YE HU, REX YUXING DU, and SINA DAMANGIR*

Unlike sales data, data on intermediate stages of the purchase funnel (e.g., how many consumers have searched for information about a product before purchase) are much more difficult to acquire. Consequently, most advertising response models have focused directly on sales and ignored other purchase funnel activities. The authors demonstrate, in the context of the U.S. automotive market, how consumer online search volume data from Google Trends can be combined with sales data to decompose advertising's overall impact into two underlying components: its impacts on (1) generating consumer interest in prepurchase information search and (2) converting that interest into sales. The authors show that this decompositional approach, implemented through a novel state-space model that simultaneously examines sales and search volumes, offers important advantages over a benchmark model that considers sales data alone. First, the approach improves goodness-of-fit, both in and out of sample. Second, it improves diagnosticity by distinguishing advertising effectiveness in interest generation from its effectiveness in interest conversion. Third, the authors find that overall advertising elasticity can be biased if researchers consider only sales data.

Keywords: Google Trends, advertising response model, market response model, product information search, dynamic linear model

Online Supplement: http://dx.doi.org/10.1509/jmr.12.0215

Decomposing the Impact of Advertising: Augmenting Sales with Online Search Data

Modeling the purchase funnel and information-processing function that leads to the sale of a product has been a central part of marketing research (Bettman 1979; Bettman, Luce, and Payne 1998; Engel and Blackwell 1982; Howard and Sheth 1969; Kotler, Rackham, and Krishnaswamy 2006). For high-involvement purchase decisions, such as those for large-ticket durable goods, consumers are highly motivated to gather product information (Alba and Hutchinson 1987; Beatty and Smith 1987; Moorthy, Ratchford, and Talukdar 1997; Punj and Staelin 1983; Ratchford, Lee, and Talukdar 2003; Srinivasan and Ratchford 1991; Zaichkowsky 1985). More broadly, in contexts in which seeking product information before purchase is the norm, the purchase funnel can be viewed, at the most rudimentary level, as consisting of two stages: the stage leading to prepurchase information search and the final purchase stage (Lilien, Kotler, and Moorthy 1992; Newman and Staelin 1972).

From such a view of the purchase funnel, advertising can drive sales in two basic ways: first, by making consumers interested enough in the focal product that they would seek information about it and, second, by converting information-seeking consumers into buyers. To isolate these two distinct sources of impact, one must be able to decompose sales into a function of (1) consumer interest in seeking information

*Ye Hu is Associate Professor of Marketing (e-mail: yehu@uh.edu), Rex Yuxing Du is Hurley Associate Professor of Marketing (e-mail: rexdu@bauer.uh.edu), and Sina Damangir is a doctoral student (e-mail: sdamangir@uh.edu), C.T. Bauer College of Business, University of Houston. The authors thank three anonymous *JMR* reviewers for their numerous inputs in improving the original manuscript. They also thank Ed Blair, Betsy Gelb, Jackie Kacen, Carl Mela, and participants at the 2012 Marketing Science Conference for helpful comments and suggestions. Fred Feinberg served as associate editor for this article.

ISSN: 0022-2437 (print), 1547-7193 (electronic)

DOI: 10.1509/jmr.12.0215

about the focal product before making a purchase decision (hereinafter referred to as "consumer prepurchase information interest," or simply "consumer interest") and (2) the extent to which consumer interest is converted into sales (hereinafter referred to as "interest conversion," or simply "conversion"). Marketers need to allow for the possibility that consumer interest and interest conversion may respond to advertising differently and follow distinct trajectories over time.

To accomplish this interest generation versus conversion decomposition, we must augment sales data with a tracking measure of consumer interest. Historically, such a measure would be available mainly through repeated cross-sectional surveys (Boyd, Ray, and Strong 1972; Newman and Lockeman 1975; Newman and Staelin 1972; Palda 1966). However, in addition to the common caveats associated with consumer self-reports (e.g., sampling, response, nonresponse errors), such surveys can be time consuming and cost prohibitive.

However, over the past decade, consumers have relied increasingly on the Internet in gathering product information, especially when considering large-ticket durable goods such as automobiles (Ratchford, Lee, and Talukdar 2003; Ratchford, Talukdar, and Lee 2007; Zettelmeyer, Morton, and Silva-Risso 2006). For example, 79% of new-vehicle buyers in the United States use the Internet to conduct prepurchase research (J.D. Power and Associates 2012). Furthermore, consumers who use the Internet to gather product information have relied increasingly on search engines to help them find the most relevant information. According to the 2012 Pew Internet & American Life Project Poll (Purcell, Brenner, and Rainie 2012), 91% of Internet users in the United States use search engines on a regular basis. Among new-vehicle buyers who use the Internet during their car shopping process, 84% rely on search engines to navigate through the wealth of information available online from automaker, dealership, and third-party websites (J.D. Power and Associates 2008).

As consumers become increasingly dependent on the Internet for product information, their reliance on search engines as a gateway to the Web grows. Such a development has opened a promising new way to track shifts in consumer interest—that is, by monitoring changes in the intensity of consumer searches for keywords related to various products. Indeed, recognizing the potential value of such tracking data to marketers, in 2008 Google introduced a Web facility called Google Trends (http://google.com/trends, previously known as Google Insights for Search). Although it is meant for marketers, any user can access it, free of charge.

As a source of consumer intelligence, Google Trends presents several appealing features. First, it allows for the tracking of various terms that consumers have typed into Google's search box, generating volume indexes going as far back as January 2004. Second, data from Google Trends are updated in near real time and aggregated on a weekly basis (or daily for the most popular queries), enabling users to track consumer interest with little time delay. Third, search volume indexes from Google Trends are highly customizable. For example, search terms can be combined or excluded to formulate composite queries, and searches can be filtered by geographic areas (e.g., countries, states, cities), time ranges (e.g., May 2004—May 2008), and categories (e.g., Beauty & Fitness, Autos & Vehicles, Computers & Electronics). Fourth, and perhaps most important, Google is by far the most dominant search engine. According to the 2012 Pew Internet & American Life Project Poll (Purcell, Brenner, and Rainie 2012), Google is the search engine most often used by 83% of U.S. Internet users, followed by Yahoo (6%) and Bing (3%). Given the ubiquity of consumer online searches and Google's dominance in this space, the volume of Google searches can plausibly be viewed as a reflection of the collective interests of Internet users.

These appealing features aside, it is an empirical question whether shifts in Google Trends indexes can be treated as a good proxy for shifts in consumer prepurchase information interest for a particular product. The answer will certainly depend on the product category under study. In this article, we focus on the U.S. automobile industry, a context in which consumers are known to conduct extensive prepurchase information search and do so increasingly on the Internet by using search engines as a gateway (J.D. Power and Associates 2008, 2012; Ratchford, Lee, and Talukdar 2003; Ratchford, Talukdar, and Lee 2007).

We are interested in jointly modeling the dynamics of (1) the volume of Google searches for a vehicle and (2) the sales of that vehicle. By treating the former as a proxy for consumer prepurchase information interest, we propose a modeling framework through which the impact of advertising on sales is decomposed into two distinct underlying components, one governing interest generation and the other governing interest conversion.

The article proceeds as follows. We begin by providing a brief overview of two literature streams: (1) studies that have modeled the impact of advertising on not only sales but also intermediate stages of the purchase funnel and (2) studies that have used Google search data as a proxy for consumer interest. We then present our proposed modeling framework for decomposing sales into interest and conversion, allowing the effectiveness of ad spend to differ between these two components. Next, we present our data, which cover monthly sales, Google search volume, and ad spend for 21 major vehicles from four popular segments in the United States (compact and midsize sport utility vehicles [SUVs] and compact and midsize sedans), between January 2004 and July 2012 (103 months). In our empirical analyses, we benchmark our proposed model against one in which only sales is modeled. We show that by augmenting sales with Google search data as a proxy for consumer interest, our proposed model offers several important advantages. First, it improves the goodness-of-fit for sales, both in and out of sample. Second, it improves diagnosticity by decomposing the overall impact of advertising into an interest generation component and an interest conversion component. Our results show that advertising elasticities for these two components often differ from each other. This finding reveals that a single measure of elasticity may paint too simplistic a picture of advertising effectiveness, which can actually differ substantially between generating consumer interest and converting that interest into sales. Third, we show that estimates of total ad elasticity, both short and long term, may be biased if we consider only sales. We conclude by discussing the main methodological and managerial implications of our study and its limitations and directions for further research.

LITERATURE REVIEW

Various paradigms have been proposed in delineating the prepurchase information-processing process (e.g., Bettman 1979; Bettman, Luce, and Payne 1998). In particular, scholars have developed numerous hierarchy-of-effects models as practical frameworks for integrating the distinct impacts of advertising on the mental and behavioral stages that consumers go through before making a purchase decision (Barry 1987; Lavidge and Steiner 1961; for a review, see Vakratsas and Ambler 1999).

Accordingly, in evaluating ad effectiveness, researchers have examined not only sales data but also data on other mental (e.g., awareness, memory, attitude) and behavioral responses (e.g., search for product information, requesting price quotes). For example, Hanssens, Parsons, and Schultz (2001, p. 9) note that practitioners have attempted to incorporate intermediate response measures in sales response models. More recently, Srinivasan, Vanhuele, and Pauwels (2010) constructed a market response model that explicitly links survey-based measures of consumer mindset with sales. Similarly, combining consumer survey data with actual purchase data, Bruce, Peters, and Naik (2012) test a theoretical framework on how advertising works, attempting to uncover the pathways for experience, cognition, and affect that influence purchase. Both studies yield new insights into how advertising affects how consumers "think" and "feel," which in turn affects what they buy (i.e., the dynamics between upper purchase funnel activities and sales).

In this study, we have a different focus. Our goal is to model the impacts of ad spend on consumer product information searches and conversion of those searches into purchases. We believe such a goal is worth pursuing for two reasons. First, actively seeking product information is often viewed as an integral behavioral predecessor to product purchase decisions (Jacoby, Szybillo, and Busato-Schach 1977). From a theoretical standpoint, it is worthwhile to quantify the impact of advertising on what consumers "search" for in addition to how they "think" and "feel." Second, consumers are increasingly relying on the Internet in gathering product information and depend on search engines as a gateway. With the advent of tracking services such as Google Trends, marketers can readily monitor consumer interest in a product by using the volume of searches for it as a proxy. Thus, from a practical standpoint, it is much easier to track consumer search interest than to track consumer mindsets through repeated surveys, which can be cost prohibitive and time consuming.

In addition to differing in focus, our study differs from Srinivasan, Vanhuele, and Pauwels (2010) and Bruce, Peters, and Naik (2012) in product context. Whereas those authors examine purchases of consumer packaged goods, we examine purchases of new vehicles. Such large-ticket durable goods represent a context in which consumers are highly motivated to gather prepurchase information and increasingly do so through online searches (J.D. Power and Associates 2008, 2012; Ratchford, Lee, and Talukdar 2003; Ratchford, Talukdar, and Lee 2007).

A key empirical question for our study is whether Google Trends data can serve as a reasonably good proxy for consumer interest in prepurchase information search, a construct that has been historically elusive (Newman and Lockeman 1975). A quick review of the emerging literature that uses Google Trends data has revealed many avenues for leveraging search volume indexes as predictors of real-world behavior. For example, in epidemiology, Ginsberg et al. (2009) and Pelat et al. (2009) show that the search volume for disease-related terms can be used as a real-time indicator of disease incidence rates, and it is cheaper and faster than measures collected through conventional epidemic surveillance methods. Choi and Varian (2009b) show that search volume data can help predict current consumer demand in a diverse set of industries including automotive, retailing, housing, and tourism. In macroeconomics, Askitas and Zimmermann (2009), Choi and Varian (2009a), Wu and Brynjolfsson (2009), and Vosen and Schmidt (2011) reveal that search volume data can be used to improve forecasts in housing prices and volumes, unemployment rates, and household expenditures. In finance, Da, Engelberg, and Gao (2011) show that search volumes for ticker names can be used to better predict stock prices. Finally, in marketing, Du and Kamakura (2012) show that seven common trends extracted from Google search data for 38 major vehicle brands can explain 74% of new car sales in the United States, highlighting the strong ties between consumer online searches and offline purchases. Joo et al. (2014) find that television advertising for financial services brands increases the number of related Google searches and searchers' tendency to use branded keywords (e.g., "Citibank").

In this study, we extend the aforementioned research. By treating the volume of Google searches as a proxy for consumer interest, we investigate how online search for a product—and, by extension, the level of consumer interest in it—responds to advertising. More importantly, by simultaneously examining search volume and sales, we are able to decompose the total impact of ad spend on sales into its effects on generating consumer interest and on converting consumer interest into sales. To the best of our knowledge, we are among the first to formally introduce search volume data in a sales response model not merely as a covariate or predictor but rather as an intermediate response measure in the purchase funnel. We hope our study can contribute to the increasing body of research by showing that online consumer interest tracking measures such as search volume data can be tapped not merely as a source of predictors but also as a source of insights into the impacts of advertising and other marketing instruments. As online consumer interest tracking data become increasingly available, we believe this stream of research will gain in importance and relevance for both marketing academics and practitioners.

MODEL

Let Q_{jt} denote the number of consumers seeking information about vehicle j before buying a car in period t. Let R_{jt} denote the fraction of Q_{jt} who actually purchase vehicle j in period t. In the context of new-vehicle shopping, in which it is the norm that consumers conduct prepurchase information search and do not purchase multiple units of the same vehicle, multiplying Q_{jt} by R_{jt} would, by definition, give rise to the sales of vehicle j in period t (i.e., $Y_{jt} = Q_{jt} \times R_{jt}$).

In this section, we propose a modeling framework that would enable us to decompose the overall impact of advertising on sales (Y_{jt}) into its impact on generating information-seeking consumers (Q_{jt}) and its impact on converting information seekers into purchasers (R_{jt}). It is beyond the scope of the current study to model contexts in which a nonnegligible portion of consumers would make a purchase without first seeking product information (i.e., $Y = Q \times R + Q'$, $Q' >> 0$), a situation we revisit when we discuss directions for further research.

Linking Latent States to Observed Variables

A key challenge is that Q_{jt} is not directly observable. One way to obtain estimates of Q_{jt} is to conduct repeated sample surveys; however, they can be costly and time consuming, rendering this approach infeasible under most circumstances. With the emergence of various online consumer tracking devices, marketers are presented with many "big data" alternatives that can be far more cost effective and timely. We posit that the amount of Googling for vehicle j in period t (which can be gathered from Google Trends in near real time for free) is highly correlated with Q_{jt} and thus can serve as a proxy.

Before proceeding, we note that a major threat to the validity of using Google search volume indexes as a proxy for Q_{jt}: consumers can Google a vehicle's name even if they are not shopping for a new car. For example, a consumer may Google the name of a vehicle because the vehicle in question is being recalled; because he or she is looking for parts and accessories, repair services, or a used car (vs. a new one); or because the consumer is simply looking for some general information about the automaker.

To address this concern, we adopted the following strategy to gather and model Google Trends data. First, in constructing the composite queries that we enter into Google Trends, we exclude keywords that are unlikely to be related to new vehicle shopping (e.g., "used," "parts," "recall," "repair"). For example, for the composite query "ford focus – used – parts – recall – repair," Google Trends would generate a volume index that includes searches containing "ford focus" but not "used," "parts," "recall," or "repair." Second, for each query, we extract two search volume indexes from Google Trends, one using "Autos & Vehicles" as the category filter and the other using "Vehicle Shopping" as the filter.[1] The first index, G_{jt}, represents the volume of searches that (1) match the composite query we constructed for vehicle j and (2) fall into the "Autos & Vehicles" category according to Google Trends, which can include shopping and nonshopping related searches. In contrast, the second index, S_{jt}, represents the volume of searches that not only match the composite query for vehicle j but also are categorized by Google Trends as vehicle-shopping related.

We postulate that the trend line of S_{jt} over time runs largely in parallel to that of Q_{jt}, with the ratio between S_{jt} and Q_{jt} following an i.i.d. log-normal distribution with mean K^S_j and variance V^S_j. Formally,

$$\ln(S_{jt}) = I_{jt} + v^S_{jt}, \tag{1}$$

where $I_{jt} \equiv \ln(Q_{jt})$, denotes the (latent) state of shopping-related consumer interest in vehicle j at time t, and $v^S_{jt} \sim N(K^S_j, V^S_j)$ acknowledges that the amount of shopping-related Google searches for vehicle j is a noisy manifestation of I_{jt}, with the noise-to-signal ratio determined by V^S_j.

Unlike S_{jt}, G_{jt} contains both shopping- and non-shopping-related searches for vehicle j, which may follow different dynamics and trend lines over time. Thus, we postulate the following:

$$\ln(G_{jt}) = I_{jt} + NI_{jt} + v^G_{jt}, \tag{2}$$

where NI_{jt} denotes the (latent) state of nonshopping interest in vehicle j at time t and $v^G_{jt} \sim N(K^G_j, V^G_j)$ acknowledges the noisy nature of the observed data.[2]

Let R_{jt} ($= Y_{jt}/Q_{jt}$) denote the fraction of Q_{jt} that converts into purchasers of vehicle j at time t, which we postulate is determined as

$$\ln(R_{jt}) = C_{jt} + \varphi_j I_{jt} + v^Y_{jt}, \tag{3}$$

where C_{jt} is a latent state variable that captures the baseline convertibility of consumer interest in vehicle j (Equations 9 and 10 give the equation of motion for C_{jt}); φ_j captures how the overall conversion rate varies as a function of I_{jt}, the latent state of shopping-related consumer interest; and v^Y_{jt} represents a contemporaneous random shock, which is distributed i.i.d. normal with mean zero and variance V^Y_j.

In Equation 3, although φ_j is unconstrained and is to be estimated empirically, we expect it to be negative because of the law of diminishing marginal returns. In other words, we expect that, holding the baseline convertibility (C_{jt}) constant, the overall conversion rate (R_{jt}) would decrease as the number of information seekers (Q_{jt}) increases. The assumption behind our expectation is that consumers are heterogeneous in their intrinsic interest in vehicle j, and the level of intrinsic interest is positively correlated with both the likelihood to seek information and the likelihood to convert after seeking information. Consequently, as the number of information seekers increases (e.g., after a major ad campaign), there should be disproportionately more low-interest consumers in the mix (i.e., those needing the extra push from the ad campaign to initiate an information search), which in turn lowers the overall conversion rate. Empirically, a negative and large (in absolute value) φ_j would indicate high diminishing convertibility as vehicle j attracts increasingly marginal information seekers.

Given Equation 3, $I_{jt} \equiv \ln(Q_{jt})$, and $Y_{jt} = Q_{jt} \times R_{jt}$, we have

$$\ln(Y_{jt}) = \varphi^*_j I_{jt} + C_{jt} + v^Y_{jt}, \tag{4}$$

where $\varphi^*_j \equiv 1 + \varphi_j$. Together, Equations 1, 2, and 4 define a system that links three latent state variables, I_{jt} (shopping interest), NI_{jt} (nonshopping interest), and C_{jt} (baseline con-

[1]For details on how Google Trends categorizes searches, see https://support.google.com/trends/answer/94792?hl=en&ref_topic=19361.

[2]An alternative specification is to remove Equation 2 (and Equations 7 and 8, which define the dynamics of nonshopping interest NI_{jt}). Empirically, we find that the full model outperforms the simpler alternative because by adding Equation 2 to Equation 1, the full model can simultaneously leverage information contained in "Vehicle Shopping" and "Autos & Vehicle" searches, both of which include shopping-related searches. Because Google Trends data are noisy, by using two correlated indicators (i.e., S_{jt} and G_{jt}, as opposed to S_{jt} alone), our full model can tap into their comovements, which send stronger/cleaner signals about the shared latent component (i.e., shopping interest I_{jt}).

vertibility of shopping interest), with three observed variables, S_{jt} (vehicle shopping–related Google searches), G_{jt} (both shopping- and non-shopping-related Google searches), and Y_{jt} (sales). Next, we introduce the system of equations that governs the dynamics of the three latent state variables.

Dynamics of the Latent States

We postulate that the dynamics of consumer shopping interest for vehicle j at time t (I_{jt}) is governed by Equations 5 and 6:

$$I_{jt} = \alpha^I_{jt} + \beta^I_j X_{jt}, \tag{5}$$

where α^I_{jt} captures the trend component in consumer shopping interest for vehicle j, whose equation of motion is given in Equation 6; $\beta^I_j X_{jt}$ captures the contemporaneous component in consumer shopping interest for vehicle j, which shifts as a function of X_{jt}, a vector of exogenous variables that includes lagged sales $\ln(Y_{j,t-1})$, consumer sentiment, gas prices, and a control for seasonality in vehicle shopping interest. We include $\ln(Y_{j,t-1})$ to allow for the possibility that lagged sales may influence current searches (e.g., through postpurchase contagion). We control for consumer sentiment, gas prices, and seasonality because these variables can potentially influence both consumer vehicle shopping interest and automakers' ad spend.

$$\alpha^I_{jt} = \delta^I_{j1}\alpha^I_{j,t-1} + \delta^I_{j2}\sum_{j'=1, j'\neq j}^{n}\alpha^I_{j',t-1} + \delta^I_{j3}\ln\left(A_{jt}\right) + \delta^I_{j4}\ln\left(\widetilde{A_{jt}}\right) + w^I_{jt}, \tag{6}$$

where α^I_{jt}, the trend component in consumer shopping interest for vehicle j at time t, is assumed to be a function of (1) its lagged value; (2) the sum of lagged trend components in consumer shopping interest for competing vehicles $j' = 1, ..., n, j' \neq j$; (3) the impact of own ad spend A_{jt}, which is log-transformed; (4) the impact of total competitive ad spend $\widetilde{A_{jt}}$, which is also log-transformed; and (5) a shock w^I_{jt}, which is assumed to be random and distributed $N(0, W^I_j)$.

In Equation 6, δ^I_{j3} and δ^I_{j4} capture, respectively, the short-term impacts of own and competitive ad spend on shopping interest for vehicle j, and δ^I_{j1} determines how quickly these impacts decay from one period to another. δ^I_{j2} allows for the possibility of "spillover" from lagged consumer interest in competing vehicles. For example, consumers who searched for information related to the Honda CR-V may go on to search for information related to the Toyota RAV4 and other vehicles from the compact SUV segment. Concurrent spillover between competing vehicles is captured by correlated w^I_{jt} and $w^I_{j't}$.

For the process governing the dynamics of NI_{jt}, nonshopping interest in vehicle j at time t, we impose a structure that is similar to Equations 5 and 6:

$$NI_{jt} = \alpha^{NI}_{jt} + \beta^{NI}_j X_{jt}, \text{ and} \tag{7}$$

$$\alpha^{NI}_{jt} = \delta^{NI}_{j1}\alpha^{NI}_{j,t-1} + \delta^{NI}_{j2}\sum_{j'=1, j'\neq j}^{n}\alpha^{NI}_{j',t-1} + \delta^{NI}_{j3}\ln\left(A_{jt}\right) + \delta^{NI}_{j4}\ln\left(\widetilde{A_{jt}}\right) + w^{NI}_{jt}, \tag{8}$$

where α^{NI}_{jt} represents the trend component, whose equation of motion is given in Equation 8; $\beta^{NI}_j X_{jt}$ captures the contemporaneous component, with X_{jt} including lagged sales, consumer sentiment, gas prices, and a seasonality control; and w^{NI}_{jt} is assumed to be distributed $N(0, W^{IN}_j)$.

For the process governing the dynamics of C_{jt}, the baseline convertibility of shopping interest in vehicle j at time t, we impose a structure that is again similar to Equations 5 and 6:

$$C_{jt} = \alpha^C_{jt} + \beta^C_j X_{jt}, \text{ and} \tag{9}$$

$$\alpha^C_{jt} = \delta^C_{j1}\alpha^C_{j,t-1} + \delta^C_{j2}\sum_{j'=1, j'\neq j}^{n}\alpha^C_{j',t-1} + \delta^C_{j3}\ln\left(A_{jt}\right) + \delta^C_{j4}\ln\left(\widetilde{A_{jt}}\right) + w^C_{jt}, \tag{10}$$

where α^C_{jt} represents the trend component, whose equation of motion is given in Equation 10; $\beta^C_j X_{jt}$ captures the contemporaneous component, with the same set of exogenous controls as in Equation 5; and w^C_{jt} is assumed to be random and distributed $N(0, W^C_j)$.

In Equation 10, δ^C_{j3} and δ^C_{j4} capture, respectively, the short-term impacts of own and competitive ad spend on converting consumer interest into sales for vehicle j, and δ^C_{j1} determines how quickly these impacts decay from one period to another. Similar to Equation 6, δ^C_{j2} allows for the possibility of competitive spillover in shopping interest convertibility.

The state-space model, which includes observation equations defined by Equations 1, 2, 4, 5, 7, and 9 and state equations 6, 8, and 10, imposes a specific structure that explicitly ties the generating process for Google shopping and nonshopping searches (S_{jt} and G_{jt}) into the generating process for sales (Y_{jt}). Because these three data-generating processes are intertwined, they must be taken into account jointly in model calibration. Hereinafter, we refer to this modeling framework as the "sales-and-search approach" or the "decompositional approach."

Endogeneity in Advertising Spending

Endogeneity can arise when lagged sales and other exogenous variables (e.g., economic conditions, seasonality) affect current ad spend as well as current consumer searches and sales. To address this issue, following Wooldridge (2008) and Ataman, Mela, and Van Heerde (2008), we explicitly model the data-generating process for ad spend as

$$\ln(A_{jt}) = \alpha^A_{jt} + \beta^A_j X_{jt} + v^A_{jt}, \text{ and} \tag{11}$$

$$\alpha^A_{jt} = \delta^A_{j1}\alpha^A_{j,t-1} + \delta^A_{j2}\ln\left(Y_{j,t-1}\right) + \delta^A_{j3}\ln\left(\widetilde{A_{j,t-1}}\right) + w^A_{jt}, \tag{12}$$

where current ad spend A_{jt} is determined by a base level α^A_{jt}, a temporary adjustment $\beta^A_j X_{jt}$, and a random shock $v^A_{jt} \sim N(0, V^A_j)$. X_{jt} contains consumer sentiment, gas prices, and seasonality. α^A_j, the base level ad spend, is modeled as a function of its lagged value $\alpha^A_{j,t-1}$, lagged sales $\ln(Y_{j,t-1})$, lagged competitive ad spend $\ln(\widetilde{A_{j,t-1}})$, and a random shock $w^A_{jt} \sim N(0, W^A_j)$. Thus, δ^A_{j1} captures the degree of inertia in ad spend decisions, δ^A_{j2} accounts for the possibility that current ad spend may be influenced by lagged sales, and δ^A_{j3}

allows for the possibility that current ad spend may be influenced by lagged competitive ad spend.

Benchmark Model: The Sales-Only Approach

An alternative to our proposed approach would be to treat the generating process for Google search data as independent of the generating process for sales data. Put differently, a simpler and potentially more robust approach would be to ignore the Google search data and focus on modeling the generating process for sales data alone. After all, as a new data source, Google Trends remains largely untested, especially when it comes to market response modeling. To facilitate comparison, we use the following sales-only model as an alternative against which we benchmark our proposed sales-and-search approach in the empirical application.

$$\ln(Y_{jt}) = \alpha^U_{jt} + \beta^U_j X_{jt} + v^U_{jt}, \text{ and} \tag{13}$$

$$\alpha^U_{jt} = \delta^U_{j1}\alpha^U_{j,t-1} + \delta^U_{j2}\sum_{j'=1,\,j'\neq j}^{n}\alpha^U_{j',t-1} + \delta^U_{j3}\ln(A_{jt}) + \delta^U_{j4}\ln(\widetilde{A_{jt}}) + w^U_{jt}, \tag{14}$$

where v^U_{jt} and w^U_{jt} are assumed to be random and distributed, respectively, $N(0, V^U_j)$ and $N(0, W^U_j)$. We note that the sales-only model is very similar to the model used by Ataman, Mela, and Van Heerde (2008). As with our proposed model, in calibrating the benchmark model, we also include Equations 11 and 12 to account for potential endogeneity in ad spend.

To facilitate the contrast between our proposed model and the benchmark, given Equations 5 and 9, we can rewrite Equation 4 as

$$\ln(Y_{jt}) = (\varphi^*_j\alpha^I_{jt} + \alpha^C_{jt}) + (\varphi^*_j\beta^I_j + \beta^C_j)X_{jt} + v^Y_{jt}, \tag{15}$$

which, compared with Equation 13, shows that our proposed approach is equivalent to decomposing the trend component of sales α^U_{jt} into two distinct parts, $\varphi^*_j\alpha^I_{jt}$ and α^C_{jt}, which are allowed to follow different dynamics.

Furthermore, given Equations 6 and 10, we can expand $\varphi^*_j\alpha^I_{jt} + \alpha^C_{jt}$ as

$$\begin{aligned}\varphi^*_j\alpha^I_{jt} + \alpha^C_{jt} &= \varphi^*_j\left(\delta^I_{j1}\alpha^I_{j,t-1} + \delta^I_{j2}\sum_{j'=1,\,j'\neq j}^{n}\alpha^I_{j',t-1}\right) \\ &+ \left(\delta^C_{j1}\alpha^C_{j,t-1} + \delta^C_{j2}\sum_{j'=1,\,j'\neq j}^{n}\alpha^C_{j',t-1}\right) + \left(\varphi^*_j\delta^I_{j3} + \delta^C_{j3}\right)\ln(A_{jt}) \\ &+ \left(\varphi^*_j\delta^I_{j4} + \delta^C_{j4}\right)\ln(\widetilde{A_{jt}}) + \left(\varphi^*_j w^I_{jt} + w^C_{jt}\right).\end{aligned} \tag{16}$$

Contrasting Equation 16 with Equation 14, we note that δ^U_{j3}, the short-term impact of own ad spend in Equation 14, is decomposed into the sum of $\varphi^*_j\delta^I_{j3}$ and δ^C_{j3} in Equation 16. Similarly, δ^U_{j4}, the short-term impact of competitive advertising in Equation 14, is decomposed into the sum of $\varphi^*_j\delta^I_{j4}$ and δ^C_{j4} in Equation 16. To the extent that ad spend has a different impact on generating consumer interest than on converting interest into sales, our decompositional approach would lead to more diagnostic (and potentially more accurate) inferences on advertising effectiveness. By contrast, the sales-only approach, which ignores the potentially distinct dynamics between interest generation and interest conversion, can produce only an overall (and thus less insightful) estimate of advertising effectiveness.

In addition to decomposing the short-term impacts of advertising into interest generation versus conversion, our proposed model also allows the decay rates of these impacts to be different (δ^I_{j1} and δ^C_{j1} in Equations 6 and 10). The sales-only approach allows for only one overall decay rate (δ^U_{j1} in Equation 14), which is more restrictive and can lead to less diagnostic (and potentially less accurate) inferences about the long-term impacts of advertising.

In conclusion, unlike the sales-only approach, our proposed model allows for the possibility that the generating processes behind search and sales data can inform each other and thus can benefit from joint calibration. If such benefit does exist and is properly captured through our proposed decompositional structure, the sales-and-search model should outperform the sales-only model in both in- and out-of-sample fit, which we test in our empirical application.

Model Calibration

To calibrate our model, we rewrite it in a state-space form for vehicles j = 1, ..., n such that

$$H_t = \theta\alpha_t + \beta X_t + v_t \quad \text{(Observation Equation), and} \tag{17}$$

$$\alpha_t = \delta_{lag}\alpha_{t-1} + \delta_z Z_t + w_t \quad \text{(State Equation),} \tag{18}$$

where
$$\begin{aligned}H_t &= [\ln(G_{1t}), \ln(S_{1t}), \ln(Y_{1t}), \ln(A_{1t}), \cdots \ln(G_{nt}), \ln(S_{nt}), \ln(Y_{nt}), \ln(A_{nt})]',\\ \alpha_t &= [\ln(\alpha^{NI}_{1t}), \ln(\alpha^I_{1t}), \ln(\alpha^C_{1t}), \ln(\alpha^A_{1t}), \cdots \ln(\alpha^{NI}_{nt}), \ln(\alpha^I_{nt}), \ln(\alpha^C_{nt}), \ln(\alpha^A_{nt})]',\\ \beta &= [\beta^{NI}_1, \beta^I_1, \beta^C_1, \beta^A_1, \cdots \beta^{NI}_n, \beta^I_n, \beta^C_n, \beta^A_n]',\\ v_t &= [v^{NI}_1, v^I_1, v^C_1, v^A_1, \cdots v^{NI}_n, v^I_n, v^C_n, v^A_n]' \sim N(0, V_{4n\times 4n}),\end{aligned}$$

$$\theta = \begin{bmatrix} \begin{matrix}1 & 1 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & \varphi^*_1 & 1 & 0\\ 0 & 0 & 0 & 1\end{matrix} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \begin{matrix}1 & 1 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & \varphi^*_n & 1 & 0\\ 0 & 0 & 0 & 1\end{matrix}\end{bmatrix},$$

$$\delta_z Z_t = \begin{bmatrix} \delta^{NI}_{13}\ln(Ad_{1t}) + \delta^{NI}_{14}\ln(\widetilde{Ad_{1t}}) \\ \delta^I_{13}\ln(Ad_{1t}) + \delta^I_{14}\ln(\widetilde{Ad_{1t}}) \\ \delta^C_{13}\ln(Ad_{1t}) + \delta^C_{14}\ln(\widetilde{Ad_{1t}}) \\ \delta^A_{12}\ln(Y_{1,t-1}) + \delta^A_{13}\ln(\widetilde{Ad_{1,t-1}}) \\ \vdots \\ \delta^{NI}_{n3}\ln(Ad_{nt}) + \delta^{NI}_{n4}\ln(\widetilde{Ad_{nt}}) \\ \delta^I_{n3}\ln(Ad_{nt}) + \delta^I_{n4}\ln(\widetilde{Ad_{nt}}) \\ \delta^C_{n3}\ln(Ad_{nt}) + \delta^C_{n4}\ln(\widetilde{Ad_{nt}}) \\ \delta^A_{n2}\ln(Y_{n,t-1}) + \delta^A_{n3}\ln(\widetilde{Ad_{n,t-1}}) \end{bmatrix},$$

$$\delta_{lag} = \begin{bmatrix} \Delta_{11} & \cdots & \Delta_{12} & \cdots & \Delta_{12} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \Delta_{j2} & \cdots & \Delta_{j1} & \cdots & \Delta_{j2} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \Delta_{n2} & \cdots & \Delta_{n2} & \cdots & \Delta_{n1} \end{bmatrix},$$

$$\Delta_{j1} \equiv \begin{bmatrix} \delta_{j1}^{NI} & 0 & 0 & 0 \\ 0 & \delta_{j1}^{I} & 0 & 0 \\ 0 & 0 & \delta_{j1}^{C} & 0 \\ 0 & 0 & 0 & \delta_{j1}^{A} \end{bmatrix}, \text{and}$$

$$\Delta_{j2} \equiv \begin{bmatrix} \delta_{j2}^{NI} & 0 & 0 & 0 \\ 0 & \delta_{j2}^{I} & 0 & 0 \\ 0 & 0 & \delta_{j2}^{C} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\forall j = 1, ..., n, \text{ and } w_t \sim N(0, W_{4n \times 4n}).$$

We estimate this system as a Bayesian dynamic linear model (West and Harrison 1997). We assume the priors on δs and βs to be normal and impose a hierarchical structure on the latter to allow within-segment pooling. We use an inverse-gamma prior for the variance terms in the error matrices (i.e., V and W). We draw the conditional posterior parameters using a Gibbs sampler with the forward-filtering-backward-smoothing procedure embedded within (Carter and Kohn 1994; Fruhwirth-Schnatter 1995). We run the Gibbs sampler with a total length of 25,000 draws, with the first 15,000 draws as burn-in. For details about the estimation algorithm, see Web Appendix A.

DATA

We focus our empirical application on the four most popular passenger-vehicle segments in the United States: compact sedan, midsize sedan, compact SUV, and midsize SUV. In each segment, we focus on the five or six best-selling models (i.e., those that had the highest sales and were available in the United States from January 2004 through July 2012). Such a focus on the major segments and established models means that we had to exclude newly launched, discontinued, and niche segments or models. The 21 vehicles we include in our analyses represent eight makes from seven automakers and account for at least 60% of sales in their respective segment over a window of 103 months. As an empirical illustration, we consider these data sufficient and leave the extension to other smaller segments and models for future researchers.

We assembled four data sets: sales (Y_t), ad spend (A_t), Google search volume indexes (G_t for general vehicle search and S_t for vehicle shopping search), and environmental controls (X_t). We gathered new vehicle sales data from *Automotive News* (www.autonews.com), which reports monthly unit sales in the United States by vehicle model. We gathered monthly model-level ad spend data from Ad$pender of Kantar Media. Most crucial to our study, we gathered search volume indexes from Google Trends for each of the vehicle models.

We used the Keyword Tool from Google AdWords to identify all the search terms that are commonly associated with each vehicle model, including popular abbreviations, nicknames, and misspellings (e.g., Volkswagen, VW, Volkswagon; Chevrolet, Cheverolet, Chevorlet, Chevy). Then, a vehicle-specific composite query encompassing all the search terms (joined by "+") was entered into Google Trends, with the filters set to "Web Search" in the "United States" from "January 2004 through July 2012" within the "Autos & Vehicles" category for general search G_t, or the "Vehicle Shopping" subcategory for shopping-related search S_t. As we noted previously, to minimize non-new-vehicle-shopping-related searches, the composite query for each vehicle excludes terms related to "used," "parts," "recalls," "repair," and so on, by using the minus sign. We obtained monthly search volume indexes from the resulting "Interest over time" charts. We normalized these indexes to a 0–100 scale, set in proportion to the volume of searches recorded in each month, with the highest being 100.[3] Figure 1 provides an illustration of the Google Trends interface through which we obtained our search volume indexes.

The last set of data we gathered relates to environmental controls (X_t), which are exogenous to the data-generating process and could temporarily shift (1) consumer interest in a vehicle, (2) the conversion of that interest into sales, and (3) the level of ad spend on the vehicle. In our empirical application, we included three such controls (in addition to lagged sales). The first is the national average gasoline price, a key factor in determining vehicle operating cost. We gathered this information from the website of the Federal Reserve Bank of St. Louis. The second environmental control is the University of Michigan Consumer Sentiment Index, a well-established barometer for the macro economy and consumer willingness to spend, especially on large-ticket durable goods such as new cars. Finally, to account for seasonality in a vehicle's sales, ad spend, general search, and shopping-related search, we used, respectively, the following four controls: industrywide sales, industrywide ad spend, Google Trends index for the whole "Autos & Vehicles" category, and Google Trends index for the whole "Vehicle Shopping" category (which can be obtained by leaving the Google Trends query box blank).

RESULTS

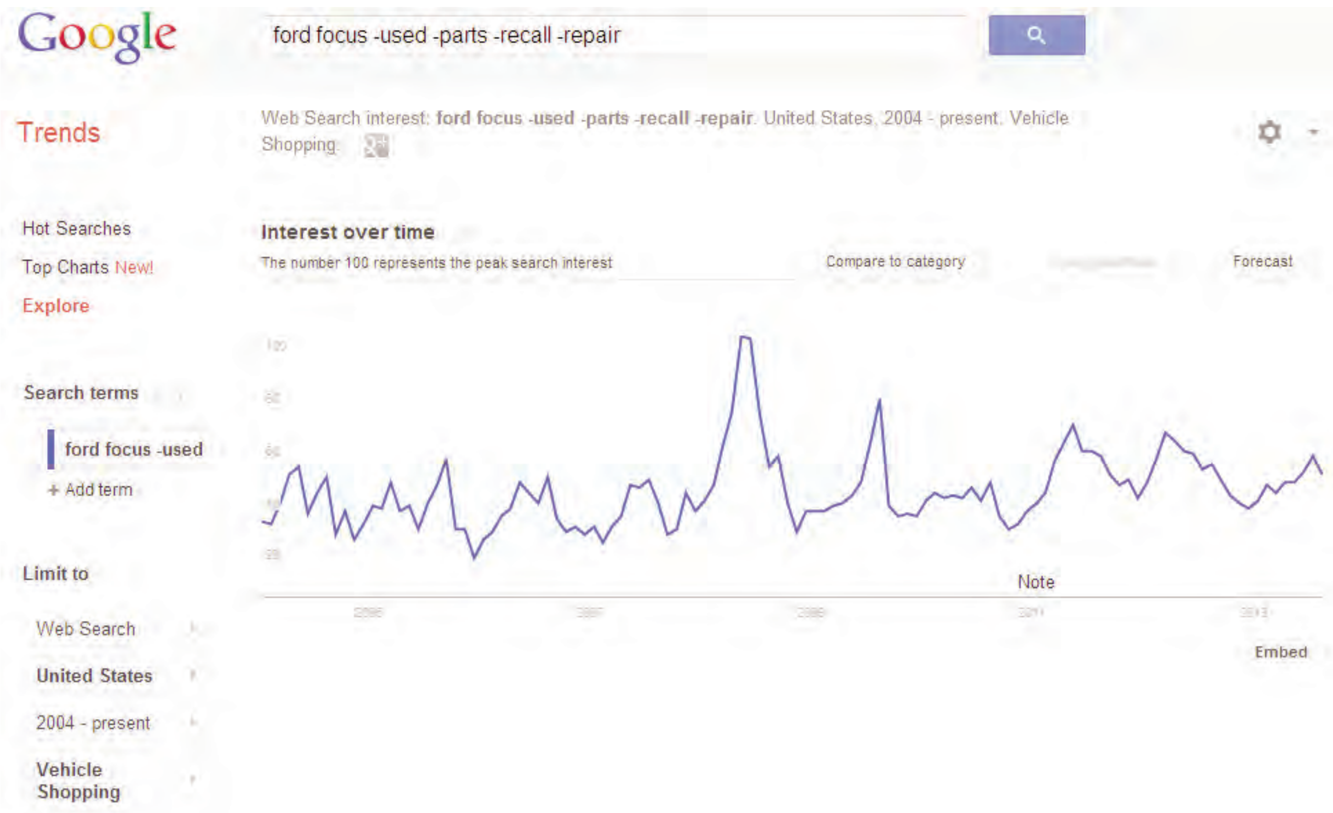
Model Performance

Before presenting our model estimates, we investigate whether the extra complexity of our decompositional approach can be justified over a simpler (and potentially more robust) alternative. As we discussed previously, a reasonable benchmark would focus on the sales data alone. Such a sales-only approach could outperform its sales-and-search counterpart in fitting the sales data if our proposed decompositional structure, which explicitly ties the generating process for search data into the generating process for sales, were misspecified and thus unwarranted.

To address this issue, we compare two models: our proposed approach, which calibrates Google search and sales

[3]Google Trends indexes all the raw search volume data for any given period in any given region. The indexation is performed by dividing the raw volume data by the total volume of Googling in that period from that region. Thus, it prevents a period or a region from having a larger/smaller index simply because there are more/fewer Google users in that period or region. As a result, strictly speaking, Google Trends indexes are calculated on the basis of proportions of all Google searches and can be interpreted as the intensity of searches among Google users in a given period and region.

Figure 1
GATHERING DATA FROM GOOGLE TRENDS



data jointly, and its sales-only counterpart (Equations 13–14). The *only* difference between these two models is whether search data are used. All other aspects are the same, including the treatment of endogeneity, spillovers, and environmental variables. Thus, such a comparison, carried out across 21 vehicles from four segments, should provide a strong test of the validity of our proposed modeling framework.

We evaluate the model performances using both in- and out-of-sample measures. For the in-sample comparison, we use the whole 103 months and compare the corrected Akaike information criterion (AIC_C) of the sales equation in both models (Hurvich and Tsai 1989; Naik and Raman 2003; Naik, Raman, and Winer 2005; Naik and Tsai 2001).[4] Table 1 shows that across the four segments, the decompositional model outperforms the sales-only model in in-sample AIC_C: compact SUV (381.48 vs. 410.70), midsize SUV (331.19 vs. 618.54), compact sedan (419.66 vs. 798.08), and midsize sedan (361.64 vs. 828.18). As a visual illustration, Figure 2 presents our model fit for the Ford Focus, which shows that both the search and sales data fit fairly well, with all the actual data points falling within the 95% confidence bands.

In addition to comparing in-sample fit, we recalibrate both models using data from the first 91 months and conduct out-of-sample forecasting using the remaining 12 months as holdout. The decompositional model outperforms the sales-only model in holdout mean absolute errors: compact SUV (.210 vs. .249), midsize SUV (.125 vs. .138), compact sedan (.235 vs. .246), and midsize sedan (.204 vs. .218). Taken together, the empirical evidence suggests that our decompositional approach is superior to its sales-only counterpart in capturing the dynamics of the underlying data-generating processes. Next, we present the parameter estimates of our proposed model and discuss their implications.

Advertising Impacts on Consumer Interest, Conversion, and Sales

Of central interest to us are five sets of parameters: δ^I_{j1} and δ^I_{j3} in Equation 6, which capture, respectively, the carryover and short-term impact of advertising on consumer interest; δ^C_{j1} and δ^C_{j3} in Equation 10, which capture the carryover and short-term impact of advertising on the baseline convertibility of consumer interest; and φ_j in Equation 3, which is expected to be negative and captures the rate of diminishing marginal convertibility of consumer interest. Equipped with estimates of these parameters (see Table 2, Panels A and B), we are in a position not only to derive the short- and long-term impacts of ad spend on sales but also to decompose them into the underlying interest- and conversion-related components. To facilitate comparison, we also report estimates of advertising effects based on the sales-only approach, δ^U_{j1} and δ^U_{j3}, which capture, respectively, the carry-

[4]As Hurvich, Shumway, and Tsai (1990) show, AIC_C is a better metric than AIC in correcting for the number of parameters in small-sample model comparisons.

Table 1
MODEL COMPARISON

	–2 Log-Likelihood[a]			*AIC*$_C$[a]			*Mean Absolute Error*[b]		
Segment/Vehicle	*Sales-Only*	*Decompositional*	*Δ*	*Sales-Only*	*Decompositional*	*Δ*	*Sales-Only*	*Decompositional*	*Δ*
Compact SUV									
Ford Escape	66.14	57.27	–8.87	86.07	79.66	–6.41	.136	.156	.019
Honda CR-V	61.85	45.75	–16.10	81.79	68.14	–13.64	.222	.144	–.078
Jeep Liberty	55.44	50.61	–4.83	75.37	73.00	–2.38	.342	.383	.041
Jeep Wrangler	66.26	52.03	–14.22	86.19	74.42	–11.77	.298	.150	–.147
Toyota RAV4	61.34	63.86	2.52	81.28	86.25	4.97	.248	.218	–.029
Segment summary	311.02	269.52	–41.50	410.70	381.48	–29.23	.249	.210	–.039
Midsize SUV									
Ford Explorer	147.24	48.07	–99.17	167.17	70.46	–96.71	.112	.113	.001
Honda Pilot	51.85	34.15	–17.70	71.79	56.54	–15.25	.180	.115	–.065
Hyundai Santa Fe	61.61	55.86	–5.75	81.55	78.26	–3.29	.158	.202	.044
Jeep Grand Cherokee	124.14	71.97	–52.17	144.08	94.36	–49.72	.167	.071	–.095
Toyota Highlander	134.01	9.18	–124.84	153.95	31.57	–122.38	.074	.125	.051
Segment summary	518.86	219.23	–299.63	618.54	331.19	–287.35	.138	.125	–.013
Compact Sedan									
Ford Focus	48.55	19.83	–28.71	68.48	42.23	–26.26	.277	.223	–.054
Honda Civic	50.47	17.15	–33.32	70.40	39.54	–30.86	.233	.204	–.029
Hyundai Elantra	316.50	106.60	–209.90	336.43	128.99	–207.44	.228	.244	.017
Toyota Corolla	83.45	14.18	–69.26	103.38	36.58	–66.81	.161	.218	.057
Toyota Prius	87.60	65.08	–22.53	107.54	87.47	–20.07	.478	.444	–.034
VW Jetta	91.90	62.46	–29.44	111.84	84.86	–26.98	.099	.075	–.025
Segment summary	678.46	285.31	–393.15	798.08	419.66	–378.41	.246	.235	–.011
Midsize Sedan									
Chevrolet Malibu	128.92	90.54	–38.37	148.85	112.93	–35.92	.314	.332	.018
Honda Accord	317.49	21.81	–295.68	337.43	44.21	–293.22	.198	.193	–.005
Hyundai Sonata	136.34	110.31	–26.04	156.28	132.70	–23.58	.216	.180	–.035
Nissan Altima	90.81	7.34	–83.47	110.75	29.73	–81.02	.229	.203	–.026
Toyota Camry	54.94	19.68	–35.26	74.88	42.07	–32.80	.131	.111	–.020
Segment summary	728.50	249.68	–478.82	828.18	361.64	–466.54	.218	.204	–.014

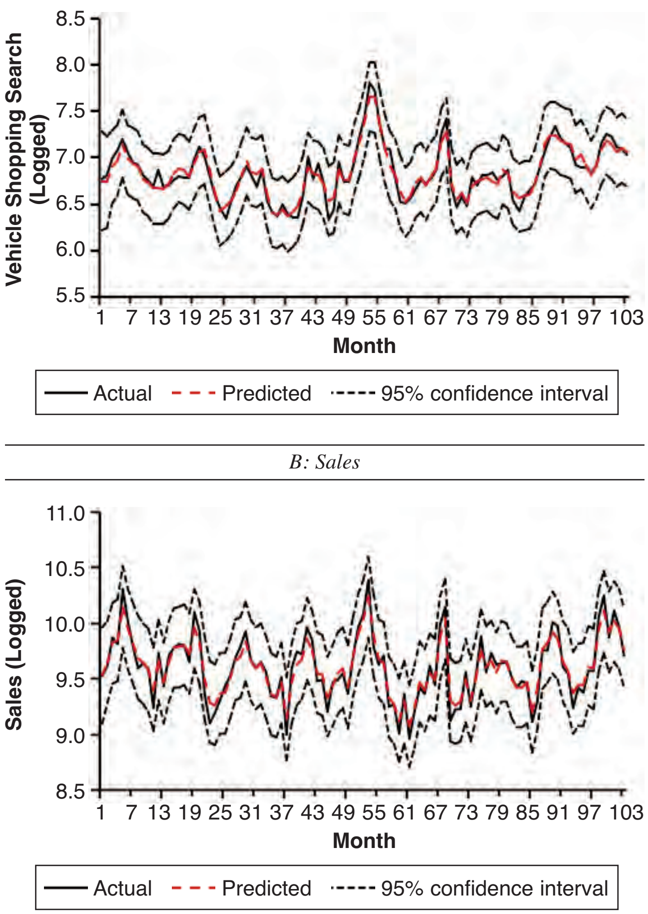
[a]In sample, the model with a smaller –2 log-likelihood or AIC$_C$ has better performance in goodness of fit.
[b]Out of sample, the model with the smaller mean absolute error has better performance in goodness of fit.

Figure 2

ACTUAL VERSUS. PREDICTED SEARCH AND SALES OF FORD FOCUS

A: Vehicle Shopping Search

B: Sales



over and short-term impact of advertising on sales (see Table 2, Panel C).

Short-Term Elasticities

Because both search data and ad spend are log-transformed in our model, δ^I_{j3} can be directly interpreted as the short-term ad elasticity of consumer shopping interest. Estimates for all 21 vehicles in our analysis are positive, 18 of which are significant ($p < .05$). We view this result as a sign of face validity, in support of the use of S_{jt}—the Google Trends indexes we have extracted using "Vehicle Shopping" as the category filter—as a proxy for consumer interest in prepurchase information search (otherwise, we might not observe such a strong and consistent pattern in the estimated effects of ad spend on S_{jt}).

We also compare δ^I_{j3} with δ^{NI}_{j3}, the short-term ad elasticity of nonshopping interest. For the latter, all 21 estimates are positive, of which only 7 are significant ($p < .05$). The mean of δ^{NI}_{j3}, .016, is significantly lower than that of δ^I_{j3}, .027 ($p < .01$). This finding suggests that in terms of short-term responsiveness to ad spend, shopping interest is much more elastic than nonshopping interest, which makes sense intuitively because ad spend is mainly intended to entice shopping-related interest. We view this finding as a sign of discriminant validity, which supports the way we constructed S_{jt} and G_{jt} and modeled their related and yet distinct dynamics (because the latter includes both shopping- and non-shopping-related searches).

Similar to δ^I_{j3}, δ^C_{j3} can be interpreted as the ad elasticity of baseline convertibility. All 21 estimates are positive and significant ($p < .05$). We interpret this strong and consistent pattern as another sign of face validity, in support of our decompositional approach and the use of Google search indexes as a proxy for consumer interest.

Comparing δ^I_{j3} with δ^C_{j3}, we note that their means (.027 vs. .030) are not significantly different ($p = .563$), indicating that consumer shopping interest and its baseline convertibility can be equally elastic to advertising. The correlation between δ^I_{j3} and δ^C_{j3}, across vehicles is .348 ($p < .05$), suggesting that the effectiveness of ad spend in generating interest is moderately tied to its effectiveness in boosting baseline convertibility. However, the lack of stronger correlation also highlights the importance of distinguishing the effectiveness of ad spend at the interest generation versus interest conversion stages of the purchase funnel, as opposed to lumping everything into a single metric.

To visualize the contrast between δ^I_{j3} and δ^C_{j3}, Figure 3 plots out the estimates, with the dotted lines representing the respective medians. For vehicles in the upper-right-hand quadrant (e.g., Accord), their ad spend has been relatively effective in the short run in both generating consumer interest and boosting baseline convertibility. The opposite can be said for vehicles in the lower-left-hand quadrant (e.g., Corolla). Although the ad spend for vehicles in the upper left-hand quadrant (Prius and Sonata) has been relatively effective in making interested consumers more likely to convert into buyers, it has been relatively ineffective in turning noninterested consumers into interested ones. The reverse is true for the vehicles in the lower-right-hand quadrant (Escape and Liberty): their ad spend is relatively effective in getting more consumers interested in the vehicle but relatively ineffective in making interested consumers more likely to convert into buyers. It is worth reiterating that such diagnosticity would be unattainable without a sales response model that can decompose the dynamics of the purchase funnel.

The parameter φ_j in Equation 3 captures how the overall conversion rate varies as a function of total consumer interest. We expect φ_j to be negative because not all interested consumers are created equal; some are bound to be intrinsically more convertible than others. All else being equal, consumers who are intrinsically more interested in the product are likely more convertible. As a result, when incremental ad spend makes more consumers become interested in and seek information about a product, those marginal information seekers should, on average, have lower convertibility than consumers who would seek product information anyway. In other words, the marginally interested consumers pulled in by the incremental ad spend dilute the average convertibility of the total pool of interested consumers. From Table 2, Panel A, we observe that, of the 21 estimates of φ_j, all are negative and 18 are significant ($p < .05$), consistent with our expectation. For vehicles with negative and large φ_j (in absolute value), such as Explorer, consumer interest generated through incremental ad spend

Table 2
ESTIMATES OF STATE EQUATION PARAMETERS (δ)

A: Decompositional Model: Shopping Interest and Nonshopping Interest									
	Shopping Interest					*Nonshopping Interest*			
Segment/Vehicle	*Carryover* (δ^{I}_{j1})	*Spillover* (δ^{I}_{j2})	*Ad* (δ^{I}_{j3})	*Competing Ad* (δ^{I}_{j4})	φ_j	*Carryover* (δ^{NI}_{j1})	*Spillover* (δ^{NI}_{j2})	*Ad* (δ^{NI}_{j3})	*Competing Ad* (δ^{NI}_{j4})
Compact SUV									
Ford Escape	**.663**	.055	**.038**	.064	**–.689**	.406	.123	.010	–.026
Honda CR-V	**.611**	.072	**.061**	.092	**–.783**	.412	.099	.015	.033
Jeep Liberty	.368	**.128**	**.029**	.022	**–.688**	.242	.093	.000	.003
Jeep Wrangler	**.759**	.041	**.014**	–.002	**–.669**	**.694**	.076	.017	.072
Toyota RAV4	**.743**	.044	**.026**	.018	**–.764**	.244	.110	**.023**	–.007
Segment mean	**.629**	.068	**.034**	.038	**–.718**	.400	.100	.013	.015
Midsize SUV									
Ford Explorer	**.487**	**.103**	**.031**	–.015	**–.917**	.213	.083	.017	.081
Honda Pilot	**.644**	**.084**	**.024**	.049	**–.915**	.226	.085	.012	–.107
Hyundai Santa Fe	**.803**	.062	.000	.007	**–.952**	**.860**	.048	.007	–.003
Jeep Grand Cherokee	**.553**	**.070**	**.030**	–.009	**–.931**	**.438**	.069	.008	**.111**
Toyota Highlander	**.813**	.044	**.014**	.017	**–.912**	.275	.088	.011	–.061
Segment mean	**.660**	.072	.020	.010	**–.925**	.402	.075	.011	.004
Compact Sedan									
Ford Focus	.290	**.104**	**.050**	**.112**	–.331	.257	.098	.015	.030
Honda Civic	.500	**.084**	**.016**	.002	–.362	.242	.067	**.020**	.098
Hyundai Elantra	**.745**	.012	**.013**	**.138**	**–.249**	**.411**	.073	**.026**	.061
Toyota Corolla	**.615**	.079	.014	–.002	–.363	.259	.084	**.040**	.058
Toyota Prius	**.883**	–.029	**.010**	**.109**	**–.728**	**.393**	.038	**.028**	.020
VW Jetta	.374	**.129**	**.024**	.017	**–.855**	.205	.085	.008	.103
Segment mean	**.568**	.063	.021	.063	–.481	.295	.074	.023	.062
Midsize Sedan									
Chevrolet Malibu	**.506**	.073	**.053**	.028	**–.754**	.380	.069	**.021**	.126
Honda Accord	.563	.094	**.052**	.016	**–.610**	.425	.090	.015	.016
Hyundai Sonata	.377	.091	.008	.028	**–.865**	**.651**	.083	.013	.063
Nissan Altima	**.578**	.042	**.050**	.092	**–.745**	.318	.087	.001	–.091
Toyota Camry	**.828**	.006	**.016**	.079	**–.574**	.351	.085	**.034**	.070
Segment mean	**.570**	.061	**.036**	.049	**–.710**	.425	.083	.017	.037

Table 2

CONTINUED

B: Decompositional Model: Conversion and Advertising

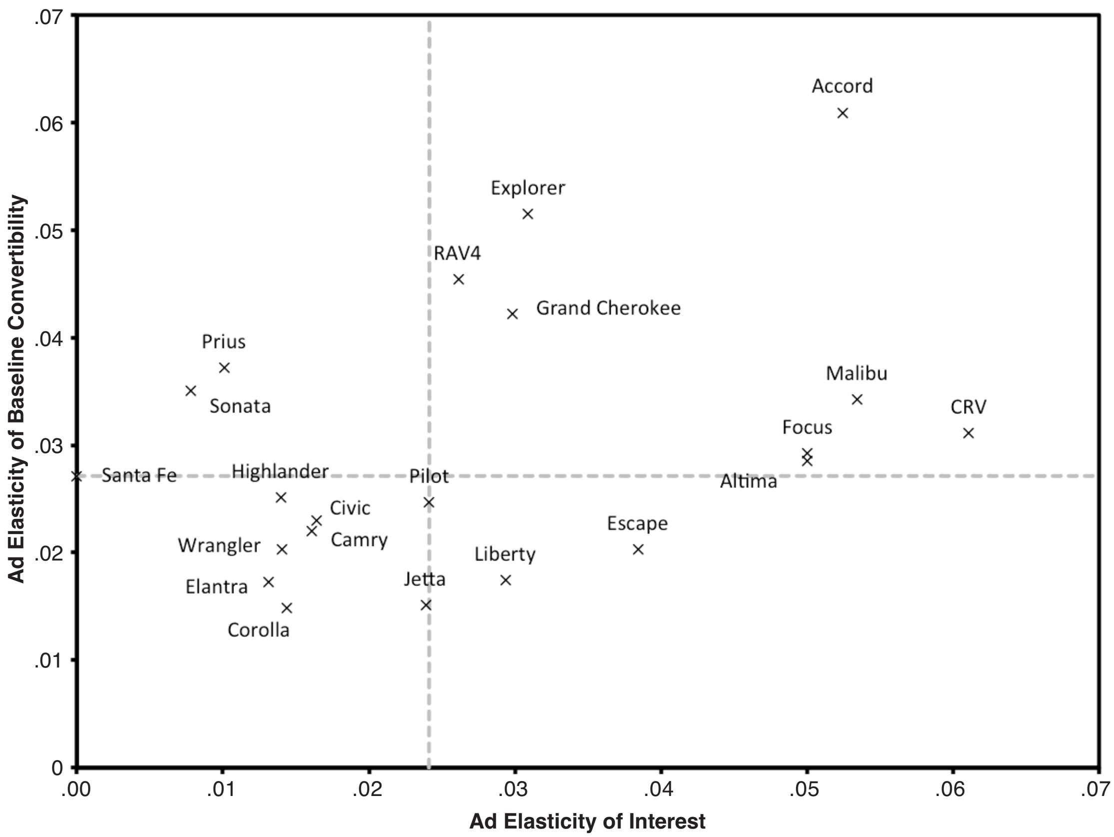
	Conversion				*Advertising (Endogeneity)*		
Segment/Vehicle	*Carryover* (δ^{C}_{j1})	*Spillover* (δ^{C}_{j2})	*Ad* (δ^{C}_{j3})	*Competing Ad* (δ^{C}_{j4})	*Carryover* (δ^{A}_{j1})	*Lagged Sales* (δ^{A}_{j2})	*Lagged Competing Ad* (δ^{A}_{j3})
Compact SUV							
Ford Escape	**.791**	.051	**.020**	.037	**.400**	**.387**	–.125
Honda CR-V	**.666**	.064	**.031**	–.027	**.566**	–.040	**–.312**
Jeep Liberty	**.793**	–.030	**.017**	–.002	**.518**	**.538**	**–.413**
Jeep Wrangler	**.753**	**.082**	**.020**	.033	**.432**	**.343**	**–.291**
Toyota RAV4	**.894**	–.009	**.045**	**–.059**	.468	–.072	–.020
Segment mean	**.779**	.032	**.027**	–.004	**.477**	.231	–.232
Midsize SUV							
Ford Explorer	**.706**	**.127**	**.051**	–.033	**.698**	**.334**	**–.356**
Honda Pilot	**.539**	.028	**.025**	–.032	.423	**–.188**	**.199**
Hyundai Santa Fe	**.721**	–.012	**.027**	**–.033**	**.626**	**.385**	**–.329**
Jeep Grand Cherokee	**.819**	.031	**.042**	–.017	**.693**	**.209**	**–.078**
Toyota Highlander	.278	**.056**	**.025**	–.070	**.587**	.116	–.002
Segment mean	**.613**	.046	**.034**	–.037	**.605**	.171	–.113
Compact Sedan							
Ford Focus	**.382**	.070	**.029**	–.022	**.677**	**–.180**	**.337**
Honda Civic	**.409**	.044	**.023**	.037	**.452**	**–.190**	**.342**
Hyundai Elantra	**.750**	–.020	**.017**	**–.053**	**.366**	**–.411**	**.234**
Toyota Corolla	**.383**	.045	**.015**	–.047	**.662**	**.372**	**.161**
Toyota Prius	**.852**	–.031	**.037**	–.017	**.668**	**–.282**	**.097**
VW Jetta	**.717**	.015	**.015**	.034	**.388**	**–.699**	**.677**
Segment mean	**.582**	.021	**.023**	–.012	**.535**	–.232	.308
Midsize Sedan							
Chevrolet Malibu	**.707**	.098	**.034**	**–.047**	**.794**	**–.842**	.029
Honda Accord	**.465**	–.005	**.061**	–.084	.428	**–.192**	.116
Hyundai Sonata	**.756**	.067	**.035**	**.067**	.321	.028	**.228**
Nissan Altima	.126	–.018	**.029**	.007	.162	.260	–.129
Toyota Camry	.340	–.036	**.022**	**–.101**	.424	–.019	–.008
Segment mean	**.479**	.021	**.036**	–.032	.426	–.153	.047

Table 2
CONTINUED

C: Sales-Only Model							
	Sales				*Advertising (Endogeneity)*		
Segment/Vehicle	*Carryover* (δ^U_{j1})	*Spillover* (δ^U_{j2})	*Ad* (δ^U_{j3})	*Competing Ad* (δ^U_{j4})	*Carryover* (δ^A_{j1})	*Lagged Sales* (δ^A_{j2})	*Lagged Competing Ad* (δ^A_{j3})
Compact SUV							
Ford Escape	**.771**	**.068**	**.020**	–.030	**.399**	**.372**	–.131
Honda CR-V	**.770**	.022	**.046**	.038	**.567**	–.045	**–.315**
Jeep Liberty	**.810**	–.019	**.022**	–.018	**.517**	**.539**	**–.414**
Jeep Wrangler	**.782**	**.102**	**.015**	**–.056**	**.428**	**.345**	**–.292**
Toyota RAV4	**.906**	–.009	**.045**	**–.070**	.466	–.067	–.015
Segment mean	**.808**	.033	**.030**	–.027	**.475**	.229	–.233
Midsize SUV							
Ford Explorer	**.785**	.063	**.051**	.048	**.701**	**.323**	**–.361**
Honda Pilot	**.536**	.028	**.027**	–.035	.425	**–.193**	**.199**
Hyundai Santa Fe	**.746**	–.015	**.022**	–.023	**.623**	**.387**	**–.331**
Jeep Grand Cherokee	**.880**	.009	**.033**	**–.076**	**.697**	**.179**	**–.096**
Toyota Highlander	.355	**.047**	**.027**	–.008	**.591**	.123	.001
Segment mean	**.660**	.026	**.032**	–.019	**.607**	.164	–.118
Compact Sedan							
Ford Focus	**.472**	.051	**.054**	–.020	**.678**	**–.176**	**.338**
Honda Civic	**.670**	.018	**.024**	.019	**.455**	**–.186**	**.343**
Hyundai Elantra	**.803**	.041	**.021**	.067	**.370**	**–.403**	**.241**
Toyota Corolla	**.902**	–.018	**.011**	**.065**	**.655**	**.348**	**.152**
Toyota Prius	**.815**	–.031	**.038**	–.003	**.654**	**–.232**	**.149**
VW Jetta	**.773**	–.011	**.015**	.033	**.387**	**–.712**	**.674**
Segment mean	**.739**	.008	**.027**	.027	**.533**	–.227	**.316**
Midsize Sedan							
Chevrolet Malibu	**.773**	.019	**.036**	–.088	**.793**	**–.859**	.018
Honda Accord	**.590**	.012	**.079**	–.056	.425	**–.165**	.135
Hyundai Sonata	**.775**	.017	**.032**	.051	.323	.034	**.233**
Nissan Altima	.244	–.009	**.053**	.080	.155	.262	–.130
Toyota Camry	**.606**	–.034	**.022**	.055	.426	–.004	.000
Segment mean	**.598**	.001	**.045**	.008	.424	–.146	.051

Notes: Boldfaced figures indicate that the 95% posterior confidence interval excludes zero ($p < .05$).

Figure 3
SHORT-TERM AD ELASTICITIES: INTEREST VERSUS BASELINE CONVERTIBILITY



Notes: Correlation = .348. The dotted lines represent median splits on either axis.

exhibits high diminishing returns. For vehicles with negative and small φ_j (in absolute value), such as Elantra, the reverse is true. We note that diagnostic insight such as this would be lost in a sales-only model.

Finally, recall Equations 15 and 16. Given the estimates of δ^I_{j3}, δ^C_{j3}, and φ^*_j ($\equiv 1 + \varphi_j$), we can calculate the short-term ad elasticity of sales as $\partial \ln(Y_{jt})/\partial \ln(A_{jt}) = \varphi^*_j\delta^I_{j3} + \delta^C_{j3}$. With the more conventional sales-only model (recall Equations 13 and 14), the short-term ad elasticity of sales is given by δ^U_{j3}. As the comparison in Table 3 shows, the sales-only model produces smaller estimates of short-term ad elasticity in 17 cases, and across the 21 vehicles, the mean of $\varphi^*_j\delta^I_{j3} + \delta^C_{j3}$(.038) is statistically greater than the mean of δ^U_{j3}(.033) ($p <$ .01). This finding suggests that, on average, the sales-only approach tends to underestimate short-term ad elasticity. The risk of underestimation aside, δ^U_{j3} is highly correlated with $\varphi^*_j\delta^I_{j3} + \delta^C_{j3}$ (r = .938, $p < .01$), which we take as an indirect sign of convergent validity, suggesting that our more complex model is robust (it did not produce elasticity estimates that drastically differ from the simpler, more established sales-only approach). Again, note that the sales-only model is less diagnostic because it lacks the capacity to decompose overall short-term ad elasticity of sales into the underlying components related to interest and conversion, in which we have observed substantial differences.

Long-Term Elasticities

The carryover effects δ^I_{j1} (M = .605) and δ^C_{j1} (M = .612) are all estimated to be positive and are significant ($p < .05$) in 15 and 18 of the 21 vehicles, respectively. This finding suggests a sizable inertia in both consumer interest and convertibility. Furthermore, the mean of δ^I_{j1} is not significantly different from that of δ^C_{j1}, and the half-life of advertising impact on consumer interest (M = 1.38 months) is similar to that on convertibility (M = 1.41 months). However, across the 21 vehicles, the two sets of carryover effects are largely uncorrelated (r = .035, $p = .909$), another strong indication that consumer interest and conversion follow distinct dynamics and are best modeled separately.

We note that, on average, both δ^I_{j1} (M = .605) and δ^C_{j1} (M = .612) are significantly ($p < .05$) smaller than δ^U_{j1} (M = .703), the carryover effect based on the sales-only model. It is worthwhile to contrast this finding with the earlier finding that $\varphi^*_j\delta^I_{j3} + \delta^C_{j3}$, the short-term ad elasticity of sales based on our model, is on average significantly greater than δ^U_{j3}, the short-term ad elasticity of sales based on the sales-only model. In other words, our model comparison shows that the sales-only approach can lead to systematic biases in two ways: first, it tends to underestimate short-term ad elasticity of sales, and second, it tends to overestimate the carryover effect. These biases can potentially lead to suboptimal temporal allocation of ad spend.

Table 3

SHORT-TERM AD ELASTICITIES OF SALES: SALES-ONLY MODEL VERSUS DECOMPOSITIONAL MODEL

	Short-Term Elasticity			
Segment/Vehicle	*Sales-Only*	*Decompositional*	*Δ*	*Δ%*
Compact SUV				
Ford Escape	.020	.032	.012	38%
Honda CR-V	.046	.044	–.001	–3%
Jeep Liberty	.022	.027	.004	16%
Jeep Wrangler	.015	.025	.010	39%
Toyota RAV4	.045	.052	.007	14%
Segment mean	.030	.036	.006	21%
Midsize SUV				
Ford Explorer	.051	.050	–.001	–3%
Honda Pilot	.027	.029	.002	8%
Hyundai Santa Fe	.022	.024	.002	8%
Jeep Grand Cherokee	.033	.044	.011	25%
Toyota Highlander	.027	.027	.000	–1%
Segment mean	.032	.035	.003	7%
Compact Sedan				
Ford Focus	.054	.063	.009	14%
Honda Civic	.024	.033	.009	28%
Hyundai Elantra	.021	.033	.012	37%
Toyota Corolla	.011	.013	.002	15%
Toyota Prius	.038	.040	.002	5%
VW Jetta	.015	.019	.003	18%
Segment mean	.027	.033	.006	19%
Midsize Sedan				
Chevrolet Malibu	.036	.047	.011	24%
Honda Accord	.079	.081	.002	3%
Hyundai Sonata	.032	.036	.004	11%
Tissan Altima	.053	.041	–.012	–28%
Toyota Camry	.022	.029	.006	22%
Segment mean	.045	.047	.002	6%

Notes: Correlation = .938. The mean elasticity based on the sales-only model is .033, which is significantly ($p < .01$) smaller than the mean elasticity based on the decompositional model (.038).

To gauge the long-term elasticity of ad spend on sales, we need to combine the model components that determine the short-term effects (δ^I_{j3}, δ^C_{j3}, and φ_j) with the carryover effects (δ^I_{j1} and δ^C_{j1}). Unfortunately, unlike short-term elasticity, there is no simple closed-form solution for our model. Instead, following Ataman, Mela, and Van Heerde (2008), we use a policy simulation to quantify the long-term impact of ad spend based on our decompositional model and compare that result with its counterpart based on the sales-only model. We first use both models to estimate each vehicle's sales within a 12-month window, setting all the control variables and ad spend to their historical averages. These estimates serve as the base (S_0). We then increase ad spend by 10% and calculate the new sales estimates (S_1), which enables us to calculate long-term ad impact on sales, $[(S_1/S_0) - 1]/10\% \equiv \Delta$. Figure 4 plots the Δs from our decompositional model and the sales-only model.

Like the short-term elasticity estimates, the correlation between the two models' long-term impact estimates remains high ($r = .801$), suggesting that our model is robust (it did not produce estimates that drastically differ from the simpler, more established sales-only model). However, unlike short-term elasticities, for which the sales-only model's estimates are systematically smaller than our model's (M = .033 vs. M = .038, difference = –.005, $p < .05$), for long-term impacts, the sales-only model underestimates in 12 cases (M = .088 vs. M = .121, difference = –.034, $p < .05$) and overestimates in 9 cases (M = .135 vs. M = .097; difference = .038, $p < .05$). Across the 21 vehicles, the two models' average long-term impact estimates are not statistically different (M = .108 vs. M = .111, difference = –.003, $p = .75$). Together, the contrasts between the elasticity estimates of the two models suggest that by relying on the sales-only approach, firms risk (1) underspending in the short run and (2) over- as well as underspending in the long run.

Endogeneity

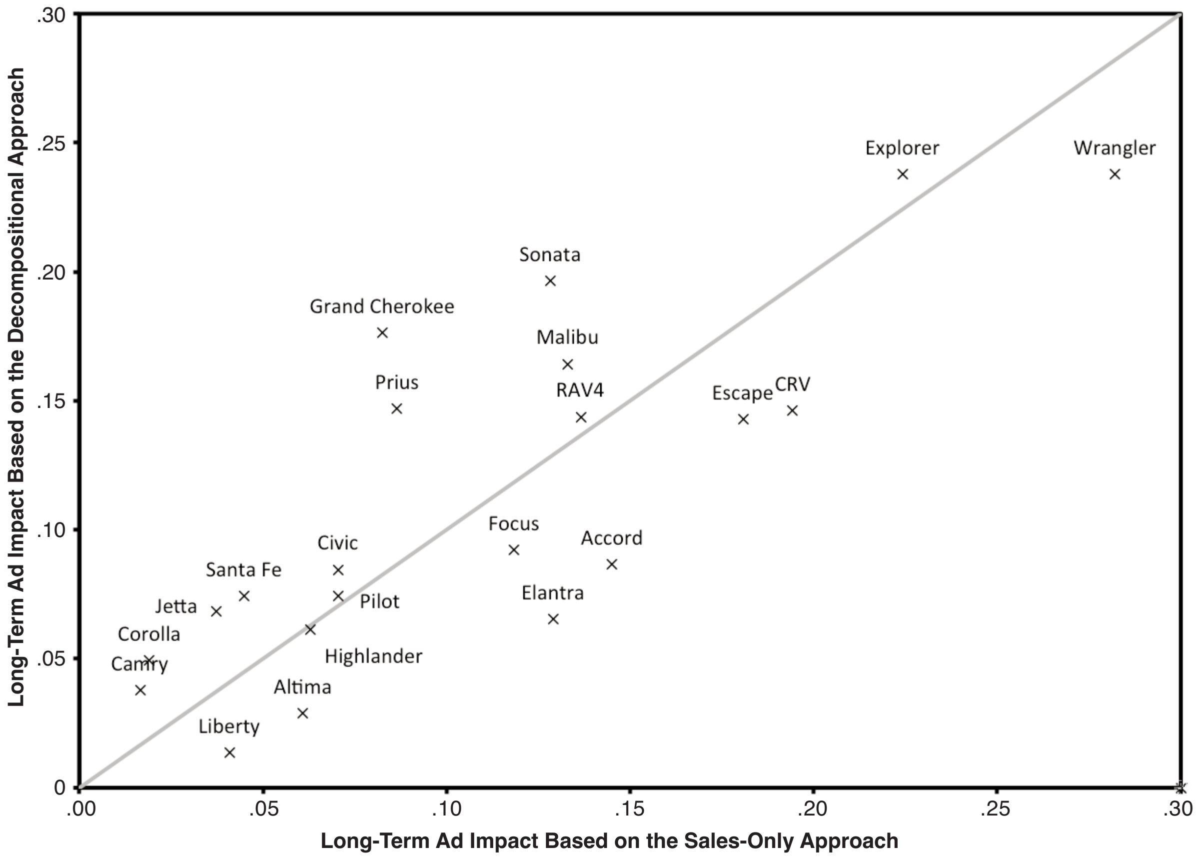
Table 2, Panel B, also reports the parameter estimates from Equation 18, where δ^A_{j1} captures the carryover in ad spend, δ^A_{j2} captures the influence of lagged sales on current ad spend, and δ^A_{j3} captures the influence of lagged competitive ad spend on current own ad spend. We observe that estimates of δ^A_{j1} are all positive, are significant in 15 cases, and have a mean of .512, indicating moderate inertia in ad spend from one month to another. There is an intriguing SUV versus sedan divergence in δ^A_{j2}, which captures how lagged sales affects current ad spend. For SUVs (both compact and midsize), in general, higher lagged sales lead to higher current ad spend (6 of 10 are significantly positive; only 1 is significantly negative). For sedans, higher lagged sales, in general, lead to lower current ad spend (7 of 11 are significantly negative; only 1 is significantly positive). This result seems to indicate that automakers make ad spend decisions differently for different vehicle body types. For cars with higher profit margins (e.g., SUVs), ad spend is strengthened when sales are strong; for less profitable cars (e.g., sedans), the strategy seems to be less ad spend when the vehicles are selling well. Similar SUV versus sedan divergence occurs in δ^A_{j3}, the effect of lagged competitive ad spend on current own ad spend. It tends to be negative among SUVs (significant in 7 of 10 cases; 6 of the 7 significant cases are negative) but positive among sedans (significant in 7 of 11 cases; all significant cases are positive). This result seems to suggest that automakers attempt to avoid fierce ad spend wars on SUVs by toning down their own ad spend in response to increased competitive spend. In contrast, for sedans, automakers increase own ad spend in response to increased competitive ad spend.

Effects of Environmental Variables, Competitive Ad Spend, and Spillovers

Table 4 reports the β estimates in Equation 17, which capture the impacts of environmental variables on consumer shopping and nonshopping interests (Panel A) and convertibility of shopping interest and ad spend (Panel B). On the one hand, we note that having accounted for carryovers and the effects of ad spend, most of the environmental controls are insignificant. The only exception is sentiments on shopping interest, many of which are negative and significant. A possible explanation is that low sentiment is accompanied by high uncertainty, which makes consumers more cautious and thus more likely to conduct extensive prepurchase information search. On the other hand, we note that the effects of sentiment on conversion are mostly positive (i.e., the more optimistic consumers are, the more likely they are to "pull the trigger" after seeking product information). However, these effects are all insignificant, except those for the Toyota Prius.

Figure 4

LONG-TERM AD IMPACT ON SALES: DECOMPOSITIONAL VERSUS SALES-ONLY MODELS



Notes: Correlation = .801. For vehicles above the 45° line, the estimated long-term impact of ad spend on sales from the decompositional model is higher than from the sales-only model.

In terms of gas price, it is reassuring to note that, for the Toyota Prius, a compact sedan known for its fuel efficiency, the effects on shopping interest and ad spend are both positive and significant (i.e., the higher the gas price, the more consumers are interested in Toyota Prius, which also receives more ad spend). In terms of seasonality, all the effects on nonshopping interest and convertibility are insignificant. The effects on shopping interest are significant and positive in only four cases. The effects on ad spend are significant and positive in ten cases. In terms of lagged sales, we observe no significant effects on current consumer interest, shopping or nonshopping, which suggests a negligible role of searches driven by lagged sales compared with searches driven by prepurchase information interest.

Finally, in terms of competitive effects, our model allows for two types. The first type arises from competing ad spend (δ^{I}_{j4}, δ^{NI}_{j4}, and δ^{C}_{j4} in Table 2). We find that the competitive effects of ad spend on shopping interest are mostly insignificant, except for three compact sedans, for which the effects are all positive. The effects of competing ad spend on conversion are significant for six vehicles, five of which are negative, indicating that increased competing ad spend tends to make interested consumers less likely to buy the focal product. The second type of competitive effects arises from interest or conversion spillovers (δ^{I}_{j2}, δ^{NI}_{j2}, and δ^{C}_{j2} in Table 2), which are mostly insignificant; however, they are all positive when they are significant, suggesting that sporadic spillover could exist among vehicles within the same segment.

CONCLUDING REMARKS

In contexts in which prepurchase information search is the norm, advertising can drive sales first by making consumers interested in seeking information about a product and then by converting information seekers into buyers. Taking such a rudimentary view of the purchase funnel, we propose a modeling framework that decomposes the overall impact of advertising on sales into interest generation versus conversion. Such decomposition is made possible by augmenting sales data with search volume indexes gathered from Google Trends, which we treat as a tracking device of consumer prepurchase information interest. To illustrate, we apply our proposed modeling framework to the new passenger vehicle market in the United States, covering 21 top-selling models from four major segments over a period of 103 months.

Our empirical analyses have led to several intriguing observations regarding the distinct dynamics of consumer interest generation versus interest conversion. We find that, on average, consumer interest and its baseline convertibility

Table 4

ENVIRONMENTAL CONTROLS IN THE DECOMPOSITIONAL MODEL (β)

A: Shopping Interest and Nonshopping Interest

	Shopping Interest				*Nonshopping Interest*			
Segment/Vehicle	*Sentiment*	*Gas*	*Seasonality*	*Lagged Sales*	*Sentiment*	*Gas*	*Seasonality*	*Lagged Sales*
Compact SUV								
Ford Escape	**–1.373**	.129	1.081	.027	.413	.072	.207	–.013
Honda CR-V	**–1.207**	–.086	.704	–.035	.578	.290	.143	.019
Jeep Liberty	**–1.058**	.221	.796	.023	.492	–.413	.138	.002
Jeep Wrangler	–.919	.904	.482	.004	.036	–1.101	.367	.014
Toyota RAV4	–1.041	.276	.658	–.032	.265	–.031	.184	–.001
Segment mean	–1.121	.289	.744	.004	.355	–.239	.208	–.013
Midsize SUV								
Ford Explorer	**–.741**	.476	.573	–.035	.154	–.567	.148	.021
Honda Pilot	**–1.116**	–.037	.529	–.037	.346	.005	.104	.006
Hyundai Santa Fe	**–1.461**	.600	.141	–.024	.221	–.353	.271	–.012
Jeep Grand Cherokee	–.736	.271	**.745**	.017	.286	–.436	.171	–.050
Toyota Highlander	**–1.128**	–.322	.780	–.057	.216	.257	.255	.021
Segment mean	–1.033	.201	.542	–.011	.243	–.215	.190	–.032
Compact Sedan								
Ford Focus	**–1.449**	.418	.691	–.017	.495	–.203	.092	.000
Honda Civic	**–1.534**	**.841**	**.967**	–.003	.456	–.932	.119	–.002
Hyundai Elantra	**–1.146**	.454	.425	–.001	–.094	–.050	.117	.026
Toyota Corolla	**–1.697**	.468	.720	–.020	.415	–.579	.077	–.001
Toyota Prius	**–2.163**	**1.418**	**.737**	.011	.126	–.309	.139	–.017
VW Jetta	**–1.469**	.099	.531	–.078	.272	–.095	.121	.001
Segment mean	**–1.578**	.616	.677	–.041	.279	–.366	.111	.034
Midsize Sedan								
Chevrolet Malibu	**–.889**	.538	.896	–.018	.062	–.355	.087	–.038
Honda Accord	–.825	.485	.828	–.055	.531	–.753	.326	.046
Hyundai Sonata	–.302	–.025	.653	–.038	–.172	–.216	.357	.020
Nissan Altima	**–.975**	.811	**1.037**	.012	.602	–.324	.173	.025
Toyota Camry	**–1.060**	.503	.884	–.005	.040	–.520	.197	–.020
Segment mean	–.800	.460	.850	–.049	.202	–.430	.228	.041

B: Conversion and Advertising

	Conversion			*Advertising (Endogeneity)*		
Segment/Vehicle	*Sentiment*	*Gas*	*Seasonality*	*Sentiment*	*Gas*	*Seasonality*
Compact SUV						
Ford Escape	.468	–.379	–.030	**1.095**	–.582	–.010
Honda CR-V	.741	.183	–.019	**1.840**	1.254	**.469**
Jeep Liberty	1.064	–.547	–.013	**2.143**	–.876	–.206
Jeep Wrangler	.195	**–1.221**	–.030	**1.550**	–.227	.040
Toyota RAV4	.835	–.104	–.022	.508	.635	**.471**
Segment mean	.718	–.397	–.023	**1.272**	.035	.135
Midsize SUV						
Ford Explorer	.917	**–.968**	.040	–.076	–.191	**.787**
Honda Pilot	.738	–.194	.025	.758	.530	.200
Hyundai Santa Fe	.892	–.738	–.033	1.005	–.823	.337
Jeep Grand Cherokee	.384	–.835	.031	–.465	.085	.459
Toyota Highlander	.707	.043	.011	.697	–.112	.144
Segment mean	.766	–.483	.015	.353	–.093	.311
Compact Sedan						
Ford Focus	.538	–.596	.040	**–1.625**	.579	**.651**
Honda Civic	.478	**–1.027**	.033	.160	.538	.242
Hyundai Elantra	–.365	–.175	.038	**–.781**	**2.001**	**.714**
Toyota Corolla	.728	–.631	.058	**–1.601**	–.238	–.042
Toyota Prius	**1.283**	–.906	.036	**–1.417**	**1.856**	**1.205**
VW Jetta	.888	.442	.040	.499	.107	.302
Segment mean	.691	–.493	.041	**–.765**	.728	.477
Midsize Sedan						
Chevrolet Malibu	1.055	–.298	–.027	–2.169	1.847	**4.057**
Honda Accord	.278	–.642	.005	–.046	**.856**	**.684**
Hyundai Sonata	–.596	.367	–.011	.303	.736	.187
Nissan Altima	.535	–.448	–.031	–.377	–.265	**.729**
Toyota Camry	.365	–.431	–.002	–.313	.992	**.727**
Segment mean	.376	–.320	–.013	–.494	.778	**1.254**

Notes: Boldfaced figures indicate that the 95% posterior confidence interval excludes zero ($p < .05$).

can be equally elastic to advertising in the short run (.027 vs. .030) and have comparable carryovers from month to month (.605 vs. .612). However, across vehicles, there is weak correlation (r = .348) between the short-term elasticities of interest and convertibility and no correlation (r = .035) between their carryovers. Such lack of strong correlation highlights the importance of separately evaluating the impacts of ad spend at different stages of the purchase funnel rather than focusing only on the total impact on sales. In our sample, we find many cases in which advertising is relatively effective in either interest generation or interest conversion, but not both. We also find many cases in which the marginal convertibility of consumer interest declines quickly, suggesting high diminishing returns on interest generation. Overall, by augmenting sales data with search data and adopting the proposed decompositional model, we have obtained many novel and more diagnostic insights about the impacts of advertising that would otherwise be unattainable.

By benchmarking our sales-and-search approach against its sales-only counterpart, we find that our approach can lead to not only better in- and out-of-sample fit but also different estimates of sales elasticity. If only sales data are considered in gauging the overall effectiveness of ad spend, marketers are likely to underestimate its short-term impact while overestimating how long that impact may linger. As for the long-term impact of ad spend on sales, there is a significant risk of both underestimation and overestimation. By augmenting sales with search data and adopting our decompositional model, these biases can be corrected, which can potentially improve ad spend decisions in terms of temporal allocation as well as setting the total budget.

A key challenge faced by our decompositional approach lies in whether it can approximate shifts in consumers' prepurchase information interest through changes in how frequently they Google the focal product. Three major threats exist. The first is that not all Google searches are shopping related, which could make Google Trends indexes a biased proxy for consumer prepurchase information interest (unless the ratio between shopping- and non-shopping-related searches remains largely stable over time).[5] In our empirical application, we take a three-pronged approach to minimize this threat. First, we use composite queries that exclude terms that are most likely unrelated to new vehicle shopping (e.g., "parts," "recall," "used"). Second, we use the "Vehicle Shopping" subcategory filter provided by Google Trends to generate search indexes that are most likely related to vehicle shopping. Finally, we use the "Autos & Vehicles" category filter to generate Google Trends indexes that cover all vehicle-related searches, which, combined with the "Vehicle Shopping" indexes, enables us to better distinguish the dynamics of shopping-related searches from those of non-shopping-related searches.

The second major threat to the validity of using Google Trends data as a proxy for consumer prepurchase information interest is that not all information-seeking consumers use Google. There are consumers who do not use the Internet when gathering product information. Furthermore, among consumers who do use the Internet, some use other search engines or visit automaker, dealer, and third-party automotive websites directly. In short, if a considerable portion of consumers are non–Google users and their prepurchase information interest differs substantially from that of Google users, the trend lines of Google searches will not run parallel to the trend lines of overall consumer interest.

In this study, we do not have a direct solution to address this second threat except to note that when buying a new car, the majority of U.S. consumers use the Internet to gather product information and rely on search engines as a gateway, an area that Google has dominated. In addition, we find it reassuring that, in many aspects, our model estimates have exhibited strong face, convergent, and discriminant validity. Finally, both in and out of sample, our model outperforms a benchmark model that considers sales data alone. Taken together, these factors should provide enough indirect evidence, at least in the empirical setting of the current study, to alleviate the concern that the search interest of Google users may not be representative of the general population.

More generally speaking, whether Google users' interests are reflective of those of the general population is ultimately a sample representativeness issue that must be determined empirically case by case. We note that sample representativeness issues also exist with survey data whenever the sampling frame does not cover the entire target population. For example, data collected from online panels may not be representative of the general population if there are significant differences between panel participants and nonparticipants or between respondents and nonrespondents.

Finally, the third major threat to the validity of using Google Trends data as a proxy for consumer prepurchase information interest lies in the possibility that consumers may conduct postpurchase information search (e.g., post hoc price checking, seeking operating information). The severity of such reverse causality depends on the ratio between pre- and postpurchase information search; the lower the ratio, the greater the threat. We speculate that prepurchase information search should dominate postpurchase information search. When conducting prepurchase information search, consumers are likely to examine many products that are in their consideration set. In contrast, when conducting postpurchase information search, consumers are likely to focus mainly on the product they have just purchased.

Empirically, the finding that our sales-and-search model outperforms the sales-only model in the out-of-sample comparison should alleviate the threat of reverse causality. If postpurchase information search played a dominant role, search data would contain little incremental information that is not already contained in sales data, which would render search data of little incremental value in out-of-sample forecasting. Furthermore, we find that lagged sales have no significant effects on current (shopping or nonshopping) search interests, which suggests a negligible role of searches caused by lagged sales. That said, our model cannot rule out reverse causality when purchase and postpurchase information search take place within the same month. To address this issue directly, researchers would probably need individual-level tracking data on searches and purchases or aggregate

[5]In general, as long as the ratio between Google search volume for any product and the level of prepurchase information interest in the population remains stable over time, Google Trends indexes are still a valid proxy of consumer interest because they can be mapped into each other by multiplying a scaling constant. In other words, bias will not be an issue as long as the trend lines of Google searches and consumer interest can be assumed to run largely parallel to each other.

tracking data that are available at higher frequencies (e.g., daily, weekly).

Directions for Extension

We imagine several promising avenues for further research. First, we developed our proposed modeling framework in line with the premise that actively seeking product information is an integral behavioral predecessor to purchase decisions. Such a premise should be valid for new-vehicle shopping because the majority of consumers are highly motivated and do conduct prepurchase information search. It would be worthwhile to extend our modeling framework to include contexts in which a nonnegligible portion of consumers would make a purchase without first seeking product information. To apply our decompositional model in such contexts, researchers would need to augment sales data with a proxy that can track a stage of the purchase funnel that all consumers would go through before making a purchase decision. For example, in contexts in which forming a consideration set is an "unskippable" stage of the purchase funnel, researchers could track how many consumers have included the focal product in their consideration set. By augmenting sales data with such a tracking measure and applying our modeling framework, sales can be decomposed into "consideration generation" versus "consideration conversion." Of course, the challenge lies in finding a reliable, timely, and cost-effective instrument for tracking consumer consideration.

In addition to using Google search data as a proxy for consumer information interest, we see the potential of tapping into other sources of online tracking data (e.g., website traffic, likes on Facebook, followers on Twitter, requests for price quotes).[6] Conceivably, these data sources may track consumer interest at different stages of the purchase funnel. For example, the number of Facebook likes or Twitter followers may track consumers' awareness and general impression of a product. The number of online requests for price quotes may track the number of consumers reaching a stage between initial information gathering and final purchase. In short, as these online tracking data become increasingly available, marketers can better leverage them by treating them as proxies for various stages of the purchase funnel and investigating the conversion between them and sales (potentially allowing for multiple pathways of conversion).

It would also be worthwhile to determine how various types of ad spend may affect consumer interest and conversion differently. For example, it is plausible that ad spend focused on brand building is more effective in generating consumer interest than ad spend focused on deal promotion, whereas the reverse might be true when converting interested consumers into buyers. One might also reasonably speculate that ad spend online and on social media could be relatively more effective in generating searches, whereas television and print ad spend could be relatively more effective in generating sales.

[6]In Web Appendix B, we present a pattern of positive and consistent correlation between vehicle sales, Google searches, online price quote requests, and website visits, which suggests the existence of a common underlying driver (i.e., the level of consumer prepurchase information interest).

Finally, the current study can be extended to include other marketing-mix variables. For example, compared with advertising, incentives and discounts may prove more effective in conversion than in interest generation. Our modeling framework can be readily extended to include additional marketing instruments.

REFERENCES

Alba, Joseph W. and J. Wesley Hutchinson (1987), "Dimensions of Consumer Expertise," *Journal of Consumer Research*, 13 (4), 411–54.

Askitas, Nikos and Klaus Zimmermann (2009), "Google Econometrics and Unemployment Forecasting," *Applied Economics Quarterly*, 55 (2), 107–120.

Ataman, M. Berk, Carl F. Mela, and Harald J. van Heerde (2008), "Building Brands," *Marketing Science*, 27 (6), 1036–54.

Barry, Thomas (1987), "The Development of the Hierarchy Effects: A Historical Perspective," *Current Issues and Research in Advertising*, 10 (2), 251–95.

Beatty, Sharon E. and Scott M. Smith (1987), "External Search Effort: An Investigation Across Several Product Categories," *Journal of Consumer Research*, 14 (1), 83–95.

Bettman, James R. (1979), *An Information Processing Theory of Consumer Choice (Advances in Marketing)*. Reading, MA: Addison-Wesley Educational Publishers.

———, Mary Frances Luce, and John W. Payne (1998), "Constructive Consumer Choice Processes," *Journal of Consumer Research*, 25 (3), 187–217.

Boyd, Harper W., Jr., Michael L. Ray, and Edward C. Strong (1972), "An Attitudinal Framework for Advertising Strategy," *Journal of Marketing*, 36 (April), 27–33.

Bruce, Norris I., Kay Peters, and Prasad A. Naik (2012), "Discovering How Advertising Grows Sales and Builds Brands," *Journal of Marketing Research*, 49 (December), 793–806.

Carter, C.K. and R. Kohn (1994), "On Gibbs Sampling for State Space Models," *Biometrika*, 81 (3), 541–53.

Choi, Hyunyoung and Hal R. Varian (2009a), "Predicting Initial Claims for Unemployment Benefits," research report, Google, (accessed March 4, 2014), [available at http://static.googleusercontent.com/external_content/untrusted_dlcp/research.google.com/en/us/archive/papers/initialclaimsUS.pdf].

——— and ——— (2009b), "Predicting the Present with Google Trends," research report, Google, (accessed March 4, 2014), [available at http://static.googleusercontent.com/external_content/untrusted_dlcp/www.google.com/en/us/googleblogs/pdfs/google_predicting_the_present.pdf].

Da, Zhi, Joseph Engelberg, and Pengjie Gao (2011), "In Search of Attention," *Journal of Finance*, 66 (5), 1461–99.

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

Engel, James F. and Roger D. Blackwell (1982), *Consumer Behavior*. Chicago: Dryden.

Fruhwirth-Schnatter, Sylvia (1995), "Bayesian Model Discrimination and Bayes Factors for Linear Gaussian State Space Models," *Journal of the Royal Statistical Society, Series B (Methodological)*, 57 (1), 237–46.

Ginsberg, Jeremy, Matthew H. Mohebbi, Rajan S. Patel, Lynnette Brammer, Mark S. Smolinski, and Larry Brilliant (2009), "Detecting Influenza Epidemics Using Search Engine Query Data," *Nature*, 457 (7232), 1012–14.

Hanssens, Dominique M., Leonard J. Parsons, and Randall L. Schultz (2001), *Market Response Models: Econometric and Time Series Analysis (International Series in Quantitative Marketing)*, 2nd ed. Boston: Kluwer Academic Publishers.

Howard, John A. and Jagdish N. Sheth (1969), *The Theory of Buyer Behavior*. New York: John Wiley & Sons.

Hurvich, Clifford M., Robert Shumway, and Chih-Ling Tsai (1990), "Improved Estimators of Kullback-Leibler Information for Autoregressive Model Selection in Small Samples," *Biometrika*, 77 (4), 709–719.

——— and Chih-Ling Tsai (1989), "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 76 (2), 297–307.

Jacoby, Jacob, George J. Szybillo, and Jacqueline Busato-Schach (1977), "Information Acquisition Behavior in Brand Choice Situations," *Journal of Consumer Research*, 3 (4), 209–16.

J.D. Power and Associates (2008), *New Autoshopper.com Study*. Agoura Hills, CA: McGraw-Hill.

——— (2012), *New Autoshopper.com Study*. Agoura Hills, CA: McGraw-Hill.

Joo, Mingyu, Kenneth C. Wilbur, Bo Cowgill, and Yi Zhu (2014), "Television Advertising and Online Search," *Management Science*, 60 (1), 56–73.

Kotler, Philip, Neil Rackham, and Suj Krishnaswamy (2006), "Ending the War Between Sales and Marketing," *Harvard Business Review*, 84 (7/8), 68–78.

Lavidge, Robert J. and Gary A. Steiner (1961), "A Model for Predictive Measurements of Advertising Effectiveness," *Journal of Marketing*, 25 (October), 59–62.

Lilien, Gary L., Philip Kotler, and K. Sridhar Moorthy (1992), *Marketing Models*. Upper Saddle River, NJ: Prentice Hall.

Moorthy, Sridhar, Brian T. Ratchford, and Debabrata Talukdar (1997), "Consumer Information Search Revisited: Theory and Empirical Analysis," *Journal of Consumer Research*, 23 (4), 263–77.

Naik, Prasad A., and Kalyan Raman (2003), "Understanding the Impact of Synergy in Multimedia Communications," *Journal of Marketing Research*, 40 (November), 375–88.

———, ———, and Russell S. Winer (2005), "Planning Marketing-Mix Strategies in the Presence of Interaction Effects," *Marketing Science*, 24 (1), 25–34.

——— and Chih-Ling Tsai (2001), "Single-Index Model Selections," *Biometrika*, 88 (3) 821–32.

Newman, Joseph W. and Bradley D. Lockeman (1975), "Measuring Prepurchase Information Seeking," *Journal of Consumer Research*, 2 (3), 216–22.

——— and Richard Staelin (1972), "Prepurchase Information Seeking for New Cars and Major Household Appliances," *Journal of Marketing Research*, 9 (August), 249–57.

Palda, Kristian S. (1966), "The Hypothesis of a Hierarchy of Effects: A Partial Evaluation," *Journal of Marketing Research*, 3 (February), 13–24.

Pelat, Camile, Clement Turbelin, Avner Bar-Hen, Antoine Flahault, and Alain-Jacques Valleron (2009), "More Diseases Tracked by Using Google Trends," *Emerging Infectious Diseases*, 15 (8), 1327–28.

Punj, Girish N. and Richard Staelin (1983), "A Model of Consumer Information Search Behavior for New Automobiles," *Journal of Consumer Research*, 9 (4), 366–80.

Purcell, Kristen, Joanna Brenner, and Lee Rainie (2012), "Search Engine Use 2012," Pew Research Internet Project, (March 9), (accessed March 6, 2014), [available at http://www.pewinternet.org/2012/03/09/search-engine-use-2012/].

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

———, Debabrata Talukdar, and Myung-Soo Lee (2007), "The Impact of the Internet on Consumers' Use of Information Sources for Automobiles: A Re-Inquiry," *Journal of Consumer Research*, 34 (1), 111–19.

Srinivasan, Narasimhan and Brian T. Ratchford (1991), "An Empirical Test of a Model of External Search for Automobiles," *Journal of Consumer Research*, 18 (2), 233–42.

Srinivasan, Shuba, Marc Vanhuele, and Koen Pauwels (2010), "Mind-Set Metrics in Market Response Models: An Integrative Approach," *Journal of Marketing Research*, 47 (August), 672–84.

Vakratsas, Demetrios and Tim Ambler (1999), "How Advertising Works: What Do We Really Know?" *Journal of Marketing*, 63 (January), 26–43.

Vosen, Simon and Torsten Schmidt (2011), "Forecasting Private Consumption: Survey Based Indicators vs. Google Trends," *Journal of Forecasting*, 30 (6), 565–78.

West, Mike and Jeff Harrison (1997), *Bayesian Forecasting and Dynamic Models*, 2nd ed. New York: Springer Verlag.

Wooldridge, Jeffrey M. (2008), *Introductory Econometrics: A Modern Approach*, 4th ed. Mason, OH: South-Western.

Wu, Lynn and Erik Brynjolfsson (2009), "The Future of Prediction: How Google Searches Foreshadow Housing Prices and Quantities," in *ICIS 2009 Proceedings*, Paper 147.

Zaichkowsky, Judith L. (1985), "Measuring the Involvement Construct," *Journal of Consumer Research*, 12 (3), 341–52.

Zettelmeyer, Florian, Fiona Scott Morton, and Jorge Silva-Risso (2006), "How the Internet Lowers Prices: Evidence from Matched Survey and Automobile Transaction Data," *Journal of Marketing*, 43 (April), 168–81.

WEB APPENDIX

Decomposing the Impact of Advertising: Augmenting Sales with Online Search Data

Ye Hu

Rex Yuxing Du

Sina Damangir

WEB APPENDIX A – MODEL ESTIMATION

We adopt a procedure similar to that of Ataman, Mela, and van Heerde (2008) in model estimation. Using a state-space model form (West and Harrison 1997, pp.100), we present the observation equation and the state equation in A1 and A2.

(A1) $H_t = \theta\alpha_t + \beta X_t + v_t$ Observation Equation

(A2) $\alpha_t = \delta_{lag}\alpha_{t-1} + \delta_z Z_t + w_t$ State Equation

where the definition of $H_t, \theta, \alpha_t, \beta, X_t, v_t, \delta_{lag}, \delta_z, Z_t$, and w_t are described in the manuscript following Equations 17 and 18. For identification purposes, K_j^S is set to zero, or $v_{jt}^S \sim N(0, V_j^S)$,

We use a Gibbs sampler to draw the conditional posteriors of the parameters.

Step 1: $\alpha_t | H_t, V, W, \beta, \delta_{lag}, \delta_z, Z_t$ (forward filtering, backward sampling)

0. Define $\widetilde{H_t} = H_t - \beta X_t$.

<u>Forward filtering</u> (West and Harrison 1997, pp. 103-104):

1. Initial condition $(\alpha_0 | D_0) \sim N(m_0, C_0)$, We set initial values as $m_0 = 0, C_0 = 1$. D_0 represents the initial information at $t = 0$. The prior at t is $(\alpha_t | D_{t-1}) \sim N(a_t, R_t)$ where $a_t = \delta_{lag} m_{t-1} + \delta_z Z_t$ and $R_t = \delta_{lag} C_{t-1} \delta'_{lag} + W$
2. The one-step-ahead forecast at time t is $\left(\widetilde{H}_t \middle| D_{t-1}\right) \sim N(f_t, Q_t)$, where $f_t = \theta a_t$ and $Q_t = \theta R_t \theta' + V$

3. The posterior distribution at time t is $\alpha_t | D_t \sim N(m_t, C_t)$, where $m_t = a_t + R_t \theta' Q_t^{-1}(\widetilde{H}_t - f_t)$ and $C_t = R_t - R_t \theta' Q_t^{-1} \theta R_t$.

<u>Backward sampling (smoothing, West and Harrison 1997, p. 113)</u>:

4. At $t = T$, we first sample from the distribution $\alpha_{t+1} \sim N(m_t, C_t)$, then backward sampling for $t = T-1, \ldots, 1$ sampling from $p(\alpha_t | \alpha_{t+1}, \text{rest}) \sim N(q_t^*, Q_t^*)$, where $q_t^* = m_t + B_t(\alpha_{t+1} - a_{t+1})$, $Q_t^* = C_t + B_t R_{t+1} B_t'$, and $B_t = C_t \delta_1' R_{t+1}^{-1}$.

Step 2: $V | \alpha_t, H_t, \beta$

The priors of V (vector) follow independent inverted Gamma distributions.
$V \sim \text{Gamma}^{-1}\left(\frac{n_{V0}}{2}, \frac{S_{V0}}{2}\right)$, $n_{V0} = 3$ and $S_{V0} = 0.001$. The posterior is $V \sim \text{Gamma}^{-1}\left(\frac{n_{V1}}{2}, \frac{S_{V1}}{2}\right)$ with $n_{V1} = n_{V0} + T$ and $S_{V1} = S_{V0} + \sum_{t=1}^{T}\left(H_t - \beta X_t - \theta\alpha_t\right)'(H_t - \beta X_t - \theta\alpha_t)$.

Step 3: $W | \alpha_t, \delta_1, \delta_2$

The prior $W \sim \text{Gamma}^{-1}\left(\frac{n_{W0}}{2}, \frac{S_{W0}}{2}\right)$, $n_{W0} = 3$ and $S_{W0} = 0.001$. The posterior is $W \sim \text{Gamma}^{-1}\left(\frac{n_{W1}}{2}, \frac{S_{W1}}{2}\right)$ with $n_{W1} = n_{W0} + T$ and $S_{W1} = S_{W0} + \sum_{t=1}^{T}\left(\alpha_t - \delta_{lag}\alpha_{t-1} - \delta_z Z_t\right)'\left(\alpha_t - \delta_{lag}\alpha_{t-1} - \delta_z Z_t\right)$.

Step 4: $\delta_{lag}, \delta_z | Z_t, \alpha_t, W$

There is a closed-form solution for δ_{lag} and δ_z to the following equation (equivalent to a simple regression):

$$\alpha_t = \delta_{lag}\alpha_{t-1} + \delta_z Z_t + w_t$$

To obtain the conditional posterior distribution of the parameters, we define $\delta_K = \begin{bmatrix} \delta_{lag} \\ \delta_z \end{bmatrix}$ and $K_T = [\alpha_{t-1}\ Z_t]$ and $W_T = W \otimes I_{T-1}$ (note the dimension is $T-1$ because of the lagging term included). We place a normal prior on the parameters, $\delta_K \sim N\left(\underline{\mu}_{\delta_K}, \underline{\Sigma}_{\delta_K}\right)$. The full conditional posterior is also normal with $\delta_K \sim N\left(\overline{\mu}_{\delta_K}, \overline{\Sigma}_{\delta_K}\right)$, where $\overline{\mu}_{\delta_K} = \overline{\Sigma}_{\delta_K}\left\{\underline{\Sigma}_{\delta_K}^{-1}\underline{\mu}_{\delta_K} + \left[K_T(W_T)^{-1}\alpha_T^{Y}\right]\right\}$, and $\overline{\Sigma}_{\delta_K} = \left\{\underline{\Sigma}_{\delta_K}^{-1} + [K_T(W_T)^{-1}K_T']\right\}^{-1}$.

Step 5: $\beta | H_t, \theta, \alpha_t, V, \mu_\beta, s_\beta$ ($\phi^* | \alpha_t$ is estimated together with β)

Define $\overline{H_t} = H_t - \theta\alpha_t$ (except for $\varphi^*\alpha_t^{I}$) and $V_T = V \otimes I_T$. We assume the coefficients for each control variable are independent from each other. A hierarchical prior governs each β_j across vehicles $j = 1, \ldots, n$, where $\beta_j \sim N\left(\mu_\beta, s_\beta\right)$, and the hierarchical priors are $\mu_\beta \sim N\left(\underline{\mu_\beta}, \underline{\Sigma_\beta}\right)$ and

$s_\beta \sim \mathrm{Gamma}^{-1}\left(\underline{\nu_\beta}, \underline{V_\beta}^{-1}\right)$. And we draw from posterior of $\beta_j \sim N\left(\overline{\beta_j}, \overline{s_j}\right)$, where $\overline{\beta_j} = \overline{s_j}\left[s_\beta{}^{-1}\mu_\beta + X_t V_T^{-1}\overline{H_t}\right]$, and $\overline{s_j} = \left(s_\beta{}^{-1} + X_t V_T^{-1} X_t'\right)^{-1}$.

Step 6: $\mu_\beta, s_\beta | \beta$

The posterior conditional hierarchical parameters are $\mu_\beta \sim N\left(\overline{\mu_\beta}, \overline{\Sigma_\beta}\right)$ and $s_\beta \sim \mathrm{Gamma}^{-1}\left(\overline{\nu_\beta}, \left[\overline{\nu_\beta}\overline{s_\beta}\,\right]^{-1}\right)$, where $\overline{\Sigma_\beta} = \left(n s_\beta^{-1} + \underline{\Sigma_\beta}^{-1}\right)^{-1}$, $\overline{\mu_\beta} = \overline{\Sigma_\beta}\left[s_\beta{}^{-1}\Sigma_{j=1}^{n}\beta_j + \underline{\Sigma_\beta}^{-1}\underline{\mu_\beta}\right]$, $\overline{\nu_\beta} = n + \underline{\nu_\beta}$, $\overline{s_\beta} = \sum_{j=1}^{n}\left(\beta_j - \mu_\beta\right)^2 + \underline{s_\beta}$. ■

WEB APPENDIX B – CORRELATION BETWEEN SALES, ONLINE SEARCHES, ONLINE PRICE QUOTE REQUESTS AND WEBSITE VISITS

A key challenge in our study lies in whether Google Trends indexes can serve as a valid proxy for consumer prepurchase information interest, a latent state that is not directly observable. To cross validate the information content of Google Trends indexes, we assembled two additional datasets for the top-two selling vehicles in each of the four segments included in our study. The first dataset is obtained from Autometrics (http://www.autometrics.com/), which aggregates online price quote requests made by U.S. consumers through third-party automotive sites (e.g., Edmunds, Kelley Blue Book). The second dataset is obtained from Compete (https://www.compete.com/about-compete/our-data/), which relies on clickstream data collected from a 2-million-U.S.-Internet-user panel in tracking web traffic. The Autometrics data goes from January 2009 through December 2011, and the Compete data goes from February 2006 through June 2012. In contrast, our sales and Google Trends data cover January 2004 through July 2012.

The table at the end of the Appendix presents the correlation between log-transformed sales, online shopping searches (Google Trends), online price quote requests (Autometrics), and vehicle website visits (Compete). We see that (a) online searches, price quote requests, and website visits are all positively correlated with sales, with correlations averaging .470, .522, and .501, respectively; and (b) online searches are also positively correlated with price quote requests (min = .319, max = .834, mean = .621) and website visits (min = .050, max = .690, mean = .470).

Given the above positive and consistent correlation pattern, a plausible inference is that online searches, price quote requests and website visits share some common drivers. Since all three activities are commonly engaged in by consumers seeking information prior to buying a vehicle, we argue the most likely common driver behind the observed correlation is the level of consumer prepurchase information interest. We take this as another piece of empirical evidence supporting our argument that the search indexes we extracted from Google Trends contain strong and genuine signals about consumer prepurchase information interest.

We also note that, for the eight vehicles for which we have complete data on online searches, price quotes and Web visits, while the average correlations with sales are quite similar (.523 vs. .522 vs. .501), online searches have the strongest correlation in four cases, price quotes in two cases, and Web visits in two cases. Furthermore, although the overall correlation between online searches and price quotes is stronger than the correlation between online searches and web visits, the latter is stronger in three out of eight cases. This pattern suggests that although there is a common underlying driver behind them, online searches, price quotes, and web visits all have their own distinct, vehicle-specific dynamics.

A key potential difference between these tracking measures lies in that they may capture consumer information interest at slightly different stages of the purchase funnel. For example, it could be that the volume of Google searches is more driven by consumers who are in the initial information gathering stage, while the number of price quote requests is more driven by consumers who have narrowed down the consideration set and are closer to making a dealership visit. We see this as a promising avenue for future research. For example, one possibility is to include all three indicators in our model, allowing for different conversion rates between them and sales and between themselves.

Finally, we note, compared with online price quote requests from Autometrics and Web visits from Compete, online searches from Google Trends have many appealing features: it is free, goes far beyond the automotive industry, is based on a much larger user base, and thus can be more reliable, especially at the more granular level (e.g., a particular metro area).

Table B1: Correlation between Sales, Online Searches, Price Quote Requests, and Website Visits*

Segment	Vehicle	Sales & Online Searches	Sales & Price Quotes	Sales & Web Visits	Online Searches & Price Quotes	Online Searches & Web Visits
Compact SUV	Ford Escape	.751	.741	.539	.834	.389
	Honda CR-V	.764	.607	.212	.694	.446
	Jeep Liberty	.623				
	Jeep Wrangler	.642				
	Toyota RAV4	.748				
Midsize SUV	Ford Explorer	.397	.859	.592	.809	.690
	Honda Pilot	.239				
	Hyundai Santa Fe	.099				
	Jeep Grand Cherokee	.291	.721	.574	.832	.450
	Toyota Highlander	.377				
Compact Sedan	Ford Focus	.528				
	Honda Civic	.628	.311	.485	.487	.580
	Hyundai Elantra	.682				
	Toyota Corolla	.548	.286	.662	.319	.606

	Toyota Prius	.325				
	VW Jetta	.510				
Midsize Sedan	Chevrolet Malibu	.172				
	Honda Accord	.496	.313	.395	.506	.050
	Hyundai Sonata	.360				
	Nissan Altima	.372				
	Toyota Camry	.312	.340	.550	.484	.552
Overall	Average	.470 .523**	.522	.501	.621	.470

*: Data on online searches are from Google Trends; data on online price quote requests are from Autometrics; data on vehicle website visits are from Compete; all correlations are based on log-transformed data **: For the eight vehicles with complete data.