LEVERAGING ONLINE CONSUMER INTEREST TRACKING DATA IN MARKET RESPONSE MODELING

A Dissertation

Presented to

the Faculty of the C.T. Bauer College of Business

University of Houston

In Partial Fulfillment

of the Requirement for the Degree

Doctor of Philosophy

By

Sina Damangir

June, 2014

DEDICATION

To Firoozeh Hoseini and Ebrahim Damangir, my parents, for their love and support.

To Soheil Damangir, my brother and my best friend.

Words cannot express how much I love you all.

ACKNOWLEDGMENT

I would like to thank my committee — Professors Edward Blair, Rex Yuxing Du, Ye Hu, Norm Johnson, and Seshadri Tirunillai — without whose advice, knowledge, and generosity this dissertation would not have been possible. But my foremost gratitude goes to my Co-chairs, Rex and Ye, for their patience, support, and guidance. Their enthusiasm for scholarship and willingness to help has been a great source of encouragement. I am truly fortunate to have had the opportunity to work with them. I owe a huge debt of gratitude to Michael Ahearne and Vanessa Patrick, my PhD coordinators during the program, for their support and providing a great environment for PhD students.

LEVERAGING ONLINE CONSUMER INTEREST TRACKING DATA IN MARKET RESPONSE MODELING

ABSTRACT

This dissertation develops techniques that use the information from online tracking data for analyzing market response. In theory, the observed market response originates from latent characteristics of the market such as consumers' preference for products and features and the competitive landscape. Understanding these latent characteristics is essential in making high quality marketing decisions. However, finding a reliable and inexpensive proxy for them is a challenge. We explore the possibility of using insights from "big data" sources to better identify these latent characteristics. We apply our techniques to analyze the market for automobiles in the US.

In Essay 1, we explore the potential of using trends in online searches for feature-related keywords as proxies for trends in the relative importance consumers place on the corresponding features. The relative importance consumers place on features may vary over time due to factors beyond the control of marketers (e.g., shifts in economic conditions, advances in technology). We make the baseline attractiveness of 70 top-selling automobiles a function of Google Trends indexes for five common features: fuel efficiency, acceleration, body type, cost to buy, and cost to operate. We find strong empirical evidence supporting the notion that the evolution of feature search intensity contains genuine information about shifting consumers' tastes.

In essay 2, we propose a model that identifies (1) the position of products on a latent perceptual map, (2) the consumer segments, and (3) the ideal point of the preference for each segment. The product positions are inferred using a novel approach using big data on online consumers' activities. We show that our proposed approach performs better than alternative approaches in identifying latent product positions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Essay 1

TRACKING MARKET SHARE DYNAMICS WITH GOOGLING FOR PRODUCT FEATURE

INTRODUCTION

Understanding the dynamics of market share is central to the practice of marketing. The most common approach is to treat the market share of a product as a function of its attractiveness in relation to that of its competition (Cooper and Nakanishi 1988). In modeling product attractiveness, marketers have often focused on two sets of drivers: the levels of features offered by the product (e.g., miles per gallon for cars, battery life for laptops), and the amount of marketing efforts promoting the product (e.g. advertising and incentives). In longitudinal market share analyses, because in most cases the feature levels of existing products do not change or change very little over time, the main focus has been on capturing the observed market share fluctuations as a result of shifts in marketing efforts. In other words, in most analyses of market share dynamics, the baseline attractiveness of a product (i.e., the part of product attractiveness that is not tied to marketing efforts) is treated as time-invariant (e.g., Ailawadi, Lehmann, and Neslin 2001; Neslin 1990).

In this study we take a different approach by allowing baseline product attractiveness to vary longitudinally even when the underlying feature levels by and large remain the same. This can occur because consumer needs and wants evolve over time. In particular, the relative importance of different features can change substantially over time, resulting in shifts in the relative standing of product baseline attractiveness. For example, growing environmental awareness coupled with new technology such as gasoline-electric hybrid can cause fuel economy for automobiles to receive increasingly more attention from consumers and weighs more prominently in their purchase decisions.

Products offerings with more appealing levels of that feature would benefit from the trend and become more attractive.

Given that the relative baseline attractiveness of a product can change as the prominence of different product features shift in consumer purchase decisions, modeling the dynamics of the product attractiveness requires a good tracking measure of the product features. To do so, one could potentially run longitudinal conjoint studies, producing a tracking measure of each feature's part-worth. The history of these part-worths would reveal how consumer preferences have evolved over time. One could include such evolution in market share analyses as a way to capture the dynamics of product attractiveness. However, using longitudinal conjoint (or other stated preference methods) to monitor which product features receive more or less attention from consumers could prove cost-prohibitive, especially if one wishes to track this information at a high frequency (say, monthly) and the number of respondents needed for a representative sample is large.

In the last few years many online consumer interest tracking services have emerged (e.g., nmincite.com, radian6.com, attensity.com, visibletechnologies.com, networkedinsights.com, sdl.com, converseon.com, synthesio.com, conversition.com, lithium.com). They provide a cost-effective platform for monitoring what consumers have to say, in their own language, about all sorts of products and services throughout the social and digital media sphere. The emergence of these services presents a potentially powerful alternative for tracking the level of attention consumers put on various product features when they make a purchase decision.

Among the existing online consumer interest tracking services, Google Trends (http://google.com/trends) is probably the best known and most widely used. It provides, free of charge, volume indexes for queries consumers have entered into the Google search engine since January 2004. Search volume indexes extracted from Google Trends are updated in real-time and aggregated on a weekly basis (or daily for the most popular queries), allowing users to track consumer interests with little time delay. Furthermore, search volume indexes from Google Trends are highly customizable. For example, search terms can be combined or excluded to formulate composite queries and can be filtered by geographic areas (e.g., countries, states, cities), time ranges (e.g., May 2004 through May 2008), and categories (e.g., beauty & fitness, autos & vehicles, computer & electronics). Last and perhaps most important, Google is by far the most dominant search engine globally. Given the ubiquity of consumer online searches and Google's dominance in this space, the volume of Google searches can plausibly be viewed as a reflection of the collective interests of Internet users.

In this study, we tap into Google Trends as a promising source of marketing intelligence for monitoring the evolution of consumer preferences. In particular, we extract search volume indexes for feature-related keywords (hereafter referred to as "feature search indexes"), which we argue can potentially serve as proxies for the relative attention consumers put on the corresponding product features.

As motivating examples, Figure 1 presents the time plots for four sets of U.S.-based feature search indexes over an eight-year span (2004 through 2011). Figure 1A plots the search indexes for four nutritional features that are commonly associated with food products: calories, carbs, cholesterol, and fiber. We see that consumer searches for calories increased steadily and substantially over the years (more than 60% higher at the end of 2011 than the beginning of 2004). By contrast, consumer searches for cholesterol declined by about 40% during the same period. As for carbs, search interest followed a U-shaped trend line, declining through mid-2005 and having had a strong comeback ever since. In contrast to the large movements in search interests for calories, cholesterol and carbs, the index for fiber remained largely stable. Figure 1B plots the search indexes for four features that are commonly associated with laptop computers. We see that as speed and memory have increased to a level that is more than adequate for most everyday computing needs, other features such as screen size and battery life have attracted substantially more attention over the years (over 200% increase for screen size and 150% increase for battery life). Similarly, Figure 1C shows a diverging pattern for two features that are commonly associated with digital cameras: weight and resolution. While the search index for resolution was more or less flat, the index for weight increased over 60%. Finally, Figure 1D plots the search indexes for five features that are commonly associated with automobiles, showing that (1) searches for keywords related to fuel efficiency went through a roller coaster ride; (2) searches for keywords related to cost-to-buy and cost-to-operate increased substantially; (3) searches for keywords related to acceleration declined substantially; and (4) searches for SUV bottomed in 2008 and had bounced back in more recent years.

The time plots presented in Figure 1 show that consumer online searches for keywords related to various product features can vary substantially over time, following very different trend lines. This raises an important question: Are trends in feature searches reflective of trends in the relative prominence of different features in consumer

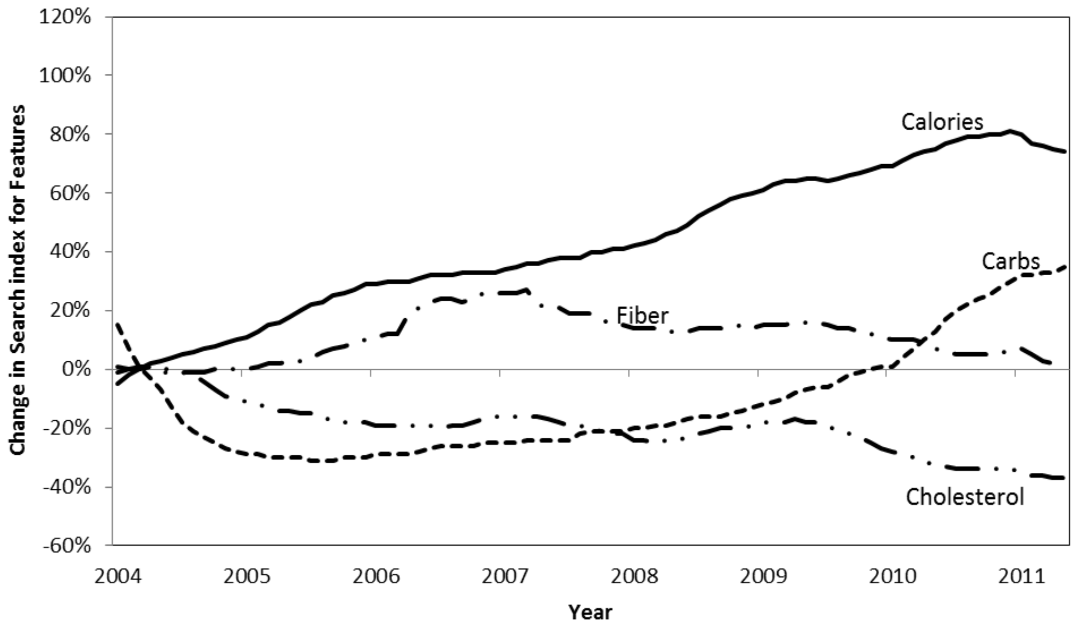


Figure 1 Panel A: Food products Google search volume restricted to the search queries in "Nutrition" category (six month moving averages of changes from initial value)

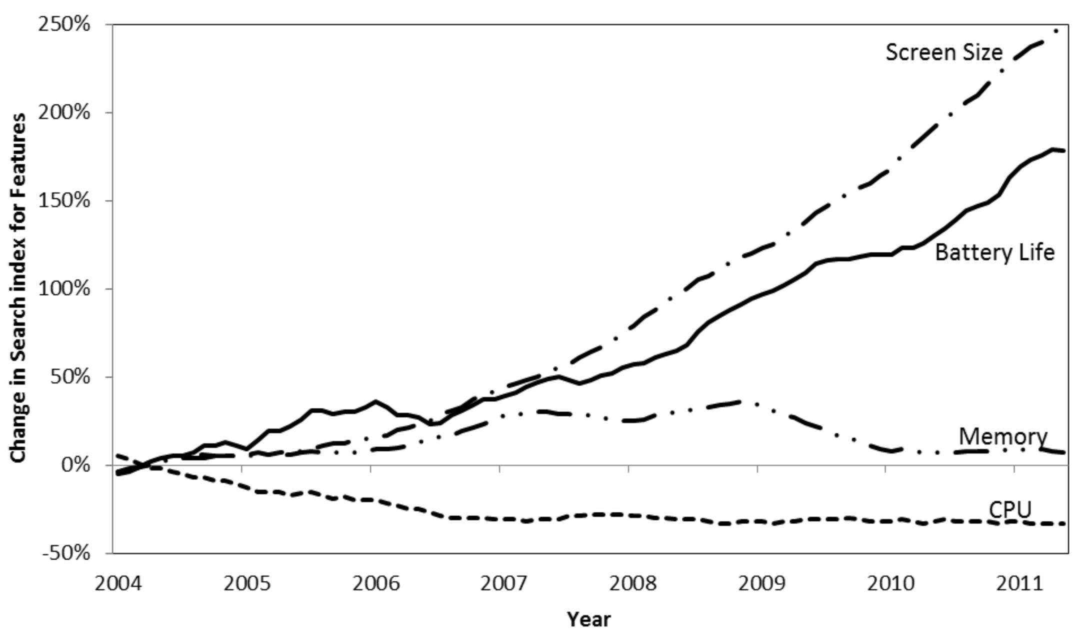


Figure 1 Panel B: Laptop Google search volume restricted to the search queries in "Nutrition" category (six month moving averages of changes from initial value)

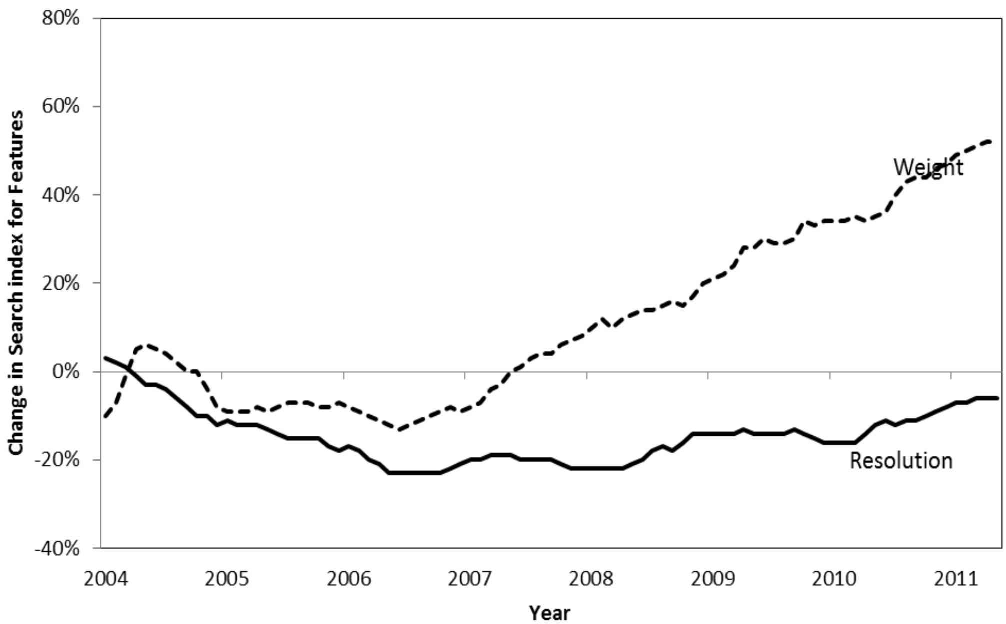


Figure 1 Panel C: Digital Camera Google search volume restricted to the search queries in "Nutrition" category (six month moving averages of changes from initial value)

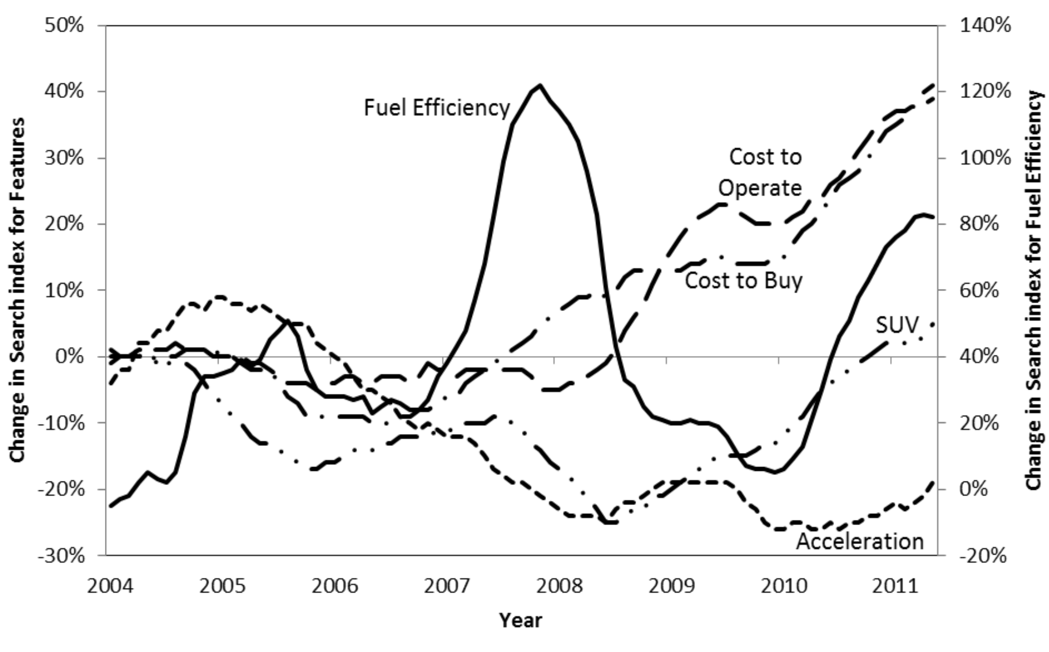


Figure 1 Panel D: Automobiles Google search volume restricted to the search queries in "Nutrition" category (six month moving averages of changes from initial value)

Figure 1 Internet Search for Features

leverage the former, which is available in real-time and for free, in monitoring the evolution of the latter, which can be otherwise hard to measure directly and costly to obtain on a regular basis. Indeed, a main motivation for our study is to address this question empirically. In particular, we set out to find out: first, to what extent evolution in feature searches can explain longitudinal variations in market shares, after controlling for marketing efforts; and second, whether trends in feature searches relate to product baseline attractiveness in any systematic fashion.

The remainder of the paper proceeds as follows. The next section provides a quick overview of two literatures, one on how existing market share analyses have modeled the dynamics of product baseline attractiveness, and the other on how Google Trends data have been used in as tracking measures of consumer interests. We propose a novel market share model where the baseline attractiveness of each competing product is allowed to vary over time as a function of consumer online feature searches. In the empirical illustration, we use market share and marketing efforts data for 70 major vehicles in the U.S. market between 2004 and 2012, augmented with Google Trends indexes for keywords that are commonly associated with fuel efficiency, acceleration, body type (SUV or sedan), cost to buy, and cost to operate. Our results indicate that, on average, 24% of market share fluctuation can be explained by trends in product feature searches, above and beyond what can be explained by marketing efforts (12%) and product brand name searches (14%). Furthermore, all else equal, we find that searches for positive features (i.e., fuel efficiency and acceleration) are positively associated with the market shares of products that have higher levels of those features. As for negative features (i.e., cost to buy and cost to operate), we find that searches for them are negatively associated the market shares of products that have higher levels of them. In concluding the paper, we discuss the managerial implications of our findings and directions for future research.

LITERATURE REVIEW

Products can often be viewed as bundles of features, and their utility modeled as a compensatory function of feature levels (Lancaster 1966). Such a view is canonical in areas of marketing research such as conjoint analysis (e.g., Bradlow, Hu, and Ho 2004; Ding 2007; Green and Rao 1971) and choice modeling (e.g., Fader and Hardie 1996). Similarly, in the context of market share analysis, the baseline attractiveness of competing products is often modeled as a function of their feature levels (Cooper and Nakanishi 1988). Because the feature levels of an existing product usually change little over time, in longitudinal analyses of market shares, product baseline attractiveness is typically treated as static, leaving market share dynamics to be explained mainly by time-varying marketing efforts (e.g., Ailawadi et al. 2001; Bucklin, Russell, and Srinivasan 1998; Bucklin, Siddarth, and Silva-Risso 2008; Gielens 2012; Khan and Jain 2005; Neslin 1990; Pollay et al. 1996).

Treating product baseline attractiveness as static is plausible when the window of observation is short and it is reasonable to assume that consumer tastes remain more or less constant. However, over time consumer needs and wants tend to evolve. When consumer tastes change substantially, the relative standing of different products in the eyes of the consumer is bound to change as well. To allow for such possibilities, a few previous studies have considered time-varying product baseline attractiveness. For example, Sriram, Balachander, and Kalwani (2007) use fixed quarterly and yearly effects to account for time-varying brand equity in a market share model of grocery products. Chintagunta (2001) and Nair, Dubé, and Chintagunta (2005) allow the baseline attractiveness of a product to evolve over time following a stochastic process (e.g., random walk). However, although these approaches can accommodate time-varying product baseline attractiveness, they are limiting in the sense that they cannot help managers understand why certain products become more or less attractive to consumers.

From the perspective that products can be viewed as bundles of features, one way for their relative attractiveness to evolve is through shifting prominence of different features in purchase decisions. All else equal, products should become relatively more (less) attractive if they offer superior (inferior) levels of features that have received increasingly more attention from consumers. In other words, product baseline attractiveness should vary systematically as the relative attention consumers put on different features shifts over time. The challenge in practice lies in finding a reliable and cost effective tracking measure of consumer attention when it comes to product features.

As noted earlier, rather than relying on longitudinal conjoint studies or tracking consumer surveys, which can be time-consuming and cost-prohibitive, the emergence of numerous online consumer interest monitoring services presents a potentially powerful alternative for tracking the relative prominence of each product feature in consumer decision-making. In this study we tap into Google Trends and treat feature search indexes as proxies for consumer interests in the corresponding features. It is obviously an empirical question as to whether and to what extent fluctuations in feature searches are systematically tied to fluctuations in product baseline attractiveness and therefore market share dynamics.

Intuitively, as consumers rely increasingly on Internet search engines such as Google in their acquisition of product information (Ratchford, Lee, and Talukdar 2003; Ratchford, Talukdar, and Lee 2007), it is conceivable that the search intensity for keywords related to a particular product feature should be positively correlated with the relative attention consumers put on that feature. In other words, when a feature gains in prominence in consumer decision-making, one should expect consumers to seek more information about it online. Indeed, our literature search has led us to a burgeoning area of research that leverages search indexes extracted from Google Trends as proxies for real-world interests.

For example, in epidemiology Ginsberg et al. (2009) and Pelat et al. (2009) show that the search volume for disease-related terms can be used as a real-time indicator of disease incidence rates, and is cheaper and faster than measures collected through conventional epidemic surveillance methods. In macroeconomics, it has been shown that search volume data can improve forecasts of housing market price and sale volume (Wu and Brynjolfsson 2009), unemployment rate (Choi and Varian 2009b; Askitas and Zimmermann 2009), and household expenditure (Vosen and Schmidt 2011). In finance, Da, Engelberg, and Gao (2011) show that search volumes for ticker names can be used to better predict stock prices.

More relevant to marketing, in the context of sales forecasting, Choi and Varian (2009a) demonstrate that search volume data can help predict current consumer demand in a diverse set of industries including retailing, automotive, housing, and tourism. Du and Kamakura (2012) show that seven common trends extracted from Google search data for 38 major vehicle brands can explain 74% of new car sales in the U.S. Both studies indicate that there can be strong ties between consumer online search interests and offline purchases.

One common aspect of the studies discussed above is that they have all focused on search terms that are directly tied to the subject of study. For example, in relating online searches to vehicle sales, both Choi and Varian (2009a) and Du and Kamakura (2012) focused on the linkage between search for a brand name (e.g., “Honda”) and the sales of that brand (hereafter, we refer to this type of online search as “brand search”). In this study, we extend beyond brand search by including feature search in an analysis of market share dynamics. Our extension is motivated by the fact that consumers often engage in both brand and feature searches when they gather product information online.

Furthermore, by modeling product market share dynamics as a function of feature searches, our approach offers two important advantages in practice. First, feature searches, unlike brand searches, are not tied to any specific product and in general have much larger volumes. For example, in the U.S. the search volume for keywords related vehicle fuel efficiency is more than five times larger than the search volume for Prius. Consequently, compared with brand searches, feature searches are less susceptible to idiosyncratic forces that affect only a few products and do not reflect shifting trends in consumer tastes (e.g., a product recall may trigger more non-purchase searches for Prius, while having little impact on searches for vehicle fuel efficiency.

Second and more important, the extension to include feature searches in market share analyses can potentially lead to more actionable insights. Relating brand searches to product performances is intuitive and useful in revealing trends in consumer interests at the brand or product level. However, it does not help in explaining why certain brands or products are receiving increasingly more or less attention from consumers. By contrast, our extension can reveal trends in consumer interests at the feature level. Knowing which features are attracting increasingly more or less consumer attention allows managers to take actions in response to the underlying shifts in consumer tastes. For example, when a feature gains prominence in consumer decision-making (as manifested in a substantial increase in consumer online searches for that feature), managers of products that are considered superior (inferior) in that feature can increase (decrease) emphasis on that feature in their communication to consumers. In other words, our extension to model product baseline attractiveness as a function of feature searches would allow managers to dynamically adjust the relative emphasis they put on each product feature in their marketing messages. They will be able to mitigate threats as well as leverage opportunities presented by shifting consumer tastes, which can now be monitored in real-time and cost-effectively through online consumer interest tracking services such as Google Trends. In the next section, we present our proposed modeling framework for tapping into this emerging source of marketing intelligence.

MODEL

We assume that in each period a consumer chooses which product to buy from all the available alternatives in the market. For consumer h ($h = 1, \ldots, H$), in each period t ($t = 1, \ldots, T$), the utility of choosing product i ($i = 0, .., n$, where 0 indicating the outside good), is given by:

$$(1) \qquad u_{hit} = v_{hit} + \varepsilon_{hit},$$

where ε_{hit} is an i.i.d. random error with a Type I extreme value distribution across consumers, and

(2) $$v_{hit} = \begin{cases} \alpha_{it} + \sum_j \beta_{ij}^M x_{ijt} + \xi_{it} & \text{for } i \neq 0 \\ 0 & \text{for } i = 0 \end{cases}$$

where α_{it} is the baseline attractiveness of the product, x_{ijt} the observed marketing mix j $(j = 1, \dots, J)$ for product i in period t, β_{ij}^M the impact of marketing mix j on the overall attractive product i, and ξ_{it} any unobserved random shock to the product attractiveness in period t, which is assumed to be i.i.d. normal.

Given Equations 1 and 2, integrating across consumers (i.e., h's) leads to the following multinomial logit choice share of product i in period t,

(3) $$P_{it} = \frac{\exp\left(\alpha_{it} + \sum_j \beta_{ij}^M x_{ijt} + \xi_{it}\right)}{1 + \sum_{i'=1}^{n} \exp\left(\alpha_{i't} + \sum_j \beta_{i'j}^M x_{i'jt} + \xi_{i't}\right)}.$$

Similarly, the choice share of the outside good in period t can be expressed as

(4) $$P_{0t} = \frac{1}{1 + \sum_{i'=1}^{n} \exp\left(\alpha_{i't} + \sum_j \beta_{i'j}^M x_{i'jt} + \xi_{i't}\right)}.$$

Thus for all $i > 0$, we have

(5) $$\ln\left(\frac{y_{it}}{y_{0t}}\right) = \ln\left(\frac{P_{it}}{P_{0t}}\right) = \alpha_{it} + \sum_j \beta_{ij}^M x_{ijt} + \xi_{it},$$

where y_{it} is the sales of product i and y_{0t} the sales of the outside good (i.e., any products that are not included in the analysis but sold in the same market during the period t).

As discussed earlier, market share analyses in the existing literature typically treat the baseline attractiveness α_{it} as a nuance parameter, either assumed to be time-invariant or controlled for through period-specific dummy effects or as a pure stochastic process.

The focus of the existing literature has been on relating market share dynamics to marketing mixes (i.e., the effects of x_{ijt}'s). In this study, we focus on modeling the dynamics of baseline attractiveness, after controlling for the effects of marketing mixes. In particular, we allow α_{it} to be a function of brand and feature searches:

$$\alpha_{it} = \beta_i^0 + \beta_i^S s_{it} + \sum_k \beta_{ik}^F z_{kt}, \tag{6}$$

where s_{it} is brand search for product i during period t; z_{kt} ($k = 1, \dots, K$) is feature search for feature k during period t; β_i^0 is the time-invariant component of product i's attractiveness; β_i^S captures to what extent search for the brand name of product i is tied to its attractiveness; β_{ik}^F ($k = 1, \dots, K$) captures the relationship between consumer search for feature k and the attractiveness of product i. To allow information pooling across products in parameter estimation, each product-specific coefficient (i.e., β_i^0, β_i^S, β_{ik}^F and β_{ij}^M) is assumed to be randomly drawn from a common normal distribution. In sum, our proposed market share model can be written as:

$$\ln\left(\frac{y_{it}}{y_{0t}}\right) = \beta_i^0 + \beta_i^S s_{it} + \sum_k \beta_{ik}^F z_{kt} + \sum_j \beta_{ij}^M x_{ijt} + \xi_{it},\ \ \xi_{it} \sim \mathrm{i.i.d.\ N}(0, \sigma_i^2), \tag{7}$$

$$\beta_i^0 \sim \mathrm{i.i.d.\ N}\left(\overline{\beta^0}, \sigma_{\beta^0}^2\right)$$

$$\beta_i^S \sim \mathrm{i.i.d.\ N}\left(\overline{\beta^S}, \sigma_{\beta^S}^2\right),$$

$$\beta_{ik}^F \sim \mathrm{i.i.d.\ N}\left(\overline{\beta_k^F}, \sigma_{\beta_k^F}^2\right), \text{for } k = 1, \dots, K, \text{and}$$

$$\beta_{ij}^M \sim \mathrm{i.i.d.\ N}\left(\overline{\beta_j^M}, \sigma_{\beta_j^M}^2\right), \text{for } j = 1, \dots, J,$$

where $\overline{\beta^0}$, $\overline{\beta^S}$, $\overline{\beta_k^F}$, and $\overline{\beta_j^M}$ are the mean and $\sigma_{\beta^0}^2$, $\sigma_{\beta^S}^2$, $\sigma_{\beta_k^F}^2$, and $\sigma_{\beta_j^M}^2$ are the variance of the corresponding normal distributions.

In Equation (7), a statistically significant β_{ik}^{F} would indicate that changes in consumer search intensity for feature k are tied to changes in the attractiveness of product i. If indeed changes in feature search intensity are caused by changes in the relative attention consumers put on different product features, one would expect the sign and size of β_{ik}^{F} to vary systematically as a function of the level of feature k offered by product i (w_{ik}). Intuitively, for example, when fuel efficiency becomes a more prominent factor in vehicle purchase decisions, it can manifest in two changes in consumer behavior: first, consumers search more for vehicle fuel efficiency related keywords, and second, more fuel efficient vehicles such as Toyota Prius will become more attractive and gas guzzlers like Toyota Sequoia will become less so. Consequently, one would expect β_{ik}^{F} associated with fuel efficiency to be positive for Toyota Prius and negative for Toyota Sequoia. To formalize the above intuition, we hypothesize that:

H_a: The volume of consumer online searches for keywords related to a positive feature is positively tied to the baseline attractiveness and thus market shares of products offering higher levels of that feature. A positive feature is one which is generally considered the more the better by most consumers (e.g., fuel efficiency, acceleration, battery life, and resolution).

Following the same logic, when operating costs becomes more of a concern in vehicle purchase decisions, it can lead to two changes: first, consumers searching more for keywords related to vehicle operating costs, and second vehicles known for low (high) maintenance costs would gain (lose) attractiveness. More formally, we hypothesize that:

H_b: The volume of consumer online searches for keywords related to a negative feature is negatively tied to the baseline attractiveness and thus market shares of products offering higher levels of that feature. A negative feature is one which is generally considered the less the better by most consumers (e.g., costs to own, costs to operate, cholesterol).

To capture the above logic and empirically test the resulting hypotheses, we model product i's coefficient for feature k, β_{ik}^{F}, as a function of product i's level of feature k, w_{ik}:

(8) $$\beta_{ik}^{F} = \overline{\beta_{ik}^{F}}, + \tau_{ik} = \gamma_{k0} + \gamma_{k1} w_{ik} + \tau_{ik},\ \tau_{ik} \sim \text{i.i.d.}\, N\left(0, \sigma_{\beta_k^F}^2\right), \text{for } k = 1, \ldots, K$$

Putting everything together, we have a two-level hierarchical market share model:

(Level 1) $$\ln\left(\frac{y_{it}}{y_{0t}}\right) = \beta_i^0 + \beta_i^S s_{it} + \sum_k \beta_{ik}^F z_{kt} + \sum_j \beta_{ij}^M x_{ijt} + \xi_{it}$$

(Level 1) $$\beta_{ik}^F = \gamma_{k0} + \gamma_{k1} w_{ik} + \tau_{ik}$$

In the above model formulation, statistically significant β_{ik}^F helps managers identify features influence product i's market share dynamics, and the size and sign of β_{ik}^F helps quantify the impact of search for feature k on the market share of product i. Moreover, γ_{k1} can reveal how this impact is systematically tied to the actual feature levels. We have hypothesized that γ_{k1} should be positive for positive features and negative for negative features. To the extent this turns out to be the case empirically, it should further validate our modeling approach and the treatment of feature search indexes as a tracking measure of underlying consumer interest in the corresponding feature.

Benchmark Models

In order to evaluate the incremental value of feature search in explaining market share dynamics, we benchmark our proposed model against two alternatives. The first benchmark does not involve any search data, assuming constant baseline attractiveness (β_i^0) and thus letting market efforts (x_{ijt},'s) to explain all the observed longitudinal variation in market shares:

(9) $$\ln\left(\frac{y_{it}}{y_{0t}}\right) = \beta_i^0 + \sum_j \beta_{ij}^M x_{ijt} + \xi_{it},\ \xi_{it} \sim \mathrm{i.i.d.}\,N(0, \sigma^2),$$

(10) $$\beta_i^0 \sim \mathrm{i.i.d.}\,N\left(\overline{\beta^0}, \sigma_{\beta^0}^2\right), \beta_{ij}^M \sim \mathrm{i.i.d.}\,N\left(\overline{\beta_j^M}, \sigma_{\beta_j^M}^2\right), \text{for } j = 1, \dots, J.$$

Compared with the first benchmark, the second benchmark relaxes the constant baseline attractive assumption, allowing it to vary as a function of the search volume for the brand name of product i (s_{it}). This addition of product brand name search as a covariate in explaining market share dynamics is consistent with the notion that how much consumers search for a product brand name online should be highly informative of the product's sales, which has been the most common way of leveraging consumer online search data in the emerging literature in this area (e.g.,Varian and Choi 2009; Du and Kamakura 2012). More formally, the second benchmark model is given by:

(11) $$\ln\left(\frac{y_{it}}{y_{0t}}\right) = \beta_i^0 + \beta_i^S s_{it} + \sum_j \beta_{ij}^M x_{ijt} + \xi_{it},\ \xi_{it} \sim \mathrm{i.i.d.}\,N(0, \sigma^2),$$

$$\beta_i^0 \sim \mathrm{i.i.d.}\,N\left(\overline{\beta^0}, \sigma_{\beta^0}^2\right),$$

$$\beta_i^S \sim \mathrm{i.i.d.}\,N\left(\overline{\beta^S}, \sigma_{\beta^S}^2\right), \text{and}$$

$$\beta_{ij}^M \sim \mathrm{i.i.d.}\,N\left(\overline{\beta_j^M}, \sigma_{\beta_j^M}^2\right), \text{for } j = 1, \dots, J.$$

In sum, by comparing our proposed model against the first benchmark we can determine how much incremental value search data (product brand name search and feature search) adds to the marketing mix variables in explaining longitudinal market share variations. By comparing our proposed model against the second benchmark we can determine the incremental value feature search adds on top of marketing mix variables and product brand name search.

DATA

We apply our proposed model to the automotive market in the U.S. In buying a large-ticket durable product such as a new vehicle, consumers are highly motivated to conduct product-related information search (Alba and Hutchinson 1987), and increasingly those searches are carried out over the Internet (Ratchford, Lee, and Talukdar 2003; Ratchford, Talukdar, and Lee 2007). Furthermore, the automotive market in the U.S. is a highly differentiated market, with a large number of competing products that can be characterized by a set of commonly observed and well defined features. Consequently, we argue this market provides a suitable context for testing our proposed modeling approach.

For our empirical illustration, we gathered monthly automobile sales from Automotive News and feature search index from Google Trends, both for the U.S. market and between January 2004 and May 2011. We focused on the 70 top-selling non-luxury automobiles that were continuously available in the U.S. market throughout the study period. These 70 vehicles accounted for approximately 60% of total industry sales. We treat all the other non-luxury automobiles as the outside good, the sales of which were also obtained from Automotive News. For each of the 70 focal vehicles, we also acquired monthly cash backs offered to consumers and monthly advertising spend as indicators of marketing support.

In selecting vehicle features to be included in our analysis, we sought the ones that are considered most relevant – according to major automotive information websites such as Edmonds.com and JDPower.com – in an average consumer's car purchasing decision. In particular, we focus on the following five features: fuel efficiency, acceleration, body configuration (passenger car vs. SUV), cost to buy, and cost to operate. The first two are considered positive features, and the last two negative features.

A key issue in constructing a search volume index for a product feature lies in that, when consumers seek information related to the same underlying feature they can use a wide variety of terms (e.g., fuel efficiency vs. fuel economy vs. gas mileage vs. miles per gallon, etc.), let alone abbreviations (e.g., mpg), minor variations, singulars/plurals, and misspellings. Fortunately, Google Trends allows one to construct composite queries by joining multiple terms with plus signs. For example, the composite query we used to extract the search index for full efficiency includes 10 different terms. Table 1 presents the actual terms used in constructing each of the five composite queries we used to extract feature search indexes from Google Trends.

To come up with the list of terms used to form the composite queries, we went through a careful multi-step procedure. First, we attempt to generate a comprehensive list of candidate terms so that we will not miss any popular terms used by consumers. We start by scanning consumer reviews on Edmonds.com and selecting terms that appear to be relevant for each feature. Each term resulted from this step is then entered into Google Adwords for suggestions of additional related keywords, which further expands the list of candidate terms. Subsequently, we focus on trimming down this list by excluding terms that can be intended for things other than the focal feature (e.g., we do not use the term “acceleration” alone since the result can be contaminated by the searches related to accelerator pedal). We used two independent judges and whenever disagreement arose, a third was used to make the decision as to whether to keep or remove a term. Finally, we remove terms that have much lower search volumes than the popular ones (both Google Adwords and Google Trends can be used to determine the relative search volume for different terms). Exactly the same procedure was followed in constructing composite queries for vehicle brand names, which consist of mainly vehicle make and model, along with their popular variations (e.g., volkswagen beetle+vw beetle+volkswagon beetle+volkswagen beatle+vw beatle+volkswagon beatle, hyundai elantra+hunday elantra+hyundai elentra+hunday elentra, chevrolet aveo+cheverolet aveo+chevorlet aveo+chevy aveo+chevrolet aveo5+cheverolet aveo5+chevorlet aveo5+chevy aveo5).

Table 1 Google Trends Queries Adopted for Feature Search Indexes

K	Feature	Query
1	Fuel Efficiency	city mileage+fuel consumption+fuel economy+fuel efficiency+fuel efficient+gas mileage+highway mileage+hybrid+mile per gallon+mpg
2	Acceleration	acceleration time+"acceleration times"+"0-60 time"+"0 to 60 time"+"0-60 times"+"0 to 60 times"+"quarter mile time"+"quarter mile times"
3	SUV	SUV
4	Cost to Buy	Price+MSRP+discount+rebate+"finance rate"+"cash back"+cashback
5	Cost to Operate	powertrain-warranty+reliability+reliable+"cost to own"+"maintenance cost"+"maintenance costs"

Notes: terms inside quotation mark match the exact phrase, terms separated by space will match queries having all of the terms, terms separated by + will match queries having any of the terms.

To make sure the feature search indexes are indeed related to automobiles (as opposed to other product categories), we rely on the category filtering function provided by Google Trends by setting it to “Autos & Vehicles”. Last but not the least, we divide each raw feature search index by the search index for the whole “Autos & Vehicles” category. This gives us a normalized measure that captures the relative search intensity for a particular feature as compared with total searches in the focal product category. In other words, the normalized measure indicates feature search *share* as opposed to level. We argue feature search share is a more reliable indicator of how the relative prominence of each feature has evolved over time, because the normalized measure has removed variations that are simply due to changes in consumers’ overall category search interest (e.g., seasonal fluctuations).

RESULTS

Model Performance

Table 2 reports an overall and vehicle-by-vehicle comparison of model performance in explaining the observed market share dynamics. On average, our proposed model explains 50% ($\mathrm{R}^2 = .50$) of variance in the data, while the two benchmark models explain 12% and 26%, respectively. Taking into account the different numbers of parameters in each model, our proposed model still outperforms with an adjusted R^2 of .46, as compared to .10 and .24 for the two benchmark models. The relatively poor performance of the first benchmark model highlights the fact that marketing mix variables can only explain a small fraction of observed market share dynamics. As we have argued earlier in the paper, over time consumer tastes are bound to evolve and when substantial changes take place in, for example, the relative prominence of different product features in consumer purchase decisions, the baseline attractiveness of existing products will shift and so will their market shares.

The significant improvement from the first to the second benchmark model is consistent with what has been found by Choi and Varian (2009a) and Du and Kamakura (2012); that is, consumer online searches for product brand names are closely tied to market demands and can explain a significant portion of variance in product sales. However, as discussed earlier, trends in online searches for product brand names can only indicate which brands are attracting more or less attention from consumers; they cannot reveal more fundamental changes in consumer tastes, such as the amount of attention she ups on of different product features, which is a void that our proposed model attempts to fill.

Compared with the second benchmark model, the improvement in explanatory power of our proposed model is remarkable: adding five non-product specific feature search indexes on top of product brand name searches almost doubles the overall variance explained (from 26% to 50%). Across the 70 vehicles included in our analysis, the goodness-of-fit improvements range from 9% (for Ford Crown Victoria) to 86% (for Honda Element), with a median of 48%. Taken together, our empirical results suggest that trends in consumer search intensity for terms that are related to product features can explain a tremendous amount of observed market share dynamics, above and beyond what can be accounted for by searches for product brand names and marketing mixes.

Table 2 Model Comparison

	Model Performance			Comparison	
Goodness of Fit	Model 1	Model 2 (Brand Search)	Model 3: (Brand & Feature Search)	Model 3 vs. Model 1	Model 3 vs. Model 2
-2 LL	5,171	3,990	1,971	3,201	2,019
AIC	5,179	4,000	1,991	3,189	2,009
AICC	5,179	4,000	1,991	3,189	2,009
BIC	5,188	4,011	2,013	3,175	1,998
Average Adjusted R^2	.10	.24	.46	.36	.22
Average R^2	.12	.26	.50	.38	.24
Vehicle	**R-squared**			**Improvement in R^2**	
Chevrolet Aveo	.02	.06	.11	.09	.05
Chevrolet Corvette	.26	.43	.69	.43	.26
Chevrolet Impala	.03	.06	.11	.08	.05
Chevrolet Malibu	.15	.17	.47	.32	.30
Chevrolet Avalanche	.04	.16	.63	.59	.47
Chevrolet Colorado	.09	.18	.65	.57	.47
Chevrolet Silverado	.11	.15	.41	.30	.26
Chevrolet Suburban	.01	.14	.36	.35	.22
Chevrolet Tahoe	.08	.08	.42	.34	.34
Chrysler 300	.19	.19	.60	.41	.41
Dodge Caravan	.01	.00	.42	.41	.41
Ford Crown Victoria	.04	.05	.09	.04	.04
Ford Focus	.05	.33	.42	.37	.09
Ford Mustang	.10	.26	.61	.51	.35
Ford Escape	.01	.38	.77	.75	.38
Ford Expedition	.35	.34	.72	.37	.37
Ford Explorer	.18	.18	.75	.57	.57
Ford F series	.12	.14	.41	.29	.27
Ford Ranger	.15	.28	.56	.41	.28
GMC Sierra	.03	.04	.24	.20	.19
GMC Yukon	.06	.07	.43	.37	.36
Honda Accord	.01	.28	.48	.47	.19
Honda Civic	.00	.18	.58	.58	.41
Honda CR-V	.00	.60	.78	.78	.19
Honda Element	.08	.76	.86	.79	.10
Honda Odyssey	.02	.05	.28	.26	.23
Honda Pilot	.00	.01	.19	.19	.18
Hyundai Accent	.36	.68	.68	.32	.00
Hyundai Elantra	.08	.43	.47	.39	.03
Hyundai Sonata	.03	.32	.49	.45	.17
Hyundai Santa Fe	.01	.05	.38	.37	.33

Table 2 Continued.

Vehicle		**R-squared**		**Improvement in R^2**	
Jeep Grand Cherokee	.22	.26	.51	.29	.25
Jeep Liberty	.23	.25	.74	.51	.49
Jeep Wrangler	.02	.37	.52	.50	.16
Kia Optima	.10	.11	.21	.11	.10
Kia Rio	.00	.06	.11	.10	.04
Kia Sedona	.18	.20	.42	.24	.22
Kia Sorento	.33	.63	.71	.38	.08
Mazda3	.25	.49	.63	.39	.15
Mazda6	.14	.17	.32	.18	.15
Mini Cooper	.01	.11	.66	.65	.55
Nissan Altima	.23	.48	.51	.28	.03
Nissan Maxima	.06	.09	.48	.42	.39
Nissan Sentra	.05	.41	.46	.40	.05
Nissan Armada	.34	.48	.65	.31	.18
Nissan Frontier	.18	.18	.41	.23	.23
Nissan Murano	.04	.18	.43	.39	.25
Nissan Pathfinder	.37	.45	.61	.24	.16
Nissan Titan	.48	.53	.73	.25	.21
Nissan Xterra	.15	.22	.70	.55	.49
Ram	.02	.29	.53	.52	.24
Scion xB	.04	.22	.44	.40	.22
Subaru Forester	.11	.76	.85	.74	.09
Subaru Impreza	.16	.16	.78	.63	.62
Subaru Legacy	.15	.15	.34	.19	.19
Toyota Avalon	.09	.20	.44	.35	.24
Toyota Camry	.10	.13	.39	.29	.25
Toyota Corolla	.18	.30	.44	.26	.13
Toyota Prius	.27	.29	.67	.40	.39
Toyota 4Runner	.17	.36	.55	.38	.19
Toyota Highlander	.02	.03	.18	.16	.15
Toyota RAV4	.29	.71	.86	.57	.15
Toyota Sequoia	.10	.38	.63	.53	.25
Toyota Sienna	.01	.16	.21	.20	.04
Toyota Tacoma	.07	.07	.34	.27	.27
Toyota Tundra	.29	.32	.34	.05	.02
VW Jetta	.00	.32	.82	.82	.49
VW New Beetle	.00	.52	.67	.67	.15
VW Passat	.19	.20	.27	.08	.07
VW Golf	.01	.01	.45	.44	.44

Parameter Estimates

Besides remarkable improvement in explanatory power, the results from our approach also add to actionable managerial insights. If changes in feature search intensity are indeed manifestations of evolving consumer tastes and the feature search indexes can serve as tracking measures of the relative attention consumers put on different features, we should expect a systematical relationship between feature search indexes and product market shares. In other words, if our model's superior performance in fitting the data is not due to "luck" but rather because trends in feature search indexes contain genuine information about shifting consumer preferences, then we should expect to see a pattern in our estimates for β_{ik}^{F} and γ_{k1} that is consistent with H_a and H_b.

Table 3 reports our model parameter estimates for the intercept (β_i^0), impact of feature search (β_{ik}^{F}), impact of product brand name search (β_i^S), and impact of marketing mixes (β_{ij}^{M}). Estimates in ***bold italic*** are significant at $p = .05$, and those in **bold** are significant at $p = .10$. For product brand name search (β_i^S), 25 (30) estimates are positive at .05 (.10) level, and 3 (3) are negative at .05 (.10) level. For incentives (β_{i1}^{M}), 10 (10) estimates are positive at .05 (.10) level, and 3 (5) are negative at .05 (.10) level. For advertising (β_{i2}^{M}), 8 (11) estimates are positive at .05 (.10) level, and 0 (1) are negative at .05 (.10) level. The fact that most of the significant parameter estimates are of the expected sign (i.e., positive) suggests a basic level of face validity of our results.

The estimates for β_{ik}^{F}'s are of the most interest to us as they capture the impacts of feature searches on product attractiveness. Take Toyota Prius as an example. Its β^F coefficient for fuel efficiency search is significant (p = .05) and has a value of 1.09, the largest among all vehicles. This indicates that, as one might have expected, the more

Table 3 Parameter Estimates

	R^2	Intercept	Feature Search					Product Search	Marketing Activities	
			Fuel Efficiency	Acceleration	SUV	Cost to Buy	Cost to Operate		Incentives	Advertising
Vehicle		β_i^0	β_{i1}^F	β_{i2}^F	β_{i3}^F	β_{i4}^F	β_{i5}^F	β_i^S	β_{i1}^M	β_{i2}^M
Chevrolet Aveo	.11	***-4.23***	**.34**	-.04	-.13	.05	-.52	.30	.05	-.13
Chevrolet Corvette	.69	***-2.96***	.22	.27	-.48	***-1.43***	***-1.41***	.41	.03	.19
Chevrolet Impala	.11	***-2.73***	-.11	-.16	-.39	.03	-.26	.66	.00	-.18
Chevrolet Malibu	.47	***-3.77***	.06	-.14	***-1.59***	***1.20***	***.84***	-.06	***.07***	.29
Chevrolet Avalanche	.63	***-3.50***	***-.79***	.02	***2.13***	-.51	***-2.50***	.26	-.02	.84
Chevrolet Colorado	.65	***-3.96***	.12	***1.10***	***1.05***	***-.97***	***-1.72***	***.70***	***.06***	.43
Chevrolet Silverado	.41	***-1.95***	***-.54***	.18	.41	.61	***-1.08***	.09	-.01	.47
Chevrolet Suburban	.36	***-3.32***	***-.75***	.13	.50	-.15	***-1.24***	.54	***-.04***	.30
Chevrolet Tahoe	.42	***-2.83***	***-.64***	.24	**.62**	-.68	**-.69**	.05	-.01	.88
Chrysler 300	.60	***-1.45***	***.35***	***.62***	***-2.05***	***-3.46***	-.15	***2.82***	.00	.00
Dodge Caravan	.42	***-3.73***	-.28	***.99***	***.92***	-.21	***-.71***	.04	-.01	.16
Ford Crown Victoria	.09	***-4.65***	.07	.27	.21	.07	-.33	.22	.01	.36
Ford Focus	.42	***-3.89***	.06	.10	***-.78***	.78	.00	***.77***	.04	.27
Ford Mustang	.61	***-2.90***	***.48***	.43	***-.97***	***-1.47***	.25	***.75***	-.01	***1.13***
Ford Escape	.77	***-4.52***	**-.41**	-.23	-.57	***.94***	***.77***	***.95***	.04	.57
Ford Expedition	.72	***-3.82***	***-.95***	***.57***	***2.55***	-.19	***-2.00***	-.38	.00	***1.49***
Ford Explorer	.75	***-4.51***	***-.66***	***1.34***	***2.87***	-.69	***-1.63***	.37	.02	***1.48***
Ford F series	.41	***-2.23***	***-.63***	.39	.50	.63	-.52	-.03	-.01	.22
Ford Ranger	.56	***-5.41***	-.20	***.60***	.50	.33	***-1.26***	***1.71***	***.09***	.53
GMC Sierra	.24	***-2.79***	***-.50***	.00	.36	.53	***-1.12***	.19	-.02	.50
GMC Yukon	.43	***-3.34***	***-.90***	.13	***.99***	-.44	***-1.38***	-.01	-.01	.86
Honda Accord	.48	***-2.89***	.12	-.25	**-.70**	.42	.02	***.90***	.02	-.25
Honda Civic	.58	***-2.75***	**.42**	-.33	***-.82***	.70	.10	.19	.02	.40
Honda CR-V	.78	***-3.91***	-.22	***-.77***	-.55	***1.13***	.31	***1.03***	.02	.21
Honda Element	.86	***-5.01***	-.28	.24	***.75***	***-.77***	***-1.07***	***1.76***	.02	.20
Honda Odyssey	.28	***-2.84***	.06	.21	**-.62**	-.26	-.36	.31	.02	.65
Honda Pilot	.19	***-3.65***	***-.50***	.03	-.21	-.42	.40	.54	.02	.83
Hyundai Accent	.68	***-5.47***	.07	-.45	***-1.13***	.54	.63	***1.57***	***.18***	**-.86**
Hyundai Elantra	.47	***-4.87***	.15	.06	-.36	.05	***1.05***	***.90***	.06	-.26
Hyundai Sonata	.49	***-3.91***	.25	-.43	***-1.51***	***.89***	***1.08***	.39	-.06	.29
Hyundai Santa Fe	.38	***-4.72***	***-.60***	***-1.04***	.06	***1.44***	***.72***	***-.47***	.01	***.77***
Jeep Grand Cherokee	.51	***-3.84***	***-.46***	.46	***1.81***	-.38	***-1.53***	.56	***-.03***	***.96***
Jeep Liberty	.74	***-3.76***	***-.63***	***.81***	***1.47***	-.58	***-1.39***	.50	***-.06***	-.22
Jeep Wrangler	.52	***-3.44***	-.20	***-.92***	***-.67***	.28	-.57	***1.32***	.04	.27

Table 3 Continued.

Vehicle	R^2	Intercept	Feature Search					Product Search	Marketing Activities	
			Fuel Efficiency	Acceleration	SUV	Cost to Buy	Cost to Operate		Incentives	Advertising
		β_i^0	β_{i1}^F	β_{i2}^F	β_{i3}^F	β_{i4}^F	β_{i5}^F	β_i^S	β_{i1}^M	β_{i2}^M
Mazda3	.63	***-3.80***	.27	**-.51**	***-1.29***	.27	***.76***	**.48**	.03	.18
Mazda6	.32	***-4.03***	-.17	.17	.36	***-1.12***	-.06	.40	.00	**1.02**
Mini Cooper	.66	***-4.52***	.25	-.44	***-1.86***	***1.81***	.25	-.20	.02	.62
Nissan Altima	.51	***-3.16***	.08	-.39	-.17	-.04	.33	***.56***	.02	.40
Nissan Maxima	.48	***-4.63***	***-.76***	.34	***-1.12***	.12	***.72***	**.78**	.01	.12
Nissan Sentra	.46	***-4.84***	.23	-.12	-.17	-.21	**.67**	***.90***	.04	.09
Nissan Armada	.65	***-4.61***	***-.41***	.03	***.88***	***-2.03***	-.39	***1.40***	**-.05**	.00
Nissan Frontier	.41	***-4.18***	-.06	.24	***.75***	***-1.48***	-.03	.51	-.05	.58
Nissan Murano	.43	***-3.38***	**-.33**	-.10	-.51	***-1.06***	-.01	***1.05***	.03	.23
Nissan Pathfinder	.61	***-3.46***	.28	.09	-.47	***-2.73***	.49	***1.35***	***-.10***	**1.09**
Nissan Titan	.73	***-3.32***	***-.47***	***.82***	***.77***	***-2.24***	***-.80***	.58	-.03	.12
Nissan Xterra	.70	***-3.24***	***-.55***	.39	***.88***	***-2.39***	***-1.18***	***.85***	.00	.70
Ram	.53	***-2.44***	***-.54***	.40	.59	.40	***-1.01***	-.22	.00	.25
Scion xB	.44	***-3.67***	***.42***	.38	***-.96***	**-.84**	-.52	**.54**	.02	.73
Subaru Forester	.85	***-4.75***	***-.53***	.09	***-1.36***	.45	-.07	***1.99***	.01	.69
Subaru Impreza	.78	***-4.13***	-.07	***-1.08***	***-1.47***	***1.38***	***.79***	-.60	***.06***	.89
Subaru Legacy	.34	***-4.15***	.09	***.69***	***-1.12***	-.03	-.55	***1.14***	***.08***	.23
Toyota Avalon	.44	***-2.93***	.28	.37	***-1.56***	***-1.27***	-.36	***1.06***	-.04	.36
Toyota Camry	.39	***-2.28***	.11	-.06	***-.78***	.63	-.14	.18	.03	.05
Toyota Corolla	.44	***-3.27***	.18	-.22	-.53	***1.02***	.01	.46	.04	-.40
Toyota Prius	.67	***-2.91***	***1.09***	***-1.31***	***-2.36***	-.37	***2.06***	-.14	***.12***	.50
Toyota 4Runner	.55	***-3.43***	***-.64***	.08	***2.99***	***-1.66***	***-1.68***	-.11	.02	***1.29***
Toyota Highlander	.18	***-3.97***	**-.34**	-.10	.44	.35	-.31	.29	.02	.02
Toyota RAV4	.86	***-5.27***	***-.66***	***-1.16***	***-1.78***	***1.56***	***1.85***	***2.23***	.04	-.42
Toyota Sequoia	.63	***-3.93***	***-.99***	-.09	***1.67***	.48	***-1.75***	***-1.42***	***.07***	***1.54***
Toyota Sienna	.21	***-3.16***	-.27	.08	.05	.38	-.34	-.43	.00	.41
Toyota Tacoma	.34	***-2.88***	.02	-.13	-.54	-.24	-.23	.39	.02	.14
Toyota Tundra	.34	***-3.52***	-.13	-.13	.06	-.10	-.53	***.61***	***.06***	.40
VW Jetta	.82	***-4.35***	-.18	**-.80**	***-1.42***	**.94**	***1.16***	.74	.02	.01
VW New Beetle	.67	***-2.24***	.26	***.92***	-.29	-.07	***-2.12***	***-7.39***	.02	.67
VW Passat	.27	***-4.45***	-.27	.10	-.01	***-1.30***	.20	***.84***	.02	***1.51***
VW Golf	.45	***-4.13***	***.49***	***-2.24***	-.18	-.10	.11	***1.00***	.02	.87

consumers search for fuel efficiency related terms, the more attractive Prius becomes, and the larger shares it grabs in the market. Furthermore, Prius has a non-significant coefficient for cost to buy, significant and negative coefficient for acceleration and SUV body type, and significant and positive coefficient for cost to operate. These are all consistent with the general perception or fact that Prius is not particular powerful, is not an SUV, and has relatively high reliability and lower long term fuel expense.

Instead of going through the estimates of β_{ik}^{F}'s for the rest of the 70 vehicles one by one, we examine the estimates of γ_{k1} reported in Table 4, which captures how β_{ik}^{F} (the impact of search for feature k on product i's attractiveness) varies as a function of w_{ik} (the level of feature k of product i, log-transformed[1]). Consistent with hypothesis H_a, γ_{11} and γ_{21} are positive and significant for fuel efficiency and acceleration, both positive features (the higher the more attractive). This implies that the impacts of searches for these features on product attractiveness are positively moderated by the actual product feature levels. Or more simply put, all else equal, for vehicles that offer higher levels of fuel efficiency and faster acceleration, their attractiveness and therefore market shares would have more to gain when consumer searches for these features intensify.

Similarly, consistent with hypothesis H_b, γ_{41} and γ_{51} are negative and significant for cost to buy and cost to operate, both negative features (the higher the less attractive). Intuitively, this indicates that for vehicles that are more costly to buy and operate, their attractiveness and therefore market shares would have more to lose when consumer searches for these features intensify. Finally, the coefficient for SUV γ_{31} is positive and

[1] We measure the levels of each feature as follows. We use vehicle i's manufacturer-specified miles per gallon as the measure of its fuel efficiency level, w_{i1}. Since vehicle acceleration is seldom included in the formal vehicle specifications—acceleration is to a great extent a function of environment and driving conditions—we use engine power (in horsepower) as measure for w_{i2}.

significant. Since SUV is dummy coded with 1 representing SUVs and 0 non-SUVs, a positive and significant γ_{31} indicates, not surprisingly, when the share of consumer searches for SUVs increases relative to non-SUVs, SUVs are expected to gain in attractiveness and market share over non-SUVs.

Table 4 Hypothesis Tests

Type of Feature	Feature	Expected Sign	Estimate	Result of Hypothesis Test
Positive Feature	Fuel Efficiency ($k = 1$)	+	1.34	supported ($p = .05$)
	Acceleration ($k = 2$)	+	0.63	supported ($p = .05$)
Categorical	SUV ($k = 3$)	+	1.39	supported ($p = .05$)
Negative Feature	Cost to Buy ($k = 4$)	–	–.81	supported ($p = .1$)
	Cost to Operate ($k = 5$)	–	–2.92	supported ($p = .05$)

In short, the fact that the estimates for the γ_{k1}'s have all turned out to be significant and of the expected signs suggests that the substantial improvement in explanatory power of our proposed model is by no means a statistical fluke. It also lends empirical support for the argument that trends in feature search indexes do contain genuine information about shifting consumer preferences and one can track these indexes as a way to monitor the relative importance of various product features in consumer purchase decisions.

Scenario Analyses

In this section, we demonstrate how our proposed model and its parameter estimates can be used in various scenario analyses. We use data from the last 12 months to illustrate how market shares would have been different if the relative intensity of consumer feature searches had been different. To establish a baseline, we calculate for each vehicle the expected market share given the actual values of all the observed predictors and the model parameter estimates reported in the previous section. After establishing the baseline, we then calculate the "what-if" market share of each vehicle by allowing search for one of the five features to increase by 10% while holding the other feature searches and marketing mix variables constant. Finally, we compare the market share under the "what-if" scenario and the baseline. The corresponding percentage changes in market shares are reported in Table 5.

Each row in Table 5 shows the sensitivity of a vehicle's market share to changes in search for each of the features. For example, the market share of Toyota Corolla is most sensitive to change in search for cost to buy. Its market share increases by 6.81% with a 10% increase in search for cost to buy, consistent with Corolla being a small economy sedan. Also, not surprisingly, Corolla's gains 1.12% and .28% in market share with a 10% increase in search for fuel efficiency and cost to operate, respectively. Finally Corolla's market share would drop 0.91% and 2.30% with a 10% increase in search for acceleration and SUV, respectively.

Table 6 lists the most positively and negatively impacted vehicles for each feature, as measured by the percentage market share change in response to a hypothetical 10% increase in feature search. For instance, faced with increasing consumer search for fuel efficiency, the top three winners are Toyota Prius, Kia Optima, and Volkswagen Golf (all fuel efficient compact cars), and the three biggest losers are Toyota Sequoia, Ford Expedition, and GMC Yukon (all gas guzzlers). Less obviously, faced with increasing consumer search for terms related to cost to buy, Mini Cooper, Toyota Rav4, and Hyundai Santa Fe turn out to be the top winners, and Chrysler 300, Nissan Pathfinder, and Nissan Xterra the biggest losers.

Table 5 Effect of Feature Search Change on Market Share

Vehicle	Base line Market Share	Fuel Efficiency	Acceleration	SUV	Cost to Buy	Cost to Operate
Chevrolet Aveo	0.41%	1.74%	0.09%	0.42%	-1.57%	-4.52%
Chevrolet Corvette	0.12%	1.29%	1.79%	-1.95%	-13.13%	-12.16%
Chevrolet Impala	1.72%	-0.03%	-0.60%	-1.34%	-1.72%	-2.21%
Chevrolet Malibu	1.82%	0.65%	-0.46%	**-9.20%**	**8.51%**	**8.38%**
Chevrolet Avalanche	0.20%	**-2.62%**	0.43%	**17.45%**	-6.14%	**-20.72%**
Chevrolet Colorado	0.28%	0.89%	**6.55%**	8.99%	-9.73%	**-14.72%**
Chevrolet Silverado	3.64%	-1.67%	1.31%	4.26%	3.21%	-9.46%
Chevrolet Suburban	0.43%	**-2.46%**	1.00%	4.92%	-3.20%	-10.81%
Chevrolet Tahoe	0.74%	-2.04%	1.65%	5.77%	-7.47%	-6.05%
Chrysler 300	0.31%	**1.80%**	3.76%	**-12.08%**	**-26.85%**	-1.17%
Dodge Caravan	0.98%	-0.68%	**5.90%**	8.02%	-3.73%	-6.22%
Ford Crown Victoria	0.36%	0.69%	1.78%	2.86%	-1.45%	-2.87%
Ford Focus	1.63%	0.64%	0.88%	-3.99%	4.68%	0.16%
Ford Mustang	0.64%	**2.30%**	2.69%	-5.25%	-13.46%	2.59%
Ford Escape	1.86%	-1.17%	-0.95%	-2.58%	6.15%	7.71%
Ford Expedition	0.38%	**-3.23%**	3.50%	**20.95%**	-3.59%	**-16.91%**
Ford Explorer	0.75%	-2.14%	**7.96%**	**23.64%**	-7.52%	-13.94%
Ford F series	4.91%	-2.01%	2.49%	4.91%	3.36%	-4.55%
Ford Ranger	0.52%	-0.38%	3.69%	4.93%	0.79%	-10.96%
GMC Sierra	1.30%	-1.52%	0.30%	3.91%	2.52%	-9.75%
GMC Yukon	0.30%	**-3.04%**	1.02%	8.57%	-5.61%	-11.97%
Honda Accord	2.91%	0.88%	-1.06%	-3.47%	1.54%	0.42%
Honda Civic	2.50%	**2.06%**	-1.52%	-4.26%	3.99%	1.11%
Honda CR-V	2.08%	-0.43%	-3.85%	-2.41%	**7.80%**	3.18%
Honda Element	0.14%	-0.67%	1.64%	6.76%	-8.18%	-9.37%
Honda Odyssey	1.05%	0.64%	1.46%	-2.90%	-4.17%	-3.11%
Honda Pilot	0.90%	-1.54%	0.46%	-0.11%	-5.45%	4.04%
Hyundai Accent	0.52%	0.68%	-2.13%	-6.30%	2.63%	6.27%
Hyundai Elantra	1.32%	0.99%	0.63%	-1.17%	-1.55%	**10.55%**
Hyundai Sonata	1.51%	1.41%	-2.03%	**-8.72%**	5.65%	**10.84%**
Hyundai Santa Fe	0.67%	-1.91%	**-5.29%**	1.75%	**10.73%**	7.14%
Jeep Grand Cherokee	0.82%	-1.37%	2.86%	**14.84%**	-5.08%	-13.15%
Jeep Liberty	0.51%	-2.03%	**4.88%**	**12.22%**	-6.72%	-12.04%
Jeep Wrangler	0.93%	-0.37%	**-4.64%**	-3.25%	0.32%	-5.03%
Kia Optima	0.37%	**2.72%**	0.19%	6.14%	-2.59%	7.17%
Kia Rio	0.25%	1.24%	-1.04%	-6.99%	0.55%	2.80%
Kia Sedona	0.20%	**-2.67%**	4.01%	3.33%	-6.10%	-12.24%
Kia Sorento	0.93%	-1.69%	3.22%	-1.50%	-2.04%	**-15.47%**
Mazda3	0.87%	1.48%	-2.45%	-7.31%	0.31%	7.52%

Table 5 Continued

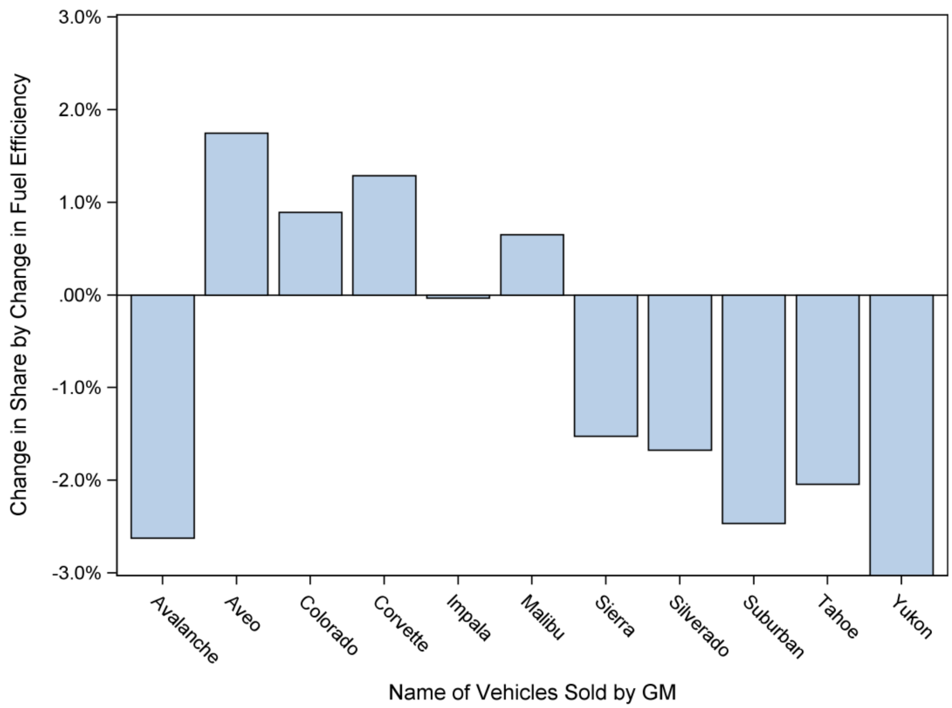
Vehicle	Base line Market Share	Fuel Efficiency	Acceleration	SUV	Cost to Buy	Cost to Operate
Mazda6	0.34%	-0.24%	1.25%	3.93%	-10.83%	-0.39%
Nissan Altima	2.23%	0.71%	-1.81%	0.16%	-2.29%	3.33%
Nissan Maxima	0.51%	-2.50%	2.21%	-6.21%	-1.00%	7.14%
Nissan Sentra	0.97%	1.30%	-0.34%	0.20%	-3.71%	6.62%
Nissan Armada	0.16%	-1.19%	0.47%	7.74%	-17.48%	-3.36%
Nissan Frontier	0.35%	0.16%	1.65%	6.75%	-13.53%	-0.14%
Nissan Murano	0.51%	-0.87%	-0.27%	-2.17%	-10.40%	0.10%
Nissan Pathfinder	0.20%	1.52%	0.81%	-1.90%	-22.19%	4.84%
Nissan Titan	0.21%	-1.43%	4.91%	6.89%	-18.89%	-6.99%
Nissan Xterra	0.18%	-1.72%	2.50%	7.69%	-19.96%	-10.28%
Ram	1.93%	-1.67%	2.54%	5.56%	1.39%	-8.82%
Scion xB	0.21%	2.06%	2.40%	-5.18%	-8.72%	-4.54%
Subaru Forester	0.74%	-1.62%	0.80%	-7.74%	1.81%	-0.43%
Subaru Impreza	0.42%	0.13%	-5.47%	-8.48%	10.19%	7.85%
Subaru Legacy	0.45%	0.75%	4.20%	-6.18%	-2.25%	-4.85%
Toyota Avalon	0.25%	1.52%	2.36%	-9.05%	-11.96%	-3.15%
Toyota Camry	3.27%	0.84%	-0.03%	-3.98%	3.38%	-1.10%
Toyota Corolla	2.74%	1.12%	-0.91%	-2.30%	6.81%	0.28%
Toyota Prius	1.37%	4.77%	-6.67%	-13.90%	-5.03%	21.52%
Toyota 4Runner	0.34%	-2.04%	0.74%	24.63%	-14.85%	-14.37%
Toyota Highlander	0.91%	-0.91%	-0.23%	4.48%	0.96%	-2.67%
Toyota RAV4	2.66%	-2.14%	-5.91%	-10.39%	11.86%	19.10%
Toyota Sequoia	0.15%	-3.38%	-0.21%	13.76%	2.11%	-14.92%
Toyota Sienna	0.92%	-0.63%	0.77%	1.71%	1.22%	-2.91%
Toyota Tacoma	1.04%	0.48%	-0.40%	-2.37%	-3.99%	-1.99%
Toyota Tundra	0.85%	-0.09%	-0.41%	1.77%	-2.83%	-4.65%
VW Jetta	1.25%	-0.29%	-4.01%	-8.14%	6.13%	11.63%
VW New Beetle	0.10%	1.41%	5.50%	-0.68%	-2.59%	-17.86%
VW Passat	0.32%	-0.64%	0.84%	1.31%	-12.19%	2.09%
VW Golf	0.24%	2.33%	-11.32%	0.10%	-2.79%	1.22%

Table 6 Top 10% Most Affected Vehicles with Feature Search Increase

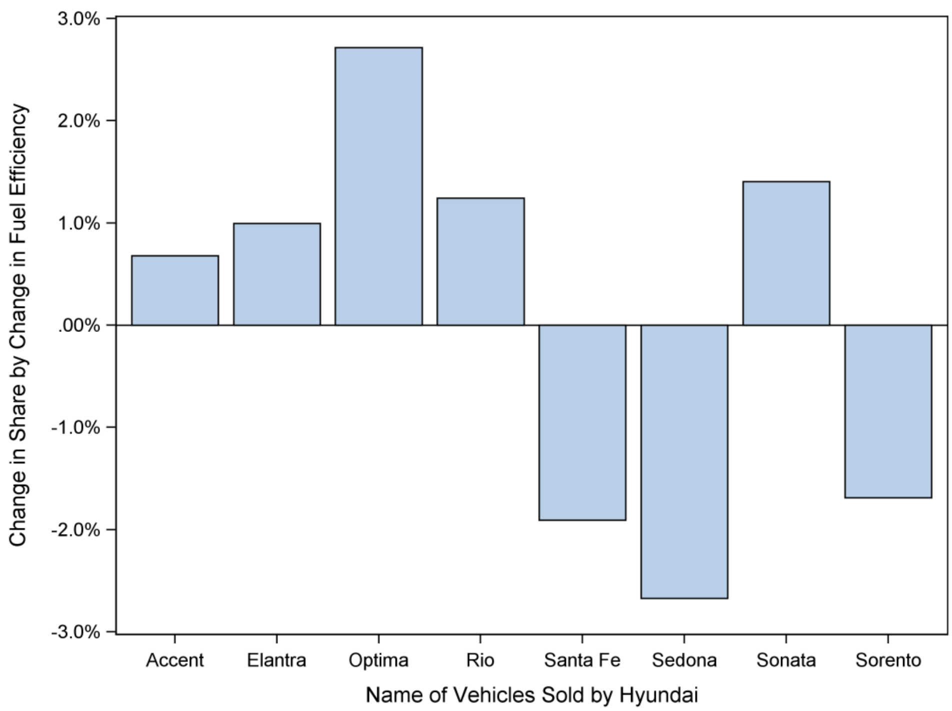
	Fuel Efficiency	**Acceleration**	**SUV**	**Cost to Buy**	**Cost to Operate**
Gaining the most	Toyota Prius	Ford Explorer	Toyota 4Runner	Mini Cooper	Toyota Prius
	Kia Optima	Chevrolet Colorado	Ford Explorer	Toyota RAV4	Toyota RAV4
	VW Golf	Dodge Caravan	Ford Expedition	Hyundai Santa Fe	VW Jetta
	Ford Mustang	VW New Beetle	Chevrolet Avalanche	Subaru Impreza	Hyundai Sonata
	Scion xB	Nissan Titan	Jeep Grand Cherokee	Chevrolet Malibu	Hyundai Elantra
	Honda Civic	Jeep Liberty	Toyota Sequoia	Honda CR-V	Chevrolet Malibu
	Chrysler 300	Subaru Legacy	Jeep Liberty	Toyota Corolla	Subaru Impreza
Loosing the most	Chevrolet Suburban	VW Jetta	Hyundai Sonata	Nissan Frontier	Toyota 4Runner
	Nissan Maxima	Jeep Wrangler	Toyota Avalon	Toyota 4Runner	Chevrolet Colorado
	Chevrolet Avalanche	Hyundai Santa Fe	Chevrolet Malibu	Nissan Armada	Toyota Sequoia
	Kia Sedona	Subaru Impreza	Toyota RAV4	Nissan Titan	Kia Sorento
	GMC Yukon	Toyota RAV4	Mini Cooper	Nissan Xterra	Ford Expedition
	Ford Expedition	Toyota Prius	Chrysler 300	Nissan Pathfinder	VW New Beetle
	Toyota Sequoia	VW Golf	Toyota Prius	Chrysler 300	Chevrolet Avalanche

Besides focusing on each individual product, managers can also use the above scenario analyses to quantify the impacts of feature search trends across product lines. Figures 2A and 2B show how vehicles from the portfolios of General Motors and Hyundai would be affected in the event of a 10% increase in consumer searches for fuel efficiency. Not surprisingly, we see that small fuel efficient cars such as Aveo and Elantra gain shares at the expense of large gas guzzling SUVs such as Yukno and Sorento. Similarly, Figures 2C and 2D show that most of Hyundai's vehicles would gain market shares as consumers search 10% more for cost to operate, while the opposite happens to GM's product portfolio. For managers, being able to quantify how market shares would change across their product lines in response to evolving consumer tastes can prove particularly useful in reallocating marketing resources across products, in order to better leverage favorable trends and alleviate the downsides of unfavorable ones.

In sum, scenario analyses such as the above can help quantify how market shares would shift in response to evolving customer tastes. These analyses can be most helpful when used in combination with forecasts of consumer search interests. Given that Google Trends data are readily available in real time, managers can update their projections of consumer trends with little time delay, quantify their market share implications and respond proactively.

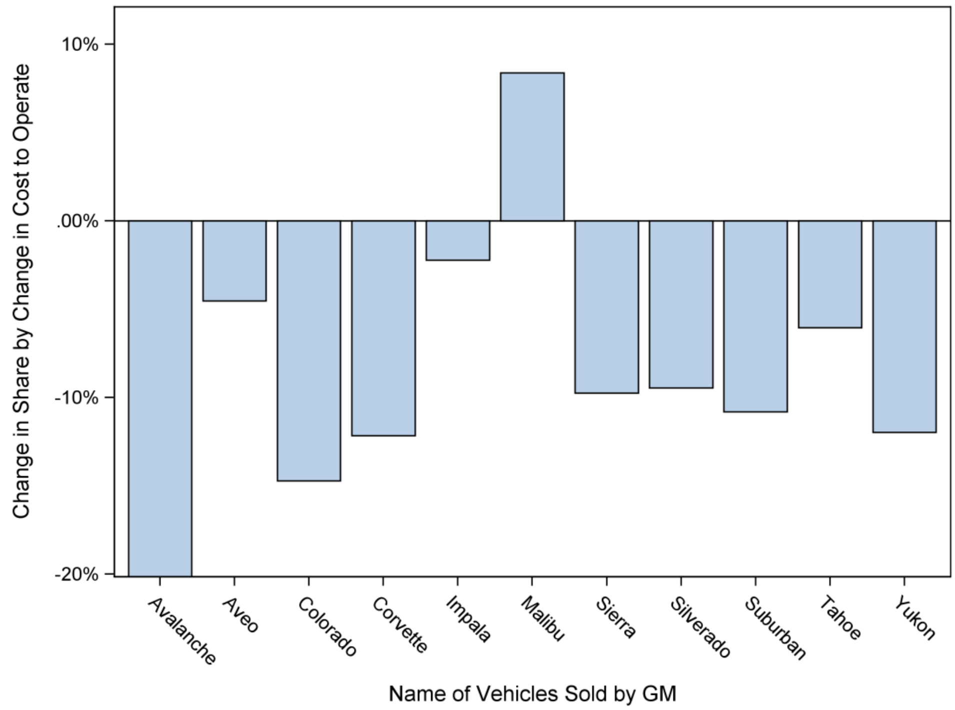


Panel A: GM Vehicles with Search for Fuel Efficiency +10%

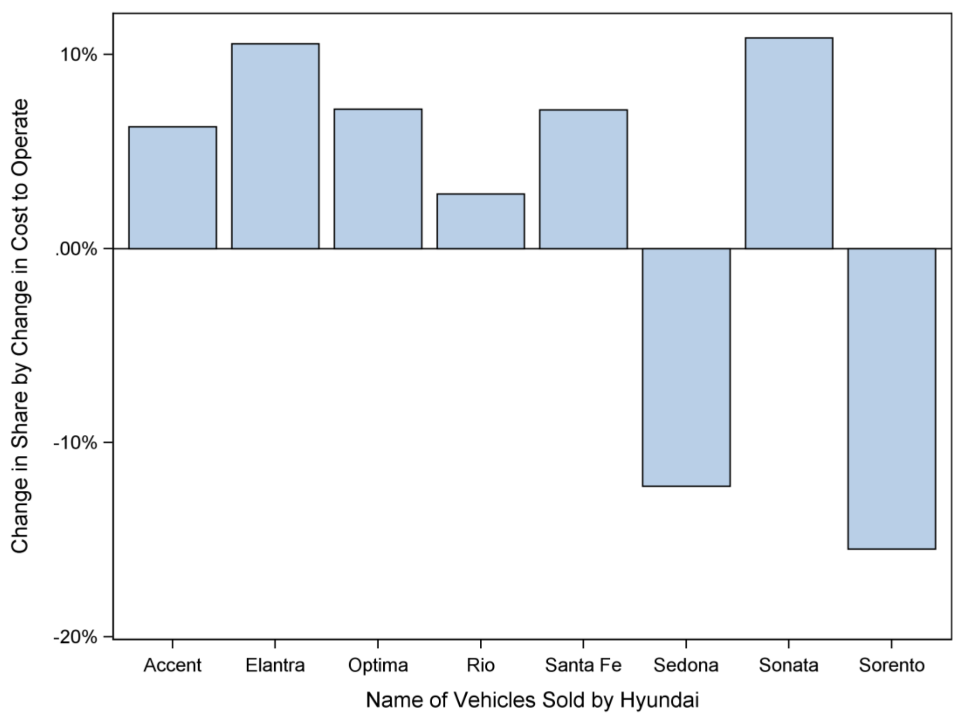


Panel B: Hyundai Vehicles with Search for Fuel Efficiency +10%

Figure 2 The Scenario Analysis of Market Shares



Panel C: GM Vehicles with Search for Cost to Operate +10%



Panel D: Hyundai Vehicles with Search for Cost to Operate +10%

Figure 2 Continued

CONCLUSION

We propose a hierarchical market share model using Internet product feature search index to capture the dynamic evolution in the baseline attractiveness of multi-feature products. Based on data from the U.S. automotive industry between 2004 and 2011, we empirically answer two important questions raised earlier. First, feature searches deliver a substantial improvement in explaining longitudinal market share variations (average $\mathrm{R}^2 = 50\%$ using our proposed model vs. 26% from the best benchmark model). Second, consistent with our expectation, the effect of feature searches on market share corresponds to actual feature levels. Specifically, an increase in the customers' interest in a positive feature such as fuel efficiency leads to an increase (decrease) in market shares of fuel efficient (inefficient) automobiles. Conversely, when people concern more about a negative feature such as the price of cars, more (less) expensive cars will lose (gain) market share. For all the five features we study, we observe the expected relationship between the actual feature level and the effect of feature search index on market shares.

We also demonstrate a wide range of applications of our approach using feature search index. First, marketers of multi-feature products can relate the market share of their products over time to evolving interests in features. Being able to dissect market taste trends by features allows marketers to fine tune their marketing mix and gain competitive advantage. When considering each product separately, marketers will be able to identify the most important features characteristizing the product, and adjust marketing messages to emphasize what counts most. When making decisions across products, marketers will be able to optimize marketing mix such as advertising spending on the whole product line. As consumer taste evolves among various features, marketers can improve their return of investment by better distributing spending on the products in advantageous situation and those in disadvantage. Over longer time harizon, marketers can even provide useful feedbacks for product designers to make product feature decisions in future markets.

Moreover, a firm can conduct a scenario analysis to quantify the effect of changes in customer taste on market share. Combining this estimate with profit margin information, marketers can take advantage of major trend shifts in the market to optimize their marketing effort. The scenario analysis can also be applied to evaluate the sensitivity of a whole product line to evolving trends. Marketers can thus better balance their marketing expense and sales force effort, and if possible, making more drastic changes (e.g., product line expansion) to retain competitive advantage. For example, suppose a product faces a declining market share. With the help of our method, the firm can diagnose whether the decline may be a result of decreasing (increasing) interest in a feature with negative (positive) effect on its market share. Examples in the automotive industry include fuel efficiency on Toyota Prius (positive) or Toyota Sequoia (negative). Using our results, it is also possible for the firm to decide the most effective way to cope with the decline. It can dynamically adjust its marketing messages to alleviate this declining interest. Sometimes, the effect of certain features may not be readily changeable overnight, as it can be tied up with the brand. Using our method, the firm can still identify strengths and weaknesses of each product at the feature level, and take extra effort to improve the product or brand image with respect to prominent features.

Given the availability of Google Trends data, the proposed method can easily supplement the current marketing research and business intelligence effort of the firm at very low cost. Although firms may be able to attempt to acquire similar information using traditional sources of data such as running waves of conjoint analysis over time, the alternatives are often difficult and sometimes may not even be possible. As the examples in Figure 1.1 shows, the applications of our proposed approach go beyond the automotive industry. It may apply to other categories of multi-feature products as well.

Our findings extend previous research (e.g. Choi and Varian 2009a; Du and Kamakura 2012) by expanding the dimension of Internet search beyond search for products themselves. Despite our effort and that of previous research, we have only scratched the surface of vast amount of Internet search data and there seems to be numerous opportunities on following the same direction in future research: going from specific search terms to generic terms. Researchers can acquire search volume for any combination of terms. It is conceivable that further expansion of the dimension of search terms, to even seemingly unrelated ones, may yield fruitful findings. In addition, we focus on the national automotive market. The evolution of interest in features can differ across geographical regions. This opens an interesting avenue for future research especially since Google Trends can readily serve as the tracking venue for different geographical regions. More specifically, a longitudinal-spatial analysis may link feature search to demographics, and provide further insights for manufacturers to adjust their product offerings and marketing communications across various markets and improve overall market demand.

LEVERAGING BIG DATA ON CO-CONSIDERATION IN MARKET RESPONSE MODELING

INTRODUCTION

The purpose of this essay is twofold: 1. Develop a method for representing the competitive interrelationship of the products using big data on consumers' online activities and 2. Use this representation of competition between the products for market structure analysis. Market structure analysis is a class of methods for representing the interrelationship of the products in a way that reflects consumers' evaluations of the products (Elrod 1991; Grover and Dillon 1985). Market structure analysis has been a pivotal subject in marketing theory and practice (Elrod 1991) since it can help answering several questions central for marketing decisions. It enables marketers to infer (1) the product positions against each other, (2) the consumers' preferences for the products given their positions in the market, and (3) the effect of competitors' marketing activities on the sales of the focal product. In this study, we contribute to this literature by showing how we can better analyze market structure using big data sources on consumers' online activities.

To represent the competitive interrelationship of the products, we develop and test a method that uses the data on consumers' online activities to measure co-consideration: the likelihood that two products are considered together by consumers. We apply our measurement method to the data on online quote requests for automobiles in the US. In this dataset, we observe the number of online quote requests over time across different geographical regions aggregated across several automotive websites. This dataset is particularly useful because it is a window into the final stages in the prospective consumers' purchase funnels. Thus, we call it Lower Funnel Prospects data (LFP data from now on).

Using LFP data has several advantages for (1) providing detailed insights about the consumers, (2) providing a holistic view about the market, and (3) potentially alleviating some common biases in empirical studies such as response bias and sampling bias. The LFP data are highly granular as a result one can focus the analysis on any product, geographical area, or time frame with no additional cost of data gathering. The data provide every single online quote requests for all automobiles available in the US market during the period of 2009 to 2011. In other words, it captures the whole phenomenon of online quote requests in the US. As a result, the insights from the data are not anecdotal or restricted to one particular subset of consumers. So, using this data set we can provide a holistic view of the market. Moreover, since auto purchase is a high involvement purchases, consumers are highly motivated to go online and get information about various aspects of the products (Ratchford, Lee, and Talukdar 2003; Ratchford, Talukdar, and Lee 2007). As a result, the LFP dataset is very informative for revealing the consumers' intentions and considerations. By capturing the consumer behavior, the LFP data can potentially reveal consumers purchase intentions better than stated intentions. Stated purchase intention in the survey studies can suffer from different form of reporting bias: participants report something different from their true intention (for example, refer to Morrison 1979). Furthermore, the LFP data can also reduce sampling bias since they records every single online quote requests. In terms of external validity, the reliability of the results from LFP data is comparable with a comprehensive survey study using a random sampling, covering the whole US market, and attaining high response rate. Such a survey needs an expensive data collection and can be cost prohibitive in practice. On the other hand, the LFP data can potentially provide comparable or even better results at a fraction of the costs.

Despite these appealing characteristics, we face several challenges in using LFP data. Users of online quote request services might use these tools for many different purposes. There is no guaranty that in every single occasion, when someone requests an online price quote, he or she is in fact considering buying the product. As a result, our data is contaminated with many sources of noises. Moreover, due to a combination of technical problems and privacy concerns, LFP data are not available at individual level. So, we are unable to directly observe co-consideration at individual level.

Given all the challenges we face, we question whether the proposed method can tease out the real signal about co-consideration. To establish the validity of our measure, we investigate the face validity of the results. We also show that additional information from the proposed measure of co-consideration can help explain market share which provides strong evidence for validity of the measure.

Particularly, we demonstrate our method in the context of the US automotive industry focusing on 23 major automobiles from sedan sector. First, we measure co-consideration between all pairs of automobiles in our set of 23 automobiles. The core idea behind the measure is as follows. Imagine an individual consumer who considers two automobiles before making the final choice. He/she is likely to request price quotes for these two automobiles within a short time interval. This behavior manifests itself in aggregate data in the form of co-occurrence: two quote requests for the two automobiles in the same geographical area and within a short time interval. The amount of co-occurrence above and beyond what we expect by chance is a measure for co-

consideration. We transform the measure of co-consideration into a perceptual map of product positions and show that the resulting perceptual map has face validity. We continue validating the measure by developing a market share model that identifies (1) multiple consumer segments that differ in their ideal points with respect to the perceptual map, (2) the relative sizes of these segments in each geographical area, and (3) the effect of marketing activities of competing products on the demand for the focal product. The model performs better when applied to the perceptual map based on our proposed measure of co-consideration than to the other alternative perceptual maps (i.e. based on market share and product features). This result shows that the perceptual map from our proposed measure of co-consideration provides useful information for predicting market share. Therefore, we have additional evidence that our measure of co-consideration captures the real latent competitive landscape of the market.

The challenges we face in and the advantages of using LFP data are not unique. In fact, they are common in many big data sources. In recent years, the capacity of computation, data storage, and communication along with availability of many devices that automatically record data makes it possible to generate, store, and analyze data at a very large scale. The combination of these technological advancements makes it possible to have new approaches in generating insights for decision making. The availability of the data on different aspects of consumers' behavior, at a wide scope potentially covering the whole market, at an unprecedented level of details opens new opportunities for marketers to better understand the consumers. But, at the same time we face several challenges in using big data sources. Big data sources tend to be noisy. As an analyzer, we can never be aware of all aspects of data generating processes. Since the data is usually gathered for other purposes, some critical pieces of information might be missing which makes inference hard. Effective use of big data sources for research and decision making is an attempt to generate meaningful signal from a noisy and messy data. Fortunately, in working with big data we usually have plenty of observations. The larger the sample size, the better chance that errors will smooth themselves out and allow reliable signals to emerge. This study is an example of dealing with such challenges in the specific context of big data on consumers' online activities.

Using the data on consumers' online activities for marketing decision making has been the subject of a growing body of research in recent years. Many studies using online consumer activities data focus on the volume of online activities related to single products (i.e. Choi and Varian 2009a and Du and Kamakura 2012). The basic premise of this body of research is that the volume of online activities is a proxy for the number of people who are considering the product. These studies, found that the volume of online activities is correlated with sales of the products in many different contexts.

In this study, we go beyond the activities related to a single product and investigate the co-occurrence of online activities related to the pairs of products. Our method of using co-occurrence is related to Netzer et. al. (2012) who applied text mining techniques to the online user generated contents. However, our study is different from Netzer et. al. (2012) in several ways. First, the way our method use co-occurrence stems from the data generating process of stream of consumers' online activities which is inherently different from the text mining approach. Second, the nature of the data we use is different from Netzer et. al. User generated contents is the voice of a relatively small but highly vocal segment of consumers whereas the data on consumer online activities we use captures the activities of the majority of the consumers at the time they are making purchase decision. As a result, our proposed method measures the competition between the products in a way that is more informative about the real purchase decisions. Finally, we go beyond only measuring the competition between the products. We propose a method for modeling the market structure based on our big data driven measure of co-consideration.

Traditionally, the studies of market structure fall into two general approaches. The first approach uses surveys of consumer consideration set to understand the extent to which the products are considered as substitutes and thus competing against each other (i.e. Urban et. al. 1984). The second approach uses purchase history to infer the underlying market structure from consumer choices (i.e. Cooper and Inoue 1996; Grover and Srinivasan 1987). It is well-known in the literature that we can have a much richer picture of market competitive structure using a combination of both approaches (Mackay, Easley, and Zinnes 1995). However, a combined approach is not always possible because it relies on surveys of consumer consideration sets which can be impractical and cost prohibitive. We contribute to this literature by showing that measure of co-consideration generated from secondary big data sources can take the roles traditionally played by surveys in market structure analysis.

The rest of the essay proceeds as follows. In the next session, we present our big data approach to measuring the co-consideration and investigate the face validity of the measure. Afterwards, we develop an empirical method that uses this measure of co-consideration as a proxy for competition between the products in a market structure analysis. Then, we present the empirical result which further validates our measure of co-

consideration and shows the usefulness of our novel approach in market structure analysis. We conclude by discussing the managerial implications of our study, its limitations, and directions for further research.

MEASURMENT

Overview

Our goal is to introduce a method to measure co-consideration by leveraging big data on consumer online activities. The underlying assumption in developing the measure is that consumers' online activities reflect products they consider. We go beyond simple consideration and investigate what products are considered together by consumers (or in other words co-considered). By measuring co-consideration we essentially measure the intensity of competition between the two products. Given the restrictions in the data, we are unable to count the exact number of people who co-consider two products. Instead, we resort to developing a measure that correlates with co-consideration. Having a measure of co-consideration is good enough for many applications.

The usefulness of our measure for decision making is a direct function of the scope of the data. The ideal situation is to build a measure that covers almost the whole market with data on online activities of users of a wide range of websites. The recent proliferation of Big Data, made such sources of data increasingly available. However, this increase in scope sometimes comes at a cost. Due to various technical and regulatory problems, these types of data in many cases are available only at aggregate level rather than individual level. Our method of measuring co-consideration should (1) be able to rigorously infer co-consideration from aggregate data (2) be computationally simple. It is impossible to implement a computationally complicated measure due to the size of the data. In short, our goal is to have a measure that balances the rigor and tractability.

The core idea is that observing consumers' online activities related to two products can be a sign that the two products are co-considered. Even in aggregate data, it is possible to find the traces of this behavior for the following reason. The more consumers co-consider two products the more likely it is to observe co-occurrences, the term we use for online activities related to the two products from the same geographical region within a close proximity of time. Depending on the nature of the data, raw co-occurrences might not be a valid measure. We need to do some adjustment to create a valid measure of co-consideration. In the following section we describe in details how we operationalize the measure of co-consideration using the specific data we have. We can modify the method presented here to suit other aggregate data on online activities related to product purchase. Regardless of the nature of the data we can keep the two following basic building blocks intact: (1) start from co-occurrence of online activity (2) do some adjustments to control for the effect of confounding variables leading co-occurrences.

Any data on consumers' online activities is inherently noisy. Consumers use online tools for many different purposes. As a result, our assumption that there is a simple link between online activities and consideration might not hold for every single observation in the data. Eventually, what matters is whether we can infer the patterns of co-consideration given all these noises.

In the absence of a true measure, to decide whether our measure is in fact a proxy for co-consideration, we rely on two pieces of evidence, (1) face validity of the measure and (2) ability to predict market share. In theory, we expect similar automobiles to be co-

considered more often than dis-similar ones. First, we investigate if our measure seems to be related to similarities between automobiles. Next, we use Multidimensional Scaling to generate a perceptual map of product positions. We investigate the face validity of the resulting perceptual map. Finally, we develop a market share model which uses the product positioning map along with marketing activities to explain market share. If our proposed perceptual map can help explain market share, the map has useful information about the true competitive structure of the market. This is strong evidence that the measure is valid and reflects the competition between the products.

Operationalization

We demonstrate our method in the context of US auto industry using the data provided to us by Autometrics Company. Autometrics data record the number of online quote requests for new vehicles on a real-time basis down to the ZIP code level. Autometrics aggregates online quotes requests across a wide range of American automotive websites such as edmunds.com, kbb.com, and cars.com. These websites offer tools to enable the consumers to find auto dealers nearby (see Figure 3 for an example for an example of the interface for requesting online price quotes). Consumers enter the car they consider and their ZIP code. Then the website connects them to the nearby auto dealers. The side product of this process is a huge dataset on consumers' intentions and considerations. While the data can have several potential applications, in this study we focus on generating competitive intelligence insights from the co-occurrence of online price quote requests.

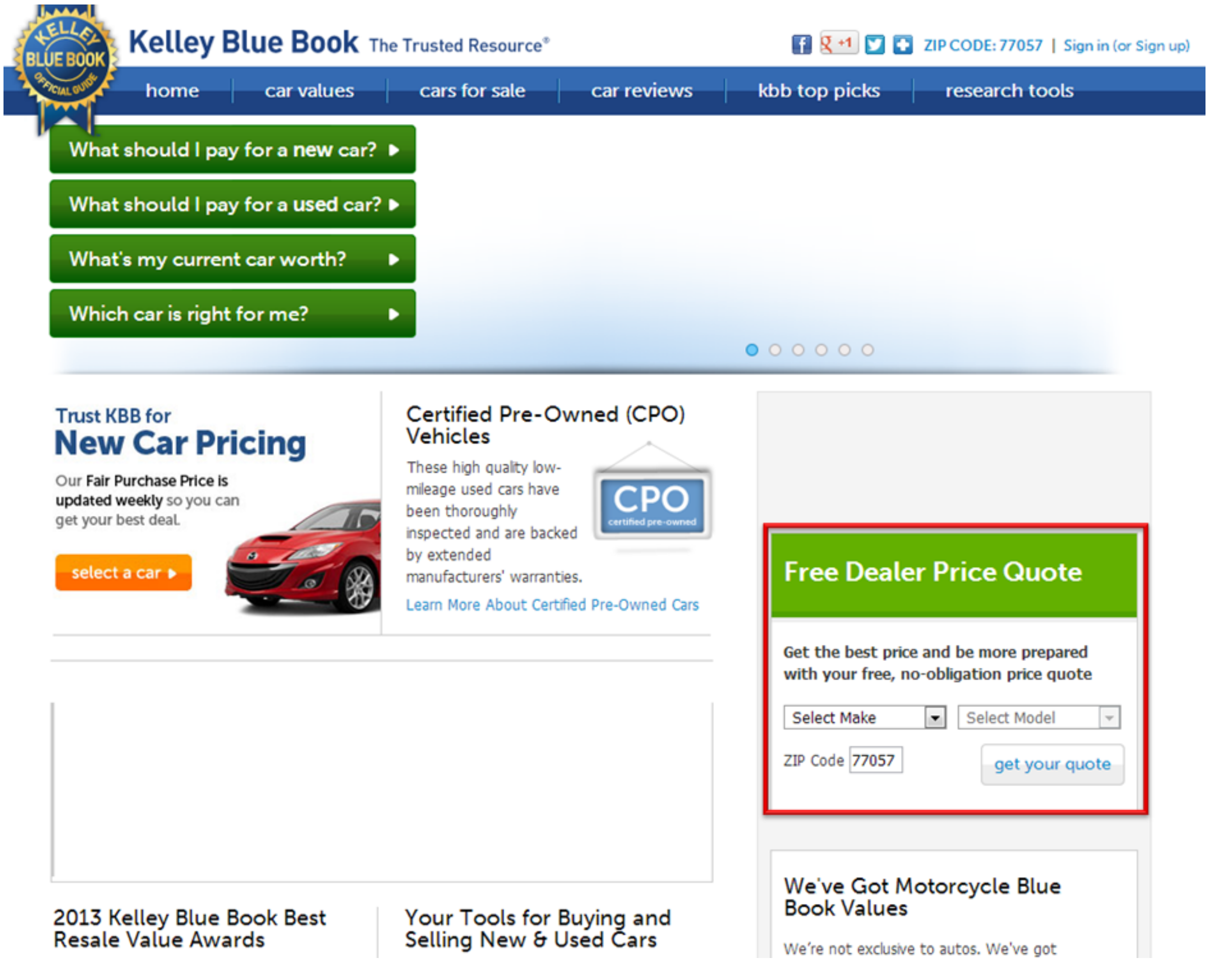


Figure 3 a snap-shot of the user interface for price quote requests

We use co-occurrence of online price quote requests as a sign of co-consideration. We operationalize co-occurrences as online price quote requests for two distinct automobiles within the same five minute interval from the same ZIP code. This method does not take into account the possibility that a consumer may consider two automobiles but request quotes for both automobiles with a larger time interval in between. We will see that our proposed operationalization result in a valid measure in terms of face validity and ability to explain market share. However, it is still unclear whether 5 minute interval is the best option or not. It is possible to run a sensitivity analysis in which one tests various time intervals and investigating whether they lead into better results or not. We leave such a sensitivity analysis for the future work.

The raw co-occurrence count is not useful directly though, because a large number of co-occurrences can happen purely by chance. Take two automobiles with a large number of price quote requests, such as Toyota Camry and Honda Accord, as an example. The large number of price quote requests makes it much more likely that quote requests for these two automobiles happen at the same time and place. An informative measure shall tell us how many co-occurrences happen above and beyond chance.

Association rule learning is the discipline in data mining that deals with such problems: finding meaningful relations between variables in large databases. One commonly used criterion for identifying strong association rules is lift, defined as the ratio of observed co-occurrence to expected co-occurrence if the two automobiles were independent. We extend this notion originally designed for shopping basket analysis to construct a lift consistent with the unique data generating process in LFP data.

We illustrates how our approach works with an example. Table 7 shows the real numbers of LFPs on December 30, 2011 at ZIP code 95832 located in Sacramento, CA. We observe that 8 quotes for Toyota Camry and Honda Accord are requested at the same time during December 30, 2011 at ZIP code 95842. To interpret this number we should ask how many co-occurrences one would observe by shear chance, given that we observed 22 quotes for Toyota Camry and 8 quotes for Honda Accord. Chance alone is enough for some of these quote requests to happen at the same time even if all of the quote requests were from independent consumers. Comparison of the observed co-occurrences and the expected co-occurrences if requests were independent gives us a measure of co-consideration.

Table 7 The Number of Quote Requests and Co-occurrences for Select Cars (on 12-30-11 at ZIP code 5842)

Number of quotes for Camry	22
Number of quotes for Accord	8
Number of quotes for Fusion	5
Number of quotes for Camry and Accord happening at the same time	8
Number of quotes for Camry and Fusion happening at the same time	5
Number of quotes for Accord and Fusion happening at the same time	2

The first step is to calculate the expected number of co-occurrences if the requests were independent. Calculating expected value of co-occurrence analytically is tricky. So, we use numerical simulation to calculate the expected value of co-occurrences. Assuming that each quote request happens at an independent random period, we can randomly simulate 22 quote requests for Toyota Camry and 8 quote requests for Honda Accord. Then, we can count the number of simulated random co-occurrences. By repeating this simulation process millions of times, we can run a Monte Carlo simulation. The mean of co-occurrences over the whole sample from Monte Carlo simulation approaches the expected value of co-occurrences assuming that quotes are generated independently.

The fact that the number of quotes fluctuates during the day should be controlled in calculating expected co-occurrences. Figure 4 shows the total number of quote requests in hour during the day. We see that the number of quote requests peaks in the afternoon and drops to the minimum around 4 o'clock in the morning. It suggests that the probability of random co-occurrence is high in the afternoon and low early in the morning. The higher the fluctuation the more likely it is to observe random co-occurrences. So, in Monte Carlo simulation, instead of assuming that it is equally likely for quote requests to happen in any time period, we assume that the probability of occurrence of quote requests follows the distribution suggested by Figure 4. Table 8 illustrates the expected value of co-occurrence for Toyota Camry, Honda Accord, and Ford Fusion on 30 December, 2011 in ZIP code 95842 based on the simulation result. We can define the ratio of actual co-occurrence to expected value of independent co-occurrence as lift between the two cars. The higher the lift, the higher co-consideration is between the two automobiles.

Table 8 The Number of Quote Requests and Expected Co-occurrences for Select Cars (on December 30, 2011 at ZIP code 95842)

Number of quotes for Camry	22
Number of quotes for Accord	8
Number of quotes for Fusion	5
Expected Value of independent random co-occurrence of Camry and Accord	.429
Expected Value of independent random co-occurrence of Camry and Fusion	.678
Expected Value of independent random co-occurrence of Accord and Fusion	.160

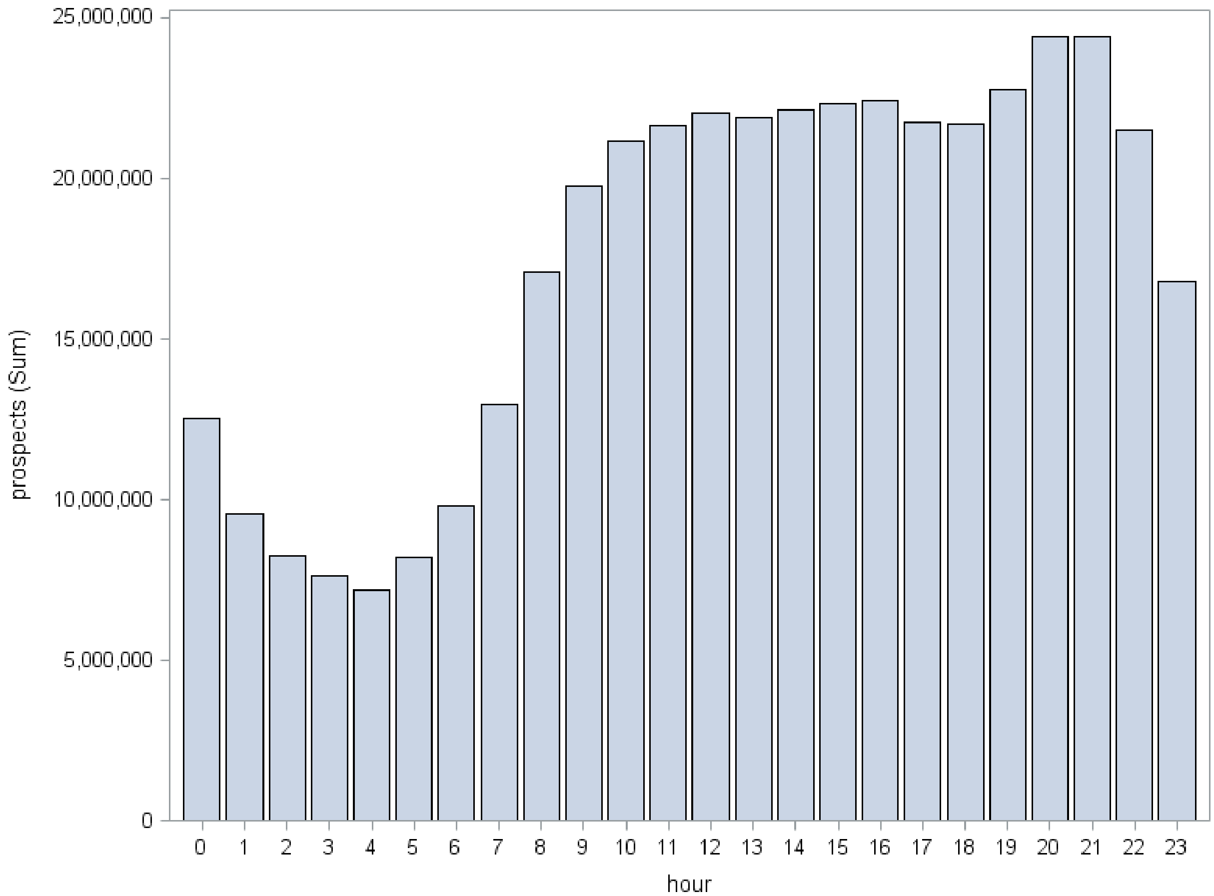


Figure 4 Variations in The Number of Quote Requests During The Day

We can calculate the lift at national level by sum up actual and expected co-occurrences over time and across ZIP codes and then taking then ratio. Table 9 reports the total number of co-occurrences and expected independent co-occurrences across all ZIP codes from January 2009 to December 2011. A higher level of the lift reflects a higher level of co-consideration, meaning that a larger number of consumers consider the two automobiles. As more people consider two cars together, the competition between them intensifies. Therefore, lift is essentially a measure of competition intensity between two products.

Table 9 Total Numbers of Co-occurrences and Expected Independent Co-occurrences (Across all ZIP codes and During the Whole Period of Analysis)

Automobiles		Expected Independent Co-occurrence	Actual Co-occurrence
Camry	Accord	3,076	8,446
Camry	Fusion	2,413	2,794
Accord	Fusion	1,636	1,793

Examining the face validity of the measure

We focus on 23 major sedan automobiles. Selecting these automobiles is for demonstration purposes only. We can replicate the analysis for any given set of automobiles. The set of 23 products result in 253 possible unique pairs of automobiles. We calculate the lift for all of these 253 pairs.

First, we investigate the face validity of the result for highest lift values. In theory, we expect consumers to co-consider similar cars. So, we expect the automobiles with high lift to be more or less similar to each other. Table 10 shows ten pairs of automobiles with the highest lift value among 253 pairs. We see that almost all pairs with high level of lift share the same body size. Usually the household economic situation and family size dictates what body size best suits the household’s needs. So, it is very likely for consumers to consider cars with similar body sizes. So consistent with theory, for cars sharing the same body size we observe high lift. Another interesting observation is that automobiles with the same brand typically have a very high lift .For examples see the lift of Volkswagen Passat and Jetta, Mazda6 and Mazda3, and Subaru Impreza, Legacy, and Forester. This observation is reasonable. It suggests that a significant portion of consumers are relatively loyal to one brand. High lift between the automobiles with same body size and high lift between the automobiles with the same brand are two patterns common across all the calculated lifts. Observing these two reasonable patterns provide some evidence that the lift has face validity as a measure of co-consideration.

Table 10 Ten Pairs of Automobiles With the Highest Lift Value

(The Label in Parenthesis Indicates The Body Size)

Pairs of Automobiles				Lift
Volkswagen Passat	(mid-size)	Volkswagen Jetta	(compact)	20.5
Mazda6	(mid-size)	Mazda3	(compact)	15.0
Subaru Legacy	(mid-size)	Mazda6	(mid-size)	13.3
Mitsubishi Lancer	(compact)	Subaru Impreza	(compact)	9.3
Subaru Legacy	(mid-size)	Kia Optima	(mid-size)	9.3
Mazda6	(mid-size)	Kia Optima	(mid-size)	8.7
Subaru Legacy	(mid-size)	Subaru Impreza	(compact)	8.1
Mitsubishi Lancer	(compact)	Mazda3	(compact)	7.3
Mitsubishi Lancer	(compact)	Nissan Sentra	(compact)	6.8
Mazda6	(mid-size)	Nissan Altima	(mid-size)	6.7
Nissan Sentra	(compact)	Toyota Corolla	(compact)	6.6
Mitsubishi Lancer	(compact)	Mazda6	(mid-size)	6.1
Volkswagen Passat	(mid-size)	Mazda6	(mid-size)	6.1
Subaru Impreza	(compact)	Subaru Forester	(full-size)	6.0
Subaru Legacy	(mid-size)	Subaru Forester	(full-size)	5.8
Toyota Avalon	(full-size)	Nissam Maxima	(full-size)	5.8
Mazda6	(mid-size)	Honda Accord	(mid-size)	5.7
Mazda6	(mid-size)	Hyundai Sonata	(mid-size)	5.5
Nissan Sentra	(compact)	Mazda3	(compact)	5.5
Toyota Corolla	(compact)	Honda Civic	(compact)	5.4

Next, we investigate whether our proposed measure provides reasonable representation of the competitive landscape of the whole market. Table 11 shows the lift for all 253 pairs. It is hard to get a holistic picture from such a table with these many figures. To overcome this problem, we use the lifts as the measure of similarity in Multidimensional Scaling (MDS). This way, we can summarize the lifts into a two dimensional perceptual map (Figure 5). In this perceptual map, the automobiles close to each other have high lift. So, two products positioned close to each other on the map compete fiercely against each other. The result is easy to interpret and have high face validity. On left bottom of the map we see large sedans. Automobiles in the center are mid-size sedans. As we continue our path to the top right of the map we reach efficient and compact sedans. This gradual shift from large to small is consistent with our expectations about sedan sector in the US auto industry. Moreover, on the right bottom corner we see automobiles with four wheel-drive and outdoor capability which form a differentiated sub-segment of sedans. Overall, these results have good face validity and give us some confidence that we successfully measure the amount of co-consideration overlap.

In what follows, we develop a model to utilize the perceptual map (such as the one in Figure 5) for analyzing the market structure and predicting market shares. The model not only provides the results useful for a wide range of marketing decisions, but also provides a framework to test the validity of our measure of co-consideration.

Table 11 Lift for All Pairs of 23 Automobiles in the Analysis

	Camry	Accord	Civic	Corolla	Altima	Prius	Focus	Sonata	Malibu	Jetta	Mazda 3	Impala	Sentra	Forester	Maxima	Impreza	Optima	Chrysler300	Avalon	Legacy	Mazda 6	Passat	Lancer
Camry		4.8	2.7	4.1	4.5	2.6	1.4	3.4	3.4	1.9	1.9	2.7	2.6	1.1	3.6	1.2	3.0	2.2	4.8	4.5	5.3	2.5	2.1
Accord	4.8		3.5	3.0	4.6	1.5	1.7	3.4	3.2	1.8	2.3	2.9	2.6	1.1	4.0	1.3	3.4	2.7	3.2	4.2	5.7	3.1	2.5
Civic	2.7	3.5		5.4	2.9	2.9	2.5	2.0	2.3	2.4	4.4	1.7	4.6	1.1	2.1	2.3	1.9	1.5	1.4	1.7	3.4	1.5	4.3
Corolla	4.1	3.0	5.4		2.8	2.5	2.7	2.1	2.4	2.6	5.2	1.6	6.6	1.1	2.0	2.0	2.1	1.2	2.3	1.7	3.3	1.7	5.1
Altima	4.5	4.6	2.9	2.8		2.1	2.0	3.5	3.7	2.0	2.7	2.9	5.0	1.1	5.1	1.6	4.0	2.4	2.5	5.2	6.7	2.9	3.0
Prius	2.6	1.5	2.9	2.5	2.1		1.2	1.7	1.5	1.9	2.1	1.0	1.7	1.0	1.0	1.4	1.4	0.9	1.9	1.1	1.5	1.6	1.4
Focus	1.4	1.7	2.5	2.7	2.0	1.2		1.6	1.8	1.8	3.6	1.3	3.6	0.7	1.1	2.0	3.9	0.9	0.7	3.7	3.9	1.5	3.9
Sonata	3.4	3.4	2.0	2.1	3.5	1.7	1.6		3.3	1.5	1.9	2.3	2.1	1.0	2.0	1.1	4.8	2.0	2.0	4.5	5.5	2.2	1.9
Malibu	3.4	3.2	2.3	2.4	3.7	1.5	1.8	3.3		1.6	2.1	4.7	2.8	1.0	2.8	1.2	3.3	3.0	1.9	2.2	4.2	1.9	2.3
Jetta	1.9	1.8	2.4	2.6	2.0	1.9	1.8	1.5	1.6		3.2	1.1	2.7	1.5	1.6	2.5	1.7	1.3	1.3	2.8	3.3	20.5	3.7
Mazda 3	1.9	2.3	4.4	5.2	2.7	2.1	3.6	1.9	2.1	3.2		1.4	5.5	1.3	2.0	4.6	2.2	1.4	1.2	2.5	15.0	1.6	7.3
Impala	2.7	2.9	1.7	1.6	2.9	1.0	1.3	2.3	4.7	1.1	1.4		1.8	0.7	3.2	0.8	2.2	3.8	2.6	1.6	2.8	1.6	1.7
Sentra	2.6	2.6	4.6	6.6	5.0	1.7	3.6	2.1	2.8	2.7	5.5	1.8		1.1	3.5	2.7	2.3	1.2	2.1	2.0	3.6	1.5	6.8
Forester	1.1	1.1	1.1	1.1	1.1	1.0	0.7	1.0	1.0	1.5	1.3	0.7	1.1		1.1	6.0	1.0	0.8	0.8	5.8	1.3	1.8	1.4
Maxima	3.6	4.0	2.1	2.0	5.1	1.0	1.1	2.0	2.8	1.6	2.0	3.2	3.5	1.1		1.5	2.5	4.1	5.8	3.8	4.8	4.1	2.6
Impreza	1.2	1.3	2.3	2.0	1.6	1.4	2.0	1.1	1.2	2.5	4.6	0.8	2.7	6.0	1.5		1.4	1.3	0.9	8.1	2.3	1.8	9.3
Optima	3.0	3.4	1.9	2.1	4.0	1.4	3.9	4.8	3.3	1.7	2.2	2.2	2.3	1.0	2.5	1.4		2.1	1.6	9.3	8.7	2.8	2.6
Chrysler300	2.2	2.7	1.5	1.2	2.4	0.9	0.9	2.0	3.0	1.3	1.4	3.8	1.2	0.8	4.1	1.3	2.1		3.8	2.5	2.8	2.1	2.6
Avalon	4.8	3.2	1.4	2.3	2.5	1.9	0.7	2.0	1.9	1.3	1.2	2.6	2.1	0.8	5.8	0.9	1.6	3.8		1.5	2.8	2.5	1.7
Legacy	4.5	4.2	1.7	1.7	5.2	1.1	3.7	4.5	2.2	2.8	2.5	1.6	2.0	5.8	3.8	8.1	9.3	2.5	1.5		13.3	4.1	3.3
Mazda 6	5.3	5.7	3.4	3.3	6.7	1.5	3.9	5.5	4.2	3.3	15.0	2.8	3.6	1.3	4.8	2.3	8.7	2.8	2.8	13.3		6.1	6.1
Passat	2.5	3.1	1.5	1.7	2.9	1.6	1.5	2.2	1.9	20.5	1.6	1.6	1.5	1.8	4.1	1.8	2.8	2.1	2.5	4.1	6.1		1.9
Lancer	2.1	2.5	4.3	5.1	3.0	1.4	3.9	1.9	2.3	3.7	7.3	1.7	6.8	1.4	2.6	9.3	2.6	2.6	1.7	3.3	6.1	1.9	

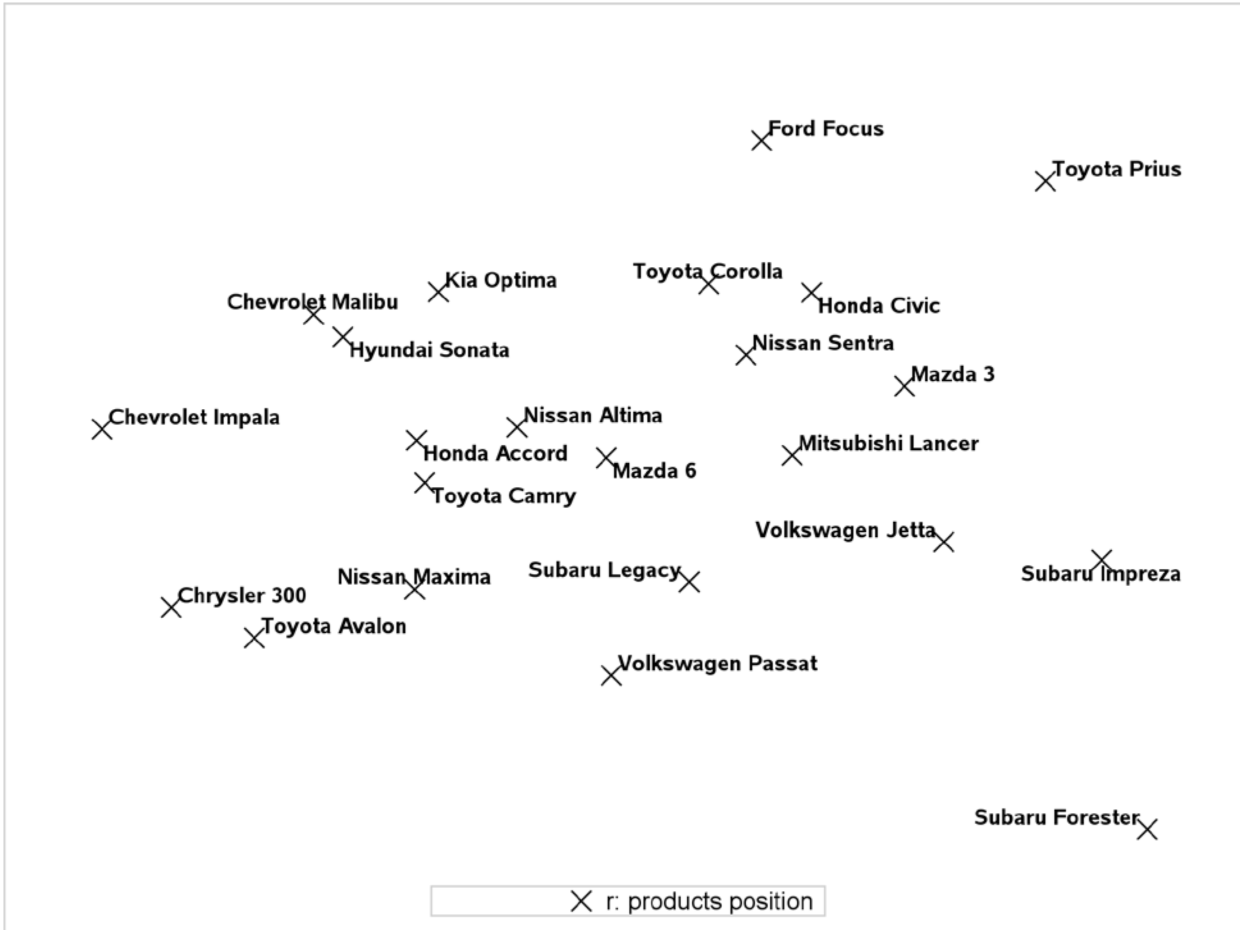


Figure 5 Perceptual Map Using Proposed Measure of Competition (LFP Map)

MODELING FRAMEWORK

In this paper, we propose a method to use the information on competitive positions of the products extracted from big data sources to improve the analysis of market structure. The market structure analysis is the process for representing the interrelationship of a set of products in a way that reflects consumers' evaluations of the product (Elrod 1991; Grover and Dillon 1985). Ideal-point models have been frequently used to describe market structure (Mackay, Easley, and Zinnes 1995). In an ideal-point model, each product is represented by a coordinate in the space. Each consumer or consumer segment has an ideal-point in the same space. The utility the consumers receive from each product is an inverse function of the distance between the product coordinate and consumers' ideal point. The previous studies using ideal-point models take several approaches to generate a spatial representation of product positions. These approaches includes using product characteristics, survey on consumers' perception of product attributes, survey on consumers' perception of differences between the products, or estimates of latent product positions from revealed preference. We propose an alternative approach in which we infer product positions from the data on online activities and then use this spatial representation of the products as the basis of our ideal-point model.

We assume that the underlying market structure can be represented in a D-dimensional preference space. We assume that each consumer segment has an ideal point represented by vector s_i where $i \in \{1 \cdots m\}$ is the index of consumer segment and each product have a coordinate represented by vector r_j where $j \in \{0 \cdots J\}$ is the index of the product. For example in Figure 2.3, we have a 2-demensional representation of preference space. The position of each product on this map is one of the r_j vectors.

We start with random utility theory and assume that each purchase decision we observe in the aggregate sales data is independent of all other purchase decisions (as if each sales is from an independent individual a reasonable assumption in the context of durable goods). Following the literature on ideal-point models, we assume that the utility from purchase increases by 1) an increase in marketing activities and 2) the r_j to s_i proximity which means how well the product matches the preference of the customers (hereinafter referred to as "preference match"). More specifically, the consumers incur a disutility which is a quadratic function of the Cartesian distance between s_i and r_j .We represent the distance between s_i and r_j by $\left|r_{jd} - s_{id}\right| = \sqrt{\sum_{d=1}^{D}\left(r_{jd} - s_{id}\right)^2}$ where r_{jd} and s_{id} are the d'th element of the vectors r_j and s_i. So, the utility from purchasing product j in each purchase occasion, h, assuming that the customer belongs to the segment i, will be

$$(1) \qquad u_{hjqt}|h\epsilon i = \rho_i\left(x_{jqt}\beta\right) - \delta_i\left(\left|r_{jd} - s_{id}\right|\right)^2 + \epsilon_{hjqt}$$

where $q \in \{1 \cdots Q\}$ and $t \in \{1 \cdots T\}$ are the indexes of market and time respectively. x_{jqt} is a k-dimensional vector of the marketing activities and β is the vector of the effect of those marketing activities. $\delta_i$and ρ_i are positive segment specific scaling parameters. The ratio of $\delta_i$and ρ_i accounts for the sensitivity of segment i to preference match relative to marketing activities. In the proposed model, we should estimate the ideal-point for each segment s_i, whereas we observe r_j's. ϵ_{hjqt} is a random disturbance.

We did not include a product specific intercept in equation 1for two reasons. First, we propose this model to identify the preference of the customers for the horizontal differentiation between products and vertical differentiation is not our focus. Second, in terms of choice probabilities there is no difference between a product with high intercept and a product with large nearby segments. In other words, we can increase the intercept and change the segment locations simultaneously without changing choice probabilities. Therefore, we cannot simultaneously identify product specific intercept and unobserved locations.

We further assume that ϵ_{hjqt} are independent, identically distributed extreme value I random variables. Therefore, P_{ijqt}, the probability of choosing product j conditional on the purchase being by a customer from segment i is as follows

(2) $$P_{ijqt}(X_{qt}, \theta_i | h\epsilon i) = \frac{\exp\left(\rho_i(x_{jqt}\beta) - \delta_i(|r_{jd} - s_{id}|)^2\right)}{\sum_{j=0}^{J} \exp\left(\rho_i(x_{jqt}\beta) - \delta_i(|r_{jd} - s_{id}|)^2\right)}$$

where $\theta_i = [\beta, \rho_i, s_i, \delta_i]$ is the vector of parameters.

We take the alternative $j = 0$ as the base alternative. Given our assumption that $r_0 = 0$, the utility derived from purchasing the base alternative is

(3) $$u_{hjqt}|h\epsilon i = \rho_i(x_{0qt}\beta) - \delta_i {s_i}^2 + \epsilon_{h0qt}$$

The unconditional probability of choosing a product for the whole market is a mixture of the probabilities from equation 2.

(4) $$P_{jqt}(x_{jqt}, \theta, \lambda_q) = \sum_{i=1}^{m} \lambda_{iq} P_{ijqt}(x_{jqt}, \theta_i | h\epsilon i)$$

where $\lambda_q = [\lambda_{1q} \cdots \lambda_{iq} \cdots \lambda_{mq}]$ is the vector of mixing probabilities. λ_{iq} is the likelihood of finding a customer from segment i in market q. Following the extensive literature on latent variable models (see, e.g., Dillon and Mu- lani 1989; Kamakura and Russell 1989), we can interpret λ_{iq} as the relative size of the segment i in market q.

We can use the unconditional probability of choice, described in equations 2 and 4, to infer both segments sizes and parameters. We observe the number of products sold in each market in each time period, y_{jqt}. We treat each product sold as an independent purchase decision. The likelihood of observing a certain sales history is

$$\mathcal{L}(y, X, \theta^*, \lambda) = \prod_t \prod_q \prod_j \prod_{h=1}^{y_{jqt}} P_{jqt}\left(x_{jqt}, \theta^*, \lambda_q\right) \tag{5}$$

Expanding the unconditional probabilities and taking log we can calculate the log-likelihood as follows:

$$\ell(y, X, \theta^*, \lambda) = \sum_t \sum_q \sum_j \sum_h^{y_{jqt}} \log\left(\sum_{i=1}^{m} \lambda_{iq} P_{ijqt}\left(x_{jqt}, \theta_i^* \middle| h\epsilon i\right)\right) \tag{6}$$

We use the EM algorithm to maximize the likelihood in equation 6.

EMPIRICAL ANALYSIS

Data

For our empirical analysis, we focus on 23 top automobiles in sedan sector of the US auto industry. The sales of these top automobiles account for 80% of total sales of sedans. Our monthly data spans 41 periods from January 2009 to May 2012. For our analysis we need data at regional level. Our regional data is at the level of Nielsen's Designated Market Area (DMA). We focus on top 50 DMAs accounting for 62% of sales and 74% of advertising spending in the US auto market. We also tried using more DMAs in the analysis and found that the main results are not sensitive to the number of DMAs.

For the 23 automobiles, we compiled historical sales figures, advertising spending, and total incentive expenditure per automobile. Incentive expenditure is the sum of all promotional expenditures including but not limited to cash back to the consumers, cost of offers for financing the cars at promotional APRs, and trade promotions paid to the dealers. All the data are monthly and at DMA level. We augment these data with the data on product positions from the perceptual map in Figure 5 Moreover, to judge whether the perceptual map using LFP data is a good representation of product positions, we also use two alternative maps of product positions which we discuss next.

Alternative Product Positioning Maps

As we demonstrate earlier, we can use our proposed measure of co-consideration to generate a perceptual map representing the competition between the products in the market. Consumers typically co-consider products that are similar to each other. Fundamentally, measure of co-consideration is closely related to measures of similarity of products. So, we can produce alternative maps of product competitive positions using other measures of similarity of products.

The first alternative product positioning map we use captures the similarity of the products in terms of their features and characteristics (the feature map hereafter). From edmonds.com, we collected the data on the three prominent features: gas mileage, price, and power. We transform the continuous feature levels to discrete values. To do so we categorize the values in the same quadrant together which result into four levels labeled as 1, 2, 3, and 4 for each of the three discrete dimensions. We use the automobile brand as the fourth discrete dimension taking values of 1, 2… 7. For each pair of automobiles, we construct the measure of similarity by adding the number of dimensions with the same values. Then, we apply MDS to these measures to generate the feature map.

The second alternative product positioning map captures the similarity in demand patterns. We use the automobiles monthly market share at DMA level and calculate the correlation between all 253 possible pairs of automobiles. The correlation is calculated over time and across DMAs. Homogeneity of preferences across DMAs and over time make it possible to treat correlation as a similarity measure. For example, consumers in Bay Area generally value environmental friendly automobiles whereas consumers in Houston prefer more powerful and large automobiles. As a result high correlation across regions conveys similarity between the products. Moreover, consumers with similar preference tend to go to the market at the same type. For example, college students, usually buy new automobiles in summer before start of fall semester. Such behavior gives raise to similar seasonality for similar products. So, we expect similar automobiles to have high correlation with each other. Again, we use MDS to transform correlations in market share into a two dimensional representation which we call the market share map from now on.

We estimate three models. We keep the model structure and the data on sales and marketing activities identical across the three models. But, for each model, we use one of the alternative positioning maps. The feature map reflects a plausible representation of the competition between sedans in US market. So, it leads to highly interpretable results and serves as a strong benchmark to test the validity of our proposed map. Market share map reflects the similarities in market share variations across regions and over time. So, it does not reflect market competitive landscape as good as the two other alternative maps. But, the market share map contains direct information about market share. This direct information about market share can substantially improve the performance of the model in predicting market shares. So, we expect it to serve as a strong benchmark for the performance of the proposed model. In the following sections, we compare the performance of these three alternative models and then we move on to present and interpret the model estimates.

Model Selection and Performance

Model parameters are estimated conditional on an assumed number of segments. To specify the model we systematically vary m, the number of segments. Following the literature on mixture models, we use Bayesian information criterion (BIC) for selecting the number of segments. On the basis of the results in Table 12 we choose a 4 segment representation of market heterogeneity.

Table 12 Model Performance for Varying Number of Segment

M	BIC
2	276591.27
3	267712.72
4	267019.78
5	267322.16

Next, we compare the performance of the three alternative models. Since all the models have the same number of parameters and observations, likelihood, BIC, and AIC are equivalent for model comparison. Table 13 shows that the proposed model, using LFP map, outperforms the benchmark models using feature and market share maps (BIC: 267019 vs. 273059 and 271349 respectively). This empirical evidence suggests that the LFP map is superior to feature and market share maps in capturing the competitive landscape underlying the data generating process. Next, we present the parameter estimates of our proposed model and discuss their implications.

Table 13 Comparison of Model Performances

	Perceptual Map from		
Criteria	**LFP**	**Features**	**Market Share**
-2 Log Likelihood	265222.67	271262.57	269552.45
AIC	265556.67	271596.57	269886.45
BIC	267019.78	273059.67	271349.55
R Squared	.69	.46	.60
Automobile		**R Squared**	
Chevrolet Impala	.18	.02	-.31
Chevrolet Malibu	.34	.31	-.10
Chrysler 300	.68	.04	.64
Ford Focus	.37	.04	.68
Honda Accord	.80	.78	.47
Honda Civic	.74	.77	.80
Hyundai Sonata	.12	.06	.76
Kia Optima	.84	.72	.79
Mazda 3	-.15	-.10	.05
Mazda 6	.89	.29	.90
Mitsubishi Lancer	.97	.95	.44
Nissan Altima	.79	.66	.78
Nissan Maxima	.79	.85	.44
Nissan Sentra	.59	.64	.81
Subaru Forester	.29	.26	.48
Subaru Impreza	.83	.71	-.17
Subaru Legacy	.96	.93	.83
Toyota Avalon	.94	.95	.94
Toyota Camry	.95	-.37	.35
Toyota Corolla	.83	.61	.87
Toyota Prius	.73	.43	.33
Volkswagen Jetta	.11	.24	.82
Volkswagen Passat	.91	.79	.95

Parameter Estimates

The estimated model has four sets of parameters (1) s_{di}'s are the ideal points of consumer segments on the perceptual map, (2) δ_i and ρ_i capture the importance of preference match relative to marketing activities for each segment, (3) β which we expect to be positive is a vector with two elements capturing the average responsiveness of the whole market to advertisements and incentive expenditures, and (4) λ_{iq}'s are the estimated sizes of market segments in each DMA. Table 14 summarizes the parameter estimates for the first three sets of parameter and Table 15 shows the estimated segment sizes in each DMA. To better interpret these results, first, we provide an intuitive map of market structure summarizing the product positions and segment ideal points. Then, we investigate the distribution of market segments across DMAs. Finally, we present the elasticity and cross elasticity patterns.

Competitive Structure Maps

We use a graphical representation of competitive market structure which gives us an intuitive way to interpreter the parameters. We overlay the estimates of s, the segment ideal points, on the positions of the products on the perceptual maps (Figure 6, Figure 8, and Figure 10). The resulting maps (competitive structure map from now on) demonstrate a complete picture showing how the products are positioned against each other and where the demand is concentrated with respect to the products. The radius of the disks is proportional to segment size in the whole market. We calculated the segment sizes for the whole market by adding up the estimated number of customers in each segment across all DMAs.

Table 14 Parameter Estimates

(in each cell, the numbers on top, middle, and bottom are point estimate, margin of error, and standard error respectively)

	LFP Perceptual Map				**Feature Perceptual Map**				**Market Share Perceptual Map**			
Parameter	**i = 1**	**i = 2**	**i = 3**	**i = 4**	**i = 1**	**i = 2**	**i = 3**	**i = 4**	**i = 1**	**i = 2**	**i = 3**	**i = 4**
	1.191	-1.013	1.090	-.032	2.192	-2.115	-1.108	.475	1.135	-1.027	.613	-.619
s_{i1}	±.009	±.009	±.012	±.000	±.102	±.265	±.014	±.018	±.045	±.052	±.003	±.008
	(.005)	(.005)	(.006)	(.000)	(.052)	(.135)	(.007)	(.009)	(.023)	(.027)	(.002)	(.004)
	-1.452	.081	1.234	.625	.327	-1.612	-.013	1.112	2.398	-.890	-.267	-.065
s_{i2}	±.007	±.006	±.011	±.000	±.054	±.146	±.013	±.064	±.091	±.051	±.002	±.007
	(.004)	(.003)	(.005)	(.000)	(.028)	(.075)	(.007)	(.033)	(.046)	(.026)	(.001)	(.004)
	1.000	.636	.824	.165	1.000	.416	.220	.083	1.000	.689	.024	.332
ρ_i	.000	±.193	±.261	±.600	.000	±.161	±.191	±.170	.000	±.166	±.139	±.285
	(.000)	(.098)	(.133)	(.306)	(.000)	(.082)	(.098)	(.087)	(.000)	(.085)	(.071)	(.145)
	3.338	2.359	1.868	155.5	.309	.893	3.491	1.755	.337	.300	2.127	15.878
δ_i	±.258	±.050	±.071	±.000	±.022	±.198	±.243	±.250	±.016	±.000	±4.638	±1.360
	(.132)	(.026)	(.036)	(.000)	(.011)	(.101)	(.124)	(.127)	(.008)	(.000)	(2.367)	(.694)
Parameters Common Across Segments												
	.067				.097				.076			
β_1	±.019				±.005				±.008			
	(.010)				(.002)				(.004)			
	.046				.064				.000			
β_2	±.014				±.008				±.000			
	(.007)				(.004)				(.000)			

Table 15 Segment Sizes

	LFP Perceptual Map				Feature Perceptual Map				Market Share Perceptual Map			
DMA	**i = 1**	**i = 2**	**i = 3**	**i = 4**	**i = 1**	**i = 2**	**i = 3**	**i = 4**	**i = 1**	**i = 2**	**i = 3**	**i = 4**
New York	14%	52%	28%	6%	26%	33%	29%	12%	73%	6%	14%	7%
Los Angeles	9%	42%	41%	8%	39%	20%	25%	16%	57%	5%	9%	29%
Chicago	12%	50%	34%	4%	40%	31%	18%	11%	60%	18%	10%	12%
Philadelphia	14%	48%	32%	6%	32%	35%	22%	11%	63%	12%	16%	10%
Detroit	7%	56%	37%	0%	48%	49%	2%	1%	25%	65%	5%	5%
Dallas	9%	51%	32%	9%	34%	26%	28%	13%	71%	14%	7%	8%
Boston	18%	41%	32%	9%	30%	36%	23%	11%	58%	8%	19%	15%
Houston	9%	51%	28%	13%	38%	21%	28%	12%	73%	12%	8%	7%
Washington DC	13%	44%	35%	8%	38%	27%	22%	13%	58%	9%	13%	20%
Miami	11%	47%	27%	16%	38%	15%	29%	18%	79%	4%	9%	8%
Bay Area	14%	33%	48%	5%	40%	20%	20%	20%	41%	5%	15%	39%
Atlanta	7%	53%	29%	11%	38%	21%	30%	11%	76%	11%	5%	7%
Tampa	7%	52%	30%	10%	47%	19%	23%	12%	69%	13%	6%	12%
Cleveland	12%	51%	35%	3%	38%	38%	16%	8%	56%	26%	11%	7%
Phoenix	9%	46%	39%	5%	40%	25%	21%	15%	58%	13%	9%	20%
Minneapolis	13%	48%	34%	4%	42%	32%	16%	10%	49%	24%	13%	14%
Orlando	7%	49%	33%	11%	42%	23%	23%	12%	70%	11%	6%	12%
Seattle-Tacoma	28%	29%	43%	0%	32%	49%	8%	12%	30%	10%	32%	29%
Pittsburgh	22%	47%	28%	3%	34%	49%	11%	6%	49%	21%	23%	6%
Denver	34%	32%	32%	2%	26%	61%	8%	5%	36%	10%	37%	17%
Baltimore	12%	48%	32%	8%	38%	28%	23%	12%	63%	12%	13%	13%
St. Louis	9%	55%	34%	3%	45%	28%	17%	10%	54%	30%	7%	9%
San Diego	12%	34%	45%	9%	38%	24%	20%	18%	50%	7%	13%	30%
West Palm Beach	9%	55%	27%	9%	45%	15%	27%	13%	74%	8%	8%	10%
Hartford & New Haven	23%	41%	29%	7%	25%	46%	20%	10%	57%	8%	25%	10%
Sacramnto	13%	41%	39%	7%	35%	25%	24%	16%	55%	7%	13%	25%
San Antonio	6%	45%	40%	9%	29%	39%	23%	9%	66%	18%	7%	9%

Table 15 Continued

DMA	LFP Perceptual Map				Feature Perceptual Map				Market Share Perceptual Map			
	i = 1	i = 2	i = 3	i = 4	i = 1	i = 2	i = 3	i = 4	i = 1	i = 2	i = 3	i = 4
Charlotte	7%	57%	29%	7%	41%	16%	30%	13%	76%	9%	5%	10%
Raleigh	8%	52%	35%	5%	37%	21%	27%	14%	65%	13%	8%	14%
Indianapolis	8%	58%	34%	0%	47%	23%	17%	12%	48%	31%	7%	14%
Portland	25%	32%	41%	1%	33%	45%	11%	11%	34%	9%	30%	27%
Cincinnati	9%	51%	37%	2%	42%	27%	18%	13%	52%	25%	10%	13%
Buffalo	13%	55%	27%	6%	52%	31%	11%	5%	45%	35%	12%	8%
Kansas City	8%	51%	37%	3%	45%	29%	17%	9%	52%	27%	7%	14%
Milwaukee	15%	46%	34%	4%	38%	35%	16%	11%	50%	22%	15%	13%
Columbus	10%	52%	36%	2%	38%	23%	23%	17%	59%	17%	9%	15%
Austin	10%	46%	38%	5%	36%	27%	21%	16%	55%	15%	14%	15%
Nashville	6%	52%	31%	12%	34%	29%	29%	9%	74%	15%	4%	6%
Salt Lake City	22%	41%	32%	5%	32%	44%	14%	10%	52%	10%	26%	12%
Albany	20%	44%	34%	2%	28%	38%	20%	14%	54%	12%	21%	13%
Harrisburg	15%	41%	39%	5%	34%	40%	15%	11%	48%	20%	16%	16%
Las Vegas	9%	44%	30%	17%	32%	32%	27%	10%	74%	10%	8%	8%
Oklahoma City	7%	56%	32%	5%	38%	25%	25%	12%	67%	19%	6%	9%
New Orleans	4%	58%	22%	17%	36%	22%	36%	6%	87%	10%	2%	0%
Norfolk	10%	50%	35%	6%	35%	26%	24%	14%	66%	13%	8%	13%
Greenvll	10%	51%	29%	9%	38%	23%	29%	11%	69%	9%	10%	13%
Jacksonville	7%	52%	30%	11%	38%	20%	29%	13%	75%	10%	5%	10%
Wilkes Barre	23%	39%	33%	5%	29%	53%	12%	6%	46%	19%	24%	10%
Birmingham	4%	62%	25%	8%	36%	15%	37%	12%	83%	12%	2%	3%
Providence	14%	48%	27%	11%	39%	26%	25%	11%	65%	9%	14%	12%

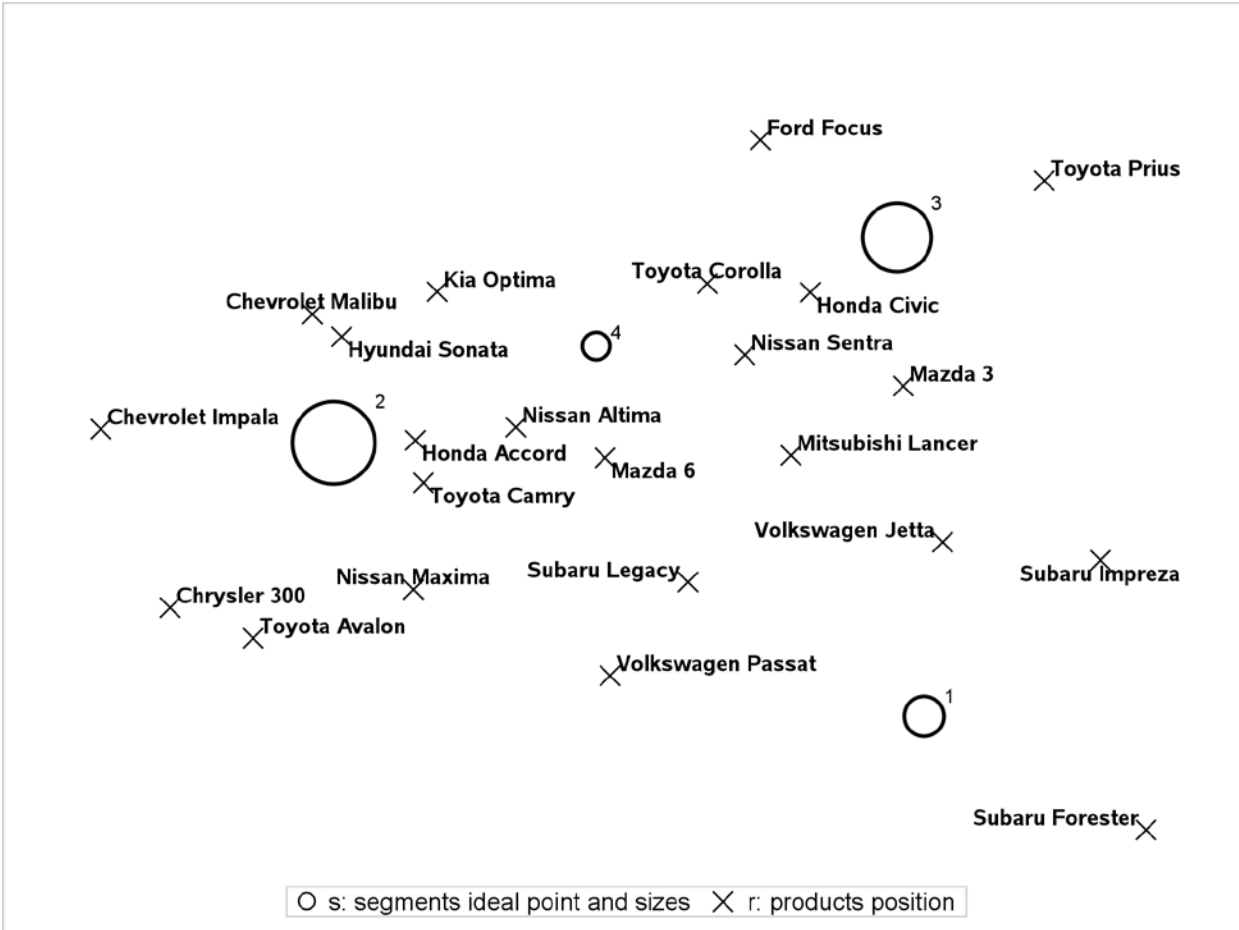


Figure 6 Segment Ideal Points and Sizes Resulting from LFP Map

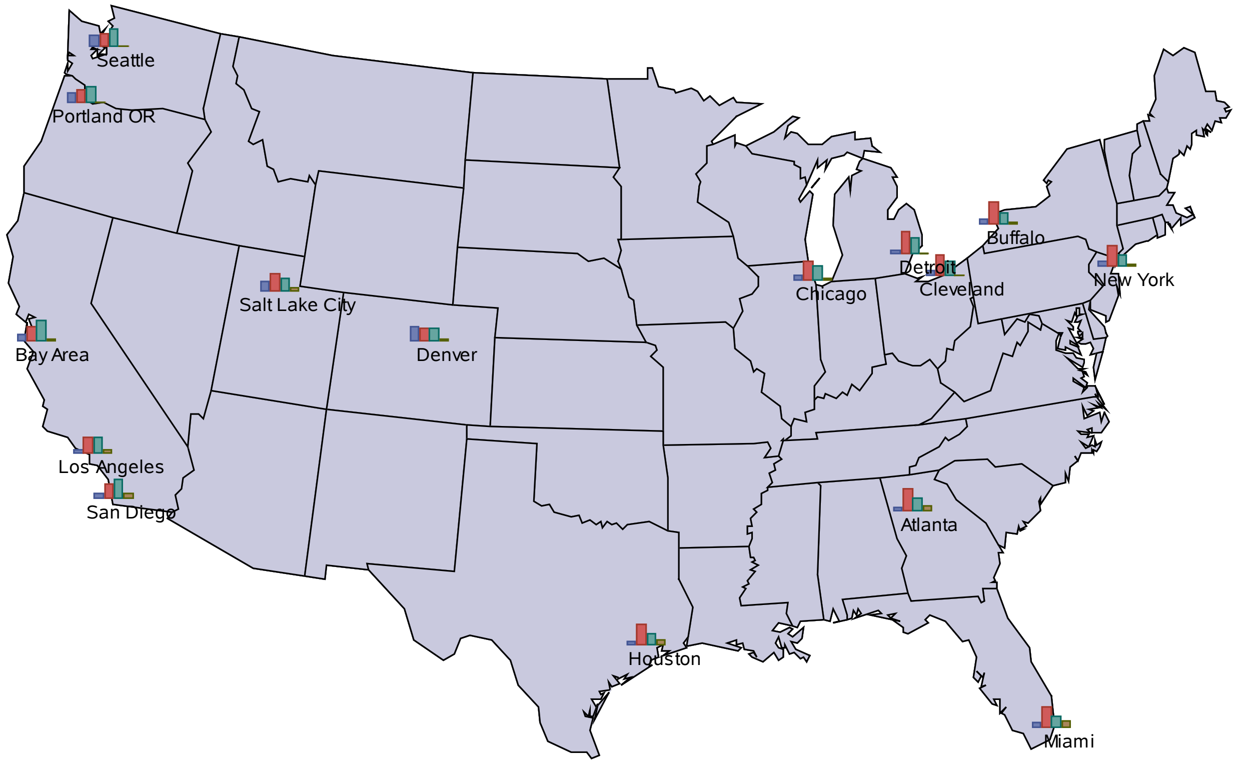


Figure 7 Segment Sizes in Select DMAs from LFP Map

(The bars from left to right represent segment 1 to 4)

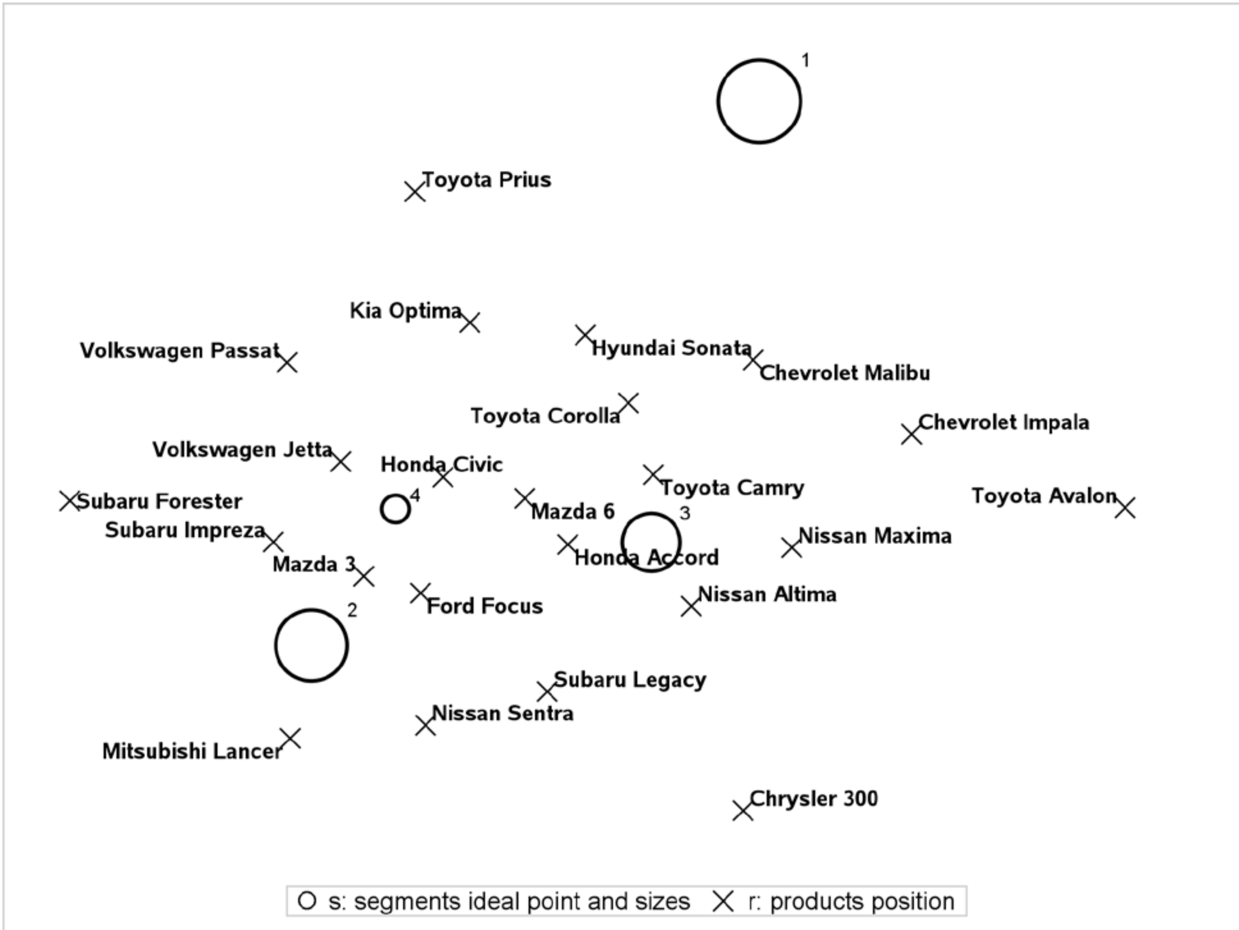


Figure 8 Segment Ideal Points and Sizes Resulting from Feature Map

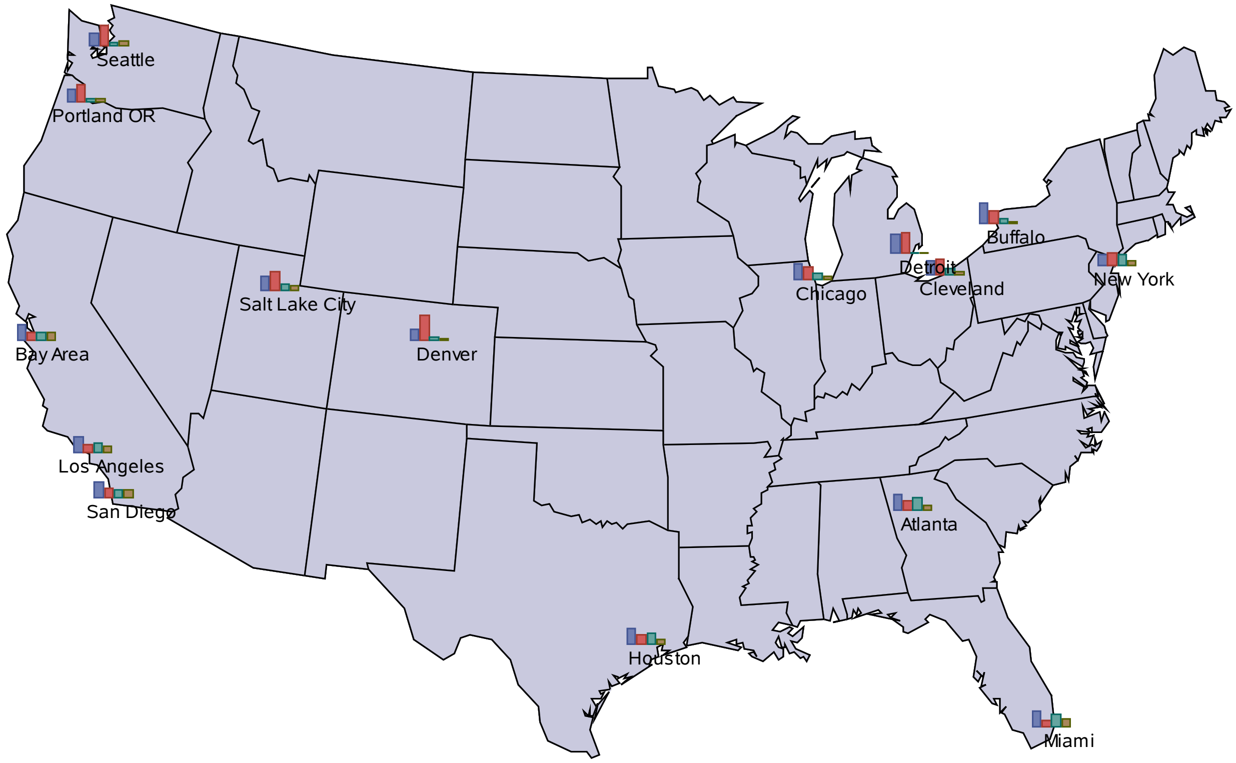


Figure 9 Segment Sizes in Select DMAs from Feature Map.

(The bars from left to right represent segment 1 to 4)

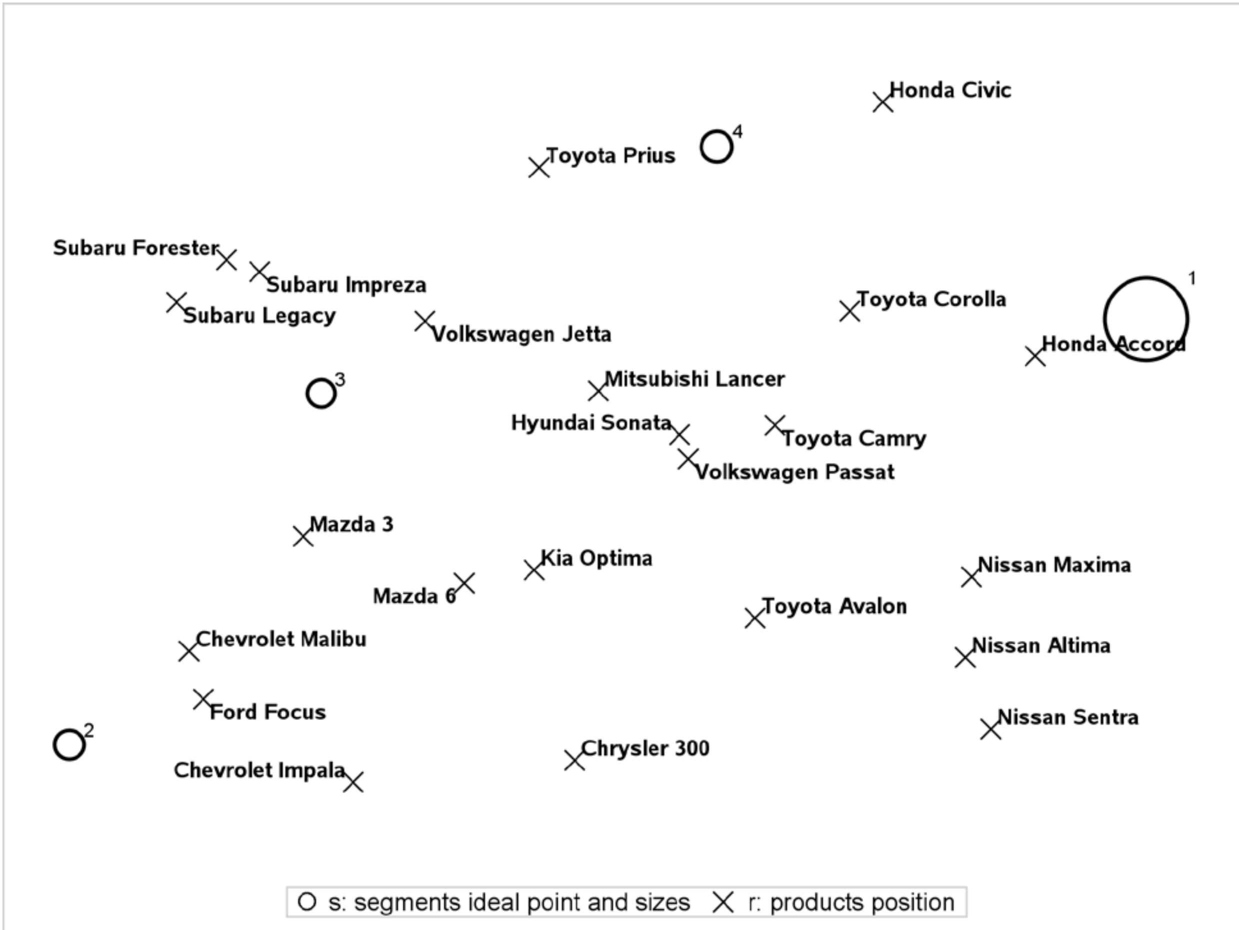


Figure 10 Segment Ideal Points and Sizes Resulting from Market Share Map

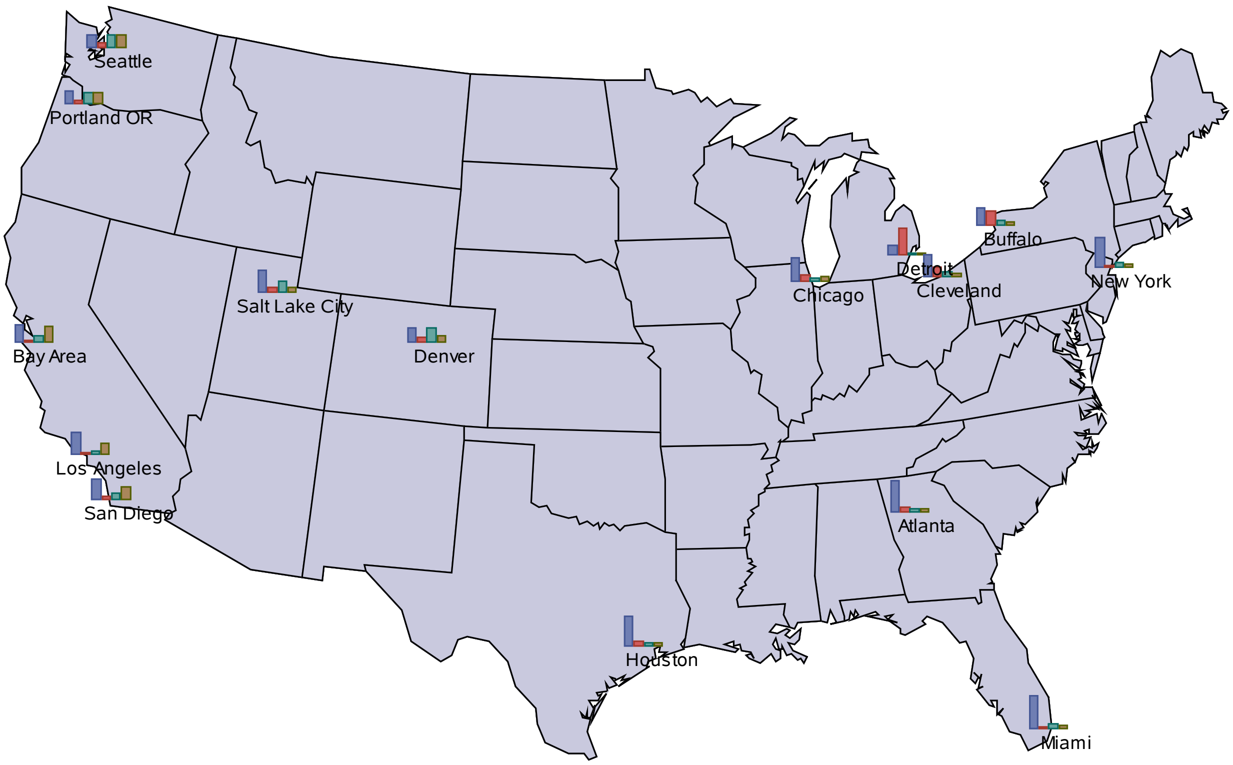


Figure 11 Segment Sizes in Select DMAs from Market Share Map.

(The bars from left to right represent segment 1 to 4)

In this figures, the centers of the disks indicate the ideal points of segments. The proximity between the segment ideal point and a product position determines the probability of choice. To calculate the probability of each segment buying each product, we aggregate the probabilities from equation 2 across DMAs and over time. The resulting choice probabilities (Table 16) give us a quantitative way to interpret the competitive structure map. As Table 15 shows the estimated segment sizes are not uniform across DMAs. To demonstrate these variations in a visual form, Figure 7, Figure 9, and Figure 11 presents a bar chart of segment sizes in select DMAs.

The model using the LFP map identifies four consumer segments as shown in Figure 6. Table 16 first 4 columns displays the probability of choice for each segment. Consumers in segment 1 are more likely to buy Volkswagen Jetta (probability of choice, P=42%), Subaru Forester (P=23%), Subaru Impreza (P=16%), Subaru Legacy (P=10%), and Volkswagen Passat (P=8%). These consumers prefer sporty and powerful automobiles with outdoor capabilities. This segment constitutes a small percentage of the market in most of the country. The exceptions to this rule are Denver, Seattle, Portland, and Salt Lake City, the mountainous DMAs where consumers need automobiles with outdoor capabilities. Segment 2, the largest segment, is close to the market leaders in this sector, Honda Accord (P=25%) and Toyota Camry (P=23%). To a lesser extent, this segment prefers other mainstream mid-size and full-size sedans such as Hyundai Sonata (P=13%), Nissan Altima (P=11%), Chevrolet Malibu (P=9%), Nissan Maxima (P=5%), Chevrolet Impala (P=5%), Kia Optima (P=4%), Mazda 6 (P=3%), Chrysler 300 (P=2%), and Toyota Avalon (P=1%). Segment 2 is the most prominent segment across the US. It is the largest segment in South East, North East, and Mid-West.

Table 16 The Choice Probability for Each Segment Estimated by the the Three Models

	LFP Perceptual Map				Feature Perceptual Map				Market Share Perceptual Map			
Automobile	**i = 1**	**i = 2**	**i = 3**	**i = 4**	**i = 1**	**i = 2**	**i = 3**	**i = 4**	**i = 1**	**i = 2**	**i = 3**	**i = 4**
Chevrolet Impala	.0%	4.7%	.0%	.0%	7.3%	.0%	.3%	.7%	.2%	14.1%	.0%	.0%
Chevrolet Malibu	.0%	9.3%	.0%	.0%	17.0%	.0%	3.3%	1.5%	.1%	35.2%	.3%	.0%
Chrysler 300	.0%	1.7%	.0%	.0%	.0%	.0%	.0%	2.7%	1.0%	1.2%	.0%	.0%
Ford Focus	.0%	.0%	13.8%	.0%	.5%	5.1%	3.8%	8.2%	.1%	33.8%	.0%	.0%
Honda Accord	.0%	25.1%	.0%	.0%	1.8%	.8%	15.9%	12.7%	18.7%	.0%	.0%	.9%
Honda Civic	.0%	.0%	31.3%	.0%	2.3%	6.4%	11.4%	6.5%	1.4%	.0%	.0%	32.1%
Hyundai Sonata	.0%	12.9%	.0%	.0%	15.4%	.2%	5.6%	1.7%	5.8%	.0%	.1%	3.6%
Kia Optima	.0%	3.7%	.1%	.0%	1.8%	.8%	3.0%	1.1%	1.7%	1.2%	.7%	.0%
Mazda 3	.0%	.0%	11.1%	.0%	.4%	11.5%	2.5%	5.8%	.4%	9.2%	18.3%	.0%
Mazda 6	.0%	2.5%	.2%	.0%	2.5%	1.2%	12.7%	8.0%	1.0%	3.0%	1.9%	.0%
Mitsubishi Lancer	.4%	.0%	2.0%	.0%	.0%	3.9%	.1%	2.3%	4.1%	.1%	1.1%	5.8%
Nissan Altima	.0%	1.7%	.1%	35.6%	1.0%	.0%	5.7%	11.6%	8.8%	.0%	.0%	.0%
Nissan Maxima	.0%	4.9%	.0%	.0%	2.3%	.0%	2.4%	4.7%	9.4%	.0%	.0%	.0%
Nissan Sentra	.0%	.1%	9.0%	6.4%	.1%	.9%	.3%	5.5%	4.3%	.0%	.0%	.0%
Subaru Forester	23.0%	.0%	.0%	.0%	.1%	23.7%	.0%	.1%	.2%	.4%	12.6%	.1%
Subaru Impreza	16.0%	.0%	.0%	.0%	.4%	19.8%	.6%	2.0%	.3%	.4%	21.2%	.1%
Subaru Legacy	1.0%	.1%	.0%	.0%	.2%	.2%	1.3%	9.1%	.1%	1.0%	13.0%	.0%
Toyota Avalon	.0%	1.3%	.0%	.0%	1.5%	.0%	.0%	.0%	3.5%	.0%	.0%	.0%
Toyota Camry	.0%	22.9%	.0%	.0%	5.1%	.2%	18.1%	8.7%	11.3%	.0%	.0%	4.1%
Toyota Corolla	.0%	.0%	14.0%	58.0%	9.5%	.1%	9.8%	3.6%	1.8%	.0%	.0%	17.5%
Toyota Prius	.0%	.0%	18.0%	.0%	17.6%	.1%	.1%	.1%	2.1%	.0%	.1%	31.2%
Volkswagen Jetta	42.5%	.0%	.2%	.0%	1.5%	15.8%	2.9%	2.8%	1.3%	.3%	3.6%	2.3%
Volkswagen Passat	8.1%	.0%	.0%	.0%	2.6%	9.1%	.5%	.5%	4.7%	.0%	.0%	2.2%

The consumers in segment 3 have a strong preference for Honda Civic (P=31%), Toyota Prius (P=18%), Toyota Corolla (P=14%), and Ford Focus (P=13%), Mazda 3 (P=11%), and Nissan Sentra (P=9%). This segment captures the consumers who are interested in compact and fuel efficient sedans. As it is evident in Figure 7, this segment is the largest segment in West Coast where consumers are more environmentally conscious and strong regulations for emissions are in place. In the rest of the country, this segment is the second largest. The smallest segment of customers, segment 4, is located between compact and mid-size sedans. This segment primarily switches between Toyota Corolla (P=58%), Nissan Altima (P=36%), and Nissan Sentra (P=6%). In summary, we have four segments with distinct tastes suggesting that the model captures the market heterogeneity well. This representation of heterogeneity in the market is in line with our expectation that popular automobiles are a close match for the preferences of large consumer segments. Moreover, the preference for differentiated automobiles such as Subaru and Volkswagen is well captured.

The feature map (Figure 8) shares some properties of the LFP map. For example, we can observe the gradual shift from compact sedans to full-size sedans and automobiles by Volkswagen and Subaru are close to each other and occupy a corner of the map. Apart from these general characteristics, the details of the map are different from the LFP map. Therefore, using the features map in our market share model, we get a different picture of the market competitive structure which we summarize in Figure 8 and middle 4 columns of Table 16. Segment 1 is the largest segment and captures the demand for a wide variety of automobiles. It primarily captures the demand for Toyota Prius (P=18%), Chevrolet Malibu (P=17%), Hyundai Sonata (P=15%), Kia Optima (P=11%), Toyota Corolla

(P=9%), and Chevrolet Impala (P=7%). This segment is the dominant segment in most DMAs except for Denver, Salt Lake City, Portland, and Seattle (refer to Figure 9). The second segment consists of consumers who prefer Subaru Forester (P=24%), Subaru Impreza (P=20%), Volkswagen Jetta (P=16%), Mazda 3 (P=11%), Volkswagen Passat (P=9%). This segment is the largest in Denver, Salt Lake City, Portland, and Seattle. In terms of the ideal point and distribution across geographical regions, this segment is similar to segment 1 from the analysis of the LFP map. However, this segment also value Ford Focus (P=5%) and is relatively large in Detroit and Cleveland. So overall it captures a mix combination of consumers' tastes. Segment 3 is located at the center of the map and constitutes of consumers who prefer the mainstream Japanese cars such as Toyota Camry (P=18%), Honda Accord (P=16%), Mazda 6 (P=13%), Honda Civic (P=11%), and Toyota Corolla (P=10%). Segment 4 is a tiny segment with a mixed preference for compact sedans and mid-size sedans.

The market share map provides a quite different picture of the market competitive structure (Figure 10 and the 4 columns on the right in Table 16). This difference is natural since the underlying perceptual map is distinct from the two other maps. This map reflects the similarity of variations in demand rather than similarity of the automobiles. So, this map can fit the market share data well. But, we do not expect the result to have the face validity we saw in the result from LFP map. Regardless of whether to accept it as true representation of market competitive structure or not, the estimated segment positions and sizes are reasonable given the perceptual map. For example, segment 1 captures the main stream demand for automobiles in this category. This segment is by far the largest and its ideal point is in close proximity of the high market share Japanese cars

Honda Accord (P=19%), Toyota Camry (P=11%), Toyota Corolla (11%), and Honda Civic (10%). Other than this large segment we have three smaller segments. Consumers in segment 2 prefer US made cars such as Chevrolet Malibu (P=25%), Ford Focus (P=34%), and Chevrolet Impala (P=14%). Segment 3 is comparable with segment 3 in the result from LFP map. This segment constitutes of consumers who are inclined toward Volkswagen Jetta (31%), Subaru Impreza (21%), Subaru Legacy (13%), and Subaru Forester (13%). At least based on our prior belief, the preference of this consumer segment for Mazda 3 (P=18%) is not consistent with their preference for the other cars. This can be a result of spurious correlations in demands for automobiles which result in the proximity of Mazda 3 and Subaru cars on the perceptual map. Consumers in segment 4 prefer compact fuel efficient sedans such as Honda Civic (P=32%), Toyota Prius (P=31%), and Toyota Corolla (P=17%). As seen in Figure 11, segment 1 is the largest segment in most DMAs. As expected, segment 2 consisting of consumers interested in American automobiles is the largest in Detroit. This segment is relatively large in a few other DMAs close to Detroit such as Buffalo, Cleveland, and Indianapolis. The fact that almost consumers interested in all top selling automobiles are grouped as segment 1, reduce the ability of the model to capture the differentiated nature of the market. The cars in the right side of the map are different from each other, but it seems that this differentiated nature of the industry is not captured using market share map. In short, analysis of these results shows that the LFP map not only fits the data the best, but also leads to a result more reasonable than the two other alternatives.

Elasticity of Demand

At the end, we calculate the elasticity of the demand to the marketing activities and cross-elasticity of the demand to competitors' marketing activities. We follow the method used in Ataman, Mela, and Van Heerde (2008) for evaluating elasticity using numerical simulation. We estimated elasticity as average change in demand if advertising or incentives levels were 10% higher than their historical level we observe in the data. In this case, the 10% change is merely for ease of interpretation. It doesn't substantially matter if we use other percentages.

Particularly, we use the estimates from the model using LFP map and hold all marketing expenditure at historical values to calculate each vehicles predict market share. This estimate serves as the base market share ($\hat{y}_{j0}$). Then, we increase the advertisement expenditures of vehicle j' by 10% and calculate the new predicted market share ($\hat{y}_{jj'1}$) which enable us to calculate the elasticity of j to change in ad spend of j', $\left[\left(\hat{y}_{jj'1}/\hat{y}_{j0}\right) - 1\right]/10\% \equiv \Delta_{jj'}$. Table 17 displays $\Delta_{jj'}$ for all possible pairs of j and j'. In the table, j' is at the rows and j is at the columns. In other words, the table shows the elasticity of market share of automobiles at the columns to the change in ad spend of the automobiles at the rows. We follow the same procedure for incentives and present the results in Table 18. For both panels the numbers along the diagonal are elasticities and the off-diagonal numbers are cross-elasticities.

Table 17 Elasticity of the Demand with Respect to Change in Advertising Expenditure

	Mazda 6	Camry	Accord	Civic	Corolla	Altima	Prius	Focus	Sonata	Malibu	Jetta	Mazda 3	Impala	Sentra	Forester	Maxima	Impreza	Optima	Chr. 300	Avalon	Legacy	Passat	Lancer
Mazda 6	.03																						
Toyota Camry	-.03	.14	-.04			-.03			-.04	-.04			-.04			-.04		-.04	-.04	-.04			
Honda Accord	-.03	-.04	.11			-.03			-.04	-.04			-.03			-.03		-.04	-.04	-.03			
Honda Civic				.11	-.03		-.05	-.05				-.05		-.04									-.04
Toyota Corolla				-.02	.06	-.01	-.02	-.02				-.02		-.02									-.02
Nissan Altima	-.01	-.01	-.01			.07			-.01	-.01			-.01			-.01		-.01	-.01	-.01			
Toyota Prius				-.02	-.01		.08	-.02				-.02		-.01									-.01
Ford Focus				-.01	-.01		-.01	.07				-.01		-.01									-.01
Hyundai Sonata	-.01	-.01	-.01			-.01			.09	-.01			-.01			-.01		-.01	-.01	-.01			
Chevrolet Malibu	-.01	-.01	-.01			-.01			-.01	.12			-.01			-.01		-.01	-.01	-.01			
Volkswagen Jetta											.08				-.06		-.06				-.06	-.06	
Mazda 3				-.01	-.01		-.01	-.01				.10		-.01									-.01
Chevrolet Impala													.01										
Nissan Sentra														.01									
Subaru Forester															.01								
Nissan Maxima																.01							
Subaru Impreza																	.02						
Kia Optima																		.03					
Chrysler 300																			.03				
Toyota Avalon																				.01			
Subaru Legacy																					.02		
Volkswagen Passat																						.04	
Mitsubishi Lancer																							.01

Note: Empty cells indicate an estimate equal to zero. The figures show the effect of advertising expenditure of automobiles at the column on the demand of the automobiles in the rows.

Table 18 Elasticity of the Demand with Respect to Change in in Incentive Expenditure

	Mazda 6	Camry	Accord	Civic	Corolla	Altima	Prius	Focus	Sonata	Malibu	Jetta	Mazda 3	Impala	Sentra	Forester	Maxima	Impreza	Optima	Chr. 300	Avalon	Legacy	Passat	Lancer
Mazda 6	.09																						
Toyota Camry	-.03	.10	-.03			-.02			-.03	-.03			-.03			-.03		-.03	-.03	-.03			
Honda Accord	-.02	-.02	.06			-.01			-.02	-.02			-.02			-.02		-.02	-.02	-.02			
Honda Civic				.10	-.03		-.05	-.05				-.05		-.04									-.04
Toyota Corolla				-.01	.04		-.01	-.01				-.01		-.01									-.01
Nissan Altima	-.02	-.02	-.02		-.01	.11			-.02	-.02			-.02			-.02		-.02	-.02	-.02			
Toyota Prius				-.01	-.01		.04	-.01				-.01		-.01									-.01
Ford Focus				-.01	-.01		-.01	.08				-.01		-.01									-.01
Hyundai Sonata	-.01	-.01	-.01			-.01			.08	-.01			-.01			-.01		-.01	-.01	-.01			
Chevrolet Malibu	-.01	-.01	-.02			-.01			-.01	.14			-.02			-.02		-.01	-.02	-.02			
Volkswagen Jetta											.07				-.05		-.05				-.05	-.05	
Mazda 3				-.01	-.01		-.01	-.01				.07		-.01									-.01
Chevrolet Impala	-.01	-.01	-.01						-.01	-.01			.12			-.01		-.01	-.01	-.01			
Nissan Sentra				-.01	-.01		-.01	-.01				-.01		.07									-.01
Subaru Forester											-.01				.04		-.01				-.01	-.01	
Nissan Maxima	-.01	-.01	-.01			-.01			-.01	-.01			-.01			.14		-.01	-.01	-.01			
Subaru Impreza											-.01				-.01		.05				-.01	-.01	
Kia Optima																		.11					
Chrysler 300																			.14				
Toyota Avalon																				.04			
Subaru Legacy											-.01				-.01		-.01				.06	-.01	
Volkswagen Passat											-.01				-.01		-.01				-.01	.12	
Mitsubishi Lancer				-.01			-.01	-.01				-.01											.25

Note: Empty cells indicate an estimate equal to zero. The figures show the effect of Incentive expenditure of automobiles at the column on the demand of the automobiles in the rows.

The elasticities and cross elasticities pass all the criteria for face validity. First, on average incentives tend to have a stronger effect on demand than advertisement which is expected. Second, elasticities tend to be larger than cross-elasticities. After all the marketing activities for the focal automobile should have a larger effect on itself and a smaller effect on the competitors. Finally, as we expect, the similar automobiles tend to have stronger effect on each other thanks to the information about the relative position of the products we get from LFP data.

CONCLUSION

At its most fundamental level, one can view the interactions of the products in a differentiated market as a result of the interplay between two latent elements of the market: 1. how consumers view the products in relation with each other and 2. the preference of the consumers for these products. Understanding these two elements play a central role in the analysis of marketing policies. The first element leads to the concept of product positioning which is based on the assumption that each product has a position in the minds of consumers. This position dictates the interrelationship of the products. For example, if two products occupy more or less similar positions they compete fiercely against each other whereas two products perceived to be different are relatively independent. The second element stems from the fact that not all positions are equal. Some positions are more attractive for consumers and hence elicit more demand. Moreover, consumers are not homogenous in their preference and they might prefer different positions. In this essay, we demonstrate that identification of product positions, consumers' ideal point, and heterogeneity of consumers is possible by augmenting sales data with a granular data source on consumer online activities.

We propose a novel ideal point market structure analysis method that identifies the product positions as well consumer segments and ideal points using aggregate sales data. For identification of consumer segments we usually need a history of consumers repeated purchases (e.g., Kamakura and Russell 1989). However, in our model identification of segments made possible by using aggregate data at various geographical regions. This way the data in each region over time can serve as repeated observations. We assume that the sizes of consumer segments vary across regions, but the parameters of each segment are identical in all the regions. So, by borrowing information across regions we can identify segment parameters using aggregate data. For identifying product positions we use a big data approach. We develop a method that uses the data on consumers' online activities for measuring *co-consideration*, the extent to which consumers consider two products together. This measure enables us to generate a perceptual map of product positions. The positions become the basis for the rest of analysis. We apply our method to the data on 23 top sedans sold in the United States over a period of 41 months. The data on online quote requests for automobiles (also called lower funnel prospects or LFP) provides us with the perfect means to identify the position of the automobiles in the marketplace.

Our empirical analyses have led to several interesting results. First, we apply our method of measuring consideration overlap to LFP data. We use the measure of consideration overlap to generate a perceptual map of products position. We observe that the result has face validity. For example, the automobiles with the same body size or with the same brand have high consideration overlap. Then, we apply our market structure analysis method to the data on perceptual map, market share, and marketing activities.

We found that the perceptual map provides genuine information that can help us in estimating market shares. This finding provides a strong evidence for validity of our measure of consideration overlap. Moreover, the model produces many additional results useful for marketing decision making, including the segment sizes across geographical areas, probability of choosing each alternative by each segment, and elasticities and cross-elasticities of market share to marketing activities. The cross-elasticities behave the way we theoretically expect: similar automobiles tend to have stronger effect on each other. We can have this feature thanks to the information about the relative position of the products we get from LFP data.

Despite all attractive features discussed above, our model has a few limitations that we can address in future extensions to this work. First, the current model assumes that the consumer is bound to select from the products explicitly included in the model. Accounting for outside goods can help improve the reliability of estimated segment locations. Second, the model currently only focuses on horizontal differentiation between the products. However, the products can also be vertically differentiated meaning that a product can generally be more attractive than the others. We can extend the modeling framework by allowing the possibility of vertical differentiation. Finally, the current model does not control for endogeneity of marketing decisions which we should control in future extensions to the model.

LFP is a versatile source of data, rich in information about consumer considerations and intentions before purchase. It is available at ZIP code level with exact time stamp (down to the second when the consumer requested the quote). This level of granularity makes it possible to extend the use of LFP data for measuring co-

consideration in several directions. Throughout this study we have focused on measuring co-consideration at national level for the whole period of the analysis. However, one can calculate the measure of co-consideration at regional level or at monthly or weekly level providing a dynamic picture of product positions. Another, intriguing possibility is analyzing order of consideration. With such a granular data we can investigate whether consumers tend to consider some product systematically before or after the others. If there is a systematic difference, the next question is whether there exists a position in consideration order leading to a better market outcome.

A threat to validity of our measure stems from the consumer behavior in requesting quotes. Our measure relies on counting the quotes happening within five minute time intervals. There is a possibility that consumers typically request quotes for multiple products with a large time interval in between. The situation becomes worst if those who quote for different automobiles in a short period of time are systematically different from the whole population. Since we do not have any data on individual consumer online quote requests, we cannot find a definitive proof whether observing quotes within a short time interval (for example 5 minutes intervals in this study) reflects real consumers co-consideration of two automobiles or not. All we can do is to see if the result appears to be valid and whether it can help us predict real market outcome. However, one more step to further validate the method is to do a sensitivity analysis which tests time intervals shorter or longer than 5 minutes. By comparing the results of alternative time intervals with our current result, one can find the optimal time interval which is the most consistent with the consumers' behavior.

One avenue for future research is to develop measures of co-consideration using other big data sources. One possible example is to use Google Trend data. We can use the name of two products together in one query to generate a proxy for the number of people who put to brands simultaneously in the search box. This behavior can be a strong sign that the consumers are considering and comparing the two products. An interesting question is whether there is a systematic difference between the perceptual map from Google Trend and the perceptual map from LFP. The differences can be partly associated to the fact that consumers usually use Google at earlier stages of the decision whereas they use online quote requests at final stages. One can even investigate how consideration overlap evolves as consumers go through purchase funnel.

In a more general context, this study demonstrates one of the first attempts in developing methods for using big data sources to improve market response modeling and market structure analysis. Big data sources on online consumer activities are a reflection of the real behavior of consumers, available at granular level and real time, and automatically generated which makes it cost effective and accurate. As more and more big data sources in various forms become available, the prospects of using such data become even more attractive for marketers. In future, we should see more academic studies following the same path of incorporating big data in tools supporting marketing policies and decisions.

LIST OF REFERENCES

Ailawadi, Kusum L., Donald R. Lehmann, and Scott A. Neslin (2001), "Market Response to a Major Policy Change in the Marketing Mix: Learning from Procter & Gamble's Value Pricing Strategy," *Journal of Marketing*, 65 (1), 44-61.

Ataman, M. Berk, Carl F. Mela and Harald J. van Heerde (2008), "Building Brands", *Marketing Science*, 27(6), 1036-1054.

Askitas, Nikos and Klaus Zimmermann (2009), "Google Econometrics and Unemployment Forecasting," *Applied Economics Quarterly*, 55 (2), 107-20.

Bradlow, Eric T., Ye Hu, and Teck-Hua Ho (2004), "A Learning-Based Model for Imputing Missing Levels in Partial Conjoint Profiles," *Journal of Marketing Research*, 41 (November), 369-81.

Bucklin, Randolph E., Gary J. Russell, and V. Srinivasan (1998), "A Relationship between Market Share Elasticities and Brand Switching Probabilities," *Journal of Marketing Research*, 35 (Febuary), 99-113.

Bucklin, Randolph E., S. Siddarth, and Jorge M. Silva-Risso (2008), "Distribution Intensity and New Car Choice," *Journal of Marketing Research*, 45 (August), 473-86.

Chintagunta, Pradeep K. (2001), "Endogeneity and Heterogeneity in a Probit Demand Model: Estimation Using Aggregate Data," *Marketing Science*, 20 (4), 442-56.

Choi, H. and H. Varian (2009a), "Predicting the Present with Google Trends," Google Inc, 1-23.

Choi, Hyunyoung and Hal Varian (2009b), "Predicting Initial Claims for Unemployment Benefits," Google.

Cooper, Lee G. and Masao Nakanishi (1988), Market Share Analysis: Evaluating Competitive Marketing Effectiveness. Boston: Kluwer Academic Publishers.

Cooper, Lee G., and Akihiro Inoue (1996), “Building Market Structures Preferences From Consumer,” *Journal of Marketing Research*, 33(3), 293–306.

Dillon, William R., and Narendra Mulani (1989), “LADI: A Latent Discriminant Model for Analyzing Marketing Research Data,” *Journal of Marketing Research*, 26(1), 15–29.

Ding, Min (2007), "An Incentive-Aligned Mechanism for Conjoint Analysis," *Journal of Marketing Research*, 44 (May), 214-23.

Du, Rex Yuxing and Wagner A. Kamakura (2012), "Quantitative Trendspotting," *Journal of Marketing Research*, 49 (4), 514-36.

Elrod, Terry (1991), “Internal Analysis of Market Structure : Recent Developments and Future Prospects,” *Marketing Letters*, 2(3), 253–66.

Fader, Peter S. and Bruce G. S. Hardie (1996), "Modeling Consumer Choice among Skus," *Journal of Marketing Research*, 33 (November), 442-52.

Gielens, Katrijn (2012), "New Products: The Antidote to Private Label Growth?," *Journal of Marketing Research*, 49 (June), 408-23.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant (2009), "Detecting Influenza Epidemics Using Search Engine Query Data," *Nature*, 457 (19), 1012-15.

Green, P.E. and V.R. Rao (1971), "Conjoint Measurement for Quantifying Judgmental Data," *Journal of Marketing Research*, 355-63.

Grover, Rajiv, and William R Dillon (1985), “A Probabilistic Model For Testing Hypothesized Hierarchical Market Structures,” *Marketing Science*, 4(4), 312–35.

Grover, Rajiv, and V. Srinivasan (1987), “A Simultaneous Approach to Market Segmentation and Market Structuring,” *Journal of Marketing Research*, 24(2), 139–53.

Kamakura, Wagner A., and Gary J. Russell (1989), “A Probabilistic Choice Model for Market Segmentation and Elasticity Structure,” *Journal of Marketing Research*, 26(4), 379–90.

Khan, Romana J. and Dipak C. Jain (2005), "An Empirical Analysis of Price Discrimination Mechanisms and Retailer Profitability," *Journal of Marketing Research*, 42 (November), 516-24.

Lancaster, Kelvin J. (1966), "A New Approach to Consumer Theory," *Journal of Political Economy*, 74 (April), 132-57.

MacKay, David B., Robert F. Easley, and Joseph L. Zinnes (1995), “Single Structure Ideal Point Model for Market Analysis,” *Journal of Marketing Research*, 32(4), 433–43.

Nair, Harikesh, Jean-Pierre Dubé, and Pradeep Chintagunta (2005), "Accounting for Primary and Secondary Demand Effects with Aggregate Data," *Marketing Science*, 24 (Summer), 444-60.

Neslin, Scott A. (1990), "A Market Response Model for Coupon Promotions," *Marketing Science*, 9 (2), 125-45.

Netzer, Oded, Ronen Feldman, Jacob Goldenberg, and Moshe Fresko (2012), “Mine Your Own Business: Market-Structure Surveillance Through Text Mining,” *Marketing Science*, 31(3), 521–43.

Morrison, Donald G. (1979),"Purchase Intentions and Purchase Behavior", *Journal of Marketing*, 43 (2), 65-74.

Pelat, C., C. Turbelin, A. Bar-Hen, A. Flahault, and A.J. Valleron (2009), "More Diseases Tracked by Using Google Trends," *Emerging Infectious Diseases*, 15 (8), 1327.

Pollay, Richard W., S. Siddarth, Michael Siegel, Anne Haddix, Robert K. Merritt, Gary A. Giovino, and Michael P. Eriksen (1996), "The Last Straw? Cigarette Advertising and Realized Market Shares among Youths and Adults, 1979-1993," *Journal of Marketing*, 60 (April), 1-16.

Ratchford, Brian T, Debabrata Talukdar, and Myung-Soo Lee (2007), "The Impact of the Internet on Consumers' Use of Information Sources for Automobiles," *Journal of Consumer Research*, 34 (June), 111-19.

Ratchford, Brian T, Myung-Soo Lee, and Debabrata Talukdar (2003), "The Impact of the Internet on Information Search for Automobiles," *Journal of Marketing Research*, 40 (2), 193-209.

Sriram, S., Subramanian Balachander, and Manohar U. Kalwani (2007), "Monitoring the Dynamics of Brand Equity Using Store-Level Data," *Journal of Marketing*, 71 (April), 61-78.

Urban, Glen L, Philip L Johnson, and John R Hauser (1984), “Testing Competitive Market Structures,” *Marketing Science*, 3(2), 83–112.

Varian, Hal R. and Hyunyoung Choi (2009), "Predicting the Present with Google Trends," Google Inc.

Vosen, Simon and Torsten Schmidt (2011), "Forecasting Private Consumption: Survey Based Indicators Vs. Google Trends," *Journal of Forecasting*, 30 (September), 565-78.

Wu, Lynn and Erik Brynjolfsson (2009), "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Quantities," in ICIS 2009 Proceedings.