

MARKET AREAS OF CAR DEALERSHIPS*

Paulo Albuquerque[†]

Bart J. Bronnenberg[‡]

September 16, 2008

*We would like to thank Dan Akerberg, Andrew Ainslie, Charles Corbett, Sanjog Misra and Minjae Song for useful comments and suggestions. We also acknowledge comments made by seminar participants at the BCRST (Binghamton, Buffalo, Cornell, Rochester, Syracuse, & Toronto) conference 2007, and Erasmus University in Rotterdam.

[†]Assistant Professor at the Simon Graduate School of Business, University of Rochester, paulo.albuquerque@simon.rochester.edu. Financial support from the Portuguese Foundation for Science and Technology is gratefully acknowledged.

[‡]Professor at CentER, Tilburg University, Tilburg, the Netherlands, bart.bronnenberg@uvt.nl.

MARKET AREAS OF CAR DEALERSHIPS

Abstract

Using transactional data, we estimate a structural model of demand and supply of sport utility vehicles (SUVs). Consumers in our model choose among SUVs at specific dealerships. By expanding the product space to include the location of point of sales, we study the size and shape of market areas - defined as the geographic area where demand for an alternative is highest - of each car model, dealership, and manufacturer. Our model is able to provide profit projections and shifts in market areas from managerial decisions such as the relocation of a dealership or changes in product assortment. We empirically apply our model for the San Diego area using a dataset that contains transaction information about locations of dealers and consumers, manufacturer prices, and retail prices. We find high disutility for travel to retailers, which geographically limits preferences to nearby alternatives. We show that most dealers have their own private 'backyard' of demand that is shared with a small set of other alternatives. As predicted by the literature on spatial competition, the size and shape of the market area is strongly dependent on competitors' location. We find that discounting prices by 10% leads to market area expansion of 5 miles.

Keywords: automobile industry, spatial competition, structural models of demand and supply

1 Introduction

The automobile industry has been the focus of numerous studies in recent years, both in economics and in marketing. This literature has covered a variety of themes, such as the analysis of demand and supply in the auto industry (Berry et al. 1995, Sudhir 2001, Berry et al. 2004), the influence of Internet on prices (e.g., Zettelmeyer et al. 2007; Scott Morton et al. 2001), and the impact of a new car form on consumer demand, for example, mini-vans (Petrin 2002) and SUVs (Luan et al. 2007). These studies provide considerable insights into how car manufacturers compete and on how consumers react to product characteristics and marketing activities. However, central to this research, they tend to disregard the role played by the location of consumers and retailers. In particular, little is known about how dealer location and the geographic distribution of consumers interrelate to shape demand and competition patterns. Indeed, demand variation across geography and how manufacturer and dealer managers can influence its patterns has so far been unstudied in the literature, and our paper aims to fill this gap.

Store location plays an important role in purchase decisions. Industry research has shown that a large percentage of variance in consumer store choice is explained by location (Progressive Grocer, 1995) and managers spend considerable time and effort to increase store attractiveness and traffic (Bell et al. 1998). Thus, we believe that this study is of importance to car manufacturers since a good understanding of the competitive environment and its characterization across geography is lacking in the literature, despite that it has been shown in other realms to serve as input for managerial decisions on pricing (e.g., Ellickson and Misra 2008), store customization (e.g., Hoch et al. 1995), and future store locations (e.g., Duan and Mela 2008).

Our objective is to evaluate and analyze geographic demand of cars at the dealer level, which is done in three ways. First, we determine how consumer demand depends on car characteristics, dealer characteristics, and geographic distance between buyers and sellers. Unique to our paper is that each alternative is defined as a combination of a car, with its product attributes, and a dealer, at a given location. Its utility is therefore informative about the trade-off between preferences for dealer location and car characteristics, including price. This approach allows us to describe market areas for each car model, dealership and manufacturer, i.e., the geographic areas where demand for an alternative is highest. Second, using estimates from the demand model, we measure substitution patterns between car dealerships and identify the set of main substitutes for each car, as judged by the intensity of cross-price elasticities. Third, our structural approach provides guidance on important manufacturer decisions - a temporary reduction in car prices, the relocation of a dealer

and the removal of a car model from the market. For each decision, we estimate new manufacturer profits and the resulting pattern of geographic demand.

To make these inferences, we estimate a random coefficients nested logit model using a unique individual-level dataset with transaction information about dealer and manufacturer prices, car characteristics, and zip code location of sellers and buyers. We augment this transactional data using Census information on consumer demographics, and estimate the demand parameters using simulated maximum likelihood, while taking into account consumer heterogeneity and endogeneity between prices and unobserved car attributes. Our method is easy to implement and uses data that are typically available to car manufacturers.

We empirically demonstrate our approach using the sport utility vehicles category in the San Diego area. Regarding demand, our results show that consumers treat alternatives of the same car type - mini, compact, and fullsize - as close substitutes and even more so if they share the same brand. We argue that the main reason for this is the fact that SUVs within type share very similar features. Thus, our results suggest that a consumer who decides to buy a SUV is likely to search among available alternatives within a car type and much less so across different car types.

When deciding where to buy a car, consumers have a strong negative sensitivity to distance to car dealerships. The majority of demand of a car dealership originates from consumers located in close proximity and we find that the probability of consumers choosing a dealer declines sharply as distance increases. While substitution is strongest among alternatives of the same SUV type, it is significantly higher if distance between dealers is small. In other words, the negative sensitivity to distance causes dealer and car substitution to be limited to short distances. As a result, each dealer has its own local demand ‘backyard’, shared with a small set of nearby alternatives of the same type and brand. As an output of our model, examples of geographic variation of demand across geography for cars, dealers, and manufacturers are provided.

Our analysis shows that the shape of market areas is strongly determined by the location of competitors. For instance, it is not uncommon to find that the highest demand is not at the location of the dealer, but instead, at locations that are furthest from direct substitutes. Although this conclusion is to be expected, the ability to quantify it in a complex scenario of a multitude of dealers and alternatives in the market is nonetheless of extreme use for managers deciding on price and promotional activities.

Additionally, we compute market areas at the manufacturer level by consolidating the market areas of its dealers and we report some interesting patterns in location decisions. For instance,

consistent with theories of spatial competition (e.g., d’Aspremont et al. 1979), we find that Honda and Toyota choose different geographic areas to target that minimize overlap and add an extra level element of differentiation between the two manufacturers. We are also able to measure the impact of temporary price reductions, such as a promotional campaign, on geographic dispersion of demand. On average, we find that a discount of 10% on prices leads to an expansion of 5 miles in the market area radius.

We also evaluate the impact of changes in product assortment and retailer locations. In order to obtain new equilibrium prices for all alternatives when market conditions change, we model the supply side and obtain estimates of manufacturer costs and margins. On average, manufacturer variable costs reach about 50% of the manufacturer price to retailers. Fullsize SUVs are more expensive to produce than compact and mini SUVs and some of the car attributes that explain cost are engine size - \$3,300 per unit of displacement - and transmission type - manual drive costs about \$1,500 less than automatic. Manufacturer margins run from \$9,000 to \$16,000, while retailer margins are a tenth of those values, with an average of \$1,500.

Finally, we quantify changes in profit and market areas from two managerial decisions: the relocation of a dealership and the removal of a car model from the market. We exemplify the first situation by moving a Toyota dealership from the suburbs to a location close to downtown. We find that the profits not including fixed costs are improved by selecting the new location due to higher population density, even though there are other Toyota dealers serving a neighboring area. We estimate that the lower bound for fixed costs preventing Toyota from relocating the dealer to be about \$500,000 per year. A second counterfactual is exemplified by removing the Ford Excursion model from the market, which actually happened in late 2005. Although we find that Ford would lose significant profits from consumer switching to other brands, the majority of Excursion consumers would still buy a SUV, thus not decreasing significantly the size of the category.

The paper continues by presenting the literature background for our paper. The description of the model is included in section 3 and Section 4 provides details about the several data sets used in the paper. The estimation algorithm is presented in section 5 and the results are analyzed in section 6. Section 7 describes managerial applications and section 8 concludes.

2 Background

Our work is related to previous papers on the car industry, spatial competition, and search and travel cost. Berry et al. (1995 and 2004) study the automotive industry in two complementary papers. The authors start by developing a model that analyzes demand and supply of differentiated cars using aggregate-level data (1995) and then expand their methodology to combine micro and macro data (2004). Among other results, they are able to produce demand elasticities of price and other observed attributes and find considerable variability across types of cars and models. Sudhir (2001) provides additional insight on manufacturer competitive behavior, showing that it may differ according to the car type: aggressive competitive behavior for smaller and cheaper car segments, cooperative behavior in the mid-size segment and Bertrand behavior in the full-size segment. Regarding the introduction of new products in the car industry, Petrin (2002) analyzes the impact of the introduction of the mini-van on consumer welfare using census data to create identification moments to be used in his estimation, while Luan et al. (2007) evaluate the evolution of consumer preferences and market structure during the introduction and growth of sales of SUVs. Respectively, these papers show that consumer welfare increases when the minivan was introduced and that the market structure significantly changed with the introduction of the sport utility vehicles. This literature on the car industry provides valuable insights on the interaction between car manufacturers and between car manufacturers and consumers, but assumes that consumers trade off all alternatives based solely on car attributes, irrespective of whether such options are conveniently located close to where a consumer lives.

The location of customers is central in the literature on spatial competition. The increasing interest in modeling spatial competition has led to recent examples covering a number of different categories. For instance, competition between movie theaters and the impact of new movie theaters on consumer welfare is studied by Davis (2001). In similar fashion to our paper, Davis uses census data to identify the location and size of the potential market and quantifies cross-revenue effects between theaters. When multiple product or service quality levels exist, Mazzeo (2002) shows that the decision on quality levels can be made endogenous, based on the presence of competing locations in the hotel industry. Thomadsen (2007) presents an example of spatial competition among retailers in the fast food industry and evaluates ownership structure and the impact of market geography on competition patterns. He uses consumer and firms' first order conditions from static Bertrand competition to estimate parameters in the absence of demand data. These papers use market-level data and lack accurate individual detail on the distance traveled to retailers. In contrast, our

estimates of spatial competition are based on actual location and distance traveled by consumers to purchase a product. Moreover, our research significantly extends these papers by studying the simultaneous impact of distance, brand and product type on demand and retailer competition.

There is a large tradition in information economics (e.g., Stigler 1961) on the effect of costly search and travel on demand and competition. In recent studies of demand for consumer durable goods (e.g., Goeree 2008, Kim et al., 2008), or non-durable goods (e.g., Bruno and Vilcassim 2008), evidence is provided that the consumer takes into account travel and search costs, which reduces the choice probabilities of some alternatives being bought and impacts inferences about competition and substitution. As an example of the recent interest in such issues in the car industry, Bucklin et al. (2008) model the impact of the number of dealers and its relative location on choice, showing that accessibility and concentration of dealers have a positive impact on car choice. We extend these papers by introducing the aspect of inconvenience due to travel between location of residence and point of sales in the demand for automobiles. This allows us to focus on the substitution patterns among options that are a combination of dealer-locations and car features. Such substitution patterns reveal geographical patterns of demand for each car dealership and can be used by the dealerships as an impetus to decisions about local marketing action, and by the manufacturer as input to decisions involving dealership location, and the local effect of changing the product mix. To the best of our knowledge, our paper is the first to study the shape of demand across geography of car dealerships.

3 Model

The two sides of the industry are developed in this section. On the demand side, we develop the choice of purchasing a car at a given dealer as a function of car characteristics and geographic distance between consumer and dealer locations. Using the results from this model, patterns of substitution across geography are obtained and described in the results section. On the supply side, we assume profit optimizing behavior of manufacturers and provide estimates of variable costs and margins. The two parts are used to run counterfactual scenarios in policy simulations and provide guidance to managerial decisions, which are described in the applications section

3.1 Demand Utility Specification

At each zip code z , there exists a number of households I_z who consider to purchase a car. The total number of households in the market for a car is given by $I = \sum_{z=1, \dots, Z} I_z$. Household i , living in zip

code z , chooses either to purchase an SUV, or not to buy from within the SUV category altogether.¹ The households who buy an SUV may choose among j alternatives, each of them characterized by its brand, car model and dealer. In other words, alternative j is a model-make-dealer combination. Each dealer sells multiple models of only one brand. For example, a Ford dealership sells cars such as the Ford Excursion, Expedition, and Escape, but no cars from other brands. Car models are further classified by car types: fullsize, compact, and mini SUVs. Types are defined by the research company that has provided the data.

We define the utility that households draw from each alternative as a function of car and dealer characteristics. We abstract from modeling the timing decision of buying a car, since we do not observe longitudinal data on individual choices. The indirect utility is given by

$$\begin{aligned} U_{ij} &= \alpha_i X_j + \beta_i P_{ij} + \gamma_1 G_{id_j} + \gamma_2 G_{id_j}^2 + \xi_{ij} + e_{ij} \\ &= V_{ij} + e_{ij} \end{aligned} \tag{1}$$

with

$$e_{ij} = v_{im_j} + (1 - \sigma_M)v_{ib_j} + (1 - \sigma_B)(1 - \sigma_M)\varepsilon_{ij} \tag{2}$$

X_j is a vector of car characteristics, such as engine size, transmission type, and number of cylinders, as well as brand specific dummies. We also include in X_j a dummy for the SUV category, which we vary across consumers. P_{ij} represents the price faced by individual i for alternative j .² G_{id_j} is the geographic distance between individual i and dealer d_j , measured as Euclidean distance in miles between the zip code centroid of i and d_j . The impact of distance on utility is modeled as a quadratic function to account for non-linear effects of distance on utility. ξ_{ij} captures the impact of car attributes considered by consumers but unobserved to the researcher, such as sun roof or satellite radio. It is likely that these attributes positively correlate with prices, causing endogeneity bias if not accounted for (Chintagunta et al., 2005). For identification purposes, the observed part of the utility of the outside good is set to zero. α_i , β_i , γ_1 , and γ_2 are parameters to be estimated.

Our model is designed to have observed and unobserved individual heterogeneity. The observed heterogeneity is based on the empirical distribution of demographic characteristics, such as income, which we observe at the zip code level. Unobserved heterogeneity is modeled by assuming a normal

¹The outside option also includes SUV purchases from consumers that travel to dealers that are not included in our analysis.

²More details about individual prices are given in the estimation section.

distribution of coefficients. For example, for price, we have

$$\beta_i = \sum_{\forall k} \beta_k \Upsilon_{ik} + \nu_i \quad (3)$$

$$\nu_i \sim N(0, \sigma_\beta) \quad (4)$$

where Υ_{ik} is an indicator function which takes the value of 1 if individual i is in income group k , 0 otherwise.

We allow for correlation within cars of the same type and within cars of same brand to be higher than across types and across brands, using a nested logit formulation for the components of the unobservable term e_{ij} . The parameter σ_B is a measure of unobserved heterogeneity in brand tastes, while σ_M comprises the heterogeneity of SUV types, with $0 \leq \sigma_B \leq 1$ and $0 \leq \sigma_M \leq 1$. The utility function is derived from a variance components formulation, described in Cardell (1997) and Richards (2007). The distributions of v_{im_j} and v_{ib_j} are assumed to be conjugate of the extreme value distribution, such that $v_{im} + (1 - \sigma_M)v_{ib} + (1 - \sigma_B)(1 - \sigma_M)\varepsilon_{ij}$ is also extreme value distributed (Cardell, 1997).

This formulation can support flexible substitution patterns. Alternatives that share the same type, will be more strongly correlated and be closer substitutes as σ_M approaches 1. Conditional on a car type, the correlation between alternatives sharing the same brand will be higher than alternatives that do not share the same brand as σ_B approaches 1. In other words, substitution will be stronger within type or within brand, than across, as parameters σ_M and σ_B get closer to 1 respectively. The model reduces to the multinomial logit model with consumer heterogeneity if both parameters are equal to 0. We note that the independence of irrelevant alternatives (IIA) property of the logit models is avoided with the inclusion of the type and brand nests, individual distance between household and retailers, and heterogeneity in preferences for the SUV category and price sensitivity.

Our choice of nests is justified by the attribute levels of the vehicles. Differences across cars of different types are much more significant than across cars of different brands, leading us to choose a first level of nests defined by car type and a second level of nests composed of alternatives of same type and brand. Consumers are expected to segment the category in similar fashion, substituting highly among alternatives of the same type (see the data section for more empirical evidence of this structure).

With these assumptions the probability of household i choosing alternative j , a car of type m

and brand b , sold at dealer d , is

$$\Pr_{ij} = \Pr_i(m) \times \Pr_i(b_m|m) \times \Pr_i(j|b_m) \quad (5)$$

where $\Pr_i(m)$ is the marginal probability of choosing the car type m , or purchasing from outside of the category; $\Pr_i(b_m|m)$ is the conditional probability of choosing brand b , conditional on choice of type m ; and finally $\Pr_i(j|b_m)$ is the conditional probability of buying at j - a unique combination of dealer, SUV type and brand - given that brand b in type m are chosen. The conditional and marginal probabilities are obtained through the following expressions:

$$\Pr_i(j|b_m) = \frac{\exp\left(\frac{1}{(1-\sigma_B)(1-\sigma_M)}V_{ij}\right)}{\sum_{\forall j \in b_m} \exp\left(\frac{1}{(1-\sigma_B)(1-\sigma_M)}V_{ij}\right)} \quad (6)$$

$$\Pr_i(b_m|m) = \frac{\exp\left(IV_{ib_m}^{(1-\sigma_B)}\right)}{\sum_{\forall b \in m} \exp\left(IV_{ib}^{(1-\sigma_B)}\right)} \quad (7)$$

$$\Pr_i(m) = \frac{\exp\left(IV_{im}^{(1-\sigma_M)}\right)}{1 + \sum_{\forall m} \exp\left(IV_{im}^{(1-\sigma_M)}\right)} \quad (8)$$

where IV_{ib_m} and IV_{im} are the inclusive values of brand nest b and type m , which are equal to:

$$IV_{ib_m} = \ln \sum_{\forall j \in b_m} \exp\left(\frac{1}{(1-\sigma_B)(1-\sigma_M)}V_{ij}\right) \quad (9)$$

and

$$IV_{im} = \ln \sum_{\forall b \in m} \exp\left(IV_{ib}^{(1-\sigma_B)}\right) \quad (10)$$

3.2 Supply

In the application section of our paper, we evaluate the impact of relocating dealerships and of withdrawing car models from a market. To properly forecast the changes that such policy shifts imply, we seek to evaluate their effects on both demand and supply. In this context, we provide a reasonable approximation of the supply side of the market. Specifically, we assume that manufacturers maximize profits by choosing car prices. The decision variable is the average price (across consumers) that manufacturers charge each dealership, W_j . Prices of cars of similar type and model can change across dealers due to different average levels of options and features of cars sold at each dealer. For example, in our data, average manufacturer prices for Jeep Wrangler vary from \$20,500 to \$23,300.

Manufacturer w sells vehicles to its dealers maximizing the following profit function

$$\pi_w = \sum_{\forall j \in b_w} (W_j - C_j) \cdot D_j - F_{b_w} \quad (11)$$

C_j includes the manufacturer's variable costs, F_{b_w} are the fixed costs associated with brand b_w , and D_j is the demand for alternative j . Assuming that manufacturers maximize the combined profit of his alternatives and play a Bertrand-Nash price game, optimal prices are given by

$$\begin{aligned} W &= - \left(\Theta \otimes \frac{\partial D}{\partial W} \right)^{-1} D + C \\ &= - \left[\Theta \otimes \left(\frac{\partial D}{\partial P} \cdot \frac{\partial P}{\partial W} \right) \right]^{-1} D + C \end{aligned} \quad (12)$$

where W , D , and C are $J \times 1$ vectors of manufacturer prices, demand and manufacturer costs respectively, and Θ is a $J \times J$ matrix, where $\Theta_{jj'} = 1$ if alternatives j and j' are sold by the same manufacturer. Finally, the symbol \otimes is used to represent element-by-element multiplication.

Next we define the dealer's pricing strategy. We assume that dealers charge a fixed markup over manufacturer price for each alternative.³ This assumption is motivated by our focus on manufacturer decision making, e.g., about the location of dealerships, and about the assortment of vehicles they should offer, instead of retailer decision making. This assumption results in the following expression:

$$P_j = W_j + M_j \quad (13)$$

where M_j is the average margin for alternative j . From this pricing strategy, note that $\frac{\partial P_j}{\partial W_j} = 1$. With this condition, and substituting equation 12 in equation 13 we obtain the price

$$P = - \left(\Theta \otimes \frac{\partial D}{\partial P} \right)^{-1} D + C + M \quad (14)$$

Our simplifications on dealer behavior are not restrictive for the purpose of measuring price reactions among manufacturers. Under our assumptions, manufacturer price adjustments still take into account (1) shifts in demand, (2) dealership location and product similarity, and (3) the prices of competitors.

4 Data

We combine several data sets to empirically test our model. Our main data set was obtained from a large automobile research company and it includes details about car transactions occurring in

³Kadyiali et al. (2000) use a similar assumption when modeling market power of manufacturers and retailers.

the San Diego area and suburbs between 2003 and 2006. For our analysis, we use the data for 2005. We have information about the car make and model, as well as car characteristics such as price, engine size, and number of cylinders. A unique feature of our data is that it contains the zip code of dealer and consumer locations. Additionally, we have retail and wholesale prices for each car, as well as any manufacturer rebate given. The data is drawn from a sample of car dealerships in the San Diego area and includes 20% of all transactions. We complemented these data with U.S. Census demographic data on income and population density at the zip code level. Finally, we also collected latitude and longitude data of both retailer and consumer zip codes from the Zipinfo database.⁴

We observe 9,717 transactions of SUVs in and around San Diego in 2005. We limit our analysis to the most important brands of SUVs in the area, removing from our dataset manufacturers with an average market share within the SUV category of less than 2.5%. Dealers with very small number of transactions (share of less than 1.5%) were also not included in our analysis. Finally, we remove from our data the transactions by consumers living in zip codes where the number of purchases is minimal (< 20 cars in the whole year). This leaves us with 6,731 observations to work with, about 70% of total observed transactions. Our data used in estimation includes 18 different dealerships, with 7 unique brands, for a total of $J = 59$ dealer-brand-SUV type unique combinations.

We now give some specific details about the data. Figure 1 shows the number of cars sold and average dealer and manufacturer price, for a sample of alternatives, by car type. The figure shows the presence of significant price differentiation between SUV types, in ascending order of mini, compact, and fullsize. Within each type, note that prices are more homogenous within the same brand than across brands. For example, among mini SUVs, Jeep dealers charge a higher price, followed by Ford and then all others. This observed price differentiation, larger across type, smaller within type, and even smaller within brands was one of the motivations for the chosen nest structure presented in the model section.

An interesting feature of our data is that it includes the location of both consumers and dealers for each transaction, allowing for a better understanding of the spatial distribution of demand and supply. Panels (a) and (b) of Figure 2 display the location of households who purchased an SUV and dealerships respectively, in the area of San Diego, where the size of the circles is proportional to the number of households and to the number of sold cars at the dealership.⁵ The majority of dealerships are located fairly close to the center of San Diego, which is located at the southwest

⁴Available at www.zipinfo.com.

⁵The small number of dealers in the map is due to the co-location of multiple dealerships in the same zipcode.

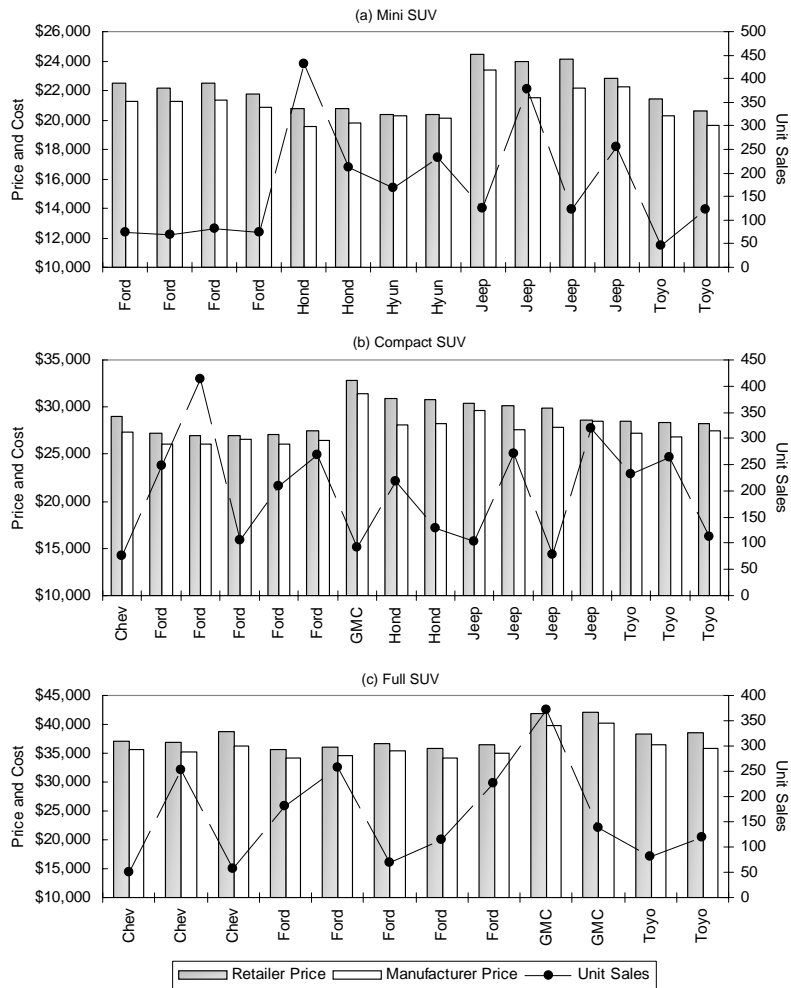


Figure 1: Unit sales, average price and average cost of cars grouped by type of SUV.

corner of the map. This is also the more densely populated area. There is a conglomeration of other dealers about 40 miles north.

Panel (c) of the same figure provides an example of the spatial distribution of Ford dealerships. Ford has three major and two small dealerships in the San Diego area, with two of the large dealerships located just a few miles North of downtown. For one of these dealers, panel (d) shows the geographic origin of its consumers. Since multiple consumers travel from each zip code to the dealers' location, we use different line styles to represent their numbers: a dotted line represents small number of consumers, while a full line is used for large number of consumers traveling from a certain zip code. 87% percent of consumers that bought a car at this dealer are located in a band of less than 20 miles, while 35% percent travel less than 10 miles to buy an SUV. Panels (e) and (f) show a similar example for the Toyota brand.

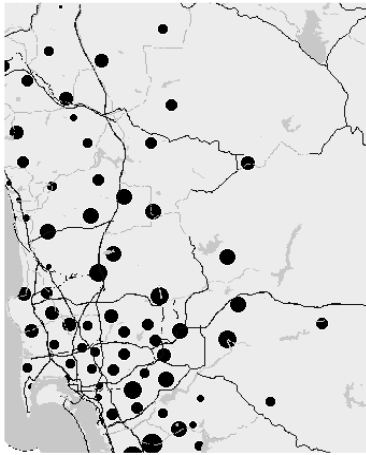
More generally, Figure 3 provides an histogram of the traveled distance between car buyers and sellers. Across all dealers included in our analysis, consumers travel an average of 11.5 miles to buy an SUV, while the median travel distance is 7.8 miles. Only a small percentage of consumers -12% - travel more than 20 miles, while about 50% of consumers purchase a car at a dealer located less the 5 miles from their location of residence.

5 Estimation

5.1 Data Augmentation

We address three aspects regarding the data before estimating the proposed model. First, the transactional data set directly provides information about the car that was purchased by households, but not about alternatives that were available but not purchased. We use the cross-sectional aspect of the data to indirectly impute this information to consumers. Second, we use Census data to estimate the number of consumers that stayed out of the category, and to represent individual consumer demographic characteristics. Finally, the transaction data does not list all car attributes. We use observed variation in manufacturer prices of cars to approximate these unobserved quantities. Once these steps are done, our individual level model can be estimated through simulated maximum likelihood.

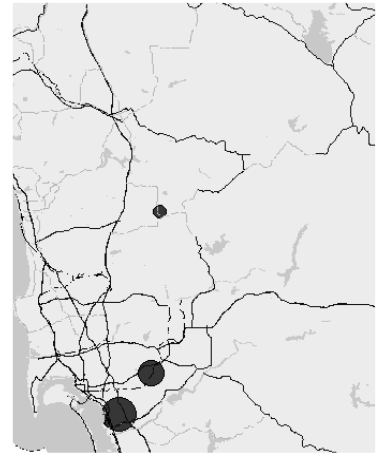
Individual Prices Transactional data sets commonly include information about the price paid by the consumer for the chosen alternative but not about prices that the same consumer would have been charged for alternatives not purchased. With reasonable assumptions about consumer knowledge of prices and car details, our objective here is to augment the data on



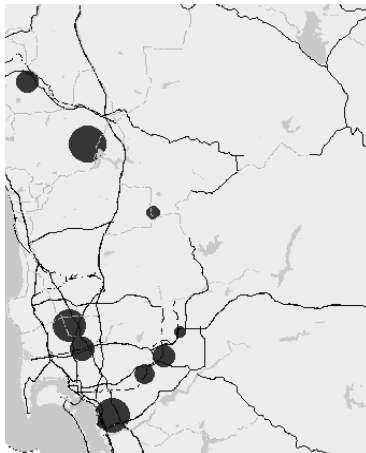
(a) Location of households



(c) Location of Ford dealers



(e) Location of Toyota dealers



(b) Location of SUV dealers



(d) Consumer travel to a Ford dealer



(f) Consumer travel to a Toyota dealer

0 9 miles

Figure 2: Spatial distribution of demand and supply for the SUV category and consumer travel to two example dealers.

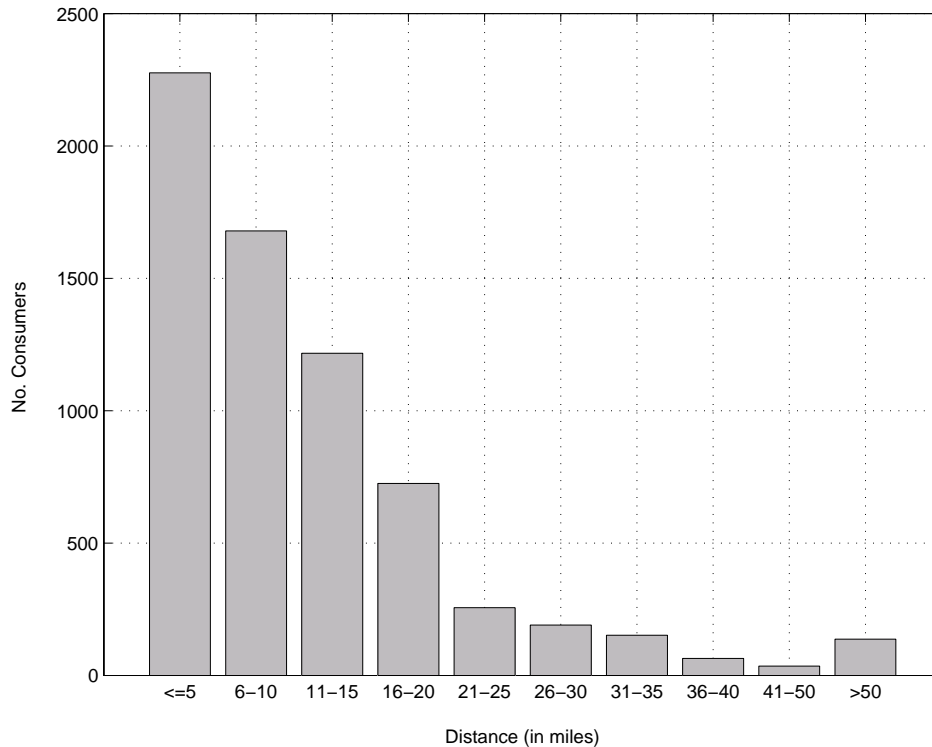


Figure 3: Histogram of distance between consumer and dealer locations.

prices of non-purchased alternatives for each individual.

We assume that consumers are aware of the average level of prices at each dealership, and that they know their own negotiation abilities, relative to the rest of the population. Accordingly, we propose to use as prices of non-purchased alternatives the average price of the cars, corrected for individual ability of negotiation. These prices are obtained in the following manner. For individual i , who purchased alternative j (a car of brand b , type m at dealer d), we identify all transactions of cars of similar make and model. For these transactions, we collect the retailer margins, obtain their empirical distribution, and find the percentile of consumer i in this distribution, based on the actual transaction margin. This gives us a measure of his negotiation ability relative to other consumers.⁶ For each non-purchased alternative, we construct similar empirical distribution of margins and in turn choose the margin value that corresponds to the observed individual negotiation percentile. We then create the final price:⁷

⁶Our assumption here is that larger margins imply lower negotiation ability. We note that the empirical distribution of margins contains cars that are very similar, which makes this assumption credible.

⁷We tested other price mechanisms for augmenting the data, such as just averaging final prices. The chosen procedure produces better fit, in terms of likelihood.

$$P_{ij} = \overline{W}_j + M_{ij} \quad (15)$$

where \overline{W}_j is the average wholesale price charged by the manufacturer to dealer d_j and M_{ij} is the margin obtained by applying the percentile of individual i to the margin distribution in dealer d_j .

Outside Good Consumers Any analysis of spatial competition must take into account the location of potential demand, as consumers have the option of purchasing a non SUV alternative. We use census data to obtain the total population size of each zip code, $\#Households_z$. The potential market for SUVs in each zip code will be a proportion of that value, based on two factors. First, our data cover only a part of all transactions and therefore we limit the potential market to the same percentage of the total number of households.⁸ Additionally, we account for the fact that consumers that have purchased an SUV recently will not be looking for a car and will not be part of the potential market. We use interpurchase time of cars to reflect this aspect on the total market potential (7 years; see Sudhir 2001, for a similar approach). Formally, the total market is given by:

$$I_z = \frac{\#Households_z \times \frac{Observed\ Transactions}{Total\ Transactions}}{Interpurchase\ time} \quad (16)$$

For each zip code z , the sum of "observed" individuals who bought an SUV and "unobserved" individuals whose choice was the outside good will be equal to the total market at that location, I_z . The Census data shows 894,400 households living in the zip codes included in our study, which results in the observed number of households for our sample of $I = \sum_z I_z = 17,888$.⁹ As mentioned in the data section, our final sample includes 6,731 households who buy from the SUV category.

Individuals that chose the outside good must also be assigned prices for SUV alternatives. We assume that consumers that remain outside of the market have the same distribution of prices as the consumers that bought an SUV. For each of these individuals, we take draws from the empirical distributions of cars from consumers who did purchase an SUV and live in the same zip code.

Unobserved Attributes One potential source of endogeneity comes from the fact that dealer prices and unobserved SUV specific characteristics that influence consumer utility, such as car accessories, may be correlated. This variation in unobserved accessories is reflected in variation in average manufacturer price of the cars. One way to avoid the bias created by this correlation is to use a control function approach (Petrin and Train, 2006; Pancras and Sudhir, 2007), exploiting the information that prices contain on unobserved attributes. This approach has two stages. In

⁸The original sample contains about 20% of all transactions. After removing the small dealers and brands, this number of observed transactions is 14%.

⁹ $894,400 (\# \text{ households}) \times 14\% (\text{percentage of observed transactions}) \div 7 (\text{interpurchase time}) = 17,888$

the first stage, we recover ξ'_{ij} , a one-to-one function of ξ_{ij} , by inverting the price equation:

$$P_{ij} = E [P_{ij}|Z_j^D, Z_j^C] + \xi'_{ij} \quad (17)$$

where Z_j^D and Z_j^C are demand and cost shifters. Z_j^D includes observed car attributes, while Z_j^C is the average wholesale price of all cars sold of the same brand and type.¹⁰ Thus, our price equation is given by:

$$P_{ij} = \omega Z_j^D + \psi Z_j^C + \xi'_{ij} \quad (18)$$

In the second stage, $\lambda \xi'_{ij} + \delta \xi'_{ij} P_{ij}$ replaces ξ_{ij} in the utility function, where λ and δ are parameters to be obtained, and the estimation proceeds as if ξ_{ij} was an observed quantity. We include the interaction between unobservables and prices $\xi'_{ij} P_{ij}$ since it is possible that car accessories affect the marginal impact of price on utility.¹¹

5.2 Demand Parameters

We start by estimating the demand parameters without making any assumptions on the behavior of dealers and manufacturers. Then, using the demand parameters, we obtain an estimate $\frac{\partial D_j}{\partial P_j}$ and get the supply parameters using equation 14.¹² The estimation of the demand parameters is done using simulated maximum likelihood, using the following likelihood function:

$$L = \prod_{\forall i} \prod_{\forall j} (\text{Pr}_{ij} | X, P, G, \xi', \theta)^{y_{ij}} \quad (19)$$

where y_{ij} is an indicator variable that takes the value of 1 for the alternative chosen by individual i and zero otherwise and θ is the vector of parameters to be estimated. Individual probabilities Pr_{ij} are obtained from the integration over the empirical distribution of demographics and normal distribution for unobserved heterogeneity, as for instance in Petrin (2002). In our algorithm, we maximize the log likelihood:

$$\log L = \sum_{\forall i} \sum_{\forall j} y_{ij} \cdot \log(\text{Pr}_{ij}) \quad (20)$$

¹⁰In Petrin and Train (2006), wholesaler prices are used in the empirical application regarding the margarine category. In our case, we have information about manufacturer prices.

¹¹By allowing the interaction between price and unobservables, we relax the additive separability between them. For more details on this approach, please see Petrin and Train (2006).

¹²Nevo (1999) uses a similar two-step approach to estimate demand and supply parameters and explain market power in the cereal industry.

5.3 Supply Parameters

As previously mentioned, we are interested in testing managerial decisions that may affect the prices of competing manufacturers. Thus, we need to know all components to obtain new equilibrium prices in the case of a change in the system. Estimation of the demand system provides $-\left(\frac{\partial D_j}{\partial P_j}\right)^{-1} D_j$. We observe is $M_j = P_j - W_j$ in our data, so we are able to find manufacturers costs through¹³

$$C_j = P_j + \left(\frac{\partial D_j}{\partial P_j}\right)^{-1} D_j - M_j \quad (21)$$

Using our estimates for C_j , we can analyze how manufacturer costs vary across j , that is, we can run a regression of C_j on car attributes

$$C_j = \rho_1 L_m + \rho_2 L_b + \rho_3 X_j + \nu_j \quad (22)$$

where L_m and L_b are dummy variables for car type and brand respectively and X_j are car characteristics such as the number of cylinders and engine displacement. The parameter vector $\rho = [\rho_1, \rho_2, \rho_3]$ relating manufacturer costs to car characteristics can be estimated through ordinary least squares.

6 Model Estimates

In this section, we present and discuss the results about the demand and supply parameter estimates, price elasticities, and market areas. The next section describes managerial applications of our model.

6.1 Demand

Table 1 presents the results for the demand parameters and log likelihoods for three alternatives models: (1) the logit model, (2) a nested logit with no control for price endogeneity and (3) a nested logit with a correction for endogeneity of prices. Comparing the log likelihood of the different formulations, we observe that the nested logit fits the data better than the base logit model. Comparing the cases that account and do not account for price endogeneity, we see a slight improvement in the log likelihood. The main difference between models (2) and (3) occurs in terms of price coefficient. As expected, it becomes significantly more negative, when endogeneity between

¹³We tested two alternatives for price setting, one where manufacturers choose prices that maximize combined profits, as described in the model section, and one where they set prices that maximize independently the profit for each alternative. Estimates of margins and costs were more reasonable for the independent maximization and we use this approach to estimate variable costs.

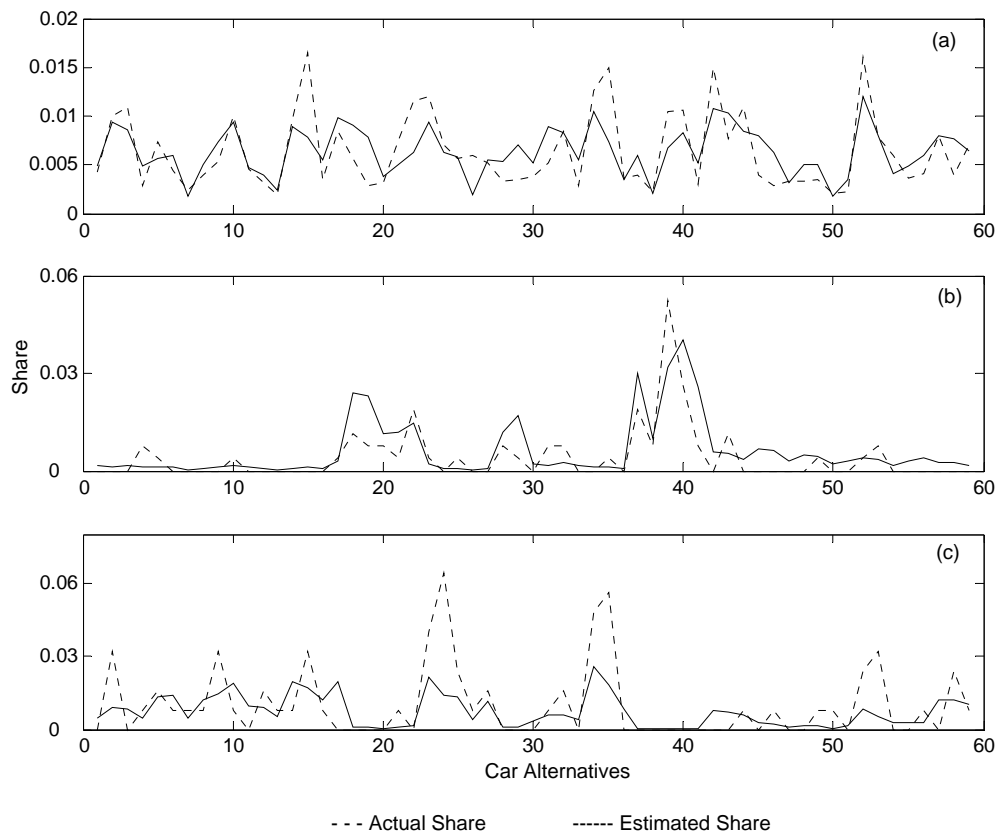


Figure 4: Actual and estimated shares: (a) average shares for each dealer across all zipcodes; (b) shares for zipcode 92008; (c) shares for zipcode 91902.

unobserved attributes and price is accounted for. The remainder of the analysis is done using the best model (3).

Regarding the fit of the model, Figure 4 shows the actual and estimated average market shares of each alternative j (excluding the outside option) for the total San Diego market (panel a) and for two selected zip codes (panel b and c). We find that the model explains the variation in car popularity well, not only at the general market level but also at the zip code level, with a good match between estimated shares and actual shares. The model does equally well for other zip codes.

We now comment on and draw some conclusions from the demand parameters. The price coefficient is negative and significant for all income levels, with the highest income segment slightly less price sensitive. The parameters translate to an average price elasticity of -4.6. We analyze cross-price elasticities in the next subsection. In terms of car attributes,¹⁴ consumers value more

¹⁴We code the variable Transmission as "1" if automatic and "0" otherwise. We code the drive type as "1" if the car has 4-wheel drive and "0" otherwise.

cylinders, automatic transmission and prefer less engine displacement while regarding the car type, consumers draw more utility from fullsize and compact SUVs than mini SUVs.

We include two variables that have the purpose of measuring the popularity of a certain car model and of a certain dealership location: the number of dealers offering alternative j and the number of alternatives at the location of j .¹⁵ We find that consumers prefer alternatives that are available at many dealers. This may be a result of consumers finding appealing to have multiple dealers to negotiate prices with or just a larger set of models and cars to choose from. Including this variable tests the existence of congestion (Akerberg and Rysman, 2005) within a brand nest, where the existence of too many dealers offering the same car may lower its appeal, which we do not find in this case. The number of cars sold at the zip code of j also has a positive effect on utility. This could be a result of consumer preference for variety of alternatives at a particular location. Variety can have strong impact on choice (Hoch et al., 1999) and in this particular case, serves as evidence of positive externalities of collocating dealers.

We also observe that the linear residuals from the control function, which represent attributes unobserved to the researcher but considered by consumers, have a positive impact on choice, i.e., cars with more unobserved accessories are likely to be more appealing to the final consumer. Finally, we find that the nest parameter for car type is large, with a value of 0.75, leading to much stronger substitution between alternatives within car type than across. For the brand nest parameter, the value is 0.16. These estimates strongly suggest that consumers segment the alternatives by car type, with additional segmentation between brands. In the next subsection, we further analyze the impact of these estimates on car substitution.

Finally, the intercept for the SUV category is negative and varies significantly across income levels. Specifically, the estimates show that the income group of \$34,000 to \$54,000 is significantly more likely to buy from this category than consumers outside of this income bracket.

6.2 Market Areas

From our estimation results, we note that distance between dealers and consumers plays an important role in the decision of buying a car and we obtain a highly significant negative parameter - the longer the distance, the lower the utility and choice probability of an alternative. The distance square parameter is positive, revealing that as distances increases, utility still declines, but at a slower pace. Figure 5 exemplifies how distance between consumers and dealers affects choice prob-

¹⁵The number of dealers and location is taken as exogenous in our model, since it is a result of an entry game that happened in the past.

Category	Variable	(1) Logit		(2) Nested Logit I		(3) Nested Logit II	
Intercept	income < 34K	-5.098	0.488	-2.862	0.153	-2.555	0.200
	34k < income <54K	-4.176	0.449	-2.089	0.125	-1.783	0.143
	income > 54K	-5.201	0.475	-2.812	0.145	-2.505	0.193
Car Type (base: Mini)	Compact	0.574	0.105	0.238	0.029	0.321	0.039
	Fullsize	0.960	0.253	-0.102	0.052	0.110	0.082
Car Brand (base: Ford)	Chevrolet	0.739	0.093	0.209	0.030	0.226	0.031
	GMC	1.148	0.106	0.197	0.026	0.268	0.037
	Honda	1.401	0.099	0.349	0.034	0.374	0.036
	Hyundai	1.024	0.090	0.281	0.032	0.267	0.032
	Jeep	0.776	0.053	0.161	0.017	0.189	0.020
	Toyota	0.049	0.065	0.006	0.018	0.029	0.022
Car Attributes	# of Cylinders	0.279	0.041	0.077	0.012	0.080	0.012
	Engine Displac. (in liters)	-0.246	0.063	-0.064	0.018	-0.080	0.016
	Transmission (1=Auto)	0.489	0.059	0.113	0.019	0.120	0.019
	Drive Type (1=4Wheel)	0.366	0.110	0.114	0.025	0.055	0.030
Price (\$10,000)	income <34K	-0.686	0.174	-0.079	0.050	-0.203	0.065
	34K < income <54K	-0.615	0.159	-0.080	0.030	-0.203	0.048
	income > 54K	-0.527	0.171	-0.036	0.040	-0.160	0.069
Location	Distance (in 100 miles)	-10.834	0.206	-3.166	0.251	-3.157	0.228
	Distance Squared	8.246	0.231	1.644	0.251	1.639	0.215
Distribution	# Dealers Selling j 's Model	0.486	0.051	0.158	0.021	0.161	0.022
	# Dealers at Location of j	0.005	0.006	0.004	0.002	0.004	0.001
Control Function	Unobservable Intercept	1.013	0.370			0.173	0.109
	Unobservables Price	-0.185	0.105			0.001	0.031
Nests	Segment Nest Coefficient			0.757	0.018	0.754	0.019
	Brand Nest Coefficient			0.142	0.031	0.159	0.033
	Log Likelihood	-49985		-49385		-49378	

Table 1: Mean and Standard Deviation of Demand Parameters.

abilities, using as example a Jeep dealer selling three different models, Grand Cherokee, Liberty and Wrangler. This figure is created by computing individual choice probabilities given the set of estimated parameters and averaging these probabilities across consumers located at rolling windows of three miles from the dealership, e.g., consumers located 0 to 3 miles, 1 to 4 miles and so on. For example, individuals located at distances between 0 to 3 miles of the Jeep dealer show on average a choice probability of about 3% of buying the Grand Cherokee, slightly less than 2% of buying the Liberty and about 1% of buying the Wrangler.

It is clear that consumers more likely to buy from this dealer are located closer, with a significant drop in choice probabilities occurring as distance increases. We note also that this decrease with distance is not smooth due to other factors that influence utility, namely the location of competitors. For example, a steep drop in choice probabilities may be due to the fact that a competitor dealer is located at that distance from the Jeep dealership, for example at distances of 13-14 miles and at 23 miles. In contrast, small peaks such as the one observed between miles 11 and 12 are caused by the non-existence of dealers located at that distance, which leads those consumers to be more likely to consider further away dealers.

Another way to look at the impact of distance on demand is through the use of geographic visualizations of market areas for each car model, dealership and manufacturer. We define market area as the geographic area where demand for an alternative is highest. We start by looking at an example of one car model, sold at two different dealers. Panels (a) and (b) of Figure 6 show the average choice probabilities for Ford Expedition, as a percentage of all fullsize SUV alternatives, at two Ford dealerships designated by A and B.¹⁶ To construct the market areas, we average the choice probabilities of individuals located at each zip code for the alternative being studied and linearly interpolate for the remaining areas in the map. The dark circles represent the dealers' locations. Other retailers are not shown for clarity of exposition.

As expected, we observe larger choice probabilities in areas surrounding the dealer locations, with the Expedition having an estimated share of about 25% within fullsize SUVs for zip codes located 5 miles or less from the dealers' locations. However, the presence and location of the other dealer has a major impact on demand. In fact, average choice probabilities of consumers buying from dealer B are highest not at the zip code of the dealership, but to the right of its location, further away from his strongest competitor, dealer A.

Borders of markets area can be defined once a lower limit of choice probabilities is chosen. For

¹⁶In this analysis, we condition on fullsize SUV alternatives because we find a high correlation within alternatives of the same type and low correlation otherwise, as presented in the next subsection.

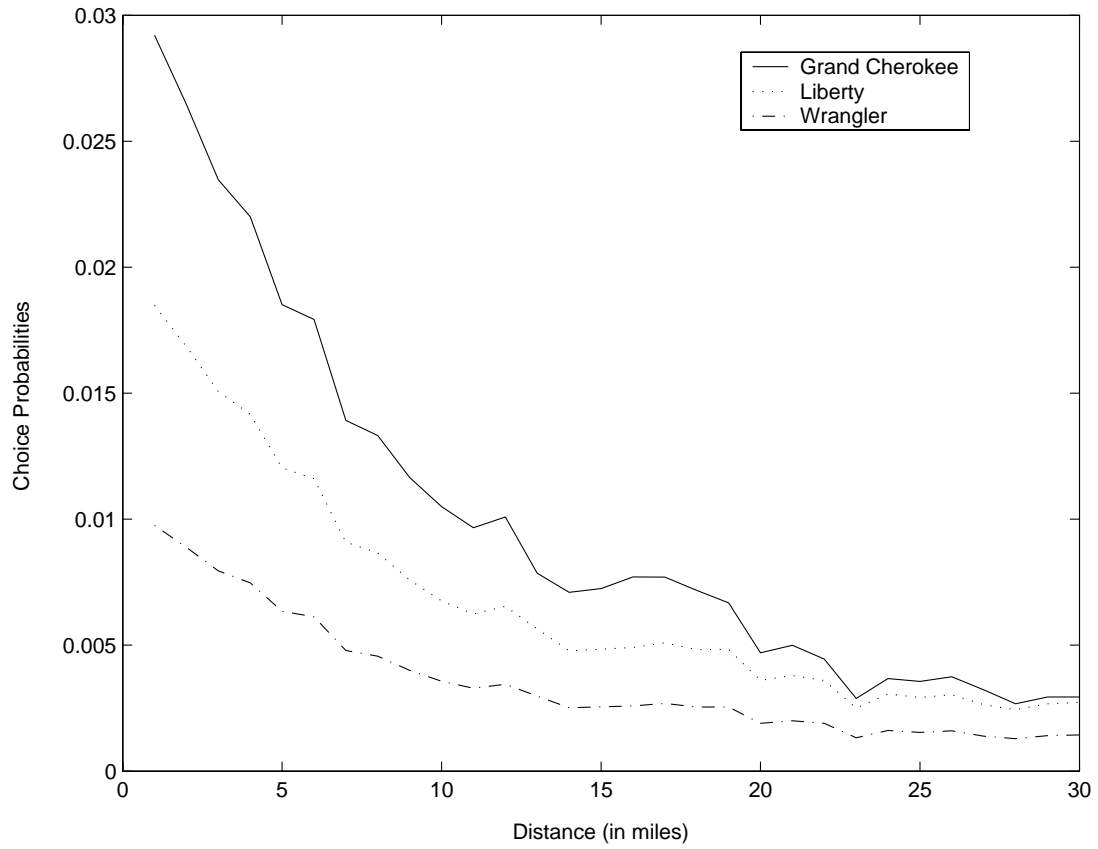


Figure 5: Estimated choice probabilities of consumers located at different distances intervals from a Toyota dealer.

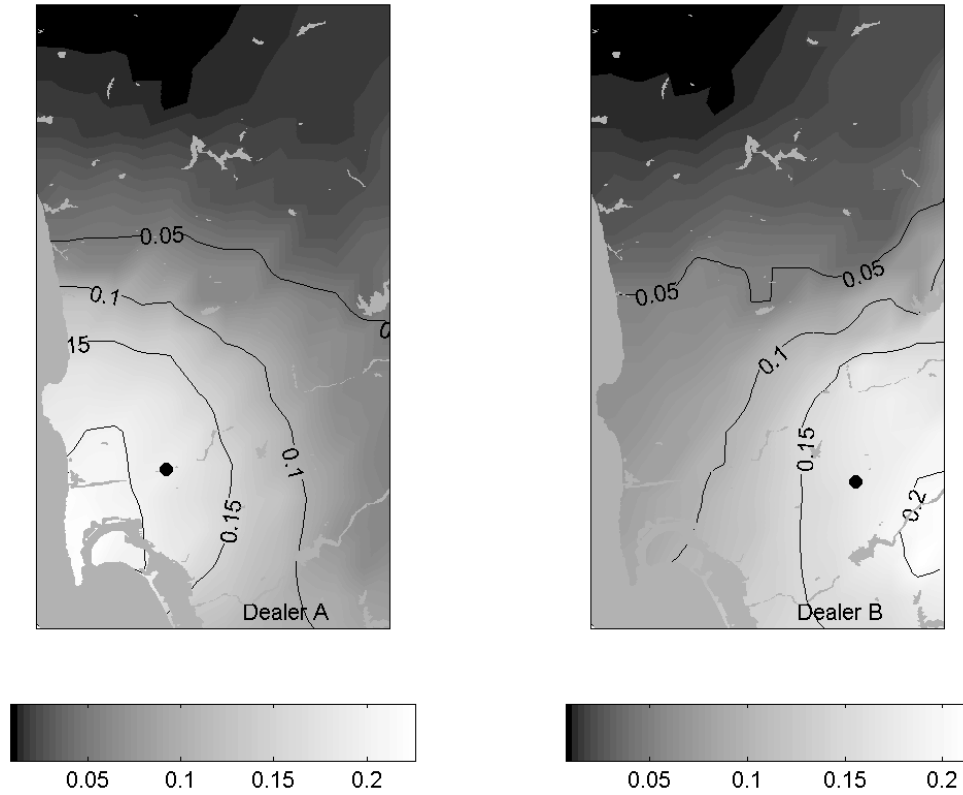


Figure 6: Market areas for Ford Expedition sold at two different dealers, designated by A and B. The choice probabilities are computed within fullsize SUVs.

example, the market area of Excursion A where choice probabilities are above 15% among fullsize SUVs has a size of approximately 100 square miles, with borders given by the isoline labeled 0.15. Choice probabilities above 10% are observed in an area of about 300 square miles.

We also provide examples of market maps for car manufacturers. To do this, we sum the choice probabilities for all alternatives of a manufacturer. Figure 7 shows the market areas of Honda and Toyota, with the circles representing dealer locations. Honda has two dealers, located in almost the same latitude, but one much closer to the coast than the other one. Toyota, on the other hand, has two dealers located closer to downtown and a third located about 20 miles north. The market areas of the two Japanese manufacturers display a very interesting pattern. While the demand for Honda is concentrated on a horizontal band, of about 10 to 15 miles north-south width, Toyota has two discontinuous patches of demand, one close to downtown and the other inland, in the area of Escondido.

In the context of these findings, we note that theoretical models of spatial competition predict spatial differentiation when manufacturers are deciding on locations for outlets. For instance,

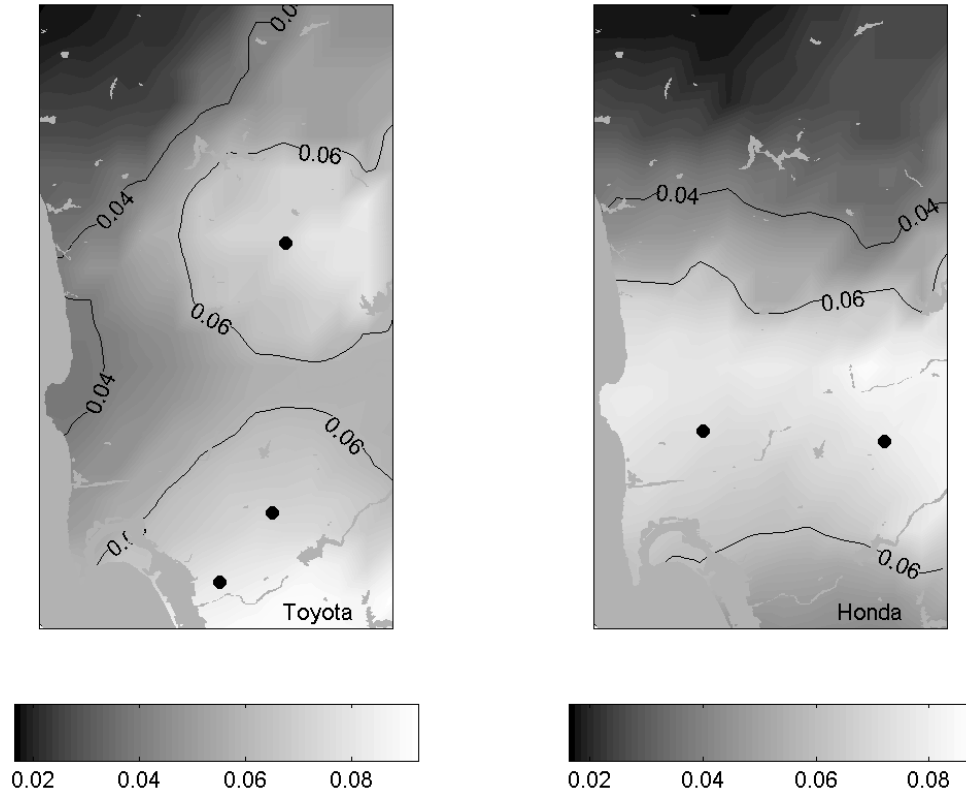


Figure 7: Market areas for Toyota and Honda in San Diego and suburbs.

d'Aspremont et al. (1979) show that in the case of quadratic cost of distance, manufacturers will choose the locations furthest apart from each other, in order to minimize the competition effects on price. Additionally, in the case of product choice involving multiple characteristics, Irmen and Thisse (1998) show that manufacturers choose one dimension to maximally differentiate while minimizing differentiation on other characteristics. Given our results, it seems that location serves as the differentiation dimension, since, within a car type, attributes of cars of different manufacturers are strikingly similar. The patterns observed in Figure 7 empirically confirm this theoretical prediction about location choice.

6.3 Substitution Patterns

In this section, we describe the general results for cross-price elasticities, which are commonly used as a measure of substitution between alternatives. Over all alternatives, the cross-price elasticities average a low value of 0.03, because a large number of values, especially across car types, are close to zero. If we restrict our analysis to cross-elasticities between cars of the same type, the average

cross-elasticity reaches 0.14 among fullsize SUVs, and 0.08 among mini SUVs and among compact SUVs. Finally, SUVs of the same type and brand have on average cross-elasticity of 0.16. Although these values already give some idea about the higher substitution occurring between cars that share type and brand, an example that also takes into account distance between dealers is presented next.

Figure 8 provides the illustrative case of a Ford dealership, which sells four different SUV models: the Escape (classified as a mini SUV in our data), the Expedition (fullsize), the Explorer and the Explorer Sport Track (both compact SUVs). In the three panels of the figure, the cross-price elasticities between each of the first three models and other cars are displayed. The selected Ford dealership is placed at the origin of the X-axis, with other car dealers located at the actual geographic distance from this dealer.

In each panel of the figure, alternatives with cross-price elasticity above 0.03 are presented. For all models, the most important substitutes are cars of the same type, with alternatives of other types having cross-price elasticities significantly lower and thus not displayed. For example, alternatives included as substitutes of Escape (mini SUV) in the top panel of Figure 8 are all mini SUVs, with other alternatives having much lower cross-price elasticities, no matter their geographic location. This is driven by the fact that we find a value for the car type nest parameter that is close to 1, which leads to high correlation within car type. We infer that consumers believe that cars of the same type are closer substitutes and it seems to be the first form of segmentation in this category. Possible reasons for this finding are that alternatives within types have very similar physical attributes, such as size, horse power and fuel efficiency, besides similar prices (see Figure 1 in the data section).

Within each SUV type, two forces impact the strength of competition: distance and brand name. It is clear that the shorter the distance, the higher the cross-price elasticities. Also, we observe that co-incidence in terms of brand name leads to an increase in product substitutability. For the three Ford cars, changes in prices at other Ford dealers have stronger impact on demand than changes in prices of other brands. For example, a Ford Escape sold at dealer 9 miles away is perceived as a stronger substitute than alternatives such as the Jeep Wrangler or Liberty sold at a dealer 3 miles away. We conclude that although distance plays an important role in decreasing substitutability between alternatives, its impact is reduced if cars share the same brand. From the highly concentrated patterns of demand surrounding each dealer's location presented earlier and the limited substitutability across car types and distance, we conclude that each dealer has its own private 'backyard' of demand, shared with a small set of competitors, where most of its sales are

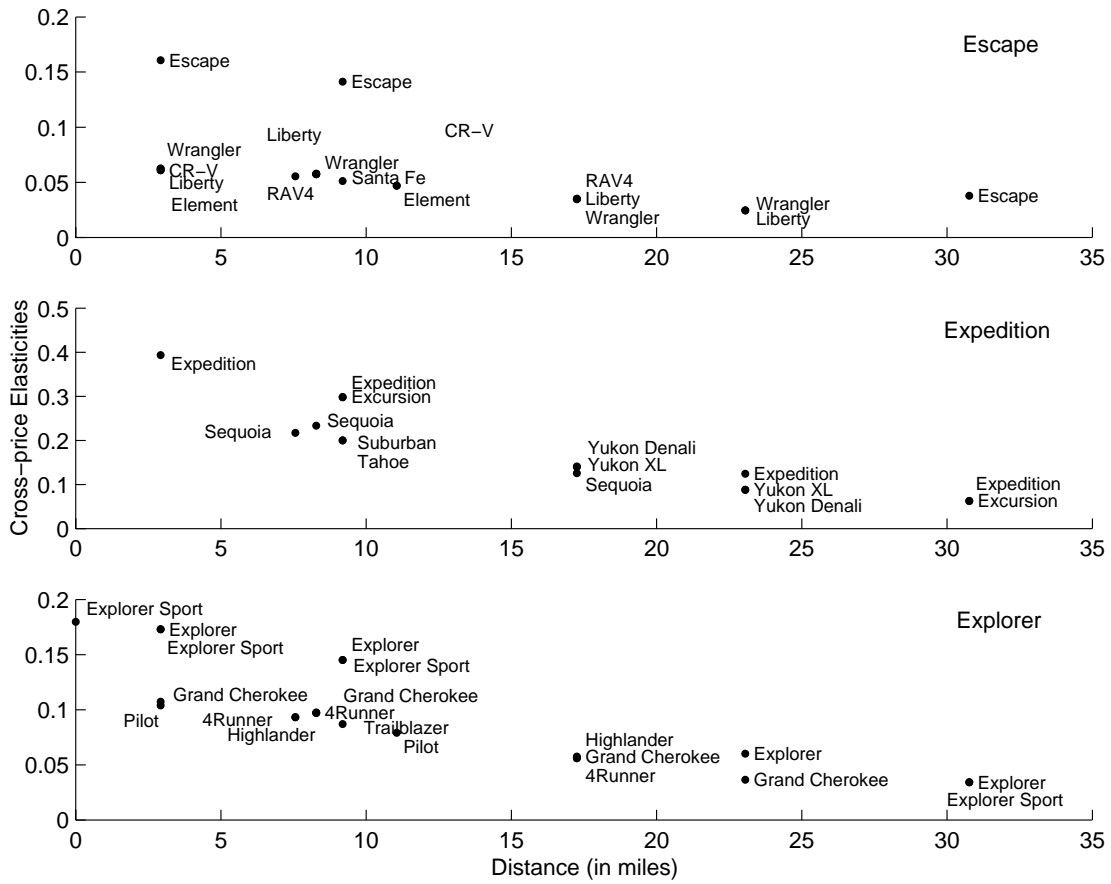


Figure 8: Cross-price elasticity between car dealerships, for a Ford dealer, for the three SUV-types sold.

Manufacturer Variable Costs	Mean	St. Dev.
Intercept	6458.1	4802.4
Compact	-4249.4	1115.6
Mini	-5179.3	1816.7
Engine Displacement	3332.6	698.1
No. Cylinders	-223.0	521.7
Transmission	1546.1	803.6
Drive	-5328.1	1216.9

Table 2: Mean and standard deviation of regression parameters of manufacturer cost.

likely to originate. Our model is able to quantify the size and shape of these market areas (see e.g., Figure 6).

6.4 Supply

Using Equation 21, we find that manufacturer costs for SUVs range from about \$11,500 to \$20,000. Table 2 presents the estimates for the supply parameters explaining the variable cost of production. From the regression of manufacturer costs on attributes, we find, as expected, that compact and mini SUVs are cheaper to produce than fullsize. Cars of larger engine size cost more to produce, with \$3,332 per unit of displacement. Cars with automatic transmission cost \$1,546 more to produce while not having 4×4 drive leads to average savings of \$5,328.

Averaging across alternatives of the same make and model, we find that Hyundai Santa Fe has the lowest cost, of about \$7,000, while the large fullsize SUVs, mostly of American origin, cost between \$20,000 and \$27,000 to produce. This leads to margins that range from \$9,000 to \$16,000 for the manufacturer. Dealer margins, which are observed in our data, are about one tenth of that, with an average of about \$1,500.

7 Managerial Applications

We use our model to test the impact of managerial decisions on geographic distribution of demand and additionally profits. We start by looking at the impact on market areas and demand of a reduction of price of 10% by a manufacturer for all its cars. We then test the relocation of a dealership and the removal from the market of a car model.

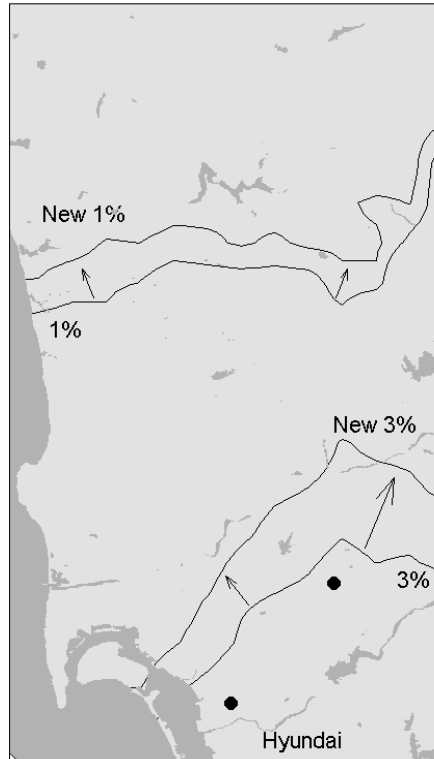


Figure 9: Map displaying the change in choice probabilities across distance from a price reduction of 10%.

7.1 Temporary Price Reduction

It is very common in the car industry to observe promotional campaigns where price is temporarily reduced to induce higher demand. We exemplify the impact on profits and on market areas of such promotional activity by dropping prices in all Hyundai dealers by 10%, which represents a reduction of about \$1,400 in the final price of the car, and recompute choice probabilities. By assuming a temporary and not a permanent price discount, prices at competitors are kept at the original levels. We display the impact on market areas in Figure 9.

The location of the two Hyundai dealers is displayed - the two dark circles - as well as the original and new isolines of average choice probabilities of 3% and 1% for Hyundai. The original lines show that locations where average choice probabilities were 3% and 1% before the discount. As in previous market areas, choice probabilities decline with distance. The drop in prices has the effect of moving those lines outward, thus increasing choice probability for alternatives sold at the Hyundai dealers across the San Diego area. On average, equivalent choice probability lines are shifted outward by about 5 miles as a result of 10% drop in car prices. This finding is of significant

use to dealers and manufacturers. For instance, most campaigns price reductions are combined with other local advertisements, and coordination in terms of market coverage is essential to maximize the impact of such marketing activities.

7.2 Relocation of a dealership

Our model treats the location decisions of dealerships as exogenous to the current consumer and manager decisions. Although this assumption was partially driven by the unavailability of entry and exit in our data, it also allows us to evaluate alternative spatial configurations of dealers. We explore a counterfactual situation that tests the profitability and spatial demand coverage of the relocation of a Toyota dealership.

The steps to obtain the profit in a potential new location are as follows. We start by choosing a new zip code and compute the geographic distance between all consumers and this new location. We replace the actual distances to this dealer with the newly computed distances and obtain estimates for market share using our demand model. We then obtain new equilibrium prices using the supply equations and iterate until convergence. Figure 10 provides details on the new and old locations of the Toyota dealers.

The actual location of the tested dealer is inland at zip code 92064. To run a counterfactual situation, we chose a new location - zip code 92111 - closer to downtown with higher population density in its surroundings, but also closer to the other two Toyota dealers. By relocating the dealer to this location, we quantify the trade-off in profits between concentrating dealers at areas of high demand and differentiation created by geographic distance between dealers.

In terms of market areas, there is an obvious shift of demand creation to areas surrounding downtown, given the concentration of all its three dealers. However, most of the area north of San Diego will now have much less coverage. The manufacturer profits not accounting for fixed costs of this new location would be \$12.1MM compared to \$11.6MM for the old location, a difference of about \$500K a year, if demand would remain at these levels. The main difference comes from the fact that, although the three dealers are now more concentrated - leading to new equilibrium prices for Toyota to be lower by about \$200 per car - they target the most populated area, which leads to an increase in the number of cars sold overall by Toyota from 953 to 1018.

Several reasons may justify Toyota's decision to have the current spatial configuration instead of the situation tested here. For instance, it is likely that in the new location closer to downtown rent would be considerably larger than in the suburbs or there is lack of availability of land or

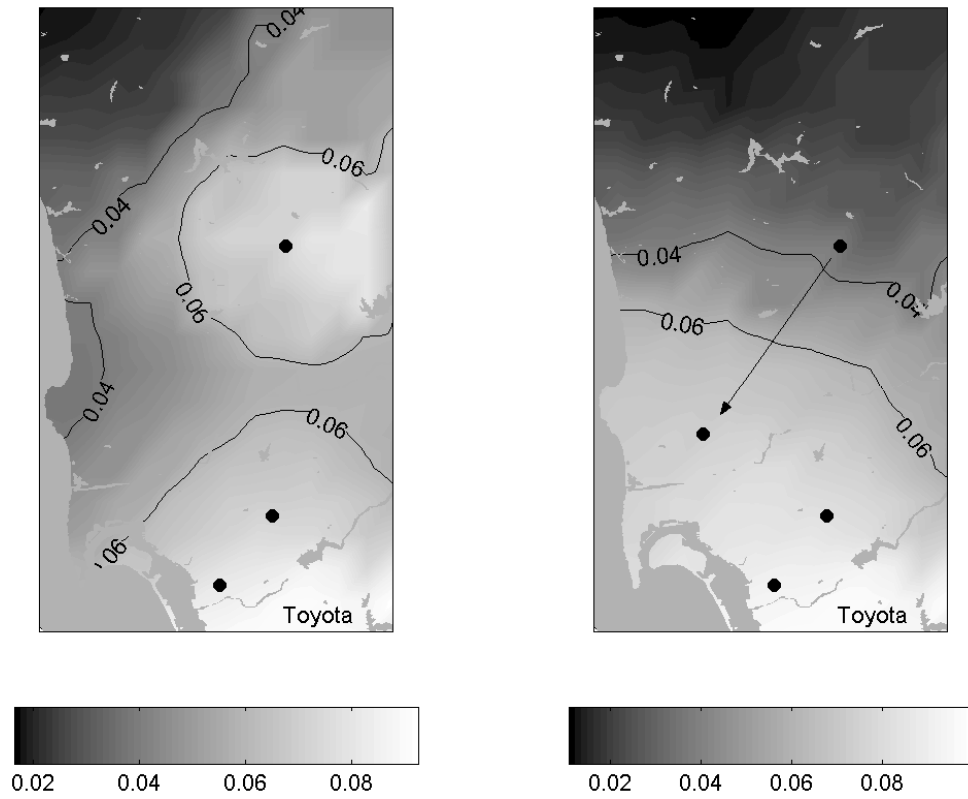


Figure 10: Original and hypothetical market areas for Toyota, showing the impact of relocating a dealership.

infrastructure for a large dealership at that zip code. The initial cost of relocation may also be excessively high and would prevent Toyota from adopting the new configuration if managers take a myopic approach. It could also lead to reactions from competitors, such as further relocations, which would limit the profit growth potential. Finally, it is also possible that this location is an undiscovered business opportunity, created by the developed of the area of San Diego post-decision of dealer locations.

The difference in profits between locations has also been used as an important piece of information in other models. For example, Pakes et al. (2008) use the assumption that firms choose the locations that optimize their profits to estimate fixed costs of firms. In this case, we find that the difference in fixed costs between old and new location should be at least \$500K per year to provide a purely quantitative rationale for Toyota's decision.

7.3 Discontinuing a Car Model

In recent years, demand for large SUVs has consistently declined over time. In fact, American manufacturers have discontinued some models and are at present discussing the possibility of removing several other large size SUVs from the market. We test the impact on demand and profit of such a decision, exemplified in this case by the removal of Ford Excursion from the market. In fact, Ford stopped producing this model in late 2005. We follow the same algorithm as in the previous counterfactual experiment - the change in demand is introduced, then new market shares are computed, new equilibrium prices are obtained based on the supply equations, and reiterations are done until convergence.

The data shows that the Ford Excursion accounted for about 4% of SUV Ford sales, and 1% of the SUV category in San Diego, selling 75 cars in 2005. By discontinuing this car model, we find that Ford sales go down by 50 cars, with two thirds of Ford consumers switching to other Ford models, mostly to the Expedition. The loss of Ford's share benefits mostly other American brands, Chevrolet and General Motors, which offer the closest substitutes to the Excursion - the GM Yukon and the Chevy Tahoe. The other fullsize SUV -the Toyota Sequoia - also gains some market share. Discontinuing the Expedition and considering the 2005 levels of demand, the profit for Ford before fixed costs goes down by \$370K per year.

It is interesting to point out that most of the demand of Ford Excursion would stay in the SUV category, with only 14 cars out of 75 - 19% - being "lost" to the outside good, which in our case, represent other non-SUV cars. We tested the removal of other models from the market and the

small switching to the outside is general to most alternatives.

Although we give the example of the removal of a car model from the market, our approach can be also used to test the introduction of new cars in the market, and predict shares and market areas for each dealer, given new car characteristics.

8 Conclusion and Future Research

This paper analyzes local demand and supply for cars using transactional data. Unique to our paper is that a purchase option is defined as a combination of a car, with its product attributes, and a dealer, at a given location. Utilities for such options therefore are informative about the trade-off between preferences for dealer location and car characteristics, including price. Estimating these utilities using a large transaction-level data set, we show that the effects of physical distance between buyers and sellers are important and can not easily be ignored when studying demand and substitution patterns in the car industry. Specifically, our analysis suggests that substitution even among pairs of the same car, quickly fades as the dealers selling them are located farther away from each other.

Our approach estimates profits for a number of managerial scenarios, taking into consideration the distribution of buyer and seller locations and the presence of neighboring dealerships. As such, our approach also provides empirical insights into spatial competition in a realistic setting.

The first application of our approach is to show the geographical extent of a car dealership's market. We find that each dealership has a localized demand area, and that choice probabilities decrease at a fast rate in distance between buyers and sellers. For most choice alternatives, we find that a limited set of close substitutes combined with concentrated demand leads to private "backyards" of higher demand for each dealer, and that the locus of highest demand intensity for a given dealership can be away from the location of the dealership itself.

Second, our model can be used as input for a number of managerial decisions, i.e., offers descriptions and some normative guidance for manufacturers and dealers about the effects of geographic distribution of competitors and buyers when negotiating prices with the consumers, setting the amount of discounts, or determining geographic areas where to advertise. We are able to measure the expansion of market areas when prices are reduced, and in the specific case of SUVs, we find that a discount of 10% expands demand of a dealer by 5 miles.

Third, we test hypothetical new locations for dealerships and estimate the minimum bound of fixed costs preventing the relocation of dealerships from improving the overall profit of the

manufacturer.

Finally, we quantify the impact of removing one car model from the market and show the reassignment of consumers to other car options. We estimate that the majority of consumers would stay in the SUV category.

Our research can be expanded in a number of ways. First, it is an interesting puzzle that geographic market expansion due to a 10% price reduction is 5 miles, a number that is surprisingly low. This finding implies that consumers do not wish to travel to dealerships farther away even at discounts that can be significant in dollar terms. In turn, this suggests the need to study the economic and behavioral underpinnings that lead consumers to focus on buying locally. Second, it is possible to include information about transactions about other car types besides SUVs, and thus measure substitution patterns across a larger sets of alternatives. This could be useful, given the current consumer switching to smaller cars due to rising gas prices.

Finally, we believe that our approach can be broadly applied to settings outside the car industry. Specifically, our approach can be used when manufacturers are interested in evaluating the effects of the location of outlets on demand and competition, e.g., the banking industry, the gasoline industry, etc., where store location plays an important role in the success of products and services of firms.

References

- [1] Akerberg, D. and M. Rysman (2005), "Unobserved Product Differentiation in discrete-choice models: estimating price elasticities and welfare effects", *The RAND Journal of Economics*, Vol. 36, No. 4, Winter, 771-788.
- [2] Bell, D., T. Ho and C. Tang (1998), "Determining where to shop: fixed and variable costs of shopping," *Journal of Marketing Research*, Vol. 35, August, 352-369.
- [3] Berry S., J. Levinsohn., and A. Pakes (1995): "Automobile Prices in Market Equilibrium," *Econometrica*, Vol. 63, 841-890.
- [4] Berry S., J. Levinsohn, and A. Pakes (2004), "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market," *Journal of Political Economy*, February, 68-105.
- [5] Bruno, H., and N. Vilcassim (2008), "Structural Demand Estimation with Varying Product Availability," forthcoming in *Marketing Science*.
- [6] Bucklin, R., S. Siddarth, and J. Silva-Risso (2008), "Distribution Density and New Car Choice," *Journal of Marketing Research*, Vol.44, 473-486.
- [7] Cardell, N.S. (1997), "Variance Components Structures for the extreme value and logistic distributions with applications to models of heterogeneity," *Economic Theory*, 13, 185-213.
- [8] Chintagunta, P., J.P. Dube and K. Y. Goh (2005), "Beyond the endogeneity bias: The effect of unmeasured brand characteristics on household-level brand choice models," *Management Science*, 51(5), 832-849.
- [9] d'Aspremont, C., J. Gabszewicz, J.-F. Thisse (1979), "On Hotelling's "Stability in Competition", " *Econometrica*, Vol. 47, No. 5, 1145-1150.
- [10] Davis, P. (2001): "Spatial Competition in Retail Markets: Movie Theaters," *The RAND Journal of Economics*, Vol. 37, No. 4, 964-982.
- [11] Duan, J. and C. Mela (2008): "The Role of Spatial Demand on Outlet Location and Pricing," forthcoming, *Journal of Marketing Research*.
- [12] Ellickson, P. and S. Misra (2008), "Supermarket Pricing Strategies," conditionally accepted at *Marketing Science*.
- [13] Goeree, M. (2008), "Limited Information and Advertising in the US Personal Computer Industry," *Econometrica*, 76 (5), September.
- [14] Hoch, S., B. Kim, A. Montgomery and P. Rossi, "Determinants of Store-Level Price Elasticity," *Journal of Marketing Research*, Vol. 32, No. 1 (Feb., 1995), 17-29.
- [15] Hoch, S., Bradlow, E., and B. Wansink (1999), "The Variety of an Assortment," *Marketing Science*, Vol. 18 , No. 4, 527-546.
- [16] Irmen, A. and J.-F. Thisse (1998), "Competition in Multi-characteristics Spaces: Hotelling Was Almost Right," *Journal of Economic Theory*, 78, 76-102.

- [17] Kadyiali, V., Chintagunta, P., and N. Vilcassim (2000), "Manufacturer-Retailer Channel Interactions and Implications for Channel Power: An Empirical Investigation of Pricing in a Local Market," *Marketing Science*, Vol. 19, No.2, Spring, 127-148.
- [18] Kim, J., Bronnenberg, B. and P. Albuquerque (2008), "Consumer Search and Online Demand for Consumer Goods," working paper.
- [19] Luan, J., Sudhir, K., and B. Norris, "Dynamic Market Structure in a Durable Goods Market: The Effect of a New Product Form," working paper.
- [20] Mazzeo, M (2002), "Product Choice and oligopoly market structure," *RAND Journal of Economics*, Vol. 33, No. 2 (Summer), 1-22.
- [21] McFadden. D, A. Talvitie et al. (1977), "*Demand Model Estimation and Validation*," Berkeley, California: Institute of Transportation Studies.
- [22] Nevo, A. (2001), "Measuring Market Power in the Ready-to-Eat Cereal Industry," *Econometrica*, Vol. 69:2, (March), 307-342
- [23] Pakes, A., J. Porter, K. Ho, and J. Ishii (2008), "Moment Inequalities and Their Application," working paper.
- [24] Pancras, J. and K. Sudhir (2007), "Optimal Marketing Strategies for a Customer Data Intermediary", *Journal of Marketing Research*, Vol. 44:4 (November), 560-578.
- [25] Petrin, A. (2002), "Quantifying the Benefits of New Products: The Case of the Minivan," *Journal of Political Economy* , 110, 705-729.
- [26] Petrin, A. and K. Train (2003), "Omitted Product Attributes in Discrete Choice Models," NBER Working Paper No. W9452.
- [27] *Progressive Grocer* (1995), April.
- [28] Richards, T. (2007), "A Nested Logit Model of Strategic Promotion," *Quantitative Marketing and Economics*, Vol. 5, 63-91.
- [29] Scott Morton, F., Zettelmeyer, F., and J. Silva-Risso (2001), "Internet Car Retailing," *The Journal of Industrial Economics*, Vol. 44:4, 501-519,
- [30] Sudhir, K.(2001), "Competitive Pricing Behavior in the Auto Market: A Structural Analysis," *Marketing Science*, Vol. 20, Winter, 42-60.
- [31] Thomadsen, R. (2007), "Product Positioning and Competition: The Role of Location in the Fast Food Industry," *Marketing Science*, 26(6), 792-804
- [32] Zettelmeyer, F., F.S. Morton, and J. Silva-Risso (2007), "How the Internet Lowers Prices: Evidence from Matched Survey and Automobile Transaction Data," *Journal of Marketing Research*, Vol. 43 (May), 168-181.