.058	0.009	0.048
No diff. for natl. security	-0.035	0.049	-0.016	0.053	-0.024	0.043
E85 better mpg	**0.116**	**0.051**	0.038	0.058	**0.119**	**0.047**
E10 better mpg	0.074	0.043	**0.102**	**0.049**	**0.087**	**0.040**
No diff. mpg	**0.152**	**0.062**	**0.161**	**0.066**	**0.176**	**0.054**
Colorado Springs	-0.023	0.063	0.014	0.066	0.021	0.056
Los Angeles	**0.450**	**0.047**	**0.446**	**0.047**	**0.420**	**0.038**
Little Rock	0.028	0.048	0.030	0.051	0.033	0.044
Sacramento	**0.355**	**0.049**	**0.363**	**0.044**	**0.353**	**0.040**
Tulsa	**0.093**	**0.046**	**0.105**	**0.049**	**0.096**	**0.041**

The E85 premium is the nominal (not energy-adjusted) per gallon E85 price minus the E10 price. Model 3.1 does not correct for the endogenous stratification in the sample. Model 3.2 uses only the subset of the data that is representative of the general population. Model 3.3 uses probability weights to correct for the endogenous stratification. Variables in bold are significant at the 5 percent level. All dummies equal zero is personal vehicle, vehicle type is car, no FFV badge, male, answers 'don't know' to all fuel opinion questions, and station region is Des Moines.

Table 3.9 Marginal effects for Models 3.4 to 3.6 with the E85 ratio and all variables

Variable	Model 3.4 no sample correction	Model 3.4 standard errors	Model 3.5 general subset	Model 3.5 standard errors	Model 3.6 survey weights	Model 3.6 standard errors
Log E85 ratio	**-0.739**	**0.193**	**-0.668**	**0.221**	**-0.651**	**0.180**
Government FFV	**0.355**	**0.083**	**0.392**	**0.078**	**0.332**	**0.071**
Company FFV	**-0.115**	**0.057**	-0.037	0.058	**-0.136**	**0.056**
Other non-personal FFV	0.000	0.061	-0.020	0.073	-0.005	0.058
FFV type: truck	0.011	0.037	0.047	0.038	0.011	0.033
FFV type: SUV	-0.036	0.037	-0.041	0.039	-0.033	0.033
FFV type: van	-0.011	0.047	-0.032	0.055	-0.036	0.045
Badge	0.026	0.030	0.033	0.033	0.029	0.028
Female	0.017	0.031	0.036	0.034	0.022	0.028
Log age	0.057	0.044	0.041	0.045	0.064	0.040
Log miles per year (k)	-0.034	0.021	-0.024	0.022	-0.026	0.019
E85 better for env.	-0.035	0.049	-0.022	0.052	-0.035	0.045
E10 better for env.	-0.169	0.088	-0.146	0.097	**-0.201**	**0.089**
No diff. for environment	-0.017	0.058	-0.042	0.061	-0.022	0.054
E85 better for engine	**0.110**	**0.045**	**0.137**	**0.049**	**0.128**	**0.042**
E10 better for engine	-0.045	0.047	-0.046	0.053	-0.031	0.045
No diff. for engine	0.000	0.051	0.038	0.054	0.012	0.048
E85 better for economy	**0.160**	**0.053**	**0.159**	**0.057**	**0.152**	**0.049**
E10 better for economy	0.017	0.061	0.020	0.067	0.033	0.055
No diff. for economy	0.123	0.065	**0.149**	**0.068**	**0.128**	**0.060**
E85 better for natl. sec.	0.039	0.038	0.036	0.039	0.042	0.034
E10 better for natl. sec.	0.023	0.054	0.049	0.059	0.008	0.048
No diff. for natl. security	-0.032	0.049	-0.014	0.053	-0.023	0.043
E85 better mpg	**0.113**	**0.051**	0.036	0.058	**0.117**	**0.047**
E10 better mpg	0.073	0.043	**0.102**	**0.049**	**0.087**	**0.040**
No diff. mpg	**0.152**	**0.062**	**0.159**	**0.066**	**0.177**	**0.054**
Colorado Springs	-0.015	0.063	0.021	0.067	0.027	0.057
Los Angeles	**0.465**	**0.046**	**0.460**	**0.046**	**0.434**	**0.037**
Little Rock	0.007	0.046	0.013	0.049	0.015	0.043
Sacramento	**0.368**	**0.048**	**0.378**	**0.043**	**0.366**	**0.040**
Tulsa	0.074	0.044	0.090	0.047	**0.081**	**0.040**

The E85 ratio is the E85 price divided by the E10 price. Model 3.4 does not correct for the endogenous stratification in the sample. Model 3.5 uses only the subset of the data that is representative of the general population. Model 3.6 uses probability weights to correct for the endogenous stratification. Variables in bold are significant at the 5 percent level. All dummies equal zero is personal vehicle, vehicle type is car, no FFV badge, male, answers ‘don’t know’ to all fuel opinion questions, and station region is Des Moines.

Besides the coefficients on the price variables, the estimated coefficients from the corresponding E85 premium and E85 ratio models match quite closely.[10] The motorists' age, gender, vehicle type, how many miles driven per year, opinion about which fuel is better for national security, and the presence of an FFV badge were not significant factors for the probability of choosing E85.

The results indicate that government vehicles are about 34 percent to 39 percent more likely to use E85 than personal vehicles, keeping all else equal. This is likely due to policies requiring government employees to choose E85 regardless of the fuel prices and the employees' personal opinions. On the other hand, we estimate company vehicles are about 11 to 13 percent less likely to use E85 than personal vehicles. This could be due to the policies of various companies or because the motorists are not familiar with E85 or FFVs because the work vehicle is not their primary vehicle, and the motorists are not financially responsible for their fuel choice. In Model 3.2 and Model 3.5, the company vehicle effect is not statistically significant.

For the question about which fuel is better for the environment, motorists who respond that E10 is better are about 15 to 20 percent less likely to choose E85 than motorists who respond that they don't know. The motorists who respond E85 or no difference for the environment are not significantly more or less likely to use E85 than the 'don't knows'. For the question about which fuel is better for the vehicle's engine, motorists who respond that E85 is better for the engine are about 11 to 14 percent more likely to choose E85 than motorists who respond that they don't know. The motorists who respond E10 or no difference for the engine are not significantly more or less likely to use E85 than the 'don't knows'.

[10] Age and number of miles driven per year (in thousands) are the only continuous variables besides the E85 premium and the E85 price ratio. For consistency with the theoretical model, the variables are logged in the price ratio models so their interpretation is different from the models for the premium.

For the motorists' opinions about which fuel is better for the economy, motorists who respond that E85 is better for the economy are about 15 to 16 percent more likely to choose E85, and motorists who respond that there is no difference for the economy are about 12 to 15 percent more likely to choose E85 than motorists who respond that they do not know which is better for the economy. Motorists who respond that E10 is better for the economy are not significantly more or less likely to use E85 than the 'don't knows'. No motorist opinion about which fuel is better for national security is significant for the probability of choosing E85.

The last question of the survey was, "Which fuel yields more miles per gallon?" Motorists who respond that E85 yields more miles per gallon are about 12 percent more likely to choose E85 than motorists who respond that they don't know. Motorists who respond that there is no difference for which fuel yields more miles per gallon are about 15 to 18 percent more likely to choose E85 than motorists who respond that they don't know. And even motorists who respond that E10 yields more miles per gallon are about 7 to 18 percent more likely to choose E85 than motorists who respond that they don't know. This is likely because the motorists who answer 'don't know' are motorists who happen to own FFVs but are mostly unaware of E85, they have never used it, and they never think about it. Thus the motorists who answer, 'don't know' are the motorists who are least likely to use E85.

For the station-location effects we find that motorists who fuel in LA are about 42 to 45 percent more likely to use E85 than motorists in DM, motorists who fuel in SAC are about 35 to 36 percent more likely to use E85 than DM motorists, and motorists in TS are about 9 to 10 percent more likely to use E85 than motorists in DM when all other factors are equal. This means that a motorist with the same vehicle and fuel opinions facing the same prices is more likely to choose E85 in TS than DM. To investigate and identify the regional effect further, we estimate Models 3.7 to 3.12 with only the vehicle ownership and station region dummies.

Tables 3.10 and 3.11 show results for Models 3.7 to 3.12, which contain only the relative fuel price variable and station location dummy variables. We find that the E85 premium and the E85 ratio are significant and the marginal effects are about 20 percent larger in magnitude than their counterparts in Models 3.1 to 3.6. Most of the location effects in Models 3.7 to 3.12 are similar in size to their counterparts in Models 3.1 to 3.6. Motorists in LA and SAC are significantly more likely to use E85 than motorists in DM. However despite the difference in opinions between the DM and TS and LR about which fuel is better for the environment or the economy, the probability of a motorist choosing E85 is not significantly different in these regions after controlling for fuel prices. Only the California effects are significant.

Table 3.10 Marginal effects for Models 3.7 to 3.9 with E85 premium and location variables

Variable	Model 3.7 no sample correction	Model 3.7 standard errors	Model 3.8 general subset	Model 3.8 standard errors	Model 3.9 survey weights	Model 3.9 standard errors
E85 premium	**-0.352**	**0.079**	**-0.306**	**0.089**	**-0.313**	**0.075**
Colorado Springs	-0.059	0.064	-0.010	0.070	-0.008	0.058
Los Angeles	**0.473**	**0.054**	**0.401**	**0.051**	**0.401**	**0.044**
Little Rock	-0.018	0.045	-0.001	0.050	-0.001	0.041
Sacramento	**0.365**	**0.053**	**0.357**	**0.050**	**0.357**	**0.046**
Tulsa	0.002	0.044	0.007	0.050	0.008	0.040

The E85 premium is the nominal (not energy-adjusted) per gallon E85 price minus the E10 price. Model 3.7 does not correct for the endogenous stratification in the sample. Model 3.8 uses only the subset of the data that is representative of the general population. Model 3.9 uses probability weights to correct for the endogenous stratification. Variables in bold are significant at the 5 percent level. When all location dummies equal zero, the station region is Des Moines.

Table 3.11 Marginal effects for Models 3.10 to 3.12 with the E85 ratio and location variables

Variable	Model 3.10 no sample correction	Model 3.10 standard errors	Model 3.11 general subset	Model 3.11 standard errors	Model 3.12 survey weights	Model 3.12 standard errors
Log E85 ratio	**-0.964**	**0.209**	**-0.848**	**0.236**	**-0.861**	**0.198**
Colorado Springs	-0.048	0.065	0.002	0.071	0.003	0.059
Los Angeles	**0.497**	**0.052**	**0.422**	**0.050**	**0.424**	**0.042**
Little Rock	-0.047	0.044	-0.025	0.049	-0.026	0.039
Sacramento	**0.384**	**0.053**	**0.373**	**0.049**	**0.374**	**0.046**
Tulsa	-0.021	0.042	-0.012	0.048	-0.012	0.038

The E85 ratio is the nominal (not energy-adjusted) per gallon E85 price divided by the E10 price. Model 3.10 does not correct for the endogenous stratification in the sample. Model 3.11 uses only the subset of the data that is representative of the general population. Model 3.12 uses probability weights to correct for the endogenous stratification. Variables in bold are significant at the 5 percent level. When all location dummies equal zero, the station region is Des Moines.

Figure 3.3 shows the predicted share of flex motorists who choose E85 given the E85 premium. Because the regional effects for the California locations are considerable, separate results are shown for motorists in Retailer A’s regions and for motorists in Retailer B’s regions. From Model 3.3, in Retailer A’s regions, about 11 percent of the general population of flex motorists choose E85 when the E85 premium is \$0/gallon, about 16 percent of flex motorists choose E85 when the E85 premium is −\$0.25/gallon, about 23 percent choose E85 when the premium is −\$0.50/gallon, and about 31 percent choose E85 when the premium is −\$0.75/gallon. The E85 premium that makes the average motorist among Retailer A’s stations indifferent and have exactly 50 percent probability of choosing E85 or E10 is −\$1.21/gallon, meaning if the E85 price per gallon were \$1.21 lower than the E10 price per gallon, we would expect about half of the general population of flex motorists in Retailer A’s urban areas to fuel with E85. Based on an average E10 price of \$2.39/gallon, it means an E85 price of \$1.18/gallon, a 51 percent discount, would induce 50 percent of motorists to purchase E85. In Retailer B’s station location areas of SAC and LA, Model 3.3 estimates about 74 percent of the general population of flex motorists choose E85 when the E85 premium is \$0/gal, 81 percent choose E85 when the premium is −\$0.25/gallon, 87 percent choose E85 when the premium is −\$0.50/gallon, and 91 percent choose E85 when the premium is −\$0.75/gallon. The E85 premium that makes the average motorist in Retailer B's regions probability of choosing E85 equal one half is \$0.61.

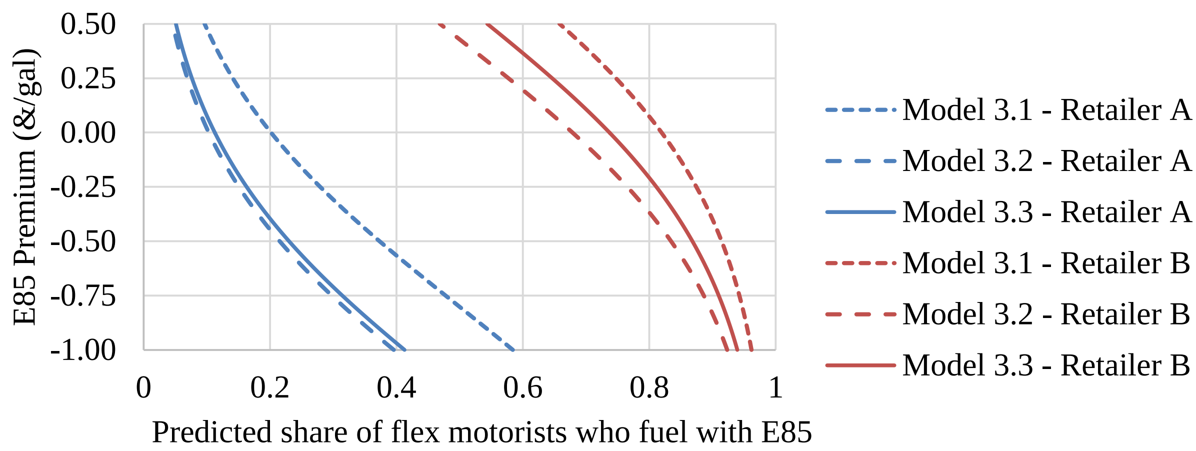

Figure 3.3 Predicted probabilities from models using E85 premium

Predicted shares are logit probabilities calculated at each price for the average motorist in each retailer's area using the results of the indicated models. Retailer A is in the Midwest and Retailer B is in California. Model 3.1 does not correct for sample selection. Model 3.2 corrects for the endogenous stratification in the sample by removing observations from motorists who drove out of their way for E85. Model 3.3 corrects for endogenous stratification by using probability weights.

Figure 3.4 shows the probability that the average motorist chooses E85 given the E85 ratio. From Model 3.6 with the probability weights to correct for endogenous stratification, in Retailer A's location areas, about 10 percent of the general population of flex motorists choose E85 when the E85 ratio is 1.0 (meaning the nominal, not energy-adjusted E85 and E10 prices are the same), about 16 percent of flex motorists choose E85 when the E85 ratio is 0.9, about 24 percent choose E85 when the ratio is 0.8, and about 38 percent choose E85 when the ratio is 0.7.

The E85 ratio that makes the average motorist among Retailer A’s stations have exactly 50 percent probability of choosing E85 or E10 is 0.63. In Retailer B’s station location areas of SAC and LA, from Model 3.6, about 74 percent of the general population of flex motorists choose E85 when the E85 ratio is 1.0, 82 percent choose E85 when the ratio is 0.9, 89 percent choose E85 when the ratio is 0.8, and 94 percent choose E85 when the ratio is 0.7. The estimated population shares for the E85 ratio models closely match the corresponding shares from the E85 premium models. The E85 ratio that makes the average motorist in Retailer B’s regions probability equal half is 1.25.

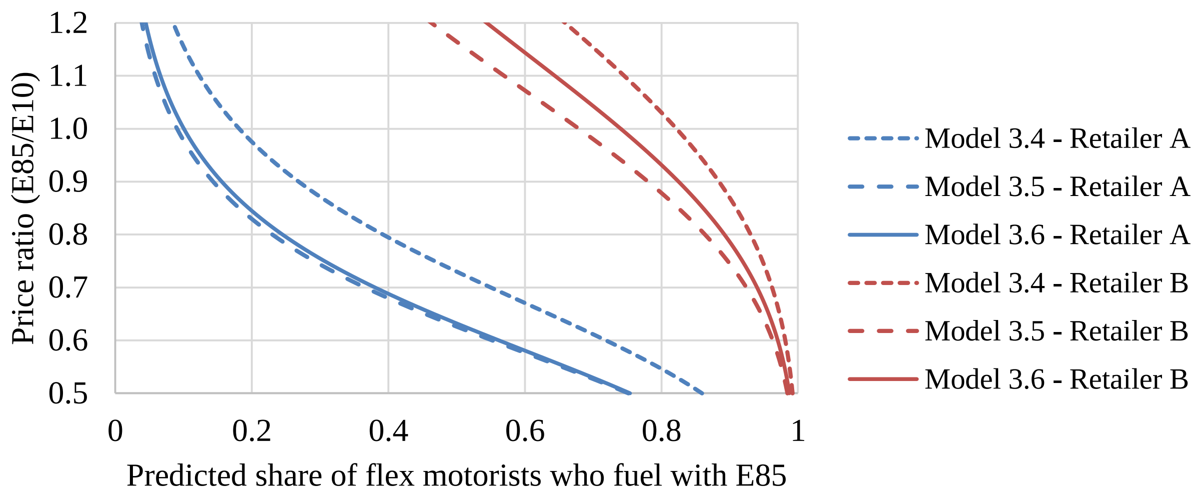

Figure 3.4 Predicted probabilities from models using E85 ratio

Predicted shares are logit probabilities calculated at each price for the average motorist in each retailer’s area using the results of the indicated models. Retailer A is in the Midwest and Retailer B is in California. Model 3.4 does not correct for sample selection. Model 3.5 corrects for the endogenous stratification in the sample by removing observations from motorists who drove out of their way for E85. Model 3.6 corrects for endogenous stratification by using probability weights.

3.5.3 Goodness of fit

We use two measures of goodness of fit to compare and analyze how well our models fit the data. First we measure how well the models predict the observed outcomes, and second we use McFadden's pseudo R-squared as a measure of how much of the observed variation in fuel choices is explained by the models. For each motorist in the sample, we calculate the probability the motorist chooses E85 given the motorist's characteristics and the coefficient estimates from the model. The predicted outcome is E85 if the predicted probability of choosing E85 is greater than 0.5. Then we compare the actual outcomes with the model's predicted outcomes. Table 3.12 shows goodness of fit measures.

Table 3.12 Comparison of goodness of fit

		Number of observations	Percentage of correct predictions	Value of log-likelihood	Pseudo R-squared
Price premium	Model 3.1	893	77.49%	-423.34	0.316
	Model 3.2	681	80.47%	-293.22	0.320
	Model 3.3	893	76.60%	-380.04	0.355
Price ratio	Model 3.4	893	77.38%	-422.97	0.316
	Model 3.5	681	80.03%	-293.24	0.320
	Model 3.6	893	76.48%	-379.80	0.356
Price premium	Model 3.7	893	72.23%	-489.29	0.209
	Model 3.8	681	78.27%	-346.54	0.197
	Model 3.9	893	71.22%	-454.82	0.229
Price ratio	Model 3.10	893	72.23%	-488.62	0.210
	Model 3.11	681	78.27%	-346.02	0.198
	Model 3.12	893	71.22%	-454.28	0.230

Predicted outcomes are the outcomes with the higher predicted probabilities of being chosen for each observation. McFadden's pseudo R-squared is $1 - \ln \mathrm{L}_{ur} / \ln \mathrm{L}_{r}$ where the unrestricted model with all parameters estimated freely, and the restricted model has only the intercept and all other parameters equal zero.

Models 3.2, 3.5, 3.8 and 3.11 with only the representative subset have the highest predictive success at about 80 percent. Models 3.1, 3.4, 3.7 and 3.10 that use the entire estimation sample and do not correct for endogenous stratification correctly predict between 73 and 77 percent of fuel choices correctly. Finally, Models 3.3, 3.6, 3.9 and 3.12 that use the probability weights also correctly predict between 73 and 77 of fuel choices. Based on the rates of correct prediction, there is no difference between the predictive power for the E85 premium models and the E85 ratio models. On this basis we cannot conclude that motorists tend to make their fuel decision based on the E85 premium or the E85 ratio.

We also measure how well the model explains observed fuel choices using McFadden's pseudo R-squared. The pseudo R-squared ${R^2}_{MCF}$ is a transformation of the log-likelihood function into an index defined as

$${R^2}_{MCF} = 1 - \frac{\ln L_{ur}}{\ln L_r},$$

where $\ln L_{ur}$ is the log-likelihood value of the unrestricted model where all parameters are chosen to maximize the function, and $\ln L_r$ is the log-likelihood value of the restricted model where all parameters are restricted to equal zero except for the intercept. 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). Maximized log-likelihood values and pseudo R-squared values are included in Table 3.12.

As with the predicted outcomes, the E85 premium and E85 ratio models are almost identical in log-likelihood and pseudo R-squared values, and the models with all independent variables perform better than the models with only the price and station location variables. The pseudo R-squared values are about 0.32 for Models 3.1, 3.2, 3.4 and 3.5. For Models 3.3 and 3.6 that use probability weights in the likelihood function, the pseudo R-squared values are slightly higher at 0.36. The results are similar for Models 3.7 to 3.12 that do not include all control variables. The pseudo R-squared values are about 0.23 for Models 3.7, 3.8, 3.10 and 3.11. For Models 3.9 and 3.12 that use probability weights in the log-likelihood function, the pseudo R-squared values are higher at 0.26. The pseudo R-squared values indicate that all models fit the data well and that including motorists' characteristics and fuel opinions adds a measurable degree of fit. On the basis of the pseudo R-squared, we cannot conclude in favor of the E85 premium models or the E85 ratio models. The reason might be that there is little variation in prices observed within each region during our visits so that we cannot identify motorists' decision rules based on the data.

3.6 Conclusion

In the study presented in this chapter, we estimate preferences for E85 relative to E10 among flex motorists in different regions of the United States. With the collaboration of two E85 retailers, we conduct an intercept survey at E85 stations to collect both revealed fuel preferences and stated motorist fuel opinions. We visit Retailer A's E85 stations in the urban areas of Colorado Springs, Des Moines, Little Rock, and Tulsa, and we visit Retailer B's E85 stations in Los Angeles and Sacramento. Fuel choices for motorists who fuel at Retailer A's stations differ significantly from fuel choices for motorists who fuel at Retailer B's stations. When the nominal E85 price per gallon is about 80 percent of the E10 price per gallon, we observe about 40 percent of flex motorists who fuel at Retailer A's E85 stations choose E85, whereas nearly 90 percent of flex motorists at Retailer B's stations choose E85. The marked difference in preferences could be because Retailer B uses specialized marketing techniques to promote biofuels to local flex motorists or because there are fewer E85 stations in Retailer B's areas than in Retailer A's areas, so each station of Retailer B's has flex motorists coming from further away, even waiting in line, to fuel with E85. Thus the flex motorists we observe at Retailer B's stations may disproportionately represent the upper tail of the distribution of WTP for E85.

We estimate preferences for about 900 flex motorists who are an endogenously stratified sample of the local population of flex motorists because we intercept them at E85 stations. Thus the probability that a motorist appears in our sample is correlated with the motorist's WTP for E85. We apply corrective probability weights for each region to the observations so that our estimates represent the general population of local flex motorists rather than the endogenously stratified sample. We find that a $0.10 increase in the E85 premium causes the probability that a motorist chooses E85 to decrease by 2.4 percent, on average.

Other significant factors affecting the probability that a flex motorist chooses E85 are the vehicle ownership (personal, government, company, etc.), the motorists' opinions about the fuels with respect to the environment, the engine, and fuel economy, and whether the motorist is in Retailer A's area or Retailer B's area. However, the regional dummy variables are not statistically significant from each other within a retailer's region. This is a key result and it means that when prices are 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 that the general opinion of ethanol among flex motorists in our sample is more favorable 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 region to project national demand, though we would necessarily need to make adjustments for California. We also find that motorists' ages, genders, vehicle types, how many miles they drive per year, opinions about which fuel is better for national security, and whether they have FFV badges on their vehicles are not significant factors.

In Colorado Springs, Des Moines, Little Rock and Tulsa we estimate that about 10 percent of the local population of flex motorists choose E85 when the nominal E85 price ratio is 1.0, about 16 percent of flex motorists choose E85 when the E85 price ratio is 0.9, about 24 percent choose E85 when the E85 price ratio is 0.8, and about 38 percent choose E85 when the E85 price ratio is 0.7. The mean of the WTP distribution, where 50 percent of motorists choose E85, is 0.63. In Sacramento and Los Angeles, about 74 percent of the local population of flex motorists choose E85 when the E85 price ratio is 1.0, 82 percent choose E85 when the price ratio is 0.9, 89 percent choose E85 when the price ratio is 0.8, and 94 percent choose E85 when the price ratio is 0.7. The mean of the distribution is 1.25.

Our models are quite successful in fitting the observed survey data both in the measure of percent correctly predicted outcomes and in McFadden's pseudo R-squared values. We estimate models where the motorists respond to the relative difference in fuel prices as well as models where the motorists respond to the absolute difference in fuel prices. We find that there is virtually no difference in how well the models fit the data in either the percentage of correctly predicted outcomes or the log-likelihood and pseudo R-squared values. Thus we cannot say whether motorists are generally responding to the E85 price ratio or the E85 price premium when they make fuel choices.

CHAPTER 4.

ESTIMATING WILLINGNESS TO PAY FOR E85 USING DATA FROM AN INTERCEPT SURVEY: EVIDENCE FROM STATED PREFERENCE

4.1 Introduction

The goal of this dissertation is to estimate the distribution of willingness to pay for E85 as a substitute for E10 among flex motorists in the United States. With the intercept survey we obtain revealed-preference data by observing motorists' actual fuel purchases as well as stated-preference data by presenting the motorists with hypothetical fuel prices and asking which fuel they would choose in the hypothetical scenario. In Chapter 3, we use only the RP data to estimate the distribution of WTP for E85, and in this chapter, we incorporate the SP data, compare results, and discuss implications.

Combining RP data and SP data has proven useful in many studies, and a brief review of the literature is in the next section. In our setting, the SP data are used to expand the amount of variation in the observed E85 and E10 prices in the RP data. When we conducted our survey, we spent a relatively short amount of time in each region except for Iowa, and we did not observe significant variation in fuel prices within any of the other regions. To expand the range of observed fuel prices, 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 use the SP data with the wider range of prices to add precision to the parameter estimates from Chapter 3, and in doing so we consider the specific nature of how the SP data are generated.

There are two main reasons why we treat the SP data differently from the RP data. First the SP data are not observations of actual fuel purchases, but rather are survey responses of what motorists say they would do in a hypothetical situation. Motorists may unintentionally or intentionally misrepresent their preferences when responding to the survey for a number of reasons which are listed in the next section. Second the hypothetical fuel prices that are presented to each motorist depend on the motorist's RP fuel choice. Motorists who choose E85 are presented with a hypothetical scenario where either the price of E85 is higher or the price of E10 is lower, and motorists who choose E10 are presented with a hypothetical scenario where either the price of E10 is higher or the price of E85 is lower. This means that the motorists' unobservable demand shifters are correlated with the hypothetical prices. We describe the issue in detail in Section 4.3 where we discuss the models.

We find that when we ignore the nature of the SP data-generating process, the coefficient estimates, especially on the price variables, are indeed biased. However, when we properly model the SP survey design, as expected, the coefficient estimates and marginal effects are similar to those estimated with the RP data alone, but the estimated standard errors of the relative price variable coefficients are significantly smaller than the estimated standard errors from the RP-only model. As in Chapter 3, we estimate two versions of the models: one where the motorists respond to the E85 premium, and one where motorists respond to the E85 price ratio. We compare measures of goodness of model fit to determine whether one decision rule prevails among motorists.

The next section of this chapter offers background information about SP choice experiments and a review of the literature related to combining RP and SP data. In Section 4.3, we describe two theoretical models for combining SP and RP data. Section 4.4 contains information about the SP data. The empirical models and estimation results are in Section 4.5, and Section 4.6 concludes.

4.2 SP Choice Experiments: Background and Related Literature

The use of SP data from choice experiments is common in many types of economic studies. An advantage of hypothetical choice experiments over actual market transactions is that they allow researchers to ‘observe’ consumer choices between products and/or attribute levels that may not exist in the market. Louviere et al. (2000) provide a review of stated choice experiment design methods, techniques for model estimation, and applications to marketing, transportation, and environmental valuation case studies. As discussed in Chapter 1, recent studies by Jensen et al. (2010), Petrolia et al. (2010), and Aguilar et al. (2015) use SP data collected by nationwide mail and online surveys to study various aspects of WTP for E85 and other fuels.

A disadvantage of SP data is that what consumers say they will do in a hypothetical setting may not be what consumers would actually do if the setting were real. In the literature, this deviation between stated survey responses and real market behavior is known as hypothetical bias. Hensher (2010) examines hypothetical bias in estimates of WTP from a number of recent studies where RP and SP data are both available and reviews possible causes and remedies.

There are many potential causes of hypothetical bias. An example that is prominent in the literature is that it may be easier for respondents to say that they would purchase a good or be willing to pay a high amount for some attribute or amenity than to actually pay for it. One common way to reduce this type of hypothetical bias has been the use of ‘cheap talk’ scripts proposed by Cummings et al. (1995). Before eliciting responses, the researcher reads a script explaining that respondents have a tendency to overstate WTP. It is possible that it is slightly easier for motorists to say that they would switch fuels under the hypothetical prices than to actually do it because switching might require using a specific island at the fuel station where there are E85 nozzles (and in rare cases this means waiting in line), but this is a minor cost that would only affect motorists switching from E10 to E85. On the other hand, the ‘inertia’ effect could go the other direction, and respondents tend to overstate their willingness to stay with their current fuel choice even when they would have actually switched. In our survey, we did not prime motorists in either direction by saying that respondents tended to overstate or understate switching behavior before eliciting the SP fuel choice.

Another reason for hypothetical bias is that respondents may misrepresent their preferences to influence the result of the study. In experimental settings, participants may know the researcher’s hypothesis and want to influence the outcome. In our setting, it is possible that motorists assume that their responses will be used for the retailer’s future pricing decisions and respond in a way that may result in lower prices rather than responding with their true behavior. For example, when we ask motorists if they would make the same fuel choice if the fuel was more expensive, they may respond negatively simply because they do not want the fuel to be more expensive the next time they visit the station. To mitigate this effect, we began each survey by clearly stating that we were conducting academic research, and the surveys were conducted by undergraduate and graduate students wearing Iowa State University jackets.

SP choice experiments can also suffer from hypothetical bias related to respondents being in the mindset of participating in an experiment rather than making consequential market choices or not truly understanding the choices in all of the unlisted attributes. However, our SP choice experiment is not a classroom/lab experiment, it is a field experiment where motorists are literally in the act of purchasing fuel so factors that affect stated preferences in experimental settings are minimized. It is the ideal setting for asking flex motorists about fuel choices, and the motorists can assume that the hypothetical fuel choices are identical to the real fuel choices in every way except for the hypothetical prices.

Nevertheless, hypothetical bias could exist in our data for other reasons still. Among them is the issue of anchoring, where respondents base their responses on the first number they are presented. In our case, some motorists may use the actual fuel prices from the RP setting as an anchor for judging how favorable or unfavorable the hypothetical prices for the fuel choices are. On the other hand, some motorists may have their own pre-calculated rules where if the price of E85 is a certain amount or percentage below the price of E10, they use E85.

Another factor contributing to hypothetical bias that may persist in our experiment is prominence, where the attribute that is varied in the hypothetical scenario (the fuel price) is thereby made more prominent to the respondent, and the respondent puts more weight on that attribute in the hypothetical SP experiment than they would otherwise. In the intercept survey of flex motorists conducted by Salvo and Huse (2013), motorists were asked for the 'main reason' motivating their fuel choice, and the overwhelming majority response was the fuel price. So while asking motorists about varied fuel prices may make the price a more prominent attribute of the fuel choice, it is likely already the most prominent factor driving the decision. Even so, it is possible that motorists are more subject to habit and routine than they realize, and if they had actually pulled in to the station and the hypothetical prices had been prevailing, motorists may never have even noticed or bothered to make a comparison before making their same usual fuel choices.

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 the fuel prices, and the SP data were generated to have wider variation in fuel prices. The traditional method for estimating models on combined RP and SP data that has been used in the transportation and the environmental economic literature is described in Ben-Akiva and Morikawa (1990), Hensher and Bradley (1993), Adamowicz et al. (1994), and Hensher et al. (1999). This traditional model of combined RP and SP data is detailed in the next section.

The intuition behind the traditional approach to combining RP and SP data is that the unobserved factors are allowed to be different for the two types of data. This is done 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 unobservable respondents' characteristics are not correlated with the hypothetical options.

Train and Wilson (2008) and Train and Wilson (2009) consider SP data constructed from RP choices, called 'SP-off-RP', and present 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. The non-independent unobserved factors persist in the hypothetical SP scenario, and we need to account for this when considering the SP choices.

4.3 Theoretical Models for Combining RP and SP Data

In this section we describe two variations of the WMLE logit model from Chapter 3 with weights to correct for endogenous stratification. The empirical models are modified to accommodate the specifics of the SP data-generating process. The first model is the traditional pooled approach where RP and SP data are combined and treated differently, but the nature of the SP-off-RP data-generating process is ignored. The second model is the SP-off-RP model of Train and Wilson (2008) which features the necessary considerations for the nature of the SP data from the intercept survey

4.3.1 Models with the E85 premium

As in Chapter 3, the indirect utility flex motorist i derives from fuel j in the RP setting is

$$V_{ij}\left(I_i, p_{ij}, \boldsymbol{x}_i, \varepsilon_{ij}\right) = v_{ij}\left(I_i, p_{ij}, \boldsymbol{x}_i\right) + \varepsilon_{ij} = \gamma_j I_i + \alpha_j p_{ij} + \boldsymbol{x}_i{}'\boldsymbol{\beta}_j + \varepsilon_{ij},$$

where I_i is the motorist's income, p_{ij} is the price of fuel j observed by motorist i, $\boldsymbol{x}_i$ is a vector of observed motorist and station characteristics, and ε_{ij} is a type 1 generalized extreme value random variable known to the motorist, but not observable to us.

As shown in Chapter 3, the probabilities of a motorist choosing E85 and E10 are respectively given by

$$\Pr(\text{RP E85}_i) = \Lambda(\alpha p_i + \boldsymbol{x}_i{}'\boldsymbol{\beta}),$$

and

$$\Pr(\text{RP E10}_i) = 1 - \Lambda(\alpha p_i + \boldsymbol{x}_i{}'\boldsymbol{\beta}) = \Lambda\left(-(\alpha p_i + \boldsymbol{x}_i{}'\boldsymbol{\beta})\right),$$

where p_i is the E85 premium faced by motorist i.

4.3.1.1 Traditional pooled model of combined RP and SP data

In the traditional pooled approach to combining RP and SP data, the utility that flex motorist i derives from fuel j in the SP setting is

$$W_{ij}\left(I_i, \bar{p}_{ij}, \boldsymbol{x}_i, \eta_{ij}\right) = v_{ij}\left(I_i, \bar{p}_{ij}, \boldsymbol{x}_i\right) + \eta_{ij} = \gamma_j I_i + \alpha_j \bar{p}_{ij} + \boldsymbol{x}_i{}' \boldsymbol{\beta}_j + \eta_{ij},$$

where $\bar{p}_{ij}$ is the hypothetical price of fuel j presented to motorist i, and η_{ij} is a random variable known to the motorist but not observed by the researcher that captures the unobservable factors in the SP setting. Note that the relationship between the observable factors that determine the choice $v_{ij}(\cdot)$ is preserved in the SP utility function.

The decision-making process in the SP setting is the same as in the RP setting so a motorist chooses E85 if $W_{ie}(\cdot) \geq W_{ig}(\cdot)$. However, to account for the differences between the RP setting and the SP setting, the distribution of η_{ij} is allowed to differ from the distribution of ε_{ij}. This is implemented by writing that $\eta_i \equiv \zeta\left(\eta_{ig} - \eta_{ie} + \tau\right)$ follows a logistic distribution with a scale equal to one. The terms τ and ζ are introduced to normalize the distribution of η_i to have a mean of zero and a scale of one by shifting and rescaling the error distributions in the SP setting. The probability that the motorist chooses E85 in the SP setting is

$$\Pr(\text{SP E85}_i) = \Lambda\left(\zeta(\alpha \bar{p}_i + \boldsymbol{x}_i{}' \boldsymbol{\beta} + \tau)\right),$$

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

$$\Pr(\text{SP E10}_i) = 1 - \Lambda\left(\zeta(\alpha \bar{p}_i + \boldsymbol{x}_i{}' \boldsymbol{\beta} + \tau)\right) = \Lambda\left(-\zeta(\alpha \bar{p}_i + \boldsymbol{x}_i{}' \boldsymbol{\beta} + \tau)\right).$$

The model is estimated on the RP and SP data pooled together. To distinguish observations, we introduce an indicator variable,

$$s_i = \begin{cases} 1 & \text{if observation is SP;} \\ 0 & \text{if observation is RP.} \end{cases}$$

The dependent variable is defined as before:

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

We maximize the WMLE log-likelihood function with sample and population probability weights to correct for the endogenously stratified sample. As in Chapter 3, Q_i is the population proportion and H_i is the sample proportion of the RP fuel choice of motorist i.

$$\text{RP SP WMLE} = \sum_i \frac{Q_i}{H_i}[(1-s_i)(1-y_i)\ln\Pr(\text{RP E10}_i) + (1-s_i)y_i \ln\Pr(\text{RP E85}_i)$$

$$+ s_i(1-y_i)\ln\Pr(\text{SP E10}_i) + s_i y_i \ln\Pr(\text{SP E85}_i)]$$

<u>4.3.1.2 Model accounting for SP-off-RP nature of data</u>

In this variation of the model, we define the utility flex motorist i derives from fuel j in the SP setting as

$$W_{ij}\left(I_i, \bar{p}_{ij}, \boldsymbol{x}_i, \varepsilon_{ij}, \eta_{ij}\right) = V_{ij}\left(I_i, \bar{p}_{ij}, \boldsymbol{x}_i, \varepsilon_{ij}\right) + \eta_{ij} = \gamma_j I_i + \alpha_j \bar{p}_{ij} + \boldsymbol{x}_i{}'\boldsymbol{\beta}_j + \varepsilon_{ij} + \eta_{ij},$$

where $\bar{p}_{ij}$ is the hypothetical price of fuel j presented to motorist i, η_{ij} is a generalized extreme random variable with scale (1/ δ) known to the motorist but not observed by us that captures additional unobservable aspects of the SP setting not present in the RP setting. Note that in this variation of the model, the relationship between both the observable and unobservable factors that determine the RP choice $V_{ij}(\cdot)$ is preserved in the SP utility function. That is, 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 $\bar{p}_{ij}$. Thus the hypothetical prices in the SP setting are endogenous as the 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

$$\eta_i \leq \delta\left[V_{ie}(I_i, \bar{p}_{ie}, \boldsymbol{x}_i, \varepsilon_{ie}) - V_{ig}\left(I_i, \bar{p}_{ig}, \boldsymbol{x}_i, \varepsilon_{ig}\right)\right]$$

where $\eta_i \equiv \delta\left(\eta_{ig} - \eta_{ie}\right)$ is symmetric with a mean of zero and follows a logistic distribution. The δ 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

$$\Pr(\text{SP E85}_i) = \Lambda\left(\delta(\alpha\bar{p}_i + \boldsymbol{x}_i{}'\boldsymbol{\beta} - \varepsilon_i)\right),$$

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

$$\Pr(\text{SP E10}_i) = 1 - \Lambda\left(\delta(\alpha\bar{p}_i + \boldsymbol{x}_i{}'\boldsymbol{\beta} - \varepsilon_i)\right) = \Lambda\left(-\delta(\alpha\bar{p}_i + \boldsymbol{x}_i{}'\boldsymbol{\beta} - \varepsilon_i)\right).$$

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. The conditional probability expressions are:

$$\Pr(\text{SP E10}_i|\text{RP E10}_i) = \Lambda\left(-\delta(\alpha\bar{p}_i + \boldsymbol{x}_i{}'\boldsymbol{\beta} - \varepsilon_i|\varepsilon_i \geq \alpha p_i + \boldsymbol{x}_i{}'\boldsymbol{\beta})\right);$$

$$\Pr(\text{SP E10}_i|\text{RP E85}_i) = \Lambda\left(-\delta(\alpha\bar{p}_i + \boldsymbol{x}_i{}'\boldsymbol{\beta} - \varepsilon_i|\varepsilon_i \leq \alpha p_i + \boldsymbol{x}_i{}'\boldsymbol{\beta})\right);$$

$$\Pr(\text{SP E85}_i|\text{RP E10}_i) = \Lambda\left(\delta(\alpha\bar{p}_i + \boldsymbol{x}_i{}'\boldsymbol{\beta} - \varepsilon_i|\varepsilon_i \geq \alpha p_i + \boldsymbol{x}_i{}'\boldsymbol{\beta})\right);$$

$$\Pr(\text{SP E85}_i|\text{RP E85}_i) = \Lambda\left(\delta(\alpha\bar{p}_i + \boldsymbol{x}_i{}'\boldsymbol{\beta} - \varepsilon_i|\varepsilon_i \leq \alpha p_i + \boldsymbol{x}_i{}'\boldsymbol{\beta})\right).$$

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 likelihood function for the entire sample is

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

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

The ε_i's that enter the conditional SP logits are not observed but we know their distribution 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

$$\Pr(\mathrm{SP\ E85}_i|\mathrm{RP\ E85}_i) = \int \Lambda\big(\delta(\alpha\bar{p}_i + \boldsymbol{x}_i{}'\boldsymbol{\beta} - \varepsilon_i)\big) f(\varepsilon_i|\varepsilon_i \le \alpha p_i + \boldsymbol{x}_i{}'\boldsymbol{\beta}) d\varepsilon_i.$$

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.

Following the method of Train and Wilson (2009), we simulate draws of ε_{ig} and ε_{ie} for each motorist by transforming draws of the uniform distribution $\mathcal{U}(0,1)$ according to the inverted conditional extreme value distribution function. Let μ_1, μ_2 be draws from the uniform distribution. Then conditional on E85 being chosen in the RP setting, a draw of ε_{ie} has the mean shifted up by $-\ln \Pr(\mathrm{RP\ E85}_i)$:

$$\varepsilon_{ie} = -\ln \Pr(\mathrm{RP\ E85}_i) - \ln(-\ln \mu_1) = -\ln \Lambda(\alpha p_i + \boldsymbol{x}_i{}'\boldsymbol{\beta}) - \ln(-\ln \mu_1).$$

Conditional on E85 in the RP setting and ε_{ie}, a draw of ε_{ig} is truncated above at $\alpha d_i + \boldsymbol{x}_i{}'\boldsymbol{\beta} + \varepsilon_{ie}$ so

$$\varepsilon_{ig} = -\ln(-\ln(m(\varepsilon_{ie})\mu_2)),$$

where $m(\varepsilon_{ie}) = \exp\big(-\exp\big(-(\alpha p_i + \boldsymbol{x}_i{}'\boldsymbol{\beta} + \varepsilon_{ie})\big)\big)$.

Similarly conditional on E10 being chosen in the RP setting, a draw of ε_{ig} is

$$\varepsilon_{ig} = -\ln \Pr(\text{RP E10}_i) - \ln(-\ln \mu_1) = -\ln \Lambda\big(-(\alpha p_i + \boldsymbol{x}_i{}'\boldsymbol{\beta})\big) - \ln(-\ln \mu_1),$$

and conditional on E10 in the RP setting and ε_{ig}, a draw of ε_{ie} is

$$\varepsilon_{ie} = -\ln\big(-\ln\big(m(\varepsilon_{ig})\mu_2\big)\big),$$

where $m(\varepsilon_{ig}) = \exp\Big(-\exp\Big(-\big(-(\alpha p_i + \boldsymbol{x}_i{}'\boldsymbol{\beta}) + \varepsilon_{ig}\big)\Big)\Big)$.

The parameters are estimated using maximum likelihood estimation over the joint probability WMLE function with sample and population probability weights to correct for the endogenously stratified sample. Again Q_i is the population proportion and H_i is the sample proportion of the RP fuel choice of motorist i.

$$\begin{aligned}\text{RP SP WMLE} = \sum_i \frac{Q_i}{H_i}\{&(1-y_{i1})(1-y_{i2})\ln[\Pr(\text{RP E10}_i)\cdot\Pr(\text{SP E10}_i|\text{RP E10}_i)]\\ &+(1-y_{i1})y_{i2}\ln[\Pr(\text{RP E10}_i)\cdot\Pr(\text{SP E85}_i|\text{RP E10}_i)]\\ &+y_{i1}(1-y_{i2})\ln[\Pr(\text{RP E85}_i)\cdot\Pr(\text{SP E10}_i|\text{RP E85}_i)]\\ &+y_{i1}y_{i2}\ln[\Pr(\text{RP E85}_i)\cdot\Pr(\text{SP E85}_i|\text{RP E85}_i)]\}.\end{aligned} \tag{4.1}$$

4.3.2 Models with the E85 ratio

As in Chapter 3, the indirect utility flex motorist i derives from fuel j in the RP setting is

$$\tilde{V}_{ij}\big(I_i, p_{ij}, \boldsymbol{x}_{i_i}, \varepsilon_{ij}\big) = \tilde{v}_{ij}\big(I_i, p_{ij}, \boldsymbol{x}_i\big)\cdot\exp\big(\varepsilon_{ij}\big) = \Big(I_i{}^{\tilde{\gamma}_j}\cdot p_{ij}{}^{\tilde{\alpha}_j}\cdot \boldsymbol{x}_i{}^{\tilde{\boldsymbol{\beta}}_j}\Big)\cdot\exp\big(\varepsilon_{ij}\big),$$

where I_i is the motorist's income, p_{ij} is the price of fuel j observed by motorist i, $\boldsymbol{x}_i$ is a vector of observed motorist and station characteristics, ε_{ij} is a type 1 extreme value random variable that is not observable, and if $\boldsymbol{x}_i$ is $k \times 1$, $\boldsymbol{x}_i{}^{\tilde{\boldsymbol{\beta}}_j} = x_{i1}{}^{\tilde{\beta}_{j1}}\cdot x_{i2}{}^{\tilde{\beta}_{j2}}\cdot \ldots \cdot x_{ik}{}^{\tilde{\beta}_{jk}}$.

Recall from Chapter 3 that the choice probabilities for the RP data model are given by:

$$\Pr(\text{RP E10}_i) = \Lambda\big(\tilde{\alpha}\ln r_i + (\ln \boldsymbol{x}_i)'\tilde{\boldsymbol{\beta}}\big),$$

and

$$\Pr(\text{RP E10}_i) = 1 - \Lambda\big(\tilde{\alpha}\ln r_i + (\ln \boldsymbol{x}_i)'\widetilde{\boldsymbol{\beta}}\big) = \Lambda\left(-\big(\tilde{\alpha}\ln r_i + (\ln \boldsymbol{x}_i)'\widetilde{\boldsymbol{\beta}}\big)\right).$$

4.3.2.1 Traditional pooled model of combined RP and SP data

In the traditional pooled approach to combining RP and SP data, the utility flex motorist i derives from fuel j in the SP setting is

$$\widetilde{W}_{ij}\big(I_i, \bar{p}_{ij}, \boldsymbol{x}_i, \tilde{\eta}_{ij}\big) = \tilde{v}_{ij}\big(I_i, \bar{p}_{ij}, \boldsymbol{x}_i\big) \cdot \exp(\tilde{\eta}_{ij}),$$

where $\bar{p}_{ij}$ is the hypothetical price of fuel j presented to motorist i, and $\tilde{\eta}_{ij}$ is a random variable known to the motorist, but not observed by the researcher. Note that the relationship between the observable factors that determine the choice $\tilde{v}_{ij}(\cdot)$ is preserved in the SP utility function.

The motorist chooses E85 in the SP setting if $\widetilde{W}_{ie}(\cdot) \geq \widetilde{W}_{ig}(\cdot)$. We can re-write this decision rule as

$$\tilde{\eta}_i \leq \tilde{\zeta}\left(\ln\frac{{I_i}^{\tilde{\gamma}_e} \cdot {\bar{p}_{ie}}^{\tilde{\alpha}_e} \cdot {\boldsymbol{x}_i}^{\widetilde{\boldsymbol{\beta}}_e}}{{I_i}^{\tilde{\gamma}_g} \cdot {\bar{p}_{ig}}^{\tilde{\alpha}_g} \cdot {\boldsymbol{x}_i}^{\widetilde{\boldsymbol{\beta}}_g}} + \tilde{\tau}\right).$$

where $\tilde{\eta}_i \equiv \tilde{\zeta}\big(\tilde{\eta}_{ig} - \tilde{\eta}_{ie} + \tilde{\tau}\big)$ and follows a logistic distribution and the introduction of $\tilde{\zeta}$ and $\tilde{\tau}$ normalize the distribution to have mean zero and scale one. Following the derivation of the choice probabilities for the RP data model, the probabilities for the SP setting are

$$\Pr(\text{SP E85}_i) = \Lambda\big[\tilde{\zeta}\big(\tilde{\alpha}\ln \bar{r}_i + (\ln \boldsymbol{x}_i)'\widetilde{\boldsymbol{\beta}} + \tilde{\tau}\big)\big],$$

and

$$\Pr(\text{SP E10}_i) = \Lambda\big[-\tilde{\zeta}\big(\tilde{\alpha}\ln \bar{r}_i + (\ln \boldsymbol{x}_i)'\widetilde{\boldsymbol{\beta}} + \tilde{\tau}\big)\big].$$

Again introduce an indicator variable,

$$s_i = \begin{cases} 1 & \text{if observation is SP;} \\ 0 & \text{if observation is RP,} \end{cases}$$

and define the dependent variable,

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

Then we use maximum likelihood estimation over the WMLE log-likelihood function analogous to the log-likelihood function for the price difference model:

$$\text{RP SP WMLE} = \sum_i \frac{Q_i}{H_i}[(1-s_i)(1-y_i)\ln \Pr(\text{RP E10}_i) + (1-s_i)y_i \ln \Pr(\text{RP E85}_i)$$

$$+ s_i(1-y_i)\ln \Pr(\text{SP E10}_i) + s_i y_i \ln \Pr(\text{SP E85}_i)]$$

<u>4.3.2.2 Model accounting for SP-off-RP nature of data</u>

In this variation of the model, we define the utility flex motorist i derives from fuel j in the SP setting as

$$\widetilde{W}_{ij}(I_i, \bar{p}_{ij}, \boldsymbol{x}_i, \varepsilon_{ij}, \eta_{ij}) = \tilde{V}_{ij}(I_i, \bar{p}_{ij}, \boldsymbol{x}_i, \varepsilon_{ij}) \cdot \exp(\eta_{ij})$$

$$= \left(I_i^{\tilde{\gamma}_j} \cdot \bar{p}_{ij}^{\tilde{\alpha}_j} \cdot \boldsymbol{x}_i^{\widetilde{\boldsymbol{\beta}}_j}\right) \cdot \exp(\varepsilon_{ij}) \cdot \exp(\eta_{ij}),$$

where $\bar{p}_{ij}$ is the hypothetical price of fuel j presented to motorist i, η_{ij} is a generalized extreme value random variable with scale $(1/\tilde{\delta})$ known to the motorist but not observed by us. Note that again, the relationship between both the observable and unobservable factors that determine the RP choice $\tilde{V}_{ij}(\cdot)$ is preserved in the SP utility function. That is, the same unobservable ε_{ij} term for motorist i that affects the RP choice carries forward to the SP setting as part of the unobservable error term and also as being correlated with the hypothetical prices.

The motorist chooses E85 in the SP setting if $\widetilde{W}_{ie}(\cdot) \geq \widetilde{W}_{ig}(\cdot)$.We can re-write this decision rule as

$$\eta_i \leq \tilde{\delta}\left(\ln\left(\frac{I_i^{\tilde{\gamma}_e} \cdot \bar{p}_{ie}^{\tilde{\alpha}_e} \cdot \boldsymbol{x}_i^{\widetilde{\boldsymbol{\beta}}_e} \cdot \exp(\varepsilon_{ie})}{I_i^{\tilde{\gamma}_g} \cdot \bar{p}_{ig}^{\tilde{\alpha}_g} \cdot \boldsymbol{x}_i^{\widetilde{\boldsymbol{\beta}}_g} \cdot \exp(\varepsilon_{ig})}\right)\right)$$

where $\eta_i \equiv \tilde{\delta}(\eta_{ig} - \eta_{ie})$ follows a logistic distribution with mean zero and scale one. Again letting $\tilde{\gamma}_e = \tilde{\gamma}_g \equiv \tilde{\gamma}$, $\tilde{\alpha}_e = \tilde{\alpha}_g \equiv \tilde{\alpha}$, $\widetilde{\boldsymbol{\beta}} \equiv \widetilde{\boldsymbol{\beta}}_e - \widetilde{\boldsymbol{\beta}}_g$, and $\bar{r}_i \equiv \bar{p}_{ie}/\bar{p}_{ig}$ simplifies the probability to

$$\Pr(\text{SP E85}_i) = \Lambda\left(\tilde{\delta}\left(\tilde{\alpha}\ln\bar{r}_i + (\ln\boldsymbol{x}_i)'\widetilde{\boldsymbol{\beta}} - \varepsilon_i\right)\right),$$

and

$$\Pr(\text{SP E10}_i) = \Lambda\left(-\tilde{\delta}\left(\tilde{\alpha}\ln\bar{r}_i + (\ln\boldsymbol{x}_i)'\widetilde{\boldsymbol{\beta}} - \varepsilon_i\right)\right).$$

For a single motorist, the joint probability of the motorist's RP and SP choices is the product of the marginal probability of the RP choice and the probability of the SP choice, conditional on the RP choice. The conditional probability expressions are:

$$\Pr(\text{SP E10}_i|\text{RP E10}_i) = \Lambda\left(-\tilde{\delta}\left(\tilde{\alpha}\ln\bar{r}_i + (\ln\boldsymbol{x}_i)'\widetilde{\boldsymbol{\beta}} - \varepsilon_i\middle|\varepsilon_i \geq \tilde{\alpha}\ln r_i + (\ln\boldsymbol{x}_i)'\widetilde{\boldsymbol{\beta}}\right)\right);$$

$$\Pr(\text{SP E10}_i|\text{RP E85}_i) = \Lambda\left(-\tilde{\delta}\left(\tilde{\alpha}\ln\bar{r}_i + (\ln\boldsymbol{x}_i)'\widetilde{\boldsymbol{\beta}} - \varepsilon_i\middle|\varepsilon_i \leq \tilde{\alpha}\ln r_i + (\ln\boldsymbol{x}_i)'\widetilde{\boldsymbol{\beta}}\right)\right);$$

$$\Pr(\text{SP E85}_i|\text{RP E10}_i) = \Lambda\left(\tilde{\delta}\left(\tilde{\alpha}\ln\bar{r}_i + (\ln\boldsymbol{x}_i)'\widetilde{\boldsymbol{\beta}} - \varepsilon_i\middle|\varepsilon_i \geq \tilde{\alpha}\ln r_i + (\ln\boldsymbol{x}_i)'\widetilde{\boldsymbol{\beta}}\right)\right);$$

$$\Pr(\text{SP E85}_i|\text{RP E85}_i) = \Lambda\left(\tilde{\delta}\left(\tilde{\alpha}\ln\bar{r}_i + (\ln\boldsymbol{x}_i)'\widetilde{\boldsymbol{\beta}} - \varepsilon_i\middle|\varepsilon_i \geq \tilde{\alpha}\ln r_i + (\ln\boldsymbol{x}_i)'\widetilde{\boldsymbol{\beta}}\right)\right).$$

Using the same notation as for the likelihood in the E85 premium model, the likelihood function for the price ratio model is

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

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

As in the model for the E85 premium, the ε_i's that enter the conditional SP logits are not observed, but their distribution is known so we can integrate by simulation by drawing conditional ε_i's as we did for the E85 premium model.

The parameters are estimated using maximum likelihood estimation over the joint probability WMLE function with sample and population probability weights to correct for the endogenously stratified sample. With the ratio model probability expressions, the WMLE function is the same as in equation (4.1).

4.4 Revealed and Stated Preference Data

The estimation sample consists of 881 observation pairs of motorists' RP and SP fuel choices, motorist characteristics, and actual and hypothetical prices faced. From the 893 observations used in Chapter 3, 12 observations are dropped because the SP choice question was asked incorrectly or the motorist was not able to answer. The motorist characteristics, survey method, and sample and population weights to correct for endogenous stratification are described in detail in Chapter 3. In the RP setting, 450 motorists chose E10 and 431 chose E85. Of the 450 motorists who chose E10, 87 (19%) said they would switch to E85 in the SP setting if the relative E85 price were more favorable, and of the 431 motorists who chose E85, 201 (47%) said they would switch to E10 in the SP setting if the relative E85 price were less favorable. Table 4.1 summarizes the RP and SP fuel choice data.

Table 4.1 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 either 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 either \$0.25, \$0.50, or \$0.75.

The relative E85 prices presented in the hypothetical scenarios were constructed so that one of the fuels' price was the same as in the RP setting, and the other fuel's price was altered in a way that would entice switching. Each motorist was presented one of six hypothetical scenarios chosen at random. In three of the versions, the fuel chosen was made more expensive, and in three of the versions, the fuel not chosen was made cheaper by either $0.25, $0.50, or $0.75. In this way we add substantial variation to the range of actual observed fuel prices.

Figure 4.1 shows the share of motorists who choose E85 given the premium in the RP setting, and Figure 4.2 shows the share of motorists who choose E85 given the hypothetical E85 premium offered to them. Figures 4.3 and 4.4 show the analogous RP and SP E85 shares with respect to the E85 price ratio. The figures show that in both the RP and SP settings, in general, a higher share of motorists state they would choose E85 when the hypothetical E85 price is more favorable. The figures also show that motorists who choose E10 in the RP setting are more likely to choose E10 in the SP setting even when presented with an extremely favorable relative E85 price. Likewise, motorists who choose E85 in the RP setting are more likely to choose E85 in the SP setting even when presented with an extremely unfavorable relative E85 price.

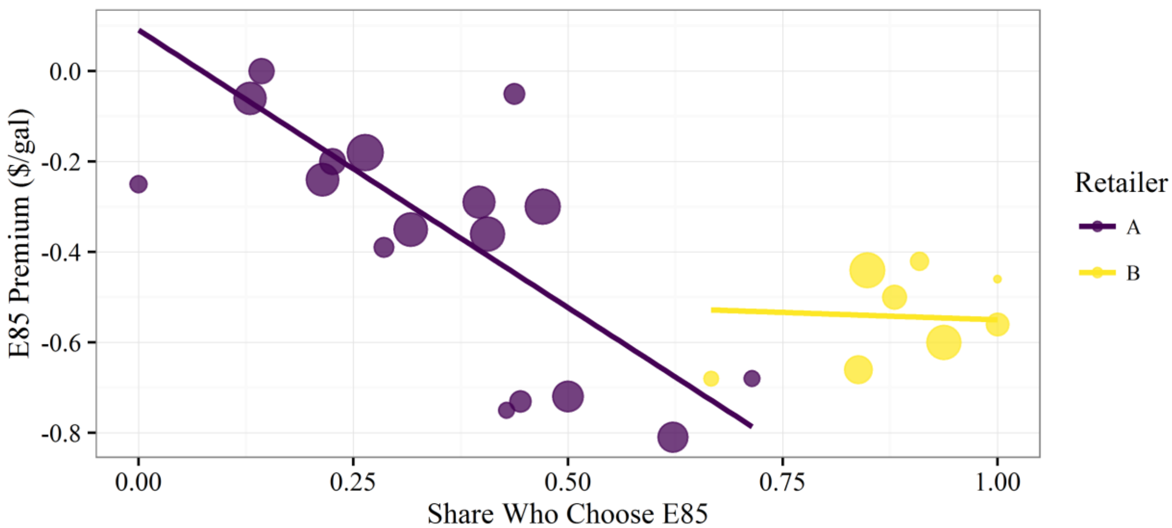

Figure 4.1 Share of motorists who choose E85 in RP setting given E85 premium

Data are observations of 881 flex motorists fueling 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.

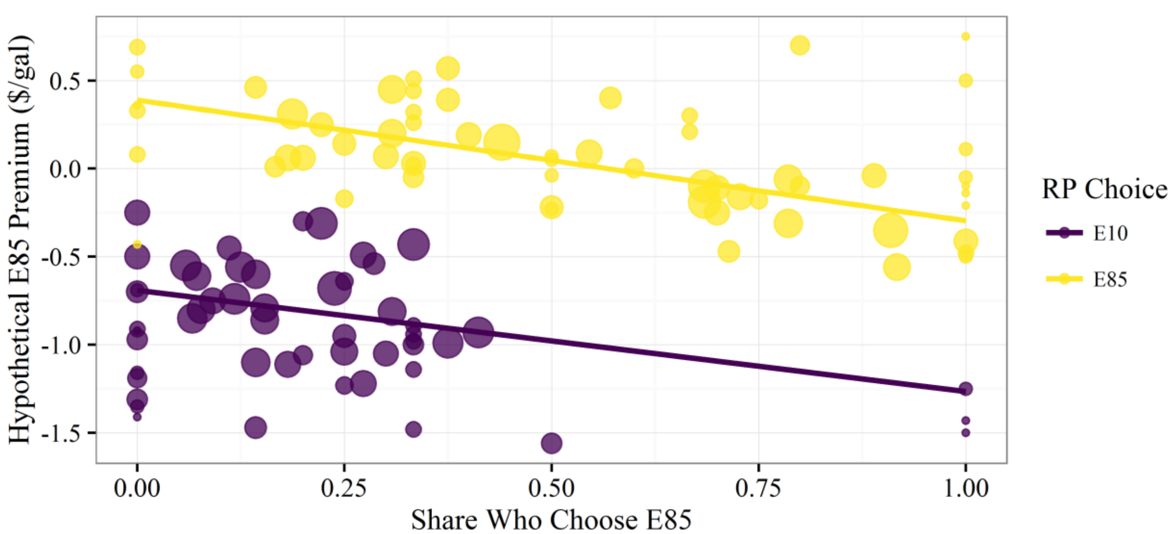

Figure 4.2 Share of motorists who choose E85 in SP setting given E85 premium

Data are responses from 881 flex motorists to a hypothetical price scenario. The motorists who chose E10 in the RP setting were offered favorable hypothetical E85 prices and motorists who chose E85 were offered favorable hypothetical E10 prices.

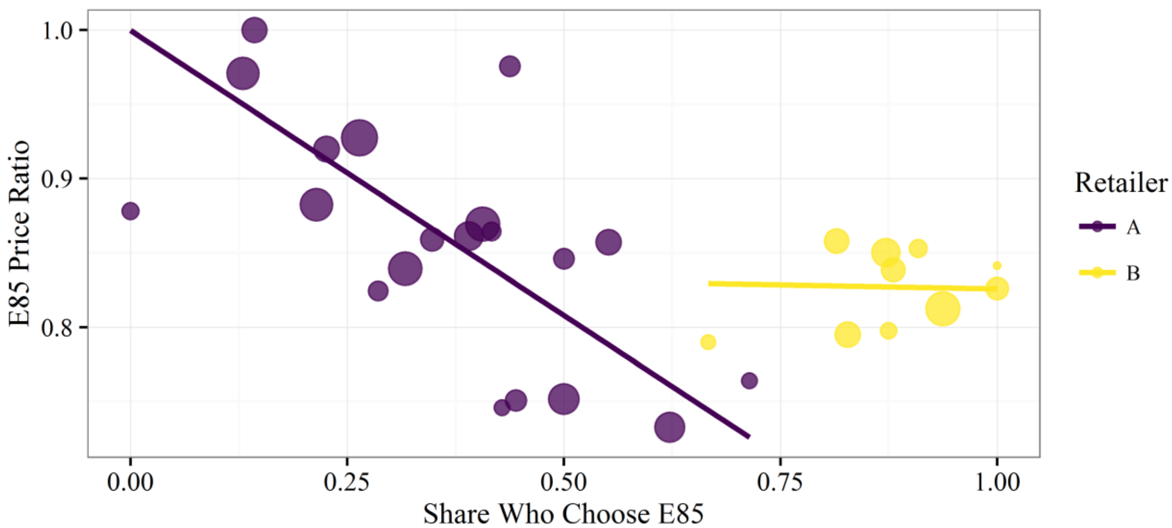

Figure 4.3 Share of motorists who choose E85 in RP setting given E85 price ratio

Data are observations of 881 flex motorists fueling 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 ratio. The lines are weighted OLS regressions.

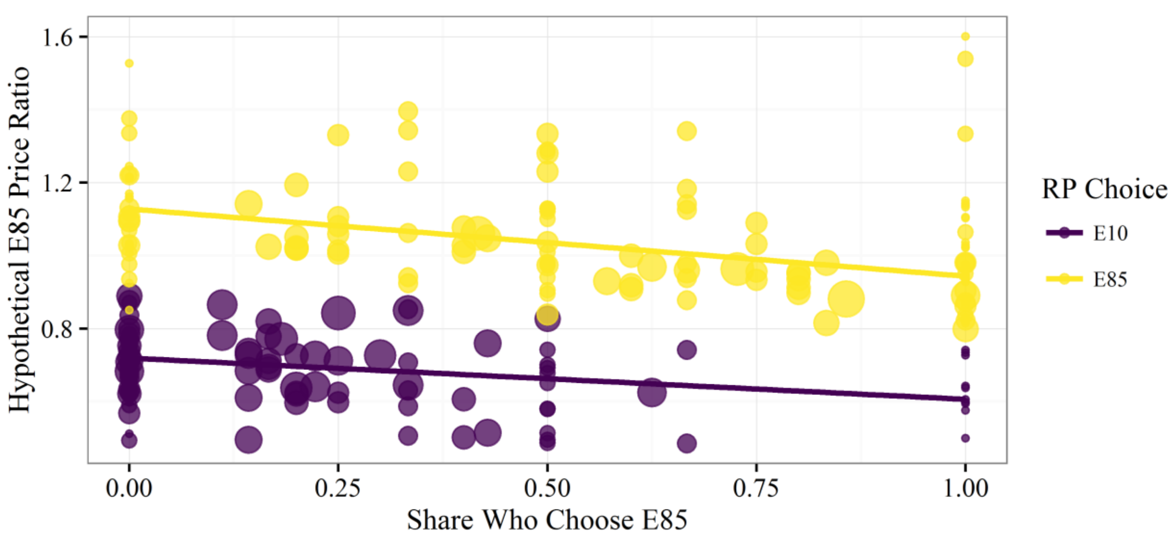

Figure 4.4 Share of motorists who choose E85 in SP setting given E85 price ratio

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 motorists who chose E85 were offered favorable E10 prices.

Table 4.1 shows that the share of E85 motorists who switch to E10 in the hypothetical scenarios is larger than the share of E10 motorists who switch to E85. There are a few possible reasons for this. One possible reason is that the E85 users were barely willing to pay for E85 in the initial RP setting whereas the E10 users were generally further away from switching. A second reason could be that flex motorists who choose E85 have a better understanding of the two fuels and treat them as near substitutes and are more responsive to price changes. Conversely, many flex motorists who choose E10 never think about E85 and are less responsive to price changes. In fact we partly impose this result with assumptions we make. In the beginning of the survey, we asked the motorists who chose E10 if their vehicle was an FFV and if they knew that the station sold E85. If the motorist responded, 'No' to either question, we did not present the hypothetical scenario. We assumed that if the motorist had arrived at the station and the relative E85 price were different, the motorist would not have noticed or cared, and would have continued to use E10.

The data indicate that preferences for ethanol are widely dispersed among flex motorists, and while some motorists are responsive to price changes and seem to view the fuels as near-substitutes, some motorists seem unwilling or unable to switch fuels regardless of the relative fuel price. In addition to the motorists with strong preferences for or against E85, some E10 users would not switch to E85 regardless of price because they do not know that their vehicle can use E85 or that the station supplies E85, and some E85 users would not switch to E10 regardless of price because they are government vehicles required to use E85.

4.5 Empirical Models and Estimation Results

We use the SP data to estimate four variations of the preferred RP-data-only WMLE models from Chapter 3: Model 3.3 for the E85 premium and Model 3.6 for the E85 price ratio. The first three variations are to view and compare bias, and the fourth variation is our preferred SP-off-RP model. In the first variation of the model, we estimate the same WMLE logit model from Chapter 3, 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 WMLE logit model again. In every variation, the probability weights on the log-likelihood function to correct for endogenous stratification are based on the motorists' RP choices. In these first two models, all of the differences between RP data and SP data are ignored, and so is the endogeneity problem prominent in our survey design where the hypothetical prices in the SP setting are correlated with the unobservable error terms in the RP setting (which are also present in the SP setting).

The third variation of the WMLE model is the traditional pooled approach to combining RP and SP data that is detailed in Section 4.3. In this variation, the RP and SP data are treated differently insofar as the unobservable error terms may have different means and variances. However, the endogeneity problem is still not addressed as the model does not account for the SP data being generated from the same motorists as the RP data where the prices of the options in the SP setting depend on the RP choices.

Finally the fourth variation is the model of Train and Wilson (2008) 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 4.3.

The estimated marginal effects and standard errors for the price variable and station region variables are shown in Table 4.2 for the E85 premium models and in Table 4.3 for the E85 price ratio models. The tables also show the estimated scale and intercept coefficients for the applicable models. Conditional on the population weights being correct, we have no specific reason to believe that the baseline Models 3.3 and 3.6 are biased. We compare the results of the SP models to the RP-only models to assess whether there is a bias from using the SP data.

Table 4.2 Marginal effects of RP and SP models with E85 premium

	Model 3.3 RP data only	Model 4.1 SP data only	Model 4.2 RP and SP data pooled	Model 4.3 RP and SP traditional	Model 4.4 SP-off-RP approach
E85 Premium	-0.248	0.061	-0.065	-0.107	-0.274
	(0.068)	(0.034)	(0.029)	(0.042)	(0.022)
Colorado Springs	0.025	-0.119	-0.070	-0.083	0.026
	(0.056)	(0.060)	(0.041)	(0.049)	(0.047)
Los Angeles	0.416	0.119	0.308	0.411	0.337
	(0.038)	(0.047)	(0.028)	(0.035)	(0.032)
Little Rock	0.031	-0.057	-0.020	-0.023	0.031
	(0.044)	(0.057)	(0.037)	(0.043)	(0.043)
Sacramento	0.348	0.043	0.244	0.318	0.288
	(0.040)	(0.052)	(0.031)	(0.036)	(0.034)
Tulsa	0.096	-0.007	0.034	0.051	0.093
	(0.041)	(0.054)	(0.034)	(0.038)	(0.037)
SP constant (τ)				-1.191	
				(0.257)	
SP scale (ζ or δ)				0.431	1.450
				(0.052)	(0.328)

The models are estimated with 881 observations and the full set of independent variables as in Chapter 3, but only marginal effects and standard errors of certain variables are displayed for clarity in comparison.

We will discuss only the results for the marginal effects for the price premium and the price ratio. Marginal effects for the other variables of the model are not sensitive to the model chosen. In the first variations of the model, labeled as Model 4.1 and Model 4.5, where we use only the SP data instead of only the RP data, the bias is so large that marginal effects for the price premium and the price ratio change sign. This is the clear result of the endogeneity problem where the hypothetical prices are correlated with the motorist's unobservable error term (which affects the decision in both the RP and SP settings).

Table 4.3 Marginal effects of RP and SP models with E85 ratio

	Model 3.6 RP data only	Model 4.5 SP data only	Model 4.6 RP and SP data pooled	Model 4.7 RP and SP traditional	Model 4.8 SP-off-RP approach
Log E85 Ratio	-0.691	0.107	-0.138	-0.330	-0.579
	(0.182)	(0.076)	(0.065)	(0.103)	(0.054)
Colorado Springs	0.032	-0.107	-0.075	-0.091	0.000
	(0.056)	(0.060)	(0.040)	(0.050)	(0.048)
Los Angeles	0.431	0.131	0.310	0.410	0.375
	(0.037)	(0.047)	(0.029)	(0.035)	(0.032)
Little Rock	0.011	-0.045	-0.027	-0.028	0.000
	(0.043)	(0.057)	(0.037)	(0.043)	(0.044)
Sacramento	0.359	0.057	0.249	0.325	0.319
	(0.040)	(0.052)	(0.031)	(0.037)	(0.033)
Tulsa	0.080	0.006	0.026	0.037	0.061
	(0.040)	(0.054)	(0.034)	(0.039)	(0.038)
SP constant ($\tilde{\tau}$)				-1.147	
				(0.252)	
SP scale ($\tilde{\zeta}$ or $\tilde{\delta}$)				0.450	1.011
				(0.058)	(0.181)

Note: The models were estimated with the full set of independent variables as in Chapter 2, but only marginal effects and standard errors of certain variables are displayed for clarity in comparison.

When we stack the SP data with the RP data and treat all the observations alike in Models 4.2 and 4.6, we find that the size of the bias diminishes somewhat, but clearly remains. The marginal effect of the E85 price variable has the proper sign, and the estimated standard error is markedly smaller: for the E85 premium model, the standard error is 0.029 compared to 0.068 in the RP-only baseline Model 3.3, and for the E85 ratio model, the standard error is 0.076

compared to 0.182 in the RP-only baseline Model 3.6. The added precision of the parameter estimates is due to the added variation in ‘observed’ fuel prices and the addition of more observations.

In Model 4.3 and Model 4.7 we use the traditional approach of pooling RP and SP data. The biases in these models are smaller still, and we know the models do not properly account for the endogeneity, although they do allow for the SP settings to have their own sets of unobservable factors and shares. The estimated SP constant which adjusts for the sample SP shares is decidedly negative for both the premium and ratio models, meaning the models estimate that motorists are less likely to choose E85 in the SP setting than in the RP setting, which is what we observe. The SP scale parameter estimate is 0.43 in the premium model and 0.45 in the ratio model, meaning the variance of the unobservable errors in the SP setting is estimated to be considerably larger than the variance of the unobservable errors in the RP setting.[11] Finally in Model 4.4 and Model 4.8, where we use our preferred approach to model the SP-off-RP data, we find that the coefficient and marginal effect estimates differ only slightly from the baseline RP-only models.

Figures 4.5 and 4.7 respectively show the estimated marginal and cumulative distribution of WTP for E85 given the E85 premium from both the preferred RP-only model, Model 3.3, as well as the preferred SP-off-RP model, Model 4.4. Figures 4.6 and 4.8 respectively show the marginal and the cumulative estimated distribution of WTP for E85 given the E85 ratio from both the preferred RP-only model, Model 3.6, as well as the preferred SP-off-RP model, Model 4.8. For the E85 premium model, the estimated scale parameter is about 1.45, meaning that although additional errors exist in the SP setting, the scale of the errors in the RP setting is 1.45

[11] The error term follows a logistic distribution, so the standard deviation of the error term is one over the scale parameter.

times larger. For the E85 ratio model, the estimated scale parameter is about 1.01. The notable gains over the RP-only baseline model are in the estimated standard errors. The standard errors of the coefficient and marginal effect estimates are about 10 to 20 percent smaller in general, and about 70 percent smaller for the price variable (0.022 compared to 0.068 for the premium model and 0.054 compared to 0.182 in the ratio model). Thus by incorporating the SP data in this way we add significantly to the precision of the estimates, especially of the effect of the fuel prices. The figures show that the estimated distribution of WTP from the SP-off-RP models is quite similar to the estimates from the RP-only models of Chapter 3.

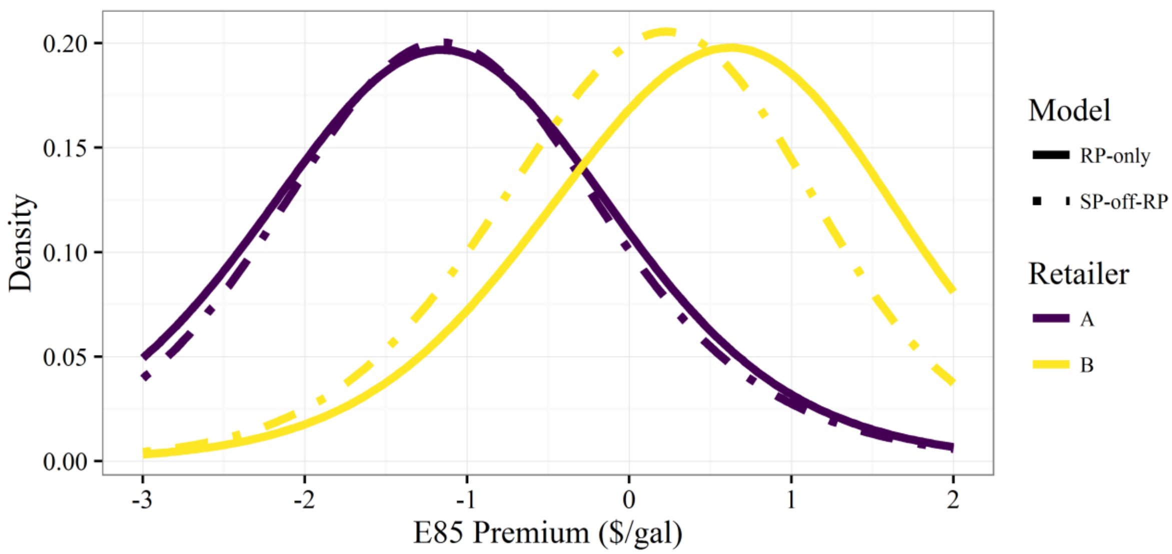

Figure 4.5 RP-only and SP-off-RP estimates of the distribution of WTP for E85 as a substitute for E10 with respect to the E85 premium (pdf)

The RP-only model is Model 3.3 and the SP-off-RP model is Model 4.4. Probabilities are calculated for each motorist at each price and then averaged over observations from each retailer. Retailer A is in the Midwest and Retailer B is in California.

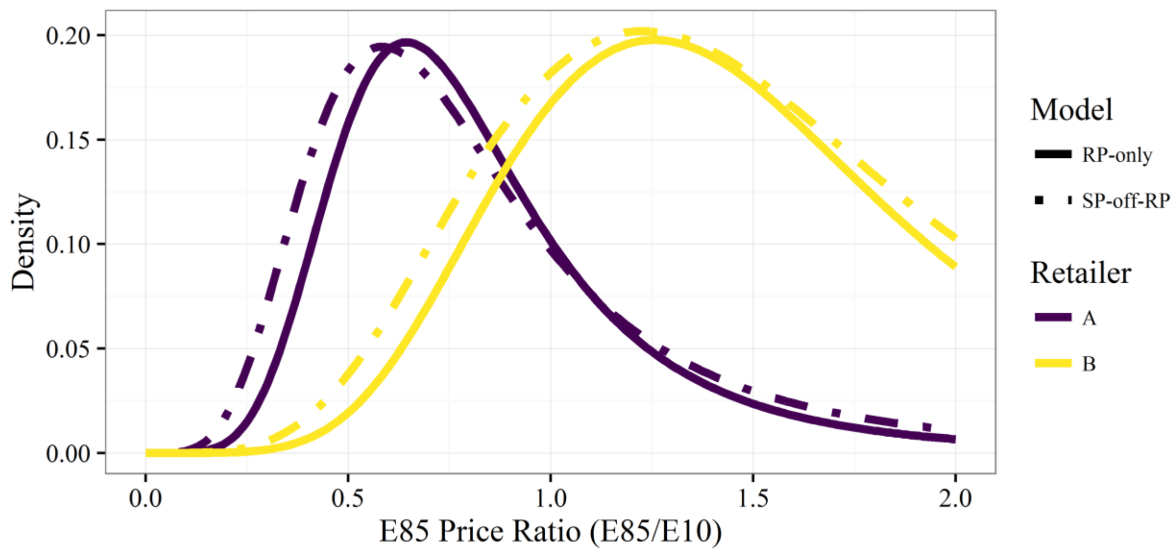

Figure 4.6 RP-only and SP-off-RP estimates of the distribution of WTP for E85 as a substitute for E10 with respect to the E85 ratio (pdf)

The RP-only model is Model 3.6 and the SP-off-RP model is Model 4.8. Probabilities are calculated for each motorist at each price and then averaged over observations from each retailer. Retailer A is in the Midwest and Retailer B is in California.

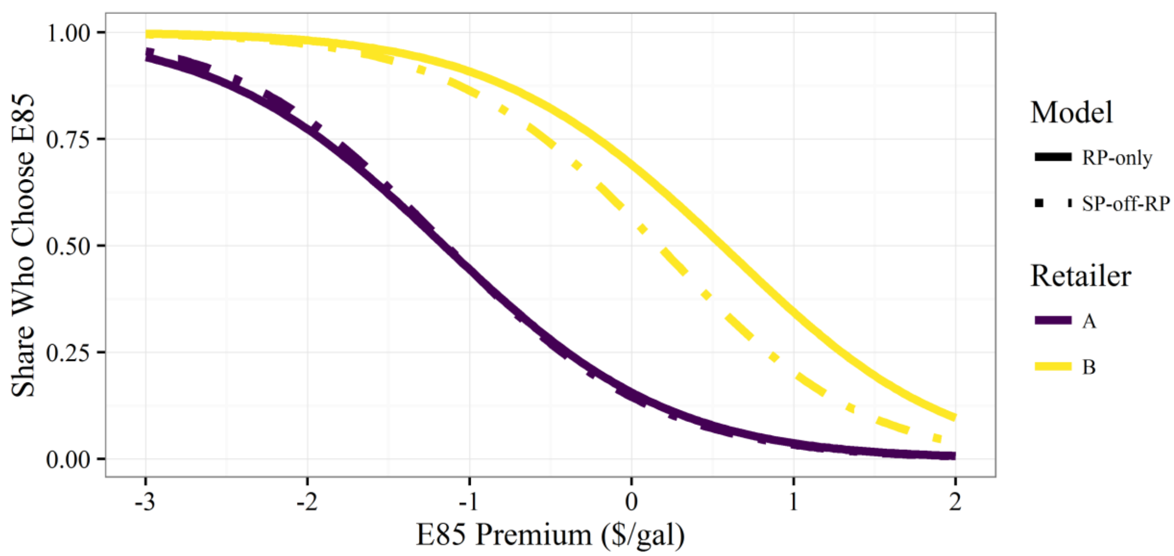

Figure 4.7 RP-only and SP-off-RP estimates of the distribution of WTP for E85 as a substitute for E10 with respect to the E85 premium (cdf)

The RP-only model is Model 3.3 and the SP-off-RP model is Model 4.4. Probabilities are calculated for each motorist at each price and then averaged over observations from each retailer. Retailer A is in the Midwest and Retailer B is in California.

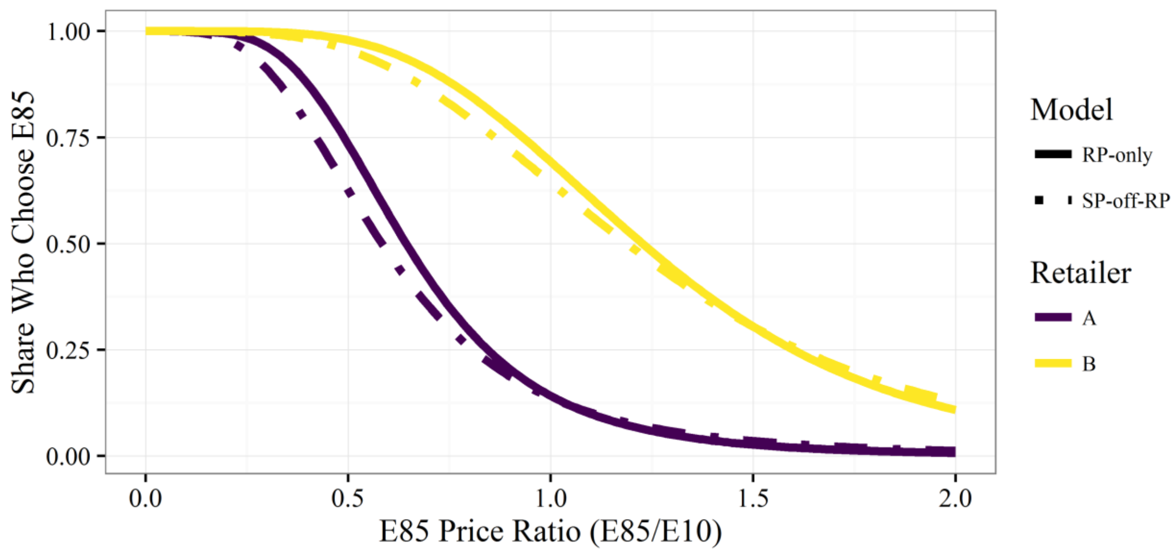

Figure 4.8 RP-only and SP-off-RP estimates of the distribution of WTP for E85 as a substitute for E10 with respect to the E85 ratio (cdf)

The RP-only model is Model 3.6 and the SP-off-RP model is Model 4.8. Probabilities are calculated for each motorist at each price and then averaged over observations from each retailer. Retailer A is in the Midwest and Retailer B is in California.

As in Chapter 3, we compare the fit of Model 4.4 and Model 4.8 to inform whether motorists make decisions based on the absolute or the relative difference in fuel prices. Incorporating the SP data allows for a wider range of relative prices where differences in decision rules might emerge. In Table 4.4 we compare the models' estimated log-likelihood values, McFadden's pseudo R-squared values, and percent correct probabilities. Recall that both models correct for the endogenous stratification to represent the general population of flex motorists rather than the sample. Applying the corrected estimates to the endogenously stratified sample results in over-prediction of motorists using E10 and under-prediction of motorists using E85. The two models fit the data about equally well, and while Model 4.4 has a slightly higher log-likelihood value, Model 4.8 correctly predicts slightly more observations. Based on these results, we still cannot say conclusively whether flex motorists base their decisions off of the absolute difference in fuel prices or the relative difference in fuel prices.

Table 4.4 Goodness of fit for SP-off-RP E85 premium model 4.4 and E85 ratio model 4.8

	Model 3.4	Model 6.4
Log-likelihood	**-868.168**	**-876.182**
Pseudo R-squared	**0.214**	**0.207**
Percent correct RP choice	**74.7%**	**74.8%**
Percent correct RP = E85	56.1%	56.1%
Percent correct RP = E10	92.4%	92.7%
Percent correct SP choice	**70.8%**	**70.4%**
Percent correct SP = E85	65.0%	67.2%
Percent correct SP = E10	74.1%	72.2%
Percent correct (RP, SP) pairs	**54.9%**	**54.5%**

Both models correct for the endogenous stratification in the sample to represent the general population. As expected, applying the estimates to the endogenously stratified sample results in relative over-prediction of motorists using E10 and under-prediction of motorists using E85.

4.6 Conclusion

Incorporating the SP data allows us to add precision to our parameter estimates when we properly model how the SP data are generated. Specifically, the unobservable factors that drive the motorists' decisions in the RP setting carry forward to the SP setting and are correlated with the hypothetical price offered in the SP setting. To correct for the endogeneity problem, we include a term in the SP utility function representing the unobservable factors from the RP setting, and we simulate its value by taking draws from its conditional distribution. The weighted log-likelihood function is the joint probability of each motorist's RP choice and the motorist's SP choice conditional on the RP choice.

When we do not account for the nature of the SP data-generating process, we observe significant biases in the estimates to the point where the sign of the price coefficients are reversed in some models. As expected, in the SP-off-RP models, the coefficient estimates are not significantly different than as in the RP-only models, but they are more precisely estimated.

The SP data feature substantially more variation in the fuel prices and accordingly the estimated standard errors for the price variable coefficients decrease by about 70 percent. We are able to more precisely estimate the effect of the relative E85 price on the probability that motorists choose E85. We estimate models where motorists choose fuels based on the E85 premium as well as models where motorists choose fuels based on the E85 price ratio. Both versions of the model fit the data quite well, and based on measures of goodness of model fit, we are not able to conclude which decision rule prevails.

The estimated distribution of preferences from the SP-off-RP models closely matches the estimated distribution of preferences from the RP-only models in Chapter 3. Using the SP-off-RP models, we find average willingness to pay of −\$1.14 in terms of the E85 premium in Retailer A's regions. Yet preferences are spread over a wide range, and 20 percent of motorists use E85 even when the premium is as high as −\$0.24, and 15 percent use E85 when the premium is zero (and E85 and E10 are the same nominal price). The results of the models using the E85 price ratio are similar. The mean WTP ratio is 0.58, about 20 percent below the energy equivalent ratio. For comparison, 20 percent of motorists use E85 in Retailer A's regions when the E85 price ratio is 0.87, and 14 percent use E85 when the price ratio is one (meaning E85 and E10 are the same nominal price). The estimated distribution of WTP from the SP-off-RP model in Retailer B's regions has a notably lower mean than the RP-only estimate; the average WTP is \$0.17 in terms of the premium and 1.19 in terms of the E85 price ratio. Overall, the estimates match closely, and increasing the E85 premium by \$0.10 decreases the probability of choosing E85 by 2.7 percent on average.

CHAPTER 5.

SUMMARY AND CONCLUSIONS

This dissertation estimates the distribution of willingness to pay for E85 as a substitute for E10 among flex motorists in the United States. The results can be used to predict the share of flex motorists who choose E85 given fuel prices and, in turn, the position of the demand curve for ethanol beyond the E10 blend wall. Estimating the demand for ethanol is a crucial piece for analysis of the biofuels mandates. The estimates presented in this dissertation can be applied to calculate RIN prices and welfare effects of RFS2.

The first study in this dissertation is an attempt to recover the distribution of motorist-level preferences from data generated by a survey of E85 stations in Minnesota. The study uses an extensive sample of recent observations from E85 stations in the Twin-Cities area. Estimates of the parameters of the WTP distribution vary substantially depending on model specification, and we cannot favor any particular estimates from the models. The conclusion is that the data are not suitable to estimate the distribution of WTP for E85.

To more accurately estimate WTP for E85 and investigate whether preferences vary across different regions of the United States, we collaborate with two E85 retailers to collect primary data from E85 stations by conducting an intercept survey. The study obtains RP data by observing actual fuel purchases and obtains SP data from responses to hypothetical choice scenarios. The SP data contribute to the range of observed fuel prices, capture the spread of fuel-switching behavior, and add precision to our parameter estimates. We use a specialized model to account for the nature of the SP-off-RP data-generating process.

The estimation sample consists of about nine hundred flex motorists refueling their vehicles at E85 stations in six urban areas in the Midwest and California: Colorado Springs, Des Moines, Little Rock, Tulsa, Los Angeles, and Sacramento. One of the E85 retailers operates the stations in the Midwest, and the other E85 retailer operates in California. Because we only survey flex motorists who fuel at E85 stations, our sample is endogenously stratified, and the probability of a motorist appearing in our sample is correlated to the motorist's WTP for E85. We apply corrective probability weights to the observations so that our estimates reflect the general population of flex motorists and not the endogenously stratified sample.

We find that an increase in the E85 premium of $0.10 decreases the probability of motorists choosing E85 by between 2.4 and 2.7 percent, on average. The estimated mean of the distribution of WTP for E85 is significantly higher for motorists in California than it is for motorists in the Midwest, but the distribution of WTP is not significantly different between the different urban areas within the Midwest. That is, given fuel prices, we do not expect the share of flex motorists who choose E85 to be significantly different in Des Moines than it is in Colorado Springs, Little Rock, or Tulsa.

We estimate models where motorists choose fuels based on the difference of the two fuel prices (the E85 premium) as well as models where motorists choose fuels based on the quotient of the two fuel prices (the E85 price ratio). Both versions of the models fit the data quite well, and, based on measures of goodness of model fit, we are not able to conclude which decision rule prevails among flex motorists.

The distribution of willingness to pay to use E85 instead of E10 is spread over a wide range of relative fuel prices. For flex motorists in the Midwest, we estimate that the mean WTP is $-\$1.14$ in terms of the E85 premium and 0.58 in terms of the E85 ratio. In California, the estimated mean WTP is \$0.17 in terms of the premium and 1.19 in terms of the price ratio, and fuel-switching behavior is similarly spread over a wide range of relative fuel prices.

From the actual observations in the sample, not weighting to correct for endogenous stratification, we observe that when the nominal price of E85 is about 80 percent of the price of E10, about 30 to 40 percent of flex motorists who fuel at E85 stations in the Midwest choose E85 while about 80 to 90 percent of flex motorists who fuel at E85 stations in California choose E85. Possible reasons for the notable difference are that E85 stations are relatively sparse in California, and each station serves a larger market of flex motorists. It could be that motorists in California are genuinely willing to pay more for E85 than motorists in the Midwest or that the E85 retailer in California effectively promotes E85 to flex motorists.

Other than the prices of the fuels and the station retailer, the significant factors affecting the probability that a motorist chooses E85 are whether the vehicle is a personal, government, or company vehicle, and motorists' opinions about which fuel is better for the environment, the engine, the economy, and which fuel yields more miles per gallon.

In conclusion, the average flex motorist in the Midwest discounts E85 relative to E10, and the average flex motorist in California puts a premium on E85 relative to E10. Motorists are diverse, and preferences are spread over a wide range of relative fuel prices. The results suggest that ethanol quantities in excess of the blend wall can be consumed in the United States through E85 if more retail fuel stations offer E85 and E85 is priced competitively with E10. The distribution of WTP for E85 is such that when E85 is priced evenly with E10 on a cost-per-mile basis, 25 percent of flex motorists in the Midwest and 75 percent of flex motorists in California choose to refuel their vehicles with E85.

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.

Baltagi, B. 2013. Econometric analysis of panel data. Chichester, UK: John Wiley & Sons Ltd.

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.

Bureau of Labor Statistics. 2014. CPI detailed report: Data for August 2014, table 24. Retrieved from http://www.bls.gov/cpi/cpid1408.pdf.

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.

Cummings, R., G. Harrison, and E. Rutström. 1995. Homegrown values and hypothetical surveys: Is the dichotomous choice approach incentive-compatible? *The American Economic Review, 85*(1), 260-266.

Department of Energy. 2008. *Handbook for handling, storing, and dispensing E85.* Publication: DOE/GO-102008-2450, National Renewable Energy Laboratory.

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

Energy Information Administration. 2013. *Gasoline prices by formulation, grade, sales type*. Retrieved from http://www.eia.gov/dnav/pet/pet_pri_allmg_d_SMN_PTA_dpgal_m.htm.

Energy Information Administration. 2014a. *Weekly retail gasoline and diesel prices*. Retrieved from http://www.eia.gov/dnav/pet/pet_pri_gnd_dcus_smn_m.htm.

Energy Information Administration. 2014b. *Refiner gasoline prices by grade and sales type*. Retrieved from http://www.eia.gov/dnav/pet/pet_pri_refmg_dcu_smn_m.htm.

Energy Information Administration. 2015. *Annual energy outlook 2015, table 12*. Retrieved from http://www.eia.gov/forecasts/aeo/tables_ref.cfm.

Houde, J. 2012. Spatial differentiation and vertical mergers in retail markets for gasoline. *The American Economic Review, 102*(5), 2147-2182.

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.

Hensher, D. A. 2010. Hypothetical bias, choice experiments and willingness to pay. *Transportation Research Part B: Methodological, 44*(6), 735-752.

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.

Liu, C. and D. Greene. 2013. Modeling the demand for E85 in the United States. Available at: http://info.ornl.gov/sites/publications/Files/Pub45616.pdf.

Louviere, J., D. Hensher, and J. Swait. 2000. Stated choice methods: analysis and applications. Cambridge University Press.

Manski, C. and S. Lerman. 1977. The estimation of choice probabilities from choice based samples. *Econometrica*, 1977-1988.

Minnesota Department of Commerce. 2014. *E85 fuel use data*. Retrieved from http://mn.gov/commerce/energy/topics/clean-energy/Vehicles-Fuels/Ethanol.jsp.

Nebraska Energy Office. 2014. *Ethanol and unleaded gasoline average rack prices.* Retrieved from http://www.neo.ne.gov/statshtml/66.html.

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. 2013. Arbitrage between ethanol and gasoline: evidence from motor fuel consumption in Brazil. Available at: http://purl.umn.edu/150964.

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. 2011. Is arbitrage tying the price of ethanol to that of gasoline? Evidence from the uptake of flexible-fuel technology. *Energy Journal, 32*(3), 119-148.

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.

United States Census Bureau. 2014. *Minnesota state and county QuickFacts.* Retrieved from http://quickfacts.census.gov/qfd/states/27000.html.

APPENDIX A: CHAPTER 2 COMPLETE ESTIMATION RESULTS

Table A.1 Model 2.1 complete results (all stations, squared premium, OLS)

Variable	Estimate	Std. Error	*t*-statistic	p-value
E85 premium	-0.8922	0.0293	-30.4286	0.0000
E85 premium squared	0.0174	0.0547	0.3180	0.7505
Log E85 stations in county	-0.1998	0.0268	-7.4499	0.0000
Second month selling E85	-0.3628	0.0481	-7.5477	0.0000
Third month selling E85	-0.2399	0.0510	-4.7002	0.0000
Fourth month selling E85	-0.1658	0.0499	-3.3210	0.0009
Month 2. February	-0.0489	0.0162	-3.0255	0.0025
Month 3. March	0.1096	0.0167	6.5717	0.0000
Month 4. April	0.1719	0.0173	9.9212	0.0000
Month 5. May	0.3341	0.0181	18.4564	0.0000
Month 6. June	0.3023	0.0190	15.8898	0.0000
Month 7. July	0.3176	0.0201	15.7941	0.0000
Month 8. August	0.2574	0.0213	12.0862	0.0000
Month 9. September	0.1247	0.0219	5.6885	0.0000
Month 10. October	0.1316	0.0232	5.6652	0.0000
Month 11. November	-0.0210	0.0245	-0.8566	0.3917
Month 12. December	-0.0629	0.0261	-2.4085	0.0161
Year 2008	-0.0866	0.0315	-2.7493	0.0060
Year 2009	-0.5330	0.0540	-9.8675	0.0000
Year 2010	-0.5476	0.0769	-7.1240	0.0000
Year 2011	-0.3173	0.1003	-3.1627	0.0016
Year 2012	-0.4403	0.1239	-3.5539	0.0004
Year 2013	-0.6683	0.1476	-4.5290	0.0000
Year 2014	-0.7795	0.1709	-4.5608	0.0000
Station 58 trend	0.0037	0.0022	1.6720	0.0946
Station 59 trend	-0.0012	0.0022	-0.5562	0.5781
⋮	⋮	⋮	⋮	⋮
Station 453 trend	-0.0007	0.0028	-0.2627	0.7928
μ	n/a	n/a		
σ	2.1260	0.2587		
R-squared	0.6680			

Data are 4,891 monthly observations from 58 Twin-Cities area E85 stations from 9/2006 to 8/2014. The dependent variable is the log of E85 sales volume. E85 volumes and prices are measured in E10 energy-equivalent terms. Station-specific fixed effects are modeled by mean-differencing the data over each station. Month and year dummies are compared to January and 2007. The parameters of the WTP distribution μ and σ are recovered from the premium coefficients.

Table A.2 Model 2.2 complete results (identified stations, squared premium, OLS)

Variable	Estimate	Std. Error	*t*-statistic	p-value
E85 premium	-0.8547	0.0276	-30.9951	0.0000
E85 premium squared	-0.0764	0.0533	-1.4340	0.1516
Log E85 stations in county	-0.2079	0.0255	-8.1532	0.0000
Second month selling E85	-0.3735	0.0447	-8.3548	0.0000
Third month selling E85	-0.2491	0.0475	-5.2485	0.0000
Fourth month selling E85	-0.1677	0.0474	-3.5385	0.0004
Month 2. February	-0.0547	0.0152	-3.5871	0.0003
Month 3. March	0.0999	0.0157	6.3546	0.0000
Month 4. April	0.1593	0.0163	9.7526	0.0000
Month 5. May	0.3219	0.0171	18.8611	0.0000
Month 6. June	0.2907	0.0180	16.1852	0.0000
Month 7. July	0.3095	0.0190	16.2711	0.0000
Month 8. August	0.2496	0.0202	12.3801	0.0000
Month 9. September	0.1172	0.0208	5.6458	0.0000
Month 10. October	0.1215	0.0220	5.5180	0.0000
Month 11. November	-0.0237	0.0232	-1.0215	0.3071
Month 12. December	-0.0655	0.0248	-2.6397	0.0083
Year 2008	-0.0865	0.0299	-2.8929	0.0038
Year 2009	-0.5361	0.0513	-10.4402	0.0000
Year 2010	-0.5525	0.0731	-7.5600	0.0000
Year 2011	-0.3212	0.0954	-3.3661	0.0008
Year 2012	-0.4256	0.1178	-3.6124	0.0003
Year 2013	-0.6380	0.1403	-4.5460	0.0000
Year 2014	-0.7618	0.1625	-4.6880	0.0000
Station 58 trend	0.0042	0.0021	2.0507	0.0404
Station 59 trend	-0.0008	0.0021	-0.3846	0.7005
⋮	⋮	⋮	⋮	⋮
Station 451 trend	0.0343	0.0027	12.8017	0.0000
Station 452 trend	0.0227	0.0027	8.4730	0.0000
μ	-1.5145	0.8787		
σ	1.7551	0.1880		
R-squared	0.7002			

Data are 4,763 monthly observations from 56 Twin-Cities area E85 stations from 9/2006 to 8/2014. The dependent variable is the log of E85 sales volume. E85 volumes and prices are measured in E10 energy-equivalent terms. Station-specific fixed effects are modeled by mean-differencing the data over each station. Month and year dummies are compared to January and 2007. The parameters of the WTP distribution μ and σ are recovered from the premium coefficients.

Table A.3 Model 2.3 complete results (all stations, squared premium, simple IV GMM)

Variable	Estimate	Std. Error	*t*-statistic	p-value
E85 premium	0.1752	0.2067	0.8476	0.3967
E85 premium squared	-3.0951	0.5545	-5.5817	0.0000
Log E85 stations in county	-0.2379	0.0601	-3.9562	0.0001
Second month selling E85	-1.4507	0.1932	-7.5090	0.0000
Third month selling E85	-0.7051	0.0969	-7.2805	0.0000
Fourth month selling E85	-0.6665	0.0992	-6.7165	0.0000
Month 2. February	-0.1985	0.0227	-8.7583	0.0000
Month 3. March	-0.0572	0.0256	-2.2378	0.0252
Month 4. April	-0.0149	0.0294	-0.5049	0.6136
Month 5. May	0.1156	0.0319	3.6263	0.0003
Month 6. June	0.1141	0.0312	3.6528	0.0003
Month 7. July	0.2451	0.0312	7.8607	0.0000
Month 8. August	0.1558	0.0323	4.8241	0.0000
Month 9. September	0.0165	0.0347	0.4749	0.6348
Month 10. October	-0.0202	0.0351	-0.5747	0.5655
Month 11. November	-0.1161	0.0386	-3.0063	0.0026
Month 12. December	-0.2325	0.0402	-5.7760	0.0000
Year 2008	-0.2837	0.0540	-5.2534	0.0000
Year 2009	-0.7932	0.0929	-8.5374	0.0000
Year 2010	-0.9899	0.1306	-7.5786	0.0000
Year 2011	-0.9942	0.1678	-5.9235	0.0000
Year 2012	-1.0015	0.2037	-4.9179	0.0000
Year 2013	-1.3580	0.2440	-5.5653	0.0000
Year 2014	-1.7671	0.2795	-6.3225	0.0000
Station 58 trend	0.0141	0.0047	2.9695	0.0030
Station 59 trend	0.0107	0.0042	2.5513	0.0107
⋮	⋮	⋮	⋮	⋮
Station 452 trend	0.0362	0.0045	8.1038	0.0000
Station 453 trend	-0.0066	0.0096	-0.6876	0.4917
μ	-0.1494	0.0885		
σ	-0.0511	0.0513		

Data are 4,891 monthly observations from 58 Twin-Cities area E85 stations from 9/2006 to 8/2014. The dependent variable is the log of E85 sales volume. E85 volumes and prices are measured in E10 energy-equivalent terms. Station-specific fixed effects are modeled by mean-differencing the data over each station. Month and year dummies are compared to January and 2007. The parameters of the WTP distribution μ and σ are recovered from the premium coefficients.

Table A.4 Model 2.4 complete results (identified stations, squared premium, simple IV GMM)

Variable	Estimate	Std. Error	*t*-statistic	p-value
E85 premium	-0.1095	0.1858	-0.5896	0.5555
E85 premium squared	-2.3356	0.4822	-4.8434	0.0000
Log E85 stations in county	-0.2210	0.0617	-3.5809	0.0003
Second month selling E85	-1.6063	0.2120	-7.5776	0.0000
Third month selling E85	-0.7981	0.1053	-7.5804	0.0000
Fourth month selling E85	-0.7475	0.1101	-6.7903	0.0000
Month 2. February	-0.1286	0.0144	-8.9330	0.0000
Month 3. March	-0.0026	0.0201	-0.1303	0.8964
Month 4. April	0.0427	0.0245	1.7405	0.0818
Month 5. May	0.1694	0.0276	6.1392	0.0000
Month 6. June	0.1550	0.0284	5.4523	0.0000
Month 7. July	0.2617	0.0286	9.1625	0.0000
Month 8. August	0.1749	0.0308	5.6837	0.0000
Month 9. September	0.0435	0.0337	1.2899	0.1971
Month 10. October	0.0006	0.0351	0.0176	0.9860
Month 11. November	-0.0926	0.0397	-2.3312	0.0197
Month 12. December	-0.1970	0.0409	-4.8128	0.0000
Year 2008	-0.3227	0.0597	-5.4058	0.0000
Year 2009	-0.8639	0.1014	-8.5222	0.0000
Year 2010	-1.0805	0.1422	-7.5963	0.0000
Year 2011	-1.1043	0.1830	-6.0339	0.0000
Year 2012	-1.1642	0.2236	-5.2060	0.0000
Year 2013	-1.5624	0.2657	-5.8814	0.0000
Year 2014	-1.9376	0.3045	-6.3641	0.0000
Station 58 trend	0.0159	0.0050	3.2063	0.0013
Station 59 trend	0.0121	0.0043	2.8025	0.0051
⋮	⋮	⋮	⋮	⋮
Station 452 trend	0.0367	0.0047	7.7974	0.0000
Station 453 trend	-0.0051	0.0098	-0.5226	0.6012
μ	0.1395	0.1785		
σ	0.0424	0.0798		

Data are 4,763 monthly observations from 56 Twin-Cities area E85 stations from 9/2006 to 8/2014. The dependent variable is the log of E85 sales volume. E85 volumes and prices are measured in E10 energy-equivalent terms. Station-specific fixed effects are modeled by mean-differencing the data over each station. Month and year dummies are compared to January and 2007. The parameters of the WTP distribution μ and σ are recovered from the premium coefficients.

Table A.5 Model 2.5 complete results (identified stations, squared premium, complex IV GMM)

Variable	Estimate	Std. Error	*t*-statistic	p-value
E85 premium	-0.2578	0.1386	-1.8598	0.0629
E85 premium squared	-2.1254	0.3611	-5.8859	0.0000
Log E85 stations in county	-0.3493	0.0552	-6.3315	0.0000
Second month selling E85	-1.6618	0.2141	-7.7610	0.0000
Third month selling E85	-0.7271	0.0871	-8.3459	0.0000
Fourth month selling E85	-0.5864	0.0771	-7.6095	0.0000
Month 2. February	-0.1263	0.0133	-9.4797	0.0000
Month 3. March	0.0026	0.0188	0.1361	0.8918
Month 4. April	0.0429	0.0228	1.8831	0.0597
Month 5. May	0.1581	0.0259	6.1059	0.0000
Month 6. June	0.1522	0.0277	5.5056	0.0000
Month 7. July	0.2403	0.0285	8.4271	0.0000
Month 8. August	0.1363	0.0312	4.3644	0.0000
Month 9. September	0.0111	0.0348	0.3199	0.7491
Month 10. October	-0.0487	0.0377	-1.2924	0.1962
Month 11. November	-0.1537	0.0431	-3.5647	0.0004
Month 12. December	-0.2516	0.0439	-5.7279	0.0000
Year 2008	-0.3715	0.0620	-5.9914	0.0000
Year 2009	-1.0035	0.1068	-9.3980	0.0000
Year 2010	-1.3002	0.1515	-8.5844	0.0000
Year 2011	-1.3802	0.1958	-7.0472	0.0000
Year 2012	-1.4721	0.2418	-6.0873	0.0000
Year 2013	-1.8901	0.2872	-6.5805	0.0000
Year 2014	-2.3476	0.3264	-7.1933	0.0000
Station 58 trend	0.0215	0.0048	4.5089	0.0000
Station 59 trend	0.0143	0.0042	3.3629	0.0008
⋮	⋮	⋮	⋮	⋮
Station 452 trend	0.0419	0.0050	8.4278	0.0000
Station 453 trend	-0.0012	0.0097	-0.1255	0.9001
μ	0.2483	0.0946		
σ	0.1083	0.0733		

Data are 4,763 monthly observations from 56 Twin-Cities area E85 stations from 9/2006 to 8/2014. The dependent variable is the log of E85 sales volume. E85 volumes and prices are measured in E10 energy-equivalent terms. Station-specific fixed effects are modeled by mean-differencing the data over each station. Month and year dummies are compared to January and 2007. The parameters of the WTP distribution μ and σ are recovered from the premium coefficients.

Table A.6 Model 2.6 complete results (identified stations, cubic premium, OLS)

Variable	Estimate	Std. Error	*t*-statistic	p-value
E85 premium	-0.8784	0.0285	-30.8696	0.0000
E85 premium squared	-0.3616	0.1011	-3.5776	0.0004
E85 premium cubed	0.4680	0.1410	3.3190	0.0009
Log E85 stations in county	-0.2056	0.0255	-8.0666	0.0000
Second month selling E85	-0.3761	0.0447	-8.4223	0.0000
Third month selling E85	-0.2505	0.0474	-5.2843	0.0000
Fourth month selling E85	-0.1685	0.0473	-3.5594	0.0004
Month 2. February	-0.0544	0.0152	-3.5740	0.0004
Month 3. March	0.1002	0.0157	6.3809	0.0000
Month 4. April	0.1619	0.0163	9.9146	0.0000
Month 5. May	0.3227	0.0171	18.9254	0.0000
Month 6. June	0.2917	0.0179	16.2580	0.0000
Month 7. July	0.3103	0.0190	16.3301	0.0000
Month 8. August	0.2495	0.0201	12.3903	0.0000
Month 9. September	0.1182	0.0207	5.7016	0.0000
Month 10. October	0.1175	0.0220	5.3339	0.0000
Month 11. November	-0.0282	0.0232	-1.2121	0.2255
Month 12. December	-0.0701	0.0248	-2.8245	0.0048
Year 2008	-0.0905	0.0299	-3.0296	0.0025
Year 2009	-0.5472	0.0514	-10.6452	0.0000
Year 2010	-0.5669	0.0731	-7.7518	0.0000
Year 2011	-0.3364	0.0954	-3.5253	0.0004
Year 2012	-0.4353	0.1177	-3.6978	0.0002
Year 2013	-0.6522	0.1403	-4.6500	0.0000
Year 2014	-0.7896	0.1625	-4.8580	0.0000
Station 58 trend	0.0031	0.0020	1.5242	0.1275
Station 59 trend	-0.0019	0.0021	-0.9266	0.3542
⋮	⋮	⋮	⋮	⋮
Station 451 trend	0.0215	0.0027	8.0091	0.0000
Station 452 trend	-0.0012	0.0027	-0.4646	0.6422
μ	-0.7377	0.1060		
σ	0.4279	0.0913		
R-squared	0.7009			

Data are 4,763 monthly observations from 56 Twin-Cities area E85 stations from 9/2006 to 8/2014. The dependent variable is the log of E85 sales volume. E85 volumes and prices are measured in E10 energy-equivalent terms. Station-specific fixed effects are modeled by mean-differencing the data over each station. Month and year dummies are compared to January and 2007. The parameters of the WTP distribution μ and σ are recovered from the premium coefficients.

Table A.7 Model 2.7 complete results (identified stations, cubic premium, complex IV GMM)

Variable	Estimate	Std. Error	*t*-statistic	p-value
E85 premium	-0.2195	0.1309	-1.6770	0.0935
E85 premium squared	-0.8250	0.9170	-0.8997	0.3683
E85 premium cubed	-1.7375	1.2516	-1.3882	0.1651
Log E85 stations in county	-0.3356	0.0552	-6.0830	0.0000
Second month selling E85	-1.6381	0.2177	-7.5230	0.0000
Third month selling E85	-0.5313	0.0637	-8.3370	0.0000
Fourth month selling E85	-0.4227	0.0611	-6.9219	0.0000
Month 2. February	-0.1273	0.0136	-9.3502	0.0000
Month 3. March	0.0127	0.0185	0.6862	0.4926
Month 4. April	0.0376	0.0231	1.6274	0.1037
Month 5. May	0.1634	0.0247	6.6199	0.0000
Month 6. June	0.1568	0.0271	5.7774	0.0000
Month 7. July	0.2346	0.0285	8.2378	0.0000
Month 8. August	0.1392	0.0311	4.4779	0.0000
Month 9. September	-0.0056	0.0356	-0.1583	0.8742
Month 10. October	-0.0245	0.0374	-0.6559	0.5119
Month 11. November	-0.1366	0.0423	-3.2315	0.0012
Month 12. December	-0.2269	0.0434	-5.2303	0.0000
Year 2008	-0.3242	0.0604	-5.3686	0.0000
Year 2009	-0.9278	0.1075	-8.6346	0.0000
Year 2010	-1.1976	0.1520	-7.8792	0.0000
Year 2011	-1.2641	0.1947	-6.4934	0.0000
Year 2012	-1.3691	0.2376	-5.7631	0.0000
Year 2013	-1.7701	0.2814	-6.2906	0.0000
Year 2014	-2.1643	0.3262	-6.6345	0.0000
Station 58 trend	0.0190	0.0048	3.9355	0.0001
Station 59 trend	0.0112	0.0045	2.4763	0.0133
⋮	⋮	⋮	⋮	⋮
Station 452 trend	0.0382	0.0045	8.4559	0.0000
Station 453 trend	-0.0044	0.0106	-0.4134	0.6793
μ	0.4786	3.6478		
σ	-0.2659	38.6005		

Data are 4,763 monthly observations from 56 Twin-Cities area E85 stations from 9/2006 to 8/2014. The dependent variable is the log of E85 sales volume. E85 volumes and prices are measured in E10 energy-equivalent terms. Station-specific fixed effects are modeled by mean-differencing the data over each station. Month and year dummies are compared to January and 2007. The parameters of the WTP distribution μ and σ are recovered from the premium coefficients.

Table A.8 Model 2.8 complete results (short-run fuel demand elasticity fixed at -0.10)

Variable	Estimate	Std. Error	*t*-statistic	p-value
Log E85 price	-0.10	n/a	n/a	n/a
E85 premium	-0.8299	0.0288	-28.8250	0.0000
E85 premium squared	-0.3620	0.1023	-3.5395	0.0004
E85 premium cubed	0.4326	0.1427	3.0326	0.0024
Log E85 stations in county	-0.2171	0.0258	-8.4182	0.0000
Second month selling E85	-0.3755	0.0452	-8.3088	0.0000
Third month selling E85	-0.2507	0.0480	-5.2262	0.0000
Fourth month selling E85	-0.1698	0.0479	-3.5453	0.0004
Month 2. February	-0.0516	0.0154	-3.3520	0.0008
Month 3. March	0.1060	0.0159	6.6757	0.0000
Month 4. April	0.1680	0.0165	10.1679	0.0000
Month 5. May	0.3330	0.0173	19.3028	0.0000
Month 6. June	0.3003	0.0182	16.5391	0.0000
Month 7. July	0.3164	0.0192	16.4569	0.0000
Month 8. August	0.2548	0.0204	12.5083	0.0000
Month 9. September	0.1235	0.0210	5.8878	0.0000
Month 10. October	0.1157	0.0223	5.1903	0.0000
Month 11. November	-0.0369	0.0235	-1.5708	0.1163
Month 12. December	-0.0833	0.0251	-3.3166	0.0009
Year 2008	-0.1002	0.0302	-3.3135	0.0009
Year 2009	-0.5957	0.0520	-11.4528	0.0000
Year 2010	-0.6141	0.0740	-8.2996	0.0000
Year 2011	-0.3780	0.0966	-3.9151	0.0001
Year 2012	-0.4964	0.1191	-4.1674	0.0000
Year 2013	-0.7262	0.1419	-5.1168	0.0000
Year 2014	-0.8821	0.1645	-5.3633	0.0000
Station 58 trend	0.0043	0.0021	2.0895	0.0367
Station 59 trend	-0.0008	0.0021	-0.3871	0.6987
⋮	⋮	⋮	⋮	⋮
Station 453 trend	-0.0001	0.0027	-0.0548	0.9563
μ	-0.7511	0.1239		
σ	0.4570	0.1027		
R-squared	0.6942			

Data are 4,763 monthly observations from 56 Twin-Cities area E85 stations from 9/2006 to 8/2014. The dependent variable is the log of E85 sales volume. E85 volumes and prices are measured in E10 energy-equivalent terms. Station-specific fixed effects are modeled by mean-differencing the data over each station. Month and year dummies are compared to January and 2007. The parameters of the WTP distribution μ and σ are recovered from the premium coefficients.

Table A.9 Model 2.9 complete results (short-run fuel demand elasticity fixed at -0.30)

Variable	Estimate	Std. Error	*t*-statistic	p-value
Log E85 price	-0.30	n/a	n/a	n/a
E85 premium	-0.1186	0.1391	-0.8527	0.3938
E85 premium squared	-1.1264	0.9817	-1.1474	0.2512
E85 premium cubed	-1.4243	1.3878	-1.0262	0.3048
Log E85 stations in county	-0.4390	0.0564	-7.7863	0.0000
Second month selling E85	-1.6485	0.2094	-7.8723	0.0000
Third month selling E85	-0.6374	0.0723	-8.8161	0.0000
Fourth month selling E85	-0.4403	0.0639	-6.8935	0.0000
Month 2. February	-0.1277	0.0137	-9.3377	0.0000
Month 3. March	0.0333	0.0186	1.7910	0.0733
Month 4. April	0.0614	0.0233	2.6316	0.0085
Month 5. May	0.2101	0.0232	9.0402	0.0000
Month 6. June	0.2050	0.0245	8.3599	0.0000
Month 7. July	0.3050	0.0247	12.3272	0.0000
Month 8. August	0.1996	0.0263	7.5852	0.0000
Month 9. September	0.0795	0.0307	2.5892	0.0096
Month 10. October	0.0347	0.0320	1.0860	0.2775
Month 11. November	-0.0569	0.0356	-1.5985	0.1099
Month 12. December	-0.1556	0.0371	-4.1947	0.0000
Year 2008	-0.2337	0.0528	-4.4250	0.0000
Year 2009	-0.8738	0.0953	-9.1669	0.0000
Year 2010	-1.0416	0.1326	-7.8583	0.0000
Year 2011	-0.9522	0.1691	-5.6313	0.0000
Year 2012	-1.0265	0.1995	-5.1446	0.0000
Year 2013	-1.3323	0.2392	-5.5705	0.0000
Year 2014	-1.6512	0.2809	-5.8788	0.0000
Station 58 trend	0.0154	0.0041	3.7175	0.0002
Station 59 trend	0.0059	0.0040	1.4900	0.1362
⋮	⋮	⋮	⋮	⋮
Station 453 trend	-0.0080	0.0088	-0.9160	0.3597
μ	-0.7720	0.1916		
σ	0.5266	0.1631		
R-squared	0.6822			

Data are 4,763 monthly observations from 56 Twin-Cities area E85 stations from 9/2006 to 8/2014. The dependent variable is the log of E85 sales volume. E85 volumes and prices are measured in E10 energy-equivalent terms. Station-specific fixed effects are modeled by mean-differencing the data over each station. Month and year dummies are compared to January and 2007. The parameters of the WTP distribution μ and σ are recovered from the premium coefficients.

Table A.10 Model 2.10 complete results (short-run fuel demand elasticity fixed at -0.50)

Variable	Estimate	Std. Error	*t*-statistic	p-value
Log E85 price	-0.50	n/a	n/a	n/a
E85 premium	-0.6360	0.0306	-20.7851	0.0000
E85 premium squared	-0.3635	0.1087	-3.3443	0.0008
E85 premium cubed	0.2913	0.1516	1.9217	0.0547
Log E85 stations in county	-0.2631	0.0274	-9.5993	0.0000
Second month selling E85	-0.3727	0.0480	-7.7615	0.0000
Third month selling E85	-0.2514	0.0510	-4.9313	0.0000
Fourth month selling E85	-0.1751	0.0509	-3.4393	0.0006
Month 2. February	-0.0405	0.0164	-2.4760	0.0133
Month 3. March	0.1295	0.0169	7.6716	0.0000
Month 4. April	0.1924	0.0176	10.9568	0.0000
Month 5. May	0.3743	0.0183	20.4158	0.0000
Month 6. June	0.3345	0.0193	17.3355	0.0000
Month 7. July	0.3409	0.0204	16.6807	0.0000
Month 8. August	0.2763	0.0217	12.7591	0.0000
Month 9. September	0.1447	0.0223	6.4917	0.0000
Month 10. October	0.1084	0.0237	4.5781	0.0000
Month 11. November	-0.0720	0.0250	-2.8814	0.0040
Month 12. December	-0.1361	0.0267	-5.0975	0.0000
Year 2008	-0.1388	0.0321	-4.3196	0.0000
Year 2009	-0.7895	0.0553	-14.2845	0.0000
Year 2010	-0.8030	0.0786	-10.2126	0.0000
Year 2011	-0.5445	0.1026	-5.3065	0.0000
Year 2012	-0.7407	0.1266	-5.8516	0.0000
Year 2013	-1.0220	0.1508	-6.7761	0.0000
Year 2014	-1.2518	0.1748	-7.1622	0.0000
Station 58 trend	0.0091	0.0022	4.1608	0.0000
Station 59 trend	0.0037	0.0023	1.6255	0.1041
⋮	⋮	⋮	⋮	⋮
Station 453 trend	0.0042	0.0029	1.4706	0.1415
μ	-0.7761	0.3037		
σ	0.6153	0.3822		
R-squared	0.6726			

Data are 4,763 monthly observations from 56 Twin-Cities area E85 stations from 9/2006 to 8/2014. The dependent variable is the log of E85 sales volume. E85 volumes and prices are measured in E10 energy-equivalent terms. Station-specific fixed effects are modeled by mean-differencing the data over each station. Month and year dummies are compared to January and 2007. The parameters of the WTP distribution μ and σ are recovered from the premium coefficients.

Table A.11 Model 2.11 complete results (short-run fuel demand elasticity estimated freely)

Variable	Estimate	Std. Error	*t*-statistic	p-value
Log E85 price	-0.10	n/a	n/a	n/a
E85 premium	0.5866	0.0327	17.9304	0.0000
E85 premium squared	-1.1628	0.0318	-36.5993	0.0000
E85 premium cubed	-0.3594	0.0978	-3.6756	0.0002
Log E85 stations in county	0.6752	0.1369	4.9321	0.0000
Second month selling E85	-0.1381	0.0249	-5.5389	0.0000
Third month selling E85	-0.3802	0.0432	-8.7984	0.0000
Fourth month selling E85	-0.2495	0.0459	-5.4399	0.0000
Month 2. February	-0.1608	0.0458	-3.5107	0.0005
Month 3. March	-0.0707	0.0148	-4.7906	0.0000
Month 4. April	0.0658	0.0153	4.2958	0.0000
Month 5. May	0.1262	0.0159	7.9211	0.0000
Month 6. June	0.2621	0.0168	15.5679	0.0000
Month 7. July	0.2416	0.0176	13.7386	0.0000
Month 8. August	0.2745	0.0185	14.8443	0.0000
Month 9. September	0.2181	0.0196	11.1501	0.0000
Month 10. October	0.0871	0.0201	4.3264	0.0000
Month 11. November	0.1281	0.0213	6.0097	0.0000
Month 12. December	0.0233	0.0227	1.0262	0.3048
Year 2008	0.0073	0.0244	0.2976	0.7660
Year 2009	-0.0339	0.0291	-1.1658	0.2437
Year 2010	-0.2629	0.0522	-5.0359	0.0000
Year 2011	-0.2898	0.0724	-4.0015	0.0001
Year 2012	-0.0923	0.0933	-0.9886	0.3229
Year 2013	-0.0770	0.1156	-0.6658	0.5056
Year 2014	-0.2184	0.1378	-1.5844	0.1132
Station 58 trend	-0.2474	0.1601	-1.5450	0.1224
Station 59 trend	-0.0039	0.0020	-1.9632	0.0497
⋮	⋮	⋮	⋮	⋮
Station 453 trend	-0.0077	0.0026	-2.9335	0.0034
μ	-0.6512	0.0424		
σ	0.3082	0.0502		
R-squared	0.7201			

Data are 4,763 monthly observations from 56 Twin-Cities area E85 stations from 9/2006 to 8/2014. The dependent variable is the log of E85 sales volume. E85 volumes and prices are measured in E10 energy-equivalent terms. Station-specific fixed effects are modeled by mean-differencing the data over each station. Month and year dummies are compared to January and 2007. The parameters of the WTP distribution μ and σ are recovered from the premium coefficients.

APPENDIX B: CHAPTER 3 COMPLETE ESTIMATION RESULTS

Table B.1 Results of Model 3.1: No sample correction, E85 premium and all variables

Variables	Coef.	SE	p-val.	ME	SE	p-val.
Intercept	**-2.931**	**0.694**	**0.000**	**-0.456**	**0.103**	**0.000**
E85 price premium	**-1.717**	**0.481**	**0.000**	**-0.267**	**0.072**	**0.000**
Government FFV	**2.335**	**0.558**	**0.000**	**0.363**	**0.083**	**0.000**
Company FFV	-0.722	0.373	0.053	-0.112	0.058	0.052
Other non-personal FFV	0.018	0.389	0.964	0.003	0.061	0.964
FFV type: truck	0.065	0.241	0.789	0.010	0.038	0.789
FFV type: SUV	-0.250	0.234	0.286	-0.039	0.036	0.284
FFV type: van	-0.084	0.303	0.782	-0.013	0.047	0.782
Badge	0.167	0.192	0.383	0.026	0.030	0.384
Female	0.117	0.201	0.560	0.018	0.031	0.559
Age	0.010	0.006	0.133	0.002	0.001	0.129
Miles per year (k)	-0.010	0.006	0.090	-0.002	0.001	0.090
E85 better for env.	-0.230	0.316	0.467	-0.036	0.049	0.466
E10 better for env.	-1.091	0.569	0.055	-0.170	0.087	0.052
No diff. for env.	-0.101	0.373	0.786	-0.016	0.058	0.786
E85 better for engine	**0.719**	**0.291**	**0.013**	**0.112**	**0.045**	**0.013**
E10 better for engine	-0.281	0.302	0.353	-0.044	0.047	0.352
No diff. for engine	0.008	0.331	0.981	0.001	0.051	0.981
E85 better for econ.	**1.042**	**0.347**	**0.003**	**0.162**	**0.053**	**0.002**
E10 better for econ.	0.111	0.390	0.776	0.017	0.061	0.776
No difference for econ.	0.796	0.421	0.059	0.124	0.065	0.058
E85 better for natl. sec.	0.240	0.244	0.324	0.037	0.038	0.322
E10 better for natl. sec.	0.147	0.344	0.669	0.023	0.054	0.668
No diff. for natl. sec.	-0.224	0.312	0.473	-0.035	0.049	0.474
E85 better mpg	**0.743**	**0.333**	**0.026**	**0.116**	**0.051**	**0.024**
E10 better mpg	0.474	0.278	0.088	0.074	0.043	0.087
No difference mpg	**0.978**	**0.400**	**0.015**	**0.152**	**0.062**	**0.014**
Colorado Springs	-0.145	0.405	0.720	-0.023	0.063	0.720
Los Angeles	**2.891**	**0.326**	**0.000**	**0.450**	**0.047**	**0.000**
Little Rock	0.180	0.308	0.558	0.028	0.048	0.557
Sacramento	**2.285**	**0.342**	**0.000**	**0.355**	**0.049**	**0.000**
Tulsa	**0.596**	**0.296**	**0.044**	**0.093**	**0.046**	**0.042**
Observations			893			
Percent correct predictions			77.492			
Log likelihood value			-423.342			
McFadden's pseudo R^2			0.316			

The E85 price premium is the nominal (not energy-adjusted) per gallon E85 price minus the E10 price. Model 3.1 does not correct for the endogenously stratified sample. Variables in bold are significant at the 5 percent level. The table shows coefficient and marginal effects estimates.

Table B.2 Results of Model 3.2: Representative subset, E85 premium and all variables

Variables	Coef.	SE	p-val.	ME	SE	p-val.
Intercept	**-4.233**	**0.798**	**0.000**	**-0.581**	**0.101**	**0.000**
E85 price premium	**-1.743**	**0.615**	**0.005**	**-0.239**	**0.083**	**0.004**
Government FFV	**2.858**	**0.597**	**0.000**	**0.392**	**0.077**	**0.000**
Company FFV	-0.216	0.428	0.614	-0.030	0.059	0.614
Other non-personal FFV	-0.126	0.526	0.811	-0.017	0.072	0.811
FFV type: truck	0.340	0.279	0.224	0.047	0.038	0.223
FFV type: SUV	-0.305	0.285	0.285	-0.042	0.039	0.281
FFV type: van	-0.241	0.403	0.550	-0.033	0.055	0.548
Badge	0.241	0.238	0.312	0.033	0.033	0.312
Female	0.256	0.247	0.301	0.035	0.034	0.297
Age	0.008	0.008	0.312	0.001	0.001	0.308
Miles per year (k)	-0.011	0.007	0.146	-0.001	0.001	0.147
E85 better for env.	-0.176	0.383	0.647	-0.024	0.053	0.647
E10 better for env.	-1.077	0.713	0.131	-0.148	0.097	0.128
No diff. for env.	-0.303	0.444	0.495	-0.042	0.061	0.494
E85 better for engine	**1.017**	**0.359**	**0.005**	**0.140**	**0.049**	**0.004**
E10 better for engine	-0.312	0.384	0.416	-0.043	0.053	0.415
No diff. for engine	0.296	0.392	0.451	0.041	0.054	0.451
E85 better for econ.	**1.163**	**0.422**	**0.006**	**0.160**	**0.058**	**0.006**
E10 better for econ.	0.135	0.492	0.784	0.019	0.068	0.784
No difference for econ.	**1.080**	**0.500**	**0.031**	**0.148**	**0.069**	**0.030**
E85 better for natl. sec.	0.250	0.289	0.387	0.034	0.040	0.385
E10 better for natl. sec.	0.360	0.425	0.397	0.049	0.058	0.397
No diff. for natl. sec.	-0.116	0.385	0.764	-0.016	0.053	0.765
E85 better mpg	0.275	0.426	0.520	0.038	0.058	0.519
E10 better mpg	**0.743**	**0.364**	**0.041**	**0.102**	**0.049**	**0.038**
No difference mpg	**1.170**	**0.483**	**0.015**	**0.161**	**0.066**	**0.014**
Colorado Springs	0.100	0.483	0.836	0.014	0.066	0.836
Los Angeles	**3.245**	**0.395**	**0.000**	**0.446**	**0.047**	**0.000**
Little Rock	0.218	0.371	0.558	0.030	0.051	0.558
Sacramento	**2.644**	**0.370**	**0.000**	**0.363**	**0.044**	**0.000**
Tulsa	**0.761**	**0.357**	**0.033**	**0.105**	**0.049**	**0.032**
Observations			681			
Percent correct predictions			80.470			
Log likelihood value			-293.219			
McFadden’s pseudo R^2			0.320			

The E85 premium is the nominal E85 price minus the E10 price. Model 3.2 corrects for the endogenous stratification by using only observations from motorists who do not drive out of their way to visit the E85 station. All dummies equal zero is a personal vehicle, the type is car, it has no FFV badge, the motorist is male and answers ‘don’t know’ to all fuel opinion questions, and the station location area is Des Moines. Bold variables are significant at the 5 percent level.

Table B.3 Results of Model 3.3: WMLE sample correction, E85 premium and all variables

Variables	Coef.	SE	p-val.	ME	SE	p-val.
Intercept	**-4.096**	**0.703**	**0.000**	**-0.562**	**0.091**	**0.000**
E85 price premium	**-1.711**	**0.498**	**0.001**	**-0.235**	**0.068**	**0.001**
Government FFV	**2.459**	**0.542**	**0.000**	**0.338**	**0.071**	**0.000**
Company FFV	**-0.962**	**0.417**	**0.021**	**-0.132**	**0.057**	**0.020**
Other non-personal FFV	-0.017	0.419	0.967	-0.002	0.057	0.967
FFV type: truck	0.073	0.242	0.763	0.010	0.033	0.763
FFV type: SUV	-0.250	0.237	0.292	-0.034	0.033	0.291
FFV type: van	-0.270	0.327	0.410	-0.037	0.045	0.410
Badge	0.210	0.202	0.299	0.029	0.028	0.300
Female	0.167	0.206	0.418	0.023	0.028	0.417
Age	0.012	0.007	0.073	0.002	0.001	0.069
Miles per year (k)	-0.009	0.006	0.133	-0.001	0.001	0.133
E85 better for env.	-0.264	0.328	0.420	-0.036	0.045	0.420
E10 better for env.	**-1.474**	**0.656**	**0.025**	**-0.202**	**0.089**	**0.023**
No diff. for env.	-0.163	0.389	0.676	-0.022	0.053	0.676
E85 better for engine	**0.944**	**0.308**	**0.002**	**0.130**	**0.042**	**0.002**
E10 better for engine	-0.211	0.325	0.517	-0.029	0.045	0.516
No diff. for engine	0.100	0.347	0.772	0.014	0.048	0.772
E85 better for econ.	**1.122**	**0.364**	**0.002**	**0.154**	**0.049**	**0.002**
E10 better for econ.	0.236	0.401	0.556	0.032	0.055	0.556
No difference for econ.	**0.932**	**0.437**	**0.033**	**0.128**	**0.060**	**0.032**
E85 better for natl. sec.	0.292	0.246	0.236	0.040	0.034	0.234
E10 better for natl. sec.	0.064	0.348	0.854	0.009	0.048	0.854
No diff. for natl. sec.	-0.176	0.314	0.576	-0.024	0.043	0.576
E85 better mpg	**0.864**	**0.346**	**0.013**	**0.119**	**0.047**	**0.012**
E10 better mpg	**0.636**	**0.296**	**0.032**	**0.087**	**0.040**	**0.031**
No difference mpg	**1.284**	**0.394**	**0.001**	**0.176**	**0.054**	**0.001**
Colorado Springs	0.151	0.410	0.712	0.021	0.056	0.713
Los Angeles	**3.058**	**0.320**	**0.000**	**0.420**	**0.038**	**0.000**
Little Rock	0.237	0.324	0.464	0.033	0.044	0.464
Sacramento	**2.574**	**0.328**	**0.000**	**0.353**	**0.040**	**0.000**
Tulsa	**0.699**	**0.304**	**0.021**	**0.096**	**0.041**	**0.020**
Observations			893			
Percent correct predictions			76.596			
Log likelihood value			-380.044			
McFadden's pseudo R^2			0.355			

The E85 premium is the nominal E85 price minus the E10 price. Model 3.3 corrects for the endogenous stratification in the sample by applying probability weights. All dummies equal zero is a personal vehicle, the type is car, it has no FFV badge, the motorist is male and answers, 'don't know' to all fuel opinion questions, and the station location area is Des Moines. Bold variables are significant at the 5 percent level.

Table B.4 Results of Model 3.4: No sample correction, E85 ratio and all variables

Variables	Coef.	SE	p-val.	ME	SE	p-val.
Intercept	**-3.542**	**1.405**	**0.012**	**-0.550**	**0.214**	**0.010**
Log E85 price ratio	**-4.756**	**1.286**	**0.000**	**-0.739**	**0.193**	**0.000**
Government FFV	**2.284**	**0.557**	**0.000**	**0.355**	**0.083**	**0.000**
Company FFV	**-0.743**	**0.371**	**0.045**	**-0.115**	**0.057**	**0.044**
Other non-personal FFV	0.002	0.391	0.997	0.000	0.061	0.997
FFV type: truck	0.073	0.241	0.762	0.011	0.037	0.762
FFV type: SUV	-0.233	0.236	0.324	-0.036	0.037	0.323
FFV type: van	-0.069	0.303	0.819	-0.011	0.047	0.819
Badge	0.170	0.192	0.375	0.026	0.030	0.376
Female	0.106	0.202	0.598	0.017	0.031	0.597
Log age	0.368	0.289	0.203	0.057	0.044	0.199
Log miles per year (k)	-0.219	0.133	0.099	-0.034	0.021	0.098
E85 better for env.	-0.223	0.318	0.483	-0.035	0.049	0.482
E10 better for env.	-1.090	0.571	0.056	-0.169	0.088	0.053
No diff. for env.	-0.108	0.376	0.773	-0.017	0.058	0.773
E85 better for engine	**0.711**	**0.291**	**0.015**	**0.110**	**0.045**	**0.014**
E10 better for engine	-0.291	0.303	0.338	-0.045	0.047	0.337
No diff. for engine	-0.002	0.331	0.996	0.000	0.051	0.996
E85 better for econ.	**1.031**	**0.347**	**0.003**	**0.160**	**0.053**	**0.003**
E10 better for econ.	0.111	0.390	0.775	0.017	0.061	0.775
No difference for econ.	0.793	0.420	0.059	0.123	0.065	0.058
E85 better for natl. sec.	0.252	0.244	0.302	0.039	0.038	0.300
E10 better for natl. sec.	0.149	0.346	0.668	0.023	0.054	0.667
No diff. for natl. sec.	-0.209	0.313	0.504	-0.032	0.049	0.505
E85 better mpg	**0.727**	**0.332**	**0.029**	**0.113**	**0.051**	**0.027**
E10 better mpg	0.471	0.279	0.091	0.073	0.043	0.090
No difference mpg	**0.982**	**0.403**	**0.015**	**0.152**	**0.062**	**0.015**
Colorado Springs	-0.097	0.409	0.812	-0.015	0.063	0.812
Los Angeles	**2.995**	**0.325**	**0.000**	**0.465**	**0.046**	**0.000**
Little Rock	0.044	0.299	0.882	0.007	0.046	0.882
Sacramento	**2.372**	**0.342**	**0.000**	**0.368**	**0.048**	**0.000**
Tulsa	**0.475**	**0.285**	**0.095**	**0.074**	**0.044**	**0.093**
Observations			893			
Percent correct predictions			77.380			
Log likelihood value			-422.975			
McFadden's pseudo R^2			0.316			

The E85 price ratio is the nominal E85 price divided by the E10 price. Model 3.4 does not correct for the endogenous stratification in the sample. All dummies equal zero is a personal vehicle, the type is car, it has no FFV badge, the motorist is male and answers, 'don't know' to all fuel opinion questions, and the station location area is Des Moines. Bold variables are significant at the 5 percent level.

Table B.5 Results of Model 3.5: Representative subset, E85 ratio and all variables

Variables	Coef.	SE	p-val.	ME	SE	p-val.
Intercept	**-4.820**	**1.600**	**0.003**	**-0.662**	**0.212**	**0.002**
Log E85 price ratio	**-4.865**	**1.644**	**0.003**	**-0.668**	**0.221**	**0.002**
Government FFV	**2.852**	**0.601**	**0.000**	**0.392**	**0.078**	**0.000**
Company FFV	-0.273	0.425	0.521	-0.037	0.058	0.521
Other non-personal FFV	-0.145	0.531	0.785	-0.020	0.073	0.785
FFV type: truck	0.343	0.279	0.219	0.047	0.038	0.218
FFV type: SUV	-0.299	0.286	0.296	-0.041	0.039	0.291
FFV type: van	-0.235	0.405	0.562	-0.032	0.055	0.560
Badge	0.244	0.238	0.305	0.033	0.033	0.305
Female	0.262	0.249	0.293	0.036	0.034	0.289
Log age	0.295	0.333	0.376	0.041	0.045	0.373
Log miles per year (k)	-0.171	0.161	0.286	-0.024	0.022	0.287
E85 better for env.	-0.158	0.382	0.678	-0.022	0.052	0.678
E10 better for env.	-1.067	0.714	0.135	-0.146	0.097	0.131
No diff. for env.	-0.305	0.442	0.490	-0.042	0.061	0.490
E85 better for engine	**0.996**	**0.360**	**0.006**	**0.137**	**0.049**	**0.005**
E10 better for engine	-0.336	0.384	0.382	-0.046	0.053	0.381
No diff. for engine	0.279	0.393	0.478	0.038	0.054	0.477
E85 better for econ.	**1.156**	**0.419**	**0.006**	**0.159**	**0.057**	**0.006**
E10 better for econ.	0.149	0.489	0.761	0.020	0.067	0.761
No difference for econ.	**1.084**	**0.496**	**0.029**	**0.149**	**0.068**	**0.029**
E85 better for natl. sec.	0.262	0.288	0.362	0.036	0.039	0.360
E10 better for natl. sec.	0.355	0.426	0.405	0.049	0.059	0.405
No diff. for natl. sec.	-0.104	0.386	0.787	-0.014	0.053	0.787
E85 better mpg	0.263	0.426	0.537	0.036	0.058	0.536
E10 better mpg	**0.740**	**0.364**	**0.042**	**0.102**	**0.049**	**0.039**
No difference mpg	**1.159**	**0.484**	**0.017**	**0.159**	**0.066**	**0.015**
Colorado Springs	0.156	0.489	0.750	0.021	0.067	0.750
Los Angeles	**3.347**	**0.393**	**0.000**	**0.460**	**0.046**	**0.000**
Little Rock	0.095	0.360	0.792	0.013	0.049	0.792
Sacramento	**2.753**	**0.369**	**0.000**	**0.378**	**0.043**	**0.000**
Tulsa	**0.655**	**0.342**	**0.055**	**0.090**	**0.047**	**0.053**
Observations			681			
Percent correct predictions			80.029			
Log likelihood value			-293.241			
McFadden’s pseudo R^2			0.320			

The E85 ratio is the nominal E85 price divided by the E10 price. Model 3.5 corrects for the endogenous stratification by using only observations from motorists who do not drive out of their way to visit the E85 station. All dummies equal zero is a personal vehicle, the type is car, it has no FFV badge, the motorist is male and answers ‘don’t know’ to all fuel opinion questions, and the station location area is Des Moines. Bold variables are significant at the 5 percent level.

Table B.6 Results of Model 3.6: WMLE sample correction, E85 ratio and all variables

Variables	Coef.	SE	p-val.	ME	SE	p-val.
Intercept	**-5.055**	**1.423**	**0.000**	**-0.694**	**0.189**	**0.000**
Log E85 price ratio	**-4.746**	**1.328**	**0.000**	**-0.651**	**0.180**	**0.000**
Government FFV	**2.416**	**0.538**	**0.000**	**0.332**	**0.071**	**0.000**
Company FFV	**-0.988**	**0.415**	**0.017**	**-0.136**	**0.056**	**0.016**
Other non-personal FFV	-0.034	0.421	0.937	-0.005	0.058	0.937
FFV type: truck	0.077	0.241	0.750	0.011	0.033	0.750
FFV type: SUV	-0.243	0.239	0.309	-0.033	0.033	0.309
FFV type: van	-0.262	0.328	0.426	-0.036	0.045	0.425
Badge	0.215	0.202	0.287	0.029	0.028	0.287
Female	0.157	0.208	0.449	0.022	0.028	0.448
Log age	0.468	0.294	0.111	0.064	0.040	0.107
Log miles per year (k)	-0.190	0.136	0.163	-0.026	0.019	0.164
E85 better for env.	-0.252	0.330	0.445	-0.035	0.045	0.445
E10 better for env.	**-1.464**	**0.655**	**0.025**	**-0.201**	**0.089**	**0.023**
No diff. for env.	-0.163	0.391	0.677	-0.022	0.054	0.677
E85 better for engine	**0.936**	**0.308**	**0.002**	**0.128**	**0.042**	**0.002**
E10 better for engine	-0.223	0.326	0.494	-0.031	0.045	0.493
No diff. for engine	0.087	0.347	0.803	0.012	0.048	0.803
E85 better for econ.	**1.109**	**0.364**	**0.002**	**0.152**	**0.049**	**0.002**
E10 better for econ.	0.241	0.401	0.548	0.033	0.055	0.548
No difference for econ.	**0.932**	**0.437**	**0.033**	**0.128**	**0.060**	**0.032**
E85 better for natl. sec.	0.306	0.247	0.216	0.042	0.034	0.214
E10 better for natl. sec.	0.055	0.349	0.875	0.008	0.048	0.875
No diff. for natl. sec.	-0.165	0.315	0.602	-0.023	0.043	0.602
E85 better mpg	**0.856**	**0.345**	**0.013**	**0.117**	**0.047**	**0.012**
E10 better mpg	**0.634**	**0.296**	**0.032**	**0.087**	**0.040**	**0.031**
No difference mpg	**1.287**	**0.395**	**0.001**	**0.177**	**0.054**	**0.001**
Colorado Springs	0.200	0.414	0.629	0.027	0.057	0.629
Los Angeles	**3.166**	**0.317**	**0.000**	**0.434**	**0.037**	**0.000**
Little Rock	0.106	0.314	0.735	0.015	0.043	0.735
Sacramento	**2.668**	**0.326**	**0.000**	**0.366**	**0.040**	**0.000**
Tulsa	**0.587**	**0.292**	**0.044**	**0.081**	**0.040**	**0.043**
Observations			893			
Percent correct predictions			76.484			
Log likelihood value			-379.803			
McFadden's pseudo R^2			0.356			

The E85 ratio is the nominal E85 price divided by the E10 price. Model 3.6 corrects for the endogenous stratification in the sample by applying probability weights. All dummies equal zero is a personal vehicle, the type is car, it has no FFV badge, the motorist is male and answers, 'don't know' to all fuel opinion questions, and the station location area is Des Moines. Bold variables are significant at the 5 percent level.

Table B.7 Results of Model 3.7: No sample correction, E85 premium and only station location variables

Variables	Coef.	SE	p-val.	ME	SE	p-val.
Intercept	**-1.252**	**0.246**	**0.000**	**-0.231**	**0.043**	**0.000**
E85 price premium	**-1.908**	**0.447**	**0.000**	**-0.352**	**0.079**	**0.000**
Colorado Springs	-0.319	0.348	0.359	-0.059	0.064	0.359
Los Angeles	**2.564**	**0.324**	**0.000**	**0.473**	**0.054**	**0.000**
Little Rock	-0.100	0.244	0.682	-0.018	0.045	0.682
Sacramento	**1.982**	**0.316**	**0.000**	**0.365**	**0.053**	**0.000**
Tulsa	0.013	0.238	0.958	0.002	0.044	0.958
Observations			893			
Percent correct predictions			72.228			
Log likelihood value			-489.291			
McFadden's pseudo R^2			0.209			

The E85 premium is the nominal E85 price minus the E10 price. Model 3.7 does not correct for the endogenous stratification in the sample. Variables in bold are significant at the 5 percent level. The table shows coefficient and marginal effects estimates. All dummies equal zero is a personal vehicle, the type is car, it has no FFV badge, the motorist is male and answers 'don't know' to all fuel opinion questions, and the station location is Des Moines.

Table B.8 Results of Model 3.8: Representative subset, E85 premium and only station location variables

Variables	Coef.	SE	p-val.	ME	SE	p-val.
Intercept	**-1.917**	**0.313**	**0.000**	**-0.316**	**0.047**	**0.000**
E85 price premium	**-1.859**	**0.552**	**0.001**	**-0.306**	**0.089**	**0.001**
Colorado Springs	-0.058	0.427	0.892	-0.010	0.070	0.892
Los Angeles	**2.432**	**0.350**	**0.000**	**0.401**	**0.051**	**0.000**
Little Rock	-0.007	0.305	0.982	-0.001	0.050	0.982
Sacramento	**2.166**	**0.343**	**0.000**	**0.357**	**0.050**	**0.000**
Tulsa	0.045	0.305	0.882	0.007	0.050	0.882
Observations			681			
Percent correct predictions			78.267			
Log likelihood value			-346.535			
McFadden's pseudo R^2			0.197			

The E85 premium is the nominal E85 price minus the E10 price. Model 3.8 corrects for the endogenous stratification by using only observations from motorists who do not drive out of their way to visit the E85 station. All dummies equal zero is a personal vehicle, the type is car, it has no FFV badge, the motorist is male and answers, 'don't know' to all fuel opinion questions, and the station location area is Des Moines. Bold variables are significant at the 5 percent level.

Table B.9 Results of Model 3.9: WMLE sample correction, E85 premium and only station location variables

Variables	Coef.	SE	p-val.	ME	SE	p-val.
Intercept	**-1.926**	**0.250**	**0.000**	**-0.320**	**0.039**	**0.000**
E85 price premium	**-1.884**	**0.455**	**0.000**	**-0.313**	**0.075**	**0.000**
Colorado Springs	-0.049	0.350	0.888	-0.008	0.058	0.888
Los Angeles	**2.417**	**0.323**	**0.000**	**0.401**	**0.044**	**0.000**
Little Rock	-0.007	0.244	0.977	-0.001	0.041	0.977
Sacramento	**2.149**	**0.318**	**0.000**	**0.357**	**0.046**	**0.000**
Tulsa	0.048	0.240	0.842	0.008	0.040	0.842
Observations			893			
Percent correct predictions			71.221			
Log likelihood value			-454.820			
McFadden's pseudo R^2			0.229			

The E85 premium is the nominal E85 price minus the E10 price. Model 3.9 corrects for the endogenous stratification in the sample by applying probability weights. All dummies equal zero is a personal vehicle, the type is car, it has no FFV badge, the motorist is male and answers, 'don't know' to all fuel opinion questions, and the station location area is Des Moines. Bold variables are significant at the 5 percent level.

Table B.10 Results of Model 3.10: No sample correction, E85 ratio and only station location variables

Variables	Coef.	SE	p-val.	ME	SE	p-val.
Intercept	**-1.341**	**0.258**	**0.000**	**-0.247**	**0.045**	**0.000**
Log E85 price ratio	**-5.237**	**1.186**	**0.000**	**-0.964**	**0.209**	**0.000**
Colorado Springs	-0.259	0.352	0.462	-0.048	0.065	0.462
Los Angeles	**2.702**	**0.323**	**0.000**	**0.497**	**0.052**	**0.000**
Little Rock	-0.254	0.237	0.285	-0.047	0.044	0.284
Sacramento	**2.085**	**0.316**	**0.000**	**0.384**	**0.053**	**0.000**
Tulsa	-0.115	0.227	0.612	-0.021	0.042	0.612
Observations			893			
Percent correct predictions			72.228			
Log likelihood value			-488.623			
McFadden's pseudo R^2			0.210			

The E85 price ratio is the nominal E85 price divided by the E10 price. Model 3.10 does not correct for the endogenous stratification in the sample. All dummies equal zero is a personal vehicle, the type is car, it has no FFV badge, the motorist is male and answers, 'don't know' to all fuel opinion questions, and the station location area is Des Moines. Bold variables are significant at the 5 percent level.

Table B.11 Results of Model 3.11: Representative subset, E85 ratio and only station location variables

Variables	Coef.	SE	p-val.	ME	SE	p-val.
Intercept	**-2.015**	**0.330**	**0.000**	**-0.332**	**0.050**	**0.000**
Log E85 price ratio	**-5.155**	**1.467**	**0.000**	**-0.848**	**0.236**	**0.000**
Colorado Springs	0.011	0.434	0.979	0.002	0.071	0.979
Los Angeles	**2.567**	**0.350**	**0.000**	**0.422**	**0.050**	**0.000**
Little Rock	-0.154	0.295	0.603	-0.025	0.049	0.602
Sacramento	**2.268**	**0.344**	**0.000**	**0.373**	**0.049**	**0.000**
Tulsa	-0.073	0.291	0.802	-0.012	0.048	0.802
Observations			681			
Percent correct predictions			80.029			
Log likelihood value			-293.241			
McFadden's pseudo R^2			0.320			

The E85 ratio is the nominal E85 price divided by the E10 price. Model 3.5 corrects for the endogenous stratification by using only observations from motorists who do not drive out of their way to visit the E85 station. All dummies equal zero is a personal vehicle, the type is car, it has no FFV badge, the motorist is male and answers 'don't know' to all fuel opinion questions, and the station location area is Des Moines. Bold variables are significant at the 5 percent level.

Table B.12 Results of Model 3.12: WMLE sample correction, E85 ratio and only station location variables

Variables	Coef.	SE	p-val.	ME	SE	p-val.
Intercept	**-2.019**	**0.261**	**0.000**	**-0.335**	**0.041**	**0.000**
Log E85 price ratio	**-5.194**	**1.199**	**0.000**	**-0.861**	**0.198**	**0.000**
Colorado Springs	0.015	0.354	0.966	0.003	0.059	0.966
Los Angeles	**2.554**	**0.321**	**0.000**	**0.424**	**0.042**	**0.000**
Little Rock	-0.158	0.237	0.506	-0.026	0.039	0.506
Sacramento	**2.255**	**0.318**	**0.000**	**0.374**	**0.046**	**0.000**
Tulsa	-0.075	0.228	0.743	-0.012	0.038	0.743
Observations			893			
Percent correct predictions			71.221			
Log likelihood value			-454.283			
McFadden's pseudo R^2			0.230			

The E85 ratio is the nominal E85 price divided by the E10 price. Model 3.6 corrects for the endogenous stratification in the sample by applying probability weights. All dummies equal zero is a personal vehicle, the type is car, it has no FFV badge, the motorist is male and answers, 'don't know' to all fuel opinion questions, and the station location area is Des Moines. Bold variables are significant at the 5 percent level.

APPENDIX C: THE INTERCEPT SURVEY

The survey uses 7 different forms, though each observation is collected entirely using one form contained on a single (double-sided) piece of paper. One of the 7 forms is a 1-page, station-level form where the interviewer can record pertinent information about the fueling station.

The next six forms are slightly different versions of the 2-page, motorist-level form. The versions are labeled A1, A2, A3, B1, B2, and B3. The forms only differ in the stated preference question (Question II). In versions with the letter A, the motorist is asked if she would still make the same fuel choice if her choice of fuel was more expensive. In versions with the letter B, the motorist is asked if she would still make the same fuel choice if the other fuel was less expensive. In versions with the number 1, the hypothetical price is \$0.25/gal different from the actual price. In versions with the number 2, the hypothetical price is \$0.50/gal different from the actual price. In versions with the number 3, the hypothetical price is \$0.75/gal different from the actual price. To summarize, the stated preference question asks if the motorist would still make the same choice if:

Version	1	2	3
A	The price of the fuel chosen was \$0.25/gal higher	The price of the fuel chosen was \$0.50/gal higher	The price of the fuel chosen was \$0.75/gal higher
B	The price of the fuel not chosen was \$0.25/gal lower	The price of the fuel not chosen was \$0.50/gal lower	The price of the fuel not chosen was \$0.75/gal lower

Instructions to the Interviewer: The motorist-level forms are completed in three stages, and there are three parts to the form that coincide with these stages. The first part of the form can (and should) be completed while you are waiting for a flex-fuel vehicle to pull alongside one of the station’s pumps. This part requires recording the fuel prices and performing addition or subtraction so that you are able to generate the appropriate stated-preference question (Question II) quickly and accurately once you observe the motorist’s fuel choice.

Fill out part 2 of the form while the motorist is preparing to fuel. Make sure to note the motorist’s fuel choice. If the motorist chooses E85, the hypothetical alternative fuel in Question II should be the least expensive gasoline option (i.e., regular grade). Remember to record the volume of fuel purchased and the expenditure once the motorist has finished.

Survey Form: Station-level

Instructions to the Interviewer: Fill out this form once for each station visit. Answer questions 1-11 upon arriving at the station, and answer question 12 when you conclude the visit.

1. Date and start time of visit: ___________________________
2. Interviewer name: ___________________________
3. Station name and brand: ___________________________
4. Station address: ___________________________
5. Initial per gallon E85 price: ___________________________
6. Gas option 1 – grade, ethanol %, and price: ___________________________
7. Gas option 2 – grade, ethanol %, and price: ___________________________
8. Gas option 3 – grade, ethanol %, and price: ___________________________
9. Number of gasoline nozzles: ___________________________
10. Number of E85 nozzles: ___________________________
11. Presence of E85 price signage ___________________________
12. Date and end time of visit: ___________________________

Before You Begin: Each station visit is assigned a 7-digit code for bookkeeping. The code is generated by concatenating today's date (MMDD) followed by your initials (First, Last) followed by the number of stations you have visited today. For example, if the date is October 15 (1015), your name is Kenneth Liao (KL), and this is the second station you have visited today (2), then the code would be, "1015KL2".

Write the 7-digit code for this station visit: ___________________________

You must write this code on each of the motorist forms you complete during this station visit.

When you are ready to begin, target the next FFV to pull alongside any of the station's pumps. When you finish one survey, target the next FFV to pull alongside any of the station's pumps. Do not survey flex motorists who are already at a pump when you arrive, and do not survey flex motorists who pull alongside a pump while you are surveying someone else. There are six versions of the motorist-level form: A1, A2, A3, B1, B2, and B3. Pick one version at random to start, and then proceed to use each version in sequence and repeat.

Write other notes (if any) about the station visit here:

Survey Form Code: M-A1 **Ref/Time:**

Part 1: (*Fill out this table while waiting for a flex-fuel vehicle to pull alongside one of the station's pumps.*)

7-digit Station-Visit Code		E85 Price	Gas 1 Price	Gas 2 Price
	Actual Prices:	Box 1	Box 2	Box 3
	Hypothetical Prices: (Add $0.25)	(Box 1 + $0.25) 4	(Box 2 + $0.25) 5	(Box 3 + $0.25) 6

Part 2: (*Fill out this table while the motorist is preparing to fuel and/or after the motorist has finished.*)

Vehicle Type	Vehicle Make	Vehicle Model	LP State	FFV Badge	Yellow Gas Cap	Motorist Sex	Volume & Expenditure
Sedan / Truck SUV / Van				Y / N	Y / N	M / F	

Fuel Choice
E85 / Gas

Part 3: (*Fill out this part of the form with assistance from the motorist.*)

"Hi, I am doing research for Iowa State University, and I am interested in your opinion on the different fuels. I have a few short questions to ask you while you are fueling, will you help me by answering?"

"Great! Are you 18 or older?" (If 'No' then STOP) (Yes) (No)

I. Is this your personal vehicle? (Yes) (No) _______________

(If company car) Are you: (a) financially responsible for your fuel choice or (b) fully reimbursed regardless?

Only ask these questions if the motorist did *NOT* choose E85:

a. Is your vehicle a flex-fuel vehicle capable of using E85? (Yes) (No) (Don't know)

b. (If 'Yes' to Q1) Have you ever fueled this vehicle with E85? (Yes) (No) (Don't know)

c. Did you know that this station supplies E85 fuel? (Yes) (No)

Only ask these questions if the motorist *DID* choose E85:

d. Did you choose to fuel at this station because it offers E85? (Yes) (No)

e. (If 'Yes' to Q4) How far out of your way did you have to drive? (minutes or miles) ______________________

Ask this question to all motorists: (*Use the values from Parts 1 and 2 to generate this question*.)

II. If the price of (*fuel chosen*) __________ had been (*$0.25/gal more expensive*) __________, would you still have purchased (*fuel chosen*) __________? (Yes) (No)

Ask these questions to all motorists

III. How many times do you fuel per month? ________________________

IV. (FFVs*) Out of those, how many times do you use E85? ________________________

V. On average, how many miles do you drive per year? ________________________

VI. How old are you? ________________________

"Thanks, we're almost done. For these last questions, please answer, '**Ethanol**', '**Gasoline**', or '**No Difference**'."

VII.	Which fuel is better for the environment?	(Eth)	(Gas)	(ND)	(DK)
VIII.	Which fuel is better for your engine?	(Eth)	(Gas)	(ND)	(DK)
IX.	Which fuel is better for the economy?	(Eth)	(Gas)	(ND)	(DK)
X.	Which fuel is better for national security?	(Eth)	(Gas)	(ND)	(DK)
XI.	Which fuel yields more miles per gallon?	(Eth)	(Gas)	(ND)	(DK)

i. What percentage of the miles per gallon that you get from gas do you get from E85? _____% (DK)

ii. What percentage of the miles per gallon that you get from E85 do you get from gas? _____% (DK)

"Thank you for your participation. Have a nice day."

APPENDIX D: INSTITUTIONAL REVIEW BOARD APPROVAL

IOWA STATE UNIVERSITY
OF SCIENCE AND TECHNOLOGY

Institutional Review Board
Office for Responsible Research
Vice President for Research
1138 Pearson Hall
Ames, Iowa 50011-2207
515 294-4566
FAX 515 294-4267

Date: 10/7/2014

To: Dr. Bruce Babcock, 468H Heady

CC: Sebastien Pouliot, 468G Heady

From: Office for Responsible Research

Title: Estimate the Willingness to Pay for E85 by US Consumers in Different Regions

IRB ID: 14-493

Study Review Date: 10/7/2014

The project referenced above has been declared exempt from the requirements of the human subject protections regulations as described in 45 CFR 46.101(b) because it meets the following federal requirements for exemption:

- (2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey or interview procedures with adults or observation of public behavior where
  - Information obtained is recorded in such a manner that human subjects cannot be identified directly or through identifiers linked to the subjects; or
  - Any disclosure of the human subjects' responses outside the research could not reasonably place the subject at risk of criminal or civil liability or be damaging to their financial standing, employability, or reputation.

The determination of exemption means that:

- **You do not need to submit an application for annual continuing review.**
- **You must carry out the research as described in the IRB application.** Review by IRB staff is required prior to implementing modifications that may change the exempt status of the research. In general, review is required for any modifications to the research procedures (e.g., method of data collection, nature or scope of information to be collected, changes in confidentiality measures, etc.), modifications that result in the inclusion of participants from vulnerable populations, and/or any change that may increase the risk or discomfort to participants. Changes to key personnel must also be approved. The purpose of review is to determine if the project still meets the federal criteria for exemption.

  Non-exempt research is subject to many regulatory requirements that must be addressed prior to implementation of the study. Conducting non-exempt research without IRB review and approval may constitute non-compliance with federal regulations and/or academic misconduct according to ISU policy.

  Detailed information about requirements for submission of modifications can be found on the Exempt Study Modification Form. A Personnel Change Form may be submitted when the only modification involves changes in study staff. If it is determined that exemption is no longer warranted, then an Application for Approval of Research Involving Humans Form will need to be submitted and approved before proceeding with data collection.

Please note that you must submit all research involving human participants for review. **Only the IRB or designees may make the determination of exemption**, even if you conduct a study in the future that is exactly like this study.

Please be aware that **approval from other entities may also be needed.** For example, access to data from private records (e.g. student, medical, or employment records, etc.) that are protected by FERPA, HIPAA, or other confidentiality policies requires permission from the holders of those records. Similarly, for research conducted in institutions other than ISU (e.g., schools, other colleges or universities, medical facilities, companies, etc.), investigators must obtain permission from the institution(s) as required by their policies. **An IRB determination of exemption in no way implies or guarantees that permission from these other entities will be granted.**

Please don't hesitate to contact us if you have questions or concerns at 515-294-4566 or IRB@iastate.edu.