Marketing Science Institute Working Paper Series 2015
Report No. 15-105

Using Forums and Search for Sales Prediction of High-Involvement Products

Tomer Geva, Gal Oestreicher-Singer, Niv Efron, and Yair Shimshoni

"Using Forums and Search for Sales Prediction of High-Involvement Products" © 2015 Tomer Geva, Gal Oestreicher-Singer, Niv Efron, and Yair Shimshoni; Report Summary © 2015 Marketing Science Institute

MSI working papers are distributed for the benefit of MSI corporate and academic members and the general public. Reports are not to be reproduced or published in any form or by any means, electronic or mechanical, without written permission.

Report Summary

In 2012 companies spent an estimated $840 million on social media monitoring, nearly one-third of their social media marketing budget. While using data from social media websites is considered less expensive than traditional market research, in practice it is often costly to collect and process data, especially when implementing complex content-processing procedures such as sentiment analysis. Additionally, sales forecasts based on social media data may be hampered by people's tendency to restrict the topics they publicly discuss.

Recently, search engine logs, aggregating billions of individual search engine queries, have been made public through tools such as Google Trends. Since these search data reflect the "true intentions" of consumers, they may serve as a proxy for consumers' interest in a product, which could be used to accurately predict true demand patterns.

In this report, Tomer Geva, Gal Oestreicher-Singer, Niv Efron, and Yair Shimshoni empirically study the interplay between search trend data and publicly available WOM from social media websites in the context of sales prediction. They seek to answer an important question for managers: Can search trend data outperform social media monitoring data or even simply enhance these data? The answer could improve sales predictions in terms of cost and scope.

Their study is based on three sources of data: monthly data for 23 car brands (all with average sales above 5,000 cars per month) sold in the U.S. between 2007 and 2010, Google's comprehensive index of internet discussion forums, and Google search trend data. They use two well-known algorithms to generate predictions: the least-squares linear regression algorithm and the back-propagation neural network algorithm (which inherently accounts for non-linear relationships and complex interactions between variables).

Findings

Their analysis provides first evidence that search trend data facilitate comparable or more accurate sales predictions compared with the more commonly used forum-based data. Further, the authors find that augmenting forum-based models with search trend data significantly improves predictive accuracy.

They also find that the difference in predictive accuracy between search-based models and forum-based models is considerably larger in the case of "value" brands. This difference in informativeness suggests that, compared with consumers of "premium" brands, customers who purchase value brands are less likely to display their purchase intentions in forums.

These findings have important implications for car manufactures and may be generalizable to a wide array of purchase decisions for high-involvement products. More accurate sales prediction models can drive better decision making in various domains such as marketing expenditure, competitive analysis, inventory management and supply chain optimization. Moreover, since this prediction method does not require proprietary data available only to the manufacturer, it can be used by upstream and downstream players, as well as by stock market investors.

Tomer Geva is Assistant Professor and Gal Oestreicher-Singer is Associate Professor, both at the Management of Technology and Information Systems Department, Recanati Business School, Tel-Aviv University. Niv Efron is a Senior Staff Software Engineer and Yair Shimshoni is a Senior Quantitative Analyst, both at Google Inc.

Acknowledgments
Tomer Geva and Gal Oestreicher-Singer acknowledge financial support from a Marketing Science Institute research grant and a grant from the Henry Crown Institute for Business Research. Tomer Geva acknowledges financial support by the Israel Science Foundation (grant number 1443/14) and a post-doctoral research scientist fellowship at Google Inc. (Tel Aviv Office).

Introduction

The availability, level of detail and scale of social media data have encouraged researchers as well as practitioners to explore means of using such data to explain and predict offline economic outcomes. As a result, in the fields of information systems (IS) and marketing research, a dominant stream of research has emerged that focuses on abstracting data from social media websites such as blogs or internet discussion forums as a measure for the word of mouth (WOM) a product enjoys. Clearly, being visible to all, publicly broadcasted opinions and conversations may influence other consumers; it is this far reach that first made researchers expect that such publicly broadcasted opinions might be useful in predicting economic outcomes. Indeed, the number of mentions (Liu 2006), as well as the sentiment (Chintagunta et al. 2011) expressed in such publicly available data have been shown to predict offline sales.

Social media monitoring, whereby companies obtain information by sifting through data from social media websites, has become a widespread practice in industry. In 2012 companies spent an estimated $840 million dollars on social media monitoring, nearly 1/3 of their social media marketing budget.[1] Social media monitoring is considered less expensive than traditional market research, which involves surveys or focus groups. In practice, however, it is often costly to collect and process social media data, especially when implementing more complex content processing procedures such as sentiment analysis. For example, simple monitoring and measurement tools cost a few thousand dollars per year,[2] and more elaborate tools offered by marketing companies may cost tens of thousands of dollars.[3]

Cost is not the only challenge to companies attempting to exploit online social media data to predict economic outcomes. In particular, it is not clear to what extent consumers indeed "speak their minds". Specifically, do we talk about every product we use? Do we talk about all products to the same extent or with the same level of excitement? Recent work (Lovett et al. 2014) has shown, for example, that brand characteristics influence WOM patterns. Therefore, it is possible that when consumers are aware that their opinions are visible to others, they limit the topics and volume of the opinions they express.

[1] Business Week, http://www.businessweek.com/stories/2010-10-20/wanted-social-media-sifters
[2] See, for example, the Oracle solution on http://www.oracle.com/us/solutions/social/overview/index.html
[3] See, for example, the Nielsen solution on http://www.nielsensocial.com/

An alternative source of online data may offer companies a better glimpse into customers' true purchase intentions—those they pursue when their activities are not visible to all. Search engine logs, aggregating billions of individual search engine queries, have recently been made public through tools such as Google Trends. It has been argued that search data actually reflect the "true intentions" of consumers (Wu and Brynjolfsson 2009), suggesting that these data could serve as a proxy for consumers' interest in a product, which could be used to accurately predict true demand patterns.

The two sources of data—publicly available WOM from social media websites such as internet blogs or discussion forums (hereafter referred to as "forum data"), and search data—are very different in nature. Each source has different characteristics in terms of its potential influence on others and the level of self-restraint it is likely to entail: Opinions posted publicly on social media sites may influence other consumers yet may not fully represent consumers' intentions, whereas privately-conducted search may better reflect actual interest in the product. Moreover, the two data sources can be thought of as representing different aspects of the consumer decision process. Marketing literature describes three pre-purchase stages: recognition of a problem or need; search for information; evaluation of alternatives (for a discussion see Blackwell et al. 2001).[4] Search engine query logs provide insights as to the number of consumers who are actively engaged in search, whereas publicly available forum data provide a glimpse at influential information that could affect consumers' evaluations, as well as need recognition.[5]

The interplay between these two data sources and its relationship to sales prediction raise an important question for managers that lies at the heart of the current paper: Can search trend data outperform social media monitoring data (as a potentially more accurate reflection of customers' "true intentions"), or even simply enhance these data (given that the two data sources might represent different aspects of the decision process)? A positive answer to this question could revolutionize sales predictions in terms of both cost and scope.

Notably, evidence suggests that the answer to this question might depend on the specific product context in which data are being collected. Specifically, prior literature establishes that consumers do not always engage in extensive and active search for information. In fact,

[4] The full model also includes purchase and post-purchase stages.
[5] We note that WOM generation is also a part of post-purchase activities (Dellarocas and Narayan 2006), yet we are interested in the pre-sale information that these data contain.

information-search behaviors have been shown to vary for different levels of product involvement. Product involvement refers to consumers' interest in a product and their perceptions regarding its importance (Blackwell et al. 2001); it is also related to the extent of perceived risk associated with the product (Dholakia 2001). Product involvement has received considerable attention from consumer behavior researchers, because the degree of personal involvement that a given product evokes is key in shaping the decision process that consumers undergo with regard to the purchase of that product. Simply stated, the more important the product is to a consumer, the more motivated the consumer is to search and be involved in the decision. Durable products such as cars or consumer electronics are typical examples of high-involvement products, while consumable products, such as groceries, movies and music, are typical examples of low-involvement products (Moorthy et al. 1997; Viswanathan et al. 2007). Aspects of consumer search behavior that are influenced by product involvement include the inclination to search, the extent to which search is active or passive, and the quantity of information the consumer is able to process (Laurent and Kapferer 1985; Zaichkowsky 1985).

Recently, Gu et al. (2012) observed that for a high-involvement product (digital cameras) consumers did not suffice with the WOM available on the retailer's website (Amazon.com) and actively searched for external sources of information online (consumer forums and opinion websites). The authors empirically showed that such external sources of WOM and their sentiment have a strong effect on sales of such products. This highlights two key premises of our work: consumers of high-involvement products actively search for information, which implies that they leave “footprints” in search logs; and at the same time, they are strongly influenced by the available WOM.

Therefore, in this work we will focus on the automotive industry, a classic example of a domain of high-involvement products, and we will study the interplay between the two above-mentioned sources of data—search trend data and forum data. Given that, for most consumers, the purchase of a car is a substantial financial expense and an information-intensive decision, we expect both WOM and search to be important. Additionally, the automotive industry's vast marketing budget (estimated to have exceeded $10 billion during 2008), as well as its importance

to the economy, render this industry an interesting test-bed with important practical implications.[6]

Finally, in this paper we focus on measuring the utility of the two sources of data for predicting sales. Accurate sales forecasting is a critical factor in a variety of key business processes, including inventory control, manufacturing decisions and marketing activities. In this context we raise the following research questions:

Our first research question examines whether predictive models using search trend data obtain similar or better accuracy than predictive models utilizing sentiment and volume of forum data when predicting sales of high-involvement products.

Our second research question examines whether predictive models based on a combination of search trend data and forum data are superior to those based on forum data alone. If the incorporation of search trend data into a model based on forum data improves the model's predictions, this could yield more accurate sales prediction models. The risk of adding a second source of data is that it can potentially lead to over-fitting and may ultimately harm out-of-sample predictive accuracy. Superiority of a model combining the two data sources would show that the two data source contain non-overlapping information and would indicate that they capture different aspects of the consumer decision process.

Our third research question explores the differences in informativeness across different types of brands, specifically, "premium" versus "value" brands. Even within a category containing solely high-involvement products—automobiles—diverse characteristics of different brands may influence the predictive power of search and forum data. For instance, some brands—e.g., brands that are more fashionable or luxurious (Porsche)—might be more frequently mentioned in forums compared with other brands (see also Lovett et al. 2014)—e.g., more common and less fashionable brands (Chevrolet). Such differences may affect the informativeness of forum data for prediction. The automotive industry also provides us with the context to answer our third research question, as it includes both "premium" and "value" brands.

[6] http://blog.nielsen.com/nielsenwire/consumer/ad-spending-in-u-s-down-11-5-percent-in-first-three-quarters-of-2009/

The modeling methodology in this study is predictive, rather than explanatory (Shmueli and Koppius 2011). The substantial differences between predictive and explanatory methodologies, their use cases and the justifications for using each approach are thoroughly detailed by Shmueli (2010) and Shmueli and Koppius (2011). Specifically, Shmueli and Koppius (2011) state that predictive methodology is particularly useful for "assessment of the predictability of empirical phenomena". Indeed, assessment of inherent predictability of sales given different online data sources is at the core of this study's research questions. Once the predictive capabilities of the data have been assessed, the outcomes can motivate new theory generation by subsequent studies.

In this work we use Google search trend data, and a unique source of data—Google's comprehensive index of internet discussion forums, which is, to the best of our knowledge, the most comprehensive forum data set that has been made available for any academic research. We also collected data about unit sales of new cars in the US between 2007 and 2010, as well as additional benchmark data (e.g., gasoline prices; consumer sentiment index; see below for further details).

We find that predictive models based on search trend data can outperform forum-data-based predictive models. This suggests that in the context of high-involvement products predictive models based on publicly available search trend data could be used —reducing the need to collect and process the more costly forum data. Furthermore, forecasting models that incorporate both forum data and search trend data provide significantly more accurate sales predictions compared with models using forum-based data alone. Interestingly, models that incorporate both forum data and search trend data do not consistently outperform models using search trends alone.

We also show that search trend-based models outperform the forum-data-based models for both "value" and "premium" brands (where "value" brands are associated with lower prices and with lower perceived quality, and "premium" brands are associated with higher prices and perceived quality). Nevertheless, this difference is considerably larger for "value" car brands, while for "premium" car brands the improvement obtained by using search trend data is more moderate.

Last, we utilize two different modeling approaches and show that predictive accuracy is strongly affected by the selection of the prediction model, with non-linear methods considerably outperforming linear methods.

Related Literature

In this work we draw on and add to three streams of research. The first studies the predictive power of social media mentions; the second includes the smaller but growing body of work that documents the predictive power of search trend data in predicting sales; and finally, we build on previous work that studies the relationship between consumers' decision processes with regard to certain products and their involvement with those products.

The prevalence of online platforms in which users can communicate product information to each other, such as discussion groups, forums and even product reviews on online sellers' websites, has led to an increase in publicly available WOM. This WOM is different from traditional person-to-person communication, which is often between familiar parties and limited in reach. Online public WOM has drawn much attention from both marketing and IS researchers, who have studied its effect on sales. For example, WOM in the form of posts on websites such as Yahoo! Movies has been shown to impact box office revenues (Duan et al. 2008a; Liu 2006); music blog buzz has been shown to impact music listening (Dewan and Ramaprasad 2012) and sales (Dewan and Ramaprasad 2009; Dhar and Chang 2009); book reviews published on a seller's own website were shown to impact the sales of reviewed books (Chevalier and Mayzlin 2006); and conversations on Usenet have been shown to affect TV ratings (Godes and Mayzlin 2004). Researchers have also studied the interplay between online WOM and critics' reviews (Chakravarty et al. 2010) and its usefulness for predicting movie revenues (Dellarocas et al. 2007); as well as the impact of internal and external WOM on sales (Gu et al. 2012). Other researchers evaluated the positive feedback of sales on WOM (Duan et al. 2008b) and the optimal response of firms to WOM (Chen and Xie 2008; Dellarocas 2006). In addition, several studies have evaluated various moderating factors that affect WOM's influence on sales; these factors include product and consumer characteristics (Zhu and Zhang 2010) as well as reviewer characteristics (Hu et al. 2008) and identity exposure (Forman et al. 2008). In the context of the

automotive industry, social media mentions of car brands were previously used to study the market structure and competitive landscape of the industry (Netzer et. al 2012).

In addition, the valence, or sentiment, of WOM has received increased attention. However, findings on this topic are somewhat varied. For instance, Liu (2006) and Duan et al. (2008b) have found that it is WOM volume, and not valence or user rating, that affects sales. In contrast, more recent studies such as those of Rui et al. (2012) and Chintagunta et al. (2011) report valence as an important factor in explaining sales. Rui et al. (2012) suggest that the difference between their outcomes and prior findings may have resulted from their use of an automated classifier, rather than reported user ratings, to measure valence. Chintagunta et al. (2011) attribute the difference in valence results to their improved modeling, which takes into account various complications of using a national-level data set that were not considered in previous studies.

The second stream of research focuses on the use of search engine logs for explaining and predicting a variety of economic and social events. Although search is conducted privately, tools such as Google Trends have recently made search logs publicly available at the aggregate level. Choi and Varian (2009, 2011) used this type of data to demonstrate contemporaneous predictive capabilities in various fields, including sales of motor vehicle parts, initial claims for unemployment benefits, travel, consumer confidence index, and automotive sales. Wu and Brynjolfsson (2009) utilized Google search data to predict future house sales and price indices as well as home appliance sales. Vosen and Schmidt (2011) used Google search data to predict private consumption, while Ginsberg et al. (2008) used Google search query data to build an early detection system for influenza epidemics whose methodology has been recently challenged by Lazer et al. 2014. Du and Kamakura (2012) developed a method for extracting latent dynamic factors in multiple time series and demonstrated their method by utilizing search trend data and predicting automotive sales. Seebach et al. (2011) also used Google data to predict automotive sales. Hu et al. (2014) constructed a model that uses search trend data and automotive sales data to decompose the impact of advertising into two components: generating consumer interest in information search and converting the interest into sales. Goel et al. (2010) used Yahoo!'s search engine data to predict various outcomes, including weekend box office revenues for feature films, video game sales and song ranks. The latter study points to several factors that can affect

predictions based on search data, including variability in the predictive power of search in different domains and possible difficulties in finding suitable query terms. It also discusses the need to utilize benchmark data when available. An explanation of why web search data are useful in predicting future sales is provided by Wu and Brynjolfsson (2009), who suggest that web search logs constitute “honest signals of decision-makers’ intentions”. That is, if buyers reveal their true intentions to purchase, future sales levels are expected to correspond to these intentions.

To our knowledge, the interplay between publicly available WOM and search trend logs for the purpose of sales prediction has not been previously studied. In fact, only one paper has previously contrasted those two sources of data: Luo et al. (2013) studied the relationship between social media data (using web blogs and consumer ratings), Google searches and web traffic in modeling firm equity value. However, their paper focuses on a substantially different domain, raises different research questions, and uses a different methodology. For example, Luo et al. jointly model the two data sources, thus capturing the "marginal" informativeness of each source given that the other source exists. In contrast, our paper compares search trend data and forum data as sources of predictive information, and measures the predictive accuracy of separate models, one model using search trend data and the other using forum data.

The third stream of literature discusses the effect of consumer involvement and perceived risk on search behavior patterns. Involvement levels range from low to high, and the degree of involvement associated with a product is determined by how important consumers perceive the product to be (Blackwell et al. 2001). Dholakia (2001) defines product involvement as "an internal state variable that indicates the amount of arousal, interest or drive evoked by a product class", and suggests that involvement is strongly linked to consumer risk perception. Involvement includes both enduring factors and situational factors (the level of interest evoked in a specific situation; Bloch and Richins 1983).

As indicated above, when a consumer feels more involved with a product, he or she is more likely to act with deliberation to minimize the risk and maximize the benefits gained from purchase and use. That is, the extent to which a product is "important" to a consumer influences the degree to which the consumer is motivated to be involved in the purchase decision, e.g., by seeking out information regarding the product.

Aspects of consumer search behavior that are influenced by product involvement include the volume of search conducted, the extent to which search is active or passive, and the quantity of information the consumer is able to process (Laurent and Kapferer 1985; Zaichkowsky 1985). Nevertheless, the effects of high involvement are not limited to search behavior; high involvement with a product has been shown to serve as a motivation for spreading WOM (Lovett et al. 2014) as well as for seeking out WOM and being influenced by it (Gu et al. 2012). Thus, in this work we focus on the context of high-involvement products, where we expect both publicly available WOM and search trend logs to be predictive of sales.

We integrate the three literature streams reviewed above to provide new insights as to the interplay between data from social media and from search engine logs in the context of predicting the sales of high-involvement products.

Last, we note that in addition to the three wide-ranging streams of research above, our study also draws on and contributes to the specific field of automotive sales prediction and modeling, which has received extensive coverage in different contexts in previous literature. For instance, Hymans et al. (1970) focused on the context of automotive expenditures to demonstrate the importance of including baseline information such as consumer sentiment index in modeling durable goods sales. Carlson (1978) used Seemingly Unrelated Regressions to model the demand for different sizes of automobiles. Urban et al. (1990) developed a behavioral state model for pre-launch market prediction in which customers move between different behavioral states toward making an automotive purchase decision. This model was later extended by Urban et al. (1993), who added additional behavioral factors such as categorization and elimination of alternatives. Greenspan and Cohen (1999) developed a macroeconomic model for forecasting aggregate new car sales in the US; their model considered the stock of vehicles and vehicle scrappage. Recently, Wang et al. (2011) applied a non-linear method, using an adaptive network-based fuzzy inference system. Landwehr et al. (2011) adopted a somewhat different perspective on the automotive sales prediction problem, incorporating visual car design parameters such as design prototypicality and design complexity into the prediction model. Most relevant to our context are papers that use Google Trends data in the context of automotive sales prediction (e.g., Choi and Varian 2009, 2011; Seebach et al., 2011; Du and Kamakura 2012). Those papers are discussed in detail above.

Data and Representation

This research uses monthly data for 23 car brands (all brands with average sales above 5,000 cars per month) sold in the US over the 4-year period between 2007 and 2010. We use three different sources of data, described below: sales, search, and forums. Note that following common practice (for example, Choi and Varian 2009; Du and Kamakura 2012; Seebach et al. 2011) we focus on brand-level sales rather than specific car model sales (e.g., the *BMW* car brand, rather than the *528i* car model).[7] See Appendix A for the list of brands and monthly sales. Nevertheless, for robustness, we also report on car-model-level sales.

Sales data

We utilize data on US unit sales of new cars and light trucks, obtained from the Automotive News website (www.autonews.com/). Automotive News provides sales data at a monthly level of aggregation. This is a well-known source for automotive sales information that has been used in various related studies such as Choi and Varian (2009) and Du and Kamakura (2012). In what follows, we use $Sales_{i,t}$ to denote the sales volume of brand *i* during month *t.*

Search data

We use Google search engine query logs. These are the same raw data that Google uses to display search engine query trends on the Google Trends website (http://www.google.com/trends/). Specifically, we collect the reported volume of monthly Google search queries for each of the car brands. We limit our data to searches originating from the US and to searches related to the automotive industry, by selecting the relevant category options in Google Trends. In what follows, we use $Search_{i,t}$ to denote the search volume of brand *i* during month *t.* [8]

Forum data

To represent forum data, we use Google's vast scan of the internet. To the best of our knowledge, this is the most comprehensive scan of forum data that has been made available for

[7] The motivation for this practice is twofold. First, brand-level data are much more abundant than car model data. Second, keyword identification is considerably more accurate at the brand level than at the car model level (see discussion below on keyword selection).

[8] As detailed in Choi and Varian (2009), Google Trends data are computed by a sampling method and therefore may contain some noise. To reduce the noise we use a similar procedure as in Preis et al. (2013) of averaging the value of multiple draws from Google Trends.

any academic research.[9] Specifically, we extracted data from all English-language forums indexed by Google's discussion forum search.[10] This index includes sites that are dedicated for, or include sections, in which users can publicly post opinions and reviews, as well as relate to previous content. (Examples include townhall-talk.edmunds.com, forums.motortrend.com, answers.yahoo.com, etc.)

Following recent literature on this topic, we extracted two aspects of forum data for each car brand: the number of times the brand was mentioned in forums ("forum mentions") and the overall sentiment (valence) of these forum mentions ("forum sentiment"). To represent brand *i*'s forum mentions in month *t* (denoted $forum\_mentions_{i,t}$), we used the number of new forum posts mentioning brand *i* during month *t*. To represent forum sentiment for brand *i* in month *t* (denoted $forum\_sentiment_{i,t}$), we used the ratio between the sums of "positive mentions" and "negative mentions" for brand *i* in month *t*. To label forum postings as "positive" or "negative", we used a dictionary-based sentiment analysis approach that is popular in the literature (see, for instance, Berger and Milkman 2012). Specifically, we utilized the extended positive and negative word dictionaries from the well-known Harvard IV-4 psychological dictionary[11] and summed the number of new forum posts mentioning "positive words" and posts mentioning "negative words" alongside brand *i* during month *t*. The advantages of using this dictionary approach are its generalizability and reproducibility over proprietary or "black box" types of sentiment analysis solutions.[12]

Keywords

In order to collect search data and forum data, it was necessary to specify keywords that could be used to identify searches or forum mentions associated with each brand. This section elaborates on the design decisions we made regarding keyword selection.

Let *K* denote a set of keywords and *B* denote a given brand. We use the term "accuracy" to denote the ratio between the number of search queries (or forum posts) that specify (any word in)

[9] Related studies that used data from forums or blogs utilized data either from a specific website, from a domain-specific search engine for forums, or from a simple hit count from Google web searches.
[10] Results for Google's discussion forum search are currently presented when selecting the *"more"* option under Google's search box, and subsequently selecting *"Discussions"*.
[11] http://www.wjh.harvard.edu/~inquirer/
[12] Our findings reported in later stages show that despite its simplicity, this type of sentiment representation considerably improves predictive accuracy (see Figures 3, 4).

K and that actually relate to brand *B*, and the total number of search queries (or forum posts) specifying any word in *K*. We use the term "coverage" to denote the ratio between the number of searches (forum posts) using any word in *K*, and the hypothetical, full number of searches (forum posts) referring to brand *B* (using any keyword).

In general, when selecting a set of keywords to identify a given brand, there is a tradeoff between accuracy and coverage. Clearly, inclusion of a larger number of keywords can increase coverage, but it may introduce noise and decrease accuracy. On the other hand, if we choose a limited set of terms for a given car brand and obtain high accuracy, we may not fully capture the brand's "online presence". For example, if one wishes to capture search queries pertaining to the Chevrolet car brand, one will most likely use the term "Chevrolet". Next, one can increase coverage by adding car model names such as "Malibu" (capturing additional searches for "Chevrolet Malibu") or "Spark" (capturing additional searches for "Chevrolet Spark"). However, adding search terms such as "Malibu" or "Spark" may also introduce a large number of irrelevant queries, e.g., queries relating to the city of Malibu, California. Note that there is no point in adding the more specific, two-word term "Chevrolet Malibu" (or "Chevrolet Spark"), to the set of keywords, as a search using this term is a subset of the searches using the keyword "Chevrolet".

To the best of our knowledge, the literature does not offer a methodology for optimal selection of keywords with the aim of achieving best predictive performance using both search and forum data, for different domains. Therefore, in this study we utilized brand-level keywords (e.g., "Chevrolet" for the Chevrolet brand), similarly to Seebach et al. (2011).[13] (See Appendix A for a detailed list of the keywords we used.)

While brand-level keywords can naturally provide high accuracy in capturing brand-related search queries, we also adapted our modeling procedures to mitigate coverage concerns. First, as mentioned above, we expected that brand-level keywords (e.g., "Chevrolet") would be considerably more commonplace than model-level keywords (e.g., "Spark"), for most car brands. Thus, the initial level of coverage was already expected to be relatively high.

[13] Seebach et al. (2011) reported that using brand-level keywords produced the best results in a similar scenario of automotive sales prediction using search trend data.

Second, we note that when a prediction model is constructed for each car brand, as long as the keyword coverage is sufficiently representative of the brand, to the point that the ratio between the volume of searches (or forum mentions) captured by the brand-level keyword and the hypothetical, unknown, "full volume of relevant searches" (or forum mentions) remains stable over time—there is actually no need to fully capture the hypothetical, unknown, "full coverage". Even simple models such as linear regression can overcome this problem by simply adjusting the coefficient values. As we are generally interested in predictive capability, rather than specific coefficient values, scaled coefficient values are not a concern.

Third, to control for different levels of baseline coverage across multiple brands, we converted the dependent and independent variables into per-brand, normalized variables and utilized the distance, in term of standard deviations, from the brand's mean, rather than the original values.

For the reasons mentioned above, in what follows we report on predictions at the brand level, using brand-level keywords. Nevertheless, in light of the managerial importance of predictions at the car model level, we subsequently report about car-model-level sales predictions as well. Additionally, for robustness we examined a different keyword methodology involving a combination of brand-level and car-model level keywords for predicting brand-level sales—and obtained similar findings. This analysis is reported in Appendix E.

Modeling

Setup

Our dependent variable is $Sales_{i,t}$—automotive sales for brand i in month t. To make a prediction for each brand's sales in month t, we use data that are available at month t - 1. Predictors include sales in previous months, forum data, and search trend data (as elaborated above), as well as benchmark data. Modeling was carried out on a monthly basis, beginning with one lag of historical data (month t - 1) and gradually incorporating additional lags (up to five lags of data, months: t - 1, …, t - 5)[14].

Following previous research in this domain, we utilized the following benchmark data:

[14] The importance of using lagged data in the prediction model setup was recently discussed in Goel et al. (2010) and Lazer et al (2014).

- Seasonality: sales in the same month, in the previous year (i.e., $Sales_{i,t-12}$). Usage of such data to represent seasonality is common in autoregressive models in this domain (see, for example, Choi and Varian 2009, 2011) due to the cyclic variation in customer automotive purchase patterns.
- Consumer sentiment index (see, for example, Hu et al. 2014; Hymans et al. 1970): This is a US national economic indicator based on a survey of a representative sample of the US population. It is designed to depict how consumers view their own financial situations and short/long-term economic conditions; thus, it has high relevance for consumer car purchase decisions. Consumer sentiment index is reported by the University of Michigan and Thompson Reuters.
- Gasoline prices (see, for example, Hu et al. 2014): This national average price information is collected and reported by the US Energy Information Administration and is based on the retail prices provided by a sample of approximately 800 gasoline stations. As gasoline prices influence the total cost of vehicle ownership, this economic indicator is expected to be associated with consumer car purchase decisions.

We also experimented with including stock market index information (as suggested by Hymans et al. 1970) using the well-known S&P500 index, but our analysis indicated that this information did not improve predictive accuracy, so we did not include it in our main analyses. Our detailed analysis of using different selections of benchmark data sets is reported in Appendix D.

Having collected the data, we defined a "*benchmark model*" as a model that utilizes consumer sentiment, gasoline price, seasonality ($Sales_{i,t-12}$), and previous sales data. Subsequently, in order to gauge the informativeness of forum-based data and search trend data, as well as the benefit of augmenting forum-based data with search trend data, we defined several additional models incorporating different sets of data, as follows: The "*forum-based model*" utilizes the benchmark model data in addition to forum mentions; the "*extended forum-based model*" adds forum sentiment data to the forum-based model; the "*search trends-based model"* utilizes both benchmark information and search trend data; and the "*combined search and forum-based model"* utilizes all the sets of information mentioned above. Table 1 summarizes the different sets of data utilized in each prediction model.

Table 1. Data Included in Each Model

	Benchmark Model	***Forum-Based Model***	***Extended Forum-Based Model***	***Search Trends-Based Model***	***Combined Search and Forum-Based Model***
$Sales_{i,t-1}$,…, $Sales_{i,t-n}$	√	√	√	√	√
Consumer $Sentiment_{t-1}$	√	√	√	√	√
Gasoline $price_{t-1}$	√	√	√	√	√
$Sales_{i,t-12}$	√	√	√	√	√
$Forum\_mentions_{i,t-1}$,…, $Forum\_mentions_{i,t-n}$		√	√		√
$Forum\_sentiment_{i,t-1}$,..., $Forum\_sentiment_{i,t-n}$			√		√
$Search_{i,t-1}$,..., $Search_{i,t-n}$				√	√

Forecasting algorithms

To better examine how the different types of data affect prediction accuracy, we utilized two well-known algorithms to generate predictions using the data models detailed above.

The first algorithm is the popular least-squares linear regression (LR) algorithm. This method has been used in the vast majority of related studies seeking to predict economic outcomes on the basis of either forum data or search trend data.

The second algorithm is the back-propagation neural network (NN) algorithm. This is a non-linear method that is estimated by the backprop algorithm (Werbos 1974). One of the strongest properties of the NN algorithm is that it inherently accounts for non-linear relationships and complex interactions between variables (Bishop 1995). Such a method may have the capacity to capture (potentially complex, or unexpected) interactions and relations that underlie

the real-life data generating process, without the need for the researcher to formally specify (or even be aware of) all existing relations. Thus, the NN approach can potentially "extract" more predictive power out of the data compared with linear methods, as it is not constrained by linearity and pre-specified interactions. This is highly important in the context of this study, which aims to evaluate the predictive capacity of the data. (See Appendix I for more details about the NN algorithm.)

To implement the NN algorithm we used the “nnet” package in R software. This implementation involves one layer of hidden nodes, and the minimization of a sum-of-square-errors criterion. In specifying the network architecture, one must choose the number of nodes in the hidden layer. While the literature does not offer clear rules about the optimal complexity of the network in terms of hidden nodes, it proposes general guidelines (Zhang et al. 1998); for instance, the number of hidden nodes should be proportional to the number of inputs. Following this guideline, we employed an NN model architecture in which the number of hidden nodes was equal to the number of inputs multiplied by 0.5.[15] Finally, while NNs are well-known for their ability to “learn” complex relations, in practice NN results may sometimes produce unstable predictions, overfit the data or converge to a local optimum. These problems can be even more pronounced in cases of relatively small datasets such as ours. As a safeguard against these problems, our specific implementation utilized the median prediction of an ensemble of 100 NNs, each using a different random seed.

The purpose of evaluating the performance of models relying on two different algorithms is to discern whether a given data source yields more accurate predictions than another, regardless of the type of algorithm (linear or non-linear) used, or whether a certain type of algorithm is better suited to exploit the information from a given data source. While it is possible to use many other algorithms, our choice reflects two standard linear and non-linear model types that are commonly used in econometric and machine-learning modeling. Nevertheless, for robustness we repeat the analysis using two additional non-linear modeling methods: SVM and RandomForest. See Appendix F for additional details.

[15] If the number of inputs is an odd number, we round up the number of hidden nodes.

Validation

Following common practice in predictive research, we measured the model's performance "out-of-sample", i.e., we used one set of data to train the model and another set to measure its performance. Specifically, we used two well-known validation methods: the "expanding window" and the "moving window" approaches.

The expanding window method has previously been used in related studies such as those of Choi and Varian (2009), Lazer et al. (2014) and Vosen and Schmidt (2011). According to Vosen and Schmidt (2011), the advantage of this validation method is that it effectively utilizes all data available up to the time of prediction.[16]

Implementing this method, we follow common practice and use (at least) 2/3 of our data as training set and 1/3 of our data as validation. We therefore report performance based on the entire out-of-sample validation period (months t = 25,…,36) as exhibited in Figure 1. That is, for each validation month t we measure performance while applying the model trained during the preceding months (months 1 to t – 1). We note that month t = 1 is January 2008 and month t = 25 is January 2010.[17]

Figure 1. Illustration of the Expanding Window Approach

[16] Lazer et al. (2014) have also emphasized the advantage of dynamically re-calibrating models used over a long period of time.

[17] Our data are for January 2007- December 2010. We "lose" 12 months' worth of data when accounting for seasonality.

For the moving (or rolling) window method we use 24 months of "rolling" training data.

Implementing this method, we follow common practice and use 2/3 of our data as a training set and 1/3 of our data as a validation set. We therefore report performance based on the entire out-of-sample validation period (months t = 25,…,36) as exhibited in Figure 2. That is, for each validation month t we measure performance while applying the model trained during the 24 preceding months (months t – 24 to t – 1). We note that month t = 1 is January 2008 and month t = 25 is January 2010.

Figure 2. Illustration of the Moving Window Approach

In what follows we first report and discuss our results using the expanding window method and subsequently report our results using the moving window method.

Performance measure

We used the mean absolute percentile error (MAPE) as our performance criterion. We made this choice for two main reasons: First, MAPE controls for volume differences across brands. For example, using MAPE, a 10% error in prediction for a large manufacturer is treated similarly to a 10% error in prediction for a small manufacturer. Second, MAPE is indifferent to the direction of the error (either overestimation or underestimation). This is appropriate for our approach, as we are interested in evaluating the predictive capacity of the data, i.e., the extent to which reliance on the data can reduce prediction error, regardless of a brand's sales volume or the direction of the error.

For robustness, we repeated the analysis using mean square error (MSE) criteria and reached similar findings in terms of the relative performance of search trends-based models and forum-data-based models. These results are provided in Appendix C.

Normalization

We transformed each of the variable representations into normalized values (for each brand). There are two motivations for normalization at the brand level. First, normalization controls for differences in sales volume across different brands. Second, as discussed in the “Data and Representation” section, normalization is a key component in our keyword handling methodology.

We note that while our models use normalized variables, in order to provide interpretable results, we calculated the MAPE according to the actual, “de-normalized” numbers. We further note that in deriving the results reported below, in order to avoid information leakage from the validation set data, we used normalization/de-normalization procedures in which the normalization factors (sample mean and standard deviation) were calculated solely on the basis of the specific training set data used in each "expanding window" iteration.

Results

Figure 3 displays the results obtained with LR using the different data representations (i.e., the different model types defined in Table 1). Figure 4 displays the results obtained with NN for the different models. Table 2 presents the differences in MAPE values between models utilizing different sets of data and the corresponding significance values, using a bootstrap confidence interval. Specifically, Table 2 displays results for the LR and NN algorithms with different numbers of lags, and for the best setup for each data representation in terms of algorithm and number of lags. The results reported include models with one lag of data and with two lags of data. Notably, although we tested prediction models using up to five lags of data, we found that adding data from lag 3 or higher actually degraded predictive accuracy for all the models.

Our first core finding is that models based on search trend data considerably outperform models based on forum data. In particular, search-trend-data based models outperform the more elaborate forum-data-based models, i.e., those that include sentiment scores.

Our second core finding relates to augmenting forum data with search trend data. Specifically, we find that augmenting forum data with search trend data further improves the prediction accuracy, even when using the more elaborate forum-data-based models that include sentiment scores. This finding suggests that search trend data contain additional valuable information not available in forum data. Interestingly, the advantage of the combined model over the model based on search trend data alone is smaller in magnitude.

An additional interesting finding that arises from our analysis is that the non-linear NN models considerably outperform the corresponding LR models for all data sources and lags. This suggests that complex relations may exist within such data sets, and that linear models—which are prevalent in many related studies that examine either forum data or search trend data—may be underutilizing the available information. For robustness we investigated this further, and repeated the analysis with two additional non-linear algorithms (SVM and RandomForest). Our results, which are reported in Appendix F, show that all three non-linear models outperform LR.

Figure 3. Prediction Results – LR

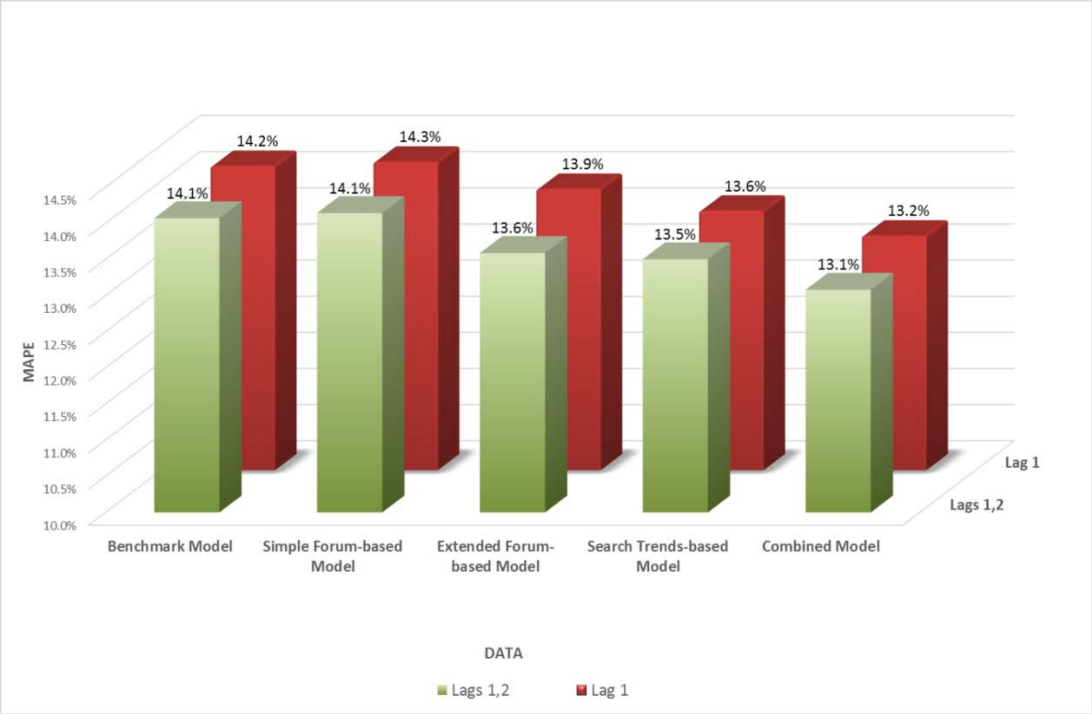


Figure 4. Prediction Results – NN

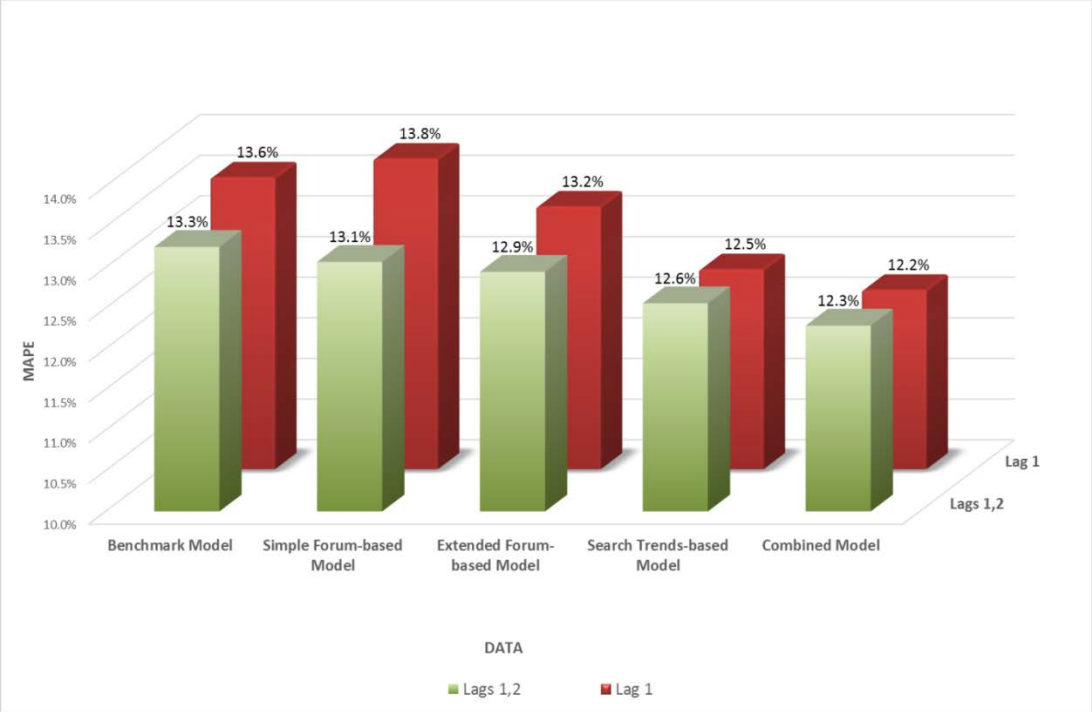


Table 2. MAPE Differences and One-Sided Confidence Intervals for the Difference in MAPE Values Using the LR, NN, and Best Setup (Forecasting Algorithm and Lags) for Each Model

Model A	**Model B**	**LR - Lag 1**	**LR - Lag 1,2**	**NN - Lag 1**	**NN - Lag 1,2**	**Best Setup**
Forum-Based Model	Benchmark Model	-0.06%	-0.07%	-0.23%	0.18%	0.18%
Extended Forum-Based Model	Benchmark Model	0.32%*	0.48%**	0.36%**	0.31%	0.31%
Search Trends-Based Model	Benchmark Model	0.62%***	0.57%***	1.13%***	0.69%***	0.79%***
Combined Model	Benchmark Model	0.96%***	0.99%***	1.38%***	0.97%***	1.05%***
Search Trends-Based Model	Forum-Based Model	0.68%***	0.64%***	1.36%***	0.51%***	0.61%***
Search Trends-Based Model	Extended Forum-Based Model	0.30%	0.09%	0.77%***	0.39%*	0.49%*
Combined Model	Forum-Based Model	1.02%***	1.06%***	1.61%***	0.78%***	0.86%***
Combined Model	Extended Forum-Based Model	0.65%***	0.51%***	1.03%***	0.66%***	0.74%***
Combined Model	Search Trends-Based Model	0.34%*	0.42%*	0.25%	0.27%	0.25%

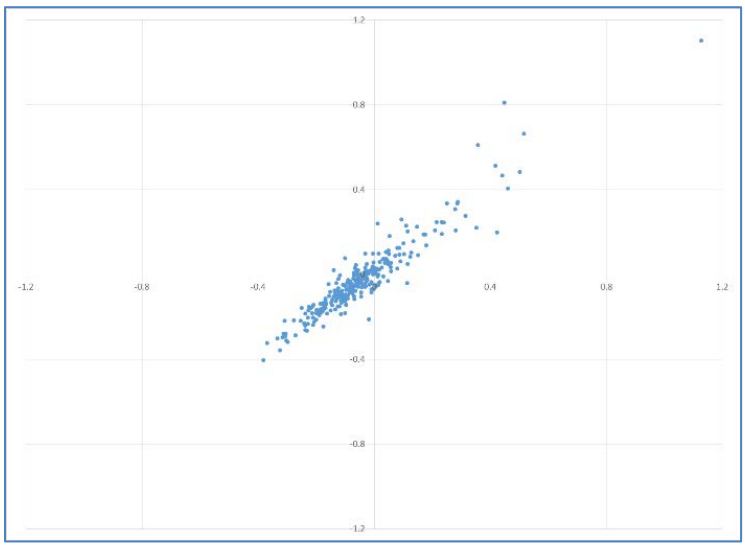
Table 2 reports the difference in MAPE using two models (*Model A* and *Model B* - each based on different data inputs) while considering 1 or 2 lags for the LR and NN algorithms, as well as when using the best setup, in terms of forecasting algorithm and number of lags for each model. Specifically, the table reports the difference:
diff = MAPE(model B) - MAPE(model A).
Lower confidence interval bounds for *diff* were calculated using 2000 iterations of the BCA bootstrapping confidence interval calculation method implemented in R software. A lower confidence interval bound for *diff*, with a positive value, provides confidence that *MAPE(Model A)* is indeed better (lower) than *MAPE(Model B)*.
We report the following lower confidence bounds:
* 0.9 lower confidence bound for *diff* is positive
** 0.95 lower confidence bound for *diff* is positive
*** 0.99 lower confidence bound for *diff* is positive

Additional analysis

A. *Direction of error*

In the section above we presented the results based on MAPE. However, as discussed, this performance measure is oblivious to the direction of error (as is MSE), which may sometimes carry practical importance. To provide insights as to the direction of the models' predictions, Figure 5 presents a scatter plot of the percentile errors for sales predictions using the best-performing search trends-based model and the best-performing extended forum-based model. Table 3 provides additional statistics regarding the models' overprediction and underprediction errors. Figure 5 suggests that both models tend to underpredict more often and that in most cases the two models err in the same direction. This observation is supported by Table 3, which shows that in 86.3% of cases the prediction errors of both models were in the same direction and that both models generally tended to under-predict (64% for the search-trends-based model and 71.8% for the extended forum-based model).

Figure 5. Percentile Errors Using Best Search and Forum Data



This figure presents a scatter plot of each prediction percentile error. The horizontal axis pertains to percentile errors for the best search trends-based model (NN, 1 lag). The vertical axis pertains to percentile errors for the best extended forum-based model (NN, 1,2 lags).

Table 3. Underprediction and Overprediction

		Extended Forum-Based Model	
		Overprediction	Underprediction
Search Trends-Based Model	Overprediction	24.3%	9.7%
	Underprediction	4.0%	62.0%

Table 3 details the percentage of cases in which each model underpredicted or overpredicted. The numbers were computed using the best search trends-based model (NN, 1 lag) and the best extended forum-based model (NN, 1,2 lags).

B. *Validation Using a Moving Window*

In addition to using “expanding window” validation methodology, for robustness we also evaluated a moving (or rolling) window approach using 24 months of "rolling" training data. Figure 6 displays the results obtained with LR using the different data representations. Figure 7 displays the results obtained with NN for the different models. Table 4 presents the differences in MAPE values (performance differences) between models utilizing different sets of data and the corresponding significance values using a bootstrap confidence interval. Specifically, Table 4 displays results for the LR and NN algorithms with different numbers of lags, and for the best setup for each data representation in terms of algorithm and number of lags. Overall, the findings obtained using the “moving window” approach are similar to those obtained using the “expanding window” approach. Note, that in the context of this work, the moving window seems to provide slightly better (although not significantly different) results.

Figure 6. Prediction Results for LR ("Moving Window")

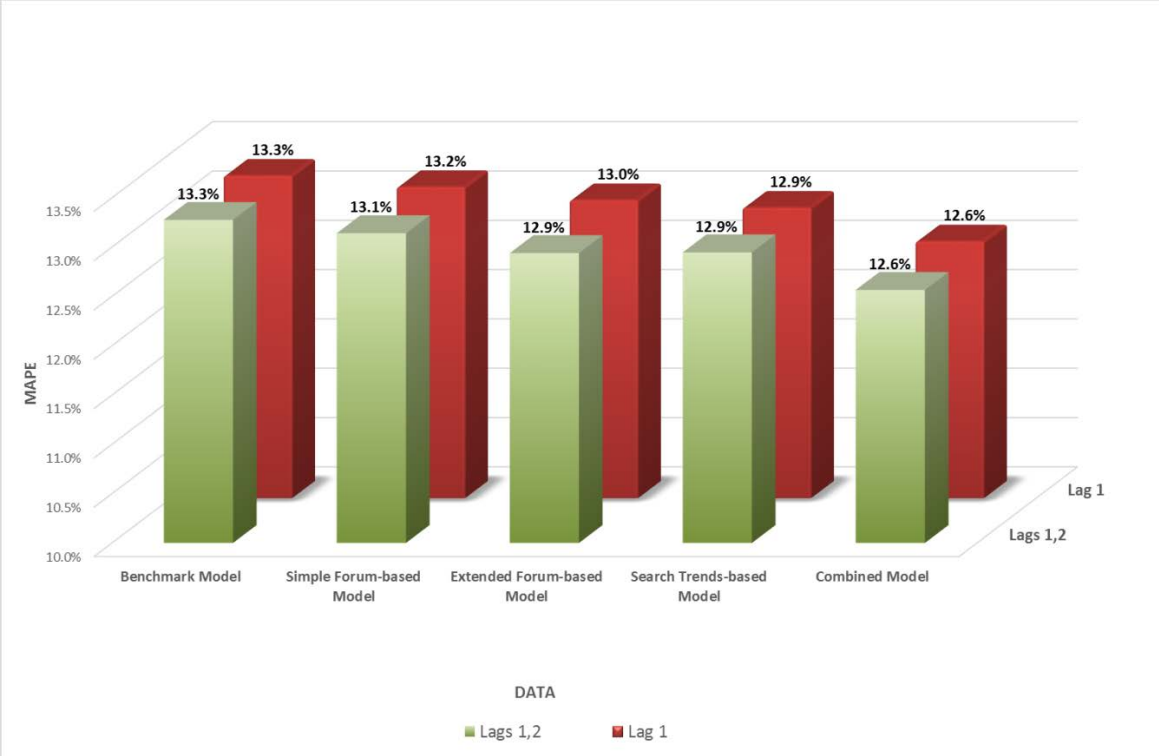


Figure 7. Prediction Results for NN ("Moving Window")

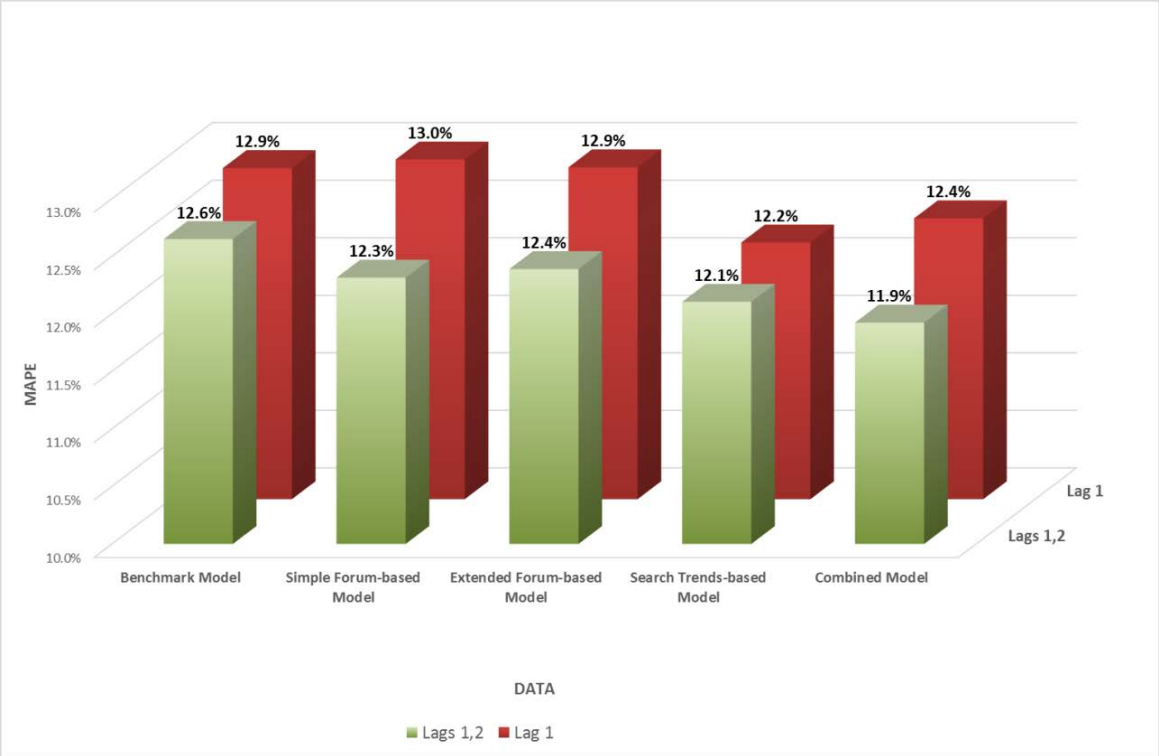


Table 4. MAPE Differences and One-Sided Confidence Intervals for the Difference in MAPE Values Using the LR, NN, and Best Setup (Forecasting Algorithm and Lags) for Each Model

Model A	Model B	LR - Lag 1	LR - Lag 1,2	NN - Lag 1	NN - Lag 1,2	Best Setup
Forum-Based Model	Benchmark Model	0.11%**	0.13%**	-0.07%	0.34%**	0.34%**
Extended Forum-Based Model	Benchmark Model	0.25%**	0.34%**	-0.01%	0.26%	0.26%
Search Trends-Based Model	Benchmark Model	0.32%***	0.31%**	0.65%***	0.55%***	0.55%***
Combined Model	Benchmark Model	0.67%***	0.71%***	0.44%**	0.73%***	0.73%***
Search Trends-Based Model	Forum-Based Model	0.21%*	0.17%	0.72%***	0.21%*	0.21%*
Search Trends-Based Model	Extended Forum-Based Model	0.08%	-0.03%	0.65%***	0.28%*	0.28%*
Combined Model	Forum-Based Model	0.55%***	0.58%***	0.51%**	0.39%***	0.39%***
Combined Model	Extended Forum-Based Model	0.42%***	0.37%***	0.44%**	0.47%***	0.47%***
Combined Model	Search Trends-Based Model	0.34%**	0.40%**	-0.21%	0.18%	0.18%

Table 4 reports the difference in MAPE using two models (Model A and Model B - each based on different data inputs) while considering 1 or 2 lags for the LR and NN algorithms, as well as when using the best setup, in terms of forecasting algorithm and number of lags for each model. Specifically, the table reports the difference:
diff = MAPE(model B) - MAPE(model A).
Lower confidence interval bounds for diff were calculated using 2000 iterations of the BCA bootstrapping confidence interval calculation method implemented in R software. A lower confidence interval bound for diff, with a positive value, provides confidence that MAPE(Model A) is indeed better (lower) than MAPE(Model B).
We report the following lower confidence bounds:
* 0.9 lower confidence bound for diff is positive
** 0.95 lower confidence bound for diff is positive
*** 0.99 lower confidence bound for diff is positive

C. *Using Less-Recent Data*

The models we have reported thus far assumed the availability of recent data (i.e., data from the previous month). However, from a managerial perspective, for reasons of data availability and collection costs, it may be important to study the predictive power of lagged search and forum data. As mentioned in the introduction, search may represent an earlier stage of the consumer decision process. Thus, one might expect that when using lagged data instead of recent data, prediction models based on search trend data might provide an even greater advantage over

forum-data-based models. We therefore repeated the analysis using different lags of both search trend data and extended forum data. Figure 8 presents the results of this analysis. As expected, for both search trend data and forum data, prediction accuracy decreases as we use more distant lags, showing a clear relationship between information speed and information quality. Interestingly, the deterioration in prediction accuracy of the search-trend-based models is more moderate than that of the forum-data-based models. Specifically, the accuracy difference (difference in MAPE) between the search-based predictions and the forum-based predictions is 0.77% when using one lag, and this difference increases to 1.52% and 1.03% when using the second and third lags, respectively. From a managerial perspective, these results demonstrate the superiority of search-trend-based prediction models over forum-data-based models for use with lagged data. From a theoretical perspective, the fact that the difference in predictive accuracy between search and forum data is greater for longer lags supports the notion that search and usage of forums represent different stages of the consumer decision process (i.e., the search and evaluation stages, respectively). This observation is in accordance with recent empirical evidence, including, most notably, the work of Luo et al. (2013), who showed that in the context of stock market predictions, social media has faster prediction value (that is, lower wear-in effect) compared with search.

Figure 8. Prediction Results Using Different Lags of Search and Extended Forum Data

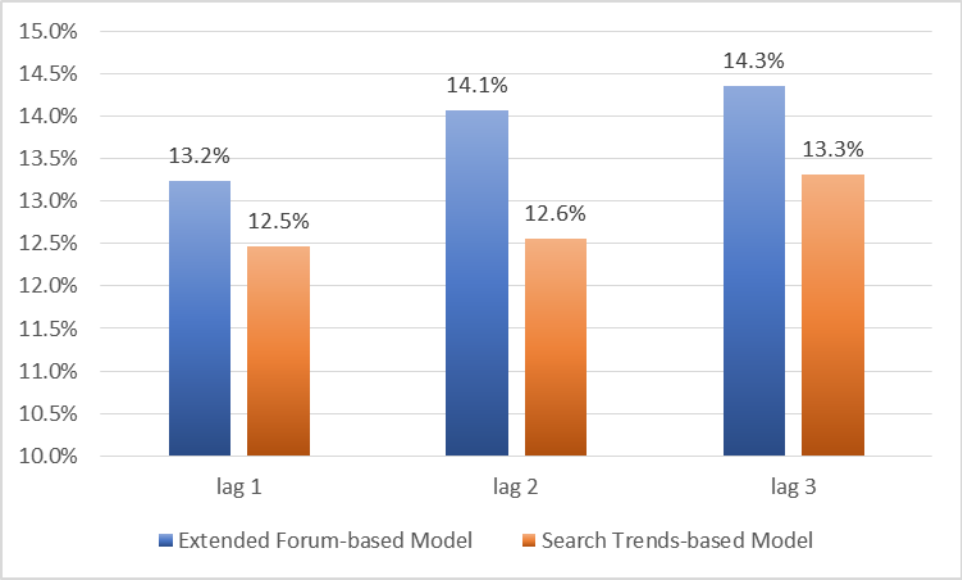


Figure 8 details MAPE for a NN model using extended forum-based data or search trend-based data from either one monthly lag, two monthly lags, or three monthly lags.

D. *Car Model-Level Analysis*

While prediction at the brand level is useful for capturing the informativeness of search trend data and forum data, there is also practical value in evaluating the predictive performance of the two data sources at the car model level. Nevertheless, prediction of car model sales is expected to be noisier than prediction of car brand sales, due to smaller sample sizes and the more volatile nature of individual car model sales.

To evaluate predictive accuracy at the car model level, we created a data set consisting of all the car models whose annual sales exceeded 10,000 units in the US, and that were sold continuously (and not replaced by a new model with the same name) during the years 2007–2010 (a total of 78 car models).

Figure 9 displays the car model-level results obtained with LR using the different data representations, and Figure 10 displays the results obtained with NN. Table 5 presents the differences in MAPE values (performance differences) between models utilizing different sets of data and the corresponding significance values using a bootstrap confidence interval. Specifically, Table 5 displays results for the LR and NN algorithms with different numbers of lags, and for the best setup for each data representation in terms of algorithm and number of lags. The results reported include models with one lag of data and with two lags of data. Notably, although we tested prediction models using up to five lags of data, we found that adding data from lag 3 or higher actually degraded predictive accuracy for all the models. (In most cases this degradation was also noticeable at lag 2.)

Figures 9 and 10 as well as Table 5 clearly show that prediction models based on search trend data or on a combination of search trend data and forum data considerably outperformed prediction models based on forum data alone. It also appears that forum-data-based models suffered from a higher level of noise at the car model level.

Figure 9. Prediction Results Car-Model Level – LR

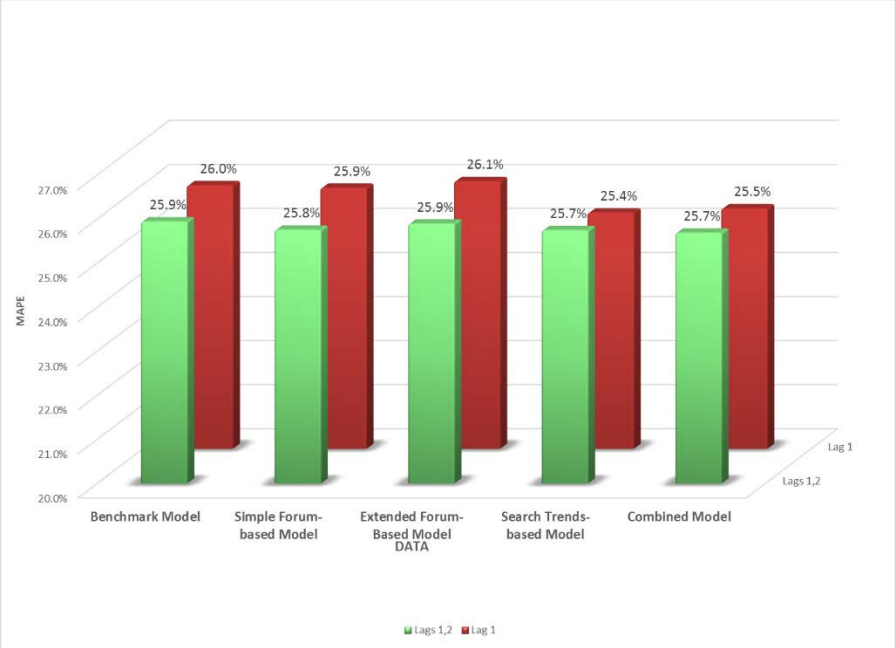


Figure 10. Prediction Results Car-Model Level – NN

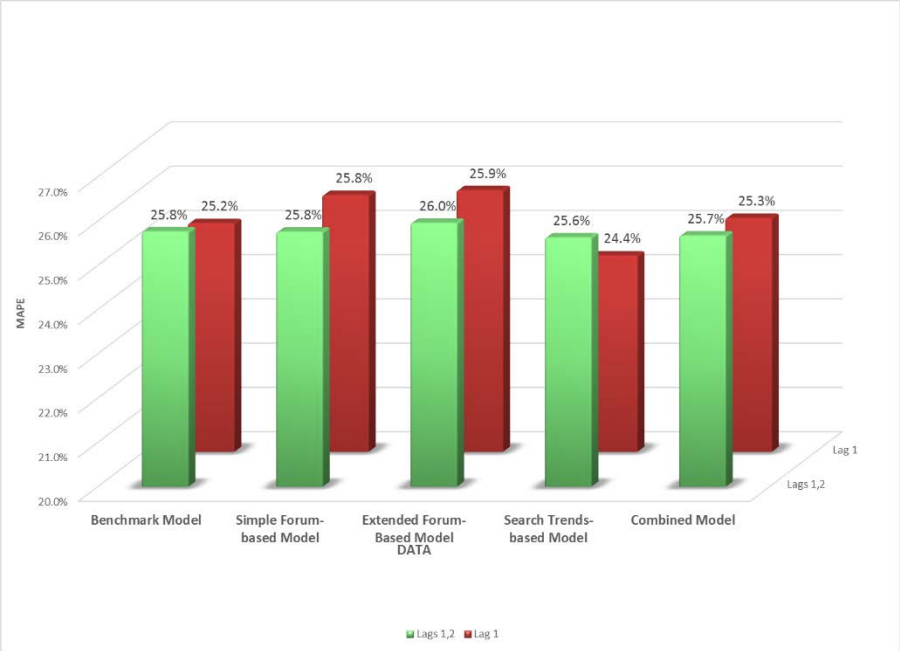


Table 5. MAPE Differences and One-Sided Confidence Intervals for the Difference in MAPE Values Using the LR, NN, and Best Setup (Forecasting Algorithm and Lags) for Each Model

Model A	**Model B**	**LR - Lag 1**	**LR - Lag 1,2**	**NN - Lag 1**	**NN - Lag 1,2**	**Best Setup**
Forum-Based Model	Benchmark Model	0.06%*	0.19%**	-0.64%	0.01%	-0.59%
Extended Forum-Based Model	Benchmark Model	-0.09%	0.05%	-0.75%	-0.19%	-0.75%
Search Trends-Based Model	Benchmark Model	0.62%***	0.20%	0.73%***	0.14%	0.73%***
Combined Model	Benchmark Model	0.53%***	0.26%	-0.12%	0.09%	-0.12%
Search Trends-Based Model	Forum-Based Model	0.56%***	0.01%	1.37%***	0.13%	1.37%***
Search Trends-Based Model	Extended Forum-Based Model	0.71%***	0.15%	1.48%***	0.33%*	1.48%***
Combined Model	Forum-Based Model	0.47%***	0.07%	0.52%***	0.09%	0.48%***
Combined Model	Extended Forum-Based Model	0.62%***	0.21%	0.63%***	0.29%*	0.63%***
Combined Model	Search Trends-Based Model	-0.10%	0.06%	-0.85%	-0.04%	-0.85%

Table 5 reports the difference in MAPE using two models (*Model A* and *Model B* - each based on different data inputs) while considering 1 or 2 lags for the LR and NN algorithms, as well as when using the best setup, in terms of forecasting algorithm and number of lags for each model. Specifically, the table reports the difference:
diff = MAPE(model B) - MAPE(model A).
Lower confidence interval bounds for *diff* were calculated using 2000 iterations of the BCA bootstrapping confidence interval calculation method implemented in R software. A lower confidence interval bound for *diff*, with a positive value, provides confidence that *MAPE(Model A)* is indeed better (lower) than *MAPE(Model B)*.
We report the following lower confidence bounds:
* 0.9 lower confidence bound for *diff* is positive
** 0.95 lower confidence bound for *diff* is positive
*** 0.99 lower confidence bound for *diff* is positive

Prediction results according to brand characteristics

To gauge additional aspects of the informativeness of the different data sources, we explored the extent to which the informativeness of forum data or of search trend data is dependent on the characteristics of the brand for which predictions are being made. Specifically, we examined whether each data source (using the best setup in terms of the number of lags and forecasting algorithm) produces more accurate predictions for "premium" (luxury) car brands (i.e., brands with higher pricing, higher perceived quality or higher willingness to recommend), or for “value” car brands.

As mentioned in the introduction of this paper, one of the fundamental differences between our two data sources—social media and search trends—is in their visibility: While search is conducted in private, social media mentions are publicly visible; in fact, they are written for others to read. Previous work in the marketing literature suggests that, because of their visibility, forum mentions may be associated with different outcomes in the cases of premium versus value brands. In effect, social media mentions have long been recognized as a modern form of WOM (Godes and Mayzlin, 2004). The rich marketing research on both online and offline WOM has provided much evidence that the motivation to engage in WOM is impacted by social drivers such as the need for self-enhancement and is used for social signaling (Sundaram et al., 1998, Henning-Thurau et al, 2004). Hence, consumers might be more inclined to converse about highly-regarded or high-quality brands (Amblee and Bui 2008), about luxury goods that signal high social status (Veblen 1994), or about brands with a high degree of differentiation, to express uniqueness (Lovett et al., 2014). In the context of forum mentions, this suggests that premium brands will be better represented in the data. Indeed, our data show that, relative to sales volume, premium brands have on average 2.1 times more forum mentions compared with value brands. We therefore expect that when predicting sales volume, forum mentions will provide better information, and hence better prediction accuracy, for premium brands than for value brands.

To gauge the differences across brands, we therefore split the car brands into two subsets based on three different factors—price, perceived quality and willingness to recommend—as follows. Car brands for which the list price for the least expensive car model of each brand was more than $20,000 were referred to as “high price”, and the rest of the car brands were referred

to as “low price”.[18] Next, we obtained survey data about perceived quality and willingness to recommend each brand. These data were obtained from YouGov plc, a market research firm that monitors a panel of 5,000 people in the US, on a daily basis, and reports on brand-related perceptions. See Appendix B for more details about the data from this survey. For this analysis, we refer to the 12 brands with lower perceived quality as brands with “low perceived quality” and the remaining 11 brands as brands with “high perceived quality”. Similarly, we refer to the 12 brands associated with lower willingness to recommend as brands with “low willingness to recommend” and the remaining 11 brands as brands with “high willingness to recommend”.[19]

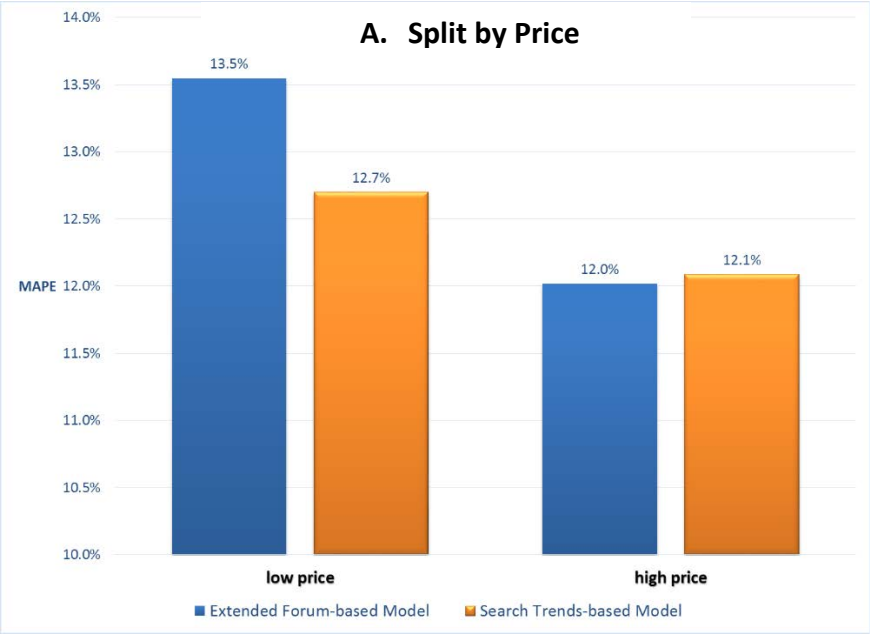
Figure 11 presents the results for the best prediction model using either search trend data or forum data, based on the three different splits: Figure 11A presents the results based on the price; Figure 11B presents the results based on perceived quality; and Figure 11C presents the results based on willingness to recommend.

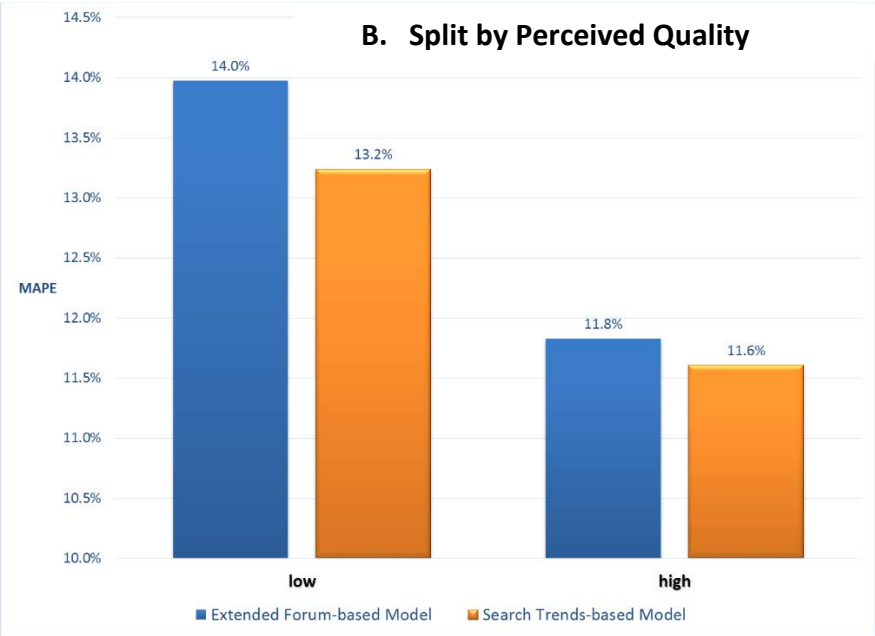
The results show that predictions for premium car brands (high price/high perceived quality/high willingness to recommend) are generally more accurate than for value brands (low price/low perceived quality/low willingness to recommend) across all prediction models.

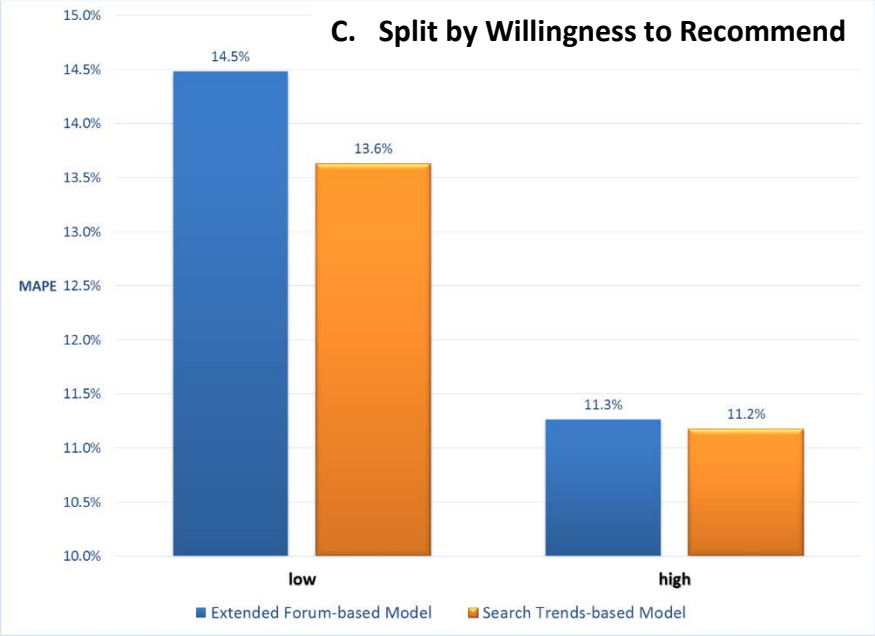
Additionally, for both premium and value brands, search trend models either outperformed the extended forum-based models or obtained a comparable level of predictive accuracy. Interestingly, for value brands, search trend data significantly outperformed forum data, whereas for premium brands differences were not significant. These results support the conjecture that for value cars, forum data may be less informative regarding true purchase intentions, compared with data on searches that are conducted in private.

[18] Prices for 2007. Source: Automotive News website.
[19] Since our data includes 23 brands it cannot be precisely split into 2 groups with equal sizes. Nevertheless result are robust to including the median brand in both groups.

Figure 11. Prediction Results for Premium vs. Value Brands



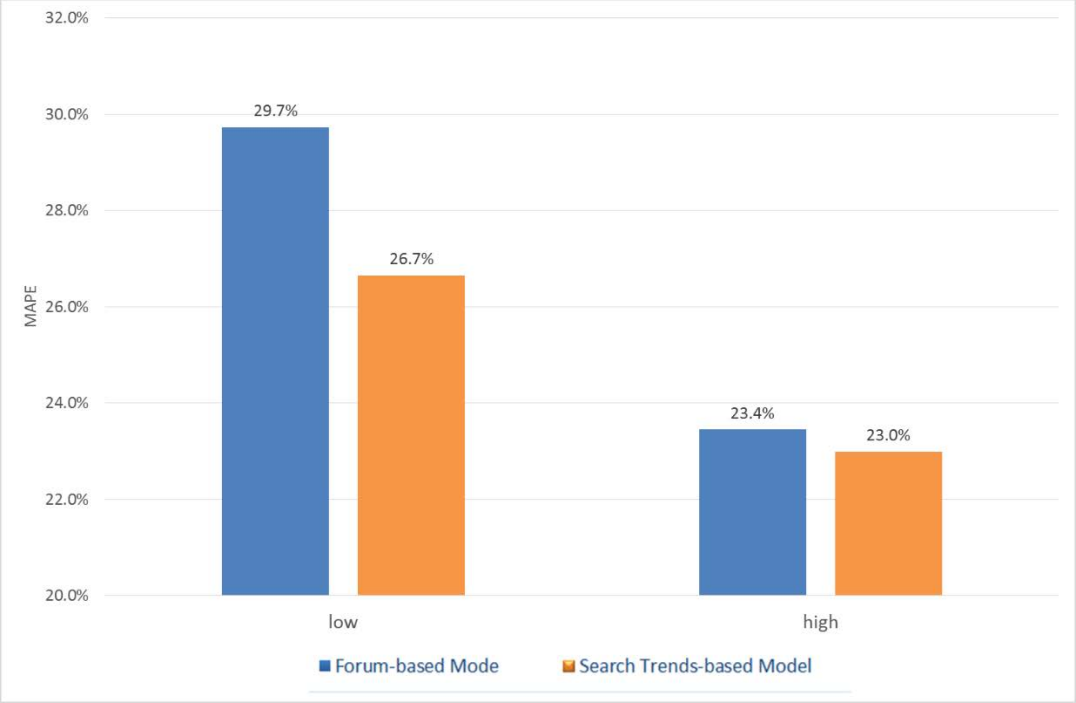




Car model level of analysis

Last, for robustness we also analyzed the respective performance of search trend data and forum data for different car characteristics at the car model level. Figure 12 presents the prediction results for "high price" and "low price" car models (defined, as above, using a price threshold of $20,000), using the forum-data-based and search trend-based models that yielded the best predictions in the "Car Model-Level Analysis" section above. Results are consistent with the brand-level results in that the difference in predictive accuracy between search trend data and forum data is considerably larger for value car models. (In Appendix H we carry out a similar analysis using a different price threshold and reach a similar conclusion.)

Figure 11. Prediction Results for High-Priced vs. Low-Priced Car Models



Conclusions

In this paper we empirically studied the interplay between search trend data and publicly available WOM from social media websites in the context of sales prediction. Marketing literature suggests that consumers considering the purchase of a product engage in a multi-stage decision process. We suggest that search and forum data may provide valuable insights about different aspects of this process—yet, thus far, previous literature on product sales prediction has studied each data source separately. In fact, the two data sources have largely been investigated in disparate literature streams and for separate purposes.

Prior literature suggests that, when attempting to obtain evidence related to customers' decision making processes, it is necessary to distinguish between different types of products. Specifically, consumers are more likely to engage in extensive and active information search in the case of high-involvement products than in the case of low-involvement products. Therefore, in this paper we focused on high-involvement products.

Using data from the automotive industry, we provide first evidence that search trend data facilitate more accurate sales predictions compared with the more commonly used forum-based data. Second, we show that augmenting forum-based models with search trend data significantly improves predictive accuracy. This evidence indicates that search data provides external and non-overlapping information, to forum data, for the purpose of sales forecasting. This finding provides reassurance that companies that have already invested in collecting forum-based data for modeling purposes have not “wasted” their money, while further suggesting that these companies can considerably improve forecasting accuracy with a relatively small additional investment in collecting search trend data.

Additionally, we find that the difference in predictive accuracy between search-based models and forum-based models is considerably larger in the case of “value” brands. This difference in informativeness suggests that, compared with consumers of “premium” brands, customers who purchase value brands are less likely to display their purchase intentions in forums, or to "speak their minds" publicly.

Finally, we find that non-linear prediction methods considerably outperform commonly used linear methods when using either source of data, suggesting that complex relations exist within these types of data.

Our work carries managerial implications for car manufactures and, more broadly, for manufacturers of high-involvement products. Moreover, the advantage of our method is that it does not require proprietary data available only to the manufacturer. Hence, it can be used by upstream and downstream players, as well as by stock market investors. Furthermore, car manufacturers can use this approach to evaluate the expected sales of their competitors. More accurate sales prediction models can, in turn, drive better decision making in various domains such as marketing expenditure, competitive analysis, inventory management and supply chain optimization.

For the specific case of automotive sales, these decisions involve the allocation of extremely large funds and, therefore, even small improvements in forecasting accuracy are expected to have a considerable effect. While we do not have a direct method to translate our results into dollar values, we calculated a conservative “ballpark” estimate for the dollar amounts associated with aggregate monthly inventory prediction errors (both overprediction and

underprediction) over a period of one year. The calculation is as follows: Approximately 15M cars are sold annually in the US alone, with sub-compact car prices close to $15,000 (the vast majority of cars sold are substantially more expensive). Therefore, at the car model level of analysis, a MAPE improvement of approximately 0.8% versus the best benchmark method will translate to a conservative estimate of $1.8 billion dollars yearly in the US alone. Given these large numbers, car manufacturers may choose to collect both forum data and search trend data in order to achieve maximum predictive accuracy; however, smaller players (for example, upstream and downstream sellers) may be able to suffice with search trend data alone.

We expect that our findings may be generalizable to a wide array of purchase decisions regarding high-involvement products, such as housing purchases and travel planning. In the case of low-involvement products, such as music, applications, and movie tickets, consumers do not conduct extensive and active search, and decisions are made more lightheartedly. The predictive power of search trends in this context is therefore unclear. This raises an interesting direction for future research.

Other possible avenues of future work include incorporating information about the popularity of different discussion forum websites in the prediction models. Additionally, it is possible to analyze the predictive value of search trend and forum data according to additional brand or model characteristics, such as vintage. It would also be interesting to study how the incorporation of predictions based on those data sources into the managerial decision making process interacts with these data and affects their predictive capability over time. Finally, from a methodological perspective, it would be interesting to compare predictions based on publicly available information to industry experts' forecasts using proprietary data.

Appendix A – List of Brands and Grouping by Price, Perceived Quality, and Willingness to Recommend

Table A1 provide details about the brands included in this study. The list of brands includes all car brands with average US sales exceeding 5,000 cars per month during 2007-2010. (Source: Automotive News). [20] Quality and Willingness to Recommend ranking is based on the YouGov BrandIndex product. (See additional details in Appendix B).

Table A1. List of Brands and Grouping by Price, Perceived Quality, and Willingness to Recommend

Brand	Keyword(s)	Price	Average Quality Ranking	Average Recommend Ranking
Acura	acura	high price	high	high
Audi	audi	high price	high	low
Bmw	bmw	high price	high	high
Buick	buick	high price	low	low
Cadillac	cadillac	high price	high	low
Chevrolet	chevrolet, chevy	low price	low	high
Chrysler	chrysler	low price	low	low
Dodge	dodge	low price	low	low
Ford	ford	low price	low	high
Gmc	gmc	low price	low	low
Honda	honda	low price	high	high
Hyundai	hyundai	low price	low	low
Infiniti	infiniti	high price	high	low
Jeep	jeep	low price	low	low
Kia	kia	low price	low	low
Lexus	lexus	high price	high	high
Lincoln	lincoln	high price	low	low
Mazda	mazda	low price	low	low
Mercedes Benz	mercedes	high price	high	high
Nissan	nissan	low price	high	high
Subaru	subaru	low price	low	high
Toyota	toyota	low price	high	high
Volkswagen	volkswagen	low price	high	high

[20] Combined search volume of multiple keywords can be obtained from Google Trends by utilizing the "+" sign between different terms.

Appendix B – YouGov Survey Data

In this work we utilized customer perception data obtained from YouGov plc. YouGov monitors a panel of 5,000 people in the US, on a daily basis, and reports on brand-related perceptions.

We used the data about "perceived quality" and “willingness to recommend” based on average ratings for the period 2008-2010.[21] Regarding product quality, panel participants were presented with the following questions on a daily basis: (a) Which of the following brands do you think represents ‘good quality’? (b.) Now which of the following do you think represents ‘poor quality’? Daily quality scores were calculated according to the following formula:

$$Perceived\ Quality = \frac{PositiveCount - NegativeCount}{PositiveCount + NegativeCount + NeutralCount}$$

Regarding willingness to recommend, panel participants were presented the following questions on a daily basis: (a.) Which of the following brands would you recommend to a friend or colleague? (b.) Which of the following brands would tell a friend or colleague to avoid? Daily quality scores were similarly calculated according to the following formula:

$$Willingness\ to\ Recommend = \frac{PositiveCount - NegativeCount}{PositiveCount + NegativeCount + NeutralCount}$$

To avoid question biases, YouGov utilizes different respondents for each question. During the relevant time period the average number of daily respondents for the quality question was 130 (s.d. 25.2). The average number of daily respondents for the willingness-to-recommend questions was 126 (s.d. 23.7).[22]

[21] For the Hyundai and Kia brands data are available only for 2010.

[22] Source: YouGov BrandIndex product documentation and data.

Appendix C – MSE results

In this appendix we provide results using Mean Squared Error (MSE) criteria rather than MAPE.

Figure C1 displays the results obtained with LR using the different data representations. Figure C2 displays the results obtained with NN for the different models. We observe similar findings in terms of the superiority of the search trend data, and combined models using search trend data and forum data over the extended forum-based model. Additionally, prediction models using NN outperform prediction models using LR.

Figure C1 - Prediction Results for LR (MSE)

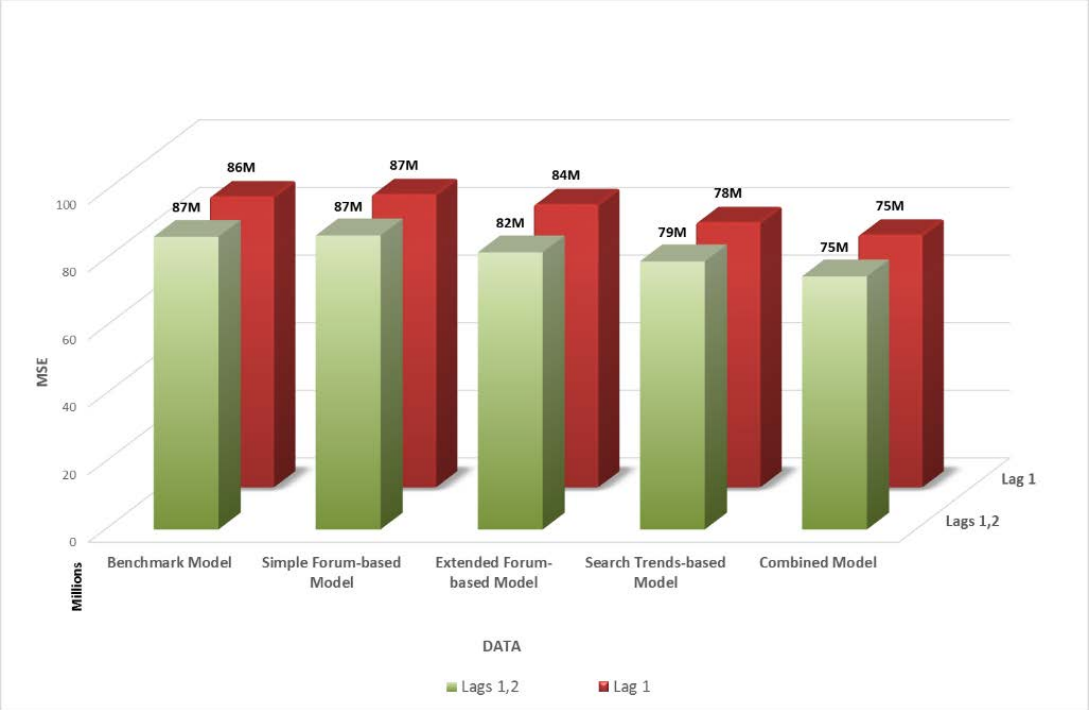
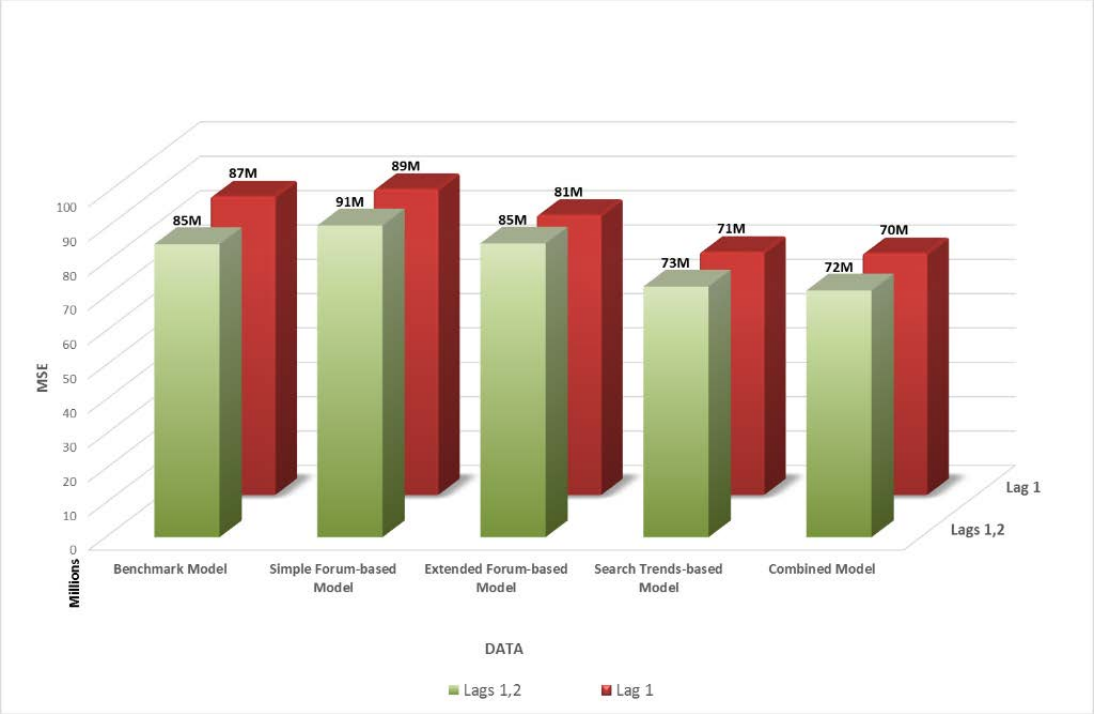


Figure C2 - Prediction Results for NN (MSE)

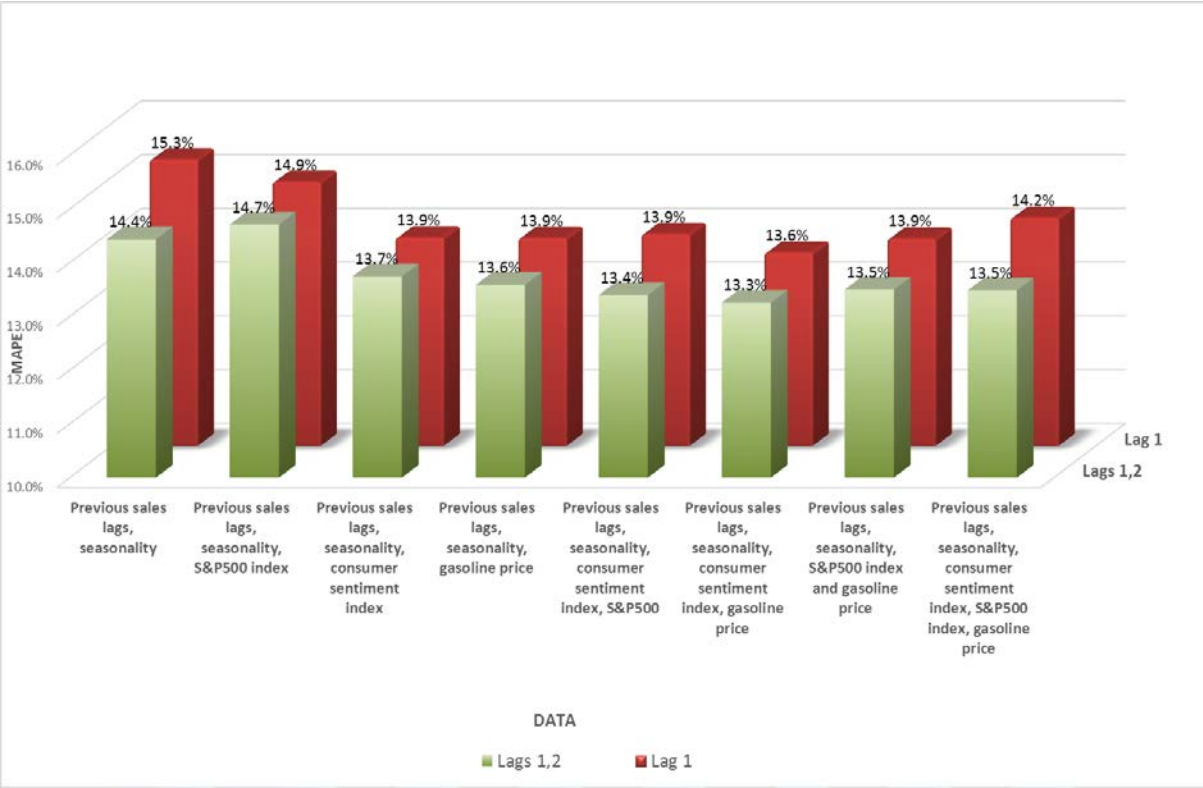


Appendix D– Benchmark Analysis

Our goal in selecting benchmark data is to create a benchmark that has high predictive capacity so that the additional contribution of utilizing search trend data and forum data can be strictly measured. The need for using strict benchmarks has been reported in related studies such as Goel et al. (2010). Therefore, in addition to using previous sales lags and seasonality (measured by sales in the same month, in the previous year as in Choi and Varian 2009), we evaluated various combinations of the latest available consumer sentiment index, stock market index (S&P500), and gasoline prices.

Figure D1 presents the predictive performance of the different benchmark data combinations using an NN algorithm. Overall, the most efficient benchmark was obtained using a combination of previous months' sales, seasonality, consumer sentiment index, and gasoline prices.

Figure D1 - MAPE Values for Different Benchmarks



Appendix E – Extended Keyword Selection

To test the robustness of our keyword selection process, in this section we repeat the analysis using a different keyword selection method. Specifically, we added predictors (explanatory variables) to the models described in the main body of the paper, based on search and forum data derived from additional keywords: specifically, keywords associated with the top selling car model for each brand. Figure E1 displays the results obtained with LR using the different data representations. Figure E2 displays the results obtained with NN using the different data representations. Overall, the findings obtained using this keyword approach are similar to those obtained using the keyword approach reported in the main body of the paper.

Figure E1 - Prediction Results for LR

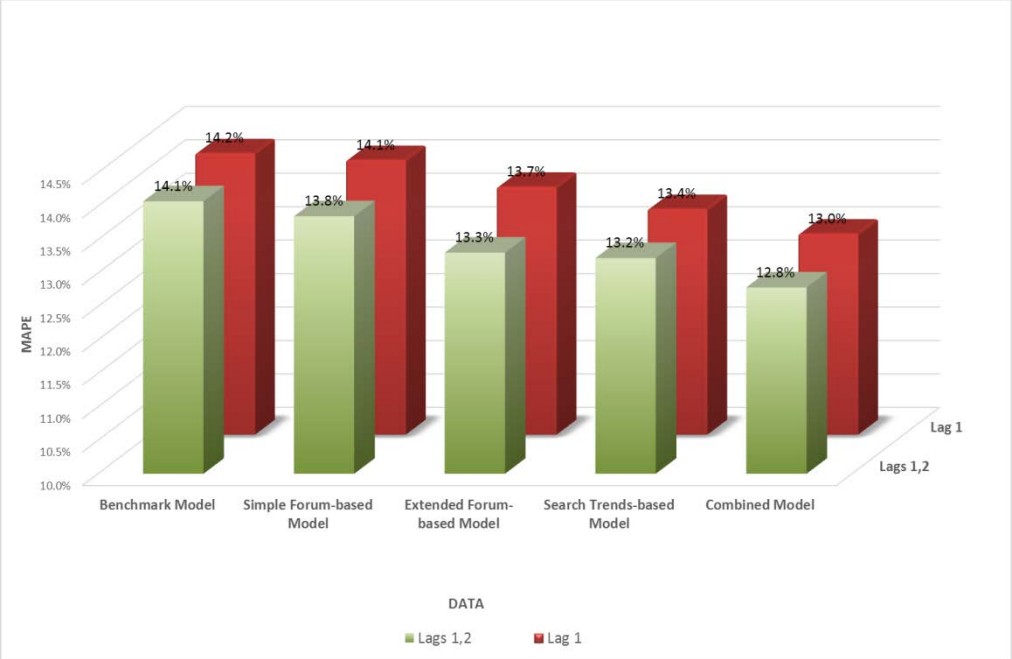
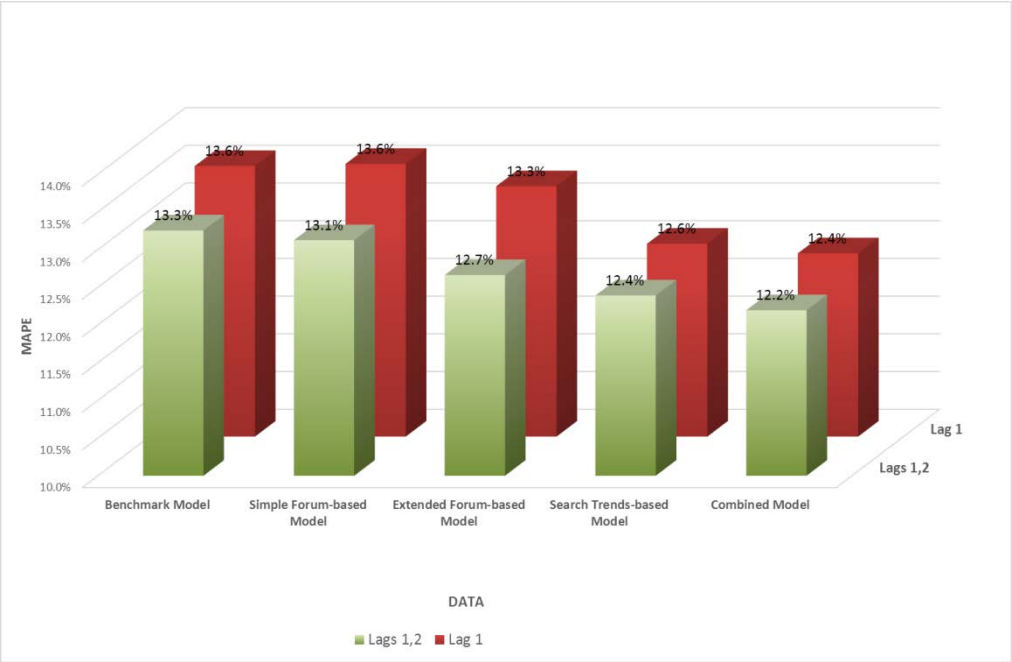


Figure E2 - Prediction Results for NN



Appendix F – Additional Forecasting Algorithms

One interesting finding of this research is that the non-linear NN model outperformed the LR model. This suggests that advanced non-linear methods can outperform the popular linear methods reported in related marketing and IS literature for such prediction tasks. To evaluate the robustness of this finding, we evaluated the predictive performance of two additional non-linear methods: a RandomForest method (using 1,000 sub-trees, implemented in R package randomForest) and a Support Vector Machine (SVM) regression using a radial kernel (implemented in R package e1071). Figure F1 presents predictive accuracy using SVM, and Figure F2 presents predictive accuracy using RandomForest. Table F1 details the MAPE values for the best-performing models. From these figures and table it is evident that the two additional non-linear methods also outperformed LR.

Table F1 - Best Performance Different Prediction Methods

Model	**MAPE**
LR	13.3%
NN	12.2%
SVM	12.3%
RandomForest	12.6%

Figure F1 - Prediction Results for SVM

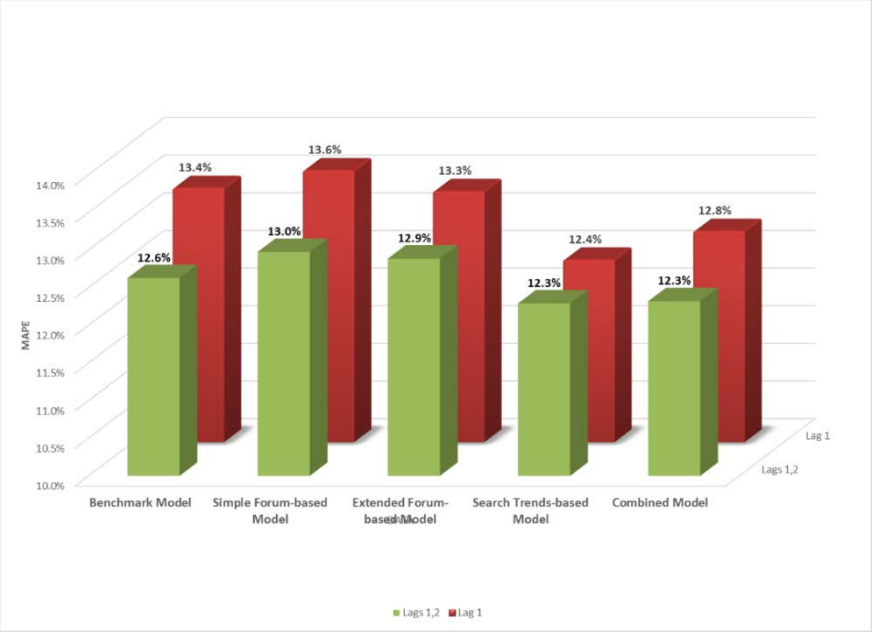
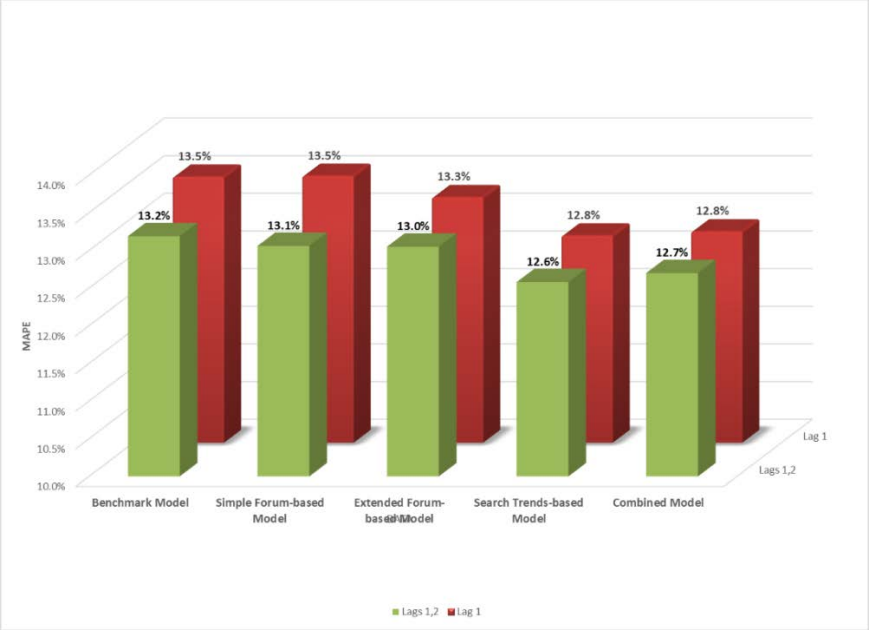


Figure F2 - Prediction Results for RandomForest



Appendix G – Forum Data Characteristics

In order to represent forum data we used Google's vast scan of the internet. To the best of our knowledge, this is the most comprehensive scan of forum data that has been made available for any academic research. Figure G1 provides a distribution of the relative volume of forum mentions per brand. Figures G2 and G3 provide distributions of the volume of forum mentions with positive or negative sentiment, respectively.

Figure G1 –Brand Forum Mention Average Volume (by Decile)

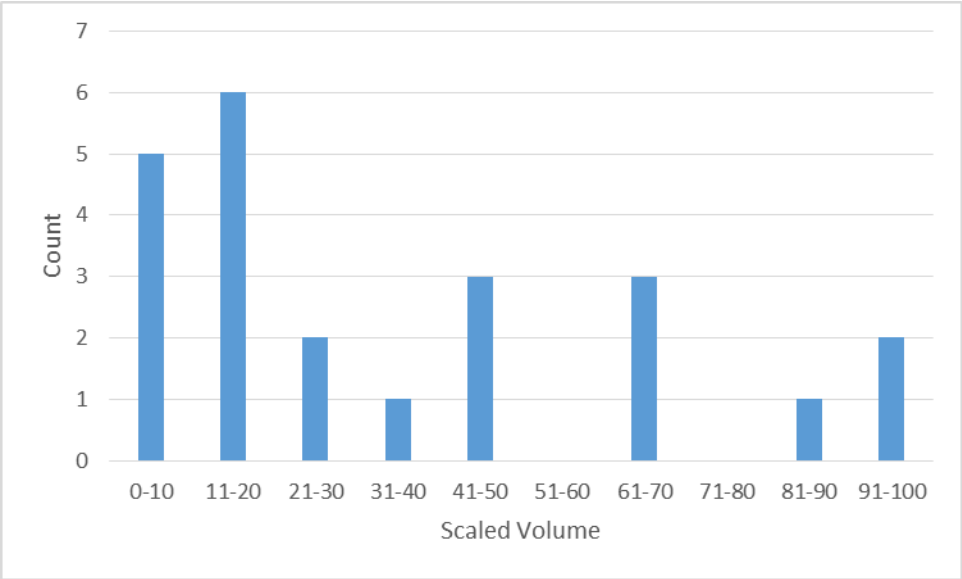


Figure G1 presents a distribution of scaled forum mention volume for the 23 different brands. To preserve data confidentiality, the volume of forum mentions for each brand is scaled using the following method:
(brand mention volume) / [(volume for the most mentioned brand) – (volume for the least mentioned brand)]

Figure G2 –Brand Forum Positive Sentiment Average Scores (by Decile)

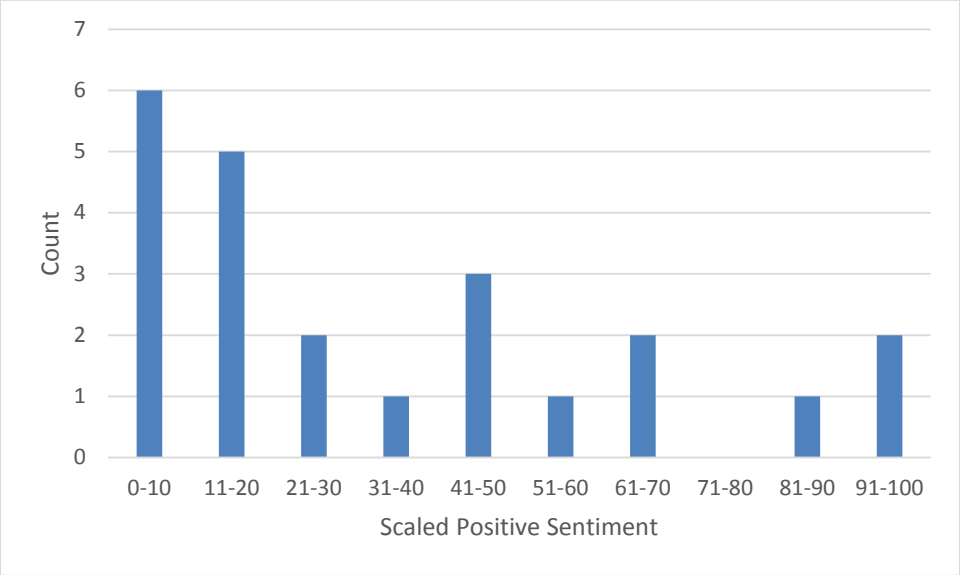


Figure G2 presents a distribution of scaled positive sentiment mentions for the 23 different brands. To preserve data confidentiality, the volume of positive mentions for each brand is scaled using the following method:
(brand positive mentions volume)/[(positive mentions for the brand with the highest volume of positive mentions)–(positive mentions for the brand with the lowest volume of positive mentions)]

Figure G3 –Brand Forum Negative Sentiment Average Scores (by Decile)

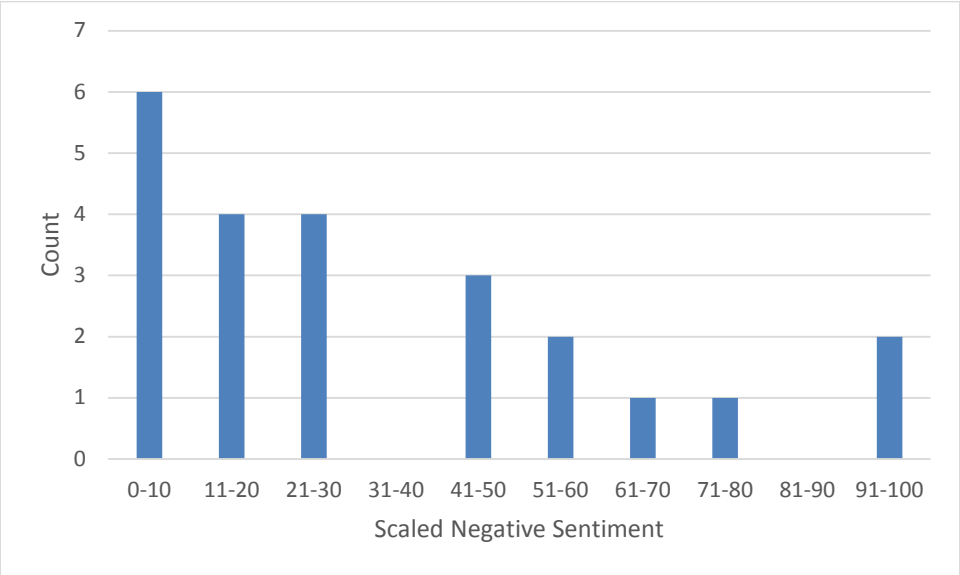
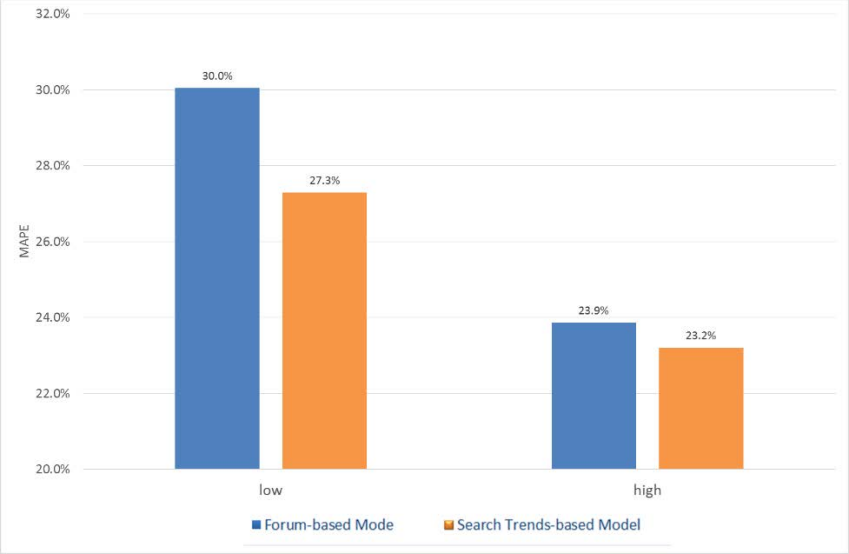


Figure G3 presents a distribution of scaled negative sentiment mentions for the 23 different brands. To preserve data confidentiality the volume of negative mentions for each brand is scaled using the following method:
(brand negative mentions volume)/[(negative mentions for the brand with the highest volume of negative mentions)–(negative mentions for the brand with the lowest volume of negative mentions)]

Appendix H – Prediction Results According to Car Model Characteristics – Using Different Price Thresholds

In the body of the paper we provide an analysis of predictive accuracy based on search trend data and forum-based data using a price threshold of $20,000. In Figure H1 we provide a similar analysis, but using different thresholds. In this case "premium" models include only models with a list price of $40,000 or above, and “value” models brands are cheaper than $30,000. Using these different thresholds we reach similar conclusions, showing that search trend data provides a greater advantage over forum data in the case of value brands (as compared with premium brands).

Figure H1. Prediction Results for Premium vs. Value Brands Based on Price



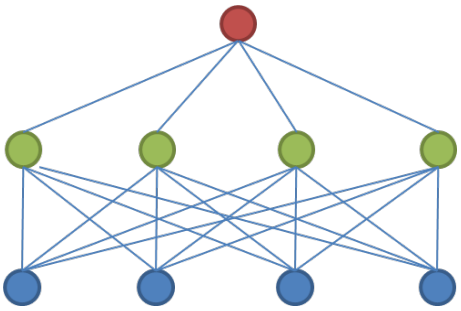
Appendix I – Neural Network

The neural network is a biologically-inspired model that attempts to learn patterns from data directly (Rumelhart et al. 1986). The NN is represented by a weighted directed graph containing three types of nodes (or “neurons”) organized in layers: the input nodes, the hidden nodes and the output node(s). A neuron receives input signals from a previous layer, aggregates those signals based on an input function, and generates an output signal based on an output (or

transfer) function. The output signal is then routed to the other nodes in the network according to the network configuration. Each link connecting any two nodes is characterized by a weight. These weights are determined through a training process in which the NN repeatedly receives examples of past data instances for which the actual output is known, thereby allowing the system to adjust the weights.

Finding the values of these weights requires solving an optimization problem. Perhaps the most common method is the backpropagation algorithm (see, for instance, Bishop ,1995). This method consists of two phases: feed-forward and backwards-propagation. In the feed-forward phase, outputs are generated for each node on the basis of the current weights and are propagated to the output nodes to generate predictions. Then, in the backwards-propagation phase, "prediction errors" are propagated back, layer by layer, and the weights of the connections between the nodes are adjusted for error minimization. The feed-forward and backwards-propagation phases are executed iteratively, until convergence.

Figure I1 - An Illustration of a Simple NN



An illustration of a NN with 4 input nodes (blue), one layer with 4 hidden nodes (green), and a single output node (red).

References

Amblee, N. and Bui, T. 2008. "Can Brand Reputation Improve the Odds of Being Reviewed On-Line?," *International Journal of Electronic Commerce* (12:3), pp. 11-28.

Berger, J., and Milkman, K. L. 2012. "What Makes Online Content Viral?," *Journal of Marketing Research* (49:2), pp. 192–205.

Bishop, C. M. 1995. *Neural Networks for Pattern Recognition*, Oxford, UK: Clarendon Press.

Blackwell D., Miniard P. W., and Engel, J. F. 2001. *Consumer Behavior*, 9th ed. Orlando, FL: Harcourt.

Bloch, P. H., and Richins, M. L. 1983. "A Theoretical Model for the Study of Product Importance Perceptions," *Journal of Consumer Research* (47), pp. 69-81.

Carlson, R. L. 1978. "Seemingly Unrelated Regression and the Demand for Automobiles of Different Sizes, 1965-75: A Disaggregate Approach," *Journal of Business* (51:2), pp. 243-262.

Chakravarty, A., Yong, L., and Mazumdar, T. 2010. "The Differential Effects of Online Word-of-Mouth and Critics' Reviews on Pre-Release Movie Evaluation," *Journal of Interactive Marketing* (24: 3), pp. 185-197.

Chen, Y., and Xie, J. 2008. "Online Consumer Reviews: A New Element of Marketing Communications Mix," *Management Science* (54:3), pp. 477–491.

Chevalier, J., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research* (43:3), pp. 345-354.

Chintagunta, P. K., Gopinath, S., and Venkataraman, S. 2011. "The Effect of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation across Local Markets," *Marketing Science* (29:5), pp. 944-957.

Choi, H., and Varian, H. 2009. "Predicting the Present with Google Trends," working paper.

Choi, H., and Varian, H. 2011. "Predicting the Present with Google Trends," working paper.

Dellarocas, C. 2006. "Strategic Manipulation of Internet Opinion Forums: Implications for Consumers and Firms," *Management Science* (52:10), pp. 1577–1593.

Dellarocas, C., Awad, N., and Zhang, X. 2007. "Exploring the Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," *Journal of Interactive Marketing* (21:4), pp. 23–45.

Dellarocas, C., and Narayan, R. 2006. "A Statistical Measure of a Population's Propensity to Engage in Post-Purchase Online Word-of-Mouth," *Statistical Science* (21:2), pp. 277–285.

Dewan, S., and Ramaprasad, J. 2009. "Chicken and Egg? Interplay between Music Blog Buzz and Album Sales," in *PACIS 2009 Proceedings*.

Dewan, S., and Ramaprasad, J. 2012. "Music Blogging, Online Sampling, and the Long Tail," *Information Systems Research* (23:3), pp. 1056–1067.

Dhar, V., and Chang, E. A. 2009. “Does Chatter Matter? The Impact of User-Generated Content on Music Sales," *Journal of Interactive Marketing* (23:4), pp. 300-307.

Dholakia, U. M. 2001. “A Motivational Process Model of Product Involvement and Consumer Risk Perception,” *European Journal of Marketing* (35:11), pp. 1340–1360.

Du, R. Y., and Kamakura, W. A. 2012. “Quantitative Trendspotting,” *Journal of Marketing Research* (49:4), pp. 514-536.

Duan, W., Gu, B., and Whinston, A. B. 2008a. “The Dynamics of Online Word-of-Mouth and Product Sales—An Empirical Investigation of the Movie Industry,” *Journal of Retailing* (84:2), pp. 233–242.

Duan, W., Gu, B., and Whinston, A. B. 2008b. “Online Reviews Matter? An Empirical Investigation of Panel Data,” *Decision Support Systems* (45:4), pp. 1007–1016.

Forman, C., Ghose, A., and Wiesenfeld, B. 2008. “Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets,” *Information Systems Research* (19:3), pp. 291–313.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., and Brilliant, L. 2008. “Detecting Influenza Epidemics Using Search Engine Query Data,” *Nature* (457:7232), pp. 1012-1014.

Godes, D., and Mayzlin, D. 2004. “Using Online Conversations to Study Word-of-Mouth Communication,” *Marketing Science* (23:4), pp. 545-560.

Goel, S., Hofman, J. M., Lahaie, S., Pennock, D. M., and Watts, D. J. 2010. “Predicting Consumer Behavior with Web Search,” *Proceedings of the National Academy of Sciences* (107:41), pp. 17486–17490.

Greenspan, A., and Cohen, D. 1999. "Motor Vehicle Stocks, Scrappage, and Sales," *The Review of Economics and Statistics* (81:3), pp. 369-383.

Gu, B., Park, J., and Konana, P. C. 2012. "The Impact of External Word-Of-Mouth Sources on Retailer Sales for High Involvement Products," *Information Systems Research* (23:1), pp. 182-196.

Hennig-Thurau, T., Gwinner, K. P., Walsh, G. and Gremler, D. D. 2004. "Electronic Word-of-Mouth Via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?,". *Journal of Interactive Marketing*, (18), pp. 38–52.

Hu, N., Liu, L., and Zhang, J. 2008. “Do Online Reviews Affect Product Sales? The Role of Reviewer Characteristics and Temporal Effects,” *Information Technology Management* (9:3), pp. 201–214.

Hu, Y, Du, R. Y., and Damangir, S. 2014. "Decomposing the Impact of Advertising: Augmenting Sales with Online Search Data," *Journal of Marketing Research*, forthcoming.

Hymans, S. H., Ackley, G., and Juster, F. T. 1970. "Consumer Durable Spending: Explanation and Prediction," *Brookings Papers on Economic Activity*, pp. 173-206.

Landwehr, J. R., Labroo, A. A., and Herrmann, A. 2011. "Gut Liking for the Ordinary: Incorporating Design Fluency Improves Automobile Sales Forecasts," *Marketing Science* (30:3), pp. 416-429.

Laurent, G., and Kapferer, J. 1985. “Measuring Consumer Involvement Profiles,” *Journal of Marketing Research* (22:1), pp. 41–53.

Lazer , D., Kennedy, R., King, G., and Vespignani, A. 2014. “The Parable of Google Flu: Traps in Big Data Analysis,” *Science* (343:6176), pp. 1203-1205.

Liu, Y. 2006. “Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue,” *Journal of Marketing* (70:3), pp. 74–89.

Lovett, M., Peres, R., and Shachar, R. 2014. “On Brands and Word-of-Mouth,” *Journal of Marketing Research*, forthcoming.

Luo, X., Zhang, J., and Duan, W. 2013. "Social Media and Firm Equity Value," *Information Systems Research* (24:1), pp.146-163.

Moorthy, S., Ratchford, B. T., and Talukdar, D. 1997. "Consumer Information Search Revisited: Theory and Empirical Analysis," *Journal of Consumer Research* (23:4), pp. 263–277.

Netzer, O., Feldman, R., Goldenberg, J., and Fresko, M. 2012. "Mine Your Own Business: Market-Structure Surveillance Through Text Mining," *Marketing Science* (31:3), pp. 521-543.

Preis, T., Moat, H. S., & Stanley, H. E. 2013. "Quantifying Trading Behavior in Financial Markets Using Google Trends," *Scientific Reports*, 3.

Rui, H., Liu, T., and Whinston, A. 2012. "Whose and What Chatter Matters? The Impact of Tweets on Movie Sales," working paper.

Rumelhart, D. E., McClelland, J. L. and Williams, R. J. 1986. "Learning Internal Representation by Error Propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Cambridge, MA: MIT Press, pp. 318–362.

Seebach, C., Pahlke, I., and Beck, R. 2011. "Tracking the Digital Footprints of Customers: How Firms Can Improve Their Sensing Abilities to Achieve Business Agility," in *ECIS 2011 Proceedings*.

Shmueli, G. 2010. “To Explain or to Predict?,” *Statistical Science* (25:3), pp. 289–310.

Shmueli, G., and Koppius, O. 2011. “Predictive Analytics in Information Systems Research,” *MIS Quarterly* (35:3), pp. 553-572.

Sundaram, D.S., Mitra, K., and Webster, C. 1998. "Word-Of-Mouth Communications: a Motivational Analysis," *Advances in Consumer Research* (25), pp. 527-531.

Urban, G. L., Hauser, J. R., and Roberts, J. H. 1990. "Prelaunch Forecasting of New Automobiles," *Management Science* (36:4), pp. 401-21.

Urban, G. L., Hulland, J. S., and Weinberg, B. D. 1993. "Premarket Forecasting for New Consumer Durable Goods: Modeling Categorization, Elimination, and Consideration Phenomena," *Journal of Marketing* (57:2), pp. 47-63.

Viswanathan, S., Gosain, S., Kuruzovich, J., and Agarwal, R. 2007. “Online Infomediaries and Price Discrimination: Evidence From the Auto-Retailing Sector,” *Journal of Marketing* (71:3), pp.89-107.

Vosen, S., and Schmidt, T. 2011. “Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends,” *Journal of Forecasting* (30:6), pp. 565-578.

Wang, F.-K., Chang, K.-K., and Tzeng, C.-W. 2011. "Using Adaptive Network-Based Fuzzy Inference System to Forecast Automobile Sales," *Expert Systems with Applications* (38:8), pp. 10587-10593.

Werbos, P. J. 1974. “Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences,” PhD dissertation, Harvard University.

Wu, L., and Brynjolfsson, E. 2009. “The Future of Prediction: How Google Searches Foreshadow Housing Prices and Quantities,” in *Proceedings of the 2009 International Conference on Information Systems*.

Zaichkowsky, J. L. 1985. ”Measuring the Involvement Construct,” *Journal of Consumer Research* (12:3), pp. 341-352.

Zhang, G., Patuwo, B. E., and Hu, M. Y. 1998. “Forecasting with Artificial Neural Networks: The State of the Art,” *International Journal of Forecasting* (14:1), pp. 35–62.

Zhu, F., and Zhang, X. M. 2010. ‘‘Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics,’’ *Journal of Marketing* (74:2), pp. 113-148.