This article was downloaded by: [136.167.230.79] On: 07 December 2022, At: 07:13
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA

Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Consumer Search and Automobile Dealer Colocation

Charles Murry, Yiyi Zhou

To cite this article:
Charles Murry, Yiyi Zhou (2020) Consumer Search and Automobile Dealer Colocation. Management Science 66(5):1909-1934.
https://doi.org/10.1287/mnsc.2019.3307

Full terms and conditions of use: https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2019, INFORMS

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

informs
http://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 66, No. 5, May 2020, pp. 1909–1934
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

Consumer Search and Automobile Dealer Colocation

Charles Murry,[a] **Yiyi Zhou**[b]

[a] Department of Economics, Boston College, Chestnut Hill, Massachusetts 02467; [b] Department of Economics and College of Business, Stony Brook University, Stony Brook, New York 11794

Contact: charles.murry@bc.edu, http://orcid.org/0000-0001-8405-2868 (CM); yiyi.zhou@stonybrook.edu, http://orcid.org/0000-0001-9052-127X (YZ)

Received: July 20, 2017
Revised: June 29, 2018; November 16, 2018
Accepted: December 26, 2018
Published Online in Articles in Advance: September 23, 2019

https://doi.org/10.1287/mnsc.2019.3307

Copyright: © 2019 INFORMS

Abstract. Retailers colocate with rivals to take advantage of economies of agglomeration even though colocation implies greater competition. Using data on all new car transactions registered in Ohio from 2007 to 2014, we estimate a structural model of consumer search for spatially differentiated products that explicitly captures the agglomeration and competition effects of retail colocation. Search frictions generate an average of $333 per car in dealer markups. Agglomeration implies that dealer closures could harm incumbent colocated dealers, even though the incumbent dealers would face less competition. Our results inform the recent policy debate surrounding the massive downsizing of car retail networks and highlight the role of contagion in brick-and-mortar retailing.

History: Accepted by Matthew Shum, marketing.

Keywords: car dealers • retail exit • agglomeration

1. Introduction

Economists and marketers have long sought to understand the location decisions of retail stores and the effects of these decisions on industry profits and consumer welfare. Particular attention has been paid to why retailers tend to locate near each other even though colocation implies fiercer competition. A classic explanation is related to consumers' limited information, for example, as in Stahl (1982) and Wolinsky (1983). If consumers must engage in costly search to resolve informational problems before purchase, then they are more likely to search areas where there is a concentration of stores to limit their search costs. Consumer search creates an agglomeration benefit for retailers to colocate. However, colocating with rival stores could intensify price competition, potentially outweighing the agglomeration benefit from colocation.

In this paper, we empirically study the underlying demand-driven reasons of retail colocation in the new car retail industry. In particular, we examine the extent to which the agglomeration and competition effects are related to consumer search. To do this, we present a structural model of consumer search for spatially differentiated products. In the model, consumers have limited information about their demand for particular new cars, but can learn their exact valuation of a car by paying a search cost to visit a geographic cluster of new car dealers. This search cost depends on the cluster size and the distance between the consumer's residence and the center of the dealer cluster. After visiting a dealer cluster, consumers learn their exact utility from all cars sold in the geographic cluster.[1] Consumers choose which clusters to search, and conditional on a search set, they choose their best option among the searched products. The model gives rise to a positive or negative agglomeration effect of dealer colocation: clusters with more dealers can be more or less likely to be searched by consumers. Whether a positive agglomeration effect exists and whether the agglomeration effect dominates the competition effect are empirical questions and depend on model primitives that we estimate.

To estimate the model, we construct a data set that includes all new car sales registered in Ohio from 2007 to 2014. For each registration, we know the make (brand), model, model year, transaction price, transaction date, identity and location of the selling dealer, and zip code of the buyer's residence. The detailed spatial nature of the data allows us to accurately capture the spatial demand substitution patterns that underlie the demand-driven motives of retail colocation. Our estimates imply that dealer colocation has a positive agglomeration effect on consumer search. Our estimation results also suggest that the car price needs to be $45 lower on average to compensate consumers from traveling an additional mile to search a dealer cluster, although consumers need to be compensated less for traveling to clusters with more available products. We also find that the consumer search frictions generate $333 on average in markups per car for car dealers, and there is significant heterogeneity in these information rents.[2]

Although our estimates suggest a positive agglomeration effect of colocation on the demand for

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

vehicles, it is not clear whether dealers would prefer to locate near each other because having closer rivals also implies fiercer price competition. To understand the balance between agglomeration and competition, we decompose these effects in a counterfactual scenario where two car brands, Pontiac and Saturn, are each shut down (before their actual dissolutions). Colocated dealers are harmed through the agglomeration effect because fewer consumers will find that location desirable to search. However, dealer closures help colocated firms through the typical competition effect: fewer close rivals implies less price competition. To quantify the competition effect, we simulate what would happen after closures when consumers are forced to choose another car within the same geographic cluster but not to reoptimize their search sets. To quantify the agglomeration effect, we simulate the equilibrium outcomes when consumers are allowed to reoptimize their search sets but not to adjust their relative choice probabilities within the cluster.

In the cases of both Pontiac and Saturn closures, our estimates suggest that the agglomeration effect harms neighboring dealers. However, we find that in the case of Pontiac closures, the competition effect dominates the agglomeration effect, and nearby dealers are better off. In the case of Saturn closures, we conclude the opposite: that nearby dealers are worse off after Saturn dealership closures. Our findings have implications for policy makers and managers. In a case when the agglomeration effect dominates the competition effect, one store's exit will make other colocated stores worse off and may result in a closure spiral, harming the local economy. In this regard, government bailout may be needed to prevent such closure spirals. In this case, moreover, it is not wise for managers to take aggressive strategies that aim to drive neighboring competitors out of business, and it is also profitable to locate near other stores.

The new car retail industry is an ideal setting for such a study. First, it is a large industry with ubiquitous retail colocation. For example, in Ohio (the geographic region of the data we use in our study), more than 85% of new car dealers are located within half a mile of a rival dealer. Along with the results of our structural model, we present evidence from car purchase patterns that illustrate the importance of retail colocation and consumer search in this industry. Second, the automobile industry has experienced massive retail closures over the past half century. In particular, car dealer closures became a debated policy question during the most recent U.S. financial crisis. During the Troubled Asset Relief Program (TARP), which provided billions of dollars in aid to the U.S. automobile manufacturing industry, Congress, state dealer associations, and car manufacturers battled over the legalities and policy implications of proposals by General Motors (GM) and Chrysler to close thousands of franchised new car dealers. However, little is known empirically about the local microeconomic ramifications of retail closures. Our study helps fill the gap in the literature and sheds some light on this policy debate.

Most prior studies on retail colocation have inferred the agglomeration–competition trade-off from the revealed entry and location decisions of profit-maximizing retailers. Some of these studies have found that the competition effect dominates, for example, Seim (2006), Jia (2008), and Zhu and Singh (2009). As a result, retailers prefer to differentiate locations, and the implication is that a retailer's exit would benefit other incumbents. On the other hand, Vitorino (2012) finds evidence that the agglomeration effect dominates in shopping malls, Ellickson et al. (2013) find that there exists a net positive agglomeration effect in the big-box retail industry, and Datta and Sudhir (2013) find evidence of the agglomeration benefits of retail colocation using variation in zoning laws. In our paper, we present and estimate a structural model of consumer search for spatially differentiated products, thus specifically modeling the demand-side mechanism for the agglomeration benefits of retail colocation.

Our paper also contributes to the burgeoning literature on empirically understanding limited consumer information. For example, Sovinsky Goeree (2008) estimates a model where advertising affects the choice sets of consumers and show how limited consumer information contributes to firms' market power. Much of this literature has used costly consumer search to explain limited consumer information, including Mehta et al. (2003), Hong and Shum (2006), Wildenbeest (2011), Seiler (2013), and Honka (2014), and others. Our study differs from previous studies by explicitly modeling the spatial colocation of products and showing how consumer search implies an agglomeration effect of retail colocation. Our model is closely related to the prior literature on consumer search. Both De los Santos et al. (2012) and Honka and Chintagunta (2016) analyze data on consumer actual searches for consumption goods and find that the simultaneous search strategy matches their data better than the sequential search strategy. Without appropriate data to test different search models, we follow Honka (2014) and Moraga-González et al. (2015) by assuming that consumers adopt the simultaneous search strategy to search for new cars.

Our model closely follows that of Moraga-González et al. (2015), who develop and estimate a structural model of consumer search for new cars in the Netherlands.[3] There are two main differences between our

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

setup and theirs: (i) in our model, search occurs at the level of a geographic dealer cluster instead of a single dealer, and (ii) we estimate the variance of search-cost shock, which is crucial to quantifying agglomeration and plays a similar role to the nesting parameter in a nested logit framework. On the first point, we validate this assumption by estimating the importance of cluster size in the search cost and presenting descriptive statistics showing that dealer colocation helps explain purchase behavior. On the second point, we also allow for agglomeration at the cluster level to be zero or negative by including a term that captures cluster size. Also, estimating the variance of search-cost shocks is crucial for understanding the agglomeration effect, as the value of this variance parameter determines whether the agglomeration effect dominates the competition effect. An additional point of contrast between our paper and theirs is that we use individual purchase data, so we observe precisely how far consumers travel to purchase cars. Accordingly, we use Goolsbee and Petrin's (2004) two-step simulated maximum likelihood method to estimate the nonlinear parameters of the model.

Finally, our counterfactual analysis contributes to the literature on retail closure and brand termination. Benmelech et al. (2014) use data across retail industries to estimate the effect of closures due to chain-level financial problems on the closure decisions of close-by incumbent retail outlets. They find that nearby retail outlets are more likely to close after a retailer's closure, which they interpret as evidence of agglomeration effects of closures of bankrupt firms' stores on nonbankrupt incumbent stores. Ozturk et al. (2016) examine the effect of Chrysler dealer closings on the prices of nearby dealers using a national sample of new car transactions in a differences-in-differences framework. They find that after the closures, nearby dealers experience a lower price increase than distant dealers, which is evidence that the agglomeration effect exists in car buying. Different from the existing studies in this field, we develop a structural model in which either positive or negative agglomeration effects can present and quantify the effects of closures by decomposing the competition and agglomeration effects.

2. Data and New Car Retail Industry

We combine several data sets for our analysis. Our primary data include detailed records of all new vehicle transactions that were registered in Ohio from 2007 to 2014. The second data source provides general information on characteristics of all vehicles sold during this time period, and the third data source provides information on all new car dealerships in Ohio. We also use American Community Survey data from the U.S. Census to measure the local demographics at the zip-code level.

2.1. Data Description

The primary data were obtained from the Ohio Bureau of Motor Vehicles and consist of all new vehicle transactions initially registered in Ohio from 2007 through 2014. For each transaction, we know the brand (car make), model, model year, transaction price, and transaction date. We also know the identity of the selling dealer and the five-digit zip code of the buyer.

Throughout this paper, we define a product by car model and model year, for example, Toyota Camry 2010 model. In total, the data include 1,892 products. We make a number of sample selection decisions for the raw data. First, we remove all commercial vehicles, motorcycles, trailers, and consumer pickup trucks.[4] Second, we drop the products with average prices above $70,000 and small dealers with annual sales below 100 units. These account for around 4% transactions over the eight years. In the end, we are left with more than 2.5 million new car transactions of 34 brands sold by 970 dealers. We define a dealer as a particular brand franchise at a particular location.

We merge the transactions data with information on car characteristics from Edmunds.com. We use three commonly used characteristics to define the characteristics of a car: acceleration (horse power divided by weight), miles per dollar (miles per gallon divided by dollars per gallon), and size (length multiplied by width and height). We also dichotomize cars into luxury and nonluxury brands and U.S. and non-U.S. brands.[5] Table 1 presents the descriptive statistics of the 2.5 million transactions included in our sample. The average transaction price is $28,253, with a standard deviation of $11,318. Among all new cars registered in Ohio over these eight years, 11% of them are luxury brands and 45% are U.S. brands.

Table 2 reports the demographics at the zip-code level in Ohio from 2007 to 2014 that we will use in our following analysis, including the median household income in thousands of dollars, share of college degree or higher, number of households, average household size, share of households with children, and the share of urban area. Overall, there is substantial variation for all variables across zip codes. There is also a substantial upward trend in income over our sample.

2.2. New Car Retail Industry

Table 3 presents the statistics of the top 10 brands, including the market share in terms of the total units sold, number of active dealers in Ohio from 2007 to 2014, share of single-brand dealerships, and the average transaction price. The top 4 brands, Honda, Chevrolet, Ford, and Toyota, accounted for half of the new car sales in Ohio during this time period. Of the

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

Table 1. Summary Statistics of New Car Transactions in Ohio from 2007 to 2014

Variable	Mean	SD	Min	Median	Max
Transaction price	28,253	11,318	15,312	25,794	65,965
Acceleration	5.91	1.57	2.85	5.57	11.08
Car size	0.84	0.14	0.52	0.81	1.50
Miles per dollar	8.07	1.72	3.63	7.90	17.46
Luxury brand	0.11	0.31	0	0	1
U.S. brand	0.45	0.50	0	0	1

Notes. Our selected sample includes 2,503,734 new car sales in Ohio from 2007 to 2014, accounting for 101,371 product-dealer-year combinations. Acceleration is the ratio of horse power over the curb weight. Miles per dollar is the average of highway and local miles per gallon divided by gasoline price per gallon in dollars. Car size is the multiplication of car length, width, and height, measured in 100 inches. SD, standard deviation.

970 dealers, 611 (63%) sold only one brand. The share of single-brand dealers varies significantly across brands. For example, among all 55 dealers that sold Honda models, 46 of them (84%) sold Honda cars exclusively. In contrast, only 4 out of 119 dealers selling Jeep were single branded. Along with geographic colocation, multiple-brand dealers are an important aspect of agglomeration in this industry.

Table 4 presents the number of active new car dealers, the total units sold, and the average transaction price in Ohio from 2007 to 2014. The new car sales dropped dramatically during the financial crisis, and this drop was particularly severe for U.S. brands. Along with the sales drop, the number of active dealerships dropped from 759 in 2007 to 635 in 2010 (a 16% drop). In particular, the number of dealers selling U.S. brand dropped from 510 in 2007 to 379 in 2010 (a 26% drop), and the total units sold of U.S. brands dropped by 30% during this time period.

Recent car dealer closures stemmed from two primary causes. First, American manufacturers discontinued a number of brands in the mid to late 2000s, starting with Oldsmobile in 2004, and continuing with Saturn and Pontiac in 2009, Mercury in 2010, and Saab in 2011.[6] These brands had seen steady declines in sales and were reported as being unpopular and out of touch with consumer needs in media and industry reports (Valdes-Dapena 2009). In Ohio, because of the terminations of product lines, 83 dealers stopped selling Pontiac, 20 dealers stopped selling Saturn, 42 dealers stopped selling Mercury, and 8 dealers stopped selling Hummer. The second cause of the dealer closures had to do with the financial crisis more directly. GM and Chrysler received TARP U.S. government loans in 2009, and because of that, their subsequent reorganization were allowed to terminate dealers.[7] In our counterfactual exercises, we will examine the effects of dealer closures and offer an explanation of why even unclosed dealers might prefer other dealers not to close. We will also show that the gain of rival dealers is exaggerated by the standard full-information demand model.

2.3. Spatial Distribution of Ohio Car Dealerships

We group the 970 dealerships into 248 clusters by using the density-based spatial clustering of applications with noise (DBSCAN) algorithm. The DBSCAN algorithm is ideal for grouping retail locations, as the objective is to partition points into dense regions separated by nondense regions. Importantly, the algorithm allows some points to be unclassified—so called noise points. Because there are many isolated car dealers in Ohio, we added a preclassification stage to the DBSCAN algorithm, where we combined dealers at very similar locations (i.e., in the same city block or multibranded dealerships located at the same address) into a single observation and preclassified all observations that

Table 2. Local Demographics in Ohio by Zip Code by Year

	Mean	SD	Min	Median	Max
Median household income ($000s)	48.28	14.05	10.06	47.23	140.56
Share of college degree or above	0.19	0.12	0.03	0.15	1
Number of households	3,850	4,932	0	1,515	26,802
Average household size	2.54	0.34	1.01	2.52	12
Share of households with children	0.30	0.07	0.01	0.29	1
Share of urban area	0.43	0.43	0	0.34	1

Source. American Community Survey.
Notes. The statistics are calculated based on the demographics of all Ohio zip codes from 2007 to 2014, including 9,418 unique zip code–year combinations. SD, standard deviation.

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

Table 3. Top 10 Brands

	Share of units sold (%)	No. of dealers	Single-brand dealers		Average trans. price ($)
Honda	14	55	46	84%	25,687
Chevrolet	13	204	145	71%	26,222
Ford	13	162	126	78%	27,097
Toyota	10	53	17	32%	25,754
Kia	5	45	29	64%	22,759
Hyundai	5	42	25	60%	22,984
Nissan	5	44	31	70%	26,202
Jeep	4	119	4	3%	30,334
Dodge	3	114	13	11%	26,838
Subaru	3	27	13	48%	25,777
All	100	970	611	63%	27,957

Source. Ohio Bureau of Motor Vehicles.
Notes. The statistics are calculated based on all new car sales that were registered in Ohio from 2007 to 2014. The sample selection is described in the text. The number of dealers in any given year is less than the total active dealers reported in Table 3 because of industry churn, primarily exit of Pontiac, Saturn, and other U.S. dealers and entry of foreign dealers.

were more than 10 kilometers from the next closest dealer as isolated dealers.[8]

We display visual results of the spatial distribution of dealership clusters using the DBSCAN algorithm in Appendices C and D. Among these 970 dealers, 74 of them are grouped into single-dealership clusters before we estimate clusters using the algorithm, and the remaining 916 dealers are grouped into 248 multi-dealer clusters. In Appendix C, we present a macro visualization of the clusters across the entire state. On the right-hand side is a map of the state of Ohio, and on the left-hand side is a graph of all the points we cluster, where the open circles are points the algorithm assigned as single dealers, and the colored crosses represent dealers in dealer clusters.

Because Ohio is a large state, it is difficult to get a sense of the clustering results without zooming into to particular geography. In Appendix D, we display the clustering results for four different cities/towns in Ohio. Each color represents a different cluster except orange, which in every case represents multiple separate clusters (there are many more clusters than clearly distinguishable colors). Each point may contain many dealer/brands, and the number of brands for each point is displayed below the marker.

Table 5 reports the descriptive statistics at the cluster-year level. On average, a cluster includes 3 physical dealers offering 35 car models of 4 brands and selling 1,403 cars annually. There is significant heterogeneity across clusters. The smallest cluster only includes 1 dealer offering 1 brand and selling 50 cars a year, whereas the largest cluster includes 19 dealers offering 188 car models of 25 brands and selling 16,909 a year. The clusters themselves are not uniformly sized, and traveling between any two dealers in any given cluster may take different time. Some clusters are trivial clusters with one location/brand. Some clusters have zero or close to zero distance between all of the dealers because they are at the same location or have a latitude/longitude that are approximately the same. Other clusters can have a width as great as 8 kilometers, the maximum size

Table 4. New Car Dealerships, Sales, and Price in Ohio from 2007 to 2014

	Number of dealers			Total units sold (000)			Average transaction price ($)		
Year	All	U.S.	Non-U.S.	All	U.S.	Non-U.S.	All	U.S.	Non-U.S.
2007	759	510	352	295	138	157	26,143	26,339	25,972
2008	738	477	357	299	131	168	25,942	26,168	25,767
2009	666	410	320	233	96	137	26,539	27,386	25,948
2010	635	379	290	267	138	159	27,710	28,905	26,902
2011	677	416	296	321	145	176	28,388	29,075	27,821
2012	689	416	305	345	151	192	28,566	29,323	27,970
2013	693	421	305	368	165	202	29,293	30,092	28,641
2014	690	416	309	376	173	203	29,798	30,736	29,002

Source. Ohio Bureau of Motor Vehicles.
Notes. The statistics are calculated based on all new car sales that were registered in Ohio from 2007 to 2014. The sample selection is described in the text.

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

Table 5. Descriptive Statistics of Dealership Clusters

	Mean	SD	Min	Median	Max
No. of dealers	3	3	1	2	19
No. of brands	4	4	1	3	25
No. of models	35	34	3	22	188
No. of products	56	55	4	35	307
No. of units sold	1,403	2,263	50	456	16,909

Notes. The statistics are calculated based on 248 dealer clusters in Ohio from 2007 to 2014, 1,784 cluster-year combinations in total. The product is defined at the car model–model year level (e.g., Toyota Camry 2010 model). SD, standard deviation.

we allow for a cluster. The average distance between dealer locations in multidealer-location clusters (i.e., not including multifranchised dealer locations) is 2.05 miles. The 25th quantile of the maximum distance between two locations in a cluster is 1.4 miles, and the 75th quantile of the maximum distance between two locations is 3.6 miles. Although making small changes to the DBSCAN parameters does not change these descriptive statistics much, we do not reestimate the model presented below for various parameter choices. See Table 5 for additional statistics.

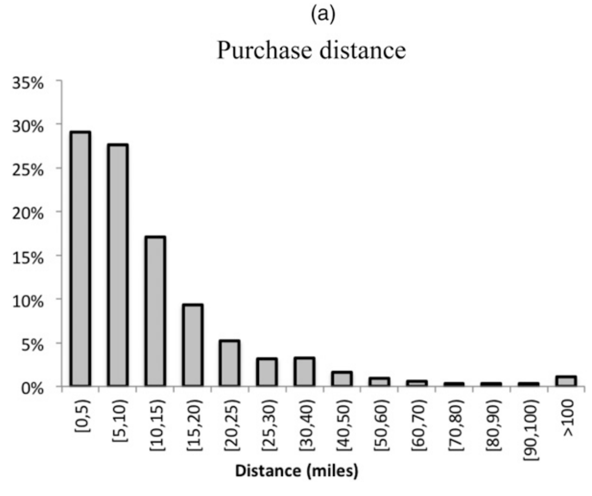
2.4. Consumer Travel and Dealer Colocation

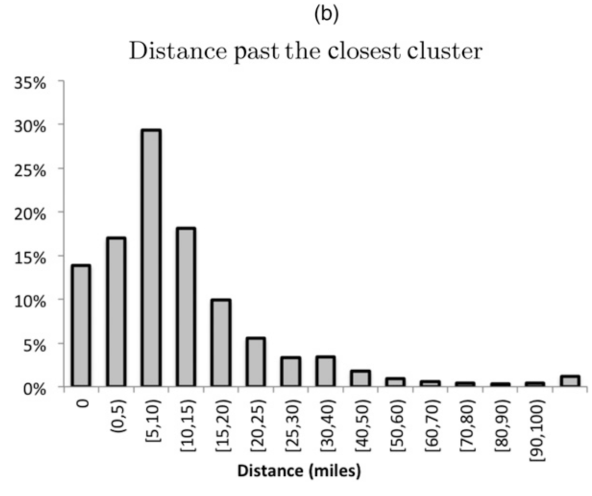
In this section, we present new car buyers' choices of dealers and their travel patterns. Figure 1(a) is a histogram of the distance from a buyer's residence to the dealer he or she purchased from. The average travel distance is 13.2 miles, the standard deviation is 18.5 miles, and the median is 8.4 miles. Twenty-nine percent of the Ohio new car buyers traveled less than 5 miles to buy their cars, 28% traveled more than 5 but less than 10 miles, 17% traveled more than 10 but less than 15 miles, and the remaining 26% traveled more than 15 miles. In particular, 90% buyers traveled less than 27 miles, and 95% buyers traveled less than 40 miles.

Figure 1(b) is a histogram of the extra distance that buyers traveled passed the closest dealer cluster. Only 14% of the Ohio new car buyers bought their cars from the nearest dealer clusters, 17% traveled less than 5 miles beyond their nearest clusters, 29% traveled more than 5 miles but less than 10 miles beyond their nearest clusters, and the remaining 40% traveled more than 10 miles beyond their nearest clusters. Our hypothesis is that one reason consumers travel is because the distant cluster that they choose offers more variety or lower prices, and they have limited ability to search all clusters. Regardless of their motives of traveling, the facts we present here suggest that new car buyers do search.

Finally, we present evidence that dealer colocation is important for consumer demand. To do this, we run a regression where the dependent variable is the distance a consumer traveled to purchase a car, and the key explanatory variable is the size of the geographic cluster, in terms of the number of dealers, where the car was purchased. The hypothesis is that consumers will travel farther to purchase cars from larger dealer clusters if colocation positively affects the purchase decision. The results are presented in Table 6. In the first column, we control for the buyer zip-code fixed effects. In the second column, we add the dummies indicating the make of the purchased car. The buyer zip-code effects control for differences in retailing environment faced by different consumers. For example, rural consumers may tend to travel farther just because dealer density is low in rural areas,

Figure 1. Distribution of Travel Distance





Source. Ohio Bureau of Motor Vehicles.
Notes. Travel distance refers to the distance between a new car buyer's residence and the selling dealer's address. Figures are drawn based on 2,503,734 new car transactions in Ohio from 2007 to 2014. Sample selection is described in the text.

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

Table 6. Regression: Travel Distance and Cluster Size

DV: *Purchase travel distance*	(1)	(2)
Number of colocated dealers	0.165	0.139
	(0.002)	(0.002)
Constant	11.954	14.417
	(0.0213)	(0.107)
Buyer zip-code effect	Yes	Yes
Car make effect	No	Yes
R^2	0.083	0.090
Observations	3,005,651	3,005,651

Notes. Unit of observation is an individual car transaction from Ohio between 2008 and 2014. The dependent variable (DV) is the distance from the buyer's residence to the dealer. *Number of colocated dealers* is the number of dealers in the same geographic cluster as the selling dealer, where clusters are defined as in Section 2.3.

and these consumers may also face dealers that tend to be less colocated with other dealers. We control for the car make because different makes have different retail network densities. U.S. brands are generally associated with lower travel distances because the retail networks are more dense. In both regressions, we find that longer purchased travel distances are associated with cars located in geographic clusters with more dealers. We take this finding as preliminary evidence that consumers value dealer colocation. The model we present in the next section provides a formal mechanism for this result.

3. Demand Model

We consider a market where differentiated cars are sold by many geographically dispersed dealers to geographically dispersed consumers. Consumers have limited information about the utility they derive from each car and must engage in costly search to resolve uncertainty before purchase.

3.1. Utility

We use subscript i to denote the consumer, subscript z to denote the zip code, subscript j to denote the product (e.g., Honda Accord 2010 model), subscript f to denote the dealer (e.g., "Bob's Honda Sales"), and subscript t to denote the year. The indirect utility that consumer i living in zip code z derives from product j sold by dealer f in year t is

$$u_{izjft} = \mathbf{x}_{jt}\tilde{\boldsymbol{\beta}}_{izt} - \tilde{\alpha}_{izt}p_{jft} + \varphi_f + \tau_t + \xi_{jft} + \varepsilon_{izjft}, \qquad (1)$$

where $\mathbf{x}_{jt}$ is a $K \times 1$ vector of observed car attributes including acceleration (horse power/curb weight), miles per dollar, car size, body style, an indicator of a luxury brand, and an indicator of a U.S. brand; p_{jft} is the average price of product j sold by dealer f in year t; φ_f is dealer-specific effects; τ_t is the yearly dummies; and ξ_{jft} is an unobserved term at the product-dealer-year level. Notice that we include dealer effects and year dummies in the utility so that ξ_{jft} captures only transitory demand shocks. The term ε_{izjft} is an idiosyncratic match value that can be ascertained only upon visiting the dealer, including the fit and comfortableness, personal image of the car, and specific way that salespeople in the dealership sell the car. Random coefficients $\tilde{\boldsymbol{\beta}}_{izt}$ and $\tilde{\alpha}_{izt}$, which capture consumer heterogeneity in tastes for product attributes and price, are assumed to take the following form:

$$\tilde{\beta}_{kizt} = \bar{\beta}_k + \boldsymbol{H}^x_{zt}\boldsymbol{\beta}^x_k + \sigma^x_k\varsigma^x_{kizt}, \text{ for } k = 1, \ldots, K,$$
$$\tilde{\alpha}_{izt} = \bar{\alpha} + \alpha^{inc}y_{zt} + \sigma^p\varsigma^p_{izt}, \qquad (2)$$

where $\boldsymbol{H}^x_{zt}$ is a vector of demographics at the consumer's residence zip code z in year t that affects her preferences, y_{zt} is the log of the median household income in zip code z in year t, and ς^x_{kizt} and ς^p_{izt} are assumed to follow standard normal distribution identically and independently distributed across characteristics and consumers, denoted by $F_\varsigma(\cdot)$.

Let δ_{jft} denote the mean utility across consumers of product j sold by dealer f in year t,

$$\delta_{jft} = \mathbf{x}_{jt}\bar{\boldsymbol{\beta}} - \bar{\alpha}p_{jft} + \varphi_f + \tau_t + \xi_{jft}, \qquad (3)$$

and let μ_{izjft} denote the heterogeneous utility that consumer i living in zip code z obtains from product j sold by dealer f in year t,

$$\mu_{izjft} = \sum_{k=1}^{K} x_{kjt}(\boldsymbol{H}^x_{zt}\boldsymbol{\beta}^x_k + \sigma^x_k\varsigma^x_{kizt}) + \left(\alpha^{inc}y_{zt} + \sigma^p\varsigma^p_{izt}\right)p_{jft}. \quad (4)$$

Then, we can write the utility Equation (1) as

$$u_{izjft} = \delta_{jft} + \mu_{izjft} + \varepsilon_{izjft}. \qquad (5)$$

Consumers have an outside option of not purchasing a new car from a dealer in Ohio. We assume that the utility from the outside choice is $u_{iz0t} = \varepsilon_{iz0t}$, where ε_{iz0t} is assumed to follow a standard type I extreme value (TIEV) distribution, identically and independently across consumers and over years.

3.2. Main Assumptions on Consumer Search

Consumers have the choice to search cars at one or multiple dealer clusters, where a dealer cluster represents a geographic area where dealers are colocated. A consumer pays a cost to search a cluster. Before searching, the consumer has expectations about the utility she will derive from each car being sold in a cluster, and once the cost is incurred, she learns the exact utility from each car in that area. We assume that consumers simultaneously decide the set of clusters to search, and conditional on that search set, they choose the best option. The model is a parametric version of the optimal portfolio choice problem described by Chade and Smith (2006), similar

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

to the consumer search application developed by Anderson et al. (1992, chapter 7), which was recently extended to empirical applications by De los Santos et al. (2012) and Moraga-González et al. (2015).

3.2.1. Simultaneous Search. The existing theoretical literature typically models consumer search strategies in two ways. One strand of the literature assumes simultaneous (or nonsequential) search, where consumers sample a fixed number of sellers and choose to purchase from the most preferred seller among those they have searched (see Stigler 1961, Burdett and Judd 1983, Janssen and Moraga-González 2004). The other strand of the literature assumes sequential search strategy, where after each search, consumers choose to purchase from the lowest price observed so far or making an additional search. Both search strategies have been adopted by empirical researchers. There are two studies we are aware of that test which search model is more consistent with the data in a retail goods setting. De los Santos et al. (2012) analyze a detailed data on the browsing and purchasing behavior of a large panel of consumers and find that the simultaneous search strategy outperforms the sequential search model in their setting. Honka and Chintagunta (2016) also find that the simultaneous search better matches their data on the demand for auto insurance, by examining the price variation in consumers' observed considerations sets. Because we do not observe individual consumer's consideration set or search process in our data, we are unable to let the data tell us which search strategy better represents our empirical setting. New car dealers are heavy advertisers, and their locations are usually well marked and near highways or transit thoroughfares. We think it is reasonable that car shoppers are well informed about dealer locations, car features, and other aspects of car buying before they plan shopping trips. Moreover, consumers could easily recall a previously searched car, which would violate an assumption of many sequential search models (see De los Santos et al. 2012).[9]

3.2.2. Information Set. We now explain consumers' information during the search process. First, consumers know the observed and unobserved product attributes at the product-dealer-year level, $(\mathbf{x}_{jt}, \varphi_f, \tau_t, \xi_{jft})$ for all j, f, t in Equation (1). Second, consumers know the average price of each product charged by each dealer in each year, p_{jft} for all j, f, t. This type of information is available on a plethora of car-buying websites. Also, advertisements may communicate this information, along with information about dealer-specific prices such as a dealer's willingness to give price discounts.[10] Finally, consumers know only the distribution of the match values ε_{izjft} before search and need to engage in costly search to know the exact values. As is common in the literature, we assume that ε_{izjft} follows a standard TIEV distribution, independently across consumers, products, dealers, and over years.

In our model, the term ε_{izjft} captures the information that can be ascertained only upon visiting the dealer. In reality, consumers search may be over these individual match values or over the product prices or some other common attributes. Unfortunately, our data do not allow us to identify the source of consumers' uncertainty. If consumers search to resolve their uncertainty on price, we can interpret the term ε_{izjft} as the deviation of a consumer's individual price from the average price. Let consumer i's (living in zip code z) utility from product j sold by dealer f in year t be $u_{izjft} = x_{jt}\beta_{izt} - \alpha\tilde{p}_{izjft} + \xi_{jft}$, where $\tilde{p}_{izjft}$ is the individual price that is unknown to consumer i before searching. We can decompose this individual price into two components: $\tilde{p}_{izjft} = p_{jft} + \vartheta_{izjft}$, where p_{jft} is the average price of product j sold by dealer f in year t, and ϑ_{izjft} is i's deviation from the average. If we set $\varepsilon_{izjft} = \alpha\vartheta_{izjft}$, the model of searching over match values in the paper is equivalent to this one of searching over price.

There are three things worth mentioning about the equivalence of match value and price search. First, to make the two models equivalent, the price coefficients have to be the same across consumers. Second, to make the two models equivalent, either we can assume ϵ_{izjft} follows a standard TIEV distribution and ϑ_{izjft} follows a TIEV distribution with location parameter 0 and scale parameter $1/\alpha$, or we can assume ϑ_{izjft} follows a standard TIEV distribution and ϵ_{izjft} follows a TIEV distribution with location parameter 0 and scale parameter α. Third, searching over prices and searching over match values are not the same if we consider the supply-side choices. In theory, variation of transaction price could help identify the source of uncertainty. However, using that variation for identification requires us to incorporate the price-bargaining process between consumers and dealers into the model, which is unachievable given our already complicated search model and our lack of search and bargaining data.

3.3. Search and Purchase Decisions

Let $\mathcal{J}_{ft}$ be the set of products from dealer f in year t, and let $\mathcal{F}_{mt}$ be the set of dealers in cluster m in year t. Let $\mathcal{S}_i$ define consumer i's set of all possible subsets of dealer clusters, with element S. For example, in the two-cluster case, the set of all possible subsets of dealer clusters, $\mathcal{S}$, includes {Ø}, {1}, {2}, {1&2}, where {Ø} represents "do not search," {1} represents "search cluster 1 only," {2} represents "search cluster 2 only," and {1&2} represents "search both clusters." Here, we assume that consumers never travel farther

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

than 40 miles to search for cars, which accounts for 95% of buyers in our sample. As a result, the potential dealer clusters that are considered vary across consumers, depending on which zip code they live in. This restriction dramatically reduces the computational burden of computing consumers' optimal search sets. Otherwise, it is computationally infeasible to compute them by allowing consumers to optimally choose their search sets among 248 clusters.

The expected gain that consumer i living in zip code z obtains from a search set $S \in \mathcal{S}_i$ is

$$U_{izt}(S) = E_\varepsilon\left[\max\{u_{iz0t}, \max_{j\in\mathcal{J}_{ft}, f\in\mathcal{F}_{mt}, m\in S} u_{izjft}\}\right] = \ln\left[1 + \sum_{j\in\mathcal{J}_{ft}, f\in\mathcal{F}_{mt}, m\in S} \exp(\delta_{jft} + \mu_{izjft})\right],$$

where the analytic expression exists because the match values ε follows a TIEV distribution.

We define the expected value that consumer i living in zip code z gets from a search set $S \in \mathcal{S}_i$, denoted by $V_{izt}(S)$, as the difference between her expected gain from S and the searching cost that she needs to pay to visit all clusters included in the search set S. Furthermore, we specify the search cost of a search plan S as

$$C_{izt}(S) = \sum_{m\in S} c_{izmt} + v_{izt} + \omega_{izSt}, \tag{6}$$

where c_{izmt} is the search cost that consumer i living in zip code z needs to pay if she visits a dealer cluster m in year t, v_{izt} is an individual-year-specific term that captures the search-cost shocks associated with that consumer, and ω_{izSt} is an individual–search set–year–specific error term that captures the unobserved search-cost shocks such as the traffic condition of that search. We assume that v_{izt} follows a normal distribution with mean zero and standard variance σ_v^2. If consumer i living in zip code z does not search any cluster, her utility is $C_{izt}(\emptyset) = \omega_{iz0t}$.[11]

Furthermore, we assume that the search cost of consumer i (living in zip code z) from visiting a dealer cluster m in year t is

$$c_{izmt} = \gamma_{iz} d_{zm} + \rho n_{mt}, \quad \text{with } \gamma_{iz} = \lambda_0 + \boldsymbol{H}^c_{zt}\boldsymbol{\lambda}^H, \tag{7}$$

where d_{zm} is the distance from consumer i's zip code to the geographic center of a dealer cluster m, n_{mt} is the number of dealers in that cluster, and $\boldsymbol{H}^c_{zt}$ is a vector of consumer i's zip-code-level demographics that may affect her search cost, including the log of median household income, share of urban area, and share of households with children under 18 years old in i's zip code in year t. Importantly, we include ρ in the model to allow the cost of search to vary with the size, in terms of dealers, of the clusters. Although this does not formally test our model against the model of Moraga-González et al. (2017) because the models are not nested, a positive estimate of ρ would support the per-dealer assumption of Moraga-González et al. (2017). Importantly, this additional search cost on the number of dealers gives the model flexibility in allowing for economies of agglomeration. As we discuss below, a very large positive ρ would imply that there is no positive demand effect of dealer colocation.

Consumer i (living in zip code z) chooses the search set S^* in year t that gives her the highest expected utility among all possible search sets $\mathcal{S}_i$; that is,

$$V_{izt}(S^*) \geq V_{izt}(S)\ \forall S \in \mathcal{S}_i.$$

In our model, the variation in the optimal search sets across consumers is generated by their different valuations for products contained within each search set $[U_{izt}(S)]$, their different distances to dealer clusters (d_{zm}), their different demographics $(\boldsymbol{H}^c_{zt})$, the number of dealers in each cluster (n_{mt}), their idiosyncratic cost shocks (v_{izt}), and their different draws of search-set-specific idiosyncratic cost shocks (ω_{izSt}). The size of a consumer's search cost ultimately determines how many clusters she will search.

However, as pointed out by Chade and Smith (2006), the optimal search set will not necessarily follow a cutoff rule of an ordering of $U_{izt}(S)$s from highest to lowest. Following De los Santos et al. (2012), among others, we analytically compute the probability of choosing each search set by assuming that ω_{izSt} follows a TIEV distribution with a location parameter of zero and a scale parameter of κ.[12] Specifically, the probability that consumer i living in zip code z chooses a search set S is

$$\mathcal{P}_{izSt} = \frac{\exp[(U_{izt}(S) - \sum_{m\in S} c_{izmt} - v_{izt})/\kappa]}{1 + \sum_{S'\in\mathcal{S}_i} \exp[(U_{izt}(S') - \sum_{m\in S'} c_{izmt} - v_{izt})/\kappa]}.$$

Let m^f denote the cluster of dealer f. The probability that consumer i living in zip code z will purchase product j from dealer f in year t conditional on a search set S follows the familiar analytical expression

$$\mathcal{P}_{izjft|S} = \begin{cases} \dfrac{\exp(\delta_{jft} + \mu_{izjft})}{1 + \sum_{j'\in\mathcal{J}_{ft}, f'\in\mathcal{F}_{mt}, m\in S} \exp(\delta_{j'f't} + \mu_{izj'f't})} & \text{if } m^f \in S, \\ 0 & \text{if } m^f \notin S. \end{cases}$$

Then, the unconditional probability that consumer i purchases product j from dealer f in year t is

$$\mathcal{P}_{izjf}(\boldsymbol{\delta}_t, \mathbf{x}_t, \mathbf{p}_t, \boldsymbol{H}_{zt}, \mathbf{d}_z; \boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = \int \sum_{S\in\mathcal{S}_i} \mathcal{P}_{izjft|S}\mathcal{P}_{izSt} dF_\varsigma(\cdot), \tag{8}$$

where $\boldsymbol{H}_{zt} = (y_{zt}, \boldsymbol{H}^x_{zt}, \boldsymbol{H}^c_{zt})$ includes the zip-code-level demographics that affect consumer preference and

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

search cost, $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}^x, \boldsymbol{\sigma}^x, \alpha^{inc}, \sigma^p, \boldsymbol{\lambda}, \rho, \kappa, \sigma_v)$ represents all "nonlinear" parameters in the model, and $\boldsymbol{\theta}_2 = (\bar{\boldsymbol{\beta}}, \bar{\alpha}, \varphi_f, \tau_t)$ represents all "linear" parameters in the model.

3.4. Discussion of the Model

3.4.1 Search Cost and Substitution Patterns. Nontrivial consumer search creates particular substitution patterns within and across clusters. All cars within a cluster are either in or out of a given consumer's choice set. In this sense, within-cluster substitution between cars is similar to standard full-information Berry–Levinsohn–Pakes (BLP; see Berry et al. 1995) models with zero search cost. In our model, however, consumers dislike traveling distance, and hence, all else equal, they are more likely to visit nearby clusters. Therefore, if two clusters are farther away, the products across these two clusters are less substitutable, all else equal.

Particularly, the magnitude of the across-cluster substitution depends on consumer search cost. In our model, the parameter γ measures the extent to which consumers dislike traveling distance. Hence, a larger γ implies a lower across-cluster substitution. Similarly, a larger ρ (the coefficient before the number of dealers in a cluster) also implies a larger search cost, and hence a lower cross-cluster substitution. Another crucial parameter in our search-cost equation is κ, the standard deviation of the idiosyncratic shock to each search set. A larger κ implies more randomness, and hence a lower importance of car attributes and traveling distance in consumers' optimal choice of search set. In other words, a larger κ implies that dealers' location is less relevant, and hence the products in different clusters are more substitutable. In the extreme case when κ is infinite, consumers choose each search set among all possible sets with equal probability, and dealers' locations will not affect the substitution at all. In this sense, the parameter κ has an interpretation similar to that of the nesting parameter in a nested logit model if nests are defined as geographic clusters. Just like the nesting parameter, κ governs the amount of substitution across dealer clusters, or nests. As we discuss later, identification of this parameter is akin to identification of the nesting parameter in Berry (1994), or any random coefficient that controls substitution patterns.[13]

3.4.2. Agglomeration, Competition, and the Economics of Retail Closures. In our model, cars in large clusters are more likely to be searched but less likely to be chosen given they are searched. The former happens because consumers find large clusters more attractive (as long as ρ is not too high), and the later happens because large clusters imply more choices and greater competition. Here we explain the agglomeration effect in more detail. Intuitively, the sign and magnitude of the agglomeration effect of dealer colocation are determined by two forces. First, the likelihood that a cluster is visited by a consumer depends on the additional gain by including it into her search set. A cluster with more dealers (and hence more products) is, all else equal, more attractive to consumers and hence more likely to be included in search. This is because a cluster with more products provides a higher chance for a consumer to find a product that has the characteristics she values and also provides more draws of the idiosyncratic shock ϵ_{izjft}, and hence a higher maximum order statistic.[14] This is the agglomeration benefit from colocation. However, the likelihood a cluster is visited by a consumer also depends on the extra search cost if she includes it in her search cost. Most likely, the colocation of more dealers leads to a higher search cost (a positive ρ). In this case, a cluster with more dealers is more costly to include in the search set and hence less likely to be visited, all else equal. This is the agglomeration cost from colocation. Whether colocation leads to a positive or negative agglomeration effect crucially depends on the value of ρ. If ρ is sufficiently small, the agglomeration benefit tends to outperform the agglomeration cost, and hence a cluster with more dealers tends to be more likely visited by consumers. In contrast, if ρ is sufficiently large, a cluster with fewer dealers tends to be more likely included in a consumer's search set.

Our model also clearly predicts how the closure of a dealer affects other colocated dealers in the same cluster. To illustrate this, below we walk through a simpler two-cluster example. Consider two dealer clusters, 1 and 2. There are n_1 dealers located in cluster 1 and n_2 dealers located in cluster 2. So the set of possible search sets is $\mathcal{S} = \{\emptyset, 1, 2, 1\&2\}$. For notation simplicity, let u_{if} denote the consumer i's utility from firm f, which follows a type I extreme value distribution with location parameter μ and scale parameter 1. For simplicity, assume that consumers have the same search cost for a cluster which is a linear function of the distance and the number of dealers in that cluster, that is, $c_m = \gamma d_m + \rho n_m$. In addition, there exists a search-cost shock for each search set, denoted by ω_S, which is assumed to follow a TIEV distribution with location parameter 0 and scale parameter κ.

The expected utility of choosing a search set $S \in \mathcal{S}$ is $U_S = \ln(1 + (\sum_{m\in S} n_m)e^{\mu})$, and the search cost is $c_S = \gamma(\sum_{m\in S} d_m) + \rho(\sum_{m\in S} n_m) + \omega_S \equiv \bar{c}_S + \omega_S$. Because of the assumption on the distribution of ω_S, the probability of choosing a search set $S \in \mathcal{S}$ is

$$\mathcal{P}_S = \frac{\exp[(U_S - \bar{c}_S)/\kappa]}{\begin{array}{c}1 + \exp[(U_1 - \bar{c}_1)/\kappa] + \exp[(U_2 - \bar{c}_2)/\kappa] \\ + \exp[(U_{1\&2} - \bar{c}_{1\&2})/\kappa]\end{array}},$$

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

and the probability of choosing a dealer in cluster m given a search set S is

$$\mathcal{P}_{m|S} = \frac{e^{\mu}}{\exp(U_S)}. \tag{9}$$

Then, the unconditional probability of buying from a dealer in cluster m is

$$\mathcal{P}_m = \sum_{S \in \mathcal{S}_m} \mathcal{P}_S \mathcal{P}_{m|S},$$

where $\mathcal{S}_m$ is the set of search sets that include cluster m.

We show that our model can predict either positive or negative agglomeration effects, depending on the parameters. Moreover, the sign of the agglomeration effect depends on the number of dealers. We define that the agglomeration effect of colocating in cluster m is positive (negative) if $\partial(\mathcal{P}_m + \mathcal{P}_{1\&2})/\partial n_m$ is positive (negative); that is, cluster m is more (less) likely to be visited if more dealers colocate in that cluster.

Proposition 1. *There exists a positive cutoff ρ^*_m such that the agglomeration effect in cluster m is positive iff $\rho < \rho^*_m$ and negative iff $\rho \geq \rho^*_m$. Moreover, the cutoff ρ^*_m is decreasing in n_1, n_2 and γ, but increasing in μ.*

Proposition 2. *In the case of $\rho < \rho^*_m$ so that the agglomeration effect is positive, there exists a cutoff κ^* such that as κ increases, the agglomeration effect becomes stronger when $\kappa \leq \kappa^*$ and weaker when $\kappa > \kappa^*$.*

The proofs of these propositions are in Appendix A. Next we can show the impacts of closing a dealer on other colocated dealers in the same cluster. Closing a dealer in cluster 1 would lower the expected utility of searching cluster 1 through U_1 and the expected utility of searching both clusters through $U_{1\&2}$. Consider an incumbent colocated dealer j in cluster 1. Lower U_1 and $U_{1\&2}$ increase the conditional probabilities $\mathcal{P}_{j|1}$ and $\mathcal{P}_{j|1\&2}$; that is, closing a dealer in cluster 1 increases the probability of purchasing from a colocated dealer conditional on cluster 1 being searched. This is the competition effect of dealer closure on a colocated dealer. Meanwhile, when ρ is below (above) ρ^*_1, lower U_1 and $U_{1\&2}$ decrease (increase) the probability that cluster 1 is included in consumers' search set, $\mathcal{P}_1 + \mathcal{P}_{1\&2}$. This is the agglomeration effect of dealer closure on a colocated dealer. In the case of $\rho \geq \rho^*_1$, the agglomeration effect is negative, and hence the total effect of closing a dealer in cluster 1 would benefit other colocated dealers in cluster 1. However, when $\rho < \rho^*_1$, the agglomeration effect is positive, and hence the total effect of dealer closure depends on the size of the agglomeration and competition effects. As κ increases, for example, when it is near or lower than κ^*, the agglomeration effect tends to dominate the competition effect, implying that the closing of a dealer in cluster 1 would make the colocated dealers in cluster 1 worse off. This is intuitive, as a low κ implies that search is less random and consumers are more willing to substitute across clusters based on cluster characteristics.

3.4.3. Cluster vs. Per-Dealer Search. Consumers in the model necessarily learn the characteristics of all cars from all dealers in a search set. This is in contrast to models of Moraga-González et al. (2015) and Moraga-González et al. (2017), who assume that consumers pay a cost to learn about all cars at a single dealer. Our model does not nest theirs. A nested model would include both a per-dealer search cost and a cost to visiting the geographic cluster. We do not do this mainly for two reasons.

First, the nested model would impose substantially greater computational burden because the set of all choice sets would become very large. Even though Moraga-González et al. (2015) and Moraga-González et al. (2017) offer a very innovative method to simplify the computation of choice probabilities, the derivations are based on the normalization of $\kappa = 1$ (or $\sigma_\lambda = 1$ in Moraga-González et al. 2015, p. 18). In contrast, our interest is to study the agglomeration effect, which is shown to be sensitive to the value of κ. So we cannot normalize κ to be one. Because of that, we cannot collapse the computation as they suggest. Nevertheless, the innovations of Moraga-González et al. (2015) and Moraga-González et al. (2017) can be extremely useful in situations where the researcher wants to model search and is not interested in a counterfactual simulation that is sensitive to the value of κ.

Second, to separately identify the cluster search cost and the per-dealer search cost in the nested model, ideally one needs information about which search set each consumer actually chose, which we do not have. In that case, the number of different searched clusters and the *identities* of the searched dealers in each cluster would identify the two different costs. For example, if consumers consistently searched all dealers from only a single, nearby area, this would imply that cluster search costs are high, but, conditional on driving to a cluster, per-dealer search costs are low. In contrast, if consumers searched one dealer each from many different disperse clusters, then that would imply the cluster search cost is low. By not modeling per-dealer search costs, we are likely overprescribing information to consumers, in the sense that consumer may not really know all information about every car in a particular dealer cluster. Because we assume consumers are overinformed within a cluster, our model likely implies too much competition between dealers in the same geographic cluster, so our markups are likely to be underestimated. However, this is speculative because estimates from a nested model might imply much more cross-cluster

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

substitution if the cluster search cost was estimated as small, thus making dealers across clusters more competitive.

4. Estimation and Results

Our estimation procedure follows Goolsbee and Petrin's (2004) two-step approach. In the first step, we use the individual transaction data to maximize the likelihood function that includes product-dealer-year-specific dummies to capture those mean utilities. This identifies all nonlinear parameters $\boldsymbol{\theta}_1$ and the mean utilities $\boldsymbol{\delta}$. To estimate linear parameters of the model $\boldsymbol{\theta}_2$, in the second step we run a regression of the estimated mean utilities on model-dealer-year characteristics, using instruments to account for correlation of price with the unobserved quality ξ_{jft}.

In the first step, for any candidate value of $\boldsymbol{\theta}_1$ and any vector of product fixed effects $\boldsymbol{\delta}$, the log-likelihood function is

$$\log L(\boldsymbol{\delta}, \boldsymbol{\theta}_1) = \sum_{i,z,j,f,t} I_{izjft} \times \log\left[\frac{\mathcal{P}_{izjf}(\boldsymbol{\delta}_t, \mathbf{x}_t, \mathbf{p}_t, \boldsymbol{H}_{zt}, \mathbf{d}_z; \boldsymbol{\theta}_1)}{\sum_{j,f} \mathcal{P}_{izjf}(\boldsymbol{\delta}_t, \mathbf{x}_t, \mathbf{p}_t, \boldsymbol{H}_{zt}, \mathbf{d}_z; \boldsymbol{\theta}_1)}\right], \tag{10}$$

where $\frac{\mathcal{P}_{izjf}(\boldsymbol{\delta}_t, \mathbf{x}_t, \mathbf{p}_t, \boldsymbol{H}_i, \mathbf{d}_i; \boldsymbol{\theta}_1)}{\sum_{j,f} \mathcal{P}_{izjf}(\boldsymbol{\delta}_t, \mathbf{x}_t, \mathbf{p}_t, \boldsymbol{H}_i, \mathbf{d}_i; \boldsymbol{\theta}_1)}$ is the probability that consumer i living in zip code z purchases product j from dealer f given that she purchases, and I_{izjft} is an indicator indicating whether consumer i living in zip code z purchases j from f in year t.

Following Goolsbee and Petrin (2004), we do not maximize the likelihood over the entire space of $(\boldsymbol{\theta}_1, \boldsymbol{\delta})$ directly. Instead, we concentrate out the likelihood function and only search over the space of $\boldsymbol{\theta}_1$. To do this, we condition on $\boldsymbol{\theta}_1$ and solve for the vector $\boldsymbol{\delta}_t(\boldsymbol{\theta}_1)$ that matches the observed market shares to those predicted by the model:

$$s_{jft} = s_{jf}(\boldsymbol{\delta}_t, \mathbf{x}_t, \mathbf{p}_t; \boldsymbol{\theta}_1) = \sum_z \frac{b_{zt}}{B_t} \int \mathcal{P}_{izjf}(\boldsymbol{\delta}_t, \mathbf{x}_t, \mathbf{p}_t, \boldsymbol{H}_{zt}, \mathbf{d}_{zt}; \boldsymbol{\theta}_1) dF(\varsigma_i, \omega_i), \tag{11}$$

where b_{zt} is the number of potential consumers in zip code z, $B_t = \sum_z b_{zt}$ is the total number of potential consumers, and $F(\cdot)$ is the distribution of random preference to product attributes and price ς_i and the random search cost term ω_i. This procedure is just the maximum likelihood estimation (MLE) analogy to the generalized method of moments (GMM) procedure proposed in BLP.

After we obtain the estimates of nonlinear parameters $\hat{\boldsymbol{\theta}}_1$ from the first step, we compute the mean utilities at the estimated value of $\hat{\boldsymbol{\delta}} = \boldsymbol{\delta}(\hat{\boldsymbol{\theta}}_1)$. In the second step, we estimate $\boldsymbol{\theta}_2$ from Equation (3). We construct instruments in a similar spirit to Gandhi and Houde (2017). For a given car characteristic, we take the squared difference of a particular product with the average characteristic in that product's cluster. We also take this difference with respect to the average characteristic in the entire state of Ohio. Let $\boldsymbol{Z}_{jft}$ denote the vector of all excluded variables including car characteristics $\mathbf{x}_{jt}$, L instrumental variables, dealer fixed effects, and year dummy variables. Our empirical moment conditions for the second stage are

$$G_N(\boldsymbol{\theta}_2) = \frac{1}{N_2} \sum_{j,f,t} \xi_{jft} \mathbf{Z}_{jft}, \tag{12}$$

where N_2 is the number of product-dealer-year-level observations.

Notice that the estimation errors of the first-stage estimation will be carried into the second-stage estimation, implying that we need to correct the standard errors of the second-stage estimates. We correct for this additional uncertainty in the second-stage estimates and provide the derivation of the asymptotic covariance matrix of the second-stage estimates in Appendix B.[15] We use antithetic acceleration when we simulate the integrals in the choice probabilities, and we do not adjust the standard errors for simulation bias (for details, see Stern and Zhou 2018).

4.1. Identification

In this section, we provide an informal discussion of the model identification. First, we should note that we do not claim to identify search, per se. In general, the model is identified subject to all of the behavioral and parametric assumptions. We do not formally test the assumption cluster search against an alternative assumption, for example, no search or per-dealer search. The identification of the parameters in the utility function is similar to the full-information BLP models. They can be identified because we observe different within-cluster market shares corresponding to different product characteristics and different sets of products available across clusters and over time. Also, we also observe consumer-level choices, and the variation in their within-cluster choices corresponding to different product characteristics, consumer characteristics, and choice sets also helps to identify those parameters in the utility function.

To identify the price coefficients, we need to address the classical endogeneity problem that arises because dealers and consumers observe the unobserved quality when making their decisions, and so the average price will adjust to the changes in unobserved quality. Relevant and valid instruments are those variables that are correlated with prices and independent of unobserved transitory demand shocks. Following Nevo (2001), Houde (2012), and Gandhi and Houde (2017), we use deviations of particular product characteristics from the averages of other

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

products as instruments. Specifically, we include four other groups of instrumental variables: (i) the deviations from the average characteristics of all other product–dealer combinations available in the same dealer cluster and in the same year, (ii) the squares of the first group variables, (iii) the deviations from the average characteristics of all other product–dealer combinations in the same year, and (iv) the squares of the third group variables. We display the results from a regression of price on all of the exogenous variables in Appendix E. Most of the excluded variables are highly significant, and the R^2 is 0.72.

The specific search mechanism is not identified per se. Because we do not observe consumer choice sets or search behavior, we cannot reject another search model, for example, a full-information model or a sequential search model, in favor of our model. Conditional on our parametric assumptions about search, the parameters in the search-cost function can be identified from covariation in individual distances and individual choices in the data. Consider two consumers with similar preferences in product attributes and price. They should have similar expected gain from each search set $U_{izt}(S)$ and similar conditional purchasing probabilities $\mathcal{P}_{izjft|S}$. The distance coefficients, $\boldsymbol{\lambda}$, can be identified from the variation in their choices corresponding to their different distances to dealers and their different demographics affecting their search costs ($\mathbf{H}^c_{zt}$). The cluster size coefficient ρ is identified from covariation in choices and the size of the chosen dealer's cluster. The identification of κ comes from the variation of individual choices among those consumers with similar demographics and similar distances to dealer clusters. For example, if those consumers make similar choices, then the variance of their search-cost shocks κ should be small. In this sense, κ plays a role similar to that of the nesting coefficient in a nested logit model. As choice sets vary across similar consumers, does the researcher observe consumers making similar decisions or apparently random decisions? The variance of ε is not identified, as is typical in discrete choice models, and, in turn, κ is not separately identified from the variance of ϵ.

The above argument of the identification of the search-cost parameters relies on the assumption that dealer entry, exit, and location choices are not correlated with the unobserved transitory demand shocks, ξ_{jft}, after controlling for dealer and time fixed effects. Because we include dealer and year effects, this assumption is valid if entry decisions are based on the long-run store characteristics and aggregate economic shocks, but not on the realization of the transitory shocks ξ_{jft}. This assumption is reasonable in our context, because the sunk cost involving the entry, exit, or location change of a dealer is substantial, partly because of regulations that limit entry and exit. Forced exit of dealers by the manufacturer is very difficult in this industry because of state laws requiring payments to dealers for the termination of franchise contracts. In addition, there are other state laws that make entry and exit difficult, including mandated exclusive territories for brands. For a discussion of the regulatory environment, see Lafontaine and Morton (2010) and Murry and Schnieder (2016). Also, to the extent the local demographics and population change over time, initial decisions about entry may not reflect current demographics, population, or other transitory factors; see Murry (2018) for evidence. As discussed in Section 2.2, the number of dealers decreased by nearly 16% in Ohio during the past financial crisis, creating sharp changes in the structure of local markets. Importantly, these changes were driven mainly by nationwide brand terminations and dealer closures due to car manufacturers' financial crises, and not by factors related to local transitory demand conditions.

4.2. Model Estimates and Fit

Table 7 reports the estimates of all parameters in our search model. As expected, the estimate of α^{inc} is positive, implying that higher-income consumers are less price sensitive. The implied consumer-model-dealer-year-level own-price elasticities of demand range from −10.73 to −1.38, with a sales-weighted average of −4.05. This suggests that consumers are price sensitive on average, but there is substantial heterogeneity. Overall, our estimates of price elasticities are consistent with those of previous studies of automobile demand. For example, the average own-price elasticity is equal to −4.1 in Albuquerque and Bronnenberg (2012), −5.3 in Murry (2015), and −3.14 in Nurski and Verboven (2016). As expected, the average consumer prefers cars with higher acceleration (horsepower divided by vehicle weight) and higher miles per dollar. Larger households like larger cars more. Consumers that live in zip codes with higher rates of education prefer U.S. brand cars less.

The distance coefficients have the expected signs and are precisely estimated. The search cost is increasing in the traveling distance. This relationship is even stronger for households with higher income and children, and also stronger for locations with more urban areas and worse traffic conditions. The coefficient before the number of dealers in a cluster (ρ) is almost zero and not significant at the 10% level. The standard deviation of the search-set shock (κ) is estimated to be 0.3716 and is estimated precisely.[16] In the following sections, we will further discuss the implications of the estimated search costs.

To examine the fit of the model, we simulate the choices of those buyers who we used to construct the likelihood function in the first-stage estimation.

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

Table 7. Model Estimates

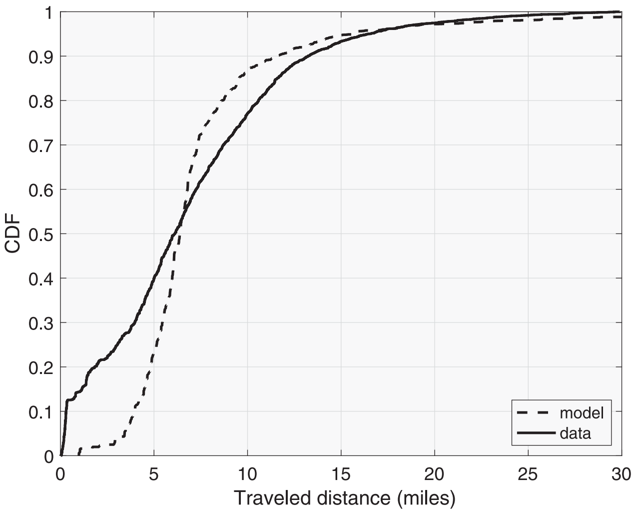
Variable	Coefficient	Estimate	SE
Utility parameters			
Price ($10,000)	$\bar{\alpha}$	–1.4201	(0.0111)
Price × log(*Income*)	α^{inc}	0.1224	(0.0169)
Price random effect	σ^{p}	0.1338	(0.0429)
log(*Acceleration*)	$\bar{\beta}_1$	1.4064	(0.0324)
log(*Car size*)	$\bar{\beta}_2$	8.355	(0.0559)
log(*Miles per dollar*)	$\bar{\beta}_3$	1.0662	(0.0532)
Luxury brand	$\bar{\beta}_4$	0.8394	(0.0298)
U.S. brand	$\bar{\beta}_5$	0.3856	(0.0248)
log(*Car size*) × *Household size*	β_2^x	1.8612	(0.3883)
log(*Miles per dollar*) random effect	σ_3^x	0.1814	(0.0991)
U.S. brand × *College degree*	β_5^x	–1.0692	(0.4201)
U.S. brand random effect	σ_5^x	0.0425	(0.0535)
Constant	$\bar{\beta}_0$	–6.9138	(0.2061)
Search parameters			
Distance (100 miles)	λ_0	13.6687	(0.6776)
Distance × log(*Income*)	λ_1^H	0.4678	(0.1159)
Distance × *Share of households with Children*	λ_2^H	–0.2601	(0.0174)
Distance × *Share of urban area*	λ_3^H	1.7977	(0.4901)
No. of dealers in cluster	ρ	0.0014	(0.0020)
SE of consumer heterogeneity	σ_v	1.2209	(0.1348)
SE of search-set heterogeneity	κ	0.3716	(0.0376)

Notes. The estimation includes 101,371 model-dealer-year combinations, 9,415 zip code–year combinations, 2,112 dealer cluster–year combinations, and 8,000 individuals. The mean utility function also includes body style dummies, dealer fixed effects, and yearly dummies. The average transaction price is measured in $10,000. Distance between a consumer zip code and a dealer cluster is measured in 100 miles. SE, standard error.

Figure 2 presents the cumulative distribution functions (CDFs) of the model predicted purchase travel distance and the distance observed in the data. Our model predicts that the average distance is 7.42 miles and the standard deviation is 4.81 miles, whereas the average distance is 6.80 miles in our estimation sample and the standard deviation is 5.20 miles. A Kolmogorov–Smirnov test rejects the null that the model predicted distribution is identical to the data.

We conjecture that the discrepancies of the two distributions are caused by the following reasons. First, we restrict consumers to travel a maximum of 40 miles, but this mileage limit may be different across different individuals. Consider a consumer who does not travel more than 30 miles to buy a car. Suppose that within 40 miles of her residence, there are two dealers: one is 1 mile away and the other is 35 miles away. Because she considers only dealers within 30 miles, she will choose the first dealer for sure, and her traveling distance in the data will be just 1 mile. However, the model assumes that all consumers will consider all dealers within 40 miles and it will predict positive probabilities for both dealers, and as a result, her predicted distance will be larger than her actual distance. This can explain why the empirical CDF is above the model CDF on the segment of small distance. Second, consumer search cost may not be linear in the traveling distance. The distance function could be concave, or more likely a step function. As a result, a model with a linear approximation of the search cost will overpredict the search cost when the distance is large, and hence tend to predict that consumers travel less than their actual distance on the segment of large distance.

Figure 2. Predicted and Empirical Distributions of Traveled Distance



Notes. The solid line represents the cumulative distribution of the actual traveling distance of the individual buyers that we used to construct the likelihood function in the first-stage estimation. The dashed line represents the cumulative distribution of their traveling distance predicted by the model.

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

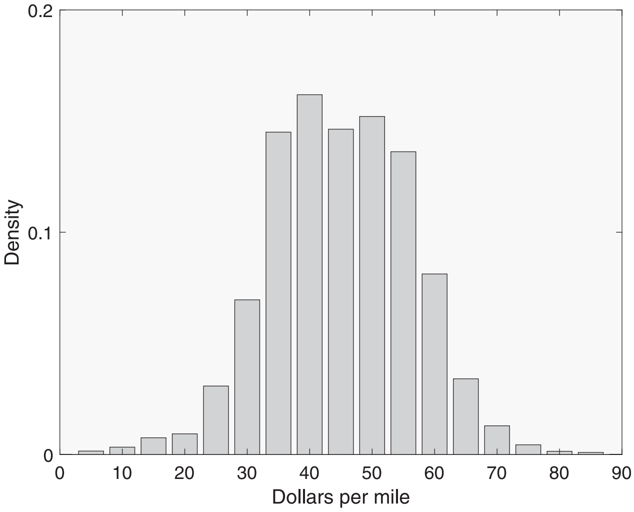
4.3. Implied Search Cost

To get a sense of the economic magnitude of the parameter estimates, it is useful to consider how much a product's price needs to be lowered to compensate consumers if they have to travel one more mile. To do that, for each product, we first calculate the sales change if it was one mile away from every consumer. Then, we calculate how much the price would need to be lowered to compensate the same sales loss. Figure 3 demonstrates the distribution of dollars per mile. Our results suggest that the average value is $45, and there is substantial heterogeneity across individuals.

Our estimate of the travel cost is lower than those reported in other studies that estimate consumer distance costs in the new car retail industry. For example, Moraga-González et al. (2015), with the paper closest to ours, report a median travel cost of €107 per kilometer. The difference could come from two reasons. First, Moraga-González et al. (2015) consider the market for cars in the Netherlands. It is reasonable to think that search costs are higher in Europe because of congestion and higher fuel prices. In general, there are drastic lifestyle differences between Ohio and the Netherlands, so it is not clear how these numbers should be compared. Second, we observe individual-level choices and the exact locations of the buyer and seller, and hence our parameter estimates could reflect important microlevel information captured in the covariation of distance and purchase probabilities.[17]

Furthermore, we examine what the estimates imply about consumers' search intensity. In general, consumer search is limited. Our estimates suggest that among those consumers who search, 46.67% of them search only one cluster, 34.41% search two clusters, 18.68% search three clusters, and fewer than 1% of them search more than three clusters. These results are consistent with industrial reports and previous studies. For example, in a survey by DME Automotive, an industry consulting group, 47% of all new car buyers visited a single dealer before purchase.[18] Moraga-González et al. (2015) report that 47% of survey respondents in their consumer survey data searched one dealer. Although our model does not have empirical content regarding specifically how many dealerships are searched because search happens at the dealer-cluster level, we can at least say that searching two or more clusters implies searching at least two dealers, and so our estimates of consumer search intensity could be consistent with those of other sources.

Figure 3. Distribution of Dollar Per Mile



5. Counterfactual Experiments

In this section, we use our estimation results to conduct two sets of counterfactual exercises. The goal of our first set of counterfactual exercises is to understand how search frictions affect the market outcomes. We examine how the market outcomes change when we change the values of those key parameters in the search-cost equation while keeping other parameters unchanged. In the second set of counterfactual exercises, we examine the impact of dealer closure on remaining dealers. In particular, we decompose the total effect of a dealer closure into an agglomeration effect and a competition effect. Simulating a dealer closure is not only a clear way to decompose the competition and agglomeration forces at work in the model; it also is a highly policy relevant exercise. As discussed in Section 2.2, the recent U.S. financial crisis of 2007–2009 saw many retail exits; however, the microeconomic effects of retail closure, the agglomeration effect in particular, have not been well documented. To do both types of counterfactual exercises, we first specify a supply-side model to describe the price setting of car dealers. This allows us to recover marginal retail costs and evaluate changes to market outcomes, including optimal prices, in counterfactual environments.

5.1. Retail Pricing

We assume that car dealers, which are multiproduct firms, play a static Nash–Bertrand pricing game by simultaneously setting the retail price for each of their cars in each year.[19] The total variable profit of dealer f is defined as

$$\pi_f(p_t) = \sum_{j \in \mathcal{J}_{ft}} (p_{jft} - mc_{jft}) q_{jf}(p_t),$$

where mc_{jft} is the constant marginal cost of product j sold by dealer f in year t. This marginal cost represents the wholesale (or "invoice") price of the car, along with other variable costs or benefits associated with car retailing, including the future warranty and service contracts for the car and the opportunity costs of the inventory management problem faced by the

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

dealer. See Albuquerque and Bronnenberg (2012) and Murry (2015) for a discussion of wholesale prices and retailing costs.

Dealers simultaneously set prices to maximize their own profits, taking into account prices and attributes of competing dealers. The first order condition for a particular dealer that defines a Nash equilibrium in prices is

$$q_{jf}(\mathbf{p}_t) + \sum_{j \in J_{ft}} (p_{jft} - mc_{jft}) \frac{\partial q_{jf}(\mathbf{p}_t)}{\partial p_{jft}} = 0. \tag{13}$$

Let $\triangle$ denote the demand price derivative matrix with the row $k = (j, f)$ and column $k' = (j', f')$ element:

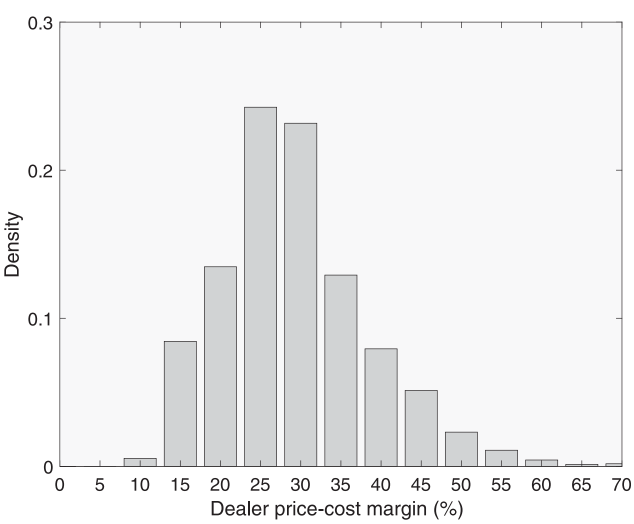
$$\triangle_{kk'} = \frac{\partial q_{jf}(\mathbf{p}_t)}{\partial p_{j'f't}} = \begin{cases} \sum_z b_{zt} \int (\bar{\alpha} + \alpha^{inc} y_{zt}) \mathcal{P}_{izjft}(1 - \mathcal{P}_{izjft}) dF_{\varsigma}(\cdot), & \text{if } k = k', \\ -\sum_z b_{zt} \int (\bar{\alpha} + \alpha^{inc} y_{zt}) \mathcal{P}_{izjft} \mathcal{P}_{izj'f't} dF_{\varsigma}(\cdot), & \text{if } k \neq k'. \end{cases}$$

We define an ownership matrix Ω^*, with $\Omega(j, j')^* = 1$ if product j and j' are sold by the same dealer and 0 otherwise. Let $\Omega = \Omega^* \times \triangle(p)$. Then, Equation (13) can be written in matrix notation as the following markup equation:

$$p - mc = \Omega^{-1} q(p). \tag{14}$$

From Equation (14), we compute the price-cost margins for each product sold by each dealer in each year, using the estimated demand parameters in Table 7. Figure 4 displays the distribution of dealer markups, defined as the ratio of price-cost margin over price. The weighted average markup (price-cost margin) is 29% and the median is 28%. These results are in line with other studies of the automobile industry, for example, 24% in Berry et al. (1995), 17% in Petrin (2002), \$6,220 (price minus marginal cost) in Albuquerque and Bronnenberg (2012), \$5,238 in Murry (2015), and 43% in Nurski and Verboven (2016) and Moraga-González et al. (2015).

Figure 4. Predicted Dealer Markup (Percentage)



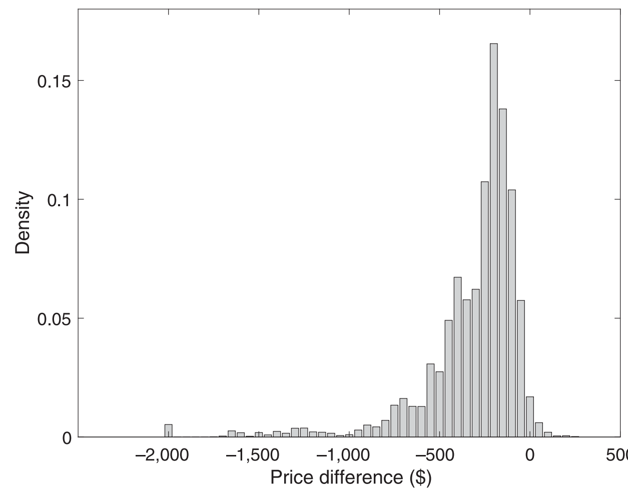
5.2. Impacts of Search Frictions

To quantify the impact of search frictions, we simulate the market outcomes assuming that consumers have full information along with the estimated preference parameters in Table 7. One way to do this in our model is the following: recall from Section 4 that when $\gamma = 0$, $\rho = 0$, and $\kappa = 1$, the individual choice probabilities from our search model are equivalent to those from a full-information model with a mean utility from outside option being log(2). Using this logic, our simulation results imply that in the full-information case, the total sales will be 33% higher than the sales predicted by our search model, and the weighted average price will be \$333 lower. Figure 5 presents the distribution of the difference between the product-dealer-year level price predicted by the full-information model and that predicted by the search model. The standard deviation is \$367, indicating that the price impacts caused by the search frictions are significantly different across products and dealers.

Next, we simulate the equilibrium price and sales by varying features of the search cost: the effect of distance,γ, the standard deviation of search-cost shocks κ, and the dealer-cluster size coefficient ρ, while holding other model parameters at their estimated values in Table 7. We report the price and sales impacts in Table 8 and Table 9.

In panel A, we report the results when we scale γ from the estimated values from 0 to twice while keeping other parameters equal to their estimates.

Figure 5. Price Difference in a Full-Information Model



Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

Table 8. Price Difference by Varying Search Cost Parameters

Panel A. Varying γ			Panel B. Varying κ			Panel C. Varying ρ		
Scale of $\hat{\gamma}$	Mean ($)	SD ($)	Scale of $\hat{\kappa}$	Mean ($)	SD ($)	Value of ρ	Mean ($)	SD ($)
0	–109	(216)	0.1	+104	(281)	0.1	143	(127)
0.25	–90	(181)	0.5	+53	(127)	0.2	244	(199)
0.50	–61	(130)	1.5	–28	(68)	0.3	314	(259)
0.75	–29	(67)	2	–44	(105)	0.4	365	(306)
1.25	+26	(65)	2.5	–54	(128)	0.5	404	(344)
1.50	+49	(123)	3.0	–60	(143)	0.6	438	(377)
1.75	+69	(175)	3.5	–64	(154)	0.7	460	(404)
2	+84	(218)	4	–68	(161)	0.8	481	(427)

Notes. Price difference is the difference in the sales-weighted average price between a model with the stated parameter changed and a model with all parameters equal to their estimated values reported in Table 7. SD, standard deviation.

The weighted average price will be lower (higher) and total sales will be higher (lower) as we scale down (up) γ. This is expected, because a higher γ discourages consumers from visiting those dealers that are farther away from their residence and leads to greater local monopoly power, and hence higher prices, for dealers because nearby residents are more captive.[20] In particular, when consumers' disutility from distance doubles the estimated value, the weighted average price will be $84 higher and the total sales will be 32% lower than the predicted when γ equals the estimated value. The standard deviation of the price difference is $218, indicating a significant heterogeneity across products and dealers. In contrast, when consumers have no disutility from distance, the weighted average price will be $109 lower and the total sales will be 38% higher than in the case when γ equals the estimated value.

Panel B reports the price and sales impacts when we scale κ from 0.1 to four times of the estimated value while keeping other parameters equal to their estimates. The weighted average price will be higher (lower) and the total sales will be lower (higher) when we scale down (up) κ, which is consistent with our analysis in Section 3.4. A larger κ implies more randomness in consumers' choice of which clusters to search. Consequently, a dealer's location is less relevant in consumers' choice, and this reduces dealers' local monopoly power. For example, compared with the predicted price and sales with κ at its estimated value, the weighted average price will be $68 lower with a standard deviation of $161, and the total sales will be 17% higher, when κ is four times of its estimated value. In contrast, when κ is 0.1 of its estimated value, the weighted average price will be $104 higher with a standard deviation of $281, and the total sales will be 10% lower.

Panel C reports the results when we set ρ (the coefficient before the number of dealers in a cluster) to various values from 0 to 0.8. A larger ρ implies that consumers pay larger search costs, resulting in less search, worse matches, and lower sales. Moreover, a larger ρ discourages consumers from traveling to search those larger but farther clusters, leading to greater local monopoly power and hence higher prices for nearby dealers. The effect of the higher prices reinforces the decrease in total sales. For example, when ρ equals 0.2, the weighted average price will be $244 higher with a standard deviation of $199, and the total sales will be 66% lower, than that when ρ equals its estimated value.

Table 9. Total Sale Change by Varying Search Cost Parameters

Panel A. Varying γ		Panel B. Varying κ		Panel C. Varying ρ	
Scale of $\hat{\gamma}$	Change (%)	Scale of $\hat{\kappa}$	Change (%)	Value of ρ	Change (%)
0	+38	0.1	–10	0.1	–44
0.25	+31	0.5	–7	0.2	–66
0.50	+22	1.5	+6	0.3	–78
0.75	+11	2	+10	0.4	–85
1.25	–10	2.5	+13	0.5	–89
1.50	–18	3.0	+15	0.6	–92
1.75	–25	3.5	+16	0.7	–94
2	–32	4	+17	0.8	–95

Note. The total sales change is the change of the total units sold predicted by a model with the stated parameter changed from that predicted by a model with all parameters equal to their estimated values reported in Table 7.

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

5.3. Single Dealer Closure

Next we use our structural model to study the effects of closing a single dealer. As we have discussed above, our proposed model implies that closing a dealership generates two effects to the remaining dealers located in the same cluster. On one hand, a closure may reduce the total attraction of the cluster, and thus reduce the sales of other dealers in this cluster through decreased consumer search. This effect puts pressure on dealers to reduce their prices to counter the negative impact of reduced search because lower prices attract more searching consumers. This is the agglomeration effect. On the other hand, closing a dealer will directly reduce the price competition among dealers in the same cluster and create an incentive for higher equilibrium prices. This is the competition effect.

5.3.1. Detailing Two Large Dealers. We consider a hypothetical scenario in which a single dealer was closed in 2007. The two dealers we consider are two of the largest dealers in 2007, labeled dealer A and dealer B. Dealer A's share in its cluster was between 10% and 20%, whereas dealer B's share in its cluster was above 50%.[21] To separately quantify the agglomeration and competition effects of dealer colocation we simulate changes to equilibrium prices and sales for the following three scenarios. In the first scenario, we quantify the agglomeration effect by allowing consumers to respond to the dealer closure by adjusting only the choice of each search set ($\mathcal{P}_{izS}$), but not the choice probability of purchasing each product conditional on a search set ($\mathcal{P}_{izjft|S}$). In this scenario, retailers adjust their prices only to attract new searchers to their dealer cluster, but not to compete against a rival in the same geographic cluster. In the second scenario, we quantify the competition effect by allowing consumers to respond to the dealer closure by adjusting only the choice of car conditional on a search set ($\mathcal{P}_{izjft|S}$), but not the choice of clusters to search ($\mathcal{P}_{izS}$). Accordingly, retailers will adjust their prices to compete with local rivals but not to attract more consumers to search their geographic cluster. In the last scenario, we allow consumers to respond to the closure by adjusting both their choice sets and the choice of car conditional on the choice set, which quantifies the total effect.

Table 10 reports the impacts of closing a dealer on the price, total sales, and total profit of two groups of dealers. The first group of dealers are those that are colocated with the closed dealer, and the second group includes all other dealers that are not located in the cluster where the closed dealer was located. In each counterfactual scenario, we simulate the equilibrium price and sales and compute the profit for each product–dealer combination. We take the difference of the simulated price over the observed price for each product–dealer. Then, we compute the sales-weighted average and standard deviations, and we report the results in the first and second columns. We also compute the total sales and total profit in the counterfactual scenario and report the

Table 10. Impacts of Closing One Dealer

	Price ($)		Total sales (%)	Total profit (%)
	Panel A. Closing dealer A			
Colocated dealers				
Agglomeration effect	−8.78	(5.16)	−10.94	−11.05
Competition effect	+10.93	(5.52)	+8.80	+8.97
Total effect	+1.06	(0.24)	−3.15	−3.14
Noncolocated dealers				
Agglomeration effect	+0.46	(1.95)	+0.28	+0.27
Competition effect	+0.17	(0.89)	+0.17	+0.16
Total effect	0.62	(2.66)	+0.43	+0.42
	Panel B. Closing dealer B			
Colocated dealers				
Agglomeration effect	−40	(37)	−26.42	−26.08
Competition effect	+502	(206)	+87.08	+97.43
Total effect	+451	(131)	+34.29	+41.05
Noncolocated dealers				
Agglomeration effect	+5.05	(21.44)	+0.49	+0.55
Competition effect	+3.14	(19.54)	+0.28	+0.32
Total effect	+8.03	(37.08)	+0.71	+0.82

Notes. The first column reports the sales-weighted average of the price difference at the product-dealer level between the counterfactual scenario and the data. The second column reports the standard deviation of that price difference. The third column reports the difference in total sales between the counterfactual scenario and the data. The forth column reports the difference in total profit between the counterfactual scenario and the data.

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

differences over the observed ones in the third and forth columns.

Because of the agglomeration effect, closing dealer A or dealer B will reduce the total attractiveness of the cluster and hence reduce the total sales of other colocated dealers (by −10% and −26%, respectively). Moreover, the closure will also induce the colocated dealers to reduce their prices to counter the reduced attractiveness of the cluster (by −\$8 and −\$40, respectively). Therefore, our results imply that in both cases of closing dealer A and closing dealer B, the agglomeration effect is positive, that is, the cluster of the closed dealer will be less likely included in consumers' search set. On the other hand, because of the competition effect, colocated dealers will be able to charge higher prices and sell more when a nearby rival exits. Other dealers not located in the same cluster will benefit from the closure both because their clusters become more attractive to consumers to visit and also because the competition becomes less intense. Unsurprisingly, the effects on nonlocated dealers are much smaller than on those colocated dealers.

Moreover, because dealer B plays a larger role in its cluster than dealer A, closing dealer B will have larger impacts on colocated dealers through both agglomeration and competition effects. Taking these two effects together, closing dealer A has a net negative effect on the incumbent dealers, suggesting the agglomeration effect dominates in this case. However, closing dealer B will slightly benefit the colocated dealers, suggesting the competition effect dominates in this case.

5.3.2. Overall Effects of Single Dealer Closures. Next, we close all dealers that existed in 2007, one at a time. This exercise informs us of the distribution of effects that could happen across different types of dealers, if they were to close. For example, we would expect the agglomeration effect to be different for a small dealer with few neighbors than for a large dealer with many neighbors. To do this, we simulate the equilibrium market outcomes by closing one dealer at each time. We do this separately for every dealer that sold cars in 2007. In Table 11, we describe the total effect of the closures by looking at the distribution across the single closures. The average dealer in 2007 had 406 sales and sold cars at an average price of \$29,142. On average, closing a dealer results in a −3.42% decrease in sales for the other dealers in the colocated cluster. This suggest that dealer B, above, is a clear outlier, with a large increase in affiliated-cluster sales. as for the results above, price changes are very small for any dealer closure, and can be either negative or positive. Most dealer closures result in the agglomeration effect outweighing the competition effect, as roughly 90% of dealer closures result in fewer sales for the remaining colocated dealers. There are some dealer closures that result in a substantial decrease in sales for remaining neighboring dealers, with over 25% of closures leading to greater than 4% fewer sales.

To gain a deeper insight into these results, we correlate the closure effect with characteristics of the closed dealer. In particular, we consider the size of the closed dealer (total 2007 sales) and the relative importance of the dealer in the cluster (own sales over total cluster sales). The results are presented in Figure 6. First, in Figure 6(a), the importance of a dealer in a cluster is clearly negatively correlated with a change in sales. The same is true with the dealer size (unconditional on within cluster importance), shown in Figure 6(b), although dealer B from above is clearly an outlier, visible in the top right corner of panel (b). The main takeaway is that if a dealer closes, the neighboring dealers will be worse if the closed dealer was relatively big within the cluster.

5.4. Terminations of Pontiac and Saturn

In this section, we focus on the termination of Pontiac and Saturn in 2009. Because the brand terminations were nationwide decisions, the dealer closures caused by the brand terminations were plausibly not correlated with the local conditions. We first simulate the market price and sales for other dealers in 2007 by hypothetically assuming that the termination of Pontiac occurred before 2007, and then compare the simulated outcomes with the observed ones. We do the same simulations assuming Saturn was closed in 2007. We decompose the effect of closures in the agglomerations effect, competition effect, and total effect like in the previous exercise in Section 5.3.

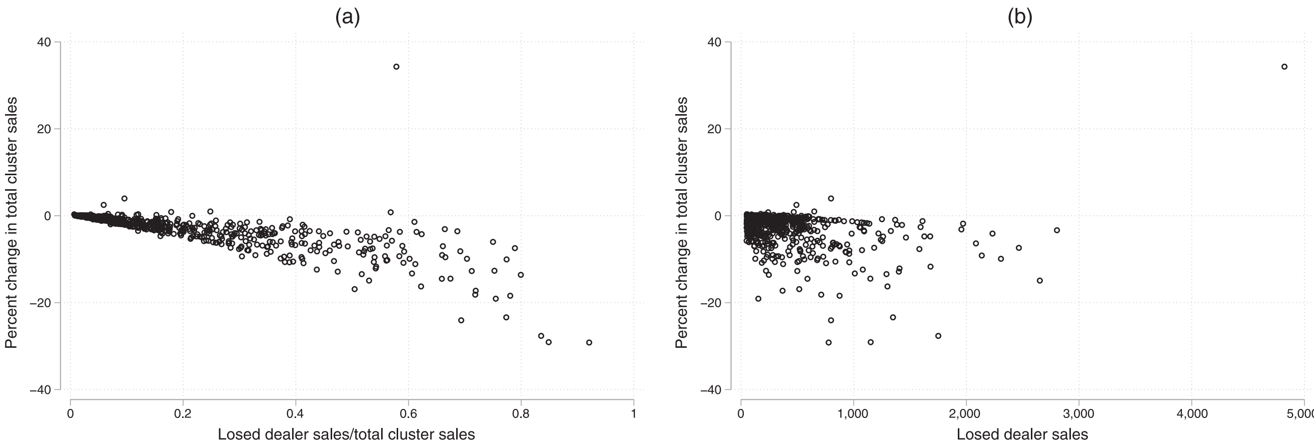
Table 11. Distribution of Effects After Single Dealer Closures

	Mean	SD	Q25	Q50	Q75
Closed dealers					
Total sales (2007)	406	404	142	286	519
Average price (\$)	29,142	6,091	25,910	27,590	30,067
Colocated dealers					
Total % change: Quantity sold (%)	−3.24	4.06	−4.27	−1.90	−0.63
Total % change: Average price (%)	−0.01	+0.09	−0.03	−0.01	+0.01

Note. SD, standard deviation; Q, percentile.

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

Figure 6. Change in Sales of All Neighboring Dealers After Single Dealer Closures



Relative size within cluster | Dealer sales

Table 12, panels A and B, reports the impacts of closing Pontiac and Saturn on the price, total sales, and total profit of the colocated dealers and noncolocated dealers. On the one hand, the agglomeration effect implies that those clusters would be less attractive to consumers and hence would reduce the total sales of those clusters. To counter this effect, the colocated dealers in those clusters would cut their prices. Our simulation results suggest that because of the agglomeration effect, hypothetically closing Pontiac would induce the dealers colocated with Pontiac dealers to cut their prices by $19 per car on average. Their total sales would be 2.14% lower, and their profit would be 2.31% lower. Meanwhile, the weighted average price, total sales, and profits of the dealers in other clusters would be higher, because those clusters would become more attractive to consumers. In the counterfactual scenario of closing Saturn, the colocated dealers suffer even more and the noncolocated dealers benefit less, indicating that the agglomeration effect is stronger in this counterfactual scenario. In short, we find that the agglomeration effect is positive in both counterfactual experiments of Pontiac closure and Saturn closure. Therefore, a full-information model that ignores the agglomeration effect would definitely overstate the positive impact of dealer closure on remaining colocated dealers.

Table 12. Impacts of Brand Closures

	Price ($)		Total sales (%)	Total profit (%)
Panel A. Closing Pontiac				
Colocated dealers				
Agglomeration effect	−19	(62)	−2.14	−2.31
Competition effect	+2	(51)	+3.24	+3.20
Total effect	−3	(58)	+0.92	+0.87
Noncolocated dealers				
Agglomeration effect	+9	(10)	+1.56	+1.68
Competition effect	+7	(6)	+1.46	+1.54
Total effect	+15	(15)	+2.90	+3.10
Panel B. Closing Saturn				
Colocated dealers				
Agglomeration effect	−10	(17)	−4.44	-4.57
Competition effect	+15	(18)	+4.49	+4.70
Total effect	+5	(16)	−0.24	−0.17
Noncolocated dealers				
Agglomeration effect	+6	(7)	+1.35	+1.39
Competition effect	+3	(4)	+1.18	+1.20
Total effect	+9	(10)	+2.42	+2.48

Notes. The first column reports the sales-weighted average of the price difference at the product-dealer level between the counterfactual scenario and the data. The second column reports the standard deviation of that price difference. The third column reports the difference in total sales between the counterfactual scenario and the data. The forth column reports the difference in total profit between the counterfactual scenario and the data.

On the other hand, closing a brand benefits colocated dealers because fewer competitors implies larger market power and more sales. In the counterfactual scenario of closing Pontiac, those colocated dealers would slightly increase their prices, and their total sales would be 3.24% higher. As a result, their profit would be 3.2% higher. Meanwhile, those noncolocated dealers would be also better off because of the lower competition, although they would benefit less than the colocated dealers would. Similarly, colocated dealers and noncolocated dealers are better off when we hypothetically close Saturn.

Taking these two effects together, closing Pontiac would make the colocated dealers slightly better off, whereas closing Saturn would slightly reduce the total profit of the colocated dealers. Our results imply that the competition effect dominates when Pontiac is closed, whereas the agglomeration effect dominates when Saturn is closed. This is quite different from the findings of other studies on dealer closures following the financial crisis. For example, Ozturk et al. (2016)

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

found evidence that the competition effect dominated when Chrysler closed dealers in 2010. Benmelech et al. (2014) documented massive retail exits during the financial crisis, for financial reasons such as bankruptcy, and found evidence for an agglomeration effect of dealer closure.[22]

The difference in the results between Pontiac and Saturn is intuitive given that these brands had much different retailing arrangements. In Ohio, 16 out of 20 dealers selling Saturn were single-brand dealers, whereas 18 out of 76 dealers selling Pontiac were single branded. Partly because of that, Saturn dealers were of more importance within the clusters where they were located than Pontiac dealers. For example, Saturn's actual sales accounted for more than 12% of all sales in their clusters, whereas Pontiac's within-cluster share was 8%. As a result, closing Saturn would have larger impacts on colocated dealers both through the agglomeration effect and through the competition effect, which is shown in Table 12. Moreover, the agglomeration effect would outperform the competition effect on colocated dealers if Saturn were closed.

6. Conclusion

In this paper, we present a structural model of consumer search for spatially differentiated products in the new car retail industry. The model explicitly captures the agglomeration and competition effects of retail colocation. We estimate the model using detailed data on all new car transactions in a single U.S. state. Our approach contributes to the literature on consumer demand with limited information and the literature on retail agglomeration.

Our results indicate that consumers' search cost is $45 per mile on average, and because of their substantial search cost, half of them only search one geographic cluster before purchase. We also show that the average price-cost markup is $333 higher in the presence of search frictions. Moreover, our counterfactual analysis suggests that both the competition and agglomeration effects matter after a dealer closure. We show that the agglomeration effect becomes stronger when consumers get larger disutility from traveling distance or when their choice of which cluster(s) to search becomes less random. In general, the results suggest that whether the agglomeration effect outweighs the competition effect largely depends on the substitution patterns across dealer clusters. We think this insight is an important result that can help inform policy makers and managers about the effects of colocation and contagion of retail closures.

Our main finding is that the agglomeration effect is positive: there are negative effects on dealers of a colocated dealer closure. Because competition also plays a role, we find cases where there exist only moderate negative net effects of closures on colocated dealers, and some cases where the competition effect dominates and closures are a net positive for colocated incumbents. However, our finding that the agglomeration effect could dominate the competition effect has important implications for policy makers and managers. For example, our results rationalize the colocation of rivals and suggest that colocated stores act (partially) as complements, and as such, rivals may not want to force closures. Also, the results suggest that if the agglomeration effect dominates, then local governments that often help organize retail landscapes should be worried about closure spirals, as each retail closure has a net negative effect on remaining stores.

To be sure, our analysis relies on particular assumptions. Although we are confident that our model captures the major features of this industry, some caveats are worth mentioning. First, although the evidence we present suggests dealer agglomeration is an important consideration during consumers' car-buying process, the search process in reality may be more complicated than our model presents. In particular, the recent proliferation of car-buying websites that aim at providing consumers with more information has likely started to change the way consumers search for cars. However, cars are experience goods, so websites could never fully inform a consumer completely about the utility, as personal interaction can. Second, consumers may search in a more complicated way, for example, nesting geographical concerns with the search for a dealer (as in Moraga-González et al. 2015) and the search for a car type. Because we do not observe search behavior explicitly, we are unable to separately identify different search mechanisms. Third, although we present a demand-driven reason for dealers to colocate, there are likely cost-driven reasons, for example, land prices, zoning, and management convenience for multidealership dealer conglomerates. Our analysis is not a full equilibrium analysis of retail location decisions and cannot be used to balance all the trade-offs associated with the optimal location decision. Instead, we focus on identifying the importance of demand-side motives that have been identified in the theoretical literature as being important determinants of colocation.

Acknowledgments

The authors thank Simon Anderson, Panle J. Barwick, Ying Fan, Paul Grieco, Jean-Francois Houde, Peter Newberry, Regis Renault, John Rust, Henry Schneider, Steven Stern, Matthijs Wildenbeest, Mo Xiao, Daniel Xu, various seminar participants, the editor Matthew Shum, an anonymous associate editor, and three anonymous referees for helpful suggestions.

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

Appendix A. Proofs of Proposition 1 and 2

Proof of Proposition 1

Without loss of generality, we will show that the proposition holds for cluster 1.

The probability of visiting a firm in cluster 1 is

$$\mathcal{P}_1+\mathcal{P}_{1\&2}=1-\frac{e^{\kappa}+\frac{n_2e^{\mu}+1}{e^{(\gamma d_2+\rho n_2)}}}{e^{\kappa}+\frac{n_1e^{\mu}+1}{e^{(\gamma d_1+\rho n_1)}}+\frac{n_2e^{\mu}+1}{e^{(\gamma d_2+\rho n_2)}}+\frac{(n_1+n_2)e^{\mu}+1}{e^{\gamma(d_1+d_2)+\rho(n_1+n_2)}}}.$$

Let Γ denote the denominator $e^{\kappa}+\frac{n_1e^{\mu}+1}{e^{(\gamma d_1+\rho n_1)}}+\frac{n_2e^{\mu}+1}{e^{(\gamma d_2+\rho n_2)}}+\frac{(n_1+n_2)e^{\mu}+1}{e^{\gamma(d_1+d_2)+\rho(n_1+n_2)}}$. Then we have

$$\frac{d(\mathcal{P}_1+\mathcal{P}_{1\&2})}{dn_1}=\frac{e^{\kappa}+\frac{n_2e^{\mu}+1}{e^{(\gamma d_2+\rho n_2)}}}{\Gamma^2}\frac{d\Gamma}{dn_1}.$$

Here,

$$\frac{d\Gamma}{dn_1}=\frac{e^{\mu}-(n_1e^{\mu}+1)\rho}{e^{(\gamma d_1+\rho n_1)}}+\frac{e^{\mu}-[(n_1+n_2)e^{\mu}+1]\rho}{e^{\gamma(d_1+d_2)+\rho(n_1+n_2)}}>0$$
$$\Leftrightarrow [(n_1+n_2)e^{\mu}+1]\rho+(n_1e^{\mu}+1)\rho e^{\gamma d_2+\rho n_2}<(1+e^{\gamma d_2+\rho n_2})e^{\mu}.$$

First, the above inequality holds when $\rho=0$. Second, the left-hand side is clearly above the right-hand side when ρ goes to infinity. Third, the left-hand side is an increasing function of ρ. Hence, there must exists a positive cutoff ρ^* such that as long as $\rho<\rho_1^*$, $\frac{d\Gamma}{dn_1}>0$, and hence $\frac{d(\mathcal{P}_1+\mathcal{P}_{1\&2})}{dn_1}>0$; otherwise, $\frac{d\Gamma}{dn_1}\le 0$, and hence $\frac{d(\mathcal{P}_1+\mathcal{P}_{1\&2})}{dn_1}\le 0$. Moreover, it is easy to show that the cutoff ρ_1^* is decreasing in n_1 and n_2.

Proof of Proposition 2

First, we can easily show that $\frac{d(\mathcal{P}_1+\mathcal{P}_{1\&2})}{dn_1}\longrightarrow 0$ as $\kappa\longrightarrow\infty$ and that $\frac{d(\mathcal{P}_1+\mathcal{P}_{1\&2})}{dn_1}>0$ as $\kappa\longrightarrow 0$.

We take partial derivative of the agglomeration effect $\frac{d(\mathcal{P}_1+\mathcal{P}_{1\&2})}{dn_1}$ with respect to κ.

$$\begin{aligned}\frac{d^2(\mathcal{P}_1+\mathcal{P}_{1\&2})}{dn_1d\kappa}&=\frac{e^{\kappa}}{\Gamma^3}\left[\Gamma-2e^{\kappa}\left(e^{\kappa}+\frac{n_2e^{\mu}+1}{e^{(\gamma d_2+\rho n_2)}}\right)\right]\frac{d\Gamma}{dn_1}\\&\propto\Gamma-2e^{\kappa}\left(e^{\kappa}+\frac{n_2e^{\mu}+1}{e^{(\gamma d_2+\rho n_2)}}\right)\\&=-2(e^{\kappa})^2-(2B-1)e^{\kappa}+A,\end{aligned}$$

where $A\equiv\Gamma-e^{\kappa}$ and $B\equiv\frac{n_2e^{\mu}+1}{e^{(\gamma d_2+\rho n_2)}}$.

Define $\kappa^*=\ln\{\frac{\sqrt{(2B-1)^2+8A}-(2B-1)}{4}\}$. Then, $\frac{d^2(\mathcal{P}_1+\mathcal{P}_{1\&2})}{dn_1d\kappa}\ge 0$ when $\kappa\le\kappa^*$, and $\frac{d^2(\mathcal{P}_1+\mathcal{P}_{1\&2})}{dn_1d\kappa}<0$ when $\kappa>\kappa^*$.

Appendix B. Derivation of the Second-Stage Estimates

First Stage

Estimated nonlinear parameter $\hat{\theta}_1$ maximizes the log-likelihood function in Equation (10). Asymptotically,

$$\sqrt{N_1}(\hat{\theta}_1-\theta_1)\sim N(0,\Sigma).$$

Then, we obtain the estimated mean utility $\hat{\delta}$, which is a function of estimated nonlinear parameters $\hat{\theta}_1$, denoted by $\hat{\delta}=g(\hat{\theta}_1)$.

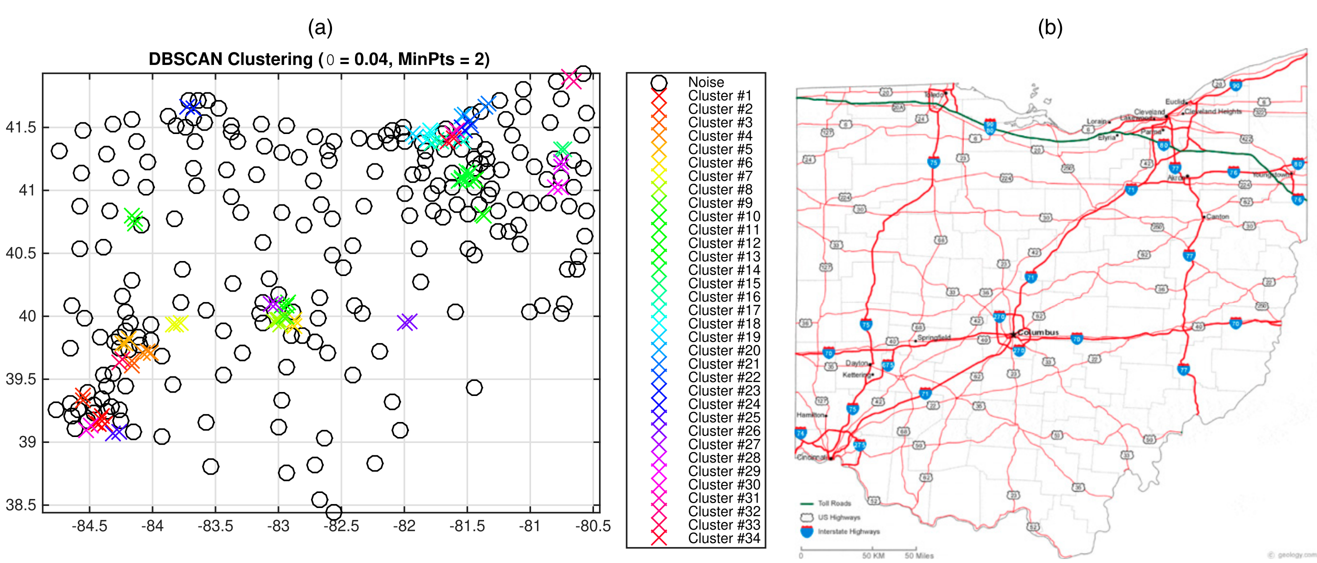
Let $C(\delta)$ denote the variance–covariance of δ. It is

$$C(\delta)=\left[\frac{\partial g(\hat{\theta}_1)}{\partial\theta_1}\right]\Sigma\left[\frac{\partial g(\hat{\theta}_1)}{\partial\theta_1}\right]'.$$

Here, $g(.)$ can be obtained from the contraction mapping $\Gamma(\delta,\theta_1)=0$ that sets the predicted market shares equal to the observed market shares. Therefore,

$$\frac{\partial g(\hat{\theta}_1)}{\partial\theta_1}=-\left[\frac{\partial\Gamma(\delta,\theta_1)}{\partial\delta}\right]^{-1}\left[\frac{\partial\Gamma(\delta,\theta_1)}{\partial\theta_1}\right].$$

Appendix C. (Color online) Spatial Distribution of Ohio Dealer Clusters

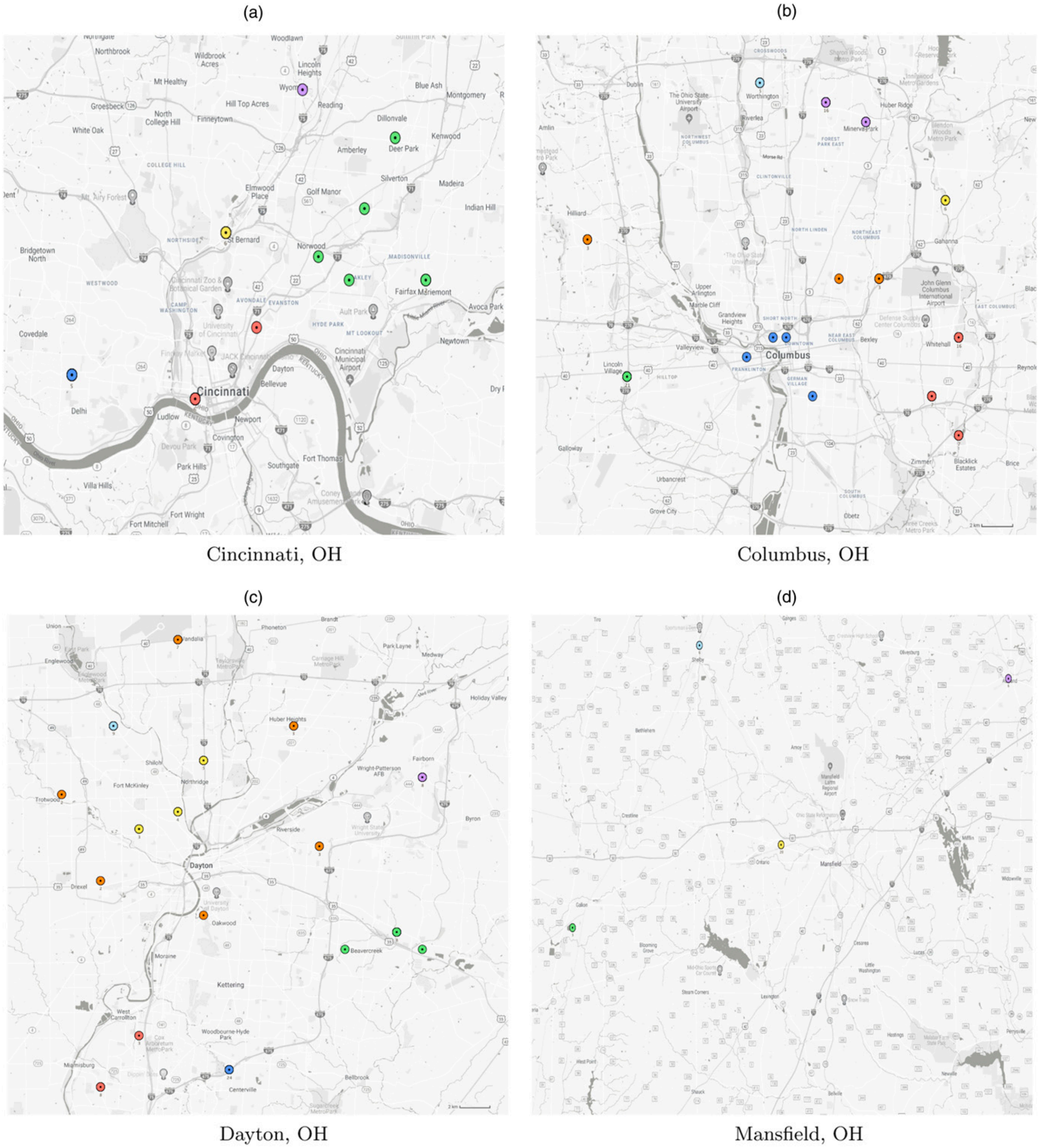


Results of DBSCAN Clustering   Map of Ohio

Note. The open circles indicate single-dealer clusters, and the colored crosses indicate multidealer clusters.

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

Appendix D. (Color online) Dealer Clusters in Four Ohio Locations



(a) Cincinnati, OH

(b) Columbus, OH

(c) Dayton, OH

(d) Mansfield, OH

Second Stage

Estimated linear parameter $\hat{\theta}_2$ solves the moment conditions (12). Asymptotically,

$$\sqrt{N_2}(\hat{\theta}_2 - \theta_2) \sim N(0, \Omega),$$

where Ω is the corrected covariance–variance of $\hat{\theta}_2$. Let $C(\xi)$ denote the covariance–variance of ξ. Then,

$$\Omega = (X'P_Z X)^{-1} X'P_Z[C(\xi) + C(\delta)]P_Z' X(X'P_Z' X)^{-1},$$

where $P_Z = Z(Z'Z)^{-1}Z'$ is the projection matrix.

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

Appendix E. First-Stage Price Regression

		Exogenous variables			
Variable		IV1	IV2	IV3	IV4
log(*Acceleration*)	3.4567	0.0073	−1.1521	−2.5431	2.2650
	(0.1809)***	(0.0987)	(0.1081)***	(0.2069)***	(0.1061)***
log(*Car size*)	−0.6958	−0.1468	1.2764	3.6415	3.3798
	(1.1531)	(0.1704)	(0.1525)***	(1.1667)***	(0.1463)
log(*Miles per dollar*)	−0.5109	−0.2047	0.0207	−0.2437	−0.0279
	(0.0608)***	(0.1502)	(0.1505)	(0.1610)	(0.1473)
Luxury brand	4.8530	−0.0139	−0.2454	−4.4358	0.8730
	(0.4045)***	(0.0513)	(0.0325)***	(0.3782)***	(0.2073)***
U.S. brand	−2.1417	−0.1679	−0.0913	2.3655	0.9695
	(0.1540)***	(0.0326)***	(0.0223)***	(0.1543)***	(0.0994)***

Overall $R^2 = 0.7131$
Weak identification test: Cragg–Donald Wald F statistic = 1,273.077
Overidentification test: p-value = 0.0000

Notes. The table shows dealer fixed effect regression of price at the product-dealer-year level on the exogenous variables, including log of acceleration, log of car size, log of miles per dollar, luxury brand, U.S. brand, body style dummies, yearly dummies, and instrumental variables. Log of acceleration, log of car size, and log of miles per dollar are normalized. IV, instrumental variable. IV1 values are the deviations from the average characteristics of product-dealer combinations available in the same dealer cluster and in the same year, IV2 values are the squares of IV1 values, IV3 values are the deviations from the average characteristics of all product–dealer combinations in the same year, and IV4 values are the squares of IV3 values. There were 101,371 product-dealer-year-level observations. $R^2 = 0.71$.

Endnotes

1 The model that we present and the subsequent results rely on the assumption that consumers search all dealers within a cluster. This assumption could be relaxed in future work with appropriate data.

2 We should note that we *do not* explicitly test our assumption of cluster search and that our estimates come from a model that is identified subject to all of the behavioral and parametric assumptions we make. However, we do provide descriptive results that suggest dealer clustering and consumer travel are important features of this market in Section 2.

3 In an updated version of their paper, Moraga-González et al. (2017) estimate a sequential search model using the same automobile data. One innovative result from their exercise is that their expression for choice probabilities looks very similar to the expression in the case of simultaneous search.

4 It is common in the literature to consider pickup trucks a different market. Additionally, some models of pickup trucks have dozens of trim levels that vary widely in price and characteristics, making it problematic to aggregate to the model level.

5 Luxury brands include Acura, Audi, BMW, Buick, Cadillac, Infiniti, Lexus, Lincoln, Mercedes-Benz, Porsche, and Volvo. U.S. brands include traditionally U.S. brands that are no longer U.S. owned, like Chrysler.

6 Saab, the major Swedish-produced car brand, was owned by General Motors until 2011. After 2011, the company reorganized, and it started producing cars again in 2014.

7 Both companies, along with Ford Motors, had a clear policy to create smaller dealer networks, but were generally unable to do so because state regulations prohibit dealer franchise contract termination by manufacturers in the automobile industry. Dealers lobbied against dealer closures, citing existing state regulations that prohibit closures. Many of the proposed closures (from both of the reasons stated above) went into legal arbitration. For example, when GM closed the Oldsmobile brand, they reportedly paid over $1 billion to their dealers. For a deeper discussion of the political economy, see Lafontaine and Morton (2010). For example of popular press coverage of dealer closures, see Terlep (2009).

8 In principle, DBSCAN can handle these isolated dealers and classify them as noise, but the algorithm was much more robust to parameter choices when we preclassified isolated dealers in this manner. We use only the latitude and longitude of dealers to classify them into groups. Like other classification algorithms, we could add other geographic features to the classification objective, and ideally one would use travel patterns data to train the algorithm. The advantage DBSCAN has over k-means clustering, for example, is that there is no need to specify the number of clusters ex ante. In practice, we use the Euclidian distance and set $\epsilon = 4\,kilometers$ and the minimum number of points equal to two. See Ester et al. (1996) or various programming languages' implementations (e.g., the "sklearn.cluster" package for Python) for more details of the algorithm.

9 Recently, Moraga-González et al. (2017) showed that a version of sequential search can be operationalized in an empirical model of differentiated product demand in a way similar to that used in previous papers that used simultaneous search, including their own earlier manuscript, Moraga-González et al. (2015).

10 We use the product-dealer-year-level average price mainly for two reasons. First, new car retail prices are always set through negotiation and hence may vary across dealers and across consumers. Modeling a bargaining protocol between consumers and dealers would severely complicate our search model, and we are not aware of other research that models search and negotiation in a differentiated product setting. Second, the use of individual prices directly would introduce a missing data problem. Because car prices are typically negotiated, we observe only the price of the car that a consumer eventually bought but not the prices faced by the consumer for other cars considered but did not purchase. Therefore, some assumption on the data-generating process for individual prices would be needed to estimate price elasticities.

11 There are two types of consumers who do not purchase: those who do not search at all and those who search and choose the outside option. If a consumer does not search at all, her search set is an empty set. In this case, her expected gain is zero, and her search cost equals a random variable, ω_{iz0t}. As long as a consumer does not search an empty set, she will get an expected gain $U_{izt}(S)$ and pay a search cost

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

$C_{izt}(S)$. Here, the expected gain $U_{izt}(S)$ includes the possibility that she ends up not buying. In particular, as we assumed, the utility from the outside option is ε_{iz0t}. Not searching and not buying are treated the same in the sense that both options are assumed to have mean zero utility. With data on failed searches, we could credibly separately identify the two means from each other, for example, as is done in Moraga-González et al. (2015).

[12] The specification is also similar to the limited information model in Sovinsky Goeree (2008), however, in that model choice sets are exogenous to the consumer.

[13] A particularly interesting case is when $\gamma = 0$, $\rho = 0$, and $\kappa = 1$. Let $\mathcal{S}_f$ denote the set of search sets that includes the dealer f's cluster. The individual choice probability in this particular case is

$$\begin{aligned}\mathcal{P}_{izjft} &= \sum_{S\in\mathcal{S}_i} \mathcal{P}_{izjft|S}\mathcal{P}_{izSt} = \sum_{S\in\mathcal{S}_f} \frac{\exp(\delta_{jft} + \mu_{izjft})}{\sum_{S'\in\mathcal{S}_i} \exp(U_{izt}(S'))}\\ &= \frac{2^{M-1}\exp(\delta_{jft} + \mu_{izjft})}{2^M + 2^{M-1}\left[\sum_{j',f'} \exp(\delta_{j'f't} + \mu_{izj'f't})\right]}\\ &= \frac{\exp(\delta_{jft} + \mu_{izjft})}{e^{\ln(2)} + \sum_{j',f'} \exp(\delta_{j'f't} + \mu_{izj'f't})}.\end{aligned}$$

Therefore, this case is equivalent to the standard full-information model with mean utility from the outside option being $ln(2)$.

[14] This is a well-known property of variants of the logit discrete choice model and has the flavor of "love of variety" in representative consumer models. See Anderson et al. (1992) for details of welfare in discrete choice demand models.

[15] We thank one of our referees for pointing this out.

[16] A previous working paper version of this paper used much fewer data (four years in a single city in Virginia) to estimate the model, and in that case, κ was very weakly identified. This reinforces the idea that variation in choice sets is crucial to identifying this parameter.

[17] Consider that a consumer bought a Toyota Camry. Because Moraga-González et al. (2015) do not observe which dealer the consumer purchased from, they assume that she purchased it from the closest Toyota dealer. In contrast, we incorporate exact purchase distance information, which may show that this consumer actually purchased from a dealer farther away than the closest Toyota dealer. See Table 3 for details.

[18] See http://www.dmeautomotive.com/announcements/1-in-6-car-buyers-skips-test-drive-nearly-half-visit-just-one-or-no-dealership-prior-to-purchase.

[19] Price here is the same concept as in the demand model: the average price for each product at each dealer in each year. Because of our data are at the transaction level, we can construct average prices for a given model at a given dealer. Therefore, the price of a single product differs across dealers.

[20] Less search also results in worse matches, which will put downward pressure on prices, but we find that the information rents dominate.

[21] Our point here is to illustrate how large the agglomeration and competition effects can be if an "anchor" store closes. Of course, if a small dealer is closed, both effects will be much smaller, although the net effect could go either way.

[22] For example, they document the complete liquidation of multiple large retailers, including Circuit City, Linens 'n Things, and The Sharper Image. Other large retail chains that experienced massive closings because of financial trouble include Kmart and Sears.

References

Albuquerque P, Bronnenberg B (2012) Measuring the impact of negative demand shocks on car dealer networks. *Marketing Sci.* 31(1):4–23.

Anderson SP, de Palma A, Thisse JF (1992) *Discrete Choice Theory of Product Differentiation* (MIT Press, Cambridge, MA).

Benmelech E, Bergman N, Milanez A, Mukharlyamov V (2014) The agglomeration of bankruptcy. NBER Working Paper No. w20254, National Bureau of Economic Research, Cambridge, MA.

Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica* 63(4):841–890.

Berry ST (1994) Estimating discrete-choice models of product differentiation. *RAND J. Econom.* 25(2):242–262.

Burdett K, Judd KL (1983) Equilibrium price dispersion. *Econometrica* 51(4):955–969.

Chade H, Smith L (2006) Simultaneous search. *Econometrica* 74(5): 1293–1307.

Datta S, Sudhir K (2013) Does reducing spatial differentiation increase product differentiation? Effects of zoning on retail entry and format variety. *Quant. Marketing Econom.* 11(1):83–116.

De los Santos B, Hortacsu A, Wildenbeest MR (2012) Testing models of consumer search using data on web browsing and purchasing behavior. *Amer. Econom. Rev.* 102(6):2955–2980.

Ellickson PB, Houghton S, Timmins C (2013) Estimating network economies in retail chains: A revealed preference approach. *RAND J. Econom.* 44(2):169–193.

Ester M, Kriegel HP, Sander J, Xu X (1996) A density-based algorithm for discovering clusters in large spatial databases with noise. *Proc. 2nd Internat. Conf. Knowledge Discovery Data Mining* (AAAI Press, Palo Alto, CA), 226–231.

Gandhi A, Houde JF (2017) Measuring substitution patterns in differentiated products industries. Working paper, University of Wisconsin–Madison, Madison.

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

Hong H, Shum M (2006) Using price distributions to estimate search costs. *RAND J. Econom.* 37(2):257–275.

Honka E (2014) Quantifying search and switching costs in the US auto insurance industry. *RAND J. Econom.* 45(4):847–884.

Honka E, Chintagunta PK (2016) Simultaneous or sequential? Search strategies in the US auto insurance industry. *Marketing Sci.* 36(1): 21–42.

Houde J-F (2012) Spatial differentiation and vertical mergers in retail markets for gasoline. *Amer. Econom. Rev.* 102(5):2147–2182.

Janssen M, Moraga-González JL (2004) Strategic pricing, consumer search and the number of firms. *Rev. Econom. Stud.* 71(4):1089–1118.

Jia P (2008) What happens when Wal-mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica* 76(6):1263–1316.

Lafontaine F, Morton FS (2010) Markets: State franchise laws, dealer terminations, and the auto crisis. *J. Econom. Perspect.* 24(3):233–50.

Mehta N, Rajiv S, Srinivasan K (2003) Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing Sci.* 22(1):58–84.

Moraga-González JL, Sándor Z, Wildenbeest MR (2015) Consumer search and prices in the automobile market. Working paper, Vrije Universiteit Amsterdam, Amsterdam.

Moraga-González JL, Sándor Z, Wildenbeest MR (2017) Consumer search and prices in the automobile market. Working paper, Vrije Universiteit Amsterdam, Amsterdam.

Murry C (2015) Advertising in vertical relationships: An equilibrium model of the automobile industry. Working paper, Penn State University, State College, PA.

Murry C (2018) The effect of retail competition on relationship-specific investments: Evidence from new car advertising. *Internat. J. Indust. Organ.* 59(July):253–281.

Murry C, Schnieder H (2016) The economics of retail markets for new and used cars. *Handbook on the Economics of Retailing and Distribution* (Edward Elgar, Cheltenham, UK).

Nevo A (2001) Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69(2):307–342.

Downloaded from informs.org by [136.167.230.79] on 07 December 2022, at 07:13 . For personal use only, all rights reserved.

Nurski L, Verboven F (2016) Exclusive dealing as a barrier to entry? Evidence from automobiles. *Rev. Econom. Stud.* 83(3):1156–1188.

Ozturk OC, Venkataraman S, Chintagunta PK (2016) Price reactions to rivals' local channel exits. *Marketing Sci.* 35(4):588–604.

Petrin A (2002) Quantifying the benefits of new products: The case of the minivan. *J. Political Econom.* 110(4):705–729.

Seiler S (2013) The impact of search costs on consumer behavior: A dynamic approach. *Quant. Marketing Econom.* 11(2):155–203.

Seim K (2006) An empirical model of firm entry with endogenous product-type choices. *RAND J. Econom.* 37(3):619–640.

Sovinsky Goeree M (2008) Limited information and advertising in the US personal computer industry. *Econometrica* 76(5):1017–1074.

Stahl K (1982) Location and spatial pricing theory with nonconvex transportation cost schedules. *Bell J. Econom.* 13(2):575–582.

Stern S, Zhou Y (2018) Antithetic acceleration and estimation of unobserved heterogeneity parameters. Working paper, Stony Brook University, Stony Brook, NY.

Stigler GJ (1961) The economics of information. *J. Political Econom.* 69(3):213–225.

Terlep S (2009) GM to close 1,100 dealerships. *Wall Street Journal* (May 16), https://www.wsj.com/articles/SB124238650972623589.

Valdes-Dapena P (2009) GM to pull the plug on Pontiac. *CNNMoney* (April 27), https://money.cnn.com/2009/04/24/autos/pontiac_obit/.

Vitorino MA (2012) Empirical entry games with complementarities: An application to the shopping center industry. *J. Marketing Res.* 49(2):175–191.

Wildenbeest MR (2011) An empirical model of search with vertically differentiated products. *RAND J. Econom.* 42(4):729–757.

Wolinsky A (1983) Retail trade concentration due to consumers' imperfect information. *Bell J. Econom.* 14(1):275–282.

Zhu T, Singh V (2009) Spatial competition with endogenous location choices: An application to discount retailing. *Quant. Marketing Econom.* 7(1):1–35.