Consumer Search in the U.S. Auto Industry: The Value of Dealership Visits *

Current version: December 2019

Dan Yavorsky† Elisabeth Honka‡

Abstract

In many markets, consumers visit stores and physically inspect products before making purchase decisions. We view the inspection of a product at a retail location as a search for product fit. We quantify the cost and benefit from searching for product fit using a discrete choice model of demand with optimal sequential search. In these models, the benefit of searching is measured by the standard deviation of the product fit and has, heretofore, been fixed to one for identification purposes. With an exogenous search cost shifter, both the cost *and* benefit of searching can be separately identified. Our empirical setting is the U.S. automotive market. We assemble a unique data set containing individual-level smartphone geolocation data that inform us about dealership visits. We also obtain information on new vehicle purchases from proprietary DMV registration data. Our exogenous cost shifter is the distance a consumer must travel to visit a dealership. Our results show that the benefit provided by dealerships to consumers is substantial. Failure to estimate the standard deviation of the product fit leads to biased search cost and consumer surplus estimates and to inaccurate predictions regarding the number of searches consumers conduct.

Keywords: Consumer Search, Automotive Industry

JEL Classification: D83, L62, M31

*We are extremely grateful to Peter Rossi. We thank Keith Chen, Brett Hollenbeck, Randy Bucklin, Sylvia Hristakeva, Sergei Koulayev, Stephan Seiler, Marco Testoni, Raluca Ursu, and Geoff Zheng for helpful conversations. We also thank participants at the University of Houston Doctoral Symposium and seminars at UCLA and UC San Diego for their comments. All errors are our own.

†University of California Los Angeles, dan.yavorsky.phd@anderson.ucla.edu.

‡University of California Los Angeles, elisabeth.honka@anderson.ucla.edu.

1 Introduction

Prior to the Internet, store visits and the physical inspection of products played a prominent role during consumers' purchase process – especially for high-ticket durable goods such as automobiles, furniture or electronics. Information on prices and product features was difficult to obtain other than via store visits. During the last two decades, a wealth of online information has become available, giving consumers an alternative method to easily gather product information including prices. Nowadays, consumers in the market for a durable good still visit brick-and-mortar stores to obtain information about product attributes which are difficult to find online. These consumer-specific preferences for certain product attributes are often called "product fit" or "match value."[1] An open question is how much consumers benefit from visiting brick-and-mortar stores today.

Across many product categories, consumers in 2019 visit fewer brick-and-mortar stores and buy more goods online than they did 15 years ago. However, looking at the number of physical store visits alone is not sufficient to determine the benefit from visiting brick-and-mortar stores. The reason is that the number of visits is *jointly* determined by both the cost and the expected benefit of a store visit. Thus consumers might be visiting fewer brick-and-mortar stores because of higher cost or lower expected benefit (or a combination of both). To answer our research question, the cost and benefit of store visits must be *separately* quantified. In this paper, we show that this can be achieved in a sequential search model for product fit with access to an exogenous search cost shifter.

Our empirical application is the new car market. Consumers typically visit dealerships before making a purchase. Prior to the Internet, Ratchford and Srinivasan (1993) find that consumers made, on average, 4.6 dealership visits in 1986. Ratchford, Talukdar, and Lee (2007) find that number to be 2.2 in 2002. By 2016, this number had decreased to 1.3.[2] This empirical

[1] Throughout this paper, we use these two terms interchangeably. In our empirical context of new car purchases, the match value captures, e.g., how spacious and comfortable a car is, how it drives, how courteous and helpful the dealership staff is, etc. to an individual consumer.

[2] During a 2016 FTC Workshop on auto distribution in the U.S., Peter Welch, President of the National Automobile Dealers Association (NADA), reported that "the average number of dealerships that a consumer visits before they make a purchase, an actual purchase, has gone down from 4.1 dealerships in 2005 to today, it's

pattern in the new car market is consistent with patterns found in many other product markets.

Our unique data on the new car market come from Texas. We observe mobile device geolocation information that captures each consumers' home locations and dealership visits, i.e., individual-level offline search behavior unintrusively collected in real time. These unique data stand in contrast to previous literature on offline search that neither observes the number nor the order of searches – especially in the car market (e.g., Nurski and Verboven 2016, Murry and Zhou 2019). An exception is Moraga-Gonzalez, Sandor, and Wildenbeest (2018). However, the authors rely on aggregate moments of the number of searches from retrospective survey data. Our data describe both the number and order of searches at the individual-level and are unobtrusively gathered while consumers engage in the actual search behavior. We combine these data with information on new car registrations from the Texas Department of Motor Vehicles (Texas DMV), i.e., new vehicle purchases. We supplement these data with consumer and vehicle characteristics, and information on the location of auto dealerships and the brands they carry.

Importantly, these data allow us to measure the distances between each consumer's home and each searchable dealership – distance being our exogenous search cost shifter. We estimate a sequential search model for product fit à la Weitzman (1979) and parametrize search cost as a function of distance. Through simulations, we show that exogenous variation in search cost identifies the standard deviation of the match value distribution (MVSD). This parameter is a direct measure of the potential benefit from search. In prior literature, this parameter was commonly fixed to one for identification reasons (see, e.g., Kim, Albuquerque, and Bronnenberg 2010, Dong et al. 2019).

We estimate the MVSD to be 8.16. This estimate is large relative to its typical normalization to 1 in search models for product fit and to the scale of the utility parameters (which is typically also normalized to 1 for identification reasons). The magnitude of the MVSD demonstrates that dealerships continue to provide substantial value to car shopping consumers today. Further, we also find that normalizing the MVSD to 1 has at least three implications. First, it results in an

1.3" (https://www.ftc.gov/system/files/documents/public_events/895193/auto_distribution_transcript.pdf).

overestimation of the impact of distance on search cost. For example, increasing the distance to a dealership from 10 to 15 miles increases estimated search cost by 9% when the MVSD is normalized to 1. The same increase in distance yields an estimated search cost increase of only 3% when the MVSD is estimated.

Second, because normalization leads to incorrect estimates of the cost and benefit of search, the normalized model yields inaccurate predictions regarding the number of searches consumers engage in and incorrect estimates of consumer surplus. Normalizing the MVSD to 1 leads to an overprediction of the proportion of consumers who search once and an underprediction of the proportion of consumers who make two or more searches, i.e., the long tail of the distribution of the number of searches. Estimating the MVSD allows the model to more accurately predict the distribution of the number of searches and especially its long tail. Furthermore, for the auto industry, we find that normalizing the MVSD to 1 severely overestimates consumer surplus.

And lastly, as a result of the biased parameter estimates, managerial and policy analyses based these estimates will be incorrect. We demonstrate this when assessing the adoption of at-home test drives. This counterfactual is inspired by a new marketing technique pioneered by Hyundai through its "Hyundai Drive" program. In this program, consumers can reap the benefits of search without incurring any distance-based search costs. Across the studied brands, we find that unilaterally initiating at-home test drives yields market share increases of 2.3–4.6 percentage points. However, the same counterfactual conducted with the MVSD normalized to 1 understates the effects of the program and suggests market share increases of only 0.8–1.6 percentage points.

This paper makes two primary contributions to the literature. First, we show that the MVSD can be separately identified from search cost with access to an exogenous search cost shifter. Further, failure to estimate the MVSD leads to incorrect model parameter and consumer surplus estimates, incorrect predictions regarding the distribution of the number of searches, and incorrect conclusions from policy analyses. Similar to search models for prices in which the standard deviation of the price distribution is commonly estimated from data, we show the importance of doing the same for search models for product fit. And second, this study is

among the first to provide a detailed assessment of physical ("offline") search using individual-level observational data. Other studies of offline consumer search behavior mostly rely on path tracking information within a store, aggregate or retrospective survey data. Other studies of consumer search using individual-level observational data do so in an online context using cookies or other web-tracking technology to assess browsing behavior. Access to data from new technologies such as mobile device geotracking gives unintrusive access to consumers and allows researchers to observe them in their daily behavior opening doors to new research opportunities.

The outline of this paper follows: we discuss this study in the context of the relevant literature in Section 2. In Section 3, we present the data. We provide reduced-form results in support of our modeling framework in the following section. In Section 5, we formally introduce the model before discussing identification and estimation in Section 6. We present the empirical results in Section 7 and a counterfactual in Section 8. Finally, we conclude and discuss future research in the last section.

2 Relationship to Existing Literature

This study contributes to the streams of literature on consumer search and on the role of dealerships in the U.S. auto industry. In the following, we review the relevant literature and delineate the positioning of our research vis-à-vis the findings from extant research.

Previous literature estimating sequential search models for product fit usually utilizes data on online browsing behavior. For example, Kim, Albuquerque, and Bronnenberg (2010) and Kim, Albuquerque, and Bronnenberg (2017) use aggregate view-rank and purchase data for camcorders sold on Amazon.com. Koulayev (2014), Chen and Yao (2017), De los Santos and Koulayev (2017), and Ursu (2018) study different aspects of online hotel bookings. Yao, Wang, and Chen (2017) offer an application to television viewing and Morozov (2019) studies new product introductions in the computer hard drive market. In contrast to the previously mentioned papers, we focus on offline consumer search.

Because of limited data availability, studies of offline search behavior are very rare. Two

exceptions are Jain, Misra, and Rudi (2016) and Seiler and Pinna (2017).[3] Jain, Misra, and Rudi (2016) assess the effects of sales assistance and search on purchase incidence and expenditure using video recordings of a retail clothing store's product display area. Seiler and Pinna (2017) measure the returns to price search by analyzing shopping cart movements in a supermarket. These two papers differ from ours in two aspects: they focus on the duration of search, i.e., time spent searching, rather than the number of searches and they study consumer search activity within a store, while we focus on search activity across stores.

Most closely related to this paper is Moraga-Gonzalez, Sandor, and Wildenbeest (2018) who incorporate sequential search for product fit directly into the framework of Berry, Levinsohn, and Pakes (1995). The authors supplement aggregate car sales data from the Netherlands with moments from a survey on Dutch consumers' search behavior. Using these data, Moraga-Gonzalez, Sandor, and Wildenbeest (2018) predict the searched dealership(s) and the search order, whereas we observe these decisions directly in the data.

This study also fits into the literature on car dealership locations and how they affect consumer demand and competition. Bucklin, Siddarth, and Silva-Risso (2008) show that, prior to the Internet, new car buyers were more likely to select cars whose dealer networks had shorter distances to the closest outlet and more dealers within a given radius from the buyer. Albuquerque and Bronnenberg (2012) estimate a model of supply and demand that does not incorporate consumer search behavior, but utilizes the locations of consumers and dealers (and thus the distance between them) and find that consumers have a strong disutility for travel. Palazzolo and Feinberg (2015) generalize a full information discrete choice model to incorporate the probability that an observed consideration set is optimal while flexibly permitting consideration set substitution among available options. The authors use their model together with survey data to study the impact of vehicle redesigns and recalls on consideration and purchase. Murry and Zhou (2019) study the agglomeration-competition trade-off of dealership co-location

[3]Seiler and Yao (2017) use supermarket path tracking data combined with advertising data to investigate how advertising affects consumers along their conversion funnel. Hui, Bradlow, and Fader (2009) and Hui et al. (2013) also study offline behavior in a supermarket using path tracking data. However, these two papers investigate vice versus virtue product purchases, the effects of other shoppers on shopping paths and purchases, and the effectiveness of mobile promotions and not consumer search behavior.

using a sequential search model and transaction-level data for new vehicles sold in Columbus, Ohio. While search behavior is modeled, the authors do not have data on the actual searches performed by consumers. In contrast to the previous literature on car dealerships, in this paper, we observe individual consumers' dealership visits and estimate a sequential search model that uses distance to dealerships as an exogenous search cost shifter.

3 Data

3.1 Data Sources and Cleaning

We combine data from several sources for the empirical analysis. Mobile device geolocation data inform us about consumers' home locations and dealership visits. However, these data do not provide information on the purchased vehicle. Therefore we combine the dealership visit data with data on new car registrations from the Texas DMV. By combining these two data sets, we observe both the search sequence and the purchased vehicle at the individual level. We supplement these data with consumer and vehicle characteristics, and information on the location of auto dealerships and the brands they carry.

3.1.1 Search Data

We obtained consumer search data from Safegraph, a company that "provides high-quality location data products" by aggregating location information from various mobile device applications.[4] For the three-months time period from November 1, 2016 to January 31, 2017, Safegraph provided us with two types of information for each (anonymized) individual mobile device: home locations and dealership visits. The home locations were generated by Safegraph's proprietary model that is based on the (im)mobility of the device, time of day, and assumed work patterns of device owners. The home locations are stored at the geohash-8 level.[5] The

[4]See https://github.com/YalePrivacyLab/tracker-profiles/blob/master/trackers/SafeGraph.md and https://www.safegraph.com/.

[5]A geohash is a public domain geocoding system that encodes location into a short string of letters and numbers. A geohash-8 is approximately 40 meters wide by 20 meters tall, or 800 square meters.

dealership visits were generated by Safegraph merging their mobile device geolocation data with their proprietary data set of U.S. dealership geospacial locations. A unique record is a mobile device at a dealership (identified by name and street address) at a specific date and time. Thus, for each mobile device, we observe the visited dealerships and the order of those visits.

The dealership visit data record all dealership visits – for any reason. As a result, they may capture behavior other than new car searches. To remove errant observations, we exclude all data for a device if the device is observed (i) at any dealership more than 25 times, (ii) at the same dealership more than 10 times, or (iii) at any dealership between 12:00am and 6:00am. These criteria are applied to exclude, for example, the device of someone who works at, delivers vehicles to, or provides janitorial services for an auto dealership. In addition, we limit the dealership visit data to device owners with home locations in the state of Texas. The remaining data consist of approximately 154,000 unique mobile devices making 277,000 dealership visits to one or more of the 1,258 dealerships identified by Safegraph.

Dealership visits of consumers shopping for a new car can spread over days or even weeks. This raises the concern that vehicles purchased at the beginning of the 3-month time period may have truncated search sequences. To investigate this concern, we calculate the average number of days between the first and last search for consumers whose last search occurred in November 2016 versus during December 2016 and January 2017. We find these values to be 1.1 and 10.2 days, respectively.[6] These numbers suggest that many search histories for consumers who made a purchase in November 2016 are truncated. As a result, we limit our analysis to only those consumers who purchased their vehicles in the latter two months ("sample period"). This reduces the sample size to approximately 127,000 consumers making 243,000 dealership visits.

[6]If we exclude consumers who only made one search, the mean and median number of days between the first and last search for consumers whose last search occurred during first month are 9.2 and 7, while the mean and median number of days between the first and last search for consumers whose last search occurred during the latter two months are 29.3 and 25.

3.1.2 Purchase Data

Our second data come from the Texas DMV. We observe all first-time titled or registered vehicles in the state of Texas during (or up to 14 days after) the sample period.[7] The data include the Vehicle Identification Number (VIN), the registrant, the registrant's address, the date that the title or registration paperwork was processed by the state of Texas, the gross sales price before any adjustment for a trade-in vehicle, the VIN for the trade-in vehicle if applicable, and the name, city, and state of the previous owner. It is important to note that information on the previous owner identifies the dealership from which a vehicle was purchased ("selling dealership"). While most of this information is available to the public through a Freedom of Information Act (FOIA) request, the personally identifying information is not. It must be obtained through a special request and its use is subject to restrictions.

To focus on new retail vehicle sales for which consumer search data may be available from Safegraph, we limit the Texas DMV data to vehicles that belong to the 35 most popular brands from model years 2015–2018 with an odometer reading of fewer than 2,000 miles, a price exceeding $5,000, and for which the date of title occurred during the sample period (or less than 14 days after its end). These criteria are used to omit used vehicles, commercial vehicles such as freight trucks, and alternative vehicles such as tractors, motor homes, and motorcycles. In addition, we constrain the Texas DMV data to only vehicles for which the registrant has a non-PO Box home address in the state of Texas; this is a necessary criterion for combining the search and purchase data. Finally, data are excluded if more than one vehicle is titled to the same individual during the sample period or if the registrant is not an individual. Approximately 195,000 vehicles meet these criteria.

Figures 1(a) and 1(b) each provide a map of individuals' home locations. Figure 1(a) displays home locations for searchers from the Safegraph data, while Figure 1(b) displays home locations of buyers from the Texas DMV data. Although the data sources are different, the figures show a similar geographic distribution of consumers across the state and major urban

[7] In a conversation with a Texas DMV employee, her experience was that title and registration paperwork is typically processed between one and fourteen days after submission to the DMV or county tax office, and that submission typically occurs immediately upon sale of a vehicle by a dealership.

areas are clearly identifiable.

==========

Insert Figure 1 about here

==========

3.1.3 Dealership Data

We manually prepared a third data set consisting of all auto dealerships in the state of Texas and the brands carried at each dealership. This data collection was necessary for two reasons. First, the colloquial and legal definition of a dealership vary and we required a definition compatible with the Safegraph data. We use the term and have organized the data such that a "dealership" represents a distinct geographic area of new vehicle retailing. For example, although Randall Noe Chrysler Dodge and Randall Noe Subaru are legally distinct dealerships, their showrooms share the same building, they have the same street address, and the vehicles in inventory are adjacently located on the same lot. We therefore categorize them as one dealership. And second, to make reasonable assumptions about vehicles that were searched but not purchased, as required for the structural search model to be introduced later, we require information on the set of brands carried at each dealership.

We identify 1,314 auto dealerships that carry new vehicles in the state of Texas. This number of dealerships is very similar to the number of dealerships (1,309) listed by the Texas Automotive Dealership Association (TADA) as of February 2018.[8] The small difference in the number of dealerships is due to (i) dealerships that operated during the sample period but closed by the 2018 TADA count and (ii) TADA and us differing in how to treat two adjacent and related retail locations (TADA may recognize them as two dealerships while we count them as one or vice versa). Figure 2(a) plots the locations of all 1,314 auto dealerships. As expected, dealership locations are highly correlated with the geographic distribution of the population. Figure 2(b) provides a map of the Houston metropolitan area onto which dealership locations

[8] https://www.tada.org/web/Online/About_TADA/Dealer_Search.aspx

are overlaid. It shows that dealerships often occur spatially clustered and are generally located along major roadways.

==========

Insert Figure 2 about here

==========

Consistently identifying dealerships across data sets is essential for combining the search and purchase data. The manually compiled data on dealerships are matched (i) to the dealership street addresses in the Safegraph search data and (ii) to the previous owner names in the DMV registrations data. An overlapping subset of 1,197 dealerships are identified from both data sets. Not all dealerships could be merged either because the previous owner names in the Texas DMV data were not sufficiently descriptive to uniquely identify a dealership in the Safegraph data or because the Safegraph data did not contain visits to that dealership. This latter case could occur if Safegraph did not have a geofence for that dealership or if no mobile device in the Safegraph data was observed to visit the dealership between November 2016 and January 2017.

3.1.4 Vehicle Characteristics Data

VinAudit.com, Inc., a leading vehicle data and software solutions provider for the U.S. automotive market, provided data on vehicle characteristics. The data were obtained by "decoding" the registered and trade-in VINs using the company's API. Collected vehicle characteristics include model year, make, model, trim, "base" MSRP, vehicle type (car, SUV, truck, van), body type (e.g., sedan), number of doors, drive type (e.g., front-wheel drive), engine size, and transmission type.[9] We supplement these data with information on vehicle horsepower collected from Google.

[9]The APIs used to gather car characteristics provided "base" MSRP. The definition of "base" is manufacturer-specific and may be at either the model-level or trim-level. These base prices do not include features that customize the vehicles beyond the set of included features in the most basic configuration of the manufacturer's model or trim. For example, if a sunroof is included as part of a specific 2017 Ford Fiesta SE but the sunroof is not a standard feature on all 2017 Ford Fiesta SEs, then the added cost of the sunroof is not included in the collected MSRP. This means that, for example, two vehicles that are largely similar (i.e., same model or same trim) but have different optional features will show the same MSRP.

We also collected data on vehicle "types" and rankings within each type from Edmunds.com. Edmunds classifies all vehicle models into one of 40 types. For example, the Honda Fit is categorized as an Extra-Small Hatchback, while the Toyota Highlander is categorized as a Midsize 3-Row SUV. Within a type, Edmunds ranks each vehicle. Their ranking process is proprietary and opaque, described as: "Each vehicle is driven on a standardized road test loop and visits our test track for instrumented testing in controlled conditions. Our time behind the wheel is used to develop ratings that describe how a car stacks up against its direct rivals in a particular size and price class."[10] The Edmunds data are used to make assumptions about (similar) searched or potentially searched, but not purchased vehicles.

3.1.5 Consumer Characteristics Data

We collected consumer demographic data from the U.S. Census Bureau's 2010 Census and American Community Survey. The data include information at the Census Blockgroup level on age, race, gender, educational attainment, income, employment status, and number of children.

3.2 Construction of Analysis Data Set

3.2.1 Combining Search and Purchase Data

There is no unique identifier in the Safegraph dealership visit data that corresponds to information in the Texas DMV registrations data to indicate which search sequence corresponds to which purchased vehicle. Therefore, to merge these two datasets, we employ the following algorithm: first, the algorithm finds a search sequence that includes a visit to the selling dealership prior to the vehicle registration date, and second, the algorithm requires that the home location of the searcher and of the buyer are approximately the same. More details on the employed algorithm are documented in Appendix A.

With this algorithm, we find the purchased vehicle for 12,065 search sequences. This represents 9.5% of the mobile devices in the Safegraph data and 6.2% of the vehicle registrations

[10] https://www.edmunds.com/new-car-ratings/

in the Texas DMV data. These figures appear reasonable for the following reasons: first, the search data also include dealership visits for consumers who were shopping for used vehicles or who were taking a currently owned vehicle to a dealership's service department for maintenance – both activities are likely to occur with greater frequency than new car shopping. Second, the search data also include dealership visits by consumers who searched, but did not purchase a new vehicle. And third, the search data purportedly represent 5%–10% of U.S. mobile devices and thus it is expected that these data capture approximately the same percentage of new car purchases.

3.2.2 The Searchable Set of Dealerships

In much the same way that a multinomial discrete choice model requires information on the chosen product as well as the products not chosen, our structural search model requires information on the set of (potentially) searchable products. Across the 35 brands, there are more than 1,100 dealerships in the state of Texas. To ease the computational burden imposed by such a large set of searchable products, we take two steps.

First, we limit the analysis to the five brands with the largest market shares in Texas during the sample period, which are Chevrolet, Ford, Honda, Nissan, and Toyota. This set of brands has a combined market share of 60%. Focusing the analysis on these brands requires that we also limit the data to consumers who purchased these brands and whose search histories only include visits to dealerships that carry these brands. This reduces the sample size to 7,051 consumers.

Second, we limit the searchable set of dealerships to any dealership within a particular radius of the consumer – twice the 70th percentile of the distance between the home location of a car buyer and the selling dealership. We do so by county and also differentiating between rural and urban consumers. Details of the employed procedure are described in Appendix A. The average radiuses for urban and rural consumers are 29.6 and 33.5 miles, respectively. We find that 7.6% of consumers search or purchase a vehicle from a dealership outside of their assigned radius and exclude these consumers from the analysis sample. This step reduces the sample

size to 6,511 consumers making 7,175 dealership visits.

3.2.3 The Searchable Set of Vehicles

Through the process of merging the search and purchase data, we know which vehicle each individual consumer purchased at which selling dealership. However, for all dealerships she visited but did not make a purchase, we do not observe which vehicle she searched. Further, for all dealerships the consumer decided not to visit, we do not observe which vehicle she would have searched if she had decided otherwise. To estimate the structural search model, we need both pieces of information. We make the following set of assumptions about searched and potentially searchable vehicles.

For each consumer and each searchable dealership identified by the process outlined in Section 3.2.1, we assume that there is one most preferred vehicle that could be searched, i.e., if the consumer were to visit a particular dealership, she would have searched only her most preferred vehicle at that dealership. It is important to note that the most preferred vehicle at a particular dealership is consumer-specific: if the same dealership is searchable by multiple consumers, each consumer has her own most preferred vehicle at that dealership. To identify this consumer-specific, most preferred vehicle at each dealership, we impose two criteria: the potentially searched vehicle must be in that dealership's inventory and it must be of the same Edmunds type as the vehicle ultimately purchased by the consumer. The latter assumption is equivalent to assuming that consumers search for a specific vehicle conditional on having decided the type of vehicle they would like to buy.

Finally, we must also infer the set of vehicle characteristics for each searchable vehicle. To do so, we compute the median values (for numeric characteristics) and modal values (for categorical characteristics) by model and dealership. We document details of the construction of the searchable set of vehicles and their characteristics in Appendix A.

3.3 Descriptive Statistics

The final analysis sample contains $6,511$ consumers who make $7,175$ visits to one of 544 dealerships. 91.1% of consumers search once, 7.8% of consumers search twice, and 1.1% of consumers search three or more times. Figure 3 shows a histogram of the number of searches. The average number of searches per consumer is 1.1. This average number of searches in our data is consistent with the most recent reports, albeit slightly smaller (see footnote 2).

==========

Insert Figure 3 about here

==========

The distance a consumer must travel to visit a dealership is our key variable. Consumers live in close proximity to many dealerships. On average, consumers in the analysis sample live within 10 miles of 6 dealerships. As we extend the radius to 20 and 30 miles, that number increases 16 and 24 dealerships, respectively. The observation that consumers live in close proximity to many dealerships, but only search a limited number of them suggests that it is important to take their search behavior into consideration when modeling demand.

Not only are many dealerships located near consumers, but consumers also tend to purchase from nearby dealerships. The median distance from a consumer's home to the selling dealership is 5.2 miles. Figure 4(a) shows the distance distribution from home to the selling dealership: most consumers purchase from dealerships within 30 miles of their home.[11] As a point of comparison, in Figure 4(b), we display the distribution of percentiles for the distance to the selling dealership in each consumer's set of searchable dealerships. For example, suppose a consumer bought from the closest dealership and had five searchable dealerships. Then the consumer purchased from a dealership in the 0.2 percentile. The distribution in Figure 4(b) shows that consumers tend to purchase from dealerships below the 0.5 percentile. The few consumers who purchase from a high percentile dealership are mostly consumers with small

[11] Albuquerque and Bronnenberg (2012) find the mean and median distance between consumers' homes and *selling* dealerships to be 10 and 7.3 miles, respectively, based on zip code centroids. These distances are also consistent with not observed, but *estimated* distance distributions reported by prior literature (e.g., Nurski and Verboven 2016, Moraga-Gonzalez, Sandor, and Wildenbeest 2018, Murry and Zhou 2019).

sets of searchable dealerships (fewer than five dealerships).

==========

Insert Figure 4 about here

==========

Consumers can purchase a vehicle from the closest dealership carrying the purchased brand or they can purchase from a more distant dealership carrying the same brand. 72% of consumers in the analysis sample buy from the closest dealer offering the purchased brand, the remaining 28% do not. Figure 5 shows the distance (beyond the closest dealership) to the selling dealership for those consumers who did not purchase from the closest dealership offering the ultimately purchased brand. The mean and median "extra" traveled distance are 7.6 and 5.7 miles, respectively. Buying from a more distant dealership suggests that that dealership offers an observed (to the consumer) prior-to-search benefit that exceeds the additional time, travel, and mental costs incurred from making the longer trip.

==========

Insert Figure 5 about here

==========

And lastly, descriptive statistics on dealerships and purchased vehicles are displayed in Table 1. Chevrolet and Ford have many more dealerships than Honda, Nissan, or Toyota. The average vehicle characteristics reported in Table 1 are influenced by pickup truck sales. Chevrolet and Ford sell a higher percentage of trucks in the Texas market than the other brands and, as a result, tend to have higher average horsepower and engine sizes, while offering lower gas mileage.

==========

Insert Table 1 about here

==========

4 Reduced-Form Evidence

In this section, we investigate the set of variables that predict consumer purchase and show that including distance to dealerships is important when modeling demand for cars. We do so by

estimating three standard multinomial logit models. Because consumers are assumed to choose a vehicle from a set of vehicles that are of the *same* Edmund's vehicle type, it is not necessary (or possible) to include vehicle characteristics such as the number of doors or engine type (gas vs electric) as these characteristics do not vary across vehicles within an Edmund's type. Instead, we focus on three major characteristics likely to impact consumer decision-making conditional on type: horsepower, engine size (i.e., displacement measured in liters), and estimated city mileage (MPG). In addition, because our data are from Texas, we include interactions with a dummy variable ("large vehicle") to indicate if the vehicle is a truck or a large SUV (most of which are built on a truck chassis). And lastly, we use MSRP as a measure of price.[12]

In Table 2, we show coefficient estimates and implied price elasticities for two full information models (a) and (b), in which the consumer is assumed to have knowledge of all searchable alternatives, and a limited information model (c), in which the consumer only chooses among the products she has searched. The two full information models differ in the included covariates with model (b) additionally including the distance to each dealership.

==========

Insert Table 2 about here

==========

Across all three logit models, the coefficient estimates are similar in terms of signs and magnitudes. Coefficient significance is also generally consistent with the exception of MPG in the limited information model (c). Price negatively affects utility. On average, consumers prefer Toyota and Honda to Nissan, Ford, and Chevrolet for small and medium-sized vehicles. For large vehicles including pickup trucks, consumers prefer Ford. Among the vehicle attributes, engine size and city mileage have the expected positive sign. Horsepower is estimated to have a negative effect on utility. A potential explanation is that, conditional on vehicle type, engine size, and city mileage, vehicles configured to provide extra horsepower may require additional maintenance and are therefore not preferred by consumers.

[12]While the data include the actual price paid for each purchased vehicle, MSRP is observed for the purchased and *not* purchased vehicles prior to search. It also better reflects consumers' knowledge of prices prior to search.

Recall that model (b) additionally includes distances to dealerships as a covariate. The distance coefficient is (as expected) negative, large (in absolute terms), and precisely estimated. Moreover, its addition provides substantial improvements in the log-likelihood and Bayesian Information Criterion (BIC) indicating that distance adds to the explanatory power of the model. In addition, price elasticities are more elastic when distance is included.

Comparing models (a) and (b) to model (c), price elasticities are smaller than one (in absolute terms) in the limited information model. The price elasticities are inelastic because of the large number of consideration sets of size one: consumers with such a consideration set would not select a different alternative even if prices increase because they only have one alternative available to them.

5 Model

5.1 Utility and Search

We model consumer search and purchase decisions using a sequential search model for match value. A product is defined as a combination of a specific vehicle and a dealership (e.g., a Honda Civic at the John Eagle Honda Dealership of Dallas). Thus the match value represents everything a consumer learns about a specific vehicle at a specific dealer by visiting the latter. This includes (but is not limited to) how spacious and comfortable the car is, the layout of the driver dashboard, how the car drives (i.e., speeds up, handles the road, brakes), how courteous and helpful the dealership staff is, etc.. Given that we use MSRP as the measure of price in the utility function (prior and post search), the match value also includes any deviation from MSRP due to, e.g., bargaining or a trade-in vehicle.

Consumer $i = 1, \ldots, N$ derives utility from product $j = 1, \ldots, J_i$ with utility u_{ij} given by

$$u_{ij} = \delta_{ij} + \varepsilon_{ij} \tag{1}$$

with

$$\delta_{ij} = \boldsymbol{x}_j \boldsymbol{\beta} + \eta_{ij},$$

$$\eta_{ij} \sim \mathrm{N}(0,1), \text{ and}$$

$$\varepsilon_{ij} \sim \mathrm{N}\left(0, \sigma\right).$$

We partition what the consumer knows and does not know prior to searching. Prior to search, δ_{ij} is known by the consumer. It is composed of a vector of observable vehicle characteristics $\boldsymbol{x}_j$, a vector of consumer preferences for those characteristics $\boldsymbol{\beta}$, and consumer i's product-specific idiosyncratic preferences η_{ij}, which are unobserved by the researcher but known by the consumer prior to search. The second component of utility, ε_{ij}, is the product fit or match value. The consumer knows the distributions of match values, $N\left(0, \sigma\right)$, but she is uncertain about her specific match value ε_{ij} and must search to discover it.

Search is performed sequentially and at a cost. Searching a product completely reveals the consumer's match value with that product, but does not reveal information about any other product. The cost of searching a product is parameterized as $c_{ij} = \exp\{\boldsymbol{d}_{ij}'\boldsymbol{\gamma}\}$. $\boldsymbol{d}$ is a vector of an intercept, the distance between the consumer's home location and the dealership of the searched product, an urban/rural indicator for the geographic location of the consumer, and an interaction between the latter two variables. $\boldsymbol{\gamma}$ is a parameter vector for the cost covariates. Consumers are assumed to have perfect recall and there is no cost for a consumer to revisit an already-searched product. Our data are conditional on search and purchase and we therefore do not model the outside option of not searching and/or not making a purchase.

5.2 Optimal Consumer Behavior

A consumer searches a product if the marginal benefit of doing so exceeds her marginal search cost. Since the match values ε_{ij} follow $N\left(0, \sigma\right)$, prior to search, the utilities u_{ij} follow $N\left(\delta_{ij}, \sigma\right)$. Define u_i^* as the highest utility among the searched products thus far.[13] Conditional on u_i^*, a

[13]For the first search, we set $u_i^* = -\infty$ since our data are conditional on search.

consumer's expected marginal benefit from searching product j is given by

$$\begin{aligned} B_{ij} &= \int_{u_i^*}^{\infty} (u_{ij} - u_i^*)\ f_{u_{ij}}(u_{ij})\ du_{ij} \\ &= \Pr\left[\varepsilon_{ij} > u_i^* - \delta_{ij}\right] \times \mathbb{E}\left[\varepsilon_{ij} - (u_i^* - \delta_{ij}) \,\middle|\, \varepsilon_{ij} > u_i^* - \delta_{ij}\right]. \end{aligned} \tag{2}$$

This marginal benefit is the probability that the realized utility for j, u_{ij}, exceeds the best realized utility among the already searched products, u_i^*, multiplied by the expected value of u_{ij} given $u_{ij} > u_i^*$. As shown in Equation (2), the marginal benefit depends on the MVSD through the integration over the utility distribution $f_{u_{ij}}(u_{ij})$. Holding everything else constant, a (symmetric) distribution with larger variance has more mass in the tails of the distribution and thus has both a higher probability that the next realized utility will exceed the currently best realized utility and a larger conditional expected value. Thus quantifying the MVSD informs us about the magnitude of the marginal benefit from searching.

Weitzman (1979) derived the rules for optimal behavior under sequential search. The rules involve "reservation utilities" z_{ij}, which are the utilities that equate the marginal cost and expected marginal benefit of search. Kim, Albuquerque, and Bronnenberg (2010) show that there is a closed-form solution for calculating reservation utilities under the assumption of normally distributed match-values, i.e.,

$$z_{ij} = \delta_{ij} + \zeta_{ij} \times \sigma \tag{3}$$

with ζ_{ij} coming from the implicit function $\frac{c_{ij}}{\sigma} = \phi(\zeta_{ij}) - \zeta_{ij} \times (1 - \Phi(\zeta_{ij}))$. We follow their approach.

Next, we formally state Weitzman's (1979) rules. Because the rank of the reservation utilities is a one-to-one mapping with the product index j, we cast the model using j as the order of the reservation utilities such that $j = 1$ is the product with the highest reservation utility for the consumer and $j = J_i$ for the product with the lowest reservation utility. Let us denote the number of searches made by a consumer as K_i. For notational simplicity, we drop the

consumer-specific subscript i hereafter.

Three rules govern consumer search and purchase behavior:

1. *Selection Rule:* A consumer searches products in a decreasing order of reservation utilities, i.e.,

$$z_1 \geq z_2 \geq \ldots \geq z_K \geq \max_{l>K} \{z_l\} . \tag{4}$$

2. *Stopping Rule:* A consumer stops searching when the maximum realized utility among the searched products is larger than the maximum reservation utility among the unsearched products, i.e.,

$$\max_{h \leq K} \{u_h\} \geq \max_{l>K} \{u_l\} . \tag{5}$$

   Equivalently, at each step during the search process, when a consumer decides to continue searching, the opposite of the stopping rule must hold, i.e.,

$$\max_{h<k} \{u_h\} < z_k \qquad \forall k = 2, \ldots, K. \tag{6}$$

3. *Choice Rule:* A consumer purchases the alternative with the highest realized utility among the searched ones, i.e.,

$$u_{j^*} = \arg\max_{h \leq K} \{u_h\} . \tag{7}$$

6 Estimation

6.1 Likelihood Function

The probability that a specific sequence of searches and an ultimate purchase are made by a consumer is the probability that each of the Weitzman (1979) rules holds at their respective steps in the consumer's shopping and purchase process, i.e.,

$$
\begin{aligned}
&L_i(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma; \boldsymbol{x}, \boldsymbol{d}) \\
&\quad = \text{Prob}\big[\text{Consumer } i \text{ selects } j = 1 \text{ on the first search,} \\
&\qquad\qquad \text{Continues to search and selects } j = k \text{ on the } k^{\text{th}} \text{ search for } k = 2, \ldots, K_i, \\
&\qquad\qquad \text{Stops searching after the } K_i^{\text{th}} \text{ search, and purchases } j_i^*\big] \\
&\quad = \int \mathbb{1}\big[z_{ij} \geq \max_{h<j}\{u_h\} \text{ for } j = 2, \ldots, K_i \bigcap z_{ij} = \arg\max_{k>j}\{z_{ik}\} \text{ for } j = 1, \ldots, K_i \bigcap \\
&\qquad\qquad \max_{h \leq K_i}\{u_{ih}\} \geq \max_{k>K_i}\{z_{ik}\} \bigcap u_{ij_i^*} = \arg\max_{h \leq K_i}\{u_{ih}\}\big]\, dF(\eta, \varepsilon).
\end{aligned} \tag{8}
$$

The model likelihood is the product of the N individual likelihoods, i.e.,

$$
L\left(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma; \boldsymbol{x}, \boldsymbol{d}\right) = \prod_{i=1}^{N} L_i\left(\boldsymbol{\beta}, \boldsymbol{\gamma}, \sigma; \boldsymbol{x}, \boldsymbol{d}\right). \tag{9}
$$

Note that we parametrize the MVSD as $\theta = \log(\sigma)$ in all estimations. Neither the search nor purchase probabilities can be expressed in closed form. We approximate the integrals in the likelihood function with averages using logit-smoothed accept-reject simulation. This simulated maximum likelihood estimation algorithm follows Train (2009). Details are provided in Appendix B.

6.2 Identification

The parameters to be estimated include the preference parameters $\boldsymbol{\beta}$, search cost parameters $\boldsymbol{\gamma}$, and the MVSD σ.

The preference parameters $\boldsymbol{\beta}$ are identified by purchase frequency, search order, and search frequency. In much the same way that the purchase decision among a set of products identifies preference parameters in a traditional discrete choice model, the purchase decision among searched products identifies preference parameters in a search model. In addition, because consumers search in order of decreasing reservation utilities $z_{ij} = \delta_{ij} + \zeta_{ij} \times \sigma$, holding everything else constant, products with higher δ_{ij} have higher reservation utility values and are searched earlier and more frequently.

With *no* exogenous search cost shifter, e.g., with fixed cost of search, only the ratio of $\frac{c}{\sigma}$ can be identified. The reason is that only the number of searches informs c and σ, but the search order does not. To see this, recall that consumers search in decreasing order of reservation utilities which are a function of search cost and the MVSD σ (see Equation (3)). If there is a common search cost across products and a common MVSD, then ζ_{ij} is the same for all products and the search order is entirely driven by δ_{ij}.

However, *with* an exogenous search cost shifter, i.e., distance to dealerships in our empirical application, both the search cost c_{ij} (and its governing parameters $\boldsymbol{\gamma}$) and the MVSD σ can be separately identified. The reason is that both the number of searches and the search order inform c and σ which, in turn, influence reservation utilities (see Equation (3)). For example, higher search costs yield lower ζ_{ij} values. Thus a consumer who must incur higher cost to search a product will assign it a lower reservation utility, rank it lower in the search order, and stop her search earlier (on average) than an otherwise identical consumer facing an identical searchable set of products but with lower search cost. Thus the observable differences in search cost due to the exogenous search cost shifter coupled with the patterns of search order and search length identify the search cost parameters $\boldsymbol{\gamma}$.

And lastly, a higher MVSD σ assigns a larger value to the second component of the reservation utility (see Equation (3)). As a result, the extent to which the search order is driven by $\zeta_{ij} \times \sigma$ rather than δ_{ij} identifies the MVSD. In addition, a larger MVSD yields a bigger probability of large positive ε draws than a smaller MVSD and thus results in earlier termination of search. Early search termination occurs because, holding everything else constant, a realized utility value will exceed the maximum reservation utility value among the unsearched products more often when the MVSD is large compared to when the MVSD is small. Thus, search length also identifies the MVSD.

6.3 Simulation Study

In this section, we describe the results from a simulation study in which we demonstrate that our estimation approach recovers preference and cost parameters as well as the MVSD parameter if an exogenous search cost shifter is included in the estimation.

For the simulation study, we generate 5,000 consumers each searching up to five brands. The number of searchable dealerships is drawn from a combination of a chi-squared distribution and an exponential distribution so as to generally mimic the observed searchable set sizes in the empirical application. These simulated searchable set sizes range from 1 to 40 with a median of 11 and a mean of 12. Observable characteristics include the brand as well as price and city mileage; the latter two are drawn from uniform distributions on the interval -2 to 2. Distances to dealerships are drawn from the absolute value of the sum of a chi-squared distribution with 8 degrees of freedom and a normal distribution with mean zero and standard deviation 16. Simulated distances to the searchable alternatives range from 0.01 miles to 102.4 miles with a median of 15.5 and a mean of 18.2 miles. For estimation, we use 1,000 draws from the distributions of the consumers' idiosyncratic preferences and match-value terms and smoothing parameters of (15, 15, 15, 5). We replicate each estimation 50 times and report mean coefficient estimates, their standard deviation, and average standard errors. The results are shown in Table 3 and show that, for a search model, the parameters are recovered well.

==========

Insert Table 3 about here

==========

7 Empirical Results

We show the results from three search model specifications in Table 4. In model (i), we fix the MVSD to 1 and estimate a fixed cost of search. In model (ii), we continue to fix the MVSD to 1, but allow search cost to vary with distance, an urban dummy, and an interaction between both variables. In model (iii), we allow the MVSD σ to enter the model as a parameter

to be estimated. In all three model specifications, we include the same set of covariates as in the reduced-form choice models in Section 4. Across the three model specifications, the utility parameter estimates are similar, generally sharing the same sign, similar magnitude, and significance with the exception of the brand intercepts for Nissan and Toyota. The utility parameter estimates also generally similar to the results from the reduced-form choice models presented in Table 2.

==========

Insert Table 4 about here

==========

Comparing models (i) and (ii) in Table 4, the addition of distance, an urban dummy, and an interaction between both variables to the search cost function leads to a much larger log-likelihood value (a change of over 2,400). This improvement is driven by a better ability of the model to fit the search order during the estimation process because consumers tend to visit dealerships close to their homes. When distance is not included in the model, two dealerships are equivalent from a modeling perspective if they offer a vehicle with the same characteristics even if one dealership is located next to the consumer's home and the other dealership is located 100 miles away. All search cost parameter estimates are positive and statistically significant. Comparing urban and rural consumers, urban consumers face higher cost than rural consumers, and that cost difference increases with distance. For example, the cost (in utils) to a rural consumer to visit a dealership 1 mile away is 1.34, while the cost to an urban consumer to do the same is 1.64 (23% higher). At a distance of 20 miles, the rural consumer's cost is 1.43 and the urban consumer's cost is 1.77 (28% higher).

In model (iii), our main model, we additionally also estimate the MVSD. The improvement of over 700 in the log-likelihood is large and shows that model (iii) fits the data better than model (ii). In addition, the correlation between $\log(\sigma)$ and the search cost intercept is 0.95. While this value is high, it is not high enough to suggest concern for identification. The correlations between all other estimated coefficients lie within ± 0.8 with most being smaller than ± 0.4.

The MVSD estimate is 8.16. This estimate is large compared to the common practice of

setting σ to 1. This large MVSD estimate indicates that consumers gain substantial benefits when visiting car dealerships. The MVSD has an estimated standard error of 0.31 and a 95% confidence interval (3.88, 17.19).[14] Note that the interval does not include 1. The parameter estimates for model (iii) also indicate that fixing the MVSD to 1 imposes a bias on the other parameters, in particular, on the cost parameters. More specifically, we find that the cost intercept is smaller, the urban intercept has a different sign, and the coefficient on distance is estimated to be almost three times as large in model (ii) as in model (iii). To demonstrate the differences visually, in Figure 6, we plot c_{ij}/σ for a rural consumer over distances ranging from 0 to 100 miles for all three search models. The c_{ij}/σ value is substantially different between model (ii) and (iii) for almost all positive distances. Thus we conclude that it is important to estimate the MVSD to correctly recover search cost parameters.

==========

Insert Figure 6 about here

==========

7.1 Model Fit

We evaluate the in-sample predictive performance of the estimated models using several measures. First, in Table 6, we show the product and brand hit rates for four models: full information multinomial logit model (model (b) in Table 2) and the three search models (models (i) to (iii) in Table 4). The "hit rate" is the percent of time that simulated behavior matches observed behavior. For example, the product hit rate is the average percent of time that the simulated-to-be-purchased product is the same as the actually purchased product.[15] The results show that all three search models considerably outperform the multinomial logit model (model (b) in Table 2). Among the three search models, added flexibility yields better hit rates. The improvement is the largest when search costs are modeled as a function of distance and

[14] We use the Delta method to calculate the standard error for σ.

[15] We calculate the hit rates as follows: for each consumer, we simulate 500 vectors of the random utility errors $\boldsymbol{\eta}_i$. For each simulation, we calculate the consumer's reservation utilities for each searchable product and allow search and choice to progress following Weitzman's rules. For each consumer and each simulation, we record the predicted search length, the searched product(s), and the chosen product.

the urban dummy (model (ii) in Table 4); the improvement in hit rates is marginal when the MVSD is estimated (model (iii) in Table 4).

==========

Insert Table 6 about here

==========

Additionally, in the bottom half of Table 6, we display the distribution of the number of searches consumers make (observed in our data) and the distributions of the number of searches that are predicted by the three search models from Table 4. While allowing search cost to vary with search cost shifters (model (ii)) enables the model to more precisely predict which products (dealerships) consumers search (see Section 7), model (ii) does not predict the distribution of the number of searches well: it overpredicts the proportion of consumers who search once and underpredicts the proportion of consumers who make two or more searches, i.e., the long tail of the distribution of the number of searches. While the maximum number of searches consumers conduct is 6+ in our data, model (ii) predicts that the maximum number of searches is three. Model (iii) – in which the MVSD is estimated – provides by far the best fit to the observed distribution of number of searches. In other words, permitting the MVSD to be freely estimated allows the model to fit the data better – in particular, the long tail of the distribution of the number of searches. Model (iii) predicts the maximum number of searches to be 5, while the maximum number of search in our data is 6+. To summarize, our results show that estimating the MVSD is crucial to correctly predict the distribution of the number of searches and especially its long tail.

7.2 Price Elasticities

We show the implied own-price elasticities for all three search models in Table 5. Elasticities are calculated by first simulating search and choice behavior from the fitted model. Then, separately for each brand, we increase the prices of available alternatives by 10% and re-simulate search and choice behavior. We repeat this exercise 500 times for each consumer. Elasticities are

computed as the average difference in % between the simulated outcomes with and without a 1% price increase.

==========

Insert Table 5 about here

==========

The own-price elasticity estimates range from -0.6 to -0.8. The elasticity estimates are influenced by the assumption that consumers make search and purchase decisions only among 5 brands and conditional on having already selected the type of vehicle they will purchase. While these assumptions were useful in easing the computational burden of fitting the model, they restrict the interpretability of the elasticity estimates because, for example, consumers may substitute to other brands and, at some level of price increase, consumers will not simply substitute to other vehicles in the same Edmunds type, but will substitute to other vehicle types.

7.3 Consumer Surplus

While the MVSD is a direct measure of the magnitude of potential benefits achievable through search (ex ante), we also calculate consumer surplus as a measure of achieved benefits through actual search behavior (ex post). For an individual consumer, the consumer surplus is defined as the utility from the chosen product net of all costs (price and search costs), i.e.,

$$CS_i = u_{ij^*} - \sum_{j=1}^{K_i} c_{ij}. \tag{10}$$

Note that, because our data are on conditional on search and on purchase and thus we do not model the no-search and no-purchase decisions, our consumer surplus estimates taking both prices and search costs into account can be negative (see, e.g., Moraga-Gonzalez, Sandor, and Wildenbeest 2017). To calculate consumer surplus for a consumer, we simulate 500 realizations from a model and average the resulting consumer surplus estimates (see Equation (10)) for each consumer. Figure 7 shows the distributions of consumer surplus from models (ii) and (iii). The

mean consumer surplus estimates for models (ii) and (iii) are -0.7 and -8.0, respectively. Thus normalizing the MVSD to 1 severely *overstates* consumer surplus.

==========

Insert Figure 7 about here

==========

To assess the relationship between consumer surplus and consumer characteristics, the consumer surplus from model (iii) is regressed on a set of consumer demographics. The results are displayed in Table 7. The included demographic variables explain a large proportion of the variation in the consumer surplus, i.e. 38.5%. We find significantly larger consumer surplus for urban and non-white consumers and significantly smaller consumer surplus for older consumers and consumers with kids.

==========

Insert Table 7 about here

==========

8 Counterfactual

In 2017, Hyundai initiated a program called "Hyundai Drive" in which a consumer can schedule a test drive of a Hyundai vehicle at a location that is convenient to her and a dealer will bring the car to that location. This program reduces travel costs to a dealer, but does not eliminate all search costs as it takes time to take the test drive and the mental costs of considering an additional alternative continue to be present. Here, we assess the potential market share changes to each brand from adopting at-home test drives.

To do so, we assume that brands adopt at-home test drives unilaterally. The brand offering the program permits the consumer to select a vehicle from that brand's closest dealership to be brought to the consumer for inspection and a test drive. Thus, the consumer incurs no travel costs to "visit" that dealership. To calculate the impact of adopting at-home test drives, we first simulate search and choice behavior for each consumer 500 times. Then, for each brand,

we set the distance of the closest dealership for each consumer to zero and re-simulate search and choice behavior. We then calculate the average difference in brand-choices from the base simulation to the counterfactual simulation. Results are presented in the top half of Table 8.

==========

Insert Table 8 about here

==========

Unilateral implementation of at-home tests drives permits brands to capture an additional 3.9–7.7 percentage points of the analyzed market. Because the five brands under study account for 60.1% of the Texas auto market, this corresponds to overall market share changes of 2.3–4.6 percentage points (holding constant the choices made by consumers who purchase other brands). This program yields the smallest market share increases for Chevrolet and Nissan and the largest market share increase for Toyota.

As a point of comparison, we conduct the same exercise with the estimates from model (ii) in which the MVSD is fixed to 1. We find the results to be roughly half as large as those from model (iii) in which the MVSD is estimated. Specifically, estimated increases in market shares under model (ii) range from 0.8 to 1.6 percentage points. Thus, failing to estimate MVSD yields underestimated market share increases from the unilateral implementation of at-home test drives.

9 Conclusions and Future Research

Prior to the Internet, store visits and the physical inspection of products played a prominent role during consumers' purchase process – especially for high-ticket durable goods such as automobiles. Information on prices and vehicle features was difficult to obtain except via a dealership visit. During the last 15 years, a wealth of online information sources have provided a low cost alternative to time consuming dealership visits. Therefore it is an important empirical question whether dealerships continue to provide (substantial) value to consumers in the Internet age.

To answer this question, it is necessary to separately quantify both the cost *and* the benefit

of search. We show that – with an exogenous search cost shifter – both the cost and the benefit of search can be identified. We estimate a sequential search model for product fit. Our empirical results show that the benefit provided by dealerships to consumers remains substantial. Consumers value visiting dealerships and learning about their liking (or dislike) of more experiental product attributes, i.e., product fit, such as how spacious and comfortable the car is, the layout of the driver dashboard, how the car drives (i.e., speeds up, handles the road, brakes), how courteous and helpful the dealership staff is, etc.. This finding points to a continuously important role of physical stores in the Internet age. The actions of online retailers in other product categories also corroborate this conclusion. Initially online-only retailers, such as Warby Parker and Bonobos, have recently opened brick and mortar stores to provide consumers with information that is difficult to communicate electronically.

Our paper is not without limitations and offers opportunities for future research. First, for vehicles not purchased, we observe that a consumer visited a dealership, but not which specific car the consumer was interested in. This is a limitation of our data. We therefore make the assumption that the consumer was searching for similar cars across different dealerships. Second, we do not model vehicle choice, only brand-dealer choice. It is left for future research to extend the model to also incorporate vehicle choice. And lastly, the benefit provided by dealers might be heterogenous varying with brand, dealership size or vehicle class. We leave it for future research to explore this type of heterogeneity.

References

Albuquerque, Paulo and Bart J Bronnenberg (2012), "Measuring the Impact of Negative Demand Shocks on Car Dealer Networks," *Marketing Science*, 31 (1), 4–23.

Berry, Steven, James Levinsohn, and Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63 (4), 841–890.

Bucklin, Randolph E, Sivaramakrishnan Siddarth, and Jorge M Silva-Risso (2008), "Distribution Intensity and New Car Choice," *Journal of Marketing Research*, 45 (4), 473–486.

Chen, Yuxin and Song Yao (2017), "Sequential Search with Refinement: Model and Application with Click-Stream Data," *Management Science*, 63 (12), 4345–4365.

De los Santos, Babur and Sergei Koulayev (2017), "Optimizing Click-Through in Online Rankings with Endogenous Search Refinement," *Marketing Science*, 36 (4), 542–564.

Dong, Xiaojing, Ilya Morozov, Stephan Seiler, and Liwen Hou (2019), "Estimation of Preference Heterogeneity in Markets with Costly Search," *Working paper*.

Honka, Elisabeth (2014), "Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry," *The RAND Journal of Economics*, 45 (4), 847–884.

Hui, Sam, Eric Bradlow, and Peter Fader (2009), "Testing Behavioral Hypotheses Using an Integrated Model of Grocery Store Shopping Path and Purchase Behavior," *Journal of Consumer Research*, 36 (3), 478–493.

Hui, Sam, Jeffrey Inman, Yanliu Huang, and Jacob Suher (2013), "The Effect of In-Store Travel Distance on Unplanned Spending: Applications to Mobile Promotion Strategies," *Journal of Marketing*, 77 (2), 1–16.

Jain, Aditya, Sanjog Misra, and Nils Rudi (2016), "Sales Assistance, Search and Purchase Decisions: An Analysis Using Retail Video Data," *Working Paper*.

Kim, Jun B, Paulo Albuquerque, and Bart J Bronnenberg (2010), "Online Demand Under Limited Consumer Search," *Marketing Science*, 29 (6), 1001–1023.

— (2017), "The Probit Choice Model under Sequential Search with an Application to Online Retailing," *Management Science*, 63 (11), 3911–3929.

Koulayev, Sergei (2014), "Search for Differentiated Products: Identification and Estimation," *The RAND Journal of Economics*, 45 (3), 553–575.

Moraga-Gonzalez, Jose, Zsolt Sandor, and Matthijs Wildenbeest (2018), "Consumer Search and Prices in the Automobile Market," *Working paper*.

Moraga-Gonzalez, Jose Luis, Zsolt Sandor, and Matthijs R Wildenbeest (2017), "Nonsequential Search Equilibrium with Search Cost Heterogeneity," *International Journal of Industrial Organization*, 50, 392–414.

Morozov, Ilya (2019), "Measuring Benefits from New Products in Markets with Information Frictions," *Working paper*.

Murry, Charles and Yiyi Zhou (2019), "Consumer Search and Automobile Dealer Co-Location," *Management Science*, forthcoming.

Nurski, Laura and Frank Verboven (2016), "Exclusive Dealing as a Barrier to Entry? Evidence from Automobiles," *The Review of Economic Studies*, 83 (3), 1156–1188.

Palazzolo, Mike and Fred Feinberg (2015), "Modeling Consideration Set Substitution," *Working paper*.

Ratchford, Brian T and Narasimhan Srinivasan (1993), "An Empirical Investigation of Returns to Search," *Marketing Science*, 12 (1), 73–87.

Ratchford, Brian T, Debabrata Talukdar, and Myung-Soo Lee (2007), "The Impact of the Internet on Consumers' Use of Information Sources for Automobiles: A Re-Inquiry," *Journal of Consumer Research*, 34 (1), 111–119.

Seiler, Stephan and Fabio Pinna (2017), "Estimating Search Benefits from Path-Tracking Data: Measurement and Determinants," *Marketing Science*, 36 (4), 565–589.

Seiler, Stephan and Song Yao (2017), "The Impact of Advertising along the Conversion Funnel," *Quantitative Marketing and Economics*, 15 (3), 241–278.

Train, Kenneth E (2009), *Discrete Choice Methods with Simulation*, Cambridge University Press, 2nd edition.

Ursu, Raluca (2018), "The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions," *Marketing Science*, 37 (4), 530–552.

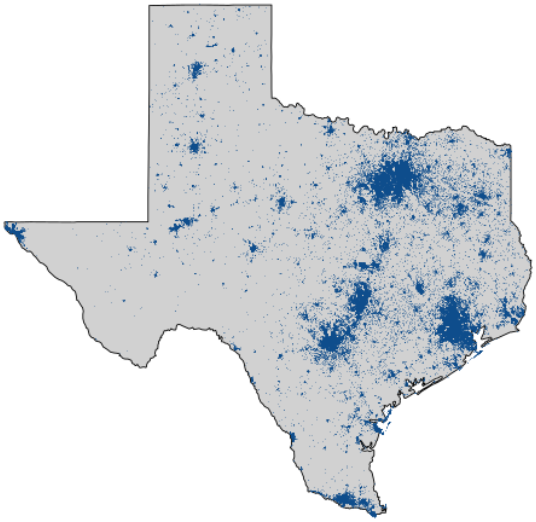
Weitzman, Martin L (1979), "Optimal Search for the Best Alternative," *Econometrica*, 47 (3), 641–654.

Yao, Song, Wenbo Wang, and Yuxin Chen (2017), "TV Channel Search and Commercial Breaks," *Journal of Marketing Research*, 54 (5), 671–686.

Figures and Tables

Figure 1: Home Locations

(a) Searcher Data
(from Safegraph)



(b) Buyer Data
(from Texas DMV)

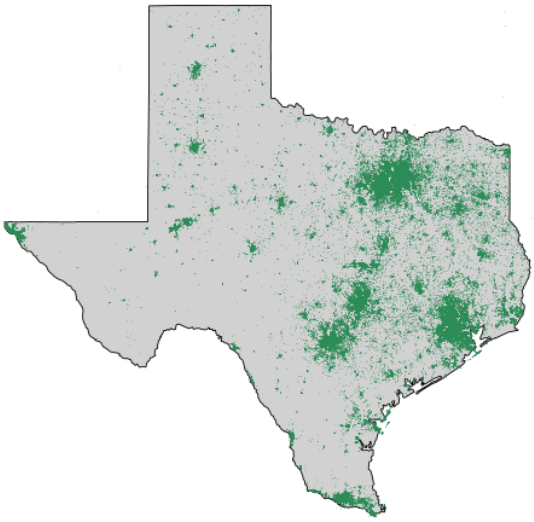
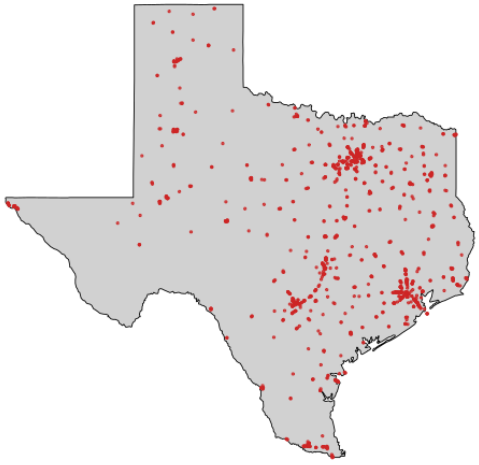


Figure 2: Dealership Locations

(a) Texas Dealerships



(b) Houston Dealerships

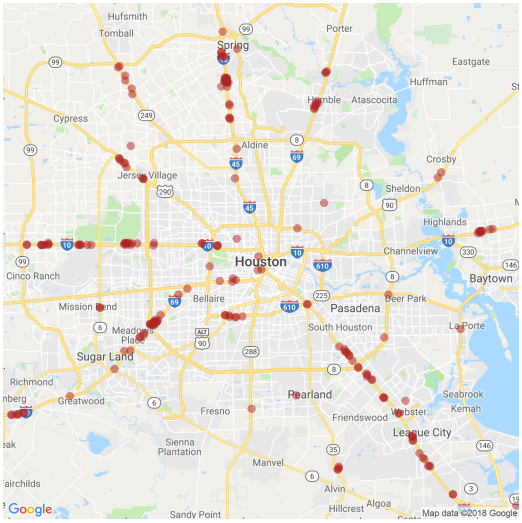


Figure 3: Distribution of Number of Searches

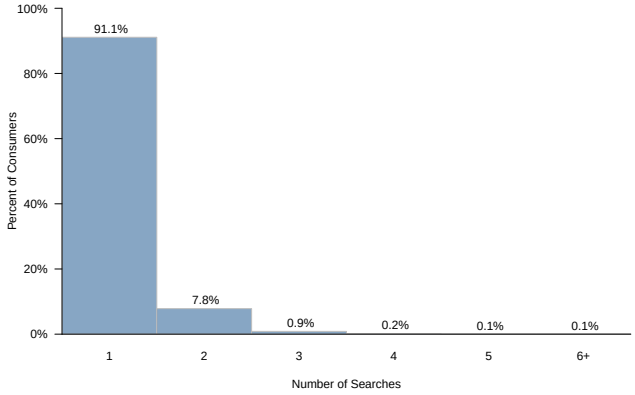
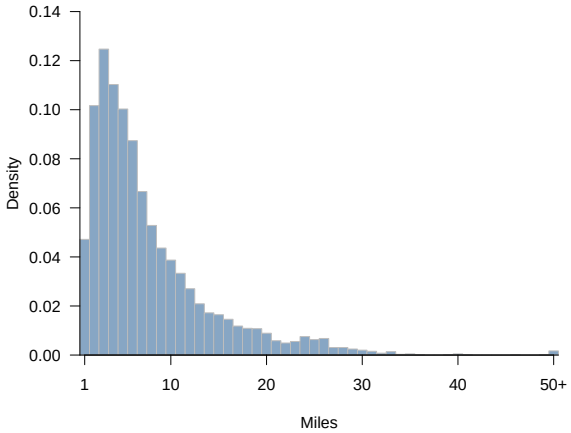


Figure 4: Consumers' Distances to Dealerships

(a) In Miles



(b) As Percentiles of Their Searchable Sets of Dealerships

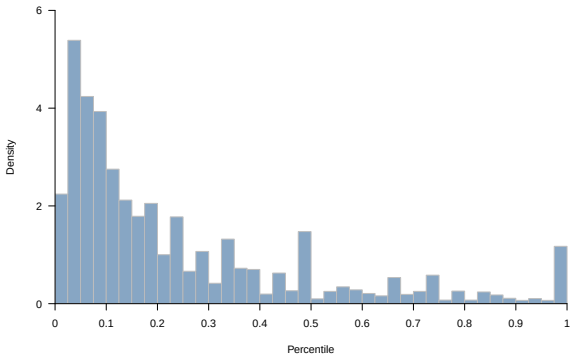


Figure 5: Miles Beyond Closest Same-Brand Dealership to Selling Dealership

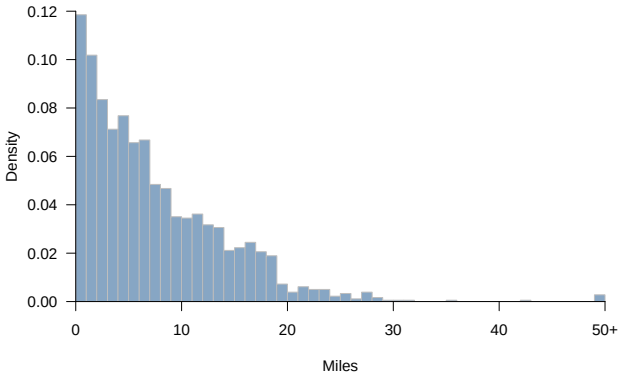


Figure 6: Comparison of Cost-Sigma Ratio vs Distance Across Fitted Search Models

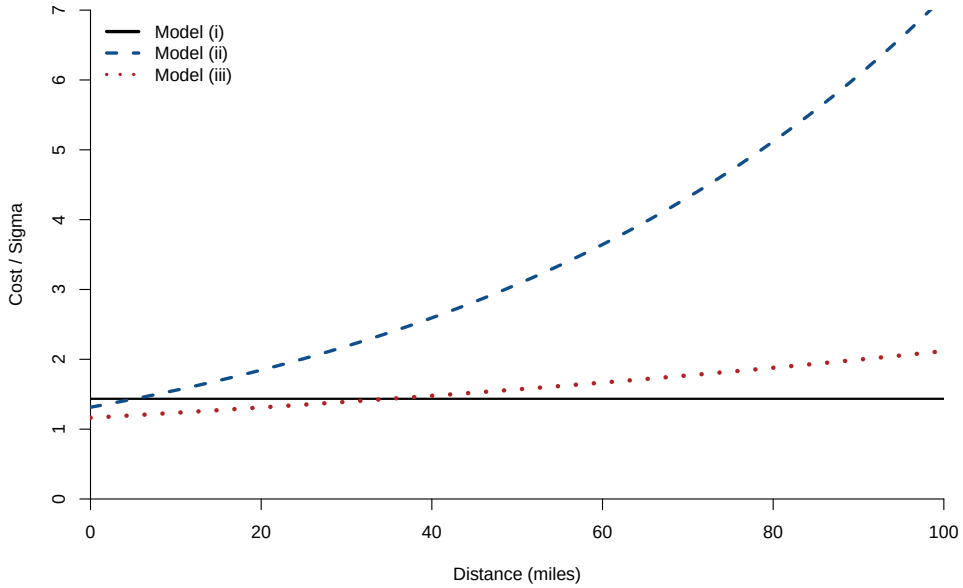


Figure 7: Consumer Surplus Based on Models (ii) and (iii)

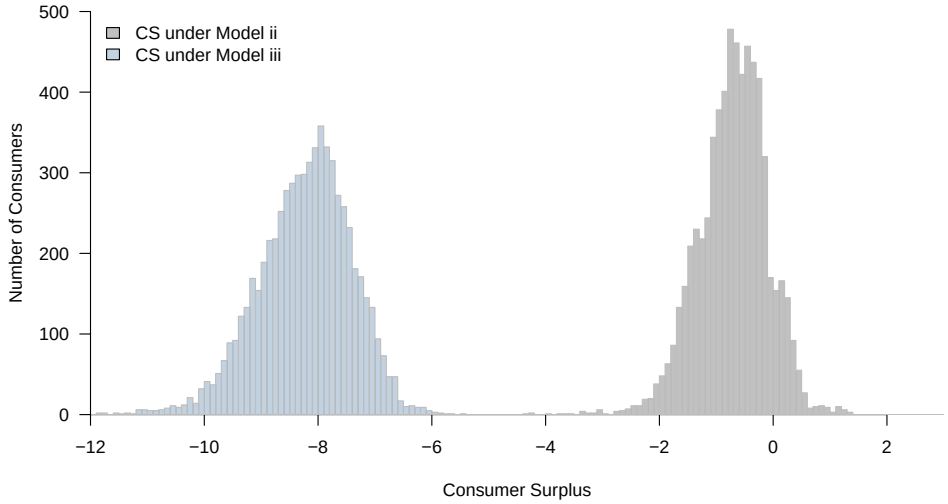


Table 1: Descriptive Statistics

	Number of Purchases in Analysis Sample	Number of Dealerships in Texas	Mean Values of Purchased Vehicles			
			MSRP	HP	Engine Size	City MPG
Chevrolet	1,222	212	$34,912	241	3.95	19.9
Ford	1,860	213	$36,670	259	3.31	18.7
Honda	981	56	$26,879	192	2.32	26.9
Nissan	652	64	$29,944	208	2.97	23.8
Toyota	1,796	79	$31,277	212	3.19	22.5

Table 2: Multinomial Logit Models

	Model (a) Full Information		Model (b) Full Information		Model (c) Limited Information	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
MSRP (in $10,000)	-0.388^{**}	0.034	-0.415^{**}	0.037	-0.441^{**}	0.169
Chevrolet	-1.786^{**}	0.069	-1.662^{**}	0.071	-2.658^{**}	0.450
Ford	-1.283^{**}	0.073	-1.104^{**}	0.075	-2.190^{**}	0.446
Nissan	-0.143	0.073	-0.118	0.076	-0.027	0.378
Toyota	0.069	0.055	0.188^{**}	0.056	-0.332	0.285
Large Vehicle x Chevrolet	1.349^{**}	0.115	1.375^{**}	0.119	2.017^{**}	0.601
Large Vehicle x Ford	1.824^{**}	0.110	1.922^{**}	0.114	2.478^{**}	0.611
Large Vehicle x Toyota	-0.010	0.098	-0.064	0.101	-0.538	0.494
Horsepower (in 100)	-1.013^{**}	0.059	-1.050^{**}	0.061	-0.989^{**}	0.303
Engine Size	0.284^{**}	0.034	0.324^{**}	0.036	0.388^{**}	0.186
MPG	0.085^{**}	0.010	0.095^{**}	0.011	0.027	0.053
MPG x Large Vehicle	0.016^{**}	0.006	0.015^{**}	0.006	0.034	0.031
Distance (in Miles)			-0.219^{**}	0.003		
Number of Consumers	6,511		6,511		6,511	
Number of Products	175,840		175,840		7,175	
Log-Likelihood	$-18,157$		$-13,071$		-381	
BIC	36,455		26,247		867	
Own-Price Elasticities						
Chevrolet	-1.18		-1.26		-0.88	
Ford	-1.21		-1.30		-0.92	
Honda	-1.00		-1.08		-0.54	
Nissan	-1.40		-1.50		-0.68	
Toyota	-1.20		-1.29		-0.81	

Note: Asterisks indicate statistical significance at the 95% confidence level. The base brand is Honda.

Table 3: Simulation Studiy

	True Values	Average Estimate	Standard Deviation of Estimates	Average Asymptotic Standard Error
Brand 1	0.2	0.178	0.088	0.038
Brand 2	0.4	0.374	0.064	0.045
Brand 3	0.6	0.540	0.066	0.042
Brand 4	0.8	0.714	0.080	0.038
MSRP	-0.5	-0.443	0.023	0.012
MPG	0.5	0.445	0.023	0.011
Cost Intercept	-2.0	-2.623	0.167	0.078
Distance	0.3	0.347	0.015	0.007
Log Sigma	0.69	0.439	0.059	0.024

Table 4: Search Model Results

	Model (i)		Model (ii)		Model (iii)	
	Coef.	Std. Err.	Coef.	Std. Err.	Coef.	Std. Err.
Preference Parameters						
MSRP (in $10,000)	-0.182^{**}	0.018	-0.186^{**}	0.019	-0.202^{**}	0.020
Chevrolet	-0.807^{**}	0.035	-0.787^{**}	0.034	-0.803^{**}	0.038
Ford	-0.602^{**}	0.037	-0.579^{**}	0.039	-0.578^{**}	0.040
Nissan	-0.068^{**}	0.038	-0.063	0.040	0.073	0.039
Toyota	0.047	0.031	0.072^{**}	0.030	0.097^{**}	0.033
Large Vehicle x Chevrolet	0.611^{**}	0.059	0.611^{**}	0.060	0.670^{**}	0.066
Large Vehicle x Ford	0.886^{**}	0.056	0.823^{**}	0.060	1.020^{**}	0.064
Large Vehicle x Toyota	0.028	0.053	0.011	0.054	0.046	0.058
Horsepower (in 100)	-0.473^{**}	0.030	-0.501^{**}	0.031	-0.518^{**}	0.034
Engine Size	0.132^{**}	0.019	0.146^{**}	0.018	0.151^{**}	0.021
MPG	0.042^{**}	0.005	0.042^{**}	0.006	0.044^{**}	0.006
MPG x Large Vehicle	0.005	0.003	0.006	0.003	0.004	0.004
Log Search Cost Parameters						
Intercept	0.361^{**}	0.018	0.273^{**}	0.043	2.250^{**}	0.040
Urban			0.203^{**}	0.048	-0.055^{**}	0.015
Distance (in Miles)			0.017^{**}	0.001	0.006^{**}	0.000
Urban x Distance			0.002^{**}	0.001	0.003^{**}	0.000
Match-Value Standard Deviation						
Sigma	1.00		1.00		8.16	0.310
Number of Consumers	6,511		6,511		6,511	
Number of Products	175,840		175,840		175,840	
Log-Likelihood	$-20,765$		$-18,362$		$-17,652$	
BIC	41,636		36,829		35,409	

Note: Asterisks indicate statistical significance at the 95% confidence level. The base brand is Honda. Optimization via BFGS with relative tolerance convergence criterion set to 1e-6. Simulated likelihood using Q=1,000 independent random draws of the random utility error and the match-value distribution. Search costs are parameterized as $c_{ij} = \exp\{\gamma_0 + \gamma_1 \text{Urban} + \gamma_2 \text{Distance} + \gamma_3 \text{Urban} \times \text{Distance}\}$.

Table 5: Own-Price Elasticities

	Model (i)	Model (ii)	Model (iii)
Chevrolet	-0.79	-0.78	-0.75
Ford	-0.69	-0.68	-0.67
Honda	-0.52	-0.58	-0.56
Nissan	-0.72	-0.72	-0.71
Toyota	-0.62	-0.63	-0.60

Table 6: In-Sample Predictive Performance

Hit Rates in %	Model (b) from Table 2	Model (i) from Table 4	Model (ii) from Table 4	Model (iii) from Table 4
Product Hit Rate	8.4	10.3	19.3	19.4
Brand Hit Rate	30.0	39.4	39.9	39.9

Number of Searches	Observed (Data)	Model (i) from Table 4 (Predicted)	Model (ii) from Table 4 (Predicted)	Model (iii) from Table 4 (Predicted)
1	91.1%	96.2%	98.7%	88.5%
2	7.8%	3.7%	1.3%	10.3%
3	0.9%	0.1%	0.1%	1.1%
4	0.2%	0.1%	-	0.1%
5	0.1%	-	-	0.1%
6+	0.1%	-	-	-

Table 7: Consumer Surplus and Demographics

	(i)	(ii)
Urban Indicator	0.836**	0.717**
	(0.022)	(0.029)
Median Age 30 - 40	−0.011	−0.063**
	(0.024)	(0.024)
Median Age 40 +	−0.103**	−0.135**
	(0.030)	(0.030)
Percent Male	−0.053	−0.048
	(0.181)	(0.172)
log(Number of Kids)	−0.069**	−0.081**
	(0.010)	(0.010)
Percent with College Degree	0.236**	−0.031
	(0.052)	(0.052)
log(Income in $1,000)	0.015**	0.001
	(0.008)	(0.007)
Unemployment Rate	0.233	0.167
	(0.210)	(0.204)
Percent Black	0.670**	0.329**
	(0.070)	(0.073)
Percent Other Race	0.960**	0.684**
	(0.081)	(0.082)
Intercept	−8.862**	−8.137**
	(0.130)	(0.136)
Month Fixed Effects	No	Yes
County Fixed Effects	No	Yes
Number of Observations	6,511	6,511
R^2	0.276	0.385

Note: Standard errors reported in parentheses. Asterisks indicate statistical significance at the 95% confidence level.

Table 8: Market Shares Under Unilateral Adoption of At-Home Test Drives

	Market Shares Based on Model (iii)		
Brand Adopting At-Home Test Drives	No At-Home Test Drives	With At-Home Test Drives	Change in Percent
Honda	0.144	0.194	34.5%
Chevrolet	0.200	0.247	23.3%
Ford	0.286	0.338	17.9%
Nissan	0.093	0.132	41.3%
Toyota	0.276	0.353	27.8%

	Market Shares Based on Model (iii)		
Brand Adopting At-Home Test Drives	No At-Home Test Drives	With At-Home Test Drives	Change in Percent
Honda	0.148	0.164	11.3%
Chevrolet	0.198	0.210	6.4%
Ford	0.283	0.298	5.2%
Nissan	0.096	0.109	13.7%
Toyota	0.275	0.302	9.8%

Appendix A: Details on Analysis Data Set Construction

Search and Purchase Data Merging Algorithm

First, recall that the home location of a mobile device user in the search data is estimated by Safegraph and provided as a geohash-8 while the purchase data from the Texas DMV provides the registrant's street address. We geocode both sets of location information as latitude and longitude coordinates. For the search data, we use the center of the geohash; for the purchase data, we use coordinates from the GoogleMaps API.

We then require that the two home locations are in close proximity. To do so, we calculate the distance between each of the 154,000 mobile device home locations in the dealership visits data and each of the 264,000 unique registrant's home addresses in the purchase data using the Haversine great-circle distance measure, which is the shortest distance between two points on a sphere. Due to computer memory constraints, for each mobile device home location, we retain up to a maximum of 50 "potential merges" (i.e., registered vehicles) where the distance between the home locations is one-fifth of a mile or closer. In short, we find up to 50 vehicle sales that could potentially be the result of each observed search sequence.

Next, we evaluate potential merges in increasing order of distance. For each potential merge, we assess whether the mobile device had visited the selling dealership and whether that visit occurred on or before the vehicle registration date. If so, we merge that search sequence with that purchased vehicle, and we discard all other potential merges involving that search sequence or vehicle. If, instead, the potential merge involved home locations that were close in proximity but where the mobile device had not visited the selling dealership on or before the registration date, we discard that potential merge. We then evaluate the next (in terms of shortest distance) potential merge. The process proceeds until all potential merges are evaluated and either accepted or discarded.

The Searchable Set of Dealerships

We limit the searchable set of dealerships to any dealership within a particular radius of the consumer. To define the radius for each consumer, we first partition consumers into urban and rural consumers. Urban consumers are those who live within the city limits of any major city (where major is defined as a city with a population exceeding 100,000 inhabitants) and rural consumers live outside of these major cities. For urban consumers in each major city and for rural consumers in each county (excluding inhabitants of the major cities in that county), we assign a radius as twice the 70th percentile of the distance between the home location of a car buyer and the selling dealership. We calculate these distance distributions from the original data set containing 195,000 new vehicle DMV registrations.

For example, suppose a consumer lives in the rural part of Dallas County and our data contain 10,000 new vehicle registrations with the Texas DMV from rural residents of Dallas County. We calculate the 70th percentile of the distance between those 10,000 residents' home and the selling dealerships. Suppose this value is 30 miles. Then for each consumer in the analysis sample who lives in the rural part of Dallas County, we assign a radius of 60 miles. Therefore, the searchable set of dealerships for rural Dallas County residents includes any dealership within 60 miles of the consumer's home location. Note that this set of dealerships may differ across rural residents of Dallas County if some dealerships are within 60 miles of some consumers' homes, but others are not.

The Searchable Set of Vehicles

For each consumer and each searchable dealership identified by the process outlined in Section 3.2.1, we assume that there is one most preferred vehicle that could be searched. That is, if the consumer were to visit a particular dealership, she would have searched only her most preferred vehicle at that dealership. To continue the example of the rural consumer in Dallas County from the previous section, suppose there are 14 dealerships selling the five focal brands within the consumer's 60 mile radius. Then this consumer faces 14 vehicles that she could search. It is

important to note that the most preferred vehicle at a particular dealership is consumer-specific: if the same dealership is searchable by multiple consumers, each consumer has her own most preferred vehicle at that dealership.

To identify the consumer-specific, most preferred vehicle at each dealership, we impose two criteria: the potentially searched vehicle must be in that dealership's inventory and it must be of the same Edmunds type as the vehicle ultimately purchased by the consumer. More specifically, "inventory" is defined as any model sold by that dealership during the 16 months for which the Texas DMV registration data are available.[16] This criteria enforces that consumers can only search for a model at a particular dealership if that dealership has ever carried and sold that model. In addition to being in inventory, the searched vehicle must be of the same Edmunds type as the vehicle ultimately purchased by the consumer. This assumption is equivalent to assuming that consumers search for a specific vehicle conditional on having decided the type of vehicle they would like to buy.

To continue the example, suppose the rural Dallas County consumer is observed to purchase a 2017 Nissan Versa Note and that her searchable set of dealerships includes XYZ Chevrolet. According to Edmunds, two Chevrolet vehicles are of the same type (type: "Extra-Small Hatchback") as the Nissan Versa Note: the Chevrolet Spark and the Chevrolet Sonic. Suppose that the Texas DMV registrations data show that the XYZ Chevrolet dealership sold zero 2017 Chevrolet Sparks and nineteen 2017 Chevrolet Sonics. Then the searchable vehicle for our example consumer at XYZ Chevrolet is assumed to be the 2017 Chevrolet Sonic because (i) it has the same model year and the same Edmunds type as the purchased vehicle and because (ii) it was in inventory at the searched dealership.[17]

Finally, we must also infer the set of vehicle characteristics for each searchable vehicle. To do so, we compute the median values (for numeric characteristics) and modal values (for categorical characteristics) by model and dealership. Thus if the rural Dallas County consumer

[16] A model is a particular combination of model year, make, and model. For example, one model is a 2017 Toyota Corolla.

[17] In the event that multiple vehicles meet these criteria, the vehicle that is ranked highest according to Edmunds.com is selected as the most-preferred vehicle.

could search a 2017 Chevrolet Sonic at two different Chevrolet dealers (say ABC Chevrolet and XYZ Chevrolet), the city mileage of the 2017 Sonic at ABC Chevrolet depends on the city mileage of the other 2017 Sonics sold by ABC Chevrolet and differs from the city mileage calculated for the 2017 Sonic at XYZ Chevrolet.

Appendix B: Kernel-Smoothed Frequency Simulator

Neither the search nor purchase probabilities can be expressed in closed form. We approximate the integrals in the likelihood function with averages using logit-smoothed accept-reject simulation. This simulated maximum likelihood estimation algorithm follows Train (2009) and is outlined in the following steps:

1. Do steps 2–4 for each consumer $i = 1, \ldots, N$

2. Draw Q values from the density $f_\varepsilon(\varepsilon)$ for each of the K_i searches for a total of $Q \cdot K_i$ draws and draw Q values from the density $f_\eta(\eta)$ for a total of $Q \cdot J_i$ draws

3. For each set of random draws:

   (a) Compute $\nu_{1,j} = z_{ij} - \max_{h \leq j} \{u_{ih}\}$ for $j = 2, \ldots, K_i$

   (b) Compute $\nu_{2,j} = z_{ij} - \max_{k > j} \{z_{ik}\}$ for $j = 1, \ldots, K_i$

   (c) Compute $\nu_3 = \max_{h \leq K_i} \{u_{ih}\} - \max_{k > j} \{z_{ik}\}$

   (d) Compute $\nu_4 = u_{ij_i^*} - \max_{h \leq K_i} \{u_{ih}\}$ for the chosen j_i^*

   (e) Compute the simulated individual likelihood given one set of draws:

$$\tilde{L}_i^q = \left(1 + \sum_{j=2}^{K_i} \mathrm{e}^{-\lambda_1 \nu_{1,j}} + \sum_{j=1}^{K_i} \mathrm{e}^{-\lambda_2 \nu_{2,j}} + \mathrm{e}^{-\lambda_3 \nu_3} + \mathrm{e}^{-\lambda_4 \nu_4}\right)^{-1}$$

4. Average the draw-specific simulated individual likelihoods over the draws:

$$\tilde{L}_i = \frac{1}{Q} \sum_{q=1}^{Q} \tilde{L}_i^q$$

5. Take logs and aggregate the individual-specific likelihoods over individuals to form the total simulated log-likelihood function:

$$\log\left(\tilde{L}\right) = \sum_{i=1}^{N} \log\left(\tilde{L}_i\right)$$

Computing time when fitting the sequential search model is directly proportional to the number of draws used to approximate the integrals in the likelihood function. However, a large number of draws is necessary to achieve a good approximation. When calculating each individual likelihood, we therefore use $1,000$ draws for η_{ij} and $1,000$ draws for ε_{ij}. By comparison, Kim, Albuquerque, and Bronnenberg (2017) use 40 draws when testing the kernel-smoothed AR estimation approach; Honka (2014) and Ursu (2018) use 50 draws to estimate their search models with the same approach.