

Singapore Management University Institutional Knowledge at Singapore Management University

Research Collection Lee Kong Chian School Of
Business

Lee Kong Chian School of Business

9-2014

Exploring the Relationship Between Online Search and Offline Sales for Better "Nowcasting"

Sandeep R. CHANDUKALA

Singapore Management University, sandeepc@smu.edu.sg

Jeffrey P. DOTSON

Brigham Young University

Qing LIU

University of Wisconsin-Madison

Stephan CONRADY

Conrady Applied Science, LLC

DOI: <https://doi.org/10.1007/s40547-014-0020-1>

Follow this and additional works at: https://ink.library.smu.edu.sg/lkcsb_research

Part of the [Advertising and Promotion Management Commons](#), [E-Commerce Commons](#), and the [Marketing Commons](#)

Citation

CHANDUKALA, Sandeep R.; DOTSON, Jeffrey P.; LIU, Qing; and CONRADY, Stephan. Exploring the Relationship Between Online Search and Offline Sales for Better "Nowcasting". (2014). *Customer Needs and Solutions*. 1, (3), 176-187. Research Collection Lee Kong Chian School Of Business.

Available at: https://ink.library.smu.edu.sg/lkcsb_research/4800

This Journal Article is brought to you for free and open access by the Lee Kong Chian School of Business at Institutional Knowledge at Singapore Management University. It has been accepted for inclusion in Research Collection Lee Kong Chian School Of Business by an authorized administrator of Institutional Knowledge at Singapore Management University. For more information, please email libIR@smu.edu.sg.

Exploring the Relationship Between Online Search and Offline Sales for Better “Nowcasting”

Sandeep R. Chandukala · Jeffrey P. Dotson · Qing Liu · Stefan Conrady

Abstract Although the conditions that motivate individuals to buy, sell, search, and post on the Internet are diverse, the information generated as a byproduct of these activities has the potential to help marketers develop a better understanding of consumer and firm behavior. In this paper, we utilize online product search data to build a model that links sales to marketing mix activity. The objectives of this research are threefold: first, to incorporate secondary data collected from online sources into a model of demand, thus improving our ability to forecast sales; second, to develop a better understanding of the role of marketing mix activity in the sales generation process; and third, generation of “nowcast” or contemporaneous prediction of sales. We illustrate the benefits of our approach using data for a luxury automobile brand where we use

advertising as an example of marketing mix activity. We show that our proposed model produces improved sales forecast relative to standard time series approaches. Further, the specification of our model is ideal for the generation of a nowcast. We discuss why nowcasting has the potential to improve operations in the automotive industry.

Keywords Bayesian estimation · Online search · Nowcasting · Marketing mix modeling

1 Introduction

Active consumer participation in online product and information markets has created a rich source of secondary data that has the potential to help marketers develop a better understanding of consumer and firm behavior. For example, researchers at Google have used search engine queries to monitor the diffusion and severity of the seasonal flu in the USA [19]. Economists have used Google Trends data to better forecast a variety of economic activities including retail, automotive, and home sales [10, 11, 20]. Research in computational engineering and finance has shown that the mood states of Twitter users can help predict the stock market [4]. In marketing, researchers have shown that there is a positive correlation between blogging activity and customer acquisition [30].

Although potentially useful, integrating secondary data from online sources into existing marketing models in order to make meaningful policy decisions presents a variety of challenges, both conceptual and methodological [14, 32]. With the advent of new web services, managers now have access to large amounts of online activity data (both duration and number of sources). In the context of the automotive industry, for example, manufacturers have begun to collect and archive information about the number and characteristics

Sandeep R. Chandukala, Jeffrey P. Dotson, and Qing Liu contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s40547-014-0020-1) contains supplementary material, which is available to authorized users.

S. R. Chandukala
Kelley School of Business, Indiana University, 328C, 1309 E. Tenth Street, Bloomington, IN 47405, USA
e-mail: sarchand@indiana.edu

J. P. Dotson
Marriott School of Management, Brigham Young University, 665 TNRB, Provo, UT 84602, USA
e-mail: jeff_dotson@byu.edu

Q. Liu (✉)
Wisconsin School of Business, University of Wisconsin–Madison, 975 University Avenue, Madison, WI 53706, USA
e-mail: qliu@bus.wisc.edu

S. Conrady
Conrady Applied Science, LLC, 312 Hamlet’s End Way, Franklin, TN 37067, USA
e-mail: stefan.conrady@conradyscience.com

of customers that have expressed interest in their vehicles. Although the general belief is that this data has value, it is not clear what the data is informative of, or how it can be incorporated into traditional marketing mix models in order to study dynamics that exist, for example, among sales, online search, and marketing mix activity. For example, in the aforementioned research using Google trends and Twitter data, online search has been incorporated directly into the response model as an independent covariate. In the automotive context described above, this approach may be problematic as online search is likely to be endogenous to offline sales; that is, there may be unobserved common shocks that lead to changes in both offline sales and online search.

In this paper, we develop a marketing mix model that accounts for the possibility of simultaneous generation of sales and search. The form of this model allows us to derive the predictive distribution of sales, conditional upon observing a contemporaneous realization of online search. This predictive distribution provides us with a theoretically coherent approach to predicting the present or “nowcasting” or “Forecasts of current events for which data has not been released” [6]. This approach is especially useful in automobile industry where online search data is available in real time but the reporting of actual sales can be significantly lagged. Managers can therefore benefit from monitoring online search data as this provides them a way to forecast both current and future sales. This could provide a manager with real-time insight into the effectiveness of a marketing campaign, thus allowing them to make modifications if results are not consistent with expectations.

In addition to improved forecasting, our proposed model allows managers to obtain a better, more substantive understanding of the role of marketing activity in the sales generation process. Substantial effort in marketing has been focused on developing an understanding of the impact of marketing activity on sales, including the relationship between advertising and sales [12, 24]. However, understanding the impact of marketing activity on non-sales-based measures like consumer search has received relatively little attention in this literature. By accounting for the (potentially simultaneous) relationship between sales and search, our proposed model makes it possible to determine if the marketing activity: (1) increases consumer search for additional product information (i.e., search escalation); (2) leads to temporal shift in sales that would have been realized in the future (i.e., purchase acceleration); (3) generates new interest for a product (i.e., interest generation).

Specifically, our modeling approach treats both offline sales and online search as dependent variables that are reflective of a common, unobserved construct that we call latent interest or interest potential. Implicit in this definition is the assumption that marketing activity can have different types of effects: interest generation, purchase acceleration, or search

escalation. Accommodating simultaneity in the sales-search relationship, in order to understand the various roles of marketing mix activity, presents a variety of challenges for existing approaches to modeling marketing mix effects. Our modeling approach allows us to integrate these seemingly disparate sources of information into a unified framework and, as such, is one of the important contributions of this research.

We demonstrate the usefulness of our proposed model by applying it to time series data from the luxury automobile segment where we integrate information from three distinct sources: sales volume, online product search, and advertising expenditures. In this context, we investigate the process through which classical forms of advertising (television, print, and radio) influence the sales generation process. Our findings provide support for the belief that the role of advertising is one of latent interest generation rather than purchase acceleration or simple search escalation.

Taken collectively, our work contributes to the literature in three ways. First, our modeling approach accommodates simultaneity that exists between online search and offline sales through the inclusion of a common unobserved factor, latent interest. Second, the structure of our model allows us to test a variety of specifications regarding the role of marketing mix activity in the sales generation process. Finally, we show that our proposed model substantially improves our ability to nowcast sales relative to standard time series approaches.

The remainder of the paper is organized as follows: In the next section, we propose our conceptual model that links online search information with offline sales and marketing mix activity. We further discuss how the structure of our model can be used to derive the predictive distribution of demand, conditional on contemporaneous realizations of search, thus providing us with a theoretically coherent approach to nowcasting. We then describe the data used in our analysis and present and discuss results generated by our model. In the subsequent section, we summarize our findings and show how the results of our model can be used to improve managerial decision making through nowcasting. We conclude the paper by discussing avenues for future research.

2 Model Development and Nowcasting

The conceptual framework used to develop our model is presented in Fig. 1. Although the form of our model generalizes to all types of marketing mix activity, we highlight the use of advertising in Fig. 1 in anticipation of our empirical example. The dotted lines in this figure indicate alternative paths of influence of marketing activities that can be examined using our model. The proposed framework is flexible, intuitive, and allows us to structurally differentiate the effects of marketing

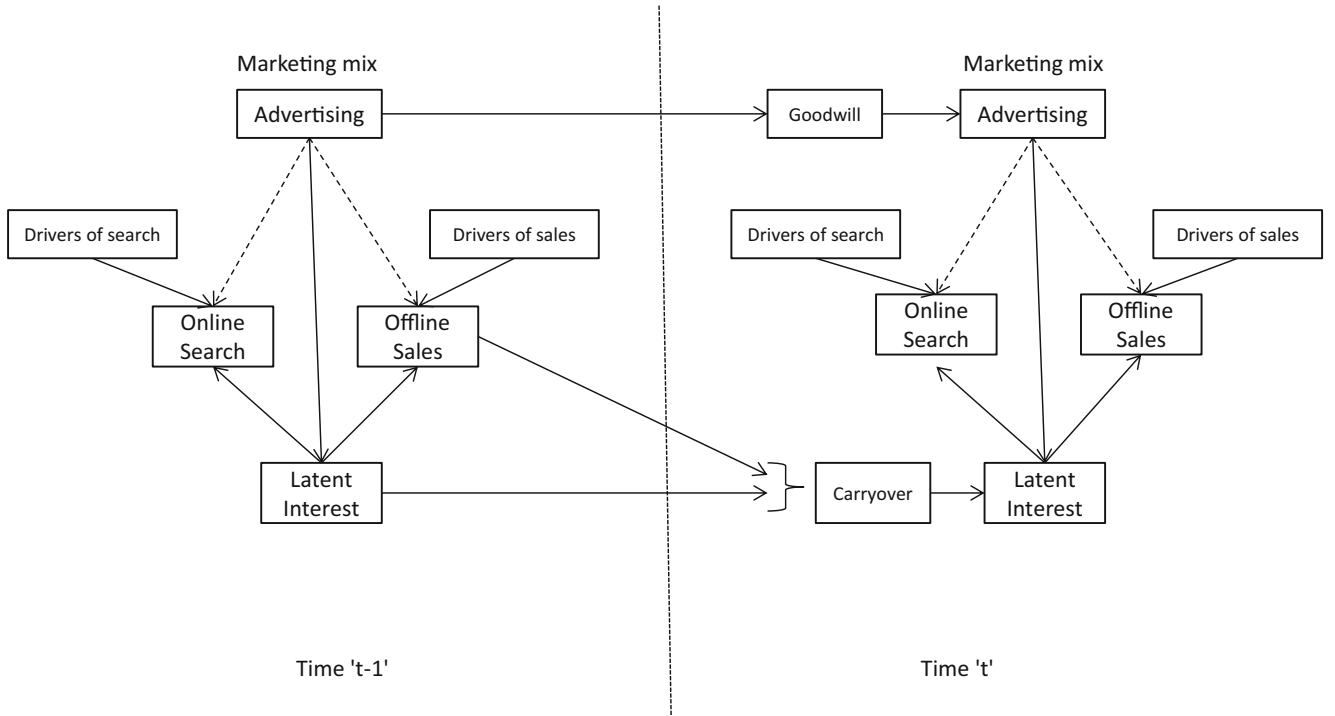


Fig. 1 Conceptual model. Linking offline sales, online search, and marketing mix activity

activity on system under study: i.e., interest generation, purchase acceleration, or search escalation.

We begin this section of the paper by considering and explaining the different components of our conceptual framework. We first discuss models that link latent interest with sales and search and then consider how advertising (or other forms of marketing mix activity) can be incorporated into the framework. By so doing, we can easily identify and test alternative model specifications of advertising effects.

2.1 A Model of Latent Interest, Search, and Sales

As seen in Fig. 1, our conceptual framework is built around the construct of latent interest. In this context, latent interest is an unobserved variable that is representative of demand potential within a market. The observation of online search behavior provides information about consumer interest in a product and is therefore a reflective measure of latent interest. In each time period, the magnitude of search is impacted by latent interest in the market and a set of independent drivers. Similarly, sales realized in each time period are also a reflective measure of latent interest. Like search, sales can also be impacted by additional marketing activity.

By treating both search and sales as (partially) reflective measures of a common construct, we control for the potential for simultaneous (i.e., endogenous) generation. This is an important feature of our model, as previous studies of online search and offline metrics consistently show strong contemporaneous correlation between the two variables. For

example, automobile search with sales [10, 11], search for cancer and cancer incidence [8], and search for jobs during an economic downturn and unemployment figures [17]. Given the preponderance of evidence, dynamic econometric models must account for the potentially simultaneous relationship between online search and offline metrics (e.g., sales). By introducing the concept of latent interest into our model, we are able to resolve the simultaneity issue both conceptually (i.e., we provide an explanation for the existence of simultaneity) and statistically.

Equation (1) formalizes the relationship between search, sales, and latent interest. The system of equations in (1) express sales in each time period, s_t , as a function of latent interest that currently exists in the market, d_t , and additional covariates, x_t , that have a short-term impact on sales. In this system, d_t can be viewed as a time-varying intercept for the sales equation that corresponds to expected sales that should be realized in each time period t , absent any additional promotional activity, x_t . Similarly, online search, q_t , in our model is driven by latent interest, d_t (albeit at a different scale) and other possible covariates, z_t , which could include, but are not limited to, the set of covariates that drive sales. Finally, we allow latent interest, d_t , to evolve from period to period through the inclusion of a carryover term, αd_{t-1} , and a (time-invariant) growth rate, λ .

$$\begin{aligned} s_t &= d_t + x_t' \beta + \varepsilon_t^s \\ q_t &= \kappa + \gamma d_t + z_t' \psi + \varepsilon_t^q \\ d_t &= \lambda + \alpha d_{t-1} + \varepsilon_t^d \end{aligned} \quad (1)$$

Equation (1) forms a system where observed sales and online search are connected through latent interest, d_t . Consistent with standard practice in the literature, we assume that the errors ε_t^s , ε_t^q , and ε_t^d are serially uncorrelated and normally distributed with mean 0 and variances equal to σ_s^2 , σ_q^2 , and σ_d^2 , respectively [31]. Dependence between sales and search is induced directly through latent interest and not through correlated random shocks.

Although the system described by Eq. (1) allows us to connect online search and corresponding offline sales, it is unable to capture a variety of patterns commonly observed in sales data. We therefore generalize the basic model of latent interest to manifest a carryover effect from the prior time period. Conceptually, latent interest carryover is the portion of expected demand from the previous period that was not realized. This effect is denoted in Fig. 1 as carryover. Mathematically, latent interest carryover is the difference between latent interest and sales from the previous period, $(d_{t-1} - s_{t-1})$. As such, latent interest in the current period, d_t , is a function of the baseline growth rate λ and the interest carried over from the previous period $(d_{t-1} - s_{t-1})$. Note that for simplification, we use the term “interest carryover” to capture both the case when latent interest from previous period was not fulfilled (i.e., $d_{t-1} > s_{t-1}$) and the case when sales from the last period exceeded latent interest (i.e., $d_{t-1} < s_{t-1}$). The system of equations in (1) with latent interest carryover can be re-specified as:

$$\begin{aligned} s_t &= d_t + x_t' \beta + \varepsilon_t^s \\ q_t &= \kappa + \gamma d_t + z_t' \psi + \varepsilon_t^q \\ d_t &= \lambda + \alpha(d_{t-1} - s_{t-1}) + \varepsilon_t^d \end{aligned} \quad (2)$$

where α attenuates the carryover of latent interest in the previous period (i.e., not all unrealized sales may pass through from period to period).

Although the introduction of carryover latent interest (demonstrated as “carryover” in Fig. 1) seems like a relatively minor change, it allows us to capture a variety of interesting features in the market place. First, consider the case where $d_{t-1} > s_{t-1}$ and $\alpha > 0$. This corresponds to a scenario where the full potential interest in the marketplace in period $t-1$ was not realized as sales, perhaps due to some random demand shock. Conditional on the value of α , a portion of this interest potential still exists in the market and can be realized in time t . A more interesting case is one where $d_{t-1} < s_{t-1}$, that is when sales exceeded expected latent interest in period $t-1$. This can occur when, for example, a firm runs a short-term sales promotion, as captured by x_{t-1} , that encourages consumers to expedite purchases they would have made in the future. Under this scenario, the firm achieves a boost to sales (relative to expectations) in time period $t-1$. Given the structure of our model in Eq. (2), interest in time period t will decrease

according to $\alpha(d_{t-1} - s_{t-1})$. This phenomenon is often seen in practice and is formally referred to as purchase acceleration or a post-promotional dip. A benefit of the model structure presented in Eq. (2) is that it allows us to capture a variety of interesting sales patterns that can be influenced by marketers. In the following section, we exploit this structure to study the effects of the marketing mix, and in particular, the advertising effects as an example.

2.2 Impact of Marketing Activity

The proliferation and wide availability of search data has resulted in a variety of investigations of the relationship of search and sales [10, 11]. That said, little research exists that has incorporated consumer online search into formal marketing mix models. We contribute to this literature by proposing a model that links offline sales and online search to marketing mix activity. Understanding the role of marketing activity in the context of a sales-search model has the potential to be a very valuable tool for marketing managers. In the discussion that follows, we will restrict our attention to advertising as element of the marketing mix of interest. That said, the form of this model will generalize to other types of marketing activity.

A key objective of this research is to develop a better understanding of the role of advertising in the sales generation process. Rather than simply including advertising expenditure as a covariate in our proposed model, we include an estimated measure of advertising stock or goodwill. This is consistent with recent studies of advertising effects that have investigated the forgetting or carryover effects of advertising on sales [2, 5]. Advertising stock is defined as the exponentially weighted average of past advertising [1, 34] and can be written as:

$$g_t = \sum_{i=0}^{t-1} \phi^i a_{t-i} \quad (3)$$

where $\phi \in [0, 1]$.

Including advertising stock in the model introduces an additional dynamic component to the model where advertising expenditures can exhibit a long-term impact on sales through both the dynamics that exist in the sales-response system and through their accumulated contribution to g_t (represented as goodwill in Fig. 1). Note that to capture diminishing returns to advertising expenditures, we operationalize advertising spend a_t as $\ln(1 + a_t)$. Further justification for the log specification is provided in Doyle and Saunders [16] where they demonstrate that for advertising response models with diminishing returns, linear and exponential functional forms cannot be used. Bass et al. [2] provides a more recent use of this specification and a detailed explanation.

Furthermore, recent research has investigated the indirect effects of advertising, i.e., the impact of advertising through consideration sets on choice [34] and through brand beliefs and consumer actions on purchase likelihood [9]. Findings from these papers support the presence of the multistage decision processes that were proposed by early work on advertising effectiveness by Howard and Sheth [22] and Lavidge and Steiner [23]. We therefore, propose an alternative explanation where advertising impacts latent interest (solid line from advertising to latent interest in Fig. 1) rather than sales. Thus, advertising has an indirect effect on sales and search through latent interest which in turn, drives both sales and online search:

$$\begin{aligned} s_t &= d_t + x_t' \beta + \varepsilon_t^s \\ q_t &= \kappa + \gamma d_t + z_t' \psi + \varepsilon_t^q \\ d_t &= \lambda + \alpha(d_{t-1} - s_{t-1}) + g_t \eta + \varepsilon_t^d \end{aligned} \quad (4)$$

where η captures the impact of advertising stock on latent interest. Thus, in our proposed model, advertising impacts latent interest and latent interest drives both sales and search. We also test a variety of models (described in Appendix A) of the presumed impact of advertising on sales and search. Next, we show how our proposed model can be used by managers in better predicting sales, given that managers have real-time access to search data while information on sales data can be significantly lagged (up to a month or more in the auto industry).

2.3 Generation of a Nowcast

Like Choi and Varian [10, 11], we find that online search in the present period is not predictive of future offline sales in our empirical application. Rather, the two variables are contemporaneously correlated (i.e., present period search is predictive of present period sales). Including search in the model, however, can still be of practical benefit to a firm. Given the correlation between search and sales, we can use online search to generate a nowcast or contemporaneous prediction of sales. In the context of the automotive industry, this is especially beneficial because there often exists a significant lag between the realization and reporting of sales. Search information, however, can be observed in real time thus, allowing us to better predict how products are currently faring in the marketplace.

Given the structure of our proposed model, this is accomplished by deriving the predictive distribution of sales in the next period, s_{t+1} , conditional upon observing search in the same time period, q_{t+1} . To ease exposition and more closely connect the system estimated in our empirical application, we omit the collection of covariates, z_t , from the equation for

search, q_t . As shown in Eq. (5), the predictive distribution of $[s_{t+1} | q_{t+1}, \text{else}]$ is proportional to the joint distribution of $[s_{t+1}, q_{t+1} | \text{else}]$ where we integrate out latent interest, d_{t+1} :

$$\begin{aligned} & \int [s_{t+1} | q_{t+1}, X_{t+1}, g_{t+1}, s_t, d_t, \beta, \lambda, \alpha, \eta, \kappa, \gamma, \sigma_s^2, \sigma_d^2, \sigma_q^2] \propto \\ & \int [s_{t+1}, q_{t+1} | X_{t+1}, g_{t+1}, s_t, d_t, \beta, \lambda, \alpha, \eta, \kappa, \gamma, \sigma_s^2, \sigma_d^2, \sigma_q^2] = \\ & \int [s_{t+1} | X_{t+1}, d_{t+1}, \beta, \sigma_s^2] [q_{t+1} | d_{t+1}, \kappa, \gamma, \sigma_q^2] \\ & [d_{t+1} | g_{t+1}, s_t, d_t, \lambda, \alpha, \eta, \sigma_d^2] dd_{t+1} \end{aligned} \quad (5)$$

Given the parametric assumptions of our model, we can generate draws of $[s_{t+1}, q_{t+1} | \text{else}]$ from a normal distribution with mean:

$$\mu_{s_{t+1}|q_{t+1}} = X'_{t+1} \beta + \frac{\sigma_d^2 \gamma (q_{t+1} - \kappa)}{\sigma_d^2 \gamma^2 + \sigma_q^2} + \frac{\sigma_q^2 (\lambda + \alpha(d_t - s_t) + g_{t+1} \eta)}{\sigma_d^2 \gamma^2 + \sigma_q^2}$$

and variance:

$$\frac{\sigma_d^2 \sigma_q^2 + \sigma_s^2 \sigma_d^2 \gamma^2 + \sigma_s^2 \sigma_q^2}{\sigma_d^2 \gamma^2 + \sigma_q^2}$$

This form of nowcasting provides us with an expectation of contemporaneous sales performance and allows us to quantify our uncertainty about the same. We will show later in the paper that the nowcast generated from our proposed model are more accurate than those developed using standard time series approaches.

3 Estimation

We estimate our proposed models using Bayesian statistical methods. By linking online search and offline sales through unobserved latent interest, our proposed models are consistent with existing research where critical elements of information are unobserved [2, 7, 33]. However, our specification of latent, realized, and carryover interest prevents us from being able to express Eq. (2) in the canonical state space form [31], thus creating a variety of estimation challenges. Specifically, our model specification induces a variety of restrictions on the coefficient matrix that prevent us from using existing software for estimation (e.g., the Kalman Filter). As such, we develop an estimation algorithm, where at each iteration in the sampler we first project out the latent interest vector, $d = (d_t, \dots, d_1)'$, and then treat it as a conditioning argument and sequentially sample from the full conditional distributions of the remaining model parameters. The draw of ϕ (Eq. 3) is executed using a Metropolis-Hastings algorithm. Details of the full estimation

algorithm appear in Appendix B. We run the chain for 20,000 iterations and use the last 5,000 iterations for estimating the posterior estimates of the parameters and use trace plots to assess the convergence criterion. The efficacy and mixing properties of our estimation routine were tested using a simulation study. Appendix C presents the results of our simulation study, where we generate the data using our proposed model (M_3 , Eq. 4) and demonstrate that we can recover the true parameter values.

4 Data

Empirically, we explore the impact of advertising on sales and online search in the context of the automotive industry. Data for this project are provided by a national marketing research consultancy that specializes in the automotive industry and consist of weekly observations of three independently collected components for a well-known luxury automobile brand: unit sales (s), online search information (q) and advertising expenditures (a). The data are specific to a particular region (i.e., one of the largest DMAs in the US market) and consist of 51 weekly observations. Descriptive statistics for the data set appear in Table 1 and time series plots of the key variables are provided in Fig. 2. We estimate our models and report our results using the first 49 weeks in our data set and withhold the remaining 2 weeks in order to assess predictive fit. To ensure that the short holdout period was not driving our results, we

Table 1 Description of data, model parameters, and summary statistics

Variable	Description	Mean	Std Dev
Observed Variables			
s_t	Unit sales	505	236
q_t	On-line search	3,679	800
a_t	Advertising expenditures (1,000's)	\$120	\$147
x_t	Covariates - End of Month indicators		
Latent Variables			
d_t	Latent interest (linking sales and search)		
g_t	Advertising stock		
Model parameters			
α	Effect of carryover interest from previous period		
β	Direct impact of covariates on sales		
γ	Effect of latent interest on online search		
δ	Effect of advertising stock on online search		
η	Effect of advertising stock on latent interest		
θ	Effect of advertising stock on sales		
κ	Baseline search rate		
λ	Baseline interest excluding any carryover effects		
ϕ	Decay rate for advertising stock		

also estimated each model using a longer holdout period (5 weeks) and obtained similar results on both model fit and cross-model comparisons.

4.1 Unit Sales

For each week in the sample, we have access to the total number of new units sold in all dealerships within the market in question. As shown in the top panel of Fig. 2, the unit sales series appears to be stationary (e.g., it does not exhibit any pronounced seasonal or trend features). There are, however, recurrent spikes in sales that correspond to the retail close date (i.e., the end of the fiscal month). The spikes in sales can be attributed to two possible causes: (1) an acceleration of promotional incentives designed to help the dealership meet its monthly sales quota and/or (2) a delayed reporting of sales that occurred earlier in the month. As we do not have direct measurements of either of these potential explanations, we control for the observed spikes in sales by including an indicator variable for the monthly close date in our set of covariates, x_t .

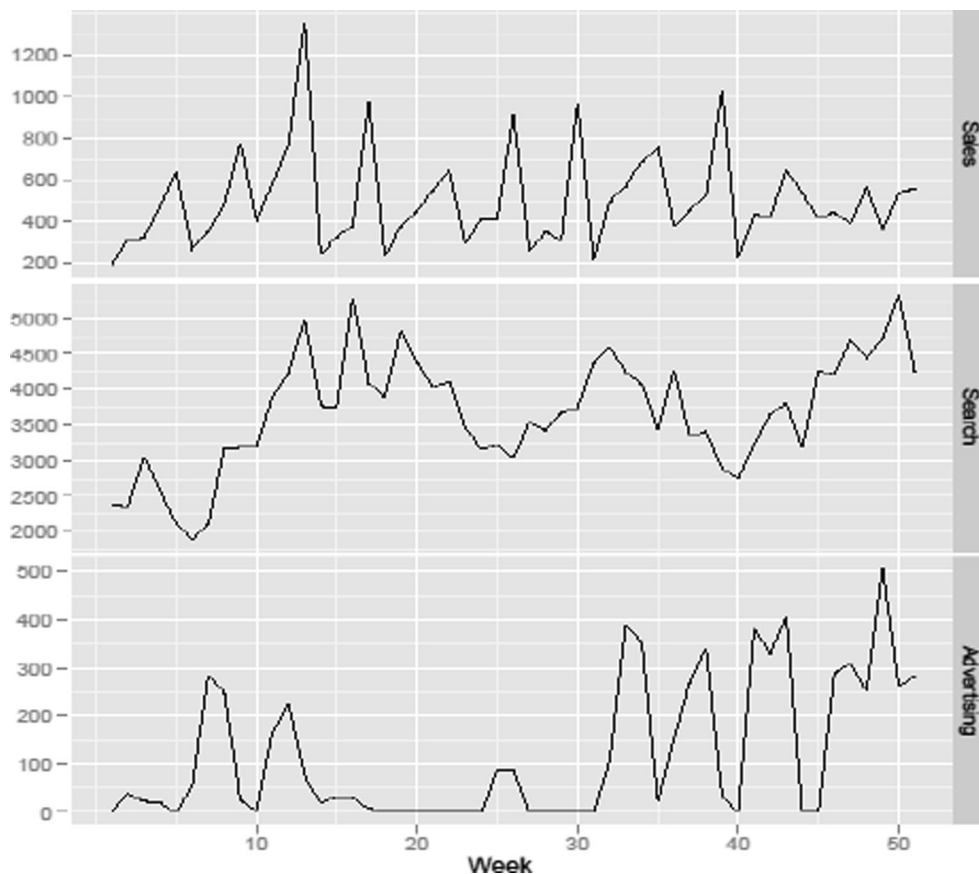
4.2 Advertising Expenditures

Advertising data consist of the total dollar amount spent on traditional media (print, radio, and television) during each week. The data are provided by a media tracking firm that monitors and records advertising broadcast/published in major media outlets (network and cable television, newspapers, and radio). Conditional on the source of the advertisement and breadth of exposure (i.e., the time it was shown on television), the firm is able to compute the expected cost associated with the advertisement. Normalizing the cost of each advertisement in terms of its dollar amount allows us to aggregate over disparate forms of marketing communication. Given the limited number of observations in our data set, we combine expenditures on print, radio, and television into a single, composite media expenditure variable.

4.3 Online Search

Search data are provided by an automotive analytics vendor that specializes in the collection and aggregation of online automotive data. Specifically, the firm collects data from prospective buyers who express desire in either purchasing or learning more about a vehicle. For example, a record is created in our data set when a prospective buyer configures and/or requests a price quote through a site like edmunds.com. In order to initiate this process, the buyer must input their zip code, thus allowing us to connect these records to specific geographical regions. In the automotive industry, this type of information is also referred to as consideration, demand, or “hand-raising” data. Our data set on search contains

Fig. 2 Time series plots of observed data



information aggregated (by the automotive analytics vendor) from 117 different online sources.

5 Alternate Models and Results

In this section, we present all the different alternate model specifications and the results for the fit and parameter estimates of our best fitting model.

5.1 Alternate Models

We compare our proposed model to a variety of models commonly used in the time series literature. Also, given our proposed model in Fig. 1, we can specify a variety of alternate models that differ in terms of the presumed impact of advertising. This is accomplished by allowing advertising to enter the system as a covariate in one or more of the three equations in the dynamic system. For example, we could allow advertising to directly impact sales s_t , or, rather, it could directly impact search. By searching over the full space of prospective models and identifying the one that best corresponds to our observed data, we can make inference about the role of advertising in the sales generation process. Our proposed

model allows us to test for purchase acceleration, interest generation, and search escalation effects of advertising expenditures.

Specifically, we test the following models: (1) an AR(1) model without any covariates (i.e., sales regressed on lagged sales); (2) an AR(1) model with online search, advertising, and covariates (contemporaneous and lagged); (3) a Vector Auto Regressive Model (VAR) with sales, online search, and advertising all included as endogenous variables and covariates included as exogenous variables; (4) model of realized interest only (M_0 —Eq. (1)); (5) model of latent interest carry-over (M_1 —Eq. (2)); (6) model of purchase acceleration (M_2 —Eq. A.1, Appendix A); (7) our proposed model of interest generation (M_3 —Eq. (4)); (8) model of search escalation (M_4 —Eq. A.2, Appendix A); (9) model of sales and search (M_5 —Eq. A.3, Appendix A); and (10) model of interest and search (M_6 —Eq. A.4, Appendix A).

The AR(1) models ignore the potential endogenous relationship between online search and offline sales. The VAR model addresses the endogeneity issue but lacks the flexibility and richness that our model provides in understanding the role of advertising in online search and offline sales. We also examine four alternative models of the relationship between advertising and sales. Each model is characterized according to its simulated impulse response function (IRF) presented in

Fig. 3. The IRF, widely used in time series literature (for example, [29]), is computed by first striking the system (i.e., creating an impulse in advertising spend) and then observing how the resulting energy diffuses throughout the system over time (i.e., what is the impact on sales?). This process is repeated and displayed for both sales (s_t) and search (q_t). For simple illustration purposes, the covariates, x_t , are assumed to be 0 and the advertising impulse, a_t , occurs only in week 4. For example, the upper left panel of Fig. 3 shows the IRF for Model 1 (M_1), a model where advertising has no impact on either sales or search, hence the horizontal line for each. For Model 4 (M_4), however, advertising impacts only online search but not sales and hence the IRF shows a horizontal line only for sales.

By examining the implied IRF for each alternative model, we are able to observe the qualitative differences in each specification as well as the empirical patterns of data required for identification. For example, Eq. (4), i.e., model M_3 , corresponds to a model of latent interest generation. This can be seen in the corresponding IRF in Fig. 3. An increase in advertising gives rise to an immediate increase in both sales and search behavior. Although the effect dissipates in

subsequent periods, it never drops below 0, thus the net effect of advertising is always positive. Specific details of each model are discussed in Appendix A.

5.2 Results

Table 2 reports the in-sample and holdout predictive fit for all models: the seven models proposed (M_0 – M_6) and the three traditional time series models. For all models presented, we compute the mean absolute percentage error (MAPE) and Theil’s U [31] for both in-sample and hold-out sample. Both MAPE [2] and Theil’s U are independent of the scale of the data, thus allowing for easy, more meaningful interpretation. We report Theil’s U , because it compares the MSEs of the proposed model with a “no-change” model [15]. Smaller values for MAPE and Theil’s U are indicative of better model fit.

Interestingly, the traditional time series models (i.e., AR(1) and VAR) do not fit the data as well as the proposed models that account for interest carryover and the indirect effect of advertising. Furthermore, we find that models with direct impact of advertising stock on sales (M_2) or online search

Fig. 3 Simulated impulse response functions (M_1 – M_6)

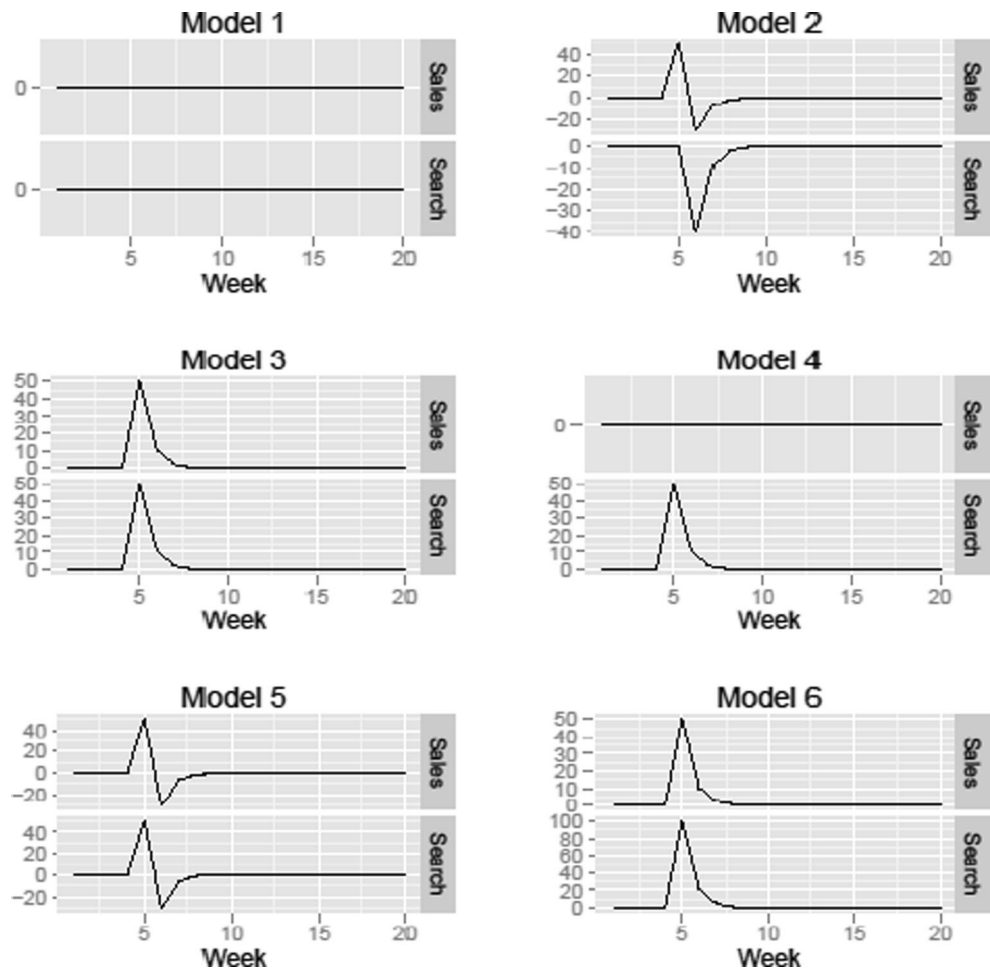


Table 2 Model fit statistics (in-sample and holdout predictions)

Model Description	In-Sample		Holdout	
	MAPE	Theil's U	MAPE	Theil's U
AR(1) without covariates	28.6 %	0.635	14.6 %	7.461
AR(1) with search and advertising	24.5 %	0.545	18.8 %	11.290
VAR with search and advertising	24.6 %	0.555	16.7 %	11.960
M ₀ —realized interest only	15.8 %	0.329	27.1 %	14.450
M ₁ —latent interest carryover	12.1 %	0.249	24.3 %	12.795
M ₂ —purchase acceleration	10.4 %	0.216	13.1 %	7.004
M ₃ —interest generation	9.9%	0.206	10.2%	5.537
M ₄ —search escalation	12.1 %	0.251	24.7 %	12.860
M ₅ —sales and search	11.0 %	0.222	13.0 %	6.857
M ₆ —interest and search	11.0 %	0.223	10.3 %	5.645

(M₄) fit the data poorly. The fit statistics shown in Table 2 indicate that the best fitting model is M₃, in terms of both in-sample and hold-out fit. This implies that primary influence of advertising on sales and online search is realized through latent interest. It is also worth noting that the fit statistics indicate that models M₆ and M₃ perform similarly for the holdout sample. This can be explained by noting that the only difference in specification between models M₃ and M₆ is the influence of advertising stock. In M₃ advertising, stock only influences latent interest while it influences latent interest and online search in M₆. Upon further investigation, we find that the parameter estimates for models M₃ and M₆ are very similar and the impact of advertising stock on online search (δ) is not significant (parameter estimates for M₆ are available from the authors). As such, M₆ reduces to something very close to M₃. This result further supports M₃ as the best fitting model. Taken collectively, these results support the belief that the role of advertising is one of interest generation (M₃) rather than purchase acceleration (M₂) or online search escalation (M₄).

Table 3 provides the coefficient estimates for M₃, our best fitting model. Reported are the posterior means and standard deviations for all the parameters in the model. The parameter estimates appear to be reasonable in terms of both their magnitude and algebraic signs. For example, the coefficient of interest carryover ($\alpha=0.259$) is positive, indicating that at every time period, there is a certain portion of the latent interest from the previous period that still exists in the market place. Therefore, interest carryover from the previous period has a large positive impact on the current period latent interest. The end of month indicator (EOM) used as a covariate is significant and positive ($\beta=0.725$), which supports the general notion that there are usually higher sales toward the end of the month when dealers are trying to meet their quotas.

In addition to the results presented above, there are three key parameters of interest that are worth discussing. First, we find that latent interest has a significant positive impact on consumers' online search behavior ($\gamma=0.287$). This result is

important, as it supports our hypothesis that both search and sales are reflective measures of underlying interest. Second, we find that advertising stock also has a significant positive effect on latent interest ($\eta=0.076$), thus influencing both sales and search. This result is also important as it demonstrates a relationship between advertising and sales (and search), something that is often difficult to show in the context of a marketing mix model. Finally, the diminishing effect of advertising over time is also significant ($\phi=0.066$) and is consistent with previous studies [2, 26]. This implies that there is some, albeit slight, contribution of advertising to an overall stock variable. This stock, however, decays rapidly, lasting just a little over a week (1.06 weeks in expectation). As such, firms must frequently replenish their stock of advertising goodwill in order to realize the benefits.

6 Discussion and Conclusion

We demonstrate that the standard models with advertising and search as covariates in an AR(1) model, or search, advertising,

Table 3 Posterior parameter estimates for the best fitting model (M₃)

Parameter	Mean	Std Dev
β —EOM indicator	0.725	0.096
λ —Baseline interest	5.886	0.071
α —Effect of latent interest carryover	0.259	0.125
κ —Baseline search rate	6.459	0.807
γ —Effect of latent interest on search	0.287	0.135
η —Effect of advertising stock on interest	0.076	0.028
ϕ —Advertising discount rate	0.066	0.156
σ_s^2 —Error variance - Sales	0.041	0.014
σ_d^2 —Error variance - Latent interest	0.030	0.012
σ_q^2 —Error variance - Search	0.048	0.010

Reported estimates have more than 95 % of the posterior mass away from zero

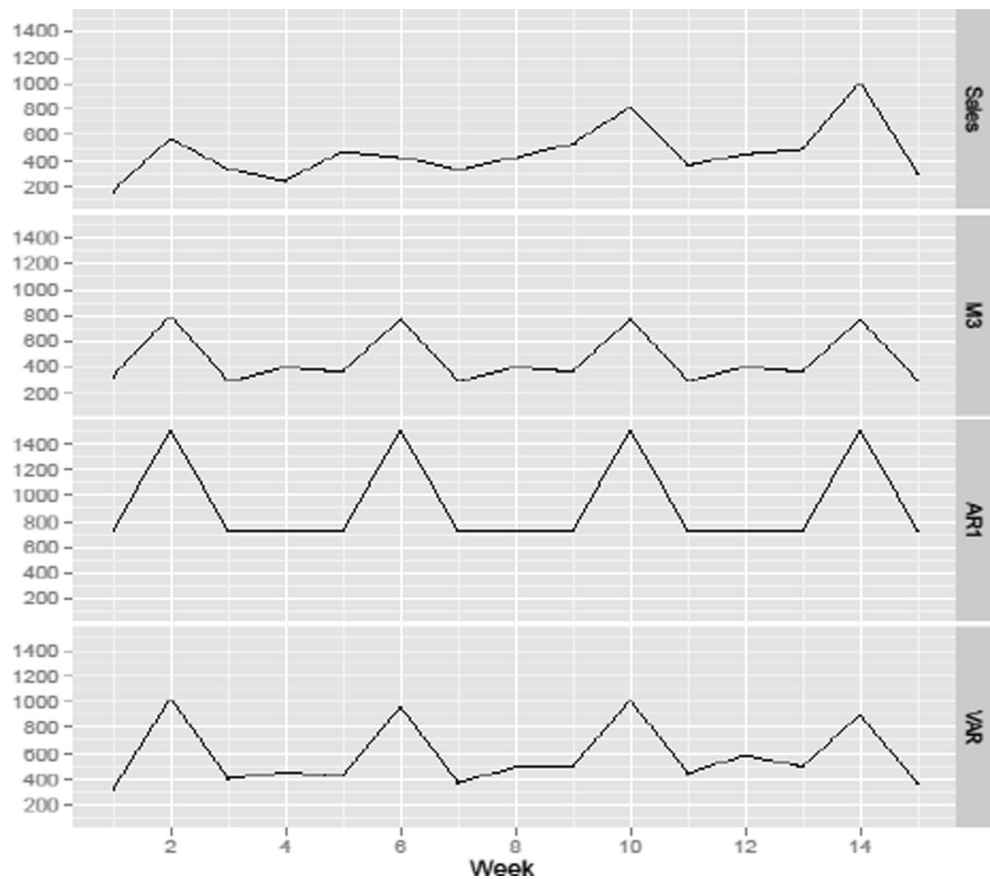
and sales in a VAR model, fit the data worse than our proposed approach. The fit statistics in Table 2 provide strong evidence that a model of interest carryover where advertising effects are manifest through latent interest provides a better explanation for the link between online search and offline sales in the luxury automobile category. The primary reason for this improved fit is that our proposed modeling approach accounts for the contemporaneous correlation between search and sales and is flexible in capturing changes inherent in automobile sales data. For example, introduction of the latent interest carryover captures instances where full interest potential is not realized ($d_{t-1} > s_{t-1}$) and scenarios where sales exceeds expected interest ($d_{t-1} < s_{t-1}$) due to short-term sales promotions.

Our modeling approach has the potential to help managers determine if their current levels of marketing mix, in particular, advertising are too high or too low based on the advertising stock and the diminishing effect of advertising over time. This can provide managers with insight into the amount of investment (based on advertising stock) and the time duration between the exposures (based on diminishing effect of advertising) required to make their brand advertising more effective. Our research also tests the direct and indirect effects of advertising on sales and search. We demonstrate that advertising has an indirect impact on sales and search through latent interest.

Furthermore, information about the number of people who specifically search online for a given product provides a good indication of the portion of sales that a manager can expect in the present or future in a given region. Our proposed methodology provides managers with a tool that allows them to incorporate secondary search data into models of sales and marketing mix to better understand and predict sales. To this effect, we propose a theoretically coherent approach to derive the predictive distribution of sales conditional upon observing a contemporaneous realization of online search. This predictive distribution therefore provides managers with a way to generate a contemporaneous forecast or nowcast which is particularly relevant in the automobile industry.

The benefits of the nowcasting approach are shown in Fig. 4 where we contrast actual sales over a 15-week holdout period to predicted sales using nowcast derived in Eq. (11), and to predicted sales generated from a simple AR(1) model which includes end of month indicators as covariates, as well as to those generated from the VAR model which treats sales, search, and advertising expenditure all as endogenous variables. We use a longer time frame to illustrate the fact that, conditional on observing search, our approach can generate an accurate forecast of sales well into the future. During the 15-week forecast period, the MAPE for the AR(1) model is 110.0 % whereas the MAPE for the VAR model is 38.2 %

Fig. 4 Illustration of nowcast



and the MAPE for our model is 30.5 %. This difference in predictive fit is a direct result of both the structure of our model and the inclusion of search as a contemporaneous predictor. In particular, the inferior performance of the AR(1) model in predictive fit relative to our model and the VAR model suggests that it is indeed beneficial to incorporate online search information into the nowcast. Between the two models that do incorporate online search information, our model structure provides a better predictive fit than does the VAR model. Thus, it clearly demonstrates that our proposed approach of using online search data to generate contemporaneous prediction of offline sales could be a valuable tool for managers.

7 Future Research

This work raises a variety of additional questions that are worthy of future research. First, we observe a strong, contemporaneous relationship between online search and offline sales. This result is similar to [10, 11] conclusion regarding the relationship between online search observed through Google Trends and a variety of offline metrics. Although we show that these contemporaneous relationships are useful for model construction and nowcasting, it would be of great benefit to both the academic and practitioner community to identify forms of online behavior that are predictive of future behavior. It is possible that consumers engage in systematically differentiated search behavior at various stages of the purchase process. If this is the case, research is needed that connects micro-level search behavior to the various stages of the purchase funnel, thus enabling the prediction of consumer attitudes and behavior.

Second, in this research, we only incorporate a specific element of the marketing mix (i.e., advertising), but it would be interesting to include and test other forms, i.e., promotions and pricing as well. It would also be useful to extend our research to incorporate a variety of forms of advertising. In this paper, we find that classical advertising (TV, radio, print) results in the creation of new interest for a brand. It is likely that the role of other types of marketing activities may influence the sales generation process in diverse, possibly interactive ways. By fitting a model that incorporates multiple marketing actions, we could begin to assess their relative strengths and weaknesses, thus allowing us to provide guidance as to the optimal allocation of resources.

Finally, we are limited by our data to sales and search for a single automotive brand. It would be interesting to explore a similar model specification constructed in a competitive context. This would allow us to determine if and to what extent own-brand activities can actually create demand for the category (i.e., category expansion effects). Also, incorporating information about competitive brand search and marketing

mix could be an interesting extension. We are not aware of research that integrates both competitive search and competitive advertising into models of own-brand effects. We believe the modeling approach developed in this paper could be used to effectively explore this area. We leave these issues for future research.

References

1. Bass FM, Clarke DG (1972) Testing distributed lag models of advertising effect. *J Mark Res* 9:298–308
2. Bass FM, Bruce N, Majumdar S, Murthi BPS (2007) Wearout effects of different themes: a dynamic Bayesian model of the advertising-sales relationship. *Mark Sci* 26(2):179–195
3. Bettman J, Park WC (1980) Effects of prior knowledge and experience and phase of the choice process on consumer decision processes: a protocol analysis. *J Cons Res* 7:234–248
4. Bollen J, Mao H, Zeng XJ (2011) Twitter mood predicts the stock market. *J Comp Sci* 2(1):1–8
5. Bruce N (2008) Pooling and dynamic forgetting effects in multi-theme advertising: tracking the advertising sales relationship with particle filters. *Mark Sci* 27(4):659–673
6. Bughin JR (2011) ‘Nowcasting’ the Belgian Economy. SSRN working paper. <http://ssrn.com/abstract=1903791>
7. Carter C, Kohn R (1994) On Gibbs sampling for state space models. *Biometrika* 81:541–553
8. Cooper C, KP Mallon, S Leadbetter, L Pollack, L Peipins (2005) Cancer internet search activity on a major search engine, United States 2001- 2003. *J Med Int Res* 7
9. Chandukala SR, Dotson JP, Brazell JD, Allenby GM (2011) Bayesian analysis of hierarchical effects. *Mark Sci* 30(1):123–133
10. Choi H, H Varian (2009a) Predicting the present with Google trends. Google technical report
11. Choi H, H Varian (2009b) Predicting initial claims for unemployment benefits. Google technical report
12. Clarke DG (1976) Econometric measurement of the duration of advertising effect on sales. *J Mark Res* 13(4):345–357
13. Danaher PJ, Rossiter JR (2011) Comparing perceptions of marketing communication channels. *E J Mark* 45:6–42
14. Dellarocas C, Zhang X, Awad NF (2007) Exploring the value of online product reviews in forecasting sales: the case of motion pictures. *J Int Mark* 21(4):23–45
15. Darrat AF, Zhong M (2000) On testing the random-walk hypothesis: a model-comparison approach. *The Fin Rev* 35:105–124
16. Doyle P, Saunders J (1990) Multiproduct advertising budgeting. *Mark Sci* 9(2):97–113
17. Ettredge M, Gerdes J, Karuga G (2005) Using web-based search data to predict macroeconomic statistics. *Comm ACM* 48(11):87–92
18. Feinberg FM (2001) On continuous-time optimal advertising under S-shaped response. *Mgmt Sci* 47(11):1476–1487
19. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L (2009) Detecting influenza epidemics using search engine query data. *Nature* 457:1012–1014
20. Goel S, JM Hofman, S Lahaie, DM Penneck, DJ Watts (2010) What can search predict. WWW 2010 April 26-30 2010 Raleigh, North Carolina
21. Hendl I, Nevo A (2003) The post-promotion dip puzzle? What do the data have to say. *Quant Mark Econ* 1:409–424
22. Howard JA, JN Sheth (1969) The theory of buyer behavior. John Wiley & Sons NY

23. Lavidge RJ, Steiner GA (1961) A model for predictive measurement of advertising effectiveness. *J Mark Res* 25(6):59–62
24. Lodish LM, Abraham M, Kalmenson S, Livelsberger J, Lubetkin B, Richardson B, Stevens ME (1995) How T.V. Advertising works: a meta-analysis of 389 real world split cable T.V. Advertising Experiments. *J Mark Res* 32(2):125–139
25. Mayzlin D, J Shin (2010) Uninformative advertising as an invitation to search. Unpublished manuscript, Yale University
26. Naik PA, Mantrala MK, Sawyer AG (1998) Planning media schedule in the presence of dynamic advertising quality. *Mark Sci* 17(3):214–235
27. Neslin S, Stone LGS (1996) Consumer inventory sensitivity and post promotion dip. *Mark Lett* 7(1):77–94
28. Newman JW, Staelin R (1973) Information sources for durable goods. *J Adv Res* 13:19–29
29. Nijs VR, Dekimpe MG, Steenkamp JEM, Hanssens DM (2001) The category demand effects of price promotions. *Mark Sci* 23(4):596–610
30. Onishi H, Manchanda P (2012) Marketing activity, blogging and sales. *Int J Res Mark* 29(3):221–234
31. Petris G, Sonia P, Campagnoli P (2009) *Dynamic linear models with R*. Springer, NY
32. Rusmevichientong P, Salisbury JA, Truss LT, Van Roy B, Glynn PW (2006) Opportunities and challenges in using online preference data for vehicle pricing: a case study at General Motors. *J Rev Pri Mgmt* 5(1):45–61
33. Rutz OJ, Bucklin RE (2011) From generic to branded: a model of spillover in paid search advertising'. *J Mark Res* 48(1):87–102
34. Terui N, Ban M, Allenby GM (2011) The effect of media advertising on brand consideration and choice. *Mark Sci* 30(1):74–91