Essays on differentiated products in oligopoly markets

by

Jinjing Luo

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Economics

Program of Study Committee:
GianCarlo Moschini, Major Professor
Gabriel E. Lade
Harvey E. Lapan
Sergio H. Lence
Amy Froelich

The student author, whose presentation of the scholarship herein was approved by the program of study committee, is solely responsible for the content of this dissertation. The Graduate College will ensure this dissertation is globally accessible and will not permit alterations after a degree is conferred.

Iowa State University

Ames, Iowa

2020

DEDICATION

To my parents and grandparents

TABLE OF CONTENTS

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Dr. GianCarlo Moschini, for his unlimited support and advice throughout my days at Iowa State University. This dissertation is not possible without his help. His guidance to be an independent and thoughtful researcher will have a life-long impact on me. I am also grateful to Dr. Edward Perry for his help on the third chapter of this dissertation. His insightful comments and inspiring advice are invaluable. I also want to express my gratitude to my department. It provides great opportunities to communicate and carry out ideas. I have received constructive comments from the department members.

I would also like to thank my committee members Dr. Harvey Lapan, Dr. Gabriel Lade, Dr. Sergio Lence, and Dr. Amy Froelich for their time and comments.

Finally, I am grateful to my family, especially my husband Li Xiang, for their company, encouragement, and support.

ABSTRACT

The common theme running through my dissertation work is differentiated products in oligopoly markets with emphasis on the demand analysis. In the markets, consumers generally present heterogeneous preferences even for the same product, which brings additional dimension of differentiation for the market and add more complexity for the equilibrium. The purpose of this dissertation is to qualify and quantify consumers' purchase behavior in relevant markets, and to further investigate the effects on firms' pricing strategy, the market structure, and the effectiveness of policy regulations.

The second chapter studies consumers' heterogeneous demands for two transportation fuels, E10 and E85, and evaluates the Renewable Fuel Standard (RFS) program that promotes the consumption of E85. The RFS sets mandate volumes for the consumption of biofuels. Because E85 contains more ethanol than E10, the Renewable Fuel Standard Program conveys a subsidy to E85, relative to E10. A central question for RFS policy concerns the pass-through of the induced subsidy to consumers. To investigate the issue, we develop a structural model suitable to study the pass-through rate at retail level. We consider two dimension of heterogeneity: consumers' randomly distributed locations, which make fuel products provided by different gas stations horizontally differentiated; and their heterogeneous propensities for the two different fuel products, which are vertically differentiated. Accordingly, we build the model from the "Hotelling" framework and extend it to also include the vertical dimension of heterogeneity. The model provides a natural way to represent the role of imperfect competition at the retail level. We then calibrate the model parameters using real world data and simulate the Nash Equilibrium results of the model. Our results show that the pass-through rate is about 0.7 at the

baseline level of model parameters. More market power, higher subsidy level, and consumers' higher preferences for subsidized product are related to lower pass-through rate.

The third chapter empirically models heterogeneity in the seed demand and also focuses on a crucial issue in farmers' purchase behavior—brand inertia. Brand inertia has long been a topic of interest in the Economics and Marketing literature, and it refers to the situation that an individual is more likely to choose the brand he/she has purchased previously. Among the potential explanations for this tendency, researchers have been particularly interested in the importance of state dependence, defined as the causal dependency of an individual's future choices on their current state. In this chapter, we develop and estimate a micro-level random coefficient logit model—the random coefficient is to capture unobserved heterogeneity over farmers and state dependence is modeled by incorporating the previous purchased brand. Our results show that on average, farmers are willing to pay an additional \$5.31/unit for a brand if it was purchased in the previous period, equivalent to about 12% of the average retail price, and there is substantial heterogeneity in the estimates. In the counterfactual analysis of two temporary shocks, price discount and late technology innovation adoption, we find that state dependence implies long-lasting effect on farmers' choice decision.

CHAPTER 1. INTRODUCTION

My dissertation has focused on problems at the intersection of industrial organization, agricultural economics, and environmental economics. The second chapter evaluates the effectiveness of the Renewable Fuel Standard (RFS) problem under imperfect competition in the transportation fuel market, which is an oligopoly market with differentiated fuel products. The third chapter studies farmers' purchase behavior in the soybean seeds market, which has been characterized by structural changes and consolidation.

The RFS program sets annual nested mandates for the consumption of renewable fuels. The mandate is implemented through Renewable Identification Numbers (RIN), the price of which quantifies the extent of the subsidy for ethanol and tax for fossil gasoline. Because E85 (51%-83% of ethanol depending on seasonality) contains substantially more ethanol and less gasoline than E10 (10% of ethanol), positive RIN prices translate into a policy-induced subsidy for E85, relative to E10. A central question for RFS policy concerns the pass-through of RIN prices to consumers. Recent empirical work, however, has raised questions on the extent of this pass-though, and further indicated that market power at retail level might be the key. To investigate the issue, we develop a structural model suitable to study the pass-through rate at retail level: our model is rooted in Hotelling's horizontal differentiation framework, which is extended to also represent the imperfect substitutability between E10 and E85 (a vertical product differentiation attribute). We solve for the Nash equilibrium results where gas stations maximize their profits and consumers choose from different fuel products, which gas station to refuel, and whether refill with E10 or E85. The model is framed under two forms of competition, one is oligopoly and the other is monopoly, to capture how market competition affects the pass-through rate; and three cases of E85 penetration, no E85 stations, one of the two gas stations with E85, and both stations with E85, to explore how specifically market competition over E85 affects the pass-through rate. In most cases, analytical results are not approachable. We then calibrate the model parameters using real world data and simulate the Nash Equilibrium results of the model under all different scenarios.

Chapter 3 focuses on the U.S. soybean seeds market and uses survey data of real-world seed purchase transactions. Over the last few decades, the seed industry has been characterized by considerable growth and consolidation. Much of this has been driven by the development and rapid diffusion of genetically engineered (GE) crops. Because of the great efficiency GE products provide for weed and pest management, they have been widely adopted since the introduction in 1996. Taking advantage of our data, which also starts from 1996, we can estimate farmers' propensity for this new technology over time. Moreover, we observe from the data that farmers are more likely to purchase the same brand over time, i.e., the brand inertia behavior. We focuses on the state dependence explanation of brand inertia, which is of particular interest to researchers and has importation implications on firm's pricing incentives, the market structure, and the persistence in brand shares over time. To estimate state dependence and GE propensities, we develop a micro-level random coefficient model, which allows substantial heterogeneity over farmers and also captures the complex substitution patterns across different products. From the model estimates, we further derive the distribution of farmers' WTP for attributes of interest and simulate the own-price and cross-price demand elasticities for each product. To investigate the implication of state dependence in the seed market, the counterfactual analysis compares the real-world scenario and the scenario without state dependence. By imposing a temporary shock, we show that state dependence implies long-lasting effects of farmers' choice decision. Specifically, in the experiment of late GT adoption as a temporary shock, firms can loss more than 50% market shares in the current year and suffer from less market shares even they pick up the GE technology late. Our results show that a major technology is crucial for seed companies, which somehow explains the rapid adoption pattern of the GE technology.

CHAPTER 2. PASS-THROUGH OF POLICY-INDUCED E85 SUBSIDY: INSIGHTS FROM THE HOTELLING'S MODEL

Modified from a manuscript published in Energy Economics

Jinjing Luo and GianCarlo Moschini

Department of Economics, Iowa State University, Ames IA 50010

Abstract

We build a structural model of imperfect competition for a retail market that supplies both low-ethanol (E10) and high-ethanol (E85) gasoline blends. The model permits us to study some impacts of the E85 subsidy induced by the U.S. Renewable Fuel Standard, specifically how the pass-through of this subsidy to retail prices is affected by market power. The model is rooted in Hotelling's horizontal differentiation framework, which is extended to also represent the imperfect substitutability between E10 and E85 (a vertical product differentiation attribute). The model naturally captures two sources of imperfect competition in the fuel market—refueling stations' market power arising from their spatial location, and limited availability of E85 stations. We derive both analytical and numerical solutions for Nash equilibrium outcomes under various scenarios. In our baseline parameterization, when the penetration of E85 stations is incomplete, we find that the pass-through rate is about 0.7. Complete penetration of E85 stations leads to near complete pass-through, notwithstanding the market power enjoyed by stations because of their spatial location. With monopolistic market power (e.g., collusion), however, with full penetration of E85 stations the pass-through rate is lower. Moreover, when market power only arises from location differentiation (duopoly model with full penetration of E85), the pass-through rate converges to one as the subsidy gets large, whereas it converges to zero if a station has

exclusivity in selling E85 (partial penetration of E85) or there is collusion/monopoly power from collusion.

1. Introduction

Over the last decade, the United States has implemented ambitious policies designed to drastically increase the share of renewable energy used for transportation fuels. In particular, the Energy Independence and Security Act of 2007 (EISA) greatly expanded the Renewable Fuel Standard (RFS), which entails a set of nested quantity “mandates” for several forms of renewable fuels. The schedule originally envisioned by EISA contemplated the use of biofuels growing to 36 billion gal by the year 2022 (Schnepf and Yacobucci, 2013). This bold target has had to be scaled back somewhat, by repeated waivers by the Environmental Protection Agency, because of the apparent failure of commercially viable cellulosic biofuel supply. Still, the non-cellulosic portion of these mandates—mainly corn-based ethanol, but also advanced biofuels such as biodiesel and sugarcane-based ethanol—are still being pursued at the full level envisioned by EISA (21 billion gal by the year 2022). As these mandate levels have grown over the years, the “blend wall” has materialized. This concept refers to the bottleneck that arises when the total quantity of ethanol to be blended into the gasoline supply exceeds 10%.

To understand the root of the blend wall problem, one must note that ethanol is blended into the fuel supply essentially by way of two distinct blends: low-ethanol E10 fuel (which contains 10% ethanol) and high-ethanol E85 fuel (which contains anywhere from 51% to 83% ethanol, depending on seasonality). E10 can be used by all cars, whereas E85 can be used only by flexible fuel vehicles (FFVs). Insofar as meeting rising EISA's mandates requires ethanol use in excess of 10% of the total gasoline use, increased consumption of E85 is necessary. By 2013 the RFS required 13.8 billion gal of ethanol, which constituted 10.3% of gasoline consumption, thus exceeding the E10 blend wall (Stock, 2015). It has become apparent, however, that increasing consumers' use of E85 is problematic (Collantes, 2010; Pouliot and Babcock, 2014). Specific constraints are due to the low number of FFVs (which at present make up approximately 8.5% of the fleet of cars and light trucks), and the limited availability of E85 service stations (E85 pumps are available at about 3% of stations).[1] Furthermore, previous work has found that the majority of FFV drivers actually do not fill up their vehicles with E85.

Demand for E85 depends crucially on the distinctive features of the choice problem faced by FFV drivers. E85 has considerably less energy content than E10—on average, a gallon of E85 only delivers about 80% of the miles of a gallon of E10. Consumers who ultimately care about the cost per mile traveled, therefore, would require the price of E85 to be suitably discounted relative to that of E10 in order to be enticed to buy (Collantes, 2010; Pouliot and Babcock, 2017). Furthermore, because a tank of E85 delivers fewer miles than E10, consumers face an additional convenience cost because of the need for more frequent refueling stops. All this suggests that, from a vertical product differentiation perspective, E85 is an inferior product relative to E10. Such a conclusion may be partially offset, however, if consumers attached some utility to the consumption of renewable energy per se, perhaps because of their beliefs about the lower carbon emission of ethanol relative to fossil gasoline (i.e., consumers may have “green” preferences).[2] But there is also a horizontal differentiation component to E85 demand, which is related to the relative scarcity of E85 refueling stations: other things equal, drivers located farther from an E85 station will be less willing to refill their tank with this fuel (relative to the ubiquitous E10).

[1] The Alternative Fuels Data Center of the U.S. Department of Energy reports 3322 E85 stations in 2017. The total number of gasoline stations in the United States, which has been declining over time, was reported to be 114,474 in 2012 by the U.S. Bureau of the Census.

[2] Indeed, in an early empirical analysis of E85 demand, Anderson (2012) found that a substantial fraction of FFV drivers were willing to pay a premium for E85 fuel.

The heterogeneity of consumers, vis-à-vis the structural determinants of demand, suggests that E10 and E85 are imperfect demand substitutes. If expanding consumption of E85 is required to overcome the blend wall, the structural determinants of E85 demand will matter in translating price signals into consumption decisions. Such price signals are supposed to be induced by the policy design of the RFS. As shown by Lapan and Moschini (2012), in a competitive setting, a quantity mandate is isomorphic to a combination of a tax (on fossil fuel) and a subsidy (for ethanol) that is revenue neutral. The use of renewable identification numbers (RINs) to enforce the mandates makes this equivalence transparent. These tradable instruments command a price whenever the mandate is binding, which reflects the cost of complying with the RFS at the margin (Korting and Just, 2017). The price of RINs quantifies the extent of the subsidy for ethanol and tax for fossil gasoline. Because E85 contains substantially more ethanol and less gasoline than E10, positive RIN prices translate into a policy-induced subsidy for E85, relative to E10.

If the policy-induced subsidy for E85 is fully reflected in retail fuel prices, then consumers would be given the proper market signal: tighter RFS mandates would lead to higher RIN prices, increasing the spread between E10 and E85 RIN prices, ultimately resulting in higher E85 consumption. Recent empirical work, however, has raised questions on the extent of this pass-though. Knittel et al. (2017) find that, whereas RIN price pass-through is fairly complete at the wholesale level, there appears to be little or no pass-through of RIN prices to retail E85 prices. Market power is the chief market imperfection typically invoked to rationalize a less-than-full pass-through of an (exogenous) cost differential. Lade and Bushnell (2019) emphasize the dynamics of the process and, unlike Knittel et al. (2017), find that pass-through of E85 subsidy is on average one-half to three-quarters. Li and Stock (2019), utilizing station-level data from Minnesota from 2012 to 2015, estimate the state-level pass-through rate for E85 to be approximately 0.5. Similar to Lade and Bushnell (2019), they also find that indicators of local monopoly for E85 stations correlate with lower pass-through.

As noted by Knittel, Meiselman, and Stock (2017, p. 1082), "… a central question for RFS policy is whether this pass-through occurs at the retail level." The empirical contributions cited in the foregoing do not provide a conclusive answer, and in fact suggest the possibility that imperfect competition at the retail level may be key. Complete pass-through is a feature of perfect competition, in the sense that goods are priced at marginal costs. The rate by which equilibrium prices are affected by a given cost change, though, in general depends on the elasticities of demand and supply. With constant marginal costs, as maintained below, the E85 subsidy is fully transmitted to retail prices under perfect competition. Under imperfect competition, however, the pass-through rate also depends on conduct parameters for market power and the shape of demand curves (Weyl and Fabinger, 2013). In the context of gasoline retail, what pass-through should one expect? And, what are the critical factors affecting the subsidy pass-through rate?

To address these questions, this paper develops a structural model suitable to study the pass-through of the policy-induced E85 subsidy to retail prices. Our model is rooted in Hotelling's (1929) spatial competition model, a standard approach in industrial organization whereby firms are endowed with some market power by the presumption of product differentiation. This structure captures in a natural way the heterogeneity of consumers, vis-à-vis the locations of refueling stations. We extend this horizontal product differentiation framework to also accommodate consumers' heterogeneous preferences with respect to E85, a vertical differentiation feature that appears essential to the context being modeled. Consideration of more than one dimension of product differentiation makes models rather unwieldy.[3] The model we develop is, inevitably, rather stylized. Yet the model captures important features of the structure of demand for E10 and E85 discussed earlier, and provides a natural way to represent the role of imperfect competition at the retail level. The advantage of a structural model, albeit a stylized one, is that of providing a vehicle by which we can isolate the effects of various factors and assess their contribution toward favoring or impeding pass-through. As such, this paper complements the emerging literature, noted earlier, on the empirical assessment of RIN pass-through to gasoline prices (Knittel et al., 2017; Lade and Bushnell, 2019; Li and Stock, 2019).

To investigate the impact of E85 availability, our benchmark model is a duopoly setting with incomplete penetration of E85 stations—specifically, two stations, only one of which sells E85. As a comparison, we also investigate the duopoly model with only E10 fuel (this is the basic Hotelling's model), and the duopoly model where both firms sell both fuels (E10 and E85). Except for the basic Hotelling duopoly model, and one of the monopoly models we consider, analytic solutions for the Nash equilibrium are not possible in our extended models. Hence, we solve our models numerically. To that end, we calibrate the values of parameters in the various models and solve for Nash equilibrium results under alternative parameter settings.

Our results show that pass-through is incomplete with incomplete penetration of E85 stations. In the benchmark model, we estimate the pass-through rate to be about 0.7. When firms have no exclusivity of selling E85 (i.e., E85 is offered at all locations), however, the pass-through is near complete even though firms still have some market power from horizontal differentiation. Moreover, our results show that the market generally exhibits lower pass-through

[3] Conceptual models include Neven and Thisse (1990), Ferreira and Thisse (1996), and Gabszewicz and Wauthy (2012). Recent applications include Brécard (2014), Di Comite et al. (2014), Norman et al. (2016), and Pennerstorfer (2017).

when the E85 subsidy is higher, and when market power is higher (the two stations are a monopoly). Our results also show that prices of E10 at both locations barely change with the E85 subsidy (under the working assumption that the cost of E10 fuel is constant). Interestingly we show that, in the model with partial penetration of E85 stations, the E10 price at the same location with E85 is higher than the E10 price at the other location in duopoly, whereas the relation reverses in monopoly. For demands, we show that E85 consumption increases with the subsidy, as expected, and the decrease in E10 consumption is larger at stations with both E10 and E85 pumps, no matter whether in duopoly or in monopoly.

The paper is organized as follows. Section 2 provides more details on framing pass-through rate in our setting. In Section 3, we specify drivers' preferences and derive demand functions for various cases of interest. In Section 4, we calibrate some of the critical parameters of the model. Section 5 presents the results for the main duopoly models we consider. Section 6 considers the issue of market power further, in the context of a monopolized market. Finally, Section 7 concludes the paper.

2. Background

We begin with a simple example that shows how subsidy passthrough works in the context of the simplest Hotelling model of horizontal product differentiation. We then review how the E85 subsidy arises from the basic mechanisms of the RFS. Both of these discussions point to the usefulness of looking at the subsidy pass-through to the spread between E10 and E85 retail prices, a feature that we will then investigate with the analytical model that we develop below.

2.1. A simple motivating example

To motivate the analysis that we propose, consider the textbook linear-city setup where two firms are located at the extreme of a line of unit length, each offering a product to a population of consumers, with unit demand, who are uniformly distributed on the unit segment (see, e.g., Tirole, 1988, pp. 279–280). The two products are perceived as imperfect substitutes, by any one consumer, because of the consumer's own location. To fix ideas, think of the product sold by firm A at location L0 as E10 gasoline, and the product sold by firm B at location L1 as E85 gasoline (presently we will discuss why this is too simplistic, and how the model needs to be generalized). The two firms compete in prices by setting p_A and p_B, respectively. If the consumers' reservation value for one unit of either good is sufficiently large, so that the market is covered, then the demand functions facing the two firms are easily obtained:

$$q_i = \frac{1}{2} + \frac{p_j - p_i}{2t} \quad , \quad i,j = A,B\,, i \neq j \tag{1}$$

where $t > 0$ is the "travel cost" parameter. Suppose the firms have constant per-unit costs c_A and c_B, respectively. It is readily found that the Nash equilibrium prices are:

$$p_i^* = t + \frac{2c_i + c_j}{3} \quad , \quad i,j = A,B\,, i \neq j \tag{2}$$

In this setting, we ask what the implications would be of a tax/subsidy on these products. Suppose first that a per-unit subsidy $s > 0$ is provided to both products. Then it is readily seen that both equilibrium prices decline by exactly the amount s (i.e., there is 100% pass-through), and the equilibrium quantities are unchanged. Alternatively, suppose that only product B enjoys the per-unit subsidy s.[4] It is easy to see that, in this case

$$\frac{\partial p_A^*}{\partial s} = -\frac{1}{3} \text{ and } \frac{\partial p_B^*}{\partial s} = -\frac{2}{3} \tag{3}$$

[4] For example, considering the RIN obligations in 2013, Stock (2015) concluded that the net effect of RIN prices was a near-zero subsidy for E10 ($0.01/gal) and a large subsidy for E85 ($0.50/gal).

Interestingly, both prices decline in equilibrium. The subsidy-induced decline in product B's marginal cost provides an incentive to reduce p_B. Furthermore, because prices are "strategic complements" in this setting, this leads firm A to also reduce its price. In equilibrium, although the subsidy does not apply to both products, the subsidy to product B reduces both prices.

If the purpose of the subsidy is to incentivize consumption of product B, then from the demand functions above it is clear that what matter is the net effect of the subsidy on the *price difference* $(p_A^* - p_B^*)$. When the subsidy only applies to product B, the pass-through rate on this price difference is:

$$\frac{\partial\left(p_A^* - p_B^*\right)}{\partial s} = \frac{1}{3} \tag{4}$$

Thus, we conclude that, in this imperfectly competitive setting, the effectiveness of the subsidy to promote use of product B is blunted by the exercise of market power from horizontal differentiation, and by the fact that prices are strategic complements.

This model is clearly too simplistic to capture the stylized facts of E10 and E85 gasoline retail: in the foregoing, the two products are sold at different locations by different firms, whereas in reality E10 and E85 are typically marketed by refueling stations that sell both products; furthermore, this purely horizontal setting does not capture the vertical differentiation dimension that is an essential feature of E10 and E85 gasoline demand. Below we provide a suitably generalized model. Yet, this simple model shows some of the reason why limited pass-through may arise in an imperfectly competitive setting, and why it is instructive to look at the subsidy pass-through on the products' equilibrium price difference.

2.2 Pass-through of energy policy effects

Conceptually, the pass-through rate measures how consumer price is affected by a small change in a per-unit tax or subsidy. In the context of the RFS, framing the pass-through of the policy-implied tax/subsidy effects requires some attention to the mechanisms by which RIN prices affect costs of producers and retailers. In Knittel, Meiselman, and Stock (2017) RIN pass-through rate is measured as the partial effect of RIN obligation spread on retail price spread for E10 and E85. Lade and Bushnell (2019) measures how retail price of E85 changes in response to the change in E85 RIN subsidy. Li and Stock (2019) note that the changes in wholesale spread of E10 and E85 are mainly driven by fluctuations in RIN prices and estimate how the E85 retail price responses to the changes in the wholesale spread. As explained in more details below, in this study we define the pass-through rate as the impact of the E85 subsidy, induced by the RFS, on the spread between E10 and E85 retail prices.

To illustrate how the RFS tax/subsidy implications filter to retail prices in the context of the model we develop, we presume a competitive refining/blending industry that operates under constant returns to scale.[5] Specifically, we consider a simplified structure where fossil gasoline and corn ethanol, which are blended into E10 and E85, can be obtained at constant per-unit costs. Such "producer prices" are denoted p_g and p_e, respectively. [All prices are expressed in natural units, e.g., $/gallon]. To translate such prices into the costs faced by retailers, we start by noting that the RFS requires obligated parties (e.g., refineries) to retire a bundle of RINs (associated with the requirements of the RFS nested mandates) for every gallon of fossil gasoline sold. Let

---

[5] This is assumption is attractive for its simplicity, but it is also consistent with the pass-through evidence presented by Pouliot, Smith, and Stock (2017) who estimate RIN pass-through at the rack and cannot reject complete RIN pass-through to wholesale fuel prices. Knittel. Meiselman, and Stock (2017) also find complete pass-through of RIN prices to wholesale gasoline prices.

the cost of this bundle be denoted by B, and let R denote the RIN price associated with ethanol (i.e., the price of D6 RINs).[6]

Throughout the paper, we assume E10 includes 10% ethanol and E85 uses 74% ethanol (as maintained by Knittel, Meiselman, and Stock 2017). Then, wholesale prices of blended fuels satisfy:

$$p_{E10}^{w} = 0.10\left(p_e - R\right) + 0.90\left(p_g + B\right) \tag{5}$$

$$p_{E85}^{w} = 0.74\left(p_e - R\right) + 0.26\left(p_g + B\right) \tag{6}$$

The retailing industry takes the wholesale prices p_{E10}^{w} and p_{E85}^{w} as given, such that their retailing costs can be represented as $c_{E10} = p_{E10}^{w} + \mu$ and $c_{E85} = p_{E85}^{w} + \mu$, where μ is the sum of motor fuel taxes and per-unit marketing/retailing costs.

In the absence of the RFS policy (i.e., $R = 0$ and $B = 0$), the cost of the two fuel blends to retailers are fully determined by the producer prices p_g and p_e (plus the aforementioned term μ). The RFS, however, introduces product-specific tax/subsidies equal to $(0.1R - 0.9B)$ for E10 fuel and $(0.74R - 0.26B)$ for E85 fuel (as in Knittel, Meiselman, and Stock 2017). A more insightful analysis of how policy measures are passed through to retail prices can be obtained by looking at the cost advantage for the E85 fuel, defined as $c_{E10} - c_{E85}$. From the foregoing, it is clear that this cost advantage is partly determined by the producer prices p_g and p_e (which we will hold constant in our analysis), and by the net subsidy to the E85 fuel (relative to E10). Specifically:

---

[6] Using the percentage standards for the 2017 year, B is determined by the obligated party's need to "retire" for each gallon of fossil fuel sold, 0.0167 D4 RINs to meet the biodiesel mandate, 0.0238 D4 or D5 RINs to meet the total advanced biofuel mandate, and 0.107 D4, D5 or D6 RINs to meet the total renewable fuel mandate.

$$c_{E10} - c_{E85} = 0.64\left(p_g - p_e\right) + s \tag{7}$$

where s denotes the per-gallon subsidy implied by the policy:

$$s \equiv 0.64\left(R + B\right) \tag{8}$$

The central question we want to address in this paper concerns how the RFS policy provides signals to consumers, vis-à-vis their choice of fuel type. A "more stringent" RFS policy would entail higher RIN prices, increasing both the D6 RIN price R and the cost of the RIN bundle obligations B, which translate directly into an increase in the relative subsidy s enjoyed by E85. As noted, with a competitive retailing sector the retail prices satisfy $p_{E10} = c_{E10}$ and $p_{E85} = c_{E85}$. Thus, from (7), the "pass-through" rate of the policy subsidy would be $\partial\left(p_{E10} - p_{E85}\right)/\partial s = 1$. That is, the policy subsidy s enjoyed by E85 would be completely reflected in the retail price spread, i.e., completely passed through to consumers. The model that we develop permits us to investigate the extent to which such a complete pass-through fails under the assumed imperfect competition setting. This way of characterizing pass-through is similar to the approach used by Knittel, Meiselman, and Stock (2017). Their (empirical) motivation for looking at the price spread between E10 and E85 was different from ours, but the fact remains that looking at the price spread between E10 and E85 provides a clean and informative summary on the nature of the pass-through effects.

3. The Model

We study the duopoly setting where two stations are located at either end of the unit segment (these locations are labeled L0 and L1, respectively). A useful model for comparison is that where both stations only sell E10 fuel, which corresponds to the basic Hotelling duopoly model. Our main model is that with incomplete penetration of E85 stations: the station at L0 offers both E10 and E85, whereas the gas station at L1 only sells the conventional E10. Finally, we also consider the case of complete penetration of E85, where both firms sell both types of fuels. The two stations maximize their separate profits, and we derive the Nash equilibrium of the non-cooperative game. To proceed, however, we first need to specify consumers' preferences.

3.1 Preferences

A unit mass of consumers are uniformly distributed on $[0,1]$. Each consumer has a car, either a normal car or an FFV. The proportion of cars that are FFV is α. Consumers can fill the tank at either station to drive a distance of M miles. Her utility from driving one mile is denoted as $\tilde{u}$. If $x \in [0,1]$ denotes a consumer's own location, she incurs a cost (disutility) of $\tilde{t}x$ when refueling at L0, and $\tilde{t}(1-x)$ when refueling at L1, where the parameter $\tilde{t} > 0$ captures the intensity of consumers' cost due to their heterogeneous location attribute. This cost is meant to capture the disutility associated with the time and travel cost associated with a refueling stop, and it is independent of the type of fuel purchased. The prices for the two fuels of interest are denoted by p_j^ℓ, where the superscript $\ell \in \{0,1\}$ denotes the location of the station and the subscript $j \in \{A,B\}$ denotes the type of fuel.[7] Note that, for notational simplicity, the subscript A will refer to E10 fuel, and the subscript B will refer to E85 fuel (as a mnemonic, B = biofuel). For a consumer located at x, if she chooses to refuel with E10 at L0, the payoff associated with this choice would be

$$U_A^0 = \tilde{u}M - p_A^0 \cdot \left(\frac{M}{\phi_A} \right) - \tilde{t}x \cdot \left(\frac{M}{k\phi_A} \right) \tag{9}$$

[7] These prices are quoted in natural units (e.g., \$/gallon). Because of the lower energy content of E85 fuels, some authors prefer to express prices (and quantities) at equal energy content. For example, Pouliot and Babcock (2014) measure the quantity of E10 in E85-equivalent units, with prices appropriately scaled. As will become apparent in what follows, in our context it is more instructive to deal with quantity and prices for both fuels in their natural units, and to separately keep track of the lower energy content of E85.

where ϕ_j is the efficiency of fuel j (miles per gallon), and k denotes the size of the tank (gallons). Thus, M/ϕ_A measures the number of gallons needed for M miles, and $M/k\phi_A$ is the number of service stops needed for M miles.

If the consumer owns an FFV and chooses to refuel with E85 at L0, her payoff from driving M miles would be

$$U_B^0 = \tilde{u}M - p_B^0 \cdot \left(\frac{M}{\phi_B} \right) - \tilde{t}x \cdot \left(\frac{M}{k\phi_B} \right) \tag{10}$$

Note that drivers who refuel with E85 are penalized from more frequent refueling (because $\phi_B < \phi_A$, then $M/k\phi_B > M/k\phi_A$), and how much they are penalized depends on their location, x.

Some normalizations permit us to simplify these payoffs without loss of generality. Specifically, let $t \equiv \tilde{t}/k$, normalize $M/\phi_A = 1$, re-define $u \equiv \tilde{u}M$, and let $\lambda \equiv \phi_A/\phi_B$. Then the payoffs in (9) and (10) can be re-written as:

$$\begin{aligned} U_A^0 &= u - p_A^0 - tx \\ U_B^0 &= u - \lambda p_B^0 - \lambda tx \end{aligned} \tag{11}$$

Here, t is a re-scaled Hotelling travel cost parameter, and the coefficient λ captures the energy efficiency of E10 relative to E85 (i.e., the energy content in one gallon of E10 is equivalent to that of λ gallons of E85). Also, $\lambda = 1.25$ is a known constant in our model.[8]

This formulation maintains a systematic vertical ranking between E10 and E85: if the two fuels were priced equally in energy equivalent terms (i.e., $p_A^0 = \lambda p_B^0$) then E85 would be dominated by E10 for all consumers (because $\lambda > 1$). But we augment this basic structure by introducing a vertical differentiation parameter to capture the fact that consumers may have

[8] Given the assumed 74% average ethanol content of E85, and the U.S. Energy Information Administration (EIA) energy content of gasoline and ethanol (see, e.g., Moschini, Lapan, and Kim 2017), then $\lambda = 1.25$.

heterogeneous attitudes towards E85. Specifically, they may perceive some extra benefit from using E85 because they consider E85 a "green" fuel with lower environmental impact, or they may associate a lower payoff to E85 because of lack of awareness of E85 (Pouliot, Liao, and Babcock 2018) or because of the lower driving range permitted by one tank of E85 (Collantes 2010). To account for these effects, we replace the term u by $(u+\theta)$ in the payoff associated with E85, where the parameter $\theta \in \left[\underline{\theta}, \bar{\theta}\right]$, with $\underline{\theta} < 0$ and $\bar{\theta} > 0$, captures drivers' heterogeneous attitude towards E85. In solving the model we will assume that θ is uniformly distributed on the support $\left[\underline{\theta}, \bar{\theta}\right]$. In the paper, we refer to θ as drivers' type and x as their location.

This representation of consumers' utility nests two dimensions of product differentiation. The parameter θ measures the degree of consumer heterogeneity with respect to vertical differentiation, whereas the location x characterizes their horizontal differentiation. As both $\bar{\theta}$ and $\underline{\theta}$ approach zero, the heterogeneous component of vertical differentiation disappears. The parameter t measures the intensity of preferences vis-à-vis horizontal differentiation. As $t \to 0$, horizontal differentiation disappears.

3.2 The model with one E85 station at location L0

This is our main model. Both gas stations provide E10, whereas only the gas station at L0 provides E85. For owners of normal cars, the payoffs associated with the two possible choices are:

$$\begin{cases} U_A^0 = u - p_A^0 - tx & \text{if refuel with E10 at L0} \\ U_A^1 = u - p_A^1 - t(1-x) & \text{if refuel with E10 at L1} \end{cases} \tag{12}$$

Drivers of FFVs face a richer set of alternatives, however: the choice of station, and whether to refuel with E10 or E85. Hence, the payoffs associated with the choices available to an FFV driver are:

$$\begin{cases} U_A^0 = u - p_A^0 - tx & \text{if refuel with E10 at L0} \\ U_A^1 = u - p_A^1 - t(1-x) & \text{if refuel with E10 at L1} \\ U_B^0 = u + \theta - \lambda p_B^0 - \lambda tx & \text{if refuel with E85 at L0} \end{cases} \quad (13)$$

For normal car drivers, their choice of E10 at L0 or E10 at L1 depends on their location and relative prices of E10 at two locations. FFV drivers can choose E10 at L0, E10 at L1, or E85 at L0 based on their type x and location x, as well as relative prices of each fuel, p_A^0, p_A^1, and p_B^0. To get the aggregate demands of each fuel, now we inquire into each driver's decision on location and type to refuel.

For given fuel prices, normal car drivers' payoffs from refueling at location L0 and L1 are shown as the orange line and the blue line respectively in Figure 1. The indifferent E10 consumer, identified by the condition $U_A^0 = U_A^1$, has location

$$\tilde{x} = \frac{1}{2}\left[\frac{p_A^1 - p_A^0}{t} + 1\right] \quad (14)$$

So, normal car drivers located at the left of $\tilde{x}$ refuel with E10 at L0, whereas they choose to refuel at L1 if located at the right of $\tilde{x}$.

Figure 1 illustrates FFV drivers' payoffs from each refueling option. If they choose to refuel with E10, the orange line and the blue line represent their payoff just as for normal car drivers. If they choose to refuel with E85, the payoff from refueling falls into the area between two parallel black lines (associated with the upper and lower bounds of the θ parameter) because of the heterogeneous preferences over E85. In Figure 1, $\left[u - \lambda p_B - \lambda tx + \bar{\theta}\right]$ represents the payoff from E85 for a consumer with the highest preference for ethanol, whereas $\left[u-\lambda p_B-\lambda tx+\underline{\theta}\right]$ represents the payoff from E85 for a consumer with the lowest preference for ethanol; $\left[u-p_A^0-tx\right]$ is the payoff from E10 at L0, and $\left[u-p_A^0-t(1-x)\right]$ is the payoff from E10 at L1. FFV drivers choose the type of fuel that maximizes their payoff.

For drivers who are located at the left of $\tilde{x}$ (defined by equation (14)), the option of E10 at L1 is dominated by E10 at L0. The fuel choice is determined by comparing payoffs from the left two refueling options at L0. The indifferent consumer, identified by the condition $U_B^0=U_A^0$, has a heterogeneity parameter that also depends on her location:

$$\tilde{\theta}(x)\equiv(\lambda-1)tx+\lambda p_B^0-p_A^0 \tag{15}$$

Hence, a driver would choose to refuel with E85 at L0 if her type satisfies $\theta>\tilde{\theta}(x)$, and she would choose E10 at L0 otherwise.

For drivers located at the right of $\tilde{x}$, the choice of E10 at L0 is dominated by E10 at L1. The coordinates of the indifferent consumer identified by the condition $U_B^0=U_A^1$, for $x\geq\tilde{x}$, are:

$$\tilde{\theta}(x)\equiv(\lambda+1)tx-t+\lambda p_B^0-p_A^1 \tag{16}$$

An FFV driver chooses to refuel with E85 at L0 if $\theta>\tilde{\theta}(x)$, and use E10 at L1 otherwise. Note that the indifferent consumer type is an increasing function of x in both equations (15) and (16), which implies that only FFV drivers with high enough preferences would choose E85 if they are located farther from the E85 station at L0. Evaluating equation (16) at $\theta=\overline{\theta}$ and inverting it yields the farthest location consistent with a possible choice of E85 at L0:

$$\overline{x}^1=\frac{p_A^1-\lambda p_B^0+\overline{\theta}+t}{(\lambda+1)t} \tag{17}$$

This point is shown in Figure 1 as the location of the intersection of U_B^0 and U_A^1 at $\theta=\overline{\theta}$.

FFV drivers' choices are depicted in Figure 2, where the x axis denotes FFV drivers' location and the y axis denotes the type θ. In the rectangle $[0,1]\times[\underline{\theta},\bar{\theta}]$, FFV drivers at the top left would choose to refuel with E85 from L0—these are FFV drivers who are close to L0 and have high preferences for E85; FFV drivers at the bottom left would choose to refuel with E10 from L0—these are FFV drivers who are close to L0 and have low preferences for E85; FFV drivers on the right portion of the rectangle would choose E10 at L1—these are FFV drivers who are close to L1. The threshold $\tilde{\theta}^0$ (type of consumer at $x=0$ who is indifferent between E10 at L0 and E85 at L0) in Figure 2 is obtained from evaluating equation (15) at $x=0$, while $\tilde{\theta}$ (type the consumer at $x=\tilde{x}$ who is indifferent between E10 at L0 and E85 at L0 and E10 at L1) is derived from evaluating the same equation at $x=\tilde{x}$.

Let d_A^0, d_A^1, and d_B^0 denote the market demands for E10 at L0, E10 at L1, and E85 at L0 respectively. These demand functions can be obtained by integrating individual demands over the distributions of individual characteristics. Given the assumed uniform distribution of x and θ, then, from Figure 2, we can express d_A^0, d_A^1, and d_B^0 as functions of the threshold levels in the (x,θ) space:

$$\begin{aligned} d_A^0 &= (1-\alpha)\tilde{x} + \alpha\tilde{x}\frac{\tilde{\theta}+\tilde{\theta}^0-2\underline{\theta}}{2(\bar{\theta}-\underline{\theta})} \\ d_B^0 &= \lambda\alpha\left[\tilde{x}\frac{2\bar{\theta}-\tilde{\theta}-\tilde{\theta}^0}{2(\bar{\theta}-\underline{\theta})} + (\bar{x}^1-\tilde{x})\frac{\bar{\theta}-\tilde{\theta}}{2(\bar{\theta}-\underline{\theta})}\right] \\ d_A^1 &= (1-\alpha)(1-\tilde{x}) + \alpha\left[(\bar{x}^1-\tilde{x})\frac{\tilde{\theta}+\bar{\theta}-2\underline{\theta}}{2(\bar{\theta}-\underline{\theta})} + 1-\bar{x}^1\right] \end{aligned} \tag{18}$$

In equation system (18), d_A^0 and d_A^1 are sums of E10 demands from both normal car drivers and FFV drivers. In d_A^0, $(1-\alpha)\tilde{x}$ is the demand of normal car drivers for E10 at L0: $(1-\alpha)$ is the fraction of normal car drivers in the market, and $\tilde{x}$ measures the fraction of normal car drivers who refuel with E10 at L0 (normal car drivers who are located at the left of $\tilde{x}$). The second element of d_A^0 involves the term $0.5\tilde{x}\left(\tilde{\theta}+\tilde{\theta}^0-2\underline{\theta}\right)/\left(\bar{\theta}-\underline{\theta}\right)$, which is the fraction of FFV drivers who choose to refuel with E10 at L0 (indicated by the area of the bottom left trapezoid in Figure 2). The demand for E10 at L1, d_A^1, is constructed in the same way. For d_B^0, λ captures the E85 demand of a single FFV driver (recall that consumers need λ gallons of E85 to drive the same number of miles as one gallon of E10). The expression in the square brackets represents the fraction of FFV drivers who choose E85 at L0 (a sum of the area of a trapezoid and the area of a triangle in Figure 2). Because we have normalized the mass of drivers in the market to one, and the market is covered, then $d_A^0+d_A^1+d_B^0/\lambda=1$.

Recalling that the threshold levels in the (x,θ) space are themselves functions of the fuel prices p_A^0, p_A^1, and p_B^0, equation (18) implicitly defines the demand functions facing the retailing stations. Actually, with different combinations of prices, the demands of FFV drivers for each fuel may differ from what is shown in Figure 2. There are five scenarios in total (in addition to the extreme cases where no FFV driver chooses E85, or all FFV drivers refuel with E85). What is illustrated in Figure 2 is the case that arises under the baseline parameter values (discussed below) and it implies that FFV drivers do not refuel with E85 if they are located far enough from the E85 station and/or they have low enough preferences for E85. The other four scenarios are shown in Figure 3. When p_B^0 is relatively high, only FFV drivers with higher preferences for E85 and who are close to L0 would choose E85, i.e., for some consumers it might be that $\tilde{\theta}>\bar{\theta}$ (recall that $\tilde{\theta}$ is from evaluating equation (15) at $x=\tilde{x}$); this is "case 1" in Figure 4. "Case 2" is the baseline scenario illustrated in Figure 2 and already discussed in the foregoing. With relatively low p_B^0, it may be the case that $\tilde{\theta}^0<\underline{\theta}$ (recall that $\tilde{\theta}^0$ is from evaluating equation (15) at $x = 0$), so that even FFV drivers with low preferences for E85 choose to refuel with E85; such a situation arises for "case 3", "case 4," and "case 5" in Figure 3. Case 5 differs from case 4 in that all FFV drivers on the left of $\tilde{x}$ choose to refuel with E85. $\underline{x}^0$ and $\underline{x}^1$ in Figure 3 are obtained from evaluating equation (15) and (16), respectively, at $\theta = \underline{\theta}$; $\overline{x}^0$ and $\overline{x}^1$ are obtained from evaluating these two equations, respectively, at $\theta = \overline{\theta}$. A summary of all threshold levels is reported in Table A1 in Appendix A, which also provides more details on how the associated demand structure evolves with different combinations of fuel prices.

3.3 The model with both E10 and E85 at both stations

To capture the effects of market penetration of E85 stations, we next consider the situation where both gas stations offer both E10 and E85. For normal car drivers, payoffs from using E10 at two locations are still as in equation (12). As for FFV drivers, their payoffs for the various choice possibilities are as follows:

$$\begin{cases} U_A^0 = u - p_A^0 - tx & \text{if refuel E10 at location0} \\ U_A^1 = u - p_A^1 - t(1-x) & \text{if refuel E10 at location1} \\ U_B^0 = u - \lambda p_B^0 - \lambda tx + \theta & \text{if refuel E85 at location0} \\ U_B^1 = u - \lambda p_B^1 - \lambda t(1-x) + \theta & \text{if refuel E85 at location1} \end{cases} \tag{19}$$

We denote the locations of indifferent consumers for E10 and E85 as $\tilde{x}_A$ and $\tilde{x}_B$, respectively. From the conditions $U_A^0 = U_A^1$ and $U_B^0 = U_B^1$, we have:

$$\begin{aligned} \tilde{x}_A &= \frac{1}{2} + \frac{1}{2t}\left(p_A^1 - p_A^0\right) \\ \tilde{x}_B &= \frac{1}{2} + \frac{1}{2t}\left(p_B^1 - p_B^0\right) \end{aligned} \tag{20}$$

In this symmetric model, without loss of generality, assume $\tilde{x}_A \geq \tilde{x}_B$, which implies $p_A^1 - p_A^0 \geq p_B^1 - p_B^0$. For drivers who are located at the left of $\tilde{x}_A$, fuel option of E10 at L1 is dominated by the option of E10 at L0; for drivers who are located at the left of $\tilde{x}_B$, the fuel option of E85 at L1 is dominated by the option of E85 at L0. The FFV driver who is indifferent between E10 and E85 at L0 is identified by equation (15), and the FFV driver who is indifferent between the two fuels at L1 is identified by the condition $U_A^1 = U_B^1$, yielding

$$\tilde{\theta}(x) \equiv (\lambda - 1)t(1 - x) + \lambda p_B^1 - p_A^1 \tag{21}$$

The demands of FFV drivers for each fuel are illustrated in Figure 4. As before, $\tilde{\theta}^0$ is still from evaluating equation (15) at $x = 0$, and $\tilde{\theta}_B$ is by evaluating the equation at $x = \tilde{x}_B$. Similarly, $\tilde{\theta}_1^1$ is obtained from evaluating equation (21) at $x = 1$, and $\tilde{\theta}_A$ is by evaluating the equation at $x = \tilde{x}_A$. FFV drivers' fuel choices, under all combinations of θ and x, are shown in Figure 4.

Demands for fuels in the market are denoted as d_A^0, d_A^1, d_B^0, and d_B^1. As before, these demands are obtained by integrating individual demands over the distribution of individual characteristics on $[0,1] \times [\underline{\theta}, \bar{\theta}]$. Summing up the demands of all FFV drivers and normal car drivers, we can get

$$\begin{aligned}
d_A^0 &= (1-\alpha)\tilde{x}_A + \alpha\left[\tilde{x}_B \frac{\tilde{\theta}_B + \tilde{\theta}^0 - 2\underline{\theta}}{2(\bar{\theta} - \underline{\theta})} + (\tilde{x}_A - \tilde{x}_B)\frac{\tilde{\theta}_B + \tilde{\theta}_A - 2\underline{\theta}}{2(\bar{\theta} - \underline{\theta})} \right] \\
d_A^1 &= (1-\alpha)(1-\tilde{x}_A) + \alpha(1-\tilde{x}_A)\frac{\tilde{\theta}_A + \tilde{\theta}_1^1 - 2\underline{\theta}}{2(\bar{\theta} - \underline{\theta})} \\
d_B^0 &= \lambda\alpha\left[\tilde{x}_B \frac{2\bar{\theta} - \tilde{\theta}_B - \tilde{\theta}^0}{2(\bar{\theta} - \underline{\theta})} \right] \\
d_B^1 &= \lambda\alpha\left[(\tilde{x}_A - \tilde{x}_B)\frac{2\bar{\theta} - \tilde{\theta}_B - \tilde{\theta}_A}{2(\bar{\theta} - \underline{\theta})} + (1-\tilde{x}_A)\frac{2\bar{\theta} - \tilde{\theta}_A - \tilde{\theta}_1^1}{2(\bar{\theta} - \underline{\theta})} \right]
\end{aligned} \tag{22}$$

Again recalling that the threshold levels in the (x, θ) space are themselves function of the fuel prices, equation (22) implicitly defines the demand functions facing the retailing stations.

Of course, in this symmetric model, in equilibrium the fuel prices will satisfy $p_A^1 = p_A^0$ and $p_B^1 = p_B^0$ (as we assume the costs of the same fuels are the same for two stations). When both stations provide E85, again we find that multiple demand configurations can arise, depending on fuel prices. What we discuss in the text is the case that arises under the baseline parameter values (see below). It implies that, with full penetration of E85 stations, any FFV driver may choose to refuel with E85 if she has a high enough type θ (regardless of her location x). There are a total of three cases. In addition to that illustrated in Figure 4, there is the scenario under high E85 price, such that $\tilde{\theta}^B > \bar{\theta}$ and $\tilde{\theta}^A > \bar{\theta}$ (only FFV drivers close to either station with high preferences for E85 would choose to refuel with E85) and the scenario under low E85 price such that $\tilde{\theta}^0 < \underline{\theta}$ and $\tilde{\theta}^1 < \underline{\theta}$ (most FFV drivers would select E85). When retail price of E85 is too high or too low relatively, there might be no consumption of E85 or all FFV drivers may choose to refuel with E85. A full discussion of all these cases can be found in Appendix A.

3.4 Nash equilibrium

Nash equilibrium requires that each station's choice of prices be a payoff-maximizing "best response" to the choices of the other station, and this must hold simultaneously for both stations. To compute the Nash equilibrium, we first derive each gas station's profit under each case. The working assumption is that of constant marginal cost c_A and c_B for E10 and E85, respectively, where the unit cost c_B for E85 embeds the subsidy s. We further assume that fuel costs are the same for all stations.

For the benchmark model without E85 fuel (no E85 stations at either location), given the demand functions in equation (1) in section 2.1, the profits of the two stations are:

$$\pi^0 = \frac{1}{2t}\left(p_A^0 - c_A\right)\left[p_A^1 - p_A^0 + t\right]$$
$$\pi^1 = \frac{1}{2t}\left(p_A^1 - c_A\right)\left[p_A^0 - p_A^1 + t\right]$$

This is the textbook parameterization of the basic Hotelling's model, yielding the Nash equilibrium solution (see, e.g., Tirole 1988):

$$p_A^0 = p_A^1 = t + c_A \quad , \quad d_A^0 = d_A^1 = \frac{1}{2} \quad , \quad \pi^0 = \pi^1 = \frac{t}{2}$$

The equilibrium values of these and other variables of interest are reported in Table 1 in section 5.1.

For the model with only one E85 station, the profits of the two gas stations are:

$$\pi^0 = \left(p_A^0 - c_A\right)d_A^0 + \left(p_B^0 - c_B\right)d_B^0$$
$$\pi^1 = \left(p_A^1 - c_A\right)d_A^1$$

where the demand functions are derived in section 3.2.

For the model with both stations offering both fuels, the profits of gas stations at locations L0 and L1 are

$$\pi^0 = \left(p_A^0 - c_A\right)d_A^0 + \left(p_B^0 - c_B\right)d_B^0$$
$$\pi^1 = \left(p_A^1 - c_A\right)d_A^1 + \left(p_B^1 - c_B\right)d_B^1$$

where the demand functions are as derived in section 3.3. In the duopoly setting, the gas station at L0 chooses p_A^0 and p_B^0, and its competitor chooses p_A^1 and p_B^1. In this symmetric market, we are looking for the symmetric equilibrium where $p_A^0 = p_A^1$ and $p_B^0 = p_B^1$, so $\tilde{x}_A = \tilde{x}_B = 1/2$. A complete list for the equilibrium values of several variables of interest is reported in Table 1 below.

Analytic solutions for the Nash equilibrium are not possible in models with E85, even the symmetric one. To proceed, we have computed the Nash equilibrium numerically, as follows.

From the payoff functions defined in the foregoing we derive analytic first order conditions (FOCs) that define the best response functions for each station. In the model with one E85 station, we have two best response functions for the gas station at L0 and one for the gas station at L1. In the model with two E85 stations, we have two best response functions for each gas station. The best response functions are solved simultaneously in Matlab using *vpasolve*. Multiple systems of solutions from the FOCs are possible, hence we relied on local second order conditions (SOCs) for a maximum to narrow the possible candidates. Eventually, only one system of solutions survives the SOCs, which is the Nash equilibrium outcome. To perform this numerical process, of course, we first need the values of all parameters, which is what we do in the next section.

4. Parameter Calibration

Two of the models discussed in the foregoing are not amenable to an analytic solution of the Nash equilibrium. To solve this model numerically, the first step is to calibrate the parameters of the model. The parameter λ captures the energy efficiency of E10 compared to E85, which, as discussed earlier, is a known constant $\lambda = 1.25$. In addition, the model has eight other parameters:

a) s, the per-unit subsidy of E85;

b) c_A, the marginal cost of E10;

c) c_B, the marginal cost of E85;

d) α, the fraction of FFVs;

e) u, consumers' reservation utility from driving one gallon of E10;

f) t, Hotelling's "travel cost" parameter;

g) $\bar{\theta}$, the upper bound of drivers' preference parameter for E85;

h) $\underline{\theta}$, the lower bound of drivers' preference parameter for E85.

Prices (p_A^0, p_A^1, p_B^0, and p_B^1), and demands (d_A^0, d_A^1, d_B^0, and d_B^1), are all endogenous variables. To calibrate these parameters, we use relevant features of the model along with market data pertaining to the year 2017.

The subsidy s captures the policy-induced subsidy for E85, relative to E10. Following the discussion in section 2.2, the subsidy is constructed by equation **(8)**. In equation **(8)**, $R = 0.72$ is the price of D6 RINs and $B = 0.0836$ is the cost of compliance for a gallon of obligated conventional gasoline. [9] So, $s = 0.5143$.

c_A and c_B correspond to the terms c_{E10} and c_{E85} of section 2.2. We note here that the producer price p_g is not observed. What we observe is the RIN-laden average gasoline wholesale price, denoted as $\tilde{p}_g$, which in 2017 was $1.689/gal.[10] In the postulated competitive refining/blending industry that operates under constant returns to scale, we should have $\tilde{p}_g = p_g + B$, where the bundle of obligations term B was introduced in section 2.2. The average ethanol wholesale price, denoted p_e in section 2.2 as, was $1.45/gal in 2017.[11] The average gasoline motor fuel tax, denoted as μ in section 2.2, is $0.449/gal.[12] Together with the values

9 The prices of D4, D5, and D6 RINs (RIN year 2017 and transfer year 2017) are $p_{D4} = 1.03$, $p_{D5} = 0.91$, and $p_{D6} = 0.72$, respectively, from EPA, https://www.epa.gov/fuels-registration-reporting-and-compliance-help/rin-trades-and-price-information. Hence, from section 2, $B = 0.0167 p_{D4} + 0.0071 p_{D5} + 0.0832 p_{D6} = 0.0836$.

10 The gasoline wholesale price is from EIA, https://www.eia.gov/dnav/pet/pet_pri_refoth_dcu_nus_a.htm. It is the "Motor Gasoline" price under "Sales for Resale" category.

11 The ethanol wholesale price is the ethanol rack price in Omaha, Nebraska.

12 The average gasoline motor fuel tax is from Moschini, Lapan, and Kim (2017). We do not consider the per-unit marketing/retailing costs. First, the marketing/retailing costs are too small to have a significant effect on the model results. Moreover, we will show next that any cost would be absorbed by the value of t in our calibration procedure.

of R and B, we calibrate the cost of E10 fuel to be $c_A = 2.0421$ and the cost of E85 fuel to be $c_B = 1.4283$ (recall that fuels are measured in volume terms, and they possess different energy content).

For the FFV fraction α, in 2017 there were 20.34 million FFVs and 215.09 million gasoline vehicles (cars and light trucks categories).[13] Hence, we estimate $\alpha = 0.0864$. As it has been shown in the indifference consumers and equations (14)-(21) in section 3, the reservation utility u has no effect on driver's choice of fuel so we do not specify an exact value for u. In the duopoly models, we assume u is large enough so that all drivers would choose to refuel. For the parameter t, in the basic Hotelling's model this parameter decides the E10 equilibrium price margin: as discussed in section 3.4, the equilibrium E10 price in the basic Hotelling model is equal to $t + c$, hence the retail price margin is t. The retail price of E10 in 2017 was $2.3625/gal.[14] Given the cost of E85 discussed earlier, the margin for E10 is $0.3204, so we set $t = 0.32$.

Concerning the bounds $\left(\underline{\theta}, \bar{\theta}\right)$ of the distribution of consumers' preferences for the high-ethanol attribute of E85, we have assumed $\underline{\theta} < 0$ and $\bar{\theta} > 0$. That is, FFV drivers with high preferences for E85 are willing to pay a premium, whereas FFV drivers with low preferences for E85 would only purchase it under some price discount. To calibrate these bound parameters, we rely on WTP estimates, as well as specific features of our model. Pouliot, Liao, and Babcock (2018) estimate the WTP for E85 in the United States using survey data. The survey targeted at FFV drivers and their estimates of WTP show that about 25% of motorists would prefer E85

13 EIA Annual Energy Outlook, 2017.

14 Quarterly nationwide average retail prices of E10 are from the Clean Cities Alternative Fuel Price Report. The annual average price is just the average of each quarter.

when E85 and E10 are equally priced at energy-equivalent unit. The result would suggest that $\bar{\theta}/(\bar{\theta}-\underline{\theta})=0.25$, implying $\underline{\theta}=-3\bar{\theta}$. Next, recall that we would like our highly stylized model to capture a realistic scenario with the baseline parameters. Specifically, we would like to ensure that, in the baseline, $\bar{x}^1<1$ and $\tilde{\theta}^0>\underline{\theta}$ (recall Figure 2). Numerical exploration of the model indicates that the value of $\underline{\theta}$ has little effect on $\bar{x}^1$ and $\tilde{\theta}^0$, but $\bar{x}^1$ and $\tilde{\theta}^0$ vary significantly with $\bar{\theta}$. To ensure that $\bar{x}^1<1$, $\bar{\theta}$ needs to be lower than 0.27. Hence, we pick $\bar{\theta}=0.25$, which satisfies this constraint and still allows for "green" drivers to have a nontrivial WTP for E85 (this value implies that the highest WTP for FFV drivers is \$0.25/gal under our model normalization; to fuel a car with capacity of 16 gal, this is equivalent to \$4/tank). Given the value of $\underline{\theta}$, from the relation $\underline{\theta}=-3\bar{\theta}$ derived earlier, we have $\underline{\theta}=-0.75$ (which also implies that the desired condition $\tilde{\theta}^0>\underline{\theta}$ is satisfied). In any event, we conduct the sensitivity analysis of model results under different $\bar{\theta}$ and $\underline{\theta}$ (Appendix D). In sum, the baseline parameter values used for our benchmark analysis are: $s=0.5143$, $c_A=2.0421$, $c_B=1.4283$, $\alpha=0.0864$, $t=0.32$, $\bar{\theta}=0.25$, and $\underline{\theta}=-0.75$. In the next section, we show the results of different models under baseline values of all parameters. We also evaluate how results are affected by changes in the subsidy of E85.

5. Results

We first present the Nash equilibrium results for duopoly models under the alternative conditions of no E85 fuel, incomplete penetration of E85 stations (only the station at L0 sells E85), and complete penetration of E85 stations (all stations sell both fuels). Next, in the comparative statics section, we evaluate how results are affected by changes in the subsidy level of E85.

5.1 Baseline results

Results for the three duopoly models are reported in Table 1. For the models with no E85, we report both the analytic solutions and the values of these solutions at the baseline parameters. This permits a straightforward comparison with the main model—where E85 is only sold in one station—for which we can only find numerical solutions. For the model with E85 at both stations, we report the numerical solutions where the analytic solutions are not applicable. For each case, we report equilibrium prices and quantities, profits for the two stations. We report the pass-through rate, defined as $\partial(p_A - p_B)/\partial s$ in section 2.2, which measures the rate at which the subsidy brought about by RINs is passed on to the retail spread between E85 and E10 prices. Because both stations provide E10, we report $\partial(p_A - p_B)/\partial s$ as the average of $\partial\left(p_A^0 - p_B^0\right)/\partial s$ and $\partial\left(p_A^1 - p_B^0\right)/\partial s$. The effects of the subsidy s on individual prices are discussed in section 5.2. The process of computing all pass-through rates is detailed in Appendix B.

From Table 1, we see the effects of competition as the possibility of E85 substituting for E10 is introduced. Adding an E85 pump at L0 causes E10 price at the same location to increase slightly while E10 price at the other location decreases. Demand for E10 at both locations decreases (recall that, with the assumed covered market condition with a given mass of consumers, the availability of E85 substitutes for some E10 at both locations). In the model with only one E85 location, E85 consumption is 0.0292 at baseline parameter values, which means that about 27% ($= d_B^0/\alpha\lambda$) of FFV drivers choose E85. Adding another E85 location increases E85 consumption by 61.6%. Having an E85 pump increases the L0 gas station's profit by 2.69% but decreases the other gas station's profits by 1.31%.

The pass-through rate of the E85 subsidy due to RIN prices, at the baseline parameters, is 0.7188 (i.e., approximately 70%). Hence, competition with incomplete penetration of E85 stations is characterized by incomplete pass-through. This result formalizes the role of retail market power in affecting pass-through of RIN prices to retail prices. The gas station at L0 has a product that the other station does not have, it is effectively a monopoly for E85. The exercises of this market power, by raising E85 prices, is constrained by the fact that E10 is also sold by the same station (in addition to being sold by a competitor station). Still, at the equilibrium solution the markup, over cost, of the E85 price is higher than for E10.

It is interesting to note that the market power that is relevant for the pass-through effect just discussed differs from the source of market power that arises from the structure of the basic Hotelling model. In the model with E85 available at both locations, the pass-through rate is still incomplete but close to one (0.9665), which implies that it is the market power from exclusivity of selling E85 that mostly determines the incompleteness of the pass-through of the E85 subsidy, rather than the market power stemming from horizontal differentiation. The incompleteness of pass-through rate is very much related to the difference in the markups of E10 and E85. In the model with pure E10, the markup over cost is $t = 0.32$; in the model with pure E85, the markup over cost is $\lambda t = 0.4$—that is, it is more profitable to sell E85 than E10 to an FFV driver. In the model with both fuels at both locations, this effect provides an incentive for stations to decrease the price of E85 while increasing the price of E10. The equilibrium markup of each fuel depends on the relative demand elasticities. The results show that, in equilibrium, the markup of E85 is 0.3713 ($= \lambda(p_B - c_B)$) while the markup of E10 is 0.3208 ($= p_A - c_A$).

Essentially, the availability of E85, together with consumer heterogeneity, brings about differentiation along a vertical attribute. This is valuable to a firm only insofar as it has some exclusivity. When all stations sell E85, alongside E10, this exclusivity vanishes and what remains is the horizontal differentiation of consumers, which is what endows firms with some limited market power in the Hotelling model. With full penetration of E85 stations, the value of the E85 subsidy is mostly captured by consumers.

Despite the fact that full penetration of E85 stations brings limited additional profit to the fuel-retailing firms, this situation should not be interpreted as a lack of incentives for retail stations to adopt E85 pumps. It is quite clear that, with incomplete penetration, the station that sells E85 enjoys higher returns than in the case when no station carries E85 (0.1644 > 0.16), and the firm who does not sells E85 in the case of incomplete penetration can increase its profit by also adopting an E85 pump (0.1614 > 0.1579). Although this model is not quite suited to investigate the conditions for optimal entry of E85 stations, the structure of the model is such that the "excess entry" result discussed by Mankiw and Whinston (1986) is expected to apply.

5.2 Comparative statics

Of all parameters, the subsidy is of the most interest. Assessing the effect subsidy provides us key implications on evaluating the effectiveness of the RIN system. For the model with E85, for which we only have numerical solutions, in this section we present some numerical comparative statics results. Here, we focus specifically on understanding how varying the subsidy may affect the pass-through rate in equilibrium outcomes for the duopoly model with one E85 station (incomplete penetration of E85). Corresponding results for the duopoly model with two E85 stations (full penetration of E85) are reported in the Appendix C. Comparative statics results for parameters other than the subsidy level are reported in Appendix D.

To evaluate the equilibrium results under alternative values of the subsidy, all other parameters (except the cost of E85, which is directly affected by the subsidy as in equation (7)) are held at their baseline values. The results, reported in tabular form in the Appendix, are summarized in Figure 5. The vertical black dashed line in Figure 5 represents the baseline subsidy value of s=0.5143. The dotted points along the separate lines mean that pass-through rates in the model are not a continuous function of the subsidy. The discontinuity is a result of the fact that alternative demand system configurations can be attained under different values of the subsidy. These alternative "cases," illustrated earlier in Figure 2 and Figure 3, are specifically labeled in Figure 5. In Appendix A, we provide the actual values of the subsidy at the kink points in Figure 5.

From Figure 5, we see that when the subsidy s is small ($s < 0.1089$, as shown in Appendix A), no E85 is sold in the market, and the pass-through rate (to the implicit choke-off prices) is equal to 1. In this case, even if the gas station passes all the subsidy to FFV drivers, the latter would still choose to refuel with E10 as the price of E85 is not low enough to compensate for its low energy content (even for the FFV driver with highest preference for E85). At $s = 0.1089$, the gas station at L0 is just indifferent between selling E10 or E85, and the FFV driver with highest preference for E85 is just indifferent between choosing E10 or E85 at L0. Then the pass-through rates jump to $2/3$ (for $0.1809 < s < 0.1569$), which is exactly the pass-through of subsidy/tax on an obligated product in equation (3). This corresponds to case 1 in Figure 3, for which $\tilde{\theta} > \bar{\theta}$. In this case, the equilibrium results show that offering E85 in the market has no effect on the equilibrium prices of E10—E10 prices at both locations are $t + c$. So, when the subsidy is low such that E85 does not directly compete with E10 at another location, the introduction of E85 in the market will not affect the equilibrium price of E10. For higher values of the subsidy, specifically over the domain $0.1569 < s < 0.5265$, the pass-through rate decreases from 0.9402 to 0.7179 as the subsidy increases. For still higher values of the subsidy, over the domain $0.5265 < s < 1.4735$, the pass-through rate stays around 0.5, and then decreases toward zero as $s > 1.4735$. Note that case 5 of Figure 3 is not depicted in Figure 5 because case 5 only materializes when $s > 2.1343$ (we observe a small jump from case 4 to case 5, and the pass-through rate then monotonically decreases to zero in case 5).

Comparative statics for the pass-through rate in the duopoly model with two E85 locations are presented in Appendix C. Similar to the duopoly model with one E85 station, we find that the pass-through rate in the model with two E85 stations is not a continuous function of the subsidy. Furthermore, the complete penetration of E85 stations has notable effects on the pass-through rate. Instead of generally decreasing with s as in the one E85 station situation (recall Figure 5), with two E85 stations the pass-through rate increases toward one.

In Figure 6, we provide some additional details by reporting the equilibrium prices, shares of FFV drivers who choose each fuel, and pass-through to individual fuels at values of the subsidy from 0 to 1.8. The orange line represents the equilibrium results of E10 at L0; the red line represents the equilibrium results of E85 at L0; and, the blue line represents the equilibrium results of E10 at L1. The vertical dashed line standing near 0.5 indicates the baseline value of s.

The top panel of Figure 6 depicts the equilibrium prices of each fuel. It shows that under the assumption of constant E10 cost, the retail prices of E10 barely change. As shown by the equilibrium results at representative values of the subsidy reported in Table C1 in Appendix C, E10 price at L0 is higher than that at L1. The decrease in E85 prices as s increases is significant, which is in line with the pass-through rate in the duopoly model with one E85 station. The middle panel of Figure 6 reports the share of FFV drivers who choose each fuel rather than the demands of each fuel. With larger subsidy, the share of FFV drivers who choose to refuel with E85 goes from 0 to 0.9, and the share of FFV drivers who choose to refuel with E10 at L0 decreases from 0.5 to 0. The share of FFV drivers who choose to refuel with E10 at L1 also decreases but slower. The bottom panel of Figure 6 reports the pass-through of the subsidy to each retail price, defined as $\partial p_A^0/\partial s$, $\partial p_B^0/\partial s$ and $\partial p_A^1/\partial s$, respectively. Note that the pass-through rate reported in Table 1, $\partial(p_A - p_B)/\partial s$, is equivalent to $0.5\left(\partial p_A^0/\partial s + \partial p_A^1/\partial s\right) - \partial p_B^0/\partial s$. The equilibrium results of the duopoly model with two E85 locations are relegated to Appendix C. Numerical equilibrium results at some representative values of s in both models are also reported in Appendix C.

The advantage of a stylized model, such as ours, is that we can evaluate the partial effect of each parameter on the model result. In additional to the subsidy, we run the model with alternative values of some key parameters—the fraction of FFVs (α), Hotelling's "travel cost" parameter (t), high type preference ($\bar{\theta}$), and low type preference ($\underline{\theta}$).[15] We analyze the impact of each parameter one at a time, holding all other parameters at their baseline values. We find that equilibrium prices and the pass-through rate barely change with α and $\underline{\theta}$. The effects of these two parameters on E85 demand are proportional: larger α relates to more FFV drivers and larger $\underline{\theta}$ in absolute value corresponds to smaller proportion of high type FFV drivers. As t increases, all equilibrium prices increase as expected (recall that t is reflected in the price margin in the Hotelling's model), along with the pass-through rate (the increase is moderate as shown in Table D4 in Appendix D). The parameter $\bar{\theta}$ has little effect on equilibrium E10 prices, whereas an increase of this preference parameter results in higher E85 price and lower pass-through rate.

6. More on Market Power: Monopoly

In Hotelling's framework, firms have some relief from the predicament of price competition. Firms enjoy some local market power because of the spatial heterogeneity of

[15] Tables of results with alternative values of these parameters are reported in Appendix D, specifically Table D1-Table D4.

consumers, vis-à-vis the location of the retailing firms. The intensity of this effect is captured by the parameter t . The possibility of selling E85 provides an additional venue for a station to extract rent from consumers' vertical differentiation, provided the station has some exclusivity in its access to E85. In our modeling framework, such a situation is captured by E85 being available at only one of the two stations. With full penetration of E85, the pass-through rate is almost complete and the profits from selling an additional fuel are quite limited.

There are reasons to believe that the characterization of market power provided by Hotelling's model may be insufficient in our setting. Firms may be able to enjoy more market power if they collude. Indeed, the possibility of tacit collusion is particularly real in settings, such as fuel retailing, where firms interact repeatedly (Tirole 1988). Furthermore, in reality, neighboring stations/brands may be owned by the same firm. As shown by Hastings (2004), the loss of independent, unbranded competitor would increase local fuel price. In such cases, market outcomes close to monopolistic may be quite plausible.

To investigate the effects that collusive behavior may have on the market outcomes of interest, in this section we solve the monopoly problem that would arise if the two stations in our model perfectly coordinated their choices (for both E10 and E85) with the objective of maximizing joint profit. The demand structures in the monopoly models with one or two E85 stations are the same as those under their duopoly counterparts. Thus, there are five cases in the monopoly with one E85 station, and three cases in the monopoly with two E85 stations. However, the values of subsidy at the kink points are different (see Appendix A). At the baseline values of all parameters, the demand structure that applies to monopoly with one E85 station is the same as in Figure 2, and in the model with two E85 stations is the same as Figure 4.

In the case of monopoly, we can find analytic solutions for the settings where there is no E85 station, and where both stations sell both fuels. When only one of the two stations sells E85, however, we again need to resort to a numerical solution. A monopoly, given the characterization of consumers' preferences used in the model, would want to charge the highest possible price, conditional on consumers' participation that ensures a covered market. This condition, in term of prices, requires $p_A^0 + p_A^1 \leq 2u - t$ (it can be verified that a covered market is indeed a profit-maximizing feature of the parameter space we investigate). Profit maximization solutions for the monopoly case when there is no E85 are $p_A^0 = p_A^1 = u - t/2$ (at these prices the consumer most distant to a station, located at $x = 0.5$, is just willing to refuel). Unlike the duopoly model, we need the value of reservation utility u to get the numerical solutions in the monopoly models. To make the result comparable with the duopoly model, we choose the value of u such that in the case with no E85, equilibrium price of E10 is same with that of the duopoly model. So, when there is no E85, $p_A^0 = p_A^1 = 2.3621$, implying $u = 2.5221$. Equilibrium values of other variables of interest are reported in Table 2.

Because of our model setup, the equilibrium price level is largely determined by the parameter u (which has no effect on the competitive duopoly equilibrium analyzed in the previous section). Adding an E85 pump only at location L0 decreases the E10 price at the same location and increases the E10 price at the other location, which contrasts with the results we found for duopoly. The price effects of introducing E85, however, are minimal, as the monopoly charges the maximum price consistent with retaining a covered market. The monopoly's total profits, reported in Table 2, show that adding E85 pumps to one or both stations increases profits. The additional profit afforded by E85, however, is minimal, a reflection of the small size of the E85 market at the baseline. E85 consumption slightly increases with implementation of

E85 at another location. Comparing of profit outcomes under duopoly and monopoly also provides some insights concerning the incentive for adoption of E85 pumps. At baseline parameter values, under duopoly we find that the station at L1 can increase its profit by 0.0035 by also adding an E85 pump. Under monopoly, the comparable additional profit is 0.0017. Hence, full penetration of E85 stations is less likely when collusive behavior at the retailing level applies.

Of more direct interest to us is the pass-through of the E85 subsidy, which turns out to be incomplete. When only one station sells E85, the pass-through rate is 60.51%, clearly lower than what is attained under duopoly. Perhaps most interestingly, full penetration of E85 stations lowers, rather than increasing, the equilibrium pass-through rate. Table 2 shows that the pass-through rate of the E85 subsidy is just 50% when both stations sell both fuels, regardless of other parameters, as long as the baseline demand configuration applies.[16] To get a full idea of how pass-through rate evolves with the subsidy under incomplete penetration, we provide the following Figure 7.

In Figure 7, we depict the equilibrium pass-through rate for the subsidy ranging from $s = 0$ to $s = 1.8$ (taking c_A as given). The dotted lines connecting the pass-through rates show that they are not continuous functions of the subsidy s. The vertical black dashed line indicates the baseline value s=0.5143. Figure 7 shows that when $0.1089 < s < 0.1569$ the pass-through rate is 2/3 as in equation (3). Indeed, for this domain the pass-through rate is 2/3 in all models that we have considered—the duopoly model with one E85 station in Figure 5, the duopoly model with two E85 stations in Figure in the Appendix, and the monopoly model with two E85 stations in

[16] In the model with two E85 stations, from the analytic solution for p_B in Table 2, the pass-through of subsidy is ½ to the E85 price and zero to the E10 price, hence the pass-through rate is ½.

Figure in the Appendix. This means that, for these relatively low subsidy levels, market structure (duopoly or monopoly) and the number of E85 stations (one or two) have no effect on the pass-through rate nor on E10 prices, as there is no direct competition between E85 and E10 at different locations. Over the domain $0.1569 < s < 0.6518$, the pass-through rate increases with the subsidy level (from around 0.5 to a little higher than 0.6). For $s > 0.6518$, the pass-through rate stays around 0.5, jumping down at $s = 1.6141$ and decreasing toward zero for large subsidy levels. In the monopoly model with two E85 stations, details for which are reported in the Appendix C, the pass-through rate exhibits a similar behavior as with the one E85 station case of Figure 7.

The foregoing results, together with Figure 5, establish that when market power arises from exclusivity of selling E85 (duopoly model with one E85 station) or general collusive/monopoly power, then the pass-through rate decreases toward zero as the subsidy level increases. Conversely, if market power arises just horizontal differentiation (duopoly model with two E85 stations), then the pass-through rate increases toward one as the subsidy level increases.

7. Conclusion

The RFS implemented by the United States over the last decade represents an ambitious policy aimed at promoting the substitution of fossil fuel with renewable fuel. To fulfill the mandates envisioned by the RFS, it is becoming necessary for the market to absorb an increasing amount of biofuel as high-ethanol blends, such as E85. The mechanism that should bring this about is rooted in RIN prices, which simultaneously constitute an implicit subsidy for biofuels and a tax on fossil fuels. The effectiveness of this mechanism, however, depends critically on the E85 subsidy, due to RIN prices, to pass through to consumers. Indeed, as noted by previous research, owners of FFVs "… will have little incentive to use E85 unless it is priced significantly lower than gasoline" (Collantes 2010). The findings of an emerging empirical literature, discussed earlier, suggest obstacles to the pass-through of RIN prices to retail E85 prices. It seems that the pass-through of the E85 subsidy, mediated by RIN prices, may be incomplete and the pass-through rate relates to the possible existence of market power.

In this paper, we build a structural model of how market power may arise in E85 retailing, and use this model to gain insights into the nature of imperfect competition in this market, and the role of the E85 subsidy in determining the market outcomes of interest. Our model is rooted in Hotelling's spatial competition framework, which provides a natural representation of gas stations' market power due to location differentiation. This basic model is extended to account for important features of the market for E85, specifically the imperfect substitution between E85 and E10 (which itself depends on consumer heterogeneity), and the limited availability of E85 stations. We specifically evaluate three duopoly models and, to gain further insights into the role of market power, three monopoly models. Analytic solutions for the Nash equilibrium are possible only for the basic Hotelling's models (and the extended model with full penetration of E85 stations under monopoly). For all other models we resort to numerical solutions (upon calibration of key models parameters, consistent with real-world data).

Results from the model suggest that pass-through of the E85 subsidy to retail prices is indeed generally incomplete. In our baseline model, which maintains the incomplete penetration of E85 refueling stations, the equilibrium pass-through rate is about 70%. With full penetration of E85 pumps (i.e., all stations offer both E10 and E85), the pass-through of the E85 subsidy to retail prices is near complete (even though gas stations retain some market power from their location differentiation). In the collusive outcome whereby gas stations act as a monopoly (as may arise from tacit collusion due to repeated interaction), the pass-through rate is significantly lower; furthermore, in this case, full penetration of E85 pumps decreases the equilibrium pass-through rate (rather than increasing it, as in the duopoly model). Noticeably, when E85 only substitutes for E10 demand at the same location but not E10 demand at the other location, the pass-through rate is 2/3 regardless of whether it is monopoly or duopoly, partial or full penetration of E85. When the subsidy is large enough (i.e., greater than some threshold levels, which take different values in different models), the pass-through rate goes to one in the duopoly model with two E85 stations, whereas in the other three models (duopoly with one E85 station, monopoly with one E85 station, and monopoly with two E85 stations) the pass-through rate goes to zero. The result highlights the different implications for market power that arise from horizontal differentiation as opposed to from stations' exclusivity (or monopoly power) in selling E85 fuel.

The model we build enables us to examine the effect of the subsidy on equilibrium fuel prices and demands. We show, as expected, E85 consumption increases with the subsidy. When the subsidy increases from 0.1 (when $s < 0.1$, there is no E85 consumption in the market) to 1, the percentage of FFV drivers who refuel with E85 goes up from 0% to 58%. However, as the FFV fleet size is quite small ($\alpha = 0.0864$), even at the subsidy level of \$1.00, only 5% of all drivers would choose to refuel with E85. We show that the introduction of E85 has little effect on E10 prices and the effect is different in duopoly and in monopoly. In duopoly, price of E10 at the same location with E85 is slightly higher than that at the other location, whereas in monopoly it reverses. This result may serve as an indicator of monopoly power. Both in duopoly and monopoly, introduction of one E85 station reduces E10 demand at the same location more than E10 demand at the other station.

References

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

Brécard, D. (2014). Consumer confusion over the profusion of eco-labels: Lessons from a double differentiation model. **Resource and Energy Economics**, 37, 64-84.

Collantes, G. (2010). Do green tech policies need to pass the consumer test?: The case of ethanol fuel. **Energy Economics**, 32(6), 1235-1244.

Di Comite, F., Thisse, J.-F., & Vandenbussche, H. (2014). Verti-zontal differentiation in export markets. **Journal of International Economics**, 93(1), 50-66.

Ferreira, R. D. S., & Thisse, J. F. (1996). Horizontal and vertical differentiation: The Launhardt model. **International Journal of Industrial Organization**, 14(4), 485-506.

Gabszewicz, J. J., & Wauthy, X. Y. (2012). Nesting horizontal and vertical differentiation. **Regional Science and Urban Economics**, 42(6), 998-1002.

Hastings, J. S. (2004). Vertical relationships and competition in retail gasoline markets: Empirical evidence from contract changes in Southern California. **American Economic Review**, 94(1), 317-328.

Hotelling, H. (1929). Stability in competition. **The Economic Journal**, 39, 41-57.
Knittel, C. R., Meiselman, B. S., & Stock, J. H. (2017). The pass-through of RIN prices to wholesale and retail fuels under the renewable fuel standard. **Journal of the Association of Environmental and Resource Economists**, 4(4), 1081-1119.

Korting, C., & Just, D. R. (2017). Demystifying RINs: A partial equilibrium model of US biofuel markets. **Energy Economics**, 64, 353-362.

Lade, G. E., & Bushnell, J. (2019). Fuel Subsidy Pass-Through and Market Structure: Evidence from the Renewable Fuel Standard. **Journal of the Association of Environmental and Resource Economists**, 6(3), 563-592

Lapan, H., & Moschini, G. (2012). Second-best biofuel policies and the welfare effects of quantity mandates and subsidies. **Journal of Environmental Economics and Management**, 63(2), 224-241.

Li, J., & Stock, J. H. (2019). Cost pass-through to higher ethanol blends at the pump: Evidence from Minnesota gas station data. **Journal of Environmental Economics and Management,** 93 (1), 1–19.

Mankiw, N. G., & Whinston, M. D. (1986). Free entry and social inefficiency. **The RAND Journal of Economics**, 17, 48-58.

Moschini, G., Lapan, H., & Kim, H. (2017). The Renewable Fuel Standard in Competitive Equilibrium: Market and Welfare Effects. **American Journal of Agricultural Economics**, 99(5), 1117-1142.

Neven, D., & Thisse, J.-F. (1990). On quality and variety competition. In: Gabszewicz, J.J., Richard, J.-F., Wolsey, L. (Eds.), **Economic Decision Making: Games, Econometrics and Optimization**. Contributions in Honour of J. Drèze, Amsterdam, North-Holland, pp. 175–199.

Norman, G., Pepall, L., Richards, D., & Tan, L. (2016). Competition and consumer data: The good, the bad, and the ugly. **Research in Economics**, 70(4), 752-765.

Pennerstorfer, D. (2017). Can competition keep the restrooms clean? Price, quality and spatial competition. **Regional Science and Urban Economics**, 64, 117-136.

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

Pouliot, S., & Babcock, B. A. (2017). Feasibility of meeting increased biofuel mandates with E85. **Energy Policy**, 101, 194-200.

Pouliot, S., Liao, K. A. & Babcock, B. A. (2018). Estimating willingness to pay for E85 in the United States using an intercept survey of flex motorists. **American Journal of Agricultural Economics,** 100(5), 1486-1509.

Schnepf, R., & Yacobucci, B. D. (2013). Renewable Fuel Standard (RFS): overview and issues. **Congressional Research Service**, Report R40155.

Stock, J. H. (2015). The Renewable Fuel Standard: A Path Forward. Center on Global Energy Policy, Columbia University.

Tirole, J. (1988). **The theory of industrial organization**. Cambridge, MA: MIT Press.

Weyl, E. G., & Fabinger, M. (2013). Pass-through as an economic tool: Principles of incidence under imperfect competition. **Journal of Political Economy**, 121(3), 528-583.

Figure

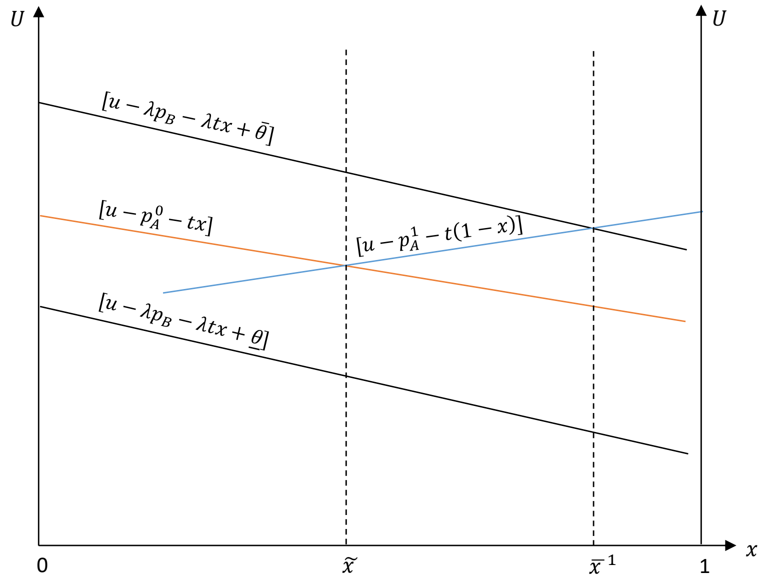


Figure 1. FFV drivers' choice of fuel at given prices p_A^0 , p_A^1, and p_B^0

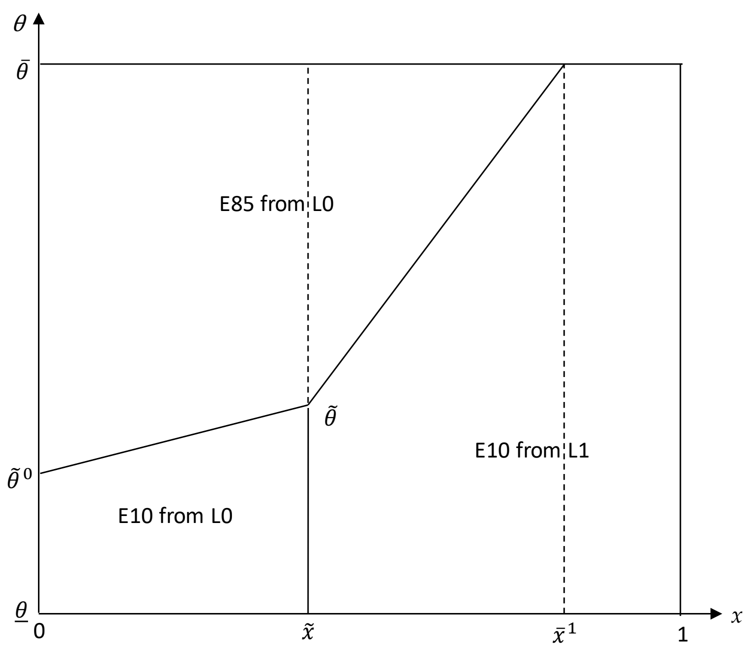


Figure 2. FFV drivers' demands in the one E85 station model, baseline scenario ("case 2")

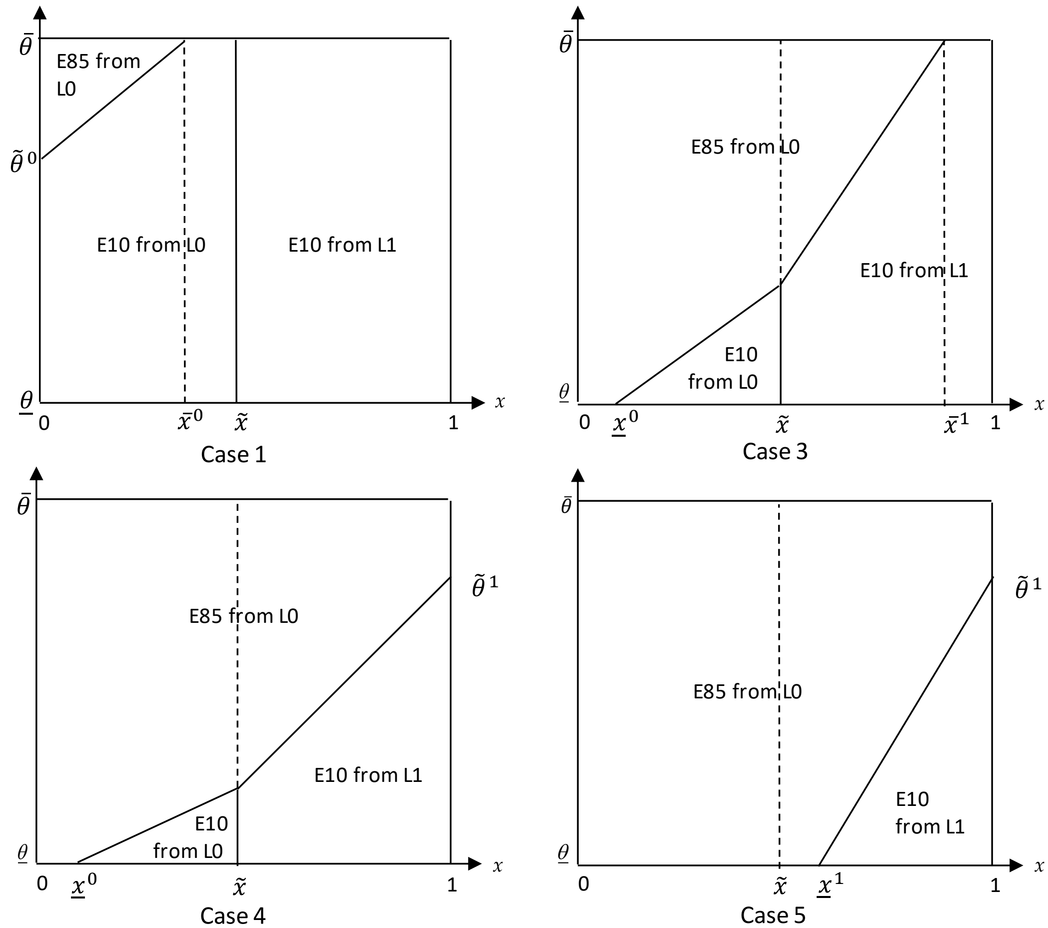


Figure 3. FFV drivers' demands in the one E85 station model, other scenarios

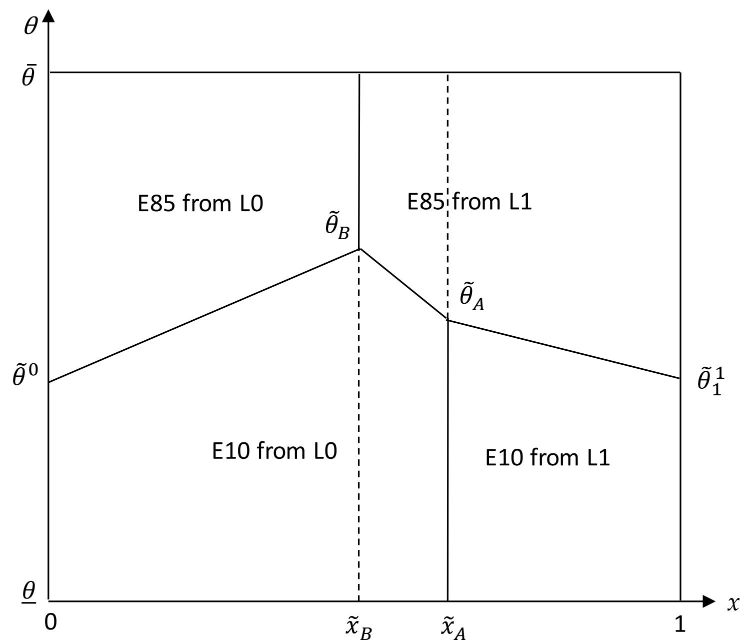


Figure 4. FFV drivers' demands in the two E85 stations model, baseline scenario

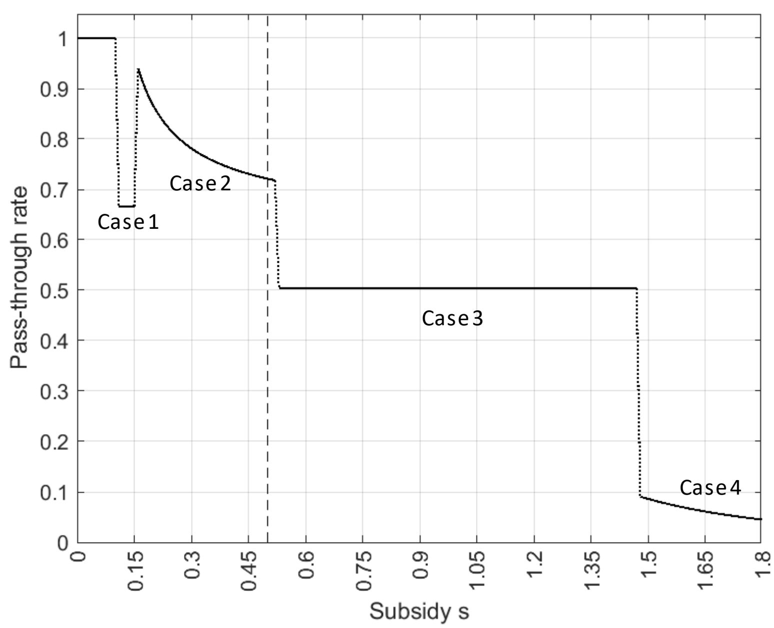


Figure 5. Pass-through rate and the E85 subsidy in the duopoly model with one E85 station

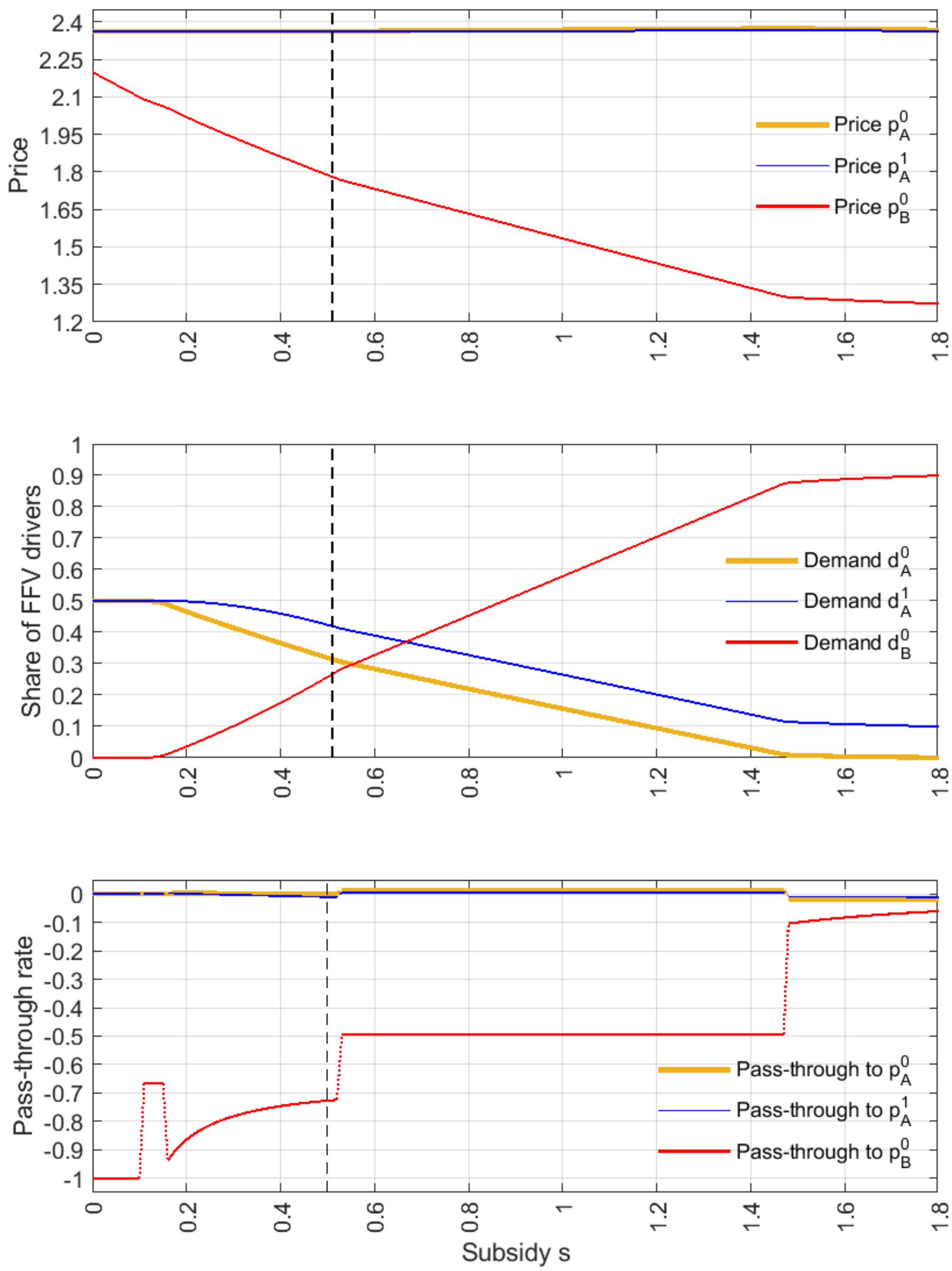


Figure 6. Simulated equilibrium results in the duopoly model with one E85 station

Table

Table 1. Nash Equilibrium Outcomes for Duopoly Models

	No E85 stations		One E85 station	Two E85 stations
	Analytic solution	Baseline values	Numerical solution	Numerical solution
p_A^0	$c_A + t$	2.3621	2.3627	2.3630
p_A^1	$c_A + t$	2.3621	2.3606	2.3630
p_B^0	-	-	1.7772	1.7251
p_B^1	-	-	-	1.7251
$\frac{\partial(p_A - p_B)}{\partial s}$	-	-	0.7188	0.9665
d_A^0	$\frac{1}{2}$	0.5000	0.4807	0.4811
d_A^1	$\frac{1}{2}$	0.5000	0.4959	0.4811
d_B^0	-	-	0.0292	0.0236
d_B^1	-	-	-	0.0236
π_0	$\frac{t}{2}$	0.1600	0.1643	0.1614
π_1	$\frac{t}{2}$	0.1600	0.1579	0.1614

Table 2. Nash equilibrium outcomes for Monopoly model

	No E85 stations		One E85 station	Two E85 stations	
	Analytic solution	Baseline values	Numerical solution	Analytic solution	Baseline values
p_A^0	$u-\frac{t}{2}$	2.3621	2.3617	$u-\frac{t}{2}$	2.3621
p_A^1	$u-\frac{t}{2}$	2.3621	2.3625	$u-\frac{t}{2}$	2.3621
p_B^0	-	-	1.8546	$\frac{2u-t+\bar{\theta}+\lambda c_B-c_A}{2\lambda}-\frac{t(\lambda-1)}{8\lambda}$	1.8790
p_B^1	-	-	-	$\frac{2u-t+\bar{\theta}+\lambda c_B-c_A}{2\lambda}-\frac{t(\lambda-1)}{8\lambda}$	1.8790
$\frac{\partial(p_A-p_B)}{\partial s}$	-	-	0.6051	$\frac{1}{2}$	0.5000
d_A^0	0.5	0.5000	0.4895	$\frac{1}{2}-\frac{\alpha}{4(\bar{\theta}-\underline{\theta})}\left[\bar{\theta}+c_A-\lambda c_B-\frac{t(\lambda-1)}{4}\right]$	0.4895
d_A^1	0.5	0.5000	0.4948	$\frac{1}{2}-\frac{\alpha}{4(\bar{\theta}-\underline{\theta})}\left[\bar{\theta}+c_A-\lambda c_B-\frac{t(\lambda-1)}{4}\right]$	0.4895
d_B^0	-	-	0.0196	$\frac{\lambda\alpha}{4(\bar{\theta}-\underline{\theta})}\left[\bar{\theta}+c_A-\lambda c_B-\frac{t(\lambda-1)}{4}\right]$	0.0131
d_B^1	-	-	-	$\frac{\lambda\alpha}{4(\bar{\theta}-\underline{\theta})}\left[\bar{\theta}+c_A-\lambda c_B-\frac{t(\lambda-1)}{4}\right]$	0.0131
π	$u-\frac{t}{2}-c$	0.32	0.3234	$\left(u-\frac{t}{2}-c_A\right)+\frac{\alpha\left[\bar{\theta}+c_A-\lambda c_B-\frac{t}{4}(\lambda-1)\right]^2}{4(\bar{\theta}-\underline{\theta})}$	0.3251

Appendices

Appendix A. Demand configurations in the market with one E85 station and two E85 stations

In this appendix, we discuss in detail the different cases in each model---duopoly model with one E85 station, duopoly model with two E85 stations, monopoly model with one E85 station, and monopoly model with two E85 stations. As mentioned in the main text, in addition to the scenario of no E85 consumption and all FFV drivers refueling with E85, there are five possible demand configurations ("cases") in the market with one E85 station, and three possible cases in the market with two E85 stations, regardless of whether we have duopoly or monopoly. An alternative scenario for case 3 in the duopoly model with one E85 station, case 3a as shown in Figure A3a below, arises under different parameter conditions (we call it case "3a" because, taking other parameters as given, either case 3 or case 3a will materialize with the increase in the subsidy). In part A1, we discuss the different cases that pertain to the market with one E85 station; in part A2, we discuss the different cases for the market with two E85 stations. Specifically, we describe the conditions under which each case arises, provide a diagrammatic illustration, and construct the corresponding demand systems. Although the markets under duopoly or monopoly share the same possible case configurations, the parametric conditions required for each case are different. The critical values of the subsidy s at which we have transition between cases are also reported in each section. Before getting into these details, however, in Table A1 we first summarize the threshold levels of drivers' characteristics (type and location) that are used to define the various cases and to derive the corresponding demand systems.

Table A1. Summary of threshold levels (drivers' type and location) for FFV drivers

FFV driver	Defined by	Expression
$\tilde{x}$	$U_A^0 = U_A^1$	$\frac{1}{2}+\frac{1}{2t}\left(p_A^1 - p_A^0\right)$
$\overline{x}^0$	$U_A^0 = U_B^0\big\vert_{\theta=\overline{\theta}}$	$\frac{p_A^0 - \lambda p_B^0 + \overline{\theta}}{(\lambda-1)t}$
$\overline{x}^1$	$U_A^1 = U_B^0\big\vert_{\theta=\overline{\theta}}$	$\frac{p_A^1 - \lambda p_B^0 + \overline{\theta} + t}{(\lambda+1)t}$
$\underline{x}^0$	$U_A^0 = U_B^0\big\vert_{\theta=\underline{\theta}}$	$\frac{p_A^0 - \lambda p_B^0 + \underline{\theta}}{(\lambda-1)t}$
$\underline{x}^1$	$U_A^1 = U_B^0\big\vert_{\theta=\underline{\theta}}$	$\frac{p_A^1 - \lambda p_B^0 + \underline{\theta} + t}{(\lambda+1)t}$
$\tilde{x}_B$	$U_B^0 = U_B^1$	$\frac{1}{2}+\frac{1}{2t}\left(p_B^1 - p_B^0\right)$
$\overline{x}_1^1$	$U_A^1 = U_B^1\big\vert_{\theta=\overline{\theta}}$	$1-\frac{p_A^1 - \lambda p_B^1 + \overline{\theta}}{(\lambda-1)t}$
$\underline{x}_1^1$	$U_A^1 = U_B^1\big\vert_{\theta=\underline{\theta}}$	$1-\frac{p_A^1 - \lambda p_B^1 + \underline{\theta}}{(\lambda-1)t}$
θ^0	$U_A^0 = U_B^0\big\vert_{x=0}$	$\lambda p_B^0 - p_A^0$
θ	$U_A^0 = U_B^0\big\vert_{x=\tilde{x}}$	$\frac{(\lambda-1)}{2}\left(p_A^1 + t\right) - \frac{(\lambda+1)}{2}p_A^0 + \lambda p_B^0$
θ_A	$U_A^1 = U_B^1\big\vert_{x=\tilde{x}}$	$\frac{(\lambda-1)}{2}\left(p_A^0 + t\right) - \frac{(\lambda+1)}{2}p_A^1 + \lambda p_B^1$
θ_B	$U_A^0 = U_B^0\big\vert_{x=\tilde{x}_B}$	$\frac{(\lambda-1)}{2}\left(p_B^1 + t\right) - p_A^0 + \frac{(\lambda+1)}{2}p_B^0$
θ^1	$U_A^1 = U_B^0\big\vert_{x=1}$	$\lambda p_B^0 - p_A^1 + \lambda t$
θ_1^1	$U_A^1 = U_B^1\big\vert_{x=1}$	$\lambda p_B^1 - p_A^1$

A1. Possible demand configurations in the market with one E85 station

In this market, normal car drivers can refuel with E10 at either L0 or L1 depending on their location and the relative retail prices of E10 at two gas stations. FFV drivers can either refuel with E10 at both locations, or E85 at L0. Their choice of fuel type and gas station relates to their type θ, location x, and the retail prices of all fuels. The key differences among all cases relate to whether $\tilde{\theta}^0 \in (\underline{\theta}, \bar{\theta})$, $\tilde{\theta} \in (\underline{\theta}, \bar{\theta})$, and $\tilde{\theta}^1 \in (\underline{\theta}, \bar{\theta})$. Here, $\tilde{\theta}^0$ is the type of the consumer at $x = 0$ who is indifferent between choosing E10 and E85 at L0; θ is the type of the consumer at $x = \tilde{x}$ who is indifferent between choosing E10 and E85 at L0; θ^1 is the type of the consumer at $x = 1$ who is indifferent between choosing E10 at L1 and E85 at L0. See Table A1 for the relevant expressions.

When p_B^0 is relatively high compared to p_A^0 and p_A^1, no FFV drivers in the market refuels with E85. As the price p_B^0 goes down, at first only high type FFV drivers with low convenience cost of refueling choose to refuel with E85. This is Case 1, which is illustrated in Figure A1. The parametric conditions for case 1 are $\underline{\theta} < \tilde{\theta}^0 < \bar{\theta}$ and $\tilde{\theta} > \bar{\theta}$. These conditions imply no direct competition between E85 at L0 and E10 at L1, and that E85 only substitutes E10 demand at the same location.

As noted in the text, we assume that FFV drivers are independently and uniformly distributed in the (θ, x) space. In Case 1, only FFV drivers at the top left corner of Figure 1 choose to refuel with E85. Other than this corner area, FFV drivers located at the left of $\tilde{x}$ would choose to refuel with E10 at L0; FFV at the right of $\tilde{x}$ choose to refuel with E10 at L1. The demand for E85 is the aggregate demand over all FFV drivers. The demands for E10 at L0 and L1 are the aggregate demands of relevant FFV drivers and normal car drivers respectively. The demand system is shown as following.

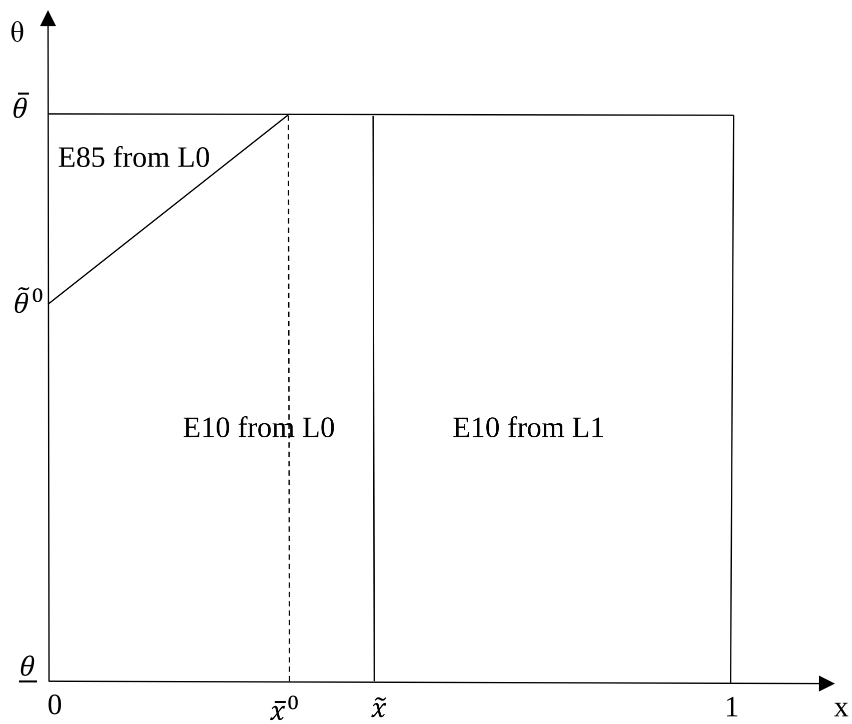


Figure A1. FFV drivers' demands in the one E85 station model ("case 1")

$$d_A^0 = (1-\alpha)\tilde{x} + \alpha\left[\left(\tilde{x} - \bar{x}^0\right) + \bar{x}^0 \frac{\bar{\theta} + \tilde{\theta}^0 - 2\underline{\theta}}{2\left(\bar{\theta} - \underline{\theta}\right)}\right]$$

$$d_B^0 = \lambda\alpha\bar{x}^0 \frac{\bar{\theta} - \tilde{\theta}^0}{2\left(\bar{\theta} - \underline{\theta}\right)}$$

$$d_A^1 = (1-\alpha)(1-\tilde{x})$$

When the price p_B^0 further decreases, FFV drivers with lower preferences of E85 and further location may choose E85. This is Case 2, the case in the main text. This case differs from Case 1 because FFV drivers on the left of $\tilde{x}$ may also refuel with E85 at L0 if they have high preferences for E85. The parametric conditions for case 2 are $\tilde{\theta}^0 > \underline{\theta}$, $\tilde{\theta} < \bar{\theta}$, and $\tilde{\theta}^0 > \bar{\theta}$. Case 2 is illustrated in Figure A2, which is exactly the same with Figure 2 in the text. The demand system is discussed in detail in the main text (section 3.3).

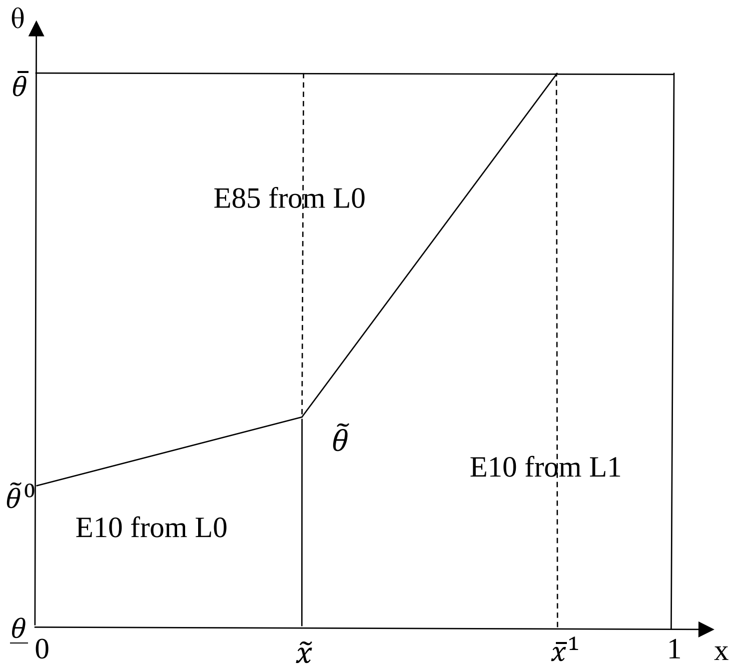


Figure A2. FFV drivers' demands in the one E85 station model ("case 2")

When p_B^0 is substantial lower than p_A^1, $\bar{x}^1$ exceeds 1, which leads to case 3 shown in Figure A3. The parametric conditions for case 3 are $\tilde{\theta}^0 > \underline{\theta}$ and $\tilde{\theta}^1 < \bar{\theta}$. We do not need the constraint on $\tilde{\theta}$ because $\tilde{\theta}^0 < \tilde{\theta} < \tilde{\theta}^1$ by definition. The demand system for this case is,

$$d_A^0 = (1-\alpha)\tilde{x} + \alpha\tilde{x}\frac{\tilde{\theta}+\tilde{\theta}^0 - 2\underline{\theta}}{2(\bar{\theta}-\underline{\theta})}$$
$$d_B^0 = \lambda\alpha\left[\tilde{x}\frac{2\bar{\theta}-\tilde{\theta}-\tilde{\theta}^0}{2(\bar{\theta}-\underline{\theta})} + (1-\tilde{x})\frac{2\bar{\theta}-\tilde{\theta}-\tilde{\theta}^1}{2(\bar{\theta}-\underline{\theta})}\right]$$
$$d_A^1 = (1-\alpha)(1-\tilde{x}) + \alpha\left[(1-\tilde{x})\frac{\tilde{\theta}+\tilde{\theta}^1-2\underline{\theta}}{2(\bar{\theta}-\underline{\theta})}\right]$$

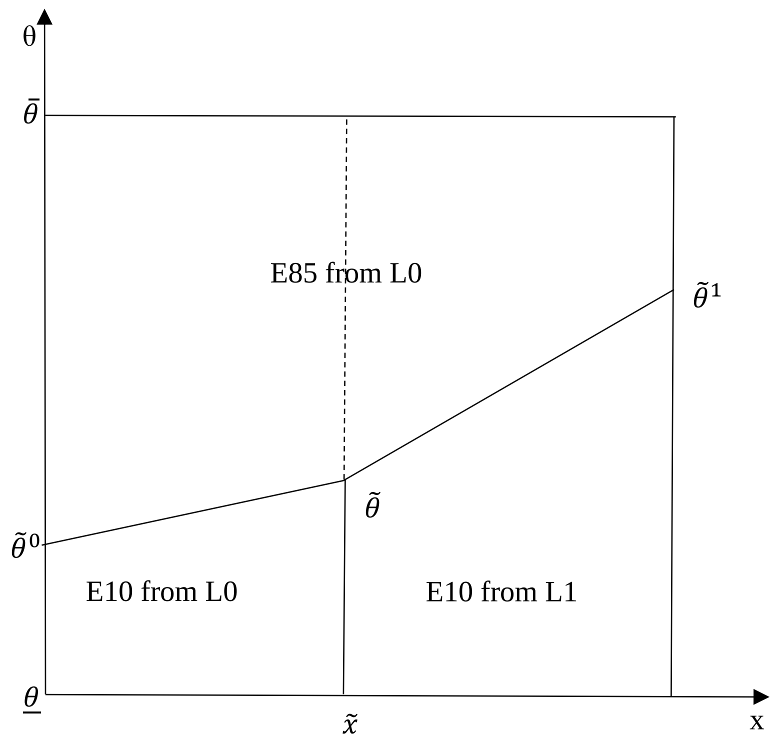


Figure A3. FFV drivers' demands in the one E85 station model ("case 3")

When p_A^0 is low enough compared to p_A^1, $\tilde{\theta}^0$ may reach $\underline{\theta}$ before $\bar{x}^1$ reaches 1, which means that all FFV drivers near L0 refuel with E85. The parametric conditions are $\tilde{\theta}^0 < \underline{\theta}$, $\tilde{\theta}^0 < \tilde{\theta} < \tilde{\theta}^1$, and $\tilde{\theta}^1 > \bar{\theta}$. This is shown as case 3a in Figure A3a. We label this case "3a" to emphasize its relationship with case 3—as the subsidy level changes, either case 3 or case 3a can arise (at the baseline values of other parameters, only case 3 can materialize). However, we will show in Table D1 in Appendix D that when $\underline{\theta}$ and $\bar{\theta}$ move away from their baseline values, case 3a may replace case 3. The demand system for case 3a is

$$d_A^0 = (1-\alpha)\tilde{x} + \alpha\left(\tilde{x} - \underline{x}^0\right)\frac{\tilde{\theta} - \underline{\theta}}{2\left(\bar{\theta} - \underline{\theta}\right)}$$

$$d_B^0 = \lambda\alpha\left[\underline{x}^0 + \left(\tilde{x} - \underline{x}^0\right)\frac{2\bar{\theta} - \tilde{\theta} - \underline{\theta}}{2\left(\bar{\theta} - \underline{\theta}\right)} + \left(\bar{x}^1 - \tilde{x}\right)\frac{\bar{\theta} - \tilde{\theta}}{2\left(\bar{\theta} - \underline{\theta}\right)}\right]$$

$$d_A^1 = (1-\alpha)(1-\tilde{x}) + \alpha\left[1 - \bar{x}^1 + \left(\bar{x}^1 - \tilde{x}\right)\frac{\bar{\theta} + \tilde{\theta} - 2\underline{\theta}}{2\left(\bar{\theta} - \underline{\theta}\right)}\right]$$

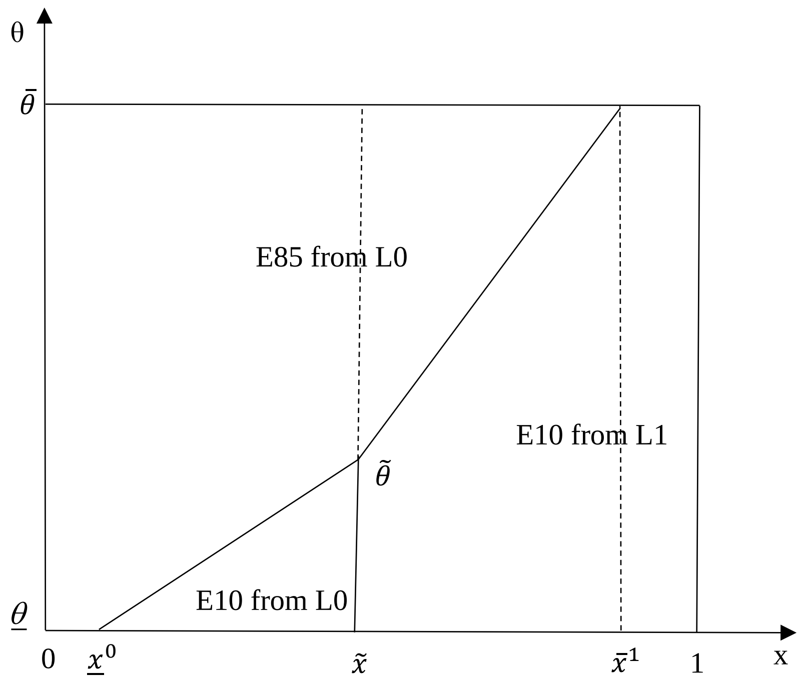


Figure A3a. FFV drivers' demands in the one E85 station model ("case 3a")

When p_B^0 further decreases, more FFV drivers choose to refuel with E85. This leads to Case 4, which has both $\bar{x}^1 > 1$ and $\tilde{\theta}^0 < \underline{\theta}$, and the parametric conditions are $\tilde{\theta}^0 < \underline{\theta}$, $\tilde{\theta} > \underline{\theta}$, and $\tilde{\theta}^1 < \bar{\theta}$. In Case 4, FFV drivers at L1 refuel with E85 if they have high θ preferences. This case is illustrated in Figure A4. The demand system for this case is:

$$d_A^0 = (1-\alpha)\tilde{x} + \alpha\left(\tilde{x} - \underline{x}^0\right)\frac{\tilde{\theta} - \underline{\theta}}{2\left(\bar{\theta} - \underline{\theta}\right)}$$

$$d_B^0 = \lambda\alpha\left[\underline{x}^0 + \left(\tilde{x} - \underline{x}^0\right)\frac{2\bar{\theta} - \tilde{\theta} - \underline{\theta}}{2\left(\bar{\theta} - \underline{\theta}\right)} + (1-\tilde{x})\frac{2\bar{\theta} - \tilde{\theta} - \tilde{\theta}^1}{2\left(\bar{\theta} - \underline{\theta}\right)}\right]$$

$$d_A^1 = (1-\alpha)(1-\tilde{x}) + \alpha(1-\tilde{x})\frac{\tilde{\theta}^1 + \tilde{\theta} - 2\underline{\theta}}{2\left(\bar{\theta} - \underline{\theta}\right)}$$

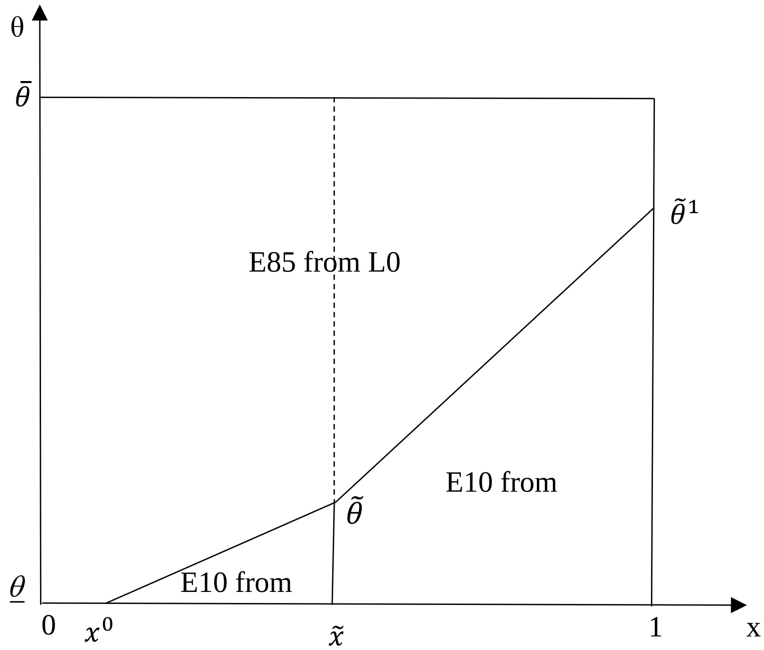


Figure A4. FFV drivers' demands in the one E85 station model ("case 4")

With even lower price of E85, only FFV drivers who locate far away from L0 and have low enough preferences for E85 refuel with E10. This is Case 5, illustrated in Figure A5. The difference between case 5 and case 4 is that now all FFV drivers on the left of $\tilde{x}$ choose to refuel with E85, even with the lowest type $\underline{\theta}$. Accordingly, the parametric conditions that pertain to this case are $\tilde{\theta} < \underline{\theta}$ and $\underline{\theta} < \tilde{\theta}^1 < \bar{\theta}$. The demand system for this case is:

$$d_A^0 = (1-\alpha)\tilde{x}$$

$$d_B^0 = \lambda\alpha\left[\underline{x}^1 + \left(1-\underline{x}^1\right)\frac{2\bar{\theta}-\underline{\theta}-\tilde{\theta}^1}{2\left(\bar{\theta}-\underline{\theta}\right)}\right]$$

$$d_A^1 = (1-\alpha)(1-\tilde{x}) + \alpha\left(1-\underline{x}^1\right)\frac{\tilde{\theta}^1-\underline{\theta}}{2\left(\bar{\theta}-\underline{\theta}\right)}$$

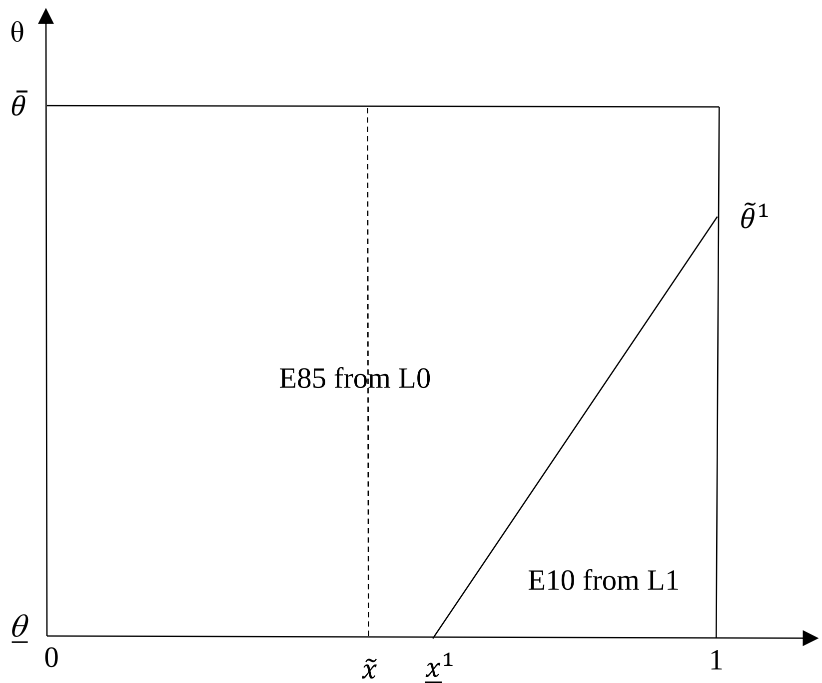


Figure A5. FFV drivers' demands in the one E85 station model ("case 5")

Eventually, when the retail price of E85 is low enough, relative to those of E10 at two locations, all FFV drivers in the market choose to refuel with E85. Under different combinations of parameters, either there is no E85 consumption in the market, or the demand for each fuel meets one of the above scenarios, or all FFV drivers in the market choose E85. In our model, all retail prices are endogenous determined. With the increase in subsidy s and decrease in E85 retail price, we would expect market equilibria to move from no E85 consumption, to case 1, to case 2, to either case 3 or case 3a, to case 4, to case 5, and then to the case where every FFV driver chooses E85. The critical values of the subsidy s that correspond to the transition between cases are reported in Table A2 (separately for the duopoly and monopoly market structures).

Table A2. Critical values of the subsidy level in the market with one E85 station

Case	Condition	Duopoly s	Monopoly s
No E85	$\tilde{\theta}^0 \geq \bar{\theta}$		
	$\downarrow$	0.1089	0.1089
Case 1	$\underline{\theta} < \tilde{\theta}^0 < \bar{\theta}$, $\tilde{\theta} > \bar{\theta}$		
	$\downarrow$	0.1569	0.1569
Case 2	$\tilde{\theta}^0 > \underline{\theta}, \tilde{\theta} < \bar{\theta}$, $\tilde{\theta}^1 > \bar{\theta}$		
	$\downarrow$	0.5265	0.6518
Case 3	$\tilde{\theta}^0 > \underline{\theta}, \tilde{\theta}^1 > \bar{\theta}$,		
	$\downarrow$	1.4735	1.6141
Case 4	$\tilde{\theta}^0 > \underline{\theta}$, $\tilde{\theta}^1 < \bar{\theta}$		
	$\downarrow$	2.1343	2.4968
Case 5	$\tilde{\theta} < \bar{\theta}$, $\underline{\theta} < \tilde{\theta}^1 < \bar{\theta}$		
	$\downarrow$	>10	>10
All E85	$\tilde{\theta}^1 < \bar{\theta}$		

A2. Possible demand configurations in the market with two E85 stations

When there are two E85 stations, no matter whether in duopoly or monopoly, the demand configurations are different. When the prices of E85 are relative high compared to E10 prices, we have case 1 for the market of two E85 stations, which is illustrated in Figure A6. Unlike the foregoing case 1 in part A1, FFV drivers with high preferences for E85 close to both stations refuel with E85. In Figure A6, $\bar{x}^0$ is the location of the FFV driver with type $\bar{\theta}$ who is indifferent between choosing E10 or E85 at L0 as in Figure A1. The threshold $\tilde{\theta}_1^1$ is newly introduced for the market with two E85 stations, and it indicates the type of FFV driver located at $x = 1$ who is indifferent between E10 and E85 at L1. As noted in the main text, we assume $\tilde{x}_A \geq \tilde{x}_B$, implying $p_A^1 - p_A^0 \geq p_B^1 - p_B^0$. The parametric conditions for case 1 are $\underline{\theta} < \tilde{\theta}^0 < \bar{\theta}$ and

$\tilde{\theta}_B > \bar{\theta}$. We do not need the restriction on $\tilde{\theta}_1^1$ and $\tilde{\theta}_A$ because in equilibrium $\tilde{\theta}_1^1 = \tilde{\theta}^0$ and $\tilde{\theta}_A = \tilde{\theta}_B$. The demand system for this case is

$$d_A^0 = (1-\alpha)\tilde{x}_A + \alpha\left[\left(\tilde{x}_A - \bar{x}^0\right) + \bar{x}^0 \frac{\bar{\theta} + \tilde{\theta}^0 - 2\underline{\theta}}{2\left(\bar{\theta} - \underline{\theta}\right)}\right]$$

$$d_B^0 = \lambda\alpha\bar{x}^0 \frac{\bar{\theta} - \tilde{\theta}^0}{2\left(\bar{\theta} - \underline{\theta}\right)}$$

$$d_A^1 = (1-\alpha)(1-\tilde{x}_A) + \alpha\left[\left(\bar{x}_1^1 - \tilde{x}_A\right) + \left(1 - \bar{x}_1^1\right)\frac{\bar{\theta} + \tilde{\theta}_1^1 - 2\underline{\theta}}{2\left(\bar{\theta} - \underline{\theta}\right)}\right]$$

$$d_B^1 = \lambda\alpha\left(1 - \bar{x}_1^1\right)\frac{\bar{\theta} - \tilde{\theta}_1^1}{2\left(\bar{\theta} - \underline{\theta}\right)}$$

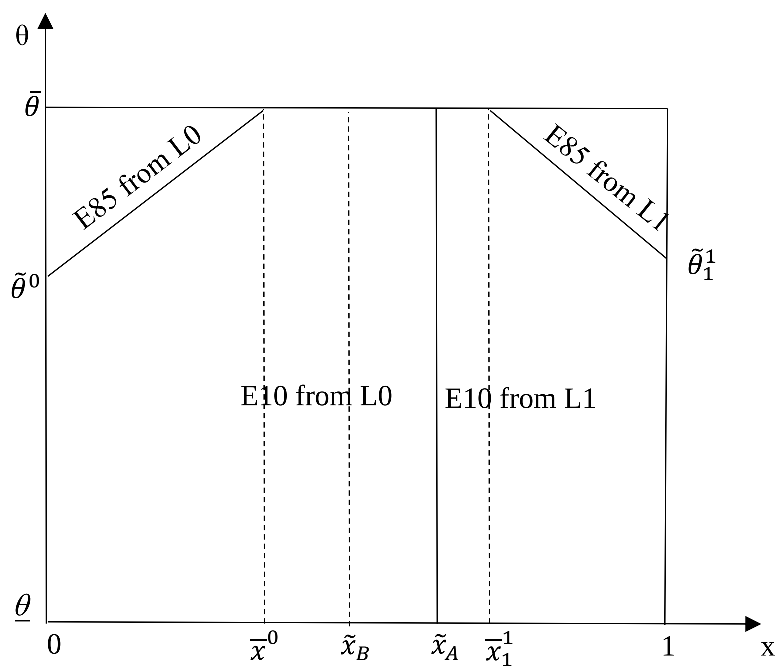


Figure A6. FFV drivers' demands in the two E85 stations model ("case 1")

When prices of E85 fall, FFV drivers with low θ preferences choose to refuel with E85, such that $\tilde{\theta}^0 > \underline{\theta}$ and $\tilde{\theta}_B < \bar{\theta}$. This is Case 2, the case discussed in the main text (section 3.3) that arises with the baseline parameter values. This case is illustrated in Figure A7 , which is exactly the same as Figure 4 in the main text (section 3.3). Section 3.3 also presents the demand system for this case.

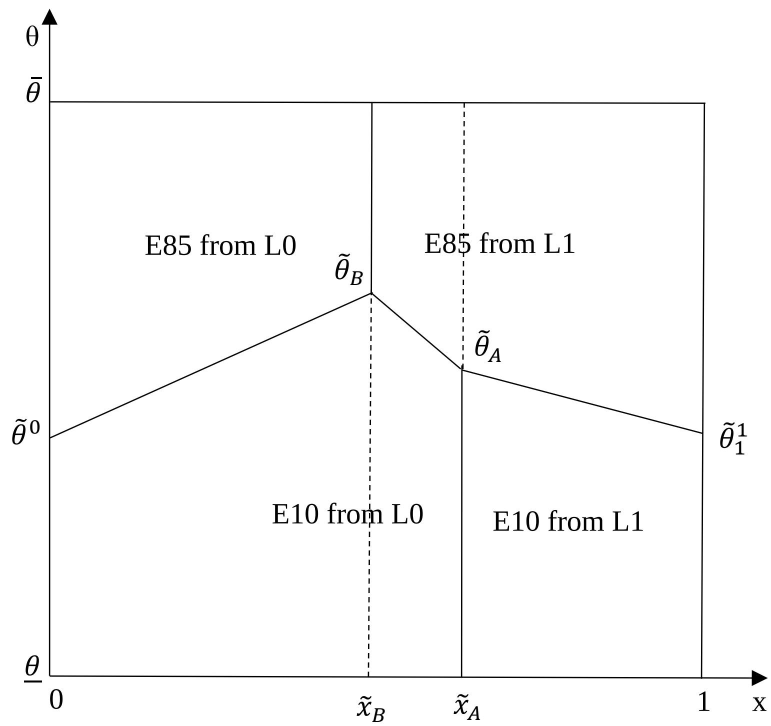


Figure A7. FFV drivers' demands in the two E85 stations model ("case 2")

When the prices of E85 continuously go down, the market moves to case 3, where $\tilde{\theta}^0 < \underline{\theta}$ and $\underline{\theta} < \tilde{\theta}_B < \bar{\theta}$. The scenario is depicted in Figure A8, and is associated with the following demand system:

$$d_A^0 = (1-\alpha)\tilde{x}_A + \alpha\left[(\tilde{x}_A - \tilde{x}_B)\frac{\tilde{\theta}_A + \tilde{\theta}_B - 2\underline{\theta}}{2(\bar{\theta} - \underline{\theta})} + \left(\tilde{x}_B - \underline{x}^0\right)\frac{\tilde{\theta}_B - \underline{\theta}}{2(\bar{\theta} - \underline{\theta})} \right]$$

$$d_B^0 = \lambda\alpha\left[\underline{x}^0 + \left(\tilde{x}_B - \underline{x}^0\right)\frac{2\bar{\theta} - \tilde{\theta}_B - \underline{\theta}}{2(\bar{\theta} - \underline{\theta})} \right]$$

$$d_A^1 = (1-\alpha)(1-\tilde{x}_A) + \alpha\left[\left(\bar{x}_1^1 - \tilde{x}_A\right)\frac{\tilde{\theta}_A - \underline{\theta}}{2(\bar{\theta} - \underline{\theta})} \right]$$

$$d_B^1 = \lambda\alpha\left[\left(1 - \bar{x}_1^1\right) + \left(\bar{x}_1^1 - \tilde{x}_A\right)\frac{2\bar{\theta} - \tilde{\theta}_A - \underline{\theta}}{2(\bar{\theta} - \underline{\theta})} + (\tilde{x}_A - \tilde{x}_B)\frac{2\bar{\theta} - \tilde{\theta}_A - \tilde{\theta}_B}{2(\bar{\theta} - \underline{\theta})} \right]$$

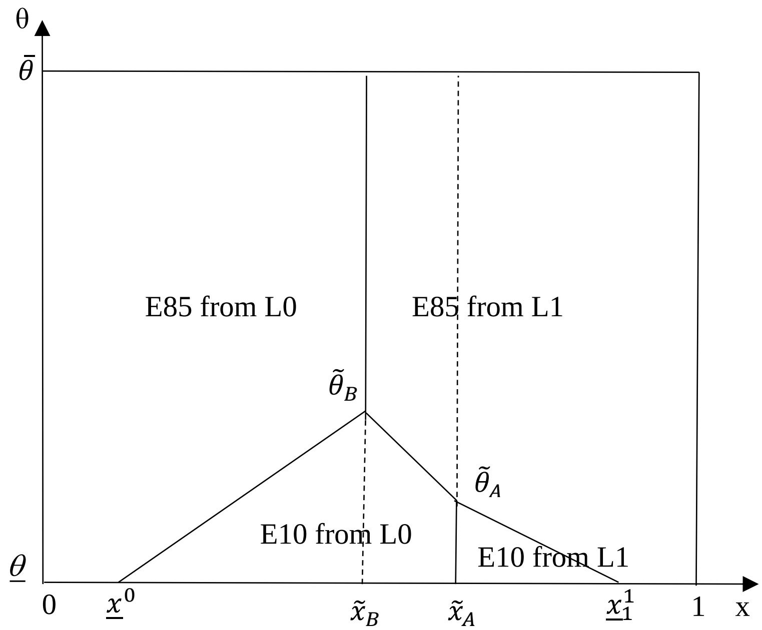


Figure A8. FFV drivers' demands in the two E85 stations model ("case 3")

When prices of E85 are very low compared to E10, all FFV drivers in the market choose to refuel with E85. Because we only sort for symmetric equilibria, we assume the prices differences between the same type of fuel at different locations are negligible compared to the price differences between E10 and E85 in all three cases above. We do not consider the scenario in which E85 price is high in one station but relatively low in another station such that the parametric requirements on $\tilde{\theta}_A$ and $\tilde{\theta}_1^1$ are different from those on $\tilde{\theta}_B$ and $\tilde{\theta}^0$ (for example, $\tilde{\theta}_B > \bar{\theta}$ whereas $\tilde{\theta}_A < \bar{\theta}$). The scenario that actually materializes in the market depends on fuel prices, which further depends on parameters. At the baseline values of all other parameters, as we increase the subsidy s from 0, the market moves from no E85 consumption, to case 1, to case 2, to case 3, and then to the situation where all FFV drivers refuel with E85. The critical values of s that correspond to these transitions are listed in Table A3.

Table A3. Critical values of the subsidy level in the market with two E85 stations

Case	Condition	Duopoly s	Monopoly s
No E85	$\tilde{\theta}^0 > \bar{\theta}$		
	$\downarrow$	0.1089	0.1089
Case 1	$\underline{\theta} < \tilde{\theta}^0 < \bar{\theta}$, $\tilde{\theta}_B > \bar{\theta}$		
	$\downarrow$	0.1569	0.1569
Case 2	$\tilde{\theta}^0 > \underline{\theta}$, $\tilde{\theta}_B < \bar{\theta}$		
	$\downarrow$	0.9581	1.693
Case 3	$\tilde{\theta}^0 < \underline{\theta}$, $\underline{\theta} < \tilde{\theta}_B < \bar{\theta}$		
	$\downarrow$	1.0049	>10
All E85	$\tilde{\theta}_B < \bar{\theta}$		

Appendix B. Pass-through of the subsidy to equilibrium prices

In the duopoly model with one E85 station, the profits of gas stations at L0 and L1 are,

$$
\begin{aligned}
\pi^0 &= \left(p_A^0 - c_A\right)d_A^0 + \left(p_B^0 - c_B\right)d_B^0 \\
\pi^1 &= \left(p_A^1 - c_A\right)d_A^1
\end{aligned}
$$

The gas station at L0 maximizes its profit with respect to p_A^0 and p_B^0, while the gas station at L1 maximizes its profit with respect to p_A^1. Their best response functions can be derived from,

$$
\begin{cases}
\dfrac{\partial \pi^0}{\partial p_A^0}\left(p_A^0, p_B^0, p_A^1 \mid s\right) = 0 \\
\dfrac{\partial \pi^0}{\partial p_B^0}\left(p_A^0, p_B^0, p_A^1 \mid s\right) = 0 \\
\dfrac{\partial \pi^1}{\partial p_A^1}\left(p_A^0, p_B^0, p_A^1 \mid s\right) = 0
\end{cases}
$$

Recall that the subsidy level s affects the equilibrium prices through c_B, the cost of E85. Equilibrium prices are functions of the subsidy s, $p_A^0(s)$, $p_A^1(s)$, and $p_B^0(s)$. Because we do not have closed-form solutions for these equilibrium prices, they are simulated using Matlab at different values of the subsidy while holding other parameters at their baseline. The pass-through rates of s to equilibrium prices, p_A^0, p_B^0, and p_A^1, are determined by comparative statics.

$$\begin{cases} \dfrac{\partial^2\pi^0}{\left(\partial p_A^0\right)^2}\dfrac{\partial p_A^0}{\partial s}+\dfrac{\partial^2\pi^0}{\partial p_A^0\partial p_B^0}\dfrac{\partial p_B^0}{\partial s}+\dfrac{\partial^2\pi^0}{\partial p_A^0\partial p_A^1}\dfrac{\partial p_A^1}{\partial s}+\dfrac{\partial^2\pi^0}{\partial p_A^0\partial s}=0 \\ \dfrac{\partial^2\pi^0}{\partial p_A^0\partial p_B^0}\dfrac{\partial p_A^0}{\partial s}+\dfrac{\partial^2\pi^0}{\left(\partial p_B^0\right)^2}\dfrac{\partial p_B^0}{\partial s}+\dfrac{\partial^2\pi^0}{\partial p_B^0\partial p_A^1}\dfrac{\partial p_A^1}{\partial s}+\dfrac{\partial^2\pi^0}{\partial p_B^0\partial s}=0 \\ \dfrac{\partial^2\pi^1}{\partial p_A^1\partial p_A^0}\dfrac{\partial p_A^0}{\partial s}+\dfrac{\partial^2\pi^1}{\partial p_A^1\partial p_B^0}\dfrac{\partial p_B^0}{\partial s}+\dfrac{\partial^2\pi^1}{\left(\partial p_A^1\right)^2}\dfrac{\partial p_A^1}{\partial s}+\dfrac{\partial^2\pi^1}{\partial p_A^1\partial s}=0 \end{cases}$$

In this system of equations, the variables to be solved for are $\partial p_A^0 / \partial s$, $\partial p_B^0 / \partial s$, and $\partial p_B^0 / \partial s$. The coefficients are all the second-order derivatives evaluated at equilibrium prices, whose values are further determined by the subsidy s. This system of equations is solved in Matlab using *vpasolve*.

The pass-through rate of interest, defined as $\partial\left(p_A - p_B\right)/\partial s$, is $0.5\left(\partial p_A^0 / \partial s + \partial p_A^1 / \partial s\right) - \partial p_B^0 / \partial s$. Pass-through rates for the other models---duopoly with two E85 station, monopoly with one E85 station, and monopoly with two E85 stations—are derived in the say way: first, best response functions are constructed under profit-maximizing conditions; then, pass-through rates are calculated by comparative statics assuming exogenous parameters other than the cost of E85. In models with two E85 stations, $\partial\left(p_A - p_B\right)/\partial s$ is directly $\partial\left(p_A^0 - p_B^0\right)/\partial s$ because in equilibrium $\partial p_A^0 / \partial s = \partial p_A^1 / \partial s$ and $\partial p_B^0 / \partial s = \partial p_B^1 / \partial s$.

Appendix C. Effects of the subsidy on equilibrium (all models)

In this appendix, we complete the analysis of comparative statics effects of the subsidy on equilibrium in all models. We first provide Figure C1 which shows how pass-through rates evolve with the subsidy level in all models. We then supplement Figure 6 of section 5.2 by Table C1 with equilibrium results at representative values of the subsidy. We also provide effects of the subsidy on equilibrium in the other models—duopoly with two E85 stations, monopoly with one E85 station, and monopoly with two E85 stations. For each model, we provide diagrams (Figure C2, Figure C3, and Figure C4) illustrating the effects on prices, demands, and pass-through rates, along with tables (Table C2, Table C3, and Table C4) of equilibrium solutions at various representative values of the subsidy.

Panels (1) and (3) in Figure C1 correspond to Figure 5 and Figure 7 in the main text, respectively. The two panels on the right in Figure C1 are their counterparts for the models with two E85 stations. It is clear that case 1 results in the same pass-through rate in all models. When $s > 0.1569$ (kink point of case 1 and case 2 in all models), the pass-through rate in panel (2) of Figure C1 jumps to 0.87 and then increases toward 1. Another discontinuity instance in panel (2) of Figure C1 arises at $s = 0.9581$, when the demand configuration changes from case 2 to case 3 (Figure A7 and A8, respectively). The pass-through rate increases and jumps back to one at $s = 1.0049$. In panel (4) of Figure C1, for the monopoly model with two E85 stations, the pass-through rate jumps down to about 0.5 for $s > 0.1569$ and falls further toward zero for $s > 1.6930$. By comparing all panels in Figure C1, we observe that, except for panel (2) that corresponds to the duopoly model with two E85 stations, pass-through rates in all other models decrease toward zero as the subsidy level increases.

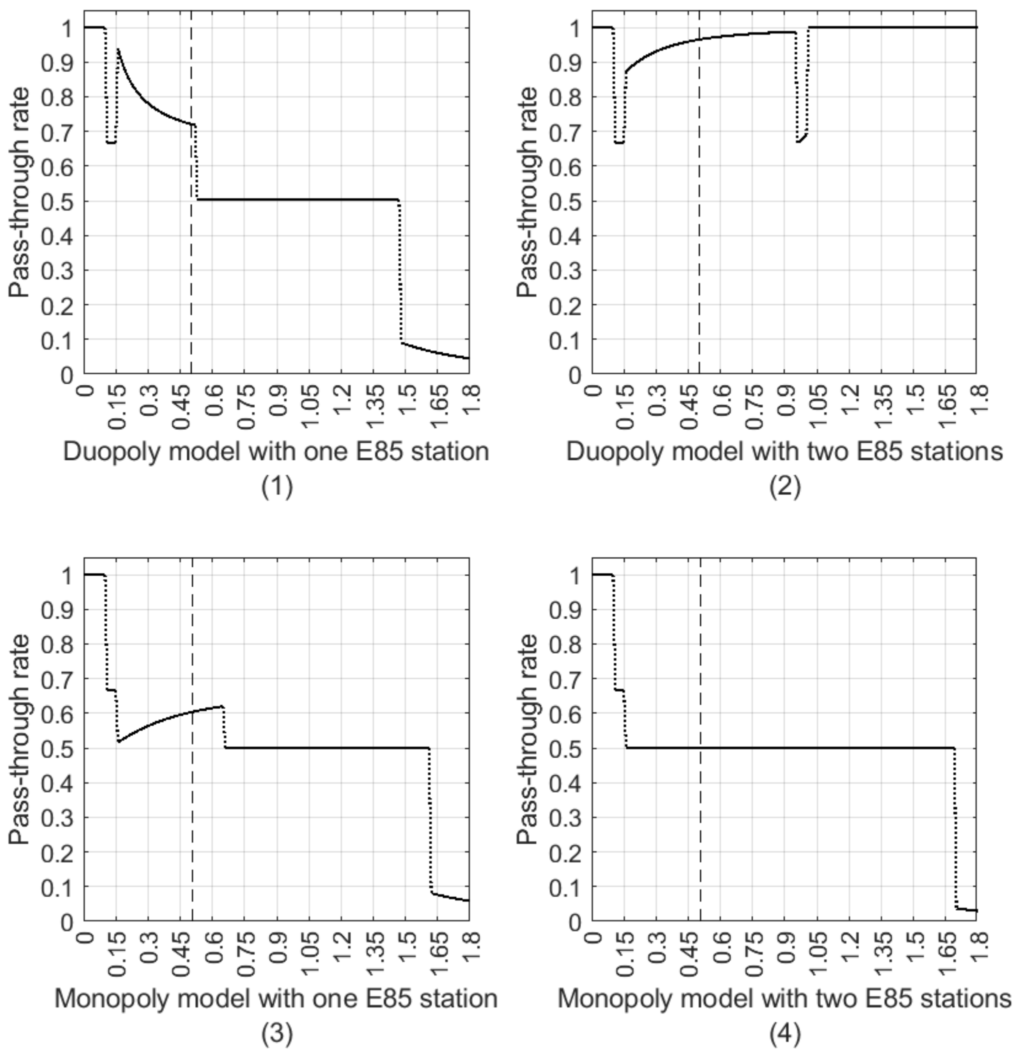


Figure C1. Pass-through rates and the E85 subsidy in all models

In the tables below, the first row indicates the "case" experienced by the market under the value of the subsidy level shown in the second row. Case "No E85" means that there is no E85 consumption with the parameters associated with this scenario (see Appendix A for a discussion of all the cases). For all figures, we have consider a wider range for the subsidy s, from zero to 1.8. In the tables, however, we report the results for subsidy levels of 0, 0.1, 0.15, 0.2, 0.4, 0.5143 (the baseline), 0.6, 0.8, and 1.

Table C1. Effect of the E85 subsidy in the duopoly model with one E85 station

Case	No E85		1	2			3		
S	0	0.1	0.15	0.2	0.4	**0.5143**	0.6	0.8	1
p_A^0	2.3621	2.3621	2.3621	2.3622	2.3627	**2.3627**	2.3635	2.3659	2.3683
p_A^1	2.3621	2.3621	2.3621	2.3622	2.3616	**2.3606**	2.3608	2.3620	2.3631
p_B^0	2.1986	2.0986	2.0623	2.0189	1.8611	**1.7772**	1.7319	1.6328	1.5338
d_A^0	0.5	0.5	0.4994	0.4970	0.4869	**0.4807**	0.4774	0.4702	0.4630
d_A^1	0.5	0.5	0.5	0.4999	0.4980	**0.4959**	0.4943	0.4907	0.4870
d_B^0	0	0	0.0008	0.0039	0.0189	**0.0292**	0.0353	0.0489	0.0625
$\frac{\partial p_A^0}{\partial s}$	0	0	0	0.0029	0.0011	**-0.0006**	0.0117	0.0119	0.0121
$\frac{\partial p_B^0}{\partial s}$	-1	-1	-0.6667	-0.8642	-0.7461	**-0.7247**	-0.4954	-0.4954	-0.4953
$\frac{\partial p_A^1}{\partial s}$	0	0	0	0.0011	-0.0069	**-0.0112**	0.0057	0.0058	0.0058
$\frac{\partial (p_A - p_B)}{\partial s}$	1	1	0.6667	0.8661	0.7431	**0.7188**	0.5041	0.5042	0.5043
π_0	0.16	0.16	0.1600	0.1602	0.1621	**0.1643**	0.1672	0.1762	0.1880
π_1	0.16	0.16	0.16	0.1600	0.1591	**0.1579**	0.1576	0.1570	0.1564

Table C2. Effect of the E85 subsidy in the duopoly model with two E85 stations

Case	0		1	2					3
s	0	0.1	0.15	0.2	0.4	**0.5143**	0.6	0.8	1
p_A^0	2.3621	2.3621	2.3621	2.3623	2.3628	**2.3630**	2.3631	2.3632	2.3622
p_B^0	2.1986	2.0986	2.0623	2.0198	1.8347	**1.7251**	1.6421	1.4464	1.2612
d_A^0	0.5	0.5	0.4994	0.4971	0.4871	**0.4811**	0.4767	0.4661	0.4568
d_B^0	0	0	0.0008	0.0036	0.0162	**0.0236**	0.0292	0.0424	0.0540
$\frac{\partial p_A^0}{\partial s}$	0	0	0	0.0037	0.0018	**0.0013**	0.0011	0.0007	-0.0266
$\frac{\partial p_B^0}{\partial s}$	-1	-1	-0.6667	-0.8906	-0.9503	**-0.9652**	-0.9725	-0.9827	-0.7160
$\frac{\partial (p_A - p_B)}{\partial s}$	1	1	0.6667	0.8943	0.9520	**0.9665**	0.9736	0.9835	0.6894
π_0	0.16	0.16	0.1600	0.1602	0.1609	**0.1614**	0.1617	0.1626	0.1634

Table C3. Effect of the E85 subsidy in the monopoly model with one E85 station

Case	0		1	2				3	
s	0	0.1	0.15	0.2	0.4	**0.5143**	0.6	0.8	1
p_A^0	2.3621	2.3621	2.3621	2.3621	2.3619	**2.3617**	2.3615	2.3614	2.3614
p_A^1	2.3621	2.3621		2.3621	2.3623	**2.3625**	2.3627	2.3628	2.3628
p_B^0	2.1986	2.0986	2.0623	2.0351	1.9227	**1.8546**	1.8022	1.6961	1.5961
d_A^0	0.5	0.5	0.4994	0.4980	0.4926	**0.4895**	0.4872	0.4817	0.4763
d_A^1	0.5	0.5	0.5	0.4999	0.4976	**0.4948**	0.4921	0.4861	0.4807
d_B^0	0	0	0.0008	0.0027	0.0123	**0.0196**	0.0260	0.0402	0.0537
$\frac{\partial p_A^0}{\partial s}$	0	0	0	-0.0005	-0.0015	**-0.0018**	-0.0020	0.0000	0.0000
$\frac{\partial p_B^0}{\partial s}$	-1	-1	-0.6667	-0.5317	-0.5865	**-0.6051**	-0.6153	-0.5000	-0.5000
$\frac{\partial p_A^1}{\partial s}$	0	0	0	0.0005	0.0015	**0.0018**	0.0020	0.0000	0.0000
$\frac{\partial (p_A - p_B)}{\partial s}$	1	1	0.6667	0.5317	0.5865	**0.6051**	0.6153	0.5000	0.5000
π_0	0.16	0.16	0.1600	0.1601	0.1622	**0.1648**	0.1675	0.1761	0.1872
π_1	0.16	0.16	0.16	0.1600	0.1593	**0.1586**	0.1578	0.1559	0.1542

Table C4. Effect of the E85 subsidy in the monopoly model with two E85 stations

Case	0		1	2					
s	0	0.1	0.15	0.2	0.4	**0.5143**	0.6	0.8	1
p_A^0	2.3621	2.3621	2.3621	2.3621	2.3621	**2.3621**	2.3621	2.3621	2.3621
p_B^0	2.1986	2.0986	2.0623	2.0361	1.9361	**1.8790**	1.8361	1.7361	1.6361
d_A^0	0.5	0.5	0.4994	0.4980	0.4926	**0.4895**	0.4872	0.4818	0.4764
d_B^0	0	0	0.0008	0.0025	0.0093	**0.0131**	0.0160	0.0228	0.0295
$\frac{\partial p_A^0}{\partial s}$	0	0	0	0	0	**0**	0	0	0
$\frac{\partial p_B^0}{\partial s}$	-1	-1	-0.6667	-0.5	-0.5	**-0.5**	-0.5	-0.5	-0.5
$\frac{\partial (p_A - p_B)}{\partial s}$	1	1	0.6667	0.5	0.5	**0.5**	0.5	0.5	0.5
π	0.32	0.32	0.3200	0.3202	0.3226	**0.3251**	0.3276	0.3354	0.3458

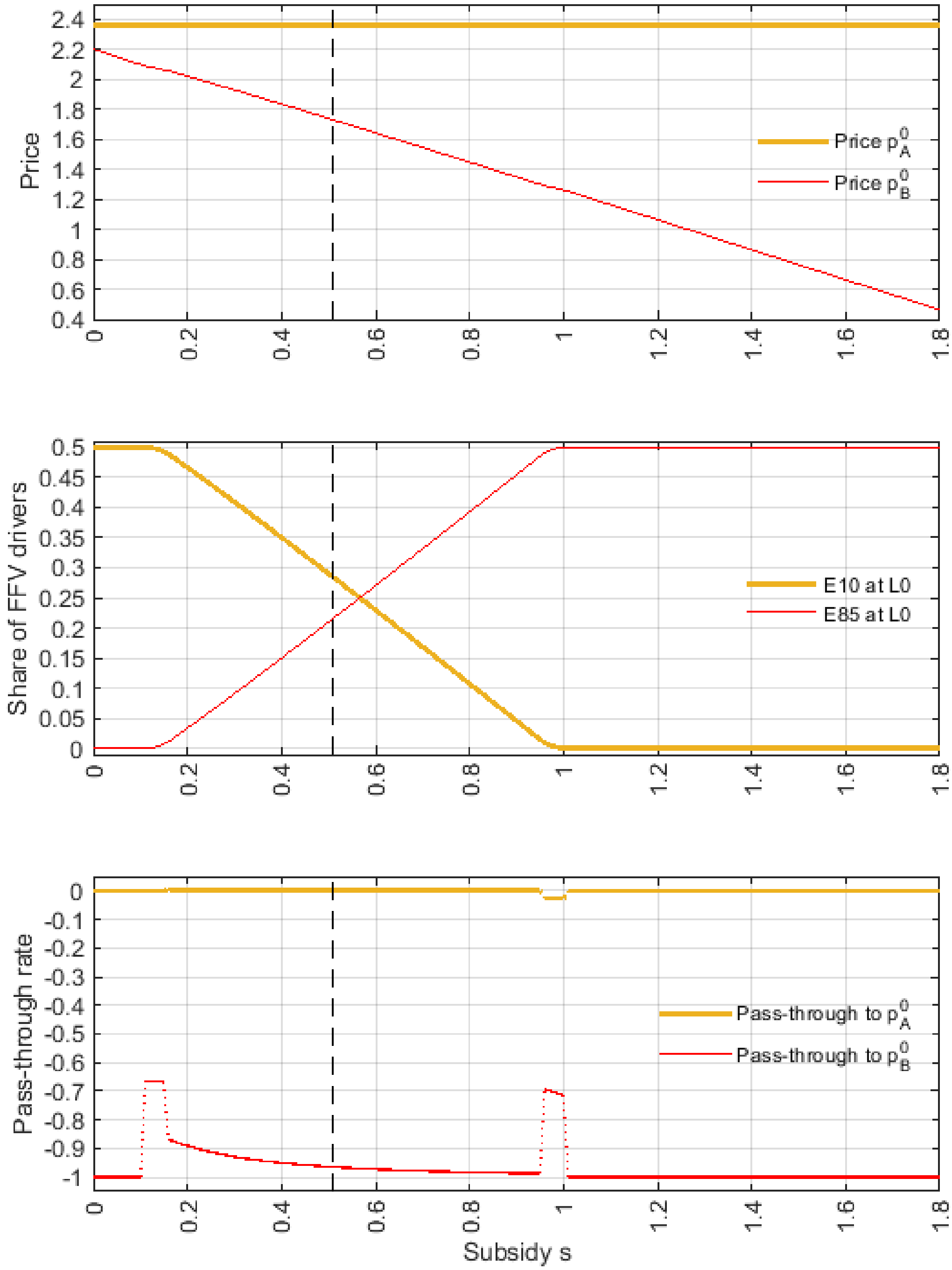


Figure C2. Simulated equilibrium results of the duopoly model with two E85 stations

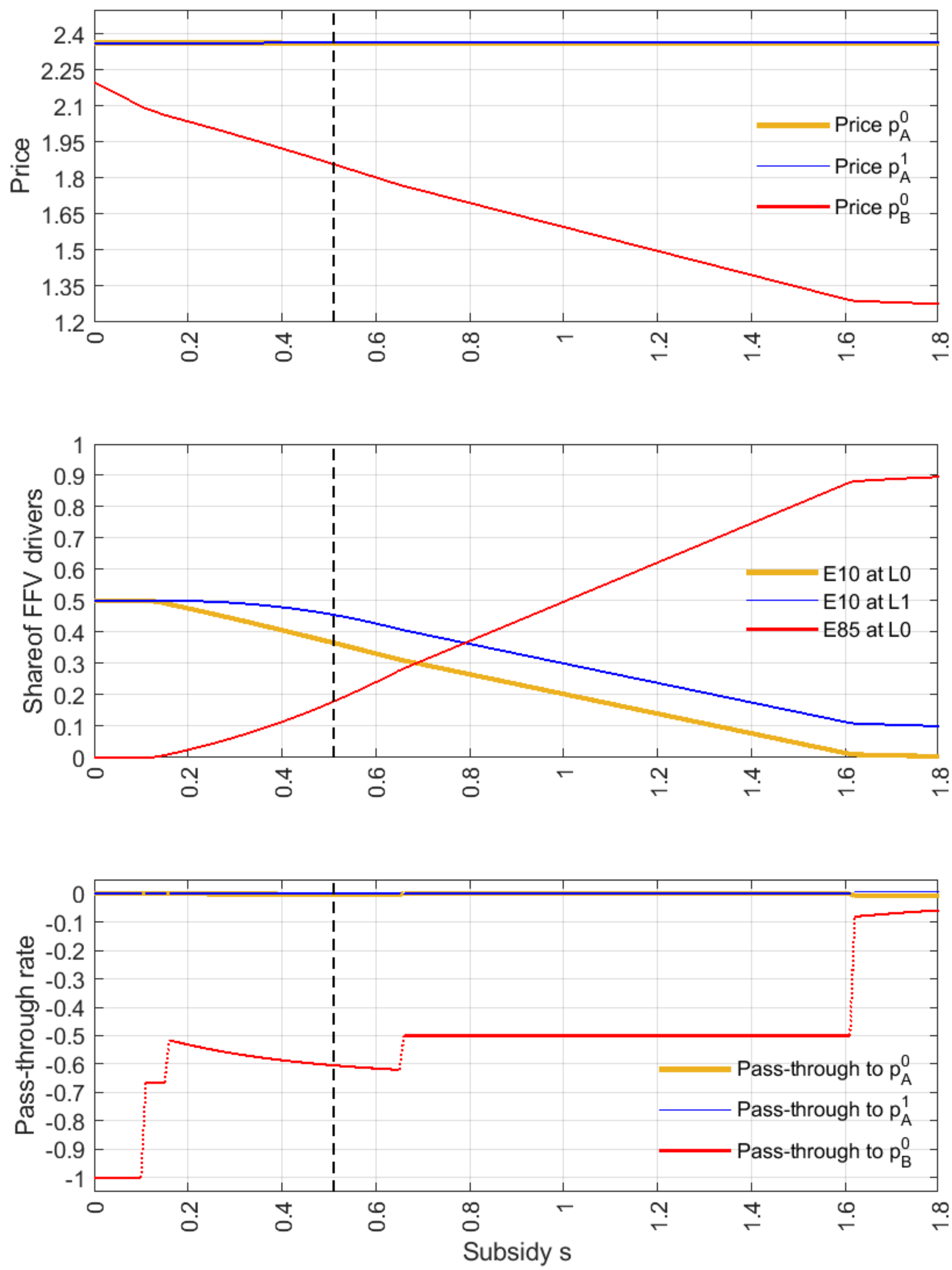


Figure C3. Simulated equilibrium results of the monopoly model with one E85 station

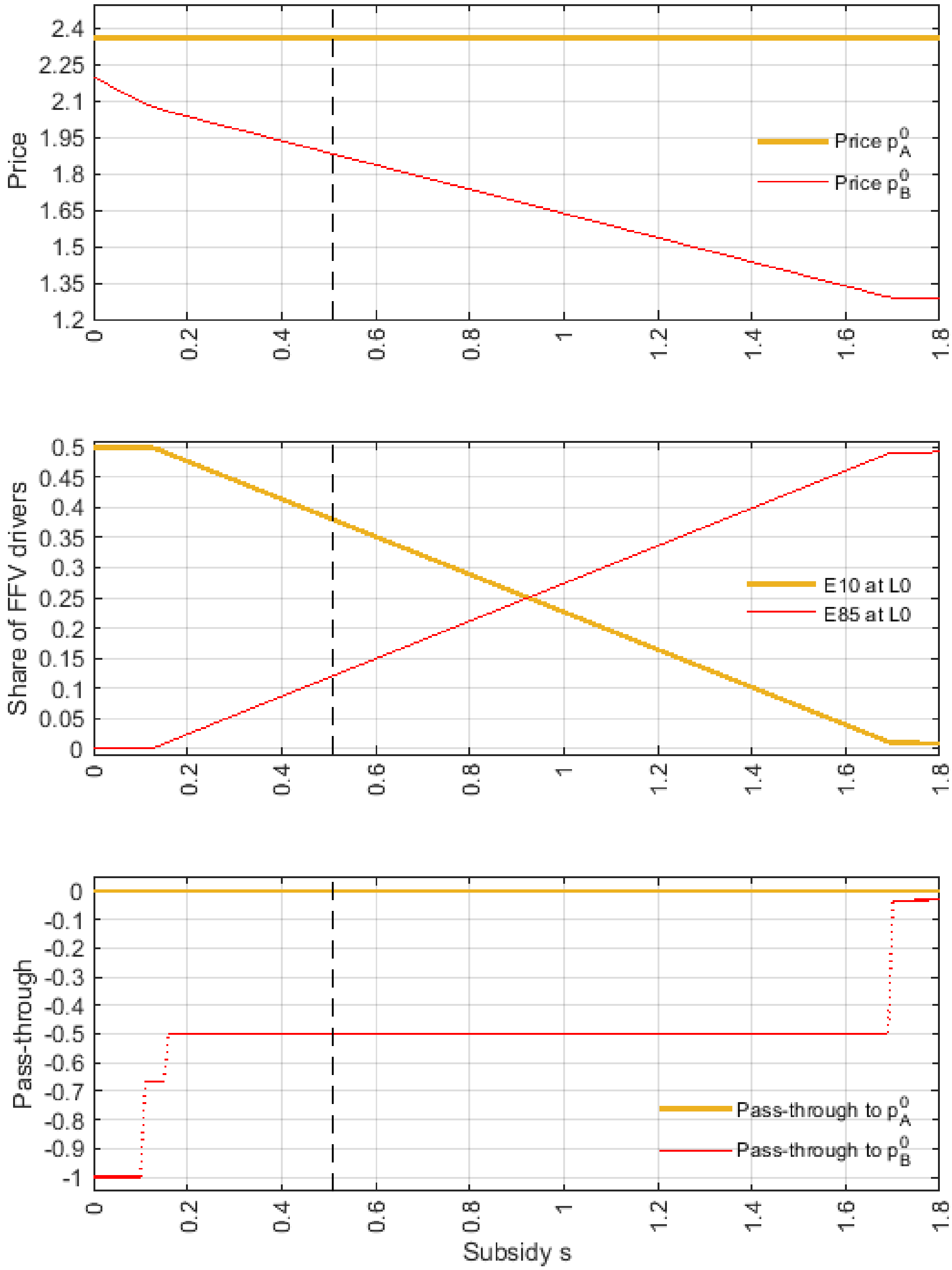


Figure C4. Simulated equilibrium results of the monopoly model with two E85 stations

Appendix D. Effects of other parameters on equilibrium (duopoly model with one E85 station)

In a stylized model, our results depend on the calibrated parameters—the subsidy s, marginal costs of each fuel, FFV fraction size α, consumers' reservation utility u, Hotelling's travel cost t, the preference upper bound $\bar{\theta}$, and lower bound $\underline{\theta}$. In this appendix, we evaluate how other parameters (beyond the E85 subsidy) affect equilibrium, and thus investigate the sensitivity of our model results. Specifically, we run the model at various representative values of FFV fraction size, Hotelling's travel cost, high type preference, and low type preference.

As stated in section 4, the calibration of preference bounds $\bar{\theta}$ and $\underline{\theta}$ relies on some model feature and previous literature, specifically Pouliot, Liao, and Babcock (2018)'s estimates of WTP for E85. To get an idea of what the equilibrium would be at higher or lower consumer preferences for E85, we simulate our results under different values of $\bar{\theta}$ and $\underline{\theta}$ in the duopoly model with one E85 station.

In Table D1, we let $\underline{\theta}$ change from -1.5 to -0.1 (with other parameters held at their baseline values). The value $\underline{\theta} = -1.5$, which is more than half of the fuel price in absolute value, means that the driver strongly dislike E85. At $\underline{\theta} = -0.1$, only a small fraction of FFV drivers has negative preferences for E85. In Table D2, we let $\bar{\theta}$ vary from 0.01 to 2 (with other parameters held at their baseline values). In the first row of Table D1 and D2, "2" means case 2, "3" means case 3, and "3a" means case 3a, and these are all cases in the model with one E85 station (see Appendix A1). As shown in these tables, we find that $\underline{\theta}$ has little effect on the equilibrium, although it may affect which case (demand configuration) arises. On the other hand, $\bar{\theta}$ does have significant effects on equilibrium. With an increase in $\bar{\theta}$, the equilibrium prices of E10 barely change but equilibrium price of E85 goes up significantly, which is consistent with the decrease in the pass-through rate. Similar to the lower bound parameter, changing the upper bound parameter also affects which case may materialize, and which demand system matters.

Table D3 reports the equilibrium results at different fractions of FFVs. As α increases from 0.01 to one, equilibrium prices and the pass-through rate of the subsidy to the price spread barely change, whereas E85 demand goes up proportionally, along with a decrease in both E10 demands. The equilibrium results under the travel cost are reported in Table D4, where we allow its representative values to vary from 0.01 to one. All prices go up with t as expected (recall in the basic Hotelling model, this parameter decides the price margin), so does the price difference between E10 and E85. Consistently, the pass-through rate goes up—with higher t, the market is more differentiated, which indicates less competition between two gas stations, which is further associated with more competition between E10 and E85 at the same location. The E85 demand, interestingly, first goes up and then goes down, with a peak at the baseline value of t.

Table D1. Effect of low type preference in the duopoly model with one E85 station

Case	2						3a
$\underline{\theta}$	-1.5	-1.25	-1	**-0.75**	-0.5	-0.25	-0.1
p_A^0	2.3624	2.3625	2.3626	**2.3627**	2.3629	2.3633	2.3623
p_A^1	2.3612	2.3611	2.3609	**2.3606**	2.3600	2.3590	2.3577
p_B^0	1.7772	1.7772	1.7772	**1.7772**	1.7771	1.7770	1.7948
d_A^0	0.4890	0.4872	0.4846	**0.4807**	0.4743	0.4615	0.4509
d_A^1	0.4977	0.4973	0.4967	**0.4959**	0.4945	0.4918	0.4886
d_B^0	0.0167	0.0195	0.0234	**0.0292**	0.0389	0.0583	0.0756
$\frac{\partial p_A^0}{\partial s}$	-0.0004	-0.0005	-0.0005	**-0.0006**	-0.0008	-0.0010	-0.0395
$\frac{\partial p_B^0}{\partial s}$	-0.7234	-0.7237	-0.7241	**-0.7247**	-0.7257	-0.7279	-0.3090
$\frac{\partial p_A^1}{\partial s}$	-0.0064	-0.0074	-0.0089	**-0.0112**	-0.0149	-0.0225	-0.0309
$\frac{\partial (p_A - p_B)}{\partial s}$	0.7200	0.7197	0.7193	**0.7188**	0.7179	0.7161	0.2738
π_0	0.1625	0.1629	0.1634	**0.1643**	0.1657	0.1686	0.1721
π_1	0.1588	0.1586	0.1583	**0.1579**	0.1572	0.1559	0.1542

Table D2. Effect of high type preference in the duopoly model with one E85 station

Case	2			3			
$\bar{\theta}$	0.01	0.1	**0.25**	0.5	1	1.5	2
p_A^0	2.3627	2.3627	**2.3627**	2.3643	2.3664	2.3675	2.3683
p_A^1	2.3620	2.3615	**2.3606**	2.3616	2.3631	2.3639	2.3644
p_B^0	1.7280	1.7454	**1.7772**	1.8752	2.0762	2.2768	2.4771
d_A^0	0.4880	0.4849	**0.4807**	0.4786	0.4765	0.4753	0.4746
d_A^1	0.4987	0.4978	**0.4959**	0.4938	0.4914	0.4901	0.4893
d_B^0	0.0166	0.0217	**0.0292**	0.0345	0.0401	0.0432	0.0452
$\frac{\partial p_A^0}{\partial s}$	0.0027	0.0014	**-0.0006**	0.0094	0.0068	0.0053	0.0044
$\frac{\partial p_B^0}{\partial s}$	-0.7716	-0.7478	**-0.7247**	-0.4963	-0.4973	-0.4979	-0.4983
$\frac{\partial p_A^1}{\partial s}$	-0.0051	-0.0079	**-0.0112**	0.0046	0.0033	0.0026	0.0022
$\frac{\partial (p_A - p_B)}{\partial s}$	0.7704	0.7446	**0.7188**	0.5033	0.5024	0.5019	0.5015
π_0	0.1614	0.1623	**0.1643**	0.1696	0.1805	0.1913	0.2022
π_1	0.1595	0.1590	**0.1579**	0.1578	0.1577	0.1577	0.1577

Table D3. Effect of the fraction of FFVs in the duopoly model with one E85 station

Case	2					
α	0.01	**0.0864**	0.25	0.5	0.75	1
p_A^0	2.3622	**2.3627**	2.3639	2.3661	2.3689	2.3725
p_A^1	2.3619	**2.3606**	2.3576	2.3530	2.3484	2.3436
p_B^0	1.7773	**1.7772**	1.7768	1.7764	1.7760	1.7757
d_A^0	0.4978	**0.4807**	0.4444	0.3892	0.3344	0.2798
d_A^1	0.4995	**0.4959**	0.4881	0.4762	0.4641	0.4519
d_B^0	0.0034	**0.0292**	0.0843	0.1682	0.2519	0.3353
$\frac{\partial p_A^0}{\partial s}$	-0.0001	**-0.0006**	-0.0011	0.0005	0.0062	0.0186
$\frac{\partial p_B^0}{\partial s}$	-0.7219	**-0.7247**	-0.7307	-0.7399	-0.7491	-0.7583
$\frac{\partial p_A^1}{\partial s}$	-0.0013	**-0.0112**	-0.0327	-0.0665	-0.1014	-0.1376
$\frac{\partial (p_A - p_B)}{\partial s}$	0.7212	**0.7188**	0.7138	0.7069	0.7015	0.6988
π_0	0.1605	**0.1643**	0.1724	0.1847	0.1969	0.2090
π_1	0.1598	**0.1579**	0.1540	0.1481	0.1421	0.1362

Table D4. Effect of travel cost in the duopoly model with one E85 station

Case	3			2			
t	0.01	0.1	0.15	**0.32**	0.5	0.75	1
p_A^0	2.0522	2.1432	2.1934	**2.3627**	2.5434	2.7940	3.0443
p_A^1	2.0522	2.1424	2.1923	**2.3606**	2.5413	2.7923	3.0431
p_B^0	1.6355	1.6763	1.6988	**1.7772**	1.9137	2.1104	2.3117
d_A^0	0.4853	0.4839	0.4831	**0.4807**	0.4820	0.4834	0.4846
d_A^1	0.4928	0.4937	0.4942	**0.4959**	0.4976	0.4986	0.4991
d_B^0	0.0274	0.0281	0.0284	**0.0292**	0.0254	0.0225	0.0203
$\frac{\partial p_A^0}{\partial s}$	0.0004	0.0037	0.0055	**-0.0006**	0.0024	0.0059	0.0089
$\frac{\partial p_B^0}{\partial s}$	-0.4999	-0.4985	-0.4978	**-0.7247**	-0.7555	-0.7910	-0.8199
$\frac{\partial p_A^1}{\partial s}$	0.0002	0.0018	0.0027	**-0.0112**	-0.0089	-0.0054	-0.0017
$\frac{\partial (p_A - p_B)}{\partial s}$	0.5001	0.5013	0.5019	**0.7188**	0.7522	0.7913	0.8235
π_0	0.0106	0.0559	0.0808	**0.1643**	0.2540	0.3788	0.5036
π_1	0.0050	0.0495	0.0742	**0.1579**	0.2484	0.3741	0.4996

CHAPTER 3. BRAND INERTIA IN SEED DEMAND: STATE DEPENDENCE AND HETEROGENEITY

Jinjing Luo,[1] GianCarlo Moschini,[1] and Edward D. Perry[2]

Department of Economics, Iowa State University, Ames IA 50010,[1] *Department of Agricultural Economics, Kansas State University, Manhattan KS 66506*[2]

Abstract

A commonly observed feature of differentiated product markets is brand inertia, the tendency of consumers to purchase brands they have purchased in the past. In this paper, we develop and estimate a micro-level random coefficients logit model to study two competing explanations of brand inertia, state dependence and heterogeneity, in the U.S. soybean seed industry. Specifically, heterogeneity is captured by brand-specific random coefficients and state dependence is incorporated through a brand purchase history variable. We further deal with two important identification issues: we apply a correction to the initial conditions problem by including variables that identify farmers' initial brand choices; and, to deal with price endogeneity, we use the control function approach. The model is estimated using a large dataset of more than 200,000 seed purchase decisions by roughly 28,000 farmers over the period 1996-2016. We find that state dependence and heterogeneity are both important features of seed demand. On average, farmers are willing to pay an additional $5.31/unit for a brand if it was purchased in the previous period, equivalent to about 12% of the average retail price. We also find that farmers are willing to pay large premiums for brand labels and the glyphosate tolerance (GT) technology, although they display considerable heterogeneity in these values. To investigate the implications of state dependence for farmers' dynamic purchase behavior, we

simulate two counterfactual scenarios in which we impose temporary shocks in the soybean seed market both with and without state dependence. In the first counterfactual, we simulate a temporary price discount for each brand, and in the second simulation we simulate a brand delaying the addition of the GT trait to their product line. These simulations show that when state dependence is present, temporary shocks have long-lasting effects on dynamic farmer behavior. The GT technology simulations, in particular, demonstrate that there are large and long-lasting brand loyalty penalties to not offering a new product innovation.

1. Introduction

The extent to which brand loyalty matters for demand has long been a motive of interest in economics and marketing (Bronnenberg, Dubé, and Moorthy 2019). In differentiated product markets, a well-established empirical regularity is brand inertia: individuals are more likely to purchase a brand if they have purchased it in the past. Among the potential behavioral explanations for this tendency, researchers have been particularly interested in the importance of *state dependence*, defined as the causal dependency of an individual's future choices on their current state (Heckman 1981; Dubé, Hitsch and Rossi 2010). A growing body of research has shown that the presence of state dependence can have important implications for the extent of market power and pricing behavior (Dubé, Hitsch and Rossi 2010), market structure (Dubé, Hitsch and Rossi 2009), the price effects of mergers (MacKay and Remer 2019), and the persistence of brand shares (Bronnenberg, Dhar, and Dubé 2009; Bronnenberg, Dubé, and Gentzkow 2012). Whereas the extant literature has focused mainly on the consumer-packaged goods (CPG) industry,[17] this paper investigates brand inertia and state dependence in the context of an important agricultural input market: the U.S. soybean seed industry.

[17] In addition to those mentioned above, other studies in the CPG industry include Keane (1997), Seetharaman and Chintagunta (1998), Seetharaman (2004), and Horsky, Misra, and Nelson (2006). Sudhir and Yang (2014) and Train

Over the last few decades, the seed industry has been characterized by considerable growth and consolidation (OECD 2018). Much of this has been driven by the development and rapid diffusion of genetically engineered (GE) crops. First introduced in the mid-1990s, GE varieties embedding herbicide tolerance and/or insect-resistance provided farmers with drastically new technological solutions for weed and pest management. As a result, GE crops were met with considerable success and now exceed 90% of planted U.S. acreage in corn, soybeans, and cotton (Barrow, Sexton, and Zilberman 2014). The commercialization of GE varieties required access to both GE traits and elite germplasm, the latter arising from decades of traditional breeding efforts. Whereas GE traits were overwhelmingly developed by one company (Monsanto), the ownership of germplasm was more dispersed. The highly complementary nature of these two building blocks (Graff, Rausser, and Small 2003) led to an early wave of acquisitions and mergers (Fernandez-Cornejo, 2004). Furthermore, the diffusion of GE crops was facilitated by Monsanto's aggressive licensing of GE traits to other seed suppliers, a contractual strategy that also benefited from a parallel major strengthening of intellectual property for plants (Clancy and Moschini, 2017).

Among the major U.S. crops, the soybean seed industry has perhaps undergone the largest transformation. The once common farming practices of saving harvested soybeans for seed use, and/or purchasing publicly developed varieties, have been replaced by the almost complete reliance on new proprietary commercial soybean varieties that embed the GE trait for glyphosate tolerance (GT).[18] An ongoing area of research has sought to assess the implications of these changes for the industry and the welfare of its main players: trait developers, seed

---

and Winston (2007) study the automobile industry, and Handel (2013) analyzes the health insurance industry.

[18] Consider, for example, that in 1970 about 70% of planted soybeans were public varieties (Fernandez-Cornejo, 2004). Based on the data used in this paper, by 2016 this fraction is less than 1%.

companies, and farmers. An essential ingredient for this research program is the estimation of seed demand. Beyond the assessment of the value of product innovation (Ciliberto, Moschini, and Perry 2019), a suitable seed demand model would permit the investigation of other questions of interest, including the exercise of market power, related antitrust concerns that may arise, and the role of brand loyalty.

In this paper, we develop and estimate a micro-level structural model of U.S. soybean seed demand. Specifically, we estimate a random coefficients logit model that allows for the presence of state dependence in farmers' preferences for brand labels. To estimate demand, we draw on a dataset containing more than 200,000 seed purchase decisions by roughly 28,000 U.S. soybean farmers during the 1996-2016 period. These unique data provide the requisite information on seed purchase histories, seed characteristics, and prices. In developing and estimating the model, our main objectives are to: (i) identify the dollar value of state dependence for brand labels in the soybean seed industry; (ii) investigate whether farmer heterogeneity is an important feature of the demand for brand labels and GE glyphosate tolerance; (iii) describe the economic features of farmers' seed demand through the derivation of WTP distribution and own-price and cross-price demand elasticities; iv) assess the implications of state dependence for farmers' purchase behavior.

The model we develop and estimate must address two important issues. The first issue concerns the identification of state dependence. The basic problem is that brand persistence or inertia (sometimes referred to as stickiness) can arise because of genuine state dependence *or* because of heterogeneity (Heckman 1981; Keane 1997). Heterogeneity describes the fact that individuals may simply have different, state-invariant, preferences for a brand. Failure to properly control for heterogeneity will tend to exaggerate the presence of state dependence. A related but distinct issue is the *initial conditions problem* (Heckman 1987; Arulampalam and Stewart 2009; Akay 2012; Simonov et al. 2019). This problem arises when the researcher does not observe an individual's entire purchase history and, if not properly accounted for, will also tend to exaggerate the extent of state dependence.

To control for heterogeneity, we permit farmers to have normally distributed preferences for all brands. To address the initial conditions problem, we apply a correction similar to the procedure outlined in Wooldridge (2005). In particular, we include brand-specific indicator variables that code for whether an individual purchased that brand in their first period of observation. Despite such control for heterogeneity, it is still possible to obtain spurious state dependence if the assumed distribution for heterogeneity deviates significantly from the true distribution. Thus, as a final check for whether we have identified genuine state dependence, we conduct a reshuffling procedure similar to Dubé, Hitsch and Rossi (2010). The basic idea of this procedure is to reshuffle each individual's choice sequence in a random way and then re-estimate the model. If structural state dependence remains, then this suggests that unobserved heterogeneity has not been sufficiently accounted for.

The second issue we face is the well-known problem of price endogeneity in demand models of differentiated products (Berry, 1994; Berry, Levinsohn, and Pakes 1995; Nevo 2000). Although our model is estimated using individual choices, which may alleviate some concerns about endogeneity (Goldberg 1995), there may still remain certain unobservable factors correlated with both the price and demand. The most common solution to this problem is to use two stage least squares (2sls) with instrumental variables (IVs). This approach, however, cannot be directly applied in non-linear individual-level discrete choice models (Train 2009). Therefore, we implement a control function approach, as outlined in Petrin and Train (2010) and

Wooldridge (2015). Much like 2SLS, this consists of running a first-stage regression of price on all model variables and a set of excluded IVs. We then compute the predicted residuals from this first-stage regression and include them as a control variable in the random coefficients logit model. For IVs, we use the previous year's soybean futures price interacted with brand and GT trait dummies. These IVs are in the spirit of the cost-brand interaction IVs used by Berto Villas-Boas (2007), and they exploit the fact that the previous year's futures price affects a seed firm's production costs.[19]

Overall, we find significant evidence of structural state dependence, even after controlling for persistent unobserved farm-level heterogeneity. On average, having a previous experience with a brand increases its value by about $5.31/unit of soybeans, equivalent to about 12% of the average price of $45/unit. This estimate is close to the dollar value of brand loyalty estimated for orange juice in Dubé, Hitsch and Rossi (2010).   We also find that, on average, farmers are willing to pay large premiums for brand labels and for GE traits, although this willingness to pay (WTP) can vary widely across farmers. For example, from 2011-2016, farmers' mean WTP for the GT trait was $23.38/unit, with 10% of farmers valuing GT at $41 or more and another 10% of farmers valuing it at $8 or less.

Using the model estimates, we assess some potential implications of state dependence for farmers' dynamic purchase behavior. In particular, we impose two types of temporary shocks in the soybean seed market and then simulate each farmers' choice probabilities over time with/without state dependence. The first shock we simulate is a temporary price discount for

[19] Seed firms contract with individual farmers to grow their future commercial seed supply and this is usually done in the region where the seed will eventually be sold (Lamkey 2004). A farmer's opportunity cost of growing seed for a company is what they could have obtained on the market. Thus, a higher futures price will tend increase production costs for seed firms. Note that by interacting the futures price with GT and brand dummies we aim to capture product-specific cost impacts.

each brand. The second shock we impose is to delay the addition of the GT trait to a brand's product line from 1996 to 1999. Our goal with this exercise is to understand the implications of not adding a new product technology when state dependence is an important determinant of demand. Overall, the results of these exercises highlight that, in the presence of state dependence, temporary shocks produce long-lasting demand effects, illuminating the rationale underlining dynamic pricing incentives for firms. The GT trait simulation, in particular, demonstrates the underlying source of what has been called the "pioneering advantage" in the previous literature (Bronnenberg, Dubé, and Moorthy (2019). Specifically, if a firm fails to offer a major new product or incorporate a new product characteristic quickly, it is penalized in *two* ways: (i) it incurs a loss in share from not offering that characteristic and (ii) it incurs a loss in share because of dynamic brand loyalty effects. This latter effect, which only materializes in markets with state dependence, can last for years, and perhaps explains why soybean seed firms were so quick to offer the GT trait in their product lines.

The analysis and results provided in this paper contribute to the literature in several ways. First, we provide new evidence on the *distribution* and heterogeneity of U.S. farmers' WTP for the major U.S. soybean seed brands, and for the GT trait, over time. This complements the results reported by Ciliberto, Moschini, and Perry (2019), whose WTP estimates for GE traits were not modeled to vary across farmers. The second contribution of this study is to provide dollar value estimates of brand loyalty in an important U.S. agricultural input market. While brand loyalty in related industries has been documented using survey and interview-based evidence (Kohls et. al., 1957; Funk and Vincent, 1978; Kool, 1994; Harbor, Martin, and Akridge, 2008; Sellars and Gunderson, 2018), there are no studies of brand inertia using revealed preference data in an agricultural context. Finally, we contribute to the implications of state dependence, especially when combined with the introduction of a major technology innovation. Whereas research in Bronnenberg, Dhar, and Dubé (2009) and Bronnenberg, Dubé, and Gentzkow (2012) find there is a large, persistent advantage to being the first brand in a particular *geographical* location, our exercise emphasizes the function of state dependence in a new dimension of the *product space*.

The rest of this paper is organized as follows. In Section 2, we provide background information on the U.S. soybean seed market. Section 3 presents the data used in the econometric regression. In Section 4, we develop the demand model, discuss the identification strategy, and present the estimation process. Section 5 presents the estimation results, followed by the implied WTP distributions and demand elasticities. Using simulation, some implications of state dependence are considered in Section 6. Section 7 concludes.

2. Soybean Seed Industry Background

The U.S. seed industry has grown considerably over the last few decades, fostered by sustained demand domestically and abroad.[20] Part of this growth has been driven by technological innovation, the result of significant research and development (R&D) investments, partly owing to the changing landscape of intellectual property rights. As the industry has grown it has also experienced considerable consolidation and rising seed prices (Fernadez-Cornejo 2004). A major development affecting seed markets, maize and soybeans in particular, has been the introduction of genetically engineered (GE) traits in the mid-1990s. By using breakthrough recombinant DNA techniques of modern biology, it became possible to integrate certain foreign genes (from bacteria) into the germplasm of elite crop varieties. These genes confer traits to the

[20] The size of the global commercial seed market was estimated at about USD 12 billion (ISAAA, 2016) in the United States and around USD 52 billion worldwide (Syngenta, 2016) in 2014.

resulting “transgenic” crops, such as herbicide tolerance and insect resistance, which are highly valued by growers (Moschini 2008).

The GE revolution in the seed industry also benefited from the general strengthening of intellectual property rights for biological innovations (Moschini 2010). This is particularly important for soybean seeds, the focus of this paper. Soybean varieties are self-pollinating, meaning that they reproduce true to type (unlike hybrid maize, for example). Thus, prior to the advent of GE varieties, farmers could rely on saved seeds (from the previous harvest) and could access, essentially at cost, varieties developed and released by public institutions (state universities). The introduction of patented GE traits, and the associated increased use of trade secrets and contracts, effectively permitted the industry to develop proprietary seed products (Clancy and Moschini 2017). This greatly increased the profitability of R&D in plant breeding, which lead to increased investments and an early wave of industry consolidation through mergers and acquisitions (Fernandez-Cornejo 2004). By the year 2000, the two largest firms (Monsanto and Dupont) accounted for about 40% of the U.S. soybean seed market, a combined share that has risen to about 60% in recent years.

Soybeans constitute the second most planted crop (after maize) in the United States. Unlike genetically-engineered corn varieties, which can have several traits—glyphosate tolerance (GT), corn borer resistance, root worm resistance, and their combinations—the only trait with major commercial relevance during our study period (1996-2016) has been glyphosate tolerance.[21] Glyphosate is a powerful, broad-spectrum herbicide used in combination with GT crops. It can kill approximately 99% of non-glyphosate resistant weeds without harming GT

[21] GE varieties tolerant to glufosinate did not achieve commercial relevance till very recently and, as in Ciliberto, Moschini, and Perry (2019), we do not distinguish between conventional and glufosinate tolerant varieties in our empirical analysis.

varieties (Wechsler, McFadden, and Smith, 2018). By reducing the need to use tillage, as well as multiple types of herbicides, GT varieties permit an extremely effective (and simplified) weed control strategy (Perry, Moschini, Hennessy, 2016). Because of this, GT soybeans were rapidly adopted: first commercially introduced in 1996, GT varieties accounted for more than 50% of the market by 1999, and more than 90% by 2007. This is despite the fact that GT soybeans command a significant price premium (Schenkelaars et al., 2011; OECD, 2018). Indeed, previous resarch has found that U.S. farmers' willingness to pay (WTP) for the GT trait far exceeds its cost, resulting in significant net economic gains, especially in recent years (Ciliberto, Moschini, and Perry, 2019; Shi, Chavas, and Stiegert, 2010; Fernandez-Cornejo, Hendricks, and Mishra, 2005).

The marketing of seed varieties relies heavily on long-standing and well-known brands such as Pioneer and Asgrow. In addition, several brands can be marketed by the same parent company. For example, the company Dupont has primarily sold varieties under the Pioneer brand, whereas Monsanto has marketed varieties under several brands such as Asgrow, DeKalb, and Channel. Each brand typically offers multiple distinct varieties that differ in characteristics such as glyphosate tolerance, soybean cyst nematode resistance, relative maturity, and tolerance to iron deficiency chlorosis. Most brands currently market both conventional and GT varieties.[22]

Average market shares for the 13 largest brands over the considered timespan are reported in Table 1. Brands are grouped by their well-known parent companies such as Monsanto, DuPont, Syngenta, and Dow AgroSciences.[23] We separately categorize public seeds

---

[22] Two exceptions are Channel, a Monsanto brand, which entered the market in 2010 and only offers GT varieties, and public providers (mainly state university programs) who offer only conventional varieties (See Fernandez-Cornejo (2004), p. 36, for a list of major public breeders).

[23] The company names in Table 1 reflect the industry configuration as of 2016, the last year of our data. Since then, major mergers and acquisitions are re-shaping the ownership structure of the industry—the acquisition by

and the seeds sold by brands not owned by the big four parent companies.[24] Table 1 also illustrates some turnover in brands. Golden Harvest was phased out by Syngenta is 2012 and, in recent years, DeKalb is also being phased out (Monsanto is focusing this brand mostly in the maize seed market).

3. Data

The data used in this paper pertain to seed purchases by a large and representative sample of U.S. soybean farmers. This data is drawn from a proprietary dataset assembled by Kynetec USA, a market research company that specializes in the collection of survey data in U.S. agriculture. The data span the 21-year period 1996-2016. For each year, the seed purchases of more than 3,500 soybean farmers are recorded. The sample itself is constructed to be representative at the crop reporting district (CRD) level.25 Each soybean farmer in the sample is observed to make one or more seed purchases and the data contains detailed information on the nature of the purchase and the variety (e.g., variety name, brand, parent company, GE traits, price, amount of seed, acres planted). Although this is not a balanced panel data set, a large portion of farmers are observed over multiple years (and multiple purchases are observed in the same year). As discussed further below, multiple observations per farmer are essential for

ChemChina of Syngenta in Apr 2017, the merger of Dow and DuPont in Sep 2017, and the acquisition by Bayer of Monsanto in June 2018. The agricultural concerns consolidated by the Dow-Dupont mergers were subsequently spun off as Corteva in 2019.

[24] The ownership of each brand as reported in Table 1 also pertains to 2016, the last year of our data. Brands' affiliation with their parent company in some cases was the result of market consolidation that took place earlier in our sample. This is particularly true for Monsanto, who acquired Asgrow in 1997, DeKalb in 1998, Channel in 2004, and Kruger in 2006. Also, DuPont acquired Pioneer in 1999; Syngenta acquired NK in 2000 and Golden Harvest in 2004; Dow acquired Mycogen in 1998.

[25] CRDs are regions identified by National Agricultural Statistics Service of the U.S. department of Agriculture (USDA). Each U.S. state comprises several CRDs, and each CRD includes multiple counties.

identifying/disentangling the elements of state dependence and time-invariant unobserved heterogeneity as drivers of farmers' purchase decisions.

3.1. Products

As noted, in this paper we develop and estimate a farm-level discrete choice model of soybean seed demand. An essential ingredient for this model is the definition of a "product." The product definition (or product space) partially defines farmers' choice sets, which include all possible alternatives that farmers may choose from. The value of modern soybean seed varieties primarily derives from two complementary sources: germplasm (i.e., the underlying genetics accumulated from past generations of selective breeding) and GE traits. The finest possible definition of a product would be in terms of individual varieties. For several reasons, however, analysis at the variety level is not feasible in our context. First, there are simply too many varieties26; the implied choice-set by a variety-level product definition would be too large to be estimated with a farm-level mixed logit model. Furthermore, individual varieties have limited geographic presence, as each is bred to be best suited to specific agro-climatic conditions (e.g., latitude). In addition, new varieties are introduced every year, and the life cycle of any given variety is relatively short (four to five years, on average).

It is perhaps more helpful to think of varieties as forming "product lines" over time, as companies introduce improved new varieties that are embedded and built on the genetics of previous varieties. We presume that this continuity is captured by the "brand" (e.g., Asgrow). Varieties marketed by any one brand at different locations may differ, even considerably, but in any one local market one can expect varieties of the same brand to share common characteristics. Hence, we choose to define products by brand, and by whether or not it includes the GT trait.

[26] There are totally 18,420 varieties in our soybean dataset.

Specifically, to make the farmers' choice set of the model tractable, but still include as many alternatives as possible, we rely on the 13 distinct brands illustrated in Table 1. Note that we treat the public/university seeds offered by all public sectors as a single brand, named Public. These 13 brands account for about 70% of the US soybean seed market, over the period analyzed. All remaining varieties are aggregated into an "Others" group. To account for the important role played by the GT trait, each brand can be associated with two products, depending on whether or not it embed the GT trait. Because Channel only provides GT seeds and Public only offers conventional seeds, we have thus identified a total of 26 distinct products. In any one choice situation, however, a farmer may not have access to all such alternatives. To be more specific about that, we next discuss the definition of "market" used in this study.

3.2 Markets and choice sets

In our model of individual choices, a market is a time-specific location where residing farmers face the same choice set. Following Ciliberto, Moschini, and Perry (2019) we define a market as a CRD-year combination. As noted, CRDs are multi-county sub-state regions identified by the USDA. This market definition is similar to the CRD-level aggregation used in market analyses by some of the major seed companies. Differentiating markets by years is a natural extension, as commercialized varieties evolve over time, and a calendar year contains a natural planting window. In our dataset, we observe a total of 3,791 markets across 233 CRDs. The number of markets in selected years, and the average number of choice alternatives (i.e., products) available to farmers, are provided in Table 2.

As noted, a farmer's choice set can contain at most 26 alternatives. The availability of a product in a market is identified by the existence of at least one purchase record. Thus, farmers residing in the same market share the same choice set. Note that, following the introduction of

GT varieties, the number of products available to farmers initially increased, but eventually decreased as GT products crowded out conventional products.[27]

3.3 Prices

A common challenge in discrete choice models of individual choices is the construction of prices with transaction data only. The basic problem is that we observe the price for the alternative actually chosen by the individual, but we do not observe the prices of the unchosen alternatives. A typical solution to this problem is to compute average transaction prices and use these prices as the prices that individuals face for each alternative. For example, Goldberg (1995) , in her nested logit model of household automobile vehicles demand, uses the market-level transaction price (net price). On the other hand, Train and Winston (2007) use retail prices in a mixed logit model to study the declining market share of U.S. automakers over time. They argue that, although discounts are common, there seems to be little difference between the discounts offered by American, Japanese, and European manufacturers. In a study of consumer choice behavior in consumer packaged goods markets, where discounts are not common, Keane (1997) uses retail prices, and notes the potential for price endogeneity when using net prices.

In our setting, discounts are common feature of the seed purchasing process—in our dataset, about 63% of 204,697 total observed purchases have a discount. A major reason is the timing of a purchase: farmers who buy earlier are often rewarded with a discount on the listed price.[28]

---

[27] Similar to Train and Winston (2007), the model we develop is a conditional demand model—only soybean seed choices are considered, conditional on the farmer having chosen to plant soybeans on a given plot (i.e., there is no "outside option"). Furthermore, as discussed further beliow, we focus only on new seed purchses (observations where a farmer uses saved seeds are dropped).

[28] The probability of getting a discount is around 30% if a farmer orders their seed in March, April, May, and June, whereas the probability is about 80% if she orders before January. In a logit regression of whether or not the farmer gets a discount, we find that the probability of getting a discount is highest in August and decreases in this order: August, September, October, November, July, December, January, February, March, April, May, June. Additionally,

To account for discounts in the construction of prices for each alternative, we adopt what we term a “contingent” price system. First, in each market, and for any given product, we create two prices: a retail price (the average of all observed retail prices for each product in each market), and a net price (the average of prices after netting out the observed discount in each market). For each farmer, we then identify whether they received a discount. If they did receive a discount, we set the unchosen alternatives’ prices to the net prices, and if they did not receive a discount, we set the unchosen alternatives’ prices to the retail prices. For the chosen alternative, we assume the farmer faced the price she actually paid, inclusive of the discount, if she received one.

Our goal with this method is to capture the unobserved factors that contributed to the farmer obtaining a discount for the observed choice (for instance, a farmer able to purchase seeds early, and observed to obtain a discount for her seed choice, most likely would have been able to obtain a discount for the other alternatives available to her in the market). We also note that our use of the observed net price for the purchased product (rather than the market price) fits the nature of the problem at hand. Unlike the case of consumer packaged goods, where shelf prices are typically common to all consumers, seed prices are typically negotiated between the farmer and the seller (e.g., dealers or seed companies’ representatives).

Finally, we note that the dataset spans 21 years, a long period during which prices changed considerably. Consistent with the homogeneity property of the per-acre profit function, described in what follows, we express all prices in real terms by deflating them by the USDA crop sector index of prices paid.[29]

---

the planting season of soybeans in the United States lasts from June to the end of October (Syngenta 2016).

[29] The Crop Sector Index is published by USDA-NASS Quick Stats. This index takes 2011 as the base year.

3.4 Inertia

A major focus of our analysis concerns the possible presence of state dependence in farmers' seed choices. To motivate this perspective, following Dubé, Hitsch, and Rossi (2010), Table 3 reports the purchase rate and repurchase rate of each brand, presented as percentages. The purchase rates are the unconditional probability of choosing each brand, calculated as the market shares over the full time period. Conditional on the previous choice, the repurchase rates show the probability of purchasing the same brand again.[30] Note also that the purchase records of a given farmer may not enter the sample in consecutive years. When this occurs, we use the most recent period's records. It is apparent from Table 3 that the repurchase rates are considerably higher than the corresponding purchase rate. In some case, the ratio of the repurchase rate to the purchase rate is extremely high. For example, only 1.28 percent of purchases were for the brand Growmark. Yet, conditional on buying Growmark, an individual had a nearly 76% probability of purchasing it again in the next period. These data are synonymous with persistence or inertia in brand choices over time. Of course, as previously noted, this could because individuals have heterogeneous preferences or because of state dependence.

3.5 Purchase sources

The distribution of seed varieties to farmers is highly localized and typically run by independent agents, such as farmer-dealers, farmers' cooperatives, company salespeople, and private wholesalers and retailers (Fernandez-Cornejo 2004). For large farm operations, seed companies may also sell directly to farmers through their sales representatives (Fernandez-Cornejo, 2004; Syngenta, 2016). The structure of this distribution system is illustrated in Figure 1. Based on this structure, we classify all observed purchases into the three sources that capture

---

[30] In forming Table 3, we use the data of "Choice in regression" as discussed in detail in section 5.1.

the main differences in seed marketing: “sales representative”, “dealer”, and “retailer”. Specifically, a purchase is designated as being from a sales representative if the farmer purchased the seed product directly from the seed company or their representative; it is classified as coming from a dealer if the farmer obtained their seed from a farmer dealer, an independent seed dealer, or if the farmer herself is a dealer; it is classified as coming from a retailer if the products came from any other source, including cooperatives, seed retailers, and grain elevators (seed retailers account for the majority).[31] In our data, 46% of purchases are made from “dealer”, 35% are made from “retail”, and 19% from “sales representative.”

In the econometric model, estimated below, we use the records of purchase sources to generate product-specific marketing variables. These variables potentially control for important sources of unobserved heterogeneity induced by marketing activities.

4. Model Specification

We develop a seed demand model under the presumption that farmers, on each of their plots, choose the preferred seed alternative to maximize expected profit. The structure of this payoff function depends on the production technology, the prices of output and all other inputs, and, of course, the price of the seed. Ciliberto, Moschini, and Perry (2019) show that, when the production function satisfies two reasonable properties—constant returns to scale in all inputs, and a fixed proportion between land and seed—per-acre expected profit is linear in the (suitably normalized) seed price, i.e.,

$$\pi_{ij} = \pi_{ij}(r,w) - \alpha p_j . \tag{23}$$

[31] Retailers differ from dealers in that they typically sell other farm inputs (e.g., fertilizers and pesticides).

Here i indexes the plot to be planted (i.e., the choice situation in what follows), r is the expected price of the output to be produced on this plot, w is the price vector of all inputs used in production (except land, meaning that the profit in (23) can be interpreted as returns to land), p_j is the price of seed alternative *j*, and the parameter α captures the (constant) amount of seed per acre (i.e., the seed density).[32]

Given this objective function, the problem for farmer h, in market m, on plot i, can be stated as that of choosing product j such that

$$\max_j \ \pi_{ij}^h, \ j \in \{1, \ldots, J_m\}, \tag{24}$$

where J_m is the number of available products in the market m pertaining plot i.

4.1 The econometric model

To make this framework operational, we need to parameterize the profit function. In addition to the seed price, which enters linearly, we approximate the other structural determinants (e.g., output and input prices) of the per-acre profit function by a set of seed, market, and farmer-specific variables, in addition to the inertia variable that captures state dependence. Specifically, the per-acre profits for farmer h from product j, for choice situation i in the corresponding market m, are

$$\pi_{ij}^h = \boldsymbol{\beta}_h' \boldsymbol{x}_{jm}^h - \alpha_h p_{jm}^h + \gamma_h I_{bt}^h + \nu_{ij} + \varepsilon_{ij} \tag{25}$$

[32] Note that Ciliberto, Moschini, and Perry (2019) express seed prices on a per-acre basis. Here, however, we express seed prices on the typical per-unit measure used in the industry (i.e., per “bag,” each containing 160,000 kernels). Given the assumed fixed proportion between land and seed, the choice of units is immaterial. To translate one into the other note that, in our data, one acre of land on average uses 1.186 units of seed.

In this equation, $\boldsymbol{x}_{jm}^{h}$ is a vector of seed characteristics possibly interacted with farmer or market-specific characteristics (the primary seed characteristics include the brand and GT trait), and $\boldsymbol{\beta}_{h}$ is a vector of coefficients that capture the impact of each variable in $\boldsymbol{x}_{jm}^{h}$. To capture unobserved heterogeneity, we allow farmers to have heterogeneous propensities preferences over brands and the GT trait, which are a subset of the variables in $\boldsymbol{x}_{jm}^{h}$. In other words, we permit a subset of $\boldsymbol{\beta}_{h}$—the coefficients of the variables just mentioned—to be random across farmers. The variable p_{mj}^{h} is the price of product j in market m for farmer h, and thus the coefficient α_{h} represents the impact of price on per-acre returns. Note that to simplify the notation in equation (25), we have not explicitly written that the farmer (h) and market (m) are uniquely identified by the choice situation (i), the brand (b) is uniquely identified by the product (j), and time (t) is uniquely identified by market (m), which is further identified by choice situation (i) as mentioned.[33]

Structural state dependence is captured by the indicator variable I_{bt}^{h}, defined as follows:

$$I_{bt}^{h} \equiv I\left\{b \in s(b)_{t-1}^{h}\right\},$$

where $s(b)_{t-1}^{h}$ is the set of brands that have been purchased by farmer h in the previous period $t-1$. The indicator variable I_{bt}^{h} takes value one if the brand associated with alternative j is in set $s(b)_{t-1}^{h}$. As noted, because the panel is unbalanced, in some cases we do not observe a farmer in consecutive periods. When this occurs, we use the most recent year in which the farmer was observed. Finally, ν_{ij} and ε_{ij} are residuals that capture any remaining unobserved variation in

33 One way to express this explicitly is to write superscripts/subscripts as $m[i]$, $h[i]$, $b[i]$, and $t[i]$.

profits. We assume that the residual v_{ij} is normal and correlated with price and that ε_{ij} is i.i.d extreme value. The fact that v_{ij} is correlated with price is synonymous with the well-known problem of price endogeneity. If this residual is not controlled for in estimation, then the estimated impact of α will be biased.

4.2 Identification

This section discusses two main issues in the identification of the empirical model: the endogeneity of seed prices and how to disentangle state dependence from unobserved heterogeneity. In section 4.2.1, we discuss potential sources of price endogeneity and introduce the control function method as a solution for this endogeneity. In section 4.2.2, using simple examples, we articulate how our model and data separately identify state dependence and heterogeneity.

4.2.1 Price endogeneity and the control function method

Price endogeneity is a common issue in the empirical industrial organization literature. At the product-market level, the basic problem is that there are unobserved factors correlated with demand. If firms account for this unobserved shocks in setting their prices, then the estimated price impacts will be biased. One partial solution to this issue is to include product fixed effects (Nevo 2000), however, there is still the likely possibility that there are product-location-time specific unobserved demand shocks that are correlated with prices. As a result, an estimation procedure that utilizes instrumental variables is usually required.

Even though we deal with individual choices and micro data, our model suffers from price endogeneity from several potential sources. Unobserved product attributes is a main source. We add product fixed effects (brand and trait intercepts) and further interact the terms with time. However, any location-specific and/or farmer-specific unobserved attributes may still cause price endogeneity. Moreover, price endogeneity can come from our construction of prices that account for discounts, which are farmer-specific. For example, some farmers may have better relationships with their dealers, which could result in pricing behavior that takes into account a particular farmer's preferences.

In the extant literature, the most common approach to dealing price endogeneity is the "BLP" approach (Train 2009). Most studies that apply this approach, however, use an aggregate discrete choice demand model. The "BLP" approach can still be applied with individual-level data; examples include Berry, Levinsohn, and Pakes (2004), Goolsbee and Petrin (2004), and Train and Winston (2007), but there are two major limitations to using it. First, there are often complications with the contraction part of the algorithm, such as non-convergence (Train 2009). Second, and more importantly, this approach does not control for endogeneity at the farmer level. An alternative approach that resolves both of these issues is the control function approach (Wooldridge 2015; Petrin and Train 2010). Loosely speaking, the control function approach is similar in nature to 2SLS, but can be applied to non-linear models, and is computationally less difficult. Given these benefits, and the fact that we use micro data, we take the control function approach to address price endogeneity. The specifics of this approach are as follows.

Recall that we assume that v_{ij} is correlated with price and that ε_{ij} is i.i.d extreme value. We assume that prices are determined as follows:

$$p_{mj}^{h} = \boldsymbol{z}'_{ij}\boldsymbol{\theta} + \mu_{ij}, \tag{26}$$

where z_{ij} includes all variables in equation (25) plus a set of excluded IVs. The residuals μ_{ij}, together with v_{ij} in equation (25), is specified as jointly normally distributed. With these assumptions, the per acre profit function can be re-written as:

$$\pi_{ij} = \boldsymbol{\beta}_h' \boldsymbol{x}_{jm}^h - \alpha_h p_{jm}^h + \gamma_h I_{bt}^h + \lambda u_{ij} + \varepsilon_{ij}. \tag{27}$$

where the distribution of ε_{ij} is still i.i.d extreme value and v_{ij} has been replaced with $\lambda\mu_{ij}$, with the non-correlated component of v_{ij} asborbed into β_h, which includes normally distributed brand and trait specific random components. In terms of estimation, we first estimate equation (26) and collect the predicted residusals $\hat{\mu}_{ij}$. These residuals are then included as a control variable in the model.

For the IVs, we exploit the fact that soybean seed firms contract out with individual farmers to grow their commercial seed supply for the following year (Lamkey 2004). The terms of contract are set such that the farmer is paid at least what they could have obtained had they planted and sold their own soybeans. This payment will therefore vary in response to changes in expected soybean output prices.[34] A standard proxy for a commodity's expected output price is the futures price corresponding to delivery in the month following the coming season's harvest. Given this, we use the *previous year's* soybean futures price as an instrument for the current year's seed prices. This IV is not only be highly correlated with costs, for the reasons just noted, but will also not affect farmers' relative demand for soybean seed products; i.e., it fulfills the exclusion restriction requirement.[35] To allow for variation across products, we interact futures prices with the brand and GT trait dummies. This is similar to the approach taken by Berto Villas

---

[34] If a seed firm buys its own land and uses that land to grow its commercial supply, the same logic still applies. Fluctuations in expected output prices will change the rental price of land and, therefore, the opportunity cost of seed production.

[35] If the previous year's futures price is correlated with the current year's futures price, it may correlate with a farmers' decision of *which crop* to plant. However, recall that the model we estimate is *conditional* soybean demand model. Thus, there is little reason to think the previous futures price correlates with the demand for particular soybean product. Moreover, we include time specific variables for both brands and GE traits, which should capture any impact of future prices on relative demand.

Boas (2007): in her paper on vertical integration in the yogurt market, she creates a set of IVs equal to the interaction of input costs with brand dummies.

4.2.2 State dependence and heterogeneity

Before proceeding to the estimation procedure, we informally discuss the intuition of how the model and data separately identify heterogeneity and state dependence. The primary source of identification for the heterogeneity parameters is the panel aspect of the data; i.e., the fact that we observe and utilize each individual's sequence of choices. Consider the following simple example of two farmers over the course of four years. In each year, each farmer makes a single choice between brand A and brand B. Let us suppose that in the pooled sample each brand has 50% market share. There are two extreme ways by which this could happen: (i) every farmer could purchase each brand 50% of the time; or (ii) %50 of farmers could *always* choose A and 50% could *always* choose B. In case (i), the estimated variance parameter would be zero, as all farmers are equally likely to purchase A or B; in case (ii), the estimated variance or heterogeneity would be very large. Thus, in the model we estimate, the heterogeneity parameters are pinned down by the degree to which the brand shares for each individual differ from the brand shares in the pooled sample. The more they differ, the greater the heterogeneity.

The state dependence coefficient, in particular, is identified by the *ordering* of brand choices over time for each individual. Let us return to the previous example. Suppose each individual chooses each brand twice, so there is little space for heterogeneity. State dependence is identified by the number switches; the fewer the number of switches, the greater the estimated coefficient for state dependence. In this example, the sequence AABB is associated with the highest level of state dependence, comparing to other possible sequences like ABBA (two switches) and ABAB (three switches).

Implicit in the previous example is that we need to observe some switching between the brands. Without any switching there is no way to distinguish between state dependence and heterogeneity. Returning to the previous example, suppose there are two farmers, one with the purchase sequence of AAAA and one with BBBB. These choice sequences could be purely the result of heterogeneity or purely the result of state dependence.

Moreover, even supposing we do observe brand switching, for that switching to identify state dependence it needs to be the result of *choice set variation.* Suppose there is a shock in the third period, for example brand B introduces GT, so the first farmer switches to brand B. If in the fourth period brand A adopts the same variety, state dependence is identified if the first farmer still purchases brand B, whereas the choice consistency should be ascribed to heterogeneity if she switches back to brand A. The case may be too specific in the real world, but it shows the principle: state dependence is identified if the unconditional (of choice set variation) probability of choosing the same brand increases after switch. The extent of state dependence is better measured with rich choice set variation: if the first farmer sticks to brand A with a relative small variation, but switches to brand B with a relative large variation.

Examples of choice set variation used in the prior literature include changes in price, advertising, and the availability of alternatives (Sudhir and Yang 2014). In our context, there are two primary sources of choice set variation. The first source is seed price variation. Relative seed prices flucuate from year to year as different brands try to attract new customers. These fluctations can be in the form of explicit discounts or due to changes in base prices. The second source of variation is changes in product attributes and the availability of alternatives. In particular, the GT trait was not added to all brands at the same time and in the same locations. Moreover, certain brands phased out there conventional varieties faster than others. These changes will have resulted in some farmers either switching to a new brand or trying a new brand. A related source of variation is in the nature of a brand. Seed varieties have relatively short commercial life-cycles. For example, the set of varieties offered under the Asgrow brand in 2000 were quite different from the set of varieties offered in 2010. Thus, for farmers not loyal or partial to Asgrow, there will be ongoing uncertainty about the quality of the brand. From time to time, therefore, such farmers may experiment with a brand like Asgrow to obtain information about it.

4.3 Estimation and the initial conditions problem

The model is estimated using simulated maximum likelihood, as outlined in Hole (2007) and Train (2009). The latent profit function, originally defined in (25), can be written more succinctly as

$$\pi_{ij} = \boldsymbol{\varphi}_{jm}^{h} + \varepsilon_{ij} \tag{28}$$

where $\boldsymbol{\varphi}_{jm}^{h}$ includes all components except the IID extreme value error term:

$$\boldsymbol{\varphi}_{jm}^{h} = \boldsymbol{\beta}_h' \boldsymbol{x}_{jm}^{h} - \alpha_h p_{jm}^{h} + \gamma_h I_{bt}^{h} + \lambda \mu_{ij} \tag{29}$$

For a *given* realization of $\boldsymbol{\varphi}_{jm}^{h}$, the probability that farmer h chooses alternative j in choice situation i is given by the familiar logit expression

$$L_{ij}\left(\boldsymbol{\theta}_h, s_{t-1}^{h}\right) = \frac{\exp\left(\boldsymbol{\varphi}_{jm}^{h}\right)}{\sum_{n=1}^{J^m} \exp\left(\boldsymbol{\varphi}_{nm}^{h}\right)} \tag{30}$$

where $\boldsymbol{\theta}_h = \{\beta_h, \alpha_h, \gamma_h, \lambda\}$ is the vector of coefficients to be conditioned on. J^m is the number of alternatives in farmer h's choice set and is market-specific in our setting. For each farmer, we observe a *sequence* of choices. Conditional on the initial state, the probability of the sequence for farmer h is given by the product of the logits conditional on $\boldsymbol{\theta}_h$ and s_0^h:

$$L^h\left(\boldsymbol{\theta}_h, s_0^h\right) = \prod_{i=1}^{I^h} L_{ij}\left(\boldsymbol{\theta}_h, s_{t-1}^h\right) \tag{31}$$

where I^h indexes the number of observed choice situations of farmer h, and the set $\left\{1,2,\ldots,I^h\right\}$ represents farmer h's choice sequence (on average, a farmer makes 23.6 purchases overall and 4 purchases per year).

We need to integrate out the random elements in $\boldsymbol{\theta}_h$ to obtain the unconditional probability of an individual's choice sequence. However, as noted by Heckman (1987), s_0^h is not exogenous and is in fact stochastically dependent on individual heterogeneous preferences. The correlation between $\boldsymbol{\theta}_h$ and s_0^h leads to the *initial conditions problem*. If not properly handled, the initial conditions problem leads to exaggerated state dependence coefficient.

To deal with the initial conditions problem, we adopt Wooldridge (2005)'s solution for its computational simplicity and good performance. The method imposes distributional assumption on unobserved heterogeneity conditional on the initial state s_0^h. Wooldridge (2005) stated that if the distribution is corrected specified, the resulting conditional MLE is consistent and asymptotically normal. Using Monto Carlo simulation, Akay (2012) and Arulampalam and Steward (2009) show that Wooldridge's method is very effective for panels longer than 5 periods.

When applying the Wooldridge method in our multinomial choice setting, we make similar assumptions on the coefficients of brand-specific intercepts. We denote the coefficients $\boldsymbol{b}_h$ to better explain the process. As we allow farmers to have heterogeneous propensities for each brand, $\boldsymbol{b}_h$ is a subset of β_h. We assume $\boldsymbol{b}_h \mid s_0^h \sim Normal\left(\boldsymbol{b}_0 + \boldsymbol{b}_1 \bullet s_0^h, \Omega_b\right)$, which means that conditional on every farmer's initial brand choices, their propensities for each brand are normally distributed. Ω_b is the variance-covariance matrix and assumed to be a diagonal matrix. Following Wooldridge (2005), we can further write

$$\boldsymbol{b}_h = \boldsymbol{b}_1 \bullet s_0^h + \boldsymbol{e}_h$$

where $\boldsymbol{e}_h \mid s_0^h \sim Normal\left(\boldsymbol{b}_0, \Omega_b\right)$, so $\boldsymbol{b}_0$ is absorbed by $\boldsymbol{e}_h$. Remind that s_0^h is a set of brands chosen by farmer h in his/her first observed period. In the above equation, we consider s_0^h as a vector of indicator variables for each brand, 0 if the brand is not chosen and 1 if it is chosen in the initial period. $\boldsymbol{b}_1 \bullet s_0^h$ is the inner product of the two vectors with the same length. Because $\boldsymbol{b}_h$ enters model (25) linearly, s_0^h enters the model linearly. Same as $\boldsymbol{b}_h$, s_0^h will interact with the brand intercepts. As a result, we can consider $\boldsymbol{b}_1$ as coefficients of the variables representing whether the farmer chose the brand of the alternative in the initial period.

Under the assumption, the unconditional probability for individual h's purchase sequence is

$$L^h = \int_{\mathbb{R}^c} L^h\left(\boldsymbol{\theta}, s_0^h\right)\Phi\left(\boldsymbol{\theta} \mid s_0^h\right) d\boldsymbol{\theta} \tag{32}$$

$\Phi(\;)$ is the joint distribution of all random components in $\boldsymbol{\theta}$, consisting of the random coefficients in $\boldsymbol{\beta}_h$, α_h, γ_h, and $\boldsymbol{e}_h$ as discussed above (the probability of any fixed components in $\boldsymbol{\theta}$ is 1, so $\Phi(\;)$ degenerates to the joint distribution of all random components). We assume normal distribution for all the parameters except α that is assumed to be lognormal distributed.

We will discuss more details about the presumed distribution of the price coefficient in section 5.6.2.

The probability is simulated using 250 Halton draws for any given value of the means and variances of the random components in $\boldsymbol{\theta}$.[36] Thus,

$$\tilde{L}^h = \frac{1}{D}\sum_{d=1}^{D} L^h\left(\boldsymbol{\theta}_d, s_0^h\right) \tag{33}$$

where D is the number of draws and is equal to 250 (d indexes each draw). $\boldsymbol{\theta}_d$ contains the d th draw of $\boldsymbol{\theta}$ from $\Phi\left(\boldsymbol{\theta} \middle| s_0^h\right)$. As noted by Train (2009), equation (33) is an unbiased estimator of L^h by construction. The log-likelihood for the model is

$$\ln L = \sum\nolimits_{h=1}^{H} \ln \tilde{L}^h \tag{34}$$

The parameters are estimated by maximizing the log-likelihood in STATA using 250 Halton draws. Specifically, we are utilize the user written “mixlogit” package by Hole (2007), outlined in Cameron and Trivedi (2005).

5. Model Results

In this section, we first discuss our preparation of the estimation data in section 5.1, then elaborate model variables in the regression and present some basic summary statistics of the variables in section 5.2. In section 5.3, we show estimation results for the basic conditional logit model and the mixed logit model of different settings. In section 5.4, we compute and discuss WTP distributions for the main variables of interest using the coefficient estimates from the full

36 Halton draws are used to approximate the distribution of random coefficients. Because there is no closed form expression for equation (32), we simulate the equation by taking the mean of Halton draws. As noted in Train & Winston (2007), Train (2009), Petrin & Train (2010), 100 Halton draws is more efficient than 1000 random draws and 250 Halton draws is sufficient for simulation.

mixed logit model. We then compute mean own-price and cross-price demand elasticities of each product in section 5.5. Finally, in section 5.6, we consider the robustness of our regression. First, we examine the state dependence coefficient by conducting a reshuffling procedure along the lines of Dubé, Hitch, and Rossi (2010). Secondly, we discuss the WTP estimates from different settings of the price coefficient.

5.1 Data preparation and summary statistics

In total, the dataset contains 213,062 seed purchase records for 28,017 farmers. We clean and reformat the dataset for estimation of the conditional and mixed logit regressions following the steps listed in Table 4.

We drop all cases in which a farmer did not purchase a new soybean variety. These cases include the following purchase classifications: “From my own farm”, “I’m a seed grower”, or “New seed that was left over from last year.” Next, because public varieties have not included the GT trait, we assume the small number of cases in which they were associated with the GT trait was an error, and therefore drop these observations (108 in total). This leaves 204,697 purchase records, which are termed “Choices in analysis” in Table 4. We highlight this number and this step because this is the dataset that provides all information used in the empirical analysis of this paper. However, the first three years of data (1996-98) are exclusively used to build farmers’ purchase history. In addition, the first year a farmer appears in the data (for many farmers this is after 1998) is used to create the “state dependence” variable, and thus such observations are not use in the logit regressions. Again, however, these records still enter the model through certain explanatory variables, such as the initial brand choices, the state dependence terms, and the marketing variables. Moreover, for robustness, we drop purchase records of farmers who appear in the sample only three or less year. The end result of this process yields 90,264 purchase records that provide the estimation data for the logit regressions (termed “Choices in regression” in Table 4). For the model to be estimated, the data needs to be further expanded such that, for each individual choice, there is also a row of information for each unchosen alternative in the corresponding market; this expansion results in 1,057,637 “observations”. Descriptive statistics, calculated from the dataset of “Choices in analysis”, are provided in the previous Tables 2 and 3, whereas Table 5 describes the dataset of “Choices in regression.”

Table 5 shows that a farmer generally chooses about two brands (and also two products) per year, and the number of chosen brands doubles for the whole observed period of a farmer, suggesting some brand switching, which aids identification of structural state dependence (discussed in section 4.2.2). Table 5 also demonstrates that, on average, a farmer chooses among 11.7 alternatives, with a minimum of two (note that we drop records in markets with only one alternative) and a maximum of 23 (recall that our product definition results in 26 possible alternatives). For a typical farmer, we observe about 8 years of data (recall the minimum is four, as discussed in the foregoing, and the maximum is 21, i.e., a farmer appears in the sample in every year over the 1996-2016 period).

5.2 Model variables

Recall that the regression model includes one primary set of explanatory variables, represented by $\boldsymbol{x}_{jm}^{h}$. Within the vector $\boldsymbol{x}_{jm}^{h}$, there are three types of variables: (i) a set of brand and trait intercepts, each interacting with time; (ii) initial condition variables, and (iii) a set of marketing variables and their interactions with individual-specific purchasing experiences. Each of these three types of variables constitute 41, 14, and 6 variables, respectively. Further details are as follows.

In the first set of variables in x_{jm}^{h}, the brand and trait intercepts capture the average profit gains of each brand and GT trait. In the estimation, we classify our regression timespan 1999-2016 into three periods 1999-2004, 2005-2010, 2011-2016 with the same length and interact the three periods with brand and trait intercepts to account for any time-variant effect. In the model they are coded as GT 1999-2004, GT 2005-2010, GT 2011-2016, and “brand”, “brand” 1999-2004, “brand” 2005-2010. “brand” refers to the brands listed in Table 1, except Public, the baseline brand, and the indicator variable for Public is dropped from regression to avoid collinearity; we further take “brand” 2011-2016 as the baseline. For time-variant brand effects, they can be caused by introduction of different new varieties over time, or any brand-specific changes that happen nationwide. The time-variant trait effect may come from the commercialization of new GT patent, the emergence of glyphosate-tolerant weed, etc; likewise, we only capture the national effect rather than the CRD-specific effect. In the logit regression, case-specific variables, like time or CRD region, will not affect the choice decision if they influence all alternatives in the same way, so they come in the model by interacting with alternative-specific brand or trait intercepts. In our case, time-variant brand effects are explicitly model, whereas CRD-variant brand or trait effects go to the error terms. There are 41 variables in this set of variables, two periods for Channel (Channel entered the market after 2005), three periods for GT trait and other brands except Public (2+3+3*12=41).

The second set of variables in x_{jm}^{h} capture the initial conditions of a farmers’ choice sequence. Following the idea of Wooldridge (2005), we include initial states as extra explanatory variables to account for the initial conditions problem of correlation between unobserved heterogeneity and the initial state. By further assuming the unobserved heterogeneity is normally distributed conditional on the initial states, we can integrate the conditional heterogeneity out by simulation and get what Wooldridge called "conditional" maximum likelihood estimates. Specifically, we add 14 brand-specific initial brand choices variables, as "brand"_initial in the model. From equation $\boldsymbol{b}_h = \boldsymbol{b}_1 \bullet s_0^h + \boldsymbol{e}_h$, these 14 variables are generated as the interaction terms of s_0^h and brand intercepts for each brand respectively, and $\boldsymbol{b}_1$ is the vector of coefficients to estimate. These 14 variables are both alternative-specific and farmer-specific.

Finally, the last set of variables in x_{jm}^h is to control for possible heterogeneity brought by marketing activities. Corresponding to the three purchase sources, we construct three marketing dummy variables—dealer, rep, and retailer in the model. For any given market, we say a source is active for a product (and thus the corresponding marketing variable take value 1) if the corresponding brand is recorded to be purchased from this source in this market in any one of last three years. Thus, these marketing variables are market-product-specific, so they are same for all alternatives of a product among the market.[37] Note that we do not use current year's records to construct these marketing variables to avoid potential endogeneity issues (the purchase source and the choice decision are made simultaneously). We further interact these three variables with three individual-specific variables indicating whether the farmer has purchased from the respective source to capture farmers' heterogeneous responses to the marketing variables. The three interaction terms are coded as ind_dealer, ind_rep, ind_retailer. For example, value "1" for variable ind_dealer means that there is an active dealer for the brand of the alternative and the farmer faced this choice set has purchased from dealer previously.

[37] Note here the marketing variables are market-product-specific rather than farmer-product-specific. To generate farmer-product-specific marketing variables, the purchase source information is needed even for not chosen alternatives, which is not available in our case.

To control for unobserved heterogeneity, our model allows heterogeneous responses to variables in $\boldsymbol{x}_{jm}^{h}$. However, to keep the computational burden of the model within tractable boundaries, a subset of $\boldsymbol{x}_{jm}^{h}$ will have random coefficients. Specifically, the subset variables consist of the GT trait intercepts over time, GT 1999-2004, GT 2005-2010, GT 2011-2016 and the brand intercepts, "brand", a total of 16 variables.

Table 6 below summarizes the descriptive information of some variables in the model from the regression sample. We do not list the brand variables in each period, which can be found in Table 1, neither the "brand"_initial variables. Table 6 shows that the mean of the state dependence term is 0.81, suggesting the existence of structural state dependence. The standard deviation of state dependence term is also relatively large, implying farmers' heterogeneous propensity for previously purchased brand. The trait variables in each period outline the trend of GT adoption—GT seeds are widely adopted in 1999-2004, adoption rate peaked in 2005-2010, and then the rate decreased in 2011-2016. Last 6 rows describe the marketing variables. The statistics show that the three sources are available in most markets but some farmer only have purchased from one or two of all three.

Additionally, the cost instruments are constructed as interactions of futures prices with brand and trait dummies. As noted by Fernandez-Cornejo and Spielman (2002), the cost of seed production can be contracted as adjusted yields times the futures prices of the product for contract farmer growers. Specifically, the futures prices are constructed from the futures contract with a delivery month of November (right after the harvest season) as the average daily closing prices from January to March (before the planting season), as in Kim and Moschini (2018), who use futures prices as expected prices for farmers. The price records are from Quandl.[38] We also deflate the futures prices by the USDA crop sector index of prices paid, consistent with all other prices in the model.

5.3 Estimation results

Table 7 presents estimation results for the logistic regression. Model (27) in section 4.2.1 is labeled as "Main" model in the last column. The first column "Clogit" is a conditional logit model where we set $\boldsymbol{\delta}_h = \mathbf{0}$ comparing to the mixlogit models in all other columns. In the second column, "No state", we set $\gamma_h = 0$ in model (27). The third column "No IV" assumes exogenous price and sets $\lambda = 0$. The forth column "No initial" ignores the initial conditions problem and sets the coefficients of all "brand"_initial variables to zero.

As noted, the final column contain estimation results for the full mixed logit model, our preferred specification and the specification we later use to conduct some counterfactual exercises. The "Coef" and "SD" reports the means and standard deviations of the corresponding coefficient. The mean coefficients for the GT (comparing to conventional product) and brand variables (comparing to Public seeds) generally conform to expectations. All suggest a positive impact on the return to a soybean product. Further discussion of these variables and their economic importance are discussed in the next section where we compute the willingness-to-pay (WTP) distributions.

The full model results demonstrate the importance of specifying random coefficients for state dependence, the GT dummies, and the various brand dummies. As evidence of this, the log likelihood further decreases by more than 10% comparing to "Clogit" and all variance parameters are significant and large in magnitude. The mean estimate for state dependence also

[38] See https://www.quandl.com.

decreases, suggesting some previous bias by not allowing for unobserved heterogeneity. Nonetheless, it still remains quite large in magnitude. This could be because it is *truly* large in magnitude or because our specification of unobserved heterogeneity is still not rich enough to capture the observed purchase patterns. We consider this issue further in section 5.6. To demonstrate the influence of state dependence on the other coefficients, the second column "No state" drops the state dependence term from the model. Here we see a sharp decrease in the log-likelihood, on the order of 10%, and there is some increase in the coefficients of marketing variables and initial conditions (shown in the Appendix A). The coefficients of brand intercepts move towards zero.

Looking across the estimation results, the coefficients are generally estimated with high precision, especially when the control function is used. The importance of the control function approach is also demonstrated by the difference in results between the "no IV" and the "Main" estimation results. The price coefficient decreases substantially from about -0.104 to -0.437, suggesting that endogeneity was indeed present. The positive and significant estimate for the control residual ("control") also suggests the existence of unobservable factors positively correlated with demand, a finding similar to that found in Petrin and Train (2009). For the remaining variables, the state dependence coefficient barely changes, most of the coefficients of initial conditions increase, whereas all other coefficients scale down with the price coefficient.

Comparing with the forth column, we found that the state dependence coefficient decreased significantly in including the initial conditions, suggesting these variables capture an important source of unobserved taste variation not captured by the random coefficients. The coefficients of the initial conditions are all significant at $p < 0.01$ level in the main model. When drop the initial conditions, the log-likelihood decreases by about 10%. We see some changes in coefficients of the marketing variables with no clear direction. The remaining coefficients barely change.

Table 7 does not contain all of the estimated coefficients; the full set of results are provided in the Appendix A. Among the omitted coefficients, there are a few noteworthy results. First, the omitted brand-period interaction coefficients indicate that, compared to the final period, all brands were valued less (relative to public varieties) in the first two periods. This is consistent with the decline in demand for public varieties. A second set of results omitted from the table are the marketing variables. The coefficients for the marketing variables are generally smaller in magnitude and less statistically significant. They also suggest that an active dealer or sales representative in the market can increase farmers' probability of choosing the brand only if the farmer has purchased from the respective source before.

5.4 WTP distributions

The estimated coefficients presented in Table 7, per se, are not terribly informative about the economic importance of the various factors that impact the profitability of soybean varieties. Therefore, we use the coefficient estimates from the mixed logit model to report the WTP distributions for the main variables of interest: the structural state dependence term, the GT coefficient, and the brand coefficients (in the last period).

The WTP for an attribute measures the maximum amount ($/unit) that a farmer is willing to part with for that characteristic, and it is obtained by dividing the coefficient of interest by the price coefficient (Train 2009). However, when allowing farmers to have heterogeneous propensities for seed price, the calculation for the WTP involves the ratio of two random variable. The choice of normal distributions for the random all random coefficients of the mixed logit model, therefore, has undesirable implications. With a normally distributed price coefficient, some farmers may have negative propensity for money. Moreover, the WTP distribution would be a Cauchy; this is a distribution with heavy tails and whose moments do not exist. Our choice of the lognormal distribution for the price coefficient thus brings two significant benefits: The sign of the estimated price coefficient accords with economics prior for all decision makers (Hole and Kolstad 2012); and, the WTP distributions have well-defined moments (Hole and Kolstad 2012; Daly, Hess, and Train 2012). In any event, in section 5.6.2, we will discuss the effect of different distribution assumptions on the model results and the WTP distributions.

There are two parts in Table 8, the "Mean" part and the "Distribution" part. The "Mean" part reports the mean of the WTP distribution and the 95% confidence interval of the mean (the standard error of the mean is calculated using the delta method). These statistics are derived analytically from the estimates in the main model. Specifically, note that, for these WTP estimates, we are interested in $E\left[X/Y\right]$, where $X \sim Normal(\mu_x, \sigma_x^2)$ and $\ln Y \equiv y \sim Normal(\mu_y, \sigma_y^2)$. Hence, given independence between the two random variables,

$$E\left[X/Y\right] = E\left[X\exp(-y)\right] = \mu_x \exp\left(-\mu_y + \frac{\sigma_y^2}{2}\right)$$

The "Distribution" part of Table 8 reports the 10 percentile, the median, and the 90 percentile of the WTP distribution. These statistics are derived by simulation. Specifically, the WTP distribution is simulated by dividing random draws from two corresponding distributions of coefficients using standard deviation estimates from the model. We take 100,000 random draws from each distribution.

We begin with the value of structural state dependence. On average, a farmer is willing to pay about 5.31 ($/unit) more for a brand if they purchased it in the previous period. Given that the mean price for a unit of soybeans is $45/unit, the state dependence effect is about 12% of the average, a relatively large effect and quite close to what Dubé, Hitsch, and Rossi (2010) find in the consumer-packaged goods industries.[39] Put differently, experience is important: for two otherwise equally valued products, a farmer is willing to pay significantly more for the one he has experience with. There is also significant heterogeneity in state dependence. A substantial number of farmers value an additional past experience at more than $9, whereas some farmers obtain little value from an additional experience.

To give a visual depiction of the WTP distributions, Figure 2 provides the numerically-computed density distributions for the estimated state dependence and GT trait WTPs. Consistent with the values in Table 8 where the mean is generally larger than the median, all distributions in the figure are right-skewed. Figure 2 demonstrates that the value of structural state dependence is positive for nearly all farmers. The same can be said for the value of GT traits, and the vast majority of farmers place positive value on GT during all sub-periods. We can also see that the value of the GT trait is larger on average and more dispersed than the value of state dependence. This suggests that some farmers place very high value on the GT trait whereas other farmers do not. Consider, for example, that in the final period about 10% of farmers valued the GT trait at $41 or more, whereas another 10% valued the GT trait at $8 or less.

The three different WTP distributions also demonstrate interesting changes over time. In the first period, 1999-2004, the average WTP is 18.95 ($/unit), this increases to 27.60 ($/unit) in the second period (2005-2010), and then decreases to 23.38 ($/unit) in the third period (2011-2016) (in $/acre terms, these values are $22.47/acre, $32.73/acre, and $27.72/acre,

---

[39] As shown in Table 5, the average retail seed price is $45/unit—$37/unit over the period 1999-2004, $44/unit in the period 2005-2010, and $53/unit in the period 2011-2016. All prices are deflated by the crop sector index of prices paid. To give these number additional context, the U.S. average soybean yield in 2018 was 51.6 bushels per acre and soybean price was $9.15/bushel, implying an average total gross revenue $398/unit (measured in 2018 dollar, adjusted by the unit/acre ratio).

respectively).[40] These changes essentially reflect the changing rate of GT soybean adoption over time. As noted in CIliberto, Moschini, & Perry (2019), the observed increase after the period 1999-2004 was likely the result of falling glyphosate prices (a complementary input to GT soybeans), rising output prices, learning, and/or an increasing number of varieties with GT. However, in the final period, the mean WTP for GT decreases and its variance increases (the orange line). This is consistent with some recent developments in the efficacy of GT soybeans. In recent years, glyphosate weed resistance has become increasing problematic (Perry, Ciliberto, Hennessy, and Moschini, 2016). Some farmers have responded by switching to non-GT varieties or by increasing glyphosate application rates (Perry, Hennessy, Moschini, 2019). Glyphosate weed resistance also varies considerably across the U.S., which may explain the increasing variance in farmers' WTP.

The remaining entries in Table 8 represent farmers' mean WTP for the each of the brands *relative* to public varieties in the last period (2011-2016). As expected, farmers are willing to pay a signficant premium for a branded product, with Asgrow and Pioneer having some of the highest WTPs. There is also considerable heterogeneity. Diminishing brands Golden Harvest and Dekalb have the most concentrated WTPs, whereas some of the mid-size to smaller brands like Beck's and Channel have widely distributed tastes.

5.5 Demand elasticities

A major advatange of our framework, as contrasted with the basic logit model, is that it can capture rich substitution patterns between seed varieties. To demonstrate these patterns, we compute and report simulated mean own-price and cross-price demand elasticities for each product. To compute these elasticities, we compute predicted market shares for each product j,

[40] In Ciliberto, Moschini, and Perry (2019), the WTP for the GT trait in soybeans was \$16.68/acre in 1996-2000, \$23.25/acre in 2001-2006, and \$24.66/acre in 2007-2011.

denoted S_j. To obtain these shares, we first predict farm-level probabilities for each product using the estimated coefficients from the full mixed logit model. Because this model includes random coefficients, this is done through simulation. Given the individual level probabilities, we then aggregate to the product-market level using the observed number of purchased units for each farmer as weights. To generate the elasticities, we change the price for each product k by a small amount $p_k' - p_k$, and then recompute the aggregate prediced market share for product j, denoted by S_j'. The elasticity of demand for product j with respect to a change in the price of product k is given by:

$$e_{jk} = \frac{\Delta\left(S_j' - S_j\right)}{\Delta\left(p_k' - p_k\right)} \frac{p_k}{S_j}, \tag{35}$$

where e_{jj} is the own-price elasticity of demand and $e_{jk}\left(j \neq k\right)$ represents the cross-price elasticity of demand j for product k.

Table 9 contains the full matrix of elasticities. Each product j is listed in the first column and each product k is listed in the top row. Thus, the own-price and cross-price elasticities of demand for product j with respect to a change in the price of product k are reported in the corresponding row. Note that we use “0” to represent conventional products and “1” to represent GT products. In the first row of Table 9, to save space, we only use the first two letters of each brand. For example, the top-left entry of “-8.63” corresponding to Asgrow0 and AS0 is the own-price elasticity of demand for Asgrow0; the value to the right of this, “0.06”, corresponding to Asgrow0 and BE0, is the cross-price elasticity of demand for Asgrow0 with respect to a change in the price of Beck’s0. We further divide Table 9 into 4 sub-matrices—the top left panel contains elasticities for conventional products with respect to other conventional products, the top right panel contains elasticities for conventional products with respect to GT products, the bottom left panel contains elasticities for GT products with respect to conventional products, and the bottom right panel contains elasticities for GT products with respect to GT products. Grouping it this way allows us to see some clear patterns in the elasticities.

The orange cells in Table 9 contain the own-price demand elasticities for each product. They are all negative and highly elastic, typically ranging from about -5 to as high as -12. Despite different models, the values are close to the estimated mean own-price elastictes reported in CIliberto, Moschini, & Perry (2019). In their preferred model, they find a mean own-price elasticity of -7.04 for corn and soybean products. We also note that GT products are slightly more elastic than conventional products, which may in part be the result of higher prices for GT products.

The blue cells highlight cross-price elasticities for different products marketed with the same brand. Because not all brands posses both GT and conventional products (Channel and Public), there is some asymmetry along the alignment of the blue cells. Cells highlighted in green identify the closest substitute for each product k: this is simply the cell with the highest value in each column (excluding the own-price elasticities). If this cell *also* happens to be the product with the same brand, then it is highlighted in blue-green.

Three intuitive regularities emerge from the cross-price elasticies. First, generally speaking, a farmer is more likely to substitute between products that contain the same trait. We term this the “trait effect”. Put differently, if a product with GT is a farmer’s most preferred variety, it is significantly more likely that their next preferred product also has GT. This can be seen by the fact that the upper left and lower right blocks of cross-price elasticities are typically larger compared to the lower left and upper right blocks (the main exception to this is the crosssubstition from GT to conventional products of the same brand). Second, individuals are typically more likely to substitute products of the same brand. We call this the “brand effect”. Consider, for example, how individuals substitute from ASo to products with GT (the first column in the lower left block of Table 9). Among all such products, farmers are most likely to substitue to Asgrow1: the value of 0.10 exceeds all other values in the lower left panel. Notice, however, that the cross-price elasticites for all conventional products from ASo are greater than 0.10. Thus, in this case, the trait effect dominates the brand effect. More generally, for conventional products, the trait effect usually dominates the brand effect (though not always). This is evidenced by the fact that the majority of green cells for conventional products are in the upper right block. Conversely, for GT products, the closest substitute is almost always the identically branded conventional version.[41] For example, the closest substitute for AS1 is Asgrowo (value of 2.48). The trait effect, however, dominates if we exclude the conventional Asgrow variety; the closest substitutes for AS1 are all non-Asgrow GT products. Finally, there is a strong assymetry in the cross-price elasticities between GT and conventional products of the same brand. For example, the cross-price elasticity of 0.10 from ASo to Asgrow1 is a small fraction of the cross-price elasticity of 2.48 from AS1 to Asgrowo. This is simply due to the fact that GT products typically have much larger shares compared to the identically branded conventional versions (specifically, the denominator of equation (35) is much smaller for conventional products).

---

[41] We suspect that the results mean that farmers are more likely to switch to the conventional version of the same brand. The high substitution rate can be a result of small share of conventional product from the formation of elasticity in equation in (35).

5.6 Robustness check

In this section, we present some robustness check for the model results. The main purpose of this section is to investigate how the state dependence coefficient and WTP estimates are affected by model specifications. The main concern for the state dependence coefficient is that unobserved heterogeneity has not been fully controlled, so we utilize the non-zero-order feature of the structural state dependence and do the experiment of reshuffling. Next, we use different specifications for the price coefficient to explore their effects on the WTP estimates. We further consider potential patterns in farmers' attitude towards state dependence, which is relegated to Appendix B.

5.6.1 State dependence and reshuffle

Although our framework permits unobserved heterogeneity for all brands, we do restrict this heterogeneity to follow a normal distribution. Previous research has shown that, even having controlled for unobserved heterogeneity, it is still possible to incorrectly find positive evidence of structural state dependence if that unobserved heterogeneity is not sufficiently flexible enough. The term for this is spurious state dependence. To check whether unobserved heterogeneity has been captured in a sufficiently rich way, we conduct a reshuffling procedure along the lines of Dubé, Hitsch, & Rossi (2010) and Bronnenberg, Dubé, & Moorthy (2019). The basic idea of this procedure is to reshuffle the choice sequences in a random way and then re-estimate the full mixed logit model. This exploits the fact that structural state dependence should in principle only be identified by non-zero order features in the data. If we have sufficiently controlled for unobserved heterogeneity, then the state dependent parameter should go to zero. On the other hand, if the estimate for γ remains large and positive, this may suggest that our original estimate is spurious and is likely due to an insufficiently rich accounting for unobserved heterogeneity.

For reshuffling, we first generate a new time variable whose values are drawn from a discrete uniform distribution with values ranging from 1996-2016. We then replace the original time variable with this new, randomly generated time variable. Consequently, we build a "reshuffled" purchase history for each farmer, which results in a new "randomized" state dependence term. All other explanatory variables are held fixed throughout the reshuffle process. In other words, instead of randomly reshuffling the purchase sequence for each individual and then reconstructing new market-specific choice sets, we maintain the original choice set for each individiual and thereby *only* reshuffle the state dependence term. We do it this way for the following reasons. The main reason is that the industry has experienced signficant structural changes during the observed timespan—some brands have exited and entered the industry, prices have risen, and the GT trait has come to dominate. For these reasons, fully reconstructed choice-sets would have unreasonable properties. Consider the following example. Suppose an individual purchased the brand Channel with GT in 2015, and that upon reshuffling the new, randomly assigned year was 1996. If choice-sets were fully recontstructed, then Channel with GT will enter the choice set of all farmers in this local CRD in 1996 and its price will be an average price of all corresponding Channel products after reshuffle. This raises three problems: (i) Channel only enters the market after 2009; (ii) seed prices in the later periods are significantly higher than in early periods, even after deflation; and (iii) the size of any choice-set is subject to change in the reconstruction process.

The model results after reshuffling are presented in Table 10. We report results for two types of models: the basic conditional logit model, where unobserved heterogeneity is not captured by random coefficients, and the mixed logit model, which controls for unobserved heteroegeneity through the inclusion of random coefficients for the brand and trait dummies.

Overall, the coefficient for structural state dependence significantly decreases in both models, suggesting that state dependence is indeed a feature of soybean seed demand. In the conditional logit model, the coefficient decreases from 2.363 to 1.268, whereas in the mixed logit model, the coefficient decrease from 2.092 to 0.384. The smaller decrease in state dependence for the conditional logit model highlights the importance of controlling for unobserved heterogeneity.

While the state dependence coefficient does decrease signficantly in both cases, particularly in the full model, it still remains positive with a coefficient of 0.384. This may suggest that the assumption of normally distirbuted random coefficients is not rich enough and therefore a portion of our original state dependence estimate is spurious. We note, however, that an alternative possiblity is that this is the result of a higher-order Markov chain (more distant purchases may still have some impact on farmers' seed choice), or even the result of our large sample size. In any case, our reshuffling procedure suggests that at least 80% of what we captured in the mixed logit model in Table 7 is the result of genuine structural state dependence. To the extent that there is some bias in the structural state dependence coefficient, it is small (less than $1/unit).

5.6.2 Specifications for the price coefficient

We dicussed the reason why we presume lognormal distribution for the price coefficient. Despite the undesirable features, the specification of normal distribution can be an alternative. As there has long been a debate over the presumed distribution of price coefficient, in this section we present the model results under different specifications and we further explore the effects on WTP estimates.[42]

---

42 For example, Revelt and Train (1998) set fixed price coefficient; Petrin and Train (2009) assume constant price coefficient but interact price with income to account for consumers' heterogeneous propensities for price; Negrin et al. (2008) specify normal distribution for all coefficients and measure WTP at the means. Regier et al. (2009) let the coefficients for cost and waiting time follow log-normal distributions and calculate WTP also at the means of the

Table 11 lists part of the model results under four different price specifications, and Table 12 reports the WTP estimates. The last column, labeled as “Log-normal”, sets the price coefficient to be log-normal distributed. This is our main model in Table 7 and the WTP estmates in Table 12 are the mean values in Table 8. In the first two columns, we restrict the price coefficient to be fixed but interact price with acre range in the second column to account for possible income effect on price. Models results in these two are not significantly different from each other. However, the log-likelihoods are lower than the last two columns where standard deviations of the distribution of price coefficient are significantly not zero, suggesting the importance of specifying heterogeneous responses to price, which can not be captured by individual acre range. The WTP estimates of the first two columns are the means of the WTP distribution, calculated as the mean of the attribute divided by the price coefficient.

The third column in Table 11 reports the model results when we specify normal distribution for the price coefficient. The moments of the WTP distributions do not exist, so in Table 12 we report the division of the mean of the attribute and the mean of the price coefficient. Comparing the third column with our main model, we see a slight increase in the log-likelihood, suggesting a better fit with normal price coefficient, which is in expectation as log-normal distribution add additional constraint on the price coefficient (forced to be negative). Anyway, the change in loglikelihood is not significant (about 0.3%). The mean of the price coefficient scales down and so does all other variables except the state dependence term, resulting in an increase in the WTP estimate for the state dependence term but similar WTP estimates for all other attributes. However, as discussed in section 5.6.1 that we may overestimate the state

coefficient distributions. See more details in Bliemer and Rose (2013), Hole and Kolstad (2012), and Louviere et al. (2005).

dependence coefficient, we have higher tolenrance for a lower WTP (Log-normal) for the state dependence term rather than a higher one (Normal).

6. Counterfactual Analysis

The overall goal of the counterfactual analysis is to identify the unique impacts of structural state dependence on dynamic farmer behavior. In other words, we want to identify how the purchasing patterns of soybean seed consumers differs depending on whether structural state dependence is present. The analysis provides insight into firm's optimal pricing strategy. Moreover, it also sheds light on the decision of technology adoption, particularly for the soybean seed market, which introduces the GT technology since 1996.

Following the ideas, we have two parts in the section. In section 6.1, we consider the counterfactual scenario of a temporary price discount in 1999, the start point of our estimation. In section 6.2, we study the counterfactual scenario of late GT adoption after 1999. In both cases, we let each brand experience each shock at one time and then simulate the after-shock market shares over time. The above process is done for both the main model and model without inertia as a comparison to investigate the function of state dependence over time.

6.1 Price discount

In this section, we study the effect of state dependence on farmers' purchasing behavior by imposing a temporary price discount of 5%. Like Keane (1997) who did the simulation of price discount in his choice model of ketchup, we aim at examining farmers' long-term behavior with state dependence. The temporary price discount takes place in 1999 and we further assume that all prices return to their original values after the discount. In other words, there is no further price adjustment for firms, so the simulated results are not equilibrium results. As noted, we focus on discovering farmers' purchase behavior, how firms react in equilibrium is studied in Dube, Hitsch and Rossi (2009) and Mackay and Remer (2019)—generally, state dependence implies dynamic pricing strategy for firms and may or may not make the market more competitive. Further note that the data in analysis are unbalanced—not all farmers are observed in 1999. The temporary price discount will not affect those farmer. As a result, the effect of the temporary price discount can be underestimated, however it will not affect the major findings in this exercise.

Results in Table 13 are predicted from model (27) using the estimates in Table 7. We make some additional assumptions to use our model estimates to do out-sample predictions for the observations in time period 1996-1998. First, we use the estimates of marginal propensity for GT trait in 1999-2004 for 1996-1998 (we have the estimates in 1999-2004, 2005-2010, and 2011-2016, respectively). Then, state dependence term in 1996 is assumed to be 0 as previous purchase records are not available.

The first column in Table 13 lists all brands except for Channel, which only enters the market since 2010. Each brand stands for two rows whose estimates are derived when the brand has a 5% price drop. The “state” row uses the main model with state dependence, whereas the “nostate” row reports results simulated from the Noinertia model when the state dependence coefficient is set to zero. The third column shows the weighted (by purchased units) market shares of each brand at 1999. This column serves as a reference for the remaining part of this table. The last 7 columns list the changes in market shares of each brand over time if its price has a temporary 5% drop in 1999 in the two scenarios respectively.

There are two main findings from Table 13. First, the temporary price discount has a larger effect on the current market shares if there is no state dependence in the market. However, without state dependence, the price discount only affects current market shares. When state dependence exists, the effect of temporary price discount is long-lasting although diminishing.

From the listed shares in the third column, we find that state dependence benefits Pioneer and Other (we treat Other as a separate brand), the two largest brands in the market, and hurts all other brands, with smaller brands suffer more.

6.2 Late adoption of GT

GT trait was introduced in the soybean seeds market since 1996 and widely adopted thereafter. [43] Taking advantage of the feature of our data that begin from 1996, we design the experiment of late GT adoption decision for each brand. In this section, we again investigate farmers dynamic purchase behavior implied by state dependence. Moreover, we try to understand the implications of a brand deciding to adopt a new technology facing its consumers' dynamic purchase behavior.

In the experiment, we let each brand (except Channel) be a late adopter of GT technology repectively—the brand only sold conventional product from 1996-1999 and picked up the new technology from 2000.[44] Table 14 follows the same structure with the table above whereas the last 7 columns record the changes in market shares over time due to late adoption of GT technology for each brand.

We have three main findings. The first one is same as what we find in section 6.1: the effect of late GT adoption is long-lasting when there is state dependence, whereas the effect is temporary without state dependence. In the second, unlike a temporary price discount late adoption of GT technology can have a larger effect on the market shares at 1999 when there is

[43] All brands except Public and Channel sold GT product since 1997, the second year after the introduction of GT. Asgrow and Pioneer are earlier and faster adopters in the first few years.

[44] The following process explains how we drop GT products in 1996-1999 for each brand. Take Asgrow for example. If the market sales "Asgrow 0" and "Asgrow 1", then we directly drop the alternative of "Asgrow 1". If the market only sales "Asgrow 1", we replace it with "Asgrow 0" and change the price to national average price of "Asgrow 0" in the corresponding year.

state dependence, which is the case for Asgrow, Growmark, Other, and Pioneer. We take this difference as a result of the accumulative effects of first three years. Finally, late adoption of GT trait causes a bigger loss for Asgrow, Other, and Pioneer, which have the largest market shares over time and are also early and major adopters of the new technology. Overall, we can see that a quick response to major technology innovation is crucial in the seed market: almost all brands lose more than half of their market shares if they fail to adopt the technology in the first few years.

7. Conclusion

In this paper, we develop and estimate a micro-level structural model of U.S. soybean seeds demand to study a recurring theme in economics and marketing, brand inertia. To disentangle unobserved heterogeneity from state dependence, we adopt the random coefficients logit model for demand estimation. For the initial conditions problem in the estimation of state dependence, we use the initial brand choices as extra explanatory variables. To further deal with price endogeneity, we apply the control function approach in this nonlinear regression, using the previous year's futures price as a cost instrument and interacting them with brand and GT trait dummies.

Our results show that structural state dependence generally exists for all farmers, with an average WTP of $5.31/unit, which is about 12% of the average soybean retail price of $45/unit. We also find that state dependence is quite heterogeneous—state dependence is valued at more than $10/unit for 10% of farmers, whereas another 10% value it at less than $2/unit. Along with demand estimation, we show that farmers' WTPs for the GT trait vary over time and over individuals. On average, the WTP for GT is $18.95/unit during 1999-2004, goes up to $27.60/unit during 2005-2010, and then declines to $23.38/unit during 2011-2016. Adjusted by unit/acre ratio, our results are consistent with the WTP estimates reported by Ciliberto, Moschini, and Perry (2019), who employ a different modeling approach. Our finding of declining WTP for the GT trait in the 2011-2016 period is also consistent with the emergence of glyphosate resistant weed (Perry, Ciliberto, Hennessy, and Moschini, 2016). We further show that farmers are quite heterogeneous towards GT traits, with much higher variance than the state dependence coefficient.

Using the estimated model, we generate own and cross-price demand elasticities for each product. We find that the own-price elasticities are, on average, -8.6 for conventional products and -10 for GT products. Generally, farmers are more likely to substitute among products of the same brand and same trait, however, the strength of these effects differ depending on the type product. In particular, a farmer is more likely to switch to another conventional product for a different brand if she chooses a conventional product, whereas she is more likely to switch to the conventional product of the same brand if she chooses a GT product.

Finally, we assess some potential implications of state dependence for farmers' dynamic purchase behavior. We find that temporary shocks lead to long-run responses from farmers. In comparison, in the simulated results from the model without state dependence any temporary shock only impacts concurrent time period. Nevertheless, the temporary effect may be larger than its counterpart in the model with state dependence. The counterfactual analysis of late GT adoption shows that early adoption of a major technology innovation is crucial for seed companies. Even if they catch up later, the loss in market shares is substantial and long-lasting.

References

Akay, A. (2012). Finite-sample comparison of alternative methods for estimating dynamic panel data models. **Journal of Applied Econometrics**, 27(7), 1189-1204.

Arulampalam, W., & Stewart, M. B. (2009). Simplified implementation of the Heckman estimator of the dynamic probit model and a comparison with alternative estimators. **Oxford Bulletin of Economics and Statistics**, 71(5), 659-681.

Barrows, G., Sexton, S. and Zilberman, D. (2014) Agricultural Biotechnology: The Promise and Prospects of Genetically Modified Crops. **Journal of Economic Perspectives**, Vol. 28, pp. 99-119.

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

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

Berto Villas-Boas, S. (2007). Vertical relationships between manufacturers and retailers: Inference with limited data. **The Review of Economic Studies**, 74(2), 625-652.

Bliemer, M. C., & Rose, J. M. (2013). Confidence intervals of willingness-to-pay for random coefficient logit models. **Transportation Research Part B: Methodological**, 58, 199-214.

Bronnenberg, B.J., Dhar, S.K. and Dubé, J.P.H. (2009). Brand history, geography, and the persistence of brand shares. **Journal of Political Economy**, 117(1), pp.87-115.

Bronnenberg, B. J., Dubé, J. P. H., & Gentzkow, M. (2012). The evolution of brand preferences: Evidence from consumer migration. **American Economic Review**, 102(6), 2472-2508.

Bronnenberg, B. J., Dubé, J.-P., & Moorthy, S. (2019). The Economics of Brands and Branding. In: Dubé, J. P. & Rossi, P. E., eds, **Handbook of the Economics of Marketing**, Elsevier B.V., chapter 6, pp. 291-358.

Cameron, A.C. and Trivedi, P.K. (2005). **Microeconometrics: methods and applications**. Cambridge university press.

CIliberto, F., Moschini, G., & Perry, E. (2019). Valuing Product Innovation: Genetically Engineered Varieties in US Corn and Soybeans. **The RAND Journal of Economics**, 50(3), 615-644.

Clancy, M. S., & Moschini, G. (2017). Intellectual property rights and the ascent of proprietary innovation in agriculture. **Annual Review of Resource Economics**, 9, 53-74.

Daly, A., Hess, S., & Train, K. (2012). Assuring finite moments for willingness to pay in random coefficient models. **Transportation**, 39(1), 19-31.

Dubé, J. P., Hitsch, G. J., & Rossi, P. E. (2010). State dependence and alternative explanations for consumer inertia. **The RAND Journal of Economics**, 41(3), 417-445.

Dubé, J.-P., Hitsch, G. J., & Rossi, P. E. (2009). Do switching costs make markets less competitive? **Journal of Marketing research**, 46(4), 435-445.

Fernandez-Cornejo, J., & Spielman, D. J. (2002). Concentration, market power, and cost efficiency in the corn seed industry, Working Paper, No. 375-2016-20005.

Fernandez-Cornejo, J. (2004). **The Seed Industry In U.S. Agriculture: An Exploration Of Data And Information On Crop Seed Markets, Regulation, Industry Structure, And Research And Development.** United States Department of Agriculture, Economic Research Service.

Fernandez-Cornejo, J., Hendricks, C., & Mishra, A. (2005). Technology Adoption and Off-Farm Household Income: The Case of Herbicide-Tolerant Soybeans. **Journal of Agricultural and Applied Economics**, 37(3), 549-563.

Funk, T. F., & Vincent, A. T. (1978). The farmer decision process in purchasing corn herbicides. No. 1620-2016-134790.

Goldberg, P. (1995). Product Differentiation and Oligopoly in International Markets: The Case of the U.S. Automobile Industry. **Econometrica**, 63(4), 891-951.

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

Graff, G. D., Rausser, G. C., & Small, A. A. (2003). Agricultural biotechnology's complementary intellectual assets. **Review of Economics and Statistics**, 85(2), 349-363.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. **American Economic Review**, 103(7), 2643-82.

Harbor, A. L., Martin, M. A., & Akridge, J. T. (2008). Assessing input brand loyalty among US agricultural producers**. International Food and Agribusiness Management Review**, 11(1030-2016-82702), 17-34.

Heckman, J.J. (1981). Heterogeneity and state dependence. In **Studies in labor markets** (pp. 91-140). University of Chicago Press.

Heckman, J. J. (1987). **The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some Monte Carlo evidence** (pp. pp-114). University of Chicago Center for Mathematical studies in Business and Economics.

Hole, A. R. (2007). Fitting mixed logit models by using maximum simulated likelihood. **The Stata Journal**, 7(3), 388-401.

Hole, A. R., & Kolstad, J. R. (2012). Mixed logit estimation of willingness to pay distributions: a comparison of models in preference and WTP space using data from a health-related choice experiment. **Empirical Economics**, 42(2), 445-469.

Horsky, D., Misra, S., & Nelson, P. (2006). Observed and unobserved preference heterogeneity in brand-choice models. **Marketing Science**, 25(4), 322-335.

ISAAA. (2016). **Global Status of Commercialized Biotech/GM Crops**: 2016.

Keane, M. P. (1997). Modeling heterogeneity and state dependence in consumer choice behavior. **Journal of Business & Economic Statistics**, 15(3), 310-327.

Kim, H., & Moschini, G. (2018). The dynamics of supply: US corn and soybeans in the biofuel era. **Land Economics**, 94(4), 593-613.

Kohls, R. L., Stucky, R. L., & Gifford, J. I. (1957). Farmers' Selection of Farm Machinery Dealers. **Journal of Marketing**, 21(4), 446-450.

Kool, M. (1994). Vendor loyalty of farmers: Characterisation, description and analysis. **European Review of Agricultural Economics**, 21(2), 287-307.

Lamkey, K. (2004). Seed production in corn and soybean. Agronomy Reports. 4. https://lib.dr.iastate.edu/agron_reports/4/ on Apr 3, 2020.

Louviere, J., Train, K., Ben-Akiva, M., Bhat, C., Cameron, T. A., Carson, R. T., . . . Waldman, D. (2005). Recent progress on endogeneity in choice modeling. **Marketing Letters**, 16(3-4), 255-265.

MacKay, A. and Remer, M., 2019. Consumer Inertia and Market Power. **SSRN 3380390**.

Moschini, G. (2008). Biotechnology and the development of food markets: retrospect and prospects. **European Review of Agricultural Economics**, 35(3), 331-355.

Moschini, G. (2010). Competition issues in the seed industry and the role of intellectual property. **Choices**, 25(2), 1-12.

Negrin, M. A., Pinilla, J., & Leon, C. J. (2008). Willingness to pay for alternative policies for patients with Alzheimer's disease. **Health Economics**, Policy and Law, 3(3), 257-275.

Nevo, A. (2000). A practitioner's guide to estimation of random-coefficients logit models of demand. **Journal of Economics & Management Strategy**, 9(4), 513-548.

OECD. (2018). **Concentration in Seed markets: Potential Effects and Policy Responses**. Paris: OECD.

Perry, E. D., Ciliberto, F., Hennessy, D. A., & Moschini, G. (2016). Genetically engineered crops and pesticide use in US maize and soybeans**. Science Advances**, 2(8), e1600850.

Perry, E.D., Moschini, G. and Hennessy, D.A., 2016. Testing for complementarity: Glyphosate tolerant soybeans and conservation tillage. **American Journal of Agricultural Economics**, 98(3), pp.765-784.

Perry, E.D., Hennessy, D.A. and Moschini, G., 2019. Product concentration and usage: Behavioral effects in the glyphosate market. **Journal of Economic Behavior & Organization**, 158, 543-559.

Petrin, A., & Train, K. (2010). A control function approach to endogeneity in consumer choice models. **Journal of Marketing Research**, 47(1), 3-13.

Regier, D. A., Ryan, M., Phimister, E., & Marra, C. A. (2009). Bayesian and classical estimation of mixed logit: an application to genetic testing. **Journal of Health Economics**, 28(3), 598-610.

Revelt, D., & Train, K. (1998). Mixed logit with repeated choices: households' choices of appliance efficiency level. **Review of Economics and Statistics**, 80(4), 647-657.

Schenkelaars, P., de Vriend, H., Kalaitzandonakes, N., Magnier, A., & Miller, D. (2011). Drivers of Consolidation in the Seed Industry and its Consequences for Innovation. **Report commissioned by COGEM**.

Seetharaman, P. B. (2004). Modeling multiple sources of state dependence in random utility models: A distributed lag approach. **Marketing Science**, 23(2), 263-271.

Seetharaman, P. B., & Chintagunta, P. (1998). A model of inertia and variety-seeking with marketing variables. **International Journal of Research in Marketing**, 15(1), 1-17.

Sellars, S. C., & Gunderson, M. A. (2018). Brand and Dealer/Retailer Loyalty among Large U.S. Farmers. Selected presentation in 2018 AAEA.

Shi, G., Chavas, J.-P., & Stiegert, K. (2010). An analysis of the pricing of traits in the US corn seed market. **American Journal of Agricultural Economics**, 92(5), 1324-1338.

Simonov, A., Dubé, J.P., Hitsch, G.J., and Ross, P. (2019) State-dependent demand estimation with initial conditions correction. **NBER Working Paper** No. 26217, September.

Sudhir, K., & Yang, N. (2014). Exploiting the Choice-Consumption Mismatch: A New Approach to Disentangle State Dependence and Heterogeneity. Accessed at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2409115 on Apr 3, 2020.

Syngenta. (2016). Our industry 2016. https://www.scribd.com/document/317849502/Syngenta-Industry-Whats-Next on Apr 3, 2020

Train, K. E. (2009). **Discrete choice methods with simulation**. Cambridge University Press.

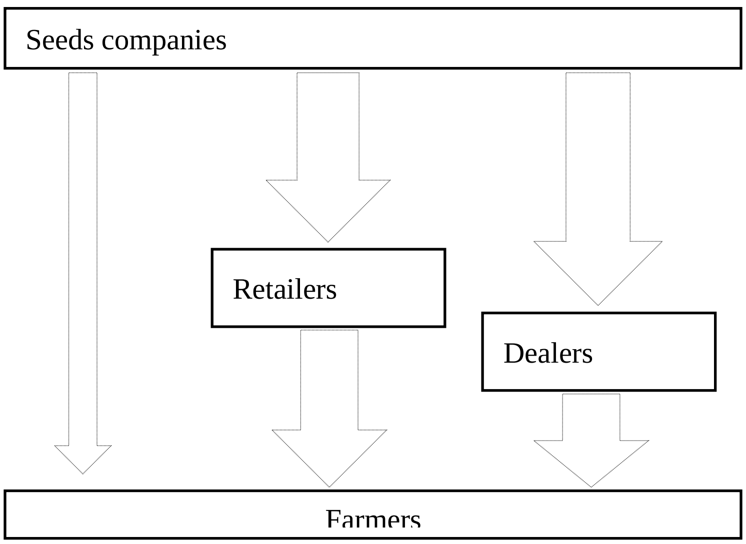
Train, K. E., & Winston, C. (2007). Vehicle choice behavior and the declining market share of US automakers. **International Economic Review**, 48(4), 1469-1496.

Wechsler, S. J., McFadden, J. R., & Smith, D. J. (2018). What do farmers' weed control decisions imply about glyphosate resistance? Evidence from surveys of US corn fields. **Pest Management Science**, 74(5), 1143-1154.

Wooldridge, J. M. (2005). Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. **Journal of Applied Econometrics**, 20(1), 39-54.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. **Journal of Human Resources**, 50(2), 420-445.

Figure



Note: Adapted from Syngenta (2016, p. 46)

Figure 1. The seed distribution structure in the United States

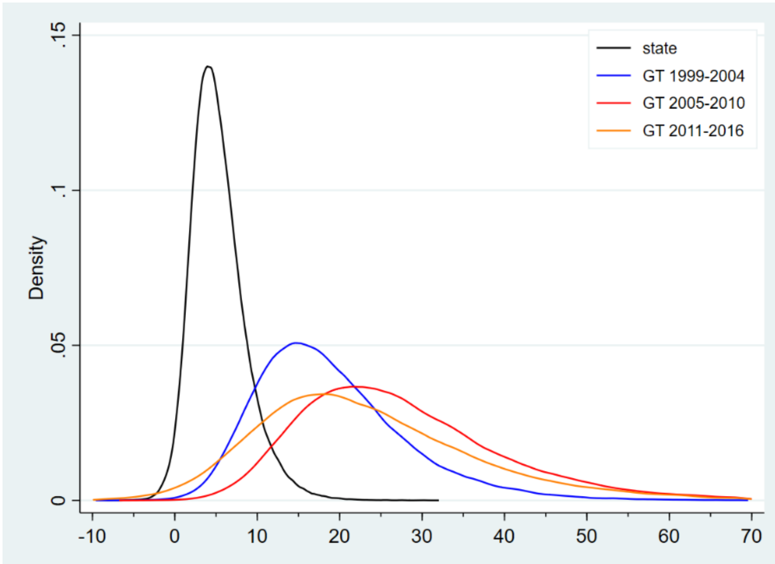


Figure 2. WTP distributions for state and trait

Table

Table 1. Brand Market shares

Market share	1996-98	1999-2004	2005-10	2011-16
Monsanto				
Asgrow	0.133	0.166	0.161	0.217
Channel	0	0	0.006	0.032
DeKalb	0.078	0.054	0.037	0.003
Kruger	0.013	0.017	0.013	0.006
DuPont				
Pioneer	0.186	0.214	0.268	0.285
Syngenta				
Golden	0.037	0.061	0.027	0.001
NK	0.051	0.054	0.087	0.079
Dow				
Mycogen	0.024	0.021	0.013	0.020
Public				
Public	0.063	0.026	0.007	0.005
Others				
Beck's	0.007	0.013	0.016	0.029
Croplan	0.012	0.025	0.033	0.030
Growmark	0.018	0.014	0.011	0.008
Stine	0.041	0.031	0.024	0.021

Table 2. Markets and Products

Year	Number of Markets	Average number of products		
		Total	GT	Conventional
1996	165	5.60	0.61	4.99
2000	182	9.02	4.95	4.07
2004	174	7.17	5.33	1.83
2008	178	6.06	5.29	0.77
2012	188	5.87	4.88	0.98
2016	189	6.24	4.81	1.42

Table 3. Purchase and repurchase rates of each brand

Brand	Purchase	Repurchase
Asgrow	18.00	78.37
Beck's	2.29	76.51
Channel	1.42	54.88
Croplan	2.78	61.72
DeKalb	3.05	65.37
Golden	3.08	72.81
Growmark	1.28	75.77
Kruger	1.49	74.75
Mycogen	1.78	73.93
NK	7.44	69.86
Other	29.6	85.41
Pioneer	24.12	86.19
Public	0.87	57.05
Stine	2.80	69.74

Table 4. Data cleaning and reformation

Deletion	No.
Original soybean choices	**213,062**
Purchase records if purchase source is “From my own farm”, “I’m a seed grower”, or “New seed that was left over from last year”	6,890
Purchase records if the seed is not newly purchased	1,367
Purchase records if the product is “Public” with GT trait	108
Choices in analysis	**204,697**
(Following deletions will not affect purchase history)	
Purchase records of zero net prices	831
Purchase records of year 1996, 1997, 1998	21,924
Purchase records in markets with only one alternative	584
(Following deletions will not affect available alternatives in one market)	
Purchase records of all farmers' first-time recorded purchase	62,548
Purchase records of all farmers if they show up in less than or equal to 3 years in the total time span	28,546
Choices in regression	**90,264**
(Expand each choice by available products in the local market)	
Observations in regression	**1,057,637**

Table 5. Descriptive information

Variable	Mean	SD	Min	Max
No. of alternatives in one market	11.717	4.393	2	23
No. of chosen brands of a farmer	3.730	1.831	1	11
No. of chosen brands in one year's purchase of a farmer	1.901	1.009	1	8
No. of chosen products of a farmer	4.659	2.538	1	17
No. of chosen products in one year's purchase of a farmer	1.994	1.084	1	8
No. of recorded years of a farmer	7.984	3.689	4	21
Retail price	44.924	10.900	8.757	89.646
Net price	40.734	9.050	8.757	85.167
Discount	4.189	4.321	0	37.363

Note: the mean is calculated by averaging over all purchase records, rather than averaging over the markets or farmers

Table 6. Descriptive information of model variables

Variable	Obs	Mean	Std. Dev.	Min	Max
price	90264	40.734	9.050	8.757	85.167
state	90264	0.811	0.392	0	1
GT 1999-2004	29284	0.768	0.422	0	1
GT 2005-2010	28300	0.960	0.196	0	1
GT 2011-2016	32680	0.938	0.242	0	1
dealer	90264	0.938	0.241	0	1
rep	90264	0.851	0.356	0	1
retailer	90264	0.948	0.222	0	1
ind_dealer	90264	0.759	0.428	0	1
ind_rep	90264	0.435	0.496	0	1
ind_retailer	90264	0.626	0.484	0	1

Table 7. Estimation Results[45]

	Clogit b/(se)	No state b/(se)	No IV b/(se)	No initial b/(se)	Main b/(se)
Coef					
price	-0.199***	-0.404***	-0.104***	-0.439***	-0.437***
	(0.016)	(0.021)	(0.006)	(0.022)	(0.022)
control	0.138***	0.315***		0.353***	0.349***
	(0.014)	(0.019)		(0.020)	(0.019)
state	2.363***		2.095***	2.500***	2.092***
	(0.040)		(0.051)	(0.058)	(0.052)
GT 1999-2004	3.442***	6.776***	3.280***	7.635***	7.470***
	(0.197)	(0.326)	(0.158)	(0.297)	(0.314)
GT 2005-2010	4.933***	9.798***	6.127***	11.256***	10.878***
	(0.246)	(0.422)	(0.301)	(0.499)	(0.501)
GT 2011-2016	3.405***	7.434***	5.811***	8.516***	9.214***
	(0.146)	(0.455)	(0.319)	(0.443)	(0.445)
SD					
price		0.132***	0.131***	0.138***	0.145***
		(0.022)	(0.011)	(0.007)	(0.006)
state			1.036***	1.267***	1.043***
			(0.007)	(0.037)	(0.038)
GT 1999-2004		2.311***	2.360***	2.450***	2.508***
		(0.208)	(0.129)	(0.141)	(0.185)
GT 2005-2010		2.822***	3.189***	3.680***	3.271***
		(0.201)	(0.293)	(0.241)	(0.242)
GT 2011-2016		3.714***	4.153***	4.046***	4.283***
		(0.266)	(0.212)	(0.272)	(0.278)
LL	-125277	-118707	-110442	-110960	-109647
N	1057637	1057637	1057637	1057637	1057637

Note: standard errors are clustered at CRD level. $^{*}p<0.05$, $^{**}p<0.01$, $^{***}p<0.001$.

[45] Note that the price coefficient is presumed to be lognormal distributed. The Coef and SD of price in mixlogit models in the table are the mean and SD of the lognormal distribution. “mixlogit” in STATA outputs the mean and SD of the natural logarithm of the price coefficient (Hole, 2007). The delta method is used for the standard errors of the estimate.

Table 8. WTP estimates derived from the main model

WTP	Mean(Analytic)			Distribution(Numerical)		
	Value	95% CI		p10	p50	p90
state	5.31	4.68	5.94	1.65	4.86	9.56
GT 1999-2004	18.95	17.56	20.34	8.81	17.51	30.90
GT 2005-2010	27.60	25.61	29.59	13.66	25.61	44.20
GT 2011-2016	23.38	20.45	26.30	8.12	21.39	41.33
Asgrow	25.82	22.94	28.71	16.05	24.50	37.34
Beck	21.27	18.32	24.21	11.36	19.81	32.97
Channel	17.48	13.90	21.07	7.55	16.10	29.34
Croplan	21.99	19.30	24.68	13.39	20.80	32.22
DeKalb	14.36	11.18	17.55	8.99	13.64	20.67
Golden	14.87	11.71	18.03	8.77	13.97	22.14
Growmark	22.73	19.67	25.78	13.88	21.45	33.19
Kruger	18.17	15.06	21.29	10.10	16.98	27.76
Mycogen	17.27	14.23	20.30	9.62	16.15	26.34
NK	22.39	19.50	25.28	13.78	21.20	32.55
Other	22.28	19.46	25.10	13.77	21.11	32.35
Pioneer	23.07	20.21	25.93	14.36	21.89	33.32
Stine	18.09	15.06	21.13	11.17	17.14	26.24

Table 9. Own-price and cross-price elasticities of demand

elasticity	AS0	BE0	CR0	DE0	GO0	GR0	KR0	MY0	NK0	OT0	PI0	PU0	ST0	AS1	BE1	CH1	CR1	DE1	GO1	GR1	KR1	MY1	NK1	OT1	PI1	ST1
Asgrow0	-8.63	0.06	0.07	0.22	0.25	0.10	0.06	0.06	0.13	1.54	0.95	0.29	0.16	2.48	0.11	0.04	0.13	0.33	0.23	0.11	0.09	0.11	0.29	1.20	1.05	0.26
Beck0	0.31	-9.24	0.07	0.06	0.10	0.04	0	0.01	0.02	1.48	1.00	0.10	0.14	0.77	3.39	0.10	0.13	0.12	0.08	0.05	0.01	0.04	0.23	1.01	0.95	0.13
Croplan0	0.42	0.06	-9.72	0.20	0.21	0.02	0.11	0.12	0.20	2.01	1.03	0.37	0.22	0.99	0.12	0.05	1.39	0.29	0.24	0.03	0.09	0.13	0.37	1.30	1.06	0.26
Dekalb0	0.93	0.04	0.15	-9.73	0.37	0.09	0.10	0.11	0.16	1.84	1.12	0.38	0.26	0.81	0.07	0	0.13	1.65	0.33	0.07	0.09	0.10	0.21	1.16	0.92	0.24
Golden0	0.64	0.04	0.09	0.23	-8.73	0.07	0.09	0.08	0.17	1.77	0.86	0.29	0.21	0.75	0.08	0	0.13	0.33	2.06	0.08	0.10	0.10	0.25	1.23	0.81	0.23
Growmark0	0.82	0.06	0.03	0.17	0.21	-9.47	0.06	0.07	0.14	1.50	1.18	0.25	0.18	0.86	0.16	0.06	0.12	0.26	0.23	2.18	0.09	0.08	0.25	1.05	1.07	0.22
Kruger0	0.65	0	0.19	0.23	0.36	0.08	-8.31	0.12	0.20	1.83	1.01	0.23	0.30	0.55	0.01	0	0.14	0.28	0.27	0.07	2.09	0.11	0.23	0.89	0.71	0.21
Mycogen0	0.48	0.02	0.14	0.20	0.25	0.07	0.10	-8.13	0.17	1.23	0.78	0.30	0.18	0.66	0.03	0	0.10	0.30	0.25	0.06	0.07	2.01	0.18	1.09	0.73	0.18
Nk0	0.53	0.01	0.13	0.15	0.26	0.07	0.08	0.08	-9.21	1.87	1.05	0.34	0.14	0.78	0.04	0.01	0.22	0.33	0.30	0.08	0.10	0.10	1.86	1.34	1.03	0.27
Other0	0.34	0.06	0.08	0.10	0.16	0.04	0.04	0.04	0.10	-6.84	0.73	0.31	0.16	0.84	0.10	0.07	0.17	0.20	0.17	0.05	0.07	0.10	0.38	3.33	1.01	0.20
Pioneer0	0.42	0.08	0.08	0.12	0.15	0.06	0.05	0.04	0.11	1.46	-8.35	0.23	0.14	0.85	0.13	0.07	0.15	0.25	0.18	0.07	0.07	0.09	0.36	1.15	3.37	0.19
Public0	0.38	0.02	0.09	0.13	0.15	0.04	0.03	0.05	0.11	1.92	0.71	-4.90	0.14	0.87	0.05	0.05	0.22	0.19	0.15	0.06	0.04	0.09	0.44	1.56	1.08	0.22
Stine0	0.53	0.09	0.13	0.22	0.28	0.08	0.10	0.08	0.12	2.41	1.08	0.35	-10.29	1.04	0.08	0.09	0.17	0.27	0.25	0.08	0.12	0.15	0.40	1.34	1.21	1.58
Asgrow1	0.10	0.01	0.01	0.01	0.01	0	0	0	0.01	0.17	0.08	0.03	0.01	-9.15	0.19	0.19	0.39	0.39	0.22	0.12	0.10	0.16	0.91	2.67	2.53	0.28
Beck1	0.04	0.23	0.01	0.01	0.01	0.01	0	0	0	0.16	0.11	0.01	0.01	1.60	-8.53	0.19	0.24	0.23	0.14	0.12	0.04	0.09	0.59	2.22	2.00	0.24
Channel1	0.02	0.01	0.01	0	0	0	0	0	0	0.19	0.08	0.02	0.02	2.35	0.26	-10.68	0.35	0.10	0.03	0.13	0.11	0.24	1.05	2.56	2.68	0.30
Croplan1	0.04	0.01	0.07	0.01	0.01	0	0	0	0.02	0.22	0.10	0.05	0.02	2.66	0.20	0.20	-12.11	0.38	0.29	0.07	0.15	0.23	1.21	3.12	2.70	0.43
Dekalb1	0.09	0.01	0.01	0.11	0.04	0.01	0.01	0.01	0.02	0.26	0.16	0.04	0.02	2.64	0.19	0.06	0.38	-11.46	0.53	0.14	0.17	0.18	0.87	2.88	2.36	0.44
Golden1	0.08	0	0.01	0.03	0.26	0.01	0.01	0.01	0.02	0.25	0.13	0.04	0.02	1.67	0.13	0.02	0.33	0.61	-9.88	0.13	0.17	0.17	0.78	2.51	1.85	0.39
Growmark1	0.08	0.01	0	0.01	0.02	0.21	0.01	0.01	0.01	0.18	0.13	0.03	0.02	2.20	0.26	0.20	0.20	0.39	0.31	-10.89	0.18	0.13	0.81	2.25	2.40	0.39
Kruger1	0.07	0	0.01	0.02	0.03	0.01	0.16	0.01	0.02	0.23	0.12	0.02	0.03	1.78	0.07	0.16	0.38	0.44	0.39	0.17	-9.90	0.20	0.85	2.27	2.14	0.45
Mycogen1	0.05	0	0.01	0.01	0.02	0	0.01	0.12	0.01	0.22	0.10	0.03	0.03	1.89	0.12	0.23	0.40	0.29	0.25	0.08	0.13	-9.86	0.86	2.47	2.20	0.35
Nk1	0.03	0	0.01	0.01	0.01	0	0	0	0.05	0.19	0.09	0.04	0.01	2.30	0.17	0.22	0.45	0.32	0.25	0.11	0.12	0.19	-10.50	2.90	2.64	0.33
Other1	0.03	0.01	0.01	0.01	0.01	0	0	0	0.01	0.45	0.08	0.03	0.01	1.86	0.18	0.15	0.32	0.29	0.23	0.09	0.09	0.15	0.80	-7.51	2.09	0.27
Pioneer1	0.03	0.01	0.01	0.01	0.01	0	0	0	0.01	0.16	0.26	0.03	0.01	2.01	0.19	0.18	0.31	0.27	0.19	0.10	0.10	0.15	0.83	2.36	-7.75	0.27
Stine1	0.09	0.01	0.01	0.02	0.03	0.01	0.01	0.01	0.02	0.31	0.15	0.06	0.15	2.25	0.22	0.21	0.49	0.50	0.39	0.17	0.20	0.24	1.05	3.13	2.77	-11.85

Table 10. Model results after reshuffling the choice sequences

	Clogit b/(se)	Mixlogit b/(se)
Coef		
price	-0.188***	-0.407***
	(0.015)	(0.019)
control	0.123***	0.315***
	(0.014)	(0.018)
state	1.268***	0.384***
	(0.028)	(0.021)
GT 1999-2004	3.305***	6.935***
	(0.189)	(0.267)
GT 2005-2010	4.792***	9.937***
	(0.235)	(0.525)
GT 2011-2016	3.316***	7.611***
	(0.147)	(0.380)
mprice_p		-0.947***
		(0.051)
SD		
price		0.129***
		(0.008)
state		0.291***
		(0.066)
GT 1999-2004		2.350***
		(0.132)
GT 2005-2010		2.819***
		(0.287)
GT 2011-2016		3.595***
		(0.214)
LL	-143542	-118238
N	1057637	1057637

Note: standard errors are clustered at CRD level.
*p<0.05, **p<0.01, ***p<0.001.

Table 11. Model results under different specifications of the price coefficient

	Fixed price b/(se)	Price by acre b/(se)	Normal b/(se)	Log-normal b/(se)
Coef				
price	-0.304***	-0.307***	-0.326***	-0.437***
	(0.022)	(0.024)	(0.026)	(0.022)
price_acre5		0.004		
		(0.005)		
control	0.225***	0.227***	0.234***	0.349***
	(0.021)	(0.022)	(0.025)	(0.019)
state	1.953***	1.936***	2.107***	2.092***
	(0.057)	(0.052)	(0.050)	(0.052)
GT 1999-2004	5.418***	5.449***	6.140***	7.470***
	(0.295)	(0.328)	(0.383)	(0.314)
GT 2005-2010	8.665***	8.717***	9.232***	10.878***
	(0.375)	(0.422)	(0.431)	(0.501)
GT 2011-2016	7.133***	7.118***	7.298***	9.214***
	(0.263)	(0.269)	(0.440)	(0.445)
SD				
price			0.142***	0.145***
			(0.004)	(0.006)
state	1.000***	0.985***	1.075***	1.043***
	(0.045)	(0.039)	(0.041)	(0.038)
GT 1999-2004	2.344***	2.379***	2.553***	2.508***
	(0.109)	(0.101)	(0.164)	(0.185)
GT 2005-2010	3.159***	3.217***	3.459***	3.271***
	(0.180)	(0.205)	(0.227)	(0.242)
GT 2011-2016	4.070***	4.061***	3.980***	4.283***
	(0.219)	(0.223)	(0.328)	(0.278)
LL	-111843	-111817	-109332	-109647
N	1057637	1057637	1057637	1057637

Note: standard errors are clustered at CRD level. *p<0.05, **p<0.01, ***p<0.001.

Table 12. WTP estimates under different specifications of the price coefficient

WTP	Fixed price	Price by acre	Normal	Log-normal
state	6.42	6.31	6.46	5.31
GT 1999-2004	17.82	17.75	18.83	18.95
GT 2005-2010	28.50	28.39	28.32	27.6
GT 2011-2016	23.46	23.19	22.39	23.38

Table 13. Changes in brand market shares for price discount

Brand	State	Share	Change in Share due to No GT from 1996-1999						
		1999	1999	2000	2001	2002	2004	2008	2016
Asgrow	state	0.138	0.065	0.046	0.035	0.032	0.020	0.015	0.010
	nostate	0.138	0.102	0	0	0	0	0	0
Beck's	state	0.012	0.005	0.001	0	0	0	0	0
	nostate	0.016	0.007	0	0	0	0	0	0
Croplan	state	0.017	0.014	0.004	0.004	0.003	0.005	0.001	0
	nostate	0.029	0.024	0	0	0	0	0	0
DeKalb	state	0.035	0.032	0.004	0.003	0.003	0.003	0.002	0
	nostate	0.068	0.063	0	0	0	0	0	0
Golden	state	0.042	0.030	0.007	0.006	0.004	0.003	0.001	0
	nostate	0.067	0.053	0	0	0	0	0	0
Growmark	state	0.017	0.009	0.000	0.001	0.001	0	0.001	0
	nostate	0.017	0.014	0	0	0	0	0	0
Kruger	state	0.012	0.007	0	0	0	0	0	0
	nostate	0.017	0.012	0	0	0	0	0	0
Mycogen	state	0.019	0.009	0.001	0.001	0	0.001	0	0
	nostate	0.022	0.014	0	0	0	0	0	0
NK	state	0.024	0.020	0.007	0.005	0.004	0.001	0.002	0
	nostate	0.038	0.035	0	0	0	0	0	0
Other	state	0.345	0.092	0.042	0.035	0.031	0.031	0.014	0.007
	nostate	0.280	0.140	0	0	0	0	0	0
Pioneer	state	0.261	0.086	0.046	0.039	0.036	0.024	0.015	0.008
	nostate	0.189	0.122	0	0	0	0	0	0
Public	state	0.046	0.013	0.001	0	0	0	0	0
	nostate	0.071	0.020	0	0	0	0	0	0
Stine	state	0.034	0.025	0.003	0.002	0.002	0.001	0.000	0.000
	nostate	0.049	0.044	0	0	0	0	0	0

Table 14. Changes in brand market shares for late GT adoption

Brand	State	Share	Change in Share due to No GT from 1996-1999						
		1999	1999	2000	2001	2002	2004	2008	2016
Asgrow	state	0.138	-0.119	-0.071	-0.059	-0.051	-0.037	-0.029	-0.017
	nostate	0.138	-0.103	0	0	0	0	0	0
Beck's	state	0.012	-0.008	-0.002	-0.001	-0.001	-0.001	-0.001	0
	nostate	0.016	-0.011	0	0	0	0	0	0
Croplan	state	0.017	-0.012	-0.003	-0.003	-0.003	-0.002	0	0
	nostate	0.029	-0.019	0	0	0	0	0	0
DeKalb	state	0.035	-0.027	-0.003	-0.002	-0.002	-0.003	-0.001	0
	nostate	0.068	-0.051	0	0	0	0	0	0
Golden	state	0.042	-0.031	-0.007	-0.007	-0.007	-0.006	-0.001	0
	nostate	0.067	-0.046	0	0	0	0	0	0
Growmark	state	0.017	-0.014	-0.005	-0.005	-0.005	-0.003	-0.002	-0.001
	nostate	0.017	-0.013	0	0	0	0	0	0
Kruger	state	0.012	-0.008	-0.003	-0.002	-0.001	-0.001	-0.001	0
	nostate	0.017	-0.010	0	0	0	0	0	0
Mycogen	state	0.019	-0.010	-0.002	-0.002	-0.001	-0.002	-0.001	0
	nostate	0.022	-0.013	0	0	0	0	0	0
NK	state	0.024	-0.020	-0.007	-0.004	-0.004	-0.003	-0.003	-0.001
	nostate	0.038	-0.029	0	0	0	0	0	0
Other	state	0.345	-0.238	-0.127	-0.097	-0.085	-0.055	-0.034	-0.024
	nostate	0.280	-0.168	0	0	0	0	0	0
Pioneer	state	0.261	-0.216	-0.154	-0.114	-0.097	-0.072	-0.052	-0.032
	nostate	0.189	-0.128	0	0	0	0	0	0
Public	state	0.046	0	0	0	0	0	0	0
	nostate	0.071	0	0	0	0	0	0	0
Stine	state	0.034	-0.027	-0.007	-0.004	-0.004	-0.003	-0.001	-0.001
	nostate	0.049	-0.035	0	0	0	0	0	0

Appendices

Appendix A. Full results of Table 7

Table A. Full estimation Results of Table 7 in the main text

	Clogit b/(se)	No inertia b/(se)	No IV b/(se)	No initial b/(se)	Main b/(se)
Coef					
price	-0.199***	-0.404***	-0.104***	-0.439***	-0.437***
	(0.016)	(0.021)	(0.006)	(0.022)	(0.022)
control	0.138***	0.315***		0.353***	0.349***
	(0.014)	(0.019)		(0.020)	(0.019)
state	2.363***		2.095***	2.500***	2.092***
	(0.040)		(0.051)	(0.058)	(0.052)

Table A. Continued

GT 1990-2004	3.442***	6.776***	3.280***	7.635***	7.470***
	(0.197)	(0.326)	(0.158)	(0.297)	(0.314)
GT 2005-2010	4.933***	9.798***	6.127***	11.256***	10.878***
	(0.246)	(0.422)	(0.301)	(0.499)	(0.501)
GT 2011-2016	3.405***	7.434***	5.811***	8.516***	9.214***
	(0.146)	(0.455)	(0.319)	(0.443)	(0.445)
Asgrow	3.477***	9.093***	2.616***	10.442***	10.179***
	(0.416)	(0.711)	(0.373)	(0.679)	(0.757)
Beck	2.929***	7.027***	0.765*	9.399***	8.382***
	(0.416)	(0.687)	(0.385)	(0.674)	(0.768)
Channel	2.741***	5.247***	-0.789	7.253***	6.891***
	(0.385)	(0.887)	(0.560)	(0.749)	(0.865)
Croplan	2.456***	7.418***	1.084**	9.186***	8.668***
	(0.427)	(0.737)	(0.377)	(0.683)	(0.737)
Dekalb	0.414	4.334***	0.05	5.970***	5.661***
	(0.377)	(0.654)	(0.422)	(0.622)	(0.729)
Golden	0.767	4.043***	-0.511	6.391***	5.861***
	(0.426)	(0.774)	(0.507)	(0.687)	(0.766)
Growmark	2.536***	7.334***	1.064**	9.869***	8.958***
	(0.408)	(0.755)	(0.398)	(0.666)	(0.767)
Kruger	2.013***	5.486***	-0.247	7.990***	7.164***
	(0.429)	(0.773)	(0.465)	(0.674)	(0.808)
Mycogen	1.718***	5.332***	0.021	7.535***	6.807***
	(0.379)	(0.695)	(0.408)	(0.660)	(0.757)
NK	2.647***	7.743***	1.497***	9.167***	8.825***
	(0.383)	(0.681)	(0.367)	(0.647)	(0.739)
Other	2.764***	7.795***	2.185***	9.146***	8.783***
	(0.371)	(0.663)	(0.355)	(0.632)	(0.708)
Pioneer	2.809***	7.794***	2.160***	9.441***	9.092***
	(0.373)	(0.658)	(0.366)	(0.635)	(0.720)
Stine	1.661***	5.753***	0.625	7.402***	7.132***
	(0.379)	(0.662)	(0.403)	(0.632)	(0.735)
dealer	-0.150**	-0.03	-0.069	-0.029	-0.073
	(0.053)	(0.079)	(0.060)	(0.060)	(0.060)
ind_dealer	0.237***	0.320***	0.197**	0.201**	0.187**
	(0.061)	(0.085)	(0.068)	(0.069)	(0.070)
rep	-0.016	-0.046	0.088**	-0.086**	-0.107***
	(0.025)	(0.040)	(0.028)	(0.030)	(0.030)
ind_rep	0.106**	0.318***	0.073	0.173***	0.180***
	(0.040)	(0.058)	(0.045)	(0.044)	(0.047)
other	0.034	0.093	0.065	-0.094	-0.03
	(0.049)	(0.068)	(0.059)	(0.061)	(0.057)
ind_other	-0.02	0.250**	0.027	0.258***	0.119
	(0.058)	(0.084)	(0.071)	(0.068)	(0.063)
Asgrow_initial	0.233***	1.560***	0.401***		0.334***
	(0.034)	(0.123)	(0.040)		(0.047)
Beck_initial	0.911***	4.684***	2.762***		2.906***
	(0.135)	(0.388)	(0.320)		(0.270)

Table A. Continued

Channel_initial	0.649**	6.169***	3.064***		2.758***
	(0.225)	(0.861)	(0.461)		(0.339)
Croplan_initial	0.701***	2.238***	1.230***		0.998***
	(0.094)	(0.553)	(0.132)		(0.138)
Dekalb_initial	0.384***	1.937***	0.631***		0.525***
	(0.061)	(0.110)	(0.075)		(0.071)
Golden_initial	0.783***	2.984***	1.549***		1.213***
	(0.081)	(0.319)	(0.167)		(0.126)
Growmark_initial	1.260***	3.568***	1.877***		1.794***
	(0.093)	(0.363)	(0.185)		(0.124)
Kruger_initial	1.504***	4.589***	2.348***		2.584***
	(0.171)	(0.452)	(0.309)		(0.229)
Mycogen_initial	1.727***	4.378***	2.565***		2.958***
	(0.119)	(0.293)	(0.254)		(0.167)
NK_initial	0.540***	1.934***	0.946***		0.766***
	(0.049)	(0.165)	(0.082)		(0.095)
Other_initial	0.365***	1.732***	0.516***		0.556***
	(0.031)	(0.103)	(0.054)		(0.045)
Pioneer_initial	0.486***	2.342***	0.798***		0.587***
	(0.032)	(0.112)	(0.042)		(0.045)
Public_initial	0.396**	1.734***	1.362***		0.852**
	(0.143)	(0.332)	(0.250)		(0.269)
Stine_initial	0.951***	2.382***	1.344***		1.047***
	(0.098)	(0.194)	(0.107)		(0.112)
Asgrow 1999-2004	-2.070***	-5.438***	-1.059**	-6.424***	-5.974***
	(0.310)	(0.572)	(0.395)	(0.585)	(0.666)
Asgrow 2005-2010	-2.503***	-5.149***	-1.719***	-6.073***	-5.565***
	(0.279)	(0.523)	(0.444)	(0.636)	(0.673)
Beck 1999-2004	-2.029***	-6.305***	-1.034*	-7.249***	-6.721***
	(0.356)	(0.648)	(0.427)	(0.632)	(0.692)
Beck 2005-2010	-2.367***	-5.289***	-1.707***	-6.106***	-5.648***
	(0.291)	(0.576)	(0.457)	(0.652)	(0.686)
Channel 2005-2010	-0.071	-0.891	0.172	-0.753	-0.401
	(0.205)	(0.500)	(0.462)	(0.552)	(0.615)
Croplan 1999-2004	-1.757***	-5.132***	-0.597	-6.170***	-5.724***
	(0.330)	(0.601)	(0.391)	(0.588)	(0.659)
Croplan 2005-2010	-2.320***	-4.939***	-1.353**	-6.064***	-5.535***
	(0.301)	(0.536)	(0.446)	(0.634)	(0.677)
Dekalb 1999-2004	0.326	-1.607**	0.895*	-2.745***	-2.337***
	(0.307)	(0.557)	(0.451)	(0.568)	(0.662)
Dekalb 2005-2010	-0.827**	-2.199***	-0.248	-3.420***	-2.891***
	(0.313)	(0.533)	(0.510)	(0.655)	(0.702)
Golden 1999-2004	0.12	-1.690**	1.217*	-3.375***	-2.935***
	(0.352)	(0.634)	(0.502)	(0.631)	(0.678)
Golden 2005-2010	-1.179***	-2.411***	-0.191	-4.089***	-3.583***
	(0.319)	(0.573)	(0.509)	(0.667)	(0.687)
Growmark 1999-2004	-1.980***	-5.381***	-0.827*	-6.702***	-6.189***
	(0.332)	(0.606)	(0.410)	(0.570)	(0.667)

Table A. Continued

Growmark 2005-2010	-2.826***	-5.524***	-1.913***	-6.838***	-6.289***
	(0.314)	(0.562)	(0.444)	(0.643)	(0.685)
Kruger 1999-2004	-2.177***	-6.043***	-0.738	-7.132***	-6.657***
	(0.395)	(0.645)	(0.441)	(0.630)	(0.744)
Kruger 2005-2010	-2.739***	-5.513***	-1.599***	-6.724***	-6.209***
	(0.351)	(0.584)	(0.482)	(0.664)	(0.736)
Mycogen 1999-2004	-1.646***	-4.604***	-0.589	-5.604***	-5.217***
	(0.319)	(0.564)	(0.408)	(0.579)	(0.674)
Mycogen 2005-2010	-2.690***	-5.198***	-1.968***	-6.120***	-5.764***
	(0.289)	(0.548)	(0.476)	(0.654)	(0.698)
NK 1999-2004	-2.067***	-5.542***	-1.113**	-6.381***	-6.008***
	(0.312)	(0.596)	(0.400)	(0.585)	(0.671)
NK 2005-2010	-1.973***	-4.442***	-1.051*	-5.429***	-4.908***
	(0.264)	(0.527)	(0.443)	(0.629)	(0.675)
Other 1999-2004	-1.601***	-4.689***	-0.613	-5.862***	-5.403***
	(0.304)	(0.577)	(0.393)	(0.577)	(0.657)
Other 2005-2010	-2.111***	-4.326***	-1.283**	-5.492***	-4.982***
	(0.261)	(0.517)	(0.440)	(0.626)	(0.664)
Pioneer 1999-2004	-1.553***	-4.698***	-0.813*	-5.521***	-5.112***
	(0.286)	(0.543)	(0.391)	(0.552)	(0.634)
Pioneer 2005-2010	-2.128***	-4.539***	-1.422**	-5.487***	-4.986***
	(0.265)	(0.518)	(0.444)	(0.626)	(0.666)
Stine 1999-2004	-1.357***	-3.853***	-0.352	-4.894***	-4.543***
	(0.316)	(0.551)	(0.414)	(0.547)	(0.661)
Stine 2005-2010	-2.295***	-4.429***	-1.556***	-5.386***	-4.940***
	(0.302)	(0.526)	(0.470)	(0.631)	(0.686)
SD					
price		0.132***	0.131***	0.138***	0.145***
		(0.022)	(0.011)	(0.007)	(0.006)
state			1.036***	1.267***	1.043***
			(0.007)	(0.037)	(0.038)
GT 1990-2004		-2.311***	2.360***	2.450***	-2.508***
		(0.208)	(0.129)	(0.141)	(0.185)
GT 2005-2010		-2.822***	3.189***	3.680***	-3.271***
		(0.201)	(0.293)	(0.241)	(0.242)
GT 2011-2016		-3.714***	4.153***	4.046***	4.283***
		(0.266)	(0.212)	(0.272)	(0.278)
Asgrow		1.336***	0.549***	0.416***	0.576***
		(0.054)	(0.033)	(0.053)	(0.036)
Beck		2.850***	1.969***	1.664***	2.060***
		(0.203)	(0.121)	(0.166)	(0.178)
Channel		3.610***	2.812***	2.436***	2.592***
		(0.398)	(0.405)	(0.362)	(0.300)
Croplan		1.488***	0.997***	0.777***	0.950***
		(0.226)	(0.103)	(0.095)	(0.066)
Dekalb		0.713***	-0.280**	-0.175	-0.115
		(0.115)	(0.104)	(0.105)	(0.264)
Golden		1.498***	0.952***	0.769***	0.903***
		(0.149)	(0.108)	(0.102)	(0.115)

Table A. Continued

Growmark		1.538***	0.943***	0.765***	0.914***
		(0.218)	(0.144)	(0.188)	(0.140)
Kruger		2.300***	1.608***	1.314***	1.526***
		(0.221)	(0.230)	(0.113)	(0.124)
Mycogen		2.072***	1.394***	1.415***	1.453***
		(0.205)	(0.150)	(0.103)	(0.131)
NK		1.251***	0.889***	0.691***	0.748***
		(0.102)	(0.069)	(0.071)	(0.085)
Other		1.205***	0.619***	0.523***	0.638***
		(0.076)	(0.055)	(0.036)	(0.054)
Pioneer		1.480***	0.520***	0.426***	0.448***
		(0.071)	(0.059)	(0.047)	(0.092)
Stine		1.201***	0.891***	0.735***	0.553***
		(0.141)	(0.096)	(0.076)	(0.138)
LL	-125277	-118707	-110442	-110960	-109647
N	1057637	1057637	1057637	1057637	1057637

Appendix B. Patterns in state dependence

In the models, we allow farmers to have heterogeneous state dependence attitude. As noted in section 5.5.1, this may be not enough so we conduct the reshuffling procedure. We find that most of what we captured is genuine state dependence rather than unobserved heterogeneity. However, there can be certain patterns in farmers’ state dependence attitude which may not only affect the distribution of state dependence coefficient but also impact the model results and our WTP estimation. In this part, we primarily consider two possible patterns in state dependence: brand specific state dependence and land size specific state dependence. As note by Dube, Hitsch and Rossi (2010), it can be expected that some brands with unique trademarks might display more state dependence. The land size specific specification is to capture any heterogeneous attitude of state dependence toward land size. It is reasonable that there is some implicit income effect. Also, as seed companies and dealers value their relationship with large farm operations, the situation may affect their brand loyalty attitude.

Table B presents the model results. The first column presents the results with brand specific state dependence. Specifically, we let the state dependence term interact with each brand-specific intercept and further choose Asgrow as the baseline. Therefore, the reported "state" is the state dependence coefficient for Asgrow; the reported "state_brand" is the difference between farmers' state dependence attitude for the brand and for Asgrow. The brand specific state dependence are generally estimated with larger variance (except for Pioneer and Other, the two largest brand in our sample). The results show that the state dependence coefficient can be as high as 3.38 (WTP approximately 7.49) for Mycogen and as low as 1.17 (WTP approximately 2.59) for Channel.

The second column considers land size specific state dependence. Same as section 5.6.2, we classify farmers into two categories: one with land size no less than 500 acres and the other with all remaining farmers. The variable is further interacted with the state dependence term where the category of no less than 500 acres is chosen as the baseline. The results show that farmers with less land display more brand loyalty—the difference is small but significant. It can be the results that farmers with larger land size can try seeds of different brands to keep updated with new varieties. Despite the results that farmers with different characteristics may have different attitude toward state dependence, we observed little improvement in the log-likelihood and little changes in all other variables. Our WTP estimates for each brand and trait are quite stable despite the model specification of patterns in state dependence.

Table B. Patterns in state dependence attitude

	By brand b/(se)	By acre b/(se)	Main b/(se)
Coef			
price	-0.451***	-0.441***	-0.437***
	(0.023)	(0.020)	(0.022)
control	0.361***	0.355***	0.349***
	(0.021)	(0.019)	(0.019)

Table B. Continued

GT 1990-2004	7.643***	7.663***	7.470***
	(0.288)	(0.327)	(0.314)
GT 2005-2010	11.585***	10.766***	10.878***
	(0.622)	(0.449)	(0.501)
GT 2011-2016	8.609***	8.853***	9.214***
	(0.478)	(0.393)	(0.445)
state	1.797***	1.921***	2.092***
	(0.070)	(0.064)	(0.052)
state_acre5		0.245***	
		(0.048)	
state_Beck	0.39		
	(0.313)		
state_Channel	-0.627*		
	(0.266)		
state_DeKalb	0.366***		
	(0.084)		
state_Golden	0.720***		
	(0.187)		
state_Growmark	0.954***		
	(0.261)		
state_Kroger	0.454		
	(0.244)		
state_Mycogen	1.583***		
	(0.260)		
state_NK	0.344**		
	(0.111)		
state_Other	-0.048		
	(0.070)		
state_Pioneer	0.281***		
	(0.075)		
state_Public	-0.046		
	(0.260)		
state_Stine	1.155***		
	(0.122)		
SD			
price	0.143***	0.139***	0.145***
	(0.008)	(0.006)	(0.006)
state	1.025***	1.031***	1.043***
	(0.037)	(0.040)	(0.038)
GT 1990-2004	2.473***	-2.833***	-2.508***
	(0.128)	(0.155)	(0.185)
GT 2005-2010	3.697***	3.124***	-3.271***
	(0.272)	(0.223)	(0.242)
GT 2011-2016	4.078***	4.092***	4.283***
	(0.304)	(0.215)	(0.278)
LL	-109261	-109567	-109647
N	1057637	1057637	1057637

CHAPTER 4.   GENERAL CONCLUSION

In this dissertation, I investigate consumers' behavior in two oligopoly markets, the fuel market and the seed market. Chapter 2 is an applied theory work, where I build, calibrate, and simulate a model of consumers' heterogeneous preferences over two differentiated products, E10 and E85. Chapter 3 empirically models farmers' heterogeneous preference over different seed products, with an emphasize on their brand inertia behavior and their WTP for the new genetically engineered innovation. Through the studies in consumer demand, the two projects further investigate some implications at the market level. The pass-through work evaluates the Renewable Fuel Standard program through the pass-through rate of the policy-induced subsidy to the retail prices. It finds that the pass-through is generally incomplete because of the monopoly power enjoyed by E85 stations, especially as a result of the scarcity of E85 stations. The brand inertia work quantifies the level of state dependence and heterogeneity and investigates the long run implications on the market demand. It shows that farmers have a substantial WTP for previously purchased brand and the GT trait over time, however there is substantial heterogeneity in these WTPs. Through counterfactual analysis, I find that state dependence implies long-lasting effect of a temporary shock on the supply side, which further suggests dynamic pricing behavior of seed companies.