Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2014 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

2014

CUSTOMER JOURNEYS ON ONLINE PURCHASE: SEARCH ENGINE, SOCIAL MEDIA, AND THIRD-PARTY ADVERTISING

Wenjing Duan
The George Washington University, wduan@gwu.edu

Jie Zhang
University of Texas at Arlington, jiezhang@uta.edu

Follow this and additional works at: http://aisel.aisnet.org/pacis2014

CUSTOMER JOURNEYS ON ONLINE PURCHASE: SEARCH ENGINE, SOCIAL MEDIA, AND THIRD-PARTY ADVERTISING

Wenjing Duan, School of Business, The George Washington University, Washington D.C., USA, wduan@gwu.edu

Jie Zhang, College of Business Administration, University of Texas at Arlington, Arlington, TX, jiezhang@uta.edu

Abstract

As the technologies and better practices become broadly available, companies are moving more quickly from a single-click or search-only model toward greater sophisticated models of informing and influencing the customer online shopping journeys. This study scrutinizes the predictive relationship between three referral channels, search engine, social medial, and third-party advertising, and online consumer search and purchase. The results derived from vector autoregressive models suggest that the three channels have differential predictive relationship with sale measures. Such differential relationship is even more pronounced for the long-term, accumulative effects. The predictive power of the three channels is also considerably different in referring customers among competing online shopping websites. This study offers new insights for IT and marketing practitioners in respect to how different channels perform in order to optimize the media mix and overall performance.

Keywords: Online purchase; marketing attribution; online consumer behavior; social media; search engines; online advertising; clickstream data; vector auto regression.

1 INTRODUCTION

Herbert Simon (1970) was the first to articulate the concept of "attention economics", speculating that "a wealth of information creates a poverty of attention." As information technology continues to advance rapidly, the way users generate and obtain information has been fundamentally changed. Online shopping websites and storefronts currently targets visitors using many types of information such as purchase history, demographic characteristics, and how the visitors arrive at the online store such as through social media recommendations, search engines, display ads, or email promotions. Consumers may browse, search, and purchase products or services guided by different sources of information through various media channels and paths. Consumers are drowning in various informational cues but short of attention to make the best sense of that information (Anderson and Palma 2012; Ghose et al. 2013).

The Internet has fostered the growth of many new distribution and advertising channels, along with the merging of transaction and interaction-based sites. Companies make significant investments in online, mobile, and offline media and channels such as search engines, social media, referral channels, and third-party advertising, etc., to draw in customers' attention to their websites, hoping to ultimately convert by purchasing product and services. The "customer journey" is unlikely to be linear, which takes places across multiple sessions, sites, and devices. Even relatively low cost products are being scrutinized, compared, and purchased across multiple touch-points in various media platforms. There are literally dozens of paths involving different combinations of channels that ultimately lead to consumer purchases.

Both traditional industry practice and academic research have been using aggregate metrics, namely the last-click analysis, which simply credit the last touch point leading consumer conversion without considering the influence of various channels may play in the purchase trips. However, according to a recent Google Analytics survey report, only 14% of respondents consider last-click analysis to be "very effective," yet over 50% of them are still using last click measurement. Today, more and more online e-commerce websites are turning to more sophisticated marketing attribution to gain more insights into their success and failure. With the availability of large amount of data of customers specifying their interactions with different channels in their search and purchase journey, there is a fast growing interest for both academics and practitioners in studying how to attribute the appropriate recognition for the conversions and revenues to different channels as well as designing and implementing the right tools to take them beyond the last click. This study would take the first initiative to contribute to such investigation. (Chan et al. 2011)

In the past decade, there is an extensive body of research on consumer online search and purchase behavior, focusing on modelling consumer browsing and search, and predicting probability of purchase (e.g., Johnson et al. 2004, Mode and Fader 2004, Park and Fader 2004, Danaher 2007). The rapid growth of online advertising market particularly propelled through search engines, also ignited fast growing body of work on examining the impact of search engine advertising on consumer search and purchase behavior. In the meantime, the proliferation of social media networks not only provides substantial valuable platforms for online advertising, but also continuously supplies unprecedented amount of opinions and experiences from a large number of active online users. There are also ample recent studies focusing on investigating the effect of social media on consumer behavior, product choices, and market performances (Tirunillai and Tellis 2012, Luo et al. 2013, Ghose et al. 2013).

Extant research, nevertheless, predominately focuses on one source of information and/or one type of channel or websites. In this study, we aim to investigate the effect of multiple sources of reference channels, namely, social media, search engines, and online third-party advertising, examine their relative importance and interrelatedness on online consumer search and purchase behavior. Specifically, this study aims to answer the following research questions: (1) can the referring paths through search engine, social media, or online third-party advertising predict online consumer purchase? 2) what is the relative importance of those referring paths in predicting consumer purchase? (3) what are the dynamics of the relationship between the three referral channels and online consumer

purchase? and (4) How do the referral channels to competing websites affect the sales of the focal website?

The rest of this paper is organized as follows. In Section 2, we review the related literature. In Section 3, we describe our data and empirical setting, the model setup, specification tests, and estimation and identification. Section 4 presents our main findings. We conclude in Section 5.

2 LITERATURE REVIEW

Echoed in Lanham (2006) to Herbert Simon (1970)'s first academic articulation of "attention economic", websites in the Information (overload) Age are aggressively competing for consumers' attention. There are a few recent theoretical developments looking into how websites can optimize the advertising and multi-channel distribution strategy to compete for consumer attention. For example, White and Jain (2010) study the incentives for multiple ad-funded websites with differing technologies to show advertisement to force the visitors to pay attention. Anderson and Palma (2012) model multiple sectors competing for customer attention, with competition in price within each sector. Anderson and Palma (2013) identified multiple equilibria in advertisers' strategies of sending different amount of advertising messages.

Search engine advertising, as one of the newest yet quickly becoming the most dominant form of online advertising, has incurred increasing interest in academic research. The majority of the theoretical literature, e.g., Edelman et al. (2007) and Katona and Sarvary (2010), emphases on optimal keywords bidding strategy and mechanism design for search websites. Yao and Mela (2009) developed a dynamic model of advertisers' bidding strategy to structurally model the competition among advertisers for search keywords.

On the other hand, empirical research on search engine advertising has focused on the profit impact on advertisers' click-through and conversion rates, and ultimately online sales. Ghose and Yang (2009) quantify the relationship between different sponsored search metrics including click-through rates, conversion rates, cost per click, and advertisement ranking using a hierarchical Bayesian modeling framework. Yang and Ghose (2010) model the interrelationship between organic listings and sponsored search advertising. They found asymmetric positive interdependence between organic and paid listings. Rutz and Bucklin (2011) examined the spillover effects between generic and branded keywords. Their findings suggest that generic keyword searches affect branded keyword searches, but not the vice versa.

Social media in the form of User Generated Content and Word-of-mouth are also shown to influence sales and conversion. For example, Chevalier and Mayzlin (2006), Clemons et al. (2006), and Dellarocas et al. (2007) show the mean rating has a significant effect on sales, and Duan et al. (2008) and Liu (2006) show that the volume of online product reviews has a significant effect on sales. All of these studies focus on examining the effect of product reviews on sales of products on the same websites. However, the impact of social media websites like facebook.com on referring consumers to retail websites to purchase still needs to be studied. Our study will fill in this gap with our consumer online clickstream data.

A few other recent empirical studies examine the relationship of multiple advertising channels. Danaher et al. (2010) develop an optimal media selection method that determines the number of advertising impressions that should be purchased and used from each chosen website. Goldfarb and Tucker (2011) explore substitution patterns across advertising platforms and found online advertising substitutes for offline advertising. Such substitution effect is the strongest in markets with fewer customers. Chan et al. (2011) develop an integrated model to measure the lifetime value of customers acquired from Google search advertising. They show customers acquired through Google search advertising have a higher transaction rate than customers acquired through other channels. Chiou and Tucker (2012) find some evidence that allowing third-party sellers to use a trademark in search engine advertising reduce the click-through rate of the paid search ads, but the click-through rate on the unpaid links are greatly increased. Ghose et al. (2013) examine the economic impact of ranking and its interaction with social media on travel search engine revenue. Xu et al. (2014) proposes a mutually

exiting point process model for online advertising and conversion. They find that display advertisements are more likely to stimulate subsequent visits through other advertisement formats than incurring direct purchase. They show that commonly used conversion measure underestimates the conversion effect of display advertisements the most.

There has been a growing literature on investigating the Internet clickstream data, which primarily focuses on the depth, width, and dynamics of online consumer search behavior. Many studies in this area have focused on behavior for a single site (Bucklin and Sismeiro 2003, Moe 2002) or for a given store over time (Moe and Fader 2004). Other papers look into behavior across sites. John et al. (2004) model an individual's tendency to search across competing e-commerce website. They found more-active online shoppers tend also to search more sites. Park and Fader (2004) develop a stochastic timing model of cross-site visit behavior to make better inferences about individual browsing and search behavior at multiple sites. Montgomery et al. (2004) demonstrate how consumer search path information can be modeled using a dynamic multinomial probit model, which can then be used to make probabilistic assessments about future paths. Danaher (2007) develop a multivariate generalization of the negative binomial distribution that models the page views across multiple websites to predict Internet reach and frequency. Huang et al. (2009) uses clickstream data to examine the differing consumer behavior for search and experience goods. Using aggregate search data from Amazon.com, Kim et al. (2010) jointly estimate consumer information search and online demand for consumer durable good goods.

There are several important differences between this paper and previous related studies. First, to the best of our knowledge, enabled by the tremendously increasing availability of data, this study is the first to examine how various referring channels (customer touch points) contribute to conversions and its implications for optimal targeting and allocation of marketing investment. Second, this is also one of the first studies to look at multiple online advertising channels, investigate their relative importance and their interrelatedness on affecting online consumer behaviors. Third, our clickstream data sample not only includes multiple websites, but also spans longer time periods. Lastly, we employ a multivariate time series technique, i.e., the vector autoregressive model with exogenous covariates, which can model both the short-term and the enduring effect of different referring channels.

3   DATA AND METHODOLOGY

3.1   Data

In this study, we use the 2011 comScore Media Metrix dataset in the United States subscribed from the Wharton Research Data Services (www.wrds.upenn.edu). The comScore Web Behavior Database is the click-stream panel data covering the browsing habits of approximately 100,000 households whose Internet surfing and purchasing behavior was recorded over time. These panelists, a sample of representative users, had agreed to install special unobtrusive software on their computers that monitored their browsing activities. Upon recruitment, each household reports a number of demographic variables, including the household size, age, region, education and income level, race, presence of children, and the speed of the Internet connection. The sequence and timing of all URLs viewed by each panel member are recorded. The collected data contain information regarding what sites each individual user visits and when they visit. The data also include the precise day and time when each individual viewed a specific URL. Purchase is defined as any visit during which a purchase occurred. Those visits in which the user saw the "confirm-order" page of the store's website were identified as purchase visits.

We use the daily click-stream panel data during the year of 2011. Comprising over 200MB per day, the data quantity is enormous. We take a representative subset of the full dataset, as we now detail. We focus on the category of general online retail merchandisers and select the top six stores ranked by daily sales. The six websites are Amazon.com, Walmart.com, Target.com, Macys.com, Sears.com, and Jcpenny.com. We chose them because (1) these stores were relatively more frequently shopped online, which ensures that we can collect enough daily transactions for analysis; (2) The relatively

popular goods in these categories also increase the probability that consumers may obtain information on these goods from various resources.

Users come to the focal retail websites by following the links on the previously visited websites, which is defined as referral websites. We classify the referral websites into three types: search engine (e.g. Google.com, Yahoo.com, Bing.com), social media (e.g. Facebook.com, Youtube.com, Bizrate.com), and third-party (e.g. DoubleClick.net, Imdb.com, Comcast.com) websites. The top referral websites of each type and the frequencies of them referring users in our session data sample are shown in Table 1.

Search Engine		Social Media		Third-Party	
Google.com	22.72%	Facebook.com	0.66%	Walmart.com	1.20%
Yahoo.com	6.00%	Youtube.com	0.33%	Amzon.com	1.14%
Bing.com	3.04%	Craigslist.org	0.10%	Comcast.com	0.30%
AOL.com	1.24%	poptropica.com	0.10%	Target.com	0.27%
Ask.com	0.83%	Bizrate.com	0.09%	DoubleClick.net	0.28%
Mywebsearch.com	0.69%	Slickdeals.net	0.08%	Imdb.com	0.09%
Live.com	0.52%	Blogspot.com	0.08%	Webseal.com	0.09%
Search-results.com	0.44%	Xegen.com	0.07%	Zappos.com	0.09%
Nextag.com	0.29%	squidoo.com	0.06%	Jr.com	0.08%
A lot.com	0.11%	Smarter.com	0.04%	eBay.com	0.08%

Table 1.   Top Referral Websites in Each Category

The endogenous variables are defined and measured in the following ways:

- Daily sales amount (*Sales*) can be directly derived from the database by summing up all of the consumer spendings at the website within each day.
- Conversion rate (*Conversion*) measures the probability of a visitor to the website from whichever channel becoming a payer customer. It can be calculated by

$$Conversion\ rate\ = \frac{Number\ of\ purchases}{Number\ of\ Visits} \tag{1}$$

- Daily sales volume (*volume*) is the number of products sold by the website within each day.
- Referral path (*Path*) refers to the type of referral website through which a visit comes with. It includes search engine, social media, third-party advertising, or direct URL. We used three dummy variables to represent the first three *Search*, *Social*, and *Third-party*, respectively.
- Average daily duration (*Duration*) and Average pages views (*Pageview*) measure user engagement in browsing, searching and interacting with the website. They are calculated by taking the average of total time (minutes) spent and unique web pages viewed of all the web visits to the website on a given day.
- Rival Sales, Conversion, Duration and Pageview are the performance and activities measure of the competitors.

We control the following variables:

- Customer demographics such as average household age, household income, and household education.
- Google search trend to control overall activities at a store.

Table 2 provides the descriptive statistics of the key variables for each websites. Figures 1 and 2 shows the time-series plot of referral paths and sales and conversion for selected websites. There is substantial variation in the raw data. The figure shows a moderate relationship between referral paths

and sales measures, particularly with the sales. Section 3.2 outline the models that examine this relationship systematically.

Variables	Daily Sales (K$)	Daily Conversion (%)	Daily Volume	Search engine	Social media	Third-party ads	Duration (K min)	Page-view (K)	Google Search
Amazon	3.56 (2.90)	3.34 (0.81)	65.36 (31.90)	3563.28 (2900.6)	54.13 (24.33)	228.72 (78.61)	12.03 (3.64)	12.52 (3.77)	63.19 (12.15)
Walmart	1.04 (1.20)	1.95 (0.78)	90.37 (137.29)	182.01 (85.94)	7.90 (8.46)	56.36 (35.41)	4.89 (2.87)	4.99 (2.56)	23.17 (6.83)
Target	0.22 (0.33)	0.68 (0.48)	2.89 (2.96)	150.98 (64.89)	6.08 (6.71)	35.27 (11.05)	2.44 (1.26)	2.62 (1.30)	1.15 (0.72)
Macy's	0.26 (0.33)	1.29 (0.94)	2.43 (2.39)	61.36 (26.17)	3.59 (4.21)	9.86 (5.89)	1.32 (0.64)	1.98 (0.76)	4.71 (1.63)
Sears	0.22 (0.38)	0.96 (0.89)	1.57 (1.83)	57.40 (22.86)	2.47 (2.46)	13.95 (7.87)	0.95 (0.42)	0.99 (0.41)	6.98 (1.32)
JcPenney	0.31 (0.31)	1.79 (1.19)	4.23 (4.23)	66.96 (25.61)	1.61 (1.60)	11.06 (5.85)	1.50 (0.64)	2.09 (0.88)	5.61 (1.27)

Note: The table reports the mean and the standard deviation is in the parenthesis.
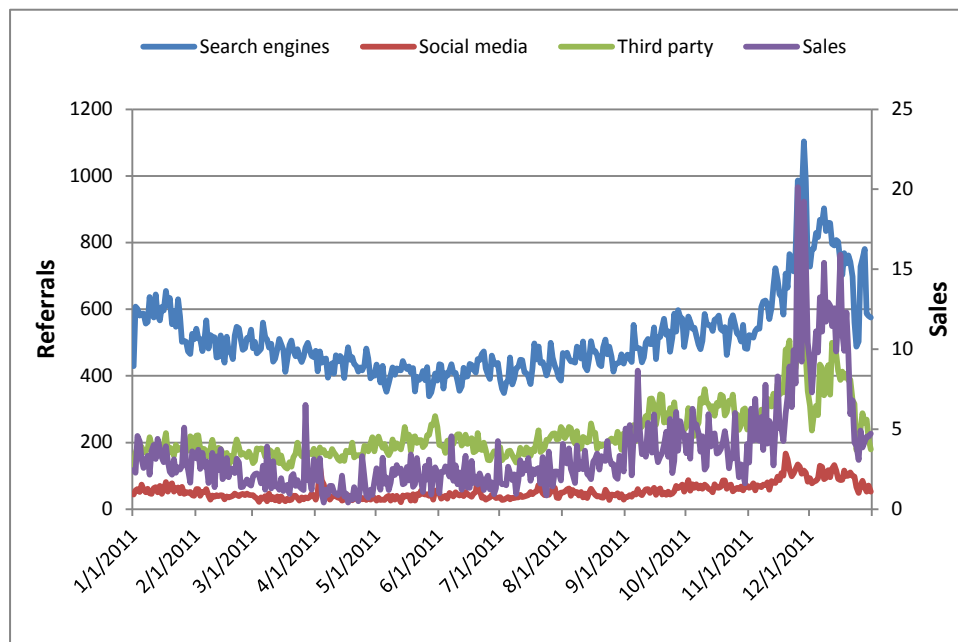
Table 2.   Descriptive Statistics of Key Variables

Figure 1.   Time-series plot of referral paths and sales (Amazon.com)
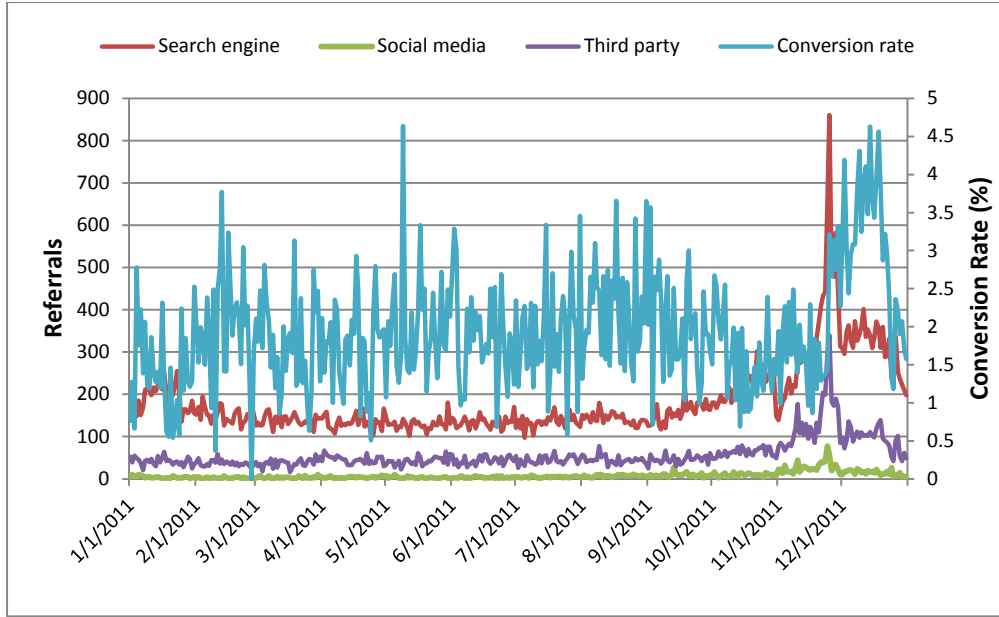
Figure 2.   *Time-series plot of referral paths and conversion (Walmart.com)*

3.2   VARX Model Specification and Estimation

We use VARX models (Dekimpe and Hanssens 1995) to capture dynamic interactions, competition effects, and feedback effects. VARX has several advantages over alternative models, because it can account for biases such as endogeneity, auto correlations, omitted variables, and reversed causality. Our empirical time-series analysis proceeds in the following steps that are applied to each website separately (Srinivasan et al. 2010). First, we estimate dynamic interactions among paths through search engine, social media and third-party ads, and online store sales and conversion rate using VARX models. Second, we quantify the relative influence of different paths on sales and conversion rate with the Generalized Forecast Error Variance Decomposition (GFEVD). Third, we track the long-term sales and conversion rate responses to a one-unit shock of the referrals from search engine, social media and third-party ads through Generalized Impulse Response Functions (GIRF).

Step 1: Vector-autoregressive Model Specification

We estimate an 11-equation VARX model per website, where endogenous variables are sales, conversion rate, volume, path, duration and pageview. We also have exogenous control variables, such as customer demographics and Google search trend. The VARX model is specified as shown in equation (2), where i ($i = 1, 2 \ldots 11$) represents a focal website, t represents time, j is lag length, and J is maximum lags. α_{ik} ($k = 1, 2\ldots 11$) denotes constant. δ_{ik}, $\phi_{ik,l}^{p}$, $\tau_{ik,s}$ ($k, l = 1, 2\ldots 11$, $s = 1, 2, 3, 4$) are coefficients: δ_{ik} reflects the seasonality effect, $\phi_{i1,3}^{j}$ is the coefficient of the search engine referral to website i j-day ago on the current sales, $\phi_{i3,1}^{j}$ reflects the feedback effect, and $\phi_{i3,3}^{p}$ reflects the reinforcing effect of the past search engine referral on the current one. ε_{kt} ($k = 1, 2\ldots 11$) represents white-noise residual. x_{ist} ($s = 1, 2, 3, 4$) represents the exogenous variables: average household age, household income, household education, and Google search trend. All the variables are on a daily basis for 365 periods in year 2011.

$$
\begin{bmatrix}
Sales_{it} \\
Conversion_{it} \\
Volume_{it} \\
Search_{it} \\
Social_{it} \\
Display_{it} \\
Duration_{it} \\
Pageview_{it} \\
Search_{-i,t} \\
Social_{-i,t} \\
Display_{-i,t}
\end{bmatrix}
=
$$

$$
\begin{bmatrix}
\alpha_{i1} + \delta_{i1} t \\
\alpha_{i2} + \delta_{i2} t \\
\alpha_{i3} + \delta_{i3} t \\
\alpha_{i4} + \delta_{i4} t \\
\alpha_{i5} + \delta_{i5} t \\
\alpha_{i6} + \delta_{i6} t \\
\alpha_{i7} + \delta_{i7} t \\
\alpha_{i8} + \delta_{i8} t \\
\alpha_{i9} + \delta_{i9} t \\
\alpha_{i10} + \delta_{i10} t \\
\alpha_{i11} + \delta_{i11} t
\end{bmatrix}
+ \sum_{p=1}^{P}
\begin{bmatrix}
\phi_{i1,1}^{p} & \cdots & \phi_{i1,11}^{p} \\
\phi_{i2,1}^{p} & \cdots & \phi_{i2,11}^{p} \\
\phi_{i3,1}^{p} & \cdots & \phi_{i3,11}^{p} \\
\phi_{i4,1}^{p} & \cdots & \phi_{i4,11}^{p} \\
\phi_{i5,1}^{p} & \cdots & \phi_{i5,11}^{p} \\
\phi_{i6,1}^{p} & \cdots & \phi_{i6,11}^{p} \\
\phi_{i7,1}^{p} & \cdots & \phi_{i7,11}^{p} \\
\phi_{i8,1}^{p} & \cdots & \phi_{i8,11}^{p} \\
\phi_{i9,1}^{p} & \cdots & \phi_{i9,11}^{p} \\
\phi_{i10,1}^{p} & \cdots & \phi_{i10,11}^{p} \\
\phi_{i11,1}^{p} & \cdots & \phi_{i11,11}^{p}
\end{bmatrix}
\begin{bmatrix}
Sales_{i,t-p} \\
Conversion_{i,t-p} \\
Volume_{i,t-p} \\
Search_{i,t-p} \\
Social_{i,t-p} \\
Display_{i,t-p} \\
Duration_{i,t-p} \\
Pageview_{i,t-p} \\
Search_{-i,t-p} \\
Social_{-i,t-p} \\
Display_{-i,t-p}
\end{bmatrix}
+
\begin{bmatrix}
\tau_{i1,1} & \cdots & \tau_{i1,4} \\
\tau_{i2,1} & \cdots & \tau_{i2,4} \\
\tau_{i3,1} & \cdots & \tau_{i3,4} \\
\tau_{i4,1} & \cdots & \tau_{i4,4} \\
\tau_{i5,1} & \cdots & \tau_{i5,4} \\
\tau_{i6,1} & \cdots & \tau_{i6,4} \\
\tau_{i7,1} & \cdots & \tau_{i7,4} \\
\tau_{i8,1} & \cdots & \tau_{i8,4} \\
\tau_{i9,1} & \cdots & \tau_{i9,4} \\
\tau_{i10,1} & \cdots & \tau_{i10,4} \\
\tau_{i11,4} & \cdots & \tau_{i11,4}
\end{bmatrix}
\begin{bmatrix}
x_{i1t} \\
x_{i2t} \\
x_{i3t} \\
x_{i4t}
\end{bmatrix}
+
\begin{bmatrix}
\varepsilon_{1t} \\
\varepsilon_{2t} \\
\varepsilon_{3t} \\
\varepsilon_{4t} \\
\varepsilon_{5t} \\
\varepsilon_{6t} \\
\varepsilon_{7t} \\
\varepsilon_{8t} \\
\varepsilon_{9t} \\
\varepsilon_{10t} \\
\varepsilon_{11t}
\end{bmatrix}
\qquad (2)
$$

Step 2: Generalized Forecast Error Variance Decomposition (GFEVD)

Based on VARX parameters, we derive GFEVD estimates to examine the following questions: to what extent do online referrals explain the variance of store sales, conversion and sales volume beyond fundamental controls? Like a dynamic R^2, GFEVD can gauge the relative power over time of shocks initiated by each endogenous variable in explaining those store sales measures, without assuming a causal ordering (Nijs et al. 2007). GFEVD estimates are derived from:

$$
\theta_{ij}^{g}(n) = \frac{\sum_{t=0}^{n} (\psi_{ij}^{g}(t))^2}{\sum_{t=0}^{n} \sum_{j=0}^{m} (\psi_{ij}^{g}(t))^2}, \quad i,j = 1,...,m. \qquad (3)
$$

The parameter $\psi_{ij}^{g}(t)$ is the value of a Generalized Impulse Response Function (GIRF) following a one-unit shock to variable i on variable j at time t (Pesaran and Shin 1998). GFEVD attributes 100% of the forecast error variance in the website sales and conversion metrics to past values of all endogenous variables. We focus on a more managerially interesting case, that is, the extent to which the channels via search engine, social media, or third-party ads explain the variance of sales and conversion. The relative importance of endogenous variables is established based on GFEVD values at 20 days (Luo 2013), which reduces sensitivity to short-term fluctuations. To establish the statistical significance of GFEVD estimates ($p = 0.05$), we obtain standard errors using Monte Carlo simulations with 1,000 runs.

Step 3: Generalized Impulse Response Functions (GIRF)

We also inspect GIRFs based on the estimated parameters of the full VARX model. The impulse response function estimates the net result of a "shock" to the channels via search engine, social media, or third-party ads on sales and conversion relative to their baselines (their expected points in the absence of the shock). Specifically, we measure cumulative sales and conversion responses to a one-unit shock with the simultaneous-shocking approach (Dekimpe and Hanssens 1999). Residual

variance-covariance matrix of equation (2) is used to derive a vector of expected instantaneous shock values. Standard errors are derived with Monte Carlo simulation 1,000 runs to test the statistical significance of parameters.

We derive the following summary statistics from each GIRF: (1) immediate impact on website sales and conversion, which is readily observable and applicable to managers; (2) total cumulative impact, which combines all effects across dust-settling periods and helps managers scrutinize whether and how much the paths of search engine, social media, and third-party ads contribute to sales and conversion in the long run.

4   EMPIRICAL FINDINGS

The process of estimating VARX models begins with unit-root tests to check whether variables are evolving or stationary, since GFEVD only allows comparable analyses with stationary variables. Stationarity implies that, although a shock to endogenous variables in VARX can cause fluctuations over time, its effects diminish ultimately. Then, endogenous variables revert back to the deterministic (mean + trend + seasonality) pattern without a permanent regime lift. The variance of stationary variables is finite and time-invariant. We conduct augmented Dickey-Fuller (ADF) tests to check stationarity (Dekimpe and Hanssens 1999). We take first difference to the variables of *Search Engine*, *Rival Search Engine*, and *Rival Social Media*. The ADF test statistics range from -2.88 to -22.16, all of which are significant ($p < 0.05$). Thus, the null hypothesis of a unit root can be rejected with a 95% confidence level, suggesting that the series are stationary and do not cointegrate in equilibrium (Hamilton 1994). Thus, we estimate VARX model with levels of endogenous variables. To report the findings, we average results across all websites of each industry (Srinivasan et al. 2010). The optimal lag length of the VARX model is 2 according to Schwartz's Bayesian Information Criterion (SBC) and final prediction error (FPE). Next we address our research questions and present empirical findings.

4.1   Test for Granger Causality

Following Tirunillai and Tellis (2012) and Luo et al. (2013), we conduct Granger Causality tests (Granger 1969). The results of the Granger Causality tests show that the three path metrics (search engine, social media and third-party ads) all have significant (p from 0.000 to 0.008) temporal-based causal relationships with the sales, conversion and volume of the focal online stores. With regard to competitive relationships, competitors' path metrics "Granger cause" the sales, conversion, and volume of the focal website for most websites ($p < 0.05$). These results confirm the significant temporal predictive relationship between the referral paths and online store sales measures.

4.2   Predictive Relationship between Referral Paths and Sales Measures

From the VARX models, impulse-response functions are derived that trace the over-time incremental effect of an unexpected change in the referral paths. Table 3 reports the immediate (the next day) and cumulative (in a relatively longer period e.g. 20 days) impulsive response elasticities. The magnitude of the elasticities reflects the changes in referral paths in response to one unit of unexpected changes in sales, conversion or volume. Figures 3 and 4 visually depicts these dynamic impulse response functions for the website Walmart.com.

As shown in Table 3, in the short run, all the three path metrics of the focal website have a significantly positive predictive relationship with the focal website's sales and volume (0.041 to 1.068, $p < 0.1$), but not significant with the conversion. Only referrals from search engines to the rival website have a significantly negative predictive relationship with the focal website's sales and volume (-0.040 and -0.445 $p < 0.1$). In the long-run, all the three path metrics have a significant positive predictive relationship with the focal website's sales, conversion and volume (0.186 to 8.894, $p < 0.1$), and all the path metrics of the rival website have a significantly negative predictive relationship with the focal website's sales, conversion and volume (-0.034 and -2.470 $p < 0.1$). In contrast, traffic metrics such as page views and duration do not show significantly predictive relationship with any of

the sales measures except that page views shows a weak predictive relationship with sales (immediate 0.036, $p < 0.1$ and cumulative 0.223, $p < 0.1$), and duration shows a weak immediate predictive relationship with sales (1.153, $p < 0.1$).

	Variables	Sales	Conversion	Volume
	Search engine	0.122***	0.035	1.597**
	Social media	0.056*	0.045	1.135*
	Third-party ads	0.041**	0.030	1.003*
	Duration	0.040	0.028	1.153*
Immediate	Pageviews	0.036*	-0.023	0.782
	Rival search	-0.040*	-0.013	-0.445*
	Rival social	-0.031	-0.011	-0.486
	Rival third	-0.008	-0.019	-0.360
	Search engine	0.700***	0.209***	8.894***
	Social media	0.521**	0.290***	5.836*
	Third-party ads	0.459***	0.186**	4.455**
	Duration	0.195	0.118	5.003
Cumulative	Pageviews	0.223*	-0.100	2.581
	Rival search	-0.128***	-0.068*	-1.482***
	Rival social	-0.099***	-0.034*	-1.724***
	Rival third	-0.129***	-0.054***	-2.470***

Note: *** p value < 0.001   ** p value < 0.05   * p value < 0.1

Table 3.   Responses of Online Store Sales and Conversion to Referral Paths
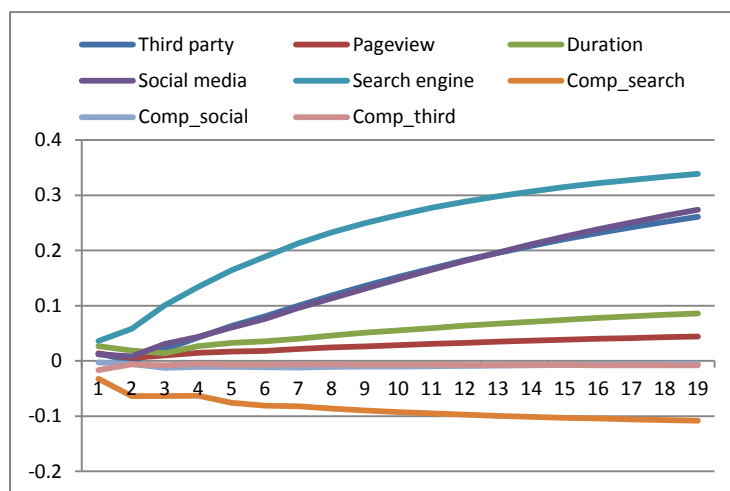

Figure 3.   Accumulated Responses of Online Store Sales to Referral Paths (Warlmart.com)
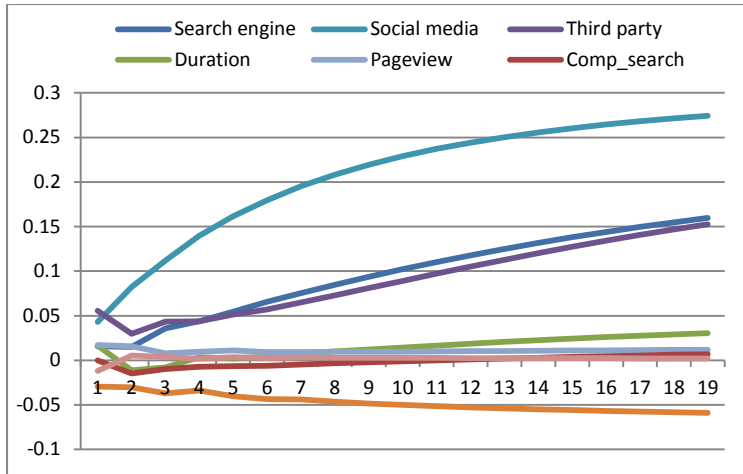
Figure 4.   Accumulated Responses of Online Store Conversion to Referral Paths (Warlmart.com)

4.3   Variance Decomposition Results

The variance decomposition results show the percentage of variances in the sales metrics explained by the path referral variables and the web traffic variables. The results in Table 4 show that the referral path metrics (search engine, social media and third-party ads) explain a higher percentage of variance in sales, conversion, and volume than the web traffic variables, or the path metrics of rival websites do.

Among the three referral path metrics, search engine explains the most of the variances (6.1%) in sales, social media the next (5.2%), and third-party ads the lowest (3.8%); while social media explains the most of the variances (1.65%) in conversion, higher than search engine (1.28%) and third-party ads (0.98%). Search engine referral of the rival websites also explains the most of the variances (0.92% and 0.65%) in sales and conversion of the focal firm, higher than social media (0.47% and 0.17%), and third-party ads (0.51% and 0.31%).

Variables	Sales	Conversion	Volume
Search engine	6.1	1.28	3.17
Social media	5.2	1.65	2.72
Third-party ads	3.83	0.98	3.69
Total Referral Paths	**15.1**	**3.91**	**9.58**
Duration	1.31	0.95	1.44
Pageviews	1.71	1.17	1.14
Total Traffic Variables	**3.02**	**2.12**	**2.59**
Rival search	0.92	0.65	0.91
Rival social	0.47	0.17	0.27
Rival third	0.51	0.31	0.37
Total Rival Paths	**1.90**	**1.13**	**1.55**

Table 4.   Variance Decomposition of Sales and Conversion Summarized by Industry

5   DISCUSSION AND CONCLUTION

The last decade has seen an overwhelming proliferation of channels, platforms, and technologies. E-commerce websites are challenged to stay on top of emerging innovations while attempting to optimize the opportunities they have already taken on. With the shift from product-centered thinking to customer-centered thinking in both research and practice (Chan et al. 2011), the means of reaching consumers has expanded enormously with lower cost of sending information as increasingly more and cheaper channels currently reach consumers. Many media and advertising channels catch the attention of prospective consumers. Such channels also interact with numerous product classes, which makes consumer attention span larger than ever before. Our research aims to contribute to the literature on more comprehensive understanding on the differential effect of online information and advertising channels, as well as offer deeper insights to managers on how to better targeting consumers through optimizing online advertising strategies.

This study aimed to look more deeply into customer journey on online search and purchase. Particularly this research was intended to investigate the predictive power of various online referral paths, their relative importance and interrelatedness, and dynamics of the relationship between different referral paths and consumer search and purchase behavior. The results suggest that different referral paths have differential predicting power on consumer purchase amount, volume, and the conversion rate. Such differences are even more prominent in examining the short and long term impact. For the short run (immediate, next day) effect, referral paths from search engines, social media, and third-party ads all have significant positive predictive relationship with online sales amount and volume, but not with the conversion. Only referral paths to competing websites from search engines have significant negative predictive relationship with both sales amount and volume. Referral paths to competing websites from social medial and third-party channels have no significant relationship with any of the sales measures for the focal website. For the cumulative and long run effect, all three referral path metrics have significant positive relationship with all the sale measures, while all the referral paths to competing websites has negative relationship with all the sale measures for the focal website. The variance decomposition results suggest that referral paths from search engines explain most of the variances in sales, while social media explain most of the variance in conversion. Referral paths to competing websites from search engines explain most of the variances of both sales and conversion for the focal website. Our results also show that the widely used conventional web traffic metrics such as page views and duration has no significant predictive relationship with sales.

This research contributes to the literature across IS and Marketing. To the best of our knowledge, our study is among the first to examine multiple referral and advertising channels, their relative importance, interrelatedness, and competition effect. In this sense, our study adds to the fast growing IS and Marketing literature on the impact of search engine, social media, and online third-party advertising. Prior IS and Marketing studies have demonstrated relationship between each individual advertising channel and website sales performance. Our study adds to those streams of research by offering a more integrative perspective and understanding in showing the differential predictive power of various channels. Managers should prioritize and allocate IT and marketing budget appropriately among various channels and platforms according to their ability to canal customers to the final point of sales. In addition, our time-series models can gauge both the short-term and the long-term and accumulative effect of various referral paths. Focusing merely on short-term effect would neglect the enduring influences and differences from different referral paths not only to the focal websites, but also to the competing websites, thus resulting in underestimating or misinterpreting their relationship with various sale metrics. The emerging online big data analytics research should therefore pay more attention to time-series modelling techniques and the long-term effects of advertising campaigns and media channels.

Our results suggest that number of customers referred from search engines is highly predictive of website sales, which suggests that search engine optimization is critical to generate the optimal amount of sales. Interestingly, our results indicate that customers referred from social media channels is the leading predictor of conversion. This suggests that customers are more likely to make the

purchase being exposed to the information (advertisement) on social media channels. This is particularly important for small players in the market, who is fighting for the visibility and every conversion is very valuable. Especially for new entrants in the market, how to allocate the limited It and marketing budget among different channels is crucial to gain sufficient early adopters in order to reach critical mass. Savvy merchants and managers understand that customers cannot be captured with just one message, one pretty picture, one nicely designed video, or even one perfectly placed advertisement. It is a complex process from carefully selecting and planting the seeds, cultivating it, and finally reaping the fruits. Hence it is crucial to take stock of all the factors that affected the consumer search and purchase results when it comes to giving credit to the various elements of marketing effort in different channels. This is also equally important to help making plans for future marketing campaign.

This study has several limitations that serve as paths for future research. First, our research focuses on one industry, which may limit the generalizability of the results. Future research could examine other leading industries to enrich the implications. Second, out research design cannot assure causality of the predicative relationship between different referral paths and sale metrics. Conducting field experiment could serve as one effective direction of future research. Third, this study focuses on investigating the differential predictive power of different referral channels. A fruitful extension of future research would explore why channels work separately and together by examining the way in which customers interact with different channels. Such an extension can be even more expanded by filling the gap between online and offline worlds.

References

Anderson, S. P., A. Palma. (2012). Competition for attention in the information (overload) age. Rand Journal of Economics, 43(1), 1-25.

Anderson, S. P., A. Palma. (2013). Shouting to Be Heard in Advertising. Management Science, 59(7), 1545-1556.

Chan, T.Y., C. Wu, Y. Xie. (2011). Measuring the Lifetime Value of Customers Acquired from Google Search Advertising. Marketing Science, 30(5), 837-850.

Chevalier, J. and Mayzlin, Dina (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," Journal of Marketing Research, 43, 3.

Chiou L., A., C. Tucker. (2012). How Does the Use of Trademarks by Third-Party Sellers Affect Online Search? Marketing Science, 31(5), 819-837.

Clemons, Eric K., Gao, Guodong, and Hitt, Lorin M. (2006), "When Online Reviews Meet Hyper Differentiation: A Study of the Craft Beer Industry," Journal of Management Information Systems, 23, 2, 149-171.

Dellarocas, Chrysanthos, Zhang, X. and Awad, Neveen (2007), "Exploring the Value of Online Reviews to Organizations: Implications for Revenue Forecasting and Planning," Journal of Interactive Marketing, 21, 4, 23-45.

Duan, W., Gu, Bin, and Whinston, Andrew (2008), "The Dynamics of Online Word-of-Mouth and Product Sales – An Empirical Investigation of the Movie Industry," Journal of Retailing, 84, 2, 233-242.

Danaher, P.J. (2010). Modeling Page Views Across Multiple Websites with an Application to Internet Reach and Frequency Prediction. Marketing Sci. 26(3) 422-437.

Danaher, P.J., J. Lee, L. Kerbache. (2010). Optimal Internet Media Selection. Marketing Science, 29(2), 336-347.

Dekimpe, M. and D. M. Hanssens. (1999). Sustained Spending and Persistent Response: A New Look at Long-term Marketing Profitability. Journal of Marketing Research, 36(4), 397-412.

Edelmen, B., M. Ostrovsky, M. Schwarz. (2007). Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. American Economic Review, 97(1), 242–259.

Ghose, A., S. Yang. (2009). An empirical analysis of search engine advertising: Sponsored search in electronic markets. Management Science, 55(10), 1605–1622.

Ghose, A., P. G. Ipeirotis, B. Li. (2013). Examining the Impact of Ranking on Consumer Behavior and Search Engine Revenue. Management Science, Forthcoming.

Ghose, A., P. G. Ipeirotis, B. Li. (2013). Surviving social media overload: predicting consumer footprints on product search engines. Working paper.

Goldfarb, A., C. Tucker. (2011). Online display advertising: Targeting and obtrusiveness. Marketing Science, 30(3), 389-404.

Granger, C. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica, 37(3), 424-438.

Hamilton, J. (1994). Time Series Analysis Princeton, N.J.: Princeton University Press.

Huang, P., Lurie, N. H., Mitra, S. (2009). Searching for Experience on the Web: An Empirical Examination of Consumer Behavior for Search and Experience Goods. Journal of Marketing, 73(2), 55-69.

Katona, Z., M. Sarvary. (2010). The race for sponsored links: Bidding patterns for search advertising. Marketing Science, 29(2), 199–215.

Kim, J.B., Albuquerque, P., Bronnenberg, B.J. (2010). Online Demand Under Limited Consumer Search. Marketing Science, 29(6), 1001–1023.

Lanham, R. (2006). The Economics of Attention: Style and Substance in the Age of Information. Chicago: University of Chicago Press.

Liu, Y. (2006), "Word-of-Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," Journal of Marketing, 70, 3, 74-89.

Luo, X., J. Zhang, W. Duan. (2013). Social media and firm equity value. Information Systems Research, 24(1), 146-163.

Johnson, E. J., W. Moe, P. Fader, S. Bellman, J. Lohse. (2004). On the depth and dynamics of online search behavior. Management Science, 50(3), 299–308.

Moe, Wendy W. (2003). Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational click-stream. Journal of Consumer Psychology, 13(1, 2), 29–40.

Moe, Wendy W., Peter F. Fader. (2004). Dynamic conversion behavior at e-commerce sites. Management Science, 50(3), 326–335.

Montgomery, A.L., S. Li, K. Srinivasan, J.C. Liechty. (2004). Modeling Online Browsing and Path Analysis Using Clickstream Data. Marketing Science, 23(4), 579-595.

Park, Young-Hoon, Peter S. Fader. (2004). Modeling browsing behavior at multiple websites. Marketing Science, 23(3), 280-303.

Rutz, O. J., R. E. Bucklin. (2011). From generic to branded: A model of spillover dynamics in paid search advertising. Journal of Marketing Research, 48 (1), 87–102.

Srinivasan, S., M. Vanhuele, and K. Pauwels. (2010). Mind-Set Metrics in Market Response Models: An Integrative Approach. Journal of Marketing Research, 47(4), 293–312.

Tirunillai S, Tellis G. (2012). Does chatter really matter? Dynamics of user-generated content and stock performance. Marketing Science, 31(2), 198–215.

White, A. and K. Jain. (2010). The Attention Economy of Search and Web Advertisement. Working paper.

Xu, Lizhen, J.A., Duan, and A. B. Whinston. 2014. Path to Purchase: A Mutually Exciting Point Process Model for Online Advertising and Conversion. Management Science, forthcoming.

Yang, S., Ghose, A. (2010). Analyzing the Relationship Between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence? Marketing Science, 29 (4), 602–623.