FUEL PRICES, REGULATIONS AND THE ADOPTION OF FUEL-SAVING TECHNOLOGIES AND ALTERNATIVE FUELS

DUONG HAI LONG

(B.Soc.Sci (Hons.), B.Eng (Hons.), NUS)

A THESIS SUBMITTED
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY
DEPARTMENT OF ECONOMICS
NATIONAL UNIVERSITY OF SINGAPORE

2017

Supervisor:
Associate Professor Alberto Salvo

Examiners:
Dr Timothy Wong Chong Ji
Associate Professor Chu Junhong
Assistant Professor Ken Onishi, Singapore Management University

Declaration

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Signed: Duong Hai Long

Date: **19 Jul, 2017**

Acknowledgements

> If I have seen further, it is by standing on the shoulders of giants.
>
> *Sir Isaac Newton*

If by end of this thesis, I could see any inch further than I had before, it is all thanks to the enormous support that I have been constantly receiving from my advisor, my thesis committee members, my faculty, my PhD cohort-mates, my friends and family.

First of all, it is with immense gratitude that I acknowledge the guidance and support of my advisor, Professor Alberto Salvo. His enthusiasm, patience, knowledge and inspiration for research have encouraged me and helped me tremendously during and beyond the working of this thesis. His expertise in applied economics has constantly challenged my thinking, improved my research skills and prepared me for future endeavour. I would never imagine having a better advisor for my PhD study.

I am also grateful for my thesis committee members, Professor Yang Nan and Professor John Ham, who have given me numerous insightful comments and helpful advices, improving not only this thesis, but also my thinking and skills in general.

I would also like to thank all my classmates and friends, seniors and juniors, from PhD room 1 and room 2, for all the struggles we have gone through

together and all the support that we have given each other during this long and winding road.

Last but not the least, I owe my deepest gratitude to my family, especially my parents, for their selfless love and endless support for me. This thesis is dedicated to them.

Contents

Summary

This thesis consists of three independent chapters: one on fuel economy in the U.S. Automobile industry and two on choices between alternative fuel and fossil fuel at fuelling stations in Brazil[1].

The first chapter investigates the influence of fuel-price and fuel-economy regulation on the adoption of fuel-saving technologies in new light-duty vehicles sold in the U.S. This question will be addressed with an empirical model of the U.S. automobile industry that incorporates demand, supply, and regulations. The novelty of this model is that it allows for endogenous choices of vehicle fuel efficiency by firms in the form of a fuel-efficiency frontier. The counterfactual studies show that both fuel prices and regulations help explain the recent acceleration of fuel economy, although the merits of each factor depend on how many fuel technologies are available to each vehicle to further improve its fuel performance.

The second chapter investigates the effect of price salience on consumer fuel choice at the pump under the working hypothesis that consumers do not always possess accurate information about prices. Utilizing data gathered from an experiment conducted in four Brazilian cities, which involved the

[1]The later two chapters use a dataset that was generously provided by my advisor, Professor Alberto Salvo

showing of salient price information to random consumers at fueling stations, this study reveals that the observed consumer choices were consistent with a model of imperfect price information. Treatments that raise price salience can be effective towards reducing such price noise. I then examine the extent this imperfect information contributes towards explaining the puzzle why Brazilian consumers often choose the more expensive fuel even when the cheaper alternative is equally accessible.

The third chapter develops a discrete-continuous choice model to study the choice between different fuels for driving in Brazil, accounting for consumer heterogeneity along both margins. The model accommodates the preference for fixed payment that was observed from the data, i.e. the fact that an abnormally large proportion of drivers chose to purchase exactly 50 reais worth of fuel. The results show that price salience can increase the quantity of purchase and affect the choice between fuels, and that the type of treatment is important to achieve this effect.

List of Figures

List of Tables

Chapter 1

The Adoption of Fuel-saving Technologies in U.S. Automobile Industry: Regulation Push and Demand Pull

1.1 Introduction

The transportation sector is a major consumer of energy and contributor of air pollutants in the U.S. In 2013, its energy consumption and greenhouse gas emission made up 27% of the U.S. total (Environmental Protection Agency, 2015). As a result, it has always been on the list of primary concerns for policy-makers regarding energy security and environmental protection. Improving the fuel efficiency of new vehicles is an important part of the U.S. strategy to reduce oil dependency, improve air quality, and slow down climate change.

From 2000 to 2015, the average fuel efficiency of new passenger cars sold in the U.S. increased by 31%, from 28.5 miles per gallon (mpg)[1,2] to 37.24 mpg. Newly sold light-duty trucks also had a similar increase over the same period, from 21.3 mpg to 27 mpg. This was in stark contrast with the situation in the preceding decade, during which the U.S. automobile industry only made meagre improvements in the fuel efficiency of its vehicle fleets (in 1990, the fuel efficiency was 28 mpg for passenger cars and 20.8 mpg for light-duty trucks).

To understand this trend, two crucial factors need to be examined: regulations and fuel prices, not only because they are two of the key incentives for manufacturers and consumers to adopt fuel-efficient vehicles, but also because both have experienced recent changes, the timings of which coincided with the turn of the previously mentioned trend in fuel efficiency. These changes are illustrated in Figure B.1. After being frozen for most of the 1990s and the early years of the 2000s, fuel-economy standards started to increase again after 2005, for trucks, and 2011, for cars. Fuel prices, due to the situation in the Middle East post-9/11 and other supply and demand factors, have fluctuated wildly since the start of the millennium. Understanding how much these movements contributed to the increase in vehicle fuel efficiency is an interesting exercise on its own. Even more importantly, however, it has great practical implica-

[1] *Mpg* measures the number of miles over which the vehicle can travel on one gallon of fuel.

[2] The average reported here is the harmonic average weighted by sales.

tions for future industry performance. Given that gasoline prices have started to drop substantially since 2014 after reaching their peak in 2012, it is important to assess whether the past performance of improving fuel efficiency can still continue under the new circumstances and the ever-increasing regulatory targets.

The ability of the industry to reduce fuel consumption hinges on its capacity to develop and adopt fuel-saving technologies. The previous empirical literature on the automobile industry often assumes exogeneity of the product characteristics space, including fuel efficiency (Berry et al., 1995; Petrin, 2002). To analyse fuel-saving technology adoption, there is a need to endogenize the choice of product characteristic, especially the fuel efficiency of the vehicles. Several recent papers have attempted to do so, such as those by Klier and Linn (2012); Gramlich (2010); Zhou (2016).

This paper assesses the impact of regulatory changes and fuel-price fluctuation on fuel efficiency in the U.S. automobile industry. A model of a differentiated-good oligopoly that allows for automakers endogenous adoption of ready-for-production fuel-saving technologies is structurally estimated with aggregate market data. Counterfactual studies are performed to estimate what would have happened if fuel prices and/or regulations had been kept unchanged, to separately identify the effect of each factor.

The primary contribution of the paper is the development of a structural model of the automobile industry with endogenous choice of vehicle fuel efficiency. In addition to making the model realistic, endogenizing fuel efficiency allows the estimation of the cost of improving fuel efficiency, with which we can conduct various counterfactual analyses quantifying the effects of changes in fuel prices or regulations, or even the interaction of both, on the firms, the consumers, and the environment.

In addition, the paper contributes to the literature on the effects of the Corporate Average Fuel Economy (CAFE) standards on the automobile industry. The model estimates a compliance cost of 230 USD per mpg-vehicle, which is comparable to estimates in previous structural studies[3,4].

This paper also extends the empirical results of the literature on the effects of fuel prices in the automobile markets. The findings here indicate that fuel costs have significant and substantial negative effects on demand for new vehicles, which is consistent with the findings from existing works, which are that consumers care about fuel prices.

The rest of the paper is structured as follows. Section 2 introduces some background information regarding regulations, fuel prices, fuel-saving tech-

[3]For example Jacobsen (2013) and Gramlich (2010)

[4]However, Anderson and Sallee (2011), using a loophole in the standards related to flexible fuel vehicles, estimated a much lower compliance cost in the range of $9 to $27.

nologies, and related literature. The structural model is specified in the next section (3), after which the estimation, identification, and estimation results are detailed (Section 4). The counterfactual studies are then described and the results reported (Section 5), after which the final conclusions are presented (Section 6).

1.2 Background

1.2.1 Regulation

There are two federal government agencies in the U.S. that regulate vehicle fuel efficiency, namely the National Highway Traffic Safety Agency (NHTSA), which oversees vehicle fuel consumption, and the Environmental Protection Agency (EPA), which monitors the emission of greenhouse gases and other pollutants. This paper will focus on the regulations from the NHTSA[5].

The NHTSA regulates vehicle fuel economy using the Corporate Average Fuel Economy (CAFE) standards, which were enacted by Congress in the 1975 Energy Policy and Conservation Act (EPCA) and amended in the 2007 Energy Independence and Security Act (EISA). The standards dictate a set of thresholds for the average fuel economy that all the fleets of new vehicles

[5]The regulations by the EPA are rarely violated, which may be because either the regulations are too loose or the cost of violation is too high. Lack of variation in compliance status makes identification difficult, and if the regulations are too loose it will not affect the firms decisions anyway

sold in the U.S. must meet or exceed. Failure to comply with the standards will incur a fine of 55 USD for each mpg below the standard per vehicle[6,7].

Before 2011, there were two separate sets of standards for cars and for trucks. In 2011, each model was assigned its own standard, which was set based on the vehicle footprint (wheelbase multiplied by track width). There is also a system of CAFE credit banking that allows manufacturers who exceed the standards to use the excessive mpg to offset any deficit from other fleets, or from future deficit. Limited borrowing from future credit is also allowed, and trading of CAFE credits between manufacturers has also been allowed since 2011, although the amount that can be traded is limited.

The standards for cars stayed unchanged at 27.5 mpg for two decades, and only have started to increase since 2011. The standards for trucks started to move earlier, rising since 2005 after a 10-year freeze at 20.7 mpg. Under the EISA and the Obama Administration directive in 2009, the two standards are to be increased even more in the future, and are expected to reach 40.3 to 41.0 mpg in 2021 and 48.7 to 49.7 mpg in 2025.

[6] The average used in CAFE calculation is the harmonic average, i.e. the inverse of the production-weighted average of the inverse

[7] There may also be other unobserved cost of violation such as reputation cost, political cost

1.2.2 Fuel Prices

Purchasers of automobiles care about fuel prices. Busse et al. (2013) found that consumers are forward looking regarding fuel cost, and that a $1 increase in the price of gasoline is associated with a $250 decrease in the prices of new cars in the lowest fuel-economy quartile and a $104 increase in the prices of new cars in the highest quartile. Klier and Linn (2010)) found that the gasoline price explains nearly half of the loss of market share of U.S. manufacturers from 2002 to 2007.

Since 2000, fuel prices have been increasingly volatile, in contrast to the flat and slightly downward trend during the 1990s. They rose quickly from 2000 to 2012, with only a temporary drop in 2008 during the Great Recession (Figure B.1). However, since 2013, the trend appears to have been reversed again, reaching a new low point, the lowest in more than a decade.

1.2.3 Fuel Saving Technologies

Knittel (2011) made two observations regarding vehicle fuel efficiency: first, there is a trade-off between fuel efficiency and performance, and second, if weight, power, and torque had been kept at the same level, fuel efficiency could have improved by nearly 60% from 1980 to 2006. Figure B.2 illustrates these two observations. At a given point in time, the scatter-plots between fuel efficiency and performance attributes such as power, torque, weight, or size are all downward sloping, suggesting trade-offs between the two variables. However, from 2006 to 2014, the curves move outward, indicating that, at the same level of performances, fuel efficiency improved over the period.

The improvement in fuel efficiency is due to various fuel-saving technological developments and adoptions over the years. Table A.1, which is taken from the National Academy of Sciences (2011), lists some of the technologies, together with the estimated cost. The adoption of these technologies is a lengthy process dependent on both the demand and supply factors of the industry. Figure B.3 plots the evolution of the adoption rates of some popular technologies. It takes a decade or more for most technologies to achieve majority adoption by the market.

1.2.4 Related Work

This paper contributes to a large body of empirical literature studying the U.S. automobile industry. The impact of gasoline prices on vehicle prices, market shares, and fuel efficiency has been studied by Pakes et al. (1993); Bento et al. (2009); Busse et al. (2013); Li et al. (2009), among others. Klier and Linn (2016) studied the effect of the CAFE standard on horsepower and torque, while Jacobsen (2013) emphasizes consumer and producer heterogeneity in studying the distributional effects of CAFE standards.

This paper is also particularly aligned with a recent trend within the above literature of emphasizing the need to consider endogenous product choice in the automobile industry. Gramlich (2010) assumed a trade-off between fuel efficiency and vehicle quality when studying the effect of gasoline prices on fuel economy. Klier and Linn (2012) studied the medium-run effects of the CAFE standard, making use of engine platformbased instruments to correct for product choice endogeneity. Zhou (2016) studied the effect of R&D and gasoline tax policies on knowledge capital and technology adoption in the automobile industry, using longer-run characteristics and grandfathered technologies as instruments for product choice endogeneity.

On the technical side, the two-stage model presented in this paper loosely follows the work of Eizenberg (2014), using Berry et al. (1995)'s specifications in the second stage and a Nash equilibrium in the first stage to model price competition with product choice. The model is, however, different from that of Eizenberg (2014) in that it focuses on variable cost instead of fixed cost, the decision is continuous so that point-identification is possible, and the first stage decision is not directly observed but only indirectly implied from the observed final fuel efficiency.

This paper also contributes to the studies of the impacts of fuel prices and CAFE standards on the automobile industry with a model that endogenizes technology adoption. It is important to put the two factors fuel prices and standards together because, as shown in Figure 1, the timing of the recent improvement in fuel efficiency coincides with substantial changes in both factors. My treatment of product choice goes beyond the research of Gramlich (2010), and, like the work published by Zhou (2016), considers technological adoption beyond the performancefuel efficiency trade-offs, and also considers the cost of adoption.

1.3 Model

1.3.1 Overview

Both supply and demand will be modelled here, and compliance cost due to regulations will also be incorporated. *Demand* will be modelled using a random-coefficient discrete choice specification, similar to that developed by Berry et al. (1995). *Supply* will be modelled with oligopolistic competition between multi-product firms setting not only prices but also fuel efficiency for their vehicles. Firms can improve the fuel efficiency of their vehicles, but need to pay an extra cost per vehicle to install such technology throughout the fleet. *Regulators* set fuel-economy standards each year; firms that violate these standards pay fines. Some firms barely meet the standards, paying no fines, but with the need to optimize under a shadow cost of compliance if they are to maintain their standard-complying status.

1.3.2 Demand

Consumer utility Market t is populated by M_t consumers. Each consumer is to purchase one vehicle from one of the new models available in the market, or to make no purchase[8]. The consumer derives utility from the characteristics of the purchased model. A subset of characteristics is observed by the econometrician; the rest are not observed. Different consumers may also have different tastes regarding each vehicle characteristic.

Suppose consumer i in market t purchasing vehicle model j receives a utility according to the following equation:

$$u_{ijt} = \alpha_{it} p_{jt} + \gamma_{it} dpm_{jt} + {x^u_{jt}}' \beta^u_{it} + \xi_{jt} + \epsilon_{ijt}$$

p_{jt} denotes the price of model j in market t (in thousands of dollars), dpm_{jt} represents the fuel cost of the same model, expressed in terms of dollars per 100 miles of travel, x^u_{jt} is the set of observed vehicle characteristics, ξ_{jt} captures the value from unobserved vehicle characteristics, and ϵ_{ijt} is idiosyncratic shock.

α_{it} and γ_{it} capture consumer marginal disutility from paying for the vehicle purchase and from paying for fuel consumption. β_{it} captures consumers

[8]The no-purchase option includes the choice of purchasing second-hand vehicles, which is not modelled explicitly due to the lack of data.

valuations of the observed vehicle characteristics. These parameters can vary across consumers to capture consumer heterogeneity - the fact that consumers may have different tastes for different vehicle characteristics.

There is substantial variation in income during the sample period, the income effects from which will be accounted for by allowing the coefficients α_{it} and γ_{it} to depend on the median income in each market y_t.

The outside good is assumed to give a zero-mean utility. This is for normalization purposes, as utility is equivalent under translation by a constant.

$$u_{i0t} = \epsilon_{i0t}$$

Distributional assumption The distribution of the parameters on the monetary variables (price and dollars per mile), i.e. α_{it} and γ_{it}, is assumed to be a log-normal distribution. There are two reasons for this choice. First, these parameters are expected to be negative (disutility from losing money), and hence it is necessary to avoid a situation in which consumers may have positive utility for losing money. A log-normal distribution guarantees that the parameters will not change sign. Note that I will allow for a constant to be multiplied with this distribution; the constant can be negative or positive, so I do not impose *a priori* the negativity of these parameters, and only impose the restriction that the parameters do not change sign. The sign of the estimates can therefore still be viewed as a test of the validity of the model. Second, these parameters are expected to interact with income, which empirically follows a log-normal distribution.

Specifically, assume $\alpha_{it} = \frac{\alpha}{y_t}\exp(\sigma_p \nu_{p,it})$ and $\gamma_{it} = \frac{\gamma}{y_t}\exp(\sigma_{dpm}\nu_{dpm,it})$, where α, σ_p, γ, σ_{dpm} are parameters to be estimated, and $\nu_{p,it}$ and $\nu_{dpm,it}$ are unobserved tastes that follow i.i.d standard normal distribution.

The taste parameters for vehicle characteristics are assumed to be normally distributed. Specifically, the taste for the k^{th} characteristics, $\beta^u_{k,it} = \beta^u_k + \sigma_k \nu_{k,it}$. β^u_k will capture the taste of the median consumer, whereas σ_k will capture how dispersed this taste is among all consumers. Note that this specification allows for some consumers to like the characteristics and the others to dislike them. How agreeable the consumers are about certain characteristics depends on how small the dispersion of taste σ_k is.

Choice probability and demand Let $V_{ijt} = \alpha_{it}p_{jt} + \gamma_{it}dpm_{jt} + x^{u\prime}_{jt}\beta^u_{it} + \xi_{jt}$, so that $U_{ijt} = V_{ijt} + \epsilon_{ijt}$. In addition, use t to denote the set of vehicle models available in market t. Consumer i will choose to purchase model j if $U_{ijt} > U_{ikt}, \forall k \in t$. From the perspective of the econometrician, the probability that consumer i will purchase model j, conditional on observables, is

$$P(U_{ijt} > U_{ikt}, \forall k \in t) = P(\epsilon_{ikt} - \epsilon_{ijt} < V_{ijt} - V_{ikt}, \forall k \in t)$$

Assuming type-1 extreme value for the idiosyncratic shocks $\{\epsilon_{i0t}, ..., \epsilon_{ijt}\}$, the above probability will take on a tractable functional form:

$$s_{ijt} = \frac{\exp(V_{ijt})}{1 + \sum_{k \in t} \exp(V_{ikt})}$$

The market share for model j can be calculated by averaging the choice probability over all consumers:

$$s_{jt} = \int \frac{\exp(V_{ijt})}{1 + \sum_{r \in t} \exp(V_{irt})} dF_t(i)$$

The estimation below simulates a population of N_t representative consumers, and the market share can be obtained by taking the average across all consumers in the simulated population:

$$s_{jt} = \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{\exp(V_{ijt})}{1 + \sum_{r \in t} \exp(V_{irt})}$$

1.3.3 Cost

As with the work of Berry et al. (1995) and the existing literature, a constant marginal cost of producing vehicles is assumed here. It is also assumed that the marginal cost can be separated into production cost and fuel-saving technological cost (or fuel-tech cost for short). Production cost is the cost of producing one vehicle, and depends on the characteristics of the vehicles. Fuel-tech cost is the amount that firms can spend on each vehicle to improve its fuel efficiency.

$$c_{jt} = c_{jt}^{\text{prod}} + c_{jt}^{\text{fst}}$$

Production cost The production cost depends on vehicle characteristics. To make sure that the cost is positive, its functional form is assumed to be as follows:

$$\ln c_{jt}^{\text{prod}} = x_{jt}^{c\,\prime}\beta^c + \omega_{jt}$$

x_{jt}^c is the set of observed vehicle characteristics that affect production cost, and ω_{jt} captures the effect of unobserved cost characteristics and technologies.

Fuel-tech cost Fuel consumption rate gpm_{jt}[9] is assumed to be the outcome of the engineering configuration of the vehicle, i.e. it can be expressed as a function of the vehicle characteristics x_{jt}^{gpm}, subject to random technological shocks τ_{jt}.

[9] Fuel consumption rate (measured in gallons per 100 miles) is the inverse of fuel efficiency. Fuel consumption rate is linearly proportionate to fuel cost, which is what consumers really care about. Expressing everything in terms of fuel consumption rate is also more tractable, and hence in the model *gpm* instead of *mpg* is used.

$$\ln g\tilde{p}m_{jt} = x_{jt}^{gpm\prime}\beta^{gpm} + \tau_{jt}$$

This is related to the concept of the fuel-efficiency frontier discussed by Knittel (2011) who observed that trade-offs exist between fuel efficiency and performance characteristics of the vehicle, e.g. vehicles with higher horsepower or greater weight tend to consume more fuel.

There exists a spectrum of technologies that can be installed in a vehicle to improve its fuel efficiency beyond what is determined by the above engineering configuration. Let e_{jt} denote the decrease in fuel saving of the vehicle resulting from firms adopting a spectrum of technologies for their vehicles, i.e.

$$\ln gpm_{jt} = \ln g\tilde{p}m_{jt} - e_{jt} = x_{jt}^{gpm\prime}\beta^{gpm} + \tau_{jt} - e_{jt}$$

e_{jt} would be proportionate to the number of technologies that firms used. In fact, given a convex cost curve for fuel-saving improvement, it is likely that there is one-to-one mapping between e_{jt} and the set of technologies that can achieve such improvement. Therefore, the fuel-tech cost can be defined using this variable:

$$c_{jt}^{\text{fst}} = c^{\text{fst}}(e_{jt})$$

Flexible specification of this function will be allowed for (i.e. polynomial with flexible order).

1.3.4 Market Structure and Competition

In each market, a number of firms, each of which manufactures several vehicle models, compete in two stages. In the first stage, without knowing the realization of the market shocks ξ, ω and τ, they simultaneously choose the fuel efficiency for all their models (by choosing the e_{jt} described above) to maximize their expected profits, which they will receive at the end of the second stage. In the second stage, after learning about the values of these shocks, they choose prices simultaneously to maximize the profit, subject to the existing regulations. Each stage is discussed in detail below, beginning with the later stage.

Second stage

Firms learn about market shocks and make their pricing decisions. In doing so, they need to consider their compliance status with regard to the fuel-economy regulations.

Firms are classified into three types according to the status of their compliance with the fuel-economy regulations:

1. *Violating firms*: those whose average fuel economy is below the standard,

and that pay the resultant penalty.

2. *Unconstrained firms*: those whose average fuel economy exceeds the standards, and that incur no penalty or compliance cost.

3. *Constrained firms*: those whose average fuel economy is binding at the standard levels, and that pay no penalty but incur a compliance cost when adjusting the fuel economy of their fleets to meet the standards.

The fuel-economy standards are set separately for cars and trucks. Let $\overline{mpg}_{ft}^{car}$ and $\overline{mpg}_{ft}^{truck}$ be the harmonic average fuel efficiency for cars and trucks manufactured by firm f[10] and $\overline{\overline{mpg}}^{car}$ and $\overline{\overline{mpg}}^{truck}$ the standards for cars and trucks respectively. If firms violate the standards they have to pay a fine of \$55 per mpg in violation of the total standards across violating vehicles, i.e. $55(\overline{\overline{mpg}}_t^{car} - \overline{mpg}_{ft}^{car})s_{ft}^{car}$, with s_{ft}^{car} being the share of cars of firm f, and $55(\overline{\overline{mpg}}_t^{truck} - \overline{mpg}_{ft}^{truck})s_{ft}^{truck}$, with s_{ft}^{truck} being the share of trucks. The profit for a violating firm will be

$$\pi_{ft} = \sum_{j \in ft} (p_{jt} - c_{jt})\, s_{jt} - 55(\overline{\overline{mpg}}_t^{car} - \overline{mpg}_{ft}^{car})s_{ft}^{car} - 55(\overline{\overline{mpg}}_t^{truck} - \overline{mpg}_{ft}^{truck})s_{ft}^{truck} \tag{1.1}$$

A constrained firm, i.e. a firm whose average fuel economy is binding at the standard level $\overline{mpg}_{ft}^{car} = \overline{\overline{mpg}}_t^{car}$ and $\overline{mpg}_{ft}^{truck} = \overline{\overline{mpg}}_t^{truck}$, pays no fine, but

[10]The harmonic average is used because it is the way the regulator calculates the average fuel economy (in miles per gallon, which is the inverse of the fuel consumption).

in optimizing its profit it needs to operate under the constraint of the regulations (assuming the firms compliance is a prior-determined commitment). In other words, the firm is solving a maximization problem under the constraints $\overline{mpg}_{ft}^{car} \geq \overline{\overline{mpg}}_{t}^{car}$ and $\overline{mpg}_{ft}^{truck} \geq \overline{\overline{mpg}}_{t}^{truck}$. There will be a shadow associated with each of these constraints, denoted by λ_{kt}^{car} and λ_{kt}^{truck} respectively, and the firm will behave as if it maximizes the Lagrange of the constrained profit:

$$\begin{aligned}\pi_{ft} = &\sum_{j \in ft} (p_{jt} - c_{jt})\, s_{jt} - \lambda_{ft}^{car} (\overline{\overline{mpg}}_{t}^{car} - \overline{mpg}_{ft}^{car}) s_{ft}^{car} \\ &- \lambda_{ft}^{truck} (\overline{\overline{mpg}}_{t}^{truck} - \overline{mpg}_{ft}^{truck}) s_{ft}^{truck}\end{aligned} \tag{1.2}$$

Maximizing these profits with respect to prices will provide a system of equations that relates prices and costs. For example, for constrained firms, the first-order condition (FOC) will be

$$\begin{aligned}\frac{\partial \pi_{ft}}{\partial p_{kt}} = &\sum_{j \in ft} \left(p_j - c_j - \frac{\lambda_{ft}}{\overline{\overline{gpm}}_{jt}^{2}} \left(\overline{gpm}_{jt} - \overline{\overline{gpm}}_{jt} \right) - \frac{\lambda_{ft}}{\overline{gpm}_{jt}^{2}} \left(gpm_{jt} - \overline{gpm}_{jt} \right) \right) \frac{\partial s_{jt}}{\partial p_{kt}} \\ &+ s_{kt} \\ = &\; 0\end{aligned} \tag{1.3}$$

First stage

Let $p^*(\xi, \omega, \tau, e)$ be the prices chosen by the firms in the second stage and $\pi^*(\xi, \omega, \tau, e) = \pi(p^*(\xi, \omega, \tau, e), \xi, \omega, \tau, e)$ be the corresponding profit. In the first stage, firms maximize the expected profit, with that profit expectation informed by the distribution of ξ, ω and τ , by choosing the optimal number of fuel-saving technologies:

$$\max_{\{e_{jt}\}_{j\in ft}} E_{(\xi,\omega,\tau)}\left[\pi^*_{ft}(\xi,\omega,\tau,e)\right]$$

The FOC with respect to e_{kt} is as follows:

$$\begin{aligned}\frac{\partial E_{(\xi,\omega,\tau)}\left[\pi^*_{ft}(\xi,\omega,\tau,e)\right]}{\partial e_{kt}} &= E_{(\xi,\omega,\tau)}\left[\frac{\partial \pi^*_{ft}(\xi,\omega,\tau,e)}{\partial e_{kt}}\right] \\ &= E_{(\xi,\omega,\tau)}\left[\frac{\partial \pi_{ft}(p^*(\xi,\omega,\tau,e),\xi,\omega,\tau,e)}{\partial e_{kt}}\right]\end{aligned} \tag{1.4}$$

The last equation is derived from the envelope theorem and the fact that $p^*(\xi,\omega,\tau)$ maximizes π_{ft}. The above FOCs can be used to solve for the marginal cost of fuel-efficiency improvement.

$$\begin{aligned} c^e_{kt} &= \frac{dc^{FST}_{kt}}{de_{kt}} \\ &= E_{(\xi,\omega,\tau)}\left[\sum_{j\in ft}\left(p_{jt} - c_{jt} - \frac{\lambda_{ft}}{\overline{\overline{gpm}}^2_{jt}}\left(\overline{gpm}_{jt} - \overline{\overline{gpm}}_{jt}\right)\right.\right. \\ &\quad \left.\left. - \frac{\lambda_{ft}}{gpm^2_{jt}}\left(gpm_{jt} - \overline{gpm}_{jt}\right)\right)\frac{\partial s_{jt}}{\partial e_{kt}} + \gamma\frac{s_{kt}gpm_{kt}}{\overline{gpm}_{kt}}\right] \Big/ E_{(\xi,\omega,\tau)}[s_{kt}] \end{aligned} \tag{1.5}$$

1.4 Data and Estimation

1.4.1 Data

The model was estimated using the market data of new light-duty vehicles sold in the U.S. from 2006 to 2014. Sales data at the nameplate level (e.g. BMW 3-series, Toyota Camry) were obtained from the Automotive News Market Data Book. Vehicle characteristics, classification, and manufacturer-suggested retail prices (which were used as a proxy for retail prices) at different trim levels (e.g. BMW 328i XDrive 2dr Coupe AWD (3.0L 6cyl 6M), Toyota Camry SE 4dr Sedan (2.5L 4cyl 6A)) were drawn from www.msn.com/en-us/autos. The annual average motor gasoline regular retail prices from the EIA were used as a proxy for fuel prices faced by consumers. The number of households from the U.S. Census Bureau was used as a measure of market size, and household median income from Fred St Louis was used for consumer income. The CAFE standards were taken from the National Highway Traffic Safety Agency.

Vehicle models with sales fewer than 1000 were excluded, resulting in a total of 2,284 model-year observations from 40 manufacturers[11]. The data on vehicle characteristics and prices were matched to the sales data using year, make, and nameplate, and the corresponding average values of all the trim

[11] There were several mergers and splits between firms during the period. These firms will be treated as distinct. In fact, due to the static nature of the model, the same firm from different years will be treated as unrelated in the estimation

levels of the same nameplate were used for estimation.

Table A.2 shows the summary statistics, grouped into Passenger Car and Light-duty Truck categories. On average, trucks have lower fuel efficiency, larger size, and higher horsepower, weight, and torque, and account for approximately 47% of the sales during the period.

Figure B.4 plots the evolution of the sales-weighted average vehicle characteristics over the period. Weight and size stay relatively flat[12], while fuel efficiency and horsepower trend upward.

1.4.2 Identification

In addition to the usual price endogeneity, there is another source of endogeneity that needs to be corrected for. In the first stage, the set of fuel-saving technologies installed in each vehicle is unknown. The selection of these technologies depends on the vehicle characteristics, and hence may potentially bias the estimation of the fuel-efficiency frontier equation.

The first stage of the model closely follows the BLP (1995), as the BLP-type is a natural set of instruments commonly used in the literature to correct for price endogeneity. The average characteristics of other models from the same manufacturer, and those of other models from the same market, are

[12] In the 1990s the opposite was true, i.e. the weight trend was upward and the fuel-efficiency trend was flat. The changes suggest that manufacturers have arrived at a new set of strategic product choices, potentially in response to new fuel prices and a new regulatory environment

used as instruments for the demand and cost equations.

To identify the level of fuel-saving-technology adoption, it is assumed that the fuel-saving-technology cost is convex, so that the marginal cost of fuel-efficiency improvement is monotonic to the level of fuel-saving technology adoption, which enables one-to-one mapping between one variable and another. Specifically, if $c^e_{jt} = c^e(e_{jt}) = \frac{dc^{FST}(e_{jt})}{de_{jt}}$ is monotonic, it can also be written as $e_{jt} = e(c^e_{jt})$, and the fuel-efficiency frontier equation becomes $gpm_{jt} = gpm\left(x^{gpm}_{jt}, \tau_{jt}\right) \exp\left(-e\left(c^e_{jt}\right)\right)$.. A flexible function can be specified to approximate $e\left(c^e_{jt}\right)$.

τ_{jt} may contain unobserved characteristics that affect fuel efficiency and hence can be correlated with c^e_{jt}. Instruments can be used to adjust for such omitted variable bias, but because c^e_{jt} enters the fuel-efficiency frontier equation via a flexible function specification, instruments of higher-order polynomial power may be needed to correct for the endogeneity from the additional terms in the flexible function. A more elegant solution is to use a control function. Suppose we have a set of instrumental variables Z^e for c^e_{jt} and assume that $c^e_{jt} = {z^e_{jt}}'\iota^z + \chi_{jt}$ and $\tau_{jt} = \chi_{jt}\iota^\chi + \epsilon_{jt}$ with ϵ_{jt} being uncorrelated with c^e_{jt}, the frontier becomes $gpm_{jt} = gpm\left(x^{gpm}_{jt}, \chi_{jt}, \epsilon_{jt}\right) \exp\left(-e\left(c^e_{jt}\right)\right)$ with χ_{jt} being estimable by the regression of Z^e on c^e_{jt}, and ϵ_{jt} being an exogenous error term. The instruments used are also of the BLP-type, i.e. the average fuel-related vehicle characteristics of other models manufactured by the same firm.

To identify the compliance cost, firms need to be classified according to their compliance status - standard violating, constrained, or unconstrained. Because of the system of credit trading, averaging, banking, and borrowing, the status is not obvious from the data on average fuel economy itself. Jacobsen (2013) developed a dynamic model of credit banking and borrowing to account for such complications. Gramlich (2010), using data from 1971 to 2007, assumed that European manufacturers were always violating, while their domestic and Asian counterparts were always constrained and unconstrained, respectively. Using a dynamic model, like Jacobsen (2013), is too computationally intensive, while the assumptions that Gramlich (2010) used no longer holdhold[13]. Instead, I will manually go through the times series of the average fuel economy for each manufacturer to identify whether: 1. The firm pays any fine; 2. The average fuel economy of the firm is consistently above or below the standards; or 3. The average fuel economy fluctuates at about the same level as the standards.

1.4.3 Estimation

Assume the following functional forms:

[13] For example, BMW, a European manufacturer, previously chose to violate the standards and pay the fine, but since 2008 have consistently met and sometimes exceeded the standards.

$$\ln gpm_{jt} = x_{jt}^{gpm\prime} \beta^{gpm} - e_{jt} + \tau_{jt}$$

$$\ln c_{jt}^{prod} = x_{jt}^{c\,\prime} \beta^{c} + \omega_{jt}$$

Let ι^e be the parameters that specify the flexible functions $e(\cdot)$. Given a vector of parameters $\theta = (\alpha, \gamma, \sigma_{price}, \sigma_{dpm}, \lambda, \beta, \beta^c, \beta^{gpm}, \iota^z, \iota^\chi, \iota^e)$, we can

- solve for $\{\xi_{jt}\}$ that equates the model-predicted share with the observed shares $s_{jt}(\xi, p, dpm, X^u; \theta_u) = \hat{s}_{jt}$

- solve for $\{c_{jt}\}$ using the second-stage system of FOCs, from which we derive ω_{jt}

- solve for $\{c_{jt}^e\}$ using the systems of FOCs of the first-stage, from which we derive $\chi_{jt} = c_{jt}^e - z_{jt}^{e\,\prime} \iota^z$ and $\tau_{jt} = \ln gpm_{jt} - x_{jt}^{gpm\prime} \beta^{gpm} - e(c_{jt}^e; \iota^e) - \chi_{jt} \iota^\chi$

The following moment conditions are assumed:

$$\begin{cases} E[\xi_{jt} \mid z_{jt}^u] & = 0 \\ E[\omega_{jt} \mid z_{jt}^c] & = 0 \\ E[\tau_{jt} \mid z_{jt}^e] & = 0 \end{cases}$$

There are two complications in estimating the above system of moments. First, the estimation of the second stage of the model requires the calculation of an expectation about the distribution of the random market shocks, which is unknown. Following Eizenberg (2014), there are two approaches: either approximating the expectation of the function by the function value at the expected value of the shocks, i.e. $E_(\xi, \omega, \tau)[f(\xi, \omega, \tau)] \approx f(E[\xi, \omega, \tau])$, or drawing from the empirical distribution estimated in the first stage. Due to computational constraints, in this version of the paper, the first approach is employed.

Second, if joint GMM were to be carried out, all the parameters in the first stage and most of the parameters in the second stage would enter the GMM objective in a non-linear manner, preventing the use of concentration out of linear parameters to simplify the estimation procedure. This would increase the computational burden substantially. Therefore, for this version of the paper, a two-step estimation strategy is employed to reduce the computational complexity, albeit at the cost of lower estimation efficiency:

1. Estimate the first stage using demand and cost moments to derive the first-stage parameters $(\alpha, \gamma, \sigma_{price}, \sigma_{dpm}, \beta^u, \sigma, \beta^c, \lambda)$ – this is similar to the standard BLP.

2. Using the fitted parameters from the first step, solve the second-stage FOCs for c^e and use it to regress the fuel-efficiency frontier equation with the control function.

1.4.4 Results and Discussion

Panels A, B, C, and E from Table A.3 display the first-step estimation results.

Utility The signs of the coefficients on price and cost of fuel consumption (*dpm* dollars per 100 miles) are negative, as expected. The average consumer values power and size and dislikes greater weight. However, the dispersion of taste for power and weight is relatively large, with the standard deviation of the taste distribution for both characteristics similar in magnitude to that of the mean taste, suggesting that a substantial portion of consumers have opposite tastes to those of the average consumer. Regarding size, consumers are generally agreeable on their liking of larger size, with the standard deviation being small compared with the magnitude of the mean taste.

The model implies an average own-price elasticity of -2. This estimate is comparable to those of the existing literature: Zhou (2016)'s estimate is -2.0, Klier and Linn (2012)'s -3.48, and Klier and Linn (2012)'s -1.4. My estimates imply an average mark-up of 46%.

Production cost Increasing power, weight, and torque increases production cost. This is to be expected, as increasing power and torque generally requires higher-quality components and materials. Increasing size decreases cost, but this effect should be interpreted as conditional on fixing weight and power.

Production cost is trending downwards, reflecting a general improvement in manufacturing technologies, among other factors.

Fuel efficient frontier An increase in power, weight, and torque decreases the fuel efficiency of the vehicles. This is similar to the results of Knittel (2011), who found a trade-off between fuel economy and vehicle performance. The time effects are trending up, suggesting an improvement in fuel technology over time. This is also in line with the outward shift of the fuel-efficiency frontier over time reported by Knittel (2011).

Compliance cost The estimated compliance cost for constrained manufacturers is 230 USD per mpg per vehicle. This is higher than the penalty for violating the standard (55 USD per mpg), but comparable with estimates from Jacobsen (2013)), who made estimates in the range of $157$264, and Gramlich (2010), who made an estimate of $347. Reasons for these high compliance costs include reputational cost and political cost (due to damaged relationships with regulators and legislators). However, these estimates are in contrast with the low compliance cost estimated by Anderson and Sallee (2011), who suggested a value of $9 to $27.

Marginal fuel-tech cost In the second stage, a polynomial is used to approximate the fuel-tech function $e\left(c^e_{jt}\right)$. Polynomials of different order have been tried, from 1 to 7. Figure B.5 plots the shape of such functions[14]. A polynomial of order 3 was chosen for the final estimation because going beyond order 3 does not provide significant additional explanatory power to the function.

Panel D from Table A.3 displays the results from the second step. The median c^e across all vehicles is 2.88, which means that, on average, manufacturers incur a cost of 288 USD to improve the fuel efficiency of their vehicles by 1% beyond what the vehicle has achieved. To put the numbers into perspective, for a vehicle with fuel efficiency of 30 mpg equipped with the median amount of fuel-saving technology, it will cost 944 USD to increase its fuel efficiency to 31 mpg.

Figure B.6 plots the distribution of c^e separately for cars and trucks. The truck distribution is more skewed to the right, suggesting that trucks have exhausted their fuel-saving options more than cars have. Almost 40% of the car models have a marginal cost of fuel-efficiency improvement of less than 1,000 USD/mpg, and 76% of less than 2,000 USD/mpg, implying that cars

[14]Note that the absolute vertical location of the curve is not important because the constant term of the function $e(c^e_{jt})$ is not separately identified from the constant term of the fuel-efficiency frontier equation.

still have a lot of room to further adopt more fuel-saving technologies at low cost. The respective numbers for trucks are only 18% and 56%.

Counterfactual Studies This section details the investigation of how fuel prices and CAFE standards, individually and together, have affected the recent changes in fuel economy of new vehicles sold in the U.S. This will be addressed by several counterfactual studies. In particular, the following three counterfactual simulations will be carried out:

1. Gasoline prices over the years are kept at the 2006 level
2. CAFE standards over the years are kept at the 2006 level
3. Both gasoline price and CAFE standards over the years are kept at the 2006 level

When estimating the model, the compliance costs are assumed to be same across fleets. In counterfactual scenarios, it is impossible to impose such constraints, because with any market changes firms, due to differences in cost and product structure, will response differently, moving the cost margins in different directions and magnitudes. Therefore, in all the simulations, I allow the compliance costs to vary across fleets such that firms maintain the original compliance status, unless it is impossible to do so (e.g. maximum mpg is less than the standard) or it is profitable to become unconstrained, not accounting for the fixed cost of non-compliance (which is not estimated in the model).

The counterfactual equilibrium is calculated using a nested iteration algorithm that sequentially solves for the equilibrium prices, the compliance costs that satisfy the compliance status, and the equilibrium choices of fuel-saving-technology adoption.

1. Outer-iteration: pick a vector of FST level $\{e_{jt}\}$

    (a) Middle-iteration: given $\{e_{jt}\}$, pick a vector of compliance cost $\{\lambda_{jt}\}$

        i. Inner-iteration: given $\{e_{jt}, \lambda_{jt}\}$ pick a vector of prices $\{p_{jt}\}$

            A. Check if $\{p_{jt}\}$ solve the first-stage FOC, up a tolerance of 1e-8

            B. If it is, go to the middle-iteration

            C. Else pick a new price vector and continue with the inner iteration

        ii. Check if the compliance constraints are all satisfied

        iii. Pick a new vector of compliance cost by tightening all the constraints that are satisfied and loosening all the constraints that are violated

        iv. If the new vector are close enough to the old vector (up to a tolerance of 1e-6), go to the outer-iteration, else continue with the middle-iteration

(b) Check if the second-stage FOCs are satisfied, up to a tolerance of 1e-4

(c) If it is, stop, else pick a new vector of $\{e_{jt}\}$, update the fuel-efficiency and FST costs accordingly and continue the outer-iteration

The counterfactual average fuel economy (in miles per gallon) for passenger cars and for light-duty trucks is plotted in Figure B.7, B.8, B.9 (for cars), B.10, B.11 and B.12 (for trucks). Two types of averages are considered: raw averages and sales-weighted averages. Raw averages are considered to separate the effect of demand from the firms choice of fuel efficiency. Thus, the raw average will reflect more of the firms adoption of fuel-tech, while the sales-weighted average will capture both the adoption by firms and the demand from consumers for fuel-efficient vehicles.

Cars Figure B.7 demonstrates the effects of changing fuel prices from the corresponding level to the 2006 level on the average fuel economy of cars. The raw average of fuel economy does not change much under counterfactual fuel prices, suggesting that fuel-price changes do not significantly affect the firms choice of fuel efficiency for passengers. The sales-weighted average fuel economy decreases slightly under 2006 fuel prices (which were mostly lower than the fuel prices in other years), indicating that consumers switch to less fuel-efficient cars when fuel cost decreases.

Figure B.8 demonstrates the effect of changing CAFE standards to 2006 levels on the overall car fuel efficiency. The raw average decreases slightly, especially after 2011 (the year the car standards started to rise), indicating that standards influenced firms adoption of fuel-tech for cars. The change in the sales-weighted fuel economy is more pronounced, even when fuel cost does not change, suggesting that consumers switch fuel-efficiency class when firms adjust their fleets to meet standards.

Figure B.9 shows the effects of changing both fuel prices and CAFE standards to the 2006 levels. The effects on both the raw average and sales-weighted average fuel economy seem to reflect the combined effects of the individual changes.

Trucks Figure B.10 shows the effect of fuel prices on truck fuel economy. Changing prices does not seem to change the average fuel efficiency firms set for their fleet, or the average fuel efficiency consumers choose for their vehicles. However, when the standards change, as shown in Figure B.11, the fuel efficiency of trucks undergoes a big change, in both the raw average and the sales-weighted averages. Thus, for trucks, fuel prices have minimal impact while CAFE standards have a big impact both on how firms adopt fuel-tech for their vehicles and on how consumers choose fuel-efficient vehicles. This is related to the results discussed above regarding the marginal fuel technological cost of trucks. The options to improve the fuel-efficiency for trucks seem to have been exhausted, and so trucks have a higher marginal cost of improving their fuel efficiency further. A small decrease in standards would relieve firms of a large amount of fuel-tech costs that they incur to make their trucks comply with regulations, and this cost-saving will be passed down to consumers and induce even more change in the sales-weighted fuel economy.

1.5 Conclusion

In this paper, a structural model of oligopolistic competition in the U.S. automobile industry that allows for the adoption of fuel-saving technologies has been developed and fitted to market data. The counterfactual results show that changes in regulations mostly explain the automakers adoption of fuel-saving technologies, but that consumers adoption of those technologies depends substantially on fuel prices. There is also a notable difference between light trucks and passenger cars, namely that the fuel efficiency of light trucks is more influenced by regulations than by fuel prices, relative to passenger cars. These results highlight the challenges to the industry to meet the fuel-efficiency target in the near future, when the fuel prices are forecasted to drop substantially. It is also suggested that regulators may need to focus more on passenger cars, for which there is still much room for improvement. This is because the cost of improving the marginal fuel efficiency of a large proportion of passenger-car models is still relatively low.

Chapter 2

Price Salience and Imperfect Information: Evidence from Experiments at Fueling Stations in Brazil[1]

2.1 Introduction

To reduce dependence on oil and address growing environmental concerns over emission from fossil fuels, many countries are exploring alternative ways to power their economies. Biofuels such as corn ethanol, sugar-cane ethanol and biodiesel are being introduced and actively promoted to consumers as substitutes to conventional fuels such as gasoline or diesel. Almost 5% of energy consumption in the US transportation sector in 2015 was powered by biofuel (US Energy Information Administration, 2017), while the European Commission targets to achieve at least 3.6% in the use of advanced biofuels for

[1]The data used in this chapter were collected and generously provided to me by my advisor, Professor Alberto Salvo. The chapter is co-authored with Professor Alberto Salvo.

transportation by 2030 (European Commission, 2016).

Our current understanding of how consumers perceive and value these new energy choices is still very limited. Recent empirical evidence suggests that consumers may not see biofuels as perfect substitutes to petroleum-based fuels. Anderson (2012) found that a substantial fraction of US households were willing to pay a premium for ethanol. In the Brazilian context, Salvo and Huse (2013) reported that about 20% of consumers still chose gasoline over ethanol, even when it was priced 20% above ethanol in terms of dollars per mile driven. Consumers seem to leave a lot of money on the table when it comes to fuel choice for their vehicles.

This imperfect substitution may be due to consumers valuing non-price characteristics (with price based on the amount of dollars per mile driven). Concerns about health, technologies, the environment and various other economic and social issues can be factored into consumer valuation of the fuels. As shown in Salvo and Huse (2013), motorists in Brazil perceive ethanol and gasoline to have differing impacts on the environment, the local economy and on the performance of the vehicle engine, which translates significantly to the variation in choices that they make.

Imperfect substitution can also be due to imperfect price information. The majority of consumers surveyed in Salvo and Huse (2013) chose the fuel that they had purchased in the last two stops, and the act of "choosing out of habit" significantly increases the likelihood of purchasing the more expensive fuel. This observed persistence in consumer fuel purchases could well be explained by the heterogeneity in consumer preference, which is consistent with the aforementioned theory of consumers valuing non-price characteristics. However, it can also be due to friction in the way consumers obtain and process information necessary to make optimal decisions about fuel choices.

Sub-optimal purchasing decisions have been studied extensively, with empirical evidence and policy implications in various settings. In the context of energy and fuel choice, Hastings and Shapiro (2013), Anderson (2012) and Rivers and Schaufele (2015) provided evidence that consumers respond to changes in carbon taxes, gasoline taxes and prices of fuels of different grades in ways that are inconsistent with optimal full-information decision-making. A common theme amongst these papers and other similar studies is the finding that information salience, provided in the forms of information on taxes or prices, even when publicly available and easily accessible, may still affect consumer choices in a meaningful way.

This paper attempts to bridge the gap between the literature on alternative fuel choices and the literature of information friction. A discrete choice model will be developed, which will capture the effect of imperfect information while controlling for consumer heterogeneous preference. The model will be used to address the following questions. First, do consumer choices reflect imperfect information about fuel prices? Second, what is the effect of increased price salience on reducing such imperfect information, if any? And finally, to what extent does price salience and imperfect information explain the substitution pattern between different fuel types?

These questions will be addressed empirically by analyzing data from a series of experiments conducted in fueling stations across four cities in Brazil. The experiments include two treatments that increase price salience. The treatments take the form of verbal statements and and printed flyers that are presented to consumers before their placement of orders. The treatments are meant to provide the consumers with accurate and easy to understand information on prices, minimizing the potential suboptimal fuel choices due to misperception about the relative cost-effectiveness of the fuels. Differences in choices between the control group and the treated group will reveal the effect of the price salience, which help us identify the existence and the magnitude of the price noise that the consumers possess.

The primary contribution of the paper is providing the empirical results that are consistent with the purchasing behaviors of consumers who possess imperfect information about fuel prices. This extends the existing body of empirical evidences on imperfect information and the effects of information salience. More importantly, these results contribute to the understanding of consumer behaviors in the retail energy market. Particularly, the paper provides a partial explanation for the puzzle why Brazillian consumers often choose the more expensive fuel even when the cheaper alternative is equally accessible. This has important policy implications. If the low adoption of alternative fuel stems from consumer valuing non-price characteristics of the fuels, there is little that the policy-makers can do to promote the alternative fuel, especially in the short-term. However, if the low adoption is because of the low awareness of the relative cost-effectiveness of the fuel, policy-makers can implement policies that raise price salience to attract consumers towards the cheaper fuels.

The secondary contribution of the paper is the development of the econometric model that captures the effects of price noise and of price salience on consumer choices. The model eventually reduces to a heteroskedastic probit choice model, which enables straightforward estimation and interpretation. The application of the model goes beyond analyzing fuel choices and the model can be employed in other settings where the researchers need to estimate the effects on price noise with data on revealed preference.

2.2 Background and related works

2.2.1 Background

In Brazil, sugar-cane ethanol is widely available as an alternative to gasoline. This is due to government policies established in the 1970s in response to the oil crisis, which mandated the supply of sugarcane ethanol at fueling stations across the country, whilst actively promoting the new fuel through various agencies and programs. It was also a requirement for gasoline to be blended with anhydrous ethanol, with an increment in the blending ratio over time, from 10% in 1976 to 27% in 2015.

Since 2003, car-makers have been marketing flexible-fuel vehicles cars that could operate on any combination of gasoline and ethanol. This new fleet of vehicles proved to be a success and rapidly replaced single-fuel vehicles. By 2012, 95% of new cars and 57% of new light commercial vehicles registered in Brazil were flexible-fuel vehicles (Brazilian Automotive Industry Association, 2013).

Due to the unique situation in Brazil, ethanol, along with gasoline, has become accessible to a large population of motorists in the country. This fact constitutes an important contribution of this paper. In other countries, even if alternative fuels are introduced, it is often not accessible to the general population, because of constraints in infrastructure (dual-fuel stations) and availability of suitable vehicles. Any study on fuel choice in such a setting needs to take into consideration consumer selection within the market, as well as the external validity of the demand estimation. Brazil provides a unique context, as the alternative fuel is as accessible as the conventional fuel, hence alleviating concerns about demand being influenced by infrastructure or technologies instead of prices in the study of fuel choices.

2.2.2 Fuel prices in Brazil

In Brazil, gasoline prices are highly regulated by the government. There were periods of time when variations in international crude oil prices did not translate into changes in gasoline prices at local fuel stations[2]. In contrast, the supply chain for the ethanol industry was deregulated in the 1990s and ethanol prices were greatly subjected to market forces. Particularly, ethanol prices in Brazil were largely influenced by supply and demand in the sugar market, as sugarcane was the main ingredient for ethanol production in the country. In fact, between 2000 and 2010, ethanol prices at the pump have peaked each time world sugar prices rose beyond a certain threshold (Salvo and Huse, 2013). This provides a source of variation in the relative prices between gasoline and ethanol, which will be exploited in this paper to identify the demand between

[2]Salvo and Huse (2013) mentioned that one such occasion was in mid-2008, when there was a peak in world oil prices, but no similar peaks were observed in gasoline prices at local stations during that period of time.

for the two fuels.

Pure gasoline is 1.41 times higher in energy content than the same volume of pure ethanol. The blend of gasoline used in Brazil during the period of study contained 18% to 25% of ethanol; which resulted in an approximate 30% difference in energy content between the blended gasoline and ethanol. This means that for ethanol prices to be equivalent to gasoline in terms of dollars per mile traveled, it needs to be priced at 0.7 the price of gasoline for the same volume. In fact, the ratio of 0.7 was widely reported in the media across the country, thus informed motorists would potentially be aware of it.

Figure D.1 plots the price path for gasoline and ethanol during the period of study, with gasoline prices scaled down by a factor of 0.7 to be comparable with ethanol prices. The figure shows that there was substantial variation in the relative prices between the two fuels over time and across cities, especially in Sao Paulo and Curitiba. In Belo Horizonte and Recife, ethanol was the more expensive fuel for most of the given period, although the price differences experienced small changes due to fluctuation in ethanol prices. In Curitiba and Sao Paulo, ethanol prices experienced a hike before March 2011, then decreased quickly to a point below the corresponding normalized gasoline price, before undergoing a smaller increase in June and stabilizing afterwards. This led to a series of changes to the prices of ethanol, from ethanol being the cheaper fuel in Curitiba and Sao Paulo, to being fairly priced, and eventually becoming the more expensive fuel relative to gasoline in the same cities. The experiments, the timing of which is indicated by the vertical line in the figure, captured this fluctuation in relative prices well.

2.2.3 Fuel characteristics and differences between ethanol and gasoline

Engineering studies have identified several important differences between ethanol and gasoline as a fuel for spark-ignition engines (Hsieh et al., 2002; Yüksel and Yüksel, 2004; Masum et al., 2013). In this section, we will attempt to summarize relevant information from these studies (mostly from section 3 of Masum et al. (2013)).

The most relevant and important difference lies in heating value. The heating value of ethanol is about 1/3 times lower than that of gasoline. Hence, for the same energy output (to enable the same distance of travel), more liters of ethanol would be required. This has two implications. Firstly, to be fairly priced (in terms of km per dollar), the same amount of ethanol has to be priced at a lower price than gasoline. Secondly, given one common tank used for both fuel in most bi-fuel vehicles, one can travel over longer distances on gasoline than on ethanol before having to visit fueling stations.

Ethanol has a higher octane number than gasoline, enabling it to withstand higher compression before detonating. This leads to better anti-knock characteristics and reduces engine damage resulting from premature fuel ignition. Ethanol, being miscible with water, can cause corrosion on mechanical components, particularly those made of copper, brass or aluminum. It can also corrode certain types of rubber and cause blockages in the fuel pipe, unless fluorocarbon rubber or appropriate materials are used in replacement. Corrosion inhibitors can be incorporated into the fuel to alleviate these problems. Modern bi-fuel vehicles are also designed with parts and materials that are resistant to corrosion. However, the inadequate management of the issues relating to corrosion in early days of bi-fuel vehicles has left certain negative impressions among older consumers.

Emission-wise, ethanol produces significantly less CO and HC, two important green-house gases (GHG), whilst moderately increasing the emission of CO_2, according to the study by Hsieh et al. (2002). Accounting for CO_2 absorbed by the crop used to produce ethanol, which can offset the life-cycle GHG emission of bio-based ethanol, Wang et al. (2012) estimated that sugarcane ethanol can reduce the life-cycle GHG by 40-62% relative to gasoline. The emission of other pollutants (NO_x, benzene, formaldehyde, acetaldehyde, acetone, etc.) has also been studied, with varying advantages and disadvantages of ethanol relative to gasoline.

Beyond chemical and physical characteristics, different fuels also possess varying social attributes. Sugar and ethanol are important products in many states in Brazil. Consumers in those states can exhibit home bias, favoring local products over the alternatives, as they care more about the local society and economy.

We will conclude this subsection by emphasizing what we mean by controlling for non-price characteristics of the fuel. As we only have 3 fuel alternatives (ethanol, gasoline and mid-grade gasoline), instead of viewing a fuel as a bundle of characteristics like what is normally done in hedonic regression, it is more efficient and transparent to use dummy variables to represent each fuel. The dummy will capture the effect of all the fuel-specific characteristics.

What is more important to be controlled for is the fact the different consumers may value the same characteristics differently. Older consumers may have more resistance adapting to new technology, and people with needs for vehicle performance or high usage may prefer fuel that they perceive to be superior for engine operation and maintenance, while those with environmental or health concerns opt for fuel with lower emission. To account for this heterogeneity, we will use observed consumer characteristics, including important demographics such as age, gender, education, as well as surveyed information about their usage and vehicle prices, together with vehicle characteristics, in conjunction with interactions with fuel types. Unobserved heterogeneity will also be partly controlled for by the use of station fixed effects.

2.2.4 Related works on alternative fuels

Demand for fuel has been extensively studied[3], but the focus has mainly been on fossil fuels. With the introduction of alternative fuels in many countries around the world, literature on alternative fuels is emerging, but is still limited in comparison to that of gasoline and diesel.

Anderson (2012) examined the demand for corn-based ethanol using data from Minnesota, US. The data is aggregated at a station-month level and consumer willingness to pay for ethanol was identified through OLS as well as 2SLS regression. The paper found that many consumers were willing to pay a premium for alternative fuel. As a substitute for gasoline (E10), demand for ethanol was estimated to have an own-price elasticity of negative 3.2-3.8 and a gasoline-price elasticity of 2.3-3.2.

Salvo and Huse (2013) investigated consumer choices between sugarcane-based ethanol and gasoline in 6 cities in Brazil. The paper found that a large proportion of consumers chose to purchase the more expensive fuel, and there was significant consumer heterogeneity in the choice between the two fuels.

[3]Some well-known surveys on this literature are: Drollas (1984), Dahl and Sterner (1991), Goodwin (1992) Espey (1998), Graham and Glaister (2002)

Salvo (2015) used the same data set as the one used in this paper and attempted to answer similar research questions, albeit from different perspective. The paper provided reduced-formed evidences to support the fact that increasing price salience can influence consumer choices of fuels at the pump. It was reported in the paper that displaying information about relative prices of the fuels prior to the purchases can increase the probability of choosing ethanol by as much as 6 percentage points when ethanol is "very favorably priced" in comparison to gasoline. The paper provided two structural models explaining the effect of increased price salience on consumer choices, namely price salience shifting consumers' sensitivity to prices and price salience shifting the consumers' consideration set. The paper also highlighted the fact that, although the effect of the price salience treatment is statistically significant, its magnitude is small compared to the effects due to consumer heterogeneity such as education.

This paper builds upon Salvo (2015). Our motivation is similar: to explain why consumers choose the more expensive fuel and to examine how price salience affects consumer choices. However, we recognize the following shortcomings of the models developed in Salvo (2015). With regards to the first model, price sensitivity is associated with preference and budget constraints, and it is not clear how price salience can affect such primitives as preference and budget. From a modelling perspective, allowing a treatment to alter preference without understanding the mechanisms in-between is troublesome for the validity of the model because it entails the possibility of endogenous preferences. Thus, the model with price salience shifting price sensitivity should best be considered only as a reduced-form analysis. With regards to the second model, it is not clear how the consideration set is formed, and why some consumers choose to exclude certain fuels from their consideration set, especially when the number of choices to be considered is small (three or four fuel types). Moreover, the number of possible consideration sets increases exponentially with the number of choices, and the model had to make certain assumptions to reduce the size of the state space, raising the concern over identification.

The present paper explores an alternative explanation for the phenomenons. We posit that some consumers choose the more expensive fuels because they do not possess accurate prices information, and price salience can affect consumer choices by improving the accuracy of the price information that the consumers possess. This explanation, if proven to be significant, provides an explicit mechanism for price salience shifting price sensitivity. There is a large literature on both the theory and the empirical evidence of imperfect information that can provide explanation and insights about the inaccurate prices that consumers possess. The size of set of parameters necessary to model consumer price information is manageable, which facilitates reliable identification strategy.

2.2.5 Related works on information friction and information salience

There is a large body of theoretical as well as empirical research on information friction. These works have identified various reasons why consumers may not have the most accurate information available, or may not react to it optimally. For examples, there can be external and internal costs associated with information acquisition (i.e. search cost) because of which consumers may find it not worthwhile to continuously acquire updated information about prices. The optimization problem may be overly complicated or may require complex input, which may dissuade consumers from solving it fully and optimally. The scarcity of various cognitive resources such as memory and attention can further impair consumer's ability to obtain and process all the information available to them. The issue has been studied in several contexts, such as in the context of health care (Handel, 2013) and cellphone plans (Miravete, 2003). In this paper, the issue will be investigated in another context, namely the fuel retail industry.

One related literature is the literature on information salience. Chetty et al.

(2009) conducted an experiment in which tax-inclusive price tags were posted on various items in grocery stores, and found that the increase in the salience of sales tax reduced consumer demand by 8%. Similarly, Busse et al. (2006) found that demand is more responsive to shipping charges than auction prices in an experiment conducted using an online auction platform; Brown et al. (2010) found that consumers react differently to rebates than to car prices in US automobile market; DellaVigna and Pollet (2009) found that investor's response to earnings announcements on Friday was less immediate and more drift than announcements on other weekdays. These findings share a common theme: two pieces of information, although being equivalent in content, can lead to significantly different responses from consumers, depending on how salient they are to the consumers.

The treatments that we conducted in the experiment involved communicating verbal and printed information to consumers. Such communication can be viewed as a form of "advertisement". Effects of advertising on consumer choice have been studied extensively. The long-standing issue in this literature is the distinction between the informative, persuasive and complementary natures of advertisement. Using scanner data of 10 brands of toilet tissues, Tellis (1988) found that advertising is less important to brand choice than other marketing variables such as price, features and displays. This is consistent with studies on brand choice of aluminum foil and dry dog food by Kanetkar et al. (1992), which also found that advertising increases price sensitivity. Although for yogurt, Pedrick and Zufryden (1991) reported stronger effect of advertising exposure on brand choice. Ackerberg (2001) suggested that one way to untangle the informative effect from the persuasive (or prestige) effect of an advertisement of experience-based good is to look at its interaction with experience, namely the first effect only influences inexperienced consumers whereas the later effect influences everyone. Utilizing scanner data, Ackerberg (2003) developed a structural model of brand choice for yogurt and found advertisement having significant informative effect but insignificant prestige effect. Anand and Shachar (2011) studied network television industry and structurally estimated a significant informative effects of advertising on consumer choice, although the effect can be negative. Surveying various studies (Bagwell, 2007) concluded that most empirical evidences point towards information as the main factor driving advertising effects. Howerver, existing studies have only focused on a small set of products, most of which are frequently purchased consumer goods, and we should be cautious when generalizing these results to other situations. To the best of my knowledge, other than the model examined in Salvo (2015), there has not been a study about the effect of advertising on choice of fuels.

2.3 Experimental design and treatment effect

The experiments were conducted at 52 different fuelling stations across four cities in Brazil - Sao Paolo, Curitiba, Belo Horizonte and Recife. There were in total 193 visits to the stations, with majority of the stations visited more than three times (up to a maximum of 5 times), resulting in a total of 10,422 subjects surveyed.

Each station visit was conducted within a day. The timings of the visits were varied so that we would have data at different times of the day on different days of the week. By design, the first 18 consumers were assigned to the control group, the next 18 to one treatment group, and the subsequent 18 to the other treatment group.

Control group After the consumer had placed his order with the station's attendant and the vehicle was in the process of being serviced, the enumerator would survey the consumer to match the observed fuel choice with demographic characteristics.

Two treatments by station visit A treated subject would hear one of the following statements, either from the attendant or from the enumerator, prior to placing his order. The statement would be consistent with actual price levels for regular-grade gasoline and ethanol posted at the station's pump on the day of the visit:

- *Hoje a gasolina está mais vantajosa, veja aqui,* loosely translated as "Today gasoline is more advantageous / the better deal, see here,"
- *Hoje o álcool está mais vantajoso, veja aqui,* translated as "Today ethanol is more advantageous / the better deal, see here,"
- *Hoje a gasolina e o álcool estão com rendimento parecido, veja aqui,* translated as "Today gasoline and ethanol oer similar yields / similar deals, see here."

As the subject hears the verbal statement regarding relative price conditions at the station on that day, he or she would be handed a flyer similar to that shown in Figure D.2, for the price-ratio relative to 70% treatment, or Figure D.3, for the km per R$50 treatment. The illustrated flyer corresponds to a station visited in Sao Paulo on June 13, 2011 in which $(p_e, p_g) = (1.649; 2.499)$. To be clear, some treated consumers were handed a price-ratio flyer and others were treated with a km-per-R$50 flyer, and in either case, the flyer was consistent with the verbal statement that introduced it.

The price-ratio flyer (Figure D.2) had a thumbs up alongside Mais vantagem status to gasoline when $p_e/p_g \geq 0.705$, and a thumbs up for ethanol when $p_e/p_g \leq 0.695$, and stated that both fuels offered similar yields (Rendimento parecido) otherwise. The flyer also reminded subjects of the media-reported parity threshold, listing specialist advice that when the price ratio (between ethanol and gasoline) is: (i) lower than 70%, ethanol is more advantageous; and when (ii) higher than 70%, gasoline is more advantageous.

The km-per-R$50 flyer (Figure D.3) had a thumbs up for gasoline when the distance to be travelled on R$50 of regular gasoline purchased at the station was expected to exceed, by some margin, the distance to be travelled on R$50 of regular ethanol, and a thumbs-up for ethanol in the opposite situation, and was neutral when gasoline and ethanol offered similar yields. A table in the flyer specially provided comparison for the estimated distances travelled using ethanol and gasoline across three of the most popular vehicle engine sizes, namely the 1.0, 1.4, and 1.8 liter engines (absolute fuel economy generally declines with engine size).

Table C.1 reports the summary statistics of variables that vary at the consumer level and the difference in means test to check for covariate balance across the control and treatment groups. Across station visits, the average number of subjects who declined to participate is 3.3 for the control group, 2.7 for those subjected to the price-ratio flyer treatment and 2.5 for those subjected to the km-per-R$50 flyer treatment. Overall, the sample appears to have the statistical properties of a randomized experiment (see Salvo (2015) for more details).

Salvo (2015) also reported that when ethanol was very favorably priced ($p_e/p_g < 0.7/1.1$), the price-ratio flyer treatment increased the proportion of ethanol purchases by 6 percentage points, and the km-per-R\$50 increased it by 5 percentage points.

2.4 Model of consumers with noisy price information

In this section, a discrete-choice random utility model with noisy information on fuel prices is introduced and then employed to investigate whether imperfect price information can explain the imperfect substitutability between gasoline and ethanol. The model will also be used to quantify the effects of the two above-mentioned price salience treatments.

Suppose that instead of observing prices with perfect accuracy, consumers' perceptions of fuel prices were subject to a random noise and they made purchasing decisions based on these perceived prices. Denote p_{jt} to be the actual prices of fuel j in station t in terms of R\$ per km of travel, $\tilde{p}_{ijt}$ the perceived price, and z_{ijt} the price noise. Assume that the price noise enters multiplicatively:$\tilde{p}_{ijt} = p_{jt} \exp(z_{ijt})$. The utility from purchasing fuel j at station t for consumer i is assumed to take the following form:

$$U_{ijt} = \alpha_i \ln \tilde{p}_{ijt} + \beta_j X_{it} + \xi_{jt} + \epsilon_{ijt}$$

X_{it} denotes the consumer characteristics, ξ_{jt} captures station-specific factors and ϵ_{ijt} captures idiosyncratic utility shocks. The term ξ_{jt} can be interpreted as product characteristics that are specific to each station, such as brands or the availability of the types of fuels. It can also capture the demographics of the neighborhood, and hence absorb certain consumer characteristics that may have been in X. If firms priced strategically according to locations, ξ_{jt} will also absorb the effects of such supply factors. α_i captures price sensitivity. β_j, which varies with fuel, captures consumer heterogeneous preference for fuels the fact that consumers with characteristics X_{it} can value different fuels differently.

The logarithm form is used for prices because we found that consumers were more responsive to price ratios than price differences. In fact, if both variables are included in the same equation, the effects of price ratios will be larger and statistically significant, whereas the effects of price differences would be smaller and statistically insignificant[4].

Rewrite the utility in terms of the actual prices and the price noise, as

[4] As we will discuss later, the empirical actually identifies relative utility instead of utility levels, which means that for estimation we will difference the utility with respect to a reference fuel. Differencing two logarithms is equal to a logarithm of ratios; so using logarithm is a convenient way of making use of price-ratio variation.

shown below:

$$U_{ijt} = \alpha_i \ln p_{jt} + \beta_j X_{it} + \xi_{jt} + \underbrace{\alpha_i z_{ijt} + \epsilon_{ijt}}_{\tilde{\epsilon}_{ijt}}$$

Denote $\tilde{\epsilon}_{ijt} = \alpha_i z_{it} + \epsilon_{ijt}$, the combination of the stochastic error term ϵ_{ijt} and the price noise z_{ijt}. The utility can now be viewed as a function of the actual price p_{jt}, but with a new stochastic term that incorporates the price noise. One way to think of the effect of price noise on fuel choice is that it introduces noise to the latent utility, which translates into more noise towards choices.

Consequently, if the price salience treatments are at all effective in informing consumers about the accurate prices of the fuels, they should remove or reduce the effect of these noises in the stochastic term of the utility. Thus, one way to identify the effectiveness of the salience treatments is to look at the variance of $\tilde{\epsilon}_{ijt}$ and see how it changes between the control group and the treated groups. Furthermore, given that the best the treatments can do is to remove all the price noise, looking at the effects of the treatment can reveal the lower bound of the amount of noise the consumers face[5].

To facilitate the following exposition, let us simplify and assume only two

[5] This is of course dependent on the condition that there is no other unintended effect of raising salience beyond its informational intent. One could worry about Hawthorne's effect - the fact that consumers change their behaviour due to being observed, or due to unfamiliar interaction during their purchase. But we did have a portion of the experiments conducted by the station attendant, who was likely to be familiar with the consumers and a usual part of the experience at the gas stations; while the rest of the experiments were conducted by trained numerators.

choices, i.e. $j \in 0, 1$. Also, we will drop the station index t for brevity.

$$U_{i0} = \alpha_i \ln p_0 + \beta_0 X_i + \xi_0 + \tilde{\epsilon}_{i0}$$

$$U_{i1} = \alpha_i \ln p_1 + \beta_1 X_i + \xi_1 + \tilde{\epsilon}_{i1}$$

Consumer i will choose option $j = 1$ when

$$U_{i1} > U_{i0} \Leftrightarrow \alpha_i \ln p_1 + \beta_1 X_i + \xi_1 + \tilde{\epsilon}_{i1} > \alpha_i \ln p_0 + \beta_0 X_i + \xi_0 + \tilde{\epsilon}_{i0}$$

$$\Leftrightarrow \tilde{\epsilon}_{i0} - \tilde{\epsilon}_{i1} < \alpha_i(\ln p_1 - \ln p_0) + (\beta_1 - \beta_0) X_i + \xi_1 - \xi_0$$

Increasing β_0 and β_1 by the same amount will not affect the above condition, which also applies for ξ_0 and ξ_1. Thus, we cannot identify both β_0 and β_1, and only the difference between them can be identified. Again, this applies similarly for ξ_0 and ξ_1. This is a common feature of discrete choice, in which the utility cannot be identified in all levels, but only in the difference compared to a reference alternative. Identifying the difference between β_0 and β_1 is equivalent to setting $\beta_0 = 0$ and identifying β_1. Likewise, we need to set $\xi_0 = 0$ and identify ξ_1.

Based on the above argument, it can be argued that identifying the distribution of both $\tilde{\epsilon}_{i0}$ and $\tilde{\epsilon}_{i1}$ is not feasible. Instead, we will identify the distribution of the differenced error term $\tilde{\varepsilon}_i = \tilde{\epsilon}_{i0} - \tilde{\epsilon}_{i1}$. Note that this new error term contains the noise about the price ratios: $\tilde{\varepsilon}_i = \alpha_i(z_{i0} - z_{i1}) + \epsilon_{i0} - \epsilon_{i1}$, and thus the logic about the effect of price salience on the variance of the error term still applies here. However, the price noise here is no longer the noise about the individual fuel price itself, but should be viewed as the noise of the price ratio between a pair of fuels.

The probability of consumer i choosing fuel $j = 1$ is[6]

$$\begin{aligned} Prob(U_{i1} > U_{i0}) &= Prob(\tilde{\varepsilon}_i < \alpha_i(\ln p_1 - \ln p_0) + \beta_1 X_i + \xi_1) \\ &= F_{\tilde{\varepsilon}_i}(\alpha_i(\ln p_1 - \ln p_0) + \beta_1 X_i + \xi_1) \end{aligned}$$

$F_{\tilde{\varepsilon}_i}$ denotes the cumulative density function of $\tilde{\varepsilon}_i$.

Price salience will affect this choice probability by affecting the distribution $F_{\tilde{\varepsilon}_i}$. Based on the assumption that price salience reduces price noise, we can expect less noise under treated groups, i.e. smaller variance of $z_{i1} - z_{i0}$, and hence smaller variance of $\tilde{\varepsilon}_i$ and less dispersed distribution $F_{\tilde{\varepsilon}_i}$, which translates into a more responsive choice probability.

To be precise, suppose the variance of the price noise is dependent on whether the consumer is subjected to the salience treatment. If the consumer is in the control group, the variance is $Var(z_{i0} - z_{i1}|control) = \sigma_C^2$, and if he or she is in the treated group, the variance $Var(z_{i0} - z_{i1}|treated) = \sigma_T^2$. Variance

[6]This is a probability under the perspective of the econometrician, as the ecnometrician does not know $\tilde{\varepsilon}_i$ and need to view it as a random error

of the combined random error is

$$
\begin{aligned}
Var(\tilde{\varepsilon}_i|control) &= Var(\alpha_i(z_{i0} - z_{i1}) + \tilde{\epsilon}_{i0} - \tilde{\epsilon}_{i1}|control) \\
&= \alpha_i^2\sigma_C^2 + Var(\tilde{\epsilon}_{i0} - \tilde{\epsilon}_{i1}|control) \\
&+ 2\alpha_i Cov(z_{i0} - z_{i1}, \tilde{\epsilon}_{i0} - \tilde{\epsilon}_{i1}|control) \\
Var(\tilde{\varepsilon}_i|treated) &= \alpha_i^2\sigma_T^2 + Var(\tilde{\epsilon}_{i0} - \tilde{\epsilon}_{i1}|treated) \\
&+ 2\alpha_i Cov(z_{i0} - z_{i1}, \tilde{\epsilon}_{i0} - \tilde{\epsilon}_{i1}|treated)
\end{aligned}
$$

With treatment randomization, $Var(\tilde{\epsilon}_{i0} - \tilde{\epsilon}_{i1}|control) = Var(\tilde{\epsilon}_{i0} - \tilde{\epsilon}_{i1}|treated)$. Given the inclusion of a rich set of controls, consumer heterogeneity that may be correlated with the degree of price noise that the consumer faced should have been removed from the unobserved error $\tilde{\epsilon}$, and hence we will assume that $Cov(z_{i0} - z_{i1}, \tilde{\epsilon}_{i0} - \tilde{\epsilon}_{i1}|treated) = 0$ and $Cov(z_{i0} - z_{i1}, \tilde{\epsilon}_{i0} - \tilde{\epsilon}_{i1}|control) = 0$.

$$Var(\tilde{\varepsilon}_i|control) - Var(\tilde{\varepsilon}_i|treated) = \alpha_i^2(\sigma_C^2 - \sigma_T^2)$$

The difference in the variances of the combined error terms between groups is proportional to the reduction in the variance of the noise due to the applied treatment.

Note that we also need to normalize the scale of the utility by setting one of the variances σ_C^2 or σ_T^2 to a fixed number. The reason is, looking at the choice probability $Prob(\tilde{\varepsilon}_i < \alpha_i(\ln p_1 - \ln p_0) + \beta_1 X_i + \xi_1) = F_{\tilde{\varepsilon}_i}(\alpha_i(\ln p_1 - \ln p_0) + \beta_1 X_i + \xi_1)$, we can observe that scaling $\tilde{\varepsilon}_i, \alpha_i, \beta_1, \xi_1$ by the same constant will not affect the choice probability, and hence the scale is not identified. We will set $\sigma_C^2 = 1$.

Identification relies on two assumptions. Firstly, the stochastic term $\tilde{\varepsilon}$ is assumed to be uncorrelated with prices. This assumption helps identify the price sensitivity parameter α_i. This assumption is justified by the inclusion of a rich set of controls, including the station-product fixed effects, which will absorb all of the strategic pricing effects by the stations and the product-specific characteristics. Secondly, the price noise z is assumed to be uncorrelated with the unobserved utility ϵ. This assumption enables the identification of the noise reduction effect (reduction in variance of z) from the reduction in the variance of the stochastic term (variance of $\tilde{\varepsilon}$. One factor that can generate correlation between these two variables is consumer characteristics. Certain consumer types may have less access to information than others, and at same time have strong preference for a specific fuel type. To control for these factors, we include consumer demographics as well as vehicle characteristics to the model, alongside station-product fixed effects.

Intuitively, identification works as follows. Changes in fuel shares due to variation in prices identify price sensitivity α_i. Assuming the treated group is equally sensitive to price as the control group, knowing α_i will help pin down the mean value that consumers in the treated groups assign to each fuels. If treatment is effective, the price noise in the treated groups will be low, and consumers in these groups will simply follow the fuel associated with the highest identified mean value, giving rise to a more concentrated market share. If we observe a more evenly distributed market shares, it will suggest a higher degree of price noise left over, and hence a lower effectiveness of the salience treatments.

2.5 Estimation

In this section, we will proceed to discuss a general case of more than two choices and how the estimation is conducted.

Similar to the case of two choices, one alternative need to be chosen as the reference alternative, the mean utility of which will be set to zero to identify the location of the utility; and one of the variance is normalized to 1 to pin down the scale of the utility.

We will also need to specify the distribution for the combined error term $\tilde{\varepsilon}_{ij}$. A popular choice is the Type-1 Extreme value, which will result in a very tractable form for the choice probability. However, assuming Type-1 Extreme value distribution will lead to Independence of Irrelevant Alternatives, which will impose a strong restriction on the substitution pattern. This is particularly relevant to this study, as we are looking at choices between a group of conventional fuels (gasoline and mid-grade gasoline) and an alternative fuel (ethanol), and it is expected that the substitution within and between groups can be different. Furthermore, the combined error contains the price-ratio noise, and we should not assume that the consumer perception about gasoline and midgrade gasoline price differences is the same as the ethanol and gasoline price differences, as the prices of the first pair are highly correlated (correlation of 0.88), while the prices of the second pair are not (correlation of 0.48).

For these reasons, we will allow flexible correlation structures between the random utility of different fuels by specifying a multivariate normal distribution for the combined error terms. We will also allow the covariance matrix of this distribution to be shifted with the treatments.

To be specific, let denote $J = 0, 1, 2$ the set of fuels. Choose $j = 0$ to be the base alternative, so that we can set $\beta_0 = 0$ and $\xi_1 = 0$

$$U_{i0} = \alpha_i \ln p_0 + \tilde{\epsilon}_{i0}$$

$$U_{ij} = \alpha_i \ln p_j + X_i'\beta_j + \xi_j + \tilde{\epsilon}_{ij}, \quad j > 0$$

Relative utility:

$$\bar{U}_{ij} = U_{ij} - U_{i0} = \alpha_i(\ln p_j - \ln p_0) + X_i'\beta_j + \xi_j + (\tilde{\epsilon}_{ij} - \tilde{\epsilon}_{i0})$$

$$= \alpha_i(\ln p_j - \ln p_0) + X_i'\beta_j + \xi_j + \tilde{\varepsilon}_{ij}$$

Let denote the mean relative utility: $V_{ij} = \alpha_i(\ln p_j - \ln p_0) + X_i'\beta_j + \xi_j$. Consumer i will choose the base choice $j = 0$ when $\bar{U}_{i1} < 0$ and $\bar{U}_{i2} < 0$, or

$$\begin{bmatrix} \bar{U}_{i1} \\ \bar{U}_{i2} \end{bmatrix} < \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} \tilde{\varepsilon}_{i1} \\ \tilde{\varepsilon}_{i2} \end{bmatrix} < \begin{bmatrix} -V_{i1} \\ -V_{i2} \end{bmatrix}$$

Similarly, consumer i will choose choice $j = 1$ if $\bar{U}_{i1} > 0$ and $\bar{U}_{i1} > \bar{U}_{i2}$

$$\begin{bmatrix} -\bar{U}_{i1} \\ \bar{U}_{i2} - \bar{U}_{i1} \end{bmatrix} < \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} -\tilde{\varepsilon}_{i1} \\ \tilde{\varepsilon}_{i2} - \tilde{\varepsilon}_{i1} \end{bmatrix} < \begin{bmatrix} V_{i1} \\ V_{i1} - V_{i2} \end{bmatrix}$$

And he will choose $j = 2$ if

$$\begin{bmatrix} -\bar{U}_{i2} \\ \bar{U}_{i1} - \bar{U}_{i2} \end{bmatrix} < \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Leftrightarrow \begin{bmatrix} -\tilde{\varepsilon}_{i2} \\ \tilde{\varepsilon}_{i1} - \tilde{\varepsilon}_{i2} \end{bmatrix} < \begin{bmatrix} V_{i2} \\ V_{i2} - V_{i1} \end{bmatrix}$$

Assume that $\begin{bmatrix} \tilde{\varepsilon}_{i1} \\ \tilde{\varepsilon}_{i2} \end{bmatrix}$ follows a multivariate normal distribution $N(0, \Sigma_i)$ with the covariance matrix dependent on the treatment received, with $\Sigma_i = \Sigma_C$ for control group and $\Sigma_i = \Sigma_T$ for the treated group.

Based on the above condition, for the consumer to choose each of the choices, the choice probability can be calculated from with the given distribution. However, similar to the problem of scaling the utility for the two choices, we need to impose certain normalizations on the covariance structure for it to be identified.

We will adopt the numerical integration method described in Train (2003). The covariance will be decomposed using the Cholesky decomposition$\Sigma_C = S_C' S_C$ and $\Sigma_T = S_T' S_T$. To normalize the matrix, we will set the first element in the Cholesky decomposition for the control group S_C to 1, which essentially results in the variance of the relative utility between the first two choices being normalized to one. We will also use the GHK simulator as suggested in the Train (2003), with 150 draws from the Halton sequence for numerical integration.

The Maximum Likelihood estimator will be employed to estimate the parameters of the model. This model can be classified as a heteroskedastic multinomial probit model, as the covariance matrix varies across treatment groups.

There are 3 ways to consider ξ_{jt}. Firstly, one could argue that fuel is a rather homogeneous product, as gasoline from station A may not substantially differ from gasoline from station B. We will consider a model with ξ_{jt}. that depends on fuel but not stations. Secondly, one could regard ξ_{jt}. as a station-fuel fixed effect, keeping in mind that fixed effects may not be consistently estimated in a non-linear model with a small sample per station. In the data, we have between 54 to 270 consumers per station, with more than 108 in majority of the stations. Thirdly, we can use additional moments to back out

ξ_{jt}. from data. Similar to the method in Berry (1994), Goolsbee and Petrin (2004) and other works, the use of moments is aimed at matching the observed shares of ethanol and gasoline purchase at each station with the predicted share calculated from the model. As the results for the later approach is similar to the results of using the fixed effects, for brevity, we will only show the results from the station fixed effect models.

2.6 Results and Discussion

Table C.2 reports the estimation results for two specifications. Both specifications include consumer demographics (including gender, age, education, an indicator for expensive vehicle as a proxy for economic status and an indicator for extensive usage of vehicle) together with vehicle characteristics (including vehicle class, age, engine size, fuel tank and fuel efficiency). Specification (1) includes city fixed effects, while Specification (2) includes station fixed effects. The included variables, together with fixed effects are meant to capture the heterogeneity in consumer preference over non-price characteristics of the fuels. The standard errors are clustered at station level.

Price effect and price elasticity Even after conditioning on consumer and vehicle characteristics, effects of prices on fuel choices remain statistically and economically meaningful. According to Specification (2), a 10% increase in gasoline price will lower its own share of purchases by 10.8 percentage points (ppt), while increasing the propensity of ethanol by 10.5 ppt and mid-grade gasoline by 0.3 ppt. Given that the market share of gasoline in the sample is 60.4%, ethanol 33.8% and mid-grade gasoline 5.7%, and assuming that the short-run demand for fuel is inelastic, this implies an own-price elasticity of -1.8 for gasoline, and a gasoline cross-price elasticity of 3.1 for the market share of ethanol and 0.5 for mid-grade gasoline. The own-price elasticity of the market share for ethanol is estimated to be -3.8, and -3.8 for mid-grade gasoline.

City effects Differences in choices across cities are significant, even after accounting for differences in prices across cities. Consumers in Belo Horizonte are 22.4 ppt less likely to choose ethanol and 23.5 ppt more likely to choose gasoline than the consumers in Sao Paulo. The numbers for consumers in Recife are 15.4 ppt and 18.6 ppt respectively. At the same time, consumers at Curitiba have more or less the same distribution of purchases as consumers in Sao Paulo, all other factors remaining equal. As relative prices of ethanol over gasoline were consistently higher in Belo Horizonte and Recife than that in Sao Paulo and Curitiba during the experiments, one can argue that these differences in fuel choices may simply be due to the residual effects of different price levels between these two pairs of cities, where the logs of prices in the two specifications were not completely captured. However, these differences remain robust even after adding a higher order of price controls to allow for more flexible price effects. A more plausible explanation is the fact that Sao Paolo and Curitiba are capitals of the two ethanol-exporting states where consumers may possess bias towards their local products, as they care more about their local society and economy. This is consistent with the results in earlier studies.

Consumer demographics and heterogeneous preference The effects of consumer characteristics are significant and substantial. Female consumers appear to favor gasoline over both ethanol and mid-grade gasoline, with a higher propensity of 4.6 ppt to purchase gasoline over male consumers, while having a lower likelihood of purchasing the other two fuels. Older age is associated with the choice of mid-grade gasoline, with consumers above the age of 65 displaying a 6.1 ppt to 6.4 ppt higher propensity to choose mid-grade gasoline over consumers below the age of 25. The effect of education is noisy, but looking at Specification (1), which does not include the station fixed effect, we can say with some reservation that, all else equal, higher education is correlated with a lower likelihood of choosing ethanol, though the marginal effect of 4.4 ppt is only significant at the 10% level.

To interpret these effects as consumer heterogeneity in preference over non-price characteristics of the fuels, one need to keep in mind that, more often than not, a demographic variable can capture multiple aspects of consumer characteristics. Age and education can be highly correlated with income, occupation, and neighborhood of residence, which in turn can be associated with either technological savviness or health or environmental awareness. Nonetheless, some patterns can be interpreted with plausibility. Concerns about the high maintenance cost of ethanol and its lower energy content (which in turn requires more volume of purchase, thus resulting in longer stopping times and/or higher frequencies of visits to fuelling stations) is consistent with a lower share of ethanol among extensive-usage consumers. Familiarity with older technology (gasoline) and skepticism against new technology (ethanol)[7] is partly reflected in the 3.4 ppt lower propensity for the choice in ethanol among consumers aged above 65, compared to those aged below 25.

Vehicle characteristics The effects of vehicle characteristics are less noticeable. The differences in choice probability between most vehicle classes and that of the compact class (the omitted class) are insignificant, with the

[7]especially those who experienced the time of introduction of ethanol-capable vehicles, when the technology was still under-developed to adapt to physical and chemical properties of ethanol, which led to negative effects on its early reputation

exception of mid-size vehicles. This is probably due to the lack of statistical power, as the numbers of observations for other vehicle classes, apart from the mid-size and sub-compact classes, are limited.

Station fixed effects Next, we look at the effects of accounting for station-specific factors by comparing Specification (1), the one without station fixed effects, against Specification (2), the one with station fixed effects included. Adding station fixed effects slightly changes the estimated price marginal effect, and hence affecting elasticity. The estimated gasoline own-price elasticity changes from -2.58 (without station fixed effects) to -1.8 (with station fixed effect), whereas the change for ethanol is from -3.37 to -3.6, and for mid-grade gasoline from -6.1 to -3.8. Mid-grade gasoline prices also become more responsive to ethanol prices with station fixed effects, compared to the estimates without station fixed effects, where there seem to be no substitution between the two fuels.

These changes reflect the possibility that there are station-specific factors that correlate with prices. This includes operating cost, which correlates with the general price-level in the neighborhood of the stations, as well as unobserved consumer taste that firms will need to take into account when setting prices. These factors will be part of the station fixed effects.

We, however, do not observe significant changes in the estimated effects of the consumer characteristics before and after adding the station fixed effects. Thus, the unobserved consumer heterogeneity, if any, does not seem to have a strong correlation with the observed heterogeneity. This reduces concerns about endogeneity, as our set of controls seem to capture the most important dimensions of consumer heterogeneity in fuel choices.

Treatment effect We now turn to the main objective of this study and examine how price salience treatment affects the degree of consumer imperfect price information.

Panel A of Table C.4 reports the estimated variance of the stochastic term $\tilde{\epsilon}_{ijt}$ for each treatment group and for each of the fuel types. The variance of the control group for ethanol utility is set to 1 to anchor the scale of the utility, and all other variances should be interpreted relative to it.

To reiterate, the variance of the stochastic term $\tilde{\epsilon}_{ijt}$ includes the utility variation due to unobserved consumer heterogeneity, conditional upon observed consumer and vehicle characteristics as well as city effects and timing effects (Specification (2) further includes station fixed effects, which would have removed variation due to station-specific factors, especially potential pricing endogeneity by the station operators, from this variance). In addition, if price noise exists, its variance will also be constituted in the variance of these stochastic terms. The extent that treatments are randomized and their effects on the variance of the stochastic term is independent of the unobserved consumer heterogeneity, and the process of differencing across treatment groups will eliminate the first source of variation and reveal the effects of the increase in price salience on the strength of the price noise.

Panel B of Table C.4 carries out the differencing of the variance of the stochastic terms across treatment groups. For convenience of interpretation (as it is difficult to interpret the unit of $\tilde{\epsilon}$ itself), we report the difference in logarithms, so that the values will reflect the percentage changes in the stochastic variance due to the salience treatment. For ethanol, as the variance of the control group is set to 1, the logarithm is 0, which is essentially equivalent to taking the logarithm of each of the variances. We can also do differencing in level, and the results will be very similar, as most of the values are close to 1 (and $\ln(1 + x) \approx x$).

Effects on ethanol-gasoline relative utility The price ratio treatment is estimated to reduce the stochastic variation in ethanol utility by 12.7% (Specification (2)) or 15.3% (Specification (1)). The effect without including station fixed effects (15.3%) is significant at 5% level, while the effect with station fixed effects included is significant at 10%. The lower statistical significance after controlling for station effects comes from both a slightly lower precision (standard error of 0.076 vs 0.072) and a lower estimated coefficient (0.127 vs 0.153). However, the change in the estimated effect (a decrease of 2.6 ppt) is moderate and within the margin of error, and we take it as evidence that these estimates are reasonably robust against station-specific unobservables.

To give some perspective on the magnitude of this effect, we can look at the actual price variation over time and across cities. The variance of the log ethanol-gasoline adjusted price ratio in the data is 0.0106, and the coefficient of the log price in the utility is estimated to be -3.961 or -3.854 (depending on which specification, (1) or (2) to be used), i.e. $U_{ijt} = \alpha_i \ln(p_{ij}/km_{it}) + ... + \tilde{\epsilon}_{ijt}$ with $\alpha_i = -3.961$ or -3.854. All else equal, the contribution of the price variation to the utility variation is $\alpha_i^2 Var(\ln(p_{ij}/km_{it}))$, which is estimated to be $3.961^2 \times 0.0106 = 0.166$ or $3.854^2 \times 0.0106 = 0.157$. In other words, the contribution of price to variation in ethanol-gasoline relative utility is 15.7% to 16.6% the contribution of the stochastic terms $\tilde{\epsilon}_{ijt}$ (variance of which is 1). Consequently, contribution of the reduction in price noise due to price salience treatment (12.7% to 15.3%) is at the same magnitude as the contribution of price variation across four cities and over the period of study contributing to the utility. The best the treatments can do is eliminate the price noise, and this reduction can act as a lower bound for the amount of noise the consumers face. This suggests that, at a given point in time, the consumers face an amount of uncertainty about price with the same magnitude as the amount of uncertainty a person faces when picking a random price point from the price history across the four cities. Looking at within-city price variation can make this reduction in noise more pronounced, as the within-city variance in log adjusted ethanol-gasoline relative price is 0.0043[8], which implies a contribution to variation in ethanol-gasoline relative of only about 6.4%-6.7% that of the stochastic term, and half of the effect of the reduction in price noise due to the price-ratio flyer.

On the other hand, the effects of the km-per-50R$ are small and insignificant statistically. The magnitude of the effect is 1.6% reduction on the variance ethanol stochastic term, or almost zero if station fixed effects are included. The standard error is too large to give a definite conclusion whether the small effect is because of the ineffective treatment or the lack of statistical power. Nonetheless, the point estimates suggest that the effects of these treatments, if any, seem to be smaller than the effects of the price-ratio treatments.

Effects on midgrade gasoline-gasoline relative utility The effect of both treatments on the variance of the mid-grade gasoline stochastic term are small and insignificant. In the sample, prices of gasoline and mid-gasoline

[8]the variance of the residuals when regressing the log relative price on city dummies

are highly correlated (correlation of 0.88, compared to the correlation of 0.48 between gasoline and ethanol prices), and hence the proportional difference between them does not change much over time (the variance of the price ratio is 0.022, compared to a variance of 0.072 for the ethanol-gasoline price ratio). Knowing that the two prices move together, consumers will face less uncertainty in terms of price information comparing between the two grades of gasoline, as opposed to comparing between ethanol and gasoline. Thus, the lower effect of the salience treatment on gasoline-midgrade gasoline relative utility is consistent with our hypothesis that the treatments work through price noise reduction.

Effects of Salience treatment on choice probability We now turn to investigating the decision-making of better-informed consumers. First, we will compare the average probability of purchasing ethanol and gasoline when all consumers in the sample are subject to different treatments. To understand how these effects interact with relative prices, we will vary the price ratios in the following ways. The prices of gasoline and mid-grade gasoline are set at the same values as the original data. The prices of ethanol will be set as a fixed ratio multiplied by the gasoline prices. The ratio will be varied, and for each ratio, we will calculate the average gasoline share and ethanol share accordingly.

Figure D.4 plots the share of ethanol (top panel) and gasoline (bottom panel) when ethanol-gasoline price ratios vary. There are two lines in each panel, the dashed line is for the control group, and the solid is for the price-ratio flyer group. Figure D.5 plots the difference between these two lines, which essentially captures the effects of the price ratio treatment on the shares of ethanol and gasoline. As expected, the treatment is most effective in shifting fuel shares when the price-ratio is at the two ends of the range (i.e. either when price of ethanol is too high, or too low in comparison to the price of gasoline). The effect is rather small though. Even when the price ratio is as high as 0.85, the price-ratio treatment lowers the ethanol share by only 1.5ppt and increases the gasoline share by 1.5ppt.

In contrast, as discussed before, the effect of gender on gasoline share is estimated to be about 4.6ppt, college degree 4.4ppt (compared to "at most primary"). To further compare the effect of price salience treatments with the effect of consumer heterogeneity, Figure D.6 plots the ethanol and gasoline shares when the price ratio varies for two representative consumers, one a gasoline "fan" (female, age > 65, college graduate from Belo Horizonte) and the other an ethanol fan (male, age 25 to 40, primary-level education in Sao Paulo). The solid lines plot the shares for the consumers in the control group, while the dashed lines plot the shares when the consumers are subject to the price-ratio flyer treatment. Consumer heterogeneity is associated with a large change in shares for both types of consumers, while changes as a result of the treatments are rather minimal.

Heterogeneous treatment effects In Table C.5, we allow the variance of the random utility to be shifted by not only the treatments, but also vehicle prices (a proxy for economic status), vehicle usage, education and the volatility of recent price movement, one at a time. There are several interesting results in this table. First, column (5) shows that during periods of unstable price movement, consumers were subject to a much higher degree of noise. Moreover, the effects of price salience treatments were also higher. This is consistent with the theory of consumers facing noisy price information, as the noise is likely to become stronger with more fluctuation in the actual prices.

Secondly, columns (3) and (4) suggest that the treatments were more effective among low-usage drivers and college degree holders, although the differences were not statistically significant. High-usage consumers may have to fuel their vehicles more often, and thus care about fuel price movements more than low-usage consumers. Thus, further subjecting them to the price salience techniques would not likely be effective, as they tend to already be familiar with the actual prices.

2.7 Conclusion

In this paper, we employ a discrete choice model to study consumer fuel choices in Brazil and address a puzzling phenomenon in this market the fact that a large proportion of consumers choose to purchase a more expensive fuel despite equal access to a cheaper alternative. The explanation that this paper examines and puts forward is that consumers may not possess accurate information about prices, and thus at the point of purchase may not be aware of the cost effectiveness of each fuel. To test this explanation, an experiment in which a random set of consumers was treated with increased price salience (in the form of a flyer handed out before the purchase) was carried out. The econometric model allows for random noise in the prices perceived by the consumers to enter the utility function, and act as an additional source of randomness in the utility. By comparing the variance of the utility noise between the control and the treatment, the effect of the price salience treatment is revealed, from which we can gauge the accuracy of the price information that consumers have.

The results show that the price salience treatments, particularly the one with a display of information that is simpler and easier to understand (the price-ratio flyer), help substantially reduce the degree of price noise that the consumers face. The amount of variance deduction in price noise due to the price-ratio flyer is equivalent (in contribution to the random utility variation) to the variation of the actual price over the course of one year across four cities.

This effect, however, does not translate into substantial changes in choice probability between control and treated groups. The price ratio flyer only increases the probability of choosing ethanol by about 2%, even when ethanol is highly more favored over gasoline. This is reflects the fact that other non-price characteristics and consumer heterogeneity seem to play a significant role in consumer choices.

In sum, there is evidence that Brazilian consumers do not often possess accurate information about fuel prices, which leads to fuel choices that do not appear to be cost effective, even after controlling for consumer heterogeneity. However, the effect of such informational imperfection on fuel demand is rather small, compared to the effect of consumer heterogeneity on fuel demand. We have also provided several suggestive evidences for the mechanisms behind the informational imperfection that the consumers face. The cognitive cost of processing information seems to play an important role. Future research, with more fine-tuned experiments, could explore and formally test for such mechanisms.

Chapter 3

Discrete-continuous choice with fixed-payment preference: Automobile fuels in Brazil[1]

3.1 Introduction

The importance of fuel to the economy and the environment cannot be overemphasized. Unsurprisingly, fuel demand has been studied extensively in the existing literature. However, most studies, especially the earlier ones, focus mainly on conventional fuels and aggregate demand. Recently, concerns about environmental impact, the arrival of modern technologies, and various economic and political factors have paved the way for the introduction and promotion of alternative fuels to the market. Consequently, a new strand of literature is emerging, examining the way consumers choose between different fuels now that more alternatives are available to them. In addition, with more

[1]The data used in this chapter were collected and generously provided to me by my advisor, Professor Alberto Salvo. The chapter is co-authored with Professor Alberto Salvo.

and better data, from fieldwork as well as administrative resources, researchers are increasingly looking at data of finer disaggregation, e.g. consumer level or transactional level, to gain a richer understanding of how consumers value and choose fuels.

This paper attempts to bridge the gap between the literature on quantitative demand for fuel and the literature on discrete choice between different fuel alternatives, using a novel dataset on consumer choice at fueling stations collected during fieldwork in Brazil. The paper will present the development of a choice model that incorporates both the discrete choice between the conventional fuel (gasoline) and the alternative fuel (ethanol), and the continuous choice of how much fuel to purchase in each visit to the stations.

Considering continuous choice alongside the discrete choices helps capture and better explain consumer heterogeneity regarding preference for different fuels. We can think of quantity affecting the discrete choice via its interaction price; thus, considering the way consumers choose quantity will help account for some aspects of consumer heterogeneity in responding to fuel price fluctuations.

Continuous choice can also have policy implications. To the extent that many energy policies aim at reducing quantitative fuel demand, a policy that promotes a certain fuel can help or compromise another type of fuel, depending on how quantities and fuel types interact.

Given this background, this chapter is organized as follows. The next section will discuss related works on discrete-continuous choice and fuel demand. Following this, the data, which is the same data set used in Chapter 2, will be described. The empirical models and the estimation procedure will then be detailed. Finally, the results will be discussed and the chapter concluded.

3.2 Related works

3.2.1 Works on discrete-continuous choice

In many choice situations, consumers must decide not only which type of goods to consume or purchase (the discrete choice) but also how much of the chosen good to consume or purchase (the continuous choice). This paper will specifically discuss the choices between different fuel alternatives that motorists can choose at refueling stations in Brazil. In making these choices, the motorists need to decide whether to refuel their vehicles with gasoline, ethanol or mid-grade gasoline etc., and decide how much of the chosen fuel they should purchase. However, the discrete-continuous choice duality can also occur in the choices of traveling (in which mode to travel and how far), recreational activities (which activity to take up and how much time to spend on the activity), electric appliances (which appliance to use and how much electricity to be consumed by it) etc.

The two choices (discrete and continuous) are often interconnected; the outcome of one choice can affect the other. For example, due to differences in prices, the optimal quantities for different alternatives can be different; a consumer purchasing an expensive product may need to reduce the quantity of the purchase to match her marginal utility from consuming the product with the excessive cost of the purchase. On the other hand, a low price for a "low-quality" alternative can increase the quantity that the consumer can purchase given her budget to the point at which she will switch away from the expensive "high-quality" product.

Another important link between discrete and continuous choices is consumer types. Certain types of consumers may have higher preference for quantity, and at the same time prefer one product over the other. For example, in the case of fuel, a consumer with high vehicle usage may prefer gasoline to ethanol, as gasoline contains more energy for the same volume, and is often perceived as being less corrosive to engine components. Thus, a decrease in the price of ethanol may attract consumers to switch to ethanol, but the proportion of consumers with high vehicle usage among them would be small, and hence the increase in market share may not be proportionate to the increase in quantity demanded.

Dubin and McFadden (1984) and Hanemann (1984) provided a unified framework to study the both discrete and the continuous choice. Dubin and McFadden (1984) developed a specific choice model for residential electric appliance holdings and consumption, while Hanemann (1984) discussed the general approach to modeling consumer demand in discrete-continuous choice situations. Both approaches were based on the process of specifying an indirect utility function and using Roys identity to obtain the demand function. Specification-wise, Dubin and McFadden (1984) used a special functional form for the indirect utility to obtain a linear-linear form for the demand equation. Hanemann (1984) provided three specifications that would result in a tractable functional form for the demand equation. Estimation-wise, Dubin and McFadden (1984) proposed a two-stage procedure, with the first stage estimating the discrete-choice equation and the second stage using the fitted values from the first as instruments for the continuous choices in the demand equation. Hanemann (1984) discussed both the two-stage estimation and the full maximum likelihood estimation.

Regarding the link between the discrete choice and the continuous choice, Dubin and McFadden (1984) assumed mutual exclusion among alternatives so that the full decision-making process could be viewed as a two-stage sequential game with oneself: the first stage is to choose one alternative to purchase and the second stage is to choose the quantity of purchase conditional on the choice in the first stage. Hanemann (1984) started with a general utility defined over a vector of products, but employed several modeling features (perfect substitution, cross-product repackaging, or simply mutual exclusion due to logical or institutional reasons), so that the eventual optimal bundle consisted of only one alternative.

A related branch of literature focuses on multiple discrete-continuous choices. The models in this literature allow for multiple alternatives to be chosen in one purchase, and often involve specifying a utility function over a vector of quantities for all the alternatives. The functional form for the utility is chosen in a way that relaxes the perfect substitution feature of Hanemann (1984) to accommodate multi-product purchases, but still generates corner solutions to accommodate cases in which some products are not chosen in the bundle. Such models have been employed by Kim et al. (2002) to study consumer demand for variety, Bhat (2005) to study time use, and Ahn et al. (2008) to study the ownership and use of alternative-fuel vehicles.

In the data, there are consumers who purchased more than one type of fuel, but the number is very small compared with that of the majority, who only purchased one type of fuel. As a simplification, it is assumed that only one alternative should be chosen in a visit to the refueling station. Thus, we will adopt a specification similar to those in the work of Hanemann (1984) and Dubin and McFadden (1984), as opposed to those from the multiple-discrete choice literature. Specifically, the model used in this paper will give rise to the indirect utility and demand, like Equation (3.21a) and (3.21b) in Hanemann (1984).

Nonetheless, the models from the multiple discrete-choice literature highlight the fact that specifying the utility function can help resolve various irregularities in the data, as such a utility function can be used to consistently model consumer choices under different constraints or scenarios.

3.2.2 Works on demand for automobile fuel

The demand for fuel, especially gasoline, has been extensively studied. Various methodologies have been applied on various types of data (cross-sectional, time series, panel of macro data, micro data etc.) to obtain estimates for the price and income elasticity of gasoline demand in the short run as well as in the long run. Well-known surveys in this body of work include those of Drollas (1984), Dahl and Sterner (1991), Goodwin (1992), Espey (1998), Graham and Glaister (2002) and Pouliot and Babcock (2014). Graham and Glaister (2002), after a comprehensive review, suggested that the long-run price elasticity of gasoline demand typically falls in the range of -0.6 to -0.8, and the short run elasticity in the range of -0.2 to -0.3, although these numbers vary greatly across countries and time periods. Hughes, Knittel, and Sperling (Hughes et al.) have reported that the short-run price elasticity in the U.S. has shifted from a range of -0.21 to -0.34 in the period 19751980 to a range of -0.034 to -0.077 in the period 20012006.

One important conclusion from this literature is that there is a distinction between short-run and long-run demand response to changes in fuel prices. In the long run, consumers have a much wider range of adjustments that they can make to accommodate a given change, from adjusting their travelling routes and modes to switching their vehicles to those of different fuel-economy classes, and even to relocating or changing their jobs. Espey (1998) summarizes the long-run/short-run classification as follows: models with time dimension and lagged structures capture both short-run and long-run elasticity estimates. Models that include controls for vehicle ownership and/or fuel efficiency tend to capture short-run to medium-run elasticity. Estimates from a static model with no controls over vehicle ownership or fuel efficiency are ambiguous and dependent on the length of the temporal dimension and the variation in prices and incomes along the cross-sectional dimension of the data used for estimation.

In this paper, experimental micro-data will be used at the transactional level of disaggregation one observation is one purchase by a motorist at a refueling station. The estimation makes use of both temporal and cross-sectional (across cities, and across stations within a city) variations in prices. The experiments were carried out over a period of over one year, and hence, according to the literature, the estimation is better qualified to capture short-run demand. The cross-sectional variation across stations and cities is substantial, and spatial variation in demand may reflect consumers adjusting their characteristics and/or traveling patterns to the local prices over an extended period. However, given that the data is from only one country and most stations are in the urban area, the spatial variation is comparable with what the literature considers short-run to medium-run dynamics, rather than long-run dynamics.

3.2.3 Works on demand for alternative fuel

The works discussed in the previous section are concerned mainly with petroleum based fuels, especially gasoline, as they are still the main source of energy for vehicles around the world. However, there is also a smaller strand of literature studying demand for alternative fuels. Anderson (2012) examined the demand for corn-based ethanol using data from Minnesota, U.S., and found that many consumers were willing to pay a premium for alternative fuel. As a substitute for gasoline (E10), ethanol was estimated to have an own-price elasticity of negative 3.2-3.8 and a gasoline-price elasticity of 2.33.2. Salvo and Huse

(2013) investigated the choice between sugarcane-based ethanol and gasoline in six cities in Brazil. Their findings included the discovery that a substantial proportion of consumers chose to purchase the more expensive fuel, and the discovery of substantial consumer heterogeneity in the choice between the two fuels.

One conclusion from this literature is that consumers do not always view ethanol as the perfect substitute for gasoline as their choices do not always align with the most cost-effective source of energy for their vehicles. There are several potential explanations for this behavior, one being that consumers may value the fuel beyond its energy content. For example, concerns about technologies, environments, and societal factors were shown to significantly influence consumer fuel choices in the work of Salvo and Huse (2013). Consumers may also exhibit state-dependence in their decision process, preventing them from flexibly responding to new prices. For example, Salvo (2016) considered a model of limited attention, with consumers having a consideration set that may exclude certain choices. Noisy or inaccurate information about prices may also prevent consumers from achieving optimal fuel choices.

3.3 Data

3.3.1 Data source and Experiments

For completeness, a summary of the data is included here. The data on payments will then be discussed in detail, and used to calculate the quantity variables for the continuous choice of the empirical model.

The data was collected from a series of experiments conducted at 53 fueling stations in four cities in Brazil during the period between March 2011 and April 2012. There were a total of 193 visits to the stations, with each station visited up to five times, resulting in a total of 10,422 subjects being surveyed.

Each station visit was conducted within a day. The timings of the visits were varied so that we would have data for various times of the day and different days of the week. By design, the first 18 consumers were assigned to control, the next 18 assigned to one treatment group, and the subsequent 18 assigned to the other treatment group.

Control group Once the consumer had placed his order with the stations attendant and the vehicle was being serviced, the enumerator would survey the consumer to for the demographic and other information.

Two treatments by station visit A treated subject would, prior to placing his order, hear one of the following statements, from either the attendant or the enumerator. The statement was consistent with the actual price levels for regular-grade gasoline and ethanol posted at the stations pump on the day of the visit:

- *Hoje a gasolina está mais vantajosa, veja aqui,* loosely translated as "Today gasoline is more advantageous / the better deal, see here,"

- *Hoje o álcool está mais vantajoso, veja aqui,* translated as "Today ethanol is more advantageous / the better deal, see here,"

- *Hoje a gasolina e o álcool estão com rendimento parecido, veja aqui,* translated as "Today gasoline and ethanol offer similar yields / similar deals, see here."

As he heard the verbal statement on the relative price conditions at the station that day, a subject would be handed a flyer similar to that shown in Figure D.2, for the price ratio relative to 70% treatment, or to that shown in Figure D.3, for the km-per-R$50 treatment. The illustrated flyer corresponds to a station visited in Sao Paulo on June 13, 2011 in which $(p_e, p_g) = (1.649; 2.499)$. To clarify, some treated consumers were handed a price-ratio flyer and others were treated with a km-per-R$50 flyer, and in each case the flyer was consistent with the verbal statement that introduced it.

The price-ratio flyer (Figure D.2) gave a thumbs-up alongside Mais vantagem status to gasoline when $p_e/p_g \geq 0.705$ and a thumbs-up to ethanol when $p_e/p_g \leq 0.695$, and stated that both fuels offered similar yields (Rendimento parecido) otherwise. The flyer also reminded subjects of the media-reported parity threshold, stating that specialists advised that when this price ratio (between ethanol and gasoline) was (i) lower than 70%, ethanol was more advantageous; and (ii) higher than 70%, gasoline was more advantageous.

The km-per-R$50 flyer (Figure D.3) gave a thumbs-up to gasoline when the distance to be travelled on R$50 of regular gasoline purchased at the station was expected to exceed, by some margin, the distance to be travelled on R$50 of regular ethanol, and a thumbs-up to ethanol for the opposite situation, and was neutral when gasoline and ethanol offered similar yields. Specifically, a table in the flyer compared the expected distances for ethanol and gasoline for three of the most popular vehicle engine sizes: 1.0, 1.4, and 1.8 liters (absolute fuel economy generally declines with engine size).

3.3.2 The choice of quantities purchased

Figure F.1 shows the distribution of payments observed from the data. Three observations can be made.

First, there is a marked spike in the distribution at 50R$. Approximately one-fifth of the consumers in the data chose to spend exactly 50R$ on the fuel. This spike stands out not only because it is the point with the highest density in the distribution, but also, and to a greater degree, because its density is several times higher than the neighborhood density. Being the highest density point can be well explained by a fitting price distribution that led to quantities chosen by consumers that fall into a range centering around 50R$. Being distinct from neighborhood density suggests a special preference for that payment that goes beyond the preference for quantity.

Second, there are spikes at payments that were in multiples of 10R$. The number of consumers who paid in multiples of 10R$ is noticeably higher than that of customers who paid with odd payments. This can be observed in the right panel of the same figure, in which the histogram of the last digit of payments is plotted. Approximately three-quarters of consumers (including the 19% that chose 50R$) paid with payments in multiples of 10R$. There are also spikes at multiples of 5R$. Although those are much less noticeable than the mutiples-of-10R$ payments, making up only 8% of the data, they are still significantly higher than the number of payments ending with digits other than 0 or 5. The most plausible explanation for this is the availability of and preference for banknotes in different denominations.

Third, there is still a small number of payments that were not in multiples of 10R$ or 5R$, and these are spread out evenly across the range of observed payments. In total, they make up 17

Payments of exactly 50R$ in value will be called fixed payments, payments in multiples of 5R$ or 10R$ will be called regular payments, and other payments will be called odd payments. If further distinctions are required, regular payments in multiples of 10R$ will be called regular-0 payments and other regular payments will be called regular-5 payments.

Odd payments versus regular payments

The choice of odd payments can be rather inconvenient in the given context. At refueling stations across Brazil, most of the transactions were conducted manually, with the actual refueling performed by a human attendantattendant[2] who would also process the payments, which would most likely be in cash. As a result, a significant cognitive and time cost, to both parties, would have been incurred for keeping track of and preparing for the right volume of fuel and the right number of notes and coins, if available at all, to be exchanged. Such a cost would reasonably outweigh the change in utility from a small deviation from the optimal quantity, in order to avoid such an odd payment. Thus, the sizable proportion of odd payments suggests that there are other considerations that need to be considered.

Figure F.2 plots the distribution of the quantities of purchase as propor-

[2] http://www.planalto.gov.br/ccivil_03/Leis/L9956.htm

tions of the consumers fuel tanks for three different subsamples in the data: a subsample of odd payments (top-left panel), a subsample of regular-5 payments (top-right panel), and a subsample of regular-0 payments (bottom panel, and excluding fixed payments). The first panel shows that odd-payment consumers tended to fill the fuel tanks of their vehicles more than halfway. Likewise, the median and the mode of the distribution for odd payments is approximately 0.7, implying that the typical odd-payment consumers would fill more than two-thirds of the fuel tank with their purchase. This is in stark contrast with the other two groups of consumers who paid with regular payments. Consumers with regular payments, regardless of whether the payments were in multiples of 10R$ or not, tended to fill less than half of their fuel tanks, with the typical person in this group purchasing about one-quarter tanks worth of fuel.

We can make a reasonable assumption that the typical consumer would not visit the fueling stations with too much or too little fuel left in their tank. With too little fuel, their vehicles would run the risk of running out before an opportunity to refill, while with too much fuel the drivers would waste their time on too many visits to the fueling stations. Under this assumption, odd-payment consumers are likely to fill their tanks completely, while regular-payment consumers are not likely to. During the experiment, we did not survey for, and hence did not observe the amount of fuel left in the tank prior to the purchases, I argue that odd payment is the best indicator that consumers will order a full tank, based on the above assumption.

50R$ fixed-payments versus other regular payments

Figure F.3 plots two histograms side by side. Both are histograms of total payments, but one is for the subsample of regular-0 payments, including the fixed payments of 50R$, and the other is for the subsample of regular-5 payments. The rationale for plotting these two histograms side by side is that the difference in frequency of choosing regular-0 payments over regular-5 payments is mostly due to the preference for or the availability of banknotes in denominations of 10, 20, 50, and 100 reais, compared with that of banknotes in denominations of 5 reais.

Assuming this effect is independent of the preference for quantity, the difference between adjacent regular-0 and regular-5 payments will be similar throughout the range of the observed payments. For example, the preference for 25R$ will be close to the preference for 20R$, subtracting the banknote effect, and the preference for 45R$ will be close to the preference for 40R$, subtracting the banknote effect. As a result, the difference between choosing 20 vs 40 would be similar to the difference between choosing 25 vs 45, because the banknote effect is canceled out by differencing. Drawing a histogram will normalize the total density to 1, and thus eliminate the effect of banknote-specific preference. Consequently, if the above assumption is true, the two histograms should be similarly shaped.

In fact, the plots do show that the two histograms are similar, except for the abnormally high concentration at 50. This highlights the fact that there is a special preference for fixed payment at 50 that goes beyond the preference for regular-0 payments over regular-5 payments. We can use the densities at 45R$ and 55R$ as good proxies for what the density at 50R$ could have been in the absence of the special preference for the fixed payment at 50R$, and approximate that the density at 50R$ has been inflated almost five times thanks to the special preference for the fixed payment. This pronounced irregularity in the distribution of payments calls for a special treatment in the empirical model.

The similar shapes of the two histograms support the assumption that the note-specific preference is independent of the preference for quantity. In fact, we can also look at the histograms of the liters of purchase as a proportion of the fuel tank and make the same observation: the two groups of consumers, i.e. those making regular-0 and regular-5 payments, are, along many dimensions, similar to each other.

Consumer Heterogeneity and Payments

As the final investigation into the ways consumers pay, the dummy variables for different types of payments on consumer demographics and vehicle characteristics are regressed. The estimation results are shown in Table E.2.

The first two columns use a dummy for odd payment as the dependent variables and estimate using all observations in the data; this is essentially a comparison between odd payments and regular payments. Age, vehicle prices, vehicle usage, and ownership of non-common vehicle brands tended to increase the likelihood of odd payments.

These variables are correlated with high income and usage, which implies high quantity demand, increasing the likelihood of hitting the ceiling of the vehicle tank.

This is consistent with the fact that odd payments tended to be made to fill up to full tank. Living in Belo Horizonte and Recife decreased the likelihood of odd payments; these are cities in which ethanol was consistently more expensive throughout the sample period, which would have led to fewer purchases of ethanol and more of gasoline. As gasoline takes up less volume than ethanol while providing the same amount of energy, this would have meant a lower chance of consumers filling a full tank for the same amount of distance traveled.

The third and fourth columns use a dummy for regular-0 payments as the dependent variables and estimate using observations of regular payments; this is essentially a comparison between regular-0 and regular-5 payments. Age, vehicle prices, vehicle usage, and ownership of non-common vehicle brands tend to increase the likelihood of odd payments, but otherwise there is no clear pattern in the way consumers chose between these two payment types.

The fifth and sixth columns compare fixed payments of 50R$ against neighborhood payments (payments from 40R$ to 60R$). Again, there is no clear pattern of how consumers chose 50R$ over other neighborhood payments.

The above observations can be interpreted as evidence supporting the view that preferences for specific banknotes are rather random and less likely to be correlated with unobservable factors that can influence prices. As discussed in detail later, these patterns in payments will be used to classify consumers into three groups.

3.4 Model

3.4.1 Utility

Consumers receive utility from purchasing a certain amount of fuel. Throughout this analysis, I remain agnostic about the nature of the continuous choice, only assuming that consumers prefer more to less. Therefore, with utility maximization, price would enter as a determinant of the chosen amount, keeping in mind that the purchased quantity of fuel at refueling stations may reflect not only the miles of travel that can be spent on the fuel but also the potential increase in the time period before the next refuel, and hence the potential decrease in the number of trips to fueling stations, or the possibility of hedging and speculation about future fuel prices etc.

To obtain tractable functional form for the demand function, an isoelastic specification for the utility function is employed:

$$U_{ijt} = \frac{\theta_{ijt}^{\frac{1}{\varepsilon}} km_j^{1-\frac{1}{\varepsilon}}}{1-\frac{1}{\varepsilon}} - p_{ijt}km_j + X_{it}^u \beta_j^u + \nu_{ijt} \tag{3.1}$$

In the above equation, index i indicates a consumer, index j a product (in this case, a type of fuel), and index t a market (or, in this case, a station visit). km_{ijt} denotes the distance of travel (in km) that consumer i can spend on the purchase of fuel j. p_{ijt} denotes the price per km (R$/km) of travel on fuel j, which is equal to the price per liter of fuel j divided by the fuel efficiency of consumer is vehicle. X_{it}^u denotes the observed consumer characteristics that directly affect discrete fuel choice. ε is a preference parameter that will eventually determine the price elasticity of demand, θ_{ijt} is another preference parameter that will determine the level of demand, and ν_{ijt} is the idiosyncratic utility shocks.

In the data, payments were observed, and quantities were calculated accordingly based on the fuel prices. Because of this, and because of the irregularities in the ways consumers pay, as mentioned in Section 3, it is more convenient for estimation purposes to work with payments rather than quantity. Theoretically, given the observed prices, using either one of the variables is equivalent to using the other. Let $E_{jt} = p_{jt}x_j$, the payment made by the consumer for fuel j, and rewrite the utility in terms of this payment:

$$U_{ijt} = \frac{\theta_i^{\frac{1}{\varepsilon}} E_{ijt}^{1-\frac{1}{\varepsilon}} p_{ijt}^{\frac{1}{\varepsilon}-1}}{1-\frac{1}{\varepsilon}} - E_{ijt} + X_{it}^u \beta_j^u + \nu_{ijt}$$

3.4.2 Income effect and other heterogeneity in consumer preference for quantity

Equation 3.1 can be understood as the result of the following utility maximization under a constrained budget. A consumer has to choose to spend her income y on either fuel or an outside good, the quantities of which will be denoted by x and z. She will receive a utility that is quasi-linear in the outside good: $U(x, z) = u(x) + z$ with $u(x) = \frac{\theta^{\frac{1}{\varepsilon}} x^{1-\frac{1}{\varepsilon}}}{1-\frac{1}{\varepsilon}}$, the first term in Equation 3.1. This implies a constant marginal utility of 1 from the outside good and a decreasing marginal utility from the fuel. The price of the outside good is normalized to 1, giving rise to the following budget constraint: $px + z \leq y$. Assuming an interior solution, we can substitute the constraint $z = y - px$ into the quasi-linear utility and derive that the consumer will receive $U(x) = u(x) - px + y$ from purchasing x amount of fuel.

The quasi-linear specification helps keep the solutions for demand and indirect utility tractable and simplify estimation substantially. However, it assumes away the income effect on fuel demand, and hence imposes a serious restriction on consumer behaviors. Most of the works on fuel demand have found a positive income elasticity of gasoline demand in the range of 1.1 to 1.3 in the long run or 0.35 to 0.55 in the short run (Espey, 1998), and omitting income from the demand equation can be considered a misspecification (Dahl and Sterner, 1991).

There are some other alternative functional forms that have a closed-form solution and result in a positive income effect. For example, $U(x,z) = \theta x^{\rho} + z^{\rho}$ will result in demand: $x = \frac{y}{p+(\theta/p)^{\frac{1}{\rho-1}}}$, but this demand still constrains the income elasticity to exactly 1 while complicating the relationship between quantity and prices.

Another strategy is therefore employed here. Consider a more general utility function. We still maintain separability but relax quasi-linearity: $U(x,z) = U(x) + f(z)$. Substituting the budget constraint $px + z = y$ into the utility function we have: $U(x) = U(x, y - px) = U(x) + f(y - px)$. It is reasonable to assume that the typical share of fuel expenditure over total consumer expenditure is small, and we can approximate the second term using Taylors expansion as follows: $f(y - px) \approx f(y) - f'(y)px$. Thus, the original utility can be approximated by: $U(x) \approx U(x) + f(y) - f'(y)px = f'(y)\left[\frac{U(x)}{f'(y)} - px + \frac{f(y)}{f'(y)}\right]$. As scaling utility by a constant factor does not affect the preference, we can rewrite the utility as $\left[\frac{U(x)}{f'(y)} - px + \frac{f(y)}{f'(y)}\right]$. The implied demand is $x = U'^{-1}(pf'(y))$, which is dependent on both prices and income.

The above analysis suggests that one way to induce the income effect in a quasi-linear specification is to scale the utility by a function of income. We can employ this strategy for the utility specified in Equation 3.1 by allowing income to shift θ, i.e. $\theta_{it} = \theta(y_{it})$. This, however, should not be interpreted as allowing income to affect preference, but rather should be understood only as a transformation of the net utility to simplify the demand and the indirect utility function while allowing for positive income elasticity. A consequence of this transformation is that θ_{it} can no longer be viewed purely as a preference parameter, because it is shifted with income.

Income was not observed. Therefore, demographics such as age, education, city, and vehicle price and usage will be used to proxy for income. In other words, θ_{it} will be written as a function of demographics.

Allowing θ to be shifted with demographics and other consumer characteristics also helps capture consumer heterogeneity along other dimensions.

Occupation and personal need may require different consumers to have different preferences regarding distance of travel. Vehicle characteristics in terms of performance and emission can also be factored into the way consumers evaluate the cost and benefit of driving. This heterogeneity in quantity preference can also indirectly affect the discrete choice between fuels through its interaction with prices, and accounting for this can also be viewed as accounting for the way consumers react to prices when making the discrete choices.

Finally, consumers may value the same distance of travel with different fuels differently, as operating and maintenance costs can vary across fuels. As discussed before, the chemical and physical characteristics of ethanol and gasoline, and even midgrade gasoline and gasoline, are different, resulting in differences in wear and tear of the parts and materials of the engine, as well as its performance. Emissions also differ between fuels, and consumers with concerns about health and/or the environment may internalize this cost and factor it into their valuation of each km of travel. For these reasons, I will also allow the preference parameters to vary across fuels by incorporating a fuel fixed effect into θ_{ijt} :

$$\ln \theta_{ijt} = \alpha_j + X_{it}\gamma + \eta_{it}$$

This is similar to an alternative specification, where we allow the effect of the payment on utility to differ across fuels, i.e. not always decrease utility -1 for each R$ paid:

$$U_{ijt} = \frac{\theta_{it}^{1/\varepsilon} km_{ijt}^{1-1/\varepsilon}}{1-1/\varepsilon} - \pi_j p_{ijt} km_{ijt} + X_{it}^u \beta_j^u + \nu_{ijt}$$

The similarity is because, if the quantity is chosen optimally, the indirect utility will be $\frac{\theta_{it}\pi_j^{1-\varepsilon} p_{ijt}^{1-\varepsilon}}{\varepsilon-1} + X_{it}^u \beta_j^u + \nu_{ijt}$, which is equivalent to having $\theta_{ijt} = \theta_{it}\pi_j^{1-\varepsilon}$. I choose to fix the coefficient of the payment at -1 and allow the preference parameters to vary across fuels because the payment can be interpreted as the opportunity cost of consuming the outside , which should be valued consistently across fuels. Furthermore, as payment enters the equation linearly with a coefficient of -1, the "unit" of the utility is pinned to Brazilian reais, giving a sense of how economically significant the estimates are.

3.4.3 Payment types

As discussed in the previous section, payments made by consumers can be classified into three important types: regular payments, odd payments, and fixed payments; these payment types seem to capture two special behaviors: preference for fixed payments and preference for full-tank purchases. To capture this pattern of payments and behaviors, the model will assume that there are three types of consumers visiting fueling stations: flexible-payment consumers, fixed-payment consumers, and full-tank consumers. These types are determined by looking at the type of payment the consumers use. Specifically, regular payments indicate flexible-payment consumers, fixed-payments of 50R$ indicate fixed-payment consumers, and odd payments indicate full-tank consumers.

Flexible payment consumers The first group will choose the quantity freely to achieve optimal utility for each fuel type :

$$\ln km^*_{ijt} = -\varepsilon \ln p_{ijt} + \ln \theta_{ijt}$$

The above equation can also be written in terms of the optimal expenditure spent on the fuel, E^*_{ijt}, which is same as the total payment for the purchase:

$$\ln E^*_{ijt} = (1 - \varepsilon) \ln p_{ijt} + \ln \theta_i$$

As discussed in the previous, let θ_i depend on observed consumer characteristics and a random idiosyncratic demand shock η_i:

$$\ln \theta_{ijt} = \alpha_j + X^e_{it} \beta^e + \eta_{it}$$

Rewrite the above optimal expenditure as:

$$\ln E^*_{ijt} = (1 - \varepsilon) \ln p_{ijt} + X^e_i \beta^e + \eta_i$$

The utility received from purchasing the optimal amount of fuel j:

$$U^*_{ijt} = \underbrace{\frac{E^*_{ijt}}{\varepsilon - 1} + X^u_i \beta^u_j}_{V_{ijt}} + \nu_{ijt}$$

Fixed-payment consumers Fixed-payment consumers would commit to the fixed payment $\overline{E} = 50$, choosing the fuel that gives them the best utility with that payment. The fixed payment, together with prices, determines how much distance the consumer can buy with each fuel: $\overline{km}_{ijt} = \overline{E}/p_{ijt}$. The only choice for these consumers is which fuel to purchase. The utility that the consumer receives is

$$U_{ijt} = \frac{\theta_i^{\frac{1}{\varepsilon}} \overline{km}_{ijt}^{1-\frac{1}{\varepsilon}}}{1 - \frac{1}{\varepsilon}} - \overline{E} + X^u_{it}\beta^u_j + \nu_{ijt}$$

Full-tank consumers Full-tank consumers. Full-tank consumers commit to fill their tanks, by whatever amount is lacking . Letting this amount be lt_{it}, this, together with the fuel efficiency $(km/lt)_{ijt}$ of the vehicle, will determine the distance the consumers can travel, $\tilde{km}_{ijt} = lt_{it}(km/lt)_{ijt}$, and also the payment, $\tilde{E}_{ijt} = p_{ijt}\tilde{km}_{ijt}$. The utility is therefore

$$U_{ijt} = \frac{\theta_i^{\frac{1}{\varepsilon}} \tilde{km}_{ijt}^{1-\frac{1}{\varepsilon}}}{1 - \frac{1}{\varepsilon}} - \tilde{E} + X^u_{it}\beta^u_j + \nu_{ijt}$$

Choice probability . In general, the utility of all the types can be expressed by $U_{ijt} = V_{ijt} + \nu_{ijt}$. As with other discrete choices, we need to choose one choice to be the base alternative to pin down the location of the utility. The β_j^u for the alternative will be set to 0, and the other utility should be interpreted as a relative utility.

Consumer i chooses fuel j when $U_{ijt} > U_{ikt}$ for all k in the choice of station t. Equivalently, the probability of choosing j is

$$P_{ijt} = Prob(U_{ijt} > U_{ikt}, \forall k) = Prob(\nu_{ijt} - \nu ikt > V_{ikt} - V_{ijt})$$

We need to specify a distribution for ν_{ijt} to obtain the functional form for this probability. This will be discussed in the next section, together with the estimation method used.

3.4.4 Identification

Maximum likelihood will be used for estimation. As the model is highly non-linear, identification comes from a combination of functional form assumptions and variations in prices and choices .

What is estimated is essentially a system of two equations, one for the discrete choice and the other for the continuous choice:

$$\ln x^*_{ijt} = -\varepsilon \ln p_{jt} + \ln \theta_{ijt} \tag{3.2}$$

$$U_{ijt} = \theta_{ijt} \frac{p_{jt}^{1-\varepsilon}}{\varepsilon - 1} + X^u_{it}\beta^u_j + \nu_{ijt} \tag{3.3}$$

In the absence of the discrete choice, Equation 3.2 could be used as a linear regression to estimate the price elasticity of demand (assuming exogenous variation in prices is justified), and ε will be identified from the response, in terms of quantity of purchase, to changes in fuel prices. θ_{ijt} is determined by the level of the quantity of purchase, i.e. a high level of purchase will imply high θ_{ijt} . Note that $\ln \theta_{ijt}$ is specified as a function of demographics, and hence the coefficients of the demographics inside θ_{ijt} are determined by the average purchase of the respective demographic group.

A commonly used specification for the random utility in a discrete model is as follows: $U_{ij} = \alpha p_{ij} + X_i\beta_j + \nu_{ij}$ with α capturing the sensitivity of choice probability with respect to price change. We can easily draw parallels between this utility and the one specified in Equation 3.3. Conditional on the demand response along the intensive margin ε, price changes will translate directly to changes in the term $\frac{p_{jt}^{1-\epsilon}}{\epsilon-1}$, which will in turn affect utility and choice probability, but only after going through θ_{ijt}. Thus, θ_{ijt} here has a similar role to that of α in the previously mentioned discrete-choice model: it reflects the sensitivity of choice probability in response to price changes. Thus, one important source of identification for θ_{ijt} is shifts in the market shares of different fuels under varying prices.

The effect of ε in the discrete-choice equation is rather subtle . We need to look at the marginal utility: $\frac{\partial U_{ijt}}{\partial p_{jt}} = -\theta_{ijt} p_{jt}^{-\varepsilon}$. The magnitude of the marginal utility decreases with price, and the rate of the decrease depends on ε. With high ε, the marginal utility will decrease quickly, and so the response of the discrete choices to the same percentage change in prices will be very different at different price levels.

In short, variation in the continuous choice in response to price variation identifies ε, while variation in the discrete choice in response to price variation helps identify θ_{ijt}. In addition, the level of purchase also helps identify θ_{ijt}, while the variation in discrete-choice responses to the same percentage change in price at different price points helps identify ε.

Self-selection in types Consumer types can be endogenous if consumers self-select into different types when price changes. It can be the case that when fuel prices rise above some certain threshold, a consumer who usually pays a fixed amount for fuel in each visit can be become price sensitive again and adjust the payment according to the prices. To examine this possibility, I regress the dummy variables for each consumer type on the fuel prices. A simple regression between types and fuel prices indicate strong correlation between the two sets of variables, especially between consumer types and the price of ethanol. However, after controlling for city fixed effects, the correlation becomes smaller and insignificant. This result suggests that the correlation between consumer type and fuel prices in the simple regression mostly reflect the difference in type composition between cities, rather than a self-selection prices. It is plausible that consumers in different cities develop different habits in paying for fuels. I can allow for the composition of consumer type to vary with city, or with any consumer characteristics.

Statistical test One question is whether we can have a formal test of the current model with three types against an alternative model with a different number of types. The test can be done if I allow for a random mixture of types, and test if the probability of a certain type is zero or not. In this version of the paper, I assume the consumer type is fixed and observable in order to simplify the estimation, and hence do not conduct the tests.

3.5 Estimation

The observed data include fuel prices, consumer demographics, consumers choices of fuel, and the payments made for the chosen fuel, from which the quantity of purchase can be calculated. Equations 3.3 and 3.2 capture the essence of the system of equations that will be estimated. However, it should be noted that for Equation 3.2 only the quantity for the chosen fuel is observed; the quantities for the other fuels are missing, and this selection bias could lead to incorrect estimates if the discrete choice were not considered. As discussed previously, we can rewrite Equations 3.3 and 3.2 in terms of payments and prices to facilitate estimation. θ_{ijt} should also be written as a function of demographics, fuel fixed effects, and idiosyncratic demand shocks $\theta_{ijt} = \alpha_j + X_{it}^e \beta^e + \eta_{ijt}$. Let j_i^* denote the fuel chosen by consumer j, $p_{it}^* = p_{j_i^* t}$ be the price of the chosen fuel, and $E_{it}^* = x_{ij_i^* t}^* p_{it}^*$ be the payment made for the chosen fuel.

Estimation equation For flexible-payment consumers, the equations can be written as

$$\ln E_{it}^* = (1-\varepsilon) p_{it}^* + \alpha_j + X_{it}^e \beta^e + \eta_{it} \tag{3.4}$$

$$U_{ijt} = \frac{E_{it}^* \left(\frac{p_{jt}}{p_{it}^*}\right)^{1-\varepsilon} e^{\alpha_j - \alpha_{j_i^*}}}{\varepsilon - 1} + X_{it}^u \beta_j^u + \nu_{ijt} \tag{3.5}$$

For fixed-payment consumers, as the continuous choice is pre-determined by the fixed payment, only the discrete equation matters for estimation:

$$U_{ijt} = \frac{e^{\frac{1}{\varepsilon}\left(\alpha_j + X_{it}^e \beta^e + \eta_{it}\right)} \left(\frac{\overline{E}}{p_{jt}}\right)^{1-\frac{1}{\varepsilon}}}{1 - \frac{1}{\varepsilon}} + X_{it}^u \beta_j^u + \nu_{ijt} \tag{3.6}$$

Denote the mean utiltity as V_{ijt}, i.e. Let the mean utility be denoted by V_{ijt}, i.e.

$$V_{ijt} = \begin{cases} \frac{E^*_{it}\left(\frac{p_{jt}}{p^*_{it}}\right)^{1-\varepsilon} e^{\alpha_j - \alpha_{j^*_i}}}{\varepsilon - 1} + X^u_{it}\beta^u_j & \text{if not paying 50R\$} \\ \frac{e^{\frac{1}{\varepsilon}\left(\alpha_j + X^e_{it}\beta^e + \eta_{it}\right)}\left(\frac{\overline{E}}{p_{jt}}\right)^{1-\frac{1}{\varepsilon}}}{1-\frac{1}{\varepsilon}} + X^u_{it}\beta^u_j & \text{if paying 50R\$} \end{cases}$$

Distributional assumption For tractability, the stochastic term in the random utility equation is assumed to follow Type-1 extreme value distribution. Let mu denote the scale parameter of the distribution. This distribution gives rise to a closed-form expression of the choice probability for fuel j, P_{ijt}, as follows:

$$P_{ijt} = \frac{\exp(V_{ijt}/\mu)}{\sum_k \exp(V_{ikt}/\mu)}$$

The stochastic term in the quantity η_{it} is assumed to follow normal distribution, although this choice of distribution is not critical to the tractability of the estimation. It is chosen because the distribution of the payments is observed to be right-skewed, and can therefore be well captured by a lognormal distribution. η_{it} enters the log equation of the payments, and hence a normal distribution is appropriate to model this term. Let σ^2 denote the variance of η_{it}; the likelihood of observing a payment E_{ijt} is then

$$l(\eta_{ijt}) = \phi(\eta_{it}/\sigma) = \phi((\ln(E_{ijt} - (1-\varepsilon)p_{jt} - X_{it}\beta^e)/\sigma)$$

with $\phi(\cdot)$ denoting the probability density function of a standard normal distribution.

Likelihood Let Θ denote the set of all the parameters in the model, i.e. $(\alpha_j, \beta_j^u, \beta^e, \varepsilon, \sigma, \mu)$. For a flexible-payment consumer, given Θ and a payment for a particular fuel, we can derive η_{jt}. Conditional on η_{jt}, we can derive the optimal payments for all other fuels, and hence the mean utility and the choice probability of other fuels. Thus, the likelihood of choosing fuel j and spending E_{ijt} on that fuel can be decomposed into the likelihood of E_{ijt} being the optimal payment for fuel j and the likelihood of j being the best fuel conditional on the optimal payment.

$$\begin{aligned}
L^{\text{flex}}(E_{ijt}, j|\Theta, X, p) = \quad & l(E_{ijt}|\Theta, X, p) \times P(j|E_{ijt}, \Theta, X, p) \\
= \quad & l(\eta_{it} = \ln E_{ijt} - \alpha_j - (1-\varepsilon)p_{jt} - X_{it}^e \beta^e|\Theta, X, p) \\
& \times P(j|\eta_{it}, \Theta, X, p) \\
= \quad & \phi((\ln(E_{ijt} - (1-\varepsilon)p_{jt} - X_{it}\beta^e)/\sigma) \\
& \times \frac{\exp(V_{ijt}/\mu)}{\sum_k \exp(V_{ikt}/\mu)}
\end{aligned}$$

For a fixed-payment consumer, as her ˙ijt is not known, we need to integrate η_{it} out to obtain the choice probability, which is also the likelihood of her purchase:

$$L^{\text{fixed}}(j|\Theta, X, p) = \int \frac{\exp(V_{ijt}/\mu)}{\sum_k \exp(V_{ikt}/\mu)} dF(\eta_{it})$$

3.5.1 Simulation study

A simulation study was conducted to verify the internal validity of the model and the estimation. Artificial data was created using the prices of gasoline and ethanol in the sample. The stochastic terms η_{it} and ν_{ijt} were randomly generated according the distributions assumed above. Choices and payments were calculated based on the demand equation and utility equation described above (Equations 3.2 and 3.3). Fixed-payment preference was assigned randomly with a probability of 0.2. For simplicity, no covariates were used, μ was set to 1, $\sigma = \exp(-0.4)$, $\alpha_{gasoline} = 0$ and $\alpha_{ethanol} = 4.5$. ε. ε was varied, taking on three different values of $e^{-0.2} = 0.82, e^{-0.5} = 0.61$ and $e^{-0.8} = 0.45$. For each set of parameters, 200 sets of data were simulated.

Estimations with and without the consideration of the preference for fixed payments were conducted on these sets of data to obtain 200 sets of estimates. Table E.1 displays the average of the estimates for each parameter, together with the average biases and the standard deviation of the estimates. Without considering preference for fixed payment, the magnitude of the price elasticity of quantity demanded, ε, was overestimated, while the variance of the stochastic term in the continuous choice equation, σ, and the difference in preference for quantity between ethanol and gasoline, $\alpha_{ethanol}$, were underestimated. Taking fixed-payment preference into account in the estimation corrected these biases.

3.5.2 Estimation Results

Table E.4 displays the estimates for four specifications. For these for specifications, I allowed vehicle characteristics (class, engine, fuel efficiency, model year, fuel tank), consumer demographics (gender, age, education), cities, day of the week, and time of the day to shift both the discrete choice utility and the continuous choice (via the effect on consumer preference type θ). In addition, station fixed effects are included in the discrete utility of specifications 2 and 4, and in the continuous choice equation of specifications 3 and 4.

The continuous choice Panel B of Table E.4 displays the estimated coefficient for the continuous choice equation, i.e. the equation for the logarithm of quantity of purchase. The income effect is very apparent from these results; increasing age, education, and vehicle price is associated with a higher quantity of purchase. Consumers aged over 65 purchase 25% to 28% more than consumers under 25 at the stations. Consumers with some college education purchase 23% to 28% more than consumers with primary education at most, and 17% to 20% more than those with some type of secondary degree. Consumers with vehicles priced in the top quartile purchase 9.3% to 9.7% more than the rest. To the extent that these variables are correlated positively with income, this can be interpreted as the income effect, which increases demand when income increases.

Vehicle characteristics are also an important determinant of the continuous choice. Larger engines and newer vehicles are associated with a higher quantity of purchase. Owners of mid-size vehicles choose 11% to 12% more quantity than owners of compact vehicles (the omitted class). Full-size vehicles are also estimated to require approximately 5% more fuel from one purchase, but the effect is not statistically significant. These variables can be correlated with income as well, but they can also reflect the occupational or personal needs of the drivers. The latter is also captured by the significant estimated effect of extensive vehicle usage on purchased quantity, which is 15% to 16%.

Consumers tend to purchase more quantity with ethanol than with gasoline or mid-grade gasoline. This, however, does not fit the argument that consumers perceive ethanol as having higher maintenance costs than gasoline (due to the perceived corrosive effects of ethanol on the engine). Thus, this effect is more likely due to consumers attaching different valuations to the same distance travelled on different fuels. Consumers may regard ethanol as cleaner due to its low GHG emission, or value its higher octane and anti-knock properties, which will offset the cost for each km of travelling on it.

In short, the continuous choice has captured a lot of consumer heterogeneity in the way consumers make quantity decisions. Most important among them seems to be the income effect. To the extent that consumers value quantity, and due to price differences, the quantity that they get for different fuels varies. This heterogeneity will also capture consumers heterogeneous responses to prices in their discrete choices between different fuel types.

Discrete choice Panels C and D of E.4 display the coefficients of the variables included in the relative utility equation. The equation has been scaled (by the scale parameter μ of the type-I extreme value distribution) such that it has the same unit as payment (i.e. in Brazilian reais).

Of the consumer demographic variables, gender and age display significant effects on discrete choice. Being female is associated with more purchases of gasoline, while the mid-grade gasolinegasoline relative utility increases with age. Consumers aged between 25 and 65 appear more likely to choose ethanol or mid-grade gasoline over gasoline, all else being equal, than consumers aged less than 25 do. Consumers from Curitiba and Sao Paulo appear to value ethanol over gasoline more than consumers from Belo Horizonte and Recife do. Consumers with high usage seem less likely to purchase ethanol over gasoline than the rest, although the difference is smaller and not precise after including the station fixed effects in the continuous choice equation.

Elasticity The estimated elasticity of purchased quantity is from -$-e^{-0.224} = -0.748$ to $-e^{-0.209} = -0.811$. This is different from the elasticity of fuel demand in the short run as stated in the literature, which is around -0.2 or -0.3, and can fall to as low as -0.02. However, we should not expect them to be similar. The elasticity estimated here should be understood as the quantity variation from one fueling-station visit to another. Instead of varying the quantity of purchases and payments, consumers can vary the frequency of their station visits. Furthermore, the data only cover consumers who pulled up at the fueling stations, and hence would not have captured the responses of people who switched to public transport or other modes of travel when fuel prices were high. In addition, the data potentially oversample frequent visitors to fueling stations, who may tend to be more price sensitive than the average consumer.

3.5.3 Estimating the effect of Price Salience Treatments

This section will address the question of how price-salience treatments affect how consumers make fuel-choice decisions along both the extensive and intensive margins. Table E.3 displays the estimation results for three different specifications. For these specifications, the full set of controls (consumer demographics, vehicle characteristics, day of week, time of day, and station fixed effects) is included in both the continuous and discrete choice relative utility equations. What changes across these three specifications is the inclusion of the dummy for the salience in either or both the continuous choice equation or/and the discrete choice equation.

It should be noted that the discrete choice equation is essentially the relative utility between the given fuel and gasoline. Even if the treatments shift the utility, they will likely shift the utilities of both the given fuel and gasoline, and the effects will be cancelled when the relative utility is considered . In addition, there is no obvious reason the price-salience treatment would shift the utilities of different fuels differently.

It would be more interesting to examine how salience treatments affect consumers when there is asymmetry between fuels. Given that price salience informs consumers about price, we can expect the treatments to benefit the consumers by making them aware of the cheaper alternative. Specifically, if ethanol is favorably priced and some consumers are not aware of that, increasing price salience should help them recognize the greater benefits of ethanol and cause some of them to switch from gasoline to ethanol. Therefore, to identify the effects of the price-salience treatments in the discrete choice equation, I use the treatments interacting with a dummy variable indicating if ethanol is more a favorable fuel than gasoline in reais-per-km terms.

Specification (2) includes the treatment dummies only in the continuous choice equation, specification (3) includes them only in the discrete choice equation, and specification (4) includes them in both equations. The results are consistent across the three specifications.

The price-ratio flyer treatment increases the quantity of purchase by between 4.6% (specification (2)) and 7.4% (specification (4)), although the latter estimate is less precise (standard error of 4.1%). The price-ratio treatment also increases the utility of ethanol relative to that of gasoline during the period when ethanol is favorable, which is consistent with the argument that price salience helps the consumers realize the economic benefit of the alternative fuel that they would have missed otherwise.

On the other hand, the effect of the km-per-50R$ flyer is small and statistically insignificant in both the continuous choice equation and the discrete choice equation. A statistical test indicates that the effect of the km-per-50R$

flyer is lower than that of the price-ratio flyer. One explanation for this difference is that the information displayed in the km-per-50R$ flyer is less familiar and more difficult to process than that of the price-ratio flyer. In fact, the ratio 0.7 is widely publicized in the media, and Salvo and Huse (2013) surveyed consumers and found that 70% of them could recall this ratio. This suggests that the effect of the price salience can be partly explained by the possibility that it lowered the mental cost of processing information that, even though available and sufficient, was not immediately apparent for decision making.

These results have two implications. The first is that, policy-wise, the choice along both dimensions matters. A policy may be designed to induce consumers to purchase more efficient fuel by making the price more salient, but price salience can have the side-effect of increasing the quantity purchased, which compromises the objective of the policy. The second is that, regarding information salience, content and presentation matter. Having sufficient but poorly presented information may diminish the effect of that information.

3.6 Conclusion

In this paper, a model of discrete-continuous choice is introduced. The data show an irregularity in the way that an abnormally large number of consumers choose to pay a specific amount, 50R$, and a simulation study shows that ignoring this irregularity can affect the accuracy of the estimation. The model was employed in the study of consumer fuel choice at fueling stations in Brazil. The results show that price salience can increase the quantity of purchase and help consumers realize more of the benefits from the favorably priced fuels. In addition, the type of treatment is important to achieve this effect.

Bibliography

Ackerberg, D. A. (2001). Empirically distinguishing informative and prestige effects of advertising. *RAND Journal of Economics*, 316–333.

Ackerberg, D. A. (2003). Advertising, learning, and consumer choice in experience good markets: an empirical examination. *International Economic Review 44*(3), 1007–1040.

Ahn, J., G. Jeong, and Y. Kim (2008). A forecast of household ownership and use of alternative fuel vehicles: A multiple discrete-continuous choice approach. *Energy Economics 30*(5), 2091–2104.

Anand, B. N. and R. Shachar (2011). Advertising, the matchmaker. *The RAND Journal of Economics 42*(2), 205–245.

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

Anderson, S. T. and J. M. Sallee (2011). Using loopholes to reveal the marginal cost of regulation: The case of fuel-economy standards. *The American Economic Review 101*(4), 1375–1409.

Bagwell, K. (2007). The economic analysis of advertising. *Handbook of industrial organization 3*, 1701–1844.

Bento, A. M., L. H. Goulder, M. R. Jacobsen, and R. H. Von Haefen (2009). Distributional and efficiency impacts of increased us gasoline taxes. *The American Economic Review*, 667–699.

Berry, S., J. Levinsohn, and A. Pakes (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Bhat, C. R. (2005). A multiple discrete–continuous extreme value model: formulation and application to discretionary time-use decisions. *Transportation Research Part B: Methodological 39*(8), 679–707.

Brazilian Automotive Industry Association (2013). Brazilian automotive industry yearbook.

Brown, J., T. Hossain, and J. Morgan (2010). Shrouded attributes and information suppression: Evidence from the field. *The Quarterly Journal of Economics 125*(2), 859–876.

Busse, M., J. Silva-Risso, and F. Zettelmeyer (2006). $1,000 cash back: The pass-through of auto manufacturer promotions. *The American Economic Review 96*(4), 1253–1270.

Busse, M. R., C. R. Knittel, and F. Zettelmeyer (2013). Are consumers myopic? evidence from new and used car purchases. *The American Economic Review 103*(1), 220–256.

Chetty, R., A. Looney, and K. Kroft (2009). Salience and taxation: Theory and evidence. *The American economic review 99*(4), 1145–1177.

Dahl, C. and T. Sterner (1991). Analysing gasoline demand elasticities: a survey. *Energy economics 13*(3), 203–210.

DellaVigna, S. and J. M. Pollet (2009). Investor inattention and friday earnings announcements. *The Journal of Finance 64*(2), 709–749.

Drollas, L. P. (1984). The demand for gasoline. *Energy Economics 6*(1), 71 – 82.

Dubin, J. A. and D. L. McFadden (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica: Journal of the Econometric Society*, 345–362.

Eizenberg, A. (2014). Upstream innovation and product variety in the us home pc market. *The Review of Economic Studies 81*(3), 1003–1045.

Environmental Protection Agency (2015). Light-duty automotive technology , carbon dioxide emissions , and fuel economy trends?: 1975 through 2012. `http://www3.epa.gov/otaq/fetrends.htm`.

Espey, M. (1998). Gasoline demand revisited: an international meta-analysis of elasticities. *Energy Economics 20*(3), 273–295.

European Commission (2016). The revised renewable energy directive.

Goodwin, P. B. (1992). A review of new demand elasticities with special reference to short and long run effects of price changes. *Journal of Transport Economics and Policy 26*(2), 155–169.

Goolsbee, A. and A. Petrin (2004). The consumer gains from direct broadcast satellites and the competition with cable tv. *Econometrica 72*(2), 351–381.

Graham, D. J. and S. Glaister (2002). The demand for automobile fuel: a survey of elasticities. *Journal of Transport Economics and Policy (JTEP) 36*(1), 1–25.

Gramlich, J. (2010). Gas prices, fuel efficiency, and endogenous product choice in the us automobile industry. *Unpublished Working Paper*.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *The American Economic Review 103*(7), 2643–2682.

Hanemann, W. M. (1984). Discrete/continuous models of consumer demand. *Econometrica: Journal of the Econometric Society*, 541–561.

Hastings, J. S. and J. M. Shapiro (2013). Fungibility and consumer choice: Evidence from commodity price shocks*. *The Quarterly Journal of Economics*.

Hsieh, W.-D., R.-H. Chen, T.-L. Wu, and T.-H. Lin (2002). Engine performance and pollutant emission of an si engine using ethanol–gasoline blended fuels. *Atmospheric Environment 36*(3), 403–410.

Hughes, J. E., C. R. Knittel, and D. Sperling. Evidence of a shift in the short-run price elasticity of gasoline demand. *The Energy Journal 29*(1).

Jacobsen, M. R. (2013). Evaluating us fuel economy standards in a model with producer and household heterogeneity. *American Economic Journal: Economic Policy 5*(2), 148–187.

Kanetkar, V., C. B. Weinberg, and D. L. Weiss (1992). Price sensitivity and television advertising exposures: Some empirical findings. *Marketing Science 11*(4), 359–371.

Kim, J., G. M. Allenby, and P. E. Rossi (2002). Modeling consumer demand for variety. *Marketing Science 21*(3), 229–250.

Klier, T. and J. Linn (2010). The price of gasoline and new vehicle fuel economy: evidence from monthly sales data. *American Economic Journal: Economic Policy 2*(3), 134–153.

Klier, T. and J. Linn (2012). New-vehicle characteristics and the cost of the corporate average fuel economy standard. *The RAND Journal of Economics 43*(1), 186–213.

Klier, T. and J. Linn (2016). The effect of vehicle fuel economy standards on technology adoption. *Journal of Public Economics 133*, 41–63.

Knittel, C. R. (2011). Automobiles on steroids: Product attribute trade-offs and technological progress in the automobile sector. *American Economic Review 2012*(101), 3368–3399.

Li, S., C. Timmins, R. H. von Haefen, et al. (2009). How do gasoline prices affect fleet fuel economy? *American Economic Journal: Economic Policy 1*(2), 113–37.

Masum, B., H. Masjuki, M. Kalam, I. R. Fattah, S. Palash, and M. Abedin (2013). Effect of ethanol–gasoline blend on nox emission in si engine. *Renewable and Sustainable Energy Reviews 24*, 209–222.

Miravete, E. J. (2003). Choosing the wrong calling plan? ignorance and learning. *The American Economic Review 93*(1), 297–310.

Pakes, A., S. Berry, and J. A. Levinsohn (1993). Applications and limitations of some recent advances in empirical industrial organization: price indexes and the analysis of environmental change. *The American Economic Review 83*(2), 240–246.

Pedrick, J. H. and F. S. Zufryden (1991). Evaluating the impact of advertising media plans: A model of consumer purchase dynamics using single-source data. *Marketing Science 10*(2), 111–130.

Petrin, A. (2002). Quantifying the benefits of new products: The case of the minivan. *Journal of Political Economy 110*(4), 705–729.

Pouliot, S. and B. A. Babcock (2014). The demand for e85: Geographical location and retail capacity constraints. *Energy Economics 45*, 134–143.

Rivers, N. and B. Schaufele (2015). Salience of carbon taxes in the gasoline market. *Journal of Environmental Economics and Management 74*, 23 – 36.

Salvo, A. (2015). Flexible fuel vehicles, less flexible minded consumers: Price salience experiments at the pump. *Working paper*.

Salvo, A. (2016). Flexible fuel vehicles, less flexible minded consumers: Price salience experiments at the pump.

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), 251–279.

Tellis, G. J. (1988). Advertising exposure, loyalty, and brand purchase: A two-stage model of choice. *Journal of marketing research*, 134–144.

Train, K. (2003). *Discrete choice methods with simulation*. Cambridge university press.

US Energy Information Administration (2017, Feb). February monthly energy review.

Wang, M., J. Han, J. B. Dunn, H. Cai, and A. Elgowainy (2012). Well-to-wheels energy use and greenhouse gas emissions of ethanol from corn, sugarcane and cellulosic biomass for us use. *Environmental Research Letters 7*(4), 045905.

Yüksel, F. and B. Yüksel (2004). The use of ethanol–gasoline blend as a fuel in an si engine. *Renewable energy 29*(7), 1181–1191.

Zhou, Y. (2016). Knowledge capital, technology adoption and environmental policies? evidence from the us automobile industry. *Unpublished Working Paper*.

Appendix A

Chapter 1: Tables

Table A.1: Examples of fuel-saving technologies, taken from EPA (2015)

TABLE 7.11 Summary of the Committee's Findings on the Costs and Impacts of Technologies for Reducing Light-Duty Vehicle Fuel Consumption

Fuel Consumption Technology	Description and Approximate Manufacturing Cost	Impact on Fuel Consumption (%)	Comments
Mass reduction (assume 3,600-pound vehicle)	1% (36 lb); $46-$55	0.25	Material substitution
	5% (180 lb); $270-$324	3-3.5[a]	Material substitution
	10% (360 lb); $648-$778	6-7[a]	Aggressive material substitution
	20% (720 lb); $1,600+	11-13[a]	Redesigned body with aluminum and composite-intensive structures
Transmission	Five-speed automatic transmissions; $133	2-3	Can also improve vehicle performance
	Six-speed automatic transmissions; $133-$215	3-5	Can also improve vehicle performance
	Seven-speed automatic transmissions; $170-$300	5-7	Can also improve vehicle performance
	Eight-speed automatic transmissions; $425	6-8	Can also improve vehicle performance
	Dual-clutch automated (DCT) manual transmissions (6/7 speed); $300 (dry clutch), –$14-$400 (wet clutch <350 N-m)	6-9	DCTs have replaced original automated manual transmissions
	Continuously variable transmissions; $150 (<2.8 L), $263 (>2.8 L)	1-7	Possible engine noise; not applicable to large engines
Aerodynamics	5 to 10% reduction in C_d (coefficient of drag); $40-$50	1-2	Wheel well and underbody covers, body shape, mirrors, etc.; bigger impact on highway drive cycle
Rolling resistance	Low-rolling-resistance tires; approximately $10 apiece ($30-$40)	1-2[b]	Stopping distance and durability can be compromised with inferior materials; optimal materials drive up costs
	Tire-inflation monitor; becoming standard equipment	0.7	Depends on monitor settings and driver behavior
	Low-drag brakes; becoming standard equipment	1	Most cars equipped already today
Electrical accessories	HVAC—variable stroke, increased efficiency (humidity control, paint, glass, etc.); $70-$90	3-4	Current FTP does not capture benefit (benefits reduced to 0.5-1.5% within Table 9.1)
	Electric and electric-hydraulic power steering; $70-$120	1-5	Electric for small cars, electric-hydraulic for bigger cars—benefits for the FTP are smaller (1-3%).

[a]With resized power train.
[b]Three percent may be feasible with resized power train.

Table A.2: Summary statistics

	Truck		Car		All	
Price ('000 $)	29.6	[7.77]	23.89	[11.03]	26.58	[10.04]
Household income ('000 $)	56.83	[1.61]	56.85	[1.59]	56.84	[1.60]
Price/HH income	0.52	[0.14)	0.42	[0.19]	0.47	[0.18]
Gasoline price ($)	2.84	[0.29]	2.83	[0.29]	2.84	[0.29]
Fuel efficiency (miles/gal)	26.95	[3.26]	36.64	[5.99]	32.07	[5.95]
Fuel consumption (gal/100 miles)	3.82	[0.89]	2.84	[0.76]	3.3	[1.08]
Fuel cost ($/mile)	0.13	[0.03]	0.09	[0.03]	0.11	[0.04]
Horsepower ('00 hp)	2.59	[0.59]	2	[0.67]	2.28	[0.70]
Weight ('000 lb)	4.56	[0.83]	3.26	[0.45]	3.87	[0.92]
Hp/Weight ('0 hp/lb)	0.57	[0.07]	0.6	[0.13]	0.58	[0.11]
Space ('0000 ft2)	1.54	[0.23]	1.3	[0.12]	1.41	[0.22]
Torque (lbft)	270.13	[72.25]	194.93	[67.86]	230.37	[79.39]
Car (dummy)	0	[0.00]	1	[0.00]	0.53	[0.50]

Table A.3: Estimation Results

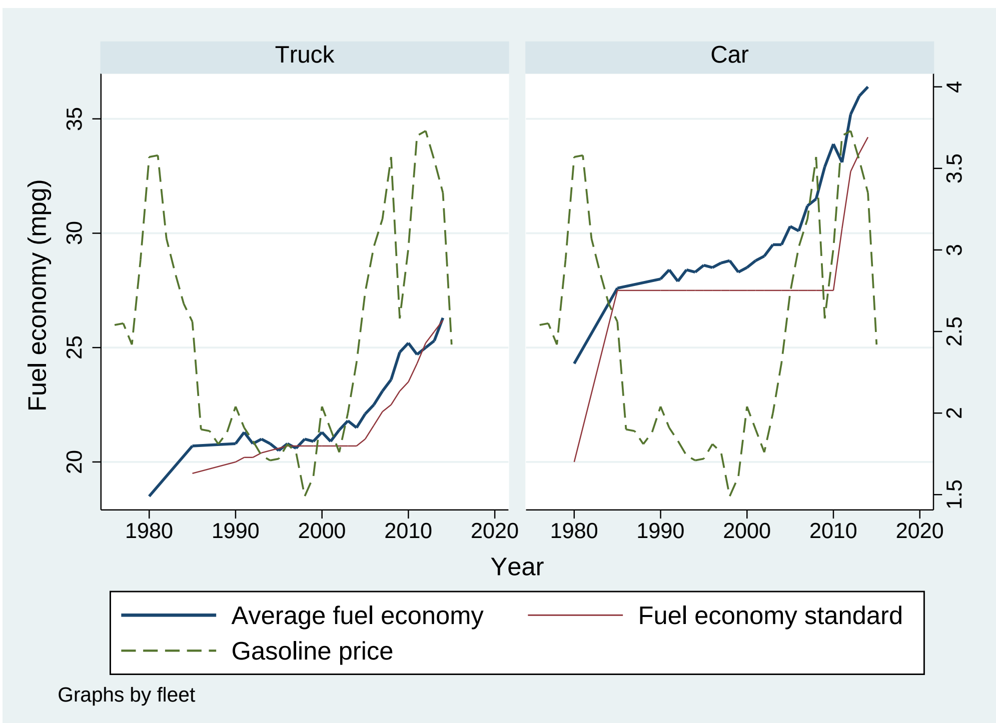
	Utility/Mean			Utility/std		
	estimate	se	t-stat	estimate	se	t-stat
price/income	-45.04	9.86	-4.57	1.47	0.16	9.20
dpm/income	-23.86	6.14	-3.89	-0.60	0.14	-4.18
hp/weight	2.77	0.70	3.97	2.26	0.40	5.64
weight	-0.46	0.15	-2.95	-0.35	0.07	-4.65
size	3.81	0.39	9.71	-0.44	0.22	-2.02
const	1.22	0.46	2.67	-14.00	0.79	-17.74
suv	-1.13	0.24	-4.65			
truck	-2.85	0.44	-6.44			
van	-3.49	0.56	-6.19			
minivan	-1.23	0.21	-5.88			

	Marginal Cost			Fuel-efficiency Frontier		
	estimate	se	t-stat	estimate	se	t-stat
log(hpwt)	1.21	0.16	7.78	-0.22	0.02	-8.88
log(weight)	2.10	0.42	5.01	-0.57	0.05	-12.27
log(size)	-1.67	0.19	-8.69	0.20	0.04	4.79
log(torque)	0.64	0.11	5.90	-0.28	0.03	-9.14
const	1.81	0.42	4.34	0.03	0.11	0.27
suv	0.02	0.04	0.41	-0.09	0.01	-13.29
truck	0.02	0.06	0.30	-0.14	0.01	-13.22
van	0.00	0.12	-0.03	-0.29	0.02	-12.92
minivan	0.12	0.06	1.92	-0.09	0.01	-7.58
trend	-0.05	0.01	-3.30			
year 2007				0.01	0.01	1.17
year 2008				0.08	0.01	7.67
year 2009				0.06	0.01	6.75
year 2010				0.09	0.01	9.37
year 2011				0.11	0.01	12.50
year 2012				0.13	0.01	13.96
year 2013				0.18	0.01	19.01
year 2014				0.18	0.01	19.11
control function				-0.03	0.01	-3.09
c_e				0.15	0.02	9.04
c_e^2				-0.02	0.00	-6.90
c_e^3				0.00	0.00	5.69

	Compliance cost		
	estimate	se	t-stat
compliance cost	0.23	0.13	1.84

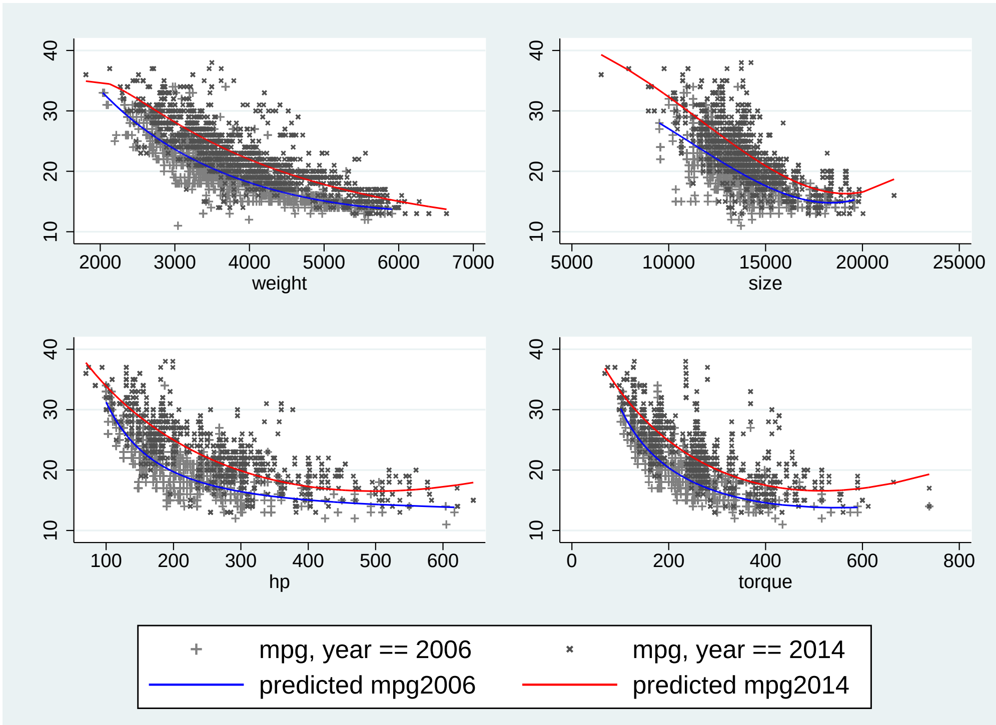
Appendix B

Chapter 1: Figures



Notes: This figure plots gasoline prices (USD/gal), CAFE standards (miles per gallon) and the harmonic average fuel economy of new light-duty vehicles sold in the US (in miles per gallon) from 1980 to 2014 (the gaosoline prices series covers some extra years). The number are reported separately for passenger cars (right panel) and light-duty trucks (left panel). Source: NHTSA

Figure B.1: Gasoline prices, average fuel economy, CAFE standards



Scatter plots between fuel economy (in miles per gallon) and weight (top left panel), size (top right panel), horse power(bottom left panel) and torque (bottom right panel) in 2006 (light gray + dots) and 2014 (black X dots). The solid lines plots the best fitted curves for each scatter plots. Each dot is a vehicle model at trim level in each year (e.g. Toyota Camry SE 4dr Sedan). Source: edmunds.com

Figure B.2: Fuel economy and vehicle characteristics

Figure 6.1

Industry-Wide Car Technology Penetration after First Significant Use

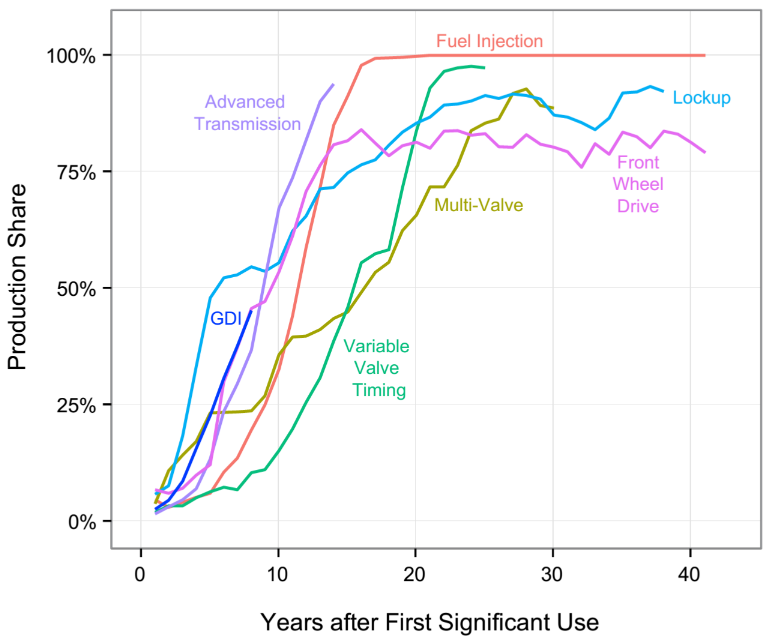
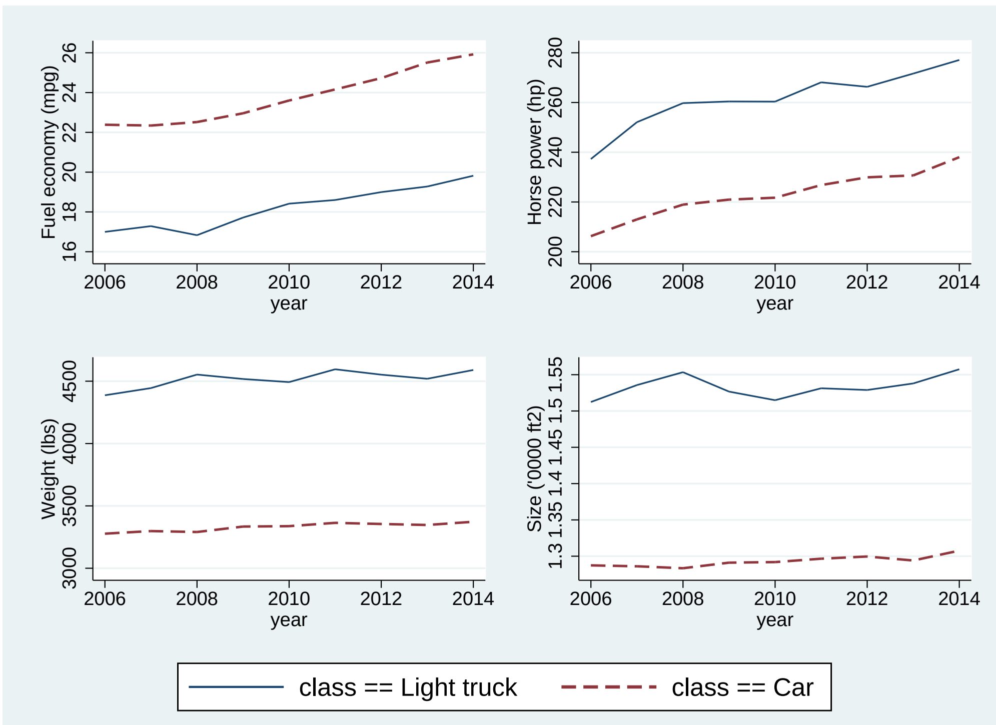
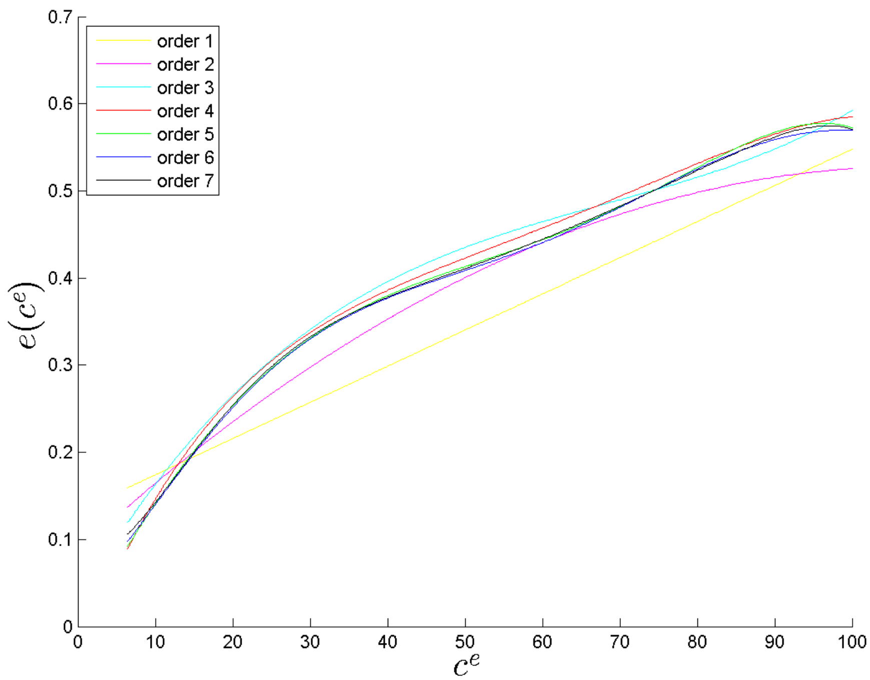


Figure B.3: Adoption rate of different fuel-saving technologies, taken from EPA (2015)



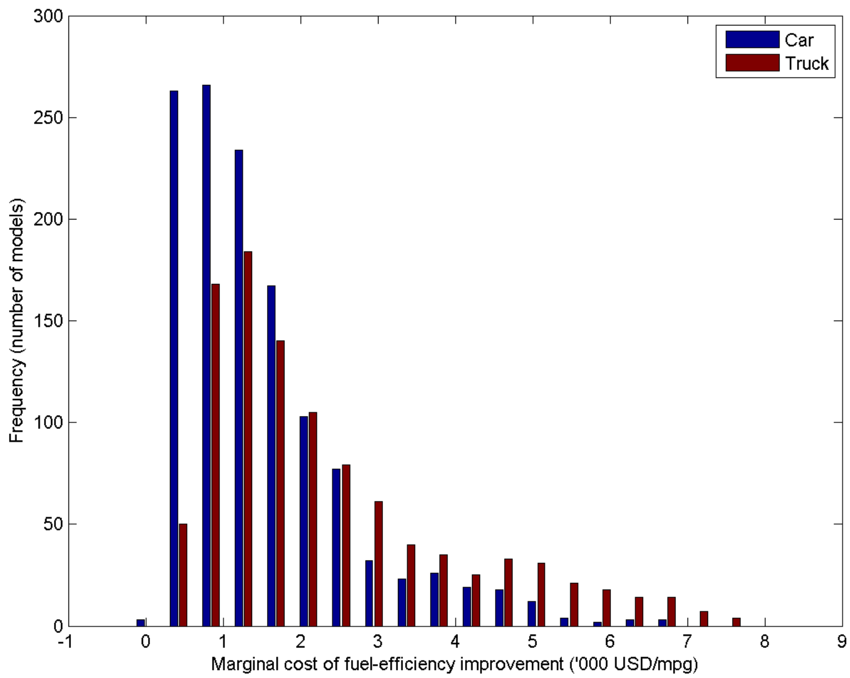
Notes: fuel economy on top left panel, horse power top right, weight bottom left, size bottom right; dahsed lines are for passenger cars and solid lines are for light-duty trucks. This is raw average (not weighted by sales) of all model at trim level. Source: edmunds.com

Figure B.4: Average vehicle characteristics over time



Notes: This figure plots the amount of improvement on fuel efficiency (in proportion, vertical axis) against the marginal technological cost of further improving the fuel-efficiency (USD per 100% improvement). Polynomials of different order are used to fit this curve, which is indicated by different colors

Figure B.5: Marginal fuel saving cost curve



Notes: this figure is based on the model's estimates for the marginal fuel technological cost, i.e. the cost of improving the fuel efficiency further by 1mpg.

Figure B.6: Distribution of the marginal fuel technological cost

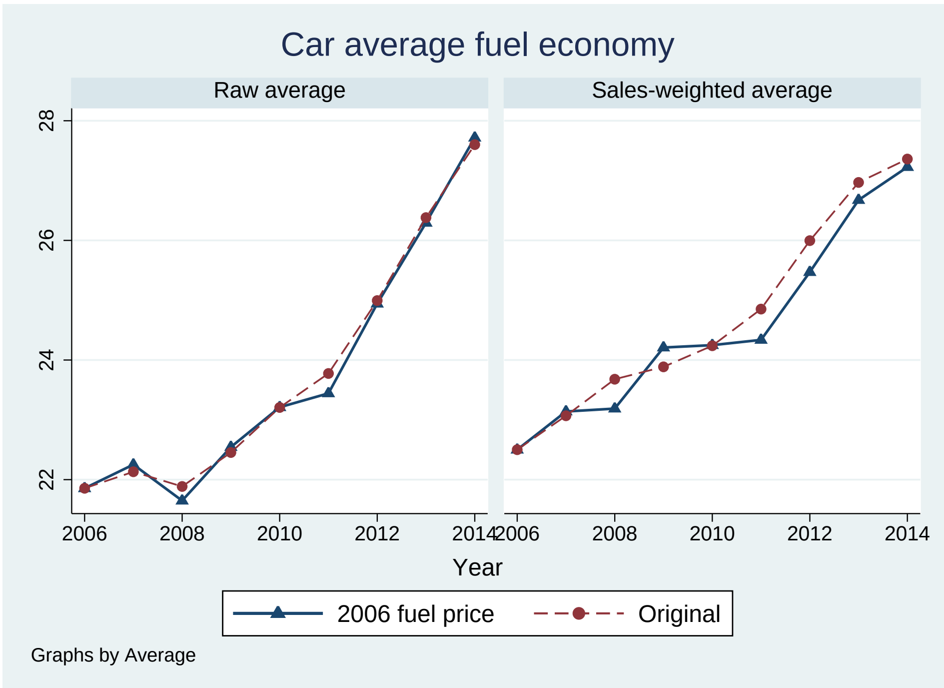


Figure B.7: Counterfactual car average fuel economy if fuel prices are set at 2006 level

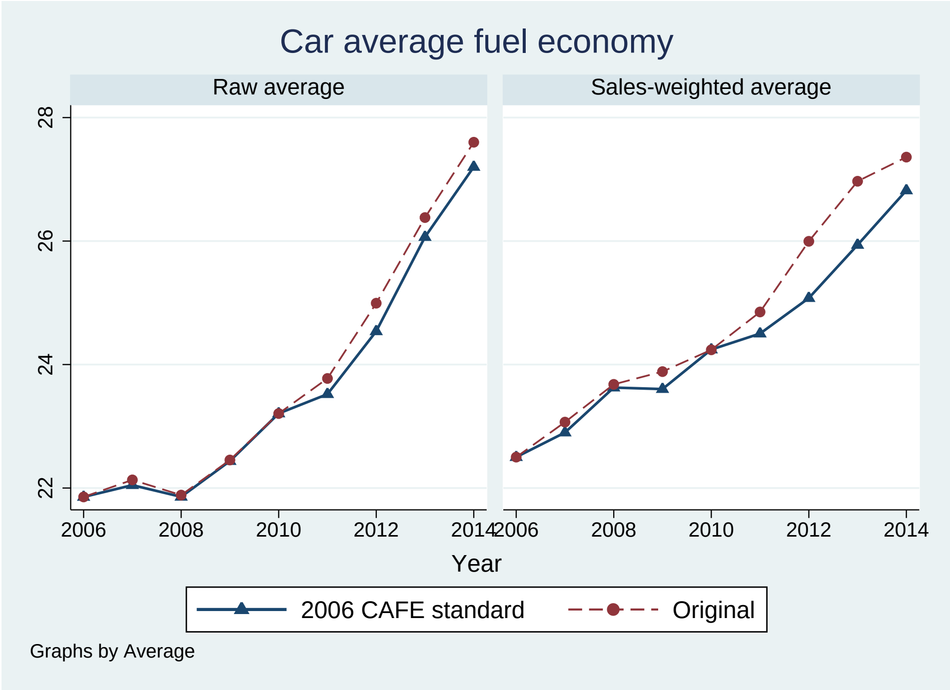


Figure B.8: Counterfactual car average fuel economy if CAFE standards are set at 2006 level

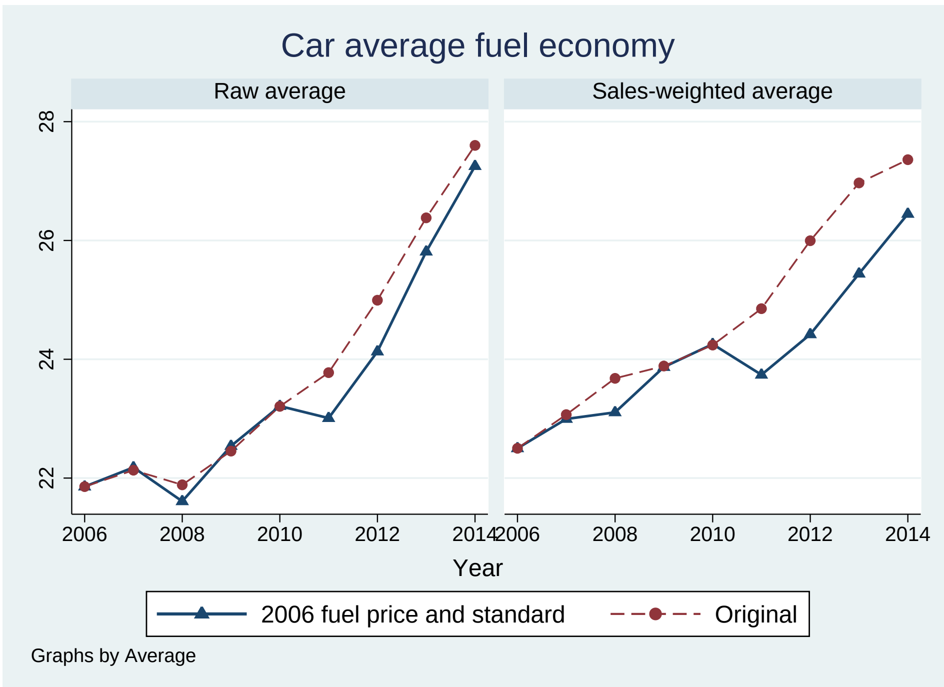


Figure B.9: Counterfactual truck average fuel economy if fuel prices and CAFE standards are set at 2006 level

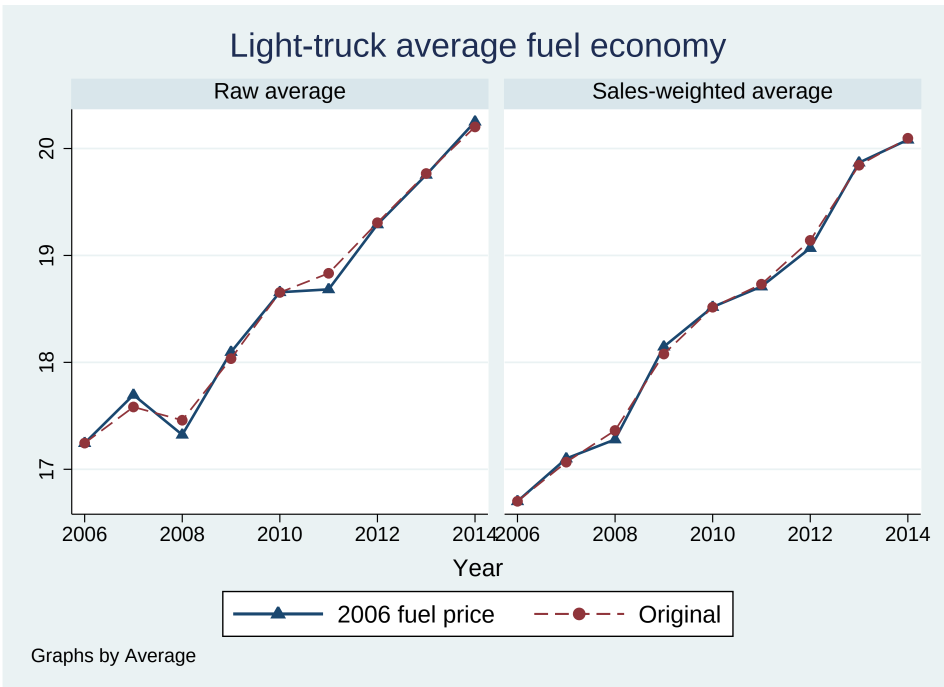


Figure B.10: Counterfactual truck average fuel economy if fuel prices are set at 2006 level

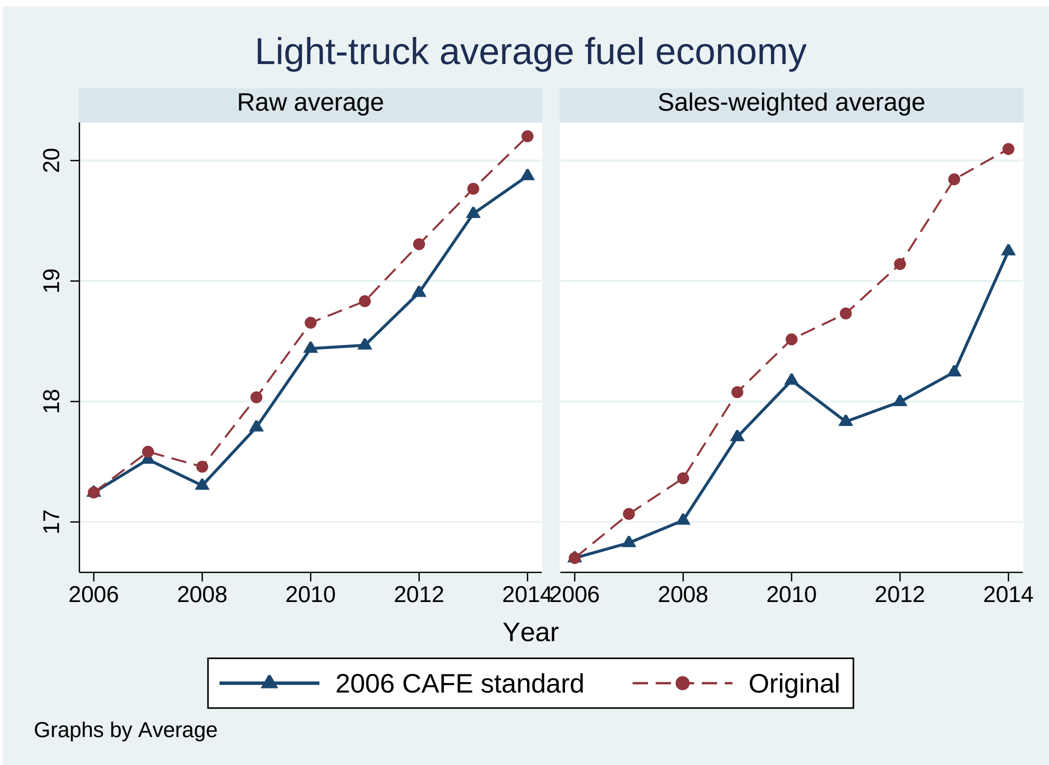


Figure B.11: Counterfactual truck average fuel economy if CAFE standards are set at 2006 level

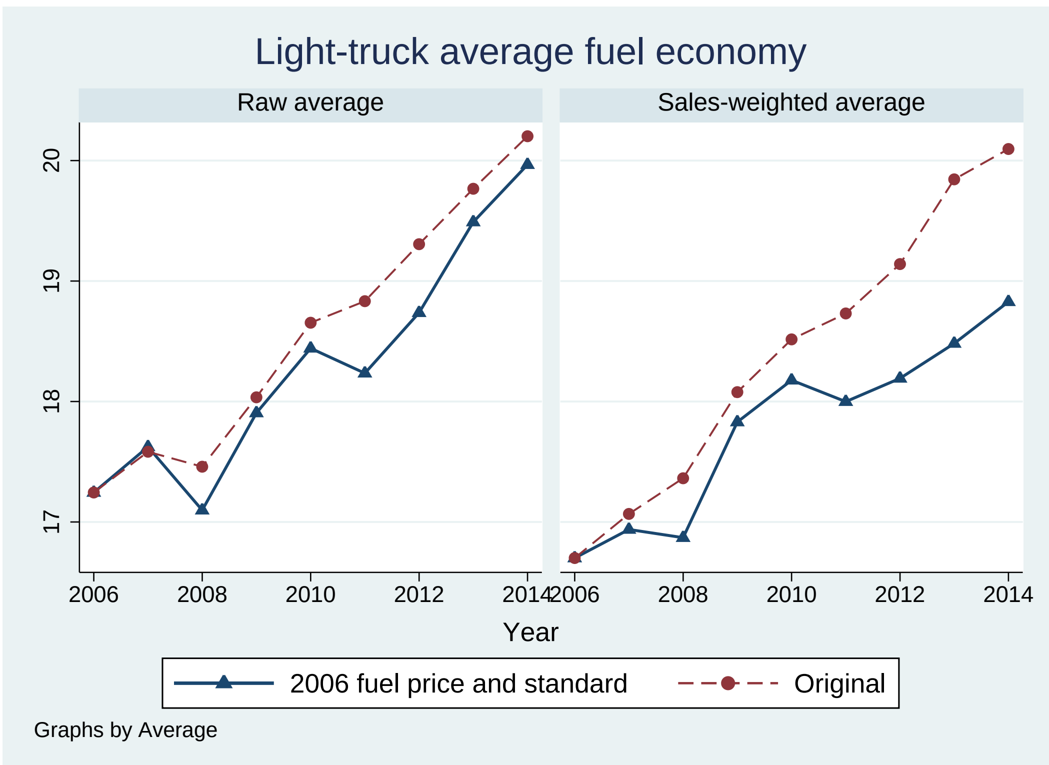


Figure B.12: Counterfactual truck average fuel economy if fuel prices and CAFE standards are set at 2006 level

Appendix C

Chapter 2: Tables

Table C.1: Summary statistics by treatment groups and Covariate balance

	Mean			Difference to control [SE]	
	Control	Price-ratio	Km/R$50	Price-ratio	Km/R$50
Female	0.340	0.345	0.347	-0.004 [0.011]	-0.007 [0.011]
Age 25 to 40	0.458	0.469	0.479	-0.011 [0.012]	-0.021* [0.012]
Age 40 to 65	0.372	0.368	0.359	0.005 [0.012]	0.013 [0.012]
Age more than 65	0.045	0.035	0.037	0.010** [0.005]	0.008* [0.005]
Secondary school	0.267	0.271	0.268	-0.003 [0.011]	-0.001 [0.011]
College	0.689	0.685	0.691	0.004 [0.011]	-0.002 [0.011]
Fiat vehicle	0.292	0.298	0.298	-0.006 [0.011]	-0.007 [0.011]
GM vehicle	0.202	0.203	0.197	-0.001 [0.010]	0.005 [0.010]
Volkswagen vehicle	0.215	0.221	0.199	-0.006 [0.010]	0.016 [0.010]
Ford vehicle	0.102	0.096	0.104	0.005 [0.007]	-0.003 [0.007]
Vehicle price ('000R$)	29.048	28.975	29.106	0.073 [0.225]	-0.058 [0.233]
Engine size (liters)	1.355	1.358	1.356	-0.003 [0.008]	-0.001 [0.009]
Observations	3474	3474	3474		

*p<0.10 **p<0.05 ***p<0.01

Table C.2: Estimation results

	(1)		(2)	
ln(price/km)	-3.854***		-3.961***	
	[0.363]		[0.419]	
Mean utility equation	**ethanol**	**mgasoline**	**ethanol**	**mgasoline**
Female (DV)	-0.128***	-0.081	-0.126***	-0.092
Age 25-40 (DV)	0.073	0.126	0.074	0.088
Age 40-65 (DV)	0.047	0.287*	0.036	0.259
Age >65 (DV)	-0.103	0.591**	-0.133*	0.593
Expensive Car (DV)	-0.082*	-0.021	-0.083*	-0.033
Extensive Usage (DV)	-0.044	0.109	-0.079**	0.092
Education: Secondary	-0.096	0.045	-0.050	0.058
Education: College	-0.147*	0.164	-0.098	0.162
ln(Fuel tank)	0.293**	-0.111	-0.001	-0.044
Engine size (lt)	-0.078	0.228	-0.094	0.271
Car Age (year)	0.001	-0.031	-0.003	-0.039
Fuel efficiency (km/lt)	0.017	-0.038	0.003	-0.038
Variance of random utility	**ethanol**	**mgasoline**	**ethanol**	**mgasoline**
Control group	1.000	1.257	1.000	1.250
	-	[1.189]	-	[1.685]
Price-ratio flyer group	0.858***	1.282	0.881***	1.225
	[0.062]	[1.245]	[0.067]	[1.667]
Km-per-50R$ group	0.984***	1.316	1.002***	1.275
	[0.088]	[1.264]	[0.092]	[1.728]
Timing FE	Yes	Yes	Yes	Yes
Vehicle brand FE	Yes	Yes	Yes	Yes
City FE	Yes	Yes		
Station FE			Yes	Yes

Station-clustered standard errors in brackets (some excluded for space)
*p<0.1, **p<0.05, ***p<0.01

Table C.3: Average marginal effects

	(1)			(2)		
	change in fuel share			change in fuel share		
	g	e	mg	g	e	mg
Price gasoline increase 10%	-0.155***	0.115***	0.037**	-0.108***	0.105***	0.003***
	[0.021]	[0.011]	[0.017]	[0.013]	[0.013]	[0.001]
Price ethanol increase 10%	0.115***	-0.114***	0.000	0.103***	-0.122***	0.019***
	[0.010]	[0.010]	[0.000]	[0.013]	[0.011]	[0.004]
Price mgasoline increase 10%	0.037**	0.000	-0.035**	0.003***	0.019***	-0.022***
	[0.016]	[0.000]	[0.016]	[0.001]	[0.004]	[0.004]
Female (vs. Male)	0.046***	-0.037***	-0.008*	0.046***	-0.035***	-0.011***
	[0.010]	[0.010]	[0.004]	[0.010]	[0.010]	[0.004]
Age 25-40 (vs. Age <25)	-0.030**	0.021	0.009	-0.028**	0.019	0.009
	[0.013]	[0.014]	[0.006]	[0.012]	[0.014]	[0.006]
Age 40-65 (vs. Age <25)	-0.038***	0.013	0.024***	-0.034***	0.012	0.022***
	[0.011]	[0.013]	[0.006]	[0.011]	[0.013]	[0.006]
Age >65 (vs. Age <25)	-0.034	-0.030	0.061***	-0.032	-0.034*	0.065***
	[0.024]	[0.021]	[0.013]	[0.024]	[0.020]	[0.013]
Edu: Secondary (vs. Primary)	0.030	-0.044*	0.014*	0.018	-0.028	0.011
	[0.025]	[0.024]	[0.008]	[0.025]	[0.024]	[0.008]
Educ: College (vs. Primary)	0.025	-0.028	0.003	0.013	-0.015	0.002
	[0.025]	[0.026]	[0.007]	[0.026]	[0.026]	[0.008]
Expensive car	0.021*	-0.015	-0.006	0.023*	-0.020	-0.003
	[0.013]	[0.012]	[0.007]	[0.013]	[0.012]	[0.008]
Extensive usage	0.002	-0.014	0.011**	0.012	-0.019**	0.007
	[0.011]	[0.009]	[0.005]	[0.010]	[0.009]	[0.005]
Car age increases 1 year	0.002	0.001	-0.003*	0.003	-0.000	-0.003
	[0.003]	[0.003]	[0.001]	[0.003]	[0.003]	[0.002]
Engine size increase 1lt	0.000	-0.024	0.023	0.010	-0.029	0.018
	[0.025]	[0.024]	[0.014]	[0.026]	[0.026]	[0.015]
Fuel tank increase 10%	-0.069*	0.087**	-0.018	-0.058	0.016	0.042
	[0.036]	[0.039]	[0.019]	[0.072]	[0.068]	[0.036]
Fuel efficiency incease 1km/lt	-0.005	0.006	-0.001	0.003	-0.001	-0.002
	[0.008]	[0.009]	[0.004]	[0.010]	[0.011]	[0.005]
Fullsize (vs Compact)	0.019	-0.030*	0.011	0.016	-0.021	0.006
	[0.016]	[0.018]	[0.011]	[0.018]	[0.019]	[0.010]
Midsize (vs Compact)	-0.028**	0.012	0.015**	-0.028**	0.014	0.014**
	[0.014]	[0.014]	[0.006]	[0.014]	[0.015]	[0.006]
Minivan (vs Compact)	-0.060*	0.023	0.035	-0.049	0.019	0.030
	[0.034]	[0.032]	[0.023]	[0.040]	[0.037]	[0.024]
Small truck (vs Compact)	-0.003	0.012	-0.008	-0.001	0.010	-0.009
	[0.021]	[0.020]	[0.009]	[0.021]	[0.020]	[0.009]
Subcompact (vs Compact)	-0.004	0.004	0.000	-0.014	0.011	0.003
	[0.010]	[0.011]	[0.005]	[0.010]	[0.011]	[0.006]
SUV (vs Compact)	-0.043	0.022	0.020	-0.033	0.026	0.008
	[0.028]	[0.027]	[0.013]	[0.027]	[0.026]	[0.011]
Belo Horizonte	0.235***	-0.224***	-0.008			
	[0.028]	[0.024]	[0.012]			
Recife	0.186***	-0.154***	-0.029***			
	[0.038]	[0.034]	[0.010]			
Curitiba	0.020	0.003	-0.021**			
	[0.029]	[0.025]	[0.011]			

Fuel: g stands for gasoline, e for ethanol, mg for midgrade gasoline
Station clustered standard errors in bracket
* p<0.1, ** p<0.05, *** p<0.01

Table C.4: Average marginal effects

	(1)		(2)	
Variance of random utility				
	ethanol	**mgasoline**	**ethanol**	**mgasoline**
Control group	1.000	1.257	1.000	1.250
	-	[1.189]	-	[1.685]
Price-ratio flyer group	0.858***	1.282	0.881***	1.225
	[0.062]	[1.245]	[0.067]	[1.667]
Km-per-50R$ group	0.984***	1.316	1.002***	1.275
	[0.088]	[1.264]	[0.092]	[1.728]
Treatment effect on log variance of random utility				
	ethanol	**mgasoline**	**ethanol**	**mgasoline**
Price-ratio flyer group	-0.153**	0.020	-0.127*	-0.021
	[0.072]	[0.073]	[0.076]	[0.069]
Km-per-50R$ group	-0.016	0.046	0.002	0.020
	[0.089]	[0.073]	[0.092]	[0.072]

Station-clustered standard errors in brackets
*p<0.1, **p<0.05, ***p<0.01

Table C.5: Estimated variance of the random utility term

	(3)		(4)		(5)		(6)	
	Car price		Usage		College		Price history	
	Low	High	Low	High	No college	College	Stable	Unstable
Panel A: Variance of random utility, ethanol								
Control (a)	1.000	1.007***	1.000	0.982***	1.000	1.093***	1.000	1.458***
Price-ratio flyer (b)	0.809***	0.875***	0.800***	0.868***	0.875***	0.860***	0.806***	1.071***
km/R$50 flyer (c)	0.966***	0.851***	0.896***	1.046***	1.014***	0.978***	0.934***	1.165***
Panel B: Variance of random utility, midgrade gasoline								
Control (d)	1.063**	0.545*	0.766**	0.619**	0.778*	0.800**	1.102**	1.879**
Price-ratio flyer (e)	0.862**	0.583**	0.601**	0.664**	0.636*	0.673**	1.081**	1.032**
km/R$50 flyer (f)	0.86**	0.518*	0.629**	0.567**	0.687*	0.644**	1.055**	1.036*
Panel C: Treatment effect on log(variance of random utility, ethanol)								
Price-ratio flyer (ln(b) - ln(a))	-0.212**	-0.141	-0.223**	-0.123	-0.133	-0.239**	-0.216*	-0.308*
	[0.105]	[0.191]	[0.105]	[0.186]	[0.167]	[0.109]	[0.114]	[0.172]
km/R$50 flyer (ln(c) - ln(a))	-0.034	-0.169	-0.110	0.064	0.013	-0.111	-0.068	-0.224
	[0.106]	[0.187]	[0.106]	[0.185]	[0.170]	[0.110]	[0.113]	[0.174]
Panel D: Treatment effect on log(variance of random utility midgrade gasoline)								
Price-ratio flyer (ln(e) - ln(d))	-0.210*	0.068	-0.242*	0.070	-0.200	-0.173	-0.019	-0.599***
	[0.127]	[0.247]	[0.134]	[0.229]	[0.197]	[0.141]	[0.141]	[0.199]
km/R$50 flyer (ln(f) - ln(d))	-0.211*	-0.050	-0.197	-0.088	-0.123	-0.218	-0.043	-0.595***
	[0.132]	[0.240]	[0.137]	[0.239]	[0.194]	[0.144]	[0.136]	[0.213]

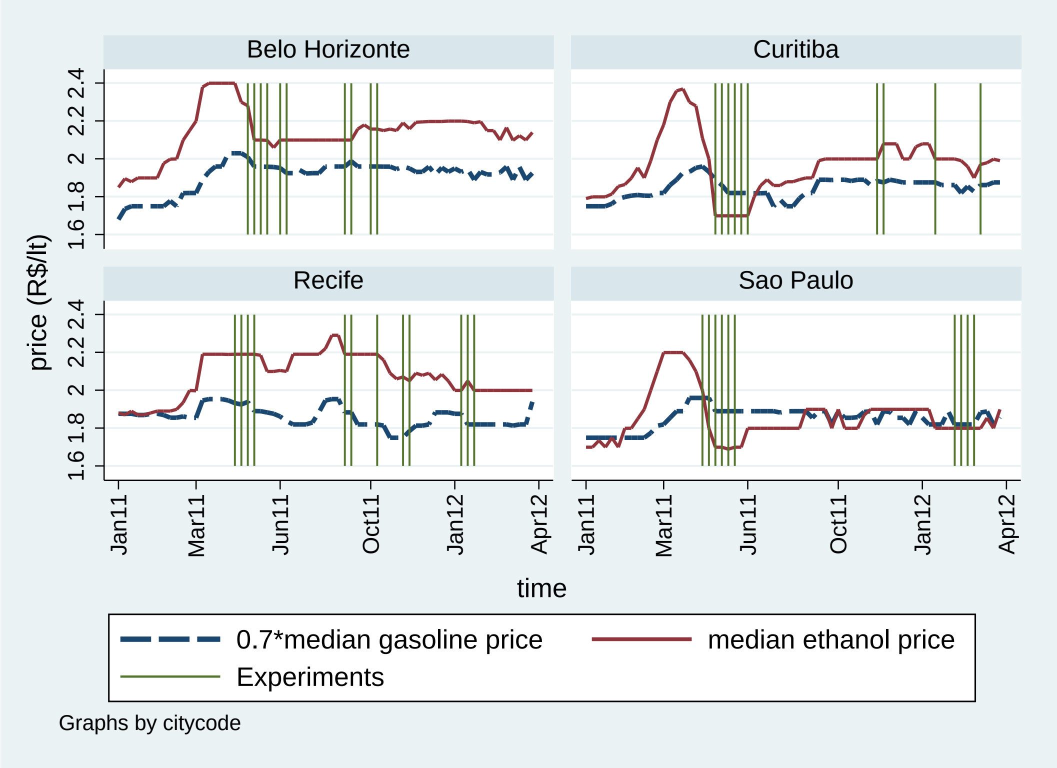
Station clustered standard errors in bracket

*p<0.1, ** p<0.05, *** p<0.01

Gasoline is the base alternative. The variance of random utility term for ethanol of the control group is set to 1. Specification (3), first 2 columns allows the covariance matrix to shift with the dummy for expensive car. Specification (4) allows the covariance matrix to shift with the dummy for extensive usage. Specification (5) for College degree and Specificaiton (6) price is stable)

Appendix D

Chapter 2: Figures



Notes: The gasoline prices are scaled down by 0.7 to be comparable to the ethanol price. All prices are median prices in a week across all sampled stations within the city. Vertical lines indicate weeks during which experiments were conducted.

Figure D.1: Price path of gasoline and ethanol by cities

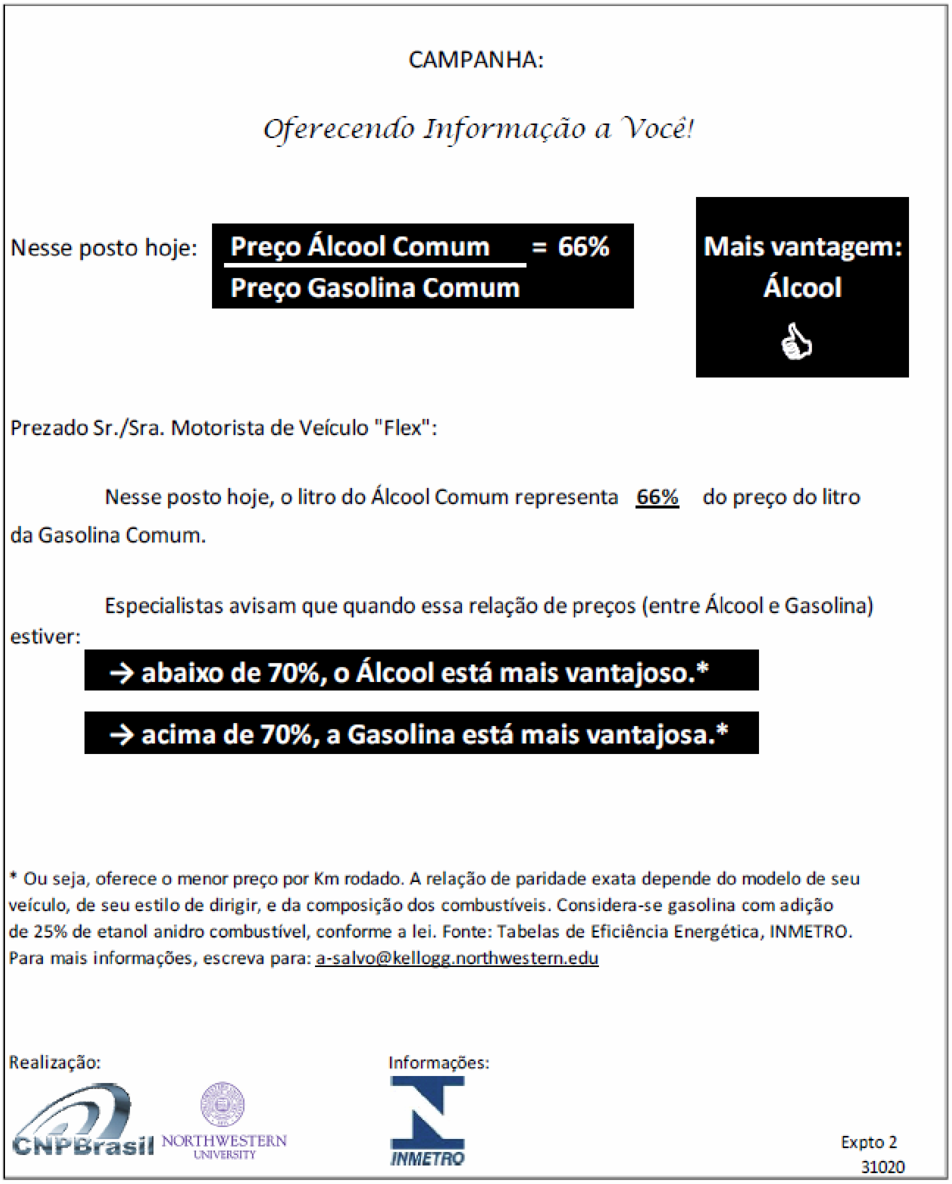
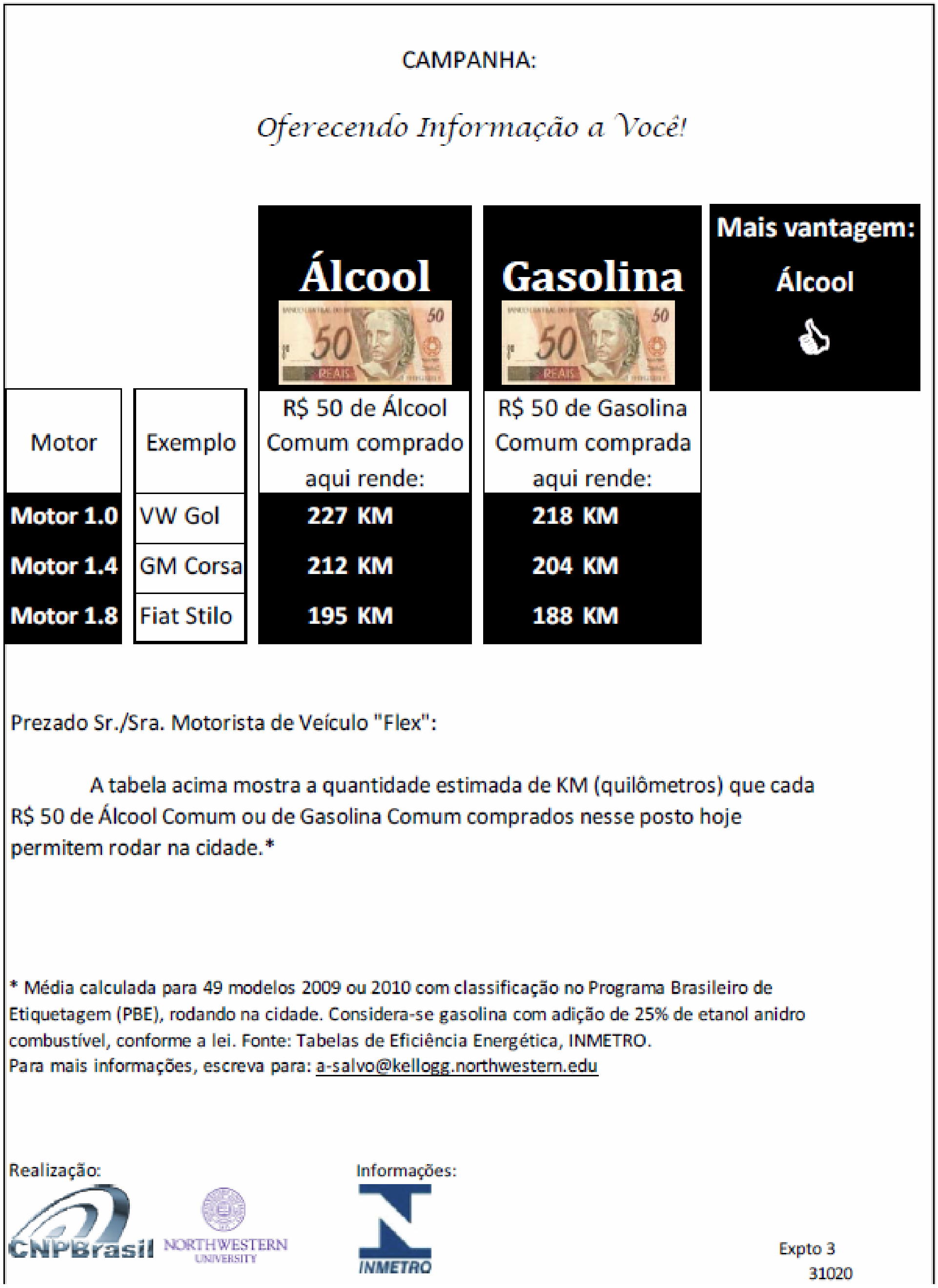


Figure D.2: A sample of Price-ratio flyer



CAMPANHA:

Oferecendo Informação a Você!

Motor	Exemplo	Álcool R$ 50 de Álcool Comum comprado aqui rende:	Gasolina R$ 50 de Gasolina Comum comprada aqui rende:	Mais vantagem: Álcool
Motor 1.0	VW Gol	227 KM	218 KM	
Motor 1.4	GM Corsa	212 KM	204 KM	
Motor 1.8	Fiat Stilo	195 KM	188 KM	

Prezado Sr./Sra. Motorista de Veículo "Flex":

A tabela acima mostra a quantidade estimada de KM (quilômetros) que cada R$ 50 de Álcool Comum ou de Gasolina Comum comprados nesse posto hoje permitem rodar na cidade.*

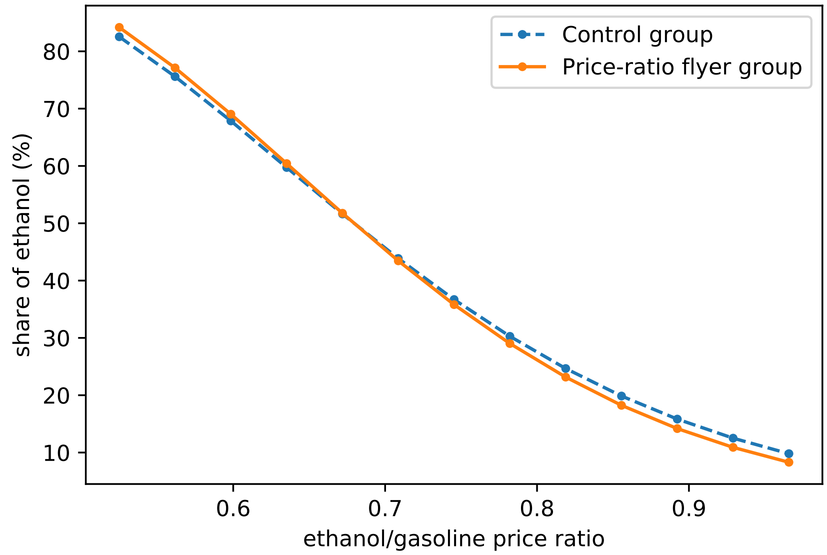
* Média calculada para 49 modelos 2009 ou 2010 com classificação no Programa Brasileiro de Etiquetagem (PBE), rodando na cidade. Considera-se gasolina com adição de 25% de etanol anidro combustível, conforme a lei. Fonte: Tabelas de Eficiência Energética, INMETRO.
Para mais informações, escreva para: a-salvo@kellogg.northwestern.edu

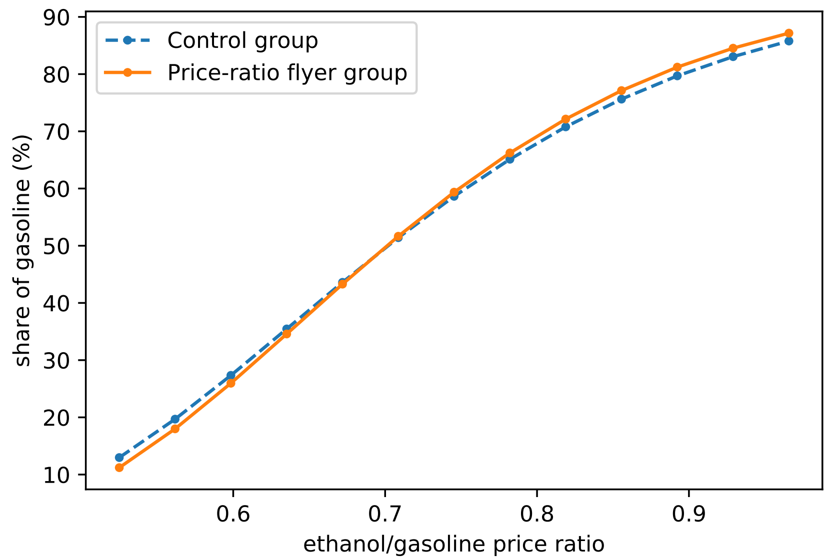
Realização: CNPBrasil NORTHWESTERN UNIVERSITY

Informações: INMETRO

Expto 3
31020

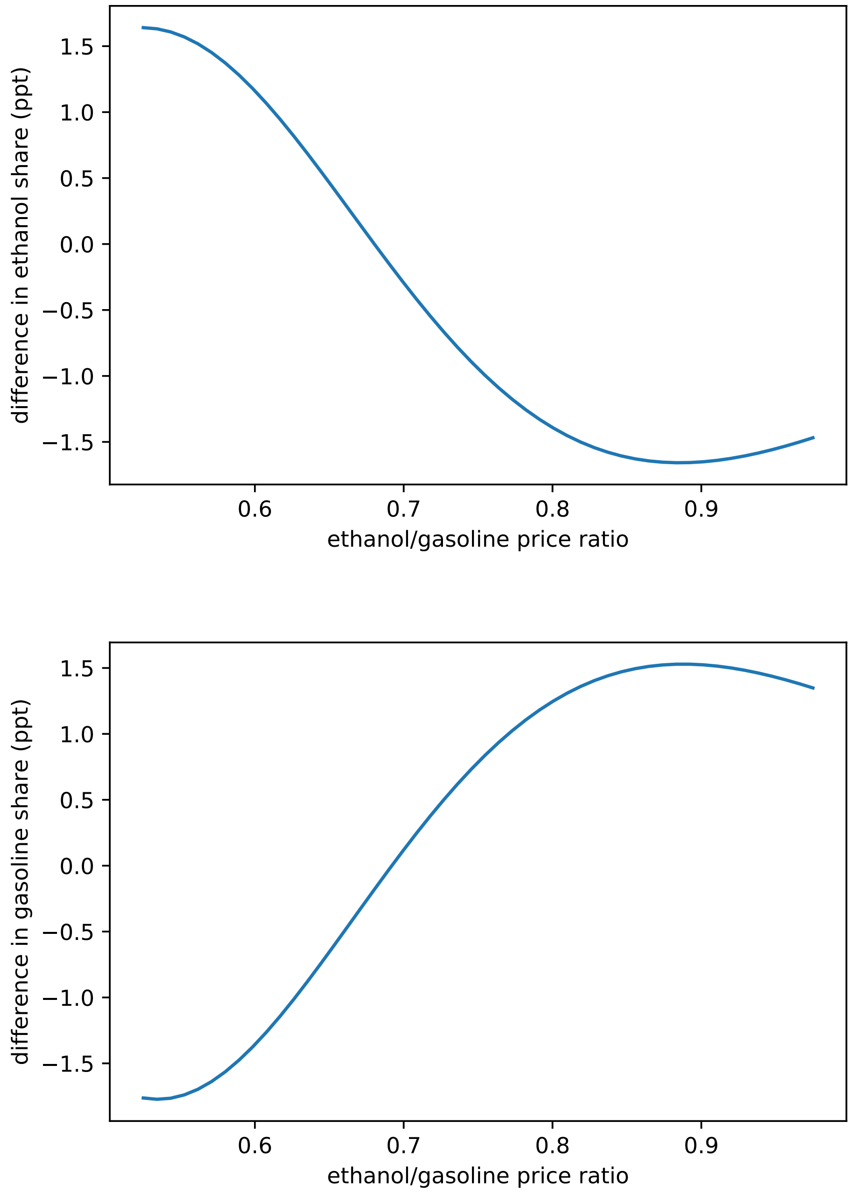
Figure D.3: A sample of Km-per-50R$ flyer





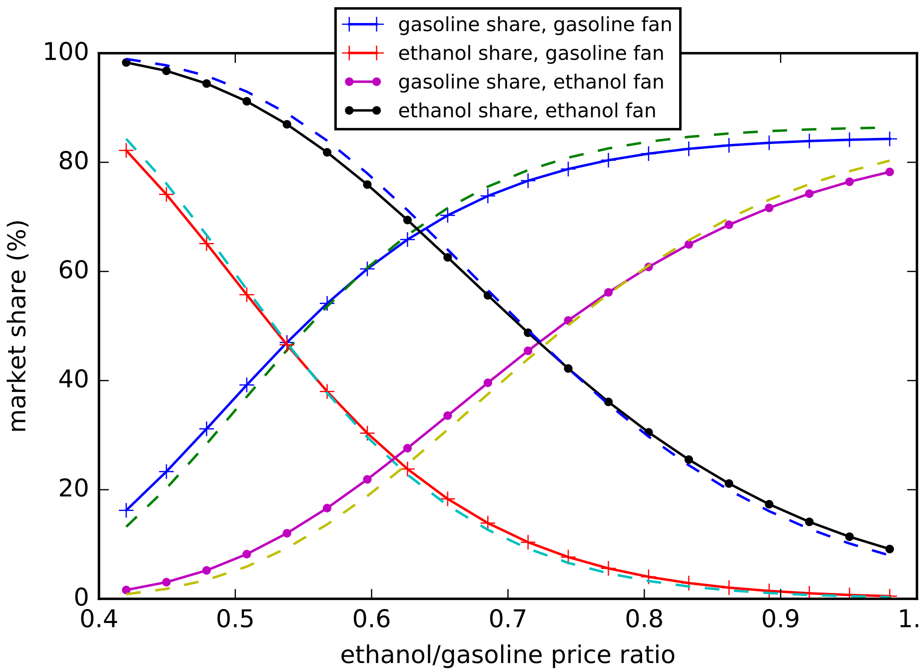
Notes:
This figures plot the ethanol share (top panel) and gasoline share (bottom panel) when ethanol-price ratio varies. Prices for gasoline and midgrade gasoline are set to the values in the data, and ethanol prices are set to the a price ratio times the gasoline price. The shares are calculated as if all consumers are in control group (the solid lines) or in price-ratio flyer group (the dashed lines)

Figure D.4: Fuel shares, control group vs price-ratio flyer group



Notes:
This figures plot changes in ethanol share (top panel) and gasoline share (bottom panel) when consumers in control group are treated with price-ratio flyer. Prices for gasoline and midgrade gasoline are set to the values in the data, and ethanol prices are set to the a price ratio times the gasoline price.)

Figure D.5: Difference in fuel shares between control and treatment groups



Gasoline fan: female, age >65, college in Belo Horizonte; ethanol fan: male, age 25 to 40, at most primary from Sao Paulo. The solid lines plot the shares as if the consumers are in the control. The dashed lines plot the shares as if the consumes are treated with price-ratio flyer.

Figure D.6: Fuel shares for a gasoline fan and an ethanol fan

Appendix E

Chapter 3: Tables

Table E.1: Simulation study results

	without fixed-payment			with fixed-payment			
	ε	σ	α	ε	σ	α	p
True value		0.67	4.50		0.67	4.50	0.20
0.45	0.56	0.60	4.39	0.46	0.67	4.50	0.20
	(0.11)	(-0.07)	(-0.11)	(0.01)	(-0.01)	(0.00)	(-0.00)
	[0.06]	[0.01]	[0.08]	[0.08]	[0.01]	[0.10]	[0.00]
0.61	0.68	0.60	4.40	0.60	0.67	4.52	0.20
	(0.07)	(-0.07)	(-0.10)	(-0.01)	(-0.00)	(0.02)	(-0.00)
	[0.07]	[0.01]	[0.09]	[0.08]	[0.01]	[0.11]	[0.00]
0.82	0.86	0.61	4.38	0.83	0.67	4.49	0.20
	(0.04)	(-0.06)	(-0.12)	(0.01)	(-0.00)	(-0.01)	(-0.00)
	[0.06]	[0.01]	[0.08]	[0.07]	[0.01]	[0.09]	[0.00]

This table reports the average of the estimates obtained from 200 simulated sets of data. The average biases are inside the brackets and the standard deviations are inside the square brackets. The true values used for simulation are reported under the header of each column.

Table E.2: Consumer characteristics and Payment types

Dependent variable	(1) odd	(2) odd	(3) 10x	(4) 10x	(5) fixed50	(6) fixed50	(7) fixed20	(8) fixed20
City: Curitiba	-0.058*		0.036*		0.134***		0.040	
	(0.024)		(0.014)		(0.030)		(0.030)	
City: Belo Horizonte	-0.109**		0.037*		0.074		0.047	
	(0.035)		(0.016)		(0.049)		(0.036)	
City: Recife	-0.166***		0.053**		0.041		0.112*	
	(0.034)		(0.019)		(0.050)		(0.045)	
Female	-0.025**	-0.026**	-0.001	-0.002	0.015	0.010	-0.004	-0.001
	(0.009)	(0.009)	(0.007)	(0.007)	(0.017)	(0.017)	(0.016)	(0.016)
Age: 25 to 40	0.028*	0.027*	0.025**	0.024*	-0.004	-0.010	0.031	0.027
	(0.010)	(0.010)	(0.009)	(0.009)	(0.028)	(0.028)	(0.024)	(0.025)
Age: 40 to 65	0.067***	0.062***	0.019	0.020*	0.063*	0.055*	0.021	0.019
	(0.011)	(0.011)	(0.010)	(0.010)	(0.027)	(0.027)	(0.025)	(0.026)
Age: >65	0.101***	0.100***	0.033*	0.038*	0.083	0.079	0.031	0.017
	(0.029)	(0.028)	(0.016)	(0.017)	(0.049)	(0.045)	(0.055)	(0.054)
Secondary	-0.044*	-0.040	0.002	-0.002	-0.091*	-0.085	-0.087*	-0.088
	(0.021)	(0.022)	(0.014)	(0.013)	(0.044)	(0.043)	(0.042)	(0.046)
College	0.028	0.025	0.012	0.009	-0.070	-0.068	-0.076	-0.073
	(0.020)	(0.022)	(0.014)	(0.014)	(0.041)	(0.041)	(0.040)	(0.045)
Expensive car	0.039***	0.038***	-0.009	-0.008	-0.038	-0.037	0.029	0.035
	(0.010)	(0.010)	(0.008)	(0.008)	(0.023)	(0.022)	(0.031)	(0.031)
Extensive usage	0.054***	0.056***	-0.009	-0.005	0.006	-0.003	0.038*	0.037*
	(0.012)	(0.012)	(0.008)	(0.008)	(0.021)	(0.021)	(0.018)	(0.017)
Car model year	0.003	0.003	0.004*	0.004*	-0.000	0.001	-0.002	-0.002
	(0.003)	(0.003)	(0.002)	(0.002)	(0.007)	(0.006)	(0.006)	(0.006)
Tank	-0.002	-0.003	-0.000	-0.000	0.001	0.001	0.003	0.002
	(0.002)	(0.002)	(0.001)	(0.001)	(0.003)	(0.003)	(0.003)	(0.003)
Fuel efficiency	-0.034	-0.046	0.002	0.008	0.039	0.054	-0.015	-0.000
	(0.024)	(0.031)	(0.013)	(0.018)	(0.036)	(0.042)	(0.028)	(0.040)
Brand: Fiat	-0.054***	-0.051**	-0.004	-0.006	-0.024	-0.029	-0.030	-0.036
	(0.015)	(0.015)	(0.010)	(0.010)	(0.029)	(0.030)	(0.027)	(0.027)
Brand: GM	-0.067***	-0.062***	-0.004	-0.003	-0.020	-0.022	-0.020	-0.023
	(0.016)	(0.015)	(0.011)	(0.011)	(0.033)	(0.034)	(0.032)	(0.032)
Brand: VW	-0.064***	-0.060***	-0.003	-0.003	-0.004	-0.009	-0.039	-0.041
	(0.016)	(0.015)	(0.009)	(0.009)	(0.031)	(0.033)	(0.027)	(0.026)
Brand: Ford	-0.052**	-0.055**	-0.013	-0.013	-0.051	-0.062	-0.019	-0.017
	(0.017)	(0.017)	(0.011)	(0.011)	(0.038)	(0.037)	(0.039)	(0.038)
Price control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Station fixed effects	no	yes	no	yes	no	yes	no	yes
Sub sample	All	All	regular payments	regular payments	payment 40 to 60	payment 40 to 60	payment 10 to 30	payment 10 to 30
N	10422	10422	8491	8491	3112	3112	3806	3806

Standard errors in parentheses
* $p<0.05$, ** $p<0.01$, *** $p<0.001$
Odd is a dummy variable for odd-payment, 10x is dummy variable for regular-0 payments, fixed50 dummy for payments exactly at 50R$ and fixed20 dummy for payments exactly 20R$. Price controls include ethanol and gasolin price and price squared.

Table E.3: Price salience treatment effects on continuous and discrete choice

	(5)	(6)	(7)
Panel A: Continuous choice equation			
Price-ratio flyer (DV) (1)	0.046**		0.074*
	[0.022]		[0.041]
km-per-50R$ flyer (DV) (2)	-0.020		-0.028
	[0.020]		[0.023]
Difference in treatment effects (1) - (2)	0.066***		0.102**
	[0.018]		[0.041]
Panel B: Discrete choice equation			
Ethanol Mean Utility			
Ethanol is favorable (DV)		17.359***	17.684***
		[3.973]	[4.090]
Price-ratio flyer (DV)		-1.840	-2.833
		[1.264]	[1.895]
km-per-50R$ flyer (DV)		0.872	1.241
		[1.405]	[1.753]
Price-ratio flyer × Ethanol is favorable (3)		4.207**	3.878**
		[1.790]	[1.794]
km-per-50R$ flyer × Ethanol is favorable (4)		0.343	0.526
		[1.981]	[2.052]
Difference in treatment effects (3) - (4)		3.863**	3.352**
		[1.516]	[1.665]
Midgrade gasoline Mean Utility			
(omitted for space)			

Station-clustered standard errors in brackets

p<0.1, **p<0.05, ***p<0.01

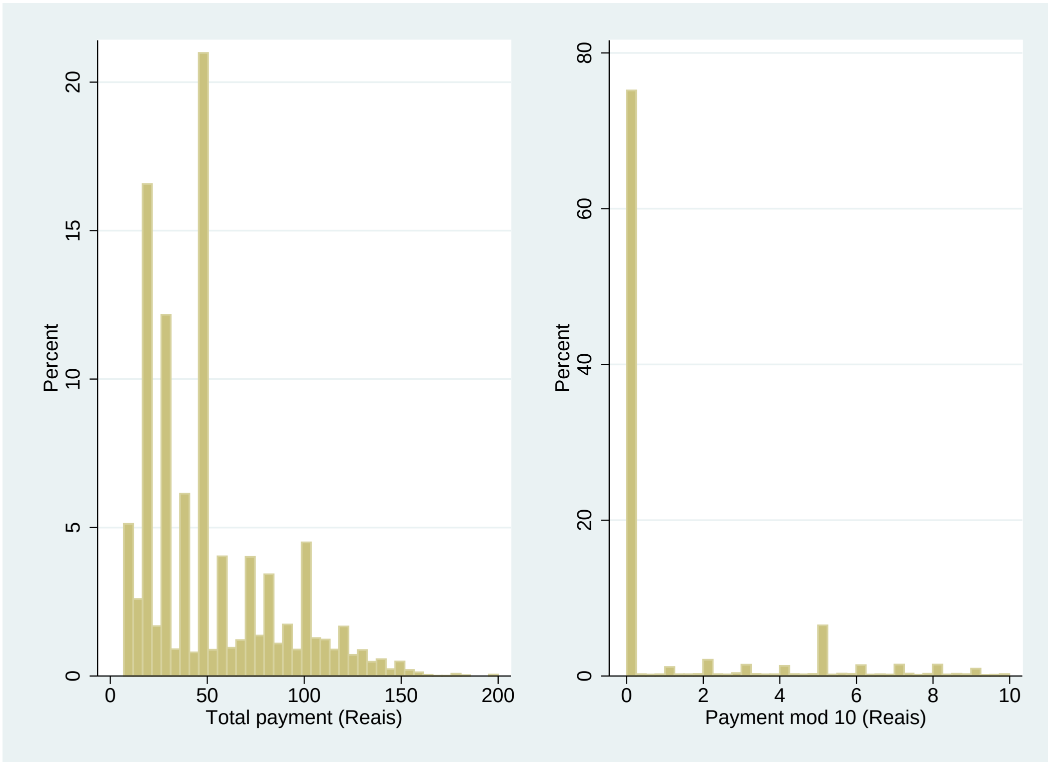
All three specifications include consumer characteristics (age, gender, education, vehicle price and usage), vehicle characteristics (engine size, fuel tank, vehicle age, vehicle class), day of week and time of day and station fixed effects to both the continuous and the discrete choice equation. "Ethanol is favorable" is a dummy varible for ethanol-gasoline price ratio less than 0.7.

Table E.4: Estimation Results

Panel A: Structural parameters				
	(1)	(2)	(3)	(4)
ln(elas)	-0.290***	-0.292**	-0.224**	-0.209*
	[0.098]	[0.123]	[0.091]	[0.110]
ln(sigma)	-0.355***	-0.354***	-0.384***	-0.384***
	[0.013]	[0.013]	[0.013]	[0.013]
ln(mu)	2.447***	2.550***	2.445***	2.547***
	[0.131]	[0.180]	[0.125]	[0.166]
logit(prob fixed-payment)	-1.391***	-1.391***	-1.391***	-1.391***
	[0.041]	[0.041]	[0.041]	[0.041]
Panel B: Continuous choice equation				
	spec 1	spec 2	spec 3	spec 4
Ethanol	0.041**	0.049*	0.031*	0.036
	[0.019]	[0.029]	[0.017]	[0.025]
Midgrade gasoline	-0.044**	-0.047	-0.034*	-0.033
	[0.020]	[0.029]	[0.018]	[0.023]
ln(fuel tank)	0.046	0.048	0.045	0.044
	[0.127]	[0.128]	[0.122]	[0.122]
Engine displacement	0.228***	0.227***	0.229***	0.229***
	[0.048]	[0.049]	[0.046]	[0.047]
Vehicle model year	0.032***	0.032***	0.035***	0.035***
	[0.006]	[0.006]	[0.006]	[0.006]
Fuel efficiency	0.013	0.012	0.004	0.003
	[0.019]	[0.021]	[0.018]	[0.020]
Vehicle class: Subcompact	-0.020	-0.020	-0.013	-0.013
	[0.017]	[0.017]	[0.018]	[0.018]
Vehicle class: Midsize	0.110***	0.111***	0.116***	0.117***
	[0.023]	[0.023]	[0.024]	[0.024]
Vehicle class: Smalltruck	0.001	-0.001	0.014	0.013
	[0.051]	[0.051]	[0.048]	[0.048]
Vehicle class: SUV	-0.021	-0.020	-0.040	-0.038
	[0.046]	[0.046]	[0.045]	[0.046]
Vehicle class: Fullsize	0.050	0.049	0.045	0.045
	[0.043]	[0.043]	[0.040]	[0.040]
Vehicle class: Minivan	0.061	0.061	0.017	0.018
	[0.068]	[0.069]	[0.065]	[0.066]
Female	0.011	0.012	-0.001	0.000
	[0.015]	[0.015]	[0.015]	[0.015]
Age: 25 to 40	0.127***	0.126***	0.119***	0.119***
	[0.025]	[0.025]	[0.024]	[0.024]
Age: 40 to 65	0.211***	0.210***	0.206***	0.206***
	[0.026]	[0.026]	[0.026]	[0.026]
Age: >65	0.279***	0.279***	0.247***	0.248***
	[0.041]	[0.041]	[0.042]	[0.042]
Education: Some secondary	0.086**	0.088**	0.057	0.059
	[0.037]	[0.037]	[0.037]	[0.038]
Education: Some college	0.285***	0.286***	0.227***	0.229***
	[0.038]	[0.038]	[0.040]	[0.040]
Vehicle price: top quartile	0.093***	0.093***	0.097***	0.097***
	[0.024]	[0.024]	[0.025]	[0.025]
Vehicle usage: top quartile	0.161***	0.161***	0.150***	0.149***
	[0.018]	[0.018]	[0.017]	[0.017]
Average vehicle price in station	0.029***	0.029***	0.003	0.003

	[0.007]	[0.007]	[0.008]	[0.008]
Curitiba	0.034	0.033		
	[0.039]	[0.039]		
Belo Horizonte	0.084	0.088*		
	[0.051]	[0.051]		
Recife	0.036	0.039		
	[0.062]	[0.062]		
Day of week	Yes	Yes	Yes	Yes
Time of day	Yes	Yes	Yes	Yes
Station effects	No	No	Yes	Yes
Panel C: Discrete choice - ethanol utility				
	spec 1	spec 2	spec 3	spec 4
ln(fuel tank)	6.756	7.033	6.653	6.951
	[4.557]	[5.291]	[4.534]	[5.214]
Engine displacement	-2.129	-1.347	-2.135	-1.377
	[1.772]	[1.925]	[1.760]	[1.910]
Vehicle model year	0.351***	0.480***	0.336***	0.466***
	[0.011]	[0.013]	[0.011]	[0.013]
Fuel efficiency	-0.722	-0.715	-0.706	-0.704
	[0.609]	[0.691]	[0.604]	[0.685]
Vehicle class: Subcompact	0.488	1.171	0.502	1.185
	[0.720]	[0.829]	[0.719]	[0.821]
Vehicle class: Midsize	1.803*	1.927*	1.764*	1.896*
	[0.993]	[1.161]	[0.980]	[1.133]
Vehicle class: Smalltruck	0.424	0.588	0.448	0.646
	[1.428]	[1.660]	[1.420]	[1.649]
Vehicle class: SUV	0.541	0.462	0.583	0.513
	[1.906]	[2.180]	[1.891]	[2.162]
Vehicle class: Fullsize	-1.320	-1.210	-1.356	-1.250
	[1.236]	[1.407]	[1.229]	[1.403]
Vehicle class: Minivan	-0.136	0.150	-0.169	0.039
	[2.265]	[2.737]	[2.259]	[2.711]
Female	-2.997***	-3.470***	-2.989***	-3.480***
	[0.786]	[1.019]	[0.779]	[0.990]
Age: 25 to 40	2.405**	2.848**	2.359**	2.788**
	[1.074]	[1.294]	[1.057]	[1.255]
Age: 40 to 65	2.684***	3.023**	2.622***	2.944**
	[1.012]	[1.272]	[0.994]	[1.226]
Age: >65	0.773	0.823	0.709	0.730
	[1.589]	[1.801]	[1.592]	[1.799]
Education: Some secondary	-1.326	-0.487	-1.363	-0.591
	[1.793]	[2.090]	[1.790]	[2.090]
Education: Some college	-0.516	0.180	-0.606	-0.021
	[1.672]	[1.902]	[1.668]	[1.898]
Vehicle price: top quartile	-0.091	-0.182	-0.115	-0.206
	[0.834]	[0.971]	[0.828]	[0.954]
Vehicle usage: top quartile	-0.668**	-0.353	-0.677***	-0.402
	[0.263]	[0.468]	[0.262]	[0.469]
Average vehicle price in station	0.693	-0.008	0.636	-0.091
	[0.666]	[0.716]	[0.674]	[0.734]
Curitiba	-1.572		-1.555	
	[1.474]		[1.460]	
Belo Horizonte	-21.213***		-21.149***	
	[3.917]		[3.759]	
Recife	-18.516***		-18.406***	
	[3.668]		[3.546]	
Day of week	Yes	Yes	Yes	Yes

Time of day	Yes	Yes	Yes	Yes
Station effects	No	Yes	No	Yes
Panel D: Discrete choice - midgrade gasoline utility				
	spec 1	spec 2	spec 3	spec 4
ln(fuel tank)	0.887	2.876	1.122	3.216
	[7.639]	[8.847]	[7.607]	[8.791]
Engine displacement	1.395	3.832	1.393	3.866
	[3.143]	[3.489]	[3.161]	[3.486]
Vehicle model year	0.494***	0.659***	0.525***	0.693***
	[0.018]	[0.022]	[0.018]	[0.022]
Fuel efficiency	-1.008	-0.661	-1.058	-0.700
	[0.968]	[1.069]	[0.978]	[1.071]
Vehicle class: Subcompact	1.140	1.749	1.137	1.703
	[1.278]	[1.411]	[1.274]	[1.402]
Vehicle class: Midsize	2.789**	3.338**	2.871**	3.428**
	[1.404]	[1.519]	[1.391]	[1.506]
Vehicle class: Smalltruck	-2.811	-3.149	-2.862	-3.183
	[2.730]	[3.151]	[2.714]	[3.115]
Vehicle class: SUV	3.373	3.738	3.219	3.648
	[2.389]	[2.786]	[2.392]	[2.770]
Vehicle class: Fullsize	1.488	0.657	1.500	0.690
	[2.469]	[2.842]	[2.436]	[2.799]
Vehicle class: Minivan	3.493	5.969	3.386	5.915
	[3.814]	[4.504]	[3.780]	[4.433]
Female	-2.652**	-3.578***	-2.651**	-3.558***
	[1.115]	[1.274]	[1.101]	[1.231]
Age: 25 to 40	3.024	2.623	3.086	2.706
	[2.071]	[2.241]	[2.054]	[2.221]
Age: 40 to 65	5.952***	6.059***	6.009***	6.156***
	[1.961]	[2.181]	[1.948]	[2.152]
Age: >65	11.077***	13.373***	11.126***	13.554***
	[2.749]	[3.703]	[2.706]	[3.603]
Education: Some secondary	-0.448	-0.421	-0.364	-0.276
	[2.226]	[2.452]	[2.217]	[2.446]
Education: Some college	0.565	0.351	0.728	0.658
	[2.187]	[2.376]	[2.155]	[2.352]
Vehicle price: top quartile	-2.162	-2.755	-2.117	-2.710
	[1.709]	[1.930]	[1.714]	[1.908]
Vehicle usage: top quartile	-1.222***	-1.101*	-1.203***	-1.044*
	[0.469]	[0.595]	[0.462]	[0.579]
Average vehicle price in station	0.866	0.570	0.933	0.719
	[1.320]	[1.557]	[1.299]	[1.499]
Curitiba	-8.492***		-8.360***	
	[2.994]		[2.977]	
Belo Horizonte	-10.793***		-10.678***	
	[3.000]		[2.963]	
Recife	-16.327***		-16.226***	
	[3.450]		[3.394]	
Day of week	Yes	Yes	Yes	Yes
Time of day	Yes	Yes	Yes	Yes
Station effects	No	Yes	No	Yes



The left panel plot the histogram of the payment itself. The right panel plots the histogram of the last digit of the payment (so multiples of 10 will be combined into the bin at 0).

Figure F.1: Histogram of total payments

Appendix F

Chapter 3: Figures

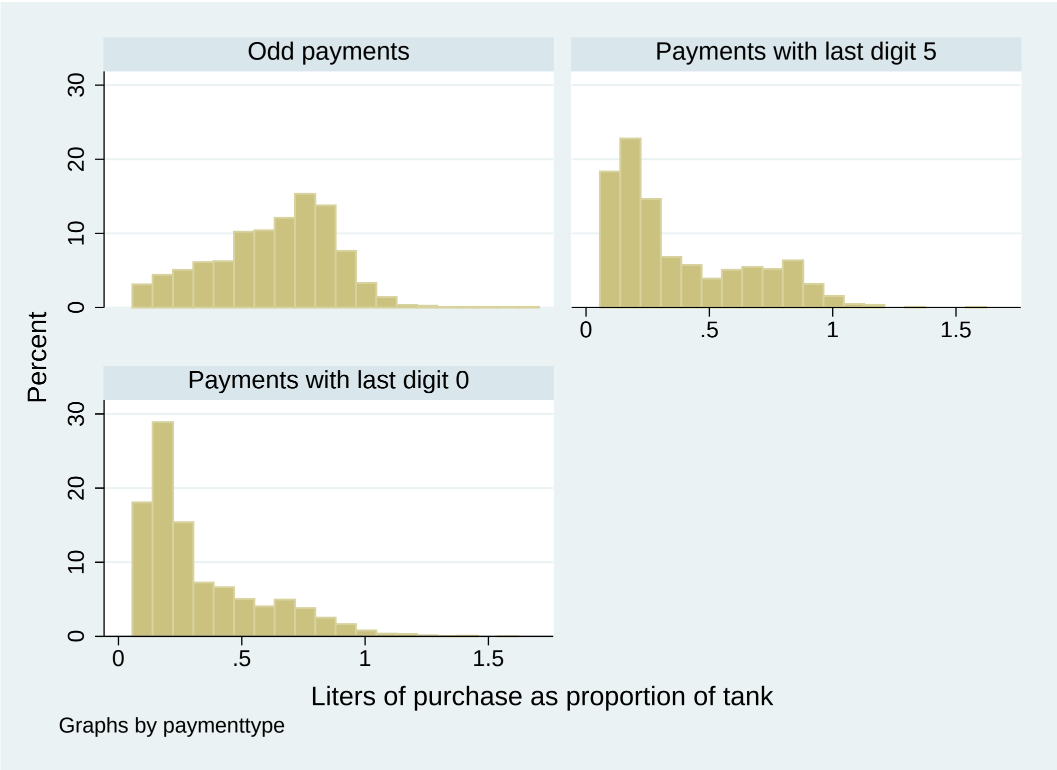


Figure F.2: Histogram of liters of purchase as proportion of fuel tank by payment types

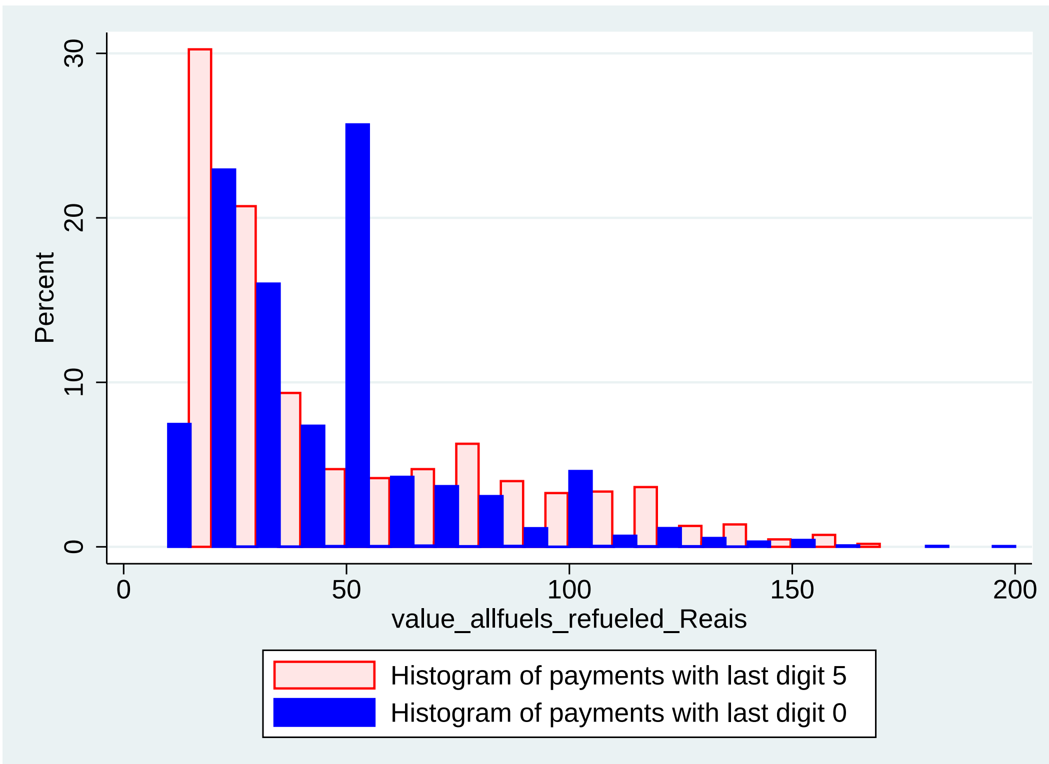


Figure F.3: Histograms of payments with last digit 5 and with last digit 0