NET Institute*

www.NETinst.org

Working Paper #10-20

September 2010

Network Effects in Alternative Fuel Adoption: Empirical Analysis of the Market for Ethanol

Scott K. Shriver
GSB, Stanford University

* The Networks, Electronic Commerce, and Telecommunications ("NET") Institute, http://www.NETinst.org, is a non-profit institution devoted to research on network industries, electronic commerce, telecommunications, the Internet, "virtual networks" comprised of computers that share the same technical standard or operating system, and on network issues in general.

Network Effects in Alternative Fuel Adoption: Empirical Analysis of the Market for Ethanol

Scott K. Shriver*

September 30, 2010

Abstract

This paper investigates the importance of network effects in the demand for ethanol-compatible vehicles and the supply of ethanol fuel retailers. An indirect network effect, or positive feedback loop, arises in this context due to spatially-dependent complementarities in the availability of ethanol fuel and the installed base of ethanol-compatible vehicles. Marketers and social planners are interested in whether these effects exist, and if so, how policy might accelerate adoption of the ethanol fuel standard within a targeted population. To measure these feedback effects, I develop an econometric framework that considers the simultaneous determination of ethanol-compatible vehicle demand and ethanol fuel supply in local markets. The demand-side of the model considers the automobile purchase decisions of consumers and fleet operators, and the supply-side model considers the ethanol market entry decisions of competing fuel retailers. I propose new estimators that address the endogeneity induced by the co-determination of alternative fuel vehicle demand and alternative fuel supply. I estimate the model using zip code level panel data from six states over a six year period. I find the network effect to be highly significant, both statistically and economically. Under typical market conditions, entry of an additional ethanol fuel retailer leads to a 12% increase in consumer demand for ethanol-compatible vehicles. The entry model estimates imply that a monopolist requires a local installed base of at least 204 ethanol-compatible vehicles to be profitable. As an application, I demonstrate how the model estimates can inform the promotional strategy of a vehicle manufacturer. Counterfactual simulations indicate that subsidizing fuel retailers to offer ethanol can be an effective policy to indirectly increase ethanol-compatible vehicle sales.

Keywords: ethanol, flex-fuel vehicles, indirect network effects, market entry

*Graduate School of Business, 518 Memorial Way, Stanford University, Stanford, CA 94305. Email: scott.shriver@gsb.stanford.edu. I wish to thank Harikesh Nair, Wes Hartmann, Peter Reiss, Han Hong, and Jon Levin for their guidance and comments. I gratefully acknowledge Margot Melendez at the National Renewable Energy Laboratory (NREL) for assistance in obtaining E85 station data and Tim Gerlach at the American Lung Association of the Upper Midwest for many helpful discussions concerning the E85 supply chain. I also thank the Stanford GSB and the NET Institute for generous financial support. All errors are my own.

The problem right now is that the supply of ethanol is not anywhere near the demand. So the vast majority of the ethanol-capable vehicles we have on the road right now do not use ethanol simply because people don't know where to buy it.

- Rick Wagoner, (former) General Motors Chairman

There's a chicken-and-egg proposition here. If you don't have enough flex-fuel vehicles, there's less incentive to make [ethanol] fuel and sell it at a retail level.

- Mark Hamerlinck, Minnesota Corn Growers Association

1 Introduction

The demand for many goods depends upon the availability of a complementary product. A canonical example is that demand for computer hardware depends upon the availability of compatible software (Katz and Shapiro, 1985, 1994; Farrell and Saloner, 1985; Church and Gandal, 1992). In these systems, fulfilling consumer demand for a "whole product" (e.g., computing services) requires provision of both a durable good (hardware) and one or more consumption goods (software) (Moore, 1995; Gupta, Jain, and Sawhney, 1999). The interdependence of demand for complementary goods implies that firms may face a chicken-or-egg problem when a new product is introduced (Katz and Shapiro, 1994; Caillaud and Jullien, 2003). That is, software firms will not enter the market if few consumers have purchased compatible hardware systems, and consumers will not purchase new hardware unless software is readily available. Complementary product suppliers may overcome this problem by entering joint marketing agreements that align the firms' financial incentives and coordinate their product distribution and promotion strategies. Agreements of this type are particularly attractive, since growth in the "installed base" of hardware (i.e., *cumulative* demand) leads more firms to enter the market for compatible software, which in turn leads to more hardware sales, and so on. This positive feedback cycle, or indirect network effect,[1] generates demand-side economies of scale for suppliers of the complementary goods, compounding investment returns on costs to implement the marketing policy (Nair, Chintagunta, and Dubé, 2004; Karaca-Mandic, 2004).

In this paper, I examine the role of indirect network effects in the market for ethanol fuel. Although ethanol is a common component in all blends of motor fuel (retail gasoline typically contains up to 10% ethanol), my inquiry relates specifically to E85, a standard blend containing 85% ethanol and 15% regular gasoline. E85 is classified as an alternative fuel that may only be used in flex-fuel vehicles (FFVs), which are engineered to accept high-alcohol fuels. In this market, flex-fuel vehicles represent the hardware side of the system, whereas locations of E85 retailers define the

[1] The effect is deemed "indirect" as consumer utility for hardware increases with greater availability of complementary software, which depends (indirectly) on the total number of consumers that have adopted the hardware system. By contrast, a "direct" network effect results when utility for a product increases in (direct) response to other consumers adopting the product. Communication devices such as fax machines and cell phones are typical examples of products that exhibit direct network effects.

availability of software. An indirect network effect arises due to the fact that as more consumers and fleets[2] purchase flex-fuel vehicles, more fuel retailers are likely to enter the E85 market, and vice-versa. Unlike the more commonly studied case of complementary high tech products, where utility for hardware depends on the distance of compatible software *manufacturers* from the consumer's ideal points in product feature "space," utility for an alternative fuel vehicle is determined by the physical proximity of alternative fuel *distributors* to the consumer's physical location. Feedback in adoption of the ethanol fuel standard[3] therefore operates within highly local markets and, due to the relative infrequency of consumer vehicle replacement, over extended periods of time. These factors make separate identification of the network effects from unobserved factors a challenging empirical exercise. My first objective for the study is therefore to consistently measure each "side" of the feedback loop: the effect of E85 availability on flex-fuel vehicle demand, and the effect of the flex-fuel vehicle installed base on the supply of E85. My second objective is to explore the marketing policy implications of these measurements for suppliers of flex-fuel vehicles and E85 fuel. A third objective of the paper is to develop new methods to permit inference of simultaneous supply and demand in markets with endogenous markets sizes, such as those governed by network effects.

To achieve these goals, I develop an equilibrium model of flex-fuel vehicle demand and E85 market entry. The model may be interpreted as a two-sided model of technology adoption, with observations of flex-fuel demand or E85 market entry representing agent decisions to adopt the ethanol fuel standard. I model flex-fuel demand as a utility-maximizing discrete choice of whether or not to purchase an ethanol-compatible vehicle. Utility for flex-fuel is a function of the availability of ethanol fuel, which I operationalize as the number of retail service stations offering E85. E85 market entry is modeled as a competitive game of complete information played amongst potentially entering retailers, in the spirit of Bresnahan and Reiss (1991). These models infer retailer profit functions from observations of the number of entering firms, using exogenous variation in the market size to separately identify variable and fixed cost factors. I extend the Bresnahan and Reiss (1991) framework to allow for an endogenously determined fuel retailer market size (the flex-fuel installed base) and to exploit demand-side data to better pin down firms' profit functions. This approach follows a recent marketing literature that demonstrates how empirical models of product introduction may be enriched by augmenting demand-side information (e.g. Reiss and Spiller, 1989; Draganska and Mazzeo, 2003; Ellickson and Misra, 2007; Musalem and Shin, 2009).

I estimate the model using an extraordinarily rich panel dataset, which encompasses the entire population of flex-fuel vehicle purchase and E85 market entry events in roughly 7,000 zip codes over the six year period 2001-2006. Access to zip code level data is critically important in my application,

[2] Fleet vehicles are owned and operated by corporations or government agencies, as opposed to individual consumers. For example, vehicle fleets are operated by rental car agencies, taxicab companies, and municipal police departments. Due to systematic differences in consumer and fleet behavior, I model demand for flex-fuel vehicles from these populations separately.

[3] In the terminology of the technology standards literature (e.g. Farrell and Saloner, 1985), the 85% (15%) blend of ethanol (gasoline) defines a standard of interoperability between the vehicle/fuel system, in the same manner that the DVD format defines an interoperability standard for media players and titles. Since flex-fuel vehicles are backwardly compatible with regular gasoline, FFV owners obtain an option value arising from the ability to use either fuel standard.

as it provides direct observation of feedback effects that operate within highly localized markets. Having repeated observations allows me to control for unobservables that are potentially correlated with entry, by using a rich specification of market and period fixed effects. Additional concerns for endogeneity may arise due to residual correlation between firms' entry decisions and market *and* time-specific unobservables. I address this additional concern using tools from the econometrics literature on panel data methods (e.g. Chamberlain, 1984; Arellano and Bond, 1991). Under the testable assumption that past levels of E85 availability do not influence current vehicle choices, appropriately lagged values of the number of E85 retailers are used as additional instruments for the current number of E85 retailers in the vehicle demand equations.

Estimation proceeds in two steps. In the first step, I estimate the flex-fuel demand parameters using the "system GMM" estimator of Blundell and Bond (1998), instrumenting for the number of E85 retailers in the market. In the second step, I estimate the market entry model conditional upon the first stage demand estimates. The two-step approach clarifies how the estimator works in practice. Intuitively, access to demand data along with instruments in the first stage enables the econometrician to consistently "back-out" the distribution of demand-side shocks. Parameters capturing the covariance of supply-side shocks with the demand shocks can then be estimated with the entry model parameters in the second step. Given a guess of these parameters, one can simulate market shocks and solve the reduced form of the system for the equilibrium number of E85 retailers. By averaging outcomes from multiple draws of the market shocks, I obtain the expected number of E85 retailers in equilibrium, given the entry and covariance parameters. Solving the reduced form fully accounts for the co-determination of E85 market entry and flex-fuel demand, and thus controls for the endogeneity of the flex-fuel installed base in the entry equation. Estimation of the entry model is then based on finding the parameter vector that minimizes the difference (in the sum of squares sense) in the observed number of E85 retailers and the model-predicted equilibrium number of firms. To correct for measurement error introduced in the second step estimation by using first stage parameters, I employ a nonparametric bootstrap over the two step procedure to obtain standard errors, in which market histories are sampled with replacement. I present several versions of the estimator that correspond to different econometric assumptions, as well as assumptions about firms' conduct.

I find evidence of a network effect with both statistical and economic significance. The demand estimates suggest that, under average market conditions, entry of an additional ethanol retailer leads to a 12.0% annual increase in consumer flex-fuel vehicle sales and a 25.6% increase in fleet flex-fuel vehicle sales. Similarly, the market entry model implies that an ethanol retailer requires an installed base of at least 204 consumer flex-fuel vehicles to operate profitably. The effect of the fleet vehicle installed base on ethanol retailer market entry is weaker than for the consumer installed base, with 7.1 fleet vehicles required to match the expected ethanol fuel consumption of one consumer flex-fuel vehicle. I find these results are robust to a variety of assumptions, including the competitive conduct of firms and the independence of market observations.

I use my estimates to assess three counterfactual policy scenarios. The first counterfactual is a descriptive exercise, aimed at quantifying the long-run impact of the network effect across a sample of heterogeneous markets. This simulation finds that in the final period studied, 27.5% of the installed base of flex-fuel vehicles and 9.4% of E85 market entry events result from the network effect. The second and third policy experiments explore strategies to improve flex-fuel vehicle manufacturer profitability by harnessing the network effect. Specifically, I evaluate two types of subsidies paid by vehicle manufacturers to fuel retailers to encourage E85 market entry. For the first type of subsidy, the vehicle manufacturer offers a fixed payment in each period to any fuel retailer that offers E85. For the second policy, the vehicle manufacturer selectively offers subsidies to markets in which the expected current period payoff is positive, i.e., profits from the incremental number of flex-fuel vehicles sold (due to the increased E85 availability) exceeds the cost to subsidize an additional entrant. In simulations, targeting incentive policies at the market level leads to a 50% increase in profitability over a fixed rate subsidy equal to 10% of E85 retailer average fixed costs.

The paper proceeds as follows. In Section 2, I discuss related literature and the positioning of my study. Section 3 describes the data collected for the analysis. I develop the econometric model in Section 4. In Section 5, I provide details of the estimation routine. Section 6 presents the main estimation results. In Section 7, I develop and summarize the counterfactual experiments. Section 8 concludes with a summary and a discussion of future directions for research.

2 Related literature

The study contributes to several streams of research. The most central of these is the empirical literature on indirect network effects.[4] Most studies of indirect network effects have focused on high technology products that conform to the standard concept of a hardware/software system advanced by Katz and Shapiro (1985). Among the products considered by these studies are VCRs (Park (2004), Ohashi (2003)), CD/DVD players (Gandal, Kende, and Rob (2000), Basu, Mazumdar, and Raj (2003), Karaca-Mandic (2004), Dranove and Gandal (2003)), PDAs (Nair, Chintagunta, and Dubé (2004)), and video game consoles (Shankar and Bayus (2003), Clements and Ohashi (2005), Liu (2006), Dubé, Hitsch, and Chintagunta (2007), Lee (2009)). Unlike these studies, I investigate network effects that operate at a local, rather than industry, level. Whereas the availability of ethanol fuel is determined by the physical location of retail E85 outlets, the relevant measure of software availability in these industries is the *variety* of software titles available, since the proliferation of retail and online outlets for these goods makes them essentially ubiquitous once released. As a consequence, most of these papers use industry aggregate time series data

[4] The network effects literature has its theoretical foundations in the works of Katz and Shapiro (1985), Katz and Shapiro (1986), Farrell and Saloner (1986) and Farrell and Saloner (1985). These papers formalize the concept of network effects as positive consumption externalities and explore the role of standards compatibility in technology adoption. Chou and Shy (1990), Church and Gandal (1992), Church and Gandal (1993), and Katz and Shapiro (1992) develop theoretical models of indirect network effects that arise through product complementary.

to identify the network effect. By contrast, my study exploits rich panel data for identification of highly localized feedback loops.

Empirical research of indirect network effects arising from spatially-dependent product complementarities has been less common. The literature on shopping malls examines retail demand externalities that arise when consumers are drawn to shopping centers to reduce their search costs for a variety of goods (e.g. Eppli and Benjamin (1994), Vitorino (2007)). The focus of these papers is to infer the benefits of agglomeration in retail outlets to the mall operators, store owners, and consumers. The motor fuel category I study is conceptually distinct from these works, however, as the key benefit to consumers is not the concentrated availability of many goods, but the distributed access to one type of good. Rysman (2004) investigates indirect network effects in the market for Yellow Pages directories, which have spatial complementarities by nature of their circulation areas. As in my study, Rysman (2004) measures both "sides" of a feedback loop, which in his context operates between the supply of advertisements in a directory and consumer use of the directory. However, Rysman (2004) does not explicitly consider firm market entry decisions, focusing instead on the role of network effects and competition among incumbent firms. More closely related to my model is that of Berry and Waldfogel (1999), who investigate the welfare implications of entry in radio broadcasting. The Berry and Waldfogel (1999) model measures the market expanding effects of station entry on the total number of radio listeners in a metro area. A limitation of the Berry and Waldfogel (1999) model is that it requires auxiliary data in order to estimate parameters that shift variable components of firm profits. My model is more suitable to contexts where both fixed and variable components of profit must be inferred solely from observations of market entry.

From a policy perspective, the study adds to the burgeoning literature on the economics of ethanol as a transportation fuel in the United States. Two recent studies include Anderson (2006), who estimates demand for E85 and finds strong evidence of fuel-switching behavior among flex-fuel vehicle owners, and Anderson and Sallee (2009) who present evidence that domestic automakers produce flex-fuel vehicles primarily as a means to lower the cost of complying with federal fuel-economy (CAFE) standards.[5] In a closely related paper, Corts (2010) investigates the influence of government fleet adoption on E85 market entry. Corts (2010) presents descriptive regressions of the number of firms on the installed bases of government fleet flex-fuel vehicles and consumer flex-fuel vehicles. Both studies find positive effects of the flex-fuel installed base on the number of E85 retailers, but differ in the estimates of the number of flex-fuel vehicles required to support an E85 retailer. These differences appear to originate in the datasets, which have similar geographic coverage but differ in time and format (Corts (2010) uses a cross-section). I compare the results in greater detail when presenting my results in Section 6.1.

[5] Vehicle manufacturers are granted generous CAFE credits for producing alternative fuel vehicles, including flex-fuel, offsetting lower fuel economy in the remainder of their product line.

3 Industry background and data

3.1 Flex-fuel vehicles

The Ford Model T, the first vehicle to be mass produced on an assembly line, could use blends of ethanol and gasoline in arbitrary proportion, similar to the flex-fuel vehicles of today. Throughout the early 20th century, service stations commonly offered blends dominated by each fuel type. However, by the 1930's gasoline became the de facto standard transportation fuel through declines in production cost, and ethanol compatibility was phased out of vehicle designs. Ethanol's resurgence in recent years has been spurred by a series of public policy interventions. The major policy instrument affecting the production of ethanol-compatible vehicles is the federal Corporate Average Fuel Economy (CAFE) standard. Enacted in the wake of the 1973 OPEC oil embargo, CAFE standards require vehicle manufacturers to maintain an overall fuel efficiency rating for the fleet of vehicles they supply. The Alternative Motor Fuels Act (AMFA) of 1988 granted vehicle manufacturers generous CAFE credits for the production of alternative fuel vehicles (AFVs), including those capable of utilizing ethanol. E85/gasoline flex-fuel vehicles have since become the dominant means by which manufacturers have captured AFV CAFE credits, representing 90% of all alternative fuel vehicles supplied from 1998 to 2006.[6] The cost of modifying an existing gasoline-only model to accept E85 is estimated at \$50-\$100 per vehicle,[7] implying production of E85 flex-fuel vehicles is an effective means to comply with the CAFE regulation (see Anderson and Sallee (2009)). Manufacturers tend to supply flex-fuel versions of their top selling models by category (pickup, sedan, SUV, and van) and typically do not mark up the flex-fuel capability. The two versions of the vehicle function identically except with respect to fuel economy. Flex-fuel vehicles running on E85 typically get about 25-30% lower fuel economy than when running on gasoline. The lower fuel economy is a consequence of ethanol's lower energy density than gasoline.

3.2 E85 Fuel

Although retail service station adoption of E85 has grown rapidly since 2005, the overall penetration of the retail fuel market remains quite small. As of September 2007, fewer than 1200 of the nation's 135,000 service stations offered E85. In order to sell E85, fuel retailers must frequently install new or upgrade existing dispensing infrastructure to accommodate E85's handling requirements. The corrosive and water-absorbing properties of alcohol require that E85 be stored in specially lined tanks dedicated for E85 use. Specialty hoses, nozzles, and handles are also required. Some existing tanks may be retrofitted for E85 use; otherwise, a new storage tank will be required for the station to offer E85. The Department of Energy estimates installation costs can range from \$5,000 to more than \$60,000, depending on the configuration required.[8] However, E85 refueling

[6] Source: http://www.afdc.energy.gov/afdc/data/vehicles.html
[7] Source: Union of concerned scientists website, http://www.ucsusa.org.
[8] Available at http://www.afdc.energy.gov/afdc/ethanol/cost.html

infrastructure has been the target of several incentive programs, both state and federal. Most notably, the Energy Policy Act of 2005 (EPACT05) provides for reimbursement of up to 30% of the costs to install E85 compliant tanks and dispensing pumps (up to $30,000). Thus, the fixed costs of entering the E85 market are to a large degree offset by available subsidies. Further, with minor modifications, E85-compliant storage tanks can generally be used to handle gasoline, implying that storage tank costs are not fully sunk, since the station may subsequently reallocate the tank to a gasoline product. Since service stations typically maintain a limited number of storage tanks (which can service multiple pumps), and only one blended product may be allocated to each tank, a major component of the cost of carrying E85 is the opportunity cost of using the tank to sell another product.[9] This opportunity cost is effectively a recurring fixed cost.

Commerical production of bulk ethanol as well as retail sales of E85 are concentrated in the Midwest. The relative abundance of ethanol in this region may be explained by the proximity to large amounts of corn, the primary feedstock for ethanol in the U.S., which results in low transportation costs for the feedstock and processed liquid fuel. Minnesota, which has long maintained pro-ethanol policies, has consistently led the nation in the number of public E85 outlets, with approximately 300 service stations (12% of the state total) online as of March 2007. In spite of the relative availability of E85 in Minnesota, it comprises a small fraction of overall fuel sales in the state. Among stations that sell E85, the average reported monthly volume from 2001-2007 was approximately 2200 gallons, about 3% of the average station volume for gasoline.[10] Anderson (2006) studies the demand for E85 in Minnesota, and finds evidence that consumers are willing to pay a small premium for E85, presumably due to perceptions about ethanol's environmental benefits relative to gasoline or preferences for a domestically produced motor fuel. He also documents evidence of fuel-switching behavior by flex-fuel vehicle owners.

3.3 Data

Data for the study are a panel of six yearly observations of 6882 Census zip code tabulation areas[11] from the states of Illinois, Indiana, Iowa, Minnesota, Texas, and California. The data span the years 2001 to 2006. Vehicle sales data comes from R.L. Polk & Company, which compiles vehicle registration information from state departments of motor vehicles. The records provided

[9] In a technical report from the National Renewable Energy Laboratory that analyzes the business case for selling E85, Johnson and Melendez (2007) state: "Gasoline stations in the United States have an estimated average of 3.3 underground storage tanks (USTs) each (Miller 2007). This average includes regular gasoline, premium, mid-grade, diesel (which tend to be concentrated at large truck stops), and kerosene. Stations often dedicate two of these tanks to regular unleaded (Kaiser 2007) and one to premium."

[10] Source: Minnesota Department of Commerce website, http://www.state.mn.us/portal/mn/jsp/home.do?agency=Commerce

[11] Henceforth, "zip code" is taken to mean a Census 2000 Zip Code Tabulation Area (ZCTA). Zip codes, which represent mail delivery routes, can overlap in space and are therefore unsuitable for defining unique spatial regions. ZCTAs created by the Census solve this issue by mapping overlapping zip codes to a single spatial region. Data from the Census is reported directly by ZCTA whereas other sources use postal zip codes. A 1999 vintage crosswalk file is available that maps postal zip codes to ZCTAs. To map newer postal zip codes to ZCTAs, I first obtained the zip code latitude and longitude coordinates from a public geocoding service. Next, using shapefiles containing the coordinates of ZCTA boundaries, I assigned the zip code to a ZCTA using a point-in-polygon routine.

are transaction level, and represent the *complete* population of flex-fuel vehicle registrations for the period of study. Record attributes include the vehicle's zip code of registration, date of registration and registrant type (consumer or fleet). Data on E85 availability comes from the National Renewable Energy Laboratory (NREL). These data includes the station name, station type (retail or private access[12]), address, E85 introduction date, and station closure date (when applicable).

Summary statistics are provided in Table 1. I also provide a visualization of the installed base of flex-fuel vehicles and the number of E85 retailers in Appendix A. The maps show the spatial distribution of the ethanol-compatible vehicles and fuel in the Midwestern states at the beginning and end of the study. Close inspection of the maps reveals a pattern of positive correlation between the flex-fuel installed base and the number of E85 retailers – this pattern is particularly clear in the southwest portion of Minnesota, which is the state's primary corn growing region.

Table 1: Summary statistics

Variable	Name	Obs	Mean	Std. Dev.	Min	Max
Flex-fuel vehicle registrations						
consumer	Q_1	41292	20.67	35.36	0	553
fleet	Q_2	41292	5.54	40.56	0	2508
E85 stations						
retail	N	41292	0.03	0.20	0	4
private	$\widetilde{N}$	41292	0.00	0.05	0	2
Consumer variables						
population	P_1	6882	4863.53	6514.06	0	47911
rural proportion	*rural*	6882	0.58	0.45	0	1
median household income ('000)	*income*	6882	50.77	20.07	0	270.5
mean age	*age*	6882	38.73	5.75	18.8	80
male percentage	*male*	6882	50.29	3.47	31.3	100
median commute time (min)	*travel_time*	6882	26.79	5.72	15	75
Fleet variables						
persons employed	P_2	41292	2053.06	4135.57	0	60548
total establishments	*total_stablishments*	41292	4907.29	8924.58	0	130757
car rental agencies	*auto_rentals*	41292	0.22	1.01	0	49
mean employee salary ('000)	*avg_salary*	6882	25.96	9.11	0	300
Vehicle/fuel supply chain variables						
car dealerships	*auto_dealers*	41292	0.96	2.15	0	28
gasoline stations	*gas_stations*	41292	4.23	5.42	0	51
ethanol plants	*ethanol_plants*	41292	0.01	0.08	0	2
corn acres planted	*corn_acres*	6882	196.76	268.11	0	1241.7
Transportation characteristics						
interstate highways	*interstates*	6882	0.29	0.53	0	3
market area (sq mi)	*land_area*	6882	79.44	136.81	0.01	3444.1

[12]Private access refueling stations are typically dedicated facilities serving a single corporate or government fleet.

In total, I observe 853,474 consumer flex-fuel registrations, 228,867 fleet flex-fuel registrations and 459 E85 entry events. To tease out the variation in E85 availability further, in Table 2, I tabulate observations of the number of E85 retailers and compute the empirical probabilities of transitions in the number of E85 retailers. The table indicates, for example, that I have 997 observations of markets with an E85 monopolist and that given a monopoly market at time t, in the next period there is 3.2% probability of observing the firm exit the market, a 90.2% probability the monopolist remains the sole incumbent, and so on.

Table 2: Observations of retail E85 station incumbency (N_t) and empirical Markov transitions (N_t to N_{t+1})

		$P(N_{mt} \longrightarrow N_{m,t+1})$ transition				
N	Obs	0	1	2	3	4
0	40168	0.988	0.011	0.001	0.000	0.000
1	997	0.032	0.902	0.054	0.012	0.000
2	101	0.000	0.073	0.745	0.164	0.018
3	21	0.000	0.000	0.167	0.333	0.500
4	5	0.000	0.000	0.000	0.000	1.000
Total	41292					

Before developing the model, I first present model free evidence for the existence of a network effect. The expected patterns are that flex-fuel demand (Q_1, Q_2) is positively correlated with the number of E85 retailers (N), and that N is positively correlated with the installed base of consumer and fleet flex fuel vehicles. Calculation of the true installed base involves modeling initial conditions and vehicle scrap rates,[13] which I ignore for the purpose of demonstrating basic relationships. Here, I use the cumulative number of consumer and fleet flex-fuel vehicles as proxies for the true installed bases.

I provide two types of preliminary evidence for the presence of the network effect. First, in Table 3, I report the conditional distributions of cumulative consumer and fleet flex-fuel registrations as a function of the number of E85 retailers, N. As expected, the installed bases increase in their mean and median values as N increases.

Table 3: Cumulative flex-fuel vehicle registrations by retail E85 stations

		Consumer FFVs		Fleet FFVs	
N	Obs	Median	Mean	Median	Mean
0	40168	70.0	132.7	17.6	128.4
1	997	131.8	180.3	49.1	224.3
2	101	209.7	184.1	57.7	88.7
3	21	281.0	241.5	67.4	50.4
4	5	312.6	185.6	104.4	76.5

As a second source of evidence, in Table 4, I report descriptive linear regressions of the dependent variables Q_1, Q_2, and N as a function of relevant covariates. I perform the regressions using the

[13] See Section ?? for a discussion of this issue.

same control variables as in the main results (including fixed effects – see Section 6 for the full specification), but for brevity I report only the key parameters of interest. The estimates provide further prima facie evidence for the presence of a network effect: all coefficients have the expected (positive) signs and are highly significant.

Table 4: Descriptive regressions of model dependent variables

	Dependent variable		
	Q_1	Q_2	N
E85 retailers (N)	4.44***	3.16***	
	(0.56)	(1.04)	
consumer FFV installed base			9.4e-05***
			(1.1e-05)
fleet FFV installed base			3.8e-06***
			(1.1e-06)
Observations	41292	41292	41292
RMSE	14.21	26.33	0.14

All regressions include market and period fixed effects
Significance levels: * p<0.05, ** p<0.01, *** p<0.001

4 Model

The model formalizes the interdependence of flex-fuel vehicle demand and E85 market entry. The demand system relates flex-fuel vehicle purchases to the availability of E85 fuel, while the entry model relates the number of retail E85 outlets to the installed base of flex-fuel vehicles. Observations of flex-fuel demand or E85 market entry represent agent decisions to adopt the ethanol fuel standard. In this sense, the model may be interpreted as a model of technology adoption: consumers comply with the standard by adopting flex-fuel technology, and fuel retailers comply by adopting E85 dispensing infrastructure.

I develop the model in three stages. In Section 4.1, I provide an overview of the model and discuss key assumptions. In Section 4.2 I present the vehicle demand system in detail, while in Section 4.3 I develop the fuel market entry model. I close the model by formalizing the equilibrium concept in Section 4.4.

4.1 Framework and assumptions

The model considers the decisions of three types of economic agents: fuel retailers, consumers and fleets. In the case of fuel retailers, the population of interest is the set of firms that may potentially enter the E85 market. For consumers, the modeled population is the set of individuals who are potential buyers of flex-fuel vehicles. In every period, each member of these populations makes a discrete choice of whether or not to adopt ethanol fuel technology, either by entering the E85 fuel market (retailers) or purchasing a flex-fuel vehicle (consumers). I model fleets in a similar fashion, abstracting away from complexities associated with centralized decision-making and bulk buying behavior. That is, fleets are effectively treated as individuals who have unit demand for flex-fuel vehicles.

The conceptual framework of the model is a simultaneous move game. The game models strategic competition among ex ante identical firms that may potentially enter the E85 market. Firm competition is strategic due to the fact that each firm's entry decision influences the level of profits realized by all firms in the market. The moves taken by agents in the game are as follows: (1) consumers/fleets choose whether or not to purchase a flex-fuel vehicle, (2) consumers/fleets who own flex-fuel vehicles set their level of E85 consumption, (3) potential entrants decide whether or not to enter the E85 market, (4) entering retailers compete in quantities to set E85 output.[14] All these actions are assumed to occur simultaneously. Note that the flex-fuel purchase decisions of consumers and fleets influence firm profits, as such decisions increase the size of the market for E85. Therefore, market equilibrium will be defined over the action space of all agents, not just firms. I assume all agents have complete information, i.e., they have full knowledge of the process determining market outcomes and directly observe stochastic shocks which are unobserved by the researcher. The solution concept of the game is a symmetric Nash equilibrium in pure strategies, which I formally define in Section 4.4.

Throughout the model development, I follow the following notational conventions. I index markets (zip codes) by $m \in \{1, ..., M\}$ and time periods (years) by $t \in \{1, ..., T\}$. I index equations by $k \in \{1, 2, 3\}$ for consumer flex-fuel demand, fleet flex-fuel demand, and E85 market entry, respectively. For example, consumer and fleet flex-fuel vehicle sales are denoted by Q_{1mt} and Q_{2mt}. I represent the number of retail and private access (fleet dedicated) E85 stations by N_{mt} and $\widetilde{N}_{mt}$. These four variables comprise the endogenous quantities observed in the data. As changes to the number of private access E85 stations are rare and not of central interest, I do not explicitly model $\widetilde{N}_{mt}$, but control for its potential endogeneity in the fleet flex-fuel demand estimation through the use of instrumental variables. Collectively I refer to predetermined market characteristics as Z. Since prior period outcomes are predetermined in the current period, the history of market outcomes is included in Z, i.e., $\{Q_{1m\tau}, Q_{2m\tau}, N_{m\tau}\}_{\tau<t} \in Z_{mt}$. I denote the model parameters collectively as $\Theta = (\theta_{12}, \theta_3)$, where θ_{12} represents all (consumer and fleet) flex-fuel demand parameters and θ_3 captures the E85 market entry and error term covariance parameters.

[14] Under the assumption of identical firms, a price setting game results in the familiar Bertrand paradox, implying marginal cost pricing for all firms. An output setting game avoids this complication and captures the key feature of interest – that profits should decline in the number of entering firms.

4.2 Flex-fuel demand

4.2.1 Consumers

Following Berry (1994) and Berry, Levinsohn, and Pakes (1995), I model consumer demand for flex-fuel vehicles as a utility maximizing discrete choice. In each period t, the population of consumers of market m, P_{1mt}, choose between purchasing a flex-fuel vehicle (1) and the outside alternative (0), which is normalized to have zero utility in expectation. The choice-specific utilities for individual consumer i are assumed to take the following form:

$$\begin{aligned} U^1_{imt} &= \alpha_1 N_{mt} + \beta_1' Z_{1mt} + \delta_{1m} + \omega_{1t} + \epsilon_{1mt} + \eta^1_{imt} \\ &\equiv \overline{U}_{1mt} + \epsilon_{1mt} + \eta^1_{imt} \\ U^0_{imt} &= \eta^0_{imt} \end{aligned} \tag{1}$$

The utility of choice 1 appearing in equation (1) captures the consumer's valuation of flex-fuel technology as a function of the availability of E85 fuel, which is given by the number of E85 retailers operating in the market, N. The associated α_1 parameter is the primary coefficient of interest – this parameter reflects the influence of E85 availability on consumer utility, and therefore captures the feedback of E85 market entry on flex-fuel vehicle adoption. The β_1 parameter captures the influence of all other relevant time and market-time varying factors. The term ϵ_{1mt} is a market and period specific shock to flex-fuel utility which is common to all consumers. The terms η^0_{imt} and η^1_{imt} capture individual $i's$ idiosyncratic preferences for the choice options.

All consumers in a market are identical up to draws of their individual valuation components, η^0_{imt} and η^1_{imt}. I assume these shocks are distributed i.i.d. extreme value. The extreme value assumption implies that choice market shares follow the standard logit formula:

$$\begin{aligned} H^1_{1mt} &\equiv \Pr\left[U^1_{imt} > U^0_{imt}\right] = \frac{\exp\left(\overline{U}_{1mt} + \epsilon_{1mt}\right)}{1 + \exp\left(\overline{U}_{1mt} + \epsilon_{1mt}\right)} = \frac{Q_{1mt}}{P_{1mt}} \\ H^0_{1mt} &\equiv \Pr\left[U^1_{imt} \le U^0_{imt}\right] = \frac{1}{1 + \exp\left(\overline{U}_{1mt} + \epsilon_{1mt}\right)} = \frac{P_{1mt} - Q_{1mt}}{P_{1mt}} \end{aligned} \tag{2}$$

H^1_1 and H^0_1 are the consumer choice shares of flex-fuel vehicles and the outside option, respectively. In this formulation, consumers are assumed to be in the market for a flex-fuel vehicle every period, allowing for the possibility of replacement sales.[15] By algebraic manipulation of the market

[15] An alternative assumption, commonly used in the durable goods literature, would be that vehicles are infintely durable and once consumers purchase a flex-fuel vehicle, they are permanently out of the market. Given the long time span of my sample (six years), I choose to allow replacement sales. This choice is consistent with my computation of the installed base of flex-fuel vehicles, which explicitly accounts for vehicle scrappage. In practice, this assumption has little effect on demand estimates, as market populations are generally very large in comparision to cumulative flex-fuel sales.

share equations above, we may write the following estimation equation[16]:

$$H_{1mt} \equiv \ln\left(H^1_{1mt}\right) - \ln\left(H^0_{1mt}\right) = \ln\left(\frac{Q_{1mt}}{P_m - Q_{1mt}}\right) = \overline{U}_{1mt} + \epsilon_{1mt} \tag{3}$$

The variable H_1 is the log-odds ratio of consumer flex-fuel vehicle purchase. Estimation of equation (3) is discussed in Section 5.

The utility specification (1) employs extremely strong controls for unobservables through the inclusion of market (δ_1) and period (ω_1) fixed effects. The market fixed effects control for all time-invariant market heterogeneity that may be correlated with Q_1. Inclusion of market fixed effects is particularly significant for my application, since estimation will be conducted under an assumption of independent markets. This assumption is far more plausible when these controls are included, since they capture all time-invariant spatial dependence in observations of consumer flex-fuel demand. The period fixed effects control for common temporal shocks to flex-fuel demand, such as changes in the variety of flex-fuel models, availability of federal tax credits, and average fuel prices. With the inclusion of these fixed effects, the variation identifying the α_1 parameter will be deviations in H_1 and N from the market-specific mean values (i.e., "within" deviations), controlling for period-specific shocks common to all markets.

4.2.2 Fleets

As previously mentioned, I simplify the treatment of fleet flex-fuel demand by ignoring bulk-buying behavior. Rather, I model fleets as individuals, who make a discrete choice of whether or not to purchase a flex-fuel vehicle. Therefore, the fleet demand system is identical in form (but not in specification) to the consumer demand system. The choice-specific utilities for fleet i are assumed to take the following form:

$$\begin{aligned} U^1_{imt} &= \alpha_{21} N_{mt} + \alpha_{22} \widetilde{N}_{mt} + \beta_2' Z_{2mt} + \delta_{2m} + \omega_{2t} + \epsilon_{2mt} + \eta^1_{imt} \\ &\equiv \overline{U}_{2mt} + \epsilon_{2mt} + \eta^1_{imt} \\ U^0_{imt} &= \eta^0_{imt} \end{aligned} \tag{4}$$

[16] In empirical specifications, I calculate H_{1mt} using $H_{1mt} = \ln\left(\frac{\kappa + Q_{1mt}}{\kappa + P_m - Q_{1mt}}\right)$ as this avoids the technical problem of infinite negative utility encountered when $Q_{1mt} = 0$. Minimizing the occurrence of this condition motivates using a yearly panel over one with a finer time frequency. Still, I observe no flex-fuel sales in 14% of observations, typically in markets with small populations. For the correction I set $\kappa = 0.5$, as this value is shown by Pettigrew, Gart, and Thomas (1986) to be the bias-minimizing value. Pettigrew, Gart, and Thomas (1986) also derive the asymptotic bias of this empirical approximation to the true log-odds ratio, and show that it is given by $\frac{2p-1}{24(np(1-p))^2} + O(n^3)$, where p is the true individual adoption probability and n is the number of trials per observation. In my application $p \sim .005$ while $n \sim 5000$, implying an asymptotic bias of order -10^{-5}, which is clearly negligible. As a robustness check of the potential bias introduced by this approximation, I estimated the consumer demand model setting $\kappa = 0$ and only using observations for which $Q_{1mt} > 0$. The estimate of α_1 from this model was consistent with the main results within one standard deviation and 4% in absolute magnitude.

This utility specification differs from the consumer utility specification in two ways. First, another potentially endogenous variable appears: the number of private access E85 refueling stations ($\widetilde{N}$), which are dedicated facilities serving a single fleet. Whereas the presence of private access refueling stations should not affect consumer utility for ethanol-compatible vehicles, these facilities may impact the observed share of flex-fuel vehicles among fleets. $\widetilde{N}$ is potentially endogenous because unobserved factors that lead to installation of a fleet-dedicated E85 station may be correlated with fleet flex-fuel demand (presumably, construction of a dedicated E85 refueling station signals a fleet's intention to increase its investment in flex-fuel vehicles). Even though I do not explicitly model the "entry" of such stations, I must control for the potential endogeneity of $\widetilde{N}$ through the use of instrumental variables. The other difference from the consumer demand specification is the factors which enter Z_2.

Following the logic of the previous section, the estimation equation for fleet flex-fuel demand is:

$$H_{2mt} = \ln\left(\frac{Q_{2mt}}{P_{2mt} - Q_{2mt}}\right) = \overline{U}_{2mt} + \epsilon_{2mt} \tag{5}$$

A final comment relates to the empirical specification of the fleet market size, P_2. Whereas the population of consumers (P_1) is well documented in Census data, the appropriate measure of P_2 is less obvious. For consistency with the assumption of fleets as individuals with unit demand for flex-fuel vehicles, I need a proxy for the total number of potential fleet sales to represent P_2. My approach is to use the number of persons *employed* in the market as this proxy. The number of employees is a reasonable measure for the potential population of fleet vehicles, as fleet vehicles are typically assigned to employees. To mitigate concerns about the market size definition, I demonstrate in Appendix D.3 that the estimated parameters are generally robust to alternative specifications of P_2.

4.3 E85 Market Entry

Fuel retailer entry into the E85 market is modeled in the tradition of Bresnahan and Reiss (1990) and Bresnahan and Reiss (1991). These models assume that the number of firms in the market, N, is the outcome of a two-stage game of complete information played among E identical potential entrants. In the first stage, firms make strategic entry decisions, anticipating the ensuing competition in output during the second stage. Potentially entering firms are assumed to have identical profit functions, which are stochastic due to the presence of market level shocks to profitability that are common to all firms.[17] Under these assumptions, N reflects bounds on a latent profit function, which may recovered through the estimation procedure. The intuition behind the approach is that in a symmetric equilibrium, an observation of $N = k$ firms implies $\Pi^k \geq 0$ and $\Pi^{k+1} < 0$, where Π^N represents profits in a market with N firms.

[17] See Bresnahan and Reiss (1990) for a discussion of the implications of this assumption.

Berry and Reiss (2007) discuss different approaches to specifying Π^N, and advocate deriving the appropriate reduced form from explicit assumptions about firm costs, the demand specification, and the second stage equilibrium process. I adopt this approach and begin by developing a model for E85 fuel demand. I decompose E85 demand (D) into the product of the market size (S) and per-capita demand, which I model as linear in E85 prices, as follows:

$$D = \sum_{f=1}^{N} d_f = S\,(a - P) \tag{6}$$

Here f indexes the fuel retailers active in a market and d_f represents the per-firm demand for ethanol fuel. The coefficient on E85 fuel price (P) is set to -1 without loss of generality, as this simply scales the quantity units in which price is quoted.[18] Next I assume second stage competition among ethanol retailers is a Cournot output game. In Appendix D, I explore the robustness of my results to alternative assumptions about firm competition. Under the assumption of independent market outcomes, strategic competition is limited to potential entrants within the market. The E85 fuel demand model is closed by specifying the firm cost function, which I take to be linear in output and common to all firms:

$$C_f = c d_f + F \tag{7}$$

Constant marginal costs are a reasonable assumption for a motor fuel retailer, since the retailer is not directly involved with ethanol production and has few mechanisms by which to achieve scale economies.

I now derive the reduced form profit function corresponding to the symmetric Cournot equilibrium. Using equation (6), the inverse demand curve may be written as:

$$P = a - \frac{D}{S} = a - \frac{1}{S}\sum_{f=1}^{N} d_f$$

In Cournot competition, firms complete in quantities. The first order condition for firm f is given by:

[18] Moreover, price effects cannot be separately identifed using observations of N alone, as the price coefficient is interacted with the market size, which must be normalized to set the scale of variable profits.

$$
\begin{aligned}
\frac{\partial \Pi_f^N}{\partial d_f} &= \frac{\partial}{\partial d_f}\left[d_f\left(P-c\right)-F\right]=0 \\
&= P-c+d_f\frac{\partial P}{\partial d_f}=0 \\
&= a-\frac{1}{S}\sum_{f=1}^{N} d_f - c + d_f\left(-\frac{1}{S}\right)=0
\end{aligned}
$$

The symmetry of firms implies that $\sum_{f=1}^{N} d_f = N d_f$. Using this condition and solving for optimal firm quantity yields:

$$
d_f^{*} = \frac{S}{1+N}\left(a-c\right) \tag{8}
$$

The equilibrium market price will therefore be:

$$
P^{*} = a - \frac{N d_f^{*}}{S} = a - \frac{N}{1+N}\left(a-c\right) \tag{9}
$$

Finally, reduced form profits may be expressed as:

$$
\Pi^N = \Pi_f^N = (P^{*}-c)d_f^{*} - F = S\frac{(a-c)^2}{(1+N)^2} - F \tag{10}
$$

To convert this model into an econometric specification, we must clarify how stochastic terms enter firm profits and how to parameterize the terms S, $(a-c)^2$, and F. Following Bresnahan and Reiss (1991), I assume unobservable firm profits enter Π^N linearly as shocks to fixed costs. The shock to fixed costs is common to all firms, and is given by ϵ_{3mt}. As firms are ex ante identical, I represent observable fixed costs as a linear function of exogenous market-level cost shifters (Z_3). Thus, fixed costs for all firms are given by $F_{fmt} = F_{mt} = \phi' Z_{3mt} + \epsilon_{3mt}$. The term $(a-c)^2$ captures the E85 fuel demand intercept and marginal costs, the constituents of firm variable profits. I parameterize $(a-c)^2$ as a function of exogenous E85 variable cost and demand shifters (Z_4). I enforce positivity of variable profits by letting $(a-c)^2_{mt} = \exp\left(\psi' Z_{4mt}\right)$.

The expression for market size, S, requires a more involved discussion. The complication which arises is due to the fact that the E85 entry model is linked to the flex-fuel demand system through the market size, which is a function current flex-fuel demand. Thus, the market size is endogenous to N. For generality, I follow Bresnahan and Reiss (1991) and assume S may be represented as a linear function of the installed base of consumer flex-fuel vehicles, the installed base of fleet flex-fuel vehicles, and exogenous market-level shifters of the effective market size (Z_5). For notational convenience, I define the *prior period* installed base of flex-fuel vehicles for agent type k ($1 =$ consumer, $2 =$ fleet), as $B_{kmt} \equiv \sum_{\tau=0}^{t-1} Q_{km\tau}$, which is predetermined in period t. Therefore, the current period installed base of agent type k flex-fuel vehicles is given by $B_{kmt} + Q_{kmt}$. With this notation, the market size may be represented as:

$$S_{mt}(N_{mt}) = (B_{1mt} + Q_{1mt}) + \gamma\,(B_{2mt} + Q_{2mt}) + \lambda' Z_{5mt} \tag{11}$$

The coefficient on the installed base of consumer flex-fuel vehicles is set to one as the expression for variable profits (which is interacted with the market size in equation (10)) includes a constant, implying a normalization of market size is required for identification. The normalization converts E85 demand into units of consumer flex-fuel vehicles. This implies, for example, that the γ coefficient in equation (11) may be interpreted as the estimated number of fleet flex-fuel vehicles required to match the fuel consumption of one consumer flex-fuel vehicle. Note that I write S as a function of N to highlight the fact that installed bases of consumer and fleet flex-fuel vehicles are implicit functions of N through current period flex-fuel vehicle demand, Q_1 and Q_2. This notation also emphasizes that it is through the N dependence of market size that positive feedback in ethanol adoption is realized in the entry model for the current period. Feedback effects in prior periods are incorporated into B_1 and B_2, which are predetermined and do not depend on the current value of N. This is the key force driving the network effect in the model.

Although agents in the market have complete information and therefore know the exact form of S, the analyst must contend with uncertainty introduced by two types of unobservables. First, since the vehicle demand shocks enter the expression for market size (11), from the econometrician's perspective S is a random variable whose distribution is induced through the joint distribution of $\begin{pmatrix}\epsilon_1 & \epsilon_2 & \epsilon_3\end{pmatrix}$. To see this, note that equations (3) and (5) imply that the current period flex-fuel sales Q_1 and Q_2 are (non-linear) functions of ϵ_1 and ϵ_2. Second, the econometrician does not know the exact values of B_1 or B_2, since vehicle scrappage is not observed, nor are the sales of flex-fuel vehicles prior to the inception of the study. The estimation procedure I develop in Section 5.2 takes account of both types of uncertainty associated with market size.

Putting the aforementioned specifications together, and defining expected firms profits as $\overline{\Pi}^N$, the model for N taken to data is:

$$N_{mt} = \sum_{k=0}^{E} k\left(\Pi_{mt}^{k} \geq 0\right)\left(\Pi_{mt}^{k+1} < 0\right) \tag{12}$$

$$where:$$

$$\Pi_{mt}^{N} = \begin{Bmatrix} 0 & if & N_{mt} = 0 \\ \overline{\Pi}_{mt}^{N} - \epsilon_{3mt} & if & 0 < N_{mt} \leq E \end{Bmatrix}$$

$$\overline{\Pi}_{mt}^{N} \equiv S_{mt}(N_{mt}) \frac{\exp\left(\psi' Z_{4mt}\right)}{\left(1 + N_{mt}\right)^2} - \phi' Z_{3mt}$$

4.4 Equilibrium

Before formally defining the equilibrium, it is important to clarify the relationship between the game strategies of individual potentially entering firms, which are unobserved, and observations of the number of firms (N). An individual firm strategy may be represented as $I_f \in \{0,1\}$ where 1 indicates entry and 0 indicates staying out of the market. Since our game has E players, a pure strategy Nash equilibrium obtains when the condition

$$\Pi_f(I_1^*, ..., I_f^*, ..., I_E^*) \geq \Pi_f(I_1^*, ..., I_f, ..., I_E^*) \ \ \forall I_f \in \{0,1\}$$

holds for all firms $f \in \{1, ..., E\}$. As shown in Bresnahan and Reiss (1990), this Nash solution concept does not uniquely determine the identity of entering firms. For example, with identical firms in a market that can profitably sustain a monopolist, there are E pure strategy solutions to the game (the $f = 1, ..., E$ strategies for which $I_f^* = 1, I_k^* = 0 \ \forall \ k \neq f$). However, since only one set of strategies will be played, the equilibrium number of entering firms, $N^* = \sum\limits_{k=1}^{E} I_k^* = 1$ is uniquely determined. The potential for multiple equilibria, which would severely complicate estimation of the model, is removed by aggregating outcomes. I take this approach and define the equilibrium in terms of the aggregated equilibrium strategies of potentially entering E85 retailers.

Uniqueness of the symmetric Nash equilibrium in N is guaranteed provided that firm profits decline in the number of firms entering the market. If this condition is violated, a unique ordering of Π^N is no longer guaranteed, and the potential for multiple equilibria once again arises. In entry models such as Bresnahan and Reiss (1991), this condition is imposed in estimation of the model of N as an ordered dependent variable. Other models satisfy the condition by specifying a profit function that declines in N, such as Berry (1992). This is essentially the approach I take, but I require a further refinement. The presence of the indirect network effect, i.e. the fact that the market size is a function of N, admits the (mathematical) possibility that profits could increase in the number of firms. I rule out this possibility by imposing the constraint that $\frac{\partial \Pi^N}{\partial N} \leq 0$ hold for all markets when estimating equation (12). That is, I impose that increases in firm profits from an expanded market size (via the network effect in $S(N)$) cannot be larger than decreases in profits from additional competition (via the $\frac{1}{(1+N)^2}$ term entering variable profits). The constraint, which is derived in Appendix B, takes the following form:

$$\frac{\alpha_1 Q_{1mt}}{1 + \exp(H_{1mt})} + \frac{\gamma \alpha_{21} Q_{2mt}}{1 + \exp(H_{2mt})} - \frac{2 S_{mt}}{1 + N_{mt}} \leq 0 \tag{13}$$

I impose this constraint when estimating equation (12). Fortunately, the data reveal ex-post that the conditions of a unique equilibrium are satisfied – i.e., the constraint does not bind for the estimated parameter vector. This result is expected, for if it were the case that the network effect were stronger than the competitive effect, the model would predict only two possible market outcomes: zero or complete market adoption of the ethanol standard by all agents, depending upon the level of market fixed costs. The data reject this notion out of hand by the very existence of a distribution in agent adoption rates of ethanol.

Since consumer and fleet flex-fuel purchase decisions influence E85 retailer profits, the equilibrium concept must also encompass their actions. Formally then, the symmetric Nash equilibrium in pure strategies for this model is defined by the number of entering E85 retailers N^*, the number of consumer flex-fuel vehicle sales Q_1^*, and the number of fleet flex-fuel vehicle sales Q_2^* that simultaneously satisfy equations (3), (5), (12), and (13).

5 Estimation

I develop a two-step procedure to estimate equations (3), (5) and (12) under constraint (13). I motivate this choice in the method overview discussion in Section 5.1. I begin with discussion of the entry model estimation in Section 5.2, which is the second step in the overall procedure, in order to defer the technicalities that accompany describing the System GMM procedure. I follow in Section 5.3 with discussion of estimation of the flex-fuel demand system, which is the first step of the overall procedure.

5.1 Method overview

The primary challenge for estimation is finding an appropriate estimator for the E85 entry model. The flex-fuel demand equations may be transformed to a linear system, and thus permit a wide variety of estimators since instruments for the endogenous variables N and $\widetilde{N}$ are available (I discuss the instruments and identification in Section 5.3 and Appendix C). The entry model, however, is highly non-linear in parameters, necessitating the use of either a likelihood-based or moment-based estimator. I therefore consider the problem of how to consistently estimate the entry equation conditional upon the flex-fuel demand parameters (θ_{12}), which constitutes the second step of the estimation procedure. Conceptually, there are two potential sources of endogeneity in the market size, S. First, referring to equations (1) and (4), N directly enters the expressions for consumer and fleet flex-fuel utility. Therefore, Q_1 and Q_2 in the expression for S are explicit functions of N. However, with consistent estimates of θ_{12} in hand, this dependency is fully accounted for when computing S for estimation of the entry model. A second source of endogeneity arises through the stochastic terms – i.e., the entry model error term ϵ_3 is potentially correlated with the demand system error terms ϵ_1 and ϵ_2, which also determine Q_1 and Q_2. These terms may be correlated through common unobservables that affect both flex-fuel demand and the profitability of E85 market entry. An example of such an unobservable might be a regional campaign of public service announcements touting the benefits of ethanol to the community. It might at first seem tractable to compute the market size in the entry model by conditioning on both the demand system residuals as well as the demand system parameter estimates. That is, given a candidate parameter vector for the entry model (θ_3), one might be tempted to express the market size $S = S(\theta_3 \mid Z, \widehat{\theta}_{12}, \widehat{\epsilon}_1, \widehat{\epsilon}_2)$. However, to do so would ignore the fact that ϵ_1 and ϵ_2 are co-determined with N, through equations (1) and (4). In other words, ϵ_1 and ϵ_2 may not be held fixed when fitting the model to data.

My approach to controlling for both sources of endogeneity mentioned above is to work with the reduced form of the model, given the demand parameters. That is, for a candidate entry model parameter vector, I solve the nonlinear system (3), (5) and (12) for the equilibrium number of entering firms, N^*. This solution process ensures consistency of the estimated parameters, since all cross-equation dependencies are reflected in the resulting value of N^*. I base estimation on minimizing the difference (in the sum squares sense) between the model predicted number of firms, N^*, and the observed number of entering firms, N. This non-linear least squares estimator is a variant of GMM.

A requirement to solve for N^* is that the error vector ϵ is known. ϵ is not directly observed, but draws of ϵ may be simulated if we are willing to take a stand on its joint distribution (as one would when pursuing a likelihood-based approach). I assume ϵ to be distributed multivariate normal:

$$\begin{pmatrix} \epsilon_{1mt} \\ \epsilon_{2mt} \\ \epsilon_{3mt} \end{pmatrix} \sim N(0, \Sigma), \;\; where \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \rho_{12}\sigma_1\sigma_2 & \rho_{13}\sigma_1 \\ \rho_{12}\sigma_1\sigma_2 & \sigma_2^2 & \rho_{23}\sigma_2 \\ \rho_{13}\sigma_1 & \rho_{23}\sigma_2 & 1 \end{pmatrix} \tag{14}$$

Unobserved E85 retailer profitability (ϵ_3), which has a natural interpretation as a shock to fixed costs, is normalized to have unit variance in order to identify the parameters of the entry model. As is customary in latent variable models, this normalization sets the scale of E85 retailer profits, which are unobserved. Note that while ϵ_1 and ϵ_2 may not be held fixed, we may use the first step residuals to estimate σ_1, σ_2, and ρ_{12}. In particular, $\widehat{\sigma}_1 = \sqrt{Var\left(\widehat{\epsilon}_1\right)}$, $\widehat{\sigma}_2 = \sqrt{Var\left(\widehat{\epsilon}_2\right)}$, and $\widehat{\rho}_{12} = Corr(\widehat{\epsilon}_1, \widehat{\epsilon}_2)$. The remaining covariance terms, ρ_{13} and ρ_{23}, are parameters to be estimated with the entry equation. I implement the estimator by drawing multiple ϵ vectors for each observation,[19] solving the reduced form for N^* for every draw of ϵ, and then averaging the result to integrate the effects of ϵ out of the objective function. In Appendix D, I also report results from estimating the entry model by maximum likelihood, under the *restriction* that $\rho_{13} = \rho_{23} = 0$, and find very close agreement with the unrestricted estimator.

A final step must be taken for valid inference. Using estimated demand parameters ($\widehat{\theta}_{12}$) in the entry model estimation introduces the potential for measurement error, since $\widehat{\theta}_{12}$ is an imprecise measurement of θ_{12}. The standard errors of the entry model parameters (θ_3) must reflect this additional uncertainty. To address this issue, I perform a nonparametric panel bootstrap over the two-step procedure. That is, I resample market histories with replacement and perform the two-

[19] This step is performed once, prior to the entry model estimation, using standard normal draws. The value of ϵ is recomputed as the entry parameter (θ_3) vector changes by taking the product of ϵ and the Cholesky decomposition of the covariance matrix implied by θ_3.

step estimation procedure BS times. Parameter standard errors are then given by the standard deviations of bootstrapped parameter values. Estimates reported in Section 6 are based upon $BS = 30$ replications. In the reported results, both flex-fuel demand and E85 entry model parameters have bootstrapped standard errors. Using bootstrapped standard errors implies inference will be robust to the effects of heteroskedasticity, provided the estimation sample is truly representative of the general population.

5.2 E85 entry equation

Here I develop a procedure to estimate the E85 entry model that allows for arbitrary correlation between the market entry equation and the flex-fuel demand equations. Promotional activity by a regional ethanol trade organization might be one example of such a common unobserved effect. In Appendix D.1, I develop an alternative entry model estimator under the assumption of no correlation between the entry model error term and the flex-fuel demand error terms.

As discussed in the estimation method overview (Section 5.1), estimation of the entry model parameters is conducted conditional upon estimates of the vehicle demand parameters (θ_{12}). Uncertainty introduced into the entry model estimation from imprecise measurement of the demand parameters is controlled for by bootstrapping over the two step estimation procedure, in which entire market histories are sampled with replacement. The key insight that motivates the chosen estimator is that, given the demand parameters and simulated draws of the error terms ϵ, the reduced form of the model may be solved for N^*, the model-predicted equilibrium number of entering E85 retailers. Consequently, if the demand parameters are correctly identified, the entry model may be consistently estimated without the use of exogenous instruments for the endogenous market size. Consistency follows from the fact that explicitly solving the model automatically incorporates all cross-equation dependencies into the resulting value of N^*. Intuitively, estimation proceeds by minimizing the squared difference between the model predicted number of firms, N^*, and the observed number of entering firms, N. This non-linear least squares estimator is given by:

$$\widehat{\theta}_3 = \arg\min_{\theta_3} \sum_{m,t} \left(N_{mt} - \frac{1}{NS} \sum_{s=1}^{NS} N^*_{smt} \right)^2 \tag{15}$$

$$subject\ to: \frac{\partial \Pi^N_{mt}}{\partial N} \leq 0 \quad \forall m, t$$

Obtaining simulated draws of the error terms requires an assumption about the joint distribution of ϵ. As discussed in Section 5.1, I assume a multivariate normal distribution for ϵ. The variance-covariance terms of the flex-fuel demand errors ($\sigma_1, \sigma_2, \rho_{12}$) may be estimated from the residuals of the flex-fuel demand estimation ($\widehat{\epsilon}_1, \widehat{\epsilon}_2$). Since the variance of the entry model error term ϵ_3 is normalized to 1 in order to identify the parameters, only the cross-equation correlations (ρ_{13}, ρ_{23}) are unknown. These parameters are to be estimated in together with the parameters that enter the E85 profit function explicitly $(\gamma, \lambda, \psi, \phi)$. If the correlations ρ_{13} and ρ_{23} turn out significant in the entry model estimation, it is evidence for the presence of common unobserved shocks that influence both flex-fuel demand and E85 entry. If these coefficients are insignificant, it suggests that the model specification already fully captures the relevant co-dependence of the demand and entry systems. Per capita variable profit factors (ψ) are identified through interactions with the market size, consisting primarily of the installed base of flex-fuel vehicles. Fixed cost factors (ϕ) are identified by the (conditional) mean values of N.

Potential concerns with a non-linear least squares estimator relate to matters of econometric efficiency and algorithm convergence. To appreciate the concerns, consider the predicted number of firms based on a single draw of ϵ. The value of N^* resulting from this draw could be very different from $E[N^*]$ if the draw originates in the tail of the distribution. The resulting objective function is unlikely to be smooth, hindering an optimizer's ability to find the optimal parameter vector. The discrete nature of N will tend to magnify this problem – since outcomes are determined by truncation points of the latent profit function, small changes in parameter values can lead to large changes in N^*. The standard errors of the estimated parameters are therefore also likely to be large, limiting the ability to conduct inference. A solution to these issues is to integrate the effect of the unobservables out of the objective function by taking multiple draws of ϵ for every observed market and averaging the resulting N^*. That is, I compute the expected equilibrium number of firms by $E[N^*] = \frac{1}{NS} \sum_{s=1}^{NS} N^*_{smt}$, where NS is the number of simulated draws of ϵ. In my empirical work, I take $NS = 30$, which should be a reasonable estimate of the asymptotic value of $E[N^*]$, given the normality assumption.

5.2.1 Computation of N^*

Next, I explain the calculation of N^*, the model-predicted equilibrium number of entering E85 retailers. First note that equations (3) and (5) imply that unit sales of flex-fuel vehicles may be written:

$$Q_1(n \mid \widehat{\theta}_{12}) = P_1 \frac{\exp\left(\widehat{\alpha}_1 n + \epsilon_1 + k_1\right)}{1 + \exp\left(\widehat{\alpha}_1 n + \epsilon_1 + k_1\right)}$$
$$Q_2(n \mid \widehat{\theta}_{12}) = P_2 \frac{\exp\left(\widehat{\alpha}_{21} n + \epsilon_2 + k_2\right)}{1 + \exp\left(\widehat{\alpha}_{21} n + \epsilon_2 + k_2\right)}$$

where I have suppressed subscripts and introduced the shorthand notations $k_1 = H_1 - \widehat{\alpha}_1 N - \widehat{\epsilon}_1$, $k_2 = H_2 - \widehat{\alpha}_{21} N - \widehat{\epsilon}_2$. The values of k_1 and k_2 are the predicted purchase log-odds ratios with the effect of the observed number of firms N partialed out. This ensures that the predicted values of $Q_1(n)$ and $Q_2(n)$ include the market fixed effects δ, which are not estimated explicitly in the demand estimation. All factors other than n are fixed in the above equations (here, ϵ_1 and ϵ_2 are simulated error terms). Given a candidate entry model parameter vector (θ_3), the market size may then be computed by equation (11) and hence the reduced form profit Π^N may be derived from equation (12). The system may not be solved analytically for N^*, but numerical solution of the system in straightforward. N^* is given by the smallest value of n (for $n > 0$) that satisfies the condition $(\Pi^n \geq 0) \cap \left(\Pi^{n+1} < 0\right)$.

The previous discussion of computing N^* ignored a technical but important detail relating to the installed base of flex-fuel vehicles. In equation (11), B denotes the prior period installed base, which is intended to represent the number of flex-fuel vehicles *on road* at the end of the previous year. Even though I observe all flex-fuel vehicle purchases throughout the period of my study, the true value of B is unobserved, for two reasons. First, I do not observe flex-fuel vehicle demand in periods prior to 2001 ($t = 1$), resulting in an initial conditions problem.[20] I employ two assumptions that allow me to integrate uncertainty about initial conditions out of the objective function. The first pertains to flex-fuel vehicle sales in unobserved periods. I assume that the *proportion* of flex-fuel vehicles produced which are sold into a market for a given year follows a truncated normal distribution with known mean and variance. In a preliminary step, I estimate these means and variances market by market using the six yearly observations of flex-fuel vehicle sales (R.L. Polk data) and the aggregate flex-fuel vehicle production quantities for the corresponding year. The latter are available from the Department of Energy and date to the first production year for flex-fuel vehicles.[21] To simulate flex-fuel vehicle sales for unobserved periods, I draw the sales proportion from the corresponding market distribution and multiply by the aggregate flex-fuel production for that period. As with draws of the error terms ϵ, I take $NS = 30$ draws per market, so that the effect of the initial conditions are integrated out of the objective function when $E[N^*]$ is computed. The second source of uncertainty relates to unobserved vehicle retirement. I take vehicle survival rates to be exogenous, common to all markets, and time-stationary (rates depend only upon the time difference between purchase and the current period). Further, I assume these rates are as quoted in the Department of Transportation's Transportation Energy Data Book (Davis and Diegel (2006)). Under these assumptions, the prior period installed base for agent type k is given by: $B_{kmt} = \sum_{\tau=-\infty}^{t-1} w_{t-\tau} Q_{km\tau}$, where w is the survival rate, $Q_{km\tau}$ is data for $\tau > 0$, and $Q_{km\tau}$ are simulation draws for $\tau \leq 0$. The fact that I explicitly model retirement of flex-fuel vehicles is consistent with the demand model assumption of allowing replacement sales.

A final implementation detail concerns the constraint in equation (15), which is expressed in terms of parameters and observables in equation (13). I implement the constraint using a penalty function during estimation. That is, I set $N^* = -\infty$ if the constraint is violated. The constraint does not bind at the converged parameter value.

[20] Commercial production of flex-fuel vehicles began in 1996, but significant quantities were not produced until 1998.

[21] Available at http://www.afdc.energy.gov/afdc/data/vehicles.html.

5.3 Flex-fuel demand equations

I begin with a discussion of the implications of the model structure for estimation of the flex-fuel demand equations. First note that collectively, equations (3), (5) and (12) represent a static model with feedback. That is, current period flex-fuel vehicle demand (Q_{1mt}, Q_{2mt}) depends only upon the current value of E85 availability (N_{mt}), but because N_{mt} is a function of the installed bases of flex-fuel vehicles, N_{mt} depends upon the entire history of vehicle demand outcomes up to the current period, i.e. $\tau : \tau \leq t$. In particular, note that current shocks to flex-fuel demand ($\epsilon_{1mt}, \epsilon_{2mt}$) will be correlated with current and *future* values of N. However, given the exogeneity of Z and predetermination of prior values of N, the following conditional expectations hold:

$$E\left(\epsilon_{1mt} \mid Z_1, \delta_{1m}, \omega_{1t}, N_{m\tau}\right) = 0 \quad \forall \tau < t \tag{16}$$

$$E\left(\epsilon_{2mt} \mid Z_2, \delta_{2m}, \omega_{2t}, N_{m\tau}, \widetilde{N}_{m\tau}\right) = 0 \quad \forall \tau < t \tag{17}$$

This structure has important consequences for estimation. In particular, the fact that ϵ_{1mt} and ϵ_{2mt} are correlated with future values of N implies that the usual time-demeaning transformation to remove the market fixed effects δ_{1m} and δ_{2m} will lead to biased estimates.[22] However, a first difference procedure to remove the market fixed effects remains consistent under the weaker assumption of sequential exogeneity of N implied by (16) and (17). Taking first differences of equations (3) and (5) yields:

$$\Delta H_{1mt} = \alpha_1 \Delta N_{mt} + \beta_1' \Delta Z_{1mt} + \Delta\omega_{1t} + \Delta\epsilon_{1mt} \tag{18}$$

$$\Delta H_{2mt} = \alpha_{21} \Delta N_{mt} + \alpha_{22} \Delta \widetilde{N}_{mt} + \beta_2' \Delta Z_{2mt} + \Delta\omega_{2t} + \Delta\epsilon_{2mt} \tag{19}$$

Henceforth, I refer to equations (16) and (17) as the "difference" equations, and equations (3) and (5) as the "level" equations. Absent contemporaneous correlation between the N variables and the error terms, OLS estimation of equations (18) and (19) would deliver consistent estimates. However, ΔN is endogenous in (18) and (19) since H_1 and H_2 are simultaneously determined with N through (12). Although entry of private E85 facilities is not explicitly modeled, a similar concern applies to $\Delta\widetilde{N}$ (here, the endogeneity concern arises from the possibility of an omitted variable correlated with fleet vehicle demand and $\widetilde{N}$). The general remedy for these complications is to employ instrumental variables. However, finding exogenous variables which meet the criteria of being correlated with ΔN and $\Delta\widetilde{N}$ but not the error terms $\Delta\epsilon_1$ and $\Delta\epsilon_2$ can be challenging. One appealing approach is to bring additional exogenous data to bear. In other words, find a set of "external" instruments W that shift N but do not enter the specification of vehicle demand (i.e., $W \in Z/Z_1 \cup Z_2$). Unfortunately, suitable instruments of this type are not available.[23] I therefore

[22] See, for example, Wooldridge (2002), Section 11.1 for a discussion of this issue.

[23] In Appendix C, I motivate several instrumentation strategies using data collected for the study and present the corresponding demand estimates. The instruments are motivated by considering the form of the E85 retailer profit function (10), which is decomposed into shifters of market size, fixed costs, and variable profits. I construct instruments corresponding to each type of profit shifter. For each set of results, I report diagnostic tests of instrument

pursue an alternative strategy that exploits the panel structure of the data and the orthogonality conditions implied by (16) and (17):

$$E\left(\Delta\epsilon_{1mt}N_{m\tau}\right)=0 \quad \forall\tau:\tau\leq t-2 \tag{20}$$

$$E\left(\Delta\epsilon_{2mt}N_{m\tau}\right)=0 \quad \forall\tau:\tau\leq t-2 \tag{21}$$

$$E\left(\Delta\epsilon_{2mt}\widetilde{N}_{m\tau}\right)=0 \quad \forall\tau:\tau\leq t-2 \tag{22}$$

In particular, these conditions mean that values of N lagged by two or more periods are valid instruments for ΔN, i.e., $\{N_{m\tau}:\tau\leq t-2\}$ may serve as an instruments for ΔN_{mt}. Similar relations hold for $\widetilde{N}$. Therefore, a valid instrument matrix for market m observations in the "difference" equation (18) may be expressed as:

$$W_{1m}\equiv\begin{pmatrix}\Delta Z_{1m3} & 1_t & N_{m1} & 0 & 0 & ... & 0\\ \Delta Z_{1m4} & 1_t & N_{m1} & N_{m2} & 0 & ... & 0\\ \vdots & & & & & & \vdots\\ \Delta Z_{1mT} & 1_t & N_{m1} & N_{m2} & N_{m3} & ... & N_{m,T-2}\end{pmatrix} \tag{23}$$

I use the notation of 1_t to represent a row vector whose i^{th} element is equal to 1 if $i=t$, where t is the period of the corresponding observation. This construct captures the period fixed effects in equation (18). The W matrix reflects the fact that for the strictly exogenous variables Z and the period fixed effects ω_{1t}, the differenced values may properly serve as their own instruments.

GMM estimation in first differences using instruments like those found in equation (23) is a variant of the Arellano and Bond (1991) estimator. The Arellano-Bond estimator is typically applied to dynamic panel data models that have lagged dependent variables on the right hand side, but it is equally applicable to static models with endogenous regressors. Two tests are available to substantiate the validity of this approach to identification. First, a formal requirement for the exogeneity of lagged levels of N is that no serial autocorrelation is present in ϵ. A test proposed by Arellano and Bond (1991), checks for serial autocorrelation in the GMM residuals by testing for second order autocorrelation (AR(2)) in $\Delta\epsilon$.[24] Second, as is standard for GMM estimators, the Hansen J test (Hansen (1982)) of overidentifying restrictions may be used to test that the instruments are properly excluded from the estimation equation. I report these tests in the main results (Section 6) and in my comparison of different instruments provided in Appendix C. As discussed in those sections, I find that the required conditions hold for my model.

As pointed out by Arellano and Bover (1995), if there is considerable persistence in the endogenous variable (i.e., it approximates a random walk), use of the standard Arellano-Bond estimator

power and proper exclusion. In all cases, the candidate instruments fail the diagnostic test for weak instruments.

[24] After transforming to first differences, testing for AR(1) in ϵ in the original estimation equation is equivalent to testing for AR(2) in $\Delta\epsilon$. The test statistic is given by $\sum_{m,t}\epsilon_{m,t}\epsilon_{m,t-2}$, which is normally distributed under the null of no serial autocorrelation.

can still result in a weak instruments problem, since lagged levels alone may have poor predictive power for first differences. This is the case in my application, as N has limited within-market variation (i.e., E85 market entry is rare). A potential solution to this problem, posed by Arellano and Bover (1995) and fully developed by Blundell and Bond (1998), is to make use of additional moment conditions to improve efficiency. The key insight of these papers is that, the equation in levels may be consistently estimated using instruments that are *orthogonal* to the fixed effects. Time invariant regressors may therefore enter the equation in levels, since such variables are clearly orthogonal to the market fixed effect. This is particularly convenient, as it provides a means to estimate the effects of time invariant regressors while maintaining controls for market-level unobservables. Under an additional assumption, described momentarily, lagged differences of the endogenous variables are valid instruments for the endogenous variable in levels. The instrument matrix for the "levels" equation may then be expressed as:

$$\Delta W_{1m} \equiv \begin{pmatrix} Z_{1m2} & 1_t & \Delta N_{m2} & 0 & 0 & ... & 0 \\ Z_{1m3} & 1_t & \Delta N_{m2} & \Delta N_{m2} & 0 & ... & 0 \\ \vdots & & & & & & \vdots \\ Z_{1mT} & 1_t & \Delta N_{m1} & \Delta N_{m2} & \Delta N_{m3} & ... & \Delta N_{m,T-2} \end{pmatrix} \tag{24}$$

Estimation proceeds by stacking the "levels" equation on top of the "differences" equation and applying the usual GMM procedure to the joint system. To summarize, the moment conditions for the consumer flex-fuel demand system may be written:

$$\begin{aligned} E[\Delta W_1 \epsilon_1] &= 0 \\ E[W_1 \Delta \epsilon_1] &= 0 \end{aligned} \tag{25}$$

The fleet demand equation takes a similar form, but the instrument matrices also include lags of dedicated-access E85 facilities ($\widetilde{N}$). The resulting "system GMM" estimator delivers highly efficient estimates under the stated assumptions.

The additional assumption necessary for doing "system GMM" is that correlation of the fixed effect with the endogenous variable is time-invariant. In the case of (18), this condition translates to $E(\delta_{1m}\Delta N_{mt}) = 0$, which implies a weak form of stationarity on the stochastic process generating N. The condition holds if there are no systematic deviations across markets in the observed *rate* of E85 retailer entry across markets with different mean levels of consumer utility for flex-fuel vehicles. In other words, departures from the steady-state of N is idiosyncratic across markets. If flex-fuel vehicle installed bases are in steady-state, we would expect this condition to hold, since firm profits would then be time invariant. To the extent that installed bases have not yet reached steady-state in the sampled data, we might expect $E(\delta_1 \Delta N) > 0$, since higher mean consumer flex-fuel utility should translate into more rapid growth of the installed base and hence a higher frequency of E85 retailer entry. Whether or not the assumption is problematic in finite samples turns on what factors are captured by the fixed effects δ across markets. However, problems arising from this condition should be reflected in tests of the exogeneity of the instruments, which are not rejected in my model.

6 Results

The results are presented in two sections. In Section 6.1, I present and discuss the main estimation results. I present robustness checks of many of the key model assumptions in Appendix D.

6.1 Main estimation results

6.1.1 Flex-fuel demand

I begin discussion of the results with the flex-fuel demand model estimates, which are presented in Table 5. The dependent variables are the log-odds ratios of flex-fuel purchase by consumers (H_1, first column) and fleets (H_2, second column). For brevity, I suppress estimates of fixed effects. I first comment on the diagnostic tests that substantiate the validity of the instruments employed. The null hypothesis of no serial autocorrelation in the residuals is not rejected at the 5% level in both the consumer and fleet equations, indicating use of the estimator is valid. For both demand systems, the test of overidentifying restrictions (H_0 : exclusion restrictions are valid) is also not rejected at the 5% level, suggesting that the lagged values of N (and $\widetilde{N}$) are properly excluded from the flex-fuel demand equation. I also perform a weak instruments test by testing the joint significance of lagged values of N in a regression of demand model controls on N. This is analogous to the F test of excluded instruments reported in the two stage least squares regressions. The large (>>10) value of this F statistic provides further evidence that the instruments have identifying power.

Focusing on the consumer flex-fuel demand estimates, the key parameter (α_1) is the coefficient on the number of E85 retailers (N). This coefficient captures the "demand side" of the network effect, i.e., the effect of E85 availability on consumer flex-fuel demand. As anticipated, this coefficient is positive and highly significant. The marginal effects of ethanol fuel entry on flex-fuel sales are compiled in the following table[25]:

[25] To deduce the marginal effect of changes E85 availability on flex-fuel vehicle sales (Q_1), two issues must be considered. First, evaluation of the marginal effect must account for the discrete nature of N. That is, the quantity of interest is not the elasticity $\frac{N}{Q}\frac{\partial Q_1(\alpha_1,N)}{\partial N}$ but rather the proportional change $\frac{Q_1(\alpha_1,N_b)-Q_1(\alpha_1,N_a)}{Q_1(\alpha_1,N_a)}$, where N_a and N_b are reference values of N. The leading case of changes in N in the data is from no incumbents to monopoly, for which $N_a = 0$ and $N_b = 1$. Second, computation of $Q_1(\alpha_1, N)$ must be in reference to some utility level that incorporates the mean contributions from Z_1 as well as the unobservable δ_1. I compute the reference utility level as sample average of the log-odds sales ratio, less the contribution from N, i.e. $\overline{U} = \frac{1}{MT}\sum_{m,t}(H_{mt} - \widehat{\alpha}_1 N_{mt})$. Normalizing the market population to $P = 1$ (it is immaterial for the calculation of percentage changes), I evaluate the marginal effect using the formula $Q_1(\alpha_1, N) = \frac{\exp\left(\alpha_1 N+\overline{U}\right)}{1+\exp\left(\alpha_1 N+\overline{U}\right)}$.

N	ΔQ_1
$0 \longrightarrow 1$	12.04%
$1 \longrightarrow 2$	12.02%
$2 \longrightarrow 3$	12.01%
$3 \longrightarrow 4$	11.99%

While the percentage increase in flex-fuel sales does decline in N, at the reference utility level the marginal effect of entry is roughly constant, with an additional ethanol fuel retailer leading to a 12.0% increase in flex-fuel sales. Thus, the model implies an economically significant increase in flex-fuel demand in response to E85 market entry.

The remaining coefficients (β_1) capture the effects of exogenous market-level observables (Z_1) on consumer flex-fuel vehicle demand. Of these, only *gas_ stations*, *ethanol_ plants*, and *auto_ dealers* are time-varying. The remaining covariates are time constant, and appear only in the "levels" equation of the GMM system. These regressors control for consumer demographics (*rural*, *income*, *age*, *male*), commuting patterns (*travel_ time*, *interstates*), retail vehicle availability (*auto_ dealers*), and fuel prices (*gas_ stations, ethanol_ plants*, *corn_ acres*). Signs of these coefficients generally conform to prior expectations. For example, flex-fuel sales are higher in markets with fewer gas stations (high gasoline prices) and more auto dealerships (greater availability of flex-fuel models). Controls for commuting patterns are insignificant. The demographic variables indicate greater demand for flex-fuel vehicles in rural, higher income regions.

Turning now to the fleet flex-fuel demand estimates, the coefficient on N (α_{21}) is also positive and significant, implying a positive feedback loop in fleet flex-fuel demand and E85 retailer entry. I compute the marginal effects of E85 retailer entry on fleet flex-fuel demand in the same manner as for the consumer model, and obtain:

N	ΔQ
$0 \longrightarrow 1$	25.64%
$1 \longrightarrow 2$	25.63%
$2 \longrightarrow 3$	25.61%
$3 \longrightarrow 4$	24.60%

The model predicts a greater demand response by fleets to E85 market entry than by consumers. This result seems intuitive, since many fleets operate under alternative fuel use mandates, and are therefore more likely to utilize E85 than consumers. The coefficient on private access E85 stations is insignificant after controlling for market fixed effects, as anticipated.

The exogenous market characteristics in the fleet demand equation (Z_2) control for firm characteristics (*avg_ salary*), commuting patterns (*interstates*), retail vehicle availability (*auto_ dealers*), and fuel prices (*gas_ stations*, *ethanol_ plants*, *corn_ acres*). The number of car rental agencies (*auto_ rentals*) is included, since the alternative fuel preferences of these firms (which constitute the majority of large fleets) may differ systematically from other types of vehicle fleets. The number of businesses in the market, *total_ establishments,* is a control for systematic differences in the market size definition, which is based on the number of persons employed in the market. In effect, this controls for the effect of concentration of employment on the potential market size. The general pattern of these parameter estimates resembles that in the consumer equation. However, the "income" effect is reversed – flex-fuel purchase is more likely in markets with lower paying firms. The *rural* variable stands out as particularly material, which may reflect systematic differences in fleet composition or usage patterns in more agrarian regions.

Table 5: Main estimation results - flex-fuel vehicle demand

	Equation	
	Consumer	Fleet
retail E85 stations (N)	0.114***	0.229***
	(0.029)	(0.077)
gas	-0.006***	-0.034***
	(0.001)	(0.003)
refineries	0.032	-0.017
	(0.061)	(0.069)
dealers	0.006**	0.025***
	(0.002)	(0.006)
corn	-0.008***	-0.058***
	(0.002)	(0.003)
interstates	-0.004	-0.096***
	(0.019)	(0.024)
rural	0.140***	1.833***
	(0.038)	(0.041)
income	0.007***	
	(0.001)	
age	0.006	
	(0.003)	
male	0.027***	
	(0.006)	
commute	-0.003	
	(0.002)	
private E85 stations ($\widetilde{N}$)		0.236
		(0.342)
rentals		0.034
		(0.023)
total_ stablishments		-0.000***
		(0.000)
avg_ alary		-0.007***
		(0.002)
Market fixed effects	Yes	Yes
Period fixed effects	Yes	Yes
Observations	41292	41292
RMSE	0.80	1.18
Weak id F	48.18	25.05
Overid Chi-squared (dof)	5.72 (2)	4.14 (2)
Overid p-value	0.06	0.13
AR(1) Z	1.685	-1.697
AR(1) p-value	0.092	0.090

Standard errors based on 30 bootstrap replications
Significance levels: * p<0.05, ** p<0.01, *** p<0.001

6.1.2 Fuel Retailer Entry

The E85 market entry estimates are presented in Table 7. The estimated correlation coefficients ρ_{13} and ρ_{23} are insignificant. I interpret this result as indicating that the model specification fully captures the relevant variation through the control variables. Thus, when exploring the robustness of my results to other assumptions in Appendix D, I use a compuptationally more convenient entry model estimator that assumes independent shocks to E85 profitability and flex-fuel vehicle demand.

Entry model regressors are divided into three categories, corresponding to fixed cost shifters (Z_3), variable profit shifters (Z_4) and market size shifters (Z_5). For fixed costs, the primary controls are state and period fixed effects. The *rural* and *income* variables are included as fixed cost shifters, under the premise that equipment and labor input prices may vary with these factors. As one might anticipate, fixed costs appear to be higher in more rural, higher income regions. The negative coefficient on *gas_stations* presumably reflects local economies of scale in servicing the infrastructure requirements of service stations (e.g., installation of tanks and pumps). Similarly, *ethanol_plants* should correlate positively with availability of ethanol-related infrastructure and installation services. However, this effect is not significant.

The variable profit shifters include demographics that may correlate with fuel type preferences (*rural*, *income*), fuel consumption patterns (*travel_time*), competition concentration (*land_area*), wholesale ethanol prices (*corn_acres*, *ethanol_plants*), and substitute fuel price (*gas_stations)*. In general, the estimates are as expected – profits are higher where wholesale ethanol prices are low (more corn, ethanol plants), competition is more geographically dispersed (larger area), and substitute fuel prices are high (fewer gas stations). The demographic factors suggest rural, lower income flex-fuel vehicle owners are more likely to purchase E85. The fact that variable profits decrease in income is somewhat counterintuitive, since presumably higher income consumers would have higher willingness to pay for E85, but *income* may also capture some type of costs particular to higher vehicle density regions. The negative coefficient on *travel_time* is also puzzling, since long travel times should reflect greater fuel consumption. However, another interpretation of this result is that fuel is purchased with greater probability outside the consumer's home market. All coefficients are statistically significant.

The market size coefficients are particularly interesting to interpret, as units have been normalized to consumer flex-fuel vehicles. The only exogenous market shifter used in the market size expression is *interstates,* which is assumed to capture E85 consumption by flex-fuel vehicles in long-range commuting patterns. The estimate implies the presence of an interstate highway has the same contribution to retailer profits as 12.2 consumer flex-fuel vehicles in the installed base. In a similar fashion, the coefficient on the fleet flex-fuel installed base (γ) parameter indicates that it takes $1/.14 = 7.1$ fleet flex-fuel vehicles to equal the profit contribution of one consumer flex-fuel vehicle. That fleet vehicles contribute less to retailer profits is somewhat surprising, given that fleets frequently operate under alternative fuel usage mandates while consumers have no such restriction. In his study, Corts (2010) finds that it takes about twice as many consumer flex-fuel

Table 6: E85 market entry thresholds in consumer flex-fuel vehicles

	Required installed base	
E85 stations	Monopoly conditions	Average conditions
1	203.6	256.2
2	458.1	576.4
3	814.3	1024.7
4	1272.4	1601.0

vehicles as government fleet flex-fuel vehicles to support one E85 station. Both the data and methods differ, so a direct comparison of the results is difficult. In particular, fleet vehicles in my study are predominantly comprised of corporate vehicles, whereas in Corts (2010) "fleets" are exclusively government vehicles. It should be noted that my result is not an artifact of the model structure, as the descriptive regressions in Table 4 show. The effect of the fleet installed base on the number of E85 retailers is smaller (and insignificant), suggesting the differences between the studies relates to the data. Inasmuch as Corts (2010) uses a cross-section of all current registrations (as opposed to my time series of vehicle registrations), there may be significant differences due to the initial conditions of the fleet installed base, which I must model indirectly.

Interpretation of the entry model results is facilitated by computing the number of flex-fuel vehicles required to support a given number of E85 retailers, or the "entry thresholds" in the language of Bresnahan and Reiss (1991). The "supply side" of the network effect essentially operates through this mechanism: as the installed base of flex-fuel vehicles increases, the market becomes more profitable to serve, and at certain threshold values of the installed base, additional entry becomes feasible. I report the entry thresholds reported in Table 6. The threshold values are computed from the entry model estimates using the formula: $S_N^* = \frac{\exp\left(\widehat{\psi}'\overline{Z}_4\right)}{\widehat{\phi}'\overline{Z}_5}(N+1)^2$. I report entry thresholds with respect to average monopoly conditions (i.e., $\overline{Z}_4$ and $\overline{Z}_5$ are computer for observations with $N = 1$) and for the full sample. The two measures are reported since there appear to be systematic differences between markets that can support E85 retailers and those which do not, much of which can be attributed to the state effects in fixed costs. Under typical monopoly conditions, the model predicts that at least 204 consumer flex-fuel vehicles are required to support a single ethanol fuel retailer. Using the entire sample average increases the monopoly thresholds by approximately 50 vehicles. These estimates compare very favorably with the observed installed bases (see Table 3) and the estimates by the Environmental Protection Agency (Department of Transportation and the Environmental Protection Agency (2002)), which suggest about 200 flex-fuel vehicles are required. Corts (2010) estimates the required number of consumer flex-fuel vehicles to support a station to be somewhat higher, in the range of 320 to 560 vehicles.

Table 7: Main estimation results - E85 market entry

	Variable profits (ψ)	Fixed costs (ϕ)	Market size (γ, λ)	Covariance (ρ)
rural	1.662***	0.595***		
	(0.114)	(0.093)		
income	-0.014***	0.005*		
	(0.002)	(0.002)		
gas_ stations	-0.049***	-0.039***		
	(0.004)	(0.006)		
ethanol_ plants	0.344	-0.259		
	(0.197)	(0.246)		
travel_ time	-0.034***			
	(0.006)			
corn_ acres	0.037***			
	(0.004)			
land_ area	-0.002***			
	(0.000)			
constant	-2.338***			
	(0.190)			
fleet FFV installed base			0.141**	
			(0.047)	
interstates			12.231**	
			(1.912)	
ρ_1				0.033
				(0.070)
ρ_2				-0.017
				(0.138)
State fixed effects	No	Yes	No	-
Period fixed effects	No	Yes	No	-

Standard errors based on 30 bootstrap replications
Significance levels: * p<0.05, ** p<0.01, *** p<0.001

7 Applications and strategic implications

The results presented in the previous section provide strong evidence of an indirect network effect in the market for ethanol-compatible vehicles and fuel. In this section, I demonstrate application of the results and discuss the strategic implications for firms. The first counterfactual experiment quantifies the cumulative effect of positive feedback arising from flex-fuel vehicle demand dependence on E85 availability. This descriptive exercise provides insight into the long-run significance of the network effect across a sample of heterogeneous markets, which cannot be easily inferred from the point estimates of marginal effects calculated in Section 6.1.1. The remaining experiments evaluate promotional strategies firms may implement to accelerate the adoption of the ethanol fuel standard, and thereby generate additional demand for their products. In these experiments, I take the perspective of a flex-fuel vehicle manufacturer and evaluate the effect of offering promotional subsidies to fuel retailers to enter the E85 market.

To reduce the computational burden of evaluating the counterfactual experiments, I focus on a random sample of 400 markets in Midwestern states.

7.1 Quantifying the network effect

Feedback in the econometric model is realized through two mechanisms. The first source of feedback originates from the simultaneous determination of flex-fuel demand and E85 entry, and operates within the period outcomes are realized. The "demand side" of this feedback loop is compactly summarized by the marginal effects of N on flex-fuel demand computed in Section 6.1.1. The second source of feedback is the persistence of flex-fuel demand in the installed base, which operates across periods. Here, I quantify the long-run impact of both effects in sample market conditions.

The method is to first simulate market outcomes assuming the model estimates are the true parameters governing the data generating process (the "baseline"). Then, the simulation is repeated under the assumption that neither consumers nor fleets have utility for E85 (i.e. I set $\alpha_1 = \alpha_{21} = 0$). Simulation of market outcomes requires the following series of steps. First, for each market observation, I draw market shocks from the distribution (14). Next, I solve the model for the equilibrium values of Q_1^*, Q_2^* and N^* period by period, updating the installed base sequentially. Third, I repeat the first two steps multiple (30) times and average the result to obtain the expected value of the market outcomes.

I summarize the counterfactual graphically in Figures (1) and (2) below. In these plots, the installed bases and number of E85 retailers are aggregated across the sample markets. At any point in time, the cumulative influence of the network effect is given by the difference between the baseline and no-feedback curves. In the final period (2006), the network effect accounts for 9.4% of the predicted number of E85 retailers. Similarly, the network effect accounts for 27.5% of the total flex-fuel vehicle installed base. Indirect network effects of this size are economically material, but moderate in comparison to those observed in high-tech product markets.[26]

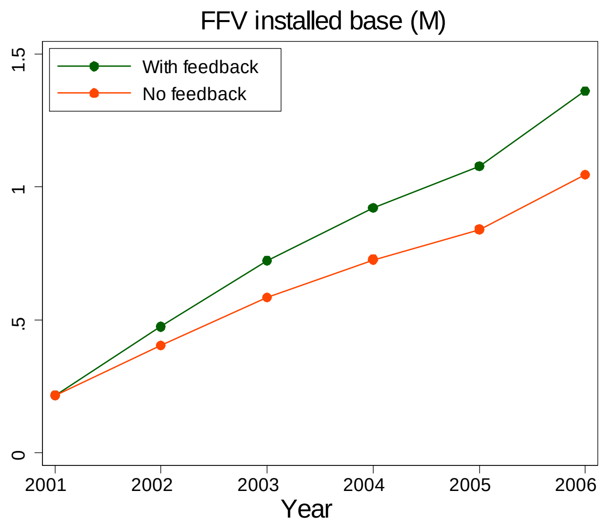

Figure 1: Flex-fuel installed base with and without feedback

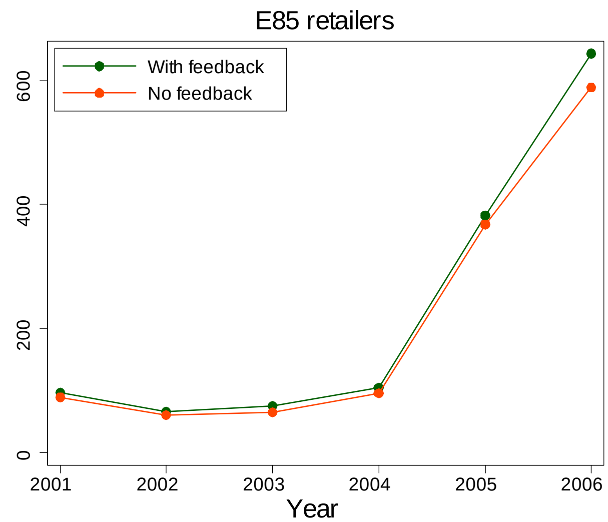

Figure 2: Number of E85 retailers with and without feedback

7.2 Subsidy of Ethanol Market Entry

From the marketer's perspective, the externalities which characterize an indirect network effect are a double-edged sword. On the one hand, existing availability of a complementary good increases demand for the marketer's product at no additional expense. On the other, when an important complement is scarce or unavailable, demand may be severely depressed or curtailed entirely. In the latter circumstance, neither producer of the complementary good pair wishes to make unilateral investments in expanding their product availability, since such efforts may not be reciprocated. A solution to this chicken-or-egg problem is to devise co-marketing strategies that "internalize" the network effect – i.e., contractual obligations that align firm incentives. I assess two arrangements of this type, in which subsidies are provided by vehicle manufacturers to fuel retailers in order to encourage E85 market entry. The first subsidy policy considers the benefits and costs of a fixed-rate subsidy that reduces the fixed cost of market entry by 10% for all potential entrants in all time periods. The second policy allows for a variable-rate subsidy, where the subsidy level is set by the vehicle manufacturer. The subsidy offered equals the amount required to induce one more entrant than the un-subsidized market equilibrium would support. The subsidy is only offered in markets where profits from additional vehicle unit sales in the current period exceed the cost of the subsidy.

[26]For example, Nair, Chintagunta, and Dubé (2004) find that indirect network effects in PDA hardware/software adoption account for 22% of the installed base of PDAs over a period roughly half the length of my study.

To assess profits in dollar terms, one would need to know the manufacturer margin per flex-fuel vehicle and the normalization required to scale E85 retailer fixed costs into dollar units. Outcomes, however, only depend on the ratio of these two quantities. I simplify matters by normalizing the unit of profit to be the average manufacturer margin, and assume a ratio of 1:100 for vehicle margins to average E85 fixed costs. This would be consistent with manufacturer profits of $50 per flex-fuel vehicle and $5000 average E85 fixed costs.[27]

The subsidy policies are computed in a manner analogous to that described in the previous section. For the "fixed-rate" subsidy, I evaluate the effect of a 10% reduction in fixed costs by adjusting the fixed cost intercept value. For each firm that enters, the vehicle manufacturer is assessed a subsidy cost of 10%*100 = 10 units. The variable rate, or "market optimized", policy allows for targeting the most profitable markets for subsidy. Manufacturers only extend the offer of subsidy to markets for which it is profitable and pay the minimum subsidy required to induce an additional entrant. I evaluate the policy by solving for the change in fixed costs required to induce $N^* + 1$ E85 retailers in equilibrium. If this cost (in vehicle units) is smaller than the incremental flex-fuel unit sales with $N^* + 1$ E85 retailers, market outcomes corresponding to $N^* + 1$ ethanol retailers are assigned and the vehicle manufacturer is credited with the resulting net profit. Otherwise, the net profit is simply the unit sales with N^* retailers.

Figures (3) to (6) below summarize the effect of these counterfactual policies. Plotted values are the incremental changes to the quantities of interest from the baseline simulation, which capture the cumulative effect of the policy. Figures (3) to (5) demonstrate that, relative to the fixed rate policy, the variable rate policy results in 40% higher consumer/fleet flex-fuel installed bases in the final period, while inducing 6% *less* E85 market entry. The profit implications of the two policies are summarized in Figure 6. Over the course of the sample period, actively targeting markets for subsidies results in a 50% increase in cumulative profits relative to the blanket "fixed-rate" policy. This simple experiment clearly demonstrates that the network effect may be harnessed to improve vehicle manufacturer profitability, and that an optimal incentive policy will incorporate knowledge of local market conditions.

[27] Margins on new vehicles sold in the US varies greatly by manufacturer. Japanese automakers lead in profitability with pretax margins in excess of $1200 per vehicle, while the Big Three domestic manufacturers (which produce the bulk of flex-fuel vehicles) typically earn less than $250 per vehicle. For example, GM actually lost $2500 per domestic vehicle sold in 2005. Source: http://www.washingtonpost.com/wp-dyn/content/article/2006/06/01/AR2006060102083.html

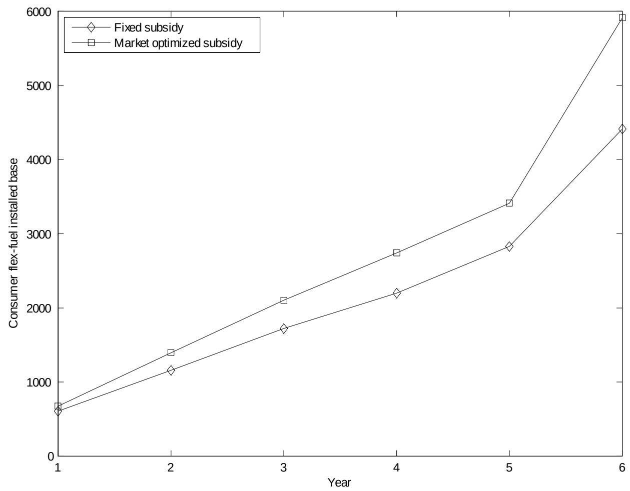

Figure 3: Subsidy counterfactual - consumer flex-fuel installed base

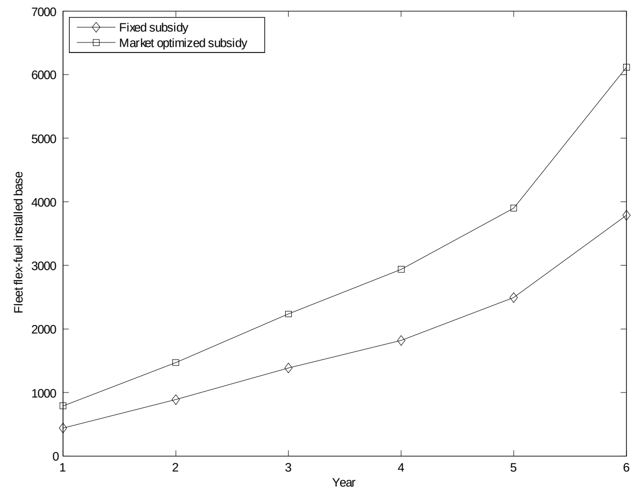

Figure 4: Subsidy counterfactual - fleet flex-fuel installed base

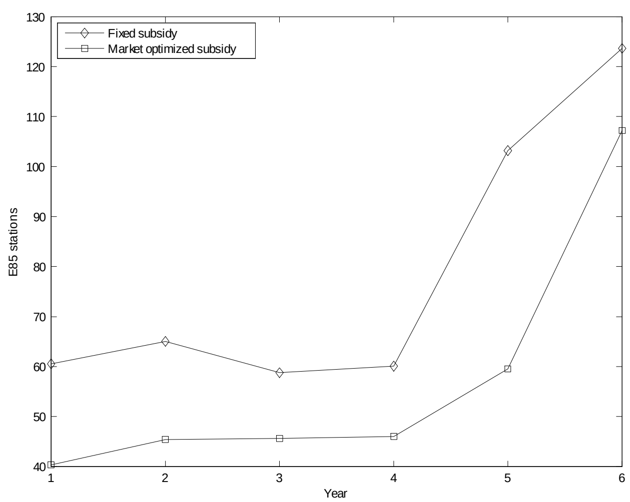

Figure 5: Subsidy counterfactual - number of retail E85 stations

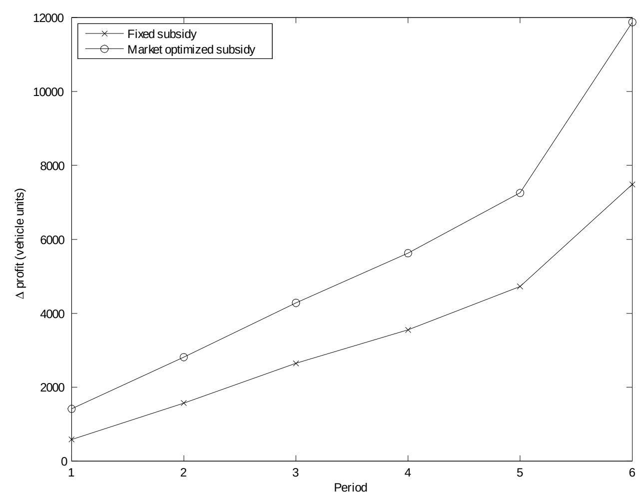

Figure 6: Subsidy counterfactual - incremental profits in vehicle units

8 Conclusion

This paper examines the role of network effects in the demand for ethanol-compatible vehicles and retailer entry into the market for ethanol fuel. The network effect arises in this market due to complementarities in the availability of ethanol fuel and the installed base of ethanol-compatible vehicles. To measure these effects empirically, I develop a simultaneous equations model of consumer flex-fuel vehicle demand, fleet flex-fuel vehicle demand, and E85 retailer market entry.

The model extends the Bresnahan and Reiss (1991) model of competitive entry to incorporate an endogenous market size. I develop a new estimator for the model that accommodates the simultaneity induced by the co-determination of ethanol-compatible vehicle demand and ethanol fuel supply. correlated shocks to the demand and entry systems, and that does not require the use of instruments for the entry equation. In contrast to most studies of indirect network effects, the feedback mechanism investigated here is spatial in nature and operates at a highly localized level. To identify these highly localized effects, I estimate the model using a panel of zip code level observations. These data incorporate the entire population of vehicle registrations and ethanol fuel market entry events in six states over six years. The rich panel structure allows me to control for unobservables and to correct for simultaneity bias in the vehicle demand estimation by using lagged endogenous variables as instruments. The model estimates provide strong evidence of a network effect, with both statistical and economic significance. Under typical market conditions, entry of an E85 retailer leads to a 12.0% increase in consumer demand for flex-fuel vehicles and a 27.5% increase among fleets. The entry model predicts that an E85 retailer requires an installed base of at least 204 flex-fuel vehicles to be profitable. I apply the estimates in a series of counterfactual policy experiments in which I further quantify the network effect and explore strategies to improve ethanol-compatible vehicle manufacturer profitability by leveraging the network effect. I find market-optimized subsidies provided by vehicle manufacturers to fuel retailers to be highly effective in enhancing profitability.

There are several possibilities for extensions to the paper. First, the assumption of identical firms could be relaxed in the entry model. One approach would restrict the set of potential E85 entrants to existing service stations, and to obtain data on those firm characteristics. These data are currently unavailable, however. Since scale economies in ethanol distribution are likely an important determinant of observed patterns of entry behavior, incorporating chain effects could have important implications for estimates of the network effect. This extension will considerably complicate an already computation-intensive estimation procedure, however, since the identities of entering firms must be determined in equilibrium. Another extension might consider in greater depth the alternative fuel vehicle choices of fleets, in combination with the decision to build dedicated refueling facilities. A wider array of alternative fuel vehicles are available to fleets, including compressed natural gas (CNG), liquefied natural gas (LNG), propane based fuels. It would be interesting to consider these fuel types (which are not backwardly compatible with gasoline) in the context of a multinomial choice model.

References

Anderson, S. (2006): "The Demand for E85 Ethanol," Discussion paper, Department of Economics, University of Michigan, Ann Arbor.

Anderson, S., and J. Sallee (2009): "Using loopholes to reveal the marginal cost of regulation: The case of fuel-economy standards," Discussion paper, Harris School of Public Policy Studies, University of Chicago.

Arellano, M., and S. Bond (1991): "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The Review of Economic Studies*, 58(2), 277–297.

Arellano, M., and O. Bover (1995): "Another look at the instrumental variable estimation of error-components models," *Journal of econometrics*, 68(1), 29–52.

Basu, A., T. Mazumdar, and S. Raj (2003): "Indirect Network Externality Effects on Product Attributes," *Marketing Science*, 22(2), 209–221.

Berry, S. (1992): "Estimation of a model of entry in the airline industry," *Econometrica*, 60(4), 889–917.

——— (1994): "Estimating discrete-choice models of product differentiation," *RAND Journal of Economics*, 25, 242–262.

Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile prices in market equilibrium," *Econometrica*, 63, 841–890.

Berry, S., and P. Reiss (2007): "Empirical models of entry and market structure," *Handbook of Industrial Organization*, 3, 1845–1886.

Berry, S., and J. Waldfogel (1999): "Free entry and social inefficiency in radio broadcasting," *The RAND Journal of Economics*, pp. 397–420.

Blundell, R., and S. Bond (1998): "Initial conditions and moment restrictions in dynamic panel data models," *Journal of econometrics*, 87(1), 115–143.

Bresnahan, T., and P. Reiss (1990): "Entry in Monopoly Markets," *The Review of Economic Studies*, 57(4), 531–553.

Bresnahan, T., and P. Reiss (1991): "Entry and competition in concentrated markets," *Journal of Political Economy*, 99(51), 977–1009.

Caillaud, B., and B. Jullien (2003): "Chicken & Egg: Competition among Intermediation Service Providers," *The RAND Journal of Economics*, 34(2), 309–328.

Chamberlain, G. (1984): "Panel data," *Handbook of Econometrics*, pp. 1247–1318.

Chou, C., and O. Shy (1990): “Network Effects without Network Externalities,” *International Journal of Industrial Organization*, 8(2), 259–270.

Church, J., and N. Gandal (1992): “Network Effects, Software Provision, and Standardization,” *The Journal of Industrial Economics*, 40(1), 85–103.

——— (1993): “Complementary Network Externalities and Network Adoption,” *International Journal of Industrial Organization*, 11, 239–260.

Clements, M., and H. Ohashi (2005): “Indirect Network Effects and the Product Cycle: Video Games in the US, 1994-2002,” *Journal of Industrial Economics*, 53(4), 515–542.

Corts, K. (2010): “Building out alternative fuel retail infrastructure: Government fleet spillovers in E85,” *Journal of Environmental Economics and Management*.

Davis, S., and S. Diegel (2006): “Transportation Energy Data Book: Edition 25,” Discussion paper, ORNL/TM-2007/070, Oak Ridge National Laboratory (ORNL).

Department of Transportation, D. o. E., and the Environmental Protection Agency (2002): “Report to Congress: Effects of the Alternative Motor Fuels Act CAFE Incentives Policy,” Discussion paper, United States Congress.

Draganska, M., and M. Mazzeo (2003): “The Impact of Quality and Variety on Product Assortment Decisions: An Empirical Investigation,” working paper, Stanford University.

Dranove, D., and N. Gandal (2003): “The Dvd-vs.-Divx Standard War: Empirical Evidence of Network Effects and Preannouncement Effects,” *Journal of Economics & Management Strategy*, 12(3), 363–386.

Dubé, J., G. Hitsch, and P. Chintagunta (2007): “Dynamic Standards Competition and Tipping: The Case of 32/64 Bit Video Game Consoles,” *manuscript, Chicago GSB.*

Ellickson, P., and S. Misra (2007): “Enriching interactions: Incorporating revenue and cost data into static discrete games,” Discussion paper, Manuscript, University of Rochester.

Eppli, M., and J. Benjamin (1994): “The evolution of shopping center research: a review and analysis,” *Journal of Real Estate Research*, 9(1), 5–32.

Farrell, J., and G. Saloner (1985): “Standardization, Compatibility, and Innovation,” *The RAND Journal of Economics*, 16(1), 70–83.

——— (1986): “Installed Base and Compatibility: Innovation, Product Preannouncements, and Predation,” *The American Economic Review*, 76(5), 940–955.

Gandal, N., M. Kende, and R. Rob (2000): “The Dynamics of Technological Adoption in Hardware/Software Systems: The Case of Compact Disc Players,” *The RAND Journal of Economics*, 31(1), 43–61.

Gupta, S., D. Jain, and M. Sawhney (1999): “Modeling the evolution of markets with indirect network externalities: An application to digital television,” *Marketing Science*, 18(3), 396–416.

Hansen, L. (1982): “Large sample properties of generalized method of moments estimators,” *Econometrica: Journal of the Econometric Society*, 50(4), 1029–1054.

Johnson, C., and M. Melendez (2007): “E85 Retail Business Case: When and Why to Sell E85,” Discussion paper, National Renewable Energy Laboratory, Technical Report TP-540-41590.

Karaca-Mandic, P. (2004): “Network Effects in Technology Adoption: The Case of DVD Players,” Discussion paper, RAND Institute.

Katz, M., and C. Shapiro (1985): “Network Externalities, Competition, and Compatibility,” *The American Economic Review*, 75(3), 424–440.

——— (1986): “Technology Adoption in the Presence of Network Externalities,” *The Journal of Political Economy*, 94(4), 822–841.

——— (1992): “Product Introduction with Network Externalities,” *The Journal of Industrial Economics*, 40(1), 55–83.

——— (1994): “Systems Competition and Network Effects,” *The Journal of Economic Perspectives*, 8(2), 93–115.

Lee, R. S. (2009): “Vertical Integration and Exclusivity in Platform and Two-Sided Markets,” Working paper series, NET Institute.

Liu, H. (2006): “Dynamics of Pricing in the Video Game Console Market: Skimming or Penetration,” .

Moore, G. (1995): *Crossing the Chasm*. HarperCollins Publishers, New York.

Musalem, A., and W. Shin (2009): “When Demand Projections Are Too Optimistic: A Structural Model Of Product Line And Pricing Decisions,” Discussion paper, Manuscript, Duke University.

Nair, H., P. Chintagunta, and J. Dubé (2004): “Empirical Analysis of Indirect Network Effects in the Market for Personal Digital Assistants,” *Quantitative Marketing and Economics*, 2(1), 23–58.

Ohashi, H. (2003): “The Role of Network Effects in the US VCR Market, 1978-1986,” *Journal of Economics & Management Strategy*, 12(4), 447–494.

Park, S. (2004): “Quantitative Analysis of Network Externalities in Competing Technologies: The VCR Case,” *Review of Economics and Statistics*, 86(4), 937–945.

Pettigrew, H., J. Gart, and D. Thomas (1986): "The bias and higher cumulants of the logarithm of a binomial variate," *Biometrika*, 73(2), 425–435.

Reiss, P., and P. Spiller (1989): "Competition and entry in small airline markets," *Journal of Law and Economics*, 32.

Rysman, M. (2004): "Competition Between Networks: A Study of the Market for Yellow Pages," *Review of Economic Studies*, 71(2), 483–512.

Shankar, V., and B. Bayus (2003): "Network effects and competition: an empirical analysis of the home video game industry," *Strategic Management Journal*, 24(4), 375–384.

Stock, J., and M. Yogo (2002): "Testing for weak instruments in linear IV regression," Discussion paper, National Bureau of Economic Research.

Vitorino, M. (2007): "Empirical entry games with complementarities: An application to the shopping center industry," Discussion paper, University of Chicago.

Windmeijer, F. (2005): "A finite sample correction for the variance of linear efficient two-step GMM estimators," *Journal of econometrics*, 126(1), 25–51.

Wooldridge, J. (2002): *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Appendices

A Visualization of flex-fuel installed base and E85 availability

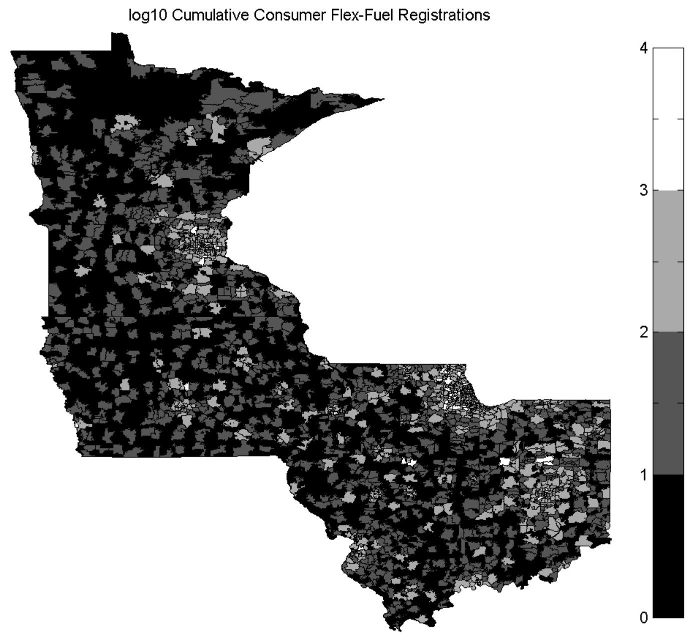

Figure 1: Log installed base of flex-fuel vehicles, 2001

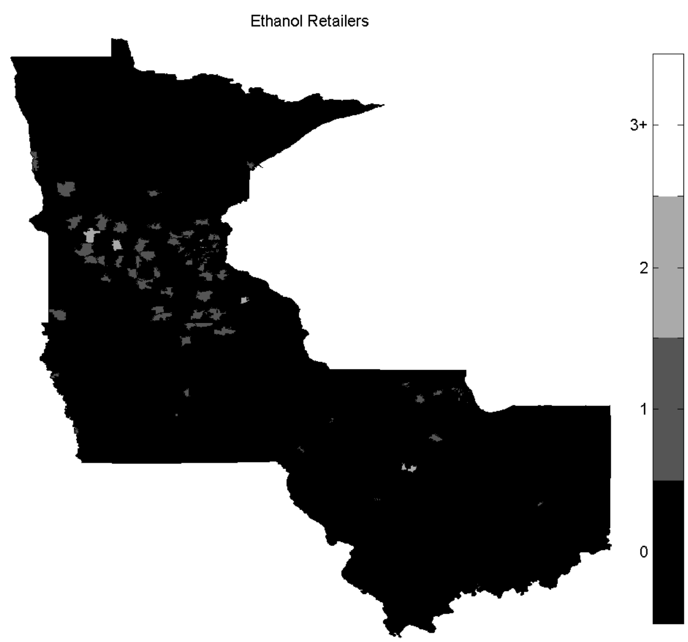

Figure 2: E85 availability (station counts), 2001

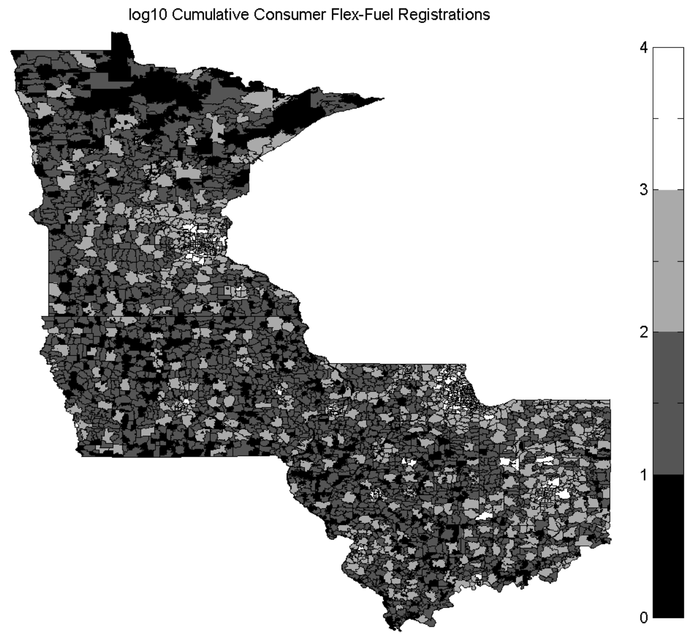

Figure 3: Log installed base of flex-fuel vehicles, 2001

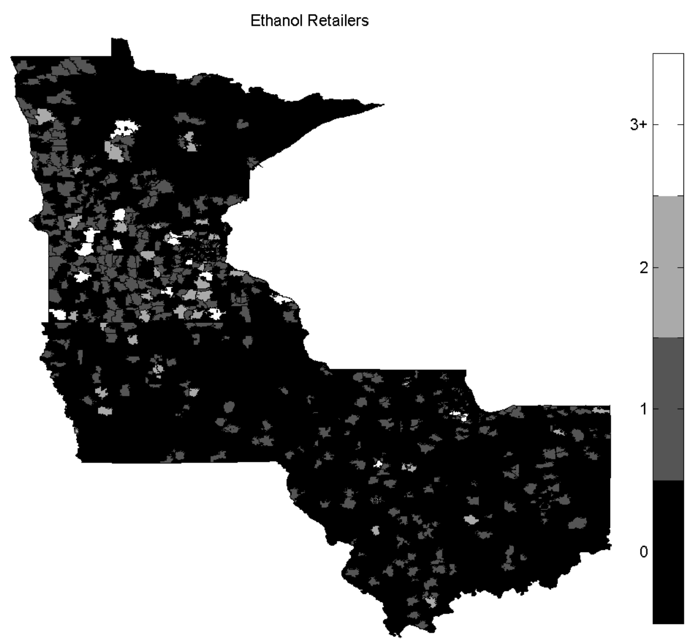

Figure 4: E85 availability (station counts), 2006

B Derivation of the Entry Equilibrium Constraint

To derive the constraint condition imposing $\frac{\partial \Pi^N}{\partial N} \leq 0$, I suppress all subscripts and introduce the following shorthand notations relating to equations (3) and (5). Let:

$$H_1 = \ln\left(\frac{Q_1}{P_1 - Q_1}\right) = \alpha_1 N + k_1$$
$$where: k_1 = \beta_1' Z_1 + \delta_1 + \omega_1 + \epsilon_1$$

$$H_2 = \ln\left(\frac{Q_2}{P_2 - Q_2}\right) = \alpha_{21} N + k_2$$
$$where: k_2 = \alpha_{22}\widetilde{N} + \beta_2' Z_2 + \delta_2 + \omega_2 + \epsilon_2$$

Then the current period consumer and fleet flex-fuel sales may be written:

$$Q_1 = P_1 \frac{\exp(\alpha_1 N + k_1)}{1 + \exp(\alpha_1 N + k_1)} = P_1 \frac{\exp(H_1)}{1 + \exp(H_1)}$$

$$Q_2 = P_2 \frac{\exp(\alpha_1 N + k_1)}{1 + \exp(\alpha_1 N + k_1)} = P_2 \frac{\exp(H_2)}{1 + \exp(H_2)}$$

Differentiating Q_1 with respect to N yields:

$$\frac{\partial Q_1}{\partial N} = \frac{\partial}{\partial N} P_1 \frac{\exp(\alpha_1 N + k_1)}{1 + \exp(\alpha_1 N + k_1)} = \alpha_1 P_1 \frac{\exp(\alpha_1 N + k_1)}{(1 + \exp(\alpha_1 N + k_1))^2} = \frac{\alpha_1 Q_1}{1 + \exp(H_1)}$$

Similarly, for Q_2 we obtain:

$$\frac{\partial Q_2}{\partial N} = \frac{\alpha_1 Q_2}{1 + \exp(H_2)}$$

From equation (11) we have:

$$\begin{aligned} \frac{\partial S}{\partial N} &= \frac{\partial}{\partial N}\big((B_1 + Q_1) + \gamma (B_2 + Q_2) + \lambda' Z_5\big) \\ &= \frac{\partial Q_1}{\partial N} + \gamma \frac{\partial Q_2}{\partial N} \end{aligned}$$

Using these side calculations and defining $V = \exp(\psi' Z_4) > 0$, the constraint $\frac{\partial \Pi^N}{\partial N} \leq 0$ takes the form:

$$
\begin{aligned}
\frac{\partial \Pi^N}{\partial N} &= \frac{\partial}{\partial N}\left[S(N)\frac{V}{(1+N)^2} - F\right] \leq 0 \\
&\implies \frac{\partial S}{\partial N}\frac{V}{(1+N)^2} - \frac{2SV}{(1+N)^3} \leq 0 \\
&\implies \left(\frac{\partial Q_1}{\partial N} + \gamma\frac{\partial Q_2}{\partial N}\right)\frac{1}{(1+N)^2} - \frac{2S}{(1+N)^3} \leq 0 \\
&\implies \frac{\alpha_1 Q_1}{1+\exp\left(H_1\right)} + \frac{\gamma\alpha_{21} Q_2}{1+\exp\left(H_2\right)} - \frac{2S}{1+N} \leq 0
\end{aligned}
$$

C Alternative flex-fuel demand instruments

In this appendix, I report flex-fuel demand system estimates using a variety of instrumentation strategies. The intent here is to demonstrate the challenge of finding valid "external" instruments for identification in my application and to motivate the use of the "system GMM" procedure of Blundell and Bond (1998), which implements a time series identification strategy based on lagged values of the instrumented variables. Correcting for the endogeneity of E85 market entry in the flex-fuel demand equations (3) and (5), requires finding variables that shift E85 availability independently of flex-fuel demand. The primary difficulty in finding "external" instruments arises due to the fact that few data sources can match the richness of variation *within* zip codes I observe with the Polk vehicle registration data. One method to address this shortcoming would be to aggregate the flex-fuel data across time into a single cross-section, and search for time-static zip code level instruments. However, a cross-sectional approach eliminates the possibility of controlling for unobserved factors in the zip code that influence flex-fuel demand, which runs counter to the goal of identifying highly localized effects (and would severely weaken the credibility of the independent markets assumption taken during estimation). An additional concern is that many shifters of E85 availability will tend to be correlated with flex-fuel demand through their influence on E85 prices.

The search for "external" instruments begins with an inspection of equation (12). The form of the E85 retailer profit function clarifies that instruments for N fall into one of three categories: shifters of market size, shifters of variable profits, or shifters of fixed costs (collectively, I refer to these as "profit shifter" instruments). Previous studies of market entry, e.g. Bresnahan and Reiss (1991), have typically used shifters of market size (such as the market population) as instruments for N. Care must be taken with this approach in my application, since the market size is defined in terms of the flex-fuel installed base, which is endogenous through current period vehicle demand. However, lagged values of the installed bases are predetermined in the current period and therefore should be valid instruments. In the estimates reported in Tables 9 and 10 below, I designate the set of instruments comprised by the lagged consumer and fleet installed bases of flex-fuel vehicles as "MS".

In my study, the most convincing category of exclusion restrictions would involve shifters of E85 retailer fixed costs. These factors would have the greatest theoretical rationale for proper exclusion from the vehicle demand equations since fixed costs should not influence E85 prices and are independent of the market size. To develop a set of fixed cost shifters, I reviewed sample installation costs for E85 infrastructure.[28] Inspection of these quotations reveals that the primary cost components are: (1) ethanol-compliant tanks and pumps, and (2) site preparation services, including excavation and concrete pouring. I therefore constructed instruments for fixed costs using yearly data on the number of establishments in a zip code engaging in petroleum equipment wholesaling (NAICS 424720), site preparation contracting (NAICS 238910), and ready-mix concrete

[28] Available at http://www.afdc.energy.gov/afdc/ethanol/cost.html.

manufacturing (NAICS 327320).[29] These data are given the variable names *petro_wholesalers*, *siteprep_contractors* and *cement_plants*, respectively. I refer to these fixed cost instruments as "FC".

Shifters of fuel retailer variable profits are the final category of potential instruments. Using variable profit shifters as instruments for N requires the strong assumption that vehicle choices depend upon E85 availability, but not its price. To see this, note that variable cost factors c enter equilibrium E85 prices, as shown in equation (9). As a practical matter, an instrument of this class should have strong explanatory power for N but little influence on vehicle demand. Referring to the descriptive regressions in Table 4, the number of ethanol plants in a market is a promising instrument that meets these criteria. In general, pure ethanol will be less expensive in markets where a refinery is present since transportation costs are lower, and where multiple refineries are present, competition should further lower E85 input costs. Similarly, the number of bulk petroleum terminals in a market should influence the variable cost of E85, since ethanol and gasoline are typically sold at such locations. I therefore construct a set of instruments, "VP", using the variables *ethanol_plants* and *fuel_terminals*. Note that in other model specifications, *ethanol_plants* appears as a regressor in the demand equations as a control for local ethanol fuel prices. Since the estimated coefficient for *ethanol_plants* is uniformly insignificant, excluding it from the demand equations is not of great concern.

Before evaluating and comparing the results, I consider the anticipated direction of the simultaneity bias the instruments seek to correct. A priori, I expect N to be positively correlated with flex-fuel vehicle demand shocks. Presumably, positive shocks to flex-fuel vehicle demand reflect unobserved market conditions which are also conducive to E85 market entry. In this case, the coefficients on N in the vehicle demand equations will be biased away from zero since the partial correlation with the demand shocks will be attributed to the (positive) coefficients on N. In other words, not correcting for the simultaneity will overstate the network effect in the demand estimates. However, one cannot rule out the possibility that the reverse relationship holds, i.e., that N could be negatively correlated with shocks to flex-fuel demand. One plausible scenario is that is an increase in demand for gasoline, which would lead to higher retail gas prices, would then drive consumers to buy more flex-fuel vehicles, while fuel retailers would be inclined to abandon selling E85 in favor of the higher margin gasoline.

Another basis by which estimates may be evaluated is economic plausibility. Theory predicts α_1 and α_{21} should be positive, by presumption of positive feedback cycles in ethanol adoption. The raw patterns of correlation in the data strongly support this prediction: the descriptive regressions in Table 4 as well as the uncorrected OLS and first difference estimates of the structural model (Tables 9 and 10 below) all predict a positive effect of N on flex-fuel vehicle demand. Thus, negative estimates of α with high significance are very suspect. To assess a reasonable upper bound for α_1, I consider its predicted impact on flex-fuel sales at varying levels of N. I compute the predicted

[29] Available from the Census zip code business patterns (ZBP) database: http://www.census.gov/epcd/www/zbp_base.html

percentage increase in consumer flex-fuel sales (Q_1) as a function of N relative to $N = 0$ at various levels of α_1, assuming a baseline utility equal to the sample log-odds ratio:

$N \setminus \alpha_1$	0.05	0.1	0.2	0.5	1
1	5%	10%	22%	64%	170%
2	10%	22%	49%	170%	619%
3	16%	35%	82%	341%	1754%
4	22%	49%	121%	619%	4325%

Table 8: Predicted % increase in consumer flex-fuel sales relative to $N = 0$ at sample mean utilities

Casual inspection of the above table suggests that a reasonable bound is $\alpha_1 < 0.5$, since vehicle sales increases in excess of 50% due to entry of an ethanol monopolist seem implausible based on the observed patterns in the data.

As a comparative baseline, I include OLS regressions and the system GMM estimates with the IV regressions reported Tables 9 and 10, which respectively contain the consumer and fleet demand models. The first three "OLS" columns show regressions of the flex-fuel log-odds ratios, H_1 and H_2, with different sets of controls. The results demonstrate that in the consumer equation, estimates of α_1 (the coefficient on N) are generally robust across specifications, whereas in the fleet equation, controlling for unobserved market heterogeneity is extremely important for estimates of α_{21} (the coefficient on N). For both demand models, the OLS estimates of the α coefficients seem plausible from an economic perspective. Also, as expected, the effect of private ethanol fueling facilities in the fleet demand system becomes insignificant after controlling for market unobservables, since most of the variation in $\widetilde{N}$ is absorbed by the fixed effects. For this reason, obtaining strong instruments for $\widetilde{N}$ is not a great concern for consistency of the fleet demand estimates.

Inspection of Tables 9 and 10 reveals the shortcomings of the "profit shifter" instruments for identification: all estimates of α_1 and α_{21} are either statistically insignificant or economically implausible. A potential solution to the identification problem is to use lagged values of N as instruments for N. The required assumption is that current flex-fuel vehicle demand does not depend on *past* levels of E85 availability. This assumption is a priori reasonable and consistent with the other assumptions of the model. The "system" GMM estimator of Blundell and Bond (1998), reported in the final column of the tables, can use two period or greater lags of N as instruments for N. To reduce the total instrument count, which can result in overfitting the endogenous regressors (see Windmeijer (2005)), I use two and three period lagged values of N as instruments in the reported regressions. As seen in Tables 9 and 10, the estimator delivers highly significant estimates for α_1 and α_{21}. Comparing with the OLS estimates (most appropriately, the third column that includes market and time fixed effects), the correction for endogeneity is the anticipated direction, although the "corrected" estimates are statistically consistent with the uninstrumented counterparts (within 95% confidence intervals). I interpret this result as evidence that the controls employed in the model are sufficiently rich to render the residuals purely idiosyncratic. The null hypothesis of

	OLS	OLS	OLS	IV MS	IV FC	IV VP	IV SGMM
e85_retail (N)	0.148***	0.149***	0.129***	-9.632***	-0.301	0.906	0.114***
	(0.018)	(0.015)	(0.025)	(2.113)	(0.306)	(0.565)	(0.029)
gas_stations		0.021***	0.007**	0.005	0.007**	0.007**	-0.000
		(0.001)	(0.002)	(0.009)	(0.002)	(0.002)	(0.001)
ethanol_plants		0.020	-0.041	0.331	-0.025		-0.004
		(0.044)	(0.058)	(0.337)	(0.062)		(0.045)
auto_dealers		-0.011***	0.000	-0.036*	-0.002	0.003	0.012***
		(0.002)	(0.005)	(0.016)	(0.005)	(0.005)	(0.002)
corn_acres		-0.007***					-0.005***
		(0.001)					(0.001)
interstates		-0.024**					-0.016
		(0.008)					(0.012)
pop_growth		1.067***					0.813***
		(0.029)					(0.050)
rural		0.741***					0.398***
		(0.015)					(0.025)
income		0.006***					0.009***
		(0.000)					(0.000)
age		0.019***					0.027***
		(0.001)					(0.002)
male		-0.001					0.010*
		(0.002)					(0.004)
travel_time		-0.011***					-0.004**
		(0.001)					(0.002)
Zip fixed effects	No	No	Yes	Yes	Yes	Yes	Yes
Period fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	40716	40716	33930	33930	33930	33930	40716
RMSE	0.90	0.80	0.71	1.47	0.71	0.72	0.68
Weak id F				11.02	23.48	5.94	48.59
Over id χ^2 (dof)				0.01 (1)	39.08 (2)	8.18 (1)	5.72 (2)
Over id p-value				0.92	0.00	0.00	0.06
AR(1) Z	77.174	76.996	4.198	2.730	4.325	3.911	1.685
AR(1) p-value	0.000	0.000	0.000	0.006	0.000	0.000	0.092

Table 9: IV comparison: consumer flex-fuel vehicle demand

no serial autocorrelation in the residuals is not rejected at the 5% level in both the consumer and fleet equations, indicating use of the estimator is valid. The test overidentifying restrictions (H_0 : exclusion restrictions are valid) is also not rejected at the 5% level, suggesting that the lagged values of N are properly excluded from the flex-fuel demand equation. I also perform the equivalent of a weak instruments test by testing the joint significance of lagged values of N in a regression of demand model controls on N. This is analogous to the Stock and Yogo (2002) test of excluded instruments reported in the two stage least squares regressions. The large (>>10) value of this F statistic provides further evidence that the instruments have identifying power. In sum, the system GMM estimation clearly has the most desirable properties among the available IV estimators, and I therefore use it to estimate flex-fuel vehicle demand in the main estimation routine.

	OLS	OLS	OLS	IV MS	IV FC	IV VP	IV GMM
e85_retail (N)	-0.231***	0.233***	0.236***	1.300	2.852***	1.352	0.229**
	(0.025)	(0.028)	(0.035)	(0.789)	(0.584)	(0.846)	(0.077)
e85_private ($\widetilde{N}$)	-0.279*	0.586***	-0.095	-0.163	-0.225	-0.165	0.402*
	(0.135)	(0.131)	(0.102)	(0.123)	(0.156)	(0.124)	(0.157)
gas_stations		-0.035***	-0.002	0.001	0.001	0.001	-0.033***
		(0.002)	(0.005)	(0.005)	(0.005)	(0.005)	(0.003)
total_establishments		-0.067	0.120	0.074	0.015	0.072	-0.021
		(0.062)	(0.090)	(0.106)	(0.133)	(0.106)	(0.079)
ethanol_plants		0.022***	0.011	0.016	0.022	0.016	0.021***
		(0.003)	(0.010)	(0.011)	(0.011)	(0.011)	(0.006)
auto_dealers		-0.056***					-0.056***
		(0.001)					(0.003)
auto_rentals		-0.079***					-0.081***
		(0.011)					(0.021)
avg_salary		1.788***					1.799***
		(0.019)					(0.039)
corn_acres		0.105***	0.032*	0.032*	0.025	0.032*	0.093***
		(0.011)	(0.014)	(0.015)	(0.016)	(0.015)	(0.017)
interstates		-0.000***	-0.001*	-0.001*	-0.001*	-0.001*	-0.000***
		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
rural		-0.005***	-0.183***	-0.175***	-0.179***	-0.175***	-0.005**
		(0.001)	(0.027)	(0.028)	(0.026)	(0.027)	(0.002)
Zip fixed effects	No	No	Yes	Yes	Yes	Yes	Yes
Period fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Observations	40716	40716	33930	33930	33930	33930	40716
RMSE	1.62	1.13	0.86	0.88	0.94	0.88	1.20
Weak id F				12.68	24.41	5.20	48.59
Over id χ^2 (dof)				14.66 (1)	0.97 (2)	11.16 (1)	4.14 (2)
Over id p-value				0.00	0.61	0.00	0.13
AR(1) Z	100.994	78.258	-2.065	-1.266	-0.805	-1.249	-1.697
AR(1) p-value	0.000	0.000	0.039	0.206	0.421	0.212	0.090

Table 10: IV comparison: fleet flex-fuel vehicle demand

D Robustness results

In this section, I explore the robustness of my results to three key assumptions: the spatial independence of markets, the competitive conduct of firms and the definition of the fleet market size. The assumption of Cournot competition only has implications for the entry model, while spatial independence affects both vehicle demand and ethanol supply estimates. I report these robustness checks in Tables 13, 14, and 15.

Since the main estimation routine found no evidence of correlated shocks to E85 market entry and flex-fuel vehicle demand, in Section D.1, I develop an alternative estimator which is computationally more attractive for testing a wide variety of assumptions. Except as where otherwise noted, I employ this estimator in all the results reported below.

D.1 Restricted maximum likelihood estimator

I develop an estimator of the E85 entry model under the restriction that $\rho_{13} = \rho_{23} = 0$. In this case, the only source of endogeneity comes from the dependence of Q_1 and Q_2 on N. Since the only stochastic dependence here comes through the shock to fixed costs ϵ_3, the likelihood of N once again takes the form of an ordered dependent variable, as in Bresnahan and Reiss (1991). Assuming that $\epsilon_3 \sim N(0,1)$, the likelihood of an observation of N can then be expressed as:

$$L_{mt}(\theta_3|\widehat{\theta}_{12}, B_{mt}) = \Pr[N_{mt}|\widehat{\theta}_{12}, B_{mt}] = \Phi\left(\overline{\Pi}_{mt}^{N}\right) - \Phi\left(\overline{\Pi}_{mt}^{N+1}\right) \tag{26}$$

where $\Phi(\cdot)$ is the standard normal cumulative density function. As in the unrestricted model, estimation is conditional upon the first step estimates ($\widehat{\theta}_{12}$) and the (unobserved) prior period installed base. Also as before, the bootstrap procedure will correct for measurement error associated with using $\widehat{\theta}_{12}$ rather than the true θ_{12}. Of course, no draws of unobservables are required for this model, but it is still necessary to integrate the uncertainty in the installed base of flex-fuel vehicles out of the likelihood function. This step is accomplished by the Monte Carlo integration:

$$\begin{aligned} L_{mt}(\theta_3|\widehat{\theta}_{12}) &= E_B\left[L_{mt}(\theta_3|\widehat{\theta}_{12}, B_{mt})\right] \\ &= \frac{1}{NS}\sum_{s=1}^{NS} L_{mt}(\theta_3|\widehat{\theta}_{12}, B_{mt}^{s}) \end{aligned} \tag{27}$$

Once again, to ensure that a unique equilibrium exists for all observations, I impose the constraint that firm profits decline in the number of firms (N), through the use of a penalty function. Thus, the estimator of θ_3 solves the constrained maximization:

$$\widehat{\theta}_3 = \arg\max_{\theta_3} \sum_{m,t} \ln L_{mt}(\theta_3 | \widehat{\theta}_{12}) \tag{28}$$
$$subject\ to: \frac{\partial \Pi^N_{mt}}{\partial N} \leq 0 \ \ \forall m, t$$

As Table 11 demonstrates, I find very close agreement in parameter values from this estimator and the unrestricted estimator.

Table 11: Comparison of E85 market entry estimates under i.i.d. and unrestricted shock assumptions

Entry, flex-fuel demand error structure:		iid	unrestricted
Variable profits	*rural*	1.708***	1.662***
		(0.176)	(0.114)
	income	-0.012***	-0.014***
		(0.002)	(0.002)
	gas_stations	-0.061***	-0.049***
		(0.005)	(0.004)
	ethanol_plants	0.353	0.344
		(0.204)	(0.197)
	travel_time	-0.032***	-0.034***
		(0.007)	(0.006)
	corn_acres	0.029***	0.037***
		(0.007)	(0.004)
	land_area	-0.001***	-0.002***
		(0.000)	(0.000)
	constant	-2.613***	-2.338***
		(0.215)	(0.190)
Fixed costs	*rural*	0.588***	0.595***
		(0.085)	(0.093)
	income	0.008**	0.005*
		(0.003)	(0.002)
	gas_stations	-0.027***	-0.039***
		(0.005)	(0.006)
	ethanol_plants	0.443	-0.259
		(0.493)	(0.246)
	State fixed effects	Yes	Yes
	Period fixed effects	Yes	Yes
Market size	*fleet FFV installed base*	0.153**	0.141**
		(0.053)	(0.047)
	interstates	10.411**	12.231**
		(3.835)	(1.912)
Covariance terms	ρ_1		0.033
			(0.070)
	ρ_2		-0.017
			(0.138)

Significance levels: * p<0.05, ** p<0.01, *** p<0.001
Standard errors based on 30 bootstrap replications

D.2 Market indepdence and competitive conduct

To assess the robustness of model estimates to the assumption of market independence, I devise an ad hoc procedure to limit the influence of out of market competitors in E85 retailing. Concerns about the violation of market independence center on the fuel retailer entry model, since the inclusion of market fixed effects largely controls for patterns of spatial dependence in the vehicle demand system. Following the logic behind Bresnahan and Reiss (1990) and Bresnahan and Reiss (1991), I estimate the model using markets which are geographically isolated from one another, in the sense that markets with E85 retailers are separated by some minimum distance criteria. As they study professional industries present in virtually all population centers, Bresnahan and Reiss are able to construct a large estimation sample using fairly stringent criteria for geographic isolation. I do not have that luxury, since markets with ethanol retailers are rare (2.7% of observations) and are geographically clustered. Imposing greater geographic isolation leads to significant loss in the number observations, particularly for highly competitive markets, as may be seen in the table below:

	Nearest competitive ethanol market:		
Retail E85 stations	5 miles	10 miles	20 miles
0	38,530	35,400	30,659
1	823	597	209
2	80	48	18
3	16	12	1
4	4	3	1
Total	39,453	36,060	30,888

Table 12: Sample sizes under restrictions on out-of-market E85 retailer proximity

I estimate the model using samples which satisfy the 5, 10, and 20 mile separation criteria. The convergence properties for the 20 mile separation model are poor – convergence was achieved for less than 50% of the bootstrap replications. I therefore discount any inference from this specification, but it is reported for completeness. Comparing these estimates to the main results, we find that estimates are quite robust to the extent to which markets are isolated, and hence independent. Estimates of the α parameters in the flex-fuel demand equations are highly robust and remain significant with E85 station separations of up to 10 miles. Parameters of the E85 entry model manifest greater variation, but generally remain consistent with 95% confidence intervals. The net effect of the differences in parameter estimates may be seen in the entry threshold estimates. As intuition might suggest, imposing greater geographic isolation lowers the point estimate of the monopoly entry threshold (by ~15%).

The Cournot assumption implies that installed bases should increase at a rate proportional to $(1 + N)^2$. However, casual inspection of cumulative consumer flex-fuel registrations conditional upon N (Table 3) suggests that installed bases may increase at a slower rate. Cartel behavior is an alternative model of competitive conduct that is consistent with a slower rise in the installed base (linear in N). The cartel model also represents a lower bound for the effect of competition on ethanol retailer profits. Referencing Table 15 and comparing the cartel model estimates (column 5) to the analogous Cournot model (column 1), we see that all parameter estimates are statistically equivalent, with the exception of the variable profit intercept. The predicted monopoly threshold is slightly higher (7.4%) in the cartel model.

Minimum E85 station separation (miles)	0	5	10	20
e85_retail (N)	0.114***	0.128***	0.134***	0.200
	(0.018)	(0.040)	(0.046)	(0.124)
gas_stations	-0.006***	-0.007***	-0.006**	-0.006**
	(0.001)	(0.002)	(0.002)	(0.002)
ethanol_plants	0.032	0.022	0.070	0.036
	(0.061)	(0.053)	(0.072)	(0.084)
auto_dealers	0.006**	0.007*	0.008**	0.008*
	(0.002)	(0.003)	(0.003)	(0.003)
corn_acres	-0.008***	-0.009***	-0.009***	-0.011***
	(0.002)	(0.001)	(0.001)	(0.002)
interstates	-0.004	-0.000	-0.004	0.007
	(0.019)	(0.014)	(0.016)	(0.017)
rural	0.140***	0.115**	0.126**	0.116
	(0.038)	(0.038)	(0.038)	(0.063)
income	0.007***	0.006***	0.006***	0.005***
	(0.001)	(0.001)	(0.001)	(0.001)
age	0.006	0.005	0.005	0.005
	(0.003)	(0.003)	(0.003)	(0.004)
male	0.027***	0.025***	0.027**	0.025**
	(0.006)	(0.007)	(0.010)	(0.009)
travel_time	-0.003	-0.003	-0.003	-0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Zip fixed effects	Yes	Yes	Yes	Yes
Period fixed effects	Yes	Yes	Yes	Yes
Observations	41292	39453	36060	30888

Table 13: Market independence robustness - consumer flex-fuel demand

Minimum E85 station separation (miles)	0	5	10	20
e85_retail (N)	0.229***	0.243*	0.239**	0.400*
	(0.077)	(0.095)	(0.085)	(0.200)
gas_stations	-0.034***	-0.038***	-0.037***	-0.038***
	(0.003)	(0.003)	(0.003)	(0.003)
ethanol_plants	-0.017	0.018	0.004	-0.050
	(0.069)	(0.088)	(0.086)	(0.122)
auto_dealers	0.025***	0.028***	0.029***	0.029***
	(0.006)	(0.007)	(0.008)	(0.007)
corn_acres	-0.058***	-0.060***	-0.062***	-0.069***
	(0.003)	(0.004)	(0.003)	(0.003)
interstates	-0.096***	-0.103***	-0.121***	-0.117***
	(0.024)	(0.024)	(0.022)	(0.031)
rural	1.833***	1.797***	1.794***	1.808***
	(0.041)	(0.037)	(0.044)	(0.036)
e85_private ($\widetilde{N}$)	0.236	-0.036	-0.278	0.417
	(0.342)	(0.609)	(0.760)	(0.910)
auto_rentals	0.034	0.019	0.020	0.034
	(0.023)	(0.021)	(0.024)	(0.027)
total_establishments	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
avg_salary	-0.007***	-0.007**	-0.007***	-0.008***
	(0.002)	(0.002)	(0.002)	(0.002)
Zip fixed effects	Yes	Yes	Yes	Yes
Period fixed effects	Yes	Yes	Yes	Yes
Observations	41292	39453	36060	30888

Table 14: Market independence robustness - fleet flex-fuel demand

Competition		Cournot	Cournot	Cournot	Cournot	Cartel
Minimum E85 station separation (miles)		0	5	10	20	0
Variable profits	*rural*	1.662***	1.755***	1.865***	1.858***	1.580***
		(0.114)	(0.149)	(0.148)	(0.331)	(0.117)
	income	-0.014***	-0.015***	-0.019***	-0.020***	-0.013***
		(0.002)	(0.002)	(0.002)	(0.005)	(0.002)
	gas_stations	-0.049***	-0.048***	-0.045***	-0.043***	-0.047***
		(0.004)	(0.010)	(0.008)	(0.010)	(0.005)
	ethanol_plants	0.344	0.379*	0.439*	0.078	0.378*
		(0.197)	(0.156)	(0.179)	(0.370)	(0.155)
	travel_time	-0.034***	-0.034***	-0.042***	-0.030*	-0.030***
		(0.006)	(0.006)	(0.010)	(0.014)	(0.006)
	corn_acres	0.037***	0.037***	0.038***	0.037***	0.038***
		(0.004)	(0.005)	(0.004)	(0.005)	(0.004)
	land_area	-0.002***	-0.002***	-0.002***	-0.002**	-0.002***
		(0.000)	(0.000)	(0.000)	(0.001)	(0.000)
	constant	-2.338***	-2.383***	-2.163***	-2.425***	-3.803***
		(0.190)	(0.202)	(0.211)	(0.337)	(0.170)
Fixed costs	*rural*	0.595***	0.862***	1.024***	1.415***	0.565***
		(0.093)	(0.094)	(0.114)	(0.219)	(0.104)
	income	0.005*	0.003	0.000	-0.007	0.005*
		(0.002)	(0.002)	(0.003)	(0.004)	(0.002)
	gas_stations	-0.039***	-0.029***	-0.022**	-0.030**	-0.036***
		(0.006)	(0.007)	(0.007)	(0.011)	(0.007)
	ethanol_plants	-0.259	-0.227	-0.119	-0.635*	-0.192
		(0.246)	(0.176)	(0.256)	(0.273)	(0.212)
	State fixed effects	Yes	Yes	Yes	Yes	Yes
	Period fixed effects	Yes	Yes	Yes	Yes	Yes
Market size	*fleet FFV installed base*	0.141**	0.212**	0.290**	0.237*	0.140**
		(0.032)	(0.070)	(0.108)	(0.105)	(0.043)
	interstates	12.231**	12.667*	15.479***	23.128	15.415***
		(1.912)	(5.001)	(4.441)	(20.849)	(4.212)

Table 15: Market independence/competition robustness - E85 market entry

D.3 Fleet market size

Table 16 reports the fleet flex-fuel vehicle demand estimates under an alternative definition of market size. Instead of using the number of employees as the measure of market size (the "structural" column), I define the market size as the total number of fleet purchased vehicles in the market. This definition of market size implies the dependent variable H_2 can be interpreted as an "inside" market share – I thus refer to this model as the "reduced form". Overall, the parameter estimates compare favorably, although a few coefficients have statistically significant differences. Importantly, the coefficient on N is robust. As the two market size definitions differ significantly, the degree of similarity in the results is reassuring that the market size definition does not have great influence on the model estimates.

Market size specification	Structural	Reduced form
e85_retail (N)	0.229***	0.319***
	(0.077)	(0.067)
gas_stations	-0.034***	-0.017***
	(0.003)	(0.003)
ethanol_plants	-0.017	0.094*
	(0.069)	(0.048)
auto_dealers	0.025***	-0.002
	(0.006)	(0.006)
corn_acres	-0.058***	-0.036***
	(0.003)	(0.001)
interstates	-0.096***	-0.102***
	(0.024)	(0.014)
rural	1.833***	1.221***
	(0.041)	(0.032)
e85_private ($\widetilde{N}$)	0.236	0.111
	(0.342)	(0.442)
auto_rentals	0.034	-0.059**
	(0.023)	(0.022)
total_establishments	-0.000***	-0.000***
	(0.000)	(0.000)
avg_salary	-0.007***	-0.005***
	(0.002)	(0.001)
Zip fixed effects	Yes	Yes
Period fixed effects	Yes	Yes

Table 16: Market size robustness - fleet flex-fuel vehicle demand