Demand for Automobiles at Point of Purchase

Paulo Albuquerque† Bart J. Bronnenberg‡

November 1st, 2007

*We would like to thank Dan Ackerberg, Andrew Ainslie, Alexei Alexandrov, Charles Corbett, Dan Horsky, Sanjog Misra and Minjae Song for useful comments and suggestions. We also acknowledge comments made by seminar participants at the BBCRST (Binghamton, Buffalo, Cornell, Rochester, Syracuse, & Toronto) conference, 2007, and at the Simon Graduate School of Business.

†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 Tilburg University, Tilburg, the Netherlands, b.j.bronnenberg@uvt.nl.
Demand for Automobiles at Point of Purchase

Abstract

We study demand patterns for automobiles in the SUV category at point of purchase. This paper has two goals: the first goal is to determine what factors influence consumer choice of both which car to buy and where to buy it. The second goal is to measure substitutability between alternatives, using the estimates of the demand model. The degree of substitution between car alternatives is measured across three dimensions: geography, brands, and car segments. We compute and analyze own- and cross-prices elasticities of demand obtained from a structural model of demand that accounts for geographic distributions of consumers and dealers, and the endogeneity between price and unobserved car attributes. We make use of a transactional dataset that contains valuable information about zip code location of dealers and consumers, about wholesale prices that dealers pay to the manufacturers and about retail prices that dealers charge to consumers. We find that car characteristics such as SUV type and brand, strongly influence not only consumer choice probabilities, but also the level of substitutability between alternatives. Additionally, we find a surprisingly strong disutility for travel distance to a retailer. Given this strong preference for near-by alternatives, each dealer has its own private "backyard" of demand situated around its location, shared with a small set of other dealers. From substitution patterns, we infer that this small set of stronger substitutes is composed of dealers of the same car segment located in close proximity, and to a less degree, from other dealers located further away, if selling cars of the same car segment and brand.

Keywords: automobile industry, store competition, spatial competition
1 Introduction

The automobile industry has been the focus of a number of 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; Morton et al. 2001), and the impact of a new car form on consumer demand, for example, mini-vans (Petrin 2001) and SUVs (Luan et al. 2007). These studies provide considerable insights on 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 car dealerships as intermediaries between supply and demand. In particular, little is known about how dealer location and the geographical distribution of potential consumers interrelate to shape demand and competition patterns.

In this article, we study the interaction between retailers and consumers, in the context of spatial substitution patterns of car dealerships of sport utility vehicles (SUVs). We have two main objectives in mind. The first objective is to analyze demand for spatially and functionally differentiated products, i.e., to determine how choice depends on: (1) car characteristics, such as car brand name, prices, and accessories, (2) SUV types,\(^1\) and (3) geographic distance between buyers and sellers. The relative importance that consumers put on each of these factors determines how far they will travel to competing dealers in order to buy a car. For example, if consumers have strong preferences for a certain car brand, they will be willing to travel longer distances to obtain it. Conversely, if they place a lot of value on the avoidance of travel inconvenience, they are likely to choose among alternatives sold at near-by dealers.

The second objective of this article is to use the estimated demand model to investigate the substitution patterns between car dealerships. This is a relevant issue, since managers of dealerships are frequently concerned with the impact of competition from neighboring retailers offering similar products to consumers. For instance, the existence of geographically close competitors is likely to have important effects on a wide range of managerial issues, and a good understanding of the competitive environment and its characterization across geography serves as input for managerial decisions on pricing (e.g., Ellickson and Misra 2006), store customization (e.g., Hoch et al. 1995), and future store locations (e.g., Duan and Mela 2006). Therefore, we evaluate the impact of three important and concurrent factors on substitution patterns, namely (1) the existence of multiple dealerships offering the same brand, (2) the similarity of vehicle types, and (3) the geo-\(^1\)We consider four car types in the SUV category: mini, compact, fullsize and luxury.
graphical proximity of rival dealerships. As a result, we are able to identify, for each combination of dealer-vehicle type-brand, the set of main substitutes, as judged from the intensity of cross-price elasticities.

We use a unique individual-level transactional dataset with information about dealer and manufacturer price, manufacturer rebates, brand and type of SUV purchased, and zip code location of sellers and buyers. We augment this transactional data using information about competing prices at other locations and Census data on consumer demographics to obtain a fully defined individual data set and estimate the demand parameters using simulated maximum likelihood, taking into account consumer heterogeneity and potential endogeneity between prices and unobserved car attributes. The individual information of price and locations allows us to obtain the consumer distribution of travel sensitivity and price elasticity, which is not possible with market-level data.

Our results show that consumers regard alternatives of the same car type as close substitutes and even more so if alternatives share both the same car type and the same brand. This suggests that a consumer who decides to buy a SUV is likely to search among available alternatives within a car type and less so across different car types. We argue that the main reason for this is the fact that SUVs of similar type share similar features, especially price. One managerial implication of this result is that car manufacturers are in fact positioning their cars correctly to cover the entire market, while avoiding high substitution across different vertical segments.

An important result of the analysis is that consumers have a strong disutility for buying at far away dealers. The majority of demand of a car dealership originates from buyers located in very close proximity and we find that the probability of consumers choosing a dealer declines significantly as distance increases. We express this unwillingness to buy from far away dealers by computing the additional miles that consumers in Southern California would travel if dealers provide an incentive of $1500, as a price reduction. We find that for a $1500 price difference, about 60% of consumers would increase their travel distance by at least 5 miles and about 20% of consumers would travel at least additional 10 miles, without change in their utility. Although the impact of this rebate on travel may at first look be seen as low, we note that in the data we observe 60% of consumers traveling 10 miles or less, revealing a strong preference for dealers situated at near-by locations. The result is also robust to a number of different specifications. The large disutility for distance can be justified by the inconvenience of traveling time and costs of trips to the dealer, not only at the moment of purchase, but may also be associated with repeated future servicing trips.

As a result of the previously described consumer behavior, we find that the closest substitutes
of a given SUV are alternatives of the same SUV type. Within each SUV type, the degree of substitutability is considerably higher if dealers sell the same brand and/or distance between dealers is small. The negative sensitivity of consumers to distance causes dealer substitutability, measured by cross-price elasticity, to be limited to short distances. However, the impact of distance is moderated by coincidence of brand. To some extent, we find that alternatives sold at dealers of the same brand are seen as substitutes for longer distances than competitors selling different brands.

In conclusion, we observe that the significant disutility of distance combined with strong consumer segmentation of products by type/brand causes each dealer to have its own local demand backyard, shared mostly with a small set of near-by dealers. To the best of our knowledge, this is the first paper to show the existence of "areas of influence" of each dealer, based on the spatial location of dealers and consumers, and quantify its impact on demand and substitution patterns. The finding of limited spatial substitution between dealers has important managerial implications for pricing and communication strategies. For instance, it is likely that advertisements regarding the dealer will be effective in neighboring areas of the dealership, but ineffective otherwise.

Our work is set within the literature on spatial competition in general and the literature on competition in the car industry. The increasing interest in modeling spatial competition has led to several recent studies. For instance, Davis (2001) studies competition between movie theaters and measures the impact of increasing capacity and the introduction of new movies on competition and consumer welfare. 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. Mazzeo (2002) focuses on spatial competition in cases where the managers must decide on the level of quality, based on the presence of competing locations in the hotel industry. Thomadsen (2005) 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. These papers all use some kind of market-level aggregate data, and lack accurate individual detail on the distance traveled by consumers. In contrast, our estimates of spatial competition are based on actual location and distance between sellers and buyers. Our research seeks to extend these papers by studying the simultaneous impact of distance, brand and product type on product substitution and indirectly on dealer competition.

The literature on competition in the auto industry has studied the behavior of consumers and manufacturers. Berry et al. (1995, 1998) develop general methods for demand estimation in the
automobile industry and apply their methods to the measurement of demand elasticities. Sudhir (2001) provides 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 product 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. Finally, Busse et al. (2007) study the degree to which dealers pass on manufacturer incentives to consumers. Relevant to our study is that such incentives are often hidden from consumers by dealers. Collectively, this literature provides insights on the interaction between car manufacturers and consumers. In contrast, our research extends the knowledge on the car industry by focusing on the role of dealers in shaping demand. To the best of our knowledge, our paper is the first to study the impact of geography on demand patterns for automobiles and the implications of these patterns for competition among car dealerships.

The paper now continues with the description of the demand model in section 2. Section 3 provides details about the several data sets used in the paper. The estimation algorithm is presented in section 4 and the results are described in section 5. Section 6 concludes.

2 Demand

2.1 Utility Specification

We focus on the demand for SUVs in Southern California, and more specifically in San Diego and its suburbs. At each zip code $z$, there exists a number of households $N_z$ who consider purchasing a car. The total number of households in the market for a car is given by $N = \sum_{z=1}^{Z} N_z$. Household $i$, living in zip code $z$, chooses either to purchase one of the SUV alternatives offered at the dealerships, or not to buy from within the SUV category altogether. The households who buy an SUV may choose from SUV vehicle type $m$ ($m = 1, \ldots, M$), which includes mini, compact or full-size SUVs,2 brand $b$ ($b = 1, \ldots, B$), and dealer $d$, ($d = 1, \ldots, D$). Dealers in our data offer

---

2The SUV category is divided into several types or models (mini, compact, full-size and luxury), in which each manufacturer competes with one alternative. SUV types are defined by the research company that provided the data. Significant differences in terms of price and attributes occur across categories, as described in the next section. Given
exclusively one SUV brand \( b \), but multiple SUV types \( m \) of that brand. The consumer’s indirect utility is given by:

\[
U_{imdb} = \alpha X_{mb} + \beta_i P_{imdb} + \gamma_i \exp(-\delta D_{id}) + e_{imdb} \tag{1}
\]

with

\[
e_{imdb} = \xi_{imdb} + v_{im} + (1 - \sigma_M)v_{ib} + (1 - \sigma_B)(1 - \sigma_M)\varepsilon_{imdb} \tag{2}
\]

The first utility component, \( X_{mb} \), is a vector of alternative-specific characteristics, where an alternative is a brand-SUV type combinations. \( P_{imdb} \) is the price faced by individual \( i \) for each alternative. \( D_{id} \) is the geographic distance between individual \( i \) and dealership \( d \), measured as the Euclidean distance, in miles, between the zip code centroid of \( i \) and \( d \). The impact of distance on utility is modeled as a negative exponential function, to account for non-linear effects, namely that the marginal impact of distance is decreasing with distance (Cressie, 1993).\(^3\) Distance plays an important role in the choice decision because it captures any inconvenience cost and time of traveling to the dealership, both at the transaction occasion and at future servicing occasions. In order to capture heterogeneity in price and distance sensitivity, we divide households into groups \( g = 0, ..., G \), based on demographic characteristics, such as income:

\[
\beta_i = \beta_0 + \sum_g \beta_g H_{ig} \tag{3}
\]

\[
\gamma_i = \gamma_0 + \sum_g \gamma_g H_{ig} \tag{4}
\]

Where \( H_{ig} \) is a dummy variable that takes the value of 1 if individual \( i \) belongs to demographic group \( g \).\(^4\)

The unobserved term includes several components. \( \xi_{imdb} \) captures the impact of unobserved (to the researcher) attributes, e.g., preference for unobserved extras or accessories offered in each car that vary across individuals. It is likely that there exists a positive correlation between the preference for extras and prices, which will lead to endogeneity if this component is not accounted for. Our estimation strategy will provide direct estimates of \( \xi_{imdb} \), using the individual variation in manufacturer prices.\(^5\) The terms \( v_{im} \) and \( v_{ib} \) represent the individual preference for the SUV-type

---

\(^3\)Besides the negative exponential, we also tested a linear function and a quadratic function to measure the distance effects. The results were consistent across these three alternative formulations in term of consumer response to distance, but the negative exponential function showed the best fit with the data.

\(^4\)We estimated an alternative model specification with additional consumer heterogeneity in distance and price sensitivity, through the inclusion of random coefficients. We found to be insignificant and chose the most parsimonious model.

\(^5\)For details, see the last subsection of the data augmentation procedures, in the Data section.
and for the brand respectively. The parameter $\sigma_B$ measures heterogeneity in brand tastes, while $\sigma_M$ measures the heterogeneity of SUV types, with $0 \leq \sigma_B \leq 1$ and $0 \leq \sigma_M \leq 1$. $\varepsilon_{imdb}$ and the combined term $v_{im} + (1 - \sigma_M)v_{ib} + (1 - \sigma_B)(1 - \sigma_M)\varepsilon_{imdb}$ are assumed to be extreme value distributed. The distributions of $v_{im}$ and $v_{ib}$ are assumed to be conjugate of the extreme value distribution, such that $v_{im} + (1 - \sigma_M)v_{ib} + (1 - \sigma_B)(1 - \sigma_M)\varepsilon_{imdb}$ is also extreme value distributed (Cardell, 1997). The utility formulation leads to the variance components formulation, described in Cardell (1997) and Richards (2007).

This formulation represents the following nested hierarchy. Dealers are grouped first in $M + 1$ nests, $M$ for the different types of SUVs plus one for the outside option. Within each SUV-type nest, dealers are then arranged by sub-nests by brand $b$. For each combination of SUV type and brand, there may exist multiple dealers located at different zip codes. This hierarchy is consistent with consumers choosing vehicle types first, brand second, and dealer third. This model is very flexible and can represent many demand patterns. For instance, substitution will be stronger within SUV type or within brand as parameters $\sigma_M$ and $\sigma_B$ get closer to 1. The model collapses to the multinomial logit with consumer heterogeneity if both parameters are equal to 0.

Our choice of nest hierarchy is justified by physical characteristics and prices of different SUV vehicles. Similarity of vehicles of the same SUV-type is higher than of SUVs of the same brand. For instance, prices of different mini SUVs range from 20 to 24 thousand dollars across all brands, but prices of Ford SUVs range from 22 to 39 thousand dollars across SUV-types (see next section for more details on dealer and manufacturer prices). We expect that consumers segment the category in similar fashion, with higher correlation for similar alternatives of the same SUV-type, thus justifying our way of capturing correlation between alternatives. Additionally, within each SUV-type, choice options that share the same brand name are identical up to unobserved extras, but not so if they are from a different brand. Consequently, we allow for increased utility dependence within vehicle type and brand.

Given these assumptions, the probability of household $i$ choosing model $m$ of brand $b$ from dealer $d$ can be written as:

$$Pr_{imd(b)} = \Pr_i(d(b)|b(m)) \times \Pr_i(b(m)|m) \times \Pr_i(m) \quad (5)$$

where $\Pr_i(d(b)|b(m))$ is the conditional probability of individual $i$ buying at dealer $d$, given that brand $b$ of SUV-type $m$ is chosen; $\Pr_i(b(m)|m)$ is the conditional probability of that individual choosing brand $b$, given that SUV-type $m$ is chosen; and finally $\Pr_i(m)$ is the marginal probability of choosing a particular SUV-type or purchasing from outside of the category.
The conditional and marginal probabilities are obtained by:

\[
Pr_i(b(m)|b(m)) = \frac{\exp \left( \frac{1}{(1-\sigma_B)(1-\sigma_M)}V_{imb} \right)}{\sum_{\forall d(b) \in b(m)} \exp \left( \frac{1}{(1-\sigma_B)(1-\sigma_M)}V_{imb} \right)} \tag{6}
\]

\[
Pr_i(b(m)|m) = \frac{\exp \left( IV_{ib(m)}^{(1-\sigma_B)} \right)}{\sum_{\forall b \in m} \exp \left( IV_{ib(m)}^{(1-\sigma_B)} \right)} \tag{7}
\]

\[
Pr_i(m) = \frac{\exp \left( IV_{im}^{(1-\sigma_M)} \right)}{1 + \sum_{m=1,...,M} \exp \left( IV_{im}^{(1-\sigma_M)} \right)} \tag{8}
\]

with \( V_{imb} = \alpha X_{mb} + \beta_i P_{imb} + \gamma_i \exp (-\delta D_{id}) + \xi_{imb} \). \( b(m) \) is equivalent to \( b \in m \). Further, \( IV_{ib(m)} \) and \( IV_{im} \) are the inclusive values of brand nest \( b \) and SUV-type \( m \), which are equal to:

\[
IV_{ib(m)} = \ln \sum_{\forall d \in b(m)} \exp \left( \frac{1}{(1-\sigma_B)(1-\sigma_M)}V_{imb} \right) \tag{9}
\]

and

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

The IIA property of logit models across car models is avoided with the inclusion of (1) the correlation between alternatives within the same brand and SUV-type, captured by parameter \( \sigma_B \) and \( \sigma_M \) and (2) the introduction of heterogeneity in the price and distance sensitivity. This formulation is flexible enough to allow us to investigate if substitution patterns are stronger at the brand-level, SUV-type level or at a geographic level, as we next demonstrate.

### 2.2 Substitution patterns in the proposed demand model

To present an example of the flexibility of our demand specification at capturing alternative substitution patterns, we generate shares from several different scenarios manipulating several parameter values and compute the cross-price demand elasticity across a number of dealers. In this simulation, price and distances between dealers and consumers are similar to our data. For ease of exposition, we use a linear formulation for distance, \( \gamma D_{id} \).\(^6\) We create six scenarios where we manipulate the parameter values of distance, brand nest and SUV-type nest effects, and display the findings in Figure 1. We note that the price coefficient is maintained constant across all scenarios. It is likely that manipulating the terms for distance, brand and SUV-type nests would lead to different price sensitivity, and thus different overall magnitude in price cross-price elasticities. The purpose here is to evaluate differences across dealers in each scenario, not the overall magnitude of cross-elasticity.

\(^6\)The conclusions do not change if we use the exponential function.
As an example, we show the cross-price elasticity between an hypothetical Ford mini SUV and eight other alternatives that vary in terms of three variables: (1) SUV type (mini vs. fullsize), (2) brand (Ford vs. Honda) and (3) distance to the dealer under study (small distance - 3 miles - and long distance - 10 miles).

In scenario (a), effects for all three variables are set to zero. The cross-price elasticities are all very similar across dealers, and follow market shares, which were kept similar to create a benchmark. In scenario (b), we introduce distance sensitivity, with a linear coefficient equal to -2, while keeping the other effects at zero. In this case, consumers prefer dealers located in their vicinity rather than far away. This preference leads to higher cross-elasticity between local dealers (d = 3 miles), while dealers located far apart (d = 10 miles) are estimated to be less cross-elastic. We note that dealers located at longer distances have the smallest cross-price elasticities, no matter if they are of the same brand or type of the dealer under study.

Next, we turn on the SUV-type nest parameter to a value of 0.4 in Scenario (c), keeping the distance parameter at -2. Price changes of cars of the same SUV-type as the alternative under study, i.e., mini SUVs, sold in the vicinity of the dealer have now the strongest impact on demand, while price changes of full size SUVs have very little impact. Given the high parameter of distance sensitivity, dealers belonging to the same car-type nest but located far away still have quite low cross-elasticity values. If the distance coefficient is smaller, it is possible that SUV-type correlation becomes dominant, as it is the case with the actual data. Scenario (d) has distance parameter equal to -2, brand nest value equal to 0.2 and the SUV-type nest parameter is now zero. In this case, correlation is higher for dealers of the same brand, which decreases over distance.\(^7\)

In Scenario (e), we include a value of 0.4 for SUV-type, 0.2 for the brand nest parameter, and zero sensitivity to distance. In contrast to the previous scenario, dealers that belong to same nest and brand will have very similar cross-price elasticities, no matter their distance to the chosen dealer. Finally, scenario (f) shows the situation where all parameters take values different from zero. We now see higher cross-elasticity in near-by dealers of the same brand and segment. Distance and brand both have an impact on cross-elasticities: we observe that alternatives of the same brand but located far away (second alternative from the left) have significant cross-elasticity, as it is the case with alternatives of a different brand but located at a smaller distance (third alternative from the left). In conclusion, our individual utility formulation is flexible enough to allow for a number

\(^7\)We note that due to the assumed nest hierarchy with SUV-type at the higher level and brand at a lower level, the model has the limitation of not allowing for significantly higher correlation between dealers of the same brand that belong to the different SUV-types. See the previous subsection for the justification of the chosen nest hierarchy.
Figure 1: Cross-elasticities between Ford mini SUVs and eight other alternatives with different values for distance, brand and SUV-type nest parameters: (a) all zero; (b) Distance parameter = -2; (c) Distance = -2, SUV-type = 0.4; (d) Distance = -2, Brand = 0.2; (e) Brand = 0.2, SUV-type = 0.4; (f) Distance = -2, Brand = 0.2, SUV-type = 0.4. In each case, whenever not mentioned, the parameter was set to zero.
of different substitution patterns between dealers.

3 Data

We combine several data sets to estimate our nested model. Our main data set was obtained from a large automobile research company and it includes details about car transactions occurring in the San Diego area and suburbs between 2001 and 2005. The transaction information includes the brand and model of the car purchased, its price and model year. A highly unique feature of our data is that we have access to zip codes of the location of dealer and of consumer (assumed to be the residence location). An equally unique feature is that we have retail and wholesale price for each car, along with manufacturer rebates and transaction date. Our sample is very large and includes about 20% of all transactions in the San Diego area. A second dataset is based on U.S. Census data. It covers demographic characteristics from the zip codes included in the transaction dataset, such as income distribution and population density. We also collected latitude and longitude values of both retailer and consumer zip codes from the zipinfo database.\(^8\)

Given data volume, our analysis uses a subset of the transactions. We chose to research the SUV category, during the year of 2003, for which we observe 9,399 transactions in and around San Diego. This category provides sufficient variation across dealers, brands and types of SUVs to estimate demand for automobiles at different sellers, while keeping the dataset at a tractable size. We limit our analysis to the most important brands of SUVs in San Diego, by removing from our dataset the brands with an average market share within the category inferior to 2%. Dealers with a small volume of sales (<40 cars per year) were also not included in our analysis. This leaves us with 7,439 observations to work with, 80% of total observed transactions in the area. We believe that the major dealers in this category will be primarily concerned with their largest competitors and consequently the final dataset will not create a bias by removing very small players. Our data used in estimation includes 20 different dealerships, with 7 unique brands, for a total of 42 dealer-brand-SUV type unique combinations.

Figure 2 shows the numbers of cars sold and average dealer and manufacturer price, for each dealer grouped by SUV-type. The figure shows the presence of price differentiation between SUV-types, with an ascending order of prices for mini, compact and full-size SUVs. Within each SUV-type, it is also noticeable that prices and costs are more homogenous within the same brand than across brands. For example, among mini-SUV dealers, we observe that Jeep dealers charge a

\(^8\)Available at www.zipinfo.com.
Figure 2: Unit sales, average price and average cost of cars sold in each retailed, group by type of SUV.

higher price, followed by Ford and then all others. As previously mentioned, this observed price differentiation, larger across SUV-types and smaller within-type and even smaller within-brands was one of the motivations for the chosen hierarchy in nested logit structure presented in the previous section.

An interesting feature of our data is that it includes the location of both the consumer and the dealer for each transaction, allowing for a better understanding of the spatial distribution of demand and supply. Panels (a) and (b) of Figure 3 display the location of households who purchased an SUV in 2003 and dealerships in and around San Diego, where the size of the circles is proportional to the number of households and to the number of cars sold at the dealership.9 The majority of

---

9The small number of dealer locations is due to the coexistence of multiple dealerships in the same zip code. The number of available cars is the total across all dealerships located in that zip code.
dealerships are located fairly close to the center of San Diego, neighboring more densely populated areas. Some dealers take exception to this city location, by conglomerating about 40 miles north. This auto center includes Ford, Jeep and Honda dealerships.

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 San Diego. For one of those 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 of this dealer are located in a band of less than 20 miles, while 60% percent travel less than 10 miles to buy an SUV there. Panels (e) and (f) show a similar example for the Toyota brand.

More generally, Figure 4 provides an histogram of the distance between buyers and sellers in the transactional data. Across all dealers included in our analysis, consumers travel an average of 10 miles to buy an SUV. Only a small percentage of consumers (7%) did travel more than 20 miles to buy an SUV, while about 4.5% of consumers purchase a car in a dealer located in the zip code of their residence.

4 Estimation

4.1 Data Augmentation

The available data sets are measured at different levels of aggregation: in the transactional data, we observe individual-level details, while in the Census data, the observations are aggregated at the zip code level. We must first decide whether to reduce our individual transactional data to aggregate quantities or augment that data with information from the aggregate distributions at our disposal. We choose the latter option, since it avoids loosing the link between individual choices, location and prices. Additionally, data augmentation is used to obtain a fully defined dataset that can be used in estimation. There are three aspects that we address here: first, the transactional data set directly provides the price paid by the consumer for the chosen car but only indirectly for other alternative cars, through data from transactions of other consumers. Second, the data do not reveal the number of consumers that stayed out of the category and, for privacy reasons, consumer demographic characteristics. Finally, it lists only a subset of car attributes, e.g. options or accessories included in each car are not part of the data.
Figure 3: Spatial distribution of demand and supply for the SUV category and consumer travel to two example dealers.
We proceed as follows. To obtain car prices at retailers where the transaction did not occur, we use the observed data from other consumer transactions, e.g., car wholesale prices and dealer margins at the time of purchase. We combine these data to construct prices for unpurchased alternatives. We use Census data, collected at the zip code level, to expand the set of individuals to include both "in-category" and "outside good" purchasers. Census data is also useful to add information about consumer heterogeneity. Finally, we obtain a direct estimate of unobserved options and accessories using the manufacturer prices of the cars, to account for the impact of unobserved attributes on utility. Once these steps are done, our final specification is defined purely at the individual level and we estimate our model through simulated maximum likelihood. We next describe these steps in more detail.

**Individual Prices**  Transactional data sets commonly include information about the price paid by the consumer for the chosen alternative, but do not provide prices that the same consumer would have been charged for alternatives not purchased. Our objective here is to augment the data on prices of unpurchased alternatives for each individual, using the information that we have on (1) the wholesale price and margins of the purchased choice, and (2) the distribution of these quantities across all individuals for the other alternatives. In essence, we are assigning to each individual the price that would most likely have been observed in unrealized transactions.

We separate the final price into two components - manufacturer price and retailer margin -
because in practice they measure different quantities that may significantly vary across consumers (see e.g., Busse et al., 2007). Conditional on a certain SUV type, brand and dealer, heterogeneity in costs reflects different levels of extras and accessories, while heterogeneity in margins reflects different negotiation ability for each consumer. For each individual, we assign prices that are determined by the observed data of the actual transaction for these two components. We now clarify what we mean by this in more detail.

For individual $i$, who purchased brand $b$, model $m$ at dealer $d$, we start by identifying the set of cars the same car-type $m$ that were sold at the dealer where individual $i$’s transaction occurred. For these cars, we obtain the empirical distributions of wholesale prices and dealer margins. We then find the percentile of consumer $i$ in each of these distributions, based on his transaction details. In turn, for each unpurchased alternative, we obtain similar empirical distributions and apply the previously obtained percentiles to these distributions. We finally compute the price as:

$$P_{imb} = W_{imb} + M_{imb}$$

To better explain this procedure, we provide a practical example of how we assign prices using information for one individual in our dataset. Denote him by individual A. We observe that individual A chose to buy a Ford mini SUV paying $22,994. The dealer invoice price for his car was $22,107, which results in a dealer margin of $887 ($22,994 - $22,107). We observe 75 other transactions of Ford mini SUVs in this dealer. Among the values in the distribution of manufacturer prices and margins of these 75 cars, individual A’s percentiles were, respectively, the 73th and the 41th (see the top row panels of Figure 5). This indicates that individual A bought a Ford mini SUV that was more expensive than average in terms of manufacturer price (73th percentile), suggesting an above-average set of options or extras, and he was charged a below average margin (41th percentile), suggesting that he is a better than average negotiator.

As an example of how to compute the prices for the unpurchased alternatives, we use the car Ford full-size SUV sold in the same dealership and compute the price that individual A would have been likely to observe for this alternative. This dealer sold 181 full-size SUV units, with empirical distributions of manufacturer prices and dealer margins presented in the panels in the bottom row of Figure 5. Next, we use the percentile information from the observed transaction to obtain draws from these distributions, that is, we take the 73th percentile of the manufacturer price distribution, $36,610 and the 41th percentile of the dealer margin distribution, $1,099. The final price will then

\[10\] We tested other price mechanisms for augmenting the data. The chosen procedure produces the better fit, in terms of likelihood.
Figure 5: Example of empirical distributions and percentiles of manufacturer prices and dealer margins for the actual transaction and one non-purchased alternative.

be $36,610 + $1,099 = $37,709. The same steps are used for all other alternatives.

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 model or nothing at all.\textsuperscript{11} For this purpose, we use Census data, which provides the total number of households inhabiting in each zip code, $N_z$. The potential market for SUVs in each zip code will be a proportion of that value, based on two factors. First, we only observe a part of all transactions and consequently, we assume that the maximum potential market in our analysis is the same proportion of the total households.\textsuperscript{12} Additionally, we must take into account 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 the interpurchase time for cars to reflect this aspect on the total market potential

\textsuperscript{11}The outside good option also includes SUV purchases from San Diego consumers that travel to dealers located outside the geographical area in our study. The percentage of transactions that result these sales is however very small, less than 5% of total transactions observed.

\textsuperscript{12}The original sample contains about 20% of all transactions. After removing the small dealers and brands, this number of observed transactions is closer to 16%.
(7 years; see Sudhir 2001, for a similar approach). Formally, the total market is given by:

$$TM_z = \frac{N_z \times Observed\ Dealers}{Total\ Dealers} \times \frac{1}{Interpurchase\ time}$$ \hspace{1cm} (12)

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, \( TM_z \).

The Census data shows about one million households living in the zip codes included in our study, which results in the observed number of households for our sample of \( \sum_z TM_z = 24,070 \).\(^{13}\) As mentioned in the data section, our final sample includes 7,439 households who buy from the SUV category, representing slightly less than one third of all consumers. This seems a reasonable approximation, since in our data which includes all car transactions, sales of SUVs as a proportion of total car sales are in fact close to one third.

Individuals that chose the outside good must also be assigned prices for SUV alternatives. For each of these individuals, we take draws from the empirical distributions of prices of consumers who purchased an SUV and live in the same zip code.

**Individual Demographic Characteristics**

In order to capture consumer heterogeneity at the zip code level, we use the marginal distribution of demographics at the zip code level and assign values to each consumer. Using the income distribution as an example, this augmentation is done with the following steps. First, for each zip code \( z \), we take the empirical distribution of discrete groups of income, with the following limits [\$0K,\$35K], [\$36K,\$54K], [\$55K,\$64K], [\$65K,+\infty].\(^{14}\) Next, each individual located at zip code \( z \) is assigned a random draw from that distribution, and a dummy variable which takes the value of 1 for the group where the individual "falls in" and 0 otherwise is constructed. Finally, we repeat for all individuals, all zip codes and all demographic variables. This is done \( R \) times, and choice probabilities are averaged across these \( R \) draws.\(^{14}\)

**Unobserved Attributes**

In our model, one potential source of endogeneity is the fact that dealer prices are individual specific and some unobserved SUV specific characteristics that influence consumer utility, such as car extras or accessories, may be correlated. The average variation across dealers of these unobservables is controlled for by alternative-specific intercepts included in the utility formulation. We must also account for variation in unobserved attributes across individuals within each dealer. We do so by exploiting the information that manufacturer prices contain on

\(^{13}\)1,053,062 (# households) \times 0.16 (percentage of observed transactions) \div 7 (interpurchase time) = 24,070

\(^{14}\)We use \( R = 10 \) is our estimation. We tested larger values of \( R \) and the difference in results is unsignificant. See the likelihood function for more details.
unobserved attributes, since cars heavily loaded with extras will have higher manufacturer prices, while cars that do not include extras will be among the cheapest alternatives in each dealer.

Our approach has two stages. In the first stage, we obtain a proxy for $\xi_{imdb}$, denoted by $\xi'_{imdb}$. For each dealer $d$, this is done by computing the difference between the individual manufacturer price and the average manufacturer price of all cars of the same type for each dealer:

$$\xi'_{imdb} = W_{imdb} - \overline{W}_{mb}$$

where $W_{imdb}$ is the manufacturer price of charged to individual $i$. $\overline{W}_{mb}$ is the average of all manufacturer price of cars that share the same brand and SUV type (across all dealers). In essence, we are computing the difference in the manufacturer price of each individual car and the average manufacturer price of cars of the same type-brand. We believe that this difference to the mean is suitable to capture the variation in the amount of options and extras included in each car since cars with the same manufacturer price are identical, as manufacturers do not discriminate prices across dealers, i.e., it is always the case that cars with more (less) options will cost more (less) to the dealer.

In the second stage, a function of the residuals $\xi'_{irbs}$ will be included in the utility function. More specifically, we can write the composite error in Equation 2 as:

$$\varepsilon_{imdb} = \lambda \xi'_{imdb} + v_i + (1 - \sigma_M) v_b + (1 - \sigma_B) (1 - \sigma_M) \varepsilon_{imdb}$$

By including $\xi'_{imdb}$ as a proxy for $\xi_{imdb}$, this component will no longer end up in $\varepsilon_{imdb}$, thus avoiding possible correlation between prices and the error term and minimizing the possible bias in the price coefficient. This approach is similar to the control function approach proposed in Petrin and Train (2006), where prices are regressed on demand and cost shifters. However, there are two important differences: first, we have additional information about the manufacturer’s price, which is not affected by margin variation and market power of retailers; second, we are accounting for individual-level unobservables, and not market-level or time varying unobservables.

4.2 Likelihood

The estimation of the demand parameters is done using simulated maximum likelihood. Our previously described steps of augmenting the data now allow us to produce the following likelihood function, using the choice probabilities given by Equation 5:

$$L = \prod_{\gamma i} \prod_{\gamma imdb} \int_h (Pr_{imdb} | X, P, D, \xi', \theta, H)^{y_{imdb}} dh$$

20
where $y_{imdb}$ is an indicator variable that takes the value of 1 for the alternative chosen by individual $i$ and zero otherwise. The integral is used to account for the observed heterogeneity, where $h$ is the empirical distribution of demographic characteristics. $\theta$ is the vector of parameters to be estimated. In our algorithm, we approximate this likelihood by:

$$L = \prod_{\forall i} \prod_{\forall mdb} \left( \frac{1}{R} \sum_{r=1 \ldots R} \Pr_{imdb}^r | X, P, D, \xi', \theta, H^r \right)^{y_{imdb}} \tag{16}$$

where we take $R$ draws of the empirical distribution and average out the choice probabilities. We then maximize the log likelihood:

$$\log L = \sum_{\forall i} \sum_{\forall mdb} \log \left( \frac{1}{R} \sum_{r=1 \ldots R} \Pr_{imdb}^r | X, P, D, \xi', \theta, H^r \right)^{y_{imdb}} \tag{17}$$

5 Results

In this section, we present the results from our estimation. The results are divided into four parts: (1) findings regarding the fit of the proposed model and comparison with alternative formulations, (2) results on the impact of distance on demand, (3) analysis of own-price elasticities across brands and car types and (4) results on substitution patterns across brands, car types and dealer locations, using cross-price elasticities.

5.1 Demand results

Table 1 presents the maximum likelihood estimates of the demand parameters and log likelihoods for three alternative models. In the first model, instead of using individually assigned prices as previously described, we use dealer-specific average price. This approach does not consider individual heterogeneity in prices across individuals, but avoids the potential endogeneity of individual unobservables, since an average price implies the same average level of extras and options across all cars within a dealer. The second model uses individual prices, but does not include the proxy for unobservables. Finally, the third model is the complete model, with both individual prices and proxy for unobservables, $\xi'_{imdb}$.\[15\]

Looking first at the log likelihood of the different formulations, we observe that including individual prices provides significantly better fit than using average prices. Comparing the cases that include and do not include the proxy for unobserved attributes, we again see an improvement in

\[15\] Besides these models, we also ran logit formulations, instead of the proposed nested logit. Log likelihoods were significantly worse.
### Table 1: Mean and Standard Deviation of Demand Parameters.

<table>
<thead>
<tr>
<th>Distance base (&lt;34K)</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.853 0.265</td>
<td>1.138 0.120</td>
<td>1.197 0.191</td>
</tr>
<tr>
<td>35K-54K</td>
<td>0.684 0.274</td>
<td>0.665 0.112</td>
<td>0.687 0.194</td>
</tr>
<tr>
<td>55K-64K</td>
<td>0.685 0.333</td>
<td>0.603 0.093</td>
<td>0.626 0.220</td>
</tr>
<tr>
<td>&gt;65K</td>
<td>-1.233 0.364</td>
<td>-0.810 0.150</td>
<td>-0.864 0.251</td>
</tr>
</tbody>
</table>

Distance rate of decline 0.114 0.012 0.093 0.007 0.096 0.016

<table>
<thead>
<tr>
<th>Price base (&lt;34K)</th>
<th>-0.076 0.010</th>
<th>-0.042 0.006</th>
<th>-0.058 0.007</th>
</tr>
</thead>
<tbody>
<tr>
<td>35K-54K</td>
<td>0.029 0.009</td>
<td>0.024 0.005</td>
<td>0.025 0.006</td>
</tr>
<tr>
<td>55K-64K</td>
<td>0.025 0.011</td>
<td>0.025 0.005</td>
<td>0.025 0.007</td>
</tr>
<tr>
<td>&gt;65K</td>
<td>0.033 0.010</td>
<td>0.020 0.005</td>
<td>0.021 0.007</td>
</tr>
</tbody>
</table>

Brand-SUV type Intercepts

#### Mini SUV

- Ford -3.273 0.169 -3.015 0.050 -2.729 0.201
- Honda -2.920 0.161 -2.845 0.047 -2.569 0.200
- Hyundai -3.116 0.162 -2.929 0.050 -2.664 0.199
- Jeep -2.897 0.163 -2.816 0.045 -2.497 0.204
- Toyota -3.259 0.168 -3.043 0.053 -2.774 0.199

#### Fullsize SUV

- Chevrolet -2.571 0.197 -2.810 0.040 -2.269 0.225
- Ford -2.531 0.194 -2.770 0.040 -2.249 0.224
- GMC -2.138 0.221 -2.600 0.052 -1.981 0.237
- Toyota -2.496 0.208 -2.830 0.041 -2.272 0.226

#### Compact SUV

- Chevrolet -2.800 0.177 -2.673 0.035 -2.276 0.208
- Ford -2.556 0.167 -2.493 0.035 -2.119 0.208
- GMC -2.496 0.182 -2.525 0.034 -2.064 0.214
- Honda -2.355 0.182 -2.491 0.031 -2.054 0.212
- Jeep -2.539 0.170 -2.475 0.031 -2.061 0.211
- Toyota -2.460 0.174 -2.512 0.032 -2.119 0.209

Brand nest 0.131 0.033 0.179 0.036 0.161 0.033

SUV type nest 0.756 0.023 0.855 0.011 0.848 0.012

Unobservables $\lambda$ 0.023 0.004

Log Likelihood -38918 -38868 -38846
the log likelihood. In terms of parameters, the main difference between models (2) and (3) occurs in terms of price coefficient, becoming more (negatively) significant, from $-0.042$ to $-0.058$.

For the complete model, the estimated parameters imply an average price of elasticity of $-6.7$. For the alternative two models, we find values of $-5.4$ for model (1) and $-4.1$ for model (2). When compared with model (1), the main difference is the level of detail on prices and allowing for individual price variation leads to higher price sensitivity. We believe that this is so because part of the price response is averaged out, when we aggregate prices to the dealer level, instead of keeping them at the individual level. Contrasting the complete model with model (2), as expected, we see that including the proxy for unobserved attributes produces a significant increase in price elasticity, eliminating the bias toward zero created by endogeneity of unobservables and prices. The remainder of the analysis is done using the best model (3) and comparisons with the other models are mentioned when necessary.

We find significant heterogeneity in the price coefficient across income groups, especially between the lowest income group (income levels of less than 34 thousand dollars per year) and the remaining groups. The base group shows a coefficient of $-0.058$, while other groups have a coefficient of $-0.033$ ($-0.058 + 0.025$). Income seems to have a significant impact on price sensitivity.

In terms of the nest parameters, the SUV-type nest parameter is estimated to be close to one, with a value of 0.85, suggesting that alternatives within the same SUV-type are perceived as very close substitutes, or, in other words, alternatives across SUV types are almost in completely separated markets. We observe a value of 0.16 for the brand nest parameter. While this value significantly differs from zero, its relative smaller size suggests that within type, consumers are typically considering multiple brands. This means that in addition to the strong substitutability within car type, cars that share both SUV type and segment are perceived as the closest substitutes. A possible reason for this finding is that alternatives within SUV-types are likely to have similar physical attributes, such as size, horse power and fuel efficiency, besides offering similar prices (see Figure 2). We provide further illustration of these substitution patterns next. An implication of this result is that car manufacturers are in fact positioning their cars correctly to cover the entire market, while avoiding high substitution across different vertical segments.

In terms of fit, Figure 6 shows the actual and estimated average market shares of each combination of dealer-brand-type (excluding the outside option) for the total San Diego market (Figure 6a) and for two selected zip codes (Figure 6b and 6c), for the proposed model (4). We observe that the model explains the variations in the average share very well, but what is more important
Figure 6: Actual and estimated shares: (a) shares for each dealer for the entire San Diego area; (b) shares for zipcode 91945; (c) shares for zipcode 92014.

is that for each zip code level market, the match between estimated and actual shares is also very good, even when the alternatives have different geographical patterns of demand. The model does equally well for other not displayed zip codes. The pseudo-\(\rho^2\) is 0.57, which is also satisfactory.

5.2 Impact of distance on demand patterns

Distance between the dealers and consumers plays an important part in the decision of choosing where to buy a car, and as expected, we obtain a highly significant negative impact of distance - the farther away the seller, the less likely a consumer is to buy. To illustrate this effect, Figure 7 exemplifies how distance between buyers and sellers affects choice probabilities, using a selected Toyota dealer selling three SUV types, mini, full-size and compact SUVs. To build this figure, we start by computing the individual choice probabilities for these three alternatives, given the set of estimated parameters. Then, we use a kernel smoother with a band width of 5 miles on the estimated choice probabilities. For example, individuals located at distances between 0 to 5 miles
of the Toyota dealer show on average a choice probability of about 6.8% of buying the compact SUV model, 2.4% of buying the mini SUV and about 4% of buying the fullsize SUV.

It is clear that consumers located in a fairly confined neighboring area around the store’s location are more likely to buy from this dealer, with a large drop in choice probabilities as distance increases. Note also that this decrease with distance is not smooth due to other factors that influence utility, namely the location of competitors. For example, lower values of choice probabilities may be due to the fact that a competitor dealer is located at that distance from the Toyota dealership. In contrast, higher values of choice probabilities such as the small peak observed around between 20-25 miles and 30-35 miles are caused by the lack of dealers located in zip codes at that distance, leading consumers in those areas to be more likely to consider further away dealers, such as the selected Toyota dealer.

Another way to look at the impact of distance on demand is to investigate the spatial distribution of choice probabilities across dealers offering similar cars. Panels (a) and (b) of Figure 8 show the
geographic distribution of choice probabilities for two Ford dealers selling fullsize SUVs, designated by A and B, as a percentage of fullsize SUV alternatives. For each zipcode, we average the choice probabilities of consumers living there. The choice probabilities in map points located within zip code centroids are interpolated. Retailers of other brands are not shown for clarity of exposition.

As expected, we observe larger choice probabilities in areas surrounding the dealers’ location with both retailers having estimated choice probabilities of about 25% to 30% for zip codes located 5 miles or less away from the dealer’s location. However, the presence and location of other dealers has a major impact on the demand distribution of each dealer. For example, the average probability of choosing dealer B is actually highest not at the zip code of the dealership, but to the right of its location, further away from his strongest substitute, dealer A. This highly concentrated choice probabilities surrounding the dealers’ locations suggest the existence of private "backyard" of demand for each dealership, where most of its sales are likely to be originated.

A final aspect related to the impact of distance on demand includes quantifying the travel elasticity of consumers. For each dealer, we hypothetically decrease its price by $1500, which represents a rebate of, on average, 5% of all prices observed. In practice, we observe that dealer rebates range from $1000 to $4000, depending mostly on the price of the car. Using the model, we compute how much consumers would be willing to additionally travel, beyond what they already travel, keeping their utility equivalent to the situation before the price decrease. The distribution is shown in Figure 9.

A $1500 rebate is incentive enough for all consumers to travel an additional 3 miles. About 60% of consumers would be willing to increase their travel distance by 5 miles, and about 20% by 10 miles, keeping their utility equivalent. Although these incremental distances are surprisingly short for a rebate of $1500, we note that in the data, we observe 60% of consumers traveling 10 miles or less, revealing a strong preference for dealers located in very close to the consumer locations. For these consumers, a five-mile increase would be a 50% or more increase in the distance traveled.

Economically, the question that our paper raises is why distance has such a large impact. It is possible that consumers are concerned mostly with monthly payments which for, a 5-year payment period, would mean savings of less than $30 per month ($1500/60). Some consumers may have binding time constraints, which would make an additional hour of time spent driving to the dealer to be extremely valuable. Another argument for such preference for near-by dealerships relates to post-purchase servicing of the cars. Periodic car servicing is usually done at the dealer where the car was purchased. The distance to dealer represents not only the inconvenience of traveling once,
Figure 8: Geographical distribution of choice probabilities for two Ford fullsize SUV, within the fullsize SUV segment.
but also regularly during the lifetime of the car. On the whole, it is clear that consumers give a strong importance to geographic distance and dealers located in the household vicinity have a significant advantage to being chosen.

5.3 Analysis of own-price elasticities

As previously mentioned, we obtain an average own-price elasticity of -6.6 across all alternatives offered in our data set. However, we find significant differences in price elasticities across individuals, locations, brands and SUV types.

Given that we observe individual prices and have buyer heterogeneity across zipcodes, we obtain individual specific price elasticities. As an example, we show in Figure 10 the own-price elasticity for two Ford dealers, for their fullsize cars, aggregated by individuals in each zipcode. The size of the circles is proportional to the own-elasticity of each zip code, with large circles reflecting more negative price elasticities. The location of dealers is denoted by a star. We see that individuals located far away from the dealers are estimated to be more price sensitive. For dealer A, located close to downtown, we see consumers in the vicinity to be less price sensitive, with elasticities around -7. For larger distances, namely buyers located in the north, the elasticity increases to -10. Similar scenario happens with dealer B, which is located in the north of San Diego. In this case,
Figure 10: Own-Elasticity for two Ford dealers, for their fullsize car. A larger circle implies more negatively significant elasticity.

households located in downtown San Diego are now the most price sensitive, with elasticities of -9.

For consumers located far from a dealer, price plays a much important role in the choice decision, if in fact the alternative located far away is to be considered. In other words, the utility of close alternatives is heightened by a positive shift resulting from the lack of distance between buyer and seller, which makes price a much less important factor in the utility of near-by consumers.

In Figure 11, we plot these elasticities by SUV type and brand. The values show the percentage decline in demand, if price is increased by 1%, averaged across all alternatives that belong to the same type, or averaged across alternatives of the same brand.

We observe larger declines in demand in the fullsize segment, justified by the fact that prices are considerably higher for that segment, and consequently a 1% price increase is a larger dollar amount. Also, buying a more expensive car is a larger investment by consumers, with additional maintenance
costs, which may increase their attention to price differences. In terms of brands, Honda and Hyundai dealer in general shows the lowest price sensitivity. Brands that offer alternatives across the three types of SUVs, such as Toyota and Ford, present a widest variation in elasticities. Hyundai only sells mini SUVs, and thus its variation of own-price elasticity across its dealers is almost nonexistent. Dealer managers must take this into consideration when offering rebates, as the level of discounts needed to increase demand is car-type and brand specific, not dealer specific.

5.4 Analysis of substitution

For each alternative, we measure substitutability through cross-price elasticities, i.e., the impact of price changes of competitors on the demand of a given dealer. We exemplify our findings by looking at a Ford dealer, which sells multiple SUV models, and plot competitor alternatives by distance and cross-price elasticity in Figure 12. This figure has three panels, one for each of the type of SUVs sold at the Ford dealer. We display the alternatives that have the highest cross-price elasticities and locate them according to geographical distance, in miles, to the selected Ford dealership and cross-price elasticity values with that dealer. The size of the circles is proportional to market share of each model. We only include alternatives that have cross-elasticity above 0.05.

First, we observe that in all three SUV-types, the most important competitors come from
Figure 12: Cross-price elasticity between car dealerships, for a Ford dealer, for the three SUV-types sold.

within the same nest. Alternatives belonging to different SUV-types all show cross-elasticities significantly lower than 0.05, and are thus not represented. For example, all alternatives displayed in the top panel of Figure are mini SUVs, with cars of other SUV-type showing very small cross-price elasticities, no matter where they are geographically located. This pattern is a result of the value for the SUV-type nest parameter being close to one and thus we see considerably larger substitutability within SUV-type than across SUV-types.

Within each SUV type, two forces impact the strength of competition: distance and brand name. On one hand, it is clear that the shorter the distance, the higher the number of competitors with significant cross-price elasticities and that these values are themselves larger. On the other hand, we observe that sharing the brand name leads to an increase in product substitutability. Specially in the fullsize and compact segments, changes in prices of the other Ford dealers have abnormally strong impact on the demand of the Ford dealer when compared with the other brands.
In a number of cases, a more distant Ford dealer is a stronger substitute than other brand dealers located at shorter distances. This leads us to conclude that substitutability decreases with distance, but this decrease is moderated by sharing or not the same brand name.

The figure visually identifies the main substitutes of the Ford dealer, which vary across each model sold. In mini SUVs, the highest impact of competitor price changes on demand comes primarily from a Honda and a Jeep dealers located three miles away. In the compact and fullsize SUV segment, two Ford dealers located three and nine miles away are the closest substitutes, while Jeep dealerships solely show up as competitors in the compact segment.

6 Managerial implications and conclusion

This paper presented an approach to the problem in which a car dealership manager wants to (1) measure the impact of distance, product types and brands on demand patterns and (2) identify substitute dealers that sell the same or other car brands at other locations. We show that both prices and distance between consumer home and dealers impact demand significantly. This suggests that each dealership has a localized area of influence, his own demand "backyard", where consumers have a high choice probability of buying from that dealer. This choice probability decreases at a fast rate as distance increases.

We also find that consumers strongly segment the SUV category by type of SUV. Two additional factors moderate substitutability: distance and brand. Within each SUV-type, sharing the same brand increases substitution degree, while increasing distance between dealers reduces it. In most cases, a dealer sells cars that span over a number of car segments. We find different sets of significant substitute dealers for each SUV-type, measured by cross-elasticities.

Dealers must be aware of these facts when negotiating the final price with the consumers, setting the amount of discounts and options, or determining geographic areas where to advertise. First, in terms of reacting to price changes of competitors, dealers should first make sure that a given car sold at a competitor dealer is in fact seen by consumers as a substitute of their own products. Substitutability is strongly and geographically localized. Second, we observe in practice considerable advertising for dealerships in both TV and press. Given the negative impact of distance on utility, advertising in areas far from the dealer’s location are likely to have an impact on choice if a considerable price incentive is also used to compensate for the inconvenience of traveling not only for the purchase occasion but future servicing trips to the dealer. Third, given our results, it is likely that other dealers selling the same brand are the most relevant substitutes, since the cars
offered are identical up to the options or extras included in each car. Accounting for these individual options and consequently individual prices influences price sensitivity, and must be accounted for when selling a car.

We believe that investigating the part played by dealers in shaping the demand and supply of the car industry can lead to several lines for future research in this area. For instance, a possible next step in the analysis of the car category would be to include the interaction between manufacturers and dealers in the supply side. An analysis of this kind would shed some light on car manufacturers and dealers conduct and the distribution of market power in the car industry.
References


